



## ObjectBoxG:基于GC3模块的目标检测算法

张建宇, 谢娟英

引用本文:

张建宇, 谢娟英. ObjectBoxG:基于GC3模块的目标检测算法[J]. *智能系统学报*, 2024, 19(6): 1385-1394.

ZHANG Jianyu, XIE Juanying. ObjectBoxG: object detection algorithm based on GC3 module[J]. *CAAI Transactions on Intelligent Systems*, 2024, 19(6): 1385-1394.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202310025>

## 您可能感兴趣的其他文章

### 改进Faster R-CNN的汽车仪表指针实时检测

Improved Faster R-CNN vehicle instrument pointer real-time detection algorithm  
*智能系统学报*. 2021, 16(6): 1056-1063 <https://dx.doi.org/10.11992/tis.202011003>

### 基于改进的Faster RCNN面部表情检测算法

Facial expression recognition based on improved Faster RCNN  
*智能系统学报*. 2021, 16(2): 210-217 <https://dx.doi.org/10.11992/tis.201910020>

### 基于级联宽度学习的多模态材质识别

Cascade broad learning for multi-modal material recognition  
*智能系统学报*. 2020, 15(4): 787-794 <https://dx.doi.org/10.11992/tis.201908021>

### 联合外形响应的深度目标追踪器

A deep object tracker with outline response map  
*智能系统学报*. 2019, 14(4): 725-732 <https://dx.doi.org/10.11992/tis.201807029>

### 自动驾驶场景下小且密集的交通标志检测

Detection of small and dense traffic signs in self-driving scenarios  
*智能系统学报*. 2018, 13(3): 366-372 <https://dx.doi.org/10.11992/tis.201706040>

### 高斯核函数卷积神经网络跟踪算法

Convolutional neural network tracking algorithm accelerated by Gaussian kernel function  
*智能系统学报*. 2018, 13(3): 388-394 <https://dx.doi.org/10.11992/tis.201612040>

DOI: 10.11992/tis.202310025

网络出版地址: <https://link.cnki.net/urlid/23.1538.TP.20240909.1108.002>

# ObjectBoxG: 基于 GC3 模块的目标检测算法

张建宇, 谢娟英

(陕西师范大学 计算机科学学院, 陕西 西安, 710119)

**摘要:** 随着对目标检测任务研究的不断深入, 以 ObjectBox 检测器为代表的无锚框方法引起了研究者的关注。然而, ObjectBox 检测器不能充分利用多尺度特征, 也未充分考虑目标中心点与全局信息关联。为此, 借助图卷积神经网络的节点相互影响原理, 提出基于图谱方法的图卷积层模块 GConv (graph convolution layer), 学习图像全局特征; 融合模块 GConv 与 C3 (cross stage partial network with 3 convolutions) 得到 GC3 (graph C3 module) 模块, 进一步提取图像原始特征、细节特征以及全局特征; 将 GC3 结合广义特征金字塔网络 GFPN (generalized feature pyramid network), 提出图广义特征金字塔网络 GGFPN (graph generalized feature pyramid network), 并嵌入 ObjectBox 算法, 设计出 ObjectBoxG 算法。经典数据集的实验测试表明, 提出的 GC3 模块比原 C3 模块具有更强特征提取能力; 提出的 GGFPN 网络比 GC3 的特征学习能力更强; 提出的 ObjectBoxG 算法具有优良的目标检测性能。

**关键词:** 图卷积神经网络; 特征提取; 特征融合; 目标检测; 深度学习; 无锚框方法; 特征金字塔网络; Object-Box 检测器; 多尺度特征; 全局特征

中图分类号: TP181 文献标志码: A 文章编号: 1673-4785(2024)06-1385-10

中文引用格式: 张建宇, 谢娟英. ObjectBoxG: 基于 GC3 模块的目标检测算法 [J]. 智能系统学报, 2024, 19(6): 1385-1394.

英文引用格式: ZHANG Jianyu, XIE Juanying. ObjectBoxG: object detection algorithm based on GC3 module[J]. CAAI transactions on intelligent systems, 2024, 19(6): 1385-1394.

## ObjectBoxG: object detection algorithm based on GC3 module

ZHANG Jianyu, XIE Juanying

(School of Computer Science, Shaanxi Normal University, Xi'an 710119, China)

**Abstract:** With the deepening development of the study on object detection tasks, anchor-free methods such as the ObjectBox detector have attracted the attention of researchers. However, the ObjectBox detector has its limitations: it does not fully utilize multiscale features or adequately consider the correlation between target center points and global information. A graph convolution layer module (GConv), which is based on the graph spectrum method, is proposed to learn global image features and address the aforementioned limitations. Additionally, a new module named GC3 combines the proposed GConv module with C3 (cross-stage partial network with 3 conversions) to further extract the original, fine, and global image features. GC3 is combined with the generalized feature pyramid network (GGFPN) to form the GGFPN. The GGFPN is then embedded into the ObjectBox detector, resulting in the ObjectBoxG algorithm. Experiments on benchmark datasets demonstrate that the proposed GC3 module has stronger feature extraction capability than the original C3 module, and the proposed GGFPN network offers superior feature learning capability to GC3. The ObjectBoxG algorithm demonstrates excellent performance in object detection.

**Keywords:** graph convolutional neural network; feature extraction; feature fusion; object detection; deep learning; anchor-free methods; feature pyramid network; Object-Box detector; multi-scale features; global features

收稿日期: 2023-10-19. 网络出版日期: 2024-09-09.

基金项目: 国家自然科学基金项目 (62076159, 61673251, 12031010); 中央高校基本科研业务费项目 (GK202105003).

通信作者: 谢娟英. E-mail: [xiejuany@snnu.edu.cn](mailto:xiejuany@snnu.edu.cn).

目标检测是计算机视觉领域的底层任务之一, 包括定位图像中目标及对目标进行识别<sup>[1]</sup>, 深度学习使目标检测成为热门研究领域之一。基于深度学习的目标检测算法的基本框架<sup>[2]</sup>, 包括骨

干架构 (backbone)、检测头 (head), 以及连接 backbone 与 head 的瓶颈结构 (neck)。其中, backbone 提取图像中目标的特征; neck 对 backbone 提取的特征进行加工, 再分配到 head; head 实现目标位置检测与类别分类。

backbone 中的卷积神经网络 (convolutional neural networks, CNNs) 是标准网络架构<sup>[3]</sup>, Transformer 架构作为 backbone 之一, 具有很优异的性能<sup>[4]</sup>。CNNs 将图像视为欧氏空间里像素的规则网格, 应用具有平移不变性和局部性的卷积核进行特征提取<sup>[5]</sup>。Transformer 将图像切分为一系列互不重叠的补丁序列, 使用多头注意力机制进行特征学习<sup>[6]</sup>。Han 等<sup>[7]</sup>提出视觉图神经网络框架 Vision GNN (vision graph neural network), 启发性地设计图模块 (Grapher), 利用 K-近邻思想 (K), 将图像构建为图结构; 使用空间图卷积方法, 聚合并更新邻居节点之间的信息, 进行特征提取。然而, Grapher 设计中每个节点的邻居数固定为 K, 虽对邻居节点的信息有较好的聚合效果, 却忽视了单个节点与全局信息的关联。

neck 结构对 backbone 各层级的特征图进行融合加工, 增强网络表达能力, 并决定 head 的数量以及分配方式。neck 结构主要分为上下采样、路径聚合以及神经结构搜索 NAS (neural architecture search) 等<sup>[8]</sup>。上下采样<sup>[9]</sup>将 backbone 的特征图进行上下采样, 得到多级特征图后分配至 head, 不具有特征层聚合性操作。路径聚合<sup>[10]</sup>基于一个基本现象, 经过层层卷积下采样的深层特征图使得小物体的信息严重丢失。因此, 路径聚合通过聚合深层特征图与浅层特征图的信息, 将 backbone 的多层级特征图进一步融合后分配至 head, 其特点是反复利用各种上下采样、拼接等操作来设计聚合策略。特征金字塔网络 (feature pyramid network, FPN)<sup>[11]</sup>是路径聚合的经典方法, 旨在聚合从主干网中提取的不同分辨率特征。PAnet (path aggregation network)<sup>[12]</sup>增加了一个额外的自下而上的路径聚合网络, 融合多尺度特征, 但计算成本更高。GFPN (generalized feature pyramid network) 充分交换高级语义信息和低级空间信息<sup>[13]</sup>, 作为 neck 实现了 SOTA (state-of-the-art) 性能。神经结构搜索 NAS<sup>[14]</sup>利用神经网络搜索方法搜索合适的聚合路径, 但搜索时间成本极高, 且可解释性低。

根据 head 的不同, 将目标检测算法分为 2 大类, 两阶段检测算法和单阶段检测算法<sup>[15]</sup>。前者在第 1 阶段选择一系列候选区域, 在第 2 阶段提

取候选区域特征, 实现目标定位与分类, 例如, 区域卷积神经网络 R-CNN (region with convolutional neural network) 系列算法<sup>[16-18]</sup>; 后者直接提取图像特征, 生成目标框, 实现目标定位和分类, 例如, SSD (single shot multibox detector) 系列算法<sup>[19-20]</sup>。主流的目标检测算法基于锚框 (Anchor) 机制, 采用预定义边框进行定位和分类。Anchor 机制引入额外的超参数: 锚框大小和长宽比。这不仅导致提取的候选区域太多, 增加了计算开销, 还引起正负样本不均衡问题, 因此, 无锚框 (Anchor-free) 方法被提出<sup>[2]</sup>。Law 等<sup>[21]</sup>提出 CornerNet, 输出目标的左上角点和右下角点来进行检测。Tian 等<sup>[22]</sup>提出全卷积单阶段目标检测 FCOS (fully convolutional one-stage object detection), 以类似于语义分割的逐像素预测方式解决目标检测问题。Zand 等<sup>[23]</sup>提出了 ObjectBox 检测器, 将目标中心位置视为与形状和大小无关的 Anchor, 并在每个目标的所有尺度上进行学习, 使用一种定制交的并比 (intersection over union, IOU) 损失处理尺度变化的目标, 在无锚框 (Anchor-free) 方法的目标检测算法中检测效果显著。然而, ObjectBox 检测器以 C3 模块为主要的特征提取模块, 未充分考虑目标中心点与全局信息的关联; 同时, ObjectBox 检测器, 以 CSPDarknet 作为 backbone, 以改进的 PAnet 结构作为 neck, 难以充分利用多尺度特征。

图卷积神经网络中, 图中每个节点均受到邻居节点影响而改变自己的状态直到最终平衡, 关系越亲近的邻居影响越大。因此, 本研究提出图谱卷积层, 考虑单个节点与全局节点的关联, 使用图谱图卷积, 学习图像的全局特征; 在 C3 模块基础上, 提出 GC3 模块 (graph C3 module) 提取特征。提出图广义特征金字塔网络 GGFPN (graph generalized feature pyramid network) 结构, 得到 ObjectBoxG 目标检测算法。在目标检测经典数据集 PASCAL VOC<sup>[24]</sup> 和 MS COCO<sup>[25]</sup> 的实验表明, 提出的 GC3 模块具有更好的特征学习能力, 提出的 GGFPN 能对多个层级的特征图进行有效地融合加工, 提出的 ObjectBoxG 算法具有更优良的目标检测性能。

## 1 相关概念及本文方法

首先介绍图的基础概念、图卷积神经网络, 接着简要介绍视觉图神经网络架构 (vision GNN)<sup>[7]</sup>, 最后详细阐述本研究的创新点: 提出的图谱图卷积模块 GConv、GC3 模块、图广义特征

金字塔网络 GGFPN, 以及 ObjectBoxG 算法。

### 1.1 图的基础概念

图表示为  $G = (V, E)$ , 包括节点集合  $V$  和边集合  $E$ 。  $\|V\| = n$ ,  $v_i \in V$  表示节点  $i$ ,  $e_{ij}$  表示连接节点  $i$  和  $j$  的边 (这里仅考虑无向图)。节点  $v$  的邻居节点为  $N(v) = \{u \in V | e_{vu} \in E\}$ 。图  $G$  的邻接矩阵  $A \in \mathbf{R}^{n \times n}$  是一个  $n$  阶矩阵, 若  $e_{ij} \in E$ , 则  $A_{ij} = 1$ ; 否则,  $A_{ij} = 0$ ; 如图  $G$  是无向图, 则其邻接矩阵是对称阵, 即  $A_{ij} = A_{ji}$ 。

$L = D - A$  为图  $G$  的拉普拉斯矩阵, 其中,  $D$  是图  $G$  的度矩阵, 一个对角阵, 其对角线元素  $D_{ii} = \sum_{j=1}^n A_{ij}$ 。图  $G$  的归一化拉普拉斯矩阵为  $L = I_n - D^{-1/2} A D^{-1/2}$ , 其中,  $I_n \in \mathbf{R}^{n \times n}$  是单位矩阵。显然,  $L$  是一个实对称矩阵, 特征分解得到:  $L = U \Lambda U^T$ ,  $U = \{u_i\}_{i=1}^n$  是  $n$  个相互正交的特征向量,  $\Lambda = \text{diag}(\{\lambda_i\}_{i=1}^n)$  表示特征值矩阵 (diag 表示对角阵),  $\lambda_i$  是  $u_i$  对应的特征值。图节点属性矩阵  $X \in \mathbf{R}^{n \times d}$ ,  $d$  是节点属性的维度,  $x_i \in \mathbf{R}^d$  表示节点  $i$  的属性向量。

### 1.2 图卷积神经网络

图卷积神经网络 GCN (graph convolutional network) 分为两大类: 基于图谱的方法和基于空间的方法。Bruna 等<sup>[26]</sup> 提出第一个基于图谱的图卷积神经网络, 结合图理论从卷积定理出发, 在频域上定义了图卷积。基于空间的图卷积神经网络最早由 Micheli 等<sup>[27]</sup> 提出, 通过复合非递归层, 解决图的相互依赖性, 同时继承消息传递思想。

图谱方法的 GCN 利用卷积定理从频域定义图卷积, 即利用卷积定理, 先利用傅里叶变换将信号从原始空间转换到频域, 再在频域将信号做乘法, 最终利用傅里叶逆变换将信号转换到原始空间, 实现图卷积可以理解在频域中的节点级别信号  $x \in \mathbf{R}^n$  与卷积核  $g_\theta = \text{diag}(\theta)$  ( $\theta \in \mathbf{R}^n$  是卷积核的参数) 的乘积

$$g_\theta \times x = U g_\theta U^T x \quad (1)$$

式中:  $U$  是图  $G$  的拉普拉斯矩阵的特征向量矩阵;  $L = I_n - D^{-1/2} A D^{-1/2}$ ,  $L = U \Lambda U^T$ ,  $\Lambda$  是特征值构成的对角矩阵,  $U^T x$  是输入信号  $x$  的图傅里叶变换;  $g_\theta$  可以理解为图  $G$  的归一化拉普拉斯矩阵的特征值函数, 即  $g_\theta(\Lambda)$ 。Hammond 等<sup>[28]</sup> 使用切比雪夫多项式  $T_k(x)$  对近似函数  $g_\theta(\Lambda)$  展开到第  $k$  阶, 降低了计算复杂度, 其数学表达式为

$$g_\theta(\Lambda) = \sum_{k=0}^K \theta'_k T_k(\tilde{\Lambda}) \quad (2)$$

式中:  $\tilde{\Lambda} = (2\Lambda/\lambda_{\max}) - I$ ,  $\lambda_{\max}$  是  $L$  的最大特征值;  $\theta' \in \mathbf{R}^K$  是切比雪夫系数的向量;  $T_k(x)$  是递归的切

比雪夫多项式, 其定义为  $T_k(x) = 2xT_{k-1}(x) - T_{k-2}(x)$ , 并且  $T_0(x) = 1$ ,  $T_1(x) = x$ 。结合以上 2 个式子, 基于频域的图卷积操作可重新定义为

$$g_\theta' \times x = \theta'_k T_k(\tilde{L}) x \quad (3)$$

式中  $\tilde{L} = (2L/\lambda_{\max}) - I$ 。根据式 (3), 每个节点仅依赖于第  $k$  阶邻域的信息。

为了定义图卷积神经网络, 根据式 (3), 可以堆叠多个卷积层, 每层后接非线性变换。Kipf 等<sup>[29]</sup> 建议, 将式 (3) 中每层的卷积数  $K$  限制为  $K = 1$ , 则每层的图卷积操作可表示为图拉普拉斯矩阵  $L$  上的线性函数, 在每个线性函数上, 进一步近似  $\lambda_{\max} \approx 2$ , 最终得到

$$g_\theta' \times x = \theta'_0 x + \theta'_1 x (L - I_n) \quad (4)$$

式中  $\theta'_0$ 、 $\theta'_1$  是整个图中可以共享的参数。实践中, 可以进一步减少参数数量, 当  $\theta = \theta'_0 = -\theta'_1$  时, 式 (4) 可表达为

$$g_\theta' \times x = f_\theta(x) \approx \theta \left( I + D^{-\frac{1}{2}} A D^{-\frac{1}{2}} \right) x \quad (5)$$

根据式 (5), 为在图  $G$  上进行卷积, 引入一个简单而灵活的模型  $f_\theta(\cdot)$ , 其中, 模型的核心在于共享参数  $\theta$  以及底层图结构的邻接矩阵  $A$ 。

### 1.3 图卷积层

Vision GNN 的核心思想是将图像转化为图  $G$ , 构建图卷积层 Grapher, 使用基于空间的图卷积, 通过聚合相邻节点的特征, 实现部分相似节点间的信息交换。然而, Grapher 固定邻居数  $K$ , 聚合  $K$  个邻居信息, 忽视单个节点与全局信息的关联。本研究在 Vision GNN 基础上, 提出基于图谱方法的图卷积层 GConv。

首先, 将图像构建为图结构。将  $H \times W \times C$  尺寸的特征图转换为特征矩阵  $X_{N \times C}$ , 其中,  $N = H \times W$ 。将  $X_{N \times C}$  每一行作为一个节点的特征向量,  $X_{N \times C}$  可被视为节点集  $V = \{v_1, v_2, \dots, v_N\}$ 。进一步确定距离矩阵, 根据 Grapher 思想, 计算节点  $v_i$  与  $v_j$  对应的特征向量  $x_i$  与  $x_j$  的欧氏距离, 得到节点  $v_i$  与  $v_j$  之间的距离  $e_{ij}$ , 遍历所有节点, 得到距离矩阵  $D_{\text{ist}} \in \mathbf{R}^{N \times N}$ 。距离  $e_{ij}$  表示节点  $v_i$  与  $v_j$  之间的距离, 也反映节点  $v_i$  与  $v_j$  之间的相似性, 相似性与距离呈负相关。

原始的 Grapher 在  $D_{\text{ist}}$  中, 对于任意一行  $i$  (对应于节点  $v_i$ ), 通过索引  $K$  个最小值, 固定节点  $v_i$  的  $K$  个邻居, 得到节点  $v_i$  的邻居点集  $N(v_i)$ ; 再使用基于空间的图卷积进行特征提取。

本研究提出的图卷积层 GConv, 在  $D_{\text{ist}}$  的基础上, 引入转化函数  $F_t(\cdot)$ , 构建邻接矩阵  $A = F_t(D_{\text{ist}}) \in \mathbf{R}^{N \times N}$ ; 边的强度考虑节点之间的相似性,

与距离呈负相关,因此,构造邻接矩阵  $A \in \mathbf{R}^{N \times N}$ :

$$A_{ij} = F_t(D_{ist_{ij}}) = \begin{cases} \lambda \bar{D}_{ist_i} - D_{ist_{ij}}, & D_{ist_{ij}} \leq \lambda \bar{D}_{ist_i} \\ 0, & \text{其他} \end{cases} \quad (6)$$

式中:  $\bar{D}_{ist_i} = \sum_{j=1}^N D_{ist_{ij}}/N$  表示节点  $v_i$  到其他节点的平均距离;  $\lambda$  是转化函数  $F_t(\cdot)$  的超参数,用以确定目标节点的邻居范围,舍弃距离过大的节点。

构建邻接矩阵  $A$  后,使用基于图谱方法的图卷积进行特征提取,数学表达式为

$$X' = \text{GraphConv}(X) = D^{-\frac{1}{2}} A D^{-\frac{1}{2}} X W \quad (7)$$

式中:  $\text{GraphConv}(\cdot)$  表示图谱方法的图卷积,  $W$  是可学习的参数矩阵,  $X'$  是图卷积后的特征矩阵。

深度 GCNs<sup>[30-31]</sup> 中的过度平滑现象将降低节点特征的显著性,并导致视觉识别性能的下降。因此,在图卷积前后分别加入一个线性层,增加特征多样性的同时,保证输入与输出的特征图尺寸和通道数一致;在图卷积后插入一个非线性激活函数,以避免梯度消失的问题;结合残差模块的思想,防止信息传递过程中造成丢失,得到图卷积层  $G\text{Conv}$ 。给定输入特征  $X \in \mathbf{R}^{N \times d}$ ,  $G\text{Conv}$  可表示为

$$Y = G\text{Conv}(X) = \sigma(\text{GraphConv}(XW_{in}))W_{out} + X \quad (8)$$

式中:  $Y \in \mathbf{R}^{N \times c}$  是输出特征,  $W_{in}$  和  $W_{out}$  分别为全连接层的输入层和输出层权重,  $\sigma$  是激活函数。基于图谱方法的图卷积,更注重每个节点与全局信息的关联,进一步提取特征图的全局信息和相似节点间的特征信息,但因其考虑全部节点并注重全局信息,导致计算量大,可能造成小目标特征缺失。

#### 1.4 GC3 模块

C3(cross stage partial network with 3 convolutions) 模块是 CNN 中的经典模块<sup>[32]</sup>。在 Object-Box<sup>[23]</sup> 算法中, C3 模块是网络的重要组成部分,其主要作用是增加网络的深度和感受野,提高网络的特征提取能力。将上述提出的图卷积层与 C3 模块结合,设计出图 1 的 GC3 模块,其中,黄色框部分是原始 C3 模块的结构,蓝色框是新增的图卷积层。

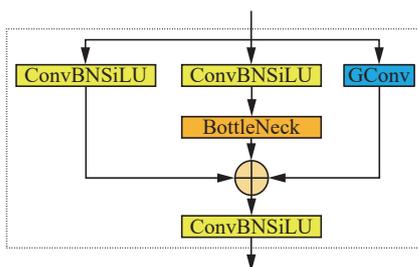


图 1 GC3 模块  
Fig. 1 GC3 module

由图 1 可知,原始的 C3 模块结构为 2 条分支,一条分支使用标准卷积层和 Bottleneck 模块,学习特征图的细节特征;另一条分支使用标准卷积层保留特征图的原始输入特征;最终将 2 分支进行拼接后,再使用标准卷积层得到输出特征图。本研究设计的 GC3 模块,在原有模块的基础上增加一条支路,通过图卷积层,聚合输入特征图的全局特征,最终将细节特征、原始特征以及全局特征进行拼接,再使用标准卷积层,得到输出特征图。GC3 模块在原有模块的基础上,借助图卷积层的特性,进一步提取特征图的全局特征。

#### 1.5 ObjectBoxG 目标检测算法

原始 ObjectBox 算法的结构如图 2 所示,neck 结构采用的是改进的 PAnet 结构。图中的黑色箭头表示 PAnet 部分,蓝色箭头表示改进的 PAnet 部分,即将深层 (Deep layer) 的特征图进行上采样和中间层的特征图拼接,提取特征后得到的新特征图信息作为一个分支,分别与 PAnet 的浅层 (Shallow layer) 特征图信息和中间层 (Medium layer) 特征图信息进行拼接和特征提取,最终通过 PAnet 输入检测头 (detector)。原始 Object-Box 算法中的 neck 结构,未充分融合深层特征图与浅层特征图的语义信息,同时,也未对特征图进行充分的特征提取。

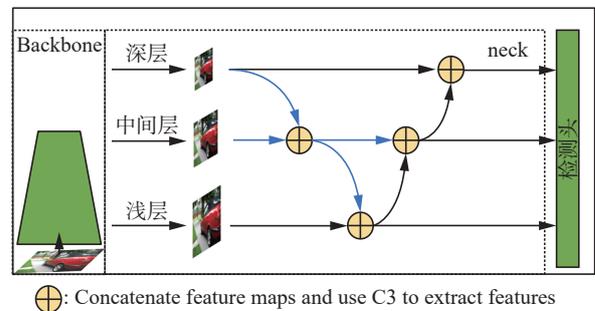


图 2 原始 ObjectBox 结构

Fig. 2 Structure diagram of original ObjectBox algorithm

鉴于 GC3 模块是即插即用的特征提取模块,可以很容易地替换卷积神经网络的任何一个卷积层。本研究结合 GFPN 结构<sup>[13]</sup> 的思想,采用 GC3 模块作为 GFPN 结构中特征提取主要模块,设计出图 3 的 GGFPN(graph generalized feature pyramid network) 结构,进一步提升目标检测算法的性能。图 3 为 GGFPN 结构,在原始 neck 结构基础上,将中间层特征图经过下采样与深层特征图进行拼接,特征提取后得到新的深层特征图信息,并且经过上下采样充分融合各层级特征图,得到新的中间层特征图信息。将各层级特征图信息通过 GC3 模块进一步提取特征,最终输入检测头,进而输出多尺度预测结果。

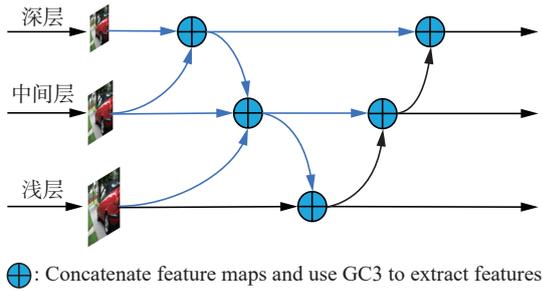


图 3 提出的 GGFPN 结构

Fig. 3 Structure diagram of GGFPN proposed in this paper

在 ObjectBox 算法基础上, 保留原有的 backbone 和 detector 结构, 将原有的 neck 部分替换为 GGFPN 结构, 并将新算法命名为 ObjectBoxG 算法, 如图 4 所示。ObjectBoxG 属于单阶段目标检测算法, 通过多尺度特征图预测, 以期对不同尺寸的目标有不错的检测效果。

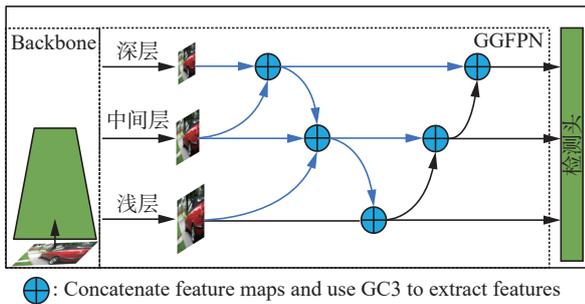


图 4 提出的 ObjectBoxG 算法的结构

Fig. 4 Structure diagram of the proposed ObjectBoxG algorithm

## 2 实验结果

实验分为 3 部分进行, 首先, 在 PASCAL VOC 数据集上, 通过设定本研究提出的图卷积层的不同  $\lambda$  值, 测试将 C3 模块替换成 GC3 模块的 ObjectBox 算法 (表示为 ObjectBox(GC3)) 以及原有 ObjectBox 算法的检测性能, 验证提出的即插即用模块 GC3 的特征提取能力和灵活性。其次, 以 yolov5s 算法<sup>[33]</sup> 为基准, 将其 C3 模块替换成 GC3 模块, 得到 yolov5s(GC3); 在 yolov5s 算法嵌入 GGFPN 结构得到 yolov5s+GGFPN, 在公开数据集 MS COCO 上进行测试, 进一步验证提出的

GC3 模块的特征提取能力, 并验证提出的 GGFPN 结构融合加工多层次特征图的能力。最后, 在 MS COCO 公开数据集, 将 ObjectBoxG 算法与原有的 ObjectBox 算法以及 ObjectBox(GC3) 算法进行比对, 验证提出的 ObjectBoxG 算法的检测性能。鉴于设备以及部分参数的设置问题, 实验无法复现出 ObjectBox 算法以及 yolov5s 算法在原文中的结果, 因此, 仅比较本研究在相同实验环境下的复现结果。

### 2.1 GC3 模块的性能测试

在 PASCAL VOC 数据集测试 ObjectBox(GC3) 算法的性能。PASCAL VOC 数据集设置 20 个类别, 有 2007 年版本和 2012 年版本 (两版本数据互不包含)。实验将 2007 年的训练集与验证集和 2012 年的训练集和验证集共同作为训练集, 共有 16 551 张图像数据, 包含 40 058 个实例目标, 由于 PASCAL VOC 2012 的测试集没有公开, 因此, 测试集仅使用 PASCAL VOC 2007 的测试集, 共有 4 952 张图像数据, 包含 12 032 个实例目标。目标检测有多项评价指标, 包括平均精度 (average precision, AP)[0.5:0.95], 代表 IoU 从 0.5 以步长 0.05 变化到 0.95 得到的平均精度均值 (mean average precision, mAP); AP[0.5], 代表 IoU 取 0.5 时的 mAP 值。这里使用的指标是 AP[0.5]。

实验代码基于 ObjectBox 的开源代码, 使用 Pytorch 框架实现 GC3 模块及整个训练测试流程。实验环境为 CentOS Linux 7.5 操作系统, GPU 为单卡 NVIDIA GeForce RTX3080ti, Pytorch 版本为 1.10.0。根据 GPU 显存大小, 训练时将 ObjectBox、ObjectBox(GC3) 以及 ObjectBoxG 的 batch size 均设置为 4。除了 epoch 设定为 300 外, 其余参数均按照 ObjectBox 设置, 训练策略参考 ObjectBox 文献 [23]。

实验首先测试本研究提出的图卷积层邻接矩阵的参数  $\lambda$  对提出的 ObjectBox(GC3) 网络的影响。表 1 为设定不同  $\lambda$  时, ObjectBox(GC3) 在 PASCAL VOC 数据集的 AP[0.5] 指标值, 加粗和下划线表示最优值。图 5 为  $\lambda = 2$  时, ObjectBox(GC3) 和 ObjectBox 在 PASCAL VOC 数据集的 AP[0.5] 指标。

表 1 不同  $\lambda$  值下 ObjectBox(GC3) 在 PASCAL VOC 数据集的 AP[0.5] 值比较

Table 1 Performance of ObjectBox (GC3) on PASCAL VOC dataset in terms of AP [0.5] under different  $\lambda$

$\lambda$	0.5	1	1.5	2	4	8
AP[0.5]	0.798	0.802	0.804	<b><u>0.811</u></b>	0.801	0.800

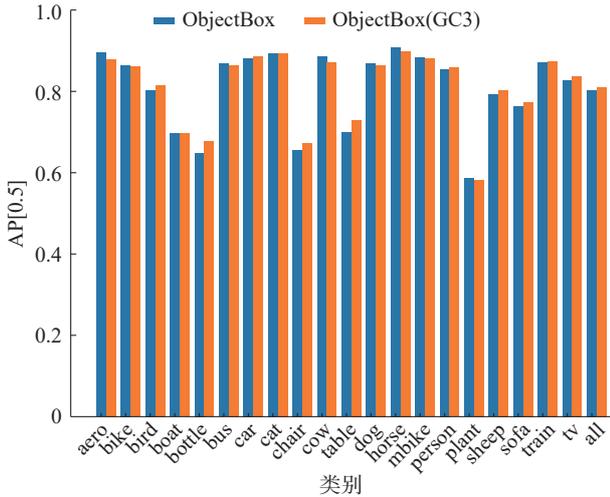


图 5 ObjectBox(GC3) 和 ObjectBox 在 PASCAL VOC 数据集的 AP[0.5] 指标比较

Fig. 5 Performance comparison of ObjectBox(GC3) and ObjectBox on PASCAL VOC dataset in terms of AP[0.5]

表 1 实验结果显示, 在 PASCAL VOC 数据集,  $\lambda = 2$  时, 本研究 ObjectBox(GC3) 模型的 AP[0.5] 指标值取得最优结果。因此, 后续实验中的  $\lambda$  均设置为 2。表 1 实验还说明, 不同的  $\lambda$  决定目标节点的邻居范围;  $\lambda$  取值越小, 确定目标节点的邻居节点越少, 这导致在图卷积过程中只聚合少量极相似节点的信息, 导致 GC3 模块提取全局特征的能力减弱。在  $\lambda$  取值大于某一临界值时, 将所有节点都归纳为目标节点的邻居, 导致图卷积过程中聚合大量非相似节点的信息, 导致 GC3 模块特征提取能力减弱。

图 5 结果显示, ObjectBox(GC3) 在 PASCAL VOC 数据集的 20 个类别中, 在 11 个类别的 AP 指标优于 ObjectBox, 且整体的 mAP 为 0.811, 高于 ObjectBox 算法在该数据集的整体 mAP 值 (0.803), 说明相比于 C3 模块, 提出的 GC3 模块具有更优的特征学习能力。本研究的 ObjectBox (GC3) 模型对 bird、bottle、chair、sheep、sofa 以及 tv 等类别的检测结果显著提升, 对 aero、bike、bus、cow、dog、horse、mbike、plant 类别的检测结

果低于 ObjectBox 算法的检测结果, 说明 GC3 模块可能导致 ObjectBox(GC3) 算法对相似目标的错误分类。

### 2.2 GGFPN 结构的性能测试

在 MS COCO 数据集测试将 C3 模块替换成 GC3 模块的 yolov5s 算法 (表示为 yolov5s(GC3))、嵌入 GGFPN 结构的 yolov5s 算法 (表示为 yolov5s+GGFPN) 的性能, 以验证本研究提出的 GGFPN 结构的性能。MS COCO 数据集包含 80 个类别, 背景复杂, 目标数量多、存在大量的小目标。实验选用 MS COCO2017 数据集, 其中训练集有 118 287 张图像, 验证集有 5 000 张图像, 测试集有 40 670 张图像。由于测试集未公开标注文件, 因此, 使用验证集来衡量模型性能。

MS COCO 数据集采用更严格的 6 项评价指标, 包括评价指标 AP[0.5:0.95] ( $S$ ), 代表 IoU 从 0.5 以步长 0.05 变化到 0.95 得到的 mAP; AP[0.5], 代表 IoU 取 0.5 时的 mAP 值, 与 PASCAL VOC 的评价指标相同; AP[0.75], 代表 IoU 取 0.75 时的 mAP 值, 是比 IoU 取 0.5 更为严格的指标; AP[ $S$ ] 代表小目标 (面积小于  $32 \times 32$ ) 的 AP[0.5:0.95] 值; AP[ $M$ ] 代表中等大小目标 (面积在  $32 \times 32$  与  $96 \times 96$  之间) 的 AP[0.5:0.95] 值; AP[ $L$ ] 代表大目标 (面积大于  $96 \times 96$ ) 的 AP[0.5:0.95] 值。

实验代码基于 yolov5s 的 5.0 版本的开源代码, 使用 Pytorch 框架实现 GGFPN 结构及整个训练测试流程。实验环境为 CentOS Linux 7.5 操作系统, GPU 为单卡 NVIDIA GeForce RTX3080ti, Pytorch 版本为 1.10.0。根据 GPU 显存大小, 训练时将 yolov5s、yolov5s(GC3) 以及 yolov5s+GGFPN 的 batch size 均设置为 8。除了 epoch 设定为 300 外, 其余参数均按照 yolov5s 设置<sup>[33]</sup>, 训练策略参考 yolov5s 模型。表 2 为 yolov5s、yolov5s (GC3) 以及 yolov5s+GGFPN 算法在 MS COCO 数据集的性能对比, 为了公平比较, 重新复现了基线算法 yolov5s, 使其环境与 yolov5s(GC3) 以及 yolov5s+GGFPN 算法完全一致。

表 2 yolov5s、yolov5s(GC3) 与 yolov5s+GGFPN 在 MS COCO 数据集的 mAP 值比较

Table 2 Performance of yolov5s, yolov5s(GC3) and yolov5s+GGFPN on MS COCO dataset in terms of mAP

%

方法	Backbone	AP[0.5:0.95]	AP[0.5]	AP[0.75]	AP[ $S$ ]	AP[ $M$ ]	AP[ $L$ ]
yolov5s	CSPdarknet	34.2	53.3	36.5	14.0	33.6	46.8
yolov5s(GC3)	CSPdarknet	34.2	53.3	36.8	13.1	33.9	47.9
yolov5s + GGFPN	CSPdarknet	<b>36.7</b>	<b>55.7</b>	<b>39.6</b>	<b>14.6</b>	<b>35.3</b>	<b>50.9</b>

注: 加粗表示最优值。

表 2 实验结果显示, 在更复杂的 MS COCO 数据集, 在集成性较高的 yolov5 算法框架下, yolov5s(GC3) 模型的 mAP 总体略高于 yolov5s 算法的 mAP, 特别是在更严格的 AP[0.75] 指标, 说明提出的 GC3 模块具有较强的特征提取能力, 但 yolov5s(GC3) 模型对小目标的检测性能不如原始 yolov5s 算法的 AP[S] 值。yolov5s+GGFPN 的各项指标分别优于 yolov5s 和 yolov5s(GC3) 的各项指标值, 说明在 GC3 模块基础上设计的 GGFPN 结构, 能有效地融合多层次特征图, 具有更强的特征学习能力, 从而可以进一步提升 yolov5s 算法的性能。

### 2.3 ObjectBoxG 算法性能测试

在 MS COCO 数据集测试提出的 ObjectBoxG 算法的性能, 并与原始 ObjectBox 算法、嵌入 GC3 的 ObjectBox 算法, 即 ObjectBox(GC3) 算法, 进行实验比较。实验代码基于 ObjectBox 开

源代码, 使用 Pytorch 框架实现 GC3 模块、GGFPN 结构及整个训练测试流程。实验环境为 CentOS Linux 7.5 操作系统, GPU 为单卡 NVIDIA GeForce RTX3080ti, Pytorch 版本为 1.10.0。根据 GPU 显存大小, 训练时将 ObjectBox、ObjectBox(GC3) 以及 ObjectBoxG 的 batch size 均设置为 4。除了 epoch 设定为 300 外, 其余参数均按照 ObjectBox 设置, 训练策略参考 ObjectBox 文献 [23]。表 3 为 ObjectBox、ObjectBox(GC3) 以及 ObjectBoxG 算法在 MS COCO 数据集的 mAP 指标。图 6 为 ObjectBox、ObjectBox(GC3) 以及 ObjectBoxG 算法在 MS COCO 数据集的部分结果。表 4 为 ObjectBox 算法、ObjectBoxG 算法与其他同系列算法在 MS COCO 数据集的性能对比。为了公平比较, 重新复现了基线算法 ObjectBox, 使其环境与 ObjectBox(GC3) 和 ObjectBoxG 算法完全一致, 其他算法的结果来自其原文。

表 3 ObjectBox、ObjectBox(GC3) 与 ObjectBoxG 在 MS COCO 数据集的 mAP 值比较  
Table 3 Performance of ObjectBox, ObjectBox(GC3) and ObjectBoxG on MS COCO dataset in terms of mAP %

方法	Backbone	AP[0.5:0.95]	AP[0.5]	AP[0.75]	AP[S]	AP[M]	AP[L]
ObjectBox	CSPdarknet	44.2	63.9	47.6	21.1	42.8	58.8
ObjectBox(GC3)	CSPdarknet	44.8	64.6	48.2	20.9	43.5	58.8
ObjectBoxG	CSPdarknet	<b>45.0</b>	<b>64.8</b>	<b>48.5</b>	<b>21.8</b>	<b>43.8</b>	<b>58.8</b>

注: 加粗表示最优值。

表 4 ObjectBoxG 与同系列算法在 MS COCO 数据集的性能比较

Table 4 Performance comparison of the ObjectBoxG and counterpart algorithms on MS COCO dataset %

方法	Backbone	AP[0.5:0.95]	AP[0.5]
SSD513 <sup>[19]</sup>	ResNet101	31.2	50.4
ASSD513 <sup>[34]</sup>	ResNet101	34.5	55.5
RFBnet512 <sup>[35]</sup>	VGG	32.3	51.8
Retinanet <sup>[36]</sup>	ResNet101	39.1	59.1
F-RCNN w/FPN <sup>[11]</sup>	ResNet101	36.2	59.1
yolov3 <sup>[37]</sup>	DarkNet53	33.0	57.9
CornerNet <sup>[21]</sup>	Hourglass104	40.6	56.4
CenterNet <sup>[38]</sup>	Hourglass52	41.6	59.4
FCOS <sup>[22]</sup>	ResNeXt101	42.1	62.1
PAA <sup>[39]</sup>	ResNet101	44.8	63.3
yolov5s	CSPDarknet	34.6	53.9
ObjectBox	CSPDarknet	44.2	63.9
ObjectBoxG	CSPDarknet	<b>45.0</b>	<b>64.8</b>

注: 加粗表示最优值。



图 6 在 MS COCO 数据集上部分实验结果

Fig. 6 Partial experimental results on the MS COCO dataset

表 3 实验结果显示, 在 MS COCO 数据集, ObjectBox(GC3) 模型的 mAP 总体高于 ObjectBox 算法的 mAP, 说明提出的 GC3 模块具有更强的特征提取能力, 但 AP[S] 值不如原始 ObjectBox 算法。提出的 ObjectBoxG 算法的各项指标分别优于 ObjectBox 和 ObjectBox (GC3) 的各项指标值, 说明 GGFPN 结构, 能有效地融合多层次特征图, 可以进一步提升算法的性能。

图 6 中, 第 1 列图片表示 MSCOCO 数据集真实目标, 第 2 列图片表示原始 ObjectBox 算法的检测结果, 第 3 列图片表示 ObjectBox(GC3) 算法的检测结果, 第 4 列图片表示 ObjectBoxG 算法的检测结果。由图 6 结果显示, ObjectBoxG 算法在 MSCOCO 数据集上整体表现较好, 但针对小目标以及被遮挡的目标存在漏检和误检的情况, 说明融合多层次特征图使模型学习背景特征, 造成小目标特征的缺失。

表 4 的实验结果显示, 本研究提出的 ObjectBoxG 模型无论在 AP[0.5:0.95] 还是 AP[0.5] 指标的值均优于其他对比模型的相应指标值, 说明本研究的 ObjectBoxG 模型具有优良的目标检测性能, 也说明提出的 GGFPN 和 GC3 的强大特征学习能力。

综合表 3 和表 4 的实验结果分析可见, 在相同的实验条件下, ObjectBox(GC3) 性能优于 ObjectBox, 说明相比于 C3 模块本研究提出的 GC3 模块具有更强的特征学习能力。另外, 在相同实验条件下, ObjectBoxG 模型的性能优于 ObjectBox(GC3) 和 ObjectBox, 并且优于其他目标检测算法, 说明 ObjectBoxG 算法具有更优良的目标检测性能。

### 3 结束语

基于图卷积层 (GConv), 并将其与 C3 模块结合, 提出了 GC3 模块, 在 GC3 模块基础上, 提出了 GGFPN 网络, 与 ObjectBox 算法结合, 提出新的基于图卷积的目标检测模型 ObjectBoxG。

GC3 模块提升了 C3 模块的特征提取能力, GGFPN 结构, 能有效地融合多层次特征图, 使 ObjectBoxG 算法取得了最优的目标检测性能。

然而, GC3 模块依旧存在对相似目标以及小目标特征学习能力低的问题。同时, 提出的模型基于无锚框 (Anchor-free) 方法, 无论是 ObjectBox(GC3) 还是本研究的 ObjectBoxG 模型在 MS

COCO 数据集的检测结果不及基于锚框 (Anchor) 机制的 SOTA 算法 yolov7。yolov7 对 MS COCO 数据集的 AP[0.5] 达到 69.7%<sup>[40]</sup>, 而本研究 ObjectBoxG 算法对 MS COCO 数据集的 AP[0.5] 只有 64.8%。这主要是由于 yolov7 是基于 Anchor 机制的新目标检测算法, 其网络的 backbone、neck、head 部分都增加了新模块。Anchor-based 网络参数量更大, 对锚框的选择更精细, 因此性能更好。另外, 除了这些与模型有关的因素外, 还与实验硬件环境有关。尽管如此, GC3 模块、GGFPN 网络的特征学习能力及其灵活性还是不可否认的。

另外, 需要说明的是 ObjectBoxG 算法的性能还依赖于主干网络。因此, 后续将进一步探索图卷积层 (GConv)、GC3 模块、GGFPN 网络结构与其他新模型的结合; 并探索引入注意力机制、动态卷积等改进 GC3 模块, 提升其特征学习能力。最后, 提出的图卷积层 GConv 也存在很大改进空间, 如何设计转化函数挖掘原始特征图隐含的拓扑信息, 如何引入新的图卷积操作来提取更准确的特征都需要进一步研究。

### 参考文献:

- [1] 张婷婷, 章坚武, 郭春生, 等. 基于深度学习的图像目标检测算法综述[J]. 电子学报, 2020, 36(7): 15.  
ZHANG Tingting, ZHANG Jianwu, GUO Chunsheng, et al. Survey of object detection based on deep learning[J]. Acta electronica sinica, 2020, 36(7): 15.
- [2] ZAIDI S S A, ANSARI M S, ASLAM A, et al. A survey of modern deep learning based object detection models[J]. *Digital signal processing*, 2022, 126: 103514.
- [3] LIU Zhuang, MAO Hanzhi, WU Chaoyuan, et al. A ConvNet for the 2020s[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans: IEEE, 2022: 11966–11976.
- [4] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers[M]//Lecture Notes in Computer Science. Cham: Springer International Publishing, 2020: 213–229.
- [5] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 770–778.
- [6] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: transformers for image

- recognition at scale[EB/OL]. (2020-10-22) [2020-10-22]. <http://arxiv.org/abs/2010.11929>.
- [7] HAN K, WANG Y, GUO J, et al. Vision Gnn: an image is worth graph of nodes[J]. *Advances in neural information processing systems*, 2022, 35: 8291-8303.
- [8] TAN Mingxing, PANG Ruoming, LE Q V. EfficientDet: scalable and efficient object detection[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 10778-10787.
- [9] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2015, 37(9): 1904-1916.
- [10] FU Chengyang, LIU Wei, RANGA A, et al. DSSD: deconvolutional single shot detector[EB/OL]. (2017-01-23) [2020-10-22]. <http://arxiv.org/abs/1701.06659>.
- [11] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 936-944.
- [12] LIU Shu, QI Lu, QIN Haifang, et al. Path aggregation network for instance segmentation[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 8759-8768.
- [13] JIANG Yiqi, TAN Zhiyu, WANG Junyan, et al. Giraffed-et: A heavy-neck paradigm for object detection[EB/OL]// (2022-02-09)[2022-12-12]. <https://arxiv.org/abs/2202.04256>.
- [14] GHIASI G, LIN T Y, LE Q V. NAS-FPN: learning scalable feature pyramid architecture for object detection[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019: 7029-7038.
- [15] XIE Juanying, LIU Ran. The study progress of object detection algorithms based on deep learning[J]. *Journal of Shaanxi Normal University (natural science edition)*, 2019, 47(5): 1-9.
- [16] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus: IEEE, 2014: 580-587.
- [17] GIRSHICK R. Fast R-CNN[C]//2015 IEEE International Conference on Computer Vision. Santiago: IEEE, 2015: 1440-1448.
- [18] REN Shaoqing, HE Kaiming, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2017, 39(6): 1137-1149.
- [19] LIU Wei, ANGUELOV D, ERHAN D, et al. SSD: single shot MultiBox detector[M]//Lecture Notes in Computer Science. Cham: Springer International Publishing, 2016: 21-37.
- [20] LI Zuoxin, YANG Lu, ZHOU Fuqiang. FSSD: feature fusion single shot multibox detector[EB/OL]. (2017-12-04) [2020-10-22]. <http://arxiv.org/abs/1712.00960>.
- [21] LAW H, DENG Jia. CornerNet: detecting objects as paired keypoints[C]//European Conference on Computer Vision. Cham: Springer, 2018: 765-781.
- [22] TIAN Zhi, SHEN Chunhua, CHEN Hao, et al. FCOS: fully convolutional one-stage object detection[C]//2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019: 9626-9635.
- [23] ZAND M, ETEMAD A, GREENSPAN M. ObjectBox: from centers to boxes for anchor-free object detection[C]//AVIDAN S, BROSTOW G, CISSÉ M, et al. European Conference on Computer Vision. Cham: Springer, 2022: 390-406.
- [24] EVERINGHAM M, VAN GOOL L, WILLIAMS C K I, et al. The pascal visual object classes (VOC) challenge[J]. *International journal of computer vision*, 2010, 88(2): 303-338.
- [25] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: common objects in context[M]//Lecture Notes in Computer Science. Cham: Springer International Publishing, 2014: 740-755.
- [26] BRUNA J, ZAREMBA W, SZLAM A, et al. Spectral networks and locally connected networks on graphs[EB/OL]. (2013-12-21) [2021-01-01]. <http://arxiv.org/abs/1312.6203>.
- [27] MICHELI A. Neural network for graphs: a contextual constructive approach[J]. *IEEE transactions on neural networks*, 2009, 20(3): 498-511.
- [28] HAMMOND D K, VANDERGHEYNST P, GRIBONVAL R. Wavelets on graphs via spectral graph theory[J]. *Applied and computational harmonic analysis*, 2011, 30(2): 129-150.
- [29] KIPF T N, WELING M. Semi-supervised classification with graph convolutional networks[EB/OL]. (2016-09-09) [2021-01-01]. <http://arxiv.org/abs/1609.02907>.
- [30] LI Qimai, HAN Zhichao, WU Xiaoming. Deeper insights into graph convolutional networks for semi-supervised

- learning[C]//Proceedings of the AAAI Conference on Artificial Intelligence. New Orleans: AAAI, 2018.
- [31] OONO K, SUZUKI T. Graph neural networks exponentially lose expressive power for node classification[EB/OL]. (2019-05-27) [2021-01-01]. <http://arxiv.org/abs/1905.10947>.
- [32] WANG C Y, MARK LIAO H Y, WU Y H, et al. CSPNet: a new backbone that can enhance learning capability of CNN[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Seattle: IEEE, 2020: 1571-1580.
- [33] ULTRALYTICS COMPANY. Yolov5. [EB/OL]. [2021-01-01]. <https://github.com/ultralytics/yolov5/2021>.
- [34] YI Jingru, WU Pengxiang, METAXAS D N. ASSD: attentive single shot multibox detector[J]. *Computer vision and image understanding*, 2019, 189: 102827.
- [35] LIU Songtao, HUANG Di, WANG Yunhong. Receptive field block net for accurate and fast object detection[M]//Lecture Notes in Computer Science. Cham: Springer International Publishing, 2018: 404-419.
- [36] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]//2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017: 2999-3007.
- [37] REDMON J, FARHADI A. yolov3: an incremental improvement[EB/OL]. (2018-04-08) [2021-01-01]. <http://arxiv.org/abs/1804.02767>.
- [38] DUAN Kaiwen, BAI Song, XIE Lingxi, et al. CenterNet: keypoint triplets for object detection[C]//2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019: 6568-6577.
- [39] KIM K, LEE H S. Probabilistic anchor assignment with IoU prediction for object detection[M]//Lecture Notes in Computer Science. Cham: Springer International Publishing, 2020: 355-371.
- [40] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. Yolov7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023: 7464-7475.

#### 作者简介:



张建宇, 硕士研究生, 主要研究方向为深度学习、计算机视觉。E-mail: [zhangjianyu06@qq.com](mailto:zhangjianyu06@qq.com)。



谢娟英, 教授, 博士生导师, 博士, 主要研究方向为机器学习、数据挖掘、生物医学大数据分析。发表学术论文百余篇。E-mail: [xiejuany@snnu.edu.cn](mailto:xiejuany@snnu.edu.cn)。