

EfficientAD: Accurate Visual Anomaly Detection at Millisecond-Level Latencies

Kilian Batzner¹

Lars Heckler^{1,2}

Rebecca König¹

¹MVTec Software GmbH, ²Technical University of Munich

{kilian.batzner, lars.heckler, rebecca.koenig}@mvtec.com

Abstract

Detecting anomalies in images is an important task, especially in real-time computer vision applications. In this work, we focus on computational efficiency and propose a lightweight feature extractor that processes an image in less than a millisecond on a modern GPU. We then use a student–teacher approach to detect anomalous features. We train a student network to predict the extracted features of normal, i.e., anomaly-free training images. The detection of anomalies at test time is enabled by the student failing to predict their features. We propose a training loss that hinders the student from imitating the teacher feature extractor beyond the normal images. It allows us to drastically reduce the computational cost of the student–teacher model, while improving the detection of anomalous features. We furthermore address the detection of challenging logical anomalies that involve invalid combinations of normal local features, for example, a wrong ordering of objects. We detect these anomalies by efficiently incorporating an autoencoder that analyzes images globally. We evaluate our method, called EfficientAD, on 32 datasets from three industrial anomaly detection dataset collections. EfficientAD sets new standards for both the detection and the localization of anomalies. At a latency of two milliseconds and a throughput of six hundred images per second, it enables a fast handling of anomalies. Together with its low error rate, this makes it an economical solution for real-world applications and a fruitful basis for future research.

1. Introduction

In the past years, deep learning methods have continued to improve the state of the art across a wide range of computer vision applications. This progress has been accompanied by advances in making neural network architectures faster and more efficient [43, 59, 61, 63]. Modern classification architectures, for example, focus on characteristics such as latency, throughput, memory consumption, and the number of trainable parameters [32, 33, 54, 59, 60, 63]. This ensures that as networks become more capable, their com-

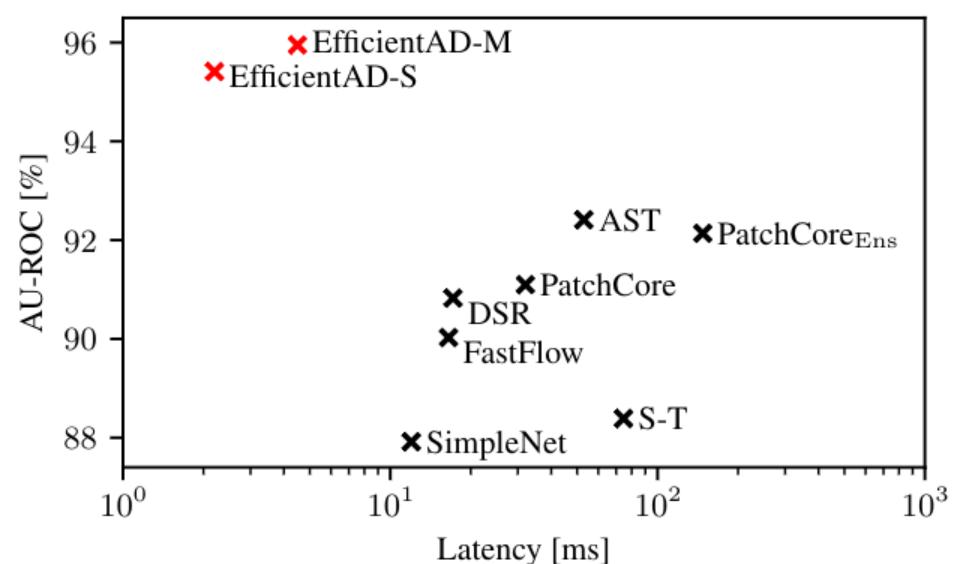


Figure 1. Anomaly detection performance vs. latency per image on an NVIDIA RTX A6000 GPU. Each AU-ROC value is an average of the image-level detection AU-ROC values on the MVTec AD [7, 9], VisA [69], and MVTec LOCO [8] dataset collections.

putational requirements remain suitable for real-world applications. The field of visual anomaly detection has also seen rapid progress in the recent past, especially on industrial anomaly detection benchmarks [7, 9, 47, 50]. State-of-the-art anomaly detection methods, however, often sacrifice computational efficiency for an increased anomaly detection performance. Common techniques are ensembling, the use of large backbones, and increasing the input image resolution to up to 768×768 pixels.

Real-world anomaly detection applications frequently put constraints on the computational requirements of a method. There are cases where detecting an anomaly too late can cause substantial economic damage, such as metal objects in a crop field entering the interior of a combine harvester. In other cases, even human health is at risk, for example, if a limb of a machine operator approaches a blade. Furthermore, industrial settings commonly involve strict runtime limits caused by high production rates [4]. Not adhering to these limits would decrease the production rate of the respective application and thus its economic viability. It is therefore essential to pay attention to the computational and economic cost of anomaly detection methods to keep them suitable for real-world applications.

In this work, we propose EfficientAD, a method that sets

EfficientAD：毫秒级延迟的准确视觉异常检测

基利安·巴茨纳 拉尔斯·赫克勒 丽贝卡·柯尼希

¹MVTec 软件有限公司，慕尼黑工业大学

{kilian.batzner, lars.heckler, rebecca.koenig}@mvtec.com

摘要

在图像中检测异常是一项重要任务，尤其是在实时计算机视觉应用中。在这项工作中，我们专注于计算效率，提出了一种轻量级特征提取器，可在现代 GPU 上在不到一毫秒的时间内处理图像。然后，我们使用学生-教师方法来检测异常特征。我们训练一个学生网络来预测正常（即无异常）训练图像的提取特征。通过学生无法预测异常特征，实现了测试时的异常检测。我们提出了一种训练损失函数，阻止学生在正常图像之外模仿教师特征提取器。这使我们能够大幅降低学生-教师模型的计算成本，同时改善异常特征的检测。我们还解决了具有挑战性的逻辑异常检测问题，这些异常涉及正常局部特征的无效组合，例如物体的错误排序。我们通过高效地结合一个全局分析图像的自动编码器来检测这些异常。我们在来自三个工业异常检测数据集集合的 32 个数据集上评估了我们的方法，称为 EfficientAD。EfficientAD 为异常的检测和定位设立了新标准。在 2 毫秒的延迟和每秒 600 张图像的吞吐量下，它能够快速处理异常。结合其低错误率，这使其成为实际应用的经济解决方案，也为未来研究提供了富有成效的基础。

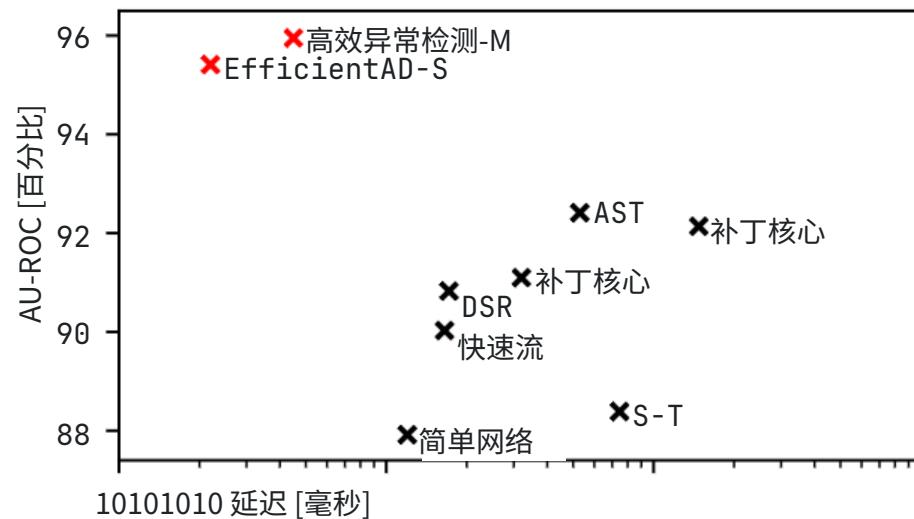


图 1. 异常检测性能与 NVIDIA RTX A6000 GPU 上每张图像的延迟对比。每个 AU-ROC 值是 MVTec AD [7, 9]、VisA [69] 和 MVTec LOCO [8] 数据集集合上图像级检测 AU-ROC 值的平均值。

算要求仍然适合实际应用。视觉异常检测领域在近期也取得了快速进展，特别是在工业异常检测基准测试上[7, 9, 47, 50]。然而，最先进的异常检测方法通常会牺牲计算效率来换取更高的异常检测性能。常见的技术包括集成、使用大型骨干网络，以及将输入图像分辨率提高到 768×768 像素。

实际的异常检测应用经常对方法的计算要求施加限制。在某些情况下，过晚检测到异常可能会造成巨大的经济损失，例如农田中的金属物体进入联合收割机内部。在其他情况下，甚至人类健康也面临风险，比如机器操作员的肢体接近刀片。此外，工业环境通常由于高生产率而存在严格的运行时限[4]。不遵守这些限制会降低相应应用的生产率，从而影响其经济可行性。因此，关注异常检测方法的计算和经济成本至关重要，以保持其适用于实际应用。

1. 引言 在过去几年中，深度学习方法持续改进了各种计算机视觉应用的最新技术水平。这一进展伴随着使神经网络架构更快速、更高效的进步[43, 59, 61, 63]。例如，现代分类架构关注延迟、吞吐量、内存消耗和可训练参数数量等特征[32, 33, 54, 59, 60, 63]。这确保了随着网络变得更加强大，它们的计算

在这项工作中，我们提出了 EfficientAD，这是一种设定

new standards for both the anomaly detection performance and the inference runtime, as shown in Figure 1. We first introduce an efficient network architecture for computing expressive features in less than a millisecond on a modern GPU. To detect anomalous features, we use a student–teacher approach [10, 50, 62]. We train a student network to predict the features computed by a pretrained teacher network on normal, i.e., anomaly-free training images. Because the student is not trained on anomalous images, it generally fails to mimic the teacher on these. A large distance between the outputs of the teacher and the student thus enables the detection of anomalies at test time. To further increase this effect, Rudolph *et al.* [50] use *architectural asymmetry* between the teacher and the student. We instead propose *loss-induced asymmetry* in the form of a training loss that hinders the student from imitating the teacher beyond the normal images. This loss does not affect the computational cost at test time and does not restrict the architecture design. It allows us to use our efficient network architecture for both the student and the teacher, while improving the detection of anomalous features.

Identifying anomalous local features enables the detection of anomalies that are *structurally* different from the normal images, for example, contaminations or stains on manufactured products. A challenging problem, however, are violations of *logical* constraints regarding the position, size, arrangement, etc. of normal objects. To address this, EfficientAD includes an autoencoder that learns the logical constraints of training images and detects violations at test time. We show how to integrate the autoencoder efficiently with a student–teacher model. Furthermore, we present a method to improve the anomaly detection performance by calibrating the detection results of the autoencoder and the student–teacher model before combining their results.

Our contributions are summarized as follows:

- We substantially improve the state of the art for both the detection and the localization of anomalies on industrial benchmarks, at a latency of 2 ms 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 in comparison to the feature extractors used by recent methods [47, 50, 64].
- 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 and propose a method for a calibrated combination of the detection results with those of a student–teacher model.

2. Related Work

2.1. Anomaly Detection Tasks

Visual anomaly detection is a rapidly growing area of research with a diverse range of applications, including medical imaging [3, 18, 36], autonomous driving [13, 23, 30], and industrial inspection [7, 17, 40]. Applications often have specific characteristics, such as the availability of image sequences in surveillance datasets [29, 35, 67] or the different modalities of medical imaging datasets (MRI [5], CT [3], X-ray [26], etc.). This work focuses on detecting anomalies in RGB or gray-scale images without conditioning the prediction on a sequence of images. We use industrial anomaly detection datasets to benchmark our proposed method against existing ones.

The introduction of the MVTec AD dataset [7, 9] has catalyzed the development of methods for industrial applications. It comprises 15 separate inspection scenarios, each consisting of a training set and a test set. Each training set contains only normal images, for example, defect-free screws, while the test sets also contain anomalous images. This represents a frequent challenge in real-world applications where the types and possible locations of defects are unknown during the development of the anomaly detection system. Therefore, it is a challenging yet crucial requirement that methods perform well when trained only on normal images.

Recently, several new industrial anomaly detection datasets have been introduced [8, 11, 25, 27, 37, 69]. The Visual Anomaly (VisA) dataset [69] and the MVTec Logical Constraints (MVTec LOCO) dataset [8] follow the design of MVTec AD and comprise twelve and five anomaly detection scenarios, respectively. They contain anomalies that are empirically more challenging than those of MVTec AD. Furthermore, MVTec LOCO contains not only structural anomalies, such as stains or scratches, but also logical anomalies. These are violations of logical constraints, for example, a wrong ordering or a wrong combination of normal objects. We refer to MVTec AD, VisA, and MVTec LOCO as dataset collections, as each scenario is a separate dataset consisting of a training and a test set. All three provide pixel-precise defect segmentation masks for evaluating the anomaly localization performance of a method.

2.2. Anomaly Detection Methods

Traditional computer vision algorithms have been applied successfully to industrial anomaly detection tasks for several decades [58]. These algorithms commonly fulfill the requirement of processing an image within a few milliseconds. Bergmann *et al.* [7] evaluate some of these methods and find that they fail when requirements such as well-aligned objects are not met. Deep-learning-based methods have been shown to handle such cases more robustly [7, 8].

新的异常检测性能和推理运行时间标准，如图 1 所示。我们首先引入了一种高效的网络架构，可以在现代 GPU 上以不到一毫秒的时间计算出富有表现力的特征。为了检测异常特征，我们使用了学生-教师方法[10, 50, 62]。我们训练一个学生网络来预测预训练教师网络在正常（即无异常）训练图像上计算的特征。由于学生网络未在异常图像上训练，它通常无法在这些图像上模仿教师。因此，教师和学生输出之间的大距离使得在测试时能够检测异常。为了进一步增强这种效果，Rudolph 等人[50]在教师和学生之间使用了架构不对称。相反，我们提出了损失诱导的不对称，即一种训练损失，它阻碍学生在正常图像之外模仿教师。这种损失不会影响测试时的计算成本，也不会限制架构设计。它允许我们为学生和教师都使用我们的高效网络架构，同时改善异常特征的检测。

识别异常的局部特征可以检测出与正常图像在结构上不同的异常，例如制造产品上的污染或污渍。然而，一个具有挑战性的问题是违反了正常物体的位置、大小、排列等方面逻辑约束。为了解决这个问题，EfficientAD 包含了一个自动编码器，它学习训练图像的逻辑约束，并在测试时检测违规情况。我们展示了如何将自动编码器有效地与学生-教师模型集成。此外，我们提出了一种方法，通过在组合结果之前校准自动编码器和学生-教师模型的检测结果来提高异常检测性能。

我们的贡献总结如下：

- 我们在工业基准测试中大幅提高了异常检测和定位的最新水平，延迟为 2 毫秒，吞吐量超过每秒 600 张图像。
- 我们提出了一种高效的网络架构，与最近方法[47, 50, 64]使用的特征提取器相比，将特征提取速度提高了一个数量级。
- 我们引入了一种训练损失，显著改善了学生-教师模型的异常检测性能，而不影响其推理运行时间。
- 我们实现了基于自动编码器的逻辑异常高效检测，并提出了一种方法，将检测结果与学生-教师模型的结果进行校准组合。

2. 相关工作

2.1. 异常检测任务

视觉异常检测是一个快速发展的研究领域，具有广泛的应用，包括医学成像[3, 18, 36]、自动驾驶[13, 23, 30]和工业检测[7, 17, 40]。应用通常具有特定的特征，如监控数据集中可用的图像序列[29, 35, 67]或医学成像数据集的不同模态(MRI [5]、CT [3]、X 射线[26]等)。本工作专注于在 RGB 或灰度图像中检测异常，而不依赖于图像序列进行预测。我们使用工业异常检测数据集来对比我们提出的方法与现有方法的性能。MVTec AD 数据集[7,9]的引入推动了工业应用方法的发展。它包含 15 个独立的检查场景，每个场景都由一个训练集和一个测试集组成。每个训练集只包含正常图像，例如无缺陷的螺丝，而测试集还包含异常图像。这代表了现实世界应用中的一个常见挑战，即在开发异常检测系统时，缺陷的类型和可能位置是未知的。因此，方法在仅使用正常图像进行训练时能够表现良好是一个具有挑战性但至关重要的要求。最近，几个新的工业异常检测数据集被引入[8, 11, 25, 27, 37, 69]。视觉异常(VisA)数据集[69]和 MVTec 逻辑约束(MVTec LOCO)数据集[8]遵循 MVTec AD 的设计，分别包含十二个和五个异常检测场景。它们包含的异常在经验上比 MVTec AD 的更具挑战性。此外，MVTec LOCO 不仅包含结构性异常，如污渍或划痕，还包含逻辑异常。这些是对逻辑约束的违反，例如错误的顺序或正常物体的错误组合。我们将 MVTec AD、VisA 和 MVTec LOCO 称为数据集集合，因为每个场景都是由训练集和测试集组成的独立数据集。这三个数据集都提供像素精确的缺陷分割掩码，用于评估方法的异常定位性能。

2.2. 异常检测方法 传统的计算机视觉算法已经成功应用于工业异常检测任务几十年了[58]。这些算法通常能够在几毫秒内处理一张图像。Bergmann 等人[7]评估了其中一些方法，发现当不能满足诸如物体对齐良好等要求时，这些方法会失效。基于深度学习的方法已被证明能更稳健地处理这些情况[7, 8]。

A successful approach in the recent past has been to apply outlier detection and density estimation methods in the feature space of a pretrained and frozen convolutional neural network (CNN). If feature vectors can be mapped to input pixels, assigning their outlier scores to the respective pixels yields a 2D anomaly map of pixel anomaly scores. Common methods include multivariate Gaussian distributions [15, 28, 45], Gaussian Mixture Models [37, 68], Normalizing Flows [21, 44, 48, 49, 64], and the k-Nearest Neighbor (kNN) algorithm [14, 38, 39, 47]. A runtime bottleneck for kNN-based methods is the search for nearest neighbors during inference. With PatchCore [47], Roth *et al.* therefore perform kNN on a reduced database of clustered feature vectors. They achieve state-of-the-art anomaly detection results on MVTec AD. In our experiments, we include PatchCore and FastFlow [64], a recent Normalizing-Flow-based method with a comparatively low inference runtime.

Bergmann *et al.* [10] propose a student–teacher (S–T) framework for anomaly detection, in which the teacher is a pretrained frozen CNN. They train student networks to mimic the output of the teacher on the training images. Because the students have not seen anomalous images during training, they generally fail to predict the teacher’s output on these images, which enables anomaly detection. Various modifications of S–T have been proposed [50, 53, 62]. Rudolph *et al.* [50] reach a competitive anomaly detection performance on MVTec AD by restricting the teacher to be an invertible neural network. We compare our method to their Asymmetric Student Teacher (AST) approach and to the original S–T method [10].

Generative models such as autoencoders [6, 8, 12, 19, 31, 41, 52] and GANs [2, 20, 42, 55, 56] have been used extensively for anomaly detection. Recent autoencoder-based methods rely on accurate reconstructions of normal images and inaccurate reconstructions of anomalous images [8, 12, 19, 41]. This enables detecting anomalies by comparing the reconstruction to the input image. A common problem are false-positive detections caused by inaccurate reconstructions of normal images, e.g., blurry reconstructions. To avoid this, GCAD [8] lets an autoencoder reconstruct images in the feature space of a pretrained network. Another recent reconstruction-based method is DSR [66], which uses the latent space of a pretrained autoencoder and generates synthetic anomalies in it. Similarly, the recently proposed SimpleNet [34] generates synthetic anomalies in a pretrained feature space to train a discriminator network for detecting anomalous features. In our experiments, we include GCAD, DSR, and SimpleNet.

3. Method

We describe the components of EfficientAD in the following subsections. It begins with the efficient extraction of features from a pretrained neural network in Sec. 3.1. We

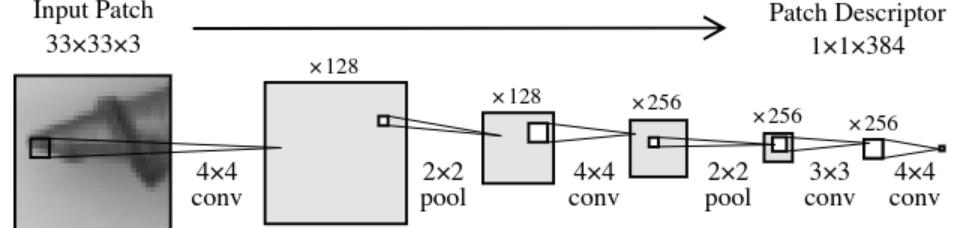


Figure 2. Patch description network (PDN) architecture of EfficientAD-S. Applying it to an image in a fully convolutional manner yields all features in a single forward pass.

detect anomalous features at test time using a lightweight student–teacher model, as described in Sec. 3.2. A key challenge is to achieve a competitive anomaly detection performance while keeping the overall runtime low. To this end, we introduce a novel loss function for the training of a student–teacher model. In Sec. 3.3, we explain how to efficiently detect logical anomalies with an autoencoder-based approach. Finally, we provide a solution for calibrating and combining the detection results of the autoencoder with those of the student–teacher model in Sec. 3.4.

3.1. Efficient Patch Descriptors

Recent anomaly detection methods commonly use the features of a deep pretrained network, such as a WideResNet-101 [47, 65]. We use a network with a drastically reduced depth as a feature extractor. It consists of only four convolutional layers and is visualized in Figure 2. Each output neuron has a receptive field of 33×33 pixels and thus each output feature vector describes a 33×33 patch. Due to this clear correspondence, we refer to the network as a patch description network (PDN). The PDN is fully convolutional and can be applied to an image of variable size to generate all feature vectors in a single forward pass.

The S–T method [10] also uses features from networks with only few convolutional layers. The computational cost of these networks is nevertheless high because of the lack of downsampling in convolutional and pooling layers. The number of parameters of the networks used by S–T is comparably low (between 1.6 and 2.7 million per network). Yet, executing a single network takes longer and requires more memory in our experiments than a U-Net [46] with 31 million parameters, an architecture used by the GCAD method [8]. This demonstrates how the number of parameters can be a misleading proxy metric for the latency, throughput, and memory footprint of a method. Modern classification architectures typically perform downsampling early to reduce the size of feature maps and thus the runtime and memory requirements [22]. We implement this in our PDN via strided average-pooling layers after the first and the second convolutional layer. With the proposed PDN, we are able to obtain the features for an image of size 256×256 in less than 800 μ s on an NVIDIA RTX A6000 GPU.

近年来，一种成功的方法是在预训练和冻结的卷积神经网络（CNN）的特征空间中应用异常值检测和密度估计方法。如果特征向量可以映射到输入像素，将其异常分数分配给相应的像素可以得到一个 2D 异常图，显示像素异常分数。常见的方法包括多变量高斯分布[15, 28, 45]、高斯混合模型[37, 68]、归一化流[21, 44, 48, 49, 64]和 k 近邻（kNN）算法[14, 38, 39, 47]。对于基于 kNN 的方法，推理过程中搜索最近邻是一个运行时瓶颈。因此，Roth 等人在 PatchCore[47]中对聚类后的特征向量进行 kNN 操作。他们在 MVTec AD 上实现了最先进的异常检测结果。在我们的实验中，我们包括了 PatchCore 和 FastFlow[64]，后者是一种基于归一化流的最新方法，具有相对较低的推理运行时间。

Bergmann 等人[10]提出了一个用于异常检测的学生-教师（S-T）框架，其中教师是一个预训练的冻结 CNN。他们训练学生网络来模仿教师在训练图像上的输出。由于学生在训练期间没有见过异常图像，他们通常无法预测教师在这些图像上的输出，从而实现异常检测。S-T 的各种修改版本已被提出[50, 53, 62]。Rudolph 等人[50]通过将教师限制为可逆神经网络，在 MVTec AD 上达到了具有竞争力的异常检测性能。我们将我们的方法与他们的非对称学生教师（AST）方法和原始 S-T 方法[10]进行比较。

生成模型如自编码器[6, 8, 12, 19, 31, 41, 52]和 GANs [2, 20, 42, 55, 56]已被广泛用于异常检测。最近的基于自编码器的方法依赖于对正常图像的准确重建和对异常图像的不准确重建[8, 12, 19, 41]。这使得通过比较重建结果和输入图像来检测异常成为可能。一个常见问题是由于正常图像的不准确重建（例如模糊重建）导致的误报。为了避免这种情况，GCAD [8]让自编码器在预训练网络的特征空间中重建图像。另一个最近的基于重建的方法是 DSR [66]，它使用预训练自编码器的潜在空间并在其中生成合成异常。类似地，最近提出的 SimpleNet [34]在预训练特征空间中生成合成异常，以训练用于检测异常特征的判别器网络。在我们的实验中，我们包括了 GCAD、DSR 和 SimpleNet。

3. 方法 我们在以下小节中描述 EfficientAD 的组成部分。它始于 3.1 节中从预训练神经网络高效提取特征。我们

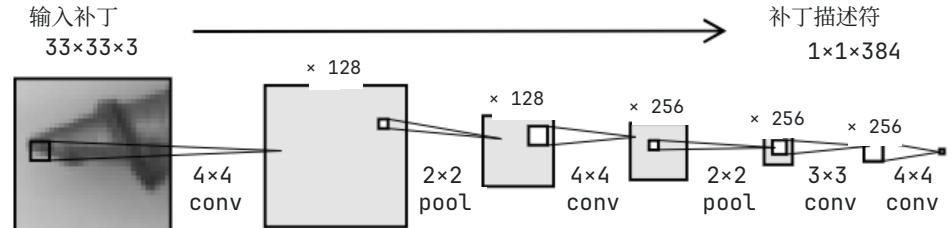


图 2. EfficientAD-S 的补丁描述网络（PDN）架构。以全卷积方式将其应用于图像可在单次前向传播中生成所有特征。

使用轻量级学生-教师模型在测试时检测异常特征，如第 3.2 节所述。关键挑战是在保持整体运行时间低的同时实现具有竞争力的异常检测性能。为此，我们引入了一种新的损失函数用于学生-教师模型的训练。在第 3.3 节中，我们解释如何使用基于自编码器的方法有效地检测逻辑异常。最后，我们在第 3.4 节中提供了一种校准和结合自编码器检测结果与学生-教师模型结果的解决方案。

3.1. 高效补丁描述符 最近的异常检测方法通常使用深度预训练网络的特征，如 WideResNet-101 [47, 65]。我们使用深度大幅减少的网络作为特征提取器。它仅由四个卷积层组成，如图 2 所示。每个输出神经元的感受野为 33×33 像素，因此每个输出特征向量描述一个 33×33 的补丁。由于这种明确的对应关系，我们将该网络称为补丁描述网络（PDN）。PDN 是完全卷积的，可以应用于可变大小的图像，在单次前向传播中生成所有特征向量。

S-T 方法[10]也使用仅有少数卷积层的网络的特征。然而，由于卷积层和池化层缺乏下采样，这些网络的计算成本仍然很高。S-T 使用的网络参数数量相对较低（每个网络在 160 万到 270 万之间）。但在我们的实验中，执行单个网络所需的时间和内存比使用 3100 万参数的 U-Net[46]（GCAD 方法[8]使用的架构）更长、更多。这表明参数数量可能是一个具有误导性的指标，不能准确反映方法的延迟、吞吐量和内存占用。现代分类架构通常在早期进行下采样以减小特征图的大小，从而降低运行时间和内存需求[22]。我们在 PDN 中通过在第一和第二卷积层之后使用步进平均池化层来实现这一点。使用所提出的 PDN，我们能够在 NVIDIA RTX A6000 GPU 上在不到 800 微秒的时间内获得 256×256 大小图像的特征。

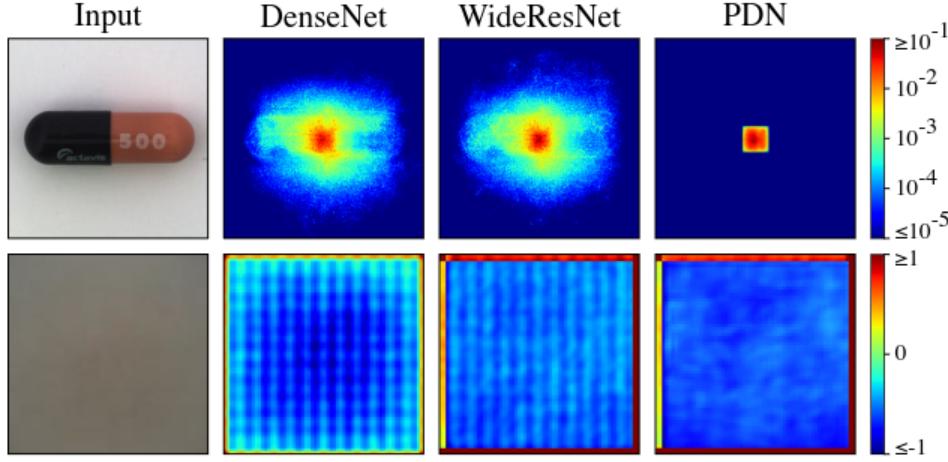


Figure 3. Upper row: absolute gradient of a single feature vector, located in the center of the output, with respect to each input pixel, averaged across input and output channels. Lower row: Average feature map of the first output channel across 1000 randomly chosen images from ImageNet [51]. The mean of these images is shown on the left. The feature maps of the DenseNet [24] and the WideResNet exhibit strong artifacts.

To make the PDN generate expressive features, we distill a deep pretrained classification network into it. For a controlled comparison, we use the same pretrained features as PatchCore [47] from a WideResNet-101. We train the PDN on images from ImageNet [51] by minimizing the mean squared difference between its output and the features extracted from the pretrained network. We provide the full list of training hyperparameters in the supplementary material. Besides higher efficiency, the PDN has another benefit in comparison to the deep networks used by recent methods. By design, a feature vector generated by the PDN only depends on the pixels in its respective 33×33 patch. The feature vectors of pretrained classifiers, on the other hand, exhibit long-range dependencies on other parts of the image. This is shown in Figure 3, using PatchCore’s feature extractors as an example. The well-defined receptive field of the PDN ensures that an anomaly in one part of the image cannot trigger anomalous feature vectors in other, distant parts, which would impair the localization of anomalies.

3.2. Lightweight Student–Teacher

For detecting anomalous feature vectors, we use a student–teacher (S–T) approach in which the teacher is given by our distilled PDN. Since we can execute the PDN in under a millisecond, we use its architecture for the student as well, resulting in a low overall latency. This lightweight student–teacher pair, however, lacks techniques used by previous methods to increase the anomaly detection performance: ensembling multiple teachers and students [10], using features from a pyramid of layers [62], and using architectural asymmetry between the student and the teacher network [50]. We therefore introduce a training loss that substantially improves the detection of anomalies without affecting the computational requirements at test time.

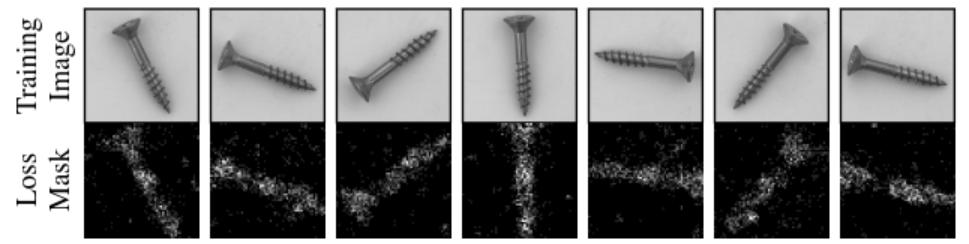


Figure 4. Randomly picked loss masks generated by the hard feature loss during training. The brightness of a mask pixel indicates how many of the dimensions of the respective feature vector were selected for backpropagation. The student network already mimics the teacher well on the background and thus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

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. Similar to Online Hard Example Mining [57], we therefore restrict the student’s loss to the most relevant parts of an image. These are the patches where the student currently mimics the teacher the least. We propose a hard feature loss, which only uses the output elements with the highest loss for backpropagation.

Formally, we apply a teacher T and a student S to a training image I , which yields $T(I) \in \mathbb{R}^{C \times W \times H}$ and $S(I) \in \mathbb{R}^{C \times W \times H}$. We compute the squared difference for each tuple (c, w, h) as $D_{c,w,h} = (T(I)_{c,w,h} - S(I)_{c,w,h})^2$. Based on a mining factor $p_{\text{hard}} \in [0, 1]$, we then compute the p_{hard} -quantile of the elements of D . Given the p_{hard} -quantile d_{hard} , we compute the training loss L_{hard} as the mean of all $D_{c,w,h} \geq d_{\text{hard}}$. Setting p_{hard} to zero would yield the original S–T loss. In our experiments, we set p_{hard} to 0.999, which corresponds to using, on average, ten percent of the values in each of the three dimensions of D for backpropagation. Figure 4 visualizes the effect of the hard feature loss for $p_{\text{hard}} = 0.999$. During inference, the 2D anomaly score map $M \in \mathbb{R}^{W \times H}$ is given by $M_{w,h} = C^{-1} \sum_c D_{c,w,h}$, i.e., by D averaged across channels. It assigns an anomaly score to each feature vector.

In addition to the hard feature loss, we use a loss penalty during training that further hinders the student from imitating the teacher on images that are not part of the normal training images. In the standard S–T framework, the teacher is pretrained on an image classification dataset, or it is a distilled version of such a pretrained network. The student is not trained on that pretraining dataset but only on the application’s normal images. We propose to also use the images from the teacher’s pretraining during the train-

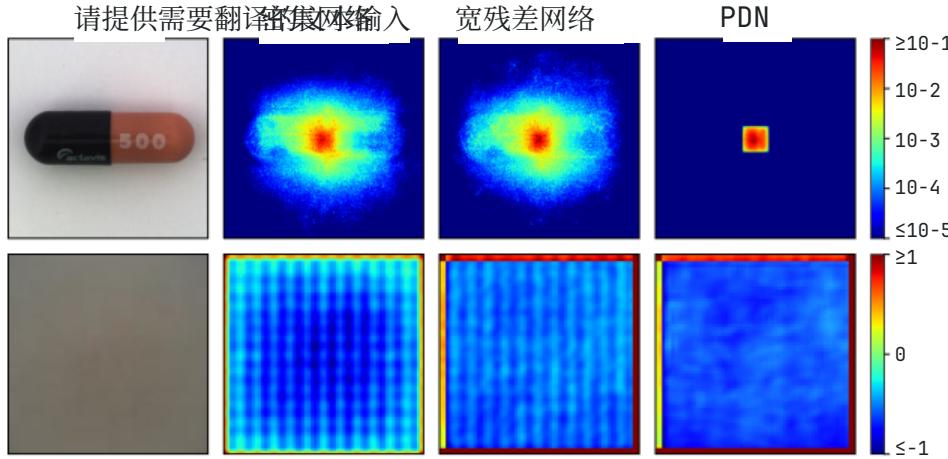


图 3。上排：单个特征向量相对于每个输入像素的绝对梯度，该特征向量位于输出的中心，在输入和输出通道之间取平均。下排：在 ImageNet [51] 中随机选择的 1000 张图像中，第一个输出通道的平均特征图。这些图像的平均值显示在左侧。DenseNet [24] 和 WideResNet 的特征图表现出强烈的伪影。

为了使 PDN 生成富有表现力的特征，我们将一个深度预训练分类网络蒸馏到其中。为了进行受控比较，我们使用与 PatchCore [47] 相同的来自 WideResNet-101 的预训练特征。我们通过最小化 PDN 输出与从预训练网络提取的特征之间的均方差，在 ImageNet [51] 的图像上训练 PDN。我们在补充材料中提供了完整的训练超参数列表。除了更高的效率外，与最近方法使用的深度网络相比，PDN 还有另一个优势。根据设计，PDN 生成的特征向量仅依赖于其各自 33×33 补丁中的像素。另一方面，预训练分类器的特征向量表现出对图像其他部分的长程依赖性。这在图 3 中有所展示，以 PatchCore 的特征提取器为例。PDN 明确定义的感受野确保了图像一部分的异常不会触发其他远处部分的异常特征向量，这将损害异常的定位。

3.2. 轻量级学生-教师 为了检测异常特征向量，我们使用学生-教师 (S-T) 方法，其中教师由我们的蒸馏 PDN 给出。由于我们可以在不到一毫秒的时间内执行 PDN，我们也为学生使用相同的架构，从而实现低整体延迟。然而，这种轻量级的学生-教师对缺乏先前方法用来提高异常检测性能的技术：集成多个教师和学生[10]，使用来自层级金字塔的特征 [62]，以及在学生和教师网络之间使用架构不对称[50]。因此，我们引入了一种训练损失，它在不影响测试时计算要求的情况下，大大提高了异常检测的性能。

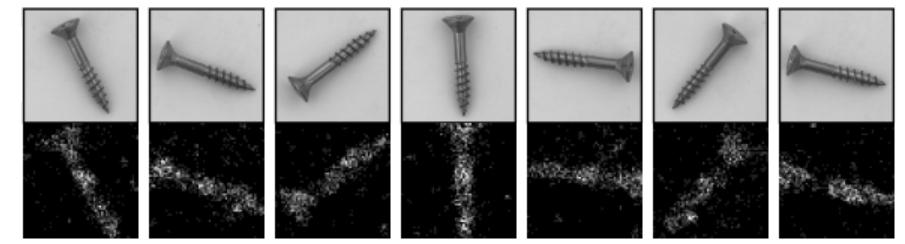


图 4. 训练期间由硬特征损失生成的随机选取的损失掩码。掩码像素的亮度表示相应特征向量的多少维度被选择用于反向传播。学生网络已经在背景上很好地模仿了教师，因此专注于学习不同旋转螺丝的特征。

我们观察到，在标准的 S-T 框架中，增加训练图像的数量可以提高学生模仿教师处理异常的能力。这会降低异常检测的性能。同时，故意减少训练图像的数量可能会抑制正常图像的重要信息。我们的目标是向学生展示足够的数据，使其能够充分模仿教师处理正常图像，同时避免泛化到异常图像。类似于在线硬样本挖掘[57]，我们因此将学生的损失限制在图像最相关的部分。这些是学生当前最不能模仿教师的区域。我们提出了一种硬特征损失，它只使用损失最高的输出元素进行反向传播。

形式上，我们将教师 T 和学生 S 应用于训练图像 I，得到 $T(I) \in R$ 和 $S(I) \in R$ 。我们计算每个元组 (c, w, h) 的平方差 $D = (T(I) - S(I))^2$ 。基于挖掘因子 $p \in [0, 1]$ ，我们然后计算 D 元素的 p 分位数。给定 p 分位数 d，我们计算训练损失 L 为所有 $D \geq d$ 的平均值。将 p 设为零将得到原始的 S-T 损失。在我们的实验中，我们将 p 设为 0.999，这相当于在 D 的三个维度中每个维度平均使用 10% 的值进行反向传播。图 4 可视化了 $p=0.999$ 时硬特征损失的效果。在推理过程中，2D 异常分数图 $M \in R$ 由 $M = C$ 给出

c_D^P ，即 D 在通道上的平均值。它为每个特征向量分配一个异常分数。

除了硬特征损失之外，我们在训练过程中还使用了一个损失惩罚，进一步阻止学生模型在非正常训练图像上模仿教师模型。在标准的学生-教师框架中，教师模型是在图像分类数据集上预训练的，或者是这种预训练网络的蒸馏版本。学生模型并不在该预训练数据集上训练，而只在应用程序的正常图像上训练。我们建议在训练期间也使用教师预训练的图像

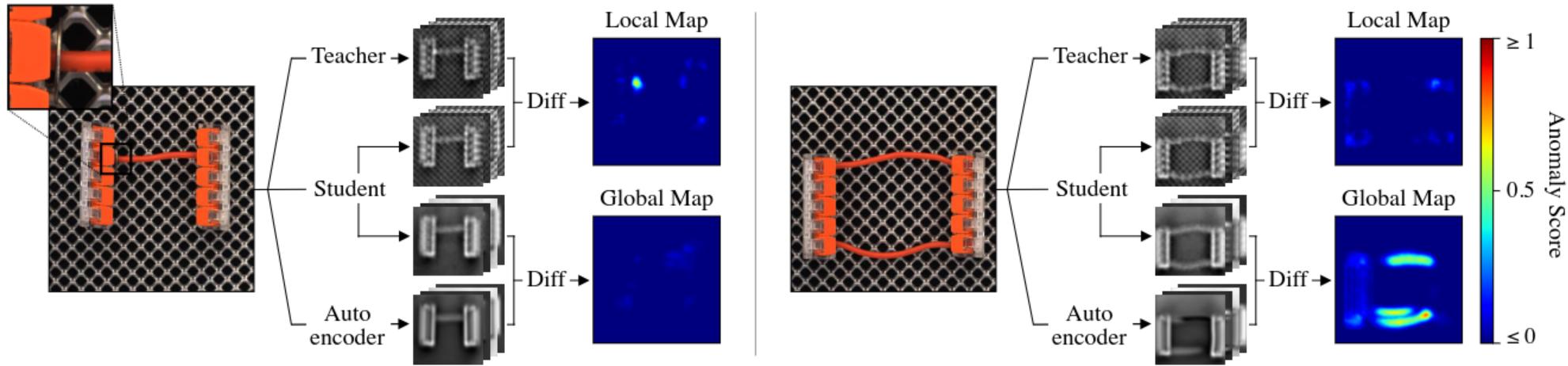


Figure 5. EfficientAD applied to two test images from MVTec LOCO. Normal input images contain a horizontal cable connecting the two splicing connectors at an arbitrary height. The anomaly on the left is a foreign object in the form of a small metal washer at the end of the cable. It is visible in the 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.

ing of the student. Specifically, we sample a random image P from the pretraining dataset, in our case ImageNet, in each training step. We compute the loss of the student as $L_{ST} = L_{\text{hard}} + (CWH)^{-1} \sum_c \|S(P)_c\|_F^2$. This penalty hinders the student from generalizing its imitation of the teacher to out-of-distribution images.

3.3. Logical Anomaly Detection

There are many types of logical anomalies, such as missing, misplaced, or surplus objects or the violation of geometrical constraints, for example, the length of a screw. As recommended by the authors of the MVTec LOCO dataset [8], we use an autoencoder for learning logical constraints of the training images and detecting violations of these constraints. Figure 5 depicts the anomaly detection methodology for EfficientAD. It consists of the aforementioned student–teacher pair and an autoencoder. The autoencoder is trained to predict the output of the teacher. Formally, we apply an autoencoder A to a training image I , yielding $A(I) \in \mathbb{R}^{C \times W \times H}$, and compute the loss as $L_{AE} = (CWH)^{-1} \sum_c \|T(I)_c - A(I)_c\|_F^2$. We use a standard convolutional autoencoder comprising strided convolutions in the encoder and bilinear upsampling in the decoder. We provide the detailed hyperparameters of its layers in the supplementary material.

In contrast to the patch-based student, the autoencoder must encode and decode the complete image through a bottleneck of 64 latent dimensions. On images with logical anomalies, the autoencoder usually fails to generate the correct latent code for reconstructing the image in the teacher’s feature space. However, its reconstructions are also flawed on normal images, as autoencoders generally struggle with reconstructing fine-grained patterns [12, 16]. This is the

case for the background grids in Figure 5. Using the difference between the teacher’s output and the autoencoder’s reconstruction as an anomaly map would cause false-positive detections in these cases. Instead, we double the number of output channels of our student network and train it to predict the output of the autoencoder in addition to the output of the teacher. Let $S'(I) \in \mathbb{R}^{C \times W \times H}$ denote the additional output channels of the student. The student’s additional loss is then $L_{STAE} = (CWH)^{-1} \sum_c \|A(I)_c - S'(I)_c\|_F^2$.

The student learns the systematic reconstruction errors of the autoencoder on normal images, e.g., blurry reconstructions. At the same time, it does not learn the reconstruction errors for anomalies because these are not part of the training set. This makes the difference between the autoencoder’s output and the student’s output well-suited for computing the anomaly map. Analogous to the student–teacher pair, the anomaly map is the squared difference between the two outputs, averaged across channels. We refer to this anomaly map as the global anomaly map and to the anomaly map generated by the student–teacher pair as the local anomaly map. We average these two anomaly maps to compute the combined anomaly map and use its maximum value as the image-level anomaly score. The combined anomaly map thus contains the detection results of the student–teacher pair and the detection results of the autoencoder–student pair. Sharing the student’s hidden layers in the computation of these detection results allows our method to maintain low computational requirements, while enabling the detection of structural and logical anomalies.

3.4. Anomaly Map Normalization

The local and the global anomaly map must be normalized to similar scales before averaging them to obtain the

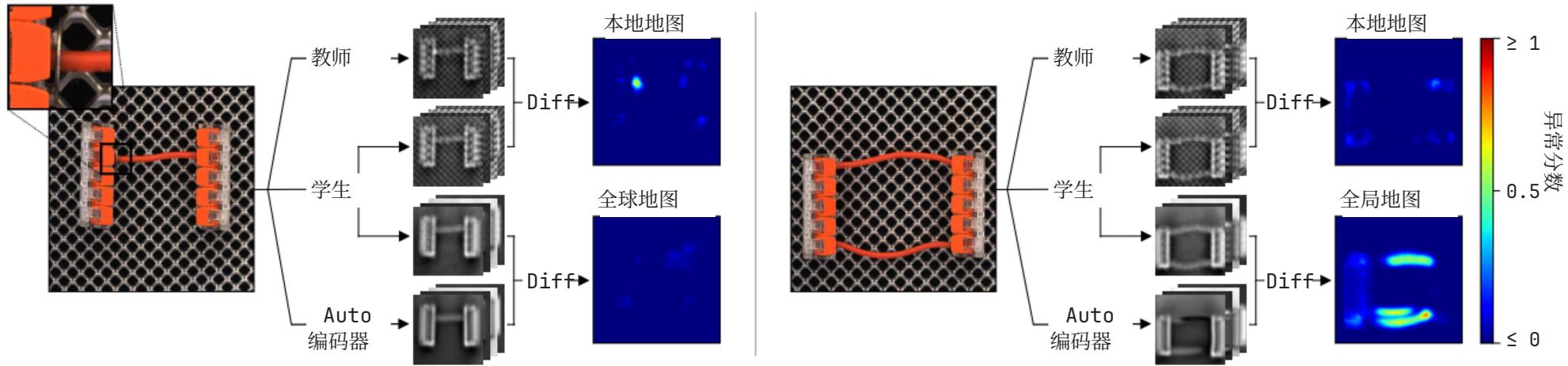


图 5. EfficientAD 应用于 MVTec LOCO 的两个测试图像。正常输入图像包含一条水平电缆，在任意高度连接两个拼接连接器。左侧的异常是一个外来物体，形式为电缆末端的小金属垫圈。它在局部异常图中可见，因为学生和教师的输出不同。右侧的逻辑异常是存在第二条电缆。自动编码器无法在教师的特征空间中重建右侧的两条电缆。学生除了预测教师的输出外，还预测自动编码器的输出。由于其感受野限制在图像的小块区域，它不受额外红色电缆存在的影响。这导致自动编码器和学生的输出不同。“差异”指的是计算两组输出特征图之间的元素级平方差，并计算其在特征图之间的平均值。为了获得像素异常分数，使用双线性插值将异常图调整大小以匹配输入图像。

学生的训练。具体来说，我们在每个训练步骤中从预训练数据集（在我们的情况下是 ImageNet）中随机采样一张图像 P 。我们计算学生的损失为 $L = L + \frac{1}{C} \|S(P) - A(P)\|_F^2$ 。这种惩罚阻碍了学生将对教师的模仿泛化到分布外的图像。

3.3. 逻辑异常检测 有许多类型的逻辑异常，如缺失、错位或多余的物体，或几何约束的违反，例如螺丝的长度。按照 MVTec LOCO 数据集[8]作者的建议，我们使用自动编码器来学习训练图像的逻辑约束并检测这些约束的违反。图 5 描述了 EfficientAD 的异常检测方法。它由上述的学生-教师对和一个自动编码器组成。自动编码器被训练来预测教师的输出。形式上，我们将自动编码器 A 应用于训练图像 I ，得到 $A(I) \in R$ ，并计算损失为 $L = \frac{1}{C} \|A(I) - S(I)\|_F^2$

$c // T(I) - A(I) //$ 。我们使用标准卷积自动编码器，包括编码器中的步进卷积和解码器中的双线性上采样。我们在补充材料中提供了其层的详细超参数。

与基于补丁的学生不同，自动编码器必须通过 64 个潜在维度的瓶颈来编码和解码完整的图像。对于具有逻辑异常的图像，自动编码器通常无法生成正确的潜在代码来在教师的特征空间中重建图像。然而，它在正常图像上的重建也存在缺陷，因为自动编码器通常难以重建细粒度的模式[12, 16]

这适用于图 5 中的背景网格。使用教师输出和自动编码器重建之间的差异作为异常图会在这些情况下导致误报检测。相反，我们将学生网络的输出通道数量加倍，并训练它除了预测教师的输出外，还预测自动编码器的输出。设 $S(I) \in R$ 表示学生的额外输出通道。学生的额外损失则为 $L = \frac{1}{C} \|S(I) - A(I)\|_F^2$

$$P = \frac{1}{C} \|A(I) - S(I)\|_F^2.$$

学生学习自编码器在正常图像上的系统性重建误差，例如模糊的重建。同时，它不会学习异常情况的重建误差，因为这些不是训练集的一部分。这使得自编码器输出和学生输出之间的差异非常适合用于计算异常图。类似于学生-教师对，异常图是两个输出之间的平方差，在通道间取平均。我们将这个异常图称为全局异常图，将学生-教师对生成的异常图称为局部异常图。我们将这两个异常图取平均值来计算组合异常图，并使用其最大值作为图像级异常分数。因此，组合异常图包含了学生-教师对的检测结果和自编码器-学生对的检测结果。在计算这些检测结果时共享学生的隐藏层，使我们的方法能够保持较低的计算需求，同时能够检测结构性和逻辑性异常。

3.4. 异常图归一化

局部和全局异常图在平均以获得之前必须归一化到相似的尺度

combined anomaly map. This is important for cases where the anomaly is only detected in one of the maps, such as in Figure 5. Otherwise, noise in one map could make accurate detections in the other map indiscernible in the combined map. To estimate the scale of the noise in normal images, we use validation images, i.e., unseen images from the training set. For each of the two anomaly map types, we compute the set of all pixel anomaly scores across the validation images. We then compute two p -quantiles for each set: q_a and q_b , for $p = a$ and $p = b$, respectively. We determine a linear transformation that maps q_a to an anomaly score of 0 and q_b to a score of 0.1. At test time, the local and global anomaly maps are normalized with the respective linear transformation.

By using quantiles, the normalization becomes robust to the distribution of anomaly scores on normal images, which can vary between scenarios. Whether the scores between q_a and q_b are normally distributed or a mixture of Gaussians or follow another distribution has no influence on the normalization. Our experiments include an ablation study on the values of a and b . The choice of the mapping destination values 0 and 0.1 has no effect on anomaly detection metrics such as the area under the ROC curve (AU-ROC). That is because the AU-ROC only depends on the ranking of scores, not on their scale. We choose 0 and 0.1 because they yield maps that are suitable for a standard zero-to-one color scale.

4. Experiments

We compare EfficientAD to AST [50], DSR [66], FastFlow [64], GCAD [8], PatchCore [47], SimpleNet [34], and S-T [10], using official implementations where available. We provide configuration details for all evaluated methods in the supplementary material. GCAD consists of an ensemble of two anomaly detection models that use different feature extractors. We find that one of the two ensemble members performs better on average than the combined ensemble and therefore report the results for this member. This reduces the latency reported for GCAD by a factor of two. For SimpleNet, we are able to reproduce the official results but find that SimpleNet tunes the training duration on the test images of a scenario. During training, the model is repeatedly evaluated on all test images and the maximum of all obtained test scores is reported after training. We disable this technique, since it overestimates the actual performance of the model on unseen images. In practice, it would furthermore require a validation set with anomalous images. MVTec AD, VisA, and MVTec LOCO do not include anomalous images in their training and validation sets to avoid defect-type-specific tuning of hyperparameters. For SimpleNet, we therefore evaluate the final trained model, following common practice.

For PatchCore, we include two variants: the default sin-

gle model variant, for which the authors report the lowest latency, and the ensemble variant, denoted by PatchCore_{Ens}. We are able to reproduce the official results but disable the cropping of the center 76.6 % of input images for a fair comparison. In the case of MVTec AD, 99.9 % of the defects lie fully or partially within this cropped area. In real-world applications, anomalies can occur outside of this area as well. We disable custom cropping, as it implies knowledge about the anomalies in the test set. For FastFlow, we use the version based on the WideResNet-50-2 feature extractor, as it is similar to the WideResNet used by PatchCore, SimpleNet, and our method. We use the implementation provided by the Intel anomalib [1] but disable early stopping, i.e., the scenario-specific tuning of the training duration on test images, analogously to SimpleNet. With early stopping enabled, EfficientAD itself achieves an image-level detection AU-ROC of 99.8 % on MVTec AD.

For our method, we evaluate two variants: EfficientAD-S and EfficientAD-M. EfficientAD-S uses the architecture displayed in Figure 2 for the teacher and the student. For EfficientAD-M, we double the number of kernels in the hidden convolutional layers of the teacher and the student. Furthermore, we insert a 1×1 convolution after the second pooling layer and after the last convolutional layer. We provide a list of implementation details, such as the learning rate schedule, in the supplementary material.

We evaluate each method on the 32 anomaly detection scenarios of MVTec AD, VisA, and MVTec LOCO. The anomaly detection performance of a method is measured with the AU-ROC based on its predicted image-level anomaly scores. We measure the anomaly localization performance using the AU-PRO segmentation metric up to a false positive rate of 30 %, as recommended by [7]. For MVTec LOCO, we use the AU-sPRO metric [8], a generalization of the AU-PRO metric for evaluating the localization of logical anomalies. The supplementary material provides the results for additional anomaly detection metrics, such as the area under the precision-recall curve and the pixel-wise AU-ROC.

When reporting the AU-ROC or AU-PRO for a dataset collection, we follow the policy of the dataset authors. For each collection, we evaluate the respective metric for each scenario and then compute the mean across scenarios. For MVTec LOCO, we use the official evaluation script, which gives logical and structural anomalies an equal weight in the computed metrics. When reporting the average AU-ROC or AU-PRO on the three dataset collections, we compute the average of the three dataset means. Thus, an overall average score weights logical anomalies and structural anomalies by roughly one-sixth and five-sixths, respectively. We provide the evaluation results for each of the 32 anomaly detection scenarios individually in the supplementary material to enable an evaluation with a custom weighting.

异常图归一化

局部和全局异常图在平均以获得组合异常图之前必须归一化到相似的尺度。这对于仅在一个图中检测到异常的情况很重要,如图 5 所示。否则,一个图中的噪声可能会使另一个图中的准确检测在组合图中变得难以辨别。为了估计正常图像中噪声的尺度,我们使用验证图像,即来自训练集的未见图像。对于两种异常图类型中的每一种,我们计算验证图像中所有像素异常分数的集合。然后我们为每个集合计算两个 p 分位数: q 和 q ,分别对应 $p = a$ 和 $p = b$ 。我们确定一个线性变换,将 q 映射到异常分数 0,将 q 映射到分数 0.1。在测试时,局部和全局异常图使用各自的线性变换进行归一化。

通过使用分位数,归一化对正常图像上异常分数的分布变得稳健,这种分布在不同场景之间可能会有所不同。无论 q 和 q 之间的分数是正态分布、高斯混合分布还是遵循其他分布,都不会影响归一化。我们的实验包括对 a 和 b 值的消融研究。选择映射目标值 0 和 0.1 对异常检测指标 (如 ROC 曲线下面积 (AU-ROC)) 没有影响。这是因为 AU-ROC 只依赖于分数的排序,而不依赖于它们的尺度。我们选择 0 和 0.1 是因为它们产生的映射适合标准的 0 到 1 颜色尺度。

4. 实验 我们将 EfficientAD 与 AST [50]、DSR [66]、FastFlow [64]、GCAD [8]、PatchCore [47]、SimpleNet [34] 和 S-T [10] 进行比较,在可能的情况下使用官方实现。我们在补充材料中提供了所有评估方法的配置详细信息。GCAD 由两个使用不同特征提取器的异常检测模型组成的集成。我们发现,两个集成成员中的一个平均表现比组合集成更好,因此报告了这个成员的结果。这将 GCAD 报告的延迟减少了一半。对于 SimpleNet,我们能够复现官方结果,但发现 SimpleNet 在场景的测试图像上调整训练持续时间。在训练过程中,模型反复评估所有测试图像,并在训练后报告所有获得的测试分数的最大值。我们禁用了这种技术,因为它高估了模型在未见过的图像上的实际性能。在实践中,这还需要一个包含异常图像的验证集。MVTec AD、VisA 和 MVTec LOCO 在其训练和验证集中不包含异常图像,以避免对特定缺陷类型的超参数进行调整。因此,对于 SimpleNet,我们评估最终训练的模型,遵循常见做法。

对于作者报告延迟最低的单一模型变体和称为 PatchCore 的集成变体,我们能够复现官方结果,但为了公平比较,禁用了输入图像中心 76.6% 的裁剪。在 MVTec AD 的情况下,99.9% 的缺陷完全或部分位于这个裁剪区域内。在实际应用中,异常也可能发生在这个区域之外。我们禁用自定义裁剪,因为它暗示了对测试集中异常的了解。对于 FastFlow,我们使用基于 WideResNet-50-2 特征提取器的版本,因为它与 PatchCore、SimpleNet 和我们的方法使用的 WideResNet 相似。我们使用 Intel anomalib [1] 提供的实现,但禁用了早停,即在测试图像上对训练持续时间进行场景特定调整,类似于 SimpleNet。启用早停后,EfficientAD 本身在 MVTec AD 上达到 99.8% 的图像级检测 AU-ROC。

对于我们的方法,我们评估了两个变体:EfficientAD-S 和 EfficientAD-M。EfficientAD-S 使用图 2 中显示的架构作为教师和学生。对于 EfficientAD-M,我们将教师和学生的隐藏卷积层中的核数量翻倍。此外,我们在第二个池化层之后和最后一个卷积层之后插入了 1×1 卷积。我们在补充材料中提供了实现细节列表,如学习率调度等。

我们在 MVTec AD、VisA 和 MVTec LOCO 的 32 个异常检测场景上评估每种方法。方法的异常检测性能是根据其预测的图像级异常分数,使用 AU-ROC 进行衡量。我们使用 AU-PRO 分割指标来衡量异常定位性能,假阳性率最高为 30%,这是[7]推荐的。对于 MVTec LOCO,我们使用 AU-sPRO 指标[8],这是 AU-PRO 指标的一种推广,用于评估逻辑异常的定位。补充材料提供了其他异常检测指标的结果,如精确率-召回率曲线下面积和像素级 AU-ROC。

在报告数据集集合的 AU-ROC 或 AU-PRO 时,我们遵循数据集作者的政策。对于每个集合,我们评估每个场景的相应指标,然后计算场景之间的平均值。对于 MVTec LOCO,我们使用官方评估脚本,该脚本在计算指标时给予逻辑异常和结构异常相等的权重。在报告三个数据集集合的平均 AU-ROC 或 AU-PRO 时,我们计算三个数据集平均值的平均值。因此,总体平均分数大约将逻辑异常和结构异常的权重分别设为六分之一和六分之五。我们在补充材料中提供了 32 个异常检测场景中每个场景的单独评估结果,以便进行自定义权重的评估。

对于 PatchCore,我们包括两种变体:默认的 sin-

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
PatchCore _{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
PatchCore _{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.

a (for q_a)	0.5	0.8	0.9	0.95	0.98	0.99
AU-ROC	95.9	95.9	96.0	95.9	95.9	95.8
b (for q_b)	0.95	0.98	0.99	0.995	0.998	0.999
AU-ROC	95.8	95.9	96.0	96.0	95.9	95.9
p_{hard}	0	0.9	0.99	0.999	0.9999	0.99999
AU-ROC	94.9	94.9	95.7	96.0	95.8	95.7

Table 3. Mean anomaly detection AU-ROC of EfficientAD-M on MVTec AD, VisA, and MVTec LOCO when varying the locations of quantiles. These are the two sampling points a and b of the quantile-based map normalization and the mining factor p_{hard} . Setting p_{hard} to zero disables the proposed hard feature loss. Default values used in our experiments are highlighted in bold.

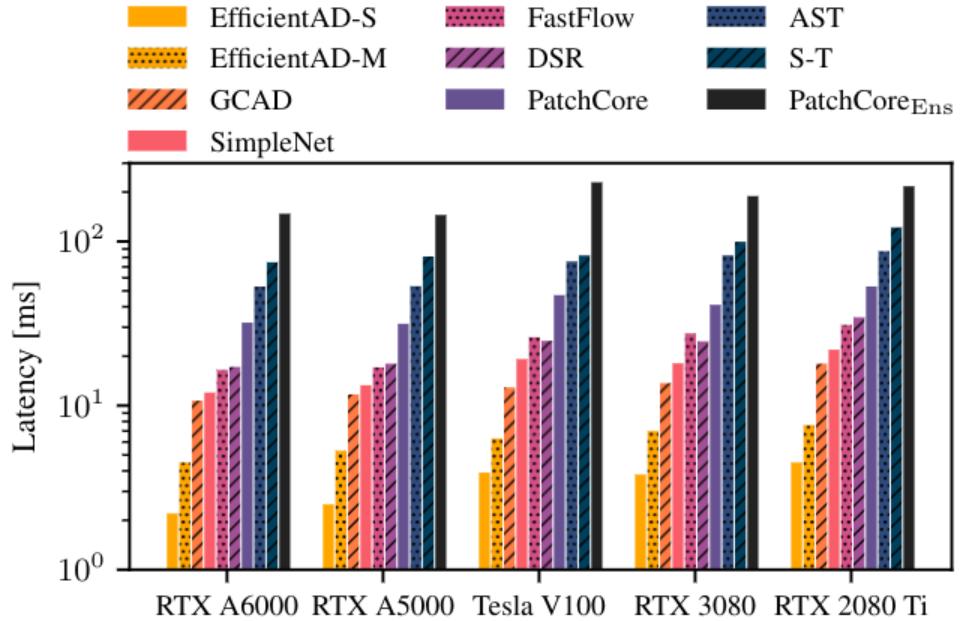


Figure 6. Latency per GPU. The ranking of methods is the same on each GPU, except for two cases in which DSR is slightly faster than FastFlow.

Table 1 reports the overall anomaly detection performance for each method. EfficientAD achieves a strong image-level detection and pixel-level localization of anomalies. Reliably localizing anomalies in an image provides explainable detection results and allows the discovery of spurious correlations in detections. It also enables a flexible postprocessing, such as excluding defect segmentations based on their size.

Table 2 breaks down the overall anomaly detection performance into the three dataset collections. It shows that the lead of EfficientAD on MVTec LOCO is in equal parts due to its performance on logical and on structural anomalies. In Table 3, we assess the robustness of EfficientAD to varying hyperparameters.

Furthermore, we measure the computational cost of each method during inference. As explained above, the number of parameters can be a misleading proxy metric for the latency and throughput of convolutional architectures since it does not consider the resolution of a convolution’s input feature map, i.e., how often a parameter is used in a forward pass. Similarly, the number of floating point operations (FLOPs) can be misleading since it does not take into account how easily computations can be parallelized. For transparency, we report the number of parameters, the number of FLOPs, and the memory footprint of each method in the supplementary material. Here, we focus on the metrics that are most relevant in anomaly detection applications: the latency and the throughput. We measure the latency with a batch size of 1 and the throughput with a batch size of 16. Table 1 reports the measurements for each method on an NVIDIA RTX A6000 GPU. Figure 6 shows the latency of each method on each of the GPUs in our experimental setup. The supplementary material contains a detailed description of our timing methodology.

方法	检测。 [图像/秒]	分割。	延迟	吞吐量
GCAD	85.4	89.0	11.0	SimpleNet 81.9 74.4 12 194
S-T	88.4	89.7	75.16	FastFlow 90.0 86.5 17 120 DSR 90.8 78.6
PatchCore	17.104	PatchCore	91.1 80.9 32 76	PatchCore 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)

表 1. 异常检测和异常定位性能与延迟和吞吐量的比较。每个 AUROC 和 AU-PRO 百分比是 MVTec AD、VisA 和 MVTec LOCO 上平均 AUROC 和平均 AU-PRO 的平均值。对于 EfficientAD，我们报告了五次运行的平均值和标准偏差。

方法	MAD	LOCO	VisA	平均值	
				逻辑。	结构。
GCAD	89.1	83.3	83.7	85.4	83.9
	98.2	77.6	87.9	87.9	71.5
	94.6	88.4	66.5	88.3	FastFlow 96.9
	90.0	75.5	82.9	DSR 98.1	82.6
	90.2	PatchCore	98.7	80.3	94.3
	PatchCore	99.3	79.4	97.7	92.1
	83.4	94.9	92.4	79.7	87.1
	90.0	97.5	95.4	85.8	94.1
	-90.7	98.1	96.0	86.8	94.7

表 2. 每个数据集集合的平均异常检测 AU-ROC 百分比（左）和 MVTec LOCO 的逻辑和结构异常（右）。对于 EfficientAD，我们报告了五次运行的平均值。仅在 MVTec AD (MAD) 上进行方法开发容易导致设计选择过度拟合剩余的少数错误分类测试图像。

a (对于 q)	0.5	0.8	0.9	0.95	0.98	0.99	AU-ROC	95.9	95.9	96.0
	95.9	95.9	95.8							
b (对于 q)	0.95	0.98	0.99	0.995	0.998	0.999	AU-ROC	95.8	95.9	96.0

p0	0.9	0.99	0.999	0.9999	0.99999	AU-ROC	94.9	94.9	95.7	96.0	95.8	95.7
----	-----	------	-------	--------	---------	--------	------	------	------	------	------	------

表 3. 在改变分位数位置时，EfficientAD-M 在 MVTec AD、VisA 和 MVTec LOCO 上的平均异常检测 AU-ROC。这是基于分位数的映射归一化的两个采样点 a 和 b，以及挖掘因子 p。将 p 设置为零会禁用所提出的硬特征损失。我们实验中使用的默认值以粗体突出显示。

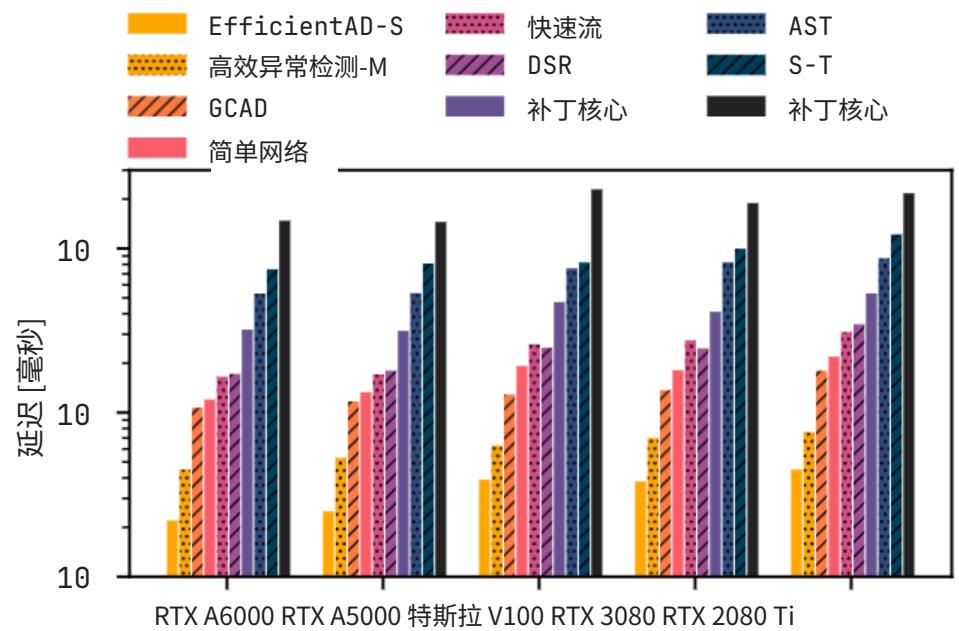


图 6. 每个 GPU 的延迟。除了两种情况下 DSR 略快于 FastFlow 外，各方法在每个 GPU 上的排名都相同。

表 1 报告了每种方法的整体异常检测性能。EfficientAD 在图像级检测和像素级异常定位方面都取得了强劲的表现。可靠地定位图像中的异常提供了可解释的检测结果，并允许发现检测中的虚假相关性。它还能实现灵活的后处理，例如根据缺陷分割的大小来排除它们。

表 2 将整体异常检测性能分解为三个数据集集合。它显示 EfficientAD 在 MVTec LOCO 上的领先地位同等程度地归功于其在逻辑异常和结构异常上的表现。在表 3 中，我们评估了 EfficientAD 对不同超参数的鲁棒性。

此外，我们测量了每种方法在推理过程中的计算成本。如上所述，参数数量可能是卷积架构延迟和吞吐量的误导性代理指标，因为它没有考虑卷积输入特征图的分辨率，即一个参数在前向传播中被使用的频率。同样，浮点运算次数 (FLOPs) 也可能具有误导性，因为它没有考虑计算可以更容易地并行化。为了透明度，我们在补充材料中报告了每种方法的参数数量、FLOPs 数量和内存占用。这里，我们专注于异常检测应用中最相关的指标：延迟和吞吐量。我们使用批量大小为 1 测量延迟，使用批量大小为 16 测量吞吐量。

表 1 报告了每种方法在 NVIDIA RTX A6000 GPU 上的测量结果。图 6 显示了每种方法在我们实验设置中每个 GPU 上的延迟。补充材料包含了我们计时方法的详细描述。

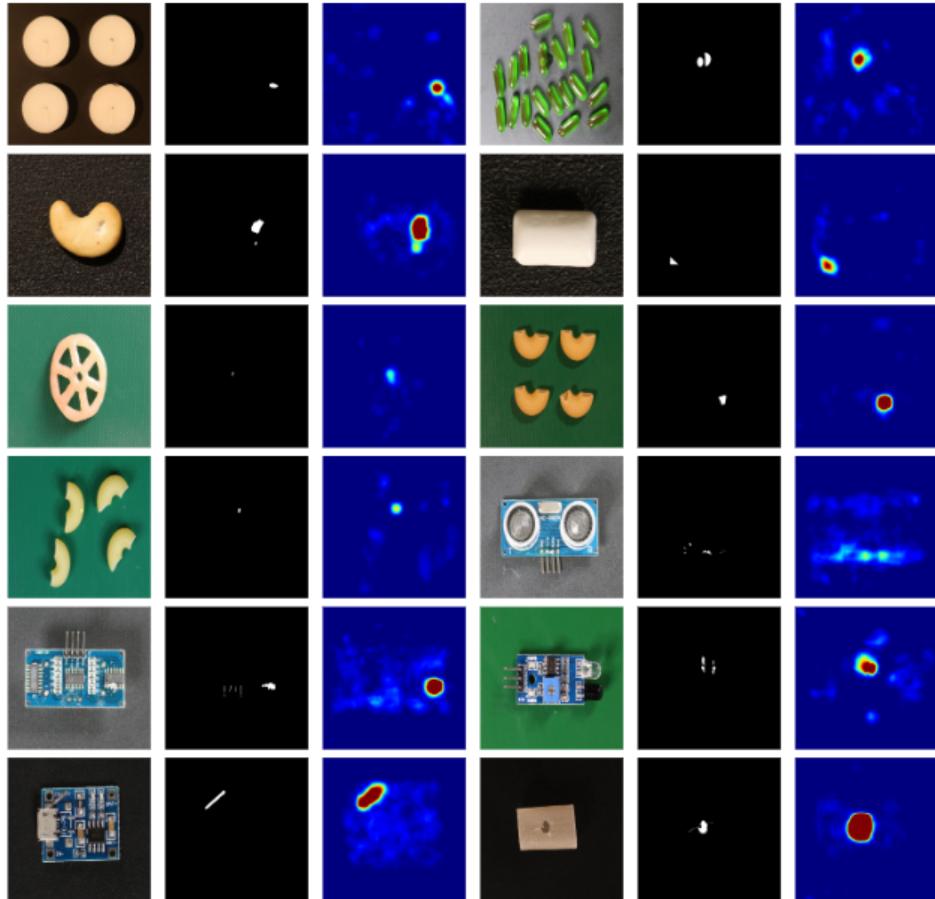


Figure 7. Non-cherry-picked qualitative results of EfficientAD on VisA. For each of its 12 scenarios, we show a randomly sampled defect image, the ground truth segmentation mask, and the anomaly map generated by EfficientAD-M.

In Figure 7, we show randomly sampled qualitative results of EfficientAD on the VisA dataset collection. The supplementary material provides qualitative results for the other evaluated methods and dataset collections as well.

We examine the effects of the components of EfficientAD in the ablation study shown in Table 4 and Table 5. For experiments without the proposed quantile-based map normalization, we use a Gaussian-based map normalization as a baseline instead. There, we compute the linear transformation parameters such that pixel anomaly scores on the validation set have a mean of zero and a variance of one. This baseline normalization is sensitive to the distribution of validation anomaly scores, which can vary between scenarios. The quantile-based normalization is independent of how the scores between q_a and q_b are distributed and performs substantially better than the baseline.

We also evaluate the effect of the two proposed loss terms for training the student–teacher pair. The hard feature loss increases the anomaly detection AU-ROC by 1.0 % in Table 4. This improvement alone is greater than or equal to each of the improvement margins between the consecutive rows of FastFlow, DSR, PatchCore, PatchCore_{Ens}, and AST in Table 1. The student’s penalty on pretraining images further improves the anomaly detection performance. Notably, the proposed map normalization, the hard feature loss, and the pretraining penalty keep the computational requirements of EfficientAD low, while creating a substantial margin w.r.t. the anomaly detection performance.

	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.

5. Conclusion

In this paper, we introduce EfficientAD, a method with a strong anomaly detection performance and a high computational efficiency. It sets new standards for the detection of structural as well as logical anomalies. Both EfficientAD-S and EfficientAD-M outperform other methods on the detection and the localization of anomalies by a large margin. Compared to AST, the second-best method, EfficientAD-S reduces the latency by a factor of 24 and increases the throughput by a factor of 15. Its low latency, high throughput, and high detection rate make it suitable for real-world applications. For future anomaly detection research, EfficientAD is an important baseline and a fruitful foundation. Its efficient patch description network, for instance, can be used as a feature extractor in other anomaly detection methods as well to reduce their latency.

Limitations. The student–teacher model and the autoencoder are designed to detect anomalies of different types. The autoencoder detects logical anomalies, while the student–teacher model detects coarse and fine-grained structural anomalies. Fine-grained logical anomalies, however, remain a challenge – for example a screw that is two millimeters too long. To detect these, practitioners would have to use traditional metrology methods [58]. As for the limitations in comparison to other recent anomaly detection methods: In contrast to kNN-based methods, our approach requires training, especially for the autoencoder to learn the logical constraints of normal images. This takes twenty minutes in our experimental setup.

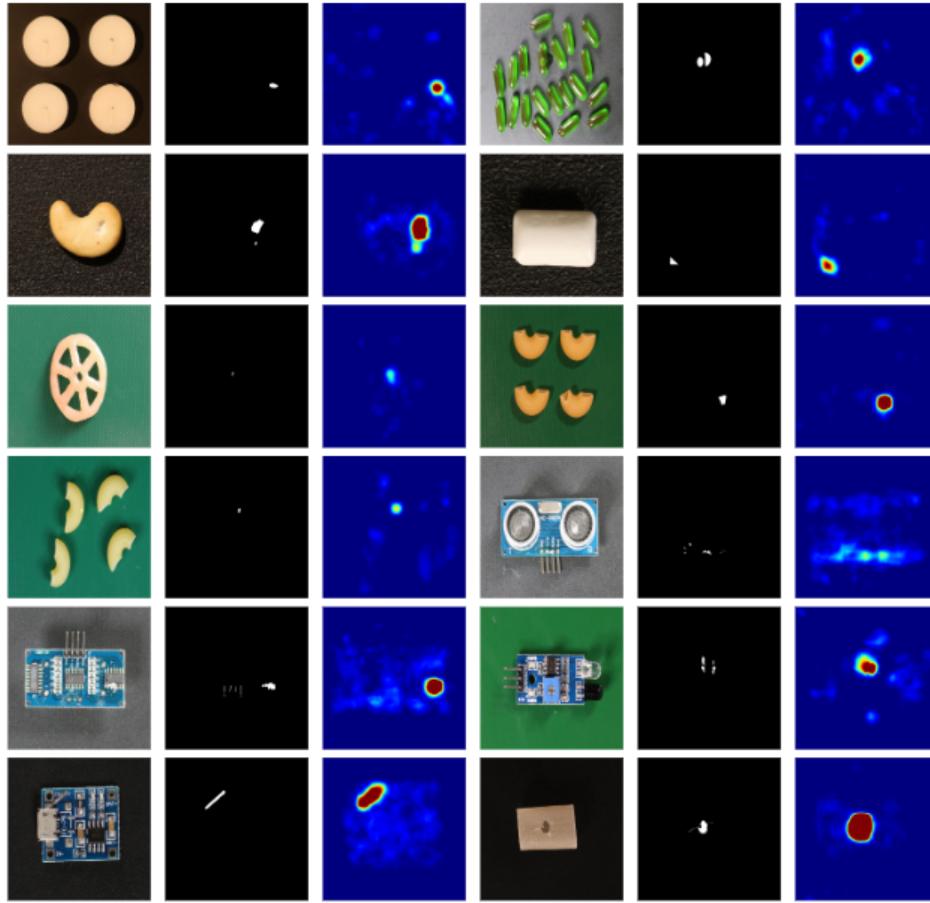


图 7. EfficientAD 在 VisA 上的非精选定性结果。对于其 12 个场景中的每一个，我们展示了随机抽样的缺陷图像、真实分割掩码以及由 EfficientAD-M 生成的异常图。

在图 7 中，我们展示了 EfficientAD 在 VisA 数据集集合上随机抽样的定性结果。补充材料还提供了其他评估方法和数据集集合的定性结果。

我们在表 4 和表 5 所示的消融研究中检验了 EfficientAD 各组成部分的效果。对于没有使用所提出的基于分位数的图归一化的实验，我们使用基于高斯分布的图归一化作为基线。在那里，我们计算线性变换参数，使验证集上的像素异常分数均值为零，方差为一。这种基线归一化对验证异常分数的分布敏感，而这种分布在不同情况下可能会有所不同。基于分位数的归一化独立于 q 和 q' 之间的分数分布方式，其表现明显优于基线。

我们还评估了用于训练学生-教师对的两个提出的损失项的效果。硬特征损失在表 4 中将异常检测 AU-ROC 提高了 1.0%。仅这一改进就大于或等于表 1 中 FastFlow、DSR、PatchCore、PatchCore 和 AST 连续行之间的每个改进幅度。学生对预训练图像的惩罚进一步提高了异常检测性能。值得注意的是，所提出的图归一化、硬特征损失和预训练惩罚保持了 EfficientAD 的低计算需求，同时在异常检测性能方面创造了显著的优势。

	检测 AU-ROC	差异	延迟 [ms]
PDN 93.2 2.2 → 使用地图归一化	94.0 + 0.8	2.2 → 使用硬特征	
损失 95.0 + 1.0 2.2 → 使用预训练惩罚	95.4 + 0.4	2.2	
EfficientAD-S 95.4 2.2	EfficientAD-M 96.0 + 0.6	4.5	

表 4. 累积消融研究，逐步结合技术以形成 EfficientAD。每个 AU-ROC 百分比是 MVTec AD、VisA 和 MVTec LOCO 平均 AU-ROC 的平均值。

	检测 [毫秒]	差异	延迟
EfficientAD-S 95.4 2.2 无地图归一化	94.7 - 0.7	2.2 无	
硬特征损失 94.7 - 0.7 2.2 无预训练惩罚	95.0 - 0.4	2.2	

表 5. 对 EfficientAD-S 单独移除技术的隔离消融研究。

5. 结论 在本文中，我们介绍了 EfficientAD，这是一种具有强大异常检测性能和高计算效率的方法。它为结构性和逻辑性异常的检测设立了新标准。EfficientAD-S 和 EfficientAD-M 在异常检测和定位方面都大幅超越了其他方法。与排名第二的 AST 方法相比，EfficientAD-S 将延迟降低了 24 倍，并将吞吐量提高了 15 倍。其低延迟、高吞吐量和高检测率使其适用于实际应用。对于未来的异常检测研究，EfficientAD 是一个重要的基准和富有成果的基础。例如，其高效的补丁描述网络可以用作其他异常检测方法的特征提取器，以降低它们的延迟。

局限性。学生-教师模型和自动编码器被设计用于检测不同类型的异常。自动编码器检测逻辑异常，而学生-教师模型检测粗粒度和细粒度的结构异常。然而，细粒度的逻辑异常仍然是一个挑战 - 例如一个长度超出 2 毫米的螺丝。要检测这些，实践者必须使用传统的计量方法[58]。与其他最新异常检测方法相比的局限性：与基于 kNN 的方法相比，我们的方法需要训练，特别是自动编码器需要学习正常图像的逻辑约束。在我们的实验设置中，这需要二十分钟。

References

- [1] Samet Akcay, Dick Ameln, Ashwin Vaidya, Barath Lakshmanan, Nilesh Ahuja, and Utku Genc. Anomalib: A deep learning library for anomaly detection. In *2022 IEEE International Conference on Image Processing (ICIP)*, pages 1706–1710. IEEE, 2022. [6](#)
- [2] Samet Akcay, Amir Atapour-Abarghouei, and Toby P Breckon. Ganomaly: Semi-supervised anomaly detection via adversarial training. In *Computer Vision–ACCV 2018: 14th Asian Conference on Computer Vision, Perth, Australia, December 2–6, 2018, Revised Selected Papers, Part III 14*, pages 622–637. Springer, 2019. [3](#)
- [3] Samuel G Armato III, Geoffrey McLennan, Luc Bidaut, Michael F McNitt-Gray, Charles R Meyer, Anthony P Reeves, Binsheng Zhao, Denise R Aberle, Claudia I Henschke, Eric A Hoffman, et al. The lung image database consortium (lidc) and image database resource initiative (idri): a completed reference database of lung nodules on ct scans. *Medical physics*, 38(2):915–931, 2011. [2](#)
- [4] Donald Bailey. *Implementing Machine Vision Systems Using FPGAs*, pages 1103–1136 in “Machine Vision Handbook” by Bruce G. Batchelor. Springer London, London, 2012. [1](#)
- [5] Spyridon Bakas, Hamed Akbari, Aristeidis Sotiras, Michel Bilello, Martin Rozycki, Justin S. Kirby, et al. Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. *Scientific Data*, 4(1), 2017. [2](#)
- [6] Christoph Baur, Benedikt Wiestler, Shadi Albarqouni, and Nassir Navab. Deep autoencoding models for unsupervised anomaly segmentation in brain mr images. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pages 161–169. Springer International Publishing, 2019. [3](#)
- [7] Paul Bergmann, Kilian Batzner, Michael Fauser, David Sattlegger, and Carsten Steger. The MVTec Anomaly Detection Dataset: A Comprehensive Real-World Dataset for Unsupervised Anomaly Detection. *International Journal of Computer Vision*, 129(4):1038–1059, 2021. [1, 2, 6](#)
- [8] Paul Bergmann, Kilian Batzner, Michael Fauser, David Sattlegger, and Carsten Steger. Beyond Dents and Scratches: Logical Constraints in Unsupervised Anomaly Detection and Localization. *International Journal of Computer Vision*, 130(4):947—969, 2022. [1, 2, 3, 5, 6](#)
- [9] Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. MVTec AD — A Comprehensive Real-World Dataset for Unsupervised Anomaly Detection. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9584–9592, 2019. [1, 2](#)
- [10] Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. Uninformed Students: Student-Teacher Anomaly Detection With Discriminative Latent Embeddings. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4182–4191, 2020. [2, 3, 4, 6](#)
- [11] Paul Bergmann, Xin Jin, David Sattlegger, and Carsten Steger. The MVTec 3D-AD Dataset for Unsupervised 3D Anomaly Detection and Localization. In *Proceedings of the 17th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 5: VISAPP*, pages 202–213. INSTICC, SciTePress, 2022. [2](#)
- [12] Paul Bergmann, Sindy Löwe, Michael Fauser, David Sattlegger, and Carsten Steger. Improving Unsupervised Defect Segmentation by Applying Structural Similarity to Autoencoders. In *Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 5: VISAPP*, pages 372–380. INSTICC, SciTePress, 2019. [3, 5](#)
- [13] Hermann Blum, Paul-Edouard Sarlin, Juan Nieto, Roland Siegwart, and Cesar Cadena. Fishyscapes: A Benchmark for Safe Semantic Segmentation in Autonomous Driving. In *IEEE International Conference on Computer Vision Workshops (ICCVW)*, pages 2403–2412, 2019. [2](#)
- [14] Niv Cohen and Yedid Hoshen. Sub-image anomaly detection with deep pyramid correspondences. *arXiv preprint arXiv:2005.02357v1*, 2020. [3](#)
- [15] Thomas Defard, Aleksandr Setkov, Angelique Loesch, and Romaric Audigier. PaDiM: A Patch Distribution Modeling Framework for Anomaly Detection and Localization. In *Pattern Recognition. ICPR International Workshops and Challenges*, pages 475–489. Springer International Publishing, 2021. [3](#)
- [16] Alexey Dosovitskiy and Thomas Brox. Generating Images with Perceptual Similarity Metrics based on Deep Networks. In *Advances in Neural Information Processing Systems*, pages 658–666, 2016. [5](#)
- [17] Thibaud Ehret, Axel Davy, Jean-Michel Morel, and Mauricio Delbracio. Image Anomalies: A Review and Synthesis of Detection Methods. *Journal of Mathematical Imaging and Vision*, 61(5):710–743, 2019. [2](#)
- [18] Tharindu Fernando, Harshala Gammulle, Simon Denman, Sridha Sridharan, and Clinton Fookes. Deep learning for medical anomaly detection – a survey. *ACM Computing Surveys*, 54(7), 2021. [2](#)
- [19] Dong Gong, Lingqiao Liu, Vuong Le, Budhaditya Saha, Moussa Reda Mansour, Svetha Venkatesh, and Anton Van Den Hengel. Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection. In *IEEE International Conference on Computer Vision (ICCV)*, pages 1705–1714, 2019. [3](#)
- [20] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative Adversarial Nets. In *Advances in Neural Information Processing Systems*, pages 2672–2680, 2014. [3](#)
- [21] Denis Gudovskiy, Shun Ishizaka, and Kazuki Kozuka. CFLOW-AD: Real-Time Unsupervised Anomaly Detection With Localization via Conditional Normalizing Flows. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 98–107, 2022. [3](#)
- [22] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016. [3](#)

参考文献

- [1] Samet Akcay、Dick Ameln、Ashwin Vaidya、Barath Lakshmanan、Nilesh Ahuja 和 Utku Genc。Anomalib：一个用于异常检测的深度学习库。在 2022 年 IEEE 国际图像处理会议 1706-1710。IEEE，2022 年。6
- [2] Samet Akcay、Amir Atapour-Abarghouei 和 Toby P Breckon。GANomaly：通过对抗训练进行半监督异常检测。在计算机视觉-ACCV 2018：第 14 届亚洲计算机视觉会议，澳大利亚珀斯，2018 年 12 月 2-6 日，修订选定论文，第 III 部分 14，页码 622-637。施普林格，2019 年。3
- [3] Samuel G Armato III、Geoffrey McLennan、Luc Bidaut、Michael F McNitt-Gray、Charles R Meyer、Anthony P Reeves、Binsheng Zhao、Denise R Aberle、Claudia I Henschke、Eric A Hoffman 等人。肺部图像数据库联盟 (LIDC) 和图像数据库资源计划 (IDRI)：一个完整的 CT 扫描肺结节参考数据库。医学物理学，38(2):915-931，2011 年。2
- [4] Donald Bailey。使用 FPGA 实现机器视觉系统，Bruce G. Batchelor 编著的“机器视觉手册”第 1103-1136 页。Springer London 出版社，伦敦，2012 年。1 [5] Spyridon Bakas、Hamed Akbari、Aristeidis Sotiras、Michel Bilello、Martin Rozycski、Justin S. Kirby 等人。通过专家分割标签和放射组学特征推进癌症基因组图谱胶质瘤 MRI 收集。科学数据，4(1)，2017 年。2 [6] Christoph Baur、Benedikt Wiestler、Shadi Albarqouni 和 Nassir Navab。用于无监督的深度自编码模型。脑部磁共振图像中的异常分割。在《脑损伤：胶质瘤、多发性硬化症、中风和创伤性脑损伤》中，第 161-169 页。施普林格国际出版社，2019 年。3
- [7] Paul Bergmann、Kilian Batzner、Michael Fauser、David Sattlegger 和 Carsten Steger。MVTec 异常检测数据集：用于无监督异常检测的综合真实世界数据集。国际计算机视觉杂志，129(4):1038-1059，2021 年。1, 2, 6 [8] Paul Bergmann、Kilian Batzner、Michael Fauser、David Sattlegger 和 Carsten Steger。超越凹痕和划痕：无监督异常检测和定位中的逻辑约束。国际计算机视觉杂志，130(4):947-969，2022 年。1, 2, 3, 5, 6
- [9] Paul Bergmann、Michael Fauser、David Sattlegger 和 Carsten Steger。MVTec AD — 一个用于无监督异常检测的综合真实世界数据集。在 IEEE 计算机视觉与模式识别会议 (CVPR) 上，第 9584-9592 页，2019 年。1, 2 [10] Paul Bergmann、Michael Fauser、David Sattlegger 和 Carsten Steger。未知学生：使用判别性潜在嵌入的学生-教师异常检测。发表于 IEEE 计算机视觉与模式识别会议 (CVPR)，第 4182-4191 页，2020 年。2, 3, 4, 6
- [11] Paul Bergmann、Xin Jin、David Sattlegger 和 Carsten Steger。MVTec 3D-AD 数据集用于无监督 3D 异常检测和定位。发表于第 17 届国际计算机视觉、成像和计算机图形理论与应用联合会议论文集第 5 卷：VISAPP，第 202-213 页。INSTICC，ger。MVTec 3D-AD 数据集用于无监督 3D 异常检测和定位。发表于第 17 届国际计算机视觉、成像和计算机图形理论与应用联合会议论文集第 5 卷：VISAPP，第 202-213 页。INSTICC，
- [12] Paul Bergmann、Sindy L'owe、Michael Fauser、David Sattlegger 和 Carsten Steger。通过将结构相似性应用于自动编码器来改进无监督缺陷分割。发表于第 14 届国际计算机视觉、成像和计算机图形理论与应用联合会议论文集 - 第 5 卷：VISAPP，第 372-380 页。INSTICC，SciTePress，2019 年。3, 5
- [13] Hermann Blum、Paul-Edouard Sarlin、Juan Nieto、Roland Siegwart 和 Cesar Cadena。Fishscapes：自动驾驶中安全语义分割的基准。发表于 IEEE 国际计算机视觉研讨会 (ICCVW)，第 2403-2412 页，2019 年。2
- [14] Niv Cohen 和 Yedid Hoshen。子图像异常检测具有深度金字塔对应关系的运动。arXiv 预印本 arXiv:2005.02357v1，2020 年 3 月
- [15] Thomas Defard、Aleksandr Setkov、Angélique Loesch 和 Romaric Audigier。PaDiM：一个用于异常检测和定位的补丁分布建模框架。发表于模式识别。ICPR 国际研讨会和挑战，第 475-489 页。施普林格国际出版社，2021 年。3
- [16] Alexey Dosovitskiy 和 Thomas Brox。生成基于深度网络的具有感知相似性度量的图像。发表于神经信息处理系统进展，第 658-666 页，2016 年。5
- [17] Thibaud Ehret、Axel Davy、Jean-Michel Morel 和 Mauricio Delbracio。图像异常：检测方法的回顾与综合。数学图像与视觉杂志，61(5):710-743，2019。2
- [18] Tharindu Fernando、Harshala Gammulle、Simon Denman、Sridha Sridharan 和 Clinton Fookes。深度学习在医学异常检测中的应用 – 一项调查。ACM 计算调查，54(7)，2021。2
- [19] Ian Goodfellow、Jean Pouget-Abadie、Mehdi Mirza、Bing Xu、David Warde-Farley、Sherjil Ozair、Aaron Courville 和 Yoshua Bengio。生成对抗网络。发表于神经信息处理系统进展，第 2672-2680 页，2014 年。3
- [20] Ian Goodfellow、Jean Pouget-Abadie、Mehdi Mirza、Bing Xu、David Warde-Farley、Sherjil Ozair、Aaron Courville 和 Yoshua Bengio。生成对抗网络。发表于神经信息处理系统进展，第 2672-2680 页，2014 年。3
- [21] Denis Gudovskiy、Shun Ishizaka 和 Kazuki Kozuka。CFLOW-AD：通过条件归一化流实现实时无监督异常检测和定位。发表于 IEEE/CVF 冬季计算机视觉应用会议 (WACV) 论文集，第 98-107 页，2022 年。3
- [22] 何恺明、张翔宇、任少卿和孙剑。深度残差学习用于图像识别。发表于 IEEE 计算机视觉与模式识别会议 (CVPR)，第 770-778 页，2016 年。3

- [23] Dan Hendrycks, Steven Basart, Mantas Mazeika, Andy Zou, Joseph Kwon, Mohammadreza Mostajabi, Jacob Steinhardt, and Dawn Song. Scaling out-of-distribution detection for real-world settings. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato, editors, *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 8759–8773. PMLR, 17–23 Jul 2022. 2
- [24] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017. 4
- [25] Yibin Huang, Congying Qiu, Yue Guo, Xiaonan Wang, and Kui Yuan. Surface defect saliency of magnetic tile. In *2018 IEEE 14th International Conference on Automation Science and Engineering (CASE)*, pages 612–617, 2018. 2
- [26] Jeremy Irvin, Pranav Rajpurkar, Michael Ko, Yifan Yu, Silvana Ciurea-Ilcus, Chris Chute, Henrik Marklund, Behzad Haghgoor, Robyn Ball, Katie Shpanskaya, et al. CheXpert: A large chest radiograph dataset with uncertainty labels and expert comparison. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 590–597, 2019. 2
- [27] Stepan Jezek, Martin Jonak, Radim Burget, Pavel Dvorak, and Milos Skotak. Deep learning-based defect detection of metal parts: evaluating current methods in complex conditions. In *2021 13th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, pages 66–71. IEEE, 2021. 2
- [28] Chun-Liang Li, Kihyuk Sohn, Jinsung Yoon, and Tomas Pfister. Cutpaste: Self-supervised learning for anomaly detection and localization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9664–9674, 2021. 3
- [29] Wei-Xin Li, Vijay Mahadevan, and Nuno Vasconcelos. Anomaly Detection and Localization in Crowded Scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 36(1):18–32, 2013. 2
- [30] Krzysztof Lis, Krishna Kanth Nakka, Pascal Fua, and Mathieu Salzmann. Detecting the unexpected via image resynthesis. In *IEEE International Conference on Computer Vision (ICCV)*, pages 2152–2161, 2019. 2
- [31] Wenqian Liu, Runze Li, Meng Zheng, Srikrishna Karanam, Ziyan Wu, Bir Bhanu, Richard J. Radke, and Octavia Camps. Towards visually explaining variational autoencoders. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8639–8648, 2020. 3
- [32] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 10012–10022, 2021. 1
- [33] Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. A convnet for the 2020s. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11976–11986, 2022. 1
- [34] Zhikang Liu, Yiming Zhou, Yuansheng Xu, and Zilei Wang. Simplenet: A simple network for image anomaly detection and localization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 20402–20411, June 2023. 3, 6
- [35] Cewu Lu, Jianping Shi, and Jiaya Jia. Abnormal Event Detection at 150 FPS in MATLAB. In *IEEE International Conference on Computer Vision (ICCV)*, pages 2720–2727, 2013. 2
- [36] Bjoern H. Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, et al. The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Transactions on Medical Imaging*, 34(10):1993–2024, 2015. 2
- [37] Pankaj Mishra, Riccardo Verk, Daniele Fornasier, Claudio Piciarelli, and Gian Luca Foresti. Vt-adl: A vision transformer network for image anomaly detection and localization. In *2021 IEEE 30th International Symposium on Industrial Electronics (ISIE)*, pages 01–06. IEEE, 2021. 2, 3
- [38] Paolo Napoletano, Flavio Piccoli, and Raimondo Schettini. Anomaly Detection in Nanofibrous Materials by CNN-Based Self-Similarity. *Sensors*, 18(1):209, 2018. 3
- [39] Tiago S Nazare, Rodrigo F de Mello, and Moacir A Ponti. Are pre-trained cnns good feature extractors for anomaly detection in surveillance videos? *arXiv preprint arXiv:1811.08495v1*, 2018. 3
- [40] Guansong Pang, Chunhua Shen, Longbing Cao, and Anton Van Den Hengel. Deep learning for anomaly detection: A review. *ACM Comput. Surv.*, 54(2), mar 2021. 2
- [41] Hyunjong Park, Jongyoun Noh, and Bumsub Ham. Learning memory-guided normality for anomaly detection. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14360–14369, 2020. 3
- [42] Pramuditha Perera, Ramesh Nallapati, and Bing Xiang. Ocgan: One-class novelty detection using gans with constrained latent representations. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2893–2901, 2019. 3
- [43] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016. 1
- [44] Danilo Rezende and Shakir Mohamed. Variational inference with normalizing flows. In *International conference on machine learning*, pages 1530–1538. PMLR, 2015. 3
- [45] Oliver Rippel., Arnav Chavan., Chucai Lei., and Dorit Merhof. Transfer learning gaussian anomaly detection by fine-tuning representations. In *Proceedings of the 2nd International Conference on Image Processing and Vision Engineering - IMPROVE*, pages 45–56. INSTICC, SciTePress, 2022. 3
- [46] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241. Springer International Publishing, 2015. 3

- 3 [23] Dan Hendrycks, Steven Basart, Mantas Mazeika, Andy Zou, Joseph Kwon, Mohammadreza Mostajedi 和 Dawn Song。扩展真实世界环境中的分布外检测。在 Kamalika Chaudhuri、Stefanie Jegelka、Le Song、Csaba Szepesvari、Gang Niu 和 Sivan Sabato 编辑的第 39 届国际机器学习会议论文集中，机器学习研究进展，第 162 卷，第 8759-8773 页。PMLR，2022 年 7 月 17-23 日。2
- [24] Gao Huang、Zhuang Liu、Laurens Van Der Maaten 和 Kilian Q Weinberger。密集连接的卷积网络。在 IEEE 计算机视觉与模式识别会议论文集中，第 4700-4708 页，2017 年。4
- [25] Yibin Huang、Congying Qiu、Yue Guo、Xiaonan Wang 和 奎元。磁砖表面缺陷显著性。2018 年 IEEE 第 14 届自动化科学与工程国际会议 (CASE) 论文集，第 612-617 页，2018 年。2
- [26] 杰里米·欧文、普拉纳夫·拉杰普拉卡尔、迈克尔·高、余一凡、西尔瓦娜·丘里亚·伊尔库斯、克里斯·丘特、亨里克·马克伦德、贝赫扎德·哈格胡、罗宾·鲍尔、凯蒂·什潘斯卡亚等人。CheXpert：一个具有不确定性标签和专家比较的大型胸部 X 光数据集。人工智能协会会议论文集，第 33 卷，第 590-597 页，2019 年。2
- [27] 斯特潘·耶泽克、马丁·乔纳克、拉迪姆·布尔格特、帕维尔·和 Milos Skotak。基于深度学习的金属零件缺陷检测：评估复杂条件下的当前方法。在 2021 年第 13 届超现代电信和控制系统及研讨会 (ICUMT) 上，第 66-71 页。IEEE，2021 年。2
- [28] Chun-Liang Li, Kihyuk Sohn, Jinsung Yoon, 和 Tomas Pfister。CutPaste：用于异常检测和定位的自监督学习。在 IEEE/CVF 计算机视觉与模式识别会议论文集中，第 9664-9674 页，2021 年。3
- [29] Wei-Xin Li, Vijay Mahadevan, 和 Nuno Vasconcelos。在拥挤场景中的异常检测和定位。IEEE 模式分析与机器智能交易 (TPAMI) , 36(1):18-32, 2013 年。2
- [30] Krzysztof Lis, Krishna Kanth Nakka, Pascal Fua, 和 Mathieu Salzmann。通过图像重合成检测意外情况。在 IEEE 国际计算机视觉会议 (ICCV) 上，第 2152-2161 页，2019 年。2
- [31] Wenqian Liu, Runze Li, Meng Zheng, Srikrishna Karanam, Ziyan Wu, Bir Bhanu, Richard J. Radke, 和 Octavia Camps。朝着视觉解释变分自动编码器的方向。发表于 2020 年 IEEE 计算机视觉与模式识别会议 (CVPR) 论文集，第 8639-8648 页。3
- [32] 刘泽、林雨桐、曹越、胡翰、魏逸轩、张政、林思贤和郭百宁。Swin transformer：使用移位窗口的分层视觉 transformer。发表于 2021 年 IEEE/CVF 国际计算机视觉会议论文集，第 10012-10022 页。1
- [33] 刘庄、毛汉子、吴朝元、克里斯托夫·费希特- enhofer、Trevor Darrell 和 Saining Xie。2020 年代的卷积网络。发表于 IEEE/CVF 计算机视觉与模式识别会议论文集，第 11976-11986 页，2022 年。1
- [34] 刘志康、周一鸣、徐元胜和王子雷。Simplenet：一个用于图像异常检测和定位的简单网络。发表于 IEEE/CVF 计算机视觉与模式识别会议 (CVPR) 论文集，第 20402-20411 页，2023 年 6 月。3, 6
- [35] 卢策吾、施剑平和贾佳亚。异常事件在 MATLAB 中以 150 FPS 进行检测。发表于 IEEE 国际计算机视觉会议 (ICCV) 论文集，第 2720-2727 页，2013 年。2
- [36] Bjoern H. Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby 等人。多模态脑肿瘤图像分割基准 (BRATS) 。IEEE 医学影像学报，34(10):1993-2024, 2015 年。2
- [37] Pankaj Mishra, Riccardo Verk, Daniele Fornasier, Claudio Piciarelli 和 Gian Luca Foresti。Vt-adl：用于图像异常检测和定位的视觉 transformer 网络。发表于 2021 年 IEEE 第 30 届工业电子国际研讨会 (ISIE) 论文集，第 01-06 页。IEEE，2021 年。2, 3
- [38] Paolo Napoletano、Flavio Piccoli 和 Raimondo Schettini。基于 CNN 自相似性的纳米纤维材料异常检测。传感器，18(1):209, 2018 年。3
- [39] Tiago S Nazare、Rodrigo F de Mello 和 Moacir A De Souza。预训练的 CNN 是否是监控视频中异常检测的良好特征提取器？arXiv 预印本 arXiv:1811.08495v1, 2018 年。3 [40] Guansong Pang, Chunhua Shen, Longbing Cao 和 Anton Van Den Hengel。深度学习用于异常检测：一个综述。ACM 计算机调查，54(2), 2021 年 3 月。2
- [41] Hyunjong Park, Jongyoun Noh 和 Bumsup Ham。学习记忆引导的正常性用于异常检测。在 IEEE 计算机视觉与模式识别会议 (CVPR) 上，第 14360-14369 页，2020 年。3
- [42] Pramuditha Perera、Ramesh Nallapati 和 Bing Xiang。Oc-gan：使用具有约束潜在表示的 GAN 进行一类新颖性检测。在 IEEE 计算机视觉与模式识别会议 (CVPR) 上，第 2893-2901 页，2019 年。3
- [43] Joseph Redmon、Santosh Divvala、Ross Girshick 和 Ali Farhadi。你只需看一次：统一的实时目标检测。在 IEEE 计算机视觉与模式识别会议论文集中，第 779-788 页，2016 年。1
- [44] Danilo Rezende 和 Shakir Mohamed。使用归一化流的变分推断。在国际机器学习会议上，第 1530-1538 页。PMLR，2015 年。3
- [45] Oliver Rippel、Arnav Chavan、Chucui Lei 和 Dorit Merhof。通过微调表示进行迁移学习高斯异常检测。在第二届国际图像处理和视觉工程会议论文集 - IMPROVE 中，第 45-56 页。INSTICC, SciTePress, 2022 年。3
- [46] Olaf Ronneberger、Philipp Fischer 和 Thomas Brox。U-net：用于生物医学图像分割的卷积网络。载于《医学图像计算与计算机辅助干预 – MICCAI 2015》，第 234-241 页。施普林格国际出版社，2015 年。3

- [47] 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, 3, 4, 6
- [48] Marco Rudolph, Bastian Wandt, and Bodo Rosenhahn. Same same but differnet: Semi-supervised defect detection with normalizing flows. In *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1906–1915, 2021. 3
- [49] Marco Rudolph, Tom Wehrbein, Bodo Rosenhahn, and Bastian Wandt. Fully convolutional cross-scale-flows for image-based defect detection. In *2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 1829–1838, 2022. 3
- [50] 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, 3, 4, 6
- [51] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3):211–252, 2015. 4
- [52] Mayu Sakurada and Takehisa Yairi. Anomaly detection using autoencoders with nonlinear dimensionality reduction. In *Proceedings of the MLSDA 2014 2nd Workshop on Machine Learning for Sensory Data Analysis*, MLSDA’14, page 4–11, New York, NY, USA, 2014. Association for Computing Machinery. 3
- [53] Mohammadreza Salehi, Niousha Sadjadi, Soroosh Baselizadeh, Mohammad H. Rohban, and Hamid R. Rabiee. Multiresolution knowledge distillation for anomaly detection. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14897–14907, 2021. 3
- [54] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4510–4520, 2018. 1
- [55] Thomas Schlegl, Philipp Seeböck, Sebastian M Waldstein, Ursula Schmidt-Erfurth, and Georg Langs. Unsupervised Anomaly Detection with Generative Adversarial Networks to Guide Marker Discovery. In *International Conference on Information Processing in Medical Imaging*, pages 146–157. Springer, 2017. 3
- [56] Thomas Schlegl, Philipp Seeböck, Sebastian Waldstein, Georg Langs, and Ursula Schmidt-Erfurth. f-AnoGAN: Fast Unsupervised Anomaly Detection with Generative Adversarial Networks. *Medical Image Analysis*, 54, 2019. 3
- [57] Abhinav Shrivastava, Abhinav Gupta, and Ross Girshick. Training region-based object detectors with online hard example mining. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 761–769, 2016. 4
- [58] Carsten Steger, Markus Ulrich, and Christian Wiedemann. *Machine Vision Algorithms and Applications*. Wiley-VCH, Weinheim, 2nd edition, 2018. 2, 8
- [59] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, pages 6105–6114. PMLR, 2019. 1
- [60] Mingxing Tan and Quoc Le. Efficientnetv2: Smaller models and faster training. In *International conference on machine learning*, pages 10096–10106. PMLR, 2021. 1
- [61] Mingxing Tan, Ruoming Pang, and Quoc V Le. Efficientdet: Scalable and efficient object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10781–10790, 2020. 1
- [62] Guodong Wang, Shumin Han, Errui Ding, and Di Huang. Student-teacher feature pyramid matching for anomaly detection. In *The British Machine Vision Conference (BMVC)*, 2021. 2, 3, 4
- [63] Hongxu Yin, Arash Vahdat, Jose M Alvarez, Arun Mallya, Jan Kautz, and Pavlo Molchanov. A-vit: Adaptive tokens for efficient vision transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10809–10818, 2022. 1
- [64] Jiawei Yu, Ye Zheng, Xiang Wang, Wei Li, Yushuang Wu, Rui Zhao, and Liwei Wu. Fastflow: Unsupervised anomaly detection and localization via 2d normalizing flows. *arXiv preprint arXiv:2111.07677v1*, 2021. 2, 3, 6
- [65] Sergey Zagoruyko and Nikos Komodakis. Wide residual networks. In Edwin R. Hancock Richard C. Wilson and William A. P. Smith, editors, *Proceedings of the British Machine Vision Conference (BMVC)*, pages 87.1–87.12. BMVA Press, September 2016. 3
- [66] Vitjan Zavrtanik, Matej Kristan, and Danijel Skočaj. Dsr—a dual subspace re-projection network for surface anomaly detection. In *Computer Vision—ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXXI*, pages 539–554. Springer, 2022. 3, 6
- [67] Yingying Zhang, Desen Zhou, Siqin Chen, Shenghua Gao, and Yi Ma. Single-image crowd counting via multi-column convolutional neural network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 589–597, 2016. 2
- [68] Bo Zong, Qi Song, Martin Renqiang Min, Wei Cheng, Cristian Lumezanu, Daeki Cho, and Haifeng Chen. Deep autoencoding gaussian mixture model for unsupervised anomaly detection. In *International conference on learning representations*, 2018. 3
- [69] Yang Zou, Jongheon Jeong, Latha Pemula, Dongqing Zhang, and Onkar Dabeer. Spot-the-difference self-supervised pre-training for anomaly detection and segmentation. In *Computer Vision—ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXX*, pages 392–408. Springer, 2022. 1, 2

- [47] Karsten Roth, Latha Pemula, Joaquin Zepeda, Bernhard Schölkopf、Thomas Brox 和 Peter Gehler。走向工业异常检测的全面回顾。发表于 IEEE/CVF 计算机视觉与模式识别会议论文集, 第 14318-14328 页, 2022 年。1, 2, 3, 4, 6
- [48] Marco Rudolph、Bastian Wandt 和 Bodo Rosenhahn。相同相同但不同: 使用归一化流进行半监督缺陷检测。发表于 2021 年 IEEE 冬季计算机视觉应用会议 (WACV) 论文集, 第 1906-1915 页, 2021 年。3
- [49] Marco Rudolph、Tom Wehrbein、Bodo Rosenhahn 和 Bastian Wandt。全卷积跨尺度流用于基于图像的缺陷检测。在 2022 年 IEEE/CVF 冬季计算机视觉应用会议 (WACV) 上, 第 1829-1838 页, 2022 年。3
- [50] Marco Rudolph, Tom Wehrbein, Bodo Rosenhahn, 和 Bastian Wandt。用于工业异常检测的非对称师生网络。在 IEEE/CVF 冬季计算机视觉应用会议论文集中, 第 2592-2602 页, 2023 年。1, 2, 3, 4, 6
- [51] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, jeev Satheesh、Sean Ma、Ziheng Huang、Andrey Karpathy、Aditya Khosla、Michael Bernstein、Alexander C. Berg 和 Li Fei-Fei。ImageNet 大规模视觉识别挑战赛。国际计算机视觉杂志, 115(3):211-252, 2015 年。4
- [52] Mayu Sakurada 和 Takehisa Yairi。使用具有非线性降维的自动编码器进行异常检测。在 MLSDA 2014 第二届机器学习用于感知数据分析研讨会论文集中, MLSDA'14, 第 4-11 页, 纽约, 纽约州, 美国, 2014 年。计算机协会。3
- [53] Mohammadreza Salehi、Niousha Sadjadi、Soroosh Baselizadeh、Mohammad H. Rohban 和 Hamid R. Rabiee。多分辨率知识蒸馏用于异常检测。发表于 IEEE 计算机视觉与模式识别会议 (CVPR) , 第 14897-14907 页, 2021 年。3
- [54] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey moginov 和 Liang-Chieh Chen。MobileNetV2: 倒置残差和线性瓶颈。发表于 IEEE 计算机视觉与模式识别会议论文集, 第 4510-4520 页, 2018 年。1
- [55] Thomas Schlegl, Philipp Seeböck, Sebastian M Waldstein, Ursula Schmidt-Erfurth 和 Georg Langs。使用生成对抗网络进行无监督异常检测以指导标记发现。发表于国际医学图像信息处理会议, 第 146-157 页。Springer 出版社, 2017 年。3
- [56] Thomas Schlegl, Philipp Seeböck, Sebastian Waldstein, Georg Langs 和 Ursula Schmidt-Erfurth。f-AnoGAN: 使用生成对抗网络进行快速无监督异常检测。医学图像分析, 54 卷, 2019 年。3
- [57] Abhinav Shrivastava, Abhinav Gupta, 和 Ross Girshick。使用在线困难样本挖掘训练基于区域的目标检测器。发表于 IEEE 计算机视觉与模式识别会议论文集, 第 761-769 页, 2016 年。4
- [58] Carsten Steger, Markus Ulrich 和 Christian Wiedemann。机器视觉算法与应用。Wiley-VCH 出版社, 魏海姆, 第 2 版, 2018 年。2, 8 [59] Mingxing Tan 和 Quoc Le。EfficientNet: 重新思考模型卷积神经网络的缩放。在国际机器学习会议上, 第 6105-6114 页。PMLR, 2019 年。1
- [60] Mingxing Tan 和 Quoc Le。EfficientNetV2: 更小的模型和更快的训练。在国际机器学习会议上, 第 10096-10106 页。PMLR, 2021 年。1 [61] Mingxing Tan, Ruoming Pang 和 Quoc V Le。EfficientDet: 可扩展和高效的目标检测。发表于 IEEE/CVF 计算机视觉与模式识别会议论文集, 第 10781-10790 页, 2020 年。1
- [62] 王国栋、韩树民、丁二锐和黄迪。用于异常检测的学生-教师特征金字塔匹配。发表于英国机器视觉会议 (BMVC) , 2021 年。2, 3, 4
- [63] 尹宏旭、Arash Vahdat、Jose M Alvarez、Arun Mallya, Jan Kautz 和 Pavlo Molchanov。A-vit: 用于高效视觉 Transformer 的自适应令牌。发表于 IEEE/CVF 计算机视觉与模式识别会议论文集, 第 10809-10818 页, 2022 年。1
- [64] 于佳伟、郑烨、王翔、李伟、吴玉双、赵睿和吴立伟。FastFlow: 通过二维归一化流进行无监督异常检测和定位。arXiv 预印本 arXiv:2111.07677v1, 2021 年。2,
- [65] Sergey Zagoruyko 和 Nikos Komodakis。宽残差网络-作品。在 Edwin R. Hancock、Richard C. Wilson 和 William A. P. Smith 编辑的《英国机器视觉会议 (BMVC) 论文集》中, 第 87.1-87.12 页。BMVA 出版社, 2016 年 9 月。3
- [66] Vitjan Zavrtanik、Matej Kristan 和 Danijel Skočaj。DSR-一种用于表面异常检测的双子空间重投影网络。在《计算机视觉-ECCV 2022: 第 17 届欧洲会议, 以色列特拉维夫, 2022 年 10 月 23-27 日, 会议记录, 第 XXXI 部分》中, 第 539-554 页。施普林格出版社, 2022 年。3, 6
- [67] Yingying Zhang、Desen Zhou、Siqin Chen、Shenghua Gao, 和易马。通过多列卷积神经网络进行单图像人群计数。发表于 IEEE 计算机视觉与模式识别会议论文集, 第 589-597 页, 2016 年。2
- [68] 宗博、宋琦、闵仁强、程伟、克里斯蒂安·卢梅扎努、赵大基和陈海峰。用于无监督异常检测的深度自编码高斯混合模型。发表于国际学习表示会议, 2018 年。3
- [69] 邹阳、郑钟宪、拉塔·佩穆拉、张冬青、Onkar Dabeer 和 Onkar Dabeer。用于异常检测和分割的不同自监督预训练。发表于《计算机视觉-ECCV 2022: 第 17 届欧洲会议, 以色列特拉维夫, 2022 年 10 月 23-27 日, 会议记录, 第 XXX 部分》, 第 392-408 页。Springer 出版社, 2022 年。1, 2