agriculture

MDPI

Article

Evaluation of Deep Learning for Automatic Multi-View Face Detection in Cattle

Beibei Xu [1], Wensheng Wang [1,2,3,*], Leifeng Guo [1,3], Guipeng Chen [4], Yaowu Wang [5], Wenju Zhang [1] and Yongfeng Li [1]

[1] Agricultural Information Institute, Chinese Academy of Agriculture Sciences, Beijing 100086, China; xuxiaobei224@163.com (B.X.); guoleifeng@caas.cn (L.G.); zhangwenju@caas.cn (W.Z.); liyongfeng_1116@163.com (Y.L.)
[2] Information Centre, Ministry of Agriculture and Rural Affairs, Beijing 100125, China
[3] Key Laboratory of Agricultural Big Data, Ministry of Agriculture and Rural Affairs, Beijing 100086, China
[4] Agricultural Economics and Information Institute, Jiangxi Academy of Agriculture Sciences, Nanchang 330200, China; chenguipeng1983@163.com
[5] Laboratory of Geo-Information Science and Remote Sensing, Wageningen University, 6708 PB Wageningen, The Netherlands; wangyaowu@caas.cn
* Correspondence: wangwensheng@caas.cn

Abstract: Individual identification plays an important part in disease prevention and control, traceability of meat products, and improvement of agricultural false insurance claims. Automatic and accurate detection of cattle face is prior to individual identification and facial expression recognition based on image analysis technology. This paper evaluated the possibility of the cutting-edge object detection algorithm, RetinaNet, performing multi-view cattle face detection in housing farms with fluctuating illumination, overlapping, and occlusion. Seven different pretrained CNN models (ResNet 50, ResNet 101, ResNet 152, VGG 16, VGG 19, Densenet 121 and Densenet 169) were fine-tuned by transfer learning and re-trained on the dataset in the paper. Experimental results showed that RetinaNet incorporating the ResNet 50 was superior in accuracy and speed through performance evaluation, which yielded an average precision score of 99.8% and an average processing time of 0.0438 s per image. Compared with the typical competing algorithms, the proposed method was preferable for cattle face detection, especially in particularly challenging scenarios. This research work demonstrated the potential of artificial intelligence towards the incorporation of computer vision systems for individual identification and other animal welfare improvements.

Keywords: cattle face detection; RetinaNet; deep learning; precision livestock

Citation: Xu, B.; Wang, W.; Guo, L.; Chen, G.; Wang, Y.; Zhang, W.; Li, Y. Evaluation of Deep Learning for Automatic Multi-View Face Detection in Cattle. *Agriculture* **2021**, *11*, 1062. https://doi.org/10.3390/agriculture11111062

Academic Editors: Gniewko Niedbała and Sebastian Kujawa

Received: 23 September 2021
Accepted: 23 October 2021
Published: 28 October 2021

1. Introduction

Animal husbandry is undergoing a transition from extensive farming to precision livestock farming and welfare breeding. However, the farming facilities and technologies play crucial parts in affecting the economic benefits of large-scale pastures. Inadequate management probably directly damages the health of livestock and is adverse to the food quality and safety, and the development of the livestock industry [1]. Therefore, there is an urgent need for cost-effective technology methods to address these challenges in animal agricultural systems, such as lack of labor and difficulties in real-time monitoring. Precision farming has aroused more interest recently due to the increasing concern over sustainable livestock and production efficiency [1–5]. Precision farming takes advantage of modern information technologies as an enabler of more efficient, productive, and profitable farming enterprises. For example, Internet of Things (IoT) are used for collecting data on the whole lifecycle of livestock, including breeding, slaughtering, meat processing, and marketing; Big Data and Artificial Intelligence (AI) can provide accurate analysis and real-time physical dynamics of each animal species as for a scientific basis for decision-making and analysis of farm managers. Among these, recognition of individual livestock

is an indispensable and significant task in precision livestock management since it has a tremendous breadth of applications. Quick and accurate recognition of individual farm animals is of major importance for illness prevention and control [6], genetic enhancement of varieties [7], quality security of dairy products [8], and reduction of agricultural fake insurance claims [9].

Classical livestock identification techniques, such as ear notching [10], ear tattooing [11], hot iron branding [12], and ear tags [13–16], is subject to equipment loss, duplication, fraud, animal welfare security, monitoring cost, and distance challenges. Instead, based on biometric traits, non-contact identification is a new trend in livestock identification due to its uniqueness, invariance, low cost and easy operation, and high animal welfare. The non-contact identification methods, such as retinal vascular patterns [17,18], iris patterns, and muzzle print patterns [19,20], utilize computer vision and pattern recognition to extract biological features of livestock for individual identification. However, as an individual's most direct external visual information, the difference in facial features allows the livestock's face to be used more extensively to identify the individual. From the perspective of farm practice, compared with the biometric recognition methods noted, face identification is more intuitive and compatible with habits. There is also no need for cooperation of livestock fixed postures. In addition, face identification has great advantages in terms of anti-interference and scalability, which is analogous to human face identification.

Detection of livestock face is often conducted prior to individual identification and tracking in biometric and surveillance systems. Many approaches have been put forward in the literature for animal face detection. Mukai et al. employed the Haar and HOG (Histogram of Oriented Gradients) feature to build the classifiers for pet faces and proved the effectiveness for detecting the cat and dog faces [21]. Local Binary Pattern (LBP) features have been used to extract local texture features from different levels of Gaussian filtered images of cattle faces for face detection [22]. Clark presented pigs' face detection by identifying the features utilizing the Viola–Jones approach for cascade classifiers and basic likelihood functions [23]. Mohammed et al. aimed to detect multi-view faces in cattle with accuracy enhancement using three classifiers and temperature thresholding. Cattle face detection was established in thermal imaging by adopting HOG as a feature and Support Vector Machine (SVM) as a classifier [24]. Akihiko et al. combined face detection with digital cameras to automatically find dogs and cats in the images with acceptable speed performance by integrating edge-based features with multi-layer classifiers [25].

However, the heavy involvement of handcrafted features prevents these approaches from application to complex scenarios in terms of speed and accuracy. The use of the Convolutional Neural Network (CNN) to detect livestock has been demonstrated as successful and promising for further research with regard to variable inputs, processing speed, and accuracy for object detection in images [26]. Alžběta et al. dealt with a reliable dog face detection approach in the images by adopting the two-step technique using the cascade of regressors [27]. The recent advances in deep learning [28–34] have shown their great potential in object detection and classification of thousands of global images due to higher accuracy, precision, and quicker processing speed. Faster R-CNN has been directly used for face detection combined with PANSNet-5 in the cow face recognition framework [35]. Considering the practical scenario of multi-face detection task of livestock cattle identity authentication, Gou et al. improved Faster R-CNN by substituting ZF network for Inception v2 as the basic network [36].

Despite these advances in livestock face detection, the subtle changes in lighting, severe pose variation, false acceptances because of complex background, color similarity between livestock and background, shape deformation, and occlusion present serious challenges to face detection in an actual setting such as a cattle feedlot. Consequently, it is highly necessary to perform a wider assessment of face detection algorithm performance across a range of livestock production settings. The rapid development of object detection with deep learning provides promising techniques for face detection. RetinaNet, a recently proposed powerful object detection framework, which surpasses the detection performance

of cutting-edge, two-stage R-CNN family object detectors and matches the speed of one-stage object detectors, appears to be the most prospective for livestock face detection. In the previous research, RetinaNet was used to explore for detection of road damages [37], automated detection of firearms in cargo X-ray images [38], and the task of indoor assistance navigation for blind and visually impaired persons [39]. Despite the general appeal of RetinaNet, it has not been evaluated in great detail for precision livestock monitoring practices. Given the urgent need to develop technologies that can assist with livestock production and welfare management, it is timely to assess the application of a state-of-the-art machine learning algorithm for precision livestock monitoring. Due to their great significance concerning animal husbandry, cattle were chosen as the case study to explore the performance of RetinaNet-based object detection for multi-view face detection.

2. Related Work

Face detection is a particular application of object detection that accurately finds the target face and its location in images. Object detection is currently a very active research field in computer vision that facilitates high-level tasks such as automatic individual identification and intelligent image recognition. The early object detection methods, including Viola–Jones detectors, HOG detector, and deformable part-based model were built based on handcrafted features, which render the time complexity high and many of the windows redundant [40]. In addition, manually designed features in the traditional object detection are not sufficiently robust to deal with the wide diversity of image changes encountered in practice; thereby, CNN was introduced into the object detection community. Due to its relatively superior performance of learning for robust and high-level feature representations of an image, CNN-based object detection prevents extracting complicated features and their reconstruction process in traditional object detection. Therefore, after R. Girshick et al. took the lead to propose the region-based CNN features for object detection in 2014, the object detection algorithms evolved from R-CNN at an unprecedented speed and have made much progress in recent years. Current state-of-the-art CNN-based object detectors can be grouped into two-stage algorithms and one-stage detection algorithms.

The two-stage detectors start with the extraction of object proposals through selective search or Region Proposal Network (RPN), and then the candidate regions are classified and regressed for precise coordinates. Regression-based algorithms such as Yolo and SSD require the sampling densely at various positions with different aspect ratios first, then provide the direct prediction of object categorization and a bounding box using CNN. Although the end-to-end procedure of the regression-based detectors outperforms the region-based detectors in processing speed, they achieve lower mean average precision because of example imbalance between object and background. As a result, T.-Y. Lin et al. designed a novel one-stage detector called RetinaNet in 2017 to address the class imbalance and increase the importance of hard examples [41]. "Focal loss" was used in RetinaNet to redefine the standard cross-entropy loss, so the training could automatically downweight the simple examples and center more on hard and misclassified examples. Focal loss enables RetinaNet to achieve comparable accuracy of two-stage algorithms and also maintains relatively high processing speed [41].

Considering the aspects of operating speed and accuracy in farming practice, RetinaNet was selected in this paper for further study. For face detection, unlike the human face, consideration should be given to changes in cattle's face and body orientation due to their random roaming. Therefore, this paper will explore the effectiveness of RetinaNet for multi-view cattle face detection. Advancements in deep learning networks present an opportunity to extend the research to the empirical comparisons of the typical CNN backbones for RetinaNet in the task of detecting multi-view cattle face.

3. Materials and Methods

3.1. Overview of the Proposed Framework

Figure 1 shows the overall workflow proposed for processing RGB images that are captured by 2-D cameras to detect multi-view cattle faces based on RetinaNet. The RGB images acquired by 2-D cameras are used as input images after image preprocessing, including image partitioning and image resize. The backbone, including ResNet, VGG, and Densenet, is selected for feature extraction, and then the Feature Pyramid Network (FPN) strengthens the multi-scale features formed in the former convolutional network to obtain more expressive feature maps, which contain a rich and multi-scale feature pyramid. The feature map selects two Fully Convolutional Network (FCN) sub-networks with the same structure but without sharing parameters for cattle face classification prediction and bounding-box prediction. Ground truth was annotated manually for every cattle face in the training sets and then network training was performed after labeling for forming the cattle face detector, followed by the output of multi-view cattle face detection in testing sets.

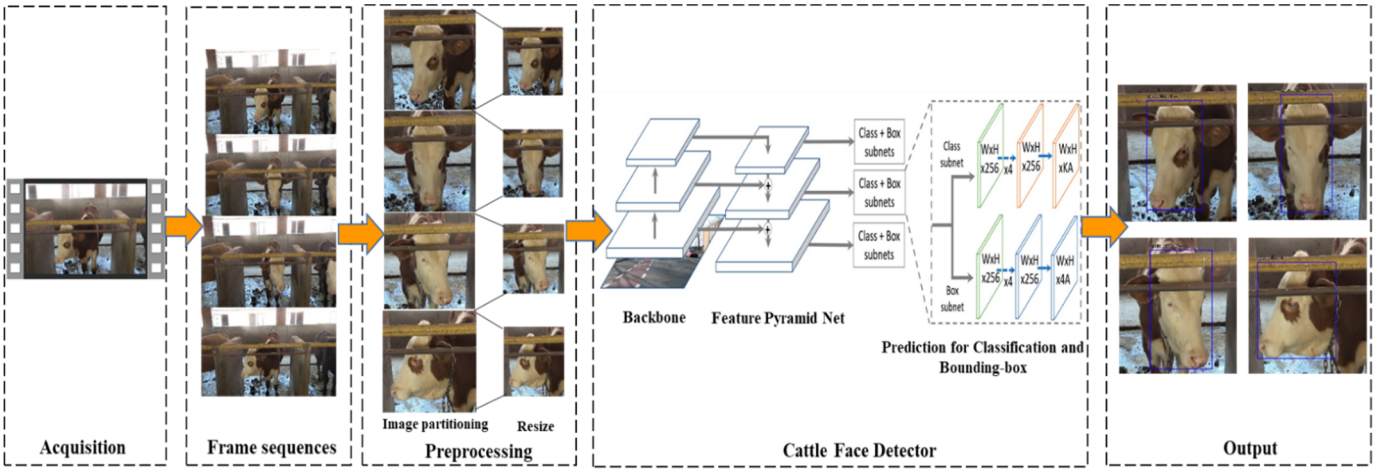


Figure 1. The proposed framework for the detection of multi-view cattle face based on RetinaNet.

3.2. RetinaNet-Based Object Detection

The name of RetinaNet comes from its dense sampling on the input image. RetinaNet is designed to evaluate the proposed focal loss for class imbalance in regression-based algorithms. The framework consists of three parts: (i) the front backbone network for feature extraction, (ii) FPN for constructing the multi-scale feature pyramids, and (iii) two subnetworks for object classification and bounding box regression. Focal loss is a newly high-sufficient loss function that replaces the training with the sampling heuristics and two-stage cascade while dealing with class imbalance. The details for backbones and FCN sub-networks, commonly used in R-CNN-like detectors, are expounded in the original papers, and this section mainly describes FPN and focal loss of the algorithm.

3.2.1. Feature Pyramid Networks

FPN is adopted to strengthen the feature extraction of backbone for weak semantic features using a top-down pyramid and lateral connections (see Figure 2). As indicated in the blue blocks, the bottom-up path is the feed-forward calculation for the main convolutional network, which calculates the feature hierarchy with different proportions. For the feature pyramid, the pyramid level is defined for each stage and the output of the last layer in each stage is chosen as the feature map because the deepest layer of each stage should have the strongest characteristics. Specifically, for the ResNet101 used in the RetinaNet, the outputs of these final residual blocks for conv2_x, conv3_x, conv4_x, and conv5_x are denoted as {C2, C3, C4, C5}. Since conv1 will occupy plenty of memory, it is not included in the pyramid.

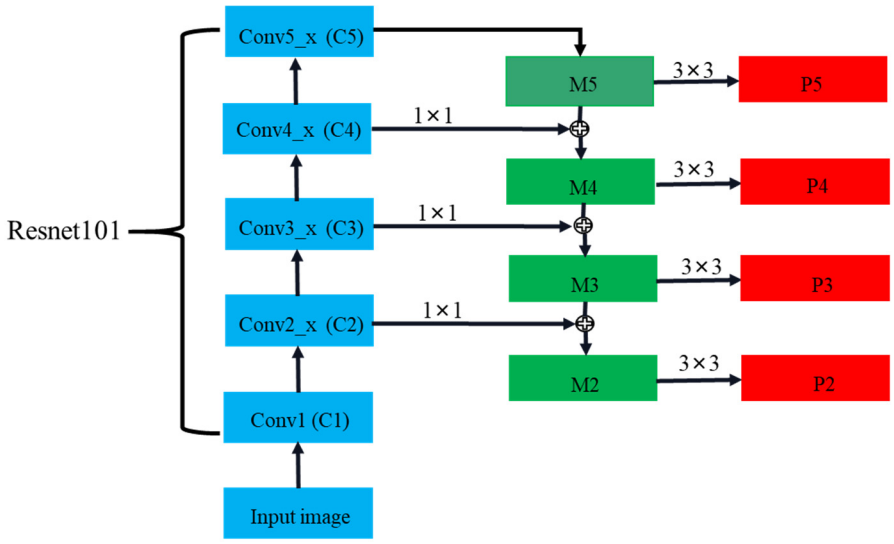


Figure 2. The architecture of FPN.

The top-down flow marked in green obtains high-resolution features by upsampling the feature maps with coarser space but stronger semantics from higher pyramid levels. Later, the bottom-up path is connected laterally to reinforce these features. Specifically, the weak feature map is upsampled twice, and then the upsampling map is merged with the corresponding bottom-up map. This cycle is repeated until the final resolution map is produced. We only need to combine a 1×1 convolutional layer with C5 to produce low-resolution images to run the iteration. Next, we append a 3×3 convolution to perform on each merged image so as to diminish the aliasing effect of upsampling. The same applies to other layers and the final feature map set is called {P2, P3, P4, P5} for object classification and bounding box regression, corresponding to {C2, C3, C4, C5}, respectively.

3.2.2. Focal Loss

The box regression sub-net and classification sub-net in the RetinaNet are implemented using the standard Smooth L_1 loss (Formula (1)) and the Focal loss (Formula (3)), respectively, as the loss functions. Focal loss is a cross-entropy loss that can be dynamically scaled. A weighting factor is added for the traditional cross-entropy function, which can automatically drop the weight of the loss contributed by simple examples and center more on hard samples to solve the class imbalance.

$$SmoothL_1(x) = \begin{cases} 0.5\,x^2 & if\ |x| < 1 \\ |x| - 0.5 & otherwise \end{cases} \tag{1}$$

$$x = f(x_i) - y_i \tag{2}$$

$$FL(P) = -\partial_t (1-P)^{\gamma} \log(P) \tag{3}$$

$$P = \begin{cases} p & if\ z = 1 \\ 1-p & otherwise \end{cases} \tag{4}$$

Here, x is the error value between the estimated value $f(x_i)$ and ground truth y_i; ∂_t and γ are two tunable focusing hypermeters and they function as the role of balancing the ratio between simple and difficult examples; p is the estimated possibility for the given label class. Thus, if the figure of math is 1, it specifies the label class and P is the same as the p in this situation.

3.3. Datasets Preparation and Preprocessing

To address the scarce dataset for cattle face detection and recognition using deep learning, datasets were collected from two housing farms located in Jiangxi Province, China,

and there were 85 healthy scalpers and Simmental ranging in age from 6 to 20 months. The experiment was conducted under various scenes such as different illumination, overlapping, and postures without human intervention, and it took three days to complete this data collection. Examples of multi-view cattle face in different scenes are displayed in Figure 3. This work aims to simulate and facilitate the detection and identification of cattle face by future mobile devices instead of surveillance cameras, and it is common to collect the images where the cattle faces occupy large areas. The cattle were filmed using a Sony FDR-AX 40 camera with MOV video format (3840 × 2160 pixels) at 25 frames per second. The camera on a tripod was fronted straight to the standing cow with a view of 3 cow's face width and 1.5 cow's face length. The original images cropped from videos were in JPG format at 3840 by 2160 pixels. After extracting valuable data frames of every video in MATLAB, the selected images were clipped using MATLAB and then be resized to 224 × 224 pixels. Notably, to ensure the effectiveness of detection performance, during the image selection, different situations of cattle faces for each cow were selected and highly similar faces, especially in consecutive frames, were avoided. The datasets contained a total of 3000 images (1000 negative images included) that were split into training and testing in the proportion 2:1.

Figure 3. Examples of cattle faces in different scenes.

LabelImg is the annotation tool that was used to label the ground truth for cattle faces using RectBox for training datasets. For labeling, the region of every cattle face was selected and annotated using the RectBox in the image. Then, the class label named cattle face needed to be marked on the bubble pop up on the screen. The details of data annotation include object name, box location, and image size, as shown in Figure 4.

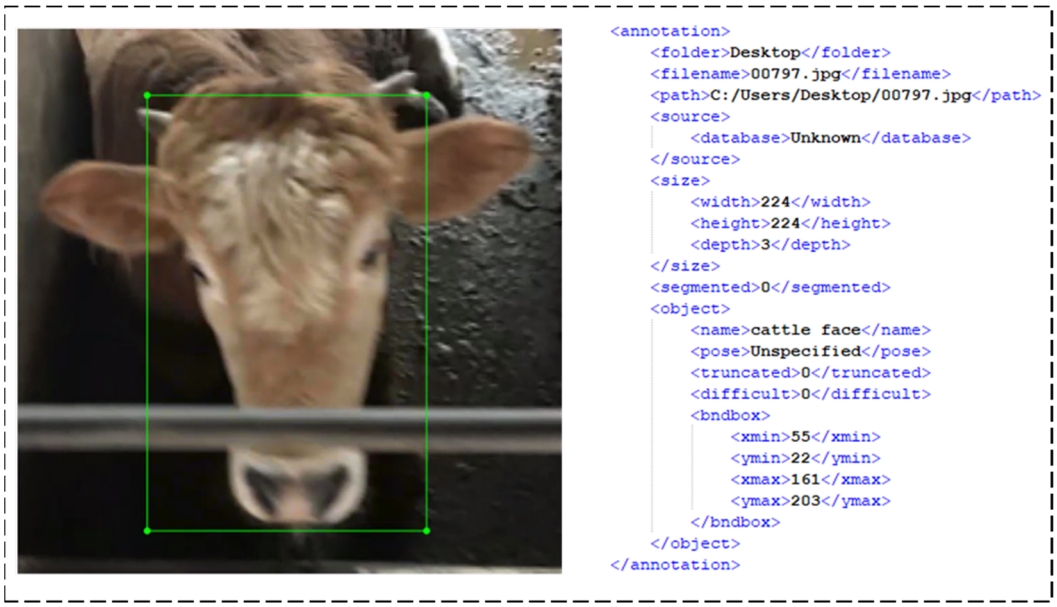


Figure 4. Example of annotation (green) for cattle face and details of a labeled cattle face.

4. Results

4.1. Implementation Details

The experiment was conducted on a desktop computer equipped with Windows 10 64-bit and an NVIDIA GeForce GTX 1080 graphics card. The proposed framework was written employing available libraries including numpy 1.16.5 and scikit-learn 0.21.3 in Python3.6. Keras 2.31 combined with tensorflow-gpu-2.1.0 was installed to provide a deep neural network framework for Python that was compatible with the Python version.

Transfer learning was adopted because of the limited computing resources and datasets for training. Transfer learning was to fine-tune a particular model for the intended task based on existing models. The backbones used in the proposed framework were initialized by ResNet-pretrained model using COCO datasets and VGG-pretrained model using ImageNet datasets and Densenet-pretrained model using ImageNet datasets. All 200,000 training iterations took approximately 17 h, and the best performing epoch for the model was chosen on testing data after the training loss converged. The threshold was set at 0.5 for the Intersection-over-Union (IoU) of confidence and bounding-box in all network models.

4.2. Performance Analysis with Different Backbones

As referred in Section 3.1., the original ResNet 50 backbone model of RetinaNet can be replaced with ResNet 101, ResNet 152, VGG 16, VGG 19, Densenet 121, and Densenet 169. The experiment compared the RetinaNet with ResNet 50 with these various backbone CNNs. The results in Figure 5 demonstrate the comparison Average Precision (AP) and Average Processing Time (Atime) between different backbones using 1000 images, including 500 positive samples with cattle face and 500 negative samples without cattle face. In addition, to better assess the performance of various models on cattle face detection in detail, we also computed True Positive (TP), False Positive (FP), and False Negatives (FN) of seven backbones and then calculated the corresponding precision, recall, and F1 score, as presented in Table 1.

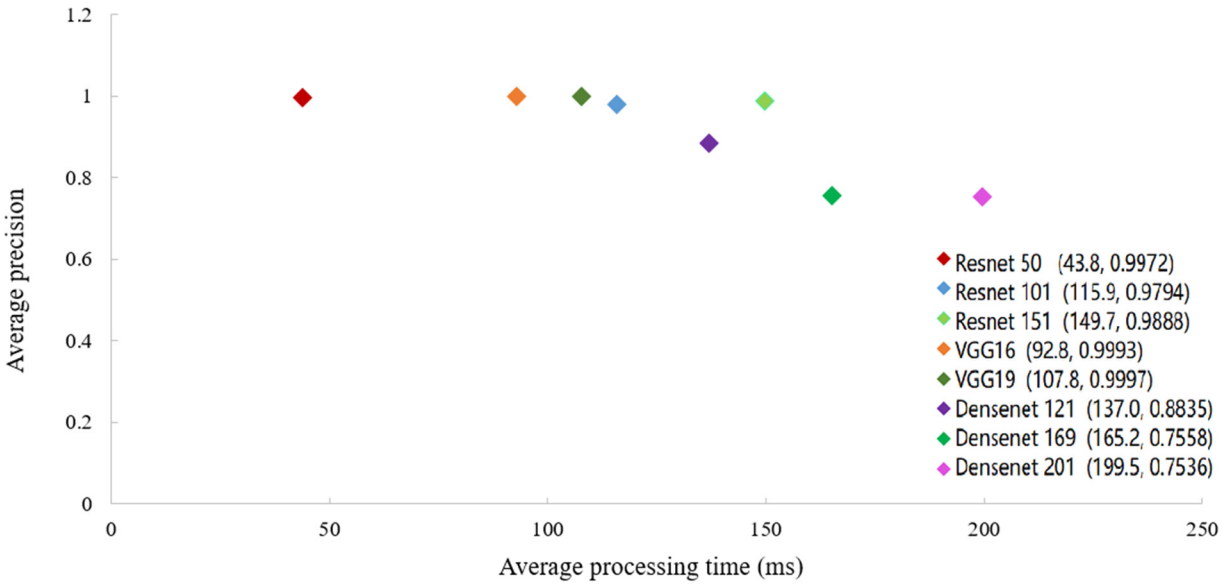


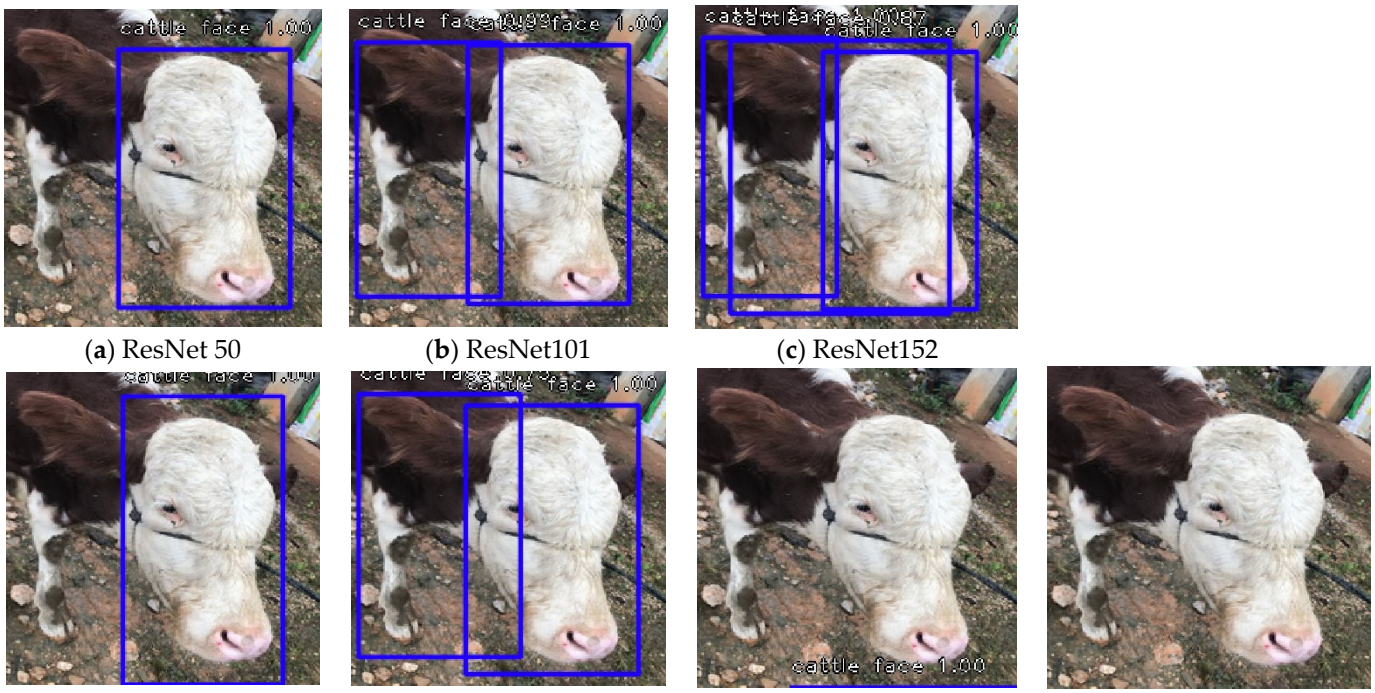
Figure 5. Average precision and average processing time of cattle face detection using different backbones.

Table 1. Comparison of detection results with different backbones.

Backbone	Precision	Recall	F1	TP	FP	FN
ResNet 50	0.9980	1.0000	0.9990	500	1	0
ResNet 101	0.9840	1.0000	0.9920	500	8	0
ResNet 152	0.9840	1.0000	0.9920	500	8	0
VGG16	0.8040	1.0000	0.8910	500	122	0
VGG19	0.8800	1.0000	0.9390	500	65	0
Densenet 121	0.3850	0.4220	0.4030	211	337	289
Densenet 169	0.6270	0.2760	0.3830	138	82	362

It can be seen from Figure 5 that the average precision of VGG 16 and VGG 19 are slightly higher than the value of ResNet 50 and achieve the best average precision, but the average processing time of ResNet 50 outperforms other backbones. As for cattle face detection, Densenet has a poor detection effect with the best average precision of 88.35% and the fastest processing time of 0.1370 s. AP and Atime are both significant metrics in the matter of how practical the system might be in actual use. Therefore, considering processing time and accuracy, the detection algorithm with ResNet 50 as the feature extraction model is regarded as having the best performance, whose AP reaches 99.8% and Atime is 0.0438 s per image.

As observed in Table 1, the cattle face detection model using ResNet 50 yields a precision of 99.8%, a 100% of recall and an F1 score of 0.9990, which are higher than other backbones. Moreover, the results concerning cattle face detection errors depict that the model achieves the lowest FP and FN rates with only 1 in 500 cattle faces potentially being misclassified in the case of ResNet 50. In contrast, although deeper ResNet including ResNet101, ResNet 152, and VGG network architectures obtain better performance on FP, they are reported to receive more falsely detected cattle face, especially using VGG. As with the results shown in Figure 5, the lowest scores on precision, recall, and F1 score are reported by employing Densenet due to the superior FP and FN rates but the lowest TP rate. Some representative examples for the prediction on the test image processed by seven different backbones is visualized in Figure 6.



(**a**) ResNet 50 (**b**) ResNet101 (**c**) ResNet152

(**d**) VGG16 (**e**) VGG19 (**f**) Densenet 121 (**g**) Densenet 169

Figure 6. Comparisons of cattle face detection processed by seven different backbones.

4.3. Comparison with Other State-of-the-Art Object Detection Algorithms

The proposed RetinaNet based multi-view cattle face detection is also compared to show its advantages over the typical existing object detection approaches. Yolov3 and Faster R-CNN are the typical works of object detectors in practice. For instance, Faster R-CNN has been attempted to explore the multi-class fruit detection [42–44], livestock detection [45], posture detection of pigs [46], and cattle face detection [35]. Yolov3 has also been applied to fruit and fruit disease detection [47–50], plant and plant disease and pest detection [51–53], livestock behavior detection [47,54], and fish detection [55]. Therefore, experiments in this paper are conducted to compare the testing results of these competing methods with the ground truth information, and the results are summarized in Table 2.

Table 2. Comparison of detection results with three competing methods.

Methods	AP	Atime	Precision	Recall	F1	TP	FP	FN
Yolov3	0.9968	0.1368	0.8700	1	0.9300	498	72	2
Faster R-CNN	0.9857	0.1526	0.9940	1	0.9970	500	3	0
RetinaNet + ResNet 50	0.9980	0.0438	0.9980	1	0.9990	500	1	0

It is observed from Table 2 that RetinaNet with ResNet 50 show better detection performance than Yolov3 and Faster R-CNN in both detection accuracy and calculation requirement for future online detection (AP of 99.8% and Atime of 0.0438 s). The results indicate that RetinaNet is most competent in real-world practice as the datasets are in different complex scenes with severe face-pose variation and different degrees of occlusion. Yolov3 and Faster R-CNN achieved nearly similar performance with RetinaNet in AP (99.68% for Yolov3 and 99.8% for RetinaNet) and F1 score (0.9970 for Faster R-CNN and 0.9990 for RetinaNet), respectively, but the F1 score is preferable as the metric for "true positive detection" whilst average precision is preferable for "boundary extraction" of cattle face. Therefore, Yolov3 and Faster R-CNN are not sufficiently reliable in complex multi-view cattle face detection.

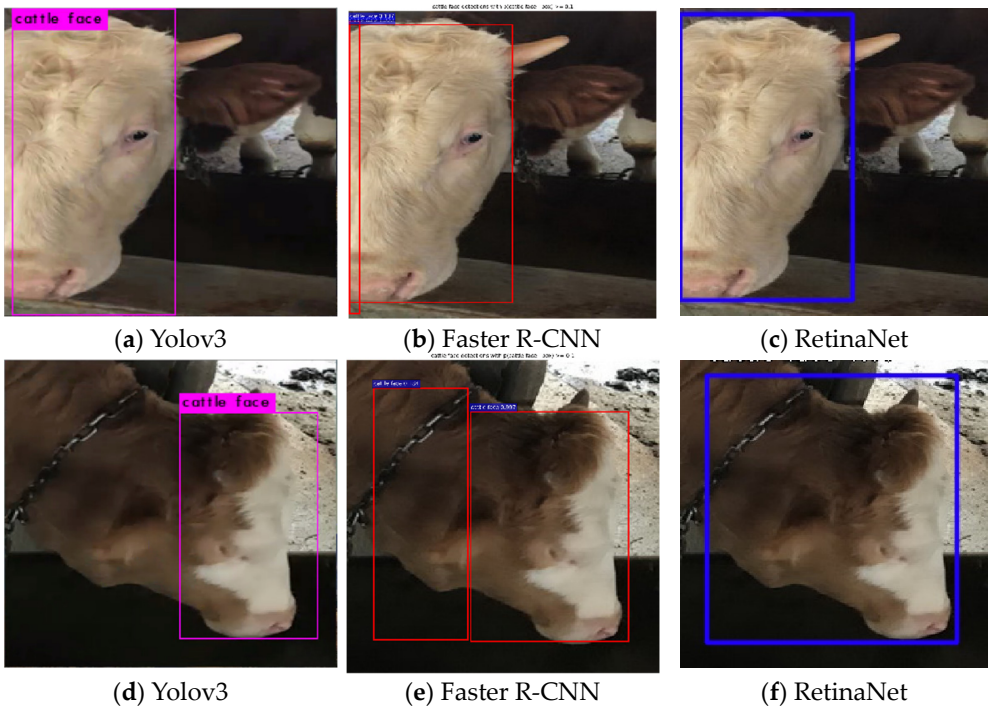
4.4. Evaluation of Multi-View Cattle Face Detection Results

The major misdetections of the abovementioned algorithms concern multi-view cattle face in complex conditions. To clearly observe the comparisons of results for multi-view cattle face detection in different scenes, 100 images were selected from 500 positive samples for three scenes of partial occlusion, light change, and posture change, and then the detection AP values and F1 scores were calculated separately for these competing detection models, as shown in Table 3.

Table 3. Comparison of detection results under different conditions.

Methods	Partial Occlusion		Light Variation		Posture Change	
	AP	F1	AP	F1	AP	F1
Yolov3	0.9980	1.0000	1.0000	1.0000	0.9720	0.9980
Faster R-CNN	0.9910	0.9990	1.0000	1.0000	0.9840	0.9980
RetinaNet + ResNet 50	1.0000	1.0000	1.0000	1.0000	0.9980	0.9990

As seen in Table 3, RetinaNet with ResNet 50 outperforms Yolov3 and Faster R-CNN under three particularly challenging situations. Three detection models all present very accurate detection results with AP of 100% and F1 score of 1.0000 in the situation with light changes, which implies that CNN-based deep learning algorithms are robust to illumination variations. However, as observed, there are inaccurate detection boundaries using Yolov3 and false cattle face detections using Faster R-CNN while the performance of RetinaNet remains relatively high in partial occlusion situation. Although three detection models do not present good detection results in posture change situations, RetinaNet achieves better performance in detection accuracy and boundary accuracy owing to the structure of FPN and focal loss in the model. Faster R-CNN presents the advantage of RPN, which is commonly used in two-stage detectors, and thus the boundary precision is higher than Yolov3. To facilitate the readers to visually observe the comparisons of results, this paper compares the predictions processed by the above-competing methods under partial occlusion and posture change situations, as shown in Figure 7.



(**a**) Yolov3 (**b**) Faster R-CNN (**c**) RetinaNet

(**d**) Yolov3 (**e**) Faster R-CNN (**f**) RetinaNet

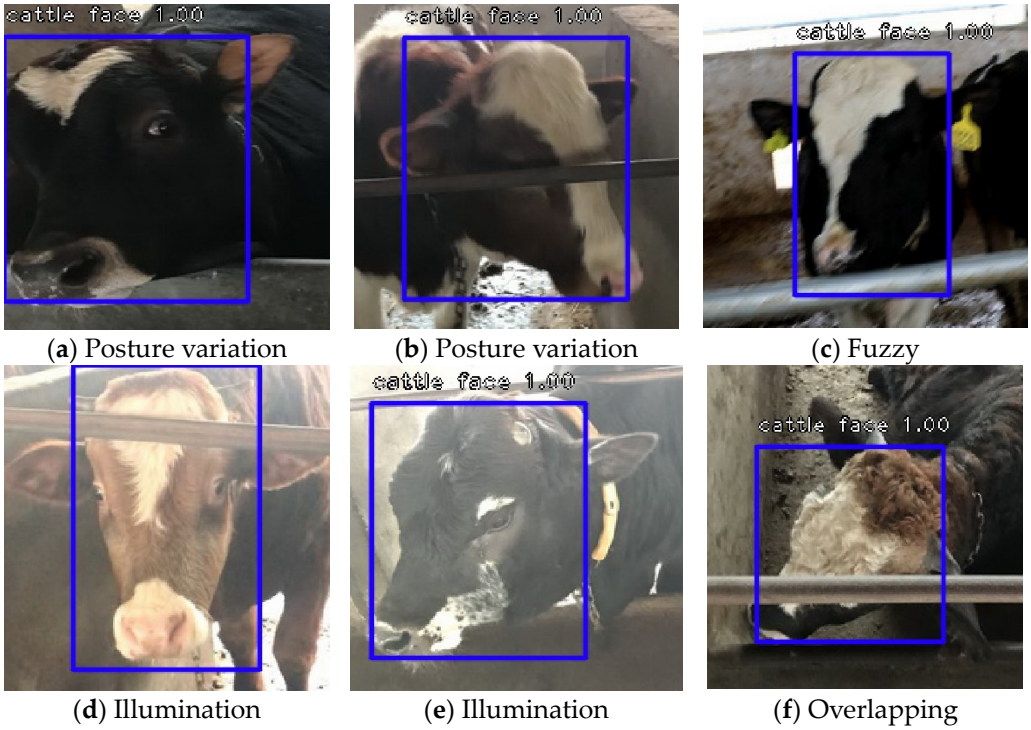
Figure 7. Comparisons of cattle face detection processed by three object detection algorithms in partial occlusion and posture change situations.

5. Discussion

This paper evaluated an up-to-date object detector, RetinaNet, to automate the face detection process for a livestock identification vision system in the farmland. The key novelty of the study is the application evaluation of the RetinaNet algorithm with various backbones and comparisons with typical competing detection models for multi-view cattle face detection in complex and relevant cattle production scenarios. The essence of the detection in this paper is bounding-box location and classification with confidence. Previous studies in cattle face suffered the deviation of the bounding-box [56] and the challenge for dataset collection from complex scenarios [35]. The strong point of the RetinaNet is the capability to perform both relatively high detection accuracy and fast processing time of cattle face within the imagery. This allows for the development of further algorithms to perform tasks such as facial expression assessment from the imagery for welfare monitoring. Cattle face detection in the paper is the first step toward real-time individual livestock identification in farming environments that have different applications, such as the cattle insurance industry, meat products traceability [57], and other animal welfare improvements.

Transfer learning is an essential part of machine learning as pretrained CNN models can be fine-tuned and re-trained to perform new tasks when limited annotated data exists for training. However, the generalization capabilities of various deep networks on different datasets might change due to their architecture [43,58,59]. Therefore, this study compared the performance quantitatively of ResNet, VGG, and Densenet with different depth to select the optimal backbone in this detection task. The results indicate that RetinaNet with ResNet 50 achieves the best performance with an average precision of 99.8%, F1 score of 0.9990, and average processing time of 0.0438 s. Since backbones with better performance can improve the accuracy of detection, and there is no agreed pretrained CNN model in object detection algorithms, this backbone could be properly adjusted and optimized depending on the circumstances and applications. For instance, Yolov3 incorporating the DenseNet for apple detection in various growth periods [49] was considered to perform well. Still, ResNet may be better for fruit detection and instance segmentation [43], and plant disease detection achieves better results using VGG architecture [60].

For demonstrating the feasibility of the proposed framework further, this study made the performance comparisons with two competitive algorithms of object detection on the same datasets. The detection results presented illustrate that the AP and Atime provided by the RetinaNet with ResNet 50 model are significantly better than the other two models, reflecting the superiority of the proposed cattle face detection model. Considering the multi-view face caused by various unstructured scenes in actual cattle production scenarios, such as overlapping, occlusion, and illumination changes, the cattle face detection accuracy could be reduced to some extent. The F1 scores and average precision metrics were assessed over unstructured scenes in the study, and it is worth mentioning that the performance of RetinaNet was better than other algorithms. Some detection results of cattle faces are shown in Figure 8. Especially for partial occlusion and light variation situations, the accuracy of cattle face detection using RetinaNet reaches 100%, but the posture change situation is particularly challenging, even using RetinaNet and computer vision in general. The suggested main reason for this performance discrepancy of posture change situation can be attributed to multiple behaviors, such as leaning over to graze or drink and lying on the side to rest, which then bring difficulties to cattle face detection.



(**a**) Posture variation (**b**) Posture variation (**c**) Fuzzy

(**d**) Illumination (**e**) Illumination (**f**) Overlapping

Figure 8. Detection results of cattle faces in various unstructured scenes.

6. Conclusions

Developing deep learning for object detection and image processing is crucial to the livestock identification system, which substitutes for wearable devices such as RFID ear tags, thus reducing the damage to animals. To establish the livestock machine vision system capable of monitoring individuals, this paper focused on cattle face detection, which is an important component of envisaged future technology. The state-of-art RetinaNet detection model proposed in this study was assessed on various unstructured scenes. The compared metrics performed successfully across a range of scenarios with an average precision score of 99.8% and an average processing time of 0.0438 s. The results presented indicate that the proposed model was particularly effective for the detection of cattle faces with illumination changes, overlapping, and occlusion. Compared to the existing algorithms, the proposed model has better universality and robustness both in accuracy and speed, which makes it generally more applicable for actual scenes. However, the conditions of training and testing are the same in this work, and the robustness of the system may be questioned; thus, further experiments are needed.

This work has potential for computer vision system integration into mobile apps to perform not only livestock detection and counting and individual identification, but also facial expression recognition for animal welfare. Despite the significantly high success of the proposed method, it is still far from being a generic tool that could be used in actual livestock production scenarios. Future work will focus on a lightweight neural network to improve the running speed of cattle face detection. In addition, future work will also concentrate on building an autonomous livestock individual identification system using facial features.

Author Contributions: Conceptualization, W.W. and B.X.; Methodology, B.X., L.G. and G.C.; Software, B.X. and Y.W.; Validation, B.X., W.Z. and Y.L.; Formal Analysis, Y.L.; Investigation, B.X. and Y.W.; Resources, L.G. and G.C.; Data Curation, W.Z. and Y.L.; Writing—Original Draft Preparation, B.X. and. W.W.; Writing—Review & Editing, W.W.; Visualization, B.X.; Supervision, W.W.; Project Administration, W.W., L.G.; Funding Acquisition, L.G. and G.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by China Scholarship Council (202003250122) and was funded by Inner Mongolia Autonomous Region Science and Technology Major Project (2020ZD0004), National Natural Science Foundation of China (32060776), Youth Science Foundation of Jiangxi Province (20192ACBL21023), and Hebei Province Key Research and Development Plan (20327202D, 20327401D).

Acknowledgments: We are grateful to two private housing farms in Jiangxi Province in China for their kindly support with data collection.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Liaghat, S.; Balasundram, S.K. A Review: The Role of Remote Sensing in Precision Agriculture. *Am. J. Agric. Biol. Sci.* **2010**, *5*, 553–564. [CrossRef]
2. Auernhammer, H. Precision farming—The environmental challenge. *Comput. Electron. Agric.* **2001**, *30*, 31–43. [CrossRef]
3. Vranken, E.; Berckmans, D. Precision livestock farming for pigs. *Anim. Front.* **2017**, *7*, 32–37. [CrossRef]
4. Andonovic, I.; Michie, C.; Cousin, P.; Janati, A.; Pham, C.; Diop, M. Precision Livestock Farming Technologies. In Proceedings of the 2018 Global Internet of Things Summit (GIoTS), Bilbao, Spain, 4–7 June 2018; pp. 1–6.
5. Xu, B.; Wang, W.; Falzon, G.; Kwan, P.; Guo, L.; Chen, G.; Tait, A.; Schneider, D. Automated cattle counting using Mask R-CNN in quadcopter vision system. *Comput. Electron. Agric.* **2020**, *171*, 105300. [CrossRef]
6. Disney, W.T.; Green, J.W.; Forsythe, K.W.; Wiemers, J.F.; Weber, S.J.R.T. Benefit-cost analysis of animal identification for disease prevention and control. *Rev. Sci. Tech. l'OIE* **2001**, *20*, 385–405. [CrossRef]
7. Gwaza, D.; Gambo, D. Application of Radio Frequency Identification to Selection for Genetic improvement of Rural Livestock Breeds in Developing Countries. *J. Anim. Husb. Dairy Sci.* **2017**, *1*, 38–52.
8. Yordanov, D.; Angelova, G. Identification and Traceability of Meat and Meat Products. *Biotechnol. Biotechnol. Equip.* **2006**, *20*, 3–8. [CrossRef]
9. Awad, A.I. From classical methods to animal biometrics: A review on cattle identification and tracking. *Comput. Electron. Agric.* **2016**, *123*, 423–435. [CrossRef]
10. Leslie, E.; Hernández-Jover, M.; Newman, R.; Holyoake, P. Assessment of acute pain experienced by piglets from ear tagging, ear notching and intraperitoneal injectable transponders. *Appl. Anim. Behav. Sci.* **2010**, *127*, 86–95. [CrossRef]
11. Jones, S.M. *Tattooing of Cattle and Goats*; University of Arkansas System: Little Rock, AR, USA, 2014.
12. Adcock, S.J.J.; Tucker, C.B.; Weerasinghe, G.; Rajapaksha, E. Branding Practices on Four Dairies in Kantale, Sri Lanka. *Animals* **2018**, *8*, 137. [CrossRef] [PubMed]
13. Stanford, K.; Stitt, J.; Kellar, J.A.; McAllister, T. Traceability in cattle and small ruminants in Canada. *Rev. Sci. Et Tech. Int. Off. Epizoot.* **2001**, *20*, 510–522. [CrossRef] [PubMed]
14. Yang, L.; Liu, X.Y.; Kim, J.S. Cloud-based Livestock Monitoring System Using RFID and Blockchain Technology. In Proceedings of the 2020 7th IEEE International Conference on Cyber Security and Cloud Computing (CSCloud)/2020 6th IEEE International Conference on Edge Computing and Scalable Cloud (EdgeCom), New York, NY, USA, 1–3 August 2020; pp. 240–245.
15. Klindtworth, M.; Wendl, G.; Klindtworth, K.; Pirkelmann, H. Electronic identification of cattle with injectable transponders. *Comput. Electron. Agric.* **1999**, *24*, 65–79. [CrossRef]
16. Zaiqiong, W.; Zetian, F.; Wei, C.; Jinyou, H. A RFID-based traceability system for cattle breeding in China. In Proceedings of the 2010 International Conference on Computer Application and System Modeling (ICCASM 2010), Taiyuan, China, 22–24 October 2010; pp. V2-567–V2-571.
17. Whittier, J.C.; Shadduck, J.A.; Golden, B.L. *Secure Identification, Source Verification of Livestock—The Value of Retinal Images and GPS*; Wageningen Academic Publishers: Wageningen, The Netherlands, 2003; pp. 167–172.
18. Gonzales Barron, U.; Corkery, G.; Barry, B.; Butler, F.; McDonnell, K.; Ward, S. Assessment of retinal recognition technology as a biometric method for sheep identification. *Comput. Electron. Agric.* **2008**, *60*, 156–166. [CrossRef]
19. Kumar, S.; Pandey, A.; Sai Ram Satwik, K.; Kumar, S.; Singh, S.K.; Singh, A.K.; Mohan, A. Deep learning framework for recognition of cattle using muzzle point image pattern. *Measurement* **2018**, *116*, 1–17. [CrossRef]
20. Kumar, S.; Singh, S.K.; Abidi, A.I.; Datta, D.; Sangaiah, A.K. Group Sparse Representation Approach for Recognition of Cattle on Muzzle Point Images. *Int. J. Parallel Program.* **2018**, *46*, 812–837. [CrossRef]
21. Mukai, N.; Zhang, Y.; Chang, Y. Pet Face Detection. In Proceedings of the 2018 Nicograph International (NicoInt), Tainan, Taiwan, 28–29 June 2018; pp. 52–57.
22. Kumar, S.; Tiwari, S.; Singh, S.K. Face recognition for cattle. In Proceedings of the 2015 Third International Conference on Image Information Processing (ICIIP), Waknaghat, India, 21–24 December 2015; pp. 65–72.
23. Clark, A.W. Calculating the Weight of a Pig through Facial Geometry Using 2-Dimensional Image Processing. Master's Thesis, Texas Tech University, Lubbock, TX, USA, 2015.
24. Jaddoa, M.; Gonzalez, L.; Cuthbertson, H.; Al-Jumaily, A. Multi View Face Detection in Cattle Using Infrared Thermography. In *Proceedings of the Applied Computing to Support Industry: Innovation and Technology*; Ramadi, Iraq, 15–16 September 2019, Springer: Cham, Switzerland, 2020; pp. 223–236.

25. Yamada, A.; Kojima, K.; Kiyama, J.; Okamoto, M.; Murata, H. Directional edge-based dog and cat face detection method for digital camera. In Proceedings of the 2011 IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, NV, USA, 9–12 January 2011; pp. 87–88.
26. Lecun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [CrossRef]
27. Vlachynska, A.; Oplatkova, Z.K.; Turecek, T. Dogface Detection and Localization of Dogface's Landmarks. In *Artificial Intelligence and Algorithms in Intelligent Systems*; Springer: Cham, Switzerland, 2019; pp. 465–476.
28. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587.
29. Girshick, R. Fast R-CNN. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 1440–1448.
30. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [CrossRef]
31. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
32. Redmon, J.; Farhadi, A. YOLO9000: Better, Faster, Stronger. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 6517–6525.
33. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. SSD: Single Shot MultiBox Detector. In Proceedings of the Computer Vision—ECCV 2016, Amsterdam, The Netherlands, 11–14 October 2016; Springer: Cham, Switzerland, 2016; pp. 21–37.
34. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
35. Yao, L.; Hu, Z.; Liu, C.; Liu, H.; Kuang, Y.; Gao, Y. Cow face detection and recognition based on automatic feature extraction algorithm. In Proceedings of the ACM Turing Celebration Conference—China, Chengdu, China, 17–19 May 2019.
36. Gou, X.; Huang, W.; Liu, Q. A Cattle Face Detection Method Based on Improved NMS. *Comput. Modernization* **2019**, *7*, 43–46.
37. Ochoa-Ruiz, G.; Angulo-Murillo, A.A.; Ochoa-Zezzatti, A.; Aguilar-Lobo, L.M.; Vega-Fernández, J.A.; Natraj, S. An Asphalt Damage Dataset and Detection System Based on RetinaNet for Road Conditions Assessment. *Appl. Sci.* **2020**, *10*, 3974. [CrossRef]
38. Yunqi, C.; Basak, O. *Automated Firearms Detection in Cargo X-Ray Images using RetinaNet*; International Society for Optics and Photonics: Bellingham, WA, USA, 2019; Volume 10999. [CrossRef]
39. Afif, M.; Ayachi, R.; Said, Y.; Pissaloux, E.; Atri, M. An Evaluation of RetinaNet on Indoor Object Detection for Blind and Visually Impaired Persons Assistance Navigation. *Neural Process. Lett.* **2020**, *51*, 2265–2279. [CrossRef]
40. Zou, Z.; Shi, Z.; Guo, Y.; Ye, J. Object detection in 20 years: A survey. *arXiv* **2019**, arXiv:1905.05055.
41. Lin, T.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal Loss for Dense Object Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 318–327. [CrossRef] [PubMed]
42. Tu, S.; Pang, J.; Liu, H.; Zhuang, N.; Chen, Y.; Zheng, C.; Wan, H.; Xue, Y. Passion fruit detection and counting based on multiple scale faster R-CNN using RGB-D images. *Precis. Agric.* **2020**, *21*, 1072–1091. [CrossRef]
43. Kang, H.; Chen, C. Fruit detection, segmentation and 3D visualisation of environments in apple orchards. *Comput. Electron. Agric.* **2020**, *171*, 105302. [CrossRef]
44. Wan, S.; Goudos, S. Faster R-CNN for multi-class fruit detection using a robotic vision system. *Comput. Netw.* **2020**, *168*, 107036. [CrossRef]
45. Xu, B.; Wang, W.; Falzon, G.; Kwan, P.; Guo, L.; Sun, Z.; Li, C. Livestock classification and counting in quadcopter aerial images using Mask R-CNN. *Int. J. Remote Sens.* **2020**, *41*, 8121–8142. [CrossRef]
46. Nasirahmadi, A.; Sturm, B.; Edwards, S.; Jeppsson, K.-H.; Olsson, A.-C.; Müller, S.; Hensel, O. Deep Learning and Machine Vision Approaches for Posture Detection of Individual Pigs. *Sensors* **2019**, *19*, 3738. [CrossRef] [PubMed]
47. Fu, L.; Feng, Y.; Wu, J.; Liu, Z.; Gao, F.; Majeed, Y.; Al-Mallahi, A.; Zhang, Q.; Li, R.; Cui, Y. Fast and accurate detection of kiwifruit in orchard using improved YOLOv3-tiny model. *Precis. Agric.* **2020**, *22*, 754–776. [CrossRef]
48. Koirala, A.; Walsh, K.B.; Wang, Z.; McCarthy, C. Deep learning for real-time fruit detection and orchard fruit load estimation: Benchmarking of 'MangoYOLO'. *Precis. Agric.* **2019**, *20*, 1107–1135. [CrossRef]
49. Tian, Y.; Yang, G.; Wang, Z.; Li, E.; Liang, Z. Detection of Apple Lesions in Orchards Based on Deep Learning Methods of CycleGAN and YOLOV3-Dense. *J. Sens.* **2019**, *2019*, 7630926. [CrossRef]
50. Kuznetsova, A.; Maleva, T.; Soloviev, V. Detecting Apples in Orchards Using YOLOv3. In Proceedings of the Computational Science and Its Applications (ICCSA) 2020, Cagliari, Italy, 1–4 July 2020; Springer: Cham, Switzerland, 2020; pp. 923–934.
51. Zhou, J.; Tian, Y.; Yuan, C.; Yin, K.; Yang, G.; Wen, M. Improved UAV Opium Poppy Detection Using an Updated YOLOv3 Model. *Sensors* **2019**, *19*, 4851. [CrossRef]
52. Liu, J.; Wang, X. Tomato Diseases and Pests Detection Based on Improved Yolo V3 Convolutional Neural Network. *Front. Plant Sci.* **2020**, *11*, 898. [CrossRef]
53. Liu, G.; Nouaze, J.C.; Touko Mbouembe, P.L.; Kim, J.H. YOLO-Tomato: A Robust Algorithm for Tomato Detection Based on YOLOv3. *Sensors* **2020**, *20*, 2145. [CrossRef]

54. Wang, J.; Wang, N.; Li, L.; Ren, Z. Real-time behavior detection and judgment of egg breeders based on YOLO v3. *Neural Comput. Appl.* **2020**, *32*, 5471–5481. [CrossRef]
55. Raza, K.; Hong, S. Fast and Accurate Fish Detection Design with Improved YOLO-v3 Model and Transfer Learning. *Int. J. Adv. Comput. Sci. Appl.* **2020**, *11*, 7–16. [CrossRef]
56. Wang, H.; Qin, J.; Hou, Q.; Gong, S. Cattle Face Recognition Method Based on Parameter Transfer and Deep Learning. *J. Phys. Conf. Ser.* **2020**, *1453*, 012054. [CrossRef]
57. Zhao, J.; Li, A.; Jin, X.; Pan, L. Technologies in individual animal identification and meat products traceability. *Biotechnol. Biotechnol. Equip.* **2020**, *34*, 48–57. [CrossRef]
58. Wang, G.; Sun, Y.; Wang, J. Automatic Image-Based Plant Disease Severity Estimation Using Deep Learning. *Comput. Intell. Neurosci.* **2017**, *2017*, 2917536. [CrossRef] [PubMed]
59. Ayan, E.; Erbay, H.; Varçın, F. Crop pest classification with a genetic algorithm-based weighted ensemble of deep convolutional neural networks. *Comput. Electron. Agric.* **2020**, *179*, 105809. [CrossRef]
60. Ferentinos, K.P. Deep learning models for plant disease detection and diagnosis. *Comput. Electron. Agric.* **2018**, *145*, 311–318. [CrossRef]