# Towards real-time analysis systems for automated phenotyping of livestock

***AI for better animal welfare and smaller footprint***

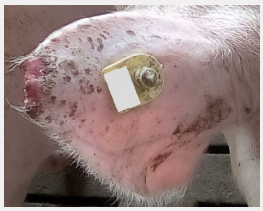
**Peter H.N. De With, Professor TU/e VCA**

With contributions from: Dr.ir.Patrick Langenhuizen, Ass. Prof. TU/e, Qinghua Guo (PhD), Shoujun Huo (PhD)
Piter Bijma, Assoc. Prof. at WUR



## 1.1 Intro: Problem statement

- **Animal health** and **welfare** are increasingly important
- Freedom to perform wide range of (social) **behaviours**
- Improvement strategies
  - **Better housing and care**
  - **Genetic selection for sociality**



# 1.2 Intro: Consortium

**Academia**

TU/e – Video Coding and Architectures

WUR – Animal Breeding and Genomics

WUR – Business Economics

**Breeding sector**

Hendrix Genetics

Topigs Norsvin

**Technology sector**

VencoMatic Group

FarmResult

Theta Vision

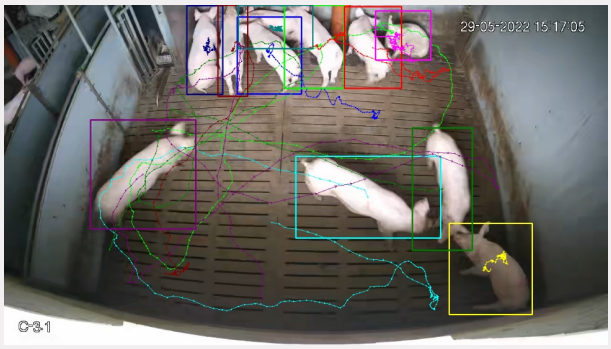




# 1.3 Intro: State of the art

Ongoing projects

- **IMAGEN**
  NWO-TTW Perspective program
- **SmartTurkeys**
  NWO-TTW Open Technology Programme



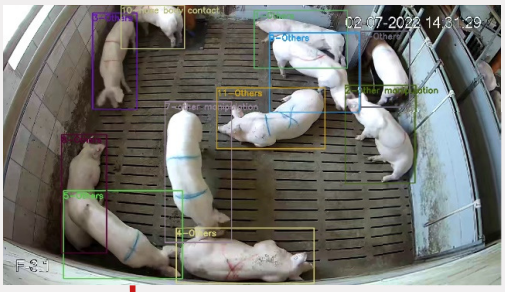

- Tracking and (damaging) behaviour detection
- Inference time approximately 10-15 FPS on single A100 GPU (NVIDIA corp.)
- TRL4 – TRL5

# 1.4 Intro: Project goal

**Fully programmable system**

- **Real-time** operation
- **Flexible** and **rapid** employment in wide range of **field environments**

Piling detection



Tracking and detection

Social genetic effects





# 1.5 Intro: Research plan - Objective

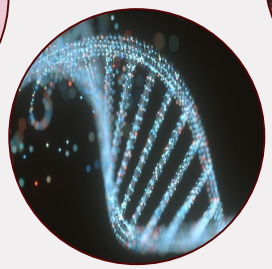
Real-time sensing and behavioural analysis using AI techniques for the automated detection and tracking of phenotypes in groups of livestock.

**Digital architecture**

**Implementation**

**Evaluation**

**Genetic effects**



**Farmer adoption**

## Required Technical Steps in Algorithm Development

- Object Detection & Localization
- Object Classification
- Object tracking
- Posture analysis & Motion analysis
- Animal behavior
- System architecture and mapping to fast computation & algorithms
- System integration & validation
- …….but, we started and are still busy with data collection!





## P1: Single-Camera Based Multi-animal Tracking

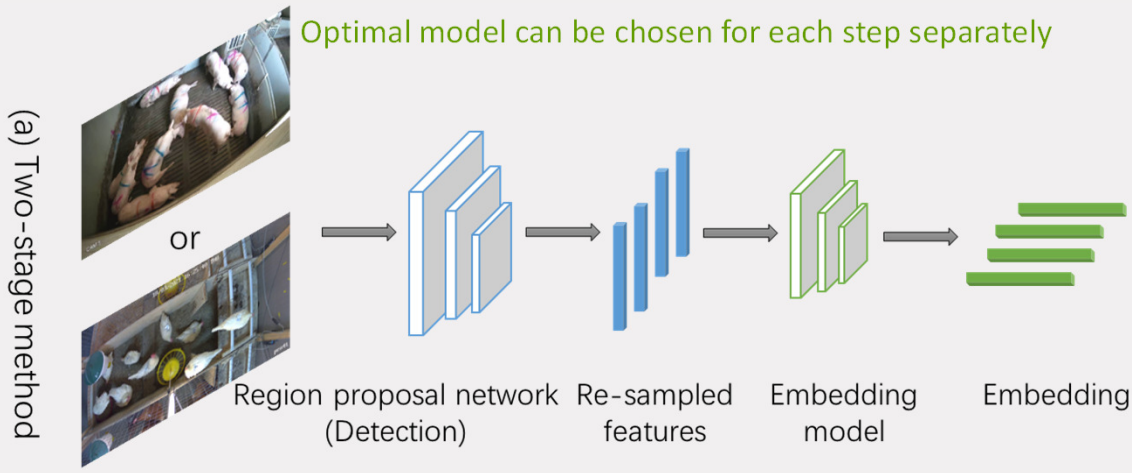
With contributions from PhD students Qinghua Guo and Shoujun Huo

# P1 Rel. work-1: Artificial Intellig. (AI) with 2D RGB cameras

## Multiple animal tracking based on single-camera views

I. Two-stage method : Expensive computation cost → slow processing speed

Optimal model can be chosen for each step separately



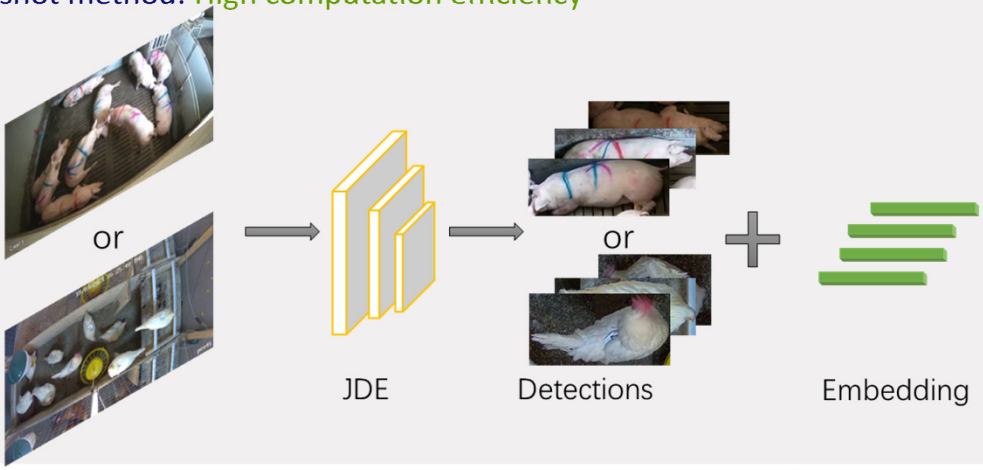





# P1 Rel. work-2: Artificial Intellig. (AI) with 2D RGB cameras

## Multiple animal tracking based on single-camera views

II. One-shot method: High computation efficiency



# P1 Multiple animal tracking - SOTA: Evaluation metrics

| Metric | Description |
|---|---|
| MOTA↑ | Multi-Object Tracking Accuracy. This measure combines three error sources: false positives, false negatives and identity switches. |
| MOTP↑ | Multi-Object Tracking Precision. The misalignment between the annotated and the predicted bounding boxes. |
| MT↑, PT, ML↓ | Number of mostly tracked, partially tracked and mostly lost trajectories. |
| IDF1↑ | ID F1 score. The ratio of correctly identified detections over the average number of ground-truth and computed detections. |
| IDs↓ | Number of identity switches. |
| FPS↑ | Execution time, frame per second. |





# P1 1. Background on Detection & Tracking

- Two primary ways to implement a tracking system

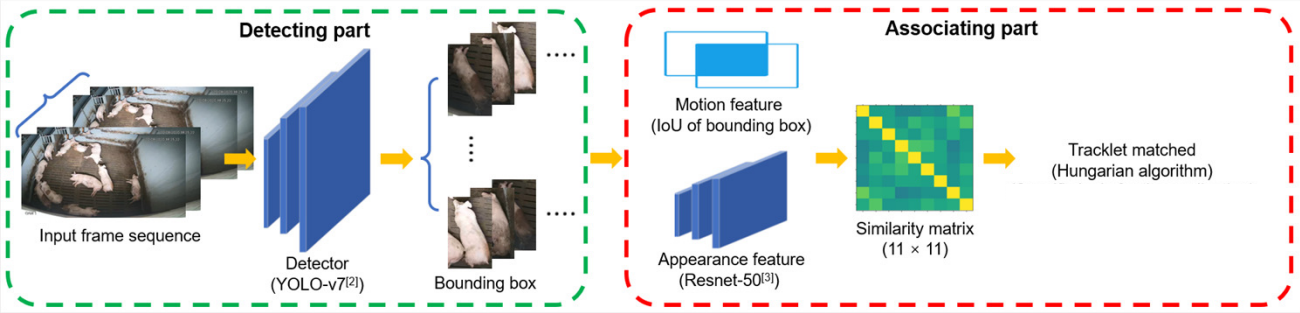
| | Performance | Computational cost |
|---|---|---|
| Joint Detection and Association | Lower | Lower |
| Separate Detection & Association | Higher | Higher |

*Why is Separate Detection and Association performing better?*

- Detection and association require different scales of features (association requires more delicate vision features than detection)
- Integrating both tasks into one-shot network: **a conflict** that will decrease the tracking performance.

# P1 2. Designing a framework for animal tracking - Method

Framework handling detection and association as distinct tasks




1) Animal detection: Identify the animals in each frame.

2) Association: Compare the similarity based on location & appearance, link the best matching pairs.





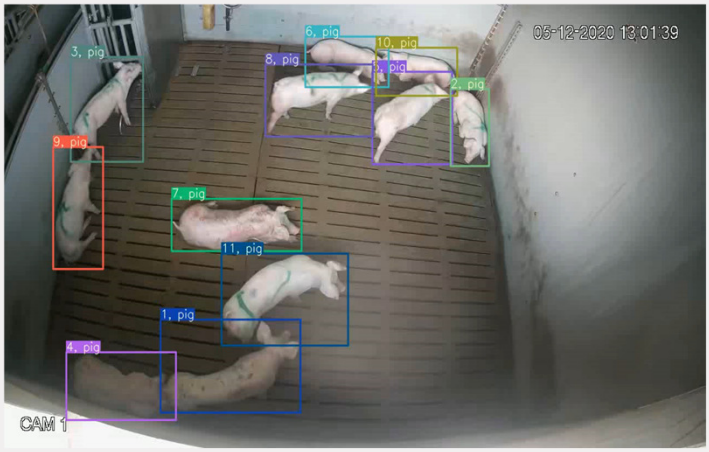
# P1 2. Designing a framework for animal tracking - Results

*Results on videos with 11 pigs*

*ID Switch: Wrongly assigned identities;*
*IDF1: The ability of a tracking system to correctly identify the objects being tracked;*
*MOTA: The ability of a tracking system to correctly detect the objects being tracked*

| | IDF1 | ID Switch | MOTA | | IDF1 | ID Switch | MOT A | | IDF1 | ID Switch | MOTA |
|---|---|---|---|---|---|---|---|---|---|---|---|
| V1 (34min) | 99.5% | 2 | 99.4% | V14 (10min) | 99.8% | 0 | 99.7% | V27 (10min) | 99.8% | 0 | 99.5% |
| V2 (34min) | 100% | 0 | 100% | V15 (10min) | 100% | 0 | 100% | V28 (10min) | 95.1% | 2 | 99.8% |
| V3 (34min) | 96.9% | 2 | 99.5% | V16 (10min) | 99.9% | 0 | 99.9% | V29 (10min) | 99.8% | 0 | 99.7% |
| V4 (3min) | 95.3% | 2 | 98.8% | V17 (10min) | 100% | 0 | 100% | V30 (10min) | 99.9% | 0 | 99.8% |
| V5 (3min) | 99.6% | 0 | 99.3% | V18 (10min) | 96.0% | 2 | 99.3% | V31 (10min) | 94.7% | 2 | 99.7% |
| V6 (3min) | 99.8% | 0 | 99.6% | V19 (10min) | 99.6% | 0 | 99.3% | V32 (10min) | 100% | 0 | 99.9% |
| V7 (3min) | 100% | 0 | 100% | V20 (10min) | 99.3% | 0 | 98.7% | V33 (10min) | 100% | 0 | 100% |
| V8 (3min) | 99.8% | 0 | 99.6% | V21 (10min) | 100% | 0 | 100% | V34 (10min) | 100% | 0 | 99.9% |
| V9 (3min) | 100% | 0 | 100% | V22 (10min) | 98.0% | 2 | 99.1% | V35 (10min) | 100% | 0 | 100% |
| V10 (3min) | 100% | 0 | 99.9% | V23 (10min) | 99.9% | 0 | 99.8% | V36 (10min) | 93.5% | 2 | 99.8% |
| V11 (3min) | 100% | 0 | 100% | V24 (10min) | 100% | 0 | 100% | V37 (10min) | 91.5% | 2 | 99.9% |
| V12 (3min) | 100% | 0 | 100% | V25 (10min) | 100% | 0 | 100% | V38 (10min) | 96.9% | 2 | 99.5% |
| V13 (3min) | 100% | 0 | 100% | V26 (10min) | 91.9% | 4 | 88.8% | **avg.** | **98.59%** | **0.63** | **99.43%** |

# P1 3. Multiple animal tracking visualization - experiment




Finding: System performs well during intense moving and when occlusion occurs.





# P1 4. Discussion on Tracking Framework

**Detection part:**

Although bounding boxes work well in the system, when pigs gather, one pig's bounding box may overlap significantly with another, making it difficult to use location and appearance cues for the subsequent association phase.

**Association part:**

We consistently match animals in the current frame with those in the previous frame. Is this approach reasonable? What if I reverse the video and apply the same method?

# P1 5. Detection: bounding-box vs. instance segmentation

For both cases, the same association strategy is applied to obtain the identities, the only thing different is the bounding-box detector and segmentation detector, respectively.

Bounding-box tracking


ID-switch occurs between pig_8, pig_10, and pig_11

Instance segmentation tracking


No ID-switch occurs





# P1 7. Context understanding – how to exploit this?

Natural Language Processing (NLP) usually uses a self-attention mechanism to understand the context.

Each row in this matrix captures not only the meaning in the sentence but also each word's interaction with other words

| | YOUR | CAT | IS | A | LOVELY | CAT |
|---|---|---|---|---|---|---|
| YOUR | 0.268 | 0.119 | 0.134 | 0.148 | 0.179 | 0.152 |
| CAT | 0.124 | 0.278 | 0.201 | 0.128 | 0.154 | 0.115 |
| IS | 0.147 | 0.132 | 0.262 | 0.097 | 0.218 | 0.145 |
| A | 0.210 | 0.128 | 0.206 | 0.212 | 0.119 | 0.125 |
| LOVELY | 0.146 | 0.158 | 0.152 | 0.143 | 0.227 | 0.174 |
| CAT | 0.195 | 0.114 | 0.203 | 0.103 | 0.157 | 0.229 |

(6, 6)

$$Attention(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

X V (6, 512) = Attention (6, 512)

## P1 9. Conclusions and Future work

- Video-based monitoring with computer vision technology gives promising results for object detection and tracking, and high scores can be obtained.
- Animal movements, trajectories, social networks at group level and individual level can be reconstructed as informative events and useful phenotyping data.
- But in near future:

  1. Stitch to swift instance segmentation as detector to increase the Signal-to-Noise Ratio

  2. Tracking the video from front to back and back to front yields different results, hence re-design an association strategy based on a better understanding of context.





## P2 Animal keypoint detection and posture estimation based on single images

- Animal keypoint detection and posture recognition at the frame level
- Later: P3 Animal behavior recognition at the segment level

With contributions from PhD student Qinghua Guo

# P2. 1. Pig keypoint detection & posture recognition - Workflow

**Keypoint detectors**

KD - SOTA 1 - YOLOv8 - m1 (pre-trained on a large dataset)

Test

- Norwegian data (keypoint, bbox, segmentation) → Evaluation metrics → PCK, PDJ, OKS
- Volmer data (keypoint, bbox, posture) → Evaluation metrics → PCK, PDJ

**Pose estimators**

- PE - SOTA 1 - YOLOv8 - m2 (image+keypoint-based)
- PE - SOTA 2 - ResNet (image-based)
- PE - SOTA 3 - MLP (keypoint-based)

Test → Volmer data (keypoint, bbox, posture) → Evaluation metrics → Top 1 accuracy



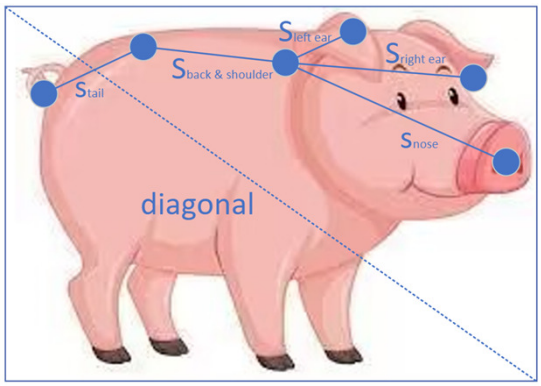


# P2 2. Keypoint detection - SOTA: Evaluation metrics

## 1. Percentage of Detected Joints (PDJ)

- $d_i$ = Euclidean distance between ground truth and predicted point
- diagonal = diagonal of the bounding box
- n = the number of keypoints
- bool = Boolean number (0|1)

$$PDJ = \frac{\sum_{i=1}^{n} bool(d_i < 0.05 * diagonal)}{n}$$

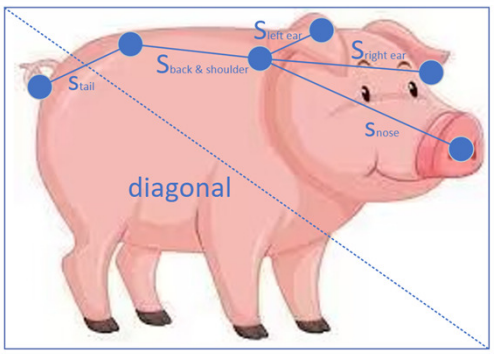


# P2. 3. Keypoint detection - SOTA: Evaluation metrics

## 2. Percentage of Detected Joints (PCK)

- $d_i$ = Euclidean distance between ground truth and predicted point
- s_i = skeleton length of the animal (PCK)
- n = the number of keypoints
- bool = Boolean number (0|1)

$$PCK = \frac{\sum_{i=1}^{n} bool(d_i < 0.5 * s_i)}{n}$$








# P2 4. Keypoint detection - SOTA: Evaluation metrics

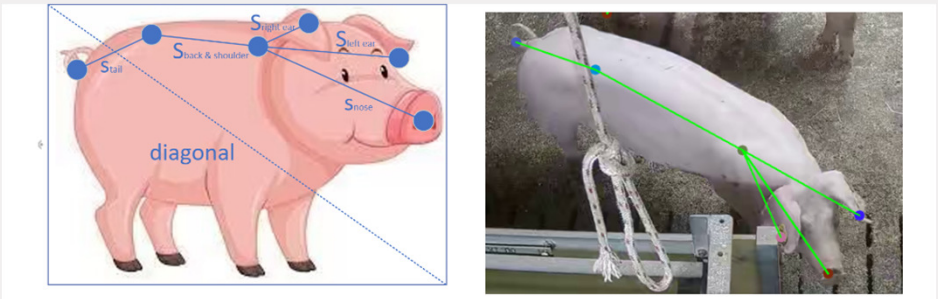
## 3. Object Keypoint Similarity (OKS)

- $d_i$ = Euclidean distance between ground truth and predicted point
- A = segmentation area of object (OKS)
- $k_i$ from human data set
- $k_i$ = [0.089,0.107, 0.079,0.026,0.035,0.035]
- n = the number of keypoints

$$OKS = \frac{\sum_{i=1}^{n} exp(-\frac{d_i^2}{2Ak_i^2})}{n}$$

# P 2 5. Pig keypoint detection + posture recognition - Dataset

- Norway dataset: 1,160 random frames, 6 annotated keypoints
  ["tail", "back", "shoulder", "nose", "left ear", "right ear"] in COCO format




(a) (b)

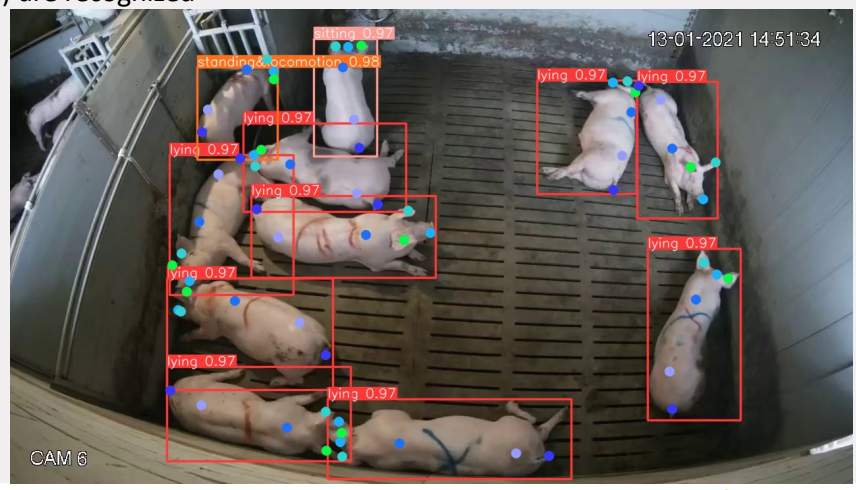
Figure 2: Localization information of six keypoints, including tail, back, shoulder, nose, left ear, and right ear. (a) is a visualized criterion for the pig keypoint annotation. (b) is a sample example in practice.





# P 2 7. Pig keypoint detection and posture recognition - Results

- Visualization of predicted keypoints and postures, 3 posture classes (lying, sitting, and standing & locomotion) are recognized



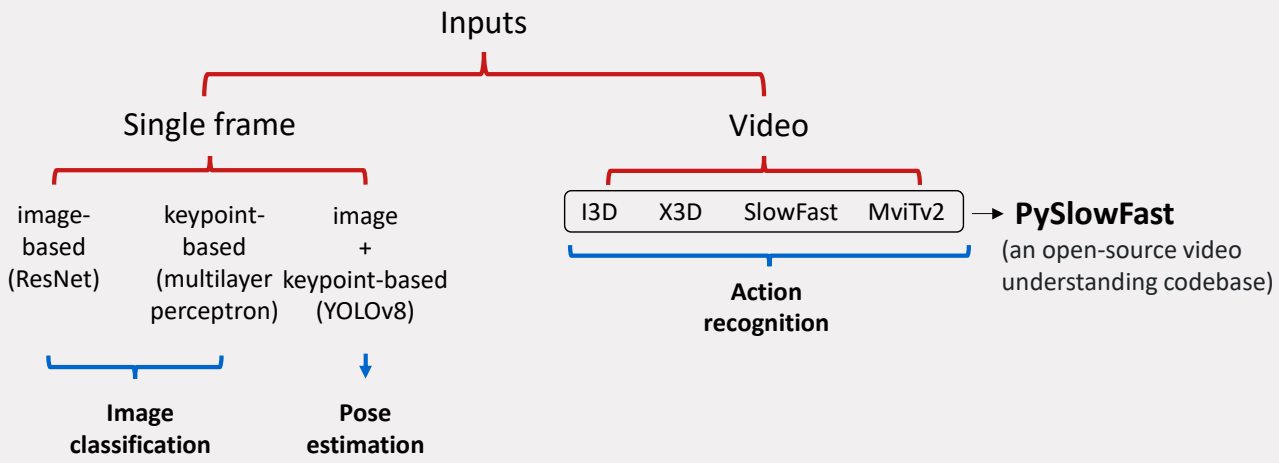
# P3 Animal behavior recognition at the video segment level

With contributions from PhD student Qinghua Guo





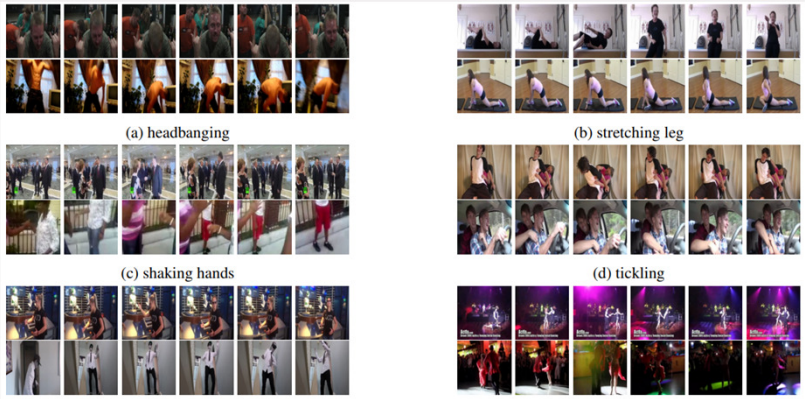
# P3 1. Action recognition/classification - SOTA methods



# P3 2. Action recognition/classification - Public human dataset

Kinetics Datasets (400/600)

The dataset contains **400 human action classes**, with **at least 400 video clips** for each action. Each clip lasts around **10s**



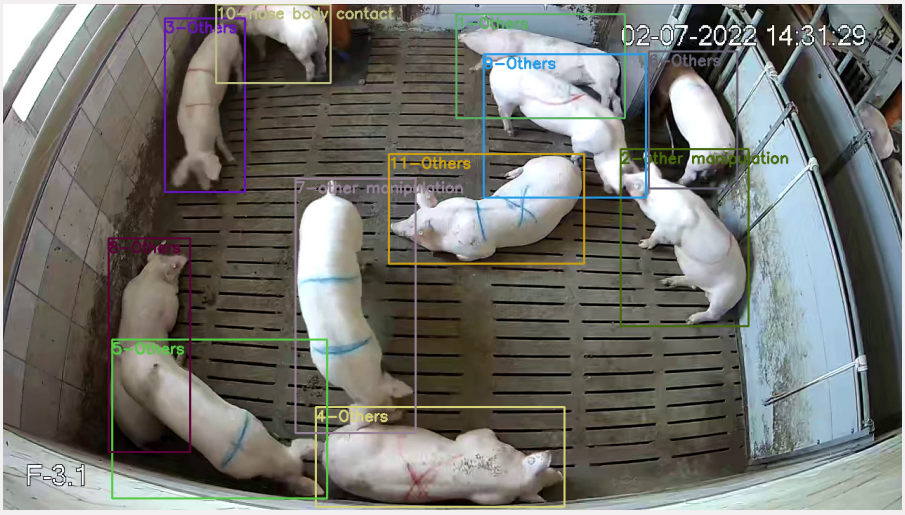





# P3 3. Action recognition/classification - Pig dataset

A small dataset with 5 behavior classes (ear manipulation, tail manipulation, other manipulation, nose body contact, Others)

**NB:** The dataset has an imbalanced distribution and insufficient data volume.



# P3 4. Action recognition/classification - SOTA: Eval. metrics

**Evaluation metrics:**

Top-1 accuracy: Amount of the model has predicted the correct label with the highest probability

$$\text{Top-1 Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}} \times 100\%$$

Top-5 accuracy: Amount of the correct label appears in the top 5 predicted classes

$$\text{Top-5 Accuracy} = \frac{\text{Number of correct predictions in top 5}}{\text{Total number of predictions}} \times 100\%$$

NB: Top 5 is reasonable in the big dataset with many (e.g. 400) classes, we use Top-2 accuracy in our case.





# P3 5. Action recognition/classification – First results

| Architecture | Top1 | Top5 | Dataset |
|---|---|---|---|
| I3D | 71.6 | 90.0 | Kinetics 400 |
| X3D -M | 76.0 | 92.1 | |
| SlowFast 8x8, R101 | 77.9 | 93.2 | |
| MViTv2 - S | 81.0 | 94.6 | |

| Architecture | Top1 | Top2 | Dataset |
|---|---|---|---|
| I3D | 86.92 | 96.26 | Our |
| X3D -M | 86.92 | 92.52 | |
| SlowFast 8x8, R101 | **87.85** | 94.39 | |
| MViTv2 - S | **87.85** | **97.20** | |

**NB:** The training dataset is enhanced with data augmentation; the testing set has insufficient samples for validating the model robustness and generalization.

# P4 Conclusions, Wrap up, and next steps

With contributions from Dr. P. Langenhuizen and Dr. Joost v.d. Putten





## P4 1. Conclusions on posture and behavior recognition

- Existing models developed for human behavior show also promising results for animal behavior classification.
  - More data is needed, fine-tuning for the animal case are both needed.
- At current status, posture recognition with key points combined with RGB information shows good performance compared to only RGB-based or only key-point-based recognition models. This will also benefit behavior recognition.
- Final objective is to finalize the processing chain of detection, tracking and posture behavior recognition make it efficient and executable.

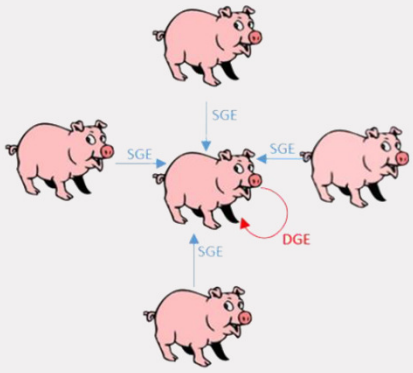
# P4 2. Research plan – Challenges

**Smart mapping**
- Rapid deployment
- Scalability

**Smart efficient algorithms**
- Robustness in identification
- Real-time execution

**Genomics**
- Integrating phenotypes and genomics
- Model development

**Final large-scale application**
- Adoption by farmers
- Impact of implementation

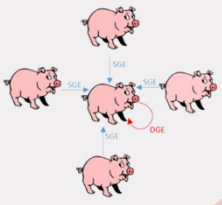







# P4 4. Key points

HENDRIX GENETICS
Topigs Norsvin
SmartTurkeys
an IMAI Group sENsor



KIC MISSIE AI for Agriculture, Horticulture, Water and Food Call

TU/e WAGENINGEN UNIVERSITY & RESEARCH
Topigs Norsvin HENDRIX GENETICS
FarmResult Vencomatic Group

# P4 5. Impact A

Develop AI solution for analyzing behaviour on individuals and on group level

- Timely intervention possible






# P4 5. Impact D

Develop AI solution for analyzing behaviour on individuals and on group level

- Insight in farm-economic advantages

# How - Programmable Falcon platform

- High-performance AI processing
- Versatile connectivity
- Pre-installed AI software
- Efficient power consumption
- Programmable functionality
- Hardware accelerated I/O
- Multi-threaded pipeline

- **AI Engine:** NVIDIA Jetson NX Orin 8GB
- **CPU:** 6-core Arm® Cortex®-A78AE 2 GHz v8.2 64-bit CPU,
- **System Memory:** 8GB 128-bit LPDDR5, 102.4 GB/s
- **Graphics:** 1024-core NVIDIA Ampere GPU with 32 Tensor Cores
- **Storage:** KIOXIA 256GB NVMe Storage
- **OS Support:** Linux with Jetpack OS 6.0
- **Security:** TPM 2.0
- **Deep Learning Accelerator:** 1x NVDLAv2 @ 614MHz
- **Vision Accelerator:** 1x PCA v2
- **Video Encoding:** Up to 6x 1080p60 (H.265)
- **Video Decoding:** Up to 9x 1080p60 (H.265)
- **Display:** 1x 8K30 multi-mode DP 1.4a





# How - Use case 1: medical research prototypes (cancer detection)

- Hybrid convolutional CNN-Transformer architecture
- Pretrained 5M endoscopy images
- 15k Labeled endoscopy images
- Real-time at 20 fps
- Multi-modal (white light, NBI)

# AI for better animal welfare and smaller footprint