arXiv:2303.04989v3 [cs.CV] 7 Apr 2024

ARS-DETR：面向航空图像有向目标检测的长宽比敏感检测变换器

曾颖[1], 陈宇石[1], *IEEE* 会员，Xue Yang[2], 李庆云[1], 严俊驰[2],*IEEE* 高级会员

摘要—近年来，航空图像中的定向目标检测取得了长足进展，并获得了良好的成果。然而，航空图像中的高精度定向目标检测仍然是一项具有挑战性的任务。近期一些研究采用基于分类的方法预测角度，以解决角度中的边界问题。但我们发现，这些研究往往忽视了不同长宽比的目标对角度的敏感性。同时，探索合适的方法改进新兴的基于变换器的方法，使其适用于定向目标检测，也值得关注。本文提出一种用于航空图像定向目标检测的长宽比敏感检测变换器，称为 ARS-DETR。具体而言，我们提出一种新的角度分类方法——长宽比感知圆形平滑标签（AR-CSL），以更合理的方式平滑角度标签，并舍弃先前工作（例如 CSL）引入的超参数。随后，我们设计了一种旋转可变形注意力模块，根据对应角度旋转采样点，从而消除区域特征与采样点之间的未对齐。此外，我们采用根据长宽比设定的动态权重系数来计算角度损失。在若干具有挑战性的数据集上进行的全面实验表明，我们的方法在高精度定向目标检测任务中取得了具有竞争力的表现。

索引词—**定向目标检测、高精度检测、检测变换器、特征对齐。**

I. 引言

航空图像中的目标检测一直是遥感领域的研究热点。随着大量可用的高分辨率航空图像迅速增加 [1]–[3],，如何准确有效地检测这些航空图像中的目标已成为一个关键问题。

得益于深度学习的发展，深度卷积神经网络（CNNs）的出现极大地影响了检测器的设计，并在通用目标检测中取得了良好的性能。与使用手工设计特征进行检测相比，后者过程繁琐且准确性不足，CNNs能够从训练数据中学习并进行迭代更新，展现出强大的能力以提取高级鲁棒特征，实现精准检测。人们还提出了许多先进的检测器，使用水平边界框（HBBs）来检测目标。

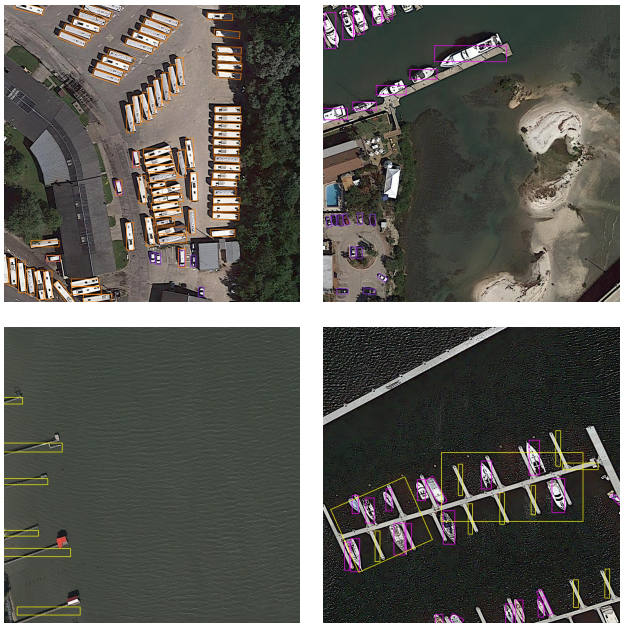


图1. 尽管角度预测不准确，但其 AP_{50} 仍然表现出色。

与通用图像相比，航空图像中的目标通常具有多种尺度、长宽比和方向，有时还会密集排列。仅使用水平边界框检测这些目标时，水平边界框往往无法很好地贴合目标，因此会包含大量背景或与其他目标重叠。因此，采用定向边界框（OBB）表示目标的定向目标检测，更适合航空目标检测。

作为遥感领域新近兴起的一项任务，定向目标检测在分析航空图像中的目标方面展现出强大的能力，许多先进的定向目标检测器已被提出，并在航空图像中取得了良好的性能 [4]–[7]。然而，许多检测器将定向目标检测视为涉及角度预测的通用目标检测。为此，它们只是在检测头中增加了一个角度参数。因此，如图1所示，这些检测器的角度预测并不十分准确。尽管如此，在当前的评价指标（即 AP_{50}，）下，这些检测器仍能取得相当不错的结果，这表明 AP_{50} 不足以准确反映定向目标检测器的性能

本工作得到中国国家自然科学基金项目（62371169、61971164和U20B2041）的资助（通信作者：陈宇石）。

曾颖、陈宇石和李庆云现供职于哈尔滨工业大学电子与信息工程学院，中国哈尔滨150001（电子邮箱：22S005064@stu.hit.edu.cn；chenyushi@hit.edu.cn；21B905003@stu.hit.edu.cn）

Xue Yang 就职于 OpenGVLab、上海人工智能实验室，地址：中国上海市 200030（电子邮箱：yangxue-2019sjtu@sjtu.edu.cn）

严俊驰就职于上海交通大学电子信息与电气工程学院、教育部人工智能重点实验室，中国上海 200030（电子邮箱：yanjunchi@sjtu.edu.cn）

，而航空图像中的高精度定向目标检测仍是一项具有挑战性的任务。

角度作为定向目标检测中的独特参数，在高精度检测中发挥着至关重要的作用。同时，角度的特性也使其难以预测，因此需要给予更多关注。首先，角度的周期性会导致边界处的不连续，从而使训练过程中的优化结果欠佳。其次，具有不同长宽比的目标对角度表现出不同的敏感性，而大多数定向目标检测器都忽略了这一点。长宽比较小的目标，尤其是近似正方形的目标，对角度变化的敏感性较低。因此，即使预测角度之间存在明显差异，即角度偏差较大，它们仍能够在目标与预测结果之间保持较高的倾斜交并比（SkewIoU）。相比之下，当目标具有较大的长宽比时，即使轻微的角度偏差也可能导致SkewIoU大幅下降。

在众多角度预测方法中，基于分类的方法表现出良好的性能[8]–[10]。具体而言，该方法将角度与边界框解耦，并将角度预测转化为分类任务，从而消除了边界问题 [11]。此外， [12]也展现了基于分类的方法在高精度定向目标检测中的巨大潜力。然而，这种方法仍存在一些问题，例如完全忽略角度与长宽比之间的相关性、引入超参数（如 [8] 中的窗口半径）等。因此，角度预测的准确性在一定程度上受到限制。

近年来，基于变换器的检测器 [13], [14] 重新激发了目标检测领域的研究热情。DEtection TRansformer（DETR [13]）无需额外复杂的手工设计组件（如预设锚框或非极大值抑制（NMS）），而是将目标检测视为集合预测任务，并通过二分图匹配分配标签，其性能可与 Faster RCNN 等经典检测器相媲美 [15]。现有的 DETR 衍生模型 [14],[16]–[20] 大幅提升了检测性能和收敛速度，展现出变换器在高精度目标检测中的巨大潜力。尽管已有一些基于 DETR 的定向目标检测方法 [21], [22] 被提出，但它们仍采用回归来预测角度，且没有考虑边界不连续所引发的问题。同时，它们以简单的方式预测角度，并未探索如何将角度信息嵌入 DETR。如何更自然地将 DETR 应用于定向目标检测，仍是一个研究课题。

本文提出了一种长宽比敏感检测变换器，用于实现航空图像中的定向目标检测，称为 ARS-DETR。具体而言，我们设计了一种无超参数的宽高比感知圆形平滑标签（AR-CSL），根据目标的长宽比表示相邻角度之间的关系。考虑到不同目标对角度的敏感性，AR-CSL 利用不同长宽比目标在各角度偏差下的 SkewIoU，对角度标签进行平滑处理。随后，我们还提出了一种旋转可变形注意力模块，将角度信息嵌入检测器中，以对齐特征。最后，我们采用长宽比敏感匹配和损失策略，实现检测器训练的动态调整，从而减轻模型训练负担。在不同航空数据集上的大量实验表明，ARS-DETR 在高精度定向目标检测方面具有良好效果。总而言之，我们的贡献主要包括以下四个方面：

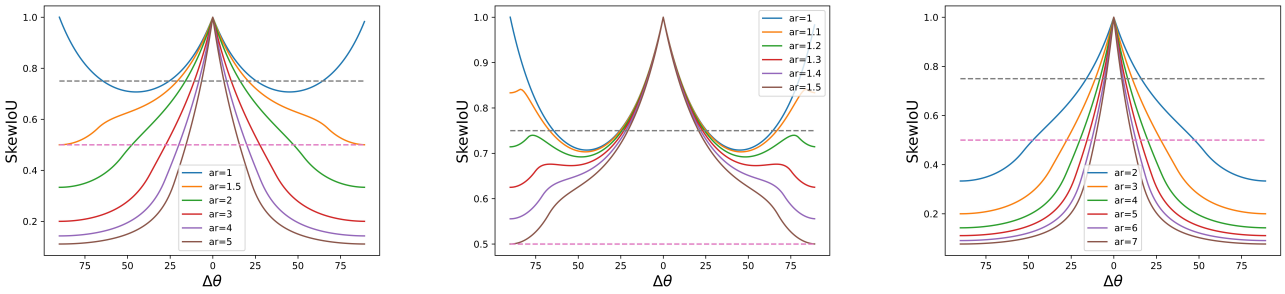
- 我们详细分析了角度偏差对定向目标检测的影响，并给出了相应公式。此外，我们还分析了当前定向目标检测评价指标（即 AP_{50}）的缺陷。
- 我们设计了一种名为 AR-CSL 的新型角度分类方法，以更合理的方式平滑角度标签。该方法采用不同角度偏差下、具有不同长宽比的目标的 SkewIoU 值，同时消除了先前工作引入的窗口半径超参数。
- 我们提出了一种嵌入角度信息的旋转可变形注意力模块（RDA），以利用角度信息提取对齐特征。同时，我们提出了长宽比敏感匹配（ARM）和长宽比敏感损失（ARL），根据目标的长宽比自适应地调整对角度的关注程度。此外，我们还结合去噪策略（DN），进一步提升基于DETR的方法在定向目标检测中的性能。
- 在三个公开航空数据集 DOTA-v1.0、DIOR-R 和 OHD-SJTU 上进行的大量实验，证明了所提模型的有效性。ARS-DETR 在所有数据集上均实现了具有竞争力的高精度定向目标检测性能。

II. 相关工作

A. 定向目标检测

作为一项新兴任务，定向目标检测近年来取得了长足进展。定向目标检测任务的一种简单解决方案是将锚框或感兴趣区域（RoIs）从水平类型改为旋转类型。RoI-Transformer [4] 构建几何变换，将候选框旋转，从而更准确地定位目标。为解决精细化单阶段检测器中的特征错位问题，R^3Det [24] 和 S^2A-Net [25] 采用特征对齐模块来获得更准确的位置。然而，这些主流的基于回归的方法常因预测结果超出定义范围而面临边界问题 [8] ，并且需要额外的复杂处理。SCRDet [5] 设计了一种新颖的 IoU-Smooth L1损失，以缓解角度周期性和边缘可交换性导致的损失骤增，从而降低模型训练的难度。

CSL [8]将角度预测从回归转为分类，从而消除了边界问题。这是通过设计圆形平滑标签实现的。滑动顶点 [6]将水平边界框的顶点滑动，以准确表示多方向目标。GWD [11],KLD [26] 和 KFIoU [27] 将旋转边界框



(a) $k \geq 1$ (b) $1 \leq k \leq 1.5$ (c) $k > 1.5$

图2. 曲线表示不同长宽比下 SkewIoU 与角度偏差 $\Delta\theta$ 之间的关系。 k 表示长宽比。

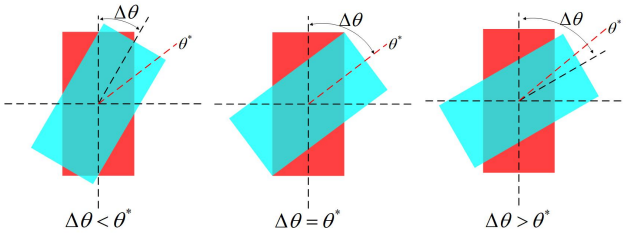
转换为高斯分布，以避免定向目标检测中的边界不连续问题和类方形问题。PSC [28]通过将不同周期的旋转周期性映射到不同频率的相位，为定向目标检测中的各种周期性模糊问题提供了统一框架。

B. 基于角度分类的有向目标检测

基于分类的角度预测方法是一种新颖且有效的方法，能够在准确预测角度的同时规避边界问题，并且也取得了长足进展。CSL [8] 将角度变量离散为 180 个类别，并通过高斯窗口函数平滑角度标签。CSL 方法直接推动了基于分类的定向目标检测算法的发展。DCL [9] 采用密集编码标签，以减少 CSL 的计算量和参数量。 [10] 采用基于 CSL 的动态加权机制，对旋转目标进行精确的角度估计。为克服角度表示中的歧义性和高昂成本， [29]提出了一种多粒度角度表示方法，由粗粒度角度分类和细粒度角度回归组成。TIOE [12] 提出了一种渐进式方向估计策略，通过 n 元编码来逼近目标的方向。AR-BCL [30] 采用基于长宽比的双向编码标签来解决类方形检测问题 [9]。相比之下，本文提出的新型角度编码技术不需要超参数，也不存在边界问题和类方形问题。此外，该技术还探索了角度分类在高精度检测中的潜力，而上述大多数方法都忽视了这一点。

*C. DETR*及其变体

DETR [13] 提出了一种基于 Transformer 的端到端目标检测器，无需使用先验锚框设计和 NMS 等手工设计组件。近年来，DETR取得了很大进展，与经典检测方法相比，也展现出强大的目标检测能力 [14], [16]–[20]。Deformable DETR [14] 提出了一种可变形注意力模块，用于对参考点周围自适应位置的值进行采样，并利用多层级特征来缓解 DETR 收敛缓慢和高复杂度的问题。



(a) 情况 1 (b) 边界条件 (c) 情况 2

图 3. SkewIoU 计算的两种情况。(a) $\Delta\theta < \theta^*$ 所在的情况；(b) 情况 1 与情况 2 之间的边界条件；(c) $\Delta\theta > \theta^*$ 所在的情况。

DAB-DETR [20]为每个查询提供明确的位置先验，通过使用锚框尺寸，让交叉注意力模块聚焦于与目标对象对应的局部区域。DN-DETR [17]和 DINO [18] 设计了一种绕过二分图匹配的去噪辅助任务。这不仅加快了训练收敛，还带来了更好的训练结果。此外，已有一些基于DETR的旋转目标检测器 [21], [22]。O^2DETR [21] 首次尝试将 DETR 应用于定向目标检测任务，而 AO2-DETR [22] 则将旋转候选框生成和精炼模块引入 Transformer架构，以精炼特征。然而，两者都采用简单的回归方式预测角度，且未解决边界不连续性问题，也未将角度信息融入 DETR。

III. 对定向目标检测的重新思考

本节分析角度与长宽比之间的关系。此外，我们还分析了当前定向目标检测中所用评价指标 AP_{50} 的不足，并强调高精度定向目标检测的重要性。

A. 角度与长宽比

长宽比不同的目标对角度的敏感度也不同。为了更好地观察角度与长宽比之间的关系，我们假设有两个中心、高度和宽度相同的边界框，并给出这两个长宽比不同的边界框在不同角度偏差下的 SkewIoU，

如图2所示。SkewIoU 可通过公式 1 计算：

$$\text{SkewIoU}(k,\Delta\theta)=\begin{cases}\frac{4k\tan\Delta\theta-m-n}{4k\tan\Delta\theta+m+n} & \Delta\theta\le 2\arctan\frac{1}{k}\\ \frac{1}{2k\sin\Delta\theta-1} & \Delta\theta>2\arctan\frac{1}{k}\end{cases},$$
$$m=(1-k\tan\frac{\Delta\theta}{2})^2\tan^2\Delta\theta,$$
$$n=(\frac{-2\sin^2\frac{\Delta\theta}{2}+k\sin\Delta\theta}{\cos\Delta\theta})^2, \tag{1}$$

其中，$k\ge 1$ 表示长宽比，$\Delta\theta\in[0°,90°]$表示角度偏差，即两个边界框之间角度差的绝对值。如图3所示，存在一个临界角边界阈值（$\theta^*=2\arctan\frac{1}{k}$）。图2所示的对称性表明，$\Delta\theta$ 的变化具有双向特征。

图2(a)展示了不同长宽比下 SkewIoU 与角度偏差之间的曲线。可以看出，根据评价指标 AP_{50} （如果预测框与真实标注之间的 SkewIoU 大于 0.5，则将其判定为真正例），不同长宽比边界框的 SkewIoU 变化趋势明显分为两类，分界线为 $k=1.5$，分别如图2(b)（$1\le k\le 1.5$）和图2(c)（$k>1.5$）所示。具体而言，图2(b)表明，当长宽比小于 1.5 时，无论角度偏差如何，SkewIoU 始终大于 0.5（见粉色虚线）。相比之下，如图2(c)所示，当长宽比大于 1.5 时，SkewIoU 会随着角度偏差增大而迅速下降，但有效角度偏差仍然保持较大的范围。总而言之，长宽比较小的目标对角度偏差不太敏感，而长宽比较大的目标则更为敏感，但在 AP_{50} 下仍表现出显著的角度偏差容忍度。

B. 高精度定向目标检测

鉴于角度是定向目标检测中的重要参数，其估计准确度将极大影响后续相关任务，例如目标细粒度识别 [31]，目标航向估计 [32],[33],等，因此，AP_{50} 似乎不足以准确反映高精度定向目标检测的性能。因此，我们提倡使用更严格的评价指标，例如通用目标检测中通常使用的 AP_{75}[1]，，来衡量这项具有挑战性的任务。在 AP_{75} 评价指标下，不仅要求预测框更接近真实标注框，对角度预测的要求也更加严格。如图 2 中的灰色虚线所示，当采用 AP_{75} 时，无论长宽比如何，角度偏差都应控制在特定范围内，否则将不会被判定为正样本。长宽比越大，该范围越窄。

[1]**注意：** 在定向目标检测中实现较高的 AP_{75} 比在水平目标检测中更困难，因为旋转边界框更加精确、冗余区域更少，因此对误差更加敏感。

表 I DOTA-V1.0 和 DIOR-R 数据集上部分定向目标检测器的精度。

方法	DOTA-v1.0			DIOR-R		
	AP_{50}	AP_{75}	$AP_{50:95}$	AP_{50}	AP_{75}	$AP_{50:95}$
旋转 FCOS [34]	72.45	39.84	41.02	62.00	36.10	37.61
S^2A-Net [25]	75.29	40.08	42.00	64.50	38.24	38.02
旋转 Faster RCNN [15]	73.96	43.44	42.93	63.41	41.80	39.72
KLD [26]	73.46	44.74	43.70	64.63	41.60	40.34
旋转 ATSS [35]	73.37	44.95	43.53	63.52	42.61	40.72
GWD [11]	73.25	45.21	44.04	60.31	40.90	39.70
Oriented RepPoints [7]	74.38	46.56	44.57	66.31	44.36	42.81

表 I 分别比较了部分定向目标检测器在 AP_{50} 和 AP_{75}，下的准确度。可以看出，所有检测器在 AP_{50}方面均取得了较高的性能，彼此之间的差距也很小。然而，采用 AP_{75} 时情况则有所不同：某些检测器的表现可能不如其他检测器，尽管它们在 AP_{50} 上的性能更低，例如 S^2 A-Net 与旋转 ATSS 的对比。因此，AP_{75} 能够进一步体现高精度定向目标检测的性能。

至于 $AP_{50:95}$，它也常用于通用目标检测，无疑是一种更严格、更全面的评价指标，但我们认为，AP_{75} 能更直接地反映高精度定向目标检测。$AP_{50:95}$ 包含大量评价指标。在 AP_{50} 或 AP_{55},这类指标下，如果角度预测不准确（如图1所示），仍可能取得相对较高的数值。相反，在 AP_{90} 或 AP_{95}, 这类指标下，定向检测器的性能会大幅下降，这对于当前研究中的比较毫无意义。因此，作为一种中间评价指标，AP_{75}，更加均衡，也更适用于高精度定向目标检测。同时，$AP_{50:95}$中存在平均运算，这使得 AP_{50}和 AP_{55} 这类指标对总体数值有很大影响。

总之，重视高精度定向目标检测是必要且有意义的。

IV. 方法

本节主要围绕角度与长宽比之间的关系设计我们的方法。我们提出了一种新的角度分类方法 AR-CSL，可根据长宽比动态调整平滑过程。随后，我们采用 Deformable DETR [14] 作为检测架构，并通过旋转可变形注意力模块、去噪训练策略以及长宽比敏感匹配和损失对其进行改进，以适应定向目标检测。

A. 概述

图 4 展示了所提出的 ARS-DETR 框架。给定一张图像，首先使用主干网络提取特征图。主干网络会生成层次化特征图，并使用最后三个阶段的输出。然后，采用 1×1 卷积将其通道映射到统一维度。此外，通过对最终特征图进行 3×3 卷积，获得分辨率最低的特征图。接着，将嵌入二维位置编码的多尺度特征图馈入编码器。在不使用自顶向下

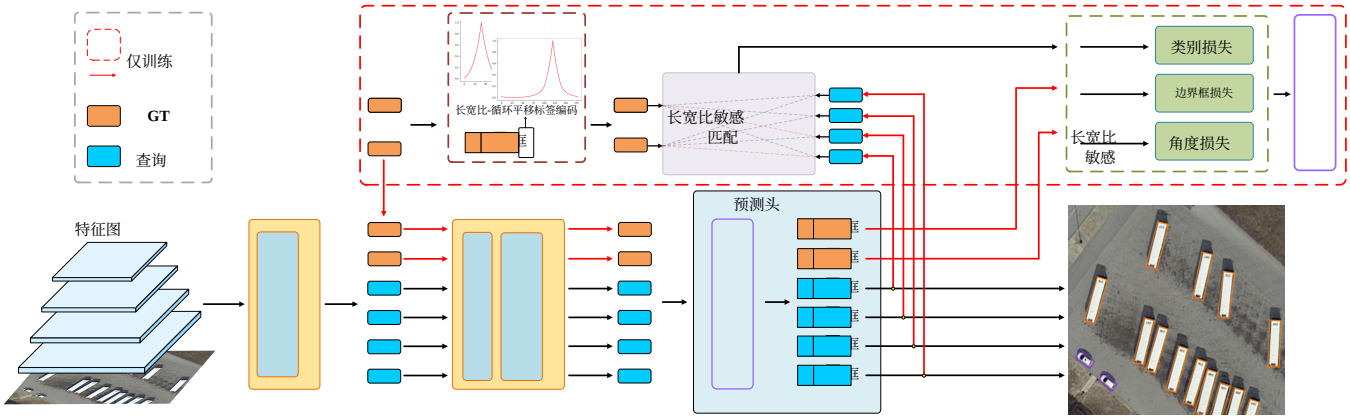


图4. 所提出的 ARS-DETR 框架。“GT”表示真实标注。“仅训练时使用”表示该模块仅在训练过程中起作用，并会在推理时移除。

FPN 结构的情况下， [36], 多尺度可变形注意力能够在多尺度特征图之间交换信息，并进一步优化这些特征图。随后，编码器的输出将生成大量候选框，并用于解码器。接着，选取得分最高的 Top-K 得分候选框作为目标查询，并通过解码器中的多头注意力和多尺度旋转可变形注意力将其转换为输出嵌入。最后，预测头进一步将解码器输出的输出嵌入解码为类别标签、角度标签和水平框坐标。此外，在训练过程中，我们使用带噪声的真实标注作为查询参与训练，以稳定训练过程。同时，我们采用长宽比敏感匹配（ARM）来调整匹配过程中角度的影响，并采用长宽比敏感损失（ARL），在计算角度损失时针对不同长宽比的目标调整训练策略。

B. 宽高比感知圆形平滑标签

1) 重新审视圆形平滑标签：圆形平滑标签（CSL）[8] 不使用基于回归的损失函数，而是将角度预测转换为分类任务，从而自然地消除了边界问题。如图5(a)和图5(b)所示，CSL 将角度划分为 180 个类别，并将第一个角度类别和最后一个角度类别视为相邻角度类别，以消除边界不连续性的影响。随后，它采用高斯窗口函数对目标的角度类别标签进行平滑，以反映相邻角度类别之间的相关性，并使其对角度估计误差具有一定的容忍度。CSL 的表达式如下：

$$CSL(t) = \begin{cases} g(t), & \theta - r < t < \theta + r \\ 0, & otherwise \end{cases}, \tag{2}$$

其中， $g()$ 是窗口函数， t 是标签表示的角度， r 是窗口函数的半径， θ 是真实标注的角度。

尽管 CSL 取得了一些进展，但仍存在两个影响其性能的缺点：

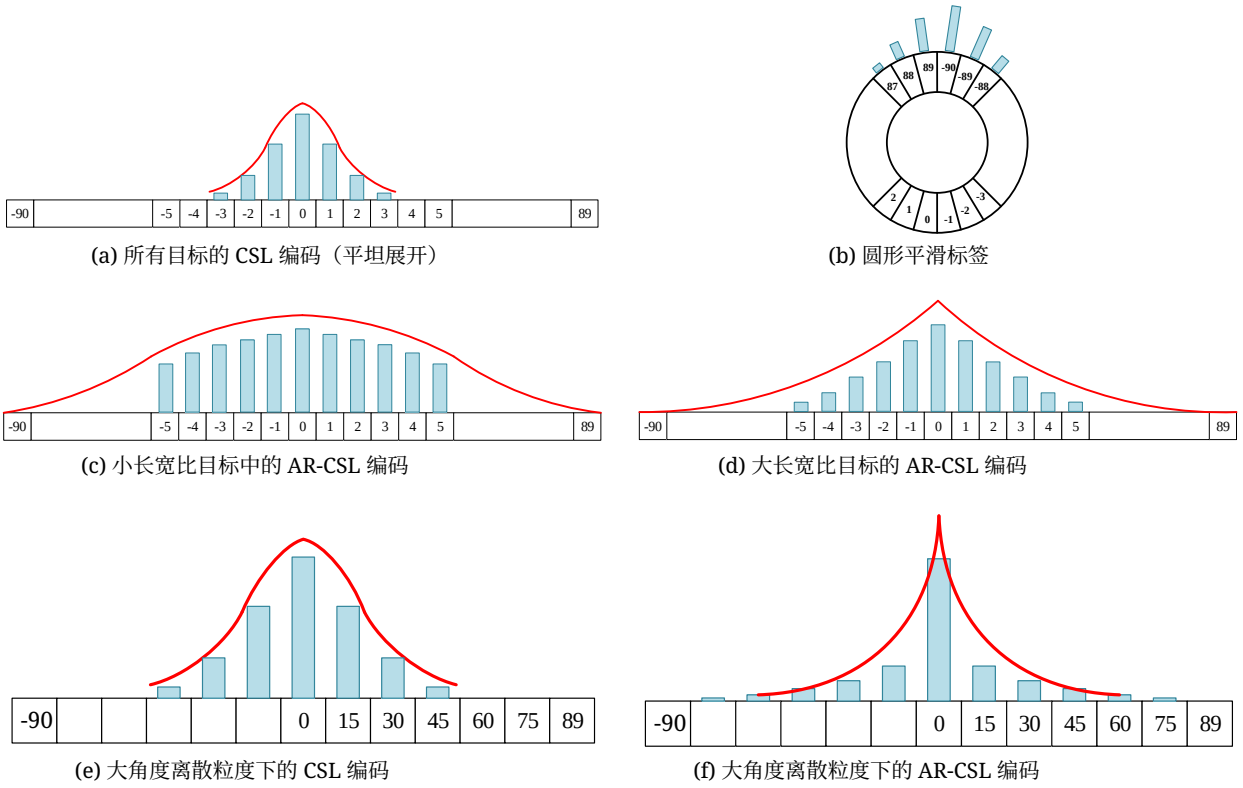
- **固定标签函数**。CSL采用固定半径的高斯函数来学习相邻角度之间的相关性并平滑标签，却未考虑目标的长宽比，如图5(a)所示。然而，从图2中可以明显看出，长宽比不同的目标在相邻角度下的 SkewIoU 差异很大。因此，相邻角度之间的相关性不应是固定不变的，而且高斯窗口可能并不适用于所有目标。

- **对角度离散粒度不敏感**。CSL 对角度离散粒度也不敏感。当角度离散粒度 ω 为 1 时，表示角度被划分为 180 个类别，平滑结果如图5(a)所示。相比之下，当角度离散粒度 ω 为 15 时，表示角度被划分为 12 个类别，平滑结果如图5(e)所示。从这两个结果可以看出，CSL 在不同角度离散粒度下的平滑结果始终一致，这显然是不合理的。随着 ω增大，相邻角度之间的相关性会减弱，而 CSL 对此并不敏感，仍会为相邻角度类别赋予相同的平滑值。因此，还应考虑不同角度离散粒度下相邻角度之间的相关性。

- **超参数介绍**。窗口函数的半径会在一定程度上影响最终性能。作为一个超参数，当 ω 发生变化时，如何确定最佳半径值是一个棘手的问题。

2）宽高比感知圆形平滑标签的设计：根据上述分析，固定的窗口函数和超参数（即半径）在一定程度上降低了基于分类的定向目标检测器的适用性。本小节将从编码形式的角度解决上述问题。

考虑到 SkewIoU 能够动态反映不同目标相邻角度之间的相关性，我们设计了一种长宽比感知圆形平滑标签（AR-CSL）技术，用 SkewIoU 代替固定窗口函数来平滑



(a) 所有目标的 CSL 编码（平坦展开）

(b) 圆形平滑标签

(c) 小长宽比目标中的 AR-CSL 编码

(d) 大长宽比目标的 AR-CSL 编码

(e) 大角度离散粒度下的 CSL 编码

(f) 大角度离散粒度下的 AR-CSL 编码

图 5. 不同角度偏差下，两种编码方法在不同长宽比目标上的比较。为便于比较，图 (a) 和图 (c)-(f) 中的标签均平铺展开；否则，它们应像 (b) 中那样呈圆形。(a) 对于 CSL，无论目标的长宽比如何，都会采用窗口半径固定的高斯窗口来平滑角度标签。(c)-(d) 对于 AR-CSL，会考虑长宽比不同的目标，并采用更合理的平滑策略来体现相邻角度之间的相关性。(e) 对于 CSL，角度离散粒度 ω 会被忽略，因此在不同角度离散粒度 ω 下会得到相同的平滑值。(d) 对于 AR-CSL，平滑值根据角度偏差动态计算，因此会随不同角度离散粒度 ω而变化。

标签。具体而言，我们根据公式 1 计算各角度偏差下边界框的 SkewIoU，并将计算得到的值作为当前角度类别区间的标签。

然后，我们采用最大最小归一化方法对 SkewIoU 值进行归一化，如下所示：

$$\begin{aligned} AR-CSL(k,t) &= \frac{SkewIoU(k,\Delta\theta)-SkewIoU(k)_{min}}{1-SkewIoU(k)_{min}}, \\ \Delta\theta &= |t-\theta|, \end{aligned} \tag{3}$$

其中，k 为真实标注的长宽比，t 为标签表示的角度，θ 为真实标注的角度，$\Delta\theta \in [0°, 90°]$ 为角度偏差。

与 CSL 相比，所提出的 AR-CSL 具有以下优势：

- **动态标签函数**。如图 5(c)–5(d) 所示，通过 SkewIoU，根据目标的长宽比动态计算平滑值。
- **角度离散粒度敏感性**。由于根据公式 3 会考虑不同角度类别的角度偏差，因此相邻类别的平滑值会随角度离散粒度的变化而改变，如图 5(d)–5(f) 所示。
- **无需超参数**。根据公式 1 和公式 3，该方法无需引入超参数，因此使用起来更加方便。

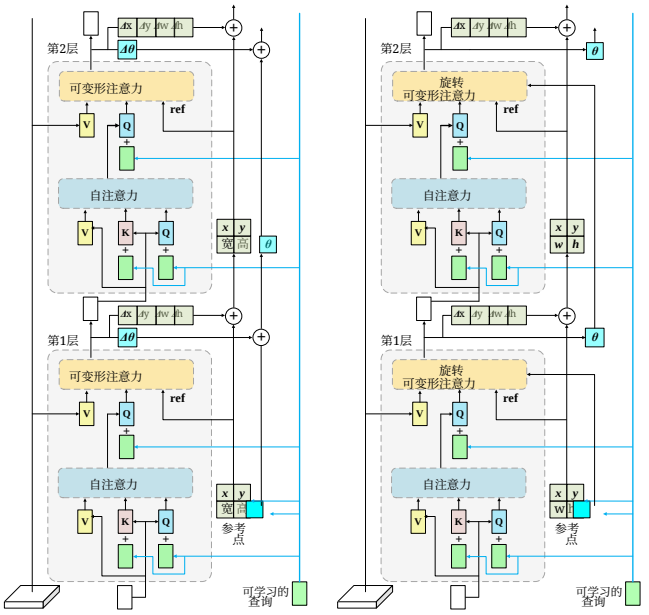
C. 旋转可变形注意力模块

图 6(a) 展示了一种简单的基于 DETR 的定向检测器 [21]，[22]。该检测器仅在预测头中增加一个角度参数，以实现旋转边界框估计。然而，它未能将角度信息融入检测器，以充分发挥检测器的潜力，从而导致特征错位。为解决这一问题，我们提出了旋转可变形注意力模块（RDA）。

给定输入特征图 $x \in \mathbb{R}^{C\times H\times W}$，令 $q \in \Omega_q$ 表示具有表示特征 $z_q \in \mathbb{R}^C$ 和参考框 $b_q = [p_q, w_q, h_q, \theta_q] = [(x_q, y_q) w_q, h_q, \theta_q]$的查询元素，其中 $p_q = (x_q, y_q) \in [0,1]^2$ 是参考框的中心点，$w_q \in [0,1]$、$h_q \in [0,1]$、$\theta_q \in \left[-\frac{\pi}{4}, \frac{\pi}{4}\right]$ 分别是参考框的宽度、高度和角度。旋转可变形注意力特征的计算方式如下：

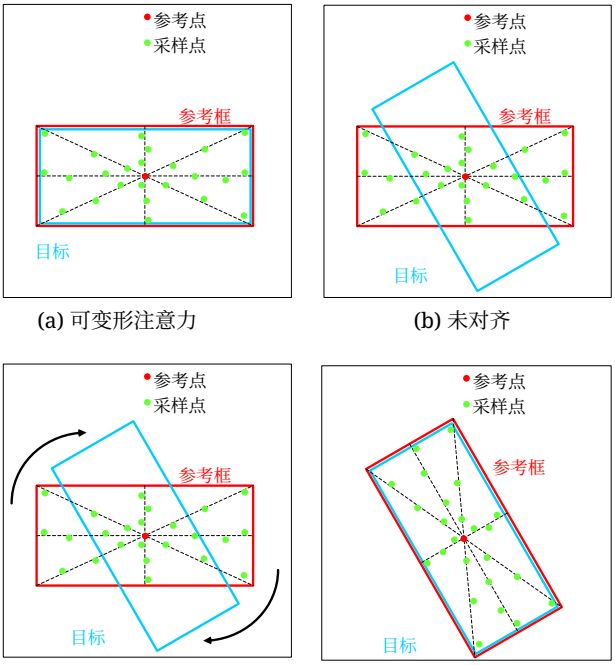
$$RDA(z_q, p_q, x) = \sum_{m=1}^{M} W_m \left[\sum_{k=1}^{K} A_{mqk} \cdot W_m^{'} x(p_q + \Delta p_{mqk})\right], \tag{4}$$

其中 m 表示注意力头的索引，k 表示采样点的索引。M 为注意力头总数，K 为采样点总数。我们采用 $M = 8$ 和 $K = 4$，遵循 [14]。$W_m^{'} \in \mathbb{R}^{\frac{C}{M}\times C}$ 和 $W_m \in \mathbb{R}^{C\times\frac{C}{M}}$ 为可学习权重。Δp_{mqk} 和 A_{mqk} 分别表示第 m^{th} 个注意力头中第 k^{th} 个采样点的采样偏移和注意力权重，



(a) 简单方法 (b) 本文方法

图 6. 在 DETR 中迭代更新角度信息的两种方法。(a) 在简单方法中，尽管角度信息在每一层之后都会迭代更新，但它并未嵌入 DETR 中。(b) 在我们提出的方法中，角度信息会在每一层之后替换为新值，并用于辅助对齐特征。



(a) 可变形注意力 (b) 未对齐

(c) 对齐 (d) 旋转可变形注意力

图 7. 可变形注意力中的未对齐与旋转可变形注意力中的对齐示意图。

标量注意力权重 A_{mqk} 的取值范围为 $[0,1]$,，并由 $\sum_{k=1}^{K} A_{mqk}=1$归一化。对于每个 Δp_{mqk}，其计算方式如下：

$$\Delta p_{mqk}=\frac{(w,h)}{2K}(z_q f_{mk}+r_{mk})R^T(\theta_q), \tag{5}$$

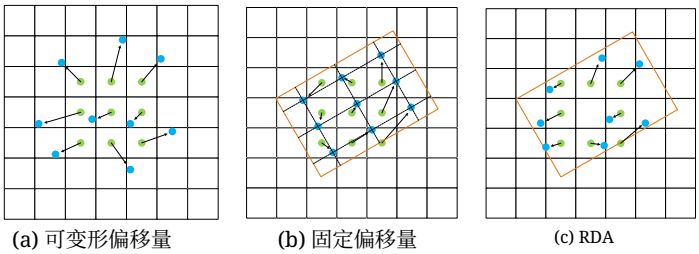
其中，$f_{mk}\in\mathbb{R}^{C\times 2}$ 是投影矩阵。r_{mk} 是 m^{th} 注意力头中第 k^{th} 个采样点的偏置，其计算方式为：

$$r_{mk}=\frac{k(cos(\frac{2\pi m}{M}),sin(\frac{2\pi m}{M}))}{max(\left|cos(\frac{2\pi m}{M})\right|,\left|sin(\frac{2\pi m}{M})\right|)}, \tag{6}$$

$R(\theta_q)=(cos\theta_q,\ -sin\theta_q;\ sin\theta_q,\ cos\theta_q)^T$是旋转矩阵。

通过这种方式，我们可以获得动态采样点$p_q+\Delta p_{mqk}$，并尽可能将其限制在b_q 范围内，以提取对齐后的特征。

如图 7(a) 所示，可变形注意力模块中的采样点会根据相应的参考框进行调整，使采样点被限制在参考框内，并尽可能落在目标内部。然而，如图 7(b) 所示，当目标为定向类型时，如果仍然使用水平参考框[22]，采样点就无法与目标准确对齐。因此，我们设计了旋转可变形注意力模块，根据嵌入的角度信息旋转采样点，使采样点与特征对齐，如图 7(c) 和图 7(d) 所示。此外，我们不逐层细化角度，而是在每一层之后独立预测一个新的角度，如图 6(b) 所示。



(a) 可变形偏移量 (b) 固定偏移量 (c) RDA

图 8. 不同采样方法的示意图。橙色矩形表示边界框。黑色箭头表示偏移场。绿色圆点表示常规采样位置，蓝色圆点表示带偏移的采样位置。(a) 可变形偏移量；(b) 边界框内带旋转的固定偏移量；(c) 边界框内带旋转的可变形偏移量。

如图 8 所示，我们将另外两种采样方法与我们的 RDA 进行了比较。图 8(a) 展示了 [37] 中的采样方法。该方法学习可变形偏移量，以扩展空间采样位置，但在弱监督下可能会从错误的位置采样，尤其是在目标密集排列时。图 8(b) 展示了 [4], [25] 中的采样方法。该方法根据边界框的角度旋转采样点，以对齐特征，但采样位置是固定的。我们的 RDA 也会旋转采样点，同时还会学习可变形偏移量，并根据边界框的宽度和高度对其进行约束。与这些方法相比，RDA 在对齐特征的同时提供了一定的灵活性。

D. 去噪训练

在 [17]中验证了，DETR 中二分图匹配的不稳定性会导致收敛缓慢，进而影响性能。所提出的基于 DETR 的检测

方法通过迭代细化粗略的目标特征和边界框，这一过程可以模拟为重构带噪声的真实标签和边界框。去噪训练作为标签和边界框去噪的辅助任务，具有固定的目标分配结果，因此能够减轻匹配不稳定性的影响并加快收敛。此外，为了模拟正负样本的预测过程，我们为每个真实目标生成正、负两种带噪目标，从而提供更合理的优化目标。此外，为了将去噪训练迁移到定向目标检测中，我们还设计了角度去噪任务用于训练。

给定真实标注 $gt = (c_{gt}, b_{gt}, \theta_{gt})$，其中 c_{gt} 是类别，$b_{gt} = (x_c, y_c, w, h)$ 是边界框，θ_{gt} 是角度，噪声真实值 $gt_n = (c_n, b_n, \theta_n)$ 的生成方式如下。

噪声标签 c_n 通过随机选取部分真实标签，并以 α 的比例将所选标签与任意目标标签叠加生成。正样本的目标标签设为真实标签，而负样本的目标标签则设为背景类别。

噪声边界框 b_n 通过随机移动真实标注边界框的四条边界生成。具体而言，将真实标注框 $b_{gt} = (x_c, y_c, w, h)$ 转换为格式 $b_{gt} = (x_l, y_u, x_r, y_b)$，其中 x_l 和 x_r 分别表示左、右边界的水平坐标，y_u 和 y_b 分别表示上、下边界的垂直坐标。负噪声边界框的噪声尺度应大于正噪声边界框，因为距离较远的候选框应预测为负样本 [18]。因此，为正噪声边界框生成随机噪声偏移量 $\hat{\epsilon} = (\Delta x_l, \Delta y_u, \Delta x_r, \Delta y_b)$，其中 $\Delta x_l, \Delta x_r \sim U(-\frac{\beta}{2}w, \frac{\beta}{2}w)$ 和 $\Delta y_u, \Delta y_b \sim U(-\frac{\beta}{2}h, \frac{\beta}{2}h)$。而对于负噪声边界框，$\Delta x_l$，$-\Delta x_r \sim \left(U(-\beta w, -\frac{\beta}{2}w) + U(\frac{\beta}{2}w, \beta w)\right)$ 以及 $\Delta y_u, \Delta y_b \sim \left(U(-\beta h, -\frac{\beta}{2}h) + U(\frac{\beta}{2}h, \beta h)\right)$。噪声边界框通过 $b_n = b_{gt} + \hat{\epsilon}$ 计算得到，然后转换为格式 $b_n = (x_{nc}, y_{nc}, w_n, h_n)$，作为初始候选框。解码器学习对 b_n 去噪并重建 b_{gt}。

噪声角度 θ_n 通过将 θ_{gt} 与$\theta_n = f(\theta_{gt} + \Delta\theta)$进行偏移生成，其中 $\Delta\theta \sim U(-\gamma\pi, \gamma\pi)$用于生成正向噪声角度，$\gamma$ 为角度噪声尺度。而对于负向角度，

$\Delta\theta \sim U(-2\gamma\pi, 2\gamma\pi)$。$f()$是一个周期函数，从而确保 $\theta_{gt} + \Delta\theta$ 保持在规定范围内。

E. 长宽比敏感匹配与损失

解码器对输出嵌入进行解码后，预测结果将与目标进行匹配，以计算训练过程中的损失。令 y 表示定向目标的真实标注集，$\hat{y}$ 表示 N 预测结果。然后，我们计算这两个集合之间的代价，并搜索N 元素 $\sigma \in O_n$ 中代价最低的排列：

$$\hat{\sigma} = \arg\min_{\sigma \in O_n} \sum_{i}^{N} L_{match}(y_i, \hat{y}_{\sigma(i)}), \tag{7}$$

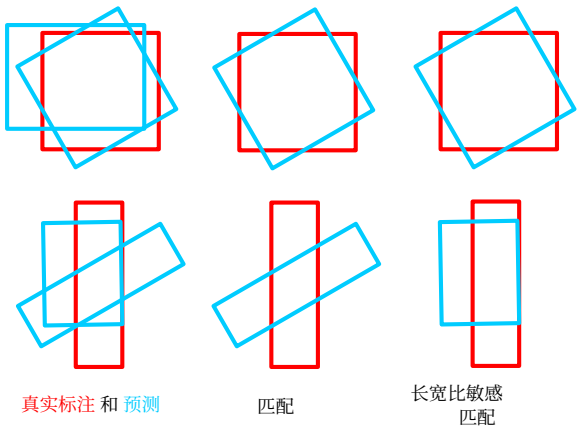


图9. 长宽比敏感匹配示意图。“GT”表示真实标注。

其中，$L_{match}(y_i, \hat{y}_{\sigma(i)})$ 是真实标注 y_i 与索引为 $\sigma(i)$的预测结果之间的成对匹配代价，该代价考虑了类别预测、角度预测以及预测水平框与真实标注水平框之间的相似度，并定义如下：

$$\begin{aligned} L_{match}(y_i, \hat{y}_{\sigma(i)}) = {} & \lambda_{cls} \cdot L_{cls}(c_i, \hat{c}_{\sigma(i)}) \\ & + \lambda_{box} \cdot L_{box}(b_i, \hat{b}_{\sigma(i)}) \\ & + \lambda_{iou} \cdot L_{iou}(b_i, \hat{b}_{\sigma(i)}) \\ & + \lambda_{\theta} \cdot L_{\theta}(\theta_i, \hat{\theta}_{\sigma(i)}), \end{aligned} \tag{8}$$

其中，c 是类别标签，b 是水平框，θ 是角度。此外，L_{cls} 是焦点损失，L_{box} 是L1损失，L_{iou} 是广义交并比损失（GIoU），L_{θ} 是交叉熵损失。

考虑到长宽比更大的目标对角度更加敏感，我们引入一个动态系数来调整角度损失，并将其命名为长宽比敏感匹配（ARM），其形式如下：

$$L(\theta_i, \hat{\theta}_{\sigma(i)}) \rightarrow \frac{2k_i}{1 + k_i} L(\theta_i, \hat{\theta}_{\sigma(i)}), \tag{9}$$

其中，k_i 是目标的长宽比，$k_i \geq 1$。

如图9所示，根据公式8，红色真实标注将与代价最低的蓝色预测相匹配。引入ARM后，长宽比较大的真实标注会在计算代价时更加考虑角度偏差，因此更可能与角度更接近自身角度的预测相匹配。

训练损失函数的定义与公式8一致。同时，我们还引入了长宽比敏感损失（ARL），其定义与公式9相同，用于动态调整角度损失。

V. 实验

A. 数据集与实现细节

DOTA-v1.0 [2] 是规模最大的定向目标检测数据集之一，包含来自不同传感器和平台的 2,806 张大型航空图像，图像尺寸约为 800 ×800 至 4,000 × 4,000 像素，共有 188,282 个实例。它包含 15 个常见

表 II 不同半径下 AR-CSL 与 CSL 在 DOTA-V1.0 上的比较。

方法		CSL				AR-CSL
		R=2	R=4	R=6	R=8	
可变形 DETR	AP_{50}	**72.57**	72.24	72.15	72.10	72.38
	AP_{75}	43.61	42.82	44.07	43.19	**45.71**

类别：飞机（PL）、棒球场（BD）、桥梁（BR）、田径场（GTF）、小型车辆（SV）、大型车辆（LV）、船舶（SH）、网球场（TC）、篮球场（BC）、储罐（ST）、足球场（SBF）、环岛（RA）、港口（HA）、游泳池（SP）和直升机（HC）。我们使用训练集和验证集进行训练，并使用测试集进行测试。我们将图像划分为 1024×1024 个子图，重叠区域为 200 像素。训练期间，仅采用随机水平、垂直和对角翻转来避免过拟合，不使用其他技巧。测试集的性能通过官方 DOTA评测服务器进行评估。

DIOR-R [3] 是一个航空图像数据集，其定向边界框标注源自 DIOR [38] 数据集。该数据集包含 23,463 张图像和 192,518 个实例，涵盖 20 个常见类别：飞机（APL）、机场（APO）、棒球场（BF）、篮球场（BC）、桥梁（BR）、烟囱（CH）、高速公路服务区（ESA）、高速公路收费站（ETS）、水坝（DAM）、高尔夫球场（GF）、田径场（GTF）、港口（HA）、立交桥（OP）、船舶（SH）、体育场（STA）、储罐（STO）、网球场（TC）、火车站（TS）、车辆（VE）和风车（WM）。DIOR-R 中目标尺寸变化很大，体现在空间分辨率以及目标间的类间和类内尺寸差异方面。不同的成像条件、天气、季节和图像质量是 DIOR-R 面临的主要挑战。我们使用训练集和验证集进行训练，并使用测试集进行测试。

OHD-SJTU [33] 是一个用于定向目标检测和目标朝向检测的公开数据集。它包含两个不同规模的数据集，分别称为 OHD-SJTU-S 和 OHD-SJTU-L。OHD-SJTU-S 收集了 43 张大型场景图像，尺寸范围从 10,000 ×10,000 到 16,000 × 16,000，并包含 4,125 个实例。相比之下，OHD-SJTU-L 增加了更多类别和实例，包含六个目标类别和 113,435 个实例。按照以往工作的 [24], [33]处理方式，我们将图像划分为 600 × 600 个子图像，重叠部分为 150 像素，并将其缩放至 800 × 800。

本文所有模型均基于 PyTorch [39]框架 MMRotate [40]实现，并使用 AdamW[41]优化器进行训练。初始学习率为 10^{-4}，在 DOTA-v1.0、DIOR-R 和 OHD-SJTU-L 上采用每个小批量 2 张图像的“3x”训练计划，在 OHD-SJTU-S 上则采用“9x”训练计划。此外，我们进行 500 次迭代的学习率预热，并在衰减步长处将学习率除以 10。

表 III 不同角度离散粒度 ω 下 AR-CSL 与 CSL 在 DOTA-V1.0 上的比较。

方法	离散粒度	AP_{50}	AP_{75}	方法	离散粒度	AP_{50}	AP_{75}
CSL(R=2)	ω=1	**72.57**	43.61	CSL(R=$\frac{6}{\sqrt{\omega}}$)	ω=1	72.15	44.07
	ω=6	72.22	42.51		ω=6	71.70	43.02
	ω=15	71.28	39.93		ω=15	70.45	40.59
	ω=30	71.39	27.62		ω=30	70.55	28.72
CSL(R=6)	ω=1	72.15	44.07	AR-CSL	ω=1	72.38	**45.71**
	ω=6	71.39	42.73		ω=6	72.44	44.32
	ω=15	72.06	33.13		ω=15	71.90	41.52
	ω=30	71.01	19.61		ω=30	71.38	30.25

表 IV 使用 CSL 和 AR-CSL 在 DOTA-V1.0 上对不同检测器进行比较。

方法($\omega = 6$)		AP_{50}	AP_{75}
Deformable DETR(R-50)	CSL(R=6)	71.39	42.73
	AR-CSL	**72.44**	**44.32**
RetinaNet(R-50)	CSL(R=6)	67.63	38.64
	AR-CSL	**67.98**	**39.18**
FCOS(R-50)	CSL(R=6)	70.60	38.42
	AR-CSL	**71.60**	**39.74**

B. 消融研究

1) 关于 *AR-CSL* 的研究：在本小节中，我们采用 De- formable DETR 作为基线，并将 AR-CSL 与 CSL 及其他角度分类方法进行比较。

CSL与Deformable DETR上的AR-CSL对比。我们在基于Deformable DETR的DOTA-v1.0上，对比了所提出的AR-CSL与采用不同半径 R、不同角度离散粒度 ω 的CSL，结果如表II和表III所示。首先，在表II中，我们将ω设为1，并比较了采用不同半径R的AR-CSL与CSL。图10展示了CSL和AR-CSL的可视化结果。 R 对CSL的影响主要体现在高精度定向检测上，最大差距可达1.25%的AP_{75}（$R = 6$ 为44.07%，而 $R = 4$ 为42.82%）。相比之下，AR-CSL无需调整任何超参数即可达到最佳性能（AP_{75}约为45.71%）。其次，在表III中，我们将CSL中的 R 设为2和6，并在不同角度离散粒度 ω下与AR-CSL进行比较。随着 ω 增大，角度间隔也随之增大，每个角度类别的角度表示变得更加模糊，因此CSL和AR-CSL在AP_{75} 上的性能迅速下降。然而，CSL在AP_{75} 上的性能对 ω的变化更加敏感，因此需要进一步调整 R。当 ω 为1和6时，最佳 R 为6（AP_{75}, 分别为44.07%和42.73%）。当 ω 为15和30时，最佳 R 为2（AP_{75}, 分别为39.93%和27.62%）。这是因为随着 ω 增大，相邻角度类别之间的相关性会降低，因此较小的 R 可能更合适。相反，AR-CSL能够随着ω 的变化动态平滑角度标签，因此受 ω 的影响较小，性能也更好。此外，我们还在CSL中将 R 修改为$\frac{R}{\sqrt{\omega}}$，使半径能够在一定程度上根据不同的角度离散粒度进行动态调整。与采用固定



图10. CSL 与 AR-CSL 在 DOTA-V1.0 上的可视化对比。

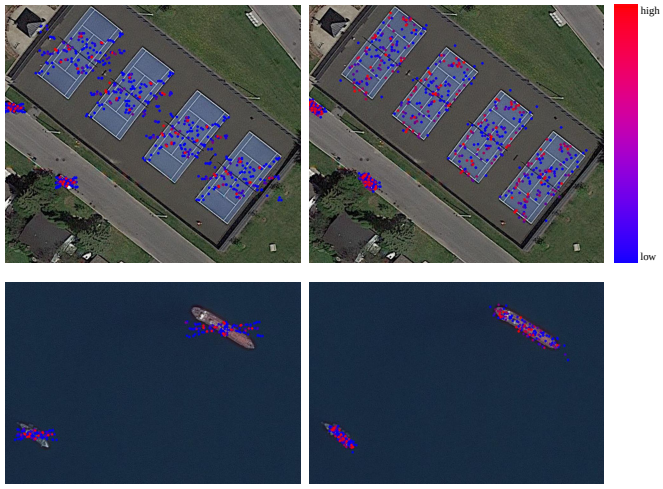


图11. 使用 RDA 模块前（左）和后（右）采样点的位置。每个采样点均以实心圆标记，其颜色表示其注意力权重。

半径的CSL相比，修改后的CSL性能更好，但仍逊于AR-CSL。

使用 CSL 和 AR-CSL 对不同检测器进行比较。我们在其他检测器上进行 CSL 和 AR-CSL 实验，以验证 AR-CSL 的泛化能力，并将 ω 设置为 6。实验结果如表 IV 所示。当检测器更换为 RetinaNet 时，AR-CSL 在 AP_{50} 上达到 67.98%，在 AP_{75}上达到 39.18%。当检测器更换为 FCOS 时，AR-CSL 在 AP_{50} 上达到 71.60%，在 AP_{75}上达到 39.74%。与 CSL 相比，在使用不同检测器时，AR-CSL 在 AP_{75} 上同样表现良好。

与 Deformable DETR 上其他角度分类方法的比较。为了探究我们提出的 AR-CSL 的有效性，我们还在 Deformable DETR 上采用不同的角度分类方法进行了多项实验，结果如表 V 所示。Reg 是采用基于回归的方法预测角度的基线，其 AP_{50} 和 AP_{75} 分别达到 69.39% 和 40.79%。CSL[8] 将角度预测转化为分类任务，并利用高斯窗口平滑角度标签，大幅

表 V DOTA-V1.0 上不同角度分类方法的比较。

方法		Reg	CSL [8]	POE [12]	AR-BCL [30]	AR-CSL
可变形 DETR	AP_{50}	69.39	72.15	72.34	72.27	**72.38**
	AP_{75}	40.79	44.07	43.75	44.67	**45.71**

表 VI DOTA-V1.0 上不同角度预测类型和方式的消融研究。

角度预测类型	角度预测方式	AP_{50}	AP_{75}
回归	$\theta = \theta_{ref} + \Delta\theta_{pred}$	69.48	40.32
	$\theta = \theta_{pred}$	69.39	40.79
分类 (AR-CSL)	$\theta = \theta_{ref} + \Delta\theta_{pred}$	70.81	40.64
	$\theta = \theta_{pred}$	**72.38**	**45.71**

表 VII ARS-DETR 在 DOTA-V1.0 上的组件消融研究。

DN	RDA	ARM	ARL	AN	AP_{50}	AP_{75}
					72.38	45.71
✓					73.14 **(+0.76)**	47.04 **(+1.33)**
✓	✓				72.80 **(+0.42)**	48.06 **(+2.35)**
✓		✓			73.41 **(+1.03)**	47.63 **(+1.92)**
✓			✓		73.31 **(+0.93)**	47.58 **(+1.87)**
✓		✓	✓		73.43 **(+1.05)**	48.13 **(+2.42)**
✓	✓	✓	✓		73.90 **(+1.52)**	48.62 **(+2.91)**
✓	✓	✓	✓	✓	**74.16**（+1.78）	**49.41**（+3.70）

表 VIII DOTA-V1.0 上不同采样方法的比较。

	方法	AP_{50}	AP_{75}
ARS-DETR	可变形偏移量	73.47	46.94
	固定偏移量	73.58	48.76
	RDA	**74.16**	**49.41**



图12. 使用 ARS-DETR 在 DOTA-v1.0 数据集上的检测结果示例。

表IX DOTA-V1.0 上不同角度噪声尺度 γ 的消融研究。

γ		0.00	0.01	0.02	0.05	0.07
ARS-DETR	AP_{50}	73.90	73.46	73.91	**74.16**	73.72
	AP_{75}	48.62	48.86	49.23	**49.41**	48.85

使定向目标检测的性能提升了 2.76%（AP_{50} ）和 3.28%（AP_{75}）。近期提出的 POE [12] 采用 n元编码预测角度，但其在高精度定向目标检测上的表现略逊于 CSL。AR-BCL [30]考虑了近似正方形目标，并引入双向角度以改进 CSL，使 AP_{75}进一步提升了 0.6%。我们提出的 AR-CSL 会根据目标的不同长宽比进行动态调整，在 AP_{75},上达到 45.71%，优于所有对比方法。

2）ARS-DETR 研究： 在本小节中，我们采用 Deformable DETR + AR-CSL 作为基线，并探究所提出方法的有效性。

DETR 中不同角度预测方式的消融研究。 角度的预测类型主要包括回归和分类，每种类型又可分为 DETR 中的两种方式：直接预测（$\theta = \theta_{pred}$）或残差预测（$\theta = \theta_{ref} + \Delta\theta_{pred}$）。表VI比较了这四种组合，发现无论是在回归还是分类中，直接预测方式的效果都是最好的。我们推测，角度的周期性会导致残差预测在优化过程中存在两个优化方向，从而产生边界问题，并在一定程度上阻碍性能提升。

旋转可变形注意力模块的消融研究。为探究旋转可变形注意力模块（RDA）的有效性，我们在表 VII 中对使用和不使用 RDA 的检测器进行了消融研究。通过对齐采样点与特征，RDA 使 AP_{75},分别从 47.04% 和 48.13% 提升至 48.06% 和 48.62%，增幅为 1.02% 和 0.49%。可视化结果见图 11。使用 RDA 之前，大量采样点分布在目标周围的背景区域或相邻目标上，导致未对齐。同时，高注意力权重的采样点相对较少且排列密集。相比之下，使用 RDA 后，采样点基本能够与目标对齐；与此同时，高注意力权重的采样点分布得更加广泛，表明模型更加关注目标的多个部分。此外，如表 VIII 所示，RDA 在 AP_{50} 和 AP_{75}, 上也优于另外两种采样方法，体现了其优越性。

长宽比敏感匹配与损失的消融研究。为探究长宽比敏感匹配（ARM）和长宽比敏感损失（ARL）的作用，我们在表 VII 中进行了详细的消融研究。结果清楚地表明，单独使用 ARM 和 ARL 均可提升性能，在 AP_{75}, 上分别提升约 0.59% 和 0.54%。同时使用 ARM 和 ARL 时，AP_{75}, 上的性能可进一步提升 1.09%，验证了 ARM 和 ARL 的有效性。

去噪训练的消融研究。在实验中，我们主要探究角度噪声的影响，并将其从 DN 中单独分离出来，记为 AN。标签噪声 α和边界框噪声 β 分别设为 0.5 和 0.4，与 DINO [18] 的设置相同。表 VII 显示，基本的 DN 训练策略（仅向标签和边界框添加噪声）可使 AP_{50} 和 AP_{75} 分别提升 0.76% 和 1.33%，达到 73.14% 和 47.04%。表 IX

表 X 与 DOTA-V1.0 上先进旋转目标检测器的比较。R-50 表示 RESNET50 [42]。Swin-T 表示 Swin Transformer [43]。RER-50 表示 RERESNET50 [25]。D-DETR 表示 Deformable DETR [14]。* 表示该模型采用 1X训练计划（12 个训练轮次）。**红色** 和**蓝色**：前两名性能。

方法	主干网络	PL	BD	BR	GTF	SV	LV	SH	TC	BC	ST	SBF	RA	HA	SP	HC	AP_{50}	AP_{75}	$AP_{50:95}$
AO2-DETR [22]	R-50	87.99	79.46	45.74	66.64	78.90	73.90	73.30	90.40	80.55	85.89	55.19	63.62	51.83	70.15	60.04	70.91	22.60	33.31
旋转 D-DETR [14]	R-50	84.89	70.71	46.04	61.92	73.99	78.83	87.71	90.07	77.97	78.41	47.07	54.48	66.87	67.66	55.62	69.48	40.32	40.27
旋转 FCOS* [34]	R-50	89.06	76.97	47.92	58.55	79.78	76.95	86.90	90.90	84.87	84.58	57.11	64.68	63.69	69.38	46.87	71.88	37.30	39.80
Rotated FCOS [34]	R-50	88.52	77.54	47.06	63.78	80.42	80.50	87.34	90.39	77.83	84.13	55.45	65.84	66.02	72.77	49.17	72.45	39.84	41.02
S^2A-Net* [25]	R-50	89.25	81.19	51.55	71.39	78.61	77.37	86.77	90.89	86.28	84.64	61.21	65.65	66.07	67.57	50.18	73.91	35.52	39.05
S^2A-Net [25]	R-50	89.26	84.11	51.97	72.78	78.23	79.41	87.46	90.85	85.62	84.09	60.18	65.90	72.54	71.59	55.31	75.29	40.08	42.00
旋转 RetinaNet* [44]	R-50	89.64	82.56	38.43	69.83	77.39	62.74	77.24	90.68	83.79	82.04	59.91	64.83	57.37	64.76	45.56	69.79	37.69	39.64
旋转 RetinaNet [44]	R-50	87.33	78.91	46.45	69.81	67.72	62.34	73.59	90.85	82.79	79.37	59.62	61.89	65.01	67.76	44.95	69.23	40.96	40.38
Gliding 滑动顶点* [6]	R-50	89.20	75.92	51.31	69.56	78.11	75.63	86.87	90.90	85.40	84.77	53.36	66.65	66.31	69.99	54.39	73.22	37.47	39.52
Gliding 滑动顶点 [6]	R-50	88.71	77.22	52.00	70.85	73.75	74.81	86.55	90.89	80.41	84.63	57.66	62.88	68.49	71.86	58.17	73.26	41.14	41.29
H2RBox [45]	R-50	88.16	80.47	40.88	61.27	79.78	75.25	84.40	90.89	80.05	85.35	58.91	68.46	63.67	71.87	47.18	71.77	41.42	41.49
R^3Det* [24]	R-50	89.29	75.21	45.41	69.23	75.53	72.89	79.28	90.88	81.02	83.25	58.81	63.15	63.40	62.21	37.41	69.80	36.59	37.82
R^3Det [24]	R-50	89.24	83.32	48.03	72.52	77.52	76.72	86.48	90.89	82.33	83.51	60.96	63.09	67.58	69.27	49.50	73.40	41.69	41.43
KFIoU [27]	R-50	89.20	76.40	51.64	70.15	78.31	76.43	87.10	90.88	81.68	82.22	64.65	64.84	66.77	70.68	49.52	73.37	42.71	41.70
旋转 Faster RCNN* [15]	R-50	89.25	82.44	50.05	69.34	78.17	73.59	85.91	90.89	84.08	85.50	57.66	60.96	66.25	69.22	57.74	73.40	39.61	40.75
旋转 Faster RCNN [15]	R-50	89.09	78.28	48.93	71.54	74.01	74.99	85.90	90.84	86.87	85.03	57.97	69.74	68.10	71.28	56.88	73.96	43.44	42.93
旋转 D-DETR（使用 CSL） [8]	R-50	86.27	76.66	46.64	65.29	76.80	76.32	87.74	90.77	79.38	82.36	54.00	61.47	66.05	70.46	61.97	72.15	44.07	42.72
SASM [46]	R-50	87.51	80.15	51.07	70.35	74.95	75.80	84.23	90.90	80.87	84.93	58.51	65.59	69.74	70.18	42.31	72.47	44.21	43.01
KLD [26]	R-50	89.08	84.18	43.77	72.33	79.85	73.58	85.69	90.88	85.14	81.96	65.86	64.60	63.60	68.26	53.19	73.46	44.74	43.70
旋转 ATSS* [35]	R-50	88.50	77.73	49.60	69.86	76.87	72.52	82.49	90.83	80.30	82.96	62.34	64.67	64.83	66.81	53.97	72.29	37.81	40.05
旋转 ATSS [35]	R-50	88.94	79.89	48.71	70.74	75.80	74.02	84.14	90.89	83.19	84.05	60.48	65.06	66.74	70.14	57.78	73.37	44.95	43.53
GWD* [11]	R-50	88.92	77.03	45.90	69.30	72.53	64.06	76.40	90.87	79.20	80.45	57.68	64.37	63.60	64.74	48.26	69.55	38.91	39.50
GWD [11]	R-50	89.06	80.56	44.27	73.02	79.51	73.53	85.55	90.89	86.21	83.26	63.17	64.24	63.56	69.04	52.92	73.25	45.21	44.04
PSC [28]	R-50	89.65	83.80	43.64	70.98	79.00	71.35	85.08	90.90	84.28	82.51	60.64	65.06	62.52	69.61	54.0	72.87	46.18	43.98
CFA [47]	R-50	88.34	83.09	51.92	72.23	79.95	78.68	87.25	90.90	85.38	85.71	59.63	63.05	73.33	70.36	47.86	74.51	46.55	44.41
Oriented Reppoints* [7]	R-50	87.78	77.67	49.54	66.46	78.51	73.11	86.58	90.86	83.75	84.34	53.14	65.63	63.70	68.71	45.91	71.71	41.39	40.88
Oriented Reppoints [7]	R-50	88.52	80.62	52.68	73.04	79.61	80.73	87.76	90.89	81.82	85.33	59.95	64.88	73.81	69.84	46.18	74.38	46.56	44.57
RoI Trans. [4]	R-50	88.70	83.66	54.65	72.72	73.77	78.05	87.39	90.90	80.64	84.76	60.73	63.98	77.61	73.32	54.48	75.03	48.86	45.84
RoI Trans. [4]	Swin-T	88.44	85.53	54.56	74.55	73.43	78.39	87.64	90.88	87.23	87.11	64.25	63.27	77.93	74.10	60.03	**76.49**	**50.15**	**47.60**
ARS-DETR	R-50	86.97	75.56	48.32	69.20	77.92	77.94	87.69	90.50	77.31	82.86	60.28	64.58	74.88	71.76	66.62	74.16	49.41	46.21
ARS-DETR	Swin-T	87.65	76.54	50.64	69.85	79.76	83.91	87.92	90.26	86.24	85.09	54.58	67.01	75.62	73.66	63.39	**75.47**	**51.77**	**47.77**

表 XI 与 DIOR-R 上先进旋转目标检测器的比较。**红色**和**蓝色**：前两名性能。

方法	旋转 FCOS [34]	S^2A-Net [25]	R^3Det [24]	Gliding 顶点 [6]	KFIoU [27]	SASM [46]	GWD [11]	KLD [26]	旋转 Faster RCNN [15]	旋转 ATSS [35]	CFA [47]	定向 RepPoints [7]	RoI 变换 [4]	ARS-DETR
APL	62.89	67.98	62.55	62.67	58.03	64.78	69.68	66.52	63.07	62.19	61.10	67.80	63.28	68.00
APO	41.38	44.44	43.44	38.56	45.41	49.90	28.83	46.80	40.22	44.63	44.93	48.01	46.05	54.17
BF	71.83	71.63	71.72	71.94	69.52	74.94	74.32	71.76	71.89	71.55	77.62	77.02	71.93	74.43
BC	81.00	81.39	81.48	81.20	81.55	80.38	81.49	81.43	81.36	81.42	84.67	85.37	81.33	81.65
BR	38.01	42.66	36.49	37.73	38.82	34.52	29.62	40.81	39.67	41.08	37.69	38.55	43.71	41.13
CH	72.46	72.72	72.63	72.48	73.36	69.21	72.67	78.25	72.51	72.37	75.71	78.45	72.69	75.66
ESA	77.73	79.03	79.50	78.62	78.08	76.28	76.45	79.23	79.19	78.54	82.68	81.13	80.17	81.92
ETS	67.52	70.40	64.41	69.04	66.41	61.37	63.14	66.63	69.45	67.50	72.03	72.06	70.04	73.07
DAM	28.61	27.08	27.02	22.81	25.23	31.66	27.13	29.01	26.00	30.56	33.41	33.67	31.42	34.89
GF	74.58	75.56	77.36	77.89	79.24	72.22	77.19	78.68	77.93	75.69	77.25	76.00	78.00	76.10
GTF	77.04	81.02	77.17	82.13	78.25	77.81	78.94	80.19	82.28	79.11	79.94	79.89	83.48	78.62
HA	40.66	43.41	40.53	46.22	44.67	44.69	39.11	44.88	46.91	42.77	46.20	45.72	49.04	36.33
OP	53.92	56.45	53.33	54.76	54.45	52.08	42.18	57.23	53.90	56.31	54.27	54.27	58.29	55.41
SH	79.41	81.12	79.66	81.03	80.78	83.64	79.10	80.91	81.03	80.92	87.01	85.13	81.17	84.55
STA	66.33	68.00	69.22	74.88	68.40	62.83	70.41	74.17	75.77	67.78	70.43	76.04	77.93	70.09
STO	67.57	70.03	61.10	62.54	64.52	63.91	58.69	68.02	62.54	69.24	69.58	65.27	62.61	72.23
TC	79.88	87.07	81.54	81.41	81.49	80.79	81.52	81.48	81.42	81.62	81.55	85.38	81.40	81.14
TS	48.10	53.88	52.18	54.25	51.64	56.54	47.78	54.63	54.50	55.45	55.51	59.76	56.05	61.52
VE	46.22	51.12	43.57	43.22	46.03	43.58	44.47	47.80	43.17	47.79	49.53	48.02	44.18	50.57
WM	64.79	65.31	64.13	65.13	59.50	63.14	62.63	64.41	65.73	64.10	64.92	68.92	66.44	70.28
AP_{50}	62.00	64.50	61.91	62.91	62.29	62.21	60.31	64.63	63.41	63.52	65.25	**66.31**	64.97	**66.12**
AP_{75}	36.10	38.24	38.40	40.00	40.20	40.40	40.90	41.60	41.80	42.61	43.41	44.36	**46.02**	**45.81**
$AP_{50:95}$	37.61	38.02	37.84	38.34	38.52	39.51	39.70	40.34	39.72	40.72	42.18	42.81	**43.31**	**43.89**

表明在角度上额外添加噪声也能 进一步提升性能。当 γ 设为 0.05 时,



图13. 使用 ARS-DETR 在 DIOR-R 数据集上的检测结果示例。

表XII 与 OHD-SJTU-L 上的先进旋转目标检测器进行比较。**红色** 和**蓝色**：前两名性能。

方法	PL	SH	SV	LV	HA	HC	AP_{50}	AP_{75}	$AP_{50:95}$
RRPN [23]	89.55	82.60	57.36	72.26	63.01	45.27	68.34	22.03	31.12
R^2CNN [48]	90.02	80.83	63.07	64.16	66.36	55.94	70.06	32.70	35.44
RetinaNet-H [24]	90.22	80.04	63.32	63.49	63.73	53.77	69.10	35.90	36.89
R^3Det [24]	89.89	87.69	65.20	78.95	57.06	53.50	72.05	36.51	38.57
旋转 Faster RCNN [15]	81.44	88.30	66.58	75.11	67.97	48.19	71.32	37.43	39.28
旋转 ATSS [35]	81.21	88.73	71.67	76.51	71.48	38.50	71.33	39.82	40.37
RetinaNet-R [24]	90.00	86.90	63.24	86.90	62.85	52.35	**72.78**	40.13	40.58
OHDet [33]	89.73	86.63	61.37	78.80	63.76	54.62	72.49	**43.60**	**41.29**
ARS-DETR	87.48	87.76	65.23	78.51	69.43	56.89	**74.20**	**46.08**	**43.22**

表 XIII OHD-SJTU-S 上与先进旋转目标检测器的比较。**红色**和**蓝色**：前两名性能。

方法	PL	SH	AP_{50}	AP_{75}	$AP_{50:95}$
RRPN [23]	90.14	76.13	83.13	27.87	40.74
R^2CNN [48]	90.91	77.66	84.28	55.00	52.80
RetinaNet-H [24]	90.86	66.32	78.59	58.45	53.07
R^3Det [24]	90.82	85.59	88.21	67.13	56.19
旋转 Faster RCNN [15]	90.83	79.18	85.01	62.73	52.17
旋转 ATSS [35]	90.81	86.49	88.65	72.53	59.51
RetinaNet-R [24]	90.82	85.59	**89.48**	74.62	61.86
OHDet [33]	90.74	87.59	89.06	**78.55**	**63.94**
ARS-DETR	90.18	89.71	**89.95**	**80.67**	**65.49**

分别以 74.16% 和 49.41% 的 AP_{50} 和 AP_{75} 取得最佳性能。表 VII 和表 IX 中的结果表明，加入噪声的真实标注框可以进一步帮助模型学习定向目标检测。

C. 与最先进方法的比较

DOTA-v1.0 上的结果。表 X 汇报了 16 种定向目标检测器的结果。由于原论文中的不同方法采用了不同的图像分辨率、数据预处理、数据增强、主干网络、训练策略及各种技巧等，我们在 MMRotate [40], 上使用相同的设置实现所有检测器，以尽可能确保比较公平。所有结果均通过单尺度训练和测试获得，并采用“1x”（12 个训练轮次）或“3x”（36 个训练轮次）训练计划。以 R-50 和 Swin-T 为主干网络时，如图12所示，我们的方法在 AP_{50}，上分别取得 74.16% 和 75.47%，在 AP_{75}, 上分别取得 49.41% 和 51.77%。使用 ResNet50 作为主干网络时，ARS-DETR 在 AP_{50} 上的表现不如许多先进旋转目标检测器，但在高精度检测方面具有明显优势，并在 AP_{75} 上超越其他先进检测器。具体而言，ARS-DETR 分别比 RoI Trans 高 0.55%（49.41% 对 48.86 %）、比 Oriented Reppoints 高 2.85%（49.41% 对 46.56%）、比 CFA 高

2.86%（49.41% 对 46.55%）、比 GWD 高 4.2%（49.41% 对 45.21%）。此外，有些检测器在 AP_{50} 上表现良好，但在 AP_{75} 上性能大幅下降（例如 S^2A-Net，在 AP_{50} 上为 75.29%，在 AP_{75} 上为 40.08%）；另一些检测器在 AP_{50} 上表现相对较差，但在 AP_{75} 上表现出色（例如 PSC，在 AP_{50} 上为 72.87%，在 AP_{75} 上为 46.18%）。这进一步证明，仅使用 AP_{50} 来评判检测器的性能并不合适。此外，ARS-DETR 的表现也优于基于 DETR 的检测器，例如旋转可变形 DETR 和 AO2-DETR。

DIOR-R上的结果。DIOR-R数据集上的结果如表XI所示。所有方法均采用"3x"训练计划，并使用R-50作为主干网络。ARS-DETR的AP_{50} 和AP_{75}，分别达到66.12%和45.81%。可视化结果如图13所示。

OHD-SJTU上的结果。我们还比较了若干有向目标检测方法在OHD-SJTU上的性能，主要包括R2CNN、RRPN、RetinaNet、R^3Det和OHDet。无需任何额外技巧，我们的ARS-DETR在OHD-SJTU-L和OHD-SJTU-S上的AP_{75} 分别达到46.08%和80.67%，超过了其他先进的有向目标检测器。详细结果见表XII和表XIII。

此外，从上述比较可以看出，AP_{50} 不足以准确反映定向目标检测器的性能，尤其是其高精度定向检测性能。简而言之，本文主要倡导采用更严格的指标（例如 AP_{75}）进一步研究高精度定向目标检测器。

VI. 结论

本文详细分析了角度与不同长宽比目标之间的相关性，并指出了当前评价指标（即 AP_{50}）在高精度定向目标检测中的缺陷。广泛使用的指标 AP_{50}，对角度偏差的容忍度较大，无法准确反映定向目标检测器的性能。因此，采用更严格的指标 AP_{75}衡量性能更为合理。随后，我们设计了一种名为 ARS-DETR 的定向目标检测器，并发现，在 DETR 中，根据不同长宽比目标对角度的敏感性，动态调整角度分类、匹配和损失计算过程中的平滑程度，能够有效提升性能。此外，对齐 DETR 解码器中的特征并采用去噪训练策略，可以进一步提升 DETR 对定向目标检测任务的适应能力。与其他先进旋转目标检测器相比，ARS-DETR 在各种数据集上均取得了更高的检测精度，尤其是在更严格的评价指标下。此外，我们希望该方法能够推动未来高精度定向目标检测研究，以及 DETR 在定向目标检测中的应用。

参考文献

[1] Z. Liu、L. Yuan、L. Weng 和 Y. Yang，"用于船舶识别的高分辨率光学卫星图像数据集及一些新基线。"发表于*ICPRAM*，2017 年，第 324–331 页。

[2] G.-S. Xia、X. Bai、J. Ding、Z. Zhu、S. Belongie、J. Luo、M. Datcu、M. Pelillo 和 L. Zhang，《DOTA：用于航空图像目标检测的大规模数据集》，载于 IEEE/CVF计算机视觉与模式识别会议论文集，2018 年，第 3974–3983 页。[3] G. Cheng、J. Wang、K. Li、X. Xie、C. Lang、Y. Yao 和 J. Han，《用于目标检测的无锚框定向候选框生成器》，IEEE地球科学与遥感汇刊，第 60 卷，第 1–11 页，2022 年。[4] J. Ding、N. Xue、Y. Long、G.-S. Xia 和 Q. Lu，《用于航空图像定向目标检测的 RoI Trans 变换器学习》，载于 IEEE/CVF计算机视觉与模式识别会议论文集，2019 年，第 2849–2858 页。[5] X. Yang、J. Yang、J. Yan、Y. Zhang、T. Zhang、Z. Guo、X. Sun 和 K. Fu，《SCRDet：面向更稳健的小型、密集且旋转目标检测》，载于 IEEE/CVF国际计算机视觉会议，2019 年，第 8232–8241 页。[6] Y. Xu、M. Fu、Q. Wang、Y. Wang、K. Chen、G.-S. Xia 和 X. Bai，《水平边界框上的滑动顶点：用于多方向目标检测》，IEEE模式分析与机器智能汇刊，第 43 卷，第 4 期，第 1452–1459 页，2020 年。[7] W. Li、Y. Chen、K. Hu 和 J. Zhu，《用于航空目标检测的 Oriented RepPoints》，载于 IEEE/CVF 计算机视觉与模式识别会议论文集，2022 年，第 1829–1838 页。[8] X. Yang 和 J. Yan，《使用圆形平滑标签的任意方向目标检测》，载于欧洲计算机视觉会议，2020 年，第 677–694 页。[9] X. Yang、L. Hou、Y. Zhou、W. Wang 和 J. Yan，《用于消除边界不连续性的密集标签编码旋转检测》，载于 IEEE/CVF计算机视觉与模式识别会议论文集，2021 年，第 15 819–15 829 页。[10] J. Wang、F. Li 和 H. Bi，《高斯焦点损失：学习用于航空图像旋转目标检测的分布极化角度预测》，IEEE 地球科学与遥感汇刊，第 60 卷，第 1–13 页，2022 年。[11] X. Yang、J. Yan、Q. Ming、W. Wang、X. Zhang 和 Q. Tian，《使用高斯 Wasserstein 距离损失重新思考旋转目标检测》，载于国际机器学习会议。PMLR，2021 年，第 11 830–11 841 页。[12] Q. Ming、L. Miao、Z. Zhou、J. Song、Y. Dong 和 X. Yang，《用于高精度定向目标检测的任务交错与方向估计》，ISPRS摄影测量与遥感学报，第 196 卷，第 241–255 页，2023 年。[13] N. Carion、F. Massa、G. Synnaeve、N. Usunier、A. Kirillov 和 S. Zagoruyko，《使用变换器进行端到端目标检测》，载于欧洲计算机视觉会议。Springer，2020 年，第 213–229 页。[14] X. Zhu、W. Su、L. Lu、B. Li、X. Wang 和 J. Dai，《Deformable DETR：用于端到端目标检测的可变形变换器》，载于国际学习表征会议，2021 年。[15] S. Ren、K. He、R. Girshick 和 J. Sun，《Faster R-CNN：借助区域候选网络实现实时目标检测》，载于神经信息处理系统进展，2015 年，第 91–99 页。[16] D. Meng、X. Chen、Z. Fan、G. Zeng、H. Li、Y. Yuan、L. Sun 和 J. Wang，《条件 DETR：实现快速训练收敛》，载于 IEEE/CVF国际计算机视觉会议，2021 年，第 3651–3660 页。[17] F. Li、H. Zhang、S. Liu、J. Guo、L. M. Ni 和 L. Zhang，《DN-DETR：通过引入查询去噪加速 DETR 训练》，载于 IEEE/CVF计算机视觉与模式识别会议论文集，2022 年，第 13 619–13 627 页。[18] H. Zhang、F. Li、S. Liu、L. Zhang、H. Su、J. Zhu、L. M. Ni 和 H.-Y. Shum，《DINO：采用改进去噪锚框的 DETR，用于端到端目标检测》，载于国际学习表征会议，2023 年。[19] Y. Wang、X. Zhang、T. Yang 和 J. Sun，《Anchor DETR：基于变换器的检测器的查询设计》，载于人工智能促进协会人工智能会议，第 36 卷，第 3 期，2022 年，第 2567–2575 页。[20] S. Liu、F. Li、H. Zhang、X. Yang、X. Qi、H. Su、J. Zhu 和 L. Zhang，《DAB-DETR：动态锚框是 DETR 更好的查询》，arXiv 预印本 arXiv:2201.12329，2022 年。[21] T. Ma、M. Mao、H. Zheng、P. Gao、X. Wang、S. Han、E. Ding、B. Zhang 和 D. Doermann，《使用变换器的定向目标检测》，arXiv 预印本 arXiv:2106.03146，2021 年。[22] L. Dai、H. Liu、H. Tang、Z. Wu 和 P. Song，《AO2-DETR：任意方向目标检测变换器》，IEEE视频技术电路与系统汇刊，2022 年。

[23] J. Ma、W. Shao、H. Ye、L. Wang、H. Wang、Y. Zheng 和 X. Xue，《通过旋转候选框实现任意方向场景文本检测》，IEEE多媒体汇刊，第 20 卷，第 11 期，第 3111–3122 页，2018 年。[24] X. Yang、J. Yan、Z. Feng 和 T. He，《R3Det：采用特征细化的改进型单阶段旋转目标检测器》，载于人工智能促进协会人工智能会议，第 35 卷，第 4 期，2021 年，第 3163–3171 页。[25] J. Han、J. Ding、J. Li 和 G.-S. Xia，《对齐用于定向目标检测的深层特征》，IEEE地球科学与遥感汇刊，第 60 卷，第 1–11 页，2021 年。[26] X. Yang、X. Yang、J. Yang、Q. Ming、W. Wang、Q. Tian 和 J. Yan，《通过 Kullback-Leibler 散度学习用于旋转目标检测的高精度边界框》，神经信息处理系统进展，第 34 卷，2021 年，第 18 381–18 394 页。[27] X. Yang、Y. Zhou、G. Zhang、J. Yang、W. Wang、J. Yan、X. Zhang 和 Q. Tian，《用于旋转目标检测的 KFIoU 损失》，载于国际学习表征会议，2023 年。[28] Y. Yu 和 F. Da，《相移编码器：预测定向目标检测中的精确方向》，载于 IEEE/CVF计算机视觉与模式识别会议论文集，2023 年。 [Online]。可访问：https://arxiv.org/abs/ 2211 06368。[29] H. Wang、Z. Huang、Z. Chen、Y. Song 和 W. Li，《用于遥感目标检测的多粒度角度表示》，IEEE地球科学与遥感汇刊，第 60 卷，第 1–13 页，2022 年。[30] Z. Xiao、B. Xu、Y. Zhang、K. Wang、Q. Wan 和 X. Tan，《基于长宽比的双向标签编码用于类正方形旋转检测》，IEEE地球科学与遥感快报，2023 年。[31] X. Sun、P. Wang、Z. Yan、F. Xu、R. Wang、W. Diao、J. Chen、J. Li、Y. Feng、T. Xu 等，《FAIR1M：高分辨率遥感影像细粒度目标识别的基准数据集》，ISPRS摄影测量与遥感学报，第 184 卷，第 116–130 页，2022 年。[32] X. Yang、H. Sun、X. Sun、M. Yan、Z. Guo 和 K. Fu，《通过多任务旋转区域卷积神经网络实现任意方向船舶的位置检测与方向预测》，IEEE Access，第 6 卷，第 50 839–50 849 页，2018 年。[33] X. Yang 和 J. Yan，《关于任意方向目标检测：重新审视基于分类的方法》，国际计算机视觉杂志，第 130 卷，第 5 期，第 1340–1365 页，2022 年。[34] Z. Tian、C. Shen、H. Chen 和 T. He，《FCOS：全卷积单阶段目标检测》，载于 IEEE/CVF国际计算机视觉会议，2019 年，第 9627–9636 页。[35] S. Zhang、C. Chi、Y. Yao、Z. Lei 和 S. Z. Li，《通过自适应训练样本选择弥合基于锚框与无锚框检测之间的差距》，载于 IEEE/CVF计算机视觉与模式识别会议论文集，2020 年，第 9759–9768 页。[36] T.-Y. Lin、P. Doll´ar、R. Girshick、K. He、B. Hariharan 和 S. Belongie，《用于目标检测的特征金字塔网络》，载于 IEEE计算机视觉与模式识别会议论文集，2017 年，第 2117–2125 页。[37] J. Dai、H. Qi、Y. Xiong、Y. Li、G. Zhang、H. Hu 和 Y. Wei，《可变形卷积网络》，载于 IEEE国际计算机视觉会议论文集，2017 年，第 764–773 页。[38] K. Li、G. Wan、G. Cheng、L. Meng 和 J. Han，《光学遥感图像中的目标检测：综述与新基准》，ISPRS摄影测量与遥感学报，第 159 卷，第 307 页，2020 年，第 296 页。[39] A. Paszke、S. Gross、F. Massa、A. Lerer、J. Bradbury、G. Chanan、T. Killeen、Z. Lin、N. Gimelshein、L. Antiga 等，《PyTorch：一种命令式风格的高性能深度学习库》，载于神经信息处理系统进展，2019 年。[40] Y. Zhou、X. Yang、G. Zhang、J. Wang、Y. Liu、L. Hou、X. Jiang、X. Liu、J. Yan、C. Lyu 等，《MMRotate：使用 PyTorch 的旋转目标检测基准》，载于第 30 届 ACM国际多媒体会议，2022 年，第 7331–7334 页。[41] I. Loshchilov 和 F. Hutter，《解耦权重衰减正则化》，载于国际学习表征会议，2018 年。[42] K. He、G. Gkioxari、P. Doll´ar 和 R. Girshick，《Mask R-CNN》，载于 IEEE/CVF国际计算机视觉会议，2017 年，第 2961–2969 页。[43] Z. Liu、Y. Lin、Y. Cao、H. Hu、Y. Wei、Z. Zhang、S. Lin 和 B. Guo，《Swin Transformer：使用移位窗口的分层视觉变换器》，载于 IEEE/CVF国际计算机视觉会议，2021 年，第 10 012–10 022 页。[44] T.-Y. Lin、P. Goyal、R. Girshick、K. He 和 P. Doll´ar，《用于密集目标检测的焦点损失》，载于 IEEE/CVF国际计算机视觉会议，2017 年，第 2980–2988 页。[45] X. Yang、G. Zhang、W. Li、X. Wang、Y. Zhou 和 J. Yan，“H2rbox：定向目标检测只需水平框标注”，载于国际学习表征会议，2023年。[46] L. Hou、K. Lu、J. Xue 和 Y. Li，“面向定向目标检测的形状自适应选择与测量”，载于人工智能促进协会年会论文集，2022年，第923–932页。[47] Z. Guo、C. Liu、X. Zhang、J. Jiao、X. Ji 和 Q. Ye，“超越边界框：面向定向和密集目标检测的凸包特征自适应”，载于*IEEE/CVF*计算机视觉与模式识别会议论文集，2021年，第8792–8801页。[48] Y. Jiang、X. Zhu、X. Wang、S. Yang、W. Li、H. Wang、P. Fu 和 Z. Luo，“R2cnn：用于方向鲁棒场景文本检测的旋转区域卷积神经网络”，*arXiv*预印本 *arXiv:1706.09579*，2017年。