Towards automatic monitoring of disease progression in sheep: A hierarchical model for sheep facial expressions analysis from video

Francisca Pessanha[1,3], Krista McLennan[2] and Marwa Mahmoud[3]
[1] Faculty of Engineering, University of Porto, Portugal
[2] Department of Biological Sciences, University of Chester, United Kingdom
[3] Department of Computer Science and Technology, University of Cambridge, United Kingdom

***Abstract*— Pain in farm animals harms the economics of farming and affects animal welfare. However, prey animals tend to not openly express signs of weakness, making the pain assessment process difficult. We propose a novel hierarchical model for disease progression evaluation, adapted for a wide range of head poses, according to which relevant information is extracted. A fine-tuned CNN is applied for face detection, followed by a CNN-based pose estimation and pose-informed landmark location method. Then multi-modal features are extracted, combining the appearance of regions-of-interest, described using a Histogram of Oriented Gradients, with geometric features and the pose values, leading to a binary Support Vector Machine classifier. To evaluate the efficiency of the complete pipeline, videos of the same sheep recorded at initial and advanced stages of treatment were tested, showing a decrease in the average pain score detected. The pain evaluation method significantly outperformed the existing state-of-the-art approach, being the first to apply a pose-based feature extraction in sheep pain detection.**

I. INTRODUCTION

Animal welfare and responsible farming have seen an increase in legislation in the past few years. Many diseases that affect animals are usually painful and cause distress. However, as prey species, sheep tend to not openly express signs of pain or weakness. The lack of human ability to recognise signs of pain is one of the most common causes of the untreated pain experienced by these animals [21], which is often associated with diseases such as footrot [5], mastitis [4] and pregnancy toxaemia [20]. The identification and quantification of pain are crucial for subsequent treatment and suffering relief [7].

Facial expressions have been used as an indicator of pain level in multiple species, including sheep [10], [22]. The Sheep Pain Facial Expression Scale (SPFES), introduced by McLennan *et al.* [22], provides a reliable tool for pain assessment associated with naturally occurring painful diseases in sheep. This scale analyses regions-of-interest, particularly the eyes, nose, ears, lips and cheeks, to determine levels of pain. It has shown high accuracy in identifying sheep pain and signs of illness. Studies also showed that changes in the facial expressions were detected after treatment. Nevertheless, the training of observers to manually assess of pain on large numbers of animals is very time-consuming, supporting a clear advantage in an automatic method.

We propose a hierarchical model for automatic pain assessment. The system is composed of a Convolutional Neural Network (CNN) - based face detection, followed by head pose estimation to assist the landmark localisation process. To automatically predict the pain level, multi-modal features are extracted then a single classifier is trained for pain prediction, thus this process removes the bias associated with individual action unit classifiers.

The main contributions of this paper can be summarised as follows:

- Proposing a robust sheep face detection model based on a fine-tuned SSD-Mobilenet networked trained on a varied dataset of sheep in farm and *in the wild* with varied head rotations. We also suggest a tracking algorithm that allows continuous detection and analysis of video.
- Introducing a pose informed automatic pain estimation method that adapts to different head rotation and consequent self-occlusion. When compared with previous work, the proposed method allows analysis of new facial regions that are only visible from profile viewpoints.
- Evaluating our proposed multi-step model on videos, thus making use of the temporal nature of facial expressions, in opposition to single image pain evaluation that has been the common evaluation method in previous state-of-the-art models. This improves the robustness of the system to momentary pain variations.
- Demonstrating the effectiveness of our global pain estimation model compared to models utilising regions-of-interest, hence, removing the bias introduced by annotations of different face areas.

To the best of our knowledge, this is the first complete pipeline for automatic analysis of animal disease progression in video.

II. RELATED WORK

Pain assessment in animals, based on their facial expressions, was firstly introduced by Langford *et al.* [16] describing a three-point scale for a set of relevant face features, such as orbital tightening, nose bulge, cheek bulge and ear position. This approach has been applied to multiple species, such as rats [25], rabbits [15], horses [3], [28] and more recently sheep [10], [22].

Manual scoring is still the usual practice for applying grimace scales, yet, it is very time-consuming and can introduce bias into the final score. A partially automated approach was proposed by Sotocinal *et al.* [25] aiming to extract *scoring-ready* images from videos of mice, replacing

the manually frame selection process. For this purpose, a Haar feature cascade classifier was used to detect the eye and ear, returning the frames where the key features were detected. Despite partially solving the labour-intensive problem of manual scoring, the pain assessment remains manual. This process was later automated [27] using a convolutional neural network based on the Inception V3 model, getting a greater proportion of images classified as "pain" following a laparotomy surgery when compared to sham surgery or post-surgical analgesic. This suggests that the proposed model provides an objective way to identify pain and pain relief in mice.

Previous work in sheep [19] showed the potential of an automatic pain assessment system, combining the pain prediction for several facial action units described in the SPFES. To detect the regions of interest, 8 facial landmarks were located using a modified version of Ensemble of Regression Trees (ERT) [14] with triplet interpolated feature (TIF) extraction [31]. However, the limited number of landmarks restricted the definition of the key facial areas and thus the pain estimation step was only defined for frontal faces. Further work on landmark detection was developed by Hewitt *et al.* [12] adding a pose estimation step to the pipeline and improving the landmark localisation for extreme poses. The main limitation found in previous work was the face detection step with both the Viola-Jones object detection framework [19], [29] and HOG-based face detection models [2], [12] proving to be insufficient to detect faces with the variety of head poses necessary.

III. DATA

In this section, we describe the dataset used in our work, according to the source and characteristics of the image. The dataset defined in [19] was augmented with new photographs leading to a final labelled set of 1306 images with bounding box annotations, with 1075 of them having face landmarks annotations. Additionally, 86 still images were taken from videos of sheep *in the wild* and fully annotated, including pain assessment following the SPFES. Moreover, 8 videos corresponding to an initial and an advanced stage of treatment of 4 different sheep were added for disease progression analysis.

A. Main Dataset and Annotations

The two subsets described by Mahmoud *et al.* [19] were revised and updated following the labelling criteria described in the present section. The two subsets include:

- **Sheep from a farm (SFF)**: 559 images taken on a similar farm setting.
- **Sheep from the internet (SFI)**: 98 images collected from the Internet.

Additionally, *in the wild* images with different head poses and scenarios were added to create a more representative set, as follows:

- **Sheep from WSID-100 (SFW)**: 239 images selected from the WSID-100 dataset [32] under the category sheep.

Fig. 1. Sample images from each subset showing the diversity in head pose and breed present in the complete dataset. From top to bottom, left to right: SFF, SFI, SFW, SFFli.

- **Sheep from Flickr (SFFli)**: 410 images extracted from Flickr under the tag "sheep".

In total, the main dataset includes 1306 images. SFF is composed of images with a consistent resolution (1333×1000 *px* or 755×1000 *px*) and background, a barn or fenced grassland. The diversity in breed and colour is limited and the illumination conditions are shared by a significant number of images. In contrast, the SFI, SFW, and SFFli showcase a mixture of sheep breeds, scenarios, and overall acquisition conditions. Regarding image resolution, all images of SFFli have 800 *px* on the longest edge while the SFW and SFI resolutions vary greatly. All subsets include a diverse number of sheep faces in each image - from cluttered images to frontal clean shots - as well as different head poses (see Fig. 1).

To guarantee the coherence of the annotations, a set of criteria was defined, with further revisions of previously made annotations [19]. Additionally, a qualitative pose (looking right, left or frontal) was assessed to all faces in the dataset.

For face bounding box annotations, the criteria proposed by Su *et al.* [26] was used as a starting point. Additionally, we defined what would be a suitable face detection for pain recognition according to the SPFES, guaranteeing that the key features were present.

The annotation criteria can be summarised as follows:

1) Regarding occlusion, only self-occlusion is accepted, not considering the examples where the face is occluded by external elements.
2) The different key features present must be distinguishable. Therefore, in photographs with a small depth of field, only the sheep near the focal plans are noted, as only on these cases it is possible to distinguish clearly the changes in the facial features. The same applies to faces that are very far from the camera.
3) Only head poses with a yaw angle of approximately 0 to 180º are considered. This excludes sheep looking backwards.
4) Lambs and goats are not labelled since the SPFES does not apply to them.

5) The face is defined from the chin level to the forehead level. If the forehead is not visible due to fur a consistent estimation of its position is made.

Faces were then annotated following the 25 landmarks scheme presented by Hewitt *et al.* [12]. Additionally, the occluded landmarks were labelled. The quantitative pose was calculated by transforming the mean face shape into the shape defined by the landmark, applying two approaches for solving the Perspective-n-Point problem: an iterative method based on Levenberg-Marquardt optimization and RANdom SAmple Consensus (RANSAC) algorithm by Fischler *et al.* [6]. The Mean Normalized Euclidean Error (MNE) between the transformed mean shape and the ground truth was calculated for both methods and the most accurate transformation was considered to be our head pose. The faces with an MNE value larger than 15% were considered invalid for the training of both pose estimator and the pose-informed landmark detector.

After the previously described data selection process, the landmarks and pose annotations subset included: 649 SFF faces, 123 SFI faces, 39 SFW faces, and 264 SFFli faces, with a total of 1075 faces.

From these, 879 correspond to poses between 0^o and 30^o, 107 to poses between 30^o and 60^o and 89 to poses between 60^o and 90^o.

B. Pain Dataset and Annotations

The pain score annotations were made by specialists in assessing facial expressions in sheep following the guidelines described by McLennan *et al.* [22].

The pain dataset contains a total of 86 images extracted from videos in a farm context (see Fig. 2), with annotations for the face and landmarks.

Each facial feature was scored individually from "0" ("pain not present") to "2" ("pain present"). The number of "2" annotations was small since it represents a higher state of pain and was always associated with pain in other facial features. Contemplating this fact, the sheep were considered in pain if the average score of the visible facial features was higher than 0.5, corresponding to signs of pain in at least two regions of interest.

Furthermore, a set of videos from 4 sheep in different stages of the disease, ranging from day 1 / 7 to day 42 of treatment, was selected to test the full pipeline. The data represents natural footage of sheep, often including extreme camera movements with the image following a specific sheep around the farm. The length of the videos is also variable, between 1 and 8 minutes.

IV. METHODOLOGY

In this section, we present the hierarchical model proposed (see Fig. 3). First, we explain how we detected the sheep faces and estimated the head pose. Then, we outlined how we used this information for facial landmark detection. Finally, we describe the normalisation and extraction of key facial features and how appearance and geometry features are used to predict pain.

Fig. 2. Sample images from the dataset collected for pain estimation. The upper row shows samples of "no pain" while the lower row shows samples of "pain". From left to right: 0 to 30 degrees; 30 to 60 degrees and 60 to 90 degrees head pose.

A. Face Detection: SSD-MobileNet

With the emergence of Convolutional Neural Networks (CNN), it is important to compare the performance of the traditional strategies applied in previous work [12], [19] and complex features based methods. Zhang *at al.* [33] studied the performance of classical detection algorithms of human faces, including Viola-Jones [29] cascade and a fine-tuned Faster R-CNN [9], for monkey face detection. Applying the previously described detectors to *real world* images the AdaBoost [8] resulted in a high number of false positives contrasting with the Faster R-CNN model, that displays a high Area Under the Curve (AUC) and detections for different head poses.

Following a similar line of thought, we propose fine-tuning a convolutional neural network model, giving preference to faster models, aiming at video applications. For this reason, a Single Shot MultiBox Detector (SSD) was used, introduced by Liu *et al.* [18], built on the MobileNets architecture [13], [17]. This method has proven to have comparable accuracy with slower object detection models such as Faster R-CNN and has shown higher accuracy than YOLO [23], the previous state-of-the-art for single-shot detectors.

B. Pose Estimation: Hopenet

Considering the satisfactory results obtained by Hewitt *et al.* [12], a Hopenet [24] was applied to the updated dataset, with 1075 faces. For this purpose, the faces were flipped according to the relative yaw, normalising the pose. Additionally, to improve the balance between the different head poses, a negatively correlated augmentation similar to the one proposed by Yang *et al.* [30] was performed. From the distribution of the absolute yaw angles, an augmentation factor for each pose bin was determined according to Equation 1 where $count_{max}$ is the maximum count for any pose bin and $count_b$ is the count for pose bin b. The level of boosting is controlled by parameter α with $0 \leq \alpha \leq 1$, in this case $\alpha = 0.6$. This method allowed to get a more balanced set of images, without losing the underlying distribution.

$$aug_b = \left[\left(\frac{count_{max}}{count_b} \right)^{\alpha} \right] \tag{1}$$

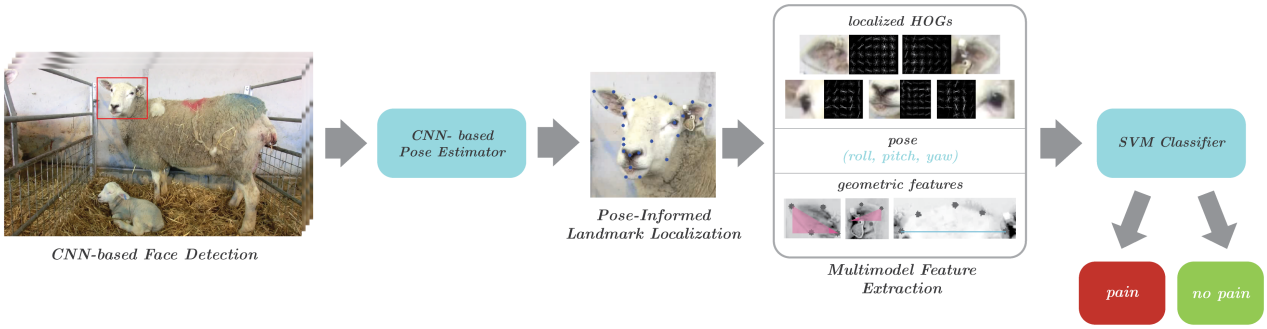

Fig. 3. The full pipeline of the proposed automatic approach for disease progression monitoring.

The augmentation algorithm applied to the images included rotation, flipping and thin-plate splines warping [1] generating slight variations from the input image, using the landmarks as a reference.

The multi-loss convolutional neural network was initialised using a pre-trained model on the 300W-LP [34], a large-pose human face images dataset.

C. Landmark Location: Ensemble of Regressions Trees

To optimise the landmark detection presented by Hewitt *et al.* [12], we propose setting three pose-informed Ensemble of Regressions Trees (PI-ERT) model [14] introducing the occlusion information for each range of poses to define the relevant landmarks for each model.

For this purpose, the faces - after pose normalisation - were divided into three bins according to their absolute yaw ($[0, 30]$; $[30, 60]$; $[60, 90]$ degrees). For each range, the landmarks that were occluded in more than 70% of the faces in a specific bin were excluded from the shape model.

D. Pain Estimation

To assess the pain based on facial expressions, the correlation between appearance/geometric features and the overall pain score was studied. As mentioned in section IV-C, the head rotation leads to self-occlusion of facial areas, therefore, for each yaw angle range, the visible facial features were defined. Accordingly, after pose normalisation, both ears and eyes, as well as the nose were defined visible for yaw values until 10 degrees. With further rotation until 60 degrees, the nose, left eye, left cheek, and both ears are visible, with the right eye occluded. Lastly, for yaw from 60 to 90 degrees, the visible facial features were the nose, left eye, left ear and left cheek, with the entire right side occluded (profile view).

Similar to what was proposed by Mahmoud *et al.* [19], Histogram of Oriented Gradients were used as an appearance feature descriptor for each facial area. This descriptor, defined by Dalal *et al.* [2], represents the distribution of local intensity gradients and edge directions, providing pertinent information regarding the shape/appearance of the object. As geometric features, the global angle of each ear, between its root and tip, and the distance between the ears roots were used. Additionally, considering the changes in the appearance of each area of interest with the point of view, the quantitative pose was also added to our feature set by concatenation.

The final feature vector was defined with the HOGs for the left and right ear, left and right eye, nose and left cheek, the geometric features, and the pose. A Support Vector Machine model, with a linear kernel, was then trained.

V. EXPERIMENTAL EVALUATION

To evaluate our models, a 5-fold cross-validation was used for all the experiments and the average value was reported. Each fold was balanced according to the classification label of interest in each section; For face detection, this was the qualitative pose of each face. For the pose detection, this was the quantitative pose. For the pain estimation, this was the pain score. When augmenting the data, the evaluation was always performed on an unaugmented test fold, not used for the training phase. Each step of the pipeline was trained and evaluated independently and then all combined in the disease progression evaluation step in the pain dataset.

A. Implementation Details

1) Sheep Face Detection: The larger side of each image was resized to 300 and the other was updated proportionally. After resizing, the bounding boxes with less than 4500 *px* of area were not considered viable for training purposes.

The SSD-Mobilenet network was initialised using a model pre-trained on the Common Objects in Context (COCO) dataset presented by Lin *et al.* [17]. The batch size used was 12 and, regarding the hyperparameters, the number of training steps (total number of training iterations) used was 10000, corresponding to 185 epochs, with 648 training images.

A detection was considered a true positive if the Intersection over Union (IoU) with a ground truth box was higher than 0.5. We only considered one true positive for each ground truth bounding box, considering the rest of the detections as false positives. Additionally, the detections with an IoU higher than 0.7 were replaced by their mean shape, since they likely refer to the same face.

For the full pipeline implementation, a tracking system using Kernelized Correlation Filters (KCF) [11] was introduced, only running the CNN-based face detector once every second.

2) Landmark Location: Different amount of perturbations were introduced on the training phase (30, 50, 70 and 90), with the final models being the ones were the performance *plateau*, with 50 perturbations for the bins [30,60] and [60, 90], with a smaller number of examples, and 30 perturbations for the bin [0, 30]. The error was normalised according to the mean edge length of the bounding box.

3) Pain Estimation: Firstly, all faces were flipped according to their relative yaw, thus normalising the pose. The facial areas were extracted and the ears and nose were normalised, rotating the ears horizontally in relation to the line defined by the ear root and tip, and rotating the nose vertically following the line defined by the nose tip and the middle of the mouth. The set dimensions for each facial feature were 100×80 *px* for the ears, 80×80 *px* for eyes and cheeks and 80×120 *px* for the nose. The hyperparameters of HOGs were defined experimentally through a nested cross-validation, according to the F1-score of the final score. The HOGs parameters used were 9 orientations, 16 pixels per cell and 2 cells per block. When a region of interest was not visible from a particular pose, it was replaced by a zeros matrix of the same size for the appearance feature assessment. For the geometric features, the missing parameters were considered as "-1". An SVM model with a linear kernel was trained and each feature was tested individually to verify their relevance and then they were all combined and evaluated.

4) Disease progression: For pain estimation algorithm, we used a leave-one-animal-out testing approach. In total, four models were trained, with each model removing one of the four sheep of interest from the training set. Considering the nature of the videos, with a moving "handheld" camera, the sheep face could be cropped or outside the frame. For this reason, only sections where the specific sheep is in focus were considered. All example video segments showed a diversity of head poses, including frontal and profile views. The full pipeline was applied every 10 frames and the average pain score in the video was then calculated.

B. Sheep Face Detection

The SSD-Mobilenet proved to be efficient in detecting sheep faces in different environments (see Fig. 4), with diverse characteristics both intrinsic (breed and pose) and extrinsic (illumination and scenario). In comparison with the models previously described, HOG-SVM [12] and a Viola-Jones based detector [19], the algorithm overcomes the lack of flexibility in regards to the head pose, recognising faces through a wide set of poses, a crucial feature for detection *in the wild*.

Our proposed model showed a precision of 94.17 %, recall of 94.02 % and an F1-score of 94.00 %. In contrast, the HOG-SVM model, trained on the same data was unable to detect the non-frontal faces, achieving a precision of 94.44% with the low recall of 8.73 %.

C. Pose Estimation

The metrics used to evaluate the pose results were the mean absolute error (MAE), the Pearson's Correlation Coefficient (PCC), that measures the correlation between the predictions and the ground truth, and the Sign Agreement metric, indicating if the prediction matches the general direction of the head.

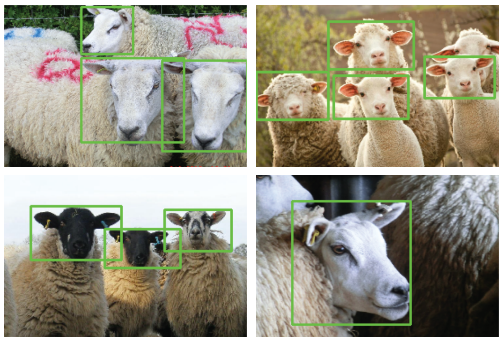

Fig. 4. Examples of the CNN-based face detection, displaying detections over a wide range of head poses on the SFF, SFI, SFW and SFFli datasets

TABLE I

POSE ESTIMATION RESULTS SHOWING RESULTS FOR PITCH, YEW AND ROLL WHEN APPLIED ON THE PROPOSED DATASET

	Yaw	***Pitch***	***Roll***	***Average***
MAE	10.38	9.42	6.78	8.86
PCC	0.82	0.42	0.43	0.59
SAGR	0.72	0.73	0.77	0.74

Considering the diversity of poses in the dataset, the results showed satisfactory values (Table I), especially for the yaw angle, with a high agreement in sign and a reasonable error for the 30 degrees bin division used in the pose-informed landmark location step. Compared with the values presented by Hewitt *et al.* [12], there is an increase of 2.28 degrees in the average MAE, which can be a consequence of the augmentation of the diversity of head poses in the dataset, in particular extreme head poses, that continue to be underrepresented after data augmentation.

D. Landmark Localisation

Our occlusion-informed method achieved a significantly higher success rate (SR) than the previous method, with an increase of 14% for faces with a head pose within 30 and 60 degrees and 21% for faces with a head pose higher than 90 degrees (Table II).

As expected the model generalised better for the 0 to 30 degrees bin, since the number of images, in this case, is 879, which is more than eight times higher than the other two: 107 and 89 for poses between 30 and 60 degrees and 60 and 90 degrees, respectively (see Fig. 5).

However, it is noticeable that the system does not generalise well for poses underrepresented in the dataset, for instance, with an extreme pitch angle, such as the one presented in the last column of Fig. 5. This suggests that, although there is an increase in the range of yaw angles represented in the dataset, it is still necessary to introduce more representation of head poses, with angle variations following all three directions, to get a more robust model.

TABLE II

LANDMARK LOCALISATION PERFORMANCE, FOR THE PI-ERT WITH ALL THE 25 LANDMARKS AND THE PROPOSED OCCLUSION-INFORMED PI-ERT (OPI-ERT)

	OPI-ERT			*PI-ERT*		
	MNE	**SR**	**AUC**	**MNE**	**SR**	**AUC**
0 - 30º	-	-	-	0.05	0.87	0.94
30 - 60º	0.10	0.64	0.90	0.11	0.50	0.89
60 - 90º	0.12	0.39	0.88	0.15	0.18	0.85

Fig. 5. Qualitative examples of the landmark location results for the occlusion-informed ERT. The right-most column shows an example where the method struggles, due to an extreme head pose. Rows (from top to bottom): ground-truth, standard PI-ERT; occlusion informed PI-ERT

E. Pain Estimation

Considering a majority vote accuracy baseline of 55%, all the three classes of features proposed showed a significant improvement over the baseline. Additionally, the combined model showed impressive results, with an F1-score of 73% (Table III). Comparing to the previous model defined by Mahamoud *et al.* [19], there is an accuracy increase of 11%, which is remarkable considering the diversity of the dataset used, with a range of different head poses in opposition to datasets with only frontal faces, which were used previously.

F. Disease progression

To evaluate the validity of the full pipeline proposed for disease progression analysis, the model was applied to videos from the same sheep at the beginning of the disease (first or seventh day) and after treatment (after 42 days). Two animals had mastitis while the other two had pregnancy toxaemia.

TABLE III

PAIN ESTIMATION RESULTS FOR THE SVM MODELS TRAINED WITH THE INDIVIDUAL FEATURES AND THE COMBINED FEATURE VECTOR WHEN APPLIED TO THE DATASET PROPOSED (MAJORITY VOTE: 55 %)

	Precision	***Recall***	***F1- score***	***Accuracy***
HOGs	0.74	0.66	0.69	0.74
Pose	0.76	0.54	0.63	0.71
Geometry	0.75	0.49	0.58	0.68
Combined	**0.83**	**0.68**	**0.73**	**0.78**

The pain score returned for each video was, on average, 0.89 in videos showing initial stages of the treatment. Then an average decrease of 0.30, was observed in the advanced treatment videos. This decline was detected in all the four examples consistently, leading us to argue that there is a clear relationship between the pain score returned and the illness progression.

The mean pain score after 42 days was 0.59, which was higher than what we expected considering the duration of the treatment by this point. However, we had a few possible explanations for that. Since the pain analysis is the last step of a hierarchical pipeline, preceded by face detection, pose estimation and landmark detection, the error obtained in each step will propagate and affect the final classification. That is normal and anticipated in any hierarchical model. Additionally, the facial pain score will not be constant, with natural fluctuations in the pain level and with sheep tending to mask their signs of pain when observed, which is an expected behaviour considering that in the videos used the camera operator is on the field, following the animal. Finally, external factors such as the wind and loud noises will have a noticeable effect on the ear position, making them go backwards in a similar manner to what happens in cases of pain.

VI. CONCLUSIONS AND FUTURE WORK

This paper introduced an optimised dataset for sheep face detection, pose estimation and landmark localisation, containing a total of 1306 images annotated for face detection and 1075 faces with both the bounding box annotation and landmark locations. Additionally, a set of 86 frames extracted from natural footage of sheep was described and used for pain estimation purposes.

The CNN-based face detection proved to be effective in detecting sheep faces through a variety of head poses, being adequate for applications *in the wild* and providing the possibility to extend the pipeline to accommodate profile faces. This model allowed the integration of the cheek region described in the SPFES scale, not seen from a frontal point of view.

The pose and landmark localisation models proposed in previous work [12] were extended based on the improved dataset defined in this paper. Three pose-informed ERT (PI-ERT) models were trained for landmark localisation, redefining their shape by removing the self-occluded landmarks for each pose's yaw range.

Finally, we propose a pain assessment model, based on pose-informed appearance and geometric features as well as the head pose. Our experiments showed an accuracy of 78% outperforming state-of-the-art models.

The performance of the hierarchical pipeline in predicting disease progression *in the wild* was evaluated using footage from the same sheep, with a total of four different animals, in an initial and advanced stage of the treatment, observing a correlation between the average pain score and the stage of the treatment in all four cases.

For future work, we recommend the implementation of a better video capturing system with a hidden camera, for example, that can be set close to a manger, to record the animals in an undisturbed manner when it is more likely for them to not hide signs of pain. Additionally, we would like to develop an interface and design user tests to better understand the needs of the farmers and consider possible points of improvement of our models based on their feedback.

REFERENCES

[1] F. L. Bookstein. Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Transactions on pattern analysis and machine intelligence*, 11(6):567–585, 1989.

[2] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, volume 1, pages 886–893. IEEE, 2005.

[3] E. Dalla Costa, M. Minero, D. Lebelt, D. Stucke, E. Canali, and M. C. Leach. Development of the Horse Grimace Scale (HGS) as a pain assessment tool in horses undergoing routine castration. *PLoS ONE*, 9(3):1–10, 2014.

[4] S. Dolan, L. Field, and A. Nolan. The role of nitric oxide and prostaglandin signaling pathways in spinal nociceptive processing in chronic inflammation. *Pain*, 86(3):311–320, 2000.

[5] S. Dolan, J. Kelly, A. Monteiro, and A. Nolan. Up-regulation of metabotropic glutamate receptor subtypes 3 and 5 in spinal cord in a clinical model of persistent inflammation and hyperalgesia. *Pain*, 106(3):501–512, 2003.

[6] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981.

[7] P. Flecknell. Analgesia from a veterinary perspective. *British Journal of Anaesthesia*, 101(1):121–124, 2008.

[8] Y. Freund, R. Schapire, and N. Abe. A short introduction to boosting. *Journal-Japanese Society For Artificial Intelligence*, 14(771-780):1612, 1999.

[9] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 580–587, 2014.

[10] C. Häger, S. Biernot, M. Buettner, S. Glage, L. Keubler, N. Held, E. Bleich, K. Otto, C. Müller, S. Decker, et al. The sheep grimace scale as an indicator of post-operative distress and pain in laboratory sheep. *PloS one*, 12(4), 2017.

[11] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista. High-speed tracking with kernelized correlation filters. *IEEE transactions on pattern analysis and machine intelligence*, 37(3):583–596, 2014.

[12] C. Hewitt and M. Mahmoud. Pose-informed face alignment for extreme head pose variations in animals. In *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*, pages 1–6. IEEE, 2019.

[13] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017.

[14] V. Kazemi and J. Sullivan. One millisecond face alignment with an ensemble of regression trees. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1867–1874, 2014.

[15] S. C. Keating, A. A. Thomas, P. A. Flecknell, and M. C. Leach. Evaluation of emla cream for preventing pain during tattooing of rabbits: changes in physiological, behavioural and facial expression responses. *PloS one*, 7(9), 2012.

[16] D. J. Langford, A. L. Bailey, M. L. Chanda, S. E. Clarke, T. E. Drummond, S. Echols, S. Glick, J. Ingrao, T. Klassen-Ross, M. L. LaCroix-Fralish, et al. Coding of facial expressions of pain in the laboratory mouse. *Nature methods*, 7(6):447, 2010.

[17] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014.

[18] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg. Ssd: Single shot multibox detector. In *European conference on computer vision*, pages 21–37. Springer, 2016.

[19] M. Mahmoud, Y. Lu, X. Hou, K. McLennan, and P. Robinson. Estimation of pain in sheep using computer vision. In *Handbook of Pain and Palliative Care*, pages 145–157. Springer, 2018.

[20] J. V. Marteniuk and T. H. Herdt. Pregnancy toxemia and ketosis of ewes and does. *Veterinary Clinics of North America: Food Animal Practice*, 4(2):307–315, 1988.

[21] K. M. McLennan. Why pain is still a welfare issue for farm animals, and how facial expression could be the answer. *Agriculture*, 8(8):127, 2018.

[22] K. M. McLennan, C. J. Rebelo, M. J. Corke, M. A. Holmes, M. C. Leach, and F. Constantino-Casas. Development of a facial expression scale using footrot and mastitis as models of pain in sheep. *Applied Animal Behaviour Science*, 176:19–26, 2016.

[23] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016.

[24] N. Ruiz, E. Chong, and J. M. Rehg. Fine-grained head pose estimation without keypoints. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 2074–2083, 2018.

[25] S. G. Sotocina, R. E. Sorge, A. Zaloum, A. H. Tuttle, L. J. Martin, J. S. Wieskopf, J. C. Mapplebeck, P. Wei, S. Zhan, S. Zhang, et al. The rat grimace scale: a partially automated method for quantifying pain in the laboratory rat via facial expressions. *Molecular pain*, 7:1744–8069, 2011.

[26] H. Su, J. Deng, and L. Fei-Fei. Crowdsourcing annotations for visual object detection. In *Workshops at the Twenty-Sixth AAAI Conference on Artificial Intelligence*, 2012.

[27] A. H. Tuttle, M. J. Molinaro, J. F. Jethwa, S. G. Sotocinal, J. C. Prieto, M. A. Styner, J. S. Mogil, and M. J. Zylka. A deep neural network to assess spontaneous pain from mouse facial expressions. *Molecular pain*, 14:1744806918763658, 2018.

[28] J. P. van Loon and M. C. Van Dierendonck. Monitoring equine head-related pain with the Equine Utrecht University scale for facial assessment of pain (EQUUS-FAP). *Veterinary Journal*, 220(January):88–90, 2017.

[29] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001*, volume 1, pages I–I. IEEE, 2001.

[30] H. Yang and X. A. Wang. Cascade classifier for face detection. *Journal of Algorithms & Computational Technology*, 10(3):187–197, 2016.

[31] H. Yang, R. Zhang, and P. Robinson. Human and sheep facial landmarks localisation by triplet interpolated features. In *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1–8. IEEE, 2016.

[32] Y. Yao, J. Zhang, F. Shen, L. Liu, F. Zhu, D. Zhang, and H. T. Shen. Towards automatic construction of diverse, high-quality image datasets. *IEEE Transactions on Knowledge and Data Engineering*, 2019.

[33] M. Zhang, S. Guo, and X. Xie. Towards automatic detection of monkey faces. In *2018 24th International Conference on Pattern Recognition (ICPR)*, pages 2564–2569. IEEE, 2018.

[34] X. Zhu, Z. Lei, X. Liu, H. Shi, and S. Z. Li. Face alignment across large poses: A 3d solution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 146–155, 2016.