

使用姿态估计和双向长短时记忆网络进行奶牛跛行检测

Helena Russello ^{a,*}, Rik van der Tol^a, Eldert J. van Henten^a, Gert Kootstra ^{a,*}

^a*Agricultural Biosystems Engineering group, Wageningen University & Research, Wageningen, The Netherlands*

Abstract

本研究提出了一种将姿态估计与双向长短期记忆 (BLSTM) 神经网络相结合的跛行检测方法。结合姿态估计和 BLSTM 分类器具有以下优点：无标记的姿态估计，通过从关键点轨迹中学习时间运动特征来消除手动特征工程，以及能够处理短序列和小型训练数据集。采用 T-LEAP 姿态估计模型从行走牛的视频中提取了九个关键点（位于牛的蹄子、头部和背部）的运动序列。然后，关键点的轨迹被用作 BLSTM 分类器的输入，该分类器经过训练可进行二分类跛行分类。我们的方法显著优于依赖手动设计的运动特征的已有方法：我们最佳的架构达到了 85 % 的分类准确率，而基于特征的方法的准确率为 80 %。此外，我们还证明了我们的 BLSTM 分类器仅用一秒钟的视频数据即可检测到跛行现象。

Keywords: lameness, cows, locomotion, pose-estimation, deep-learning, lstm

1. 介绍

跛足是奶牛常见的状况，其特点是由于蹄或肢体的病变导致步态异常。跛足对福利和经济有负面影响，因为它影响奶牛的产奶量、繁殖能力和生活质量 (Enting et al., 1997; Whay and Shearer, 2017)。跛足通常是在视觉步态评分过程中被检测到，这些过程中，训练有素的观察员会评估牛群中

*Corresponding authors

Email addresses: `firstname.lastname@wur.nl` (Helena Russello),
`firstname.lastname@wur.nl` (Gert Kootstra)

Preprint submitted to Elsevier

August 15, 2025

跛足的流行情况，但这些过程耗时且执行频率不高 (O’leary et al., 2020)。自动化步态评分，例如通过持续相机监控，可以实现跛足的早期检测，从而及时进行治疗。

视频摄像机是用于自动运动得分的有吸引力的传感器，因为它们相对便宜，不具侵入性，而且在大群体中具有良好的扩展性。从视频中自动检测跛行通常通过使用临床步态评分方法中的运动特征 (Sprecher et al., 1997) 来对奶牛的跛行程度进行分类。例如，Poursaberi et al. (2010); Viazzi et al. (2014); Van Hertem et al. (2018) 的研究使用背部姿势，Song et al. (2008) 使用追踪距离，Wu et al. (2020); Zheng et al. (2023) 使用步长，Kang et al. (2020) 使用支撑阶段来分类跛行。其他研究，如 Zhao et al. (2023); Barney et al. (2023); Taghavi et al. (2024); Russello et al. (2024) 将多种运动特征结合作为跛行分类的输入特征。将背部姿势、头部位置、追踪距离和步幅结合起来，比使用单一特征可以提高分类性能 (Russello et al., 2024)。尽管使用运动特征的模型在检测跛行方面表现出令人鼓舞的结果，但它们仍然有一些限制。例如，这些运动特征假定数据噪声较低，在高噪声数据下可能表现不佳 Taghavi et al. (2023)。此外，手动选择运动特征作为特征可能会限制提供给这些模型的信息，而仅使用人为设计的特征可能无法捕捉复杂的模式。

为了克服这些限制，一些研究使用深度神经网络来进行特征提取和跛行分类。这些方法大多包含两到三个步骤：首先从视频帧中分离出动物的身体结构，然后进行特征提取和跛行分类（后两者通常合并为一步）。例如，Karoui et al. (2021) 跟踪放置在牛腿上的粘附性物理标记，并训练一个卷积神经网络 (CNN) 以根据这些标记的运动预测跛行。Wu et al. (2020) 使用 YOLOv3 (Farhadi and Redmon, 2018) 提取腿部坐标以创建步长特征。然后在长短期记忆神经网络 (LSTM) 和双向 LSTM (BLSTM) 中使用步长矢量序列来检测跛行。Arazo et al. (2022) 应用分割模型及 SlowFast 视频识别模型 (Feichtenhofer et al., 2019)，将基于牛的身体轮廓时间序列的跛行分类。Jiang et al. (2020) 提取光流图序列，捕捉视频帧之间的运动，并在 BLSTM 网络中使用它来进行跛行分类。Arazo et al. (2022) 和 Jiang et al. (2020) 表明，他们的模型只使用 RGB 视频数据即可有效分类跛行（一种一步法）。然而，他们发现当使用结合分割掩膜 (Arazo et al., 2022) 或光流图 (Jiang et al., 2020) 的两步法以引导模型关注动物的身体结构时，分类精度有所提高。

即使使用深度学习模型减少了广泛的预处理和人工特征工程的需求，同时提高了鲁棒性，上段中讨论的多步方法仍然存在固有限制。仅腿部跟踪 (Wu et al., 2020; Karoui et al., 2021) 可能忽略其他身体部位的有价值的运动模式，而全身分析 (Jiang et al., 2020; Arazo et al., 2022) 在跟踪

关键解剖特征方面缺乏特异性，并且由于其高维性通常需要更多的训练数据，这使得学习过程更加困难。

为了克服这些限制，我们提出了一种结合姿势估计和 BLSTM 模型优势的两步方法 (Graves and Schmidhuber, 2005)。我们使用 T-LEAP 姿势估计模型 (Russello et al., 2021)，该模型允许跟踪多个关键点，包括腿部关节、头部位置和脊柱运动，从而为分类模型提供来自生物力学相关关键点的紧凑却全面的运动数据。在 Russello et al. (2024) 的工作基础上，该工作展示了关键点轨迹在计算手工制作运动特征方面的有效性，我们使用 BLSTM 模型直接从关键点轨迹序列中学习运动特征并分类跛行。BLSTM 是一种神经网络，擅长于捕捉序列数据中的长期关系和建模复杂的时间模式。BLSTMs 已成功应用于多个领域，包括人体运动分析 (Du et al., 2019; Battistone and Petrosino, 2019)，以及跛行检测 (Jiang et al., 2020; Wu et al., 2020)。在这些应用领域 (Du et al., 2019; Jiang et al., 2020; Wu et al., 2020) 中，BLSTMs 一直优于常规 (单向) LSTMs。因此，我们使用 BLSTMs 根据行走奶牛的关键点轨迹检测跛行。

总之，我们提出的方法具有以下优点：姿态估计为来自多个身体部位的时空运动提供了一种紧凑的数据表示，而 BLSTM 模型消除了对噪声过滤以及手动特征工程的需要。我们的贡献如下：

1. 我们提出了一种使用 BLSTM 神经网络的两步方法，用于检测九个关键点的时间序列数据中的跛行。我们的研究探讨了通过学习的运动特征检测跛行是否优于先前建立的手工设计的运动特征。为了确保公平评估，我们使用相同的数据集直接与 Russello et al. (2024) 进行了比较。
2. 我们评估了我们的方法在多个 BLSTM 架构中的性能 (不同层数和层大小)。此外，我们评估了 BLSTM 分类器在不同序列长度 (30、60 和 90 帧，相当于在 30 FPS 下的视频中的 1、2 和 3 秒) 时的性能，因为健康和跛脚的奶牛倾向于以不同的速度行走 (Flower et al., 2005)。
3. 为了促进可重复性并推动未来的研究，我们将在发表后公开我们的数据和代码。

2. 材料和方法

我们的方法，如图 1 所示，使用 T-LEAP 从行走的奶牛视频中提取关键点轨迹。关键点轨迹被裁剪为固定的序列长度，并传递给 BLSTM 分类模型。数据和方法在以下的子章节中进一步描述。

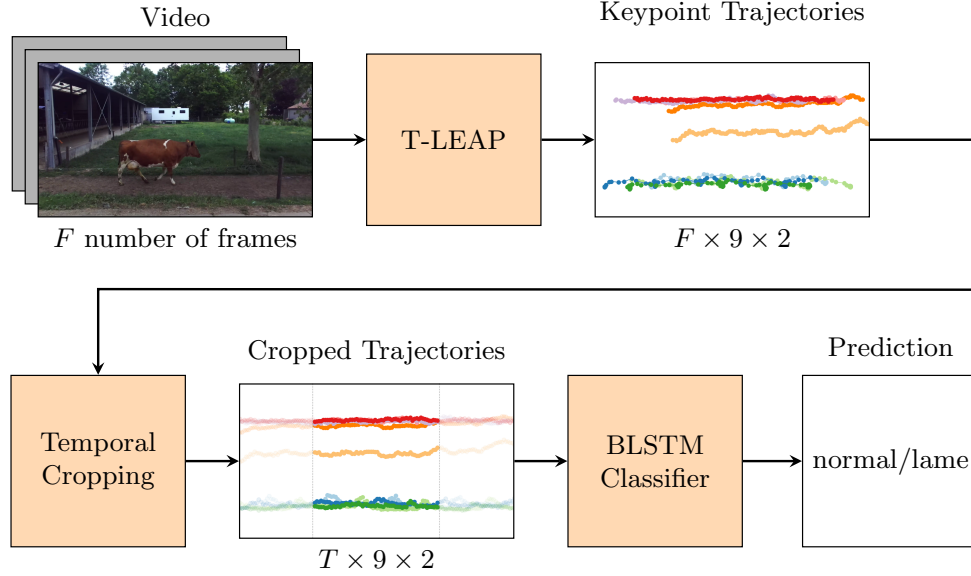


Figure 1: 我们跛行检测方法的概要。关键点轨迹通过 T-LEAP 从视频中提取（图 3 和 4）。然后，关键点轨迹被裁剪到一个固定的序列长度 T ，并传递给 BLSTM 分类器。

2.1. 数据集

实验使用了来自数据集 Russello et al. (2024) 中的关键点轨迹，该数据集包括 272 个视频，记录了 98 头荷斯坦-弗里斯牛。使用 ZED 相机¹ 在户外走道上自由行走时从侧面拍摄这些牛。这些视频以每秒 30 帧的速度录制，长度从 90 到 207 帧不等（平均长度：134 帧）。

来自 Russello et al. (2024) 的数据集包含从 272 个视频中提取的关键点轨迹。对于每个视频，四名观察者使用 Sprecher et al. (1997) 量表对步态进行评分。步态评分被合并为二进制跛行标签（正常/跛行）。在 272 个关键点轨迹中，143 个被标记为正常，129 个被标记为跛行，形成一个相对平衡的数据集。

使用在 Russello et al. (2021) 中介绍的 T-LEAP 姿态估计模型，关键点轨迹被自动提取。简而言之，T-LEAP 是一种姿态估计模型，经过在行走奶牛视频上的训练，用于追踪奶牛的蹄子、头部和背部的九个解剖标志（关键点）（见图 2）。关键点轨迹，如图 3 和 4 所示，表示通过视频的每一帧中的这九个关键点在二维图像平面上的运动。图 3 展示了正常步态的

¹<https://www.stereolabs.com/zed>

关键点轨迹例子，而图 4 则展示了跛足步态的例子。有关数据集的详细描述，请参考 Russello et al. (2024)。

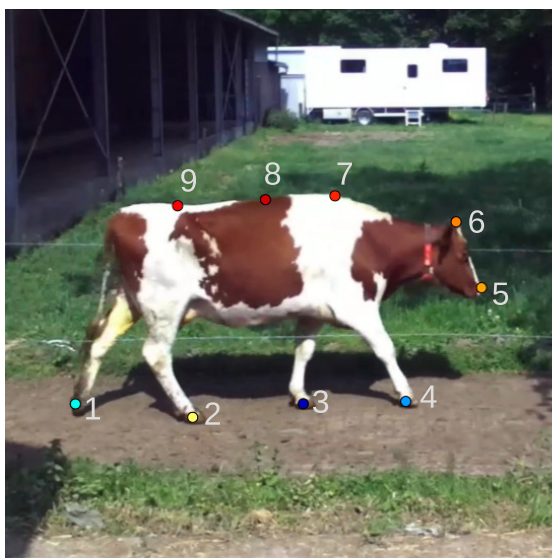


Figure 2: 如 Russello et al. (2024) 中所述的 9 个关键点。关键点命名如下：1：左后蹄，2：右后蹄，3：左前蹄，4：右前蹄，5：鼻子，6：前额，7：马肩隆，8：尾端胸椎，9：骶骨。

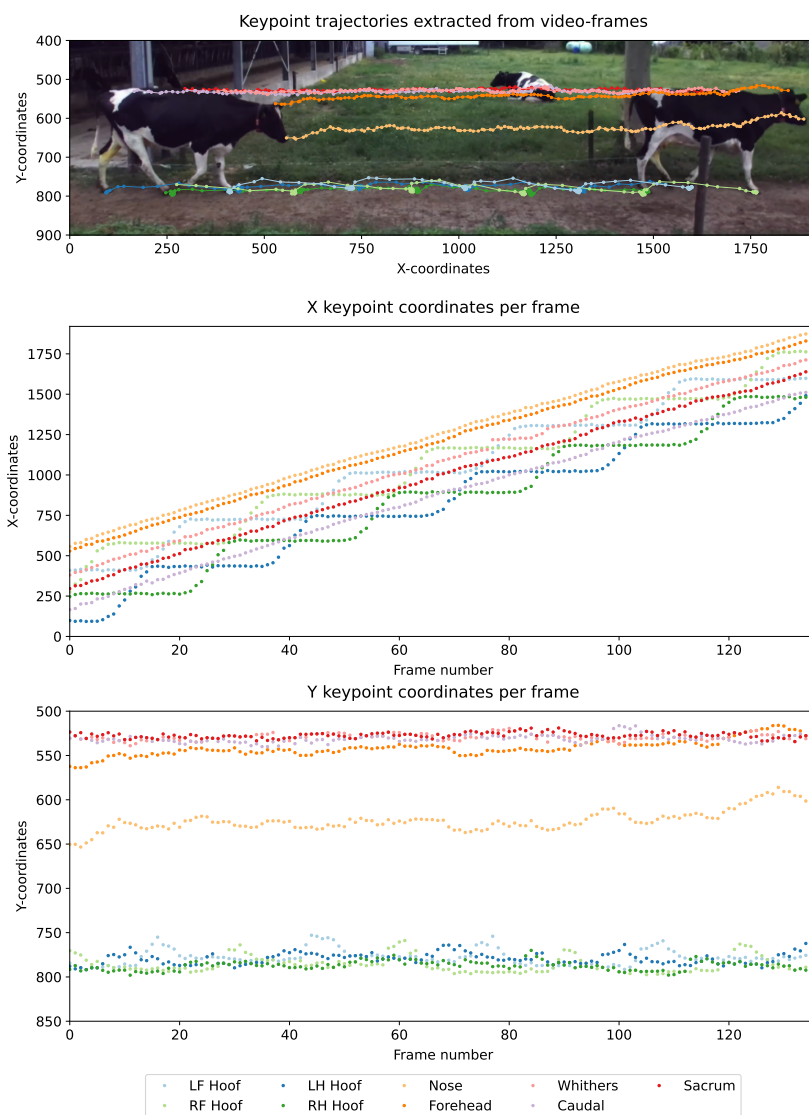


Figure 3: 使用 T-LEAP 从一段展示正常步态的奶牛视频中提取的关键点轨迹示例。顶部图形增添了视频的首帧和末帧，并显示了从所有视频帧中提取的每个关键点的 (x, y) 坐标。中间和底部图形分别显示每帧的关键点的 x 坐标和 y 坐标。

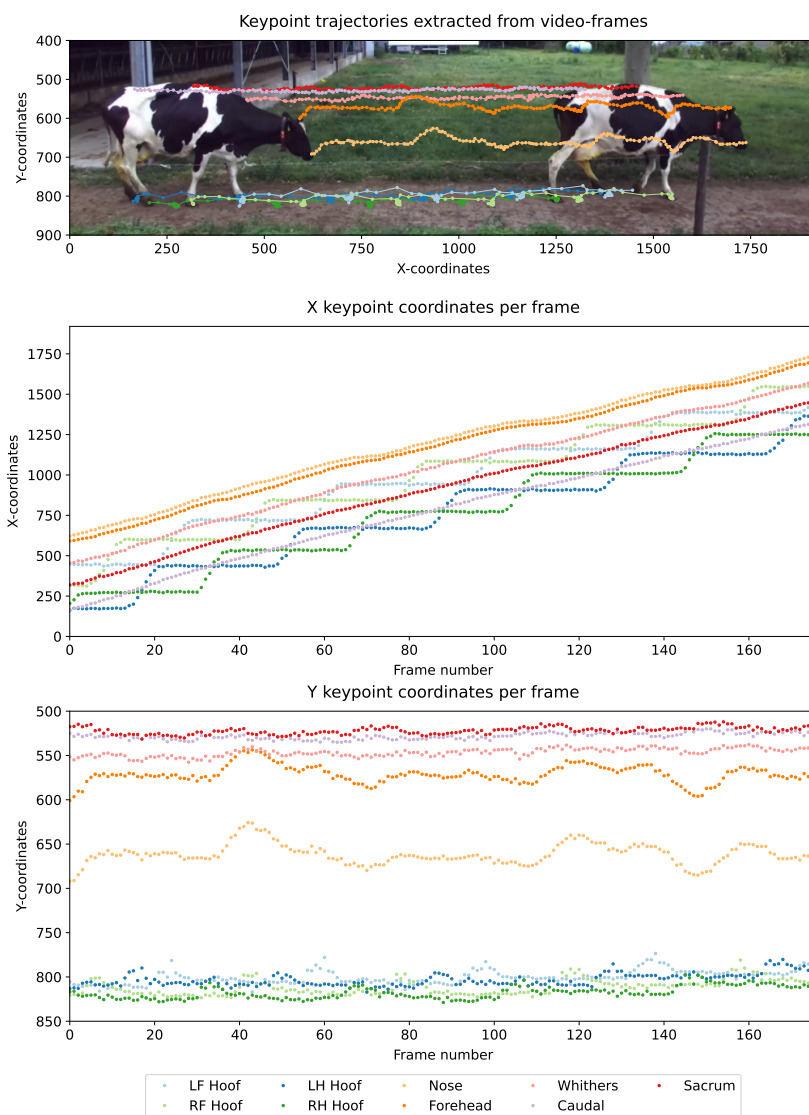


Figure 4: 通过 T-LEAP 从一段展示严重跛行步态的奶牛视频中提取的关键点轨迹示例。
 上图补充了视频的第一帧和最后一帧，显示了从所有视频帧中提取的每个关键点
 的 (x, y) 坐标。中间和底部图分别显示了每帧关键点的 x 坐标和 y 坐标。

2.2. 数据增强

在训练集上通过如下所述的时间裁剪和随机抖动进行了数据增强。对于测试集，没有应用随机抖动，并从关键点轨迹的中间部分提取了长度为 30、60 和 90 帧的固定序列。

随机时间裁剪. 由于关键点轨迹的帧数在 90 到 207 帧之间变化，而当前的实验需要固定序列长度为 $T \in \{30, 60, 90\}$ 帧，因此对数据应用了随机裁剪，以允许固定序列长度，同时增加数据量。也就是说，假设有一个包含 F 帧的关键点轨迹，起始帧是从 $[0, F - T]$ 的离散均匀分布中随机选择的。这种方法确保模型在每个训练时期遇到关键点轨迹的不同片段。

类似于 Karoui et al. (2021)，我们通过添加随机高斯噪声来增强关键点坐标，从而提高模型的鲁棒性。具体来说，我们对每个关键点坐标应用抖动，方法是 从一个均值设置为原始坐标值且标准差定义为头长（以头部与鼻子关键点的欧几里得距离测量）百分之一的正态分布中随机采样。

2.3. 使用双向长短时记忆网络检测跛行

在这项研究中，我们使用了一个双向长短期记忆 (BLSTM) 分类模型 (Graves and Schmidhuber, 2005) 来基于行走牛的关键点轨迹检测跛行。LSTM (Hochreiter and Schmidhuber, 1997) 是一种递归神经网络 (RNN)，擅长捕捉序列数据中的长期关系和学习复杂的时间模式。顾名思义，RNN 使用递归连接，即，一个神经元在某一时间步的输出反馈回到下一个时间步作为输入。这允许捕捉序列中的时间依赖关系和模式。LSTM 是一种 RNN，旨在克服 RNN 的局限性，并允许在序列数据中捕捉长期依赖关系 (Hochreiter and Schmidhuber, 1997)。一个 LSTM 单元通常由一个通过三个门控控制的单元状态组成：一个遗忘门、一个输入门和一个输出门。遗忘门控制要从之前的状态中丢弃哪些信息，输入门控制要在单元状态中存储哪些新信息，输出门控制要输出单元状态中的哪些信息。在每个时间步，LSTM 接收当前输入和之前的隐藏状态，然后使用这些门选择性地更新其记忆并产生输出。这种门控机制允许 LSTM 在序列中学习长期依赖关系 - 它们能够记住很多时间步前的重要信息，同时忘记不相关的细节 (Gers et al., 2002)。这使得它们在分析步态模式时非常有用 (Lefebvre et al., 2013; Battistone and Petrosino, 2019)，因为动物姿态序列变化与跛行存在的复杂关系是复杂的。

单向和双向 LSTM 之间的主要区别在于它们如何利用输入序列中的时间信息。单向 LSTM 以单一方向处理输入序列，可以从序列的开始到结束（前向方向）或者从序列的结束到开始（后向方向）。前向 LSTM 只包含来自过去的信息来预测当前时间步，而后向 LSTM 则包含来自未来的信息来预测当前时间步。顾名思义，双向 LSTM 结合了一个前向 LSTM 和一

个后向 LSTM，两者都连接到同一个输出层。这意味着对于序列中的任何一个时间步，BLSTM 利用该时间步之前和之后的点的序列信息 (Graves and Schmidhuber, 2005)。这种情况在未来特征提供给当前时间步的信息与过去特征不同的情况下很有用，比如在步态这样的不对称信号中，这是预期的。

2.4. BLSTM 模型架构

图 5 所示的提议 BLSTM 分类模型接收关键点序列作为输入，并将该序列分类为正常或跛行。其模型架构由两部分组成：一个 BLSTM 神经网络，接着是一个全连接网络 (FCN) (图 5)。我们通过经验选择了不同的 BLSTM 架构，由两层或三层包含 128 或 256 个隐藏单元的层组成 (h)。图 5 中展示的 BLSTM 神经网络由两个 BLSTM 层组成，而我们三层架构中有一个额外的 BLSTM 层，但 FCN 保持不变。

输入 (x) 由形状为 $T \times 18$ 的序列组成，其中 T 对应于序列的长度，18 对应于包含 9 关键点的二维图像坐标的平坦化向量。输入被顺序传递给 BLSTM，也就是说，从 x_0 到 x_T 。一个 BLSTM 层结合了一个前向和后向的 LSTM，每个包含 h 个隐藏单元。前向 LSTM 在前向方向处理输入序列 (从序列的第一个项目 x_0 到最后一个项目 x_T)，后向 LSTM 在后向方向处理输入序列 (从 x_T 到 x_0)。前向和后向 LSTM 的输出被连接并传递到下一个 BLSTM 层。最后一个 BLSTM 层的输出被连接并传递给分类层 (FCN)，该层由两个输入大小为 $2h$ 和 h 的全连接层组成。分类层输出一个未归一化的预测值 ($\hat{o}$)。在推理过程中，使用一个 S 形函数 ($\hat{y} = \sigma(\hat{o})$) 将输出规范化到 0 和 1 之间，表示跛脚的概率。如果 $\hat{y} \geq 0.5$ ，则将奶牛分类为跛脚。

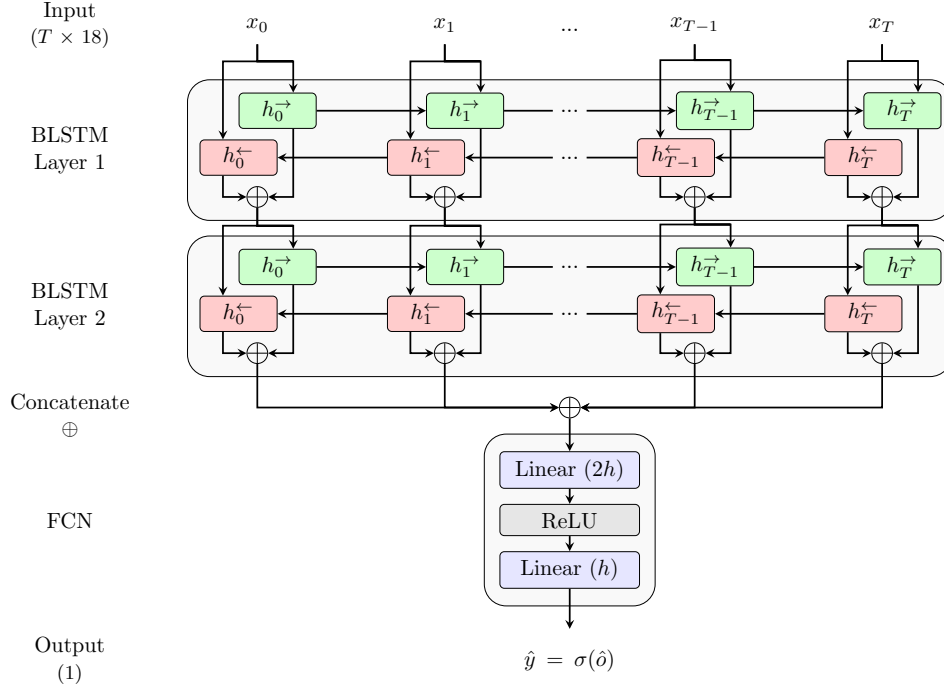


Figure 5: 2 层 BLSTM 分类器的架构，其中 T 对应序列长度 (30, 60 或 90)， h 对应隐藏单元的数量 (128 或 256)。绿色矩形表示前向 LSTM，红色矩形表示后向 LSTM。请注意，3 层 BLSTM 架构有一个额外的 BLSTM 层，但它们的 FCN 保持不变。

2.5. 训练过程

这些模型使用 PyTorch 深度学习框架 (版本 2.1.0) (Ansel et al., 2024) 实现。实验是在一台配备 AMD Ryzen 9 5900X CPU、NVIDIA GeForce RTX 4070 Ti-SUPER GPU 和 64 GB RAM 的计算机上运行的。

鉴于数据集的规模有限，为了增加结果的可靠性，我们遵循了与 Russello et al. (2024) 相同的程序，并通过分层 5 倍交叉验证进行分组，将数据集分为训练集和验证集。这些折叠是基于牛的 ID 进行分组，确保个别牛只出现在训练集或验证集中，而不是两个同时出现。为了在每个训练折叠中确保类别的均匀分布，我们使用了 PyTorch 的加权随机采样器 (Ansel et al., 2024)，通过少数类别过采样来解决类别不平衡问题。所有模型使用大小为 8 的批次进行了 100 个 epoch 的 5 倍交叉验证训练。我们使用 AdamW-AMSGrad 优化器 (Reddi et al., 2019; Loshchilov and Hutter, 2017)，学习率调度器每 50 个 epoch 将学习率减少一半。应用了梯度裁剪 (Pascanu et al., 2013)，阈值为 0.5，以防止梯度消失和爆炸，这是

LSTM (Bengio et al., 1994) 的常见问题。学习率、权重衰减和 dropout 超参数使用平铺交叉验证方法进行了调优。即，超参数首先通过 5 倍交叉验证的网格搜索在每个折叠上进行优化，然后使用最佳的超参数集合对相同的 5 倍交叉验证的模型进行重新训练。我们使用平铺交叉验证，因为它在计算上比嵌套交叉验证更为经济，并且通常会选择出与嵌套交叉验证选择的算法质量相似的算法 (Wainer and Cawley, 2021)。

2.6. 评估指标

模型的性能使用以下指标进行评估：准确率、宏观 F1-score、灵敏度和特异性。F1-score 采用宏观平均，也就是说，该指标先按类别（正常/跛行）计算，然后取平均值。评估指标在五折交叉验证中取平均值。不同模型之间的差异使用 McNemar 检验和二项式分布进行统计测试，设定 $p < 0.05$ 为显著性水平。

2.7. 实验

2.7.1. 不同的模型架构

我们首先探讨了不同的网络架构，分别使用两个和三个具有 128 和 256 个隐藏单元的 BLSTM 层，后续称为 2×128 、 3×128 、 2×256 和 3×256 。虽然这些架构是凭经验选择的，但它们从 Graves et al. (2013) 中获得了灵感，该作者提出了包括 2 层和 3 层 250 个隐藏单元的 BLSTM 架构用于语音识别。最初的实验使用了 $T = 90$ 帧的序列，因为这是数据集中最短视频的长度。

2.7.2. 与参考模型的比较

本研究旨在比较通过学习到的运动特征（建议的模型）与先前建立的手动设计运动特征（参考模型）在检测跛行方面的性能。然后，我们与 Russello et al. (2024) 进行了直接比较，并将我们的结果与他们的最佳模型进行了比较，该模型是使用相同数据集上的六个运动特征训练的支持向量机 (SVM) 分类器。

简而言之，Russello et al. (2024) 使用了相同的数据集进行关键点轨迹分析，并应用异常检测和过滤以纠正和平滑有噪声的关键点轨迹。这些关键点轨迹随后被用来计算六个运动特征，即背姿势、头部晃动、跟踪距离、步幅长度、站立时间和摆动时间。然后将这些运动特征用作输入特征来训练多种分类器，如决策树或支持向量机，以检测跛行。

与 Russello et al. (2024) 不同，BLSTM 输入数据未进行滤波，因为 BLSTM 通常可以处理某种程度的噪声。为了比较滤波的影响，我们用从滤波后的关键点轨迹计算出的运动特征以及从原始、未滤波的关键点轨迹计算出的运动特征训练了参考 SVM 分类器。

2.7.3. 不同的序列长度

步伐持续时间是检测跛行时需要考虑的重要因素，因为健康的奶牛往往步伐较快（每步 1.3 秒），而跛行的奶牛步伐较慢（每步 1.5 秒）（Flower et al., 2005）。

第一组实验使用了 90 帧（即三秒视频数据）的序列，因为这是数据集中最短的视频长度。我们通过对更短的序列进行评估，即序列长度为 30 帧和 60 帧（分别对应一秒和两秒的视频数据），进一步扩展了我们的分析。为此，我们在 30 帧和 60 帧的序列上重新训练和评估了表现最佳的 BLSTM 结构。

3. 结果

不同 BLSTM 架构和序列长度的结果，相较于在相同数据集上使用 SVM 方法的效果，将在接下来的小节中展示和讨论。

3.1. 不同的 BLSTM 架构

表格 1 的上部展示了使用 90 帧序列的各种 BLSTM 架构的结果。正确率从 83 % 到 85 %，宏观 F1 分数在 83 % 到 84 % 之间。灵敏度从 78 % 到 83 %，特异性从 85 % 到 89 %。整体表现最佳的是 BLSTM 2×256 和 BLSTM 3×128 架构，它们的 F1 分数为 84 %。BLSTM 2×256 的准确性较高，而 BLSTM 3×128 的灵敏度和特异性较高。BLSTM 2×128 和 BLSTM 3×256 的表现略逊于其他两者。总体而言，不同架构之间的性能差异较小，只有 BLSTM 2×128 与其他架构有显著不同。

Table 1: 不同 BLSTM 架构在序列长度为 90 帧时的评估结果，接下来是同一数据集上 SVM 方法的 Russello et al. (2024) 结果。结果以百分比 (%) 表示。最佳结果以粗体显示，次佳结果为 underlined。

Model	N # frames	Accuracy	F1 (macro)	Sensitivity	Specificity
BLSTM 2×128	90	83.48	82.63	78.48	88.74
BLSTM 3×128	90	<u>84.47</u>	83.70	82.74	<u>87.78</u>
BLSTM 2×256	90	84.59	83.70	<u>81.94</u>	87.07
BLSTM 3×256	90	83.11	82.37	81.77	85.07
Gait features + SVM (filtering)	124 ($\pm$ 24)	80.07	78.70	76.78	81.15
Gait features + SVM (no-filtering)	124 ($\pm$ 24)	75.17	73.38	70.70	76.26

3.1.1. 与参考模型的比较

表 1 的下半部分展示了相同数据集的 Russello et al. (2024) 结果。评估结果展示了他们的最佳模型，即径向核 SVM 分类器，分别有和没有关键点过滤的情况。

SVM 分类器在所有指标上的性能较低，无论是否使用关键点过滤方法。例如，在没有对关键点轨迹进行过滤的情况下，F1 分数比 BLSTM 低了大约九个百分点。在应用过滤后，SVM 分类器的性能显著提高，但仍然明显低于 BLSTM，其 F1 分数低了五个百分点。BLSTM 模型的预测与 SVM 模型的预测之间存在显著差异。

3.2. 不同的序列长度

表 2 展示了用 90、60 和 30 帧序列训练的 BLSTM 3×128 模型的结果。使用 90 帧序列提供了最佳的整体性能。虽然使用 60 帧和 30 帧序列时，在所有指标上的性能较低，但差异没有统计上的显著性。

Table 2: 使用不同序列长度训练的 BLSTM 模型的评估结果。结果以百分比表现 (%)。最佳结果以粗体突出显示。

Model	N # frames	Accuracy	F1 (macro)	Sensitivity	Specificity
BLSTM 3×128	90	84.47	83.70	82.74	87.78
	60	83.08	82.23	82.60	83.73
	30	83.07	82.23	82.07	83.91

4. 讨论

4.1. 跛行检测

本研究的主要目的是调查基于学习的端到端方法是否能够在最近的文献中相比于传统的特征提取方法提高跛行检测能力 (Russello et al., 2024)。我们在相同的实验条件下，对我们的 BLSTM 方法与基于特征的 SVM 方法 Russello et al. (2024) 进行了直接比较。具体而言，我们使用了相同的 272 个奶牛视频数据集、相同的标签，以及由 T-LEAP 姿势估计模型生成的相同关键点轨迹。

所有提出的 BLSTM 架构的表现都显著优于 SVM 模型，其中 BLSTM 3×128 和 2×256 实现了最高的准确率。这些结果表明，通过 BLSTM 直接学习时间步态模式，相较于手动特征工程方法，性能得到了改进。关于对噪声的鲁棒性，一个关键发现是：SVM 模型需要一个预处理步骤来校正异常值和过滤关键点轨迹，当使用未过滤的数据时，其准确率下降了 5 个百分点。相比之下，我们的 BLSTM 模型即使在未过滤的轨迹上训练，也能取得更好的结果。这种对噪声关键点检测的鲁棒性表明，BLSTM 在面对诸如遮挡或损坏的视频帧等现实世界的挑战时会更具弹性。进一步的研究可以通过系统地评估 BLSTM 在不同的验证数据噪声水平下的表现

来探讨这种鲁棒性的程度。这样的分析有助于确立跛行检测应用中姿态估计模型的最低准确率要求。

我们将跛足检测视为二元分类任务，即“正常”或“跛足”行走，而不是多类别运动评分。将 Sprecher et al. (1997) 的 5 级运动量表（正常（1），轻微跛足（2），中度跛足（3），跛足（4），严重跛足（5））合并为二元分类（正常（1）对比跛足（2-5））的决定是在 Russello et al. (2024) 中提出的，目的是解决数据集的局限性：其相对较小的规模和向较低跛足水平倾斜的分布。正如在 Russello et al. (2024) 中讨论的，较精细的运动评分可能需要一个更大且更好地代表严重跛足病例的数据集。此外，分类器的性能可能会受到真值数据质量的限制，因为即使在数据整理之后，运动评分的主观性质所带来的某些偏差可能仍然存在。

最初，T-LEAP 姿态估计模型是在 17 个关键点 (Russello et al., 2021) 上训练的，其中包括其在髋关节和腕关节/踝关节处的两个额外腿部关键点。在这里，我们使用其中的九个关键点（四个蹄子，三个背部关键点，额头和鼻子）来表示步态运动学。我们使用这个关键点子集来直接比较我们的 BLSTM 方法和 Russello et al. (2024) 中使用九个关键点计算手工制作步态特征的 SVM 方法。由于 BLSTM 能够捕捉复杂的时间模式，结合额外的关键点可以实现更丰富的步态运动学表示。例如，包含在髋关节和腕关节/踝关节处的关键点将包括触地和离地角度的信息，以及关节角速度，这些可能会随着跛行严重程度而变化 (Pluk et al., 2012)。研究额外关键点对跛行检测的好处留待未来工作进行。

Russello et al. (2024) 使用的数据集中包含的是在奶牛连续行走、没有分心或中断时拍摄的视频。因此，模型在假设不间断运动的情况下进行了训练和评估。然而，奶牛停止行走对跛行检测性能的影响尚未探讨。这种潜在的选择性偏差可能是一个限制，因为现实世界的情况下经常涉及奶牛暂停行走。未来的研究应评估当奶牛表现出自然行为时模型性能的下降情况。这将为这些检测系统在农场环境中的适用性提供见解。

4.2. 不同的序列长度

在 3×128 BLSTM 模型上评估序列长度对跛行检测的影响，测试了 30、60 和 90 帧的序列（分别对应于 30 FPS 视频中的 1、2 和 3 秒）。虽然 90 帧序列达到了最佳性能，但较短序列的结果也具有可比性，这表明仅凭一秒钟的视频数据可能就能实现跛行检测。

查看我们数据集中步幅的持续时间，健康的奶牛平均需要 31 帧（1.03 秒）来完成一个步幅（蹄打击到蹄打击），而严重跛行的奶牛需要多达 45 帧（1.5 秒）。尽管 30 帧的序列可能无法捕捉到完整的步态周期（即，每条腿一次步幅），但它们似乎为准确检测提供了足够的信息。这可以归因于几

个因素。首先，即使在 30 帧的窗口内，模型也可以评估行走速度，而这种速度在中度到重度跛行时会下降 (Blackie et al., 2013)。其次，对于较慢的、跛行的奶牛，这些 30 帧的序列捕捉到的步幅更少，虽然它们可能错过了受影响腿的步幅，但可能包括奶牛用来缓解疼痛的健康肢体中的代偿性动作 (Blackie et al., 2013)。

尽管较短序列的结果很有前景，但我们建议在可能的情况下使用较长序列。较短的序列有可能错过关键的步态异常，特别是在考虑到奶牛作为坚忍的动物时，常常试图隐藏它们的疼痛。较长的序列提供了更多机会来识别步态模式中的细微异常，并建立对动物运动的更全面评估。

进一步研究了解所需的最低序列长度可能会使农场上的相机布置更优化。例如，在空间有限或杂乱的环境中，动物可能只能在短时间内可见。

4.3. 与相关工作的比较

由于数据、方法和评估指标的差异，我们的做法与相关工作进行直接比较存在显著挑战。在小节 4.1 中，我们通过使用相同的数据集并利用他们的开源代码，与 Russello et al. (2024) 进行了直接比较，这使我们能够客观地展示我们方法的改进。

在以下段落中，我们将讨论我们的方法与其他直接相关研究的关系。然而，这些比较受限于这些工作中公开数据和代码的缺失。如果不重复他们的实验（这将耗费大量时间），直接的一对一比较是不切实际的。尽管如此，我们在这些限制内比较了我们的结果，并对我们的发现进行了对比。为了推进可重复性并促进开源研究实践，我们承诺在发表时公开我们的数据和代码，并鼓励在未来的研究中保持类似的透明度。

Karoui et al. (2021) 开发了一种基于物理反光标记放置在牛腿上的关键点轨迹的跛行检测系统。他们为每条腿提取了四个关键点轨迹，并通过向关键点轨迹添加随机噪声生成了额外的合成数据，从而得到了一个包含 24000 个样本的数据集。这些关键点轨迹用于训练一个 LeNet CNN LeCun et al. (1998) 进行二元分类，实现了 91 % 的准确率和 F1 分数。我们的方法不同之处在于使用了一种无标记的姿态估计模型，而不是使用物理标记，这种方法在实践中侵入性更小且更易于部署。此外，在运动数据中，物理标记更容易受到皮肤移动错误的影响，因为在行走时皮肤会位移 (Bergh et al., 2014)。除了腿部关键点，我们还包括了背部和头部区域的关键点，使我们能够覆盖更广泛的步态模式。他们的实验表明，通过添加 5 % 噪声变异生成新数据可以产生最佳结果，虽然我们同样使用 1 % 标准差的高斯噪声增强了我们的数据，但我们并未广泛研究其影响。然而，基于 Karoui 等人的研究结果，我们相信这一增强技术有助于提高我们模型的鲁棒性和泛化能力。

Wu et al. (2020) 开发了一种跛行检测系统，使用 YOLOv3 在行走奶牛的视频中进行腿部定位。该方法通过计算每帧中左右腿之间的水平距离来提取相对步长。这些步长矢量序列被用作各种分类模型的步态特征。他们在 700 个视频上使用 10 折交叉验证进行评估，在不同的架构上取得了高准确率：决策树 (90.5 %)，支持向量机 (94.3 %)，LSTM (98.6 %)，以及 BLSTM (99 %)。他们的 SVM 和 BLSTM 之间的性能差距 (4-5 个百分点) 也与我们的研究发现一致。尽管他们的总体准确度超过了我们的结果，这可以归因于他们更大的训练数据集：他们的 10 折交叉验证使用了 630 个视频进行训练，而我们的 5 折交叉验证平均使用了 228 个视频进行训练。当他们将训练集缩减到 350 个视频并使用 2 折交叉验证时，性能显著下降 (SVM: 82.9 % , BLSTM: 86.7 %)。此外，尽管他们使用的序列明显较短 (30-90 帧与他们的 500 帧相比)，我们的系统仍达到了类似的性能。我们的方法在以下几个方面不同。虽然 Wu et al. (2020) 仅关注腿部的水平运动，我们纳入了腿部、背部和头部关键点的水平和垂直运动，提供了更丰富的空间信息。此外，他们首先从腿部坐标计算步长矢量，并将此步态特征作为模型输入，而我们使用原始关键点轨迹，从而无需手动特征工程和预处理。

Jiang et al. (2020) 提出了一个跛行分类方法，该方法结合了光流图和 BLSTM 分类器。他们的研究使用了一个包含 1080 段行走奶牛视频的数据集 (其中 756 段用于训练，324 段用于测试)，视频长度从 125 帧到 1000 帧，跛行严重程度被分为 4 个等级 (正常、轻微、中度和严重)。从视频帧生成光流图序列，并使用 DenseNet CNN (Huang et al., 2017) 从光流图的每个时间步中提取空间特征。然后训练一个 BLSTM 分类器来从空间特征序列中对跛行的 4 个等级进行分类，达到了 95 % 的分类准确率。我们的方法主要在于从视频帧提取空间数据的方式有所不同。我们专注于跟踪九个特定的关键点坐标，而他们的方法则使用光流图从整个身体捕捉运动信息。尽管他们的方法提供了更丰富的时空信息，但导致 BLSTM 的输入维度显著增加 (他们的方法是 $T \times 224 \times 224$ ，而我们的方法是 $T \times 18$)。这种复杂性的增加使得他们的学习过程在计算上要求更高，需要预训练和更大的训练数据集。此外，他们研究了序列长度对分类准确性的影响，从使用 30 帧序列的 85 % 提高到使用 60 帧序列的 95 %。值得注意的是，他们展示了较长的序列尤其有利于中度和重度跛行病例的检测，这可以归因于跛行奶牛较慢的行走速度。我们的发现在这方面与他们有所不同，因为我们观察到 BLSTM 模型在使用 30 帧、60 帧和 90 帧时，准确性差异最小。这些不同的发现可能归因于数据集组成的不同。首先，他们在四个均衡的跛行严重程度等级上进行分类，而我们进行的是二元分类。尽管我们的二元标签数据集相对均衡，这些二元标签是从一个五级跛行程度量表中合并得来，偏向于非跛行和轻微跛行奶牛，并且重度病例较少。合并跛行

程度降低了数据的细粒度，从而可能掩盖了使用较长序列检测重度病例的好处。

总结来说，这一小节将我们的姿态估计和 BLSTM 方法与三个直接相关的研究进行了比较：Karoui et al. (2021) 在牛腿上使用物理反射标记，LeNet CNN 实现了 91 % 的准确度，但需要侵入式标记放置；Wu et al. (2020) 采用 YOLOv3 进行腿的定位，并在 700 个视频上结合 BLSTM 达到了最高 99 % 的准确度，但在较小的训练集上性能显著下降；Jiang et al. (2020) 将光流图与 BLSTM 结合应用于四级跛行分类，实现了 95 % 的准确度，但需要更高的计算复杂性和更大的数据集。

5. 结论

在本文中，我们开发了一种结合姿态估计和 BLSTM 的跛行检测方法。通过 T-LEAP 姿态估计模型，从行走奶牛的视频中提取了位于奶牛蹄、头部和背部的九个关键点的运动序列。然后将关键点的轨迹用作 BLSTM 分类器的输入，该分类器经过训练可执行二元跛行分类。实验包括将我们提出的四种 BLSTM 架构与依赖于人工设计的步态特征的既定方法进行比较，以及比较在短序列长度（1、2 和 3 秒的视频数据）期间的性能。

在所提出的模型架构中，BLSTM 3×128 和 2×256 导致了整体上最好的性能，准确率为 84.5 %，而基于特征的方法的准确率为 80.1 %。此外，我们的姿态估计和 BLSTM 方法与使用 BLSTM 跛行分类器的其他研究达到了相当的性能（Wu et al., 2020; Jiang et al., 2020）。在不同的序列长度上，BLSTM 分类器表现出类似的性能。我们的模型在使用 3 秒序列时达到 84.5 % 的准确率，在使用 2 秒和 1 秒序列时均达到 83.1 % 的准确率，表明跛行可以通过仅仅一秒的视频数据检测出来。

结合无标记姿态估计从多个关键点提取运动，以及 BLSTM 学习时间步态特征，提供了几个优势。使用原始关键点轨迹消除了手动特征工程的需要，同时包括腿部、背部和头部关键点提供了丰富的空间信息。此外，我们的方法比先前使用 BLSTM 跛行分类器的研究更为高效（Wu et al., 2020; Jiang et al., 2020），因为我们处理的序列显著更短，训练数据集也更小。

虽然我们的方法在从视频中自动检测跛行方面提供了一个有前景的方向，但进一步的研究方向包括：(1) 系统地评估在不同噪声水平下 BLSTM 的稳健性，以确定姿态估计模型的最低准确性要求，(2) 探讨额外关键点带来更好检测效果的好处，(3) 在一个多点刻度上探索细粒度跛行分类，而不是二元分类，(4) 评估跛行检测在“真实世界”条件下的性能，以及 (5) 确定视频片段的最短持续时间要求。

6.

致谢 本出版物是研究项目“人类与动物健康深度学习”（项目编号 EDL P16-25-P5）的一部分，该项目属于高效深度学习研究计划（<https://efficientdeeplearning.nl>），由荷兰研究理事会（NWO）（部分）资助。

7.

利益声明 作者声明，他们没有已知的竞争性财务利益或个人关系可能会影响本文中报告的工作。

8.

在写作过程中声明生成式人工智能和人工智能辅助技术的使用 在撰写本文的过程中，作者使用了 Anthropic 公司的 Claude (Sonnet 4) 工具以优化选定的句子，使其更加清晰。使用该工具后，作者对内容进行了审查和编辑，并对发表文章的内容承担全部责任。

References

- Ansel, J., Yang, E., He, H., Gimelshein, N., Jain, A., Voznesensky, M., Bao, B., Bell, P., Berard, D., Burovski, E., et al. (2024). Pytorch 2: Faster machine learning through dynamic python bytecode transformation and graph compilation. In *Proceedings of the 29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 2*, pages 929–947.
- Arazo, E., Aly, R., and McGuinness, K. (2022). Segmentation Enhanced Lameness Detection in Dairy Cows from RGB and Depth Video. arXiv:2206.04449 [cs].
- Barney, S., Dlay, S., Crowe, A., Kyriazakis, I., and Leach, M. (2023). Deep learning pose estimation for multi-cattle lameness detection. *Scientific Reports*, 13(1):4499.
- Battistone, F. and Petrosino, A. (2019). TGLSTM: A time based graph deep learning approach to gait recognition. *Pattern Recognition Letters*, 126:132–138.
- Bengio, Y., Simard, P., and Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE transactions on neural networks*, 5(2):157–166.
- Bergh, A., Egenvall, A., Olsson, E., Uhlhorn, M., and Rhodin, M. (2014). Skin displacement in the equine neck. *Equine Veterinary Journal*, 46:36–36.
- Blackie, N., Bleach, E., Amory, J., and Scaife, J. (2013). Associations between locomotion score and kinematic measures in dairy cows with varying hoof lesion types. *Journal of Dairy Science*, 96(6):3564–3572. ISBN: 0022-0302 Publisher: Elsevier.
- Du, X., Vasudevan, R., and Johnson-Roberson, M. (2019). Bio-lstm: A biomechanically inspired recurrent neural network for 3-d pedestrian pose and gait prediction. *IEEE Robotics and Automation Letters*, 4(2):1501–1508.
- Enting, H., Kooij, D., Dijkhuizen, A., Huirne, R., and Noordhuizen-Stassen, E. (1997). Economic losses due to clinical lameness in dairy cattle. *Livestock production science*, 49(3):259–267.
- Farhadi, A. and Redmon, J. (2018). Yolov3: An incremental improvement. In *Computer vision and pattern recognition*, volume 1804, pages 1–6. Springer Berlin/Heidelberg, Germany.
- Feichtenhofer, C., Fan, H., Malik, J., and He, K. (2019). Slowfast networks for video recognition. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 6202–6211.
- Flower, F., Sanderson, D., and Weary, D. (2005). Hoof pathologies influence kinematic measures of dairy cow gait. *Journal of dairy science*, 88(9):3166–3173.
- Gers, F. A., Schraudolph, N. N., and Schmidhuber, J. (2002). Learning precise timing with lstm recurrent networks. *Journal of machine learning research*, 3(Aug):115–143.
- Graves, A., Mohamed, A.-r., and Hinton, G. (2013). Speech recognition with deep recurrent neural networks. In *2013 IEEE international conference on acoustics, speech and signal processing*, pages 6645–6649. Ieee.
- Graves, A. and Schmidhuber, J. (2005). Framewise phoneme classification with bidirectional lstm and other neural network architectures. *Neural networks*, 18(5-6):602–610.
- Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8):1735–1780.
- Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708.
- Jiang, B., Yin, X., and Song, H. (2020). Single-stream long-term optical flow convolution

- network for action recognition of lameness dairy cow. *Computers and Electronics in Agriculture*, 175(April):105536. Publisher: Elsevier.
- Kang, X., Zhang, X. D., and Liu, G. (2020). Accurate detection of lameness in dairy cattle with computer vision: A new and individualized detection strategy based on the analysis of the supporting phase. *Journal of Dairy Science*, 103(11):10628–10638. Publisher: Elsevier.
- Karoui, Y., Jacques, A. A. B., Diallo, A. B., Shepley, E., and Vasseur, E. (2021). A Deep Learning Framework for Improving Lameness Identification in Dairy Cattle. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(18):15811–15812. Number: 18.
- LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324.
- Lefebvre, G., Berlemont, S., Mamalet, F., and Garcia, C. (2013). BLSTM-RNN Based 3D Gesture Classification. In *Artificial Neural Networks and Machine Learning – ICANN 2013*, pages 381–388. Springer.
- Loshchilov, I. and Hutter, F. (2017). Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*.
- McNemar, Q. (1947). Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika*, 12(2):153–157.
- O’leary, N., Byrne, D., O’Connor, A., and Shalloo, L. (2020). Invited review: Cattle lameness detection with accelerometers. *Journal of dairy science*, 103(5):3895–3911.
- Pascanu, R., Mikolov, T., and Bengio, Y. (2013). On the difficulty of training recurrent neural networks. In *International conference on machine learning*, pages 1310–1318. Pmlr.
- Pluk, A., Bahr, C., Poursaberi, A., Maertens, W., Van Nuffel, A., Berckmans, D., Pluk, A., Poursaberi, A., Berckmans, D., and Van Nuffel, A. (2012). Automatic measurement of touch and release angles of the fetlock joint for lameness detection in dairy cattle using vision techniques. *Journal of dairy Science*, 95(4):1738–1748. Publisher: Elsevier.
- Poursaberi, A., Bahr, C., Pluk, A., Nuffel, A. V., Berckmans, D., Van Nuffel, A., and Berckmans, D. (2010). Real-time automatic lameness detection based on back posture extraction in dairy cattle: Shape analysis of cow with image processing techniques. *Computers and Electronics in Agriculture*, 74(1):110–119. ISBN: 0168-1699 Publisher: Elsevier B.V.
- Reddi, S. J., Kale, S., and Kumar, S. (2019). On the convergence of adam and beyond. *arXiv preprint arXiv:1904.09237*.
- Russello, H., van der Tol, R., Holzhauer, M., van Henten, E. J., and Kootstra, G. (2024). Video-based automatic lameness detection of dairy cows using pose estimation and multiple locomotion traits. *arXiv preprint arXiv:2401.05202*.
- Russello, H., van der Tol, R., and Kootstra, G. (2021). T-LEAP: occlusion-robust pose estimation of walking cows using temporal information. *arXiv:2104.08029 [cs]*. arXiv: 2104.08029.
- Song, X., Leroy, T., Vranken, E., Maertens, W., Sonck, B., and Berckmans, D. (2008). Automatic detection of lameness in dairy cattle-Vision-based trackway analysis in cow’s locomotion. *Computers and Electronics in Agriculture*, 64(1):39–44. ISBN: 0168-1699 _eprint: 9809069v1.

- Sprecher, D., Hostetler, D., and Kaneene, J. (1997). A LAMENESS SCORING SYSTEM THAT USES POSTURE AND GAIT TO PREDICT DAIRY CATTLE REPRODUCTIVE PERFORMANCE. *Science*, (97).
- Taghavi, M., Ouweltjes, W., Klandermans, B., and Kamphuis, C. (2024). Keeping an eye on locomotion and behavior using computer vision. In *11th European Conference on Precision Livestock Farming*, pages 1268–1275. European Association for Precision Livestock Farming.
- Taghavi, M., Russello, H., Ouweltjes, W., Kamphuis, C., and Adriaens, I. (2023). Cow key point detection in indoor housing conditions with a deep learning model. *Journal of Dairy Science*.
- Van Hertem, T., Tello, A. S., Viazzi, S., Steensels, M., Bahr, C., Romanini, C. E. B., Lokhorst, K., Maltz, E., Halachmi, I., Berckmans, D., Schlageter Tello, A., Viazzi, S., Steensels, M., Bahr, C., Romanini, C. E. B., Lokhorst, K., Maltz, E., Halachmi, I., and Berckmans, D. (2018). Implementation of an automatic 3D vision monitor for dairy cow locomotion in a commercial farm. *Biosystems Engineering*, 173:166–175. ISBN: 1537-5110 Publisher: Elsevier.
- Viazzi, S., Bahr, C., Van Hertem, T., Schlageter-Tello, A., Romanini, C. E. B., Halachmi, I., Lokhorst, C., and Berckmans, D. (2014). Comparison of a three-dimensional and two-dimensional camera system for automated measurement of back posture in dairy cows. *Computers and Electronics in Agriculture*, 100:139–147. ISBN: 0168-1699 Publisher: Elsevier B.V.
- Wainer, J. and Cawley, G. (2021). Nested cross-validation when selecting classifiers is overzealous for most practical applications. *Expert Systems with Applications*, 182:115222.
- Whay, H. R. and Shearer, J. K. (2017). The impact of lameness on welfare of the dairy cow. *Veterinary Clinics: Food Animal Practice*, 33(2):153–164.
- Wu, D., Wu, Q., Yin, X., Jiang, B., Wang, H., He, D., and Song, H. (2020). Lameness detection of dairy cows based on the YOLOv3 deep learning algorithm and a relative step size characteristic vector. *Biosystems Engineering*, 189:150–163. Publisher: Elsevier Ltd.
- Zhao, K., Zhang, M., Ji, J., Zhang, R., and Bewley, J. M. (2023). Automatic lameness scoring of dairy cows based on the analysis of head- and back-hoof linkage features using machine learning methods. *Biosystems Engineering*, 230:424–441.
- Zheng, Z., Zhang, X., Qin, L., Yue, S., and Zeng, P. (2023). Cows’ legs tracking and lameness detection in dairy cattle using video analysis and Siamese neural networks. *Computers and Electronics in Agriculture*, 205:107618.