Cleanliness evaluation of poultry carcasses using deep learning analysis of images

Vitaly Belik[1], Lina Buhre[2], Claudia Ruhland[2], Diana Meemken[2], Nina Langkabel[2]

[1] System Modeling Group, Institute of Veterinary Epidemiology and Biostatistics, School of Veterinary Medicine, Freie Universität Berlin

[2] Institute of Food Safety and Food Hygiene, School of Veterinary Medicine, Freie Universität Berlin

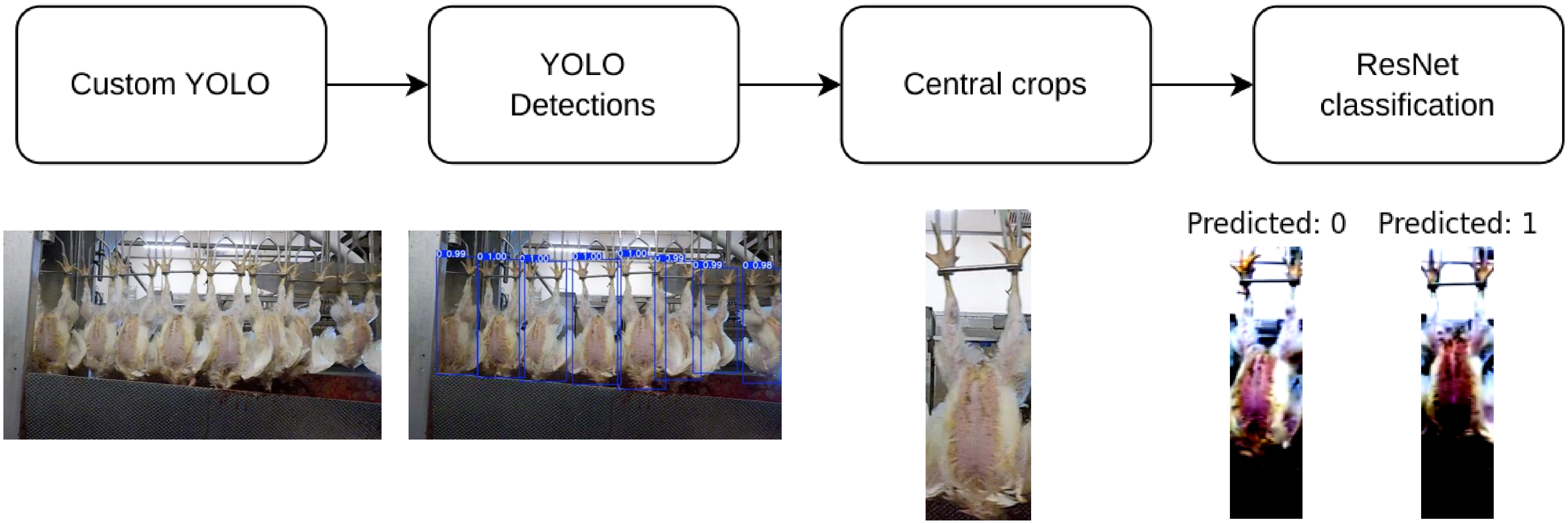
Introduction

- Cleanliness of poultry carcasses is required for welfare assessment and management of meat hygiene at slaughterhouses (Reg. (EC) No 852/2004, Reg. (EC) No 853/2004, AVV LmH, 2022)
- Assesment of dirtiness compared to cleanliness of broiler carcasses should be performed at the slaughterhouse
- But manual evaluation is tedious and could be performed only on some small sample of carcasses
- Automated evaluation would evaluate all carcasses and does not require a human observer

Dataset

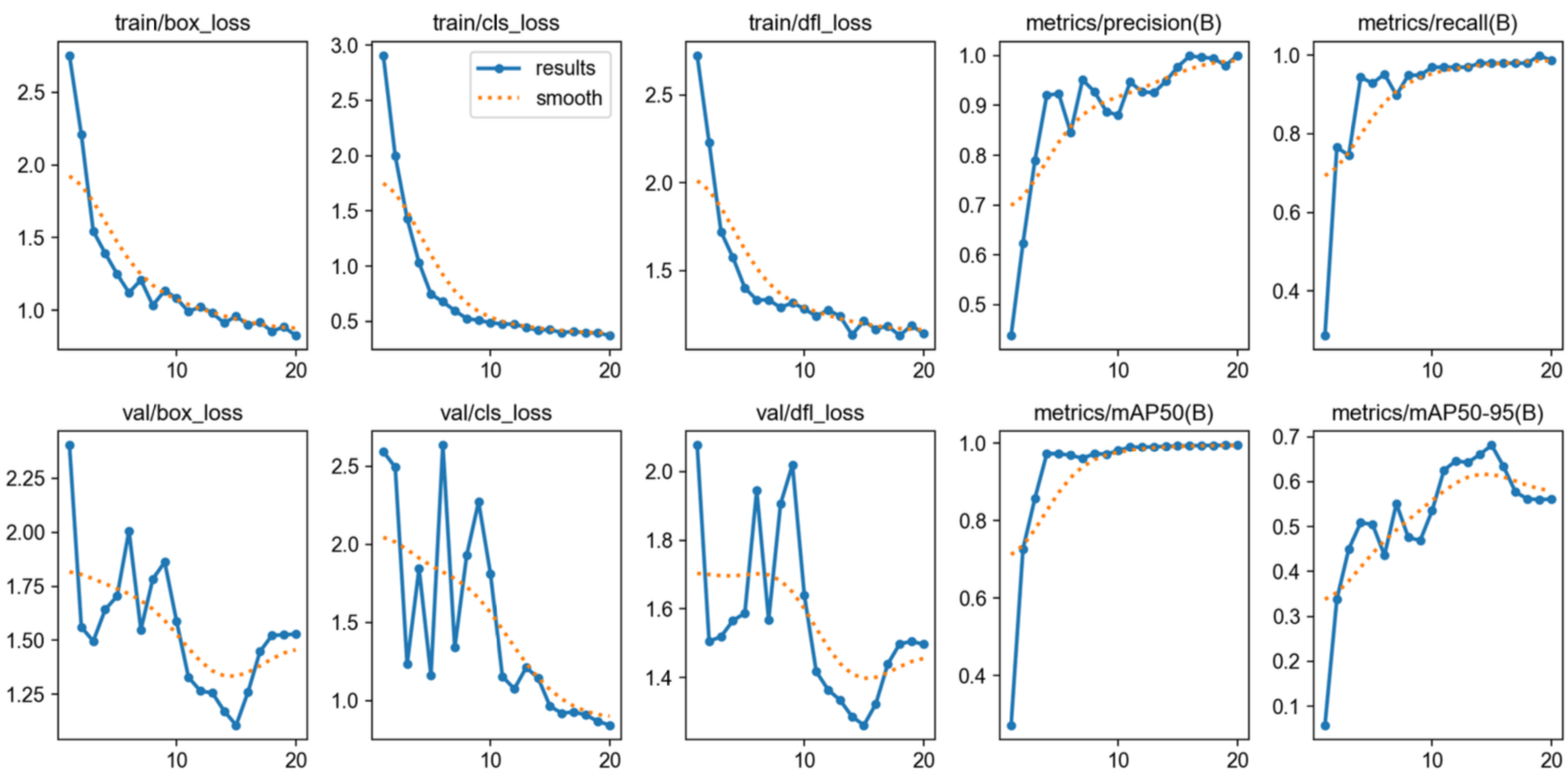
- We recorded 10 short (3-5 minutes long) high frequency (118 fps) videos of two herds in a slaughter house. The first video was used for model training.
- For YOLO training we used just 37 images with ca. 8 annotated objects (bounding boxes of carcasses) each.
- For ResNet training we had 487 human (by a trained veterinary student) labelled instances of non-clean carcass images and 715 clean carcass images.

Computational pipeline



We trained a two stage model for detection of carcasses (YOLO + center bbox) and their classification into non-clean and clean class (ResNet).

Detection of carcasses (YOLO)



https://github.com/ultralytics/ultralytics

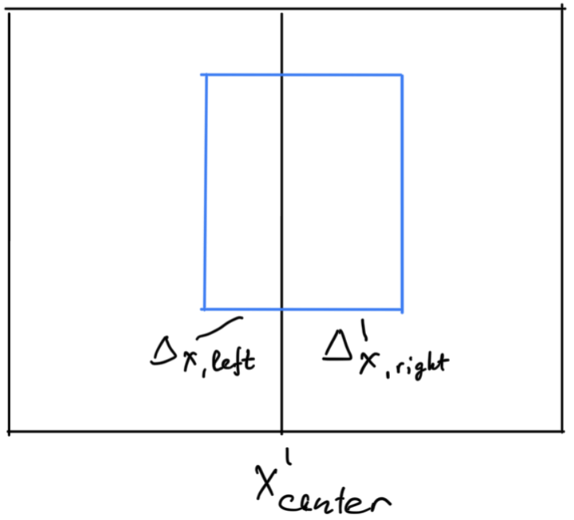
Filtering central carcasses

To guarantee the assessment of a carcass occurs only once, we chose only the most central bounding box from a sequence of consecutive frames.

We run YOLOv8m to get all bounding boxes of carcasses on a video frame. Afterwards we identified the most central bounding box. This was done finding the minima of a quantity describing the centrality

$$\zeta = \Delta_{x,\text{left}} - \Delta_{x,\text{right}}$$

where $\Delta_{x,\text{left}} = x_{\text{center}} - x_{\text{min}}$ and
$\Delta_{x,\text{right}} = x_{\text{max}} - x_{\text{center}}$

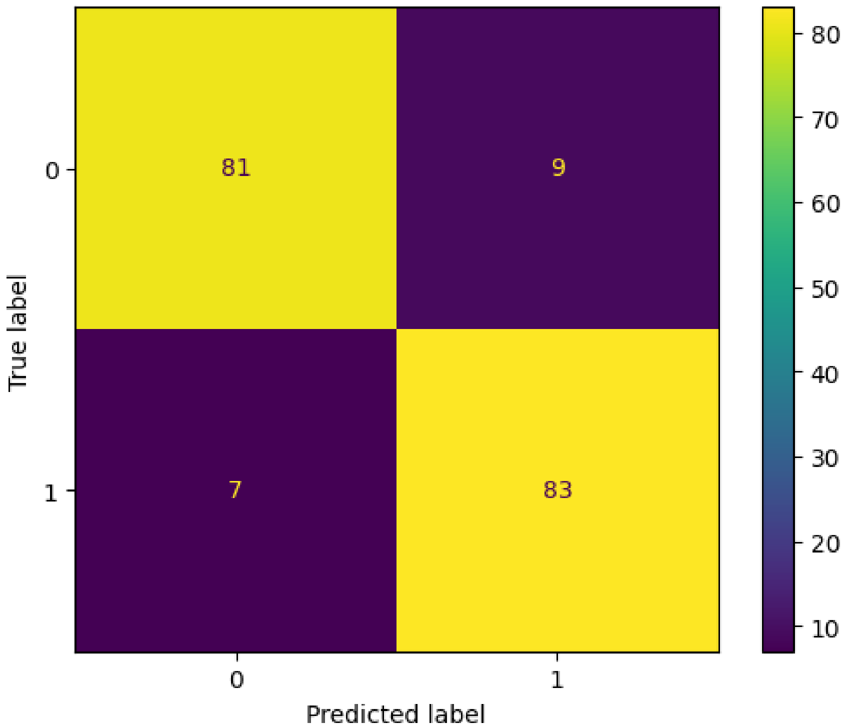


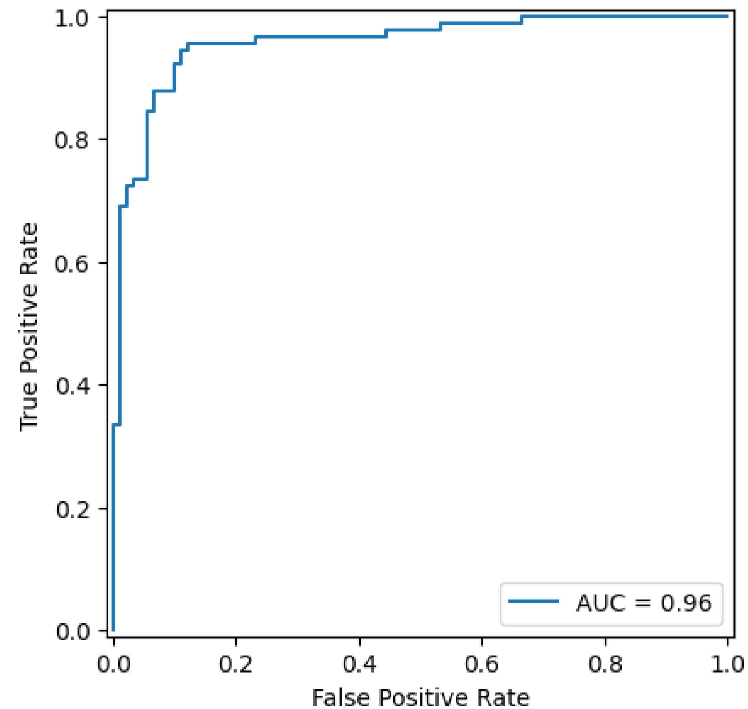
Classification

- Crops from a frame containing single carcasses are prepared for further classification.
- We fine tuned a ResNet implemented in pytorch pretrained on ImageNet to classify carcasses into non-clean and clean classes. For the training (80/20 split) we undersampled the overrepresented class.
- ResNet34 with 21.8 Mio parameters was a suitable choice given limited annotations

https://doi.org/10.48550/arXiv.1512.03385

Classification (ResNet)





Accuracy $= \frac{\mathrm{TP+TN}}{\text{Total count}} = 0.911$; Sensitivity (Recall) $= \frac{\mathrm{TP}}{\mathrm{TP+FN}} = 0.922$;

Specificity = $= \frac{\mathrm{TN}}{\mathrm{TN+FP}} = 0.900$

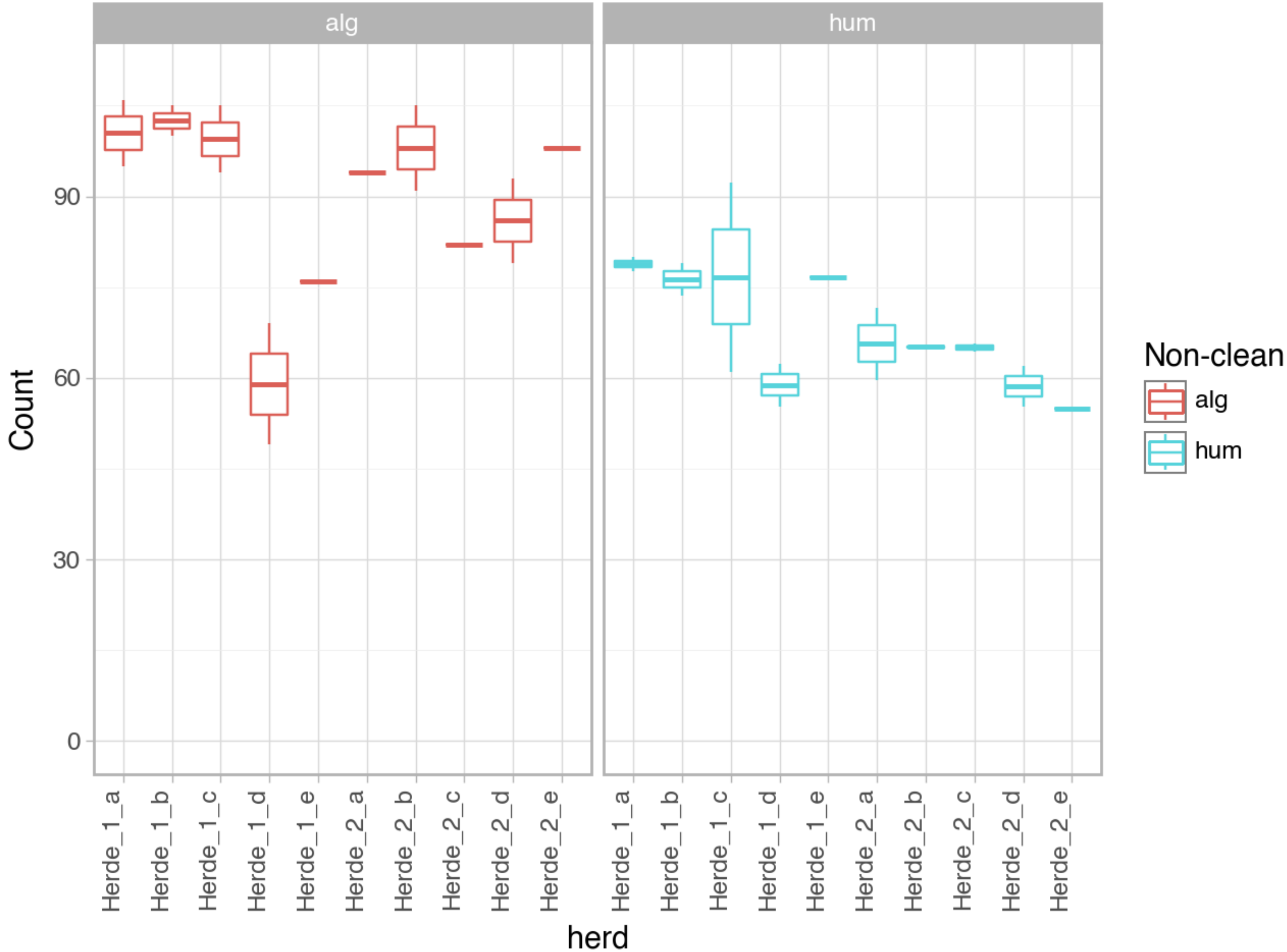
Results

To compared our algorithm with human evaluations we calculated *inter observer agreement*

$$IOA = \frac{\text{Smaller count}}{\text{Larger count}}$$

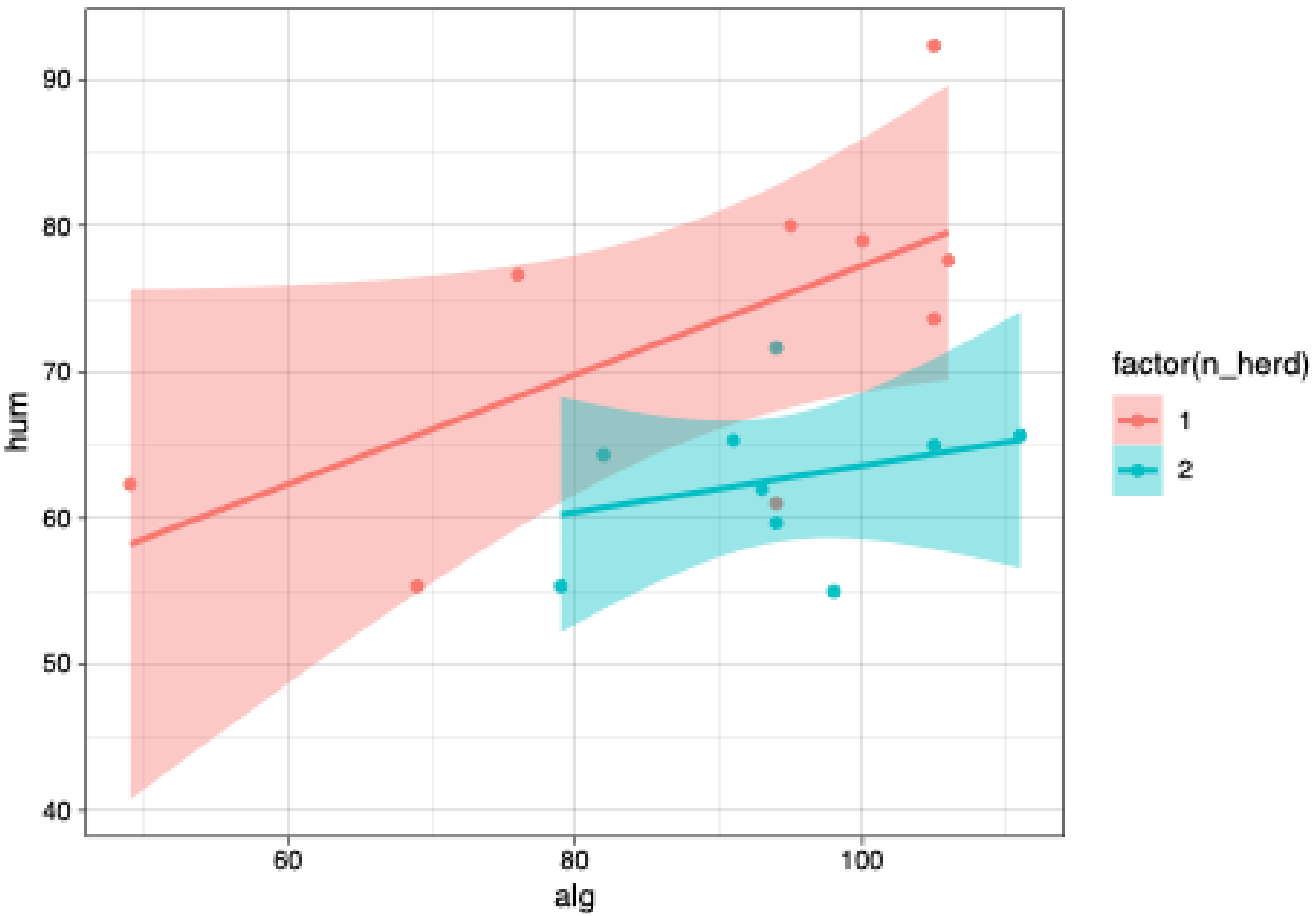
Herd	Run	Alg	Hum	IOA
Herde_1_a	1	95	80.00	0.84
Herde_1_a	2	106	77.67	0.73
Herde_1_b	1	100	79.00	0.79
Herde_1_b	2	105	73.67	0.70
Herde_1_e	1	76	76.67	0.99
Herde_1_d	1	49	62.33	0.79
Herde_1_d	2	69	55.33	0.80
Herde_1_c	1	94	61.00	0.65
Herde_1_c	2	105	92.33	0.88
Herde_2_a	1	94	71.67	0.76
Herde_2_a	2	94	59.67	0.63
Herde_2_b	1	91	65.33	0.72
Herde_2_b	2	105	65.00	0.62
Herde_2_c	1	82	64.33	0.78
Herde_2_c	2	111	65.67	0.59
Herde_2_d	1	93	62.00	0.67
Herde_2_d	2	79	55.33	0.70
Herde_2_e	1	98	55.00	0.56

Results



alg
hum
Count
90
60
30
0
Herde_1_a
Herde_1_b
Herde_1_c
Herde_1_d
Herde_1_e
Herde_2_a
Herde_2_b
Herde_2_c
Herde_2_d
Herde_2_e
herd
Non-clean
alg
hum

Results



Conclusion

- An automated pipeline was implemented using custom trained YOLO and ResNet
- Agreement so far between the algorithm and humans was IOA=0.74±0.11 (0.8±0.1 and 0.67±0.08)
- Allows to access not only a sample but essentially every carcass

Outlook

– Expanding the annotation base / Validation with additional videos
– Using further DL architectures for classification
– Incorporating a tracking algorithm to evaluate carcasses using ensemble approach
– Refining the model to detect different grades of non-cleanness
– Incorporating XAI
– Improvement of the pipeline latency to be able to work with a video stream

Acknowledgements

- This project was partly funded by the QS Wissenschaftsfonds.

- Thank you for your attention!

vitaly.belik@fu-berlin.de