基于改进 YOLOX 的安全帽佩戴检测算法

熊 伟[1] 张中天[1] 路 鑫[1] 和宝同[1]
XIONG Wei ZHANG Zhongtian LU Xin HE Baotong

摘 要 针对安全帽佩戴检测任务中的小目标密集导致的遮挡和难以检测等问题，提出了一种基于 YOLOX 和 transformer 的安全帽佩戴检测算法。首先使用融合了 CNN 和 transformer 的网络 swin Csp 进行特征提取，提升模型获取全局上下文信息的性能；然后采用简化的重复加权双向特征金字塔网络 BiFPN，使算法更好地融合不同尺度的特征；最后在 neck 和 YOLO head 的连接部分加入 CBAM 注意力机制，加强安全帽特征与位置信息的关联程度，进一步增强网络的性能和鲁棒性。基于 SHWD 数据集进行实验，结果表明，算法在 hat 类上提高了 0.99% 的精度，在两种类型的召回率上分别有 3.16% 和 1.7% 的提升，在 mAP 方面则有 1.79% 的提升，有效实现了快速且精度较高的安全帽佩戴检测。

关键词 深度学习；目标检测；特征融合；注意力机制；安全帽识别

doi：10.3969/j.issn.1672-9528.2024.02.005

0 引言

安全帽在保护人员生命安全方面有着至关重要的作用。然而在实际的生产过程中，由于人员流动性大、工人安全意识不强等原因，许多工人在进行生产活动时不规范佩戴安全帽，由此造成过很多具有严重后果的事件。提高工人安全帽的佩戴可以减少工地安全事故中的人员伤亡，有效提升生产安全性。单纯地依靠人工进行检测存在成本高昂和效率低下的缺陷 [1]。故在减少人工监控的同时确保安全帽的正确佩戴已经成为亟待解决的问题。

现阶段的目标检测算法可以归为两种类型，一种是以 Faster R-CNN[2] 等算法为例的两阶段（two stage）算法。这一系列算法要先通过启发式的方法或是卷积神经网络生成候选区域，之后在候选区域上做分类与回归。这一系列方法具有较高的准确度和较强的可扩展性，但其识别速度慢，在工程实践项目上表现欠佳。另外一种是以 SSD 和 YOLO 算法为代表的一阶段（one stage）算法。该系列算法使用一个 CNN 对不同目标的类型和空间位置进行直接预测。前者使用多尺度特征层进行目标检测，将 Faster R-CNN 的 anchor 机制与 YOLO grid 的思想相结合。后者使用多尺度特征映射对整幅图像的每个位置都进行特征提取和回归分析。与 SSD 相比，YOLOv3 实现了更优秀的速度与精度的权衡 [3]。

近年来，基于深度学习的目标检测方法被广泛应用于各种检测任务中。吴冬梅等人 [4] 改进了 YOLOv4 中的 NMS 算法，并增加主干提取网络中的残差块个数以提升检测精度；石家玮等人 [5] 合并 BN 层和卷积层减少前向推理的计算量，再用 K-means 聚类改进先验框的维度；苏文俊等人 [6] 采用深度可分离卷积压缩 YOLOv4 的网络结构，在颈部网络添加 CSP 结构使网络更易优化；韩贵金等人 [7] 针对夜间光线不佳的情景，添加了多尺度残差模块和并行池化注意力模块以降低局部模糊和颜色丢失的影响。综上所述，以往的研究不够关注上下文信息的获取及目标与位置信息的关联程度，文本提出一种基于改进 YOLOX[8] 的安全帽佩戴检测算法，主要贡献如下。

（1）提出了一种新型的主干特征提取网络，将 swin-transformer 块添加到原 YOLOX 的主干特征提取网络中，以得到更多的全局上下文信息，进行更有区分度的特征提取。

（2）提出了一种轻量化的三层 BiFPN，以用于更有效地融合跨尺度的特征，提升模型对小目标的检测能力。将 CBAM 注意力机制嵌入到 neck 和 YOLO head 的连接处，以更准确地捕捉输入特征图中的关键信息，提高了模型的准确性和鲁棒性。

1 YOLOX 算法

YOLO 是一种经典的一阶段目标检测算法。它将分类问题转化为回归问题，各个类型的边界框坐标及置信度会直接采用回归的方法生成。与 Fast R-CNN 相比，大大提高了检测速度。

YOLOX 以 YOLOv3-Darknet53 作为基准模型，融合了 Mosaic、解耦头和修改后的 Mixup 数据增强、无锚框以及 SimOTA 等方法，达到了兼具速度和精度的一流检测水平。YOLOX 的网络结构图如图 1 所示。

在目标检测任务中分类和定位一直存在着冲突 [9]。在以

1. 华北电力大学计算机系 河北保定 071000

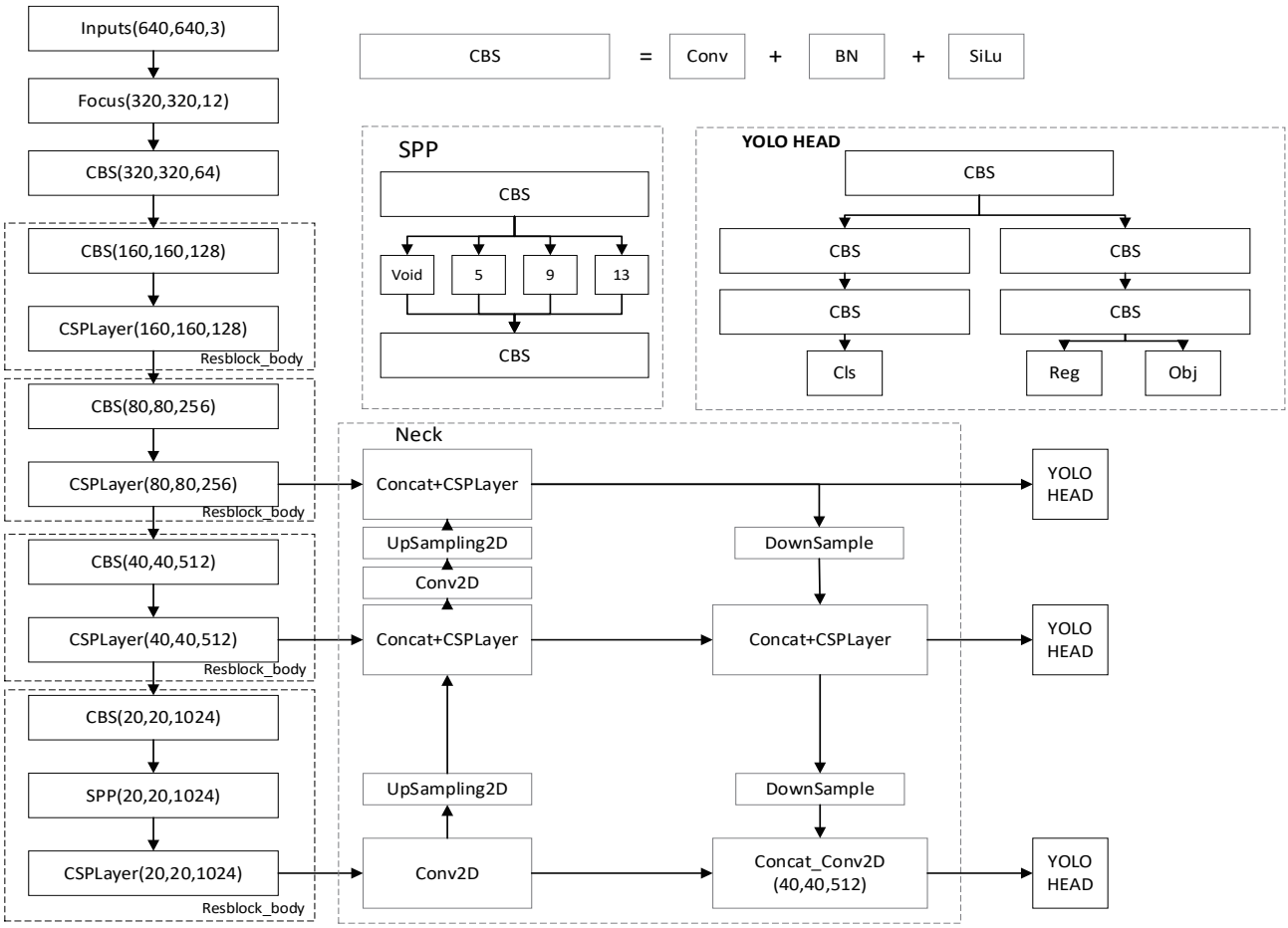

图 1 YOLOX 网络结构图

往的 YOLO 系列算法中都采用了相同的耦合检测头。为了缓解耦合检测头对目标检测器的负面影响，YOLOX 使用了轻型的解耦检测头。

在 YOLOX 中，会先通过 1*1 卷积的方式减少 YOLO neck 部分送来特征图的通道数，再让其经过两个并行的分支，每个分支由两个 3*3 的卷积层组成。它们分别进行分类和定位，且在定位的分支上还增加了 IoU 分支。得益于此种结构，YOLOX 实现了较大的性能提升，但其参数量只增加得很少。

近些年来的研究表明，无锚检测器的性能可以达到与基于锚的检测器同等的水平 [10-11]。所以为了减少参数量、提升检测速度和精度，YOLOX 使用了无锚的设计。此外 YOLOX 还采用了马赛克和经过修改的 Mixup 对数据进行增强，并调用 SimOTA 标签样本匹配算法等技术，全面提升了检测速度与精度，再加上其易于部署的特点，在工业领域具有一定的实践意义。

2 改进 YOLOX 的安全帽佩戴检测算法

本文提出了一种包含了卷积和 transformer 的混合网络结构，如图 2 所示。网络分为三部分，第一部分以 swin-CspDarkNet 为基础，集成了 swin transfomer 以获取更多的上下文信息且保留更多的特征信息。第二部分本文将 FPN 换成了能融合更多特征信息的自适应双向特征金字塔 BiFPN。本文还将 CBAM 注意力机制应用在了 neck 和 YOLO head 的连接部分，使模型可以关注到更重要的信息，忽略无关的信息，以此提高模型的性能。第三部分即采用解耦的检测头来检测安全帽。

2.1 backbone

安全帽佩戴检测任务经常面临小目标密集、遮挡和复杂背景等情况，这对目标检测模型而言是很棘手的问题。为了提高网络的特征提取能力，保留更多的全局上下文信息，本文将 swin transformer 融合到安全帽佩戴检测模型的主干特征提取网络中。

在传统的卷积神经网络中一般使用池化操作进行下采样，以此来扩大特征图的感受野，得到不同尺度的特征。在 swin transformer 中则使用 patch merging 来模拟卷积中的池化操作，其通过把多个近邻的小 patch 合成一个大 patch，扩大 patch 的感受野，同时也抓住了多尺度的特征。在网络相对较浅而特征映射相对较大时，过早应用 swin transformer 可能会丢失一些有意义的上下文信息。因此在 swin CSPDarkNet

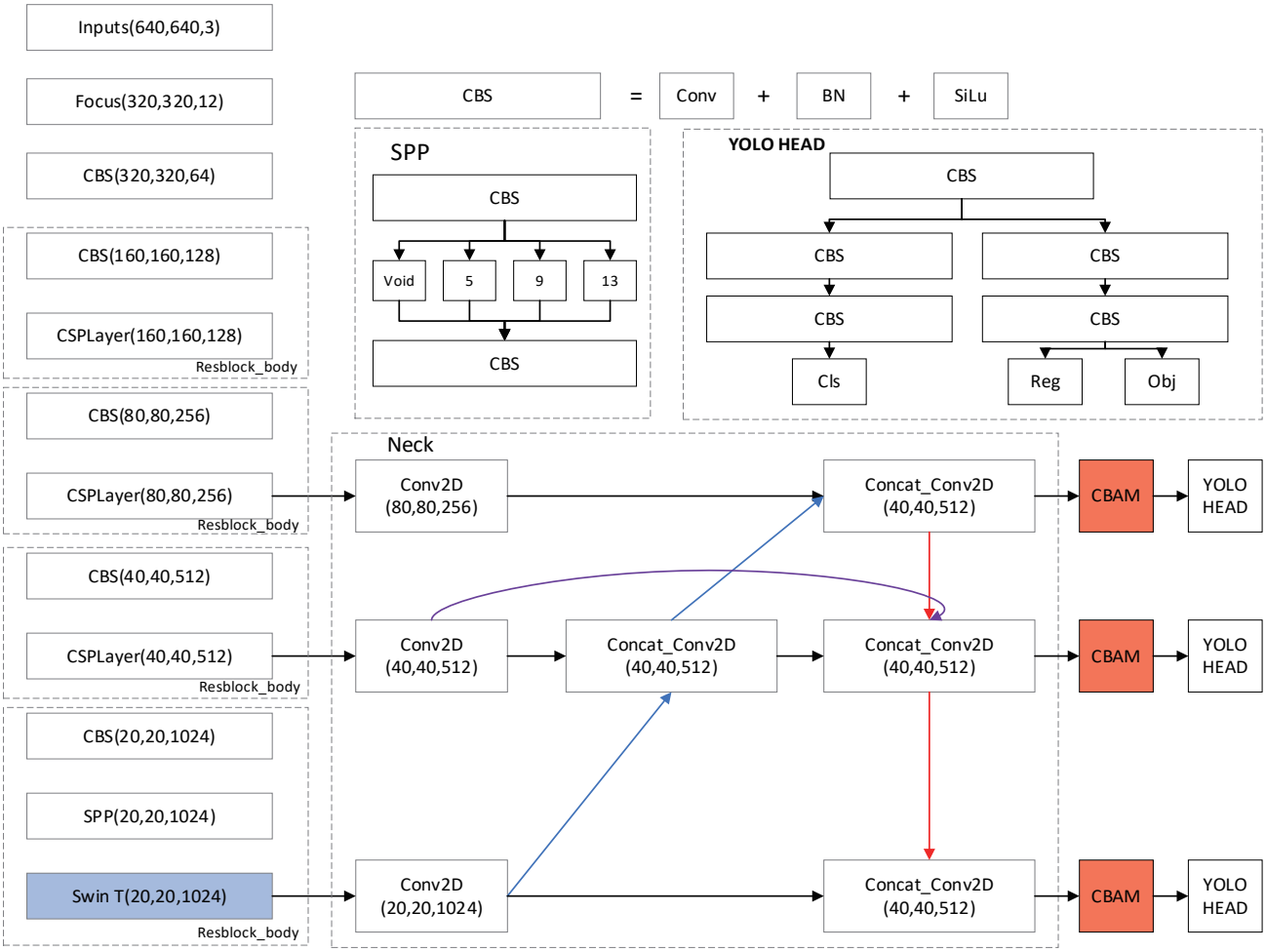

图 2 本文算法网络结构图

中，本文把 CspDarkNet 中的最后一个特征层替换为 swin transformer 层。

focus 模块和多个 Resblock_body 模块的堆叠构成了整个主干特征提取网络。focus 模块的主要工作是将输入张量按通道进行分组，并对每组通道做一定的变换，最终把一个新的张量输出。下采样和复数个残差结构共同组成了 Resblock_body 模块。本文使用 swin transformer 层对最后一个 Resblock_body 中的 CSPLayer 做了替换。

swin transformer 将输入信息分割为 4×4 大小的块，在每个块中分别计算自注意力，然后通过滑动窗口的方法在不同块之间实现了交互，以此来实现全局的自注意力，获得了更多的全局上下文信息。swin transformer 块的结构如图 3 所示，其由两部分构成，第一部分是窗口多头自注意力模块（Window-multi-head self attention），第二部分是滑动窗口多头自注意力模块（shifted Window-multi-head self attention）。每个模块包含两个子模块，其中一个子模块由归一化层和自注意力模块组成，另一个子模块由归一化层和 MLP 组成，使用残差连接两个子模块。

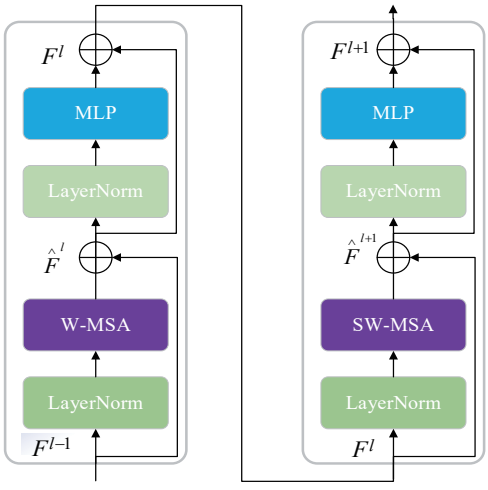

图 3 swin transformer 块

swin transformer 块的计算过程公式为：

$$\begin{aligned}\hat{F}^l &= \text{W-MSA}(\text{LN}(F^{l-1})) + F^{l-1},\\ F^l &= \text{MLP}(\text{LN}(\hat{F}^l)) + \hat{F}^l,\\ \hat{F}^{l+1} &= \text{SW-MSA}(\text{LN}(F^l)) + F^l,\\ F^{l+1} &= \text{MLP}(\text{LN}(\hat{F}^{l+1})) + \hat{F}^{l+1}\end{aligned} \quad (1)$$

式中：$\hat{F}^{l}$ 和 F^{l} 分别代表窗口多头自注意力模块和 MLP 模块的输出特征，$\hat{F}^{l+1}$ 和 F^{l+1} 代表第二个模块中滑动窗口多头自注意力模块和 MLP 模块的输出特征。

2.2 neck

在安全帽佩戴检测领域中，由于环境复杂、遮挡、距离过远等原因，摄像头捕捉到的图像中目标尺度经常有较大差异，故对于多尺度特征的把握十分关键。在以往的 YOLO 系列算法中，neck 部分一般会选用 FPN[12] 和 PANet[13] 特征融合网络。

BiFPN[14] 是一种基于特征金字塔网络（FPN）的神经网络结构，用于解决计算机视觉任务如目标检测、语义分割等中的特征融合问题。有别于传统的特征金字塔网络，BiFPN 在每个尺度上都进行了多次特征融合，以更有效地利用多层次的特征信息。BiFPN 不仅高效，而且准确和通用，在目标检测、语义分割等领域都被广泛使用。BiFPN 的结构如图 4 所示。多个 BiFPN 模块组成了 BiFPN 网络的主要结构，其中每个 BiFPN 模块都包含了一个特征金字塔、一个上采样网络和一个下采样网络。特征金字塔结构由多个分别对应不同尺度的特征层组成。上采样网络会将低分辨率的特征图上采样到高分辨率，而下采样网络则用于将高分辨率的特征图下采样到更低的分辨率。如此就可以在不同尺度的特征层之间建立起双向的连接，从而使得特征信息在不同尺度之间不受限制地流动和交互。

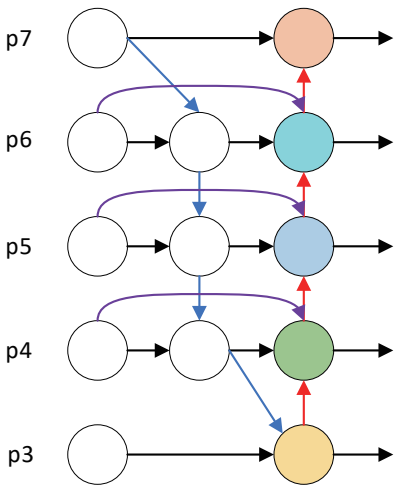

图 4 BiFPN 结构图

区别于传统的特征融合网络，BiFPN 在 p4、p5、p6 三层各添加一条额外的边以融合更多的特征信息，同时也不会增加过多的计算成本。此外，在把高级特征和低级特征相融合的时候，BiFPN 并非是简单相加或者拼接，而是采用可变的权重来学习不同的输入特征。以第 6 层为例，BiFPN 公式为：

$$P_6^{td}=\mathrm{Conv}\left(\frac{\omega_1 g P_6^{in}+\omega_2 g\,\mathrm{Resize}(P_7^{in})}{\omega_1+\omega_2+\varepsilon}\right),$$
$$P_6^{out}=\mathrm{Conv}\left(\frac{\omega_1' g P_6^{in}+\omega_2' g P_6^{td}+\omega_3' g\,\mathrm{Resize}(P_5^{out})}{\omega_1'+\omega_2'+\omega_3'+\varepsilon}\right) \quad (2)$$

在本文算法中，由于只有主干特征网络的最后三个特征层被使用，所以将 BiFPN 简化成三层的结构。此步骤可以在本文算法网络结构中加入 BiFPN，以实现对同级节点最大程度的特征融合，同时减少模型的计算复杂度。具体网络结构如图 5 所示。相比于五层的 BiFPN 网络，三层 BiFPN 网络的特征融合方式减小了来自浅层特征的噪声对模型的干扰，可以显著提高模型的检测速度。后续的消融实验也证明了 BiFPN 结构的变化对算法整体检测效果的影响，其具有更快的检测速度、更轻量的模型，同时还有更优秀的检测精度。

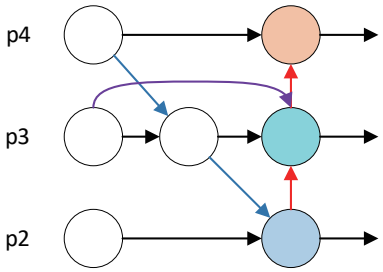

图 5 三层 BiFPN

总体而言，BiFPN 网络的主要优点在特征金字塔多尺度信息中充分利用，并通过双向连接对多尺度的特征图进行融合及增强。这种结构让网络可以更准确地理解图像中的目标，使目标检测网络的性能获得提升。

2.3 CBAM 注意力机制

在真实的工程实践中，注意力机制是深度学习中一种关键的技术，其可以让模型处理数据时关注更加重要的信息而选择性忽略无关的信息，以此提高模型的性能。

CBAM[15] 即卷积块的注意力模块，是一种兼具了通道和空间注意力机制特点的注意力机制，用于增强深度卷积神经网络的性能。其使用在卷积层后添加 CBAM 模块的方式，对特征图进行两方面的调节，使模型的精度得到提升。

通道注意力模块（channel attention module）和空间注意力模块（spatial attention module）两个模块共同构成了 CBAM 注意力机制。CBAM 的结构如图 6 所示。

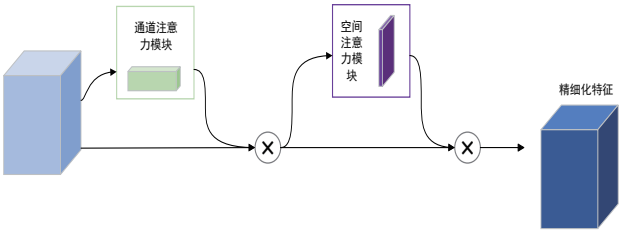

图 6 CBAM 注意力模块

其中 CAM 学习每个通道的特征，使得模型更加关注重要的特征通道。首先 CAM 将输入特征图按照深度 (通道) 维度分解成多个通道，并在每个通道上做最大池化和平均池化，得到两个池化结果。然后将这两个结果进行拼接后，经过两个全连接层得到通道注意力权重。最后将得到的权重乘上输入特征图，得到加权的特征图。通道注意力模块计算过程如图 7 所示。

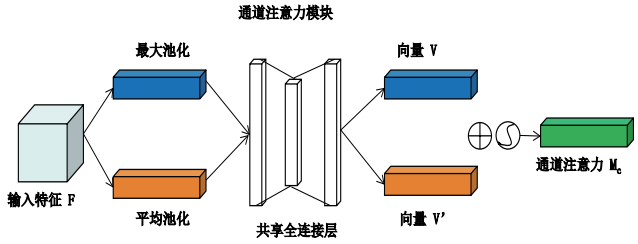

图 7 通道注意力

CAM 的计算公式如式（3）所示。其中 F 代表输入，σ 代表 sigmoid 激活函数。

$$\begin{aligned} A &= \mathrm{AvgPool}(F), \\ M &= \mathrm{MaxPool}(F), \\ M_C(F) &= \sigma(\mathrm{MLP}(A) + \mathrm{MLP}(M)) \end{aligned} \tag{3}$$

SAM 学习每个空间位置的特征，以使模型对重要的空间位置给予更多关注。首先 SAM 对输入特征图按照空间维度做全局平均池化，得到一个全局特征描述向量。然后经过两个全连接层做处理以得到空间注意力权重。最后将空间注意力权重与加权的特征图相乘，得到最终的特征图输出。空间注意力计算过程如图 8 所示。

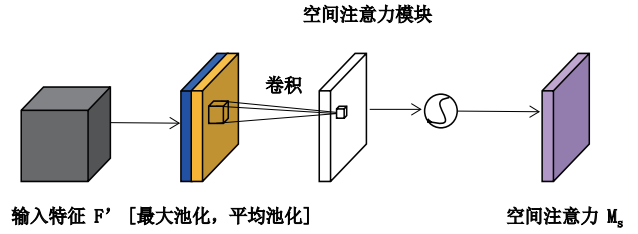

图 8 空间注意力

SAM 的计算公式为：

$$M_S(F) = \sigma(f^{n\times n}([A;M])) \tag{4}$$

式中：$f^{n\times n}$ 表示 $n\times n$ 的卷积计算。

可以同时学习通道和空间注意力权重是 CBAM 注意力机制的优点，这也意味着它能够自适应地对特征图中的不同通道和空间位置做调整，从而更加准确地捕获到输入特征图中的关键信息，提高模型的准确性和鲁棒性。尽管目前在目标检测任务中注意力机制被添加在 backbone 部分的用法居多，但是在安全帽佩戴检测任务中，尺度差异较大的输入信息不会获得很理想的实际效果。故本文将 CBAM 注意力模块添加在 neck 和 YOLO head 的连接之处。后续的消融实验证明，CBAM 模块在较小影响模型大小的前提下，增强了模型学习到重要信息的能力。CBAM 在本文算法结构的具体位置如图 2 中所示。

3 实验结果与分析

本文基于网络获取和现场实拍构建了安全帽佩戴检测数据集。为验证本文算法各部分模块的有效性，进行了详细的消融实验。与当前主流的目标检测算法 Faster R-CNN、YOLOv4、YOLOv5 以及 YOLOX 进行了详细的对比实验。接下来，本章将对数据集以及实验进行详细的介绍。

3.1 数据集

本文使用某基建现场实拍的数据以及网络得到的数据，融合了 SHWD 而成的数据集。为缓解 SHWD 数据集中正负样本差距过大数据不平衡的问题，本文只融合了部分 SHWD 数据集并将其命名为 helmet 数据集。数据集共有 8496 张图片，其中涵盖了 14 785 个规范佩戴安全帽的对象和 44 196 个未佩戴安全帽的对象。本文将数据集以 8:1:1 的比例随机分成训练集、测试集和验证集。

3.2 评估指标

在安全帽佩戴检测模型的实验中，本文选取精度（precision）、召回率（recall）、平均精度均值（mAP）以及 FPS 作为评估指标。其计算方法为：

$$P = \frac{T_P}{T_P + F_P} \tag{5}$$

$$R = \frac{T_P}{T_P + F_N} \tag{6}$$

$$mAP = \frac{\sum_{i=0}^{n} AP(i)}{n} \tag{7}$$

$$FPS = \frac{1}{T} \tag{8}$$

式中：P 表示在所有预测为正类的样本中，原本就为正类的样本所占的百分比，T_P 代表真正例数，F_P 代表假正例数；召回率（recall）是指在所有目标中被无误地检测出的比例，其中 F_N 是指未被模型成功检测的真实目标数量。平均精度 mAP 是对检测器性能进行综合评估的指标。每个类别的 AP 值即为其精度和召回率之间的面积，计算所有类别的 AP 值再取其平均值即是 mAP 值；FPS 衡量算法处理视频帧的速度，即每秒钟可以处理的视频帧数量，T 表示处理一帧所需要的时间，单位为秒。

3.3 实验配置及相关参数设置

本文采用 PyTorch 作为安全帽佩戴检测模型的深度学习框架。实验环境的基本配置如表 1 所示。

表 1 实验环境配置表

名称	版本
操作系统	Windows 10
CPU	i5-12490F
GPU	NVIDIA 3060Ti
RAM	16 GB
CUDA	CUDAv11.0 CuDNNv8.0.5.39
Python	3.7
PyTorch	1.7.1
torchvision	0.8.2

模型训练所用参数如表 2 所示。

表 2 模型训练参数表

参数	值
Input_shape	[640×640]
Batch_size	8
epochs	300
Init_lr	0.01(SGD)
Weight_decay	0.000 5
lr_decay_type	cos
Num_workers	2
Momentum	0.937

3.4 训练过程

通过 300 个 epoch 的训练过程后，得到如下所示的详细结果。图 9 为损失函数图。训练损失和验证损失在图中使用实线表示，二者的平滑曲线则使用虚线表示。在训练 250 个 epoch 以后，训练损失在 2.90 左右收敛，验证损失在 2.85 左右收敛。网络收敛情况良好，训练成功。

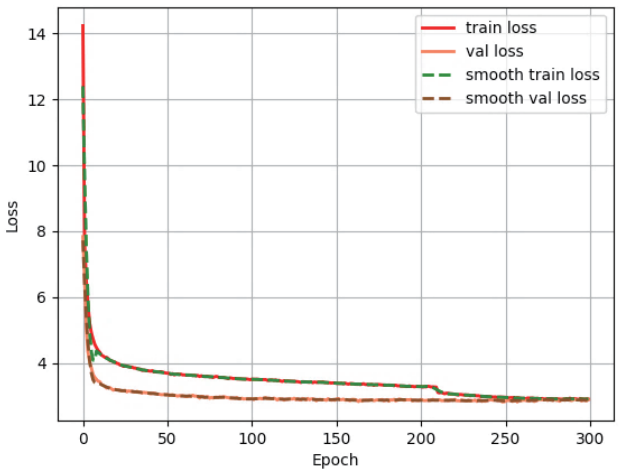

图 9 训练过程

本文模型部分识别效果如图 10 所示。

图 10 识别结果

3.5 消融实验

本文算法在 YOLOX 的基础上，将 Csp Darknet 的最后一个特征层使用 swin transformer 块进行了替换，且在 neck 的部分使用了 BiFPN 特征融合网络。最后在 neck 和 YOLO head 的连接之处加入了 CBAM 注意力模块。为了验证模型中每个改进点给模型性能造成的影响，本文进行了彻底的消融实验。实验结果如表 3。其中 P^h 和 P^p 分别代表 hat 类型和 person 类型的精度，R^h 和 R^p 分别代表 hat 类型和 person 类型的召回率。精度和召回率的门限值皆为 0.5。

表 3 消融实验

swin Csp	BiFPN	CBAM	P^h/%	P^p/%	R^h/%	R^p/%	mAP @0.5	FPS
—	—	—	94.58	92.51	88.60	89.82	93.50	90.41
	—	—	95.27	93.26	90.51	91.02	94.56	83.25
—	√	—	94.88	92.69	90.31	91.27	94.21	102.31
—	—	√	94.33	93.31	91.20	90.38	93.98	88.96
√	√	—	94.89	93.30	90.29	91.32	94.72	98.39
√	—	√	95.31	92.89	90.98	91.12	94.92	80.63
—	√	√	95.10	92.55	91.32	90.69	94.79	96.72
√	√	√	95.57	92.29	91.76	91.52	95.29	87.83

根据消融实验可知，初始 YOLOX-s 算法的 mAP 为 93.50%，FPS 为 90.41。在分别加入了 swin Csp、BiFPN 和 CBAM 后，mAP 分别获得了 1.06%、0.71%、0.48% 的提升，其中 swin Csp 和 CBAM 会使 FPS 稍有下降，但是 BiFPN 提升了 11.90 的 FPS。在将这三者中的两个加入 YOLOX 后，都在 mAP 指标上获得了提升。三者都加入之后，在 YOLOX-s 的基础上，mAP 获得了 1.79% 的提升。FPS 虽然比原算法低 2.58，但是也充分满足实时检测的需求。

3.6 对比实验

本文与常用的目标检测算法进行了对比实验，以检验算法的有效性。本文选取了 Faster R-CNN、YOLOv4、YOLOv5、YOLOX 与本文算法在 helmet 数据集上进行对比实验。对照实验的参数全部使用默认值，同时采用官方预设的预训练权重。实验结果如表 4。其中最好结果的字体做了加粗处理，精度和召回率的门限值都为 0.5。

表 4 对比实验

方法	P^h/%	P^p/%	R^h/%	R^p/%	mAP@0.5	FPS
Faster R-CNN	67.56	51.88	88.82	83.39	82.92	23.13
YOLOv4	94.63	80.34	77.47	72.22	82.94	45.04
YOLOv5s	93.47	**93.32**	88.64	74.50	90.67	**94.71**
YOLOXs	94.58	92.51	88.60	89.82	93.50	90.41
本文算法	**95.57**	92.29	**91.76**	**91.52**	**95.29**	87.83

4 总结

本文提出了一种改进 YOLOX 的安全帽佩戴检测算法，在原 YOLOX 的卷积神经网络中加入 swin transformer，以提高网络的特征提取能力。特征融合网络使用轻量化后的三层 BiFPN，以对特征金字塔的多尺度信息进行充分利用，再利用双向连接的方法融合增强多个尺度的特征图。把 CBAM 注意力模块加入 neck 和 YOLO head 的连接部分，以同时学习通道注意力和空间注意力权重，自适应地调整特征图中的不同通道和空间位置，从而更准确地捕获到输入特征图中的关键信息，以提高模型的性能及鲁棒性。后续考虑在安全帽检测场景下使用更轻量化的模型部署到嵌入式设备，以解决硬件资源紧张的问题，进一步提升模型的实用性。

参考文献：

[1] 乔炎，甄彤，李智慧 . 改进 YOLOv5 的安全帽佩戴检测算法 [J]. 计算机工程与应用，2023, 59(11): 203-211.

[2] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [J].IEEE transactions on pattern analysis and machine intelligence, 2017, 39(6): 1137-1149.

[3] 陈科圻，朱志亮，邓小明，等 . 多尺度目标检测的深度学习研究综述 . 软件学报，2021, 32(4): 1201-1227.

[4] 吴冬梅，闫宗亮，宋婉莹，等 . 改进 YOLOv4 的安全帽佩戴检测及身份识别方法 [J]. 计算机仿真 ,2022,39(12):290-293+377.

[5] 石家玮，杨莉琼，方艳红，等 . 基于改进 YOLOv4 的安全帽佩戴检测方法 [J]. 计算机工程与设计，2023,44(2):518-525.

[6] 苏文俊，张学军，许先富，等 . 改进 YOLOv4 的人脸口罩检测与硬件加速 [J]. 计算机工程与设计，2023,44(3):798-806.

[7] 韩贵金，王瑞萱，徐午言，等 . 改进 YOLOX 的夜间安全帽检测算法 [J/OL]. 计算机工程与应用 : 1-10 [2023-08-29]. http://kns.cnki.net/kcms/detail/11.2127.TP.20230626.1846.020.html.

[8] GE Z, LIU S, WANG F, et al. Yolox: Exceeding yolo series in 2021[EB/OL].(2021-07-18)[2023-10-02]. https://arxiv.org/abs/2107.08430.

[9] WU Y, CHEN Y, YUAN L, et al. Rethinking classification and localization for object detection[C] //Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).Piscataway:IEEE,2020: 10183-10192.

[10] LAW H, DENG J J I J O C V. CornerNet: detecting objects as paired keypoints [C]//European Conference on Computer Vision.Berlin:Springer,2018:765-781.

[11] TIAN Z, SHEN C, CHEN H, et al. FCOS: fully convolutional one-stage object detection[C] //Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV).Piscataway:IEEE, 2019:9626-9635.

[12] LIN T Y, PIOTR D, ROSS G, et al. Feature pyramid networks for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition: CVPR 2017. Piscataway: IEEE,2017:1-751.

[13] LIU S, QI L, QIN H F, et al. Path aggregation network for instance segmentation[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition.Piscataway: IEEE, 2018: 8759-8768.

[14] TAN M, PANG R, LE Q V. Efficientdet: scalable and efficient object detection[C]//Proceedings of the Proceedings of the IEEE/CVF conference on computer vision and pattern recognition.Piscataway:IEEE,2020:10778-10787.

[15] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C]//Computer vision-ECCV 2018: 15th European Conference.Berlin:Springer,2018:3-19.

【作者简介】

熊伟（1976—），男，辽宁锦州人，博士，高级工程师，研究方向：计算机网络、机器学习、自然语言处理。

张中天（2000—），通信作者（email：zhangzhongtian@qq.com），男，河南安阳人，硕士，研究方向：计算机视觉、自然语言处理。

路鑫（2000—），男，河南新乡人，硕士，研究方向：计算机视觉。

和宝同（1997—），男，河北保定人，硕士，研究方向：计算机视觉。

（收稿日期：2023-11-16）