doi:10.6041/j.issn.1000-1298.2026.17.026

关注细粒度特征的复杂场景牛只头部位置轻量化检测模型 VSF－YOLO 研究

李永亭[1,2]　郭雨鑫[1]　齐咏生[1,2]　刘利强[1,3]　王朝霞[1,3]

（1.内蒙古工业大学电力学院，呼和浩特 010080；2.机电控制自治区高等学校重点实验室，呼和浩特 010080；
3.内蒙古自治区高等学校智慧能源技术与装备工程研究中心，呼和浩特 010080）

摘要：当前牲畜检测算法常因复杂场景的干扰而面临采集图像模糊、残缺遮挡等问题，导致检测精度降低。同时检测模型通常参数量大，难以部署至便携式设备。为此提出一种轻量化、应对复杂场景和聚焦面部细粒度特征的 VSF－YOLO 网络架构应用于牛只面部检测。算法首先提出一种 VegsNet 主干网络，将其作为 YOLO v10n 的主干网络以减少模型复杂度，降低运行参数；其次，为捕捉图像细粒度特征，提出一种空间感受野卷积高效地捕捉全图关键信息，并设计一种全文特征锚点注意力机制，利用全文锚点特征信息对细粒度特征进行准确抓取；最后，通过优化损失函数增强卷积对细粒度特征的提取，定位关键区域。为检测算法的有效性和实时性，在特定数据集上进行消融实验和与多种经典算法的对比实验，并部署至手机 Android 系统进行测试。结果表明，与传统 YOLO v10n 相比参数量下降 30%，mAP@0.5 提升 3.55 个百分点，得益于细粒度特征抓取，在相同模糊、遮挡复杂场景下牛脸的检测置信度平均提高 36%，且移植到手机应用程序后运行效率提升，相比原模型运行帧率平均提升 30%。

关键词：牛只面部检测；细粒度特征；轻量化；复杂场景；Android 部署；VSF－YOLO

中图分类号：TP391.4　　**文献标识码**：A　　**文章编号**：1000-1298(2026)17-0276-12　　**OSID**：

Research on Lightweight Detection Model VSF－YOLO for Cattle Head Position in Complex Scenes on Fine-grained Features

Li Yongting[1,2]　Guo Yuxin[1]　Qi Yongsheng[1,2]　Liu Liqiang[1,3]　Wang Zhaoxia[1,3]

(1. *School of Electric Power*, *Inner Mongolia University of Technology*, *Hohhot* 010080, *China*
2. *Autonomous Region Key Laboratory of Electromechanical Control in Higher Education Institutions*, *Hohhot* 010080, *China*
3. *Inner Mongolia Autonomous Region Higher Education Engineering Research Center for Smart Energy Technology and Equipment*, *Hohhot* 010080, *China*)

Abstract: In practical applications, current livestock detection algorithms often suffer from issues such as blurred, incomplete, or occluded captured images due to interference from complex scenes, leading to reduced detection accuracy. Additionally, detection models typically have a large number of parameters, making them difficult to deploy on portable devices. To address these challenges, a lightweight VSF－YOLO network architecture designed for cattle facial detection was proposed, which was capable of handling complex scenes and focusing on fine-grained facial features. The algorithm firstly introduced a VegsNet backbone network to replace the original YOLO v10n backbone, thereby reducing model complexity and operational parameters. Secondly, to capture fine-grained image features, a spatial receptive field convolution was proposed to efficiently extract key information from the entire image. Additionally, a full-feature anchor attention mechanism was designed to accurately capture fine-grained features by leveraging anchor point feature information across the entire image. Finally, the loss function was optimized to enhance the convolution's ability to extract fine-grained features and locate critical regions. To validate the algorithm's effectiveness and real-time performance, ablation studies and

收稿日期：2025－04－23　修回日期：2025－05－26

基金项目：国家自然科学基金项目(62363029)、内蒙古“英才兴蒙”工程团队项目(2025TEL07)、内蒙古教育厅创新团队项目(NMGIRT2508)和“政产学研推用银”创新联合体项目(2023RC－联合体－10)

作者简介：李永亭(1975—)，女，教授，主要从事新能源系统状态评估、故障监测与故障诊断研究，E-mail：lyt@imut.edu.cn

通信作者：齐咏生(1975—)，男，教授，博士，主要从事机器人协同控制技术研究，E-mail：qys@imut.edu.cn

comparative experiments with various classical algorithms were conducted on a specific dataset. The model was also deployed on an Android mobile system for testing. The results indicated that compared with the traditional YOLO v10n, the parameter count was decreased by 30%, while the average accuracy was increased by 3.55 percentage points. Benefiting from fine-grained feature extraction, the detection confidence for cattle faces was improved by an average of 36% under the same complex scenarios involving blurring and occlusion. Moreover, after being ported to mobile applications, the operational efficiency was increased, with the average frame rate rising by 30% compared with that of the original model.

Key words: cattle facial detection; fine-grained features; lightweight; complex scenarios; Android deploy; VSF－YOLO

0　引言

在牧业智能化发展中，奶牛作为牧场的核心生产力，其稳定生长直接关系到牧场的经济效益。准确地检测牛只头部不仅能够有效地获取牛只的身份特征信息便捷牧民管理，更能助力牧场实现经济效益的显著提升[1-2]。然而传统的人工观测方法往往费时费力，在采集过程中牛只运动的不稳定性和牲畜的不配合性会对采集造成较大影响。因此实时准确地检测牛只头部对身份识别至关重要。

为了突破这一难题，研究者开始尝试将深度学习技术应用于奶牛特征检测中。掩膜 R－CNN 算法开始应用在奶牛饲养领域[3]，卷积神经网络（CNN）等深度学习模型具有强大的特征学习和识别能力[4]，CNN 结合迁移学习的策略也逐渐步入大众视野[5]，多尺度融合算法也应用于牛只领域[6]，为提取图像中的有用信息提供新思路。

然而，这些方法仍然存在一些局限性，国内外研究者积极投身于奶牛面部检测技术的研发工作中[7-9]。Lodkaew 等[10]提出了 CowXNet 算法，该算法通过设计多个特征来检测奶牛状态。Siregar 等[11]提出了一种两阶段分割方法提升深度学习模型的准确性，适当比例和像素缩放可优化模型泛化能力。Dac 等[12]借助深度学习开发了一种非侵入式牲畜面部检测系统，同时 Hou 等[13]提出了一种轻量化的非接触式牛臀识别方法并部署在边缘计算设备中。Bergman 等[14]则将人脸检测系统和损失函数引入奶牛面部检测，实现了 20 ms 内处理 10 头牛只面部的视频帧，定位牛只面部。

目前，图像检测领域的研究中，YOLO 系列因其高集成、高速、高准确性等特点[15-16]受到研究者的一致认同。Guo 等[17]使用 YOLO v3 进行体温等参数的自动检测识别；而 Wang 等[18]则对 YOLO v5l 模型进行改进，加入了 Atrous 空间金字塔和注意力机制，该方法的平均准确率（mAP）达到 94.3%。Kawagoe 等[19]根据奶牛特征使用 YOLO 检测器从喂食过程中捕获的视频图像中提取奶牛头部区域位置，通过大量实验验证，该方法的准确率高达 90% 以上；Qin 等[20]则将 YOLO v5 轻量化部署到 Android 端程序中，展现出出色的实用性和推广价值。同时 Luan 等[21]对 YOLO v8 算法进行轻量化和加权特征融合探测头的优化；同时 Mu 等[22]提出了轻量级多尺度特征融合检测头模块，充分利用深度卷积的优势，实现轻量级网络架构，有效减少冗余信息。

目前已有的算法在检测牛只头部位置方面展现出了初步成效，实际应用中仍然面临着多重挑战。实际中算法运行参数量大、使用成本高、检测效率低；而且牛只运动容易造成图像残缺和模糊，导致遮挡环境下采集效率低[23]；同时牛只生活环境干扰因素较多，牛场中一般有较多干扰因素，牛只面部的污垢杂毛等都会遮盖有用信息，在牛只面部检测时无法有效感知和捕捉细粒度特征[24]，大量的冗余特征会干扰算法的检测。

为解决上述问题，本文设计一种专注于面部细粒度特征，能够实时检测牛只头部位置的 VSF－YOLO 算法。为了便于算法嵌入于手机端 APP 前端应用程序中进行采集，通过 VegsNet 主干网络将算法轻量化，减少运算量；使用空间感受野卷积 SRFConv 代替传统卷积模块，专注于图像细粒度特征，通过自适应权重提高对复杂场景的鲁棒性；通过优化损失函数，帮助 SRF 卷积定位图中关键区域；同时结合一种全文锚点注意力机制 FFAA，在强调细节特征同时关注上下文特征之间的关系，通过感知锚点关系，强化重要细粒度特征的同时弱化环境特征，提高模型的表达能力。

1　数据采集与预处理

1.1　数据采集

于内蒙古自治区呼和浩特市欧太牧场采集了 1 000 头牛只右脸、中脸和左脸视频（分辨率为 720 像素 ×1 280 像素），并从中逐帧选取各 15 幅不同图

像,共得 15 000 幅牛只头部图像。通过水平翻转、±15°旋转、中心裁剪等数据增强方法,将数据集扩展为包含 30 000 幅图像的 3 角度牛脸图像集。

1.2 基于 Auto－MSRCR 的数据增强

数据集处理中,通过 Auto－MSRCR 算法对捕获的图像进行增强,本文拟解决图像残缺、模糊以及面部遮挡等复杂场景干扰下的图像检测问题,要求数据集亮度达到基本标准,效果如图 1 所示。



(a) 原始图像 (b) 数据增强的图像

图 1 基于 Auto－MSRCR 算法的增强图像

Fig. 1 Enhanced image based on Auto－MSRCR algorithm

采用 Auto－MSRCR 算法实现了图像的分层强化,提高了对牛只头部的精确检测。增强后的数据按照比例 8∶1∶1构建训练集、验证集和测试集,对改进后的模型进行训练。

2 基于 YOLO v10 细粒度牛只头部检测模型

YOLO v10 算法,作为深度学习领域目标检测技术的重要里程碑,其展现出的核心特性和创新改动具有显著进步,在目标检测与实例分割任务上均展现出了卓越性能[25]。随着 YOLO 系列的发展,在检测网络中引入更深层次的网络结构和更有效的特征融合机制,显著增强了细粒度特征提取能力,在复杂场景中的目标区分度和检测精度均得到提升[26]。尽管如此,在实际图像采集过程中,由于图像残缺、运动模糊,以及牲畜面部的脏污、长毛等干扰信息,仍可能影响检测效率。基于上述问题,根据牛只头部不同位置的特征改进网络,改进后 VSF－YOLO 结构如图 2 所示。

VSF－YOLO 算法的主要改进分为 4 部分:①YOLO v10n的 Backbone 替换为结合轻量化网络 VegsNet 的主干网络,增加主干网络非线性能力的同时减少整体的复杂程度。②在 Neck 部分使用 SRFConv(Receiving field attention conv)卷积替换基础卷积,能够动态地为不同特征赋予权重,从而增强牛只头部细粒度特征的提取能力。③在检测头前提出了一种 FFAA(Full-text feature anchor attention)注意力机制,用于捕获长距离像素间的上下文信息,并收敛特征关注区域。④优化损失函数,使用 SoftMax 损失函数结合 SRF 卷积,高亮区分特征区域,专注图像间的细粒度特征。

2.1 VegsNet 主干网络

为了使模型能够更好地适应各种不同的终端设

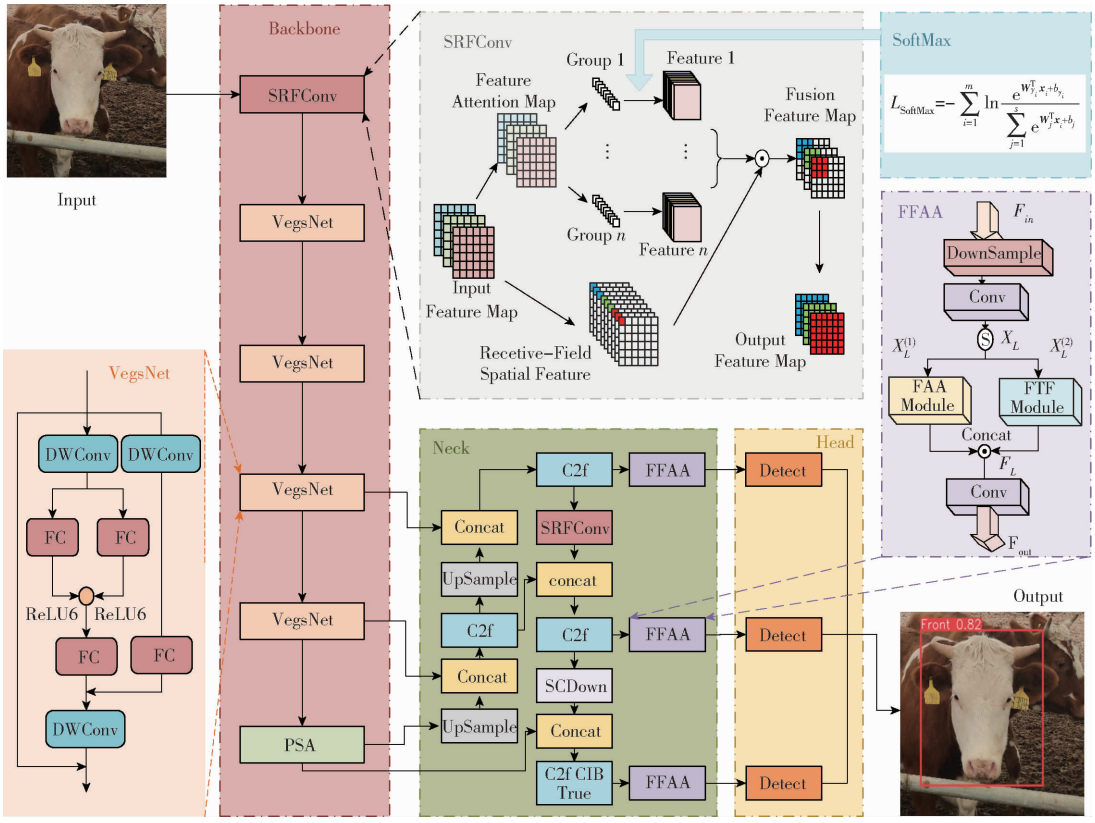

图 2 VSF－YOLO 网络整体结构图

Fig. 2 Overall structure diagram of VSF－YOLO network

备，并且减轻计算负载和减少模型参数量，受 Vanillanet 轻量化结构[27]启发，本文提出一种 VegsNet 网络结构并集成到 YOLO v10n 的主干网络中。模块结构如图 3 所示。

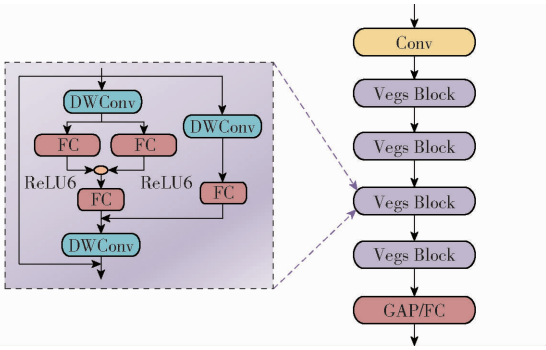

图 3　VegsNet 主干网络结构图

Fig. 3　VegsNet backbone network structure diagram

VegsNet 模块通过元素级乘法融合不同子空间特征，采用四阶段分层结构构建，利用卷积层下采样，并结合改良模块提取特征。每个模块内在深度可分离卷积后应用批量归一化，并在末尾再次引入深度可分离卷积。通道扩展因子设为 4，各阶段网络宽度依次翻倍。Vegs 模块采用 ReLU6 激活函数，主干结构中引入三通路特征提取方式，特征经深度可分离卷积与全连接层处理后融合输出。该设计在提升速度的同时，将参数量减少 33%。

新主干网络因层数较少且结构简洁，其非线性表达能力相对有限。因此通过元素间乘法融合了双通路的线性变换特征，提升特征提取维度，增强了网络非线性能力。定义单个通道特征系数矩阵和对应特征值分别表示为 $\boldsymbol{W}=[w_1,w_2,\cdots,w_d]$、$\boldsymbol{X}=[x_1,x_2,\cdots,x_d]$，其中 d 是输入通道数。从而得到 $(\boldsymbol{W}_1^{\mathrm{T}}\boldsymbol{X})(\boldsymbol{W}_2^{\mathrm{T}}\boldsymbol{X})$。为了简化，此模块专注于涉及单输出通道变换和单个元素输入的情况。定义 w_1、w_2 和 $x\in\mathbf{R}^{(d+1)\times 1}$，通过单层 Vegs 模块可将输出扩展到多通道的输出 $\boldsymbol{W}_1$、$\boldsymbol{W}_2\in\mathbf{R}^{(d+1)\times(d'+1)}$，一般情况下通道维度数扩展式为

$$\begin{aligned}\boldsymbol{w}_1^{\mathrm{T}}x*\boldsymbol{w}_2^{\mathrm{T}}x^{(i)}&=\left(\sum_{i=1}^{d+1}\boldsymbol{w}_1^{(i)}x^{(i)}\right)*\left(\sum_{j=1}^{d+1}\boldsymbol{w}_2^{(j)}x^{(j)}\right)=\\&\sum_{i=1}^{d+1}\sum_{j=1}^{d+1}\boldsymbol{w}_1^{(i)}\boldsymbol{w}_2^{(j)}x^{(i)}x^{(j)}=\\&\alpha_{(1,1)}x^{(1)}x^{(1)}+\cdots+\alpha_{(5,6)}x^{(5)}x^{(6)}+\cdots+\\&\alpha_{(d+1,d+1)}x^{(d+1)}x^{(d+1)}\end{aligned}\tag{1}$$

式中，i、j 代表对应通道数，α 为每个通道项目的系数，则 α 可以表达为

$$\alpha_{(i,j)}=\begin{cases}\boldsymbol{w}_1^{(i)}\boldsymbol{w}_2^{(j)} & (i=j)\\ \boldsymbol{w}_1^{(i)}\boldsymbol{w}_2^{(j)}+\boldsymbol{w}_1^{(j)}\boldsymbol{w}_2^{(i)} & (i\neq j)\end{cases}\tag{2}$$

上述操作通过将 d 个维度展开，将输出的维度扩展为 $\frac{(d+2)(d+1)}{2}$ 个不同项的组合，即 $\frac{(d+2)(d+1)}{2}$ 个维度。如式(1)所示，除了特定项外，每个项都与 x 呈非线性关系，从而增强了网络的非线性能力。各项都与 x 为非线性关系，提高了网络的非线性能力。

本文构建了一个由 1 个标准卷积层和 4 个 VegsNet 层组成的骨干网络。这种设计显著降低了模型的计算需求和参数量，便于部署到边缘性设备，如手机等。但是参数量的下降会导致模型精度下降，整体网络性能降低，针对此弊端本文使用改进网络 Neck 端的卷积，加入注意力机制、损失函数调整等方式提高整体模型性能。

2.2　SRFConv 卷积

在采集时牛只头部图像容易出现残缺、信息不完整的情况，如牲畜活动的不确定性会导致采集到的图像存在漏拍残缺或栏杆等遮挡情况，采集数据时存在部分干扰图像如图 4 所示。



(a) 正常图像　(b) 残缺图像　(c) 遮挡图像

图 4　采集到的正常图像、残缺图像和遮挡图像

Fig. 4　Collected normal images, incomplete images, and occluded images

图像的遮挡残缺会导致在特征提取时无法完全高效地获取有用特征，基于此本文提出空间感受野卷积(SRFConv)模块结合分组卷积与空间感受野，并通过损失函数个性化调整卷积核参数，提高同类图像之间的细粒度特征敏感度，通过不同的特征权重为有效信息赋予高权重，减少冗余信息的干扰。

为了实现更有效地提取细粒度区分特征，SRF 卷积采用双通道设计，一条为空间感受野通道，使用卷积滑动窗口获取图像中广泛的特征信息；另一条通道为细粒度特征通道，通过分组卷积获取图像中具有区分度的特征信息，并通过损失函数为其赋予相应权重。两通道融合形成最终特征图，共同捕获图像中的细粒度特征。SRFConv 卷积结构如图 5 所示(注：Group(N)表示大小为 $g\times g$ 的分组卷积，g 表示卷积核大小，X_i 表示输入的特征图，Adjust Shape 操作将结合的特征信息调整维度。F 是将细粒度特征注意力图与空间感受野特征图相乘得到的结果)。

在空间感受野通道中为了更有效地增强特征感

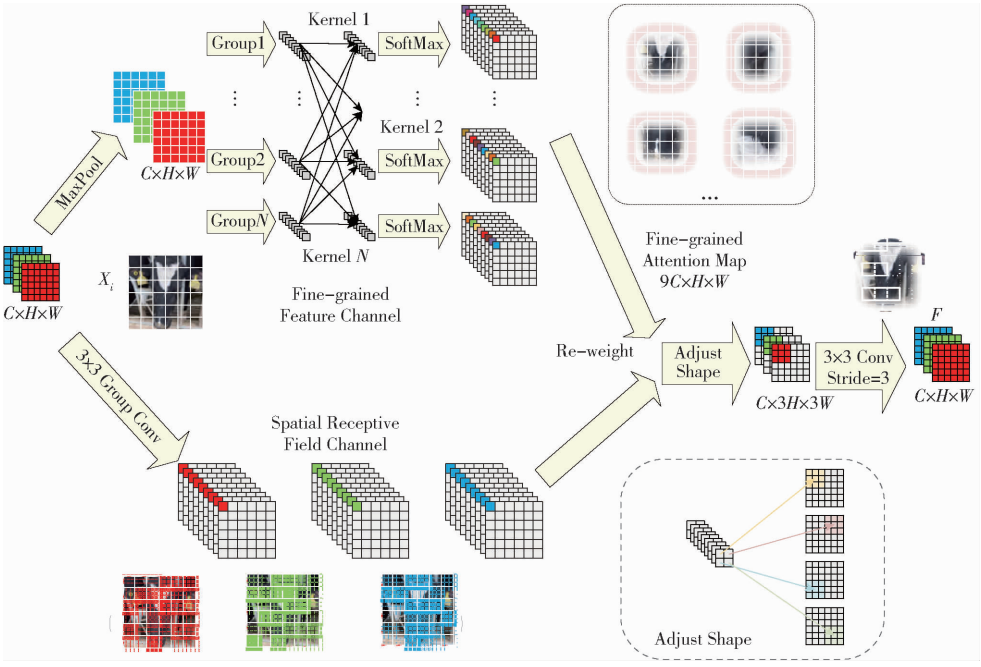

图 5 SRFConv 卷积结构图

Fig. 5 SRFConv convolutional structure diagram

知能力,精确捕捉图像中的特征信息,本文采用了 3×3 卷积核。每个 3×3 窗口在空间特征中代表一个感受野滑块,利用非重叠的滑动窗口机制,对特征图进行精确地感受野提取,从而确保网络能够全面捕捉到图像中的特征信息。

在细粒度特征通道中为实现细粒度特征的高效提取,采用分组卷积技术进行处理。采用最大池化(MaxPool)对输入特征的关键信息进行筛选聚合,使用 N 个卷积分组对图像特征进行分组提取后,原始特征被映射至不同的特征分组。通过 1×1 分组卷积操作实现信息交互后,使用优化后的 SoftMax 函数为每组的特征赋予权重,关注输入图像中具有细粒度特征的定位关键特征区域,生成用于后续卷积核权重分配的权重注意力特征图。

结合细粒度特征通道获得的权重注意力特征图和从空间感受野通道得到的特征图,借助细粒度特征信息获得专注于关键有效信息的特征图,并将其调整到适当尺寸(Adjust Shape),最后通过步长为 3 的 3×3 卷积将特征图尺寸还原,得到感受野注意力卷积的最终输出 F。

通过空间感受野卷积进行特征提取可以更好地关注到图像中的关键性区域,更多地聚焦图像之间的相似但有区别度的特征,借此对残缺或遮挡的图像中的有效信息分配更高的注意力特征权重,借此排除部分图像中的杂物遮挡信息和残缺信息,实现对遮挡、残缺情况的应对能力。

2.3 FFAA 注意力机制模块

在实际采集时牛只的运动会导致牛只头部图像产生运动模糊,特征的锐利度较差,面部的部分细节信息也无法完整地抓取,拍摄中产生运动模糊前后的图像如图 6 所示。



(a) 正常图像 (b) 模糊图像

图 6 采集到的正常图像和模糊图像

Fig. 6 Collected normal images and blurred images

对模糊图像的特征提取,要求准确捕捉输入图像的全局特征和关键区域信息,基于此本文提出一种全文特征锚点注意力机制 FFAA(Full-text feature anchor attention),旨在通过不同大小的深度可分离卷积、小核卷积和轴向卷积的结合,实现高效率获取多尺度全文信息的同时捕获全文锚点信息,减少计算成本。FFAA 注意力机制由全文特征提取模块 FTF 和全文锚点注意力模块 FAA 组成,在 FTF 模块中使用不同大小的深度可分离卷积进行特征提取融合;同时在 FAA 模块中结合平均池化和最大池化,使用轴向卷积提取特征并融合。FFAA 注意力机制结构如图 7 所示。

2.3.1 全文特征提取 FTF 模块

为了定位并识别模糊图像中对象的特征位置,本文提出 FTF 模块以捕获输入图像的多尺度纹理

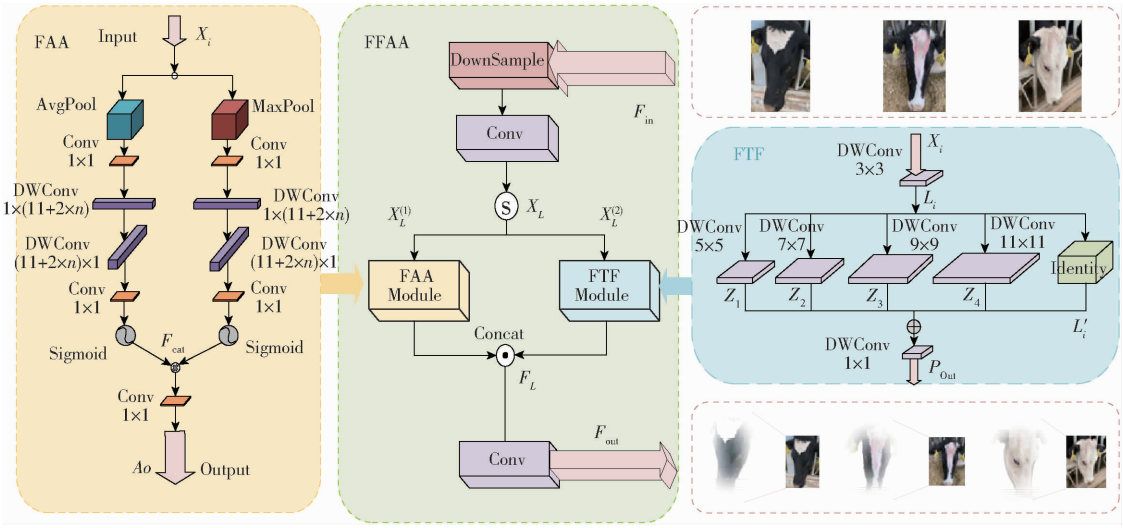

图 7　FFAA 注意力机制结构图

Fig. 7　FFAA attention mechanism structure diagram

信息，结合整体图像获得有效特征，在减少干扰的同时减少运动模糊对特征提取的干扰。FTF 模块包括一个小核卷积来捕捉局部信息，随后是一系列并行的深度可分离卷积以跨多个尺度捕获上下文信息，表达式为

$$L_i = \mathrm{Conv}_{3\times3}(X_i) \tag{3}$$

$$Z_i = \mathrm{DWConv}_{k(h)\times k(h)}(L_i) \tag{4}$$

其中，L_i为输入通过 3×3 卷积提取的特征，Z_i代表第 i 个深度可分离卷积提取到的上下文特征信息，深度分离卷积的大小由 $k(h)$ 函数决定，在本文中，$k(h)$满足函数 $k(h)=2h+3$。最后使用 1×1 卷积将不同大小深度可分离卷积提取到的局部特征和上下文特征融合，得到的输出可以表征不同通道之间的相互关系，同时保证两模块输出结构相似，便于后续融合。

FTF 模块通过其独特的提取机制，有效解决了牛只面部采集时因面部毛发和运动模糊导致的图像质量下降问题，在此复杂场景下，能够精确捕获输入图像的多尺度纹理信息，从而高效提取有效的面部信息。

2.3.2　FAA 全文锚点注意力模块

FFAA 注意力机制的核心是通过 FAA 模块捕捉图像中的锚点特征信息。一条通路借助最大池化等操作突出重要细节特征，另一条支路则通过平均池化等模块维持全局特征。在特征融合后，可以更精细地区分细节差异，从而实现更精准的检测与区分。

在通道处理过程中，分别对输入进行平均池化和最大池化操作，并通过 1×1 卷积核进行优化特征融合。接着，使用两个深度方向的轴向卷积来代替传统的大核卷积操作，对融合特征在两个深度方向上进行卷积运算。轴向卷积的使用可以显著减少参数量，从而提升模型效率。

最终，通过 Sigmoid 激活函数处理注意力特征图，使其数值范围限制在 0 到 1 之间。分别对两条不同通路的特征图使用 Sigmoid 函数处理后，采用 Fmix 方法进行特征融合。将这两条通路的特征图进行拼接，得到通道数为原来两倍的特征图，最后使用一个 1×1 小核卷积操作将其恢复到原始的通道数，从而高亮重要的特征区域。

2.3.3　全文特征锚点注意力 FFAA

为应对实际场景中运动模糊图像的特征提取，通过融合上述两模块，结合两模块优势构成 FFAA 注意力机制模块，更好地抓取图像中牛脸特征，更快速准确地识别牛脸。

首先经过一个下采样模块和一个卷积模块，初步对输入的图像提取特征，此处卷积设定为 3×3 卷积，初步提取输入特征，获得的X_L特征图沿主通道维度被分成两半，并输入到两个路径，分别送入两条支路进入 FAA 模块和 FTF 模块中，在 FAA 模块中获得图像整体特征和高注意值特征的联合关系，从 FTF 中获得全图特征，将二者结合得到输出F_L为融合特征，在整个特征图中给有效特征更高的注意力值。

通过全特征锚点注意力可以更好地在模糊图像中提取到牛只面部信息，准确地获取有效特征。同时在较为清晰的图像中也可以借助图像锚点信息和全图特征更好地将特征区域收敛，对关键信息有更好的抓取能力。

2.4　损失函数优化

损失函数优化可以促进模型训练的速度和准确度，改进空间感受野卷积对细粒度特征的抓取，增强复杂场景下的特征表达能力。在 SRF 卷积中，需要

确保注意力图能够聚焦于具有区分度的细节区域。此外,为了解决图像中因多姿态和多角度导致的细粒度细节难以捕捉、检测不准确等问题,本文采用优化 SoftMax 损失函数L_{SoftMax}来识别细粒度图像中的对象,表达式为

$$L_{\mathrm{SoftMax}} = -\sum_{i=1}^{m} \ln \frac{\mathrm{e}^{\boldsymbol{W}_{y_i}^{\mathrm{T}}\boldsymbol{x}_i + b_{y_i}}}{\sum_{j=1}^{s} \mathrm{e}^{\boldsymbol{W}_j^{\mathrm{T}}\boldsymbol{x}_i + b_j}} \tag{5}$$

其中,m 表示给定批次大小,$\boldsymbol{W}_{yi}^{\mathrm{T}}\boldsymbol{x}_i$表示上一层的输出,$y_i$表示第 i 幅图像的检测分类类别,$\boldsymbol{x}_i$表示第 i 幅图像的特征向量,b 表示网络偏移量,s 表示当前批次图像中有效检测的数量。本文使用 SoftMax 用于优化网络,在 SRF 卷积中在分组卷积处理后,对交互后的特征定位细粒度特征区域高效地抓取有效信息。

同时设计子损失函数,以持续优化和强化网络中的定位相关信息。为掩码分配了更为精确的位置,从而使得互补网络能够更好地区分并学习次要图像特征,表达式为

$$L_Z = \max\{0, p^k - p^{k+1}\} \tag{6}$$

式中,p^k表示来自分组卷积认定特征有效的概率,p^{k+1}表示认定非关键特征的概率,L_Z表示次要特征训练损失值。当$p^k \geqslant p^{k+1}$时,$L_Z > 0$,没有损失,但当$p^k \leqslant p^{k+1}$时,$L_Z \leqslant 0$,会发生训练损失。损失函数有助于强化后的卷积层找到更为精确的特征,而在分组提取特征后,损失函数能够更精确地定位目标。然而,区分主要和次要特征对于分组卷积提供的精确局部区域的权重分配至关重要,为了确保细粒度特征通道能够定位关键部分,模型的总损失L_{total}为

$$L_{total} = L_{\mathrm{SoftMax}} + \partial L_Z \tag{7}$$

式中,∂用于平衡两个损失函数的调衡系数,优化后的 SoftMax 损失函数稳定性好,结合概率输出的损失值,对筛选特征图中的关键区域有较好效果,为图像中有细粒度区分特征的区域赋予高注意力值,强化对这部分细节的提取区分,实现更好的细粒度特征提取。结合 SRF 卷积中细粒度特征通道,根据细粒度特征的关注为图像中的有效相似信息赋予较高的特征值,减少遮挡残缺带来的影响。

3 实验

3.1 实验环境和参数设置

为保证算法的公平性,本文算法训练环境均在 Ubuntu 18.04 Linux 操作系统下完成,Python 版本为 3.9,Pytorch 版本为 1.8.1,Cuda 版本为 11.8。训练硬件方面 CPU 采用 i9 - 14900K@2.50 GHz,GPU 使用 NVIDIA GeForce RTX 4090 显卡。

模型训练策略采用 SGD 优化器,总迭代次数设定为 2 000 次,批量大小为 64,权重衰减系数为 0.000 5。模型的初始学习率设置为 0.01,最佳权重文件用于模型评估。所有算法均在未加载预训练权重的情况下,在同一数据集上进行对比。训练数据集来自自建牛脸数据库,对牛只面部的左、中、右 3 角度面部分别赋予不同的标签进行训练。

3.1.1 模型评价指标

本文使用参数量(Params)和每秒千兆浮点运算次数(GFLOPs)评估模型的复杂性。为了评估奶牛头部的准确性,采用平均检测精度均值 mAP@0.5、精准率(Precision)和召回率(Recall)作为模型的性能评价指标。

3.1.2 模糊度评价指标

输入图像的模糊程度可以通过分析其清晰度特征(例如边缘强度、频率分布等)实现,将输入图像转换成灰度图像,再运用 Laplacian 算子来计算出图像梯度,从而精准地提取边缘信息。将 Laplacian 变换结果的方差作为图像清晰度的度量值,方差越小表明图像越模糊。模糊度指标 B 函数为

$$B = 1 - \frac{S - S_{\min}}{S_{\max} - S_{\min}} \tag{8}$$

式中 S——整体清晰度

$S_{\max}$——最大清晰度

$S_{\min}$——最小清晰度

通过参数的计算将其归一化为整体的模糊度指标 B,B 值越接近 1,图像越模糊。

3.1.3 残缺度评价指标

图像的遮挡或残缺程度通过分析图像的结构完整性和区域一致性来实现。通过区域分析找出大面积的单一颜色、噪点或结构丢失等异常来判断是否有遮挡残缺。定义残缺遮挡评价指标 Z 的计算函数为

$$Z = \frac{D - D_{\min}}{D_{\max} - D_{\min}} \tag{9}$$

式中 D——遮挡或残缺的总度量值

$D_{\min}$——最小遮挡程度

$D_{\max}$——最大遮挡程度

Z 作为整体遮挡残缺的评价值,当 Z 值趋近于 1 时,表明图像的遮挡残缺情况越严重。

3.2 消融实验

本文提出的细粒度网络 VSF - YOLO 使用 VegsNet 主干,SRF 卷积、FFAA 卷积改进以及损失函数的优化,改进算法的优越性效果通过消融实验测试,测试结果如表 1 所示。同时选择不同干扰程度图像测试复杂场景下算法的稳定性,结果如图 8 所示(复杂场景分析均为加入损失函数优化后的运行结果)。

表 1　模型消融实验结果

Tab. 1　Results of ablation experiments for different models

模型	mAP@0.5/%	精准率/%	召回率/%	网络层数	参数量	浮点运算次数
YOLO v10n	89.54	87.26	84.59	385	2.71×10^6	8.4×10^9
YOLO v10n + VegsNet	88.40	87.12	82.62	309	1.81×10^6	6.4×10^9
YOLO v10n + SRFConv	90.42	89.56	83.38	402	2.73×10^6	8.9×10^9
YOLO v10n + FFAA	91.58	90.24	84.51	415	2.97×10^6	9.5×10^9
YOLO v10n + VegsNet + SRFConv	90.19	89.39	85.84	326	1.84×10^6	6.8×10^9
YOLO v10n + VegsNet + FFAA	90.74	90.11	86.05	339	1.83×10^6	7.5×10^9
YOLO v10n + SRFConv + FFAA	92.86	92.72	88.98	462	2.98×10^6	9.8×10^9
YOLO v10n + VegsNet + SRFConv + FFAA	92.69	92.31	89.36	366	1.91×10^6	7.9×10^9
YOLO v10n + VegsNet + SRFConv + FFAA + loss function	93.09	92.91	90.01	366	1.91×10^6	7.9×10^9



图 8　复杂场景分析

Fig. 8　Complex scenario analysis

如表 1 所示，当加入 VegsNet 主干网络后模型参数量和浮点运算次数明显减少，分别达到 1.81×10^6 和 6.4×10^9，同时不可避免的参数下降导致精度略微下降；添加 SRFConv 模块时增加了模型参数量，mAP@0.5 提高了 0.88 个百分点；添加 FFAA 注意力机制模块后，mAP@0.5 明显提高到 91.58%。在添加 SRFConv 模块 + FFAA 注意力机制模块时，模型的参数量和浮点运算次数虽然远高于基线网络，但其 mAP@0.5 值得到显著提升，达到 92.86%。VegsNet 主干网络、SRFConv 模块和 FFAA 注意力机制模块的组合表现出最佳性能和参数量下降之间的平衡，参数量相比于原始 YOLO 网络减少了 29.5%，同时将 mAP@0.5、精准率和召回率提高了 3.15、5.05、4.77 个百分点，整体运行速度也提升了 7%。优化后损失函数的加入使网络在参数量不变的基础上，mAP@0.5 再度提升至 93.09%，损失函数的使用提高了卷积对图像特征的抓取能力，在精准率和召回率上都有提升。

图 8 可以看出，同样程度的干扰下 VSF－YOLO 相较于原 YOLO v10n 网络显示出了更好的检测效果。在 YOLO v10n 网络加入 VegsNet 主干网络后 mAP@0.5 有不同程度的下降，在加入 SRFConv 卷积和 FFAA 注意力机制后检测能力显著提升并超越原 YOLO 网络，在复杂场景干扰下也拥有更强的稳定性和抗干扰性。

同时本文利用 P－R 曲线对 YOLO v10n 网络和 VSF－YOLO 网络的模型性能进行验证，得到结果如图 9 所示。VSF－YOLO 模型的 P－R 曲线整体位于原模型之上，在各置信度阈值下均表现出更优的精准率与召回率平衡。表明 VSF－YOLO 模型的检测性能整体获得较大提升，特别是在保持较高召回率的同时依然维持了更好的精度。

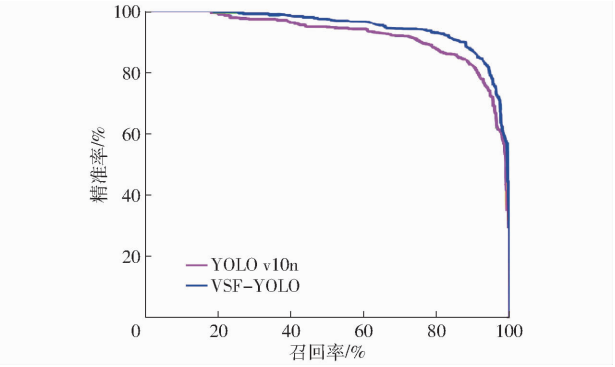

图 9 精准率-召回率曲线

Fig. 9 Precision - recall curves

综上可知,VSF - YOLO 相较原网络在模型轻量化上有较大的提升,网络层数下降,参数量下降约 30%,同时在牛脸检测方面的平均检测精度 mAP@0.5 提升 3.55 个百分点,且在检测任务的复杂场景干扰中,有效特征的抓取能力明显提高,得到的检测结果更优秀,证明 VSF - YOLO 在牛脸检测领域可以在参数量减少的同时拥有较好的检测能力。

3.3 对比实验

为了验证改进 VSF - YOLO 算法的综合性能,在数据集上将 VSF - YOLO 算法与目前其他主流轻量化检测算法(YOLO v4 - tiny[28]、YOLO v5n[29]、YOLOX - tiny[30]、YOLO v7 - tiny[31]、YOLO v8n[32]和 MobileNet[33])进行比较。由此证明所提 VSF - YOLO 系列算法的优势,同时通过挑选 100 幅牛脸图像批量验证。通过统计总运行时间后除以图像数,得到每幅图像的平均运行速度指标(单位:s),以此展示模型轻量化运行能力,结果如表 2 所示。

表 2 与其他轻量化算法比较结果

Tab. 2 Comparison results with other lightweight algorithms

模型	mAP@0.5/%	精准率/%	召回率/%	参数量	浮点运算次数	平均运行速度/s
YOLO v4 - tiny	89.73	88.27	86.98	6.06×10^6	6.960×10^9	0.023 3
YOLO v5n	91.11	92.39	86.82	7.28×10^6	1.716×10^{10}	0.024 9
YOLOX - tiny	91.89	90.60	87.90	5.06×10^6	1.536×10^{10}	0.023 9
YOLO v7 - tiny	92.23	91.21	88.41	6.23×10^6	1.386×10^{10}	0.023 5
YOLO v8n	90.73	92.15	89.07	3.01×10^6	8.200×10^9	0.023 4
MoblieNet	91.64	90.97	89.79	3.27×10^6	1.050×10^{10}	0.024 1
VSF - YOLO	93.09	92.91	90.01	1.91×10^6	7.900×10^9	0.022 9

从表 2 中可以看出,所有模型都是参数量小于 8×10^6 的轻量级检测模型,尽管 YOLO v5n、YOLO v7 - tiny 和 YOLOX - tiny 在精度上表现优异,但其参数量和浮点运算次数却远超其他算法,YOLO v5n 运行速度在所有测试算法中平均耗时最长。同时 YOLO v4 - tiny 算法以 6.96×10^9 的浮点运算次数在所有对比模型中脱颖而出,其低运算次数和少参数的特性使其成为移动终端部署的理想选择,但该算法的 mAP@0.5 稍低。此外,YOLO v8n 算法参数量仅为 3.01×10^6,在模型复杂度方面表现良好,但检测精度相对不足。在 MoblieNet 网络测试中,mAP@0.5 也达到了较高的 91.64%,精准率和召回率也较高,相比于 YOLO v8n 精度更高但是参数量也更大,同时算法的平均运行时间也以 24.1 ms 长于 YOLO v8n 的 23.4 ms。而 YOLO v7 - tiny 算法是除 VSF - YOLO 算法外检测精度最高的模型,其 mAP@0.5 达到 92.23%,参数量为 6.23×10^6,在模型复杂度和检测精度之间取得了良好平衡。相比之下,本文模型在头部检测任务中表现得更加出色,各项指标均取得了较好的结果,mAP@0.5 达到了 93.09%,并且参数量仅为 1.91×10^6,在测试预测任务中平均耗时最低,实现了最高的精度和最低的模型复杂度。

根据图 10 中特征热力图可看出,相较于其他轻量化网络,VSF - YOLO 的特征抓取和重要信息捕捉能力都有明显的优势。与前代 YOLO v7 - tiny 和 YOLO v8n 相比,VSF - YOLO 对面部区域的捕获能力明显提高,对图像中有效特征区域的收敛能力也有显著优化。同时得益于细粒度特征和全文锚点信息的感知应用,与 MoblieNet 网络比较也展现出更为优秀的特征提取能力。

4 系统部署测试

4.1 算法手机 Android 端部署

系统使用 Android Studio 平台部署[34],调用手机摄像头获取图像进行检测,通过算法标出检测框后输出在手机屏幕上。项目设置兼容 Android 版本最低为 10.0,SDK 版本为 31.0.0,NDK 使用 24.0.8215888 版本,CMake 使用 3.10.2 版本,ncnn 版本为 20231027 - android - vulkan。测试阶段通过 Google USB Driver 模块连接手机,测试机为红米 K60pro,Android 版本为 14.0,通过 USB 连接测试算法。

将训练优化后的目标检测模型导出为 ONNX

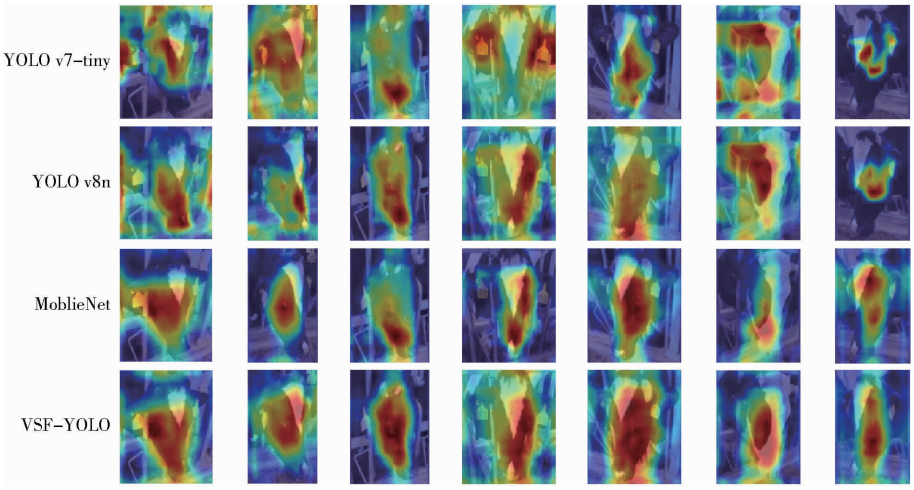

图 10　特征热力图对比

Fig. 10　Comparison of feature heatmaps

文件，以确保其可部署性和跨平台可用性。为了将其部署在 Android 设备上，使用 NCNN 将 ONNX 文件转换为 param 文件和 bin 文件，随后将这些文件加载到 Android Studio 项目中，编译并导入到移动设备上。部署后的 APP 界面如图 11 所示。

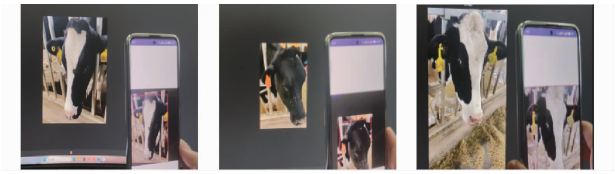

图 11　APP 使用示例

Fig. 11　APP usage example

算法部署后对数据集外的牛只进行拍摄测试，这种方法充分利用了设备的处理器架构和计算能力，提高了模型的运行效率和性能。

4.2　移动端应用测试

在实际应用场景中常常遇到遮挡或牛只乱动导致图像有用信息残缺，导致面部特征提取效率低下。在 Android 系统上部署的 VSF－YOLO 算法，能够有效地采集并分析复杂场景下的牛只图像，移植软件设定采集图像为 320 像素 ×320 像素，调用后置摄像头采集。将 YOLO v10n 和 VSF－YOLO 模型分别部署到移动端 APP 中，测试结果如图 12 所示。

在软件测试环节，选取了面部特征不完整、运动状态下模糊以及面部存在大量干扰信息的 30 头牛只作为测试对象，图 12 所示为其中代表性的 3 幅，其中牛只 1 和牛只 2 的遮挡残缺评价指标分别达到 0.78 和 0.69，显示出较为严重的遮挡情况；而牛只 3 的模糊度评价指标为 0.75，图像模糊程度较高。

在测试机上部署基于 YOLO v10n 原版训练的模型，在上述程度的复杂场景下无法检测到测试牛只面部（牛只 1 和牛只 3 未检测到面部）或检测置信度较低，仅为 0.44 的置信度，同时软件运行平均

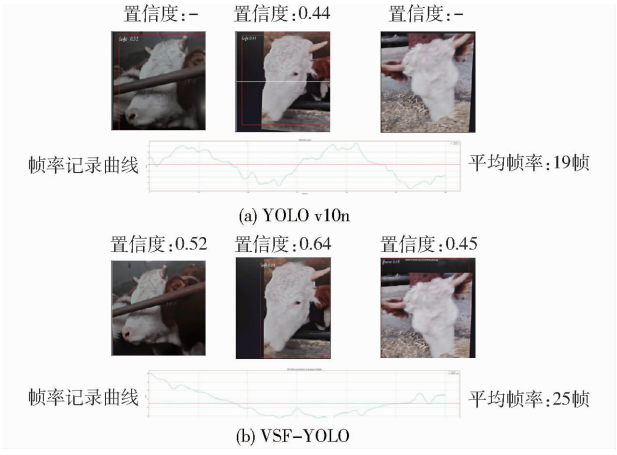

图 12　移动端实际测试结果

Fig. 12　Actual testing on mobile devices

帧率较低，帧率跟踪曲线平均仅有 19 帧；在未连接充电线运行软件 10 min 耗电 15%，原 YOLO v10n 模型在手机端的运行能力有待优化。而在同一测试机中部署了 VSF－YOLO 网络模型的应用程序检测能力明显增强，在测试检测中 30 头牛只的置信度平均提升 36%，优化后的软件平均运行帧率达到 25 帧，整体的运行流畅度提高；同时满电 10 min 的持续运行测试后电量剩余 90%，在运行效率和能效上都有明显提升。

上述实验可以看出在相同的测试平台上使用相同的部署策略，在手机这类算力相对较低的平台上使用也能看出，优化后的模型相比原模型运行帧率提升约 30%，模型的准确率有明显的提高，更能适应手机端的算力，更好地应对拍摄中的复杂场景干扰。

5　结论

（1）本文将改进 VSF－YOLO 算法应用到复杂场景下的牛脸检测任务中。主干网络改进为自研的

轻量化 VegsNet 主干网络,模型参数量相比原始网络减少了 29.5% 并流畅部署于移动端。

(2)为应对复杂场景中牛只面部遮挡、残缺导致检测精确度下降的问题,本研究提出一种关注细粒度差异特征的空间感受野卷积 SRFConv 和改进 SoftMax 损失函数,更好的关注图像中的关键区域,提高对遮挡、残缺干扰的应对能力。

(3)针对牛场中牛只乱动导致拍摄模糊的状况,提出一种全文特征锚点注意力机制 FFAA,捕捉并结合图像的多尺度信息和全文锚点信息,在模糊的图像中可以找到牛只面部,增加模型在实际使用中对复杂场景的抗性。

参 考 文 献

[1] Yang W. Research on key technologies of smart breeding platform based on 5G[J]. Applied Mathematics and Nonlinear Sciences, 2024, 9(1): 118 - 132.

[2] Wang F, Chen W. Exploring the path of development of international trade of agricultural products in the context of digitalization [J]. Applied Mathematics and Nonlinear Sciences, 2024, 9(1): 201 - 213.

[3] Zhang Q, Ahmed K, Sharda N. A comprehensive survey of animal identification: exploring data sources, AI advances, classification obstacles and the role of taxonomy[J]. International Journal of Intelligent Systems, 2024, 1(1): 1 - 20.

[4] Bi Y, Campos L M, Wang J, et al. Depth video data-enabled predictions of longitudinal dairy cow body weight using thresholding and Mask R - CNN algorithms[J]. Smart Agricultural Technology, 2023, 6: 100352.

[5] Cui G, Qiao L, Li Y. Division of cow production groups based on SOLOv2 and improved CNN - LSTM[J]. Agriculture, 2023, 13(8): 1562.

[6] 齐咏生,全家乐,栾浩天,等. 基于多尺度融合网络的轻量化牛脸识别算法[J]. 农业机械学报, 2025,56(12): 615 - 622,644.
Qi Yongsheng, Quan Jiale, Luan Haotian, et al. Lightweight cattle face recognition algorithm based on multi-scale fusion network [J]. Transactions of the Chinese Society for Agricultural Machinery, 2025,56(12):615 - 622,644. (in Chinese)

[7] Sun L, Liu G, Yang H, et al. LAD - RCNN: a powerful tool for livestock face detection and normalization[J]. Animals, 2023, 13(9): 1446.

[8] Zhao Z. Application of improved CNN-based face detection technology in public administration[J]. Journal of Computational Methods in Sciences and Engineering, 2023, 23(4): 1985 - 1997.

[9] Abdulhussien D M, Saud L J. An evaluation study of face detection by Viola-Jones algorithm[J]. International Journal of Health Sciences, 2022, 6(S8): 4174 - 4182.

[10] Lodkaew T, Pasupa K, Loo C K. CowXNet: an automated cow estrus detection system[J]. Expert Systems with Applications, 2023, 211: 118550.

[11] Siregar A M, Purwanto Y A, Wijaya S H, et al. Two-stages of segmentation to improve accuracy of deep learning models based on dairy cow morphology[J]. International Journal of Electrical and Computer Engineering, 2023, 13(2): 2093.

[12] Dac H H, Gonzalez Viejo C, Lipovetzky N, et al. Livestock identification using deep learning for traceability[J]. Sensors, 2022, 22(21): 8256.

[13] Hou H, Shi W, Guo J, et al. Cow rump identification based on lightweight convolutional neural networks[J]. Information, 2021, 12(9): 361.

[14] Bergman N, Yitzhaky Y, Halachmi I. Biometric identification of dairy cows via real-time facial recognition[J]. Animal, 2024, 18(3): 101079.

[15] Hussain M. YOLO - v1 to YOLO - v8, the rise of YOLO and its complementary nature toward digital manufacturing and industrial defect detection[J]. Machines, 2023, 11(7): 677.

[16] 潘浩文,何梦腾,邓洪兴,等. 基于 DCT - YOLO 的轻量化奶牛爬跨行为检测方法[J]. 农业机械学报, 2026, 57(2): 245 - 255.
Pan Haowen, He Mengteng, Deng Hongxing, et al. Lightweight cow mounting behavior detection method based on DCT - YOLO[J]. Transactions of the Chinese Society for Agricultural Machinery, 2026,57(2):245 - 255. (in Chinese)

[17] Guo S S, Lee K H, Chang L, et al. Development of an automated body temperature detection platform for face recognition in cattle with YOLO V3 - tiny deep learning and infrared thermal imaging[J]. Applied Sciences, 2022, 12(8): 4036.

[18] Wang R, Bai Q, Gao R, et al. Oestrus detection in dairy cows by using atrous spatial pyramid and attention mechanism[J]. Biosystems Engineering, 2022, 223: 259 - 276.

[19] Kawagoe Y, Kobayashi I, Zin T T. Facial region analysis for individual identification of cows and feeding time estimation[J]. Agriculture, 2023, 13(5): 1016.

[20] Qin B, Zeng Y, Wang X, et al. Lightweight DB - YOLO mask-wearing intelligent detection and android application based on bidirectional weighted feature fusion[J]. Electronics, 2023, 12(24):4936.

[21] Luan T, Zhou S, Liu L, et al. Tiny-object detection based on optimized YOLO - CSQ for accurate drone detection in wildfire scenarios[J]. Drones, 2024, 8(9): 454.

[22] Mu Y, Hu J, Wang H, et al. Research on the behavior recognition of beef cattle based on the improved lightweight CBR－YOLO model based on YOLO v8 in multi-scene weather[J]. Animals, 2024, 14(19): 2800.

[23] 齐咏生,张新泽,张嘉英,等. 基于特征掩膜的局部遮挡牛脸识别方法[J]. 农业机械学报, 2024, 55 (11): 93－102.
Qi Yongsheng, Zhang Xinze, Zhang Jiaying, et al. Local occluded cattle face recognition method based on feature masking [J]. Transactions of the Chinese Society for Agricultural Machinery, 2024, 55(11): 93－102. (in Chinese)

[24] 李祥霞, 吉晓慧, 李彬. 细粒度图像分类的深度学习方法[J]. 计算机科学与探索, 2021, 15(10): 1830.
Li Xiangxia, Ji Xiaohui, Li Bin. Deep learning methods for fine-grained image classification[J]. Computer Science and Exploration, 2021,15(10):1830. (in Chinese)

[25] Yu H, Zhang X, Wang Y, et al. Fine-grained accident detection: database and algorithm[J]. IEEE Transactions on Image Processing, 2024, 33: 1059－1069.

[26] Wang A, Chen H, Liu L, et al. YOLO v10: real-time end-to-end object detection[J]. Advances in Neural Information Processing Systems, 2024, 37: 107984－108011.

[27] Chen H, Wang Y, Guo J, et al. Vanillanet: the power of minimalism in deep learning[J]. Advances in Neural Information Processing Systems, 2023, 36: 7050－7064.

[28] Liu Q, Fan X, Xi Z, et al. Object detection based on YOLO v4－Tiny and improved bidirectional feature pyramid network[C]// Journal of Physics: Conference Series. IOP Publishing, 2022, 2209: 012023.

[29] Zhao Y, Hu C, Li H, et al. The wind turbine nacelle oil-level detection algorithm based on improved YOLO v5n－Seg[C]// Journal of Physics: Conference Series. IOP Publishing, 2024, 2782: 012071.

[30] Ru C, Zhang S, Qu C, et al. The high-precision detection method for insulators' self-explosion defect based on the unmanned aerial vehicle with improved lightweight ECA－YOLOX－Tiny model[J]. Applied Sciences, 2022, 12(18): 9314.

[31] Yang Y, Qiang Q. An improved PCB defect detection algorithm for YOLO v7－tiny[C]// International Conference on Mechatronic Engineering and Artificial Intelligence (MEAI 2023). SPIE, 2024: 646－652.

[32] Jiang T, Li Y, Feng H, et al. Research on a trellis grape stem recognition method based on YOLO v8n－GP[J]. Agriculture, 2024, 14(9): 1449.

[33] Xu B, Wang W, Guo L, et al. CattleFaceNet: a cattle face identification approach based on RetinaFace and ArcFace loss[J]. Computers and Electronics in Agriculture, 2022, 193: 106675.

[34] Abbaz P. Real time logo recognition using YOLO on android[J]. Eurasian Journal of Academic Research, 2024, 4(7S): 1094－1099.