Hochschule
Bonn-Rhein-Sieg
University of Applied Sciences

b-it Bonn-Aachen International Center for Information Technology

Master Thesis Proposal

Automated gait and pose analysis for lameness detection in cows

Wei-Chan Hsu

Supervised by

Prof. Dr. Rainer Herpers
Prof. Dr. Wolfgang Koch
M. Sc. Michael Schleiss

February 2019

1. Introduction

Lameness is a serious disorder in dairy farms that increases the risk of culling of cattle as well as economic losses, such as decrease in milk production and the cost of medication [1, 2]. This issue is addressed by lameness detection using locomotion scoring [1, 3], which assesses the lameness level of cattle by monitoring their pose and gait patterns. The scoring is typically estimated by veterinarians through visual inspection, which is not only time-consuming and laborious, but also subjective and may vary from time to time. Researchers have proposed different approaches to automatic lameness detection and scoring. Computer vision techniques are particularly suitable candidates for the task since such systems can operate remotely without changing the daily routine of the animals [4]. The detection is often based on the back posture of the cows [4, 5, 6, 7]. However, most previous studies [4, 6, 7, 8] employ traditional computer vision techniques and take into account only a few hand-drafted features, which may not be enough to accurately predict the score for cows. Deep learning can tackle this issue because of its automatic feature representation, enabling more features to be learned for prediction.

This study aims to evaluate various vision-based approaches to automated locomotion scoring. In particular, some deep learning architectures that have been applied to similar domains will be tested; the architecture may contain convolutional neural networks (CNNs), which has been mainly used to solve vision tasks [9, 10], or recurrent neural networks (RNNs), which are effective for sequential data [11, 12]. The proposed approach takes a series of frames as input and outputs the predicted locomotion score. While there are many types of locomotion scoring systems, this study will adopt the 5-point scale established by Sprecher et al. [1]. The score ranges from 1 (normal) to 5 (severely lame), which is mainly based on the archness of cow's back as well as the walking patterns, as described in Table 1.1 and Figure 1.1. In this work, the terms of lameness detection and locomotion score prediction will be used interchangeably.

To validate the efficacy of deep learning in lameness detection, a traditional machine learning technique, i.e. support vector machine (SVM), will be used as the baseline. Unlike deep learning, feature engineering is necessary for SVM, so several keypoints on the cow's body will be selected for pose estimation, in which spatial-temporal features can be extracted for lameness detection. These keypoints serve as the elements for gait and pose analysis of cows. As the amount of data for this study is limited, pose estimation may improve feature representation of deep learning approach. Hence, investigating how pose estimation can facilitate lameness detection is also one of the goals in this study.

To summarize, the contributions of this study include:

- Survey of state of the art in automated lameness detection
- Investigate gait and pose features of cows that influence the lameness detection, and use them to predict locomotion score based on the 5-point scale.
- Create a dataset of cows annotated with keypoints for lameness detection
- Investigate methods from similar tasks and transfer them to automated lameness detection.

Table 1.1: Criteria used to assign a locomotion score and clinical description [1].

Locomotion Score	Clinical Description	Assessment Criteria
1	Normal	The cow stands and walks with a level-back posture. Her gait is normal.
2	Mildly lame	The cow stands with a level-back posture but develops an arched-back posture while walking. Her gait remains normal.
3	Moderately lame	An arched-back posture is evident both while standing and walking. Her gait is affected and is best described as short-striding with one or more limbs.
4	Lame	An arched-back posture is always evident and gait is best described as one deliberate step at a time. The cow favors one or more limbs/feet.
5	Severely lame	The cow additionally demonstrates an inability or extreme reluctance to bear weight on one or more of her limbs/feet.

2. Related Work

This section reviews some previous research of automated lameness detection and two relevant topics: deep learning and motion analysis. Due to the substantial success in computer vision tasks, deep learning approach has great potential and will be tested and evaluated in this study. In order to detect the lameness of cows, it is necessary to analyze the cow's pose and gait to extract spatio-temporal features; hence, motion analysis is necessary, especially with regard to gait and pose.

(a) LS = 1 (b) LS = 2

(c) LS = 3 (d) LS = 4

(e) LS = 5

Figure 1.1: Locomotion scores (LS) of cows when standing and moving [13].

2.1. Automated lameness detection

In order to deal with the lameness issues, a common method is to manually identify the variations in gaits or postures of the cows and assign locomotion scores. While it is quick to apply and easy to perform, it is time-consuming, subjective, and can lead to inconsistency [5]. Many researchers have proposed different approaches to automatic lameness detection by analyzing various measurements. One approach is to use load cells or pressure sensors to measure the ground reaction forces of the cows, such that the gait patterns of the cows, such as the stance time or asymmetry in weight bearing between the right and left limbs, can be observed [14, 15].

Another popular approach makes use of computer vision techniques to visually observe the poses or gait features of the cows. Song et al. associated the scale of lameness with the step overlap of the cows from side-view images; the results show a positive linear relationship between cows' trackways overlap and locomotion scores by human observation [8]. Poursaberi et al. developed a vision system to assess the lameness of cows by analyzing the parameters of two ellipses fitted on the back of the cows [6]. However, the aforementioned work relies on image segmentation, which is strongly affected by the changing lighting conditions. Some researchers proposed methods of back arch estimation of cows by using 3D cameras [4, 16]. Since the 3D cameras were used to capture top-view images, they cannot directly access the gait features. Additionally, thermography has been increasingly used to detect lameness in cattle [17]; it is based on the fact of positive correlation between cow's sole temperature and the locomotion score. Nonetheless, thermography is more of a tool for lameness prevention than for detection, since foot lesions can be present without any sign of lameness [17]. A recent work [18] utilized 2D near infrared images to train a Faster R-CNN to detect the hooves and carpal/tarsal joints of cows, such that gait features can be obtained for lameness detection.

Most of the proposed methods only considered either the back posture [6, 19, 16] or gait [8, 14, 15] for lameness detection. Due to the lack of features, these methods usually adopted a three-point [6, 14] or even binary [15, 19] locomotion scoring instead of the standard five-point scale. This work will combine the features extracted from both the back pose and the gait for lameness detection based on the five-point scoring scale.

2.2. Deep learning

Deep learning is a branch of machine learning that has been applied to solve various tasks efficiently, such as computer vision and natural language processing. Unlike traditional machine learning techniques, deep learning can learn representations of data without feature engineering, which is achieved through the architecture of deep neural networks. With multiple layers, a deep network can learn features in a hierarchical fashion; the farther a layer is from the input layer, the more abstract the feature it can extract [11].

One of the most popular forms of deep learning are convolutional neural network (CNNs), which contain a series of convolutional and pooling layers whose primary goals are feature extraction and dimensionality reduction, respectively. Rectified linear unit (ReLU) is usually used between layers to introduce nonlinearity. At the end of the network are fully connected layers that predict a class score or other forms of output. CNNs are largely used for vision tasks, and have many variations for addressing different issues [20, 21]. Recurrent neural network (RNNs), designed to process sequence data, are another popular form of deep learning. They have feedback loops and maintain in the hidden layers a state vector that capturing the history of the past elements of the sequence [11]. Long short-term memory (LSTM) networks [22] are a variation of RNNs with special hidden units that determine the amount of the past and new information to remember. They circumvent the vanishing and exploding gradients problem of conventional RNNs, so that are more effective to learn long sequences [11]. This work will try to leverage the ability of deep learning to extract spatio-temporal features for automated lameness detection.

2.3. Motion analysis

Motion analysis is a popular topic of computer vision because of its wide applications [23], including surveillance, human-machine interaction, athletic movement analysis, clinical diagnosis, among others. Generally, there are three major tasks (Figure 2.1): detection, tracking, and recognition or behavior understanding [24, 25]. Detection involves motion segmentation to differentiate the objects from the background as well as object classification to determine the type of the objects; tracking is a temporal process concerned with objects matching in consecutive frames using visual features; recognition aims to interpret the behaviors of the objects through the sequence of frames [24, 25]. Since this thesis is concerned with pose estimation along with gait and pose analysis, this section focuses on these two topics; the former can be considered as the next step or part of object tracking [26], while the latter belongs to behavior understanding.

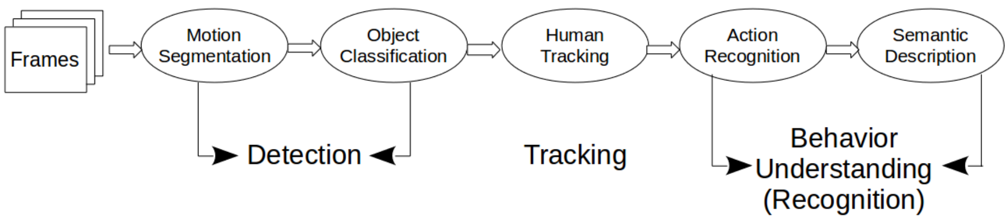


Figure 2.1: A general framework for human motion analysis (adapted from [25]).

2.3.1. Pose estimation

Pose estimation is the process of detecting the configuration, either joints or keypoints, of an object of interest [26]. Both spatial and temporal features can be extracted from the estimated poses over time, and can be further used for behavior understanding. The approaches can be divided into traditional and deep learning-based, as discussed below.

Traditional approaches

Traditional pose estimation approaches can be divided into two classes: top-down [27] and bottom-up [28]. Top-down estimation tries to match the projection of body with the image observation; it has the drawbacks of manual initialization of estimate in the first frame, and occlusion issue. Bottom-up estimate assemblies the body parts into a whole body, requiring detectors for body parts [29]. The two classes can be combined for better estimate [30, 31].

Deep learning-based approaches

DeepPose [10] is a human pose estimation algorithm by applying a cascade of convolutional neural networks, each of which being a regressor. The prediction is refined at each stage in which the processed image is cropped around the predicted joints location from the preceding stage. Another CNN-based system, called DensePose [32], is based on the mask-RCNN architecture that contains a cascade of proposing regions-of-interest (ROI); it was designed to map all human pixels of an image to the 3D surface of the human body. While most studies focus on human pose estimation, Pereira et al. developed a CNN-based animal pose estimation system called LEAP (LEAP Estimates Animal Pose) [33]. The authors adopted a fully convolutional network architecture that maps from raw images to a set of confidence maps for each body part. The locations of body parts are obtained by minimizing the errors between ground truth and predicted confidence maps. Another study [34] implemented transfer learning to track body parts of the animals with minimal training data. The authors showed that 100 frames can achieve good tracking performance with an accuracy of five pixels. The model, called DeepLabCut, is a deep residual neural network (ResNet) based on the architecture of DeeperCut [35], which is a human pose estimation model.

DeepLabCut is an open-source pose estimation framework that takes the advantage of transfer learning. As one goal of this study is to do gait and pose analysis, this study will make use of DeepLabCut to estimate cow's pose for analysis.

2.3.2. Gait and pose analysis

Gait and pose analysis is the study of human's or animal's motion with an aim to detect biomedical information, which can be utilized to identify individuals, biomechanical abnormalities or diagnose diseases [36]. Gait provides particularly important biomedical information for analysis because of its repeatability, robustness, and remote accessibility [37]; moreover, gait is a reliable indicator of health condition [38, 39]. Gait analysis can be categorized into sensor-based and vision-based [41]; the former utilizes micro-electro-mechanical systems (MEMS) such as inertial measurement units (IMUs) to track the motion of lower-limb joints [42], while the latter tracks the motion from sequence of images. Vision-based systems can be further divided into marker-based and markerless systems [42].

Many gait analysis methods have been proposed, and vision-based is preferred over sensor-based since it is nonintrusive and easier to set up. Vision-based gait analysis consists of three main steps: image acquisition, data processing, and gait analysis [38]. Most studies used gait for biometric recognition [36, 37, 40]. Wang [43] tried to recognize walking patterns by applying principal component analysis (PCA) to the frame-to-frame optical flows. Due to the automatic feature learning ability, deep learning has been applied to gait analysis for feature extraction. One study [36] combined a gait energy image (GEI) based Siamese neural network and a three-dimensional CNN (C3D) to extract spatial-temporal gait features. The authors in [12] proposed a feature learning method for gait recognition: a CNN is used to extract joint heatmaps, which are fed into an LSTM network that models high-level motion features.

As gait consists of vital information about health conditions, there is a growing interest in medical applications [39, 38]. One study used linear discriminant analysis (LDA) to extract gait features from binary silhouette to detect walking patterns of Parkinson's disease [44]. Cortés [45] applied random forest to gait pathology classification, using the lower extremity range of motion as the features; an accuracy of 95 % was reached under a five-fold cross validation. Due to the difficulty in obtaining large amount of medical data, deep learning has yet to be applied in medical gait analysis.

This work will investigate a few types of deep networks that were proposed for gait [12] or action recognition [36, 46, 47, 48], and applied them to lameness detection, which belongs to medical domain. Although action recognition is different from gait analysis, it also requires feature extraction from video frames. In other word, both the tasks of gait and action recognition share a common trait with lameness detection: the spatio-temporal features that are critical to analysis. Furthermore, gait recognition and lameness detection have similar context of data which contain a walking object

Despite the research of vision-based lameness detection and the popularity of deep learning in various fields, to the extent of our knowledge there are currently no published works on the combination of both. One reason may be the lack of data for model training. Besides, the majority of relevant deep learning-based studies focus on biometric recognition rather than medical gait analysis. More generally, there is a lack of research using deep learning for gait and pose analysis in medical field. This thesis aims to analyze the efficacy of deep learning techniques for lameness detection of cows using a limited amount of data, which will be approached by employing the methods from similar tasks. To further understand the detection of lameness, gait and pose analysis of cows is another important issue and will be addressed in the study as well.

From what was discussed above, there are some research questions to be investigated:

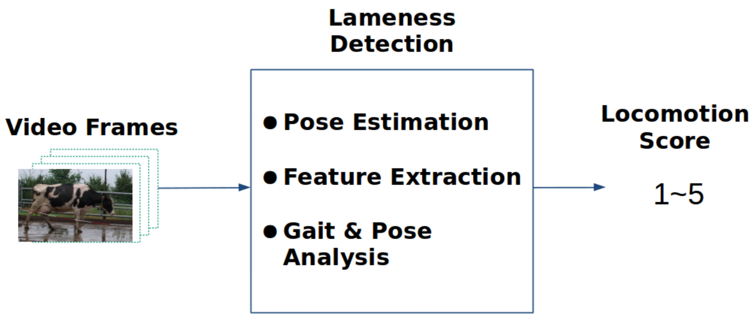
- Given that deep neural networks can automatically learn features, does it help to supply frames of estimated pose of cows as input? How can the estimated pose be applied to lameness detection?
- Which features are the most influential in lameness detection? Do spatial and temporal features have comparable importance?
- Which kind of models are suitable for lameness detection? Can deep networks outperform traditional machine learning using limited amount of data in this task?

3. Methodology

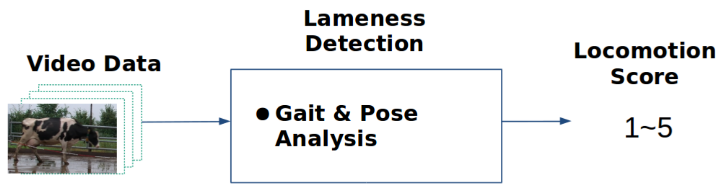
This section discusses the methods used for lameness detection, including how the data are pre-processed, feature extraction, the techniques employed to predict locomotion score, and evaluation method. Figure 3.1 demonstrates the workflow of the proposed automated locomotion scoring system. There are two variants: Given a series of video frames, one is to first extracts features from the estimated pose of the cow in the frames, and use them to predict a locomotion score for the cow, while the other is to directly use raw video frames to learn features and predict a score.

3.1. Data acquisition and preprocessing

The data are side-view videos collected using Samsung FULL HD Camcorder at the Teaching and Research Station Frankenforst of Bonn University on six different days in two weeks. Each video ranges from around three to ten minutes long and contains several walking cows, each of which was recorded separately on a walkway around eleven meters from the camera. There are 65 cows in total, each cow has her own ID number for recognition, and was examined and assigned a locomotion score by a veterinarian on three different days, including the first and final day. The



(a) Approach with pose estimation.



(b) End-to-end approach.

Figure 3.1: Flowchart of the proposed automated locomotion scoring system.

scores are considered as the ground truth, and are used to calculate the scores for the other days by interpolation.

Each video is first trimmed into shorter video clips, each of which contains only one cow walking from left to right of the frame. However, in some of the videos there is a person guiding the cow. Furthermore, these video clips are cropped in such a way that the cow is at the center of each frame, which also reduces the size of the data. This cropping step is achieved by applying the TensorFlow Object Detection API [49] to the original videos and using OpenCV toolbox to extract the region of interest and reduce the frame rate from 50 to 20 frame per second. The output of the preprocessing step are the cropped video clips with a resolution of 680×420. It should be noted that the video frames shown in Figure 3.1 are obtained by this preprocessing step discussed above.

3.2. Pose estimation

Pose estimation is considered as part of lameness detection as shown in Figure 3.1 since there is a strong correlation between the pose of cows and the locomotion score. This step estimates the pose of cows by extracting a few keypoints from their body. As described in Section 2.1, the lameness is closely related to the back posture and gait patterns, so the keypoints on a cow's back and hooves can be selected as shown in Figure 3.2. To optimize the performance of lameness detection, different choice of keypoints will be explored. Besides, the keypoints can also be used to construct the cow's skeleton. The deep learning-based pose estimation framework for animals, called DeepLabCut [34], is used in this project. The framework provides a graphical user interface (GUI) using wxPython for keypoints annotation. One advantage of DeepLabCut is that it supports both manual and automatic frame extraction for data labeling; moreover, it can produce great results without large amount of data thanks to transfer learning. Figure 3.2 shows a cow with ten keypoints estimated using DeepLabCut, which was trained with a few hundred of annotated frames of different cows.

Figure 3.2: Ten keypoints (colored crosses) on a cow for pose estimation: head, neck, withers, three points on the back, and one on each hoof.

3.3. Gait and pose analysis for locomotion score prediction

The estimated keypoints or skeleton from the last step are used as the features for gait and pose analysis. Two approaches (demonstrated in Figure 3.1) will be employed in the thesis: the first approach employs pose estimation, including support vector regression (SVR) as the baseline and some deep neural networks such as graph convolutional network [47] that requires objects with skeleton as input. The second approach directly uses raw images to predict locomotion score, including various architectures of deep networks.

Support vector regression (SVR), a variation of support vector machine (SVM), requires hand-crafted features for data discrimination. Different spatial-temporal features, such as the curvature of cow's back, the stride of cow's gait, and the motion of cow's head, will be explored for lameness detection. In the deep learning-based

Table 3.1: Candidate deep learning architecture for lameness detection.

Model	**Original Task**
CNN-LSTM [12]	Gait Recognition
C3D [36]	Action Recognition
Two-Stream [46]	Action Recognition
GCN [47]	Action Recognition
TSN [48]	Action Recognition

approach, some of the architectures listed in Table 3.1 will be tested, each of which is a combination of CNN and LSTM except for C3D. While most of these architectures were designed for action recognition, they are able to extract spatio-temporal features for video analysis. As the amount of data in this work is much less than most open-source datasets, transfer learning will be applied. The models take as input a sequence of images, each of which contains a cow with or without estimated pose. While most vision-based methods mentioned in Section 2.1 used single frame as input, this study applies sequence of frames as input to make use of temporal features which may allow better lameness detection. A video clip of one cow along with the locomotion score are considered as a training sample. As mentioned in the last section that it is unclear if pose estimation facilitate the prediction for deep networks, both the frames with and without estimated pose will be separately used as input. For traditional machine learning approach, the input also consists of extracted features to discriminate between different lameness levels. The output is a predicted locomotion score in the range of one and five. It should be noted that the locomotion score prediction is considered as a regression problem in this study even though the original scoring system [1] has five distinct scores; the reason is that the score describes a cow's health condition which should take a continuous value.

3.4. Evaluation

Different methods will be tested and evaluated for locomotion score prediction. The data will be split into training and test sets, and the mean squared errors will be used as the metric of scoring regression evaluation for comparing methods. For traditional machine learning methods, k-fold cross validation will be applied, and various sets of features will be tested in order to find out the most important features. As for deep learning-based methods, the training/test data will be separated in various ways in addition to random split: one way is to split the data based on the cow's ID, and another is based on the consistency of the cow's locomotion score. The reason is to check if the models can predict locomotion score based on the behaviors as well as the gait and pose features of cows rather than identify individual ones.

4. Project Plan

4.1. Work Packages

Starting date: 25 February, 2019
Completion date: 25 August, 2019

4.1.1. Task Description

Work package 1: Literature survey

- **1.1** Automatic lameness detection
- **1.2** Pose estimation
- **1.3** Gait analysis
- **1.4** Deep learning based motion analysis
- **Milestone 1:** Comprehensive state of the art

Work package 2: Data preprocessing

- **2.1** Video trimming and cropping for cow extraction
- **2.2** Data labeling
- **2.3** Data analysis
- **Milestone 2:** Labeled training data

Work package 3: Gait and Pose analysis

- **3.1** Data annotation for pose estimation
- **3.2** Pose analysis using different features
- **Milestone 3:** Annotated videos with keypoints/skeletons

Work package 4: Automatic locomotion scoring

- **4.1** Design and program of model training
- **4.2** Data training and validation
- **4.3** Error analysis and quality assurance
- **4.4** Model adjustment with different hyperparameters
- **Milestone 4:** Trained models with results

Work package 5: Evaluation

- **5.1** Model testing
- **5.2** Comparison of different approaches
- **Milestone 5:** Evaluation results

Work package 6: Result analysis

- **6.1** Quantitative analysis of models
- **6.2** Plots and tables: method comparison
- **Milestone 6:** Visual evaluation of methods

Wrok package 7: Report

- **7.1** Report draft
- **7.2** Report revision
- **Milestone 7:** Final report

4.2. Project Schedule

The project schedule is shown in Figure 4.1.

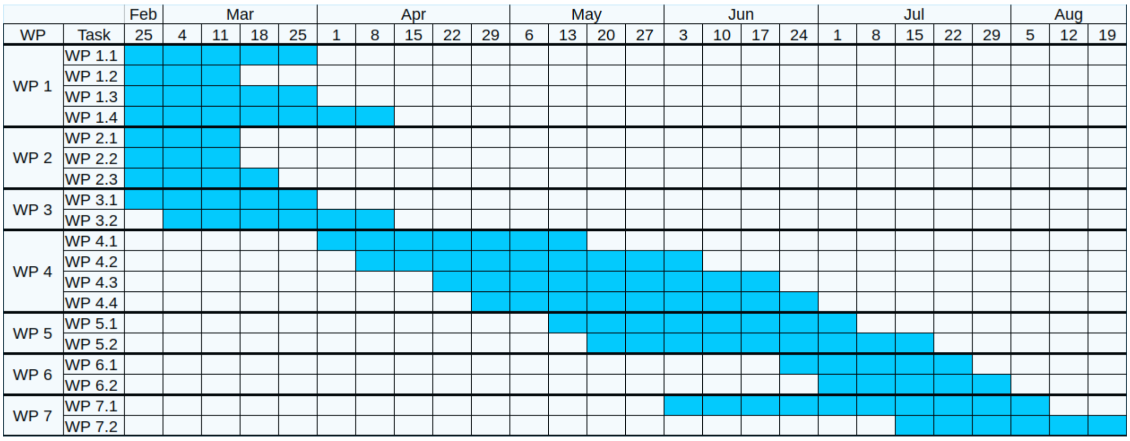


Figure 4.1: Work packages of the project.

4.3. Deliverables

4.3.1. Minimums

- Comprehensive state of the art in lameness detection, gait and pose analysis
- Dataset of cow videos with labels
- Annotated data with keypoints
- Evaluation of features for lameness detection

4.3.2. Expected

- A running automated locomotion-scoring algorithm
- Comparison between traditional and deep learning-based approaches
- Assessment of deep learning in lameness detection

4.3.3. Maximum

- Convincing experimental evidence, that the approach can be adopted to real agricultural business

References

[1] D. Sprecher, D. Hostetler, and J. Kaneene, "A lameness scoring system that uses posture and gait to predict dairy cattle reproductive performance," *Theriogenology*, vol. 47, no. 6, pp. 1179–1187, 1997.

[2] L. Warnick, D. Janssen, C. Guard, and Y. Gröhn, "The effect of lameness on milk production in dairy cows," *Journal of dairy science*, vol. 84, no. 9, pp. 1988–1997, 2001.

[3] H. Whay, "Locomotion scoring and lameness detection in dairy cattle," *In Practice*, vol. 24, no. 8, pp. 444–449, 2002.

[4] K. Abdul Jabbar, "3D video based detection of early lameness in dairy cattle," PhD thesis, University of the West of England, 2017.

[5] A. Van Nuffel, I. Zwertvaegher, S. Van Weyenberg, M. Pastell, V. Thorup, C. Bahr, B. Sonck, and W. Saeys, "Lameness detection in dairy cows: Part 2. use of sensors to automatically register changes in locomotion or behavior," *Animals*, vol. 5, no. 3, pp. 861–885, 2015.

[6] A. Poursaberi, C. Bahr, A. Pluk, D. Berckmans, I. Veermäe, E. Kokin, and V. Pokalainen, "Online lameness detection in dairy cattle using body movement pattern (bmp)," in *Intelligent Systems Design and Applications (ISDA), 2011 11th International Conference on*, pp. 732–736, IEEE, 2011.

[7] F. Schächter, "Automatische Rückenlinienanalyse bei Milchkühen aus Bilddaten," Bachelorarbeit, Rheinische Friedrich-Wilhelms-Universität Bonn, 2017.

[8] X. Song, T. Leroy, E. Vranken, W. Maertens, B. Sonck, and D. Berckmans, "Automatic detection of lameness in dairy cattle—vision-based trackway analysis in cow's locomotion," *Computers and electronics in agriculture*, vol. 64, no. 1, pp. 39–44, 2008.

[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, pp. 1097–1105, 2012.

[10] A. Toshev and C. Szegedy, "Deeppose: Human pose estimation via deep neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1653–1660, 2014.

[11] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, p. 436, 2015.

[12] Y. Feng, Y. Li, and J. Luo, "Learning effective gait features using lstm," in *Pattern Recognition (ICPR), 2016 23rd International Conference on*, pp. 325–330, IEEE, 2016.

[13] S. L. Berry, "Locomotion scoring of dairy cattle." `https://www.vetmed.wisc.edu/dms/fapm/fapmtools/6lame/New5point_locomotionscoreguide.pdf`.

[14] K. C. Mertens, J. Vangeyte, S. Van Weyenberg, S. De, and A. V. Campeneere, "Automatic lameness detection in dairy cows using artificial neural networks," in *Infomation Technology, Automation and Precision Farming. International Conference of Agricultural Engineering-CIGR-AgEng 2012: Agriculture and Engineering for a Healthier Life, Valencia, Spain, 8-12 July 2012*, CIGR-EurAgEng, 2012.

[15] M. Pastell and M. Kujala, "A probabilistic neural network model for lameness detection," *Journal of Dairy Science*, vol. 90, no. 5, pp. 2283–2292, 2007.

[16] T. Van Hertem, S. Viazzi, M. Steensels, E. Maltz, A. Antler, V. Alchanatis, A. A. Schlageter-Tello, K. Lokhorst, E. C. Romanini, C. Bahr, *et al.*, "Automatic lameness detection based on consecutive 3D-video recordings," *Biosystems Engineering*, vol. 119, pp. 108–116, 2014.

[17] M. Alsaaod, A. Schaefer, W. Büscher, and A. Steiner, "The role of infrared thermography as a non-invasive tool for the detection of lameness in cattle," *Sensors*, vol. 15, no. 6, pp. 14513–14525, 2015.

[18] J. Gardenier, J. Underwood, and C. Clark, "Object detection for cattle gait tracking," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2206–2213, May 2018.

[19] S. Viazzi, C. Bahr, T. Van Hertem, A. Schlageter-Tello, C. Romanini, I. Halachmi, C. Lokhorst, and D. Berckmans, "Comparison of a three-dimensional and two-dimensional camera system for automated measurement of back posture in dairy cows," *Computers and Electronics in Agriculture*, vol. 100, pp. 139–147, 2014.

[20] S. Ji, W. Xu, M. Yang, and K. Yu, "3d convolutional neural networks for human action recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 1, pp. 221–231, 2013.

[21] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[22] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[23] M. Ye, Q. Zhang, L. Wang, J. Zhu, R. Yang, and J. Gall, "A survey on human motion analysis from depth data," in *Time-of-flight and depth imaging. sensors, algorithms, and applications*, pp. 149–187, Springer, 2013.

[24] J. K. Aggarwal and Q. Cai, "Human motion analysis: A review," *Computer vision and image understanding*, vol. 73, no. 3, pp. 428–440, 1999.

[25] L. Wang, W. Hu, and T. Tan, "Recent developments in human motion analysis," *Pattern recognition*, vol. 36, no. 3, pp. 585–601, 2003.

[26] T. B. Moeslund and E. Granum, "A survey of computer vision-based human motion capture," *Computer vision and image understanding*, vol. 81, no. 3, pp. 231–268, 2001.

[27] D. Gavrila and L. Davis, "Tracking of humans in action: A 3d model-based approach," in *Proc. ARPA Image Understanding Workshop, Palm Springs*, pp. 737–746, 1996.

[28] G. Mori, X. Ren, A. A. Efros, and J. Malik, "Recovering human body configurations: Combining segmentation and recognition," in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, vol. 2, pp. II–II, IEEE, 2004.

[29] R. Poppe, "Vision-based human motion analysis: An overview," *Computer vision and image understanding*, vol. 108, no. 1-2, pp. 4–18, 2007.

[30] M. W. Lee, I. Cohen, and S. K. Jung, "Particle filter with analytical inference for human body tracking," in *Motion and Video Computing, 2002. Proceedings. Workshop on*, pp. 159–165, IEEE, 2002.

[31] M. W. Lee and R. Nevatia, "Human pose tracking using multi-level structured models," in *European conference on computer vision*, pp. 368–381, Springer, 2006.

[32] R. A. Güler, N. Neverova, and I. Kokkinos, "Densepose: Dense human pose estimation in the wild," *arXiv preprint arXiv:1802.00434*, 2018.

[33] T. D. Pereira, D. E. Aldarondo, L. Willmore, M. Kislin, S. S.-H. Wang, M. Murthy, and J. W. Shaevitz, "Fast animal pose estimation using deep neural networks," *bioRxiv*, p. 331181, 2018.

[34] A. Mathis, P. Mamidanna, K. M. Cury, T. Abe, V. N. Murthy, M. W. Mathis, and M. Bethge, "Deeplabcut: markerless pose estimation of user-defined body parts with deep learning," tech. rep., Nature Publishing Group, 2018.

[35] E. Insafutdinov, L. Pishchulin, B. Andres, M. Andriluka, and B. Schiele, "Deepercut: A deeper, stronger, and faster multi-person pose estimation model," in *European Conference on Computer Vision*, pp. 34–50, Springer, 2016.

[36] W. Liu, C. Zhang, H. Ma, and S. Li, "Learning efficient spatial-temporal gait features with deep learning for human identification," *Neuroinformatics*, pp. 1–15, 2018.

[37] L. Lee and W. E. L. Grimson, "Gait analysis for recognition and classification," in *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on*, pp. 155–162, IEEE, 2002.

[38] A. Leu, D. Ristić-Durrant, and A. Gräser, "A robust markerless vision-based human gait analysis system," in *Applied Computational Intelligence and Informatics (SACI), 2011 6th IEEE International Symposium on*, pp. 415–420, IEEE, 2011.

[39] J. Ortells, M. T. Herrero-Ezquerro, and R. A. Mollineda, "Vision-based gait impairment analysis for aided diagnosis," *Medical & biological engineering & computing*, pp. 1–12, 2018.

[40] J. Wang, M. She, S. Nahavandi, and A. Kouzani, "A review of vision-based gait recognition methods for human identification," in *Digital Image Computing: Techniques and Applications (DICTA), 2010 International Conference on*, pp. 320–327, IEEE, 2010.

[41] M. Kyrarini, X. Wang, and A. Gräser, "Comparison of vision-based and sensor-based systems for joint angle gait analysis," in *2015 IEEE International Symposium on Medical Measurements and Applications (MeMeA) Proceedings*, pp. 375–379, IEEE, 2015.

[42] T. Seel, J. Raisch, and T. Schauer, "Imu-based joint angle measurement for gait analysis," *Sensors*, vol. 14, no. 4, pp. 6891–6909, 2014.

[43] L. Wang, "Abnormal walking gait analysis using silhouette-masked flow histograms," in *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, vol. 3, pp. 473–476, IEEE, 2006.

[44] C.-W. Cho, W.-H. Chao, S.-H. Lin, and Y.-Y. Chen, "A vision-based analysis system for gait recognition in patients with parkinson's disease," *Expert Systems with applications*, vol. 36, no. 3, pp. 7033–7039, 2009.

[45] G. Cortés, "A machine learning approach for gait deficiencies classification," Master thesis, Hochschule Bonn-Rhein-Sieg, 2019.

[46] K. Simonyan and A. Zisserman, "Two-stream convolutional networks for action recognition in videos," in *Advances in neural information processing systems*, pp. 568–576, 2014.

[47] S. Yan, Y. Xiong, and D. Lin, "Spatial temporal graph convolutional networks for skeleton-based action recognition," in *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[48] L. Wang, Y. Xiong, Z. Wang, Y. Qiao, D. Lin, X. Tang, and L. Van Gool, "Temporal segment networks: Towards good practices for deep action recognition," in *European conference on computer vision*, pp. 20–36, Springer, 2016.

[49] J. Huang, V. Rathod, C. Sun, M. Zhu, A. Korattikara, A. Fathi, I. Fischer, Z. Wojna, Y. Song, S. Guadarrama, *et al.*, "Speed/accuracy trade-offs for modern convolutional object detectors," in *IEEE CVPR*, vol. 4, 2017.