文章编号:1671-4598(2024)07-0030-06 DOI:10.16526/j.cnki.11-4762/tp.2024.07.005 中图分类号:U495 文献标识码:A

基于 YOLOX 的轻量级毫米波雷达和相机融合检测算法

金建鸿[1,2],张 勇[3],戴 喆[4],李 孔[5]
(1. 浙江公路水运工程咨询集团有限责任公司,杭州 310000;
2. 浙江大学,杭州 310058;3. 新奇点智能科技集团有限公司,杭州 311199;
4. 长安大学 运输工程学院,西安 710064;5. 长安大学 信息工程学院,西安 710064)

摘要:为满足交通系统全天候高效准确的感知需求,对基于 YOLOX 的轻量级毫米波雷达和相机融合检测算法进行了研究;该研究主要包括异构传感器信息融合和模型轻量化两个方面;异构传感器信息融合主要针对雷达信息对视觉信息辅助能力不足的问题,设计了一种雷达空间注意力模块有效提取雷达空间特征,以此辅助视觉在低能见度场景下学习到鲁棒的特征表达;在自制数据集和 NuScenes 数据集上训练并测试,所提出的 RV-YOLOX 算法相比 YOLOX 算法 AP 指标提升约 3~4 左右,说明算法的全天候检测能力得到了提升;模型轻量化则针对终端算力设备对算法部署的限制,采用了结构重参数化的方式对 RV-YOLOX 进行了优化,轻量级 RV-YOLOX 在提高推理速度的同时实现了与 RV-YOLOX 相当的检测精度。

关键词:交通系统;融合检测;YOLOX;深度学习

Lightweight Millimeter Wave Radar and Camera Fusion Detection Algorithm Based on YOLOX

JIN Jianhong[1,2], ZHANG Yong[3], DAI Zhe[4], LI Kong[5]
(1. Zhejiang Highway and Water Transport Engineering Consulting Co., Ltd., Hangzhou 310000, China;
2. Zhejiang University, Hangzhou 310058, China;
3. Xinqidian Intelligent Technology Group Co., Ltd., Hangzhou 311199, China;
4. School of Transportation Engineering, Chang'an University Xi'an, Xi'an 710064, China;
5. School of Information engineering College of Chang'an University Xi'an, Xi'an 710064, China)

Abstract: In order to meet the needs of all-weather, efficient, and accurate perception in traffic systems, a lightweight millimeter-wave radar and camera fusion detection algorithm based on YOLOX is studied; The study mainly includes two aspects: the fusion of heterogeneous sensor information and model lightweight; The fusion of heterogeneous sensor information primarily addresses the insufficient auxiliary ability of the radar information to the visual information, and the radar spatial attention module is designed to effectively extract the radar spatial features, thus help the vision to learn the robust feature expression in low visibility scenarios; Training and testing are carried out on the self-made datasets and NuScenes datasets, the proposed RV-YOLOX algorithm increases the AP index by approximately 3~4 times, compared with the YOLOX algorithm, indicating an enhancement in all-weather detection capability; The lightweight model addresses the restrictions of algorithm deployment on terminal computing devices, the structural reparameterization is used to optimize the RV-YOLOX, the lightweight RV-YOLOX improves inference speed while achieving detection accuracy comparable to the RV-YOLOX.

Keywords: traffic system; fusion detection; YOLOX; deep learning

0 引言

感知系统通常配备有多个传感器,以便在监测对象时具有更好的精度和鲁棒性。在许多情况下,这些传感器的信息需要互补融合,以实现所需的精度。相机传感器由于其低成本和丰富的语义信息,成为目标检测应用中最常用的传感器之一[1-2]。然而,在雨、雪、雾、灰尘和强/弱照明条件等具有挑战性的环境下,相机的性能可能受到限制。激光雷达虽然在上述环境中能表现出一定优势,但价格昂贵。毫米波雷达则不受天气和光照的影响,能全天候工作,但其检测精度较低[3]。目前,传感器信息融合的相关研究主要集中在自动驾驶车载端,例如通过融合相机和激光雷

收稿日期:2023-07-04; **修回日期**:2023-08-24。
基金项目:浙江省 2021 年度交通运输厅科技计划项目(2021022)。
作者简介:金建鸿(1974-),男,正高级工程师。
通讯作者:戴 喆(1993-),男,博士,讲师。
引用格式:金建鸿,张 勇,戴 喆,等.基于 YOLOX 的轻量级毫米波雷达和相机融合检测算法[J].计算机测量与控制,2024,32(7):30-35.

达的信息提高目标检测精度[4-7]。而路侧端视角下的传感器信息融合工作相当匮乏，特别是因为缺乏毫米波雷达的基准数据集，与毫米波雷达相关的信息融合工作非常有限。

随着 NuScenes[8]、Astyx HiRes[9]、Oxford RobotCar[10]、SCORP[11]等毫米波雷达相关数据集的开源，使用深度学习算法实现毫米波雷达和相机信息融合的研究取得了很大的进展。文献［12］提出了雷达区域建议网络(RRPN，radar region proposal network)，以雷达检测点作为图像目标检测算法 Fast R-CNN[13]产生锚框的中心点，有效解决了 Fast R-CNN 的速度瓶颈问题。文献［14］在 RRPN 的基础上优化了锚框策略，并构建了由池化层组成的雷达特征提取分支来提取雷达特征，最终通过注意力机制实现了雷达特征与图像特征的融合。文献［15］设计了多层融合架构，使检测网络能自动调整至有利于整体损失最小化的融合方式。另外，文献［16］则提出了一种基于不确定性的相机和毫米波雷达融合方法，取得了优于单传感器基线的性能。

尽管锚框融合、级联融合和相加融合等方式在实现雷达信息辅助视觉信息方面取得了不错的效果。但这些融合方式对雷达特征的挖掘仍不够充分，雷达信息对视觉信息的辅助能力还有提升的空间。此外，现有工作使用的目标检测框架性能略显不足，且很难满足终端设备部署使用的需求。为了实现路侧端视角下交通目标全天候稳定感知，从而提高道路运输网的安全保障能力。本文旨在利用毫米波雷达和相机两种传感器感知信息，设计一种能有效融合这两种信息的高性能检测框架。在此基础上，利用结构重参数化的方式对所设计的框架进行轻量化处理，以实现轻量级的多传感器融合检测框架。

鉴于毫米波雷达和相机特点的互补性，本研究提出了一个端到端的信息融合检测框架 RV-YOLOX，用于高效融合毫米波雷达和相机信息，实现交通目标的实时检测。RV-YOLOX 是基于 YOLOX［17］框架的单阶段融合检测网络，包含两个特征提取分支和一个输出分支。两个特征提取分支分别用于处理相机采集的 RGB 图像和毫米波雷达采集的雷达信息，输出分支则延续了 YOLOX 网络解耦头的结构。本文还设计了一种雷达空间注意力模块，通过结合级联融合和相加融合的特点，以提取毫米波雷达中多感受域的空间信息。由于网络结构中很难确定毫米波雷达和 RGB 图像信息融合的最优层级，因此注意力融合采用了多层自适应的方式，由网络根据训练数据的分布情况自适应的对信息融合的层级进行调整。

由于相机和雷达传感器对目标物体的空间描述坐标系以及数据采集频率都有所差异，因此两源传感器的信息在进入 RV-YOLOX 网络之前，首先要对时域上不同步，空域上属于不同坐标系的两源数据进行时空配准。经过时空配准处理后，将雷达数据投影到像素坐标系中就能得到稀疏的雷达图像，RV-YOLOX 就可以对雷达信息进行特征提取。

YOLOX 常用的主干网络包括 ResNet、DarkNet、ReNetX 等，这些网络训练得到的权重会占用大量存储，且需要做大量的运算。为了压缩模型的大小并提高其推理速度，本文选用 RepVGG［18］网络作为图像特征提取网络，旨在保证检测精度的同时加速推理过程。RepVGG 是以 VGGNet［19］为基础演化而来的网络，VGGNet 经过结构重参数化得到的 RepVGG 不仅没有损失检测精度，而且还获得了一倍的推理速度提升。因此，本文将 RV-YOLOX 中的主干网络替换为 RepVGG 网络，以实现对 RV-YOLOX 的轻量化处理。此外，通过在公开数据集 NuScenes 和自制的路侧端数据集上对 RV-YOLOX 算法进行评估，实验结果表明 RV-YOLOX 可以有效地融合相机和毫米波雷达的特征，检测性能优于基线算法和部分融合检测框架。

1　RV-YOLOX

RV-YOLOX 采用单阶段目标检测算法 YOLOX 为基础架构，通过对相机和毫米波雷达传感器的感知信息进行融合，实现对小车、卡车、公交车等目标类型的检测。RV-YOLOX 框架主要由两个输入分支和一个输出分支组成：两个输入分支分别负责提取 RGB 图像和毫米波雷达的特征信息，输出分支则通过融合两种特征进行目标检测，以输出目标的分类和定位结果。为了满足实时性的需求，本研究在 YOLOX-Tiny 网络的基础上进行了改进，增加了雷达特征分支和雷达空间注意力分支以有效地融合毫米波雷达和 RGB 图像的特征。相较于两阶段融合检测框架，这一单阶段融合检测框架具有速度优势，并且其结构相对简洁。

1.1　数据配准

为了在同一描述空间下对雷达和相机数据进行处理，需要对雷达和相机进行时空配准。具体来说，数据配准是将时域上不同步，空域上属于不同坐标系的雷达和视频数据进行时空维度上的对齐。用齐次坐标表示雷达坐标系到像素坐标系的映射关系时，可以用以下矩阵进行描述：

$$\boldsymbol{p} = \boldsymbol{HP} \tag{1}$$

$$s\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix}\begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \tag{2}$$

式（2）中 s 表示尺度因子，$\boldsymbol{P}=\begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}$ 表示雷达点的齐次坐标，$\boldsymbol{p}=\begin{bmatrix} u \\ v \\ 1 \end{bmatrix}$ 表示雷达点映射到像素平面的齐次坐标，$\boldsymbol{H}$ 矩阵表示二者之间的投影关系。本文采用了通用性好，标定精度高的直接线性变换标定方法来求解从 $\boldsymbol{H}$ 矩阵，这种方法最大的优点就是只需要至少 6 对控制点对即可求解出模型参数。

1.2　雷达图像描述

利用数据配准技术，雷达数据可以从雷达坐标系映射到

像素坐标系，从而得以构建稀疏的雷达伪图像。如图 1 所示，图中矩形则表示与 RGB 图像相同尺度的零矩阵。图 1 中稀疏雷达图像具有 2 个通道，这两个通道分别是以车辆类别和车辆速度信息作为像素值，赋值到雷达点在像素空间的投影位置得到的。雷达图像的稀疏性既体现了雷达数据的固有特性，也反映了道路交通环境中的目标分布较为稀疏的特点。值得一提的是，在真实道路交通环境中得到的雷达图像要比图 1 示例中绘制的更为稀疏。

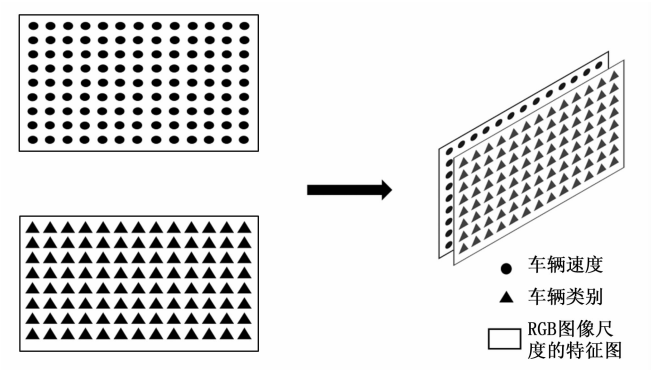

图 1 稀疏雷达图像示例

1.3 RV-YOLOX 架构

RV-YOLOX 的网络架构如图 2 所示，其采用两个独立的输入特征提取分支，分别从 RGB 图像和稀疏雷达图像中提取图像特征和雷达特征。其中图像特征提取分支使用的是 YOLOX 默认的特征提取架构，这一设计便于利用 YOLOX 模型在 COCO 数据集上的预训练权重，以便于对模型进行初始化。另一方面，雷达特征提取分支则采用了本研究设计的网络架构。由于雷达图像稀疏性的特点，雷达特征提取分支主要通过 2D 卷积层来计算雷达特征，并通过最大池化层和平均池化层对特征图进行下采样处理。在雷达特征提取分支中，Block2、Block3 和 Block4 的结构相同，而 Block1 则在输入端多连接了一个 3×3 的卷积层。这 4 个 Block 的卷积核个数分别为 64、256、256 和 128。

检测头部分采用了 YOLOX 原始的解耦检测头设计，以平衡分类任务和回归任务之间的冲突。如图 2 所示，YOLOX 检测头首先通过一个 1×1 的卷积层减少通道维度，然后使用包含两个 3×3 卷积层的平行分支进行特征提取。YOLOX 检测头对预测特征图的每个位置都预测了 $num_{cls}+4+1$ 个参数，其中 num_{cls} 代表目标类别数，4 代表边界框参数，1 代表是否包含物体。不同的预测特征图采用的检测头不同，检测头之间参数也不共享。

损失函数部分使用了两个损失函数评估预测与 ground truth 之间的损失。第一个损失函数 BCE loss 是分类损失，用于评估 Cls 和 IoU 分支。第二个损失函数 IoULoss，用于评估 Reg 分支。YOLOX 的总损失函数包括上述三部分损失，定义如式（3）：

$$L_{\text{total}} = L_{\text{cls}} + L_{\text{obj}} + L_{\text{reg}} \tag{3}$$

1.4 雷达空间注意力

由于雷达点能够检测到目标车辆在空间中的位置信息，因此本研究试图将此空间位置信息从雷达信息中提取出来，进而指导视觉特征提取更有效的信息流。在实际交通场景中，经常会遇到许多小目标或因光照、天气等因素导致的模糊目标。针对这样的情况，如果能先定位到小目标和模糊目标的大概位置，让视觉特征聚焦于在这些位置提取特征，对于提高检测的性能是极其有效的。基于以上分析，本文提出了一种雷达空间注意力模块，期望能对 RGB 图像分支的特征提取产生积极影响。

雷达空间注意力模块的结构如图 3 所示。首先，雷达特征图 F 经过最大池化层和平均池化层处理得到两张通道

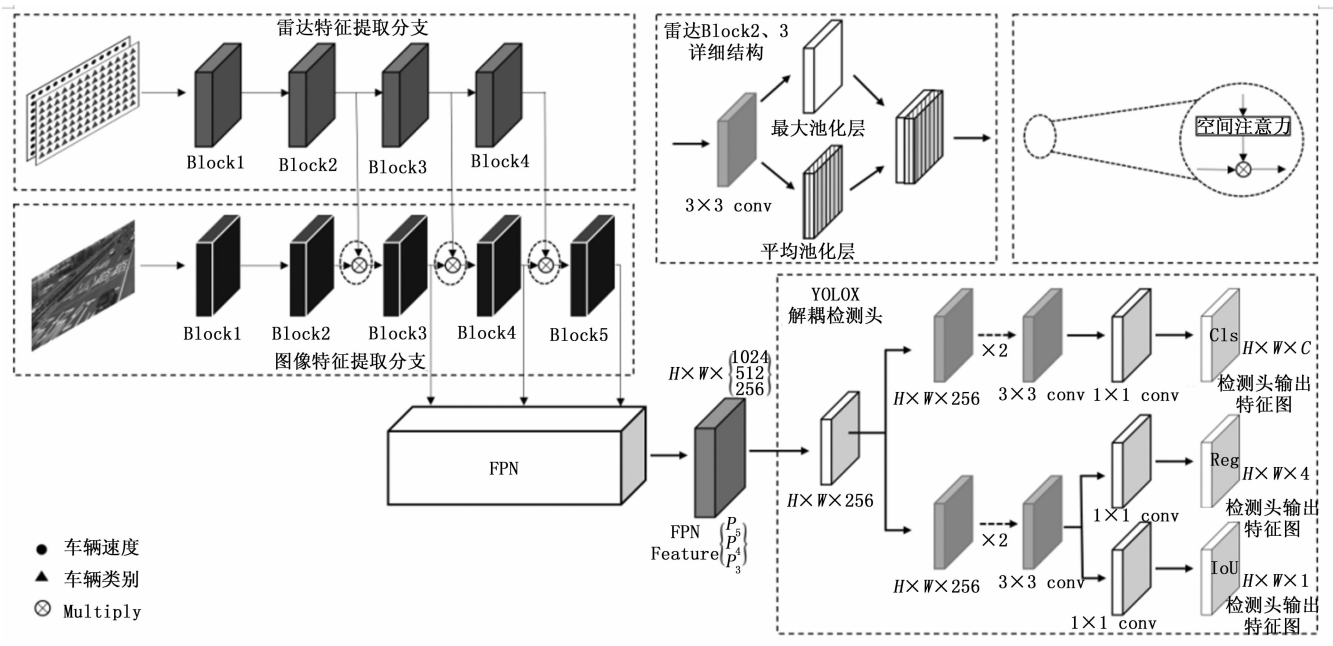

图 2 RV-YOLOX 模型架构

数为 1，尺度为 $H\times W$ 的特征图。这两种不同的池化层可以从不同描述角度提取雷达特征图的空间信息。因此，选择将这两张特征图进行拼接。接着，为了获取多感受域的空间信息，使用 3 种不同的卷积层进一步提取特征。最后，将提取到的 3 种空间信息进行相加，得到最终的雷达空间注意力特征图 $M_S(F)$。计算公式可以表示为：

$$M_S(F)=\sigma\left\{\sum_{i=1}^{3}\mathrm{conv2d}_{2i-1}[\mathrm{AvgPool}(F);\mathrm{MaxPool}(F)]\right\} \quad (4)$$

式（4）中，σ 表示 Sigmoid 函数，$\mathrm{conv2d}_{2i-1}$ 表示卷积核为 $2i-1$ 大小的 $2D$ 卷积层，AvgPool 表示全局平均池化，MaxPool 表示全局最大池化，F 表示雷达特征图。使用 $M_S(F)$ 对图像特征图 F' 进行空间维度的加权，即可得到雷达空间注意力引导下的图像特征信息流，计算公式为：

$$F'_S=M_S(F)F' \quad (5)$$

雷达空间注意力具有多尺度空间感受域，可以增强图像特征图中对目标车辆特征表示的信息流。

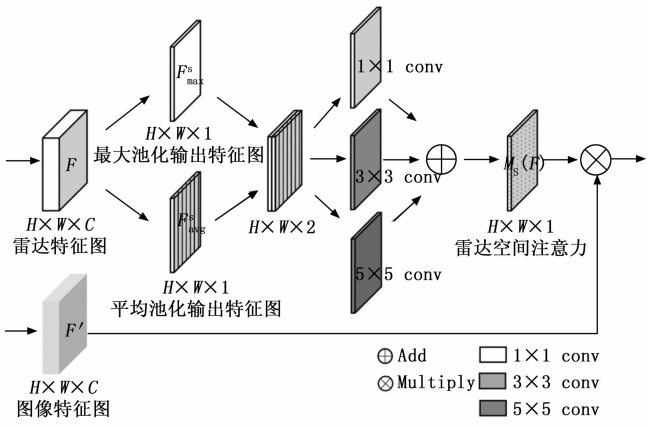

图 3　雷达空间注意力模块

1.5　轻量级 RV-YOLOX

神经网络模型分为训练和推理两个阶段。训练阶段是根据训练数据学习模型的参数，推理阶段则是利用学习好的模型参数对新数据计算得出结果。根据现有的一些经验来看，训练阶段并行多个分支一般能够增加模型的表征能力，扩大模型的搜索空间。但是训练完成进入推理阶段，则并不需要这么多的参数。因此，如果能在训练阶段构造复杂的模型来捕捉数据中的微小信息，在模型部署前对其进行简化将会有效解除终端算力设备对神经网络部署方面的限制。RepVGG 网络正是在上述理论启发下，对 VGG 网络通过结构重参数化获得的轻量级网络结构。

RepVGG 网络通过结构重参数的操作，在训练阶段采用多分支模型，推理阶段将多分支模型转换为单分支模型，从而实现推理速度的提升和内存占用的降低。在多分支模型向单分支模型的转换过程中，通过多个算子的融合，使得计算量得以减少。随着计算量的减小，硬件的使用次数也相应减少，进一步促进了推理过程的加速。RepVGG 网络的结构重参数化过程如图 4 所示，主要包括 3 个步骤：第一步，在只有 BN 算子的分支上构建一个实现恒等映射的卷积核大小为 3 的 Conv2d 算子，并将卷积核大小为 1 的 Conv2d 算子转换成卷积核大小为 3 的 Conv2d 算子。第二步，将卷积核大小为 3 的 Conv2d 算子和 BN 算子融合成一个算子。第三步，将 3 个卷积核大小为 3 的 Conv2d 算子融合成一个算子。

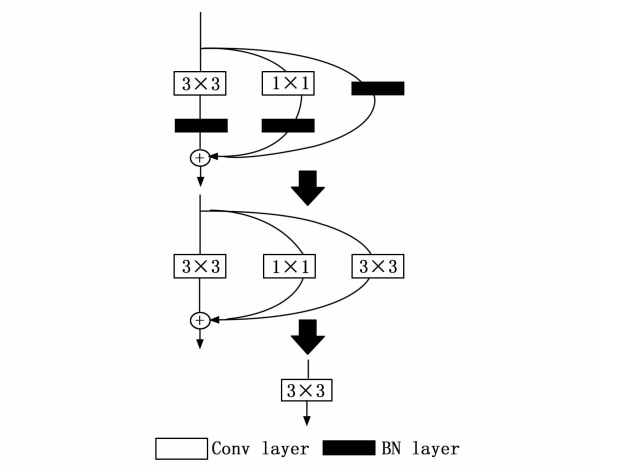

图 4　结构重参数化

为了适应一些终端算力设备的硬件要求，本文将 RV-YOLOX 中图像特征提取分支的网络结构换成 RepVGG-A0 网络，以实现轻量级的 RV-YOLOX 网络。RepVGG-A 的网络配置如表 1 所示，表中 1，2，4，14，1 分别表示每个 stage 重复 block 的次数，a 代表模型 stage2～stage4 的宽度缩放因子，b 代表 stage5 的宽度缩放因子。RepVGG-A0 网络中 a 取值为 0.75，b 取值为 2.5。

表 1　RepVGG-A 网络配置表

Stage	输出大小	RepVGG-A
1	112×112	1×
2	56×56	2×64 a
3	28×28	4×128 a
4	14×14	14×256 a
5	7×7	1×512 b

2　实验结果与分析

2.1　数据集介绍

为了评估 RV-YOLOX 算法在车载端视角下的性能表现，本研究采用了涵盖多种室外场景的 NuScenes 数据集。该数据集的数据采集车辆搭配有 5 个毫米波雷达、6 个摄像头和 1 个激光雷达。数据集共包括 1 000 个场景，约 140 万张相机图像，140 万帧毫米波雷达扫描数据，39 万帧激光雷达扫描数据和 1.4 万个目标边界框。NuScenes 数据集开源了传感器套件的全部采集数据，极大推动了毫米波雷达与其他传感器信息融合相关研究的进展。

为了评估 RV-YOLOX 算法在路侧端视角下的性能表现，本研究在真实交通场景中将相机和毫米波雷达两种传感器固定好位置，同步采集一段时间得到了自制数据集的

原始数据。原始数据还需经过数据清洗、时空配准、图片标注等工作才能得到最终的数据集。为了获得目标车辆真实的位置和类别标签，本研究对所采集的数据进行了手动标注。经过数据清洗和数据整理，自制数据集中共包含路侧端视角下白天和晚上两个场景各 10 000 帧相机和毫米波雷达的同步数据。

2.2 实验设置

实验包括在自制数据集和 NuScenes 数据集上对 FFPN[20]、RANet[14]、BIRANet[14]、YOLOX[17]、RV-YOLOX 和轻量级 RV-YOLOX 算法的评估。其中，FFPN 和 YOLOX 属于基于视觉的检测网络。RANet 和 BIRANet 是基于 FFPN 构建的两阶段融合检测网络。RANet 仅使用了锚框策略对毫米波雷达和 RGB 图像的信息进行融合，BIRANet 则在锚框和网络结构两个方面对两源信息进行融合。RV-YOLOX 和轻量级 RV-YOLOX 是本研究提出的单阶段融合检测网络，通过雷达注意力的方式实现了两源信息的融合。

RV-YOLOX 和轻量级 RV-YOLOX 使用配对的 RGB 图像和稀疏雷达图像进行训练，其中输入图像的大小为 1 024×1 024，批量大小设置为 2，其余超参数均与 YOLOX 的默认设置相同。图像特征提取分支和检测头部分采用 YOLOX 在 COCO 数据集上的预训练权重进行初始化，雷达特征提取分支则随机初始化权重。实验的硬件环境包括 2 块 NVIDIA GeForce GTX 1080 Ti GPU，1 块 Intel Core i7-6800K CPU 以及 64 GB 的内存。

mAP（Mean Average Precision，均值平均精度）指标常用于评估检测算法的性能，其定义如下：

$$mAP = \frac{\sum_{q=1}^{Q} AP(q)}{Q} \tag{6}$$

在式（6）中，q 代表不同的类别，Q 代表类别的总数，AP（q）代表类别 q 的平均精度。在评估实验结果时，使用了与 COCO AP 相关的评估指标。COCO AP 使用 10 个 IoU 阈值（范围 0.5 到 0.95，步长为 0.05）计算所有类别的 mAP。AP_{50}和 AP_{75}分别是 IoU 阈值为 0.5 和 0.75 的 COCO AP。AP_S、AP_L 和 AP_M 则是像素区域小于 32^2、大于 96^2 以及在两个阈值之间时计算得到的 COCO AP。

2.3 对比实验分析

自制数据集上不同算法的性能比较如表 2 所示。前 3 个算法基于 Faster-RCNN 架构，后 3 个算法基于 YOLOX 架构。综合分析可见，YOLOX 架构的算法在性能上优于 Faster-RCNN 架构，这验证了本研究选择 YOLOX 架构的合理性，并反映了采用较新网络架构所带来的明显性能提升。在基于 Faster-RCNN 的算法中，BIRANet 表现最优，与仅使用 RGB 图像信息的 FFPN 及仅使用雷达点锚框的 RANet 相比，在 AP 性能指标上提高了 4～5 个百分点。这表明，选择合适的策略来融合雷达和相机的信息，可以有效提升检测效果。同时，本文设计实现的 RV-YOLOX 算法与原始 YOLOX 算法相比，性能提升了约 3 个百分点。这也证实了，在 YOLOX 架构下，本研究所提出的注意力融合方式可以较好地融合雷达和相机的信息，取得良好的检测性能。

NuScenes 数据集上不同算法的性能比较如表 3 所示。首先，BIRANet 在 AP 指标上较仅使用 RGB 图像的 FFPN 提高了约 3.6%，这进一步证实了融合雷达信息可以有效提升检测效果的结论。其次，YOLOX 在小目标检测上表现出优势，AP_S 指标达到了 0.587。这一发现在自制数据集上并不明显，可能是由于两个数据集的分布差异，在自制数据集中小目标检测的难度更大，仅依靠视觉信息难以有效检测。此外，BIRANet 和 RV-YOLOX 之间的性能对比同样证实了雷达和 RGB 图像信息融合对提高检测性能的关键作用。总的来说，在 NuScenes 数据集上，算法间的表现差异与自制数据集的情况非常相似，可以得出雷达数据确实有助于提升检测效果，具体提升程度则取决于信息融合的策略和方式。本研究提出的 RV-YOLOX 算法正是对信息融合方法的一次有效探索。

2.4 消融实验分析

表 2 中，展示了 YOLOX、RV-YOLOX 和轻量级 RV-YOLOX 三个算法的消融实验。与仅使用 RGB 图像的 YOLOX 相比，RV-YOLOX 在 AP 值上提升了约 3.1%，在 AP_{50}和 AP_{75}上分别提升了 6.3%和 4.2%。值得注意的是，RV-YOLOX 的推理速度为 95ms，相比 YOLOX 的 76ms 有所下降。RV-YOLOX 在检测精度和推理速度方面的变化都是由于在网络中增加了雷达信息导致的。为了使 RV-YOLOX 能满足部署要求，本研究将图像分支的主干网络换成了 RepVGG 网络。可以看到轻量级 RV-YOLOX 在 AP 值上再次提升了 1%，推理速度则提高到了 80 ms。相较于 RV-YOLOX，轻量级 RV-YOLOX 保持了精度上的增益，在速度上也有明显的性能优势。

表 2 自制数据集上算法性能比较

评价指标 / 检测方法	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L	推理速度/ms
FFPN(RGB)[20]	0.530	0.789	0.681	0.383	0.545	0.584	182
RANet(Radar)[14]	0.524	0.779	0.643	0.398	0.552	0.567	158
BIRANet (RGB+Radar)[14]	0.571	0.822	0.713	0.444	0.598	0.637	232
YOLOX(RGB)[17]	0.594	0.829	0.738	0.439	0.600	0.646	**76**
RV−YOLOX (RGB+Radar)	0.625	0.892	0.780	0.489	0.648	0.679	95
轻量级 RV−YOLOX (RGB+Radar)	**0.635**	**0.892**	**0.785**	**0.490**	**0.654**	**0.687**	80

表 3 中，基于 YOLOX 架构的 3 个算法之间也展现了明显的性能差异。与仅使用 RGB 图像的 YOLOX 相比，RV-YOLOX 和轻量级 RV-YOLOX 通过融合雷达和 RGB 图像信息，在各项指标上均有显著提升。具体来说，RV-YOLOX 在 AP 值上提升了约 2.8%，0AP_{50}和 AP_{75}上分别提升了 3.4%和 1.3%。对于不同大小的目标，*RV-YOLOX* 的 AP_S、AP_M 和 AP_L 也分别提升了 4.1%、3.6%和 0.5%。

此外，通过进一步优化，轻量级 RV-YOLOX 在 AP 值上再次提升到了 0.4%，并在 AP_{50} 和 AP_S 上分别达到 0.927 和 0.632。总体而言，表 2 和表 3 中的消融实验充分证实了本研究提出的注意力融合方式在提高检测性能方面的关键作用，并验证了轻量级 RV-YOLOX 算法的有效性和实用性。

表 3　NuScenes 数据集上算法性能比较

检测方法 \ 评价指标	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
FFPN(RGB)[20]	0.697	0.882	0.820	0.503	0.680	0.731
RANet(Radar)[14]	0.690	0.839	0.801	0.448	0.678	0.733
BIRANet(RGB+Radar)[14]	0.723	0.889	0.843	0.535	0.701	0.769
YOLOX(RGB)[17]	0.734	0.891	0.886	0.587	0.750	0.789
RV-YOLOX (RGB+Radar)	0.762	0.925	0.899	0.628	0.786	0.794
轻量级 RV-YOLOX (RGB+Radar)	0.766	0.927	0.893	0.632	0.784	0.793

3　结束语

RV-YOLOX 是基于 YOLOX 网络架构提出的一种毫米波雷达和相机融合检测框架。该框架使用独立的图像特征提取分支和雷达特征提取分支，以分别提取 RGB 图像特征和雷达图像特征。为了充分利用雷达的空间位置信息引导 RGB 图像提取更有效的信息流，RV-YOLOX 中专门设计了雷达空间注意力模块，以将雷达信息传递给图像分支。此外，通过将图像特征提取分支更换为 RepVGG-A0 网络，便可构建轻量级 RV-YOLOX。在 NuScenes 数据集和自制数据集上与基线算法的比较分析显示，所提出的改进策略可以有效提升全天候复杂场景下的检测效果。未来工作中将考虑结合多传感器信息和文本信息，以实现一种基于对比学习的多模态网络架构，从而进一步提高道路交通场景中车辆的检测精度。

参考文献：

[1] 诸葛晶昌，李　想. 基于改进 YOLOv5s 的机坪特种车辆检测算法研究 [J]. 计算机测量与控制，2023，31 (6)：27－33.

[2] 郭昊琰，兰国峰，赵　辉，等. 基于线阵相机的高速小目标提取算法研究 [J]. 计算机测量与控制，2023，31 (2)：262－268.

[3] 罗　辉. 基于数据优先级的雷达目标跟踪偏差补偿方法 [J]. 计算机测量与控制，2020，28 (6)：222－225.

[4] 黄　兴，应群伟. 应用激光雷达与相机信息融合的障碍物识别 [J]. 计算机测量与控制，2020，28 (1)：184－188.

[5] DU X，ANG M H，KARAMAN S，et al. A general pipeline for 3d detection of vehicles [C] //2018 IEEE International Conference on Robotics and Automation (ICRA). IEEE，2018：3194－3200.

[6] LIANG M，YANG B，WANG S，et al. Deep continuous fusion for multi-sensor 3d object detection [C]. Proceedings of the European Conference on Computer Vision (ECCV). 2018：641－656.

[7] CUI Y，CHEN R，CHU W，et al. Deep learning for image and point cloud fusion in autonomous driving：A review [J]. IEEE Transactions on Intelligent Transportation Systems，2021，23 (2)：722－739.

[8] CAESAR H，BANKITI V，LANG A H，et al. nuscenes：A multimodal dataset for autonomous driving [C] //Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020：11621－11631.

[9] MEYER M，KUSCHK G. Automotive radar dataset for deep learning based 3d object detection [C] //2019 16th European Radar Conference (EuRAD). IEEE，2019：129－132.

[10] BARNES D，GADD M，MURCUTT P，et al. The oxford radar robotcar dataset：A radar extension to the oxford robotcar dataset [C] // 2020 IEEE International Conference on Robotics and Automation (ICRA). IEEE，2020：6433－6438.

[11] NOWRUZI F E，KOLHATKAR D，KAPOOR P，et al. Deep open space segmentation using automotive radar [C] // 2020 IEEE MTT-S International Conference on Microwaves for Intelligent Mobility (ICMIM). IEEE，2020：1－4.

[12] NABATI，RAMIN，HAIRONG QI. Rrpn：Radar region proposal network for object detection in autonomous vehicles [C] // 2019 IEEE International Conference on Image Processing (ICIP). IEEE，2019：3093－3097.

[13] GIRSHICK，ROSS. Fast r-cnn [C] // Proceedings of the IEEE International Conference on Computer Vision，2015：1440－1448.

[14] YADAV R，VIERLING A，BERNS K. Radar+ RGB fusion for robust object detection in autonomous vehicle [C] // 2020 IEEE International Conference on Image Processing (ICIP). IEEE，2020：1986－1990.

[15] NOBIS，FELIX，et al. A deep learning-based radar and camera sensor fusion architecture for object detection [J]. Sensor Data Fusion：Trends，Solutions，Applications (SDF). IEEE，2019：1－7.

[16] KOWOL K，ROTTMANN M，BRACKE S，et al. YOdar：uncertainty-based sensor fusion for vehicle detection with camera and radar sensors [J]. ArXiv Preprint ArXiv：2010.03320，2020：1－10.

[17] GE Z，LIU S，WANG，et al. Yolox：Exceeding yolo series in 2021 [J]. ArXiv Preprint ArXiv：2107.08430，2021：1－9.

[18] DING X，ZHANG X，MA N，et al. Repvgg：Making vgg-style convnets great again [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021：13733－13742.

[19] SIMONYAN K，ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [J]. ArXiv Preprint ArXiv：1409.1556，2014：1－10.

[20] REN S，HE K，GIRSHICK R，et al. Faster r-cnn：Towards real-time object detection with region proposal networks [J]. Advances in Neural Information Processing Systems，2015：91－99.