arXiv:2304.02199v2 [cs.CV] 4 May 2023

通过知识组合，在无旋转标注的情况下学习旋转目标检测

Tianyu Zhu[1,2,*] Bryce Ferenczi[2] Pulak Purkait[1] Tom Drummond[3]
Hamid Rezatofighi[2] Anton van den Hengel[1]Amazon IML。[2]莫纳什大学[3]
墨尔本大学
tianyuzhu52@gmail.com {alanzty, purkaitp, hengelah}@amazon.com tom.drummond@unimelb.edu.au {tianyu.zhu, bryce.ferenczi, hamid.rezatofighi}@monash.edu

摘要

旋转边界框大幅减少了细长物体的输出歧义，因此优于轴对齐边界框。尽管旋转检测器效果显著，但其应用并不广泛。标注旋转边界框十分费力，因此许多目标检测数据集并未提供这类标注，而是采用轴对齐标注。本文提出一种框架，只需对目标数据集 [1]进行成本更低的轴对齐标注，即可使模型预测出精确的旋转框。

为实现这一目标，我们利用了这样一个事实：神经网络能够学习比任务实际使用的更丰富的目标域表征。这些未充分利用的表征可用于解决更精细的任务。我们提出的框架结合了来自域外源数据集的任务知识（该数据集具有更强的标注），以及来自目标数据集的域知识（该数据集的标注较弱）。我们采用新颖的分配过程和投影损失，以支持在源数据集和目标数据集上进行协同训练。因此，该模型无需在推理期间增加额外计算开销，就能在目标域中解决更精细的任务。我们在多个目标数据集上对该方法进行了全面评估，包括新鲜农产品数据集、*HRSC2016* 和 *SSDD*。结果表明，所提出的方法始终达到与全监督方法相当的性能。

致谢 本文的灵感来自亚马逊开展的一个计算机视觉项目。我们衷心感谢以下为该研究项目作出贡献的亚马逊同事：*Gil Avraham*、*Hisham Husain*、*Chenchen Xu*、*Ravi Garg*、*Shatanjay Khandelwal* 和 *Philip Schulz*。在整个研究过程中，他们的支持、见解和反馈都弥足珍贵，我们由衷感谢他们的帮助。

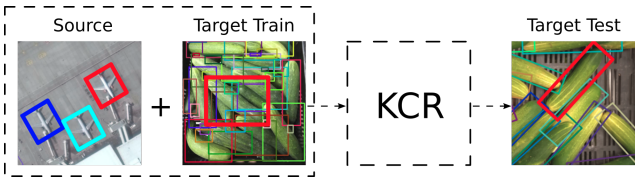

图 1. KCR 结合了具有更强旋转标注的源数据集中的任务知识，以及具有较弱轴对齐标注的目标数据集中的域知识，从而使模型能够在目标域中进行旋转目标检测。

1. 引言

近期研究中提出的旋转检测器 [17, 20, 32]因其在俯视图像 [15, 33, 34]上的出色表现而受到关注。它们减少了细长物体在下游任务中的输出歧义，因此在遮挡严重的密集场景中优于轴对齐检测器 [18]。然而，与轴对齐标注相比，旋转标注的成本更高。此外，常见的二维标注工具（如 Sagemaker Groundtruth [2] 和 VGG app [3]）不支持旋转边界框标注。因此，许多常见的目标检测数据集只有轴对齐标注 [3, 4, 11]。这些问题限制了旋转检测器的应用潜力。在这项工作中，我们提出了**知**识**组**合来学习**旋**转目标检测，这是一种训练方案，只需使用成本较低的目标数据集轴对齐标注，即可预测旋转框。

神经网络将数据编码到潜在空间中，再对其解码以优化给定任务。潜在嵌入是数据的一种抽象表示，包含的信息比输出丰富得多 [29]。深度学习早期研究表明，模型会隐式学习检测边缘和角点等图像特征 [10, 12], , 这些特征

*通讯作者。[1]代码地址：
https://github.com/alanzty/KCR-Official

[2]https://aws.amazon.com/sagemaker/data-labeling/[3]
https://www.robots.ox.ac.uk/vgg/software/via/

如果解码得当，可用于更精细的任务。我们认为，可以通过与强标注源数据集协同优化，学习如何在目标域中解码出更精确的任务。我们设计了一个框架，将源数据集中的旋转目标检测任务知识，与目标数据集中仅带有轴对齐标注的非重叠类别目标的域知识相结合，如图 1 所示。这种方法兼具弱监督学习和迁移学习的优势。

我们遵循一项设计原则：框架应在最大限度地让模型学习目标域知识的同时，尽量减少较弱标签带来的负面影响。为此，我们通过投影损失和一种新颖的分配过程，对源数据集和目标数据集进行协同训练。我们通过消融实验验证了这些设计选择。我们开展了大量实验，证明该框架能够应对源数据集与目标数据集之间较大的域差距。因此，借助 KCR，边界框方向实际上可以免费学习，这得益于 DOTA 等提供旋转标注的免费公开源数据集 [32] 。我们在一个目标密度高且遮挡严重的新鲜农产品数据集上展示了该方法的有效性。对于具有挑战性的黄瓜数据集，所提出的方法利用较弱的轴对齐框进行学习，与利用较强旋转标注进行学习的全监督模型之间的性能（AP50）差距仅为 3.2%。我们将相同的框架应用于 HRSC2016 [16]和 SSDD [27]数据集，结果表明该方法的表现始终与全监督模型相当。对于 SSDD，性能差距缩小至 1.0%。我们相信，我们的方法能够大幅提升旋转目标检测器的应用范围和影响力。我们将公开源代码，以帮助社区降低未来的标注成本。总之，我们的主要贡献如下：

1) 我们提出一种框架，将强标注源数据集中的任务知识与弱标注目标数据集中的域知识相结合。2) 我们将该方法应用于二维旋转目标检测任务，使模型仅凭轴对齐标注即可预测旋转边界框，并通过多个数据集验证该方法的通用性。3) 我们证明了该框架能够应对源数据集与目标数据集之间各种域差距。因此，在实际应用中，无需增加标注成本即可学习边界框方向。

2. 相关工作

旋转目标检测任务。 旋转目标检测要求模型预测具有五个自由度的最小面积矩形，即包围感兴趣目标的旋转边界框 [32]。在轴对齐目标检测中，输出矩形有四个自由度，与图像坐标轴对齐 [13]。旋转框所占面积要小得多，尤其是在估计如图 1 所示的斜置细长物体的位置状态时。旋转目标检测严格优于轴对齐检测，因为框内背景更少，而且方向还可能传达目标姿态信息。然而，与大量现成的大规模水平框数据集 [4, 11, 13, 19]相比，带有旋转标注 [16, 27, 32] 的数据集寥寥无几。这一现象的潜在原因之一是，使用当前标注工具，只需简单点击并拖动即可轻松标注轴对齐框。旋转框则需要额外一个自由度，因此要紧密包围目标需付出更多努力。AWS Sagemaker 和 VGG app 等常用标注工具不支持旋转框。为了获得更贴合目标的旋转标注，用户必须购买实例分割服务，而这项服务成本高得多，对最终任务也没有必要。因此，我们提出这项工作，通过让方向能够自由学习，解决旋转目标检测任务流程中的不足。

弱监督学习。 弱监督学习介于全监督学习与无监督学习之间，因为它仅使用弱标签。标签之所以较弱，是因为它们不完整、不准确或含有噪声 [1, 6, 9, 23, 30, 31]。在计算机视觉中，常见的弱监督学习任务包括仅使用图像级标注的目标检测 [5,35] ，以及仅使用框标注的实例分割 [2, 21]。由于像素级预测难度较大，弱监督实例分割在新颖目标上的表现仍明显落后于全监督模型 [28]。本文关注在训练期间仅使用轴对齐标注来学习旋转目标检测：轴对齐标注提供四个参数，而旋转目标检测需要五个参数。两者相差一个参数，因此我们的方法属于弱监督方法。据我们所知，这一问题仅在特定类别的目标上得到过尝试 [8]，尚未得到普遍解决。对于细长目标，与直接解决弱监督实例分割问题相比，解决这一问题更为合适。

迁移学习。 计算机视觉模型通常会初始化为在 ImageNet 或 COCO 上预训练的骨干网络 [7, 14, 22]，这是一种基础的迁移学习形式，也是常见做法。当目标数据集样本量较小时，例如医学影像 [24]，迁移学习非常有效。尽管传统迁移学习减少了特定任务所需的数据样本数量，但模型仍需要目标域的精细标注才能学习具体任务。在本文中，我们采用协同训练策略，将源域中进行更精细预测的能力迁移到标注较弱的目标数据集。

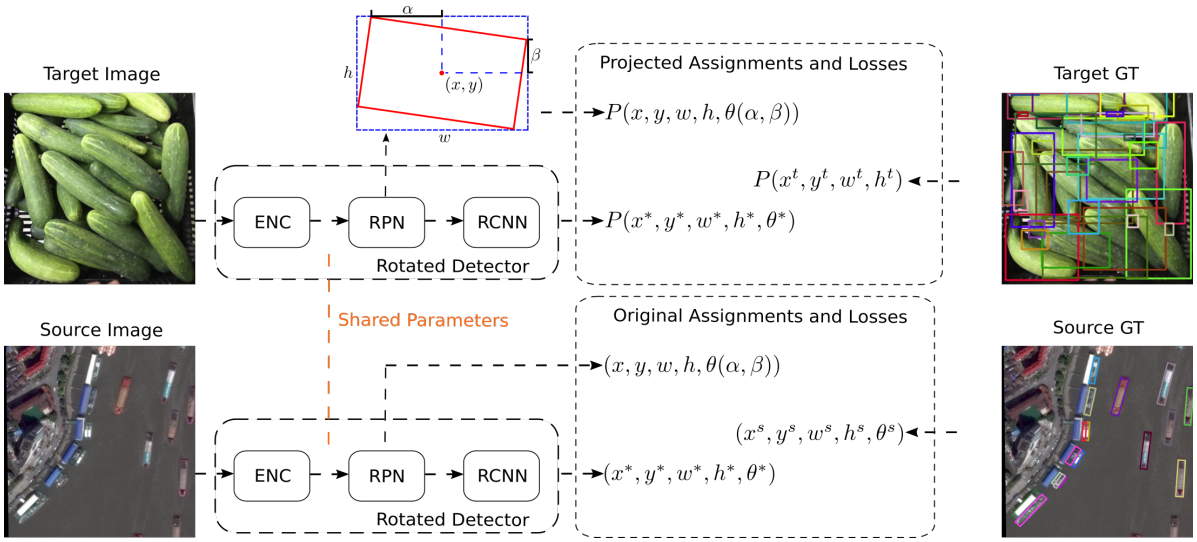

图 2. KCR 的整体框架：该方法结合源数据集中的旋转目标检测知识与轴对齐目标数据集中的域知识，以推断目标对象的旋转边界框。旋转检测器 [33] 接收由卷积神经网络编码的图像，并生成第一阶段 候选区域（x， y， w， h， α， β， p）和第二阶段 精细化结果（x^*， y^*， w^*， h^*， θ^*， p^*）。我们使用函数 P，在分配过程和损失函数中对边界框表示进行投影或扩展。

3. KCR

我们的工作旨在学习如何在一个目标 数据集上预测旋转边界框，而该数据集的训练样例中只有轴对齐边界框。我们设计了一种协同训练方案，利用一个域外但标注充分的源域 数据集，学习细长物体的精确旋转角度。一个可能的源域与目标数据集组合示例是卫星图像数据集 DOTA [32], 和新鲜农产品数据集，如图 1 所示。本节首先简要介绍检测器的工作流程，随后介绍我们用于实现弱监督学习和知识迁移的训练方案。

3.1. 旋转目标检测概述

在这项工作中，我们采用与 [33] 相同的架构选择，整体流程如图 2 所示。前向传播分为两个阶段：先是定向区域提议网络（oriented RPN），然后是 Oriented R-CNN；两种方法都包含一个分类头和一个回归头。RPN 接收一张图像并生成 N 定向区域候选框 $\{R_i\}_{i=1}^N$，每个候选框的形式为 $R_i = (x_i, y_i, w_i, h_i, \alpha, \beta, p_i)$，其中 (x_i, y_i) 表示中心， w_i 和 h_i 分别表示最紧密轴对齐外接框的宽度和高度。 α 和β 分别是相对于外接矩形上边和右边中点的偏移量。 p_i 是目标得分。随后，利用特征空间中的旋转感兴趣区域 [33] 裁剪所提出的旋转区域。接着，将候选区域输入 Oriented R-CNN；它由一个卷积神经网络以及后续的另一个分类头和一个边界框回归头组成，用于校正空间位置。我们将第二阶段的输出记为 $\{R_i^*\}_{i=1}^N$ 和 $R_i^* = (x_i^*, y_i^*, w_i^*, h_i^*, \theta_i^*, p_i^*, c_i^*)$，其中 (x_i^*, y_i^*) 表示中心， w_i^*和 h_i^*分别表示宽度和高度， θ_i^* 表示最终预测旋转框的旋转角度。 p_i^* 是第二阶段的目标得分， c_i^* 表示分类得分。

3.2. 旋转区域提议学习

我们设计了一种同时利用源数据集和目标数据集的协同训练方案。我们将源图像的旋转边界框记为 $\{B_j^s\}_{j=1}^m$，其中 $B_j^s = (x_j^s, y_j^s, w_j^s, h_j^s, \theta_j^s)$ 表示一个旋转框。对于目标数据集，目标图像上的轴对齐框记为 n，其中 $B_j^t = (x_j^t, y_j^t, w_j^t, h_j^t)$ 表示一个轴对齐框。

在第一阶段，会生成一组 N 定向区域，并为每个候选区域 $R_i = (x_i, y_i, w_i, h_i, \alpha, \beta, p_i)$ 根据交并比（IoU）匹配结果分配真实标签 $B_{\sigma(i)}$，其中分配结果 $\sigma(i)$ 和匹配得分 $\tau(i)$ 为

$$\sigma(i) = \underset{j \in \{1,\dots,m\}}{\arg\max}\ iou((x_i, y_i, w_i, h_i), P(B_j)), \quad (1)$$

$$\tau(i) = \max_{j \in \{1,\dots,m\}} iou((x_i, y_i, w_i, h_i), P(B_j)). \quad (2)$$

由于在第一阶段，(x_i, y_i, w_i, h_i)表示的是最紧凑的外接轴对齐框，而非旋转区域本身，因此我们需要一个变换函数 P，将源域旋转数据集中的真实标注投影到轴对齐框。需要注意的是， $P(B_j)$严格来说等于或大于对象的规范轴对齐框。因此，对于目标域轴对齐数据集，我们需要放大 B_j^t。这里

形式化地定义变换函数

$$P(B)=\begin{cases}(x_{min},y_{min},x_{max},y_{max})\\ \qquad\text{for } B=(x,y,w,h,\theta),\\ (x_{min}-\gamma w_i,y_{min}-\gamma h_i,x_{max}+\gamma w_i,\\ \qquad y_{max}+\gamma h_i)\text{ for } B=(x,y,w,h),\end{cases}\tag{3}$$

其中，γ 只是一个可调节的放大因子。该区域提议网络的损失定义如下:

$$L_{\mathcal{S}}=\frac{1}{N}\sum_{i=1}^{N}-\mathbf{1}_{\tau(i)\geq 0.5}\log(p_i)+l_1(R_i,B^s_{\sigma(i)})\tag{4}$$

$$L_{\mathcal{T}}=\frac{1}{N}\sum_{i=1}^{N}-\mathbf{1}_{\tau(i)\geq 0.5}\log(p_i)+l_1((x_i,y_i,w_i,h_i),P(B^t_{\sigma(i)})),$$

$$L=L_{\mathcal{S}}+L_{\mathcal{T}},\tag{5}$$

其中，目标域和源样本的损失均由二元交叉熵（BCE）损失以及针对空间输出的 l_1 回归损失构成。BCE 的标签取决于候选区域与任一投影后的真实标注的重叠度是否大于 0.5。对于源数据集，我们可以根据 θ^s计算真实标注 (α^s,β^s)，其中 θ^s是针对旋转表示的回归。然而，对于目标数据集，我们只计算 (x_i,y_i,w_i,h_i) 的回归损失。模型必须从源数据集中学习旋转知识。为简化数学符号表示，我们省略分类损失，因为与目标得分相比，它与我们的贡献关系较小。

3.3. 学习旋转 R-CNN

在第二阶段，Oriented R-CNN 接收第一阶段生成的一部分候选区域，并给出最终预测$R_i^*=(x_i^*,\ y_i^*,\ w_i^*,\ h_i^*,\ \theta_i^*,\ p_i^*,\ c_i^*)$。边界框回归不那么重要，原因有二。首先，裁剪后的候选区域已经是近似检测结果。这一阶段回归的目标是进行微调。其次，我们在此采用类别无关的旋转边界框回归，因为模型能够从源数据集中学习通用的边界框回归。第二阶段的分类也很容易训练，因为标准轴对齐检测器在这方面的机制本质上相同。最重要且最具挑战性的方面是生成准确的目标得分 p_i^*，它直接取决于真实标注分配过程。我们首先将源数据集的分配过程表述为

$$\sigma_s^*(i)=\underset{j\in\{1,\dots,m\}}{\arg\max}\ iou((x_i,y_i,w_i,h_i,\theta_i),B_j^s),\tag{6}$$

$$\tau_s^*(i)=\max_{j\in\{1,\dots,m\}}iou((x_i,y_i,w_i,h_i,\theta_i),B_j^s).\tag{7}$$

请注意，分配过程基于候选区域，而非最终的精细化预测。与第一阶段的区别在于，这里我们使用准确的旋转框真实标注 B_j^s 而不是 R_i 外部包围轴对齐框。我们使用 $(\alpha,\ \beta)$，结合 [33]，计算第一阶段的 θ。这在源数据集上效果良好。然而，对于缺少旋转信息的目标数据集，这会带来挑战。如果我们沿用与源数据集完全相同的分配过程，那么一个非常接近正确结果的候选区域可能会被分类为负样本，因为旋转边界框与轴对齐边界框之间的重叠度较低。因此，这种训练方案的假阴性率会升高，从而降低召回率。为此，我们提出两种额外策略：投影分配和启发式选择。

投影分配。为避免上述假阴性问题加剧，我们可以将预测的 $(x_i,y_i,w_i,h_i,\theta_i)$ 投影为轴对齐框，然后计算投影框与轴对齐真实标注之间的重叠度。于是有，

$$\sigma_t^*(i)=\underset{j\in\{1,\dots,m\}}{\arg\max}\ iou(P(x_i,y_i,w_i,h_i,\theta_i),P(B_j^t)),\tag{8}$$

$$\tau_t^*(i)=\max_{j\in\{1,\dots,m\}}iou(P(x_i,y_i,w_i,h_i,\theta_i),P(B_j^t)).\tag{9}$$

其中，P是上一节中定义的变换。通过这种方式，由于旋转框可以被正确分配给轴对齐对应框，我们降低了假阴性率，并提高了模型的召回率。然而，考虑一个旋转角度为 $\pi-\theta$ 或 $-\theta$的旋转框。它投影得到的轴对齐框与旋转角度为θ 的旋转框投影得到的轴对齐框相同，但这两个旋转框完全不同。因此，这一过程可能会将负样本误分类为正样本，从而降低精确率。损失如下:

$$L_{\mathcal{S}}^*=\frac{1}{N}\sum_{i=1}^{N}-\mathbf{1}_{\tau_s^*(i)\geq 0.5}\log(p_i^*)+l_1(R_i^*,B^s_{\sigma(i)}),\tag{10}$$

$$L_{\mathcal{T}}^*=\frac{1}{N}\sum_{i=1}^{N}-\mathbf{1}_{\tau_t^*(i)\geq 0.5}\log(p_i^*)\tag{11}$$

$$L^*=L_{\mathcal{S}}^*+L_{\mathcal{T}}^*.\tag{12}$$

启发式选择。除了投影分配之外，我们还提出了一种并行策略。我们采用与源数据集 7 相同的分配规则。然而，我们根据轴对齐真实标注的宽高比和面积大小评估每项分配的可靠性，并尝试从可靠的真实标注中学习更多信息。对于某个轴对齐真实标注框，我们可以计算其宽高比 r_j，其定义为 $r_j=\max(w_j,\ h_j)/\min(w_j,\ h_j)$。宽高比越大，真实标注框就越可靠。此外，我们可以在目标数据集中找到一个面积阈值 $a_{threshold}$，即未被遮挡的典型目标面积的最小值。对于

$w_j * h_j < a_{threshold}$，它很可能是一个被遮挡的框。根据这些启发式规则，我们为目标数据集定义了一个二元可靠性开关和损失，表示如下:

$$g_i = \begin{cases} 1 \text{ if } (r_{\sigma^*(i)} > 3 \text{ or } w_{\sigma^*(i)} * h_{\sigma^*(i)} < a_{threshold}) \\ 0 \text{ otherwise} \end{cases} \tag{13}$$

$$L^*_{\mathcal{T}} = \frac{1}{N} \sum_{i=1}^{N} -g_i \mathbf{1}_{\tau^*_t(i) \geq 0.5} log(p^*_i). \tag{14}$$

并且我们选择屏蔽不可靠样本。尽管启发式选择不如投影分配通用，但对于目标对象宽高比已知的有限工业应用，这种方法可能很有益，例如检测传送带上的瓶子。

4. 实验与讨论

本节首先介绍实现细节和实验设置，包括数据集。随后，我们通过消融实验展示所提方法的有效性，并比较不同的目标域和源数据集。最后，我们给出一些定性结果。

4.1. 实验设置

源数据集。 本文中的源数据集是指带有框方向标注的数据集。模型可以使用源数据集来学习旋转框检测任务。我们选择了多种源数据集，包括 DOTA [32], COCO [13] 和新鲜农产品数据集 1，其中 DOTA [32] 是一个广受欢迎的卫星图像数据集，带有旋转标注；COCO [13] 则是一个广受欢迎的通用目标检测数据集。在 COCO 中，可以通过对每个实例分割真实标注掩码求最小外接矩形，生成旋转边界框真实标注。新鲜农产品数据集 1 是一个颇具挑战性的数据集，具有较高的目标密度和严重遮挡。图像和实例的数量见表 1。该数据集包含三个长形子集，分别为香蕉、黄瓜和胡萝卜。

目标数据集。 目标数据集的训练子集仅包含用于模型学习的轴对齐边界框真实标注。验证子集和测试子集则包含旋转框真实标注，用于评估模型性能。我们选择 HRSC2016 [16],、SSDD [27], 、黄瓜和胡萝卜数据集 1 作为目标数据集，以涵盖多种领域组合，包括卫星、单通道、自然场景以及不同的目标密度。

实现细节。 我们使用 DOTA [32]对检测器 Oriented-RCNN [33] 进行预训练。由于船舶类别与部分目标数据集存在类别重叠，本文在预训练以及将 DOTA 作为源数据集时，移除了 DOTA 中的船舶类别及包含船舶的图像。本文用于评估该方法的主要指标是 AP50。

表 1. 新鲜农产品数据集。

类别	图像数量	实例数量
香蕉	158	7391
黄瓜	48	2036
胡萝卜	47	4647

该指标使用交并比阈值为 0.5 时的平均精度。平均精度（AP）是精确率-召回率曲线下的面积。我们使用 mmrotate 框架 [36] and [33] 进行训练。我们从目标数据集中取 2 张图像，从源数据集中取 2 张图像，组成一个包含 4 张图像的小批量。两个批次的损失和前向传播分别独立计算，随后将损失相加并通过网络反向传播。我们在目标数据集上最多训练 50 轮。推理时，我们将非极大值抑制阈值设为 0.5，而非航空影像数据集中通常使用的 0.1，因为目标密度更高的新鲜农产品数据集中，目标重叠情况要多得多。实验使用一块显存为 11GB 的 2080ti 图形处理器进行。训练时间取决于数据集大小，约为 30 分钟至 2 小时。每个数据集的测试可在 5 分钟内完成，速度为 15 帧/秒，与原始检测器相同。

4.2. 消融实验

为了解决仅有轴对齐真实标注时旋转边界框的弱监督学习问题，我们逐步构建了我们的方法。本节以黄瓜作为轴对齐训练目标数据集，以 DOTA [32] 作为源数据集，展示该方法的演进过程。我们选择这一特定组合，是因为二者在目标外观、密度和遮挡程度方面存在较大的域差距。如果模型能够在源域与目标域之间存在如此大域差距的情况下学会旋转，结果将更有说服力。

我们首先使用轴对齐的黄瓜训练数据集训练一个旋转检测器，并在旋转黄瓜测试集上进行测试，以此建立基线。该基线对应表 2 的第一行，AP50 为 0.491。随后，我们在不修改训练方案的情况下引入源数据集进行协同训练。这意味着我们将源数据集和目标数据集一视同仁，并完全遵循 [33]的训练方案。这样得到的 AP50 为 0.542，较基线有所提升。

RPN投影5将性能提升至0.581（表2）。它没有从轴对齐数据集中学习错误的旋转监督，而是选择不学习。之后，我们尝试单类别策略：对源数据集中的所有类别仅使用一个类别标签。这有效地将性能提升至0.633，因为模型专注于学习目标得分和旋转角度，而非类别分类

表 2. 所提方法的消融实验。我们使用 DOTA 作为源域，新鲜黄瓜作为目标数据集。本表分别分析了旋转协同训练、RPN投影损失、单类别、启发式选择、投影分配以及使用其他目标进行轴对齐预训练的有效性。

旋转协同训练	RPN投影	单类别	启发式选择	投影分配	预训练	AP50↑
-	-	-	-	-	-	0.491
✓	-	-	-	-	-	0.542
✓	✓	-	-	-	-	0.581
✓	✓	✓	-	-	-	0.633
✓	✓	✓	✓	-	-	0.666
✓	✓	✓	-	✓	-	0.664
✓	✓	✓	-	✓	✓	**0.683**

表3. 放大因子 γ 的分析。我们在此分析中使用 DOTA 作为源域，使用 HRSC 作为目标数据集。

γ	1.00	1.05	1.10	1.15	1.20	1.25
AP50↑	**0.791**	0.763	0.756	0.733	0.718	0.650

表4。我们使用密集且存在严重遮挡的黄瓜和胡萝卜数据集作为目标数据集，评估KCR。测试时使用紧致旋转框真实标注评估模型的AP50性能。我们将KCR方法与使用弱轴对齐标注进行训练的原始Oriented-RCNN [33]以及使用强旋转框真实标注训练的全监督模型进行比较。结果表明，KCR使模型的表现可与全监督模型相当，并且在域差距较大的情况下也能取得合理效果。

	方法	源域	目标域训练	AP50↑
Cucumber	原始 [33]	-	轴对齐	0.491
	KCR（我们的方法）	DOTA [32]	轴对齐	0.683
	KCR（我们的方法）	香蕉	轴对齐	**0.788**
	全监督	-	旋转	0.820
Carrot	原始 [33]	-	轴对齐	0.560
	KCR（我们的方法）	DOTA [32]	轴对齐	0.614
	KCR（我们的方法）	香蕉	轴对齐	**0.723**
	全监督	-	旋转	0.765

源数据集的分类问题。如果目标数据集是多类别的，我们可以对源数据集应用单类别策略，因为分类并不重要。

投影分配和启发式选择方法分别实现。第 3.3 节所述的投影分配策略将性能提升至 0.664，因为它降低了假阴性率。但由于假阳性分配问题，该策略无法进一步提升性能。启发式选择策略将模型性能提升至 0.666。最后，我们在其他轴对齐目标上对该模型进行预训练，将性能提升至 0.683，比基线高 0.192。我们采用表 2 的最后一行作为最终方案，并在第 4.2 节将其应用于目标数据集和源数据集的各种排列组合。

我们还对放大因子进行了消融实验

表5。在测试中，使用 HRSC2016 [16]和 SSDD [27]作为目标数据集，以紧致旋转框真实标注评估 KCR方法。通过使用经过旋转增强的 COCO [13], COCO 和 DOTA [32]研究源数据集的影响。我们还将 KCR 与使用弱轴对齐标注进行训练的原始 Oriented-RCNN [33], Grabcut [25]以及使用强旋转框真实标注训练的全监督模型进行比较。除标记为 HS（表示启发式选择）的方法外，KCR 均采用投影分配方法。

	方法	源域	目标域训练集	AP50↑
HRSC2016 [16]	原始 [33]	-	轴对齐	0.175
	GrabCut [25]	-	轴对齐	0.240
	Grabcut 训练集	-	轴对齐	0.629
	KCR（我们的方法）	COCO [13]	轴对齐	0.579
	KCR（我们的方法）	COCO Aug	轴对齐	0.783
	KCR（我们的方法）	DOTA [32]	轴对齐	**0.791**
	KCR-HS（本文方法）	DOTA [32]	轴对齐	**0.778**
	全监督	-	旋转	0.903
SSDD [27]	原始 [33]	-	轴对齐	0.432
	Grabcut 训练集	-	轴对齐	0.585
	KCR（我们的方法）	COCO [13]	轴对齐	**0.888**
	KCR（我们的方法）	COCO Aug	轴对齐	0.874
	KCR（我们的方法）	DOTA [32]	轴对齐	0.881
	KCR-HS（本文方法）	DOTA [32]	轴对齐	0.874
	全监督	-	旋转	0.898

$\gamma \geq 1$。正如表3所示，$\gamma = 1.00$ 效果最佳，这也是我们在所有实验中采用的设置。

4.3. 主要结果

在本节中，我们将KCR框架应用于不同的源域和目标数据集组合，以考察该方法在不同域差距下的普适性。我们还将KCR与一种流行的前景分割器 [25]进行比较。

新鲜农产品数据集。表4显示，随着逐步引入更接近目标结果的信息，结果也逐步提升。首先，我们评估未使用知识组合的方法；也就是说，对于黄瓜和胡萝卜，我们使用轴对齐边界框，并允许网络学习错误的 α, β 和 θ 值。

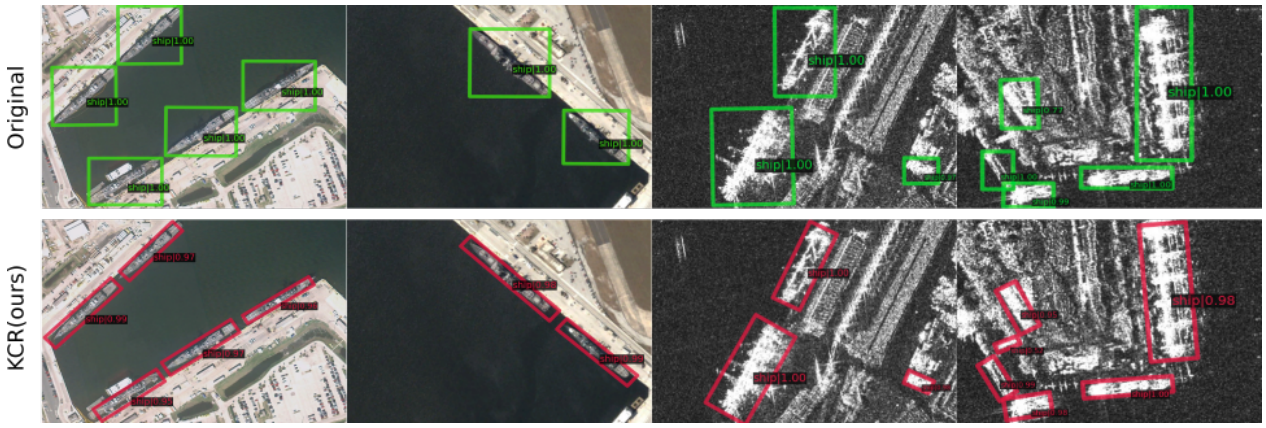

图 3. KCR 与使用弱轴对齐标注训练的原始 Oriented-RCNN 的性能可视化。图像来自 HRSC2016 [16] 和 SSDD [27]测试集。我们使用 COCO 作为源数据集，它与目标域存在较大的域差距。使用 KCR 方法训练的模型能够预测出准确的旋转边界框，比原始模型精确得多。

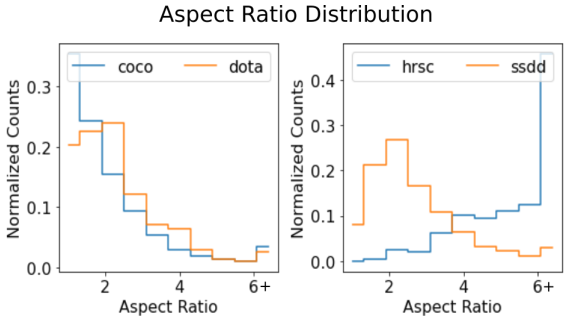

图 4. 4 个数据集上的长宽比分布。

在下一轮迭代中，我们使用 DOTA [32] 航空卫星图像进行知识组合。这些图像与黄瓜和胡萝卜之间存在较大的域差距，因为 DOTA 中的目标较为稀疏。尽管如此，即使数据集存在差异，引入旋转边界框信息仍能显著提升黄瓜和胡萝卜的检测性能，如表4第二行所示。接下来，我们将知识与带有旋转边界框标注的香蕉结合起来。这进一步提升了结果，也表明任务相似性有助于加强知识组合。香蕉数据集让模型有机会学习在严重遮挡情况下进行旋转目标检测。作为参考，表4最后一行展示了直接使用旋转边界框真实标注训练模型的结果，即全监督训练。我们注意到，我们的方法与全监督训练之间的性能差距很小。

HRSC2016。HRSC2016 [16] 是一个专注于船舶的旋转卫星图像数据集。与DOTA [32] 中的目标相比，HRSC2016中的目标通常具有更大的宽高比，并且在图像中占据更大的区域。为避免类别重叠，在将DOTA用作源数据集时，我们移除了包含船舶的图像。我们使用原始训练方案，并仅使用HRSC2016的轴对齐真实标注建立基线，得到的AP50为0.175，如表5所示，比全监督结果低72.8%。由于COCO [13],仅包含自然图像，且COCO与HRSC2016之间存在显著的视觉域差距，我们将COCO作为源数据集，研究KCR的性能。我们从实例分割掩码生成旋转边界框真实标注。使用KCR框架后，结果显著提升至0.579。尽管源数据集中使用了旋转边界框，但COCO中的目标通常是轴对齐的，因此产生的旋转训练信号较弱。因此，我们将图像旋转 $0-180$度，然后进行水平翻转和对角线翻转增强，以增加源数据集中的旋转样本数量。 ϑედეგად，经过强增强的COCO使AP50提升至0.783。由此可见，KCR能够在较大域差距下实现额外参数的迁移学习。最后，我们使用DOTA数据集作为源数据集，取得了0.791的AP50，仅比全监督模型低11.2%。性能差距可能是由于源数据集中缺少同样长且具有较大宽高比的目标。

SSDD。SSDD [27]是一个合成孔径雷达数据集，同样侧重于船舶检测。SSDD 中的图像为单通道，其频率会因采集所用的传感器而异。这些图像通常分辨率较低且含有高频噪声，因此与任何源数据集之间都存在明显的域差距。我们沿用之前相同的知识组合策略。通过 KCR，使用三个不同源数据集（DOTA、COCO 和 COCO 增强数据集）训练的旋转框模型均能取得几乎相同的表现。最终 AP50 达到 0.888，比基线高 45.6%，仅比全监督模型低 1%。

使用 GrabCut 获得旋转框。 GrabCut [25] 是一种计算机视觉算法，它以感兴趣区域为初始化条件预测前景。随后，可使用各种启发式算法对该前景进行分割，并据此预测旋转框。这种旋转预测可用于推理

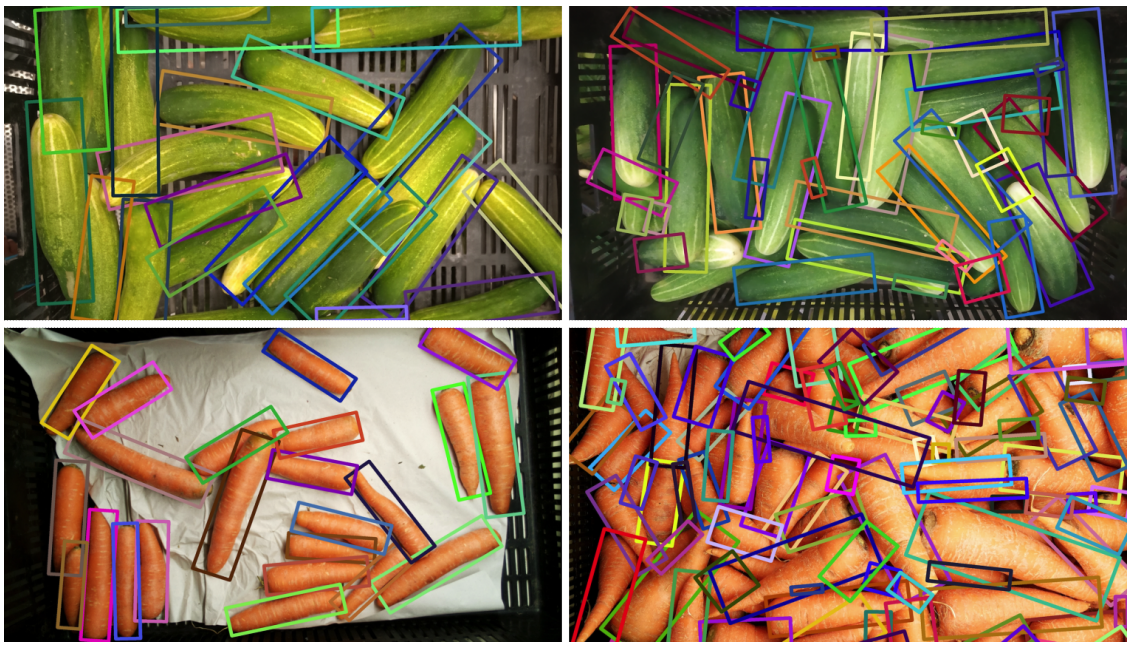

图 5. 仅使用轴对齐数据集训练的 KCR 性能可视化。为确保结果具有普遍性，图像选自未标注数据集。即使在目标密度高且存在遮挡的情况下，该模型也显然能够预测紧致的旋转边界框。

时的后处理步骤，将轴对齐边界框转换为旋转边界框。在测试时，这种方法仅使基线性能提升 6.5%，却将推理吞吐量降至 0.3 帧/秒。另一种方法是离线处理轴对齐真实标注，生成带噪声的旋转边界框真实标注用于训练。该方法效果显著更好：在旋转框真实标注上测试时，HRSC 的 AP50 达到 0.629，SSDD 达到 0.585。然而，在 HRSC 上，KCR 的表现比使用带噪声真实标注进行训练高 16.2%；在 SSDD 上则高 30.3%。

宽高比分布分析。我们展示了图 4 中的宽高比直方图。当实例的宽高比较大且经过旋转时，轴对齐框与旋转框之间的差异最大。大多数 COCO 目标既不细长**也不**旋转，而 DOTA 中细长和旋转目标的分布较为均衡。以原始 COCO 和 DOTA 作为源数据集时，HRSC 上的性能差距大于 SSDD，因为 HRSC 的宽高比分布尾部更重。

4.4. 定性结果

我们在图 3 中展示了 KCR 在 HRSC2016 [16] 和 SSDD [27] 上的表现。该模型成功学会了利用弱轴对齐标注预测准确的旋转边界框。我们使用 COCO [13]，作为源数据集来获取旋转知识，但它与目标数据集之间存在较大的域差距。与使用轴对齐框进行训练和预测的模型相比，我们框架生成的旋转预测框更加紧密。

由于目标密度更高且遮挡严重，新鲜农产品数据集更具挑战性。我们展示了 KCR 在以黄瓜和胡萝卜为目标数据集、以香蕉数据集为提供旋转知识的源数据集时的表现。图中图像来自单独的无标注数据集，以确保结果具有普遍性。如图 5 所示，在具有挑战性的场景中，该模型显然能够预测紧密的旋转边界框，这是本文的核心贡献。即使在遮挡频繁且程度极高、光照条件多变、目标尺寸各异的场景中，该模型也能以较高的精确率和召回率完成任务。对于左列的两幅图像，该模型几乎检测到了所有目标。

5. 结论

旋转目标检测通过缩小包围框的总体面积，提高前景与背景的比例，从而提升下游任务的性能。这对于目标密集且遮挡复杂的场景尤为重要。然而，现有的大多数数据集仅提供轴对齐标注，缺乏标注旋转框的能力。本文提出 KCR 来解决这一问题。KCR 是一种新颖的知识组合训练方案，仅需目标对象类别的轴对齐标注即可训练模型。在推理时，该模型能够预测精确的旋转边界框，达到与全监督方法相当的水平。该方法使检测器能够预测一个额外但至关重要的参数——

这一参数。我们认为，通过降低标注成本，这项工作能够大幅拓展旋转目标检测的应用场景。

参考文献

[1] Jiwoon Ahn 和 Suha Kwak。利用图像级监督学习像素级语义亲和性，以实现弱监督语义分割。载于 *IEEE* 计算机视觉与模式识别会议论文集，第 4981–4990 页，2018年。

[2] Jifeng Dai、Kaiming He 和 Jian Sun。Boxsup：利用边界框监督用于语义分割的卷积网络。载于 *IEEE* 国际计算机视觉会议，第 1635–1643 页，2015年。

[3] Patrick Dendorfer、Hamid Rezatofighi、Anton Milan、Javen Shi、Daniel Cremers、Ian Reid、Stefan Roth、Konrad Schindler 和 Laura Leal-Taixé。MOT20：拥挤场景中多目标跟踪的基准。*arXiv* 预印本 *arXiv:2003.09003*，2020年。

[4] Mark Everingham、Luc Van Gool、Christopher KI Williams、JohnWinn 和 Andrew Zisserman。Pascal 视觉对象类别（VOC）挑战赛。《国际计算机视觉杂志》，88(2):303–338，2010年。

[5] Xiaoxu Feng、Xiwen Yao、Gong Cheng 和 Junwei Han。弱监督旋转不变航空目标检测网络。载于电气电子工程师学会/计算机视觉基金会计算机视觉与模式识别会议论文集，第 14146–14155 页，2022年。

[6]Ramazan Gokberk Cinbis、Jakob Verbeek 和 Cordelia Schmid。用于弱监督目标定位的多折多实例学习训练。载于电气电子工程师学会计算机视觉与模式识别会议论文集，第 2409–2416 页，2014年。

[7] Kaiming He、XiangyuZhang、Shaoqing Ren 和 Jian Sun。用于图像识别的深度残差学习。载于电气电子工程师学会计算机视觉与模式识别会议论文集，第 770–778 页，2016年。

[8] Javed Iqbal、Muhammad AkhtarMunir、Arif Mahmood、Afsheen Rafaqat Ali 和 Mohsen Ali。利用方向进行弱监督目标检测，并应用于枪械定位。《神经计算》，440：310–320，2021年。

[9] Alexander Kolesnikov 和 Christoph H Lampert。《种子、扩展与约束：弱监督图像分割的三项原则》。载于欧洲计算机视觉会议，第 695–711 页。Springer，2016年。

[10]Alex Krizhevsky、Ilya Sutskever 和 Geoffrey E Hinton。《使用深度卷积神经网络进行 ImageNet 分类》。《style id='2'>ACM 通讯，60(6)：84–90，2017年。

[11] Alina Kuznetsova、Hassan Rom、Neil Alldrin、Jasper Uijlings、Ivan Krasin、Jordi Pont-Tuset、Shahab Kamali、Stefan Popov、Matteo Malloci、Alexander Kolesnikov 等。《Open Images 数据集 v4》。《国际计算机视觉期刊》，128(7):1956–1981，2020年。

[12] Yann LeCun、Léon Bottou、Yoshua Bengio 和 Patrick Haffner。应用于文档识别的基于梯度的学习。《*IEEE*论文集》，86(11):2278–2324，1998年。

[13] Tsung-Yi Lin、Michael Maire、Serge Belongie、James Hays、Pietro Perona、Deva Ramanan、Piotr Dollár 和 C Lawrence Zitnick。Microsoft COCO：情境中的常见物体。载于欧洲计算机视觉会议，2014年。

[14] Ze Liu、Yutong Lin、Yue Cao、Han Hu、Yixuan Wei、Zheng Zhang、Stephen Lin 和 Baining Guo。Swin Transformer：使用移位窗口的层次化视觉变换器。载于电气电子工程师学会/计算机视觉基金会国际计算机视觉会议，第10012–10022页，2021年。

[15] Zikun Liu、Jingao Hu、Lubin Weng 和 Yiping Yang。《基于旋转区域的卷积神经网络船舶检测》。载于*2017*年*IEEE*国际图像处理会议（*ICIP*），第 900–904 页。IEEE，2017年。

[16] Zikun Liu、Hongzhen Wang、Lubin Weng 和 Yiping Yang。《用于从复杂背景的高分辨率光学卫星图像中提取船舶的旋转边界框空间》。《style id='2'>《IEEE地球科学与遥感快报》，13(8)：1074–1078，2016年。

[17] Zikun Liu、Liu Yuan、Lubin Weng 和 Yiping Yang。《用于船舶识别的高分辨率光学卫星图像数据集及若干新基线方法》。载于国际模式识别应用与方法会议，第 2 卷，第 324–331 页。SciTePress，2017年。

[18] Shijie Looi。《旋转 Mask R-CNN：从边界框到旋转边界框》。https://github. com/mrlooi/rotated_maskrcnn，2019年。

[19] Anton Milan、Laura Leal-Taixé、Ian Reid、Stefan Roth 和 Konrad Schindler。MOT16：多目标跟踪基准。*arXiv* 预印本 *arXiv:1603.00831*，2016年。

[20] Nibal Nayef、Fei Yin、Imen Bizid、Hyunsoo Choi、Yuan Feng、Dimosthenis Karatzas、Zhenbo Luo、Uma- pada Pal、Christophe Rigaud、Joseph Chazalon 等。ICDAR2017 多语言场景文本检测与文字体系识别鲁棒性挑战赛（RRC-MLT）。载于*2017*年 第*14*届 *IAPR* 国际文档分析与识别会议（*ICDAR*），第1卷，第1454–1459页。IEEE，2017年。

[21] George Papandreou、Liang-Chieh Chen、Kevin P Mur- phy 和 Alan L Yuille。用于语义图像分割的深度卷积网络的弱监督与半监督学习。载于 *IEEE* 国际计算机视觉会议论文集，第 1742–1750 页，2015年。

[22] Joseph Redmon。Darknet：用 C 编写的开源神经网络，2013年。

[23] Zhongzheng Ren、Zhiding Yu、Xiaodong Yang、Ming- Yu Liu、Yong Jae Lee、Alexander G Sc hwing 和 Jan Kautz。具备实例感知、聚焦上下文且内存高效的弱监督目标检测。载于*IEEE/CVF* 计算机视觉与模式识别会议论文集，第 10598–10607 页，2020年。

[24] Olaf Ronneberger、Philipp Fischer 和 Thomas Brox。U-Net：用于生物医学图像分割的卷积网络。载于国际医学图像计算与计算机辅助干预会议，第 234–241 页。Springer，2015年。

[25] Carsten Rother、Vladimir Kolmogorov 和 Andrew Blake。《GrabCut：使用迭代图割进行交互式前景提取》。*ACM* 图形学汇刊，23(3):309–314，2004年。

[26] Olga Russakovsky、Jia Deng、Hao Su、Jonathan Krause、Sanjeev Satheesh、Sean Ma、Zhiheng Huang、Andrej Karpathy、Aditya Khosla、Michael Bernstein 等。《ImageNet 大规模视觉识别挑战赛》。国际计算机视觉期刊，115(3):211–252，2015年。

[27] J. Li、T. Zhang、X. Zhang 和 X. Xu。《合成孔径雷达船舶检测数据集（SSDD）：正式发布与全面数据分析》。遥感，13(18):3690，2021年9月。

[28] Zhi Tian、Chunhua Shen、Xinlong Wang 和 Hao Chen。《BoxInst：基于边界框标注的高性能实例分割》。载于*IEEE/CVF* 计算机视觉与模式识别会议论文集，第5443–5452页，2021年。

[29] Naftali Tishby 和 Noga Zaslavsky。《深度学习与信息瓶颈原理》。载于*2015 ieee* 信息论研讨会（*itw*），第 1–5 页。IEEE，2015 年。

[30] Yude Wang、Jie Zhang、Meina Kan、Shiguang Shan 和 Xilin Chen。用于弱监督语义分割的自监督等变注意力机制。载于*IEEE/CVF* 计算机视觉与模式识别会议论文集，第 12275–12284 页，2020年。

[31] Yunchao Wei、Huaxin Xiao、Honghui Shi、Zequn Jie、Jiashi Feng 和 Thomas S Huang。重新审视空洞卷积：一种用于弱监督和半监督语义分割的简单方法。载于*IEEE* 计算机视觉与模式识别会议论文集，第 7268–7277 页，2018年。

[32] Gui-Song Xia、Xiang Bai、Jian Ding、Zhen Zhu、Serge Belongie、Jiebo Luo、Mihai Datcu、Marcello Pelillo 和 Liangpei Zhang。DOTA：用于航空图像目标检测的大规模数据集。载于*IEEE* 计算机视觉与模式识别会议论文集，第 3974–3983 页，2018年。

[33] XingxingXie、Gong Cheng、Jiabao Wang、Xi wenYao 和 Junwei Han。《Oriented r-cnn 目标检测》。载于《电气电子工程师学会/计算机视觉基金会国际计算机视觉会议论文集》，第 3520–3529 页，2021年。

[34] Xue Yang、Xiaojiang Yang、Jirui Yang、QiMing、Wen- tao Wang、Qi Tian 和 Junchi 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。

[35] Dingwen Zhang、Junwei Han、Gong Cheng 和 Ming- Hsuan Yang。弱监督目标定位与检测：综述。《*IEEE*模式分析与机器智能汇刊》，44(9):5866–5885，2021年。

[36] Yue Zhou、Xue Yang、Gefan Zhang、Jiabao Wang、Yanyi Liu、Liping Hou、Xue Jiang、Xingzhao Liu、Junchi Yan、Chengqi Lyu 等。Mmrotate：基于 PyTorch 的旋转目标检测基准。*arXiv* 预印本 *arXiv:2204.13317*，2022年。