基于 JDE 模型的群养生猪多目标跟踪

涂淑琴，黄 磊，梁 云※，黄正鑫，李承桀，刘晓龙
（华南农业大学数学与信息学院，广州 510642）

摘 要：为实现群养生猪在不同场景下（白天与黑夜，猪只稀疏与稠密）的猪只个体准确检测与实时跟踪，该研究提出一种联合检测与跟踪（Joint Detection and Embedding，JDE）模型。首先利用特征提取模块对输入视频序列提取不同尺度的图像特征，产生 3 个预测头，预测头通过多任务协同学习输出 3 个分支，分别为分类信息、边界框回归信息和外观信息。3 种信息在数据关联模块进行处理，其中分类信息和边界框回归信息输出检测框的位置，结合外观信息，通过包含卡尔曼滤波和匈牙利算法的数据关联算法输出视频序列。试验结果表明，本文 JDE 模型在公开数据集和自建数据集的总体检测平均精度均值（mean Average Precision，mAP）为 92.9%，多目标跟踪精度（Multiple Object Tracking Accuracy，MOTA）为 83.9%，IDF1 得分为 79.6%，每秒传输帧数（Frames Per Second，FPS）为 73.9 帧/s。在公开数据集中，对比目标检测和跟踪模块分离（Separate Detection and Embedding，SDE）模型，本文 JDE 模型在 MOTA 提升 0.5 个百分点的基础上，FPS 提升 340%，解决了采用 SDE 模型多目标跟踪实时性不足问题。对比 TransTrack 模型，本文 JDE 模型的 MOTA 和 IDF1 分别提升 10.4 个百分点和 6.6 个百分点，FPS 提升 324%。实现养殖环境下的群养生猪多目标实时跟踪，可为大规模生猪养殖的精准管理提供技术支持。

关键词：目标检测；目标跟踪；联合检测与跟踪；数据关联；群养生猪
doi：10.11975/j.issn.1002-6819.2022.17.020
中图分类号：TP391.4 **文献标志码**：A **文章编号**：1002-6819(2022)-17-0186-10

涂淑琴，黄磊，梁云，等. 基于 JDE 模型的群养生猪多目标跟踪[J]. 农业工程学报，2022，38(17)：186-195.
doi：10.11975/j.issn.1002-6819.2022.17.020 http://www.tcsae.org
Tu Shuqin, Huang Lei, Liang Yun, et al. Multiple object tracking of group-housed pigs based on JDE model[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2022, 38(17): 186-195. (in Chinese with English abstract) doi：10.11975/j.issn.1002-6819.2022.17.020 http://www.tcsae.org

0 引 言

生猪产业一直是国内畜牧业的支柱产业，其发展关系到国家食物安全、社会稳定及国民经济的协调发展。生猪养殖业正朝着规模化、专业化、智能化和精细化发展。目前，在劳动力短缺的情况下，智能与精准畜牧业对帮助农户实现畜牧业规模化生产具有重要作用[1]。通过视频摄像头，采用计算机视觉技术获取每头猪每天的体重变化、运动轨迹、饮食情况和行为变化等数据，监测猪只行为和健康，预测猪只个体异常情况，实现生猪生产过程的精确控制[2]，对提高生猪的福利具有重要价值[3]。因此，采用多目标跟踪技术，准确跟踪群养生猪中的个体，识别猪只行为变化，对提高农场的智能化管理水平和生产力具有重要意义。

目前，国内外研究者在禽畜跟踪的方面进行很多研究。有些研究者通过给禽畜穿戴自动跟踪设备实现跟踪禽畜。如 Zambelis 等[4]使用耳标加速计对饲养奶牛的喂养和活动行为进行观察。Giovanetti 等[5]将三轴加速度计传感器安装在羊的身体上，然后测量羊在牧场的行为。Krista 等[6]将运动能耗仪安装在母羊的项圈上，以此评估绵羊行为活动水平。这些方法在某些情况下对于禽畜的观察是可行的，但是，使用可穿戴自动跟踪设备会影响禽畜的行为，严重情况下会影响其自由活动，降低动物福利。另外，大量可穿戴自动跟踪设备会增加生产的成本。

近年来，使用计算机视觉技术进行猪只日常行为监控取得了多方面的研究成果，例如猪的攻击行为[7-10]、饮食饮水行为[11-15]、母猪行为检测[16]、攀爬和玩耍行为[17-18]，猪只姿态识别[11,19-22]，早期发现呼吸道疾病[23-24]。

多目标跟踪的性能在很大程度上取决于其检测目标的性能。传统的目标检测算法，如 Zhao 等[25]使用背景减法来检测移动奶牛目标，Zhang 等[26]提出了一种基于光流估计的运动目标检测方法，于欣等[27]提出一种基于光流法与特征统计的鱼群异常行为检测方法，这些算法在速度和准确性方面不能满足实际场景要求。目前，基于深度学习的目标检测算法不断完善，其准确性和速度都有显著提升，能够满足实际应用。深度学习的目标检测算法主要分为一阶段和二阶段算法。二阶段算法在检测时首先生成候选区域，之后对候选区域进行分类和校准，准确率相对较高，典型的有 R-CNN（Region Convolution Neural Network）算法[28]，Fast R-CNN 算法[29]，Faster R-CNN 算法[30]。如王浩等[31]利用改进的 Faster R-CNN 算

收稿日期：2022-04-19 修订日期：2022-08-16
基金项目：广东省省级科技计划项目（2019A050510034）；广州市重点科技计划项目（202206010091）；大学生创新创业大赛项目（202110564025）
作者简介：涂淑琴，博士，讲师，研究方向为图像处理与计算机视觉。Email：tushuqin@163.com
※通信作者：梁云，博士，教授，研究方向为图像处理与计算机视觉。Email：yliang@scau.edu.cn

法定位群养生猪的圈内位置，识别准确率可达 96.7%。一阶段算法在检测时无需生成候选区域，直接对目标类别和边界进行回归，如 YOLO 系列算法[32-35]。如金耀等[36]利用 YOLOv3 算法[32]对生猪个体进行识别，对母猪的识别精度均值达 95.16%。相较于二阶段算法，一阶段算法的检测速度更快。

在多目标跟踪方面，现有多目标跟踪算法的应用大多是基于检测跟踪（Tracking by Detection，TBD）范式，即 SDE（Separate Detection and Embedding）模型，先用检测器输出检测结果，再用基于卡尔曼滤波和匈牙利算法的后端追踪优化算法进行跟踪，如使用 SORT（Simple Online and Realtime Tracking）[37]、DeepSORT [38]算法来提取目标的表观特征进行多目标重识别进行跟踪，其中 DeepSORT 算法在 SORT 算法的基础上，通过提取深度表观特征提高了多目标的跟踪效果。如张宏鸣等[39]利用改进 YOLOv3 算法结合 DeepSORT 算法进行肉牛多目标跟踪，张伟等[40]利用基于 CenterNet 结合优化 DeepSORT 算法进行断奶仔猪目标跟踪。上述研究的算法是两阶段过程，先检测再跟踪，目标检测和跟踪模块分离导致跟踪速度慢，达不到实时跟踪效果。

本研究将目标检测与跟踪融合在一个过程中，提出一种实时、非接触的群养生猪多目标跟踪 JDE（Joint Detection and Embedding）算法，通过一个端对端网络同时输出多目标的分类信息、边界框回归信息和外观信息，以减少算法的运行时间，达到实时跟踪的效果。在相同的公开试验数据集中将 JDE 算法与 SDE 算法进行对比，以验证本文算法的速度，同时与 TransTrack 算法[41]对比，进一步验证本文算法的准确性与实时性。

1　基于 JDE 的群养生猪多目标跟踪算法

1.1　多目标跟踪算法概述

基于 JDE 的群养生猪多目标跟踪算法如图 1 所示。该算法以群养生猪视频序列为输入；采用特征提取模块提取不同尺度的图像特征，得到 3 个不同尺度特征图的预测头，输入数据关联模块；预测头的分类信息和边界框回归信息用于得到检测框的位置结果，在跟踪部分，利用外观信息结合检测框，通过包含卡尔曼滤波和匈牙利算法的数据关联算法，输出检测与跟踪的视频序列结果。

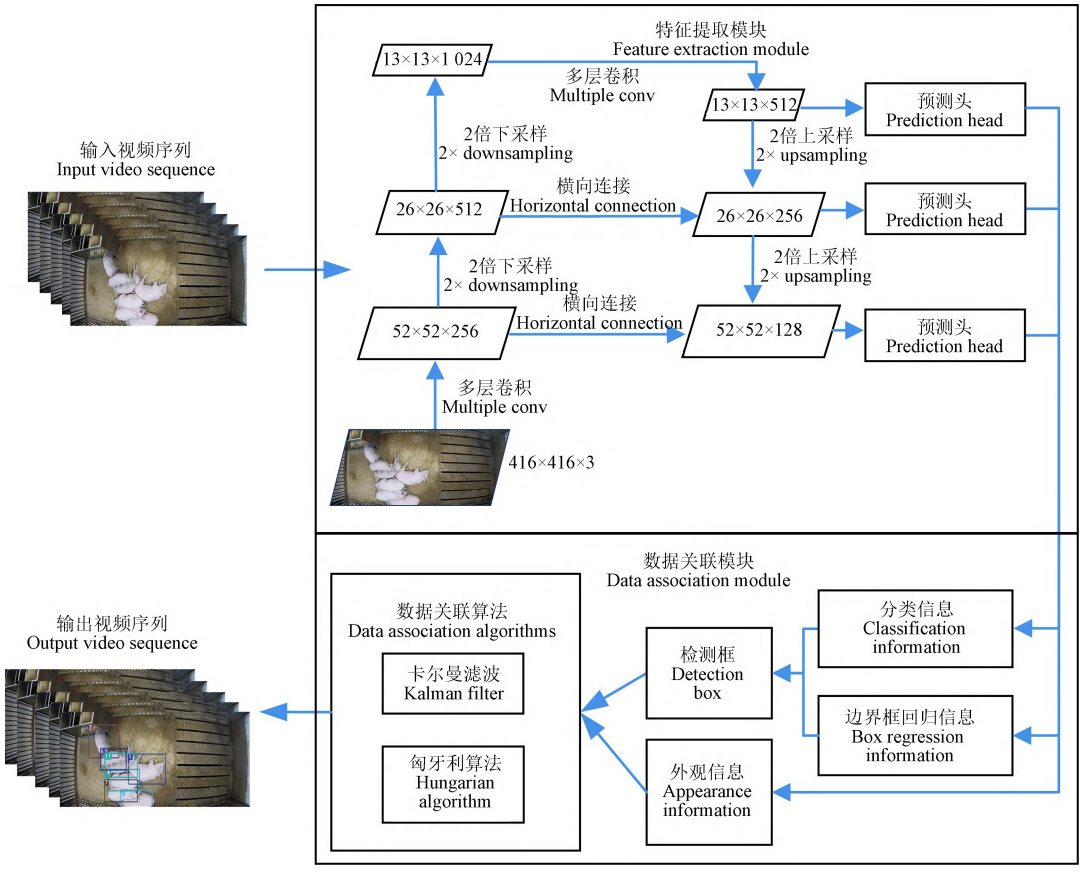


图 1　基于 JDE 的群养生猪多目标跟踪算法

Fig.1　Multiple object tracking algorithm for group-housed pigs based on Joint Detection and Embedding (JDE)

1.2　特征提取模块

特征提取模块由 Darknet-53 网络和多尺度模块特征金字塔构成，如图 2 所示。Darknet-53 网络包括 6 个卷积层和 5 个残差层，其中卷积层和残差层的大小和数量见表 1。卷积层由卷积层、批量归一化层和激活函数层共同构成，残差层由一个 1×1 大小的卷积层和 3×3 大小的卷积层构成。

特征金字塔采用同一图像的不同尺度来检测目标，有助于检测小目标。本文特征金字塔利用 Darknet-53 网络中的第 3、4 和 5 个残差块进行特征融合，产生 3 个输出预测头，分别输出分类信息、边界框回归信息和外观信息。

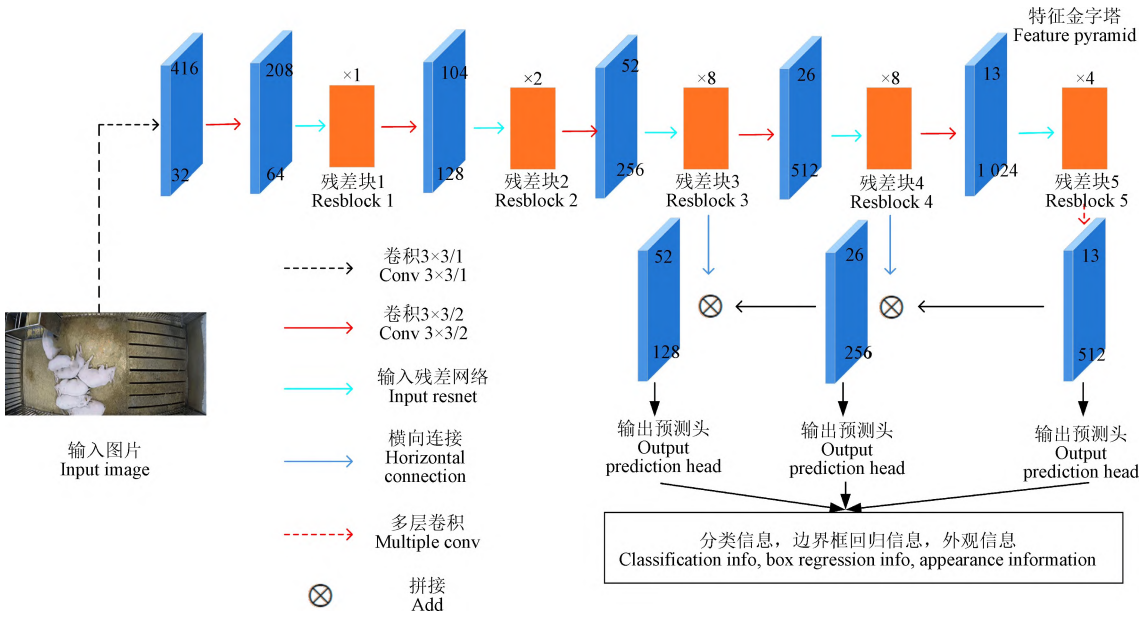


图 2 特征提取网络结构

Fig.2 Diagram of feature extraction network structure

表 1 Darknet-53 网络结构参数

Table 1 Darknet-53 network structure parameters

名称 Name	输出图片大小 Output image size/Pixel	滤波器个数 Number of filters	滤波器大小 Filters size/Pixel	数量 Numbers
卷积层 1 Conv layer 1	416×416	32	3×3	1
卷积层 2 Conv layer 2	208×208	64	3×3/2	1
残差层 1 Residual layer 1	208×208	32，64	1×1，3×3	1
卷积层 3 Conv layer 3	104×104	128	3×3/2	1
残差层 2 Residual layer 2	104×104	64，128	1×1，3×3	2
卷积层 4 Conv layer 4	52×52	256	3×3/2	1
残差层 3 Residual layer 3	52×52	128，256	1×1，3×3	8
卷积层 5 Conv layer 5	26×26	512	3×3/2	1
残差层 4 Residual layer 4	26×26	256，512	1×1，3×3	8
卷积层 6 Conv layer 6	13×13	1 024	3×3/2	1
残差层 5 Residual layer 5	13×13	512，1 024	1×1，3×3	4

1.3 数据关联模块

本文 JDE 算法的学习目标为多任务协同学习，其总体损失 L_{total} 为分类损失、边界框回归损失和外观信息学习损失之和，如式（1）所示。

$$L_{total}=\omega_\alpha L_\alpha+\omega_\beta L_\beta+\omega_\gamma L_\gamma \tag{1}$$

式中 ω_α、ω_β、ω_γ 分别为分类、边界框回归和外观信息学习的权重值，L_α 为分类损失，L_γ 为外观信息学习损失，其中损失均为交叉熵损失，计算公式如式（2）所示。

$$L_\alpha=L_\gamma=-\frac{1}{N}\sum_{i}^{}\sum_{c=1}^{M}y_{ic}\lg p_{ic} \tag{2}$$

式中 M 为类别的数量，N 为样本数，y_{ic} 为符号函数（0 或 1），c 为类别数。如果样本 N 的真实类别等于 c，则 y_{ic}=1，否则 y_{ic}=0。p_{ic} 为观测样本 i 属于类别 c 的预测概率。

L_β 为边界框回归损失，为 smooth-L1 损失，计算公式如式（3）所示。

$$L_\beta=\begin{cases}0.5x^2, |x|<1\\ |x|-0.5, \text{其他}\end{cases} \tag{3}$$

式中 x 为输入样本。

算法采用基于任务的不确定性计算加权系数，最终自动加权的损失 L_{total} 如式（4）所示。

$$L_{total}=\frac{1}{2}\left(\frac{1}{e^{s_\alpha}}L_\alpha+s_\alpha\right)+\frac{1}{2}\left(\frac{1}{e^{s_\beta}}L_\beta+s_\beta\right)+\frac{1}{2}\left(\frac{1}{e^{s_\gamma}}L_\gamma+s_\gamma\right) \tag{4}$$

式中 s_α、s_β、s_γ 为每个个体损失的任务依赖的不确定性，为可学习参数。

模型通过分类损失和回归损失学习到的分类信息和回归信息生成检测框对视频帧中每个猪只进行定位，外观学习损失得到的外观信息包括每个猪只的外观特征，二者通过数据关联，对每头猪分配 ID，实现多目标跟踪。猪只多目标跟踪的具体实现流程如图 3 所示，具体步骤如下：

1）创建初始跟踪轨迹。对于给定的视频帧序列，第一帧将根据视频帧序列的检测结果利用卡尔曼滤波对轨迹进行初始化，并维护一个跟踪轨迹池，包含所有可能与预测值相关联的轨迹。

2）数据关联。对于下一帧的输出结果，利用卡尔曼滤波进行轨迹预测，计算出预测值与轨迹池之间的运动亲和信息和外观亲和信息，其中外观亲和信息采用余弦相似度计算，运动亲和信息采用马氏距离计算，然后利用匈牙利算法的代价矩阵进行轨迹分配。

3）更新轨迹。如果出现在 2 帧内的预测值没有被分配给任何一个轨迹池中的轨迹，那么这条轨迹将被初始

化为新的轨迹，然后根据卡尔曼滤波进行所有匹配轨迹状态的更新，如果某条轨迹在连续 30 帧内没有被更新，则终止该轨迹，所有视频帧处理完毕后，输出视频帧序列。

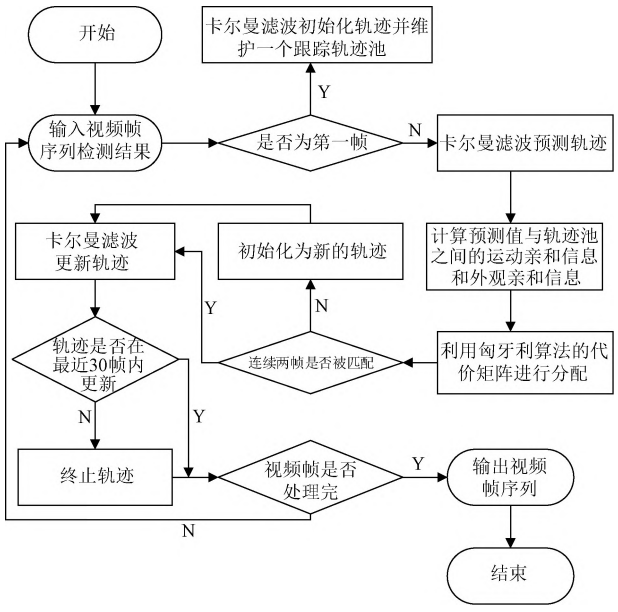


图 3　卡尔曼滤波结合匈牙利算法的猪只目标跟踪流程

Fig.3　Pig object tracking process of Kalman filter combined with Hungarian algorithm

2　数据准备与评价指标

2.1　数据集

本试验采用的数据集包括 2 部分：一部分为 Psota 等[42]提供的公开数据集，包含不同日龄、大小、数量和不同环境的猪只视频，其中，视频 1、2、4、5 为保育猪（3～10 周龄），视频 6、7、8、9、10 为早期育成猪（11～18 周龄），视频 12、15 为晚期育成猪（19～26 周龄）。根据时间段的不同将猪只的活动水平分为 3 类：白天的高活动、白天（或夜晚）的中等活动、白天（或夜晚）的低活动，详表 2。同时，根据人工观察，将猪只个数较多且黏连遮挡情况较为严重的视频定义为稠密视频，反之为稀疏视频，见表 2。另外一部分为自建数据集[43]。两部分数据集均为俯拍视频片段，由于摄像头高度及焦距的影响，不可避免拍摄到猪圈外的物品，因此，在试验中采用视频裁剪方法将视角固定为猪圈内，以减少外部环境的影响。

表 2　公开数据集

Table 2　Public dataset

视频 Videos	白天 Day	黑夜 Night	稀疏 Sparse	稠密 Dense	活动水平 Activity level	猪只个数 Number of pigs
1	√	—	√	—	高	7
2	√	—	√	—	低	7
4	√	—	—	√	中	15
5	—	√	√	—	中	8
6	√	—	—	√	高	16
7	√	—	—	√	中	12
8	—	√	—	√	低	13
9	√	—	—	√	中	14
10	—	√	—	√	中	14
12	√	—	—	√	低	15
15	—	√	—	√	中	16

首先，利用 FFmpeg 软件完成视频剪辑，从中截取稠密、稀疏、白天、黑夜的视频，2 部分数据集共 21 个视频。然后利用 DarkLabel 软件对数据进行标注，其中，公开数据集 11 个视频，共 3 300 张图像，自建数据集 10 个视频，共 1 000 张图像。部分数据集如图 4 所示。为对比不同场景下模型的检测和跟踪能力，选取不同的视频进行模型训练和测试，参与训练的视频不参与测试。本文共设计 3 个试验，其中试验 1 以视频 4、6、12 为测试集，这些视频均为白天稠密，其余视频为训练集。试验 2 以视频 2、5、8 为测试集，其中视频 5、8 分别为夜晚稀疏与夜晚稠密，视频 2 为白天稀疏，其余视频为训练集。试验 3 以自建数据集的 7 个视频为测试集（视频 3、11、14、16、18、19、21），另外 3 个视频为测试集（视频 13、17、20）。其中猪只活动水平定义如下：根据视频的人工观察结果，在白天（10:00－12:30）猪只的饮食和玩耍等行为较频繁，此时间段定义为猪只白天的高活动水平。在白天（12:30－17:00）或夜晚（17:00－20:00）猪只的饮食和玩耍等行为没有白天（10:00－12:30）高，此时间段定义为白天或夜晚的中等活动水平。在白天（7:00－10:00）或夜晚（20:00－7:00）猪只的饮食和玩耍等行为较少，躺卧行为较多，此时间段定义为白天或夜晚的低活动水平。

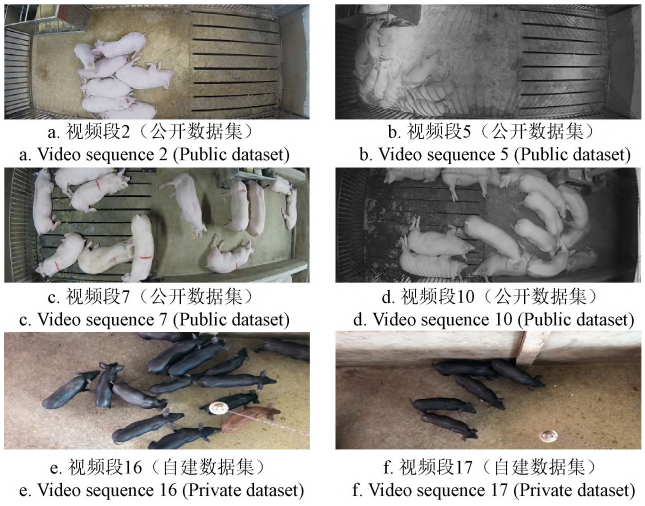


图 4　部分数据集

Fig.4　Part of the dataset

2.2　试验环境

本文所有试验在同一计算机上完成，硬件配置为 12th Gen Intel(R) i9-12900KF CPU，NVIDIA GeForce RTX 3090 GPU，32GB 内存，64 位 Linux 操作系统，Pytorch 版本 1.7.1，Python 版本 3.8，CUDA 版本 11.0。

训练过程中设置图片尺寸为 416×416（像素），批处理大小（Batchsize）设置为 32，初始学习率（Learning Rate）为 0.01，动量（Momentum）设置为 0.9，共训练 30 个时期（Epoch），使用随机梯度下降法（Stochastic Gradient Descent, SGD）进行优化，保存训练过程中精度最高的模型参数进行模型测试。

2.3 评价指标

选择精确率（Precision，P），召回率（Recall，R）和平均精度均值（mean Average Precision，mAP）3 个指标评判模型的检测性能。精确率衡量模型对猪只目标检测的精确程度，如式（5），其中 DTP 是检测正确的目标数量，DFP 是检测错误的目标数量。

$$P=\frac{\mathrm{DTP}}{\mathrm{DTP+DFP}}\times 100\% \tag{5}$$

召回率衡量模型对猪只目标检测的覆盖能力，如式（6），其中 DFN 是漏检的目标数量。

$$R=\frac{\mathrm{DTP}}{\mathrm{DTP+DFN}}\times 100\% \tag{6}$$

平均精度均值是对检测的类别对应的精度均值取平均，如式（7），其中 $P(R)$是以召回率 R 为自变量，精确率 P 为因变量的函数。

$$\mathrm{mAP}=\int_0^1 P(R)\mathrm{d}R \tag{7}$$

选择多目标跟踪精度（Multiple Object Tracking Accuracy，MOTA）和 IDF1 得分（ID F1 Score）作为多目标跟踪的主要评价指标。MOTA 衡量跟踪器检测目标和保持轨迹跟踪的性能。IDF1 为引入跟踪目标标号 ID 的 F1 值，由于引入了跟踪目标标号 ID，IDF1 更重视目标的轨迹跟踪能力。MOTA 计算公式如式（8）所示。

$$\mathrm{MOTA}=1-\frac{\sum_t(\mathrm{FP+FN+IDS})}{\sum_t g_t} \tag{8}$$

式中 FP 为在第 t 帧中目标误报总数（假阳性）；FN 为在第 t 帧目标丢失总数（假阴性）；IDS 为在第 t 帧中跟踪目标标号 ID 发生切换的次数；g_t 是 t 时刻观测到的目标数量。

IDF1 计算公式如式（9）所示。

$$\mathrm{IDF1}=\frac{2\mathrm{IDTP}}{2\mathrm{IDTP+IDFP+IDFN}} \tag{9}$$

式中 IDTP 为 ID 保持不变的情况下正确跟踪到的目标总数，IDFP 为 ID 保持不变的情况下跟踪错误的目标总数，IDFN 为 ID 保持不变的情况下跟踪目标丢失总数。

此外，其他相关指标还有碎片数（Fragmentation，FM）、主要跟踪到的目标（Mostly Tracked Target，MT）（被跟踪到的轨迹比例大于 80%）、主要丢失目标（Mostly Lost Target，ML）（被跟踪到的轨迹比例小于 20%）、部分跟踪到的目标（Partially Tracked Target，PT）（被跟踪到的轨迹比例不大于 80%且不小于 20%）、一条跟踪轨迹改变目标标号 ID 的次数（Identity Switches，IDS）以及平均每秒传输帧数（Frames Per Second，FPS）。

本文对群养生猪目标跟踪模型性能的分析选择 MOTA、IDF1 和 FPS 作为主要评价指标，辅助以 FP、FN、FM、IDS、MT、ML 等指标进行模型的性能评估。其中 MOTA、IDF1、MT 和 FPS 数值越高模型性能越好，FP、FN、FM、IDS 和 ML 数值越低模型性能越好。

3 结果与分析

3.1 JDE 模型试验结果

JDE 模型的检测结果见表 3。可以发现，本文算法在公开数据集中的 mAP 平均值达到 92.5%，测试集 2、4、6、8、12 视频的 mAP 分别为 96.2%、95.6%、96.1%、98.0%、92.2%。对于视频 5，其 mAP 为 77.0%，主要原因是该视频的场景与其他视频相比差异较大，增加了目标检测的难度；在自建数据集中的 mAP 平均值达到 93.8%，总体平均 mAP 达到 92.9%，表明本文 JDE 算法对于不同复杂场景具有较好的检测能力。

表 3 JDE 模型的目标检测试验结果

Table 3 Object detection experiment results of the Joint Detection and Embedding (JDE) model

%

测试集 Test set	视频 Video	精确率 Precision P	召回率 Recall R	平均精度均值 Mean Average Precision mAP
公开数据集 Public dataset	2	96.0	94.8	96.2
	4	84.7	96.1	95.6
	5	81.2	79.2	77.0
	6	96.2	95.1	96.1
	8	99.2	87.9	98.0
	12	83.2	93.6	92.2
自建数据集 Private dataset	13	99.5	90.9	94.8
	17	93.1	99.0	98.6
	20	94.5	80.1	88.0

JDE 模型的跟踪结果如表 4 所示。可以发现，在公开数据集中，视频 2、4、5、6、8、12 的 MOTA 分别为 91.4%、82.5%、59.2%、90.8%、94.2%、74.4%，平均 MOTA 为 82.1%，在自建数据集中，视频 13、17、20 的 MOTA 分别为 84.4%、88.1%、90.2%，平均 MOTA 为 87.6%，总体平均 MOTA 为 83.9%。不同视频的 MOTA 产生差别的主要原因是每个视频的环境不同，如视频背景、白天、黑夜、稀疏、稠密和猪只的活动状态，在视频背景干扰严重、猪只活动较为频繁（如饮食，玩耍等行为）情况下，MOTA 相对较低，在夜晚视频 8 中，猪只活动较少且背景对猪只的干扰较小，MOTA 最高，为 94.2%。在夜晚视频 5 中，视频背景干扰严重，MOTA 较低，为 59.2%，根据 IDF1 和 FPS 可以看出，本文 JDE 模型在公开数据集中的 IDF1 平均值为 77.7%，FPS 平均值为 74.26 帧/s，在自建数据集中的 IDF1 平均值为 83.5%，FPS 平均值为 73.19 帧/s，总体平均 IDF1 值为 79.6%，总体平均 FPS 值为 73.9 帧/s。可以发现，本文 JDE 模型对猪只目标的 ID 跟踪精度和 FPS 均达到较高水平，能够实现实际养殖环境下的群养猪多目标快速实时跟踪，为实际群养猪养殖场的精准管理提供技术支持。

表 4　JDE 模型的多目标跟踪试验结果

Table 4　Multiple object tracking experiment results of the JDE model

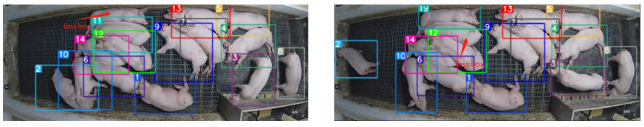
视频 Video	猪只个数 Pig numbers	主要跟踪到的目标 Mostly tracked target	部分跟踪到的目标 Partially tracked target	主要丢失目标 Mostly lost target	假阳性 False positive	假阴性 False negative	ID 跳变 ID switch	碎片数 Fragmentation	多目标跟踪精度 Multiple object tracking accuracy/%	IDF1 得分 IDF1 score/%	每秒检测帧数 Frames Per Second FPS/(帧·s^{-1})
2	7	6	1	0	29	136	14	19	91.4	80.3	77.64
4	15	12	3	0	377	384	26	30	82.5	75.3	72.47
5	8	4	4	0	438	497	40	51	59.2	59.2	76.46
6	16	14	2	0	115	285	39	63	90.8	75.1	71.56
8	13	13	0	0	104	108	15	26	94.2	94.1	73.88
12	15	12	3	0	847	286	14	45	74.4	82.2	73.55
13	14	11	3	0	24	185	9	14	84.4	79.1	71.19
17	5	5	0	0	24	31	4	9	88.1	88.3	75.73
20	8	6	1	1	6	57	6	5	90.2	83.1	72.66

猪只白天稀疏和稠密 2 种分布情况的可视化分析结果如图 5 所示。

第100帧 100th frame　　第217帧 217th frame

a. 视频2（白天，稀疏）

a. Video 2 (daytime, sparse)



第9帧 9th frame　　第46帧 46th frame

b. 视频4（白天，稠密）

b.Video 4 (daytime, dense)

注：图中数字表示猪只 ID 号，算法中第一帧图像的检测会对每头猪只分配一个从 1 递增的 ID 号，例如（1、2、3…），对后续帧进行检测和跟踪时，由于猪只的移动，可能会对某个猪只的 ID 识别错误，此时把这个猪只识别为新的猪只，则该猪只的 ID 号就变为错误的 ID 号，直至所有视频帧处理完毕。下同。

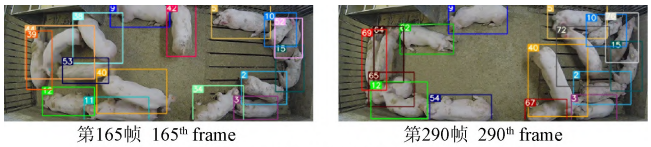
Note: The number in the figure indicates the pig ID No., the first image detection frame of the algorithm will assign an incremental ID No. from 1 to each pig, for example (1, 2, 3...), when detecting and tracking the subsequent frames, due to the movement of the pig, the ID of a pig may be identified incorrectly, at this time to identify this pig as a new pig, the ID No. of the pig will change to the wrong ID No. until all video frames are processed. Same below.

图 5　猪只白天稀疏和稠密分布情况的可视化分析结果

Fig.5　Results of the visualization analysis of the sparse and dense distribution of pigs during the day

对于猪只白天稀疏的视频 2，本文算法可以准确地检测和跟踪每一只猪，如图 5a。但是，对猪只白天稠密且猪只粘连遮挡情况较为严重的视频 4 存在漏检，如图 5b 中箭头标识的猪。这说明在猪只白天稠密的环境下，由于猪只目标出现漏检，从而影响了算法的跟踪性能。

对猪只白天和夜晚情况下的可视化分析如图 6 所示，可以发现，在猪只白天稠密且有遮挡的情况下，本文 JDE 模型可以很好地跟踪到每一只猪，如图 6a。在夜晚视频背景比较黑暗且猪只密集有遮挡的情况下，JDE 模型也可以准确地跟踪每一只猪，如图 6b。但在猪只夜晚稀疏的视频 5 中，由于所有猪只都分布于猪圈的左方，且视频背景和猪只颜色相似，这使得检测器和跟踪器较难检测和跟踪这些猪只目标，出现猪只漏检的情况，如图 6c 所示。总体上，本文 JDE 模型对于不同场景下的群养生猪多目标跟踪达到较好水平。



第165帧 165th frame　　第290帧 290th frame

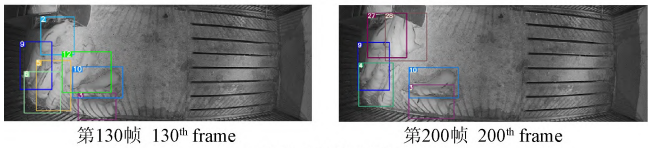
a. 稠密（白天,视频6）

a. Dense (daytime, video 6)

第167帧 167th frame　　第229帧 229th frame

b. 稠密（夜晚,视频8）

b. Dense (night, video 8)



第130帧 130th frame　　第200帧 200th frame

c. 稀疏（夜晚,视频5）

c. Sparse (night, video 5)

图 6　猪只白天和夜晚不同分布情况的的可视化分析结果

Fig.6　Results of the visualization analysis of the different distribution of pigs during the day and night

3.2　SDE 模型试验结果

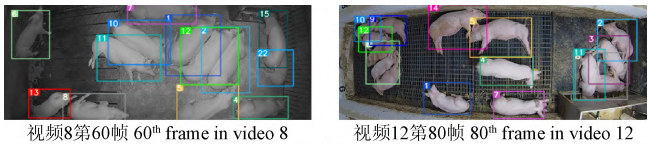
为验证本文 JDE 模型的多目标跟踪性能，与经典的 SDE 模型进行对比试验。SDE 检测器与本文 JDE 模型相同，跟踪器使用 DeepSORT，采用相同的公开数据集进行训练和测试，试验结果如表 5 所示。可以发现，SDE 模型的 MOTA 和 IDF1 平均值分别为 81.6%和 78.2%，对比表 4，本文 JDE 模型的 MOTA 提升了 0.5 个百分点。从总体性能指标来看，本文 JDE 模型的 MT、PT、ML、FN、MOTA 和 FPS 指标均优于 SDE 模型。在速度方面，SDE 模型的 FPS 均值为 16.88 帧/s，本文 JDE 模型的 FPS 均值达到 74.26 帧/s。总体来说，二者在跟踪准确度和跟踪精度接近情况下，本文 JDE 模型的视频处理速度比 SDE 模型提升了 340%，这对于实现养殖场长时间群养生猪视频的实时多目标跟踪有重要意义。

表 5 SDE 模型的多目标跟踪试验结果

Table 5 Multiple object tracking experiment results of the Separate Detection and Embedding (SDE) model

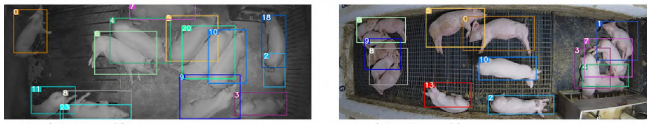
视频 Video	猪只个数 Pig numbers	主要跟踪到的目标 Mostly tracked target	部分跟踪到的目标 Partially tracked target	主要丢失目标 Mostly lost target	假阳性 False positive	假阴性 False negative	ID 跳变 ID switch	碎片数 Fragmentation	多目标跟踪精度 Multiple object tracking accuracy/%	IDF1 得分 IDF1 score/%	FPS/(帧·s^{-1})
2	7	6	1	0	128	138	8	13	87.0	85.1	21.18
4	15	15	0	0	173	215	33	30	90.6	81.3	17.00
5	8	3	4	1	216	824	19	39	55.9	60.4	15.28
6	16	13	3	0	22	636	32	53	85.6	74.2	16.16
8	13	13	0	0	430	74	9	10	86.8	87.1	16.52
12	15	10	3	2	45	1 075	8	29	74.9	76.0	15.11

选取部分数据集进行可视化分析，结果如图 7 所示，在猪只夜晚稠密的视频 8 中，SDE 模型存在错检情况，如图 7b 左下角第二头猪出现 2 个跟踪框，而本文 JDE 模型没有错检情况，如图 7a 所示。在猪只白天稠密的视频 12 中，由于猪只密集躺在一起，检测器较容易发生漏检，如图 7a、7b，JDE 模型漏检 2 头猪，SDE 模型漏检 3 头猪，JDE 比 SDE 模型具有更好的检测跟踪结果。



视频8第60帧 60th frame in video 8　视频12第80帧 80th frame in video 12

a. JDE



视频8第60帧 60th frame in video 8　视频12第80帧 80th frame in video 12

b. SDE

图 7 JDE 与 SDE 模型对猪只不同分布情况的可视化结果对比

Fig.7 Comparison of visualization results of JDE and SDE models for different distribution of pigs

此外，文献[40]采用基于 SDE 模型对猪只目标检测的平均精度均值达 99.0%，多目标跟踪精度 MOTA 为 96.8%，但文献[40]的数据场景单一，无法应对其他场景。尽管包括白天和黑夜（光照变化），但训练和测试场景相同。本文数据集包含不同情况下的场景，共有 11 个视频场景，各个场景环境不同，猪只大小也不同，训练和测试场景完全不相同。

3.3 TransTrack 试验结果

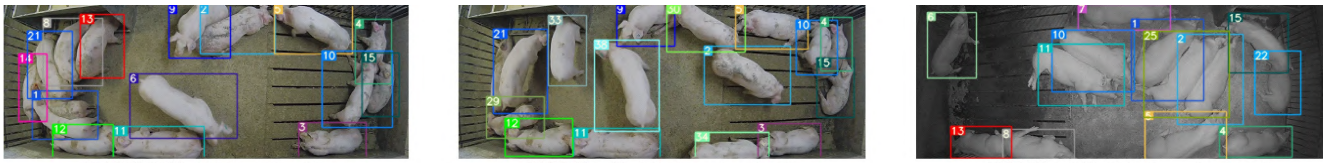
为进一步验证本文算法在群养猪多目标跟踪方面的性能，与 TransTrack 模型在相同的公开数据集上进行对比试验，试验结果如表 6 所示。TransTrack 模型的平均 MOTA、IDF1 和 FPS 分别为 71.7%、71.1%和 17.53 帧/s，与表 4 结果比较发现，本文 JDE 模型比 TransTrack 模型的 MOTA 和 IDF1 分别提升 10.4 和 6.6 个百分点，同时 FPS 提升 324%。从性能指标 MT、PT、ML、FP、FN、IDS、FM、MOTA、IDF1 和 FPS 的数值对比可以发现，本文 JDE 模型性能均优于 TransTrack 模型。

表 6 TransTrack 模型的试验结果

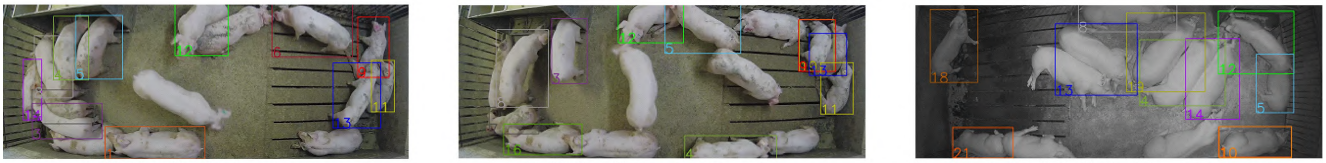
Table 6 Experimental results of the TransTrack model

视频 Video	猪只个数 Pig numbers	主要跟踪到的目标 Mostly tracked target	部分跟踪到的目标 Partially tracked target	主要丢失目标 Mostly lost target	假阳性 False positive	假阴性 False negative	ID 跳变 ID switch	碎片数 Fragmentation	多目标跟踪精度 Multiple object tracking accuracy/%	IDF1 得分 IDF1 score/%	FPS/(帧·s^{-1})
2	7	6	1	0	98	130	19	20	88.2	78.9	17.29
4	15	14	1	0	487	337	33	67	81.0	84.1	17.64
5	8	4	3	1	109	621	31	54	68.3	73.9	17.66
6	16	7	9	0	516	1 370	94	166	58.8	46.6	17.83
8	13	9	3	1	507	762	41	64	66.4	69.9	17.53
12	15	11	2	2	546	909	16	33	67.3	73.4	17.21

对 2 种模型的跟踪结果选取部分数据进行可视化分析，结果如图 8 所示。对比发现，相较于 TransTrack 模型，JDE 模型对猪只严重遮挡情况有更好的检测和跟踪能力，如图 8a。而 TransTrack 模型在猪只严重遮挡情况下，会出现猪只的漏检或者是猪只追踪的缺失，如图 8b。可以看出，本文算法在不同场景中，检测框更加贴合猪只目标，对于严重遮挡的猪只目标具有更强的检测跟踪能力。



a. JDE



b. TransTrack

图 8　JDE 与 TransTrack 模型的可视化结果对比

Fig.8　Visualization comparison of JDE model and TransTrack model

4　结　论

1）本文 JDE 模型在二阶段目标检测和跟踪分离框架的基础上进行改进，在输出检测框的同时，给网络增加目标外观信息学习损失对应的输出分支，实现检测和跟踪的多任务协同学习，实现联合目标检测和跟踪。

2）本文制作了 2 个数据集，分别为公开数据集和自建数据集。其数据场景复杂多样，各个场景的猪只大小、数量、日龄和光照条件都不同，并在公开数据集中与 SDE 模型和 TransTrack 模型进行了对比。

3）试验结果表明，本文 JDE 模型在 2 个数据集的总体平均精度均值 mAP 为 92.9%，平均多目标跟踪精度 MOTA 为 83.9%，平均 IDF1 得分为 79.6%，平均每秒检测帧数 FPS 为 73.9。在公开数据集中与 TransTrack 模型进行对比，本文 JDE 模型的 MOTA 和 IDF1 分别提升 10.4 和 6.6 个百分点，FPS 提升 324%。在公开数据集中与 SDE 模型进行对比，本文 JDE 模型在 MOTA 和 IDF1 的数值接近下，FPS 提升 340%，解决了 SDE 模型目标检测和跟踪模块分离导致目标跟踪速度慢的问题，这对于养殖场群养生猪长时间视频的实时多目标跟踪具有重要意义。

[参　考　文　献]

[1] Rowe E, Dawkins M S, Gebhardt-Henrich S G A. Systematic review of precision livestock farming in the poultry sector: Is Technology focussed on improving bird welfare?[J]. Animals (Basel), 2019, 9(9): 614.

[2] Cowton J, Kyriazakis I, Plotz T , et al. A combined deep learning GRU-autoencoder for the early detection of respiratory disease in pigs using multiple environmental sensors[J]. Sensors (Basel), 2018, 18(8): 2521.

[3] Sébastien F, Alain N R, Benoit L. Rethinking environment control strategy of confined animal housing systems through precision livestock farming[J]. Biosystems Engineering, 2017, 155: 96-123.

[4] Zambelis A, Wolfe T, Vasseur E. Technical note: Validation of an ear-tag accelerometer to identify feeding and activity behaviors of tiestall-housed dairy cattle[J]. Journal of Dairy Science, 2019, 102(5): 4536-4540.

[5] Giovanetti V, Decandia M, Molle G, et al. Automatic classification system for grazing, ruminating and resting behaviour of dairy sheep using a tri-axial accelerometer[J]. Livestock Science, 2017, 196: 42-48.

[6] Krista M M, Elizabeth A S, Carlos J B R, et al. Technical note: Validation of an automatic recording system to assess behavioural activity level in sheep (Ovis aries)[J]. Small Ruminant Research, 2015, 127: 92-96.

[7] Chen C, Zhu W X, Ma C H, et al. Image motion feature extraction for recognition of aggressive behaviors among group-housed pigs[J]. Computers and Electronics in Agriculture, 2017, 142: 380-387.

[8] Chen C, Zhu W X, Guo Y Z, et al. A kinetic energy model based on machine vision for recognition of aggressive behaviours among group-housed pigs[J]. Livestock Science, 2018, 218: 70-78.

[9] Chen C, Zhu W X, Liu D, et al. Detection of aggressive behaviours in pigs using a RealSence depth sensor[J]. Computers and Electronics in Agriculture, 2019, 166: 105003.

[10] Chen C, Zhu W X, Steibel J, et al. Recognition of aggressive episodes of pigs based on convolutional neural network and long short-term memory[J]. Computers and Electronics in Agriculture, 2020, 169: 105166.

[11] Alameer A, Kyriazakis I, Bacardit J. Automated recognition of postures and drinking behaviour for the detection of compromised health in pigs[J]. Scientific Reports, 2020, 10(1): 13665.

[12] Lao F, Brown B, Stinn J P, et al. Automatic recognition of lactating sow behaviors through depth image processing[J]. Computers and Electronics in Agriculture, 2016, 125: 56-62.

[13] Zhu W X, Guo Y Z, Jiao P P, et al. Recognition and drinking behaviour analysis of individual pigs based on machine vision[J]. Livestock Science, 2017, 205: 129-136.

[14] Leonard S M, Xin H, Brown-Brandl T M, et al. Development and application of an image acquisition system for characterizing sow behaviors in farrowing stalls[J]. Computers and Electronics in Agriculture, 2019, 163: 104866.

[15] Yang A Q, Huang H S, Zheng B, et al. An automatic recognition framework for sow daily behaviours based on

motion and image analyses[J]. Biosystems Engineering, 2020, 192: 56-71.

[16] Zhang Y Q, Cai J H, Xiao D Q, et al. Real-time sow behavior detection based on deep learning[J]. Computers and Electronics in Agriculture, 2019, 163: 104884.

[17] Nasirahmadi A, Hensel O, Edwards S, et al. Automatic detection of mounting behaviours among pigs using image analysis[J]. Computers and Electronics in Agriculture, 2016, 124: 295-302.

[18] Li D, Chen Y F, Zhang K F, et al. Mounting beaviour recognition for pigs based on deep learning[J]. Sensors (Basel), 2019, 19(22): 4924.

[19] Nasirahmadi A, Sturm B, Olsson A, et al. Automatic scoring of lateral and sternal lying posture in grouped pigs using image processing and support vector machine[J]. Computers and Electronics in Agriculture, 2019, 156: 475-481.

[20] Zheng C, Zhu X M, Yang X F, et al. Automatic recognition of lactating sow postures from depth images by deep learning detector[J]. Computers and Electronics in Agriculture, 2018, 147: 51-63.

[21] Zhu X M, Chen C X, Zheng B, et al. Automatic recognition of lactating sow postures by refined two-stream RGB-D faster R-CNN[J]. Biosystems Engineering, 2020, 189: 116-132.

[22] Zheng C, Yang X F, Zhu X M, et al. Automatic posture change analysis of lactating sows by action localisation and tube optimisation from untrimmed depth videos[J]. Biosystems Engineering, 2020, 194: 227-250.

[23] Jorquera-Chavez M, Fuentes S, Dunshea F R, et al. Remotely sensed imagery for early detection of respiratory disease in pigs: A pilot study[J]. Animals (Basel), 2020, 10(3): 451.

[24] Jorquera-Chavez M, Fuentes S, Dunshea F R, et al. Using imagery and computer vision as remote monitoring methods for early detection of respiratory disease in pigs[J]. Computers and Electronics in Agriculture, 2021, 187: 106283.

[25] Zhao K X, He D J. Target detection method for moving cows based on background subtraction[J]. International Journal of Agricultural and Biological Engineering, 2015, 8(1): 42-49.

[26] Zhang Y G, Zheng J, Zhang C, et al. An effective motion object detection method using optical flow estimation under a moving camera[J]. Journal of Visual Communication and Image Representation, 2018, 55: 215-228.

[27] 于欣，侯晓娇，卢焕达，等. 基于光流法与特征统计的鱼群异常行为检测[J]. 农业工程学报，2014，30(2)：162-168.
Yu Xin, Hou Xiaojiao, Lu Huanda, et al. Anomaly detection of fish school behavior based on features statistical and optical flow methods[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2014, 30(2): 162-168. (in Chinese with English abstract)

[28] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]// Columbus, OH, USA, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014: 580-587.

[29] Girshick R. Fast R-CNN[C]// Santiago, Chile, IEEE International Conference on Computer Vision (ICCV), 2015: 1440-1448.

[30] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.

[31] 王浩，曾雅琼，裴宏亮，等. 改进 Faster R-CNN 的群养猪只圈内位置识别与应用[J]. 农业工程学报，2020，36(21)：201-209.
Wang Hao, Zeng Yaqiong, Pei Hongliang, et al. Recognition and application of pigs'position in group pens based on improved Faster R-CNN[J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2020, 36(21): 201-209. (in Chinese with English abstract)

[32] Redmon J, Farhadia A. YOLOv3: An incremental improvement [EB/OL]. 2018-04-08, https://pjreddie.com/media/files/papers/YOLOv3.pdf.

[33] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]//Las Vegas, NV, USA, Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 779-788.

[34] Redmon J, Farhadi A. YOLO9000: Better,faster,stronger[C]// Honolulu, HI, USA, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 7263-7271.

[35] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: Optimal speed and accuracy of object detection[EB/OL]. 2020-04-23, https://arxiv.org/pdf/2004.10934.pdf.

[36] 金耀，何秀文，万世主，等. 基于 YOLO v3 的生猪个体识别方法[J]. 中国农机化学报，2021，42(2)：178-183.
Jin Yao, He Xiuwen, Wan Shizhu, et al.Individual pig identification method based on YOLOv3[J]. Journal of Chinese Agricultural Mechanization, 2021, 42(2): 178-183. (in Chinese with English abstract)

[37] Bewley A, Ge Z Y, Ott L, et al. Simple online and realtime tracking[C]//Phoenix, Arizona, USA. IEEE International Conference on Image Processing (ICIP), 2016: 3464-3468.

[38] Wojke N, Bewley A, Paulus D. Simple online and realtime tracking with a deep association metric[C]//Beijing, China. IEEE International Conference on Image Processing (ICIP), 2017: 3645-3649.

[39] 张宏鸣，汪润，董佩杰，等. 基于 DeepSORT 算法的肉牛多目标跟踪方法[J]. 农业机械学报，2021，52(4)：249-256.
Zhang Hongming, Wang Run, Dong Peijie, et al. Multi-object tracking method for beef cattle based on DeepSORT algorithm[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(4): 249-256. (in Chinese with English abstract)

[40] 张伟，沈明霞，刘龙申，等. 基于 CenterNet 搭配优化 DeepSORT 算法的断奶仔猪目标跟踪方法研究[J]. 南京农业大学学报，2021，44(5)：973-981.
Zhang Wei, Shen Mingxia, Liu Longshen, et al. Research on weaned piglet target tracking method based on CenterNet collocation optimized DeepSORT algorithm[J]. Journal of Nanjing Agricultural University, 2021, 44(5): 973-981. (in Chinese with English abstract)

[41] Sun P Z, Cao J K, Jiang Y, et al. TransTrack: Multiple object tracking with transformer[EB/OL]. 2021-05-04, https://arxiv.org/abs/2012.15460v1.

[42] Psota E T, Schmidt T, Mote B, et al. Long-term tracking of

group-housed livestock using keypoint detection and MAP estimation for individual animal identification[J]. Sensors (Basel), 2020, 20(13): 3670.

[43] Tu S Q，Yuan W J，Liang Y，et al. Automatic detection and segmentation for group-housed pigs based on PigMS R-CNN[J]. Sensors (Basel), 2021, 21(9): 3251.

Multiple object tracking of group-housed pigs based on JDE model

Tu Shuqin, Huang Lei, Liang Yun※, Huang Zhengxin, Li Chengjie, Liu Xiaolong

(*College of Mathematics and Informatics*, *South China Agricultural University*, *Guangzhou* 510642, *China*)

Abstract: Pig production has been always the pillar of the industrial livestock industry in China. Therefore, the pig industry is closely related to food safety, social stability, and the coordinated development of the national economy. An intelligent video surveillance can greatly contribute to the large-scale production of animal husbandry under labor shortage at present. It is very necessary to accurately track and identify the abnormal behavior of group-housed pigs in the breeding scene. Much effort has been focused on Multiple Object Tracking (MOT) for pig detection and tracking. Among them, two parts are included in the Tracking By Detection (TBD) paradigm, e.g., the Separate Detection and Embedding (SDE) model. Previously, the detector has been developed to detect pig objects. And then the tracking models have been selected for the pig tracking using Kalman filter and Hungarian (Sort or DeepSORT). The detection and association steps have been designed to increase the running and training time of the model in the dominant MOT strategy. Thus, real-time tracking cannot fully meet the requirement of the group-housed pigs. In this study, a Joint Detection and Embedding (JDE) model was proposed to automatically detect the pig objects and then track each one in the complex scenes (day or night, sparse or dense). The core of JDE model was to integrate the detector and the embedding model into a single network for the real-time MOT system. Specifically, the JDE model incorporated the appearance model into a single-shot detector. As such, the simultaneous output was performed on the corresponding appearance to improve the runtime and operational efficiency of the model. An overall loss of one multiple task learning loss was utilized in the JDE model. Three loss functions were included classification, box regression and appearance. Three merits were achieved after operations. Firstly, the multiple tasks learning loss was used to realize the object detection and appearance to be learned in a shared model, in order to reduce the amount of occupied memory. Secondly, the forward operation was computed using the multiple tasks loss at one time. The overall inference time was reduced to improve the efficiency of the MOT system. Thirdly, the performance of each prediction head was promoted to share the same set of low-level features and feature pyramid network architecture. Finally, the data association module was utilized to process the outputs of the detection and appearance head from the JDE, in order to produce the position prediction and ID tracking of multiple objects. The JDE model was validated on the special dataset under a variety of settings. The special dataset was also built with a total of 21 video segments and 4 300 images using the dark label video annotation software. Two types of datasets were obtained, where the public dataset contained 11 video sequences and 3 300 images, and the private dataset contained 10 video segments and 1 000 images. The experimental results show that the mean Average Precision (mAP), Multiple Object Tracking Accuracies (MOTA), IDF1 score, and FPS of the JDE on all test videos were 92.9%, 83.9%, 79.6%, and 73.9 frames/s, respectively. A comparison was also made with the SDE model and TransTrack method on the public dataset. The JDE model improved the FPS by 340%, and the MOTA by 0.5 percentage points in the same test dataset, compared with the SDE model. It infers the sufficient real-time performance of MOT using the JDE model. The MOTA, IDF1 metrics, and FPS of the JDE model was improved by 10.4 and 6.6 percentage points, and 324%, respectively, compared with the TransTrack model. The visual tracking demonstrated that the JDE model performed the best detection and tracking ability with the SDE and TransTrack models under the four scenarios, including the dense day, sparse day, dense night, and sparse night. The finding can also provide an effective and accurate detection for the rapid tracking of group-housed pigs in complex farming scenes.

Keywords: object detection; object tracking; joint detection and tracking; data association; group-housed pigs