计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2023,32(1):127−134 [doi: 10.15888/j.cnki.csa.008885]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

# 面向目标检测的尺度增强特征金字塔网络①

张　轩[1,2], 王晓峰[1,2], 张文尉[1,2], 黄煜婷[1,2], 陈东方[1,2]

[1](武汉科技大学 计算机科学与技术学院, 武汉 430070)
[2](武汉科技大学 智能信息处理与实时工业系统湖北省重点实验室, 武汉 430070)
通信作者: 王晓峰, E-mail: wangxiaofeng@wust.edu.cn

**摘　要**: 基于特征金字塔网络的目标检测算法没有充分考虑不同目标间的尺度差异以及跨层特征融合过程中高频信息损失问题, 使网络无法充分融合全局多尺度信息, 导致检测效果不佳. 针对这些问题, 提出了尺度增强特征金字塔网络. 该方法对特征金字塔网络的侧向连接和跨层特征融合方式进行了改进, 设计具有动态感受野的多尺度卷积组作为侧向连接来充分提取每一个目标的特征信息, 引入基于注意力机制的高频信息增强模块来促进高层特征与底层特征融合. 基于 MS COCO 数据集的实验结果表明, 该方法能有效提高各尺度目标的检测精度, 整体性能优于现有方法.

**关键词**: 目标检测; 特征金字塔网络; 动态感受野; 多尺度融合; 注意力机制; 深度学习

引用格式: 张轩,王晓峰,张文尉,黄煜婷,陈东方.面向目标检测的尺度增强特征金字塔网络.计算机系统应用,2023,32(1):127–134. http://www.c-s-a.org.cn/1003-3254/8885.html

## Scale-enhanced Feature Pyramid Network for Object Detection

ZHANG Xuan[1,2], WANG Xiao-Feng[1,2], ZHANG Wen-Wei[1,2], HUANG Yu-Ting[1,2], CHEN Dong-Fang[1,2]

[1](School of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan 430070, China)
[2](Hubei Province Key Laboratory of Intelligent Information Processing and Real-time Industrial System, Wuhan University of Science and Technology, Wuhan 430070, China)

**Abstract**: The object detection algorithms based on the feature pyramid network do not give due consideration to the scale differences among different objects and the high-frequency information loss during cross-layer feature fusion, denying the network sufficient fusion of global multi-scale information and consequently resulting in poor detection effects. To solve these problems, this study proposes a scale-enhanced feature pyramid network. This method improves the lateral connection and cross-layer feature fusion modes of the feature pyramid network. Specifically, a multi-scale convolution group with the dynamic receptive field is designed to serve as a lateral connection so that the feature information of each object can be extracted sufficiently, and a high-frequency information enhancement module based on the attention mechanism is introduced to promote the fusion of high-layer features with low-layer ones. The experimental results on the MS COCO dataset show that the proposed method can effectively improve the detection accuracy on objects at each scale and its overall performance is better than that of the existing methods.

**Key words**: object detection; feature pyramid network; dynamic receptive field; multi-scale fusion; attention mechanism; deep learning

目标检测旨在从图像中定位并识别感兴趣的目标, 是计算机视觉领域的一项重要研究任务, 广泛应用于人脸识别、行人跟踪、无人驾驶等领域. 在当前的目标检测任务中, 不同的目标之间往往存在着较大尺度

① 基金项目: 国家自然科学基金 (U1803262)
收稿时间: 2022-05-12; 修改时间: 2022-06-15; 采用时间: 2022-06-27; csa 在线出版时间: 2022-08-26
CNKI 网络首发时间: 2022-11-15

差异, 而单一特征图难以同时有效表征多个尺度的目标, 这使如何有效检测这些多尺度目标成为难点. 针对这一问题, 目前主流方案是在网络中构建特征金字塔, 利用聚合有全局上下文信息的多个特征层表征不同尺度的目标, 如 FPN[1]、Nas-FPN[2]、Aug-FPN[3]、FPG[4] 等. 上述算法虽然取得了较好的检测效果, 但仍然存在一些不足: 1) 上述特征金字塔网络没有认识到不同特征层上的同一目标和同一特征层上的不同目标有着较大尺度差异, 其侧向连接只有有限的非动态感受野, 难以充分提取多尺度目标的特征信息; 2) FPN 和 Aug-FPN 对不同层的特征进行融合时, 仅基于线性插值方法对高层特征进行上采样, 这使高层特征的高频信息丢失, 引起图像边缘模糊和失真等问题.

特征信息的不充分提取和部分丢失问题使特征金字塔网络不能充分融合全局多尺度信息, 为此本文提出了尺度增强特征金字塔网络 (scale-enhanced feature pyramid network, SE-FPN), 其创新点有: 1) 针对特征信息的不充分提取问题, SE-FPN 改进了 FPN 的侧向连接, 其侧向连接基于动态多感受野卷积组设计, 通过接受目标尺度变化刺激并对应调整各感受野的权重, 充分提取特征层内多个尺度目标的特征信息; 2) 针对特征信息的部分丢失问题, SE-FPN 基于联合注意力机制对高层特征进行通道特征和空间特征强化后才进行特征的上采样操作, 以减少采样后高频信息的损失.

## 1 相关工作

### 1.1 目标检测

目前的基于深度学习的目标检测方法可分为两类: 单阶目标检测器和二阶目标检测器. 单阶目标检测器基于像素进行目标分类, 检测速度快而检测精度相对较低, 如 YOLO[5]、RetinaNet[6] 等. 二阶目标检测器则大多基于候选区域网络进行目标分类, 检测精度高而检测速度相对较慢, 如 Faster R-CNN[7]、Mask R-CNN[8]、Cascade R-CNN[9] 等. 然而, 在面对目标尺度变化较大的数据集时, 单阶目标检测算法和二阶目标检测算法的检测精度都不理想.

### 1.2 特征金字塔网络

为了提高单阶和二阶目标检测器的多尺度目标检测性能, FPN[1] 利用基于 1×1 卷积的侧向连接和基于线性插值方法的跨层特征融合分支构建融合有全局多尺度信息的多个特征层来表征不同尺度目标, 并在不同的特征层上对不同尺度目标进行检测. FPN 的成功使构建高效特征金字塔网络成为研究热门, 大量工作都在研究如何有效构建多路径特征金字塔. PANet[10] 在 FPN 中添加额外的自上向下特征融合路径, 进一步增强底层特征的语义信息. Bi-FPN[11]、MpFPN[12]、ZigZagNet[13] 都构建了多个自下向上和自上向下路径, 在每个特征层中反复融入全局上下文信息, 以增强每个特征层的语义信息和图像信息. Nas-FPN[2] 使用神经搜索架构自搜索最佳的特征融合路径. FPG[4] 则手工设计基于自下向上、自上向下和横向残差传播 3 种特征融合路径的多特征金字塔网络. Aug-FPN[3]、CE-FPN[14] 则结合注意力机制或亚像素卷积等方法来减少顶层特征在侧向传播中的损失和非线性地进行跨层特征融合. 上述方法专注构建复杂特征融合路径的特征金字塔, 而忽视了每个特征层都含有丰富多尺度信息, 而简单的侧向连接结构不能实现多尺度特征提取. 此外, FPN、Aug-FPN 等算法仅基于线性插值方法对高层特征进行上采样, 这会使高层特征的部分高频信息丢失.

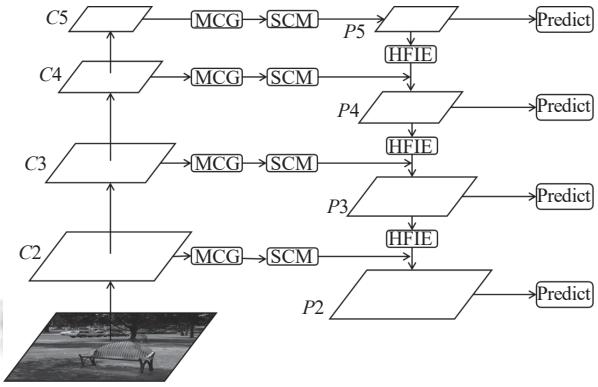



图 1 SE-FPN 的整体结构图

## 2 尺度增强特征金字塔网络

SE-FPN 的整体结构如图 1 所示, 由自下而上的骨干网络、自左向右的侧向连接和自上而下的跨层特征融合路径 3 部分组成. 首先, 以 ResNet[15] 作为骨干网络, 自下而上获取输入图像的不同深度和不同尺度的特征图, 记为{$C2$, $C3$, $C4$, $C5$}. 然后, 为了能有效提取各特征层内多尺度信息, 基于多尺度卷积组 (multi-scale convolutional group, MCG) 和尺度校准模块 (scale calibration module, SCM) 构建侧向连接. MCG 具有连续的大范围感受野, 能提取不同尺度的特征. SCM 基

于通道注意力机制来接受特征层和目标的尺度变化刺激, 对 MCG 不同尺度特征分配高或低的权重, 实现感受野同目标尺度的匹配, 充分提取多尺度目标的特征信息. 最后, 为了缓解跨层特征融合过程中高层特征的高频信息损失问题, 引入基于联合注意力机制的高频信息增强模块 (high frequency information enhancement, HFIE), 利用通道注意力机制强化边缘信息良好的通道、利用空间注意力机制突出前景特征. 在侧向传播和跨层特征融合过程中, 全局多尺度信息逐步融合并形成 4 个不同尺度的检测层, 记作{$P2$, $P3$, $P4$, $P5$}.

### 2.1 多尺度卷积组

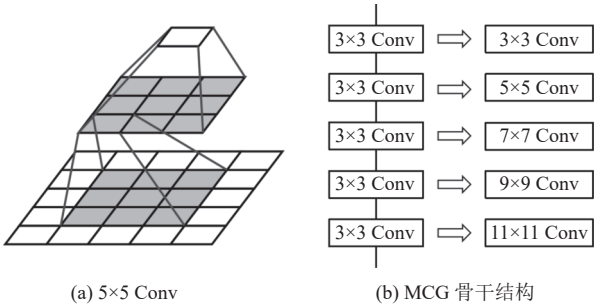
面对特征金字塔网络中复杂的多尺度目标分布问题, 为了能提取到更多的多尺度信息, MCG 需要有连续的大范围感受野来密集覆盖尽可能多的目标.

为了减少计算开销, 本文首先对大感受野卷积进行分解[16], 由若干个 3×3 卷积等效代替, 如图 2(a) 所示, 2 个 3×3 卷积堆叠即可代替 5×5 卷积. 其次, 在堆叠 3×3 卷积形成最大感受野的过程中, 共享中间卷积结果, 作为较小感受野的特征, 这使 $N$ 个 3×3 卷积即可表示 3×3 至 (2$N$+1)×(2$N$+1) 的 $N$ 个不同的感受野, 如图 2(b) 所示. 卷积的分解和共享使 MCG 在不引入大核卷积和膨胀卷积的条件下, 也能以较小的计算开销连续覆盖多个感受野, 整个 MCG 表述为:

$$f(2N+1)=\begin{cases} C_{1,1}(x), & N=0 \\ C_{1,3}(x), & N=1 \\ C_{N,3}(f(2N-1)), & N\geqslant 2 \end{cases} \tag{1}$$

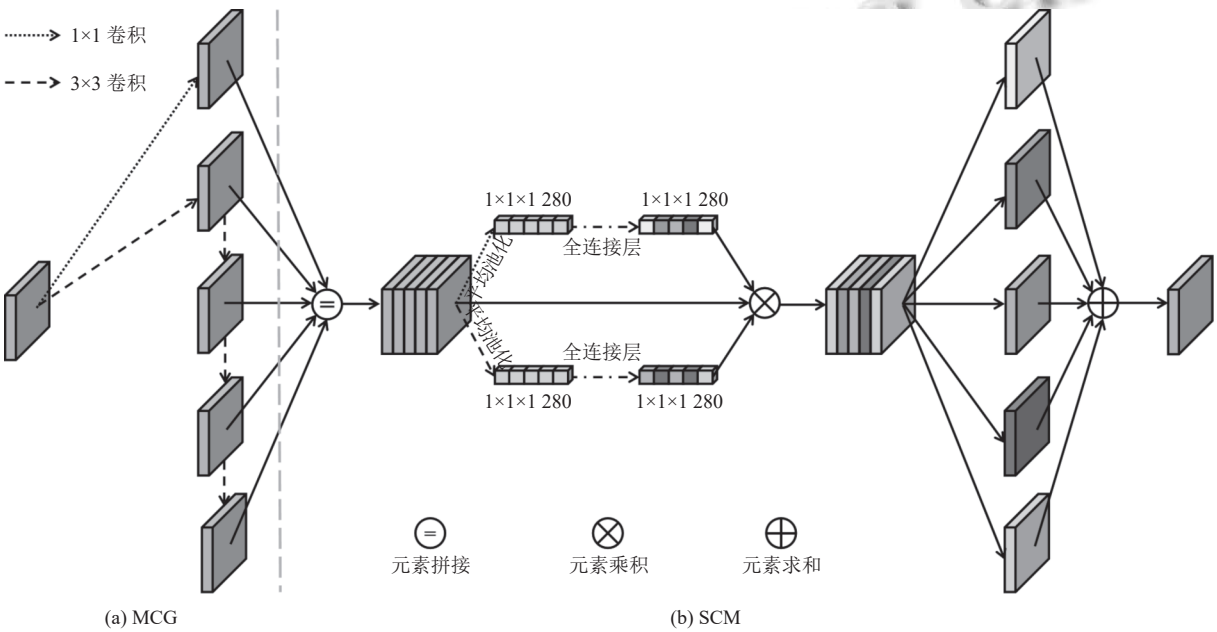
其中, $f$(2$N$+1) 表示感受野为 (2$N$+1)×(2$N$+1) 的特征图, $x$ 表示输入特征图, $C_{i,j}(x)$ 表示第 $i$ 个感受野为 $j$×$j$ 的卷积, $N$ 代表 MCG 中 3×3 卷积的数量, 改变 $N$ 值即可改变 MCG 的最大感受野大小.

MCG 还起到降低输入特征图的通道数, 减轻后续计算负担的作用, 所以 1×1 感受野的特征会丢失部分通道信息. 为了尽可能捕获多的特征信息, 提高 MCG 的特征表示能力, MCG 的首个 3×3 卷积会直接对输入特征图进行卷积操作, 而不在经过 1×1 卷积降维后的特征图上进行, 这使 MCG 不能形成完全的串行结构, SE-FPN 中 MCG 的实际结构如图 3(a) 所示.




(a) 5×5 Conv (b) MCG 骨干结构

图 2 卷积分解和 MCG 的骨干结构示意图




(a) MCG (b) SCM

图 3 SE-FPN 中自左向右的特征传播路径结构

### 2.2 尺度校准模块

SCM 基于通道注意力[17] 来自适应选择具有不同感受野的特征, 在高质量通道被增强, 低质量通道被抑制后, 兴趣特征图的通道权重总和将高于非兴趣特征图的通道权重总和, 实现 SE-FPN 对不同尺度目标的关注和自适应特征提取, SCM 的结构如图 3(b) 所示. SCM 由以下 4 个步骤实现.

1) 特征拼接: MCG 输出的不同感受野特征图在通道维上被拼接为具有全局多尺度特征信息的特征图 $f$.

2) 空间特征编码: 通道注意力机制需要对每个通道的空间特征进行压缩编码, 生成描述该通道的标识符. 为了能准确标识每一个通道, SCM 运用全局最大池化操作 ($gap$) 和全局平均池化操作 ($gmp$) 分别生成通道标识符 $m_c$ 和 $g_c$. 对于通道数、高度和宽度分别为 [$C$, $H$, $W$] 的输入特征 $f$, 每个通道的空间编码过程表述为:

$$m_c = Max(f_c(i,j)),\ 0<i<H\ \&\ 0<j<W \tag{2}$$

$$g_c = \frac{1}{H\times W}\sum_{i=0}^{H}\sum_{j=0}^{W} f_c(i,j) \tag{3}$$

3) 通道关系校准: 通道注意力机制通过两个全连接层交互通道标识符 $m_c$ 和 $g_c$, 学习和校准通道间的非线性相互依赖关系, 并为每个通道匹配一个权重. 该权重由 Sigmoid 函数激活后与对应通道进行乘法运算, 整个过程表述为:

$$f_c = f_c\times\sigma(W_1(\delta(W_2(m_c))) + W_1(\delta(W_2(g_c)))) \tag{4}$$

其中, $\sigma$ 表示 Sigmoid 激活函数, $W_1\in R^{C\times C/r}$, $W_2\in R^{C/r\times C}$ 表示全连接层, 本文实验中 $r$ 值设置为 16, $\delta$ 表示 LeakyReLU[18] 操作.

4) 特征分离和融合: 加权后的特征图将沿通道维度进行分割, 还原为特征拼接前的多个不同尺度的特征图, 以减少输出特征图的通道数. 分离后的特征进行加法融合, 实现多尺度特征增强.

### 2.3 高频信息增强模块

不同层特征之间存在尺度差异, 跨层融合过程中需要对高层特征进行两倍上采样, 常采用双线性插值方法. 然而, 双线线性插值方法具有低通滤波器效应, 会使生成图像的高频分量受到损失, 引起目标边缘模糊、图像失真等问题. 为了减少高频信息的损失, 如图 4 所示, 本文设计基于联合注意力机制[16] 的 HFIE, 利用通道注意力机制强化边缘信息良好的通道, 利用空间注意力机制强化前景特征, 抑制背景噪音, 表述为:

$$f_1 = f\times\sigma(W_1(\delta(W_2(gap(f)))) + W_1(\delta(W_2(gmp(f))))) \tag{5}$$

$$f_2 = f_1\times\sigma(Conv(gap(f_1)\oplus gmp(f_1))) \tag{6}$$

HFIE 的通道注意力结构同 SCM 中的一致, $f_1$ 表示通道增强后的特征, $f_2$ 表示空间增强后的特征, $Conv$ 表示 9×9 卷积, 用于捕获全局空间信息间的相互依赖关系, 空间权重由 Sigmoid 函数激活. 不同于通道注意力机制对空间特征取均值和最大值后的加法融合, 空间注意力机制利用特征拼接操作聚合通道特征的均值和最大值, 由符号“⊕”表示.

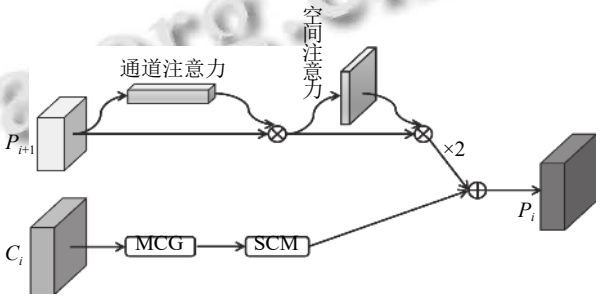



图 4 SE-FPN 中自上向下的跨层特征融合过程

## 3 实验验证

### 3.1 实验细节

数据集和评价指标: 各实验均基于 MS COCO 数据集[19] 进行, 该数据集分为有 11.8 万张图像的训练集, 有 5 千张图像的验证集和有 4 万张图像的测试集. 本文在训练集上训练神经网络, 在验证集上进行消融实验, 在测试集上报告最终实验结果, 并将该结果同其他模型比较. 评价指标为模型的参数量 (Params, 百万计); 模型的计算量 (FLOPs, 十亿计); 不同 IoU 阈值下的平均精度 (average precision, AP), 如 $AP_{50}$、$AP_{75}$; 不同尺度目标的平均精度, 如 $AP_S$、$AP_M$ 和 $AP_L$.

训练参数设置: 优化器为随机梯度下降优化器 (SGD), 初始学习率为 0.2, 训练轮数 (epoch) 为 12, 第 8 epoch 和第 11 epoch 时, 学习率递减至 0.02 和 0.002, 输入图像的分辨率为 800×1333, 批次数量 (batch-size) 为 2, MCG 的 $N$ 值设置为 4.

服务器配置: CPU 为 Xeon E5-2678 v3, 内存为 128 GB, 硬盘为 1 TB, 显卡为 8 张 Nvidia RTX 2080Ti, 操作系统为 Ubuntu 20.04, 神经网络基于 PyTorch 实现.

### 3.2 同其他特征金字塔网络的比较

为了证明 SE-FPN 的先进性, 本文基于 ResNet, 在单阶检测器 RetinaNet 和两阶检测器 Faster R-CNN 上

进行了实验, 并同 Aug-FPN、FPG、CE-FPN 等先进方法进行比较, 实验结果如表 1 所示.

以基于 Faster R-CNN 和 ResNet-50 的实验为例, SE-FPN 的平均检测精度为 39.1% , 超过 FPN 1.4%.

表 1 基于 MS COCO Test-Dev 数据集的各特征金字塔网络的参数量、计算量和精度的比较结果

| Method | Backbone | Params (M) | FLOPs (G) | AP (%) | $AP_{50}$ (%) | $AP_{75}$ (%) | $AP_S$ (%) | $AP_M$ (%) | $AP_L$ (%) |
|---|---|---|---|---|---|---|---|---|---|
| FPN* | RetinaNet-ResNet-50 | 37.4 | 250.3 | 36.9 | 56.2 | 39.3 | 20.5 | 39.9 | 46.3 |
| CE-FPN[14] | | — | — | 37.8 | 57.4 | 40.1 | 21.3 | 40.8 | 46.8 |
| Aug-FPN[3] | | — | — | 37.5 | **58.4** | 40.1 | 21.3 | 40.5 | 47.3 |
| SE-FPN | | 45.5 | 295.9 | **38.0** | 58.0 | **40.4** | **21.7** | **40.9** | **47.7** |
| FPN* | Faster-R-CNN-ResNet-50 | 41.5 | 216.7 | 37.7 | 58.7 | 40.8 | 21.7 | 40.6 | 46.7 |
| CE-FPN[13] | | 65.0 | 271.3 | 38.8 | 60.5 | 41.7 | 22.5 | 41.7 | 48.0 |
| Nas-FPN[2] | | 68.2 | 666.9 | 39.0 | 59.5 | 42.4 | 22.4 | **42.6** | 47.8 |
| Aug-FPN[3] | | — | — | 38.8 | 61.5 | 42.0 | **23.3** | 42.1 | 47.7 |
| SE-FPN | | 55.6 | 378.3 | **39.1** | **61.7** | **42.5** | 23.1 | 42.1 | **48.6** |
| FPN* | Faster-R-CNN-ResNet-101 | 60.5 | 312.5 | 39.7 | 60.7 | 43.2 | 22.5 | 42.9 | 49.9 |
| FPG[4] | | 98.8 | 716.9 | 40.6 | 62.3 | 44.3 | 23.4 | 43.5 | 51.7 |
| CE-FPN[14] | | — | — | 40.9 | 62.5 | 44.4 | 23.5 | 44.2 | 51.4 |
| Nas-FPN[2] | | 87.2 | 746.6 | 40.3 | 62.3 | 44.3 | 23.4 | 44.2 | 51.7 |
| Aug-FPN[4] | | — | — | 40.6 | **63.2** | 44.0 | 24.0 | 44.1 | 51.0 |
| SE-FPN | | 75.0 | 474.1 | **40.9** | 62.8 | **44.5** | **24.2** | **44.3** | **51.8** |

注: “*”表示该结果为本文复现结果, “—”表示论文中未给出相关数据

在小、中、大目标的检测上, 对比 FPN, SE-FPN 的平均检测精度分别提高了 1.4%, 1.5% 和 1.9%. 对比其他的先进特征金字塔网络模型, 如 Nas-FPN、Aug-FPN、CE-FPN, SE-FPN 平均检测精度更高. 综合对比不同条件下的多个实验可以认为, SE-FPN 的整体检测性能要优于当前主流的基于复杂多融合路径的特征金字塔网络模型, 能有效提高单阶检测器和二阶检测器的性能.

### 3.3 消融实验

为了证明本文提出各个方法的有效性, 本文基于 ResNet-50、Faster R-CNN 和 MS COCO 数据集的验证集进行消融实验.

表 2 是总体消融实验的结果, 本文以 FPN 作为基准, 逐步在网络中添加 MCG 和 SCM 和 HFIE, 3 个模块相互组合所带来的检测精度提升展示了这些方法的有效性.

表 2 整体消融实验结果 (%)

| 模型 | AP | $AP_S$ | $AP_M$ | $AP_L$ |
|---|---|---|---|---|
| FPN | 37.4 | 21.2 | 41.0 | 48.1 |
| FPN+HFIE | 37.8 | 21.5 | 41.8 | 48.7 |
| FPN+MCG | 38.4 | 22.5 | 42.3 | 49.1 |
| FPN+MCG+SCM | 38.7 | **23.0** | 42.7 | 49.5 |
| FPN+MCG+SCM+HFIE | **38.8** | 22.8 | **42.9** | **49.8** |

为了选取 MCG 的最佳 $N$ 值, 本文在不同 $N$ 值下进行了实验, 结果如表 3 所示. 随着 $N$ 值的增大, SE-FPN 的小尺度目标检测精度会逐渐降低, 中尺度和大尺度目标检测精度会逐渐升高, 整体的平均精度表现为先升后降. 当 $N$ 取值为 4 或 5 时, SE-FPN 的平均精度达到最高, 为 38.8%. 综合考虑参数量、计算量和平均检测精度, MCG 的 $N$ 值被设置为 4, 即 MCG 的感受野范围为 1×1、3×3、5×5、7×7、9×9.

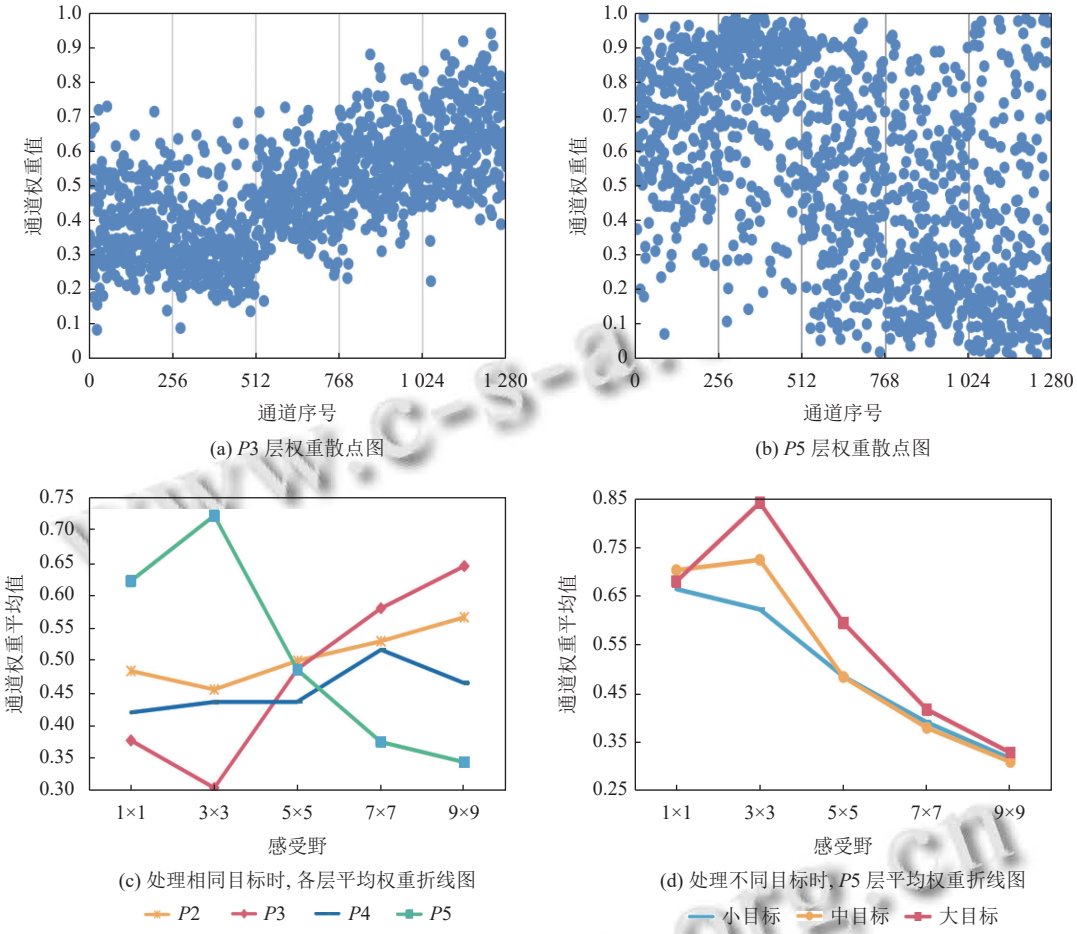
表 3 MCG 中 $N$ 值对 SE-FPN 检测精度的影响结果

| $N$ | Params (M) | FLOPs (G) | AP (%) | $AP_S$ (%) | $AP_M$ (%) | $AP_L$ (%) |
|---|---|---|---|---|---|---|
| 1 | **47.9** | **227.8** | 38.2 | **23.5** | 42.0 | 48.7 |
| 2 | 50.4 | 278.0 | 38.5 | 23.4 | 42.4 | 49.2 |
| 3 | 53.0 | 328.1 | 38.7 | 23.0 | 42.8 | 49.6 |
| 4 | 55.6 | 378.3 | **38.8** | 22.8 | **42.9** | 49.8 |
| 5 | 58.7 | 428.5 | **38.8** | 22.7 | 42.7 | 49.9 |
| 6 | 61.5 | 478.7 | 38.6 | 22.5 | 42.4 | **50.1** |

为了证明 SCM 能接受特征层和目标的尺度变化刺激并对 MCG 的感受野进行针对性调整, 本文将包含有不同尺度目标的图像送入 SE-FPN 中, 获取了不同特征层上 SCM 生成的通道权重 ($P2$ 层至 $P5$ 层的分辨率分别为 200×250、100×125、50×62、25×31, MCG 输出特征的通道数为 256).

图 5(a) 和图 5(b) 分别是 $P3$ 层和 $P5$ 层上 SCM 生成的全部通道权重的散点图 (横坐标表示不同特征的通道序号, 1–256、257–512、513–768、769–1024、1025–1280 分别属于 1×1 感受野特征、3×3 感受野特

征、5×5 感受野特征、7×7 感受野特征、9×9 感受野特征; 纵坐标为 Sigmoid 函数激活后的通道权重值). 对于大尺度特征层 *P*3, SCM 给予 7×7 和 9×9 这样的大感受野特征更多的高权重. 对于小尺度特征图 P5, SCM 更加关注于 1×1 和 3×3 这样小感受野. 整体上, 对比兴趣尺度特征图, 非兴趣尺度特征图的高权重通道数量更少.




(a) *P*3 层权重散点图 (b) *P*5 层权重散点图

(c) 处理相同目标时, 各层平均权重折线图 (d) 处理不同目标时, *P*5 层平均权重折线图

图 5 不同条件下的权重分布

为了更直观地观察不同特征图所获得的通道权重的整体性变化, 本文对每个特征图所获得的全部通道权重进行求和操作和归一化操作, 如图 5(c) 和图 5(d) (横坐标表示不同特征的感受野大小, 纵坐标表示该特征所获得的通道权重平均值).

图 5(c) 所示为不同特征层的 SCM 处理同一目标时的权重均值折线图, 结果表明随着特征层尺度的增大, SCM 的兴趣感受野会逐渐从小感受野向大感受野偏移, 权重的整体性变化显著.

图 5(d) 所示为 *P*5 层上 SCM 处理不同尺度目标时的权重均值折线图. 由于 *P*5 层尺度最小, SCM 的兴趣感受野为 1×1 和 3×3 感受野. 当处理一个小目标时, SCM 将最高权重分配给 1×1 感受野. 当处理大尺度目标时, SCM 将最高权重分配给 3×3 感受野, 并增加 5×5 感受野和 7×7 感受野的权重. 当处理中尺度目标时, SCM 集中关注 1×1 和 3×3 感受野. 上述分析表明 SCM 能对特征层内目标尺度变化做出针对性响应, 但此时权重的整体性变化相对较弱.

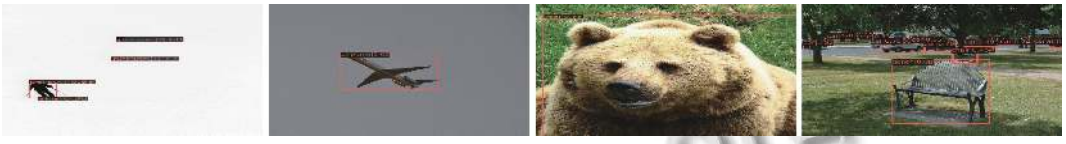
以上实验和分析表明, SCM 可以感知并正确地响应特征层和目标的尺度变化. 但是, 由于特征层之间的尺度差异一直存在, SCM 能持续接受特征层尺度变化刺激, 且层级跨越较大时, 特征层之间的尺度差异会非常显著, 所以 SCM 对特征层尺度变化的响应比对特征层中不同目标的尺度变化的响应更加显著.

### 3.4 训练和特征可视化

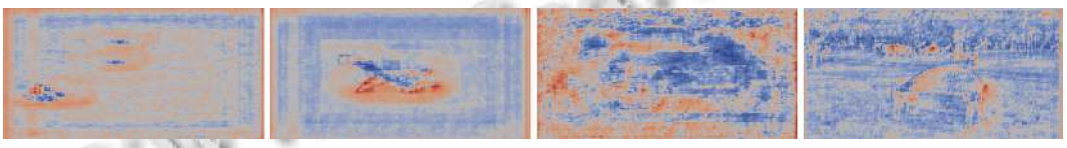
图 6 为 SE-FPN 的实际检测效果图和 SE-FPN 与

FPN 的可视化特征图, 本文选取了小目标图像、中等目标图像、大目标图像和复杂背景下多个尺度目标混合的图像作为对比. 图 6(a) 表明, 在实际的图像检测中, SE-FPN 能有效定位各尺度和复杂背景下的目标. 图 6(b) 和图 6(c) 表明, 对比 FPN, 输入图像经 SE-FPN 处理后, 其前景特征得到了显著增强, 背景噪音得到了有效抑制. 这表明充分融合全局多尺度特征信息, 使 SE-FPN 能有效区分图像的前景和背景, 这是 SE-FPN 的多尺度目标检测性能大幅提升的关键原因.

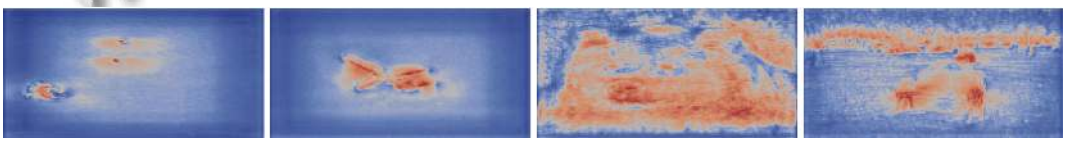
图 7 为 FPN 和 SE-FPN 在训练过程中的损失函数变化曲线图, 分别选取了训练过程中的边界回归损失函数、目标分类损失函数和总损失函数作对比, 实验结果表明, 两者在进行 6 万次左右的迭代训练后逐渐开始收敛, 而 SE-FPN 的 3 项损失值均要优于 FPN.



(a) SE-FPN 的实际检测效果图



(b) FPN 的 *P*2 特征层可视化图



(c) SE-FPN 的 *P*2 特征层可视化图

图 6 SE-FPN 实际检测效果图和 SE-FPN 与 FPN 的可视化特征对比

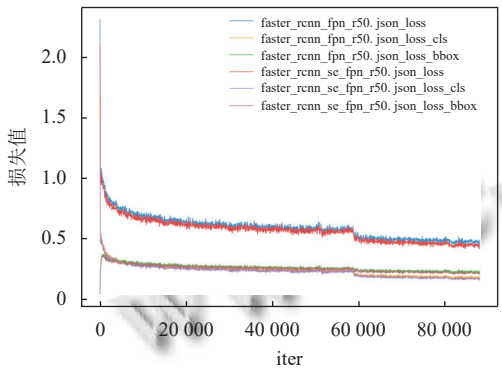



图 7 FPN 和 SE-FPN 的各损失函数变化曲线图

## 4 结束语

针对现有特征金字塔网络没有充分考虑不同目标间的尺度差异以及跨层特征融合过程中的高频信息损失问题, 使网络难以充分融合全局多尺度信息, 本文提出了面向目标检测的尺度增强特征金字塔网络 (SE-FPN). SE-FPN 基于 ResNet 构建自下向上的特征提取路径; 基于多尺度卷积组 (MCG) 和尺度校准模块 (SCM) 设计侧向连接, 自适应提取特征层内多尺度特征; 基于高频信息增强模块 (HFIE) 构建自上向下的跨层特征路径, 以减少跨层融合过程中高频信息的损失.

实验结果表明, 通过促进全局多尺度信息充分融合, SE-FPN 能显著关注前景特征, 抑制背景噪音, 大幅提高基础目标检测器对多尺度目标的检测性能, 其整体性能要优于现有的基于各类特征金字塔网络的目标检测算法.

## 参考文献

1 Lin TY, Dollár P, Girshick R, *et al*. Feature pyramid networks for object detection. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 936–944.

2 Ghiasi G, Lin TY, Le QV. NAS-FPN: Learning scalable feature pyramid architecture for object detection. Proceedings of 2019 IEEE/CVF Conference on Computer

Vision and Pattern Recognition. Long Beach: IEEE, 2019. 7029–7038.
3 Guo CX, Fan B, Zhang Q, *et al*. AugFPN: Improving multi-scale feature learning for object detection. Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020. 12592–12601.
4 Chen K, Cao YH, Loy CC, *et al*. Feature pyramid grids. arXiv:2004.03580, 2020.
5 Redmon J, Divvala S, Girshick R, *et al*. You only look once: Unified, real-time object detection. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 779–788.
6 Lin TY, Goyal P, Girshick R, *et al*. Focal loss for dense object detection. Proceedings of 2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017. 2999–3007.
7 Ren SQ, He KM, Girshick R, *et al*. Faster R-CNN: Towards real-time object detection with region proposal networks. Proceedings of the 28th International Conference on Neural Information Processing Systems. Montreal: MIT Press, 2015. 91–99.
8 He KM, Gkioxari G, Dollár P, *et al*. Mask R-CNN. Proceedings of 2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017. 2980–2988.
9 Cai ZW, Vasconcelos N. Cascade R-CNN: Delving into high quality object detection. Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 6154–6162.
10 Wang KX, Liew JH, Zou YT, *et al*. PANet: Few-shot image semantic segmentation with prototype alignment. Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2019. 9196–9205.
11 Tan MX, Le QV. EfficientNet: Rethinking model scaling for convolutional neural networks. Proceedings of the 36th International Conference on Machine Learning. Long Beach: PMLR, 2019. 6105–6114.
12 苏凯祺, 阎维青, 徐金东. 基于立体图像多路径特征金字塔网络 3D 目标检测. 北京航空航天大学学报, 2021: 1–9. [doi: 10.13700/j.bh.1001-5965.2021.0525]
13 Lin D, Shen DG, Shen ST, *et al*. ZigZagNet: Fusing top-down and bottom-up context for object segmentation. Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 7482–7491.
14 Luo YH, Cao X, Zhang JT, *et al*. CE-FPN: Enhancing channel information for object detection. Multimedia Tools and Applications, 2022: 1–20. [doi: 10.1007/s11042-022-11940-1]
15 He KM, Zhang XY, Ren SQ, *et al*. Deep residual learning for image recognition. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 770–778.
16 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556, 2014.
17 Woo S, Park J, Lee JY, *et al*. CBAM: Convolutional block attention module. Proceedings of the 15th European Conference on Computer Vision. Munich: Springer, 2018. 3–19.
18 Xu J, Li ZS, Du BW, *et al*. Reluplex made more practical: Leaky ReLU. Proceedings of 2020 IEEE Symposium on Computers and Communications. Rennes: IEEE, 2020. 1–7.
19 Lin TY, Maire M, Belongie S, *et al*. Microsoft COCO: Common objects in context. Proceedings of the 13th European Conference on Computer Vision. Zurich: Springer, 2014. 740–755.

(校对责编: 牛欣悦)