

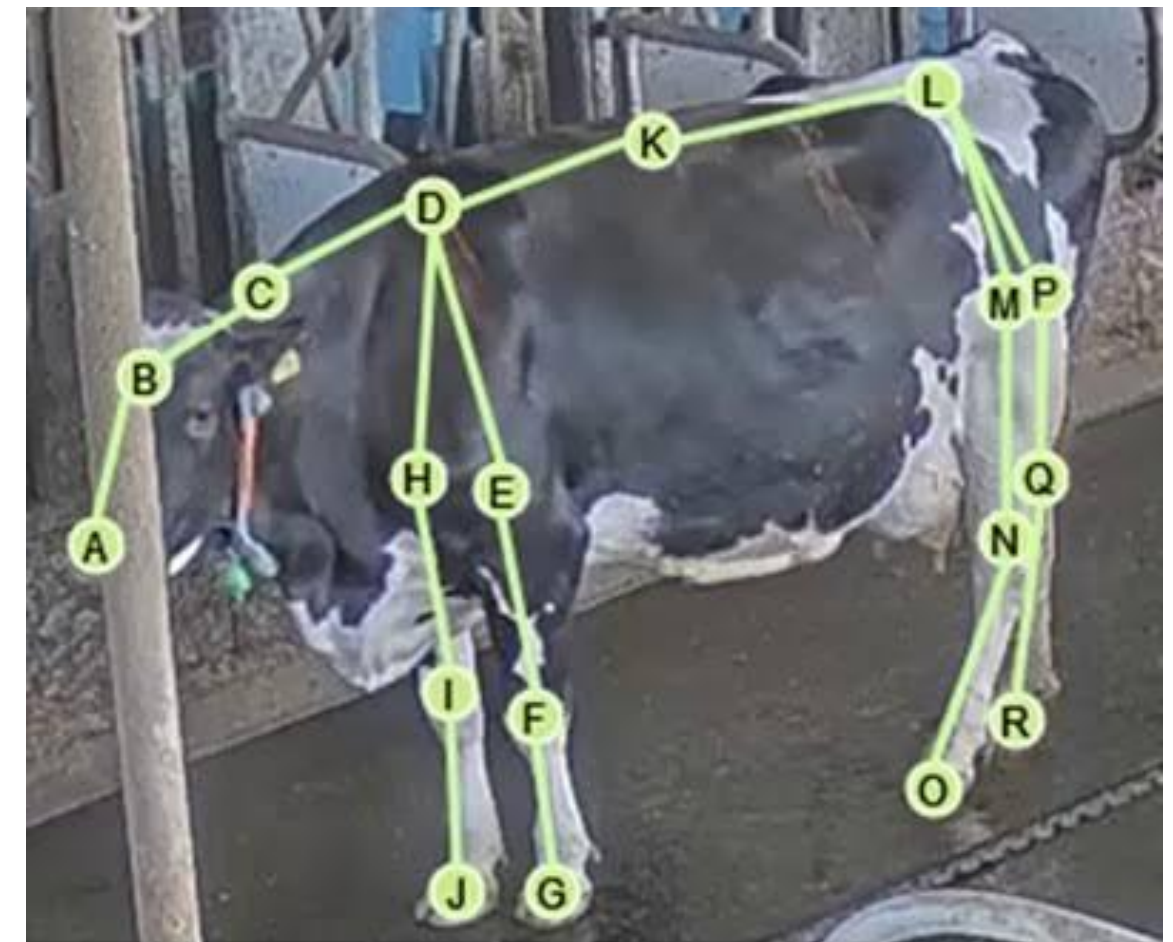
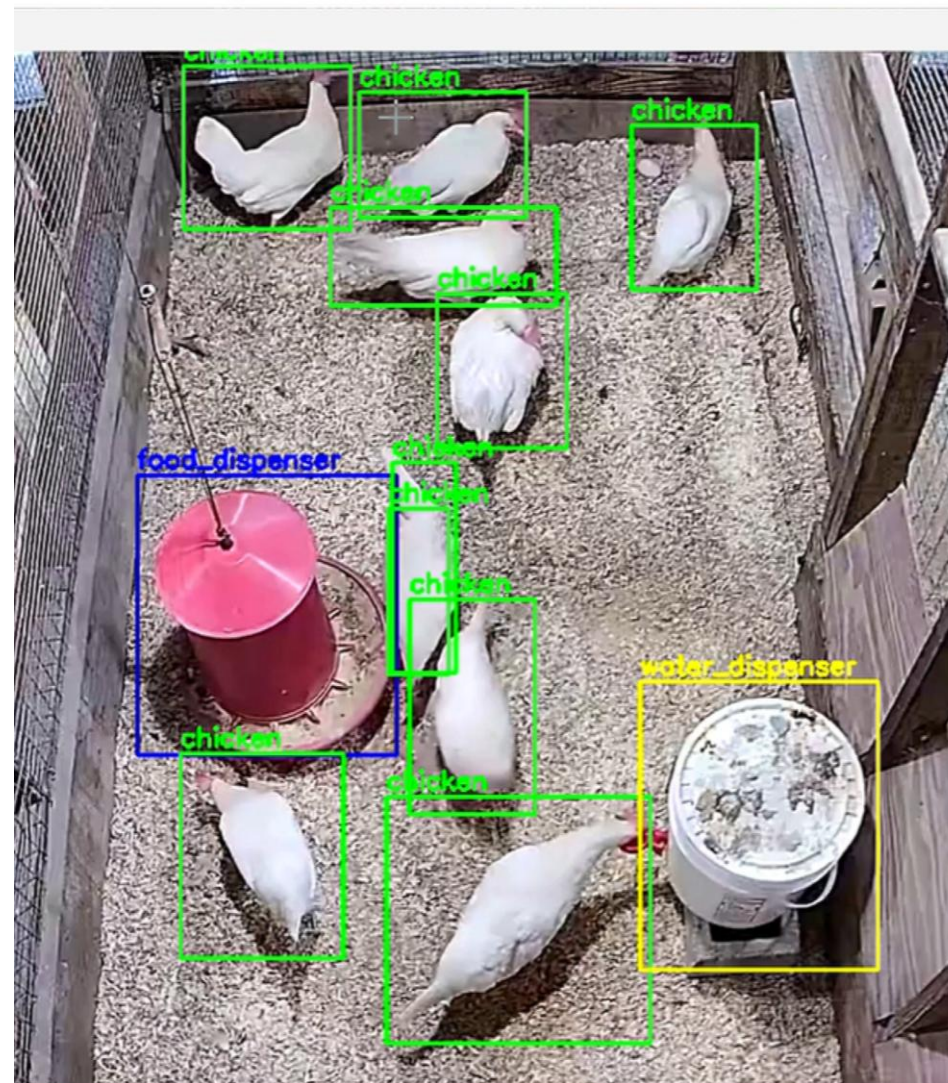
# FROM PIXELS TO INSIGHT: CHALLENGES AND OPPORTUNITIES

## FOR COMPUTER VISION IN ANIMAL MONITORING

Sam Leroux, Thorsten Cardoen

# IS PRECISION LIVESTOCK MONITORING JUST APPLIED COMPUTER VISION ?

- Most task boil down to “traditional” computer vision tasks:
  - Object detection
  - Object tracking
  - Action recognition
  - Pose estimation





# ISN'T COMPUTER VISION A SOLVED PROBLEM ?

- Trend towards very large foundation models
- Pretrained on huge amounts of data
- Can generalize to different domains and tasks



# CHALLENGES

- Lack of high quality representative datasets
  - Real world stocking densities
  - Covers entire animal lifetime
  - Labelled annotations
  - Rare events
- Scaling to the real world
  - Computational cost
  - Large setups
  - Robustness

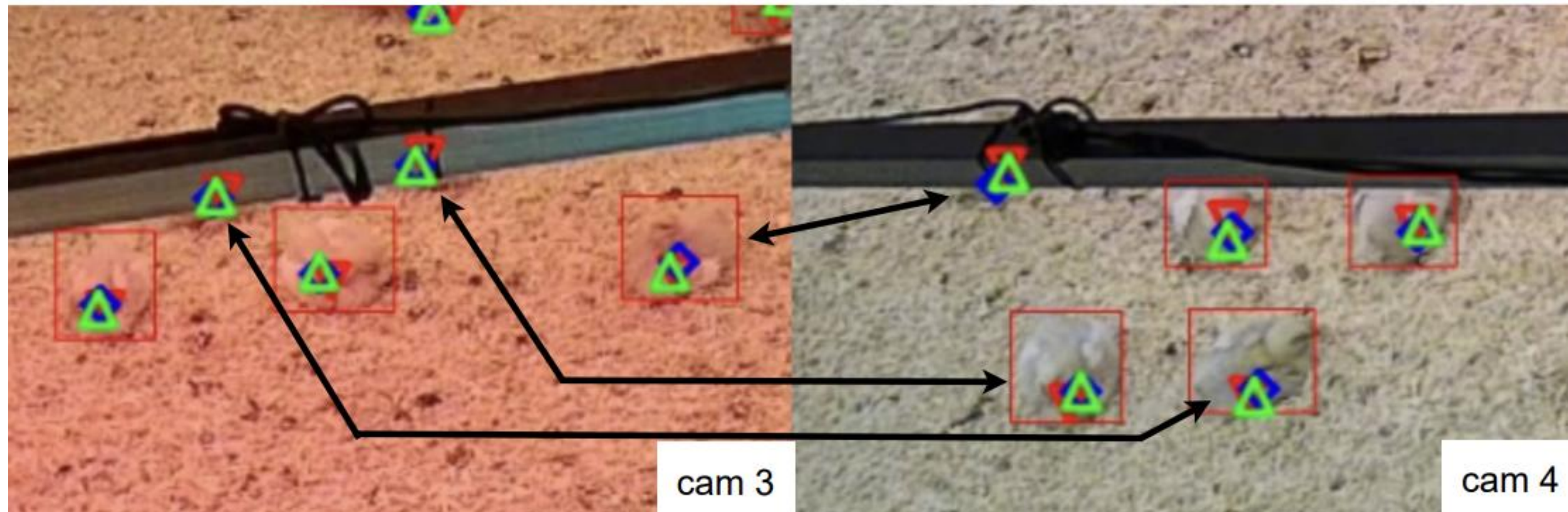
# EXAMPLE 1: MULTI-CAMERA DETECTION AND TRACKING





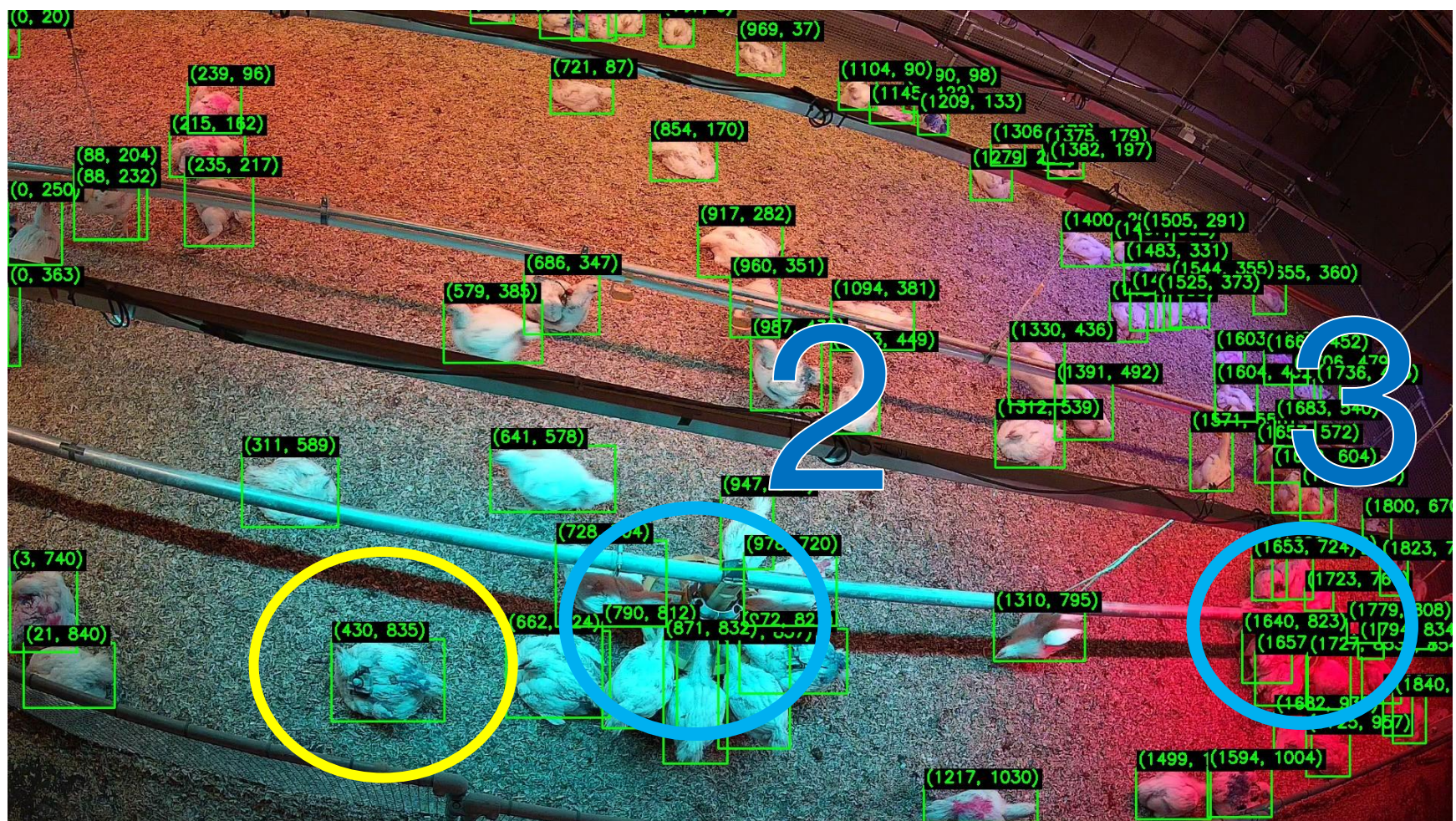
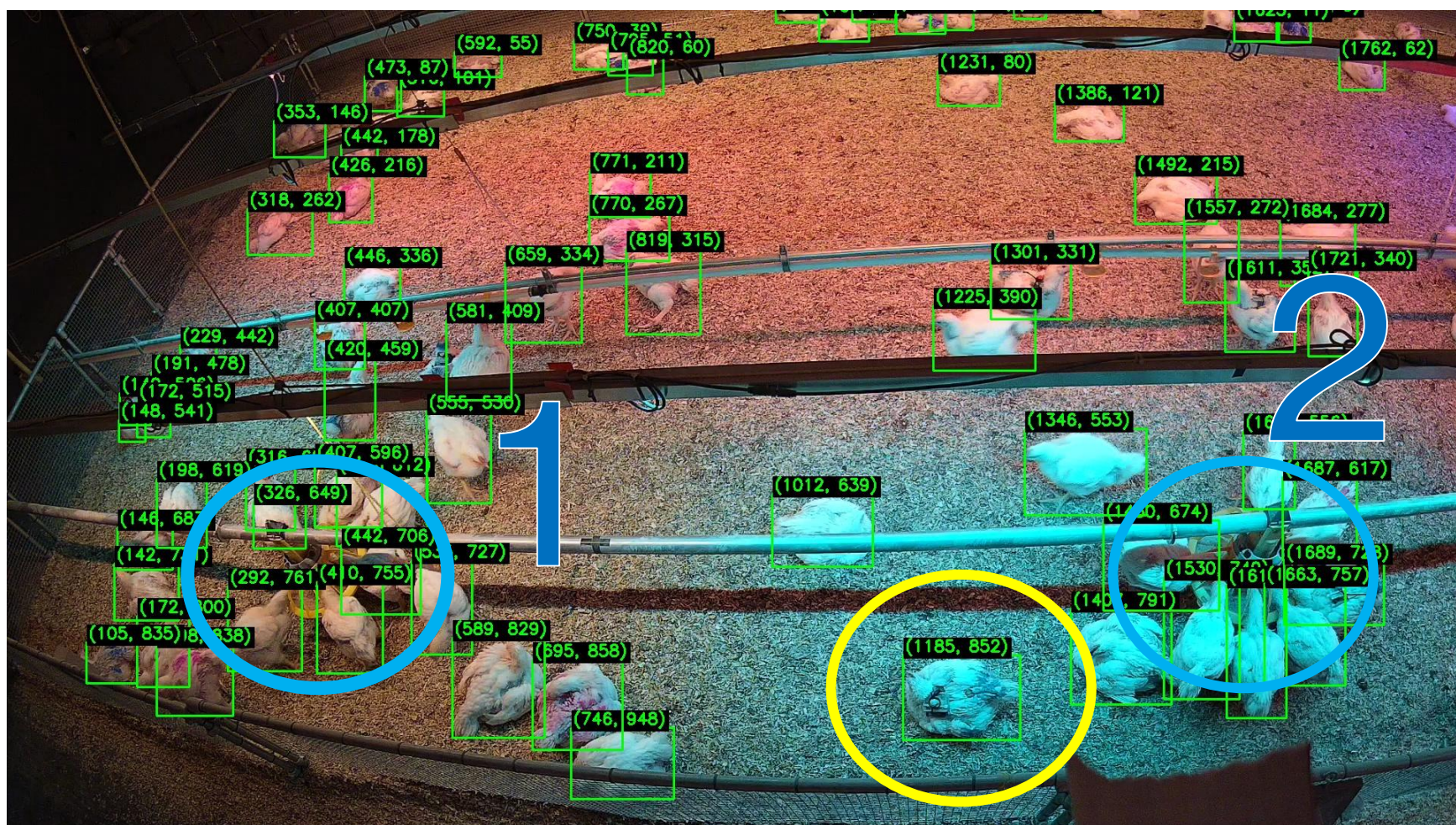


# Multi-camera setups can mitigate occlusions



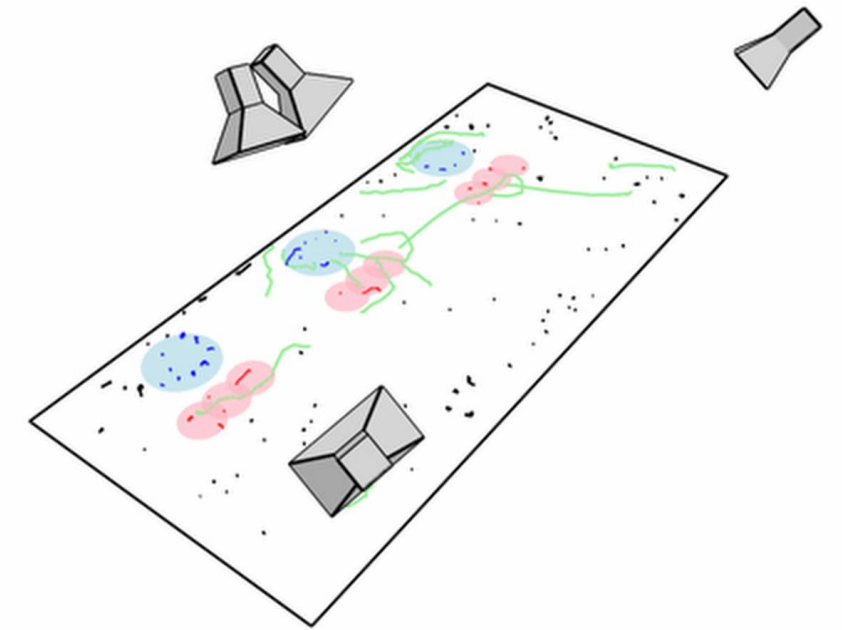
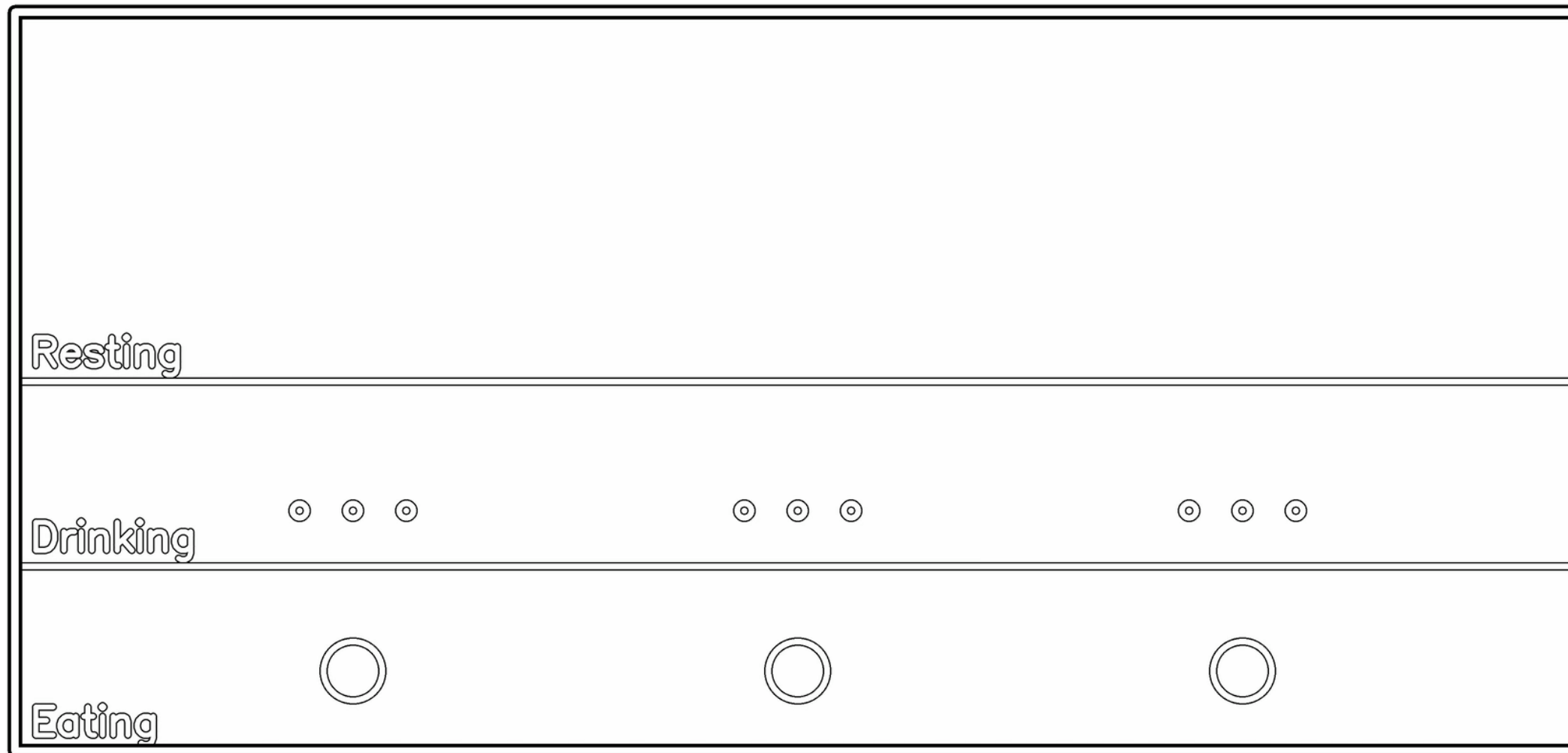


# Common reference frame is needed



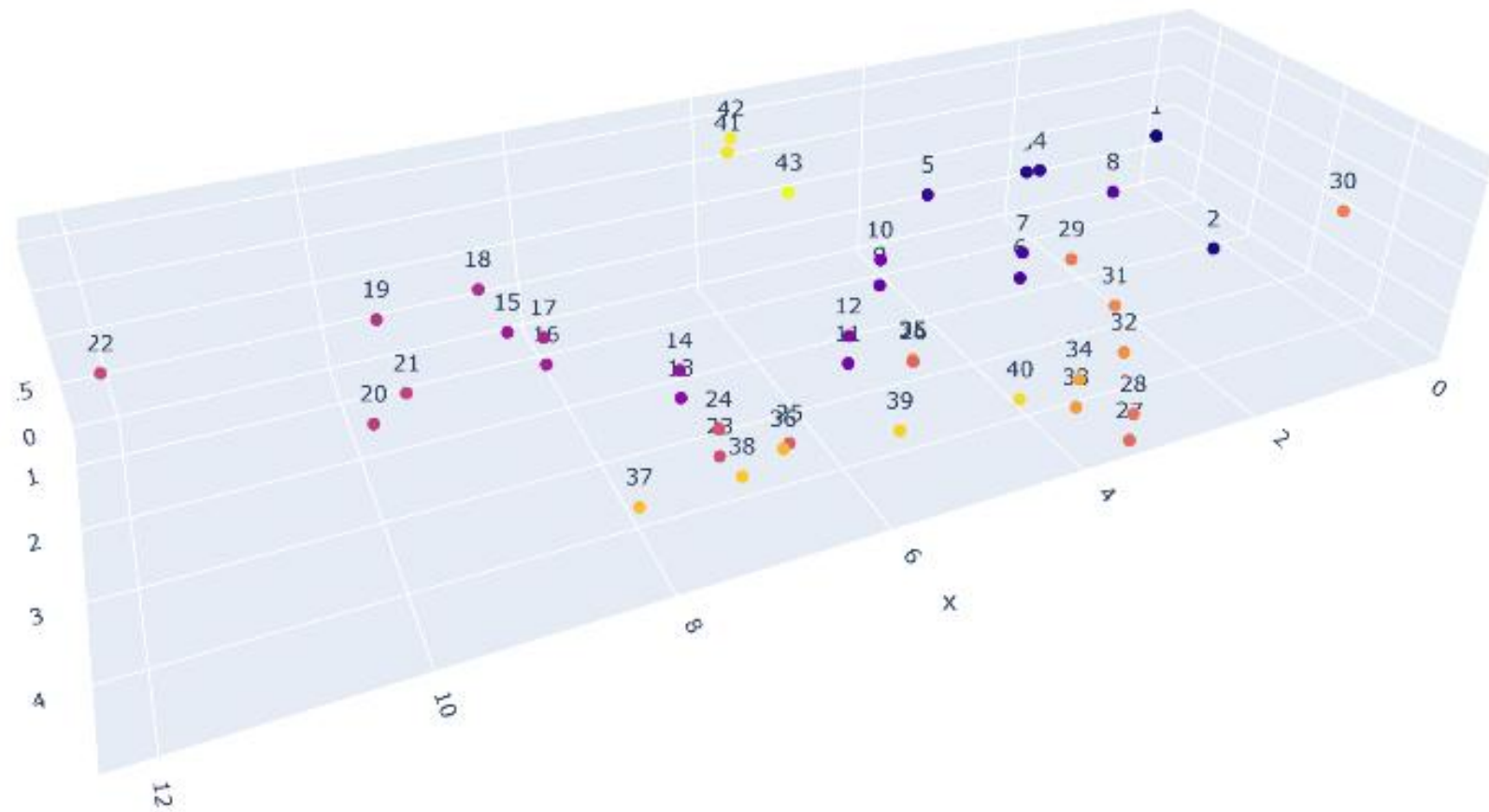


# The ground plane is used as a common reference frame

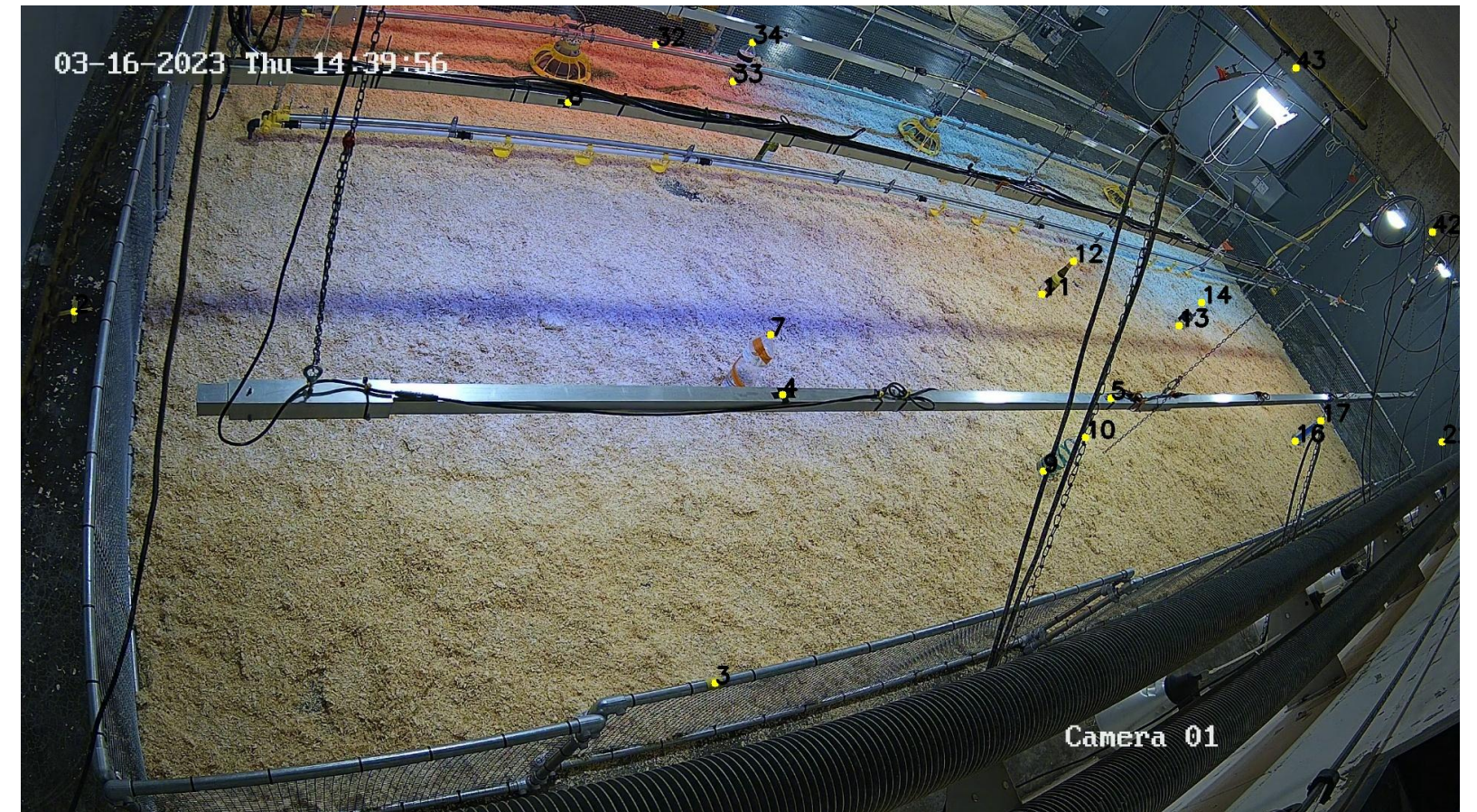




# Camera calibration through 3D to 2D correspondences



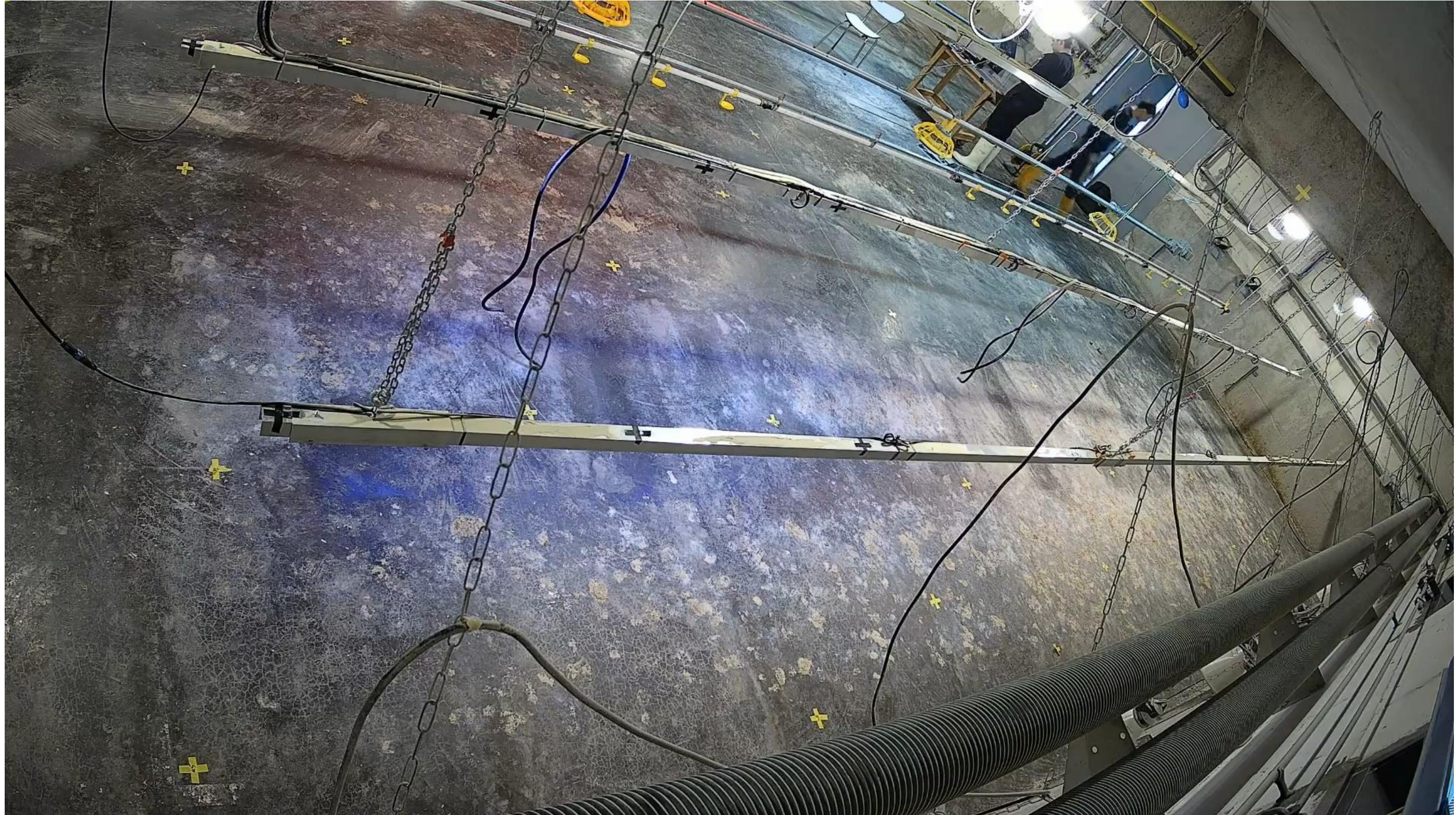
3D locations in metric units



2D locations in pixel space

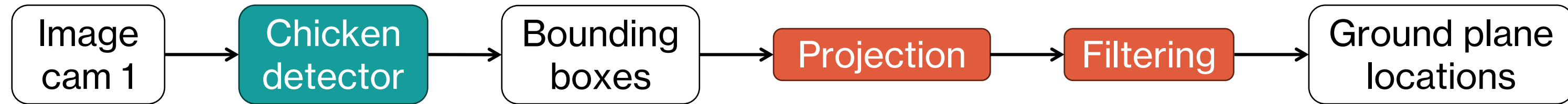


# Camera calibration through 3D to 2D correspondences



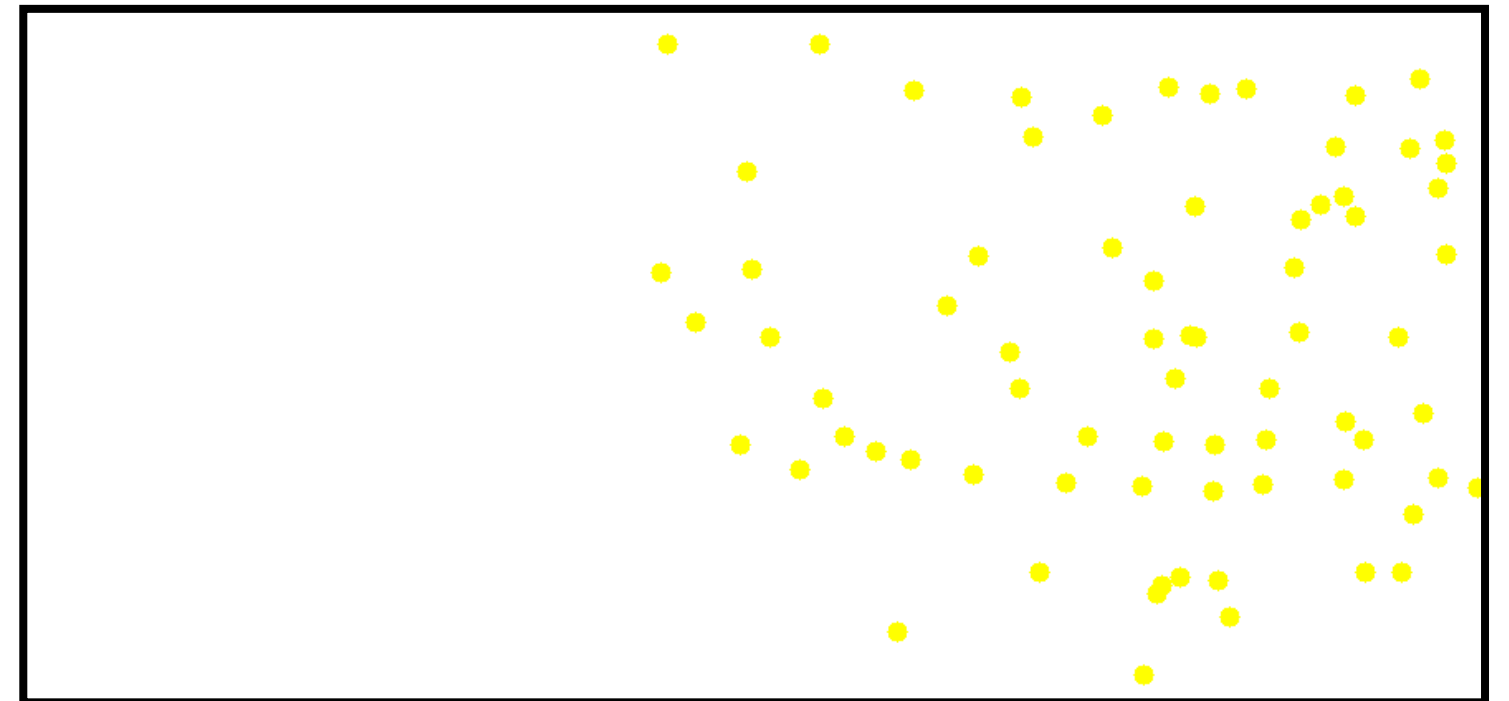
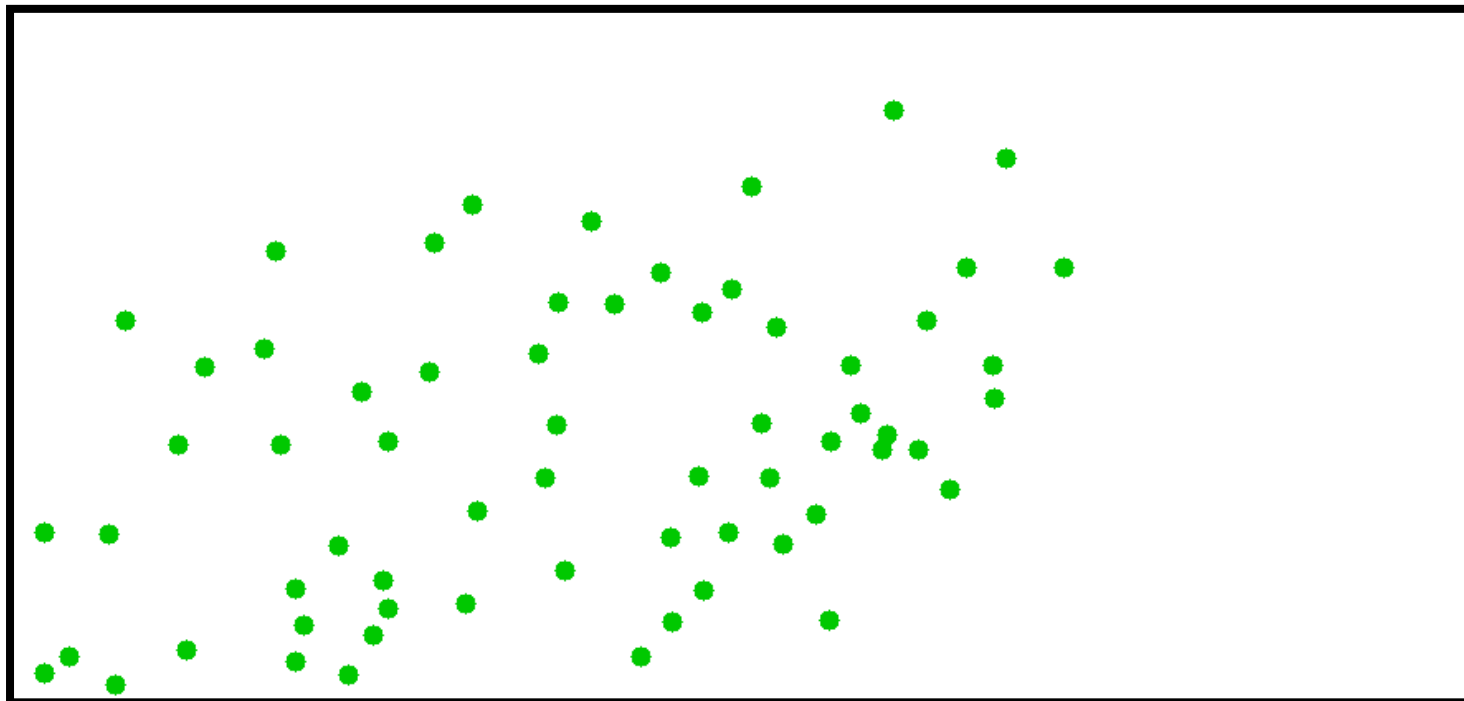
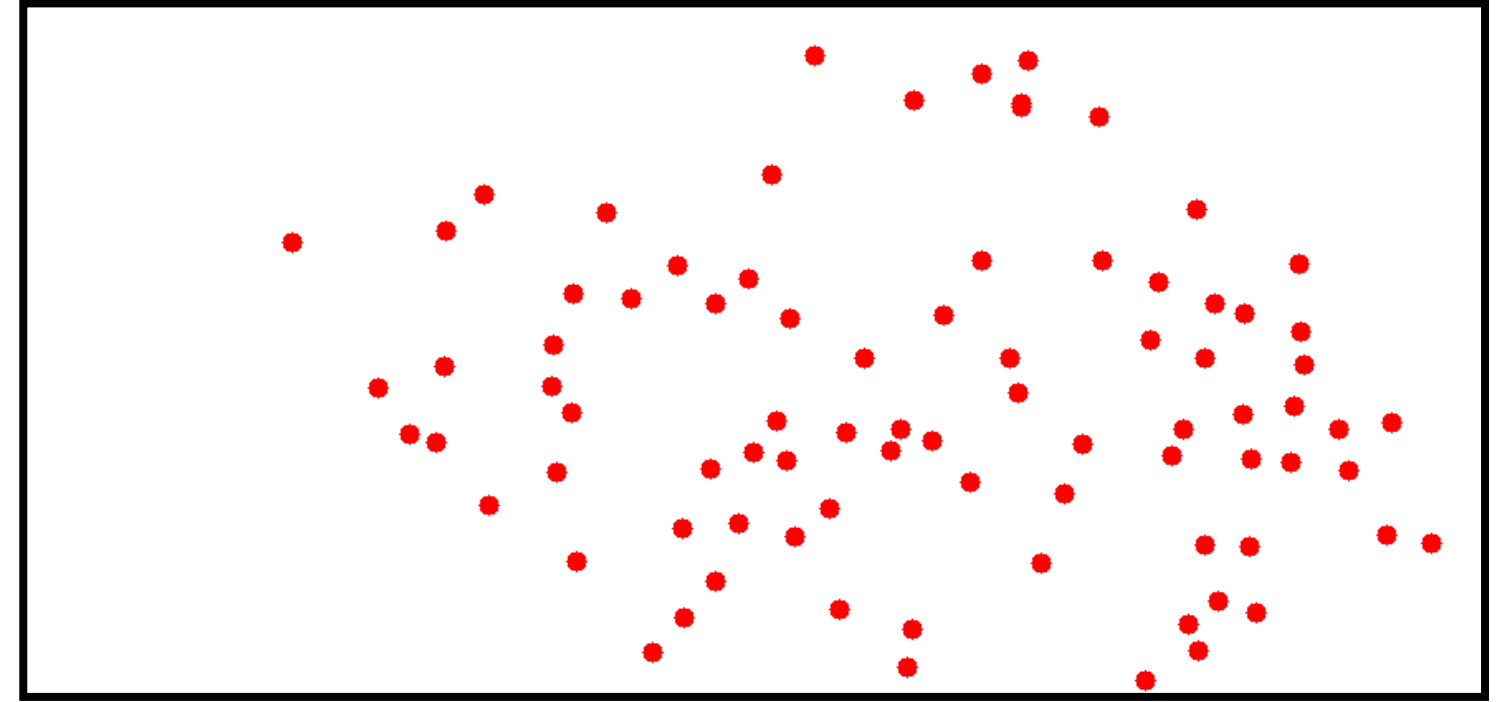
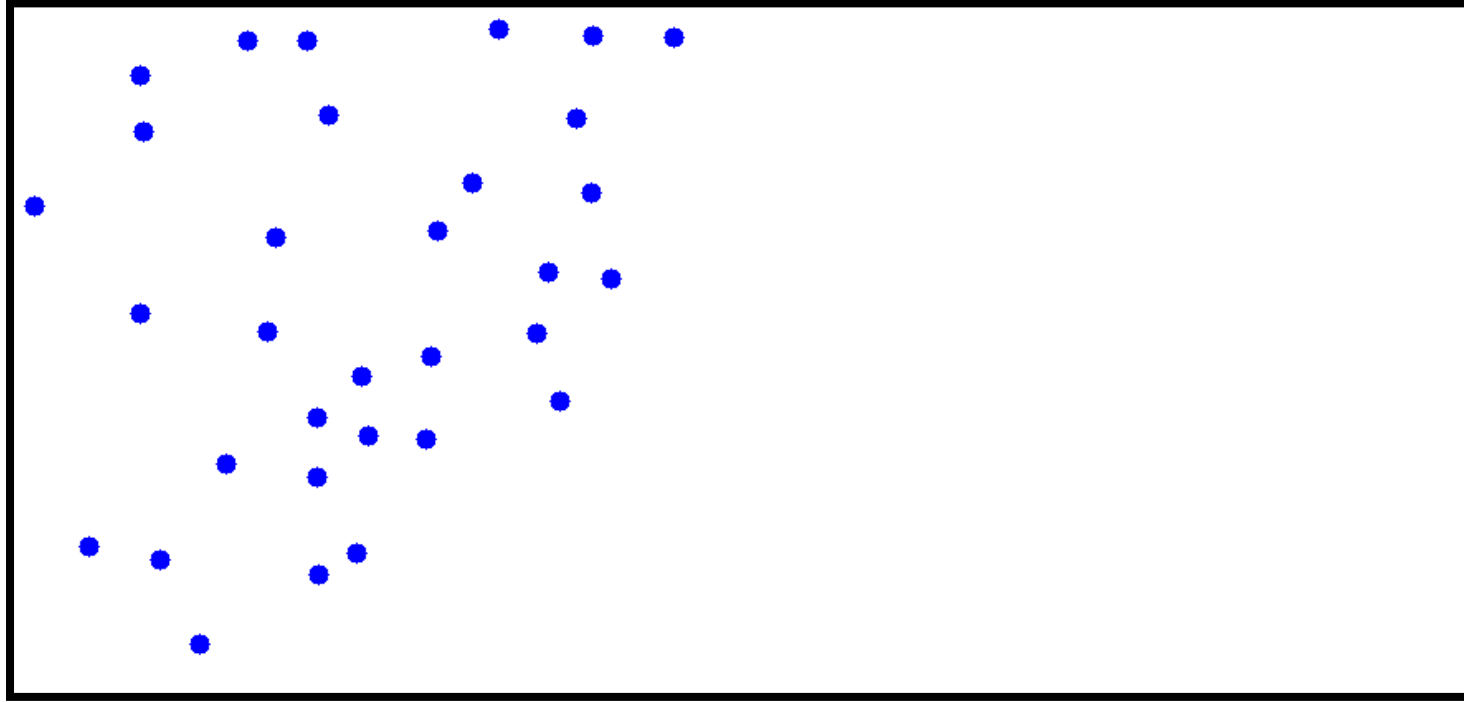


# Obtaining ground plane locations



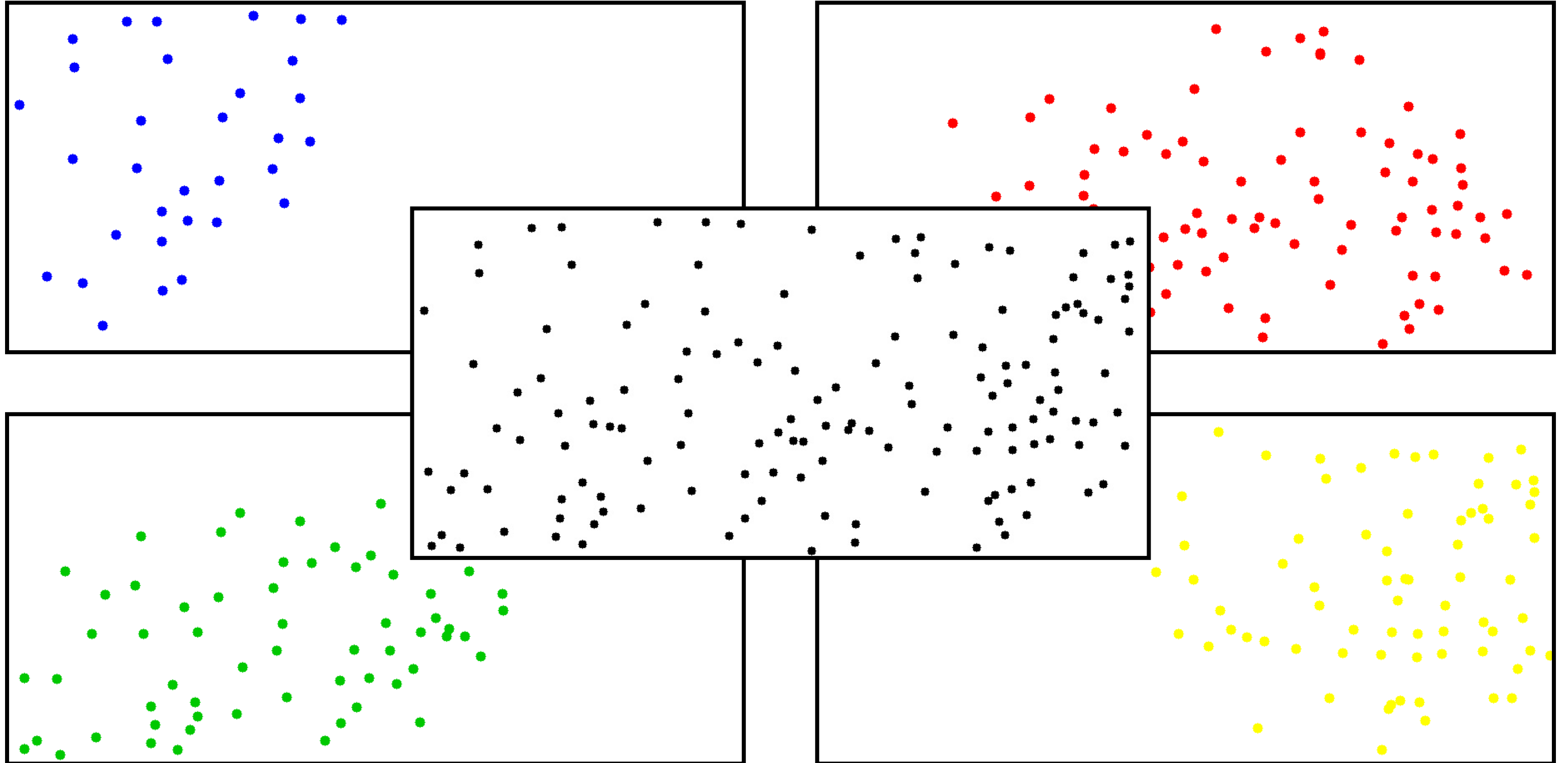


# Four sets of ground plane locations



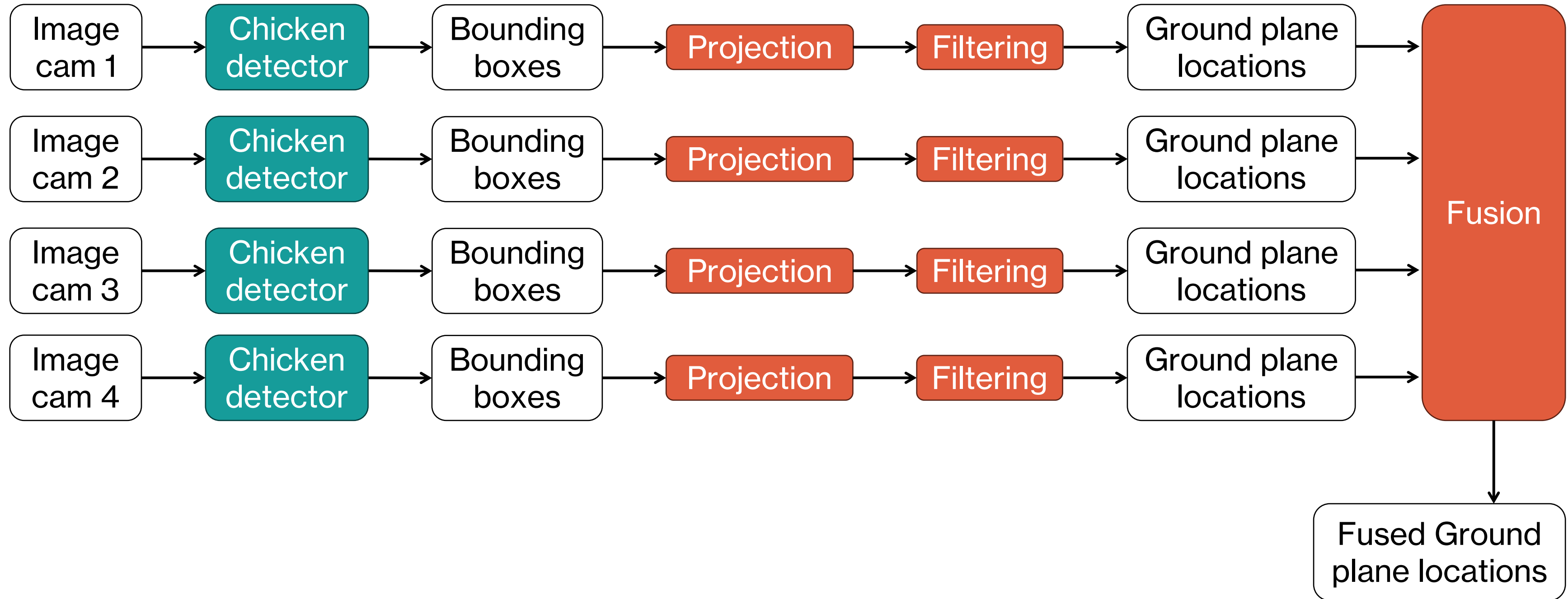


# Detections are fused on the ground plane



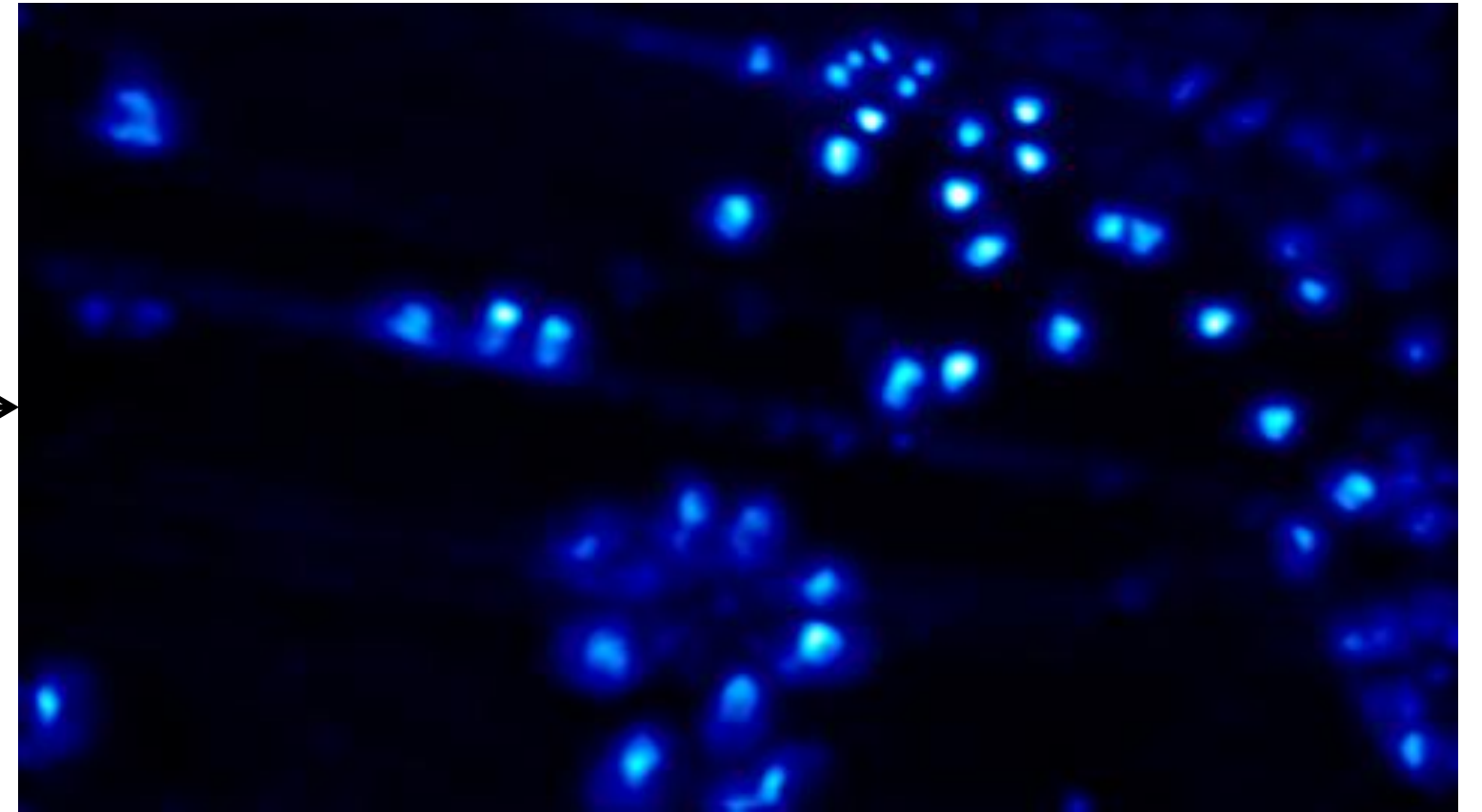
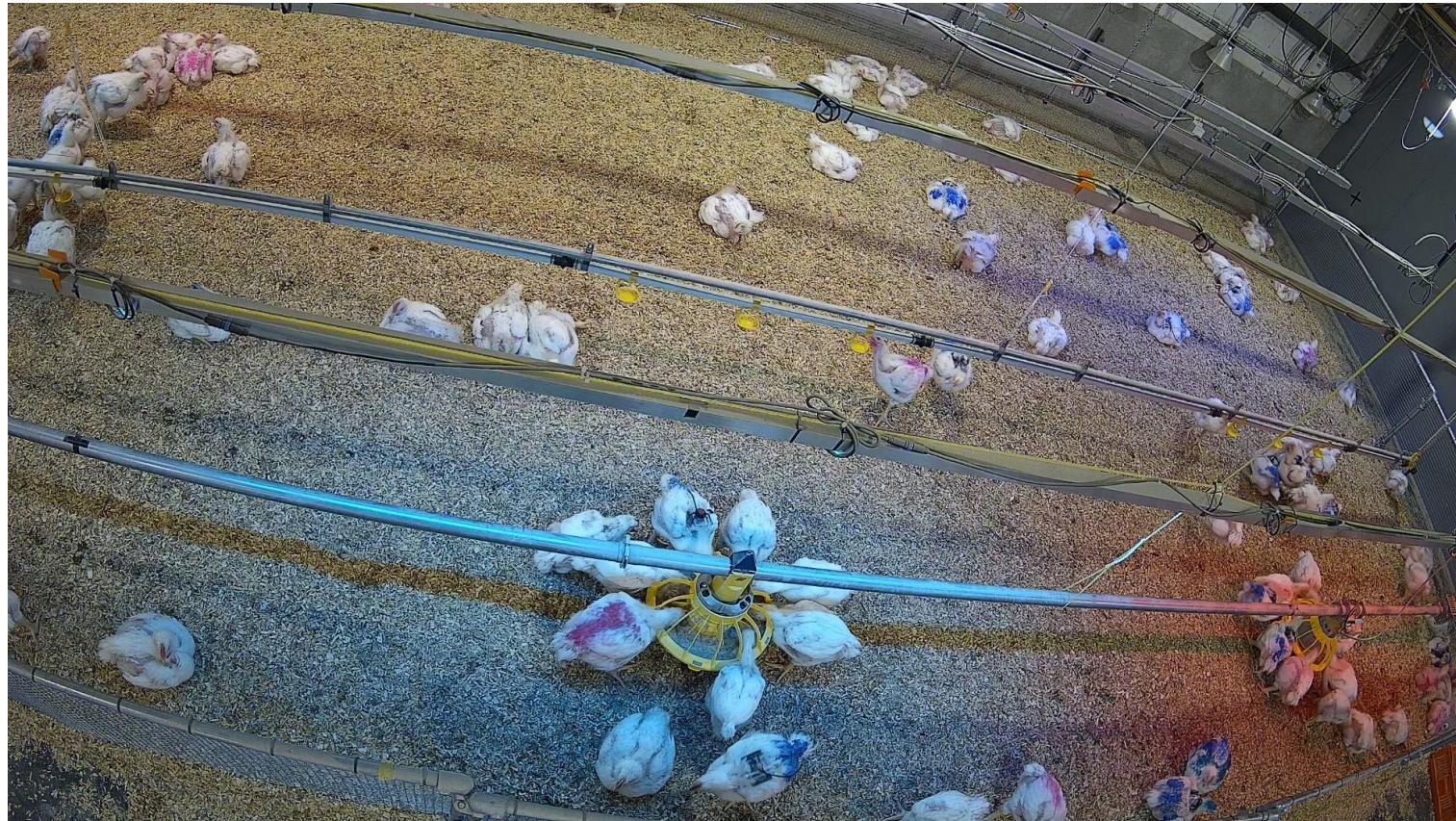
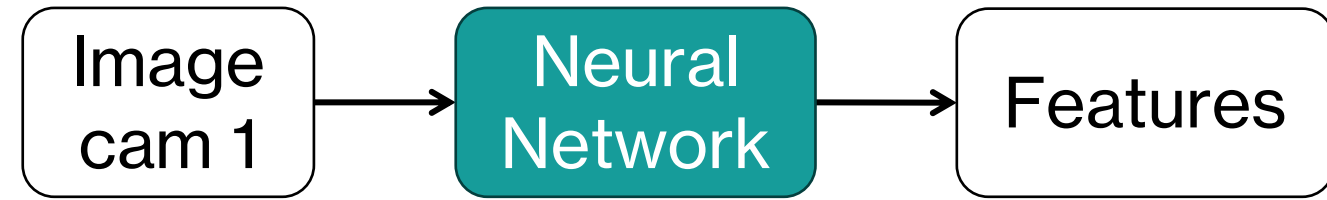


# Geometric fusion



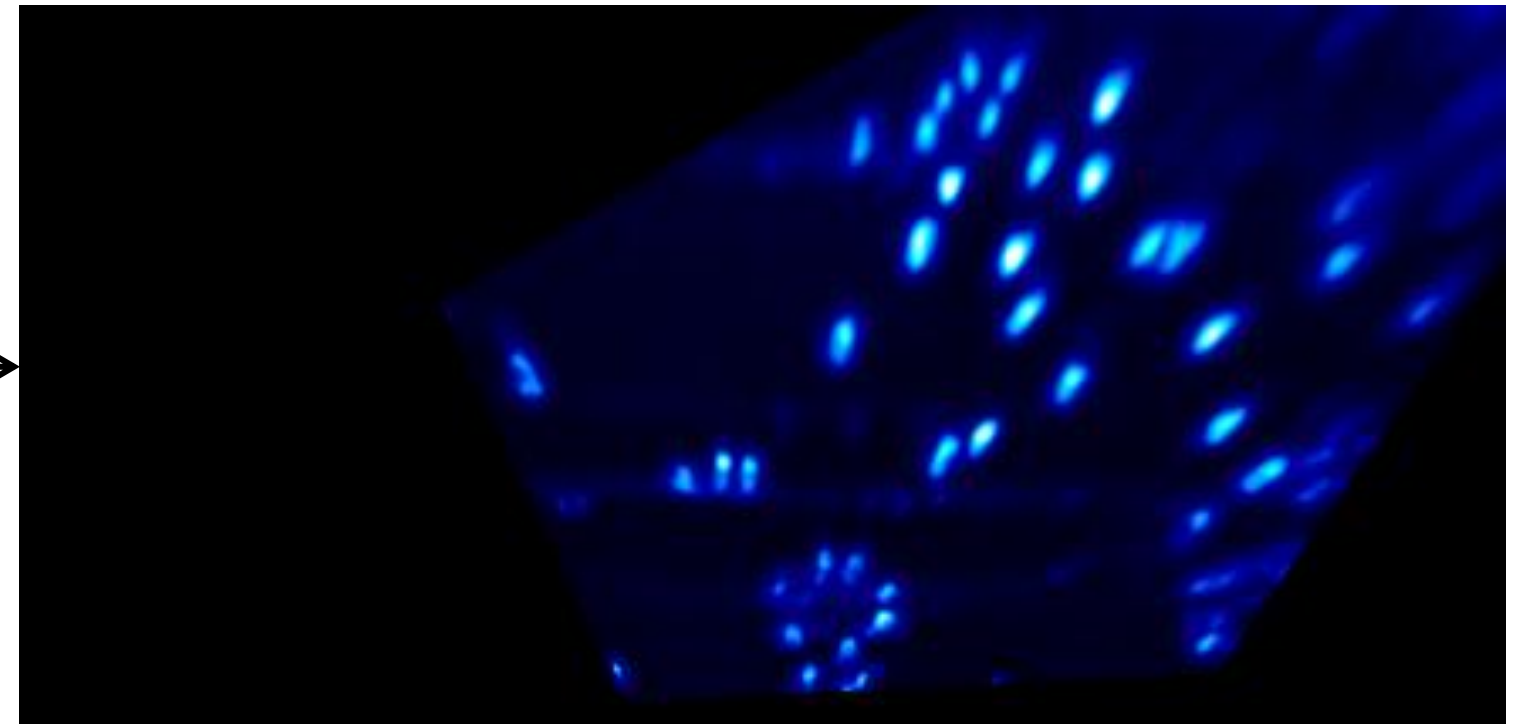
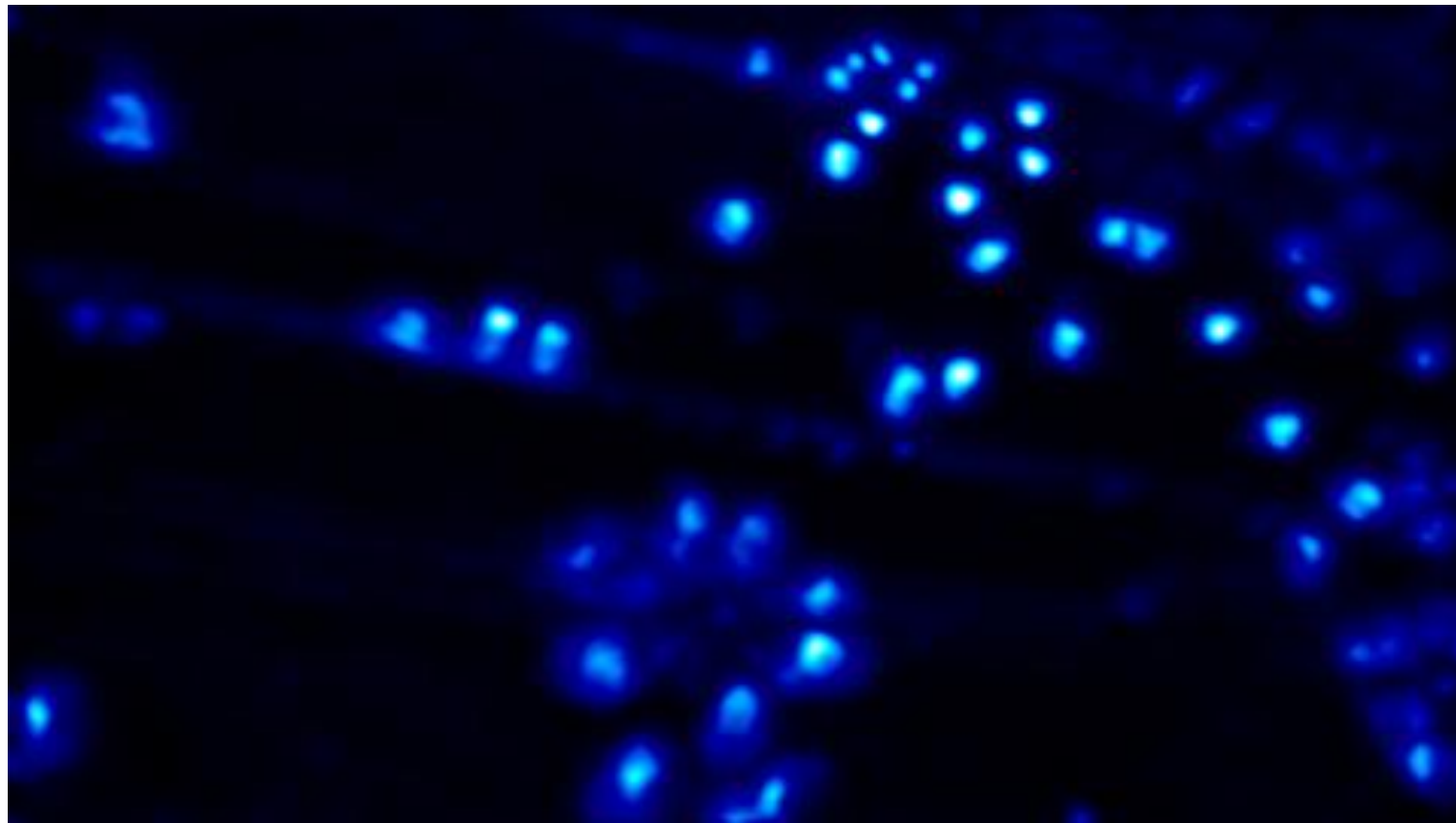
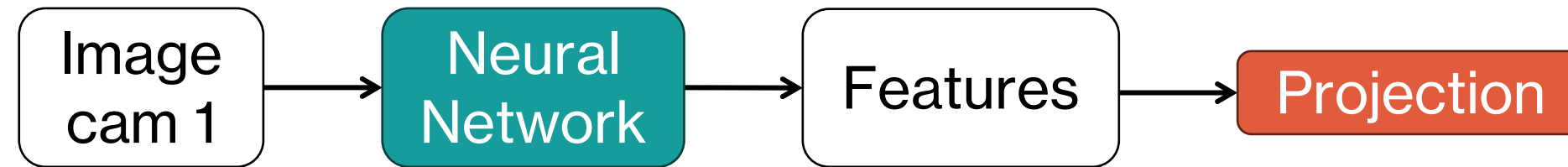


# Extracting features preserves more input information

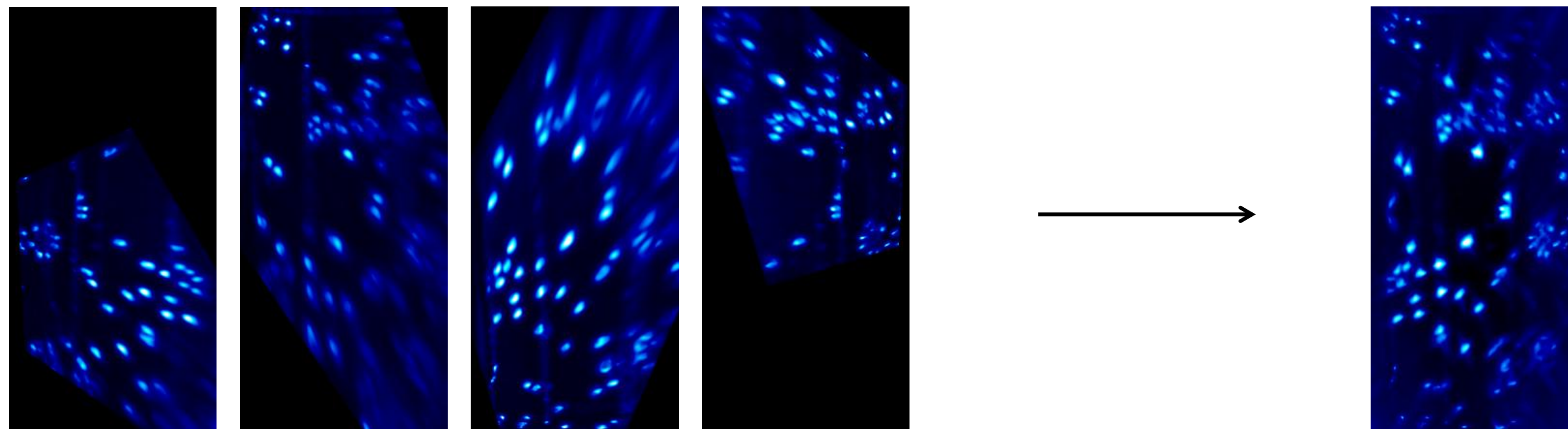
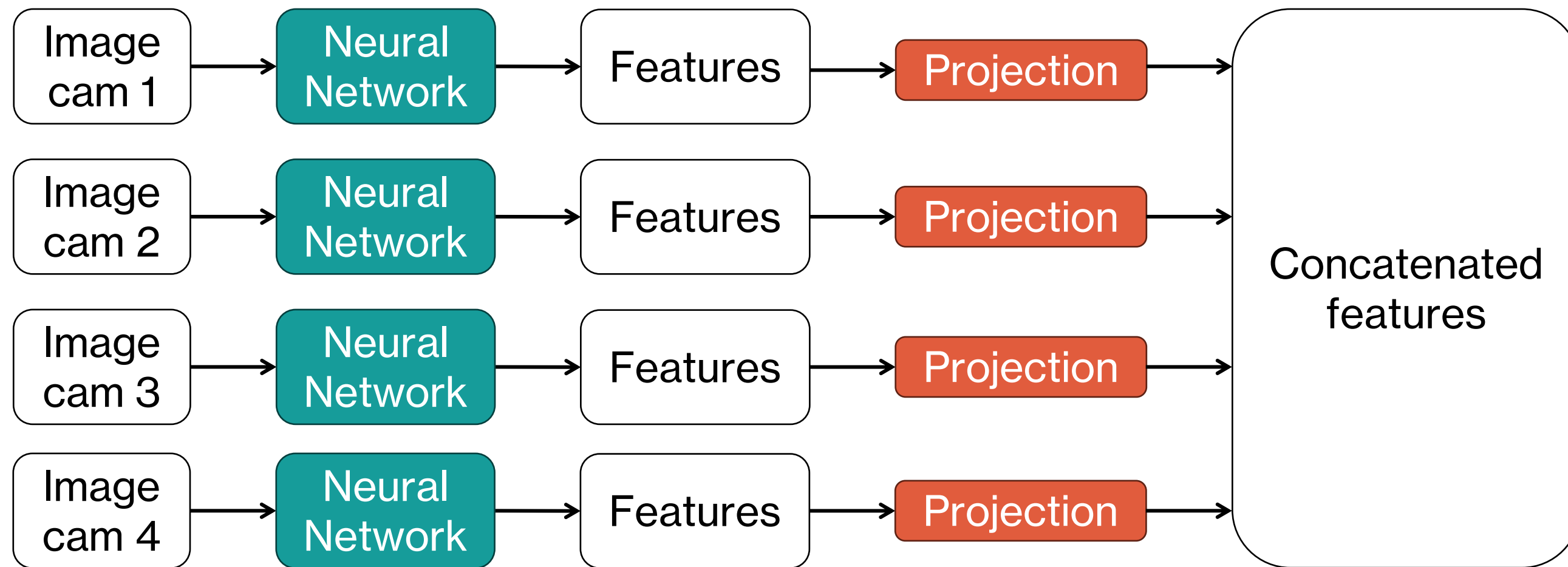




# Features can be projected to the ground plane

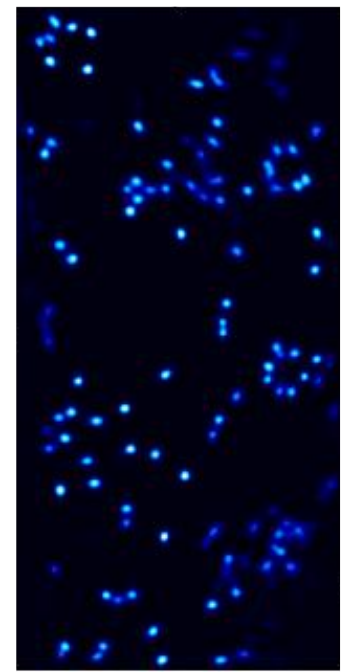
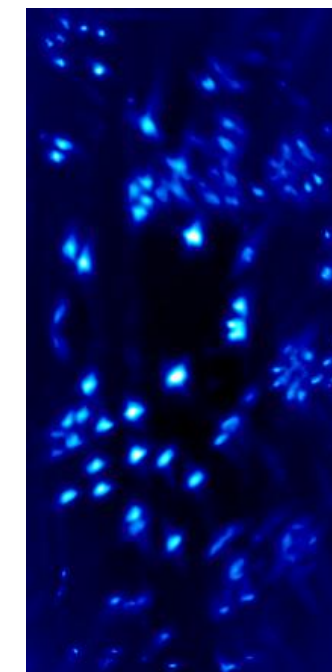
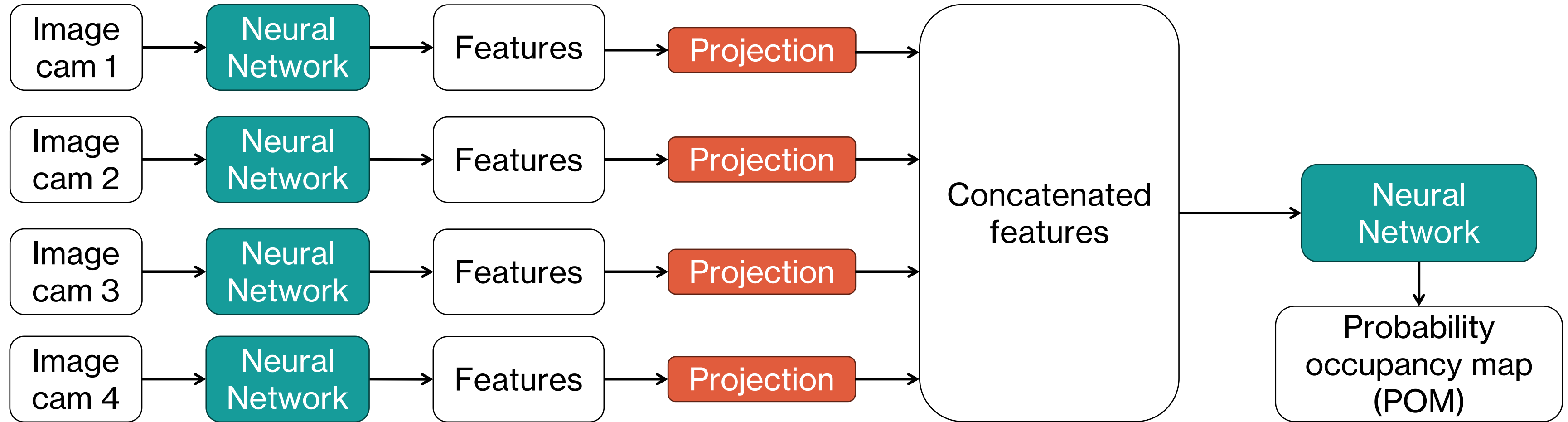






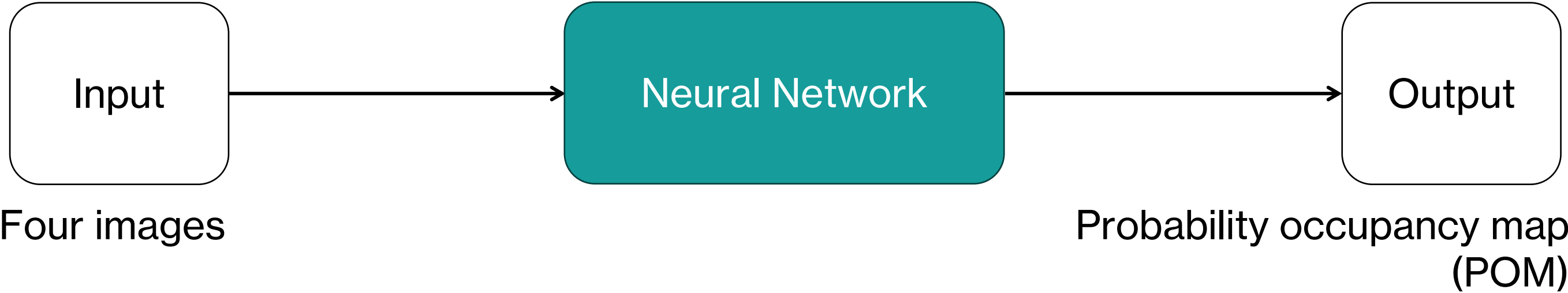


# Fusion across cameras can instead be learned





# Geometric fusion can used to obtain pseudo labels

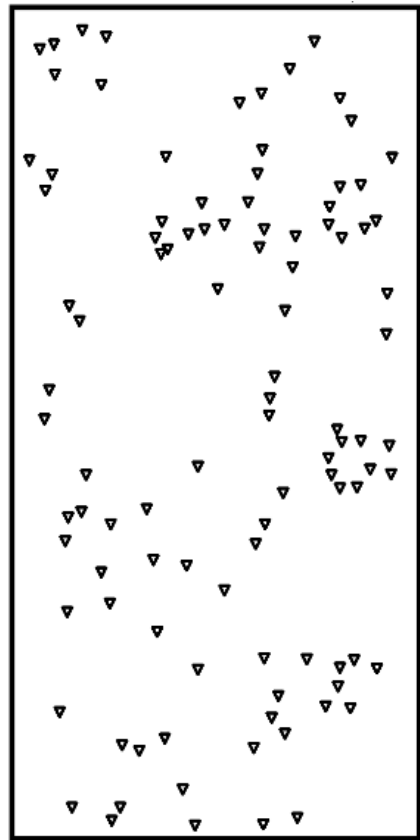


Diverse Dataset

|         | Real labels | Pseudo labels    |
|---------|-------------|------------------|
| Made by | Person      | Geometric fusion |
| Effort  | High        | Low              |
| Quality | High        | Medium           |



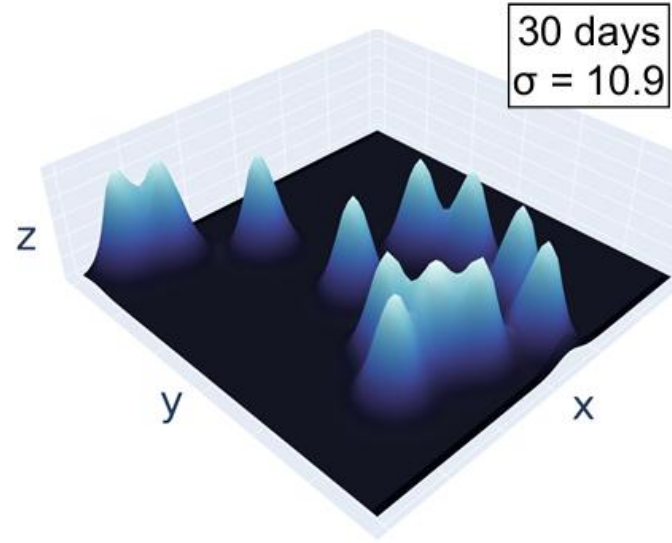
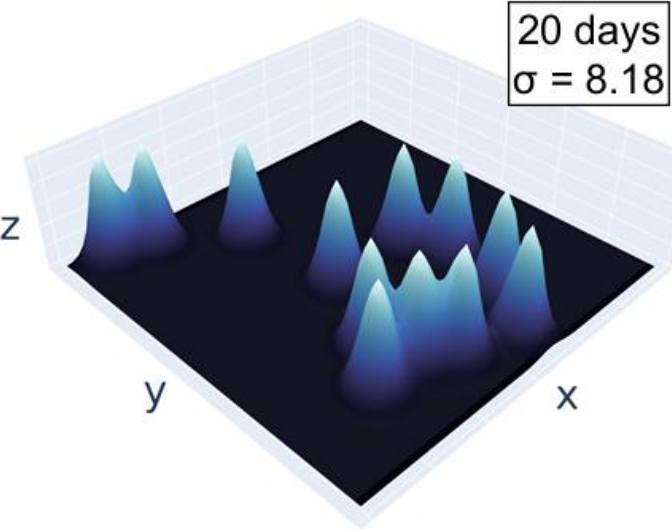
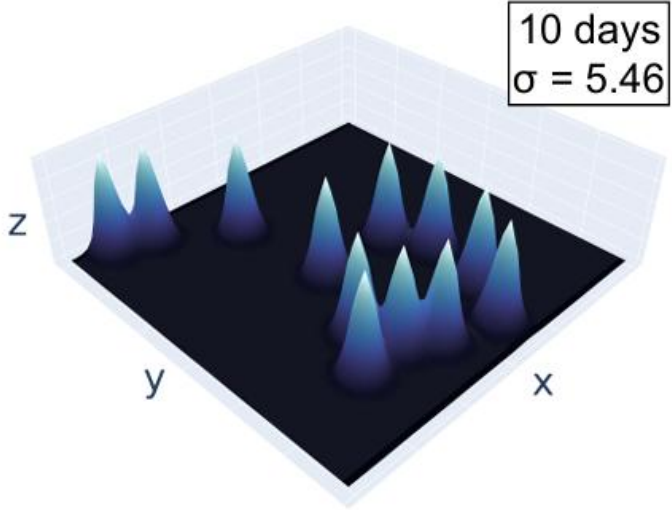
# Age-specific pseudo labels



Ground plane predictions  
from geometric fusion



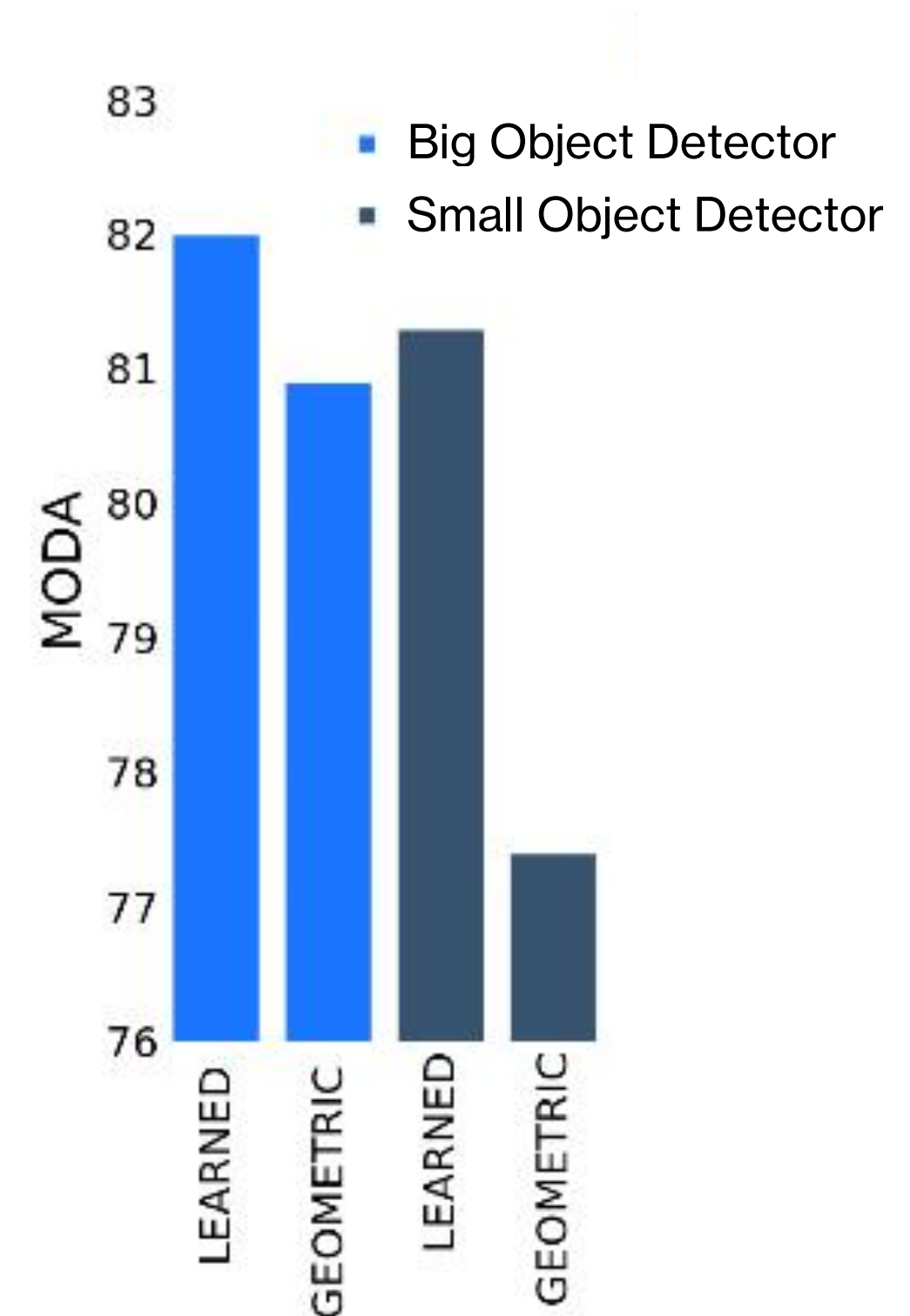
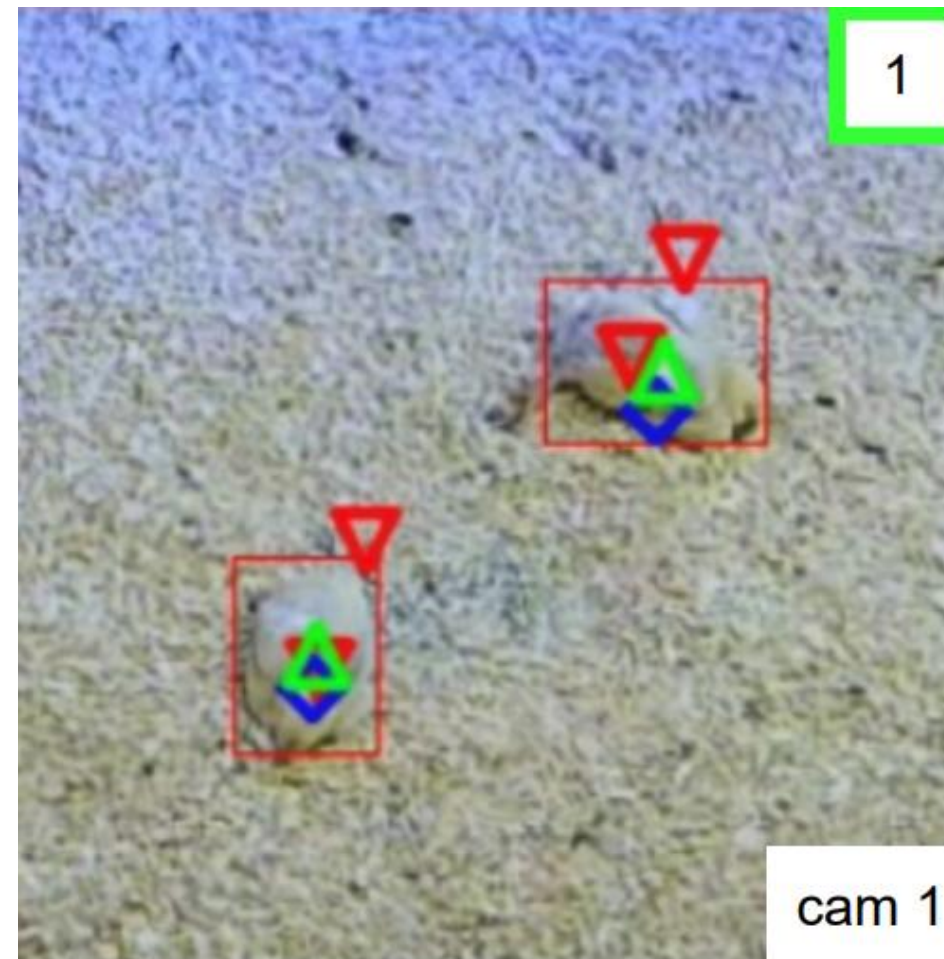
POM





# Learned fusion can improve the detection results

- ▽ Geometric fusion
- △ Learned fusion
- ◇ Ground truth

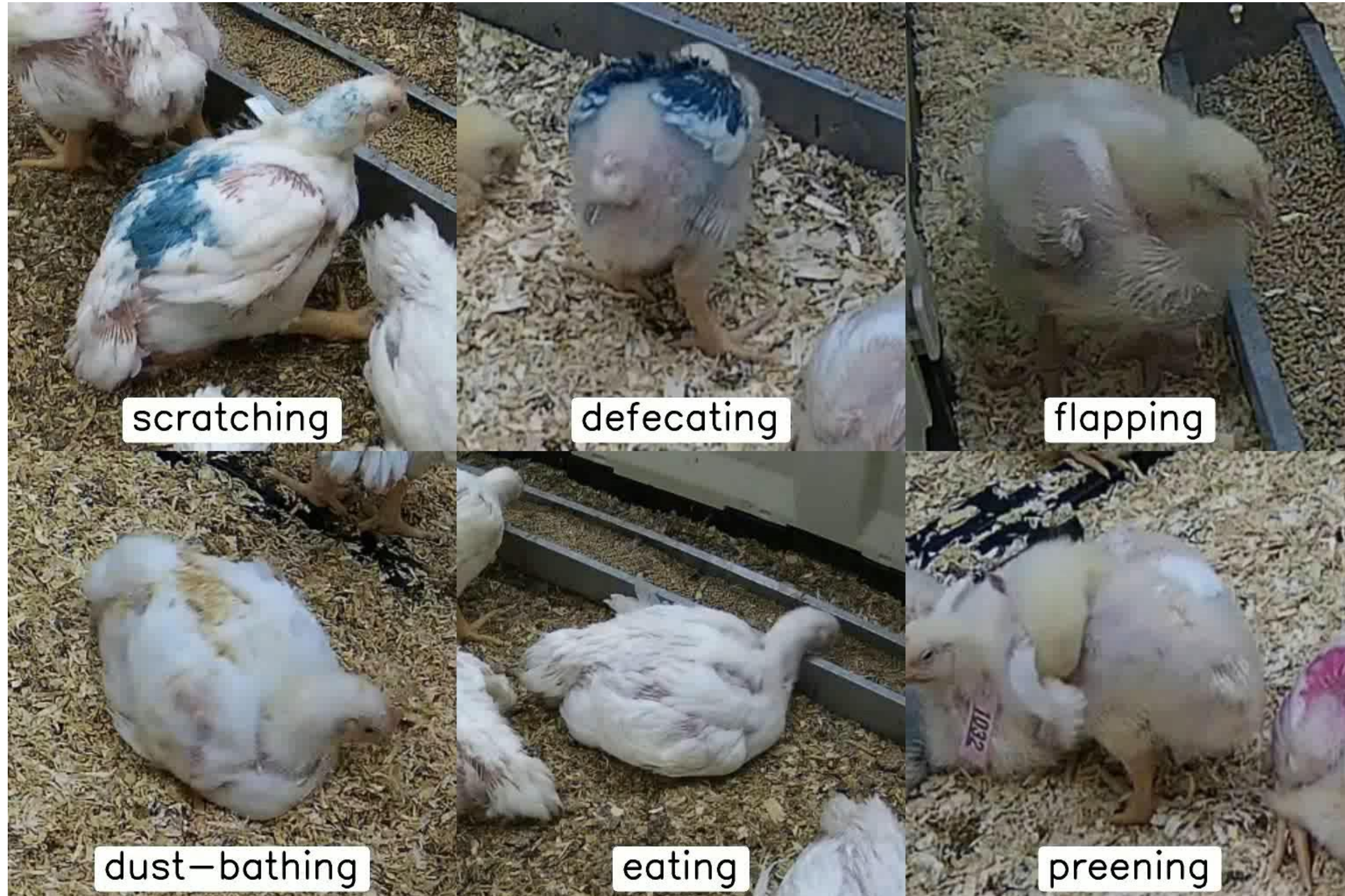


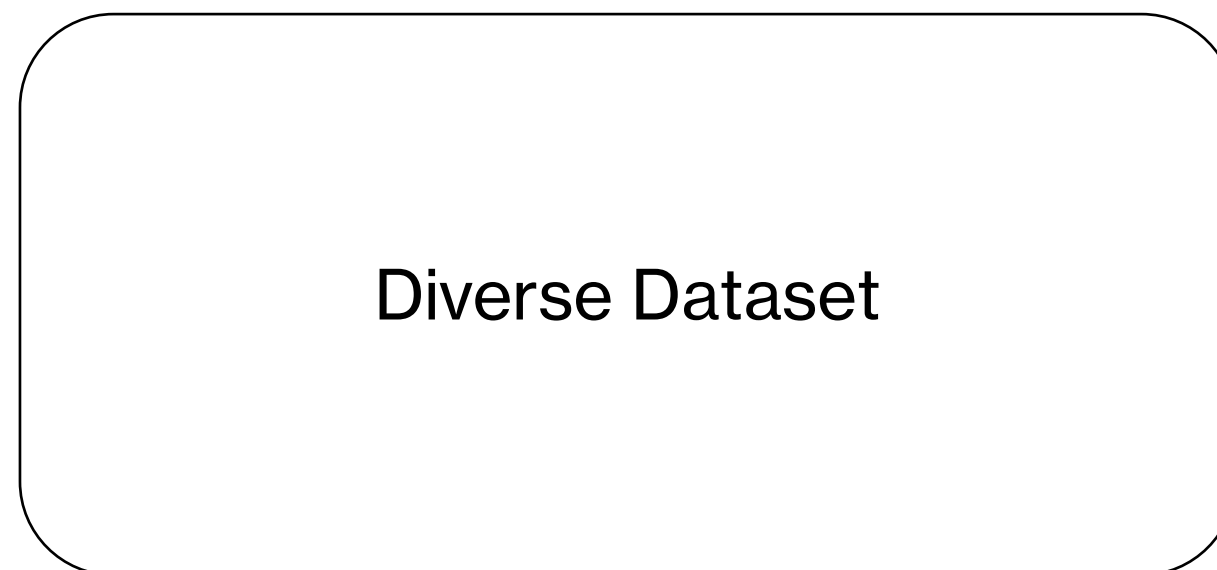
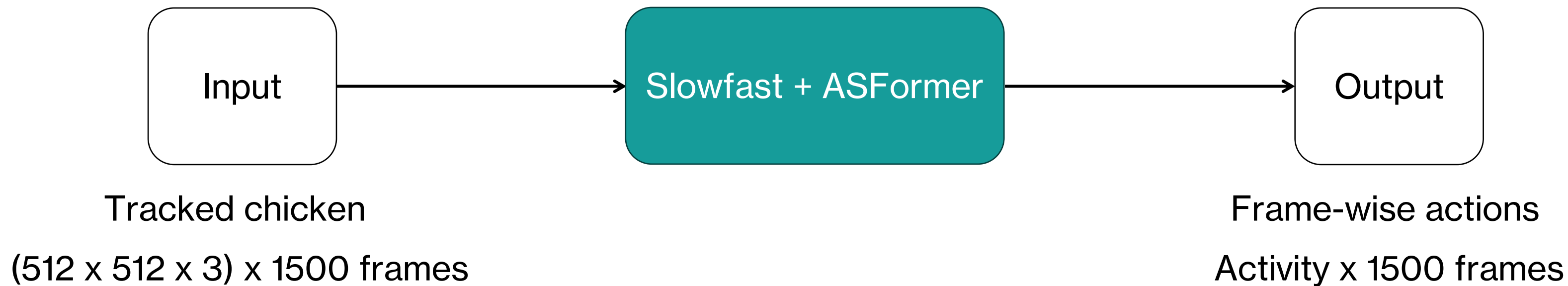


## Example 2: Action recognition



# Disease detection requires detailed actions







# A custom labelling tool is required

Id: 33



Duration: 00:33

Id: 153



Duration: 00:36

Id: 183



Duration: 00:36

Id: 213



Add Activity

NEG INTERACT

DRINK

INACTIVE

SNEEZING

FLUFFY\_FEATHERS

SLEEPING

DEFECATING

FLAPPING

HEAD\_SHAKING

PREENING

PANTING

RUNNING

FEATHER\_PECKING

VIDEO\_PROBLEM

STANDING

YAWN

EAT

EXPLOR\_FORAG

BREATHING\_DIFF

SWALLOWING

SHORT\_LEG\_MOVE

SCRATCHING

DUST\_BATHING

BODY\_SHAKING

STRETCHING

PLAY

LOCOMOTION

OUT\_OF\_SIGHT

OTHER

SITTING

LYING

DISTURBANCE

0.5X

1X

4X

1x

◀◀

◀

▶

▶▶

TOGGLE CROSSHAIR ON/OFF



◻

◀

▶

Short\_Leg\_Move

◻

◀

▶

Stretching

◻

◀

▶

Lying

◻

◀

▶

Inactive

◻

◀

▶

0:31 (777)

0:02 (50)

0:27 (699)

0:29 (737)

0:32 (819)

0:32 (819)

0:04 (124)

0:27 (699)

0:29 (737)

0:32 (819)

0:36 (901)

0:36 (901)



# We created a large action recognition dataset

- 3,560 tracks
- 884,006 frames
- 6,987,833 activities



## Big Broiler: A Temporal Action Segmentation and Tracking Dataset

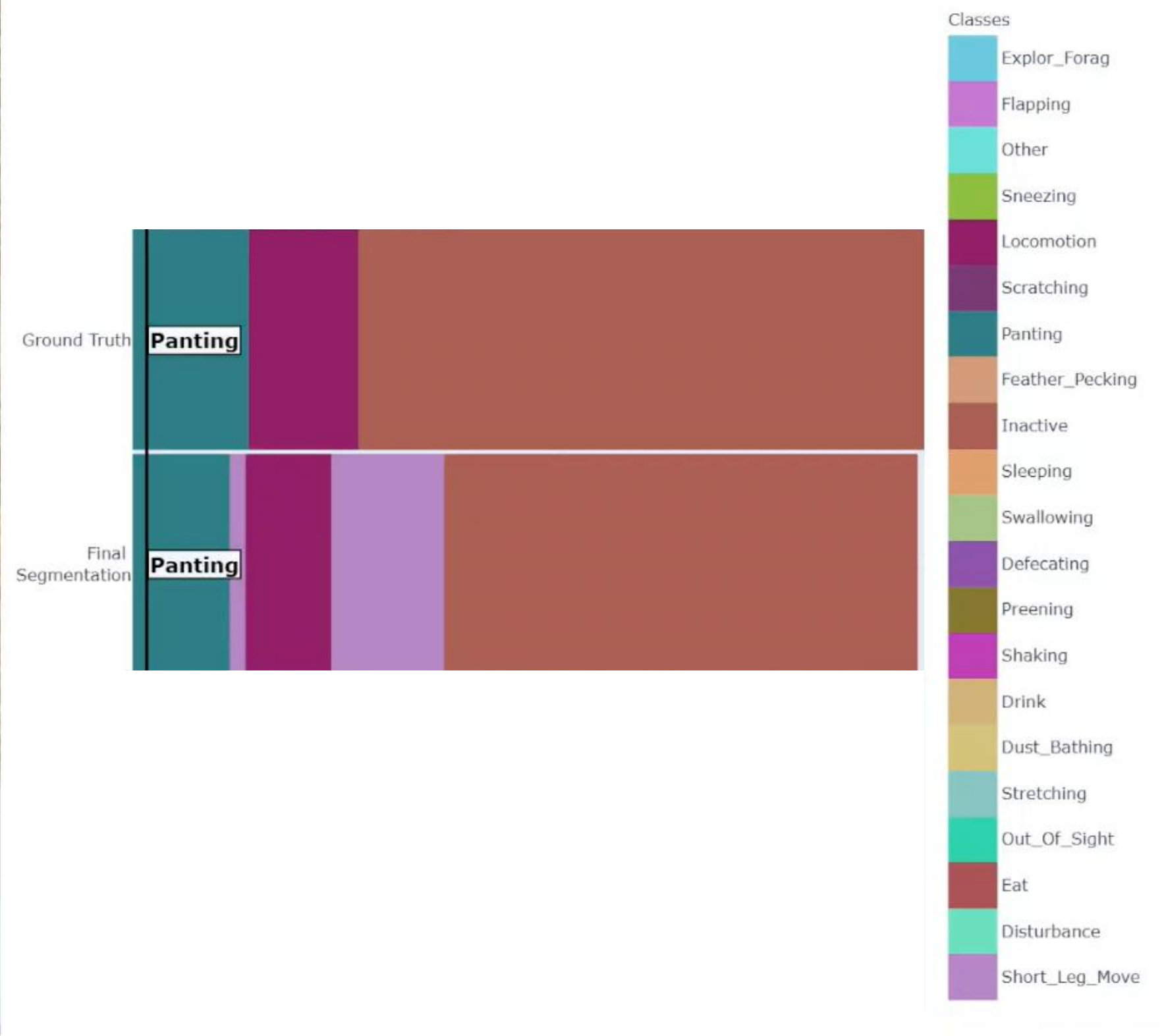
Thorsten Cardoen<sup>1</sup>      Patricia Soster de Carvalho<sup>2,3</sup>      Camilla Lopes Carvalho<sup>2</sup>  
Gunther Antonissen<sup>2</sup>      Frank A. M. Tuytens<sup>2,4</sup>      Sam Leroux<sup>1</sup>      Pieter Simoens<sup>1</sup>  
<sup>1</sup>Ghent University - imec      <sup>2</sup>Ghent University      <sup>3</sup>Poulpharm      <sup>4</sup>ILVO  
thorsten.cardoen@ugent.be



# Tracking + action segmentation can offer detailed insights









# Infection trial statistics

**176 broilers**

(4 rounds × 4 pens × 11 chickens per pen)

**3 different diseases**

Coccidiosis, Infectious Bronchitis + Ecoli and Reovirus

**~762 hours total video processed**

(~1 TB raw video data, 4572 segments of 10 minutes)

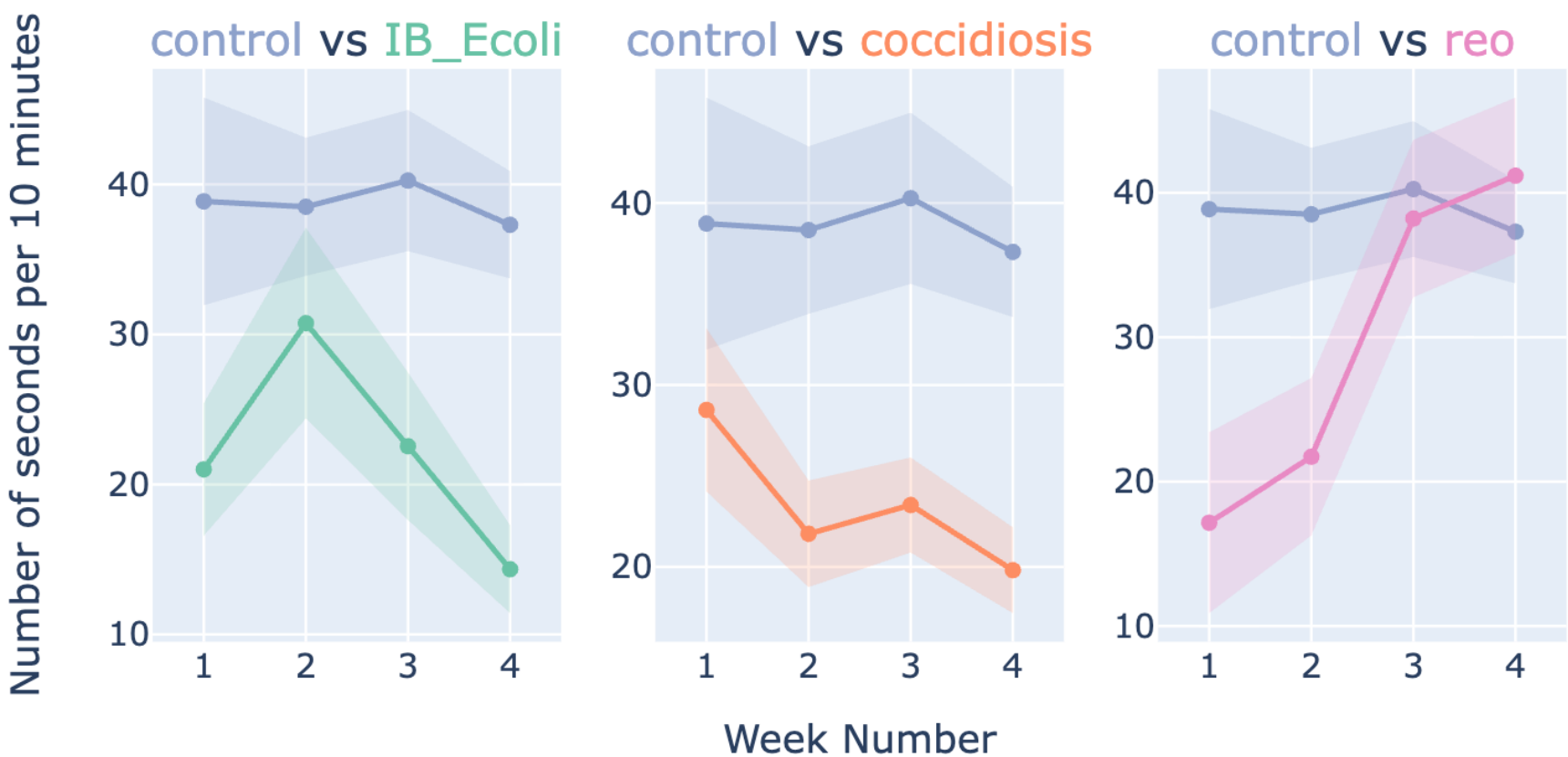
**~68.5 million frames**

(analysed at 25 fps)

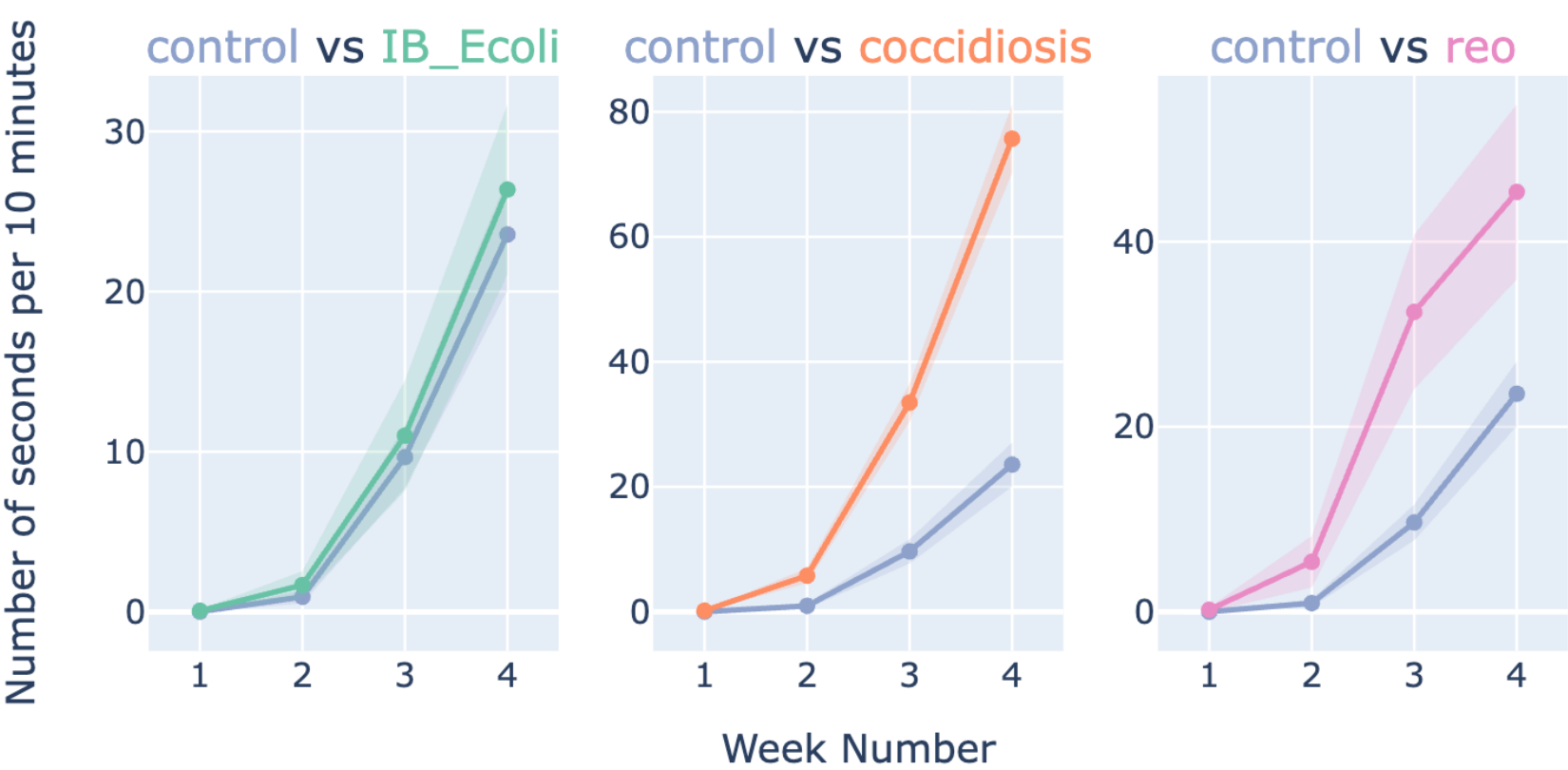


# Differences in behaviour can be identified by action recognition

Sleeping



Panting



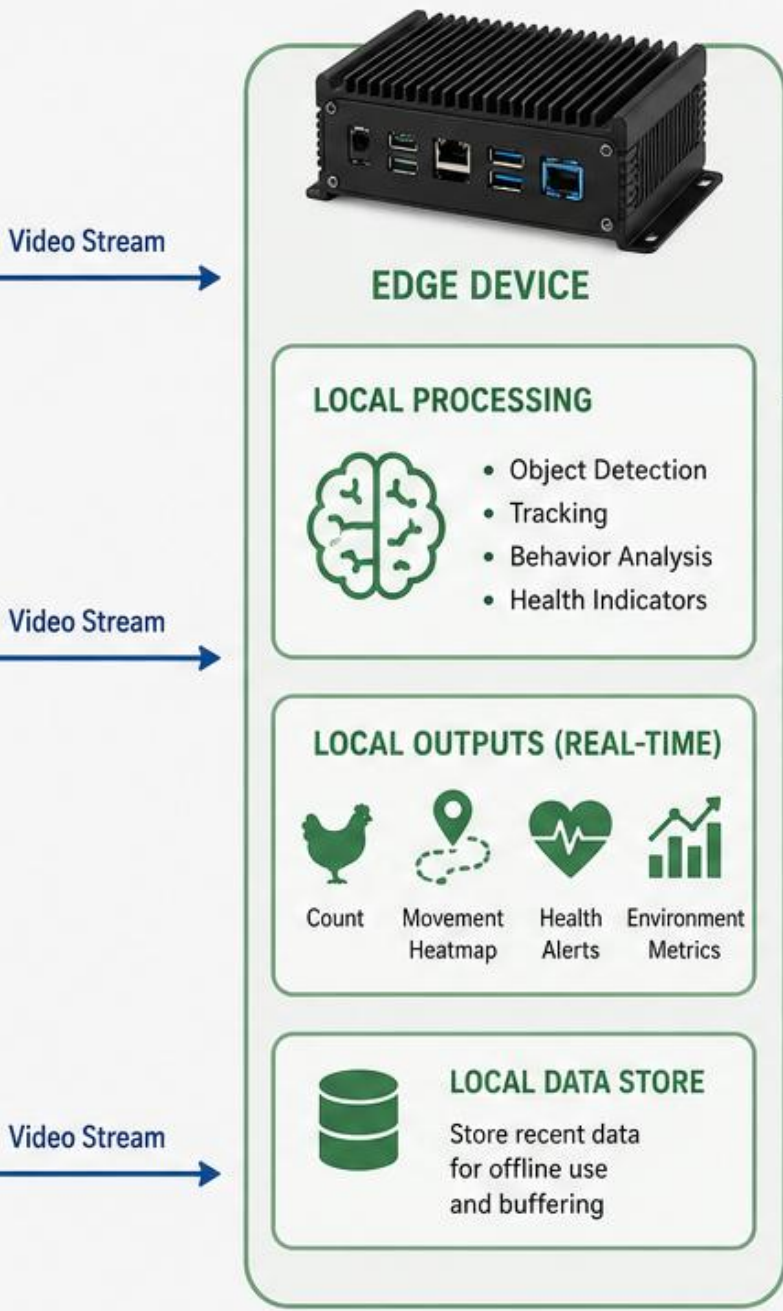


**Future directions**

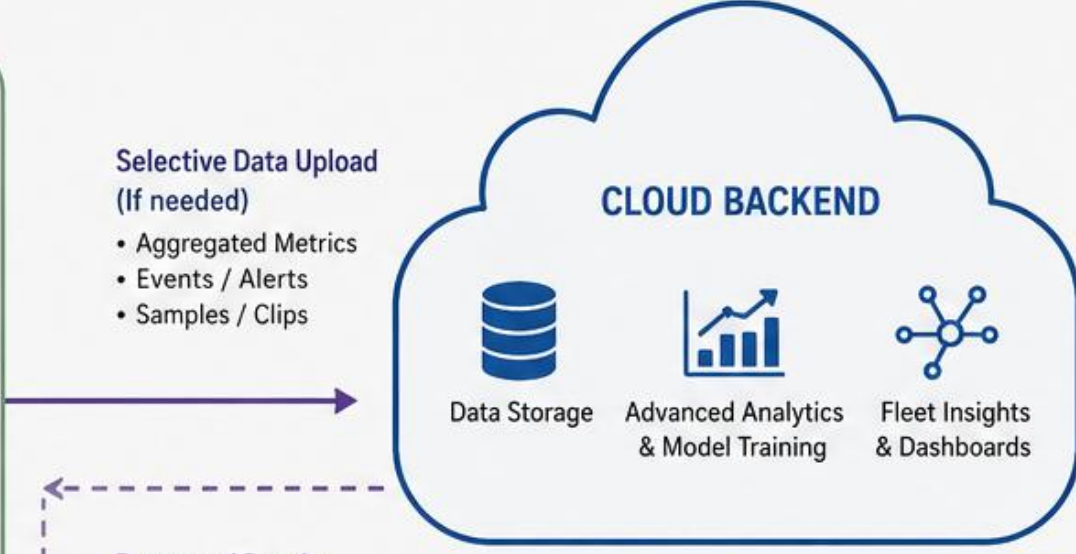
1. MULTIPLE CAMERAS CAPTURE CHICKEN DATA



2. EDGE DEVICE: LOCAL PROCESSING (ON-PREMISE)



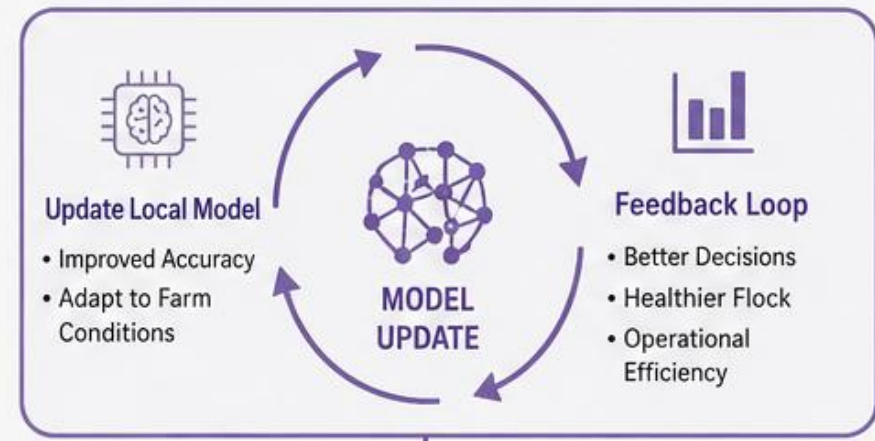
3. CLOUD BACKEND: ADVANCED PROCESSING (AS NEEDED)



Processed Results

- Deep Insights
- Updated Models
- Global Benchmarks

4. MODEL UPDATE & CONTINUOUS IMPROVEMENT





# Sam Leroux

DEPARTMENT OF INFORMATION ENGINEERING  
TECHNOLOGY

sam.leroux@gent.be

[www.ugent.be](http://www.ugent.be)

[www.imec.be](http://www.imec.be)



Universiteit Gent



@ugent



Ghent University