基于改进 Tiny-YOLO 模型的群养生猪脸部姿态检测

燕红文[1]，刘振宇[1]，崔清亮[2※]，胡志伟[1]，李艳文[1]
（1. 山西农业大学信息科学与工程学院，太谷 030801；2. 山西农业大学工学院，太谷 030801）

摘 要：生猪脸部包含丰富的生物特征信息，对其脸部姿态的检测可为生猪的个体识别和行为分析提供依据，而在生猪群养场景下，猪舍光照、猪只黏连等复杂因素给生猪脸部姿态检测带来极大挑战。该文以真实养殖场景下的群养生猪为研究对象，以视频帧数据为数据源，提出一种基于注意力机制与 Tiny-YOLO 相结合的检测模型 DAT-YOLO。该模型将通道注意力和空间注意力信息引入特征提取过程中，高阶特征引导低阶特征进行通道注意力信息获取，低阶特征反向指引高阶特征进行空间注意力筛选，可在不显著增加参数量的前提下提升模型特征提取能力、提高检测精度。对 5 栏日龄 20～105 d 的群养生猪共 35 头的视频抽取 504 张图片，共计 3 712 个脸部框，并标注水平正脸、水平侧脸、低头正脸、低头侧脸、抬头正脸和抬头侧脸 6 类姿态，构建训练集，另取 420 张图片共计 2 106 个脸部框作为测试集。试验表明，DAT-YOLO 模型在测试集上对群养生猪的水平正脸、水平侧脸、低头正脸、低头侧脸、抬头正脸和抬头侧脸 6 类姿态预测的 AP 值分别达到 85.54%、79.30%、89.61%、76.12%、79.37%和 84.35%，其 6 类总体 mAP 值比 Tiny-YOLO 模型、仅引入通道注意力的 CAT-YOLO 模型以及仅引入空间注意力的 SAT-YOLO 模型分别提高 8.39%、4.66%和 2.95%。为进一步验证注意力在其余模型上的迁移性能，在同等试验条件下，以 YOLOV3 为基础模型分别引入两类注意力信息构建相应注意力子模型，试验表明，基于 Tiny-YOLO 的子模型与加入相同模块的 YOLOV3 子模型相比，总体 mAP 指标提升 0.46%～1.92%。Tiny-YOLO 和 YOLOV3 系列模型在加入注意力信息后检测性能均有不同幅度提升，表明注意力机制有利于精确、有效地对群养生猪不同类别脸部姿态进行检测，可为后续生猪个体识别和行为分析提供参考。

关键词：图像处理；模型；目标检测；Tiny-YOLO；通道注意力；空间注意力
doi：10.11975/j.issn.1002-6819.2019.18.021
中图分类号：TP391 **文献标志码**：A **文章编号**：1002-6819(2019)-18-0169-11

燕红文，刘振宇，崔清亮，胡志伟，李艳文．基于改进 Tiny-YOLO 模型的群养生猪脸部姿态检测[J]．农业工程学报，2019，35(18)：169－179． doi：10.11975/j.issn.1002-6819.2019.18.021 http://www.tcsae.org
Yan Hongwen, Liu Zhenyu, Cui Qingliang, Hu Zhiwei, Li Yanwen. Detection of facial gestures of group pigs based on improved Tiny-YOLO[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2019, 35(18): 169－179. (in Chinese with English abstract) doi：10.11975/j.issn.1002-6819.2019.18.021 http://www.tcsae.org

0 引 言

随着生猪养殖规模不断扩大，养殖密度不断增加，对群养环境中的生猪个体进行自动有效识别，为其建档立卡并构建养殖可追溯系统，对实现猪场养殖的精准管理具有重要意义。生猪脸部的眼、鼻、耳等可辨识性个体信息对生猪个体识别至关重要，准确有效地生猪脸部检测可为生猪个体识别与规模化养殖决策制定提供技术支撑[1-2]。

随着深度学习技术的成熟，基于卷积神经网络 CNN（convolutional neural network）的视觉分析技术在生猪姿态检测[3]、生猪图像分割[4-7]、生猪个体识别[8]等诸多领域取得较大进展，并在目标检测领域同样表现出优越性能[9-13]。CNN 检测框架包括基于区域[14-15]和基于回归两大系列。其中以 YOLOV1[16]、YOLOV2[17]、YOLOV3[18]和 Tiny-YOLO[19]为代表的基于回归的方法能在保证检测精度的同时提升检测速度，适用于规模化生产环境中，并已被用于芒果[20]、苹果[21]、生猪个体[22]等目标的检测定位。而在生猪脸部姿态检测方面，经现有文献查证，未有学者对此领域做过相关研究。此外上述基于 YOLO 的系列方法认为特征图中每个区域对模型最终检测结果贡献度相同，而在群养生猪脸部姿态检测中，猪体、猪粪、猪食等噪声信息均不利于脸部位置精确提取，若可有效抑制此类信息，并对猪脸所在区域特征施以较高权重则可更好的提升检测精度。注意力机制的出现可有效解决此类问题，该机制在处理信息时只关注部分有利于任务实现的区域信息，滤除次要信息以提升模型效果，并已在图像分类[23]、图像分割[24]领域得到成熟应用，而在目标检测领域仍处于探索阶段[25]，因而探讨将该机制用于猪舍场景下群养生猪脸部定位成为可能。

基于此，本文提出一种基于 Tiny-YOLO 模型的非接触、低成本的群养生猪脸部姿态检测新方法，该方法将通道注意力和空间注意力机制相结合以构建双重注意力子模型 DAT-YOLO 进行端到端训练，实现对群养生猪水平正脸、水平侧脸、低头正脸、低头侧脸、抬头正脸和

收稿日期：2019-08-16 修订日期：2019-08-28
基金项目：国家高技术研究发展计划（863 计划）资助项目（2013AA102306）；国家自然基金面上项目资助（31772651）；山西省重点研发计划专项（农业）（201803D221028-7）
作者简介：燕红文，博士生，主要研究方向为农产品加工新技术及装备、计算机视觉技术。Email: yhwhxh@126.com
※通信作者：崔清亮，教授，博导，主要从事旱作农业机械化关键技术与装备的研究。Email: qlcui@126.com

抬头侧脸 6 类姿态高精度检测，避免猪只黏连、猪舍光照等干扰因素对检测效果的影响，以期为生猪智能养殖与管理提供技术参考。

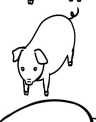
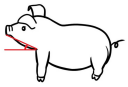
1 试验数据

1.1 数据采集

数据采集自山西省汾阳市冀村镇东宋家庄村，为获取不同猪舍场景的生猪图像，于 2019 年 6 月 1 日 9:00-14:00（晴，光照强烈）选取 3 个猪场进行视频采集，每个猪场由 10～30 间猪栏构成，每栏数量 5～8 只不等，猪栏大小约为 3.5 m×2.5 m×1 m。选取 5 栏日龄 20～105 d 的群养生猪共计 35 头，采用佳能 700D 防抖镜头，移动拍摄时生猪距离镜头 0.3～3 m 不等，因而可用于采集不同大小猪脸区域。因实际场景下群养生猪脸部姿态具有随机性，并非均是正脸面朝镜头，故将脸部姿态细分为正脸与侧脸，同时，不同角度生猪个体脸部蕴含信息差异较大，故最终将脸部姿态细化标注为水平正脸、水平侧脸、低头正脸、低头侧脸、抬头正脸和抬头侧脸 6 类。标注时将耳部作为脸部与身体部位分界点，且对未出现在采集范围或眼部未出现在镜头中的脸部不做任何标注，其每类姿态详细标注原则如表 1 所示。

表 1 生猪脸部 6 类姿态定义

Table 1 6 types of gestures definition of pig face

姿态 Gestures	姿态示意图 Gestures sketch	姿态描述 Gestures description
水平正脸 Horizontal face		两眼均位于镜头范围，嘴部最低点与颈部间连线和颈部水平方向夹角在±10°以内
水平侧脸 Horizontal side face		单眼位于镜头范围，嘴部最低点与颈部间连线和颈部水平方向夹角在±10°以内
低头正脸 Bow down face		两眼均位于镜头范围，嘴部最低点与颈部间连线和颈部水平方向夹角超过-10°
低头侧脸 Bow down side face		单眼位于镜头范围，嘴部最低点与颈部间连线和颈部水平方向夹角超过-10°
抬头正脸 Look up face		两眼均位于镜头范围，嘴部最低点与颈部间连线和颈部水平方向夹角超过+10°
抬头侧脸 Look up side face		单眼位于镜头范围，嘴部最低点与颈部间连线和颈部水平方向夹角超过+10°

注：夹角的正负定义为嘴部最低点与颈部连线的夹角在颈部水平方向上方为正，在颈部水平方向下方为负。

Note: The positive and negative angle is defined as positive when the angle between the lowest point of mouth and the line of neck is above the horizontal direction of neck and negative below the horizontal direction of neck.

1.2 数据预处理

将采集的视频数据进行下列操作以构建生猪脸部姿态检测数据集：

1）对采集的视频做切割视频帧处理，对获取到的 1 920×1 080 分辨率图像边缘添加黑色像素值操作以使其宽高比为 2:1，完成后像素值变为 2 048×1 024，并采用 labelImg[26]作为脸部姿态标注工具，其过程如图 1a～1b 所示。

2）因本文检测模型输入分辨率为 416×416，故对上述处理后的图片每 2 张进行拼接操作得到方形图片，其分辨率为 2 048×2 048，对获取的图片做放缩操作，将分辨率最终转换为 416×416，以减少运算量，提高模型训练速度，同时对步骤 1）中所标注的脸部位置进行相应坐标变换，以获取放缩后图像对应的脸部坐标信息，其过程如图 1c～1d 所示。由图 1 可见，虽然猪的头部也位于镜头采集范围内，但其眼部并未呈现，故本文未对其进行标注。

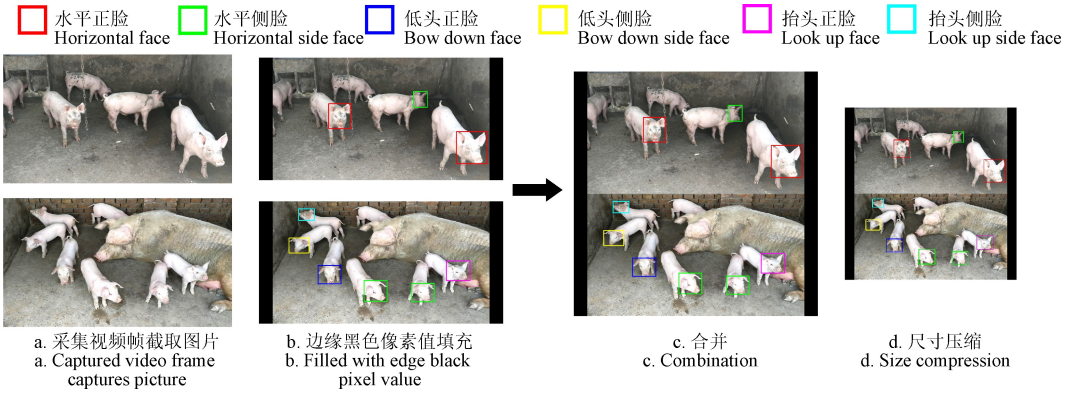


图 1 数据处理过程

Fig.1 Data processing

经上述 2 步处理后，本文共得到标注图像 989 张，按照通用数据集划分策略[27]，将其中 504 张作为训练集，65 张作为验证集，420 张作为测试集。训练集包含 3 712 个生猪脸部框，测试集包含 2 106 个猪脸框，对训练集与测试集中每类姿态标注框数量统计结果如表 2 所示。由表 2 可知，训练集与测试集上 6 种姿态数目不等会带来数据不均衡问题，本研究对该问题的处理见 5.1 讨论部分所示。

表 2　训练集测试集各个姿态类别数量

Table 2　Numbers of categories gestures on training and test set

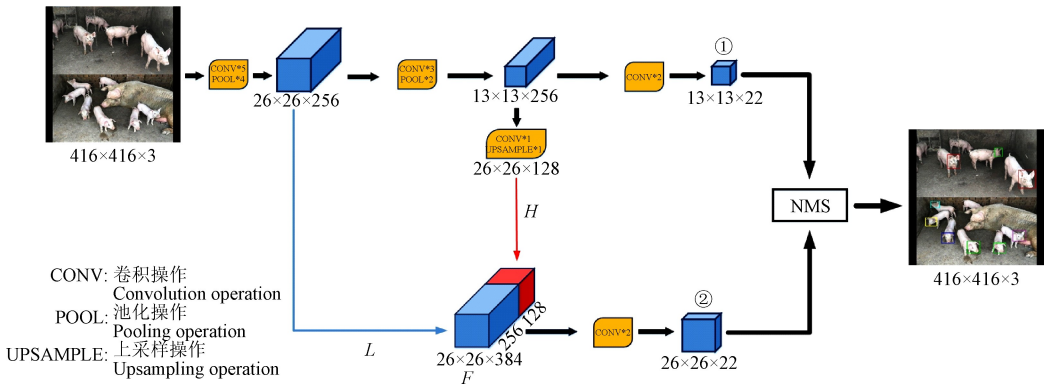
数据集 Data set	水平正脸 Horizontal face	水平侧脸 Horizontal side face	低头正脸 Bow down face	低头侧脸 Bow down side face	抬头正脸 Look up face	抬头侧脸 Look up side face
训练集 Training set	636	1 288	315	561	326	586
测试集 Test set	422	697	194	284	220	289

2　检测模型

2.1　Tiny-YOLO 模型

YOLOV1、YOLOV2 和 YOLOV3 是 Joseph 等[16-18]提出的目标检测通用模型，Tiny-YOLO 是轻量化 YOLOV3，其融合了最新的特征金字塔网络[12]（feature pyramid networks，FPN）和全卷积网络[28]（fully convolutional networks，FCN）技术，模型结构更简单，检测精度更快[29]。

Tiny-YOLO 模型主要由卷积层与池化层拼接构成，其模型结构如图 2 所示。模型输入图像分辨率为 416×416×3，经过一系列 3×3 和 1×1 卷积、池化以及上采样操作，可对输入图像进行特征提取，每步提取完成后特征图尺寸如图 3 中长方体底部数字所示，数字项分别表示特征图分辨率宽×分辨率高×通道数。由于不同种类目标在原始图中所占比例差异较大，Tiny-YOLO 引入多尺度特征提取模块以保证对不同大小目标均具有较强的检测性能。模型最终输出 2 个尺度特征图，如图 2 中①、②所示。多尺度特征图对于同一目标可能会有多个检测结果，Tiny-YOLO 引入非极大值抑制（non maximum suppression, NMS）[30]剔除冗余的检测框以使得对于每个目标均有唯一检测框，使其位置信息更准确，置信度更高。但 Tiny-YOLO 模型在特征提取过程中，对特征图中的每个区域均赋予相同权重值，而在生猪脸部姿态检测中，图像中的猪脸、猪蹄和猪舍等部位对猪脸区域精确定位贡献度不同，应赋予不同权重值。降低猪蹄、猪舍等噪声信息的影响，强化猪脸区域特征，可提升定位准确度。



注：L 表示低阶特征图，H 表示高阶特征图信息，F 表示高低阶融合特征图，①、②分别表示对输入图像的 2 种不同尺度检测结果。

Note: L represents low-order feature map, H represents high-order feature map information, F represents high-low-order fusion feature map, ① and ② represents two different scale detection results for input images.

图 2　Tiny-YOLO 模型结构图

Fig.2　Model structure of Tiny-YOLO

2.2　对 Tiny-YOLO 模型的改进

2.2.1　引入通道注意力模块

因不同通道信息对检测结果贡献不同，本文引入通道注意力（channel attention block, CAB）模块对特征图各个通道间的依赖性进行建模，可使同一特征图的不同位置具有相同的通道权重信息，使模型能选择性地强化重要信息并抑制弱相关特征，进而提高模型表征能力，可精确定位生猪脸部，其结构如图 3 所示。全局平均池化[31-32]常被用于汇集空间通道信息，该操作通过压缩输入特征图空间维度生成特征图像素点反馈信息以计算通道注意力，但其断然将特征图中的每一点对通道注意力信息的获取视为具有同等作用，削弱了特征强度较大区域对通道注意力信息的影响。而全局最大池化在梯度反向传播过程中仅计算响应最大地方的梯度反馈，可进一步强化敏感区域以弥补全局平均池化的短板。为此，本文 CAB 模块在传统通道注意力模块中加入全局最大池化操作，通过对高阶特征进行全局平均与最大池化融合得到通道权重向量以引导低阶特征图进行通道选择，实现特征响应与特征重校准效果。

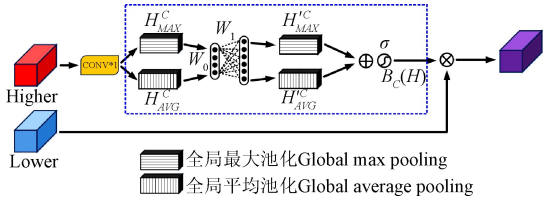
其核心操作如图 3 中虚线框部分所示，其计算为式（1）～式（3）。

$$B_C(H)=\sigma\left(H'^{C}_{AVG}+H'^{C}_{MAX}\right) \tag{1}$$

$$H'^{C}_{AVG}=W_1^{\mathrm{T}}\cdot W_0^{\mathrm{T}}\cdot H^{C}_{AVG} \tag{2}$$

$$H'^{C}_{MAX} = W_1^{\mathrm{T}} \cdot W_0^{\mathrm{T}} \cdot H^{C}_{MAX} \quad (3)$$

其中 $B_C(H)$表示最终获取的通道注意力信息，其由高阶特征生成并用于引导低阶特征图进行通道选择，H^{C}_{AVG} 与 H^{C}_{MAX} 分别表示全局平均池化与全局最大池化操作结果，H'^{C}_{AVG} 与 H'^{C}_{MAX} 分别表示经隐藏层处理后的全局平均池化与全局最大池化值，其维度均为 $H^{C}_{AVG}, H^{C}_{MAX}, H'^{C}_{AVG}, H'^{C}_{MAX} \in \mathbb{R}^{c\times1\times1}$，其中 c 表示低阶特征图通道数量，σ表示 sigmoid 激活函数，用于将通道权重值映射到 0～1 之间，W_0 与 W_1 分别表示隐藏层参数矩阵，全局平均与最大池化操作共享该部分权重值，W_0^{T} 与 W_1^{T} 分别为 W_0 与 W_1 的转置，其维度分别为 $W_0 \in \mathbb{R}^{c\times\frac{c}{s}}$，$W_1 \in \mathbb{R}^{\frac{c}{s}\times c}$，其中 s 表示放缩因子，实际应用中通常选取为 8，$\bullet$ 表示矩阵相乘操作。



注：Higher 表示高阶特征，Lower 表示低阶特征。H^{C}_{AVG} 与 H^{C}_{MAX} 分别表示全局平均池化与全局最大池化，H'^{C}_{AVG} 与 H'^{C}_{MAX} 分别表示经隐藏层处理后的全局平均池化与全局最大池化值，W_0 与 W_1 分别表示隐藏层参数矩阵，σ表示 sigmoid 激活函数，$B_C(H)$表示最终获取的通道注意力信息。

Note: Higher represents high-order features, Lower represents low-order features, H^{C}_{AVG} and H^{C}_{MAX} represent global average pooling and global maximum pooling, respectively, and H'^{C}_{AVG} and H'^{C}_{MAX} represent global average pooling and global maximum pooling values after hidden layer processing, respectively, W_0 and W_1 respectively represent the hidden layer parameter matrix, σ represents the sigmoid activation function, and $B_C(H)$ represents the finally obtained channel attention information.

图 3 通道注意力模块

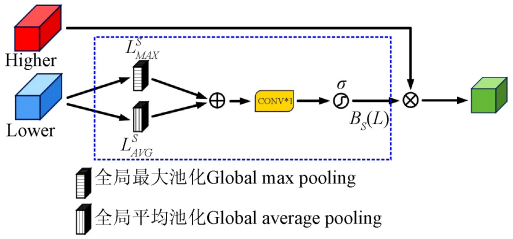
Fig.3 Channel attention block

2.2.2 引入空间注意力模块

为有效编码特征图内部像素点间关系，本文引入空间注意力（spatial attention block，SAB）模块以对特征图内部元素进行建模，不同通道特征图中相同位置处具有相同的空间权重信息。不同于通道注意力对特征图中的每一通道内部的所有特征点共享相同权重，空间注意力区别对待于特征图中的每一点，将所有通道中相同位置处的值进行全局平均池化与最大池化融合操作以获取指定位置处的空间权重值，用以补充通道注意力机制无法较好获取的位置关系信息，进而用于对特征图中各个位置特征值进行筛选以突出适用于生猪脸部姿态检测的特征，其核心操作如图 4 中虚线框所示，其计算为式（4）。

$$B_S(L) = \sigma\left(\mathrm{CONV}\left(L^{S}_{AVG} + L^{S}_{MAX}\right)\right) \quad (4)$$

其中 $B_S(L)$表示最终生成的空间注意力信息，其由低阶特征生成并用于引导高阶特征图进行空间信息筛选，L^{S}_{AVG} 与 L^{S}_{MAX} 分别表示全局平均池化与全局最大池化操作结果，其维度均为 $L^{S}_{AVG}, L^{S}_{MAX} \in \mathbb{R}^{1\times h\times w}$，$h$ 和 w 分别表示高阶特征图的高度与宽度，σ表示 sigmoid 激活函数，CONV 表示卷积核大小为 7×7，卷积核个数为 1 的卷积操作。



注：L^{S}_{AVG} 与 L^{S}_{MAX} 分别表示全局平均池化与全局最大池化操作，$B_S(L)$表示最终生成的空间注意力信息

Note: L^{S}_{AVG} and L^{S}_{MAX} respectively represent the global average pooling and global maximum pooling operations, and $B_S(L)$ represents the resulting spatial attention information.

图 4 空间注意力模块

Fig.4 Spatial attention block

2.2.3 融合通道与空间注意力的 DAT-YOLO 模型

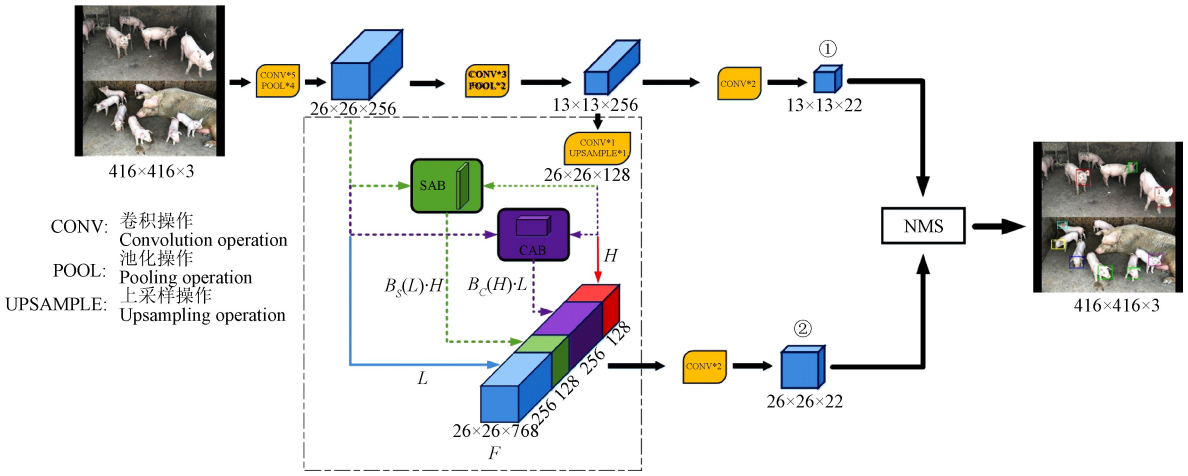
本文提出融合 CAB 与 SAB 模块的 DAT-YOLO（dual attention tiny-YOLO）模型，对 Tiny-YOLO 模型进行改进，用于群养环境下多角度生猪脸部姿态检测，其模型结构如图 5 所示。DAT-YOLO 在 Tiny-YOLO 模型中引入通道注意力块与空间注意力块两类模块以选择性融合深浅层特征，高阶特征引导低阶特征进行通道注意力获取，低阶特征反向指引高阶特征进行空间注意力筛选，可在不显著增加计算量和参数量的前提下提升模型特征提取能力。因群养状态下生猪个体距摄像仪位置不同，故其所采集的生猪脸部面积差异较大，DAT-YOLO 保留了 Tiny-YOLO 的多尺度特征提取模块以保证对不同大小脸部有较强的检测性能。DAT-YOLO 模型核心部件如图 5 中最外层虚框所示，其计算为式（5）。

$$F = H + L + B_C(H)\cdot L + B_S(L)\cdot H \quad (5)$$

其中 H 表示高阶特征，L 表示低阶特征，$B_C(H)$表示由高阶特征生成的通道注意力信息，其用于引导低阶特征进行通道选择，维度为 $B_C(H) \in \mathbb{R}^{c\times1\times1}$，$B_S(L)$表示由低阶特征生成的空间注意力信息，其维度为 $B_S(L) \in \mathbb{R}^{1\times h\times w}$，$F$ 表示经过特征融合后的特征图，其维度大小为 $F \in \mathbb{R}^{c'\times h\times w}$，其通道数量 c'大小是高阶、低阶、CAB 以及 SAB 模块通道数之和，融合后的特征图作为尺度②特征图的输入。

2.3 模型评价指标

本文采用目标检测领域公认的平均检测精度 mAP 以及精确率-召回率（precision-recall，*P-R*）曲线变化情况作为评价标准以衡量 4 种模型对生猪脸部姿态检测性能。*P-R* 曲线反映的是不同召回率与对应召回率下最大精确率间的关系变化情况，检测精度 AP 指 *P-R* 曲线下方面积，mAP 指同一模型对 6 种生猪脸部姿态类别的 AP 平均值。Precision、Recall、AP 及 mAP 定义如式（6）～式（9）所示。



注：$B_C(H)\cdot L$ 表示经过高阶特征通道筛选后的低阶特征信息，$B_S(L)\cdot H$ 表示经过低阶特征空间筛选后的高阶特征信息

Note: $B_C(H)\cdot L$ indicates low-order feature information after high-order feature channel filtering, and $B_S(L)\cdot H$ indicates high-order feature information after low-order feature space filtering.

图 5　DAT-YOLO 模型结构图

Fig.5　Model structure of DAT-YOLO

$$\text{Precision}=\frac{\text{TP}}{\text{TP}+\text{FP}} \tag{6}$$

$$\text{Recall}=\frac{\text{TP}}{\text{TP}+\text{FN}} \tag{7}$$

$$\text{AP}=\int_0^1 PR\mathrm{d}r \tag{8}$$

$$\text{mAP}=\frac{1}{C}\sum_{C_i\in C}\text{AP}_{(C_i)} \tag{9}$$

其中 TP 表示模型预测为正实际为正的样本数量；FP 表示模型预测为正实际为负的样本数量；FN 表示模型预测为负实际为正的样本数量；P，R 分别表示精确率与召回率；C 表示姿态类别总数，本文取 6，C_i 表示当前第 i 个类别，i 的取值范围为 1～6。

3　试验平台参数

试验平台配置为 Intel(R) Core(TM)i7-6700CPU@3.40 GHz 处理器，8 GB 运行内存，1 T 硬盘容量，12 GB GTX Titan X GPU，系统为 CentOS7.4。采用 keras[33]框架进行模型代码的编写。将数据集分为训练集、验证集及测试集 3 个部分，其中训练集大小为 504，验证集大小为 65，测试集大小为 420。为避免内存溢出，采取批训练方式对 Tiny-YOLO 与 YOLOV3 两类系列 8 个子模型在训练集和验证集上进行试验，训练时一个批次（batch）包含 16 张图片，遍历 1 次全部训练集数据称为 1 次迭代，本文设置迭代次数为 300。8 个子模型均采用与 Redmon[18]一致的 loss 损失函数，采用自适应矩阵估计算法（adaptive moment estimation，Adam）[34]优化模型，初始学习率设置为 0.000 1，每次更新权值时使用 BN（batch normalization）[35]进行正则化。为使模型能检测不同大小的生猪脸部，引入 Faster R-CNN 的锚框（anchor boxes）[15]思想，通过使用 K-means 算法对训练集锚框进行聚类，针对图5 中 Tiny-YOLO 系列模型①与②每种尺度分别生成 2 种不同大小的潜在待识别目标的锚框，最终获得 4 个锚点，其大小分别为（23×27）（37×58）（81×82）（135×169），其中尺度① 使用后 2 个锚点，适合检测较大脸部对象，尺度② 使用前 2 个锚点，适合检测较小对象，针对 YOLOV3 系列模型，共有 9 个锚点框，其大小分别为（10×13）（16×30）（33×23）（30×61）（62×45）（59×119）（116×90）（156×198）（373×326），其中前 3 个锚点框适合检测较小脸部对象，中间 3 个锚点框适合检测中等大小脸部框，后 3 个锚点框适合检测较大脸部对象。在计算 mAP 指标值时，参照 PASCAL VOC2012 mAP 评价指标[36]定义方式，设置当检测框与手动标注框的 IOU[37]值超过 0.5 且检测类别相同时视为检测正确。

4　结果与分析

本文移除图 5 中 CAB 模块构建 SAT-YOLO（spatial attention tiny-YOLO）子模型、移除 SAB 模块构建 CAT-YOLO（channel attention tiny-YOLO）子模型以测试 2 类注意力模块在 Tiny-YOLO 系列模型上的有效性。同时为进一步验证注意力信息在其余模型上的迁移性能，基于 YOLOV3 模型分别构建 CA-YOLOV3（channel attention YOLOV3）、SA-YOLOV3（spatial attention YOLOV3）与 DA-YOLOV3（dual attention YOLOV3）子模型，并评估其各自性能状况。

4.1　模型的检测精度分析

表 3 为 Tiny-YOLO 模型 8 种子模型对生猪 6 种脸部姿态识别的 AP 以及总体 mAP，不同模型间采用相同的试验训练参数，并在相同测试集上验证模型的有效性，其中每一系列模型均进行完全不引入注意力、仅引入通道注意力、仅引入空间注意力与同时引入 2 种注意力各 4 组试验，以验证注意力机制有效性。

试验结果表明：

1）基于 Tiny-YOLO 的子模型在水平侧脸、低头正脸、抬头侧脸姿态类别的 AP 值均优于基于 YOLOV3 的子模型，且对低头正脸类别的提升幅度最大，由表 2 可知，在对低头正脸类别预测的 AP 值方面，Tiny-YOLO 系列模型相较于 YOLOV3 系列模型提高了 2.20%～8.59%。Tiny-YOLO 系列模型虽对水平正脸和低头侧脸姿态类别未能取得最佳 AP 值，但其值与 YOLOV3 系列加入相同注意力模块的子模型相比仍具有较强竞争力。Tiny-YOLO 系列模型的总体 mAP 指标与加入相同注意力模块的 YOLOV3 系列模型相比提高了 0.46%～1.92%，且对同时加入 2 类注意力模块的子模型，DA-YOLOV3 模型预测 mAP 值达到 81.92%，DAT-YOLO 模型预测 mAP 值达到 82.38%，取得了相应系列模型最优值，而对于单独引入通道或者空间注意力的 mAP 指标，

SA-YOLOV3 预测 mAP 值达到 78.30%，CA-YOLOV3 预测 mAP 值达到 75.80%，SA-YOLOV3 预测效果优于 SA-YOLOV3 预测效果，SAT-YOLO 模型预测 mAP 值达到 79.43%，CAT-YOLO 模型预测 mAP 值达到 77.72%，SAT-YOLO 预测效果优于 CAT-YOLO 预测效果，这表明注意力信息尤其是空间注意力对 YOLOV3 和 Tiny-YOLO 系列模型性能影响更大。虽然 YOLOV3 网络层次更深，但在本文试验数据环境中，其性能劣于 Tiny-YOLO 系列模型，理论上网络层次越深，其特征表征能力越强，但深层网络可能会带来梯度消失问题，导致反向传播算法无法将梯度有效传回。此外，本文群养生猪图片具有背景可控、猪只所占图片像素比例较大的特点，对场景语义丰富度较弱的图片，浅层网络往往具有更优的检测性能。基于上述试验结果，本文后续其他指标分析主要集中于性能较优的 Tiny-YOLO 系列模型。

2）Tiny-YOLO 系列模型的 DAT-YOLO 子模型在多数类别上均能取得最佳 AP 值。与同系列模型其余 3 种子模型相比，DAT-YOLO 模型的 mAP 值提高 2.95%～8.39%。DAT-YOLO 对抬头侧脸类别预测的 AP 值比其余 3 种子模型最优值提高 6.59%，提升幅度最大。虽然 DAT-YOLO 对低头侧脸类别未能达到最优效果，但仅比其余 3 种子模型的最佳值低 1.31%，说明 DAT-YOLO 子模型最适用于群养状态下生猪脸部姿态识别。

3）引入注意力机制在很大程度上提升了生猪脸部姿态检测准确率。在 Tiny-YOLO 模型中引入通道注意力或空间注意力后的子模型，SAT-YOLO 比 Tiny-YOLO 模型的 mAP 提高 5.44%，CAT-YOLO 比 Tiny-YOLO 模型的 mAP 提高 3.73%，DAT-YOLO 比 Tiny-YOLO 模型的 mAP 提高 8.39%，其性能均优于 Tiny-YOLO 模型，且 DAT-YOLO 子模型的效果最优，其 mAP 分虽较 CAT-YOLO、SAT-YOLO 提高 4.66%和 2.95%，这是因为通道注意力可对特征图的不同通道赋予不同特征，选择性增大包含生猪脸部通道的权重值，空间注意力对同一特征图不同位置特征点给予不同权重，区别对待特征图内部像素点，强化脸部特征值贡献率，两者结合可总体提升检测准确率，这表明了注意力机制对生猪脸部姿态检测的有效性。对仅引入一种注意力的子模型，SAT-YOLO 提升幅度一般较高于 CAT-YOLO，其 mAP 较 CAT-YOLO 子模型提高 1.71%，具体到单一生猪脸部姿态类别，SAT-YOLO 子模型相较于 CAT-YOLO 子模型对低头侧脸类别的提高值达到 11.40%，提升幅度最大，说明在群养生猪脸部姿态检测中，空间注意力效果更加明显，这是因为不同于通道注意力信息，空间注意力将权重施加于特征图中的每一点，对每个特征点区别对待，自学习脸部框边界权重，可进一步提高脸部边界位置的准确率。

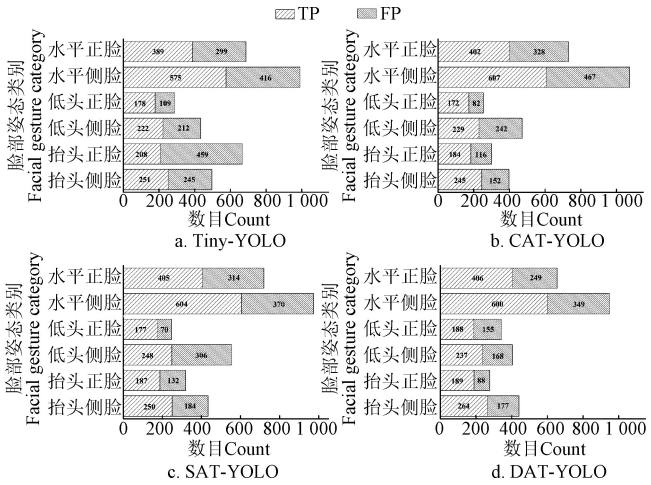
表 3 测试集不同子模型对生猪脸部姿态类别的检测精度与平均检测精度

Table 3 Detection precision(AP) and mean detection precision (mAP) indicator values of test set for different sub models in facial gesture categories of pig

模型 Model	子模型 Sub model	检测精度 Detection precision/%						平均检测精度 Mean detection precision/%
		水平正脸 Horizontal face	水平侧脸 Horizontal side face	低头正脸 Bow down face	低头侧脸 Bow side face	抬头正脸 Looking up face	抬头侧脸 Looking up side	
YOLOV3	YOLOV3	74.45	72.57	77.66	70.71	67.27	74.99	72.96
	CA-YOLOV3	77.73	78.11	75.26	73.71	73.58	76.42	75.80
	SA-YOLOV3	81.68	76.42	78.96	78.69	77.04	77.01	78.30
	DA-YOLOV3	86.58	78.27	87.41	76.76	79.04	83.46	81.92
Tiny-YOLO	Tiny-YOLO	78.63	72.33	85.42	66.03	66.56	74.96	73.99
	CAT-YOLO	82.60	78.19	83.85	70.53	73.39	77.76	77.72
	SAT-YOLO	83.07	79.20	86.50	77.43	72.82	77.56	79.43
	DAT-YOLO	85.54	79.30	89.61	76.12	79.37	84.35	82.38

4.2 模型预测结果 TP 与 FP 数目分析

为明确 Tiny-YOLO 系列模型的 4 种子模型对 6 种姿态类别的预测结果，对式（6）、式（7）中的 TP、FP 中间指标值采用柱状图表示。通常 TP、FP 值的获取，需对模型预测的结果类别框进行过滤操作，首先去除置信度低于某一值的预测框（本文置信度阈值设置为 0.3），接着将筛选过后的预测框按照置信度值进行降序排列，随后计算最高置信度值预测框与真实框间的 IOU 值，若 IOU 大于设定阈值（本文 IOU 阈值设为 0.5），则将当前预测框加入 TP 中，同时将对应真实框标注为已检测，后续对该真实框的其余预测框均被列入 FP 中，最终 Tiny-YOLO 系列模型 4 种子模型的 TP 与 FP 值如图 6 所示。



注：TP 表示模型预测为正实际为正的样本数量；FP 表示模型预测为正实际为负的样本数量。

Note: TP represents the number of samples that are predicted to be positive and actually positive, FP represents the number of samples that are predicted to be positive and actually negative.

图 6　Tiny-YOLO 系列模型 4 种子模型对生猪 6 种脸部姿态类别预测的 FP 与 TP 值

Fig.6　FP and TP values predicted by four sub models of Tiny-YOLO series models for six facial gestures categories of pigs

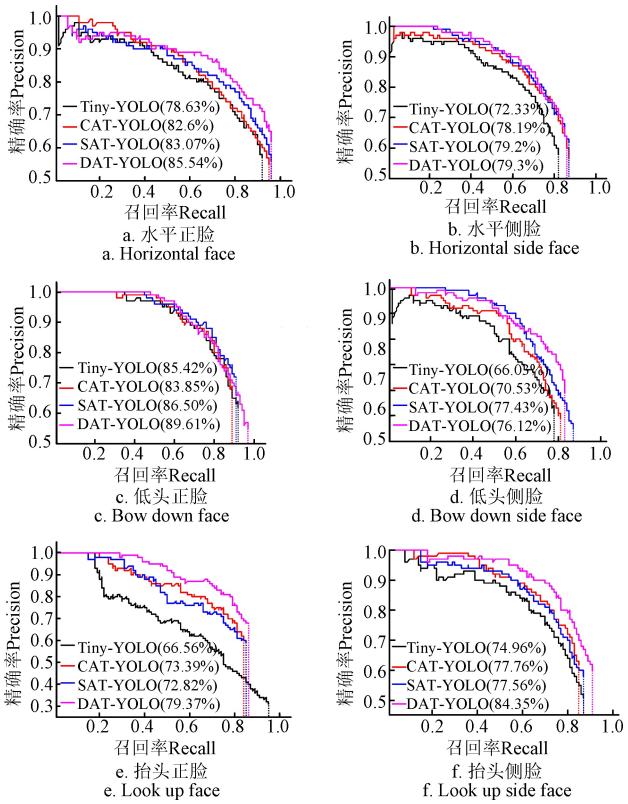
试验结果表明：

1）对于 TP 数目而言，其值越高，模型越优。由图 6 可见加入注意力机制的 3 种子模型其值相差不是很大，但 3 种子模型预测 TP 指标值均高于未加入注意力信息的 Tiny-YOLO 模型。加入注意力信息的 3 种子模型对 6 种脸部姿态类别的 TP 总数目排序依次是 DAT-YOLO> SAT-YOLO>CAT-YOLO，可见同时加入 2 种注意力对生猪脸部姿态检测的预测效果最佳。对于单一姿态类别，加入注意力的 3 种子模型 TP 值呈现出不同特点，对差异最大的低头侧脸和抬头侧脸类别，SAT-YOLO 的低头侧脸 TP 数目比 CAT-YOLO 多 19 个，DAT-YOLO 的抬头侧脸 TP 值比 CAT-YOLO 多 19 个。

2）对于 FP 数值而言，其值越低，模型效果越佳。4 种子模型预测结果值排序依次为 DAT-YOLO< SAT-YOLO <CAT-YOLO<Tiny-YOLO，可见 DAT-YOLO 子模型 FP 数值最低，说明加入注意力信息能有效降低 FP 指标值，进而侧面促进预测精度的提升。对于单独引入一种注意力的情形，SAT-YOLO 子模型一般优于 CAT-YOLO 子模型，在部分姿态类别上 SAT-YOLO 略逊于 CAT-YOLO 子模型，但其均能产生有竞争力的 FP 指标值。具体到生猪脸部单一姿态类别，DAT-YOLO 子模型对抬头正脸类别的 FP 值不足 Tiny-YOLO 子模型的五分之一，提升幅度最大。虽然引入注意力的 3 种子模型在 TP 值上未有明显差异，但 FP 值上的较大差异最终使得 DAT-YOLO 子模型优势更加明显。

4.3　模型精确率-召回率曲线分析

图 7 为 Tiny-YOLO 系列模型 4 种子模型对生猪 6 种脸部姿态识别的 *P-R* 曲线图。由式（6）、式（7）可知，Precision 与 Recall 值的获取需要首先计算 TP、FP 以及 FN 的个数，每种子模型对不同种类的预测结果 TP、FP 值可见图 6 所示。对于 FN 数值，图像中某一类别的真实框数量是已知的，TP 数量可从 4.1 部分求出，两者差值即为 FN 结果值。每个类别中 *P-R* 曲线下方面积即为表 3 中对应模型在该类别上的 AP 值，曲线越靠近右上角，模型效果越优。



注：括号中数据表示对应模型对当前类别的 AP 值，其值为对应模型在当前类别 *P-R* 曲线下方面积。

Note: The data in the brackets indicates the AP value of the current model for the corresponding model, and the value is the area under the current model *P-R* curve of the corresponding model.

图 7　Tiny-YOLO 系列模型 4 种子模型对生猪 6 种脸部姿态识别的 *P-R* 曲线

Fig.7　*P-R* curves of four sub models of Tiny-YOLO series models for six facial gestures categories of pigs

从图 7 中可看出，加入注意力机制后的 3 种子模型 *P-R* 曲线均位于未加注意力信息的 Tiny-YOLO 模型上方，这是因为 Tiny-YOLO 检测模型所提取的卷积特征并未对卷积核中不同位置处特征进行区别对待，认为每个区域对检测结果贡献度相同，但实际中，待检测物体周围往往具有复杂且丰富的语境信息，对目标区域特征施以权重可使模型更好地定位至待检测特征上，说明注意力信息有利于对群养生猪脸部姿态进行检测。对于单独引入 1 种注意力机制的情况，SAT-YOLO 子模型在除抬头正脸类别外的其余类别上其曲线均位于 CAT-YOLO 子模型上方，说明与通道注意力相比，空间注意力对区域检测效果更为明显。此外，4 种子模型对低头正脸、抬头正脸、抬头侧脸 3 个姿态类别的 Recall 值小于 0.1 时，Precision 值均维持在 1.0 附近，差异不大，但随着 Recall 值的减小，

DAT-YOLO 子模型优势逐渐明显，其对应 Precision 值均不逊于其余 3 种子模型，并在抬头正脸、抬头侧脸 2 个类别上优势最为明显。虽然 DAT-YOLO 子模型对水平正脸、水平侧脸、低头侧脸类别的表现并不突出，但其对应类别曲线状态均不弱于其余子模型，说明在同时加入 2 种注意力的情况下可最大限度提升模型性能，能充分融合通道注意力和空间注意力优势。

4.4 模型预测结果分析

为进一步展示模型预测效果，在测试集上对 Tiny-YOLO 系列模型 4 种子模型进行预测，选取其中 3 幅可视化结果如图 8 所示。其中每行所表示子模型类别见图左侧注释所示，预测结果中预测框左上角部分显示值表示预测为当前姿态类别的置信度。

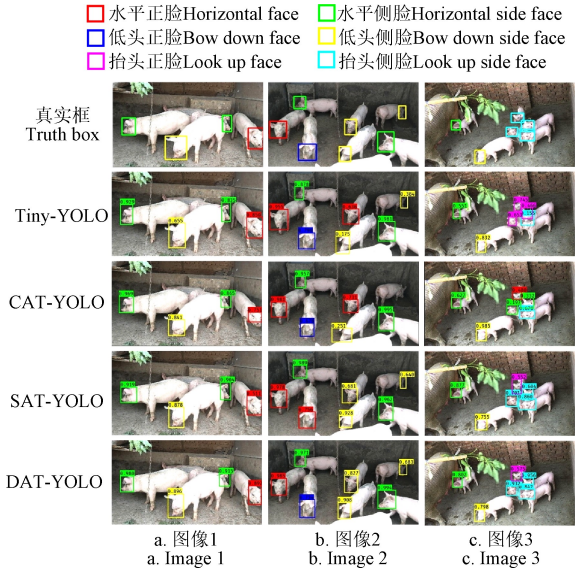


图 8 部分测试集预测结果

Fig.8 Partial test set prediction results

由图 8 可知，对于 8a 对应图像，4 种子模型均能正确预测出生猪脸部姿态类别及其位置，其中 DAT-YOLO 子模型能够取得最佳的类别置信度。对于 8b 对应的图像，Tiny-YOLO、SAT-YOLO 和 DAT-YOLO3 种子模型均能对生猪脸部较小或者不明显区域做出正确预测，且 DAT-YOLO 预测置信度最高，且除 DAT-YOLO 子模型外，其余 3 种子模型在预测过程中均有预测错误的情形出现，表明 DAT-YOLO 子模型在小目标脸部姿态检测的可用性。8c 对应图像，小猪较为密集黏连，4 种子模型虽然均有预测错误的情况出现，但对于预测正确的情况，DAT-YOLO 子模型预测结果最接近真实框。

5 讨 论

5.1 数据不均衡问题处理

本研究中 6 种脸部姿态的数量差异会产生数据不均衡问题，理论上可从数据和损失函数 2 种角度解决该问题[38-39]。结合文中模型的适用场景，本文采用损失函数对数据进行伪均衡化处理，采用与 Redmon 等[18]一致的损失函数，其形式化表示如公式（10）所示。

$$\begin{aligned}\text{loss}=&\lambda_{coord}\sum_{i=0}^{N\times N}\sum_{j=0}^{K}1_{ij}^{obj}\left[\left(t_x-t_x'\right)^2+\left(t_y-t_y'\right)^2\right]+\\&\lambda_{coord}\sum_{i=0}^{N\times N}\sum_{j=0}^{K}1_{ij}^{obj}\left[\left(t_w-t_w'\right)^2+\left(t_h-t_h'\right)^2\right]-\\&\sum_{i=0}^{N\times N}\sum_{j=0}^{K}1_{ij}^{obj}\left[c_i'\log(c_i)+(1-c_i')\log(1-c_i)\right]-\\&\lambda_{noobj}\sum_{i=0}^{N\times N}\sum_{j=0}^{K}1_{ij}^{noobj}\left[c_i'\log(c_i)+(1-c_i')\log(1-c_i)\right]-\\&\sum_{i=0}^{N\times N}1_i^{obj}\sum_{c\in classes}\left[p_i'(c)\log(p_i(c))+(1-p_i'(c))\log(1-p_i(c))\right]\end{aligned}\tag{10}$$

式（10）由 3 部分构成，第 1 行表示边框位置误差，第 2～3 行表示置信度误差，第 4 行表示对象分类误差。置信度误差和对象分类误差均采用交叉熵方式计算得到，式中 $N\times N$ 表示最终输出特征图尺寸，由图 5 可知，本研究输出 2 个尺度特征图，N 取值为 13 和 26，特征图中的行像素点与列像素点交错将其划分为多个网格区域，K 表示每个网格所包含的边框数目，本文采用与 Tiny-YOLO 相同设置，K 取值为 2，1_{ij}^{obj} 表示第 i 个网格的第 j 个边框内是否存在目标对象，存在则取值为 1，否则取值为 0，1_{ij}^{noobj} 表示第 i 个网格的第 j 个边框内是否不存在对象，不存在则取值为 1，否则取值为 0，1_i^{obj} 表示第 i 个网格中是否存在对象，只有在第 i 个网格中存在对象的前提下才会计算其对象类别误差。λ_{noobj} 用于协调网格中存在对象与不存在对象间损失函数值的比重，该值可调整网格存在对象和不存在对象 2 种情况对最终损失函数的影响程度，参照 Tiny-YOLO 模型，λ_{noobj} 取值为 0.5，即调低不存在对象的边框置信度误差权重。λ_{coord} 用于调整位置误差的权重，本文将其设置为 5（与 Tiny-YOLO 模型保持一致），以提高位置误差在整个损失函数中的比重，从而提升位置边框检测精度。采用公式（10）可以解决以下 2 种不均衡问题。

1）λ_{noobj} 可用于解决包含脸部区域和不包含脸部区域网格数量差异问题。不包含脸部对象的网格与包含脸部对象的网格数目相差悬殊，通过设置较小的 λ_{noobj} 值可使模型更关注包含脸部目标的网格，减少不包含脸部目标网格对损失函数的贡献度。

2）1_{ij}^{obj} 可用于解决网格中已包含脸部区域情况下脸部姿态类别数量差异问题。只有将其设置为 1 的对应网格的特定边框才会计入误差。通常在以下两种情况下会将其设置为 0 以忽略当前网格对检测损失函数值的影响，一种情况是当前网格对所有脸部姿态类别的预测置信度均小于阈值（本文与 Tiny-YOLO 一致选取为 0.3），另一种情况是即使置信度超过指定阈值，但预测边框与真实边框的 IOU 未超过 IOU 阈值（本文参考 Tiny-YOLO 选取为 0.5）。置信度可通过 sigmoid 激活函数得到，该激活函数在类间不产生竞争，不受脸部姿态类别数量不同的影响，即使对同一网格会出现多种姿态类别预测置

信度均高于阈值的情况，在进行 IOU 筛选过程中也会通过 IOU 阈值筛选出最合适的脸部姿态类别。

5.2　不同姿态检测精度差异研究

由表 3 可知，模型对不同类别姿态检测精度差异较大，该问题属于深度学习可解释性研究的最前沿领域，最新研究主要集中于对卷积核与图像特征点部位间关系的探讨[40-42]。模型的不同层次可用于提取不同特征，每层卷积核与图像的特定特征相关联，并具有一定的通用性，浅层卷积核可提取纹理、边缘等特征，深层卷积核可提取抽象特征，暂无文献研究表明提取的特定特征与最终检测精度间的定量关系，对于某些脸部姿态类别，在特征提取过程中，模型所关注区域不同，暂时无法从理论上解释模型各层卷积核学习的特征对特定脸部姿态检测类别的贡献程度。

5.3　注意力特征图数量取值原则

图 5 中 DAT-YOLO 模型除注意力模块外的其余特征图数目选取均与图 2 中 Tiny-YOLO 模型均保持一致，对通道与空间注意力特征图数量的选取原则是基于经验的直观准则及约定俗成的默认设置（取 2 的整数次幂），目前尚无严密且令人信服的数学解释，对该部分探索性的研究集中于神经网络架构搜索（neural architecture search, NAS）[43]，其可用于自动搜索最优卷积核大小及特征图数量。本研究更关注于不同注意力信息对检测效果的影响程度，故未将其列入研究范畴。

6　结　论

本文在 Tiny-YOLO 模型中引入通道注意力和空间注意力，对 Tiny-YOLO 模型进行了改进，建立检测模型 DAT-YOLO，用于群养场景下生猪不同脸部姿态检测，主要结论如下：

1）与 YOLOV3 系列模型相比，Tiny-YOLO 系列模型具有更强检测性能，且在加入注意力信息后，2 类系列模型检测精度均有不同程度提升。

2）Tiny-YOLO 系列模型中，引入注意力机制的 CAT-YOLO、SAT-YOLO 和 DAT-YOLO 3 种子模型的 mAP 值相较于未引入注意力机制的 Tiny-YOLO 模型分别提高了 3.73%、5.44%和 8.39%，表明注意力机制对生猪脸部姿态检测的有效性，可很大程度提升通用卷积网络的特征提取能力。

3）SAT-YOLO 检测效果总体优于 CAT-YOLO。其在低头侧脸类别上优势最为明显，相较于 CAT-YOLO 子模型，其检测精度提高 6.90%。表明空间注意力信息更适用于生猪脸部姿态检测。

4）同时引入 2 种注意力的 DAT-YOLO 子模型无论在各个类别的检测精度、所有类别的平均检测精度、FP/TP 指标值以及 *P-R* 曲线中，效果均优于 CAT-YOLO 和 SAT-YOLO 模型，表明同时引入 2 种注意力信息对生猪脸部姿态检测效果更佳，可为生猪脸部姿态检测提供方法和思路，为群养生猪个体识别提供有益参考。

［参　考　文　献］

[1] 孙龙清，李玥，邹远炳，等. 基于改进 Graph Cut 算法的生猪图像分割方法[J]. 农业工程学报，2017，33(16)：196－202.
Sun Longqing, Li Yue, Zou Yuanbing, et al. Pig image segmentation method based on improved Graph Cut algorithm[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2017, 33(16): 196－202. (in Chinese with English abstract)

[2] 邹远炳，孙龙清，李玥，等. 基于分布式流式计算的生猪养殖视频监测分析系统[J]. 农业机械学报，2017，48(S)：365－373.
Zou Yuanbing, Sun Longqing, Li Yue, et al. Video monitoring and analysis system for pig breeding based on distributed flow Computing[J]. Transactions of the Chinese Society for Agricultural Machinery, 2017, 48(S): 365－373. (in Chinese with English abstract)

[3] 薛月菊，朱勋沐，郑婵，等. 基于改进 Faster R-CNN 识别深度视频图像哺乳母猪姿态[J]. 农业工程学报，2018，34(9)：189－196.
Xue Yueju, Zhu Xunmu, Zheng Chan, et al. Lactating sow postures recognition from depth image of videos based on improved Faster R-CNN[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2018, 34(9): 189－196. (in Chinese with English abstract)

[4] 胡志伟，杨华，娄甜田，等. 基于全卷积网络的生猪轮廓提取[J]. 华南农业大学学报，2018，39(6)：111－119.
Hu Zhiwei, Yang Hua, Lou Tiantian, et al. Extraction of pig contour based on fully convolutional networks[J]. Journal of South China Agricultural University, 2018, 39(6): 111－119. (in Chinese with English abstract)

[5] 杨阿庆，薛月菊，黄华盛，等. 基于全卷积网络的哺乳母猪图像分割[J]. 农业工程学报，2017，33(23)：219－225.
Yang Aqing, Xue Yueju, Huang Huasheng, et al. Lactating sow image segmentation based on fully convolutional networks[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2017, 33(23): 219－225. (in Chinese with English abstract)

[6] Yang Aqing, Huang Huasheng, Zheng Chan, et al. High-accuracy image segmentation for lactating sows using a fully convolutional network[J]. Biosystems Engineering, 2018, 176: 36－47.

[7] Psota E, Mittek M, Pérez L, et al. Multi-pig part detection and association with a fully-convolutional network[J]. Sensors, 2019, 19(4): 852.

[8] Wang Jianzong, Liu Aozhi, Xiao Jing. Video-Based Pig Recognition with Feature-Integrated Transfer Learning[C]// Biometric Recognition, 2018: 620－631.

[9] Chen Zuge, Wu Kehe, Li Yuanbo, et al. SSD-MSN: An improved multi-scale object detection network based on SSD[J]. IEEE Access, 2019, 7: 80622－80632.

[10] GhiasiG, Lin T Y, Le Q V. Nas-fpn: Learning scalable feature pyramid architecture for object detection[C]// Proceedings of the IEEE Conference on Computer Vision

and Pattern Recognition (CVPR). IEEE Computer Society: Piscataway, NJ. 2019: 7036－7045.

[11] Law H, Deng J. Cornernet: Detecting objects as paired keypoints[C]//Proceedings of the European Conference on Computer Vision (ECCV). Cham:SpringerInternational Publishing, 2018: 734－750.

[12] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). IEEE Computer Society: Piscataway, NJ. 2017: 2117－2125.

[13] Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector[C]//European conference on computer vision(CVPR). IEEE Computer Society: Piscataway, NJ. 2016: 21－37.

[14] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[J]. Computer Science, 2013: 580－587.

[15] Ren S, He K, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, 39(6): 1137－1149.

[16] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR). IEEE Computer Society: Piscataway, NJ. 2016: 779－788.

[17] Redmon J, Farhadi A. YOLO9000: Better, Faster, Stronger[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society: Piscataway, NJ. 2017: 6517－6525.

[18] Redmon J, Farhadi A. Yolov3: An incremental improvement[J/OL]. [2019-07-10]. USA: 2018. https://arxiv.org/abs/1804.02767

[19] Pedoeem J, Huang R. YOLO-LITE: A real-time object fetection algorithm optimized for non-GPU computers[J/OL]. [2019-07-10]. USA: 2018. https://arxiv.org/abs/1811.05588

[20] 薛月菊，黄宁，涂淑琴，等. 未成熟芒果的改进 YOLOv2 识别方法[J]. 农业工程学报，2018，34(7)：173－179.
Xue Yueju, Huang Ning, Tu Shuqin, et al. Immature mango detection based on improved YOLOv2[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2018, 34(7): 173－179. (in Chinese with English abstract)

[21] 赵德安，吴任迪，刘晓洋，等. 基于 YOLO 深度卷积神经网络的复杂背景下机器人采摘苹果定位[J]. 农业工程学报，2019，35(3)：164－173.
Zhao Dean, Wu Rendi, Liu Xiaoyang, et al. Apple positioning based on YOLO deep convolutional neural network for picking robot in complex background[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2019, 35(3): 164－173. (in Chinese with English abstract)

[22] Ju M, Choi Y, Seo J, et al. A kinect-based segmentation of touching-pigs for real-time monitoring[J]. Sensors, 2018, 18(6): 1746.

[23] Wang Fei, Jiang Mengqing, Qian Chen, et al. Residual Attention Network for Image Classification[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society: Piscataway, NJ. 2017: 3156－3164.

[24] Yu Changqian, Wang Jingbo, Peng Chao, et al. BiSeNet: Bilateral segmentation network for real-time semantic segmentation[C]//Proceedings of the European Conference on Computer Vision (ECCV). Cham:SpringerInternational Publishing, 2018: 325－341.

[25] 徐诚极，王晓峰，杨亚东. Attention-YOLO：引入注意力机制的 YOLO 检测算法.计算机工程与应用[J]，2019，55(6)：13－23.
Xu Chengji, Wang Xiaofeng, Yang Yadong. Attention-YOLO: YOLO detection algorithm that introduces attention mechanism[J]. Computer Engineering and Applications, 2019, 55(6): 13－23. (in Chinese with English abstract)

[26] TzuTa L. LabelImg [CP/DK]. (2017-01-09) [2019-06-20] https:// github.com/tzutalin/labelImg

[27] Raykar V C, Saha A. Data Split Strategies for Evolving Predictive Models[C]//Machine Learning and Knowledge Discovery in Databases. 2015: 3－19.

[28] Long Jonathan, Shelhamer Evan, Darrell Trevor. Fully convolutional networks for semantic segmentation[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 39(4): 640－651.

[29] 刘军，后士浩，张凯，等. 基于增强 Tiny YOLOV3 算法的车辆实时检测与跟踪[J]. 农业工程学报，2019，35(8)：118－125.
Liu Jun, Hou Shihao, Zhang Kai, et al. Real-time vehicle detection and tracking based on enhanced Tiny YOLOV3 algorithm[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2019, 35(8): 118－125. (in Chinese with English abstract)

[30] Neubeck A, Van Gool L. Efficient non-maximum suppression[C]//18th International Conference on Pattern Recognition (ICPR). Springer: Berlin, German. 2006, 3: 850－855.

[31] Lin Min, Chen Qiang, Yan Shuicheng. Network in network[J/OL]. [2019-07-20]. USA: 2014. https://arxiv.org/abs/1312.4400

[32] Zhou Bolei, Khosla A, Lapedriza A, et al. Learning deep features for discriminative localization[C]//In: Computer Vision and Pattern Recognition (CVPR), IEEE Computer Society: Piscataway, NJ. 2016: 2921－2929.

[33] Chollet F. Keras[CP/DK]. (2015-03-28)[2019-07-05]. https://github.com/keras-team/keras/

[34] Kingma D P, Ba J. Adam: A method for stochastic optimization[J/OL]. [2019-07-20]. https://arxiv.org/ abs/1412.6980

[35] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift[C]// International Conference on Machine Learning(ICML). 2015: 448－456.

[36] Microsoft. PASCAL-VOC2012 [DB/OL]. (2012-02-20) [2019-08-02]. http://host.robots.ox.ac.uk/pascal/VOC/voc2012

[37] Rezatofighi H, Tsoi N, Gwak J Y, et al. Generalized

intersection over union: A metric and a loss for bounding box regression[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, IEEE Computer Society: Piscataway, NJ. 2019: 658－666.

[38] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]//Proceedings of the IEEE International Conference on Computer Vision. 2017: 2980－2988.

[39] Li B, Liu Y, Wang X. Gradient harmonized single-stage detector[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2019, 33: 8577－8584.

[40] Zhang Quanshi, Wu Yingnian, Zhu Songchun. Interpretable Convolutional Neural Networks[C]// The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018: 8827－8836.

[41] Zhang Quanshi, Yang Yu, Ma Haotian, et al. Interpreting CNNs via Decision Trees[C]// The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019: 6261－6270.

[42] Bolei Zhou, David Bau, Aude Oliva, et al. Interpreting deep visual representations via network dissection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 41(9): 2131－2145.

[43] Jin H, Song Q, Hu X. Auto-keras: An efficient neural architecture search system[C]//Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, ACM, 2019: 1946－1956.

Detection of facial gestures of group pigs based on improved Tiny-YOLO

Yan Hongwen[1], Liu Zhenyu[1], Cui Qingliang[2※], Hu Zhiwei[1], Li Yanwen[1]

(1. *College of Information Science and Engineering*, *Shanxi Agricultural University*, *Taigu* 030801, *China*;
2. *College of Engineering*, *Shanxi Agricultural University*, *Taigu* 030801, *China*)

Abstract: The face of the pig contains rich biometric information, and the detection of the facial gestures can provide a basis for the individual identification and behavior analysis of the pig. Detection of facial posture can provide basis for individual recognition and behavioral analysis of pigs. However, under the scene of group pigs breeding, there always have many factors, such as pig house lighting and pig adhesion, which brings great challenges to the detection of pig face. In this paper, we take the group raising pigs in the real breeding scene as the research object, and the video frame data is used as the data source. Latter we propose a new detection algorithm named DAT-YOLO which based on the attention mechanism and Tiny-YOLO model, and channel attention and spatial attention information are introduced into the feature extraction process. High-order features guide low-order features for channel attention information acquisition, and low-order features in turn guide high-order features for spatial attention screening, meanwhile the model parameters don't have significant increase, the model feature extraction ability is improved and the detection accuracy is improved. We collect 504 sheets total 3 712 face area picture for the 5 groups of 20 days to 3 and a half months of group health pig video extraction, the number of pigs is 35. In order to obtain the model input data set, we perform a two-step pre-processing operation of filling pixel values and scaling for the captured video. The model outputs are divided into six classes, which are horizontal face, horizontal side-face, bow face, bow side-face, rise face and rise side-face. The results show that for the test set, the detection precision(AP) reaches 85.54%, 79.3%, 89.61%, 76.12%, 79.37%, 84.35% of the horizontal face, horizontal side-face, bow face, bow side-face, rise face and rise side-face respectively, and the mean detection precision(mAP) is 8.39%, 4.66% and 2.95% higher than that of the general Tiny-YOLO model, the CAT-YOLO model only refers to channel attention and the SAT-YOLO model only introduces spatial attention respectively. In order to further verify the migration performance of attention on the remaining models, under the same experimental conditions, two attentional information were introduced to construct the corresponding attention sub-models based on the YOLOV3-based model. The experiment shows that compared to the YOLOV3 submodel, the sub-model based on Tiny-YOLO increase by 0.46% to 1.92% in the mAP. The Tiny-YOLO and YOLOV3 series models have different performance improvements after adding attention information, indicating that the attention mechanism is beneficial to the accurate and effective group gestures detection of different groups of pigs. In this study, the data is pseudo-equalized from the perspective of loss function to avoid the data imbalance caused by the number of poses of different facial categories, and actively explore the reasons for the difference in the accuracy of different facial gesture detection. The study can provide reference for the subsequent individual identification and behavior analysis of pigs.

Keywords: image processing; models; object detection; Tiny-YOLO; channel attention; spatial attention