Attention Guided Anomaly Localization in Images

Shashanka Venkataramanan*[0000−0003−1096−1342], Kuan-Chuan Peng†[0000−0002−2682−9912], Rajat Vikram Singh‡[0000−0002−1416−8344], and Abhijit Mahalanobis*[0000−0002−2782−8655]

*Center for Research in Computer Vision, University of Central Florida, Orlando, FL
†Mitsubishi Electric Research Laboratories, Cambridge, MA
‡Siemens Corporate Technology, Princeton, NJ
shashankv@Knights.ucf.edu, kpeng@merl.com, singh.rajat@siemens.com, amahalan@crcv.ucf.edu

Abstract. Anomaly localization is an important problem in computer vision which involves localizing anomalous regions within images with applications in industrial inspection, surveillance, and medical imaging. This task is challenging due to the small sample size and pixel coverage of the anomaly in real-world scenarios. Most prior works need to use anomalous training images to compute a class-specific threshold to localize anomalies. Without the need of anomalous training images, we propose Convolutional Adversarial Variational autoencoder with Guided Attention (CAVGA), which localizes the anomaly with a *convolutional latent variable* to preserve the spatial information. In the unsupervised setting, we propose an *attention expansion loss* where we encourage CAVGA to focus on all normal regions in the image. Furthermore, in the weakly-supervised setting we propose a *complementary guided attention loss*, where we encourage the attention map to focus on all normal regions while minimizing the attention map corresponding to anomalous regions in the image. CAVGA outperforms the state-of-the-art (SOTA) anomaly localization methods on MVTec Anomaly Detection (MVTAD), modified ShanghaiTech Campus (mSTC) and Large-scale Attention based Glaucoma (LAG) datasets in the unsupervised setting and when using only 2% anomalous images in the weakly-supervised setting. CAVGA also outperforms SOTA anomaly detection methods on the MNIST, CIFAR-10, Fashion-MNIST, MVTAD, mSTC and LAG datasets.

Keywords: guided attention, anomaly localization, convolutional adversarial variational autoencoder

1 Introduction

Recognizing whether an image is homogeneous with its previously observed distribution or whether it belongs to a novel or anomalous distribution has been identified as an important problem [5]. In this work, we focus on a related task, anomaly localization in images, which involves segmenting the anomalous regions

Attention Guided Anomaly Localization in Images

沙山卡·文卡塔拉马南⋆[0000−0003−1096−1342]、关川·彭†[0000−0002−2682−9912]、拉贾特·维克拉姆·辛格‡[0000−0002−1416−8344]和阿比吉特·马哈拉诺比斯⋆[0000−0002−2782−8655]

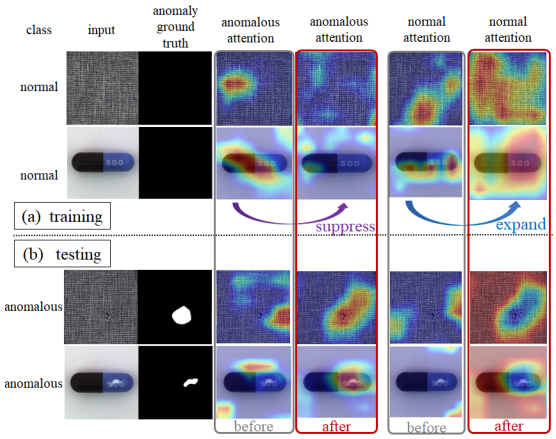
⋆中佛罗里达大学计算机视觉研究中心，奥兰多，佛罗里达州 †三菱电机研究实验室，剑桥，马萨诸塞州 ‡西门子企业技术部门，普林斯顿，新泽西州
shashankv@Knights.ucf.edu, kpeng@merl.com, singh.rajat@siemens.com, amahalan@crcv.ucf.edu

Abstract. 异常定位是计算机视觉中的一个重要问题，涉及在图像中定位异常区域，应用于工业检测、监控和医学成像等领域。由于现实场景中异常样本数量少且像素覆盖范围小，该任务具有挑战性。大多数先前工作需要利用异常训练图像计算类别特定阈值以定位异常。我们提出的卷积对抗变分自编码器与引导注意力机制（CAVGA）无需异常训练图像，通过*convolutional latentvariable*来保留空间信息以定位异常。在无监督设置中，我们提出一种*attention expansion loss*，促使CAVGA聚焦图像中所有正常区域。此外，在弱监督设置中，我们提出*complementary guided attention loss*，引导注意力图聚焦所有正常区域，同时最小化图像中异常区域对应的注意力图。在无监督设置及仅使用2%异常图像的弱监督设置中，CAVGA在MVTec异常检测数据集（MVTAD）、改进版上海科技大学校园数据集（mSTC）和大规模注意力青光眼数据集（LAG）上的异常定位性能均优于当前最先进方法。CAVGA在MNIST、CIFAR-10、Fashion-MNIST、MVTAD、mSTC和LAG数据集上的异常检测性能也超越了现有最优方法。

Keywords: 引导注意力，异常定位，卷积对抗变分自编码器

1 Introduction

识别图像是否与其先前观察到的分布同质，或是否属于新颖或异常分布，已被确认为一个重要问题[5]。在这项工作中，我们专注于一个相关任务——图像中的异常定位，这涉及对异常区域进行分割。



(i) CAVGA main idea

dataset	task	improvement
MVTAD [5]	l	(9/15; 4∼85%)
MVTAD [5]	d	(9/15; 2∼30%)
mSTC [31]	l	(7/12; 2∼42%)
mSTC [31]	d	(8/12; 1∼38%)
LAG [29]	l	(1/1; 16%)
LAG [29]	d	(1/1; 1.1%)
MNIST [27]	d	(8/10; 0.1∼2.5%)
CIFAR-10 [25]	d	(7/10; 3∼31%)
F-MNIST [57]	d	(8/10; 2∼24%)

l: localization; d: detection
F-MNIST: Fashion-MNIST [57]

- metric for l : IoU
- metric for d in MVTAD, mSTC, and LAG: classification accuracy
- metric for d in MNIST, CIFAR-10 and Fashion-MNIST: area under ROC curve

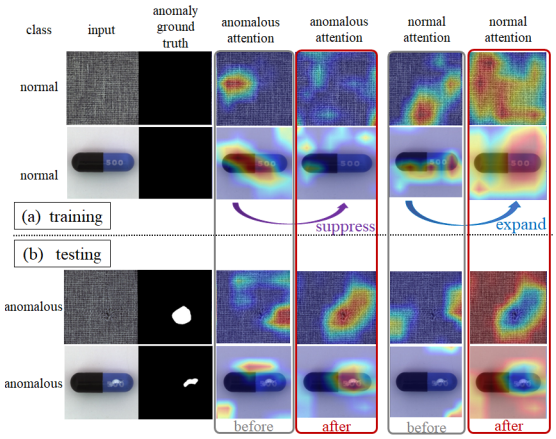
(ii) improvement summary

Fig. 1: (i) CAVGA uses the proposed complementary guided attention loss to encourage the attention map to cover the entire normal regions while suppressing the attention map corresponding to anomalous class in the training image. This enables the trained network to generate the anomalous attention map to localize the anomaly better at testing (ii) CAVGA's improvement over SOTA in the form of (number of outperforming/total categories; improvement (%) in its metric)

within them. Anomaly localization has been applied in industrial inspection settings to segment defective product parts [5], in surveillance to locate intruders [38], in medical imaging to segment tumor in brain MRI or glaucoma in retina images [4, 29], etc. There has been an increase in analysis towards segmenting potential anomalous regions in images as acknowledged in [13].

Existing state-of-the-art (SOTA) anomaly localization methods [6, 47] are based on deep learning. However, developing deep learning based algorithms for this task can be challenging due to the small pixel coverage of the anomaly and lack of suitable data since images with anomalies are rarely available in real-world scenarios [5]. Existing SOTA methods tackle this challenge using autoencoders [15, 47] and GAN based approaches [3, 43, 59], which use a thresholded pixel-wise difference between the input and reconstructed image to localize anomalies. But, their methods need to determine class-specific thresholds using anomalous training images which can be unavailable in real-world scenarios.

To tackle these drawbacks of using anomalous training images, we propose Convolutional Adversarial Variational autoencoder with Guided Attention (CAVGA), an unsupervised anomaly localization method which requires no anomalous training images. CAVGA comprises of a *convolutional latent variable* to preserve the spatial relation between the input and latent variable. Since real-world applications may have access to only limited training data [5], we propose to localize the anomalies by using supervision on attention maps. This



(i) CAVGA main idea

dataset	task	improvement
MVTAD [5]	l	(9/15; 4∼85%)
MVTAD [5]	d	(9/15; 2∼30%)
mSTC [31]	l	(7/12; 2∼42%)
mSTC [31]	d	(8/12; 1∼38%)
LAG [29]	l	(1/1; 16%)
LAG [29]	d	(1/1; 1.1%)
MNIST [27]	d	(8/10; 0.1∼2.5%)
CIFAR-10 [25]	d	(7/10; 3∼31%)
F-MNIST [57]	d	(8/10; 2∼24%)

l: localization; d: detection
F-MNIST: Fashion-MNIST [57]

- metric for l : IoU
- metric for d in MVTAD, mSTC, and LAG: classification accuracy
- metric for d in MNIST, CIFAR-10 and Fashion-MNIST: area under ROC curve

(ii) improvement summary

图1：(i) CAVGA采用提出的互补引导注意力损失，促使注意力图覆盖训练图像中的全部正常区域，同时抑制与异常类别对应的注意力图。这使得训练后的网络能够生成异常注意力图，从而在测试时更好地定位异常。(ii) CAVGA相对于SOTA的改进表现为（表现优异的类别数/总类别数；其指标提升百分比）

其中。异常定位已应用于工业检测场景以分割有缺陷的产品部件[5]，在监控中用于定位入侵者[38]，在医学影像中用于分割脑部MRI中的肿瘤或视网膜图像中的青光眼[4, 29]等。正如[13]所指出的，针对图像中潜在异常区域分割的分析研究正日益增多。

现有的最先进（SOTA）异常定位方法[6, 47]基于深度学习。然而，由于异常区域的像素覆盖范围较小，且缺乏合适的数据——因为在真实场景中很少能获得包含异常的图像[5]，为此任务开发基于深度学习的算法可能具有挑战性。现有的SOTA方法通过使用自编码器[15, 47]和基于GAN的方法[3, 43, 59]来应对这一挑战，这些方法利用输入图像与重建图像之间基于阈值的逐像素差异来定位异常。但是，它们的方法需要使用异常训练图像来确定类别特定的阈值，而这在真实场景中可能无法获得。

为了解决使用异常训练图像的这些缺点，我们提出了带有引导注意力的卷积对抗变分自编码器（CAVGA），这是一种无需异常训练图像的无监督异常定位方法。CAVGA包含一个*convolutional latent vari- able*，以保持输入与潜在变量之间的空间关系。由于实际应用可能只能获取有限的训练数据[5]，我们提出通过对注意力图施加监督来定位异常。这

is motivated by the finding in [28] that attention based supervision can alleviate the need of using large amount of training data. Intuitively, without any prior knowledge of the anomaly, humans need to look at the entire image to identify the anomalous regions. Based on this idea, we propose an *attention expansion loss* where we encourage the network to generate an attention map that focuses on all normal regions of the image.

Since annotating segmentation training data can be laborious [22], in the case when the annotator provides few anomalous training images without ground truth segmented anomalous regions, we extend CAVGA to a weakly supervised setting. Here, we introduce a classifier in CAVGA and propose a *complementary guided attention loss* computed only for the normal images correctly predicted by the classifier. Using this complementary guided attention loss, we expand the normal attention but suppress the anomalous attention on the normal image, where normal/anomalous attention represents the areas affecting the classifier's normal/anomalous prediction identified by existing network visualization methods (e.g. Grad-CAM [49]). Fig. 1 (i) (a) illustrates our attention mechanism during training, and Fig. 1 (i) (b) demonstrates that the resulting normal attention and anomalous attention on the anomalous testing images are visually complementary, which is consistent with our intuition. Furthermore, Fig. 1 (ii) summarizes CAVGA's ability to outperform SOTA methods in anomaly localization on industrial inspection (MVTAD) [5], surveillance (mSTC) [31] and medical imaging (LAG) [29] datasets. We also show CAVGA's ability to outperform SOTA methods in anomaly detection on common benchmarks.

To the best of our knowledge, we are the first in anomaly localization to propose an end-to-end trainable framework with attention guidance which explicitly enforces the network to learn representations from the entire normal image. As compared to the prior works, our proposed approach CAVGA needs no anomalous training images to determine a class-specific threshold to localize the anomaly. Our contributions are:

- **An attention expansion loss (L_{ae})**, where we encourage the network to focus on the entire normal images in the unsupervised setting.
- **A complementary guided attention loss (L_{cga})**, which we use to minimize the anomalous attention and simultaneously expand the normal attention for the normal images correctly predicted by the classifier.
- **New SOTA**: In anomaly localization, CAVGA outperforms SOTA methods on the MVTAD and mSTC datasets in IoU and mean Area under ROC curve (AuROC) and also outperforms SOTA anomaly localization methods on LAG dataset in IoU. We also show CAVGA's ability to outperform SOTA methods for anomaly detection on the MVTAD, mSTC, LAG, MNIST [27], CIFAR-10 [25] and Fashion-MNIST [57] datasets in classification accuracy.

2 Related Works

Often used interchangeably, the terms anomaly localization and anomaly segmentation involve pixel-accurate segmentation of anomalous regions within an

这一思路的提出，源于文献[28]的发现：基于注意力的监督机制能够减少对大量训练数据的需求。直观而言，在没有异常先验知识的情况下，人类需要观察整幅图像才能定位异常区域。基于此，我们提出了一种*attention expansionloss*方法，通过促使网络生成覆盖图像中所有正常区域的注意力图来实现这一目标。

由于标注分割训练数据可能非常耗时[22]，在标注者仅提供少量无真实分割异常区域标注的异常训练图像的情况下，我们将CAVGA扩展至弱监督设置。在此，我们在CAVGA中引入分类器，并提出一种仅针对被分类器正确预测的正常图像计算的*complementaryguided attention loss*。通过这种互补引导注意力损失，我们在正常图像上扩展正常注意力同时抑制异常注意力——其中正常/异常注意力指代现有网络可视化方法（如Grad-CAM[49]）所识别的、影响分类器正常/异常预测的区域。图1(i)(a)展示了训练过程中我们的注意力机制，图1(i)(b)则表明在异常测试图像上生成的正常注意力与异常注意力在视觉上具有互补性，这与我们的直觉一致。此外，图1(ii)总结了CAVGA在工业检测（MVTAD）[5]、监控（mSTC）[31]和医学影像（LAG）[29]数据集的异常定位任务中超越现有前沿方法的能力。我们同时展示了CAVGA在常见基准测试的异常检测任务中优于前沿方法的表现。

据我们所知，我们是异常定位领域首个提出端到端可训练注意力引导框架的研究，该框架明确强制网络从完整的正常图像中学习表征。与先前工作相比，我们提出的CAVGA方法无需异常训练图像即可确定类别特定阈值以定位异常。我们的贡献如下：

– **An attention expansion loss (L_{ae})**，在此无监督设置中，我们鼓励网络专注于完整的正常图像。
– **A complementary guided attention loss (L_{cga})**，我们用它来最小化异常注意力，同时为被分类器正确预测的正常图像扩展正常注意力。
– **New SOTA**：在异常定位方面，CAVGA在MVTAD和mSTC数据集上的IoU和平均ROC曲线下面积（AuROC）指标超越了当前最先进方法，并且在LAG数据集上的IoU指标也优于最先进的异常定位方法。我们还展示了CAVGA在MVTAD、mSTC、LAG、MNIST[27]、CIFAR-10[25]和Fashion-MNIST[57]数据集上的异常检测分类准确率超越当前最先进方法的能力。

2 Related Works

异常定位与异常分割这两个术语常被互换使用，它们都涉及对图像中异常区域进行像素级精确分割。

Table 1: Comparison between CAVGA and other anomaly localization methods in the unsupervised setting in terms of the working properties. Among all the listed methods, only CAVGA satisfies all the listed properties

Does the method satisfy each property?	[3, 48] [6, 43]	[4]	[47]	[54] [50]	[13, 32] [2]	CAVGA
not using anomalous training images	N	N	Y	Y	Y	Y
localize **multiple** modes of anomalies	Y	N	N	N	Y	Y
pixel (not patch) based localization	Y	Y	N	Y	Y	Y
use **convolutional latent variable**	N	Y	N	N	N	Y

image [5]. They have been applied to industrial inspection settings to segment defective product parts [5], medical imaging to segment glaucoma in retina images [29], etc. Image based anomaly localization has not been fully studied as compared to anomaly detection, where methods such as [3, 4, 6, 43, 48] employ a thresholded pixel wise difference between the input and reconstructed image to segment the anomalous regions. [47] proposes an inpainter-detector network for patch-based localization in images. [13] proposes gradient descent on a regularized autoencoder while Liu *et al.* [32] (denoted as ADVAE) generate gradient based attention maps from the latent space of the trained model. We compare CAVGA with the existing methods relevant to anomaly localization in the unsupervised setting in Table 1 and show that among the listed methods, only CAVGA shows all the listed properties.

Anomaly detection involves determining an image as normal or anomalous [3]. One-class classification and anomaly detection are related to novelty detection [41] which has been widely studied in computer vision [3, 20, 35, 37, 53] and applied to video analysis [10], remote sensing [36], etc. With the advance in GANs [17], SOTA methods perform anomaly detection by generating realistic normal images during training [21, 22, 42, 46, 48]. [12] proposes to search the latent space of the generator for detecting anomalies. [41] introduces latent-space-sampling-based network with information-negative mining while [30] proposes normality score function based on capsule network's activation and reconstruction error. [2] proposes a deep autoencoder that learns the distribution of latent representation through autoregressive procedure. Unlike [7, 11, 44, 55] where anomalous training images are used for anomaly detection, CAVGA does not need anomalous training images.

3 Proposed Approach: CAVGA

3.1 Unsupervised Approach: CAVGA$_u$

Fig. 2 (a) illustrates CAVGA in the unsupervised setting (denoted as CAVGA$_u$). CAVGA$_u$ comprises of a convolutional latent variable to preserve the spatial information between the input and latent variable. Since attention maps obtained from feature maps illustrate the regions of the image responsible for specific

表1：在无监督设置下，CAVGA与其他异常定位方法在工作特性方面的比较。在所有列出的方法中，只有CAVGA满足所有列出的特性。

Does the method satisfy each property?	[3, 48] [6, 43]	[4]	[47]	[54] [50]	[13, 32] [2]	CAVGA
not using anomalous training images	**N**	**N**	**Y**	**Y**	**Y**	**Y**
localize **multiple** modes of anomalies	**Y**	**N**	**N**	**N**	**Y**	**Y**
pixel (not patch) based localization	**Y**	**Y**	**N**	**Y**	**Y**	**Y**
use **convolutional latent variable**	**N**	**Y**	**N**	**N**	**N**	**Y**

图像[5]。它们已被应用于工业检测场景以分割有缺陷的产品部件[5]，医学影像中用于分割视网膜图像中的青光眼[29]等。与异常检测相比，基于图像的异常定位尚未得到充分研究，其中如[3, 4, 6, 43, 48]等方法利用输入图像与重建图像之间的阈值化像素级差异来分割异常区域。[47]提出了一种用于图像中基于补丁定位的修复-检测网络。[13]提出在正则化自编码器上进行梯度下降，而Liu *et al.* [32]（记为ADVAE）则从训练模型的潜在空间生成基于梯度的注意力图。我们在表1中将CAVGA与无监督设置下异常定位相关的现有方法进行比较，结果表明，在所列方法中，仅CAVGA同时具备所有列出的特性。

异常检测涉及将图像判定为正常或异常[3]。单分类与异常检测均与新颖性检测相关[41]，该领域在计算机视觉中已得到广泛研究[3, 20, 35, 37, 53]，并应用于视频分析[10]、遥感[36]等领域。随着GANs的发展[17]，当前最优方法通过在训练中生成逼真正常图像进行异常检测[21, 22, 42, 46, 48]。[12]提出通过搜索生成器的潜在空间来检测异常。[41]引入了基于潜在空间采样的网络并结合信息负挖掘，而[30]提出基于胶囊网络激活与重构误差的常态评分函数。[2]提出通过自回归过程学习潜在表示分布的深度自编码器。与使用异常训练图像进行检测的方法[7, 11, 44, 55]不同，CAVGA无需异常训练图像。

3 Proposed Approach: CAVGA

3.1 Unsupervised Approach: CAVGA$_u$

图2 (a) 展示了无监督设置下的CAVGA（记为CAVGA$_u$）。CAVGA$_u$包含一个卷积潜变量，用于保持输入与潜变量之间的空间信息。由于从特征图获得的注意力图能说明图像中对特定区域负责的

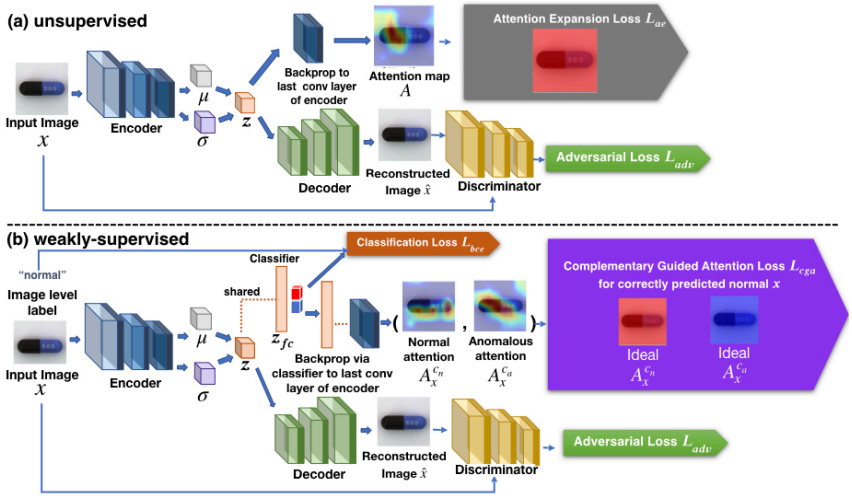


Fig. 2: (a) The framework of CAVGA$_u$ where the attention expansion loss L_{ae} guides the attention map A computed from the latent variable z to cover the entire normal image. (b) Illustration of CAVGA$_w$ with the complementary guided attention loss L_{cga} to minimize the anomalous attention $A_x^{c_a}$ and expand the normal attention $A_x^{c_n}$ for the normal images correctly predicted by the classifier

activation of neurons in the feature maps [58], we propose an attention expansion loss such that the feature representation of the latent variable encodes all the normal regions. This loss encourages the attention map generated from the latent variable to cover the entire normal training image as illustrated in Fig. 1 (i) (a). During testing, we localize the anomaly from the areas of the image that the attention map does not focus on.

Convolutional latent variable Variational Autoencoder (VAE) [23] is a generative model widely used for anomaly detection [24, 40]. The loss function of training a vanilla VAE can be formulated as:

$$L = L_R(x, \hat{x}) + KL(q_\phi(z|x)||p_\theta(z|x)), \tag{1}$$

where $L_R(x, \hat{x}) = \frac{-1}{N}\sum_{i=1}^{N} x_i log(\hat{x}_i) + (1-x_i) log(1-\hat{x}_i)$ is the reconstruction loss between the input (x) and reconstructed images ($\hat{x}$), and N is the total number of images. The posterior $p_\theta(z|x)$ is modeled using a standard Gaussian distribution prior $p(z)$ with the help of Kullback-Liebler (KL) divergence through $q_\phi(z|x)$. Since the vanilla VAE results in blurry reconstruction [26], we use a discriminator ($D(.)$) to improve the stability of the training and generate sharper reconstructed images $\hat{x}$ using adversarial learning [34] formulated as follows:

$$L_{adv} = -\frac{1}{N}\sum_{i=1}^{N} log(D(x_i)) + log(1 - D(\hat{x}_i)) \tag{2}$$

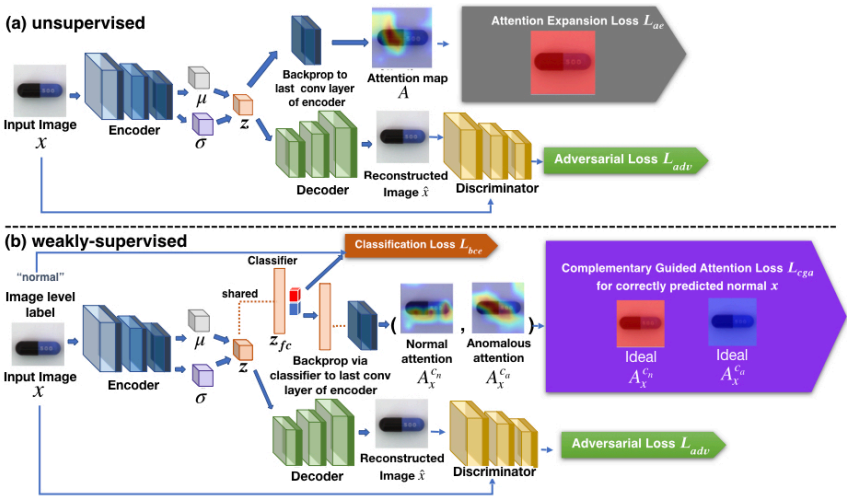


图2：(a) CAVGA$_u$的框架，其中注意力扩展损失L_{ae}引导从潜变量z计算得到的注意力图A覆盖整个正常图像。(b) CAVGA$_w$示意图，其通过互补引导注意力损失L_{cga}来最小化异常注意力$A_x^{c_a}$，并对被分类器正确预测的正常图像扩展正常注意力$A_x^{c_n}$

在特征图中神经元的激活[58]基础上，我们提出了一种注意力扩展损失，使得潜在变量的特征表示能够编码所有正常区域。该损失促使从潜在变量生成的注意力图覆盖整个正常训练图像，如图1(i)(a)所示。在测试过程中，我们通过注意力图未聚焦的图像区域来定位异常。

Convolutional latent variable 变分自编码器（VAE）[23]是一种广泛用于异常检测的生成模型[24, 40]。训练基础VAE的损失函数可表述为：

$$L = L_R(x, \hat{x}) + KL(q_\phi(z|x)||p_\theta(z|x)), \tag{1}$$

其中 $L_R(x,\hat{x}) = \frac{-1}{N}\sum_{i=1}^{N} x_i log(\hat{x}_i) + (1-x_i) log(1-\hat{x}_i)$ 是输入图像 (x) 与重建图像 $(\hat{x})$ 之间的重建损失，N 为图像总数。后验分布 $p_\theta(z|x)$ 通过 $q_\phi(z|x)$ 借助 Kullback-Liebler (KL) 散度，以标准高斯分布先验 $p(z)$ 进行建模。由于原始 VAE 会导致重建图像模糊 [26]，我们采用判别器 $(D(.))$ 通过对抗学习 [34] 提升训练稳定性并生成更清晰的重建图像 $\hat{x}$，其公式如下：

$$L_{adv} = -\frac{1}{N}\sum_{i=1}^{N} log(D(x_i)) + log(1 - D(\hat{x}_i)) \tag{2}$$

Unlike traditional autoencoders [6, 18] where the latent variable is flattened, inspired from [4], we use a convolutional latent variable to preserve the spatial relation between the input and the latent variable.

Attention expansion loss L_{ae} The main contribution of our work involves using supervision on attention maps to spatially localize the anomaly in the image. Most methods [3, 48, 53] employ a thresholded pixel-wise difference between the reconstructed image and the input image to localize the anomaly where the threshold is determined by using anomalous training images. However, CAVGA$_u$ learns to localize the anomaly using an attention map reflected through an end-to-end training process without the need of any anomalous training images. We use the feature representation of the latent variable z to compute the attention map (A). A is computed using Grad-CAM [49] such that $A_{i,j} \in [0, 1]$, where $A_{i,j}$ is the (i, j) element of A.

Intuitively, A obtained from feature maps focuses on the regions of the image based on the activation of neurons in the feature maps and its respective importance [58, 60]. Due to the lack of prior knowledge about the anomaly, in general, humans need to look at the entire image to identify anomalous regions. We use this notion to learn the feature representation of the entire normal image by proposing an attention expansion loss, where we encourage the network to generate an attention map covering all the normal regions. This attention expansion loss for each image $L_{ae,1}$ is defined as:

$$L_{ae,1} = \frac{1}{|A|} \sum_{i,j} (1 - A_{i,j}) \tag{3}$$

where $|A|$ is the total number of elements in A. The final attention expansion loss L_{ae} is the average of $L_{ae,1}$ over the N images. Since the idea of attention mechanisms involves locating the most salient regions in the image [29] which typically does not cover the entire image, we use L_{ae} as an additional supervision on the network, such that the trained network generates an attention map that covers all the normal regions. Fig. 1 (i) (a) shows that before using L_{ae} i.e. training CAVGA$_u$ only with adversarial learning ($L_{adv} + L$) does not encode all the normal regions into the latent variable, and that the attention map fails to cover the entire image, which is overcome after using L_{ae}. Furthermore, supervising on attention maps prevents the trained model to make inference based on incorrect areas and also alleviates the need of using large amount of training data as shown in [28], which is not explicitly enforced in existing methods [3, 6, 47].

We form the final objective function L_{final} below:

$$L_{final} = w_r L + w_{adv} L_{adv} + w_{ae} L_{ae}, \tag{4}$$

where w_r, w_{adv}, and w_{ae} are empirically set as $1, 1$, and 0.01 respectively.

During testing, we feed an image x_{test} into the encoder followed by the decoder, which reconstructs an image $\hat{x}_{test}$. As defined in [48], we compute the pixel-wise difference between $\hat{x}_{test}$ and x_{test} as the anomalous score s_a. Intuitively, if x_{test} is drawn from the learnt distribution of z, then s_a is small. Without

与传统将潜在变量展平的自编码器[6, 18]不同，受[4]启发，我们采用卷积潜在变量以保持输入与潜在变量{v*}之间的空间关系。

Attention expansion loss L_{ae} 我们工作的主要贡献在于利用注意力图的监督来在图像中空间定位异常。大多数方法[3, 48, 53]采用重建图像与输入图像之间的阈值化像素差异来定位异常，其阈值需借助异常训练图像确定。然而，CAVGA$_u$通过端到端训练过程中反映的注意力图，无需任何异常训练图像即可学习定位异常。我们利用潜变量z的特征表示来计算注意力图(A)。A通过Grad-CAM[49]计算，使得$A_{i,j} \in$满足[0, 1]，其中$A_{i,j}$是A的(i, j)元素。

直观上，从特征图获得的A基于特征图中神经元的激活及其各自的重要性[58, 60]，聚焦于图像的区域。由于缺乏关于异常的先前知识，通常人类需要观察整个图像以识别异常区域。我们利用这一概念，通过提出一种注意力扩展损失来学习整个正常图像的特征表示，该损失鼓励网络生成覆盖所有正常区域的注意力图。每张图像$L_{ae,1}$的注意力扩展损失定义为:

$$L_{ae,1} = \frac{1}{|A|} \sum_{i,j} (1 - A_{i,j}) \tag{3}$$

其中$|A|$是A中元素的总数。最终的注意力扩展损失L_{ae}是$L_{ae,1}$在N张图像上的平均值。由于注意力机制的核心思想在于定位图像中最显著的区域[29]，这些区域通常不会覆盖整个图像，因此我们使用L_{ae}作为对网络的额外监督，使得训练后的网络能生成覆盖所有正常区域的注意力图。图1(i)(a)显示，在使用L_{ae}之前（即仅通过对抗学习($L_{adv} + L$)训练CAVGA$_u$时），模型未能将所有正常区域编码到潜变量中，且注意力图未能覆盖整个图像；这一问题在使用L_{ae}后得到解决。此外，对注意力图进行监督能防止训练后的模型基于错误区域进行推断，并如[28]所示减少对大量训练数据的需求，而现有方法[3,6,47]并未明确强制执行这一点。

我们构建最终的目标函数 L_{final} 如下:

$$L_{final} = w_r L + w_{adv} L_{adv} + w_{ae} L_{ae}, \tag{4}$$

其中w_r、w_{adv}和w_{ae}根据经验分别设置为1,1和0.01。

在测试过程中，我们将图像x_{test}输入编码器，随后通过解码器重构出图像$\hat{x}_{test}$。根据[48]的定义，我们计算$\hat{x}_{test}$与x_{test}之间的像素级差异作为异常分数s_a。直观而言，若x_{test}来自已学习的z分布，则s_a值较小。若无

using any anomalous training images in the unsupervised setting, we normalize s_a between $[0, 1]$ and empirically set 0.5 as the threshold to detect an image as anomalous. The attention map A_{test} is computed from z using Grad-CAM and is inverted ($\mathbf{1}$ - A_{test}) to obtain an anomalous attention map which localizes the anomaly. Here, $\mathbf{1}$ refers to a matrix of all ones with the same dimensions as A_{test}. We empirically choose 0.5 as the threshold on the anomalous attention map to evaluate the localization performance.

3.2 Weakly Supervised Approach: CAVGA$_w$

CAVGA$_u$ can be further extended to a weakly supervised setting (denoted as CAVGA$_w$) where we explore the possibility of using few anomalous training images to improve the performance of anomaly localization. Given the labels of the anomalous and normal images without the pixel-wise annotation of the anomaly during training, we modify CAVGA$_u$ by introducing a binary classifier C at the output of z as shown in Fig. 2 (b) and train C using the binary cross entropy loss L_{bce}. Given an image x and its ground truth label y, we define $p \in \{c_a, c_n\}$ as the prediction of C, where c_a and c_n are anomalous and normal classes respectively. From Fig. 2 (b) we clone z into a new tensor, flatten it to form a fully connected layer z_{fc}, and add a 2-node output layer to form C. z and z_{fc} share parameters. Flattening z_{fc} enables a higher magnitude of gradient backpropagation from p [49].

Complementary guided attention loss L_{cga} Although, attention maps generated from a trained classifier have been used in weakly supervised semantic segmentation tasks [39, 49], to the best of our knowledge, we are the first to propose supervision on attention maps for anomaly localization in the weakly supervised setting. Since the attention map depends on the performance of C [28], we propose the complementary guided attention loss L_{cga} based on C's prediction to improve anomaly localization. We use Grad-CAM to compute the attention map for the anomalous class $A_x^{c_a}$ and the attention map for the normal class $A_x^{c_n}$ on the normal image x ($y = c_n$). Using $A_x^{c_a}$ and $A_x^{c_n}$, we propose L_{cga} where we minimize the areas covered by $A_x^{c_a}$ but simultaneously enforce $A_x^{c_n}$ to cover the entire normal image. Since the attention map is computed by backpropagating the gradients from p, any incorrect p would generate an undesired attention map. This would lead to the network learning to focus on erroneous areas of the image during training, which we avoid using L_{cga}. We compute L_{cga} only for the normal images correctly classified by the classifier i.e. if $p = y = c_n$. We define $L_{cga,1}$, the complementary guided attention loss for each image, in the weakly supervised setting as:

$$L_{cga,1} = \frac{\mathbb{1}\left(p = y = c_n\right)}{|A_x^{c_n}|} \sum_{i,j} (1 - (A_x^{c_n})_{i,j} + (A_x^{c_a})_{i,j}), \tag{5}$$

where $\mathbb{1}(\cdot)$ is an indicator function. L_{cga} is the average of $L_{cga,1}$ over the N images. Our final objective function L_{final} is defined as:

$$L_{final} = w_r L + w_{adv} L_{adv} + w_c L_{bce} + w_{cga} L_{cga}, \tag{6}$$

在无监督设置中使用任何异常训练图像时，我们将s_a归一化至[0,1]区间，并经验性地将0.5设为判定图像异常的阈值。注意力图A_{test}通过Grad-CAM从z计算得出，并通过取反(**1** - A_{test})获得用于定位异常的异常注意力图。此处**1**表示与A_{test}维度相同的全1矩阵。我们经验性地选择0.5作为异常注意力图的阈值以评估定位性能。

3.2 Weakly Supervised Approach: CAVGA$_w$

CAVGA$_u$可以进一步扩展到弱监督设置（表示为CAVGA$_w$），在此设置中我们探索使用少量异常训练图像来提升异常定位性能的可能性。在训练过程中，给定异常与正常图像的标签但无需异常的像素级标注，我们通过在图2(b)所示z的输出端引入二元分类器C来修改CAVGA$_u$，并使用二元交叉熵损失L_{bce}训练C。给定图像x及其真实标签y，我们将$p \in \{c_a, c_n\}$定义为C的预测结果，其中c_a和c_n分别对应异常类与正常类。根据图2(b)，我们将z复制为新的张量，将其展平以形成全连接层z_{fc}，并添加一个2节点输出层以构成C。z与z_{fc}共享参数。展平操作z_{fc}使得来自p的梯度反向传播能够获得更大的量级[49]。

Complementary guided attention loss L_{cga} 尽管从训练好的分类器中生成的注意力图已被用于弱监督语义分割任务[39, 49]，但据我们所知，我们是首个在弱监督设置下提出对注意力图进行监督以进行异常定位的方法。由于注意力图依赖于C的性能[28]，我们基于C的预测提出互补引导注意力损失L_{cga}以改进异常定位。我们使用Grad-CAM计算异常类$A_x^{c_a}$的注意力图以及正常图像x $(y = c_n)$上正常类$A_x^{c_n}$的注意力图。利用$A_x^{c_a}$和$A_x^{c_n}$，我们提出L_{cga}，其中我们最小化$A_x^{c_a}$覆盖的区域，同时强制$A_x^{c_n}$覆盖整个正常图像。由于注意力图是通过从p反向传播梯度计算的，任何错误的p都会产生不理想的注意力图。这将导致网络在训练过程中学习聚焦于图像的错误区域，而我们通过使用L_{cga}避免了这一问题。我们仅对分类器正确分类的正常图像计算L_{cga}，即当$p = y = c_n$时。我们在弱监督设置下将每张图像的互补引导注意力损失定义为$L_{cga,1}$:

$$L_{cga,1} = \frac{\mathbb{1}\left(p = y = c_n\right)}{|A_x^{c_n}|} \sum_{i,j} (1 - (A_x^{c_n})_{i,j} + (A_x^{c_a})_{i,j}), \tag{5}$$

其中 $\mathbb{1}(\cdot)$ 是一个指示函数。L_{cga} 是 $L_{cga,1}$ 在 N 张图像上的平均值。我们最终的损失函数 L_{final} 定义为:

$$L_{final} = w_r L + w_{adv} L_{adv} + w_c L_{bce} + w_{cga} L_{cga}, \tag{6}$$

Table 2: Our experimental settings. Notations: u: unsupervised; w: weakly supervised; D_M: MNIST [27]; D_F: Fashion-MNIST [57]; D_C: CIFAR-10 [25]

property \ dataset	MVTAD [5]		mSTC [31]		LAG [29]	D_M	D_F	D_C
setting	u	w	u	w	u	u	u	u
# total classes	15	15	13	13	1	10	10	10
# normal training images	3629	3629	244875	244875	2632	∼6k	6k	5k
# anomalous training images	0	35	0	1763	0	0	0	0
# normal testing images	467	467	21147	21147	800	∼1k	1k	1k
# anomalous testing images	1223	1223	86404	86404	2392	∼9k	9k	9k

where w_r, w_{adv}, w_c, and w_{cga} are empirically set as $1, 1, 0.001$, and 0.01 respectively. During testing, we use C to predict the input image x_{test} as anomalous or normal. The anomalous attention map A_{test} of x_{test} is computed when $y = c_a$. We use the same evaluation method as that in Sec. 3.1 for anomaly localization.

4 Experimental Setup

Benchmark datasets: We evaluate CAVGA on the MVTAD [5], mSTC [31] and LAG [29] datasets for anomaly localization, and the MVTAD, mSTC, LAG, MNIST [27], CIFAR-10 [25] and Fashion-MNIST [57] datasets for anomaly detection. Since STC dataset [31] is designed for video instead of image anomaly detection, we extract every 5^{th} frame of the video from each scene for training and testing without using any temporal information. We term the modified STC dataset as mSTC and summarize the experimental settings in Table 2.

Baseline methods: For anomaly localization, we compare CAVGA with AVID [47], AE_{L2} [6], AE_{SSIM} [6], AnoGAN [48], CNN feature dictionary (CNNFD) [37], texture inspection (TI) [8], γ-VAE grad [13] (denoted as γ-VAE_{g}), LSA [2], ADVAE [32] and variation model (VM) [52] based approaches on the MVTAD and mSTC datasets. Since [13] does not provide the code for their method, we adapt the code from [1] and report its best result using our experimental settings. We also compare CAVGA_u with CAM [60], GBP [51], SmoothGrad [50] and Patho-GAN [54] on the LAG dataset. In addition, we compare CAVGA_u with LSA [2], OCGAN [41], ULSLM [56], CapsNet PP-based and CapsNet RE-based [30] (denoted as $\text{CapsNet}_{\text{PP}}$ and $\text{CapsNet}_{\text{RE}}$), AnoGAN [48], ADGAN [12], and β-VAE [21] on the MNIST, CIFAR-10 and Fashion-MNIST datasets for anomaly detection.

Architecture details: Based on the framework in Fig. 2 (a), we use the convolution layers of ResNet-18 [19] as our encoder pretrained from ImageNet [45] and finetuned on each category / scene individually. Inspired from [9], we propose to use the residual generator as our residual decoder by modifying it with a convolution layer interleaved between two upsampling layers. The skip connection added from the output of the upsampling layer to the output of the convolution layer, increases mutual information between observations and latent variable and also avoids latent variable collapse [14]. We use the discriminator of DC-GAN

表2：我们的实验设置。符号说明：u：无监督；w：弱监督；D_M：MNIST [27]；D_F：Fashion-MNIST [57]；D_C：CIFAR-10 [25]

property \ dataset	MVTAD [5]		mSTC [31]		LAG [29]	D_M	D_F	D_C
setting	u	w	u	w	u	u	u	u
# total classes	15	15	13	13	1	10	10	10
# normal training images	3629	3629	244875	244875	2632	∼6k	6k	5k
# anomalous training images	0	35	0	1763	0	0	0	0
# normal testing images	467	467	21147	21147	800	∼1k	1k	1k
# anomalous testing images	1223	1223	86404	86404	2392	∼9k	9k	9k

其中，w_r、w_{adv}, w_c和w_{cga}根据经验分别设置为1,1,0.001和0.01。在测试过程中，我们使用C来预测输入图像x_{test}为异常或正常。当$y = c_a$时，计算得到x_{test}的异常注意力图A_{test}。我们采用与第3.1节相同的评估方法进行异常定位。

4 Experimental Setup

Benchmark datasets: 我们在MVTAD [5]、mSTC [31]和LAG [29]数据集上评估CAVGA的异常定位性能，并在MVTAD、mSTC、LAG、MNIST [27]、CIFAR-10 [25]和Fashion-MNIST [57]数据集上评估其异常检测性能。由于STC数据集[31]专为视频异常检测而非图像异常检测设计，我们从每个场景的视频中每隔5$^{\text{th}}$帧提取图像用于训练和测试，且未使用任何时序信息。我们将修改后的STC数据集称为mSTC，并在表2中总结了实验设置。

Baseline methods: 对于异常定位，我们在MVTAD和mSTC数据集上将CAVGA与AVID [47]、AE_{L2} [6]、AE_{SSIM} [6]、AnoGAN [48]、CNN特征字典（CN-NFD）[37]、纹理检测（TI）[8]、γ-VAE梯度[13]（记为γ-VAE_{g}）、LSA [2]、ADVAE [32]以及基于变分模型（VM）[52]的方法进行比较。由于[13]未提供其方法的代码，我们采用[1]的代码并根据我们的实验设置报告其最佳结果。我们还在LAG数据集上将CAVGA_u与CAM [60]、GBP [51]、Smooth-Grad [50]和Patho-GAN [54]进行比较。此外，我们在MNIST、CIFAR-10和Fashion-MNIST数据集上，将CAVGA_u与LSA [2]、OCGAN [41]、ULSLM [56]、基于CapsNet PP和基于CapsNet RE的方法[30]（记为$\text{CapsNet}_{\text{PP}}$和$\text{CapsNet}_{\text{RE}}$）、AnoGAN [48]、ADGAN [12]以及β-VAE [21]进行比较，以进行异常检测。

Architecture details: 基于图2（a）的框架，我们采用ResNet-18[19]的卷积层作为编码器，该编码器在ImageNet[45]上预训练并在每个类别/场景上单独微调。受[9]启发，我们提出将残差生成器作为残差解码器，通过在两个上采样层之间插入卷积层进行改进。从上采样层输出到卷积层输出添加的跳跃连接，增加了观测值与潜在变量之间的互信息，同时避免了潜在变量坍缩[14]。我们采用DC-GAN的判别器

Table 3: Performance comparison of anomaly localization in category-specific IoU, mean IoU ($\overline{\text{IoU}}$), and mean AuROC ($\overline{\text{AuROC}}$) on the MVTAD dataset. The darker cell color indicates better performance ranking in each row

Category	AVID [47]	AE_{SSIM} [6]	AE_{L2} [6]	AnoGAN [48]	γ-VAE_{g} [13]	LSA [2]	ADVAE [32]	CAVGA -D_u	CAVGA -R_u	CAVGA -D_w	CAVGA -R_w
Bottle	0.28	0.15	0.22	0.05	0.27	0.27	0.27	0.30	0.34	0.36	0.39
Hazelnut	0.54	0.00	0.41	0.02	0.63	0.41	0.44	0.44	0.51	0.58	0.79
Capsule	0.21	0.09	0.11	0.04	0.24	0.22	0.11	0.25	0.31	0.38	0.41
Metal Nut	0.05	0.01	0.26	0.00	0.22	0.38	0.49	0.39	0.45	0.46	0.46
Leather	0.32	0.34	0.67	0.34	0.41	0.77	0.24	0.76	0.79	0.80	0.84
Pill	0.11	0.07	0.25	0.17	0.48	0.18	0.18	0.34	0.40	0.44	0.53
Wood	0.14	0.36	0.29	0.14	0.45	0.41	0.14	0.56	0.59	0.61	0.66
Carpet	0.25	0.69	0.38	0.34	0.79	0.76	0.10	0.71	0.73	0.70	0.81
Tile	0.09	0.04	0.23	0.08	0.38	0.32	0.23	0.31	0.38	0.47	0.81
Grid	0.51	0.88	0.83	0.04	0.36	0.20	0.02	0.32	0.38	0.42	0.55
Cable	0.27	0.01	0.05	0.01	0.26	0.36	0.18	0.37	0.44	0.49	0.51
Transistor	0.18	0.01	0.22	0.08	0.44	0.21	0.30	0.30	0.35	0.38	0.45
Toothbrush	0.43	0.08	0.51	0.07	0.37	0.48	0.14	0.54	0.57	0.60	0.63
Screw	0.22	0.03	0.34	0.01	0.38	0.38	0.17	0.42	0.48	0.51	0.66
Zipper	0.25	0.10	0.13	0.01	0.17	0.14	0.06	0.20	0.26	0.29	0.31
$\overline{\text{IoU}}$	0.26	0.19	0.33	0.09	0.39	0.37	0.20	0.41	0.47	0.50	0.59
$\overline{\text{AuROC}}$	0.78	0.87	0.82	0.74	0.86	0.79	0.86	0.85	0.89	0.92	0.93

[42] pretrained on the Celeb-A dataset [33] and finetuned on our data as our discriminator and term this network as CAVGA-R. For fair comparisons with the baseline approaches in terms of network architecture, we use the discriminator and generator of DC-GAN pretrained on the Celeb-A dataset as our encoder and decoder respectively. We keep the same discriminator as discussed previously and term this network as CAVGA-D. CAVGA-D_u and CAVGA-R_u are termed as CAVGA_u in the unsupervised setting, and CAVGA-D_w and CAVGA-R_w as CAVGA_w in weakly supervised setting respectively.

Training and evaluation: For anomaly localization and detection on the MVTAD, mSTC and LAG datasets, the network is trained only on normal images in the unsupervised setting. In the weakly supervised setting, since none of the baseline methods provide the number of anomalous training images they use to compute the threshold, we randomly choose 2% of the anomalous images along with all the normal training images for training. On the MNIST, CIFAR-10 and Fashion-MNIST datasets, we follow the same procedure as defined in [12] (training/testing uses single class as normal and the rest of the classes as anomalous. We train CAVGA-D_u using this normal class). For anomaly localization, we show the AuROC [5] and the Intersection-over-Union (IoU) between the generated attention map and the ground truth. Following [5], we use the mean of accuracy of correctly classified anomalous images and normal images to evaluate the performance of anomaly detection on both the normal and anomalous images on the MVTAD, mSTC and LAG datasets. On the MNIST, CIFAR-10, and Fashion-MNIST datasets, same as [12], we use AuROC for evaluation.

表3：在MVTAD数据集上，按类别特定IoU、平均IoU（IoU）和平均AuROC（AuROC）进行异常定位的性能比较。每行中单元格颜色越深表示该指标性能排名越优

Category	AVID [47]	AE_{SSIM} [6]	AE_{L2} [6]	AnoGAN [48]	γ-VAE_g [13]	LSA [2]	ADVAE [32]	CAVGA -D_u	CAVGA -R_u	CAVGA -D_w	CAVGA -R_w
Bottle	0.28	0.15	0.22	0.05	0.27	0.27	0.27	0.30	0.34	0.36	0.39
Hazelnut	0.54	0.00	0.41	0.02	0.63	0.41	0.44	0.44	0.51	0.58	0.79
Capsule	0.21	0.09	0.11	0.04	0.24	0.22	0.11	0.25	0.31	0.38	0.41
Metal Nut	0.05	0.01	0.26	0.00	0.22	0.38	0.49	0.39	0.45	0.46	0.46
Leather	0.32	0.34	0.67	0.34	0.41	0.77	0.24	0.76	0.79	0.80	0.84
Pill	0.11	0.07	0.25	0.17	0.48	0.18	0.18	0.34	0.40	0.44	0.53
Wood	0.14	0.36	0.29	0.14	0.45	0.41	0.14	0.56	0.59	0.61	0.66
Carpet	0.25	0.69	0.38	0.34	0.79	0.76	0.10	0.71	0.73	0.70	0.81
Tile	0.09	0.04	0.23	0.08	0.38	0.32	0.23	0.31	0.38	0.47	0.81
Grid	0.51	0.88	0.83	0.04	0.36	0.20	0.02	0.32	0.38	0.42	0.55
Cable	0.27	0.01	0.05	0.01	0.26	0.36	0.18	0.37	0.44	0.49	0.51
Transistor	0.18	0.01	0.22	0.08	0.44	0.21	0.30	0.30	0.35	0.38	0.45
Toothbrush	0.43	0.08	0.51	0.07	0.37	0.48	0.14	0.54	0.57	0.60	0.63
Screw	0.22	0.03	0.34	0.01	0.38	0.38	0.17	0.42	0.48	0.51	0.66
Zipper	0.25	0.10	0.13	0.01	0.17	0.14	0.06	0.20	0.26	0.29	0.31
$\overline{\text{IoU}}$	0.26	0.19	0.33	0.09	0.39	0.37	0.20	0.41	0.47	0.50	0.59
$\overline{\text{AuROC}}$	0.78	0.87	0.82	0.74	0.86	0.79	0.86	0.85	0.89	0.92	0.93

[42] 在 Celeb-A 数据集 [33] 上预训练并在我们的数据上微调作为判别器，将此网络称为 CAVGA-R。为了在网络架构方面与基线方法进行公平比较，我们使用在 Celeb-A 数据集上预训练的 DC-GAN 的判别器和生成器分别作为我们的编码器和解码器。我们保持与先前讨论相同的判别器，并将此网络称为 CAVGA-D 。在无监督设置中，CAVGA-D_u 和 CAVGA-R_u 被统称为 CAVGA$_u$；而在弱监督设置中，CAVGA-D_w 和 CAVGA-R_w 则分别被称为 CAVGA$_w$。

Training and evaluation: 在MVTAD、mSTC和LAG数据集上进行异常定位与检测时，网络仅在无监督设置下使用正常图像进行训练。在弱监督设置中，由于基线方法均未提供其用于计算阈值的异常训练图像数量，我们随机选取2%的异常图像与全部正常训练图像共同参与训练。对于MNIST、CIFAR-10和Fashion-MNIST数据集，我们遵循文献[12]定义的方法（训练/测试时使用单个类别作为正常类，其余类别作为异常类。我们使用该正常类训练CAVGA-D{v*}）。在异常定位方面，我们展示生成注意力图与真实标注之间的AuROC[5]和交并比（IoU）。依据文献[5]，我们采用异常图像与正常图像正确分类准确率的均值，来评估MVTAD、mSTC和LAG数据集上正常与异常图像的异常检测性能。对于MNIST、CIFAR-10和Fashion-MNIST数据集，与文献[12]相同，我们使用AuROC进行评估。

5 Experimental Results

We use the cell color in the quantitative result tables to denote the performance ranking in that row, where darker cell color means better performance.

Performance on anomaly localization: Fig. 3 (a) shows the qualitative results and Table 3 shows that CAVGA$_u$ localizes the anomaly better compared to the baselines on the MVTAD dataset. CAVGA-D$_u$ outperforms the best performing baseline method (γ-VAE$_g$) in mean IoU by 5%. Most baselines use anomalous training images to compute class-specific threshold to localize anomalies. *Needing no anomalous training images*, CAVGA-D$_u$ still outperforms all the mentioned baselines in mean IoU. In terms of mean AuROC, CAVGA-D$_u$ outperforms CNNFD, TI and VM by 9%, 12% and 10% respectively and achieves comparable results with best baseline method. Table 3 also shows that CAVGA-D$_w$ outperforms CAVGA-D$_u$ by 22% and 8% on mean IoU and mean AuROC respectively. CAVGA-D$_w$ also outperforms the baselines in mean AuROC. Fig. 4 illustrates that one challenge in anomaly localization is the low contrast between the anomalous regions and their background. In such scenarios, although still outperforming the baselines, CAVGA does not localize the anomaly well.

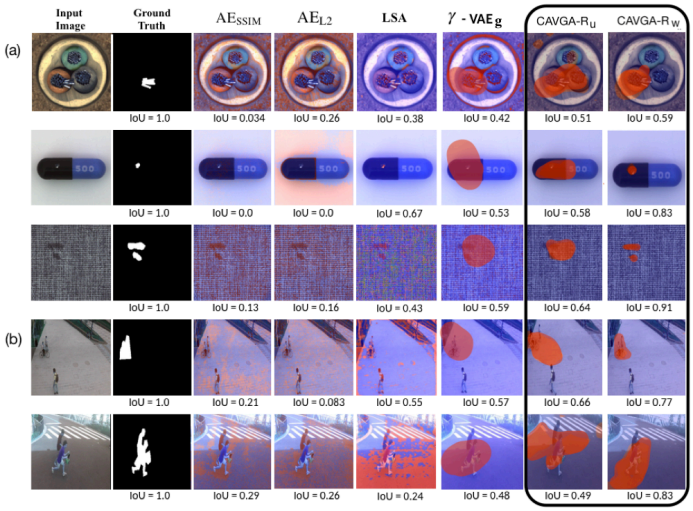


Fig. 3: Qualitative results on (a) MVTAD & (b) mSTC datasets respectively. The anomalous attention map (in red) depicts the localization of the anomaly

Fig. 3 (b) illustrates the qualitative results and Table 4 shows that CAVGA also outperforms the baseline methods in mean IoU and mean AuROC on the mSTC dataset. Table 5 shows that CAVGA outperforms the most competitive baseline Patho-GAN [54] by 16% in IoU on the LAG dataset. CAVGA is practically reasonable to train on a single GTX 1080Ti GPU, having comparable training and testing time with baseline methods.

5 Experimental Results

我们在定量结果表中使用单元格颜色来表示该行的性能排名，其中单元格颜色越深表示性能越好。

Performance on anomaly localization: 图3（a）展示了定性结果，表3显示在MVTAD数据集上，CAVGA$_u$相比基线方法能更好地定位异常区域。CAVGA-D$_u$在平均IoU指标上优于表现最佳的基线方法（γ-VAE$_g$），提升幅度达5%。大多数基线方法使用异常训练图像计算类别特定阈值以定位异常。尽管如此，CAVGA-D$_u$在平均IoU上仍优于所有提及的基线方法。在平均AuROC方面，CAVGA-D$_u$分别以9%、12%和10%的优势超过CNNFD、TI和VM，并与最佳基线方法取得相当的结果。表3还显示，CAVGA-D$_w$在平均IoU和平均AuROC上分别比CAVGA-D$_u$高出22%和8%。CAVGA-D$_w$在平均AuROC上也优于基线方法。图4说明异常定位的一个挑战在于异常区域与背景之间的低对比度。在此类场景中，尽管CAVGA仍优于基线方法，但其异常定位效果并不理想。

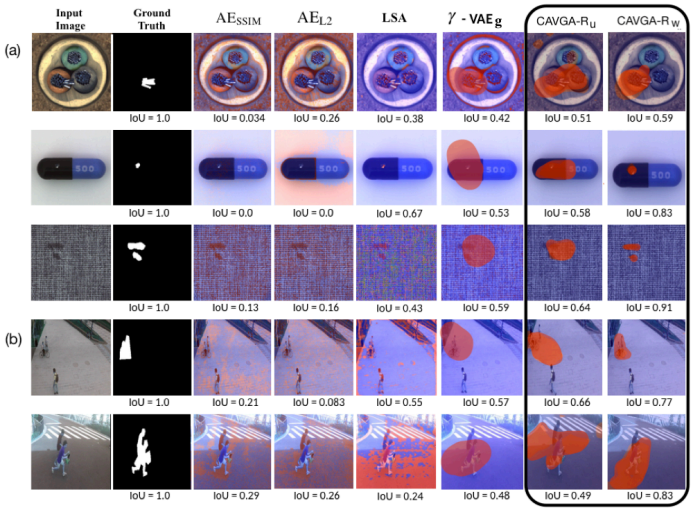


图3：分别在(a) MVTAD和(b) mSTC数据集上的定性结果。异常注意力图（红色部分）展示了异常定位情况

图3(b)展示了定性结果，表4表明CAVGA在mSTC数据集上的平均IoU和平均AuROC也优于基线方法。表5显示，在LAG数据集上，CAVGA在IoU指标上比最具竞争力的基线方法Patho-GAN [54]高出16%。CAVGA在单块GTX 1080Ti GPU上训练具有实际可行性，其训练和测试时间与基线方法相当。

Table 4: Performance comparison of anomaly localization in IoU and its mean ($\overline{\text{IoU}}$) along with anomaly detection in terms of mean of accuracy of correctly classified anomalous images and normal images on the mSTC dataset for each scene ID s_i. For anomaly localization, we also list the mean AuROC ($\overline{\text{AuROC}}$)

Task \ Method	s_i	γ-VAE$_g$ [13]	AVID [47]	LSA [2]	AE$_{\text{SSIM}}$ [6]	AE$_{\text{L2}}$ [6]	CAVGA -D$_u$	CAVGA -R$_u$	CAVGA -D$_w$	CAVGA -R$_w$
Localization	01	0.239	0.182	0.244	0.201	0.163	0.267	0.316	0.383	0.441
	02	0.206	0.206	0.183	0.081	0.172	0.190	0.234	0.257	0.349
	03	0.272	0.162	0.265	0.218	0.240	0.277	0.293	0.313	0.465
	04	0.290	0.263	0.271	0.118	0.125	0.283	0.349	0.360	0.381
	05	0.318	0.234	0.287	0.162	0.129	0.291	0.312	0.408	0.478
	06	0.337	0.314	0.238	0.215	0.198	0.344	0.420	0.455	0.589
	07	0.168	0.214	0.137	0.191	0.165	0.198	0.241	0.284	0.366
	08	0.220	0.168	0.233	0.069	0.056	0.219	0.254	0.295	0.371
	09	0.174	0.193	0.187	0.038	0.021	0.247	0.284	0.313	0.365
	10	0.146	0.137	0.146	0.116	0.141	0.149	0.166	0.245	0.295
	11	0.277	0.264	0.286	0.101	0.075	0.309	0.372	0.441	0.588
	12	0.162	0.180	0.108	0.203	0.164	0.098	0.141	0.207	0.263
	$\overline{\text{IoU}}$	0.234	0.210	0.215	0.143	0.137	0.239	0.281	0.330	0.412
	$\overline{\text{AuROC}}$	0.82	0.77	0.81	0.76	0.74	0.83	0.85	0.89	0.90
Detection	01	0.75	0.68	0.75	0.65	0.72	0.77	0.85	0.84	0.87
	02	0.75	0.75	0.79	0.70	0.61	0.76	0.84	0.89	0.90
	03	0.81	0.68	0.63	0.79	0.71	0.82	0.84	0.86	0.88
	04	0.83	0.71	0.79	0.81	0.66	0.80	0.80	0.81	0.83
	05	0.86	0.59	0.68	0.71	0.67	0.81	0.86	0.90	0.94
	06	0.59	0.62	0.58	0.47	0.55	0.64	0.67	0.65	0.70
	07	0.59	0.63	0.63	0.36	0.59	0.60	0.64	0.75	0.77
	08	0.77	0.73	0.75	0.69	0.70	0.74	0.74	0.76	0.80
	09	0.89	0.88	0.79	0.84	0.73	0.87	0.88	0.90	0.91
	10	0.64	0.80	0.84	0.83	0.88	0.88	0.92	0.94	0.94
	11	0.78	0.68	0.71	0.71	0.75	0.79	0.81	0.83	0.83
	12	0.71	0.66	0.63	0.65	0.52	0.76	0.79	0.81	0.83
	avg	0.75	0.70	0.71	0.68	0.67	0.77	0.80	0.83	0.85

Table 5: Performance comparison of anomaly localization in IoU along with anomaly detection in terms of classification accuracy on the LAG dataset [29]

Task \ Method	CAM [60]	GBP [51]	SmoothGrad [50]	Patho-GAN [54]	CAVGA-D$_u$
Localization	0.13	0.09	0.14	0.37	0.43
Detection	0.68	0.84	0.79	0.89	0.90

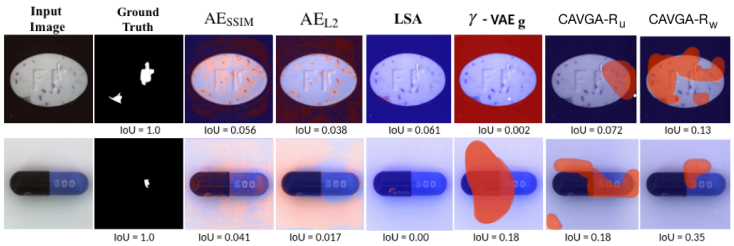


Fig. 4: Examples of incorrect localization of the anomaly on the MVTAD dataset by CAVGA-R$_u$ and CAVGA-R$_w$

表4：在mSTC数据集上各场景ID s_i 的异常定位性能对比（IoU及其均值IoU），以及异常检测性能（异常图像与正常图像正确分类的平均准确率）。对于异常定位，我们还列出了平均AuROC（AuROC）。

Task \ Method	s_i	γ-VAE$_g$ [13]	AVID [47]	LSA [2]	AE$_{SSIM}$ [6]	AE$_{L2}$ [6]	CAVGA -D$_u$	CAVGA -R$_u$	CAVGA -D$_w$	CAVGA -R$_w$
Localization	01	0.239	0.182	0.244	0.201	0.163	0.267	0.316	0.383	0.441
	02	0.206	0.206	0.183	0.081	0.172	0.190	0.234	0.257	0.349
	03	0.272	0.162	0.265	0.218	0.240	0.277	0.293	0.313	0.465
	04	0.290	0.263	0.271	0.118	0.125	0.283	0.349	0.360	0.381
	05	0.318	0.234	0.287	0.162	0.129	0.291	0.312	0.408	0.478
	06	0.337	0.314	0.238	0.215	0.198	0.344	0.420	0.455	0.589
	07	0.168	0.214	0.137	0.191	0.165	0.198	0.241	0.284	0.366
	08	0.220	0.168	0.233	0.069	0.056	0.219	0.254	0.295	0.371
	09	0.174	0.193	0.187	0.038	0.021	0.247	0.284	0.313	0.365
	10	0.146	0.137	0.146	0.116	0.141	0.149	0.166	0.245	0.295
	11	0.277	0.264	0.286	0.101	0.075	0.309	0.372	0.441	0.588
	12	0.162	0.180	0.108	0.203	0.164	0.098	0.141	0.207	0.263
	$\overline{\text{IoU}}$	0.234	0.210	0.215	0.143	0.137	0.239	0.281	0.330	0.412
	$\overline{\text{AuROC}}$	0.82	0.77	0.81	0.76	0.74	0.83	0.85	0.89	0.90
Detection	01	0.75	0.68	0.75	0.65	0.72	0.77	0.85	0.84	0.87
	02	0.75	0.75	0.79	0.70	0.61	0.76	0.84	0.89	0.90
	03	0.81	0.68	0.63	0.79	0.71	0.82	0.84	0.86	0.88
	04	0.83	0.71	0.79	0.81	0.66	0.80	0.80	0.81	0.83
	05	0.86	0.59	0.68	0.71	0.67	0.81	0.86	0.90	0.94
	06	0.59	0.62	0.58	0.47	0.55	0.64	0.67	0.65	0.70
	07	0.59	0.63	0.63	0.36	0.59	0.60	0.64	0.75	0.77
	08	0.77	0.73	0.75	0.69	0.70	0.74	0.74	0.76	0.80
	09	0.89	0.88	0.79	0.84	0.73	0.87	0.88	0.90	0.91
	10	0.64	0.80	0.84	0.83	0.88	0.88	0.92	0.94	0.94
	11	0.78	0.68	0.71	0.71	0.75	0.79	0.81	0.83	0.83
	12	0.71	0.66	0.63	0.65	0.52	0.76	0.79	0.81	0.83
	avg	0.75	0.70	0.71	0.68	0.67	0.77	0.80	0.83	0.85

表5：在LAG数据集[29]上，异常定位的IoU性能比较以及异常检测的分类准确率对比

Task \ Method	CAM [60]	GBP [51]	SmoothGrad [50]	Patho-GAN [54]	CAVGA-D$_u$
Localization	0.13	0.09	0.14	0.37	0.43
Detection	0.68	0.84	0.79	0.89	0.90

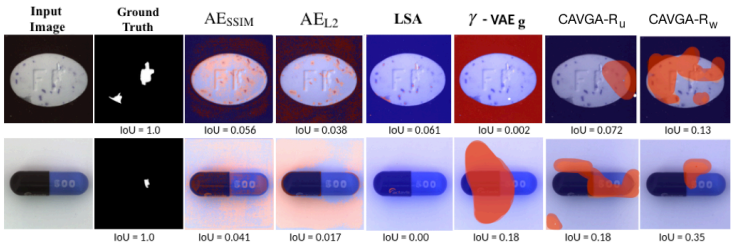


图4：CAVGA-R$_u$和CAVGA-R$_w$在MVTAD数据集上异常定位错误的示例

Table 6: The mean of accuracy of correctly classified anomalous images and normal images in anomaly detection on the MVTAD dataset

Category	AVID [47]	AE_{SSIM} [6]	AE_{L2} [6]	AnoGAN [48]	γ-VAE_g [13]	LSA [2]	CAVGA -D_u	CAVGA -R_u	CAVGA -D_w	CAVGA -R_w
Bottle	0.88	0.88	0.80	0.69	0.86	0.86	0.89	0.91	0.93	0.96
Hazelnut	0.86	0.54	0.88	0.50	0.74	0.80	0.84	0.87	0.90	0.92
Capsule	0.85	0.61	0.62	0.58	0.86	0.71	0.83	0.87	0.89	0.93
Metal Nut	0.63	0.54	0.73	0.50	0.78	0.67	0.67	0.71	0.81	0.88
Leather	0.58	0.46	0.44	0.52	0.71	0.70	0.71	0.75	0.80	0.84
Pill	0.86	0.60	0.62	0.62	0.80	0.85	0.88	0.91	0.93	0.97
Wood	0.83	0.83	0.74	0.68	0.89	0.75	0.85	0.88	0.89	0.89
Carpet	0.70	0.67	0.50	0.49	0.67	0.74	0.73	0.78	0.80	0.82
Tile	0.66	0.52	0.77	0.51	0.81	0.70	0.70	0.72	0.81	0.86
Grid	0.59	0.69	0.78	0.51	0.83	0.54	0.75	0.78	0.79	0.81
Cable	0.64	0.61	0.56	0.53	0.56	0.61	0.63	0.67	0.86	0.97
Transistor	0.58	0.52	0.71	0.67	0.70	0.50	0.73	0.75	0.80	0.89
Toothbrush	0.73	0.74	0.98	0.57	0.89	0.89	0.91	0.97	0.96	0.99
Screw	0.66	0.51	0.69	0.35	0.71	0.75	0.77	0.78	0.79	0.79
Zipper	0.84	0.80	0.80	0.59	0.67	0.88	0.87	0.94	0.95	0.96
mean	0.73	0.63	0.71	0.55	0.77	0.73	0.78	0.82	0.86	0.90

Performance on anomaly detection: Table 6 shows that CAVGA$_u$ outperforms the baselines in the mean of accuracy of correctly classified anomalous images and normal images on the MVTAD dataset. CAVGA-D$_u$ outperforms the best performing baseline (γ-VAE$_g$) in mean of classification accuracy by 1.3%. Table 4 and Table 5 show that CAVGA outperforms the baseline methods in classification accuracy on both the mSTC and LAG datasets by 2.6% and 1.1% respectively. Furthermore, Table 7 shows that CAVGA-D$_u$ outperforms all the baselines in mean AuROC in the unsupervised setting on the MNIST, CIFAR-10 and Fashion-MNIST datasets. CAVGA-D$_u$ also outperforms MemAE [16] and β-VAE [21] by 1.1% and 8% on MNIST and by 21% and 38% on CIFAR-10 datasets respectively. CAVGA-D$_u$ also outperforms all the listed baselines in mean AuROC on the Fashion-MNIST dataset.

6 Ablation Study

All the ablation studies are performed on 15 categories on the MVTAD dataset, of which 5 are reported here. The mean of all 15 categories is shown in Table 8. We illustrate the effectiveness of the convolutional z in CAVGA, L_{ae} in the unsupervised setting, and L_{cga} in the weakly supervised setting. The qualitative results are shown in Fig. 5. The column IDs to refer to the columns in Table 8.

Effect of convolutional latent variable z: To show the effectiveness of the convolutional z, we flatten the output of the encoder of CAVGA-R$_u$ and CAVGA-R$_w$, and connect it to a fully connected layer as latent variable. Following [6], the dimension of latent variable is chosen as 100. We call these network as CAVGA-R$^*_u$ and CAVGA-R$^*_w$ in the unsupervised and weakly supervised settings respectively. In the unsupervised setting, we train CAVGA-R$_u$ and CAVGA-R$^*_u$ using $L + L_{adv}$ as our objective function and compute the anomalous attention

表6：MVTAD数据集中异常检测中正确分类的异常图像与正常图像准确率的平均值

Category	AVID [47]	AE_{SSIM} [6]	AE_{L2} [6]	AnoGAN [48]	γ-VAE_g [13]	LSA [2]	CAVGA -D_u	CAVGA -R_u	CAVGA -D_w	CAVGA -R_w
Bottle	0.88	0.88	0.80	0.69	0.86	0.86	0.89	0.91	0.93	0.96
Hazelnut	0.86	0.54	0.88	0.50	0.74	0.80	0.84	0.87	0.90	0.92
Capsule	0.85	0.61	0.62	0.58	0.86	0.71	0.83	0.87	0.89	0.93
Metal Nut	0.63	0.54	0.73	0.50	0.78	0.67	0.67	0.71	0.81	0.88
Leather	0.58	0.46	0.44	0.52	0.71	0.70	0.71	0.75	0.80	0.84
Pill	0.86	0.60	0.62	0.62	0.80	0.85	0.88	0.91	0.93	0.97
Wood	0.83	0.83	0.74	0.68	0.89	0.75	0.85	0.88	0.89	0.89
Carpet	0.70	0.67	0.50	0.49	0.67	0.74	0.73	0.78	0.80	0.82
Tile	0.66	0.52	0.77	0.51	0.81	0.70	0.70	0.72	0.81	0.86
Grid	0.59	0.69	0.78	0.51	0.83	0.54	0.75	0.78	0.79	0.81
Cable	0.64	0.61	0.56	0.53	0.56	0.61	0.63	0.67	0.86	0.97
Transistor	0.58	0.52	0.71	0.67	0.70	0.50	0.73	0.75	0.80	0.89
Toothbrush	0.73	0.74	0.98	0.57	0.89	0.89	0.91	0.97	0.96	0.99
Screw	0.66	0.51	0.69	0.35	0.71	0.75	0.77	0.78	0.79	0.79
Zipper	0.84	0.80	0.80	0.59	0.67	0.88	0.87	0.94	0.95	0.96
mean	0.73	0.63	0.71	0.55	0.77	0.73	0.78	0.82	0.86	0.90

Performance on anomaly detection: 表6显示，在MVTAD数据集上，CAVGA$_u$在异常图像与正常图像正确分类的平均准确率方面超越了基线方法。CAVGA-D$_u$在平均分类准确率上比表现最佳的基线方法（γ-VAE$_g$）高出1.3%。表4和表5表明，在mSTC和LAG数据集上，CAVGA的分类准确率分别比基线方法高出2.6%和1.1%。此外，表7显示，在MNIST、CIFAR-10和Fashion-MNIST数据集的无监督设置中，CAVGA-D$_u$在平均AuROC指标上均优于所有基线方法。CAVGA-D$_u$在MNIST数据集上分别比MemAE[16]和β-VAE[21]高出1.1%和8%，在CIFAR-10数据集上分别高出21%和38%。在Fashion-MNIST数据集上，CAVGA-D$_u$的平均AuROC也优于所有列出的基线方法。

6 Ablation Study

所有消融研究均在MVTAD数据集的15个类别上进行，其中5类结果在此报告。全部15个类别的平均值展示于表8。我们验证了CAVGA中卷积z的有效性、无监督设置下的L_{ae}以及弱监督设置下的L_{cga}。定性结果如图5所示。列ID对应表8中的各数据列。

Effect of convolutional latent variable z: 为了展示卷积z的有效性，我们将CAVGA-R$_u$和CAVGA-R$_w$编码器的输出展平，并将其连接到一个全连接层作为潜变量。参照[6]，潜变量的维度设定为100。在无监督和弱监督设置下，我们分别将这些网络称为CAVGA-R$^*_u$和CAVGA-R$^*_w$。在无监督设置中，我们使用$L + L_{adv}$作为目标函数训练CAVGA-R$_u$和CAVGA-R$^*_u$，并计算异常注意力。

Table 7: Performance comparison of anomaly detection in terms of AuROC and mean AuROC with the SOTA methods on MNIST (D_M) and CIFAR-10 (D_C) datasets . We also report the mean AuROC on Fashion-MNIST (D_F) dataset

Dataset	Class	γ-VAE$_g$ [13]	LSA [2]	OCGAN [41]	ULSLM [56]	CapsNet$_{PP}$ [30]	CapsNet$_{RE}$ [30]	AnoGAN [48]	ADGAN [12]	CAVGA -D$_u$
D_M [27]	0	0.991	0.993	0.998	0.991	0.998	0.947	0.990	0.999	0.994
	1	0.996	0.999	0.999	0.972	0.990	0.907	0.998	0.992	0.997
	2	0.983	0.959	0.942	0.919	0.984	0.970	0.888	0.968	0.989
	3	0.978	0.966	0.963	0.943	0.976	0.949	0.913	0.953	0.983
	4	0.976	0.956	0.975	0.942	0.935	0.872	0.944	0.960	0.977
	5	0.972	0.964	0.980	0.872	0.970	0.966	0.912	0.955	0.968
	6	0.993	0.994	0.991	0.988	0.942	0.909	0.925	0.980	0.988
	7	0.981	0.980	0.981	0.939	0.987	0.934	0.964	0.950	0.986
	8	0.980	0.953	0.939	0.960	0.993	0.929	0.883	0.959	0.988
	9	0.967	0.981	0.981	0.967	0.990	0.871	0.958	0.965	0.991
	mean	0.982	0.975	0.975	0.949	0.977	0.925	0.937	0.968	0.986
D_C [25]	0	0.702	0.735	0.757	0.740	0.622	0.371	0.610	0.661	0.653
	1	0.663	0.580	0.531	0.747	0.455	0.737	0.565	0.435	0.784
	2	0.680	0.690	0.640	0.628	0.671	0.421	0.648	0.636	0.761
	3	0.713	0.542	0.620	0.572	0.675	0.588	0.528	0.488	0.747
	4	0.770	0.761	0.723	0.678	0.683	0.388	0.670	0.794	0.775
	5	0.689	0.546	0.620	0.602	0.635	0.601	0.592	0.640	0.552
	6	0.805	0.751	0.723	0.753	0.727	0.491	0.625	0.685	0.813
	7	0.588	0.535	0.575	0.685	0.673	0.631	0.576	0.559	0.745
	8	0.813	0.717	0.820	0.781	0.710	0.410	0.723	0.798	0.801
	9	0.744	0.548	0.554	0.795	0.466	0.671	0.582	0.643	0.741
	mean	0.717	0.641	0.656	0.736	0.612	0.531	0.612	0.634	0.737
D_F [57]	mean	0.873	0.876	-	-	0.765	0.679	-	-	0.885

Table 8: The ablation study on 5 randomly chosen categories showing anomaly localization in IoU on the MVTAD dataset. The mean of all 15 categories is reported. CAVGA-R$^*_u$ and CAVGA-R$^*_w$ are our base architecture with a flattened z in the unsupervised and weakly supervised settings respectively. "conv z" means using convolutional z

Method / Category	CAVGA -R$^*_u$	CAVGA -R$^*_u$ + L_{ae}	CAVGA -R$_u$ + conv z	CAVGA -R$_u$ + conv z + L_{ae}	CAVGA -R$^*_w$	CAVGA -R$^*_w$ + L_{cga}	CAVGA -R$_w$ + conv z	CAVGA -R$_w$ + conv z + L_{cga}
Column ID	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8
Bottle	0.24	0.27	0.26	0.33	0.16	0.34	0.28	0.39
Hazelnut	0.16	0.26	0.31	0.47	0.51	0.76	0.67	0.79
Capsule	0.09	0.22	0.14	0.31	0.18	0.36	0.27	0.41
Metal Nut	0.28	0.38	0.34	0.45	0.25	0.38	0.28	0.46
Leather	0.55	0.71	0.64	0.79	0.72	0.79	0.75	0.84
mean	0.24	0.34	0.33	0.47	0.39	0.52	0.48	0.60

map from the feature map of the latent variable during inference. Similarly, in the weakly supervised setting, we train CAVGA-R$_w$ and CAVGA-R$^*_w$ using $L + L_{adv} + L_{bce}$ as our objective function and compute the anomalous attention map from the classifier's prediction during inference. Comparing column c_1 with

表7：在MNIST（D_M）和CIFAR-10（D_C）数据集上，异常检测在AuROC和平均AuROC方面与SOTA方法的性能对比。我们同时报告了在Fashion-MNIST（D_F）数据集上的平均AuROC。

Dataset	Class	γ-VAE$_g$ [13]	LSA [2]	OCGAN [41]	ULSLM [56]	CapsNet$_{PP}$ [30]	CapsNet$_{RE}$ [30]	AnoGAN [48]	ADGAN [12]	CAVGA -D$_u$
D_M [27]	0	0.991	0.993	0.998	0.991	0.998	0.947	0.990	0.999	0.994
	1	0.996	0.999	0.999	0.972	0.990	0.907	0.998	0.992	0.997
	2	0.983	0.959	0.942	0.919	0.984	0.970	0.888	0.968	0.989
	3	0.978	0.966	0.963	0.943	0.976	0.949	0.913	0.953	0.983
	4	0.976	0.956	0.975	0.942	0.935	0.872	0.944	0.960	0.977
	5	0.972	0.964	0.980	0.872	0.970	0.966	0.912	0.955	0.968
	6	0.993	0.994	0.991	0.988	0.942	0.909	0.925	0.980	0.988
	7	0.981	0.980	0.981	0.939	0.987	0.934	0.964	0.950	0.986
	8	0.980	0.953	0.939	0.960	0.993	0.929	0.883	0.959	0.988
	9	0.967	0.981	0.981	0.967	0.990	0.871	0.958	0.965	0.991
	mean	0.982	0.975	0.975	0.949	0.977	0.925	0.937	0.968	0.986
D_C [25]	0	0.702	0.735	0.757	0.740	0.622	0.371	0.610	0.661	0.653
	1	0.663	0.580	0.531	0.747	0.455	0.737	0.565	0.435	0.784
	2	0.680	0.690	0.640	0.628	0.671	0.421	0.648	0.636	0.761
	3	0.713	0.542	0.620	0.572	0.675	0.588	0.528	0.488	0.747
	4	0.770	0.761	0.723	0.678	0.683	0.388	0.670	0.794	0.775
	5	0.689	0.546	0.620	0.602	0.635	0.601	0.592	0.640	0.552
	6	0.805	0.751	0.723	0.753	0.727	0.491	0.625	0.685	0.813
	7	0.588	0.535	0.575	0.685	0.673	0.631	0.576	0.559	0.745
	8	0.813	0.717	0.820	0.781	0.710	0.410	0.723	0.798	0.801
	9	0.744	0.548	0.554	0.795	0.466	0.671	0.582	0.643	0.741
	mean	0.717	0.641	0.656	0.736	0.612	0.531	0.612	0.634	0.737
D_F [57]	mean	0.873	0.876	-	-	0.765	0.679	-	-	0.885

表8：在MVTAD数据集上对随机选取的5个类别进行异常定位IoU的消融研究。报告了全部15个类别的平均值。CAVGA-R$_u^*$和CAVGA-R$_w^*$分别是在无监督和弱监督设置下采用扁平化z的我们的基础架构。“conv z”表示使用卷积z。

Method	CAVGA -R$_u^*$	CAVGA -R$_u^*$ + L_{ae}	CAVGA -R$_u$ + conv z	CAVGA -R$_u$ + conv z + L_{ae}	CAVGA -R$_w^*$	CAVGA -R$_w^*$ + L_{cga}	CAVGA -R$_w$ + conv z	CAVGA -R$_w$ + conv z + L_{cga}
Category								
Column ID	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8
Bottle	0.24	0.27	0.26	0.33	0.16	0.34	0.28	0.39
Hazelnut	0.16	0.26	0.31	0.47	0.51	0.76	0.67	0.79
Capsule	0.09	0.22	0.14	0.31	0.18	0.36	0.27	0.41
Metal Nut	0.28	0.38	0.34	0.45	0.25	0.38	0.28	0.46
Leather	0.55	0.71	0.64	0.79	0.72	0.79	0.75	0.84
mean	0.24	0.34	0.33	0.47	0.39	0.52	0.48	0.60

从推理过程中潜在变量的特征图映射。类似地，在弱监督设置中，我们使用$L + L_{adv} + L_{bce}$作为目标函数训练CAVGA-R$_w$和CAVGA-R$_w^*$，并在推理过程中根据分类器的预测计算异常注意力图。将列c_1与

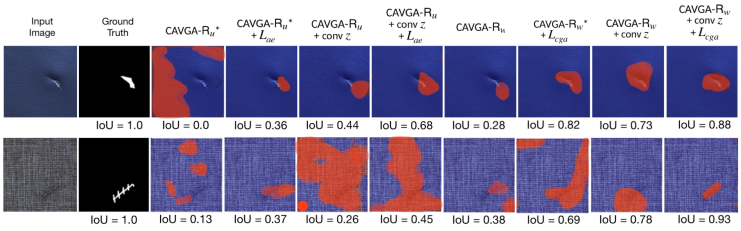


Fig. 5: Qualitative results of the ablation study to illustrate the performance of the anomaly localization on the MVTAD dataset

c_3 and c_5 with c_7 in Table 8, we observe that preserving the spatial relation of the input and latent variable through the convolutional z improves the IoU in anomaly localization without the use of L_{ae} in the unsupervised setting and L_{cga} in the weakly supervised setting. Furthermore, comparing column c_2 with c_4 and c_6 with c_8 in Table 8, we observe that using convolutional z in CAVGA-R$_u$ and CAVGA-R$_w$ outperforms using a flattened latent variable even with the help of L_{ae} in the unsupervised setting and L_{cga} in the weakly supervised setting.

Effect of attention expansion loss L_{ae}: To test the effectiveness of using L_{ae} in the unsupervised setting, we train CAVGA-R$_u^*$ and CAVGA-R$_u$ with eq. 4. During inference, the anomalous attention map is computed to localize the anomaly. Comparing column c_1 with c_2 and c_3 with c_4 in Table 8, we observe that L_{ae} enhances the IoU regardless of a flattened or convolutional latent variable.

Effect of complementary guided attention loss L_{cga}: We show the effectiveness of L_{cga} by training CAVGA-R$_w^*$ and CAVGA-R$_w$ using eq. 6. Comparing column c_5 with c_6 and c_7 with c_8 in Table 8, we find that using L_{cga} enhances the IoU regardless of a flattened or convolutional latent variable.

7 Conclusion

We propose an end-to-end convolutional adversarial variational autoencoder using guided attention which is a novel use of this technique for anomaly localization. Applicable to different network architectures, our attention expansion loss and complementary guided attention loss improve the performance of anomaly localization in the unsupervised and weakly supervised (with only 2% extra anomalous images for training) settings respectively. We quantitatively and qualitatively show that CAVGA outperforms the state-of-the-art (SOTA) anomaly localization methods on the MVTAD, mSTC and LAG datasets. We also show CAVGA's ability to outperform SOTA anomaly detection methods on the MVTAD, mSTC, LAG, MNIST, Fashion-MNIST and CIFAR-10 datasets.

Acknowledgments : This work was done when Shashanka was an intern and Kuan-Chuan was a Staff Scientist at Siemens. Shashanka's effort was partially supported by DARPA under Grant D19AP00032.

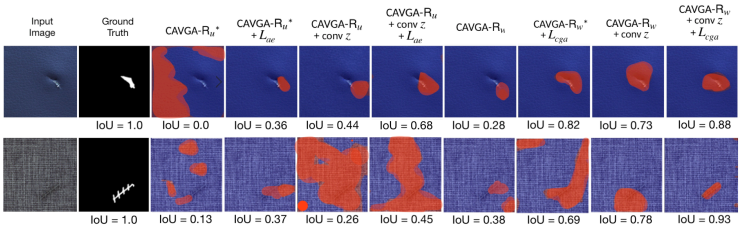


图5：消融研究的定性结果，以说明在MVTAD数据集上异常定位的性能

[c]在表8中对比第3列、c_5列与c_7列，我们观察到：通过卷积z保持输入与隐变量的空间关系，能在无监督场景下不使用L_{ae}、弱监督场景下不使用L_{cga}的情况下提升异常定位的IoU指标。此外，对比表8中c_2列与c_4列、c_6列与c_8列可知，在CAVGA-R$_u$和CAVGA-R$_w$中使用卷积z的策略，即便在无监督场景下借助L_{ae}、弱监督场景下借助L_{cga}，其效果仍优于使用扁平化隐变量的方法。

Effect of attention expansion loss L_{ae}: 为了测试在无监督设置中使用L_{ae}的有效性，我们使用公式4训练了CAVGA-R$_u^*$和CAVGA-R$_u$。在推理过程中，通过计算异常注意力图来定位异常。对比表8中的列c_1与c_2以及列c_3与c_4，我们观察到无论潜在变量是扁平结构还是卷积结构，L_{ae}都能提升IoU。

Effect of complementary guided attention loss L_{cga}: 我们通过使用等式6训练CAVGA-R$_w^*$和CAVGA-R$_w$来展示L_{cga}的有效性。对比表8中的列c_5与c_6以及列c_7与c_8，我们发现无论使用扁平化还是卷积化的潜变量，采用L_{cga}均能提升IoU。

7 Conclusion

我们提出了一种使用引导注意力的端到端卷积对抗变分自编码器，这是该技术在异常定位中的一种新颖应用。适用于不同的网络架构，我们的注意力扩展损失和互补引导注意力损失分别提升了无监督和弱监督（仅额外使用2%的异常图像进行训练）设置下的异常定位性能。我们通过定量和定性分析表明，CAVGA在MVTAD、mSTC和LAG数据集上优于当前最先进的异常定位方法。我们还展示了CAVGA在MVTAD、mSTC、LAG、MNIST、Fashion-MNIST和CIFAR-10数据集上超越最先进异常检测方法的能力。

Acknowledgments : 这项工作是在Shashanka实习期间以及Kuan-Chuan担任西门子科学家时完成的。Shashanka的研究工作部分由DARPA根据拨款D19AP00032支持。

Bibliography

[1] Code for iterative energy-based projection on a normal data manifold for anomaly localization. https://qiita.com/kogepan102/items/122b2862ad5a51180656, accessed on: 2020-02-29

[2] Abati, D., Porrello, A., Calderara, S., Cucchiara, R.: Latent space autoregression for novelty detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 481–490 (2019)

[3] Akcay, S., Atapour-Abarghouei, A., Breckon, T.P.: GANomaly: Semi-supervised anomaly detection via adversarial training. In: Asian Conference on Computer Vision. pp. 622–637. Springer (2018)

[4] Baur, C., Wiestler, B., Albarqouni, S., Navab, N.: Deep autoencoding models for unsupervised anomaly segmentation in brain mr images. In: International MICCAI Brainlesion Workshop. pp. 161–169. Springer (2018)

[5] Bergmann, P., Fauser, M., Sattlegger, D., Steger, C.: MVTec AD–a comprehensive real-world dataset for unsupervised anomaly detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 9592–9600 (2019)

[6] Bergmann, P., Löwe, S., Fauser, M., Sattlegger, D., Steger, C.: Improving unsupervised defect segmentation by applying structural similarity to autoencoders. In: International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP). vol. 5 (2019)

[7] Bian, J., Hui, X., Sun, S., Zhao, X., Tan, M.: A novel and efficient cvae-gan-based approach with informative manifold for semi-supervised anomaly detection. IEEE Access **7**, 88903–88916 (2019)

[8] Böttger, T., Ulrich, M.: Real-time texture error detection on textured surfaces with compressed sensing. Pattern Recognition and Image Analysis **26**(1), 88–94 (2016)

[9] Brock, A., Donahue, J., Simonyan, K.: Large scale GAN training for high fidelity natural image synthesis. In: International Conference on Learning Representations (2019)

[10] Cheng, K.W., Chen, Y.T., Fang, W.H.: Abnormal crowd behavior detection and localization using maximum sub-sequence search. In: Proceedings of the 4th ACM/IEEE international workshop on Analysis and retrieval of tracked events and motion in imagery stream. pp. 49–58. ACM (2013)

[11] Daniel, T., Kurutach, T., Tamar, A.: Deep variational semi-supervised novelty detection. arXiv preprint arXiv:1911.04971 (2019)

[12] Deecke, L., Vandermeulen, R., Ruff, L., Mandt, S., Kloft, M.: Image anomaly detection with generative adversarial networks. In: Joint European Conference on Machine Learning and Knowledge Discovery in Databases. pp. 3–17. Springer (2018)

Bibliography

[1] 用于异常定位的正态数据流形上迭代能量投影的代码。 https://qiita.com/kogepan102/items/122b2862ad5a51180656，访问日期：2020-02-29[2] Abati, D., Porrello, A., Calderara, S., Cucchiara, R.: 用于新颖性检测的潜在空间自回归。见：IEEE计算机视觉与模式识别会议论文集。第481–490页 (2019)[3] Akcay, S., Atapour-Abarghouei, A., Breckon, T.P.: GANomaly：通过对抗训练进行半监督异常检测。见：亚洲计算机视觉会议。第622–637页。Springer (2018)[4] Baur, C., Wiestler, B., Albarqouni, S., Navab, N.: 用于脑部MR图像中无监督异常分割的深度自编码模型。见：国际MICCAI脑部病灶研讨会。第161–169页。Springer (2018)[5] Bergmann, P., Fauser, M., Sattlegger, D., Steger, C.: MVTec AD——一个用于无监督异常检测的综合性真实世界数据集。见：IEEE计算机视觉与模式识别会议论文集。第9592–9600页 (2019)[6] Bergmann, P., Löwe, S., Fauser, M., Sattlegger, D., Steger, C.: 通过将结构相似性应用于自编码器来改进无监督缺陷分割。见：计算机视觉、成像和计算机图形学理论与应用国际联合会议（VISIGRAPP）。第5卷 (2019)[7] Bian, J., Hui, X., Sun, S., Zhao, X., Tan, M.: 一种基于CVAE-GAN的具有信息流形的高效半监督异常检测新方法。IEEE Access **7**, 88903–88916 (2019)[8] Böttger, T., Ulrich, M.: 基于压缩感知的纹理表面实时纹理错误检测。模式识别与图像分析 **26**(1), 88–94 (2016)[9] Brock, A., Donahue, J., Simonyan, K.: 用于高保真自然图像合成的大规模GAN训练。见：国际学习表征会议 (2019)

[[10] Cheng, K.W., Chen, Y.T., Fang, W.H.: 使用最大子序列搜索进行异常人群行为检测与定位。见：第四届ACM/IEEE图像流中事件与运动分析与检索国际研讨会论文集。第49–58页。ACM (2013)

[11] Daniel, T., Kurutach, T., Tamar, A.: 深度变分半监督新颖性检测。arXiv 预印本 arXiv:1911.04971 (2019)

[12] Deecke, L., Vandermeulen, R., Ruff, L., Mandt, S., Kloft, M.: 使用生成对抗网络进行图像异常检测。见：欧洲机器学习与数据库知识发现联合会议。第3–17页。Springer (2018)

[13] Dehaene, D., Frigo, O., Combrexelle, S., Eline, P.: Iterative energy-based projection on a normal data manifold for anomaly localization. International Conference on Learning Representations (2020)
[14] Dieng, A.B., Kim, Y., Rush, A.M., Blei, D.M.: Avoiding latent variable collapse with generative skip models. In: The 22nd International Conference on Artificial Intelligence and Statistics. pp. 2397–2405 (2019)
[15] Dimokranitou, A.: Adversarial autoencoders for anomalous event detection in images. Ph.D. thesis (2017)
[16] Gong, D., Liu, L., Le, V., Saha, B., Mansour, M.R., Venkatesh, S., Hengel, A.v.d.: Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 1705–1714 (2019)
[17] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: Advances in neural information processing systems. pp. 2672–2680 (2014)
[18] Gutoski, M., Aquino, N.M.R., Ribeiro, M., Lazzaretti, E., Lopes, S.: Detection of video anomalies using convolutional autoencoders and one-class support vector machines. In: XIII Brazilian Congress on Computational Intelligence, 2017 (2017)
[19] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 770–778 (2016)
[20] Hendrycks, D., Mazeika, M., Dietterich, T.G.: Deep anomaly detection with outlier exposure. In: International Conference on Learning Representations (2019)
[21] Higgins, I., Matthey, L., Pal, A., Burgess, C., Glorot, X., Botvinick, M., Mohamed, S., Lerchner, A.: beta-VAE: Learning basic visual concepts with a constrained variational framework. International Conference on Learning Representations **2**(5), 6 (2017)
[22] Kimura, D., Chaudhury, S., Narita, M., Munawar, A., Tachibana, R.: Adversarial discriminative attention for robust anomaly detection. In: The IEEE Winter Conference on Applications of Computer Vision (WACV) (March 2020)
[23] Kingma, D.P., Welling, M.: Auto-encoding variational bayes. In: International Conference on Learning Representations (2014)
[24] Kiran, B., Thomas, D., Parakkal, R.: An overview of deep learning based methods for unsupervised and semi-supervised anomaly detection in videos. Journal of Imaging **4**(2), 36 (2018)
[25] Krizhevsky, A., Hinton, G., et al.: Learning multiple layers of features from tiny images. Tech. rep., Citeseer (2009)
[26] Larsen, A.B.L., Sønderby, S.K., Larochelle, H., Winther, O.: Autoencoding beyond pixels using a learned similarity metric. In: International Conference on Machine Learning (2016)
[27] LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., et al.: Gradient-based learning applied to document recognition. Proceedings of the IEEE **86**(11), 2278–2324 (1998)

[13] Dehaene, D., Frigo, O., Combrexelle, S., Eline, P.: 基于迭代能量投影于正常数据流形的异常定位。国际学习表征会议 (2020)[14] Dieng, A.B., Kim, Y., Rush, A. M., Blei, D.M.: 使用生成跳跃模型避免隐变量坍缩。见：第22届人工智能与统计学国际会议。第2397–2405页 (2019)[15] Dimokranitou, A.: 用于图像中异常事件检测的对抗自编码器。博士论文 (2017)[16] Gong, D., Liu, L., Le, V., Saha, B., Mansour, M.R., Venkatesh, S., Hengel, A.v.d.: 记忆常态以检测异常：用于无监督异常检测的记忆增强深度自编码器。见：IEEE国际计算机视觉会议论文集。第1705–1714页 (2019)[17] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: 生成对抗网络。见：神经信息处理系统进展。第2672–2680页 (2014)[18] Gutoski, M., Aquino, N.M.R., Ribeiro, M., Lazzaretti, E., Lopes, S.: 使用卷积自编码器和一类支持向量机进行视频异常检测。见：第十三届巴西计算智能大会，2017 (2017)[19] He, K., Zhang, X., Ren, S., Sun, J.: 用于图像识别的深度残差学习。见：IEEE计算机视觉与模式识别会议论文集。第770–778页 (2016)[20] Hendrycks, D., Mazeika, M., Dietterich, T.G.: 基于离群暴露的深度异常检测。国际学习表征会议 (2019)[21] Higgins, I., Matthey, L., Pal, A., Burgess, C., Glorot, X., Botvinick, M., Mohamed, S., Lerchner, A.: beta-VAE：使用约束变分框架学习基本视觉概念。国际学习表征会议 **2**(5), 6 (2017)[22] Kimura, D., Chaudhury, S., Narita, M., Munawar, A., Tachibana, R.: 用于鲁棒异常检测的对抗判别注意力。见：IEEE冬季计算机视觉应用会议 (WACV) (2020年3月)[23] Kingma, D.P., Welling, M.: 自动编码变分贝叶斯。见：国际学习表征会议 (2014)[24] Kiran, B., Thomas, D., Parakkal, R.: 基于深度学习的视频无监督和半监督异常检测方法综述。影像学杂志 **4**(2), 36 (2018)[25] Krizhevsky, A., Hinton, G., 等：从微小图像中学习多层特征。技术报告，Citeseer (2009)[26] Larsen, A.B.L., Sønderby, S.K., Larochelle, H., Winther, O.: 使用学习到的相似性度量进行超越像素的自编码。见：国际机器学习会议 (2016)[27] LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., 等：基于梯度的学习应用于文档识别。IEEE会刊 **86**(11), 2278–2324 (1998)

[28] Li, K., Wu, Z., Peng, K.C., Ernst, J., Fu, Y.: Tell me where to look: Guided attention inference network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 9215–9223 (2018)
[29] Li, L., Xu, M., Wang, X., Jiang, L., Liu, H.: Attention based glaucoma detection: A large-scale database and cnn model. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2019)
[30] Li, X., Kiringa, I., Yeap, T., Zhu, X., Li, Y.: Exploring deep anomaly detection methods based on capsule net. International Conference on Machine Learning 2019 Workshop on Uncertainty and Robustness in Deep Learning (2019)
[31] Liu, W., Luo, W., Lian, D., Gao, S.: Future frame prediction for anomaly detection–a new baseline. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 6536–6545 (2018)
[32] Liu, W., Li, R., Zheng, M., Karanam, S., Wu, Z., Bhanu, B., Radke, R.J., Camps, O.: Towards visually explaining variational autoencoders. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2020)
[33] Liu, Z., Luo, P., Wang, X., Tang, X.: Deep learning face attributes in the wild. In: Proceedings of International Conference on Computer Vision (ICCV) (December 2015)
[34] Makhzani, A., Shlens, J., Jaitly, N., Goodfellow, I., Frey, B.: Adversarial autoencoders. In: International Conference on Learning Representations (2016)
[35] Masana, M., Ruiz, I., Serrat, J., van de Weijer, J., Lopez, A.M.: Metric learning for novelty and anomaly detection. In: British Machine Vision Conference (BMVC) (2018)
[36] Matteoli, S., Diani, M., Theiler, J.: An overview of background modeling for detection of targets and anomalies in hyperspectral remotely sensed imagery. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing **7**(6), 2317–2336 (2014)
[37] Napoletano, P., Piccoli, F., Schettini, R.: Anomaly detection in nanofibrous materials by CNN-based self-similarity. Sensors **18**(1), 209 (2018)
[38] Nguyen, P., Liu, T., Prasad, G., Han, B.: Weakly supervised action localization by sparse temporal pooling network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 6752–6761 (2018)
[39] Oquab, M., Bottou, L., Laptev, I., Sivic, J.: Is object localization for free?-weakly-supervised learning with convolutional neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 685–694 (2015)
[40] Pawlowski, N., Lee, M.C., Rajchl, M., McDonagh, S., Ferrante, E., Kamnitsas, K., Cooke, S., Stevenson, S., Khetani, A., Newman, T., et al.: Unsupervised lesion detection in brain CT using bayesian convolutional autoencoders. In: Medical Imaging with Deep Learning (2018)
[41] Perera, P., Nallapati, R., Xiang, B.: OCGAN: One-class novelty detection using GANs with constrained latent representations. In: Proceedings of the

[28] Li, K., Wu, Z., Peng, K.C., Ernst, J., Fu, Y.: 告诉我看哪里：引导注意力推断网络。见：IEEE计算机视觉与模式识别会议论文集。第9215–9223页 (2018)[29] Li, L., Xu, M., Wang, X., Jiang, L., Liu, H.: 基于注意力的青光眼检测：大规模数据库与CNN模型。见：IEEE计算机视觉与模式识别会议 (CVPR) (2019年6月)[30] Li, X., Kiringa, I., Yeap, T., Zhu, X., Li, Y.: 探索基于胶囊网络的深度异常检测方法。机器学习国际会议2019深度学习不确定性与鲁棒性研讨会 (2019)[31] Liu, W., Luo, W., Lian, D., Gao, S.: 用于异常检测的未来帧预测——一个新基线。见：IEEE计算机视觉与模式识别会议论文集。第6536–6545页 (2018)[32] Liu, W., Li, R., Zheng, M., Karanam, S., Wu, Z., Bhanu, B., Radke, R.J., Camps, O.: 走向视觉解释变分自编码器。IEEE计算机视觉与模式识别会议论文集 (2020)[33] Liu, Z., Luo, P., Wang, X., Tang, X.: 在野外深度学习人脸属性。见：国际计算机视觉会议 (ICCV)论文集 (2015年12月)[34] Makhzani, A., Shlens, J., Jaitly, N., Goodfellow, I., Frey, B.: 对抗自编码器。见：国际学习表征会议 (2016)[35] Masana, M., Ruiz, I., Serrat, J., van de Weijer, J., Lopez, A.M.: 用于新颖性与异常检测的度量学习。见：英国机器视觉会议 (BMVC) (2018)[36] Matteoli, S., Diani, M., Theiler, J.: 高光谱遥感图像中目标与异常检测的背景建模概述。IEEE应用地球观测与遥感专题期刊 **7**(6), 2317–2336 (2014)[37] Napoletano, P., Piccoli, F., Schettini, R.: 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)[38] Nguyen, P., Liu, T., Prasad, G., Han, B.: 通过稀疏时序池化网络进行弱监督动作定位。见：IEEE计算机视觉与模式识别会议论文集。第6752–6761页 (2018)[39] Oquab, M., Bottou, L., Laptev, I., Sivic, J.: 物体定位是免费的吗？——基于卷积神经网络的弱监督学习。见：IEEE计算机视觉与模式识别会议论文集。第685–694页 (2015)[40] Pawlowski, N., Lee, M.C., Rajchl, M., McDonagh, S., Ferrante, E., Kamnitsas, K., Cooke, S., Stevenson, S., Khetani, A., Newman, T., 等：使用贝叶斯卷积自编码器进行脑部CT无监督病灶检测。见：医学影像深度学习会议 (2018)[41] Perera, P., Nallapati, R., Xiang, B.: OCGAN：使用具有约束潜在表征的GAN进行单类新颖性检测。见：

IEEE Conference on Computer Vision and Pattern Recognition. pp. 2898–2906 (2019)
[42] Radford, A., Metz, L., Chintala, S.: Unsupervised representation learning with deep convolutional generative adversarial networks. In: International Conference on Learning Representations (2016)
[43] Ravanbakhsh, M., Sangineto, E., Nabi, M., Sebe, N.: Training adversarial discriminators for cross-channel abnormal event detection in crowds. In: 2019 IEEE Winter Conference on Applications of Computer Vision (WACV). pp. 1896–1904. IEEE (2019)
[44] Ruff, L., Vandermeulen, R.A., Görnitz, N., Binder, A., Müller, E., Müller, K.R., Kloft, M.: Deep semi-supervised anomaly detection. International Conference on Learning Representations (2020)
[45] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al.: ImageNet large scale visual recognition challenge. International journal of computer vision **115**(3), 211–252 (2015)
[46] Sabokrou, M., Khalooei, M., Fathy, M., Adeli, E.: Adversarially learned one-class classifier for novelty detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3379–3388 (2018)
[47] Sabokrou, M., Pourreza, M., Fayyaz, M., Entezari, R., Fathy, M., Gall, J., Adeli, E.: Avid: Adversarial visual irregularity detection. In: Asian Conference on Computer Vision. pp. 488–505. Springer (2018)
[48] Schlegl, T., Seeböck, P., Waldstein, S.M., Schmidt-Erfurth, U., Langs, G.: Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In: International Conference on Information Processing in Medical Imaging. pp. 146–157. Springer (2017)
[49] Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-cam: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 618–626 (2017)
[50] Smilkov, D., Thorat, N., Kim, B., Viégas, F., Wattenberg, M.: SmoothGrad: removing noise by adding noise. arXiv preprint arXiv:1706.03825 (2017)
[51] Springenberg, J.T., Dosovitskiy, A., Brox, T., Riedmiller, M.: Striving for simplicity: The all convolutional net. arXiv preprint arXiv:1412.6806 (2014)
[52] Steger, C.: Similarity measures for occlusion, clutter, and illumination invariant object recognition. In: Joint Pattern Recognition Symposium. pp. 148–154. Springer (2001)
[53] Vu, H.S., Ueta, D., Hashimoto, K., Maeno, K., Pranata, S., Shen, S.M.: Anomaly detection with adversarial dual autoencoders. arXiv preprint arXiv:1902.06924 (2019)
[54] Wang, X., Xu, M., Li, L., Wang, Z., Guan, Z.: Pathology-aware deep network visualization and its application in glaucoma image synthesis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 423–431. Springer (2019)
[55] Wang, Z., Fan, M., Muknahallipatna, S., Lan, C.: Inductive multi-view semi-supervised anomaly detection via probabilistic modeling. In: 2019 IEEE

IEEE计算机视觉与模式识别会议。第2898–2906页 (2019) [42] Radford, A., Metz, L., Chintala, S.: 使用深度卷积生成对抗网络的无监督表示学习。见：国际学习表征会议 (2016) [43] Ravanbakhsh, M., Sangineto, E., Nabi, M., Sebe, N.: 训练对抗性判别器用于人群中的跨通道异常事件检测。见：2019年IEEE冬季计算机视觉应用会议 (WACV)。第1896–1904页。IEEE (2019) [44] Ruff, L., Vandermeulen, R.A., Görnitz, N., Binder, A., Müller, E., Müller, K.R., Kloft, M.: 深度半监督异常检测。国际学习表征会议 (2020) [45] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., 等：ImageNet大规模视觉识别挑战赛。国际计算机视觉期刊 **115**(3)，第211–252页 (2015) [46] Sabokrou, M., Khalooei, M., Fathy, M., Adeli, E.: 对抗性学习的一类分类器用于新颖性检测。见：IEEE计算机视觉与模式识别会议论文集。第3379–3388页 (2018) [47] Sabokrou, M., Pourreza, M., Fayyaz, M., Entezari, R., Fathy, M., Gall, J., Adeli, E.: Avid：对抗性视觉不规则性检测。见：亚洲计算机视觉会议。第488–505页。Springer (2018) [48] Schlegl, T., Seeböck, P., Waldstein, S.M., Schmidt-Erfurth, U., Langs, G.: 使用生成对抗网络进行无监督异常检测以指导标记发现。见：医学图像信息处理国际会议。第146–157页。Springer (2017) [49] Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-CAM：通过基于梯度的定位从深度网络进行可视化解释。见：IEEE国际计算机视觉会议论文集。第618–626页 (2017) [50] Smilkov, D., Thorat, N., Kim, B., Viégas, F., Wattenberg, M.: SmoothGrad：通过添加噪声去除噪声。arXiv预印本 arXiv:1706.03825 (2017) [51] Springenberg, J.T., Dosovitskiy, A., Brox, T., Riedmiller, M.: 追求简洁性：全卷积网络。arXiv预印本 arXiv:1412.6806 (2014) [52] Steger, C.: 用于遮挡、杂波和光照不变物体识别的相似性度量。见：联合模式识别研讨会。第148–154页。Springer (2001) [53] Vu, H.S., Ueta, D., Hashimoto, K., Maeno, K., Pranata, S., Shen, S.M.: 使用对抗性双自编码器进行异常检测。arXiv预印本 arXiv:1902.06924 (2019) [54] Wang, X., Xu, M., Li, L., Wang, Z., Guan, Z.: 病理感知深度网络可视化及其在青光眼图像合成中的应用。见：医学图像计算与计算机辅助干预国际会议。第423–431页。Springer (2019) [55] Wang, Z., Fan, M., Muknahallipatna, S., Lan, C.: 通过概率建模进行归纳式多视图半监督异常检测。见：2019年IEEE

International Conference on Big Knowledge (ICBK). pp. 257–264. IEEE (2019)
[56] Wolf, L., Benaim, S., Galanti, T.: Unsupervised learning of the set of local maxima. International Conference on Learning Representations (2019)
[57] Xiao, H., Rasul, K., Vollgraf, R.: Fashion-MNIST: a novel image dataset for benchmarking machine learning algorithms. arXiv preprint arXiv:1708.07747 (2017)
[58] Zagoruyko, S., Komodakis, N.: Paying more attention to attention: Improving the performance of convolutional neural networks via attention transfer. In: International Conference on Learning Representations (2017)
[59] Zenati, H., Foo, C.S., Lecouat, B., Manek, G., Chandrasekhar, V.R.: Efficient GAN-based anomaly detection. arXiv preprint arXiv:1802.06222 (2018)
[60] Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A.: Learning deep features for discriminative localization. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2921–2929 (2016)

国际大数据知识会议（ICBK）。第257–264页。IEEE（2019）[56] Wolf, L., Benaim, S., Galanti, T.: 局部最大值集合的无监督学习。国际学习表征会议（2019）[57] Xiao, H., Rasul, K., Vollgraf, R.: Fashion-MNIST：一种用于基准测试机器学习算法的新型图像数据集。arXiv预印本 arXiv:1708.07747（2017）[58] Zagoruyko, S., Komodakis, N.: 更加关注注意力：通过注意力转移提升卷积神经网络的性能。载于：国际学习表征会议（2017）[59] Zenati, H., Foo, C.S., Lecouat, B., Manek, G., Chandrasekhar, V.R.: 基于GAN的高效异常检测。arXiv预印本 arXiv:1802.06222（2018）[60] Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A.: 学习用于判别性定位的深度特征。载于：IEEE计算机视觉与模式识别会议论文集。第2921–2929页（2016）