Equine Pain Behavior Classification via Self-Supervised Disentangled Pose Representation
Supplementary Materials

Maheen Rashid[1,2] Sofia Broomé[3] Katrina Ask[4] Elin Hernlund[4] Pia Haubro Andersen[4]
Hedvig Kjellström[3,5] Yong Jae Lee[1,6]
[1] UC Davis, USA [2] Univrses AB, Sweden maheen.rashid@univrses.com
[3] KTH Royal Institute of Technology, Sweden sbroome, hedvig@kth.se
[4] SLU, Sweden katrina.ask, elin.hernlund, pia.haubro.andersen@slu.se
[5] Silo AI, Sweden [6] UW Madison, USA yongjaelee@cs.wisc.edu

Horse	# No-Pain	# Pain	%age No-Pain	%age Pain
Aslan	414	376	52.41%	47.59%
Brava	335	436	43.45%	56.55%
Herrera	470	567	45.32%	54.68%
Inkasso	330	439	42.91%	57.09%
Julia	465	433	51.78%	48.22%
Kastanjett	413	357	53.64%	46.36%
Naughty but Nice	405	394	50.69%	49.31%
Sir Holger	351	308	53.26%	46.74%
Total	3183	3310	-	-

Table 1. Number of pain and no-pain video segments per horse.

Horse	# No-Pain	# Pain	%age No-Pain	%age Pain
Aslan	50974	53396	48.84%	51.16%
Brava	54355	49651	52.26%	47.74%
Herrera	46194	51964	47.06%	52.94%
Inkasso	55805	53948	50.85%	49.15%
Julia	48695	50902	48.89%	51.11%
Kastanjett	55028	55990	49.57%	50.43%
Naughty but Nice	54571	51117	51.63%	48.37%
Sir Holger	55617	56129	49.77%	50.23%
Total	421239	423097	-	-

Table 2. Number of pain and no-pain frames per horse.

1. Further implementation details

The dataset presented many practical challenges in terms of preprocessing. Videos from each camera were manually offset when necessary to sync temporally with other cameras in the stall. Time periods with humans present in the stall or corridor were manually marked for exclusion when not recorded in the experiment log. Technical faults led to intermittent recording from some cameras. Only time periods with footage from all cameras were used.

The cameras were calibrated to recover their intrinsic and extrinsic parameters by use of a large checkerboard pattern of known dimensions, solving Perspective-and-Point (PnP) problem using RANSAC in OpenCV [2], and bundle adjustment.

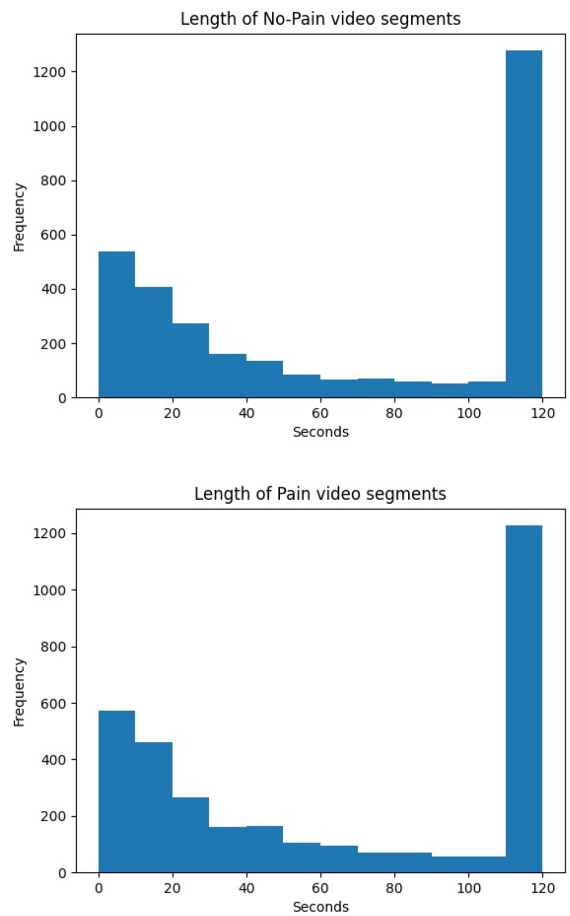

Figure 1. Histogram of video segments' length for pain and no-pain. Most data points have length 2 minutes.

Table 1 provides the number of video segments for each horse used for pain classification. Table 2 provides the same information information in terms of frames. The corresponding pain and no-pain percentages varies between these two tables since the segment length is variable.

We used 'Naughty but Nice' as our validation horse, as it has the most balanced class distribution for video segments (Table 1). When testing on 'Naughty but Nice', we used the first horse, 'Aslan', as our validation subject. To keep results comparable we used the same subjects for validation when training on frames, even though the class distribution is different (Table 2).

In Figure 1 we show the distribution of video segments' length in seconds. Most segments are two minutes in length. A slightly larger proportion of no-pain videos have 2 minutes length compared to their proportion in pain videos. This may be because horses display restlessness when in pain which makes their consistent detection in every frame more difficult.

When detecting horsing with MaskRCNN [6], we noticed high confusion between 'horse' and 'cow' categories, and included high confidence detections from both categories.

To compensate for unbalanced training data we use a weighted cross-entropy loss. The weights for each class are calculated as follows, where p and np are the number of training data points for pain and no-pain respectively:

$$
\begin{aligned}
w_{pain} &= 2(1 - \frac{p}{p+np}), \\
w_{no-pain} &= 2(1 - \frac{np}{p+np}).
\end{aligned} \tag{1}
$$

Multiplication by 2 – the number of classes – keeps the weight around 1, hence maintaining the overall magnitude of the loss.

The maximum length of each video segment is 2 minutes. At 2 fps this equals 240 frames. The minimum length is 5 seconds (10 frames). The frame based model was trained at 1 fps. This was done to avoid very repetitive frames, and to speed up training.

The model trained from scratch was trained using a different schedule as the entire network had to be learned, and not just the pain classification head. We trained the model at 0.0001 learning rate for 50 epochs. We evaluated on validation and testing data after every 5 epochs, and used the results to determine both the 'True' and 'Oracle' performance.

The code base is available at `https://github.com/menorashid/gross_pain`. The EOP dataset can be accessed by agreement for collaborative research with the authors of [1].

2. Qualitative examples

Figure 2 shows further qualitative examples of video segments correctly classified as painful.

Rows 1-2 show 'lowered ears' [5], rows 3-4 show 'lying down' [7], and row 5 shows stretching [4] similar to results in the main paper. In addition, we observe 'looking at painful area', and 'lowered head' in rows 6-7 [5], and 'frequent tail flicking' in row 8 [8]. 'lowered ears' [5] (second row), a lifted left hind limb (first row), corresponding to 'non-weight bearing' [3], 'lying down' (third row), 'looking at flank' (fourth row) [7], and one example of gross pain behavior, 'stretching' (last row) [4].

References

[1] Katrina Ask, Marie Rhodin, Lena-Mari Tamminen, Elin Hernlund, and Pia Haubro Andersen. Identification of body behaviors and facial expressions associated with induced orthopedic pain in four equine pain scales. *Animals*, 10(11), 2020.

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

[3] G Bussieres, C Jacques, O Lainay, G Beauchamp, Agnès Leblond, J-L Cadoré, L-M Desmaizières, SG Cuvelliez, and E Troncy. Development of a composite orthopaedic pain scale in horses. *Research in veterinary science*, 85(2), 2008.

[4] KB Gleerup and Casper Lindegaard. Recognition and quantification of pain in horses: A tutorial review. *Equine Veterinary Education*, 28(1), 2016.

[5] Karina B Gleerup, Björn Forkman, Casper Lindegaard, and Pia H Andersen. An equine pain face. *Veterinary anaesthesia and analgesia*, 42(1), 2015.

[6] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask rcnn. In *ICCV*, 2017.

[7] Lori C Pritchett, Catherine Ulibarri, Malcolm C Roberts, Robert K Schneider, and Debra C Sellon. Identification of potential physiological and behavioral indicators of postoperative pain in horses after exploratory celiotomy for colic. *Applied Animal Behaviour Science*, 80(1), 2003.

[8] Johannes PAM van Loon and Machteld C Van Dierendonck. Monitoring acute equine visceral pain with the equine utrecht university scale for composite pain assessment (EQUUS-COMPASS) and the equine utrecht university scale for facial assessment of pain (EQUUS-FAP): a scale-construction study. *The Veterinary Journal*, 206(3), 2015.

Figure 2. Painful behavior correctly classified by our network