

基于图像处理技术的动物行为识别研究进展



■资料图片

□李林葳 宋鑫悦 张智盛 段永鹏 侯欣宇 籍海娜 宋艳波 李菊霞 李艳文 刘振宇

基于手工特征的动物行为识别

传统的动物行为识别方法主要依赖于手工设计特征和传统机器学习算法。典型的传统方法主要包括3个步骤:特征提取、特征分类和特征融合。特征提取旨在从原始图像中提取具有代表性的特征信息,以便后续分类和识别。特征分类是指将提取到的特征变换为适合传统机器学习算法处理的数据形式,并使用这些特征来训练分类器。常用的传统机器学习算法包括支持向量机(Support Vector Machine, SVM)、K 最邻近法(K-Nearest Neighbor, KNN)等。特征融合旨在综合利用图像的多维特征,提高模型的鲁棒性和准确性,将不同来源的特征信息结合起来,实现优势互补,从而提升机器学习和模式识别任务的性能。

1、特征提取

方向梯度直方图(Histogram of Oriented Gradient, HOG):HOG 特征通过计算图像中各个局部区域的梯度方向,并对这些方向信息进行直方图统计,从而描述图像中物体的轮廓等特征,适用于检测具有复杂形状或纹理的物体。Jaddoa 等提出了一种基于 HOG 和 SVM 的牛脸多视角检测方法,该方法灵敏度为 0.9780,精密度可达 0.7212。白云翔通过斑马鱼姿态图片中的 HOG 特征进行了分类,个体识别正确率达到 60%,追踪正确率可达到 95%以上。宋琳等提出了一种融合 HOG 和 SIFT 特征的方法检测动物姿态,该方法可以忽略动物细微肢体动作对检测结果的影响,行为识别准确率为 93.1%,识别效果显著好于传统的单一特征算法,识别率较未加权特征融合方法提高了 1.4%。

局部二进制模式(Local Binary Pattern, LBP):LBP 特征能清晰描绘图片局部纹理,具有旋转不变性和灰度不变性,对光照变化具有较高鲁棒性,计算便捷,分类效果较好,在机器视觉领域中被广泛应用。

张满园等将改进的 LBP 纹理特征和 HOG 特征融合,对奶牛头部进行快速精准识别,该方法识别正确率超过 99%,可成功应用到奶牛识别领域。宋瑞金等提出了一种鱼群摄食行为检测算法,该算法对捕食和非捕食鱼群图像数据进行滤波和 LBP 特征提取,通过使用 BP 神经网络模型进行训练,该算法对鱼类摄食行为的识别准确率达 97.23%。

尺度不变特征变换(Scale-invariant Feature Transform, SIFT):SIFT 特征是局部关键点检测器和描述符,主要用于检测和描述图像中尺度不变性的关键点,在目标识别中应用广泛。

Tharwat 等采用 SIFT 特征处理二维斑马图像,基于 SVM、决策树(Decision Tree, DT)算法和模糊 k 最近邻(Fuzzy k-Nearest Neighbor, Fk-NN)分类器计算训练图像和测试图像之间的匹配度,该方法下 SVM 分类器的准确率在 99.30%左右,而 DT 和 Fk-NN 分类器的准确率分别达到 96.4%和 97.9%。Awad 等提出了一种基于局部不变特征的牛只行为识别方法,通过使用 SIFT 进行图像特征点匹配,结合随机抽样一致(Random Sample Consensus, RANSAC)算法提高了该识别方法的鲁棒性,平均识别准确率达 93.3%。

HOG、LBP 和 SIFT 特征提取方法在动物行为识别中都有广泛应用,HOG 特征可以捕捉动物的运动方向和速度,LBP 特征可用于识别动物的纹理模式,SIFT 特征则在动物运动、姿态和局部特征进行精确匹配时效果显著,3 种特征提取方法特点见表 1。

2、特征融合

根据融合时间分类:根据融合时间,特征融合可分为早融合和晚融合 2 类。早融合是特征级的融合,包括连接和相加 2 种方式。连接操作可直接

将 2 个特征拼接以完成信息融合,而相加操作是将 2 个向量组合成复合向量,实现特征融合。肖文卿等发现,融合脑电信号及与癫痫特征波相关的小波系数的线性和非线性特征,并使用 SVM 分类器,能有效实现小鼠癫痫脑电的自动分类;融合各相关小波系数的标准差和样本熵,分类正确率分别达到 86.60%和 88.60%;融合全部相关小波系数的线性和非线性特征后,分类正确率为 99.80%。

晚融合是指在预测分数上进行融合,即首先对多个模型进行训练,之后对训练后所有模型的结果进行融合以得到最终预测结果。李亿杨等将颜色特征与目标轮廓形心特征融合,基于粒子滤波算法实现生猪采食行为跟踪,并对得到的目标跟踪矩形框中心坐标进行二次修订,提高了目标生猪跟踪的可靠性和鲁棒性,有效跟踪平均精度为 93.4%。张雯雯等设计了一种基于堆叠沙漏网络特征融合和单个沙漏模块特征融合的多级特征融合方法,充分提取和融合动物姿态特征信息,该模型估计性能优于 Syn、BDL、Cy-CADA 和 CG-SSL 模型。

根据模型结构角度分类:根据模型结构,特征融合可分为串行、并行 2 种。串行策略下,模型只有一条支路,特征融合按照一定的顺序进行的,一个模型的输出作为下一个模型的输入,该策略常通过堆叠模型实现。堆叠模型中,Level 0 可包含 1 个或多个强学习器,而 Level 1 只能包含 1 个学习器。并行策略指模型有多个支路,且每个支路处理不同的特征,可同时对特征进行处理和融合,通过集成模型来实现。Bagging 结构为集成模型的实现方式之一,能从原始训练数据集中有放回地进行自助采样,构建出多个不同的训练集,可将每个训练集用于训练独立的基础模型。

早融合将不同特征在底层或输入层合并成一个特征向量,有助于完整地保留信息和降低计算复杂性,适用于具有强关联性或资源有限的情况。晚融合将不同特征分别输入处理单元,然后在更高层次上进行独立学习和融合,适用于处理关联性不强或需要更大自由度的情况。串行策略通常在特征有依赖关系的情况下使用,需要先提取、选择、转换特征,然后再进行融合。并行策略更适合在计算资源充足的情况下,能同时处理多个特征,提高计算效率。2 种特征融合方法的特点见表 2。

3、特征分类

KNN 算法:KNN 算法应用广泛,该算法 3 要素为 K 值的选取、距离度量的方式和分类决策规则。李光昌等运用 KNN 对猪只的站立、卧卧、犬坐行为进行分类,准确率分别达 94.2%、95.7%以及 94.6%,相较于采用 DeepCut、Associative Embedding 和 DeeperCut 姿态估计算法,平均识别准确率提高 4%。在样本分布不平衡时,KNN 识别准确率偏低。Tian 等通过融合 KNN 和随机森林(Random Forest, RF)算法对奶牛行为进行识别,并使用不平衡样本集进行训练和测试,KNN-RF 融合模型平均识别准确率为 96.87%,超过了 KNN 模型(95.37%)和 LVQ 模型(80.81%)的平均识别准确率。

SVM 算法:SVM 算法可分成 2 部分:找到 1 个由训练集的特定确定的决策面,执行 2 个类之间的模式识别,称为支持向量;对支持向量进行分类的机器学习算法。将多种核函数分别加入分类器后可训练出更为精准的分类模型。

何炜婷等研究了一种基于深度学习和 SVM 的动物姿态行为估计与心情识别方法,该方法能更加准确地对动物兴奋、害怕、生气等心情以及奔跑、慢走、静止等进行识别。赵珊珊提出一种边缘检测的 SVM 方法,对梅花鹿异常取食、反刍、卧息、移动等行为进行识别分析,最终识别准确率为

80%—90%。

DT 算法:DT 算法是一种树形结构的监督学习算法,在构建过程中主要分 2 步执行:先进行特征选择,找到对分类或回归任务最具区分能力的特征;然后根据特征选择的结果将数据集划分成子集,从而递归构建决策树。DT 算法在每个节点处选择最佳特征进行划分并生成子节点,当达到预定深度、节点中的样本数小于阈值或无法继续降低不纯度等条件时,停止生成子节点。

Tamura 等将 38 头荷斯坦奶牛放养在 4 个不同农场(A、B、C、D),通过 DT 算法学习和计算牛的活动变化,对奶牛的进食、反刍和卧躺行为进行分类,最终 A、B 农场奶牛行为的识别率分别为 99.2%、100%。Khanh 等通过 DT 算法对牛躺卧、站立和进食行为进行分类,平均识别准确率达到 82.24%。

K-Means 算法:陈紫城采用 HOG 描述生猪行为,随后使用 K-Means 算法对这些行为进行聚类,实现了特征提取并降低了数据复杂性。肖夏伟引入 K-Means++算法调整 Anchor 框,解决了猪体边缘识别不清晰的问题,为后续识别猪只饮水行为提供了基础。庄晓霖使用 K-Means 将家禽从图像背景中准确地分离出来,以便于后续的行为分析和异常行为检测。

传统的机器学习分类算法可以分为监督学习算法和无监督学习算法。典型的监督学习方法包括 SVM、DT、贝叶斯网络、神经网络等。监督学习方法需要手动标注动物行为来获取训练数据。无监督学习算法亦可探索动物行为的数据结构和模式,可揭示新行为特征,常用到的无监督学习包括 K-Means 等算法。有关无监督学习方法的研究较少,且该方法通常不直接用作分类器,而是结合其他算法用于减少数据的复杂性和进行图像分割。无监督学习方法未来以提高行为识别的综合性能为目标,发展方向包括半监督学习和多模态数据融合。不同特征分类方法的特点见表 3。

基于特征和机器学习方法的动物行为识别具有可解释性强、计算效率高、小数据集可用、领域知识融合便捷等优点,适用于特定问题和资源受限的情况。然而,这些方法常需要手工设计特征,泛化能力有限,难以处理复杂的时序和空间一时间关系,因此在大规模数据和自动特征学习等方面存在局限性。

基于深度学习的动物行为识别方法

近年来,随着深度学习技术的飞速发展,各种深度学习架构也应运而生,并展现出强大的表征能力和优越的性能。当前主要的学习框架包括双流 2D 卷积神经网络(Convolutional Neural Network, CNN)、循环神经网络(Recurrent Neural Network, RNN)、3D CNN、2D CNN 以及 Transformer 等。

1、双流 2D CNN 架构

该架构有 2 个独立的 CNN 以识别空间流和时间流,在网络尾部用 Softmax 融合在一起。该方法首先计算视频中每 2 帧的密集光流,从而获取包含时间信息的密集光流序列(Temporal);然后分别训练处理空间信息和处理时序信息的 2 个 CNN 模型,并对 2 个网络的训练结果进行直接融合后,得到最终分类结果。

Zhang 等将视频中的图像帧和光流作为 Two-Stream 输入,提出基于深度学习的双流卷积网络模型,其特征提取网络为 ResNet101 的 TSN 模型,对猪只行为的识别率可达到 98.99%。Hao 等提出了一种深度互学增强型双流猪行为识别方法,该方法中 2 个分支网络通过协作学习,有效提取外观或运动特征,提高了猪行为的识别性能。

2、RNN 网络架构

CNN-LSTM 网络通过融合 CNN 和 RNN 架构中经典的短时记忆网络

(Long Short-Term Memory, LSTM)来提取视频的时间特征和空间特征,这种串联结构在早期应用广泛且识别精度较高。

Yin 等从奶牛视频帧中提取空间特征,结合双向长短时记忆,在时间序列中汇聚奶牛视频帧的空间特征,对奶牛躺卧、站立、行走、饮水和进食行为进行识别,最终识别准确率可达 95.20%。Qiao 等在牛视频数据集上使用 CNN 提取特征,并通过 LSTM 模型捕获时间信息,该方法在 15 帧和 20 帧视频中识别准确率分别为 88%和 91%。Chen 等基于 CNN 和 LSTM 的深度学习方法,可提高猪攻击事件的时空特征识别及准确性,同时利用最大熵分割、HSV 颜色空间变换和模板匹配等技术确定每头猪的身份和喂食时间。Wu 等基于 YOLO v3 深度学习算法和相对步长特征向量对跛行奶牛和非跛行奶牛进行分类,结果显示,基于 LSTM 的跛行检测准确率为 98.57%,比 SVM、KNN 和决策树分类器(Decision Tree Classifier, DTC)分别高 2.93%、3.88%和 9.25%。

3、3D CNN 网络架构

3D 卷积将多个连续帧堆叠成立方体,利用相邻帧之间的关联性和运动信息,对每个卷积层的特征图与上一层中的多个连续帧进行卷积操作,进而从连续帧的变化中学习到运动特征和时间序列。引入 3D 卷积后,CNN 在视频处理任务中能够更有效地提取时空特征,提高视频分析、行为识别、动作检测等性能。

Li 等提出了一种基于奶牛骨架和稀疏采样的混合卷积算法(Hybrid CNN),该算法通过在每个三维卷积后串联一个深层二维卷积,增加了三维 CNN 的深度,提升了网络的性能,在奶牛的行走、站立和躺卧行为的识别中取得了优异结果,经过 5 次交叉验证,该方法的最终分类准确率为 91.80%,比混合 3D/2D 卷积管高 3.40%。针对目前已有模型无法处理大参数的问题,Wang 等提出了一种轻量化的 E3D 模型,该模型通过引入高效通道注意力来过滤通道信息并提高准确性,能快速准确地识别奶牛躺卧、站立、行走、饮水和进食行为。Qiao 等提出了一种结合了 C3D 网络和卷积长短时记忆的深度学习框架(C3D-ConvLSTM),该框架对奶牛进食、探索、梳理、行走和站立行为的识别准确率达到了 86.67%。

4、2D CNN 网络架构

2D CNN 被最早应用于动物行为识别中,常见的网络模型有 YOLO 和 ResNet 残差网络。YOLO 系列算法是一种基于单阶段的目标检测方法、深度 CNN 且快速准确的目标检测算法,其核心思想是将目标检测任务视为一个回归问题,在图像中仅进行一次前

动物行为识别旨在理解动物的行为,并对每种行为贴上类别标签,具有广泛的应用前景。近年来,动物行为识别在计算机视觉领域受到了越来越多地关注。动物行为识别技术的发展经历了从传统的基于手工特征的方法向基于深度学习的方法演变。本文根据动物行为识别技术发展的基本脉络,综述了传统方法和深度学习方法中的主要流派和技术手段,概述了各种技术手段优缺点及其在动物行为研究中的应用情况。

向传递即可同时预测目标的位置和类别,具有高实时性和低计算成本的优势。杜妍茹提出 YOLO v5-Deepsort 算法,利用该算法对多目标牛身份进行追踪并识别牛的站立、卧卧、采食和饮水行为,平均识别准确率为 35.2%。徐文凯在 YOLO v5 算法的基础上提出了 SO-YOLO v5 和 CA-YOLO v5 2 种模型,实现了对海参的小目标识别和多目标识别。Tu 等提出 YOLO v5S 和 YOLOX-S-DeepSORT 算法,实现了对猪只躺卧、进食、站立和其他行为的追踪识别;与经典 DeepSORT 算法相比,YOLOX-S-DeepSORT 算法多目标跟踪准确率和 IDF1 分别提高了 1.8%和 6.8%,ID 切换次数下降了 80%。张宏鸣等使用 YOLO v3 识别肉牛进食行为,平均精确度为 79.7%、平均召回率为 73.0%、平均准确率为 74.3%。

ResNet 残差网络在动物行为识别任务中同样应用广泛。翟惠莹利用深度相机采集视频数据,制作小鼠静止、行走、蜷缩、有支撑直立和无支撑直立 5 种行为的 RGBD 图库,发现 Res Net50 和 Mobile Net V2 深度学习框架,在 RGB 数据集上准确率分别达 91.2%、88.6%,在 Depth 数据集上准确率分别达到 90.8%、87.6%。李怡使用 AlexNet 和 ResNet50 网络对奶山羊的站立、蜷缩、卧爬、进食 4 类行为进行识别和监测,得到的最佳识别结果为 98.52%。欧士路提出一种基于注意力机制的时间偏移模块(Temporal Shift Module, TSM),将 ResNet-50 作为主干提取网络,对湖羊反刍、进食、奔跑 3 种行为进行准确识别,最终的准确率为 97.57%。此外,史怀等使用 SSD 算法对猪个体的进食、排泄行为进行识别,最终的识别准确率为 88.23%。

5、Transformer 网络架构

Transformer 模型的核心是自注意力机制,在处理序列数据时能够同时考虑序列中的所有位置。与传统的循环 RNN 或 CNN 不同,Transformer 没有固定的处理顺序。通过添加位置编码到输入嵌入中表示相对位置,可以确保模型在处理序列数据时考虑到每个元素的位置信息。动物骨骼关键点检测是动物行为和姿态检测的前提。张飞宇等通过在 HRNet 网络特征提取层中加入改进的 Transformer 编码器,并引入多尺度信息融合模块提高模型对于不同维度的特征学习,最终在羊和东北虎的关键点数据集上,分别取得 77.1%和 89.7%的准确率。

CNN 作为最早出现的网络架构,被广泛应用于静态图像的纹理、形状等视觉特征的识别,并可协助识别不同的动物行为。然而,CNN 主要关注空间信息,难以捕捉时间序列信息,在处理动态行为时可能存在局限性。双流网络和 3D CNN 网络架构捕捉时空特征需要大量计算资源和训练数据,在

资源受限的情况下效果不佳。Transformer 最初是为自然语言处理设计的,后被引入到机器视觉领域且效果显著。未来可进一步研究适用于视频序列的 Transformer 变体以及其他更好地捕捉动态信息的方法。

相较于传统方法,深度学习方法可以实现对动物行为的细粒度识别,在饮水和进食等相似行为识别时,深度学习方法表现更加优异。另外,由于深度学习算法通常采用层级结构,可以提取更深层次的抽象特征,因此识别效果较传统方法更好。然而,深度学习算法往往需要大量参数,这增加了计算资源的开销,在某些特定领域,传统方法表现更加优异。

小结与展望

基于图像处理的动物行为识别是一个具有高挑战性、高价值的研究方向。传统的基于人工特征的机器学习方法通过特征的提取、融合和分类来实现行为识别。其中,特征提取包括 HOG、LBP、SIFT 等特征,特征融合包括早晚融合、串行融合等策略,特征分类又通过 SVM、决策树等分类方法。这些方法在一定程度上取得了较好效果,但准确率和鲁棒性仍有待提高。深度学习方法在近年来得到了广泛关注,不断涌现出新的算法,可有效提高动物行为识别的准确率和鲁棒性,但需要更多的数据集和更强的计算能力。

基于图像处理的动物行为识别未来可以从以下几个方面深入研究:1、鲁棒性方面,动物姿态遮挡、低分辨率图像造成的噪声、细节信息丢失等问题上的识别准确率还有较大提升空间;2、应用场景方面,大场景下小目标动物的行为识别,以及动物细粒度动作识别的数据集标注、模型设计、参数调整等研究还有待深入;3、泛化能力方面,不同模式的动物行为具有巨大差异,目前行为识别多为集中于人和猪、牛、羊等特定种类的专一识别,存在一定局限性,未来可开发能处理不同物种种类、各种行为类型的通用算法和模型,并进一步提升算法的泛化能力和鲁棒性;4、多模态融合方面,应专注于结合空间信息和时序信息,将动物的低级动作特征与高级行为模式层次信息融合,研究多种类型传感器信息的整合与识别。总之,基于图像处理的动物行为识别在未来将继续得到更为广泛的关注,并伴随着技术的不断改进和完善,在智能养殖研究与实践中广泛应用。

(李林葳 宋鑫悦 张智盛 段永鹏 籍海娜 李菊霞 李艳文单位:山西农业大学信息科学与工程学院;侯欣宇单位:海南大学信息与通信工程学院;宋艳波单位:山西农业大学大学生命科学学院;刘振宇单位:山西农业大学农业工程学院)

表 1 3 种特征提取方法特点					
特征	提取速度	常见应用领域	优点		缺点
HOG	较快	目标检测	对光照变化相对不敏感	在处理复杂纹理或形状变化较大的场景时效果变差	
LBP	较快	纹理分析	对纹理具有较好的鲁棒性	对于光照变化较为敏感,处理光照变化较大的图像时可能效果不佳	
SIFT	较慢	图像匹配	对尺度和旋转变化具有不变性	计算复杂,不适合实时应用;算法受专利限制,在商业应用中可能有一定局限性	

表 2 2 种特征融合方法的特点					
分类角度	融合方法	常见应用领域	融合速度	优点	缺点
根据融合先后时间角度分类	早融合	简单的任务	较快	可以充分利用底层特征的信息	会丢失一些高层次的语义信息,不够灵活,对于复杂任务效果有限
	晚融合	复杂任务的建模	较慢	可以更灵活地融合不同层次的特征	计算开销较大,对硬件要求高
根据模型结构角度分类	串行策略	复杂任务的建模	较慢	可以逐步提取和整合特征	计算开销可能较大,需要仔细调整参数以避免信息丢失
	并行策略	多模态任务	较快	充分利用多个来源的信息	需要管理好不同来源的特征,确保它们能够有效地融合

表 3 不同特征分类方法的特点					
特征分类方法	代表算法	分类速度	优点		缺点
有监督学习	SVM DT KNN	较快	有标签的数据可以帮助模型更好地泛化到新数据	获取大量标签数据的经济和实践成本较高,在面对未标记数据时性能受限	
无监督学习	K-Means PCA 层次聚类	较慢	在处理大规模数据时更具可扩展性	模型可能对噪声敏感;在没有标签的情况下,较难评估和解释模型性能	