计算机工程与应用
Computer Engineering and Applications
ISSN 1002-8331,CN 11-2127/TP

《计算机工程与应用》网络首发论文

题目：结合 YOLO-FGE 网络的商标检测与分类
作者：缪春沅，王修晖
网络首发日期：2023-08-18
引用格式：缪春沅，王修晖．结合 YOLO-FGE 网络的商标检测与分类[J/OL]．计算机工程与应用. https://link.cnki.net/urlid/11.2127.TP.20230817.1234.002

网络首发时间：2023-08-18 11:00:16
网络首发地址：https://link.cnki.net/urlid/11.2127.TP.20230817.1234.002

结合 YOLO-FGE 网络的商标检测与分类

缪春沅， 王修晖

中国计量大学 信息工程学院浙江省电磁波信息技术与计量检测重点实验室，杭州 310018

摘要:为了解决商标样式众多、背景复杂、尺度变化大等问题，基于 YOLOv5 框架，提出了一种 YOLO-FGE 网络模型，以更精确地分辨出商标类别信息。首先，提出一种新的特征增强模块来提升特征层对不同类型商标的适应性，使网络更多关注待检测商标的有用信息。其次，在 YOLOv5 的 C3 模块中嵌入全局注意力模块对骨干网络和颈网络进行优化。最后，提出了一种增强空间注意力模块，利用空洞卷积扩大感受野，并结合通道注意力和 Transformer 模块来提升商标检测精度。在图形类商标数据集上的实验结果表明，该模型将 mAP 提升至 92.3%，比大多数现有方法具有更高的检测精度。

关键词：商标检测；特征增强；全局注意力；空间注意力

文献标志码：A 中图分类号：TP391.4 doi：10.3778/j.issn.1002-8331.2305-0513

Trademark Detection and Classification Based on YOLO-FGE

MIAO Chunyuan, WANG Xiuhui

Key Laboratory of Electromagnetic Wave Information Technology and Metrology of Zhejiang Province, College of Information Engineering, China Jiliang University, Hangzhou 310018

Abstract: In order to solve the trademarks' problems about their numerous styles, complex backgrounds, and large-scale changes, a YOLO-FGE network model based on the YOLOv5 framework was proposed to distinguish trademark category information more accurately. First, a feature enhancement module was put forward to enhance the adaptability of the feature layer to different kinds of trademarks, making the network pay more attention to the useful information of trademarks to be detected. Second, the global information attention module was embedded in the C3 module of YOLOv5 to optimize the backbone and neck network. Finally, an enhanced spatial attention module was raised, which used dilated convolution to expand the receptive field, combining channel attention and Transformer module to improve the detection accuracy. The experimental results on the graphic trademark dataset show that the model improves mAP to 92.3%, which has higher detection accuracy than most existing methods.

Key words: trademark detection; feature enhancement; global attention; spatial attention

基金项目：国家重点研发计划课题(2021YFC3340402)。

作者简介：缪春沅(1999－),女,硕士研究生,研究方向为计算机视觉,E-mail: P21030854032@cjlu.edu.cn;王修晖(1978－),男，博士,教授，研究方向为模式识别、计算机视觉、计算机图形学,E-mail: wangxiuhui@cjlu.edu.cn。

商标作为一种特殊的标志，在现代社会中起着越来越大的作用。近年来，随着知识经济的快速发展，商标申请量不断上升，截止 2022 年底，国内注册商标总量高达 4600 多万。面对数量如此巨大的待保护商标，如何对商标进行自动检测和分类成为亟待解决的热点问题。在 20 世纪 90 年代，基于计算机视觉的商标检测与分类技术首次被提出，其在本质上仍属于目标检测技术。在很长一段时间，科学家们将各种人工特征与传统的机器学习[1, 2]方法相结合，来提取商标中的信息，以达到商标检测与分类的目的。然而，自 2012 年起，随着深度学习的广泛应用与计算设备的大力发展，计算机视觉[3]这一领域在性能上得到了前所未有的提升。

商标检测与分类技术就是在图像中找出商标的位置及其对应的商标类别，本质上属于目标检测[4]的子领域。目标检测技术是计算机视觉中的任务之一，为计算机视觉中的目标分类[5, 6]、语义分割[7, 8]及实例分割[9, 10]等问题奠定了重要的基础。由于商标场景多样、商标数量庞大、商标设计多元化等一系列问题，商标检测与分类技术仍存在一定的挑战。因此，我们将从目标检测的角度来探讨商标检测的发展历程，先介绍传统的商标检测技术发展状况，再介绍基于卷积神经网络[11]的商标检测技术发展状况。

传统的基于人工特征算子的商标检测本质上是对商标的纹理、形状、颜色和特征点等共性特征来进行提取，然后通过一系列机器学习算法进行分类，以达到检测的目的。2003 年，Den 等人[12]通过模板匹配来完成视频截图中的商标检测任务，特别研究了从视频中提取的商标字符串与原始商标中的字符串相匹配的问题，但是该方法的误报率较高。2007 年，Ballan 等人[13]提出了一种基于 SIFT[14]特征点的商标检测系统，此方法具有较好的鲁棒性，能够有效地检测和分类商标，但是检测时间却大大增加。2008 年，Xie 等人[15]引入了一种新的数据挖掘方案，即空间金字塔挖掘，通过图像金字塔提取商标的局部特征，该方法可以检测自然环境下的各种商标类型，但是受图像分辨率的影响较大，因此检测的准确率不是很高。目前，传统的基于人工特征算子的商标检测技术虽在精度上有所提高，但在检测时间上却不能满足需求。

自 2012 年起，卷积神经网络开始兴起，基于卷积神经网络的商标检测技术成为大量研究者们关注的重点。2015 年，Iandolas 等人将深度卷积神经网络用于商标识别，当时在商标识别数据集上的准确性得到大大提升。2018 年，Sharma 等人[16]利用 YOLOv2 和 Faster R-CNN 模型对扫描文档中的商标进行检测，为未来的研究奠定了基础。2021 年，Sahel 等人[17]采用 RCNN、FRCNN 和 RetinaNet 模型用作商标检测，证明了随着新的 CNN 模型的出现，将其用作商标检测会产生不同的效果。同年，Yousaf 等人 [18]提出了一种针对小型商标的无分割框架，使用小型商标进行训练来解决错误分类的问题。2022 年，Hayfa 等人[19]利用两个预训练的卷积神经网络（VGGNet[20]和 ResNet）来提取商标信息，提出一种基于形状相似性的商标检测系统。同年，Trappey 等人[21]研究了两种深度学习的模型，一种模型通过 YOLOv4 来检测和定位商标，另一种模型通过三重卷积神经网络来进行商标相似度的分析，取得了较好的成果。Wang 等[22]人将 Focal_Loss 与 CIoU_Loss 合并应用到 YOLOv3 框架中，并在 LogoDet-3K 的数据集上进行验证，实验证明该模型具有较好的泛化能力。

本文的主要工作是基于 YOLOv5 目标检测网络，将图形类商标分为：圆形、矩形、三角形、扇形与五角星形，提出了一种改进的 YOLOv5 商标检测与分类算法，开展了对特征增强、注意力机制等处理方法的研究，以助于后续在现有商标库中更精确地检索类似图形商标。我们的主要研究内容包括以下几个方面：

（1）提出了一种新的特征增强模块（Feature Enhancement Module，FEM）。通过多次堆叠不同卷积核的卷积来加深网络的深度，使得网络能够融合不同复杂程度的商标特征，关注到更有用的信息，进而提高网络对商标的学习能力。

（2）提出了一种新的融合空间与通道信息的全局注意力机制（Global Information Attention Module，GIAM）。通过对坐标注意力机制的改进，使得网络既可以捕获方向感知与精确的位置信息，又可以捕获跨通道信息，有助于模型更准确地定位商标区域。我们将该注意力机制嵌入至 C3 模块中，通过实验验证 C3-GIAM 模块嵌入到算法网络的骨干网络、颈网络在数据集检测上的性能。

（3）提出了一种融合空洞卷积的增强空间注意力模块（Efficient Spatial Attention Module，ESAM）。通过不同空洞率的空洞卷积来扩大感受野，提出一种增强空间注意力模块，并与通道注意力模块、Transformer 模块混合使用，改善商标检测效果。

论文的剩余部分内容如下：第 1 节介绍了本文提出的 YOLO-FGE 网络结构的各个组成部分和相关改进方法，第 2 节展示实验结果和分析，第 3 节归纳总结。

1 本文方法

YOLOv5 算法于 2020 年 6 月提出，是目前主流的单阶段目标检测框架之一，其具有检测精度高、速度快等优点。YOLOv5 网络由输入端、骨干网络、特征融合网络和输出端组成。其中，在数据的输入部分采用了 Mosaic 数据增强，将四张图片通过随机裁剪、随机排布等操作拼接成一张图片，大大丰富了检测数

据集，增强了鲁棒性，使得模型可以在更小的范围内识别目标。同时，采用了自适应锚框计算和自适应图片缩放，减少了计算成本，提升了检测速度。在骨干网络部分，将原本的 Focus 模块替换成卷积核大小为 6×6 的卷积层，采用了 C3 模块进行特征提取，在保证检测精度没有下降的同时，减少计算量，提升推理速度。同时采用了 SPPF 模块，融合不同感受野的特征图，丰富特征图的表达能力。在 Neck 部分，采用了 FPN（Feature Pyramid Network，FPN）+PAN（Path Aggregation Network，PAN）的结构，实现浅层与深层的信息共享，增强网络的特征融合能力。在输出端，包含三个检测层，分别对应三种不同尺寸的特征图。将 CIoU_Loss 作为边界框回归损失函数，并通过非极大值抑制消除多余预测框。

本文以 YOLOv5 算法为基本网络，针对特征增强、注意力机制等进行构建与改进，提升算法对商标特征的提取能力，进而提升检测的精确度，降低检测的误报率。改进后的 YOLO-FGE 结构如图 1 所示。

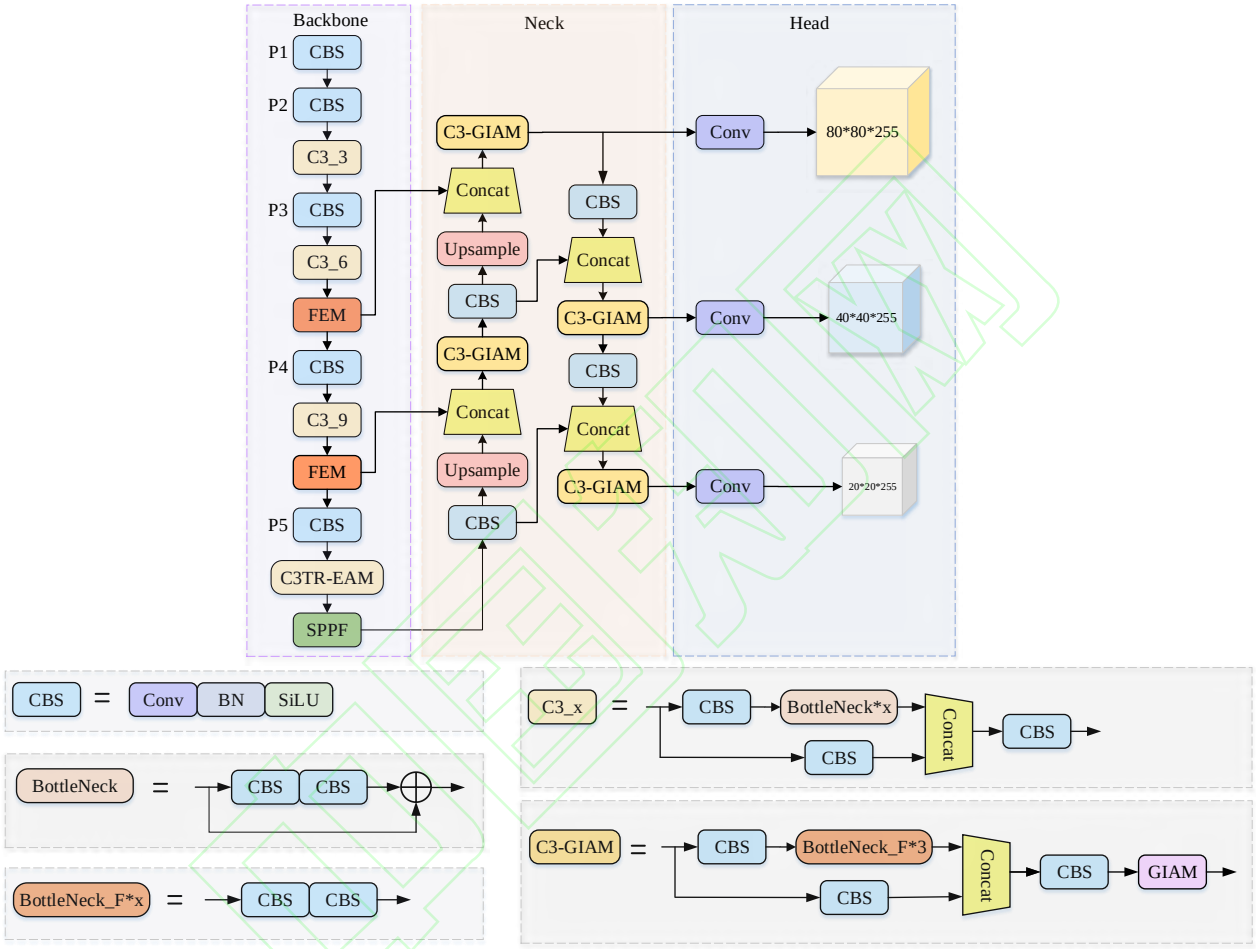


图 1　YOLO-FGE 结构图

Fig.1　Structure diagram of YOLO-FGE

1.1 特征增强模块

在卷积神经网络中，通过多次堆叠不同卷积核大小的卷积层可以提升网络的深度，获得不同感受野下的特征信息，使得网络能够更好地适应不同大小的特征图。因此，本文借鉴 Inception[23]结构与 Inception-ResNet[24]结构，提出了特征增强模块，结构如图 2 所示。

图 2 中，特征增强模块结构分为两个子模块：多层卷积融合子模块和残差连接子模块。多层卷积融合子模块由 4 条支路组成，分别对同层次下的不同特征进行提取并将这些特征进行通道级联，从而获得不同感受野的特征，增强模型的特征表达能力。值得一提的是，我们在每条支路都会首先使用卷积核大小为 1×1 的卷积来进行降维，这样可以有效减少参数量。而使用卷积核大小为 3×3、5×5、7×7 的卷积，可以在加深网络深度的同时使得网络获得多尺度的特征图，并增强对特征图的适应性。在对 4 条支路进行 Concat 操作后，我们加入了 CBAM[25]（convolutional block attention module，CBAM）注意力机制，自适应地从通道和空间两个方面进行调节，以此提升模型的鲁棒性。残差连接子模块将输入经过 SiLu 激活函数后与多层卷积融合部分的输出进行 add 操作，避免梯度消

失的发生。所有的卷积层后都会经过 SiLu 激活函数，以增强模型的泛化能力。

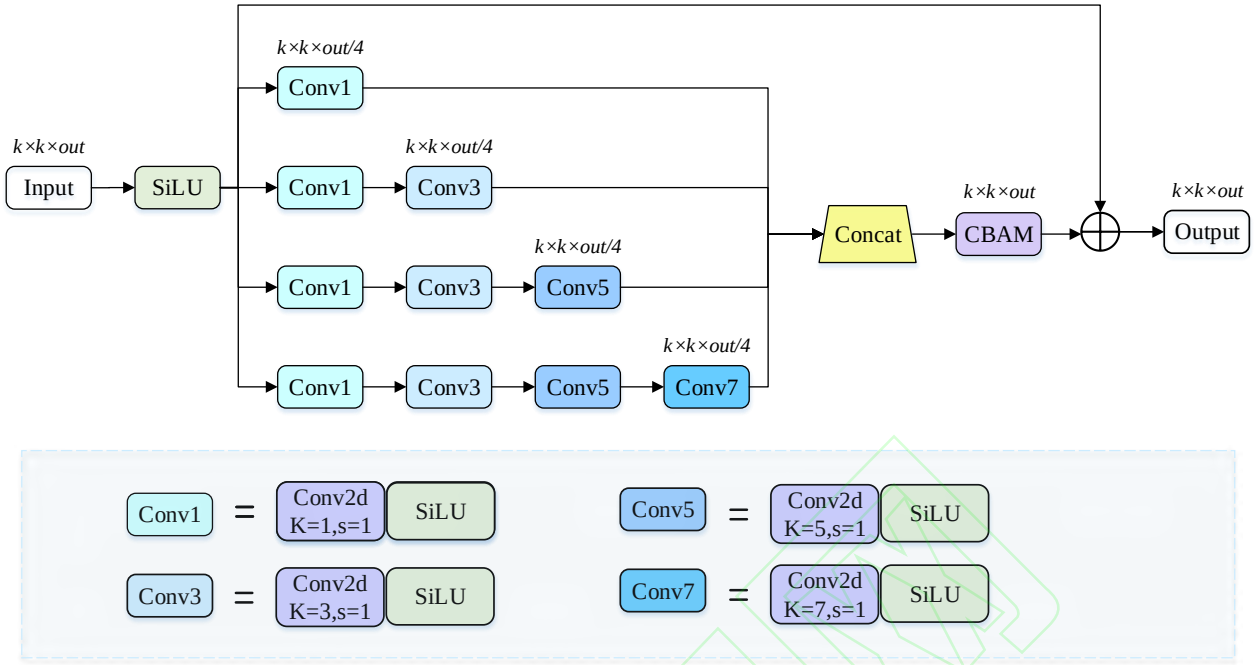


图 2 特征增强模块结构图

Fig.2 Structure diagram of FEM

1.2 全局注意力机制

在神经网络学习中，通过引入注意力机制[26,27,28]，使模型聚焦于对当前任务更为关键的信息，而忽略其他无用信息，甚至过滤掉无关信息，以此提高任务处理的效率。因此，本文构建了全局注意力模块，如图 3 所示。其中，全局注意力机制可以看成两部分：特征提取部分和残差连接部分。特征提取部分含 3 条支路，分别从 X、Y、Z 三个方向进行空间与通道信息的提取，其中 X、Y 负责提取空间信息，Z 负责提取通道信息。

对于输入特征图 $X \in R^{C\times H\times W}$，假设其维度为 $C\times H\times W$，首先在 X、Y 方向上进行两个池化大小为 $1\times 1\times W$、$1\times H\times 1$ 的自适应平均池化操作，沿着两个空间方向聚合特征，得到一对方向感知的特征图，使得该注意力模块可以很好地捕捉长程依赖关系与精确的位置关系，有助于更准确地定位商标区域；Z 方向上先进行池化大小为 $1\times H\times W$ 的自适应平均池化操作，再依次经过卷积层与激活层。计算公式如下所示：

$$F_C^X = AdaptiveAvgPool^{1\times 1\times W}(X) \tag{1}$$

$$F_C^Y = AdaptiveAvgPool^{1\times H\times 1}(X) \tag{2}$$

$$F_C^Z = ConvUnit\left(AdaptiveAvgPool^{1\times H\times W}(X)\right) \tag{3}$$

其中，$F_C^X \in R^{C\times H\times 1}$、$F_C^Y \in R^{C\times 1\times W}$、$F_C^Z \in R^{C\times 1\times 1}$ 分别代表 X、Y、Z 方向的输出，$AdaptiveAvgPool$ 代表自适应平均池化操作，$ConvUnit$ 代表一个卷积单元，包含卷积层与 Sigmoid 激活层。

接着，将 X、Y、Z 方向上的特征进行特征融合，具体操作为将 Z 方向获取的信息权重分别与 X、Y 方向上的信息相乘后进行 Concat 操作，以此促进不同方向的信息交流，并保证三个输出方向有一致的量纲。计算公式如下所示：

$$f_X = F_C^X(X)\times F_C^Z(X) \tag{4}$$

$$f_Y = F_C^Y(X)\times F_C^Z(X) \tag{5}$$

$$f = ConvUnit\left(Concat\left(f_X{}^T, f_Y\right)\right) \tag{6}$$

其中，$f_X \in R^{C\times H\times 1}$、$f_Y \in R^{C\times 1\times W}$ 代表 X 方向与 Z 方向、Y 方向与 Z 方向通道乘法之后结果，$f \in R^{C/r\times 1\times (H+W)}$ 代表 f_X 与 f_Y 结合后的特征图，$ConvUnit$ 代表一个卷积单元，包含卷积层与 h-swish 激活层。

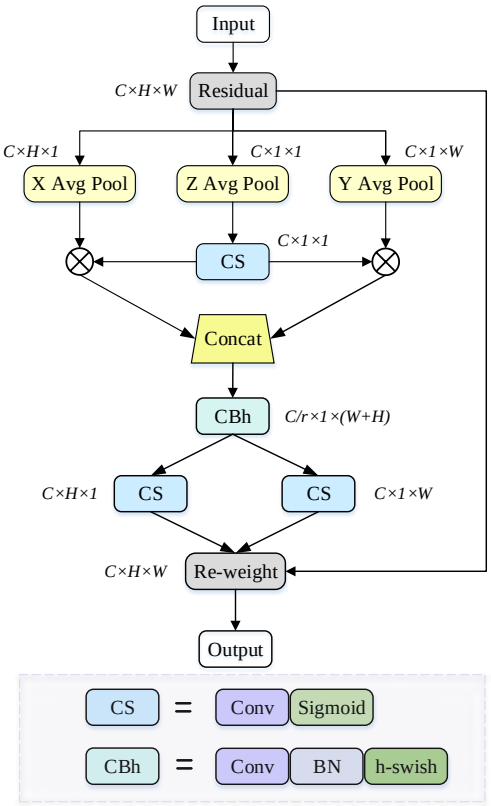


图 3 全局注意力模块结构图

Fig.3 Structure diagram of GIAM

最后，将 f 沿着空间维度，对 f 进行 split 操作，拆分成 $f_X' \in R^{C/r\times H}$ 与 $f_Y' \in R^{C/r\times W}$，分别进行卷积与激活操作后，将得到的输出用以注意力权重，和输入特征 X 进行点乘，得到该注意力模块最后的输出，计算公式如下：

$$f_X' = ConvUnit\left(f_X'\right) \tag{7}$$

$$f_Y' = ConvUnit\left(f_Y'\right) \tag{8}$$

$$f_{out}' = X \times f_X' \times f_Y' \tag{9}$$

其中，$ConvUnit$ 代表一个卷积单元，包含卷积层与 Sigmoid 激活层。

在深度卷积神经网络中，特征图不仅存在空间方向上的空间信息，还有通道方向上的通道信息。对于特征图数据，通过不同的卷积模块学习不同方向上的信息，然后通过加权融合方式可以得到特征图的权重，以进行空间和通道的信息交流。

本模块将和 C3 模块结合使用，命名为 C3-GIAM，为了探究 C3-GIAM 嵌入到网络的哪一部分能使得模型性能最佳，我们将其嵌入到骨干网络和颈网络，嵌入骨干网络的模型命名为 YOLO-CG1，嵌入颈网络的模型命名为 YOLO-CG2，同时嵌入骨干网络和颈网络的模型命名为 YOLO-CG3，具体结构如图 4 所示。全局注意力模块不仅考虑了空间方向的特征信息，还考虑了通道信息对输出特征的影响，可以使模型进一步关注特征图中的有效信息，能够提高特征选取的效率，提升商标识别结果的准确性。

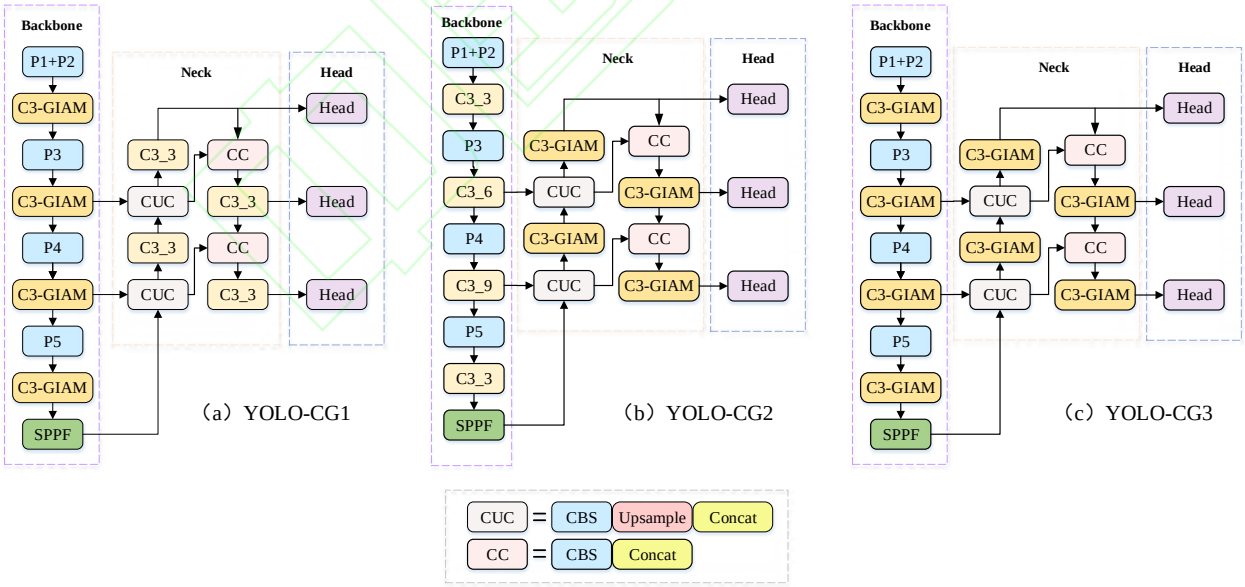


图 4 YOLOv5+C3-GIAM 网络结构图

Fig.4 Structure diagram of YOLOv5+GIAM network structure

1.3 C3TR-EAM 模块

本文借鉴了 CBAM 注意力机制中空间注意力机制的思想，提出了一种增强空间注意力模块。该模块利用了不同空洞率的空洞卷积，以获得更大的感受野，使模型能在空间维度上提取更有价值的信息。

首先，对特征图进行全局最大池化与全局平均池

化操作压缩通道，并将特征图送入卷积核大小为 3×3 的标准卷积与卷积核大小为 3×3、空洞率为 2 的空洞卷积中，通过并行运算来获得空间信息，对特征图进行通道拼接后送入卷积核大小为 1×1 的标准卷积中还原通道数。然后，再经过 Sigmoid 激活函数，与输入特征图做乘法操作，生成最终的空间特征图。本文将该模块与 CBAM 注意力机制中的通道注意力模块相结合，构成增强注意力模块，结构如图 5 所示。

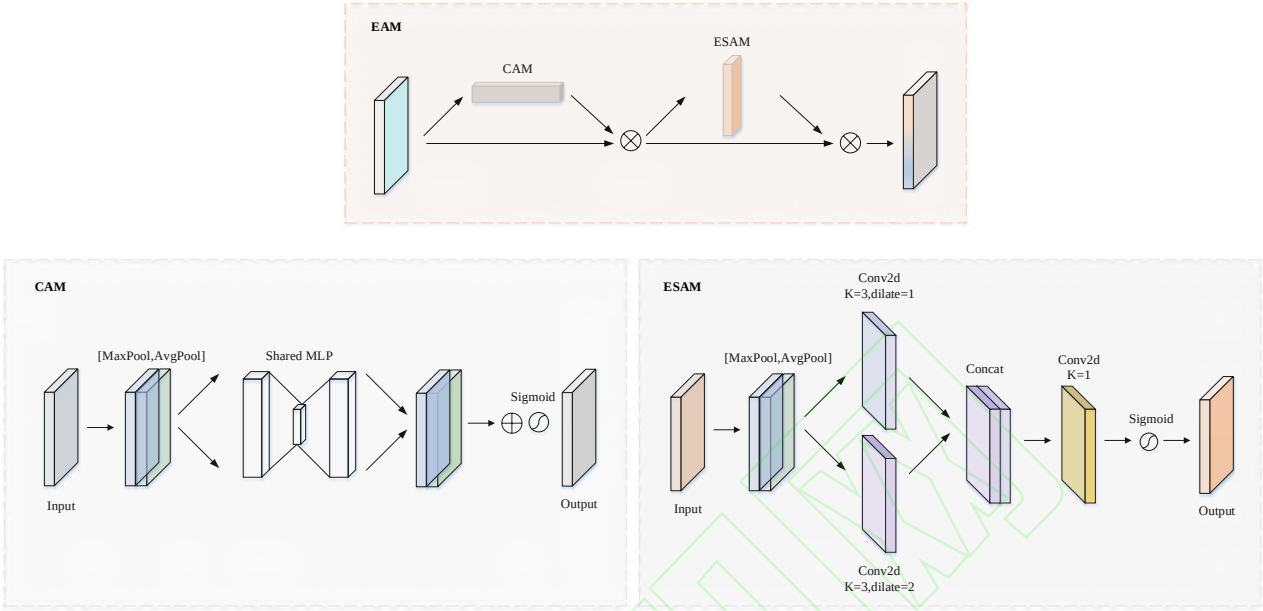


图 5 增强注意力模块结构图

Fig.5 Structure diagram of EAM

同时，由于近年来 Transformer[29]结构不仅在 NLP（Natural Language Processing，NLP）[30]领域取得可观结果，其在计算机视觉领域中也取得了很好的进展，例如 Face-book AI 提出的 DETR[31]框架、Google 提出的 ViT 模型、微软提出的 Swin Transformer[32]模型等，均是借鉴了 Transformer 的思想。本文将 Backbone 中的 C3 模块与 Transformer 结构相结合，Transformer 结构如图 6 所示，构建 C3TR 模块，并与上述的增强注意力模块相结合，构建 C3TR-EAM 模块。但考虑到 Transformer 模块对硬件配置要求较高，如果骨干网络的所有 C3 模块均替换，反而会导致训练效果不佳，我们选择将骨干网络的最后一层 C3 模块替换为 C3TR-EAM 模块。

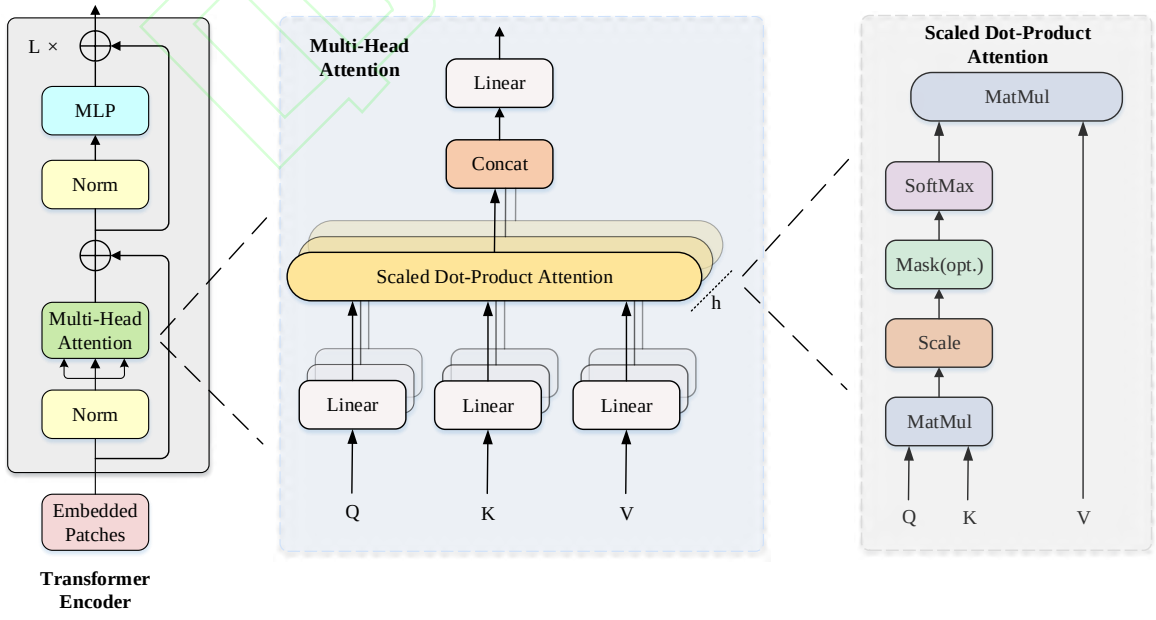
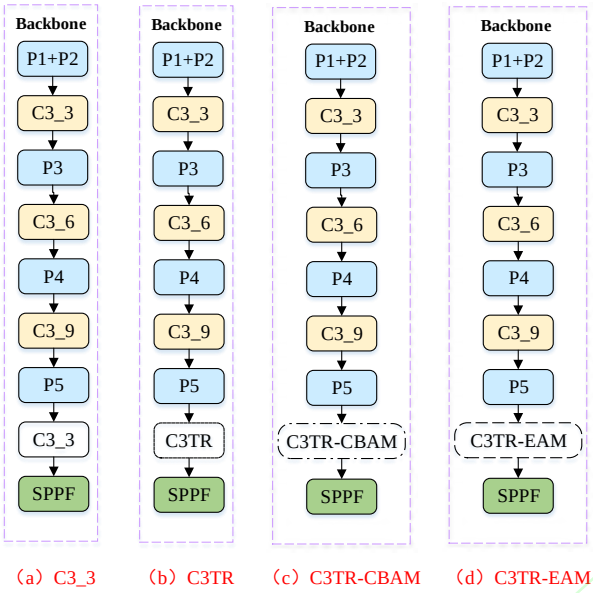


图 6 Transformer 结构图

Fig.6 Structure diagram of Transformer

由于本文的 C3TR-EAM 模块是在 C3TR 模块的基础上对 C3TR-CBAM 的改进，为了评估 C3TR-EAM 模块的有效性，我们将骨干网络的最后一层 C3 模块分别替换为 C3TR 模块和 C3TR-CBAM 模块，具体结构如图 7 所示。



（a）C3_3 （b）C3TR （c）C3TR-CBAM （d）C3TR-EAM

图 7 四种骨干网络结构图

Fig.7 Four kinds of backbone network structure

2 实验及结果分析

2.1 数据集与实验设置

本文对于图形类商标的分类为：圆形、矩形、三角形、扇形与五角星形，而对于这种分类当前并没有公开的商标数据集，因此选择自定义数据集，部分图形类商标数据集图片如图 8 所示。

（a）复杂背景 （b）遮挡

（c）多角度 （d）多目标

图 8 图形类商标数据集实例

Fig.8 Examples of graphic trademark dataset

我们从公开的商标数据集中人工筛选了 5255 张商标图片，其中圆形商标 3328 张，矩形商标 461 张，三角形商标 549 张，扇形商标 493 张，五角星形商标 424 张。将数据集按 9：1 的比例分为训练集和测试集，使用 Labelimg 对筛选后的数据打标签，标签文件为 txt 格式。在训练样本和测试样本的分割中，我们采用交叉验证的方法来保证测试的公平性和有效性。

为了评估本文提出 YOLO-FGE 架构的鲁棒性，本文在公开数据集 FlickrSportLogos-10-master 进行了验证。FlickrSportLogos-10 数据集是一个包含 361、Adidas、Anta、Erke 和 Kappa 等 10 种体育运动品牌的数据集，共有 2038 张图片。

实验设置基本采用 YOLOv5 的官方推荐参数设置，采用自适应 anchor 以及 mosaic 数据增强，输入图像尺寸大小为 640×640，单批次训练量为 32，最大迭代次数为 300，初始化学习率为 0.01。

2.2 实验环境与评估标准

本实验的开发环境为 Anaconda 4.12.0 与 PyTorch 1.10.2。实验平台操作系统为 Centos® 7.9.2009，CPU 为 Intel® Xeon® CPU E5-2630 v4 (10 core，2.4 GHz)，GPU 为 NIVIDA® GeForce® RTX 2080 Ti（11G 显存）。

为了客观评价模型的鲁棒性，在模型检测精度方面，本实验中用到的评价指标有准确率（Precision）、召回率（Recall）、平均精度均值（Mean of average precision，mAP）、浮点运算次数。其中，准确率、召回率、平均精度均值的计算公式如下所示：

$$Precision=\frac{TP}{TP+FP}\times 100\% \tag{10}$$

$$Recall=\frac{TP}{TP+FN}\times 100\% \tag{11}$$

$$mAP=\frac{\sum_{i=1}^{N}AP_i}{N} \tag{12}$$

其中，TP（True Positives）表示被正确检测到的目标数量，FP（False Positives）表示被网络错误认为是目标的数量，FN（False Negatives）表示漏检的目标数量，N 表示共需要分类的类别数，AP_i 表示某个目标类的平均精度。

2.3 C3-GIAM 不同位置的对比实验

为了探究 C3-GIAM 的有效性以及该模块的最佳嵌入方式，我们设计了三种网络结构进行对比实验，

网络结构如图 4 所示，实验结果如表 1 所示。

表 1 C3-GIAM 不同位置的对比实验

Table 1 Comparative experiments on different positions of C3-GIAM

Module	Precision	Recall	mAP%@0.5	GFLOPS
original	86.5	88.8	89.6	15.8
YOLO-CG1	86.0	90.5	91.1	17.6
YOLO-CG2	**91.0**	**90.6**	**91.3**	**14.1**
YOLO-CG3	89.5	86.7	90.2	15.9

由表 1 可知，将 C3-GIAM 分别嵌入骨干网络和颈网络 mAP 指标均有提升，其中 YOLO-CG1 模型的 mAP 提升了 1.5%，YOLO-CG2 模型的 mAP 提升了 1.7%，YOLO-CG3 模型的 mAP 提升了 0.6%。该数据证明了 GIAM 模块可以很好地捕捉长程依赖关系与精确的位置关系，能够使网络更加关注于有用的信息，从而更准确地定位商标区域。根据三种嵌入方式的实验结果可知，嵌入颈网络的检测精度提升的更高，且计算量较小。因此，我们选择了将 C3-GIAM 嵌入颈网络，我们认为将 C3-GIAM 嵌入这部分可以实现注意力重构，突出重要信息。

总的来说，GIAM 模块存在以下优势：①GIAM 模块方便灵活，对于任何一个经典模型都可以实现即插即用的功能；②GIAM 模块可以使模型聚焦于对当前任务更为关键的信息，从而更准确地实现目标定位。

为了更直观地验证该模块的优越性，选取数据集中具有代表性的图片进行验证，使用 Grad-CAM 做热力图的可视化，结果如图 9 所示。从图中可以看出，GIAM 模块可以更加有效地聚焦商标的特征，在图中则表现为商标与背景界限更加清晰，颜色与环境差异明显。

（a）原始

（b）原始+GIAM

图 9 添加 GIAM 前后的热力图

Fig.9 Thermodynamic diagram before and after adding GIAM

2.4 C3TR-EAM 的对比实验

为了评估 C3TR-EAM 模块的有效性，如图 7 所示，我们设计了三种骨干网络的结构，在图形类商标数据集上进行对比实验，结果如表 2 所示。

表 2 C3TR-EAM 有效性实验

Table 2 Effectiveness experiments of C3TR-EAM

C3 Module	Precision	Recall	mAP%@0.5	GFLOPS
original	86.5	88.8	89.6	15.8
C3-EAM	89.0	88.5	91.7	16.4
C3TR	91.0	87.3	91.6	15.6.
C3TR-CBAM	90.3	88.7	91.7	16.0
C3TR-EAM	**90.4**	**86.8**	**92.0**	**16.0**

从表 2 中可以看出，在 C3 模块的基础上引入 EAM 模块，mAP 与准确率分别得到了 2.1%、2.5%的提升；用 C3TR 模块代替 C3 模块，mAP 可以得到 2.0%的提升，准确率提升了 4.5%；在此基础上引入 CBAM 模块，mAP 会再提升 0.1%；而在 C3TR 模块的基础上引入 EAM 模块，mAP 会再提升 0.4%，且相较于引入 CBAM 模块，在计算量没有增加的同时提升了精度，与原有模型相比，mAP、准确率分别得到了 2.4%、3.9%的提升，这些数据证明了 C3TR-EAM 的有效性。

2.5 消融实验

为了进一步验证改进算法的有效性，对算法所提的添加增强注意力模块、替换 C3-GIAM 模块、替换 C3TR-EAM 模块的方法设计了 5 组消融实验，每组实验所使用的环境和训练技巧相同。结果如表 3 所示。训练轮数与改进模型指标之间的关系如图 10 所示。

表 3 消融实验结果

Table 3 Ablation results

FEM	C3-GIAM (Neck)	C3TR-EAM	Precision	Recall	mAP%@0.5	GFLOPS
			86.5	88.8	89.6	15.8
√			91.5	87.8	91.8	18.2
	√		91.0	90.6	91.3	14.1
		√	90.4	86.8	92.0	16.0

√	√		90.7	89.1	92.1	16.5
√	√	√	**92.1**	**90.8**	**92.3**	**16.7**

从上表中的结果中可以看出，添加 FEM 模块后，相比于原始的 YOLOv5s 网络模型，mAP 提高了 2.2%，准确率上涨了 5.0%，由于在 2 个特征层都加入了该模块，会导致模型参数量增加，检测速度有所下降，浮点运算次数由原来的 15.8 变为 18.2，但明显的变化是 mAP 得到了提升，说明 FEM 模块的引入增强了特征层对商标的适应性，可以使网络更加关注商标的重要信息，进而达到提升模型精确度的效果。经过 C3-GIAM 位置对比实验后，我们将 GIAM 嵌入 Neck 中的 C3 模块，较于原始 YOLOv5s 网络模型，mAP 提高了 1.7%，准确率与召回率分别上涨了 4.5%、1.8%，使模型不仅考虑了空间方向的特征信息，还考虑了通道信息对输出特征的影响，进而更加关注特征图中的有效信息。在 Backbone 中仅将 C3 模块替换为 C3TR-EAM 模块，较于原始 YOLOv5s 网络模型，mAP 提高了 2.4%，准确率提升了 3.9%。接着将 FEM 与 C3-GIAM 模块结合后，相较于原始 YOLOv5s 网络模型，mAP 提升了 2.5%，准确率与召回率分别上涨了 4.2%、0.3%。最后，将三个模块相结合，mAP 最终提升了 2.7%，准确率与召回率分别上涨了 5.6%、2.0%。由此可见，本文所提出的每个改进策略均能提高网络模型的性能，进而提升商标的检测精度。

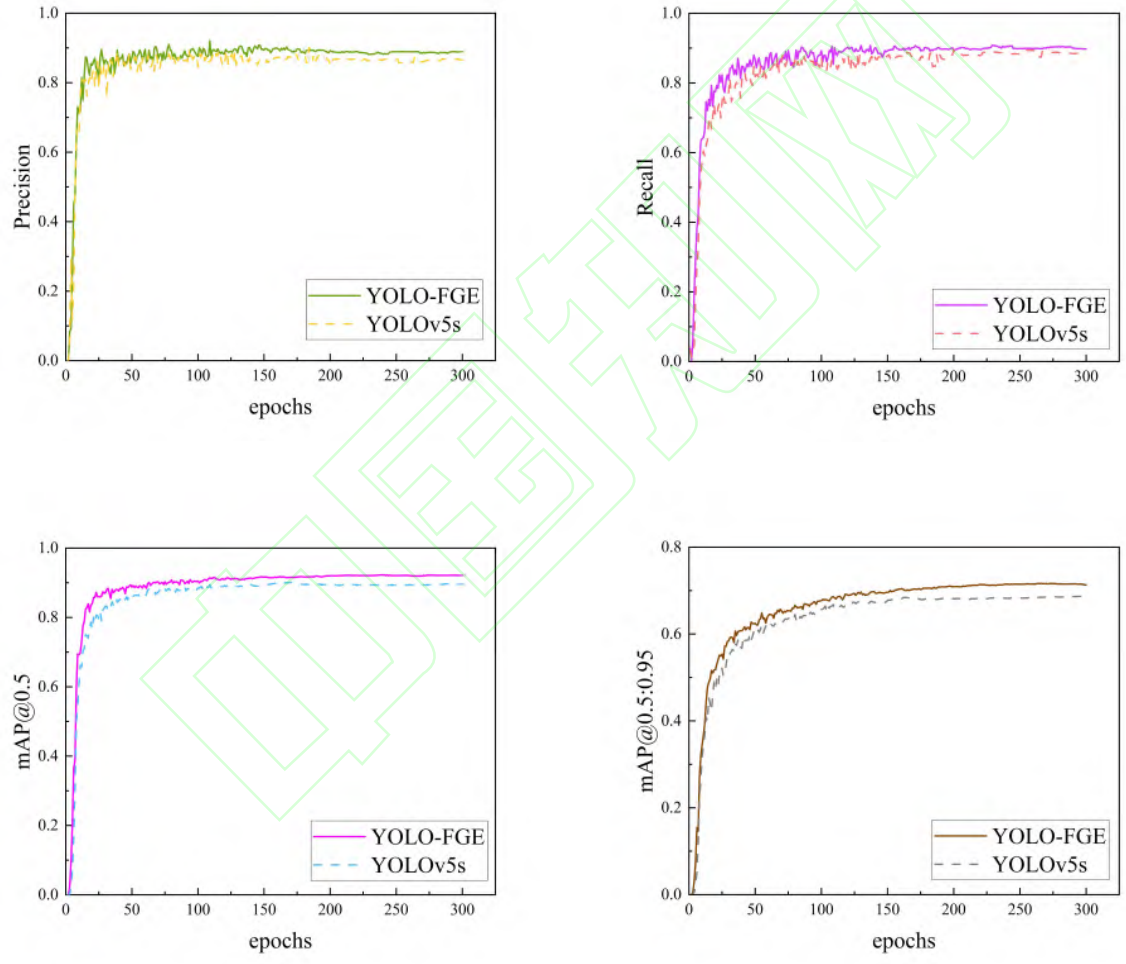


图 10 YOLO-FGE 指标曲线图

Fig.10 Index change curves of YOLO-FGE

2.6 对比实验

为了验证 YOLO-FGE 模型的有效性，本实验将该模型与 YOLOv3、YOLOv5s、YOLOv5m、YOLOv5l、YOLOv5x、YOLOv7、YOLOv8、文献[22]模型进行对比实验，实验结果如表 4 所示。

从表 4 中的结果中可以看出，本文的改进方法相较于原始模型，运行后 mAP、精确率和召回率分别得到了 2.7%、5.6%、2.0%的提高。相较于 YOLOv3 模型与 YOLOv5m 模型，mAP 分别提高了 1.1%、1.4%，精确率均提高了 2.1%，召回率分别提高了 2.3%、4.8%。而对于模型较大的 YOLOv5l 模型与 YOLOv5x 模型，这里我们不做比较。相较于最新的 YOLOv7 与

YOLOv8 模型，mAP 分别提升了 0.9%、5.8%，精确率分别提升了 5.3%、3.6%，召回率分别提升了 2.4%、11.1%。相较于文献[22]中提出的方法，mAP、精确率、召回率分别提升了 1.3%、2.9%、1.5%。本实验证明本文提出的模型相对于传统的算法，融合了更多特征，保持了较高的识别精度，相比于其他网络有较为明显的优势。

表 4　主流网络模型检测能力的对比实验

Table 4　Comparative results of mainstream network model detection capability

Method	Precision	Recall	mAP% @0.5	GFLOPS
YOLOv3	90.0	88.5	91.2	154.6
YOLOv5s	86.5	88.8	89.6	15.8
YOLOv5m	90..0	86.0	90.9	47.9
YOLOv5l	89.5	90.5	93.3	107.7
YOLOv5x	89.5	92.6	93.6	203.8
YOLOv7	86.8	88.4	91.4	105.2
YOLOv8	88.5	79.7	86.5	28.4
文献[22]	89.2	89.3	91.0	155.4
YOLO-FGE	**92.1**	**90.8**	**92.3**	**16.7**

相较于普通的目标检测任务，商标检测对网络结构的特征提取能力要求较高，传统的目标检测算法难以满足需求。本文提出的 YOLO-FGE 模型通过加深网络深度、提取全局信息、扩大感受野等方式，有效地提升了对商标的检测能力，通过增加较小的计算量，精度得到较高的提升。

图 11 为图形类商标中部分商标改进前后检测结果的比较。从图中可以看出，YOLOv5s 模型有遗漏和错误的检测，而 YOLO-FGE 模型对商标的检测更充分，检测结果更好。

图 11　图形类商标数据集的对比实验结果图

Fig.11　Comparative experimental results on graphic trademark dataset

同时，商标的大小对检测的精确度和鲁棒性也有一定的影响，针对这个问题我们进行了一系列的实验，可视化结果如图 12 所示。下列的商标图片中既包含了对大目标的检测，也包含了对小目标的检测，可以看出，我们的模型对小目标的检测精度是有明显提升的。

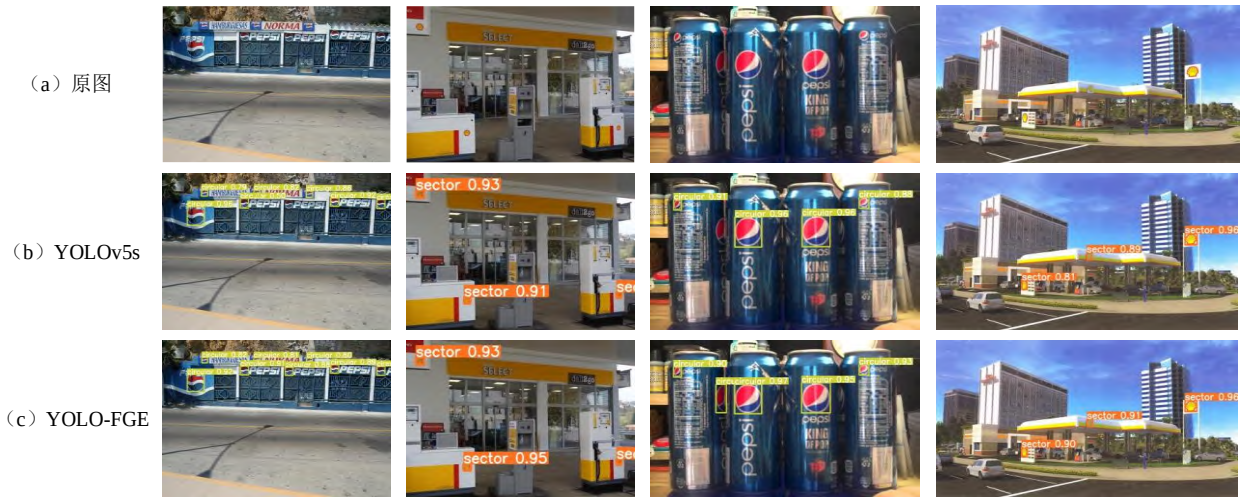


图 12　大小不同商标的对比实验结果图

Fig.12 Comparative experimental results of different sizes of trademarks

2.7 通用性实验

表 5 总结了在 FlickrSportLogos-10-master 数据集的实验结果。从表中可以看出，相较于 YOLOv3 模型，本模型在精确率、召回率和 mAP 指标上分别提升了 3.5%、7.5%、4.4%；相较于 YOLOv5s、YOLOv5m、YOLOv5l、YOLOv5x 模型，mAP 分别提高了 3.7%、3.5%、2.3%、1.7%，精确率分别提升了 3.1%、7.7%、2.5%、2.5%，召回率分别提升了 6.9%、3.8%、6.3%、4.7%；相较于最新的 YOLOv7 与 YOLOv8 模型，mAP 分别提升了 2.2%、5.2%，精确率分别提升了 6.9%、7.1%，召回率分别提升了 4.2%、8.4%。

表 5 FlickrSportLogos-10-master 数据集检测结果

Table 5 Detection results on the FlickrSportLogos-10-master dataset

Method	Precision	Recall	mAP% @0.5	GFLOPS
YOLOv3	91.3	79.3	85.2	154.6
YOLOv5s	91.7	79.9	85.9	15.8
YOLOv5m	87.1	83.0	86.1	47.9
YOLOv5l	92.3	80.5	87.3	107.7
YOLOv5x	92.3	82.1	87.9	203.8
YOLOv7	87.9	82.6	87.4	105.2
YOLOv8	87.7	78.4	84.4	28.4
YOLO-FGE	**94.8**	**86.8**	**89.6**	**16.7**

在公共数据集 FlickrSportLogos-10-master 上，YOLO-FGE 算法、YOLOv5s 算法、YOLOv3 算法、YOLOv7 算法与 YOLOv8 算法在训练 300 轮之后的效果对比如图 13 所示。从图中可以看出，本文提出的算法具有较好的鲁棒性。

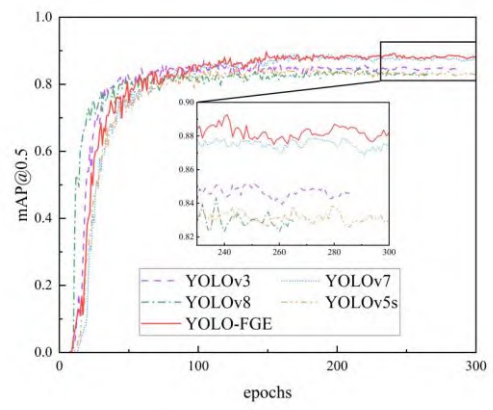


图 13 mAP 曲线

Fig.13 mAP curves

3 总结

本文提出了一种结合特征增强和注意力机制的目标检测算法 YOLO-FGE 网络。在原有模型中引入一种新的特征增强模块，通过多次堆叠不同卷积核的卷积来加深网络的深度，使得网络关注到更有用的信息，并设计了一种全局注意力机制，使得网络既可以捕获方向感知与精确的位置信息，又可以捕获跨通道信息，并与 C3 模块结合，探究 C3-GIAM 嵌入到算法网络对网络性能的影响。此外，构建了一种融合空洞卷积的增强空间注意力模块，通过不同空洞率的空间卷积来扩大感受野，并与通道注意力模块、Transformer 模块结合使用，提高网络对商标的检测能力。实验结果表明，本文提出的算法较原始 YOLOv5s 算法，mAP 提升了 2.7%，同时准确率与召回率分别提升了 5.6%、2.0%，对于图形类商标的识别有较好的鲁棒性。

参考文献：

[1] Jordan M I, Mitchell T M. Machine learning: Trends, perspectives, and prospects[J]. Science, 2015, 349(6245): 255-260.

[2] Mahesh B. Machine learning algorithms-a review[J]. International Journal of Science and Research (IJSR)[Internet], 2020, 9: 381-386.

[3] Voulodimos A, Doulamis N, Doulamis A, Protopapadakis E. Deep learning for computer vision: A brief review[J]. Computational intelligence and neuroscience, 2018, 2018.

[4] Hu H, Gu J, Zhang Z, Dai J, Wei Y. Relation networks for object detection[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2018: 3588-3597.

[5] Chen S, Wang H, Xu F, Jin Y-Q. Target classification using the deep convolutional networks for SAR images[J]. IEEE Transactions on Geoscience and Remote Sensing, 2016, 54(8): 4806-4817.

[6] Doan V-S, Huynh-The T, Kim D-S. Underwater acoustic target classification based on dense convolutional neural network[J]. IEEE Geoscience and Remote Sensing Letters, 2020, 19: 1-5.

[7] Yu C, Wang J, Peng C, Gao C, Yu G, Sang N. Bisenet: Bilateral segmentation network for real-time semantic segmentation[C]. Proceedings of the European

conference on computer vision (ECCV), 2018: 325-341.

[8] Fan M, Lai S, Huang J, Wei X, Chai Z, Luo J, Wei X. Rethinking bisenet for real-time semantic segmentation[C]. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2021: 9716-9725.

[9] Wang X, Zhang R, Kong T, Li L, Shen C. Solov2: Dynamic and fast instance segmentation[J]. Advances in neural information processing systems, 2020, 33: 17721-17732.

[10] Bolya D, Zhou C, Xiao F, Lee Y J. Yolact: Real-time instance segmentation[C]. Proceedings of the IEEE/CVF international conference on computer vision, 2019: 9157-9166.

[11] Albawi S, Mohammed T A, Al-Zawi S. Understanding of a convolutional neural network[C]. 2017 international conference on engineering and technology (ICET), 2017: 1-6.

[12] Den Hollander R J, Hanjalic A. Logo recognition in video stills by string matching[C]. Proceedings 2003 International Conference on Image Processing (Cat No 03CH37429), 2003: III-517.

[13] Bagdanov A D, Ballan L, Bertini M, Del Bimbo A. Trademark matching and retrieval in sports video databases[C]. Proceedings of the international workshop on Workshop on multimedia information retrieval, 2007: 79-86.

[14] Zhou H, Yuan Y, Shi C. Object tracking using SIFT features and mean shift[J]. Computer vision and image understanding, 2009, 113(3): 345-352.

[15] Kleban J, Xie X, Ma W-Y. Spatial pyramid mining for logo detection in natural scenes[C]. 2008 IEEE International Conference on Multimedia and Expo, 2008: 1077-1080.

[16] Sharma N, Mandal R, Sharma R, Pal U, Blumenstein M. Signature and logo detection using deep CNN for document image retrieval[C]. 2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR), 2018: 416-422.

[17] Sahel S, Alsahafi M, Alghamdi M, Alsubait T. Logo detection using deep learning with pretrained CNN models[J]. Engineering, Technology & Applied Science Research, 2021, 11(1): 6724-6729.

[18] Yousaf W, Umar A, Shirazi S H, Khan Z, Razzak I, Zaka M. Patch-CNN: deep learning for logo detection and brand recognition[J]. Journal of Intelligent & Fuzzy Systems, 2021, 40(3): 3849-3862.

[19] Alshowaish H, Al-Ohali Y, Al-Nafjan A. Trademark image similarity detection using convolutional neural network[J]. Applied Sciences, 2022, 12(3): 1752.

[20] Sengupta A, Ye Y, Wang R, Liu C, Roy K. Going deeper in spiking neural networks: VGG and residual architectures[J]. Frontiers in neuroscience, 2019, 13: 95.

[21] Trappey A J, Trappey C V, Lin E. Intelligent trademark recognition and similarity analysis using a two-stage transfer learning approach[J]. Advanced Engineering Informatics, 2022, 52: 101567.

[22] Wang J, Min W, Hou S, Ma S, Zheng Y, Jiang S. Logodet-3k: A large-scale image dataset for logo detection[J]. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 2022, 18(1): 1-19.

[23] Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision[C]. Proceedings of the IEEE conference on computer vision and pattern recognition, 2016: 2818-2826.

[24] Szegedy C, Ioffe S, Vanhoucke V, Alemi A. Inception-v4, inception-resnet and the impact of residual connections on learning[C]. Proceedings of the AAAI conference on artificial intelligence, 2017.

[25] Woo S, Park J, Lee J-Y, Kweon I S. Cbam: Convolutional block attention module[C]. Proceedings of the European conference on computer vision (ECCV), 2018: 3-19.

[26] Niu Z, Zhong G, Yu H. A review on the attention mechanism of deep learning[J]. Neurocomputing, 2021, 452: 48-62.

[27] 夏鸿斌, 肖奕飞, 刘渊. 融合自注意力机制的长文本生成对抗网络模型[J]. 计算机科学与探索, 2022, 16(7).

XIA Hongbin, XIAO Yifei, LIU Yuan. Long Text Generation Adversarial Network Model with Self-Attention Mechanism[J]. Journal of Frontiers of Computer Science & Technology, 2022, 16(7).

[28] 程艳, 蔡壮, 吴刚, 等. 结合自注意力特征过滤分类器和双分支 GAN 的面部表情识别[J]. 模式识别与人工智能, 35(3): 243-253.

CHENG Yan, CAI Zhuang, WU Gang, et al. Facial Expression Recognition Combining Self-Attention Feature Filtering Classifier and Two-Branch GAN[J]. Pattern Recognition and Artificial Intelligence, 35(3): 243-253.

[29] Han K, Xiao A, Wu E, Guo J, Xu C, Wang Y. Transformer in transformer[J]. Advances in neural information processing systems, 2021, 34: 15908- 15919.

[30] Chowdhary K, Chowdhary K. Natural language processing[J]. Fundamentals of artificial intelligence, 2020: 603-649.

[31] Dai Z, Cai B, Lin Y, Chen J. Up-detr: Unsupervised pre-training for object detection with transformers[C]. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2021: 1601-1610.

[32] Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, Lin S, Guo B. Swin transformer: Hierarchical vision transformer using shifted windows[C]. Proceedings of the IEEE/CVF international conference on computer vision, 2021: 10012-10022.