EfficientAD: Accurate Visual Anomaly Detection at Millisecond-Level Latencies

Kilian Batzner, Lars Heckler, Rebecca Konig
MVTec Software GmbH

WACV 2024

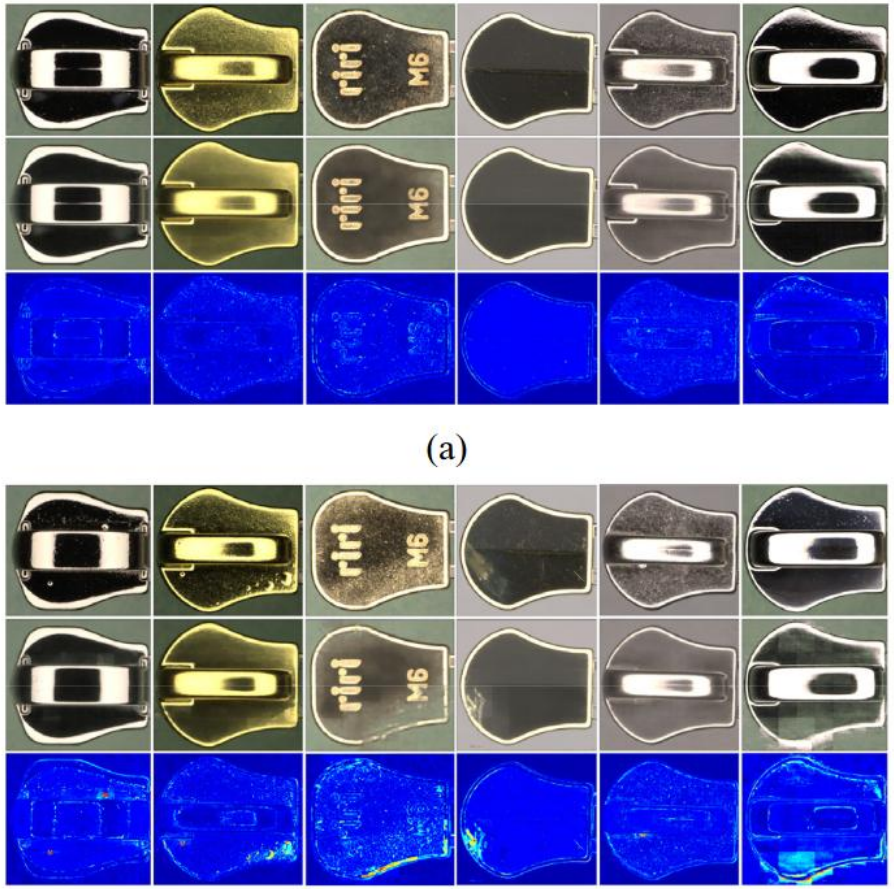
Introduction

Industrial Anomaly Detection Requirements:

- Computational Efficiency
- Economic Cost

Contribution:

- We substantially improve the state of the art for both the detection and the localization of anomalies on industrial benchmarks, at a latency of 2ms and a throughput of more than 600 images per second.
- We propose an efficient network architecture to speed up feature extraction by an order of magnitude.
- We introduce a training loss that significantly improves the anomaly detection performance of a student-teacher model without affecting its inference runtime.
- We achieve an efficient autoencoder-based detection of logical anomalies.



(a)

(b)

Figure 9: Anomaly detection results for (a) normal and (b) defected samples, top to bottom: input image, reconstructed image using AE, and anomaly map; left to right, sets #1 to #6 of zipper cursor dataset.

Patch Description Network

- four convolutional layers
- a feature vector generated by the PDN only depends on the pixels in its respective 33×33 patch
- obtain the features for an image of size 256×256 in less than 800 µs on an NVIDIA RTX A6000 GPU
- use the same pretrained features as PatchCore from a WideResNet-101.
- train the PDN on images from ImageNet by minimizing the MSE between its output and the features extracted from the pretrained network
- PDN ensures that an anomaly in one part of the image cannot trigger anomalous feature vectors in other distant parts.

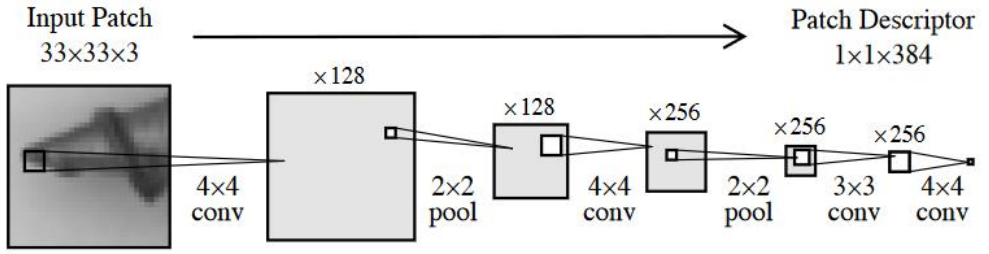


Figure 2. Patch description network (PDN) architecture of EfficientAD-S. Applying it to an image in a fully convolutional manner yields all features in a single forward pass.

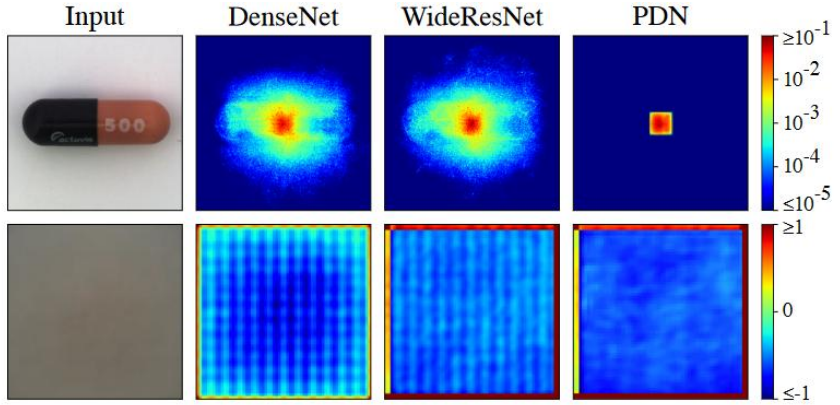


Figure 3. Upper row: absolute gradient of a single feature vector, located in the center of the output, with respect to each input pixel, averaged across input and output channels. Lower row: Average feature map of the first output channel across 1000 randomly chosen images from ImageNet [55]. The mean of these images is shown on the left. The feature maps of the DenseNet [25] and the WideResNet exhibit strong artifacts.

Lightweight Student-Teacher

- Teacher : distilled PDN
- Student : PDN

- We observe that in the standard S–T framework, increasing the number of training images can improve the student's ability to imitate the teacher on anomalies. This worsens the anomaly detection performance. At the same time, deliberately decreasing the number of training images can suppress important information about normal images. Our goal is to show the student enough data so that it can mimic the teacher sufficiently on normal images while avoiding generalization to anomalous images.

$$T(I) \in R^{C\times W\times H}$$
$$S(I) \in R^{C\times W\times H}$$

$$D_{c,w,h} = (T(I)_{c,w,h} - S(I)_{c,w,h})^2$$

$p_{hard}(0.999)$对应的值为d_{hard}

L_{hard}: $D_{c,w,h} \geq d_{hard}$

$$L_{ST} = L_{hard} + (CWH)^{-1}\sum_C ||S(P)_c||_F^2$$

sample a random image P from the pretraining dataset

During inference, the 2D anomaly score map $M \in R^{W\times H}$is given by $M_{w,h} = C^{-1}\sum_c D_{c,w,h}$

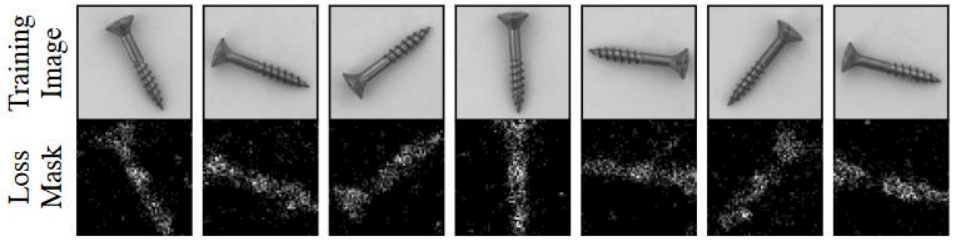


Figure 4. Randomly picked loss masks generated by the hard feature loss during training. The brightness of a mask pixel indicates how many of the dimensions of the respective feature vector were selected for backpropagation. The student network already mimics the teacher well on the background and thus focuses on learning the features of differently rotated screws.

Logical Anomaly Detection

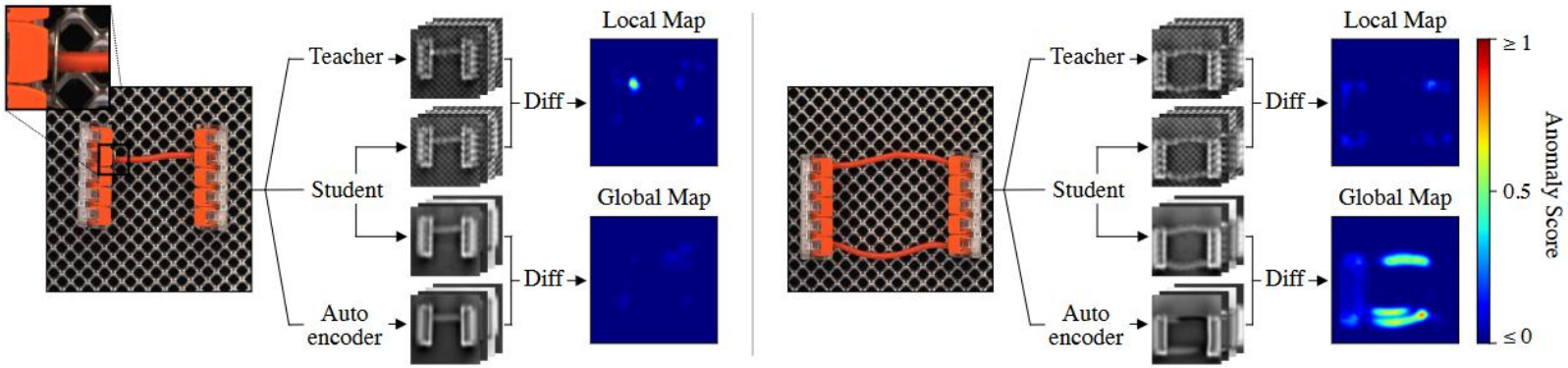


Figure 5. EfficientAD applied to two test images from MVTec LOCO. Normal input images contain a horizontal cable connecting the two splicing connectors at an arbitrary height. The anomaly on the left is a foreign object in the form of a small metal washer at the end of the cable. It is visible in the local anomaly map because the outputs of the student and the teacher differ. The logical anomaly on the right is the presence of a second cable. The autoencoder fails to reconstruct the two cables on the right in the feature space of the teacher. The student also predicts the output of the autoencoder in addition to that of the teacher. Because its receptive field is restricted to small patches of the image, it is not influenced by the presence of the additional red cable. This causes the outputs of the autoencoder and the student to differ. "Diff" refers to computing the element-wise squared difference between two collections of output feature maps and computing its average across feature maps. To obtain pixel anomaly scores, the anomaly maps are resized to match the input image using bilinear interpolation.

$$A(I) \in R^{C \times W \times H}$$
$$L_{AE} = (CWH)^{-1} \sum_C ||T(I)_c - A(I)_c||_F^2$$

$$S'(I) \in R^{C \times W \times H}$$
$$L_{STAE} = (CWH)^{-1} \sum_C ||A(I)_c - S'(I)_c||_F^2$$

Experiments

Method	Detect. AU-ROC	Segment. AU-PRO	Latency [ms]	Throughput [img / s]
GCAD	85.4	88.0	11	121
SimpleNet	87.9	74.4	12	194
S–T	88.4	89.7	75	16
FastFlow	90.0	86.5	17	120
DSR	90.8	78.6	17	104
PatchCore	91.1	80.9	32	76
$\text{PatchCore}_{\text{Ens}}$	92.1	80.7	148	13
AST	92.4	77.2	53	41
EfficientAD-S	95.4 (± 0.06)	92.5 (± 0.05)	**2.2** (± 0.01)	**614** (± 2)
EfficientAD-M	**96.0** (± 0.09)	**93.3** (± 0.04)	4.5 (± 0.01)	269 (± 1)

Table 1. Anomaly detection and anomaly localization performance in comparison to the latency and throughput. Each AU-ROC and AU-PRO percentage is an average of the mean AU-ROCs and mean AU-PROs, respectively, on MVTec AD, VisA, and MVTec LOCO. For EfficientAD, we report the mean and standard deviation of five runs.

Method	MAD	LOCO	VisA	Mean	LOCO Logic.	LOCO Struct.
GCAD	89.1	83.3	83.7	85.4	83.9	82.7
SimpleNet	98.2	77.6	87.9	87.9	71.5	83.7
S–T	93.2	77.4	94.6	88.4	66.5	88.3
FastFlow	96.9	79.2	93.9	90.0	75.5	82.9
DSR	98.1	82.6	91.8	90.8	75.0	90.2
PatchCore	98.7	80.3	94.3	91.1	75.8	84.8
$\text{PatchCore}_{\text{Ens}}$	**99.3**	79.4	97.7	92.1	71.0	87.7
AST	98.9	83.4	94.9	92.4	79.7	87.1
EfficientAD-S	98.8	90.0	97.5	95.4	85.8	94.1
EfficientAD-M	99.1	**90.7**	**98.1**	**96.0**	**86.8**	**94.7**

Table 2. Mean anomaly detection AU-ROC percentages per dataset collection (left) and on the logical and structural anomalies of MVTec LOCO (right). For EfficientAD, we report the mean of five runs. Performing method development solely on MVTec AD (MAD) becomes prone to overfitting design choices to the few remaining misclassified test images.

Experiments

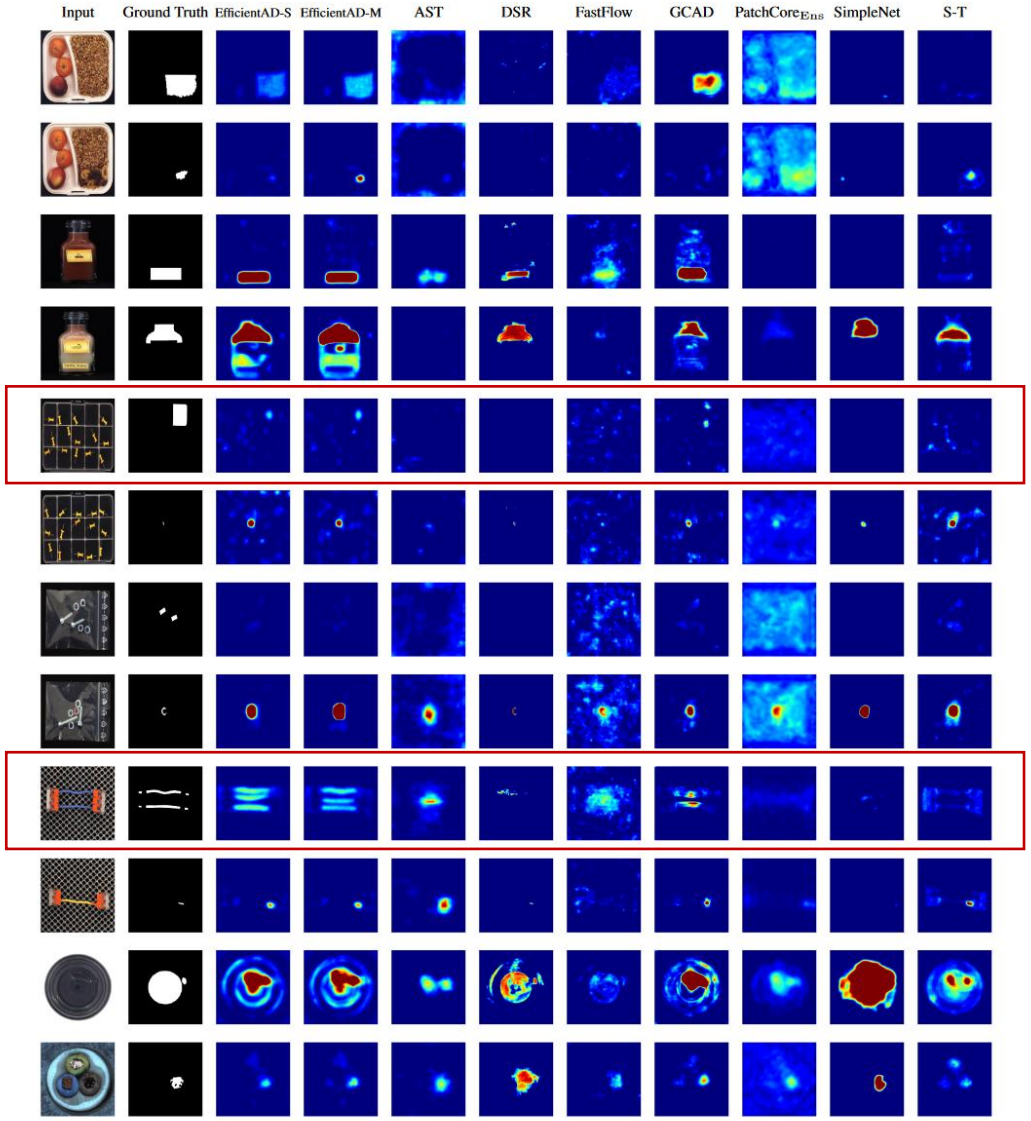


Figure 9. Anomaly maps on anomalous images from MVTec LOCO and MVTec AD. For MVTec LOCO, we show a logical anomaly (upper row) and a structural anomaly (lower row) for each scenario. The receptive field of AST's features is large enough to detect some logical anomalies, while PatchCore$_{\text{Ens}}$ and S–T struggle with logical anomalies.

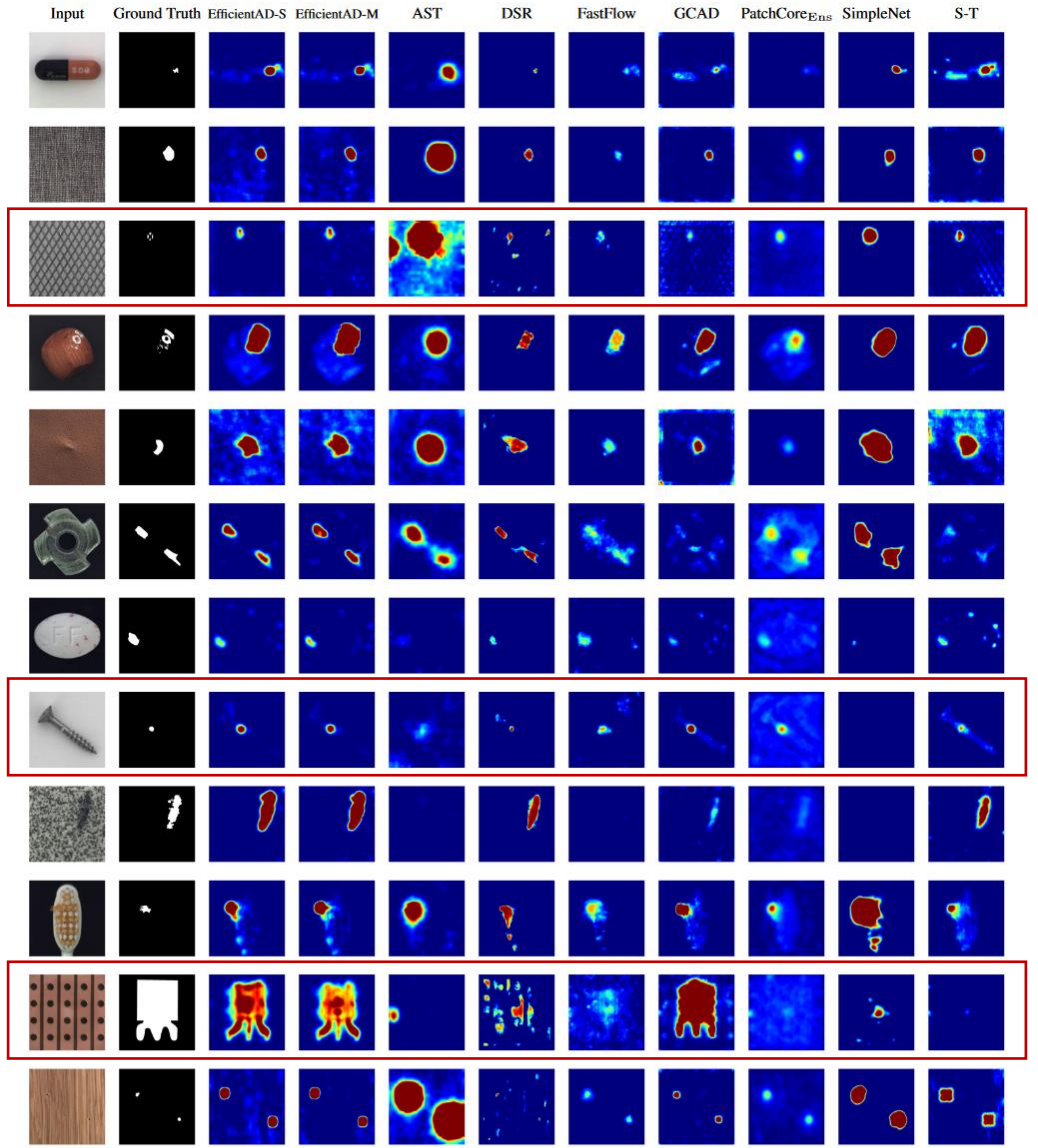


Figure 10. Anomaly maps on anomalous images from MVTec AD. Almost all anomalies are detected by every method, but the separability of pixel anomaly scores varies between methods. For example, PatchCore$_{\text{Ens}}$ detects the anomaly on the capsule in the first row but the pixel anomaly scores are in a similar range as the false positive detections in the background of the screw image.

Experiments

Method	Detect. AU-ROC	Segment. AU-PRO	Latency [ms]	Throughput [img / s]
GCAD	85.4	88.0	11	121
SimpleNet	87.9	74.4	12	194
S–T	88.4	89.7	75	16
FastFlow	90.0	86.5	17	120
DSR	90.8	78.6	17	104
PatchCore	91.1	80.9	32	76
$\text{PatchCore}_{\text{Ens}}$	92.1	80.7	148	13
AST	92.4	77.2	53	41
EfficientAD-S	95.4 (± 0.06)	92.5 (± 0.05)	**2.2** (± 0.01)	**614** (± 2)
EfficientAD-M	**96.0** (± 0.09)	**93.3** (± 0.04)	4.5 (± 0.01)	269 (± 1)

Table 1. Anomaly detection and anomaly localization performance in comparison to the latency and throughput. Each AU-ROC and AU-PRO percentage is an average of the mean AU-ROCs and mean AU-PROs, respectively, on MVTec AD, VisA, and MVTec LOCO. For EfficientAD, we report the mean and standard deviation of five runs.

	Detection AU-ROC	Diff.	Latency [ms]
PDN	93.2		2.2
↪ with map normalization	94.0	+ 0.8	2.2
↪ with hard feature loss	95.0	+ 1.0	2.2
↪ with pretraining penalty	95.4	+ 0.4	2.2
EfficientAD-S	95.4		2.2
EfficientAD-M	96.0	+ 0.6	4.5

Table 4. Cumulative ablation study in which techniques are gradually combined to form EfficientAD. Each AU-ROC percentage is an average of the mean AU-ROCs on MVTec AD, VisA, and MVTec LOCO.

	Detection AU-ROC	Diff.	Latency [ms]
EfficientAD-S	95.4		2.2
Without map normalization	94.7	- 0.7	2.2
Without hard feature loss	94.7	- 0.7	2.2
Without pretraining penalty	95.0	- 0.4	2.2

Table 5. Isolated ablation study in which techniques are separately removed from EfficientAD-S.

Experiments

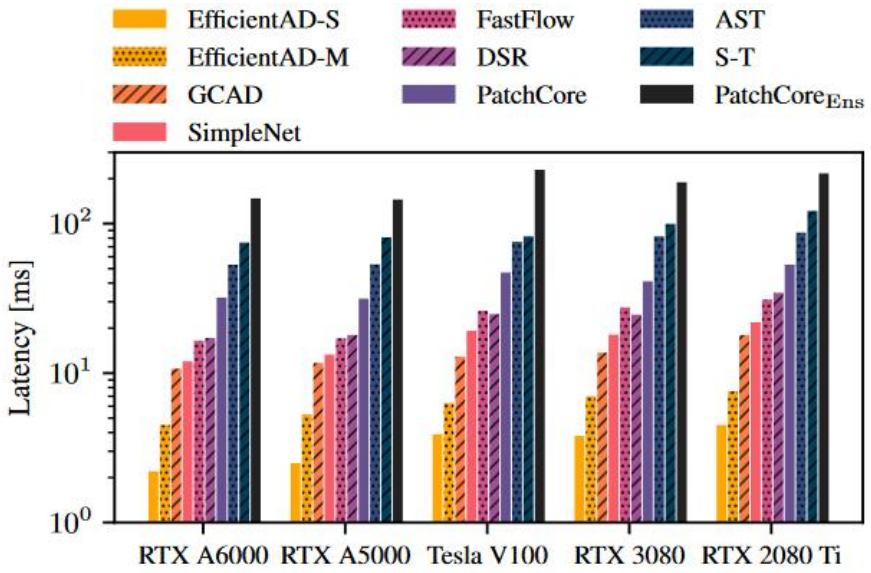
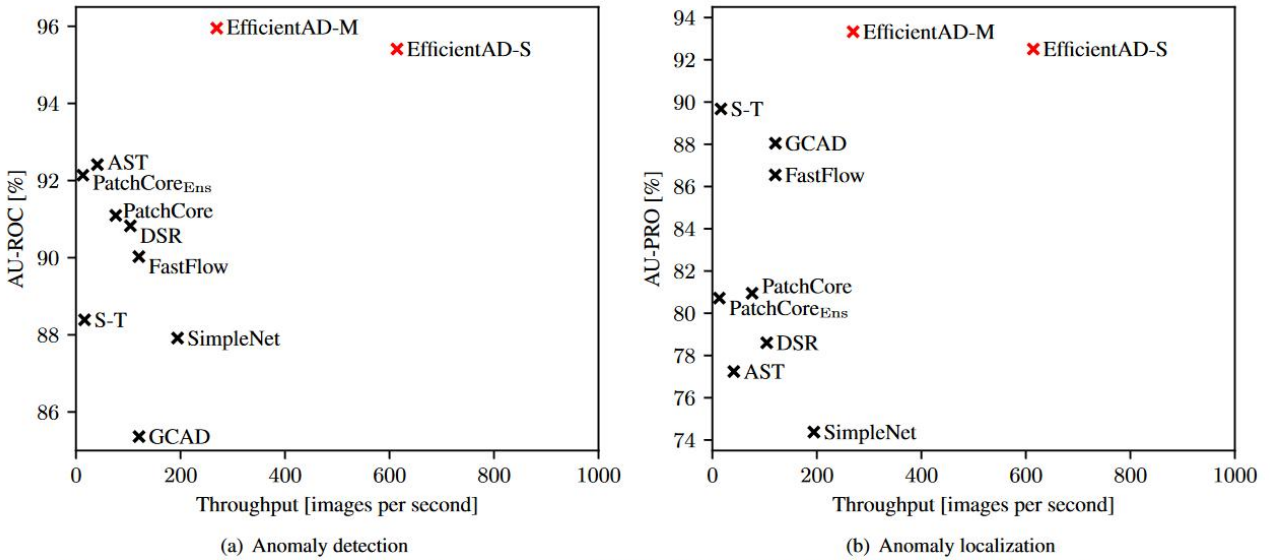


Figure 6. Latency per GPU. The ranking of methods is the same on each GPU, except for two cases in which DSR is slightly faster than FastFlow.



(a) Anomaly detection (b) Anomaly localization

Figure 8. Anomaly detection performance vs. throughput on an NVIDIA RTX A6000 GPU. We report the image-level anomaly detection performance on the left using the image-level AU-ROC. On the right, we report the anomaly localization performance using the pixel-level AU-PRO segmentation metric up to a FPR of 30 %. Each AU-ROC and AU-PRO value is an average of the values on MVTec AD, VisA, and MVTec LOCO. We measure the throughput using a batch size of 16.

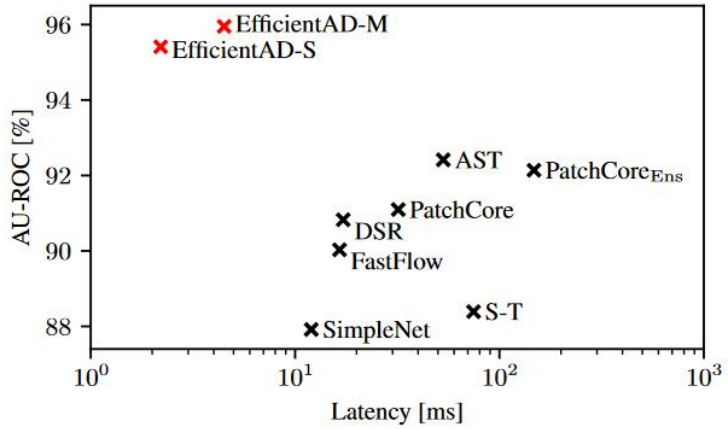


Figure 1. Anomaly detection performance vs. latency per image on an NVIDIA RTX A6000 GPU. Each AU-ROC value is an average of the image-level detection AU-ROC values on the MVTec AD [7, 9], VisA [74], and MVTec LOCO [8] dataset collections.

Thanks