Vision-based detection of pain and nest-building behaviors in sows within commercial farrowing pens

Peter Helf
Precision Livestock Farming Hub
University of Veterinary Medicine Vienna
Veterinärplatz 1, 1210 Vienna, Austria
peter.helf@vetmeduni.ac.at

Maciej Oczak
Precision Livestock Farming Hub
University of Veterinary Medicine Vienna
Veterinärplatz 1, 1210 Vienna, Austria
maciej.oczak@vetmeduni.ac.at

Abstract

Pain indicators and preparturient nest-building are precursors of farrowing, yet continuous quantification remains challenging. We present a non-invasive computer-vision system that detects pain-associated behaviors (back-arching, tail-flicking, back leg forward, trembling) and nest-building behaviors (manipulation of pen components, pawing, exploration) from top-view videos. We analyzed 748 h of RGB footage (25 fps) from 11 sows on a single farm, spanning 64 h pre-farrowing to 4 h post birth of the first piglet. Using a defined ethogram, 46,010 events were annotated with inter-annotator agreement $\kappa = 0.724$. To assess generalization, we used a sow-level split, i.e., 8 in the training set and 3 in validation. Behaviors were detected with a modified DeepEthogram architecture, combining RGB data and optical flow. Both streams were processed by separate ResNet3D-34 encoders. Optical flow was estimated using a state-of-the-art DPFlow model. Training employed focal loss to address class imbalance, alongside geometric and photometric augmentations for robustness to camera placement and lighting. Clips of 11 frames at 8.33 fps ($\approx$1.32 s) were used. On held-out sows, per-class F1 scores were 0.875 for manipulation of pen, 0.634 for pawing, 0.820 for exploration, 0.639 for back-arching, 0.778 for tail-flicking, 0.871 for back leg forward, and 0.443 for trembling. These results indicate that pen-installed vision can identify key behaviors in a non-invasive way, supporting scalable monitoring. Limitations include modest dataset size and limited diversity, i.e., a single farm and a single breed. Ongoing work will expand the dataset and leverage behavior dynamics for time-to-farrowing estimation.

1 Introduction

Accurate, continuous monitoring of the preparturient period is critical for timely interventions at the onset of farrowing to improve sow welfare and piglet survival. Pain indicators (e.g., back arch, tail flick, trembling) and nest building behaviors (e.g., manipulation of pen, pawing, exploration) are possible precursors to the onset of farrowing [1, 2]. However, routine quantification of these events remains challenging because manual assessment is labor intensive and difficult to sustain at scale.

Sensor based approaches have estimated farrowing proximity by measuring overall activity with accelerometers attached to sows in farrowing pens [3]. While increases in activity can signal nest building, these methods require animal handling and device attachment, which may introduce stress and demand upkeep. Vision based, non-invasive alternatives have inferred activity from video, either by tracking sow position via object detection [4] or by estimating motion with optical flow [5]. Both strategies have shown promise for farrowing prediction but largely operate on coarse activity proxies rather than specific behaviors. Recognizing discrete pain and nest building behaviors may provide more reliable and interpretable cues for time-to-farrowing estimation.

The Third Austrian Symposium on AI, Robotics, and Vision (AIROV26).

Recent advances in video-based behavior recognition enable automatic detection of a range of sow behaviors across modalities (RGB, grayscale, depth) and tasks (e.g., posture, drinking, aggression) with real time performance on farm [6, 7]. To our knowledge, no prior work has targeted pain and nest building behaviors. We address this gap with a non-invasive computer vision system that detects specific pain-associated and nest-building behaviors from top-view pen videos.

2 Methods

2.1 Data and annotations

Data were collected at Medau, the pig research and teaching farm of the University of Veterinary Medicine Vienna, using top-view RGB cameras. A total of 748 hours of video, recorded at 25 frames per second (fps), from 11 individually housed Large White sows was used. Each sow was monitored continuously from approximately 64 hours before farrowing until 4 hours after the birth of the first piglet, capturing both the preparturient period and the onset of farrowing under routine farm conditions. All sows were kept in BeFree pens (Schauer Agrotronic, Prambachkirchen, Austria).

Table 1: Ethogram used for the labeling of behavior events

Behavior	Definition
Pawing	Raking movement of either forelimb against the floor or the pen
Manipulation of pen	The sow touches the pen component with some part of her head and performs up- and down-movements with her head towards the pen components.
Exploratory behavior	Movements of the head while the snout of the sow is directed towards the ground
Tremble	Visible shaking as if shivering when in a lateral lying position
Back leg forward	In a lateral lying position, the back leg is pulled forward and/or in toward the body
Back arch	In a lateral lying position, one or both sets of legs become tense and are pushed away from the body and/or inwards toward the center
Tail flick	Tail is moved rapidly up and down

An ethogram was defined to cover two behavior groups relevant to farrowing onset. Pain-associated behaviors included back arch, tail flick, pulling the back leg forward, and trembling. Nest building behaviors included manipulation of pen components, pawing, and exploration. Definitions are provided in Table 1. Because multiple behaviors can occur simultaneously, the task is multi-label. Using this ethogram, annotators labeled discrete behavior events throughout the recordings, producing a total of 46,010 events. To assess annotation reliability, a subset of the data, totaling 9 hours of video footage, was double-annotated, yielding inter-annotator agreement with κ=0.724. The dataset exhibited class imbalance, driven primarily by the higher frequency of back leg forward events relative to other categories. To partially mitigate this imbalance while preserving ethogram breadth, we downsampled the event pool to 23,326 by randomly removing 80% of back leg forward instances. The class composition of the resulting dataset is summarized in Table 2. To evaluate generalization and prevent data leakage, we used a sow-level split with 8 sows for training and 3 sows held out for validation.

Table 2: Number of occurrences for each behavior in the dataset

Behavior	Count
Pawing	1,950
Manipulation of pen	5,428
Exploratory behavior	4,938
Tremble	1,267
Back leg forward	5,671
Back arch	2,045
Tail flick	2,027

2.2 Neural network and training

Raw RGB streams at 25 fps were downsampled by keeping every third frame (≈8.33 fps) to reduce redundancy and computational load. Fixed-length clips of 11 consecutive frames (duration ≈1.32 s) were extracted within the start–end timestamps of annotated events. Although the model receives an 11-frame clip, the ground-truth labels of the center frame (6th) are assigned to each clip to align supervision with the temporal midpoint while providing context on both sides. Images were resized to a width of 416 pixels while keeping the aspect ratio. This resulted in an image height of 499 pixels. The images were then normalized using the ImageNet [8] mean and standard deviation. For robustness to real-world variability, we applied geometric and photometric augmentations during training, including random horizontal flips, small rotations and scalings, and brightness/contrast jitter. Because multiple behaviors can co-occur, the ground truth is multi-label.

To capture motion cues critical for transient behaviors, we adopted the dual-stream architecture of DeepEthogram [9]. The RGB stream uses 11-frame clips, and the motion stream uses corresponding optical-flow stacks computed with DPFlow [10]. A pretrained model provided by the Python package PyTorch Lightning Optical Flow [11] and pretrained on the Spring dataset [12]. A representative flow stack is shown in Figure 1. Each stream is encoded with a separate ResNet3D-34 backbone [13] to produce feature embeddings. Pretrainted weights provided by DeepEthogram [9], trained on Kinetics700 [14], were used. Each stream outputs class-wise logits for the seven behaviors. Late fusion combines the per-class logits from both streams using learnable weights before the final sigmoid:

$$l_c = \alpha_c l_c^{RGB} + (1 - \alpha_c) l_c^{Flow}, \quad y_c = \sigma(l_c) \tag{1}$$

where c indexes behaviors, $\alpha_c \in [0, 1]$ are learnable fusion weights, and $\sigma(\cdot)$ denotes the sigmoid. A threshold of 0.5 was used to determine the presence of a behavior.

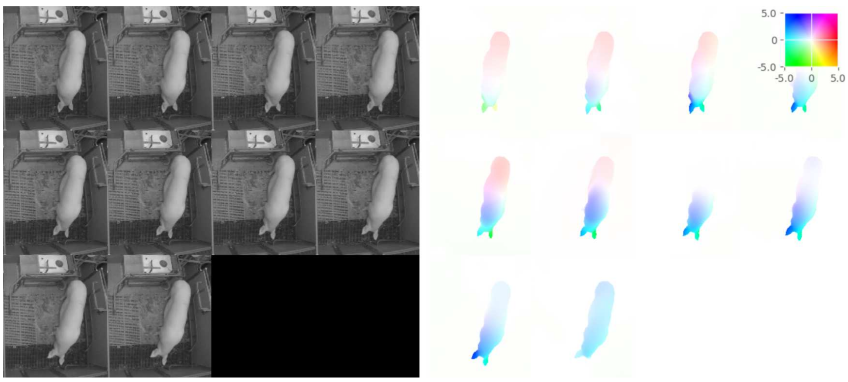

Figure 1: Optical flow stack. Left: RGB frame sequence (recording taken at night, thus a grayscale image) Right: computed optical flow. Pixel color encodes direction of movement with higher saturation indicating larger movements.

Given the remaining class imbalance, the training objective employed focal loss to down-weight common examples and emphasize minority-class instances [9, 15]. We used an initial learning rate of 10^{-3} with a ReduceLROnPlateau scheduler (patience: 5, factor: 0.1, min_lr=10^{-6}). Training stopped once the minimum learning rate was reached, and the checkpoint with the lowest validation loss was selected.

Training was performed in Python 3.11.10 with PyTorch 2.9.0 [16] using OpenCV 4.10.0.84 [17] for video I/O. Kornia 0.7.4 [18] was used for image transformation and augmentation. Experiments ran on an NVIDIA RTX 6000 Ada GPU (NVIDIA, Santa Clara, USA) in a server with an AMD EPYC 9354 32-core CPU (AMD, Santa Clara, USA).

3 Results

On the sow-held-out validation set (3 sows), the model achieved strong performance on nest-building behaviors and mixed performance on pain associated behaviors. For nest building, per-class F1 scores were 0.875 for manipulation, 0.634 for pawing, and 0.820 for exploration. For pain behaviors, F1 scores were 0.639 for back arch, 0.778 for tail flick, 0.871 for back leg forward, and 0.443 for trembling. Averaged over all seven classes, the macro F1 was 0.723. Table 3 summarizes per-class precision, recall, and F1 score. These results indicate that behaviors with more distinctive or sustained kinematics tended to yield higher F1 scores.

Table 3: F1 score, precision, and recall for the detected behaviors

Behavior	F1	Precision	Recall
Pawing	0.634	0.704	0.576
Manipulation of pen	0.875	0.835	0.919
Exploratory behavior	0.820	0.827	0.814
Tremble	0.443	0.383	0.524
Back leg forward	0.871	0.907	0.838
Back arch	0.639	0.854	0.511
Tail flick	0.778	0.833	0.730
Average	0.723	0.763	0.702

4 Discussion

Our results demonstrate that a dual-stream model can reliably detect several pre-parturient nest-building behaviors, with mixed performance on pain-associated behaviors. High F1 for manipulation and exploration suggests pronounced movements and interactions with pen fixtures are well captured by short clips. Tail flick, which is most often a short event, was detected well, indicating that distinctive movement can suffice. In contrast, trembling and back arch, often subtler movements, achieved lower scores.

The current clip length ($\approx$1.32 s) and frame rate ($\approx$8.33 fps) balance context and efficiency but may under-represent higher frequency motions (e.g., fine tremors) or longer sequences needed for some events. Extending temporal context, adopting other temporal models (e.g., Transformers), or using multi-rate inputs (retaining higher fps for flow while keeping RGB downsampled) may improve sensitivity to subtle and longer events.

Limitations include a modest dataset from a single farm, breed, and pen type, limiting visual diversity. Despite downsampling and focal loss, residual class imbalance persists and may bias decision boundaries. The single top-view perspective limits cues available for fine-grained posture changes, which may be relevant for some behaviors. Annotation noise, while mitigated by substantial agreement (κ=0.724), can still affect the results. Finally, training and validation used clips within behavior intervals without explicit background sampling. Continuous-video evaluation is needed to assess real-world prevalence and precision.

Future work should expand data across farms, breeds, camera placements, and lighting. Including background clips without any behaviors in the dataset will support more realistic evaluation. Longer-horizon temporal models (e.g., Transformers), may enhance sensitivity to some behavior indicators. Finally, the change in preparturition behavior frequency detected by the model needs to be further processed to achieve a time-to-farrowing estimation.

Acknowledgments and Disclosure of Funding

This research was funded in whole or in part by the Austrian Science Fund (FWF) [https://doi.org/10.55776/DFH34]

References

[1] Sarah H Ison, Susan Jarvis, and Kenneth MD Rutherford. The identification of potential behavioural indicators of pain in periparturient sows. *Research in veterinary science*, 109: 114–120, 2016.

[2] BI Damm, L Lisborg, KS Vestergaard, and J Vanicek. Nest-building, behavioural disturbances and heart rate in farrowing sows kept in crates and schmid pens. *Livestock Production Science*, 80(3):175–187, 2003.

[3] Maciej Oczak, Kristina Maschat, and Johannes Baumgartner. Dynamics of sows' activity housed in farrowing pens with possibility of temporary crating might indicate the time when sows should be confined in a crate before the onset of farrowing. *Animals*, 10(1):6, 2019.

[4] Maciej Oczak, Kristina Maschat, and Johannes Baumgartner. Implementation of computer-vision-based farrowing prediction in pens with temporary sow confinement. *Veterinary Sciences*, 10(2):109, 2023.

[5] Kejian Liu, Yigui Huang, Junbin Liu, Zujie Tan, and Deqin Xiao. Prediction of sow farrowing onset time using activity time series extracted by optical flow estimation. *Animals*, 15(7):998, 2025.

[6] Ziting Zhang, Hang Zhang, Yuxiang He, and Tonghai Liu. A review in the automatic detection of pigs behavior with sensors. *Journal of Sensors*, 2022(1):4519539, 2022.

[7] Yuanqin Zhang, Jiahao Cai, Deqin Xiao, Zesen Li, and Benhai Xiong. Real-time sow behavior detection based on deep learning. *Computers and Electronics in Agriculture*, 163:104884, 2019.

[8] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252, 2015. doi: 10.1007/s11263-015-0816-y.

[9] James P Bohnslav, Nivanthika K Wimalasena, Kelsey J Clausing, Yu Y Dai, David A Yarmolinsky, Tomás Cruz, Adam D Kashlan, M Eugenia Chiappe, Lauren L Orefice, Clifford J Woolf, et al. Deepethogram, a machine learning pipeline for supervised behavior classification from raw pixels. *elife*, 10:e63377, 2021.

[10] Henrique Morimitsu, Xiaobin Zhu, Roberto M Cesar, Xiangyang Ji, and Xu-Cheng Yin. Dpflow: Adaptive optical flow estimation with a dual-pyramid framework. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 17810–17820, 2025.

[11] Henrique Morimitsu. Pytorch lightning optical flow. `https://github.com/hmorimitsu/ptlflow`, 2021.

[12] Lukas Mehl, Jenny Schmalfuss, Azin Jahedi, Yaroslava Nalivayko, and Andrés Bruhn. Spring: A high-resolution high-detail dataset and benchmark for scene flow, optical flow and stereo. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4981–4991, 2023.

[13] Kensho Hara, Hirokatsu Kataoka, and Yutaka Satoh. Learning spatio-temporal features with 3d residual networks for action recognition. In *Proceedings of the IEEE international conference on computer vision workshops*, pages 3154–3160, 2017.

[14] Joao Carreira, Eric Noland, Chloe Hillier, and Andrew Zisserman. A short note on the kinetics-700 human action dataset. *arXiv preprint arXiv:1907.06987*, 2019.

[15] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, 2017.

[16] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.

[17] G. Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000.

[18] Edgar Riba, Dmytro Mishkin, Daniel Ponsa, Ethan Rublee, and Gary Bradski. Kornia: an open source differentiable computer vision library for pytorch. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 3674–3683, 2020.