基于级联式逆残差网络的遥感图像轻量目标检测算法

陈 立[1],张 帆[2],郭 威[2],黄 赟[1],李继中[3]
(1. 信息工程大学, 河南郑州 450001;2. 国家数字交换系统工程技术研究中心,河南郑州 450002;
3. 郑州战略投送基地,河南郑州 450002)

摘 要: 遥感场景下的高实时目标检测任务具有重要的研究价值与应用意义. 针对当前遥感图像目标检测模型由于目标多角度、排列密集以及背景复杂从而导致检测速度慢的问题, 提出一种级联式逆残差卷积结构(Cascaded Inverted Residual Convolution, CIRC). 该结构采用深度可分离卷积作为基本卷积单元, 快速提升模型计算能力; 在此基础上, 通过转置通道矩阵与级联深度卷积, 并增加残差连接层数, 达到强化目标多维特征的目的;进一步,进行多级模块堆叠, 提高模型对目标的检测效果. 本文在RetinaNet基础上, 利用CIRC设计了一个快速的轻量化目标检测网络—CIRCN(Cascaded Inverted Residual Convolution Net). 同时, 在训练阶段引入角度变量并参与反向传播, 在推理阶段对水平框加入角度偏置, 有效提高定向目标与检测框匹配度. 在DOTA数据集上的实验结果表明, CIRCN在精度略受损失的情况下, 检测速度达到42 fps, 比基准算法提高了3.5倍. 结果验证了所提算法的有效性与可靠性.

关键词: 遥感图像;目标检测;模型轻量化;深度可分离卷积;级联式逆残差卷积;通道混排
基金项目: 国家自然科学基金创新研究群体项目(No.61521003)
中图分类号: TP391 **文献标识码:** A **文章编号:** 0372-2112(2023)09-2588-10
电子学报 URL:http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20210831

Cascaded Inverse Residual Network for Lightweight Object Detection Model in Remote Sensing Image

CHEN Li[1], ZHANG Fan[2], GUO Wei[2], HUANG Yun[1], LI Ji-zhong[3]
(1. *Information Engineering University*, *Zhengzhou*, *Hennan* 450001, *China*;
2. *National Digital Switching System Engineering Technology Research Center*, *Zhengzhou*, *Hennan* 450002, *China*;
3. *Zhengzhou Strategic Delivery Base*, *Zhengzhou*, *Hennan* 450002, *China*)

Abstract: The task of high real-time object detection in remote sensing scenes has important research value and application significance. Aiming at the slow detection speed of the current remote sensing image target detection model due to multiple angles, dense arrangement and complex background, a cascaded inverted residual convolution (CIRC) is proposed. This structure uses depthwise separable convolution as the basic convolution unit to quickly improve the model´s computing power. On this basis, the multi-dimensional features of the object are enhanced by transposing the channel matrix with cascaded depth convolution and increasing the number of residual connection layers. Further, multi-level module stacking is carried out to improve the detection effect of the model on the object. Based on RetinaNet, this paper uses CIRC to design a fast lightweight object detection network—CIRCN (Cascaded Inverted Residual Convolution Net). At the same time, the angle variable is introduced in the training phase and participates in back propagation, and the angle offset is added to the horizontal frame in the inference phase, which effectively improves the matching degree of the directional target and the detection frame. The experimental results on the DOTA dataset show that the detection speed of CIRCN reaches 42 fps with a slight loss of accuracy, which is 3.5 times higher than the benchmark algorithm. The results verify the effectiveness and reliability of the proposed algorithm.

Key words: remote sensing image; object detection; model lightweight; depthwise separable convolution; cascaded inverted residual convolution; channel mixing

收稿日期:2021-07-02;修回日期:2021-10-15;责任编辑:梅志强

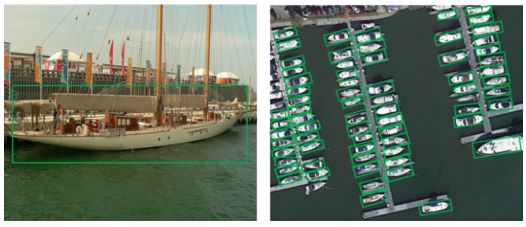
Foundation Item(s): National Natural Science Foundation of China (No.61521003)

1 引言

目标检测(object detection)是计算机视觉和图像处理领域最具挑战性的分支之一,旨在提取目标特征并得到感兴趣区域,完成对象的检测与识别. 近年来,随着遥感技术的快速发展,针对遥感图像的目标检测方法由于能够按照特定需求对海量数据进行自动化分析,因此被大量应用于城镇建设、交通规划、地表勘测等重要方面,具有重大的民生意义.

传统的目标检测方法,一般按照人工选择检测窗口、获取图像特征、设计分类器的步骤进行检测,这种方式会产生大量无关特征,增加算法复杂度的同时影响模型鲁棒性[1]. 随着深度学习在人工智能领域的全面铺开,基于卷积神经网络[2](Convolutional Neural Networks, CNN)的目标检测方法利用高效卷积运算的方式显著降低模型参数量,在水下生物检测[3]、车道线检测[4]、人脸识别[5]等诸多视觉场景中均取得良好结果. 目前按照检测阶段的划分,将基于CNN的目标检测方法分为双阶段检测器和单阶段检测器[6]. 双阶段检测器使用区域候选网络(Region Proposal Network, RPN)生成候选兴趣区域,相比于单阶段检测器更具有精度优势;单阶段检测器则在一个阶段内完成目标的分类与回归操作,因此在检测速度方面更胜于双阶段检测器,代表算法有YOLO(You Only Look Once)[7]系列与SSD(Single Shot MultiBox Detector)[8]系列等.

尽管CNN配合GPU(Graphics Processing Unit)等计算设备已经大幅提高目标检测模型的检测速度,但是在遥感图像领域,卷积神经网络的计算量与存储量依然是制约其工业落地的一大瓶颈. 在图1所示的两种场景下,与自然场景下的水平拍摄角度不同,遥感图像以俯瞰视角成像,所包含的地表物体属于特定目标,具有角度任意性、空间密集性、尺度多变性以及背景复杂性等特点,为解决这些难点,研究者们对此付出了巨大努力. 对于小目标的检测改进,现有的遥感图像检测方法往往会引入特征金字塔网络[9](Feature Pyramid Network, FPN)并在其基础上进一步改良,用以更好的融合多维度特征,较为出名的有PANet[10]、AugFPN[11]以及BiFPN[12]等. 然而,FPN及其变体引入大量计算参数,导致模型结构冗余;为适应目标的多角度特性, Ma等[13]提出的RRPN首次采取角度枚举法,利用密集锚框贴合目标的旋转角度,类似的方法还有EAST(Efficient and Accuracy Scene Text)[14]以及R2CNN(Rotational Region CNN)[15]等等,但是枚举过程需要大量的卷积计算支撑,频繁的数据输入输出增加多余计算量;在暴力枚举的思想基础上,Yang等[16]提出R^{3}Det算法,利用特征精修模块(Feature Refined Module,FRM)同时提高检测速度与精度;Ding等[17]则是使用快速的旋转池化方式(PS RoI Pooling),在不需要增加锚框数的情况下,通过削减特征通道数来提高双阶段检测器的检测效率. 上述方法从训练方式、特征融合、锚框设计等不同角度对遥感图像检测方式进行探索,在一定程度上改善了遥感图像的多类难题,但是由于网络仍然采用大量的普通卷积计算,在不考虑计算载体性能的情况下,当图像大小以及像素密度增加时,其时间复杂度将迅速提高,使得检测模型的推理时间很难满足实际应用场景下实时性的要求. 因此,如何在保证检测精度的情况下,实现一种快速的遥感图像目标检测方法是值得研究的一个方向.



(*a*) 水平拍摄角度　　(*b*) 俯瞰拍摄角度

图1　两种拍摄场景下的目标对比

鉴于遥感场景下目标检测模型效率不高的问题,为减少模型计算量,降低算法实际运行时间,结合模型轻量化思想,提出一种小体积计算、低延时功耗的级联式逆残差(Cascaded Inverted Residual Convolution, CIRC)轻量卷积方式,并利用CIRC的叠加性构建新的特征提取网络,同时对h-swish函数进行改进来优化基础网络的性能. 本文在RetinaNet[18]的结构基础上引入旋转角度变量并结合CIRC,设计了基于级联式逆残差结构的目标检测网络(Cascaded Inverted Residual Convolution Net, CIRCN),实现对遥感图像目标的快速精准检测.

2 模型轻量化

相较于传统神经网络,卷积神经网络利用稀疏连接和权值共享有效减少模型训练参数[19],在图像处理方面拥有独特的优越性. 可是,普通卷积难以满足特定场景及移动端设备对模型低延时的要求,这促使卷积神经网络模型结构开始朝轻量化方向发展. 模型轻量化的方式主要包括结构轻量化、模型压缩、知识蒸馏、量化剪枝等方面. 其中,结构轻量化直接从模型设计层面出发,设计轻便式网络结构来有效减少计算量,并降低模型实际运行时间. 2017年,Landola等[20]提出

SqueezeNet，该网络使用 卷积代替部分 卷积，并通过减少输入通道数以及后置降采样来减少参数；同年，谷歌提出 MobileNet[21]，引入深度可分离卷积[22]构建网络，大大减少模型大小和计算量；为解决 MobileNet 使用 ReLU（Rectified Linear Unit）[23]函数破坏低维度特征信息的问题，MobileNetV2[24]在其基础上提出倒残差块（Inverted Residuals Block），该残差块结构只在输入与输出矩阵形状相同时使用，对输入通道先升维后降维，并使用 Linear 函数替换最后一层的激活函数 ReLU6，从而避免非线性函数破坏太多信息；MobileNetV3[25]则使用自动网络架构搜索技术[26]（Neural Architecture Search，NAS）来寻找最佳的神经网络结构，但这种方式在搜索阶段需要耗费巨大计算量，对设备要求较高；除了 MobileNet 系列外，旷视提出的 ShuffleNet[27]系列网络提出分组点卷积（pointwise group convolutions）和通道混排（channel shuffle）方法，增加通道间信息交互的同时有效降低计算量．此外，Wang 等[28]提出的 PeleeNet 网络以及 Han 等[29]提出的 GhostNet 都在上述轻量化代表模型的思想上进行改进，并取得较好的效果．

3 RetinaNet 算法

2018 年，Lin 等[18]提出一种新的损失函数 Focal Loss，用以解决单阶段检测器速度快但是精度落后于双阶段检测器的问题．Lin 等人认为，造成精度落后问题的核心原因在于图像中前背景（foreground-background）样本的极端不均衡导致的．为了解决这一问题，Lin 等这样定义 Focal Loss：

$$\mathrm{FL}\left(p_t\right)=-\left(1-p_t\right)^{\gamma}\log\left(p_t\right) \tag{1}$$

$$p_t=\begin{cases}p & ,\text{if } y=1\\ 1-p & ,\text{otherwise}\end{cases} \tag{2}$$

其中，γ 为常数，p 表示类别 y 为 1 时所得到的预测概率．

可以发现，当 $\gamma=0$ 时，Focal Loss 即为普通的交叉熵函数，若 $\gamma>0$，p_t 的增加会导致系数 $\left(1-p_t\right)^{\gamma}$ 的降低，这就有效降低了简单样本的影响，使得模型更注重于困难样本的训练．为了验证 Focal Loss 的效果，Lin 等设计了 Retinanet 算法，RetinaNet 网络结构可分为残差网络[30]（Residual Network，ResNet）、特征金字塔网络（Feature Pyramid Network，FPN）、分类子网络和回归子网络四个部分．首先，图像作为 ResNet 的输入，有效提取相应特征；其次，将相应特征通过 FPN 进行多尺度提取并强化利用，从而获得包含多层上下文语义信息的特征图（feature map）；最后，将特征图送入分类子网络（class subnets）和回归子网络（box subnets）进行对象的分类与边框回归．

RetinaNet 算法在单阶段检测算法中取得了较好的效果．原文实验结果表明，当骨干网络（backbone）选取为 Resnet-101，图片输入的分辨率为 800×800 时，RetinaNet 的平均精准率（Average Percision，AP）超越双阶段检测器中的 Faster R-CNN 算法，使得单阶段检测器在耗时更低的情况下，也能具备比双阶段检测器更优的性能．

4 CIRCN 网络

4.1 网络整体结构

图 2 展示了 CIRCN 的网络整体结构，网络共分为 3 个部分：基础网络（CIRC 网络）、特征金字塔网络以及分类与回归子网络．CIRC 网络主要由 卷积核（图例蓝箭头所示）、卷积核（图例绿箭头）以及 CIRC 模块（图例红箭头所示，具体操作详见 4.2 节）按照不同步长及重复次数顺序堆叠，用以扩张特征图通道，提取图像初始特征；特征金字塔网络将基础网络生成的特征提取层 $\boldsymbol{C}_3(100\times100)$、$\boldsymbol{C}_4(50\times50)$ 和 $\boldsymbol{C}_5(25\times25)$ 作为输入，同时以 $\boldsymbol{C}_5$ 为基本特征层，首先进行上采样操作（图例黄箭头所示），两者进行加法操作（图例蓝方块所示）得到 $\boldsymbol{P}_3(100\times100)$、$\boldsymbol{P}_4(50\times50)$ 与 $\boldsymbol{P}_5(25\times25)$ 三个尺度的特征图；之后进行卷积操作（图例紫箭头所示），得到尺寸更小的 $\boldsymbol{P}_6(13\times13)$ 和 $\boldsymbol{P}_7(25\times25)$ 两个特征图．分类与回归子网络分别由五个 3×3 大小的卷积层顺序组成，其中，除最后一层外，其余四层均添加 Relu 函数．分类网络产生不同类别的可能得分，并通过 Sigmoid 函数形成概率分布；回归子网络则是在原有四元组坐标偏置基础上增加方向信息，产生偏置五元组 $(d_x,d_y,d_w,d_h,d_\theta)$，利用五参数表示法生成旋转检测框并对其可视化．

4.2 卷积轻量化

4.2.1 深度可分离卷积单元

为有效减少计算参数，本文使用深度可分离卷积替代标准卷积对 CIRC 网络进行相关设计．基于通道域与空间域相互独立的假设，深度可分离卷积将标准卷积拆分为深度卷积（depthwise convolution）与点卷积（pointwise convolution）两个部分．图 3 和图 4 描述了普通卷积与深度可分离卷积的执行过程，假设输入特征图 F 的尺寸为 $H_F\times W_F\times C_m$，经过尺寸为 $H_K\times W_K\times C_m\times C_n$ 的标准卷积核 K 后得到尺寸为 $H_F\times W_F\times C_n$ 的输出特征 F'，所需要的计算量为：

$$H_K\cdot W_K\cdot C_m\cdot C_n\cdot H_F\cdot W_F \tag{3}$$

其中，$H_K\cdot W_K$ 与 $H_F\cdot W_F$ 分别是卷积核 K 和输入特征 F 的大小，C_1 与 C_2 是输入特征和输出特征的通道数．

在输入与输出特征尺寸相同的情况下，若使用深度可分离卷积进行操作，深度卷积首先对每个输入通道应用单个卷积核，此时所需要的计算量为：

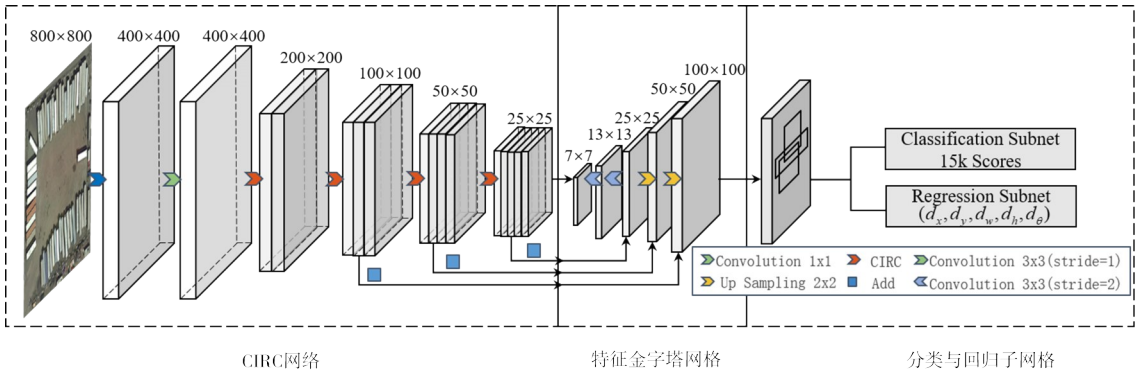


图2　CIRCN的网络整体结构

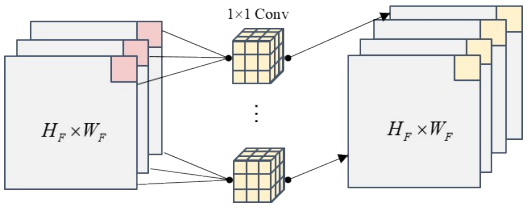


图3　标准卷积执行过程

$$H_K \cdot W_K \cdot C_m \cdot H_F \cdot W_F \tag{4}$$

此时,再使用大小为1×1的卷积核(点卷积)对深度卷积输出结果进行线性组合,此时需要的计算量为:

$$C_m \cdot C_n \cdot H_F \cdot W_F \tag{5}$$

因此,通过拆分卷积为深度方向和点方向两个步骤的方式,深度可分离卷积相较于标准卷积的计算量减少率为:

$$\frac{H_K \cdot W_K \cdot C_m \cdot H_F \cdot W_F + C_m \cdot C_n \cdot H_F \cdot W_F}{H_K \cdot W_K \cdot C_m \cdot C_n \cdot H_F \cdot W_F} = \frac{1}{C_n} + \frac{1}{H_K \cdot W_K} \tag{6}$$

式6中可以看出,深度可分离卷积大大减少了原有标准卷积的计算量,例如MobileNet使用3×3的卷积核时,在精度损失很小的情况下降低了近9倍的计算时间. 因此,本文使用深度可分离卷积作为基础卷积单元是有效的.

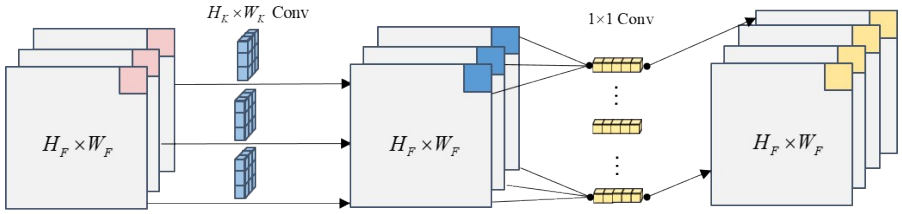


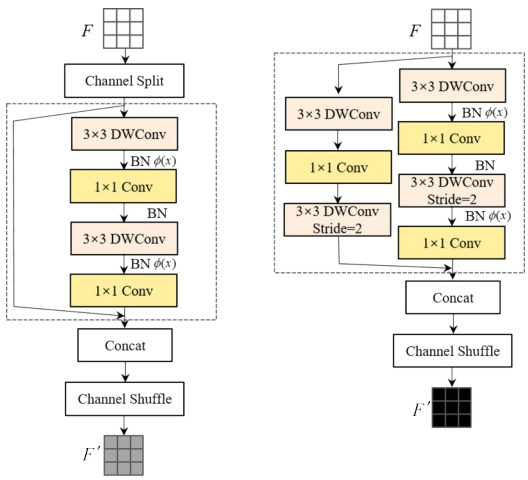
图4　深度可分离卷积执行过程

4.2.2　级联式逆残差结构

深度可分离卷积可以有效降低模型FLOPs(Floating Point of Operations),但是它却忽略了必要的逐元素操作(element-wise operation)所带来的内存访问成本(Memory Access Cost, MAC)提高,包括激活函数、张量加法等等.(除此之外,频繁的IO读写、GPU并行策略以及卷积核的加载等因素均会影响模型的检测速度.)因此,综合考虑上述因素及遥感图像的相关特性,本文提出包含通道混排的级联式逆残差模块,如图5所示.

4.2.2.1　通道拆分与混排

根据输入输出通道数相同可以最大化降低内存访问成本[29]的原则,如图5(*a*)所示,在模块初始阶段(步长为1)进行通道拆分操作. 假设输入特征F通道数为L,在通道拆分层将其均分为$L_1=L_2=L/2$,L_1进入后续卷积层得到高维特征$\widetilde{L_1}$,L_2利用跳跃连接直接与$\widetilde{L_1}$作



(*a*) 步长为1　　(*b*) 步长为2

图5　两种步长的CIRC结构

拼接,使得输出特征通道数等同于输入特征.在图5(b)情况下,由于卷积步长为2,此时通道数需要扩充至原有数目的2倍,因此并不需要进行通道拆分.此外,由于遥感图像目标的特征通道繁杂,在骨干网络生成的$\boldsymbol{C}_3$、$\boldsymbol{C}_4$和$\boldsymbol{C}_5$层,通道数分别达到了512、1 024和2 048,若在通道拼接后不作任何处理,会导致通道信息沟通阻塞,弱化重要的目标特性,包括边缘轮廓以及内部纹理等等.因此,为增强多通道间联系,在拼接操作之后使用ShuffleNet的通道混排(channel shuffle)操作:

$$H_F\times W_F\times C_m\xrightarrow{\text{reshape}}H_F\times W_F\times(g\cdot n) \tag{7}$$

$$H_F\times W_F\times(g\cdot n)\xrightarrow{\text{transpose}}H_F\times W_F\times(n\cdot g) \tag{8}$$

其中,C_m表示特征图F的通道数,$(g\cdot n)$表示对C_m分成g组,每组包含n个通道;之后对四维矩阵$H_F\times W_F\times g\times n$进行$g$和$n$维度上的转置,最后对矩阵重新进行维度融合,得到混排后的特征图矩阵$H_F\times W_F\times C_t$,通道混排的流程如图6所示.

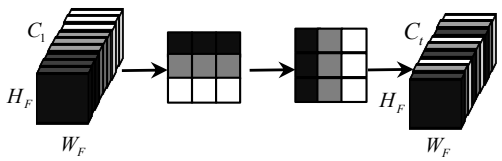


图6 通道混排流程

4.2.2.2 逆残差卷积级联

正向残差块利用逐点卷积对特征图通道先降维再升维的做法,在自然图像的检测模型中已被证明有助于提高精度[28].然而,卷积层提取的特征取决于原始特征维度,由于遥感图像的目标尺度差距大,排列密集程度高,因此需要更丰富的特征.若利用正向残差先对特征图进行压缩,卷积层所提取的特征会更加有限,从而导致检测精度的降低.因此,本文对残差模块进行改进,结合SandGlass模块思想[31],设计图5(a)、图5(b)虚线框内所示的级联式逆残差结构,该结构分为顺序分支和跳跃分支,分别可作如下数学表达:

$$E=\delta_{1,a}(\delta_{1,b}(F)) \tag{9}$$

$$\tilde{E}=\delta_{2,a}(\delta_{2,b}(E))+F \tag{10}$$

$$\hat{E}=\delta_{2,b}(E)+F \tag{11}$$

其中,F表示输入特征图,$\delta_{i,a}$与$\delta_{i,b}$函数表示第i次点卷积和深度卷积,$\tilde{E}$和$\hat{E}$表示顺序分支和跳跃分支的计算值.顺序分支由一个深度卷积级联一个逆残差模块组成,首先,深度卷积属于轻量化卷积方式,几乎不会对计算成本造成影响,同时可以弥补1×1卷积无法编码空间特征信息的局限性,保证目标空间信息的完整性;其次,逆残差模块中的两个逐点卷积对特征图维度先升维再降维,这一过程可以看作是残差的"逆过程",目的是为了增加深度卷积所处理的通道数,从而丰富特征数量,进一步提高检测精度.跳跃分支(步长为2)由两个深度卷积和一个逐点卷积组成,末尾的深度卷积用于特征图的缩放.本文相比ShuffleNet增加了跳跃分支的连接宽度,这有助于顶层高维特征保留更多的底层信息,从而有效缓解遥感图像中小目标特征的丢失问题,降低检测误差率.

4.2.2.3 激活函数改进

在低维空间尽可能保留感兴趣特征是十分必要的.常用的激活函数Relu因其以0为界的分段特性,使得对低维空间的特征流形破坏较大[25].因此,在最后一层逐点卷积之后,使用线性激活函数替代原有的Relu函数.同时,Relu函数的输出上限为无穷大,这会导致在低精度的情况下出现数值溢出,从而造成精度损失.因此,考虑到激活函数的缺陷以及计算成本,本文将h-swish函数设置为深度卷积之后新的激活函数,并对其进行简单改进,将h-swish函数的上限值设置为6.改进后的h-swish函数公式如下:

$$\phi(x)=\begin{cases}x\dfrac{\text{ReLU6}(x+3)}{6} & ,\phi(x)\leqslant 6\\ 6 & ,\phi(x)>6\end{cases} \tag{12}$$

4.2.3 整体流程

基于本文所设计的CIRC模块,将轻量级骨干网络的整体架构按表1所示方案设计如下.网络主要由级联式逆残差模块按照不同步长堆叠而成,前两层采用标准卷积,主要目的是对输入图像进行初始降采样与通道线性变换.阶段2至阶段5采用CIRC模块,根据已有研究[24],逆残差的卷积通道扩张倍数设置为5~10区间内时,会产生几乎相同的性能,因此,本文将通道扩张倍数设置为6.每个模块均按照步长先2后1的顺序进行卷积操作,考虑到计算量,将步长为2的卷积操作重复次数设置为1,步长为1的卷积操作次数在阶段2和阶段3设置为7,阶段4和阶段5设置为3.

5 实验评估

为了更好的探究基于CIRC的轻量化模型CIRCN对遥感图像目标的检测性能,在DOTA[32]数据集上设计精度与速度的对比实验.实验运行环境见表2.

5.1 数据集介绍

DOTA(Dataset for Object Detection in Aerial Images)是遥感图像领域的大规模公开基准数据集,用以评估计算机视觉任务的模型性能.本文使用的DOTA-v1.0版本包含来自多个传感器平台的约2 800张图像,图像大小从800×800像素到4 000×4 000像素不等,并按照1∶2∶3的数量比例分为验证集、测试集和训练集,数据集共标记15种常见类别的188 282个检测对象,包

表1　基于CIRC的基础网络整体流程

Layers	输入	算子类型	步长	输出	重复次数	扩张倍数	输出通道
Conv2d 3×3	800×800	3×3	2	400×400	1	—	64
Conv2d 1×1	400×400	1×1	1	400×400	1	—	128
阶段2($\boldsymbol{C}_2$)	400×400	bottleneck	2	200×200	1	6	256
	200×200	bottleneck	1	200×200	7	6	
阶段3($\boldsymbol{C}_3$)	200×200	bottleneck	2	100×100	1	6	512
	100×100	bottleneck	1	100×100	7	6	
阶段4($\boldsymbol{C}_4$)	100×100	bottleneck	2	50×50	1	6	1 024
	50×50	bottleneck	1	50×50	3	6	
阶段5($\boldsymbol{C}_5$)	50×50	bottleneck	2	25×25	1	6	2 048
	25×25	bottleneck	1	25×25	3	6	

表2　实验运行环境

类别	环境条件
服务器	H3C UniServer R4960 G3
CPU	Intel(R) Xeon(R) Gold 5218 @ 2.30 GHz×64
GPU	NVIDIA Tesla V100-SXM2-16 GB
内存	253 GiB
操作系统	Ubuntu 18.04.5 LTS
AI框架	Tensorflow1.13
CUDA版本	Cuda11.1
cuDNN版本	cuDNN8.0.2
运行环境	Python3.6

括直升机(Helicopter, HC)、游泳池(Swimming Pool, SP)、港口(Harbor, HA)、环岛(Roundabout, RA)、足球场(Soccer Ball Field, SBF)、储罐(Storage Tank, ST)、篮球场(Basketball Court, BC)、网球场(Tennis Court, TC)、船舶(Ship, SH)、大型车辆(Large Vehicle, LV)、小型车辆(Small Vehicle, SV)、田径场(Ground Track Field, GTF)、桥梁(Bridge, BR)、棒球场(Baseball Diamond, BD)和飞机(Plane, PL).

5.2　评估标准

评估标准按照性能指标分为精度评估标准与速度评估标准两大类. 精度评估根据平均精度均值(mean Average Precision, mAP)进行评判,平均精度均值是平均检测精度(Average Percision, AP)在多类别条件下的平均值,该指标融合召回率(Recall)与精准率(Percision),是目前目标检测模型最重要的精度评估指标;速度评估标准包括局部指标与整体指标,局部指标包括模型参数量、图像预处理时间、内存读取时间、实际运行时间,整体速度指标使用每秒检测帧数展开评估. 其中,模型参数量表征模型容量与计算量,包括图节点的权重偏置以及卷积层参数;图像预处理时间指的是测试集原始图像被裁切为800×800 pixel规格子图的过程用时;内存读取时间包括模型加载和逐元素操作等消耗用时,实际运行时间指的是模型利用GPU在网络图结构上的实际推理时间,反映检测网络的实际效能;每秒检测帧数通过计算每秒处理的图像数,进而判断模型的整体检测速度.

5.3　参数设置

合理的参数调整对神经网络模型的训练与推理是有利的. 数据预处理阶段,由于单张遥感图像分辨率过大,将数据集图像沿正向横纵轴统一裁剪为800×800像素的子图作为模型输入,子图像重叠步长为150像素. 锚框(Anchor)设计阶段,为适应遥感图像目标的尺度特点,在五层特征图($\boldsymbol{P}_3$、$\boldsymbol{P}_4$、$\boldsymbol{P}_5$、$\boldsymbol{P}_6$、$\boldsymbol{P}_7$)的每一个像素点预设15个锚框,锚框尺度像素设置为32、64、128、256和512,横纵比设置为{1/5, 1/3, 1/2, 1, 2, 3, 5},缩放比为$\{1, 2^{1/3}, 2^{2/3}\}$. 实验使用4块GPU开展训练与推理,批处理大小(batchsize)设置为8. 模型进行810 003次训练迭代,初始学习率(Learning Rate, LR)设置为$8e^{-5}$,在23 k次迭代间均匀攀升至$5e^{-4}$并保持不变,在650 k次迭代降至$5e^{-5}$.

5.4　对比实验

5.4.1　速度对比实验

为评估本文模型性能,在上述参数配置下,进行检测速度对比实验. 实验采用DOTA训练集与验证集开展训练过程,推理图像则采用测试集. 实验共分析三类不同基础网络配置的RetinaNet-R模型以及两类常用遥感目标检测方法与CIRCN在DOTA上的速度性能差异. 其中,RetinaNet-R表示加入角度信息后的RetinaNet算法,三类基础网络分别为ResNet50、MobileNetV2以及DarkNet53[33],两种遥感图像目标检测算法为R2CNN算法[15]和RRPN算法[13],模型共进行900k次训练迭代,测试集图片共划分为15 655张800×800像素的子图. 记录检测模型的骨干网络参数量(Backbone Network Parameters, BNP)、网络参数量(Network Parameters, NP)、图像预处理时间(Preprocess Time, PT)、内存读取时间(Memory Access Cost, MAC)、实际运行时间(Running Time, RT)以及每秒检测帧数如表3所示.

表3 不同模型的速度性能比较

Models(backbone)	BNP	NP	PT	MAC	RT	Speed
RetinaNet-R(基准)(ResNet50)	101 MiB	378 MiB	7.94 s	211.74 s	1 032.20 s	12 fps
RetinaNet-R(MobileNetV2)	**8.3** MiB	**92** MiB	7.54 s	60.52 s	**265.97** s	**45** fps
RetinaNet-R(DarkNet53)	216 MiB	464 MiB	8.12 s	124.85 s	537.34 s	23 fps
R^2CNN	—	353 MiB	8.37 s	3 452.07 s	15 746.39 s	<1 fps
RRPN	—	348 MiB	8.28 s	674.85 s	2 412.92 s	5 fps
CIRCN(CIRC)	11.3 MiB	97 MiB	**7.46** s	**47.56** s	294.12 s	**42** fps

实验结果可由两方面展开分析:(1)整体上,相较于两种常见的遥感图像目标检测算法,本文算法的参数量只有97 MiB,整体缩减了约4倍体积,这极大提高模型的存储性能,使得工业设备便于开展进一步的多模型部署优化;同时,CIRCN在测试集上的内存读取和实际运行时间达到约47.46 s和294.12 s,耗时只有RRPN算法的7%和12%,检测速度达到42 fps,是RRPN算法速度的8倍,R2CNN算法速度的40倍,这表明本文算法在遥感图像领域具有明显的效率优势.(2)局部上,相较于不同基础网络配置的RetinaNet-R算法,CIRC参数量只有ResNet50与DarkNet53的11%与5%,但略高于MobileNetV2;算法的内存读取时间和实际运行时间相比基础网络为ResNet50的基准算法,减少了约近80%耗时,每秒检测帧数达到基准算法的3.5倍,这证明了CIRC模块在存储与运算上的高效性.

5.4.2 精度对比实验

为全面可靠的对模型性能进行评估,继续采用上述算法开展精度对比实验. 其中,R2CNN算法和RRPN算法均以Faster-RCNN作为基础网络,在相同实验环境基础上重新训练至收敛,符合控制变量要求. 所有算法模型均在DOTA测试集上推理,结果交由DOTA官方服务器获得不同类别的平均检测精度AP与mAP,如表4所示. 图7展示了CIRCN在DOTA上的类别可视化检测结果.

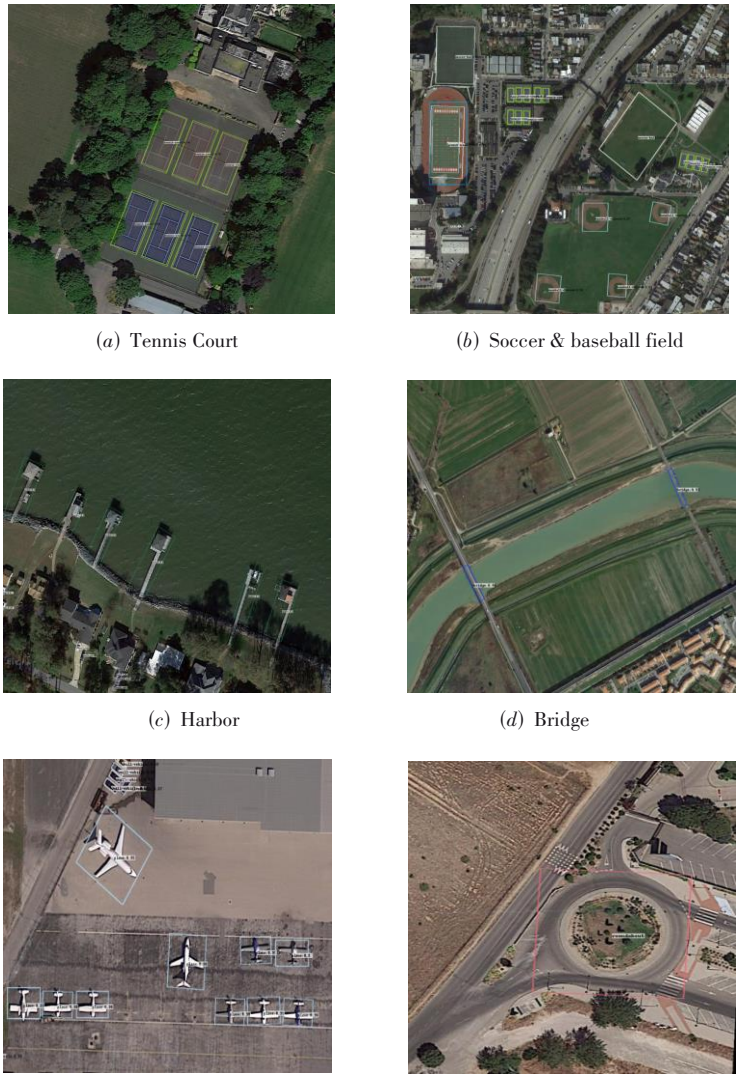
表4 不同算法在DOTA数据集上的检测精度(mAP) 单位:%

类别	RetinaNet-R (ResNet50)	RetinaNet-R (DarkNet53)	RetinaNet-R (MobileNetV2)	R^2CNN	RRPN	CIRCN
PL	**92.63**	90.95	86.81	80.94	88.52	84.52
BD	**74.13**	69.01	61.13	65.67	71.20	59.84
BR	**37.26**	36.16	26.38	35.34	31.66	30.03
GTF	44.67	42.19	36.59	**67.44**	59.30	47.45
SV	52.38	50.13	41.42	**59.92**	51.85	47.30
LV	63.79	62.84	52.43	50.91	56.19	**69.42**
SH	57.61	56.81	47.09	55.81	57.25	**64.48**
TC	**91.64**	90.96	82.77	90.67	90.81	85.78
BC	66.70	66.36	55.62	66.92	**72.84**	62.63
ST	**84.15**	83.76	73.34	72.39	67.38	78.70
SBF	**67.39**	66.17	55.88	55.06	59.69	61.49
RA	68.44	**68.63**	59.11	52.23	52.84	59.13
HA	47.53	**48.43**	36.32	55.14	53.08	57.51
SP	52.19	**55.92**	41.49	53.35	51.94	45.30
HC	35.16	35.16	18.25	48.22	**53.58**	43.51
mAP	**62.37**	61.57	51.64	60.67	61.01	**59.14**

速度和精度对比实验的综合结果表明:(1)相比R2CNN算法与RRPN算法,本文算法在速度提升40倍与8倍的情况下,mAP下降了约2%和1%. 这证明本文算法在精度损失很小的情况下,极大提升了遥感图像目标的检测速度;(2)相较于不同基础网络配置的RetinaNet-R算法,基础网络为MobileNetV2的检测速度虽然略高于CIRC,但是两者的检测精度相比基准ResNet分别下降10.73%与3.2%,前者的精度降幅远高于后者,难以达到实际应用的要求,CIRC的精度损失对于高实时场景下的遥感图像目标检测可视化是可以接受的,这也反映出CIRC模块的轻量化结构设计更适用于遥感图像的目标特点.

由此可知,本文算法与其他算法在遥感图像不同

指标的比较上均取得较优结果,生成模型在高实时检测状态下保证了检测精度,具备良好的可靠性.



(*a*) Tennis Court　(*b*) Soccer & baseball field

(*c*) Harbor　(*d*) Bridge

(*e*) Plane　(*f*) Roundabout

图7　DOTA部分类别检测结果

5.5　消融实验

为进一步判别CIRC内部模块对遥感图像目标检测的性能影响,本文消融实验按如下步骤进行:设置基准模型、分离网络模块、进行模型训练和测试图像推理. 实验以普通逆残差结构为基准模块,激活函数采用Relu函数,之后在基准模块上依次加入通道混排、级联逆残差以及改进后的h-swish函数. 为保证实验结果的可对比性,模型的训练与测试参数均保持严格一致. 表5给出不同子配置网络消融后的精度变化.

实验结果显示,通道混排操作通过融合遥感图像的多通道特征来增强目标的多维信息,达到初步强化的效果,提高模型2.13%的mAP值;级联逆残差模块通过更宽的跳跃连接融合高维度与低纬度的特征,对检测精度的提升作用最为显著,提高模型4.45%的mAP值;由于实验数据采用较高的float32精度,因此激活函数的作用并不显著,提高0.85%的mAP值. 因此,三种模块均促进了模型检测精度的提升,这证明了CIRC结构设计的有效性.

表5　CIRC模块消融实验

Methods	Channel Shuffle	CIRC	h-swish-6	mAP/%
Benchmark	Exclude	Exclude	Exclude	51.71
1	Include	Exclude	Exclude	53.84
2	Include	Include	Exclude	58.29
3(CIRCN)	Include	Include	Include	59.14

6 结束语

本文提出了一个基于级联式逆残差网络的遥感图像轻量目标检测算法,在基础网络中使用深度可分离卷积快速提升模型计算能力,并提出级联式逆残差卷积结构,通过强化目标多维特征,提高模型对目标的检测效果. 在DOTA数据集上的实验结果表明,在精度略受损失的情况下,本文算法相比基准算法大幅提升了对遥感图像目标的检测速度,相比常用的遥感图像目标检测算法同样具备较大的速度优势. 由于级联式逆残差卷积结构属于轻量级卷积,在后续研究中将继续优化该模块的设计架构,例如进行更深层次的堆叠,或者引入注意力机制,从而弥补模型在精度损失上的不足;同时尝试移植到其它检测网络,并采用结构重参数化的训练方法,提升算法的普适性与鲁棒性.

参考文献

[1] 罗会兰, 陈鸿坤. 基于深度学习的目标检测研究综述[J]. 电子学报, 2020, 48(6): 1230-1239.
LUO H L, CHEN H K. Survey of object detection based on deep learning[J]. Acta Electronica Sinica, 2020, 48(6): 1230-1239. (in Chinese)

[2] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

[3] 冀大雄, 方文巍, 朱华, 等. 基于相对测量的水下机器人主动定位方法研究[J]. 电子学报, 2021, 49(7): 1249-1256.
JI D X, FANG W W, ZHU H, et al. Active localization of autonomous underwater vehicle using noisy relative measurement[J]. Acta Electronica Sinica, 2021, 49(7): 1249-1256. (in Chinese)

[4] 徐频捷, 陈逸杰, 李之南, 等. 基于事件驱动的车道线识别算法研究[J]. 电子学报, 2021, 49(7): 1379-1385.
XU P J, CHEN Y J, LI Z N, et al. Research on event-driven lane recognition algorithms[J]. Acta Electronica Sinica, 2021, 49(7): 1379-1385. (in Chinese)

[5] 李倩玉, 蒋建国, 齐美彬. 基于改进深层网络的人脸识别算法[J]. 电子学报, 2017, 45(3): 619-625.
LI Q Y, JIANG J G, QI M B. Face recognition algorithm based on improved deep networks[J]. Acta Electronica Sinica, 2017, 45(3): 619-625. (in Chinese)

[6] OKSUZ K, CAM B C, KALKAN S, et al. Imbalance problems in object detection: A review[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43 (10): 3388-3415.

[7] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 779-788.

[8] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[C]//ECCV 2016. Cham: Springer International Publishing, 2016: 21-37.

[9] 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. Piscataway: IEEE, 2017: 936-944.

[10] LIU S, QI L, QIN H F, et al. Path aggregation network for instance segmentation[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 8759-8768.

[11] GUO C X, FAN B, ZHANG Q, et al. AugFPN: Improving multi-scale feature learning for object detection[C]// 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 12592-12601.

[12] TAN M X, LE Q V. EfficientNet: Rethinking model scaling for convolutional neural networks[EB/OL]. (2019-05-28)[2021-06-01]. https://arxiv.org/abs/1905.11946.

[13] MA J Q, SHAO W Y, YE H, et al. Arbitrary-oriented scene text detection via rotation proposals[J]. IEEE Transactions on Multimedia, 2018, 20(11): 3111-3122.

[14] ZHOU X Y, YAO C, WEN H, et al. EAST: An efficient and accurate scene text detector[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 2642-2651.

[15] JIANG Y Y, ZHU X Y, WANG X B, et al. R2CNN: Rotational region CNN for orientation robust scene text detection[EB/OL]. (2017-06-29) [2021-06-01]. https://arxiv.org/abs/1706. 09579.

[16] YANG X, YAN J C, FENG Z M, et al. R3Det: Refined single-stage detector with feature refinement for rotating object [EB/OL]. (2019-08-15)[2021-06-01]. https://arxiv.org/abs/1908. 05612.

[17] DING J, XUE N, LONG Y, et al. Learning RoI transformer for detecting oriented objects in aerial images [EB/OL]. (2018-12-01) [2021-06-01]. https://arxiv. org/abs/1812.00155.

[18] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]//2017 IEEE International Conference on Computer Vision. Piscataway: IEEE, 2017: 2999-3007.

[19] ALBAWI S, MOHAMMED T A, AL-ZAWI S. Understanding of a convolutional neural network[C]//2017 In-

ternational Conference on Engineering and Technology (ICET). Piscataway: IEEE, 2017: 1-6.

[20] IANDOLA F N, HAN S, MOSKEWICZ M W, et al. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size[EB/OL]. (2016-02-24) [2021-06-01]. https://arxiv.org/abs/1602.07360.

[21] HOWARD A G, ZHU M L, CHEN B, et al. MobileNets: Efficient convolutional neural networks for mobile vision applications[EB/OL]. (2017-04-17) [2021-06-01]. https://arxiv.org/abs/1704.04861.

[22] SIFRE L, MALLAT S. Rigid-motion scattering for texture classification[EB/OL]. (2014-03-07) [2021-06-01]. https://arxiv.org/abs/1403.1687.

[23] AGARAP A F. Deep learning using rectified linear units (ReLU)[EB/OL]. (2018-03-22)[2021-06-01]. https://arxiv.org/abs/1803. 08375.

[24] SANDLER M, HOWARD A, ZHU M L, et al. MobileNetV2: Inverted residuals and linear bottlenecks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 4510-4520.

[25] HOWARD A, SANDLER M, CHEN B, et al. Searching for MobileNetV3[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2019: 1314-1324.

[26] TAN M X, CHEN B, PANG R M, et al. MnasNet: Platform-aware neural architecture search for mobile[C]// 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2019: 2815-2823.

[27] ZHANG X Y, ZHOU X Y, LIN M X, et al. ShuffleNet: An extremely efficient convolutional neural network for mobile devices[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 6848-6856.

[28] WANG R J, LI X, LING C X. Pelee: A real-time object detection system on mobile devices[EB/OL]. (2018-04-18)[2021-06-01]. https://arxiv.org/abs/1804.06882.

[29] HAN K, WANG Y H, TIAN Q, et al. GhostNet: More features from cheap operations[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 1577-1586.

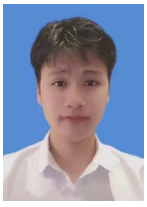
[30] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 770-778.

[31] DAQUAN Z, HOU Q B, CHEN Y P, et al. Rethinking bottleneck structure for efficient mobile network design [EB/OL]. (2020-07-05) [2021-06-01]. https://arxiv.org/abs/2007.02269.

[32] XIA G S, BAI X, DING J, et al. DOTA: A large-scale dataset for object detection in aerial images[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 3974-3983.

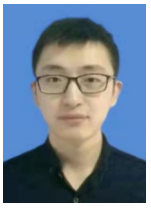
[33] REDMON J, FARHADI A. YOLOv3: An incremental improvement[EB/OL]. (2018-04-08)[2021-06-01]. https://arxiv.org/abs/1804.02767.

作者简介



陈　立　男,1997年2月出生于浙江省义乌市. 信息工程大学硕士生. 主要研究方向为计算机视觉.
E-mail: 2464863136@qq.com

张　帆(通讯作者)　男,1981年9月出生. 博士. 现为国家数字交换系统工程技术研究中心副研究员、硕士生导师. 主要研究方向为主动防御、人工智能、高性能计算. 中国电子学会会员编号:E190013697M.
E-mail: 17034203@qq.com



郭　威　男,1990年8月出生. 博士. 现为国家数字交换系统工程技术研究中心助理研究员. 主要研究方向为主动防御、人工智能、高性能计算. 中国电子学会会员编号:E190029991M.
E-mail: guowjss@126.com

黄　赟　男,1993年9月出生于江西省新余市. 信息工程大学硕士生. 主要研究方向为神经网络模型量化压缩、网络内生安全.
E-mail: yyhuangz@163.com

李继中　男,1983年4月出生于河南省上蔡市. 博士. 主要研究方向为大数据与人工智能.
E-mail: zhongzhong_hero@163.com