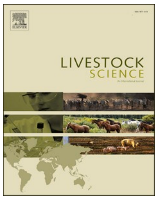
Contents lists available at ScienceDirect

Livestock Science

journal homepage: www.elsevier.com/locate/livsci



A review of deep learning algorithms for computer vision systems in livestock

Dario Augusto Borges Oliveira[a], Luiz Gustavo Ribeiro Pereira[a,b], Tiago Bresolin[a], Rafael Ehrich Pontes Ferreira[a], Joao Ricardo Reboucas Dorea[a,*]

[a] *Department of Animal and Dairy Sciences, 1675 Observatory Drive, 266 Animal Sciences Building, Madison, WI 53706-1205*

[b] *Embrapa Dairy Cattle, Av. Eugênio do Nascimento, 610 - Aeroporto, Juiz de Fora - MG, 36038-330, Brazil*

HIGHLIGHTS

- Review of animal science studies that used deep learning in computer vision systems.
- Greater adoption of deep learning algorithms for image classification.
- The phenotype with greater interest was animal behavior..
- Swine was the most frequent species found in the reviewed articles.

ARTICLE INFO

Keywords:
Artificial intelligence
Cattle
Machine learning
Precision farming
Swine

ABSTRACT

In livestock operations, systematically monitoring animal body weight, biometric body measurements, animal behavior, feed bunk, and other difficult-to-measure phenotypes is manually unfeasible due to labor, costs, and animal stress. Applications of computer vision are growing in importance in livestock systems due to their ability to generate real-time, non-invasive, and accurate animal-level information. However, the development of a computer vision system requires sophisticated statistical and computational approaches for efficient data management and appropriate data mining, as it involves massive datasets. This article aims to provide an overview of how deep learning has been implemented in computer vision systems used in livestock, and how such implementation can be an effective tool to predict animal phenotypes and to accelerate the development of predictive modeling for precise management decisions. First, we reviewed the most recent milestones achieved with computer vision systems and the respective deep learning algorithms implemented in Animal Science studies. Then, we reviewed the published research studies in Animal Science which used deep learning algorithms as the primary analytical strategy for image classification, object detection, object segmentation, and feature extraction. The great number of reviewed articles published in the last few years demonstrates the high interest and rapid development of deep learning algorithms in computer vision systems across livestock species. Deep learning algorithms for computer vision systems, such as Mask R-CNN, Faster R-CNN, YOLO (v3 and v4), DeepLab v3, U-Net and others have been used in Animal Science research studies. Additionally, network architectures such as ResNet, Inception, Xception, and VGG16 have been implemented in several studies across livestock species. The great performance of these deep learning algorithms suggests an improved predictive ability in livestock applications and a faster inference. However, only a few articles fully described the deep learning algorithms and their implementation. Thus, information regarding hyperparameter tuning, pre-trained weights, deep learning backbone, and hierarchical data structure were missing. We summarized peer-reviewed articles by computer vision tasks (image classification, object detection, and object segmentation), deep learning algorithms, animal species, and phenotypes including animal identification and behavior, feed intake, animal body weight, and many others. Understanding the principles of computer vision and the algorithms used for each application is crucial to develop efficient systems in livestock operations. Such development will potentially have a major impact on the livestock industry by predicting real-time and accurate phenotypes, which could be used in

[*] Corresponding author.
E-mail address: joao.dorea@wisc.edu (J.R. Reboucas Dorea).

https://doi.org/10.1016/j.livsci.2021.104700
Received 4 November 2020; Received in revised form 20 July 2021; Accepted 6 September 2021
Available online 11 September 2021

the future to improve farm management decisions, breeding programs through high-throughput phenotyping, and optimized data-driven interventions.

1. Introduction

The data revolution undergoing in almost every industry sector is causing an expressive shift in science, technology, and education. Several digital technologies such as wearable sensors (Neethirajan, 2017; Rutten et al., 2013), robotic milking systems (Rodenburg, 2017), infrared spectrometry (Bresolin and Dorea, 2020), and computer vision systems (Fernandes et al., 2020b; Wurtz et al., 2019) have been developed and deployed in livestock operations. The data generated by these technologies carries an incredible value, and it can be used to generate difficult-to-measure animal-level phenotypes. However, to extract the full potential of such datasets, precise phenotypes need to be predicted and used for optimized data-driven decisions. To accomplish that, appropriate analytical tools should be correctly implemented.

Among the 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 genetic improvement in breeding programs. The amount of information carried in a single image usually goes beyond the developers primary interest when computer vision systems are created. For example, suppose the images presented in Fig. 1 are 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. Interestingly, additional information present in the image 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 primary interest is related to animal behavior and identification, the amount of information in the image can allow for future development, 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.

Three recent review articles (Fernandes et al., 2020a; Wurtz et al., 2019; Nasirahmadi, Edwards, Sturm) explored the potential of computer vision systems for high-throughput phenotyping. Wurtz et al. (2019) and Nasirahmadi et al. (2017) discussed the advances that occurred in 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., 2020a 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 through image analyses.

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. The rapid and recent implementation of such algorithms for image analyses and the great predictive ability reported by the published literature indicates a powerful analytical tool to predict animal-level phenotype. The objectives of this article were: (1) to provide an overview of the main deep learning algorithms used in computer vision systems that are commonly implemented in Animal Sciences or that demonstrate potential to be applied on it; and (2) to perform a systematic review of Animal Science studies that used deep learning as the main algorithm to predict phenotypes of livestock animals.

This article is divided into two main sections. The first section aims to provide an overview of the state-of-the-art deep learning algorithms used in computer vision systems and their primary milestones, considering four different tasks: image classification, object detection, object segmentation, and feature extraction. In the second section, a systematic review of studies using deep learning in computer vision systems for phenotype prediction is provided, and we finish presenting our concluding remarks. That structure intends to bring to light the gaps in livestock applications that could be potentially solved or improved with state-of-the-art computer vision algorithms. It is important to mention that this review does not aim at proposing new solutions but serves as a reference for future works building them.

2. Deep learning for computer vision systems

2.1. Deep learning basics

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 connections 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 connections between neurons to

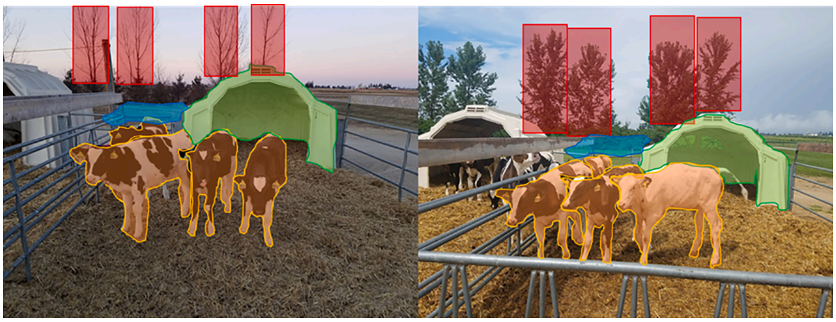


Fig. 1. Group of four Holstein dairy calves housed in a super-hutch. Figure 1a (left side) winter period, trees without leaves, and clear sky. Figure 1b (right side) summer period, trees with green leaves, and cloudy sky. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

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 parameters in the network architecture, resulting in accurate and precise predictions given a new data point or image.

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 convolutions using kernels, where each node convolves its kernel with 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 upsampling 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 image segmentation, for instance. Many other different types of layers have been continuously proposed in the literature, and the reader should refer to Goodfellow et al. (2016) and Computer Vision venues 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 label. 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 Fig. 2. 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 and 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 (Acc) for digit images classification in unseen test images.

Public databases and benchmarks

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 (2). 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 1464 images for training, 1449 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 consisted of a simple classification task, where each image was labeled to a single category among several hundred. 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 few datasets designed specifically for use in livestock computer vision systems, as seen in Table 1. One can observe that most of these datasets were published by the University of Bristol and regards cattle or cows' detection and classification. The Holstein Cattle Recognition dataset (Bhole et al., 2021) consists of thermal and RGB images from 136 animals. The

Table 1
Databases in livestock using computer vision systems.

Database	Species	Task	Image type
Holstein Cattle Recognition (Bhole et al., 2021)	Cattle	Classification	Thermal and RGB
Newcastle (Alameer, 2020)	Pig	Detection	Top-View RGB
FriesianCattle2015 (Andrew et al., 2016)	Cattle	Classification	Top-View RGB
FriesianCattle2017 (Andrew et al., 2017)	Cattle	Classification	Top-View RGB
Cows2021 (Gao et al., 2021)	Cows	Detection Classification	Top-View RGB
AerialCattle2017 (Andrew et al., 2019)	Cattle	Detection	Aerial RGB from UAV
Zenodo (Benitez Pereira et al., 2020)	Cows	Detection	Video Sequences
OpenCows2020 (Andrew et al., 2020)	Cows	Detection Classification	Top-View RGB

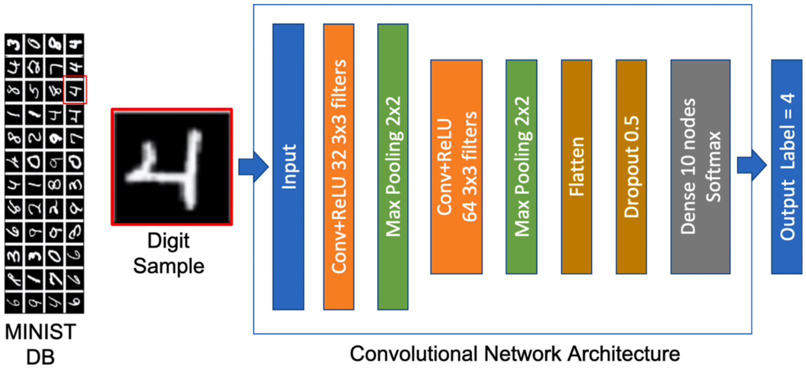


Fig. 2. Example of a simple convolutional neural network architecture for digit images classification.

Newcastle dataset (Alameer, 2020) contains 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 consist of depth-segmented RGB images of Friesian Cattle. The Cows2021 dataset (Gao et al., 2021) consists of top-view images from a herd of 186 Holstein-Friesian cattle with manual annotation of bounding boxes and animal identities. The AerialCattle2017 (Andrew et al., 2019) dataset comprises images containing tracked Friesian cattle ROIs filmed by UAV. The Zenodo dataset (Benitez Pereira et al., 2020) consists of video data over two-month period of cows in front of automatic milking stations and 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 was designed for detection, localization, and identification tasks.

In the next subsections, we will explore different models for handling different image analysis tasks using deep learning, and briefly present the advantages and progress implemented in each algorithm.

2.2. Image classification

Image classification is one of the most popular tasks for computer vision applications and its main goal is identifying 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. The task can also be expanded 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 used convolutional and fully connected layers to handle feature extraction and classification in a single model. Such architecture brought a leap in performance that began a revolution in image analysis. Inspired by Krizhevsky et al. (2012); Lecun et al. (1998) proposed a model that outperformed the previous best model based on contemporary state-of-the-art feature extraction by a considerable margin in the ImageNet challenge, of almost 11%. The AlexNet model, currently considered a simple architecture, comprises five consecutive convolutional filters, max-pool layers, and three fully connected layers for classification. Since this breakthrough, different research groups have extensively explored the development of new network architectures. In 2014, K. Simonyan and A. Zisserman published the VGG16 model (Simonyan and Zisserman, 2014), still used today for image classification problems (Figure 3). It comprises sixteen convolutional layers, multiple max-pool layers, and three final fully connected layers. They also proposed to use rectified linear unit activation functions to chain the numerous convolutional layers, creating means for highly nonlinear transformations in the model. Additionally, they offered to use smaller convolution kernels and showed they could extract the same features using much fewer parameters. This model delivered an error rate of 7.3% in the 2014 ImageNet challenge, reducing the AlexNet model's error by a factor of 2 .

In 2014, Lin et al. (2013) proposed the concept of inception modules, conceived to focus on massive local feature extraction. Such network architecture was proposed to achieve the same performance of sequential stacked layers by simultaneously training multiple convolutional layers and stacking their feature maps linked with a multi-layer perceptron. C. Szegedy et al. have explored this idea in GoogLeNet (Szegedy et al., 2014), commonly known as the first Inception network, with 22 layers using inception modules that contain a total of 50 convolution layers (Fig. 4). Each inception module implemented convolution layers with different kernel sizes to extract features at different scales. The feature maps produced were then concatenated and analyzed by the next inception module, and a final dense layer was used to map features into class labels. The GoogLeNet model achieved a 6.7% error rate over the 2014 ImageNet challenge, comparable to the VGG16 result, but using only nearly 10% of its parameters.

In 2015, Szegedy et al. (2015) developed the Inception-v2 model, which was mostly inspired by the first version, but with a significant modification that replaced the convolutional layers in the inception module with a combination of convolutional and fully-connected layers. They called this modification a convolution factorization, decreasing the number of parameters in each inception module, reducing the computational cost. This model reached a top-5 error rate of 5.6% on the 2012 ImageNet challenge. C. Szegedy et al. also further proposed the famous Inception-v3 model, fine-tuning the batch-normalization, using a higher resolution input, reducing the strides of the first two layers, and removing a max-pool layer to analyze images with higher precision. This model reached a top-5 performance at the 2012 ImageNet challenge with an error rate of 3.58%.

In 2015, He et al., 2015a noticed that merely increasing the depth of models also increased their error rate, not due to overfitting but to the difficulties of training and optimizing very deep models. The authors proposed the residual learning strategy (ResNet), through a connection between the output of one (or multiple) convolutional layers and their original input with an identity mapping, which would allow the inbetween layers to learn residuals (Fig. 5). The model could then learn a residual function between inputs and outputs that keeps most of the input information and produces only slight output changes. Consequently, patterns from the input image can be held to deeper layers. The

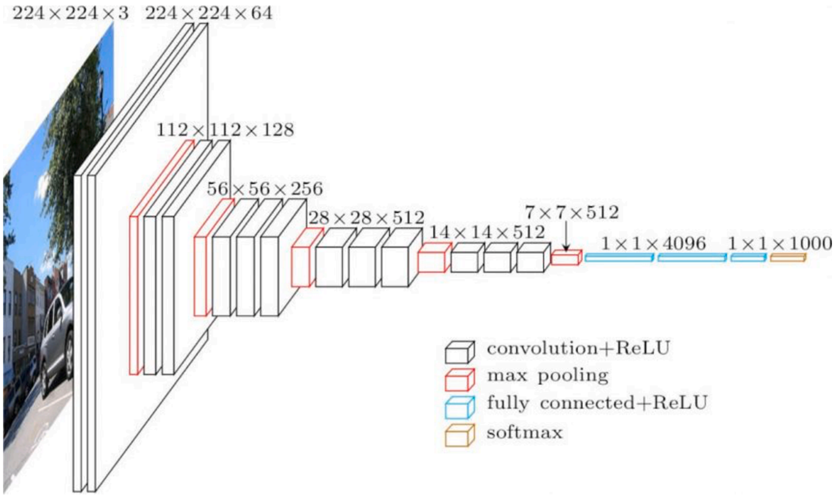
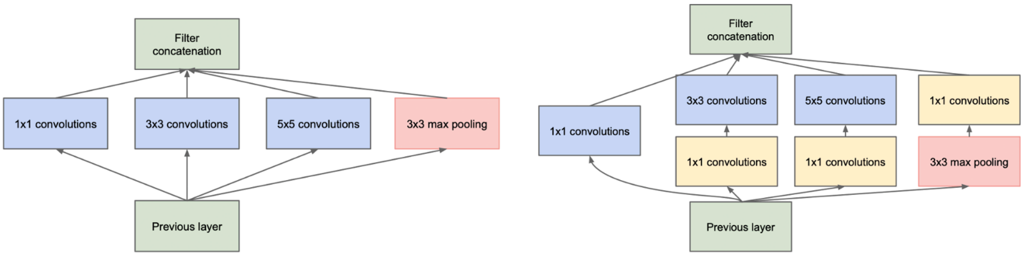


Fig. 3. VGG-16 architecture. Image from Simonyan and Zisserman (2014).



(a) Inception module, naïve version (b) Inception module with dimension reductions

Fig. 4. Inception modules. Image from Szegedy et al. (2014).

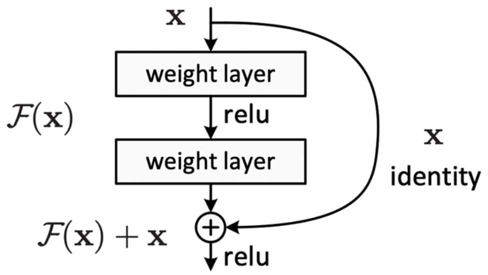

x
weight layer
$\mathcal{F}(\mathbf{x})$
relu
weight layer
x
identity
$\mathcal{F}(\mathbf{x}) + \mathbf{x}$
relu

Fig. 5. Residual learning modules. Image from He et al., 2015a.

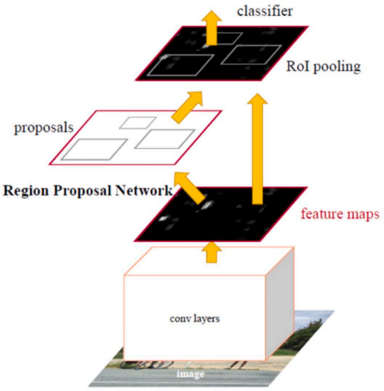


Fig. 6. Faster R-CNN architecture. Image from Ren et al. (2015).

original ResNet model used 152 convolutional layers organized in blocks of two layers with residual learning and delivered an error rate of 3.57% on ImageNet 2015 challenge.

In 2016, C. Szegedy et al. combined inception modules and residual blocks, building residual inception blocks. The resulting Inception-v4 (Inception-ResNet) model (Szegedy et al., 2016) allowed faster training and outperformed other models in the 2012 ImageNet challenge with an error rate of 3.08%. While most of the recent advances in image classification combine some of the algorithms previously discussed, some of them focus on better understanding the models and proposing more efficient approaches. In this sense, in 2017, Coudet et al. proposed xCeption (Chollet, 2016) (from extreme inception), which was inspired based on Inception-v3 and featured a separable convolution component to run convolutions in images more efficiently. It achieved the state-of-the-art results in 2017 for ImageNet with much smaller models in terms of the number of parameters. MobileNet (Howard et al., 2017) also proposed to use separable convolutional layers to build a family of lightweight networks designed to work primarily in mobile applications. More recently, NASNet (Pham et al., 2018) proposed a neural architecture search methodology to create optimal neural networks by searching the best assembling combinations from a set of commonly used operations. The authors proposed the use of a controller neural network to form components with high performance instead of designing the network manually for a given task. While NASNet can deliver very efficient networks for a specific task, the whole process requires a massive computational effort for exploring different design possibilities (Pham et al., 2018).

2.3. Object detection

Object detection is another essential research topic in computer vision covered by extensive literature. 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 proposes a classification of object regions for one or more categories in an image while the regression-based method detects objects by treating their coordinates as a regression problem. One of the first thriving region proposal-based methods was the region-based Convolutional Neural Network (CNN) introduced by Girshick et al. (2016). Such algorithm generates a large number of region proposals using selective search (Uijlings et al., 2013), then extracts deep convolutional features from these regions using CNN, and finally trains a support vector machine to label object candidates into PASCAL VOC classes.

Many improvements were added to this initial idea of region-based CNN, resulting in new approaches. With Fast R-CNN (Girshick, 2015), for example, the authors proposed to feed the input image into a CNN, generating a convolutional feature map to find the region proposals,

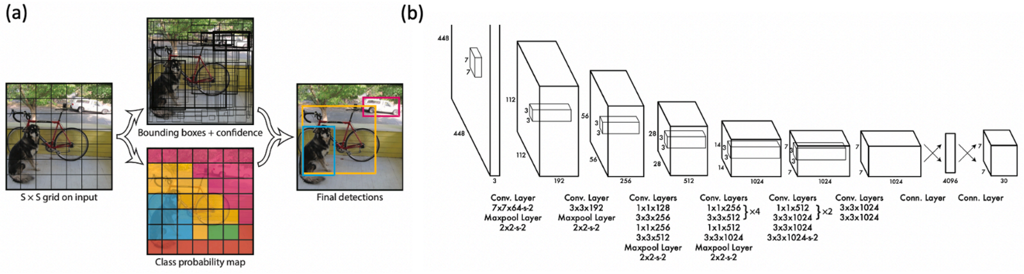


Fig. 7. YOLO architecture. Images from Redmon et al. (2015).

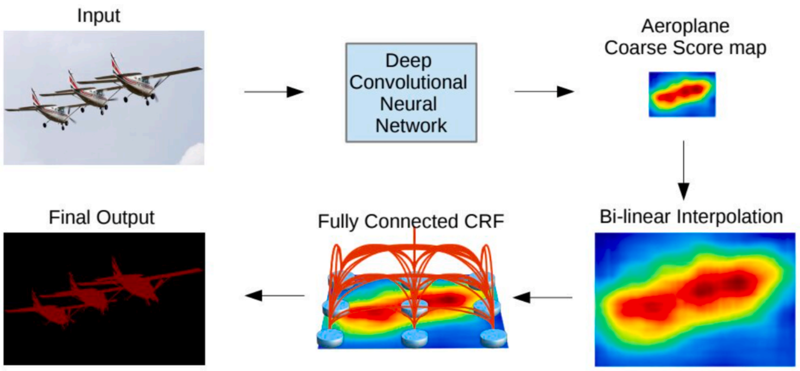


Fig. 8. CNN + CRF architecture. Image from Chen et al., 2014a.

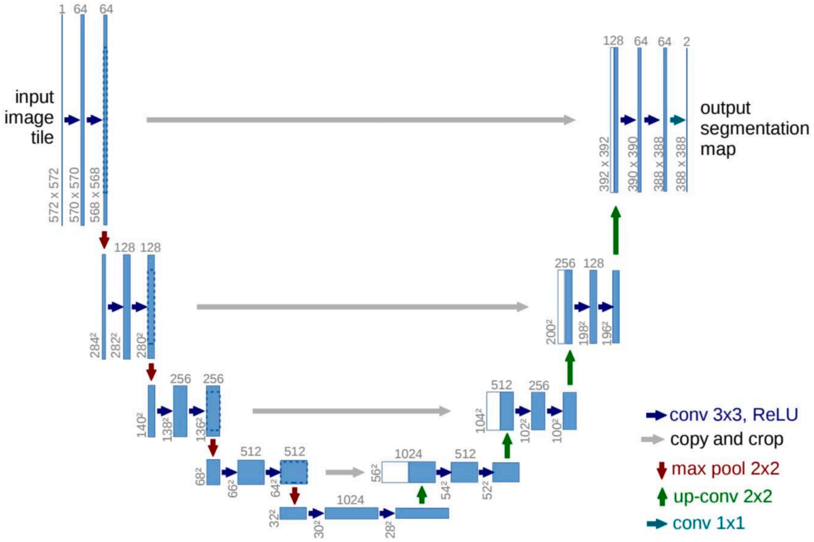


Fig. 9. unet architecture. Image from Ronneberger et al. (2015).

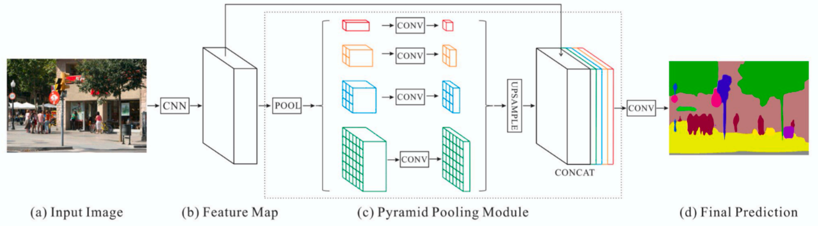


Fig. 10. PSPNet architecture. Image from Zhao et al. (2017).

instead of feeding only the region proposals from the original image into a CNN. In Faster R-CNN (Ren et al., 2015) the authors decided to use a separate network to predict the region proposals instead of using selective search algorithm on the feature map (Fig. 6). For the Cascaded R-CNN (Cai and Vasconcelos, 2018) the authors proposed to use a sequence of cascaded object detectors with increasing accuracy based on R-CNN-like networks. To be more selective against close false positives, the authors adopted this sequence of detectors trained with increasing IoU (Intersection over Union) thresholds.

Regression-based networks for object detection were firstly proposed in OverFeat (Sermanet et al., 2014) and became very popular with YOLO (Redmon et al., 2015) and RetinaNet (Lin et al., 2017). YOLO is a regression network that uses a single CNN backbone to predict bounding boxes with label probabilities in a single evaluation schema (Redmon et al., 2015). This architecture creates a regular grid to split the input image space and locate objects at the grid centers (Fig. 7). Each cell derives several bounding boxes and class probabilities, and the network training consists of maximizing these values for each cell. In YOLOv2, many modifications from the original architecture (YOLO) were proposed to improve its precision, including replacing fully connected layers by anchor boxes (has fewer parameters and is more robust) for predicting bounding boxes, similar to the proposed in Single Shot Multibox detector - SSD (Liu et al., 2016). For YOLOv3, the idea of using anchor boxes was improved using dimension clusters to predict bounding boxes and logistic classifiers to produce class probabilities for each bounding box, instead of the usual softmax. RetinaNet uses ResNet (He et al., 2015a) and Feature Pyramid (Lin et al., 2017a) networks as the backbone for feature extraction concatenated to two task-specific

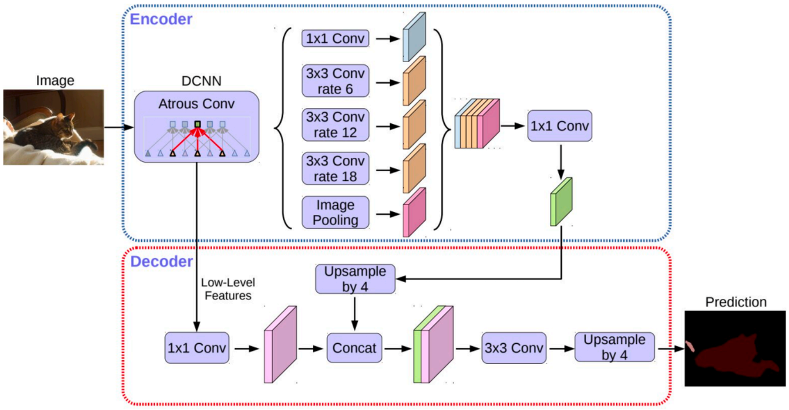


Fig. 11. DeepLab v3+ architecture. Image from Chen et al. (2018).

subnetworks for classification and bounding box regression using focal loss, in a single-stage regression training schema, outperforming Faster R-CNN. Tiny-YOLO (Redmon et al., 2015) was proposed as smaller version of YOLO, which requires less time to train and test but is usually less accurate than the original models.

2.4. *Image segmentation*

Semantic segmentation is a fundamental part of many computer vision systems and involves partitioning images into multiple segments or objects and labeling such segments with known classes. 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. One of the first studies using deep learning for semantic segmentation was a fully convolutional network (FCN), proposed by Long et al. (2015). The model included only convolutional layers applied sequentially to an input image with arbitrary size and produced a consistent segmentation image where pixels were mapped to different classes.

The main limitation of classic FCN models was the inability to handle contextual information efficiently, and different studies explored such drawback. Chen et al., 2014a investigated the spatialization of the final layer outcome in deep CNN showing that they were not reliable for accurate object segmentation. Chen et al., 2014a further proposed to feed the CNN outcome to fully connected conditional random fields (CRF) as a way to enable better definition of segmented boundaries (Fig. 8). Their results overcame the existing methods for image segmentation at that time. To optimize CNN and CRF together, different initiative were introduced. Schwing and Urtasun (Schwing and Urtasun, 2015) and Zheng et al. (2015) proposed a fully connected deep structured network for image segmentation integrating CRF with CNN. Lin et al. (2016) proposed an efficient algorithm for semantic segmentation based on contextual deep CRF, where they explored different regions in the same images for retrieving contextual information for training. Liu et al. (2015) proposed a deterministic end-to-end approach called Parsing Network for embedding information extracted using CNN into Markov random fields models.

Encoder-decoder models are also widely used for semantic segmentation. Such models are composed of an encoder, which creates an efficient representation of the input image and a decoder that decodes the representation into target segmentation maps. This simple yet powerful framing enables different networks to collaborate and create highly optimized mappings between input and labeled data. One of the first encoder-decoder models was proposed by Noh et al. (2015) and consisted of a VGG 16-layer network for encoding input data into a feature vector, and a deconvolution network composed of deconvolution and pooling layers to identify pixel-wise labels and create segmentation masks. SegNet, which is a deep convolutional neural network architecture for semantic pixel-wise segmentation, proposed by Badrinarayanan et al. (2017), comprises an encoder based on a VGG 16-layer network, and a mirrored structure serving as a decoder. This model is lightweight, mainly due to its mechanism to upsample encoded feature maps using pooling layers convolved with trainable convolutional filters to produce dense feature maps.

These early encoder-decoder models were known to struggle for segmenting small structures since they needed to hold information of the first layers in the subsequent feature maps, which were increasingly downsampled. To overcome such limitations, different approaches used skip connections between encoder layers and decoder layers, which reinforced features at different scales. One of the most popular encoder-decoder models using skip connections is the U-Net architecture, proposed by Ronneberger et al. (2015). It comprises two branches: an encoder branch to capture context, and a symmetric decoder branch to enable precise localization (Fig. 9). Feature maps from the encoder were copied to the decoder to reinforce patterns at different scales, and a final convolution layer creates a segmentation map. Multi-scale feature information was further explored in the Feature Pyramid Network (FPN) proposed by Lin et al. (2017b), which was developed initially for object detection but was later extended to perform segmentation. In this network, the authors proposed to a multi-scale structured network using feature pyramids, where low and high-resolution features were merged using lateral connections between downsampling and upsampling branches. The extension for image segmentation was implemented using two multi-layer perceptrons to generate the segmentation masks. Focusing on multi-scale image segmentation, Zhao et al. (2017) proposed the PSPNet (Pyramid Scene Parsing Network), where patterns at different scales are extracted from the input image using residual branches and a pyramid pooling module (Fig. 10). Four different scales were considered by the authors, each one corresponding to a pyramid level, and were concatenated into a single matrix comprising both local and global features at different scales. A final convolutional layer produces the pixel-wise predictions.

Some extensions of R-CNN presented in Section 2.3 simultaneously perform object detection and semantic segmentation. He et al. (2017) proposed Mask R-CNN, a model that detects objects in images and simultaneously generates segmentation masks for each instance found (instance segmentation). The Mask R-CNN architecture uses a Faster R-CNN backbone, but with three output branches: the first computes the bounding box coordinates, the second computes the associated classes, and the third computes a binary mask segmenting the object found. Its loss function combines the bounding box coordinates, the predicted class, and the segmentation mask losses into a joint optimization schema, delivering strong results.

DeepLab models (L.-C. Chen et al., 2018; L. Chen et al., 2018; Chen et al., 2014a; Chen, Papandreou, Schroff, Adam) uses dilated or atrous convolutions, where a dilation rate defines a spacing between the kernel weights for an efficient convolution at configurable scales. For example, a 3 × 3 kernel with a dilation rate of 2 will process the same image area as a 5 × 5 kernel using only 9 parameters instead of 25. This allows enlarging the receptive field with no increase in computational cost. DeepLab V1 (Chen et al., 2014a) first combined atrous convolution to address problems regarding the decreasing resolution with the depth in the network, and fully connected conditional random fields (CRF) as a post-processing step for a class coherent smoothed outcome. DeepLab V2 (Chen et al., 2018) added atrous spatial pyramid pooling, which samples convolutional feature layers with filters at multiple sampling rates, capturing image context at different scales to segment objects robustly. The best DeepLab reached a 79.7% mIoU (mean IoU) score on the 2012 PASCAL VOC challenge. DeepLab V3, proposed by Chen et al. (2017), goes deeper and uses cascaded modules of atrous convolutions increasing with depth, and then concatenates and processes the outcomes using a final convolutional layer to create the segmentation map. In 2018, Chen et al. (2018) proposed a model, called DeepLab V3+, that uses an encoder-decoder architecture with atrous separable convolution, inspired by the xCeption architecture (Fig. 11). They proposed to use DeepLab V3 as the encoder, and a decoder with skip connections at some of the encoder layers. The best DeepLab V3+ obtained an 89.0% mIoU score on the 2012 PASCAL VOC challenge. It is currently one of the most robust models for semantic segmentation, widely used in different applications.

2.5. *Feature extraction*

Extracting features from images is useful for classifying objects and connecting image content to previous consolidated knowledge. For instance, regression networks can identify known biological or structural features in images while encoder-decoder networks can find the most efficient set of features that describe a set of images. Most of the convolutional networks used for image classification Section 2.2 can also be used to solve regression problems for feature extraction. In such cases, the softmax layer is commonly replaced by a fully connected regression layer with linear or sigmoid activations. Several deep learning algorithms were proposed in this regard, obtaining strong results in image regression problems like head pose estimation (Liu et al., 2016) or facial landmark detection (Sun et al., 2013).

In problems where the features' semantic meaning is not essential, auto-encoders are an efficient tool for computing image features. These models are encoder-decoder networks that learn to map input data in the output (Vincent et al., 2010). The encoder compresses the input into a lower-dimensional code, and the decoder reconstructs the output from the code. The code consists of the innermost layer of the network and represents a compact "summary" or "compression" of the input, also called the latent-space representation. Autoencoders are fully unsupervised, which means one does not need to provide any annotation to train the model, and are also inherently lossy, which means that the input data's reconstruction is usually a degraded version of the input (Vincent et al., 2010). In general, Autoencoders are powerful feature extractors because they learn the minimum necessary representation of the input for reconstructing the input data. This information tends to hold the essential description of input data, an excellent characteristic for feature extractors. One of the most popular auto-encoders is the stacked denoising Autoencoder (Vincent et al., 2010), where many Autoencoders are stacked together to perform image denoising. Variational Autoencoders (Kingma and Welling, 2013) propose to force a prior distribution on the latent representation, which allows not only the inspection of latent variables but also realistic sample synthesis. Similarly, adversarial Autoencoders (Makhzani et al., 2015) use an adversarial loss on the latent representation to promote an approximation to a given distribution that can be potentially linked to specific features.

Another very important group of deep neural networks is Generative Adversarial Network. Although none of the reviewed articles evaluated such algorithm, we included a brief description of this neural network, given the rapid and large implementation in other fields of research such as healthcare (Oliveira, 2020), generation of animation models (Vougioukas et al., 2019), and image translation and edition (Park et al., 2019). Generative adversarial networks (GANs), introduced by Goodfellow et al. (2014) uses adversarial training involving two convolutional networks: the generator, responsible for synthesizing the images, and the discriminator that learns if a given image is real or synthesized. In Animal Sciences, a possible application would be on image generation to complement unbalanced training sets and potentially improve prediction quality in real testing sets. Additionally, GANs can be potentially used in animal biomedical image analysis for lesion segmentation (Luc et al., 2016), synthesis of realistic disease evolution scenarios and others.

3. Applications of deep learning in livestock

The increase in herd sizes observed in recent years has created operational difficulties in monitoring animal health and nutrition in livestock farms. The use of deep learning for monitoring animal health and other phenotypes is very recent and is expected to increase over the next years, following the pattern of algorithms development and the expansion of computational resources.

For each discussed task (image classification, object detection, image segmentation and feature extraction), the main deep learning algorithms used for image analyses were reviewed and discussed. It was not our aim to review all possible applications of deep learning for each task, as each of those comprises a whole research field. However, in each section, we presented a selection of emblematic works and aimed to enable the reader interested in computer vision systems and deep learning to navigate through the different algorithms and link them to potential applications in livestock systems. Additionally, we connected the deep learning strategy used in each computer vision task with the main biological problem related to the phenotypes of interest for livestock animals.

3.1. *Search criteria and overview of deep learning-based computer vision systems articles*

We performed a systematic review of peer-reviewed articles in which deep learning algorithms were primarily used for image analyses of livestock animals. Literature was reviewed following Moher et al. (2009) guidelines for Systematic Reviews and Meta-Analyses (Preferred Reporting Items for Systematic Reviews and Meta-Analyses - PRISMA). The PRISMA flow diagram is shown in Fig. 12, along with the number of retained publications or retrieved at each stage. The eligibility criteria adopted in the current review were: (1) only peer-reviewed publications written in English; (2) research studies focused on dairy and beef cattle, swine, goat and poultry; (3) only studies on computer vision (machine vision) that used deep learning algorithms as the primary approach for image analyses.

The terms computer vision, deep learning, animal, livestock, cattle, dairy, beef, swine, pigs, poultry, and broiler, as well as their random combinations, were used to search the articles in the Web of Science platform in October 2020. In addition, we searched for extra literature using Google and Google Scholar to find additional papers ($n = 9$). All retrieved titles ($n = 118$) were recorded, and non peer-reviewed articles ($n = 13$) and duplicates ($n = 4$) were removed based on source, author name, year, and article title, reducing the number of retrieved articles to 101. The focus of this article was to review the published literature on computer vision and deep leaning in livestock animals; as such, we removed: (1) publications unrelated to livestock animals that were selected even using the searched terms ($n = 28$); (2) publications that did not include images and/or deep learning algorithms ($n = 23$); (3) review articles ($n = 3$); and (4) articles not written in English ($n = 3$). After implementing all the mentioned criteria, we reduced the number of retrieved articles to 44 as shown in the PRISMA flow diagram Fig. 12.

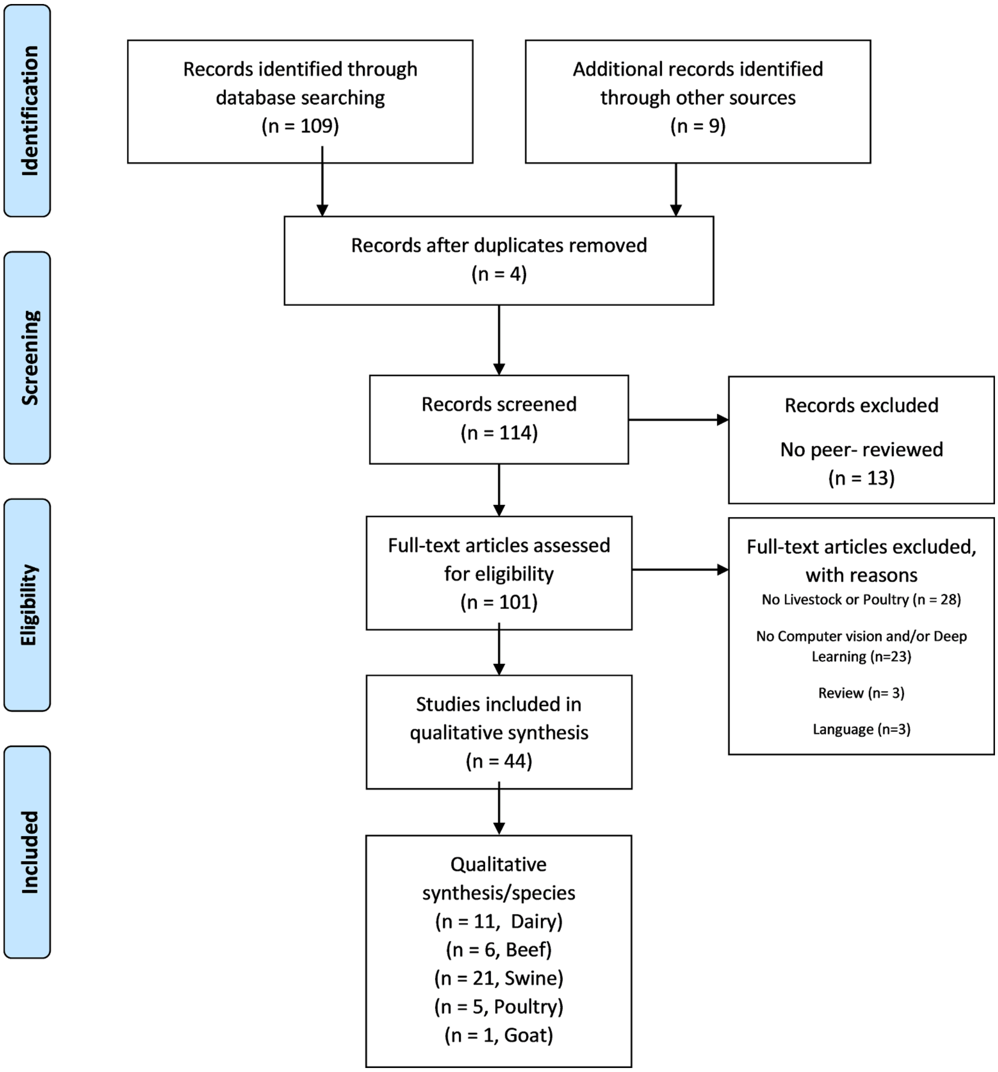

Identification
Screening
Eligibility
Included
Records identified through database searching (n = 109)
Additional records identified through other sources (n = 9)
Records after duplicates removed (n = 4)
Records screened (n = 114)
Records excluded
No peer- reviewed (n = 13)
Full-text articles assessed for eligibility (n = 101)
Full-text articles excluded, with reasons
No Livestock or Poultry (n = 28)
No Computer vision and/or Deep Learning (n=23)
Review (n= 3)
Language (n=3)
Studies included in qualitative synthesis (n = 44)
Qualitative synthesis/species
(n = 11, Dairy)
(n = 6, Beef)
(n = 21, Swine)
(n = 5, Poultry)
(n = 1, Goat)

Fig. 12. Search flow used in the review.

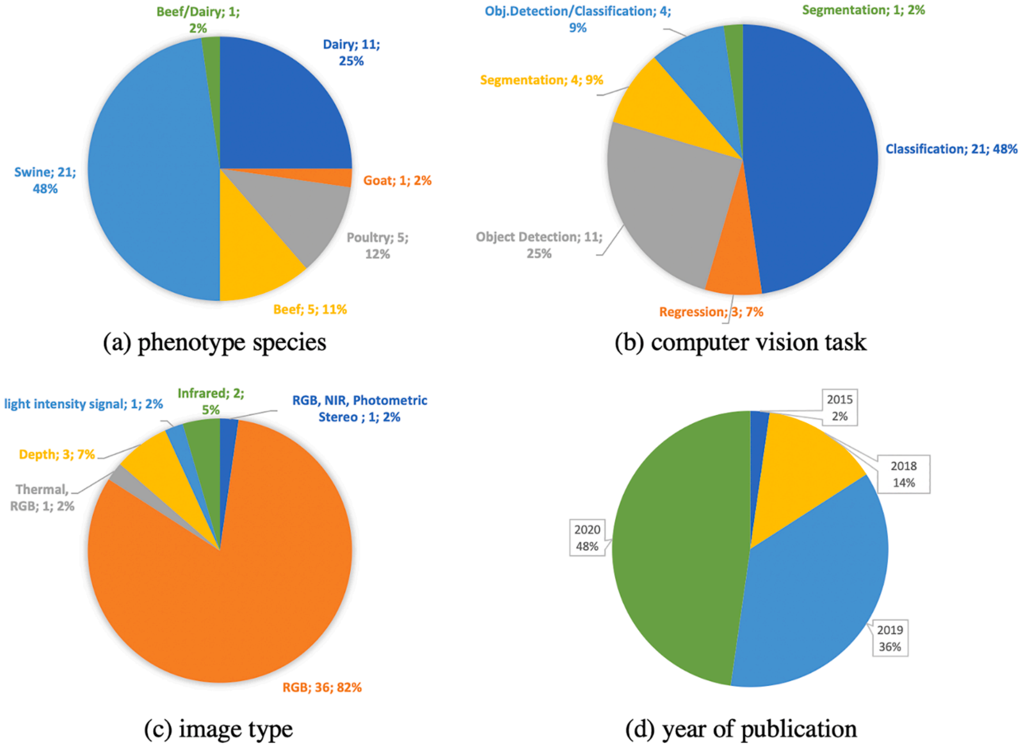
The mentioned review papers (Fernandes et al., 2020a; Wurtz et al., 2019; Nasirahmadi, Edwards, Sturm) provided an extensive overview of computer vision systems applied to processing animal phenotypes but did not focus on the deep learning strategies. Most of the published studies used traditional image analysis and processing, as shown in the 153 articles review from Wurtz et al. (2019). In this review article, we filtered the selected articles ($n = 44$) based on the following computer vision tasks: image classification, object detection, object segmentation, and feature extraction. Such filter was essential to reveal the most used computer vision tasks in Animal Science studies and hence to determine which algorithms should be reviewed within each task.

3.2. *Applying image classification in animal science*

From the total reviewed articles, 48% implemented deep learning algorithms designed to perform image classification. 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%) (Fig. 13(a)). It is important to highlight that 7% of the articles used deep neural network architecture for image classification to accommodate a regression problem. 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. Depth and infrared images were very little studied with 7% (3 articles) and 5% (2 articles), respectively (Fig. 13(c)).

The oldest peer-reviewed article retrieved in our list after applying the selection criteria (Fig. 12) 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 ($n = 6$) (Fig. 13(d)). These results can confirm the rapid and recent interest in deep learning as the primary algorithm for analyzing images in computer vision systems. Most of the articles were published in the Journal of 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 ($n = 1$) (Fig. 13 (a)). 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 ($n = 4$), weight/body condition score assessment ($n = 3$), and others ($n = 5$) (Fig. 13(b)).

In computer vision, image classification was the most frequent task ($n = 27$ among the 44 articles reviewed) found in Animal Science studies



(a) phenotype species

(b) computer vision task

(c) image type

(d) year of publication

Fig. 13. Charts with descriptive information about the article corpus gathered in our search about computer vision applied to livestock species.

(Table 2). Image classification has been used for different livestock applications including animal identification (Bezen et al., 2020; Marsot et al., 2020), posture (Riekert et al., 2020), and behavior (Chen et al., 2020a–d). Animal identification is a crucial step for a full implementation of an automated computer vision system in livestock operations. Additionally, accurate individual animal identification through image analyses could open new avenues for animal traceability programs, which is extremely important given the growing demand for food safety by final consumers around the world. If accurate through the entire animal life cycle, such image-based predictions would allow a greater level of data security and reduce the chance of fraud throughout the supply chain. Several articles proposed the use of deep learning for animal identification in cattle (Bezen et al., 2020; Kumar et al., 2018; Qiao et al., 2019a; Zhao and He, 2015) and swine (Hansen et al., 2018; Marsot et al., 2020; Zheng et al., 2015). Some of the deep learning algorithms described in Section 2.2 were reported in livestock works, including Inception-v3 (Chen et al., 2020d), ResNet (ResNet50; Chen et al., 2020c), Xception (Chen et al., 2020b), NASNet (Barbedo et al., 2017), and others such as LeNet-5 (Zhao and He, 2015). Those authors reported accuracy values ranging from 83.8 to 99.0% for animal recognition (Table 2). It is important to note that some studies did not consider individual animal classification as a multiclass problem, but rather as a binary classification problem (presence of an animal in the image). Only RGB images of animals or pictures of digital numbers presented in collar tags Bezen et al. (2020) or parts of animals (Kumar et al., 2018; Marsot et al., 2020) have been used for animal recognition. Despite the high accuracy reported in the reviewed papers, the identification of individual animals of single-color breeds (e.g. Angus, Jersey, and Landrace) based only in RGB images can be a challenge. In this context, more research evaluating different algorithms that use 3D representations as input should be considered in the future.

3.3. *High-throughput phenotyping using deep learning-based computer vision systems*

Deep learning-based computer vision systems have been used for high-throughput phenotyping as a strategy to collect animal behaviors in cattle, swine, broiler, and goats (Tables 2 and 3). In 2018, studies published by Zheng et al. (2018) and Yang et al. (2018) proposed the use of deep learning algorithms to identify pig postures (drinking, urination, and mounting) and individual ID, and to recognize feeding behavior. Recently, Chen et al. (2020a, 2020b, 2020c, 2020d) implemented different CNN architectures (VGG16, ResNet50, Inception-v3, and Xception) for handling pig behavior, combining them with a Recurrent Neural Network (RNN) called Long Short-Term Memory (LSTM). LSTM (Hochreiter and Schmidhuber, 1997) networks are usually used to process time-dependent sequences of data, as found in speech and video data, even though inputs other than time series can be used in such neural networks. To combine such algorithms, the spatial features extracted from the CNN are used as inputs in the LSTM, which then extracts spatial-temporal features. Using this combination (CNN + LSTM), the authors found high accuracies to predict aggressive episodes of pigs (Acc = 97.2%), drinking episodes (Acc = 92.5%), feeding behavior (Acc = 98.4%) and the recognition of pigs engaging with different enrichment objects (Acc = 95.6–97.6%). Chen et al. (2020d) compared different deep learning architectures (VGG16, ResNet50, Inception-v3, and Xception) combined with LSTM to predict pig feeding behavior and they concluded that the combination Xception + LSTM presented the most accurate predictions. Although most of the published articles focused on animal identification and behavior, other important phenotypes were investigated, as reported by Atkinson et al. (2020), who demonstrated the potential of Resnet-101 to predict large particle content in dairy cattle feces (90% detection rate, using NIR and 3D images) and Huang et al. (2019) who reported high predictive accuracies (98.5 and 99.1%) to classify dairy cow body condition score using Single Shot MultiBox Detector (SSD) algorithms in RGB images.

Some of the reviewed articles (Bezen et al., 2020; Cang et al., 2019; Fernandes et al., 2020b; Tian et al., 2019) used deep neural networks to solve a regression problem. Dairy cow feed intake, and pig body weight, muscle depth, and backfat thickness were the phenotypes predicted through regression-based deep learning algorithms. Through the modification of the last activation function, the deep neural network outputs

Table 2
Classification and regression in livestock.

Paper	Type	Phenotype	Algorithm	Backbone	Score
Atkinson et al. (2020)	Dairy	Grain/fiber detection in feces	Resnet-101		Acc = 90%
Barbedo et al. (2017)	Beef	Animal detection in pastures	CNN	15 different architectures (best NASNet)	Acc = 99.2–96.4%
Chen et al. (2020c)	Swine	Drinking behavior	ResNet50 LSTM		Acc = 92.5%
Chen et al. (2020d)	Swine	Feed behavoir	Inception-v3 LSTM		Acc = 95.2–97.9%
Chen et al. (2020b)	Swine	Action recognition (agressive episodes)	xCeption LSTM		Acc = 98.4%
Chen et al. (2020a)	Swine	Action recognition (engagement with objects)	VGG-16 LSTM		Acc = 98.2%
de Freitas et al. (2019)	Beef	Kerato conjunctivitis identification	CNN		Acc = 79.57%
Hansen et al. (2018)	Swine	Face recognition	CNN		Acc = 96.7%
Kumar et al. (2018)	Beef	Animal identification	CNN; SDAE; RBM		Acc = 95.98%; 96.92%; 98.99%
Li et al. (2019)	Beef Dairy	Cattle pose estimation	CPMs; Stacked hourglass; Heatmap regression		PCKh0.5 = 88.29%; 90.39%; 83.52%
Marsot et al. (2020)	Swine	Face recognition	CNN		Acc = 83.75%
McKenna et al. (2020)	Swine	Postmortem liver and heart pathologies Inspection	CNN	AlexNet	AUC = 0.89–0.96
Qiao et al., 2019	Beef	Animal Recoginition	CNN LSTM	Inception-v3	Acc = 78-91%
Riekert et al. (2020)	Swine	Position and posture of pigs	Faster RCNN	NASNet	Positions 67.7% AP, Position and posture detection mAP of 80.2%
Wang et al. (2018)	Swine	Animal recognition	CNN	Inception-v3 + Xception + DPNs131	Acc = 96.41%
Ye et al. (2020)	Poultry	Stunned condition	FasterRCNN+MRMnet (Multi-Layer Residual Module)		Acc = 98.06%
Zhao and He (2015)	Dairy	Animal recognition	CNNs	LeNet-5	Acc = 90.55–93.33%
Bezen et al. (2020)	Dairy	Feed Intake	CNN	Resnet50	AE = 0.241 kg / MSE = 0.106 kg2
Cang et al. (2019)	Swine	Body weight	Faster R-CNN + VGG		AE = 0.644 kg / ARE = 0.374%
Fernandes et al., 2020b	Swine	Body composition traits	MLP		MASE = 2.69–13.56; R2 = 0.45–0.86
Tian et al. (2019)	Swine	Animal counting	Counting CNN + ResNeXt		MAE = 1.67; RMSE = 2.13
Geng et al. (2019)	Poultry	Hatching egg activity	FCNs GRUs		Acc = 99.69%
Fang et al. (2020)	Poultry	Animal tracking	CNN	AlexNet	OR = 0.758; MTPE = 0.730
Zhuang and Zhang (2019)	Poultry	Sick or health classification	SSD	Inception-v3	mAP = 48.1%-99.7%
Huang et al. (2019)	Dairy	Body condition score	Improved SSD; SSD; YOLO-v3	DenseNet and Inception-v4; VGG-16	Acc = 88.84–99.10%
Yang et al. (2018)	Swine	Recognition of nursing interactions	FCN + SVM	VGG16	Acc = 97.6%
Zhang et al. (2019)	Swine	Behavior recognation	MobileNet and SDD		Precision = 91.4–96.5%

the predicted values as a continuous numerical value instead of class probabilities.

3.4. *Applying object detection in animal science*

Tasks related to object detection were implemented for multiples purposes in studies involving livestock animals. As previously described in the section Object Detection - Deep Learning Algorithms, the main goal of this task is to detect one or more objects in an image. In Animal Science studies, the algorithms designed for object detection were mostly used for animal detection (Cowton et al., 2019; Lee et al., 2019; Psota et al., 2019; Seo et al., 2020), mainly swine. Other applications were found, such as the detection of lameness (Kang et al., 2020) and digital dermatitis (Cernek et al., 2020) in dairy cattle (Tables 2 and 3). Cernek et al. (2020) implemented YOLOv2 in RGB images and found an accuracy of 88%. Such results indicate great potential for computer vision systems to identify cows with digital dermatitis, reducing dermatitis prevalence and improving animal welfare. Kang et al. (2020) developed a lameness scoring system for dairy cows using the Receptive Field Block Net Single Shot Detector deep learning network to locate cow hooves in the video with 87.0% of mean average precision. The located legs were then used as input for an algorithm proposed to calculate the supporting phase, which is the difference between the hoof lifting time and the hoof load time.

Articles using deep neural networks in combination with other algorithms were found among the reviewed papers. In some articles, deep learning was used to remove background or detect an object as a way to remove unnecessary noise for posterior analyses. For example, Lee et al. (2019) proposed a hybrid method composed by an image processing step, followed by a deep learning algorithm. They used Gaussian Mixture Model to detect the moving frame for 24 h, TinyYOLOv3 to detect individual pigs in each selected frame, and lastly they segmented the pig body implementing Otsus method, which performed automatic image thresholding to calculate pig size. Instead of using end-to-end deep learning methods, Lee et al. (2019) proposed these hybrid models and reported greater analysis speed and adequate accuracy when implemented in single board GPUs, such as Jeston TX2, and compared with deep learning strategies like Mask R-CNN.

YOLO and Faster R-CNN were the main algorithms used for object detection in Animal Science studies. The deep learning architecture

Table 3
Object detection in livestock.

Paper	Type	Phenotype	Algorithm	Backbone	Score
Bezen et al. (2020)	Dairy	Tag number identification	Faster RCNN	Resnet	Acc = 93.6%
Zheng et al. (2018)	Swine	Lactating postures	Faster RCNN	ZFNet	Acc = 93%
Cowton et al. (2019)	Swine	Animal detection	Faster R-CNN		mAP = 0.901
Geffen et al. (2020)	Poultry	Counting hens/cage	Faster R-CNN	ResNet101	Acc = 89.6%
Guzhva et al. (2018)	Dairy	Animal tracking	CNN	VGG	225 average tracking time per starting point
Jiang et al. (2019)	Dairy	Key parts of dairy cows	FLYOLOv3		Precision = 99.18%; Recall = 97.51%
Kang et al. (2020)	Dairy	Lameness scoring	RFB_Net_SSD	VGG16	Acc = 93-96%
Lee et al. (2019)	Swine	Detection of undergrown pigs	TinyYOLO	Tiny Darknet	NA
Nasirahmadi et al. (2017)	Swine	Behavior	Faster R-CNN	Inception-v2, ResNet50, ResNet101 and Inception-ResNet-v2	AP = 0.92–0.95
Psota et al. (2019)	Swine	Animal detection	Hourglass networks		Precision = 0.91–1.0; Recall = 0.66–0.96
Seo et al. (2020)	Swine	Animal detection	TinyYOLO	Tiny Darknet	Acc = 97.6%
Tsai et al. (2020)	Dairy	Drinking behavior	Tiny YOLOv3	Tiny Darknet	F1 score = 0.987
Cernek et al. (2020)	Dairy	Digital dermatitis scoring	YOLO		Acc = 71-88%; Cohen's kappa = 0.36–0.51
Jiang et al. (2020)	Goat	Behavior recognation	Faster R-CNN, YOLOv3, YOLOv4		Acc = 90.46–98.27%

varied across studies, with ResNet, Xcpetion, VGG16, Inception, and Darknet featuring among the most used. Barbedo et al. (2017) used an UAV to obtain aerial images of beef farms and tested 15 CNN architectures to detect animals. These authors concluded that many CNN architectures were robust enough to detect cattle on farm, and they highlighted great performance of NASNet and Xception network. Geffen et al. (2020) used Faster R-CNN, with ResNet-101 as the backbone network, to detect and count hens per cage with a detection accuracy of 89.6%. Ye et al. (2020) using R-CNN, reported 98.6% of accuracy to predict stunned condition in broilers. It is important to highlight that few articles with poultry were found and retrieved in this review.

3.5. *Applying object segmentation in animal science*

Object segmentation was the computer vision task with the fewest number of published studies in Animal Science, as observed in Table 4. Four publications with object segmentation were selected and retrieved in this review, 2 of them using Mask R-CNN, 1 using U-Net, and 1 using DeepLab V3+. Qiao et al., 2019 used a Mask R-CNN based cattle instance segmentation and contour line extraction method. The proposed approach resulted in accurate cattle segmentation, with Mean Pixel Accuracy (MPA) of 0.92, and achieved contour extraction with an Average Distance Error (ADE) of 33.5 pixels, overperforming SharpMask and DeepMask instance segmentation methods. Wu et al. (2020) used a DeepLab V3+ semantic segmentation model to perform cow body segmentation. Subsequently, a phase-based video magnification algorithm was applied in processed images to amplify the weak breathing movements, and the Lucas–Kanade optical flow algorithm was used to detect breathing direction. In addition, a respiration rate detection model was used to detect the respiratory rate of dairy cows.

In this review, we identified several potential applications of computer vision system based on deep learning algorithms to support livestock operations. For instance, we did not find articles in Animal Science using Autoencoders for feature extraction. However, we believe that such an analytical tool has a great potential to be implemented in livestock image data to reduce data dimensionality and select important features for phenotype prediction. We observed that most articles using deep learning for image classification implemented standard CNN models, but papers related to object detection are already applying competitive models such as ResNet, Inception, and Xception, and have reported good results. However, further development for more complex tasks related to detection, like modeling behavior, will need to embed more complex models from related areas like object tracking. Image segmentation seems to be the least explored field for livestock applications among the ones evaluated in our review. The papers identified for segmenting livestock images used Mask R-CNN, U-Net, and Deeplab models, but some other state-of-the-art models for semantic segmentation, such as PSPNet, were not yet considered, and we believe such algorithms could help improve results in real application settings (e.g. in commercial farms).

4. Concluding remarks

The recent adoption of deep learning algorithms in computer vision systems designed for livestock applications and its great predictive ability reported in the published studies have demonstrated the potential of such analytical approach for high-throughput phenotyping. Such computer vision systems can bring real-time, non-invasive, and precise predictions related to health, welfare, nutrition, and reproduction at group and animal level.

The number of studies using deep learning as the main framework for image analyses in computer vision systems is still small and very recent in Animal Sciences, given that 84% of the publications with livestock animals were published between 2019 and 2020. However, it is

Table 4
Image segmentation in livestock.

Paper	Type	Phenotype	Algorithm	Backbone	Score
Nye et al. (2020)	Dairy	Heritability of conformational traits	Mask R-CNN/ CNN*	Resnet	Proportion of White color: R2 = 0.926; Heritabilities: h2 = 0.180.82
Bruenger et al. (2020)	Swine	Posture estimation	Unet	Inception-ResNet-v2, ResNet34, EfficientNet	F1-score = 95%
Qiao et al., 2019	Beef	Cattle segmentation	Mask R-CNN	Resnet	Acc = 92%; Average Distance Error = 53.05
Wu et al. (2020)	Dairy	Animal detection	DeepLab V3+	ResNet-101	Acc = 99.40%; IoU = 0.987

important to point out that most of the proposed deep learning algorithms used in the reviewed studies were released very recently. For example, xCeption (Chollet, 2016) and Mask R-CNN (He et al., 2017) were published in 2016 and 2017, respectively. Thus, it is expected to observe such a delay between the development of deep learning algorithms and their wide implementation in other research areas, such as agriculture and livestock.

The great adoption of deep learning algorithms for image classification tasks compared to object detection and image segmentation is expected when computer vision starts to be implemented in a new field. Usually, object segmentation has a posterior implementation compared to image classification, as more complex problems cannot be solved by image classification alone, and thus more refined image analyses are necessary. For example, a deep learning algorithm trained exclusively to assert if there is a cow in an image will not predict exact locations, distances among cows, body areas for each animal, etc. As such, moving from image classification to instance segmentation is an expected step as more complex problems are present. Another limiting factor for the quick implementation of deep learning algorithms for object segmentation is image labeling and annotation, which is more labor-intensive and costly compared to image classification.

Few reviewed articles provided detailed information regarding the use of pre-trained weights. Although several articles used a well-known deep learning architecture, such as ResNet, Inception, and Xception, very few reported if the network was entirely re-trained or if some layers were frozen and some re-trained. The use of pre-trained weights can be an effective strategy to speed up the training process and improve network convergence, especially when given small datasets. Reporting such information is important to ensure research reproducibility and critical evaluation. As computer vision advances in a new field, such as livestock, more customized image analyses are required, and the full benefit of using pre-trained networks is reduced. In this regard, the need for large datasets for specific applications (e.g. agricultural datasets) would become very important.

Most of the articles used Holdout as the data split strategy used to create training and testing sets. In general, very little attention has been put into the hierarchical structure present in the dataset. The paper published by Psota et al. (2019) demonstrated the importance of validating the trained algorithm in an unseen group of images from new environments, where the light conditions, background, and time were different from images in the training set. Psota et al. (2019) reported a reduction in precision and recall from 100% and 96% (randomly data split - images collected in the same environment) to 91% and 67% (unseen images: new environment). Fernandes et al., 2020b also reported an increase in root mean squared error (from 4.74 to 6.34 kg) of body weight prediction when the deep learning algorithm was trained using finishing pigs from two commercial lines and validated using another line, compared to k-fold cross-validation strategies. The studies from Psota et al. (2019) and Fernandes et al., 2020b demonstrated that hierarchical data structure present in the image dataset could produce over-optimistic predictions if not carefully considered during algorithms validation. This is a critical factor in developing trustworthy technologies and avoiding frustration by end-users in the livestock industry.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to acknowledge the Dairy Innovation Hub for providing the funding for D. A. B. Oliveira, and also thank the financial support from the USDA National Institute of Food and Agriculture (Washington, DC; grant 2020-67015-30831) and Hatch project (WIS03085).

Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.livsci.2021.104700

References

Alameer, A., 2020. Automated recognition of postures and drinking behaviour for the detection of compromised health in pigs. 10.25405/data.ncl.13042619.v1.

Andrew, W., Burghardt, T., Campbell, N., Gao, J., 2020. Opencows2020. 10.5523/bris.10m32xl88x2b61zlkkgz3fml17.

Andrew, W., Greatwood, C., Burghardt, T., 2017. Visual localisation and individual identification of Holstein Friesian cattle via deep learning. 2017 IEEE International Conference on Computer Vision Workshops (ICCVW), pp. 2850–2859. https://doi.org/10.1109/ICCVW.2017.336.

Andrew, W., Greatwood, C., Burghardt, T., 2019. Aerial animal biometrics: individual Friesian cattle recovery and visual identification via an autonomous UAV with onboard deep inference. 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 237–243. https://doi.org/10.1109/IROS40897.2019.8968555.

Andrew, W., Hannuna, S., Campbell, N., Burghardt, T., 2016. Automatic individual Holstein Friesian cattle identification via selective local coat pattern matching in RGB-D imagery. 2016 IEEE International Conference on Image Processing (ICIP), pp. 484–488. https://doi.org/10.1109/ICIP.2016.7532404.

Atkinson, G.A., Smith, L.N., Smith, M.L., Reynolds, C.K., Humphries, D.J., Moorby, J.M., Leemans, D.K., Kingston-Smith, A.H., 2020. A computer vision approach to improving cattle digestive health by the monitoring of faecal samples. Sci. Rep. 10 (1), 17557. https://doi.org/10.1038/s41598-020-74511-0.

Badrinarayanan, V., Kendall, A., Cipolla, R., 2017. Segnet: a deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 39 (12), 2481–2495.

Barbedo, J.G.A., Gomes, C.C.G., Cardoso, F.F., Domingues, R., Ramos, J.V., McManus, C. M., 2017. The use of infrared images to detect ticks in cattle and proposal of an algorithm for quantifying the infestation. Vet. Parasitol. 235, 106–112. https://doi.org/10.1016/j.vetpar.2017.01.020.

Benitez Pereira, L. S., Koskela, O., Plnen, I., Kunttu, I., 2020. Data set of labeled scenes in a barn in front of automatic milking system. 10.5281/zenodo.3981400.

Bezen, R., Edan, Y., Halachmi, I., 2020. Computer vision system for measuring individual cow feed intake using RGB-D camera and deep learning algorithms. Comput. Electron. Agric. 172, 105345. https://doi.org/10.1016/j.compag.2020.105345.

Bhole, A., Falzon, O., Biehl, M., Azzopardi, G., 2021. Holstein Cattle Recognition. 10.34894/O1ZBSA.

Bresolin, T., Dorea, J.R.R., 2020. Infrared spectrometry as a high-throughput phenotyping technology to predict complex traits in livestock systems. Front. Genet. 11, 923. https://doi.org/10.3389/fgene.2020.00923.

Bruenger, J., Gentz, M., Traulsen, I., Koch, R., 2020. Panoptic segmentation of individual pigs for posture recognition. Sensors 20 (13). https://doi.org/10.3390/s20133710.

Cai, Z., Vasconcelos, N., 2018. Cascade R-CNN: delving into high quality object detection. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 6154–6162.

Cang, Y., He, H., Qiao, Y., 2019. An intelligent pig weights estimate method based on deep learning in sow stall environments. IEEE Access. https://doi.org/10.1109/ACCESS.2019.2953099.

Cernek, P., Bollig, N., Anklam, K., rte Dã, 2020. Hot topic: detecting digital dermatitis with computer vision. J. Dairy Sci. https://doi.org/10.3168/jds.2019-17478.

Chen, C., Zhu, W., Oczak, M., Maschat, K., Baumgartner, J., Larsen, M.L.V., Norton, T., 2020. A computer vision approach for recognition of the engagement of pigs with different enrichment objects. Comput. Electron. Agric. https://doi.org/10.1016/j.compag.2020.105580.

Chen, C., Zhu, W., Steibel, J., Siegford, J., Han, J., Norton, T., 2020. Classification of drinking and drinker-playing in pigs by a video-based deep learning method. Biosyst. Eng. https://doi.org/10.1016/j.biosystemseng.2020.05.010.

Chen, C., Zhu, W., Steibel, J., Siegford, J., Han, J., Norton, T., 2020. Recognition of feeding behaviour of pigs and determination of feeding time of each pig by a video-based deep learning method. Comput. Electron. Agric. https://doi.org/10.1016/j.compag.2020.105642.

Chen, C., Zhu, W., Steibel, J., Siegford, J., Wurtz, K., Han, J., Norton, T., 2020. Recognition of aggressive episodes of pigs based on convolutional neural network and long short-term memory. Comput. Electron. Agric. https://doi.org/10.1016/j.compag.2019.105166.

Chen, L., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L., 2018. Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE Trans. Pattern Anal. Mach. Intell. 40 (4), 834–848.

Chen, L.-C., Papandreou, G., Schroff, F., Adam, H., 2017. Rethinking atrous convolution for semantic image segmentation. arXiv:1706.05587.

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (Eds.), Computer Vision – ECCV 2018. Springer International Publishing, Cham, pp. 833–851.

Chollet, F., 2016. Xception: deep learning with depthwise separable convolutions. Cite arXiv:1610.02357.

Cowton, J., Kyriazakis, I., Bacardit, J., 2019. Automated individual pig localisation, tracking and behaviour metric extraction using deep learning. IEEE Access 7, 108049–108060, 10.1109/ACCESS.2019.2933060.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L., 2009. Imagenet: a large-scale hierarchical image database. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp. 248–255.

Everingham, M., Gool, L., Williams, C.K., Winn, J., Zisserman, A., 2010. The pascal visual object classes (VOC) challenge. Int. J. Comput. Vis. 88 (2), 303–-338. https://doi.org/10.1007/s11263-009-0275-4.

Fang, C., Huang, J., Cuan, K., Zhuang, X., Zhang, T., 2020. Comparative study on poultry target tracking algorithms based on a deep regression network. Biosyst. Eng. https://doi.org/10.1016/j.biosystemseng.2019.12.002.

WordNet: An Electronic Lexical Database, Language, Speech, and Communication. In: Fellbaum, C. (Ed.), 1998. MIT Press, Cambridge, MA.

Fernandes, A.F., Dorea, J.R.R., Rosa, G.J.M., 2020a. Image Analysis and Computer Vision Applications in Animal Sciences: An Overview. Front. vet. sci. 11, 2–20. https://doi.org/10.3389/fvets.2020.551269.

de Freitas, D.S., Camargo, S.D.S., Comin, H.B., Domingues, R., Gaspar, E.B., Cardoso, F. F., 2019. Recognition of bovine infectious keratoconjunctivitis using thermographic imaging and convolutional neural networks. Braz. J. Appl. Comput. 11 (3), 133–145. https://doi.org/10.5335/rbca.v11i3.9210.

Gao, J., Burghardt, T., Andrew, W., Dowsey, A. W., Campbell, N. W., 2021. Towards self-supervision for video identification of individual Holstein-Friesian cattle: The cows2021 dataset. arXiv:2105.01938.

Fernandes, A.F., Dórea, J.R., Valente, B.D., Fitzgerald, R., Herring, W., Rosa, G.J., 2020b. Comparison of data analytics strategies in computer vision systems to predict pig body composition traits from 3D images. J. Anim. Sci. 98 (8), 1–9. https://doi.org/10.1093/jas/skaa250.

Geffen, O., Yitzhaky, Y., Barchilon, N., Druyan, S., Halachmi, I., 2020. A machine vision system to detect and count laying hens in battery cages. Animal 1–7. https://doi.org/10.1017/S1751731120001676.

Geng, L., Wang, H., Xiao, Z., Zhang, F., Wu, J., Liu, Y., 2019. Fully convolutional network with gated recurrent unit for hatching egg activity classification. IEEE Access. https://doi.org/10.1109/ACCESS.2019.2925508.

Girshick, R., 2015. Fast r-cnn. Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV). IEEE Computer Society, USA, pp. 1440–-1448. https://doi.org/10.1109/ICCV.2015.169.

Girshick, R., Donahue, J., Darrell, T., Malik, J., 2016. Region-based convolutional networks for accurate object detection and segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 38 (1), 142–-158. https://doi.org/10.1109/TPAMI.2015.2437384.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press.http://www.deeplearningbook.org

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative Adversarial Nets. In: Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N.D., Weinberger, K.Q. (Eds.), Advances in Neural Information Processing Systems, 27. Curran Associates, Inc., pp. 2672–2680

Guzhva, O., Ardö, H., Nilsson, M., Herlin, A., Tufvesson, L., 2018. Now you see me: convolutional neural network based tracker for dairy cows. Front. Robot. AI 5 (SEP), 107. https://doi.org/10.3389/frobt.2018.00107.

Hansen, M.F., Smith, M.L., Smith, L.N., Salter, M.G., Baxter, E.M., Farish, M., Grieve, B., 2018. Towards on-farm pig face recognition using convolutional neural networks. Comput. Ind. 98, 145–152. https://doi.org/10.1016/j.compind.2018.02.016.

He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask R-CNN. 2017 IEEE International Conference on Computer Vision (ICCV), pp. 2980–2988.

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9 (8), 1735–1780.

He, K., Zhang, X., Ren, S., Sun, J., 2015a. Deep residual learning for image recognition. arXiv:1512.03385.

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H., 2017. Mobilenets: efficient convolutional neural networks for mobile vision applications. CoRR abs/1704.04861.

Huang, X., Hu, Z., Wang, X., Yang, X., Zhang, J., Shi, D., 2019. An improved single shot multibox detector method applied in body condition score for dairy cows. Animals. https://doi.org/10.3390/ani9070470.

Jiang, B., Wu, Q., Yin, X., Wu, D., Song, H., He, D., 2019. FLYOLOv3 deep learning for key parts of dairy cow body detection. Comput. Electron. Agric. https://doi.org/10.1016/j.compag.2019.104982.

Jiang, M., Rao, Y., Zhang, J., Shen, Y., 2020. Automatic behavior recognition of group-housed goats using deep learning. Comput. Electron. Agric. https://doi.org/10.1016/j.compag.2020.105706.

Kang, X., Zhang, X., Liu, G., 2020. Accurate detection of lameness in dairy cattle with computer vision: a new and individualized detection strategy based on the analysis of the supporting phase. J. Dairy Sci. 0 (0) https://doi.org/10.3168/jds.2020-18288.

Kingma, D. P., Ba, J., 2017. Adam: a method for stochastic optimization. arXiv:1412.6980.

Kingma, D. P., Welling, M., 2013. Auto-encoding variational Bayes. arXiv:1312.6114.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. In: Pereira, F., Burges, C.J.C., Bottou, L., Weinberger, K.Q. (Eds.), Advances in Neural Information Processing Systems, 25. Curran Associates, Inc., pp. 1097–1105

Kumar, S., Pandey, A., Sai Ram Satwik, K., Kumar, S., Singh, S.K., Singh, A.K., Mohan, A., 2018. Deep learning framework for recognition of cattle using muzzle point image pattern. Measurement 116, 1–17. https://doi.org/10.1016/j.measurement.2017.10.064.

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

Lee, S., Ahn, H., Seo, J., Chung, Y., Park, D., Pan, S., 2019. Practical monitoring of undergrown pigs for IoT-based large-scale smart farm. IEEE Access. https://doi.org/10.1109/ACCESS.2019.2955761.

Li, K., Wan, G., Cheng, G., Meng, L., Han, J., 2020. Object detection in optical remote sensing images: asurvey and a new benchmark. ISPRS J. Photogramm. Remote Sens. 159, 296–-307. https://doi.org/10.1016/j.isprsjprs.2019.11.023.

Li, X., Cai, C., Zhang, R., Ju, L., He, J., 2019. Deep cascaded convolutional models for cattle pose estimation. Comput. Electron. Agric. https://doi.org/10.1016/j.compag.2019.104885.

Lin, G., Shen, C., van den Hengel, A., Reid, I., 2016. Efficient piecewise training of deep structured models for semantic segmentation. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3194–3203.

Lin, M., Chen, Q., Yan, S., 2013. Network in network. arXiv:1312.4400.

Lin, T., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S., 2017. Feature pyramid networks for object detection. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 936–944.

Lin, T., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S., 2017. Feature pyramid networks for object detection. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 936–944.

Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollár, P., 2017. Focal loss for dense object detection. arXiv:1708.02002.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L., 2014. Microsoft coco: common objects in context. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (Eds.), European Conference on Computer Vision (ECCV). Springer International Publishing, Cham, pp. 740–755.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S.E., Fu, C.-Y., Berg, A.C., 2016. SSD: single shot multibox detector. ECCV. Springer, pp. 21–37.

Liu, X., Liang, W., Wang, Y., Li, S., Pei, M., 2016. 3D head pose estimation with convolutional neural network trained on synthetic images. 2016 IEEE International Conference on Image Processing (ICIP), pp. 1289–1293.

Liu, Z., Li, X., Luo, P., Loy, C., Tang, X., 2015. Semantic image segmentation via deep parsing network. 2015 IEEE International Conference on Computer Vision (ICCV), pp. 1377–1385.

Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3431–3440.

Luc, P., Couprie, C., Chintala, S., Verbeek, J., 2016. Semantic segmentation using adversarial networks. arXiv:1611.08408.

Makhzani, A., Shlens, J., Jaitly, N., Goodfellow, I., Frey, B., 2015. Adversarial autoencoders. arXiv:1511.05644.

Marsot, M., Mei, J., Shan, X., Ye, L., Feng, P., Yan, X., Li, C., Zhao, Y., 2020. An adaptive pig face recognition approach using convolutional neural networks. Comput. Electron. Agric. https://doi.org/10.1016/j.compag.2020.105386.

McKenna, S., Amaral, T., Kyriazakis, I., 2020. Automated classification for visual-only postmortem inspection of porcine pathology. IEEE Trans. Autom. Sci. Eng. 17 (2), 1005–1016. https://doi.org/10.1109/TASE.2019.2960106.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med. 6 (7), e1000097. https://doi.org/10.1371/journal.pmed.1000097.

Nasirahmadi, A., Edwards, S. A., Sturm, B., 2017. Implementation of machine vision for detecting behaviour of cattle and pigs. 10.1016/j.livsci.2017.05.014.

Neethirajan, S., 2017. Recent advances in wearable sensors for animal health management. Sens. Bio-Sensing Res. 12, 15–29. https://doi.org/10.1016/j.sbsr.2016.11.004.

Noh, H., Hong, S., Han, B., 2015. Learning deconvolution network for semantic segmentation. 2015 IEEE International Conference on Computer Vision (ICCV), pp. 1520–1528.

Nye, J., Zingaretti, L.M., Pérez-Enciso, M., 2020. Estimating conformational traits in dairy cattle with deepaps: a two-step deep learning automated phenotyping and segmentation approach. Front. Genet. https://doi.org/10.3389/fgene.2020.00513.

Oliveira, D.A.B., 2020. Controllable skin lesion synthesis using texture patches, Bzier curves and conditional GANs. 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), pp. 1798–1802. https://doi.org/10.1109/ISBI45749.2020.9098676.

Park, T., Liu, M.-Y., Wang, T.-C., Zhu, J.-Y., 2019. Semantic image synthesis with spatially-adaptive normalization. arXiv:1903.07291.

Pham, H., Guan, M. Y., Zoph, B., Le, Q. V., Dean, J., 2018. Efficient neural architecture search via parameter sharing. arXiv:1802.03268.

Psota, E.T., Mittek, M., Pérez, L.C., Schmidt, T., Mote, B., 2019. Multi-pig part detection and association with a fully-convolutional network. Sensors (Switzerland). https://doi.org/10.3390/s19040852.

Qiao, Y., Su, D., Kong, H., Sukkarieh, S., Lomax, S., Clark, C., 2019. Individual cattle identification using a deep learning based framework. IFAC-PapersOnLine. Elsevier B.V., pp. 318–323. https://doi.org/10.1016/j.ifacol.2019.12.558

Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2015. You only look once: unified, real-time object detection. arXiv:1506.02640.

Qiao, Y., Truman, M., Sukkarieh, S., 2019. Cattle segmentation and contour extraction based on mask R-CNN for precision livestock farming. Comput. Electron. Agric. 165, 104958. https://doi.org/10.1016/j.compag.2019.104958.

Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster R-CNN: towards real-time object detection with region proposal networks. In: Cortes, C., Lawrence, N.D., Lee, D.D., Sugiyama, M., Garnett, R. (Eds.), Advances in Neural Information Processing Systems 28. Curran Associates, Inc., pp. 91–99

Riekert, M., Klein, A., Adrion, F., Hoffmann, C., Gallmann, E., 2020. Automatically detecting pig position and posture by 2D camera imaging and deep learning. Comput. Electron. Agric. https://doi.org/10.1016/j.compag.2020.105391.

Rodenburg, J., 2017. Robotic milking: technology, farm design, and effects on work flow. J. Dairy Sci. 100 (9), 7729–7738. https://doi.org/10.3168/jds.2016-11715.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Springer International Publishing, Cham, pp. 234–241.

Rutten, C. J., Velthuis, A. G., Steeneveld, W., Hogeveen, H., 2013. Invited review: sensors to support health management on dairy farms. 10.3168/jds.2012-6107.

Schwing, A. G., Urtasun, R., 2015. Fully connected deep structured networks. arXiv:1503.02351.

Seo, J., Ahn, H., Kim, D., Lee, S., Chung, Y., Park, D., 2020. Embeddedpigdet-fast and accurate pig detection for embedded board implementations. Appl. Sci. (Switzerland). https://doi.org/10.3390/APP10082878.

Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., LeCun, Y., 2014. OverFeat: Integrated Recognition, Localization and Detection using Convolutional Networks.

Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556.

Sun, Y., Wang, X., Tang, X., 2013. Deep convolutional network cascade for facial point detection. 2013 IEEE Conference on Computer Vision and Pattern Recognition, pp. 3476–3483.

Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A., 2016. Inception-v4, inception-resnet and the impact of residual connections on learning. arXiv:1602.07261.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2014. Going deeper with convolutions. arXiv:1409.4842.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 2015. Rethinking the inception architecture for computer vision. arXiv:1512.00567.

Tian, M., Guo, H., Chen, H., Wang, Q., Long, C., Ma, Y., 2019. Automated pig counting using deep learning. Comput. Electron. Agric. https://doi.org/10.1016/j.compag.2019.05.049.

Tsai, Y.-C., Hsu, J.-T., Ding, S.-T., Rustia, D.J.A., Lin, T.-T., 2020. Assessment of dairy cow heat stress by monitoring drinking behaviour using an embedded imaging system. Biosyst. Eng. https://doi.org/10.1016/j.biosystemseng.2020.03.013.

Uijlings, J., van de Sande, K., Gevers, T., Smeulders, A., 2013. Selective search for object recognition. Int. J. Comput. Vis. https://doi.org/10.1007/s11263-013-0620-5.

Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., Manzagol, P.-A., 2010. Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion. J. Mach. Learn. Res. 11, 3371—3408.

Vougioukas, K., Petridis, S., Pantic, M., 2019. Realistic speech-driven facial animation with GANs. arXiv:1906.06337.

Wang, J., Liu, A., Xiao, J., 2018. Video-based pig recognition with feature-integrated transfer learning. In: Zhou, J., Wang, Y., Sun, Z., Jia, Z., Feng, J., Shan, S., Ubul, K., Guo, Z. (Eds.), Biometric Recognition. Springer International Publishing, Cham, pp. 620–631.

Wu, D., Yin, X., Jiang, B., Jiang, M., Li, Z., Song, H., 2020. Detection of the respiratory rate of standing cows by combining the DeepLab V3+ semantic segmentation model with the phase-based video magnification algorithm. Biosyst. Eng. 192, 72–89. https://doi.org/10.1016/j.biosystemseng.2020.01.012.

Wurtz, K., Camerlink, I., D'Eath, R.B., Fernández, A.P., Norton, T., Steibel, J., Siegford, J., 2019. Recording behaviour of indoor-housed farm animals automatically using machine vision technology: a systematic review. PLoS One 14 (12), e0226669. https://doi.org/10.1371/journal.pone.0226669.

Yang, A., Huang, H., Zhu, X., Yang, X., Chen, P., Li, S., Xue, Y., 2018. Automatic recognition of sow nursing behaviour using deep learning-based segmentation and spatial and temporal features. Biosyst. Eng. https://doi.org/10.1016/j.biosystemseng.2018.09.011.

Ye, C., Yousaf, K., Qi, C., Liu, C., Chen, K., 2020. Broiler stunned state detection based on an improved fast region based convolutional neural network algorithm. Poult. Sci. https://doi.org/10.3382/ps/pez564.

Zhang, Y., Cai, J., Xiao, D., Li, Z., Xiong, B., 2019. Real-time sow behavior detection based on deep learning. Comput. Electron. Agric. https://doi.org/10.1016/j.compag.2019.104884.

Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J., 2017. Pyramid scene parsing network. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6230–6239.

Zhao, K., He, D., 2015. Recognition of individual dairy cattle based on convolutional neural networks. Trans. Chin. Soc. Agric. Eng. https://doi.org/10.3969/j.issn.1002-6819.2015.05.026.

Zheng, C., Zhu, X., Yang, X., Wang, L., Tu, S., Xue, Y., 2018. Automatic recognition of lactating sow postures from depth images by deep learning detector. Comput. Electron. Agric. https://doi.org/10.1016/j.compag.2018.01.023.

Zheng, S., Jayasumana, S., Romera-Paredes, B., Vineet, V., Su, Z., Du, D., Huang, C., Torr, P.H.S., 2015. Conditional random fields as recurrent neural networks. 2015 IEEE International Conference on Computer Vision (ICCV), pp. 1529–1537.

Zhuang, X., Zhang, T., 2019. Detection of sick broilers by digital image processing and deep learning. Biosyst. Eng. https://doi.org/10.1016/j.biosystemseng.2019.01.003.

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A. L., 2014a. Semantic image segmentation with deep convolutional nets and fully connected CRFs. arXiv: 1412.7062.