arXiv:2603.12409v1 [cs.CV] 12 Mar 2026

ABRA：跨域传送微调知识，实现开放词汇目标检测

MattiaBernardi、Chiara Cappellino Matteo Mosconi、
Enver Sangineto、Angelo Porrello 和 Simone Calderara

摩德纳与雷焦艾米利亚大学
name.surname@unimore.com

摘要。 尽管近期的开放词汇目标检测架构（如 Grounding DINO）展现出强大的零样本能力，但在域偏移下，其性能会显著下降。此外，许多具有实际应用价值的域（例如夜间或雾天场景）缺少大规模标注数据集，无法直接进行微调。本文提出**用于适应的对齐基迁移（ABRA）**，这是一种将类别特定检测知识从有标注的源域迁移到目标域的方法；在目标域中，无法获取包含这些类别的训练图像。ABRA 将这种适应过程表述为预训练检测器权重空间中的几何传输问题，通过对齐源域专家与目标域专家来传送类别特定知识。针对各种具有挑战性的域偏移开展的大量实验表明，在多种不利条件下，ABRA 都能成功传送类别级专门化。我们的代码将在论文被接收后公开。

1 引言

近期开放词汇目标检测（OVD）的进展表明，视觉语言模型 [25]能够通过自然语言提示，在推理时定位一系列广泛的概念，从而有效摆脱固定训练类别的限制。尽管取得了这一进展，但在域偏移情况下，性能通常会大幅下降，例如从白天场景转向低光照场景，或从晴朗天气转向雾天。此外，在此类不利情境下，获取大规模、完整标注的检测数据集成本极高，甚至不可行，尤其是对于稀有类别而言。

域自适应目标检测（DAOD）方法通常通过在带标注的源域上训练检测器，并在无标注的目标图像上生成伪边界框以进行微调，来解决这一问题 [7, 30]。然而，在严重的域偏移情况下（例如夜间图像），这些伪标签往往不可靠。此外，要求目标域图像包含所有感兴趣类别——即使没有边界框标注——也会隐含地引入一种弱监督形式，因为必须预先筛选图像，以确保涵盖所有类别 [28]。

本文提出了一种不同的适应场景：训练期间只能获取目标域类别的一个子集（例如，汽车、自行车），而其他目标类别（例如，公交车、卡车）仅在源域中有标注，且目标域中没有可用的训练数据：也就是说，**既没有边界框，也没有图像**。这种设定反映了一个常见的现实问题：常见类别通常有现成的标注数据，而稀有类别则没有。由于标准的域自适应目标检测方法不适用于这些稀有类别，因此从完全标注的源域迁移知识至关重要。形式上，我们假设一个在所有类别上都有完整监督的源域，而目标域类别则被划分为目标域可用和目标域不可用两组。我们的目标是专门将知识从源域迁移到目标域不可用类别。

为应对上述情境，我们以开放词汇检测骨干网络（即，Grounding DINO [18]）为基础，将适应过程分解为两个互补的专业层面：域和类别。在我们的框架中，这种分解有助于将类别特定知识从源域有效迁移至目标域。在域层面，我们引入对象化这一过程，用以构建一个类别无关的域专家，使其能够捕捉域特定特征，例如光照条件和纹理模式。在类别层面，我们采用 [16] 提出的方法，生成轻量的类别专家，以编码类别特定语义知识。

基于这一模块化结构，我们提出 **ABRA**（**用于适应的对齐基迁移**），这是一种将类别层级知识从源域迁移到目标域的方法，即使目标域中没有这些类别的训练图像也能实现迁移。具体而言，ABRA 将这一适应问题建模为权重空间中的几何传输，通过闭式旋转对齐源域和目标域专家。由此得到的迁移后类别特定残差可直接接入目标域专家，从而构建出有效模型，无需访问这些类别的数据。

我们在 Cityscapes [5] → Foggy Cityscapes [27] 迁移上评估 ABRA；此外，我们在单源域泛化目标检测（SDGOD）设定 [33],中进一步研究了另外四种域偏移，涵盖光照与天气条件的变化。在这些基准测试中，我们将我们的方法与若干基于迁移学习和权重对齐的基线方法进行比较，结果表明 ABRA 在多种场景下均持续优于它们。特别地，它展现出强大的零样本与少样本迁移能力，为下游适应任务提供了更好的权重初始化。总之，我们的贡献如下：

– 我们提出一个能够解耦域与类别知识的框架，从而可将专门知识跨域重新定位。– 我们提出ABRA（用于适应的对齐基迁移），这是一种具有数学基础的方法，通过域对齐和类别级残差传输迁移类别知识，从而能够适应目标训练集中不存在的目标域类别。

– 我们在多种域偏移上验证了ABRA的有效性，其持续优于现有基线方法，并在具有挑战性的条件下展现出鲁棒性。

2 相关工作

开放词汇目标检测。开放词汇目标检测旨在识别并定位预训练期间未见过的目标类别，使用户能够在推理时通过文本提示指定任意类别。早期方法 [8, 21, 37, 40, 41] 利用强大的预训练编码器（如 CLIP [25]）对齐视觉和语言特征。后续工作 [14,18]则强调采用专为目标检测设计的大规模预训练方案，并纳入大规模视觉定位数据集。尽管这些模型旨在检测几乎所有目标，但它们在识别稀有类别以及在不同域中进行稳健检测方面往往存在局限。

尽管先前研究 [4, 6] 对 Grounding DINO [18] 等模型进行适应，以应对增量学习中的类别稀缺问题， [32], 但它们未能充分应对在不同域上下文中检测同一类别这一挑战。本文通过提出一种适应方法来弥补这一不足，该方法将域特定语义与类别特定语义解耦，从而实现跨域传送。

模块化深度学习与参数高效微调。模块化深度学习 [24] 与传统的整体式训练范式形成对比，它将神经网络视为可组合的实体，并可有选择地对其进行适应。秉持这一理念，参数高效微调技术（PEFT）应运而生。例如，LoRA [10] 将低秩矩阵注入预训练骨干网络，显著降低了下游适应所需的内存。类似 LoRA 的方法 [13,15–17,20,39]进一步优化了可训练参数与性能之间的权衡。例如，SVFT [16], 在每个权重矩阵的奇异值分解子空间中应用低秩残差，以更少的参数实现了与 LoRA 相当的性能。

我们秉持模块化深度学习理念，但我们的方法有所不同：我们明确将适应划分为域级模块和类别级模块。较大的组件负责全面的域适应，而较小的专用模块则编码类别特定知识。

域自适应目标检测。大多数最先进的域自适应目标检测方法依赖于平均教师 [31] 范式，其中教师检测器生成伪边界框，用于训练学生网络 [30]。两个网络均由源域检测器初始化，教师通过学生网络参数的移动平均进行更新。由于在强烈的域偏移下，伪边界框可能变得不可靠，近期方法采用自适应数据增强，以减轻向噪声预测漂移的问题。例如， [12] 利用 MixUp [38] 组合不同类别的边界框，并通过动态类别混淆矩阵估计相似度。类似地， [19] 使用检测不确定性作为混合系数，对来源图像和目标域图像进行融合，其中不确定性由一个额外的网络分支建模，该分支预测边界框坐标上的高斯分布

分布。在 [7], 中，数据增强与类别特定自适应阈值相结合，以筛除低置信度边界框并缓解类别不平衡问题。

相关研究探索了单域泛化开放集目标检测（SG-OSOD） [36], ，该任务要求模型在应对域偏移的同时，将完全未见过的目标对象归入单一“未知”类别。相比之下，我们的方法应对的是一个本质上不同的挑战：我们假设感兴趣的类别在源域中已知且有标注，但在目标域中完全不可用（没有图像，也没有标签）。因此，我们关注的不是检测新颖目标，而是明确地将微调知识跨域迁移，以便在不利的目标环境中检测这些特定类别（第 1 节和第 3 节）。

权重对齐。 以往研究探索了迁移已学习知识的策略。模型拼接 [2,9,23] 通常会直接拼接网络，而模型重新基准化方法 [1,26,29] 表明，可以通过权重对齐（例如置换），在不同预训练模型之间迁移微调模块。此外，任务算术 [11,22] 尝试通过简单的算术运算迁移不同微调模块中的知识。

我们的方法与模型重新基准化相似：首先计算一个最优旋转，对齐源域和目标域的空间。随后，将这一旋转应用于类别特定模块，确保它们与新目标域完全兼容。此过程使得已学习的类别知识能够直接从一个域迁移到另一个域。

3 预备知识

问题设定。在我们的框架中，我们采用预训练的开放词汇检测器，例如 Grounding DINO [18]，其参数化为 θ_0。如图1所示，我们的目标是将类别特定知识从源 域迁移到目标 域。在标准的全监督设置中，每个域由

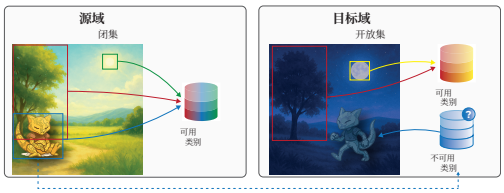

图1：目标。 在无法访问目标域数据的情况下，将类别特定知识从源域迁移到不可用的目标类别。

由一组标注图像构成的数据集：源域为 $\mathcal{D}_S = \{(x_i^S, y_i^S)\}_{i=1}^{N_S}$，目标域为 $\mathcal{D}_T = \{(x_i^T, y_i^T)\}_{i=1}^{N_T}$，其中 x_i 表示图像， y_i 表示其真实目标标注（边界框和类别标签）。

在我们的设定中，我们假设目标域中有些类别不可用。具体而言，我们用 $\mathcal{C}_S$ 和 $\mathcal{C}_T$ 分别表示源域和目标域中定义的目标类别集合。随后，我们将目标类别划分为两个互不相交的集合，即., 一个可用 集合 $\mathcal{C}_T^{\text{avail}}$和一个不可用 集合 $\mathcal{C}_T^{\text{unav}}$，其中 $\mathcal{C}_T = \mathcal{C}_T^{\text{avail}} \cup \mathcal{C}_T^{\text{unav}}$和 $\mathcal{C}_T^{\text{avail}} \cap \mathcal{C}_T^{\text{unav}} = \emptyset$。 $\mathcal{C}_T^{\text{avail}}$中的类别有标注的目标域图像作为支持， $\mathcal{D}_T^{\text{avail}}$；而 $\mathcal{C}_T^{\text{unav}} \subseteq \mathcal{C}_S$ 中的类别则

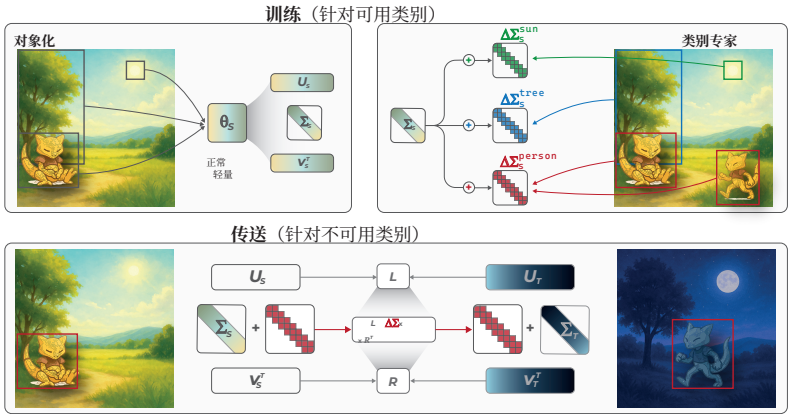

图2：迁移流程概览。从通过对象化获得的域专家出发，将在源域中通过 SVFT 学到的类别特定残差以解析方式传送到未见过的目标域。

仅在源域中出现，缺少任何目标域标签或视觉数据。这与标准的域自适应目标检测（DAOD）场景有着根本区别：在标准场景中，目标域中的每个类别都有可用样本，只是缺少监督。

如图1所示，$\mathcal{C}_T^{\text{avail}} = \{\text{树、月亮}\}$ 和 $\mathcal{C}_T^{\text{unav}} = \{\text{人}\}$，这一配置反映了现实场景：稀有类别（例如，夜间的摩托车）面临数据缺乏，而常见类别（例如，夜间的汽车）则仍然容易获取。此外，$\mathcal{C}_T^{\text{avail}}$和 $\mathcal{C}_S$ 的交集不受限制，可能为空，也可能非空。我们强调，目标域中完全没有用于迁移的类别。

奇异向量引导微调。为了高效地使预训练检测器适配新类别，我们采用奇异向量引导微调（SVFT） [16], ，这是一种参数高效微调策略。SVFT 不更新完整的权重矩阵，而是在谱域中对预训练矩阵的奇异值学习简洁的残差 $\Delta\Sigma$
$\theta_0 = U\Sigma V^\top$。由此得到适配后的参数化表示 $\theta_{FT} = U\left(\Sigma + \Delta\Sigma\right)V^\top$。这种紧凑表示便于创建轻量类别模块，使其能够在不同域之间高效迁移。

4 面向适应的对齐基迁移

我们提出了 ABRA，这是一种新颖的方法，旨在将特定类别的知识从源域迁移到目标域，且整个过程无需类别特定的目标域数据或训练迭代。近期的理论见解 [26] 表明，模型适配可以重新表述为一个运输问题。基于这一视角，我们从源域和目标域模块入手，分别记为 θ_S 和 θ_T，二者均

通过类别无关微调获得。随后，我们将特定类别的残差从源模块迁移到目标模块。具体而言，我们为目标类别 c推导出一个类别特定专家，记为 $\theta_S^{(c)} = \theta_S + \tau_{(c)\,S}$，其中 $\tau_S^{(c)}$表示仅在源域中观测到的类别 c 上进行微调所获得的残差。我们的最终目标是将这一微调位移 $\tau_S^{(c)} = \theta_S^{(c)} - \theta_S$迁移到目标域模块 θ_T上，从而得到理想检测器 $\theta_T^{(c)}$ 的估计值 $\hat{}\theta_T^{(c)}$ 。关键在于，这一理想检测器对应于这样一种模型：如果能够获得目标域中类别 c 的数据，直接在该类别上进行微调，便会自然得到该模型。图2全面展示了所提出的方法。

组织结构。 在第4.1节中，我们详细介绍了域专家模型 θ_S 和 θ_T的推导过程，旨在促使模型形成强大的类别无关表征。此外，第4.2节概述了在源域中推导类别专家 $\theta_S^{(c)} = \theta_S + \tau_{(c)\,S}$所采用的具体流程。最后，ABRA 实现的知识迁移过程可概括如下：

$$\hat{\theta}_T^{(c)} = \theta_T + \pi_{S\to T}(\tau_S^{(c)}), \quad \text{where } \tau_S^{(c)} = \theta_S^{(c)} - \theta_S \tag{1}$$

其中， $\pi_{S\to T}(\cdot)$ 表示将源域 S连接到目标域 T的传输函数，详见第4.3节。

4.1 通过对象化构建域专家

在图2左上部分，我们展示了域专家 θ_S 的推导过程，这是我们方法的第一阶段。我们将域专家定义为一种检测模型，它能够捕获特定域的底层视觉统计特征（例如雾天或雨天条件），同时严格不考虑单个类别。我们的目标是使该模型内化光照、纹理和结构特征，而非

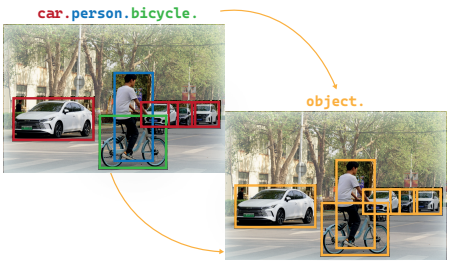

图3：对象化。 通过使用统一的“对象”标签对每个域进行微调，以捕获域外观。

而非依赖语义性、特定类别的特征。为了训练该域专家，我们引入了一种称为对象化的流程，如图 3 所示。在此过程中，训练集的真实标注经过修改，以消除类别级监督，从而突出域级视觉线索。

实际上，对于属于最具代表性的前三类类别的每个真实边界框，原始语义标签（例如，汽车、人、自行车）会被替换为单一的中性“物体”类别。该策略迫使模型定位感兴趣的通用物体，而不是识别其精确类别。形式上，给定特定域，我们通过将最具代表性的前三类的边界框重新标注为“物体”，同时丢弃所有其他标注，构建对象化数据集$\tilde{\mathcal{D}}_S$ 。具体而言：

$$\theta_S = \text{微调}(\theta_0, \quad \tilde{\mathcal{D}}_S), \quad \text{其中} \quad \tilde{\mathcal{D}}_S = \{(x_i, \text{“对象”})\}_{i=1}^{N_S^{\text{前3名}}} \tag{2}$$

其中 θ_0 表示Grounding DINO的预训练参数，x_i 表示输入图像，$N_S^{\text{top-3}}$ 是至少包含三个主导类别之一的图像总数。所得模型 θ_S充当源域专家。采用相同流程可得到目标域专家 θ_T，并可推广以获得任意感兴趣域的专家。

4.2 通过奇异向量实现类别专家

在明确域专家的作用与推导过程后，接下来便是从拟迁移的类别中提取类别特定知识。为清晰起见，我们继续将源域中的类别称为源类别，以区别于它们将被迁移到的目标域。不过，以下原则普遍适用于任何域中的任何类别。

令 c_S =公交车表示一个特定类别，该类别在源数据集中有监督标注，但在目标域中完全没有训练数据。基于域专家 θ_S，我们的目标是训练一个专用于检测此类别实例的轻量级类别特定模块。此外，我们要求由此得到的类别专家能够在域之间完全迁移，而无需任何额外的目标域训练。

为实现这一目标，我们首先筛选训练集，仅保留至少包含一个目标类别实例的图像，并严格屏蔽属于其他任何类别的所有边界框。随后，采用第3节中详述的技术，在域专家 θ_S 上高效地微调一个紧凑残差。具体而言，对于第 ℓ 层的源域专家权重矩阵 $\theta_{S,\ell}$，我们首先对其进行奇异值分解，得到$U_{S,\ell}\Sigma_{S,\ell}V_{S,\ell}^\top$。然后，我们初始化一个可训练的类别特定残差矩阵 $\Delta\Sigma_{S\ell}^{(c)}$，同时冻结原始组件 $U_{S,\ell}$、$\Sigma_{S,\ell}$ 和 $V_{S,\ell}^\top$。在优化过程中，仅微调这个最小残差 $\Delta\Sigma_{S\ell}^{(c)}$。因此，第 ℓ 层的前向传播 f 可表示为：

$$f_\ell(x) = \underbrace{U_{S,\ell}}_{\text{frozen}} \Big(\underbrace{\Sigma_{S,\ell}}_{\text{frozen}} + \underbrace{\Delta\Sigma_{S,\ell}^{(c)}}_{\text{learnable}} \Big) \underbrace{V_{S,\ell}^\top}_{\text{frozen}} x. \tag{3}$$

这种训练范式自然会对残差矩阵 $\Delta\Sigma_{S\ell}^{(c)}$ 施加低秩约束。实践中，这些残差可以限制为对角结构（参数化为向量）或带状结构（例如，三对角结构）。这大幅减少了每个类别专家所需的可学习参数数量，同时保持了模型整体的高效性和紧凑性。

完成这一微调阶段后，我们获得了一个类别专用模型$\theta_S^{(c)} = \theta_S + \tau_S^{(c)}$，该模型专为域 S中的类别 c 量身打造，如图2右上部分所示。接下来，提取出的残差 $\tau_S^{(c)}$ 可用于

在源域上直接推理，也可将其迁移至一个全新的目标域，具体内容将在下一节中详述。

4.3 通过权重对齐实现传送

有了类别无关域专家（θ_S 和 θ_T）以及类别特定残差 $\tau_S^{(c)}$，我们现在着手解决如何适应新的目标域。我们的目标是迁移在这一新环境中完全未被观测到的类别知识。为此，我们提出一种基于权重空间旋转的闭式解，从而能够将类别特定微调直接从源模型传送到目标模型。假设可以访问源类别专家和目标类别专家的权重分解，我们可以将给定层 ℓ的权重定义如下：

$$\theta_{S,\ell}^{(c)} = U_{S,\ell}(\Sigma_{S,\ell} + \Delta\Sigma_{S,\ell}^{(c)})V_{S,\ell}^\top, \qquad \theta_{T,\ell}^{(c)} = U_{T,\ell}(\Sigma_{T,\ell} + \Delta\Sigma_{T,\ell}^{(c)})V_{T,\ell}^\top. \tag{4}$$

在此公式中，目标类别特定项 $\Delta\Sigma_{T,\ell}^{(c)}$ 表示未知量，因为目标域没有可用的类别特定数据。不过，我们拥有源类别专家 $\theta_{S,\ell}^{(c)}$ 以及目标模型的域特定组件——具体而言，是从目标域专家 θ_T中导出的奇异基（$U_{T,\ell}$，$V_{T,\ell}$）。我们利用这些可用组件，通过闭式解求得未知量。

闭式对齐。 为清晰起见，在后续公式中省略层索引 ℓ。我们的核心直觉是将源残差 $U_S\Delta\Sigma_S^{(c)}V_S^\top$旋转至目标域的谱域（$U_T$，$V_T$），从而得到以下关系：

$$U_S\,\Delta\Sigma_S^{(c)}\,V_S^\top \;=\; U_T\,\pi_{S\to T}(\Delta\Sigma_S^{(c)})\,V_T^\top. \tag{5}$$

为了将这一直觉形式化，我们必须确定一个满足公式 (5) 的变换 $\pi_{S\to T}(\cdot)$。我们将这种映射参数化为源残差的双线性变换：

$$\pi_{S\to T}(\Delta\Sigma_S^{(c)}) = L\,\Delta\Sigma_S^{(c)}\,R^\top, \quad \Rightarrow \quad U_S\,\Delta\Sigma_S^{(c)}\,V_S^\top \;=\; U_T\,L\,\Delta\Sigma_S^{(c)}\,R^\top\,V_T^\top, \tag{6}$$

其中，L和 R 表示两个旋转矩阵，旨在对齐来源域与目标域的奇异子空间。这些矩阵的精确值可通过**解析**求解以下优化问题来确定，从而确保公式 (5) 中的变换保持精确：

$$\min_L \|U_S - U_T L\| \quad \text{s.t. } L^\top L = I, \qquad \min_R \|V_S - V_T R\| \quad \text{s.t. } R^\top R = I. \tag{7}$$

这两种表述都体现了经典的**正交 Procrustes 问题**，其目标是求出最优旋转矩阵 L 和 R，以尽可能有效地对齐两个域中的奇异向量子空间。最优解可以直接通过闭式形式求得：

$$U_S = U_T L^* \;\Rightarrow\; U_T^\top U_S = L^*, \qquad V_S = V_T R^* \;\Rightarrow\; V_T^\top V_S = R^*. \tag{8}$$

算法 1 ABRA：用于适应的对齐基迁移

输入： 骨干网络权重 θ_0；源数据集 $\mathcal{D}_S$；目标数据集 $\mathcal{D}_T$；类别索引 c。
$\theta_S \leftarrow$微调(θ_0, $\tilde{\mathcal{D}}_S$) ▷ 通过对象化得到的源域专家
$U_S, \Sigma_S, V_S^\top \leftarrow$ 奇异值分解(θ_S)
$\Delta\Sigma_c \leftarrow$ SVFT(θ_S, $\mathcal{D}_S^{(c)} = \{(x_i,\ y_i = c)\}$) ▷ 在来源上训练类别专家
$\theta_T \leftarrow$微调(θ_0, $\tilde{\mathcal{D}}_T$) ▷ 目标域专家
$U_T, \Sigma_T, V_T^\top \leftarrow$ 奇异值分解(θ_T)
$(L, R) \leftarrow$ 普罗克鲁斯特斯(U_S, U_T), 普罗克鲁斯特斯(V_S, V_T)
$\Delta\Sigma_c^{(t)} \leftarrow L\,\Delta\Sigma_c\,R^\top$ ▷ 通过权重对齐实现传送
$\Sigma_T' \leftarrow \Sigma_T + \Delta\Sigma_c^{(t)}$
返回 $\theta_T^{(c)} \leftarrow U_T\Sigma_T'V_T^\top$ ▷ 类别适配的目标专家 $\theta_T^{(c)}$ 。

求得最优旋转 $L^* = U_T^\top U_S$ 和 $R^* = V_T^\top V_S$后，我们便可将类别特定残差从源域无缝迁移至目标域。由此可得目标域类别专家的最终形式：

$$\theta_{T,\ell}^{(c)} \approx U_T(\Sigma_{T,\ell} + \pi_{S\to T}(\Delta\Sigma_S^{(c)}))V_T^\top \approx U_T(\Sigma_T + U_T^\top U_S\Delta\Sigma_S^{(c)}V_S^\top V_T)V_T^\top \quad (9)$$

这些公式的完整推导和扩展讨论见补充材料。总而言之，所提出的基于旋转的对齐方法首先系统地构建一个经过旋转的类别特定残差，使其能够与目标专家无缝衔接。随后，我们将这一适配后的残差整合到目标域专家之上。最终得到一个能力出众的检测器，即使目标域中从未观测到特定类别，也能够识别这些类别。图2下半部分展示了这一完整流程，算法 1 对其进行了详细解析。

5 实验

数据集。 据我们所知，由于我们的设定引入了一种新颖的域迁移范式，目前尚无标准化基准。为填补这一空白，我们构建了一个基于两个不同数据集的评估框架。我们的主要基准聚焦于广泛采用的从 Cityscapes [5] 到 Foggy Cityscapes [27]的域偏移。原始 Cityscapes 数据集提供在晴朗天气下拍摄的高分辨率城市街景，而 Foggy Cityscapes（强度为 0.02）则提供了对应的合成雾场景。为了进一步验证我们的方法在更多样的环境变化下的表现，我们采用了 SDGOD 基准 [34], ，该基准涵盖五种不同的天气和光照条件：白天晴朗、白天有雾、黄昏有雨、夜间有雨和夜间晴朗。需要注意的是，我们使用的仅是这些数据集的原始数据，而非其标准验证协议。在

在我们的设定中，类别级评估严格仅针对至少包含一个目标类别真实标注实例的图像子集进行，所有其他边界框均被屏蔽；这与现有基准有着根本性的不同。

实现细节。 ABRA 和所有对比基线方法均采用 Grounding DINO 作为预训练的开放词汇检测骨干网络。领域检查点通过仅微调编码器注意力层 10 个 epoch 得到，学习率为 1e-4 ，批量大小为 $\mathcal{B}=2$。以源领域专家为起点，我们通过 SVFT 对类别进行专门化：以更高的学习率 1e-2 和 $\mathcal{B}=4$，对相同的注意力层额外训练 12 个 epoch。更多细节见补充材料。

5.1 基准评测

为严格评估我们的方法，我们将其与受近期任务算术进展启发的多种基线方法进行基准比较。性能以平均精度均值（mAP）和 AP50 表示；前者是对 10 个递增的交并比（IoU）阈值取平均，后者则在固定的 50% IoU 阈值下计算。以下将详细介绍这些方法。

微调（上界）。 为确定我们方法的理论上界，我们针对目标域中的每个特定类别训练一个专用的全量微调模型。从 θ_0出发，我们通过筛选目标域数据集，仅保留包含类别 c 的图像，从而提取给定目标类别 c 的所有实例，并主动屏蔽其中同时出现的任何非目标类别的边界框。仅对这些隔离出的实例进行标准优化后，所得专门化模型在类别 c 上达到预期的最高性能，因此可作为我们的经验上界。

Grounding DINO 零样本（基线）。冻结的预训练模型 θ_0。
来源（基线）。该基线直接在目标域上评估模型 $\theta_{FT}^{(c)}$ 。其中， $\theta_{FT}^{(c)}$ 是通过 SVFT，仅在源域的类别 c上对预训练模型 θ_0 进行微调而得到的。由于这些专家仅接触过该类别 c的源分布，因此这一设置量化了将一个在源域上训练的模型迁移到新的目标域、且不进行任何适应所导致的性能下降。

通过任务算术构建类比模型 [11]（竞争方法）。如第 2 节所述，任务算术 [11] 是一种强大的范式，能够即时构建新模型，无需样例或额外训练步骤。例如，它可以建立不同任务之间的**类比**，形式为“A 之于 B，如同 C 之于 D。”如果已知模型 B、 C和 D 的权重，该范式便可通过直接在权重空间中进行加法、减法等简单运算，估算模型 A 的权重。

受这一机制的启发，我们提出一种简洁的基于类比的迁移方法，作为具有竞争力的基线。具体而言，给定源域专家 θ_S、目标域专家 θ_T 和源

类别专家 $\theta_S^{(c)}$，我们的目标是利用以下类比来估计目标类别专家 $\theta_T^{(c)}$ 的权重：

$$\theta_T^{(c)} : \theta_T \;=\; \theta_S^{(c)} : \theta_S. \tag{10}$$

具体而言，正如原论文所倡导的，我们计算它们对应的任务向量，分别记为 $\tau_S^{(c)} = \theta_S^{(c)} - \theta_0$、$\tau_S = \theta_S - \theta_0$ 和 $\tau_T = \theta_T - \theta_0$。随后，按如下方式组合这些向量，得到目标类别模型：

$$\theta_T^{(c)} = \theta_0 + \tau_T + \tau_S^{(c)} - \tau_S. \tag{11}$$

直观地说，以目标域专家为起点，这一操作会添加残差信息，用以区分源域中的类别专家与域专家。换言之，我们将源域中学到的类别特定偏差迁移到目标表征上。这个算术基线提供了一个轻量的参考，用于评估仅靠线性组合是否足以实现跨域类别迁移。

参数Δ [3] 用于直接权重混合（竞争方法）。近期针对大语言模型（LLMs）的研究表明，可以通过直接传送任务向量，在不同模型之间迁移微调知识。给定一个基础模型 θ_A 及其微调对应模型 θ_A^*，参数 Δ [3] 会计算相应的任务向量 $\tau_A = \theta_A^* - \theta_A$，并将其注入另一个具有相同架构的模型 θ_B，从而得到 $\theta_B^* = \theta_B + \tau_A$。这种直接在权重空间中进行的操作，无需额外训练即可在模型之间迁移任务特定更新。在我们的设定中，如果将模型 A和 B 分别视为源域专家 θ_S 和目标域专家 θ_T，那么参数 Δ自然就成为我们方法的另一种竞争方法。具体而言，这相当于对从 S 到 T的映射 $\pi_{S\to T}(\cdot)$ 使用恒等函数，这意味着迁移类别特定信息时不进行旋转或对齐。形式上，参数 Δ 可简化为：

$$\widehat{\Delta\Sigma}_T^{(c)} = \Delta\Sigma_S^{(c)}, \Longrightarrow \theta_{T,\ell}^{(c)} \approx U_T(\Sigma_T + \Delta\Sigma_S^{(c)})V_T^\top, \tag{12}$$

在目标域中直接复用源残差。

5.2 主要结果

Cityscapes→Foggy Cityscapes. 表 1 中的结果凸显了 ABRA 在所有迁移类别上的有效性。所有竞争方法都面临显著困难，无法达到来源 基线的性能。具体而言，任务类比 [11] 的表现低于零样本基线，而 ParamΔ [3] 相较于零样本仅带来微小提升。相比之下，**ABRA** 轻松超越了这两个基线和其他竞争方法，性能接近由完整微调所确立的理论上限。这些结果表明，我们的方法能够有效地将类别级专门化迁移至目标域。

表1：Cityscapes → Foggy Cityscapes。不可用类别上平均精度均值与AP_{50}的比较。ABRA优于所有基线方法。最佳结果以**粗体**标出。

方法	Bus		摩托车		骑行者		火车		卡车		平均值	
	平均精度均值	AP_{50}	平均精度均值	AP_{50}	平均精度均值	AP_{50}	平均精度均值	AP_{50}	平均精度均值	AP_{50}	平均精度均值	AP_{50}
微调	58.75	73.50	31.22	55.01	43.87	68.06	32.95	58.75	40.03	57.08	41.36	62.48
零样本	48.63	61.10	23.20	41.91	18.96	31.70	16.31	44.02	31.20	41.88	27.66	44.12
来源	54.77	66.93	29.62	50.96	40.25	61.99	29.40	55.30	37.23	51.54	38.25	57.34
任务类比 [11]	41.14	48.21	10.24	20.32	9.35	16.75	10.10	24.00	19.77	24.70	18.12	26.79
参数Δ [3]	50.23	60.41	20.59	40.59	18.53	31.44	21.70	49.42	30.42	40.26	28.29	44.42
ABRA（我们的方法）	**57.24**											

表 2：SDGOD。在具有挑战性的条件划分下对比平均精度均值和 AP_{50} 。ABRA 在所有基线方法中的总体平均值最高。最佳结果以**粗体**显示。

方法	白天雾天		黄昏雨天		夜间晴天		夜间雨天		平均值	
	mAP	AP_{50}	mAP	AP_{50}	mAP	AP_{50}	mAP	AP_{50}	mAP	AP_{50}
微调	36.37	57.83	26.77	50.65	36.86	69.40	16.81	29.85	29.20	51.93
零样本	26.36	41.10	19.55	34.93	27.50	49.63	9.19	13.60	20.65	34.82
来源	31.75	51.41	27.63	50.34	**36.38**	**66.70**	15.28	27.52	27.76	48.99
任务类比 [11]	26.49	41.27	18.97	34.69	27.38	49.44	9.61	14.27	20.61	34.92
参数Δ [3]	17.68	26.68	5.07	8.09	4.86	8.01	7.86	12.60	8.87	13.85
ABRA（我们的方法）	**32.35**	**53.81**	**27.99**	**51.39**	35.94	66.11	**16.13**	**30.97**	**28.10**	**50.57**

SDGOD 划分。此外，我们还在 SDGOD 数据集划分上评估了所提方法，结果见表 2。具有挑战性的不利条件，尤其是夜间雨天，导致零样本 基线的性能大幅下降。与此前的观察一致，现有任务向量方法无法有效适应：任务类比 [11] 的表现停留在零样本水平，而 ParamΔ [3]在所有划分上的性能都严重崩溃。相反，**SDGOD** 表现出很强的鲁棒性，整体超越来源 基线，并在白天雾天、黄昏雨天和夜间雨天条件下取得最高性能。尽管来源 基线在夜间晴天划分上的表现略胜一筹，**SDGOD** 仍然极具竞争力，且取得了最高的总体平均精度均值，与微调 上限十分接近。

5.3 补充分析

领域设计选择。为了验证域专家背后的设计选择，我们开展了一项消融研究，将采用所提出的对象化策略训练的模型与两种替代方案进行比较。首先，我们通过将对象化生成的边界框与 θ_0（Grounding DINO 的预训练版本）以零样本方式生成的边界框相结合，建立初始基线（对象化零样本）。其次，我们使用监督类别的实际语义标签对网络进行微调，而不是使用统一的通用标签（监督）。如表 3 所示，保留原始接地——

表 3：领域专家消融。评估对象化和类别可见性策略。ABRA 在未见类别上取得最高性能。

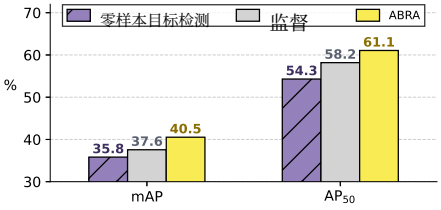

表 4：初始化的影响。使用 ABRA 初始化下游任务（FFT 或 FDA [35]）的表现始终优于标准 θ_0 预训练。

方法	初始化。	mAP	AP_{50}
FFT	θ_0	41.36	62.48
FFT	ABRA4		
FDA	θ_0	38.25	57.80
FDA	ABRA	**40.74**	(

真值语义以及依赖零样本预测，天然限制了向不可用类别的迁移能力。最终，对象化取得了最高的整体性能，证实隔离类别无关视觉线索对于有效的零样本跨域适应至关重要。

用于下游任务的 ABRA。接下来，我们将证明 ABRA 是下游优化的更优基础。为验证这一点，我们使用可迁移的类别特定残差来初始化两种不同策略：有监督全量微调（FFT）和无监督域适应；对于后者，我们采用广受认可的傅里叶域适应（FDA） [35] 方法。如表 4 所示，通过 ABRA 初始化的权重在这两种方法中，性能均持续优于标准预训练骨干网络 θ_0 ，由此为目标特定学习奠定了更为稳健的基础。

专门化的效果。为展示我们按类别定制方法的灵活性和有效性，我们评估了一种基线配置：训练单个统一专家，使其同时理解并迁移所有目标类别（记为“合并”）。如图 4 所示，即使没有类别级专门化，我们的框架仍能正常运行，并在一定程度上适应目标域。然而，为每个类别配备一个独立的轻量专家，始终

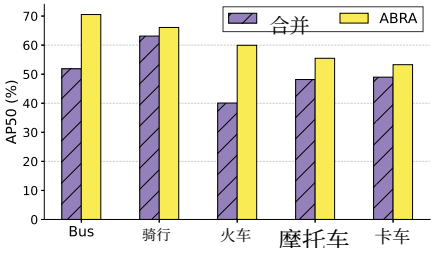

图 4：专门化对 Cityscapes 的影响。每个类别使用一个专家（ABRA）比使用单个统一专家（合并）取得更高的 AP_{50} 。

在所有类别中都取得了更高的 AP_{50} 分数。这种隔离的优势在复杂类别中尤为明显；例如，虽然 Rider 等类别的性能差距较为有限，但对于 Train 和 Bus，专门化方法带来了显著提升。这证实了，尽管我们的框架能够使用合并的专家运行，但为各个类别分配紧凑残差仍是最优策略。

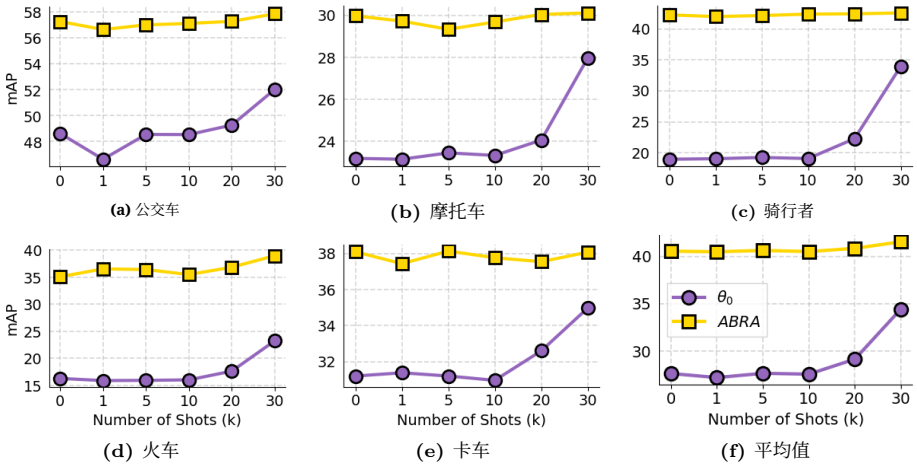

图 5：少样本迁移评估。标准预训练骨干网络（θ_0）与 ABRA 在不同样本数下的 Cityscapes 对比（平均精度均值）。

少样本迁移评估。最后，为了进一步研究我们的方法在迁移学习方面的能力，我们在少样本设置下评估其性能，分别以标准预训练检查点（θ_0）和我们提出的方法作为起点。我们重点考察使用 1、5、10、20 和 30 个样本适应目标域类别的效果。结果如图 5 所示，其中展示了特定类别的平均精度均值以及总体平均值表现。如图所示，在所有少样本场景和所有评估类别中，ABRA 始终优于标准 θ_0 初始化。随着样本数量增加，两种方法的检测性能都逐步提升。然而，在测试的每一个增量阶段，ABRA 都保持着相对于 θ_0 的显著性能优势。图 5f 中的平均精度均值结果清楚表明，在整个少样本学习过程中，ABRA 都呈现出更优的性能提升轨迹。

6 结论

我们提出了**用于适应的对齐基迁移（ABRA）**，这是一种用于跨域知识迁移的新型框架，旨在应对无法获取任何目标域数据这一具有挑战性的场景。我们的方法基于对象化 的概念来构建域专家，并利用 SVFT 实现专门化，从而在源域中高效编码类别知识。随后，我们提出了一种基于奇异值分解和旋转对齐的闭式传输机制，无需额外训练，即可将类别特定信息迁移至目标域。在多种域偏移场景下进行的全面实验表明，ABRA 在多种不利条件下均能实现强劲且稳定的迁移性能。展望未来，我们希望开发更稳定、更具表达力的域表征，以进一步提升模型在更极端域偏移场景下的鲁棒性。

参考文献

1. Ainsworth, S.K., Hayase, J., Srinivasa, S.：Git re-basin：在置换对称性下合并模型。国际学习表征会议研讨会（2023） 2. Bansal, Y., Nakkiran, P., Barak, B.：重新审视模型拼接以比较神经表征。神经信息处理系统（2021） 3. Cao, S., Wu, M., Prasad, K., Tian, Y., Liu, Z.：用于直接权重混合的参数差分：零成本后训练大语言模型。国际学习表征会议研讨会（2025） 4. Cappellino, C., Mancusi, G., Mosconi, M., Porrello, A., Calderara, S., Cucchiara, R.：Dithub：用于增量开放词汇目标检测的模块化框架。神经信息处理系统（2025） 5. Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., Schiele, B.：《用于语义城市场景理解的 Cityscapes 数据集》。载于：IEEE计算机视觉与模式识别会议论文集（2016） 6. Deng, J., Zhang, H., Ding, K., Hu, J., Zhang, X., Wang, Y.：面向视觉语言目标检测的零样本可泛化增量学习。载于：神经信息处理系统（2024） 7. Deng, J., Li, W., Duan, L.：用于无源目标检测的平衡教师。IEEE 视频技术电路与系统汇刊（2024） 8. Gu, X., Lin, T.Y., Kuo, W., Cui, Y.：通过视觉与语言知识蒸馏实现开放词汇目标检测。载于：国际学习表征会议研讨会（2022） 9. Hernandez, A., Dangovski, R., Lu, P.Y., Soljacic, M.：模型拼接：探寻表征之间的功能相似性。载于：神经信息处理系统研讨会（2022） 10. Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W., 等：LoRA：大语言模型的低秩自适应。国际学习表征会议研讨会（2022） 11. Ilharco, G., Ribeiro, M.T., Wortsman, M., Schmidt, L., Hajishirzi, H., Farhadi, A.：使用任务算术编辑模型。载于：国际学习表征会议研讨会（2023） 12. Kennerley, M., Wang, J.G., Veeravalli, B., Tan, R.T.：CAT：利用类间动态进行域自适应目标检测。载于：IEEE计算机视觉与模式识别会议论文集（2024） 13. Kopiczko, D.J., Blankevoort, T., Asano, Y.M.：VeRA：基于向量的随机矩阵适配。载于：国际学习表征会议研讨会（2024） 14. Li, L.H., Zhang, P., Zhang, H., Yang, J., Li, C., Zhong, Y., Wang, L., Yuan, L., Zhang, L., Hwang, J.N., 等：基于定位的语言图像预训练。载于：IEEE计算机视觉与模式识别会议论文集（2022） 15. Li, Y., Han, S., Ji, S.：VB-LoRA：利用向量库实现极致参数高效微调。神经信息处理系统（2024） 16. Lingam, V.C., Neerkaje, A., Vavre, A., Shetty, A., Gudur, G.K., Ghosh, J., Choi, E., Dimakis, A., Bojchevski, A., Sanghavi, S.：SVFT：使用奇异向量进行参数高效微调。神经信息处理系统（2024）

17. Liu, S.Y., Wang, C.Y., Yin, H., Molchanov, P., Wang, Y.C.F., Cheng, K.T., Chen, M.H.：Dora：权重分解低秩适配。在：国际机器学习会议（2024） 18. Liu, S., Zeng, Z., Ren, T., Li, F., Zhang, H., Yang, J., Jiang, Q., Li, C., Yang, J., Su, H. 等：Grounding DINO：将 DINO 与基础定位预训练相结合，实现开放集目标检测。在：欧洲计算机视觉会议论文集（2024） 19. Mattolin, G., Zanella, L., Ricci, E., Wang, Y.：Confmix：通过基于置信度的混合实现无监督域适应目标检测。在：冬季计算机视觉应用会议（2023） 20. Meng, F., Wang, Z., Zhang, M.：Pissa：大语言模型的主奇异值与奇异向量适配。神经信息处理系统（2024） 21. Minderer, M., Gritsenko, A., Stone, A., Neumann, M., Weissenborn, D., Dosovitskiy, A., Mahendran, A., Arnab, A., Dehghani, M., Shen, Z. 等：简单的开放词汇目标检测。在：欧洲计算机视觉会议论文集（2022） 22. Ortiz-Jimenez, G., Favero, A., Frossard, P.：切空间中的任务算术：改进预训练模型的编辑。神经信息处理系统（2023） 23. Pan, Z., Cai, J., Zhuang, B.：可拼接神经网络。在：IEEE计算机视觉与模式识别会议论文集（2023） 24. Pfeiffer, J., Ruder, S., Vulić, I., Ponti, E.：模块化深度学习。机器学习研究汇刊（2023） 25. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J. 等：从自然语言监督中学习可迁移视觉模型。在：国际机器学习会议（2021） 26. Rinaldi, F., Capitani, G., Bonicelli, L., Crisostomi, D., Bolelli, F., Ficarra, E., Rodola, E., Calderara, S., Porrello, A.：将变换器更新至最新版本：任务向量的重新基化。国际机器学习会议（2025） 27. Sakaridis, C., Dai, D., Van Gool, L.：利用合成数据进行语义雾天场景理解。国际计算机视觉杂志（2018） 28. Sangineto, E., Nabi, M., Culibrk, D., Sebe, N.：用于弱监督目标检测的自步深度学习。IEEE模式分析与机器智能汇刊（2019） 29. Stoica, G., Bolya, D., Bjorner, J.B., Ramesh, P., Hearn, T., Hoffman, J.：Zipit! 无需训练即可合并来自不同任务的模型。在：国际学习表征会议研讨会（2024） 30. Tang, S., Yang, J., Ye, M., Wang, B., Gan, Y., Zhu, X.：通过语义补偿实现无源域自适应目标检测。arXiv预印本 arXiv:2410.05557（2025） 31. Tarvainen, A., Valpola, H.：均值教师是更好的榜样：权重平均一致性目标改善半监督深度学习结果。在：神经信息处理系统（2017） 32. Van de Ven, G.M., Tuytelaars, T., Tolias, A.S.：增量学习的三种类型。自然·机器智能（2022） 33. Wu, A., Deng, C.：通过循环解耦自蒸馏实现城市场景中的单域泛化目标检测。在：IEEE计算机视觉与模式识别会议论文集（2022）

34. Wu, A., Deng, C.：通过循环解耦自蒸馏实现城市场景中的单域泛化目标检测。在：IEEE计算机视觉与模式识别会议论文集（2022） 35. Yang, Y., Soatto, S.：FDA：用于语义分割的傅里叶域适应。在：IEEE计算机视觉与模式识别会议论文集（2020） 36. Yuan, Y., Tang, L., Chen, Y., Chen, C., Huang, Y., Ding, X.：ASGS：通过自适应子图搜索实现单域可泛化开放集目标检测。在：IEEE国际计算机视觉会议（2025） 37. Zang, Y., Li, W., Zhou, K., Huang, C., Loy, C.C.：带条件匹配的开放词汇 DETR。在：欧洲计算机视觉会议论文集（2022） 38. Zhang, H., Cisse, M., Dauphin, Y.N., Lopez-Paz, D.：MixUp：超越经验风险最小化。在：国际学习表征会议研讨会（2018） 39. Zhang, R., Chen, T., Shen, Y., Carin, L., Wang, W., Chen, W.：预训练模型的自适应低秩适配。在：国际学习表征会议研讨会（2023） 40. Zhong, Y., Yang, J., Zhang, P., Li, C., Codella, N., Li, L.H., Zhou, L., Dai, X., Yuan, L., Li, Y. 等：RegionCLIP：基于区域的语言-图像预训练。在：IEEE计算机视觉与模式识别会议论文集（2022） 41. Zhou, X., Girdhar, R., Joulin, A., Krähenbühl, P., Misra, I.：利用图像级监督检测两万类。在：欧洲计算机视觉会议论文集（2022）

ABRA：跨域微调知识传送，用于开放词汇目标检测

补充材料

A 奇异子空间的正交对齐

给定源奇异子空间和目标奇异子空间

$$U_S, U_T \in \mathbb{R}^{n\times k}, \qquad U_S^\top U_S = U_T^\top U_T = I_k,$$

我们的目标是通过作用于降维后的 k 维空间中的正交变换 $L \in \mathbb{R}^{k\times k}$ 对它们进行对齐。我们寻求能够将目标基最佳映射到源基的正交矩阵：

$$\min_{L^\top L = I_k} \|U_S - U_T L\|_F^2. \tag{13}$$

展开弗罗贝尼乌斯范数并利用 $U_S^\top U_S = U_T^\top U_T = I_k$，可得标准化简：

$$\|U_S - U_T L\|_F^2 = 2k - 2\,\mathrm{tr}(U_S^\top U_T L). \tag{14}$$

由于 $2k$ 为常数，求解(13)等价于

$$\max_{L^\top L = I_k} \mathrm{tr}(L^\top M), \qquad M := U_T^\top U_S. \tag{15}$$

设 $M = U\Sigma V^\top$ 为奇异值分解。正交普罗克鲁斯特解直接给出

$$L^\star = UV^\top. \tag{16}$$

我们对齐的使用。 我们使用 $L^\star$ 将目标奇异基旋转到源空间：

$$\tilde{U}_T = U_T L^\star,$$

从而确保两个子空间在最小二乘意义上最优对齐，同时保持几何结构（正交性、尺度）。这一对齐后的基正是我们输入适配模块的内容。

特殊情形。 若 U_T 与 U_S 张成相同子空间，则 M 已是正交的，解简化为：

$$L^\star = U_T^\top U_S.$$

这只是将普罗克鲁斯特解应用于一对已经完全匹配的子空间。

B 实现细节

ABRA 和所比较的所有基线方法都将 Grounding DINO 作为预训练的开放词汇检测骨干网络。领域检查点通过仅对编码器注意力层进行 10 个 epoch 的微调得到，采用 1e-4 的标准学习率、 $\mathcal{B} = 2$ 的批量大小和 1e-4 的权重衰减。从来源域专家出发，我们通过 SVFT 对类别进行专门化：将这些相同的注意力层额外训练 12 个 epoch，并将学习率提高至 1e-2 and $\mathcal{B} = 4$。对于这些类别专家，我们不使用权重衰减，并采用 SVFT 的带状模式，其非对角值（off_diag）为 2。在域专家和类别专家的两个训练阶段中，我们都使用多步学习率调度器，在第 7 和第 9 个 epoch 将学习率降低 10 倍。此外，特定层 ref_point_head 和 sampling_offset 在两个阶段的初始学习率始终设为标准学习率的十分之一。

C 定性结果

在本节中，我们提供了更多定性结果，以进一步证明所提出方法 ABRA 的有效性。表 Tab. 5 展示了仅来源基线、我们的方法与真值（GT）在目标域中具有挑战性的样例上的视觉对比，这些样例具有严重雾气和能见度降低的特点。如最左列所示，基线模型在这些恶劣天气条件下表现明显不佳。它无法检测或正确分类困难且严重遮挡的物体，完全漏检了公交车（第一行）、骑行者（中间行）和火车（最后一行）。相反，我们提出的方法（中间列）成功定位并分类了这些具有挑战性的实例。尽管存在浓雾和域偏移，ABRA 生成的边界框预测与真值标注（最右列）高度吻合。

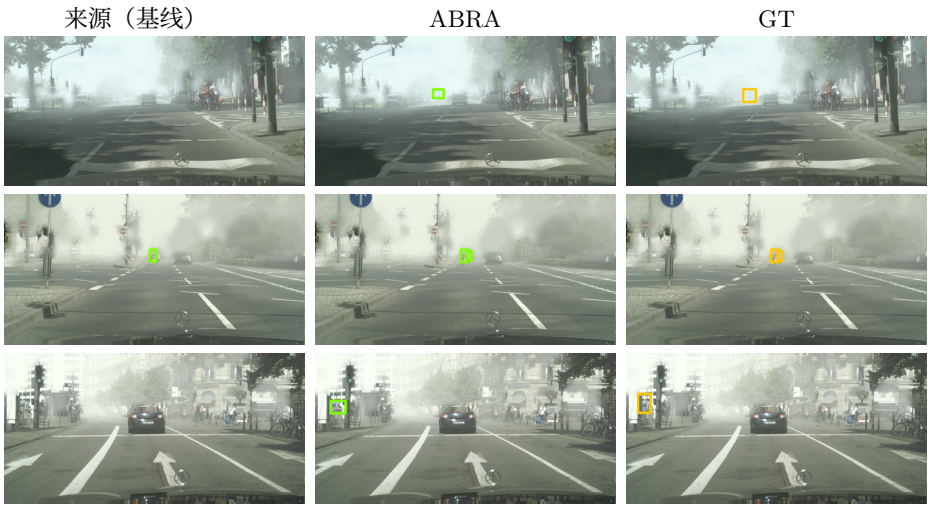

表 5： 从上到下：公交车、骑行者、火车。从左到右：来源（基线）、ABRA、真值