AI for Equine Welfare: Detecting Ear Movements for Affective State Assessment

João Alves and **Rikke Gade** {jmal,rg}@create.aau.dk
Visual Analysis and Perception Lab, Aalborg University, Pioneer Centre for AI, DK
Pia Haubro Andersen pia.haubro.andersen@slu.se
Department of Veterinary Clinical Sciences, University of Copenhagen, DK

Abstract

The Equine Facial Action Coding System (EquiFACS) enables the systematic annotation of equine facial movements through distinct Action Units (AUs) and serves as a crucial tool for assessing affective states in horses by identifying subtle facial expressions associated with discomfort. In this work, we study different methods for specific ear AU detection and localization from horse videos. We achieve 87.5% classification accuracy of ear movement presence on a public horse video dataset, demonstrating the potential of our deep learning based approach. Our code is publicly available at https://github.com/jmalves5/read-my-ears.

Keywords: Animal Welfare, Equine Facial Action Unit Detection, Equine Affective State Assessment

1 Introduction

Horses play an important role in multiple areas of our society, and thus, we have societal responsibility to ensure their well-being. Horses express pain through subtle facial movements that are often overlooked by the untrained human eye, potentially leading to late diagnoses [Broomé et al., 2022].

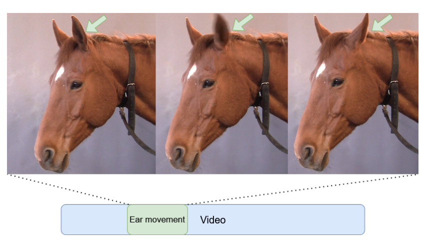


Figure 1: Ear rotator action unit (EAD104) example.

Works focusing on facial expressions of pain have mostly used the Facial Action Coding System (FACS) developed by [Ekman et al., 2002] in humans and the Equine Facial Action Coding System (EquiFACS) by [Wathan et al., 2015] in horses, which allows for an objective evaluation of the animal's facial movements, based on it's facial musculature (see Figure 2 and Figure 3). However, EquiFACS presents significant practical challenges. The subtlety of horse

AI in Science
Conference 2025
EU2025DK

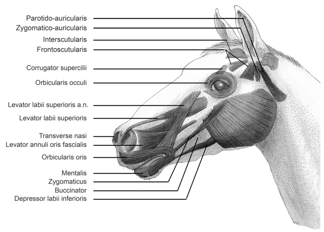


Figure 2: Horse facial muscles from [Wathan et al., 2015].

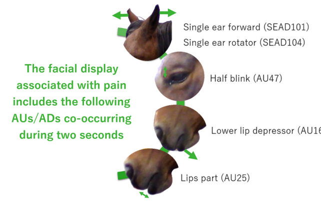


Figure 3: Example EquiFACS AUs from [Ask et al., 2024].

facial movements requires skilled expert annotation, a task that can take hours for even short video clips, increasing dataset cost [Andersen et al., 2021]. Ear-related AUs (see Figure 1) have been associated with equine affective states, like stress and pain, see [Dalla Costa et al., 2014, Rashid et al., 2020, Lundblad, 2024]. With that in mind, this work focuses on the video clip classification of ear-related movements from horse video data with the goal of preforming action unit detection for horse affective state assessment. We propose and study different methodologies to automate the extraction of these movements and evaluate our methods on a publicly available dataset from [Rashid et al., 2020]. Several works such as [Rashid et al., 2020] and [Ask et al., 2024] have studied the correlation between different EquiFACS actions with different types of equine pain to try and understand which EquiFACS actions best indicate the presence of pain. Both works found that facial action units related to the ears, eyes and lower-face were important indicators of pain. In [Andersen et al., 2021], the authors explored the feasibility of an automated pain detection system for horses by integrating EquiFACS AU detection with machine learning techniques. Their approach yielded promising classification rates, demonstrating the potential for machine learning to advance automated equine pain recognition.

The main contributions of this work are as follows: a) We propose a baseline approach and adapt two deep learning-based architectures (I3D+LSTM, VideoMAE+LSTM) for fine-grained equine AU identification; b) We demonstrate potential solutions to overcome the critical challenge of limited annotated data availability using data-efficient AU detection models; c)We take a step towards advancing animal welfare through the automated detection of key affective state indicators.

2 Methodology

For our experiments, we will use the data introduced in [Rashid et al., 2020], which consists of 12 videos of horses (S1-S12) from different breeds recorded during a study on horses experiencing acute short-term pain by [Gleerup et al., 2015]. This dataset contains expertly annotated EquiFACS labels for each of the 12 videos, providing a comprehensive resource for AU analysis. In this work

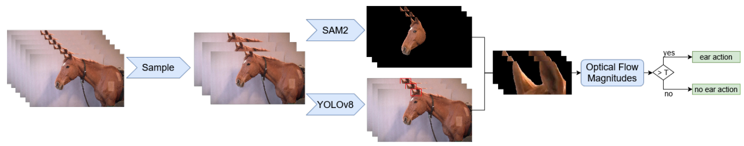
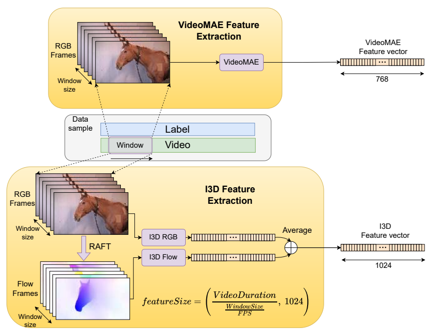
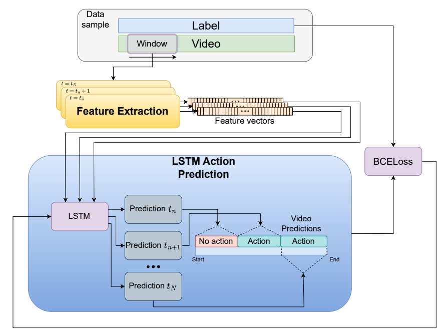


Figure 4: Pipeline for the baseline optical flow based ear movement detection (movDet).



(a) Feature extraction pipelines (I3D case shown)



(b) LSTM architecture pipeline for movement detection classification.

Figure 5: Pipelines for the proposed framework: (a) feature extraction (b) classification.

we adapt this data for the video classification task by clipping the relevant ear action sequences, as well as a close to equal number of background clips of similar lengths, ensuring a balanced dataset.

First, as a baseline, we propose an optical flow based method that analyses the magnitude of optical flow vectors within a defined region of interest, in our case the horse's ears.

Then, we adopt the approach proposed by [Wang et al., 2019], where features extracted using I3D (see [Carreira and Zisserman, 2017]), are further processed by a Long Short-Term Memory (LSTM) network for binary classification, in our case of ear movement (action vs. no-action) for each video clip. We test our method using both RGB and optical flow streams (extracted via RAFT [Teed and Deng, 2020]) separately, as well as a late fusion strategy that averages each stream's feature vectors before feeding them into the LSTM.

Finally, we propose replacing the I3D feature extractor with a VideoMAE (ViT-B) model (see Figure 5a). Features are then fed to the same LSTM architecture (see Figure 5a and Figure 5b). A more detailed version of our methods can be found in our original work [Alves et al., 2025].

Method	FPS	# Layers	LR	Window	Step	Accuracy	F1
movDet (Flow)	25	-	-	-	-	**0.75**	**0.739**
I3D+LSTM (Flow)	50	3	0.001	32	16	**0.813**	**0.816**
I3D+LSTM (Mixed)	50	3	0.005	32	16	0.75	0.769
I3D+LSTM (RGB)	50	3	0.005	32	16	0.625	0.679
VMAE+LSTM (RGB)	50	2	0.001	-	-	**0.875**	**0.870**

Table 1: Test results on movDet, I3D+LSTM and VideoMAE+LSTM. Best stream configuration accuracy is presented for each method, when applicable.

3 Experiments and Results

To assess the effectiveness of our proposed methods for equine ear movement detection, we conducted a series of experiments evaluating movement presence classification across dataset clips. The dataset consists of 283 clips of varying lengths (0.5s-3s), with 135 clips containing ear movements and 145 representing background (no ear movement). Expert annotated EquiFACS ear related AUs or EADs labels served as ground truth for classification (from [Rashid et al., 2020]). You can find additional experimental details in the original paper. Table 1 presents the classification accuracy and F1-score for each method's best-performing configuration. We selected the models that achieved the best validation accuracy for each method. Our optical flow-based approach (movDet) achieved an accuracy of 0.75 and an F1-score of 0.739. The I3D+LSTM model, leveraging deep spatiotemporal features, significantly improved upon this baseline, reaching 0.8125 accuracy and an F1-score of 0.816. Finally, the VideoMAE+LSTM model attained the highest accuracy of 0.875 and an F1-score of 0.869, demonstrating the efficacy of VideoMAE's transformer-based spatiotemporal representations in this context. For more in-depth explanations and results, please check the our full paper [Alves et al., 2025].

4 Discussion & Conclusion

Compared to our optical flow-based baseline, both the I3D+LSTM and VideoMAE+LSTM models improve accuracy by leveraging spatiotemporal features. Our results with a VideoMAE feature extractor combined with a LSTM classifier suggest that transformer architectures may be instrumental in addressing the limited availability of datasets, which is one of the primary challenges in equine affective computing. The methodologies developed here could be adapted for other species with similar affective state indicators, advancing animal welfare monitoring across various domains. Future research should continue exploring transformer-based models to enhance real-world applicability and improve the accuracy and efficiency of automated action unit detection systems, ultimately fostering a broader understanding of affective states across different species.

Acknowledgements. This work has been funded by the Independent Research Fund Denmark under grant ID 10.46540/3105-00114B.

References

J. Alves, P. H. Andersen, and R. Gade. Read my ears! horse ear movement detection for equine affective state assessment. In *Proceedings of the Computer Vision and Pattern Recognition Conference (CVPR) Workshops*, pages 5681–5689, June 2025.

P. H. Andersen, S. Broomé, M. Rashid, J. Lundblad, K. Ask, Z. Li, E. Hernlund, M. Rhodin, and H. Kjellström. Towards Machine Recognition of Facial Expressions of Pain in Horses. *Animals*, 11(6):1643, June 2021. ISSN 2076-2615. doi: 10.3390/ani11061643.

K. Ask, M. Rhodin, M. Rashid-Engström, E. Hernlund, and P. H. Andersen. Changes in the equine facial repertoire during different orthopedic pain intensities. *Scientific Reports*, 14(1):129, Jan. 2024. ISSN 2045-2322. doi: 10.1038/s41598-023-50383-y.

S. Broomé, K. Ask, M. Rashid-Engström, P. Haubro Andersen, and H. Kjellström. Sharing pain: Using pain domain transfer for video recognition of low grade orthopedic pain in horses. *PLOS ONE*, 17(3):e0263854, Mar. 2022. ISSN 1932-6203. doi: 10.1371/journal.pone.0263854.

J. Carreira and A. Zisserman. Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset, 2017.

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):e92281, Mar. 2014. ISSN 1932-6203. doi: 10.1371/journal.pone.0092281.

P. Ekman, W. V. Friesen, and J. C. Hager. Facial action coding system (FACS). *Research Nexus*, 2002.

K. B. Gleerup, B. Forkman, C. Lindegaard, and P. H. Andersen. An equine pain face. *Veterinary Anaesthesia and Analgesia*, 42(1):103–114, Jan. 2015. ISSN 14672987. doi: 10.1111/vaa.12212.

J. Lundblad. *Exploring Facial Expressions in Horses : Biological and Methodological Approaches*, volume 2024:38 of *Acta Universitatis Agriculturae Sueciae*. Swedish University of Agricultural Sciences, 2024. ISBN 978-91-8046-341-6. doi: 10.54612/a.15rcen99qf.

M. Rashid, A. Silventoinen, K. B. Gleerup, and P. H. Andersen. Equine Facial Action Coding System for determination of pain-related facial responses in videos of horses. *PLOS ONE*, 15(11):e0231608, Nov. 2020. ISSN 1932-6203. doi: 10.1371/journal.pone.0231608.

Z. Teed and J. Deng. RAFT: Recurrent All-Pairs Field Transforms for Optical Flow, 2020.

X. Wang, Z. Miao, R. Zhang, and S. Hao. I3D-LSTM: A New Model for Human Action Recognition. *IOP Conference Series: Materials Science and Engineering*, 569(3): 032035, July 2019. ISSN 1757-8981, 1757-899X. doi: 10.1088/1757-899X/569/3/032035.

J. Wathan, A. M. Burrows, B. M. Waller, and K. McComb. EquiFACS: The Equine Facial Action Coding System. *PLOS ONE*, 10(8):e0131738, Aug. 2015. ISSN 1932-6203. doi: 10.1371/journal.pone.0131738.