doi:10.6041/j.issn.1000-1298.2025.04.037

# 基于 FCM－SimCC 的猪只面部关键点定位方法

张惠莉　王光远　员玉良　代晨龙　滕　飞　任景龙
（青岛农业大学机电工程学院，青岛 266109）

**摘要**：随着生猪养殖业向规模化和集约化转型，非侵入式个体识别技术对于追踪溯源、食品安全、疾病控制等方面至关重要，而猪只面部关键点定位技术是实现猪只非侵入式个体识别的前提。本研究基于 SimCC 关键点定位算法提出一种猪只面部关键点定位模型 FCM－SimCC，使用 FasterNet 代替原算法的 CSPDarkNet 作为特征提取网络；通过在 FasterNet 中嵌入 CA 注意力机制，提高模型对长距离特征的捕获能力；使用 MLT 自适应权重多任务损失函数联合 KL 散度损失函数与 Wing Loss 损失函数对模型进行监督。在包含多个猪只品种、多种面部姿态的 4 861 幅图像的数据集上进行实验，结果表明本研究模型平均精度均值、50% 平均精度、75% 平均精度分别为 76.12%、93.44%、83.25%，相比原模型分别提升 3.14、1.77、4.47 个百分点，浮点运算量为 $2.79\times10^9$，参数量为 $1.38\times10^7$，浮点运算量减少 38.68%，参数量减少 20.16%。并与 DarkPose、HRNet、YOLO X－Pose 等主流关键点定位方法进行对比，实验结果表明 FCM－SimCC 模型能够在较低的浮点运算量与较少模型参数量的基础上实现快速精准的猪只面部关键点定位，为猪只面部关键点定位及后续的猪只个体身份识别等提供技术支持。

**关键词**：猪脸关键点定位；SimCC；FasterNet；坐标注意力机制
**中图分类号**：TP391.41；S828　　**文献标识码**：A　　**文章编号**：1000-1298(2025)04-0397-11　　**OSID**：

# Pig Facial Landmark Detection Method Based on FCM－SimCC

ZHANG Huili　WANG Guangyuan　YUN Yuliang　DAI Chenlong　TENG Fei　REN Jinglong
(*College of Mechanical and Electrical Engineering*, *Qingdao Agricultural University*, *Qingdao* 266109, *China*)

**Abstract**: With the transformation of pig breeding industry to large-scale and intensive, non-intrusive individual identification technology is very important for traceback, food safety, disease control and scientific breeding. Pig facial landmark detection serves as a fundamental requirement for achieving non-invasive pig identification. A pig facial landmark detection model named FCM－SimCC was introduced, building upon the SimCC landmark detection algorithm. The model replaced CSPDarkNet with FasterNet for feature extraction and incorporated the CA attention mechanism within FasterNet to enhance the capture of long-distance features. Supervision of the model was achieved through the MLT adaptive weight multi-task loss function combined with KL divergence loss and Wing Loss. Test on a dataset of 4 861 images was done, representing a variety of pig breeds and facial poses, the FCM－SimCC model attained mean average precision, 50% average precision, and 75% average precision of 76.12%, 93.44%, and 83.25%, respectively. These results indicated improvements of 3.14, 1.77, and 4.47 percentage points over the original model, with a reduced computational demand to $2.79\times10^9$ and a parameter count of $1.38\times10^7$, marking a 38.68% decrease in floating-point operations and a 20.16% reduction in parameters. When compared with mainstream landmark detection methodologies such as DeepPose, HRNet, and YOLO X－Pose, the FCM－SimCC model showcased its ability to provide rapid and precise pig facial landmark detection with lower computational resources and fewer parameters, offering valuable insights for similar tasks in pig facial landmark detection and individual pig identification.

**Key words**: pig facial landmark detection; SimCC; FasterNet; coordinate attention

收稿日期：2024－02－21　修回日期：2024－03－14
**基金项目**：山东省自然科学基金项目(ZR2022MC152)和山东省重点研发计划项目(2018GGX104012)
**作者简介**：张惠莉(1973—)，女，副教授，博士，主要从事精准农业关键技术与应用研究，E-mail：zhhuili73@163.com
**通信作者**：员玉良(1981—)，男，副教授，博士，主要从事精准农业系统集成与智能仪器研究，E-mail：ylyun1981@163.com

## 0 引言

随着生猪养殖业向规模化和集约化转型,非侵入式的个体识别技术对于追踪溯源、食品安全、疾病控制以及科学选育等方面至关重要。基于视觉识别的身份识别、行为分析等技术,不仅可以有效降低成本,还能减少猪的应激反应,提升动物福利。但随着养殖规模的扩大和数据多样性的增加,传统的闭集识别方法面临挑战,鲁棒性更高的开集身份识别算法成为研究热点,其中猪只面部关键点的定位技术,作为猪只非侵入式身份识别任务中不可或缺的前处理步骤获得了更多关注。

目前,国内外基于深度学习的动物面部关键点定位算法已经取得了一定的进展。文献[1-13]论证了通过深度学习技术对猪只面部关键点定位的可行性,但大多都是基于 YOLO-Face 的 5 点关键点定位,关键点数量较少,仅能用于面部姿态估计;或者依赖仿射变换,直接将猪面部映射到人脸或其他动物面部,泛化性较差;或者使用单阶段定位方法直接对关键点进行回归,在单个猪只面部上进行定位时有良好表现,却无法处理多个猪脸目标中关键点的从属关系;或者在特征提取过程中,为提取更多细节信息一直维持高分辨率特征图,导致模型规模较大。

针对上述问题,本研究提出一种基于改进 SimCC 算法的猪只面部关键点定位模型 FCM-SimCC,通过引入 PConv 卷积代替原特征提取网络中的标准卷积生成 FasterBlock,进而构成 FasterNet 特征提取网络,提高模型的关键点定位速度;通过将 CA 注意力机制模块串联在 FasterBlock 与梯度截流模块之后构成 CA-FasterLayer,帮助模型捕获长距离坐标依赖,提高关键点定位精度;使用 MLT 自适应多任务损失函数联合 Wing Loss 损失与 KL 散度损失,监督模型学习过程,提高模型收敛速度及猪只面部关键点定位精度。

## 1 数据集构建

### 1.1 数据获取

本研究数据集为自建数据集与公开数据集共同构成。自建数据集部分,采集新希望六和农牧科技有限公司夏津县育肥舍养殖的 44 头成年长白猪视频,每段视频 4 min,采样频率 60 f/s。通过 Open CV 对视频进行分帧后,使用结构相似性指标 SSIM[14] 对猪只面部图像进行筛选,去除相似度较高的图像。对于给定的两幅面部图像 $x$、$y$,SSIM 计算公式为

$$\mathrm{SSIM}(x,y)=\frac{(2\mu_x\mu_y+c_1)(2\sigma_{xy}+c_2)}{(\mu_x^2+\mu_y^2+c_1)(\sigma_x^2+\sigma_y^2+c_2)} \tag{1}$$

式中 $\mu_x$、$\mu_y$ 分别为 $x$、$y$ 均值,$\sigma_x^2$、$\sigma_y^2$ 分别为 $x$、$y$ 方差;$c_1=(k_1L)^2$,$c_2=(k_2L)^2$ 为标量,用于维持图像对比度稳定;$L$ 为图像中像素的动态范围,当图像为 RGB 标准时,权重参数 $k_1=0.01$,$k_2=0.03$;$\sigma_{xy}$ 表示 $x$ 与 $y$ 之间的协方差。

取相似性阈值 SSIM 大于等于 0.7,得到筛选后的自建数据集共 1 397 幅。为保证数据集的丰富程度、提高模型泛化能力,还选取京东猪脸识别数据集及来自 Kaggle 等数据公开平台的不同品种、不同姿态、不同光照条件的猪只图像共 3 464 幅对数据集进行扩充,最终得到包含 4 861 幅图像的猪只面部关键点定位数据集。

### 1.2 数据标注

KHAN 等[7] 通过研究 22 400 幅动物面部图像与人脸的面部特征的相似性,在保持面部特征覆盖率前提下,提出了动物面部关键点的 9 点标注策略,并指出双侧耳尖、耳根部是包含较多哺乳动物面部特征的关键点。HANSEN 等[15] 在研究中指出,猪只耳尖较为活跃、动作幅度大,不建议选作关键点。在以上研究的基础上,本研究选取猪只面部的两侧瞳孔、两侧嘴角、鼻尖、两侧鼻孔、两侧耳基上下连接点及下唇缘中心点,共 12 个关键点进行标注。标注格式采用 COCO 格式[16],使用 Labelme 软件标注,面部标注框及关键点标注如图 1 所示。

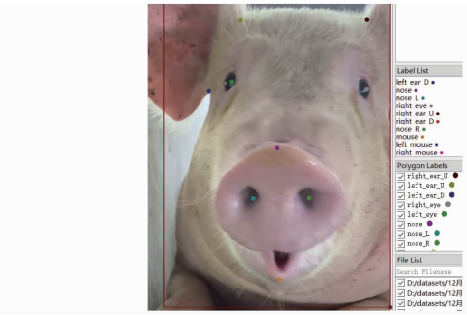



图 1 猪只面部关键点标注

Fig. 1 Pig facial landmarks labeling

### 1.3 数据预处理

实际养殖环境中猪只比较活跃,识别过程中面临污渍遮挡、光线变化、采集设备差异等多种因素的影响[8]。本研究在训练数据加载阶段使用 Albumentations[17] 工具进行数据增强,具体的增强方法包括几何变换、自适应直方图均衡、随机噪声、亮度调节、随机矩形遮盖等,提高模型对猪只面部关键点定位的泛化能力,为模拟自然环境条件,不对图像进行降噪处理。并使用 PDB(Pose based data balancing)增强策略[18] 优化图像增强过程,通过监

督模型训练过程中不同面部姿态样本的收敛程度，自适应选择增强比例，减少模型过拟合风险。

此外，本研究手工筛选 480 幅包含不同猪只品种、面部姿态、光照等条件的猪脸图像，作为预留测试集，用以评估模型对未知数据的泛化能力，保证测试集的丰富程度以模拟真实环境的复杂性。为保证测试有效性，测试集不出现在训练过程中。使用 5 折交叉验证分配训练集与验证集，确保模型训练和验证过程的严谨性和可靠性。

## 2　模型改进

### 2.1　SimCC 关键点检测算法

面部关键点检测算法主要分为直接坐标回归（Direct regression）算法[19]和热图编解码（Heatmap regression）算法[20-21]。前者直接预测图像中关键点的坐标，具有较高的训练和推理速度，但易受到噪声和异常值的影响；后者通过高斯编码器对每一个关键点匹配该关键点的空间维度特征，生成一幅二维高斯热图，通过热图来量化每个像素与所属关键点的置信度，实现关键点坐标的预测。在相同特征输入尺寸的前提下，Heatmap 算法比直接坐标回归算法有更高的定位精度。虽然 Heatmap 算法拥有检测精度的优势，但在特征提取阶段需要大量卷积和下采样操作，也带来了量化误差问题。SimCC 算法在 Heatmap 算法的基础上，将坐标分类的思想引入关键点定位算法，将水平和竖直坐标轴均匀划分为等宽的编码容器，并为编码容器赋予一个坐标标签，以建立编码容器与坐标的对应关系。模型通过学习关键点所属编码容器的类别，实现对坐标的分类，由解码器对命中的编码容器进行解码得到坐标值，实现关键点定位。通过调整编码容器宽度，将量化误差降低到亚像素级，以应对 Heatmap 算法因采样率、分辨率不足导致的量化误差的问题，如图 2 所示。

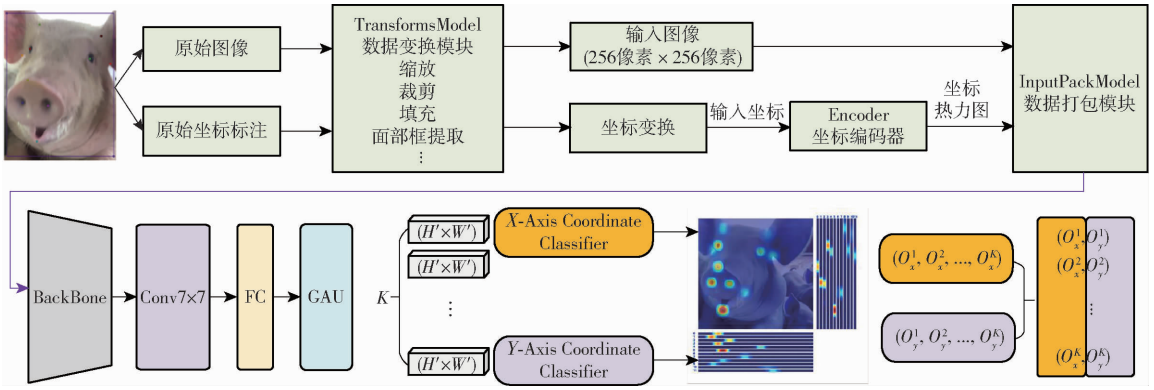



图 2　SimCC 结构图

Fig. 2　Structure of SimCC

对于共有 $k$ 个关键点的输入尺寸为 $(H, W, 3)$ 的图像，坐标编码器将第 $k$ 个关键点坐标 $(x_k, y_k)$ 经过编码后得到关于 $k$ 个关键点的特例图表征 $(H, W, K)$，编码过程公式为

$$Heatmap_{i,j,k} = \exp\left(-\frac{(i-x_k)^2+(j-y_k)^2}{2\sigma^2}\right) \quad (2)$$

式中 $1 \leqslant i \leqslant W$，$1 \leqslant j \leqslant H$，$1 \leqslant k \leqslant K$，$\sigma$ 为控制热力图分布范围的标准差，$\exp(\cdot)$ 为高斯函数。

将编码后的关键点热图二维表征输入特征提取网络，得到关于 $k$ 个关键点的尺寸为 $(K, H', W')$ 的关键点表征，传入一个由全 7×7 的标准卷积和连接层构成的线性映射模块实现特征图的展平操作，得到的对于 $k$ 个关键点的一维向量 $(K, H' \times W')$ 作为关键点的一维表征。SimCC 通过预测每一个一维表征所属的坐标标签实现关键点定位，后通过解码器解码为真实坐标。解码器模块在解码阶段负责完成坐标分类和坐标解码两个子任务，通过将两个全连接层分别作为水平和竖直坐标分类器构成检测头。具体公式为

$$\begin{cases} Class_x \in [1, N_x] & (N_x = W\lambda) \\ Class_y \in [1, N_y] & (N_y = H\lambda) \end{cases} \quad (3)$$

式中 $\lambda$ 为编码容器的缩放因子，用于控制量化误差。在预测过程中获得最大坐标分类概率所属的编码标签后，通过解码操作得到第 $p$ 个关键点真实坐标 $\hat{O}_p = (\hat{O}_x^{(p)}, \hat{O}_y^{(p)})$，有

$$\begin{cases} \hat{O}_y^{(p)} = \dfrac{\text{argmax}(\hat{O}_y(p))}{\lambda} \\ \hat{O}_x^{(p)} = \dfrac{\text{argmax}(\hat{O}_x(p))}{\lambda} \end{cases} \quad (4)$$

$\text{argmax}(\cdot)$ 为取最大值函数。

### 2.2　SimCC 关键点检测模型改进

为了进一步提高 SimCC 对猪只面部关键点的定位精度及速度，本研究在 SimCC 关键点定位算法的基础上，进行了 3 项改进，最终提出的 FCM－SimCC 模型结构如图 3 所示。

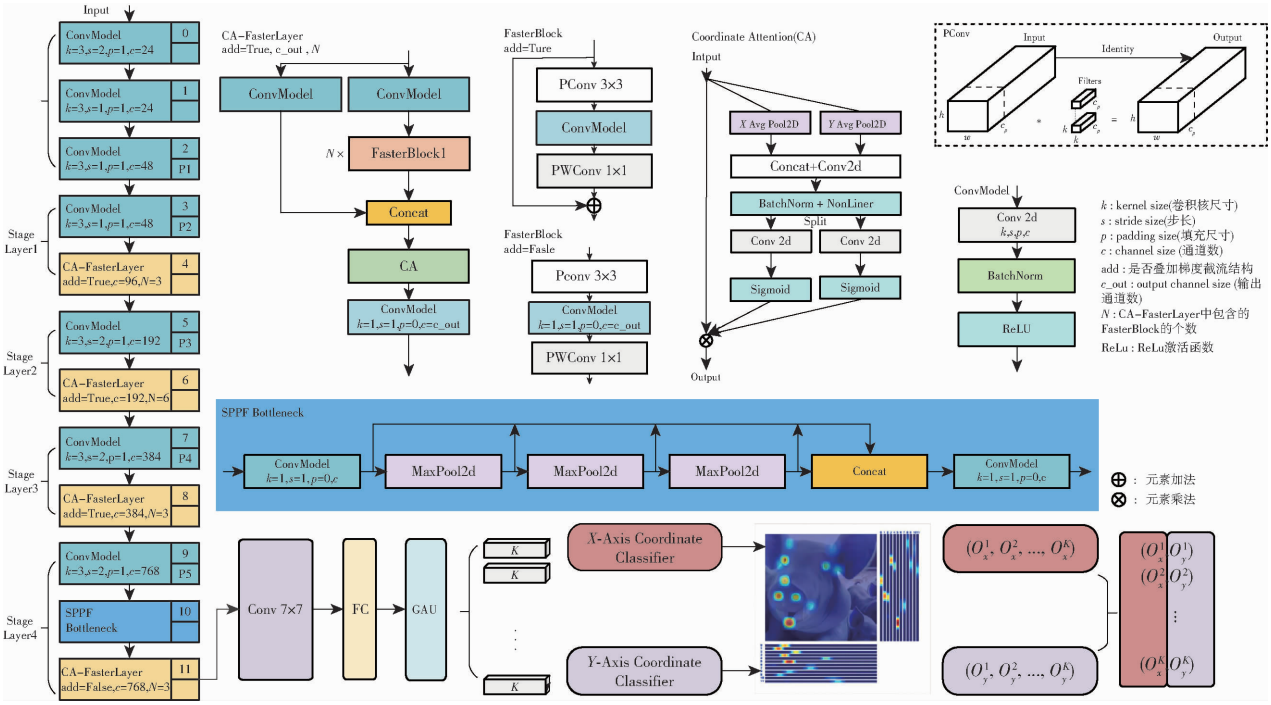



图 3 FCM－SimCC 模型整体结构图

Fig. 3 General structure of FCM－SimCC model

### 2.2.1 FasterNet 特征提取网络

为了在 SimCC 模型基础上进一步提升猪只面部关键点检测的推理速度，降低算力开销，本研究引入 FasterNet[22] 中的部分卷积（Perturbed convolution，PConv）算子结合逐点卷积（Point wise convolution，PWConv）算子构成 FasterBlock 模块，代替原模型特征提取网络 CSPDarkNet53[23] 的瓶颈层，形成 FasterNet 特征提取网络以减少算力开销。

FasterNet 借鉴 GhostNetV2[24] 的思想，使用 PConv 算子代替深度可分离卷积[25]（Depth wise separable convolution，DWConv），仅对部分输入通道应用标准卷积进行特征提取，剩余通道的特征执行恒等映射，在特征提取能力保持不变的基础上，减小内存访问频率和计算冗余，以提高训练和推理速度。对于输入尺寸为 $(h,w,c)$ 的 $k\times k$ 的标准卷积，标准卷积与 PConv 的浮点运算量为

$$FLOPs_{Conv}=h\times w\times k^2\times c^2 \tag{5}$$

$$FLOPs_{PConv}=h\times w\times k^2\times c_p^2 \tag{6}$$

式中 $c_p$ 为卷积参与率，表示部分卷积的卷积通道数，$h$、$w$ 分别表示输入特征的宽度、高度。当卷积参与率 $c_p$ 取经验数值为输入通道数 $c$ 的 1/4 时，PConv 的浮点运算量仅为标准卷积的 1/16。

输入 FasterBlock 的特征，通过 3×3 的 PConv 处理部分通道的特征，随后串联 1 个反向残差模块，该反向残差模块由 2 个 1×1 的 PWConv 构成，实现对所有通道中特征信息的有效利用。为避免因过度使用归一化层和激活层而导致特征多样性的减少，仅在第 2 个 PWConv 卷积层之后使用标准化层和激活层，通过反向残差连接缓解反向传播过程中的梯度消失问题，最终形成本研究的 FasterBlock，如图 4 所示。图中 Identity 函数表示恒等映射；Filters 为滤波器，由卷积核为 $k\times k$、通道数为 $c_p$ 的 Conv 2D 模块实现。

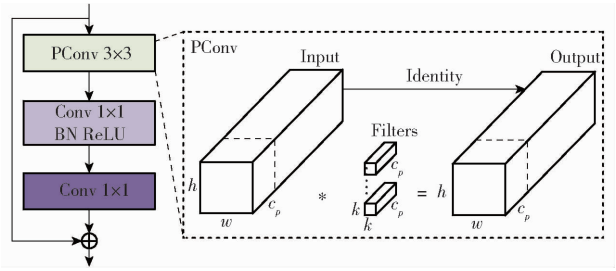



图 4 FasterBlock 模块结构图

Fig. 4 Structure of FasterBlock model

### 2.2.2 CA 坐标注意力机制

关键点定位是对空间信息高度敏感的任务[26]，在空间特征及细节纹理提取过程中发挥着至关重要的作用。为了提高特征提取网络对猪只面部关键点的特征提取能力，保留更多猪只面部关键点与局部空间细节，建立关键点与坐标细节间的长距离依赖。本研究通过引入 CA（Coordinate attentio）坐标注意力机制[27]，将 CA 注意力机制模块嵌入到 FasterBlock 的 CSP 截流网络后方，提高 FasterNet 网络对猪只面部关键点与纹理细节语义的特征表达能力，提高模型对猪只面部坐标信息的关注能力，减少

外部干扰，提高关键点定位的准确性。CA 注意力机制结构如图 5 所示。

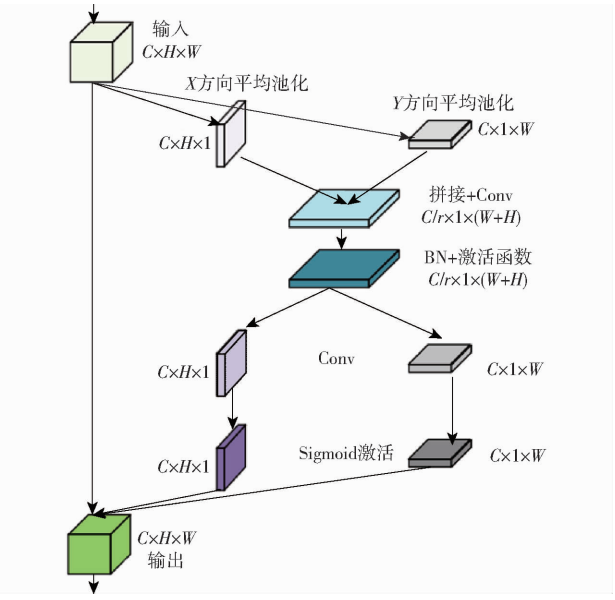



图 5　CA 注意力机制

Fig. 5　Structure of Coordinate attention

CA 注意力机制通过引入可学习参数，对竖直和水平坐标特征进行学习，实现对空间语义的提取。随后将 $z^{(h)}$、$z^{(w)}$ 按空间维度串联，由 1×1 逐点卷积进行跨通道全连接，经过归一化函数 BN 与非线性激活函数 $\delta(\cdot)$，得到关于竖直与水平维度特征图 $f^{(h)}$，$f^{(w)}$。随后在坐标注意力生成过程中，经过切片函数的分割后得到关于水平和竖直方向的特征张量，通过 $\sigma(\cdot)$ 激活函数对水平和竖直坐标分配注意力权重，注意力权重与各自通道数相同。

将输入特征与水平竖直注意力权重逐通道相乘，完成对原始特征的重校准，得到输出注意力特征，公式为

$$Y = Xs^{(h)}s^{(w)} \tag{7}$$

式中　$s^{(h)}$——竖直方向注意力权重

$s^{(w)}$——水平方向注意力权重

将 CA 注意力机制模块嵌入到 FasterNet 后的 CSP 截流模块构成 CA－FasterLayer。文献[28]指出将 SPPF 金字塔池化模块纳入网络中可以避免因网络尺寸调整导致的失真，SPPF 模块负责调整网络输出特征图形状以匹配检测头的输入，最终形成本研究的特征提取网络 CA－FasterNet。

CA－FasterLayer 特征提取网络可以利用坐标注意力机制的空间语义提取能力，准确地获得目标对象的位置，与 SimCC 算法坐标分类思想相搭配，以捕获长距离信息，能更好地在水平与竖直维度上提取关于猪只面部关键点位置的信息，以极小的计算代价和参数开销有效地增强模型的特征表达与空间泛化能力。

### 2.2.3　联合 Wing Loss 与 KL 散度损失的 MLT 自适应权重损失函数

SimCC 算法在对猪只面部关键点进行坐标编码过程中，会不可避免地导致空间信息的丢失。文献[29－30]指出关键点的概率分布空间并非服从单一分布，使用高斯分布损失和直接回归损失构成的联合损失函数可以更好地模拟现实检测场景的分布规律，保留更多的真实语义信息以实现对模型的约束。

为改进猪只面部关键点检测模型，本研究引入 MLT 多任务联合损失函数[31]对模型的关键点定位进行约束，通过 MLT 损失函数的自适应权重分配策略调节权重，联合 Wing Loss 损失函数[18]与 KL 散度损失函数[32]构成猪只面部关键点定位损失函数，鼓励模型学习关键点之间的空间关系，帮助模型更好地模拟关键点的真实分布，提高关键点之间的一致性。MLT 损失函数为

$$MLT = \sum_{i=1}^{N}\left(\frac{0.5}{p_i^2} + \lg(1+p_i^2)\right) \tag{8}$$

式中　$p_i$——权重参数

KL 散度损失通过比较模型学习的近似分布与数据真实分布之间的差异来惩罚模型对负样本的学习，有助于模型更好地学习猪只面部关键点的概率分布，在确保关键点的位置符合真实分布的前提下提高定位精度。计算式为

$$KL = \sum_{i=1}^{K} P(i)\lg\frac{P(i)}{Q(i)} \tag{9}$$

式中　$Q(i)$——第 $i$ 个关键点预测概率分布

$P(i)$——第 $i$ 个关键点真实概率分布

由编码器模块对关键点分布编码得到。对于第 $i$ 个关键点预测的概率分布为

$$P(i) = \frac{1}{2\pi\sigma_i^2}\exp\left(-\frac{d(O_i,\hat{O}_i)}{2\sigma_i^2}\right) \tag{10}$$

式中　$\sigma_i$——第 $i$ 个关键点真实分布的标准差

$d$——欧氏距离

Wing Loss 通过平滑欧氏距离度量预测值与真实值之间的差距，构成对数损失来度量模型预测误差，这种设计方式有助于在离群点上提供更小的梯度和更大的步长，帮助模型快速摆脱离群点；在较小的预测偏差时分配更小的步长以放大损失，帮助模型取得局部或整体最优，第 $i$ 个关键点上的 Wing Loss 计算公式为

$$L_W(O_i,\hat{O}_i) = \begin{cases}\omega\left(x - \ln\left(1+\dfrac{d(O_i,\hat{O}_i)^{\varepsilon}}{\omega}\right)\right) & (v_i>0)\\ 0 & (v_i=0)\end{cases} \tag{11}$$

其中 $$c = \omega - \omega\ln(1 + \omega/\varepsilon) \tag{12}$$

式中 $\omega$——非线性限制参数

$\varepsilon$——曲率参数

$K$ 个关键点的总损失为每一个关键点的 Wing Loss 之和，即

$$WING = \sum_{i=1}^{K} L_w(O_i, \hat{O}_i) \tag{13}$$

$O(x_i, y_i, v_i)$ 为真实关键点坐标，$\hat{O}(\hat{x}_i, \hat{y}_i, \hat{v}_i)$ 为预测关键点坐标，$\varepsilon = 2$ 为曲率参数，用于约束非线性区域的曲率，$\omega = 10$ 为非线性限制参数，将非线性部分限制在$[-\omega, \omega]$范围内，$c$ 用于将损失函数的线性部分与非线性部分连接成连续函数。

MLT 联合损失函数通过联合 KL 散度损失函数与 Wing Loss 损失函数，自适应地为每一个损失分配权重，综合考虑猪只面部关键点的分布概率空间模型与真实坐标回归模型对模型进行监督，以提高精确性和一致性，有助于改进猪脸面部关键点检测模型的性能，使其更适应真实情况下的猪只面部关键点检测任务。

### 2.3 评价指标

本研究使用关键点相似度 OKS 指标评价猪只面部关键点检测模型的检测结果。OKS 通过计算猪只面部关键点的预测值与其真实值之间的欧氏距离来衡量两者的相似度，编号为 $p$ 的猪只面部的 OKS 分值 $OKS_p$ 计算公式为

$$OKS_p = \frac{\sum_{i=1}^{K} \exp\left(\frac{-d_{pi}^2}{2S_p^2\sigma_i^2}\right)\delta(v_{pi} \geqslant 1)}{\sum_{i=1}^{K}\delta(v_{pi} \geqslant 1)} \tag{14}$$

式中：$d_{pi}$ 为第 $i$ 个关键点真实值与预测值间的欧氏距离；$v_{pi}$ 为关键点可见性指标，$v_{pi} = 0$ 表示该点不存在，$v_{pi} = 1$ 表示该点存在但不可见，$v_{pi} = 2$ 表示该点存在且可见；$\delta(v_{pi} \geqslant 1)$ 为选取可见点的计算函数，当可见性指标 $v_{pi} \geqslant 1$ 时 $\delta(v_{pi} \geqslant 1) = 1$，否则不计算 OKS；$S_p$ 为第 $p$ 个猪脸检测框面积的平方根；$\sigma_i$ 为关键点归一化因子，表示第 $p$ 号猪面部的第 $i$ 个关键点标注值与真实值分布间的标准差。

评价指标包括平均精度均值（Mean average precision，mAP）、50% 平均精度（50% Average precision，$AP_{0.5}$）、75% 平均精度（75% Average precision，$AP_{0.75}$）、平均召回率（Average recall，AR）。选择浮点运算量（FLOPs）作为检测速度的评价指标，模型参数量（Parameters）作为模型容量和复杂度指标以评价模型轻量化程度。

## 3 结果与分析

### 3.1 实验环境配置与训练参数设置

实验操作平台为 Intel（R）Xeon（R）Platinum 8358P CPU @ 2.60 GHz；NVIDIA GeForce RTX 3090（24 GB）；pyTorch 2.0.1@ CUDA 118，初始学习率 0.01，目标学习率 0.001，使用 Adam 优化器，权重衰减 0.005，余弦退火算法周期为 10，配置 5 折交叉验证，训练轮次为 200，图像分辨率为 256 像素 × 256 像素，批次大小为 64，取平均结果形成猪只面部关键点检测模型。

### 3.2 关键点定位实验

在实际环境中，猪只品种分布多样、面部姿态多变，且检测效果易受面部污渍、遮挡、光照变化等外部条件的影响。为验证模型在不同条件下的关键点定位能力，本研究在测试集中选取不同品种、不同面部姿态、面部污渍较多、面部被遮挡等情况的猪脸图像，以模拟真实环境，并通过可视化热力图直观表现模型在复杂情况下的关键点定位效果与泛化能力，检测结果如图 6 所示。从定位结果来看，FCM – SimCC 模型在各种条件下均表现出较好的识别性能。

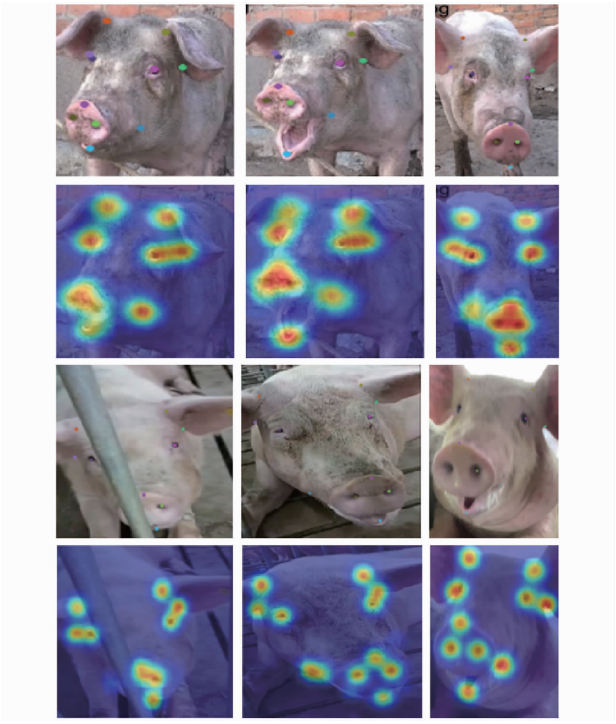


图 6 FCM – SimCC 模型检测结果

Fig. 6 Recognition effect of FCM – SimCC

交叉验证广泛用于有限样本下泛化误差估计[33]，本研究使用 5 折交叉验证，以提高数据样本的方差稳定性，提升模型泛化能力。并且绘制精度曲线，统计 5 折实验曲线，计算精度标准差，考察模型在

整个数据集上的一致性和泛化能力。由图 7 可知，随着训练过程的进行，标准差呈现稳定缩小趋势，表明模型在不同的数据分割上具有较高的一致性，模型的学习能力对数据变化不敏感，反映出良好的稳定性和泛化能力。图中第 1 ~5 折为交叉验证中第 1 ~5 折平均精度曲线，mAP 为 5 折交叉验证平均精度均值，作为模型的平均精度。

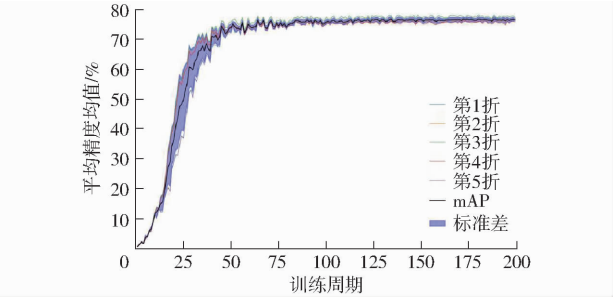



图 7　平均精度均值曲线

Fig. 7　Mean average precision curves

### 3.3　不同特征提取网络对比

为验证引入 CA－FasterNet 作为特征提取网络的有效性，分别选取 VipNAS[34]、ResNet18[35]、MobilenetV2[36]、ShuffleNetV2[37]，以及 HRNet[38] 算法的特征提取网络 HRNet－W32，对比本研究的 CA－FasterNet 作为特征提取网络时的性能。在保持 SimCC 检测头及 MLT 损失函数不变情况下训练 200 轮，在测试集上进行测试，结果如表 1 所示。

改进后 CA－FasterNet 的 mAP 与 $AP_{0.75}$ 最优，较 VipNAS、ResNet18、MobilenetV2、ShuffleNetV2 及 HRNet－W32 分别高出 6.55、3.93、14.07、13.48、0.79 个百分点与 6.02、3.40、13.46、13.59、1.46 个百分点。HRNet－W32 使用多分辨率并行子网堆叠形成整体网络，在特征提取过程中维持较高的特征分辨率，可以帮助网络学习更多的语义信息，即使在未嵌入任何注意力机制的情况下也拥有最高的

**表 1　不同特征提取网络对比实验结果**

**Tab. 1　Comparative experimental results of different feature extraction networks**

| 特征提取网络 | mAP/% | $AP_{0.5}$/% | $AP_{0.75}$/% | AR/% | 浮点运算量 | 参数量 |
|---|---|---|---|---|---|---|
| VipNAS | 69.57 | 88.34 | 77.23 | 75.57 | $8.00\times10^{8}$ | $8.65\times10^{6}$ |
| ResNet18 | 72.19 | 89.72 | 79.85 | 78.12 | $5.49\times10^{9}$ | $2.57\times10^{7}$ |
| MobilenetV2 | 62.05 | 85.54 | 69.79 | 67.81 | $3.10\times10^{8}$ | $2.23\times10^{6}$ |
| ShuffleNetV2 | 62.64 | 86.20 | 69.66 | 68.79 | $1.35\times10^{9}$ | $6.94\times10^{6}$ |
| HRNet－W32 | 75.33 | 93.51 | 81.79 | 80.11 | $1.61\times10^{10}$ | $6.36\times10^{7}$ |
| CA－FasterNet | 76.12 | 93.44 | 83.25 | 79.73 | $2.79\times10^{9}$ | $1.38\times10^{7}$ |

$AP_{0.5}$ 和 AR，$AP_{0.75}$ 精度与改进后 FasterNet 网络相差无几，但 HRNet 需要一直维持高分辨率特征图，导致模型规模过大，浮点运算量较高。MobilenetV2 作为轻量化特征提取网络，模型参数量仅 $2.23\times10^{6}$，浮点运算量仅为 $3.1\times10^{8}$，但由于在网络中频繁使用降采样操作，对特征图分辨率压缩严重，导致语义特征丢失，精度下降明显。综上，综合考虑检测精度与轻量化程度，改进后的 CA－FasterNet 网络在关键点定位性能、模型规模、推理性能等方面更加适合猪只面部关键点定位任务的要求。

### 3.4　消融实验

本研究设计了消融实验，以验证 FasterNet 特征提取网络、CA 注意力模块以及 MLT 损失函数改进对猪只面部关键点定位模型的有效性。由表 2 可知，由 SimmCC 与 CSPDarkNet53 构成的基线模型在测试集上 $AP_{0.5}$ 为 91.67%，但是 $AP_{0.75}$ 不足 80%，表明基线模型可有效地学习到关键点位置的语义信息，却无法提取更深层的语义特征，难以获得较好的定位精度。当特征提取网络更换为 FasterNet 时，得益于 PConv 算子的卷积策略改进，在计算过程减小了内存访问频率，模型的参数量和浮点运算量均大幅下降。并且模型相比于基线模型出现了精度下降现象，分析原因可知，FasterNet 中的恒等映射分支是由 PConv 后串联的 2 个 1×1 PWConv 卷积模块构成，在保证不缩减通道数的前提下模拟了逆残差结构，虽然有效地抑制了冗余特征对通道的占用，但也导致了部分语义信息丢失。在添加 CA 注意力机制模块后，CA 注意力机制模块与 CSP 结构的梯度节流分支共同作用，通过引入少量参数的代价捕获更多空间位置信息，进而补足因引入 PConv 算子构成恒等映射而丢失的语义。在引入 MLT 损失函数后，在基线模型及 CA－FasterNet 上的 $AP_{0.75}$ 分别提升 3.41、5.26 个百分点，表示 MLT 损失函数有助于建立关键点之间的位置联系，帮助模拟猪只面部关键点的真实分布，达到对预测定位的精度提升。在基线模型上同时引入 CA 注意力机制与 MLT 损失函数时，模型的 mAP、$AP_{0.5}$、$AP_{0.75}$ 分别提升 2.89、1.50、4.04 个百分点，表示模型对关键点的位置信息给予了更多的关注，MLT 损失对猪只面部关键点的定位误差产生了正面约束效果，尤其是 $AP_{0.75}$ 提升较大，表明 MLT 损失可监督模型向全局最优或局部最优收敛。同时引入 3 项改进后，FCM－SimCC

表 2 消融实验结果

Tab. 2 Ablation test results

| FasterNet | CA | MLT Loss | mAP/% | $AP_{0.5}$/% | $AP_{0.75}$/% | AR/% | 浮点运算量 | 参数量 |
|---|---|---|---|---|---|---|---|---|
| × | × | × | 72.98 | 91.67 | 78.78 | 75.04 | $4.55\times10^9$ | $1.73\times10^7$ |
| √ | × | × | 71.43 | 90.41 | 76.70 | 74.41 | $2.12\times10^9$ | $1.24\times10^7$ |
| × | √ | × | 73.33 | 91.66 | 79.68 | 75.62 | $4.95\times10^9$ | $1.84\times10^7$ |
| × | × | √ | 75.32 | 92.20 | 82.18 | 78.43 | $4.55\times10^9$ | $1.80\times10^7$ |
| √ | √ | × | 73.92 | 92.23 | 81.33 | 76.36 | $2.79\times10^9$ | $1.38\times10^7$ |
| √ | × | √ | 75.90 | 92.68 | 81.96 | 78.32 | $2.12\times10^9$ | $1.24\times10^7$ |
| × | √ | √ | 75.87 | 93.17 | 82.82 | 78.94 | $4.85\times10^9$ | $1.84\times10^7$ |
| √ | √ | √ | 76.12 | 93.44 | 83.25 | 79.73 | $2.79\times10^9$ | $1.38\times10^7$ |

注：×表示没有加入该模块；√表示加入该模块。

模型的 mAP、$AP_{0.5}$、$AP_{0.75}$、AR 分别相比基线模型增加 3.14、1.77、4.47、4.16 个百分点，参数量缩减 20.16%，浮点运算量减少 38.68%，在定位精度的提升与模型轻量化之间取得了较好平衡，满足猪只面部关键点定位算法的精度及轻量化优化目标。

### 3.5 不同关键点定位算法对比

为验证本研究提出的 FCM－SimCC 模型改进的效果，设计了不同关键点定位算法对比实验，选取了具有代表性的几个关键点检测模型，其中 YOLO v8－Pose－tiny[39]、YOLO v8n－Pose[40]为一阶段直接坐标回归算法；Integral Pose Regression（IPR）[41]为逻辑坐标回归（Logic regression）算法；HRNet、DarkPose[42]为基于热力图编解码回归算法，HRNet 使用带多尺度特征融合的 Heatmap 检测头与 HRNet－W32 特征提取网络；DeepPose 为二阶段直接坐标回归算法。保持相同条件，与本研究 FCM－SimCC 模型进行对比实验。

由表 3 可知，HRNet 算法为基于热力图编解码的定位算法，HRNet－W32 特征提取网络得益于其维持高分辨率特征的优势以及 Heatmap 算法提取的多种不同尺度的高分辨率二维高斯热图，拥有最高的平均精度和平均召回率，同样也导致其为了保证特征提取网络与检测头的特征尺寸匹配需要进行频繁采样运算，浮点运算量和参数量过于庞大，并不满足真实环境中实时监测的部署要求。YOLO X－Pose－tiny 作为轻量化一阶段网络，模型参数量最小，但 mAP 仅为 52.66%，满足实际需求。YOLO v8n－Pose 相比于 YOLOX 系列算法拥有更大的网络规模，并引入 OKS 损失函数监督模型，通过对关键点位置进行直接坐标回归实现关键点定位，在模型规模少量增加的前提下，相较于 YOLO X－Pose－tiny 算法，mAP 提升 6.81 个百分点。IPR 算法在维持较为均衡的召回率和检测精度的情况下，浮点运算量最大，分析原因可知，由于其引入的 SoftArgmax 函数计算较为复杂，计算过程中需要频繁的积分操作，导致了昂贵的算力开销。DarkPose 算法通过独立地对每个关键点进行特征提取和回归获得了较好的表现，在未进行任何优化改进的情况下以较低的浮点运算量，取得的 mAP 为 71.48%，但 DarkPose 为姿态关节关键点定算法，模型不能很好地适应动物面部同时具有刚体与非刚体属性的特点，在猪只面部关键点定位任务上并未表现出其原有的优势。综合以上分析，本研究提出的算法最能适应猪只面部关键点定位任务的需求，得益于改进 FCM－SimCC 算法对量化误差的控制、联合 MLT 损失监督模型学习过程与 CA－FasterNet 网络的优秀特征提取性能，在精度与轻量化两方面相比其他定位算法均取得较大优势。

表 3 不同关键点定位算法对比实验结果

Tab. 3 Comparison of test results with different landmarks detection methods

| 模型 | mAP/% | AR/% | 浮点运算量 | 参数量 |
|---|---|---|---|---|
| YOLO X－Pose－tiny | 52.66 | 57.19 | $3.77\times10^9$ | $6.07\times10^6$ |
| YOLO v8s－Pose | 59.47 | 59.47 | $3.30\times10^9$ | $9.20\times10^6$ |
| HRNet | 77.67 | 81.00 | $6.45\times10^{10}$ | $9.48\times10^7$ |
| DarkPose | 71.48 | 76.21 | $8.35\times10^9$ | $2.96\times10^7$ |
| DeepPose | 72.20 | 64.40 | $7.67\times10^9$ | $4.27\times10^7$ |
| IPR | 67.45 | 76.41 | $7.33\times10^{10}$ | $3.40\times10^7$ |
| FCM－SimCC | 76.12 | 79.73 | $2.79\times10^9$ | $1.38\times10^7$ |

为更加直观反映各模型在不同情况下的关键点定位效果，分别选择基于 DarkPose 检测算法、HRNet 检测算法与 FCM－SimCC 算法，选择多个猪只面部、趴卧姿态、光照不足等困难样本进行对比实验，并对关键点定位结果进行热力图可视化，结果如图 8～10 所示。

如图 8 所示，DarkPose 对猪只眼周轮廓的特征提取能力较好，能提取到关于猪只面部关键点的空间语义信息。但在侧脸的检测情况下，对猪只鼻尖、下唇缘的关注程度不足，对两侧鼻孔的定位产生了

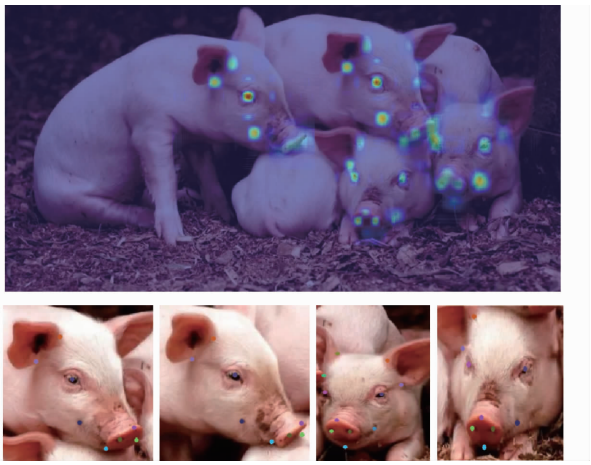


图 8　DarkPose 检测结果及可视化热力图

Fig. 8　Recognition effect of DarkPose



图 9　HRNet 检测结果及可视化热力图

Fig. 9　Recognition effect of HRNet

小范围的漂移现象。由可视化热力图可见，在面部较多污渍的侧脸情况下，DarkPose 对鼻尖及下唇缘部分特征提取能力不足。如图 11a 所示，在侧卧等较大面部姿态变化的情况下，对鼻尖位置的关注程度有限，关键点定位产生了漂移。

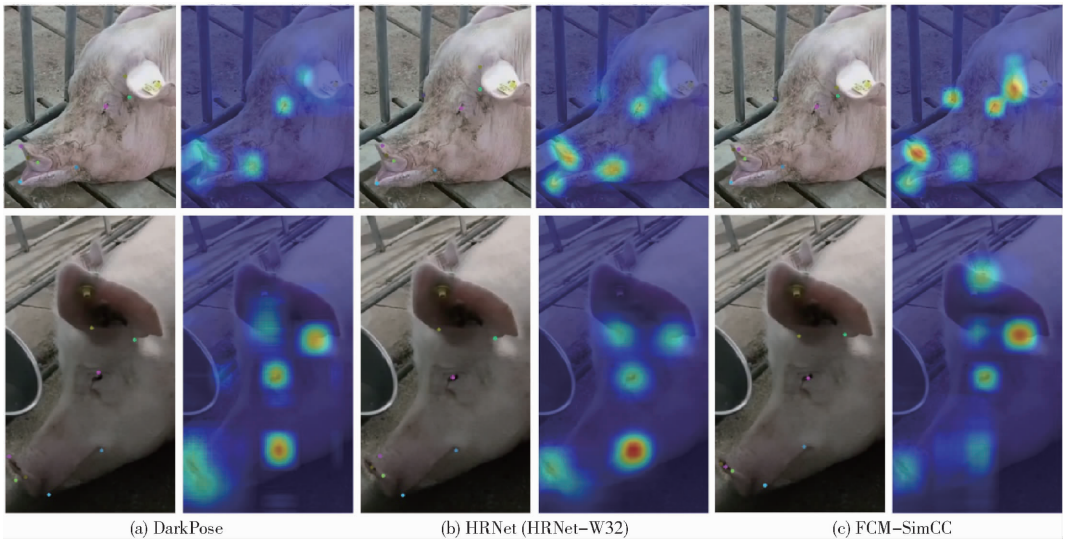


图 10　FCM－SimCC 检测结果及可视化热力图

Fig. 10　Recognition effect of FCM－SimCC

在 HRNet 算法中，得益于其维持全局高分辨率特征图的能力和最大的网络规模，HRNet 对拥有明显几何特征的面部区域提取了较多信息，在边缘轮廓及纹理交叉变化等位置的关注明显，能很好地提取嘴角、耳基点等有明显纹理交叉的部位。同时如图 9 所示，HRNet 对面部几何特征提供了过多的关注，对猪只前肘与躯干连接部分、下颌线位置、鼻梁褶皱位置产生了关注，导致关键点之间相互孤立，并未学习到关键点之间的位置关系，容易被面部褶皱和污渍影响。图 11b 也证明了 HRNet 对污渍较多的耳基点附近特征关注度过高，且对下颌线位置提供了不应该存在的关注，在模型已对嘴角部分提供了充足关注的前提下，对其他关键点的关注程度依旧不足。

如图 10 所示，相比于其他算法，本研究的 FCM－SimCC 模型提取了更多的语义，在不同面部姿态情况下的关注程度较为均匀。由热力图可见，在鼻尖到双侧嘴角范围内，模型关注区域沿关键点关联方



(a) DarkPose　(b) HRNet (HRNet-W32)　(c) FCM-SimCC

图 11　面部污渍及光照不足情况下的测试结果

Fig. 11　Recognition effect and visualized heatmap of facial stains and insufficient lighting conditions

向延伸,成功建立了关键点之间的联系,模型对于定位难度较高的两侧嘴角位置预测更加准确。如图 11c 所示,在较多污渍且较难定位的困难样本中,FCM - SimCC 对双耳基点注意力充足,表明本研究的改进有效地提取了猪只面部关键点真实分布的信息。同时,FCM - SimCC 在图 10 上对右侧嘴角的定位产生了较小偏差,但依旧在嘴角范围之内,并没有同 HRNet 算法一样,出现将前肘弯曲位置误认为嘴角的情况。在正脸和侧脸姿态下,对猪脸关键点提供的较为平均的关注程度,表明 FCM - SimCC 模型更适应猪只面部关键点定位的精度要求,在各种检测环境下均有较好的表现,能适应多种复杂的关键点定位情况。

## 4 结束语

在 SimCC 算法基础上进行改进,提出了 FCM - SimCC 模型。首先,将 FasterNet 作为特征提取网络,通过部分卷积有效减少模型浮点运算量,提高模型训练和推理速度。其次,嵌入 CA 注意力模块,能有效提取空间内长距离特征依赖,通过搭配 FasterLayer 中的梯度截流网络,在仅引入少量可学习参数的前提下,构成 CA - FasterNet 特征提取网络,提升模型对空间语义信息的保留能力。最后,通过 MLT 多任务损失联合 Wing Loss 与 KL 散度损失函数对模型进行监督,通过 Wing Loss 惩罚离群点,借助 KL 散度损失监督关键点概率分布,帮助模型更好地学习猪只面部关键点真实分布规律,提高模型的定位精度。结果表明,FCM - SimCC 模型相比基线模型,mAP、$AP_{0.75}$ 分别提升 3.14、4.47 个百分点,参数量缩减 20.16%,浮点运算量减少 38.68%。满足定位精度及定位速度的优化目标。可视化热力图能直观地展示模型的特征提取能力,表明本研究的 FCM - SimCC 在面部污渍、姿态变化、环境光变化等困难检测条件下,对比目前主流关键点定位方法,在定位精度和鲁棒性上具有一定优势,更能适应现实情况下猪只面部关键点定位的需求。

## 参 考 文 献


[1] 宋怀波,华志新,马宝玲,等. 基于 SimCC - ShuffleNetV2 的轻量化奶牛关键点检测方法[J]. 农业机械学报, 2023, 54(10): 275 - 281, 363.
SONG Huaibo, HUA Zhixin, MA Baoling, et al. Lightweight keypoint detection method of dairy cow based on SimCC - ShuffleNetV2[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(10): 275 - 281, 363. (in Chinese)

[2] 黄志杰,徐爱俊,周素茵,等. 融合重参数化和注意力机制的猪脸关键点检测方法[J]. 农业工程学报, 2023, 39(12): 141 - 149.
HUANG Zhijie, XU Aijun, ZHOU Suyin, et al. Key point detection method for pig face fusing reparameterization and attention mechanisms[J]. Transactions of the CSAE, 2023, 39(12): 141 - 149. (in Chinese)

[3] 李向宇,李慧盈. 基于卷积神经网络的猪脸特征点检测方法[J]. 吉林大学学报(理学版), 2022, 60(3): 609 - 616.
LI Xiangyu, LI Huiying. Feature point detection method of pig face based on convolutional neural network[J]. Journal of Jilin University(Science Edition), 2022, 60(3): 609 - 616. (in Chinese)

[4] 李玉龙. 基于深度学习的动物面部及关键点检测算法设计与实现[D]. 哈尔滨:哈尔滨理工大学, 2021.
LI Yulong. Design and implementation of animal face and key point detection algorithm based on deep learning[D]. Harbin: Harbin University of Science and Technology, 2021. (in Chinese)

[5] 杨家林. 基于注意力机制的轻量级羊脸识别方法研究与实现[D]. 杨凌:西北农林科技大学, 2022.
YANG Jialin. Research and implementation of lightweight sheep face recognition method based on attention mechanism[D]. Yangling: Northwest A&F University, 2022. (in Chinese)

[6] MARTVEL G, FARHAT N, SHIMSHONI I, et al. Catflw: cat facial landmarks in the wild dataset[J]. arXiv Preprint, arXiv: 2305.04232, 2023.

[7] KHAN M, MCDONAGH J, KHAN S, et al. Animalweb: a large-scale hierarchical dataset of annotated animal faces[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020.

[8] 谢秋菊,吴梦茹,包军,等. 融合注意力机制的个体猪脸识别[J]. 农业工程学报, 2022, 38(7): 180 - 188.
XIE Qiuju, WU Mengru, BAO Jun, et al. Individual pig face recognition combined with attention mechanism[J]. Transactions of the CSAE, 2022, 38(7): 180 - 188. (in Chinese)

[9] 张建龙,周康,庄晏榕,等. 育肥猪生长过程中脸部变化对识别模型准确率的影响[J]. 中国农业大学学报, 2021, 26(6): 180 - 186.
ZHANG Jianlong, ZHOU Kang, ZHUANG Yanrong, et al. Effect of facial changes on the accuracy of the recognition model during the growth of finishing pigs[J]. Journal of China Agricultural University, 2021, 26(6): 180 - 186. (in Chinese)

[10] WANG R, GAO R H, LI Q F, et al. Pig face recognition based on metric learning by combining a residual network and attention mechanism[J]. Agriculture, 2023, 13(1): 144.

[11] WANG Z, ZHOU S, YIN P, et al. Ganpose: pose estimation of grouped pigs using a generative adversarial network[J]. Computers and Electronics in Agriculture, 2023, 212: 108119.

[12] YANG H, ZHANG R, ROBINSON P. Human and sheep facial landmarks localisation by triplet interpolated features[C]// 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), 2016: 1 - 8.

[13] WANG X P, WANG W, LIU J S, et al. HRST: an improved HRNet for detecting joint points of pigs[J]. Sensors, 2022,

22(19)：7215.
[14] WANG Z, BOVIK A, SHEIKH H R, et al. Image quality assessment：from error visibility to structural similarity[J]. IEEE Trans Image Process, 2004, 13(4):600－612.
[15] HANSEN M, SMITH M L, SMITH L N, et al. Towards on-farm pig face recognition using convolutional neural networks[J]. Computers in Industry, 2018, 98：145－152.
[16] SILVIA R, ALEXANDER S, CHRISTOPHER I H. Coco (creating common object in context) dataset for chemistry apparatus [J]. Procedia Computer Science,2020, 171：2445－2452.
[17] BUSLAEV A, IGLOVIKOV V I, PARINOV A, et al. Albumentations：fast and flexible image augmentations[J]. Information (Switzerland), 2020, 11(2)：125.
[18] FENG Z H, KITTLER J, AWAIS M, et al. Wing loss for robust facial landmark localisation with convolutional neural networks [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018：2235－2245.
[19] TOSHEV A, SZEGEDY C. Deeppose：human pose estimation via deep neural networks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014：1653－1660.
[20] BULAT A, TZIMIROPOULOS G. Human pose estimation via convolutional part heatmap regression[J]. arXiv Preprint, arXiv:1609.01743, 2016.
[21] QU H, XU L, CAI Y, et al. Heatmap distribution matching for human pose estimation[J]. arXiv Preprint, arXiv:2210.00740, 2022.
[22] CHEN J, KAO S H, HE H, et al. Run, don't walk：chasing higher FLOPS for faster neural networks[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023：12021－12031.
[23] REDMON J, FARHADI A. Yolov3：an incremental improvement[J]. arXiv Preprint, arXiv:1804.02767, 2018.
[24] TANG Y, HAN K, GUO J, et al. GhostNetV2：enhance cheap operation with long-range attention[J]. arXiv Preprint, arXiv:2211.12905, 2022.
[25] CHOLLET F. Xception：deep learning with depthwise separable convolutions[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017：1251－1258.
[26] 张飞宇,王美丽,王正超. 引入 Transformer 和尺度融合的动物骨骼关键点检测模型构建[J]. 农业工程学报, 2021, 37(23)：179－185.
ZHANG Feiyu, WANG Meili, WANG Zhengchao. Construction of the animal skeletons keypoint detection model based on Transformer and scale fusion[J]. Transactions of the CSAE, 2021, 37(23)：179－185. (in Chinese)
[27] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021：13713－13722.
[28] REN S, HE K, GIRSHICK R, et al. Faster R－CNN：towards real-time object detection with region proposal networks[J]. arXiv Preprint, arXiv：1506.01497, 2015.
[29] CARREIRA J, AGRAWAL P, FRAGKIADAKI K, et al. Human pose estimation with iterative error feedback[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015：4733－4742.
[30] XIAO S, XIAO B, SHUANG L, et al. Integral human pose regression[C]// Proceedings of the European Conference on Computer Vision (ECCV), 2017：529－545.
[31] KENDALL A, GAL Y, CIPOLLA R. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018：7482－7491.
[32] KULLBACK S, LEIBLER R A. On information and sufficiency[J]. The Annals of Mathematical Statistics, 1951, 22(1)：79－86.
[33] 杨柳,王钰. 泛化误差的各种交叉验证估计方法综述[J]. 计算机应用研究, 2015, 32(5)：1287－1290, 1297.
YANG Liu, WANG Yu. Survey for various cross-validation estimators of generalization error[J]. Application Research of Computers, 2015, 32(5)：1287－1290, 1297. (in Chinese)
[34] XU L, GUAN Y, JIN S, et al. Vipnas：efficient video pose estimation via neural architecture search[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021：16072－16081.
[35] TARG S, ALMEIDA D, LYMAN K, et al. Resnet in resnet：generalizing residual architectures[J]. arXiv Preprint, arXiv：1603.08029, 2016.
[36] SANDLER M, ANDREW H, ZHU M, et al. MobilenetV2：inverted residuals and linear bottlenecks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018：4510－4520.
[37] MA N, ZHANG X, ZHENG H T, et al. ShufflenetV2：practical guidelines for efficient CNN architecture design[C]// Proceedings of the European Conference on Computer Vision (ECCV), 2018：116－131.
[38] SUN K, XIAO B, LIU D, et al. Deep high-resolution representation learning for human pose estimation[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019：5693－5703.
[39] GE Z, LIU S, WANG F, et al. Yolox：exceeding Yolo series in 2021[J]. arXiv Preprint, arXiv：2107.08430, 2021.
[40] MAJI D, NAGORI S, MATHEW M, et al. Yolo-pose：enhancing Yolo for multi person pose estimation using object keypoint similarity loss[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022：2637－2646.
[41] SUN X, XIAO B, WEI F, et al. Integral human pose regression[C]// Proceedings of the European Conference on Computer Vision (ECCV), 2018：529－545.
[42] GENG Z, SUN K, XIAO B, et al. Bottom-up human pose estimation via disentangled keypoint regression[J]. arXiv Preprint, arXiv：2104.02300, 2021.