

# 改进 YOLOv8n-Pose 的形变 QR 码校正与识别<sup>①</sup>



刘云<sup>1,2</sup>, 邹复民<sup>1,2</sup>, 蔡祈钦<sup>2</sup>, 李俊清<sup>3</sup>, 钟继雄<sup>3</sup>

<sup>1</sup>(福建理工大学 电子电气与物理学院, 福州 350118)

<sup>2</sup>(福建省汽车电子与电驱动技术重点实验室, 福州 350118)

<sup>3</sup>(福建中科兰剑智能装备科技有限公司, 福州 350108)

通信作者: 刘云, E-mail: [2221905047@smail.fjut.edu.cn](mailto:2221905047@smail.fjut.edu.cn)

**摘要:** 针对 QR 码图像检测过程中因环境复杂性和拍摄角度变化等因素导致 QR 码读取率低的问题, 本文提出一种基于改进 YOLOv8n-Pose 的形变 QR 码校正与识别算法. 首先, 在主干网络引入高效通道注意力机制 (efficient channel attention, ECA) 模块, 该模块通过不降维的方式实现跨通道交互, 有效提升网络的特征提取能力和检测精度. 其次, 采用 Slim-neck 架构重构颈部网络, 减少模型的复杂度, 提高对不同尺度 QR 码的检测能力. 最后, 通过模型检测到的 QR 码角点, 结合逆透视变换对 QR 码进行校正, 并使用 ZBar 算法进行读取. 实验结果表明, 在公开的 QR 码数据集上, 改进的算法相比原算法, mAP50 和 mAP50-95 分别提升 1.6% 和 1.1%, 模型参数数量和模型计算量分别降低 6.5% 和 9.5%, 在 CPU 和 GPU 上检测速度分别提升 0.3 f/s 和 0.7 f/s, 达到 14.2 f/s 和 59.6 f/s, 能够高效地满足 QR 码角点检测需求. 此外, 在自制的形变 QR 码数据集上, 基于改进 YOLOv8n-Pose 的 QR 码识别方法相比单独使用 ZBar 算法的 QR 码识别方法, QR 码读取率提高 23.66%, 达到 87.41%. 该方法仅需拍摄一张照片就可识别所有货物的信息, 能够有效提高货物管理的效率.

**关键词:** QR 码; 关键点预测; YOLOv8n-Pose; 注意力机制; 定位与校正

引用格式: 刘云, 邹复民, 蔡祈钦, 李俊清, 钟继雄. 改进 YOLOv8n-Pose 的形变 QR 码校正与识别. 计算机系统应用. <http://www.c-s-a.org.cn/1003-3254/9715.html>

## Deformed QR Code Correction and Recognition Based on Improved YOLOv8n-Pose

LIU Yun<sup>1,2</sup>, ZOU Fu-Min<sup>1,2</sup>, CAI Qi-Qin<sup>2</sup>, LI Jun-Qing<sup>3</sup>, ZHONG Ji-Xiong<sup>3</sup>

<sup>1</sup>(School of Electronics, Electrical Engineering and Physics, Fujian University of Technology, Fuzhou 350118, China)

<sup>2</sup>(Fujian Provincial Key Laboratory of Automotive Electronics and Electric Drive Technology, Fuzhou 350118, China)

<sup>3</sup>(Fujian Zkljan Intelligent Equipment and Technology Co. Ltd., Fuzhou 350108, China)

**Abstract:** To address the problem of low QR code reading rates caused by complex environments and changes in shooting angles during QR code detection, this study proposes an algorithm for correcting and recognizing deformed QR codes based on an improved YOLOv8n-Pose algorithm. First, the efficient channel attention (ECA) module is introduced into the backbone network. This module achieves cross-channel interaction without dimensionality reduction, effectively enhancing the feature extraction capabilities and detection accuracy of the network. Secondly, the Slim-neck architecture is adopted to reconstruct the neck network, reducing model complexity and improving the detection capability for QR codes of different scales. Finally, detected QR code corner points are used for correction through inverse perspective transformation, and the corrected QR codes are read using the ZBar algorithm. Experimental results show that, on a public QR code dataset, the improved algorithm increases mAP50 and mAP50-95 by 1.6% and 1.1%, respectively, compared to the original algorithm. Model parameters and computational costs are reduced by 6.5% and 9.5%, respectively. Detection

<sup>①</sup> 基金项目: 福建理工大学科研计划 (GYZ20072)

收稿时间: 2024-05-22; 修改时间: 2024-06-28; 采用时间: 2024-07-11; csa 在线出版时间: 2024-10-31

speed on CPU and GPU is improved by 0.3 f/s and 0.7 f/s, reaching 14.2 f/s and 59.6 f/s, respectively, meeting the requirements for efficient detection of QR code corner points. In addition, on a custom-made dataset of deformed QR codes, the proposed method based on the improved YOLOv8n-Pose algorithm enhances the QR code reading rate by 23.66% compared to the standalone ZBar algorithm, achieving a recognition rate of 87.41%. This method only requires one photo to recognize all the information about the goods, which can effectively improve the efficiency of goods management.

**Key words:** QR code; keypoint prediction; YOLOv8n-Pose; attention mechanism; location and correction

快速响应码具有信息容量大、可靠性强、纠错性高和成本低等优点,在生产、医疗、物流、销售和支付等领域得到广泛应用<sup>[1,2]</sup>。随着拍摄角度的变化,附着在任何物体上的QR码可能出现模糊、旋转、倾斜或反光等问题。所以准确地定位、分割和校正图像中的QR码是识别过程中的关键步骤。

传统QR码识别算法主要通过其形态特征来定位。张民等<sup>[3]</sup>利用QR码的3个位置探测图形对其定位,并通过双线性变换来校正形变QR码,但是如果形变程度过大,则难以准确定位QR码。吴邦政<sup>[4]</sup>采用形态学方法来定位形变QR码的角点,并进行校正,但是如果图像中存在噪点,角点的定位会存在误差,造成校正后的QR码仍存在形变。王雄华等<sup>[5]</sup>通过7条直线从图像的7个方向搜索QR码的3个角点,采用间隔抽样和斜率容错确定第4个角点,并通过逆透视变换来校正形变QR码,但是在复杂背景下效果不佳。传统算法依赖于低效的手工特征提取,没有充分利用QR码的特征信息,仅在简单背景和形变程度不大的单一QR码上具有较好的识别效果,难以适应复杂自然场景中不同形态的QR码识别需求。

深度学习作为机器学习领域中的一种新兴研究方向,通过自动提取图像的多层次特征信息,极大地提高QR码检测的效率。Zhao等<sup>[6]</sup>提出基于双金字塔结构的分割网络(barcode network),由先验金字塔池化模块(prior pyramid pooling module, P3M)和金字塔细化模块(pyramid refine module, PRM)组成,用于检测日常生活中的条形码和QR码。Chen等<sup>[7]</sup>通过多阶段逐步判别(multistage stepwise discrimination, MSD)和压缩移动网络(compressed mobile network),实现对多个QR码进行准确和快速的检测。Jia等<sup>[8]</sup>利用卷积神经网络(convolutional neural network, CNN)算法检测

QR码的4个角点,有效识别形变的QR码。叶枫等<sup>[9]</sup>通过改进的YOLOv4算法对多个QR码进行定位与分割,并结合ZBar算法实现在复杂背景下批量识别图像中的QR码。王伟彤等<sup>[10]</sup>提出一种轻量化的Centernet网络,使用CSPDarkNet53-tiny作为主干网络并添加空间金字塔池化(spatial pyramid pooling, SPP)模块,以提高对QR码的检测精度。这些方法使用深度学习来检测QR码,解决了传统QR码识别算法在复杂场景下适应性差的问题,但是大部分算法只考虑到QR码检测的准确性,缺乏对形变QR码的校正。

为了更高效地识别形变的QR码,本文提出一种基于改进YOLOv8n-Pose的形变QR码校正与识别算法。该算法能定位、分割和校正图像中处于不同位置和大小形变的QR码,提升了复杂背景下多个形变QR码的识别准确率和读取率。

## 1 YOLOv8n-Pose 关键点检测算法

YOLOv8支持多种计算机视觉任务,包括目标检测、实例分割、关键点检测、图像分类和旋转目标检测。针对不同场景需求,YOLOv8推出了5种不同参数的模型:n、s、m、l、x,其中nano版专为边缘识别部署设计,拥有最小参数量及最快检测速度。本文使用YOLOv8n-Pose模型来高效地定位QR码和其关键点。YOLOv8n-Pose的网络结构主要包括4个部分,如图1所示。

(1) 输入端(Input): 输入部分负责将输入图像缩放到训练所需的大小,并进行数据预处理和增强操作。预处理包括图像的归一化和缩放,确保输入大小的一致性。对于数据增强包括旋转、翻转遮挡、裁剪、色调调整、Mosaic<sup>[11]</sup>和Mixup<sup>[12]</sup>等,通过这些操作能够有效地提升模型的精度。



励网络 (squeeze-and-excitation network, SE-Net)<sup>[18]</sup> 的改进网络, 使用一维卷积取代 SE-Net 中的全连接层 (fully connected layer, FC), 所以 ECA 无需降维操作即可就可实现跨通道交互, 减轻 SE-Net 对网络预测性能的负面影响, 同时降低模型的计算量和复杂度, 并显著提升了特征提取能力和检测精度。

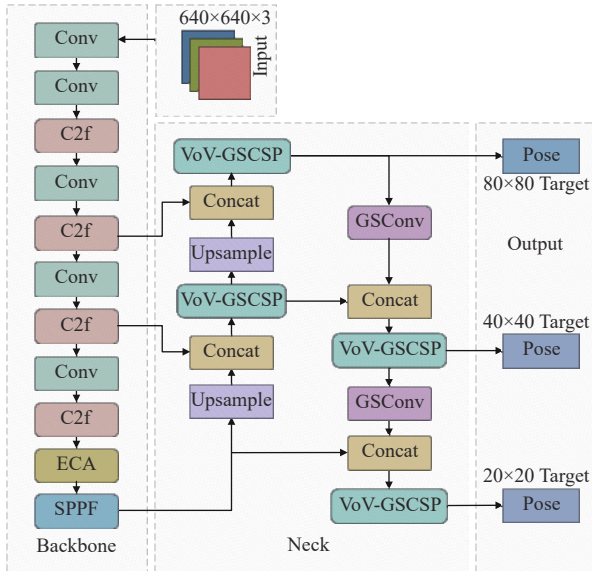


图 2 改进的 YOLOv8n-Pose 网络结构图

ECA 注意力模块如图 3 所示, 其工作原理: 首先, 通过全局平均池化 (global average pooling, GAP) 将输入特征图的每个通道简化为标量, 获取通道的全局信息。如式 (1) 所示:

$$y = \frac{1}{H \times W} \sum_a^H \sum_b^W x_i(a, b) \quad (1)$$

其中,  $y$  为全局特征,  $x_i$  为输入尺寸大小为  $H \times W$  的第  $i$  个特征图。

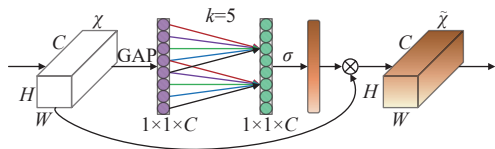


图 3 ECA 注意力模块

其次, 一维卷积层通过对 GAP 输出的特征向量进行线性变换, 学习不同通道间的交互权重, 从而捕捉通道间的依赖关系。如式 (2) 所示:

$$w = \sigma(C1D_k(y)) \quad (2)$$

其中,  $\sigma$  为 Sigmoid 激活函数,  $w$  为通道权重。

最后将通道权重向量与输入特征逐通道相乘, 得到具有通道注意力的特征, 其中每个通道的特征被赋予不同的重要度, 实现对重要特征的区分和提取。为提高局部交互效率, 采用通道维度  $C$  的映射自适应确定  $k$  的值, 进而优化特征提取过程, 如式 (3) 所示:

$$k = \psi(C) = \left\lfloor \frac{\log 2C}{\gamma} + \frac{b}{\gamma} \right\rfloor_{\text{odd}} \quad (3)$$

其中,  $\gamma$ 、 $b$  为调节参数, 取值为 2 和 1,  $k$  为卷积核大小,  $|t|_{\text{odd}}$  表示与  $t$  最接近的奇数。

尽管大量研究证明了将注意力机制引入卷积神经网络能显著提升计算机视觉任务的性能, 但是关于如何选择注意力机制在网络中最佳嵌入位置以优化网络性能的问题, 目前还没有统一的答案。所以本文采用高效的 ECA 注意力模块与 YOLOv8n-Pose 相结合构建了 4 种不同的网络结构, 如图 4 所示。将 ECA 模块嵌入主干网络的 SPPF 模块前一层的方式称为 A 方式; 将 ECA 模块嵌入主干网络的 SPPF 模块后一层的方式称为 B 方式; 将 ECA 模块嵌入 Pose 检测头的前一层的方式称为 C 方式; 将 ECA 模块嵌入主干网络的 C2f 模块中的 Add 操作之前的方式称为 D 方式。

## 2.2 轻量化 Slim-neck 架构

### (1) GSConv 卷积

鬼影混洗卷积 (ghost-shuffle conv, GSConv) 利用空间压缩和通道扩展保持语义信息, 同时通过维护通道间连接, 有效防止密集卷积可能引起的特征提取效率下降, 但是过度使用则会增加网络深度和推理时间。所以仅在颈部特征融合网络使用 GSConv 是最佳选择, 此阶段通道维度达到最大, 宽度维度达到最小, 特征图呈现细长状态, 不再需要进一步变换处理。此时使用不仅能够达到最优的工作效率, 而且计算成本相较于标准卷积可降至 60%–70%, 能够实现计算成本的显著降低。

GSConv 结构如图 5 所示, 首先, 使用标准卷积实现跨通道信息交互; 其次, 采用分组卷积提高结构效率; 最后, 将两种卷积操作的特征图级联并进行通道混洗, 通过重新排列特征通道来提高特征之间的信息流动性, 最终输出增强后的特征图。通过这种方式减少冗余信息, 实现网络的轻量化, 同时提升检测精度。

### (2) VoV-GSCSP 模块

GSBottleneck 模块通过 GSConv 降低模块的复杂度和参数量, 同时通过残差连接与特征融合加强特征

之间的传递,优化梯度之间的流动,有效防止训练过程中的信息丢失.结合GSbottleneck模块和1次聚合(one-shot aggregation, OSA)方法设计名为VoV-GSCSP的高效跨级部分网络(GSCSP)模块<sup>[19]</sup>,如图6所示.首

先,将输入特征图分为两部分,一部分特征通过Conv提取,另一部分则通过GSbottleneck模块进行更细致的特征提取.接着,两部分通过Conv级联,实现跨通道的信息交互,从而增强网络对特征信息的获取能力.

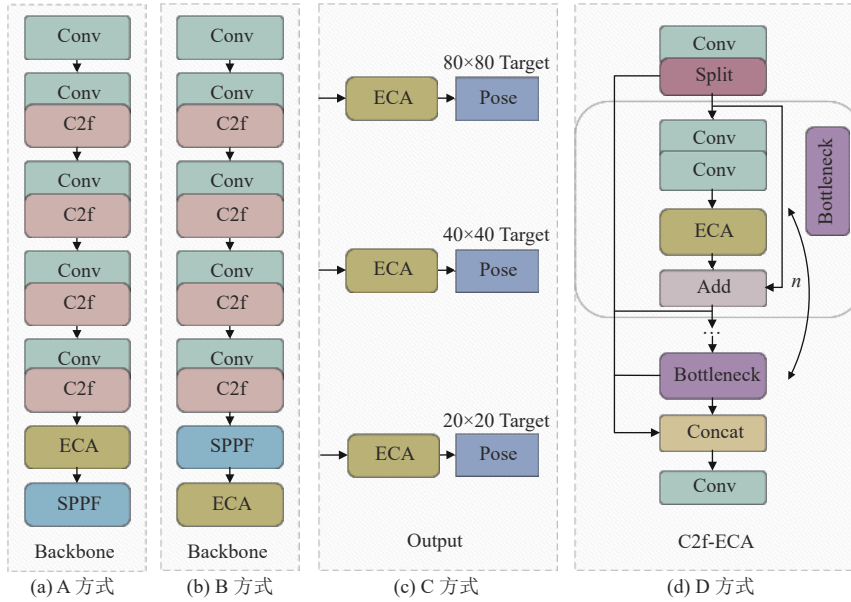


图4 ECA注意力模块4种嵌入方式

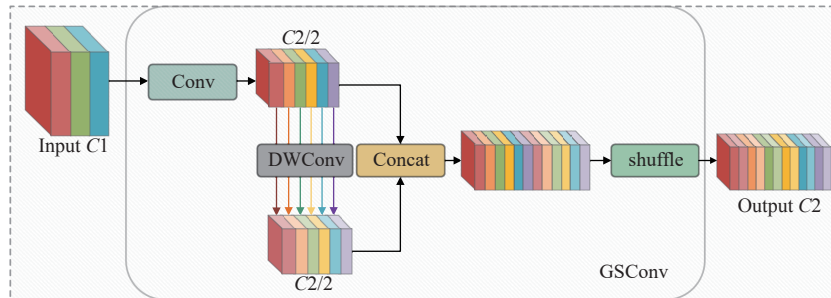


图5 GSConv卷积结构图

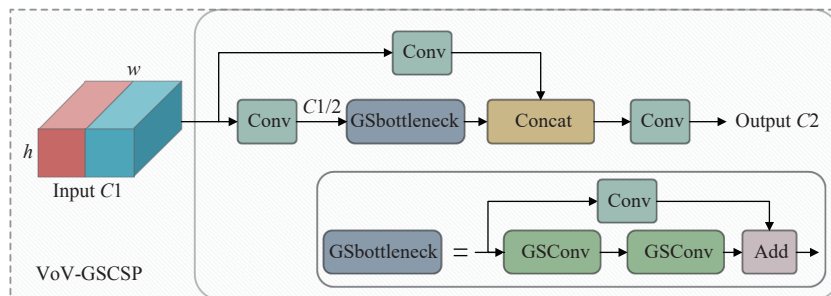


图6 VoV-GSCSP模块结构图

采用轻量化卷积GSConv和VoV-GSCSP模块重构YOLOv8n-Pose的颈部网络,命名为Slim-neck.首先,通过上采样和拼接操作将不同尺度的特征图结合;

接着,通过GSConv进行轻量化和高效的特征提取.随后,使用VoV-GSCSP模块来进一步对特征进行提取和融合.最后,通过Pose对目标进行检测.通过这种模块化

和分层的方法使 Slim-neck 架构能提取到更精细的特征,有效防止网络退化,同时提高模型对不同尺度 QR 码的检测能力和泛化能力。

### 3 基于 YOLOv8n-Pose 的 QR 码识别方法

#### 3.1 QR 码图像预处理

##### 3.1.1 图像灰度化

采集到的 QR 码图像通常为彩色,这在 QR 码读取过程中会消耗大量的内存和时间,而将彩色图像转换为灰度图像,可以有效提高 QR 码读取效率。所以采用加权平均法对 QR 码图像进行灰度化处理。该方法通过为 RGB 各通道施加不同权重计算灰度值,从而完成图像的灰度化处理,如图 7 所示。

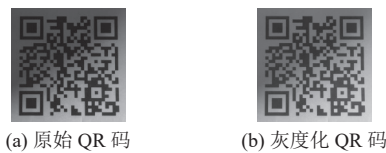


图 7 图像灰度化前后对比

##### 3.1.2 图像滤波去噪

在采集 QR 码图像过程中,环境光照和摄像头的抖动等不确定因素会导致图像中带有大量噪声。光学上,这些噪声被视为椒盐噪声或脉冲噪声,这会对 QR 码的读取产生不利影响。所以采用中值滤波处理灰度化的 QR 码图像,能够有效去除特定频率的噪声并提高图像质量,如图 8 所示。使用 Python 模拟 QR 码采集过程中椒盐噪声的干扰。通过对比滤波前后的图像可以明显看到,中值滤波有效减少椒盐噪声干扰,使图像更加平滑。



图 8 图像中值滤波前后对比

##### 3.1.3 图像对比度增强

在采集 QR 码图像过程中,相机离焦或光线较弱会导致图像模糊,为了增强图像细节和提高成像效果,需要采用图像增强技术来提高图像的对比度。所以,对滤波后的 QR 码图像采用直方图均衡化处理,通过非线性地调整灰度值,使像素均匀分布至 0-255 动态范围,能够有效提升对比度和层次感。如图 9 所示,其中图 9(c) 和图 9(d) 为图像直方图,横坐标为图像像素的

灰度值范围,纵坐标则为相应灰度数值在图像中呈现的次数,用于反映图像对比度的基本情况。

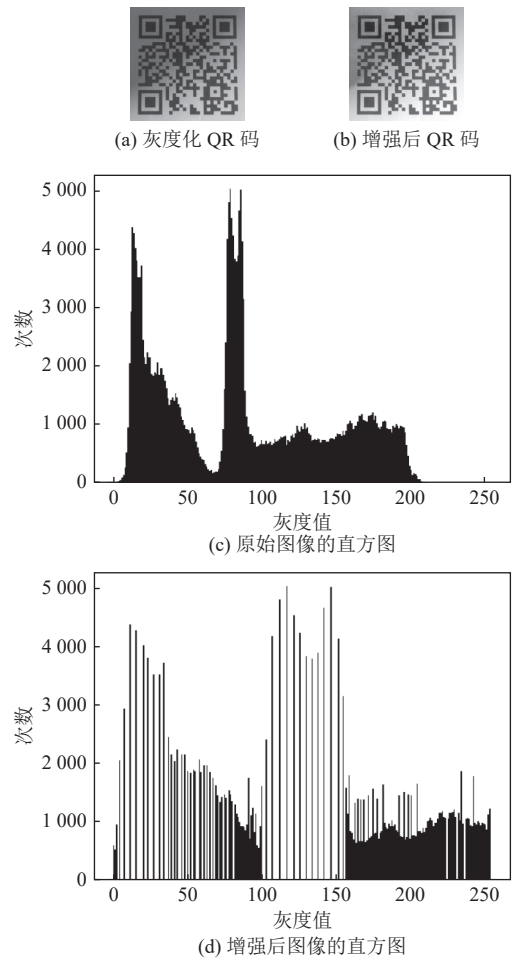


图 9 图像直方图均衡化前后对比

##### 3.1.4 图像二值化

在采集 QR 码图像过程中,不同种类和位置的光源会导致图像存在光照不均的现象。使用 Otsu 算法对图像进行二值化,虽然在光照均匀时效果较好,但是 Otsu 算法仅考虑整体图像像素分布,这会导致图像二值化阈值的唯一性。所以在光照不均的情况下,二值化后的图像可能会出现黑色方块,导致 QR 码重要信息的丢失。为了提高 QR 码的读取率,采用分子块与 Otsu 结合的方法处理对比度增强后的 QR 码,9 子块 Otsu 算法的原理如图 10 所示,该算法将图像进行均匀的分块,减少固定阈值对图像二值化过程影响。

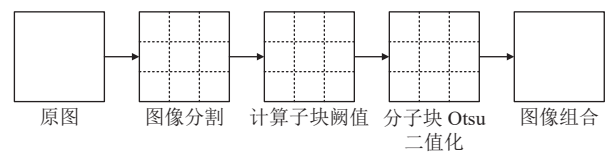


图 10 9 子块 Otsu 算法的原理图

当 QR 码分割块数不足时, 其子块中的阈值不能使图像中 QR 码的纹路完整分割, 如图 11(a)、图 11(b) 和图 11(c) 都存在不同程度的信息丢失, 而当分割块数为 16 块时, 如图 11(e) 所示, QR 码的分割效果最优, 其次是分割块数为 9 块时, 如图 11(d) 所示, QR 码的分割效果仅次于 16 块, 且不存在明显的信息丢失. 考虑增加分割块数会提高运算成本, 所以采用 9 子块 Otsu 算法对 QR 码进行二值化处理.



图 11 图像二值化结果

### 3.2 QR 码图像校正处理

透视变换是使用变换公式将图像投影到一个新的视平面上, 这过程也可称为投影映射. 通用变换公式如式 (4) 所示:

$$\begin{pmatrix} x' \\ y' \\ w' \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} u \\ v \\ w \end{pmatrix} \quad (4)$$

在齐次坐标系中, 二维平面的点  $(u, v)$  表示为  $(u, v, w)$ , 其中  $w$  通常被设置为 1.  $(u, v)$  表示原图像的

坐标点. 当要对图像进行校正时, 可采用逆透视变换来校正 QR 码. 该方法的原理如图 12 所示. 任意四边形的坐标点  $(u_i, v_i)$  经过式 (5) 坐标转换转换成  $x$  和  $y$  坐标系上的点  $(x_i, y_i)$ , 从而得出校正后图像.

$$\begin{cases} x = \frac{x'}{w'} = \frac{a_{11}u + a_{12}v + a_{13}}{a_{31}u + a_{32}v + a_{33}} \\ y = \frac{y'}{w'} = \frac{a_{21}u + a_{22}v + a_{23}}{a_{31}u + a_{32}v + a_{33}} \end{cases} \quad (5)$$

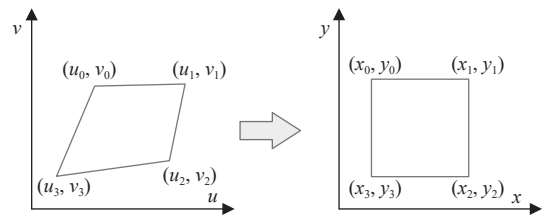


图 12 逆透视变换

### 3.3 QR 码识别流程

ZBar 是一个广泛用于扫描 QR 码和条形码的开源库, 能同时检测图像中多个 QR 码, 但是识别准确率不高, 存在部分形变 QR 码难以识别, 而本文提出的 QR 码识别方法通过多尺度的预测, 能够准确识别不同尺寸的 QR 码和精准定位其角点, 同时对 ZBar 算法读取困难的 QR 码进行校正, 以提高 QR 码读取率, QR 码识别流程如图 13 所示.

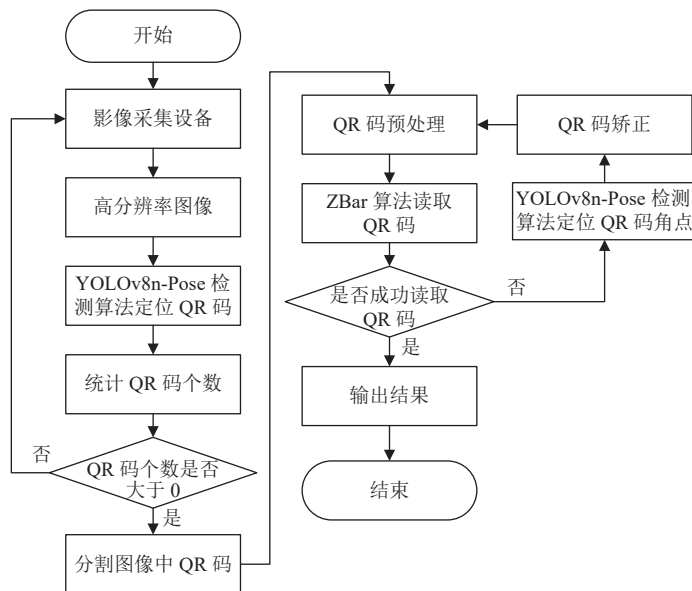


图 13 基于 YOLOv8n-Pose 的 QR 码识别流程

以一张包含多个形变 QR 码的高分辨率图像为例, 具体的 QR 码识别过程如图 14 所示. 首先, 通过改进

的 YOLOv8n-Pose 检测算法预测 QR 码的边框和角点. 然后, 对图像中的 QR 码进行分割和预处理. 接着, 使

用 ZBar 算法进行读取. 最后, 对于 ZBar 算法难以读取的 QR 码, 利用其角点坐标进行逆透视变换校正, 校正后的 QR 码经过图像预处理算法后再次使用 ZBar 算法进行读取.

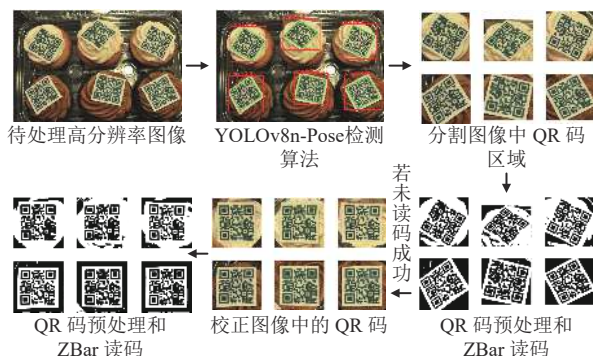


图 14 基于 YOLOv8n-Pose 的 QR 码识别过程

## 4 算法实验

### 4.1 实验数据集

本文使用百度飞桨 AI Studio 平台提供的 QR 码数据集, 该数据集包含了多种环境下的 QR 码图像. 考虑到本文的研究目的, 从 1994 张原始图像中挑选出 1059 张受光线变化、复杂背景和拍摄角度变化影响的 QR 码图像, 形成新的数据集. 针对形变 QR 码角点定位的问题, 通过图像旋转、添加噪声以及调整亮度和对比度等方法对数据集进行扩充, 最终获得 2 118 张图像. 其中对比度调整的方法采用 Gamma 变换<sup>[20]</sup>. 数据集使用 LabelMe 软件标注, 按照 8:1:1 比例分为训练集、验证集和测试集. 为了增强数据集图像的质量, 采用 Mosaic 数据增强方法, 从训练集选取 4 张图像进行拼接, 增加了目标数量, 让小目标分布更加均匀, 同时丰富了检测到的物体的背景, 其效果如图 15 所示.

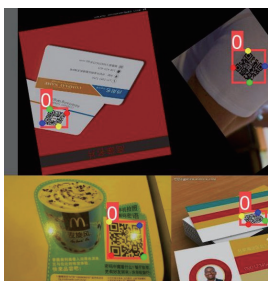


图 15 Mosaic 数据增强

### 4.2 实验环境与参数设定

本文的实验环境是 64 位的 Windows 10, 采用

PyTorch 1.12.0 深度学习框架, CUDA 11.6, 编程语言为 Python 3.8.18. 硬件配置包括 AMD Ryzen 7 5800H CPU 和 NVIDIA GeForce RTX 3060 GPU, 改进前后算法使用的超参数相同, 输入图像尺寸为 640×640, 采用 SGD 优化器优化神经网络, 采用余弦退火策略调整学习率, 初始学习率 (lr0) 为 0.01, 权重衰减率 (weight decay) 为 0.0005, 动量参数 (momentum) 为 0.937. 训练轮次 (epoch) 为 300, 批次大小 (batchsize) 为 8.

### 4.3 评价指标

选取平均精度均值 (mAP)、模型参数量 (Params)、模型计算量 (FLOPs)、每秒处理的图像帧数 (FPS) 作为模型改进前后性能评价的客观指标. OKS<sup>[16]</sup>用于衡量预测关键点与真实关键点之间的相似度, 取值范围为 0-1. 值越接近 1, 表明模型预测关键点与真实标注越相似, 其计算如式 (6) 如下:

$$OKS_p = \frac{\sum_i \exp(-d_{pi}^2/2s_p^2\sigma_i^2)\delta(v_{pi} > 0)}{\sum_i \delta(v_{pi} > 0)} \quad (6)$$

式中,  $p_i$  表示第  $p$  个目标的第  $i$  个关键点,  $d_{pi}$  表示关键点预测值与真实值之间的欧拉距离,  $s_p$  表示目标的尺度因子,  $v_{pi}$  表示关键点的可见性,  $\sigma_i$  表示第  $i$  个关键点标注值与真实值间的标准差.  $\delta$  为可见点计算函数. 其中 QR 码角点的  $\sigma_i$  取值按 YOLOv8n-Pose 的取值, 即  $\sigma_i = 1/n$ , 其中  $n$  为关键点个数.

OKS 用于计算关键点检测的 AP, 当 OKS 超过阈值 T 时, 检测结果被认定为真阳性. 精确率 (Precision) 指实际预测正确的关键点占有所有被预测为正确的关键点的比例, 召回率 (Recall) 指实际预测的正确关键点占有所有真实关键点的比例, 然后通过计算 P-R 曲线下的面积得出 AP. 其中 mAP50 是 OKS=0.50 时测得所有类别下的 AP 均值, mAP50-95 则是 OKS=0.50, 0.55, ..., 0.95 (步长为 0.05) 时所有类别的 AP 均值的平均值.

### 4.4 算法实验结果与分析

#### 4.4.1 注意力机制嵌入方式实验

为探究注意力机制在网络不同部位对模型性能的影响, 以第 2.1 节提出的 4 种不同网络与原网络进行对比, 实验结果如表 1 所示.

实验结果表明, 将注意力机制以 A 方式嵌入原网络中, 其 mAP50 和 mAP50-95 均高于其他改进的网络. 与原网络相比, mAP50 和 mAP50-95 分别提高 1.1% 和 0.7%. 通过在主干网络的 SPPF 模块前一层嵌入注



注意力机制,能够使改进网络在特征提取过程中获得更精细和更丰富的 QR 码角点特征,从而有效提升网络的检测效果,由此证明 YOLOv8n-Pose-A 嵌入方式的有效性。

表 1 注意力机制不同嵌入方式实验结果

网络模型	mAP50 (%)	mAP50-95 (%)	FPS (f/s)	
			CPU	GPU
YOLOv8n-Pose	95.8	93.2	13.9	58.9
+ECA-A	96.9	93.9	13.9	58.6
+ECA-B	95.9	93.1	13.8	58.4
+ECA-C	96.2	93.3	13.7	57.9
+ECA-D	96.3	93.5	13.7	56.5

#### 4.4.2 不同注意力机制实验

为探究不同注意力机制对模型性能的影响,以 YOLOv8n-Pose-A 的嵌入方式分别将 4 种主流注意力机制 CA、SE、ECA 和 EMA 融入到 YOLOv8n-Pose 网络中,并与原网络进行对比,实验结果如表 2 所示。

表 2 不同注意力机制实验结果

网络模型	mAP50 (%)	mAP50-95 (%)	FPS (f/s)	
			CPU	GPU
YOLOv8n-Pose	95.8	93.2	13.9	58.9
+CA	96.7	93.9	13.9	57.1
+SE	96.6	93.7	13.9	57.8
+ECA	96.9	93.9	13.9	58.6
+EMA	96.4	93.4	13.8	56.8

实验结果表明,相比较于原网络,改进后的 4 种网络在 mAP50 和 mAP50-95 上都有不同程度的提升,其中 ECA 注意力机制与 CA 注意力机制在关注 QR 码角点特征方面明显优于其他注意力机制,而两者相比较,ECA 注意力机制对检测精度提升的效果最为明显,且在检测速度上也高于其他改进网络。由此证明 ECA 注意力机制能够在保持检测速度的前提下,有效提高检测的精度。

#### 4.4.3 轻量化实验

为探究不同轻量化网络对模型性能的影响,分别选用 4 种主流轻量化网络对 YOLOv8n-Pose 网络进行轻量化改进,并与原网络进行对比,实验结果如表 3 所示。

表 3 不同轻量化网络实验结果

网络模型	mAP50 (%)	mAP50-95 (%)	FPS (f/s)		参数量 (M)	计算量 (G)
			CPU	GPU		
YOLOv8n-Pose	95.8	93.2	13.9	58.9	3.1	8.4
+Slim-neck	96.0	93.3	14.2	59.8	2.9	7.6
+GhostNetv2	95.5	92.2	14.1	57.7	2.7	7.3
+MobileNetv3	94.0	89.7	14.2	54.9	2.3	5.9
+ShuffleNetv2	93.2	88.5	18.3	62.8	1.8	5.2

实验结果表明,将 YOLOv8n-Pose 的主干网络替换 ShuffleNetv2 和 MobileNetv3 后,该模型的参数量和计算量明显减少,同时检测速度提升,但是检测精度有所降低。相比之下,采用 GhostNetv2 的 BottleNeck 替换主干网络 C2f 模块的 BottleNeck,虽然在轻量化方面差于前两种改进网络,但是 Ghost 模块通过标准卷积和廉价操作得到更多的特征,由此检测精度有相对的提升。采用 Slim-neck 架构重构颈部网络后,在检测精度和检测速度上明显高于其他改进网络。与原网络相比,mAP50 和 mAP50-95 分别提升 0.2% 和 0.1% 的,同时在轻量化上也取得显著的效果,模型参数量和模型计算量分别降低 6.5% 和 9.5%,在 CPU 和 GPU 上检测速度分别提升 0.3 f/s 和 0.9 f/s。由此证明 Slim-neck 架构相比其他轻量化网络具有更好的表现能力,在减少网络复杂度的同时,能够有效提高模型对 QR 码角点的识别精度和检测速度。

#### 4.4.4 消融实验

为了更好地验证同时引入 ECA 注意力机制和采用轻量级 Slim-neck 架构对改进 YOLOv8n-Pose 网络带来的性能提升和验证各组件的有效性,采用消融实验进行对比,实验结果如表 4 所示。

表 4 消融实验结果

网络模型	mAP50 (%)	mAP50-95 (%)	FPS (f/s)		参数量 (M)	计算量 (G)
			CPU	GPU		
YOLOv8n-Pose	95.8	93.2	13.9	58.9	3.1	8.4
+ECA	96.9	93.9	13.9	58.6	3.1	8.4
+Slim-neck	96.0	93.3	14.2	59.8	2.9	7.6
+ECA+Slim-neck	97.4	94.3	14.2	59.6	2.9	7.6

实验结果表明,在原网络中引入不同模块后,模型的各项指标都出现不同程度的提升,其中在原网络中同时引入 ECA 注意力机制与 Slim-neck 架构,能够使改进后的网络各项指标提升效果明显,其中 mAP50 和 mAP50-95 分别提高 1.6% 和 1.1%,达到 97.4% 和 94.3%,在 CPU 和 GPU 上检测速度分别提升 0.3 f/s 和 0.7 f/s,达到 14.2 f/s 和 59.6 f/s。由此证明 Slim-neck 架构可以显著提高 ECA 注意力机制的特征提取能力,使改进后的网络既满足轻量化的要求,同时也在检测精度和速度方面也有一定的优势,能够有效地满足 QR 码角点检测任务需求。

#### 4.4.5 对比实验

为了进一步比较本文对网络改进的效果,将改进前后的算法和目前主流的关键点检测算法 HRNet<sup>[21]</sup>、

YOLOv5s6-Pose<sup>[16]</sup>与 YOLOv7-w6-Pose<sup>[22]</sup>进行比较, 实验结果如表 5 所示。

表 5 对比实验结果

网络模型	mAP50 (%)	mAP50-95 (%)	FPS (f/s)		参数量 (M)	计算量 (G)
			CPU	GPU		
YOLOv8n-Pose	95.8	93.2	13.9	58.9	3.1	8.4
HRNet-W32	96.1	86.5	5.5	46.2	28.5	17.4
YOLOv5s6-Pose	96.9	94.1	7.1	51.1	15.7	25.4
YOLOv7-w6-Pose	97.8	94.6	2.9	38.6	80.2	101.6
改进算法	97.4	94.3	14.2	59.6	2.9	7.6

实验结果表明, 本文改进的算法在检测精度和速度的综合衡量下展现出比其他算法更好的性能, 并在模型参数量、计算量和检测速度方面均表现最佳。虽然在检测精度稍低于 YOLOv7-w6-Pose 算法, 但是在轻量化和检测速度方面具有明显优势。综上所述, 改进网络在极少量参数和极快的检测速度的前提下, 保证更高的检测精度, 与其他算法相比具有更高的通用性和实用价值。

#### 4.4.6 检测结果分析

如图 16 所示, 选取 3 张具有代表性的图像进行实验, 其中第 1 行图像受到光线强度变化的影响, 第 2 行图像受到复杂背景和拍摄角度变化的影响, 第 3 行图像受到微小目标和多目标等影响。使用 Grad-CAM 的方法生成关键点检测的热力图, 以可视化的方式展示模型对图像不同区域的关注程度。在热力图中, 颜色越红的区域表示模型对该区域分配的相对权重越大, 对该区域的信息关注程度越高。

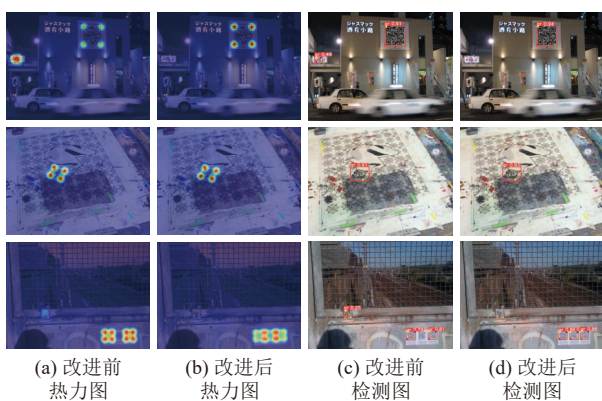


图 16 改进前后模型检测结果对比

可视化检测结果表明, 受光线强度变化、多目标和小目标的影响下, 改进网络在 QR 码的检测效果上具有明显优势, 解决原网络漏检和误检的问题, 同时提高 QR 码边框定位的准确性, 使其更接近于真实 QR 码

的边框, 边缘偏差也相对减小。受复杂背景和拍摄角度变化的影响下, 改进网络在 QR 码角点定位上比原网络更接近于真实 QR 码角点。这些检测结果进一步验证了算法改进的可行性。

#### 4.4.7 QR 码校正实验

本文选取 3 张发生形变的 QR 码进行实验, 将训练生成的 PyTorch 模型转为 ONNX 格式的模型进行加速推理, 并通过逆透视变换对 QR 码进行校正实验。图 17 为实验结果展示。实验结果表明, 在复杂背景下, 对不同数量和不同角度拍摄的 QR 码, 该算法的检测与校正结果都表现良好, 这对 QR 码的读取具有一定的帮助, 并进一步证明算法的通用性和实用价值。



图 17 QR 码校正实验结果

## 5 QR 码读取实验

### 5.1 形变 QR 码数据集

如图 18 所示, 本实验使用 H 级容错的 QR 码, 包含内容“Happy, New Year!”“Thinking of you!”和“Happy, birthday!”的 3 张 QR 码。为研究本文方法对形变 QR 码的读取率, 对 QR 图像进行随机角度的透视变换并并顺时针旋转 0°、45°、135°、225°和 315°共 5 次, 每种 QR 码生成 200 张形变图像, 共生成 600 张图像。



图 18 数据集 QR 码

为研究本文方法对日常环境中的形变 QR 码读取率, 使用 SAM 算法<sup>[23]</sup>对形变 QR 码进行语义分割并获

取掩膜,随机将1-3张QR码掩膜添加到日常场景图像中,以提高QR码读取的复杂度,共生成300张图像,如图19所示。



图19 模拟场景QR码

为研究本文方法对日常环境中QR码读取率,使用爬虫技术在百度图库随机下载400张QR码图像,该数据集分辨率大小不一,拍摄环境复杂多样.将3种数据集合并后,总计1300张图像。

## 5.2 QR码读取实验结果与分析

为进一步验证本文方法对QR码读取率的提升效果.选择4种不同的QR码读取方法在不同的数据集上进行读取实验,如表6所示.其中YOLOv8n代表QR码定位,YOLOv8n-Pose代表QR码定位和校正。

表6 QR码读取实验结果

数据集	图像数 (张)	QR码 (个)	读取比例(%)			
			ZBar	YOLOv8n +ZBar	YOLOv8n +Hough +ZBar	YOLOv8n- Pose +ZBar
形变QR码	600	600	64.17	75.67	88.33	91.83
模拟场景	300	437	48.74	62.47	65.45	82.61
真实场景	400	552	75.18	79.71	82.97	86.41
总数	1300	1589	63.75	73.44	80.18	87.41

由表6可知,在形变QR码数据集中,本文方法通过对QR码区域分割,有效提高其读取率.此外,再对读取困难的QR码进行校正,可以进一步提升读取率。

在模拟场景的形变QR码数据集中,由于仅对QR实体进行分割,这可能出现与模拟场景中的背景相连通.同时模拟场景中的背景噪声会使得QR码边缘出现毛刺和缺口,进而使得Hough变换无法准确检测直线,导致在定位QR码角点时存在误差,使校正后的图像仍存在形变.相比之下,本文方法能在复杂背景中准确识别QR码和定位其角点,并对其进行有效的校正,从而提高QR码读取率。

从3种数据集的QR码读取率比较中,本文方法具有明显优势,相比于其余3种方法,QR码读取率分别提高23.66%、13.97%和7.23%.这表明在复杂背景中读取多个QR码以及校正形变QR码方面,本文方法

具有高鲁棒性和实用性。

## 6 结论与展望

(1)本文在YOLOv8n-Pose网络上同时引入ECA注意力机制和Slim-neck架构,由此提出一种高效的QR码角点检测算法.ECA注意力机制能够增加对QR码角点的关注度,有效提升网络的特征提取能力和检测准确性.Slim-neck架构能够提高网络对不同尺度QR码的检测能力.改进后的模型在QR码数据上展现出良好的QR码角点检测能力,mAP50和mAP50-95分别达到97.4%和94.3%,同时在CPU和GPU上检测速度分别提升到14.2 f/s和59.6 f/s,在满足实时性的同时,能够提高QR码角点检测的准确性,有助于校正形变QR码。

(2)现有的目标检测算法主要使用单个矩形框来定位目标,无法校正倾斜的目标.所以本文借鉴人体姿态估计方法,提出一种基于改进YOLOv8n-Pose的形变QR码校正与识别算法.该方法能够在复杂背景下提升多个形变QR码的读取率.然而,在QR码磨损、遮挡、运动模糊或分辨率低的情况下,对其校正后仍难以读取.所以下一步将继续深入研究QR码的校正算法,以适应更多复杂场景。

## 参考文献

- 朱万山,刘兴刚,崔永玲,等.城市公共交通中非接触支付技术研究进展.电子测量技术,2020,43(9):121-126.[doi:10.19651/j.cnki.emt.1903917]
- de Seta G. QR code: The global making of an infrastructural gateway. Global Media and China, 2023, 8(3): 362-380. [doi:10.1177/20594364231183618]
- 张民,郑建立.基于符号特征的QR码识别算法.计算机工程,2011,37(4):278-280
- 吴邦政. QR码图像恢复与快速识别技术的研究与实现[硕士学位论文].福州:福州大学,2013.
- 王雄华,张昕,朱同林.快速定位的QR码校正方法.计算机工程与设计,2015,36(8):2185-2190.[doi:10.16208/j.issn1000-7024.2015.08.035]
- Zhao QJ, Ni F, Song Y, et al. Deep dual pyramid network for barcode segmentation using barcode-30k database. arXiv:1807.11886, 2018.
- Chen RJ, Huang HX, Yu YX, et al. Rapid detection of multi-QR codes based on multistage stepwise discrimination and a compressed MobileNet. IEEE Internet of Things Journal,

- 2023, 10(18): 15966–15979. [doi: [10.1109/JIOT.2023.3268636](https://doi.org/10.1109/JIOT.2023.3268636)]
- 8 Jia J, Zhai GT, Ren P, *et al.* Tiny-BDN: An efficient and compact barcode detection network. *IEEE Journal of Selected Topics in Signal Processing*, 2020, 14(4): 688–699. [doi: [10.1109/JSTSP.2020.2976566](https://doi.org/10.1109/JSTSP.2020.2976566)]
- 9 叶枫, 林好青. 复杂背景下的 QR 码批量定位与识别算法. *浙江工业大学学报*, 2023, 51(4): 355–361. [doi: [10.3969/j.issn.1006-4303.2023.04.001](https://doi.org/10.3969/j.issn.1006-4303.2023.04.001)]
- 10 王伟彤, 李长峰, 朱华波, 等. 轻量化 CenterNet 网络的二维条码定位算法. *电子测量与仪器学报*, 2022, 36(5): 128–135. [doi: [10.13382/j.jemi.B2104821](https://doi.org/10.13382/j.jemi.B2104821)]
- 11 Bochkovskiy A, Wang CY, Liao HYM. YOLOv4: Optimal speed and accuracy of object detection. *arXiv:2004.10934*, 2020.
- 12 Zhang HY, Cissé M, Dauphin YN, *et al.* Mixup: Beyond empirical risk minimization. *Proceedings of the 6th International Conference on Learning Representations*. Vancouver: OpenReview.net, 2018.
- 13 Wang CY, Liao HYM, Wu YH, *et al.* CSPNet: A new backbone that can enhance learning capability of CNN. *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. Seattle: IEEE, 2020. 390–391.
- 14 Lin TY, Dollár P, Girshick R, *et al.* Feature pyramid networks for object detection. *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu: IEEE, 2017. 2117–2125.
- 15 Liu S, Qi L, Qin HF, *et al.* Path aggregation network for instance segmentation. *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City: IEEE, 2018. 8759–8768.
- 16 Maji D, Nagori S, Mathew M, *et al.* YOLO-Pose: Enhancing YOLO for multi person pose estimation using object keypoint similarity loss. *Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. New Orleans: IEEE, 2022. 2637–2646.
- 17 Wang QL, Wu BG, Zhu PF, *et al.* ECA-Net: Efficient channel attention for deep convolutional neural networks. *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle: IEEE, 2020. 11534–11542.
- 18 Hu J, Shen L, Sun G. Squeeze-and-excitation networks. *Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City: IEEE, 2018. 7132–7141.
- 19 Li HL, Li J, Wei HB, *et al.* Slim-neck by GSConv: A lightweight-design for real-time detector architectures. *arXiv:2206.02424*, 2022.
- 20 Shi ZH, Feng YN, Zhao MH, *et al.* Normalised gamma transformation-based contrast-limited adaptive histogram equalisation with colour correction for sand-dust image enhancement. *IET Image Processing*, 2020, 14(4): 747–756. [doi: [10.1049/iet-ipr.2019.0992](https://doi.org/10.1049/iet-ipr.2019.0992)]
- 21 Sun K, Xiao B, Liu D, *et al.* Deep high-resolution representation learning for human pose estimation. *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach: IEEE, 2019. 5693–5703.
- 22 Wang CY, Bochkovskiy A, Liao HYM. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Vancouver: IEEE, 2023. 7464–7475.
- 23 Kirillov A, Mintun E, Ravi N, *et al.* Segment anything. *Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision*. Paris: IEEE, 2023. 4015–4026.

(校对责编: 张重毅)