

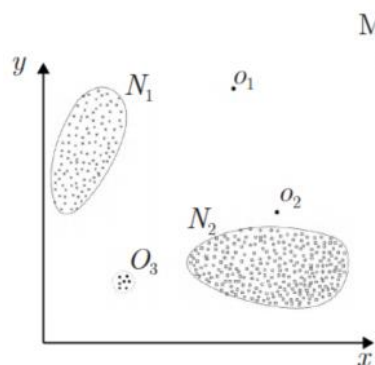
1. 背景概述



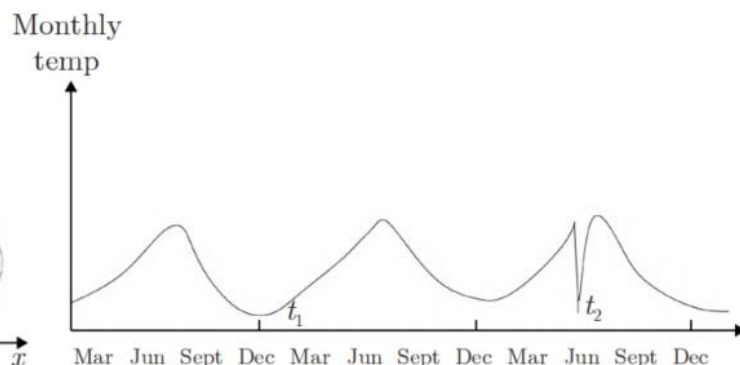
北京交通大学

图像异常检测

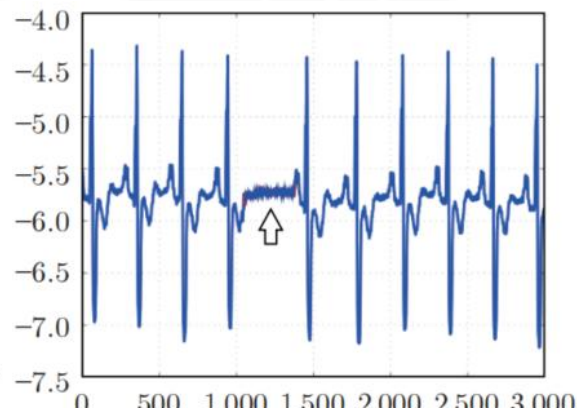
- ▲ 基本定义：是一种利用**无标注样本或正常样本**构建检测模型，检测与期望模式存在差异的异常样本的方法，是计算机视觉领域的重要研究方向。
- ▲ 前景概述：图像异常检测在各领域中都有广泛应用，前景广阔，以在工业领域的应用为例：全球工业视觉检测的市场规模**2025年预计达到192亿美元**，其中我国占比约为30%，将持续保持14%的年度平均增长率。
- ▲ 相关术语：
 - 异常：**离群值**，是与其他观测结果完全不同，以至于推断其是由不同机制产生的观测值。
异常分为**点异常、上下文异常与集群异常**。



(a) 点异常



(b) 上下文异常



(c) 集群异常

1. 背景概述

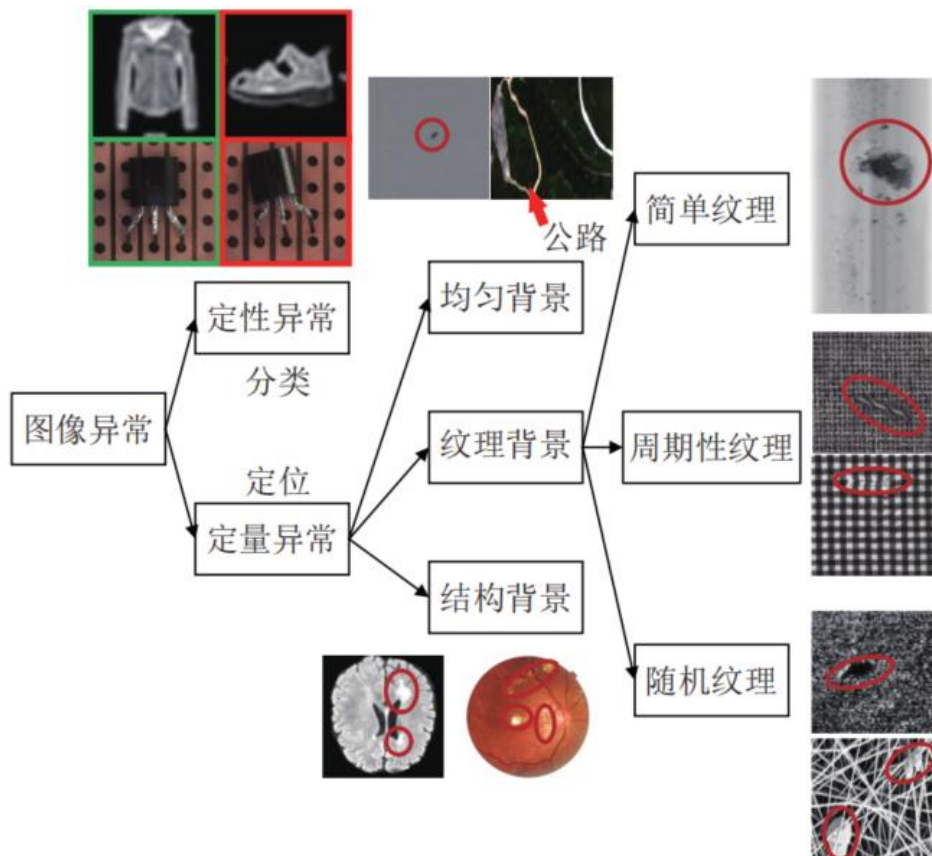


北京交通大学

图像异常检测

▲ 相关术语:

- 图像中的异常: 图像**每个像素点的像素值对应不同观测结果**, 由于像素值的多样性, 仅分析单个点的像素值很难判断是否异常。



▲ 相关领域存在的问题:

(以交通、工业、医学三大领域为例)

- 交通监测: 异常图像与背景差异小, **类型多样**, 存在**大量噪声**, 在自然环境下成像存在大量干扰。
- 工业检测: 生产环境下收集的图像中**无缺陷样本占多数**, 缺陷样本占少数。
- 医学图像: 医学影像中的带有病变区域样本稀少; 医学病变样本**人工标注耗时耗力**。

2. 经典方法

传统图像 异常检测 方法

训练模型以描述正常图像，
根据待检图像与模型之间的
差异程度进行异常检测

基于 模板匹配

设计思路：建立待测图像和模板图像的对应关系
通过比较识别异常区域；
典型方法：点匹配、线匹配、区域匹配；
优点：简单有效，适用采集环境高度可控场景；
缺点：不适用多变的场景目标。

基于 统计模型

设计思路：通过统计模型描述正常图像的特征分布，
远离该分布的图像区域判定异常；
典型方法：高斯混合模型、马尔科夫随机场；
优点：检测速度快；有严格的数学理论支撑；
缺点：需大量样本训练；仅适用简单背景图像。

基于 图像分解

设计思路：分解图像为代表背景的低秩矩阵和代表异常
区域的稀疏矩阵；
典型方法：低秩分解；
优点：无需样本训练；数学理论支撑完整；
缺点：速度慢，不适用结构复杂图像。

基于 频域分析

设计思路：通过编辑图像的频谱信息以消除重复背景纹理，
凸显异常区域；
典型方法：背景频谱消除法，纯相位傅里叶变换法POFT；
优点：无需样本训练；检测速度快；
缺点：通用性不好；仅适用重复纹理图像。

基于 稀疏编码

设计思路：通过稀疏编码学习正常样本表示，通过重构误差
和稀疏度检测异常；
典型方法：稀疏编码；
优点：通用性好，适用各类图像；
缺点：保存字典占用额外空间；速度慢。

基于 分类面构建

设计思路：在特征空间建立分类面以区分现有正常样本
和潜在异常样本；
典型方法：OC-SVM、SVDD；
优点：通用性好；检测速度快；
缺点：各项参数的设计和选择复杂。



北京交通大学

2. 经典方法



北京交通大学

基于深度学习的图像异常检测

基于距离度量

设计思路：通过特征提取网络将正常图像**映射到特征空间**，减小正常特征之间距离，根据待测图像特征到正常特征中心的**距离**进行异常检测；

典型方法：Deep SVDD；

优点：适用广泛；模型结构简单；

缺点：模型易退化；定位异常区域困难。

基于分类面构建

设计思路：通过**几何变换**增广数据，训练分类模型，利用置信度检测异常；寻找与正常样本近似的图像作为负样本，**构建分类面**；

典型方法：多类别训练法；结合OC-SVM或SVDD法；

优点：训练简单，检测精度高；

缺点：模型复杂；不适用纹理图像。

基于图像重构

设计思路：通过**自编码器**或**生成对抗网络**重构图像；根据**重构误差**进行异常检测；

典型方法：Memory-AE；GANomaly；

优点：训练阶段**无需额外样本**；异常定位精度高；

缺点：AE重构图像**效果欠佳**；**训练复杂**。

结合传统方法

设计思路：**特征提取阶段**利用预训练网络或自编码器，**决策阶段**利用传统方法进行异常判别；

典型方法：特征提取网络+OC-SVM；

优点：**通用性好**，**速度快**，检测精度高于传统方法精度；

缺点：算法复杂度、检测精度

3. 研究数据



数据集名称	年份	说明
MNIST	1998	MNIST是一个手写数字数据集，包括60,000个训练样本和10,000个测试样本。该数据集广泛用于计算机视觉领域的基准测试。
CIFAR-10	2009	CIFAR-10是一个包含10个类别的彩色图像数据集，每个类别有6,000张32x32像素的图像。该数据集广泛用于图像分类和物体识别任务。
CIFAR-100	2009	CIFAR-100是一个包含100个类别的彩色图像数据集，每个类别有600张32x32像素的图像。该数据集更具挑战性，适用于更复杂的图像分类和异常检测任务。
KITTI	2012	KITTI是一个用于自动驾驶场景的数据集，包括城市街道中的RGB图像、深度图像、3D点云以及车辆位置和运动信息。
COCO	2014	COCO是一个广泛用于物体检测、分割和姿态估计任务的数据集，包括330,000张图像和250,000个注释物体。
Anomaly Detection Image Dataset (ADID)	2016	ADID是一个专门用于图像异常检测的数据集，包括原始图像和异常图像。该数据集广泛用于研究异常检测算法的性能。
MVTec AD	2019	MVTec AD是一个包含多种物体和场景的图像异常检测数据集，包括纹理、物体形态、灰度、划痕等缺陷类型。该数据集广泛用于评估现代异常检测算法的性能。
SVHN Anomaly Detection	2020	SVHN Anomaly Detection是一个来自于街景房屋数字(SVHN)数据集的异常检测数据集，包括原始图像和椒盐噪声、模糊等异常图像。
ImageNet-C	2018	ImageNet-C是一个广泛用于图像分类和鲁棒性评估的数据集，包括被添加噪声、扭曲和干扰的图像。
LSUN Anomaly Detection	2020	LSUN Anomaly Detection是一个用于异常检测的LSUN数据集，包括原始图像和添加异常物体、奇异光照等异常图像。

3. 研究数据



应用场景	数据集名称	数据集描述
	TILDA	包含8种代表性的纹理图像总计3200张，提供正常和缺陷图像，没有提供像素级标注。
工业-布匹	PFID	包含3种花纹织布的图像，每种提供数十张正常和异常图像，提供像素级标注。
	MT	包含5类从不同光照条件下采集到的磁瓦表面缺陷图像，提供像素级标注。同时提供大量正常图像。
工业-金属	RSDD	钢轨表面缺陷数据集，包含两类图像共计195张，并提供像素级标注。
	NEU	包含6类图像，每类有300张热轧带钢表面缺陷图像，以边界框的形式提供标注，也提供像素级标注。
工业-纳米材料	NanoTWICE	包含45张扫描电子显微镜得到的纳米材料图像，其中5张为正常图像，其余40张带有各种缺陷，提供像素级标注。
工业-综合	MVTec AD	综合了工业生产中各种常见材质的图像异常检测数据集，包含5种纹理图像和10种物体图像，提供像素级人工标注。
医学-大脑	BraTS	包含65张低级别和高级别神经胶质瘤的多模态核磁共振图像，并提供精细的人工标注。
医学-视网膜	AMD	包含来自115个正常眼部和269个患者眼部的总计38400张谱域光学相干层析扫描图像(SD-OCT)，并提供精细的像素级标注。
高光谱-混合	AVIRIS	具有高光谱信息的遥感图像数据集，包含了海岸、城市和机场等各种场景的高光谱图像。
	ABU	用于遥感图像异常检测算法的训练和评估，可帮助识别遥感图像中的异常情况。

4. 评价指标



北京交通大学

▲ 常用的评价指标： (包括但不限于以下这些~)

- **Accuracy**准确率
- **Precision**精度
- **Recall**召回率
- **AUC (AUROC)**受试者工作特性曲线下面积 (ROC 曲线下的面积)
- **PRO-Score**区域重叠分数

➤ **ROC**: 模型在不同分类阈值下真阳性率和假阳性率的变化曲线

➤ **AUC**: **ROC 曲线下的面积**, 用来评估分类器对于预测结果的准确程度。

在ROC 曲线上, 横轴表示假阳性率 (False Positive Rate), 即预测为正类但实际为负类的样本数量与所有负类样本数量之比; 纵轴表示真阳性率 (True Positive Rate), 即预测为正类且实际为正类的样本数量与所有正类样本数量之比。

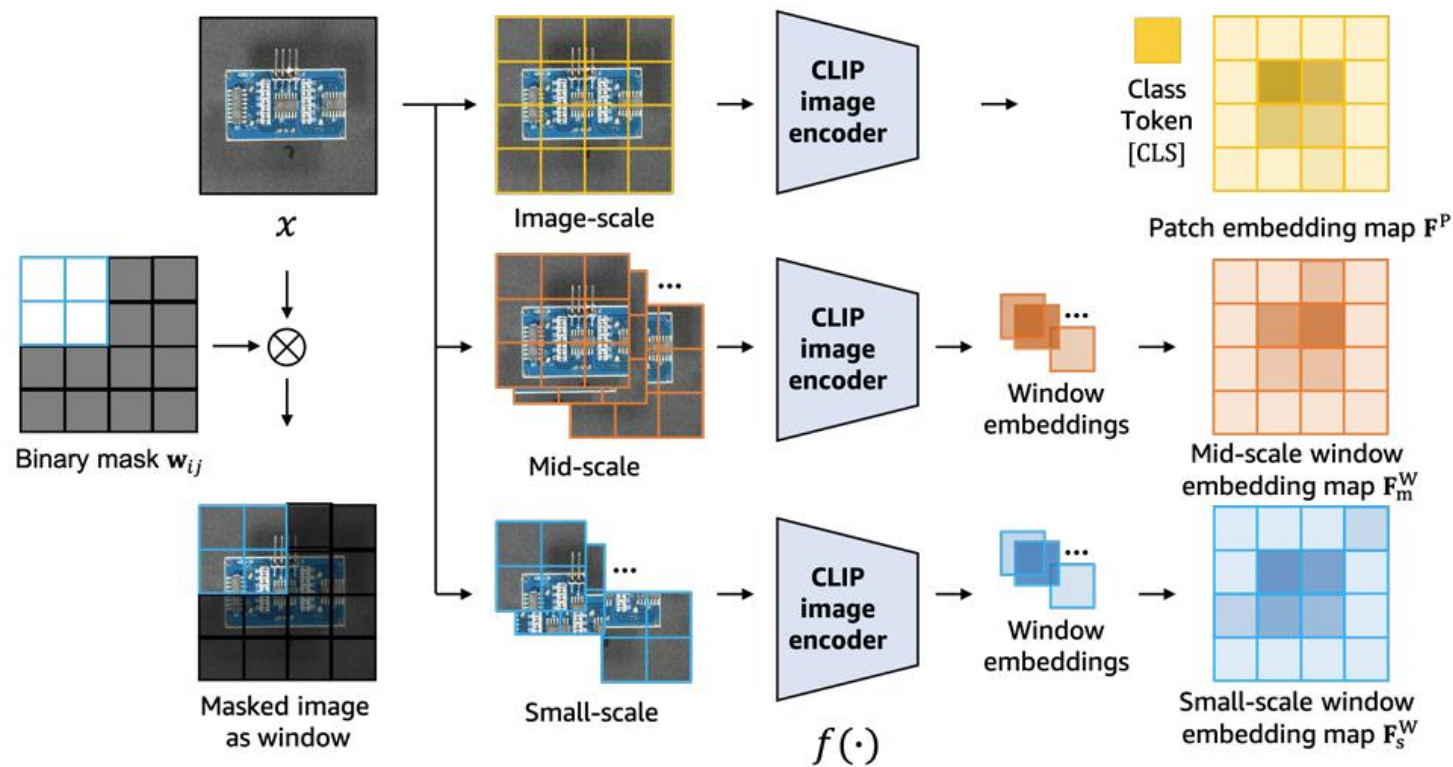
AUC 的取值范围在 **0.5 到 1 之间**, 值越大表示分类器的性能越好。

➤ **PRO-Score**: 在一系列阈值下构建性能曲线, 并以曲线下面积作为综合评估指标。

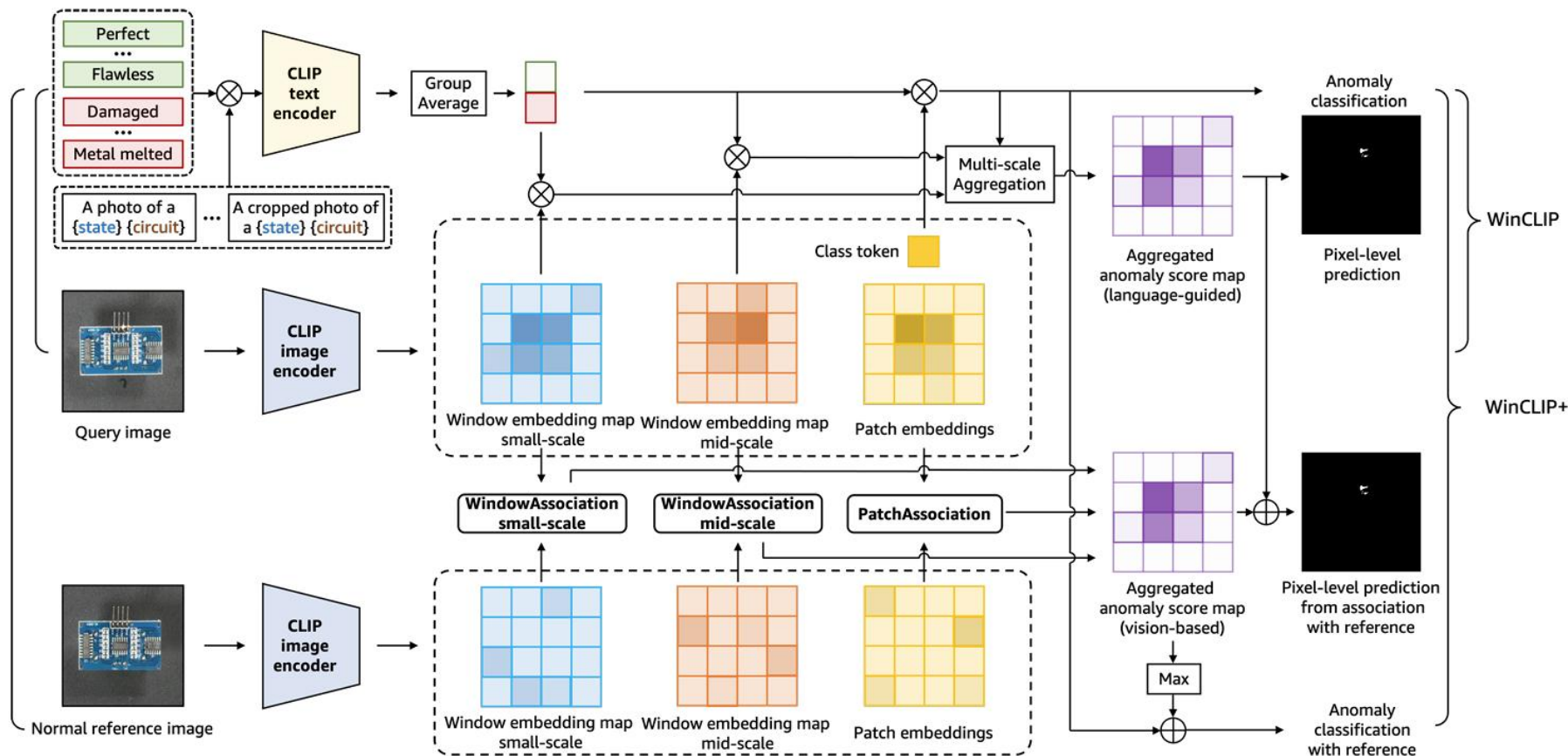
具体统计的是不同阈值下的**区域重叠率** (Per-region-overlap, PRO),

PRO 是以二值化后连通域和真值图之间的相对重叠率作为每一个阈值下的模型分类性能。

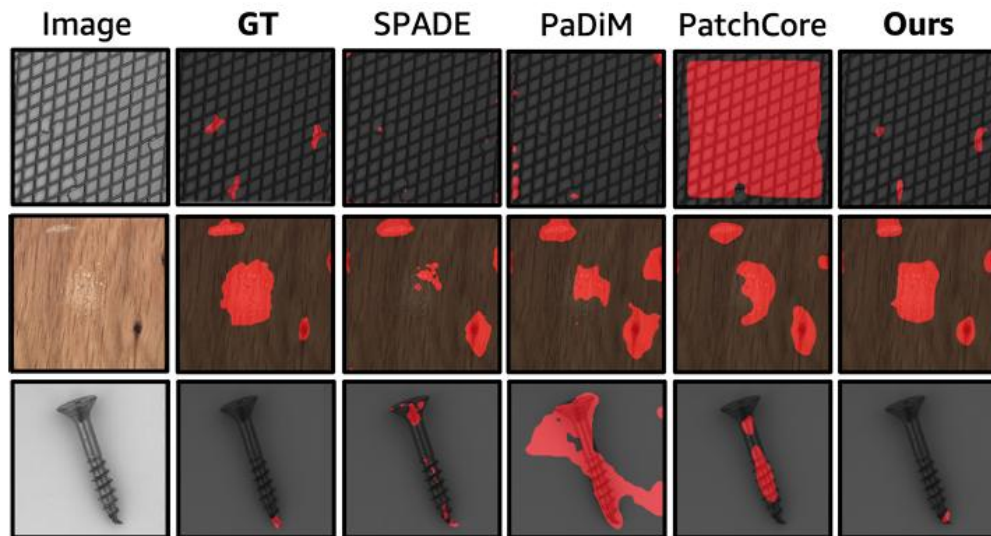
区域级性能指标。



- 动机1：数据与范式局限。
缺陷罕见，缺乏异常样本，现有无监督方法依赖大量正常数据，在小样本场景下性能受限。
- 动机2：任务扩展性挑战。
场景定制化模型难以扩展，转向零/小样本检测。但缺陷需要正常样本作参照，且无分割监督。
- 动机3：视觉-语言模型的机遇。
CLIP等视觉-语言模型通过大规模图文预训练，能够实现无需微调的零/小样本迁移。

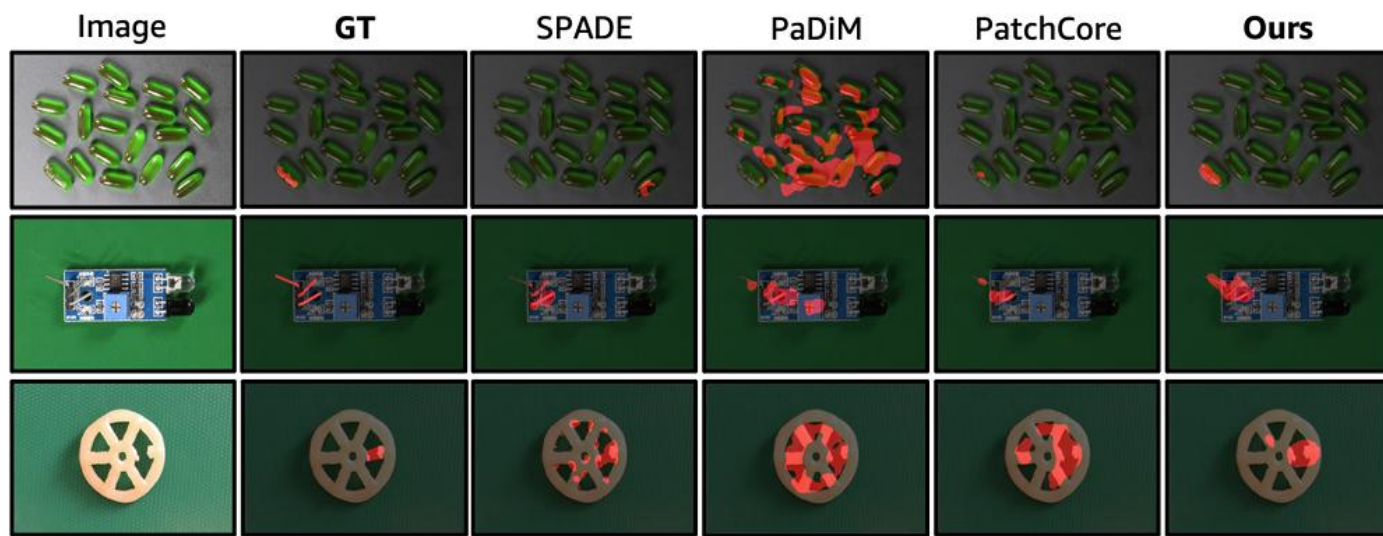


- 创新1：引入了一种组合提示集合，改进了基于CLIP的零样本分类，改进了零样本设置下的异常分类。
- 创新2：使用预训练的CLIP模型，提出WinCLIP架构，有效地提取和聚集与语言一致的多尺度空间特征，用于零样本异常分割。是第一个探索文本引导的零样本异常分类和分割的工作。
- 创新3：提出了一种简单的参考关联方法，将其应用于基于图像的多尺度特征图的小样本异常分割
WinCLIP+结合了文本引导和纯视觉方案，用于少量正常样本的异常识别



(a) MVTec-AD (1-shot)

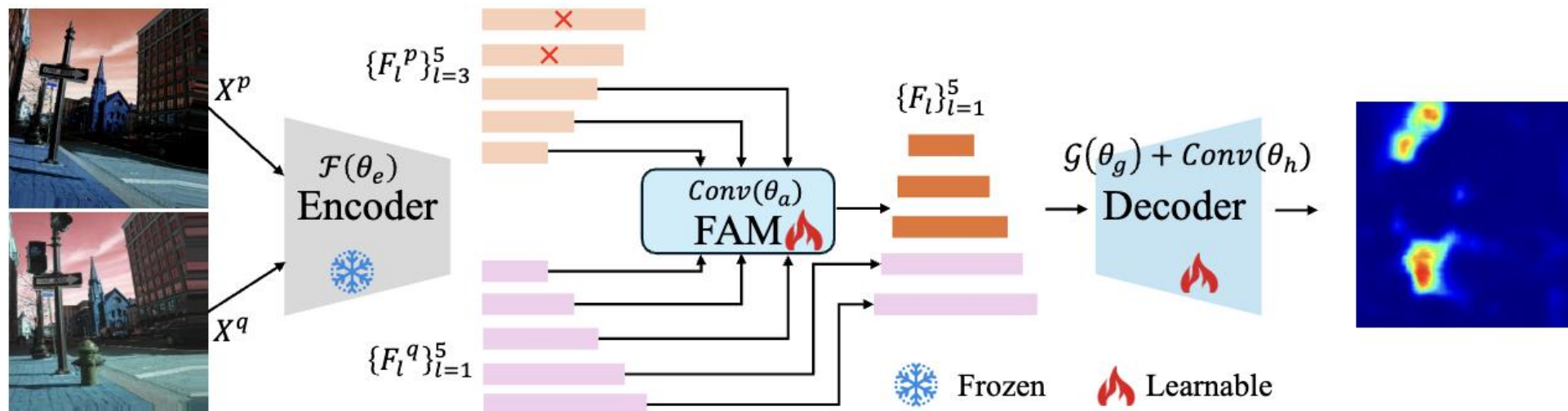
Anomaly Classification		MVTec-AD			VisA		
Setup	Method	AUROC	AUPR	F_1 -max	AUROC	AUPR	F_1 -max
0-shot	CLIP-AC [27]	74.0±0.0	89.1±0.0	88.5±0.0	59.3±0.0	67.0±0.0	74.4±0.0
	+ Prompt ens. [27]	74.1±0.0	89.5±0.0	87.8±0.0	58.2±0.0	66.4±0.0	74.0±0.0
	WinCLIP (ours)	91.8±0.0	96.5±0.0	92.9±0.0	78.1±0.0	81.2±0.0	79.0±0.0
1-shot	SPADE [7]	81.0±2.0	90.6±0.8	90.3±0.8	79.5±4.0	82.0±3.3	80.7±1.9
	PaDiM [8]	76.6±3.1	88.1±1.7	88.2±1.1	62.8±5.4	68.3±4.0	75.3±1.2
	PatchCore [31]	83.4±3.0	92.2±1.5	90.5±1.5	79.9±2.9	82.8±2.3	81.7±1.6
	WinCLIP+ (ours)	93.1±2.0	96.5±0.9	93.7±1.1	83.8±4.0	85.1±4.0	83.1±1.7
2-shot	SPADE [7]	82.9±2.6	91.7±1.2	91.1±1.0	80.7±5.0	82.3±4.3	81.7±2.5
	PaDiM [8]	78.9±3.1	89.3±1.7	89.2±1.1	67.4±5.1	71.6±3.8	75.7±1.8
	PatchCore [31]	86.3±3.3	93.8±1.7	92.0±1.5	81.6±4.0	84.8±3.2	82.5±1.8
	WinCLIP+ (ours)	94.4±1.3	97.0±0.7	94.4±0.8	84.6±2.4	85.8±2.7	83.0±1.4
4-shot	SPADE [7]	84.8±2.5	92.5±1.2	91.5±0.9	81.7±3.4	83.4±2.7	82.1±2.1
	PaDiM [8]	80.4±2.5	90.5±1.6	90.2±1.2	72.8±2.9	75.6±2.2	78.0±1.2
	PatchCore [31]	88.8±2.6	94.5±1.5	92.6±1.6	85.3±2.1	87.5±2.1	84.3±1.3
	WinCLIP+ (ours)	95.2±1.3	97.3±0.6	94.7±0.8	87.3±1.8	88.8±1.8	84.2±1.6



(b) VisA (1-shot)

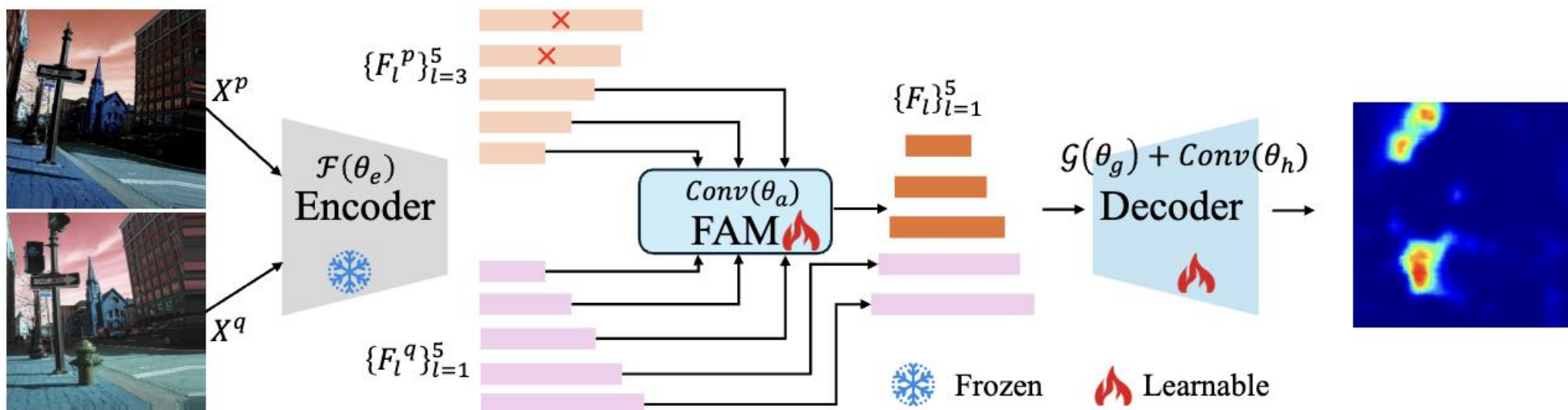
Anomaly Segmentation		MVTec-AD			VisA		
Setup	Method	pAUROC	PRO	F_1 -max	pAUROC	PRO	F_1 -max
0-shot	Trans-MM [5]	57.5±0.0	21.9±0.0	12.1±0.0	49.4±0.0	10.2±0.0	3.1±0.0
	MaskCLIP [57]	63.7±0.0	40.5±0.0	18.5±0.0	60.9±0.0	27.3±0.0	7.3±0.0
	WinCLIP (ours)	85.1±0.0	64.6±0.0	31.7±0.0	79.6±0.0	56.8±0.0	14.8±0.0
1-shot	SPADE [7]	91.2±0.4	83.9±0.7	42.4±1.0	95.6±0.4	84.1±1.6	35.5±2.2
	PaDiM [8]	89.3±0.9	73.3±2.0	40.2±2.1	89.9±0.8	64.3±2.4	17.4±1.7
	PatchCore [31]	92.0±1.0	79.7±2.0	50.4±2.1	95.4±0.6	80.5±2.5	38.0±1.9
	WinCLIP+ (ours)	95.2±0.5	87.1±1.2	55.9±2.7	96.4±0.4	85.1±2.1	41.3±2.3
2-shot	SPADE [7]	92.0±0.3	85.7±0.7	44.5±1.0	96.2±0.4	85.7±1.1	40.5±3.7
	PaDiM [8]	91.3±0.7	78.2±1.8	43.7±1.5	92.0±0.7	70.1±2.6	21.1±2.4
	PatchCore [31]	93.3±0.6	82.3±1.3	53.0±1.7	96.1±0.5	82.6±2.3	41.0±3.9
	WinCLIP+ (ours)	96.0±0.3	88.4±0.9	58.4±1.7	96.8±0.3	86.2±1.4	43.5±3.3
4-shot	SPADE [7]	92.7±0.3	87.0±0.5	46.2±1.3	96.6±0.3	87.3±0.8	43.6±3.6
	PaDiM [8]	92.6±0.7	81.3±1.9	46.1±1.8	93.2±0.5	72.6±1.9	24.6±1.8
	PatchCore [31]	94.3±0.5	84.3±1.6	55.0±1.9	96.8±0.3	84.9±1.4	43.9±3.1
	WinCLIP+ (ours)	96.2±0.3	89.0±0.8	59.5±1.8	97.2±0.2	87.6±0.9	47.0±3.0

----NeurIPS2024 《MetaUAS: Universal Anomaly Segmentation with One-Prompt Meta-Learning》



纯视觉模型引领异常检测新范式

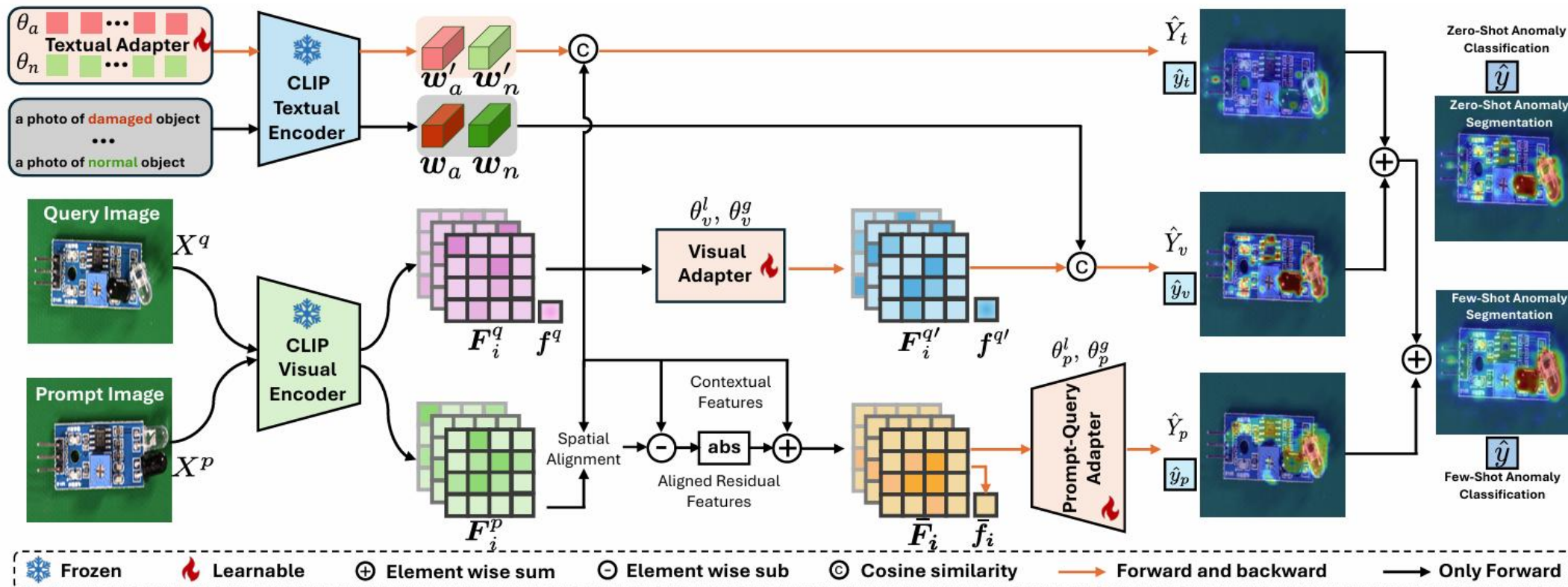
- 动机1：视觉-语言模型 (VLMs) 在异常检测中的固有局限性。
文本描述难以覆盖多样的视觉异常，提示设计耗时且泛化有限。
- 动机2：缺乏大规模、多样化的异常分割训练数据。
像素级异常标注数据少，定义开放，制约监督学习泛化能力。
- 动机3：探索纯视觉模型在通用异常分割中的潜力。
探索不依赖文本、仅凭视觉信息实现通用异常分割的新路径。
- 动机4：提升对几何变化的鲁棒性。
应对实际应用中视角、尺度等变化，提升特征对齐与语义感知能力。



- **创新1：将异常检测统一为变化检测的新范式。**
核心是利用正常参考图像识别差异，而非直接定义异常，极大提升泛化能力。
- **创新2：利用大规模合成图像对进行训练。**
使用多样化的合成图像对进行训练，数据生成独立于特定异常，有效解决标注数据稀缺问题。
- **创新3：提出单提示元学习框架 (MetaUAS)。**
设计单提示元学习框架，使模型学会如何从单个样本中学习变化，快速适应新场景中的异常。
- **创新4：引入软特征对齐模块。**
通过软特征对齐模块处理图像间的几何不对齐问题，精准区分真实语义变化与几何扰动。
- **创新5：高效且无需额外训练的推理过程。**
高效零样本推理，仅需一张正常提示图像即可完成分割，无需针对下游任务进行额外微调。

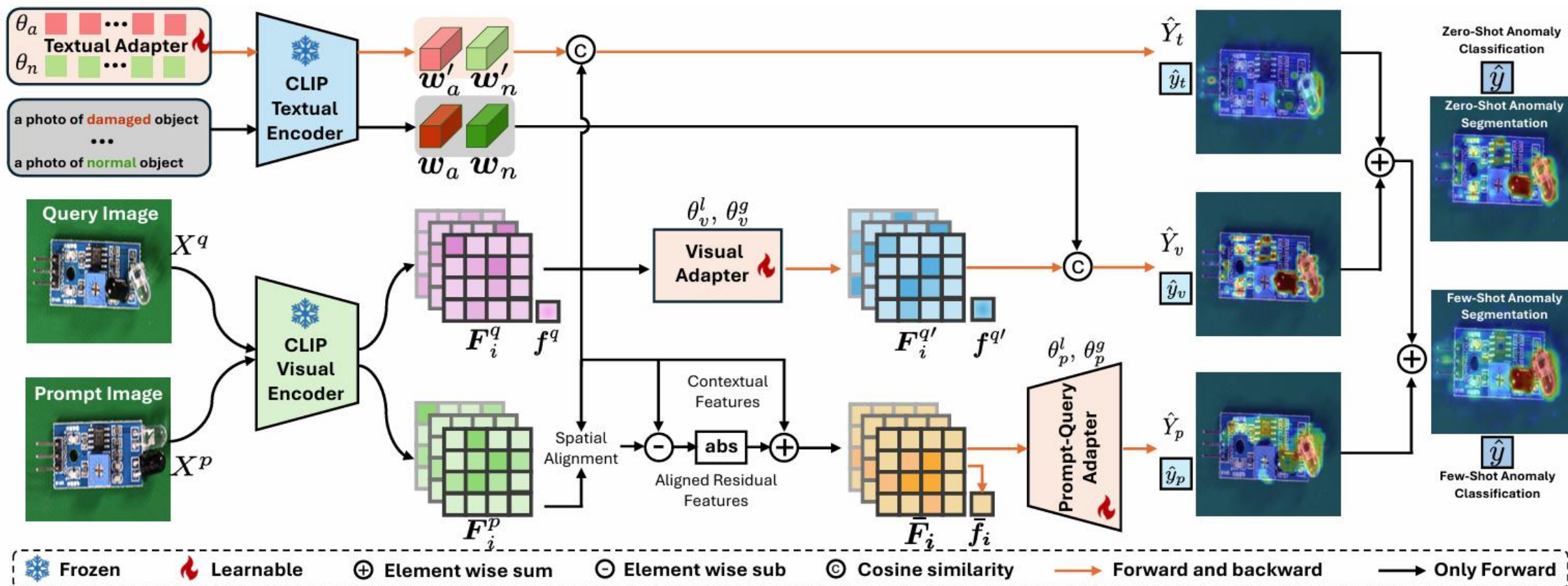
▲ AdaptCLIP

----AAAI2025 《AdaptCLIP: Adapting CLIP for Universal Visual Anomaly Detection》



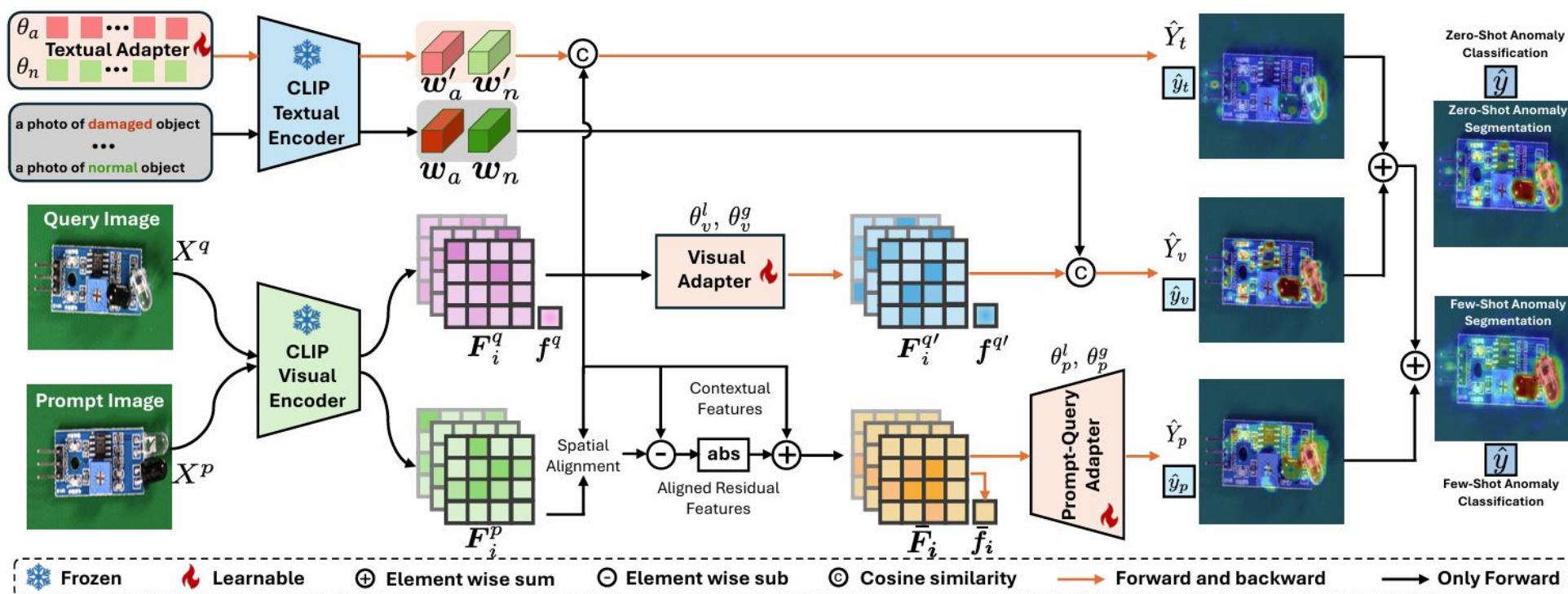
- **动机1：开放场景下的通用视觉异常检测。**
面对新的或未见的视觉领域，能够识别出异常图像和像素，无需对具体领域进行额外微调。
- **动机2：现有方法的提示模板与标记交互的局限性。**
基于视觉-语言模型（如CLIP）的预训练方法虽然展现出强大的泛化能力，但存在设计提示模板复杂、复杂的标记交互以及需要额外微调等问题，导致灵活性受限。

▲ AdaptCLIP



- 框架设计: AdaptCLIP通过在CLIP模型的输入或输出端添加三个简单的适配器来实现:
- 1' 视觉适配器: 适应视觉标记, 与固定的文本嵌入对齐。
 - 2' 文本适配器: 学习文本提示嵌入, 与固定的视觉标记对齐。
 - 3' 提示-查询适配器: 基于查询图像和正常图像提示的联合分布, 学习比较能力, 以用于少样本异常检测。

▲ AdaptCLIP



➤ 创新1：多模态交替学习。

自适应的视觉和文本表示应该交替学习，而不是联合学习。

充分利用CLIP模型的强大视觉-文本表示，避免过拟合，提高跨领域的泛化能力。

➤ 创新2：上下文信息比较学习。

在比较查询图像和正常图像提示时，应该结合上下文特征和对齐后的残差特征，而不仅仅是依赖于残差特征。这有助于更好地识别异常区域，同时保留上下文信息。

高效零样本推理，仅需一张正常提示图像即可完成分割，无需针对下游任务进行额外微调。

5. 应用领域

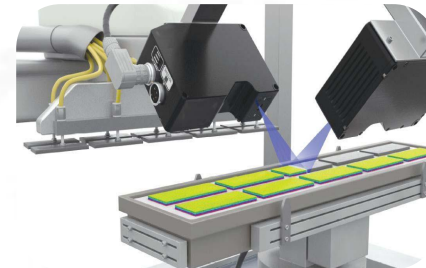
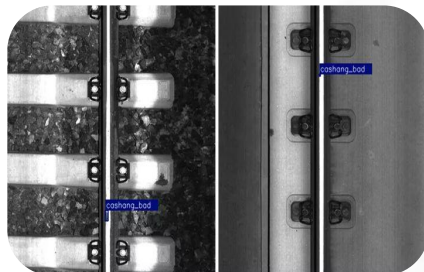
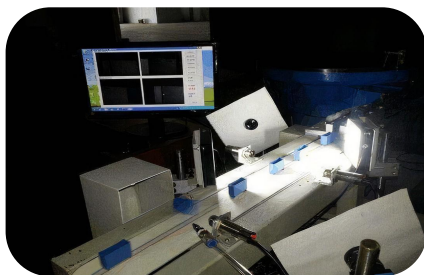
国防军事：异常敌情监测；无人侦察.....



北京交通大学

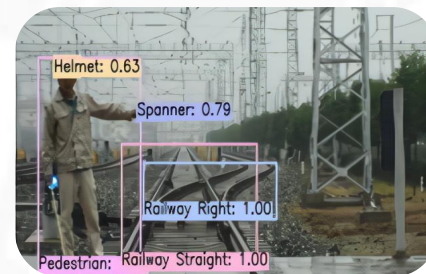
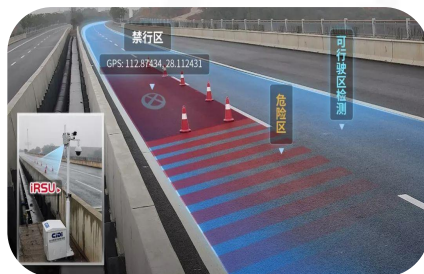
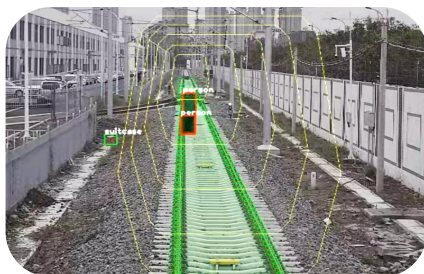
应用领域	具体对象
交通领域	轨道列车、高速公路、机场飞机、船舶
工业领域	布匹、玻璃、钢板、水泥、机械设备
航空航天领域	遥感图像、飞行器关键部件
军事领域	无人战车、军用无人机
医学领域	核磁共振图、虹膜图、视网膜图
社会领域	城市安检、家电通讯、街道监控

➤ **工业制造：** **产品外观缺陷检测**；生产线异物监测..... **智能安防：** 智能安检；入侵检测；视频监控预警；异常追踪等



➤ **智慧交通：** 自动驾驶（障碍物识别；**疲劳检测**；路况异常预警.....）

医学诊断： 病灶检测；辅助手术；病理分析；影像处理.....



6. 研究瓶颈



北京交通大学

当前研究瓶颈

异常样本的未知性

- 实际问题中异常样本未知，**基于监督学习的算法应用性差**
- 仅用正常样本训练的模型存在对异常样本的**漏检**风险

异常定义的不清晰性

- **异常程度定义困难**：
定义苛刻——>因噪声导致正常被误检为异常
定义宽松——>微弱异常被误检为正常

微弱异常的定位

- 面积小、亮度弱
- **与周围区域差异小**

高维数据特征问题

- 图像属于高维数据，**特征向量复杂**
- 常规检测算法难以直接应用到图像数据

当前算法的通用性

- 图像类型庞大，异常目标多样
- 当前检测算法**通用性差**，存在局限性

7. 未来展望



北京交通大学

➤ 设计高效的异常检测算法：

通过研究适当的算法和**优化模型的计算性能**，以**兼顾检测精度和实时性**。

例如，基于深度学习的方法通过采用端到端的训练方式来实现高效的异常检测。

此外，一些基于特征提取和降维的方法也可以用于快速的异常检测。

➤ 提高异常定位的精度：

可以探索**引入上下文信息、多尺度分析、异常建模与生成等方法**，以提高异常定位的精确度；

例如：利用图像的全局上下文信息，建模异常纹理的特征，辅助实现异常区域的准确定位。

➤ 结合小样本学习与弱监督学习：

真实场景下存在少量的缺陷样本可以用来训练模型，而**弱监督学习**则适用于严格无监督的环境；

通过**标注少量正常和异常样本**来增强对潜在异常样本的检测能力。

➤ 设计与应用轻量化模型：

设计轻量化的网络结构，以**减少计算量**并降低硬件设备的要求和成本。

轻量化网络设计可以包括设计高效的网络计算方法、**知识蒸馏**以及**模型压缩**等技术。