

Article

Towards Facial Expression Recognition for On-Farm Welfare Assessment in Pigs

Mark F. Hansen ^{1,*}, Emma M. Baxter ², Kenneth M. D. Rutherford ², Agnieszka Futro ², Melvyn L. Smith ¹
and Lyndon N. Smith ¹

¹ Centre for Machine Vision, BRL, UWE Bristol, Bristol BS16 1QY, UK; melvyn.smith@uwe.ac.uk (M.L.S.); lyndon.smith@uwe.ac.uk (L.N.S.)

² Animal Behaviour and Welfare, Animal and Veterinary Sciences Research Group, SRUC, West Mains Road, Edinburgh EH9 3JG, UK; Emma.Baxter@sruc.ac.uk (E.M.B.); Kenny.Rutherford@sruc.ac.uk (K.M.D.R.); Agnieszka.Futro@sruc.ac.uk (A.F.)

* Correspondence: mark.hansen@uwe.ac.uk

Abstract: Animal welfare is not only an ethically important consideration in good animal husbandry but can also have a significant effect on an animal's productivity. The aim of this paper was to show that a reduction in animal welfare, in the form of increased stress, can be identified in pigs from frontal images of the animals. We trained a convolutional neural network (CNN) using a leave-one-out design and showed that it is able to discriminate between stressed and unstressed pigs with an accuracy of >90% in unseen animals. Grad-CAM was used to identify the animal regions used, and these supported those used in manual assessments such as the Pig Grimace Scale. This innovative work paves the way for further work examining both positive and negative welfare states with the aim of developing an automated system that can be used in precision livestock farming to improve animal welfare.

Keywords: animal welfare; pigs; deep learning; computer vision; stress detection; facial expression recognition



Citation: Hansen, M.F.; Baxter, E.M.; Rutherford, K.M.D.; Futro, A.; Smith, M.L.; Smith, L.N. Towards Facial Expression Recognition for On-Farm Welfare Assessment in Pigs.

Agriculture **2021**, *11*, 847. <https://doi.org/10.3390/agriculture11090847>

Received: 27 July 2021

Accepted: 1 September 2021

Published: 4 September 2021

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



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Animal welfare has become increasingly important over recent years due to societal ethical concerns, consumer demand [1] and also because improving welfare can improve farm production efficiency [2].

Along with physical illness and injury, another major contributor to negative welfare is stress as it threatens an animal's homeostasis and can trigger a variety of behavioural, neuroendocrine and immunological responses [3] as the animal tries to restore balance. If stress becomes a chronic condition, it can have significant pathological consequences. Thus, being able to quickly and accurately assess stress in individual animals would allow the farmer to make a specific and timely intervention and hopefully identify and mitigate the source of stress. Such a capability potentially offers a novel and valuable tool in precision animal husbandry, whereby the observation of the animal's expression might itself offer insight into its emotional state. This would allow the more appropriate and targeted management of individuals, reducing veterinary costs, improving farm productivity, and greatly enhancing the welfare of individual animals.

Being able to accurately evaluate animal welfare and determine an animal's quality of life requires a certain degree of scientific objectivity [4]. Currently, on-farm welfare assessment is often hampered by inter-observer variability, due to factors such as subjectivity and observer bias [5]. The time available for animal monitoring is also often limited and assessment may only be conducted at the group level and intermittently, only offering snapshots during an animal's life. The goal would be to provide near real-time

assessment which enhances and supports traditional human stockpersonship, allowing rapid intervention if an animal is showing signs of distress.

This paper will provide an overview of the current state of the art in this area and a review of the relevant methods that are employed. Section 3 will discuss how the data are captured, cleaned and organised, as well as the specific deep-learning architecture chosen for this work. The results in Section 4 demonstrate the efficacy of this approach before discussing the features that the network has learnt and how such a system might be deployed to provide fast and accurate management information for the farmer.

Contributions

1. A first attempt to automate stress detection using facial expression in pigs on the farm via machine vision using a convolutional neural network.
2. It is demonstrated that we can do so with >90% accuracy on animals that are not part of the model's training set.

2. Background

Attempts to estimate the emotional state from expressions has for the most part used humans as participants. From early approaches that identified the existence of universal expressions [6] to more advanced video-based methods [7] aiming at automated emotion recognition, it has been demonstrated that it is possible to infer expressions reliably and accurately. One of the most successful approaches breaks the face down into facial action units (FAUs) and codes the relative positional movements of facial features into expressions ([8,9]). This system was primarily designed to train humans to manually measure expression in a more objective method. Known as the Facial Action Coding System (FACS), it has also been successfully used in animals such as chimpanzees, horses, cats and dogs (see [10] for details). The assessment of pig facial expressions has been applied in studies of aggressive intent [11], as well as in studies using FAUs to categorise pain-levels [12,13]. What these papers have in common are the regions analysed—eyes (in terms of orbital tightening), snout and cheek muscle tightening and ear positioning. One potential issue with using FACS/FAUs, particularly when applied to animals, is that it relies on the expression always being present during observation (either in the live animals or via images/video), whereas expressions are often fleeting (at least in humans); this issue has placed limitations on the application of facial expression assessment in practical (as opposed to research) contexts. It also requires manual coding or that any automated system can find these facial units and assess them—something that is relatively straightforward in human subjects who are participating, but which is much more difficult in animals where there may be many uncontrollable variables and the subjects are unaware of their participation.

Stress can be defined as a “*cognitive perception of uncontrollability and/or unpredictability that is expressed in a physiological and behavioural response*” [14]. In animals, acute stress is often equated with the activation of the hypothalamic pituitary adrenocortical (HPA) axis and is therefore commonly physiologically assessed by sampling for circulating levels of cortisol or corticosterone (in blood, saliva, urine) or their metabolites (in faeces) [15]. Behavioural quantification may also be applied in order to identify stress and often allows for a more specific characterisation of the nature of the stress (e.g., pain, fear, social stress) being experienced.

However, neither the assessment of glucocorticoid levels nor detailed behavioural appraisal are very suitable for practical on-farm application. Whilst measuring cortisol via blood sampling is still widely considered to be the “gold standard” physiological indicator of stress, sampling is invasive [16] and often difficult in pigs, which limits its use outside of research, particularly if multiple sampling is required, e.g., for on-going monitoring. The practical application of physiological sampling, whether using blood or other tissues, for instance under farm conditions, is also limited by the fact that results are retrospective (i.e., the time needed for processing and analysis means that information is

only provided about an animal's state in the past) and that many physiological indicators alter in response to challenges with positive or negative valence. Similarly, a detailed assessment of behaviour is not feasible due to time constraints. As a result, new approaches that allow for the fast (real-time) and accurate identification of stress or other welfare problems in individual animals are required. Deep learning has meant that, in recent years, computer vision approaches can be deployed in far more demanding environments that are typically encountered on farms. While traditional methods have been extremely susceptible to many natural variances, e.g., changes in ambient light levels, changes in camera position, etc., deep learning models, have proven their resilience/generalising capabilities in many real-world situations, for example, from self-driving cars to face recognition to generating artwork. We used three such models: two readily available (Mask-RCNN [17] for segmentation, tiny-YOLO-v3 [18] for eye detection) and one of our own to accomplish the main aim of this paper. We therefore aimed to test whether a CNN is capable of "learning" the required features to allow it to discriminate between stressed and unstressed pigs without relying on manually coded FAC units.

3. Methods

The following section is divided into three subsections that cover how the data were collected, were preprocessed, and then the details of the convolutional neural network architecture and training procedure.

3.1. Data Collection and Organisation

3.1.1. Ethical Approval

To ground-truth the machine vision and learning techniques, facial images of pigs experiencing a negative affective state of stress was required. A social stress model developed by the authors [19,20] was refined and used here to impose a profound social subordination stress. It is perhaps the most well-known, reliable and commercially relevant method for producing a profound, acute response in pigs [21–24]. Social stress can be achieved when unfamiliar pigs are mixed together as a consequence of the aggression displayed by dominant animals towards subordinates. Therefore, a high stress situation was created when older multiparous sows were mixed with younger primiparous sows (i.e., gilts) who were the subjects of this study. The mixing of gilts was closely supervised and specific end-points were put in place to safeguard pig welfare. The original social stress model was refined to reduce the frequency of mixing, reduce the duration of the mix period, use non-resident multiparous sows and mix in the final third of pregnancy (but not within three weeks of the predicted parturition date) to reduce the risk of harm to foetal development. This study underwent internal ethical review by both SRUC's and UWE Bristol's Animal Welfare and Ethical Review Bodies (ED AE 16-2019 and R101) and was carried out under the UK Home Office license (P3850A80D).

3.1.2. Animals and Housing

Eighteen primiparous sows (hereafter gilts – Large White × Landrace × Duroc—“WhiteRoc”—Rattlerow Farms Ltd., Suffolk, UK) in seven batches were the subjects of this study. Prior to selection, gilts were housed in groups of 4–6 pigs. Each batch of selected gilts were moved from their home pens in the main farm building to an experimental building with similar housing and husbandry conditions. Each pen had a deep straw-bedded, part-covered kennel area (2.5 m long), a dunging passage (2.35 m long, equipped with a drinker allowing ad libitum access to fresh water), and 6 individual feeding stalls (1.85 m long, 0.5 m wide). A standard ration (2.5–3.0 kg per sow depending on body condition) of commercial concentrate feed for gestating sows was provided once a day for each pig (ForFarmers Nova UltraGest). Data that were collected from the first two of the seven batches were not usable due to technical issues and streamlining the collection process, so only the results from the final five batches are reported here (twelve gilts). The total number of images per condition, and gilts per batch (2–3) can be seen in Table 1.

Table 1. The statistics of the full dataset after cleaning showing how many pigs (N Pigs), and the number of images (N Images) for each condition are present in each batch.

Batch	N Pigs	Condition	N Images
1	2	Stressed	5957
		Unstressed	1980
2	3	Stressed	3588
		Unstressed	1222
3	2	Stressed	1282
		Unstressed	443
4	3	Stressed	1170
		Unstressed	672
5	2	Stressed	1501
		Unstressed	346

3.1.3. Image Collection and Social Stress Application

In front of the individual feeding stalls, cameras were set-up to collect still frame images (see Figure 1). Logitech C920 HD Pro Webcams (Logitech Europe S.A, EPFL—Quartier de l’Innovation, 1015 Lausanne, Switzerland) were used to capture images mounted out of reach of the pigs using Tencro adjustable gooseneck stands. The cameras were connected to Dell precision computers running “iSpy Connect” software to allow the motion-detection capture of the pigs each time they voluntarily entered their individual feeding stalls.



Figure 1. Image capture set-up. Six Logitech C920 HD Pro webcams positioned in front of individual feeding stalls. Image captured via iSpy Connect software installed on Dell Precision computers.

As images would need to be correctly assigned to individuals after data capture, gilts were given an individual identification mark on their bodies using Magnum Chisel Tip Sharpie black marker pens. These marks were only placed on the rear of the gilts so that they were not visible in the face-on images (i.e., to ensure markings were not picked up by automated image processing) but such that experimenters could correctly identify the pigs as they entered and exited the field of view.

Gilts were moved to the experimental building and allowed to settle in over a weekend period. The main experiment ran over a five-day period (Monday to Friday). Each gilt served as its own control for the study, therefore once settled into their new home pens, baseline images were collected for approximately 24 h (i.e., “unstressed” images) on the

Monday. In order to establish a “stressed” state, older multiparous sows were selected from the breeding herd to be mixed with the younger gilts.

These sows were moved to the experimental building at the same time as the gilts (i.e., given time to settle over the weekend) but were given residence over the test pen in order for them to gain a sense of ownership prior to the gilts being added.

On the day of the mix (Tuesday—MIX day) the gilts were mixed into the test pens containing the sows (see Figure 2). Mixing was monitored throughout the day to ensure severity thresholds were not exceeded. The aim was to establish social defeat in the gilts. When this happens, gilts are displaced from the high value areas of the pen (i.e., bedded area) and this was visible after the mix (see Figure 2).



Figure 2. Mix procedure. Older sows are mixed with younger sows (left-hand image) in order to establish social defeat (right-hand image) and a “stressed” state.

After the MIX day, “stressed” images were collected for a further 2.5 days (i.e., POST MIX 1, POST MIX 2 and POST MIX 3) on the Wednesday to Friday morning before both sows and gilts were split, inspected by a named veterinary surgeon (Home Office Licence procedure) and returned to their home pens on the Friday afternoon.

3.1.4. Image Identification and Cleaning

On average, each camera, set to motion detect, took over 20,000 images per day. These raw images were screened to remove any images of low quality as a result of poor lighting or focus and anywhere the pig was not clearly visible. However, images where only parts of the face were visible were kept in case composite facial features were later deemed useful. The usable images from this initial screening were then labelled according to the gilt identification number, before being assigned as either “unstressed” (i.e., PRE-MIX) or “stressed” (i.e., POST MIX1 + POST MIX2).

3.2. Dataset and Image Preprocessing

Once the data were collected and organised as detailed in Section 3.1, a number of image preprocessing steps were performed in order to further clean the dataset. Examples of each step can be seen in Figure 3 and the motivation is as follows:

1. To remove extraneous information from the images that might provide reliable but undesired discriminatory information (i.e., different objects in the background or different ambient illumination on different days);
2. To remove secondary animals from the automatically detected masks (i.e., there may be a second pig which is behind a gate/fence which the instance segmentation may detect);

- To remove any animals that are too far from the camera and therefore too small for any sort of useful facial analysis to be performed.

Whilst every care was taken to keep the conditions identical between acquisition sessions, realistically, over the duration of months that the trials were performed, the background inevitably changed. It was therefore necessary to remove this from the images. There are potentially many methods that can be used to separate the sow from its background, and for the purposes of this experiment, we used instance segmentation via Mask-RCNN. This network is capable of pixel level segmentation and object classification. Unfortunately, amongst the 90 object classes that the Common Objects in Context (COCO) dataset [25] used for training the model, “pig” is not represented. However, there are 10 classes for living animals, and by lowering the detection confidence to 0.5, the model was able to reliably detect and segment sows from their backgrounds.

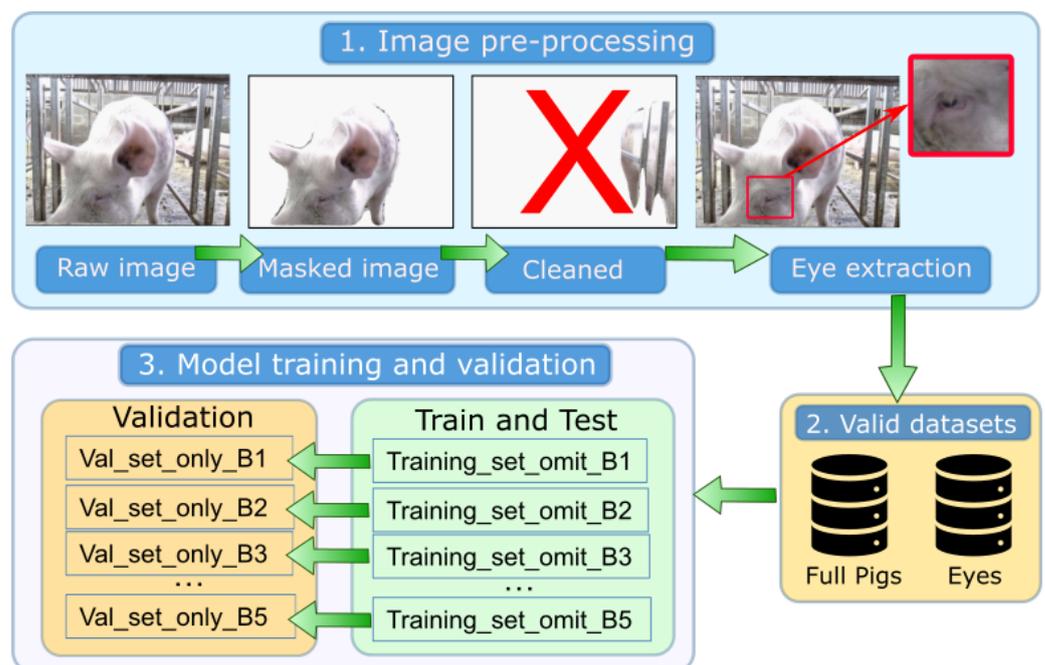


Figure 3. Pipeline of the process showing the preprocessing of the raw images to generate two datasets of valid images that are then used to create leave-one-out training and validation datasets where BN (e.g., B1, B2, etc.) represents the batch number.

With the backgrounds removed from the images, we are left with two further problems: that more than one animal may be detected, and the primary animal may appear too small to effectively provide reliable facial features.

For the first of these, the secondary animal will be less central in the image, so a small 20×20 px region in the centre of the field of view is checked for the presence of masked pixels that correspond to an animal. If there are none present, then that image is removed from the dataset.

To mitigate the second issue of animals that are too far away from the camera, a naive approach of a minimum pixel area (representing the segmented pig) was initially used. While this was successful in removing such pigs, it highlighted a further problem that occurs when pigs are too close—they obscure the entire field of view, often with no facial features showing (e.g., only the forehead of the animal). An object detector (tiny-YOLO-v3) was therefore trained to detect pigs’ eyes. Two-thousand and eleven (2011) images were randomly selected and bounding boxes were used to annotate the eye regions in the images. Pre-trained weights MS-COCO weights were used to initialise the model and after 4000 training iterations, loss was at 0.5077 and the mean average precision (mAP) was 97%. The

model was then used to detect eyes across the entire dataset. Images in which no eyes were detected were then excluded from the dataset.

The dataset statistics for the remaining data used in the following experiments can be seen in Table 1. While there is clearly a data imbalance in the image numbers between the “stressed” and “unstressed” (typically ~2:1) (due to baseline recording of “unstressed” being 1 day and “stressed” recorded over 2.5 days), the results presented in Section 4 show that it does not have a detrimental effect (i.e., we are not seeing a significant imbalance on precision/recall between the classes) on the training of the model, but methods to address this, such as class weights, could be employed and may further improve results.

3.3. Description of our CNN and the Leave-One-Out Cross-Validation Paradigm

This section details the architecture and hyper-parameters of the CNN used as well as the methodology for partitioning the dataset.

The chosen architecture is very much based on the model successfully used for biometric pig face recognition in [26] and consists of six convolutional blocks comprised of convolution layers with ReLU activation and then alternating max-pooling (2×2) and drop-out (20%) layers. The 256×256 px image size used as input is far larger than those in the previous work to help reduce the likelihood that potentially important small animal features are lost. The features extracted by the convolutional layers are then fed to a fully connected network with one output that represents the “stressed” (1) or “unstressed” (0) classes. The architecture can be seen in Figure 4. Whilst we demonstrate that this choice of architecture delivers some encouraging results, as it did when used for pig face recognition, we did not experiment with optimising hyper-parameters, so it is likely that further efficiency and accuracy improvements can be made.

Various batch sizes were explored, and 80 was chosen as optimal for all experiments. One hundred epochs (complete training and update runs), an ADAM optimizer and a learning rate of 0.001 were used.

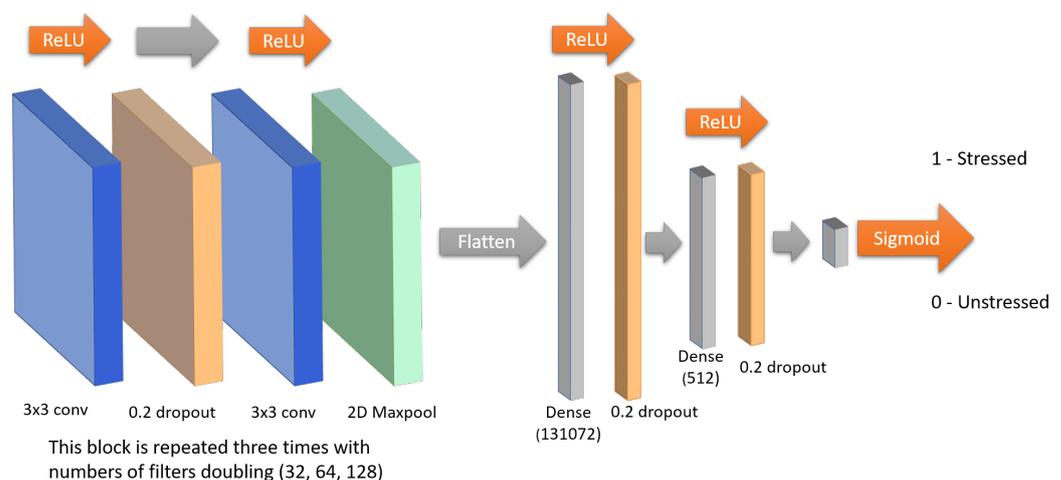


Figure 4. Architecture of our CNN consisting of three repeating blocks of a convolutional layer, a max-pooling layer, a convolutional layer and a dropout layer. The output consists of one node with a sigmoid activation function representing the two classes—stressed and unstressed.

For training the model, we selected a 90:10 ratio for the training:testing split which was randomly selected from the entire dataset. One important aspect that can be overlooked when training CNNs on image classification tasks using images from video sequences when using this paradigm of data partitioning is that the training and testing dataset can contain extremely similar images. Therefore, assumptions about the generalisability of the model can be incorrect. In [26], the dataset was analysed in terms of the structural similarity (SSIM) between sequential images, and included only those which were sufficiently different. Whilst this approach may have been effective in that work, we selected here to use a

leave-one-out cross-validation approach, as there are sufficient data, and we need to be confident that whatever features the network extracts from training are generalisable to unseen animals, i.e., we need to discount the possibility that the network has learnt features related to identity or features that relate to specific animals. A key point here is that it would not be ideal to have a model that was only capable of correctly identifying a stressed pig if it had already been trained to recognise stress in that particular pig. Rather, it must be able to detect stress in previously unseen animals that are not part of the training set.

The leave-one-out cross-validation paradigm is implemented at a batch level (batches contain two or more pigs, each recorded under stressed and unstressed conditions over different days), so that if there are five batches [1,2,3,4,5], the model would be trained on batches [1,2,3,4] and evaluated on batch 5. This is repeated so that each batch is used as the evaluation batch, and the remaining batches are used for training. In the example, this would mean that five models would be generated and evaluated against the omitted batch. The training process on batches (e.g., [1,2,3,4]) uses this paradigm, with 100% of these four batches used for training, and then, the actual validation of the model at the end of each epoch is performed on completely unseen animals from different acquisition days (e.g., batch [5]). For completeness, we also show the results of training the model against all data (e.g., batches [1,2,3,4,5]) using a data split of 90:10 (train:validation) to ensure that the model does not overfit the data. An example of this training run can be seen in Figure 5, which shows good correlation between the train and validation loss over 100 epochs. In this particular figure, the model looks as though it has not quite fully converged as the loss gradient is not zero, and may benefit from further training; however, the improvement in accuracy is likely to be minimal.

Figure 6 shows the equivalent, from one of the leave-one-out cross-validation sets, and while the training shows a very similar pattern, the validation loss remains considerably higher than that of the training set. This is to be expected because the validation data are considerably different to the training set (different pigs on different days), and the validation data have no influence on the training. However, the fact that it drops to a low level, remains relatively stable, and shows no signs of increasing is indicative of the fact that the model has learnt to extract features that allow it to infer whether an animal is stressed or unstressed, and that these features are generalisable to an unseen dataset. The fact that the validation accuracy is very similar to the training accuracy is also very encouraging.

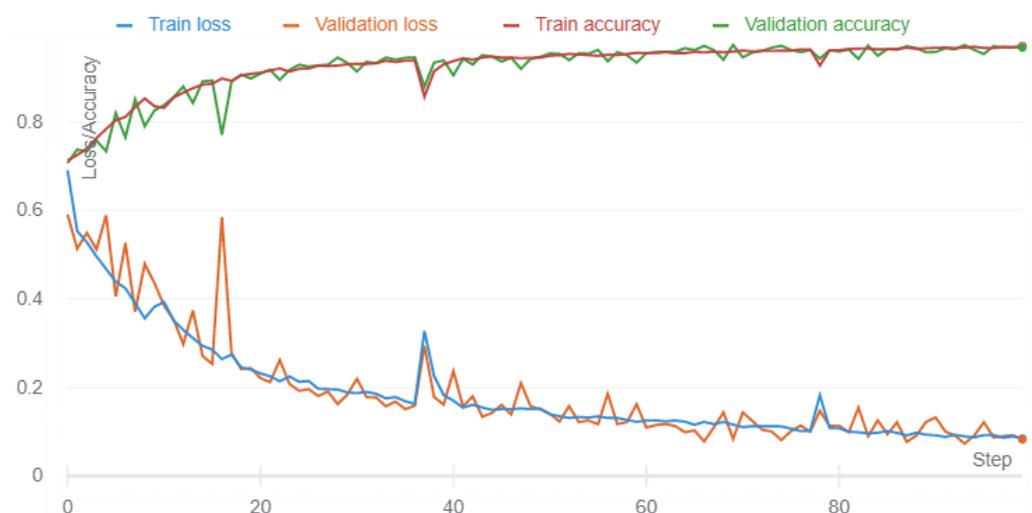


Figure 5. Training loss and accuracy as well as validation loss and accuracy from the dataset containing all the data. This pattern is representative of all leave-one-out batch training and indicates that the model has not overfitted the training set. Generated using “Weights and Biases” integration [27].

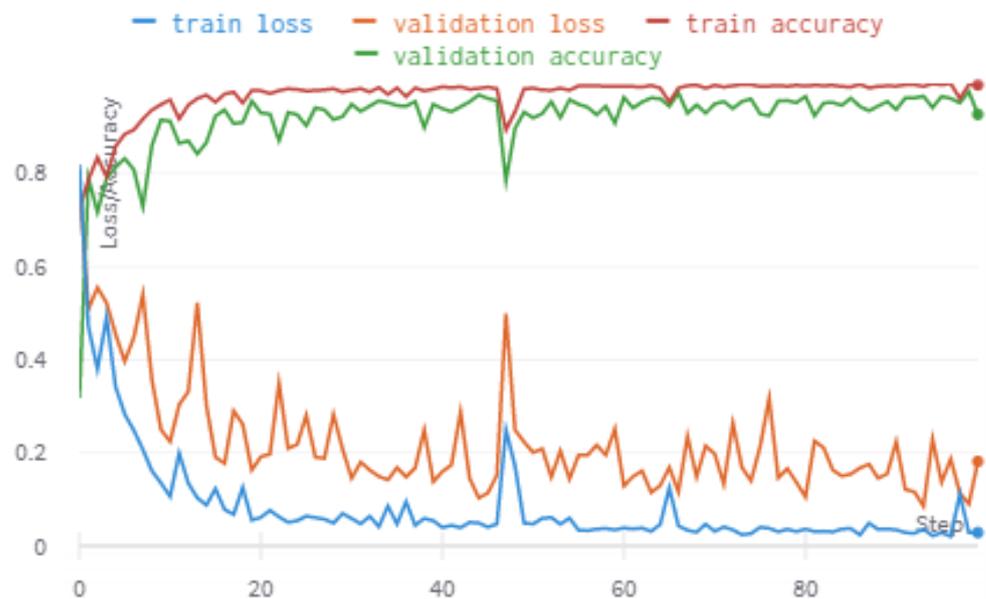


Figure 6. Training loss and accuracy using the data omitting batch 5, and validation loss and accuracy from the dataset containing only batch 5. Note that the loss is expectedly slightly higher for the validation dataset but nonetheless shows that the loss decreases and stabilises, indicating that the model has learned to extract generalisable features. Generated using “Weights and Biases” integration [27].

All preprocessing and training were performed on a workstation with an Intel I9 CPU, 64GB ram and an NVIDIA TitanX (Maxwell) GPU. Training time was approximately 5 h for a leave-one-out run (i.e., a run that omits a batch).

4. Results

We present our validation results in Table 2 in terms of precision (Equation (1)—what proportion of positive identifications were actually correct); recall (Equation (2)—what proportion of actual positives were identified correctly); and F_1 (Equation (3)—the harmonic mean of the two which provides an additional measure of the accuracy). The first column in Table 2, “Acc” presents the overall accuracy, i.e., the number of correct identifications out of the total number of images:

$$Precision = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \quad (1)$$

$$Recall = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \quad (2)$$

$$F_1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3)$$

We can see that the overall accuracy from the first row across all data is 99%. While this is an excellent result and indicates that there is some discernible difference between the two classes, it is still possible that the model is not actually relying on useful features, i.e., it could be using similarity between images or the environmental conditions particular to the days that the images were acquired.

To rule this out and test the generalisability of the model to new data, the leave-one-out paradigm was used as described in Section 3.3. Results for individual runs for this can also be seen in Table 2 and are similar but slightly lower in performance than the entire dataset. Nonetheless, considering that the models are being validated against completely unseen pigs with images captured on entirely different dates, these results are very encouraging. In comparison to the accuracy across all data of 99%, the leave-one-out models performed with a mean accuracy of 96%.

Table 2. Results for all runs where numbered rows represent the batch that was omitted in training and validated against (leave-one-out). “None” represents training on 90% of the dataset and validation against the remaining 10%. “Cumulative” represents the cumulative metrics for all leave-one-out (i.e., numbered) rows.

Omitted	Unstressed					Stressed			
	Acc.	Prec.	Recall	F1	N	Prec.	Recall	F1	N
None	0.99	0.99	0.98	0.98	4663	0.99	1.00	0.99	13,498
1	0.98	0.97	0.94	0.95	1980	0.98	0.99	0.98	5957
2	0.94	0.91	0.83	0.87	1222	0.94	0.97	0.96	3588
3	0.96	0.92	0.93	0.92	443	0.98	0.97	0.97	1282
4	0.91	0.88	0.86	0.87	672	0.92	0.93	0.93	1170
5	0.98	0.96	0.91	0.93	346	0.98	0.99	0.99	1501
Cumulative	0.96	0.93	0.90	0.91	4663	0.96	0.98	0.97	13,498

Table 3 shows that we are able to accurately estimate whether a sow is in a stressed or unstressed state in over 90% of images for pigs that have never been seen by the model. This gives us some confidence that the model has determined features that are generalisable across pigs and it is not merely learning certain features relating to specific individuals. In [28], Selvaraju et al. present a method of producing a coarse localisation map for a given class that the network has been trained on (gradient-weighted class activated mapping, Grad-CAM). Essentially, this shows which regions of an input image activate the network for a given class. The results resemble heatmaps or thermographic images where blue areas represent regions that contain less discriminative information and red represents regions with highly discriminative information. Figure 7 shows the results of applying the Grad-CAM technique to highlight regions which are activated for a given image of a given class. Regardless of the condition, the Grad-CAM heatmaps appear to show that the main regions used are the eyes, ears, shoulders/top of legs, snout and forehead. In the last of the stressed images, it is possible to see that the Grad-CAM has highlighted a bruised region (below the ear), but has also highlighted other regions indicating that it is not solely relying on the presence/visibility of a bruise.

Table 3. Normalised confusion matrix for cumulative data.

		Predicted	
		Stressed	Unstressed
GT	Stressed	0.90	0.10
	Unstressed	0.02	0.98

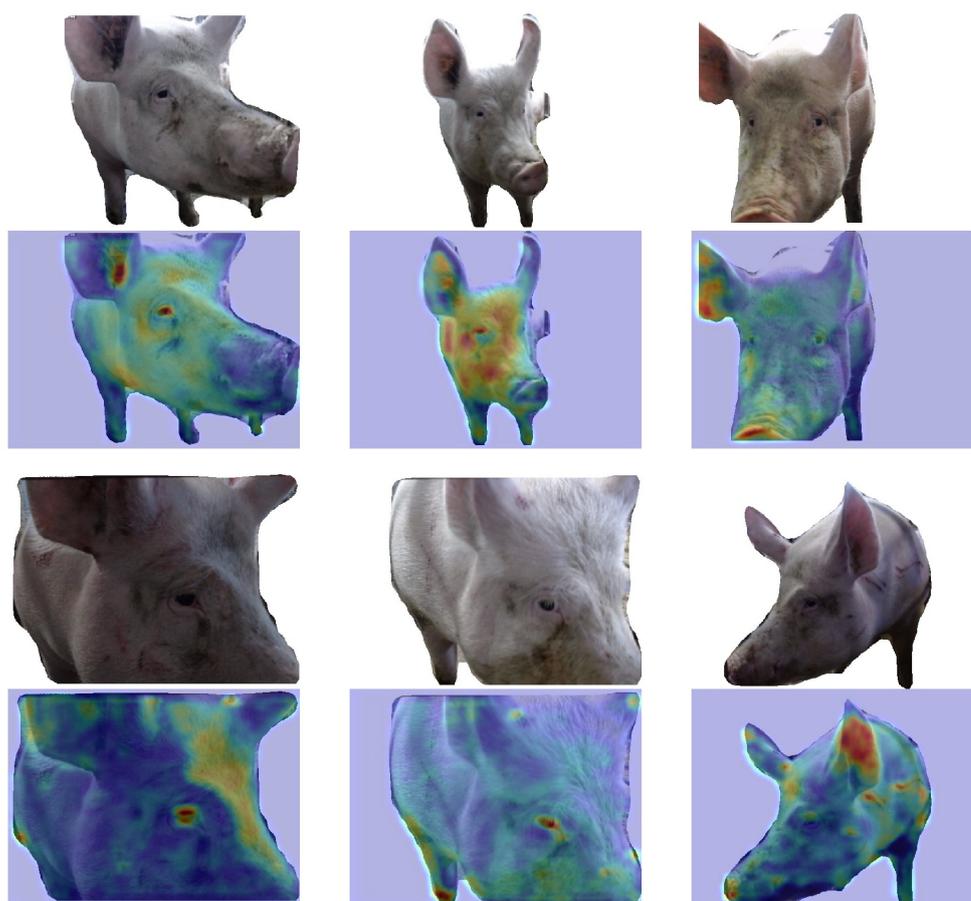


Figure 7. Six Grad-CAM heatmaps from correctly predicted conditions for unstressed (top three), and stressed pigs (bottom three). The input image is also included for easier comparison above each heatmap. Note the repeated use of the eye region, although it appears that other regions are also used. These images also show the range of lighting conditions that the model is able to cope with. NB red regions indicate a high level of activation; blue regions indicate a low level of activation.

5. Discussion

The results show that we were able to train a CNN to discriminate between images of pigs before and after they were exposed to stress. Remarkably, the network is able to generalise to pigs that it has never seen and is able to predict whether they are stressed or unstressed with $\sim 90\%$ accuracy.

Figure 7 shows representative examples of Grad-CAM output for correctly classified images. The highlighted regions are those which are most activated by the image for the given class. This shows that features such as eyes are heavily used for discriminating classes. There are other features that the model has also learnt to use, but these may not be as useful in terms of generalisability such as injuries on the animal (i.e., bruising as a result of sow-on-gilt aggression). Features such as bruising are less useful because whilst all animals that are bruised are likely to be experiencing (or have experienced) some form of negative affective state such as stress, not all animals that are stressed are bruised.

The eyes, the ears, forehead, snout and even legs/shoulders, all appear to be part of the overall information that the CNN is using. This supports previous research, such as the Pig Grimace Scale [12,13], which specifically analysed these regions (with the exception of the legs/shoulders) and the technique of qualitative behavioural assessment which uses whole animal body language to assess welfare [29]. However, out of all of the repeated regions that appear in the Grad-CAM images, the region surrounding the eye(s) is the most common. We therefore decided to see how much contribution the eyes alone make to the prediction accuracy. Using the regions that the eye detector found when cleaning the

data via the tiny-Yolo-v3 network and retraining the model using these as input (scaled to 32×32 px) gives the results shown in Table 4. These results show that the eyes, while not quite as accurate as the full image in most cases, very significantly contribute to the classifier. Interestingly, the only batch which performs better using eye region data compared to whole pig face is Batch 4, which performed the worst in terms of accurately analysing the full pig image (91%). While it is not clear what the cause of the increased performance on just the eye data is, Batch 4 does have amongst the fewest numbers of stressed animals (1170), as shown in Table 1. Manually inspecting the images shows nothing unusual in comparison with the other batches, with seemingly good variations in exposure, pose and positioning.

Grad-CAM results applied to the eye regions alone can be seen in Figure 8 and show that very similar regions are used in determining whether the model classifies the image as being stressed or unstressed. Along with the region of the actual eye and eyelids, the region below the tear-ducts seems to feature predominantly. This may be due to the presence of tear staining, which has been suggested as an indicator of negative welfare in pigs [30] but so far has not been validated as an indicator of stress [31].

Table 4. Table comparing the accuracy of the whole pig image (i.e., those used in the previous experiments) with only the eye regions to see how much contribution to prediction accuracy comes from such a small region. It can be seen that whilst not as accurate as using the full image of the pig, the eyes do significantly contribute to being able to determine whether pigs are stressed or unstressed.

Omitted Batch	Full Pig Acc.	Eyes Only Acc.
None	0.99	0.98
Batch 1	0.98	0.92
Batch 2	0.94	0.90
Batch 3	0.96	0.95
Batch 4	0.91	0.94
Batch 5	0.98	0.95

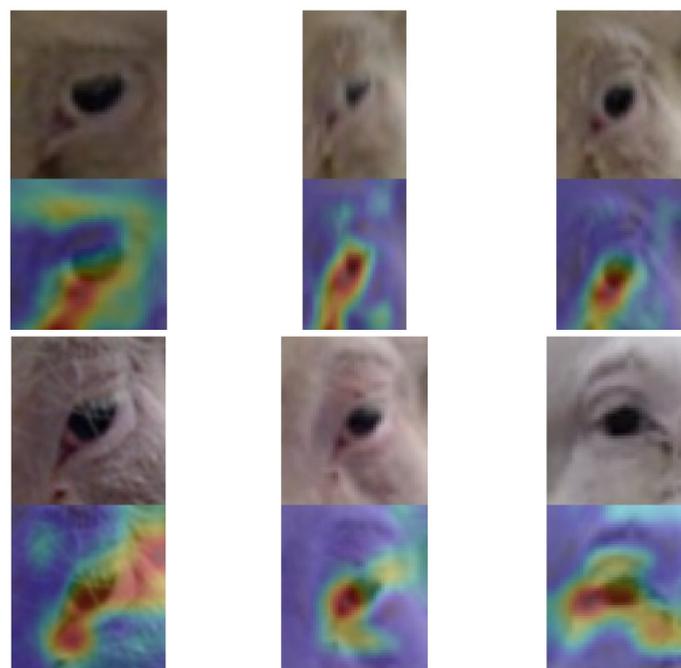


Figure 8. Six Grad-CAM heatmaps from correctly predicted conditions for unstressed (top three), and stressed pigs (bottom three). The input image is also included for easier comparison above each heatmap. The eye and eye-lid regions feature heavily as does the region below the tear-ducts, possibly indicative of using tear staining as a feature (although none of these example images show evidence of tear-staining).

The reason that the shoulders/upper legs appear so frequently in the Grad-CAM images is unknown. It is possible that they are a proxy to the position of the head, i.e., if the head is down, then less of the upper leg will be visible and vice versa. It may be that pigs experiencing low mood or that are socially subordinate exhibit a lowering of the head as many other animals do (e.g., horses [32], cows [33], humans [34]) and further work will seek to examine whether this is the case.

While the results are very promising, especially those on unseen pigs, they are probably insufficiently accurate to be used as a tool per se. The precision/recall rates are too low, indicating high numbers of false positives (~10%). There could be many reasons for this, but one of the most obvious is that if the model is learning facial features linked to stress, then these are likely not to be permanently present on the animal's face, but fleeting, indicating that a longer-term averaging across multiple images for a given animal may be helpful. The model we use forces a binary output, but we can amend this to give a probability of confidence score, so that we only make a judgement if the confidence is above a certain threshold. Figure 9 shows violin plots of the confidence score plotted against correctly and incorrectly classified images. It shows that when the model is correct, it is very certain, and predicts with a high confidence, but when it is incorrect, the model is very much less certain (mean confidence for correctly classified images are 98% and 99% for unstressed and stressed pigs, respectively, and for incorrectly classified images are 84% and 86%, respectively). This knowledge could be used to choose a threshold that would drastically reduce the false positive/negative rate and have very little impact on the accuracy. Another potential source of confusion could be that although the animal is assumed to be in a particular state, they may not be. This is especially true for the unstressed state, where the animal may be stressed for another reason that has not been accounted for. For example, although all animals were health checked prior to selection, it was not possible to discount an underlying, sub-clinical health condition that may affect their mood or a chronic social "condition" within their home pen (e.g., subordinate within their home pen prior to the mix).

As seen in [26], it is possible to use a similar system and architecture to identify a sow. Combining this functionality with the same hardware used in this experiment would create a machine vision system capable of detecting stress in individual animals that could then be identified. Future work will combine the systems and also seek to identify other emotional states such as "happiness" and pain that could be used as a means to further improve the animals' welfare.

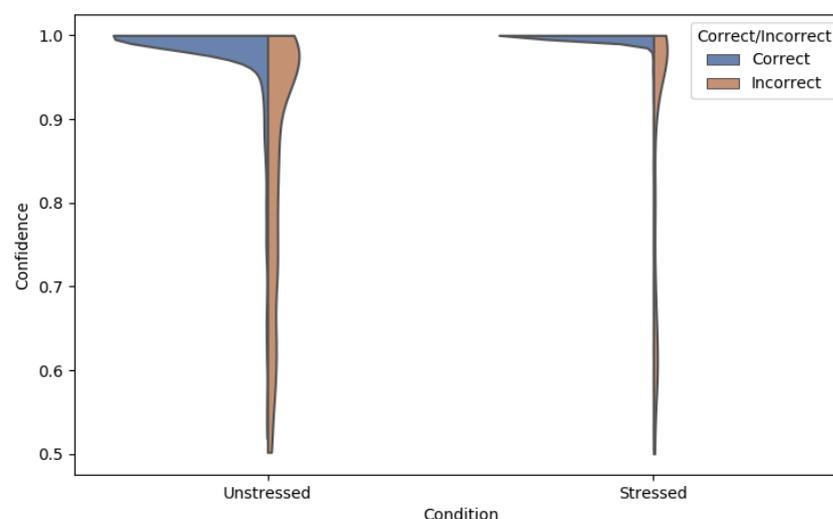


Figure 9. Violin plots showing the difference in distribution in confidence levels between correct and incorrect classifications. This indicates that the model gives far higher confidence scores when it is correct, meaning that it should be possible to set a suitable threshold to remove most misclassifications.

6. Conclusions

This paper has shown for the first time that a CNN is able to reliably distinguish whether a pig is stressed or unstressed in unseen animals using features extracted from a front view of the animal. The results show that the main regions involved in this classification match those commonly seen in the literature (such as eyes, ears, snout) and we show that the eyes regions alone significantly contribute to the overall accuracy of the system. Combining this work with biometrics could allow for the non-invasive monitoring of individuals, whereby farmers might be quickly alerted if an individual animal is showing signs of stress. We suggest that future work should analyse the regions in more detail in order to better understand the features used and how they fit with the existing literature as well as attempting to identify other expressions which may provide insights into pain and happiness as general indicators of an animal's welfare.

Author Contributions: Conceptualization, M.F.H., E.M.B., K.M.D.R., M.L.S. and L.N.S.; methodology, M.F.H., E.M.B., K.M.D.R., and M.L.S.; software, M.F.H., validation, M.F.H., E.M.B., K.M.D.R., M.L.S., formal analysis, M.F.H., E.M.B., K.M.D.R., M.L.S.; investigation, M.F.H., E.M.B., K.M.D.R., A.F.; resources, E.M.B., K.M.D.R., A.F.; data curation, E.M.B., K.M.D.R., A.F.; writing—original draft preparation, M.F.H.; writing—review and editing, M.F.H., E.M.B., K.M.D.R., M.L.S. and L.N.S.; visualization, M.F.H.; supervision, E.M.B., M.L.S.; project administration, E.M.B., M.L.S.; funding acquisition, M.F.H., E.M.B., K.M.D.R., M.L.S. and L.N.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Biotechnology and Biological Sciences Research Council, UK (Grant Reference: BB/S002138/1 and BB/S002294/1).

Institutional Review Board Statement: This study underwent internal ethical review by both SRUC's and UWE Bristol's Animal Welfare and Ethical Review Bodies (ED AE 16-2019 and R101) and was carried out under the UK Home Office license (P3850A80D).

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing not applicable.

Acknowledgments: This research was funded by the Biotechnology and Biological Sciences Research Council, UK (Grant Reference: BB/S002138/1 and BB/S002294/1). We also gratefully acknowledge the support of NVIDIA Corporation with the donation of the Maxwell Titan X GPU used for this research and are extremely grateful to farm and technical staff at SRUC's Pig Research Centre.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

- Alonso, M.E.; González-Montaña, J.R.; Lomillos, J.M. Consumers' Concerns and Perceptions of Farm Animal Welfare. *Animals* **2020**, *10*, 385. [\[CrossRef\]](#)
- Dawkins, M.S. Animal welfare and efficient farming: Is conflict inevitable? *Anim. Prod. Sci.* **2017**, *57*, 201–208. [\[CrossRef\]](#)
- Martínez-Miró, S.; Tecles, F.; Ramón, M.; Escribano, D.; Hernández, F.; Madrid, J.; Orengo, J.; Martínez-Subiela, S.; Manteca, X.; Cerón, J.J. Causes, consequences and biomarkers of stress in swine: An update. *BMC Vet. Res.* **2016**, *12*, 171. [\[CrossRef\]](#)
- Serpell, J.A. How happy is your pet? The problem of subjectivity in the assessment of companion animal welfare. *Anim. Welf.* **2019**, *28*, 57–66. [\[CrossRef\]](#)
- Tuytens, F.A.M.; de Graaf, S.; Heerkens, J.L.; Jacobs, L.; Nalon, E.; Ott, S.; Stadig, L.; Van Laer, E.; Ampe, B. Observer bias in animal behaviour research: Can we believe what we score, if we score what we believe? *Anim. Behav.* **2014**, *90*, 273–280. [\[CrossRef\]](#)
- Ekman, P. Universal facial expressions of emotions. *Calif. Ment. Health Res. Dig.* **1970**, *8*, 151–158.
- Kaya, H.; Gürpınar, F.; Salah, A.A. Video-based emotion recognition in the wild using deep transfer learning and score fusion. *Image Vis. Comput.* **2017**, *65*, 66–75. [\[CrossRef\]](#)
- Ekman, R. *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)*; Oxford University Press: Oxford, UK, 1997.
- Lien, J.J.; Kanade, T.; Cohn, J.F.; Li, C.C. Automated facial expression recognition based on FACS action units. In Proceedings of the Third IEEE International Conference on Automatic Face and Gesture Recognition, Nara, Japan, 14–16 April 1998; pp. 390–395.

10. Waller, B.M.; Julle-Daniere, E.; Micheletta, J. Measuring the evolution of facial ‘expression’ using multi-species FACS. *Neurosci. Biobehav. Rev.* **2020**, *113*, 1–11. [[CrossRef](#)]
11. Camerlink, I.; Coulange, E.; Farish, M.; Baxter, E.M.; Turner, S.P. Facial expression as a potential measure of both intent and emotion. *Sci. Rep.* **2018**, *8*, 17602. [[CrossRef](#)] [[PubMed](#)]
12. Vullo, C.; Barbieri, S.; Catone, G.; Graic, J.M.; Magaletti, M.; Di Rosa, A.; Motta, A.; Tremolada, C.; Canali, E.; Dalla Costa, E. Is the Piglet Grimace Scale (PGS) a Useful Welfare Indicator to Assess Pain after Cryptorchidectomy in Growing Pigs? *Animals* **2020**, *10*, 412. [[CrossRef](#)] [[PubMed](#)]
13. Di Giminiani, P.; Brierley, V.L.M.H.; Scollo, A.; Gottardo, F.; Malcolm, E.M.; Edwards, S.A.; Leach, M.C. The Assessment of Facial Expressions in Piglets Undergoing Tail Docking and Castration: Toward the Development of the Piglet Grimace Scale. *Front. Vet. Sci.* **2016**, *3*, 100. [[CrossRef](#)]
14. Koolhaas, J.M.; Bartolomucci, A.; Buwalda, B.; de Boer, S.F.; Flügge, G.; Korte, S.M.; Meerlo, P.; Murison, R.; Olivier, B.; Palanza, P.; et al. Stress revisited: A critical evaluation of the stress concept. *Neurosci. Biobehav. Rev.* **2011**, *35*, 1291–1301. [[CrossRef](#)] [[PubMed](#)]
15. Moberg, G.P.; Mench, J.A. *The Biology of Animal Stress: Basic Principles and Implications for Animal Welfare*; CABI: Oxon, UK, 2000.
16. Cook, N.J. Minimally invasive sampling media and the measurement of corticosteroids as biomarkers of stress in animals. *Can. J. Anim. Sci.* **2012**, *92*, 227–259. [[CrossRef](#)]
17. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. *arXiv* **2017**, arXiv:1703.06870.
18. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
19. Ison, S.H.; Donald, R.D.; Jarvis, S.; Robson, S.K.; Lawrence, A.B.; Rutherford, K.M.D. Behavioral and physiological responses of primiparous sows to mixing with older, unfamiliar sows. *J. Anim. Sci.* **2014**, *92*, 1647–1655. [[CrossRef](#)]
20. Jarvis, S.; Moinard, C.; Robson, S.K.; Baxter, E.; Ormandy, E.; Douglas, A.J.; Seckl, J.R.; Russell, J.A.; Lawrence, A.B. Programming the offspring of the pig by prenatal social stress: Neuroendocrine activity and behaviour. *Horm. Behav.* **2006**, *49*, 68–80. [[CrossRef](#)]
21. Ison, S.H.; D’Eath, R.B.; Robson, S.K.; Baxter, E.M.; Ormandy, E.; Douglas, A.J.; Russell, J.A.; Lawrence, A.B.; Jarvis, S. ‘Subordination style’ in pigs? The response of pregnant sows to mixing stress affects their offspring’s behaviour and stress reactivity. *Appl. Anim. Behav. Sci.* **2010**, *124*, 16–27. [[CrossRef](#)]
22. Rutherford, K.; Donald, R.; Arnott, G.; Rooke, J.; Dixon, L.; Mehers, J.; Turnbull, J.; Lawrence, A. Farm animal welfare: Assessing risks attributable to the prenatal environment. *Anim. Welf.* **2012**, *21*, 419–429. [[CrossRef](#)]
23. Rutherford, K.M.D.; Piastowska-Ciesielska, A.; Donald, R.D.; Robson, S.K.; Ison, S.H.; Jarvis, S.; Brunton, P.J.; Russell, J.A.; Lawrence, A.B. Prenatal stress produces anxiety prone female offspring and impaired maternal behaviour in the domestic pig. *Physiol. Behav.* **2014**, *129*, 255–264. [[CrossRef](#)] [[PubMed](#)]
24. Otten, W.; Kanitz, E.; Tuchscherer, M. The impact of pre-natal stress on offspring development in pigs. *J. Agric. Sci.* **2015**, *153*, 907–919. [[CrossRef](#)]
25. Lin, T.Y.; Maire, M.; Belongie, S.; Bourdev, L.; Girshick, R.; Hays, J.; Perona, P.; Ramanan, D.; Zitnick, C.L.; Dollár, P. Microsoft COCO: Common Objects in Context. *arXiv* **2015**, arXiv:1405.0312.
26. Hansen, M.F.; Smith, M.L.; Smith, L.N.; Salter, M.G.; Baxter, E.M.; Farish, M.; Grieve, B. Towards on-farm pig face recognition using convolutional neural networks. *Comput. Ind.* **2018**, *98*, 145–152. [[CrossRef](#)]
27. Biewald, L. Experiment Tracking with Weights and Biases, 2020. Available online: wandb.com (accessed on 10 June 2020).
28. Selvaraju, R.; Das, A.; Vedantam, R.; Cogswell, M.; Parikh, D.; Batra, D. Grad-CAM: Why Did You Say That? *arXiv* **2016**, arXiv:1611.07450.
29. Wemelsfelder, F.; Hunter, E.A.; Mendl, M.T.; Lawrence, A.B. Assessing the ‘whole animal’: A free-choice-profiling approach. *Anim. Behav.* **2001**, *62*, 209–220. [[CrossRef](#)]
30. Telkänranta, H.; Marchant-Forde, J.N.; Valros, A. Tear staining in pigs: A potential tool for welfare assessment on commercial farms. *Animal* **2016**, *10*, 318–325. [[CrossRef](#)] [[PubMed](#)]
31. Larsen, M.L.V.; Gustafsson, A.; Marchant-Forde, J.N.; Valros, A. Tear staining in finisher pigs and its relation to age, growth, sex and potential pen level stressors. *Animal* **2019**, *13*, 1704–1711. [[CrossRef](#)]
32. Fureix, C.; Jégo, P.; Henry, S.; Lansade, L.; Hausberger, M. Towards an Ethological Animal Model of Depression? A Study on Horses. *PLoS ONE* **2012**, *7*, e39280. [[CrossRef](#)] [[PubMed](#)]
33. de Oliveira, D.; Keeling, L.J. Routine activities and emotion in the life of dairy cows: Integrating body language into an affective state framework. *PLoS ONE* **2018**, *13*, e0195674. [[CrossRef](#)] [[PubMed](#)]
34. Veenstra, L.; Schneider, I.K.; Koole, S.L. Embodied mood regulation: The impact of body posture on mood recovery, negative thoughts, and mood-congruent recall. *Cogn. Emot.* **2017**, *31*, 1361–1376. [[CrossRef](#)]