文章编号:1007－6735(2025)05－0578－12　DOI: 10.13255/j.cnki.jusst.20240827002

基于 YOLOv8n 和 RexNet 的轻量化动态手势识别网络设计

廖良进[1]，　陈晓荣[1]，　倪　明[2]，　李文莎[1]，　李柏杨[1]，　倪思佳[1]
（1. 上海理工大学 光电信息与计算机工程学院，上海 200093；2. 上海交通大学医学院附属瑞金医院，上海 200025）

摘要：针对手术室环境下的动态手势人机交互问题，提出了一种基于深度学习的手势识别算法 YOLO-RexNet。基于 YOLOv8n 的动态手势检测部分，在网络中融入 DualConv 卷积模块，利用 GhostNetV2 重构 C2f 模块，并应用模型剪枝技术，在减少模型参数量和计算量的同时保持较高的检测精度。实验结果表明，改进后的算法参数量减少了 70.9%，计算量降低了 69.7%，模型权重大小下降了 66.2%，检测耗时降低了 41.3%，动态手势识别的准确率达到 99%。基于 RexNet 的手部关键点检测部分，将损失函数替换为 Huber_Loss，使关键点的预测更加准确。所提算法在保证动态手势识别精度的同时，满足轻量化要求，实现了对医疗设备的便捷操作。在 Jetson orin nano 嵌入式边缘设备上部署时，每秒帧数能够达到 65，具备良好的应用前景。

关键词：动态手势识别；关键点检测；YOLOv8n；RexNet；模型剪枝
中图分类号：TP 399　　**文献标志码：**A

Design of lightweight dynamic gesture recognition network based on YOLOv8n and RexNet

LIAO Liangjin[1], CHEN Xiaorong[1], NI Ming[2], LI Wensha[1], LI Baiyang[1], NI Sijia[1]
(1. *School of Optical−Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China*; 2. *Ruijin Hospital, Shanghai Jiaotong University School of Medicine, Shanghai 200025, China*)

Abstract: To address the challenges of dynamic gesture human-computer interaction in the operating room environment, a gesture recognition algorithm named YOLO-RexNet, based on deep learning, was proposed. The dynamic gesture detection component utilized YOLOv8n, incorporated the DualConv convolution module into the network, reconstructed the C2f module with GhostNetV2, and applied

收稿日期: 2024−08−27
基金项目: 国家自然科学基金资助项目(52175513)
第一作者: 廖良进(2001—)，男，硕士研究生．研究方向：计算机视觉、图像处理．E-mail：llj18878368816@163.com
通信作者: 陈晓荣(1974—)，女，副教授．研究方向：自动测量与在线检测、信号与信息处理、图像处理技术．E-mail：chenxr@usst.edu.cn

引文格式: 廖良进，陈晓荣，倪明，等．基于 YOLOv8n 和 RexNet 的轻量化动态手势识别网络设计[J]．上海理工大学学报，2025，47(5)：578-589.
Citation: LIAO Liangjin, CHEN Xiaorong, NI Ming, et al. Design of lightweight dynamic gesture recognition network based on YOLOv8n and RexNet[J]. Journal of University of Shanghai for Science and Technology, 2025, 47(5): 578-589.

model pruning techniques to reduce parameters and computation while maintaining high detection accuracy. Experimental results demonstrate a 70.9% reduction in parameters, a 69.7% reduction in computation, a 66.2% reduction in model size, and a 41.3% decrease in inference latency, with dynamic gesture-recognition accuracy reaching 99%. In the hand key point detection component, the Huber_Loss function was employed to enhance the precision of key point predictions. The proposed algorithm achieves lightweight design while ensuring the accuracy of dynamic gesture recognition, aiming to facilitate the gesture-controlled medical operations. When deployed on Jetson orin nano embedded edge devices, the frames per second can reach 65, which demonstrates practical potential.

Keywords: *dynamic gesture recognition*; *key point detection*; *YOLOv8n*; *RexNet*; *model pruning*

在手术室场景中，医护人员与患者之间的交叉感染是一个重要的健康隐患，特别是在操作计算机等相关医疗设备时，传统的接触式交互方式可能会增加这种风险，因此，减少医护人员对设备的直接接触变得至关重要。然而，当医生在手术过程中不便直接操作计算机鼠标时，如何有效地控制医疗设备仍然是一个亟待解决的问题[1]。在这种背景下，动态手势作为一种非接触式交互方式，能够有效减少医护人员与设备的直接接触，从而降低交叉感染的风险，并为医生提供一种更加灵活、便捷的操作方式。

一般情况下，动态手势识别可分为基于传感器[2-4]和基于计算机视觉两类。尽管基于传感器的动态手势识别能够实现较高的准确率，但需佩戴大量传感器，影响使用体验且成本较高。Wang 等[5]提出了一种基于表面肌电（sEMG）的动态手势识别方法，将 sEMG 信号转换为彩色图像，建立基于深度神经网络的动态手势识别。但该方法依赖于特定硬件设备获取信号，不利于实际应用。

相比之下，基于计算机视觉的动态手势识别逐渐成为主流。彭玉青等[6]提出了一种手势识别算法 HGDR-Net，利用经过优化的 YOLO（you only look once）算法来检测手势的边界和位置，并通过卷积神经网络（convolutional neural network，CNN）对检测到的手势进行精细的识别和分类，效果良好；然而，在背景剧烈变化和画面中存在多个手部的情况下，手势识别算法性能有所不足。Peng 等[7]提出了一种基于关键点统计变换特征的改进动态手势识别框架，降低了对背景和光照条件的敏感性。然而，该算法计算复杂度较高，实时性较差。

为解决上述问题，本文以实现动态手势控制计算机等医疗设备为背景，提出了 YOLO-RexNet 算法。本文提出的算法采用改进的 YOLOv8n 目标检测算法进行动态手势检测、分类、定位，借助轻量化网络 RexNet[8]进行关键点检测，获取不同类别动态手势以及手部关键点坐标信息，并根据提取到的相关信息实现动态手势控制医疗设备。

1　YOLO-RexNet 算法

YOLO-RexNet 算法在现实中的应用场景如图 1 所示。RGB 高清摄像头采集医生的手势图像并传输至计算机系统，通过本文提出的算法实现无接触式动态手势识别的人机交互。当系统在画面中识别到医生各种动态手势时，可依据提取到的手部关键点的坐标信息，取代医疗设备中鼠标的控制。放大、缩小、旋转术中医疗设备投射到显示器上的图像，更好地观察图像中细节，进而指导手术操作。

图 1　手术室内动态手势控制场景

Fig.1　Dynamic gesture control scene in the operating room

YOLO-RexNet 具体的算法流程如图 2 所示，可分为动态手势检测和手部关键点检测两个部分。

图 2 上半部分为动态手势检测，用于分类和

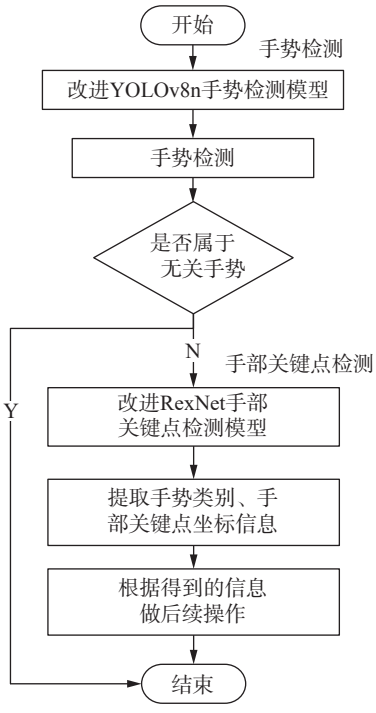


图 2 YOLO-RexNet 算法流程

Fig.2 Flow chart of YOLO-RexNet algorithm

定位动态手势区域，并记录动态手势区域预测框的位置信息。通过改进的 YOLOv8n 算法训练模型，对图像中的动态手势进行检测，定位动态手势区域。然而在手术室场景中，画面中往往会出现多只手，这些“无效手”可能对术中医生操作设备产生干扰。因此，在模型训练中引入“无效手”作为检测目标，旨在提高模型区分目标动态手势与干扰项的能力，进而降低检测错误率。

图 2 下半部分为手部关键点检测，用于检测动态手势区域中的手部关键点，提取关键点的坐标信息。利用基于 RexNet 设计的轻量化网络训练模型，对目标动态手势区域进行关键点检测。本文提出的算法基于手部的 21 个关键点构造，实现全面捕捉和解析手部姿态信息，利用这些动态手势分类信息和手部关键点坐标数据，支持医疗手术室场景中无接触控制医疗设备的人机交互任务，提高操作的精确性和效率。

2 基于改进的 YOLOv8n 的动态手势检测

2.1 改进的 YOLOv8n 模型

YOLOv8n 拥有复杂的网络架构、优化的训练流程和强大的特征提取能力，可以完成实时的目标检测任务。主干网络、颈部网络和头部网络组成了其模型的主要架构：主干网络负责从输入图像中提取特征；颈部网络将不同尺度的特征信息与上采样得到的特征信息结合；头部网络用于目标输出预测。

本文算法在 YOLOv8n 算法上进行优化，改进后的结构如图 3 所示。首先，在网络模型中融入 DualConv[9] 卷积模块；其次，利用 GhostNetV2[10] 中的 GhostV2 bottleneck 结构重构 C2f 模块，并将原始网络模型中的 C2f 模块替换成 Ghost_C2f 模块；最后，与 DualConv 融合形成 DG_C2f 模块并嵌入到主干网络中，轻量化的同时保证特征信息的提取。

2.1.1 轻量化卷积模块 DualConv

为了在实际应用中能够部署于资源有限的边缘设备上，需要考虑尽可能降低模型的参数量和计算量。由于深度神经网络中大部分的参数量和计算量来自于卷积操作，对 YOLOv8n 中的卷积模块进行优化。

DualConv 是一种轻量级卷积模块，能够在显著降低深度神经网络计算成本和参数数量的同时保证检测精度。DualConv 结合了深度可分离卷积[11]、组卷积[12]、异构卷积[13] 的优点进行卷积操作。

在标准卷积中，假设输出特征图的维度为 $D_o \times D_o$，卷积核个数为 N，输入图像通道数为 M，输入特征映射在卷积层通过 N 个大小为 $K \times K \times M$ 的卷积核进行运算。此时，在标准卷积层 F_{LSC} 中执行的 FLOPs（floating point operations）总数为

$$F_{\mathrm{LSC}} = D_o^2 K^2 MN \tag{1}$$

在 DualConv 中，$K \times K$ 卷积核比例受卷积核组个数 G 控制。在所有通道中 $(K \times K + 1 \times 1)$ 的卷积核组合的比例为 $1/G$，剩余 1×1 卷积核的比例为 $(1-1/G)$[9]。因此，在由 G 个卷积核组组成的双卷积层中，组合卷积核的 FLOPs 数、剩余 1×1 卷积核的 FLOPs 数、FLOPs 总数分别为

$$F_{\mathrm{LCC}} = \left(D_o^2 K^2 MN + D_o^2 MN\right)/G \tag{2}$$

$$F_{\mathrm{LPC}} = \left(D_o^2 MN\right)(1 - 1/G) \tag{3}$$

$$F_{\mathrm{LDC}} = F_{\mathrm{LCC}} + F_{\mathrm{LPC}} \tag{4}$$

比较 DualConv 与标准卷积的计算成本（FLOPs），可以明显看出，DualConv 能够减少卷积运算的计算成本。本文算法中设置 K=3，G=2。当 G 明显增

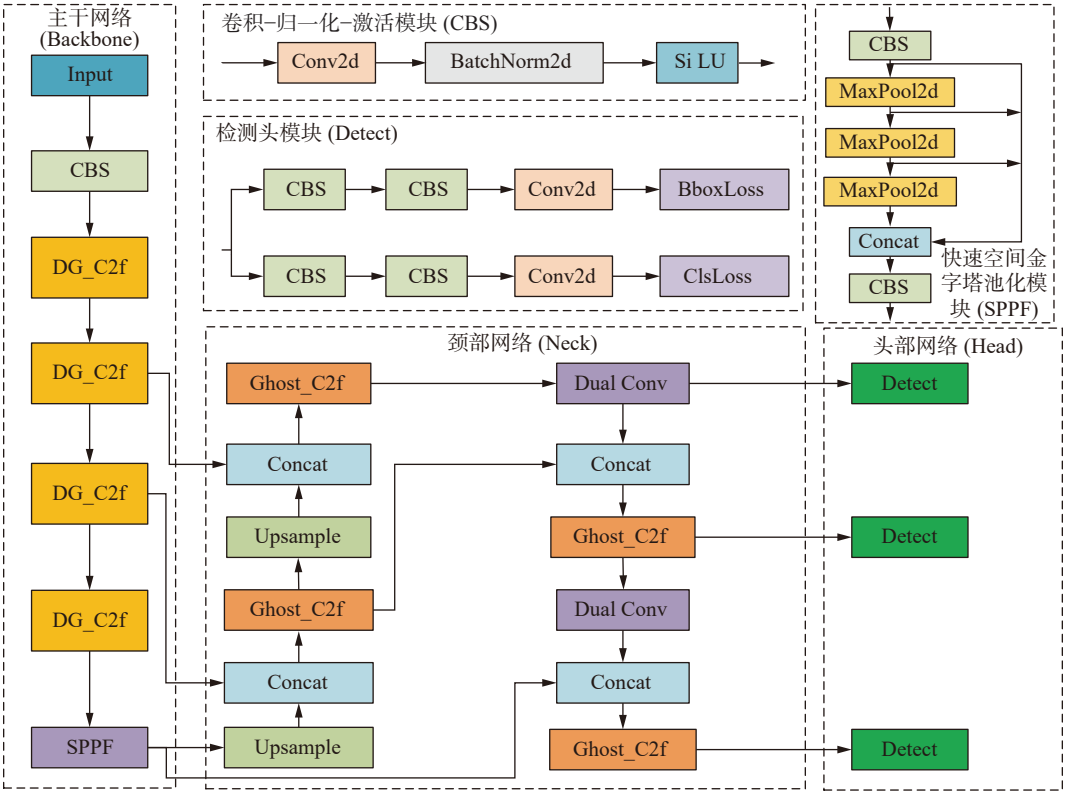


图 3　改进的 YOLOv8n 网络模型结构

Fig.3　Improved YOLOv8n network model structure

大时，计算成本的降低可以使得运算过程加速。其计算公式为

$$R_{\mathrm{DC/SC}}=\frac{F_{\mathrm{LDC}}}{F_{\mathrm{LSC}}}=\frac{1}{G}+\frac{1}{K^2} \tag{5}$$

DualConv 基本结构如图 4 所示。首先，根据组卷积的思想，将卷积核分为 G 组，每组卷积核处理对应的一组输入特征映射通道；然后，根据深度可分离卷积和异构卷积的思想，3×3 和 1×1 卷积核在输入特征映射通道上并行布局。G_1 组卷积核上端部分的一些卷积核同时进行 3×3 卷积（蓝色卷积块）和 1×1 卷积（橙色卷积块），在之后 G_2、G_3、…、G_x 不同的分组中同时进行 3×3 和 1×1 卷积，对应的特征映射通道不同，有助于同时利用大尺寸卷积核的空间特征提取能力和小尺寸卷积核的计算效率，从而在保持准确性的同时减少模型的参数数量和计算成本。

2.1.2　利用 GhostNetV2 重构 C2f 模块

为了实现高精度的检测指标，大多数深度学习网络模型具有较多的参数和较高的计算需求，这使得它们难以部署到边缘设备上，轻量级网络模型的出现有效解决了这一问题。

GhostNetV2 是一种新型的轻量级网络，能够

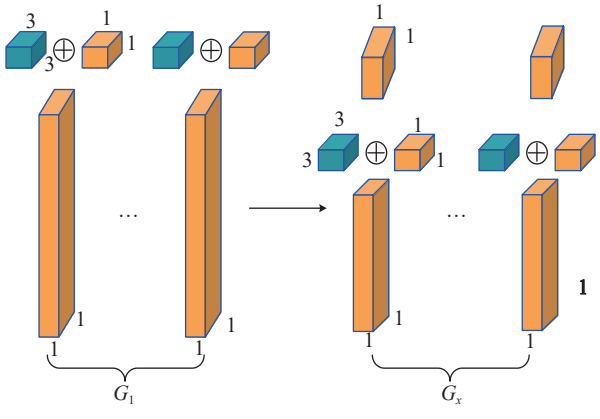


图 4　DualConv 基本结构

Fig.4　Basic structure of DualConv

在保证检测精度的同时，实现更轻量化、更快的检测速度。轻量级网络 GhostNetV2 主要由一个 DFC（decoupledfully connected）模块和两个 Ghost Module 组成。

Ghost Module 基本结构如图 5 所示，通过使用较少的计算量来生成更多的特征映射。首先对源图像进行传统的卷积计算得到本征特征图，将部分的本征特征图映射（φ_1、φ_2、φ_3）保留；然后，通过一系列的简单线性变换，得到 Ghost 特征图；最后，将其拼接至一起生成最终的特征图并输出。

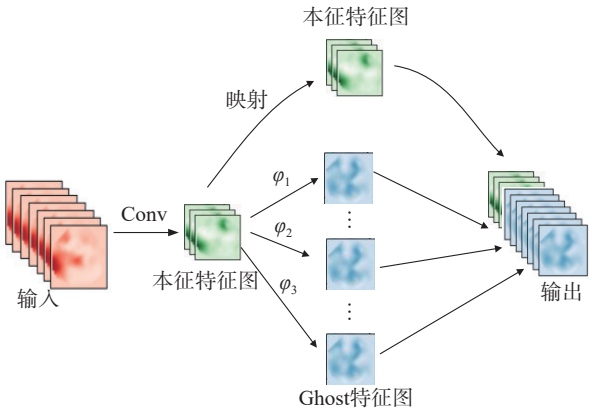


图 5 Ghost Module 基本结构

Fig.5 Basic structure of Ghost Module

DFC 模块通过捕捉不同空间像素之间的长期依赖关系，提高轻量级模型的表达能力。DFC 模块的输入为图像的原始特征 $Z\in\mathbb{R}^{H\times W\times C}$，其中，$H$、$W$、$C$ 分别为高度、宽度、通道数。在水平、垂直方向上进行特征聚集，以此减少全连接层中的计算，并实现对远程依赖关系的获取，计算公式为

$$a'_{hw}=\sum_{h'=1}^{H}F^{H}_{h,h'w}\odot z_{h'w},\quad h=1,2,\cdots,H,\quad w=1,2,\cdots,W \tag{6}$$

$$a_{hw}=\sum_{w'=1}^{W}F^{W}_{w,hw'}\odot a'_{hw'},\quad h=1,2,\cdots,H,\quad w=1,2,\cdots,W \tag{7}$$

式(6)表示水平方向的全连接过程，式中：$F^{H}_{h,h'w}$为可学习的变换权重；$z_{h'w}$表示输入特征图在位置(h',w)的特征向量；a'_{hw}表示水平方向聚合后的特征；h、w分别表示输出特征图的垂直、水平索引；h'、w'分别表示输入特征图的垂直、水平索引；$\odot$表示逐元素相乘。式(7)表示垂直方向的全连接过程，式中：$F^{W}_{w,hw'}$为另外一组可学习的变换权重；a_{hw}表示最终的注意力特征。DFC 模块通过共享部分权重，减少了计算复杂度。

GhostV2 bottleneck 结构如图 6 所示，为了增强扩展的特征，GhostNetV2 先将 DFC 模块与第一个 Ghost Module 并行，然后再将增强后的特征传递到第二个 Ghost Module，以生成相关的输出特征[14]。GhostNetV2 通过创新的 Ghost Module 设计，实现了在减少计算量和参数量的同时，保持模型的性能，特别适用于资源受限的环境。

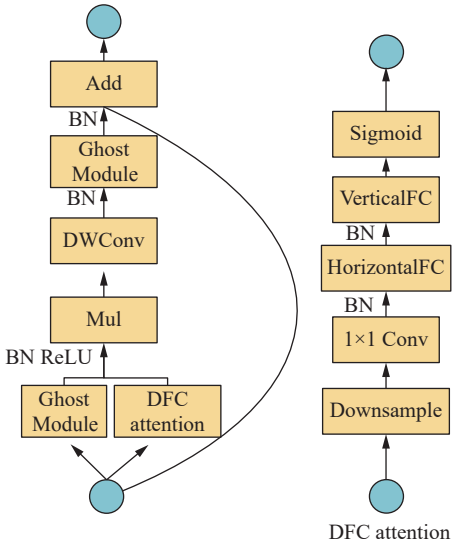


图 6 GhostV2 bottleneck 基本结构

Fig.6 Basic structure of GhostV2 bottleneck

本文利用 GhostNetV2 重构 YOLOv8n 中的 C2f 模块，在 C2f 中叠加两个 GhostV2 bottleneck 结构，形成全新的 Ghost_C2f 架构，其基本结构如图 7 所示。

在颈部网络部分，将 YOLOv8n 颈部网络中的 C2f 模块替换为设计的 Ghost_C2f 模块；在主干网络部分，Ghost_C2f 替换全部的 C2f 模块，与 DualConv 卷积模块融合形成 DG_C2f 模块，并嵌入到主干网络中。在保证模型精度的同时，大幅度地降低网络模型的参数量和计算量，加快网络的检测速度。

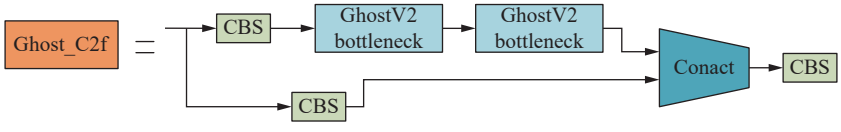


图 7 Ghost_C2f 基本结构

Fig.7 Basic structure of Ghost_C2f

2.2 模型剪枝

目前，深度神经网络在各种检测任务方面已经表现出了优秀性能，但是深度神经网络对于硬件设备的算力、内存等都有巨大的消耗，应用具

有一定的难度[15]。本文提出的改进 YOLOv8n 网络模型已经实现了一定的轻量化操作，但是对于边缘设备的部署，还有一定的差距。

Filter pruning 是一种结构化剪枝策略，通过计算每个卷积核的权重重要性评分来决定哪些卷积核可以被剪除[16]。Filter pruning 采用了 L1 范数作为卷积核的重要性评分。在卷积神经网络中，一个卷积核可以看作是一个三维张量，用 $F\in\mathbb{R}^{C\times k\times k}$ 表示。其中：C 为输入通道数；$k\times k$ 为卷积核的空间维度。对于每个卷积核 $\boldsymbol{F}$，将其视为一个由所有权重组成的向量，计算其 L1 范数为

$$\|\boldsymbol{F}\|_1=\sum_{c=1}^{C}\sum_{i=1}^{k}\sum_{j=1}^{k}\left|F_{c,i,j}\right| \tag{8}$$

式中，$F_{c,i,j}$ 表示卷积核在通道 c、位置(i,j)处的权重。

L1 范数衡量了卷积核所有权重的总强度，L1 范数越大意味着贡献越明显，反之则可以认为对网络结构的影响较小，可以优先被剔除。

Filter pruning 在单一层中对卷积核的 L1 范数进行评估，最后将其按照 L1 范数进行排序，将 L1 范数较小的卷积核视为对网络结构不重要的卷积核，并将其剔除。在剪枝完成之后，对网络进行再训练微调，以此恢复剪枝带来的性能损失。

本文采用 Filter pruning 剪枝策略对改进后的模型进行剪枝操作。模型剪枝方法能够在显著减少计算量和参数量的同时，基本保持模型的性能目标，提升模型的实际应用价值。

3　基于改进的 RexNet 手部关键点检测

3.1　轻量化网络 RexNet

RexNet 网络是在原有的 MobileNetV2[17] 轻量化网络基础上进一步改进得到的轻量化网络模型，可以在参数量减小的前提下消除特征表示瓶颈问题[18]。RexNet 主要由倒残差结构、线性瓶颈结构、挤压激励模块 SE[19](squeeze-and-excitation) 组成。这 3 部分形成 RexNet 轻量型网络的基础模型结构，如图 8 所示。

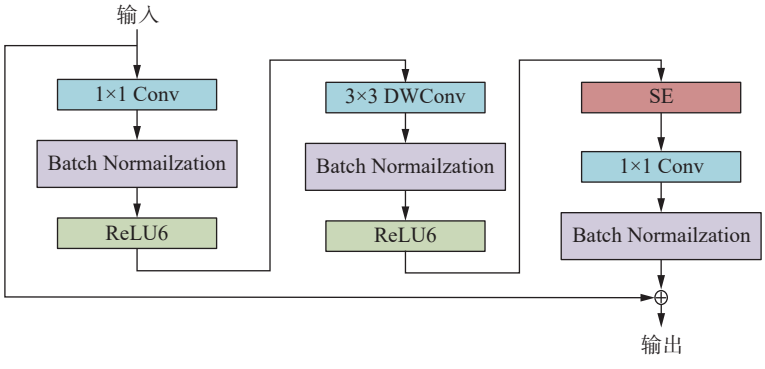


图 8　瓶颈模块基本架构

Fig.8　Basic structure of bottleneck residual block

3.2　改进的 RexNet 网络

3.2.1　Huber_Loss 损失函数

在手术室场景中，手部关键点检测因其目标小、数量多、背景环境复杂等特点，极易出现误检、关键点预测不到位等问题。Huber_Loss 常用于回归问题，是一种带参损失函数[20]。它结合了均方误差和绝对误差的优点，能增强平方误差损失函数对离群点的鲁棒性，同时有效减少异常值对模型的影响。在本文提出的关键点检测任务中，相比于均方误差，Huber_Loss 对离群点具有更强的抗干扰性。它在确保模型准确找到全局最优值的同时，能够更快速地更新模型参数，计算公式为

$$L_\delta(y,f(x))=\begin{cases}\dfrac{1}{2}(y-f(x))^2, & |y-f(x)|\leqslant\delta\\ \delta\left(|y-f(x)|-\dfrac{1}{2}\delta\right), & \text{其他}\end{cases} \tag{9}$$

式中：δ 为 Huber_Loss 的参数，本文取 δ=1.0；y 为真实值；$f(x)$ 为模型的预测值。

4　实验设计及结果分析

4.1　实验硬件配置

本文所有实验均在同一平台上进行，相关配置如表 1 所示。

表 1 实验配置

Tab.1 Experimental configuration

配置项	具体型号
操作系统	Ubuntu20.04
CPU	12 vCPU Intel(R) Xeon(R)Platinum
GPU	NVIDIA GeForce RTX 4090(24GB)
编译环境	Python3.8+Pycharm专业版
学习框架	PyTorch 2.0.0+CUDA11.8

4.2 数据集

改进的 YOLOv8n 网络模型实验使用 HaGRID[21] 数据集进行动态手势检测模型训练，对动态手势区域进行分类和定位。数据选自于 HaGRID 数据集，为精简样本，选取了共 18000 张图像，训练集、验证集、测试集按 7∶2∶1 的比例划分，检测类别共 19 种。HaGRID 数据集示例如图 9 所示。

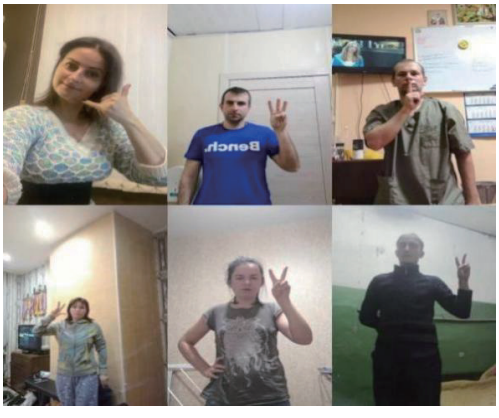


图 9 HaGRID 数据集示例

Fig.9 HaGRID dataset examples

改进 RexNet 网络实验使用的数据集包括网络图片和数据集 Large-scale Multiview 3D Hand Pose Dataset[22] 中重复度相对较低的图片，共 49062 个样本。数据集样本出自于 handpose_x 项目。数据集示例图如图 10 所示。

图 10 LM3DH 数据集示例

Fig.10 LM3DH dataset examples

4.3 评价指标

为了能够更好地评估改进模型和网络的有效性，在改进 YOLOv8n 网络模型部分，本文主要选取精确率（P）、召回率（R）、全部类别平均精度（mAP）、计算量（FLOPs）、参数量、模型权重大小、单张图片检测耗时作为实验结果的评价指标。在改进 RexNet 网络部分主要选取各个关键点损失的平均值作为实验结果的评价指标。

4.4 实验结果分析

4.4.1 改进的 YOLOv8n 网络模型消融实验

消融实验用于确定一个条件或参数对结果的影响程度。为了验证本文提出的改进模型的有效性，在本研究设置的实验条件和数据集配置条件下设计了消融实验。每一组实验的训练轮次设置为 100，学习率为 0.001，优化器设置为 Adam，超参数 batch size 设置为 16，超参数 workers 设置为 4，选取训练中最佳的权重在测试集上进行测试，得到的实验数据如表 2 所示。表中：“√”表示此项作了修改；“×”表示此项未作修改。

表 2 消融实验性能对比

Tab.2 Performance comparison of ablation experiment

方案	DualConv	Ghost_C2f	P	R	mAP@0.5	参数量/M	计算量/G	权重大小/M	单张图片检测耗时/ms
1	×	×	0.990	0.979	0.992	3.16	8.9	5.94	2.66
2	√	×	0.990	0.977	0.991	2.78	7.7	5.49	2.63
3	√	√	**0.986**	**0.975**	**0.991**	**1.90**	**5.5**	**3.85**	**2.33**

从表 2 可以看出，在 YOLOv8n 网络中引入 DualConv 后，精确率保持不变，召回率仅降低 0.2%，mAP@0.5 仅降低了 0.1%，参数量下降 12%，计算量下降 13.5%，权重大小降低 7.6%，单张图片检测耗时略微降低了 1%。利用 GhostNetV2 重构 C2f 结构并将 DG_C2f 架构嵌入到网络中后，精

确率、召回率有略微的下降。但与基线模型相比，其参数量下降了 39.9%，计算量下降了 38.2%，权重大小降低了 35.2%，单张图片检测耗时降低了 12.4%，模型的 mAP@0.5 依然能达到 99.1% 的识别准确度。这些结果证明了改进模型在实现轻量化基础上保证了检测的准确度，计算量降低至只有 5.5 G，更加有利于在边缘设备中的部署。

部分测试集的检测结果如图 11 所示。从图中可以看出，在复杂的背景环境、人体肤色与手部肤色相近、不同光照环境等影响因素下，测试集中各类手势都能够正确地识别并精确定位，同时图像中的无关手也能精确识别出。例如在图 11 中，粉色预测框就能把无效手精确检测出来，且在复杂背景下具备良好的鲁棒性。

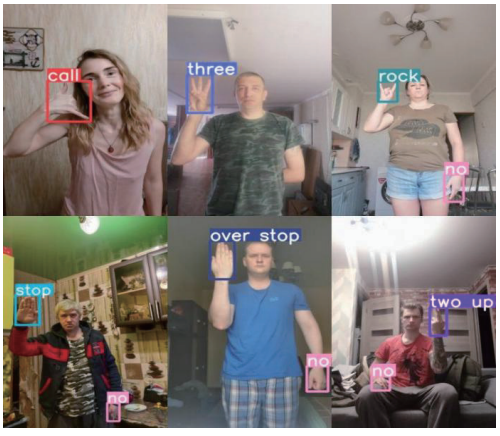


图 11　部分测试集检测结果

Fig.11　Detection results of partial test sets

4.4.2　Ghost_C2f 模块放置位置对比实验

在主干网络部分，Ghost_C2f 模块替代了原主干网络中的 C2f 模块，并与 DualConv 模块组合形成 DG_C2f 结构；在颈部网络部分，Ghost_C2f 模块替代了所有 C2f 模块。为了验证上述 Ghost_C2f 模块放置方案的有效性，在实验条件及数据集配置下设计了对比实验，每组实验训练轮次设为 100，学习率为 0.001，优化器为 Adam，batch size 为 16，workers 为 4。选取最佳权重在测试集上进行测试，实验结果如表 3 所示。具体的实验方案如下：1 号方案为仅替换主干网络中的 C2f 模块；2 号方案为仅替换颈部网络中的 C2f 模块；3 号方案为仅替换主干网络与颈部网络连接部分的 3 个 C2f 模块；4 号方案为仅替换颈部网络与检测头连接部分的 3 个 C2f 模块；5 号方案为同时替换主干网络与颈部网络连接部分的 3 个 C2f 模块以及颈部网络与检测头连接部分的 3 个 C2f 模块；6 号方案为本文提出的方案，即将主干和颈部网络中的所有 C2f 模块替换为 Ghost_C2f 模块。

从表 3 中的实验数据可以看出，本文提出的优化方案在保持最低参数量和计算量的同时，精度指标变化较小，验证了本文提出的 Ghost_C2f 模块置换优化方案（6 号方案）的有效性。

4.4.3　模型剪枝方法对比实验

为使模型适合在资源有限的边缘设备上部署，需要在保证检测精度的情况下，降低模型的大小，本文运用 Filter pruning 剪枝策略对改进后的模型进行剪枝处理。同时在实验中应用了 LAMP[23]、Group_taylor[24] 等多种模型剪枝策略进行对比。实验结果如表 4、图 12 所示。

从表 4 中可以看出，在前文改进模型的基础上，采用 Filter pruning 剪枝方法并设置剪枝率为 2.0 后，模型在精度与大小之间达到了最优的平衡效果。相比于未剪枝的基础模型，参数量减少了 70.9%，计算量降低了 69.7%，模型权重下降了 66.2%，检测耗时降低了 41.3%，动态手势识别的准确率达到了 99%，同时，计算量降低至只有 2.7 G，更加有利于在边缘设备中的部署。在图 12 中，实线为改进后模型各项指标，虚线为对改进模型剪

表 3　Ghost_C2f 模块放置位置对比

Tab.3　Comparison of Ghost_C2f module placement positions

方案	*P*	*R*	mAP@0.5	参数量/M	计算量/G	权重大小/M	单张图片检测耗时/ms
1	0.988	0.979	0.991	2.40	6.5	4.92	2.8
2	0.984	0.970	0.990	2.58	7.3	5.13	2.9
3	0.990	0.977	0.991	2.56	7.1	5.10	2.9
4	0.988	0.981	0.992	2.66	7.8	5.28	2.9
5	0.986	0.977	0.990	2.20	6.4	4.42	2.6
6（本文）	**0.983**	**0.967**	**0.989**	**1.90**	**5.5**	**3.85**	**2.5**

表 4 剪枝实验结果对比分析

Tab.4 Comparative analysis of pruning experiment results

是否改进	剪枝策略	*P*	*R*	mAP@0.5	参数量/M	计算量/G	权重大小/M	单张图片检测耗时/ms
×	×	0.990	0.979	0.992	3.16	8.9	5.94	2.66
√	×	0.986	0.975	0.991	1.90	5.5	3.85	2.33
√	LAMP-2.0	0.983	0.969	0.990	0.92	2.7	2.01	1.57
√	LAMP-3.0	0.977	0.966	0.988	0.62	1.8	1.44	1.43
√	Group_taylor-2.0	0.983	0.978	0.990	0.92	2.7	2.01	1.65
√	Group_taylor-3.0	0.979	0.966	0.989	0.62	1.8	1.44	1.43
√	**Filter pruning-2.0**	**0.986**	**0.978**	**0.990**	**0.92**	**2.7**	**2.01**	**1.56**
√	Filter pruning-3.0	0.972	0.966	0.987	0.62	1.8	1.44	1.44

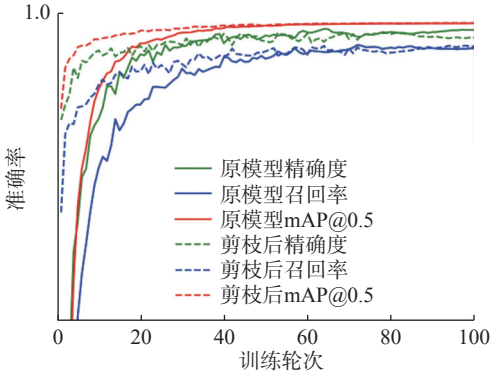


图 12 剪枝后微调模型对比

Fig.12 Comparison of fine tuned models after pruning

枝并 Fine tuning 后的模型各项指标。从图中可以看出，模型剪枝完成后在 Fine tuning 过程中，虽然检测精度略有下降，但收敛速度更快。

4.4.4 不同光照环境对比实验

在手术室实际的应用场景中，不同的光照环境可能对手势的识别造成影响。为了验证本文提出的模型在不同的光照环境下的鲁棒性，在实验条件和数据集配置条件下设计了对比实验。首先，将测试集的图片进行不同程度的曝光处理，作为实验检测对象；然后应用本文提出的模型对其进行检测，得到相关的数据指标，具体的实验结果如表 5 所示。

从表 5 中可以看出，在曝光强度变化为±50%的情况下，模型检测结果的各项指标并没有太大的变化，验证了本文提出的模型在不同的光照环境下的鲁棒性，能够应对实际应用中手术室环境下光照环境的变化。

4.4.5 与其他目标检测算法的对比实验

为更有效地综合评估本文所提改进模型的有效性，将改进并且剪枝后的模型与原始的

表 5 不同光照环境对比

Tab.5 Comparison of different lighting environments

光照强度	*P*	*R*	mAP@0.5
−50%	0.982	0.971	0.987
−40%	0.985	0.974	0.989
−20%	0.986	0.978	0.990
–	0.986	0.978	0.990
20%	0.983	0.970	0.988
40%	0.967	0.929	0.970
50%	0.917	0.848	0.907

YOLOv5n、YOLOv8n、YOLOv8s、YOLOv9c、YOLOv8n（ShuffleNetv2[25]为主干）、Faster-Rcnn[26]、SSD[27]等目标检测算法在 4.1 的实验条件、4.2 的数据集配置条件下进行对比实验，具体实验结果如表 6 所示。

表 6 不同算法性能对比

Tab.6 Performance comparison of different algorithms

模型	mAP@0.5	参数量/M	计算量/G	权重大小/M	单张图片检测耗时/ms
SSD	0.907	25.46	61.3	117.00	50.00
Faster_Rcnn	0.921	41.39	215.6	316.00	90.90
YOLOv5n	0.988	2.51	7.2	5.30	2.36
YOLOv8n	0.992	3.16	8.9	5.94	2.66
YOLOv8s	0.992	11.14	28.7	22.50	5.99
YOLOv9c	0.994	25.54	103.8	51.60	18.18
YOLOv8n（ShuffleNetv2）	0.986	1.86	5.7	3.46	2.32
本文	**0.990**	**0.92**	**2.7**	**2.01**	**1.56**

从表 6 中可以看出，本文提出的模型与传统

目标检测算法 Faster-Rcnn、SSD 相比，无论是在检测精度还是在轻量化方面都具有明显优势；与同类型的 YOLOv5n、YOLOv8n、YOLOv8s、YOLOv9c、YOLOv8n（ShuffleNetv2 为主干）等相比，精度与大部分模型相当，同时参数量、计算量、权重大小、单张图片检测耗时均是各个模型中的最优方案，有利于模型在资源有限的边缘设备上进行部署。

4.4.6　嵌入式边缘设备部署

本文提出的优化 YOLOv8n 模型在轻量化方面取得了显著的效果，适合在资源有限的边缘设备上部署。为了验证本文所提模型在边缘设备上部署的优势，进行了一系列的边缘设备部署实验。实验采用 NVIDIA 公司生产的 Jetson orin nano 人工智能开发板作为基础边缘设备，采用塔菲克 1080p 高清 USB 摄像头采集视频流画面，并在 NVIDIA 推出的用于深度学习推理加速的高性能推理引擎 TensorRT 的辅助下，验证本文所提优化 YOLOv8n 模型的轻量化优势。具体实验数据如表 7 所示，表中，FPS（frames per second）表示每秒帧数。

表 7　嵌入式边缘设备部署实验

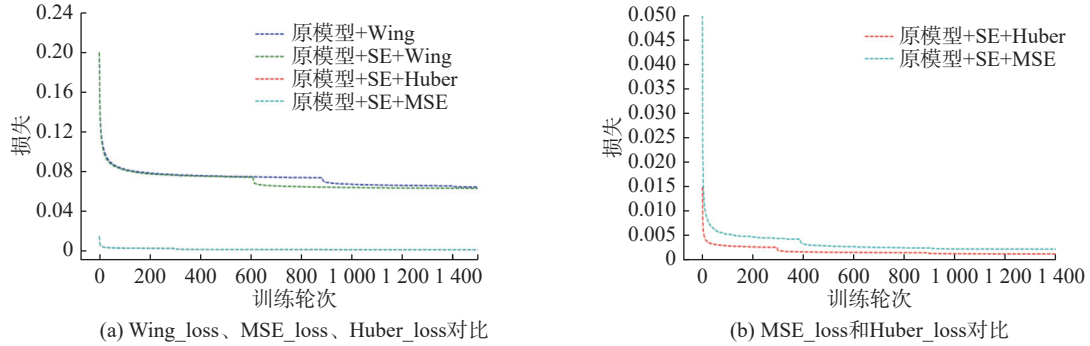
Tab.7　Embedded edge device deployment experiment

模型	是否应用 TensorRT 加速	FPS
YOLOv8n	×	26
YOLOv8n	√	51
本文	√	**65**

从表 7 中可以看出，在 TensorRT 的加速辅助下，检测的速度有了明显提升，证明了 TensorRT 在边缘设备部署加速中的有效性。同时，在此基础上部署本文提出的优化模型，与原模型相比，检测速度得到了一定的提升，适用于在实际场景下的边缘设备部署。

4.4.7　改进的 RexNet 网络消融实验

为了验证本文提出的基于 RexNet 的手部关键点检测网络的有效性，在实验条件和数据集配置条件下设计了消融实验。训练轮次设置为 1 500，学习率设置为 0.001，batch size 设置为 8，workers 设置为 2，实验结果如图 13 所示。



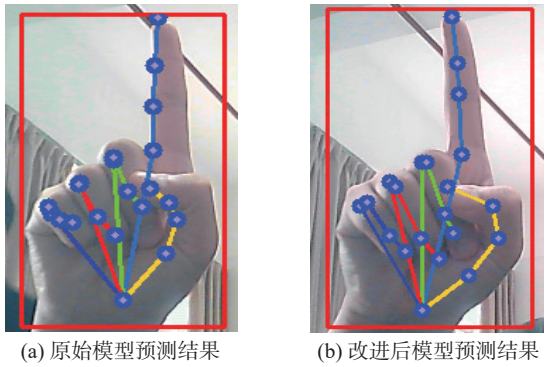
(a) Wing_loss、MSE_loss、Huber_loss对比　　(b) MSE_loss和Huber_loss对比

图 13　应用不同的损失函数后模型的损失变化趋势对比

Fig.13　Comparison of loss trend of models after applying different loss functions

从图 13 中可以看出，应用 SE 注意力机制的瓶颈模块使模型损失略有下降，表明 SE 注意力机制对模型的收敛性具有积极影响。在此基础上，本文对 Wing_loss、MSE_loss 和 Huber_loss 这 3 种损失函数进行了对比实验。实验结果表明，MSE_loss 和 Huber_loss 相比 Wing_loss 表现更优。进一步对比发现，Huber_loss 的损失值约为 MSE_loss 的一半，验证了本文采用 Huber_loss 的优越性。

图 14 为在实时摄像头下对动态手势手部关键点的检测结果。由图可以看出，改进后的关键点检测网络可以更好地预测出准确的关键点位置。



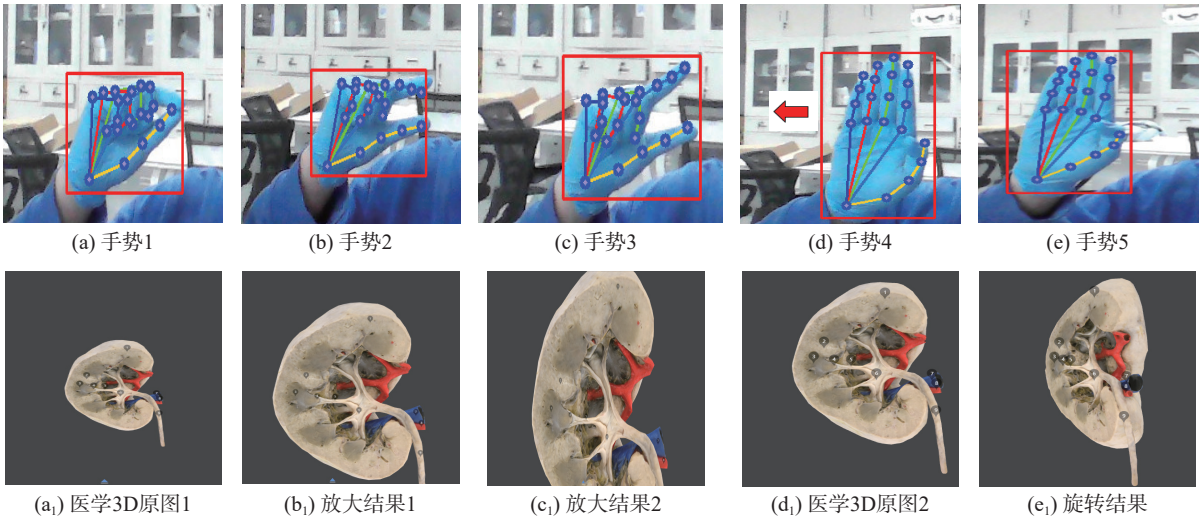
(a) 原始模型预测结果　　(b) 改进后模型预测结果

图 14　改进前后动态手势关键点检测效果对比

Fig.14　Comparison of dynamic gesture key point detection effects before and after improvement

图 15 展示了应用本文算法在实时摄像环境下，佩戴常用外科手术手套进行动态手势识别和手部关键点检测，并控制医疗设备上的肾脏三维结构图像的结果。在实际应用中，系统能够根据用户的动态手势执行相应操作。

例如，当在画面中识别到放缩手势时，同时提取和分析食指指尖和大拇指指尖的坐标位置信息，并计算两者之间的距离变化。当食指和大拇指指尖之间的距离增大(缩小)时，相应地放大(缩小)图像，实现对医疗设备显示器上图像的实时放大和缩小，便于术中医生观察图像细节。当在画面中识别到旋转手势时，以手部尾指顶部的关键点为基准，提取该关键点的坐标位置信息。当手部移动时其坐标位置信息对应发生变化时，根据其坐标位置信息的变化对应控制画面中的图像进行旋转，实现对医疗设备显示器上图像的实时旋转，便于术中医生转换图像的观察视角。具体操作变化如图 15 所示。



(a) 手势1 (b) 手势2 (c) 手势3 (d) 手势4 (e) 手势5

(a_1) 医学3D原图1 (b_1) 放大结果1 (c_1) 放大结果2 (d_1) 医学3D原图2 (e_1) 旋转结果

图 15 动态手势识别及手部关键点检测结果

Fig.15 Dynamic gesture recognition and hand key point detection results

5 结 论

本文提出了一种基于 YOLOv8n 和轻量级网络 RexNet 的动态手势识别方法，改进并优化了 YOLOv8n 模型，结合 RexNet 实现了动态手势控制医疗设备。针对医疗环境中的动态手势人机交互问题，本文提出的算法在保证检测精度的同时，实现了轻量化，适用于资源有限的边缘设备部署，为手术室环境中的人机交互方式提供了新的思路和解决方案。

参考文献：

[1] 孟青云, 戴佳蔚, 查佳佳, 等. 基于 YOLOv8 算法的常用手势识别 [J]. 现代仪器与医疗, 2023, 29(4): 12−20.

[2] 项融融, 李博, 赵桥. 基于 Leap Motion 手势识别的三维交互系统 [J]. 电子设计工程, 2024, 32(1): 44−48.

[3] 胡学政, 陶庆, 赵暮超, 等. 采用迁移学习的表面肌电信号手势识别方法 [J]. 科学技术与工程, 2024, 24(12): 5044−5050.

[4] 洪姣, 张雪娇, 刘国瑞, 等. 基于 Kinect 动态手势识别的改进 DTW 算法研究与实现 [J]. 天津城建大学学报, 2024, 30(2): 133−137.

[5] WANG F, AO X H, WU M, et al. Explainable deep learning for sEMG–based similar gesture recognition: a shapley-value-based solution[J]. Information Sciences, 2024, 672: 120667.

[6] 彭玉青, 赵晓松, 陶慧芳, 等. 复杂背景下基于深度学习的手势识别 [J]. 机器人, 2019, 41(4): 534−542.

[7] PENG Z L, YANG Z D, XIAHOU J B, et al. Recognizing teachers' hand gestures for effective non-verbal interaction[J]. Applied Sciences, 2022, 12(22): 11717.

[8] HAN D, YUN S, HEO B, et al. Rethinking channel dimensions for efficient model design[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville: IEEE, 2021: 732–741.

[9] ZHONG J C, CHEN J Y, MIAN A. DualConv: dual convolutional kernels for lightweight deep neural networks[J]. IEEE Transactions on Neural Networks and Learning Systems, 2023, 34(11): 9528−9535.

[10] TANG Y H, HAN K, GUO J Y, et al. GhostNetv2: enhance cheap operation with long-range attention [C]//Proceedings of the 36th International Conference on

Neural Information Processing Systems. New Orleans: Curran Associates Inc., 2022, 35: 9969–9982.

[11] HOWARD A G, ZHU M L, CHEN B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications[EB/OL]. (2017-04-17)[2024-08-27]. https://doi.org/10.48550/arXiv.1704.04861.

[12] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[C]//Proceedings of the 26th International Conference on Neural Information Processing Systems. Lake Tahoe: Curran Associates Inc., 2012: 1097–1105.

[13] SINGH P, VERMA V K, RAI P, et al. HetConv: heterogeneous kernel-based convolutions for deep CNNs[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach: IEEE, 2019: 4830–4839.

[14] 周思瑜, 徐慧英, 朱信忠, 等. 基于改进 YOLOv8n 的手机屏幕瑕疵检测算法: PGS-YOLO[J]. 计算机工程, 2025, 51(5): 326−339.

[15] 孙阳, 李佳. 基于通道剪枝的 YOLOv7-tiny 输电线路异物检测算法 [J]. 计算机工程与应用, 2024, 60(14): 319−328.

[16] LI H, KADAV A, DURDANOVIC I, et al. Pruning filters for efficient ConvNets[EB/OL]. (2017-03-10)[2024-08-27]. https://doi.org/10.48550/arXiv.1608.08710.

[17] SANDLER M, HOWARD A, ZHU M L, et al. MobileNetV2: inverted residuals and linear bottlenecks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 4510–4520.

[18] 孙楠, 杨煜戎, 杨哲, 等. 机器视觉中小目标检测实验优化模型设计与实现 [J]. 实验室研究与探索, 2023, 42(3): 32−39.

[19] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 7132–7141.

[20] LAINA I, RUPPRECHT C, BELAGIANNIS V, et al. Deeper depth prediction with fully convolutional residual networks[C]//2016 Fourth International Conference on 3D Vision (3DV). Stanford: IEEE, 2016: 239–248.

[21] ALEXANDER K, KARINA K, ALEXANDER N, et al. HaGRID – HAnd gesture recognition image dataset[C]//2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Waikoloa: IEEE, 2024: 4560–4569.

[22] GOMEZ-DONOSO F, ORTS-ESCOLANO S, CAZORLA M. Large-scale multiview 3D hand pose dataset[J]. Image and Vision Computing, 2019, 81: 25−33.

[23] LEE J, PARK S, MO S, et al. Layer-adaptive sparsity for the magnitude-based pruning[EB/OL]. (2021-05-09)[2024-08-27]. https://doi.org/10.48550/arXiv.2010.07611.

[24] MOLCHANOV P, MALLYA A, TYREE S, et al. Importance estimation for neural network pruning[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach: IEEE, 2019: 11256–11264.

[25] MA N N, ZHANG X Y, ZHENG H T, et al. ShuffleNet V2: practical guidelines for efficient CNN architecture design[C]//Proceedings of the 15th European Conference on Computer Vision – ECCV 2018. Munich: Springer, 2018: 122–138.

[26] 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.

[27] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]//Proceedings of the 14th European Conference on Computer Vision – ECCV 2016. Amsterdam: Springer, 2016: 21–37.

（编辑：丁红艺）