光电技术及应用

DOI：10.16818/j.issn1001－5868.20241008002

基于CCL-YOLO的洗手动作检测算法

陈昌川[1*]，周鑫蔚[1]，龙虹毓[2]，郭中原[3]，祝　鹤[4]

（1. 重庆邮电大学 通信与信息工程学院，重庆 400065；2. 重庆邮电大学 自动化学院，重庆 400065；3. 西南大学 电子信息工程学院，重庆 400715；4. 重庆邮电大学 国际学院，重庆 400065）

摘　要： 针对现有YOLOv7模型在洗手动作检测中存在的检测精度低、抗环境干扰能力弱、相似动作区分度低等问题，提出一种基于改进YOLOv7的CCL-YOLO目标检测算法。该算法通过引入增强型轴向局部注意力机制，有效增强模型对长距离上下文信息的捕捉能力；采用CARAFE算子代替原算法的最近邻插值上采样方法，实现更高效的内容感知与特征重组，在并未增加模型参数的情况下显著提高对动作检测的精度；通过将SPPCSPC模块优化为SPPFCSPC结构，在保持相同感受野的前提下，检测精度提高了2.9%，帧速率提高了10；设计轻量级自适应解耦检测头代替传统耦合检测头，有效缓解了检测任务的训练难度，在损失2%检测精度的情况下，召回率提高了7.6%，mAP@0.5值提高了2%。在自制数据集上进行对比实验，结果表明改进算法平均精度mAP@0.5达到81.2%，相比于YOLOv7算法检测精度提升7.2%，精确率和召回率分别提高2.9%和11%，能有效满足实际洗手动作检测需求。

关键词： 目标检测；YOLOv7；特征融合

中图分类号： TP391.41　**文章编号：** 1001－5868(2025)01－0157－10

Hand-Washing Action Detection Algorithm Based on CCL-YOLO

CHEN Changchuan[1], ZHOU Xinwei[1], LONG Hongyu[2], GUO Zhongyuan[3], ZHU He[4]

(1. School of Communications and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, CHN; 2. School of Automation, Chongqing University of Posts and Telecommunications, Chongqing 400065, CHN; 3. School of Electronic and Information Engineering, Southwest University, Chongqing 400715, CHN; 4. International College, Chongqing University of Posts and Telecommunications, Chongqing 400065, CHN)

Abstract: To address the limitations of existing YOLOv7 models in handwashing action detection, including low detection accuracy, weak environmental interference resistance, and insufficient discrimination of similar actions, this paper proposes an enhanced CCL-YOLO object detection algorithm based on improved YOLOv7. The proposed algorithm introduces three key innovations: (1) An Enhanced Axial Local Attention mechanism is incorporated to strengthen the model's capability in capturing long-range contextual dependencies; (2) The CARAFE operator replaces conventional nearest-neighbor interpolation for upsampling, enabling more effective content-aware feature reorganization without increasing model parameters; (3) Structural optimization from SPPCSPC to SPPFCSPC improves detection accuracy by 2.9% and frame rate by 10 while maintaining equivalent receptive fields. Additionally, a lightweight adaptive decoupled detection head is designed to replace traditional coupled detection heads, achieving a 7.6% recall improvement and

收稿日期：2024－10－08.

基金项目：重庆市教委项目(KJQN202200627)；陕西省重点研发计划项目(2023-YBNY-222).

*通信作者：陈昌川

2% mAP@0.5 enhancement at the cost of only 2% precision reduction. Experimental results on a custom dataset demonstrate that the improved algorithm achieves 81.2% mAP@0.5, representing a 7.2% accuracy improvement over baseline YOLOv7, with precision and recall rates increased by 2.9% and 11% respectively. The proposed method effectively meets practical requirements for real-world handwashing action detection while maintaining computational efficiency.

Key words: object detection; YOLOv7; feature fusion

0 引言

手部卫生对于防止细菌和病毒传播极为重要，它可以避免细菌和病毒从污染物表面传播到口、鼻或眼睛粘膜。根据国家卫生健康委员会2020年发布的《医务人员手卫生规范》(WS/T 313-2020)，医务人员手卫生操作标准包含两个技术体系：(1)卫生洗手七步法；(2)外科冲洗手消毒法。其主要内容包括：掌心相对搓揉；手指交叉，掌心对手背搓揉；手指交叉，掌心对掌心搓揉；双手互握，相互揉搓指背；拇指在掌中转动揉搓；指尖在掌心揉搓；旋转揉搓腕部直至肘部。目前，大部分场所仍采取人工监管的方式，存在监管不足、人工成本高等问题。随着深度学习的发展，对人类或动物行为的检测逐渐深入，洗手行为识别的理论基础也越来越完善。洗手动作识别具有较高的实用性和可行性，与人们日常生活息息相关，在医疗卫生、疾病防控、个人健康管理等多个领域都有广泛需求。

Li等[1]设计了一种腕带式传感平台WristWash，整个系统由电池供电，集成了惯性测量单元和基于隐藏马尔可夫模型的分析方法，使用了三轴加速度计和三轴陀螺仪采集数据，通过I^2C通信协议连接MCU将传感器数据存储在SD卡上，能根据世界卫生组织的建议自动评估洗手程序。Zhang等[2]提出一种使用可穿戴设备和机器学习的算法检测洗手动作，通过具有高斯核的SVM模型对加速度计和陀螺仪数据进行分类，但对拇指在掌中搓揉等特定步骤的分类结果较低。Ivanovs等[3]使用两种预训练网络：MobileNetV2[4]和Xception[5]，在自制数据集上对不同洗手任务进行分类，准确率分别达64%和67%。Yue[6]等提出基于YOLOv3-tiny[7-8]模型的手势识别技术成功识别和检测卫生部规定洗手法的几个步骤，使用TensorRT加速训练，并将其部署到NVIDIA的Jetson Nano智能移动终端上运行，实时监测帧率达到35 FPS以上。Wang等[9]提出一种基于三维卷积神经网络来提取洗手视频数据时间和空间特征的方法，结合概率平滑方法来优化洗手步骤的识别。Llorca等[10]针对测量洗手动作提出一种稳定的手部分割系统，将视频序列中提取的图像描述符和从光流中提取的运动描述符整合创建为特征向量，使用两个独立的向量机进行一对一的多类别分类，该方法对不同的光照、遮挡、反射及水槽表面颜色变化等具有较高的鲁棒性。

根据上述分析，基于深度学习的洗手动作识别比使用可穿戴设备的识别方法具有更高的灵活性和准确性。由于光照变化、手部区域在图像中所占比例小、水槽表面颜色变化等环境噪声的影响，以及不同洗手动作之间像素级差异较小，传统检测网络在识别不同洗手动作时效果不佳。本文基于YOLOv7[11]算法，引入Contextual Transformer Networks[12]的上下文信息提取模块，将其添加到BackBone部分的ELAN模块以增强信息表征能力。此外，特征融合网络使用了CARAFE[13]轻量级上采样算子，提高特征提取能力。使用SPPFCSPC替代SPPCSPC，减少模型的参数量，所获得的感受野相同且速度更快。采用轻量级非对称多级压缩检测头(LADH-Dectect)，对不同类别的特征进行非对称压缩，在保持轻量化的同时获得更高效的检测性能。改进后的CCL-YOLOv7在Hand-Washing Detection数据集中可进一步提高目标检测精度。

1 YOLOv7网络

YOLOv7是一种具有高精度、易训练、易部署的目标检测网络，其网络结构如图1所示。主要包括三部分，分别是Input(输入端)、BackBone(主干网络)和Head(检测头)。输入端负责接收并预处理图像，主干网络用于特征提取，而检测头则进行最终的预测。主干网络包含CBS、ELAN模块和MP结构，CBS通过在特征图上执行卷积、批量归一化(Batch Normalization，BN)和激活函数操作，有效捕捉不同尺度的图像特征；ELAN模块由多种卷积组成，在不破坏原有梯度路径的前提下，增强了网络

学习能力。头部网络包括 SPPCSPC、CBS、MP、UPSample 和三个后续的 RepConv 模块：SPPCSPC 模块通过最大池化来获得不同感受野，使得算法适应不同分辨率图像；RepConv 模块在训练阶段和推理阶段拥有不同的结构，在训练阶段中上支路的 3×3 卷积用于特征提取，中间支路的 1×1 用于平滑特征，下支路不做卷积操作，而若其作为推理模块则包含一个步长为 1 的 3×3 卷积。

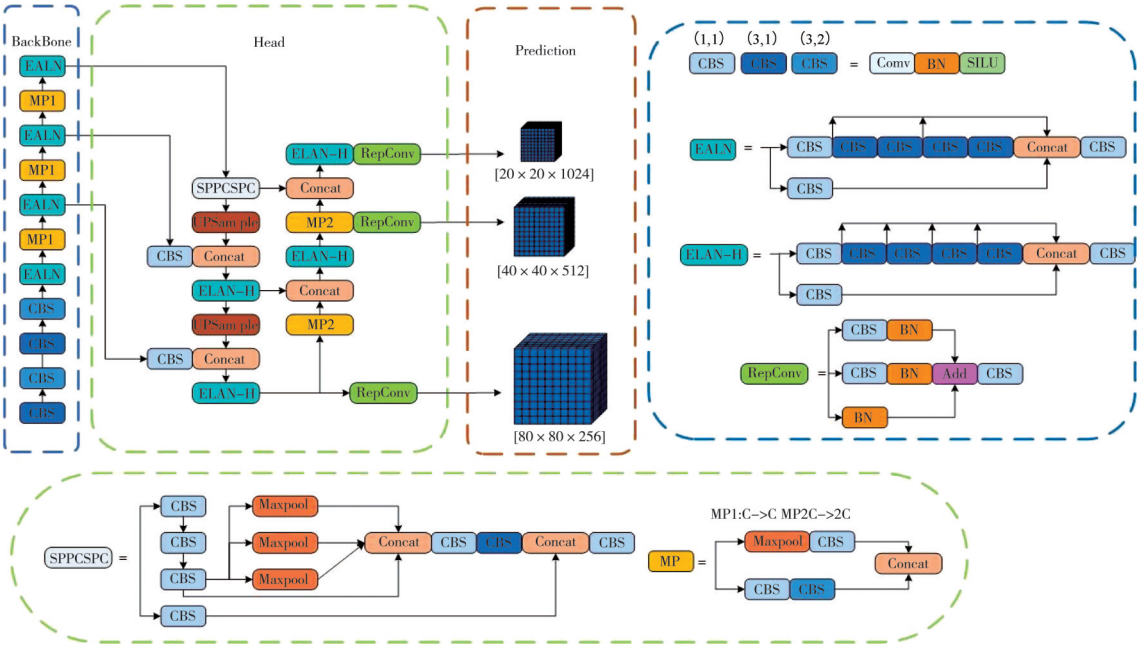


图 1　YOLOv7 网络结构示意图

2　CCL-YOLOv7 网络

基于 YOLOv7 算法设计出一种基于上下文信息的注意力机制及轻量化上采样与检测头的 CCL-YOLOv7 网络，其结构如图 2 所示。

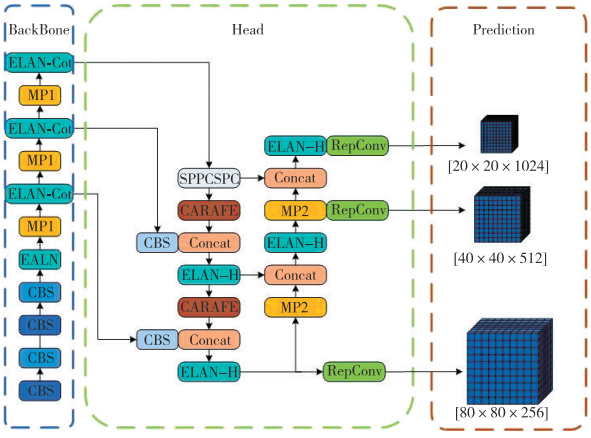
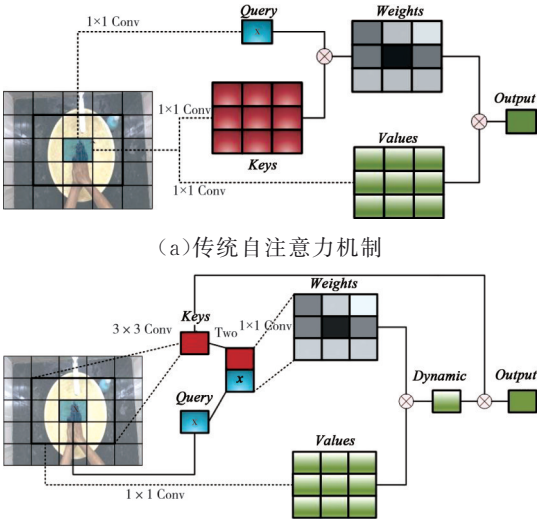


图 2　CCL-YOLOv7 结构示意图

2.1　CotAttention 自注意力机制

近年来，Transformer 在自然语言处理（Natural Language Processing，NLP）任务中的表现出色。受 Transformer 的启发，研究者将卷积神经网络（CNN）架构与 Transformer 风格的模块相结合，不断推动计算机视觉（CV）任务的发展。然而，大多数使用自注意力的 Transformer 是在二维特征图上基于空间位置得到注意力矩阵，未能充分利用丰富的上下文信息。本文采用上下文 Transformer 网络（CotContextual Transformer Networks，CoTNet）中引入的一种自注意力机制——CotAttention。传统注意力机制与 CotAttention 的对比如图 3 所示。



（a）传统自注意力机制

（b）CotAttention

图 3　传统自注意力机制与 CotAttention 任务处理流程

CotAttention将输入图像划分为若干网格，相邻三个网格的特征向量通过一个3×3卷积，生成静态表示特征向量***Key***，表示特征向量之间的静态上下文特征。基于此，将***Key***与查询特征向量***Query***级联，并输入两个连续的1×1卷积以生成注意力矩阵。该过程利用了每个***Query***和***Key***之间的关系以增强自注意力学习能力。随后，利用注意力矩阵聚合所有的物体本身的特征向量***Values***，得到输入的动态表示特征向量。最后，将静态特征向量和动态特征向量的组合作为CotAttention的输出。

CotAttention的结构如图4所示。通道数为C，高度为H，宽度为W的二维特征图X，通过三个转换矩阵$\boldsymbol{W}_q$，$\boldsymbol{W}_k$，$\boldsymbol{W}_v$，输出为编码矩阵$\boldsymbol{K}^1=X\boldsymbol{W}_k$，查询矩阵$\boldsymbol{Q}=X\boldsymbol{W}_q$和值矩阵$\boldsymbol{V}=X\boldsymbol{W}_v$。其中，$\boldsymbol{W}_q$是单位矩阵，$\boldsymbol{W}_k$表示$k\times k$的卷积，$\boldsymbol{W}_v$表示1×1的卷积。$X$在空间上采用了$k\times k$网格，以便$\boldsymbol{K}^1$从临近像素捕获重要的静态上下文信息。随后，将$\boldsymbol{K}^1$和$\boldsymbol{Q}$按通道维度拼接后，通过两个连续的1×1卷积得到注意力矩阵。如下式所示：

$$\boldsymbol{A}=[\boldsymbol{K}^1,\boldsymbol{Q}]\boldsymbol{W}_r\boldsymbol{W}_n \tag{1}$$

式中，$\boldsymbol{W}_r$表示带有激活函数的1×1卷积，$\boldsymbol{W}_n$表示不带不带激活函数的1×1卷积。利用矩阵$\boldsymbol{A}$和$\boldsymbol{V}$进行矩阵乘法，得到一个捕获了输入之间动态特征交互的矩阵$\boldsymbol{K}^2=\boldsymbol{A}\boldsymbol{V}$。模块的输出公式为：

$$\boldsymbol{Y}=\boldsymbol{K}+\boldsymbol{A}\boldsymbol{V} \tag{2}$$

经过多次对比，将CotAttention添加到主干网络的ELAN模块中，构建ELAN_CoTAttention模块，如图5所示，以获取更多的上下文信息，加强信息表征能力，达到更优的特征提取效果。

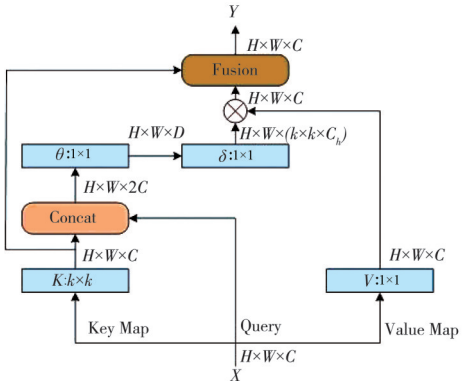


图4 CotAttention结构示意图

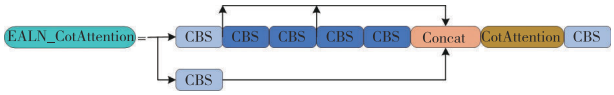


图5 ELAN_CoTAttention结构示意图

2.2 轻量级上采样算子CARAFE

在目标检测网络中，特征上采样是扩展特征图的重要操作，可实现跨不同层次特征的有效匹配。

YOLOv7原使用最近邻插值法进行上采样，通过计算像素之间的最近距离确定灰度值。然而，当计算最近邻值时，不同灰度值的相邻像素可能导致锯齿状边缘，丢失图像细节。为解决上述问题，引入了CARAFE上采样算子。该算子通过底层内容信息预测重组内核，并在预定义区域内重组特征。该机制包括两个主要部分，重组内核预测模块和内容感知重组模块，如图6所示。

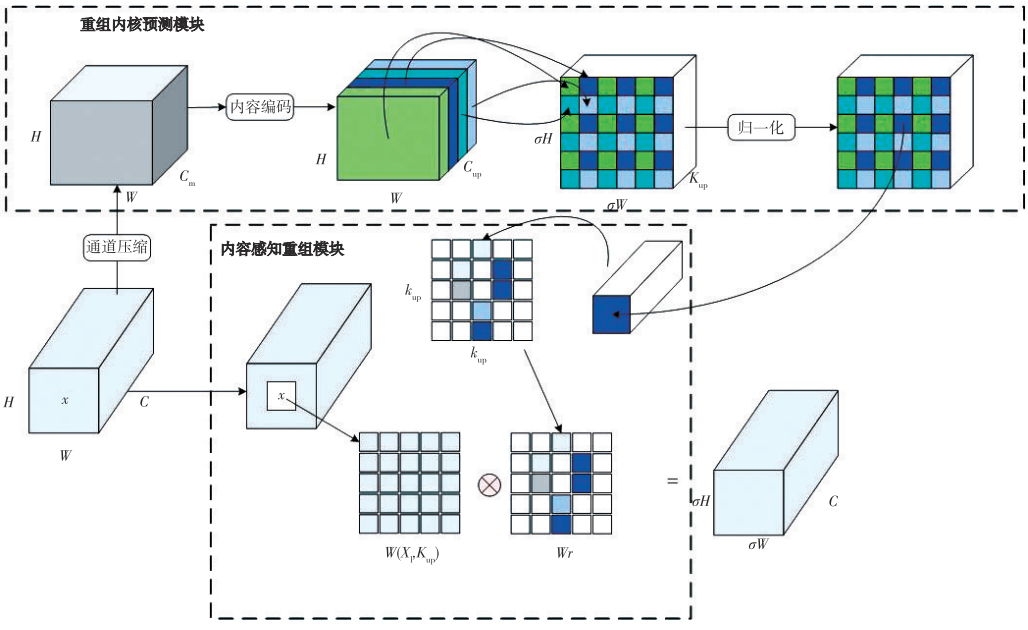


图6 CARAFE流程图

一个特征图X尺寸为$C\times H\times W$，给定为上采样率σ，CARAFE将输出$C\times\sigma H\times\sigma W$的新特征图$X'$。该特征图的任意位置$I'=(i',j')$，对应输入$X$的初始位置$I=(i,j)$，表示为：

$$i=[\frac{i'}{\sigma}] \tag{3}$$

$$j=[\frac{j'}{\sigma}] \tag{4}$$

$N(X_l,k)$定义为X以位置I为中心的$k\times k$区域。重组内核预测模块φ利用这个子区域计算每个目标位置I'的上采样核$W_{l'}$，再结合内容感知重组模块ϕ产生位置I'的上采样特征，表示为：

$$W_{l'}=\varphi(N(X_l,k_{\text{encoder}})) \tag{5}$$

$$X'_{l'}=\phi(N(X_l,k_{\text{up}}),W_{l'}) \tag{6}$$

重组内核预测模块φ包含通道压缩模块、内容编码模块和内核归一化模块。对于大小为$C\times H\times W$的特征图，1×1卷积核将输入通道数从C减少为C_{m}，以降低后续步骤的计算成本。随后，使用k_{encoder}大小的卷积进行内容编码，输出为$C_{\text{up}}\times H\times W$的特征图，其中$C_{\text{up}}$表示为：

$$C_{\text{up}}=\sigma^2k_{\text{up}}^2 \tag{7}$$

式中，k_{up}为重组卷积核的大小。最后在空间维度上对重组卷积核进行归一化处理，使卷积核权重和等于1。

内容感知重组模块ϕ对每一个重组内核$W_{l'}$进行加权求和运算，对于目标位置I'和以$I=(i,j)$为中心的相应正方形区域$N(X_l,K_{\text{up}})$。运算式为：

$$X'_{l'}=\sum_{n=-r}^{r}\sum_{n=-r}^{r}W_{l'(n,m)}\times X_{(i+n,j+\text{m})} \tag{8}$$

式中，$r=k_{\text{up}}/2$，$W_{l'}$为重组卷积核预测模块输出，X为输入特征。

所提出算法引入轻量级上采样算子CARAFE，通过其自适应上采样方法，促进了高效的内容感知与特征重组，显著提高了检测精度。

2.3　SPPFCSPC

He等[14]提出的空间金字塔池化（Spatial Pyramid Pooling，SPP）模块将经典的金字塔池化结构引入到卷积神经网络中，减少了对图相关重复特征的提取。在YOLOv5中，Glenn Jocher参考SPP的结构进行了改进，如图7所示，在相同目标的情况下减少了模型的计算量，速度远快于SPP结构，称为SPPF[15]。

YOLOv7中采用的SPPCSPC结构结合了SPP和跨阶段部分连接（Cross Stage Partial Connection，CSPC），显著提升了感受野范围，有助于检测不同大小的目标。其包含三个大小不同的最大池化层（Maxpool），有效避免了图像处理操作引起的失真问题，解决了神经网络提取重复特征问题。

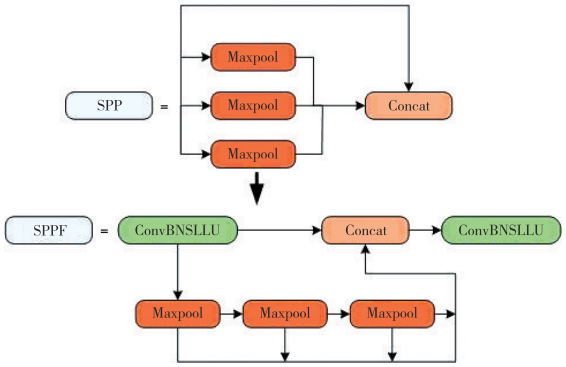


图7　SPP与SPPF结构示意图

按照SPPF的思路，对SPPCSPC改良为SPPFCSPC，结构如图8所示，将5×5，9×9，13×13并联的Maxpool改进为三个5×5串联的Maxpool。该结构整合了网络的深浅层特征并优化了梯度变化，牺牲了较小的速度并提高了模型的稳定性和准确性，减少了模型的计算量并获得了相同的感受野。

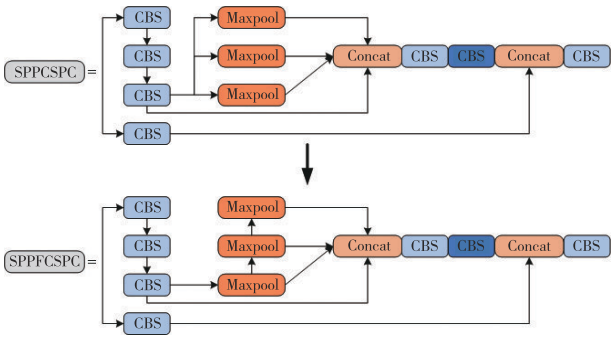


图8　SPPCSPC与SPPFCSPC结构示意图

2.4　LADH-Dectect

分类和回归之间的冲突在目标检测任务中普遍存在。在图像处理任务中，前者主要关注样本的纹理特点，而后者则更关注其边缘特征。随着YOLO系列的主干和特征金字塔结构的不断演化，它们的检测头仍是耦合的，在网络的顶部使用相同的卷积层同时处理分类和回归任务。

YOLOX[16]引入了解耦头部框架，旨在解决分类和回归任务中间的冲突，结构如图9所示。其通过一个1×1卷积降低通道维度，随后，使用两个并行的3×3卷积分别处理分类和回归任务，并在回归

中引入了IoU以指示每个特征点处的对象存在。虽然这显著提高了模型的检测精度，但这种结构大大增加了网络的参数数量，进而导致推理速度的降低。因为损失函数是正负样本都参与的损失计算，若采用解耦检测头计算反向传播的损失值，则可能导致损失分布不均匀，这种模型可能造成次优误差的反向传播。在洗手动作检测任务中，正负样本分布极不均匀，且图像中背景复杂，目标大小差别较大，这导致检测准确性降低。为了解决这些问题，根据非对称解耦头的思路，采用了一种轻量级非对称检测头（LADH-Head），其结构如图10所示。

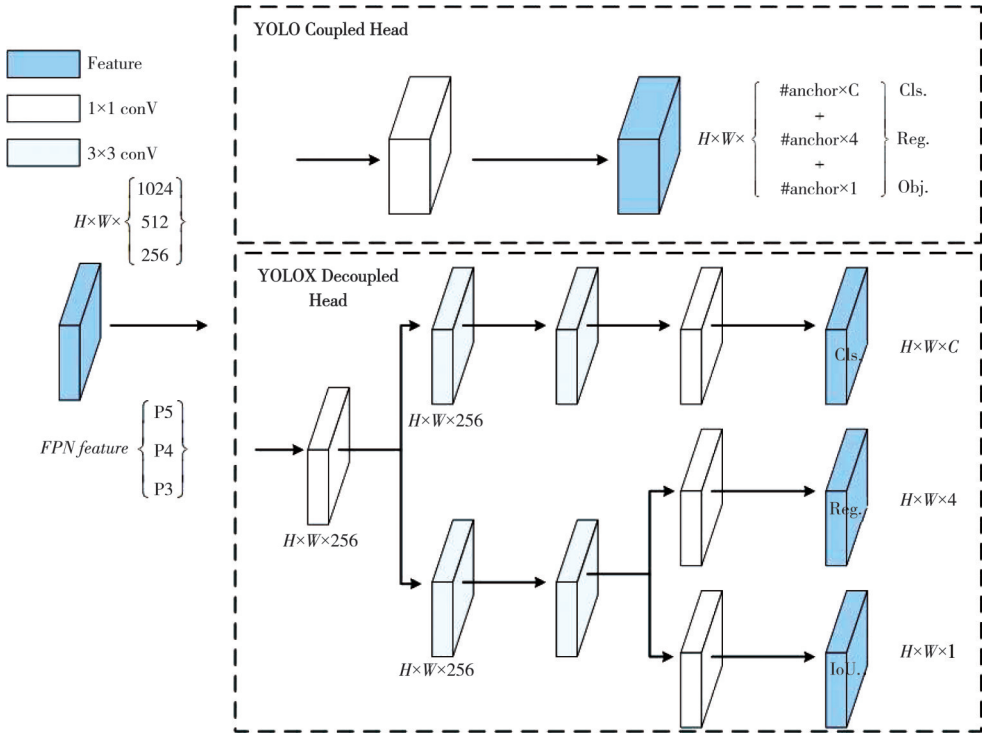


图9 YOLO耦合检测头与YOLOX解耦检测头结构示意图

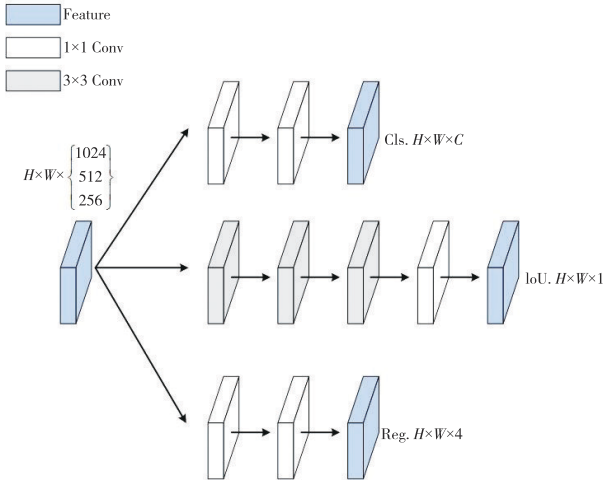


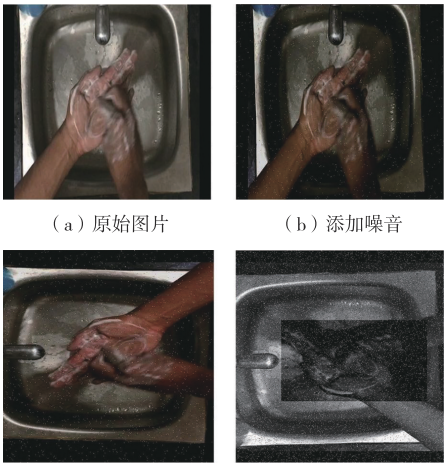
图10 LADH-Head结构示意图

LADH-Head利用三个独立的通道执行不同任务。该结构通过加深用于洗手动作识别的网络路径，并通过三个卷积层扩展感受野，从而缓解训练难度并提高模型性能。卷积层在通道维度上也进行压缩，减少了解耦头模块的参数，提高检测速度。

3 实验结果与分析

3.1 Hand-Washing Detection数据集

构建了一个新的Hand-Washing Detection数据集，用于洗手动作检测。鉴于背景环境、照明和肤色等条件不同，实验收集了292个单独洗手视频，总共3 504个片段，每个片段约在9 s。考虑到单次动作的多次重复，实验将原视频按照30帧的时长进行了切分，得到了11 433张720×480的图片帧集，再对该集内容进行筛选，删除了模糊和重复度较高的图片。为了扩充训练数据尽可能接近真实分布，从而提高检测精度，并对图片帧集中部分图片进行了裁剪、旋转、镜像、改变亮度、加噪声等数据增强操作。经增强后，数据集包含5289张图片，图11分别为原始的数据集图片和经过数据增强后的数据集图片。



（a）原始图片　（b）添加噪音　（c）旋转　（d）灰度化

图11　数据集样本

参照PASCAL VOC目标检测数据集标注的标准格式，采用Labelimg工具标注手部区域，标注了洗手中的七类洗手动作和一类无关动作或错误动作，数据集各动作类别分布如图12所示。训练集和测试集按照4∶1的比例随机分配，得到Hand-Washing Detection数据集。

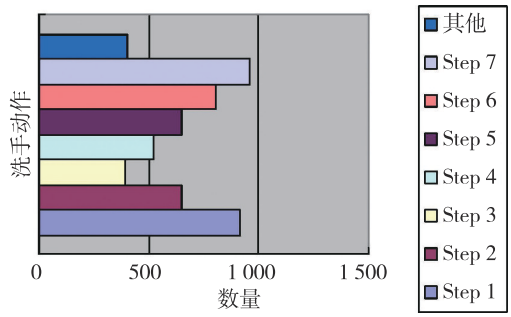


图12　数据集类别分布情况

3.2　实验环境与衡量指标

实验使用的环境参数如表1所示。训练迭代次数为150次，迭代批量大小设置为16。用优化器选择Adam。

表1　实验环境参数

参数	配置
操作系统	Windows 11 Enterprise LTSC
CPU	Intel Core i5-13400
GPU	GeForce RTX 4070(12 GB)
编程语言	Python3.8.11
深度学习框架	Pytorch
CUDA版本	Cuda12.1

为了便于评价模型改进后的性能，将精确率（P）、召回率（R）、平均精度（mAP）、Giga浮点运算每秒（$GFLOPs$）、权重文件大小作为的评价指标。

准确率P用于衡量在所有预测为正样本的结果中，预测正确的比率，其计算式为：

$$P=\frac{TP}{TP+FP} \tag{9}$$

召回率R是衡量模型在识别所有实际正样本中正确识别出的比例，计算式为：

$$R=\frac{TP}{TP+FN} \tag{10}$$

式中，TP表示在检测结果中预测正确的样本数量，FP表示在检测结果中预测错误的样本数量，FN表示所有正确目标中未被检测到的样本数量。

平均精度mAP是基于目标检测中m个类别各自的精度-召回曲线来计算的，能够较全面地反应网络的性能，其定义式如下：

$$mAP=\frac{1}{m}\sum_{i=1}^{m}\int_{0}^{1}P(R)\mathrm{d}R \tag{11}$$

$GFLOPs$与权重文件的大小是衡量模型复杂度的指标。$GFLOPs$与权重文件越小，表示模型所需要的算力就越小。由于数据集中非洗手动作的其他类别同属一个类别，特性相似度差，网络识别率极低，不具有参考意义。计算相关参数时取洗手七步法的算术平均值。

3.3　对比实验

3.3.1　不同模型对比分析

为验证本算法的先进性，使用Hand-Washing Detection数据集，选择与之相似目标检测网络进行训练及测试，包括YOLO系统网络，RTDER-l以及YOLOv7-ResNet101形成对比实验，其中YOLOv7+ResNet101是将YOLOv7的BackBone部分替换为ResNet网络而构建的，这样的对比实验有助于展示改进模型在相似类别中的优势和特点。相关实验结果如表2所示。

通过对比实验可知，除YOLOv6外的网络在mAP@0.5略有差异，但整体都在70%以上。YOLOv3有较高的准确率，但模型参数量较大，YOLOv5l的mAP@0.5优于其他模型，但在参数量上高于YOLOv7，YOLOv8l及RTDER-l网络，但在前向传播过程中需要更多的浮点运算数量。YOLOv6l的参数量和$GFLOPs$较高，但精确率和召回率却略低。YOLOv7的mAP@0.5、精确率及召回率略低于YOLOv5l和YOLOv3，但减少了模型的参数量与计算复杂度。RTDER-l精确率、召回率

与*mAP*@0.5均高于70%，且模型参数和计算复杂度都较低。所提出的改进网络模型在*mAP*@0.5和召回率上表现最佳，分别为81.2%和84.6%，尽管YOLOv7+ResNet101，RTDER-l在精确率上高于CCL-YOLOv7，但其*mAP*@0.5，召回率均低于改进网络。

表2 不同网络实验对比表

网络	*P*/%	*R*/%	*mAP*@0.5/%	*Params*/M	模型大小/MB	*GFLOPs*
YOLOv3[17]	67.7	75.0	75.1	103.7	202.9	282.2
YOLOv5l[18]	67.1	80.8	77.8	53.1	104.3	134.7
YOLOv6l[19]	55.0	72.9	67.1	110.9	217.0	391.2
YOLOv7	65.2	73.6	74.0	36.5	73.1	103.3
YOLOv8l[20]	64.5	69.6	72.7	43.6	85.6	164.8
RTDER-l[21]	71.8	76.2	75.7	32.0	64.6	103.5
YOLOv7+ResNet101[22]	69.5	69.2	75.7	66.7	132.3	165.3
本文算法	68.1	84.6	81.2	70.7	140.2	215.7

3.3.2 不同注意力机制对比分析

注意力机制能使模型更专注感兴趣的目标，提高权重聚焦重要信息。为验证CotAttention机制的有效性，对比不同注意力机制，匹配洗手动作检测任务中最佳的注意力机制设计方案，表3对几种常用注意力机制实验结果进行了统计，重点关注准确率、召回率等参数。

实验结果表明，CotAttention自注意力机制对于提升模型的准确率最为明显，其精确率为68%，仅低于CA注意力机制，召回率为77.5%，*mAP*@0.5为76.5%，均高于其他改进方案。

表3 不同注意力机制实验对比表

注意力机制	*P*/%	*R*/%	*mAP*@0.5/%
+CotAttention	68.0	77.5	76.5
+CBAMA	56.3	68.9	61.9
+SE	64.8	76.7	75.4
+CA	69.9	73.5	74.2
+ECA	68.5	74.1	75.2
+SimAM	67.5	73.8	75.6

3.4 消融实验

为了验证模型各项改进的有效性，以YOLOv7模型为Baseline进行不同改进的对比实验，训练数据集为Hand-Washing Detection数据集，表4给出了相关实验结果。

表4 消融实验对比

方法	*P*/%	*R*/%	*mAP*@0.5%	*Params*/M	*FPS*/(f·s^{-1})
Baseline	65.2	73.6	74.0	36.5	144.9
Cot	68.0	77.5	76.5	57.2	144.5
Cot + CARAFE	67.0	80.8	77.8	57.3	137.2
Cot + CARAFE + SPPFCSPC	70.1	77.0	79.2	57.3	147.2
Cot + CARAFE + SPPFCSPC + LADH-Dectect	68.1	84.6	81.2	70.7	126.5

实验结果表明，原始的YOLOv7的精确率为65.2%，召回率为73.6%，*mAP*@0.5为74%。以此为基础，将ELAN改进为ELAN_Cotattation模块，精确率、召回率、*mAP*分别提高了2.9%，3.5%，2.5%，增强了网络模型对图像上下文特征的关注度，有效提高了网络模型的特征提取能力。加入CARAFE上采用算子，参数量基本不变而指标有一定提升，但*FPS*降低了10 f/s，这表明CARAFE算子有效增加了模型的感受域，网络取得质量更高的上采样特征图，提升了模型的检测精度，但也一定程度上增大了模型推理时延。将YOLOv7模型中的SPPCSPC替换为SPPFCSPC模块后，精确率提高了2.9%，*mAP*提升了1.4%，但召回率降低了3.8%，*FPS*提升了10 f/s，在基本不影响网络参数的情况下保留不同阶段的特征信息，提高模型的精度与检测速度。将检测头更换为LADH-Dectect后，精确率虽然降低了2%，但召回率和*mAP*分别提高了8.6%与2%。CCL-YOLOv7与YOLOv7网络相比，*mAP*提高7.2%，精确率提高2.9%，召回率提高11%，网络参数量增加了34.2 M，*FPS*降低了18.4 f/s，但也能满足目标的实时检测需求。

图13显示了所提出模型与YOLOv7模型的部分预测结果对比，结果表明，在相同检测动作中，本文模型识别精度较高。



（a） YOLOv7　　（b） CCL-YOLOv7

图13　预测结果对比

4　结论

本文提出了一种基于改进YOLOv7的洗手动作识别模型CCL-YOLOv7，该模型首先引入了CotNet中的自注意力机制，更注重数据的上下文信息。其次采用了CARAFE上采样代替原始的最近邻插值上采样方法，以进一步提高网络的特征融合能力。面对外部复杂环境，在模型参数增强较少的情况下有效降低了模型漏检率。随后将YOLOv7中SPPCSPC改进为SPPFCSPC模块，提高了检测速度。最后提出了一种轻量化非对称多级压缩检测头LADH-Dectect，加深对象检测任务的网络路径，提高了模型的训练速度。通过在自制的Hand-Washing Detection数据集上进行对比试验，验证了相关改进的有效性。所提出方法在精确率、召回率及$mAP@0.5$相较于原始网络分别提升了2.9%，11%，7.2%，在满足实时性要求的延迟下获得了较高的识别精度。

参考文献：

[1] Li H, Chawla S, Li R, et al. Wristwash: towards automatic handwashing assessment using a wristworn device [C]// Proceedings of the 2018 ACM International Symposium on Wearable Computers, 2018: 132-139.

[2] Zhang Y, Xue T, Liu Z, et al. Detecting hand washing activity among activities of daily living and classification of WHO hand washing techniques using wearable devices and machine learning algorithms [J]. Healthcare Technology Letters, 2021, 8(6): 148-159.

[3] Ivanovs M, Kadikis R, Lulla M, et al. Automated quality assessment of hand washing using deep learning [J]. arXiv preprint arXiv: 2011.11383, 2020.

[4] Sandler M, Howard A, Zhu M, et al. Mobilenetv2: Inverted residuals and linear bottlenecks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 4510-4520.

[5] Chollet F. Xception: Deep learning with depthwise separable convolutions [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 1251-1258.

[6] Yue Z, Wang F, Liu Z. An intelligent hand-washing monitoring platform based on gesture recognition technology [C]// 2021 China Automation Congress (CAC), 2021: 940-944.

[7] Fu L, Feng Y, Wu J, et al. Fast and accurate detection of kiwifruit in orchard using improved YOLOv3-tiny model[J]. Precision Agriculture, 2021, 22: 754-776.

[8] 韩晓良，陈佳昌，周伟松. 基于3D注意力的MobileNet图像分类算法改进[J]. 重庆邮电大学学报(自然科学版)，2023，35(3)：513-519.
Han Xiaoliang, Chen Jiachang, Zhou Weisong. Improvement of MobileNet image classification algorithm based on 3D attention [J]. Journal of Chongqing University of Posts and Telecommunications: Natural Science Edition, 2023, 35(3): 513-519.

[9] Wang T, Xia J, Wu T, et al. Handwashing quality assessment via deep learning: A modelling study for monitoring compliance and standards in hospitals and communities [J]. Intelligent Medicine, 2022, 2(3): 152-160.

[10] Llorca D F, Parra I, So telo M Á, et al. A vision-based system for automatic hand washing quality assessment [J]. Machine Vision and Applications, 2011, 22(2): 219-234.

[11] Wang C Y, Bochkovskiy A, Liao H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 7464-7475.

[12] Li Y, Yao T, Pan Y, et al. Contextual transformer networks for visual recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45(2): 1489-1500.

[13] Wang J, Chen K, Xu R, et al. Carafe: Content-aware reassembly of features [C]// Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 3007-3016.

[14] Kaiming H, Xiangyu Z, Shaoqing R, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.

[15] Glenn Jocher. YOLOv5: A state-of-the-art object detection and segmentation model [EB/OL]. (2022-11-22) [2024-8-10].

[16] Kai Y, Yu Z, Xin Z, et al. YOLOX with CBAM for insulator detection in transmission lines[J]. Multimedia Tools and Applications, 2023, 83(14): 43419-43437.

[17] 夏　英，黄秉坤. 采用改进YOLOv3的高分辨率遥感图像

目标检测[J]. 重庆邮电大学学报(自然科学版), 2022, 34(3): 383-392.
Xia Ying, Huang Bingkun. Object detection in high-resolution remote sensing images using improved YOLOv3 [J]. Journal of Chongqing University of Posts and Telecommunications: Natural Science Edition, 2022, 34(3): 383-392.

[18] Li G, Fu L, Gao C, et al. Multi-class detection of kiwifruit flower and its distribution identification in orchard based on YOLOv5l and Euclidean distance [J]. Computers and Electronics in Agriculture, 2022, 201: 107342.

[19] Li C, Li L, Jiang H, et al. YOLOv6: A single-stage object detection framework for industrial applications [J]. arXiv preprint arXiv: 2209.02976, 2022.

[20] Wang G, Chen Y, An P, et al. UAV-YOLOv8: A small-object-detection model based on improved YOLOv8 for UAV aerial photography scenarios [J]. Sensors, 2023, 23(16): 7190.

[21] Zhao Y, Lv W, Xu S, et al. Detrs beat YOLOs on real-time object detection [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 16965-16974.

[22] Narayana C L, Ramana K V. An efficient real-time weed detection technique using YOLOv7[J]. International Journal of Advanced Computer Science and Applications, 2023, 14(2): 550-556.

作者简介:

陈昌川(1978—),男,四川省广安市人,副教授,研究领域为智能信息处理、图像处理与人工智能、特高频局放检测、红外成像与测温。