Article

Towards facial expression recognition for on-farm welfare assessment in pigs

Mark F. Hansen [1,*], Emma M. Baxter [2], Kenny M.D. Rutherford[2], Agnieszka Futro[2,r], Melvyn L. Smith[1] and Lyndon N. Smith, [1]

[1] Centre for Machine Vision, BRL, UWE Bristol, Bristol BS16 1QY, UK; e-mail@e-mail.com
[2] Animal Behaviour and Welfare, Animal and Veterinary Sciences Research Group, SRUC, West Mains Road, Edinburgh EH9 3JG, UK; e-mail@e-mail.com
* 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 is 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 train a Convolutional Neural Network (CNN) using a leave-one-out design and show that it is able to discriminate between stressed and unstressed pigs with an accuracy of >90% in unseen animals. Grad-CAM is used to identify the animal regions used, and these support 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 a view to the development of 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

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 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 observation of the animal's own expression might itself offer insight into its emotional state. This would allow more appropriate and targeted management of individuals, reducing veterinary costs, improving farm productivity, and greatly enhancing the welfare of individual animals. The aim of this paper is to show that we are able to accurately classify adult female pigs as either stressed or unstressed using a Convolutional Neural Network (CNN) under real-world farm conditions.

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

The 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 is 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.

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. *Preprints* **2021**, *1*, 0. https://doi.org/

Received:
Accepted:
Published:

 doi:10.20944/preprints202108.0405.v1

1.1. Contributions

1. A first attempt to automate stress detection using face 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 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], aimed an 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 (FAU), 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). 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/FAU, 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 much harder 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 activation of the hypothalamic pituitary adrenocortical (HPA) axis and therefore is commonly assessed physiologically 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 stress (e.g. pain, fear, social stress, etc.) being experienced.

However, neither assessment of glucocorticoid levels nor detailed behavioural appraisal are very suitable for practical on–farm application. Whilst measuring cortisol via blood sampling is still considered widely 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. 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. 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, detailed assessment of behaviour is not feasible due to time constraints. As a result, new approaches that allow for 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 the 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 use 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 aim to

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 August 2021 doi:10.20944/preprints202108.0405.v1

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 was collected, how it was pre-processed, 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. 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. Mixing of gilts was closely supervised and specific end-points 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 to mix in the final third of pregnancy (but not within three weeks of predicted parturition date) to reduce the risk of harm to fetal 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 UK Home Office license (P3850A80D).

3.1.2. Animals and housing

Eighteen primiparous sows (hereafter gilts - Large White x Landrace x Duroc – "WhiteRoc" – Rattlerow Farms Ltd, Sufflok, 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 (2.5m long), a dunging passage (2.35m long, equipped with a drinker allowing ad libitum access to fresh water), and 6 individual feeding stalls (1.85m long, 0.5m wide). A standard ration (2.5-3.0kg per sow depending on body condition) of commercial concentrate feed for gestating sows was provided once a day for each pig (ForFarmers Nova UltraGest).

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 were used to capture images (max resolution: 1080 p/30 fps - 720 p/30 fps) 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 motion-detection capture of the pigs each time they voluntarily entered their individual feeding stalls.

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 identify pigs correctly 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 24h (i.e. "Unstressed" images) on the Monday. In order to establish a "Stressed" state, older multiparous sows were selected from the breeding herd to

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 August 2021 doi:10.20944/preprints202108.0405.v1

Figure 1. Image capture set-up. Six Logitech 920 webcams positioned in front of individual feeding stalls. Image captured via iSpy Connect software installed on Dell Precision computers.

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. On the day of the mix (Tuesday – MIX day) the gilts were mixed into the sows' new home pen (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.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 August 2021 doi:10.20944/preprints202108.0405.v1

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 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 pre-processing

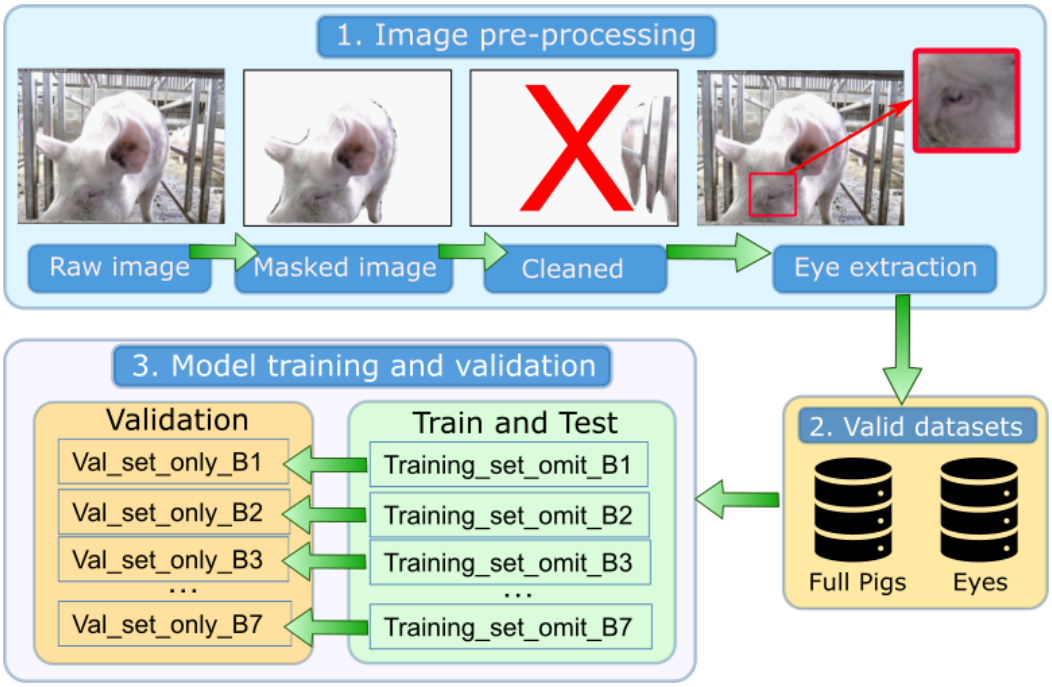


Figure 3. Pipeline of the process showing the pre-processing of the raw images to generate two datasets of valid images that are then used to create leave-one-out training and validation datasets where BN (e.g. B1, B2 etc.) represents the batch number.

Once the data had been collected and organised as detailed in Section 3.1, a number of image pre-processing 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).
3. 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 COCO (Common Objects in COntext) dataset [21] used for training the model, "pig" is not represented. However, there are 10 classes for living animals, and by lowering the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 August 2021 doi:10.20944/preprints202108.0405.v1

detection confidence to 0.5, the model was able to reliably detect and segment sows from their backgrounds.

With the backgrounds removed from the images, we are left with a further two 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 20×20px 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. Images in which no eyes were detected could then be 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 image numbers between "Stressed" and "Unstressed" (typically ∼2:1), 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 improve results further.

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

3.3. 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 architecture chosen is very much based on the model successfully used for biometric pig face recognition in [22] and consists of six convolutional blocks comprised of convolution layers with ReLU activation and then alternating max-pooling (2×2) and dropout (20%) layers. The 256×256px 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 "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 have not experimented with optimising hyper-parameters, so it is likely that further efficiency and accuracy improvements can be made.

Various batch sizes were explored, and 80 chosen as optimal for all experiments. 100 epochs, an ADAM optimizer and a learning rate of 0.001 were used.

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

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 August 2021 doi:10.20944/preprints202108.0405.v1

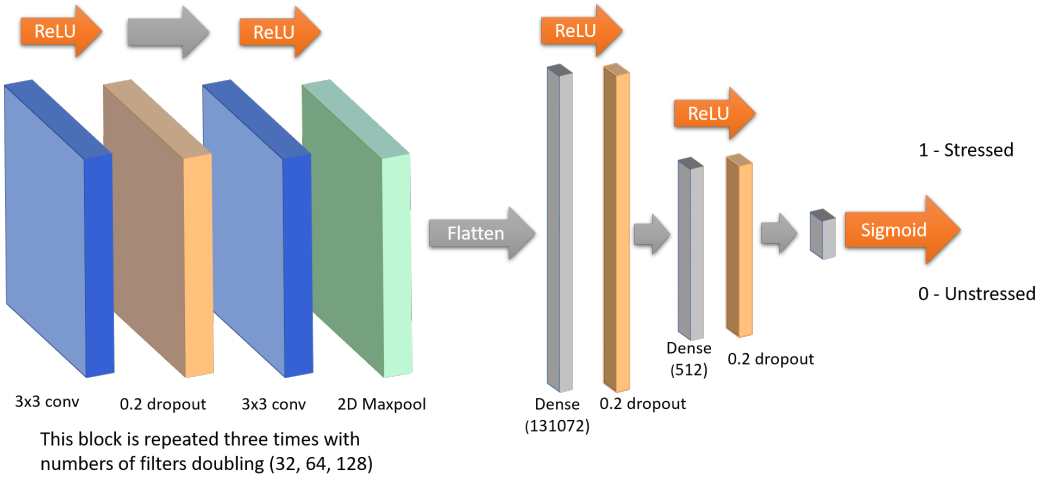


Figure 4. Architecture of our CNN consisting of three repeating blocks of a convolutional layer, max-pooling layer, convolutional layer and dropout layer. The output consists of one node with a sigmoid activation function representing the two classes - stressed and unstressed.

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 generalisablity of the model can be incorrect. In [22], 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 have selected here to use a leave-one-out cross-validation approach, as there is 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 a 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 a good correlation between 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 is considerably different to the training set (different pigs on different days), and the validation data has 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 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

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 August 2021 doi:10.20944/preprints202108.0405.v1

generalisable to an unseen dataset. The fact that the validation accuracy is very similar to the training accuracy is also very encouraging.

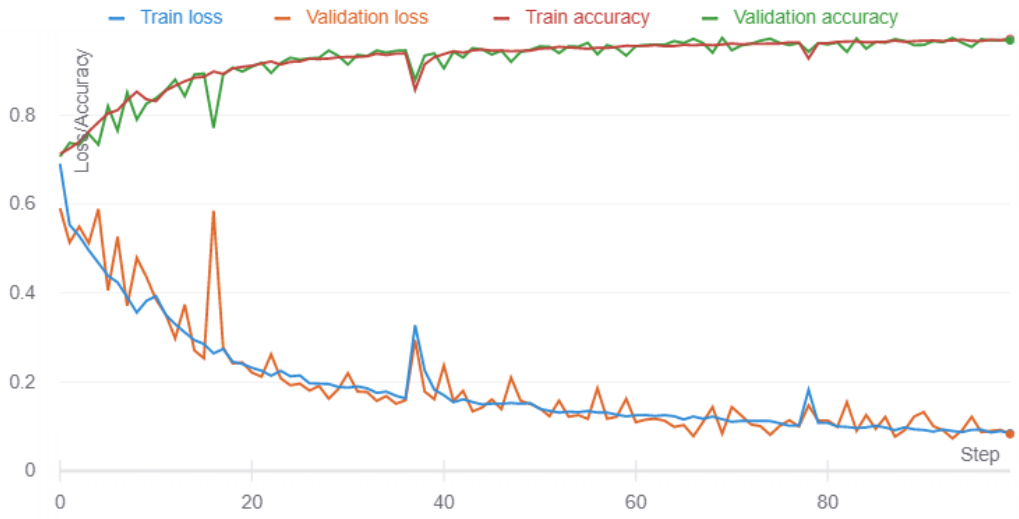


Figure 5. Training, validation loss and accuracy from the dataset containing all the data. This pattern is representative of all leave-one-out batch training and indicates that the model has not overfit the training set. Generated by [23].

All preprocessing and training was performed on a workstation with an Intel I9 CPU, 64GB ram and an NVIDIA TitanX (Maxwell) GPU.

4. Results

We present our validation results in Table 2 in terms of precision (Eq.1 – what proportion of positive identifications were actually correct), recall (Eq.2 – what proportion of actual positives were identified correctly) and F_1 (Eq.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.

$$Precision = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \tag{1}$$

$$Recall = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \tag{2}$$

$$F_1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3}$$

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. 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 perform with a mean 96% accuracy.

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 individuals. In [24], Selvaraju *et*

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 August 2021 doi:10.20944/preprints202108.0405.v1

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 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 by [23].

al. present a method of producing a course localization 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 are activating the network for a given class. Fig 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, but has also highlighted other regions indicating that it is not solely relying on the presence/visibility of a bruise.

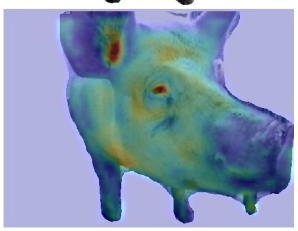
5. Discussion

The results show that we are able to train a CNN to discriminate between images of pigs before and after they have been 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 ~90% accuracy.

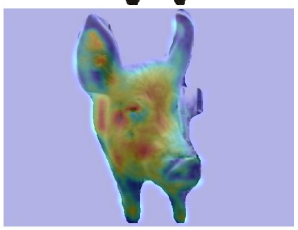
Figure 7 shows representative examples of Grad-CAM output for correctly classified images. The highlighted regions are those which are most activated by the image for the given class. It shows that features such as eyes are heavily used for discriminating classes. There are other features that the model 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 analyse 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 [25]. However, out of all of the repeated regions that appear in the Grad-CAM images, the region surrounding the eye(s) is most common. We

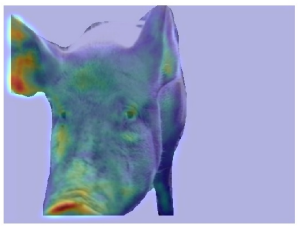
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 August 2021 doi:10.20944/preprints202108.0405.v1



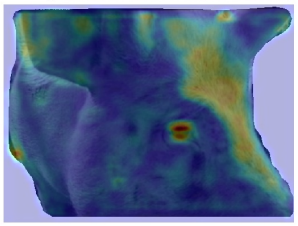
2 2019-09-30$_1$3 − 08 − 44$_6$11$_h$*eatmap* − *Copy.jpg* 2



2019-09-30$_1$5 − 50 − 51$_1$15$_h$*eatmap* − *Copy.jpg*



32019-09-30$_0$9 − 23 − 59$_5$11$_h$*eatmap* − *Copy.jpg*



3 2019-10-02$_1$0 − 39 − 38$_8$60$_h$*eatmap* − *Copy.jpg* 4

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 August 2021 doi:10.20944/preprints202108.0405.v1

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.

		Unstressed				Stressed			
Omitted	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	13498
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	13498

Table 3: Normalised confusion matrix for cumulative data.

		Predicted	
		Stressed	Unstressed
GT	Stressed	0.90	0.10
	Unstressed	0.02	0.98

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, retraining the model using these as input (scaled to 32×32px) 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, contribute very significantly to the classifier. 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 [26] but so far has not been validated as an indicator of stress [27].

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 contribute significantly 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	0.91	0.95
Batch 4	0.96	0.94
Batch 5	0.98	0.95

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[28], cows [29], humans [30]) and further work will look 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 are 98% and 99% for unstressed and stressed pigs, and for incorrectly classified are 84% and 86%). 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 [22], 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.

6. Conclusion

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 contribute significantly to the overall accuracy of the system. Combining this work with biometrics could allow for non-invasive monitoring of individuals, whereby farmers might be alerted quickly if an individual animal is showing signs of stress. We suggest 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.

7. Acknowledgements

This research has been 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.

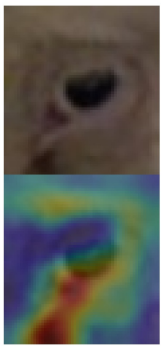
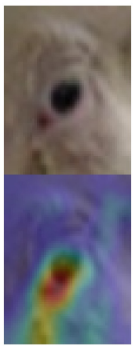
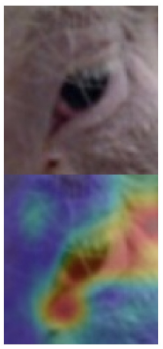
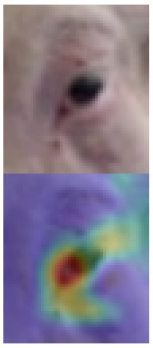
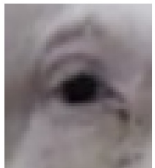
References

1. Alonso, M.E.; González-Montaña, J.R.; Lomillos, J.M. Consumers' Concerns and Perceptions of Farm Animal Welfare. *Animals* **2020**, *10*, 385. Number: 3 Publisher: Multidisciplinary Digital Publishing Institute, doi:10.3390/ani10030385.
2. Dawkins, M.S. Animal welfare and efficient farming: is conflict inevitable? *Animal Production Science* **2017**, *57*, 201–208. Publisher: CSIRO PUBLISHING, doi:10.1071/AN15383.
3. 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 Veterinary Research* **2016**, *12*, 171. Publisher: Springer.

4. Serpell, J.A. How happy is your pet? The problem of subjectivity in the assessment of companion animal welfare. *Animal Welfare* **2019**, *28*, 57–66. Publisher: Universities Federation for Animal Welfare.
5. Tuyttens, 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? *Animal Behaviour* **2014**, *90*, 273–280. Publisher: Elsevier.
6. Ekman, P. Universal facial expressions of emotions. *California mental health research digest* **1970**, *8*, 151–158.
7. Kaya, H.; Gürpınar, F.; Salah, A.A. Video-based emotion recognition in the wild using deep transfer learning and score fusion. *Image and Vision Computing* **2017**, *65*, 66–75. doi: 10.1016/j.imavis.2017.01.012.
8. Ekman, R. *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*; Oxford University Press, USA, 1997.
9. Lien, J.J.; Kanade, T.; Cohn, J.F.; Li, C.C. Automated facial expression recognition based on FACS action units. Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition. IEEE, 1998, pp. 390–395.
10. Waller, B.M.; Julle-Daniere, E.; Micheletta, J. Measuring the evolution of facial 'expression' using multi-species FACS. *Neuroscience & Biobehavioral Reviews* **2020**, *113*, 1–11. doi: 10.1016/j.neubiorev.2020.02.031.
11. Camerlink, I.; Coulange, E.; Farish, M.; Baxter, E.M.; Turner, S.P. Facial expression as a potential measure of both intent and emotion. *Scientific Reports* **2018**, *8*, 17602. Number: 1 Publisher: Nature Publishing Group, doi:10.1038/s41598-018-35905-3.
12. Vullo, C.; Barbieri, S.; Catone, G.; Graïc, 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. Number: 3 Publisher: Multidisciplinary Digital Publishing Institute, doi:10.3390/ani10030412.
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. *Frontiers in Veterinary Science* **2016**, *3*. Publisher: Frontiers, doi:10.3389/fvets.2016.00100.
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.; Richter-Levin, G.; Sgoifo, A.; Steimer, T.; Stiedl, O.; van Dijk, G.; Wöhr, M.; Fuchs, E. Stress revisited: A critical evaluation of the stress concept. *Neuroscience & Biobehavioral Reviews* **2011**, *35*, 1291–1301. doi:10.1016/j.neubiorev.2011.02.003.
15. Moberg, G.P.; Mench, J.A. *The biology of animal stress: basic principles and implications for animal welfare*; CABI, 2000.
16. Cook, N.J. Minimally invasive sampling media and the measurement of corticosteroids as biomarkers of stress in animals. *Canadian Journal of Animal Science* **2012**, *92*, 227–259. Publisher: NRC Research Press.
17. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. *arXiv:1703.06870 [cs]* **2017**. arXiv: 1703.06870.
18. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767* **2018**.
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. *Journal of Animal Science* **2014**, *92*, 1647–1655. Publisher: Oxford Academic, doi:10.2527/jas.2013-6447.
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. *Hormones and Behavior* **2006**, *49*, 68–80. Publisher: Elsevier.
21. 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:1405.0312 [cs]* **2015**. arXiv: 1405.0312.
22. 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. *Computers in Industry* **2018**, *98*, 145–152. doi:10.1016/j.compind.2018.02.016.
23. Biewald, L. Experiment Tracking with Weights and Biases, 2020. Software available from wandb.com.

24. Selvaraju, R.; Das, A.; Vedantam, R.; Cogswell, M.; Parikh, D.; Batra, D. Grad-CAM: Why Did You Say That? *arXiv preprint arXiv:1611.07450* **2016**.
25. Wemelsfelder, F.; Hunter, E.A.; Mendl, M.T.; Lawrence, A.B. Assessing the'whole animal': a free-choice-profiling approach. *Animal Behaviour* **2001**, *62*, 209–220. Publisher: Elsevier.
26. 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. Publisher: Cambridge University Press.
27. 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. Publisher: Cambridge University Press (CUP).
28. Fureix, C.; Jego, P.; Henry, S.; Lansade, L.; Hausberger, M. Towards an ethological animal model of depression? A study on horses. *PloS one* **2012**, *7*, e39280. Publisher: Public Library of Science.
29. Oliveira, D.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. Publisher: Public Library of Science, doi:10.1371/journal.pone.0195674.
30. 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. *Cognition and Emotion* **2017**, *31*, 1361–1376. Publisher: Taylor & Francis.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 August 2021 doi:10.20944/preprints202108.0405.v1

5	2020-01-13$_1$1 − 40 − 31$_8$80 − $Copy_batch8_eye0_heatmap.jpg$		5
	2020-01-13$_1$1 − 40 − 33$_7$48 − $Copy_batch8_eye1_heatmap.jpg$		5
	2020-01-13$_1$1 − 40 − 34$_9$51 − $Copy_batch8_eye1_heatmap.jpg$		
5	2020-01-15$_0$8 − 45 − 45$_5$14$_b$$atch8_eye0_heatmap - Copy.jpg$		5
	2020-01-15$_1$1 − 32 − 09$_2$22$_b$$atch8_eye1_heatmap - Copy.jpg$		4
			

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 19 August 2021 doi:10.20944/preprints202108.0405.v1

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.