http://dx.doi.org/10.52810/JIR.2024.004

ARTICLE

YOLOv7-BW: 基于遥感图像的密集小目标高效检测器

葛旭东[1], **金学波**[1,*], **马慧鋆**[1], **邹天畅**[2]

[1] 北京工商大学，计算机与人工智能学院，北京 100048
[2] 考克大学食品科学专业，考克，爱尔兰，T12 HY8E

学术编辑： Shenglun Yi；**收稿日期：** 2024-03-07；**录用日期：** 2024-05-12；**发布日期：** 2024-05-30

***通讯作者：** ✉金学波，jinxuebo@btbu.edu.cn

文章引用
葛旭东, 金学波, 马慧鋆, 邹天畅. YOLOv7-BW: 基于遥感图像的密集小目标高效检测器. 智能机器人, 2024, **1**(1): 39–54.

Citation
Ge, X., Jin, X., Ma, H., & Zou, T. (2024). YOLOv7-BW: Efficient Detector for Dense Small Targets Based on Remote Sensing Images. Journal of Intelligent Robots, 1(1), 39–54.

摘　要

近年来，深度学习技术已经越来越广泛应用于遥感图像的检测。然而，遥感图像普遍目标大小差距大同时分布密集，对检测算法性能的要求高。目前的检测方法普遍效率低，容易出现漏检以及检测框不准确的情况。为此，本文基于 YOLO 算法进行改进，提出了一种基于 YOLOv7 的算法 YOLOv7-bw，实现了对遥感图像的高效率检测，促进了目标检测在遥感行业的应用和发展。YOLOv7-bw 在原始的池化金字塔 SPPCSPC 网络中添加了 Bi-level Routing Attention 模块，对目标集中区域重点关注，以提高网络提取特征的能力；并引入动态非单调的 WIoUv3 替换原本的 CIoU 损失函数，使得损失函数在每一时刻都能做出最符合当前情况的梯度增益分配策略，以提高对检测目标的聚焦能力。通过对 DIOR 遥感图像数据集进行对比实验发现，我们的 YOLOv7-bw 具有较高的 mAP@0.5 和 mAP@0.5：0.95，在数据集上表现为 85.63% 和 65.93%，高于 YOLOv7 源码的 83.7% 和 63.9% 分别 1.93%、2.03%。同时，对比目前常用算法，我们的 YOLOv7-bw 均表现更好，证明了我们提出的算法是可行的，可以更好的应用于遥感图像检测。

关键词: 遥感图像，YOLO，目标检测，mAP

YOLOv7-BW: Efficient Detector for Dense Small Targets Based on Remote Sensing Images

Xudong Ge[1], **Xuebo Jin**[1,*], **Huijun Ma**[1], **Tianchang Zou**[2]

[1]School of Computer and Artificial Intelligence, Beijing Technology and Business University, Beijing 100048, China
[1]University College Cork, Food Science Department, Cork, Ireland, T12 HY8E

Academic Editor: Shenglun Yi; **Submitted:** 2024-03-07; **Accepted:** 2024-05-12; **Published:** 2024-05-30

***Correspondence Author:** ✉Xuebo Jin, jinxuebo@btbu.edu.cn

Abstract

In recent years, deep learning techniques have been increasingly widely used in the detection of remote sensing images. However, the general target size gap of remote sensing image is large and densely distributed, which requires the performance of detection algorithm. The current detection methods are generally low efficiency, prone to miss detection and inaccurate detection box. Therefore, this paper improves the YOLO algorithm and proposes a YOLOv7-based algorithm YOLOv7-bw, which realizes the efficient detection of remote sensing images and promotes the application and development of target detection in the remote sensing industry.YOLOv7-bw adds Bi-level Routing Attention module to the original pooling pyramid SPPCSPC network, focusing on the target concentration area to improve the network ability to extract features; and introduces dynamic non-monotonic WIoUv3 to replace the original CIoU loss function, so that the loss function can make the gradient gain allocation strategy most consistent with the current situation at every moment, so as to improve the focus ability on the detection target.Through comparative experiments on the DIOR remote sensing image dataset, we found that our YOLOv7-bw had high mAP @ 0.5 and high mAP @ 0.5:0.95, which showed 85.63% and 65.93% on the dataset, higher than 83.7% and 63.9% about 1.93% and 2.03%, respectively. What’ s more, compared with the current commonly used algorithms, our YOLOv7-bw all performs better, which proves that our proposed algorithm is feasible and can be better applied to remote sensing image detection.

Keywords: Remote sensing image, YOLO, target detection, mAP

1 引言

光学遥感图像 [1] 是由空中飞行器或卫星对地面进行拍摄的一种俯视角度的图像，其拍摄方式较为特殊，导致成像效果与日常拍摄的图片差异巨大。随着遥感技术的不断升级，成像的效果越来越好，因此这也对遥感图像处理技术提出了更高的要求。遥感图像的应用领域众多，在军事方面和民用方面都具有重要价值；在军事方面，遥感图像数据可以用来对收集的情报和侦察的信息进行处理分析，根据结果，可以调整作战计划以及军事部署等各种情况。在民用方面，从土地利用 [2]、城市规划、交通监测 [3]、灾害防治 [4]、生态保护 [5] 和工业制造 [38–47] 等多方面应用都能得到较好的效果。

遥感图像具有覆盖区域极大、目标种类繁多、目标密集、背景复杂度高等特点，进而给检测任务带来了极大的挑战 [48, 49]。传统遥感图像检测方法大致可以分为 4 类：基于模板匹配的方法 [6]、基于形状纹理的方法 [7]、基于图像分割的方法 [8] 和基于视觉显著性的方法 [9]。从上面这些方法可以看出，传统的常规方法 [50–53] 一般是先构建出通用的目标模板，然后再进行全局图像匹配；或者先分割出潜在目标区域，然后仅利用简单的特征规则进行判别。此类方法检测结果中容易掺杂大量错误实例，检测结果精度偏低，适用范围小，只能用于简单统一的背景下对目标进行检测。

得益于数据的海量增长和硬件计算能力的提升，深度学习的理论与技术也迅速发展 [54–60]，越来越多的深度学习方法被应用于遥感图像目标检测领域。基于深度学习的目标检测算法可以根据是否生成区域建议 [10] 分为基于区域建议的方法 (双阶段方法) 和基于回归的方法 (单阶段方法)。双阶段目标检测器，如 Faster R-CNN[11]、Libra R-CNN[12]、Mask R-CNN[13] 等都先要提取感兴趣的区域，之后针对每个区域进行进一步的检测和识别。虽然检测精度整体较高，但由于需要首先提取感兴趣区域，并对每个区域分别进行分类和回归，增加了额外的计算量，速度上不够快，对于实时性要求较高的系统难以应用。而单阶段目标检测器不需要单独生成候选区域，将整个检测过程看为一个整体，从输入图像的多个位置直接回归分析出目标的边界框与类别，典型的代表算法有 YOLO 系列 [14–19]、SSD[20]、FCOS[21] 等。单阶段算法目标检测速度较快，基本满足实时系统的要求，但是检测精度略低于双阶段目标检测方法。总的来说，基于深度学习的方法可以通过训练自动获取图像的深层语义特征，具有比手工设计特征更强大的表达能力。此外，其对目标在图像中的空间和密集分布等因素较为敏感，而对目标的类别敏感度低。因此，基于深度学习的方法通常不针对单一种类目标进行检测，而是可以对多类目标进行检测，更符合遥感图像实际应用，成为遥感图像目标检测的主流发展方向。

近年来，随着遥感图像检测技术的提升，其应用范围也逐渐扩大，各种改进算法也层出不穷，检测精度与检测效率都有了大幅的提升。其中，Li 等人 [22] 提出了一种双通道特征融合网络，可以沿着两条独立的路径学习局部和上下文属性特征，从而形成一个强大的联合表示，实现对遥感图像目标的有效检测；Yang 等人 [23] 提出了一种端到端旋转检测框的目标检测算法，提高了舰船的检测精度；张等人 [24] 设计了一种基于 YOLOv5s 模型的多尺度检测网络结构，提高了监控场景下目标的检测性能；Jiang 等人 [25] 结合了双射神经网络和错位定位策略，解决了遥感小目标边界框窄小的问题；Wang 等人 [26] 在浅层和深层特征图之间建立了密集连接，解决了船舶尺度变化大的问题；Yang 等人 [27] 将多层特征与有效的锚点采样相融合，提高了对小物体的敏感度；Yao 等人 [28] 通过在特征金字塔网络中引入扩张的瓶颈结构，生成了高质量的语义特征；闫等人 [28] 通过跨层级通道特征融合，保留弱小目标精确的位置信息，改善了小目标检测的效果；张等人 [30] 通过将不同层级的特征进行融合获得多个感受野特征，并构建新的级联注意力机制，加强了对遥感小目标特征的捕获能力。

综合之前的分析，本篇文章采用 YOLOv7 作为基础算法，提出 YOLOv7-bw。为了在精度和密集目标预测任务上有更好的表现，YOLOv7 使用“扩展”和“缩放”。同时解决了动态标签分配分配问题和重新参数化模块的替换问题，使目标检测器更快、更有效。其次，根据遥感图像普遍拍摄距离远，成像模糊的问题，在损失函数的边界盒回归部分，采用了 wiouv3[31] 损失函数，更好的聚焦、定位所需检测的目标，提升检测精度。最后，针对遥感图像目标较小，容易造成目标聚集扎堆的问题。引入双级路由自注意力模块 (Bi-level Routing Attention,BRA)，更好的关注目标密集区域，解决算法对于密集区域小目标无法识别的问题。

2 相关工作

2.1 注意力机制

图像处理中的注意力机制已成为深度学习领域中流行且重要的技术之一，因其具有优秀的即插即用的便利性，被广泛应用于图像处理领域的各种深度学习模型中。注意力机制通过对输入特征进行加权处理，将模型的注意力集中于最重要的区域，以提升图像处理任务的准确性和性能。早年间的注意力机制基本思路都是先输入 Query、Key、Value；

根据 Query 和 Key 计算两者之间的相关性，得到注意力得分；再对注意力得分进行缩放后（除以维度的根号），softmax 归一化，得到权重系数；最后根据权重系数对 Value 值进行加权求和，得到 Attention Value，从而关注重点区域和忽略不相关区域。随着时间的推移，Vaswani 等人 [32] 首次在 NLP(Natural Language Processing) 领域应用自注意力机制，并将其成功引入计算机视觉领域中，展现出自注意力模型的巨大潜力。区别于普通的

注意力机制，自注意力机制减少了对外部信息的依赖，更擅长捕捉数据或特征的内部相关性。自注意力机制的关键点在于，Q、K、V 为同一变量，或者三者来源于同一个 X，三者同源。通过 X 找到 X 里面的关键点，从而更关注 X 的关键信息，忽略 X 的不重要信息。不是输入语句和输出语句之间的注意力机制，而是输入语句内部元素之间或者输出语句内部元素之间发生的注意力机制。然而，像 vanilla attention 这种典型的全局上下文建模的自注意力在所有空间位置上计算成对的特征的亲和性，会导致较高的计算负担和沉重的内存占用，特别是对于高分辨率的输入。因此，近年来关于自注意力模块的研究都致力于缓解这种高计算负担，越来越多的工作中开始引入不同的手工制作的稀疏模式，如 local attention[33]，将注意力放在 query 所在的当前窗口；axial stripe[34], 关注于 query 所在当前窗口的横向和纵向延申窗口；dilated window[35]，将每个窗口分为几个不同部分，关注最近的几个窗口中与当前窗口 query 所在部分位置相同的部分。

总的来说，这些引入手工制作稀疏模式的注意力机制有局限性，在不同场景下效果不同，对于遥感图像密集较小的特性，不能很好的发挥功效。此外，这些注意力通过不同的合并或选择策略减少了键/值令牌的数量，但这些键/值令牌将由图像上的所有查询共享，效率不高。

2.2 YOLO

YOLO[14–19] 系列是当前目标检测领域性能最优算法之一，其在识别准确率和识别速度方面有着显著优势，可以实时进行目标检测，因此在各行各业得到广泛应用。其中，YOLOv7[19] 在保持速度优势的同时，准确度几乎达到最佳水平。其结构主要分为三个模块：输入端、特征提取主干网络 (Backbone) 和检测头输出端 (Head)。YOLOv7 首先将输入的图片 resize 为 640x640 大小，然后将其输入到 backbone 网络中进行特征提取，再经过 head 层网络输出三个不同大小的特征图，分别用于检测大、中、小目标，最后通过 Rep 和 Conv 层输出预测结果。

YOLO 的实时检测器自成立以来已得到研究人员的广泛认可并应用于许多场景。它采用了由边界框回归 (Bounding Box Regression,BBR) 损失、分类损失和目标损失加权的损失函数来构建模型。迄今为止，这种结构仍然是目标检测任务中最有效的损失函数范式，其中 BBR 损失直接影响模型的定位性能。为了进一步提高模型的定位性能，需要设计一个良好的 BBR 损失。

IoU[36](Intersection over Union) 用于衡量目标检测任务中预测框与真实框的重叠程度，其计算方式是预测框与真实框的交集面积与并集面积的比值。然而，IoU 作为损失函数存在一个致命缺陷，即在边界框之间没有重叠时，反向传播的梯度消失。这导致训练过程中无法更新边界框之间的重叠区域宽度。为解决这一问题，已有的研究考虑了许多与边界框相关的几何因素，并构建了惩罚项。目前的边界框损失都是基于加法的损失，采用了 CIoU[37](Complete IoU)，这种损失函数符合目标框回归的机制，考虑了目标与预测之间的距离、重叠度、尺度和纵横比，使得目标框回归更加稳定，不会像 IoU 一样在训练过程中出现发散等问题。然而，YOLOv7 当前的缺点之一是 CIoU 在处理长宽比较大的边界框时仍存在一定的误差，可能导致边界框的准确性不佳。

综上所述，虽然 YOLOv7 作为目标检测算法在速度和准确度方面具有显著优势，但它仍然存在着一些缺点，如在处理长宽比较大的边界框时的精度问题。未来的研究可以进一步改进边界框损失函数和处理长宽比的方法，以提高 YOLOv7 的性能。

3 工作原理

3.1 WIoU

3.1.1 WIoUv1

BBR 损失的框图如图1所示。

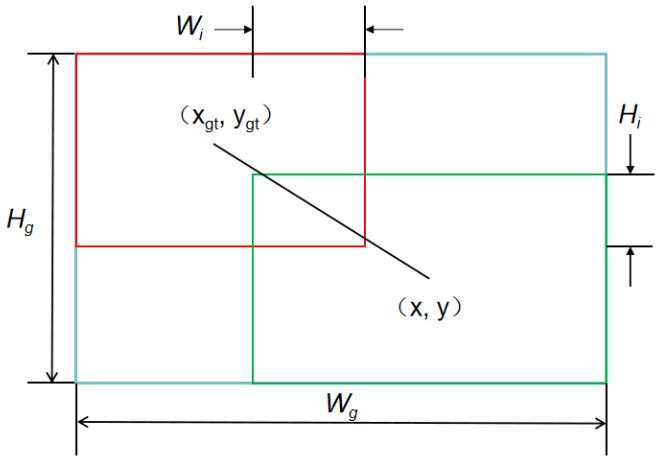


图 1. BBR 损失示例

YOLOv7 中使用的损失函数为 CIoU，其在中心点连接的归一化长度的基础上增加了对纵横比一致性的考虑，尽管解决了当负梯度 $\frac{\partial RDIoU}{\partial W_g}$ 与 $\frac{\partial LIoU}{\partial W_g}$ 抵消时，预测框无法优化的问题。但不可避免的是，预测过程中会产生许多低质量的锚框，所以加入纵横比或者距离等此类几何度量因素都会加剧低质量锚框的惩罚，从而使模型的泛化能力下降。一个好的损失函数应该在锚框与目标框较好地重合时削弱几何度量的惩罚，不过多地干预训练。因此，在这些基础之上，以距离度量构建了距离注意力，打破以往锚框损失都是基于加法的损失，二者相乘，得到了具有两层注意力机制的 $WIoUv1$：

$$L_{WIoUv1} = R_{\text{WIoU}} L_{\text{IoU}} \tag{1}$$

$$R_{WIoU} = \exp\left(\frac{(x-x_{gt})^2 + (y-y_{gt})^2}{\left(W_g^2 + H_g^2\right)^*}\right) \tag{2}$$

其中，L_{IoU} 是距离度量项，R_{WIoU} 是注意力项，将显著放大普通质量锚框的 L_{IoU}，将显著降低高质量锚框的 R_{WIoU}。为了防止 R_{WIoU} 产生阻碍收敛的梯度，将 W_g，H_g 从计算图（上标 * 表示此操作）中分离。

3.1.2 WIoUv3

尽管 $WIoUv1$ 已经可以应用于大多数场景，但对于遥感图像目标小的问题还不能很好的解决。为了更好的聚焦小的目标，选择使用在 v1 基础上通过构造梯度增益（聚焦系数）的计算方法来附加聚焦机制的 $WIoUv3$ 来替换 YOLOv7 中的 CIoU 损失函数。$WIoUv3$ 定义了离群度以描述锚框的质量，其具体定义为：

$$\beta = \frac{L_{IoU}^*}{\overline{L_{IoU}}} \in [0, +\infty) \tag{3}$$

其中，$\overline{LIoU}$ 代表动量为 m 的滑动平均值。离群度小意味着锚框质量高，为其分配一个小的梯度增益。同时，对离群度较大的锚框也分配较小的梯度增益，有效防止低质量示例产生较大的有害梯度，最后使边界框回归聚焦到普通质量的锚框上。利用 β 构造了一个非单调聚焦系数并将其应用于 $WIoUv1$：

$$L_{WIoUv3} = rL_{WIoUv1}, \quad r = \frac{\beta}{\delta \alpha^{\beta-\delta}} \tag{4}$$

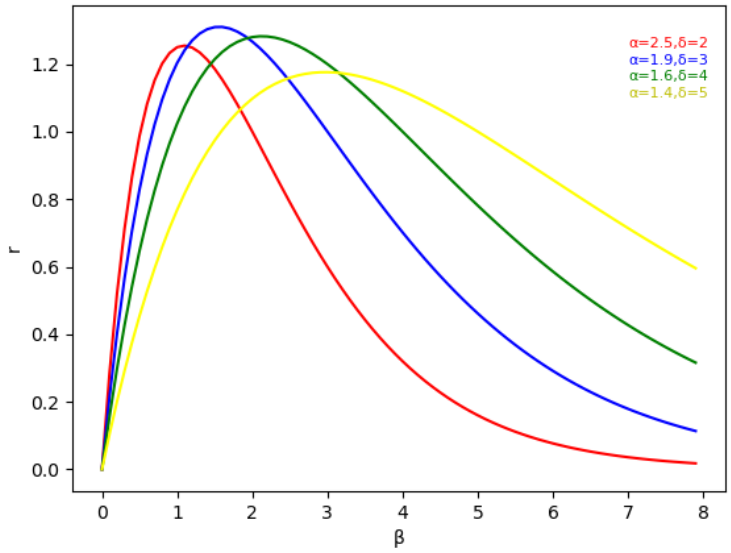


图 2. 由超参数 α、δ 控制的 β 和梯度增益 r 的映射

经过我们绘制出几组不同的 α 和 δ 对离群度 β 和梯度增益 r 的影响后，如图2 所示，可以看出蓝色曲线拥有较好的性能，在离群度低和高的时候都有较小的梯度增益，使损失函数更关注于普通质量的锚框，最终选定超参数 $\alpha = 1.9$ 以及 $\delta = 3$ 应用到最后的实验（超参数确定的实验在第 3 部分给出）。同时为了预防低质量的锚框在早期训练被留下，我们初始化了 $\overline{LIoU} = 1$，使得当 $LIoU = 1$ 时，具有最大的梯度增益。

遥感图像本身就具有目标小的特点，尤其很多时候会受天气影响，目标还可能会在阴影当中，这极大的加大了检测的难度，图3中 (a) 即为 YOLOv7 源码跑出的检测结果，结果里漏掉了 3 辆小车没有检测出来。将设置好的 $WIoUv3$ 应用到 YOLOv7 后，虽然还是漏检了一辆小车，但因为其独特的动态非单调聚焦机制，可以看到 (b) 中无论是对于物体检测的置信度，还是阴影中模糊的小目标，都有了一定程度的改善。

(a) (b)

图 3. 替换 WIoU 对 YOLO 的影响

3.2 Bi-level Routing Attention（BRA）

前文中提到的自注意力模块大都是引入一些手工制作的稀疏模式。但是，当它们使用不同的策略来合并或选择键、值标记时，这些标记是与查询无关的，也就是说，它们被所有查询共享。然而，不同语义区域的查询实际上关注着相当不同的键值对。因此，强制所有查询关注同一组令牌可能是次优的。区别于其他注意力模块，BRA

是一种动态的、查询感知的稀疏注意机制，目标是让每个查询都关注语义上最相关的键值对中的一小部分；其具体核心思想是在粗区域级别上先过滤掉最不相关的键值对，这样就只保留一小部分路由区域。然后，在这些路由区域的并集中应用细粒度的标记到标记的注意。因为 BRA 只涉及密集矩阵乘法，所以在具有良好的性能的同时还兼顾高计算效率。其具体步骤大致可以分为以下 3 个部分：

1）区域划分和输入投影

输入一张二维特征图，$X \in \mathbb{R}^{H\times W\times C}$，首先将其划分为 $S\times S$ 个不重叠的区域，其中每个区域包含 $\frac{HW}{S^2}$ 个特征向量，即将 X 变为 $X^r \in \mathbb{R}^{S^2\times \frac{HW}{S^2}\times C}$。然后推导出查询 Q、键 K 和值 V 的线性投影为：

$$Q = X^r W^q, \quad K = X^r W^k, \quad V = X^r W^v \tag{5}$$

其中，$W^q, W^k, W^v \in \mathbb{R}^{C\times C}$ 分别是 query, key, value 的投影权重。

2）具有有向图的区域到区域的路由

在第一步的基础上，我们通过构造一个有向图来找到参与关系。先通过分别对 Q 和 K 应用每个区域的平均值，得到 Q^r、$K^r \in \mathbb{R}^{S^2\times C}$，然后计算 Q^r 和 K^r 的区域间相关性的邻接矩阵 A^r：

$$A^r = Q^r (K^r)^T \tag{6}$$

邻接矩阵 $A^r \in \mathbb{R}^{S^2\times S^2}$，度量了两个区域在语义上的相关程度。接下来只保留每个区域的前 k 个连接来修剪相关性图，具体来说，以路由索引矩阵 $I^r \in \mathbb{N}^{S^2\times k}$，使用行向算子 $topk$，逐行保存前 k 个连接的索引：

$$I^r = \text{topkIndex}(A^r) \tag{7}$$

其中，I^r 的第 i 行包含第 i 个区域的前 k 个最相关区域的索引。

3）令牌到令牌的注意

利用区域到区域的路由索引矩阵 I^r，我们就可以应用细粒度的标记注意。对于区域 i 中的每个查询标记，它将关注所有位于 k 个路由区域的并集中的键值对。具体来说，先收集键和值张量：

$$\begin{aligned} K^g &= \text{gather}(K, I^r) \\ V^g &= \text{gather}(V, I^r) \end{aligned} \tag{8}$$

其中，K^g 和 V^g 是聚集后的 key 与 value 的张量，然后再对聚集后的键值对使用注意力操作：

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{C}}\right)V \tag{9}$$

$$O = \text{Attention}(Q, K^g, V^g) + \text{LCE}(V) \tag{10}$$

此处引入了一个上下文增强项 $LCE(V)$，函数 $LCE(\bullet)$ 用深度可分离卷积进行参数化，我们将卷积核大小设置为 5。

总的来说，BRA 通过在前 k 个相关窗口中收集键值对，利用稀疏性跳过最不相关区域的计算，而只涉及对 gpu 友好的密集矩阵乘法，如图4所示：

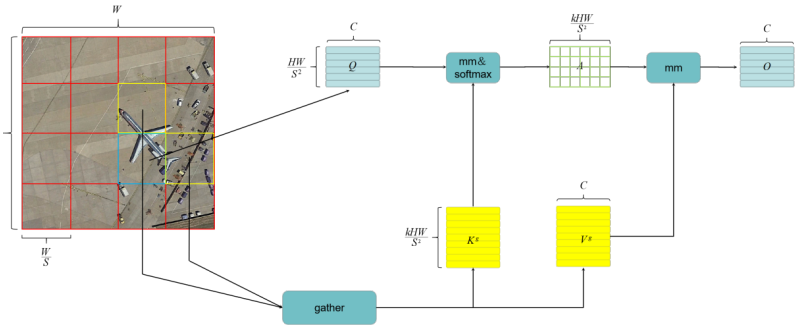


图 4. BRA 原理示意图

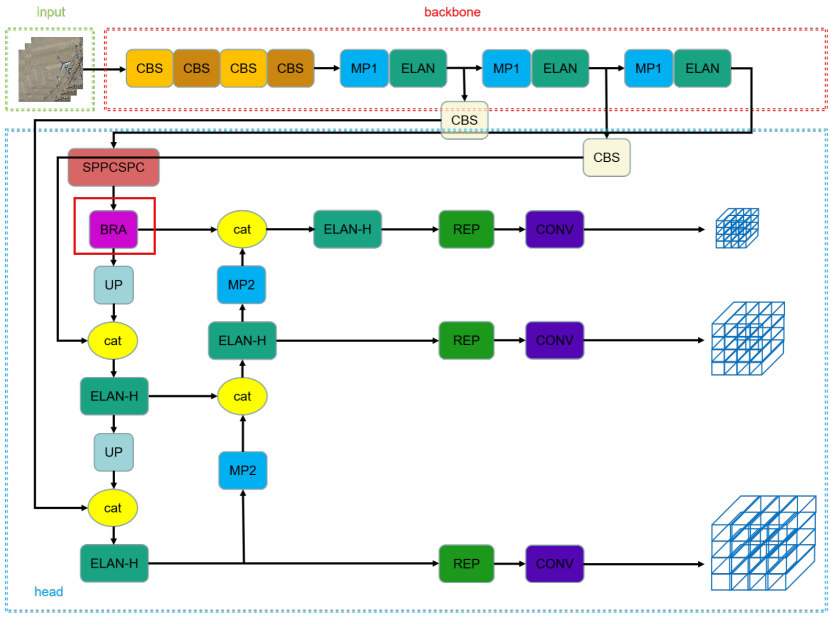


图 5. BRA 在 YOLOv7 中位置

具体将 BRA 插入到 YOLOv7 的网络中，它是先对之前提取的特征图进行区域到区域路由，再应用令牌到令牌的注意，得到新的输出特征图。考虑到其是对特征图进行操作，又因为空间池化金字塔 SPPCSPC 的作用就是避免由于图像处理操作所造成的图像失真以及以及图片特征重复提取，所以决定在 SPPCSPC 后插入 BRA 模块，具体插入 YOLOv7 网络中，位置如图5所示。

遥感图像中许多情况便是密集的小目标扎堆在某一个区域，BRA 正是对某个区域重点关注，应用细粒度级别的注意机制，符合遥感图像特点。实际情况如图6所示，左图 (a) 中为 YOLOv7 源码跑出的检测结果，右图 (b) 中为在源码基础上添加 BRA 模块的检测结果。可以看出，对于密集成堆的小车，添加 BRA 模块可以让检测出来的小车变多，同时，错检的屋顶数量也更少，对整体准确度有一定的提升。

4 实验

4.1 评价指标

为了验证模型的有效性，我们使用了 TP(True Positive)、TN(True Negative)、FP(False Positive)、FN(False Negative)、精度 (Precision)、召回率 (Recall)、AP(Average Precision)、P-R 曲线、mAP(mean Average Precision) 评价指标，其中 TP、TN、FP、FN 表示样本类别与其预测结果的关系:

TP：样本的真实类别是正例，并且模型预测的结果也是正例。

(a) (b)

图 6. 添加 BRA 对 YOLO 的影响

TN：样本的真实类别是负例，并且模型预测的结果成为负例。

FP：样本的真实类别是负例，但是模型预测的结果成为正例。

FN：样本的真实类别是正例，但是模型预测的结果成为负例。

因此，精度即为预测的真阳性样本占整个预测结果为正例样本的百分比，计算结果为：

$$\text{Precision} = \frac{TP}{TP + FP} \tag{11}$$

召回率即为预测的真阳性样本占整个真实结果为正例样本的百分比，计算结果为：

$$\text{Recall} = \frac{TP}{TP + FN} \tag{12}$$

P-R 曲线代表的是精度与召回率的关系，将 Recall 设置为横坐标，Precision 设置为纵坐标，绘制的某一类别的曲线，同时，此曲线下围成的面积即为该类别的 AP 值。

mAP 为目标检测算法对某一数据集的整体评价指标，是所有类别的 AP 值的平均值，是目前评价模型性能好坏的最主要指标，表示为：

$$mAP = \frac{1}{C}\sum_{i=0}^{C} AP_i \tag{13}$$

4.2 WIoU 超参数实验

为了检验什么样的超参数最适合应用到 WIoUv3，我们设计了一组对比试验，基于 PyTorch 框架。选择 MS-COCO 数据集中的 20 个类别，28474 张图像作为训练数据，1219 张图像作为验证数据。对于模型，我们选择 YOLOv7-w6，层通道倍数为 0.75 进行训练。这些模型被训练为 120 个 epoch，批次大小为 32 和不同的 BBR 损失。实验结果如表1所示：

由表 1 我们可以看出，拥有动态非单调聚焦的 WIoU v3 的表现效果均好于仅有聚焦机制的 WIoU v1。同时，当 $\alpha = 1.9, \delta = 3$ 时，WIoU v3 有着在不同 IoU 阈值下最好的表现，分别为 54.50、64.20、45.68。最终，我们决定以 $\alpha = 1.9, \delta = 3$，为最后的超参数设定参与之后的实验。

表 1. 超参数实验对比

	Ap^{val} (75)	Ap^{val} (50)	Ap^{val}
WIoU v1	52.82	63.15	44.87
WIoU v3($\alpha = 1.4, \delta = 5$)	53.75	64.05	45.15
WIoU v3($\alpha = 1.6, \delta = 4$)	53.91	64.16	45.44
WIoU v3($\alpha = 1.9, \delta = 3$)	**54.50**	**64.20**	**45.68**

表 2. 数据集类别统计

C1	C2	C3	C4	C5
Express-toll-station	vehicle	golffield	trainstation	chimney
C6	C7	C8	C9	C10
storagetank	ship	harbor	airplane	groundtrackfield
C11	C12	C13	C14	C15
Expressway-Service-area	dam	basketballcourt	tenniscourt	stadium
C16	C17	C18	C19	C20
baseballfield	windmill	bridge	airport	overpass

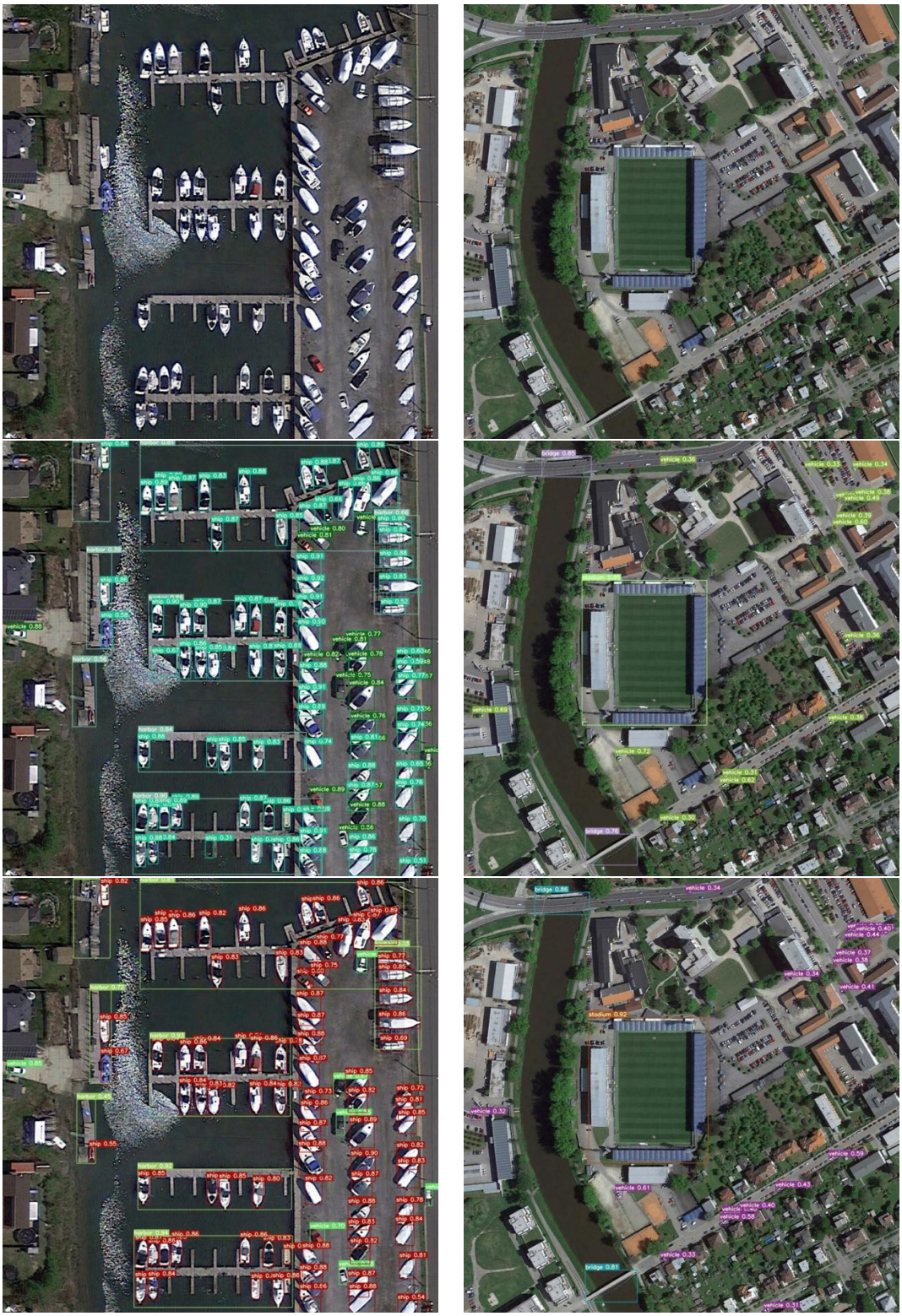
表 3. 对比试验

method	Backbone	mAP
R-CNN	VGG-16	37.7
Faster R-CNN	VGG-16	54.1
SSD	VGG-16	58.6
RetinaNet	ResNet-101	66.1
PANet	ResNet-101	66.1
CornerNet	Hourglass-104	64.9
YOLOv7	Elan	83.7
YOLOv7-bw (ours)	Elan	85.6

4.3 消融实验

为了验证各个模块对模型的有效性，我们设置了一组消融实验。数据集采用由地球观测解释领域的专家从谷歌 Earth 采集的 DIOR 遥感图像数据集，包含 23463 张遥感图像和 190288 个目标实例，这些目标实例用轴向对齐的边界框手动标记，覆盖 20 个常见对象类别，分别为飞机、机场、棒球场、篮球场、桥梁、烟囱、水坝、高速公路服务区、高速公路收费站、港口、高尔夫球场、地面田径场、天桥、船舶、体育场、储罐、网球场、火车站、车辆和风磨，其中每个类别对应的编号如表2所示。

我们挑选了几张数据集测试集内的图片，对 YOLOv7 源码以及我们改进过后的 YOLOv7-bw 进行实际效果对比，对比效果如图7所示。其中第一张为数据集中原始图片，第二张为 YOLOv7 效果演示，第三张为我们的 YOLOv7-bw 效果演示。从 (a) 图片可以看出，YOLOv7-bw 对比 YOLOv7，更少的将船错检为车；从 (b) 图片可以看出，虽然还有许多车辆未被检测出，但 YOLOv7-bw 检测出来了 18 辆车，高于 YOLOv7 检测出的 15 辆车，对于车辆的定位、聚焦有了更好的效果。



(a) (b)

图 7. YOLOv7 和 YOLOv7-bw 检测结果

4.4 与其他检测方法的比较

为了进一步验证本文改进的 YOLOv7 的效果, 选取目标检测领域常用算法与本文的改进进行比较, 包括 RCNN、SSD、RetinaNet、CornerNet 等经典算法。数据集依旧选取实验 3.3 节中的 DIOR 遥感数据集, 在相同条件下

进行训练、测试，结果如表3所示：

从表中可以看出，我们提出的 YOLOv7-bw 算法，在 IoU 阈值为 0.5 的情况下，拥有着最佳的 mAP0.5，意味着改进的 YOLOv7-bw 算法优化了遥感图像的检测效果，使得总体检测性能有所提高，让模型各有竞争性，特别是对小目标以及密集目标优化效果明显，满足遥感图像的实际检测的需求。

5 结束语

为了解决遥感图像目标密集、模糊的问题，本文基于 YOLOv7 模型提出了 YOLOv7-bw 算法，引入 BRA 注意力机制，关注目标密集区域；此外将损失函数替换为 WIoUv3，更好聚焦、定位待检测目标，并在 DIOR 遥感图像数据集上测试，试验结果表明：我们 YOLOv7-bw 的 mAP0.5 和 mAP0.5：0.95 分别为 85.63% 和 65.93%，均为最优结果，证明我们的算法是可行的。但实验过程中，我们也发现了算法的不足，例如图 9(b) 中的检测结果，虽然整体检测出车辆的数目多于 YOLOv7，但还有许多微小型车辆未被检测出，可以进行优化。在未来的研究中，我们将更致力于研究模型如何更好识别到微小目标，使其在遥感图像领域有更好的发挥。

参考文献

[1] Nie, G. T., & Huang, H. (2021). A survey of object detection in optical remote sensing images. Acta Automatica Sinica, 47(8), 1749-1768.

[2] Chen, X. L., Zhao, H. M., Li, P. X., & Yin, Z. Y. (2006). Remote sensing image-based analysis of the relationship between urban heat island and land use/cover changes. Remote sensing of environment, 104(2), 133-146. [CrossRef]

[3] Lenhart, D. O. M. I. N. I. K., Hinz, S. T. E. F. A. N., Leitloff, J. E. N. S., & Stilla, U. (2008). Automatic traffic monitoring based on aerial image sequences. Pattern Recognition and Image Analysis, 18, 400-405. [CrossRef]

[4] Liu, Y., & Wu, L. (2016). Geological disaster recognition on optical remote sensing images using deep learning. Procedia Computer Science, 91, 566-575.[CrossRef]

[5] Frohn, R. C. (2018). Remote sensing for landscape ecology: new metric indicators for monitoring, modeling, and assessment of ecosystems. CRC Press.

[6] Liu, G., Sun, X., Fu, K., & Wang, H. (2012). Aircraft recognition in high-resolution satellite images using coarse-to-fine shape prior. IEEE Geoscience and Remote Sensing Letters, 10(3), 573-577. [CrossRef]

[7] Liu, Q., Xiang, X., Wang, Y., Luo, Z., & Fang, F. (2020). Aircraft detection in remote sensing image based on corner clustering and deep learning. Engineering Applications of Artificial Intelligence, 87, 103333. [CrossRef]

[8] Zhu, C., Zhou, H., Wang, R., & Guo, J. (2010). A novel hierarchical method of ship detection from spaceborne optical image based on shape and texture features. IEEE Transactions on geoscience and remote sensing, 48(9), 3446-3456. [CrossRef]

[9] Bi, F., Zhu, B., Gao, L., & Bian, M. (2012). A visual search inspired computational model for ship detection in optical satellite images. IEEE Geoscience and Remote Sensing Letters, 9(4), 749-753. [CrossRef]

[10] Hosang, J., Benenson, R., Dollár, P., & Schiele, B. (2015). What makes for effective detection proposals?. IEEE transactions on pattern analysis and machine intelligence, 38(4), 814-830. [CrossRef]

[11] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in neural information processing systems, 28. [CrossRef]

[12] Pang, J., Chen, K., Shi, J., Feng, H., Ouyang, W., & Lin, D. (2019). Libra r-cnn: Towards balanced learning for object detection. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 821-830). [CrossRef]

[13] Nie, X., Duan, M., Ding, H., Hu, B., & Wong, E. K. (2020). Attention mask R-CNN for ship detection and segmentation from remote sensing images. Ieee Access, 8, 9325-9334. [CrossRef]

[14] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788). [CrossRef]

[15] Redmon, J., & Farhadi, A. (2017). YOLO9000: better, faster, stronger. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 7263-7271). [CrossRef]

[16] Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767. [CrossRef]

[17] Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934. [CrossRef]

[18] Ge, Z., Liu, S., Wang, F., Li, Z., & Sun, J. (2021). Yolox: Exceeding yolo series in 2021. arXiv preprint arXiv:2107.08430. [CrossRef]

[19] Wang, C. Y., Bochkovskiy, A., & Liao, H. Y. M. (2023). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 7464-7475). [CrossRef]

[20] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016). Ssd: Single shot multibox detector. In Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14 (pp. 21-37). Springer International Publishing. [CrossRef]

[21] Tian, Z., Shen, C., Chen, H., & He, T. (1904). FCOS: Fully convolutional one-stage object detection. arXiv 2019. arXiv preprint arXiv:1904.01355.

[22] Li, K., Cheng, G., Bu, S., & You, X. (2017). Rotation-insensitive and context-augmented object detection in remote sensing images. IEEE Transactions on Geoscience and Remote Sensing, 56(4), 2337-2348. [CrossRef]

[23] Yang, X., Sun, H., Sun, X., Yan, M., Guo, Z., & Fu, K. (2018). Position detection and direction prediction for arbitrary-oriented ships via multitask rotation region convolutional neural network. IEEE access, 6, 50839-50849. [CrossRef]

[24] Xu, S. Y., Chu, K. B., Zhang, J., & Feng, C. T. (2022). An improved YOLOv3 algorithm for small target detection. Electro-Opt. Control, 29, 35-39.

[25] Jiang, S., Yao, W., Wong, M. S., Li, G., Hong, Z., Kuc, T. Y., & Tong, X. (2020). An optimized deep neural network detecting small and narrow rectangular objects in Google Earth images. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 13, 1068-1081. [CrossRef]

[26] Wang, Y., Li, W., Li, X., & Sun, X. (2018, August). Ship detection by modified RetinaNet. In 2018 10th IAPR workshop on pattern recognition in remote sensing (PRRS) (pp. 1-5). IEEE. [CrossRef]

[27] Yang, X., Yang, J., Yan, J., Zhang, Y., Zhang, T., Guo, Z., ... & Fu, K. (2019). Scrdet: Towards more robust detection for small, cluttered and rotated objects. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 8232-8241). [CrossRef]

[28] Yao, Q., Hu, X., & Lei, H. (2020). Multiscale convolutional neural networks for geospatial object detection in VHR satellite images. IEEE Geoscience and Remote Sensing Letters, 18(1), 23-27. [CrossRef]

[29] Junhua, Y. A. N., Zhang, K., & Tianjun, S. H. I. (2022). Multi-level feature fusion based dim small ground target detection in remote sensing images. Chinese Journal of Scientific Instrument, 43(03), 221-229.

[30] Zhang Y Z, Guo W, Li W B. Omnidirectional accurate detection algorithm for dense small objects in remote sensing images, 2023: 1-9.

[31] Tong, Z., Chen, Y., Xu, Z., & Yu, R. (2023). Wise-IoU: bounding box regression loss with dynamic focusing mechanism. arXiv preprint arXiv:2301.10051. [CrossRef]

[32] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in neural information processing systems, 30. [CrossRef]

[33] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., ... & Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 10012-10022). [CrossRef]

[34] Dong, X., Bao, J., Chen, D., Zhang, W., Yu, N., Yuan, L., ... & Guo, B. (2022). Cswin transformer: A general vision transformer backbone with cross-shaped windows. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 12124-12134). [CrossRef]

[35] Wang, W., Chen, W., Qiu, Q., Chen, L., Wu, B., Lin, B., ... & Liu, W. (2023). Crossformer++: A versatile vision transformer hinging on cross-scale attention. IEEE Transactions on Pattern Analysis and Machine Intelligence. [CrossRef]

[36] Yu, J., Jiang, Y., Wang, Z., Cao, Z., & Huang, T. (2016, October). Unitbox: An advanced object detection network. In Proceedings of the 24th ACM international conference on Multimedia (pp. 516-520). [CrossRef]

[37] Zheng, Z., Wang, P., Liu, W., Li, J., Ye, R., & Ren, D. (2020, April). Distance-IoU loss: Faster and better learning for bounding box regression. In Proceedings of the AAAI conference on artificial intelligence (Vol. 34, No. 07, pp. 12993-13000). [CrossRef]

[38] Fang, F. A. N. G., Tan, W., & Liu, J. Z. (2005). Tuning of coordinated controllers for boiler-turbine units. Acta Automatica Sinica, 31(2), 291-296.

[39] Fang, F., Jizhen, L., & Wen, T. (2004). Nonlinear internal model control for the boiler-turbine coordinate systems of power unit. PROCEEDINGS-CHINESE SOCIETY OF ELECTRICAL ENGINEERING, 24(4), 195-199.

[40] Lv, Y., Lv, X., Fang, F., Yang, T., & Romero, C. E. (2020). Adaptive selective catalytic reduction model development using typical operating data in coal-fired power plants. Energy, 192, 116589.

[41] Fang, F., & Xiong, Y. (2014). Event-driven-based water level control for nuclear steam generators. IEEE Transactions on Industrial electronics, 61(10), 5480-5489.

[42] Liu, J., Zeng, D., Tian, L., Gao, M., Wang, W., Niu, Y., & Fang, F. (2015). Control strategy for operating flexibility of coal-fired power plants in alternate electrical power systems. Proceedings of the CSEE, 35(21), 5385-5394.

[43] Liu, J., Song, D., Li, Q., Yang, J., Hu, Y., Fang, F., & Joo, Y. H. (2023). Life cycle cost modelling and economic analysis of wind power: A state of art review. Energy Conversion and Management, 277, 116628.

[44] Wang, W., Liu, J., Zeng, D., Fang, F., & Niu, Y. (2020). Modeling and flexible load control of combined heat and power units. Applied Thermal Engineering, 166, 114624.

[45] Wei, L., & Fang, F. (2016). H_∞-LQR-Based Coordinated Control for Large Coal-Fired Boiler–Turbine Generation Units. IEEE Transactions on Industrial Electronics, 64(6), 5212-5221.

[46] Zhang, J., Feng, J., Zhou, Y., Fang, F., & Yue, H. (2012). Linear active disturbance rejection control of waste heat recovery systems with organic Rankine cycles. Energies, 5(12), 5111-5125.

[47] Liu, J., Wang, Q., Song, Z., & Fang, F. (2021). Bottlenecks and countermeasures of high-penetration renewable energy development in China. Engineering, 7(11), 1611-1622.

[48] Wang, N., Fang, F., & Feng, M. (2014, May). Multi-objective optimal analysis of comfort and energy management for intelligent buildings. In The 26th Chinese control and decision conference (2014 CCDC) (pp. 2783-2788). IEEE.

[49] Lv, Y., Fang, F. A. N. G., Yang, T.,& Romero, C. E. (2020). An early fault detection method for induced draft fans based on MSET with informative memory matrix selection. ISA transactions, 102, 325-334.

[50] Fang, F., & Wu, X. (2020). A win–win mode: The complementary and coexistence of 5G networks and edge computing. IEEE Internet of Things Journal, 8(6), 3983-4003.

[51] Fang, F., Zhu, Z., Jin, S., & Hu, S. (2020). Two-layer game theoretic microgrid capacity optimization considering uncertainty of renewable energy. IEEE Systems Journal, 15(3), 4260-4271.

[52] Zhang, X., Fang, F., & Liu, J. (2019). Weather-classification-MARS-based photovoltaic power forecasting for energy imbalance market. IEEE Transactions on Industrial Electronics, 66(11), 8692-8702.

[53] Liu, Y., Fang, F., & Park, J. H. (2018). Decentralized dissipative filtering for delayed nonlinear interconnected systems based on T–S fuzzy model. IEEE Transactions on Fuzzy Systems, 27(4), 790-801.

[54] Cheng, L., Kalapgar, A., Jain, A., Wang, Y., Qin, Y., Li, Y., & Liu, C. (2022). Cost-aware real-time job scheduling for hybrid cloud using deep reinforcement learning. Neural Computing and Applications, 34(21), 18579-18593.

[55] Guo, J., Cheng, L., & Wang, S. (2023). CoTV: Cooperative control for traffic light signals and connected autonomous vehicles using deep reinforcement learning. IEEE Transactions on Intelligent Transportation Systems.

[56] Mao, Y., Sharma, V., Zheng, W., Cheng, L., Guan, Q., & Li, A. (2022). Elastic resource management for deep learning applications in a container cluster. IEEE Transactions on Cloud Computing.

[57] Mao, Y., Fu, Y., Zheng, W., Cheng, L., Liu, Q., & Tao, D. (2021). Speculative container scheduling for deep learning applications in a kubernetes cluster. IEEE Systems Journal, 16(3), 3770-3781.

[58] Liang, S., Liu, C., Wang, Y., Li, H., & Li, X. (2020, November). Deepburning-gl: an automated framework for generating graph neural network accelerators. In Proceedings of the 39th International Conference on Computer-Aided Design (pp. 1-9).

[59] Li, W., Wang, Y., Li, H., & Li, X. (2019, January). P3M: a PIM-based neural network model protection scheme for deep learning accelerator. In Proceedings of the 24th Asia and South Pacific Design Automation Conference (pp. 633-638).

[60] Liu, B., Chen, X., Wang, Y., Han, Y., Li, J., Xu, H., & Li, X. (2019, January). Addressing the issue of processing element under-utilization in general-purpose systolic deep learning accelerators. In Proceedings of the 24th Asia and South Pacific Design Automation Conference (pp. 733-738).

葛旭东，2024 年毕业于北京工商大学控制工程专业，获硕士学位。研究方向为图像检测模式识别与信息融合、机器学习等。

Xudong Ge, graduated from Beijing University of Technology and Business in 2024 with a master's degree in Control engineering. His research focuses on image detection, pattern recognition and information fusion, machine learning, and other related fields.

金学波 教授，博士生导师.1994 年毕业于吉林大学（原吉林工业大学）获学士学位，1997 年毕业于吉林大学（原吉林工业大学）获硕士学位，2004 年获得浙江大学控制科学与工程博士学位，导师为孙优贤院士. 研究方向为信息融合、模式识别与预测、大数据分析、深度学习等. 近年来在相关领域主持了 1 项国家科技支撑计划课题、4 项国家自然基金面上项目等多项研究课题. 获 2021 年度中国粮油学会科学技术奖一等奖。在时序信号模式识别、图像目标检测与识别等研究领域，已发表 SCI、EI 收录等高水平学术论文 159 篇，其中 7 篇为 ESI 高被引论文（前 1%）、3 篇 ESI 热点论文（前 0.1%），已授权国家发明专利 20 余项，出版关于传感器信号识别与状态估计、多传感器信息融合的学术专著 3 部. 担任 SCI 收录期刊 Sensors 编委，为 IEEE/CAA Journal of Automatica Sinica、Knowledge-Based Systems 等中科院一区 SCI 期刊审稿人.

Xuebo Jin (Fellow, ASP) received the B.S. and M.S. degrees in control theory and control engineering from Jilin University, Changchun, China, in 1994 and 1997, and the Ph.D. degree in control theory and control engineering from the University of Zhejiang, Zhejiang, China, in 2004.,She was a Senior Visiting Scholar with the University of Illinois at Chicago, Chicago, IL, USA, in 2007. From 2009 to 2012, she was an Assistant Professor with Zhejiang Sci-tech University. Since 2012, she has been a Professor with Beijing Technology and Business University, Beijing, China. Her research includes a variety of areas in information fusion, big data analysis, condition estimation, and video tracking.

马慧鋆 2010 年毕业于长春光机学院原子与分子物理专业，获硕士学位，目前为北京工商大学系统科学专业在职博士生。研究方向为复杂系统建模、模式识别与信息融合、机器学习等。

Huijun Ma graduated from Changchun Institute of Optics and Mechanics with a master's degree in atomic and molecular physics in 2010. She is currently an on-the-job doctoral student in systems science at Beijing Technology and Business University. Research directions include complex system modeling, pattern recognition and information fusion, machine learning, etc.

邹天畅 就读于爱尔兰考克大学食品科学专业。研究方向为食品感官评测和产品配方改造。

Tianchang Zou, studied food science at University College Cork, Ireland. The research direction is food sensory evaluation and product formula modification.