Object tracking

Outline

* Introduction
* Object tracking vs. Object detection
* Stages of the object tracking process
* Single- & Multi-object tracking
* Object tracking challenges
* Deep learning-based approaches

What is object tracking?

* Object tracking is a (deep learning) process where the algorithm tracks the movement of an object.

* Task of estimating object positions and other relevant information of moving objects in a video.

IMAGEN Pig project

Overview of tracking steps

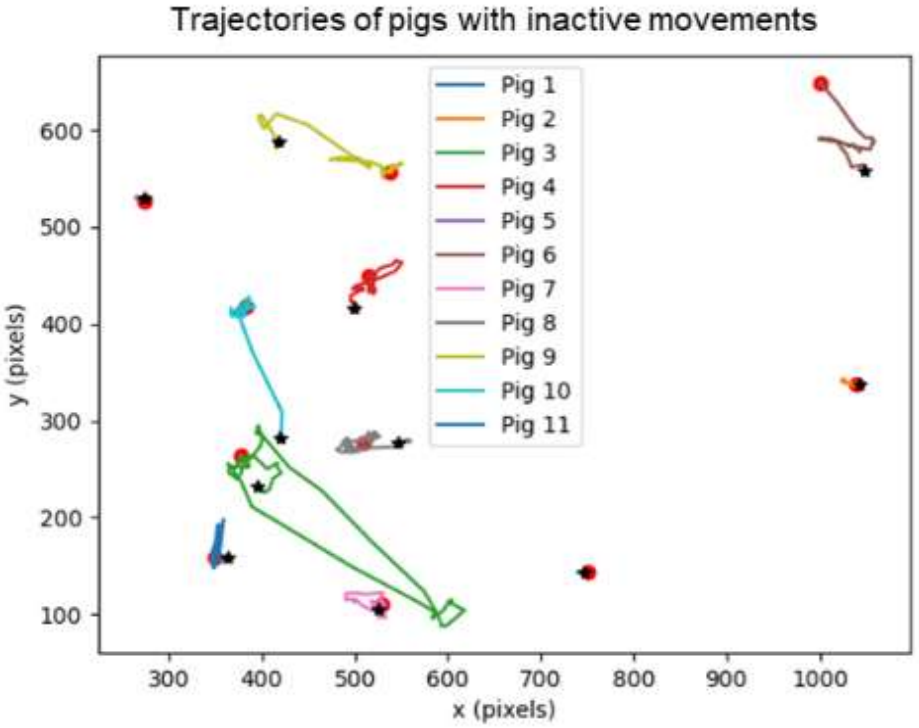
1. Object detection, where the algorithm classifies and detects the object by creating a bounding box around it.

2. Assigning unique identification for each object (ID).

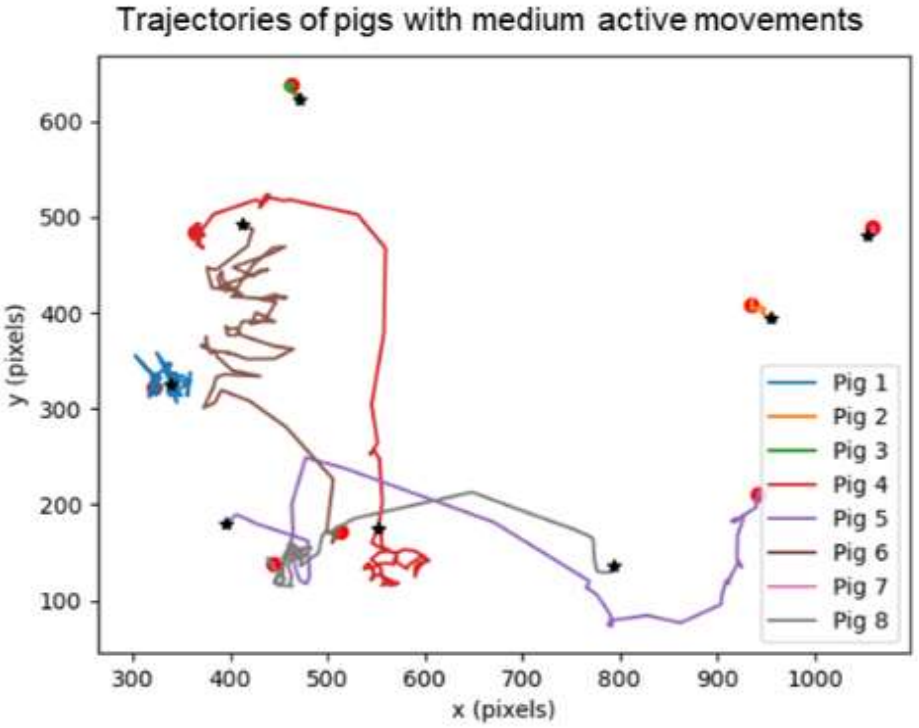
3. Tracking the detected object as it moves through frames while storing the relevant information.

Object tracking vs. Object detection

* Object tracking refers to the ability to estimate the position of a target object in each consecutive frame in a video once the initial position of the target object is defined.
* Object tracking is trained to track the trajectory of the object despite the occlusions.



(a)



(b)

Object tracking vs. Object detection

* Object detection is the process of detecting target objects in an image or a single frame of the video.
* Object can be detected only if the target image is visible on the given input.
* For a computer, to “detect objects” include two tasks - classification and localization.

Video-based tracking

* Task of tracking one/multiple moving object(s) in a video.
* To associate target objects as it appears in each video frame by analyzing the video frames sequentially and stitching the past location of the object with the present location.
* Video tracking is widely used in animal breeding, traffic monitoring, self-driving cars, and security since it can process real-time footage.

Stages of the object tracking process
Target initialization

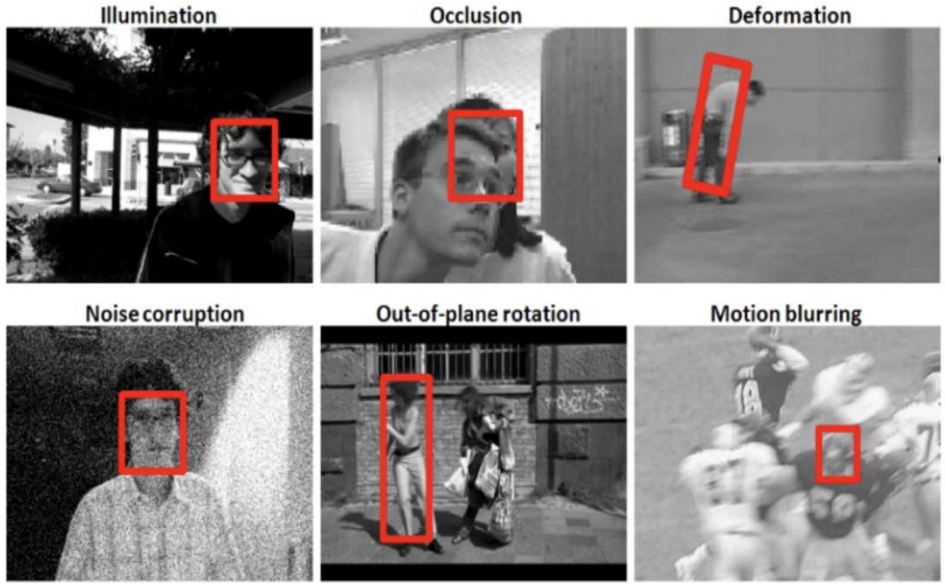
* Incorporate the process of drawing bounding boxes around objects in the initial frame of the video.
* The tracker must then estimate the objects' positions in the remaining frames while simultaneously drawing the bounding boxes.

Stages of the object tracking process
Appearance modeling

* **Deal with modeling the visual appearance of the object.**
* **Appearance modeling has to be conducted so that modeling algorithms can capture various changes and distortions.**
* **Consist of two components:**

  Visual representation &

  Statistical modeling

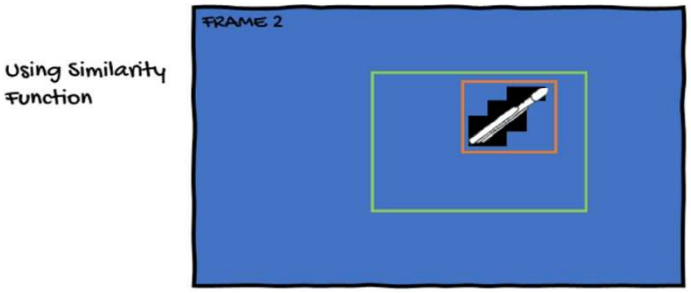


https://arxiv.org/pdf/1303.4803.pdf

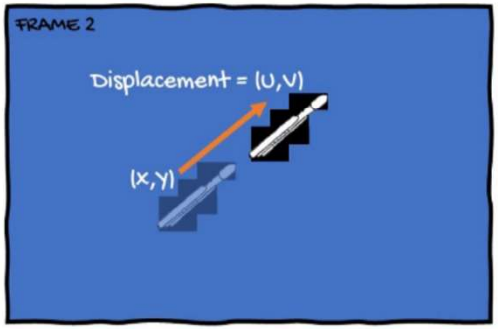
Stages of the object tracking process
Motion estimation

* **Motion estimation usually infers the predictive capability of the model to predict the object's future position accurately.**

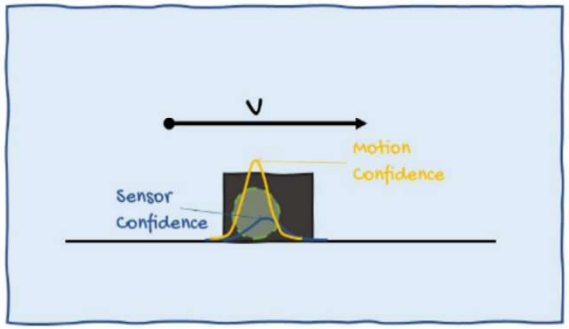
Mean Shift



Optical Flow



Kalman Filter

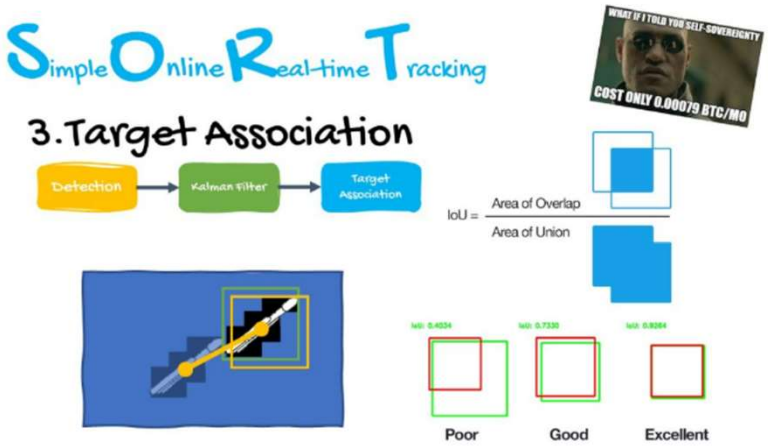


From Ritesh Kanjee

Stages of the object tracking process
Target association

* Motion estimation approximates the possible location.

* We can then use a visual model to lock down the exact location of the target.



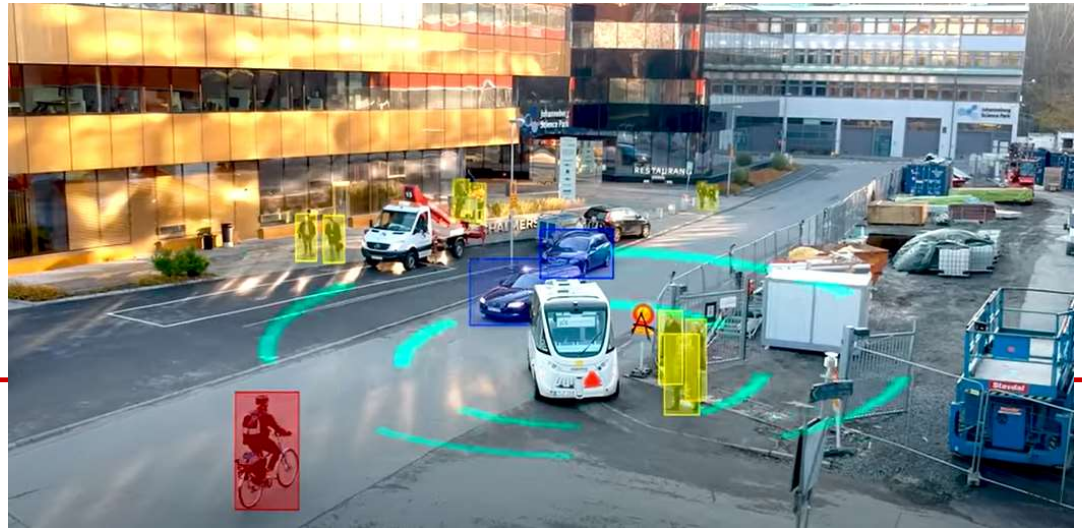
From Ritesh Kanjee

Levels of object tracking

* **Object tracking can be defined by two levels:**

  – Single Object Tracking (SOT)

  – Multiple Object Tracking (MOT): It aims to track objects of multiple classes as we see in self-driving cars.



Single Object Tracking

* Single Object Tracking (SOT) aims to track an object of a single class instead of multiple objects.

* SOT can use the category of detection-free tracking, and one has to manually provide the first bounding box to the tracker.

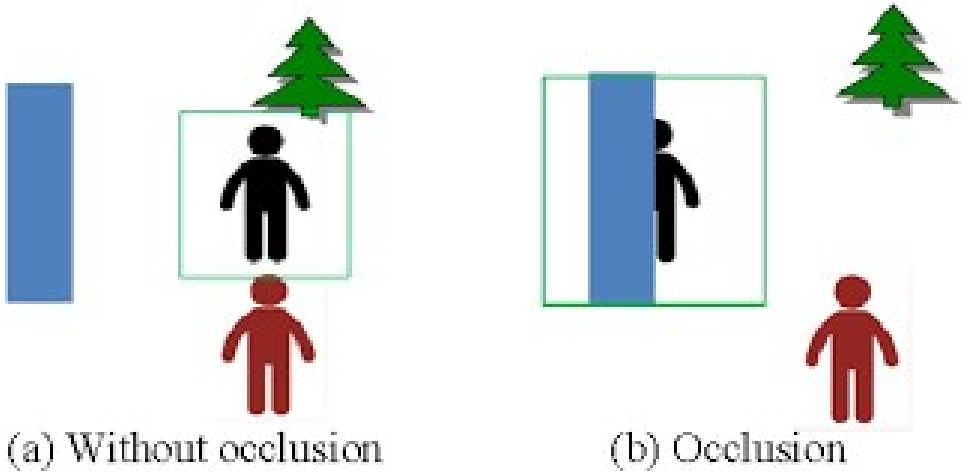
Multiple Object Tracking

* Multiple Object Tracking (MOT) refers to the approach where the tracking algorithm tracks every single object of interest in the video.

* Initially, the tracking algorithm determines the number of objects in each frame, following that it keeps track of each object's identity from one frame to the next frame until they leave the frame.

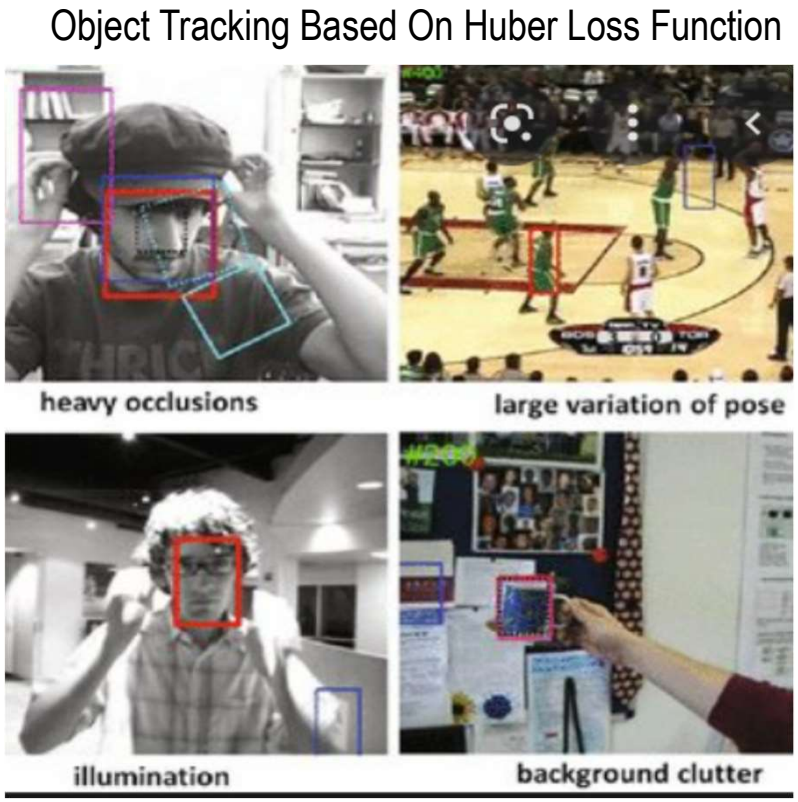
Object tracking challenges
Occlusion

* Occlusion refers to an interference phenomenon, where the object is affected by the background or foreground in which the tracking algorithm loses track of the object.



Object tracking challenges
Background clutter

* The more densely populated the background, the more difficult to extract features.
* A densely populated background introduces redundant information that makes the network less receptive to features that are important.
* To prevent background clutter, one can use a well-curated dataset that has a sparse background.



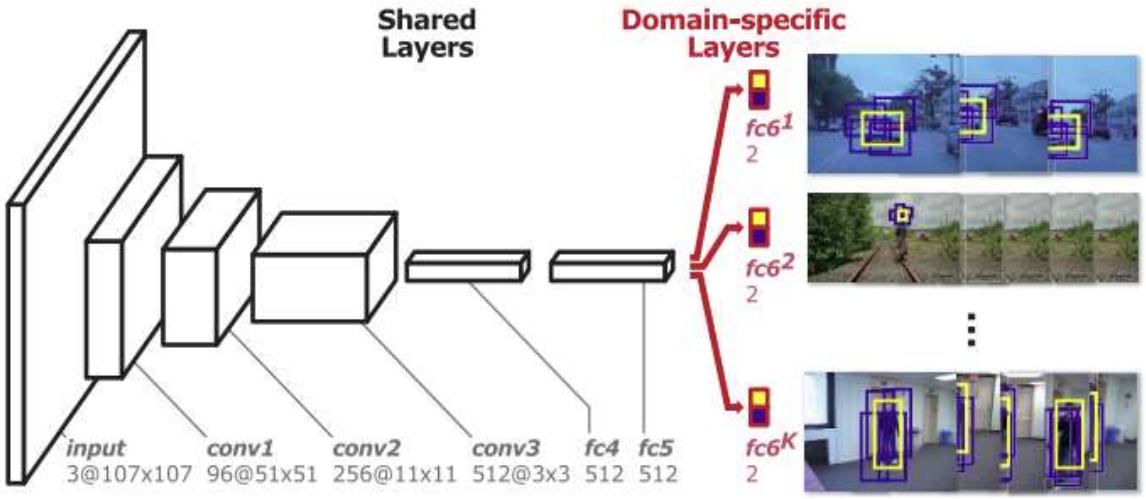
Deep Learning-based approaches to object tracking
Feature extraction

* The convolutional neural network (CNN) can efficiently extract spatial information for object detection and tracking.

* These networks, must be able to extract multi-scale spatial information. When the object is smaller in the given input image, the network may lose too much signal during. As a result, the network will not be able to detect and track smaller objects during the inference.

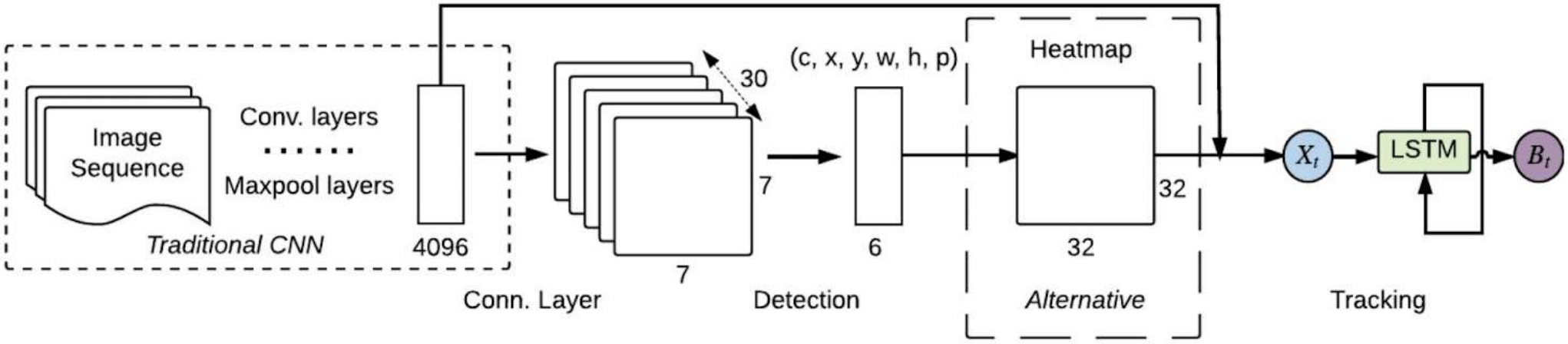
Deep Learning-based approaches to object tracking

* **Multi-Domain Net is a type of object tracking algorithm which leverages large-scale data for training. Its objective is to learn vast variations and spatial relationships.**



ROLO—Recurrent YOLO

* ROLO combines two types of neural networks: one is CNN which is used to extract spatial information while the other is an LSTM network which is used for finding the trajectory of the target object.
* At each time step, spatial information is extracted and sent to the LSTM, which then returns the location of the tracked object.

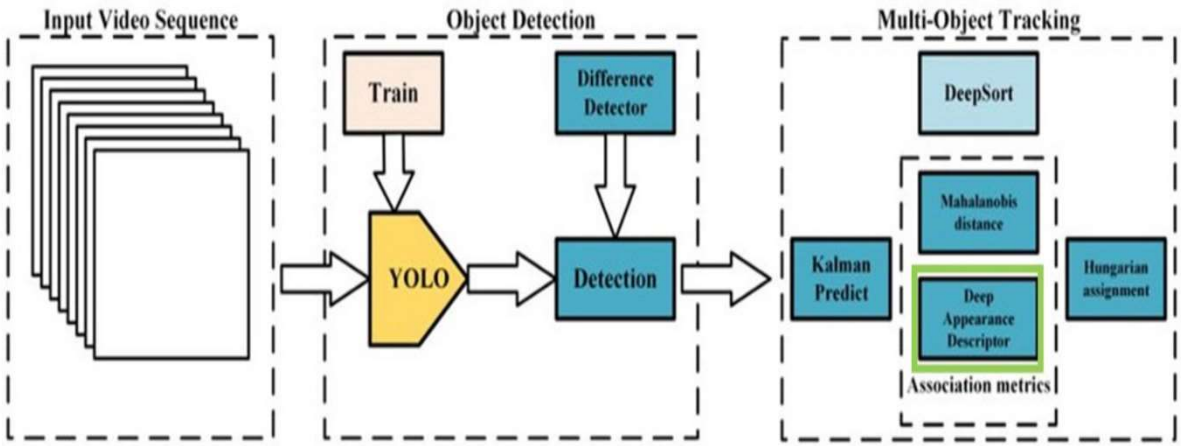


SORT

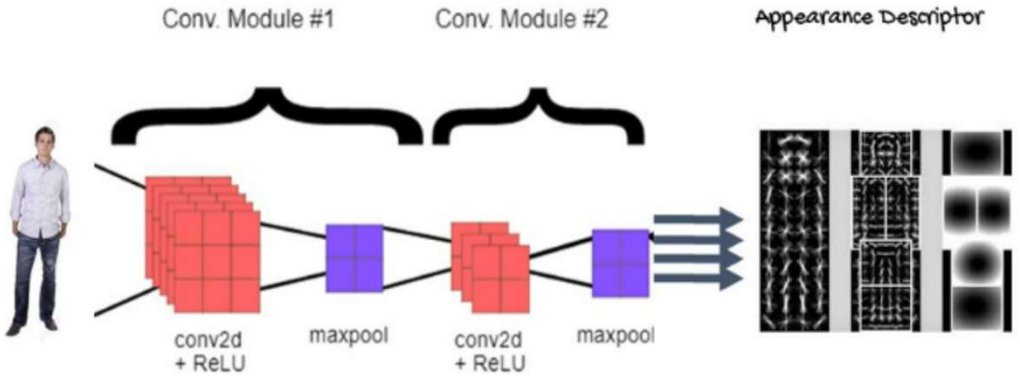
* **SORT comprises of three components:**
  - Detection: Detecting the object of interest in the initial stage i.
  - Estimation: Predicting the future location i+1 of the object from the initial stage using the Kalman filter.
  - Association: As the Kalman filter estimates the future location of the object i+1, it needs to be optimized using the correct position. This is usually done by detecting the position of the object in that position i+1. The problem is solved optimally using the Hungarian algorithm.

DeepSORT - The appearance feature vector

* Despite the effectiveness of Kalman filter, SORT returns a relatively high number of identity switches and has a deficiency in tracking through occlusions etc.



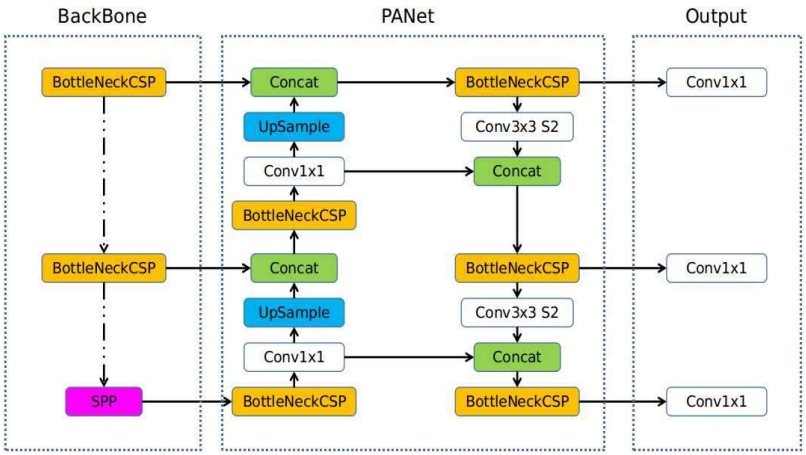
* DeepSORT introduces another distance metric based on the "appearance" of the object.



Methods: YOLOv5-DeepSORT

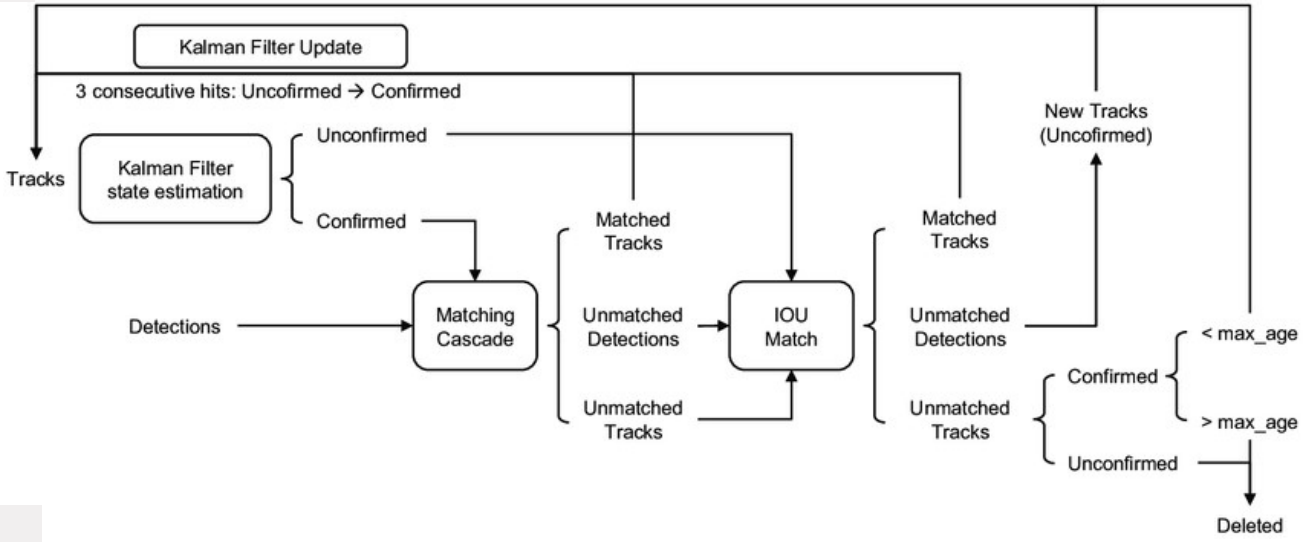
- **Network archtectures of YOLOv5 and DeepSORT**

Overview of YOLOv5



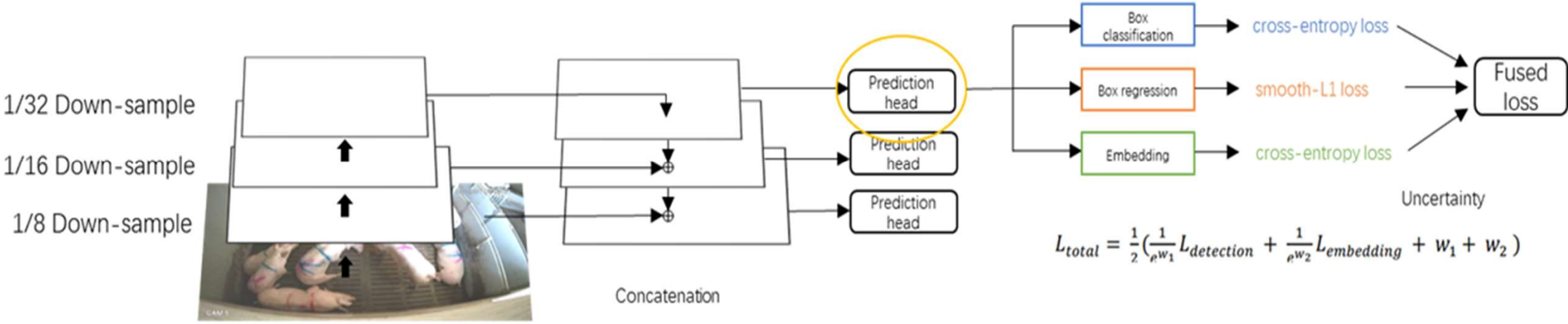
Regression-based method and fast for predicting multi-class detections

Overview of DeepSORT



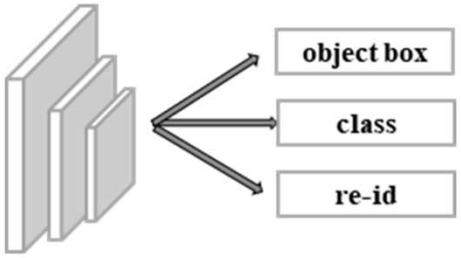
VCA
Video Coding & Architectures research group

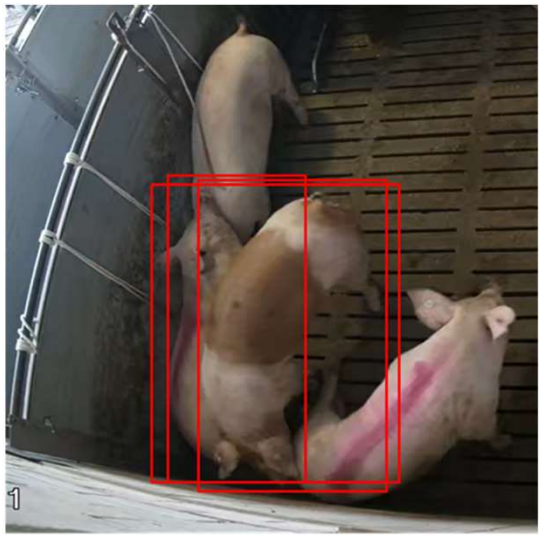
Methods: Joint Detection and Embedding (JDE) network



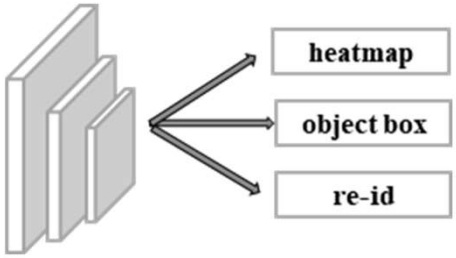
Anchor-based (JDE) vs. Anchor-free (FairMOT)

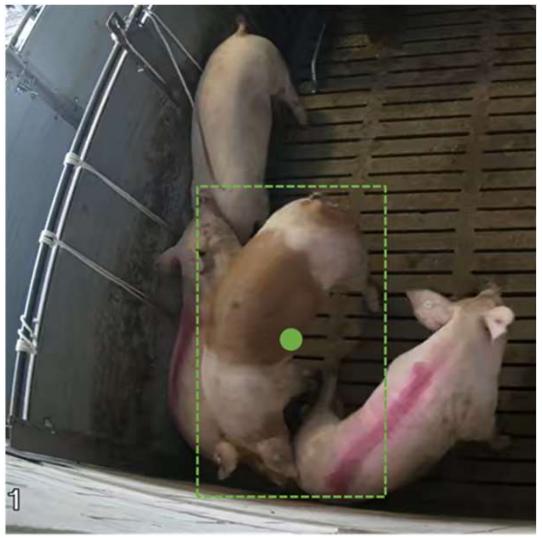
(a) Anchor-based





(b) Anchor-free

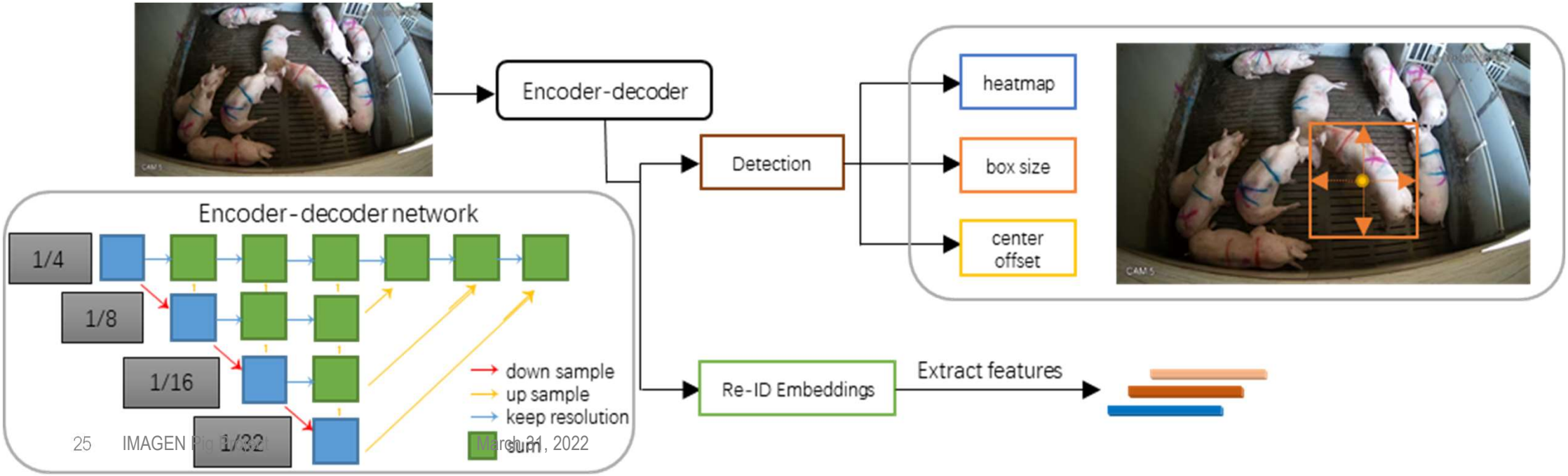




Unfair:
Duplicated detections for a single object,
multiple objects included in a single anchor box

FairMOT:
Extract features only at the object center,
avoid the unfair problems

Methods: FairMOT



The FairMOT model is proposed to only extract features at the estimated object centers. Therefore, it is so-called anchor free.

VCA
Video Coding & Architectures
research group

Conclusions

* Object tracking vs. Object detection
* Stages of the object tracking process
* Single- & Multi-object tracking
* Object tracking challenges (Occlusion, Background clutter)
* Deep learning-based approaches

Exercise

* https://pyimagesearch.com/2018/09/03/semantic-segmentation-with-opencv-and-deep-learning/

* https://www.tensorflow.org/tutorials/images/segmentation