

融合FMix及注意力层叠加态的FSC-YOLOv5s模型

张俪馨, 曾庆涛*, 齐亚莉, 董武, 余丽琴

北京印刷学院, 信息工程学院, 出版物数据资产评估实验室, 北京

收稿日期: 2023年7月10日; 录用日期: 2023年8月22日; 发布日期: 2023年8月30日

摘要

在目标检测方面, 针对YOLOv5s算法处理小数据集的薄弱性和局限性, 设计了FSC-YOLOv5s算法。当数据采集受限时, 采用数据增强方式, 就可以在不增加原始数据量的基础上扩张特征点数量, 为检测效果带来正向增益效果, 是轻量型网络的最优化选择。首先, 在数据预处理时, 在原有结构基础上叠加引入一种混合样本数据增强方法FMix, 用以增强数据量, 能够降低小数据集对检测精度的影响。其次, 对YOLOv5s算法的网络结构进行改进, 在获取网络输出内容时, 增加可以统一权值的SimAM无参数注意力层。无需添加额外参数量, 便可以加强模型对于数据关键特征的关注。同时, 在主干网络部分加入CBAM注意力层, 通过学习自动提取重要特征和抑制次要特征, 进一步加强对于有限数据特征量的深入学习。通过注意力机制改善重要特征的筛选能力, 能够有效提升检测目标的完整性。然后, 分别加入SGD、Adam、Adamw优化算法进行实验, 择优选择可以提高FSC-YOLOv5s的计算效率和适应性的优化器。最后, 通过实验显示, FSC-YOLOv5s在两个数据集上的mAP@.5:.95分别提高了30.3%和5.1%, 验证了FSC-YOLOv5s算法的有效性。

关键词

目标检测, 数据增强, YOLOv5s, 注意力机制

FSC-YOLOv5s Model Integrating FMix and Attention Stacking States

Lixin Zhang, Qingtao Zeng*, Yali Qi, Wu Dong, Liqin Yu

Publication Data Asset Valuation Laboratory, School of Information Engineering, Beijing Institute of Graphic Communication, Beijing

Received: Jul. 10th, 2023; accepted: Aug. 22nd, 2023; published: Aug. 30th, 2023

*通讯作者。

文章引用: 张俪馨, 曾庆涛, 齐亚莉, 董武, 余丽琴. 融合 FMix 及注意力层叠加态的 FSC-YOLOv5s 模型[J]. 软件工程与应用, 2023, 12(4): 651-662. DOI: 10.12677/sea.2023.124063

Abstract

In terms of target detection, the FSC-YOLOv5s algorithm is designed to address the weaknesses and limitations of the YOLOv5s algorithm in handling small data sets. When data acquisition is limited, the use of data augmentation can expand the number of feature points without increasing the original data volume, bringing a positive gain effect to the detection effect, which is the most optimal choice for lightweight networks. First, a mixed-sample data enhancement method FMix is introduced to enhance the data volume during data preprocessing, which can reduce the impact of small data sets on the detection accuracy. Second, the network structure of the YOLOv5s algorithm is improved by adding a SimAM parameter-free attention layer that can unify the weights when obtaining the network output content. The attention of the model to the key features of the data can be enhanced without adding an additional number of parameters. At the same time, the CBAM attention layer is added to the backbone network part to further enhance the in-depth learning of limited data features by automatically extracting important features and suppressing minor features through learning. Improving the screening ability of important features through the attention mechanism can effectively improve the integrity of detection targets. Then, it is combined with SGD, Adam, and Adamw optimization algorithms, respectively, to select optimizers that can improve the computational efficiency and adaptability of FSC-YOLOv5s selectively. Finally, the experiments showed that FSC-YOLOv5s improved the mAP50-95 by 30.3% and 5.1% on both datasets, respectively, verifying the effectiveness of the FSC-YOLOv5s algorithm.

Keywords

Object Detection, Data Enhancement, YOLOv5s, Attention Mechanism

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

目标检测是计算机视觉领域的重要研究方向,在很多实际应用中都发挥了关键作用。例如:在印刷行业需要利用目标检测来进行印刷品检验,航行管理需要利用目标检测检验设备安全,自动驾驶需要利用目标检测实时监测路面状况等等。随着各种移动终端设备的普及,人们开始利用以终端为载体的目标检测来满足日常生活中的检测需求。目前,已有的算法主要分为传统、基于深度学习的两种类型的目标检测算法,对于传统目标检测算法的采用率已经逐渐降低了,传统目标检测算法也被称为是“两阶段算法”,主要是依靠人工或是算法寻找到目标对象的某种属性特征,依赖目标对象的特征来实现算法的检测,较为常见的传统算法有:Viola Jones 检测器、HOG 检测器,以及基于部件的可变形模型等等。基于深度学习机制的目标检测算法也被称为“一阶段算法”,大致可以分为两类:一是基于选定区域的目标检测算法,二是基于回归的目标检测算法。以目标检测算法发展史的时间为轴,“二阶段方法”在前,“一阶段方法”在后。“二阶段方法”因为其处理过程分为提取和检测两个阶段而得名,比较著名的“二阶段方法”是诞生于 2014 年的 CVPR R-CNN [1] 算法,该算法逐一对候选区域进行特征提取,对于一些重叠区域进行特征提取时,会造成大量的计算冗余,从而影响计算速度。YOLOv5 算法为“一阶段方法”,“一阶段方法”是基于“二阶段方法”的技术进步,“一阶段方法”在基于前者的算法基础上,去掉了

提取选定区域特征的步骤，直接进行回归计算。Redmon 等人在 2016 年时首次提出了 YOLO 算法，标志着单阶段目标检测的时代来临。此后 YOLOv 系列算法蓬勃发展，从 YOLOv0 只能做单目标检测，到 YOLOv3 基本确定 YOLO 的基本思路，再到改进后的 YOLOv5 已经拥有较为成熟的网络结构，并拥有良好的检测效果。YOLO 系列算法以轻量级网络立足，适用于各种数据集的目标检测。但检测算法如何在基于小数量级数据集的基础上，进行提升检测效率的同时保证良好的检测性能，是亟待解决的问题。

结合小数据集的特点和利用单阶段 YOLO 系列算法的实时性和准确性，采用了 YOLOv5 算法中最为轻型的网络结构 YOLOv5s 结构进行改进，充分利用其优势。解决了小数据集特征数量窄少，目标对象针对性差，以及分类精度欠佳等问题，有效提高了小数量级数据集目标检测的精度，主要创新点包括以下几项：

一为了解决小数据集特征数量欠佳，分析了深度学习网络模型中的数据增广的原理，在不引入额外数据集的情况下增加数据多样性。提出了可以多种数据增广方式叠加，将 FMix [2] 模块引入数据增广部分，以加强特征融合。

二是将轻量型 SimAM [3] 和 CBAM [4] 注意力模块引入网络结构，在不增加目标检测参数量的情况下提升检测性能，进一步改进模型的实时性与识别精度。

三是尝试 Adam [5]、Adamw [6]、SGD [7] 不同的优化算法对于检测效果的影响。在确保检测网络实时性的前提下，最终得到最优化效果。

2. 基本原理

2.1. YOLOv5 算法的基本结构

YOLOv5 [8] 算法模型是单阶段算法，主要是通过主干网络直接提取不同层次所拥有的高级特征，再通过使用多级融合网络将位置信息和高级语义信息进行充分融合，最后采用回归算法计算，从而实现目标检测的最终目标。YOLOv5 算法机构简洁，如图 1 所示，主要可分为四个部分：输入端、主干网络、neck 网络部分以及输出端部分。YOLOv5 算法是在 YOLOv3 算法结构上进行的进一步修改，YOLOv5 算法分有四种结构：YOLOv5s、YOLOv5x、YOLOv5m、YOLOv5l。其中，YOLOv5s 模型所带来的检测性能较为突出，且其为最轻型的网络结构。

2.2. 输入端

输入端部分主要是针对数据进行了数据增强、自适应锚框计算以及自适应图片缩放等操作。数据增强部分采用了 Mosaic 方法，综合了以往数据增强方式的优势，采取了随机排布、随机裁剪、随机缩放等等操作。增加训练样本数量，提升模型训练的鲁棒性。经过数据增强后的数据特征更具有随机性和不确定性，有利于提升数据集的复杂度。自适应图片缩放操作时会产生图片填充黑边，而过多的图像黑边会造成计算冗余，拖慢检测速度。自适应锚框计算针对此种情况进行了算法改进，在对原始图像进行缩放操作时会参考图像本身的尺寸，使得计算量得到减少，从而提高目标检测的速度。

2.3. Backbone 主干网络

Backbone 主干网络包括 Focus 结构和 CSP (Cross Stage Partial Network) [9] 结构。Focus 结构最为关键的是切片操作，将特征图进行切片处理后，再进行卷积操作。在 YOLOv5s 中，Focus 结构使用了 32 个卷积核进行卷积操作。而 CSP 结构在 YOLOv5s 中有两处应用，分别在 Backbone 主干网络和 Neck 部分。利用 CSP 结构可以增强网络的学习能力，持续轻量化网络结构，提高检测准确性，降低计算的内存成本。

2.4. Neck 部分

YOLOv5 算法在 Neck 层采用了自顶而下的 FPN (Feature Pyramid Network) [10]结构和自上而下的 PAN (Path Aggregation Network) [11]结构。FPN 通过上采样操作, 采集高层语义信息与底层相融合。PAN 结构通过下采样操作, 采集底层的特征信息传至高层, 以便输出预测的特征信息。YOLOv5s 的 Neck 部分设计是为了充分利用 Backbone [12] [13]主干网络所提取的特征信息, 传达强语义特征, 保证目标特征的提取检测。

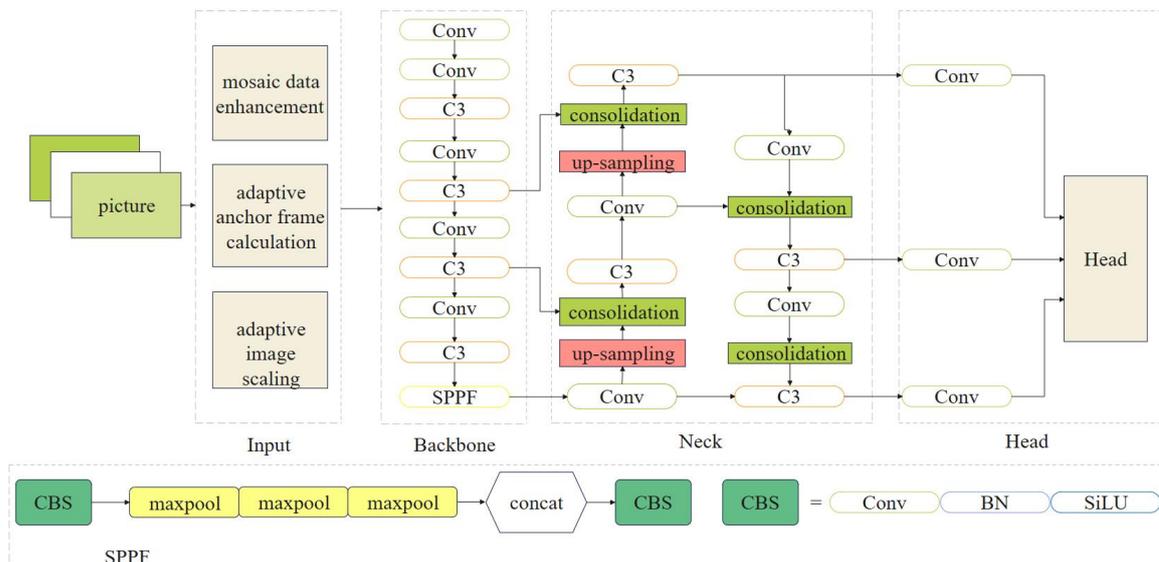


Figure 1. YOLOv5 basic network structure

图 1. YOLOv5 基本网络结构

2.5. FMix 算法模块

数据增强方式可以大致分为两类: 基础数据增强和高级数据增强。基础数据增强包含 Rotation、Translation 等几何数据增强方式, 以及 Cutout、GridMask 是侧重于图像特征视觉外观的非几何数据增强方式, 还有 Mixup、CutMix 等图像混合数据增强方式。高级数据增强中最具代表性的就是自动数据增强, 可以分为基于强化学习、非强化学习、风格迁移、特征空间四个部分。

FMix [2]是非单图像的混合数据增强算法, 由 Ethan Harris 等人提出。FMix 算法是基于 CutMix 算法进行的改进, 不仅保留了模型对真实数据的良好理解能力以及可以通过动态增加降低数据集的过拟合风险, 还改进了 CutMix 算法困限于固态矩形框。如图 2 所示, FMix 算法采用更加柔和的拼接方式, 从而保证图像的局部一致性, 提高特征语义信息的保留率。

引入了 FMix 模块使得数据特征可以得到更好地融合, 进一步扩展网络学习的数据特征数量。FMix 算法是一种混合样本数据增强, 先从 beta 分布中获取 lambda (λ)值, 再从傅里叶空间低频采样其灰度掩码, 对其进行二值化处理, 最后利用该掩码对像素进行加权。计算公式如下所示:

$$G = \Re(F^{-1}(\text{filter}(z, \sigma))) \quad (1)$$

$$\text{filter}(z, \sigma)[i, j] = \frac{z[i, j]}{\text{freq}(w, h)[i, j]^\sigma} \quad (2)$$

式中, λ 表示为复数随机变量, g 表示为衰减功率, \mathcal{F}^{-1} 表示为离散傅里叶反变换, \Re 表示其输入的实部, ω 表示与离散傅里叶变换相对应的样本频率的大小。然后将灰度图像转化为具有阈值的二进制格式掩码, 具体公式如下:

$$\text{mask}(\lambda, g)[i, j] = \begin{cases} 1, & ij \ g[i, j] \in \text{top}(\lambda wh, g) \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

将相关参数值和输入变量进行计算掩码混合函数:

$$\text{mix}(x_1, x_2, m) = m \odot x_1 + (1 - m) \odot x_2 \quad (4)$$

x_1, x_2 是输入图片, 为计算得出的二进制格式掩码, 最后将输入图像进行拼接处理, 输出最终目标图像, 如图 2 所示:

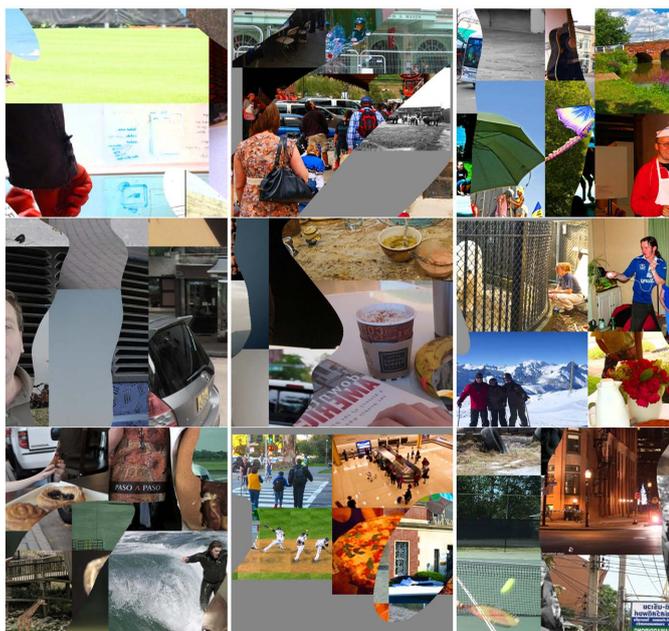


Figure 2. Image of the data enhancement effect processed by adding the FMix algorithm
图 2. 加入 FMix 算法处理后的数据增强效果图像图示

2.6. 注意力机制

计算机视觉中所应用的注意力机制[14] (Attention Mechanisms)是模仿人类视觉在大脑皮层的注意力机制[15], 可以更快速高效地分析复杂场景信息。注意力机制可以看作是检测网络对于输入图像的重要特征信息的筛选过程, 注意力机制的提出是计算机视觉技术的一次飞跃发展, 可以提高运算结果的性能。注意力机制按照侧重权重的不同大致可以分为: 通道注意力机制、空间注意力机制、时间注意力机制、分支注意力机制, 还有两种混合注意力机制: 通道空间混合注意力机制和空间时间注意力机制。

FSC-YOLOv5s 算法引入了两种注意力机制: CBAM 注意力机制和 SimAM 无参数注意力机制, 将两种模块分别插入 Backbone [16]主干网络和 Head 部分, 以增强目标检测效果的性能。

CBAM (Convolutional Block Attention Module)是一个简洁高效的前馈卷积神经网络注意力模块。如图 3 所示。CBAM 包含两个相互独立的子模块, 分别是通道注意力模块(Channel Attention Module, CAM)和空间注意力模块(Spatial Attention Module, SAM), 如图 4、图 5 所示。这两个模块可以并行或是串行的方

式组合，先利用通道注意力模块，再采用空间注意力机制，可以更好地提升网络结构的性能。

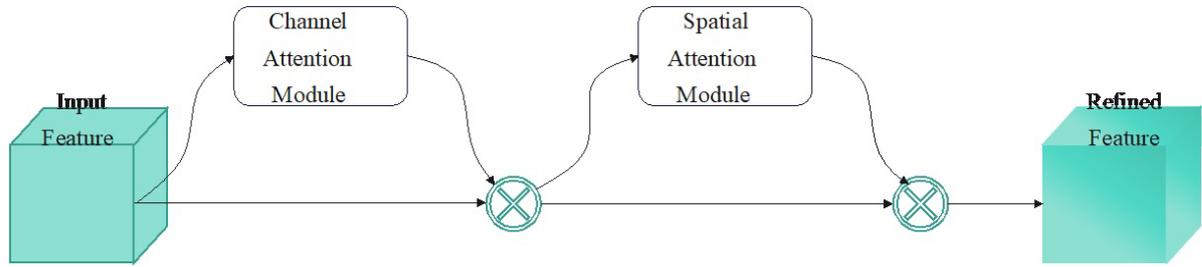


Figure 3. CBAM module flow signal

图 3. CBAM 模块流程示意

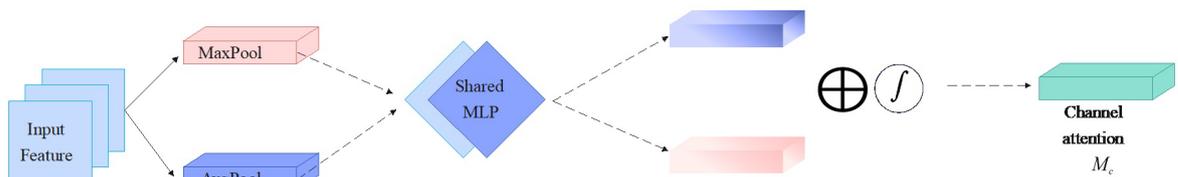


Figure 4. Channel attention module

图 4. 通道注意力模块

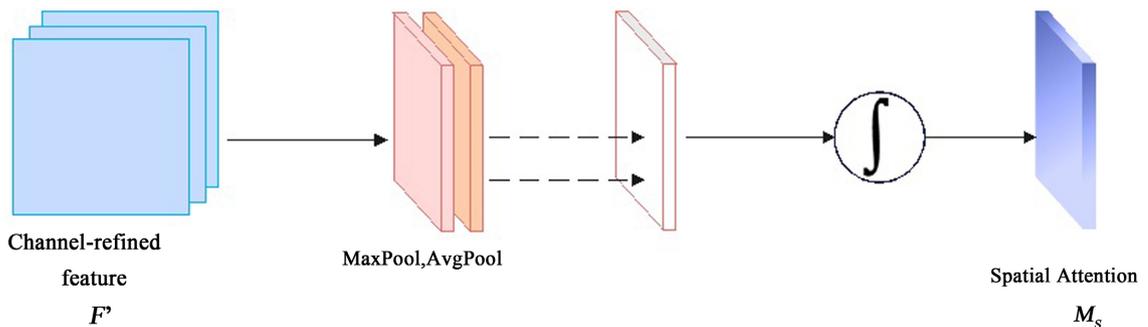


Figure 5. Spatial attention module

图 5. 空间注意力模块

CBAM 模块的计算流程是先对主干网络生成的特征图产生 1D 通道注意力特征图和 2D 通道注意力特征图，如公式 5、6 所示。分别沿着通道和空间注意力模块推算注意力权重，然后再将所得到的权重和输入的特征图相乘，进行自适应特征优化。

$$F' = M_c(F) \otimes F \tag{5}$$

$$F'' = M_s(F') \otimes F' \tag{6}$$

式中， $M_c \in R^{C \times 1 \times 1}$ ， $M_s \in R^{1 \times H \times W}$ ， $F \in R^{C \times H \times W}$ 。

SimAM [17] 注意力机制是一种简单有效的 3D 注意力模块。CBAM 模块类似于人脑的注意力机制，会在估计 1D 和 2D 的注意力权重后，再将其结合起来。所以 CBAM 注意力机制并不会产生真实的三维权重。而 SimAM 注意力机制 [18] 可以有效地产生真实的三维权重。在 SimAM 模块 [19] 的计算中，需要单估每个神经元的重要性，估算前需要找到这些神经元，测量目标神经元与其余神经元之间的线性可分离性，对于每一个神经元的能量量化，SimAM 以公式 7 为计算函数。

$$e_t(w_t, b_t, y, x_i) = (y_t - \hat{t})^2 + \frac{1}{M-1} \sum_{i=1}^{M-1} (y_0 - \hat{x}_i)^2 \quad (7)$$

式中, $\hat{t} = \omega_t t + b_t$, $\hat{x}_i = \omega_t x_i + b_t$, ω_t 和 b_t 为转换的权重和偏置。

通过最小化公式 7, 可以找到目标神经元 t 与同一通道中其他神经元之间的线性可分离性。对 y_t 和 y_0 采用二进制标签, 添加一个正则化器。每个通道在理论上都有 $M = H \times W$ 个能量函数, 最终的能量函数为公式 10 所示。

$$e_t(\omega_t, b_t, y, x_i) = \frac{1}{M-1} \sum_{i=1}^{M-1} (-1 - (\omega_t x_i + b_t))^2 + (1 - (\omega_t t + b_t))^2 + \lambda \omega_t^2 \quad (8)$$

$$\text{其中, } \omega_t = -\frac{2(t - \mu_t)}{(t - \mu_t)^2 + 2\sigma_t^2 + 2\lambda}, b_t = -\frac{1}{2}(t + \mu_t)\omega_t\lambda。$$

因此, 可以用下面公式计算最小能量。

$$e_t^* = \frac{4(\delta^2 + \lambda)}{(t - \mu)^2 + 2\sigma^2 + 2\lambda} \quad (9)$$

综上所述, 神经元的能量越低, 其重要性越高。每个神经元的重要性可以通过计算 $\frac{1}{e_t^*}$ 得到。使用缩放算符来进行特征细化, sigmoid 是一个单调函数:

$$\tilde{X} = \text{sigmoid}\left(\frac{1}{E}\right) \odot X \quad (10)$$

通过结合 SimAM 注意力模块, 可以提升检测算法的性能。

2.7. Adam、Adamw、SGD 优化算法

Adam [20] 优化算法, 是一种梯度下降优化算法的扩展, 又称之为自适应运动估计算法, 由 Diederik Kingma 和 Jimmy Lei Ba 提出。Adam 优化算法加速了优化过程, 具有高效的计算效率, 对于计算内存要求较小, 适合应用于计算数据和或是参数调节的问题, 适用于计算梯度较为稀疏问题。Adam 优于传统的随机梯度下降算法不同的是, 它可以随着展开学习行为的进程, 参考每一个网络权重参数, 从而单独调整相对应的学习速率, 从梯度的第一次和第二次矩值估算来计算不同参数的自适应学习效率。

假设一个存在一个有噪声的目标函数 $A_t(x)$, 运用 Adam 优化算法计算目标函数, m_t 定义为动量, v_t 可以理解为梯度变化的方差, 分别是 g_t 的一阶矩估计向量和二阶矩估计向量。计算第 t 步梯度的一阶矩估计向量 m_t 和二阶矩估计向量 v_t 分别是:

$$m_t = \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t \quad (11)$$

$$v_t = \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot g_t^2 \quad (12)$$

Adam 算法主要是在 RMSprop 的基础上增加了动量, 并进行了偏差修正。修正之后, 在时刻 t 的一阶矩估计量和二阶矩估计量分别是:

$$m_t = (1 - \beta_1) \sum_{i=1}^t \beta_1^{t-i} \cdot g_i \quad (13)$$

$$v_t = (1 - \beta_2) \sum_{i=1}^t \beta_2^{t-i} \cdot g_i^2 \quad (14)$$

对于式中参数, 可以设置参考值为 $\alpha = 0.001, \beta_1 = 0.9, \beta_2 = 0.999, \varepsilon = 10^{-8}$ 。

综上所述, Adam 优化器可以根据计算梯度的震荡情况和过滤震荡后的真实梯度对变量参数进行实时更新。

Adamw 优化算法包括权重衰减与 L2 正则化, 是在 Adam 算法和 L2 正则化的基础上进行改进的算法。L2 正则化是减少过拟合的经典方法, 对于优化器中的阶梯函数, 需要使用梯度修正参数, 在计算的时候会加上对正则项求梯度的结果。Adamw 函数与 Adam 函数直接加上 L2 正则项不同, 不是在损失函数的位置直接加上 L2 正则项, 而是将正则项的梯度加入反向传播的公式中, 代替在损失函数处添加正则项的操作。如下式所示:

$$X_t = X_{t-1} - \eta_t \left(\alpha m_t / (\sqrt{v_t} + \varepsilon) + \lambda \theta_{t-1} \right) \quad (15)$$

综上所述, Adamw 函数是基于 Adam 函数的基础上产生的变体, 用于处理大型数据集。Adamw 函数以一定比率来缩减模型参数的梯度, 从而减少计算量, 提升训练速度。

SGD 是一种最简单的优化算法。它采用了最简单的梯度下降法, 随步更新梯度, 算法的收敛速度受制于学习率。SGD 优化算法结构简单, 没有过多的参数需要调整, 较少的计算量可以获得较好的计算结果。其原理如下式所示:

$$W \leftarrow W - \eta \frac{\partial L}{\partial W} \quad (16)$$

式中, W 表示为更新的权重参数, η 表示学习率。

3. 分析与讨论

3.1. 实验设置

实验平台由某云服务器提供, 使用 GPU 配置为 NVIDIA GeForce RTX 3080, 采用 CUDA 11.1 加速库进行实验并加速, 采用基于 python3.8 的 PyTorch 深度学习框架实现。为了保证实验的实时性以及公平性, 各个模型输入图像尺寸均为 $640 * 640$, batch size 设置为 16, epoch 设置为 200, 使用 Warm up 和余弦退火的学习率衰减, 初始学习率为 0.01, 动量为 0.937。

3.2. 评价指标

实验所用的检测精度的评价指标选用 mAP50-95 值(mean Average Precision), 代表 IoU 阈值为 0.95 时各个类别 AP 的均值。mAP 值表示所有类别的平均精度值(Average Precision, AP)的平均值。在实验中将精确度和召回率作为检测效果的辅助参考指标。 P 为准确率, 在预测中该值理想状态为 1, R 为召回率, 在预测效果中的理想状态为 1。其计算公式如下:

$$P = \frac{TP}{TP + FP} \quad (17)$$

$$R = \frac{TP}{TP + FN} \quad (18)$$

式中, TP [21] 为正确分类的正样本数, FP 为误报的负样本数, FN 为漏报的正样本数。以下实验分别使用两个自建数据集进行实验测试。

3.3. 数据增强实验

数据增强[22]是深度学习对于数据预处理采用的一种较为常见的扩充数据的方法, 对于训练时参考数

数据集的数量级较小, 数据增广方法可以在不引入额外的数据集的情况下, 在现有的数据集现状上增加训练数据量, 降低数据采集成本。改进 YOLOv5s 网络模型, 在原先算法基础上引入 FMix 算法。

如表 1 所示, 加入 FMix 结构后, 在数据集 1 的实验中, P 值提升了 15.6%, R 值微值浮动, AP50-95 值提升 16.5%。数据集 2 的 R 值提升了 4.3%, AP50-95 值提升 1.3%。引入 FMix 结构, 可以在不加入额外数据集的情况下增加训练数据特征量, 对目标对象的检测效果的精度也有一定的提升效果。FMix 处理后的数据特征图, 在模板融合时选择了任意形状, 而非是固定的矩形方块, 保留了更多的语义信息分布, 使其更加适合针对小数据集的数据增强。

Table 1. Comparison of the quantitative results before and after the addition of the FMix module

表 1. 加入 FMix 模块前后的定量结果对比

数据集	网络结构	精确率(P)	召回率(R)	精度 AP50-95
Dataset1	YOLOv5s	0.711	1	0.585
	YOLOv5s + FMix	0.867	0.984	0.75
Dataset2	YOLOv5s	1	0.957	0.642
	YOLOv5s + FMix	0.99	1	0.655

3.4. 优化器选择实验

实验针对 Adam 优化器、Adamw 优化器、SGD 优化器进行了对比实验论证。根据实验效果数据, 选择出最适合的优化算法。

从表 2 中的结果对比可知, 在参数量和计算量基本没有发生改变的情况下, 引用 SGD 优化算法使得模型检测精度达到较好效果。在两个数据集的实验中应用 SGD 优化器都比应用 Adam、Adamw 优化器的检测精度高, 数据集 1 采用 SGD 优化器的 AP 值分别比采用 Adam 优化器和 Adamw 优化器提高了 60.3 个百分点和 49.8 个百分点。数据集 2 采用 SGD 优化器的 AP 值分别比采用 Adam 优化器和 Adamw 优化器提高了 13.5 个百分点和 28.5 个百分点。可以看到, 数据集 2 采用 SGD 优化器的检测效果比其余两种要更好, 其 P 值、R 值、AP 值、AP50-95 值均得到提升。因此, 据实验结果显示, 选择 SGD 优化算法是最有利于提升检测性能。

Table 2. Optimization algorithm module comparison experiment

表 2. 优化算法模块对比实验

数据集	网络结构	精确率(P)	召回率(R)	精度 AP50	精度 AP50-95
Dataset1	YOLOv5s + FMix + SGD	0.867	0.984	0.995	0.75
	YOLOv5s + FMix + Adam	0.177	1	0.392	0.276
	YOLOv5s + FMix + Adamw	0.187	0.75	0.497	0.276
Dataset2	YOLOv5s + FMix + SGD	0.99	1	0.995	0.655
	YOLOv5s + FMix + Adam	0.892	0.7	0.86	0.504
	YOLOv5s + FMix + Adamw	0.525	1	0.71	0.282

3.5. 注意力模块对比实验

为了验证注意力模块对目标检测算法的有效性以及探究注意力模块的最佳嵌入方式, 设计了对比实验。在以上实验的基础上, 针对加入 CBAM 模块的位置以及是否加入 SimAM 模块进行了探究。

表 3 中 CBAM_L1、CBAM_L2、CBAM_L3 对比可得, 不同位置插入 CBAM 注意力模块, 效果会不同。在三种嵌入方式中, 第三种嵌入方式明显优于其余两种, 即在 Backbone 的每一个卷积层后都加入一个 CBAM 注意力模块。在数据集 1 上, 选择 CBAM_L3 方案的 AP 值比 CBAM_L2 提高 8.3%, 比 CBAM_L1 的 AP 值提升 20.5%, 而且 CBAM_L3 方案的 P 值为三种方案中最高值, 其 R 值为 1。CBAM_L3 方案是三种植入注意力模块方案中效果最好的一种, 所以在原网络的 Backbone 主干网络中引入 CBAM 注意力模块, 且在特征提取部分加入注意力机制可以更好地提取特征, 提升检测性能。

Table 3. The CBAM attention module contrast experiment

表 3. CBAM 注意力模块对比实验

数据集	网络结构	精确率(P)	召回率(R)	精度 AP50	精度 AP50-95
Dataset1	FS-YOLOv5s + CBAM_L1	0.557	1	0.912	0.696
	FS-YOLOv5s + CBAM_L2	0.346	0.75	0.79	0.611
	FS-YOLOv5s + CBAM_L3	0.775	1	0.995	0.821
Dataset2	FS-YOLOv5s + CBAM_L1	0.981	0.9	0.986	0.665
	FS-YOLOv5s + CBAM_L2	0.882	1	0.968	0.673
	FS-YOLOv5s + CBAM_L3	0.997	1	0.995	0.634

YOLOv5s 网络中的 Head 部分主要包含了三个检测器、损失函数部分以及 Head 侧的优化策略, 即利用网络的锚框在不同尺度的特征图上进行目标检测的过程。在 Head 部分引入 SimAM 无参数注意力机制, 可以提升目标检测精度。实验结果如表 4 所示, 在数据集 1 的表现上, 于 Head 部分引入无参数注意力机制 SimAM 后, AP50-95 提高了 3.8%。在数据集 2 的实验中, AP50-95 提升了 5.9%。较之前实验效果来说有显著提升。实验发现在不同部分引入不同注意力模块, 对于深度学习网络都有正向增益效果。

Table 4. Attention module lateral contrast experiment

表 4. 注意力模块横向对比实验

数据集	网络结构	精确率(P)	召回率(R)	精度 AP50-95
Dataset1	FSCL3-YOLOv5s	0.775	1	0.821
	FSCL3-YOLOv5s + SimAM	0.926	0.85	0.908
Dataset2	FSCL3-YOLOv5s	0.997	1	0.634
	FSCL3-YOLOv5s + SimAM	0.984	1	0.693

由表 5 的消融实验可知, 引入 FMix 算法, 用以数据增强。在 Backbone 部分引入 CBAM 注意力模块, 在 Head 部分引入 SimAM 无参数注意力机制, 选择适合的 SGD 优化器, 会使得 P 值和 R 值达到一个相对理想状态的数值, 且 AP50-95 值得到提升。因此, FSC-YOLOv5s 网络结构有效提升了小数量级数据集的算法性能。

Table 5. Ablation experiment

表 5. 消融实验

数据集	网络结构	精确率(P)	召回率(R)	精度 AP50-95
Dataset1	YOLOv5s	0.711	1	0.585
	YOLOv5s + FMix + SGD	0.867	0.984	0.75
	FS-YOLOv5s + CBAM_L3	0.775	1	0.821
	FSCL3-YOLOv5s + SimAM	0.926	0.85	0.908

Continued

	YOLOv5s	1	0.957	0.642
Dataset2	YOLOv5s + FMix + SGD	0.99	1	0.655
	FS-YOLOv5s + CBAM_L3	0.997	1	0.634
	FSCL3-YOLOv5s + SimAM	0.984	1	0.693

4. 结论

针对目标检测任务中数据集数据量薄弱对目标检测造成困扰的问题, 提出引入 FMix 算法, 以增强训练数据样本, 同时引入了 CBAM 以及 SimAM 注意力模块, 增强了网络模型对于目标特征提取能力, 提高了模型的计算速度。选择了 SGD 优化算法, 提出的改进模型对在不加入额外的学习参数的情况下, 与原网络结构在小数量级的数据集上检测效果相比有了显著性提升。FSC-YOLOv5s 算法可以继续推广在动态图像的目标检测领域, 尝试在较高速或是图像固定时长较短的类型视频上, 以提升其检测精度和加快其检测速度。

基金项目

北京印刷学院科研创新团队项目: 面向按需出版的彩色高精度喷墨印刷系统关键技术研究(20190122019)、北京印刷学院学科建设和研究生教育专项: 电子信息研究生实验实践教学深化改革与质量提升项目(21090122012)、北京印刷学院校级项目(Ec202303)、北京印刷学院校级项目(Ea202301)、北京市数字教育研究重点课题(BDEC2022619027)、北京印刷学院学科建设与研究生教育专项(21090323009)。

参考文献

- [1] Girshick, R., Donahue, J., Darrell, T. and Malik, J. (2014) Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. 2014 *IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, 23-28 June 2014, 580-587. <https://doi.org/10.1109/CVPR.2014.81>
- [2] Harris, E., Marcu, A., Painter, M., et al. (2020) Fmix: Enhancing Mixed Sample Data Augmentation. <https://doi.org/10.48550/arXiv.2002.12047>
- [3] Yang, L., Zhang, R.Y., Li, L., et al. (2021) Simam: A Simple, Parameter-Free Attention Module for Convolutional Neural Networks. *International Conference on Machine Learning. PMLR*, 18-24 July 2021, 11863-11874.
- [4] Woo, S., Park, J., Lee, J.Y., et al. (2018) Cbam: Convolutional Block Attention Module. *Proceedings of the European Conference on COMPUTER Vision (ECCV)*, Munich, 8-14 September 2018, 3-19. https://doi.org/10.1007/978-3-030-01234-2_1
- [5] Kingma, D.P. and Ba, J. (2014) Adam: A Method for Stochastic Optimization. <https://doi.org/10.48550/arXiv.1412.6980>
- [6] Loshchilov, I. and Hutter, F. (2017) Decoupled Weight Decay Regularization. <https://doi.org/10.48550/arXiv.1711.05101>
- [7] Gower, R.M., Loizou, N., Qian, X., et al. (2019) SGD: General Analysis and Improved Rates. *International Conference on Machine Learning. PMLR*, Long Beach, 9-15 June 2019, 5200-5209.
- [8] 秦晓辉, 黄启东, 常灯祥, 等. 基于改进 YOLOv5 的露天矿山目标检测方法[J]. 湖南大学学报(自然科学版), 2023, 50(2): 23-30. <https://doi.org/10.16339/j.cnki.hdxzbk.2023261>
- [9] 皇甫俊逸, 孟乔, 孟令辰, 等. 基于 GhostNet 与注意力机制的 YOLOv5 交通目标检测[J/OL]. 计算机系统应用, 2023, 32(4): 149-160. <https://doi.org/10.15888/j.cnki.csa.009048>, 2023-04-13.
- [10] 易图明, 王全全, 袁威, 等. 一种基于 YOLOv5s 的改进装甲目标检测算法[J]. 现代信息科技, 2023, 7(5): 73-77. <https://doi.org/10.19850/j.cnki.2096-4706.2023.05.017>
- [11] 李小军, 邓月明, 陈正浩, 等. 改进 YOLOv5 的机场跑道异物目标检测算法[J]. 计算机工程与应用, 2023, 59(2): 202-211.
- [12] 谢椿辉, 吴金明, 徐怀宇. 改进 YOLOv5 的无人机影像小目标检测算法[J]. 计算机工程与应用, 2023, 59(9):

- 198-206. <http://kns.cnki.net/kcms/detail/11.2127.TP.20230214.1523.050.html>, 2023-04-13.
- [13] 张智华. 基于 YOLOv5 的智能图像识别检测[J]. 信息技术与信息化, 2023(2): 200-203.
- [14] 王年涛, 王淑青, 汤璐, 马丹. 基于 EfficientNet-YOLOv5s 的绝缘子缺陷检测[J]. 湖北工业大学学报, 2023, 38(1): 21-26.
- [15] 朱开. 基于改进 YOLOv5 的交通标志识别模型[J]. 信息技术与信息化, 2023(2): 30-33.
- [16] 张忠民, 吴泽. 基于改进 YOLOv5 的密集行人检测方法[J/OL]. 应用科技 1-7. <http://kns.cnki.net/kcms/detail/23.1191.u.20221101.1549.002.html>, 2023-04-13.
- [17] 刘丽伟, 侯德彪, 侯阿临, 等. 基于 SimAM-YOLOv4 的自动驾驶目标检测算法[J]. 长春工业大学学报, 2022, 43(3): 244-250. <https://doi.org/10.15923/j.cnki.cn22-1382/t.2022.3.09>
- [18] 郭聪, 杨敏. 基于无参注意力和特征融合的图像去噪算法[J]. 计算机技术与发展, 2023, 33(2): 50-56.
- [19] 郭玲, 于海雁, 周志权. 基于 SimAM 注意力机制的近岸船舶检测方法[J/OL]. 哈尔滨工业大学报, 2023, 55(5): 14-21. <http://kns.cnki.net/kcms/detail/23.1235.T.20220826.1508.010.html>, 2023-04-13.
- [20] 李梓毓, 赵月爱. 改进 Adam 优化算法的人脸检测方法[J]. 太原师范学院学报(自然科学版), 2022, 21(4): 58-63.
- [21] 马良荔, 李陶圆, 刘爱军, 等. 基于迁移学习的小数据集命名实体识别研究[J]. 华中科技大学学报(自然科学版), 2022, 50(2): 118-123. <https://doi.org/10.13245/j.hust.220218>
- [22] 皮洲, 奚雪峰, 崔志明, 等. 一种面向长文本小数据集自动摘要任务的数据增强策略[J]. 中文信息学报, 2022, 36(9): 46-56.