文章编号:1671-4598(2025)10-0225-10 DOI:10.16526/j.cnki.11-4762/tp.2025.10.029 中图分类号:TP389.1 文献标识码:A

基于 YOLOv8 的轻量级自适应权重手势识别模型

徐慕君，葛　艳，时东亮
（青岛科技大学 信息科学技术学院，山东 青岛　266000）

摘要：近年来，针对手势识别模型的轻量化研究层出不穷，但常以牺牲识别精度为代价；为此，提出一种基于YOLOv8模型的轻量级手势识别模型LAW-YOLO；设计轻量级自适应权重卷积LAWC并引入特征提取网络，通过自适应权重机制在复杂背景中精准聚焦手势区域，并有效降低计算成本；采用双向特征金字塔网络BiFPN优化模型特征融合网络并减少冗余通道，在尽可能小的精度损失下大幅减少模型参数量和计算量；使用融入了自校正卷积SC-Conv的自校正模块SC2f替换检测头前的C2f模块，增强模型特征融合网络性能，弥补轻量化带来的精度损失；改进模型在HaGRID手势数据集上验证，改进后模型参数量和模型大小较改进前分别减少41.20%和41.94%，识别精度高达98.63%，显著提升模型计算效率和识别精度。

关键词：手势图像识别；YOLOv8；轻量级网络；自适应权重；双向特征金字塔网络；自校准卷积

Model for Lightweight Adaptive-weight Gesture Recognition Based on YOLOv8

XU Mujun，GE Yan，SHI Dongliang
（School of Information Science & Technology，Qingdao University of Science and Technology，Qingdao　266000，China）

Abstract：In recent years，there is an increasing amount of research on lightweight gesture recognition models，but often at the cost of sacrificing recognition accuracy；Therefore，a lightweight gesture recognition model LAW-YOLO based on the YOLOv8 model is proposed；Firstly，a lightweight adaptive weight convolution（LAWC）is designed，and a feature extraction network is introduced to accurately focus gesture regions in complex backgrounds through an adaptive weight mechanism，effectively reducing computational costs；Secondly，a bidirectional feature pyramid network（BiFPN）is used to optimize the model feature fusion network and reduce redundant channels，significantly reducing the number of model parameters and computation complexity with minimal accuracy loss；Finally，the self-calibrated module（SC2f）incorporating self-calibrated convolution（SC-Conv）is used to replace the C2f module in front of the detection head，enhancing the performance of the model feature fusion network and compensating for the accuracy loss caused by lightweight design；Through verification on the HaGRID gesture dataset，the results show that compared with the model before the improvement，the improved model reduces the number of model parameters and model size by 41.20% and 41.94%，respectively，with a recognition accuracy of 98.63%，significantly improving the computational efficiency and recognition accuracy of the model.

Keywords：gesture image recognition；YOLOV8；lightweight network；adaptive-weight；BiFPN；SC-Conv

0　引言

手势识别技术的发展在计算机视觉和人机交互领域取得巨大的突破，为我们提供与计算机系统更加自然和直观的交互方式。这一技术在虚拟现实、智能设备、辅助技术等领域具有广泛的应用前景。然而，随着对轻量

收稿日期：2024-09-08；　**修回日期**：2024-10-21。
作者简介：徐慕君(1999-)，男，硕士研究生。
通讯作者：葛　艳(1976-)，女，博士，副教授。
引用格式：徐慕君，葛　艳，时东亮. 基于 YOLOv8 的轻量级自适应权重手势识别模型[J]. 计算机测量与控制，2025，33(10)：225-234.

化、高效的模型需求的增加，我们不得不正视当前轻量化手势识别模型的一个普遍问题，即在大幅降低参数量和计算量的同时，往往伴随着精度的明显下降。因此，设计一种能保证精度的轻量化手势识别算法具有十分重要的意义。

手势识别方法主要可分为基于可穿戴设备的方法、基于机器学习的方法和基于深度学习的方法。基于可穿戴设备的手势识别方法例如 Kinect[1]、Myo[2]等使用红外或光学技术来测量手到传感器的距离，捕捉手势形状从而实现手势识别，但高质量的传感器需要较高成本且精度易受到环境影响。传统的机器学习方法如支持向量机[3]（SVM，support vector machine）、K 近邻算法[4]（KNN，k-nearest neighbor）等，需要先使用分割方法将复杂背景和人手分离，结合方向梯度直方图[5]（HOG，histogram of oriented gradient）、尺度不变特征变换[6]（SIFT，scale-invariant feature transform）等方法进行特征提取，这些方法在多个窗口进行特征提取和分类的滑动窗口策略会导致高人工成本和高时间复杂度，且环境变化适应性和鲁棒性较差。

近年来，随着卷积神经网络（CNN，convolutional neural networks）的迅猛发展，基于深度学习的目标检测领域取得巨大突破。根据目标检测流程的复杂性，基于深度学习的目标检测分为两类：1）双阶段（Two-stage）目标检测，具体如 R-CNN[7-10]（Region-CNN）系列网络等。首先通过区域生成网络（RPN，region proposal network）生成一组候选框，然后在候选框上执行目标分类和位置回归。由于其两步骤的流程，双阶段方法通常具有更高的识别精度，但也使模型的设计更加复杂，需要耗费更多训练和推理时间。2）单阶段（One-stage）目标检测，具体如 SSD[11-13]（Single Shot Multibox Detector）、YOLO[14-20]（You Only Look Once）等，单阶段目标检测模型能够一次性完成检测任务，而不需要生成候选框。通过单个前向传播过程直接检测图像中的目标位置和类别。由于其跳过生成候选框的步骤，单阶段方法具有更加轻量的模型以及更快的训练和推理速度，使它们适用于需要实时性能的应用。

基于深度学习的方法也广泛应用到手势识别领域，尤其是单阶段目标检测算法，基于单阶段目标检测改进的手势识别算法也层出不穷。文献［21］针对复杂背景下手势识别率低的问题，提出结合自适应卷积注意力机制 SKNet 的 HD-YOLOv5s 算法，提升小尺度手势的识别精度，在自制数据集和公共数据集 NUS-Ⅱ上，平均精度均值（mAP，mean average precision）达到 99.5% 和 98.9%，但参数量和模型大小相对较大，不利于部署在边缘设备上。文献［22］针对手势识别模型复杂度高的问题，提出一种融合 Ghost 模块的轻量级 YOLOv3 模型，在 NUS-Ⅱ数据集上 *mAP* 为 99.5%，虽较原模型更为轻量，但该网络模型权重大小为 17.9 MB。文献［23］提出一种结合 ShuffleNetV2 和 Slim-neck 的轻量级手势识别算法 GS-YOLOv5，但该模型在 HaGRID 数据集上 *mAP* 为 95.2%，相比原模型精度相差较大。目前，现有的模型用于手势识别存在以下问题：

1）复杂背景和光照变化导致模型在实际应用中的表现不稳定，使模型泛化能力和鲁棒性下降；

2）远距离或小尺度手势会导致特征丢失、定位困难等问题，降低了模型的识别精度和敏感性；

3）现有手势识别模型虽具备较高精度，但复杂度过高，难以部署在资源受限的边缘设备上，而低复杂度模型又因精度不足难以满足实际需求。

针对上述问题，本文提出一种基于 YOLOv8 的轻量级手势识别模型 LAW-YOLO，主要贡献如下：

1）针对复杂背景和光照变化的问题，设计轻量级自适应权重模块 LAWC 并引入至模型特征提取网络中，通过自适应权重下采样机制对不同区域特征进行加权，以有效捕捉静态手势的关键特征。该模块的轻量化设计在保持高识别精度的同时，有效减少计算和存储资源的消耗。

2）针对远距离或小尺度手势的问题，将特征融合网络中的 PANet 替换为双向特征金字塔网络结构 BiFPN，更有效融合不同层级的特征，提升对不同尺寸和距离手势的识别能力。此外，通过降低特征融合网络的通道数，显著减少了模型的参数量和计算量。

3）为弥补参数量减少所导致的信息丢失和精度损失，在特征融合网络中引入自校正卷积 SC-Conv 改进 C2f，通过自适应调整卷积核的权重分布，确保模型在保持轻量化的同时，仍能维持高识别精度和鲁棒性。

1 YOLOv8 算法

YOLOv8 目前是 YOLO 系列中广泛用于检测和分割的算法，其性能优于 YOLOv5，YOLOv8 根据网络深度和宽度不同，分为 n、s、m、l、x。基于轻量化考虑，本文选择使用模型复杂度最低的 YOLOv8n 作为基线模型。YOLOv8 的网络结构如图 1 所示。

YOLOv8 模型网络主要由以下部分组成：

1）特征提取网络（Backbone）。特征提取网络包含 CBS、C2f、SPPF 等模块。其中 CBS 用于对输入图像进行卷积、批量归一化（BN，batch normalization）和 SiLU 激活操作；C2f 通过增加跨层分支连接，丰富梯度流，增强了残差特征的学习能力，从而构建成一个更具特征表示能力的神经网络模块。SPPF（Spatial Pyramid

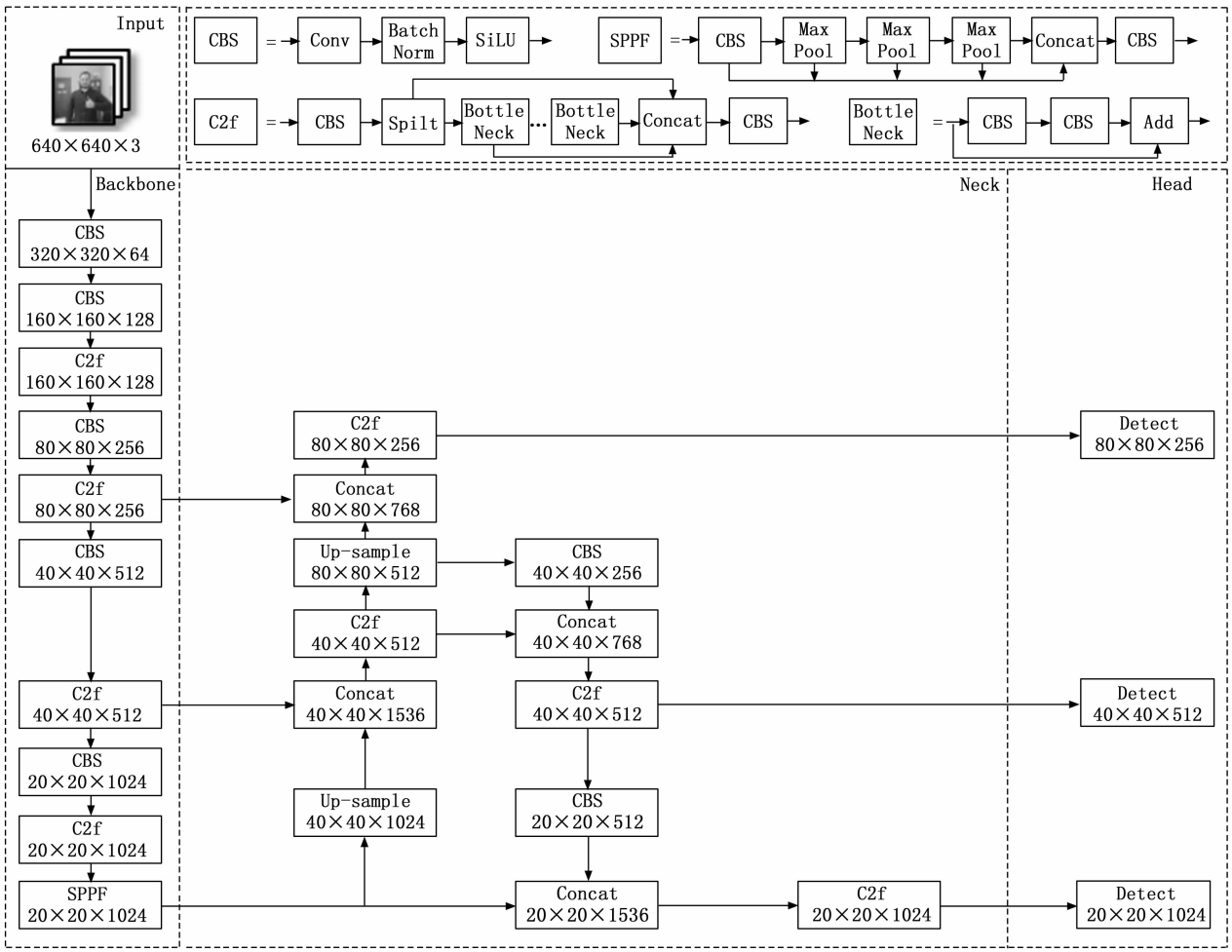

图 1　YOLOv8 网络结构图

Pooling-Fast）为空间金字塔池化模块，用于在多个尺度聚合特征。

2）特征融合网络（Neck）。特征融合网络采用路径聚合网络[24]（PANet，path aggregation network），其中包含特征金字塔网络[25]（FPN，feature pyramid network）。FPN 通过多次上采样、下采样和横向连接，实现深层与浅层特征图的融合。此外，PANet 还通过自下而上的路径进行特征融合，利用元素级加法操作，将底层与顶层特征相结合，增强模型对多尺度特征的理解能力。

3）检测部分（Head）。检测部分在特征融合网络之后，使用 3 个检测头分别检测不同尺度的目标，检测头使用分类头和检测头分离的解耦头结构（Decoupled-Head），同时采用不依赖锚框（Anchor-free）目标检测方法。

2　改进的 YOLOv8 手势识别模型

针对现有手势识别模型在复杂背景、光照变化、小尺度手势识别和模型复杂度与精度之间权衡中的不足，本文提出一种基于 YOLOv8 的轻量级手势识别模型 LAW-YOLO，本文所提出的模型网络结构如图 2 所示，主要有 3 个改进部分。针对复杂背景和光照变化的问题，设计轻量级自适应权重模块 LAWC 并引入至模型特征提取网络中，有效减少计算和存储资源的消耗并提升识别精度；针对远距离或小尺度手势的问题，将特征融合网络中的 PANet 替换为双向特征金字塔网络结构 BiFPN 并压缩特征融合网络的通道数，显著减少了模型的参数量和计算量；为弥补参数量减少所导致的信息丢失和精度损失，在特征融合网络中设计并引入融合了自校正卷积 SC-Conv 的 SC2f 替换检测头前的 C2f 模块，以较少的计算代价有效提升模型的识别精度。

2.1　轻量级自适应权重卷积 LAWC

在手势识别任务中，手势特征是图像中的关键特征。然而，复杂背景和光照变化会使得关键特征在图像中更加难以分辨和提取，导致模型难以准确捕捉到手势特征。传统卷积运算在下采样过程中常会引发信息丢失，这是由于其固定的感受野和卷积核权重，无法根据

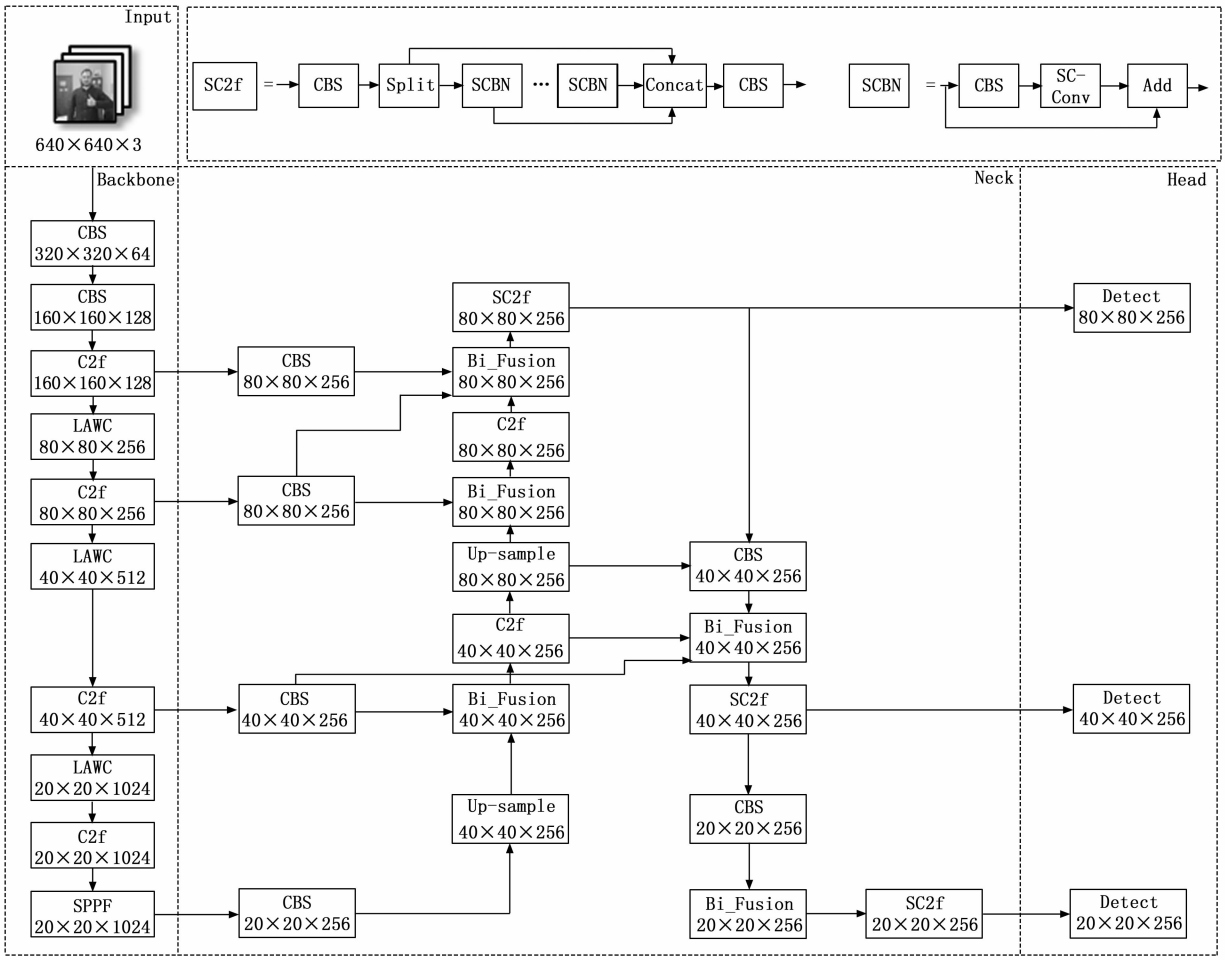

图 2 LAW-YOLO 网络结构图

输入特征进行动态调整，导致模型对图像不同区域特征的适应性不足，从而影响其对关键特征的理解与识别能力。此外，普通卷积操作通常伴随较高的参数量和计算量，随着网络深度的增加，卷积所带来的参数和计算量呈指数级增长，对于计算和存储资源有限的边缘设备而言，极大地加重计算负担，显著降低手势识别模型的运行效率。

注意力机制的引入部分解决了传统卷积在下采样过程中对关键特征不敏感的问题，使网络能够更有效聚焦于关键特征，同时抑制不重要的信息。例如，SE[26]、CBAM[27]等模块以即插即用的方式展现了良好的性能。然而，需要注意的是，注意力机制在减少模型参数量和计算量方面效果有限，反而可能增加计算负担，从而在一定程度上限制了模型的轻量化设计。

为解决上述局限性，本文设计了轻量级自适应权重卷积 LAWC 以替代 Backbone 中的 CBS 进行下采样操作。该模块通过自适应调整权重，实现了一种高效的下采样机制，能够在降低空间维度的同时保留关键特征信息。LAWC 结构如图 3 所示。

LAWC 输入特征图 $X \in R^{C\times H\times W}$，其中 C、H 和 W 分别表示通道数、输入特征图的高度和宽度。LAWC 由两个关键部分组成：空间注意力部分和通道特征整合部分，过程如下：

1）在空间注意力部分中，为生成空间注意力权重图而不引入过多的参数量和计算量，首先对输入特征图进行二维平均池化 AvgPool，捕捉局部上下文信息；然后通过 1×1 卷积进行通道间的线性组合；随后利用 Pytorch 的 rearrange 方法，将大小为 $C\times H\times W$ 的原始特征图重组为 $C\times \frac{H}{2}\times \frac{W}{2}\times 4$ 特征图，新的维度“4”存储了每个 2×2 局部区域的空间特征信息；最后，对新维度进行 softmax 操作，提供自适应机制，基于输入特征动态调整每个位置的权重，从而使重要的手势区域获得更高的权重。

2）在下采样部分中，采用分组数为 16 的分组卷积 GroupConv 对输入特征图进行高效特征提取，同时在空间下采样过程中将通道数扩展至原来的 4 倍，将大小为 $C\times H\times W$ 的原始特征图生成为 $4C\times \frac{H}{2}\times \frac{W}{2}$ 的特征图；

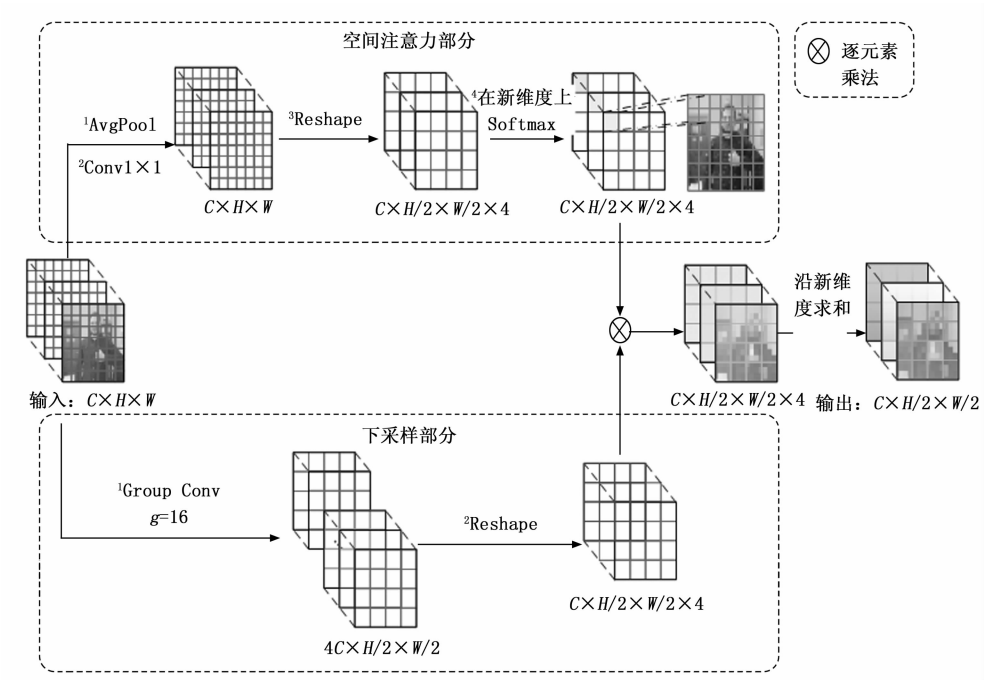

图 3　LAWC 模块结构图

随后，将大小为 $4C\times\frac{H}{2}\times\frac{W}{2}$ 的特征图重组为 $C\times\frac{H}{2}\times\frac{W}{2}\times 4$ 的特征图，新的维度“4”存储了每个通道在分组卷积后保存的信息。

3）将空间注意力部分得到的注意力权重图与下采样部分卷积后的结果进行逐元素相乘，动态调整每个位置的特征权重。

4）最终，沿新维度进行求和操作，形成 $C\times\frac{H}{2}\times\frac{W}{2}$ 大小的输出特征图。通过求和，融合了来自不同通道的局部特征信息。

LAWC 计算公式如下：

$$F_{\text{Spatial}}=\text{Soft}\{R_{\text{range}}\{Conv_{1\times1}[Avg(X)]\}\} \tag{1}$$

$$F_{\text{channel}}=R_{\text{range}}[G_{Conv}(X)] \tag{2}$$

$$F_{\text{LAWC}}=\sum_{i=1}^{4}(F_{\text{Spatial}}\times F_{\text{channel}}) \tag{3}$$

其中：F_{Spatial}、F_{channel} 和 F_{LAWC} 分别为空间注意力部分、通道特征整合部分以及 LAWC 的公式计算结果；在公式（1）中，Avg 为对输入特征图进行全局平均池化操作，$Conv_{1\times1}$ 为 1×1 卷积，R_{range} 表示进行 Rearange 操作增加新的维度，Soft 表示进行 Softmax 操作，生成自适应权重图；在公式（2）中，G_{Conv} 为分组卷积；在公式（3）中，F_{Spatial} 与 F_{channel} 逐元素相乘，并沿着新的维度“4”进行求和，得到最终的结果 F_{LAWC}。

值得注意的是，LAWC 模块采用的分组卷积，减少了参与计算的通道数，使得每次卷积操作的计算负担显著降低；简化的空间注意力机制在增加自适应特征选择能力的同时，并没有引入过多的计算复杂度。因此，LAWC 降低了计算负担并实现了更高的计算效率。

2.2　双向特征金字塔网络 BiFPN

在手势识别任务中，小尺度手势在图像中占据的像素区域较少，其特征在高分辨率特征图中更加明显，而低分辨率特征图则包含更多的语义信息。YOLOv8 采用路径聚合网络 PANet 作为特征融合的网络结构，PANet 的网络结构如图 4 所示。

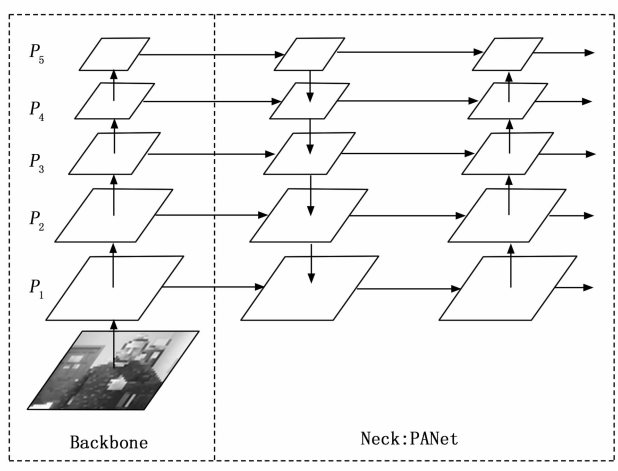

图 4　PANet 结构图

尽管 PANet 通过路径聚合增强了多尺度特征的融合，但其单向的特征流动机制在特征融合过程中可能会导致信息损失，此外，在融合不同分辨率的特征图时，

PANet 采用简单的相加（Add）或连接（Concat），未充分考虑各特征在融合过程中的相对重要性。因此，本文采用双向特征金字塔网络 BiFPN[28] 替换 PANet 作为 YOLOv8n 的特征融合网络，其结构如图 5 所示。

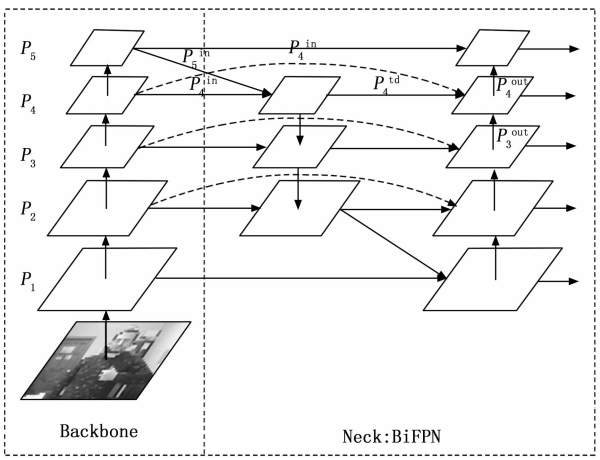

图 5 BiFPN 结构图

BiFPN 在同一层级的输入特征图与输出特征图之间增加一条额外的连接路径，进一步丰富了信息流动，使特征融合不局限于跨层次的上下传递，还实现了同层级内的横向信息交互，从而更有效地结合细节信息和语义信息，特别是在小尺度目标检测任务中表现优异。此外，与 Add 和 Concat 不同的是，BiFPN 的特征图融合操作（Bi _ Fusion）引入了自适应融合权重机制，为每个输入特征图分配动态权重，确保了在特征融合过程中关键特征得到强调。通过迭代使用 Bi _ Fusion，信息在高层与低层之间得以更高效地传递与整合。

在图 5 中，以第 4 层 P_4 为例，BiFPN 的两个融合特征 P_4^{td} 和 P_4^{out}：

$$P_4^{td} = Conv\left(\frac{w_1 \times P_4^{in} + w_2 \times R(P_5^{in})}{w_1 + w_2 + \varepsilon}\right) \tag{4}$$

$$P_4^{out} = Conv\left(\frac{w'_1 \times P_4^{in}}{w'_1 + w'_2 + w'_3 + \varepsilon} + \frac{w'_2 \times P_4^{td} + w'_3 \times R(P_3^{out})}{w'_1 + w'_2 + w'_3 + \varepsilon}\right) \tag{5}$$

公式（4）中，P_3^{out}、P_4^{in}、P_4^{td}、P_4^{out}、P_5^{in} 分别表示第 3 层输出特征、第 4 层输入特征、第 4 层中间特征、第 4 层输出特征、第 5 层输入特征；$Conv$ 表示进行卷积操作；R 表示特征图上采样或下采样操作；W_i 和 W'_i 表示可学习的权重参数；ε 设置为 0.000 1 确保数值的稳定性。

中间特征 P_4^{td} 通过两个输入特征 P_4^{in} 和 P_5^{in} 的权重参数 W_1 和 W_2，使 P_4^{td} 能够自适应地学习来自 P_4、P_5 两个尺度中的重要特征信息；P_4^{out} 能够学习来自 P_3、P_4、P_5 三个尺度的重要特征信息。类似地，其他层的中间特征和输出特征也通过此方式得到。需要注意的是，公式中涉及相加操作，因此在特征融合时，不仅要求输入特征图的空间维度相同，通道数也必须一致。本文将特征融合网络的统一设定为 256，以便进行特征融合的同时，也显著降低参数量和计算量。需要注意的是，若两个输入特征图通道数不相同或不等于 256，则需使用 1×1 卷积调整通道数。

2.3 自校正模块 SC2f

尽管 BiFPN 增强了模型性能，但降维操作却导致了明显的精度下降。为了在不额外引入参数和增加复杂性的情况下，尽可能弥补特征融合网络中通道数缩减所导致的精度损失，本文在检测头前引入自校正卷积 SC-Conv[29] 改进 C2f 模块，设计自校正模块 SC2f 模块，即在 C2f 内部将 CBS 替换为 SC-Conv。结构如图 6 所示。

为了有效收集每个空间位置丰富的上下文信息，SC-Conv 在两个不同的尺度空间中进行卷积特征转换：原始尺度空间中的特征图和下采样后的具有较小分辨率的潜在空间（自校正空间），利用下采样后特征具有较大的感受野，因此在较小的自校正空间中进行变换后的

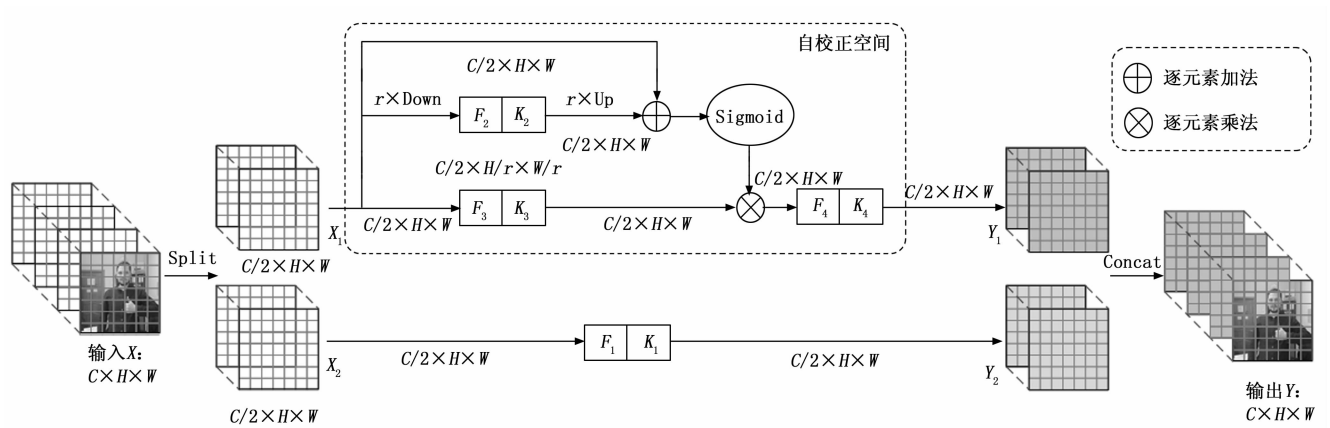

图 6 SC-Conv 结构图

嵌入将用作参考，以指导原始特征空间中的特征变换过程。输入特征图 $X \in R^{C\times H\times W}$，在通道维度拆分成两个 $\frac{C}{2}\times H\times W$ 大小的 X_1 和 X_2，卷积核 K_1，K_2，K_3，K_4，大小均为 $\frac{C}{2}\times\frac{C}{2}\times H\times W$，作用各不相同。SC-Conv 对特征图的处理过程如下：

1）在自校正空间中，首先对特征图 X_1 进行二维平均池化，将特征图尺寸缩小，保留局部上下文信息，得到 T_1：

$$T_1 = Avg_r(X_1) \tag{6}$$

式中，Avg_r 表示下采样倍率为 r 的二维平均池化。池化后的特征图经过卷积核 K_2 进行卷积操作，生成低分辨率的全局特征图，随后，使用双线性插值方法，将全局特征图映射为原始特征图相同的分辨率，即：

$$X'_1 = Up[F_2(T_1)] = Up(T_1 \times K_2) \tag{7}$$

式中，Up 表示双线性插值。X'_1 作为残差来形成用于校准的权重。自校准操作如下：

$$Y'_1 = F_3(X_1)\times\sigma(X_1 + X'_1) \tag{8}$$

式中，$F_3(X_1) = X_1 \times K_3$，$\sigma$ 表示 Sigmoid 函数。校准后的最终输出为：

$$Y_1 = F_4(Y'_1) = Y'_1 \times K_4 \tag{9}$$

2）原始尺度空间的特征图 Y_2 由 X_2 得到：

$$Y_2 = F_1(X_2) = X_2 \times K_1 \tag{10}$$

最终的输出特征图 Y：

$$Y = \mathrm{Concat}(Y_1 + Y_2) \tag{11}$$

如图 7 所示，SC2f 的 Bottleneck（SCBN）首先经过一个 CBS 对特征图进行通道维度压缩；接着经过 SC-Conv 引入自校正机制，确保重要区域的特征在卷积后能够得到保留；最后将经 SC-Conv 得到的特征图与输入特征进行残差连接。

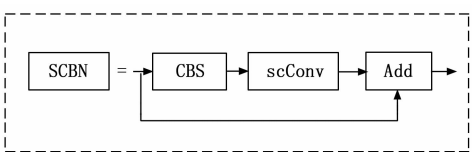

图 7　SC2f 的 Bottleneck（SCBN）结构图

如图 8 所示，SC2f 将 YOLOv8 的 C2f 模块中的 Bottleneck 全部替换为 SCBN，增强网络的学习能力，使得模型能够在不引入过多参数量、计算量的情况下更好地捕捉重要的手势特征，能够有效提升模型在边缘设备上的运行性能。

3　实验结果

3.1　实验设置

3.1.1　实验环境

本文实验于 Windows 11 专业版 64 位操作系统下进行，处理器为 Intel i5-13600KF，GPU 为 NVIDIA GeForce RTX 4070 Advanced OC，GPU 显存大小为 12 GB，使用 CUDA11.7 和 CuDNN8.9.3 对 GPU 进行加速，本文基于 PyTorch 2.0.0 深度学习框架下，以 Anaconda3.0、PyCharm2022.3、Python3.9 作为实验的软件环境。

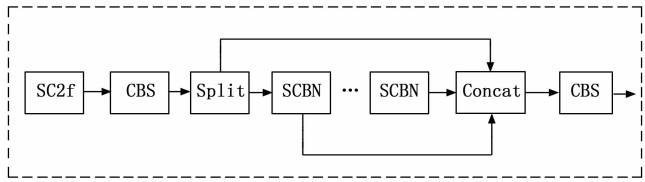

图 8　SC2f 结构图

3.1.2　实验参数设置

本文实验使用初始动量为 0.937 的随机梯度下降算法（SGD，stochastic gradient descent）优化器，输入图像尺寸为 640×640，初始学习率为 0.000 01，批处理大小为 32，训练轮数为 300 次。

3.1.3　实验数据集

本文实验使用的手势图像来自公开数据集 HaGRID[30]。如图 9 所示，HaGRID 共包含 18 个类别的手势图像，这些图像由众多不同种族的拍摄对象于不同光线条件及不同场景下拍摄，受到光线、角度、遮挡、场景以及肤色的影响。本文从每个类别中随机选取 300 张图像，共计 5 400 张。为了提高模型的鲁棒性和泛化能力，防止训练过程中出现过拟合现象，本文使用平移、旋转、增加噪声、随机裁剪的方法对所选图像进行数据增强，每个类别扩充到 1 200 张，共计 21 600 张。并按照 6∶2∶2 的比例将所扩充的数据集划分为训练集、验证集和测试集。

图 9　HaGRID 数据集部分手势图像

3.1.4　评估指标

本文使用准确率（P，precision）、召回率（R，re-

call)、平均精度均值 *mAP*、参数量（Pa，parameters)、每秒浮点运算数（FLOPs，floating-point operations per second)、模型权重大小（Weight Size）作为基本指标用于评估模型的性能。其中准确率 *P* 表示被检测到的目标中真实目标的比例；召回率 *R* 表示所有真实目标中被正确检测到的比例；平均精度均值 *mAP* 表示在多个类别上 *IoU* 阈值为 50%的平均精度；参数量 *Pa* 表示模型所包含的参数的数量；每秒浮点运算数 FLOPs 表示模型每秒浮点数计算量；模型权重大小 W_s 表示模型权重文件大小。

3.2 对比实验

为证明研究提出的本文改进模型的有效性，将本文改进模型与主流的轻量级模型、主流目标检测模型、目前最新改进的手势识别模型进行对比，主流轻量级模型选取以 Mobilenetv3[31]、Shufflenetv2[32]、Pplcnet[33] 为特征提取网络的 YOLOv8 改进模型；主流目标检测模型选取 YOLOv3-tiny、YOLOv5n、YOLOv8n、YOLOv8s；最新改进的手势识别模型选取 HD-YOLOv5s[21] 和 GS-YOLO[23]。实验结果如表 1 所示。

表 1 对比实验结果

模型名称	*P*/%	*R*/%	*mAP*/%	*Pa*/M	*FLOPs*/G	*Ws*/MB
Mobilenetv3	95.4	93.9	98.0	2.18	5.4	4.7
Shufflenetv2	94.9	91.7	97.1	1.83	5.1	3.9
Pplcnet	95.7	91.0	97.4	1.85	5.2	3.9
YOLOv3-tiny	95.4	91.5	96.7	12.14	18.9	23.3
YOLOv5n	96.1	93.0	97.5	1.78	4.3	3.8
YOLOv8s	97.4	95.0	98.5	11.13	28.5	22.5
HD-YOLOv5s	96.9	95.6	98.7	13.60	17.9	14.1
GS-YOLO	93.2	88.1	94.4	2.27	2.9	4.4
YOLOv8n	95.9	94.1	98.1	3.01	8.1	6.2
Ours	**97.2**	**94.9**	**98.6**	**1.77**	**7.1**	**3.6**

表 1 表明，本文改进模型表现突出，主要体现在以下方面：

1）本文改进模型在 *mAP* 上达到 98.6%，相较于 YOLOv8n 提升了 0.5%，能够与 YOLOv8s 的 98.5%和 HD-YOLOv5s 的 98.7%相媲美。虽然本文改进模型准确度 *P* 和召回率 *R* 上略低于 YOLOv8s 和 HD-YOLOv5s，但超越了其他对比模型，其 *mAP* 的优势表明其在复杂场景下的手势识别能力表现优越。不同模型的识别对比效果如图 10 所示。

图 10 展示了不同模型对图像中手势识别的效果，其中从左至右的五列示例分别展示了各个模型的识别结果。可以看出，本文提出的改进模型 LAW-YOLO 在手势识别任务中表现出较高的识别精度，且在所有示例中均未出现误检或漏检情况。

图 10 不同模型识别对比效果图

2）在计算复杂度和模型大小方面，本文提出的改进模型优于所有对比模型。具体而言，相较于 YOLOv8n，本文改进模型在参数量、计算量和模型大小上分别降低了 41.20、12.35 和 41.94%。特别是在参数量方面，本文改进模型仅为 YOLOv8s 的约 16%和 HD-YOLOv5s 的约 13%，而在平均精度均值上，则实现了几乎相同的性能。这表明，尽管模型规模显著减小，但在精度和性能方面并未降低，体现了本文改进模型在效率和效果上的优越性。

综上所述，本文改进模型能够有效提升手势识别的整体性能，充分验证了其设计的有效性和优势。然而，尽管其在精度、参数量和模型大小上均表现优异，但仍有进一步降低计算量的空间。需要进一步优化和研究以实现更全面的性能提升。

3.3 消融实验

由于在将 BiFPN 引入本文研究的改进模型时，统一了特征融合网络的通道数，不同通道数的设置对模型性能的影响不可忽视。较高的通道数可能提供更多的特征信息，但同时也增加了计算量和参数量；而较低的通道数则可能简化模型，降低计算复杂度，但可能会丧失一部分关键信息，因此，通道数的优化需在保证模型性能的同时，考虑到计算资源的限制和实际应用需求。表 2 展示了不同通道数对模型性能的影响。

表 2 引入 BiFPN 后不同通道数对模型性能的影响

数量	*P*/%	*R*/%	*mAP*/%	*Pa*/M	*FLOPs*/G
512	96.4	96.8	98.8	3.47	15.1
256	96.4	94.6	98.3	1.99	7.1
128	96.1	92.6	97.8	1.58	5.0

表 2 中，通道数设置为 512 时，P、R 和 mAP 均达到最高，然而，这一设置也伴随着模型参数量和计算量的显著增加；当通道数减少至 128 时，虽然模型的参数量和计算量显著减少，达到表 2 中最小值，但 mAP 也显著下降。这表明，较低的通道数可能导致特征表达能力不足，从而影响模型性能。为了在精度和计算复杂度之间实现平衡，本文选择将通道数设置为 256，以在保证模型性能的同时，优化计算资源的使用。

为验证各改进模块对手势识别的有效性，将基于 YOLOv8 模型进行消融实验，消融实验结果如表 3 所示，其中“√”表示采用该方法。“×”代表不采用该方法。实验结果如表 4 所示。

表 3　改进模型消融实验设计

模型	LAWC	BiFPN	SC2f
①	×	×	×
②	√	×	×
③	√	√	×
④	√	√	√

表 4　消融实验结果

模型	P/%	R/%	mAP/%	Pa/M	FLOPs/G	Ws/MB
①	95.9	94.1	98.1	3.01	8.1	6.2
②	96.6	94.8	98.5	2.73	7.9	5.7
③	97.0	94.7	98.2	1.72	6.9	3.7
④	97.2	94.9	98.6	1.77	7.1	3.6

表 3 和表 4 显示了各模型的性能差异。对比模型①和②，模型②在所有指标上均有所提升，其中 mAP 提升 0.4%，参数量下降 9.3%，计算量减少 2.4%，模型大小缩减 8.1%。这表明 LAWC 模块在提升模型性能方面发挥了重要作用，通过有效整合特征和优化特征图处理方式，不仅提高了识别精度，还在不显著增加计算负担和模型体积的情况下显著降低了参数量和计算开销。对比模型②和③，尽管模型③的 mAP 略微下降了 0.2%，但参数量减少了 37.0%，计算量减少了 12.7%，模型大小下降了 35.1%。这表明引入 BiFPN 并减少通道数在显著降低模型复杂度的同时，优化了特征融合策略，使得模型在处理小尺度特征时更为高效。尽管 mAP 略有下降，但整体效果的提升仍然显著。对比模型③和④，模型④在参数量、计算量和模型大小上分别上升了 2.9%、2.9%和 2.7%，但 mAP 提升了 0.3%。这表明引入 SC2f 模块后，虽然模型在参数量、计算量和模型大小上有所增加，但该模块有效地弥补了由于特征融合网络通道数减少导致的精度下降问题，从而提升了模型的整体识别精度。

4　结束语

本文提出了一种基于 YOLOv8 的轻量级自适应权重手势识别模型，首先，设计并引入了轻量级自适应权重模块 LAWC，该模块通过自适应权重下采样机制有效地处理复杂背景和光照变化，对不同区域特征进行加权，从而在保持高识别精度的同时显著减少计算和存储资源消耗。其次，将特征融合网络中的 PANet 替换为双向特征金字塔网络 BiFPN，有效提升了对远距离或小尺度手势的识别能力，同时通过降低通道数显著减少了模型的参数量和计算量。最后，使用融入了自校正卷积 SC-Conv 的自校正模块 SC2f 替换检测头前的 C2f 模块，通过自适应调整卷积核的权重分布，弥补了参数量减少带来的信息丢失和精度下降问题，确保了模型在轻量化的同时维持高识别精度和鲁棒性。综上所述，本文提出的改进模型不仅有效地解决了传统手势识别模型中的多个关键问题，还在处理高精度和高效率要求的任务时展现了作为高效手势识别模型的广泛应用潜力。

参考文献：

[1] 林清宇. 基于 Kinect 的手势检测与追踪研究 [D]. 南京：南京邮电大学，2020.

[2] SATHIYANARAYANAN M，RAJAN S. MYO Armband for physiotherapy healthcare：a case study using gesture recognition application [C] //2016 8th International Conference on Communication Systems and Networks (COMSNETS). IEEE，2016：1－6.

[3] YAN F，MEI W，CHUNQIN Z. SAR image target recognition based on Hu invariant moments and SVM [C] //2009 Fifth International Conference on Information Assurance and Security. IEEE，2009，1：585－588.

[4] AL-FAIZ M Z，ALI A A，MIRY A H. A k-nearest neighbor based algorithm for human arm movements recognition using EMG signals [C] //2010 1st International Conference on Energy，Power and Control (EPC-IQ). IEEE，2010：159－167.

[5] LUKAS P，YUJI O，YOSHIHIKO M，et al. A HOG-based hand gesture recognition system on a mobile device [C] // IEEE International Conference on Image Processing (ICIP)，IEEE，2014.

[6] SUTTAPAK W，AUEPHANWIRIYAKUL S，THEERA-UMPON N. Incorporating SIFT with hard c-means algorithm [C] //2010 The 2nd International Conference on Computer and Automation Engineering (ICCAE). IEEE，2010，4：437－441.

[7] AGRAWAL P，GIRSHICK R，MALIK J. Analyzing the performance of multilayer neural networks for object recog-

nition [C] //Computer Vision-ECCV 2014: 13th European Conference, Zurich, Switzerland, Proceedings, Part VII 13. Springer International Publishing, 2014: 329 - 344.

[8] GIRSHICK R. Fast r-CNN [J]. Arxiv Preprint Arxiv: 1504.08083, 2015.

[9] 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 and Machine Intelligence, 2016, 39 (6): 1137 - 1149.

[10] HE K, GKIOXARI G, DOLLáR P, et al. Mask r-CNN [C] //Proceedings of the IEEE International Conference on Computer Vision, 2017: 2961 - 2969.

[11] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector [C] //Computer Vision-ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, Proceedings, Part I 14. Springer International Publishing, 2016: 21 - 37.

[12] FU C Y, LIU W, RANGA A, et al. DSSD: Deconvolutional single shot detector [J]. Arxiv Preprint Arxiv: 1701.06659, 2017.

[13] LI Z, YANG L, ZHOU F. FSSD: feature fusion single shot multibox detector [J]. Arxiv Preprint Arxiv: 1712.00960, 2017.

[14] REDMON J. You only look once: unified, real-time object detection [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016.

[15] REDMON J, FARHADI A. YOLO9000: better, faster, stronger [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 7263 - 7271.

[16] FARHADI A, REDMON J. Yolov3: An incremental improvement [C] //Computer Vision and Pattern Recognition, Berlin/Heidelberg, Germany: Springer, 2018: 1 - 6.

[17] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. Yolov4: optimal speed and accuracy of object detection [J]. Arxiv Preprint Arxiv: 2004.10934, 2020.

[18] ZHENG G, SONGTAO L, FENG W, et al. YOLOX: exceeding YOLO series in 2021 [J]. Arxiv Preprint Arxiv: 2107.08430, 2021.

[19] ZHANG Y, GUO Z, WU J, et al. Real-time vehicle detection based on improved yolov5 [J]. Sustainability, 2022, 14 (19): 12274.

[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] 闫颢月, 王 伟, 田 泽. 复杂环境下基于改进 YOLOv5 的手势识别方法 [J]. 计算机工程与应用, 2023, 59 (4): 224 - 234.

[22] 牛雅睿, 武 一, 孙 昆, 等. 基于轻量级卷积神经网络的手势识别检测 [J]. 电子测量技术, 2022, 45 (4): 91 - 98.

[23] GUO J, LEI X, LI B. A lightweight gesture recognition network [J]. Journal of Visual Communication and Image Representation, 2025, 107: 104362.

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

[25] LIN T Y, DOLLáR P, GIRSHICK R, et al. Feature pyramid networks for object detection [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2117 - 2125.

[26] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132 - 7141.

[27] WOO S, PARK J, Lee J Y, et al. Cbam: Convolutional block attention module [C] //Proceedings of the European Conference on Computer Vision (ECCV). 2018: 3 - 19.

[28] TAN M, PANG R, LE Q V. Efficientdet: Scalable and efficient object detection [C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 10781 - 10790.

[29] LIU J J, HOU Q, CHENG M M, et al. Improving convolutional networks with self-calibrated convolutions [C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 10096 - 10105.

[30] KAPITANOV A, KVANCHIANI K, NAGAEV A, et al. HaGRID-hand gesture recognition image dataset [C] //Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2024: 4572 - 4581.

[31] HOWARD A, SANDLER M, CHU G, et al. Searching for mobilenetv3 [C] //Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 1314 - 1324.

[32] MA N, ZHANG X, ZHENG H T, et al. Shufflenet v2: Practical guidelines for efficient CNN architecture design [C] //Proceedings of the European Conference on Computer Vision (ECCV). 2018: 116 - 131.

[33] CUI C, GAO T, WEI S, et al. PP-LCNet: a lightweight CPU convolutional neural network [J]. Arxiv Preprint Arxiv: 2109.15099, 2021.