

基于改进 Mosaic 数据增强和特征融合的 Logo 检测

陈翠琴, 范亚臣, 王 林

(西安理工大学 自动化与信息工程学院, 西安 710048)

摘要: 近年来, Logo 检测在知识产权保护和产品品牌管理等领域得到了广泛应用; 针对 Logo 检测中的复杂背景和多尺度问题, 提出了一种改进 Mosaic 数据增强和特征融合的 Logo 检测算法; 将 6 张原始图片随机翻转、缩放和拼接构成合成图像, 与单张图像和由 4 张原始图片合成的图像一起作为 YOLOv4 模型的训练输入, 并确定 3 种输入形式的最优比例, 同时使用一种新的训练策略, 改进的 Mosaic 数据增强方法丰富了 Logo 对象的尺度和背景, 使模型更好地学习全局和局部特征; 在路径整合网络 (PANet) 的基础上引入跨层连接、重复堆叠、直接连接和加权特征融合等操作, 改进的 PANet 扩大了模型感受野, 增强了模型的多尺度特征表达能力; 实验结果表明, 提出的 MP-YOLOv4 算法在减小 21.7% 模型大小的同时, IoU (intersection of union) 等于 0.5 时的平均精度上达到了 67.4%, 较 YOLOv4 提高了 2.4%, 同时在多尺度目标上的检测性能得到了改善。

关键词: Logo 检测; YOLOv4; Mosaic 数据增强; 特征融合; 多尺度

Logo Detection Based on Improved Mosaic Data Enhancement and Feature Fusion

CHEN Cuiqin, FAN Yachen, WANG Lin

(School of Automation and Information Engineering, Xi'an University of Technology, Xi'an 710048, China)

Abstract: Logo detection has been widely used in intellectual property protection and product brand management in recent years. Aiming at the complex background and multi-scale problems in Logo detection, a Logo detection algorithm based on the improved Mosaic data enhancement and feature fusion is proposed. Six original images are randomly flipped, scaled and combined to form a composite image, which is used as the training input of the YOLOv4 model together with single image and composite image of four original images, and the optimal proportion of the three input is determined. Meanwhile, a new training strategy is used. The improved Mosaic data enhancement further enriches the scale and context of Logo objects, the model is used to learn the global and local features better. Based on the path integration network (PANet), some operations such as cross-layer connection, repeated stacking, direct connection and weighted feature fusion are introduced. The improved PANet enlarges the receptive field of the model and enhances the multi-scale feature expression ability of the model. Experimental results show that the proposed MP-YOLOv4 algorithm can reduce the model size by 21.7%, and the average precision reaches 67.4% when the Intersection of Union (IoU) equals 0.5, the average precision of the proposed MP-YOLOv4 algorithm is 2.4% higher than that of the YOLOv4 algorithm. At the same time, the detection performance of the multi-scale targets is improved.

Keywords: Logo detection; YOLOv4; Mosaic data enhancement; feature fusion; multi-scale

0 引言

标志 (Logo) 是将企业、公共机构、事业单位或个人的产品和服务标识出来的独特的图形文字或图形符号的组合。Logo 检测是目标检测的一个具体应用。它在知识产权保护、产品品牌识别、电商平台保护、智能交通车辆标识检测、社交媒体产品品牌管理等多个领域有着广泛的应用。自然图像中的 Logo 检测方法大致分为基于手工设计特征的方法和基于深度特征的方法。Sanyal 等人^[1]提出了一种基于 harris 仿射检测器获得的尺度不变特征变换 (SIFT, scale-invariant feature transform) 关键点的方法。Gao 等

人^[2]提出了一种通过空间光谱显著性来发现 Logo 的检测方法, 然后对查询图像中使用的这些区域提取加速鲁棒特征 (SURF)。然后根据提取的 SURF 特征发现数据集图像与查询图像之间的相似度。为了减少误报, 使用了局部空间上下文。Chinmoy 等人^[3]提出了一种基于 SIFT、SURF 和 HOG 描述符的 Logo 融合识别方法。自 2012 年以深度学习为主的图像分类以来, 深度卷积神经网络 (CNNs) 的使用在计算机视觉领域变得普遍。CNNs 在目标特征提取和表达方面比人工选择更合理、更强。S. C. H. Hoi 等人^[4]创建了大规模 Logo 图像数据集 LOGO-Net 以促进标志检测和产品

收稿日期: 2022-03-22; 修回日期: 2022-04-20。

基金项目: 陕西省科技计划重点项目 (2017ZDCXL-GY-05-03)。

作者简介: 陈翠琴 (1962-), 女, 山西运城人, 大学本科, 高级实验师, 主要从事深度学习、人工智能方向的研究。

王 林 (1962-), 男, 江苏东台人, 博士, 教授, 主要从事深度学习、目标检测和跟踪等方面的研究。

引用格式: 陈翠琴, 范亚臣, 王 林. 基于改进 Mosaic 数据增强和特征融合的 Logo 检测[J]. 计算机测量与控制, 2022, 30(10): 188-194, 201.

品牌识别的研究, 通过探索 RCNN^[5], Fast RCNN^[6] 和 SP-Net^[7] 等几种最先进的基于区域的深度卷积网络技术来解决深度 Logo 检测和品牌识别任务。Oliveira 等人^[8] 使用迁移学习来利用强大的卷积神经网络模型 Fast RCNN 来训练大规模的数据集, 并将它们重新用于图形 Logo 的检测。C. Eggert 等人^[9] 将 Faster R-CNN^[10] 应用到公司 Logo 检测任务中, 引入了一种改进的生成锚点建议的方案, 并提出了一种对 Faster R-CNN 的修改, 它利用了小物体的高分辨率特征地图, 而提高小目标检测的性能。Yang 等人^[11] 针对机动车 Logo 检测任务的 YOLOv3 模型^[12] 进行修改, 通过难样本训练解决小目标检测问题。

Logo 经常出现在较复杂的背景中, 同时 Logo 对象具有多尺度特性, 因此本文提出了 MP-YOLOv4 (improved mosaic and PANet YOLOv4) 算法。相比于原始的 YOLOv4 算法^[13], 本文主要做了以下的改进。为了进一步丰富 Logo 对象的尺度和背景, 提出了一种改进的 Mosaic 数据增强方法, 将 6 张原始图片进行随机缩放、裁剪并拼接构成新的训练数据, 与单张图片 and 4 张原始图片拼接一起作为模型的训练输入, 并确定 3 种输入形式的相对最优比例, 同时采用在训练结束的前 30 个世代关闭 Mosaic 数据增强的训练策略。为了进一步加强多尺度特征融合, 本文在路径整合网络 (PANet, path aggregation network)^[14] 的基础上, 结合跨层连接、重复堆叠、直接连接和加权特征融合等操作, 设计了一种新的特征金字塔网络, 增强了网络的特征融合和特征表达能力。

1 相关工作

1.1 YOLOv4 网络

YOLOv4 是 YOLO 系列中的一种新的目标检测方法, 它的网络结构如图 1 所示。

YOLOv4 目标检测网络主要由输入 (input)、骨干特征提取网络 (backbone)、特征融合部分 (neck) 和预测头 (head) 四部分组成。模型对于输入采用了 Mosaic 数据增强方法, 丰富了对象的上下文, 提高了训练效率。在骨干特征提取网络中, 在 YOLOv3 中 Darknet 的基础上, 融入了 CSPnet, 从而整合成新的骨干网络 CSPDarknet, 并在其中用 Mish 激活函数代替了 LeakyReLU 激活函数。在 Neck 部分, 使用空间金字塔 (SPP, spatial pyramid pooling) 和 PANet 对从骨干网络中获取的 3 个有效特征层来进行多尺度特征融合。在预测部分仍然使用 YOLOv3 中的 Head 对来自不同尺度的融合特征进行预测。本文主要 Logo 检测任务中的复杂背景和多尺度问题, 对 YOLOv4 算法中的 Input 部分和 PANet 部分进行改进。

1.2 Mosaic 数据增强方法

Mosaic 数据增强算法参考了 CutMix^[15] 数据增强算法, 是对 CutMix 数据增强算法的进一步扩展。一般的数据增强方法是对一幅图像进行翻转、色域变换、缩放等操作, 而 CutMix 的数据增强方法是将两幅图像进行拼接, 并将拼接后的图像直接传输到神经网络中进行训练。Mosaic 数据增强算法利用 4 幅图像进行拼接, 形成包含 4 幅原始图像的合成图像, 它可以在一幅合成图像中训练出多个不同的目标; 这使得对象出现在它们正常出现的背景之外, 为模型提供更加复杂和有效的训练背景; 同时, 在批处理归一化^[16] 操作中, 可以同时计算 6 张图像的数据, 这意味着不需要将超参数批大小 (batch_size) 设置得太大, 就可以有效地训练模型, 这样可以在单个 GPU 下训练目标检测算法, 提高了模型训练的效率, 节省计算开销。此外, 在原始的 YOLOv4 模型训练过程中, 模型有 0.5 的概率输入单张原始图像进行训练, 同时有 0.5 的概率输入由四张图像拼接而成的合

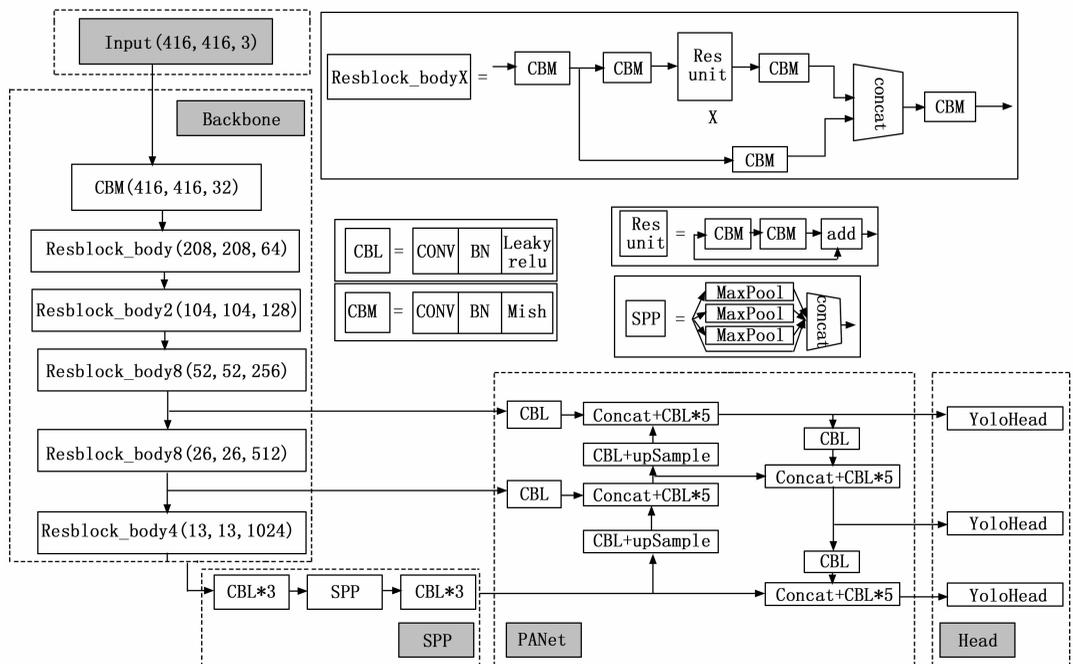


图 1 YOLOv4 的整体网络结构

成图像来进行训练。

1.3 多尺度特征融合

不同图像之间以及同一张图像内部多个 Logo 对象的相对尺度差别较大, Logo 检测任务中的多尺度检测问题面临着很大的挑战。为了在网络内部融合多尺度特征, 获得多尺度特征表达, Lin 等人^[17]提出了著名的特征金字塔网络(FPN, feature pyramid networks)。文献 [18] 使用 FPN 来融合高低层级的语义信息, 提升滤袋开口检测问题中对小目标的检测效果。Liu 等人提出了 PANet。PANet 在 FPN 的原自顶向下金字塔方法的基础上增加了一个自底向上的信息流, 重新构建了一个强化了空间定位信息的特征金字塔。在 YOLOv4 目标检测模型中就是使用 PANet 来进行特征融合。Pang 等人^[19]认为无论是 FPN 还是 PANet 在构建特征金字塔时都是使用自上而下或者自底而上的路径来传递特征, 这个过程会导致信息的丢失, 因此他们提出了平衡特征金字塔(BFP, balanced feature pyramid), 以同等重视多尺度特征图。它通过缩放、整合、精炼和增强 4 个步骤来获得相对平衡地兼顾所有特征层的特征图。

2 面向 Logo 检测的 YOLOv4 算法的改进

2.1 改进的 Mosaic 数据增强方法

原始的 Mosaic 数据增强方法对四张原始图像进行翻转、缩放、色域变换和拼接, 从而形成包含 4 幅原始图像的合成图像。为了进一步丰富 Logo 对象的背景, 提高模型在复杂背景下检测 Logo 的鲁棒性, 同时提高训练效率, 本文提出改进的 Mosaic 数据增强方法, 如图 2 所示。为了便于阐述, 将单张原始图像记为 mos1, 由 4 张、6 张图像混合而成的合成图像分别记为 mos4 和 mos6。上面和中间的实线箭头表示的通道是原始的 Mosaic 方法, 然而改进的 Mosaic 数据增强方法增加了下面虚线箭头表示的通道, 即使用 6 张图像拼接成的合成图像(mos6)作为模型的训练数据, 与 mos1 和 mos4 一起作为模型的训练输入。

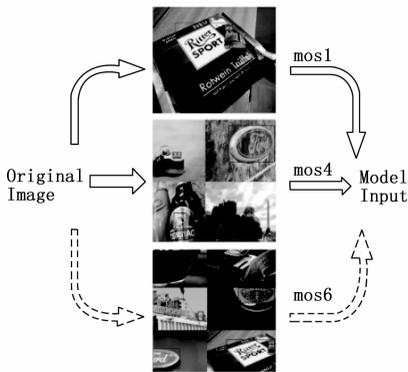


图 2 改进的 Mosaic 数据增强方法

2.1.1 由 6 张原始图像合成训练数据

1) 生成图像随机缩放的大小、随机粘贴的位置和裁剪的范围。

首先按公式 (1) 生成 6 张图像分别缩放的随机值:

$$mws = [int(w * rand(0.4, 1)), int(w * rand(0.4, 1)),$$

$$\begin{aligned} &int(w * rand(0.4, 1)), int(w * rand(0.4, 1)), \\ &int(w * rand(0.4, 1)), int(w * rand(0.4, 1))] \\ nhs = &[int(h * rand(0.4, 0.8)), int(h * rand(0.4, 0.8)), \\ &int(h * rand(0.4, 0.8)), int(h * rand(0.4, 0.8)), \\ &int(h * rand(0.4, 0.8)), int(h * rand(0.4, 0.8))] \end{aligned} \quad (1)$$

式中, $rand()$ 函数以均匀分布从一定范围内取出随机值, $int()$ 函数是对小数进行四舍五入的操作, (w, h) 表示网络要求输入的高和宽。 mws 和 nhs 中都包含了 6 个元素, 依次对应 6 张图像随机缩放之后的宽和高。

其次要生成分别沿 x 方向的一个偏移比例和 y 方向的一个偏移比例, 计算方法如公式 (2) 所示:

$$\begin{aligned} offset_x &= rand(1/4, 3/4) \\ offset_y1 &= rand(2/9, 4/9) \\ offset_y2 &= rand(5/9, 7/9) \end{aligned} \quad (2)$$

接着按照公式 (3) 生成 6 张图片的粘贴位置:

$$\begin{aligned} place_x &= [int(w * offset_x) - mws[0], \\ &int(w * offset_x) - mws[1], \\ &int(w * offset_x) - mws[2], int(w * offset_x), \\ &int(w * offset_x), int(w * offset_x)] \\ place_y &= [int(h * offset_y1) - nhs[0], \\ &int(h * offset_y2) - nhs[1], \\ &int(h * offset_y2), int(h * offset_y2), \\ &int(h * offset_y2) - nhs[4], \\ &int(h * offset_y1) - nhs[5]] \end{aligned} \quad (3)$$

式中, $place_x$ 和 $place_y$ 分别都包含了 6 个元素, 分别对应 6 张图片随机粘贴的横坐标和纵坐标。

最后, 利用偏移比例按照公式 (4) 计算出图像裁剪的界限:

$$\begin{aligned} cutx &= int(w * offset_x) \\ cuty1 &= int(h * offset_y1) \\ cuty2 &= int(h * offset_y2) \end{aligned} \quad (4)$$

式中, $cutx$ 是 x 方向的裁剪界限, $cuty1$ 和 $cuty2$ 是 y 方向的两条裁剪界限。利用 3 条界限分别裁剪出 6 张图像的相应部分并拼接在一起, 从而构成一张新的图像。

2) 按照索引从训练集中选取一条训练数据, 然后随机选取 5 条训练数据, 共获得 6 条数据。

3) 分别对每一张训练图像进行一系列基本的数据增强。以 0.5 的概率对图像进行左右翻转, 按照公式 (1) 对图像进行高和宽的随机缩放。

4) 6 张图像按照公式 (3) 各自粘贴到一张像素值为 (128, 128, 128), 大小为 [416, 416] 的图像上的指定位置。得到的 6 张图像如图 3 所示。

5) 按照公式依次裁取 6 张图像中的指定区域拼接成一张新的图像。再对合成图像色度 (hue)、饱和度 (sat) 和明度 (val) 进行扭曲, 扭曲系数分别是 $hue=0.1$, $sat=1.5$, $val=1.5$ 。生成的训练数据及标注情况如图 4 所示。

2.1.2 确定相对更优的输入比例

在网络训练过程中 mos1、mos4 和 mos6 这 3 种形式的

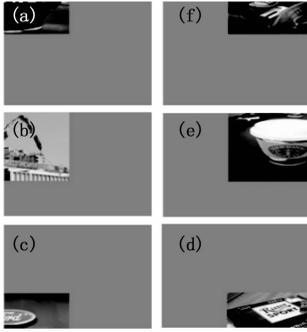


图 3 6 张图像的随机粘贴位置



图 4 合成图像及标注信息

输入所占的比例记为 α , 这种组合在一定程度上使训练数据集的尺度变化特征更加多样化, 从而进一步衰减背景对目标特征的干扰。在本文中通过枚举的方法来获得三者相对更优的输入比例。

2.1.3 改进的 Mosaic 训练策略

使用 Mosaic 合成的训练图片, 远远脱离了自然图片的真实分布, 即 Mosaic 合成的图片与自然图片存在较大的语义鸿沟。同时, Mosaic 数据增强过程中存在大量的裁剪操作会带来很多不准确的标注框。因此, 本文对 YOLOv4 中 Mosaic 训练策略进行了改进。在训练结束前的 30 个世代关闭 Mosaic 数据增强, 即只使用数据集中单张原始图像来训练模型。这样训练策略下数据集更专注于原始图片, 使得模型能很好地学习到目标的总体特征, 在此基础上, 用 Mosaic 数据增强合成的图像拥有更加复杂的背景, 用来增强模型对局部特征的学习能力, 从而提高整个模型的泛化能力。

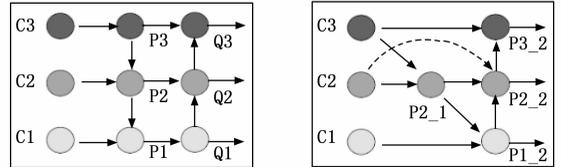
2.2 改进的多尺度特征融合

在 YOLOv4 中, 当输入图像大小为 416×416 时, 骨干网络从输入图像中提取出 3 个层次的有效特征层 C1, C2 和 C3。为了让整个模型学习到更加多样化的特征, 提高模型的检测性能, 目标检测网络的特征融合部分对不同层级的特征做进一步的增强和融合。本文对 YOLOv4 中的 PANet 部分进行改进。改进主要包括了跨层连接、重复堆叠、直接连接和加权特征融合 4 个方面。在本文所有的实验中, 当两个或多个特征进行融合的时候, 采用的是元素对位相

加 (element-wise add) 的方式, 而不是特征图堆叠 (concat) 的方式。

2.2.1 跨层连接

在图 5 (a) PANet 的基础上, 首先去除了未经过特征融合的即只有一个输入的节点, 即 P3 和 P1, 因为未经过特征融合的特征图对多尺度预测的贡献较小。其次在同一尺度的输入特征图和输出特征图之间增加一条新的连接, 以融合更丰富的特征, 称为跨层连接, 最终的网络结构如图 5 (b) 所示, 图中的短划线表示引入的跨层连接。



(a) PANet 的结构

(b) 跨层连接

图 5 PANet 和跨层连接的结构

以特征图 $P2_1$ 为例, 使用跨层连接之后, 特征图 $P2_1$ 的计算方法如公式 (5) 所示:

$$P2_1 = C2 + upSample(C3) \quad (5)$$

式中, $upSample()$ 是上采样函数, 实验中使用缩放因子为 2 的最近邻上采样。

特征图 $P2_2$ 的计算方法如公式 (6) 所示:

$$P2_2 = C2 + P2_1 + downSample(P1_2) \quad (6)$$

式中, $downSample()$ 是下采样函数, 在实验中使用步长为 2 的卷积进行下采样。

2.2.2 重复堆叠

为了进行充分的加强特征提取, 在 Neck 部分将设计的特征金字塔结构多次堆叠。如图 6 所示, 将设计的跨层连接结构重复堆叠了 3 次, 以获得更有效的特征融合和特征表达。

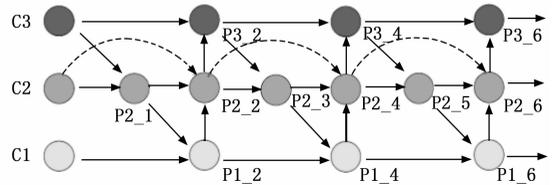


图 6 重复堆叠后的网络

2.2.3 直接连接

针对卷积过程中小对象容易丢失的问题, 在特征融合网络中设计了信息直接连接层。在跨尺度连接特征融合金字塔的每个迭代输出节点中, 我们使用信息直接连接层与初始输入特征图进行特征融合, 使小目标特征信息在特征提取过程中保持完整, 如图 7 所示。图中的 6 条点划线表示 6 处直接连接。

以特征图 $P2_4$ 为例, 使用信息直接连接之后, 特征图 $P2_4$ 的输出为:

$$P2_4 = C2 + P2_2 + P2_3 + downSample(P1_4) \quad (7)$$

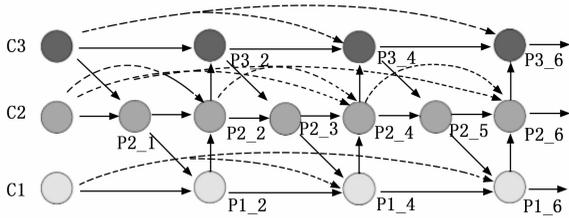


图 7 使用直接连接之后的网络结构

2.2.4 加权特征融合

当融合具有不同分辨率的特征时，常见的方法是先将它们的大小调整为相同的分辨率，然后直接再对其求和。在 PANet 中，网络均等地对待所有输入特征，而不同的输入特征具有不同的分辨率，它们对输出特征的贡献通常是不相等的。为了解决这个问题，为每个输入添加额外的权重，并让网络学习每个输入特征的重要性。采用快速归一化融合^[20]的方法来进行加权特征融合，它的计算方法如公式 (8) 所示：

$$O = \sum_i \frac{\omega_i}{\epsilon + \sum_j \omega_j} \cdot I_i \quad (8)$$

式中， O 是输出特征， I_i 是要第 i 个要融合的特征， ω_i 是第 i 个要融合特征的权重，在每个 ω_i 后应用 ReLU 激活函数来确保 ω_i 大于 0。以及 ϵ 设置为 0.000 1 是为了避免数值不稳定。网络在训练的过程中能够学习到对于各个输入特征的权重，从而实现有侧重地进行特征融合。

以特征图 $P2_4$ 为例，使用加权特征融合之后，特征图 $P2_4$ 的输出为：

$$P2_4 = \frac{\omega_1}{\epsilon + \sum_{j=1}^4 \omega_j} \cdot C2 + \frac{\omega_2}{\epsilon + \sum_{j=1}^4 \omega_j} \cdot P2_2 + \frac{\omega_3}{\epsilon + \sum_{j=1}^4 \omega_j} \cdot P2_3 + \frac{\omega_4}{\epsilon + \sum_{j=1}^4 \omega_j} \cdot \text{downSample}(P1_4) \quad (9)$$

2.3 改进的 YOLOv4 算法整体框架

最终，本文将提出的算法称为 MP-YOLOv4，MP-YOLOv4 算法的整体框架如图 8 所示。首先，使用改进的 Mosaic 数据增强算法为模型提供单张原始图像、4 张图片拼接成的合成图像和 6 张图片拼接成的合成图像共 3 种形式的训练输入数据；其次由 Backbone 从输入图像中提取特征并获得 3 个初步的有效特征层用于构建特征金字塔；然后使用改进的路径整合网络 PANet 来进行多尺度特征增强和融合；最后使用 YoloHead 得出预测结果。

3 实验验证和分析

3.1 实验设置

本文实验的硬件环境为：处理器为 Inter (R) Xeon (R) CPU E5-2640 v4 @2.4 GHz，显卡为 NVIDIA 1080Ti (11G)。网络训练阶段的软件开发环境为：操作系统是 64 位 CentOS Linux 7；驱动版本是 460.80；CUDA 版本是

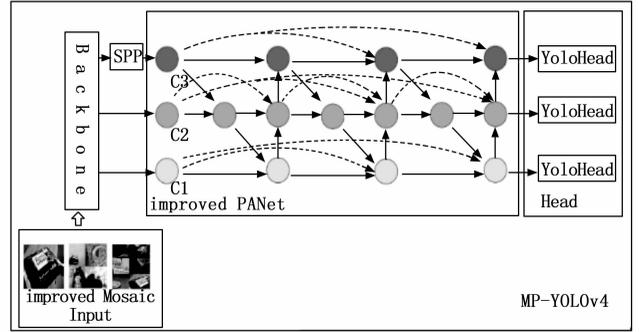


图 8 MP-YOLOv4 的整体框架

11.2；深度学习框架是 Pytorch1.4.0；编程语言是 Python3.7。关于目标检测模型训练过程中，网络参数的设置为：使用在 COCO 数据集上训练获得的 YOLOv4 整体网络的预训练权重；网络输入图像的高和宽是 416 和 416；总共训练 100 个世代 (epoch)；YOLOv4 网络训练分为两个阶段，冻结训练阶段网络的主干部分被冻结，特征提取网络的参数不发生改变，解冻训练阶段整个网络的参数都会发生改变；冻结训练阶段的 batch_size 设置为 8，解冻训练阶段的 batch_size 设置为 4；冻结训练阶段初始学习率为 10^{-3} ，解冻训练阶段的初始学习率设置为 10^{-4} ，学习率策略使用余弦退火学习率；使用多线程读取数据，num_workers 设置为 4；使用 Adam 优化算法。

3.2 实验数据集

FlickrLogos-32 数据集^[21]由从 Flickr 官网上收集的真实世界的图像组成，共包含了 32 种 Logo。整个数据集被分割为 3 个不相交的子集 P1、P2 和 P3。第 P1 是训练集，由每个类 10 张人工精心挑选的图像组成，这些图像中仅包含单个 Logo，且背景干扰较少。另外 P2 (验证集) 和 P3 (测试集) 每个类包含 30 张图像。与 P1 不同的是，这些图像包含了一个或多个 Logo 实例，且实例的背景更复杂。

3.3 评价指标

在本文中使用的 COCO 评价指标，包括 AP、AP50、AP75、APS、APM 和 APL。其中 AP 为 0.50 到 0.95 之间 10 个不同 IOU 设置下平均准确率的平均值；AP50 为 IOU 等于 0.5 时所有类别上的平均准确度；AP75 指标更加严格，表示 IOU 等于 0.75 时所有类别上的平均准确度；APS、APM 和 APL 分别描述模型在小目标、中目标和大目标上的准确度。使用模型大小 (model size) 来评估模型所占的内存空间。

3.4 实验过程

3.4.1 验证改进的 Mosaic 数据增强方法

本节将首先确定相对较好的比例设置，其次验证提前 30 个 epoch 终止训练的有效性。

在不同比例设置下的 YOLOv4 模型上的实验结果如表 1 所示。

实验 M1 (1 : 0 : 0) 中仅使用 mos1 图像训练，实验 M2 (1 : 1 : 0) 中均等地使用 mos1 和 mos4 图像训练，这是

表 1 不同比例设置下的模型性能

No.	$o : f : s$	AP/%	AP50/%	AP75/%	APS/%	APM/%	APL/%
M1	1 : 0 : 0	21.9	50.7	12.8	0.1	8.1	30.0
M2	1 : 1 : 0	34.5	65.0	32.5	0.1	14.9	45.4
M3	1 : 0 : 1	33.6	63.5	32.3	0.4	15.0	43.9
M4	1 : 1 : 1	35.8	66.3	33.2	0.2	17.2	46.1
M5	2 : 1 : 1	34.9	66.0	31.8	0.5	15.5	45.4
M6	1 : 2 : 1	35.1	66.0	32.9	1.4	16.0	45.5
M7	1 : 1 : 2	36.0	66.9	35.9	2.5	17.0	47.3

原始 YOLOv4 算法中的设置, 实验 M3 (1 : 0 : 1) 中均等地使用 mos1 和 mos6 图像训练, 对比这 3 个实验可以发现, M2 和 M3 都可以提升模型的性能, 且 M2 的提升要相对更明显; 值得注意的是, M3 虽在其他指标上不如 M2, 但是 M3 (使用了 mos6 数据) 提高了模型在小目标 (APS) 检测方面的性能。实验 M4 (1 : 1 : 1) 中均等地使用 mos1、mos4 和 mos6 数据进行训练, 比 M2 和 M3 的效果都要好。这证明了在原始 Mosaic 基础上, 添加 mos6 数据进行训练能够改善模型性能。

为了验证数据集应该更侧重于 mos1、mos4 和 mos6 中的哪一种, 开展了实验 M5 (2 : 1 : 1), M6 (1 : 2 : 1), M7 (1 : 1 : 2), 与实验 M4 (1 : 1 : 1) 分别进行对比, 可以得出结论, 当使用比例 1 : 1 : 2, 即当数据集更侧重本文提出的 mos6 数据时, 模型在牺牲很少 APM 的情况下, 在其他指标上的性能都获得了大幅度提升。因此, 选择 1 : 1 : 2 这个相对来说最优的比例作为模型训练过程中的输入设置, 即有 0.25 的概率输入 mos1 数据, 同样有 0.25 的概率输入 mos4 数据, 有 0.5 的概率输入本文提出的 mos6 数据。

为了验证提出的 Mosaic 训练策略的有效性, 即在训练结束的前 30 个 epoch 关闭 Mosaic 数据增强技术是否能够提升模型的性能, 在两种情况下分别进行了实验, 实验结果如表 2 所示。2 : 2 : 1 (w) 表示使用提出的训练策略, 2 : 2 : 1 (wo) 表示不使用。通过实验结果可以得出结论, 在损失很小 AP50 和 APM 的条件下, 其他指标获得了较大幅度的提升。因此, 本文提出的提前终止 Mosaic 数据增强的训练策略是有效的。

表 2 是否使用训练策略的对比实验结果

$o : f : s$	AP/%	AP50/%	AP75/%	APS/%	APM/%	APL/%
1 : 1 : 2 (wo)	36.0	66.9	35.9	2.5	17.0	47.3
1 : 1 : 2 (w)	36.8	66.7	37.0	2.8	16.9	47.7

表 3 特征融合实验

	①	②	③	④	AP/%	AP50/%	AP75/%	APS/%	APM/%	APL/%	模型大小/MB
BL	×	×	×	×	34.5	65.0	32.5	0.1	14.9	45.4	257.3
N1	√	×	×	×	33.9	64.7	32.1	0.05	14.0	44.8	188.3
N2	√	√	×	×	34.8	65.5	32.8	0.15	15.0	44.7	201.4
N3	√	√	√	×	35.0	66.0	32.7	0.15	15.3	45.0	201.4
N4	√	√	√	√	35.3	66.2	32.9	0.2	15.7	45.9	201.4

3.4.2 验证改进的 PANet

为了验证改进的 Neck 中跨层连接 (①)、反复堆叠 (②)、直接连接 (③) 和加权特征融合 (④) 4 个方面分别对于网络的有效性, 下面对改进的 Neck 进行了消融实验分析。实验结果如表所示。实验 BL 是基线实验, 未使用任何一个改进, 实验 N1 使用了跨层连接 (①), 实验 N2 在 N1 基础上多次堆叠了特征金字塔模块 (②), 实验 N3 进一步引入了跨层连接 (③), 实验 N4 在 N3 的基础上引入了加权特征融合 (④)。通过分析表 3 发现, 依次引入每一个改进都在一定程度上改善了模型性能。同时引入以上 4 个方面的改进, 在减少 21.7% 模型大小的同时, 在所有指标上的精度都获得了提升, 平均精度 (AP) 提高了 0.8%, IOU 等于 0.5 时的平均精度 (AP50) 提高了 1.2%。

3.4.3 总体实验

为了验证文中提出 MP-YOLOv4 目标检测算法的有效性, 在 FlickrLogos-32 数据集上对 YOLOv4 算法和 MP-YOLOv4 算法进行了训练和测试, 并与 YOLOv3、SSD^[22] 和 Faster R-CNN 等经典的目标检测算法进行对比实验。实验结果如表 4 所示。与 YOLOv4 相比, 本文提出的 MP-YOLOv4 方法在 IOU 等于 0.5 时的平均精度值 (AP50) 达到了 67.4, AP50 提高了 2.4%, 模型大小减小了 21.7%。

与 YOLOv3、以 VGG 为骨干的 Faster R-CNN 和以 Resnet50 为骨干的 Faster R-CNN 以及 SSD 相比, 本文提出的 MP-YOLOv4 算法在精度方面达到了最高, 同时在模型大小方面也达到了可观的水平, 因此在模型大小和模型精度之间获得了一个较好的平衡。

3.4.4 消融实验

为了分别验证文中提出的改进的 Mosaic 数据增强方法和改进的 PANet 的有效性, 开展了以下的消融实验。在基线实验 T1 中未采用本文提出的任何一个改进, 在实验 T2 中仅采用了改进的 Mosaic (IM, improved mosaic), 设置 3 种输入的比例为 1 : 1 : 2, 在实验 T3 中仅采用了改进的特征融合部分 (IP, improved PANet), 在实验 T4 中同时采用了本文提出的两个改进。实验结果如表 5 所示。通过分析表 5 可以得出结论, 引入两个改进中的任意一个都在一定程度上提高模型在所有指标上的性能, 同时引入这两个改进, 能在最大程度上提升模型性能。

通过表 4 可以得出结论, 本文提出的 MP-YOLOv4 算法在平均精度 (AP) 上达到了 36.7%, 较 YOLOv4 提高了 2.2 个百分点, IOU 等于 0.5 时的平均精度 (AP50) 达到了

表 4 YOLOv4 和 MP-YOLOv4 的对比实验结果

	骨干网络	输入尺寸	AP50 /%	模型大小 /MB
YOLOv3	Darknet53	[416,416]	63.0	248.1
Faster R-CNN	Resnet50	[600,600]	66.5	114.7
Faster R-CNN	VGG	[600,600]	67.0	549.3
SSD	VGG	[300,300]	58.4	105.2
YOLOv4	CSPDarknet53	[416,416]	65.0	257.3
MP-YOLOv4	CSPDarknet53	[416,416]	67.4	201.4

表 5 消融实验研究

	IM	IP	AP/%	AP50/%	AP75/%	APS/%	APM/%	APL/%
T1	×	×	34.5	65.0	32.5	0.1	14.9	45.4
T2	√	×	36.8	66.5	37.0	2.8	16.9	47.7
T3	×	√	35.3	66.2	32.9	0.2	15.7	45.9
T4	√	√	36.7	67.4	35.0	1.0	17.9	47.5

67.4%，较 YOLOv4 提高了 2.4 个百分点。同时在 APS、APM 和 APL 等指标上均有提高，说明 MP-YOLOv4 算法相比于 YOLOv4 在多尺度检测问题方面的性能得到了改善。

3.4.5 可视化研究

除了定量的实验结果之外，图 9 展示了 YOLOv4 算法和 MP-YOLOv4 算法对于 FlickrLogo-32 数据集中一些有挑战性样本的检测结果。其中图 (a)、(b)、(c) 为 YOLOv4 网络的检测结果，图 (d)、(e)、(f) 为 MP-YOLOv4 算法的检测效果对比。从图中可以得出结论，本文提出的 MP-YOLOv4 算法对于小尺寸目标、密集目标和光照条件差等条件下有着更好的检测性能，存在更少漏检和误检等情况。

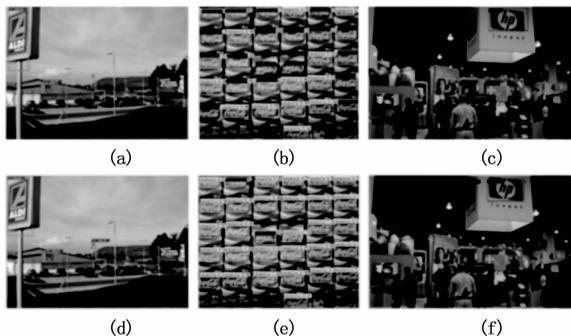


图 9 YOLOv4 和 MP-YOLOv4 的检测效果对比

4 结束语

针对 Logo 检测任务中出现的复杂背景干扰和多尺度目标等问题，本文提出了一种改进目标检测算法 MP-YOLOv4。改进了 YOLOv4 中的 Mosaic 数据增强算法，提出使用 6 张图片混合、四张图片混合和单张图片 3 种形式来共同作为模型的训练输入，并确定了 3 种输入的相对最优比例，同时采用了在训练结束前 30 个 epoch 关闭 Mosaic 数据增强方法训练策略，改进的 Mosaic 方法丰富了 Logo 对象出现的背景，使得模型更好地学习到 Logo 对象的全局特

征和局部特征，并优化了模型训练。结合跨层连接、重复堆叠、直接相连和加权特征融合等操作重新设计了网络的加强特征融合部分，增强了网络的多尺度特征表达能力。实验结果表明，相较于 YOLOv4 算法，本文提出的方法压缩了 21.7% 的模型大小，在平均精度上提高了 2.2 个百分点，在 IOU 等于 0.5 时的平均精度提高了 2.4 个百分点。同时，在小、中和大目标检测方面的性能都有提升。这说明，本文提出的 MP-YOLOv4 算法能更好地解决 Logo 检测任务中的复杂背景和多尺度问题。

参考文献:

[1] SANYAL S, SENGAMEDU S H. LogoSeeker: a system for detecting and matching logos in natural images [C] // Proceedings of the 15th International Conference on Multimedia 2007. Augsburg, Germany, 2007: 24-29.

[2] GAO K, LIN S, ZHANG Y, et al. Logo detection based on spatial-spectral saliency and partial spatial context [C] // International Conference on Multimedia and Expo., New York, NY, USA: IEEE, 2009: 322-329.

[3] BISWAS C, MUKHERJEE J. Logo recognition technique using sift descriptor, surf descriptor and hog descriptor [J]. International Journal of Computer Applications, 2014, 117 (22): 34-37.

[4] HOI S C H, WU X, LIU H, et al. LOGO-Net: large-scale deep logo detection and brand recognition with deep region-based convolutional networks [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, 46 (5): 2403-2412.

[5] GIRSHICK R, DONAHUE J, DARREL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C] // Conference on Computer Vision and Pattern Recognition, Columbus, USA: IEEE, 2014: 580-587.

[6] GIRSHICK R. Fast R-CNN [C] // International Conference on Computer Vision, Santiago, Chile: IEEE, 2015: 1440-1448.

[7] HE K, ZHANG X, REN S, et al. Spatial Pyramid pooling in deep convolutional networks for visual recognition [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 37 (9): 1904-16.

[8] OLIVERIRA G, FRAZO X, PIMENTEL A, et al. Automatic graphic logo detection via fast region-based convolutional networks [C] // International Conference on Document Analysis & Recognition, IEEE, 2007: 864-868.

[9] EGGERT C, ZECHA D, BREHM S, et al. Improving small object proposals for company logo detection [J]. ACM, 2017: 167-174.

[10] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39 (6): 1137-1149.

(下转第 201 页)