CAPÍTULO 3

Computer vision and image analysis in livestock production

João R. R. Dorea, Guilherme J. M. Rosa, Arthur F. A. Fernandes, Tiago Bresolin

1. Computer Vision System

Computer Vision can be defined as the field of study that seeks to describe the world through images by interpreting, reconstructing, and extracting properties from images, such as shapes, textures, densities, and distances (SZELISK, 2011). Computer Vision Systems are also known as machine vision systems, visual image systems, or just image systems. Therefore, Computer Vision is essentially the implementation of artificial systems to handle visual problems of interest by using image processing and analytical techniques. Along with image analysis and processing, Machine Learning and Pattern Recognition are also highly interconnected with Computer Vision. Pattern Recognition is a field that studies not only images but also other signals, such as sound and text. As the name suggests, it is an area dedicated to the study of patterns that may appear in a given signal. In the context of imaging, pattern recognition is generally studied within image analysis as the development of mathematical methods for the identification of simple geometrical structures such as lines and circles (HOUGH, 1962; ATHERTON *et al*., 1999) or key-point features that can be jointly used to identify more complex objects or patterns (BAY *et al*. 2006; LEUTENEGGER *et al*. 2011). Machine Learning is a broader field that is concerned with the development and application of algorithms for extracting information from the most diverse data sets (MURPHY, 2012). Several machine learning algorithms have been developed or adapted specifically for solving computer vision problems.

The popularization of digital cameras in smartphones and other mobile devices is directly connected to the increasing volume of data (photos and videos) generated over the last few years in many fields. Such large and heterogeneous image data have also driven the development of cutting-edge computer vision systems to solve the most diverse problems through sophisticated analytical techniques. This increased interest in computer vision and related areas can be illustrated by the increasing number of publications in the last decade, reported by Fernandes *et al*. 2020 (Figure 1).

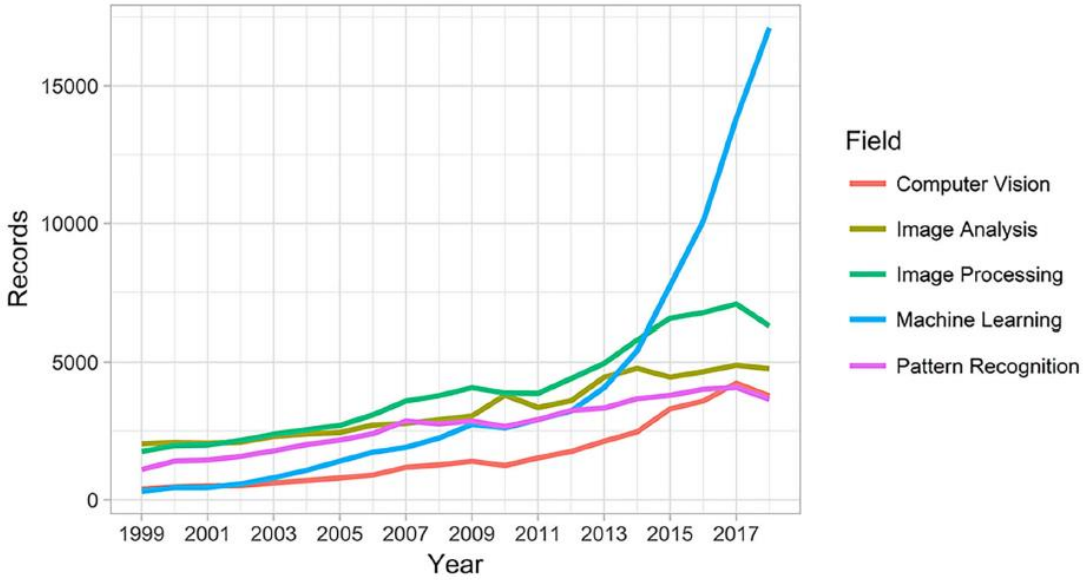

Figure 1. Count of publications hits in “Web of Science” for computer vision, image analyses, image processing, machine learning, and pattern recognition. Adapted from Fernandes *et al*., 2020.

Computer vision and image analysis have been undergoing a great revolution in the last few years mainly due to the advent of specific machine learning applications as deep learning. These methods show better performance than classic techniques for many challenging computer vision problems such as natural image segmentation, optical character recognition, and object classification (LECUN *et al*. 2015). Nowadays, computer vision systems using deep learning algorithms show comparable or better accuracy than humans in several visual recognition tasks, including recognizing traffic signs (CIRESAN *et al*., 2012), faces (TAIGMAN *et al*., 2014), and handwritten characters (CIRESAN *et al*., 2012; WANG *et al*., 2013).

2. Computer Vision System for High-Throughput Phenotyping in Livestock

Among the various digital technologies in livestock, computer vision systems are emerging as a powerful solution for high-throughput phenotyping, which is crucial to create optimized farm management decisions and for genetic improvement. The amount of information carried in a single image usually goes beyond the developer's primary interest when a computer vision system is created. For example, suppose that the images presented in Figure 2 were analyzed using a computer vision system built to identify individual animals and to predict their behavior. The predicted phenotypes would be four individual IDs with their respective behavior activity: in this case, standing. However, there is additional information available in the image that is not being used, such as the housing system, the presence of trees, the green leaves in the trees (indicating season), the sky condition (rainy, cloudy, or sunny), the animal stock density (area of pen/animal), animal social network, etc. Even if the computer vision system primary goal was related to animal identification and behavior, the amount of information carried by the image can allow future developments, as new ideas are created and more sophisticated data analytics tools become available. Very few sensing technologies can generate such rich data source from a single device.

Figure 2. Group of four dairy calves housed in a super-hutch. Figure 1a (left side): winter period, trees without leaves, and clear sky. Figure 1b (right side): summer period, trees with green leaves, and cloudy sky. Adapted from Oliveira *et al*. (2021).

Image analysis has been applied in different fields of dairy science and production such as lameness classification (ZHAO *et al*., 2015), body condition score (HALACHMI *et al*., 2013), and behavior monitoring (TSAI; HUANG, 2014; GUZHVA *et al*., 2018). However, most of the studies published in the literature used manual feature extraction, such as animal biometric measurements (e.g., distance from the head to the tail, hip height, spine curvature, etc.), and then fit those features using linear models (GOMES *et al*., 2016) or other types of predictive models (PORTO *et al*., 2015; SALAU *et al*., 2017). In addition, there are studies showing the applicability of thermal cameras for evaluation of lameness and mastitis (HOVINEN *et al*., 2008; TEDÍN *et al*. 2012; BYRNE *et al*., 2017).

Recently, RGB-D sensors have gained attention due to the incorporation of the depth information to the digital color image. The image data from these sensors have been shown to be more robust for animal identification or for body parts recognition (FERNANDES *et al*., 2020). The main applications have been on identification of animal locomotion and back posture (VIAZZI *et al*., 2014; HERTEM *et al*., 2017), body condition score (BERCOVICH *et al*., 2013; SPOLIANSKY *et al*., 2016), leg angle and udder depth (SALAU *et al*., 2017), and BW (SONG *et al*., 2018) of dairy cows. In Song *et al*. (2018), three morphological characteristics (hip height, hip width, and rump length) were measured using image analysis on an automated way. Those variables were included as additional predictors in a linear regression model along with days in milk (DIM), age, and parity (called full model). The authors reported a mean absolute percentage error of 5.5% for the full model. In addition, the authors also presented a linear regression that included DIM and age only, and the mean absolute error was 6.1%. The small improvement of adding image-based morphological characteristics to predict BW might be related to the small number of features extracted from the 3D images (only three features), and to the type of predictive model used. Therefore, type and amount of feature extracted from 3D images (FERNANDES *et al*., 2019) and modeling approach (DÓREA *et al*., 2018; FERNANDES *et al*., 2020) are important factors that may improve prediction quality in image analyses. Fernandes *et al*. (2020) and Cominotte *et al*. (2020) evaluated the impact of different combinations of image features on the predictive model quality in pigs and cattle. The authors used features related to biometric body measurements (dorsal area and length, projected body volume, height, and width in 11 points across the dorsal area) and body shape discriminator (Fourier descriptors and eccentricity) and found that models including both shape descriptors and body measures presented the best performance.

To highlight the recent growth and importance of Computer Vision in livestock, four recent review articles (NASIRAHMADI *et al*., 2017; WURTZ *et al*., 2019; FERNANDES *et al*., 2020; OLIVEIRA *et al*., 2021) explored the potential of computer vision systems for high-throughput phenotyping. Wurtz *et al*. (2019) and Nasirahmadi *et al*. (2017) discussed the advances on automated and high throughput image detection of farm animal behavioral traits, focusing on animal welfare and production. The article published by Fernandes *et al*. (2020a) provided an overview of key concepts related to computer vision, image processing, image analyses, and the types of devices and image ranging systems. Fernandes *et al*. (2020) also provided important key metrics for prediction quality assessment, and a compilation of animal phenotypes used for management purpose (e.g., body condition score, body weight, animal behavior, etc.), predicted using image analyses. An overview on how deep learning has been implemented and how it can be an effective tool to predict animal phenotypes, accelerating the development of predictive modeling for precise management decisions in livestock, was presented by Oliveira *et al*. (2021).

3. Deep Learning for Computer Vision Systems in Livestock

Several papers using deep learning algorithms as the main framework of image analyses in computer vision systems have been published in the last few years. Oliveira *et al*. (2021) performed a systematic review of peer-reviewed articles in which deep learning algorithms were primarily used for image analyses of livestock animals. From the total reviewed articles, 48% implemented deep learning algorithms designed to perform image classification (Figure 3b). In comparison, 25% used algorithms for object detection and only 9% for object segmentation. Some articles combined algorithms to perform image analyses, such as object detection and classification (9%) and object detection and segmentation (2%).

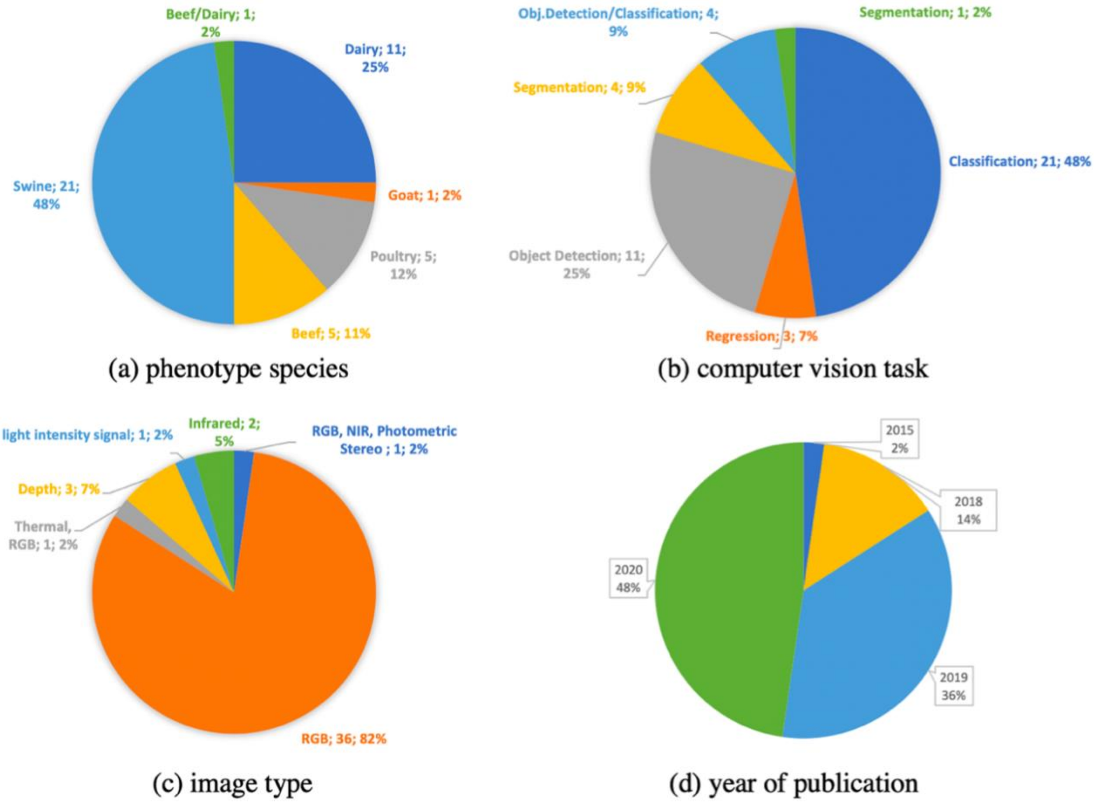

Figure 3. Descriptive information of the reviewed papers about computer vision applied to livestock species. Adapted from Oliveira *et al*. (2021).

The image analyses tasks previously mentioned (i.e., image classification, object detection, and image segmentation) are the most used in computer vision systems. Image classification is one of the most popular tasks and its main goal is to determine if a given object appears in an image. For example, image classification can be used to predict if there is a calf in the image or not (Figure 4A). Image classification can be extended to a multiclass problem, in which more than two classes could be used, and a deep learning algorithm could be applied to classify if there is a calf, a cow, or no animal in the image. Several deep learning approaches using different strategies or architectures have been proposed for image classification. One of the first deep learning architectures proposed employed convolutional and fully connected layers to handle feature extraction and classification in a single model. Such architecture brought a leap in performance that sparked a revolution in image analysis (OLIVEIRA *et al*., 2021). Object detection is another essential research topic in computer vision covered by extensive literature (Figure 4B). Deep learning approaches consistently rank among the state-of-the-art for object detection tasks and can be roughly divided into region proposal- and regression-based methods (LI *et al*., 2020). The first method aims to classify objects regions for one or more categories in the image while the regression-based method detects objects by treating their coordinates as a regression problem (OLIVEIRA *et al*., 2021). Semantic segmentation is a fundamental component of many computer vision systems, and it involves partitioning images into multiple segments or objects and labeling such segments with known classes (Figure 4C). While this field has a long history of research, deep learning networks delivered models with remarkable performance for segmentation in the last few years, becoming the new standard for image segmentation.

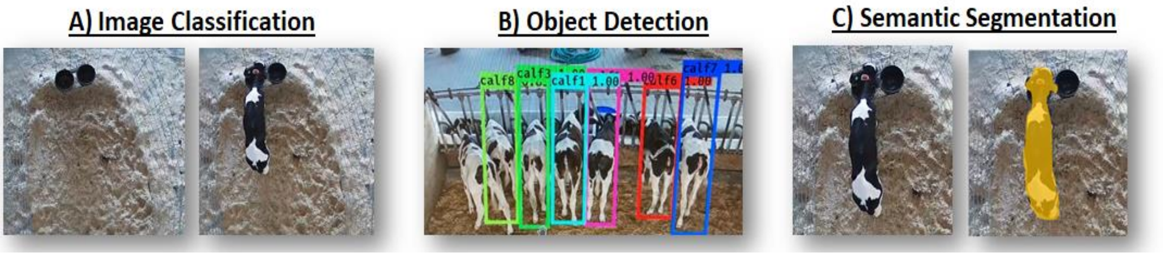

Figure 4. Image tasks using deep learning algorithms. Image Classification (Image A): classify if there is a calf in the image or not; Object detection (Image B): detect individual calves in the image; Semantic Segmentation (Image C): segmenting the calf body by detecting the pixels related to the target class (calf body).

Oliveira *et al*. (2021) reported that 7% of the articles had used deep neural network architecture for image classification to accommodate a regression problem (Figure 3b). In such cases, the major change in the deep learning algorithms was the last activation layer, modified to output the predicted values as a continuous numeric value instead of a class probability. These computer vision tasks were implemented mostly using RGB images, which represented 82% (36 articles) of the total type of images used in the reviewed articles (Figure

3c). Depth and infrared images were very little studied, with 7% (3 articles) and 5% (2 articles), respectively.

The use of deep learning for image analyses in animal sciences is very recent and it is growing extremely fast. For example, in the review published by Oliveira *et al*. (2021), the oldest peer-reviewed article retrieved was published in 2015 (Zhao and He, 2015). Additionally, 84% of the total reviewed articles (n = 44) were published in 2019 (n = 17) and 2020 (n = 20), while 16% were published in 2015 (n = 1) and 2018 (n = 6) (Figure 3d). These results confirm the rapid and recent interest in deep learning as the primary algorithm for analyzing images in computer vision systems. Most of the articles have been published in the journal Computers and Electronics in Agriculture (n = 13), followed by Biosystems Engineering (n = 6), IEEE Access (n = 4), and Sensors (n = 4). Swine (n = 21) was the most frequent species found in the reviewed articles, followed by dairy (n = 11) and beef (n = 5) cattle, poultry (n = 4), and goat (n = 1) (Figure 3a). The most frequently investigated scenario was animal behavior monitoring (n = 12), followed by animal detection/counting (n = 8), animal recognition (n = 8), health status and lameness detection (n = 4), animal pose estimation (n = 4), body weight and body condition assessment (n = 3), and others (n = 5) (Fig. 13(b)). Based on the Oliveira *et al*. (2021) review paper, it is clear that Deep Learning for image analyses have not been fully explored yet in Animal Sciences. Here, we will provide a briefly description on fundamental concepts of deep neural networks, or the so-called “deep learning” algorithms.

Deep learning algorithms were inspired by how the human brain works, using an enormous number of neurons linked by a massive number of connections to execute complex activities including speaking, moving, thinking, and seeing (GOODFELLOW *et al*., 2016). Most deep learning architectures are artificial neural networks composed of multiple layers, thus being called "deep", and a basic element called neuron (GOODFELLOW *et al*., 2016). Neurons are commonly grouped in layers, where all neurons have the same function, but each of them learns different parameters. A sequence of layers will continuously transform the input data and map it into a desired outcome in a process called feedforward. The weights of each connection between different neurons are optimized through a learning process called backpropagation (GOODFELLOW *et al*., 2016). During this optimization process, the error (difference between the observed and predicted outcome) is computed and backpropagated through the network using gradients. Those gradients are used to update the weights of each connection between neurons to minimize the error observed in the outcome. Thus, the network learns the optimal parameters for the neurons and the weight that each connection requires to predict the desired outcome. The error minimization usually leads to the convergence of parameter values in the network architecture, resulting in accurate and precise predictions of new data points or images. Different layers are commonly used to build deep networks: fully connected or dense, convolutional, deconvolutional, pooling, recurrent, and others. Fully connected or dense layers are composed of neurons with a single activation function that receives a numerical value as input, applies the function, and outputs the resulting value. Convolutional layers implement convolution operations using kernels, where each node convolves its kernel on the input image and outputs the convolved image. Convolutional layers can also be used to change image scales through the network using strides that create output images smaller than the input or using the transpose of convolutions to create output images larger than the input (often called deconvolutional layers). Pooling and up sampling layers do the same by aggregating input image values into smaller images, or interpolating smaller images' values into bigger images, respectively. Those are often used together with convolutional layers to create encoders and decoders, for example, for image segmentation. Many other different types of layers have been proposed in the literature, and the reader should refer to Goodfellow *et al*. (2016) for more detailed information. The concatenation of layers allows the creation of a complex deep network for a given task. To illustrate that, we present a basic network architecture to identify digits using the public MNIST dataset, which comprises many digit images (ranging from 0 to 9) and their corresponding labels. The task to be tackled is to identify the correct label given an input image using a convolutional neural network architecture for digit image classification, as depicted in Figure 8.

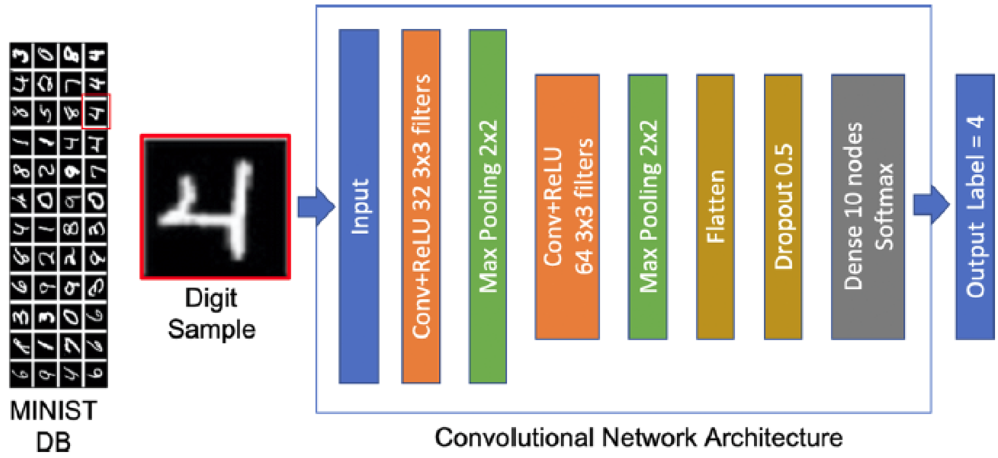

Figure 8. Example of a simple convolutional neural network architecture for digit images classification. Adapted from Oliveira *et al*. (2021)

It comprises an input layer that receives samples of digit images, followed by two blocks of convolutional layers with rectified linear unit as the activation function and max pooling layers. The compact image feature maps are then flattened to derive a feature array, and finally used for classification through a fully connected or dense layer with *softmax* activation function, which represents the probability distribution over *n* different classes (GOODFELLOW *et al*., 2016). The network also contains a dropout layer, which removes some of the connections between nodes to improve network generalization by reducing overfitting (GOODFELLOW *et al*., 2016). The loss function used was the categorical cross-entropy and the optimizer Adam (KINGMA; BA, 2017), with default values. Training consisted of presenting a batch of 128-digit image samples and computing the error using the categorical cross-entropy loss function that was backpropagated to optimize the network weights for 15 epochs or iterations. That straightforward procedure manages to deliver 99% of overall accuracy for digit images classification in unseen test images.

4. Public Databases for Image Analyses

It is important to highlight that most of the progress observed in the computer vision community was boosted by several publicly available datasets, challenges, and benchmarks, as the MNIST dataset presented above. Several other datasets are available such as PASCAL VOC (Visual Object Classes) (EVERINGHAM *et al*., 2010), ImageNet (DENG *et al*., 2009) and MSCOCO (LIN *et al*., 2014). PASCAL VOC is a popular dataset with annotated images available for different tasks: classification, segmentation, detection, action recognition, and person layout. The segmentation task comprises 21 classes of object labels with 1,464 images for training, 1,449 for validation, and a private test set for the actual challenge. The ImageNet dataset was also created as a collaboration between Stanford University and Princeton University, currently holding around fourteen million images initially labeled with synsets, or semantically meaningful set of words, from the WordNet (FELLBAUM, 1998) lexicon tree. The first challenge was to perform a simple classification task, in which each image was labeled to a single category among several hundreds. Although this challenge is still ongoing, it has further evolved into a multi-classification task where individual instances of the objects in the images were classified and located with bounding boxes. MSCOCO is a large-scale dataset for object detection, segmentation, and captioning. It includes scene imagery containing everyday objects in their natural contexts, with a total of 2.5 million labeled instances in 328,000 images. The detection challenge comprises more than 80 classes, providing more than 82,000 images for training, 40,500 for validation, and more than 80,000 images for testing. Although many of these public image datasets contain images of different species of animals, including pig, cattle and poultry, there are still few datasets designed specifically for use in livestock computer vision systems as presented in Table 1.

One can observe that most of these datasets have been published by the University of Bristol and refer to cattle detection and classification. The Holstein Cattle Recognition dataset (BHOLE *et al*., 2021) consists of thermal and RGB images from 136 animals. The Newcastle dataset (ALAMEER, 2020) is based on frames of pigs manually annotated into one of five categories for postures and drinking. The FriesianCattle2015 (ANDREW *et al*., 2016) and FriesianCattle2017 (ANDREW *et al*., 2017) datasets have depth-segmented RGB images of Friesian Cattle. The Cows2021 dataset (GAO *et al*., 2021) consists of top-view images from 186 Holstein-Friesian cattle with manual bounding boxes annotation and animal identities. The Aer-ialCattle2017 (ANDREW *et al*., 2019) dataset comprises images of tracked Friesian cattle ROIs filmed by Unmanned Aerial

Vehicle (UAV). The Zenodo dataset (PEREIRA *et al*., 2020) holds video data over a two-month period of cows in front of automatic milking stations with manually annotated behavioral classes. The OpenCows2020 dataset (ANDREW *et al*., 2020) consists of top-down images of Holstein cattle taken both indoors and outdoors, and it was designed for detection, localization, and identification tasks.

Table 1. Image datasets of livestock animals categorized by author, animal species, image analysis tasks, and image type.

Database	Authors	Species	Task	Image Type
Holstein Cattle Recognition	Bhole *et al*., 2021	Cattle	Classification	Thermal and RGB
FresianCattle2015	Andrew *et al*., 2016	Cattle	Classification	Top-View RGB
FresianCattle2017	Andrew *et al*., 2017	Cattle	Classification	Top-View RGB
AerialCattle2017	Andrew *et al*., 2019	Cattle	Detection	Aerial RGB from UAV
Zenodo	Pereira *et al*., 2020	Cattle	Detection	Video Sequences - RGB images
Cows2021	Gao *et al*., 2021	Cattle	Detection Classification	Top-View RGB
OpenCows2020	Andrew *et al*., 2020	Cattle	Detection Classification	Top-View RGB
Newcastle	Alameer, 2020	Pig	Detection	Top-View RGB

Public datasets in livestock are still limited and, therefore, a recent USDA grant was awarded to four universities: Michigan State University, University of Wisconsin-Madison, and University of Nebraska-Lincoln in the USA, and University of Leuven in Belgium, with the main objective to build the largest public database on livestock images. This research group will create and make publicly available a large dataset with images and videos, and articulate deficiencies in the current computer vision algorithms for livestock.

5. Final Considerations

Studies developing modeling strategies to analyze longitudinal imaging data in animal science are scarce. Moreover, the literature lacks studies developing integrated and long-term computer vision systems. The need for a long-term computer vision framework that can add new animal phenotypic measurements at a minimal cost, by leveraging the computational resources previously developed (image acquisition and storage, pre-processing, and automation for analyzes and data transfer) is critical to advance precision livestock farming.

Initiatives led by private and public institutions to create a large animal science community working on artificial intelligence is crucial to advance cost-effective and reliable applications of digital technologies in commercial farms. Animal scientists interested in artificial intelligence should seek for formal training at educational institutions but should also leverage on existing public and open-source algorithms, databases, and educational material to further develop their analytical skills on AI technologies.

6. References

ALAMEER, A.; KYRIAZAKIS, I.; BACARDIT, J. Automated recognition of postures and drinking be- haviour for the detection of compromised health in pigs. **Scientific Reports**, v. 10, p. 13665, 2020. https://doi.org/10.25405/data.ncl.13042619.v1. 2020.

ANDREW, W. *et al*. **Opencows** 2020. https://doi.org/10.5523/bris.10m32xl88x2b61zlkkgz3fml17. 2020.

ANDREW, W.; GREATWOOD, C.; BURGHARDT, T. Aerial animal biometrics: Individual friesian cattle recovery and visual identification via an au- tonomous uav with onboard deep inference, In: **IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)**, p. 237–243, 2020. https://doi.org/10.1109/IROS40897.2019.8968555. 2019.

ANDREW, W.; GREATWOOD, C.; BURGHARDT, T. Visual localisation and individual identification of holstein friesian cattle via deep learning. In: **IEEE**

International Conference on Computer Vision Workshops (ICCVW), p. 2850–2859, 2017. https://doi.org/10.1109/ICCVW.2017.336.

ANDREW, W. *et al*. Automatic individual holstein friesian cattle identification via selective local coat pattern matching in rgb-d imagery, In: **IEEE International Conference on Image Processing** (ICIP), p. 484–488, 2016. https://doi.org/10.1109/ICIP.2016.7532404.

ATHERTON, T. J.; KERBYSON, D. J. Size invariant circle detection. **Image and Vision Computing**, v. 17, p. 795–803, 1999. doi: https://doi.org/10.1016/S0262-8856(98)00160-7

BAY, H.; TUYTELAARS, T.; VAN GOOL L. SURF: Speeded Up Robust Features. In: **European Conference on Computer Vision**. Berlin, Heidelberg: Springer, p. 404–417, 2006. https://doi.org/10.1007/11744023_32.

BERCOVICH, A. *et al*. Development of an automatic cow body condition scoring using body shape signature and Fourier descriptors. **Journal of Dairy Science**, v. 96, p. 8047–8059, 2013.

BHOLE, A. *et al*. **Holstein cattle recognition** URL, 2021. https://doi.org/10.34894/O1ZBSA.

BYRNE, D. T. *et al*. Temporal, spatial, inter-, and intra-cow repeatability of thermal imaging. **Journal of Animal Science**, v. 95, p. 970–979, 2017.

CIRESAN, D.; MEIER, U.; SCHMIDHUBER, J. Multi-column deep neural networks for image classification. **IEEE Conference on Computer Vision and Pattern Recognition**, 2012. https://doi.org/10.1109/cvpr.2012.6248110.

COMINOTTE, A. *et al*. Automated computer vision system to predict body weight and average daily gain in beef cattle during growing and finishing phases. **Livestock Science**, v. 232, p. 103904, 2020.

DENG, J. *et al*. Imagenet: a large-scale hierarchical image database. **IEEE Conference on Computer Vision and Pattern Recognition** (CVPR). IEEE, p. 248–255, 2009.

DÓREA, J. R. R. *et al*.Use of 3-dimensional camera to predict body weight in pre-weaned dairy calves. **ADSA Annual Meeting**, Knoxville – TN, 2018.

EVERINGHAM, M. *et al*. The pascal visual object classes (VOC) challenge. **International Journal of Computer Vision**, v. 88, v. 2, p. 303—338, 2010. https://doi.org/10.1007/s11263-009-02754.

FELLBAUM, C. (ED.). **WordNet**: An Electronic Lexical Database, Language, Speech, and Communication. MIT Press, Cambridge, MA, 1998.

FERNANDES, A. F. A; DOREA, J. R. R.; ROSA, G. J. M. Image Analysis and Computer Vision Applications in Animal Sciences: An Overview. **Frontiers in Veterinary Science**, v. 21; n. 7, p. 551269, 2020a. https://doi.org/10.3389/fvets.2020.551269.

FERNANDES, A. F. A. *et al*. A novel automated system to acquire biometric and morphological measurements, and predict body weight of pigs via 3D computer vision. **Journal of Animal Science**, v. 97, p. 496-508, 2019.

GAO, J. *et al*. **Towards self-supervision for video identification of individual holstein- friesian cattle**: 2021.The cows2021 dataset. arXiv:2105.01938.

GOMES, R. A. *et al*. Technical note: Estimating body weight and body composition of beef cattle trough digital image analysis. **Journal of Animal Science**, v. 94, p. 5414–5422, 2016.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. Cambridge, MS: MIT Press, 2016.

GUZHVA, O. *et al*. Now you see me: convolutional neural network based tracker for dairy cows**. Frontiers in Robotics and AI**, v.19, 2018. https://doi.org/10.3389/frobt.2018.00107.

HALACHMI, I. *et al*. Automatic assessment of dairy cattle body condition score using thermal imaging. **Computers and Electronics in Agriculture, v.** 99, p. 35–40, 2013. https://doi.org/10.1016/j.compag.2013.08.012

HERTEM, T. V. *et al*. Implementation of an automatic 3D vision monitor for dairy cow locomotion in a commercial farm. **Biosystems Engineering**, v. 173, p. 166-175, 2017.

HOUGH, P. V. C. **Method and means for recognizing complex patterns**. 1962. Available online at: https://www.osti.gov/scitech/biblio/4746348 (accessed April 17, 2017)

HOVINEN, M. *et al*. Detection of Clinical Mastitis with the Help of a Thermal Camera. **Journal of Dairy Science**, v. 91, p. 4592-4598, 2008.

KINGMA, D. P.; BA, J. Adam: **A Method for Stochastic Optimization**. 2017. Available in https://arxiv.org/abs/1412.6980. Accessed in november,1, 2021.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **Nature**, v. 521, p. 436–444, 2015.

LEUTENEGGER, S.; CHLI, M.; SIEGWART, R.Y. BRISK: Binary robust invariant scalable keypoints. In: International Conference on Computer Vision, **Proceedings...** (Barcelona: IEEE). p. 2548–2555, 2011. https://doi.org/10.1109/ICCV.2011.6126542.

LI, K. *et al*. Object detection in optical remote sensing images: a survey and a new benchmark. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 159, p. 296-307, 2020. https://doi.org/10.1016/j.isprsjprs.2019.11.023.

LIN, T. Y. *et al*. **Microsoft COCO: Common Objects in Context. arXiv**. Available online at: https://arxiv.org/abs/1405.0312 (accessed December 5, 2018), 2014.

MURPHYM, K. P. **Machine Learning: A probabilistic perspective**. Cambridge, Massachusetts: MIT Press, 2012. Available at: http://mitpress.mit.edu (accessed June 18, 2019).

NASIRAHMADI, A.; EDWARDS, S. A.; STURM, B. Implementation of machine vision for detecting behaviour of cattle and pigs. **Livestock Science**, v. 202, p. 25-38, 2017.https://doi.org/10.1016/j.livsci.2017.05.014

OLIVEIRA, D. A. B. *et al*. A review of deep learning algorithms for computer vision systems in livestock. **Livestock Science**, v. 253, p.104700, 2021.

PEREIRA, L. S. B. *et al*. **Data set of labeled scenes in a barn in front of automatic milking system**. 2020. https://doi.org/10.5281/zenodo.3981400.

PORTO, S. M. *et al*. The automatic detection of dairy cow feeding and standing behaviours in freestall barns by a computer vision based system. **Biosystems Engineering**, v. 133, p. 46–55, 2015.

SALAU, J. *et al*. Automated calculation of udder depth and rear leg angle in Holstein-Friesian cows using a multi-Kinect cow scanning system. **Biosystems Engineering**, v. 160, p. 154–169, 2017.

SONG, X. *et al*. Automated body weight prediction of dairy cows using 3-dimensional vision. **Journal of Dairy Science**, v. 101, p. 4448-4459, 2018.

SZELISK, R. **Computer vision:** algorithms applications. In: Gries, D.; Schneider, F. B. London: Springer-Verlag, 2011. https://doi.org/10.1007/978-1-84882-935-0

SPOLIANSKY, R. *et al*. Development of automatic body condition scoring using a low-cost 3-dimensional Kinect camera. **Journal of Dairy Science**, v. 99, p. 7714–7725, 2016.

TAIGMAN, Y. *et al*. DeepFace: Closing the Gap to Human-Level Performance in Face Verification. IEEE Conference on Computer Vision and Pattern Recognition, **Proceedings...** 2014. https://doi.org/10.1109/cvpr.2014.220

TEDÍN, R. *et al*. **Towards automatic estimation of the body condition score of dairy cattle using hand-held images and active shape models**. Universidade da Coruña, 10 p, 2012.

TSAI, D. M.; HUANG, C. Y. A motion and image analysis method for automatic detection of estrus and mating behavior in cattle. **Computers and Electronics in Agriculture**, v. 104, p. 25–31, 2014.

VIAZZI, S. *et al*. Comparison of a three-dimensional and two-dimensional camera system for automated measurement of back posture in dairy cows. **Computers and Electronics in Agriculture**, v. 100, p. 139–147, 2014.

WANG X. *et al*. Visualizing quantitatively the freshness of intact fresh pork using acousto-optical tunable filter-based visible/near-infrared spectral imagery. **Computers and Electronics in Agriculture**, v. 99, p. 41-53, 2013. https://doi.org/10.1016/j.compag.2013.08.025

WURTZ, K. *et al*. Recording behavior of indoor- housed farm animals automatically using machine vision technology: A systematic review. **Plos One**, v. 14, p. e0226669, 2019. https://doi.org/10.1371/journal.pone.0226669.

ZHAO, K. *et al*. Automatic lameness detection in dairy cattle based on leg swing analysis with an image processing technique. **Computers and Electronics in Agriculture**, v. 148, p. 226–236, 2018.

ZHAO, K.; HE, D. Recognition of individual dairy cattle based on convolutional neural networks. **Transactions of the Chinese Society of Agricultural Engineering**, 2015. https://doi.org/10.3969/j.issn.1002-6819.2015.05.026.

Autores

João R. R. Dorea[1], Guilherme J. M. Rosa[1], Arthur F. A. Fernandes[2], Tiago Bresolin[1]

1. Department of Animal and Dairy Sciences, University of Wisconsin-Madison, U.S.A.
2. Cobb-Vantress