Pose estimation for behavioural anomaly detection in pigs: comparative analysis of key point configuration and neural networks

K. Ivanov[a], Valentina Bonfatti[a], Claudia Kasper[b], Hassan-Roland Nasser[c]

[a] *Department of Comparative Biomedicine and Food Science, University of Padova, Viale dell'Università 16, 35020, Legnaro (PD), Italy*
[b] *Animal GenoPhenomics, Agroscope, Posieux, Switzerland*
[c] *Digital Production, Agroscope, Posieux, Switzerland*

Effective monitoring of tail-biting behavior in pigs is crucial for ensuring animal welfare and reducing economic losses in livestock farming. This study evaluates the accuracy and computational performance of pose estimation techniques in detecting behavioral anomalies, specifically interactions involving pig's snouts and tails. A custom video dataset of 100 annotated frames, tracking 10 key points per animal, was used. Five neural networks (AnimalTokenPose, HRNet-W32, ResNet-50, and ResNet-101) were tested for tracking accuracy, computational efficiency, and robustness under challenging conditions (e.g., occlusions and lighting variability). The analysis revealed distinct movement patterns, with the tail and body (snout, head, ears, and five points along the dorsal axis) exhibiting the most concentrated trajectories on the 2D plane, suggesting frequent movement in monitored videos. Confidence scores for tracking accuracy were evaluated, and a confidence score cutoff (p_cutoff) was set to exclude unreliable key points. Key points with confidence scores below this threshold were excluded from further analysis. Models for which at least 90% of the key points had a confidence score higher than the p_cutoff threshold demonstrated higher reliability for pose estimation (confidence score>0.6). Additionally, precision and recall metrics were defined based on the detection of key points (tail and snout positions) across all frames. Precision was calculated as the proportion of correctly detected key points relative to the total number of detected key points. Recall was calculated as the proportion of correctly detected key points relative to the total number of actual key points. The best-performing model in terms of precision (mAP) and recall (mAR) was ResNet-101 pre-trained on ImageNet weights, achieving a precision of 93.68% and a recall of 94.17%. Conversely, the worst-performing model was HRNet-W32 pre-trained on the Super Animal Quadruped weights, with a precision and a recall of 66.85% and 77.0% respectively.