Department of Anaesthesiology and Perioperative Intensive Care Medicine
Department for Companion Animals

Development and Validation of an Automated Video Tracking Model for Stabled Horses

Doctoral Thesis

submitted for the fulfillment of the requirements for the degree of

DOCTOR MEDICINAE VETERINARIAE
at the University of Veterinary Medicine Vienna

submitted by

Mag. Med. Vet. Nuray KIL

Vienna, July 2022

SUPERVISOR:

Priv. Doz. Dr. Med. Vet. Ulrike Auer

Clinical Unit of Companion Animals

Department of Anaesthesiology and Perioperative Intensive Care Medicine

University of Veterinary Medicine Vienna, Austria

CO-SUPERVISOR:

Dipl. ECEIM Univ-Prof. Dr. med. vet. Jessika-Maximiliane Cavalleri

University Equine Clinic

Clinical Unit of Equine Internal Medicine

University of Veterinary Medicine Vienna, Austria

Table of Contents

1.Introduction

As defined by the International Association for the Study of Pain (IASP), pain is "an unpleasant sensory and emotional experience linked to actual or possible tissue damage or described in terms of such damage." (Merskey and Bogduk 1994). When talking about animal pain, a common definition was formulated by Molony and Kent (1997). They postulate that pain "is an unpleasant sensory and emotional experience, representing an animal's sense of damage or threat to the integrity of its tissues; it changes the animal's physiology and behavior to reduce or avoid damage, reduce the likelihood of recurrence, and to promote recovery; unnecessary pain occurs when the integrity or duration of the experience is inappropriate to the harm that was done." Pain is a subjective experience that animals are not able to talk about and it has both sensory and emotional components.

With the growing interest in animal welfare, studies on pain related behavior have increased gradually in recent years. Historically, practitioners sought to understand pain through examining physical factors. However, as implied by the preceding description, pain is also an emotional experience that has an effect on the animal's behavior. Behavioral change is an essential metric for assessing animal discomfort and pain, as well as animal welfare. Animals are unable to communicate their feelings verbally, therefore, their behavior is the key factor to understand their physiology. Painful situations can lead to behavioral changes (Price et al. 2003, Van Loon et al. 2010, Lesimple et al. 2012, Dalla Costa et al. 2014, Gleerup et al. 2015). In order to manage pain effectively, one must first have a thorough awareness of what causes it. As a result, assessing behavior is critical not just for identifying pain but also for gauging overall well-being.

Considering the limitations of physiological indications (including heart rate, respiratory rate, biological and hormonal factors), researchers assume that behavior-based pain management is more sensitive and clinically more practical. Recently, subjective and unbiased pain-related behaviors have been investigated in farm, laboratory, and domestic animals (Conzemius et al. 1997, Holton et al. 1998, Cambridge et al. 2000, Molony et al. 2002). Some researchers have also investigated behavioral responses to experimentally induced pain (Kalpravidh et al. 1984, Robertson and Muir 1983). Similarly, in other studies, to investigate the effects of analgesia in colic or post-operative pain, they objectively studied "pain behavior" by scoring certain behaviors over a period of time (Jochle 1989, Jochle et al. 1991, Johnson et al. 1993, Hamm et al. 1995, Raekallio et al. al.1997a,b).

Despite the fact that there are several ways for detecting pain in horses, most of the approaches are difficult to apply in practice due to their proneness to bias and time consumption (Bussieres et al. 2008, Van Loon et al. 2010, Dalla Costa et al. 2014, Gleerup et al. 2015, Wathan et al. 2015). Recently developed automated tracking methods demonstrated that computer-based behavior monitoring is more reliable than direct observation in many animal species, as demonstrated by recent advancements in automated tracking technologies (Bussieres et al. 2008, Van Loon et al. 2010, Dalla Costa et al. 2014, Gleerup et al. 2015, Wathan et al. 2015). Video tracking is locating one or more moving objects over an extended period of time. The goal of video tracking is to establish a connection between target objects appearing in subsequent video frames. By marking an object in the first frame and searching for it in all subsequent frames until the object is marked again. When the objects are moving at a quicker pace than the frame rate, this association becomes more difficult. It is also possible that the tracked item will alter its orientation over time, which would add to the complexity of the task. This is why video tracking systems employ a motion model, which defines how the target picture may change as a result of potential objects moving about. In this way, it is possible to determine the angle of an item based on its three-dimensional location and orientation (https://en.wikipedia.org/wiki/Video_tracking).

In this study we present the development and validation of an automated video tracking model for stabled horses. Our aim was to investigate an automated video tracking model for horses in a clinical context. In our study, it is expected that automated detection of horse movement and therefore behavior in the box stall will be a significant breakthrough. The development followed three stages: annotating the critical spots, training the network to build the model, and finally testing the model for its accuracy.

2. History of Development in Tracking Models

For the computer-aided behavioral studies, there are two main approaches that were found to be effective: Sensor-based and video-based methods. Sensors for body composition, body temperature, heart rate, and respiration rate, as well as electronic noses for health and fertility monitoring, are some of the sensors that can provide crucial information for physiological state, as reported by Frost et al. in 1997. Docking a sensor on an animal, on the other hand, is a specialized, labor-intensive activity on the farm that is never a typical agricultural practice since converters must be maintained close or in touch with one or more precise spots of an animal. Frost et al. (2000) has created a robot sensors positioning system that allows a sensor to remain in touch with any of a variety of specified body postures.

Many different species have been used to evaluate automatic animal monitoring (Brendle and Hoy 2011, Venter and Hanekom 2010). Camera technology, in particular, has made it feasible to track animal behavior in real time and has proved particularly useful for studying group behavior (Pastorelli et al. 2006).

In 1990, a computer-video interface was devised that could quantify filmed motion without requiring physical touch (Morrel-Samuels and Krauss 1990). It constructs artificial X-, Y-, and Z-coordinates around a moving object, according to the research, by juxtaposing front and side views in a single video clip (Morrel-Samuels and Krauss 1990). This two-part picture is overlayed upon the microcomputer's graphics output by a special effects generator (SEG). The user watches the SEG output and uses a computer-generated cursor to trace the movement of the object of interest in real time (Morrel-Samuels and Krauss 1990). It also assesses each monitored movement's start, finish, duration, linear displacement (i.e. length), and velocity. (Morrel-Samuels and Krauss 1990)

After a few years, another way for tracking a fish's three-dimensional location was described by Pereiara and Oliveira in 1994. They employed one camera and a mirror to record, with a virtual picture obtained by a second camera (Pereiara and Oliveira 1994). They placed a mirror at the side of an aquarium to obtain accurate 3D coordinates, which they then used to apply Wolf's (1985) photogrammetry formula. (Pereiara and Oliveira 1994).

The stereoscopic techniques are based on the comparison of two pictures obtained from different perspectives. As a result, numerous strategies have been developed. Cullen et al. (1965) utilized a stereo prism coupled to a camera, some researchers employed various mirror systems, and others have used a pair of picture or video cameras (Pitcher 1973, Pitcher 1975, Symons 1971a and 1971b).

Morris (1981) developed a unique approach for measuring spatial learning and memory in mice in 1981, and this method, known as the Morris water maze, has since been utilized in numerous studies. Morris (1981) has shown that, as long as an object is kept in a stable spatial location in relation to distal room cues, rats can quickly learn to find it even when they cannot see, hear, or smell it. Four groups of rats were given permission to jump out of the dark water and onto a platform that was either fixed or mobile, immediately above or below the water's surface. With the exception of the group whose escape platform was submerged in the sea and which traveled about from location to location, learning happened quickly. Transfer testing showed that the group for whom the platform was submerged but in a fixed place used a spatial location search method. In a follow-up trial, instantaneous transfer was demonstrated when the rats had to approach the platform from a different starting location. The results of both trials were addressed in light of recent rat research on spatial memory (Morris 1981).

A low-cost and reliable computer-aided approach for evaluating the Morris water maze were described. (Santucci 1995). SigmaScan digitizing software (Jandel Scientific, San Rafael, CA) was used to assemble the raw data. Without investing in such a system, a very cheap approach utilizing a regular video camera and digitizing software was developed in an effort to reap some of the benefits of computerized video tracking. Even while the system described in Santucci`s (1995) study was not nearly as complex or versatile as the tracking systems that were commercially available, it simplified and partially automated some of the labor-intensive tasks associated with testing the Morris water maze. More crucially, the current approach made it possible for the experimenter to gather certain data that would otherwise be impossible to obtain. (Santucci 1995)

Computer-assisted environmental systems have risen in popularity as a result of the need for environmental control (Berckmans 1986, McLendon et al. 1990, Allison and White 1991). Bloemen (1997) devised another way to examine animal behavior in their natural surroundings in the late 1990s, which used a camera system and a digital board. Even though some automated systems had been developed, data collection was still difficult since it had to be manually written down or entered into an event-recording computer (Noldus 1991, Noldus et al. 2000).

Some motions in animal behavior, such as speed and distance, cannot be quantified by human observation, automated observation overcomes this constraint (Buresová et al. 1986, Spruijt et al. 1998, Spruijt et al. 1990). Automated observation systems can be used to track both short-term and long-term activity. (Martin et al. 1992, Olivo and Thompson 1988, Spruijt and Gispen 1983, Spruijt and Gispen 1983).

Automated recording of animal behavior and movements has been more common in recent years. Frame grabbers are used in a number of regular video tracking systems to digitize analog video streams. This allows for high-speed data capture and, as a result, tracking of fast-moving animals. However, the majority of these systems have significant flaws. Hard-wired electrical equipment can only record a single animal in a certain location. It cannot distinguish particular objects from each other even if it can monitor many objects. Systems that work with frame grabbers can be functional in simple backgrounds (in terms of gray scale values) and can only deal with a small number of experimental configurations. Color video is likewise not supported by these platforms. EthoVision (EthoVision XT, L.P.J.J. Noldus Information Technology b.v., P.O. Box 268, 6700 AG Wageningen, the Netherlands), unlike some other video analysis systems (e.g., Mukhina et al. 2001, Pan et al. 1996, Spooner et al. 1994), is a general tool that could be utilized in a number of settings and applications. Noldus et al. (2001) introduced the EthoVision video tracking system, which was built to address the limitations of the methodologies and systems outlined above. EthoVision was a video tracking, movement analysis, and behavior recognition system that was used for a variety of purposes (Noldus et al. 2001). It is a robust image processing program that utilizes a high-resolution color video frame grabber and adaptable software to automate behavioral observation and movement monitoring on several animals against a range of diverse backdrops (Noldus et al. 2001). It captures the animal using a CCD (charged-coupled device) camera and then transfers it to a computer where it is digitized using software. It distinguishes the item from the backdrop using either brightness or color. The data is then subjected to extensive statistical processing in order to measure the animal's speed, position, and distance from other animals, among other characteristics. EthoVision was used to investigate the effects of stress on a variety of species, including farm animals (Bokkers and Koene 2000, Sustr et al. 2000), laboratory animals (Ruis et al. 1999), and zoo animals (Fuchs et al. 1996, van Kampen et al. 2000). For instance, EthoVision was used to follow battery chickens as they went toward food. The latency to peck the food and the speed with which the animals approached the meal were measured and used to determine the animals' motivation and ability to move (Bokkers and Koene 2000). Sustr et al. (2000) used EthoVision to color-code the front and rear of pigs. They were then able to assess pig play and pig fights in terms of mutual spatial location and orientation using the EthoVision data. Additionally, EthoVision was sensitive enough to identify age-related motor function losses in non-human primates (Walton et al. 2006).

Lind and colleagues created a technique for autonomously recording pig locomotor behavior (Lind et al. 2005). Adapting approaches for video-based rodent behavior monitoring resulted in a slew of issues. As a result, they have upgraded existing image-subtraction algorithms to provide more flexibility and accuracy while monitoring large animals in scenarios with a continually shifting backdrop (Lind et al. 2005). The enhanced tracking techniques included the creation of an automated threshold detection algorithm and a reference frame that did not contain the animal and was automatically updated. This made the device more resistant to the tracking surroundings, which may have been the same color as the animal and allowed it to alter while recording. As a result, their digital video-based tracking system for automatically monitoring pig locomotor behavior was very dependable and accurate, and it was able to identify well-known apomorphin effects in pig locomotor activity (Lind et al. 2005).

Authors previously shown that camera technology may be useful for tracking and identifying pigs as well as monitoring their behavior (Kashiha et al. 2013b). The goal of Kashiha´s (2014) research was to see whether an automated image processing system could be utilized to identify pig locomotion in a group housing setting and under experimental circumstances. Although this approach should theoretically work with a broad range of light intensity and contrast between the pigs and the ground, background removal and pig body segmentation proved difficult. As a result, the authors proposed utilizing the previous day's data to establish the Image Locomotion Status (ImLS) threshold (Kashiha et al. 2014).

Another study built a camera-based farrowing process monitoring system for sows (Oczak et al. 2016). Because piglet crushing is a significant welfare and economic issue in the pig industry, a method that counts piglets during farrowing is thought to assist to alleviate these issues. Sow behavior was video captured using 2D cameras from the time they were introduced to the farrowing cages until they were weaned in order to establish a data set that could be labeled. Manual labeling of recorded footage was done in order to produce a reference data set from which an algorithm for autonomous farrowing tracking could be constructed. The time of each individual sow's farrowing was labeled in the first phase of the labeling procedure (Oczak et al. 2016). The number of piglets in the pen was successfully estimated using the devised image analysis process and Transfer Function model (Young et al. 2011), with standard errors of 1.73 piglets in the training set and 1.72 in the validation set. Pigs had to be color-coded in order to determine their position and be distinguished from other animals. During farrowing, however, this was not a realistic strategy since it would need a person standing in front of each farrowing enclosure and marking pigs as they are born (Oczak et al. 2016).

3.Behavioral Studies in Horses

Despite the clinical value of a horse's demeanor, there was a scarcity of scientific research on equine pain behavior, especially in other equine species including the donkey, mule, zebra, and wild ass. The focus of this review was on the observation of posture, event behaviors, thresholds to aversive stimuli, lameness assessment grading, and classified pain measures and scoring systems in horses and donkeys. According to Ashley pain evaluation should include all of its dimensions, including severity, frequency, duration, and quality (Ashley et al. 2005).

The behavior-based assessment of pain in horses was first published in 2003 (Price et al. 2003). The study aim was to develop a method for the unbiased detection of post-operative pain in horses. Horses post arthroscopy were compared to pain-free horses and were recorded up to 48 hours with both direct observation and time lapse video recordings. They assumed that analyzes of activity budgets could be a more sensitive method of describing behavioral changes than direct observing (Price et al. 2003).

For the purpose of creating time budgets, which may be used to gauge the well-being of horses in terms of management and housing, observers track the animals' locomotor activity and resting patterns (McCall et al. 1985, Christensen et al. 2002, Heleski et al. 2002, Kurvers et al. 2006, Visser et al. 2008, Werhahn et al. 2012). As an alternative to tedious and time-consuming direct observations, the use of pedometers is prevalent (Holland et al. 1996, Hoffmann et al. 2009, Aguirre and Orihuela 2010, Warren-Smith and McGreevy 2010). As a pedometer disadvantage, data is lost owing to data pooling, which also prevents the difference between different gaits. In the work of Burla et al., an accelerometer (MSR145 data logger) was used to assess its potential for autonomous gait determination (Burla et al. 2014). That`s why Burla et al. (2014), has addressed the objectives included defining stand, walk, trot, and gallop acceleration value ranges and comparing acceleration data with a regularly used pedometer (ALT-Pedometer). ALT-Pedometer showed a positive association between stand and walk acceleration readings, but not trot or gallop, which were monitored with the MSR145 data logger. The MSR145 accelerometer was found to be well suited for determining the horse's gait. As an improved alternative to pedometers for monitoring horse locomotor activity, the MSR145 was demonstrated in research to be a very precise measuring instrument.

Regardless of the fact that animal posture is recognized to indicate emotional condition, no objective instrument for measuring, quantifying, and comparing postures has been developed (Sénèque et al. 2019). An efficient approach of differentiating groups with known differences in welfare states is to utilize morphometric geometrics (GM) to characterize horse posture (Sénèque et al. 2019). Sénèque et al. (2019), studied 85 pictures of riding school horses to see whether a certain posture (modeled by GM) related to changed wellbeing. Prevalence of stereotypic or aberrant repeated behaviors, depressed posture, and ear placements. These horses had a flatter, or even hollow, dorsal profile, notably at the neck and croup. These altered profiles may be used as a field or owner signal of low wellbeing (Sénèque et al. 2019).

According to clinical experience of Torcivia et al. (2020), when people approach or engage with a horse, the horse "perks up" and the discomfort behavior stops for the most part, regardless of the patient's continuous discomfort behavior. This apparent propensity to cease evident discomfort behavior in the presence of humans has been addressed in the literature, but seems to be underappreciated in horse clinical practice (McDonnell 2005, Gleerup et al. 2016). These objective pain scoring techniques that involve behavioral evaluations are especially worrisome in that they are generally performed during an in-person caregiver visit, which necessitates direct contact with the animal (Bussieres et al. 2008, Price et al. 2003). It has been thought to evaluate the continuous discomfort behavior while individuals are present (Torcivia et al. 2020). Therefore, the goal of Torcivia`s research was to quantify the impact of a caregiver visit on the continued discomfort behavior of horse patients. A propensity to halt obvious discomfort behavior when humans are around has been noted, but it did not seem to be frequently acknowledged in horse therapeutic practice (McDonnell 2005, Gleerup et al. 2016, Torcivia et al. 2020). Even more worrisome was that current state of the art objective pain scoring techniques that incorporate behavioral evaluations were generally performed during in-person caregiver visits, requiring direct connection with the horse (Bussieres et al. 2008, Price et al. 2003). Analysis of the clinical perspective of people interfering with chronic pain behavior would therefore be essential. When assessing pain, one has to be aware of this considerable decrease in apparent pain behavior when a caregiver or visitor is present and the subsequent possible underestimating of discomfort and the potential influence on clinical case management.

Equids are naturally prey animals and exhibit the fewest pain signals to deter predators. The urge to conceal pain is still there in domesticated horses, and people may interfere with ongoing pain behavior. The horse may only sometimes and subtly express pain, which could make the damage go unnoticed. (Torcivia et al. 2020). Orthopedic issues are widespread in horses and often result in euthanasia. These diseases may cause severe discomfort. It's difficult to train a spatial pain detection system using video data since the pain behavior is modest. Even for an experienced practitioner is difficult to get relevant data. In a freshly published paper, researchers revealed that a model trained exclusively on horses with acute experimental pain can detect more subtle orthopedic pain presentations. They also evaluated how well various strategies operate in other areas and how well a system trained on clean experimental pain in the orthopedic dataset detects pain. They also discussed dealing with real-world animal behavior datasets and doing fine-grained action recognition (Broomé et al. 2022).

4. Deep Learning

The field of machine learning, called the Neural Network (NN), has created the field of deep learning. Learning is a process of estimating model parameters in order to make the learned model (algorithm) better able to carry out a task (Alom et al. 2018). A deep learning system consists of several layers between the input and output layers, which allows for the incorporation of numerous non-linear information processing units with hierarchical architectures for learning and classifying features. (Schmidhuber 2015, Bengio et al. 2015). There are a variety of deep learning approaches, each with its own strengths and weaknesses. Supervised, semi-supervised, and partially supervised learning are the most common, while unsupervised learning is less common but still has its own benefits. Supervised learning is a learning technique that uses labeled data.There are different supervised learning methods for deep learning, including deep neural network (DNN), convolutional neural network (CNN), recurrent neural network (RNN), including long short-term memory (LSTM), and gated recurrent unit (GRU). CNN structure was first proposed by Fukushima in 1988, but it did not become popular until later when better computing hardware was available. In the 1990s, Lecun et al. (1998), conducted research on the effectiveness of artificial intelligence. Applied a gradient-based learning algorithm to CNNs and achieved successful results for the handwritten digit classification problem (Lecun et al. 1998). Next, the researchers optimized the CNNs and reported the latest results in several recognition tasks (Alom et al. 2018).

Convolutional neural networks have been the most effective type of image analysis models to date (CNNs). CNNs have numerous layers that, to a limited extent, use convolution filters to transform their input.

In order to create 'deep networks,' simple units are arranged in layers and then serially stacked to form deep neural networks (DNNs). As a result of training on data, the connections between the units have learned how to glean information from unstructured data in order to complete tasks (Mathis 2020).

With the development of sensors in mobile devices, human activity analysis has become common in areas like healthcare and fitness tracking. This approach has spread to equestrian sports since tracking behavior may provide valuable information about a horse's health (van Loon and van Dierendonck 2015). Among the equine behavioral researches many technologies have been created to monitor characteristics such as activity, elevation, heart rate, and so on (Langrock et al. 2012, Burla et al. 2014, Bidder et al. 2014). However, additional behaviors including rolling, pawing, and flank watching have not yet been investigated using wearable accelerometers (Burla et al. 2014). Researchers advise using machine learning technologies to identify accelerometer data more correctly. The offered approaches still need feature extraction. Constant computational power allows a convolutional neural network (CNN) to automatically extract features. Deep learning-based classifiers can learn features. According to some researches, deep learning models have been shown to be capable of learning and discriminating between human activities such as sitting, walking, climbing upstairs, walking downstairs, and falling, among others, but have not yet been applied to the detection of equine activities. (Ravi et al. 2017, Eerdekens et al. 2020). Eerdekens et al. (2020) suggested an experimentally verified CNN to automatically distinguish seven separate horse behaviors using data from two accelerometers by examining the effect of sampling rate, time series length and underground. The findings showed that the suggested CNN-based model has good accuracies across a wide range of time periods. Less than 1% of the data from one accelerometer seems to be adequate to categorize horse behaviors.

Speed was another important factor in analyzing and evaluating horse movement and behavior (Meira et al. 2014, Witte et al. 2004). Global positioning systems (GPS) or inertial measurement units (IMUs) have been used to assess speed (Williams et al. 2019, Farries et al. 2019, Kingston et al. 2006, Bazzano et al. 2016, Fonseca et al. 2010, Parkes et al. 2019, Vermeulen et al. 2006, Han et al. 2020, Best et al. 2019). Estimating horse speed might be done using GPS speed as a ground-truth and IMUs for mobility. However, IMU data is multi-dimensional, making a link to one-dimension linear speed data difficult. By transforming non-linear and high-dimensional data into an optimal model, machine learning (ML) solves this problem (Phinyomark et al. 2018). As a result, with the goal of assessing horse speed, Darbandi et al. (2021) has utilized the body-attached IMUs and machine learning approaches. The impact of the number and position of IMUs, ML techniques, gaits, and breeds (with various movement patterns) on the estimate model were explored and contrasted in order to produce an accurate and comprehensive model. This research produced three unique results. First, they found that inserting just one IMU on any limb, withers, or sacrum resulted in good speed estimate accuracy. Second, its algorithm can estimate speed for two breeds and five gaits. Finally, five popular ML approaches for assessing horse speed were evaluated.

Previously, motor control experiments were carried out with the use of markers, which is an intrusive procedure. DeepLabCut (https://github.com/DeepLabCut), a deep learning approach with markerless posture estimation, was developed as a result of this (Mathis et al. 2018). To train the dataset, it was necessary to separate frames from numerous videos and manually identify certain body parts. To elicitate the labelled data and train the network to the desired characteristics, the toolkit employs DeeperCut detectors. DeepLabCut can then be applied to new recordings after a few hours of labeling and network training (Mathis et al. 2018). The study of Mathis (2018) used the DeeperCut posture estimation model based on the ResNet architecture (Insafutdinov et al. 2016, He et al. 2016). This option is beneficial since it allows for transfer learning by using a pre-trained model, but it is also devastating because the model has over 25 million parameters. Overparameterization allows the model to make accurate predictions at the price of sluggish inference. Insafutdinov et al. (2016) found that faster inference time for DeeperCut may be achieved at the expense of higher prediction error. Beginners may struggle with Mathis et al. (2018)'s early efforts, however recent software improvements have aimed to address this (Nath et al. 2019).

Lameness is another important issue that should be focused on early diagnosis. Studies with trackable markers and high-speed cameras are time-consuming (Keegan et al. 2011). As the deep learning becomes popular it was thought to be used to track motion of horses with lameness. Wang et al. (2021) trained a neural network to detect the bodily components of a horse and then assessed their motion trajectory to determine whether the animal was lame. Using two state-of-the-art convolutional network models, they investigated several strategies for estimating animal poses (Eederkens et al. 2020, Wang et al. 2021). They demonstrated a marker-free technique using DeepLabCut that utilizes ResNet (pre-trained network) 50 (number of layers) to execute deep learning on hundreds of photos labeled with user-defined equine body parts (Nath et al. 2019, Wang et al. 2021). They discovered that the trained model was capable of properly identifying equine bodily components. Based on the identified body components, they conducted posture estimation and movement analysis. This study established that it is certainly possible to accurately assess whether a horse is lame from video footage using deep learning (Wang et al. 2021).

The complexity of behavioral data and analytics has raised the necessity for automated and efficient tracking systems. Pereira et al. (2019), then proposed the LEAP (LEAP Estimates Animal Pose), a deep-learning-based approach for predicting animal body component placements. With this framework, you can name body sections and train the network visually. LEAP predicts new data quickly, and training with just 100 frames achieves 95% peak performance. They tested LEAP on footage of free-behaving fruit flies, tracking 32 unique locations to depict the head, body, wings, and legs. In summary, they demonstrated how to monitor moving animals' bodily parts with little human effort and without physical marks. Graving et al. (2019), then with a simpler interface, outperformed the existing algorithms from Mathis et al. (2018) and Pereira et al. (2019). They developed a novel deep learning framework for evaluating animal posture and utilized convolutional neural networks to obtain quick inference without compromising accuracy (Graving et al. 2019).

In biomechanics, and neuroscience, extracting animal positions without utilizing markers is typically necessary for evaluating behavioral consequences. Extracting detailed postures without markers in constantly shifting backgrounds, on the other hand, has proven difficult. Nath et al have released the DeepLabCut, an open-source toolkit that uses a state-of-the-art human pose-estimation algorithm to enable users to train a deep neural network with minimum training data to accurately follow user-defined features with human-like labeling accuracy. Authors presented an upgraded toolbox, written in Python, with additional features including graphical user interfaces (GUIs), speed enhancements, and activelearning-based network refining. They described a step-by-step technique for utilizing DeepLabCut that leads the user through developing a customized, reusable analytic pipeline using a GPU in 1–12 hours (depending on frame size). They also provided Docker environments and Jupyter Notebooks that may be used on cloud resources like Google Colaboratory.

They provided a complete technique and DeepLabCut extension that enables researchers to predict a subject's position, allowing them to quickly quantify behavior. The DeepLabCut toolkit was created with the goal of providing a reliable and fast tool for high-throughput video analysis, in which strong feature detectors of user-defined body parts must be trained for a given context. The toolkit aimed to handle the challenge of recognizing body parts in dynamic visual situations when the presence of a changing backdrop, reflected walls, or motion blur makes typical approaches like thresholding or regression based on visual cues ineffective (Dell et al. 2014, Anderson et al. 2014, Egnor et al. 2016, Dollár et al. 2010, Gomez Marin et al. 2012, Matsumoto et al. 2013, Uhlmann et al. 2017, Ben-Shaul et al. 2017). DeepLabCut is best suited to actions that can be caught continuously by one or more cameras with low occlusions. DeepLabCut would be used to analyze behaviors with intermittent occlusions since it conducts frame-by-frame prediction.

Quantitative behavioral assessments are significant in domains ranging from neurology to ecology. Although recently published animal posture estimation systems were designed to minimize human participation, physical intervention was sometimes necessary to fix problems (Mendes et al. 2013, Kain et al. 2013, Uhlmann et al. 2017). It took a lot of work to adapt any of these tactics to new conditions.

The ongoing deep learning uprising has started with the aim of achieving human-level accuracy for object detection. DeepLabCut has been proved to reach human labeling accuracy. (Mathis et al. 2018). While computer-based video tracking has improved gradually some challenges still remain. Mathis et al., has predicted that the fast-paced development of new deep learning tools would fastly change the scenery of redeemable real-world neuroscience. (Mathis et al. 2020)

So far there have been different algorithms used in video tracking. Those are background substruction, adaptive thresholding, image recognition and deep learning. With the increase number of studies on tracking the need for new software for video processing are also increased. Recently a new software called ThruTacker has been introduced to serve the purpose. TruTracker is an open-source software program for 2D and 3D video tracking. Its features are calibration of intrinsic properties of cameras, tracking and counting hundreds of animals at the same time. (Corcoran et al. 2021)

5. Publication

Development and Validation of an Automated Video Tracking Model for Stabled Horses

Kil N, Ertelt K, Auer U. 2020. Animals 10(12), 2258; https://doi.org/10.3390/ani10122258

animals

MDPI

Article

Development and Validation of an Automated Video Tracking Model for Stabled Horses

Nuray Kil *, Katrin Ertelt and Ulrike Auer

Department of Anaesthesiology and Perioperative Intensive Care Medicine, Department for Companion Animals, Vetmeduni Vienna, 1210 Viena, Austria; Katrin.Ertelt@vetmeduni.ac.at (K.E.); Ulrike.Auer@vetmeduni.ac.at (U.A.)
* Correspondence: Nuray.Kil@vetmeduni.ac.at

Received: 16 November 2020; Accepted: 27 November 2020; Published: 30 November 2020

check for updates

Simple Summary: Although there are some methods to detect pain in horses, because of bias and time-consumption, those methods are practically challenging. However, in recent years rapidly developed automated tracking methods have proven that computer-based behaviour monitoring is more reliable in many animal species. That is why in this study we aimed to investigate an automated video tracking model for horses in a clinical context. The findings will help to develop the automated detection of daily activity, to meet the ultimate objective of objectively assessing the pain and wellbeing of horses. An initial analysis of the obtained data offers the opportunity to construct an algorithm to track automatically behaviour patterns of horses.

Abstract: Changes in behaviour are often caused by painful conditions. Therefore, the assessment of behaviour is important for the recognition of pain, but also for the assessment of quality of life. Automated detection of movement and the behaviour of a horse in the box stall should represent a significant advancement. In this study, videos of horses in an animal hospital were recorded using an action camera and a time-lapse mode. These videos were processed using the convolutional neural network Loopy for automated prediction of body parts. Development of the model was carried out in several steps, including annotation of the key points, training of the network to generate the model and checking the model for its accuracy. The key points nose, withers and tail are detected with a sensitivity of more than 80% and an error rate between 2 and 7%, depending on the key point. By means of a case study, the possibility of further analysis with the acquired data was investigated. The results will significantly improve the pain recognition of horses and will help to develop algorithms for the automated recognition of behaviour using machine learning.

Keywords: equine behaviour; image processing; automated video tracking; machine learning; pain assessment

1. Introduction

It is currently well established in veterinary medicine that pain triggers behavioural changes in animals, and their monitoring becomes relevant in the assessment of pain and in the evaluation of an animal's welfare state. Detailed knowledge of both normal and pain-related behaviours in equines is imperative to properly evaluate pain. Although the presence of strangers or unfamiliar surroundings may mask pain-related changes, even subtle variations may become apparent if behaviour is thoroughly analysed [1].

In horses, pain is typically scored manually. Various pain assessment scales, such as the composite pain score [2,3] and the horse grimace scale [4], have been developed and proven useful in the assessment of postoperative pain. Horses have rich facial expressions based on 17 facial action units recently decoded [5]. Five of these units change in a typical manner during painful stimuli and interventions [4,6]. Composite pain scores include variables such as specific pain behaviour or posture that are scored using defined classes by means of simple descriptive scales [3,7]. A recently developed ethogram for pain scoring in ridden horses was introduced to determine musculoskeletal pain [8]. The fact that posture reflects the emotional state of a horse led to the attempt to use geometric morphometrics to assess back pain and poor welfare [9,10]. Yet, access by measuring activity patterns over a longer period, which changes in response to acute postoperative pain, seems to be more sensitive than composite pain scores [1].

All the methods have a limitation: they only observe the horse for a very short period. Executing repeated pain assessments increases the risk for noncompliance in clinics. Despite good validity for discrimination between not-in-pain and in-pain horses, inexperience in the common behaviour of a horse can lead to inconsistent results. This increases the risk of underestimation of pain or misinterpretation of behaviour in context of welfare and quality of life assessment.

Traditionally, behavioural studies and also activity pattern studies have been carried out manually or semi-manually with the limitation of subjective human observation [11–13]. Human observation to score behaviour imposes a number of limitations, such as the speed of scoring behaviour on video, the precision and recognizing a behavioural pattern that is rapid and variable [14]. Moreover, the manual analysis of such videos that were recorded over long periods (6 to 24 h) carries the disadvantage of being time-consuming and cumbersome.

The introduction of computational ethology and biotelemetry overcomes these limitations [15,16]. Recent developments in computer vision, image processing and deep learning have improved research in behaviour sciences [17–19]. The developments in veterinary medicine in recent years have increasingly moved in the direction of replacing humans with their subjective and often non-reproducible assessments of behaviour—for example, regarding lameness in horses—with computer technology. Recently, studies on activity measurement and gait analysis using accelerometer data and convolutional neural networks have been published [20,21]. The positive results of these studies suggest that artificial intelligence should also be used for continuous observation of a horse's behaviour. The use of video-based, automatic image analysis has the advantage of not only categorizing the behaviour automatically by means of an algorithm, but also of being able to review the video in case of ambiguities and deviations. These computer-aided analyses are not only interesting for questions in behavioural research but also for veterinary medicine for a more objective evaluation of pain and animal welfare.

The first video-tracking systems were introduced in the early 1990s. Due to their popularity for laboratory study, image analysis has been used extensively for video tracking of rodents under laboratory conditions [22]. Various automated methods—many prototypes—based on image subtraction, grayscale threshold, statistical models or colour tracking, have been used to study animal performance [22–24]. The advantage of image analysis is that it allows monitoring in a non-invasive way, without human presence or with limited human presence [25]. This is especially important for achieving high accuracy of pain assessments, as the animals feel unobserved and are less likely to mask pain behaviour.

The purpose of image analysis is to extract the information automatically to reduce manual workload, to increase objectivity and to quantify changes which are too slight for the human eye [26]. It allows one to study behaviours that occur briefly and are interspersed with long periods of inaction [27] or occur over many hours, such as diurnal variation.

Image analysis methods that have been developed so far for monitoring livestock, rodents and other animals are mostly based on "manual engineering of features," also known as feature engineering (FE), followed by classification or regression modelling [27–32]. This differs from image analysis with deep learning techniques, developed recently for monitoring behaviour in different species [18,33–36]. Developments of deep learning techniques have introduced a new approach to image analysis in which feature engineering before classification or regression is not necessary, but rather, a "direct from video" approach is used. In this approach, the trained algorithm uses the raw video data to identify the animal and automatically detects the behaviour or action of an animal.

The objective of this study was to evaluate how the video-based automatic tracking tool performs in the recognition of activity of stabled horses in a hospital setting. In addition, we give an example of how the deployment of the model could detect the behaviour of a horse in pain.

2. Materials and Methods

2.1. Horses

A total of 34 horses were used for this study. All horses were patients of the university's equine teaching hospital and were housed in 3.5 meter-squared box stalls with free access to water, and roughage feed four times per day. Criteria for inclusion of horses in the study was different fur colour, fur badge, body size and shape and background, along with the incidence of light for the box. Informed owner consent was given prior to the start of the study. The university's ethics committee and data protection committee approved the study.

2.2. Camera Setup and Video Recordings

An action camera with a broad wide-angle (Gopro Hero 4, Gopro, Inc, San Mateo, CA, USA) mounted at a height of two meters in the right or left corner on the front side of the box stall was used. The camera was connected to a power bank for continuous power supply. Cameras were set up to see the whole area of the box stall. To reduce data space, recordings were taken in time-lapse mode (TLV) (two images per second) with a resolution of 2.7K. In total, 65 videos were used. In order to train and test the neuronal network, initially eight horses were taped over 24 h (in total 39 videos). Twenty-six horses were recorded only during daytime (in total 26 videos) and used for the validation process.

2.3. Image Analysis

For pose detection, we used the convolutional neuronal network Loopy (http://loopb.io, Loopbio GmbH, Vienna, Austria).

First, videos were uploaded to the server of Loopy and split into a training set of eight videos of eight different horses. The process of image analysis started with the labelling of markers on video files of 11.52 min corresponding to 21,342 images. In the selection of the video, the main focus was not on specific behaviour; instead, it was on the colour of the horse and the background of the box stall. Every 10th frame of the first 2500 frames (in total 249 frames per video) was chosen and prepared in the annotation section of Loopy. For each horse, the following anatomical landmarks were marked manually following the rules of good data in the documentation of the software. Three markers were annotated in each frame and the visibility of each marker classified as visible; 25, 50 or 75% occluded; or not visible. The annotated marker "wither" was defined as the highest point in the shoulder area where the neck merges into the back and the mane ends; "tail" was the area of the tail base; and "nose" meant the area including the mouth and nostrils.

The information in the dataset about the markers' locations and whether they were visible or occluded, was entered into the deep-learning training tool for pose tracking to generate version 1 (V1)

Subsequently, with this version (V1) we ran a prediction on the remaining, not in the training set, used images. The results of the prediction are provided in Loopy as a video with the predicted markers, as a csv file for download and further analysis, and in a plot selection. The video generated by the software showed image by image the markers for nose, withers and tail. These markers were checked for correctness and mislabelling by means of visual inspection. Mislabelling was defined as a missing prediction of the visible markers over more than 30 consecutive images or repeated missing and repeated misplacement of a marker. Each sequence with misclassifications or missing predictions of visible markers over more than 30 consecutive images, or repeated missing and misplaced markers on wrong body parts or objects in the box, was identified. These sections in videos were manually re-annotated and incorporated in a new training process.

2.4. Assessment of Robustness

The robustness of version (V2) was tested with the remaining unlabelled 31 videos of the 8 horses. After downloading the data with the pixel information of the markers, the numbers of images with predicted markers (PM) for nose, withers and tail (PMn, PMw, PMt) divided by the sum of all images (TI) per video were calculated.

In this phase of the model creation, the misplacement of markers, defined as a difference of more than 200 pixels between the x and y pixels of each marker in two consecutive video frames, was checked mathematically. To detect these jumps the formula,

$$\sqrt{((X2 - X1) \times (Frame2 - Frame1))^2 + ((Y2 - Y1) \times (Frame2 - Frame1))^2} \quad (1)$$

Was used and data were visually inspected and compared to the corresponding graphical representations. The number of wrongly predicted images, referred to as wrongly predicted (WP), was documented along with the reason for the mislabelling or non-labelled frames.

Re-annotating of the mislabelled section was repeated and version 3 (V3) was generated with the additional annotated images.

Model V3 was validated on 26 videos of different horses recorded during the daytime, which had not been annotated formerly. The mathematical analysis of mislabelled markers was repeated. A detailed performance analysis included the observation of the first and last 1000 images from the videos for counting each marker with the following classification: true-positive marker (TP = correctly predicted marker), true-negative marker (TN = correctly not predicted hidden marker), false-positive marker (FP = mislabelled or predicted but not visible) and false-negative marker (FN = visible but not predicted). With these figures, sensitivity was calculated utilizing the formula

$$TP/(TP + FN) \quad (2)$$

To assess the error rate, FP/(TP + FP) was determined. The accuracy defined as the percentage of correctly classified markers (TP, TN) was calculated with the formula of (TP + TN)/(TP + FP + TN + FN).

2.5. Statistical Data Analysis

For the comparison of the V2 and V3, a Student's t-test was utilized for the total sum of PM, WP, the ratio of PMn, w, t/IE and PMn, w, t/WP for each video and marker. A one–way analysis of variance (ANOVA) was carried out to compare the quality of the prediction (sensitivity, error rate and accuracy) for all three markers. Statistical significance was accepted at $p < 0.05$. Statistical analysis was performed using NCSS 2020 (Statistical Software (2020)—NCSS, LLC, Kaysville, UT, USA, ncss.com/software/ncss.

2.6. Case Study to Demonstrate

Two videos of a horse recorded between 7 a.m. and 11 a.m. on day 2 post colic surgery and on postoperative day 8, respectively, were chosen to demonstrate the usability of the automated video tracking tool assessing the behaviour. A detailed analysis of the outcome parameter x and y pixel coordinates of the marker "nose" was done with the plot function in Loopy.

A position scatter plot; a position heat map; a time spent in region calculation; and an x, y time series plot were generated and downloaded. The behavioural patterns observed with an ethogram, including behaviour classification for (1) standing observation, (2) sleeping or rest, (3) feeding and (4) moving, were scored image by image in the behavioural scoring tool in Loopy.

Both the behavioural score and the x, y pixel value were downloaded as csv files. Both curves, behaviour score and timeline of x and y pixels, generated in Excel, were plotted together and represented in a graph.

3. Results

The analysis of version 1, generated with 1990 images, revealed a continuous and accurate tracking of the three markers, at least in images similar to the annotated one, and a good resolution during movement of the horse (see Figure 1).

Figure 1. Screen shot of a video of the horse from the case example with the visible predicted marker nose (green), withers (red) and tail (blue).

The visual video inspection of V1 resulted in reannotation and additional annotation of 1548 images in mislabelled sections to generate V2. The quality check of V2 revealed good overall performances for the markers "nose" and "withers" with mean PM/TI ratios of 0.7 and 0.4. The percentages of wrong markers were 1.1%, 3% and 2.8% for withers, tail and nose, respectively. In a number of 10 videos of four horses, video sequences with more than 2% WP for the marker withers could be detected. In 19 videos of five horses, the marker nose was predicted wrongly between 2 and 8% of the time. The percentage of WP for tail was similar to that of nose despite one horse with more than 20% WP in all five videos.

The reasons for mislabelling were people's interventions, lack of lighting conditions, horses' angles to the cameras and blanket wearing. An additional annotation process increased the labelled images to 6244 (nose), 7276 (withers) and 5181 (tail). The resulting V3 performed with a significantly higher PM/TI ratio for all three markers and had significantly fewer wrongly predicted markers (see Table 1).

Table 1. Comparison of version 2 and version 3 in regard to total sums of predicted markers (PM) and wrongly predicted markers (WP), the ratio (PM/TI) of PM divided by total image (TI) and the percentage of wrongly predicted markers (%WP) in relation to number of predicted markers per video given as mean ± SD. Presented are data of the 39 training and test videos.

Version	Marker	TI	PM	WP	PM/TI	% WP
	Withers		471,544	7813	0.7 ± 0.2	1.8 ± 1.5
V2	Tail	704,608	288,499	4366	0.4 ± 0.2	3 ± 2.8
	Nose		511,817	12,908	0.7 ± 0.2	2.8 ± 2.3
	Withers		611,183	2561	0.9 ± 0.2	0.4 ± 0.3
V3	Tail	704,608	495,963	2985	0.7 ± 0.2	1.2 ± 2.2
	Nose		542,390	7541	0.8 ± 0.3	1.5 ± 1

The mean values for sensitivity, accuracy and error rate of the 26 videos that were not included in the training process are given in Table 2, The detailed results per horse/video can be seen in Table S1.

Table 2. Mean ± SD of performance criteria for machine vison technics.

Marker	Sensitivity	Accuracy	Error Rate
Withers	0.88 ± 0.2	0.82 ± 0.3	0.02 ± 0.03
Tail	0.79 ± 0.3	0.82 ± 0.2	0.06 ± 0.20
Nose	0.94 ± 0.1	0.94 ± 0.1	0.07 ± 0.17

Case Example

The positions of the study horse during four hours in the mornings of days 2 and 8 postoperation are shown in Figure 2 as a heat map and in Figure 3 as a position scatter plot. The calculations of the preferred location in the box were different for the two days. On day two, the horse spent 33% of time looking to the sidewall compared to 9% on day eight.

Figure 2. Heat map of the horse on day 2 and on day 8 post colic surgery. The horse spent 67% and 91% of the time in front of the box on days 2 and 8 respectively. Food was provided early in the morning and recording started at 06:30.

Figure 3. Position scatter plot of the horse on day 2 after colic surgery, and day 8. This plot gives additional information about the movement in the box over time in comparison to Figure 2.

The comparison between the curve x for any pixels over four hours revealed a good accordance of the variance of x, y pixels with the corresponding behaviour score. (Figures 4 and 5). A higher variance of mainly the y-coordinate during the resting periods could be also observed and seen in the video. This pattern could connected to an unsettled standing with frequent weight shifting and of the extension of one forelimb while reducing weight on that limb. This specific behaviour was seen continuously on day 2 during rest but was less noticeable on day 8.

Figure 4. x, y time series plus behaviour score of the horse on day 2 post colic surgery. The oscillation mainly in the y pixel could be revealed as a behaviour during resting classified as unsettled standing with frequent weight shifting and extension of one forelimb while reducing weight on that limb.

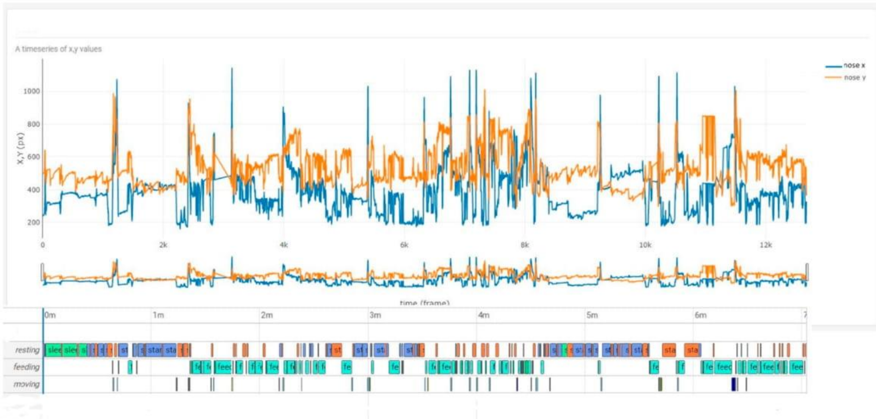

Figure 5. x, y time series plus behaviour score of the horse on day 8 post colic surgery. Weight shifting and extension of the forelimb was less frequent but still observable.

4. Discussion

This study presents a model for automated video tracking of horses to estimate behavioural changes in a clinical context. Annotating a small number of images was sufficient for the deep learning training of a predictor, which was able to mark three different body markers on time-lapse videos. The results revealed a high accuracy and sensitivity of more than 80% in pose estimation. An initial analysis of the obtained data offers the opportunity to construct an algorithm to automatically track the behavioural patterns of horses.

Several open source video tracking software systems with special hardware requirements were recently reported [17,19,35]. The Web-based tool Loopy offers a complete set of image processing, tracking and behavioural analysis tools. The software includes a number of general-purpose algorithms for tracking the positions of subjects in videos. It offers custom training for a deep learning tracking solution for single and multiple animal-pose tracking. In the author's opinion, the advantages are that Loopy is easy to use, does not need a special processor and computers requirements and always provides the latest technology in this area. As a user, no special knowledge in IT and coding languages is necessary. To generate a tracking model, an annotated ground-truth dataset for training that is generated with the annotation tool must be provided. The complexity of the scene, the visual similarity of individuals and a detection error tolerance determine how many training data are required to obtain a good result.

The decision to use the action cameras was based on their wide angle, flexible handling and easy installation. The quality of the video recordings and the time-lapse mode have proven to be suitable for generating the deep learning model and did not represent an obstacle for the prediction. Images can be recorded via simple camcorders [30], 3D cameras [36] or CCVT cameras [37]. Bonneau et al. (2020) used time-lapse cameras for object detection (goats in the field). The advantage of using action cameras in time-lapse mode is the low amount of memory required space [35].

In a first approach we decided to annotate only three markers on the horse. To have a better view of the behaviour of the horse, we decided to choose a camera angle of 60 degrees instead of a top down view. The markers for nose, withers and tail were almost always visible from this perspective. Only the tail marker was often not visible due to the preferred position of the horse with the head looking to the door, and therefore it was not predictable. The first version resulted in a nearly perfect prediction of all visible markers in images similar to the annotated one. In contrast to the position of the horse, the colour of the coat and the influx of the light had no negative effect on the prediction quality and accuracy in V1. The markers were also recognized during fast movement. An increase in the number of annotated images brought a clear improvement in terms of the position of the horse to the camera. With the second version we could concentrate on detecting the incorrectly labelled markers by means of mathematical analysis. This method, together with an optical control of the affected video section and the corresponding graphical representation, made it possible to identify the wrongly predicted marker. With this method, however, it is not possible to clearly distinguish between the false-positive predicted markers and the true-positive markers. The latter can show a greater distance between two images when the horse moves. We chose this method primarily to document an improvement between V2 and V3.

By further annotating images, we were able to achieve a significant reduction in the number of incorrect markers with version 3. The step-by-step procedure for improving the model is also described for other systems [38]. As was done here, it is also recommended that the videos be split into a training set and a test set—considered the gold standard in the development of deep learning models [18,19,34]. Following the recommendation of Graving et al. 2019 for machine learning models, we carried out the actual validation of V3 on videos that were not included in the annotation process [18].

Sensitivity and specificity are two statistical measures of the performance of classification test in the medical field but also for image processing [35–39]. Accuracy was used to evaluate the performance of automated activity tracking of goats using drones [40]. To the best of our knowledge, the present study is the first to deal with automated video tracking in horses. The values for accuracy and sensitivity are quite good and in line with the outcome of other deep learning tools [1,19]. Although the prediction rates of markers were low for some horses, additional annotation of the misclassified section and including in a new training model will improve the prediction for these horses. The main cause for misclassification was found to be the horses' special features. For example, one of the horses had a brown coat colour but the nose area was partly brown and white; that is why the marker was classified as a false-negative. Automatic waterers, piles of manure and shoes belonging to people were often marked as a false-positives, especially in V2. However, further improvement of the model largely solved this mislabelling with V3. In general, however, it must be questioned whether incorrect predictions are not tolerable to a certain extent. The image resolution of our time-lapse video was two images per second. This means that 90 images of the TLV correspond to a real time of one minute. Mislabelling of a few images will have little influence on further analysis of behaviour and calculation of activity time budgets.

No doubt, if the percentage of incorrect markings is high in future predictions due to special coat colour, position of the horse or background colour, these sections should be annotated, and further deep learning training should follow to improve the model. This procedure is also recommended for other tools and in the documentation of Loopy software.

Based on the case study, we could get insights into the possibilities that arise with the data resulting from the automated video tracking. For example, the position of a horse in the box in relation to the door can be determined over a longer period of time; thus, validation of this parameter that occurs in many composite pain scales is possible with this method [7,41]. With the help of manual video analysis, Price et al. (2003) were able to determine that horses in the postoperative phase were standing with their heads in the back of the boxes more [1]. On day eight after the colic surgery, our horse stood with its head to the sidewall much less often than two days immediately after surgery. The apparent variance of the x, y pixels clearly can be assigned to different behavioural categories. Special behaviour such as frequent weight shifting during rest could be detected with the marker nose. Additional annotation of other markers, such as the hooves or ears, will improve the observation of behaviour. The realization of an algorithm for automated behavioural recognition is very promising. The automated calculation of time budgets for behaviour classes will be possible without the need for manually scoring the behaviour. In a further step an activity time budget can be related to a pain score to test the hypothesis of Price et al. (2003): whether this approach to pain recognition is more sensitive than pain scores [1].

The limitations in this study are primarily to be found in the video recording. For a continuous evaluation of the behaviour during the night, cameras with infrared technology are better suited than the action cameras chosen here. This limited the possibility of calculations within 24 hour time budgets.

5. Conclusions

The results demonstrate that the automated tracking of body parts using time-lapse videos is possible. The information generated with the deep learning module can be used to develop an algorithm for the automated classification of horse behaviour with high accuracy and resolution in a clinical setting. In the long term, this technology will not only improve the detection of acute and chronic pain in veterinary medicine, but also provide improved and new insights for behavioural research in horses.

Supplementary Materials: The following are available online at http://www.mdpi.com/2076-2615/10/12/2258/s1, Table S1: Detailed results of the counting first and last 1000 images per horse video. TN = true negative, TP = true positive, FN = false negative, FP = false positive.

Author Contributions: Study design, data collection and data handling, manuscript preparation, N.K. and U.A.; data collection and data handling, K.E. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding. NK was funded by "Gesellschaft für Pferdemedizin".

Acknowledgments: Open Access Funding by the University of Veterinary Medicine Vienna, Author wish to thank. Jessika-M. Cavalleri (co-investigator) for her support in gaining the financial sponsorship for N.K. and manuscript preparation.

Conflicts of Interest: The authors declare no conflict of interest

References

1. Price, J.; Catriona, S.; Welsh, E.M.; Waran, N.K. Preliminary evaluation of a behaviour–based system for assessment of postoperative pain in horses following arthroscopic surgery. *Vet. Anaesth. Analg.* **2003**, *30*, 124–137. [CrossRef] [PubMed]
2. Van Loon, J.P.M.; Back, W.; Hellebrekers, L.J.; Van Weeren, P.R. Application of a composite pain scale to objectively monitor horses with somatic and visceral pain under hospital conditions. *J. Equine Vet. Sci.* **2010**, *30*, 641–649. [CrossRef]
3. Van Loon, J.P.A.M.; Van Dierendonck, M.C. Objective pain assessment in horses (2014–2018). *Vet. J.* **2018**, *242*, 1–7. [CrossRef] [PubMed]
4. Dalla Costa, E.; Minero, M.; Lebelt, D.; Stucke, D.; Canali, E.; Leach, M.C. Development of the Horse Grimace Scale (HGS) as a Pain Assessment Tool in Horses Undergoing Routine Castration. *PLoS ONE* **2014**, *9*, e92281. [CrossRef]
5. Wathan, J.; Burrows, A.M.; Waller, B.M.; McComb, K. EquiFACS: The Equine Facial Action Coding System. *PLoS ONE* **2015**, *10*, e0137818.
6. Gleerup, K.; Forkman, B.; Lindegaard, C.; Andersen, P. An Equine Pain Face. *Vet. Anaesth. Analg.* **2015**, *42*, 103–114. [CrossRef]
7. Gleerup, K.B.; Lindegaard, C. Recognition and quantification of pain in horses: A tutorial review. *Equine Vet. Educ.* **2016**, *28*, 47–55. [CrossRef]
8. Dyson, S.; Berger, J.; Ellis, D.A.; Mullard, J. Development of an ethogram for a pain scoring system in ridden horses and its application to determine the presence of musculoskeletal pain. *J. Vet. Behav.* **2018**, *23*, 47–57. [CrossRef]
9. Lesimple, C.; Fureix, C.; De Margerit, E.; Seneque, E.; Menguy, H.; Hausberger, M. Towards a postural indicator of back pain in horses (Equus caballus). *PLoS ONE* **2012**, *7*, e44604. [CrossRef]
10. Sénèque, E.; Lesimple, C.; Morisset, S.; Hausberger, M. Could posture reflect welfare state? A study using geometric morphometrics in riding school horses. *PLoS ONE* **2019**, *14*, e0211852. [CrossRef]
11. Morrel-Samuels, P.; Krauss, R.M. Cartesian analysis: A computer-video interface for measuring motion without physical contact Behavior Research. *Methods Instrum. Comput.* **1990**, *22*, 466–470. [CrossRef]

12. Pereira, P.; Oliveira, R.F. A simple method using a single video camera to determine the three-dimensional position of a fish Behavior Research. *Methods Instrum. Comput.* **1994**, *26*, 443–446. [CrossRef]
13. Santucci, A.C. An affordable computer-aided method for conducting Morris water maze testing. *Behav. Res. Methodsinstrum. Comput.* **1995**, *27*, 60–64. [CrossRef]
14. Levitis, D.A.; Lidicker, W.Z., Jr.; Freund, G. Behavioural biologists don't agree on what constitutes behaviour. *Anim. Behav.* **2009**, *78*, 103–110. [CrossRef] [PubMed]
15. Anderson, D.J.; Perona, P. Toward a science of computational ethology. *Neuron* **2014**, *84*, 18–31. [CrossRef] [PubMed]
16. Cooke, S.J.; Hinch, S.G.; Wikelski, M.; Andrews, R.D.; Kuchel, L.J.; Wolcott, T.G.; Butler, P.J. Biotelemetry: A mechanistic approach to ecology. *Trends Ecol. Evol.* **2004**, *19*, 334–343. [CrossRef]
17. Mathis, A.; Mamidanna, P.; Cury, K.M.; Abe, T.; Murthy, V.N.; Mathis, M.W.; Bethge, M. DeepLabCut: Markerless pose estimation of user-defined body parts with deep learning. *Nat. Neurosci.* **2018**, *20*, 1282–1289. [CrossRef]
18. Graving, J.M.; Chae, D.; Naik, H.; Li, L.; Koger, B.; Costelloe, B.R.; Couzin, I.D. Deepposekit, a software toolkit for fast and robust animal pose estimation using deep learning. *Elife* **2019**, *8*, e47994. [CrossRef]
19. Mathis, M.W.; Mathis, A. Deep learning tools for the measurement of animal behaviour in neuroscience. *Curr. Opin. Neurobiol.* **2020**, *60*, 11. [CrossRef]
20. Burla, J.B.; Ostertag, A.; Westerath, H.S.; Hillmann, E. Gait determination and activity measurement in horses using an accelerometer. *Comput. Electron. Agric.* **2014**, *102*, 127–133. [CrossRef]
21. Eerdekens, A.; Deruyck, M.; Fontaine, J.; Martens, L.; Poorter, E.D.; Joseph, W. Automatic equine activity detection by convolutional neural networks using accelerometer data. *Comput. Electron. Agric.* **2020**, *168*, 105–139. [CrossRef]
22. Noldus, L.P.J.J.; Spink, A.J.; Tegelenbosch, R.A.J. EthoVision: A versatile video tracking system for automation of behavioral experiments. *Behav. Res. Methodsinstrum. Comput.* **2001**, *33*, 398–414. [CrossRef] [PubMed]
23. Lind, N.M.; Vinther, M.; Hemmingsen, R.P.; Hansen, A.K. Validation of a digital video tracking system for recording pig locomotor. *Behav. Neurosci. Methods* **2005**, *143*, 123–132. [CrossRef] [PubMed]
24. Walton, A.; Branham, A.; Gash, D.M.; Grondin, R. Automated video analysis of age-related motor deficits in monkeys using Etho Vision. *Neurobiol. Aging* **2006**, *27*, 1477–1483. [CrossRef] [PubMed]
25. Bloemen, H.; Aerts, J.M.; Berckmans, D.; Goedseels, V. Image analysis to measure activity index of animals. *Equine Vet. J.* **1997**, *29*, 16–19. [CrossRef]
26. Frost, A.R.; Tillett, R.D.; Welch, S.K. The development and evaluation of image analysis procedures for guiding a livestock monitoring sensor placement robot. *Comput. Electron. Agric.* **2000**, *28*, 229–242. [CrossRef]
27. Martin, B.R.; Prescott, W.R.; Zhu, M. Quantification of rodent catalepsy by a computer imaging technique. *Pharmacol. Biochem. Behav.* **1992**, *43*, 381–386. [CrossRef]
28. Aydin, A.; Cangar, O.; Ozcan, S.E.; Bahr, C.; Berckmans, D. Application of a fully automatic analysis tool to assess the activity of broiler chickens with different gait scores. *Comput. Electron. Agric.* **2010**, *73*, 194–199. [CrossRef]
29. Burke, J.; Nuthall, P.L.; McKinnon, A.E. An analysis of the feasibility of using image processing to estimate the live weight of sheep. *FHMG Res. Rep.* **2004**, *2*, 6.
30. Kashiha, M.A.; Bahr, C.; Ott, S.; Moons, C.P.; Niewold, T.A.; Tuyttens, F.; Berckmans, D. Automatic monitoring of pig locomotion using image analysis. *Livest. Sci.* **2014**, *159*, 141–148. [CrossRef]
31. Oczak, M.; Maschat, K.; Berckmans, D.; Vranken, E.; Baumgartner, J. Automatic estimation of number of piglets ina pen during farrowing, using image analysis. *Biosyst. Eng.* **2016**, *151*, 81–89. [CrossRef]
32. Van Hertem, T.; Rooijakkers, L.; Berckmans, D.; Peña Fernández, A.; Norton, T.; Berckmans, D.; Vranken, E. Appropriate data visualisation is key to Precision Livestock Farming acceptance. *Comput. Electron. Agric.* **2016**, *138*, 1–10. [CrossRef]
33. Venter, P.J.; Hanekom, J.J. Automatic detection of African elephant (Loxodonta africana) infrasonic vocalisations from recordings 2010). *Biosyst. Eng.* **2010**, *106*, 286–294. [CrossRef]
34. Pereira, T.D.; Aldarondo, D.E.; Willmore, L.; Kislin, M.; Wang, S.S.-H.; Murthy, M.; Shaevitz, J.W. Fast animal pose estimation using deep neural networks. *Nat. Methods* **2019**, *16*, 117–125. [CrossRef]
35. Bonneau, M.; Vayssade, J.A.; Troupe, W.; Arquet, R. Outdoor animal tracking combining neural network and time-lapse cameras. *Comput. Electron. Agric.* **2020**, *168*, 6. [CrossRef]

36. Nath, T.; Mathis, A.; Chen, A.C.; Patel, A.; Bethge, M.; Mathis, M.W. Using DeepLabCut for 3D markerless pose estimation across species and behaviors. *Nat. Protocol.* **2019**, 1–27. [CrossRef] [PubMed]
37. Zhou, C.; Zhang, B.; Lin, K.; Xu, D.; Chen, C.; Yang, X.; Sun, C. Near-infrared imaging to quantify the feeding behavior of fish in aquaculture. *Comput. Electron. Agric.* **2017**, *135*, 233–241. [CrossRef]
38. Nasirahmadi, A.; Edwards, S.A.; Sturm, B. Implementation of machine vision for detecting behaviour of cattle and pigs. *Livest. Sci.* **2017**, *202*, 25–38. [CrossRef]
39. Anamika, A. Study of Techniques used for Medical image Segmentation and Computation of Statistical Test for Region Classification of Brain MRI. *Int. J. Inf. Technol. Comput. Sci.* **2013**, *5*, 44–53. [CrossRef]
40. Vayssade, J.-A.; Arquet, R.; Bonneau, M. Automatic activity tracking of goats using drone camera. *Comput. Electron. Agric.* **2019**, *162*, 767–772. [CrossRef]
41. Bussieres, G.; Jacques, C.; Lainay, O.; Beauchamp, G.; Leblond, A.; Cadore, J.L.; Desmaizieres, L.M.; Cuveilliez, S.G.; Troncy, E. Development of a composite orthopaedic pain scale in horses. *Res. Vet. Sci.* **2008**, *85*, 294–306. [CrossRef] [PubMed]

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

6.Conclusion

This paper proposes, a methodology for automated video surveillance of horses to evaluate behavioral changes in a therapeutic scenario. Annotating a limited number of images was adequate for a deep learning training of a predictor, which was able to label three distinct body markers on time-lapse films. The findings demonstrated a high accuracy and sensitivity of more than 80 percent in posture estimation. An initial study of the data acquired allows for the creation of an algorithm to follow the behavioral patterns of horses automatically.

An action camera with a time-lapse mode was used to shoot videos of horses at an animal hospital for this study. For automated body component prediction, these videos were analyzed using the Loopy convolutional neural network (http://loopb.io, Loopbio GmbH, Vienna, Austria. Loopy is a web-based program that was founded in 2015 with the goal of providing natural sciences includes image processing, tracking, deep learning, video recording and behavioral analysis features. A number of general-purpose algorithms for tracking the position of subjects in videos are included in the package. It provides a bespoke deep learning tracking method for tracking single and multiple animal poses. Traditional image processing necessitates rigorous parameter tuning by hand and significant visual variances between the animal and the background, resulting in unsatisfactory results. Meanwhile, when given a small portion of the annotated videos, deep learning algorithms learn how to locate the animals. Deep learning necessitates the use of training data that depicts the position of objects or their properties. The study of version 1, which was developed from 1990 images, demonstrated a continuous and accurate monitoring of the three markers, at least in images identical to the annotated image, as well as a high level of resolution during the horse's movement.

We chose to annotate three markers on the horse. To get a better view of the horse's behavior, we chose a camera angle that gives us a view from 60 degrees instead of looking down on it from above. The markers for nose, withers, and tail were often visible from this perspective. Only the tail marker was often not visible, because the horse preferred to be in a position where it wasn't predictable. The first version almost perfectly predicts all visible markers in images that are similar to the annotated image. The color of the coat and the inflow of light did not negatively affect the prediction quality compared to the horse's position, and the accuracy in version 1. The markers were also seen when the person was moving quickly. The more annotated images we had, the better the horse's position looked in the images. With the second version, we used math to help us detect the markers that were incorrectly labeled. The method used together with an optical control of the affected video section and a corresponding graphical representation made it possible to identify the wrong predicted marker.

The findings show that automated body part tracking using time-lapse films is achievable. The data obtained by the deep learning module may be utilized to create an algorithm for the automated categorization of horse behavior in a clinical environment with high accuracy and resolution. In the long run, this technology will not only enhance the identification of acute and chronic pain in veterinary care, but it will also give improved and novel insights for equine behavioral study.

7. Summary

In this study, we aimed to develop an automated video tracking model in order to observe behavioral changes of horses in clinical context. This project has been operated on Loopy by using deep learning and image processing method. The full scope of this project has been completed by using 34 horses in four phases. The process of image analysis started with the labelling of markers on video files of 11.52 min corresponding to 21,342 images.

In first phase, the system has been calibrated by annotating 2500 frames from 8 videos of 8 different horses. In the second phase, we ran a prediction on the remaining of unlabeled videos. The third phase which was robustness of version (V2) was tested with the remaining unlabeled videos. We ran mathematical calculation to see accuracy. In phase 4, validation has been done via by running automatic annotation on 26 videos of different horses recorded during the daytime, which had not been annotated formerly.

In conclusion, behavioral change in medical context has been successfully tracked by video processing using deep learning. This research can lead to predicting the symptoms in advance in future with right configuration.

8. Zusammenfassung

In dieser Studie haben wir uns zum Ziel gesetzt, ein automatisiertes Video-Tracking-Modell zu entwickeln, um Verhaltensänderungen von Pferden im klinischen Kontext zu beobachten. Dieses Projekt wurde auf Loopy unter Verwendung von Deep Learning und Bildverarbeitungsmethoden durchgeführt. Der gesamte Umfang dieses Projekts wurde anhand von 34 Pferden in vier Phasen abgeschlossen. Der Prozess der Bildanalyse begann mit der Beschriftung von Markern auf Videodateien von 11,52 Minuten Dauer, die 21.342 Bildern entsprechen.

In der ersten Phase wurde das System durch die Beschriftung von 2500 Bildern aus 8 Videos von 8 verschiedenen Pferden kalibriert. In der zweiten Phase haben wir eine Vorhersage für die verbleibenden nicht beschrifteten Videos durchgeführt. Die dritte Phase, die Robustheit der Version (V2), wurde mit den verbleibenden unmarkierten Videos getestet. Wir führten mathematische Berechnungen durch, um die Genauigkeit zu ermitteln. In Phase 4 erfolgte die Validierung durch die automatische Beschriftung von 26 tagsüber aufgenommenen Videos verschiedener Pferde, die zuvor nicht beschriftet worden waren.

Zusammenfassend lässt sich sagen, dass Verhaltensänderungen im medizinischen Kontext erfolgreich durch Videoverarbeitung mit Deep Learning verfolgt werden konnten. Diese Forschung kann dazu führen, dass die Symptome in Zukunft mit der richtigen Konfiguration im Voraus vorhergesagt werden können.

9.List of abbreviations

CNN convolutional neural network

DNN deep neural network

GPS global positioning systems

GRU gated recurrent unit

GUIs graphical user interfaces

IMUs inertial measurement units

LSTM long short-term memory

ML machine learning

NN neural network

RNN recurrent neural network

10.References

Aguirre V, Orihuela A. 2010. Effect of work experience and position in the plow harness on some physiological parameters of horses under field conditions. J. Appl. Anim. Welf. Sci. 3, 231-238.

Allison JM, White JM. 1991. Hardware for microcomputer control of the environment of a production broiler house. Appl. Engineering Agric. 7,119- 123.

Anamika A. 2013. Study of Techniques used for Medical image Segmentation and Computation of Statistical Test for Region Classification of Brain MRI. Int. J. Inf. Technol. Comput. Sci., 5, 44–53.

Anderson DJ, Perona P. 2014. Toward a science of computational ethology. Neuron, 84, 18–31.

Ashley FH, Waterman-Pearson AE, Whay HR. 2005. Behavioural assessment of pain in horses and donkeys: application to clinical practice and future studies. Equine Vet J, 37:565–575.

Ben-Shaul Y. 2017. OptiMouse: a comprehensive open source program for reliable detection and analysis of mouse body and nose positions. BMC Biol. 15, 41.

Bengio Y, LeCun Y, Hinton G. 2015. Deep Learning. Nature. 521, 436–444.

Berckmans D. 1986. Analyse van de kliinaatbeheersing in dierlijke pmductie eenheden ter optimalisering van de regeling. PhD Thesis. K.U. Leuven. p 374.

Bidder OR, Campbell HA, Gómez-Laich A, Urgé P, Walker J, Cai Y, Gao L, Quintana F, Wilson RP. 2014. Love thy neighbour: automatic animal behavioural classification of acceleration data using the k-nearest neighbour algorithm. PloS One 9 (2), e88609.

Bloemen H, Aerts JM, Berckmans D, Goedseels V. 1997. Image analysis to measure activity index of animals. Equine Vet. J., 29, 16–19.

Bokkers E, Koene P. 2000. Motivation and ability to walk in broilers and layer chicks on two diets. In Proceedings of the 34th InternationalCongress of the International Society of Applied Ethology (Florianopolis, 17–20 October 2000), p. 115. Florianopolis: UFSC

Brendle J, Hoy S. 2011. Investigation of distances covered by fattening pigs measured with VideoMotionTracker. Appl. Anim. Behav. Sci. 132, 27-32.

Broomé S, Ask K, Rashid-Engström M, Haubro Andersen P, Kjellström H. 2022. Sharing pain: Using pain domain transfer for video recognition of low-grade orthopedic pain in horses. PLOS ONE 17(3): e0263854. https://doi.org/10.1371/journal.pone.0263854

Buresová O, Bolhuis JJ, Bure Ï, Ïs J. 1986. Differential effects of cholinergic blockade on performance of rats in the water tank navigation task and in a radial water maze. Behavioral Neuroscience, 100, 476-482.

Burla JB, Ostertag A, Westerath HS, Hillmann E. 2014. Gait determination and activity measurement in horses using an accelerometer. Comput. Electron. Agric., 102, 127–133.

Bussieres G, Jacques C, Lainay O, Beauchamp G, Leblond A, Cadore JL, Desmaizieres LM, Cuveilliez SG, Troncy E. 2008. Development of a composite orthopaedic pain scale in horses. Res. Vet. Sci., 85, 294–306.

Cambridge AJ, Tobias KM, Newberry RC, Sarkar DK. 2000. Subjective and objective measurements of postoperative pain in cats. J Am Vet Med Assoc., 1;217(5):685-90. doi: 10.2460/javma.2000.217.685. PMID: 10976300.

Christensen JW, Ladewig J, Søndergaard E, Malmkvist J. 2002. Effects of individual versus group stabling on social behaviour in domestic stallions. Appl. Anim. Behav. Sci. 75, 233–248.

Conzemius MG, Hill CM, Sammarco JL, Perkowski SZ. 1997. Correlation between subjective and objective measures used to determine severity of postoperative pain in dogs. J Am Vet Med Assoc., 1;210(11):1619-22. PMID: 9170089.

Corcoran AJ, Schirmacher MR, Black E, Hedrick TL. 2021. ThruTracker: Open-Source Software for 2-D and 3-D Animal Video Tracking. bioRxiv 2021.05.12.443854; doi: https://doi.org/10.1101/2021.05.12.443854.

Cullen JM, Shaw E, Baldwin HA. 1965. Methods for measuring the three-dimensional structures offish schools. Animal Behaviour, 13, 534-54

Dalla Costa E, Minero M, Lebelt D, Stucke D, Canali E, Leach MC.2014. Development of the Horse Grimace Scale (HGS) as a Pain Assessment Tool in Horses Undergoing Routine Castration. PLoS ONE, 9, e92281.

Darbandi H, Serra Bragança F, van der Zwaag J, Voskamp J, Gmel AI, Haraldsdóttir EH, Havinga P. 2021. Using Different Combinations of Body-Mounted IMU Sensors to Estimate Speed of Horses—A Machine Learning Approach. Sensors, 21, 798. https://doi.org/10.3390/s21030798

Darwin C. 1872. The Expression of the Emotions in Man and Animals. John Murray, London, UK.

Dell AI, Bender JA, Branson K, Couzin ID, de Polavieja GG, Noldus LP, Pérez-Escudero A, Perona P, Straw AD, Wikelski M, Brose U. 2014. Automated image-based tracking and its application in ecology. Trends Ecol Evol., 29(7):417-28. doi: 10.1016/j.tree.2014.05.004. Epub 2014 Jun 5. PMID: 24908439.

Dollár P, Welinder P, Perona P. 2010. Cascaded pose regression. in IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2010 1078–1085 (IEEE, Piscataway, NJ, USA, 2010).

Eerdekens A, Deruyck M, Fontaine J, Martens L, Poorter ED, Joseph W. 2020. Automatic equine activity detection by convolutional neural networks using accelerometer data. Comput. Electron. Agric., 168, 105–139.

Farries G, Gough KF, Parnell AC, McGivney BA, McGivney CL, McGettigan PA, MacHugh DE, Katz LM, Hill EW. 2019. Analysis of genetic variation contributing to measured speed in Thoroughbreds identifies genomic regions involved in the transcriptional response to exercise. Anim. Genet. 50, 670–685.

Frost AR, Schofield CP, Beaulah SA, Mottram TT, Lines JA, Wathes CM. 1997. A review of livestock monitoring and the need for integrated systems. Comput. Electron. Agricult. 17 (2), 139–159.

Frost AR, Tillett RD, Welch SK. 2000. The development and evaluation of image analysis procedures for guiding a livestock monitoring sensor placement robot. Computers and Electronics in Agriculture 28, 229–242

Fuchs E, Kramer M, Hermes B, Netter P, Hiemke C. 1996. Psychological stress in tree shrews: Clomipramine counteracts behavioral and endocrine changes. Pharmacology, Biochemistry & Behavior, 54, 219-228.

Fukushima K. 1988. Neocognitron: A hierarchical neural network capable of visual pattern recognition. Neural networks 1.2, 119- 130.

Gleerup K, Forkman B, Lindegaard C, Andersen P. 2015. An Equine Pain Face. Vet. Anaesth. Analg., 42, 103–114.

Gleerup KB, Lindegaard C. 2016. Recognition and quantification of pain in horses: A tutorial review. Equine Vet.Educ., 28, 47–55.

Gomez-Marin A, Partoune N, Stephens GJ, Louis M, Brembs B. 2012. Automated tracking of animal posture and movement during exploration and sensory orientation behaviors. PLoS One 7, e41642.

Graving JM, Chae D, Naik H, Li L, Koger B, Costelloe BR, Couzin ID. 2019. Deepposekit, a software toolkit for fast and robust animal pose estimation using deep learning. Elife, 8, e47994.

Hamm D, Turchi P, Jochle W. 1995. Sedative and analgesic effects of detomidine and romidine in horses. Vet Record 132,324-326.

He K, Zhang X, Ren S, Sun J. 2016. Deep residual learning for image recognition. in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 770–778 (IEEE, Piscataway, NJ, USA, 2016).

Heleski CR, Shelle AC, Nielsen BD, Zanella AJ. 2002. Influence of housing on weanling horse behavior and subsequent welfare. Appl. Anim. Behav. Sci. 78, 291–302.

Hoffmann G, Bockisch FJ, Kreimeier P. 2009. Influence of the husbandry system on the movement activity and stress exposure of horses in discharge husbandry systems. Landbauforsch. Volk. 59, 105–111.

Holton LL, Scott EM, Nolan AM, Reid J, Welsh E, Flaherty D. 1998. Comparison of three methods used for assessment of pain in dogs. J Am Vet Med Assoc., 1;212(1):61-6. PMID: 9426779.

Insafutdinov E, Pishchulin L, Andres B, Andriluka M, Schiele B. 2016. DeeperCut: a deeper, stronger, and faster multi-person pose estimation model. in European Conference on Computer Vision 34–50 (Springer, New York, 2016).

Jochle W, Moore JN, Brown J. 1991. Comparison of detomidine, butarphanol, flunixin meglumine and xylazine in clinical cases of equine colic. Equine Vet J 7 (Suppl.), 111-116.
Jochle W. 1989. Field trial evaluation of detomidine as a sedative and analgesic agent in horses with colic. Equine Vet J 7 (Suppl.),117-120.

Johnson CB, Taylor PM, Young SS, Brearley JC. 1993. Postoperative analgesia using phenylbutazone, flunixin or carprofen in horses. Vet Rec., 2;133(14):336-8. doi: 10.1136/vr.133.14.336. PMID: 8236675.

Kain J, Stokes C, Gaudry Q, Song X, Foley J, Wilson R, De Bivort B. 2013. Leg-tracking and automated behavioural classification in Drosophila. Nature communications, 4:1910.

Kalpravidh M, Lumb WV, Wright M, Heath RB. 1984. Effects of butorphanol, flunixin, levorphanol, morphine, and xylazine in ponies. Am J Vet Res., 45(2):217-23. PMID: 6711945.

Kashiha MA, Bahr C, Ott S, Moons CP, Niewold TA, Tuyttens F, Berckmans D. 2014. Automatic monitoringof pig locomotion using image analysis. Livest. Sci., 159, 141–148.

Kashiha MA, Bahr C, Ott S, Moons CPH, Niewold TA, Ödberg FO, Berckmans D. 2013b. Automatic identification of marked pigs in a pen using image pattern recognition. Comput. Electron. Agric. 93, 111-120.

Keegan KG, Kramer J, Yonezawa Y, Maki H, Pai PF, Dent EV, Kellerman TE, Wilson DA, Reed SK. 2011. Assessment of repeatability of a wireless, inertial sensor–based lameness evaluation system for horses. American journal of veterinary research, 72(9), pp.1156-1163.

Kurvers C, Van Weeren PR, Rogers CW, Van Dierendonck MC. 2006. Quantification of spontaneous locomotion activity in foals kept in pastures under various management conditions. Am. J. Vet. Res. 67, 1212–1217.

Langrock R, King R, Matthiopoulos J, Thomas L, Fortin D, Morales JM. 2012. Flexible and practical modeling of animal telemetry data: hidden markov models and extensions. Ecology 93 (11), 2336–2342.

Lecun Y, Bottou L, Bengio Y, Haffner P. 1998. Gradient-based learning applied to document recognition. Proceedings of the IEEE 86.1, 2278-2324.

Lesimple C, Fureix C, De Margerit E, Seneque E, Menguy H, Hausberger M. 2012. Towards a postural indicator of back pain in horses (Equus caballus). PLoS ONE, 7, e44604.

Lind NM, Vinther M, Hemmingsen RP, Hansen AK. 2005. Validation of a digital video tracking system forrecording pig locomotor. Behav. Neurosci. Methods, 143, 123–132.

Martin BR, Prescott WR, Zhu M. 1992. Quantification of rodent catalepsy by a computer-imaging technique. Pharmacology, Biochemistry & Behavior, 43, 381-386.

Mathis A, Mamidanna P, Cury KM, Abe T, Murthy VN, Mathis MW, Bethge M. 2018. DeepLabCut: Markerless pose estimation of user-defined body parts with deep learning. Nat. Neurosci., 20, 1282–1289.

Mathis MW, Mathis A. 2020. Deep learning tools for the measurement of animal behaviour in neuroscience. Curr. Opin. Neurobiol., 60, 11.

Matsumoto J, Urakawa S, Takamura Y, Malcher-Lopes R, Hori E, Tomaz C, Ono T, Nishijo H. 2013. A 3D-video-based computerized analysis of social and sexual interactions in rats. PLoS One 8, e78460.

McDonnell S. 2005. Is it psychological, physical, or both? Proc. Am. Ass. Equine Practnrs. 51, 231-238.

McLendon BD, Allison JM, Steele CB. 1990. Microcomputer-based monitoring system for broiler housing. Appl. Engineering in Agric. 6,345-348.

Meira CT, Fortes MR, Farah MM, Porto-Neto LR, Kelly M, Moore SS, Pereira GL, Chardulo LAL, Curi RA. 2014. Speed Index in the Racing Quarter Horse: A Genome-wide Association Study. J. Equine Vet. Sci. 34, 1263–1268.

Mendes CS, Bartos I, Akay T, Márka S, Mann RS. 2013. Quantification of gait parameters in freely walking wild type and sensory deprived Drosophila melanogaster. elife. 2:e00231.

Merskey H, Bogduk N. 1994. Classification of chronic pain, IASP Task Force on Taxonomy. Seattle, WA: International Association for the Study of Pain Press (Also available online at www. iasp-painorg).

Molony V, Kent J. 1997. Assessment of acute pain in farm animals using behavioral and physiological measurements. Journal of animal science: 266– 272. Available: http://www.journalofanimalscience.org/content/75/1/266.short. Accessed 30 January 2013.

Molony V, Kent JE, McKendrick IJ. 2002. Validatiom of a method for assessment of an acute pain in lambs. Appl Anim Behav Sci, 76,215-238.

Morrel-Samuels P, Krauss RM. 1990. Cartesian analysis: A computer-video interface for measuring motion without physical contact. Behavior Research Methods. Instruments. & Computers, 22 (5). 466-470

Morris RGM. 1981. Spatial location does not require the presence of local cues. Learning & Mot., 12, 239-260.

Mukhina TV, Bachurin SO, Lermontova NN, Zefirov NS. 2001. Versatile computerized system for tracking and analysis of water maze tests. Behavior Research Methods, Instruments, & Computers, 33, 371-380.

Nath T, Mathis A, Chen AC, Patel A, Bethge M, Mathis MW. 2019. Using DeepLabCut for 3D markerlesspose estimation across species and behaviors. Nat. Protocol., 1–27.

Noldus LPJJ, Spink AJ, Tegelenbosch RAJ. 2001. EthoVision: A versatile video tracking system for automation of behavioral experiments. Behav. Res. Methodsinstrum. Comput., 33, 398–414.

Noldus LPJJ, Trienes RJH, Hendriksen AHM, Jansen H, Jansen RG. 2000. The Observer Video-Pro: New software for the collection, management, and presentation of time-structured data from videotapes and digital media files. Behavior Research Methods, Instruments, & Computers, 32, 197-206.

Noldus LPJJ. 1991. The Observer: A software system for collection and analysis of observational data. BehaviorResearch Methods, Instruments, & Computers, 23, 415-429

Oczak M, Maschat K, Berckmans D, Vranken E, Baumgartner J. 2016. Automatic estimation of number of piglets ina pen during farrowing, using image analysis. Biosyst. Eng., 151, 81–89.

Pan WHT, Lee CR, Lim LH. 1996. A new video path analyzer to monitor travel distance, rearing, and stereotypic movement of rats. Journal of Neuroscience Methods, 70, 39-43.

Pastorelli G, Musella M, Zaninelli M, Tangorra F, Corino C. 2006. Staic spatial requirements of growing-finishing and heavy pigs. Livest. Sci. 105, 260-264.

Pereira TD, Aldarondo DE, Willmore L, Kislin M, Wang SSH, Murthy M, Shaevitz JW. 2019. Fast animal pose estimation using deep neural networks. Nat. Methods, 16, 117–125.

Phinyomark A, Petri G, Ibáñez-Marcelo E, Osis ST, Ferber R. 2018. Analysis of big data in gait biomechanics: Current trends and future directions. J. Med. Biol. Eng. 38, 244–260.

Pitcher T J. 1975. A periscopic method for determining the threedimensional positions of fish in schools. Journal ofFisheries Research Board ofCanada, 32,1533-1538

Pitcher TJ. 1973. The three-dimensional structure ofschools in the minnow, Phoxinus phoxinus (L.). Animal Behaviour, 21, 673- 686.

Price J, Catriona S, Welsh EM, Waran NK. 2003. Preliminary evaluation of a behaviour–based system for assessment of postoperative pain in horses following arthroscopic surgery. Vet.

Anaesth. Analg, 30, 124–137.

Raekallio M, Taylor PM, Bennet RC. 1997. Preliminary investigations of pain and analgesia assessment in horses given phenylbutazone or placebo after arthroscopic surgery. Vet Surg 26,155.

Raekallio M, Taylor PM, Blomfield M. 1997b. A comparison of methods for evaluation of pain and distress after orthopaedic surgery in horses. J Vet Anaesthesia 24,17-20.

Ravi D, Wong C, Lo B, Yang GZ. 2017. A deep learning approach to on-node sensor data analytics for mobile or wearable devices. IEEE J. Biomed. Health Informatics 21 (1), 56–64.

Robertson JT, Muir WW. 1983. Anew analgesic drug combination in the horse. Am J Vet Res, 44,1667-1669.

Ruis MAW, te Brake JHA, Buwalda B, de Boer SF, Meerlo P, Korte SM, Blokhuis HJ, Koolhaas JH. 1999. Housing familiar male wildtype rats together reduces the long-term adverse behavioural and physiological effects of social defeat. Psychoneuroendocrinology, 24, 285-300.

Santucci AC. 1995. An affordable computer-aided method for conducting Morris water maze testing. Behav. Res. Methods, instrum. Comput., 27, 60–64.

Sénèque E, Lesimple C, Morisset S, Hausberger M. 2019. Could posture reflect welfare state? A study using geometric morphometrics in riding school horses. PLoS ONE, 14, e0211852.

Schmidhuber J. 2015. Deep Learning in Neural Networks: An Overview. Neural Networks. 61, 85–117.

Spooner RIW, Thomson A, Hall J, Morris RGM, Salter SH. 1994. The Atlantis Platform: A new design and further developments of Buresová's on-demand platform for the water Ï maze. Learning & Memory, 1, 203-211.

Spruijt BM, Buma MOS, van Lochem PBA, Rousseau JBI. 1998. Automatic behavior recognition: What do we want to recognize and how do we measure it? In Proceedings of MeasuringBehavior '98, 2nd InternationalConference on Methods and Techniques in Behavioral Research (Groningen, 18–21 August 1998), pp. 264- 266. Wageningen, the Netherlands: Noldus Information Technology.

Spruijt BM, Gispen WH. 1983. Prolonged animal observations by use of digitized video displays. Pharmacology, Biochemistry & Behavior, 19, 765-769.

Spruijt BM, Pitsikas N, Algeri S, Gispen WH. 1990. Org2766 improves performance of rats with unilateral lesions in the fimbria fornix in a spatial learning task. Brain Research, 527, 192-197.

Sustr P, Spinka M, Newberry RC. 2000. Automatic computer analysis of pig play. In Proceedings of Measuring Behavior 2000, 3rd International Conference on Methods and

Techniques in Behavioral Research (Nijmegen, 15–18 August 2000), pp. 307-308. Wageningen, the Netherlands: Noldus Information Technology.

Symons PEK. 1971a. Estimating distances between fish schooling in an aquarium. Journal ofFisheries Research Board ofCanada, 28, 1805-1806.

Symons PEK. 1971b. Spacing and density in schooling threespine sticklebacks (Gasterosteus aculeatus) and mummichog (Fundulus heteroclistus). Journal ofFisheries Research Board ofCanada, 28, 999-1004.

Torcivia C, McDonnell S. 2020. In-Person Caretaker Visits Disrupt Ongoing Discomfort Behavior in Hospitalized Equine Orthopedic Surgical Patients. Animals, 10, 210. https://doi.org/10.3390/ani10020210.

Uhlmann V, Ramdya P, Delgado-Gonzalo R, Benton R, Unser M. 2017. FlyLimbTracker: An active contour based approach for leg segment tracking in unmarked, freely behaving Drosophila. PLoS One, 12(4):e0173433.

Van Kampen M, Schmitt U, Hiemke C, Fuchs E. 2000. Diazepam has no beneficial effects on stress-induced behavioral and endocrine changes in male tree shrews. Pharmacology, Biochemistry & Behavior, 65, 539-546.

Van Loon JPM, Back W, Hellebrekers LJ, Van Weeren PR. 2010. Application of a composite pain scale to objectively monitor horses with somatic and visceral pain under hospital conditions. J. Equine Vet. Sci.,30, 641–649.

Van Loon JPAM, Van dierendonck MC. 2015. Monitoring acute equine visceral pain with the Equine Utrecht University Scale for Composite Pain Assessment (EQUUS-COMPASS) and the Equine Utrecht University Scale for Facial Assessment of Pain (EQUUS-FAP): A validation study. Vet J. 216: 175–177. pmid:27687948.

Venter PJ, Hanekom JJ. 2010. Automatic detection of African elephant (Loxodonta africana) infrasonic vocalisations from recordings 2010). Biosyst. Eng., 106, 286–294.

Visser EK, Ellis AD, Van Reenen CG. 2008. The effect of two different housing conditions on the welfare of young horses stabled for the first time. Appl. Anim. Behav. Sci. 114, 521–533.

Walton A, Branham A, Gash DM, Grondin R. 2006. Automated video analysis of age-related motor deficits in monkeys using EthoVision. Neurobiology of Aging 27, 1477–1483

Wang Y, Li J, Zhang Y, Sinnott RO. 2021. Identifying Lameness in Horses through Deep Learning. In The 36th ACM/SIGAPP Symposium on Applied Computing (SAC '21), March 22–26, 2021, Virtual Event, Republic of Korea. ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3412841.3441973.

Warren-Smith A, McGreevy P 2010. The use of pedometers to estimate motor laterality in grazing horses. J. Vet. Behav. 5, 177–179.

Wathan J, Burrows AM, Waller BM, McComb K. 2015. EquiFACS: The Equine Facial Action Coding System. PLoS ONE, 10, e0137818.

Werhahn H, Hessel EF, Van den Weghe HFA. 2012. Competition horses housed in single stalls (II): Effects of free exercise on the behavior in the stable, the behavior during training, and the degree of stress. J. Equine Vet. Sci. 32, 22–31.

Williams J, Kenworth K, Jones T. 2019. The role of heart rate monitoring to assess workload during maintenance interval training in national hunt racehorses. J. Vet. Behav. 29, 150.

Young PC. 2011. Recursive estimation and time-series analysis: An introduction for the student and practitioner. Springer Science & Business Media, Verlag Berlin Heidelberg.