



## 隔级融合特征金字塔与CornerNet相结合的小目标检测

赵文清, 孔子旭, 赵振兵

引用本文:

赵文清, 孔子旭, 赵振兵. 隔级融合特征金字塔与CornerNet相结合的小目标检测[J]. 智能系统学报, 2021, 16(1): 108–116.  
ZHAO Wenqing, KONG Zixu, ZHAO Zhenbing. Small target detection based on a combination of feature pyramid and CornerNet[J]. *CAAI Transactions on Intelligent Systems*, 2021, 16(1): 108–116.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202004033>

## 您可能感兴趣的其他文章

### 基于注意力机制的显著性目标检测方法

Salient object detection method based on the attention mechanism

智能系统学报. 2020, 15(5): 956–963 <https://dx.doi.org/10.11992/tis.201903001>

### 基于注意力融合的图片描述生成方法

An image caption generation method based on attention fusion

智能系统学报. 2020, 15(4): 740–749 <https://dx.doi.org/10.11992/tis.201910039>

### 基于反卷积和特征融合的SSD小目标检测算法

SSD small target detection algorithm based on deconvolution and feature fusion

智能系统学报. 2020, 15(2): 310–316 <https://dx.doi.org/10.11992/tis.201905035>

### 基于跳跃连接金字塔模型的小目标检测

Skip feature pyramid network with a global receptive field for small object detection

智能系统学报. 2019, 14(6): 1144–1151 <https://dx.doi.org/10.11992/tis.201905041>

### 基于双向消息链路卷积网络的显著性物体检测

Salient object detection based on bidirectional message link convolution neural network

智能系统学报. 2019, 14(6): 1152–1162 <https://dx.doi.org/10.11992/tis.201812003>

### 多层卷积特征的真实场景下行人检测研究

Research on pedestrian detection based on multi-layer convolution feature in real scene

智能系统学报. 2019, 14(2): 306–315 <https://dx.doi.org/10.11992/tis.201710019>

微信公众平台



关注微信公众号, 获取更多资讯信息

DOI: 10.11992/tis.202004033

# 隔级融合特征金字塔与 CornerNet 相结合的小目标检测

赵文清<sup>1,3</sup>, 孔子旭<sup>1</sup>, 赵振兵<sup>2</sup>

(1. 华北电力大学 控制与计算机工程学院, 河北 保定 071003; 2. 华北电力大学 电气与电子工程学院, 河北 保定 071003; 3. 复杂能源系统智能计算教育部工程研究中心, 河北 保定 071003)

**摘要:** 为弥补 CornerNet 中小目标语义信息弱的缺陷, 提出隔级融合特征金字塔的方法, 提高小目标平均准确率。对骨干网络后半部分融合后的 4 个特征图进行提取, 将尺寸较小的特征图进行 2 次卷积, 得到 2 个新的特征图; 运用上下融合、隔级融合和旁路连接的思想, 生成融合后的特征图并将其组成特征金字塔。将改进后的算法与当前主流 CornerNet、Faster RCNN、RetinaNet 算法在 MS COCO 数据集上进行比较, 结果表明, 改进后算法在对小目标进行检测时, 小目标平均准确率有较大提高。隔级融合特征金字塔在 CornerNet 上能有效融合高低层特征图, 使融合后的特征图有较强的语义信息, 提高 CornerNet 网络的小目标平均准确率。

**关键词:** CornerNet; 小目标检测; 卷积; 特征图; 隔级融合; 上下融合; 旁路连接; 特征金字塔

**中图分类号:** TP391    **文献标志码:** A    **文章编号:** 1673-4785(2021)01-0108-09

中文引用格式: 赵文清, 孔子旭, 赵振兵. 隔级融合特征金字塔与 CornerNet 相结合的小目标检测 [J]. 智能系统学报, 2021, 16(1): 108-116.

英文引用格式: ZHAO Wenqing, KONG Zixu, ZHAO Zhenbing. Small target detection based on a combination of feature pyramid and CornerNet[J]. CAAI transactions on intelligent systems, 2021, 16(1): 108-116.

## Small target detection based on a combination of feature pyramid and CornerNet

ZHAO Wenqing<sup>1,3</sup>, KONG Zixu<sup>1</sup>, ZHAO Zhenbing<sup>2</sup>

(1. School of Control and Computer Engineering, North China Electric Power University, Baoding 071003, China; 2. School of Electrical and Electronic Engineering, North China Electric Power University, Baoding 071003, China; 3. Engineering Research Center of the Ministry of education for Intelligent Computing of complex energy system, Baoding 071003)

**Abstract:** To improve the problem of the weak semantic information of the small target in CornerNet, a method of the hierarchical fusion feature pyramid is proposed to increase the average accuracy of the small target. The method first extracts the four feature maps after the fusion of the second half of the backbone network, then convolves the feature maps with a smaller size twice to obtain two new feature maps, and finally uses the ideas of the upper and lower fusion, inter-level fusion, and bypass connection to generate a fused feature map and form it into a feature pyramid. The result shows that the average accuracy for small targets obtained by our algorithm has been greatly improved compared with those by current mainstream algorithms, such as CornerNet, Faster RCNN, and RetinaNet on the MS COCO dataset, which demonstrates great superiority. The inter-level fusion feature pyramid can effectively fuse high-level and low-level feature maps on CornerNet, so that the fused feature maps have strong semantic information, and improve the average accuracy of the small targets of the CornerNet network.

**Keywords:** CornerNet; small target detection; convolution; feature map; interval fusion; upper and lower fusion; bypass connection; feature pyramid

收稿日期: 2020-04-27.

基金项目: 国家自然科学基金项目 (61871182); 中央高校基本科研业务费面上项目 (2020MS153).

通信作者: 赵文清. E-mail: jbzqw@126.com.

目标检测的目的是使计算机像人类一样, 从图像中获取所需要的信息。这对于人眼来说一件很简单的事情, 通过观察目标的颜色、形状等特

点就能轻易地判定目标的类别及位置;但计算机“看到的”却是一个个的像素值,很难从这些像素值中抽象出目标的特征并确定目标的位置<sup>[1]</sup>。目标检测技术在日常生活中广泛应用,例如安装摄像头检测可疑人物携带的诸如作案工具等小目标,分析异常行为,从而做到对危险的预判;在智能自动驾驶中,检测前方目标并进行识别,系统作出判断并采取相应措施;在交通视频监控中,车牌、车标等小目标需要准确检测;航拍图片因为拍摄距离比较远而造成目标像素过低等等,这些问题都需要采用目标检测技术,因此,对小目标进行准确快速的检测尤为必要。小目标主要分为绝对小目标(像素低)和相对小目标(相对于原图的长宽比例小),图像大小为原图像的0.1倍或图像像素低于 $32 \times 32$ ,即可认定为小目标<sup>[2]</sup>。

目前主流的目标检测算法主要分为两大类。第1类是双阶段检测方法,代表性算法有 Faster R-CNN(towards real-time object detection with region proposal networks)<sup>[3]</sup>、SPP-Net(spatial pyramid pooling in deep convolutional networks for visual recognition)<sup>[4]</sup>、Mask R-CNN<sup>[5]</sup>等。双阶段检测算法首先针对输入图像选择候选区域,然后用卷积对产生的候选区域提取特征,最后对候选区域进行分类和位置回归。双阶段检测算法由于检测框与小目标的不匹配等问题,在进行特征提取的过程中会造成边缘信息丢失,导致检测率低下<sup>[6]</sup>。第2类是单阶段检测方法,代表性的算法有 SSD(single shot multiBox detector)<sup>[7]</sup>、YOLO(you only look once)<sup>[8]</sup>、Retina-Net<sup>[9]</sup>等。单阶段算法相较于双阶段算法不需要计算候选框区域,直接对输入图片进行回归,检测目标的类别概率和位置坐标<sup>[10]</sup>。

由于特征图语义信息弱,导致小目标检测效果差。2017年Lin等<sup>[11]</sup>提出了FPN(feature pyramid networks for object detection)网络。该算法有2条分支,第1条分支采用自上而下网络,第2条分支采用自下而上网络。通过旁路连接将2个特征图融合,从而实现多尺度目标检测<sup>[12]</sup>。本文借鉴FPN算法的核心思想,在其基础上进行改进并运用到CornerNet(detecting objects as paired key-points)算法中,用来对小目标进行检测。

## 1 CornerNet 算法

### 1.1 基于沙漏网络的 CornerNet 网络模型

CornerNet<sup>[13]</sup>模型采用沙漏网络(stacked hourglass networks for human pose estimation)<sup>[14]</sup>作为基

础模型,如图1所示。首先经过一系列降采样,将图像分辨率降低4倍(文献[13]中输入图像大小为 $511 \times 511$ ,输出图像大小为 $128 \times 128$ ),然后经过特征提取网络提取特征。该网络通过串联2个沙漏网络模块组成,沙漏网络模块中间用一个中继监督进行连接。每个沙漏网络模块都是先通过一系列步长为2的降采样操作缩小输入的大小,获得一些分辨率较低的特征,从而使计算复杂度降低,然后通过上采样恢复到输入图像大小,提高图像的分辨率,更好地预测物体的准确位置。整个沙漏网络的深度是104层,最后将该特征图作为2个预测模块的输入,预测的关键点为左上角和右下角,因此是2个模块输入。在每个模块里面,首先经过角池化,然后输出3个分支:热图、嵌入矢量和偏移。

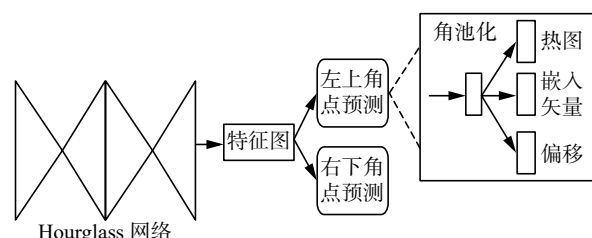


图1 CornerNet 检测模型

Fig. 1 CornerNet detection model

#### 1.1.1 预测模块详细结构

图2为左上角点的处理过程。图2中虚线框分为2路,2路处理过程一致,最后再用 ReLu 激活函数处理。2路特征图经处理后进行融合,然后用 $3 \times 3$ 的卷积核和批量归一化处理融合后的特征图,最后将处理后的特征图与骨干网络输出,并经 $1 \times 1$ 卷积和批量归一化处理后的特征图进行融合,融合后的特征图进行 ReLu 激活函数处理,经处理后的特征图再次进行 $3 \times 3$ 的卷积、批量归一化、ReLu 激活函数处理,最后3个输出分别为热图、嵌入矢量和偏移、右下角点与左上角点处理过程一致。

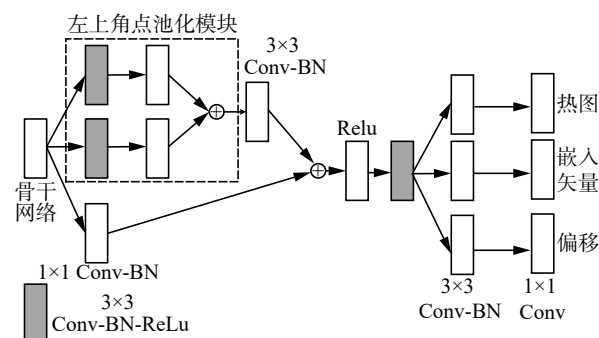


图2 角点检测模块

Fig. 2 Corner detection module

### 1.1.2 角池化

角池化通过编码来更好地定位角点。角池化层首先最大池化  $f_i$  中在  $(i,j)$  与  $(i,H)$  之间所有的特征向量,使之成为特征向量  $t_{ij}$ ,然后最大池化  $f_i$  中在  $(i,j)$  与  $(W,j)$  之间所有的特征向量,使之成为  $l_{ij}$ 。最后把  $t_{ij}$  和  $l_{ij}$  加在一起:

$$\begin{aligned} t_{ij} &= \begin{cases} \max(f_{t_{ij}}, t_{(i+1)j}), & i < H \\ f_{t_{ij}}, & \text{其他} \end{cases} \\ l_{ij} &= \begin{cases} \max(f_{l_{ij}}, l_{i(j+1)}), & j < W \\ f_{l_{ij}}, & \text{其他} \end{cases} \end{aligned} \quad (1)$$

### 1.1.3 热图

热图:预测角点位置,特征图的每个通道的值表示所定位点为角点的分数:

$$L_{\text{det}} = \frac{-1}{N} \sum_{c=1}^C \sum_{i=1}^H \sum_{j=1}^W \begin{cases} (1-p_{cij})^\alpha \log(1-p_{cij}), & y_{cij} = 1 \\ (1-y_{cij})^\beta (p_{cij})^\alpha \log(1-p_{cij}), & \text{其他} \end{cases} \quad (2)$$

式(2)是针对角点预测的损失函数,整体上是改良版的 Focal Loss(focal loss for dense object detection)。式中:  $p_{cij}$  表示预测的热图在  $C$  类的  $(i,j)$  位置的值;  $y_{cij}$  表示位置  $(i,j)$  的 ground-truth;  $N$  为图中待检测目标的数量。

### 1.1.4 偏移

偏移:输出从输入映射到特征图的误差信息。

在神经网络中存在着下采样层,因此从原始的图像输入到最后的偏移的过程会产生累计误差,因此引入偏移修正进行修正:

$$o_k = \left( \frac{x_k}{n} - \left\lfloor \frac{x_k}{n} \right\rfloor, \frac{y_k}{n} - \left\lfloor \frac{y_k}{n} \right\rfloor \right) \quad (3)$$

由于输入图像到不同尺寸的特征图之间会有尺寸缩小,因此在设计特征图时  $P_1 \sim P_6$  之间都是2的倍数。假设缩小倍数是  $n$ ,那么输入图像上的  $(x,y)$  点对应到特征图上为

$$\left( \left\lfloor \frac{x}{n} \right\rfloor, \left\lfloor \frac{y}{n} \right\rfloor \right) \quad (4)$$

式(4)中的符号是向下取整,取整会带来精度丢失问题,所以通过式(3)计算偏移,然后通过式(5)的 Smooth  $L_1$  损失函数监督学习该参数:

$$L_{\text{off}} = \frac{1}{N} \sum_{k=1}^N \text{Smooth}L_1 \text{Loss}(o_k, \hat{o}_k) \quad (5)$$

### 1.1.5 嵌入矢量

嵌入矢量:将角点配对,如果一个左上角点和一个右下角点属于一个目标,那么它们的嵌入矢量距离较小。

嵌入矢量的训练是由2个损失函数实现的。 $e_{tk}$  表示第  $k$  个目标的左上角角点的嵌入矢量,  $e_k$  表示第  $k$  个目标的右下角角点的嵌入矢量,  $e_k$  表

示  $e_{tk}$  和  $e_k$  的平均值。式(6)用来缩小属于同一个目标(第  $k$  个目标)2个角点的嵌入矢量( $e_{tk}, e_k$ )距离。式(7)用来扩大不属于同一个目标的2个角点的嵌入矢量距离。

$$L_{\text{pull}} = \frac{1}{N} \sum_{k=1}^N [(e_{tk} - e_k)^2 + (e_{bk} - e_k)^2] \quad (6)$$

$$L_{\text{push}} = \frac{1}{N(N-1)} \sum_{k=1}^N \sum_{j=1, j \neq k}^N \max(0, \Delta - |e_k - e_j|) \quad (7)$$

最后的网络损失函数为式(2)、式(5)~(7)相加的形式:

$$L = L_{\text{det}} + \alpha L_{\text{pull}} + \beta L_{\text{push}} + \gamma L_{\text{off}} \quad (8)$$

式中:  $\alpha=\beta=0.1$ ;  $\gamma=1$ 。

## 1.2 FPN 网络

2017年 Lin 等<sup>[11]</sup>提出了特征图金字塔算法。FPN 算法可以将浅层与深层的特征图进行融合,利用邻近特征图的语义信息,通过融合上下2层的特征,得到语义信息更加丰富的特征图供后续的检测,通过融合这些不同层的特征图来达到预测效果,解决了多尺度下小目标准确检测的问题<sup>[15]</sup>。目前很多算法都引入 FPN 模型来提高其检测率,例如,文献[16]在 SSD 算法中引入 FPN 算法,在 PASCAL VOC 2007 数据集上平均准确率较 SSD 提高 1.7%;文献[17]在 YOLO V3 算法中引入 FPN 算法,在 PASCAL VOC 2007 数据集上平均准确率较 YOLO V3 提高 1.58%;事实表明单一特征检测的算法引入 FPN 后平均准确率有较大提升。浅层特征图虽然具有较少的语义信息,但是保留较为准确的位置信息,而在小目标检测中,小目标对于位置的敏感度要远远大于大目标。同时,对于深层的特征图随着特征图的下采样过多导致在深层特征图丰富的语义信息中保留了大目标的特征而丢失了小目标的特征信息。文献[13]中 CornerNet 仅是对一张特征图进行检测,其忽略了浅层特征,导致小目标效果检测不是很理想。初期为了验证引入 FPN 算法的有效性,在骨干网络输出一张特征图的基础上,增加一个浅层特征图形成 2 路分支供检测,发现准确率有所提升,从而验证本文想法的正确性。本文基于 FPN 思想,充分利用 CornerNet 骨干网络输出的浅层特征图与深层特征图进行融合,形成语义信息更加丰富的 4 层与 6 层特征金字塔供检测,进而提升准确率。

## 2 隔级融合特征金字塔与 CornerNet 相结合的小目标检测模型

CornerNet 模型采用 Hourglass 作为其骨干网



络,只用该网络的最后输出图作检测。虽然特征图语义信息比较丰富,但是目标位置不精准。由于小目标尺寸过小,极易与背景融合,导致小目标检测率低下<sup>[18]</sup>。针对 CornerNet 算法对小目标检测出现漏检误检的问题,本文将骨干网络部分语义信息丰富且尺寸不同的特征图提取出来。FCOS(fully convolutional one-stage object detection)<sup>[19]</sup>算法同样是无框检测,该算法在引入特征金字塔的基础上对最小的特征图又进行了

2次卷积,目的是使得可供检测的特征图尺度更加丰富。本文基于其再次进行卷积的思想,将特征金字塔中尺寸最小的特征图进行2次卷积,进而得到更多尺度的特征图为隔级融合做准备,最后将对组成特征金字塔的特征图进行隔级融合、上下融合、旁路连接,融合后的特征图语义信息更加丰富。

图3为本文提出的隔级融合特征金字塔与 CornerNet 结合的小目标检测模型。

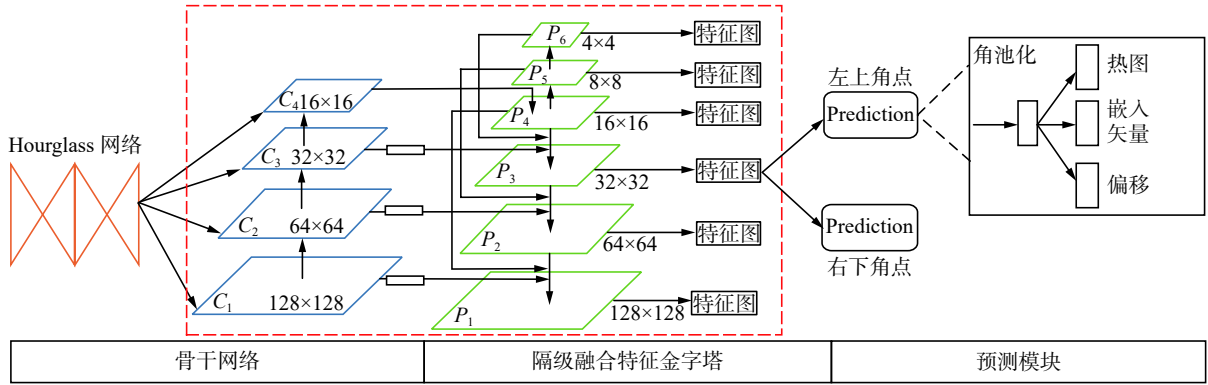


图3 隔级融合特征金字塔与 CornerNet 结合的小目标模型

Fig. 3 Small target model based on the combination of connected feature pyramid and CornerNet

## 2.1 改进后的沙漏网络

图4为一个沙漏网络模块,内部大多采用残差网络的残差块进行特征提取。过程如下:对输入的特征图进行一系列的下采样(红色方框),下采样前分出一条之路保留上采样前的特征图,然

后进行上采样(蓝色方块),上采样之后将特征图与上一个支路处理后的特征图进行融合,融合后再经过一个残差块进行特征提取,2次下采样之间有3个残差模块进行特征提取。 $C_1 \sim C_4$ 是提取出来的特征图,为隔级融合特征金字塔做准备。

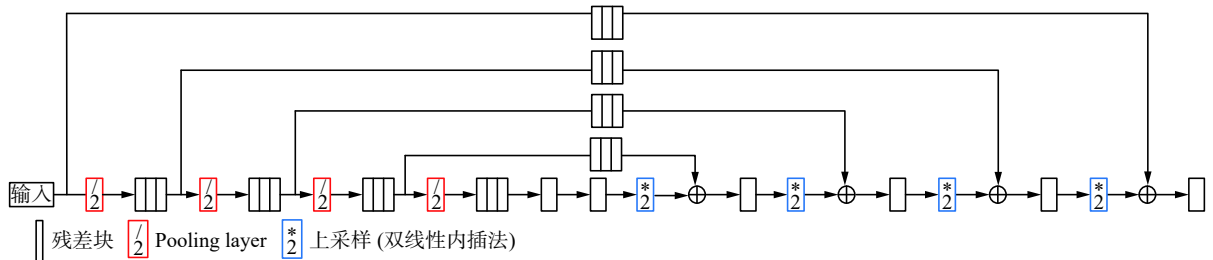


图4 沙漏网络模型

Fig. 4 Hourglass network model

图5为2种残差模块示意图。CornerNet中的沙漏网络的残差块采用图5(a)基本块,其由2个 $3 \times 3$ 卷积组成。由于改进后的网络添加了特征金字塔,导致整个模型推理速度变慢,因此本文骨干网络中的残差模块采用图5(b)中的瓶颈块。

采用图5(b)瓶颈块的残差模块,假设输入是一个 $3 \times 3 \times 128$ 的特征图,首先采用64个 $1 \times 1$ 的卷积将128通道的特征图降到64通道,然后用 $3 \times 3$ 卷积进行特征提取,最后通过 $1 \times 1$ 卷积将特征图恢复到128通道。此过程涉及到的参数数目为 $1 \times 1 \times 128 \times 64 + 3 \times 3 \times 64 \times 64 + 1 \times 1 \times 64 \times 128 = 53\,248$ 。如

不使用 $1 \times 1$ 卷积,参考图5(a),假设同样输入 $3 \times 3 \times 128$ 的特征图,第1步经过128个 $3 \times 3$ 卷积核,第2步再经过128个 $3 \times 3$ 卷积核对特征图进行卷积。整体过程的参数数目为 $3 \times 3 \times 128 \times 128 \times 2 = 294\,912$ 。2种不同情况,参数数目相差5.5倍。因此,改进后的沙漏网络在保证准确度的前提下,参数量大为减少,推理速度变快。

## 2.2 隔级融合特征金字塔模型

在卷积神经网络中,网络越深,特征图拥有的抽象特征信息就越多。因此运用隔级连接来融合高层与低层的特征图,融合出语义信息更丰富的

特征图,不仅能融合不同特征图的尺度信息,还能有效融合高低层的细节信息<sup>[20]</sup>。图6为隔级融合特征金字塔模型。

图6中骨干网络是沙漏网络,特征金字塔分为3部分:自底向上部分(红色框左半部分),中间连接部分(通过尺寸大小为 $1\times 1$ 、通道为256的卷积核对特征图进行卷积),隔级连接与特征融合部分(红色框右半部分)。

具体步骤如下:

CornerNet 骨干网络由2个沙漏网络组成。从第2个沙漏网络后半部分中提取出不同尺寸的

特征图:  $C_1$ 、 $C_2$ 、 $C_3$ 、 $C_4$ , 其尺寸大小分别为  $128\times 128$ 、 $64\times 64$ 、 $32\times 32$ 、 $16\times 16$ , 如图6所示。

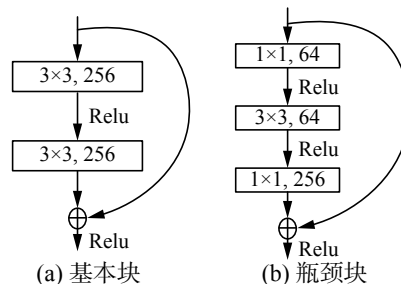


图5 残差模块

Fig. 5 Residual module

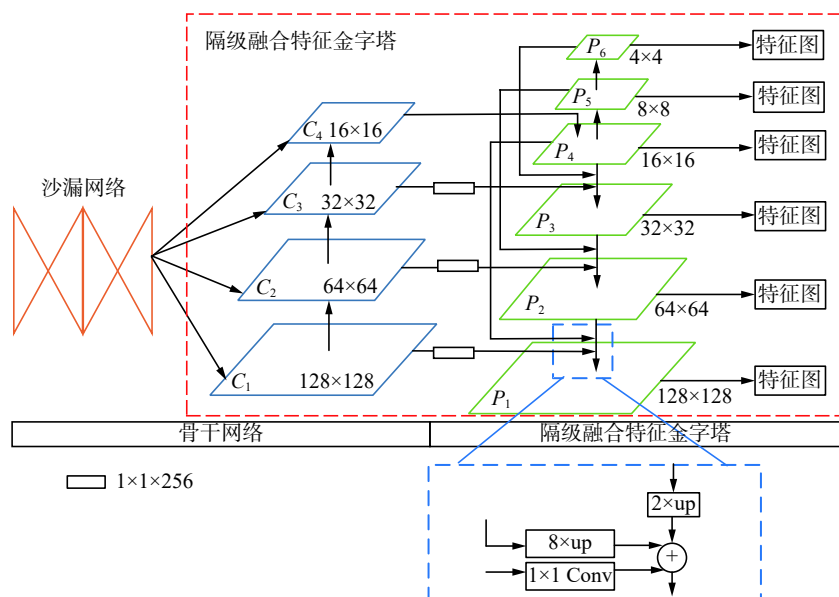


图6 隔级融合特征金字塔模型

Fig. 6 Pyramid model of feature fusion

$P_4$  就是  $C_4$ (文献[13]在预处理时已经将通道卷积成256,所以不用做任何操作)。 $P_5$ 是由 $P_4$ 经过卷积核尺寸为 $3\times 3$ 步长为2卷积得到,尺寸大小为 $8\times 8$ ;  $P_6$ 由 $P_5$ 经过卷积核尺寸为 $3\times 3$ 步长为2再次进行卷积得到,尺寸大小为 $4\times 4$ ,目的是为获得深层更加鲁棒的语义信息,为接下来的隔层连接做准备。

首先 $P_6$ 运用双线性内插法进行上采样,将其尺度扩大到 $32\times 32$ , $P_4$ 运用双线性内插法进行上采样,将其尺度扩大到 $32\times 32$ ;然后用尺寸大小为 $1\times 1$ 的卷积核改变 $C_3$ 的通道,使其通道数与 $P_4$ 保持一致;最后将处理后的3个特征图相加得到特征图 $P_3$ (如图蓝色虚线框)并采用 $3\times 3$ 卷积核对其卷积,目的是消除特征图的混叠效应。 $P_1$ 、 $P_2$ 特征图的得到的流程与 $P_3$ 流程一致。

对特征图 $P_1\sim P_6$ ,每一个特征图有2个模块:左上角和右下角的预测模块,2个模块的结构相同,每个预测模块中先经过角池化,然后输出热

图、嵌入矢量和偏移,3个分支。

热图负责预测角点位置,嵌入矢量负责将角点进行配对,偏移进行位置修订。

### 3 实验及结果分析

本文实验使用的操作系统为Ubuntu 16.04LTS, GPU为NVIDIA GeForce RTX 2080Ti,深度学习框架为pytorch1.0。

#### 3.1 实验数据集

为了避免过拟合,本文先采用VOC2007的部分数据集进行预训练,然后再采用与CornerNet相同的MS COCO数据集进行实验。MS COCO<sup>[21]</sup>数据集有80类,包含80 K的训练集、40 K的验证集以及20 K的测试集。

#### 3.2 评价指标

本文目的是提升CornerNet算法的小目标准确率,因此选取MS COCO最重要的评价指标

APS(小目标准确率) 和 AP(平均准确率)。该数据集规定目标小于  $32 \times 32$  即为小目标, 其评价指标为 APS。MS COCO 数据集的评价指标跟其他数据集评价指标有所不同, 在 MS COCO 数据集中 AP 表示在各种阈值下的平均准确率, 默认 AP 就是 mAP,  $mAP@.5IoU=AP@.5IoU$ ,  $mAP@.75IoU=AP@.75IoU$ , 以此类推。P-R 曲线指的是 Precision Recall(准确率-召回率) 曲线。准确率: 正确检测为正占全部检测为正的的比例; 召回率: 正确检测为正占全部正样本的比例。TP: 样本为正, 预测结果为正; FP: 样本为负, 预测结果为正; TN: 样本为负, 预测结果为负; FN: 样本为正, 预测结果为负, 则准确率为  $TP / (TP+FP)$ , 召回率为  $TP / (TP+FN)$ 。对于每一个阈值 ( $i$ ) 都会有相应的 P-R 曲线, 对该曲线做积分 (曲线下面的面积) 即为在该阈值下的  $AP_i$  值, 在 MS COCO 数据集中阈值一共有 10 个, 从  $IoU_1=0.5$  开始每次增加 0.05 直到  $IoU_{10}=0.95$ , 每一个阈值都有其相对应的  $AP_i$  数值, 最后将 10 个  $AP_i$  值累加除以 10 即为 MS COCO 评价指标 AP。

### 3.3 实验结果

本文以沙漏网络作为 CornerNet 的骨干网络, 采用平均准确率 (AP) 和小目标准确率 (APS) 作

为主要评价指标, 以传统 CornerNet 算法作为基准方法, 对模型各部分改进前后的效果进行比较和分析。FPN(6) 参数设置方面, 首先动量参数和权重衰减分别设为 0.9 和 0.0005, 鉴于硬件设施的局限性, 批次设置为 16, 学习率设置为 0.00025, 训练 230 K 次, 再降低 10 倍学习率, 训练 60 K 次, 当迭代 290 K 次时模型达到稳定。由于 FPN(4) 与 FPN(6) 采用相同的骨干网络, 因此将 FPN(6) 训练好的骨干网络参数导出并迁移到 FPN(4) 的骨干网络上, 动量参数和权重衰减分别设为 0.9 和 0.0005, 批次设置为 16, 学习率设置为 0.00025, 训练 120 K 次, 再降低 10 倍学习率, 训练 30 K 次, 当迭代 150 K 次时模型达到稳定。

#### 1) FPN 层数对识别结果的影响

传统的 CornerNet 算法仅对沙漏网络输出的一张特征图进行检测, 本文引入 FPN 并进行隔级融合得到多张语义信息丰富的特征图供检测。本文实验的模型有 2 个, 一个是对  $P_1 \sim P_4$ , 4 个特征图进行检测, 如图 7 红色实框所示; 另一个是  $P_1 \sim P_6$ , 6 个特征图进行检测, 如图 8 红色实框所示。通过对比 2 个模型最终平均准确率与小目标检测准确率, 得出层数对准确率影响的相关结论。

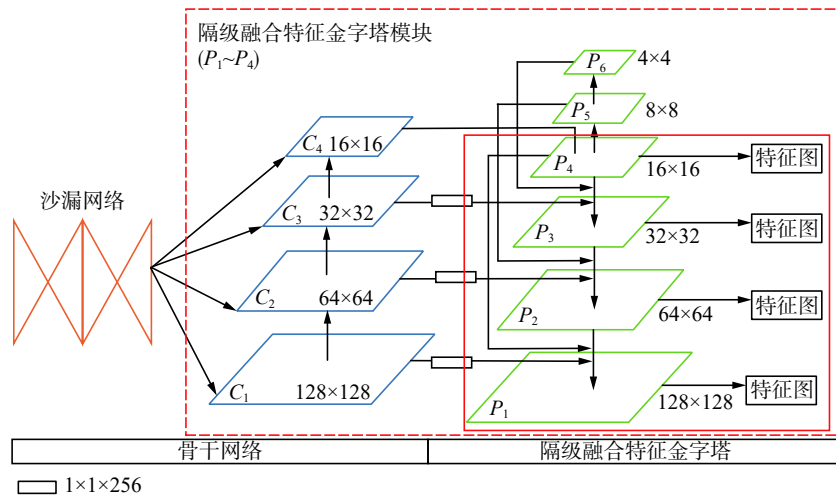


图 7 隔级融合特征金字塔 ( $P_1 \sim P_4$ )

Fig. 7 Feature pyramid of interval fusion ( $P_1 \sim P_4$ )

$P_5$ 、 $P_6$  是由  $P_4$  单纯进行卷积得到的特征图, 这 2 个特征图是为隔级融合做准备的; 但是在对这 2 个特征图进行检测时发现效果良好。表 1 为 2 种模型的实验结果。

由表 1 可以看出, 6 层特征金字塔的平均准确率比 4 层特征金字塔高 0.8%, 小目标检测准确率高 0.9%。数据结果表明,  $P_5$ 、 $P_6$  对目标准确率改善较为明显。

#### 2) 实验结果及对比

本文对当前主流的 FPN 进行了改进, 将隔级融合特征金字塔引入 CornerNet 算法中。进行实验的模型有 2 个, 一是对  $P_1 \sim P_4$ , 4 个特征图进行检测; 另一个是  $P_1 \sim P_6$ , 6 个特征图进行检测。表 2 为本文的实验结果和其他网络检测结果的对比。

本文骨干网络采用沙漏网络, 与残差网络系列相比, 沙漏网络能使待检测特征图的语义更加

丰富。通过对比结果可以看出, CornerNet 模型中引入隔级融合特征金字塔对准确度有较好地提高; FPN(4) 平均准确率为 43.5%, 较 CornerNet 算法提高 2.9%; FPN(4) 在小目标准确率上较 CornerNet 算法提高 2.9%。FPN(6) 平均准确率为

44.3%, 较 CornerNet 算法提高 3.7%。FPN(6) 在小目标准确率上较 CornerNet 算法提高 4.2%。FPN(6) 与一阶段系列代表 SSD513 对比, 平均准确率提高 13.1%, 与二阶段系列代表 Faster R-CNN w FPN 对比, 平均准确率提高 8.1%。

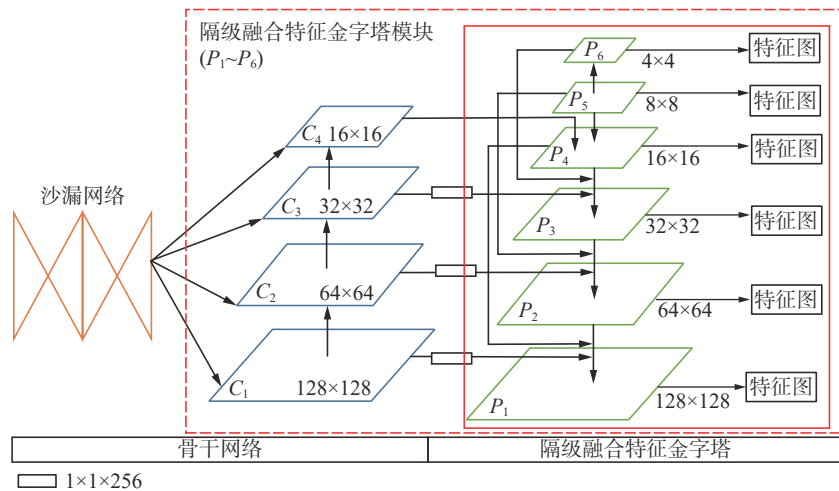


图 8 隔级融合特征金字塔 ( $P_1 \sim P_6$ )

Fig. 8 Feature pyramid of interval fusion ( $P_1 \sim P_6$ )

表 1 MS COCO 数据集不同 FPN 层数检测准确率  
Table 1 Detection accuracy of different FPN layers in MS COCO dataset %

方法	网络	AP <sub>0.5:0.95</sub>	APS
本文模型	Hourglass-104-FPN(4)	43.5	22.4
本文模型	Hourglass-104-FPN(6)	44.3	23.3

表 2 MS COCO 数据集不同算法检测准确率对比  
Table 2 Comparison of detection accuracy of different algorithms in MS COCO dataset %

方法	网络	AP <sub>0.5:0.95</sub>	APS
YOLO v2	DarkNet-19	21.6	5.0
SSD513	ResNet-101-SSD	31.2	10.2
DSSD513	ResNet-101-DSSD	33.2	13.0
Faster R-CNN+++	ResNet-101	34.9	15.6
Faster R-CNN w FPN	ResNet-101-FPN	36.2	18.2
Mask R-cnn	ResNeXt-101	39.8	22.1
RetinaNet	ResNet-101-FPN	39.1	21.8
CornerNet511	Hourglass-104	40.6	19.1
本文模型	Hourglass-104-FPN(4)	43.5	22.4
本文模型	Hourglass-104-FPN(6)	44.3	23.3

数据结果表明, 引入隔级融合特征金字塔对小目标准确率的提高有较大作用。相较于其他算法, 本算法模型的主要特点是待检测的特征图语义信息更加丰富。无论是 Faster R-CNN 系列算法还是 RetinaNet 算法, 尽管都采用了 FPN, 但是由于语义信息不够丰富, 导致算法在小目标检测准确率以及平均准确率方面并不是很出色。隔级融合特征金字塔运用在其他算法上可能也会提高检测准确率。

为验证本文算法检测小目标的有效性, 挑选了含有被云遮盖、边缘信息模糊、目标相对较小的图进行了测试对比。通过图 9、10 可以看出, 隔级融合特征金字塔与 CornerNet 相结合的算法模型对物体的识别置信度更高、定位更加准确, 并且能够识别出不易被人眼察觉的小目标, 特别是对背景复杂、遮挡、目标尺寸较小等因素造成的漏检, 改善效果较为显著<sup>[22]</sup>。



图 9 传统 CornerNet 识别结果

Fig. 9 Recognition results of traditional CornerNet





图 10 改进后模型识别结果

Fig. 10 Improved model recognition results

## 4 结束语

针对 CornerNet 模型对小目标检测平均准确率较低的问题,提出了一种隔级融合特征金字塔与 CornerNet 相结合的识别方法,骨干网络 Hourglass 中引入参数量更小的瓶颈残差模块以缩短网络的推理时间。与传统的特征金字塔旁路连接、上下融合不同,引入隔级融合特征金字塔,得到语义信息更丰富的多尺度特征图。经过实验对比,本文 4 层 FPN 模型与 6 层 FPN 模型相较于传统 CornerNet 算法具有较好的定位和识别准确度,尤其是 6 层 FPN 模型较传统 CornerNet 算法在小目标准确率方面提高 4.2%,平均准确率方面提高 3.7%,表明了本文所提出模型的有效性。

## 参考文献:

- [1] 赵永强, 饶元, 董世鹏, 等. 深度学习目标检测方法综述 [J]. 中国图象图形学报, 2020, 25(4): 629–654.  
ZHAO Yongqiang, RAO Yuan, DONG Shipeng, et al. Survey on deep learning object detection[J]. Journal of image and graphics, 2020, 25(4): 629–654.
- [2] 赵文清, 周震东, 翟永杰. 基于反卷积和特征融合的 SSD 小目标检测算法 [J]. 智能系统学报, 2020, 15(2): 310–316.  
ZHAO Wenqing, ZHOU Zhendong, ZHAI Yongjie. SSD small target detection algorithm based on deconvolution and feature fusion [J]. CAAI transactions on intelligent systems, 2020, 15(2): 310–316.
- [3] GIRSHICK R. Fast R-CNN[C]//Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago, Chile, 2015: 1440–1448.
- [4] HE K, ZHANG X, REN S, 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.
- [5] HE Kaiming, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[C]//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy, 2017: 2961–2969.
- [6] 段仲静, 李少波, 胡建军, 等. 深度学习目标检测方法 & 主流框架综述 [J]. 激光与光电子学进展, 2020, 57(12): 1–16.  
DUAN Zhongjing, LI Shaobo, HU Jianjun, et al. Review of deep learning target detection methods and mainstream frameworks [J]. Progress in laser and optoelectronics, 2020, 57(12): 1–16.
- [7] LIU Wei, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]//Proceedings of 14th European Conference on Computer Vision. Amsterdam, The Netherlands, 2016: 21–37.
- [8] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA, 2016: 779–788.
- [9] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy, 2017: 2980–2988.
- [10] 刘俊明, 孟卫华. 基于深度学习的单阶段目标检测算法研究综述 [J]. 航空兵器, 2020, 27(3): 44–53.  
LIU Junming, MENG Weihua. Review of single-stage target detection algorithm based on deep learning [J]. AERO weapons, 2020, 27(3): 44–53.
- [11] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA, 2017: 2117–2125.
- [12] 陈景明, 金杰, 王伟锋. 基于特征金字塔网络的改进算法 [J]. 激光与光电子学进展, 2019, 56(21): 165–170.  
CHEN Jingming, JIN Jie, WANG Weifeng. Improved algorithm based on characteristic pyramid network[J]. Laser and optoelectronics progress, 2019, 56(21): 165–170.
- [13] LAW H, DENG Jia. CornerNet: detecting objects as paired keypoints[C]//Proceedings of the 15th European Conference on Computer Vision (ECCV). Munich, Germany, 2018: 734–750.
- [14] NEWELL A, YANG Kaiyu, DENG Jia. Stacked hourglass networks for human pose estimation[C]//Proceedings of the 14th European Conference on Computer Vision. Amsterdam, The Netherlands, 2016: 483–499.
- [15] 和超, 张印辉, 何自芬. 多尺度特征融合工件目标语义分割 [J]. 中国图象图形学报, 2020, 25(3): 476–485.  
HE Chao, ZHANG Yinhuai, HE Zifen. Multi scale feature

- fusion for object semantic segmentation[J]. Journal of image and graphic, 2020, 25(3): 476–485.
- [16] CAO Guimei, XIE Xuemei, YANG Wenzhe, et al. Feature-fused SSD: fast detection for small objects[C]//Proceedings of SPIE 10615, Ninth International Conference on Graphic and Image Processing. Qingdao, China, 2018: 106151E.
- [17] 鞠默然, 罗江宁, 王仲博, 等. 一种融合注意力机制的多尺度目标检测算法 [J]. 光学学报, 2020, 40(13): 126–134.
- JU Moran, LUO Jiangning, WANG Zhongbo, et al. Multi-scale target detection algorithm based on attention mechanism[J]. Acta optica sinica, 2020, 40(13): 126–134.
- [18] 王慧玲, 蔡小龙, 武港山. 基于深度卷积神经网络的目标检测技术的研究进展 [J]. 计算机科学, 2018, 45(9): 11–19.
- WANG Huiling, QI Xiaolong, WU Gangshan. Research progress of target detection technology based on deep convolution neural network[J]. Computer science, 2018, 45(9): 11–19.
- [19] TIAN Zhi, SHEN Chunhua, CHEN Hao, et al. FCOS: fully convolutional one-stage object detection[C]//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea, 2019: 9627–9636.
- [20] 李晓光, 付陈平, 李晓莉, 等. 面向多尺度目标检测的改进 Faster R-CNN 算法 [J]. 计算机辅助设计与图形学学报, 2019, 31(7): 1095–1101.
- LI Xiaoguang, FU Chenping, LI Xiaoli, et al. Improved faster R-CNN for multi-scale object detection[J]. Journal of computer-aided design and computer graphics, 2019, 31(7): 1095–1101.
- [21] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: common objects in context[C]//Proceedings of the 13th European Conference on Computer Vision. Zurich, Switzerland, 2014: 740–755.
- [22] 李航, 朱明. 基于深度卷积神经网络的小目标检测算法 [J]. 计算机工程与科学, 2020, 42(4): 649–657.
- LI Hang, ZHU Ming. A small object detection algorithm based on deep convolutional neural network[J]. Computer engineering and science, 2020, 42(4): 649–657.

#### 作者简介:



赵文清, 教授, 主要研究方向为人工智能与图像处理, 主持或参与国家自然科学基金、河北省自然科学基金以及省部级项目 10 余项, 获河北省科技进步二等奖 1 项、河北省科技进步三等奖 1 项。发表学术论文 30 余篇, 出版学术专著 1 部。



孔子旭, 硕士研究生, 主要研究方向为深度学习和目标检测。



赵振兵, 副教授, 主要研究方向为深度学习与计算机视觉, 主持或参与国家自然科学基金、河北省自然科学基金、北京市自然科学基金以及省部级项目 10 余项, 获河北省科技进步一等奖 1 项。发表学术论文 20 余篇, 出版学术专著 3 部。