arXiv:2601.02016v1 [cs.CV] 5 Jan 2026

利用特权信息增强目标检测：一种模型无关的教师–学生方法

Matthias Bartolo，Dylan Seychell，*IEEE*高级会员，Gabriel Hili，Matthew Montebello，*IEEE*高级会员，Carl JamesDebono，*IEEE*高级会员，Saviour Formosa，以及 Konstantinos Makantasis，*IEEE*会员

摘要—本文研究了将利用特权信息学习（LUPI）范式应用于目标检测，以利用训练期间可用但推理阶段不可用的细粒度描述性信息。我们提出了一种通用的模型无关方法，通过教师–学生架构将特权信息——例如边界框掩码、显著性图和深度线索——注入基于深度学习的目标检测器中。我们在五种最先进的目标检测模型和多个公开基准数据集上开展实验，包括基于无人机的垃圾检测数据集和 Pascal VOC 2012，以评估其对准确率、泛化能力和计算效率的影响。结果表明，LUPI训练的学生模型始终优于其基线模型，在不增加推理复杂度或模型大小的情况下，显著提升了检测准确率。对于中大型目标，性能提升尤为明显；而消融研究表明，对教师指导采用中等权重，能够在利用特权输入和标准输入进行学习之间实现最佳平衡。这些发现证实，LUPI 框架为推进资源受限环境和现实场景中的目标检测系统提供了一种有效且实用的策略。

索引词—计算机视觉、知识蒸馏、利用特权信息学习、垃圾检测、目标检测

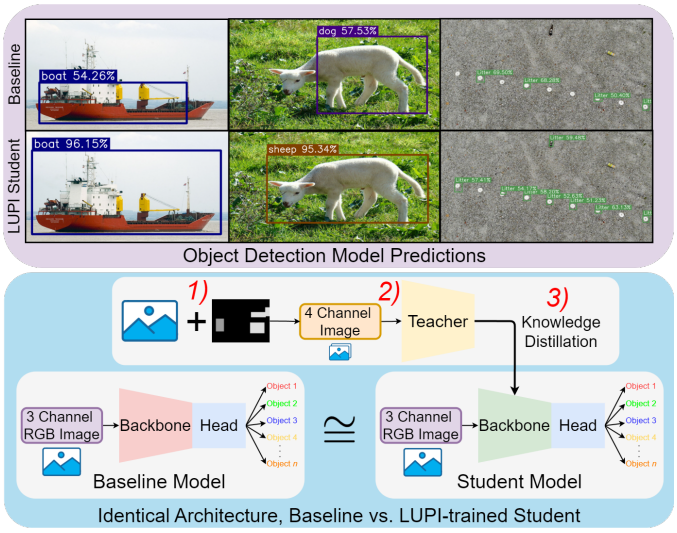

图1. 基线目标检测预测与经LUPI训练的学生模型预测的视觉对比，展示了在保持相同架构的同时准确率有所提升。该图还展示了LUPI训练流程，包括特权信息、教师模型和知识蒸馏；这些提升完全源自学习过程的强化。

I 引言

ADVANCEMENTS 计算硬件的进步，尤其是图形处理器的发展，推动了人工智能和自动化技术的迅速普及。在这一背景下，目标检测已成为一项核心问题，推动了自主系统、环境监测和机器人学等领域的应用。在过去十年中，YOLO [1]，Faster R-CNN [2],和 RetinaNet [3]等模型实现了快速且准确的检测能力，使目标检测成为一种可广泛部署的技术。尽管取得了这些进展，要持续实现较高的检测准确率仍然是一项挑战。许多最先进的检测器依赖日益复杂的架构 [4], [5],，仍需使用大型标注数据集针对特定领域应用场景进行微调 [6]，，而这两方面都会带来显著的实际限制。深度模型通常需要大量训练时间和计算资源，而大规模数据集则需要耗费大量资金和人力进行标注，才能提高检测准确率 [6], [7]。

然而，标注图像包含极其丰富的信息，而当前最先进的目标检测模型尚未充分利用这些信息。在本研究中，我们检验这样一个假设：可以在训练期间自动构建并利用描述性极强的细粒度信息，以提升目标检测器的性能。基于我们的初步结果 [8],，我们采用特权信息学习（LUPI）范式 [9], [10], [11]，并针对目标检测对其组成部分进行调整，使其能够得到有效应用。

LUPI范式旨在解决训练与测试之间存在信息不对称的问题：训练期间可以获取补充信息，但推理时无法获取。通过利用测试时无法访问的高信息量数据流，LUPI 能够显著降低对大型标注数据集的需求，同时不牺牲模型准确率。特权信息可以有多种形式，

（通讯作者：*Matthias Bartolo*。）Matthias Bartolo、Dylan Seychell、Gabriel Hili、Matthew Montebello 和 Konstantinos Makantasis 任职于马耳他大学信息与通信技术学院人工智能系，地址为马耳他 MSD 2080（电子邮箱：matthias.bartolo@um.edu.mt；dylan.seychell@um.edu.mt；gabriel.hili@um.edu.mt；matthew.montebello@um.edu.mt；konstantinos.makantasis@um.edu.mt）。Carl James Debono 任职于马耳他大学信息与通信技术学院通信与计算机工程系，地址为马耳他 MSD 2080（电子邮箱：carl.debono@um.edu.mt）。Saviour Formosa 任职于马耳他大学社会福祉学院犯罪学系，地址为马耳他 MSD 2080（电子邮箱：saviour.formosa@um.edu.mt）。

本研究是“航空废弃物识别与地理定位系统”（*AWIGS*）项目的一部分，该项目由 Xjenza Malta 通过 FUSION：研发与创新技术开发计划 Lite 资助。

包括深度线索、显著性图、高分辨率图像或特定领域标注 [12], [13], [14]。通过纳入此类信号，模型能够在训练期间学习到更丰富的特征表示，从而提升泛化能力并加快收敛，同时保持推理要求不变（见图1）。

我们的工作在多个方面具有创新性。首先，我们提出并开发了一种通用方法，可将特权信息注入任何基于深度学习的目标检测器。该方法具有模型无关性，不受架构选择的限制。其次，我们使用多个基于无人机的垃圾检测数据集和 Pascal VOC 基准，研究了该方法在五种开源先进预训练目标检测模型上的影响。第三，我们在早期工作 [8]的基础上进行了大幅扩展，分析了不同目标尺度、标准 COCO 指标以及不同形式的特权信息（包括深度、显著性及其组合）下的性能，同时考察了推理时间和模型大小等实际因素。最后，通过大量实验验证，我们证明了特权信息对于提升模型性能的重要性，并深入探讨了基于 LUPI 的方法在通用目标检测中的可行性，同时评估了这一范式在科学和实践层面的意义。

II. 相关工作

目标检测是一项复杂的任务，涉及分类和定位两方面 [7]。该领域历史悠久，从早期采用传统特征匹配 [15], [16]和机器学习技术 [17], [18] 的研究，逐步发展到融入深度学习方法 [19],；如今，深度学习方法已达到最先进的性能。本节回顾了基于深度学习的目标检测以及 LUPI 在计算机视觉应用中的相关研究。

A. 用于目标检测的深度学习

目标检测是一项监督学习任务，包含若干关键挑战，包括在复杂背景和干扰因素下检测物体、应对尺度变化、处理遮挡、缓解类别不平衡和数据集偏差，以及检测小目标 [6], [20]。应对所有这些挑战都需要稳健的学习算法。目前最先进的目标检测模型利用各种深度学习架构，输出边界框及其对应的类别标签。这些网络通常分为四类：单阶段、两阶段、基于 Transformer 的方法以及其他深度学习方法。

单阶段检测器使用单个网络解决定位和分类问题。常见示例包括 YOLO（You Only Look Once） [1]和 SSD（Single Shot MultiBox Detector） [21]。两阶段检测器使用独立的网络执行定位和分类，示例包括 Faster R-CNN [2] 和 Mask R-CNN [22]。基于 Transformer 的检测器（如 DETR [4] 和 RT-DETR[5],）利用自注意力机制进行目标检测。其他深度学习方法（如 CenterNet [23],或 SAHI[24]）融入了强化学习技术。

这些多样化的方法凸显了为应对目标检测挑战而发展出的各种策略。单阶段检测器优先考虑速度和效率，而两阶段检测器则在复杂场景中具有更高的准确率。基于 Transformer 的模型在对目标关系进行建模方面展现出灵活性，说明不同的架构范式能够应对检测问题的不同方面。

B. 计算机视觉中的特权信息学习

在计算机视觉中使用 LUPI 仍相对缺乏研究，尤其是在目标检测领域。我们早期的工作 [8] 代表了该领域最早的研究成果之一。文献中的早期工作主要聚焦于目标定位任务。Feyereisl 等人（2014 年） [25]使用分割掩码和 SURF 特征作为特权信息，并在加州理工学院-加州大学圣迭戈分校鸟类数据集上，利用结构化支持向量机+ 算法进行目标定位 [26]。改进幅度有限，而且深度学习模型在当时尚未得到广泛采用。同样，Sun 等人（2018 年） [27] 研究了在同一数据集上利用特权信息进行目标定位的方法，但取得的改进也十分有限。

LUPI 在图像分类领域得到了更广泛的探索。Sharmanska 等人 [28], [29] 研究了语义属性、边界框、文本描述和标注者推理依据作为特权信息的作用，结果表明，在 SVM+ 框架中，这些信息带来了可量化的改进。Wang 等人 [13] 将高分辨率图像和标签等特权数据纳入多标签分类，证明利用此类信息能够提升性能。Makantasis 等人[30], [31] 将音频和生理特征用作特权信息，以开发基于视觉的情感模型，试图弥合体外与体内情感建模任务之间的差距。

LUPI 与知识蒸馏 [32],密切相关；知识蒸馏是一种近年来日益流行的技术，尤其是在计算机视觉领域 [33]。在这一框架中，高容量网络（教师）将信息传递给较小的网络（学生），使学生能够学习到更丰富、更有信息量的表征 [34]。Hinton 提出的广义蒸馏 [35] 对这一过程进行了形式化。在计算机视觉应用中，常采用学生网络与教师网络之间的特征匹配和逻辑值匹配等方法；而定位蒸馏则专门针对空间信息丰富的区域，帮助学生模型聚焦于最相关的区域，从而提升检测性能 [33], [34], [35]。

然而，知识蒸馏与 LUPI 在目标和信息需求方面存在根本差异。知识蒸馏的主要目标是构建一个紧凑的学生模型，使其性能与规模大得多的教师模型相当，而两个模型使用相同的输入信息。相比之下，LUPI [9], [10], [11]的目标并非压缩大型网络，而是将使用信息量丰富的特权信息进行训练的教师模型中的知识迁移给学生模型，使学生模型在缺少这些特权信息的情况下进行预测。因此，知识蒸馏解决的是模型

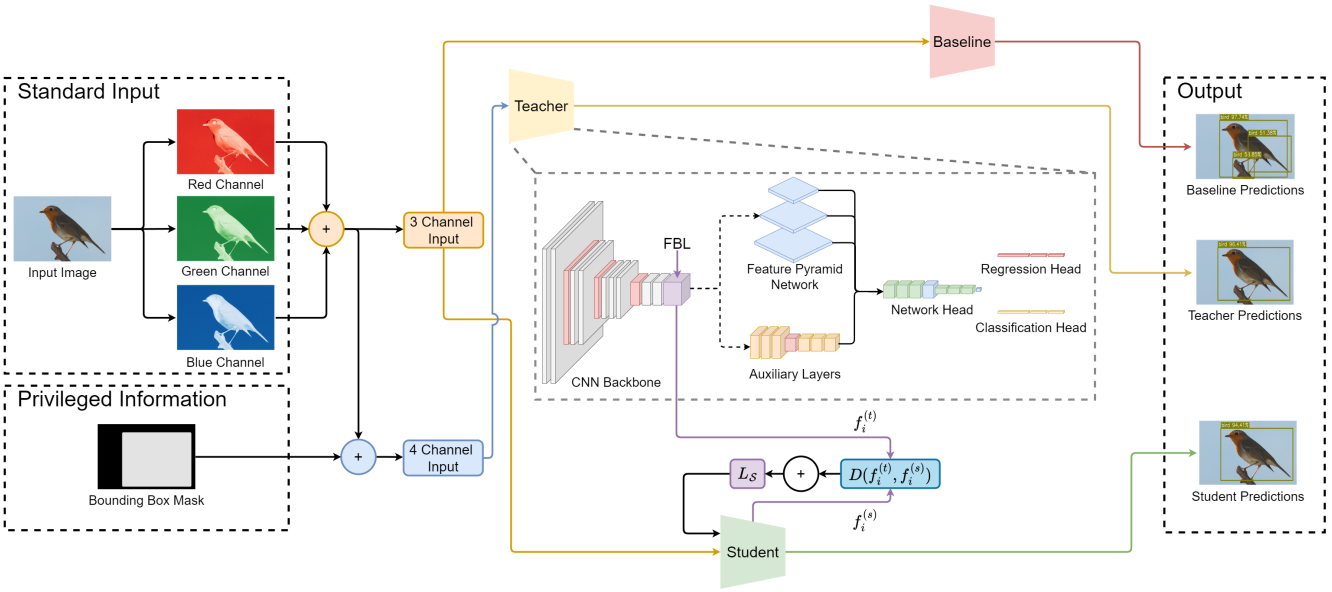

图2. 训练设置的详细架构。教师网络同时接收 RGB图像和特权输入通道，从而生成更丰富的中间表示。学生网络仅处理 RGB图像，但通过教师的知识蒸馏获得额外监督。文中还纳入了一个仅使用 RGB 的基线模型以供比较。与基线模型相比，学生模型的预测更加精细。

压缩问题，而 LUPI 解决的是训练与推理之间的信息不对称问题。

III. 方法

为了检验我们的假设——高度描述性、细粒度的信息可以在训练过程中自动构建并加以利用，从而提升目标检测器的性能——我们遵循 [11] 中的工作，并采用教师-学生框架。本节对问题进行形式化定义，并介绍所提出的方法。

A. 问题形式化

我们考虑一种有监督的目标检测场景，其中每个训练样本由一张标准输入图像$x \in X$、仅在训练期间可用的额外特权信息 $x^* \in X^*$ 以及相应的真实标签$y \in Y$组成；该标签针对图中每个物体包含一个边界框 b 和一个类别标签 l。因此，训练数据集是由三元组构成的集合

$$\mathcal{D}_{\text{train}} = \{(x_i, x_i^*, y_i)\}_{i=1}^{N}, \tag{1}$$

其中，图像 i的真实标签 y_i 为集合

$$y_i = \{(b_j, l_j)\}_{j=1}^{M}. \tag{2}$$

在（2）中， M 表示图像中所描绘物体的数量。我们的目标是估计函数

$$f_w : X \to Y \tag{3}$$

由 w 参数化，使得

$$w := w(X, X^*, Y). \tag{4}$$

在我们的案例中，函数 f_w 由神经网络实现，而参数 w 对应于该网络的权重。公式（3）和（4）表明，该网络仅使用 X进行预测，而其参数的估计不仅使用 X 和 Y，还使用了额外的特权信息 X^*。

B. 提出的方法

我们采用教师–学生范式来利用特权信息。教师网络 f_{teacher}： $X \cup X^* \to Y$ 可以同时访问标准输入和特权输入，从而学习更丰富、信息量更大的中间表示。相比之下，学生网络 f_{student} ： $X \to Y$ 只能观察标准输入，无法直接访问特权信息。然而，在训练期间，我们会促使学生在中间层 l复现教师的潜在表示，从而间接受益于额外的特权上下文。

f_{teacher} 和 f_{student} 均由 L 层神经网络实现：

$$f_{\text{teacher}} := f_1^{(t)} \circ f_2^{(t)} \circ \cdots \circ f_l^{(t)} \circ \cdots \circ f_L^{(t)}, \tag{5}$$

$$f_{\text{student}} := f_1^{(s)} \circ f_2^{(s)} \circ \cdots \circ f_l^{(s)} \circ \cdots \circ f_L^{(s)}. \tag{6}$$

此处，“∘”表示函数复合， $f_{(t)i}$，$f_i^{(s)}$ 分别表示教师网络和学生网络的第 i层。两个网络的第 l层被限制为具有相同数量的隐藏神经元，从而能够直接比较它们的潜在表示。

对于每个三元组 $(x_i, x_i^*, y_i) \in \mathcal{D}_{\text{train}}$，学生经过训练，使其潜在特征 $f_l^{(s)}(x_i)$ 与相应的教师特征 $f_{(t)l}(x_i, x_i^*)$ 对齐。这种对齐构成了知识迁移过程的基础，使学生仅依赖标准输入，也能近似教师更丰富的中间表示。

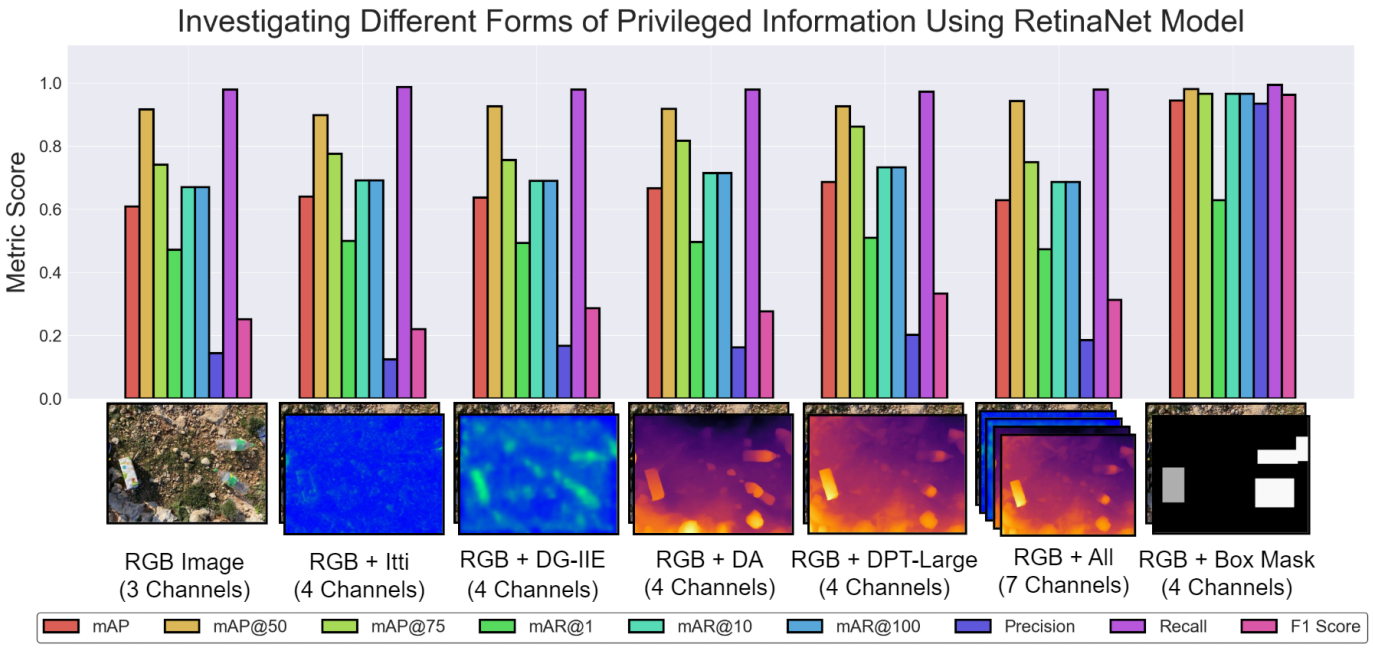

图 3. 使用 RetinaNet 模型在 SODA 1米数据集上研究不同形式的特权信息。比较内容包括显著性、深度、融合和边界框掩码表征。边界框掩码带来了最高的目标检测准确率提升。

学生通过组合损失函数进行优化，在标准检测监督与来自教师的知识迁移之间取得平衡：

$$L_{\mathcal{S}} = (1-\alpha)\cdot L_{\text{det}} + \alpha \cdot D(f_l^{(t)}, f_l^{(s)}), \tag{7}$$

其中，L_{det} 是标准检测损失，$D(\cdot)$用于衡量教师与学生特征向量之间的余弦距离：

$$D(f_l^{(t)}, f_l^{(s)}) = 1 - \frac{f_l^{(t)} \cdot f_l^{(s)}}{\left\|f_l^{(t)}\right\| \left\|f_l^{(s)}\right\|} \tag{8}$$

在（7）中，$\alpha \in [0,1]$ 控制真实标签监督与教师指导之间的相对权重。教师-学生框架的成功与训练中所用特权信息的类型和质量密切相关。接下来，我们将介绍如何构建并整合这些信息，以加强学习过程。

C. 用于目标检测的特权信息

为目标检测选择有效的特权信息形式，需要找到能够切实促进定位和分类的线索。认知研究表明，人类在识别物体时会依赖物理推理，例如估计物体的质心 [36]。这一观察意味着，结构化的空间信号能够提升检测模型的性能。

在深度学习中，先前的研究考察了显著性 [37], [38], [39] 和深度图 [40],[41]，等辅助源，其中显著性与检测性能的相关性更强 [42]。然而，直接将这些信号添加为输入通道，带来的性能提升有限。基于这些发现，本研究系统地评估了多种形式的特权信息在教师–学生训练框架中的作用。研究探索了基于显著性和深度的表征能否提升检测准确率，但它们的贡献仍然有限。在所考察的替代方案中，先前提出的边界框掩膜 [8] 带来了最高的提升。图3 展示了不同形式的特权信息，以及相应的教师模型性能。

掩膜的构造方式通过在单一、结构化的表示中同时嵌入定位信息和类别线索，有效引导网络关注目标区域。每张掩膜图像均以黑色为背景，并用灰度值填充边界框，灰度值与对应类别标签成正比。这种简洁而信息丰富的表示在简洁性、可解释性和整体检测准确率之间实现了最佳平衡。

掩膜由仅在训练期间可用的真实标注生成，因此满足 LUPI 条件。绘制边界框时按大小降序排列，以尽量减少遮挡。虽然也考虑了多边形掩膜和分割掩码 [43], [22], [44] ，但最终优先采用边界框掩码，因为它们更简单、与现有数据集一致，并且在不同实验中表现稳定。

IV. 实现

在概述了所提出的基于 LUPI 的检测框架的理论基础后，下一节将介绍其实际实现和训练设置。

A. LUPI 的教师–学生框架

为了利用上一节引入的特权信息，我们实现了一个教师–学生框架

（见图 2）。教师网络接收 RGB 图像以及额外的特权输入，例如边界框掩膜。为了适配这一额外通道，我们将教师的输入层扩展为四个通道，并使用 Kaiming正态初始化对新增权重进行初始化 [45],其余层则保留预训练的 COCO权重。通过这种调整，教师能够利用更丰富的特征表示，同时不改变整体架构。

学生网络仅处理 RGB图像，但除此之外，其架构与教师网络相同。其训练目标由检测损失和知识迁移项组成（见（7）），该项通过比较学生和教师在最后一个骨干网络层的特征表示计算得出，采用公式（8）。权重参数 α 用于控制直接监督与教师指导之间的平衡。在推理阶段，学生仅使用 RGB 输入，但仍能受益于从教师迁移而来的知识。

B. 目标检测模型与训练方案

基于上一节所述的教师–学生框架，所有模型均采用 torchvision[1] 库中的开源架构实现。完整的训练流程已在 GitHub[2] 上公开。

评估选用了五种目标检测模型：Faster R-CNN [2], SSD [21], RetinaNet [3],SSDLite [46],和 FCOS [47]。这些架构涵盖单阶段和两阶段检测范式，计算复杂度也各不相同。教师网络经过调整，可接收额外的特权输入通道，而学生网络则保留标准 RGB 输入。按照（7），知识迁移使用最后一个骨干网络层的特征进行，该层能够捕获语义丰富的表征 [48]。具体而言，Faster R-CNN、FCOS 和 RetinaNet 使用 FPN 之前的最后一个卷积层；SSD 和 SSDLite 则选用辅助头之前的最后一个卷积层（见图 2）。这一设计确保性能提升可归因于特权信息与知识迁移的整合，而非架构变更。

为确保不同实验之间的一致性，所有模型均采用相同的训练、预处理和后处理流程。模型使用 Adam优化器进行 100 训练轮数，学习率固定为 1×10^{-3}，并根据验证损失进行早停和检查点保存。输入图像（包括特权通道）经最小-最大缩放归一化后，调整至 800×800 像素大小，并按通道标准化为均值为零、方差为一。对最终预测应用 IoU 阈值为 0.5 的非极大值抑制，以去除冗余检测结果。

通过统一架构、训练设置和预处理流程，此实现隔离了特权信息和知识迁移的影响，确保公平且易于解读地评估它们对目标检测性能的贡献。

[1]https://docs.pytorch.org/vision/main/models.html#object-detection [2] https://github.com/mbar0075/lupi-for-object-detection

V. 评估策略

本节介绍了对所提出的基于 LUPI 的目标检测框架进行的实验评估，概述了用于评估该方法鲁棒性的数据集、评估指标和实验流程，并进一步分析了不同模型和条件下的性能。

A. 数据集和评估指标

在定义 LUPI 框架并将特权信息融入教师—学生设置后，评估重点聚焦于基于无人机的垃圾检测 [49], [50]——由于目标较小、背景复杂且场景变化大，这是一项具有挑战性且实用的应用。我们选择了 SODA [49],BDW [51], 和 UAVVaste [52], 等公开数据集，因为它们具有高质量标注且与现实场景密切相关。我们使用 SODA 的子集开展数据集内实验，以在受控环境下分析模型性能；同时，在 BDW 和 UAVVaste 上进行跨数据集评估，以考察模型对未见环境的泛化能力。此外，我们还纳入 Pascal VOC 2012 数据集 [53] ，以评估所提方法在更广泛物体类别中的普遍适用性。

针对每种模型架构，我们将基线仅使用 RGB 的检测器与其对应的 LUPI 教师–学生模型进行比较。我们还单独评估了教师网络，以确认特权信息在训练过程中的作用。此外，本研究还考察了运行时性能，评估学生模型能否在不增加推理成本的情况下提升检测准确率。我们进行了消融实验，以探究损失平衡参数 α 的影响；定性分析则使用 Grad-CAM 可视化结果 [54] 检查模型的注意力。评估指标包括标准目标检测指标——平均精度均值、精确率、召回率、F1分数和平均召回率，以及 COCO 风格指标 [55] ，用于评估不同目标尺度下的检测质量。这一评估策略使我们能够以一致且可控的方式分析特权信息如何提升不同检测模型的准确率、泛化能力和效率。

B. 数据集内与跨数据集实验

数据集内实验聚焦于基于无人机的垃圾检测，使用 SODA 数据集的子集，在受控条件下评估模型性能。实验考察了三种场景：(i) 在 1 米飞行高度下进行不使用切片的二分类垃圾检测，(ii) 在多个飞行高度下进行使用 3×3 切片的二分类检测，以及 (iii) 在多个飞行高度下进行使用 3×3 切片的多标签检测。在每种场景中，选定的五种目标检测架构均分别训练为教师模型和学生模型。按照 [48], 的方法，将参数 α 从 0 到 1 按 0.25 的步长进行调整，以分析教师监督强度的影响。这一实验设计使我们能够系统地评估 LUPI 框架如何提升学生模型的性能。

跨数据集实验评估了在 SODA 上训练的模型应用于其他垃圾数据集时的泛化能力。对于 BDW，直接测试了在 SODA 上以

表 I 使用 COCO指标（精确到小数点后两位）比较所有实验中的教师模型性能。涵盖同数据集、跨数据集以及 Pascal VOC 2012 评估。Faster R-CNN 的平均性能最高，RetinaNet 和 FCOS 的表现相近，而 SSD 和 SSDLite 的结果较低。请注意，SODA 1米子集不包含小目标。在跨数据集评估中，还生成了特权信息，并据此评估教师模型。

模型	数据集	mAP			mAP			mAR			mAR		
		mAP	@50	@75	@小目标	@中目标	@大型	@1	@10	@100	@小型	@中型	@大型
RetinaNet	1 米处的 SODA	0.94	0.98	0.96	–	0.93	0.95	0.63	0.96	0.96	–	0.94	0.97
FCOS		0.96	0.98	0.97	–	0.93	**0.97**	**0.63**	0.97	0.97	–	0.94	**0.98**
Faster R-CNN		**0.96**	**0.99**	**0.98**	–	**0.98**	0.96	0.63	**0.98**	**0.98**	–	**0.99**	0.97
SSD		0.78	0.96	0.94	–	0.78	0.78	0.54	0.81	0.81	–	0.82	0.81
SSDLite		0.61	0.73	0.72	–	0.00	0.73	0.48	0.63	0.63	–	0.00	0.77
RetinaNet	SODA Tiled Binary	0.90	0.95	0.94	0.78	0.98	0.97	0.34	0.83	0.91	0.84	0.98	0.98
FCOS		0.89	0.94	0.93	0.80	0.95	0.97	0.34	0.82	0.90	0.83	0.97	0.98
Faster R-CNN		**0.96**	**0.99**	**0.98**	**0.92**	**0.99**	**0.99**	**0.35**	**0.87**	**0.97**	**0.94**	**0.99**	**0.99**
SSD		0.49	0.62	0.59	0.19	0.77	0.75	0.27	0.51	0.51	0.21	0.80	0.80
SSDLite		0.18	0.23	0.19	0.00	0.05	0.80	0.17	0.19	0.19	0.01	0.07	0.83
RetinaNet	SODA 瓦片式多标签	0.88	0.92	0.91	0.75	0.98	0.98	0.66	0.89	0.89	0.77	0.98	0.99
FCOS		0.91	0.95	0.94	0.83	0.97	0.97	0.68	0.92	0.92	0.85	0.98	0.98
Faster R-CNN		**0.95**	**0.99**	**0.98**	**0.91**	**0.98**	**0.98**	**0.70**	**0.96**	**0.96**	**0.93**	**0.99**	**0.99**
SSD		0.36	0.49	0.45	0.15	0.55	0.55	0.33	0.41	0.41	0.16	0.59	0.62
SSDLite		0.11	0.13	0.13	0.00	0.00	0.46	0.13	0.13	0.13	0.00	0.00	0.54
RetinaNet	BDW	0.46	0.96	0.32	0.00	0.35	0.52	0.38	0.54	0.54	0.00	0.40	**0.59**
FCOS		0.49	0.96	0.43	0.10	0.37	0.55	0.39	0.56	0.56	0.10	0.43	0.61
Faster R-CNN		0.48	**0.97**	0.34	**0.20**	0.40	0.52	0.38	0.54	0.54	**0.20**	0.41	0.59
SSD		**0.55**	0.95	**0.62**	0.00	**0.45**	**0.58**	**0.43**	**0.59**	**0.59**	0.00	**0.48**	0.64
SSDLite		0.23	0.37	0.27	0.00	0.01	0.31	0.21	0.24	0.24	0.00	0.01	0.34
RetinaNet	UAVVaste	0.40	0.78	0.37	0.31	0.72	**0.93**	0.13	0.44	**0.47**	**0.41**	0.74	**0.95**
FCOS		0.42	0.71	**0.47**	0.36	0.74	0.90	0.14	**0.46**	0.46	0.39	0.76	0.90
Faster R-CNN		**0.44**	**0.84**	0.44	**0.37**	**0.75**	0.85	**0.15**	0.46	0.46	0.39	**0.77**	0.85
SSD		0.13	0.24	0.12	0.06	0.45	0.80	0.08	0.15	0.15	0.08	0.50	0.80
SSDLite		0.01	0.03	0.01	0.00	0.06	0.20	0.01	0.01	0.01	0.00	0.07	0.20
RetinaNet	Pascal VOC 2012	0.77	0.86	0.79	0.28	0.63	0.80	0.60	0.81	0.81	0.30	0.67	0.84
FCOS		**0.80**	**0.88**	**0.82**	**0.56**	**0.67**	**0.83**	**0.61**	**0.84**	**0.84**	**0.57**	0.72	**0.86**
Faster R-CNN		0.77	0.91	0.82	0.51	0.66	0.79	0.59	0.82	0.82	0.56	**0.72**	0.85
SSD		0.42	0.56	0.49	0.00	0.06	0.48	0.41	0.48	0.48	0.00	0.07	0.56
SSDLite		0.49	0.61	0.54	0.00	0.00	0.58	0.46	0.55	0.55	0.00	0.00	0.65

1 米飞行高度训练的模型，未进行重新训练；而 UAVVaste 评估则使用了在多个飞行高度下的 3×3SODA 切片图像上训练的模型。所有实验均聚焦于二元检测，从而分析 LUPI 训练的学生模型在多大程度上能够适应未见过的环境和不同的目标分布。这些评估还考察了运行时因素，重点关注在不增加模型大小或推理时间的情况下所取得的性能提升。

C. Pascal VOC 2012 实验

为评估所提出 LUPI 框架的更广泛适用性，研究人员在 Pascal VOC 2012 数据集上开展了实验。该数据集涵盖 20 个不同物体类别的多标签检测任务。此次评估考察了学生模型能否在包含多个物体且尺度各异的复杂场景中有效利用教师指导，从而将分析范围扩展到无人机特定垃圾检测之外。研究人员使用相同的架构和训练方案，对仅使用 RGB 的基线模型、LUPI 学生模型和教师网络进行了比较，确保能够在类别更加多样的数据集上受控地评估泛化能力。

D. 教师–学生平衡消融研究

消融研究考察了平衡参数 α 对学生性能的影响，该参数调节教师监督相对于真实标签的贡献。在 SODA 数据集（二分类和多标签场景）以及 Pascal VOC 2012 上，测试了从 0 到 1、以 0.25 为增量的取值（与[48], 一致），以确定不同任务的最佳权重。这些实验揭示了教师依赖程度如何影响学习动态，并为选择 α 取值提供了指导，使模型在避免过度依赖特权信息的同时实现性能最大化。

VI. 结果与讨论

基于上述评估设置，现将结果汇总呈现，以便直接比较不同实验，涵盖性能、消融实验、可解释性和效率。

A. 教师模型性能

在确定了特权信息的最佳形式（图 3）之后，有必要评估其在不同架构和数据集上的影响。

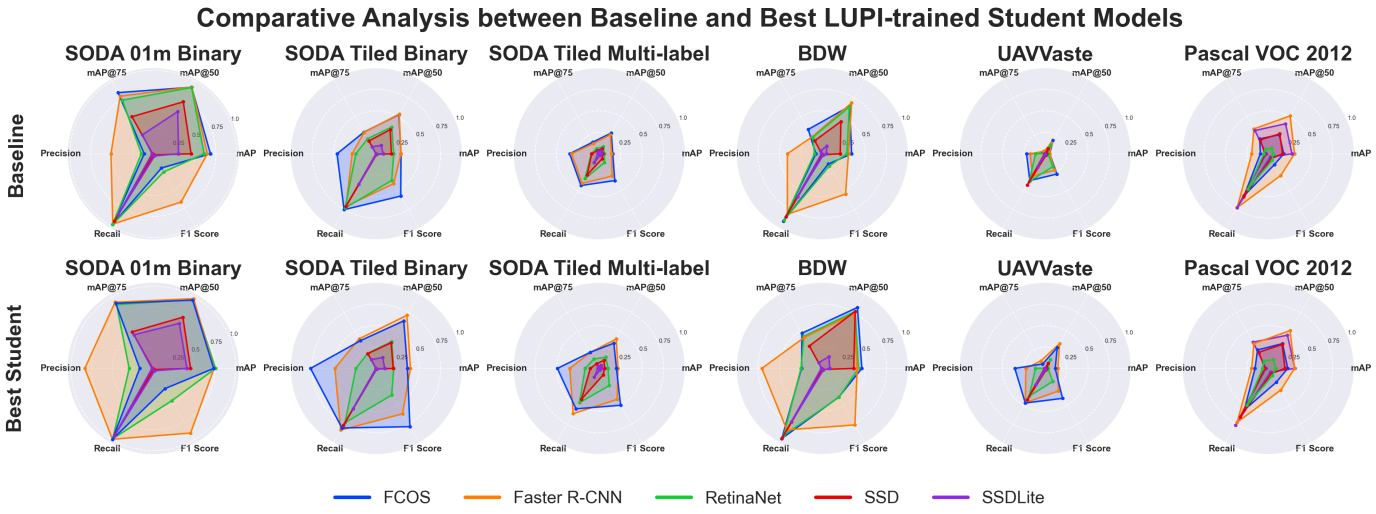

图4. 展示了五种架构在所有数据集上的基线与经过 LUPI 训练的最佳学生模型的比较分析，并以多雷达图呈现。该图突出了严格平均精度均值和F1分数方面的显著提升，其中数据集内评估的增幅最大；而在采用相同架构时，其他数据集也呈现出幅度较小但具有实际意义的改进。

表 I 总结了所有实验中教师模型的表现，包括数据集内、跨数据集和 Pascal VOC 2012 实验，涵盖选定的五种检测架构。结果表明，融入信息丰富的特权输入可显著提升教师模型的准确率，严格 mAP 和 mAR 指标均接近 1，体现出很高的可靠性。尽管小目标上的提升不那么明显，反映出这类目标检测难度较高，但中大型目标的表现始终强劲，说明特权信号有助于模型聚焦于更容易检测的目标。在跨数据集泛化和多标签检测等更具挑战性的条件下，性能虽略有下降，教师模型仍保持了稳健的准确率。总体而言，Faster R-CNN 的平均精度均值最高，RetinaNet 和 FCOS 紧随其后，而 SSD 和 SSDLite 的表现相对较低。

B. 基线与学生模型比较

在评估教师模型后，我们将LUPI训练的学生模型与仅使用RGB的基线模型进行比较，评估其性能。图4展示了所有数据集和检测架构上的比较结果。总体而言，学生模型的表现始终优于基线模型，尤其是在严格的mAP和F1分数评估指标上。最显著的提升出现在同一数据集内的无人机垃圾检测任务中，而跨数据集评估中的提升虽较小，但仍有意义。在基于无人机的场景中，Faster R-CNN、FCOS和RetinaNet从教师指导中获益最多；而在Pascal VOC上，SSD和SSDLite的提升更为明显。尽管在多标签检测和跨数据集泛化能力等更具挑战性的任务中，相对提升有所减弱，但结果证实，LUPI能有效提升不同架构和领域中学生模型的性能。

C. 平衡 α 参数的影响

尽管教师模型表现出较高的准确率，学生模型也始终优于各自的基线模型，但分析学生学习对教师指导依赖程度的变化所产生的影响仍然十分重要。这种依赖程度由参数 α控制，该参数决定知识迁移过程中教师贡献的权重。实验采用了 $\alpha \in \{0, 0.25, 0.5, 0.75, 1\}$，其中 $\alpha = 0$对应基线，$\alpha = 1$ 对应完整的教师监督。图5 使用COCO指标汇总了所有数据集和架构上的结果，红色向下箭头表示严格的mAP@50–95指标下的最佳表现。中间取值 α （0.25和0.5）通常能取得最佳表现，在利用真实标签和教师知识进行学习之间实现平衡；而 $\alpha = 0.75$ 有时表现良好，$\alpha = 1$ 则往往表现不佳。这些趋势在不同数据集和模型上均一致。

较小目标仍然带来挑战，改进幅度有限；相比之下，中大型目标受益更为明显，这反映了从教师模型传递而来的更丰富语义特征。此外还观察到，SSD 和 SSDLite 在某些数据集上的表现相对较差，这与其架构限制相一致。对于 Faster R-CNN，只有 Pascal VOC 基线的表现优于其学生模型，这可能是因为教师模型额外生成的区域提议给监督过程引入了歧义，从而限制了学生模型的收益，不过这种差异仍然很小。

D. 可解释性分析

为了进一步了解 LUPI训练的学生模型所呈现的性能提升，我们使用 Grad-CAM 可视化进行了可解释性分析 [54]，分析对象为最后一个骨干网络层。图 6 展示了SODA 1米数据集上的结果，并对比了基线模型和学生模型。可视化结果表明，LUPI训练的学生

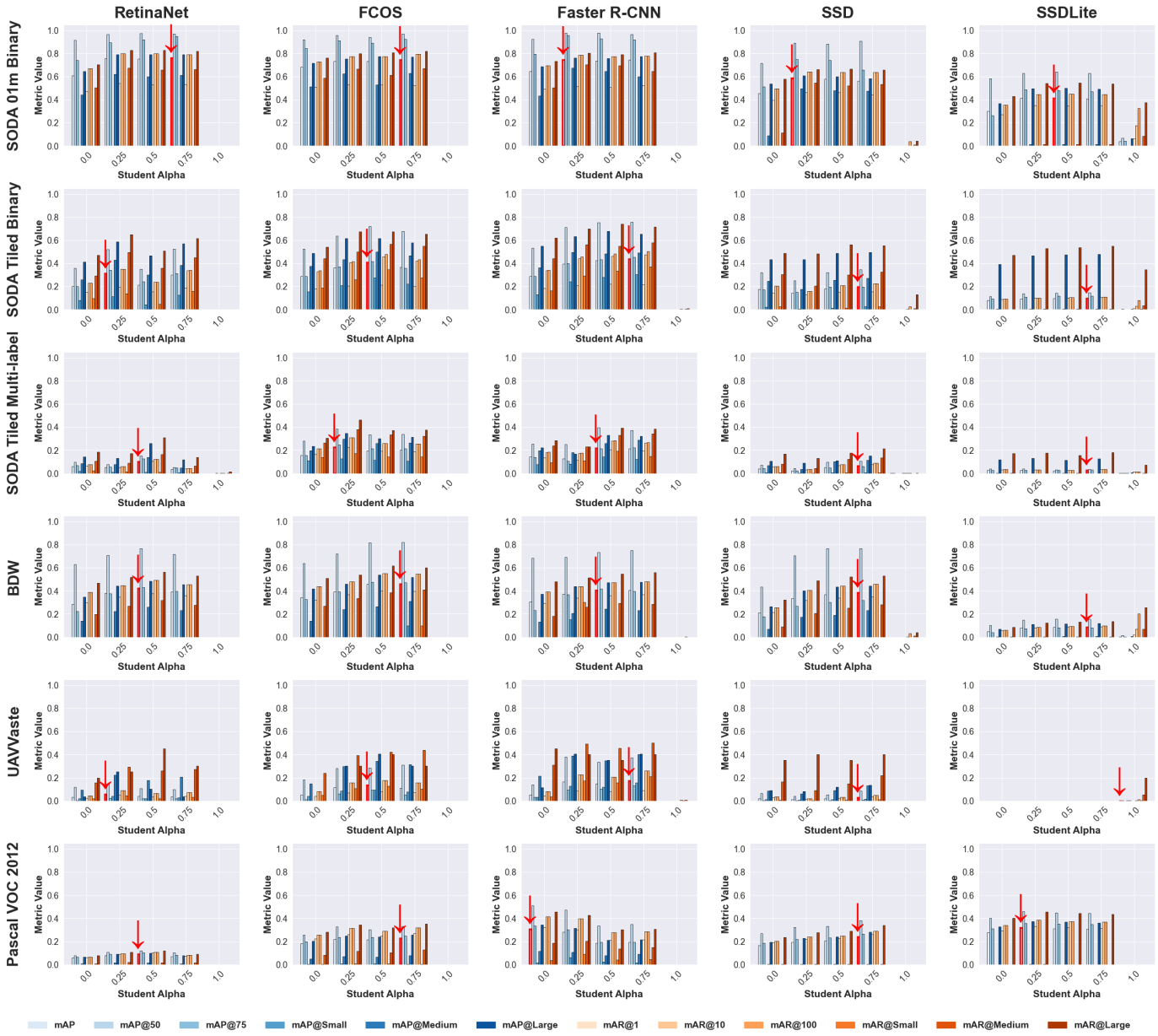

图5. 使用 COCO指标在所有数据集和实验上的消融研究结果。基线模型对应于 $\alpha = 0$；其他曲线表示学生模型。红色向下箭头表示在严格的map@50–95指标下的最佳性能。总体而言， $\alpha = 0.25$ 和 0.5 的结果通常最佳，$\alpha = 0.75$ 的表现也很好，而 $\alpha = 1$ 的平均性能较低。

模型会精准聚焦于垃圾物体，生成置信度更高、误分类更少的检测结果；而基线模型的注意力则不够集中，且经常落在背景中的无关区域。这种有针对性的注意力与严格评估下平均精度均值和 F1分数的提升相一致，表明学生模型不仅在量化指标上表现更好，也学到了更有意义且与任务更相关的特征表示。在其他数据集上也观察到了类似模式，但为简洁起见，未展示这些结果；这表明这一效果具有一致性。

E. 性能与效率分析

虽然我们一贯证明，所提出的基于 LUPI 的目标检测方法能够提高准确率，但有人可能会质疑，这些结果是否以显著增加计算成本为代价。一项主要局限是训练时间增加，因为教师模型和学生模型都必须训练，实际上使工作量翻倍，如图7所示。然而，在实际部署中，推理的执行频率通常远高于训练。表II 显示，基线模型与经 LUPI 训练的学生模型在大小、参数数量、GFLOPS 和帧每秒方面几乎完全相同，速度仅有细微差异，且这一差异在多次运行中表现一致。这表明，尽管训练需要更多时间，推理速度、模型大小和效率均不受影响，

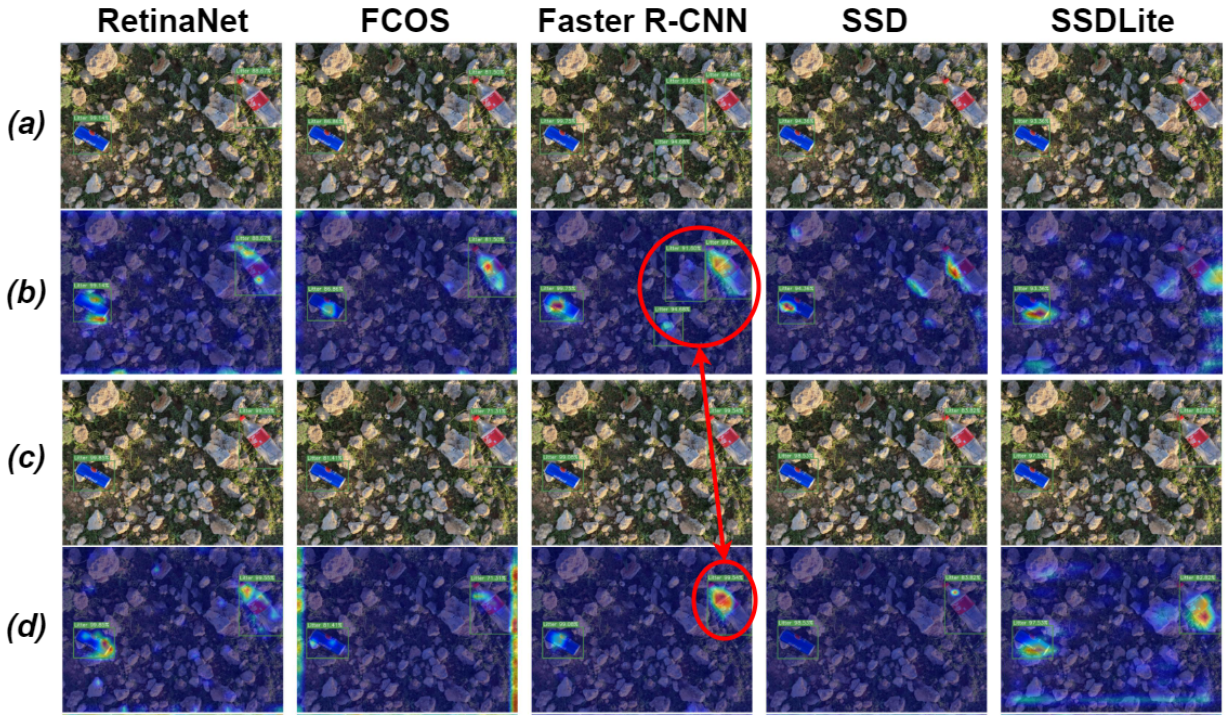

图 6. SODA 1米数据集实验中模型预测与可解释性结果的视觉比较。(a) 基线检测结果。(b) 基线 Grad-CAM 可视化。(c) 最佳 LUPI训练的学生模型检测结果。(d) 最佳学生模型 Grad-CAM 可视化。与基线相比，LUPI训练的学生模型能够更准确地预测垃圾物体。对于应用于相应蒸馏层的 Grad-CAM 可视化，学生模型的注意力更集中于垃圾物体，而基线模型在背景上的激活则更为分散。

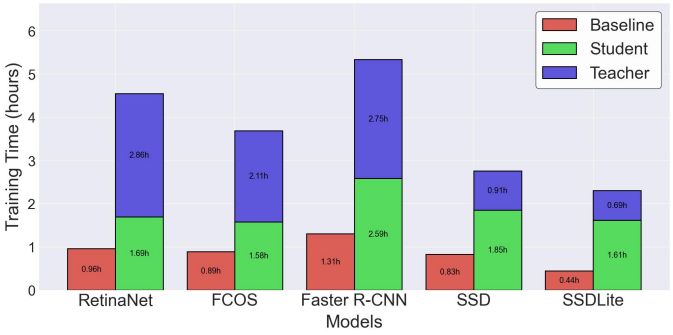

图 7. Pascal VOC 2012 数据集上的训练时间比较，突出显示了 LUPI教师–学生训练所需时间更长。

表II：基线与学生模型在 Pascal VOC 2012 上的运行时间比较，展示模型类型、大小、参数量、GFLOPS 和帧每秒。学生模型的性能提升无需增加推理成本。

模型	Type	大小（MB）	参数量（M）	GFLOPS	FPS
基线	RetinaNet	124.22	32.56	265.10	39.74
	FCOS	122.48	32.11	252.94	39.55
	Faster R-CNN	157.92	41.40	268.75	30.66
	SSD	100.27	26.29	62.94	67.44
	SSDLite	9.42	2.47	0.95	36.74
学生	RetinaNet	124.22	32.56	265.10	38.00
	FCOS	122.48	32.11	252.94	34.65
	Faster R-CNN	157.92	41.40	268.75	30.39
	SSD	100.27	26.29	62.94	67.67
	SSDLite	9.42	2.47	0.95	36.75

从而确保经 LUPI 训练的学生模型所提升的准确率能够在部署中得到充分利用。为简洁起见，这里仅展示 Pascal VOC 2012 的结果，不过在其他数据集上也观察到了类似趋势。

F. 讨论

目标检测是一个多方面的问题，因为不同架构处理定位和分类的方式各不相同。采用区域建议网络和特征金字塔的两阶段检测器会利用空间线索和语义线索，而采用辅助层或轻量化设计的单阶段模型则更多依赖端到端特征提取，这会影响它们对额外指导的响应方式。在我们的实验中，通过以下方式整合特权信息

LUPI 框架始终能够提升学生的学习效果，但提升幅度和性质因模型和数据集而异。Faster R-CNN、FCOS 和 RetinaNet 在无人机垃圾检测中表现更佳，体现了它们利用空间上下文的能力；而 SSD 和 SSDLite 在 PascalVOC 任务中相对表现更好，凸显了特征聚合和感受野方面的差异。消融研究表明，适度的教师权重有助于学生学习，因为它能平衡对真实标注监督和教师指导的依赖；而过度依赖教师有时会使学生在复杂的多标签场景中产生困惑。Grad-CAM 可视化显示，LUPI训练的学生模型能够更精准地关注相关目标，生成更具判别性且语义连贯的表征，

而基线模型的注意力则更加分散。值得注意的是，这些提升只需对架构进行极少改动即可实现，说明该框架能够与各模型的内在特性相辅相成，同时推理速度和效率保持稳定。总体而言，结果揭示了架构、数据集特征和教师指导之间复杂的相互作用，并强调特权信息带来的益处取决于模型设计和任务复杂度；LUPI 能够增强底层模型已有的能力。

VII. 实际应用

本研究提出的目标检测领域中的LUPI范式，可应用于各种目标检测和地理定位系统，并尤其适用于轻量级部署。通过利用紧凑型模型，该方法在保持高准确率的同时降低推理成本，使其即使在资源受限平台上也能高效处理数据。因此，该方法适用于需要快速、可靠且一致检测的现实应用，例如无人机监测、交通分析和监控系统；在这些场景中，速度和精确率都至关重要。

VIII. 结论

本研究通过在多种架构和数据集上开展一系列广泛实验，探讨了目标检测领域中的LUPI范式，并采用严格的COCO指标进行评估。结果一致表明，在训练期间引入特权信息能够提升检测准确率，而且无需增加模型深度、参数量或推理时间即可提高准确率。在该范式下训练的学生模型，其架构和效率均与对应的基线模型完全相同，但通过训练期间获得额外的教师指导，实现了更高的准确率。

仍存在一些局限性。特权信息的生成可能受到同类别重叠目标、较大边界框造成的遮挡，以及掩码表示中色彩区分度有限等因素的影响。此外，与传统的单模型设置相比，同时训练教师模型和学生模型会延长训练阶段。

未来可从多个方向拓展这项工作，包括将该方法与更新的检测架构相结合，例如 YOLOv12 [56] 和 RF-DETR [57]，；探索更丰富、更多样的特权信息形式，例如语义图或基于注意力的线索；以及将该框架应用于目标分割等相关任务。总体而言，本研究结果再次证实，LUPI 为在计算资源受限的环境中提升目标检测性能提供了一种实用且有效的策略，在保持推理效率不变的同时实现更高的准确率。

参考文献

[1] J. Redmon、S. Divvala、R. Girshick 和 A. Farhadi，《You Only Look Once：统一的实时目标检测》，2016年，引用 arxiv:1506.02640。[2] S. Ren、K. He、R. B. Girshick 和 J. Sun，《Faster R-CNN：借助区域建议网络实现实时目标检测》，CoRR，第 abs/1506.01497 卷，2015年。[3] T. Lin、P. Goyal、R. B. Girshick、K. He 和 P. Doll´ar，《密集目标检测的焦点损失》，载于《2017 IEEE国际计算机视觉会议》（ICCV）。意大利威尼斯：IEEE计算机学会，2017年，第2999–3007页。[4] N. Carion、F. Massa、G. Synnaeve、N. Usunier、A. Kirillov 和 S. Zagoruyko，《利用Transformer实现端到端目标检测》，CoRR，第 abs/2005.12872 卷，2020年。[5] Y. Zhao、W. Lv、S. Xu、J. Wei、G. Wang、Q. Dang、Y. Liu 和 J. Chen，《DETR在实时目标检测方面胜过YOLO系列模型》，2024年。[6] Z. Li、Y. Dong、L. Shen、Y. Liu、Y. Pei、H. Yang、L. Zheng 和 J. Ma，《目标检测的发展与挑战：综述》，Neurocomputing，第598卷，第128102页，2024年。[7] D. Prasad，《真实图像中目标检测问题综述》，《国际图像处理杂志》（IJIP），第6卷，441，2012年12月，第 .。[8] M. Bartolo、K. Makantasis 和 D. Seychell，《利用特权信息进行垃圾检测学习》，载于《2025年第13届欧洲视觉信息处理研讨会》（EUVIP）论文集，马耳他瓦莱塔，2025年。[9] V. Vapnik 和 A. Vashist，《一种新的学习范式：利用特权信息学习》，《神经网络》，第22卷，557，2009年5月6日，544，第 - 页– 。[10] D. Pechyony 和 V. Vapnik，《关于利用特权信息学习的理论》，《神经信息处理系统进展》，第23卷，2010年。[11] V. Vapnik、R. Izmailov 等，《利用特权信息学习：相似性控制与知识迁移》，《机器学习研究杂志》，第16卷，第1期，第2023–2049页，2015年。[12] V. Vapnik 和 A. Vashist，《一种新的学习范式：利用特权信息学习》，《神经网络：国际神经网络学会官方期刊》，第22卷，第544–57页，2009年7月。[13] S. Wang、S. Chen、T. Chen 和 X. Shi，《利用特权信息进行多标签分类》，《模式识别》，第81卷，第60–70页，2018年。[14] V. Sharmanska、N. Quadrianto 和 C. H. Lampert，《学习迁移特权信息》，CoRR，第 abs/1410.0389 卷，2014年。[15] H. Bay、T. Tuytelaars 和 L. Van Gool，《SURF：加速稳健特征》，《IEEE欧洲计算机视觉会议论文集》，第404–417页，2006年。[16] E. Rublee、V. Rabaud、K. Konolige 和 G. Bradski，《ORB：SIFT或SURF的高效替代方案》，载于《2011年国际计算机视觉会议》，2011年，第2564–2571页。[17] P. Viola 和 M. Jones，《利用简单特征的提升级联快速检测目标》，载于《2001年IEEE计算机学会计算机视觉与模式识别会议论文集》（CVPR 2001），第1卷。IEEE，2001年，第 I–511 页。[18] H. Bhatt、V. Shah、K. Shah、R. Shah 和 M. Shah，《黑色素瘤皮肤癌检测与分类的最新机器学习技术：全面综述》，《智能医学》，第3卷，第3期，第180–190页，2023年。[19] Y. LeCun、Y. Bengio 和 G. Hinton，《深度学习》，Nature，第521卷，第436–44页，2015年5月。[20] G. Chen、H. Wang、K. Chen、Z. Li、Z. Song、Y. Liu、W. Chen 和 A. Knoll，《小目标检测的四大支柱综述：多尺度表示、上下文信息、超分辨率和区域建议》，《IEEE系统、人类与控制论汇刊：系统》，第52卷，第2期，第936–953页，2022年。[21] W. Liu、D. Anguelov、D. Erhan、C. Szegedy、S. E. Reed、C. Fu 和 A. C. Berg，《SSD：单次多框检测器》，载于《计算机视觉——ECCV 2016——第14届欧洲计算机视觉会议，阿姆斯特丹，荷兰，2016年10月11日至14日，论文集，第一部分》，《计算机科学讲义》系列，B. Leibe、J. Matas、N. Sebe 和 M. Welling 编。施普林格，2016年，第9905卷，第21–37页。[22] K. He、G. Gkioxari、P. Doll´ar 和 R. B. Girshick，《Mask R-CNN》，载于《2017 IEEE国际计算机视觉会议》（ICCV）。意大利威尼斯：IEEE计算机学会，2017年，第2980–2988页。[23] X. Zhou、D. Wang 和 P. Kr¨ahenb¨uhl，《目标即点》，CoRR，2019年。[24] F. C. Akyon、O. Altinuc 和 S. Temizel，《切片辅助超推理与微调用于小目标检测》，载于《2022 IEEE国际图像处理

会议》（ICIP）。中国上海：IEEE，2022年10月。[25] J. Feyereisl、S. Kwak、J. Son 和 B. Han，《基于利用特权信息的结构化支持向量机进行目标定位》，载于《神经信息处理系统进展》，Z. Ghahramani、M. Welling、C. Cortes、N. Lawrence 和 K. Weinberger 编，第27卷。Curran Associates, Inc.，2014年。[26] C. Wah、S. Branson、P. Welinder、P. Perona 和 S. Belongie，《Caltech-UCSD Birds-200-2011数据集》，加州理工学院，技术报告 CNS-TR-2011-001，2011年。[27] S. Sun、C. Zhang 和 Y. Tian，《一种利用特权信息进行结构化学习的新方法》，载于《计算科学——ICCS 2018》，Y. Shi、H. Fu、Y. Tian、V. V. Krzhizhanovskaya、M. H. Lees、J. Dongarra 和 P. M. A. Sloot 编。尚：Springer国际出版公司，2018年，第453–461页。[28] V. Sharmanska、N. Quadrianto 和 C. H. Lampert，《利用特权信息进行排序学习》，载于《2013年IEEE国际计算机视觉会议》，2013年，第825–832页。[29] ——，《学习迁移特权信息》，arXiv预印本 arXiv:1410.0389，2014年。 [在线]。可访问：https://arxiv.org/abs/1410. 0389[30] K. Makantasis、D. Melhart、A. Liapis 和 G. N. Yannakakis，《利用特权信息对真实环境中的情感进行建模》，载于《2021年第9届情感计算与智能交互国际会议》（ACII）。IEEE，2021年，第1–8页。[31] K. Makantasis、K. Pinitas、A. Liapis 和 G. N. Yannakakis，《从实验室走向真实环境：通过特权信息进行情感建模》，《IEEE情感计算汇刊》，第15卷，第2期，第380–392页，2023年。[32] D. Lopez-Paz、L. Bottou、B. Sch¨olkopf 和 V. Vapnik，《统一蒸馏与特权信息》，arXiv预印本 arXiv:1511 03643，2015年。[33] G. Habib、T. jan Saleem、S. M. Kaleem、T. Rouf 和 B. Lall，《计算机视觉中的知识蒸馏全面综述》，2024年。[34] Z. Zheng、R. Ye、Q. Hou、D. Ren、P. Wang、W. Zuo 和 M.-M. Cheng，《目标检测的定位蒸馏》，2022年。[35] G. E. Hinton、O. Vinyals 和 J. Dean，《蒸馏神经网络中的知识》，CoRR，第 abs/1503.02531 卷，2015年。[36] T. Boger 和 T. Ullman，《什么是“位置”：物理推理为目标位置提供依据》，《Open Mind》（剑桥），第7卷，第130–140页，2023年5月。[37] L. Itti、C. Koch 和 E. Niebur，《一种基于显著性的视觉注意力模型，用于快速场景分析》，《IEEE模式分析与机器智能汇刊》，第20卷，第11期，第1254–1259页，1998年11月。[38] A. Linardos、M. K¨ummerer、O. Press 和 M. Bethge，《针对最先进显著性建模的域内与域外校准预测》，CoRR，第 abs/2105.12441 卷，2021年。[39] D. Seychell 和 C. J. Debono，《RGB-D内容中视觉显著性区域的排序》，载于《2018年3D沉浸国际会议》（IC3D），2018年。[40] L. Yang、B. Kang、Z. Huang、X. Xu、J. Feng 和 H. Zhao，《Depth Anything：释放大规模无标签数据的力量》，2024年。[41] R. Ranftl、A. Bochkovskiy 和 V. Koltun，《用于密集预测的视觉Transformer》，CoRR，第 abs/2103.13413 卷，2021年。[42] M. Bartolo 和 D. Seychell，《目标检测性能与视觉显著性及深度估计的相关性》，2024年。[43] D. Seychell、M. Kenely、M. Bartolo、C. J. Debono、M. Bugeja 和 M. Sacco，《自动高效标注二值掩码以增强计算机视觉模型训练》，载于《2023年IEEE国际多媒体研讨会》（ISM），2023年，第256–259页。[44] A. Kirillov、E. Mintun、N. Ravi、H. Mao、C. Rolland、L. Gustafson、T. Xiao、S. Whitehead、A. C. Berg、W.-Y. Lo、P. Doll´ar 和 R. Girshick，《 Segment Anything 》，2023 年。[45] K. He 、 X. Zhang 、S. Ren 和 J. Sun，《深入研究整流器：在ImageNet分类中超越人类水平的性能》，CoRR，第 abs/1502.01852 卷，2015年。[46] M. Sandler、A. Howard、M. Zhu、A. Zhmoginov 和 L.-C. Chen，《MobileNetV2：倒置残差与线性瓶颈》，2019年。[47] Z. Tian、C. Shen、H. Chen 和 T. He，《FCOS：全卷积单阶段目标检测》，载于《2019 IEEE/CVF国际计算机视觉会议》（ICCV）。韩国首尔：IEEE，2019年，第9626–9635页。[48] K. Makantasis、K. Pinitas、A. Liapis 和 G. N. Yannakakis，《从实验室走向真实环境：通过特权信息进行情感建模》，《IEEE情感计算汇刊》，第15卷，第2期，第380–392页，2024年。[49] D. Pisani、D. Seychell、C. J. Debono 和 M. Schembri，《SODA：无人机拍摄图像中小目标检测数据集》，载于《2024 *IEEE*国际图像处理会议 *(ICIP)*，2024，页码 151–157。[50] M. C´ordova、A. Pinto、C. C. Hellevik、S. A.-A. Alaliyat、I. A. Hameed、H. Pedrini 和 R. d. S. Torres，《深度学习垃圾检测：一项比较研究》，《传感器》，第 22 卷，第 2 期，2022年。[51] J. Wang、W. Guo、T. Pan、H. Yu、L. Duan 和 W. Yang，《使用低空无人机在野外进行瓶子检测》，载于 *2018* 年第 *21* 届国际信息融合会议（*FUSION*），2018 年，页码 439–444。[52] M. Kraft、M. Piechocki、B. Ptak 和 K. Walas，《无人机采集的低空航拍图像中基于机载视觉的自主垃圾检测》，《遥感》，第 13 卷，5，2021 年第 . 期。[53] M. Everingham、L. Van Gool、C. K. I. Williams、J. Winn 和 A. Zisserman，《PASCAL视觉目标类别挑战赛2012（VOC2012）结果》，http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html，2012年。[54] R. R. Selvaraju、A. Das、R. Vedantam、M. Cogswell、D. Parikh 和 D. Batra，《Grad-CAM：你为何这么说？通过基于梯度的定位解释深度网络的视觉结果》，*CoRR*，第 abs/1610.02391 卷，2016年。[55] T.-Y. Lin、M. Maire、S. J. Belongie、J. Hays、P. Perona、D. Ramanan、P. Doll´ar 和 C. L. Zitnick，《Microsoft COCO：上下文中的常见物体》，载于《计算机视觉——*ECCV 2014*——第 *13* 届欧洲会议，瑞士苏黎世，*2014* 年 *9* 月 *6–12* 日，论文集，第 *V* 部分》，《计算机科学讲义》系列，D. J. Fleet、T. Pajdla、B. Schiele 和 T. Tuytelaars 编，第 8693 卷。尚，瑞士：施普林格，2014 年，页码 740–755。[56] Y. Tian、Q. Ye 和 D. Doermann，《YOLOv12：以注意力为核心的实时目标检测器》，2025年。[57] I. Robinson、P. Robicheaux 和 M. Popov。（2025）RF-DETR。https://github. com/roboflow/rf-detr。最先进的实时目标检测模型。



Matthias Bartolo 获得了马耳他大学人工智能专业的信息技术理学学士学位（荣誉），以一等荣誉毕业，成绩在年级中名列第一，并因优异的学业表现入选院长荣誉名单。随后，他获得了马耳他大学的理学硕士学位。他的研究兴趣包括计算机视觉、应用机器学习，以及开发协助和支持人们的人工智能解决方案。他参与了多个研究项目，并发表了多篇同行评审出版物。此外，Matthias 还积极参与志愿服务活动，充分体现了他致力于运用专业技术服务他人的坚定承诺。



Dylan Seychell （IEEE 高级会员）在马耳他马耳他大学姆西达校区获得了计算机视觉专业的信息技术理学学士（荣誉）、理学硕士和哲学博士学位。他目前是马耳他大学人工智能系的常驻学术人员和高级讲师。他的研究兴趣包括计算机视觉、应用机器学习和《遥感》，重点关注如何在实际运行环境中部署人工智能系统。Seychell 博士担任多个国家资助研究项目的首席研究员，这些项目聚焦于从事环境可持续性和媒体诚信方面的工作。他同时也是马耳他数字创新管理局（MDIA）的技术专家。



Gabriel Hili 获得了马耳他大学颁发的理学硕士学位，研究方向为新闻媒体领域的应用计算机视觉。目前，他是马耳他大学的一名研究人员，参与与马耳他政府清洁与维护司合作开展的研究项目。他的研究兴趣包括计算机视觉和语言技术，重点关注人工智能在公共基础设施和媒体中的实际应用。Hili先生还通过协助高级教职人员开展讲座，为学术课程作出贡献以及教学材料的开发。



Matthew Montebello （IEEE高级会员）是马耳他大学的学者和研究人员，在高等教育领域拥有逾35年的经验。他拥有计算机科学和教育学博士学位，在个性化、人工智能和数字教育方面具有专业研究专长。他的工作聚焦于高等教育中的人工智能、生成式人工智能、电子档案袋，以及合乎伦理、以人为本的教育技术。他曾领导和协调多个由国家及国际机构资助的项目，在知名学术出版社发表著作，并因创新和研究成果获得奖项。他积极参与课程设计、博士生指导及机构人工智能战略的制定。



Carl James Debono （IEEE高级会员）于1997年获得马耳他大学电气工程工学学士（荣誉）学位，并于2000年获得意大利帕维亚大学电子学与计算机工程哲学博士学位。1997年至2001年，他在马耳他大学担任集成电路设计领域的研究工程师。2001年，他受聘为马耳他大学通信与计算机工程系讲师，目前任教授。他目前担任担任马耳他大学信息与通信技术学院院长。Debono教授参与了多个本地和欧洲研究项目，研究领域涉及通信系统以及图像/视频处理。他的研究兴趣包括多视点视频编码、弹性多媒体传输和计算机视觉。



Saviour Formosa 是马耳他大学的高级学者，同时也是国家和国际公共部门机构的顾问。他曾任犯罪学系主任（2016年至2020年），拥有时空环境犯罪学博士学位、地理信息系统理学硕士学位以及社会学文学荣誉学士学位。他的研究重点包括时空分析、三维场景重建、先进扫描技术以及安全与安保。他曾参与并领导多个欧盟和国家级重大项目，包括InMotion、SIntegraM和SpatialTRAIN，成功获得超过8000万欧元的资金。他通过马耳他大学的沉浸实验室开展沉浸式研究和数字孪生研究。



Konstantinos Makantasis （IEEE会员）获得了克里特技术大学计算机工程专业的文凭、理学硕士和博士学位。他是马耳他大学人工智能系的讲师。他获得了著名的玛丽·斯克沃多夫斯卡-居里行动国际人才引进拓展奖学金，以开展基于张量的情感建模机器学习方法研究。他在计算机视觉、信号与图像处理以及机器学习领域的国际期刊和会议上发表了80多篇论文。他主要从事并关注计算机视觉，机器学习/模式识别以及概率编程。