

# PaDim : a Patch Distribution Modeling Framework for Anomaly Detection and Localization

本文提出一个名为PaDiM (Patch Distribution Modeling) 的新框架，它用于在单类学习设置中同时检测和定位图像中的异常。PaDiM利用预训练的卷积神经网络 (CNN) 进行图像词向量表示，并使用多变量高斯分布来获得正常类别的概率表示。此外，PaDiM还利用CNN不同语义层次之间的相关性来更好地定位异常。

## introduction

介绍了图像异常检测的任务，并说明必要性。

还讨论了现有方法的局限性，包括它们可能需要深度神经网络训练，或者在测试时使用K-最近邻 (K-NN) 算法，这可能会随着训练数据集的增长而增加时间和空间复杂性。

作者提出了PaDim模型，PaDim使用预训练的CNN进行词向量提取，并具有两个特点：每个像素块位置由多变量高斯分布描述；PaDiM考虑了预训练CNN不同语义层次之间的相关性。

## RELATED WORK

作者详细讨论了异常检测和定位方法，并将这些方法分为两大类：基于重建的方法和基于词向量相似性的方法

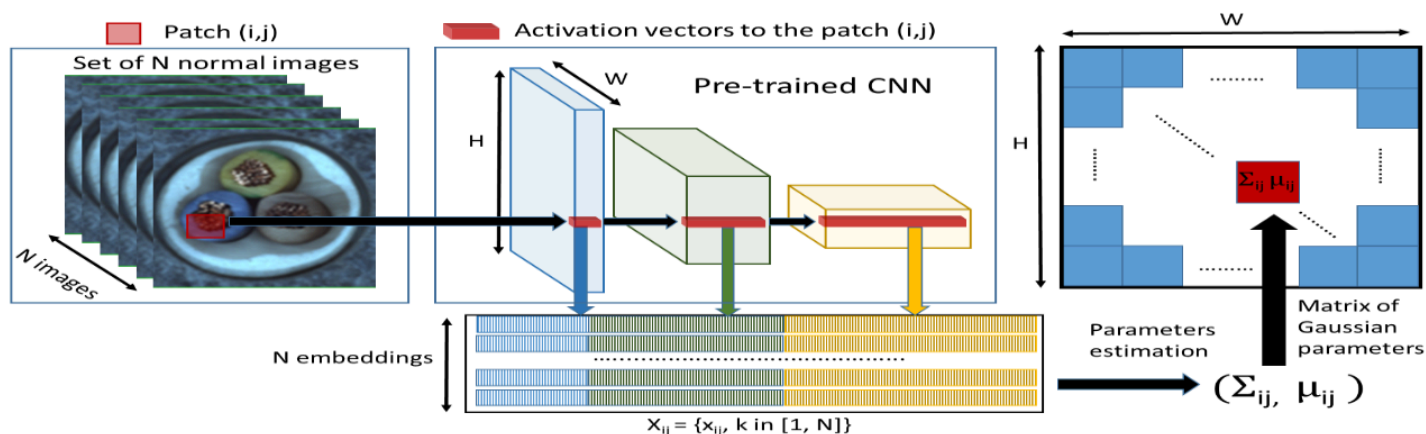
1. **基于重建的方法**：这类方法广泛用于异常检测和定位。使用的神经网络架构包括自编码器 (AE)、变分自编码器 (VAE) 和生成对抗网络 (GAN)。这些模型仅在正常训练图像上进行训练以重建图像。因此，异常图像可以作为重建效果不佳的图像被识别出来。在图像级别上，最简单的方法是将重建误差作为异常分数。此外，还可以使用潜在空间、中间激活或判别器的额外信息来更好地识别异常图像。然而，基于重建的方法的性能受限于自编码器有时也能为异常图像提供良好的重建结果。

2. **基于词向量相似性的方法**：这些方法使用深度神经网络提取有意义的向量，描述整个图像进行异常检测或图像块进行异常定位。这些方法通常给出有希望的结果，但往往缺乏可解释性，因为无法知道异常图像的哪一部分导致了高异常分数。在这种情况下，异常分数是测试图像的嵌入向量与代表训练数据集中正常性的参考向量之间的距离。正常的参考可以是包含正常图像嵌入的n-sphere的中心、高斯分布的参数或整个正常嵌入向量集合。

最后，作者提出的方法PaDiM生成用于**异常定位的补丁词向量**，类似于上述方法。然而，PaDiM通过一组高斯分布来描述正常类别，这些分布还模拟了使用的预训练CNN模型的语义层次之间的相关性。作者使用预训练网络来训练，所以预测阶段的时间复杂性低，且与训练数据集的大小无关。预训练网络包括ResNet、Wide-Resnet和EfficientNet。

## PATCH DISTRIBUTION MODELING

作者详细介绍了PaDiM框架的三个核心组成部分：嵌入提取（Embedding extraction）、正常性的学习（Learning of the normality）以及推断：异常图的计算（Inference: computation of the anomaly map），模型如下：



### 1. 嵌入提取 (Embedding extraction)

PaDiM使用预训练的卷积神经网络（CNN）来生成补丁嵌入向量。在训练阶段，正常图像的每个补丁与其在预训练CNN激活图中空间对应的激活向量相关联。不同层的激活向量被连接起来，形成包含不同语义层次和分辨率信息的嵌入向量，以编码细粒度和全局上下文。由于激活图的分辨率低于输入图像，许多像素具有相同的嵌入，从而在原始图像分辨率中形成无重叠的像素补丁。因此，输入图像可以被划分为一个网格，每个网格位置(i, j)与一个嵌入向量 $x_{ij}$ 相关联。

### 2. 正常性的学习 (Learning of the normality)

为了学习位置(i, j)的正常图像特征，首先计算N个正常训练图像在(i, j)处的补丁嵌入向量集合 $X_{ij}$ 。通过假设 $X_{ij}$ 是由多变量高斯分布 $N(\mu_{ij}, \sigma_{ij})$ 生成的，其中 $\mu_{ij}$ 是 $X_{ij}$ 的样本均值，样本协方差 $\sigma_{ij}$ 是这样估计的：

$$\Sigma_{ij} = \frac{1}{N-1} \sum_{k=1}^N (x_{ij}^k - \mu_{ij})(x_{ij}^k - \mu_{ij})^T + \varepsilon I$$

这里的正则化项 $\varepsilon I$ 确保样本协方差矩阵 $\Sigma_{ij}$ 是满秩且可逆的。最终，每个可能的补丁位置与一个多变量高斯分布相关联，该分布捕获了不同层次的信息以及层间的相关性。

### 3. 推断：异常图的计算 (Inference: computation of the anomaly map)

PaDiM使用马氏距离 (Mahalanobis distance)  $M(x_{ij})$ 来给测试图像中位置(i, j)的补丁赋予异常分数。 $M(x_{ij})$ 可以解释为测试补丁嵌入 $x_{ij}$ 与学习到的分布 $N(\mu_{ij}, \sigma_{ij})$ 之间的距离，计算如下：

$$M(x_{ij}) = \sqrt{(x_{ij} - \mu_{ij})^T \sigma_{ij}^{-1} (x_{ij} - \mu_{ij})}$$

## EXPERIMENTS

在实验部分，作者详细描述了如何评价PaDim框架的性能，包括所使用的数据集、度量标准、实验设置以及实验结果。

### 1. 数据集和度量标准 (Datasets and metrics)

- **度量标准：** 为了评估定位性能，作者计算了两个不受阈值影响的指标：接收者操作特征曲线下面积 (AUROC) 和每个区域重叠分数 (PRO-score)。AUROC考虑了正确分类为异常像素的百分比，而PRO-score则考虑了在假阳性率在0到0.3之间的正确分类像素率的平均值。
- **数据集：** 作者首先在MVTec AD数据集上评估模型，这是一个为工业质量控制设计的异常定位算法测试而设计的单类学习数据集，包含15个类别，大约240张图像。此外，为了更现实地评估模型性能，作者创建了MVTec AD的一个修改版本 (RdMVTec AD)，在该版本中，对训练集和测试集都应用了随机旋转和裁剪。作者还在Shanghai Tech Campus (STC) 数据集上测试了PaDiM，该数据集模拟了静态摄像头的视频监控，包含13个场景的274,515个训练帧和42,883个测试帧。

### 2. 实验设置

- **训练配置：** PaDiM使用不同的骨干网络进行训练，包括ResNet18、Wide ResNet-50-2和EfficientNet-B5，所有这些网络都预先在ImageNet上进行了训练。作者从ResNet的第一个三层提取补丁嵌入向量，以结合不同语义层次的信息，同时保持足够高的分辨率以进行定位任务。对于EfficientNet-B5，作者从特定的层提取嵌入向量。
- **预处理和数据增强：** 作者对MVTec AD图像进行了大小调整和中心裁剪，对STC图像仅进行了大小调整。使用双三次插值进行图像和定位图的大小调整，并在异常图上使用高斯滤波。

# result

1. **消融研究：** 作者首先评估了在PaDiM中模拟语义层次之间的相关性的影响，并通过降维来简化方法的可能性。通过表格展示了在MVTec AD上使用不同CNN层进行异常定位性能的结果，并讨论了聚合不同层次信息的优势。
2. **与最先进技术的比较：** 作者将PaDiM与当前最先进方法在MVTec AD上的异常定位性能进行了比较，并展示了AUROC和PRO-score结果。PaDiM在多个类别上超越了其他方法，特别是在纹理类别上表现尤为出色。
3. **在非对称数据集上的异常定位：** 为了估计异常定位方法的鲁棒性，作者在修改版的MVTec AD（Rd-MVTec AD）上训练并评估了PaDiM和其他几种最先进方法的性能。结果表明，PaDiM在纹理和对象类别的PRO-score和AUROC上优于其他模型，并且在非对齐图像上比其他测试方法更鲁棒。
4. **复杂度：** 进行了时间和空间复杂度的分析，在PaDim中，训练时间复杂度与数据集大小线性相关，因为使用了预训练的CNN，在空间复杂度上依赖于图像分辨率。