Politécnico
de Viseu

Escola Superior
Agrária de Viseu

Computer vision tools to detect the precursors of tail biting in pigs: first steps towards methodological framework

VERSÃO FINAL

Ana Margarida Gomes Farinha Paula (24432)

ESAV, 2025

Politécnico de Viseu

Escola Superior Agrária de Viseu

Computer vision tools to detect the precursors of tail biting in pigs: first steps towards methodological framework

VERSÃO FINAL

Ana Margarida Gomes Farinha Paula (24432)

Master's Thesis

Animal Production Technologies

Supervised by:

Jorge Oliveira

Claudia Kasper

Hassan Nasser

ESAV, 2025

Acknowledgements

I would like to thank my supervisors Claudia Kasper and Hassan Roland, who accompanied me throughout this work and in discovering this world of research and computer vision, which until then had been completely unknown to me. Special thanks also go to Professor Jorge Oliveira from the Polytechnic University of Viseu, who allowed me to take part in this project as my official supervisor at the Agrarian School of Viseu, without whom it would not have been possible to consolidate the knowledge acquired. I must also thank the competent and essential Posieux livestock team led by Guy Maïkoff, without whom it would not have been possible to obtain the data needed to carry out the project. I am grateful to my colleagues at SAAV, who were essential in keeping me motivated throughout the internship. I dedicate this work to my mum and dad, my rare treasures, always available on the right side, at the right time, and with the right words. I also dedicate this work to my children and thank them for their patience in putting up with my 'absence', but I hope that these last few months will serve as an example for their future.

The results of this work supported partially, or in its entirety, the following presentations at scientific events:

Ana Gomes, Claudia Kasper, Hassan Nasser & Jorge Oliveira (2024). Computer vision and machine learning tools to detect the precursors of tail biting in pigs: A methodological framework. Poster presentation at *XII Jornadas Internacionais de Suinicultura*, 15 e 16 de março, Aula Magna, UTAD, Vila Real, Portugal.

Ana Gomes, Hassan Nasser & Claudia Kasper (2024). Detection of pig behaviour – Groundwork for ground truth. Oral presentation at *Data day, Computer Vision for Agriculture*. 14 June, ETH Zurich, Zurich, Suisse.

Ana Gomes, Claudia Kasper, Hassan Nasser & Jorge Oliveira (2024). Outils de vision par ordinateur pour la détection des précurseurs de la caudophagie chez les porcs : première approche d'un cadre méthodologique. Poster presentation at *Journée de conférences Interne, Réseau Animaux de Rente*. 8 novembre, Aula Paul Bourqui, Grangeneuve, Posieux, Suisse.

Abstract

Tail biting is one of the most significant problems affecting production in terms of animal welfare, productivity, and health. In response, this project presents the application of existing computer vision and machine learning tools to detect the precursors of tail biting in pigs under field conditions. Computer vision systems can facilitate phenotyping to select more stress-resistant pigs and help detect problems at an early stage, both at the individual and group levels. Behavioural changes, such as activity in the pen, frequency of contact between pigs in a group, or handling of objects, can be indicative of the development of tail biting, among other problems.

The proposed work curated an animal identification dataset that will support the creation of a reference database for detecting these changes. The data included approximately 1,800 h of video recordings of the daily lives of two groups of 12 fattening pigs, each weighing 100–140 kg, at Agroscope's experimental station in Switzerland. A bibliographical survey of existing computer vision systems for pigs allowed for the selection of the annotation and training tools and the development of an ethogram that was best suited to the project. From the initial set of images, we obtained a first data subset of 2,500 images automatically selected using Lightly for annotation. The annotation algorithm chosen was CVAT (Computer Vision Tool), which enabled 280 images to be annotated with bounding boxes and 520 with semi-automatic segmentation using the Segment Anything Model (SAM) for the classes of objects present in each image (ID of each pig, head, tail), location, and additional attributes, such as poses (lying down, standing, or sitting/kneeling). The 280 images annotated with bounding boxes were used to train YOLOv8, achieving an accuracy of 0.93 for the heads and 0.84 for the tails of the pigs. Although we did not achieve the desired accuracy, the results were quite satisfactory, given the number of frames used for training. Although we could not train the object detection model with the images annotated with the Segment Anything Model within the allocated project time, we expect such an effort to yield better results, as the annotation is more accurate and provides greater amounts of ground truth, hence better accuracy.

Keywords: Pigs, behaviour, tail biting, prediction, dataset, annotations, computer vision, machine learning.

Resumo

A caudofagia é um dos problemas que mais afeta a produção em termos de bem-estar animal, produtividade e saúde, o projeto a apresentar consiste na aplicação de ferramentas existentes de computer vision e de machine learning para detetar os precursores da mordedura da cauda em suínos em condições de campo. Os sistemas de computer vision podem facilitar a fenotipagem para selecionar suínos mais resistentes ao stress e ajudar a detetar problemas numa fase precoce, tanto a nível individual como de grupo.

A observação de alterações comportamentais, como a atividade no parque, a frequência de contacto entre suínos de um grupo ou o manuseamento de objetos, pode ser indicativa do desenvolvimento da mordedura da cauda, entre outros problemas.

O trabalho proposto consiste principalmente na preparação de um conjunto de dados para identificação animal que permitirá a criação de uma base de dados de referência. Com cerca de 1800 horas de vídeos gravados do quotidiano de dois grupos de 12 porcos de engorda, cada um com 100-140 kg, na estação experimental AGROSCOPE na Suíça, e após um levantamento bibliográfico dos sistemas de computer vision existentes para porcos, foi possível selecionar as ferramentas de anotação e treino e desenvolver um etograma mais adequado ao projeto.

A partir do conjunto inicial de imagens, obtivemos um primeiro subconjunto de dados de 2.500 imagens selecionadas automaticamente utilizando o Lightly, para anotação. O algoritmo de anotação escolhido foi o CVAT (Computer Vision Tool), que permitiu anotar 280 imagens com caixas delimitadoras (Bounding box) e 520 com segmentação semi-automática utilizando o Segment Anything Model (SAM) para as classes de objetos presentes em cada imagem (ID de cada porco, cabeça, cauda), localização e atributos adicionais como a pose (deitado, de pé ou sentado/joelhado).

Foi possível treinar o YOLOv8 com as 280 imagens anotadas com Bounding box, obtendo uma precisão de 0,93 para a cabeça e 0,84 para a cauda. Embora não tenhamos atingido a precisão desejada, os resultados são bastante satisfatórios dado o número de imagens utilizadas para o treino.

O tempo de projeto não nos permitiu fazer o treino com as imagens anotadas com o SAM, mas espera-se que obtenhamos melhores resultados, uma vez que a anotação é mais precisa e fornece uma maior quantidade de dados concretos, logo uma melhor precisão.

Palavras-chave: Porcos, comportamento, mordedura de cauda, previsão, conjunto de dados, anotações, computer vision, machine learning.

Index of figures

Index of tables

Table of contents

Introduction

Pork is one of the most consumed meats globally (Odo et al., 2024), and with the predicted population growth in the coming years (Collins & Smith, 2022), the intensification of pig production has increased and will require increasingly efficient, profitable, and sustainable production. Even with the evolution of management techniques, more efficient and sustainable structures and equipment, improvements in the genetic potential of animals, and growing concern for animal health and welfare, finding a balance between environmental protection and sustainability, animal welfare and health, economic sustainability, and food security (Farahnakian et al., 2024) remains a challenging objective.

Although best management practices are increasingly respecting animal welfare, pigs continue to be raised generally under intensive conditions (Odo et al., 2024), which are susceptible to stress, leading to the appearance of diseases and abnormal behaviours and aggression, such as tail biting or ear biting. In fact, tail biting is one of the most significant welfare and economic problems affecting current pig production (Taylor et al., 2012), causing additional stress and pain to animals (Larsen et al., 2019; Munsterhjelm et al., 2016), impairing their welfare, leading to the appearance of other diseases, growth delays, rejections at slaughter (Larsen et al., 2019), and thus significant economic losses. Current research in Switzerland shows that 37% of slaughtered pigs have lesions (23.7% with healed lesions, 1.3% with acute lesions, and 12.0% with chronic lesions) on their tails caused by bites from fellow pigs (Gerster et al., 2022). van Staaveren et al. (2021) studied the impact of these injuries on economic results in Ireland and concluded that farms with a higher prevalence (≥ 0.86%) of severe tail injuries had a 4.8% lower average daily gain and needed an additional 7 days to reach the target slaughter weight. This resulted in using 3.6% more weaning feed and 1.4% more finishing feed per year, increasing feed costs by 1.5%, and reducing the average annual profit of these farms by 15.1%, hence the need to prevent these outbreaks in farms.

Currently, abnormal behaviours in groups of animals are observed directly by humans, which often makes preventive action impossible, allowing only correction to avoid greater harm. Computer vision (CV) is a subset of artificial intelligence that performs calculations on images or video data, enabling the quantitative analysis of visual information (Pangal et al., 2021). CV tools allow for the development of continuous and real-time monitoring systems through video recording, which can identify predictive behavioural patterns, such as those indicative of diseases and aggression, including ear biting and tail biting. These tools are also of great utility for animal behaviour and welfare researchers, as they allow deeper analysis, making it possible to manage a larger quantity of data with less consumption of time and resources. Additionally, they will enable the phenotyping of individuals with the studied behavioural patterns and thus

improve genetic improvement programs (Fernandes et al., 2020), for example, by selecting pigs more resistant to stress.

In this context, the theoretical part of this work aims to frame the problem of tail biting in the project carried out and highlight the importance of CV in pig production research, summarise the current state of research, and outline the main objectives and steps of this work.

1. Tail biting

Tail biting is characterised by the exploratory behaviour of a pig (the biter), which intensifies abnormally and is redirected towards manipulating and biting the tails of its conspecifics (victims) (Hakansson & Houe, 2020). This behaviour is often a manifestation of stress caused by internal frustration when biological needs cannot be met due to inadequate environmental stimuli (Schrøder-Petersen & Simonsen, 2001; Taylor et al., 2010), identified as risk factors. This stress reaction in response to exposure to one or more risk factors represents the neurophysiological reactions that are present in all circumstances in which an animal needs to adapt physiologically to restore the body's homeostasis in a very short period of time (on the order of a few minutes); however, prolonging this response over time is harmful to the animal's health.

1.1. Risk factors that lead to its occurrence

Given the significance of tail biting, this phenomenon has been studied for several decades, and many researchers have identified factors associated with the occurrence of outbreaks and best practices for reducing/preventing its incidence. Drexl et al. (2023) distinguished the risk factors for tail biting into two groups: variable factors and constant factors, and they were classified as internal and external risk factors by Schrøder-Petersen and Simonsen (2001). Production conditions and techniques that the pig farmer can alter or improve constitute constant or external factors, which include, among others, feeding systems (Moinard et al., 2003) and watering, food quality, mixing of animals (Wutke et al., 2021), temperature, and humidity, ventilation, stocking densities (D'Eath et al., 2014), herd size, type of flooring, and other equipment. The variable or internal factors are those directly linked to the animals and can be observed and analysed to enable direct intervention; they include genetics (Sonoda et al., 2013), gender, age, weight, and health status (Schrøder-Petersen & Simonsen, 2001).

1.2. Prevention or reducing tail biting

Widely studied and documented methods for preventing tail biting include controlling the health status of the animals, providing adequate feeding (Moinard et al., 2003) and water supply, maintaining ideal temperatures, avoiding the mixing of animal groups, providing enrichment materials, and tail docking (Larsen et al., 2018). Currently, the most important method is monitoring the stocking densities of pens (Grümpel et al., 2018). The provision of enrichment materials (such as straw) keeps the animals occupied (Schrøder-Petersen & Simonsen, 2001) and satisfies their foraging instinct, thereby reducing stress. Tail docking is currently one of the most common preventive measures for the majority of conventionally raised pigs in the EU, although routine tail docking is prohibited by European legislation (Directive 2008/120/EC, 2008), as the procedure is associated with pain and stress and can have long-term implications for the animals' health (Hakansson & Houe, 2020). For this reason, in some European countries, particularly Switzerland, this practice is prohibited by animal protection ordinance (OPAn) (Ordonnance sur la protection des animaux, 2008).

Despite the reduction of some of these risk factors and the application of the previously mentioned preventive methods, outbreaks of tail biting continue to be frequent, particularly in conventional production systems (Larsen et al., 2019). This is due to the fact that tail biting is a multifactorial event (D'Eath et al., 2014; Schrøder-Petersen & Simonsen, 2001; Taylor et al., 2012), and therefore difficult to predict, often being detected only when the pig farmer observes the lesions. At this point, although it is possible to prevent further consequences for the animals, the economic losses are already significant, and the welfare of the animals is already compromised (Statham et al., 2009).

1.3. Behavioural precursors of tail biting

Given the need to predict a tail biting outbreak, the behavioural changes that precede such events have been studied, and several authors have concluded that the level of activity and exploratory behaviour of pigs in the period leading up to a tail biting event is a strong predictive indicator of an outbreak (Statham et al., 2009; Zonderland et al., 2011; D'Eath et al., 2014; Larsen et al., 2016, 2019). It has even been described that the main warning signs can appear days or even weeks before the first signs of blood on the tail are observed. According to D'Eath et al. (2014), the main behavioural changes that precede a tail biting outbreak include an increase in general activity, particularly of the aggressor (biter) (Statham et al., 2009; Zonderland et al., 2011), an increase in tail-in-mouth events without causing damage (Schrøder-Petersen &

Simonsen, 2001), changes in tail posture (tails held down or tucked under) (Statham et al., 2009; Zonderland et al., 2009) and alterations in feeding patterns (Ollagnier et al., 2023).

General activity can be analysed through various behaviours; for example, the time spent in different postures, as demonstrated by Statham et al. (2009), who found that the time spent in the standing posture was higher than the time spent in sitting and lying positions in the four days preceding an outbreak. Social interactions among pigs, which involve mouth and nose contact with a specific body region of another pig, can indicate changes in health status or the occurrence of abnormal behaviours, such as tail biting, in one or more pigs within a group (Alameer et al., 2022; Camerlink & Turner, 2013). The posture of the tail tucked between the legs can indicate a tail biting episode 2–3 days in advance (Zonderland et al., 2009).

2. Computer vision in pig production

Currently, the observation of animal behaviour and activity is carried out by pig farmers during their daily tasks, observing the animals one by one in each pen several times a day, which is time-consuming (Mattina et al., 2023). In large-scale commercial farms, the difficulty of individual monitoring is even more pronounced (Guo et al., 2023), as the limitations of human observation hinder timely diagnosis and intervention for affected animals. Technological advancements have become essential for more efficient and precise animal management (Siegford et al., 2023). A precision farming objective frequently involves the placement of electronic devices on animals, such as ultra-wideband (UWB), accelerometers, microchips for body temperature, microphones, and GPS, as well as passive electronic devices like radio-frequency identification (RFID) chips (Psota et al., 2020). These devices increase the capacity for observation and data recording, thereby enhancing the production and efficiency of daily processes. However, in addition to financial constraints and concerns about durability and hardware management, these devices need to be attached to the animals, potentially influencing their normal behaviour and compromising their welfare.

The development of continuous automated monitoring systems for animal activities based on video recording, that is, CV systems, and achieving methods to manage these data to predict outcomes for each animal, is currently one of the greatest challenges for livestock farming (Mora et al., 2024). CV systems with deep learning methods will enable greater efficiency and effectiveness in management, leading to improved animal welfare (Parmiggiani et al., 2023), as they are non-invasive and allow for in-depth and advanced analysis of collected data, facilitating intelligent decision-making that surpasses the performance of the farmer. These tools are particularly valuable in precision livestock farming, enabling, for example, the

estimation of live weight, the recognition and monitoring of pig behaviour, the observation of sows during farrowing and lactation, the analysis of animal facial expressions allowing the assessment of their emotions (fear, anger) and physical well-being and intentions (signs of aggression), which are transmitted through the positions of the ears, lips, eyes, and other facial muscles (Nie et al., 2024), and the anticipation of pig diseases (Bao et al., 2024). For example, if an aggressive event occurs, pig farmers can receive an alert that includes the location and trajectory of the aggressor, allowing immediate, appropriate measures to be taken to reduce such behaviour (Guo et al., 2023).

However, CV systems are not yet widely used in commercial pig production, as creating these automated monitoring systems requires animal detection, which is very important in precision pig farming and in situations that require pig identity, a crucial and highly challenging step. First, it is difficult to individually separate the animals into images, as pigs spend most of their time in groups. Second, under real production conditions, there are frequent fluctuations in image quality, such as dirt on cameras or changes in lighting (Mattina et al., 2023). Therefore, for CV technology to be considered valid and viable under commercial conditions, its performance must be tested in multiple practical scenarios across different types of production systems and in various housing environments (Gómez et al., 2021).

2.1. Computer vision in the investigation of behaviours

Researchers have successfully identified the links between the health and behaviour of animals. However, simultaneous direct human observation of many pigs and for the necessary duration to obtain useful data requires the work of specialised ethologists in visual classification and is highly time-consuming. Sensors frequently used on animals as data sources are physically attached to them and can therefore influence their behaviour. These devices can be substituted by video surveillance technology, an emerging method that has been widely accepted by the animal production industry (Yang & Xiao, 2020) and research that provides a non-invasive and visible method for detecting animal behaviour (Wutke et al., 2021). Through continuous video recording, it is possible to locate each pig, identify its posture, and quantify its movement, offering an effective way of recognising pig behaviour. However, manual processing of images is a highly laborious, subjective, and time-consuming task, and therefore prone to errors by the annotator (Chen et al., 2021; Han et al., 2023).

In recent years, CV systems have been developed as an alternative or complement to other technologies (Mora et al., 2024) for automated detection of animal behaviours, utilising artificial intelligence to process images and videos (Ji et al., 2022; Devi et al., 2024) to achieve

a database of pig behaviours and provide a foundation for their investigation. Compared to more traditional techniques that require human effort and time, these systems have the advantage of being less costly, not requiring direct intervention with the animals, being programmable to generate continuous information (Chen et al., 2021), and simultaneously monitoring multiple animals with even a single camera (Mora et al., 2024). The video data only need to be converted into quantitative values, carried out by CV systems, to assist behavioural researchers in conducting more in-depth analyses. Establishing a database of pig behaviours requires collaboration between animal behaviour researchers and automation technology experts. It must be determined which behaviours need to be retrieved and at what time intervals they should be quantified.

The development of CV systems utilising deep learning that are able to track pigs and their behaviours individually will enable the automatic analysis and processing of images, facilitating high-throughput phenotyping and generating large quantities of data. These data can be used in the development of smart farming tools for the investigation of animal behaviour and as tools to advance breeding and genetic improvement programs (Fernandes, Dórea, Valente, et al., 2020). Animal tracking refers to the process of determining the position and identity of an animal in consecutive images or videos, enabling the identification or movement trajectories of one or several target animals in sequences of images or videos and maintaining consistent identification of the target animals over time. Most of these tracking algorithms are based on initially detecting objects in the images or video frames and then tracking them (Odo et al., 2024; Xu et al., 2024), for example, to estimate their activity and allow the calculation of proximity to pen mates and location preferences (van der Zande et al., 2021).

To build efficient tools, we need reference datasets for pig farming (Yang & Xiao, 2020) that allow us to design convolutional neural network (CNN) models and compare the performance of different tasks and the comparability between models trained on different datasets. Currently, there is a wide variety of datasets available for pigs but for specific tasks, as each study collects its own images or videos to create its training datasets, which are limited in size and insufficient for CNN models to adapt to different environmental conditions. Medium-sized datasets (tens of thousands to hundreds of thousands of images) collected over a long period of time, including different animal postures, such as lying and standing, different breeds, different ages, and different complex scenes, are needed to create a commonly available reference dataset for use in designing robust models and comparing their performance in different tasks and scenarios (Xu et al., 2024).

2.2. Current state of research

The development of automated video analysis of animal behaviours using CV technology has advanced rapidly in recent years, allowing optimism among ethologists that they will no longer need to spend countless hours manually decoding videos. Instead, a computer processes this information (Siegford et al., 2023). The development of CV is largely attributed to the use of deep learning (DL), based on deep convolutional network models (Amrish & Shwetank, 2024) where a machine learns from raw data (Han et al., 2023), extracting hidden information from complex data such as images or sounds collected in pig production units (Bao et al., 2024). Deep learning has demonstrated a strong capability for feature extraction (Wang et al., 2024) like colour, texture, and shape of an object in the image (Yang & Xiao, 2020), but it requires large quantities of data. For small-scale projects with limited resources, however, it can be challenging to compile such large data sets (Amrish & Shwetank, 2024). The data used for the development of these systems are typically divided into subsets for the training, validation, and testing of the models. The training set is used for the initial development and adjustment of the model, while the validation set is used to evaluate the model's performance during training. For the testing phase, an independent dataset, separate from the initial training and validation sets, is frequently used (Siegford et al., 2023).

In CV systems applied to animal behaviour, the detection of the animal is the primary task (object detection); consequently, its accuracy results in improved accuracy of subsequent tasks (Xu et al., 2024) such as object tracking. The identification of individual pigs is also an important task in the study of behaviours, but it is challenging due to several factors intrinsic to the species, such as the difficulty of predicting the direction of their movements, their normal behaviour of staying close to each other, and their close morphological similarities, often lacking distinct skin/fur patterns, leading to the loss of tracking and frequent identity switches (Odo et al., 2024).

Automatic monitoring of animal behaviours encompasses multiple tasks, such as animal detection, individual animal identification, segmentation, tracking, and the extraction of higher-order features (Pangal et al., 2021). For each task, specific models are designed (Xu et al., 2024) that utilise algorithms based on CNNs (Wei et al., 2023). These models can be used independently or in combination, according to the necessary requirements, to achieve the final goals of monitoring, such as classification networks for animal identification, detection models for animal localisation, segmentation models for isolating animals in images or videos, and tracking models for real-time monitoring (Xu et al., 2024). The criteria for selecting the models to use, in addition to the specific requirements of the application, often include the trade-off between speed and accuracy (Odo et al., 2023), which are interdependent

parameters that are inversely related. For example, attempting to increase accuracy by making the architecture more complex results in the loss of speed, and similarly, increasing speed by making the architecture simpler causes a loss of accuracy. Therefore, it is necessary to select a model that achieves the right combination of balance between speed and accuracy for the application and hardware in question. For example, for real-time monitoring of pigs, speed is important, while in the detection of pig body parts, accuracy is specifically optimised. When greater speed and accuracy are required, the right hardware needs to be used to quickly process the most complex data. The object detection models R-CNN, Fast R-CNN, Faster Region CNN (Faster R-CNN) (Xu et al., 2024), YOLO series (Bao et al., 2024; Xu et al., 2024), SSD, and RetinaNet have demonstrated excellent performance and are among the most popular in deep learning research for animal detection.

Pipeline CV systems frequently involve identifying the location and orientation of each animal and its body parts. The detected parts are used to monitor interactions, such as tail–mouth and ear–mouth pairs (Odo et al., 2023). In pigs, these models have been applied in the research and study of posture (lying, standing, sitting), tail posture (tail curled, tail straight) (Devi et al., 2024; Liu et al., 2023), behaviours such as feeding, drinking, and aggression (Wei et al., 2023), social interactions (Alameer et al., 2022; Gan et al., 2021), individual identification (Kashiha et al., 2013; Wang et al., 2024; Zhou et al., 2023), and the detection of pig body parts (ears, shoulders, tails) (Psota et al., 2019).

In recent years, several CV-based methods have been developed specifically for the tracking and detection of pig behaviours (Table 2 in the annexes). Riekert et al. (2020) designed a deep-learning-based system for detecting the position and posture of pigs using images from standard 2D cameras. They employed Faster R-CNN for object detection and the neural architecture search network for feature extraction, achieving an average precision (AP) of 87.4% for pig position detection and a mean average precision (mAP) of 80.2% for the combined detection of position and posture.

Alameer et al. (2022) developed an automated method that enabled the quantification of the frequency of a pig's head (including the snout) in contact with the rear region of another pig (including the tail). The method involved two phases using videos recorded with 2D cameras collated from experimental trials at two sites. In the first phase, the system detected and identified parts of the pig's body, specifically the head and rear. In the second phase, the interactions between pigs were quantified. In the first phase, a set of 2781 images made up from the two sites were annotated, with each pig in the image manually identified using a bounding box (BB) for the head and another for the rear. In the second phase, 670 images made up of images from the two sites were annotated and scored in spreadsheet format with any

contact between one pig head and another pig rear. The detection model used was the YOLO network to quantify the interaction between all pigs in a pen. They developed a method that examines all the parts of the detected pigs to identify any possible contact and calculate the intersection over union (IoU) between each detected pig's head and the back of the pig. The automated system was validated for classifying interactions between individual pigs within a group and achieved an average precision of 92.65±3.74%.

Other authors, such as Ocepek et al. (2022), have employed YOLO for the detection of pigs or body parts to develop an automated monitoring system for detecting bodies, heads, and tails. The objective of the first part of the study was to recognise individual pigs (in lying and standing positions) in groups and their body parts (head/ears and tails) using machine learning algorithms (feature pyramid networks). In the second part of the study, the aim was to improve the detection of tail posture (straight and curled tail) during activity (standing/moving) using neural network analysis (YOLOv4). The dataset of 583 images was annotated in Labelbox from videos recorded with 2D cameras of groups of weaned pigs using polygons, BBs, and key points as annotation techniques. The model recognised the body of each pig with an accuracy of 96%, while the accuracy for tails was 77% and for heads was 66%, thus achieving human-level precision using the Mask R-CNN-based deep learning model. In the second part of the study, by employing the YOLOv4 neural network, the researchers were able to recognise tail postures with a high level of accuracy (90%). They tested the detection of curled or straight tails in YOLOv4 as an alternative to Mask R-CNN, which achieved an accuracy of 77%. The algorithm detected tail postures with an average precision of 90% using only 30 annotated images. The authors further concluded that their new method could be explored in the detection of behavioural sequences, group synchrony, and the quantification of positive welfare indicators (play, exploration, curled, and wagging tail).

Guo et al. (2023) also developed a pig-tracking model using videos collected with 2D cameras. They chose CVAT as the annotation tool, which annotated segments of videos with active pigs across multiple frames using BBs to locate each animal. They tested three models —JDE (Joint Detection and Embedding), FairMOT, and YOLOv5s with DeepSORT— and concluded that FairMOT achieved the best performance in tracking individual pigs in a real farm setting, with a performance of 80.94%.

Hakansson and Jensen (2023) developed a video-based deep learning approach to detect tail-biting behaviour in groups of pigs without implementing a prior tracking algorithm. This was based on a complete dataset composed of 332,666 images, of which 5,330 exhibited biting behaviour. To achieve this goal, they tested the applicability of two models with their data: a convolutional neural network + long short-term memory (CNN-LSTM) and a CNN-CNN applied

to spatial feature representations extracted from images. Due to its lower complexity and computational workload and a sensitivity of 89% on new data, CNN-LSTM appeared to be the most promising method for predicting tail-biting events.

The CNN-LSTM model was also used to detect the posture and movements of sows (Wang et al., 2021), and was successfully applied by Han et al. (2023) to classify agonistic behaviours of pigs in single-space feeding pens with an accuracy of 96.8%. Wang et al. (2021) classified postural behaviours in sows with accuracies of 95.33% and 92.67% in videos without piglets and in all data (including and not including piglets), respectively. They further tested their model with 500 new videos from their experiment, achieving an accuracy of 90.60%, indicating that the method they proposed can be generalised to new data.

More recently, Odo et al. (2024) proposed a CV system to detect and track all pigs and ear-biting events. They employed a tracking-by-detection system that allowed for the individual identification of pigs involved in an ear-biting event. The Segment Anything Model (SAM) was used to create a dataset for training with the YOLOv5L model. The detector was trained to identify instances where the heads of two pigs were in a posture likely to indicate ear-biting.

3. Objectives and work steps

This work is part of an ongoing project by Agroscope and involves applying existing CV and machine learning (ML) tools to detect the precursors of tail biting in pigs under field conditions, mainly in the preparation of a data set that will allow the creation of a reference database. The proposed work aims to provide a first step towards a methodological framework for preparing a dataset that will allow the creation of a reference database. To achieve our goal, we implemented the following set of sequential work procedures:

A literature survey of existing machine learning models that have already been tested on pigs (individual tracking, laying estimation, etc.) or utilised and tested in different applications, which allowed us to select the annotation and training tools and develop an ethogram that was best suited to the project.

Direct observations of housed animals and their behaviours annotated according to an existing ethogram. This phase served as training for the next steps.

Video acquisition with surveillance cameras installed to continuously record videos of the daily lives of the animals chosen for the study.

Image selection from the set of videos acquired in the previous step, thereby creating our annotation dataset and train the model.

Development of an ethogram and the selection of the annotation and training tools best suited to the project, which will be applied to images automatically selected from the recorded videos.

Annotation of the dataset, which was often achieved through well-defined procedures in various stages.

Training a model, including providing it with data and adjusting its parameters to support accurate predictions.

Validation for evaluating the model's performance.

We expect to contribute to validating our model for the detection and individual identification of each pig, as well as their body parts of interest (tail and head), for the remainder of the project. This is the primary task in CV systems applied to animal behaviour.

4. Material and methodology

4.1. Location of animal experiments

The present project was carried out at the experimental pig farm of Agroscope in Posieux, Switzerland. Agroscope is the Swiss Confederation's competence centre for agricultural and food research (Figure 1), with the mission to conducting research for the benefit of agriculture, providing a basis for federal decision-making, and execute tasks arising from agricultural legislation.

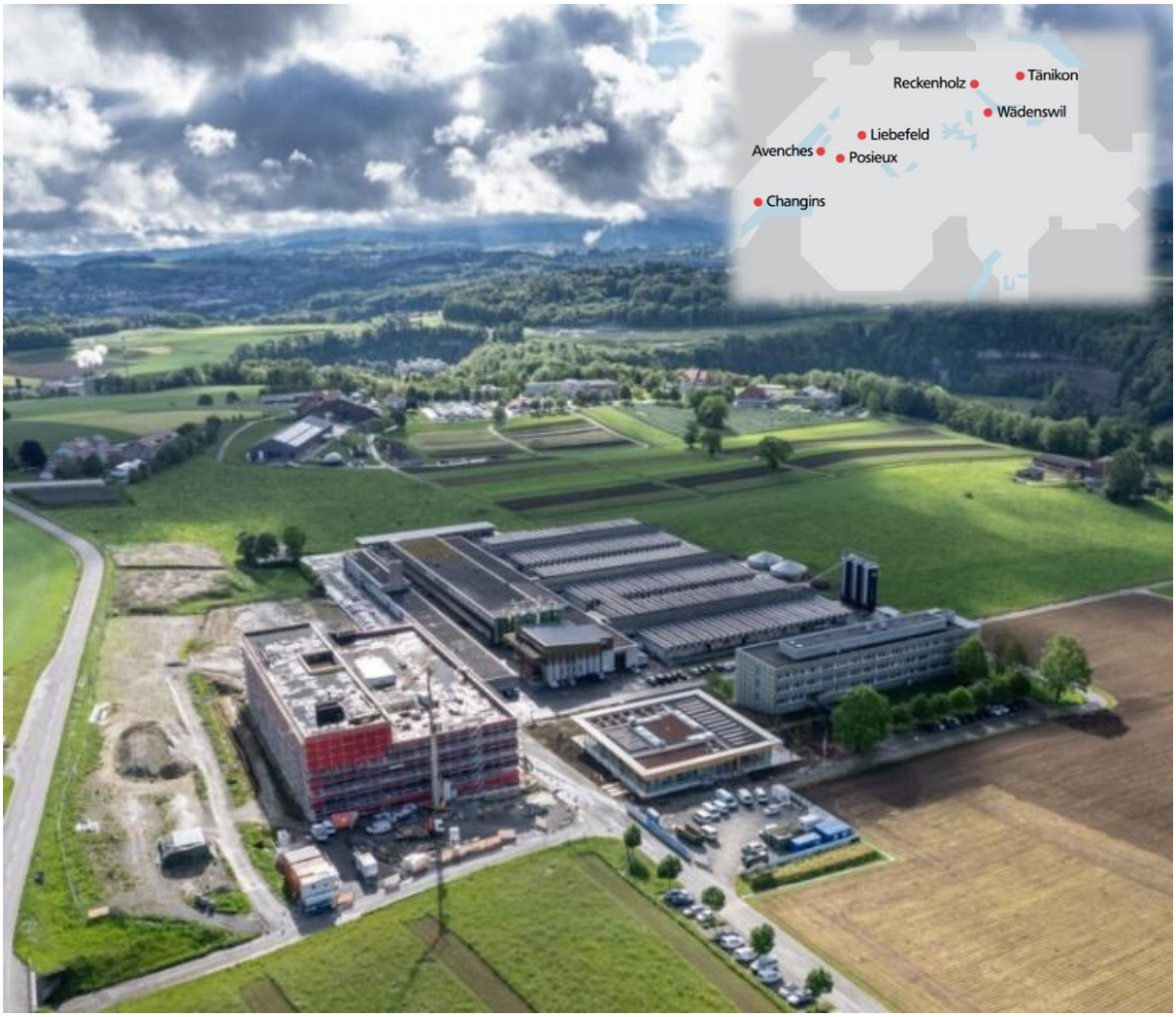


Figure 1. Location and overall image of the Agroscope Research Center (source: https://about.agroscope.ch/wp-content/uploads/2024/05/20240516_travaux_posieux-_Johannmarmy_DJI_0073-2-scaled.jpg).

4.2. Animals and housing

The experiment included 24 castrated male pigs with undocked tails (Ordonnance sur la protection des animaux, 2008) of Swiss Large White breed, born on the farm, and individually identified with an RFID tag in the ear. The observations of the pigs commenced at around 100 kg live weight and 140 days old. These pigs were also monitored by another study conducted at the station, which aimed to evaluate the effects of diets containing different amounts of former food products (FFP). These pigs were slaughtered at a target weight of around 140 kg.

The pigs were housed in the same room and equally distributed in two pens (Figures 2 and 3), with a total area of 32.14 m^2/pen, of which 19.12 m^2 was a resting area, where they received *ad libitum* water distributed in four troughs and dry feed (7:30–23:30) through two automatic feeders (Schauer Electronic MLP) in each pen. Each pen (Figure 2) was composed of two contiguous zones that were identical in dimensions and equipment, which were always available simultaneously for the same group of animals. Each zone included a resting area with a continuous floor (9.56 m^2) and an elimination area with iron grates, two nipple drinkers, two fixed straw baskets and one flexible basket suspended from the ceiling, a feeding station, and two metal chains attached to the side bars (Figure 2).

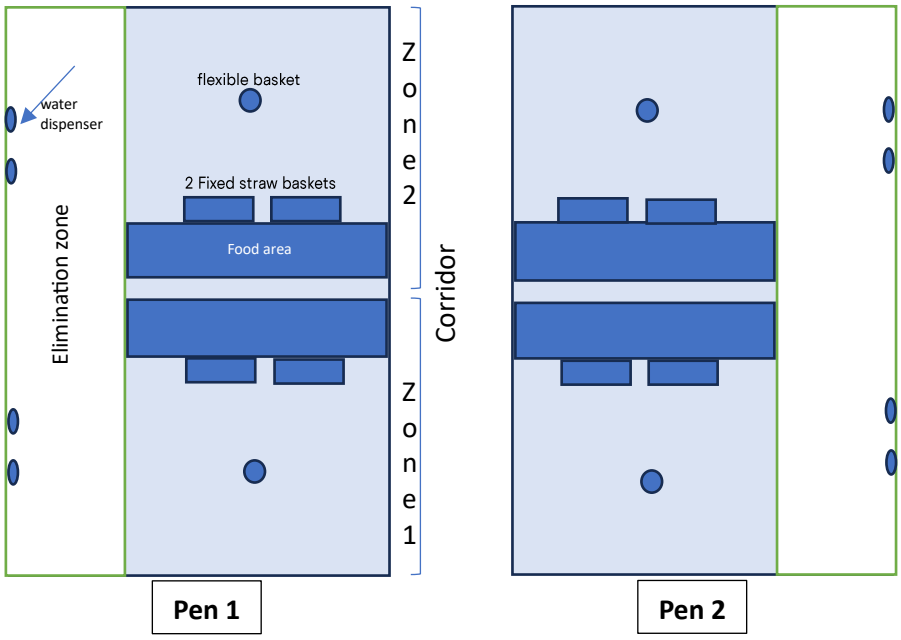


Figure 2. Lay-out of the pens (not to scale). Each pen was composed of two contiguous zones (zone 1 and 2) that were identical in dimensions and equipment. The light blue area corresponds to the resting area.

Figure 3. Fattening room: on the left Pen 1 (above zone 1 and below zone 2), in the middle corridor (view from both sides of the room) on the left Pen 2 (above zone 1 and below zone 2).

4.3. General procedures with animals

The aim of the associated project, which took place at the same time as ours with the same animals, was to evaluate the effects of diets containing different amounts of FFP on growing pigs from 20 to 140 kg. The effects on growth performance, nutrient efficiency, feeding behaviour, energy metabolism, intestinal morphology, body composition, carcass characteristics, meat quality, and metabolic profile (tissues and blood) were evaluated (Mazzoleni et al., 2023). This trial included three groups of 12 growing pigs, each ranging from 20–140 kg. The first group received complete feed with no FFP, the second group received feed with 20% FFP, and the third group received feed with 40% FFP.

For the study of FFP, all pigs were individually weighed once a week every Monday at 7:30 AM. At this time, we also performed individual markings for our project, as explained in Section 4.4. Therefore, there were no further direct interventions with the animals during the course of our work. Pigs from Groups 1 and 2 were filmed. The sole criterion for this selection was the presence of pre-installed camera supports in the pens where these groups were housed.

In addition to the weekly weighing and marking of the animals, as indicated earlier, the general management of these growing pigs was similar in all respects to that of the other growing pigs on this farm. Every morning, between 7:30 AM and 10:00 AM, the pens were cleaned, and the occupational material (straw) was replenished in the respective baskets and distributed on the floor.

4.4. Pig marks

As mentioned earlier, each pig was spray-marked every week during the weighing process, which allowed for no further direct interventions with the animals. This also took advantage of the fact that the pigs were individually separated and contained, enabling the verification of the ear tag number and ensuring the most accurate and precise marking possible. This re-marking ensured that the mark remained clearly visible throughout the filming period.

The colour chosen for the spray was blue, as in practice, blue stands out the best and remains visible for the longest time. Additionally, it did not interfere with other markings in use, namely green for the pigs under study for another project and red for animals undergoing medical treatment. We followed the same symbolism and marking method as Roch et al. (2023), which involved awing lines from one side of the pig's body to the other. This method, in addition to being simple to apply to a small group of animals, proved to allow easy identification of each animal regardless of its posture or location, as this marking was visible both laterally and vertically, as can be seen in Table 1.

As exemplified in Table 1 of the annexes, in each pen, each RFID ear tag corresponded to a number from 1 to 12, and each number corresponded to a symbol marked on the pig. This individual identification remained unchanged throughout the filming period.

Table 1. Symbols marked on the pigs.

Pig Number	Mark			Example
	Shoulder	Back	Rear	
1			I	
2			II	
3			III	
4			IIII	
5		/		
6		/	I	
7		/	II	
8		/	III	
9		/	IIII	
10	I			
11	I		I	
12	I		II	

4.5. Direct observations of animals

The direct observations aimed to train the observer to perform better during the video annotation phase and to verify whether the chosen symbolism for animal marking was suitable

for the work. Additionally, these observations aimed to refine the initially developed ethogram (Table 2), which was based on the ethograms used by Hakansson and Jensen (2023), Liu et al. (2020), and Roch (2021).

Table 2. Initial ethogram.

Behaviour/Posture	Description
Standing	Body supported by four stretched legs (1)
Lying	Lying on one side with no legs tucked underneath the body or lying ventrally with at least two legs tucked underneath the body (1)
Sitting/kneeling	Body supported by hindquarters and stretched front legs or by hind legs and bent front legs (1)
Oral and nasal actions	**Biting** (opens and closes jaw with force at least once, and uses teeth), **Seizing** (opens and closes the jaw without force at least once, not using teeth) or **Manipulating** (keeps the jaw closed, manipulates with the snout, the snout is mobile. The jaw may be relaxed and mobile, but the individual does not use the mouth to grasp) **the pen mates with head-to-head, head-to-trunk, head-to-rear.**
active pig	The actor in the action
passive pig	May be subject to the action
action positive	Social behaviours: Manipulating penmate: touching, sniffing, rooting, licking, biting or chewing the penmate's body, excluding tail and ears (distance from snout to skin 0-5cm) (1)(3)
action negative	Aggressive behaviours: Confrontations between pen mates physically oppose each other with face-to-face or head-to-body contact, with a push that can be gentle to strong
Tail-in-mouth (TM)	Gentle nibbling, suckling, or chewing of another pig's tail, without causing a reaction in the other pig (2)
Tail biting (TB)	Chewing at the tail of a penmate, causing a reaction from the other pig (2)
Basic behaviours	
Feeding	Head in the food trough (1)
Drinking	Head near the water nipple (1)
Urinating/defecating	Discharges urine or faeces
Rooting	touching, sniffing, rooting, licking or biting any part of the pen, walls, bars, feeder, chain, enrichment (straw) (1)(3)
Tail Posture	
Corkscrew	
Tail straight or hanging	
Tail tucked between legs	

(1) Zonderland et al. (2011); (2) Hakansson & Jensen (2023); (3) Munsterhjelm et al. (2016).

The activity of the pigs was observed by a single observer, me, in two groups of 12 animals ranging from 100–140 kg. Nine observations were conducted for each group. The observation of each group was carried out simultaneously for all animals over a period of 30 min on either Monday or Wednesday. The observation time was random but took into account the daily task schedule to minimise disruption to the pigs' daily routine.

To conduct the observations, a tall ladder was placed between the two feeding stations, as shown in the photographs in Figure 3. In this location, the observer did not need to enter the pen and was able to fully visualise both zones of each pen, thereby observing all 12 pigs in each group simultaneously. This setup minimised the observer's influence on the animals as much as possible. Before the observer took the position for the observations, an initial observation was conducted at the level of the animals from outside the pens, walking slowly to check for signs of aggression, tail biting, or other relevant injuries that needed to be noted. Once in the observation post, it was verified that 5 min, the necessary time for the animals to show no further interest in the observer, passed before the direct observation could begin. The objective of this stage was not data collection but rather, as mentioned earlier, the training of the observer. Therefore, since the activity of the pigs at the time of observations was not crucial/of immediate interest, the observation was conducted by group rather than individually, allowing for the registration of all interactions and behaviours according to the chosen ethogram.

4.6. Development of an ethogram for annotations

The final ethogram (Table 3) was developed from the ethogram used for direct observations (Table 2), considering the precursor behaviours of tail biting that were previously studied and considered most relevant. This development also incorporated the advice and experience of animal welfare experts at Agroscope, other ethograms used in CV, field perception, and the ease of annotating behaviours. Additionally, it considered operationalisation and applicability in relation to the subsequent phase and what we considered could be feasibly annotated within the available time for the present work. As the final objective was to measure the activity of the pigs in social contacts, tail biting and tail-in-mouth, and the different postures, we decided to simplify the initial ethogram to make the task of annotation easier, more precise, and objective. Therefore, we grouped the contacts between the animals. We had previously distinguished which pig was active and which was passive, and whether these contacts were social or aggressive, which was difficult to apply in the annotation, as it would be necessary to measure reaction times in social behaviours, distinguishing only when a pig touches its snout

to the snout, body, or back of another congener. Basic behaviours such as urinating, defecating, eating/drinking and rooting were also excluded, as we felt that these behaviours were not representative for measuring pig-to-pig contact and because it would make the task of annotation more difficult, as it would also be necessary to identify areas of interest in the images.

Using this method, we developed the ideal ethogram to achieve the objectives of the project and the parts of the ethogram that would be used progressively at each stage. Thus, in this work, as well as in identifying the individual pigs, we assigned the different postures (Table 3) adopted by each pig identified in the images.

Table 3. Final ethogram.

Behaviour/Posture	Description
Posture	
Standing	Body supported by four stretched legs (1)
Lying	Lying on one side with no legs tucked underneath the body or lying ventrally with least two legs tucked underneath the body (1)
Sitting/kneeling	Body supported by hind-quarters and stretched front legs or by hind legs and bent front legs (1)
Social behaviours	Manipulating penmate: touching, sniffing, rooting, licking, biting or chewing the penmate's body, excluding tail and ears (distance from snout to skin 0-5cm) (1)(3)
head-posterior	Touching, sniffing, rooting, licking, biting or chewing the penmate's posterior, excluding tail
head-trunk	Touching, sniffing, rooting, licking, biting or chewing the penmate's body
head-anterior	Touching, sniffing, rooting, licking, biting or chewing the penmate's head excluding ears
Tail-in-mouth (TM)	Gentle nibbling, suckling, or chewing of another pig's tail, without causing a reaction in the other pig (2)
Tail biting (TB)	Chewing at the tail of a penmate, causing a reaction from the other pig (2)
Ear biting	Biting of one of a penmate's ear with a sudden reaction of the penmate (1)
Agressive behaviours	Pushing (moving a penmate from its location by non-forceful pushing with the head), fighting (forceful pushing of a penmate with or without biting (excluding ear biting and tail biting), chasing (chasing a penmate for at least 2 seconds) (1)

(1) Zonderland et al. (2011); (2) Hakansson & Jensen (2023); (3) Munsterhjelm et al. (2016).

4.7. Technological devices and procedures

Video images were recorded using five Panasonic I-Pro Mega Super Dynamic WV-SW316L surveillance cameras (Figure 4A) installed at different points above each pen (resting area and feeding stations). The cameras were installed at a height of about 2.5 m above the floor of pens in pre-existing supports (Figure 4B, and C) and at different angles relative to the animals (Figure 5): 3 cameras in Pen 1 with an angle of less than 45° almost in top view (zone 1: camera 1, feeding stations: camera 2, zone 2: camera 4; Figure 4B) and two cameras in Pen 2 with an angle of about 45° (zone 3: camera 6, zone 4: camera 5; Figure 4C).

Figure 4. Model of the cameras used (A) and locations in the pens: B - Pen 1; C - Pen 2.

Figure 5. View of camera 4 almost in top view (A) and view of camera 5 about 45° (B).

In this positioning, we were able to capture the maximum area of each pen with the least background and occlusions (animals, walls, dividers, bars, other equipment) possible. The cameras were connected to a data storage device, QNAP NAS (Network-Attached Storage),

which accommodates digital files, such as videos, with a storage capacity of 12 TB. These files were managed by the VioStor (System Network Access Storage) software (Figure 6), which allowed for the playback of recorded videos with options to search by date, time, or camera. Through this software, we configured the IP cameras, including resolution, frame rate, and recording schedule. The videos were recorded from 7:00 AM to 8:00 PM from April 25, 2023, to June 12, 2023 in both pens under varying lighting conditions (natural light through windows, artificial light, and twilight). To ensure that all videos could be stored without compromising image quality for the subsequent stages, the capture parameters chosen were a frame rate of 15 fps and a resolution of 320 x 240 pixels. The videos were recorded in AVI format with a duration of 15 minutes each.

Figure 6. VioStor display the five cameras in real time.

As it was not possible to use the videos directly from the storage software or to connect the QNAP NAS directly to the computer with the software to be used later, it was necessary to transfer the data using external drives from the QNAP NAS to the PC on which the following analyses were done. This operation took approximately 6 weeks.

4.8. Videos acquisition

A total of 12,415 videos were recorded, corresponding to approximately 3,103 h of footage. The number of animals recorded in each frame varied, as the two zones within each pen were permanently interconnected, allowing the animals to move freely from one zone to another. Consequently, the animals were visible in one camera or another. Some recording issues were identified with camera 6, such as image blockages making it impossible to read the videos continuously, which could compromise the quality of the images for subsequent processing. Therefore, 2,472 corresponding videos from this camera were removed from the sample.

We found it necessary to convert the remaining 9,943 videos from AVI to MP4. Although the quality of AVI videos is slightly superior, MP4 files are smaller and more efficient for storage, facilitating their operationalisation and enabling their use in subsequent annotation software. This conversion also helped reduce the training time of the CNN and thus increase computational efficiency (Wutke et al., 2021). After converting the files to MP4 format, a new selection of videos was performed. The objective of this task was to remove videos from the sample that were not relevant to the annotation stage. Therefore, all videos in which no animals were present or in which staff were inside the pens, particularly during the morning cleaning period (a total of 347 videos), as well as videos in which the animals were not marked, from April 25, 2023 to May 1, 2023, and from June 8, 2023 to June 12, 2023 (a total of 2,235 videos), were also removed from the sample, totalling 2,582 videos. Thus, our final set contained the following videos: camera 1: 1841 videos, camera 2: 1846 videos, camera 4: 1840 videos, and camera 5: 1834 videos. This totalled 7361 videos of 15 min each.

4.9. Selection of the dataset for annotations

For the annotation stage, we had to select a subset of the most representative images for two reasons. First, a machine learning model can only be as good as the data it is trained on. Second, the dataset is very large, and it is impractical to use 70% of the images as a training set and perform annotations on them. Therefore, we used the embedding method, whereby images are transformed into a feature space using a Resnet-18 convolutional neural network. Each image is transformed to a vector of dimensions (1 × 512). We call the representation of the image in the feature space 'Embedding'. A dimensionality reduction technique (van der Maaten & Hinton, 2008) was applied to the vectors of size 512 to represent them in the 2-dimensional space. More information about this method can be found in a previous work (Sener & Savarese, 2018). For practical reasons, we used a commercial tool called 'Lightly' to

perform these embedding calculations and selection. In practice, the images are stored in a blob storage in Microsoft Azure, and the computation of embedding can be run in a cloud virtual machine with GPU capabilities. Lightly provides a docker container to run the embedding computations. Upon completion of the computation, the results can be viewed in a web interface, and the selection can be made (Figure 7).

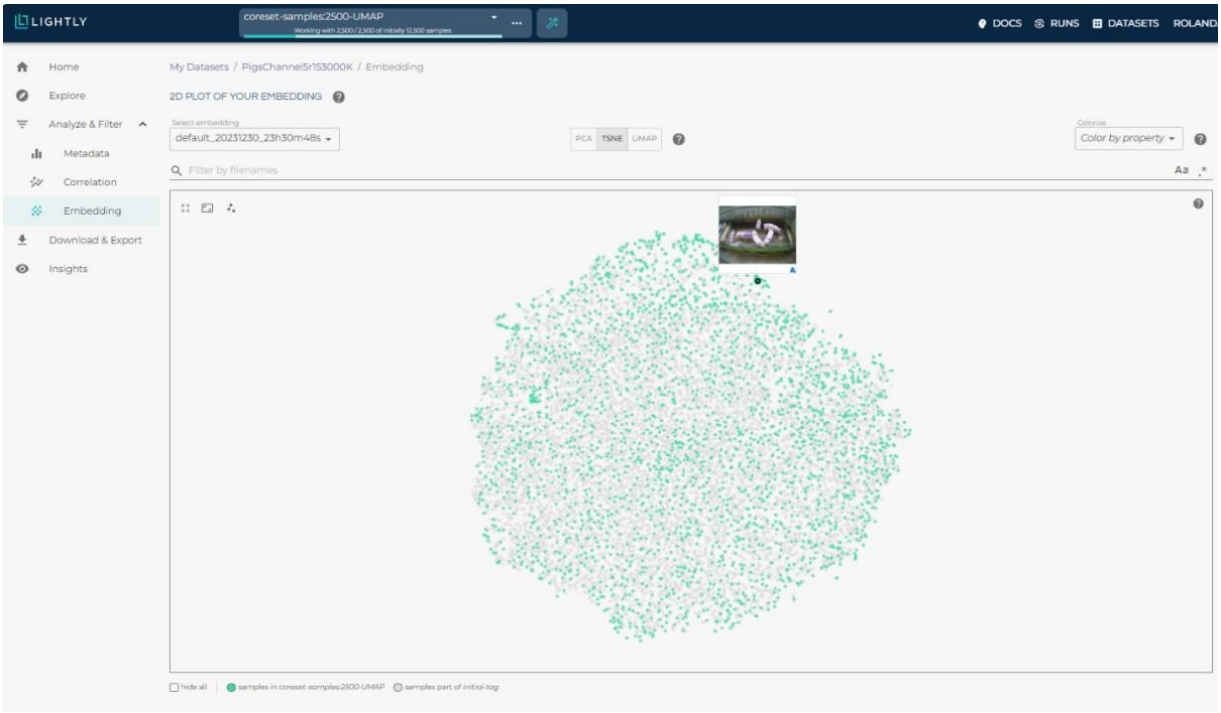


Figure 7. Lightly's web interface: Each point represents a frame, that is, 412,650 points of which 2,500 marked in green are the more diverse frames for annotation.

From the initial number of 412650 frames (1834 videos), we decided to select 2500 images, the most representative. The picture above shows the chosen images in green, whereas the grey dots represent the images that were not selected. As shown in the figure 7, the green images span the whole space of the image distribution, which improves the model's generalisability and reduces the initial set of annotations.

4.10. Choose the annotation tool

According to Amrish and Shwetank (2024), the choice of image annotation tool is generally based on three criteria: its popularity, the availability of stand-alone or web-based interfaces, and the ability to integrate with cloud-based services. We considered the criteria mentioned above, as well as the fact that they were available free software, allowing us to test the

annotation with a few videos before making our final choice. The three annotation tools tested were the VGG Image Annotator (VIA), Darwin, and CVAT.

VIA was the first to be tested. It did not accept files in AVI format; thus, at this point in the work, we converted our videos into MP4 format. VIA is open source software developed by the Visual Geometry Group (VGG); it is a simple and standalone manual annotation software for image, audio, and video (Dutta & Zisserman, 2019). VIA runs in a web browser and does not require any installation or setup. Regarding user experience, I find this tool to be very unintuitive, particularly regarding the definition of classes and their attributes. Its toolbar is not visible, and often, the object to be annotated is partially hidden by the window that contains the options for setting the tags. In general, its display is very confusing (Figure 8).

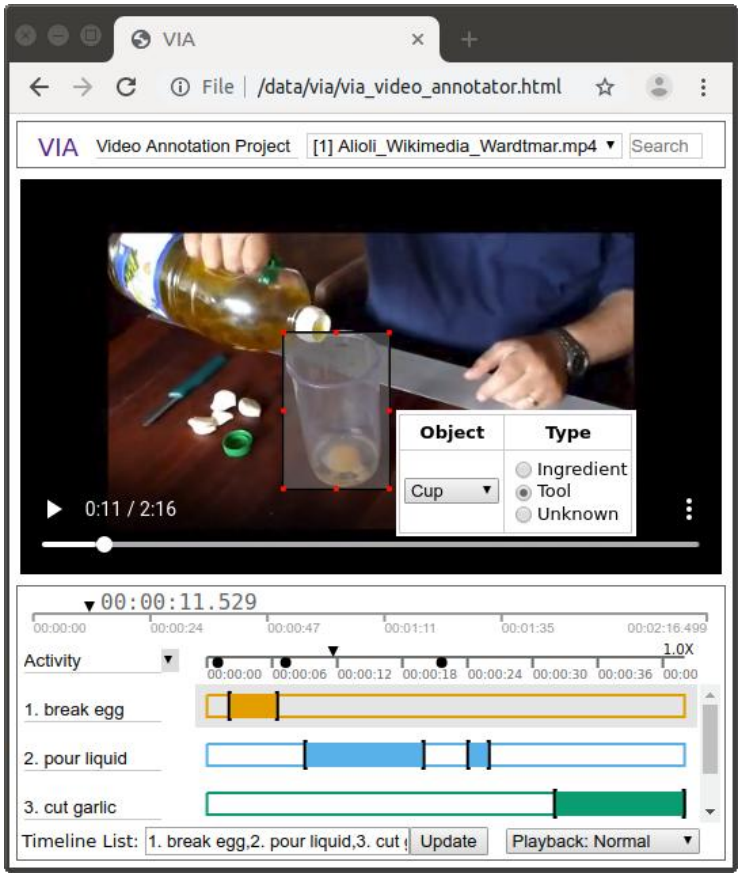


Figure 8. Video annotation with VIA (source: (Dutta & Zisserman, 2019).

The V7 Darwin software (V7, s. d.) was the second tool to be tried. Although it is not open-source software, it was already being used by Agroscope for other projects to annotate images. When the MP4 videos became available, we studied and experimented with Darwin's video annotation. This tool is more intuitive than the previous one, but even so, for beginners, annotating videos is a bit complicated because for tracking, the user has to draw a new BB in

every frame to be annotated, making the process more time-consuming. During the course of the present project, the licence for V7 Darwin expired. Therefore, we decided to try annotating the videos and images using CVAT, another annotation tool mentioned in some works (Guo et al., 2023).

CVAT eventually became our tool of choice. It is a free, online, open-source, interactive video and image annotation tool that speeds up the process of annotating video and images for use in training CV algorithms (Boesch, 2023c). During our experimentation with the application using videos, we found that CVAT was very intuitive to use, with a simple interface and semi-automatic annotation that was easy to apply, which allowed us to increase the efficiency of the task. For these reasons and to follow Pangal et al. (2021) and Guo et al. (2023), we used CVAT (computer vision annotation tool) as the annotation application.

CVAT and the other tools were initially tested on videos. However, we later decided to annotate images because we wanted to test the 'object detection' models first, which only accept images, and images are easier to annotate than videos. Nevertheless, being able to select the most diverse images in all the videos yielded a greater variety of information, as a video is a sequence of images, and very often, the pigs did not move, making the images similar.

4.11. Annotation

As mentioned in the previous section, the detection of the animal was our primary task (object detection); consequently, its accuracy resulted in better precision for subsequent tasks. However, to achieve higher accuracy in this task, it is necessary to train these CV algorithms, which fundamentally depends on the quality, size, and accuracy of the manually annotated datasets used for their training (Amrish & Shwetank, 2024; Pangal et al., 2021; Xu et al., 2024). The primary objective of the annotation procedure is to create a ground truth dataset. Image annotation of a dataset involves creating labels for the important parts of the images (ground truth), known as classes. These labels provide the system with information about the classes of objects present in each image, as well as their shapes, locations, and additional attributes, such as posture. The images with this information are used to train machine learning models, a process known as supervised learning (Amrish & Shwetank, 2024; Boesch, 2023b). The trained CV model is then implemented to predict and recognise these predetermined features in new, unannotated images (Boesch, 2023b).

To create labels, various annotation techniques can be used to provide the appropriate means for annotating images, such as BB annotation, point annotation, cuboid annotation, polygon

annotation, polyline annotation, and ellipse annotation. These techniques are provided by data annotation tools and are chosen based on the objectives of each model and the characteristics of the data.

The annotation was performed solely by the operator, who conducted the previous stage of direct animal observation. The primary objective of this stage was the individual identification of each pig, as well as their head and tail. To achieve this, we decided that each number marked on each pig (Table 1) corresponded to a class (12 classes: Pig 1 to Pig 12), the head to another class, and the tail to a different class. Additionally, another class was created for pigs whose marked number (no ID) was impossible to identify, as explained in the next section. These 15 classes were tested using the annotation tools.

4.11.1. Classes (labels) and attributes used (according to the ethogram)

Before starting the annotation, we defined the classes, as mentioned above. These are the characteristics we desired to be recognised by the system (Boesch, 2023a), which allowed us to identify each animal individually, indicating their head and tail (classes) and assigning each of them properties, called attributes, which defined the posture of each individual according to the defined ethogram (Table 3): lying, sitting/kneeling, and standing.

Definition of Classes

Class "no ID": This tag was used for pigs that had no identification (this was the case for two pigs in each pen that were still being used and tagged for the other study until 8 May 2023, but which had already been assigned numbers) and for pigs whose identification mark was not visible or partially visible, so that it was impossible even for the annotator to identify them (Figure 9). The attributes were also annotated for this class.

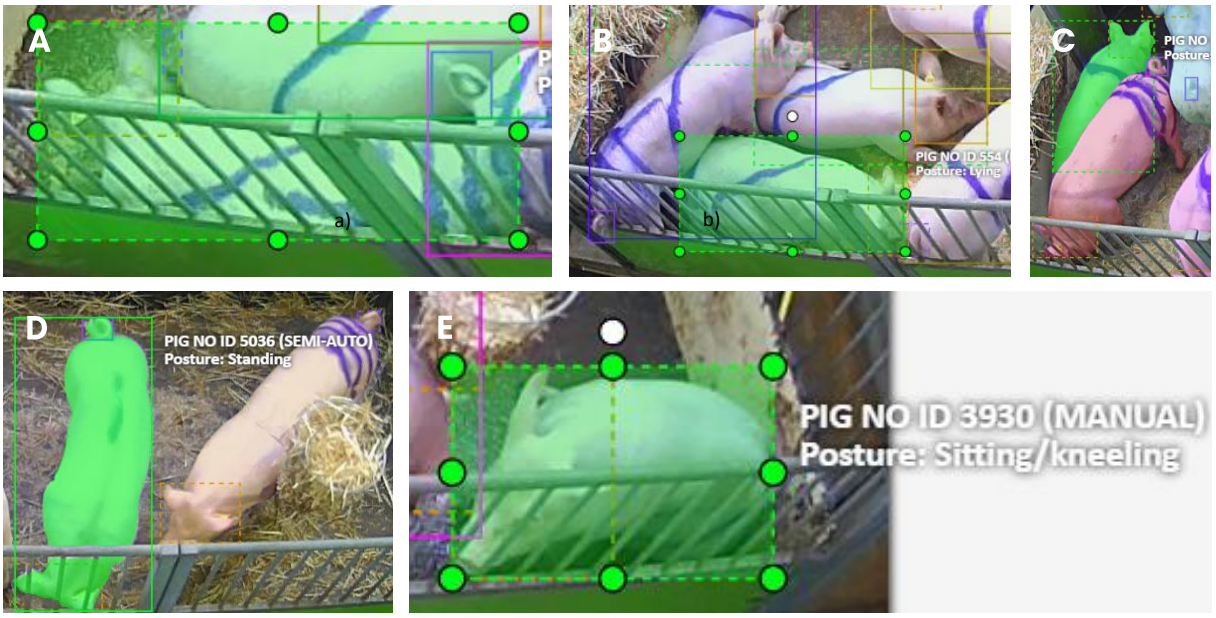


Figure 9. Examples of pig cases identified with the class 'no ID' (green). Their identification mark was not visible in D and E or partially visible but not possible identify them in A, B and C.

Classes "Pig 1" to "Pig 12": 12 classes were annotated to identify each pig with its marked number, as show in Table 1, which was identifiable to the annotator.

Class "Head": This class comprised the two ears, the muzzle, including the nose, and the cheeks, that is, making a square from the tip of the ears, through the base of the ears, to the end of the throat, and forward to the tip of the muzzle (Figure 10).

Figure 10. Heads annotated in different views, with two ears, the muzzle (including the nose) and the cheeks visible.

Class "Tail": This indicates the distal end to the proximal end, including the base of the tail, when in the straight or hanging positions. In the corkscrew posture, when at least the distal end and the proximal end were visible, no other part of the tail was hidden by another object (Figure 11).

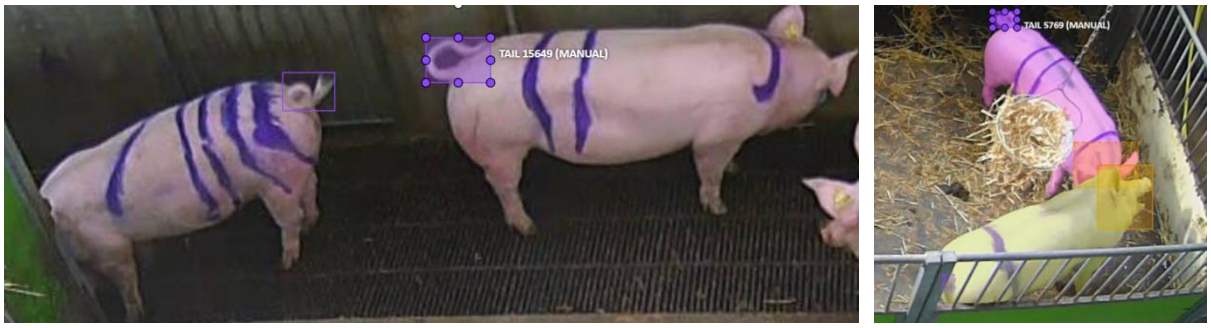


Figure 11. Tails annotated from the distal end to the proximal end, including the base of the tail, when in straight or hanging positions.

The gregarious instinct of pigs meant that most of the time they were in groups (Mattina et al., 2023), making it difficult to separate them individually in the images; often one pig hid another (fully or partially), and even their postures could hide (partially or completely) parts of their own body, namely their head and tail. Sometimes, their location in the pen also caused occlusions, such as lying against the railings, where it was often possible to identify the pig, but it was completely hidden. Figure 12 shows some of these situations.

In these cases, CVAT made it possible to classify the annotated object as hidden. This "Occluded property" function is defined as an attribute used if an object is occluded by another object or is not fully visible on the frame (CVAT, n. d.). To always apply the same criteria throughout the annotation, we defined when to consider an object occluded.

Classes "No ID" and "Pig 1" to "Pig 12" was noted as "Occluded property" (Figure 12) when the identification was visible to the annotator but was partially hidden by other objects (park fence, straw baskets, other animals).

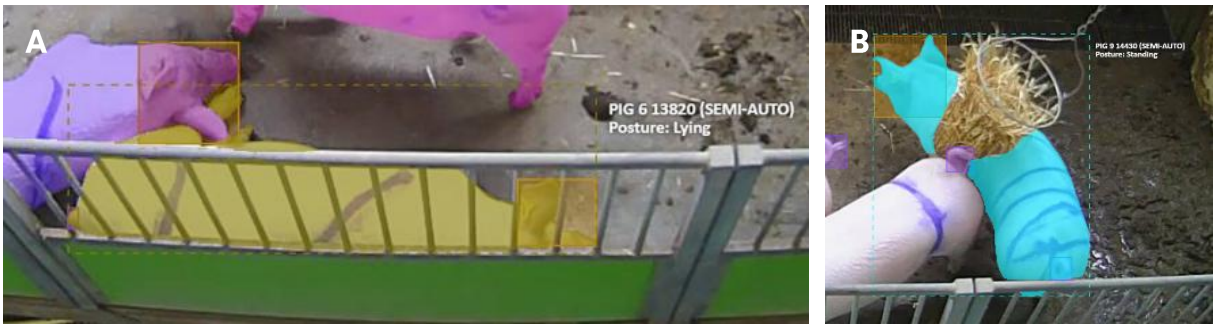


Figure 12. Pig 6 (yellow) annotated as an occluded property (bounding box with discontinuous line), the mark is visible but hidden by park fence (A). Pig 9 (blue) annotated as an occluded property (bounding box with discontinuous line); the mark visible but hidden by straw baskets (B).

Class “Tail” is annotated “occluded property” in cases the tail was not entirely visible, for example, when it was in a straight or hanging position or when it was visible to the annotator but was partially hidden by other objects, such as pen railing (Figure 13), straw baskets, and other animals. In the curled position when it is partially hidden.

Figure 13. Tail (red arrow) and head (green arrow), partially hidden by pen railing.

Class “Head” was annotated as an occluded property when the two ears and snout were not visible (Figure 14) or were visible to the annotator but were partially hidden with other objects (pen railing, straw baskets, animal's posture) as show in Figure 14.

Figure 14. (A) the two ears and snout were not visible or partially visible hidden by pen railing (red arrow). (B) the two ears and snout were not visible hidden with animal’s posture (orange BB).

4.11.2. Annotation techniques used

From the sample of 2,500 frames, 280 were annotated with a BB and 520 with semi-automatic segmentation using SAM (Kirillov et al., 2023). SAM allows for the semiautomatic creation of polygons with a pre-trained foundation model. The interactive platform uses a foundation model, such as Segment Anything (Kirillov et al., 2023) to obtain a mask for an object using positive and negative points that determine the shape of the polygon (the positive points are those related to the object) that delimits the outline of the animals (Boesch, 2023a). In this project, we used the mouse to click on the object to be identified, and the algorithm automatically identified the equivalent pixels, assigning them the same colour and thus drawing a mask over the entire object (Figure 15).

The initial idea was to annotate only with semi-automatic segmentation, as this would allow for finer and faster annotation; however, contrary to what Odo et al. (2024) reported, for our type of images, in which animals are very close and often hidden by each other, it was necessary to edit each mask. This increased the annotation time three times. Thus, we decided to use BB to save time.

Figure 15. Frame annotated with SAM: each pig with the colour of the mask corresponding to its class. For example, a standing pig in yellow corresponds to class Pig 6, and a lying pig with a red mask corresponds to class Pig 11 (red arrows).

Annotation with Bounding Box

Annotating an object with BBs involves drawing a rectangle (box) around the object to be detected with the help of the mouse and selecting the class label for each drawn box; it allows for selecting the attributes. In this work, we drew BB with two points that coincided with the two diagonally opposite ends of the pigs: heads and tails. After drawing all the boxes and before moving on to the next frame, we checked that the BB of the head and tail of a given individual was entirely included in the BB of the corresponding pig. This tool is easy to apply, but as Psota et al. (2019) and Brünger et al. (2020) pointed out, each box includes a lot of background. Thus, in our case, where the head and tail needed to be assigned to the respective individual, very often, the BB of a pig included other complete BBs such as two tails or two heads, as shown in Figure 16, which made it difficult to assign the parameters for the inclusion of the tail and head in the corresponding individual.

Figure 16. (A) In the same frame, there are several situations in which there are two BB heads and two BB tails on the same BB. (B) Two sub-reply bounding boxes.

Even if we rotate the BB along the animal's axis, even though the background is smaller, the situation of having a head or tail on the same BB as a pig remains. Despite this finding, we decided to keep the 280 frames with this annotation and continue with the remaining frames with the SAM model for the identification of each pig and BB for the tail and head and compare the accuracy of the two techniques.

Annotation with SAM

Next, we annotated 520 frames using SAM, which involved drawing a mask by clicking with a mouse on the object to be identified. Once the mask was drawn, we edited it using 'brush' and 'eraser' tools with pixel-level precision (Kirillov et al., 2023), as the complexity of the majority of the images meant that not all animals were always fully included, or multiple parts of different animals that were close together were included in the same mask, along with parts of the enclosure material (Figure 17).

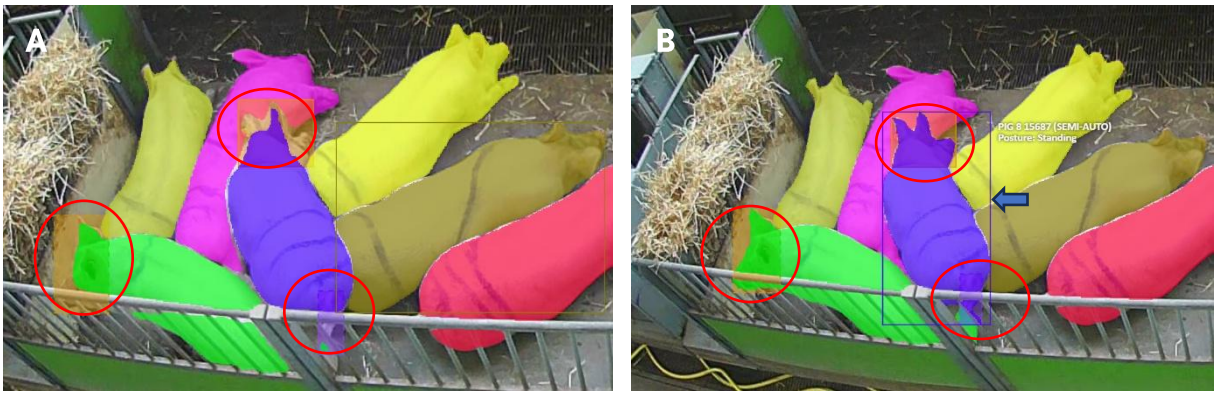


Figure 17. (A) Frame annotated with SAM before mask editing (areas to be edited identified with a red circle). (B) Same frame after editing (edited areas identified with a red circle) and the automatic bounding box of the respective pig identified with a blue arrow.

The corresponding class and attribute were then assigned in the same way as in the BB technique. The occluded objects were also identified for the masks using the same criteria. After editing, a BB is automatically formed around the identified object. Before completing the frame, as with the BB annotation, we checked that all the BBs of the head and tail were correctly inserted into the automatic BB of the respective pig (Figure 17, right image).

4.12. Model training and validation

For this phase, only the 280 images annotated with BBs were exported in YOLO format. The training model chosen for the detection of individual pigs, head, and tail was YOLO (You Only Look Once) (Ultralytics, n. d.) in the eighth-generation version of the YOLO family of models. Based on the latest advances in deep learning and CV, YOLOv8 is state of the art for its speed and accuracy, as it is the leading model of its kind (Wei et al., 2023). YOLOv8 contains five different configurations, YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l, and YOLOv8x, which gradually increase according to the depth and width of the network and, for this reason, increase as well as computing resources needed. The version used was YOLOv8m because it provides a good balance between computational complexity and accuracy.

From the set of exported images, 70% were used to train the model, and 30% were used for its validation. There are different metrics to express and evaluate the results obtained from the object detection model (Ultralytics, n. d.), indicating the precision and effectiveness with which the model can identify and locate objects in images, and allowing an understanding of how the model handles false positives (when an object of interest is predicted but is not actually present) and false negatives (when an object of interest is not predicted but is actually present). These insights are crucial for evaluating and improving the model's performance.

In this project, we used numerical metrics for each specific class, as well as visual metrics that can provide a more intuitive understanding of the model's performance, as described below (Ultralytics, n. d.).

- **Instances:** This provides a count of how many times the class appears across all images in the validation set.
- **P (Precision):** The accuracy of the detected objects, indicating how many detections were correct, that is, quantifies the proportion of true positives among all positive predictions, assessing the model's capability to avoid false positive
- **R (Recall):** The ability of the model to identify all instances of objects in the images, that is, calculate the proportion of true positives among all actual positives, which measures the model's ability to detect all instances of a class.
- **mAP50:** Mean average precision calculated at an (IoU) threshold of 0.50. It is a measure of the model's accuracy, considering only the "easy" detections.
- **mAP50-95:** The average of the mean average precision calculated at varying IoU thresholds, ranging from 0.50 to 0.95. It provides a comprehensive view of the model's performance across different levels of detection difficulty.
    - **Average Precision (AP):** AP computes the area under the precision-recall curve, providing a single value that encapsulates the model's precision and recall performance.
    - **Mean Average Precision (mAP):** mAP extends the concept of AP by calculating the average AP values across multiple object classes. This is useful in multi-class object detection scenarios to provide a comprehensive evaluation of the model's performance.
    - **Intersection over Union (IoU):** IoU is a measure that quantifies the overlap between a predicted bounding box and a ground truth bounding box. It plays a fundamental role in evaluating the accuracy of object localization.
- **F1 Score Curve (F1_curve.png):** This curve represents the F1 score across various thresholds. Interpreting this curve can offer insights into the model's balance between false positives and false negatives over different thresholds.
    - **F1 Score:** The F1 Score is the harmonic mean of precision and recall, providing a balanced assessment of a model's performance while considering both false positives and false negatives.
- **Precision-Recall Curve (PR_curve.png):** An integral visualization for any classification problem. This curve showcases the trade-offs between precision and

recall at varied thresholds. This becomes especially significant when dealing with imbalanced classes.

- **Precision Curve (P_curve.png):** A graphical representation of precision values at different thresholds. This curve helps in understanding how precision varies as the threshold changes.
- **Recall Curve (R_curve.png):** Correspondingly, this graph illustrates how recall values change across different thresholds.
- **Confusion Matrix (confusion_matrix.png):** The confusion matrix provides a detailed view of the outcomes, showing the counts of true positives, true negatives, false positives, and false negatives for each class.
- **Normalized Confusion Matrix (confusion_matrix_normalized.png):** This visualization is a normalized version of the confusion matrix. It represents the data in proportions rather than raw counts. This format makes it simpler to compare the performance across classes.
- **Validation Batch Labels (val_batchX_labels.jpg):** These images depict the ground truth labels for distinct batches from the validation dataset. They provide a clear picture of what the objects are and their respective locations as per the dataset.
- **Validation Batch Predictions (val_batchX_pred.jpg):** Contrasting the label images, these visuals display the predictions made by the YOLO 8 model for the respective batches. By comparing these to the label images, you can easily assess how well the model detects and classifies objects visually.

5. Results and discussion

During the literature review phase, we observed, as noted by Amrish & Shwetank (2024) and Pangal et al. (2021), that the annotation procedures, including the development of clear and specific techniques and methodologies, were frequently omitted in the descriptions of developed machine learning models. This omission makes it difficult to replicate or fully evaluate the research. Given the rapid development of state-of-the-art CV algorithms, it is challenging to stay updated and even more difficult to adapt new algorithms to specific animal applications, as most datasets are not publicly available (Han et al., 2023; Li et al., 2022; Siegford et al., 2023). Access to rigorously annotated datasets that contain sufficient metadata to explain how and where the data were collected, annotated, and processed will enable others to create new CV systems for behavioural analysis without having to start from scratch.

Despite the increasing amount of datasets produced by livestock farms, providing a wide range of information, CV systems for behavioural phenotyping are still in the phase of evaluating the performance of algorithms for prediction. There is a lack of validation studies on the predictive capability of these CV systems across this variability of information (Han et al., 2023). This can be justified by the fact that the models are still not sufficiently robust (Siegford et al., 2023), as there is still a need for the creation and sharing of benchmark datasets for behaviour detection, including image data, annotations, metadata, and baseline analyses for comparative evaluation. These datasets should contain many images but should also be sufficiently varied and well-annotated, representing animals of all types, that is, in different growth stages and group sizes, various breeds and colours, and under various environmental conditions to capture variations in dust, occlusion, flooring, lighting, etc. (Han et al., 2023; Siegford et al., 2023).

5.1. Annotation

In the 800 frames annotated in this study, a total of 15,646 objects were labelled (Table 4), of which 11,530 were labelled using the BB technique, and the remaining 4,116 were labelled using segmentation (SAM). In the 280 frames annotated using only BBs—that is, the dataset used for training the object detection model—a total of 4,873 objects were labelled. Of these, 3,119 labels corresponded to the ‘Head’ and ‘Tail’ sets, and only 1,754 were in the remaining classes (Pig 1 to Pig 12 and Pig no ID) (Table 4).

For the 520 frames annotated with SAM, the ‘Head’ and ‘Tail’ classes had zero annotations; as for these two objects, the BB technique was always used across all annotated frames. Nevertheless, we observed that for a total of 4,116 objects annotated with the classes Pig 1 to

Pig 12 and Pig no ID, there were 6,657 corresponding head and tail labels. This significant difference in the number of annotated objects for each class is justified by the fact that for each of the classes Pig 1 to 12 and Pig no ID, there were almost always corresponding 'Head' and Tail classes. Therefore, these two classes account for 62.48% (9,776 labels) of the total annotated objects, while the remaining 13 classes represent only 37.52% (5,870 labels).

The classes 'Pig 12' and 'Pig 9' had zero labels when we used the BB technique, as the 280 frames annotated with this technique corresponded exactly to the period when there were 2 pigs in the pens that had not yet been marked, as explained earlier, and were therefore identified as belonging to the class 'Pig no ID'.

Table 4. Annotation statistics from CVAT. Labels annotated for each annotation technique. Classes Head and tail annotated in the 800 frames with the same technique, Bounding Box.

Annotation technique	Bounding Box (BB)	Shapes Mask (SAM)	TOTAL
N° annotated frames	280	520	800
Label/Classe			
Pig 1	182	357	539
Pig 2	102	279	381
Pig 3	101	283	384
Pig 4	78	172	250
Pig 5	80	229	309
Pig 6	138	314	452
Pig 7	164	221	385
Pig 8	193	364	557
Pig 9	0	305	305
Pig 10	154	297	451
Pig 11	178	318	496
Pig 12	0	211	211
Pig no ID	384	766	1150
Sub-Total	4873		
Head	5003 (800 frames)		5003
Tail	4773 (800 frames)		4773
TOTAL	11530	4116	15646

5.2. Model training and validation

The set of 280 frames annotated exclusively with BB was used to train in the chosen object detection model. Despite the small sample of 4,873 annotated objects, the YOLOv8m model

achieved 93% accuracy in predicting heads and 84% accuracy for tails, as illustrated in the confusion matrix shown in Figure 19, where it is simpler to compare the performance across classes. These results are promising if we compare, for example, with the results obtained by Alameer et al. (2022), who used 2502 images to train the YOLOv4 model, achieving 87.2% accuracy for tail detection and 85.3% for head detection. Ocepek et al. (2022) was better results for tail detection using YOLOv4 with 90% accuracy using only 30 images with visible tails during the pigs' active phase, using the Mask R-CNN based deep learning model they predicted the tails with an accuracy of 77% and for heads was 66% in 583 images annotated. For the remaining classes, the accuracy was lower, indicating that the model had difficulties in their detection. However, as can be observed in Table 5, accuracies above 70% were achieved for some classes.

In general, the classes for which higher accuracy was achieved correspond to those with more annotated objects (Table 5); that is, more ground truths were provided for the model's training (Figure 18), which is unequivocally the case for the 'Head' and 'Tail' classes, as expected. For the remaining classes, this was not always the case, which can perhaps be explained by the fact that some annotations were prone to confusion, either due to their similarity, the same marking area on the pig's body, or imperfect annotations that made them confusing.

Table 5. Summary of the relationship between objects annotated with Bounding Box (ground truth) and the accuracy (Accuracy) obtained for each class (Label).

Label	Bounding Box (BB)	Accuracy (%)
Pig 1	182	77
Pig 2	102	67
Pig 3	101	70
Pig 4	78	55
Pig 5	80	45
Pig 6	138	72
Pig 7	164	64
Pig 8	193	84
Pig 9	0	0
Pig 10	154	74
Pig 11	178	62
Pig 12	0	0
Pig no ID	384	77
Head	3119	93
Tail		84
TOTAL	4873	

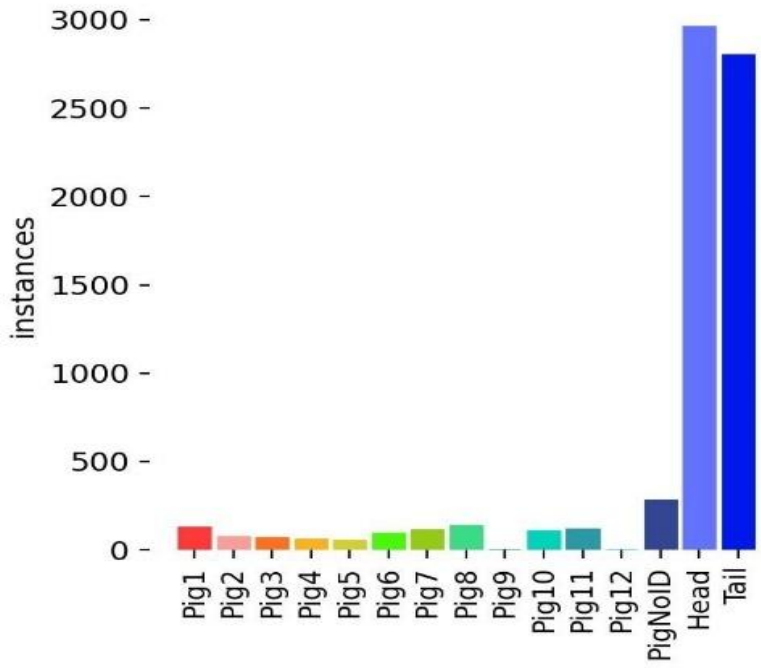


Figure 18. Count of how many times the class appears across all images in the validation set.

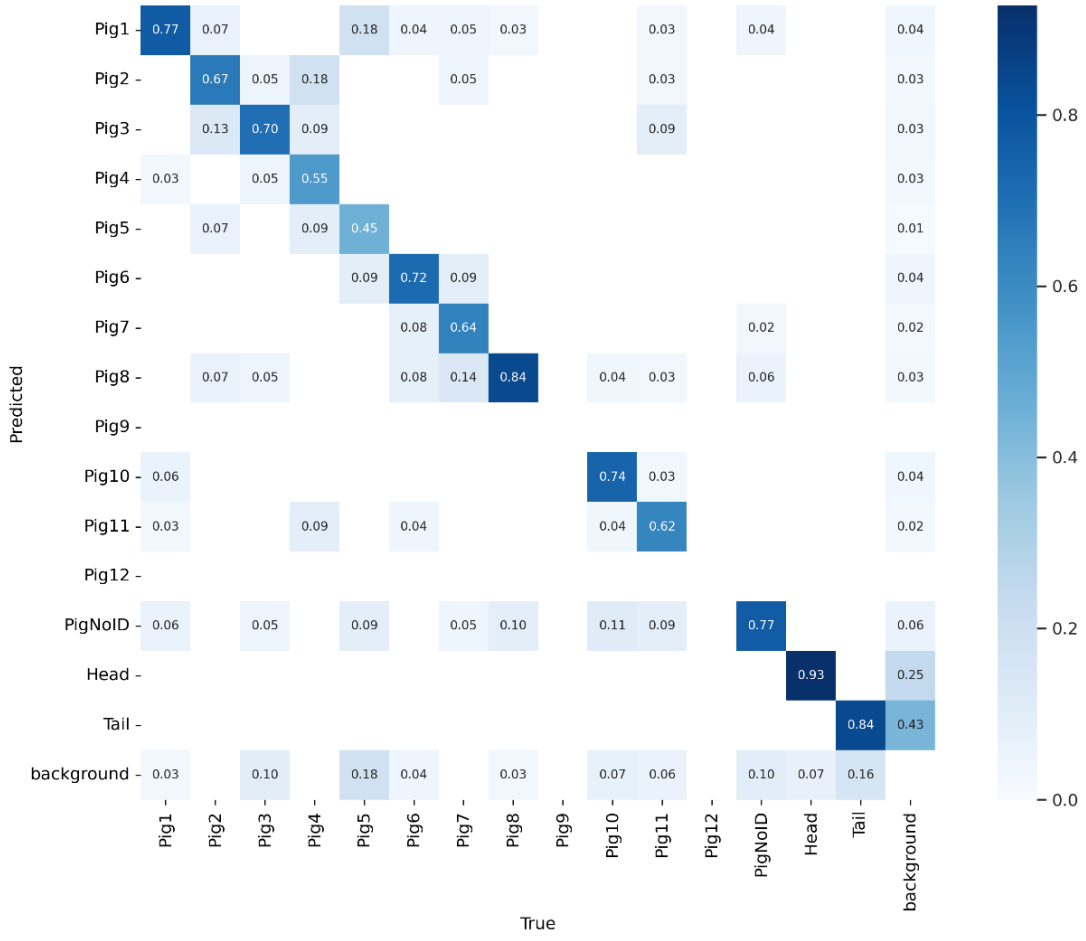


Figure 19. Confusion Matrix Normalized: The number of times pigs correctly detected (Predicted) by the model compared to the actual number of times that each pig is present (True).

The marking method may also have influenced the results. The main problem with this method is that the paint patterns fade over time, and the pig's movements can result in unclear paint patterns, and in some conditions, such as low-light environments or sunlight falling directly on the camera or animals, the marks become imperceptible, as Zhang et al. (2019) mentioned in their study. During the annotation, I noticed some correctable flaws in our marking method, which could be considered and corrected in future uses.

Some markings are not discernible (strokes not continuous from one side to the other, paint drips joining strokes) (Figure 20).

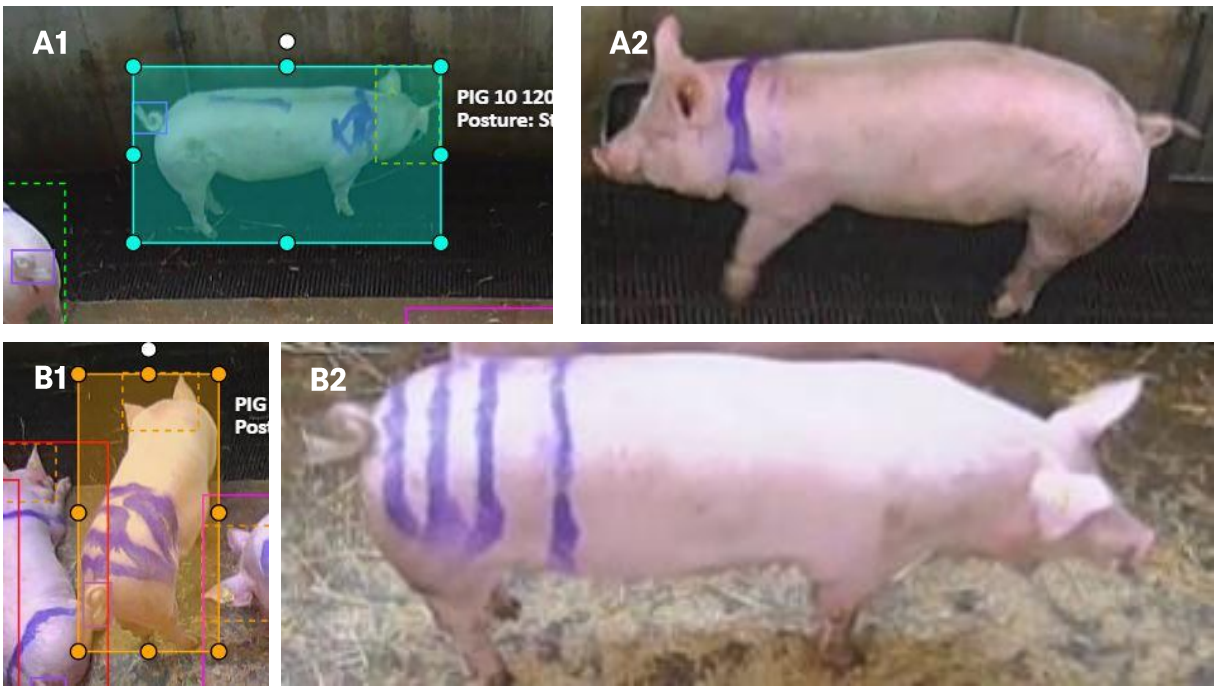


Figure 20. Examples of inaccurate markings. (A1) sloppy marking of pig n.° 10; (A2) well-defined marking of pig n°10; (B1) sloppy marking of pig n°4 (orange BB); (B2) image well-defined marking of pig n°4.

When the traces began to disappear between the days of marking with the light it became difficult to distinguish the pigs (Figure 21).

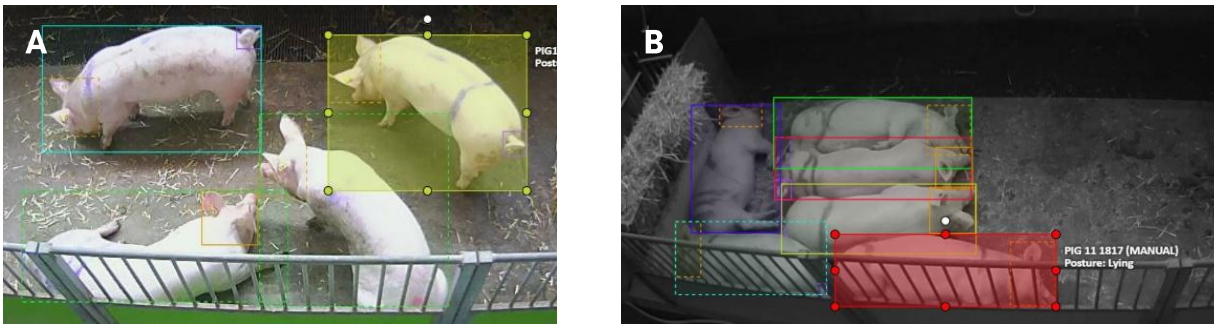


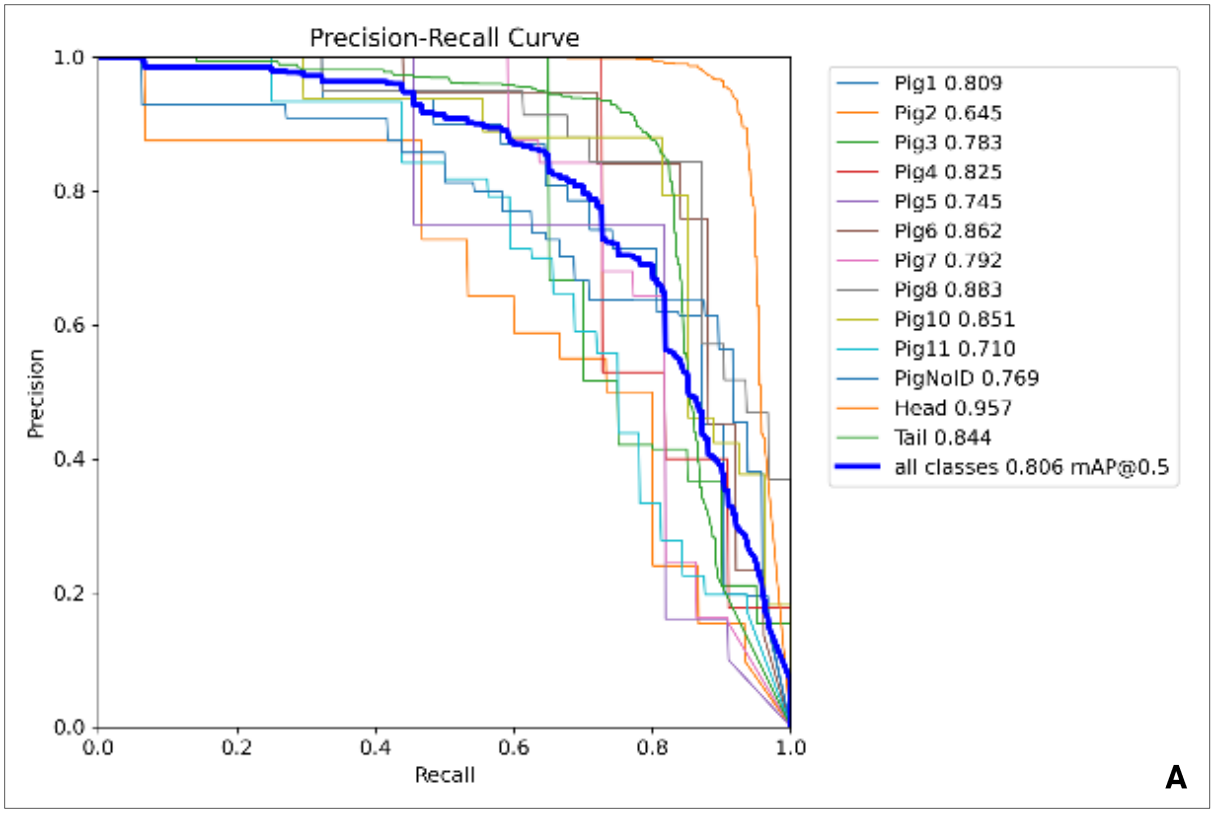
Figure 21. (A) Faded markings, which, when the light shines directly on the animals, are almost imperceptible. (B) night image.

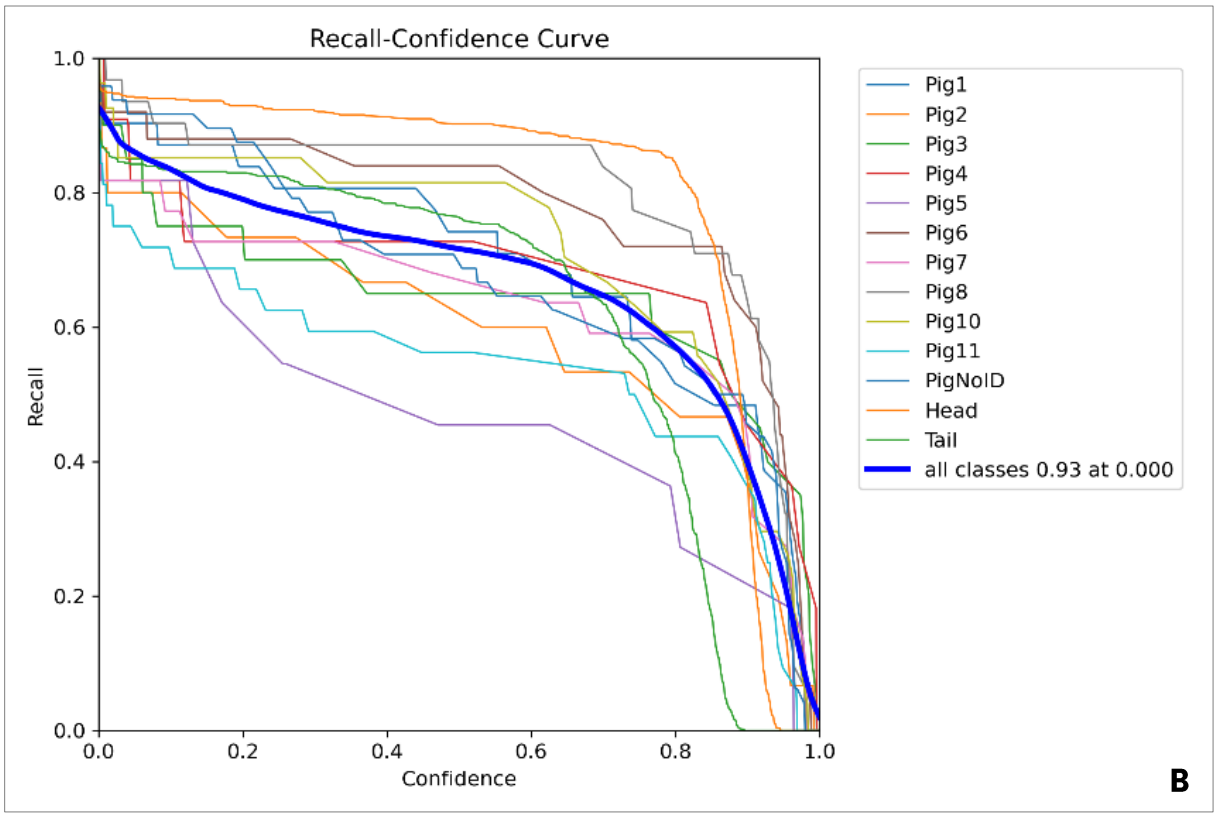
The marks that were too far down, when the pig was perpendicular to the angle of the camera were not visible (Figure 22).

Figure 22. Pig No. 4 (in orange) whose fourth line is too far back to be visible, is confused with Pig 3. (A) Pig No. 4 (in orange) fourth line is too far back partially visible to the annotator but possibly difficult for the object detector, is confused with Pig 3 (B).

The reduction of annotated objects with the occluded property, even without knowing its influence on the accuracy of the object detection model used, may improve the results, as mentioned by Odo et al. (2024). This depends on the camera's point of view and the occupancy density of the pens, that is, by placing the video recording cameras vertically in relation to the animals, thereby reducing occlusions due to the sides of the pens and animals that are very close together.

Despite these findings, the model performed well, as shown in Figure 23A. The Precision-Recall curve shows a differentiated understanding of the performance of the YOLOv8 model in the different classes (Nie et al., 2024), reaching an mAP0.5 of 0.806. This underscores the robustness of the model and the satisfactory overall accuracy of the detection, demonstrating a better performance in the detection of pig heads (0.957) than pig tails (0.844). Generally, a higher confidence score is synonymous with a superior detection capacity (Nie et al., 2024). In the Recall-Confidence curve in Figure 23B, the value 0.93 for all classes indicates the model's target recognition ability. The Accuracy–Confidence curve (Figure 23C) shows how accuracy changes with different confidence levels. In ideal cases, the accuracy should be high at all confidence levels, which is the case with our model.


Precision-Recall Curve
Precision
Recall
Pig1 0.809
Pig2 0.645
Pig3 0.783
Pig4 0.825
Pig5 0.745
Pig6 0.862
Pig7 0.792
Pig8 0.883
Pig10 0.851
Pig11 0.710
PigNoID 0.769
Head 0.957
Tail 0.844
all classes 0.806 mAP@0.5
A


Recall-Confidence Curve
Recall
Confidence
Pig1
Pig2
Pig3
Pig4
Pig5
Pig6
Pig7
Pig8
Pig10
Pig11
PigNoID
Head
Tail
all classes 0.93 at 0.000
B

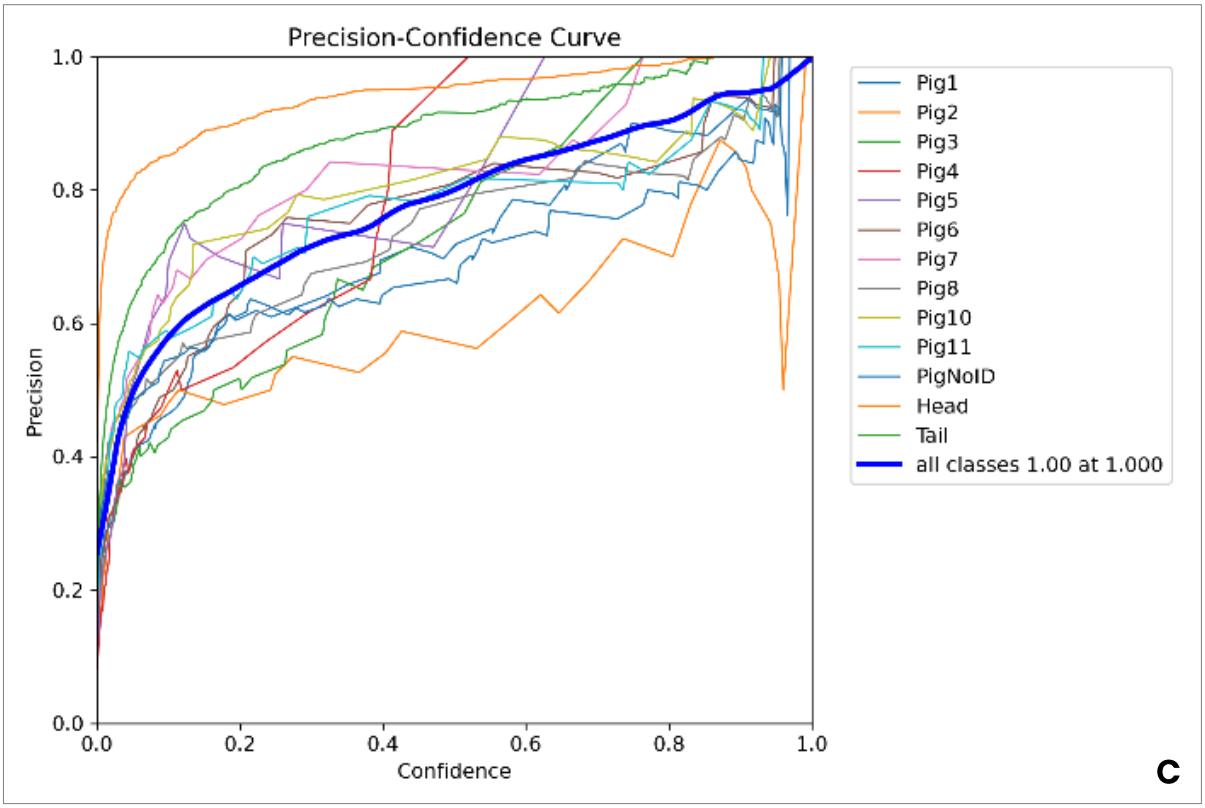


Figure 23. Visual model evaluation metrics. (A) Precision-recall curve showcases the trade-offs between precision and recall at varied thresholds. The blue line represents the mean average precision (mAP) of the model, denoting an overall detection accuracy of 0.806 across all categories. (B) Recall rate, achieving a high mark of 0.99, indicating the model's proficiency in identifying relevant instances. (C) The Precision-confidence curve displays how precision changes with different confidence levels. Ideally, you want high precision across all confidence levels.

6. Conclusions

Of the three annotation tools tested, CVAT was chosen because it allowed for the creation of suitable training data and was easy to use with a better interface. Although only the images annotated with the BB technique were used to train the model, we infer that the use of SAM will be more precise in terms of individual identification and, at a later stage, relating all parts of a pig as belonging to a single individual. The results obtained with the YOLO v8 object detection model showed that the proposed method predicted the head with 93% accuracy and the tail with 84% accuracy, yielding a baseline model with only 280 frames. Even individual identifications were detected (indicating that the model was appropriate), although with a lower accuracy (45% to 84%), probably not due to the method used, but rather to a lower number of ground truths provided for training the model, as shown by the model evaluation. Training on the full set of available annotated frames may yield promising results. These findings also suggest that our final ethogram (at least in part) and annotation methodology was correctly developed, allowing for highly satisfactory training results, even with a small sample of annotated data.

Our collected dataset with 3 months of videos can be used to make inferences to study 'friendship' (detection of emerging problems via changes in behaviours) in pigs. They can also be used to investigate the influence of different diets on behaviours. For example, in the case of the animals filmed in our work, which were also part of the FFP study, the images can be used to study the effects of diets containing different amounts/types of food on the activity levels of each group of pigs.

References

Alameer, A., Buijs, S., O'Connell, N., Dalton, L., Larsen, M., Pedersen, L., & Kyriazakis, I. (2022). Automated detection and quantification of contact behaviour in pigs using deep learning. *Biosystems Engineering*, *224*, 118-130. https://doi.org/10.1016/j.biosystemseng.2022.10.002

Amrish & Shwetank. (2024). Comparative analysis of manual and annotations for crowd assessment and classification using artificial intelligence. *Data Science and Management*, 17. https://doi.org/10.1016/j.dsm.2024.04.001

Bao, Y., Llagostera, P., & Plà-Aragonès, L. M. (2024). Is Deep Learning useful for decision making in pig production? *Internet of Things*, *26*, 25. https://doi.org/10.1016/j.iot.2024.101229

Bergamini, L., Pini, S., Simoni, A., Vezzani, R., Calderara, S., D'Eath, R., & Fisher, R. (2021). *Extracting Accurate Long-term Behavior Changes from a Large Pig Dataset*. 524-533. https://doi.org/10.5220/0010288405240533

Boesch, G. (2023a). *CVAT : Computer Vision Annotation Tool - 2024 Guide*. Viso.Ai. https://viso.ai/computer-vision/cvat-computer-vision-annotation-tool/

Boesch, G. (2023b). *Image Annotation : Best Software Tools and Solutions in 2024*. Viso.Ai. https://viso.ai/computer-vision/image-annotation/

Boesch, G. (2023c, décembre 20). *CVAT : Computer Vision Annotation Tool - 2024 Guide*. Viso.Ai. https://viso.ai/computer-vision/cvat-computer-vision-annotation-tool/

Brünger, J., Gentz, M., Traulsen, I., & Koch, R. (2020). Panoptic Instance Segmentation on Pigs. *ArXiv*, 18.

Camerlink, I., & Turner, S. P. (2013). The pig's nose and its role in dominance relationships and harmful behaviour. *Applied Animal Behaviour Science*, *145*(3), 84-91. https://doi.org/10.1016/j.applanim.2013.02.008

Chen, C., Zhu, W., & Norton, T. (2021). Behaviour recognition of pigs and cattle : Journey from computer vision to deep learning. *Computers and Electronics in Agriculture*, *187*, 106255. https://doi.org/10.1016/j.compag.2021.106255

Collins, L. M., & Smith, L. M. (2022). Review : Smart agri-systems for the pig industry. *Animal*, *16*, 100518. https://doi.org/10.1016/j.animal.2022.100518

CVAT. (s. d.). *Manual CVAT*. CVAT. Consulté 31 juillet 2024, à l'adresse https://docs.cvat.ai/docs/manual/

D'Eath, R. B., Arnott, G., Turner, S. P., Jensen, T., Lahrmann, H. P., Busch, M. E., Niemi, J. K., Lawrence, A. B., & Sandøe, P. (2014). Injurious tail biting in pigs : How can it be controlled in existing systems without tail docking? *Animal: An International Journal of Animal Bioscience*, *8*(9), 1479-1497. https://doi.org/10.1017/S1751731114001359

Devi, S. J., Doley, J., Bharati, J., Mohan, N. H., & Gupta, V. K. (2024). Analysis of pig posture detection in group-housed pigs using deep learning-based mask scoring instance segmentation. *Animal Science Journal = Nihon Chikusan Gakkaiho*, *95*(1), e13975. https://doi.org/10.1111/asj.13975

Doornweerd, J. E., Veerkamp, R. F., De Klerk, B., Van Der Sluis, M., Bouwman, A. C., Ellen, E. D., & Kootstra, G. (2023). Tracking individual broilers on video in terms of time and distance. *Poultry Science*, 103185. https://doi.org/10.1016/j.psj.2023.103185

Drexl, V., Dittrich, I., Wilder, T., Diers, S., & Krieter, J. (2023). Identifying Early Indicators of Tail Biting in Pigs by Variable Selection Using Partial Least Squares Regression. *Animals*, *13*(1), Article 1. https://doi.org/10.3390/ani13010056

Dutta, A., & Zisserman, A. (2019). The VIA Annotation Software for Images, Audio and Video. *Proceedings of the 27th ACM International Conference on Multimedia*, 2276-2279. https://doi.org/10.1145/3343031.3350535

Farahnakian, F., Farahnakian, F., Björkman, S., Bloch, V., Pastell, M., & Heikkonen, J. (2024). Pose estimation of sow and piglets during free farrowing using deep learning. *Journal of Agriculture and Food Research*, *16*, 101067. https://doi.org/10.1016/j.jafr.2024.101067

Fernandes, A. F. A., Dórea, J. R. R., & Rosa, G. J. de M. (2020). Image Analysis and Computer Vision Applications in Animal Sciences : An Overview. *Frontiers in Veterinary Science*, *7*. https://doi.org/10.3389/fvets.2020.551269

Fernandes, A. F. A., Dórea, J. R. R., Valente, B. D., Fitzgerald, R., Herring, W., & Rosa, G. J. M. (2020). Comparison of data analytics strategies in computer vision systems to predict pig body composition traits from 3D images. *Journal of Animal Science*, *98*(8), skaa250. https://doi.org/10.1093/jas/skaa250

Gan, H., Ou, M., Zhao, F., Xu, C., Li, S., Chen, C., & Xue, Y. (2021). Automated piglet tracking using a single convolutional neural network. *Biosystems Engineering*, *205*, 48-63. https://doi.org/10.1016/j.biosystemseng.2021.02.010

Gerster, U., Sidler, X., Wechsler, B., & Nathues, C. (2022). Prevalence of tail lesions in Swiss finishing pigs. *Schweiz Arch Tierheilkd*, *164*(4), 339-349. https://doi.org/10.17236/sat00352

Gómez, Y., Stygar, A. H., Boumans, I. J. M. M., Bokkers, E. A. M., Pedersen, L. J., Niemi, J. K., Pastell, M., Manteca, X., & Llonch, P. (2021). A Systematic Review on Validated Precision Livestock Farming Technologies for Pig Production and Its Potential to Assess Animal Welfare. *Frontiers in Veterinary Science*, *8*. https://www.frontiersin.org/articles/10.3389/fvets.2021.660565

Grümpel, A., Krieter, J., Veit, C., & Dippel, S. (2018). Factors influencing the risk for tail lesions in weaner pigs (Sus scrofa). *Livestock Science*, *216*, 219-226. https://doi.org/10.1016/j.livsci.2018.09.001

Guo, Q., Sun, Y., Orsini, C., Bolhuis, J. E., De Vlieg, J., Bijma, P., & De With, P. H. N. (2023). Enhanced camera-based individual pig detection and tracking for smart pig farms. *Computers and Electronics in Agriculture*, *211*, 108009. https://doi.org/10.1016/j.compag.2023.108009

Hakansson, F., & Houe, H. (2020). Risk factors associated with tail damage in conventional non-docked pigs throughout the lactation and rearing period. *Preventive Veterinary Medicine*, *184*, 105160. https://doi.org/10.1016/j.prevetmed.2020.105160

Hakansson, F., & Jensen, D. B. (2023). Automatic monitoring and detection of tail-biting behavior in groups of pigs using video-based deep learning methods. *Frontiers in Veterinary Science*, *9*, 1099347. https://doi.org/10.3389/fvets.2022.1099347

Han, J., Siegford, J., Colbry, D., Lesiyon, R., Bosgraaf, A., Chen, C., Norton, T., & Steibel, J. P. (2023). Evaluation of computer vision for detecting agonistic behavior of pigs in a single-space feeding stall through blocked cross-validation strategies. *Computers and Electronics in Agriculture*, *204*, 107520. https://doi.org/10.1016/j.compag.2022.107520

Ji, H., Yu, J., Lao, F., Zhuang, Y., Wen, Y., & Teng, G. (2022). Automatic Position Detection and Posture Recognition of Grouped Pigs Based on Deep Learning. *Agriculture*, *12*(9), Article 9. https://doi.org/10.3390/agriculture12091314

Kashiha, M., Bahr, C., Ott, S., Moons, C. P. H., Niewold, T. A., Ödberg, F. O., & Berckmans, D. (2013). Automatic identification of marked pigs in a pen using image pattern recognition. *Computers and Electronics in Agriculture*, *93*, 111-120. https://doi.org/10.1016/j.compag.2013.01.013

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y., Dollár, P., & Girshick, R. (2023). Segment Anything. *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, 3992-4003. https://doi.org/10.1109/ICCV51070.2023.00371

Larsen, M. L. V., Andersen, H. M.-L., & Pedersen, L. J. (2016). Can tail damage outbreaks in the pig be predicted by behavioural change? *The Veterinary Journal*, *209*, 50-56. https://doi.org/10.1016/j.tvjl.2015.12.001

Larsen, M. L. V., Andersen, H. M.-L., & Pedersen, L. J. (2018). Which is the most preventive measure against tail damage in finisher pigs : Tail docking, straw provision or lowered stocking density? *Animal*, *12*(6), 1260-1267. https://doi.org/10.1017/S175173111700249X

Larsen, M. L. V., Andersen, H. M.-L., & Pedersen, L. J. (2019). Changes in activity and object manipulation before tail damage in finisher pigs as an early detector of tail biting. *Animal*, *13*(5), 1037-1044. https://doi.org/10.1017/S1751731118002689

Li, J., Green-Miller, A. R., Hu, X., Lucic, A., Mahesh Mohan, M. R., Dilger, R. N., Condotta, I. C. F. S., Aldridge, B., Hart, J. M., & Ahuja, N. (2022). Barriers to computer vision applications in pig production facilities. *Computers and Electronics in Agriculture*, *200*, 107227. https://doi.org/10.1016/j.compag.2022.107227

Liu, D., Oczak, M., Maschat, K., Baumgartner, J., Pletzer, B., He, D., & Norton, T. (2020). A computer vision-based method for spatial-temporal action recognition of tail-biting behaviour in group-housed pigs. *Biosystems Engineering*, *195*, 27-41. https://doi.org/10.1016/j.biosystemseng.2020.04.007

Liu, D., Parmiggiani, A., Psota, E., Fitzgerald, R., & Norton, T. (2023). Where's your head at? Detecting the orientation and position of pigs with rotated bounding boxes. *Computers and Electronics in Agriculture*, *212*, 108099. https://doi.org/10.1016/j.compag.2023.108099

Mattina, M., Benzinou, A., Nasreddine, K., & Richard, F. (2023). An efficient anchor-free method for pig detection. *IET Image Processing*, *17*(2), 613-626. https://doi.org/10.1049/ipr2.12659

Mazzoleni, S., Tretola, M., Luciano, A., Lin, P., Pinotti, L., & Bee, G. (2023). Sugary and salty former food products in pig diets affect energy and nutrient digestibility, feeding behaviour but not the growth performance and carcass composition. *animal*, *17*(12), 101019. https://doi.org/10.1016/j.animal.2023.101019

Moinard, C., Mendl, M., Nicol, C. J., & Green, L. E. (2003). A case control study of on-farm risk factors for tail biting in pigs. *Applied Animal Behaviour Science*, *81*(4), 333-355. https://doi.org/10.1016/S0168-1591(02)00276-9

Mora, M., Piles, M., David, I., & Rosa, G. J. M. (2024). Integrating computer vision algorithms and RFID system for identification and tracking of group-housed animals : An example with pigs. *Journal of Animal Science*, skae174. https://doi.org/10.1093/jas/skae174

Munsterhjelm, C., Heinonen, M., & Valros, A. (2016). Can tail-in-mouth behaviour in weaned piglets be predicted by behaviour and performance? *Applied Animal Behaviour Science*, *184*, 16-24. https://doi.org/10.1016/j.applanim.2016.08.013

Nie, L., Li, B., Du, Y., Jiao, F., Song, X., & Liu, Z. (2024). Deep learning strategies with CReToNeXt-YOLOv5 for advanced pig face emotion detection. *Scientific Reports*, *14*(1), 1679. https://doi.org/10.1038/s41598-024-51755-8

Ocepek, M., Žnidar, A., Lavrič, M., Škorjanc, D., & Andersen, I. L. (2022). DigiPig : First Developments of an Automated Monitoring System for Body, Head and Tail Detection in Intensive Pig Farming. *Agriculture*, *12*(1), Article 1. https://doi.org/10.3390/agriculture12010002

Odo, A., McLaughlin, N., & Kyriazakis, I. (2024). Automated Monitoring of Ear Biting in Pigs by Tracking Individuals and Events. *2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 7080-7088. https://doi.org/10.1109/WACV57701.2024.00694

Odo, A., Muns, R., Boyle, L., & Kyriazakis, I. (2023). Video Analysis Using Deep Learning for Automated Quantification of Ear Biting in Pigs. *IEEE Access*, *11*, 59744-59757. IEEE Access. https://doi.org/10.1109/ACCESS.2023.3285144

Ollagnier, C., Kasper, C., Wallenbeck, A., Keeling, L., Bee, G., & Bigdeli, S. A. (2023). Machine learning algorithms can predict tail biting outbreaks in pigs using feeding behaviour records. *PLOS ONE*, *18*(1), Article 1. https://doi.org/10.1371/journal.pone.0252002

Ordonnance sur la protection des animaux, RS 455.1, 2985 (2008). https://www.fedlex.admin.ch/eli/cc/2008/416/fr#a18

Pangal, D. J., Kugener, G., Shahrestani, S., Attenello, F., Zada, G., & Donoho, D. A. (2021). A Guide to Annotation of Neurosurgical Intraoperative Video for Machine Learning Analysis and Computer Vision. *World Neurosurgery*, *150*, 26-30. https://doi.org/10.1016/j.wneu.2021.03.022

Parmiggiani, A., Liu, D., Psota, E., Fitzgerald, R., & Norton, T. (2023). Don't get lost in the crowd : Graph convolutional network for online animal tracking in dense groups. *Computers and Electronics in Agriculture*, *212*, 108038. https://doi.org/10.1016/j.compag.2023.108038

Psota, E., Mittek, M., Pérez, L., Schmidt, T., & Mote, B. (2019). Multi-Pig Part Detection and Association with a Fully-Convolutional Network. *Sensors*, *19*, 852. https://doi.org/10.3390/s19040852

Psota, E., Schmidt, T., Mote, B., & Pérez, L. (2020). Long-Term Tracking of Group-Housed Livestock Using Keypoint Detection and MAP Estimation for Individual Animal Identification. *Sensors*, *20*, 3670. https://doi.org/10.3390/s20133670

Riekert, M., Klein, A., Adrion, F., Hoffmann, C., & Gallmann, E. (2020). Automatically detecting pig position and posture by 2D camera imaging and deep learning. *Computers and Electronics in Agriculture*, *174*, 105391. https://doi.org/10.1016/j.compag.2020.105391

Roch, L. (2021). *Étude des comportements potentiellement délétères des porcs à l'engrais – Relation avec l'efficience protéique* [Master thesis]. Universität Zürich/Bern.

Roch, L., Ewaoluwagbemiga, E. O., & Kasper, C. (2023). Social interactions, precursors of damaging behaviours, object manipulation, straw rooting, and activity : A detailed data set in undocked pigs under protein restriction. *Animal - Open Space*, *2*, 100044. https://doi.org/10.1016/j.anopes.2023.100044

Schrøder-Petersen, D. L., & Simonsen, H. B. (2001). Tail Biting in Pigs. *The Veterinary Journal*, *162*(3), 196-210. https://doi.org/10.1053/tvjl.2001.0605

Sener, O., & Savarese, S. (2018). *Active Learning for Convolutional Neural Networks : A Core-Set Approach* (arXiv:1708.00489). arXiv. https://doi.org/10.48550/arXiv.1708.00489

Siegford, J. M., Steibel, J. P., Han, J., Benjamin, M., Brown-Brandl, T., Dórea, J. R. R., Morris, D., Norton, T., Psota, E., & Rosa, G. J. M. (2023). The quest to develop automated systems for monitoring animal behavior. *Applied Animal Behaviour Science*, *265*, 106000. https://doi.org/10.1016/j.applanim.2023.106000

Sonoda, L. T., Fels, M., Oczak, M., Vranken, E., Ismayilova, G., Guarino, M., Viazzi, S., Bahr, C., Berckmans, D., & Hartung, J. (2013). Tail Biting in pigs-Causes and management intervention strategies to reduce the behavioural disorder. A review. *Berliner und Münchener Tierärztliche Wochenschrift*, *126*. https://www.vetline.de/tail-biting-in-pigs-causes-and-management-intervention-strategies-to-reduce-the-behavioural

Statham, P., Green, L., Bichard, M., & Mendl, M. (2009). Predicting tail-biting from behaviour of pigs prior to outbreaks. *Applied Animal Behaviour Science*, *121*(3-4), 157-164. https://doi.org/10.1016/j.applanim.2009.09.011

Taylor, N. R., Main, D. C. J., Mendl, M., & Edwards, S. A. (2010). Tail-biting : A new perspective. *The Veterinary Journal*, *186*(2), 137-147. https://doi.org/10.1016/j.tvjl.2009.08.028

Taylor, N. R., Parker, R. M. A., Mendl, M., Edwards, S. A., & Main, D. C. J. (2012). Prevalence of risk factors for tail biting on commercial farms and intervention strategies. *The Veterinary Journal*, *194*(1), 77-83. https://doi.org/10.1016/j.tvjl.2012.03.004

Ultralytics. (s. d.). *Ultralytics YOLO Docs*. Consulté 30 juin 2024, à l'adresse https://docs.ultralytics.com/#yolo-a-brief-history

V7. (s. d.). *V7 Darwin Resources & Documentation*. V7 Darwin Resources & Documentation. Consulté 17 juillet 2024, à l'adresse https://docs.v7labs.com/

van der Maaten, L., & Hinton, G. (2008). Visualizing Data using t-SNE. *Journal of Machine Learning Research*, *9*, Article 9.

van der Zande, Lisette. E., Guzhva, O., & Rodenburg, T. B. (2021). Individual Detection and Tracking of Group Housed Pigs in Their Home Pen Using Computer Vision. *Frontiers in Animal Science*, *2*. https://www.frontiersin.org/articles/10.3389/fanim.2021.669312

van Staaveren, N., Boyle, L. A., Manzanilla, E. G., O'Driscoll, K., Shalloo, L., & Díaz, J. A. C. (2021). Severe tail lesions in finisher pigs are associated with reduction in annual profit in farrow-to-finish pig farms. *Veterinary Record*, *188*(8), e13. https://doi.org/10.1002/vetr.13

Wang, F., Fu, X., Duan, W., Wang, B., & Li, H. (2024). The Detection of Ear Tag Dropout in Breeding Pigs Using a Fused Attention Mechanism in a Complex Environment. *Agriculture*, *14*(4), Article 4. https://doi.org/10.3390/agriculture14040530

Wang, M., oczak, M., Larsen, M., Bayer, F., Maschat, K., Baumgartner, J., Rault, J.-L., & Norton, T. (2021). A PCA-based frame selection method for applying CNN and LSTM to classify postural behaviour in sows. *Computers and Electronics in Agriculture*, *189*, 106351. https://doi.org/10.1016/j.compag.2021.106351

Wei, J., Tang, X., Liu, J., & Zhang, Z. (2023). Detection of Pig Movement and Aggression Using Deep Learning Approaches. *Animals*, *13*(19), Article 19. https://doi.org/10.3390/ani13193074

Wutke, M., Schmitt, A., Heinrich, F., Das, P., Lange, A., Gentz, M., Traulsen, I., Warns, F., & Gültas, M. (2021). Detecting Animal Contacts—A Deep Learning-Based Pig Detection and Tracking Approach for the Quantification of Social Contacts. *Sensors*, *21*, 16. https://doi.org/10.3390/s21227512

Xu, P., Zhang, Y., Ji, M., Guo, S., Tang, Z., Wang, X., Guo, J., Zhang, J., & Guan, Z. (2024). Advanced intelligent monitoring technologies for animals : A survey. *Neurocomputing*, *585*, 127640. https://doi.org/10.1016/j.neucom.2024.127640

Yang, Q., & Xiao, D. (2020). A review of video-based pig behavior recognition. *Applied Animal Behaviour Science*, *233*, 105146. https://doi.org/10.1016/j.applanim.2020.105146

Zhang, L., Gray, H., Ye, X., Collins, L., & Allinson, N. (2019). Automatic Individual Pig Detection and Tracking in Pig Farms. *Sensors*, *19*, 1188. https://doi.org/10.3390/s19051188

Zhou, H., Li, Q., & Xie, Q. (2023). Individual Pig Identification Using Back Surface Point Clouds in 3D Vision. *Sensors*, *23*(11), Article 11. https://doi.org/10.3390/s23115156

Zonderland, J. J., Schepers, F., Bracke, M. B. M., Den Hartog, L. A., Kemp, B., & Spoolder, H. A. M. (2011). Characteristics of biter and victim piglets apparent before a tail-biting outbreak. *Animal*, *5*(5), 767-775. https://doi.org/10.1017/S1751731110002326

Zonderland, J. J., Van Riel, J. W., Bracke, M. B. M., Kemp, B., Den Hartog, L. A., & Spoolder, H. A. M. (2009). Tail posture predicts tail damage among weaned piglets. *Applied Animal Behaviour Science*, *121*(3-4), 165-170. https://doi.org/10.1016/j.applanim.2009.09.002

Annexes

Table 1. Correspondence of the RFID number with the assigned work number and the respective mark drawn on each pig.

Pen 1 (area 1 e 2): cameras 1,2 and 4				
RFID	N° pig	Mark		
		Shouderls	Back	Rear
3236	1			I
3210	2			II
3191	3			III
3264	4			IIII
3275	5		/	
3205	6		/	I
3255	7		/	II
3183	8		/	III
3250	9		/	IIII
3194	10	I		
3272	11	I		I
3285	12	I		II

Pen 2 (area 3 and 4): Cameras 5 e 6				
RFID	N° pig	Mark		
		Shouderls	Back	Rear
3181	1			I
3252	2			II
3192	3			III
3227	4			IIII
3209	5		/	
3274	6		/	I
3287	7		/	II
3271	8		/	III
3254	9		/	IIII
3206	10	I		
3263	11	I		I
3196	12	I		II

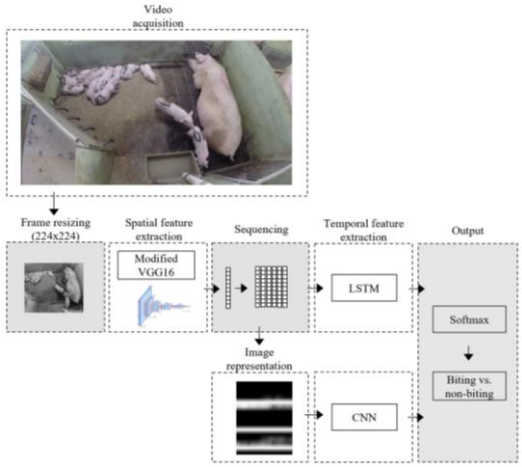
Table 2. Schematic summary of the bibliography consulted to define annotation models, annotation techniques and object detection models.

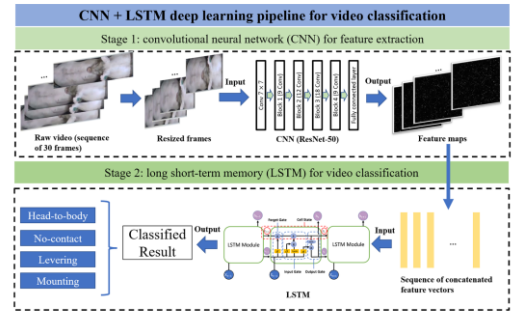
Reference	Species/ category	Objective	Annotation	Remarks	Conclusions
Tracking individual broilers on video in terms of time and distance (**Doornweerd et al., 2023)**	Broilers (39)	- Analyze broiler tracking broiler on vídeo: number of ID-switches, tracking time and distance - examine potential tracking errors (ID-losses – location, proximity, kinematics) in an experimental pen to enable broiler locomotion phenotyping	CVAT (n=200 frames)	- The YOLOv7-tiny (training=140/validation=30/ testing n=30); AP@0.5: 0.99; AP@0.75: 0.98 - **YOLOv7-tiny model** pre-trained on the Microsoft Common Objects in Context (COCO) dataset (Lin et al., 2014) was fine-tuned to **detect broilers** - A multi-object tracker (**SORT)** was used **to track the detected broilers** on video over consecutive frames. SORT associates the broiler detections between consecutive frames using Kalman Filters (Kalman, 1960) and the Hungarian algorithm (Kuhn, 1955). - This study provides insight into the video-tracking of white broilers using a popular tracking-by-detection algorithm (SORT; Bewley et al., 2016), identifies bottlenecks, and proposes potential solutions, and implications for phenotyping broiler locomotion.	- the potential tracking errors (ID-losses) revealed that the majority were associated with the location in the pen (occluded by the drinker) and the proximity to other broilers (relative to no ID-losses, within 10 cm) - Kinematics appeared to play a less predominant role in the occurrence of ID-losses. - Future broiler locomotion phenotyping will not only require addressing ID-switches, either reduce them or favor ID-losses, and the optimization of the tracking algorithm but also the use of an external animal identification system (e.g., passive radio frequency identification).

Reference	Species/ category	Objective	Annotation	Remarks	Conclusions
Enhanced camera-based individual pig detection and tracking for smart pig farms / **(Guo et al., 2023)**	pigs	Investigation and comparison of 3 state-of-the-art Development of data association strategies to optimize multi-object re-identification	CVAT -Video segments showing active pig movements are selected, followed by annotating the pig location in each video frame with a consistent identity associated for each individual pig. CVAT is capable of labeling object-location information using a rectangular-shape bounding box, and also provides the options for adding occlusion conditions. Aiming at more effective annotation work, we have chosen to annotate bounding boxes rather than object contours. CVAT supports saving the frame ID, object ID, bounding-box location, and size of the object.	In terms of the manual annotation effort, the procedure for collecting the appropriate amount of annotation as ground truth is very time-consuming. Considering the difference in moving speed between pedestrians and pigs, we have adopted an annotation interval of 2 s for pigs to improve annotation efficiency. Continuous annotation is expected to yield a more precise tracking system.	We have investigated three state-of-the-art automated multi-object tracking methods on 2 pig datasets. Both datasets contain manual annotations of pigs in real farms. The video segments have diverse challenging conditions such as occlusion, active and high-speed movements. In this way, the generalization and robustness of the tracking models are evaluated based on K-fold cross-validation. We have proposed a weighted association strategy to enhance the association algorithm of animal re-ID on JDE and FairMOT methods, which increase the performance of IDF1 by 1.97% at most, and reduces the mean number of identity switches by 46 at most. It

Reference	Species/ category	Objective	Annotation	Remarks	Conclusions
					can be concluded that the enhanced FairMOT performs the best in terms of multi-object tracking, indicated by an IDF1 of 80.94%, MOTA of 88.55%, MOTP of 82.60%, and number of identity switches of 213. All tracking systems achieve a nearly and/or real-time execution rate. For the purpose of a continuous MOT system, all proposed methods are sufficient in terms of the execution rate. In conclusion, the experimental results of evaluation metrics demonstrate the effectiveness and robustness of the three proposed methods on multi-object tracking systems. FairMOT with the proposed weighted association strategy achieves the best tracking

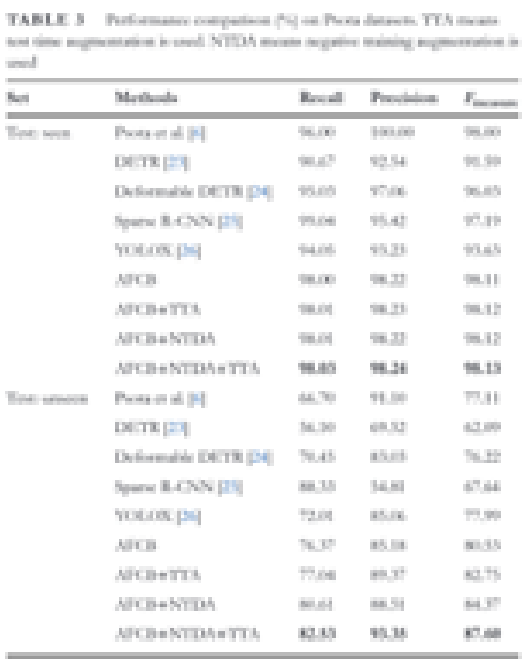
Reference	Species/ category	Objective	Annotation	Remarks	Conclusions
					performance for individual pigs in a real farm.
Automatic monitoring and detection of tail-biting behavior in groups of pigs using video-based deep learning methods / **(Hakansson & Jensen, 2023)**	Pigs /piglets prior to weaning	developing a video-based deep learning approach for detecting tail-biting behaviour in groups of pigs without the implementation of a prior tracking algorithm.	The included videos were selected based on the following inclusion criteria: (1) no interference with the stockperson working in the pen, (2) minimum recording of 55 min and (3) piglets were visible for more than 80 % of the video recording. - The full dataset consisted of 332.666 images, of which 5.330 images showed biting behaviour	Based on the respective optimal parameter settings, our final CNN-LSTM network converged with a training accuracy of 98.5% and a validation accuracy of 78.2%, while the final CNN-CNN network converged to a training accuracy of 99.5% and a validation accuracy of 99%. The results indicate that both methods are able to learn from the given data and that combining the pre-trained model VGG-16 with a LSTM or a secondary CNN without the use of prior tracking can be used to detect tail-biting behaviour in groups of pigs. - The results of the prediction of the LSTM and CNN models in relation to the different behavioural categories are shown in Table 9. While the LSTM model achieves a sensitivity of 89% in predicting tail-biting events, the specificity to exclude other behaviours is lower. In contrast, the final CNN predicts tail-biting events with a sensitivity of 37%, while the specificity to exclude other behaviours reaches values above 90%. - While random validation produced an acceptable accuracy of 96.8%, using validation	There clearly is potential for on-farm use of the proposed methods, especially the CNN-LSTM, further improvements of the methods are needed. To improve the generalizability of the proposed methods in the future, the potential of data augmentation, as well as utilizing varying pretrained models to extract spatial features should be explored. - Future studies should evaluate the generalizability of the methods of farm setting and of animals at different stages of the pig production.

Reference	Species/ category	Objective	Annotation	Remarks	Conclusions
				strategies that blocked data over time or by pen/feeder location had poorer performance To achieve this aim, we intend to apply a pre-trained CNN for extraction of latent features from the video frames, combined with two secondary models to analyse sequential data. Specifically, the two secondary models are a LSTM network applied to sequences of the extracted image features (CNNLSTM) and a CNN applied to image representations of extracted spatial features (CNN-CNN). 	

Reference	Species/ category	Objective	Annotation	Remarks	Conclusions
Evaluation of computer vision for detecting agonistic behaviour of pigs in a single-space feeding stall through blocked cross-validation strategies / **(Han et al., 2023)**	Pigs grow-finish	1) develop a CV approach to classify pigs' interactive behaviours in single-space feeding stalls, and 2) test the algorithm through random validation and two blocked validation strategies (Roberts et al., 2017), where the data were split temporally and spatially. We also present the importance of algorithm evaluation as well as diagnostics through multiple training-validation scenarios that are more practical in animal farming.	Focused on video segments when there were at least two pigs present in the feeding stall. Such events with two or more pigs were passed to the observers. - Prior to further processing, each segment of video (when there were two or more pigs in the feeding stall) was cut into 30-frame video episodes labelled with one of the four behaviour classes (NC, HB, L or M) following annotation by a trained human observer. - After performing data augmentation for L and M, we obtained a total of 15,679 30-frame episodes	Ethogram for the agonistic behaviors in pigs. *: ear-to-body was merged into head-to-body. **Behavior** / **Description** / **Code** No contact — Two pigs were in view at the feeding stall. The behind pig had at least both ears in the feeding stall but there was no physical contact between the behind pig and the body of front pig. — NC Ear-to-body* — The behind pig had at least both ears in the feeding stall and unintentional contact was made. The behind pig might be nosing the floor or eating displaced feed and making slight, non-forceful contact with the front pig. This often appeared as the behind pig's ears grazing the front pig or the behind pig's nose bumping the rear legs of the front pig while investigating the floor. — EB* Head-to-body* — The behind pig used its head to make intentional contact (greater than 1 s) with the body of the front pig. Quick (less than 1 s) bumps/run-ins by the behind pig were not recorded. — HB* Levering — The behind pig's snout was under the body of the front pig and the front pig was lifted from the ground vertically. Any lifting of the front pig that involved a behind pig was considered levering. Typically, only the back half of the front pig was lifted. This often manifested as the behind pig pushing forward under the front pig, but it could also appear as the front pig backing up and over the head of the behind pig. — L Mounting — The behind pig lifted its two front legs and put the two legs or its breast on the rear part of the front pig. The mounting pig may sit down during the mounting. Mounting commenced when the two front legs of or the breast of the behind pig contacted the front pig and terminated as soon as the mounting pig was no longer on top of the front pig even some contact was still maintained. — M  Fig. 3. The CNN + LSTM deep learning pipeline for pig's agonistic behavior classification based on videos. The input was a 30-frame video episode and the output was a behavior category. In Stage 1, the raw video episode was resized to 224 × 224 × 3, and (spatial) features of individual frames of the episode were extracted through CNN (transfer learning); in Stage 2, (spatial) features for each individual frame were concatenated as a 1-d vector, and the concatenated feature vectors were sorted and fed to LSTM for temporal representation learning.	In the future development and validation of computer vision models should focus on alternative blocked validation strategies that account for known structures within the dataset to better represent real-world conditions, which will be informative to practical applications of animal phenomics and precision livestock farming.

Reference	Species/ category	Objective	Annotation	Remarks	Conclusions
Automated detection and quantification of contact behaviour in pigs using deep learning / (**Alameer et al., 2022**)	Pigs/ finishing (tail docked)	- we aim to construct a more holistic and representative dataset for training, validation and testing the proposed method developed an automated method that enabled us to quantify the frequency of the contact of one pig's head (including snout) with another pig's rear (including tail) - we aimed to construct a more holistic and representative dataset for training, validation and testing the proposed method	- The detection dataset comprised a total of 51,193 instances (26,533 AFBI þ 24,660 AUF) across 2781 images (1556 AFBI þ 1225 AUF); each pig within an image was manually annotated into two parts: head and rear. A bounding box1 was applied manually on the head and rear of all pigs in a pen. The bounding box denotes the location and size of each pig part. - Contact between pigs. An additional dataset was annotated to validate the interaction method, i.e., the processing stage that feeds from the detection method. This dataset consisted of images from both farms used in this framework. The total number of images of this dataset was 670 images; with sets of 376 and 294 images to	1. We developed a system that detects high-level pig behaviours, i.e., interaction between any two pigs within a group, using only RGB and Infrared cameras. 2. We investigated the characteristics of existing detectors (Bochkovskiy et al., 2020; Liu et al., 2016; Redmon and Farhadi, 2018; Ren et al., 2017; Tan et al., 2020), in terms of speed and detection precision, showing that the configuration of the YOLO network is more suitable for our task. 3. We modified the YOLO network (Redmon and Farhadi, 2018; Bochkovskiy et al., 2020) architecture by adding an additional detection subnetwork to its baseline network, enabling the method to better detect smaller objects, in this case pig heads. We implemented a data-driven process to obtain the optimal layer of the baseline network for feature extraction. We then calculated relevant sets of anchor boxes using the K-means clustering algorithm. 4. We developed an additional processing module that feeds from the detection network and automatically scores interactions between pigs. 5. We applied the proposed system to a significant welfare challenge in the management of pigs, that of the detection of tail-biting outbreaks in pigs. 6. We produced and made	This paper proposed a novel solution that enables quantifying interactions (head-to-rear contact) between group-housed pigs. The method was based on machine learning and image processing whereby highly established deep learning networks were developed to detect and associate pig parts. We developed a lightweight processing module that rapidly scores intersections between pigs. The paper introduced a practical implementation for detecting interactions between multiple pigs using only video surveillance (infrared and RGB) and suitable to be used in commercial settings, as it was applied in diverse conditions.

Reference	Species/ category	Objective	Annotation	Remarks	Conclusions
			represent AFBI and AUF datasets, respectively. - The code and dataset with all its annotations can be found online at https://doi.org/10.17866/rd.salford.21346767.	publicly available a large, annotated dataset for pig parts identification and pig interactions.	
An efficient anchor-free method for pig detection / **(Mattina et al., 2023)**	pigs	· We propose a new detector known as anchor-free center based (AFCB) that is designed for the overlapping and crowded scenario of pig detection. · We propose a combination of novel negative data augmentation during training and test time augmentation to improve the robustness and generalization ability of our AFCB	two pig detection datasets provided by the authors Bergamini et al. (2021) and Psota et al. (2019	TABLE 1 CNN architectures used in the literature for pig detection since 2019 Anchor-based / Anchor-free Authors — two stage — one stage — key point-based — centre-based Cownton et al. [2] — X Nasirahmadi et al. [3] — X Zhang et al. [4] — X Li et al. [5] — X Psota et al. [6, 7] — X Seo et al. [8] — X Liu et al. [9] — X Riekert et al. [10] — X Bergamini et al. [11] — X Wutke et al. [12] — X **Ours** — X	As with many CNN frameworks, our AFCB model suffers from performance degradation under uncontrolled conditions. To tackle these problems, we propose a combination of a novel negative training data augmentation and test time augmentation. We conduct experiments on two pig detection datasets available online and find that our network surpasses state-of-the-art results on both datasets. - To monitor the level of activity of the pigs

Reference	Species/ category	Objective	Annotation	Remarks	Conclusions
		framework. · The proposed method was tested on two pig detection datasets and shown to be more effective than state-of-the-art detectors.		 Several anchor-free methods have been proposed to perform well in crowded environments. DETR [23]and deformable DETR [24] are two recent detectors that directly predict a set of box predictions by combining a CNN and a transformer-based backbone. Sparse R-CNN [25] uses a feature pyramid network (FPN) based on ResNet architecture to extract multi-scale feature maps. A dynamic head	automatically, we intend to add a branch that predicts their posture.

Reference	Species/ category	Objective	Annotation	Remarks	Conclusions
				outputs a set of object proposals. YOLOX [26] is an anchor-free version of YOLO series [27]. These methods achieve good results on the CrowdHuman dataset [28], a benchmark for detecting humans in a crowd. However, they have never been used for the challenging task of pig detection. Compared to CrowdHuman dataset, where each individual image has a different background, these methods could not be effective in pig environments due to the lack of variety in the images.	
Extracting Accurate Long-Term Behavior Changes from a Large Pig Dataset **/ (Bergamini et al., 2021)**	Growing pigs	In summary, the main contributions of our work are: A behaviour analysis pipeline that focuses on individual pig behaviours to infer statistics about 5 different individual behaviours and how these change through time; Evidence that the behaviour statistics	VaticJS (Bolkensteyn, 2016) https://aimagelab.ing.unimore.it/go/pigs-behaviours. In each frame, the annotator: Draws a rectangular bounding box around each visible pig; Associates each bounding box with one of the 8 pigs using a numeric identifier; Selects a behaviour among a list of 5 options (lie, move, eat, drink and stand). 	**Behaviour Analysis:** Behaviour analysis - use the detections and trackless identified by the algorithms to detect and tracking Detection: YOLO v3 (Redmon and Farhadi, 2018). We pre-train it on the ImageNet dataset (Deng et al., 2009) and fine tune it for pig detection by replacing the classification grid layer to predict only 2 classes (background and pig).	The conclusions drawn from the aggregated data match the expectations of experts and justify our claim that collective behaviour statistics are accurate, even though individual frame level labels may not always be as accurate. This is valid not only for actions performed frequently (e.g. lying), but also for those occurring less often (e.g. eating or drinking).

<table>
<tr><th>Reference</th><th>Species/ category</th><th>Objective</th><th>Annotation</th><th>Remarks</th><th>Conclusions</th></tr>
<tr>
<td></td>
<td></td>
<td>at the aggregated week level are reliable and robust to error in the various steps of the pipeline; A public available dataset comprising 7200 fully annotated frames. https://aimagelab.ing.unimore.it/go/pigs-behaviours.</td>
<td></td>
<td>
Table 2: Metrics from the detector on the validation set. We report results for individual sequences as well as those from the whole validation set
<table>
<tr><th>Validation sequence</th><th>AP (%)</th><th>TP (%)</th><th>FP (%)</th><th>Missed (%)</th></tr>
<tr><td>A</td><td>84.63%</td><td>89.18%</td><td>10.82%</td><td>1.16%</td></tr>
<tr><td>B</td><td>97.28%</td><td>99.59%</td><td>0.41%</td><td>2.24%</td></tr>
<tr><td>C</td><td>100.00%</td><td>100.00%</td><td>0.00%</td><td>0.00%</td></tr>
<tr><td>D</td><td>95.75%</td><td>96.97%</td><td>3.03%</td><td>0.85%</td></tr>
<tr><td>E</td><td>98.38%</td><td>99.45%</td><td>0.55%</td><td>1.03%</td></tr>
<tr><td>Whole set</td><td>95.21%</td><td>97.04%</td><td>2.96%</td><td>1.06%</td></tr>
</table>
Table 2 shows results in terms of Average Precision (AP), number of true positives (TP), false positives (FP) and missed detections on the validation set. We report statistics for the individual sequences and the average on the full validation set.

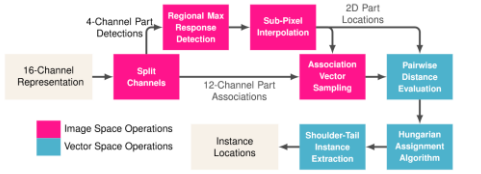
As tracker, we employ the MOSSE (Bolme et al., 2010) algorithm.

Table 3: Metrics from the tracker on the validation set. We report results for individual sequences as well as those from the whole validation set
<table>
<tr><th>Validation sequence</th><th>MOTA (%)</th><th>IDF$_1$ (%)</th><th># Switches</th><th># Fragmentations</th><th># Tracklets</th><th>Avg. tracklet length (# frames)</th></tr>
<tr><td>A</td><td>76.78%</td><td>55.10%</td><td>23</td><td>187</td><td>24</td><td>597</td></tr>
<tr><td>B</td><td>97.35%</td><td>88.39%</td><td>12</td><td>13</td><td>17</td><td>834</td></tr>
<tr><td>C</td><td>100%</td><td>100.00%</td><td>0</td><td>0</td><td>8</td><td>1800</td></tr>
<tr><td>D</td><td>92.97%</td><td>88.46%</td><td>9</td><td>43</td><td>24</td><td>597</td></tr>
<tr><td>E</td><td>97.92%</td><td>78.29%</td><td>12</td><td>18</td><td>13</td><td>1104</td></tr>
<tr><td>Whole set</td><td>93.00%</td><td>82.05%</td><td>11.2</td><td>52.2</td><td>17.2</td><td>986.4</td></tr>
</table>
</td>
<td>Future improvements can be envisioned for this challenging task. On the one hand, single components (e.g. the detection algorithm) could be specialized for the setting. On the other hand, given that the errors in the different stages of the pipeline compound, a single end-to-end method for detection-tracking-behaviour is also a possible future outcome. The detection ground truth could be refined to use ellipses instead of axis aligned bounding boxes. Another direction for extensions is increasing the number or breakdown of the behaviour classifications.</td>
</tr>
</table>

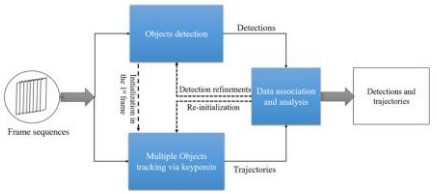
Reference	Species/ category	Objective	Annotation	Remarks	Conclusions
Long-Term Tracking of Group-Housed Livestock Using Keypoint Detection and MAP Estimation for Individual Animal Identification / **(Psota et al., 2020)**	Pigs (nursery, early finisher phase, finisher phase)	A probabilistic tracking-by-detection method is proposed. The tracking method uses, as input, visible keypoints of individual animals provided by a fully-convolutional detector	human-annotations, where both the shoulder-tail location and ear tag ID are provided for each animal in each frame. -Ear Tag Classification: http://psrg.unl.edu/Projects/Details/12-Animal-Tracking. 15 videos, each of which is 30 min in duration (The five rows correspond to high activity during the day, medium activity during the day, low activity during the day, medium activity during the night, and low activity during the night). The resolution of the videos is 2688 × 1520 and each was captured and annotated at 5 frames per second (fps)	**Table 2.** Precision/recall results for all 15 videos in the human-annotated dataset. The precision/recall results in "Location" do not require the tracker to provide the correct ID for animals. Instead, it is only required that each animal's location is matched with a detection. The "Location and ID" results require the tracker to correctly identify the location and correct ID of a pig in order to be counted as a true positive. The "(Uninitialized)" variant does not provide the location and ID of each pig in the first frame, whereas the "(Initialized)" variant does. **Location** Age \ Activity: High (Day) / Medium (Day) / Low (Day) / Medium (Night) / Low (Night) / Average Nursery: 0.9267 / 0.9964 / 0.9985 / 0.9548 / 0.8405 / **0.9434** Early Finisher: 0.9961 / 0.9973 / 1.0000 / 0.9549 / 1.0000 / **0.9857** Late Finisher: 0.9907 / 0.9890 / 0.9969 / 0.9564 / 1.0000 / **0.9866** **Average**: **0.9711** / **0.9943** / **0.9984** / **0.9487** / **0.9468** / **0.9719** **Location and ID (Initialized)** Age \ Activity: High (Day) / Medium (Day) / Low (Day) / Medium (Night) / Low (Night) / Average Nursery: 0.8893 / 0.9941 / 0.9933 / 0.8958 / 0.6256 / **0.8796** Early Finisher: 0.9949 / 0.9847 / 1.0000 / 0.8716 / 1.0000 / **0.9702** Late Finisher: 0.9836 / 0.9580 / 0.9897 / 0.8569 / 0.8462 / **0.9269** **Average**: **0.9559** / **0.9789** / **0.9943** / **0.8748** / **0.8239** / 0.9256 **Location and ID (Uninitialized)** Age \ Activity: High (Day) / Medium (Day) / Low (Day) / Medium (Night) / Low (Night) / Average Nursery: 0.8893 / 0.9941 / 0.6940 / 0.7927 / 0.6108 / **0.7962** Early Finisher: 0.9948 / 0.9718 / 1.0000 / 0.8946 / 0.5888 / **0.8900** Late Finisher: 0.9836 / 0.8176 / 0.9897 / 0.6290 / 0.5252 / **0.7890** **Average**: **0.9559** / **0.9278** / **0.8946** / **0.7721** / **0.5749** / **0.8251**	This paper presents a method for long-term tracking of individual livestock in group-house settings. This method takes advantage of the power of deep convolutional neural networks to detect individual targets and classify their identities. A probabilistic framework is used to efficiently combine per-frame detection and classification across long frame sequences. The publicly available, human-annotated dataset introduced in this work can be used to evaluate performance for long-term tracking of group-housed livestock. By representing a variety of different environments, ages/sizes of animals, activity levels, and lighting conditions, the

Reference	Species/ category	Objective	Annotation	Remarks	Conclusions
					dataset exposes the strengths and weaknesses of tracking methods. Results demonstrate that the method achieves an average precision and recall greater than 0.9 across a variety of challenging scenarios. While this work focuses on pigs, it is expected that the underlying techniques could easily be adopted to a variety of other livestock animals. This location and orientation tracking method could be used as the foundation for a more sophisticated tracker of activity and behaviour. In terms of extracting activities, it would be relatively straight-forward to convert the image-space tracking outputs to pen-space distance traveled using known camera parameters

Reference	Species/ category	Objective	Annotation	Remarks	Conclusions
					and pose estimation to the pen space. Eating, drinking, and social interactions can be approximated from proximity of targets to fixed landmarks and other targets. In this work, industry-standard ear tags were used for visual identification. Ideally, long-term tracking of individuals could be achieved without augmenting targets. However, the homogeneity of livestock populations makes it difficult to discern differences between individuals. Preliminary work suggests that this might be possible using facial recognition [71], but applications to long-term tracking are untested and facial recognition would likely require addition

Reference	Species/ category	Objective	Annotation	Remarks	Conclusions
					cameras in the pen space to get close-up shots.
Multi-Pig Part Detection and Association with a Fully-Convolutional Network / **(Psota et al., 2019)**	Pigs (ages from 1.5 to 5.5 months)	The first stage of the process aims to find the location of all pertinent body parts, while the second stage aims to associate them with one another to form whole instances.	(http://psrg.unl.edu/Projects/Details/12-Animal-Tracking) - Each unique image was randomly extracted from video recordings spanning multiple weeks in each location. More than two hours, on average, existed between samples at each location - In each of the images, a user manually annotated the location of the left ear (red), right ear (green), shoulder (blue), and tail (yellow) for each visible animal in that order. Annotations belonging to the same instance are connected with a continuous black line. If ears were not visible, they were not annotated, however, emphasis was placed on annotating both shoulders and tail for each instance even when these locations are	The method was implemented in Matlab 2018b using the deep learning toolbox.  **Figure 4.** Flow diagram of the proposed method for converting the 16-channel image space representation to a set of 2D coordinates of each visible instance. - 1600 images for training and 400 images for testing. Furthermore, the 400 testing images were subdivided into two additional subsets: 200 captured in the same environments seen in the training set (test: seen), and 200 images from environments previously unseen in the training set (test: unseen). - Results demonstrate that the method is capable of achieving over 99% precision and over 95% recall at the task of instance detection when the network is tested and trained under the same environmental conditions. When testing on environments and lighting conditions that the network had not been trained to handle, the results dropped significantly to 91% precision and 67% recall.	A major contribution to this work is the introduction of an image space representation of each pig as a collection of body parts along with a method to join parts together to form full instances.

Reference	Species/ category	Objective	Annotation	Remarks	Conclusions
			occluded, i.e., both shoulder and tail were annotated as long as they are located in the pen of interest and their estimated positions were within the field of view of the camera. - areas of interests were defined by polygons for each image in the dataset and masking out was done by setting all pixels outside the area of interest to pure black		
Automatic Individual Pig Detection and Tracking in Pig Farms / **(Zhang et al., 2019)**	Finisher pigs (9)	We propose a data association algorithm to complementarily bridge the detection and tracking processes based and track on the contextual cues in sequential frames		- there are a total of 18,000 frames for algorithm training and a total of 4200 frames coming from five different sequences for testing - Our method is trained and evaluated on 22,200 frames captured from a commercial farm. Overall, the evaluation results in a precision of 94.72%, recall of 94.74%, and MOTA of 89.58% We implemented the individual pig detection and tracking method using Matlab and MatConvNet CNN. - three CNN detection architectures (Faster-RCNN, R-FCN and SSD) were implemented and	

Reference	Species/ category	Objective	Annotation	Remarks	Conclusions
				compared. The backbone architecture for the three detection networks is VGG16. - We evaluated our method using MOT evaluation metrics  **Figure 2.** The general framework of proposed method, which couples a CNN-based detection and a multiple objects tracker via a novel hierarchical data association algorithm.	