Computer Vision: Part 1 (2D CV)

Dr. Nathanael L. Baisa

Lecture Content

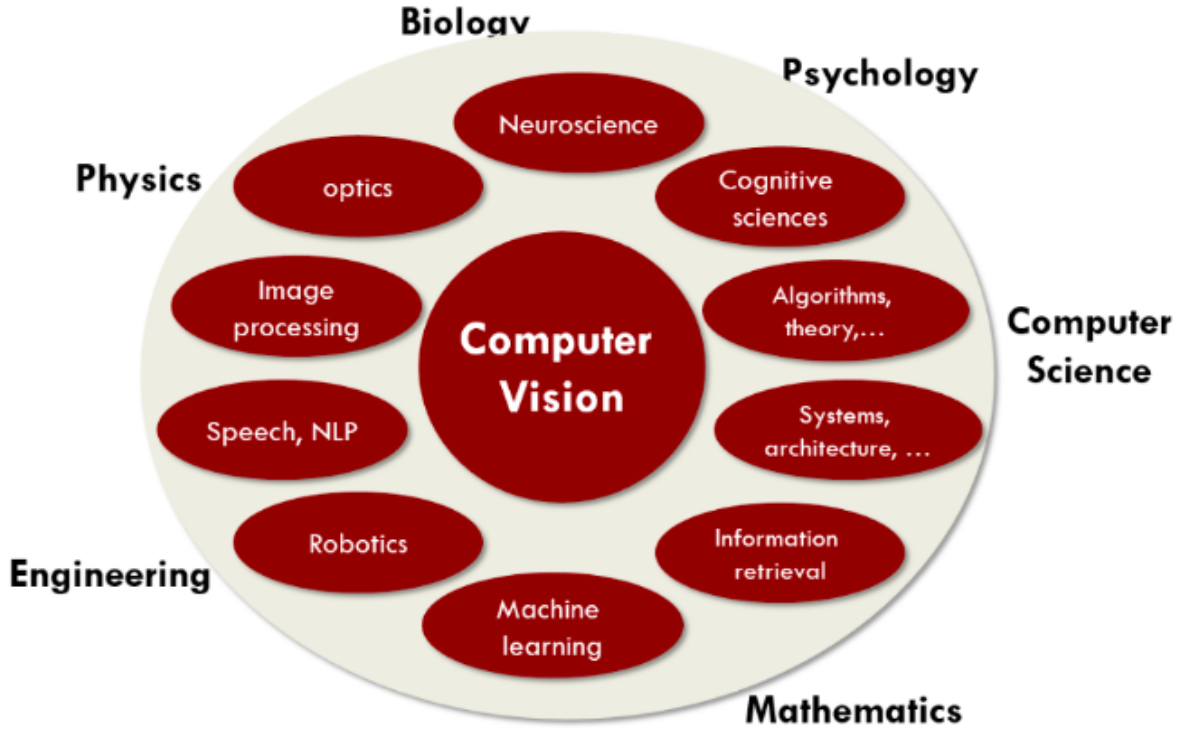
- What is Computer Vision?
- What is Image?
- Feature Detection and Matching
- Classification, Object Detection, Segmentation and Pose Estimation
- Visual Tracking
- Computer Vision Applications
- Conclusion
- References

Session Outcomes

- Acquire basic knowledge of computer vision as an interdisciplinary field.
- Understand feature detection and matching, and its applications.
- Gain basic understanding of Computer vision tasks.

What is Computer vision?

- **Computer vision** is an interdisciplinary field that deals with how computers can be made to gain high-level understanding from digital images or videos. It is a sub-field of AI.



What is image?

What we see

```
  0   2  15   0   0  11  10   0   0   0   0   9   9   0   0   0
  0   0   0   4  60 157 236 255 255 177  95  61  32   0   0  29
  0  10  16 119 238 255 244 245 243 250 249 255 222 103  10   0
  0  14 170 255 255 244 254 255 253 245 255 249 253 251 124   1
  2  98 255 228 255 251 254 211 141 116 122 215 251 238 255  49
 13 217 243 255 155  33 226  52   2   0  10  13 232 255 255  36
 16 229 252 254  49  12   0   0   7   7   0  70 237 252 235  62
  6 141 245 255 212  35  11   9   3   0 115 236 243 255 137   0
  0  87 252 250 248 215  60   0   1 121 252 255 248 144   6   0
  0  13 113 255 255 245 255 182 181 248 252 242 208  36   0  19
  1   0   5 117 251 255 241 255 247 255 241 162  17   0   7   0
  0   0   0   4  58 251 255 246 254 253 255 120  11   0   1   0
  0   0   4  97 255 255 255 248 252 255 244 255 182  10   0   4
  0  22 206 252 246 251 241 100  24 113 255 245 255 194   9   0
  0 111 255 242 255 158  24   0   0   6  39 255 232 230  56   0
  0 218 251 250 137   7  11   0   0   0   2  62 255 250 125   3
  0 173 255 255 101   9  20   0  13   3  13 182 251 245  61   0
  0 107 251 241 255 230  98  55  19 118 217 248 253 255  52   4
  0  18 146 250 255 247 255 255 255 249 255 240 255 129   0   5
  0   0  23 113 215 255 250 248 255 255 248 248 118  14  12   0
  0   0   6   1   0  52 153 233 255 252 147  37   0   0   4   1
  0   0   5   5   0   0   0   0   0  14   1   0   6   6   0   0
```

What a computer sees

What is image?

- Images are just numbers. i.e. a matrix of 2-dimensional numbers in this case (gray scale image).
- An image is made of pixels. Pixel value: 0 – 255.
- A video is a sequence of frames (images).

BW

Gray

RGB

What is image?

- RGB color image has 3 channels:

Red

Green

Blue

Feature Detection and Matching

- A pair of images to be matched. What kinds of features might one use to establish a set of correspondences between these images?

Feature Detection and Matching

- A pair of images to be matched. What kinds of features might one use to establish a set of correspondences between these images?

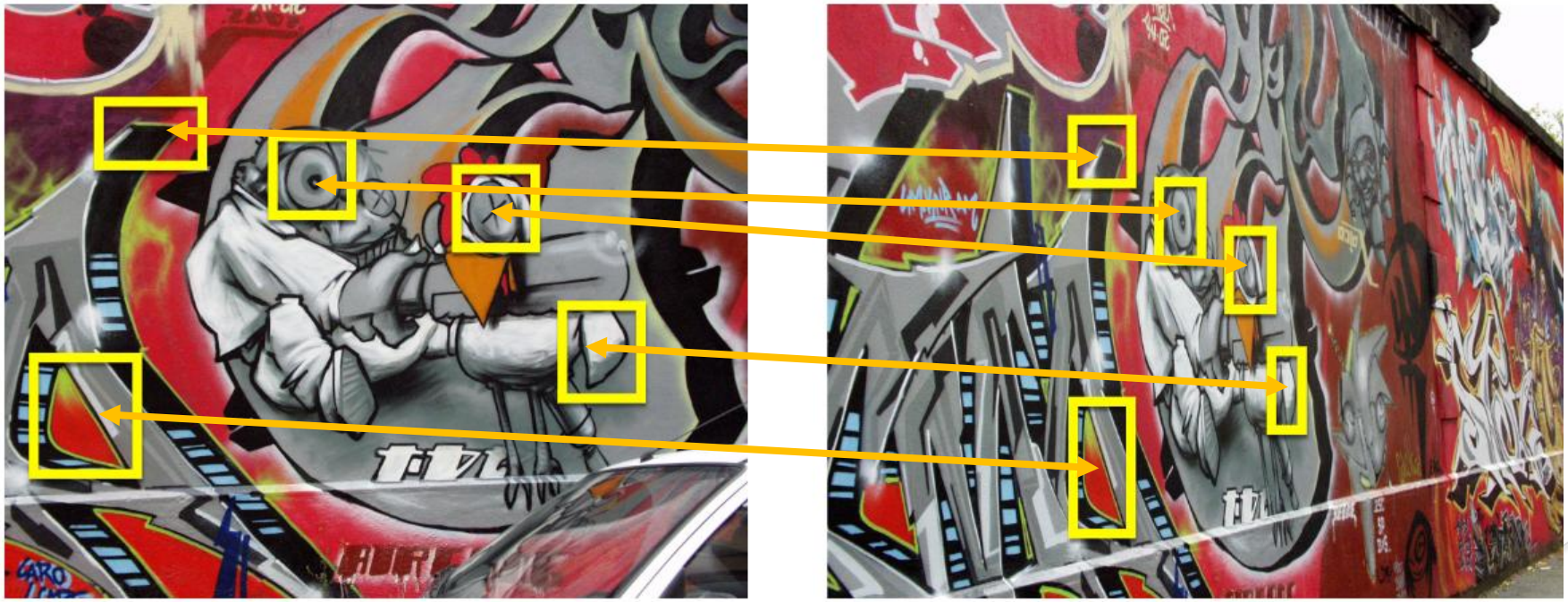
- The first kind of features that you may notice are specific locations in the images, such as building corners, doorways, etc.

Feature Detection and Matching

- These kinds of localized features are often called **keypoint features** or **interest points** (or even **corners**) and are often described by the appearance of pixel patches surrounding the point location.
- **Four** separate stages in keypoint detection and matching pipeline:
  - Feature detection (extraction) - each image is searched for locations that are likely to match well in other images.
  - Feature description - each region around detected keypoint locations is converted into a more compact and stable (invariant) descriptor that can be matched against other descriptors. There is a descriptor vector for each keypoint feature.
  - Feature matching - efficiently searching for likely matching candidates in other images.
  - Feature tracking – is more suitable for video processing i.e. for video tracking applications.

Feature Detection and Matching

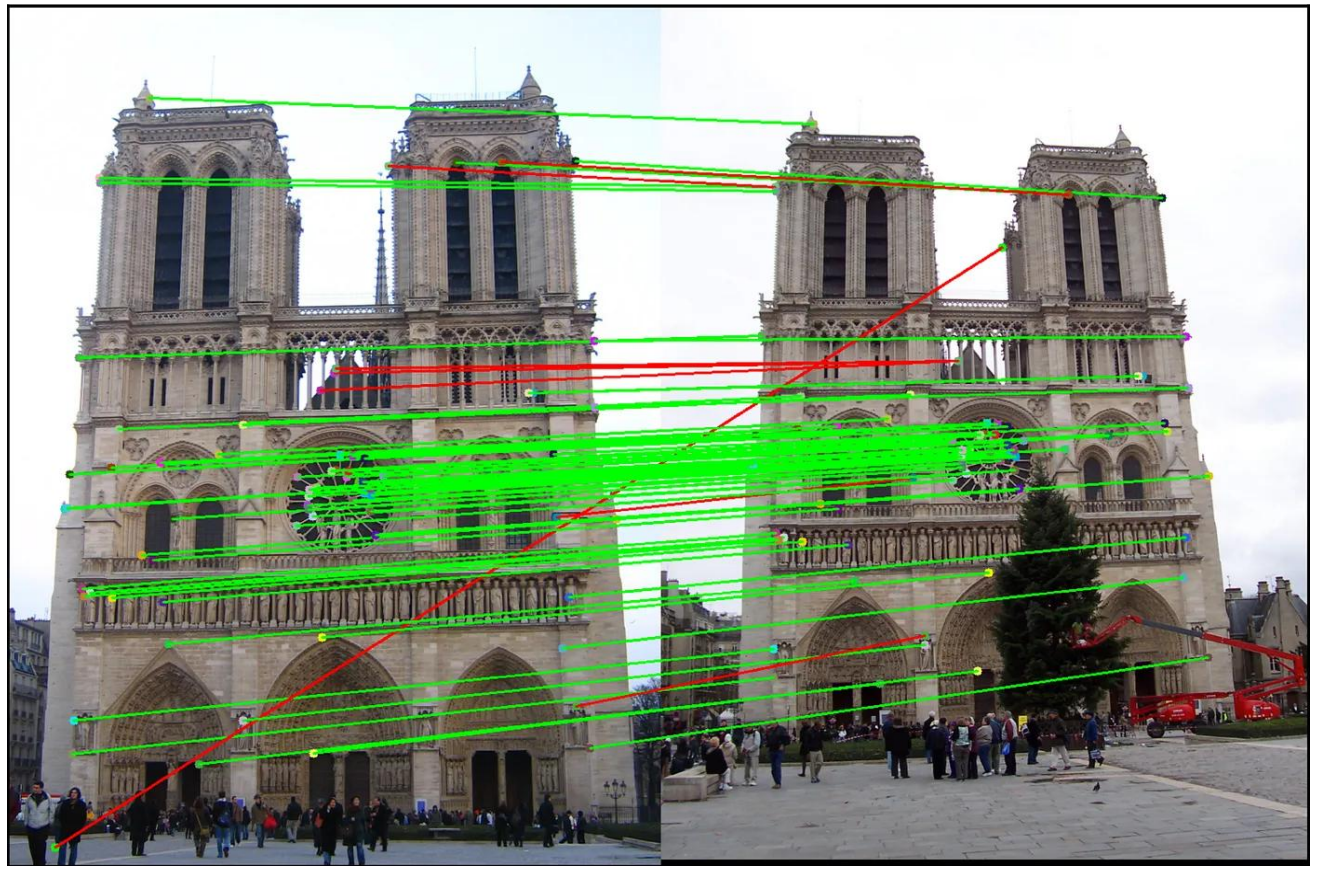
- Feature detection and matching example:



- False positive (outlier) matches can be detected and then removed using RANSAC.

Feature Detection and Matching

- Feature detection and matching example:



What are the **red** matches?
How do you remove them?

Feature Detection and Matching

- Most descriptors are computed in a local manner, hence a description is obtained for every interest point or keypoint identified. These descriptors are (ideally) invariant under changes in illumination, translation, scale, and in-plane rotation, such that they can be reliably computed with a high degree of repeatability.

- Common feature descriptor algorithms:
  - Scale Invariant Feature Transform (SIFT) – length of each descriptor vector is 128 (128-D).
  - Speeded up robust features (SURF) – length of each descriptor is 64.
  - Features from Accelerated Segment Test (FAST).
  - Binary Robust Independent Elementary Features (BRIEF).
  - Oriented FAST and Rotated BRIEF (ORB).

Feature Detection and Matching

- Common feature matching algorithms:
  - Brute-Force Matcher.
  - Fast Library for Approximate Nearest Neighbors (FLANN) Matcher.
- Applications of feature matching or image matching include:
  - Image registration
  - image mosaic or image stitching
  - Object recognition
  - Image retrieval
  - Camera calibration
  - 3D reconstruction
  - Visual SLAM e.g. https://www.youtube.com/watch?v=G-5jesjNfLc
  - Vision-based robot localization and navigation, etc.

Classification, Object Detection, Segmentation and Pose Estimation

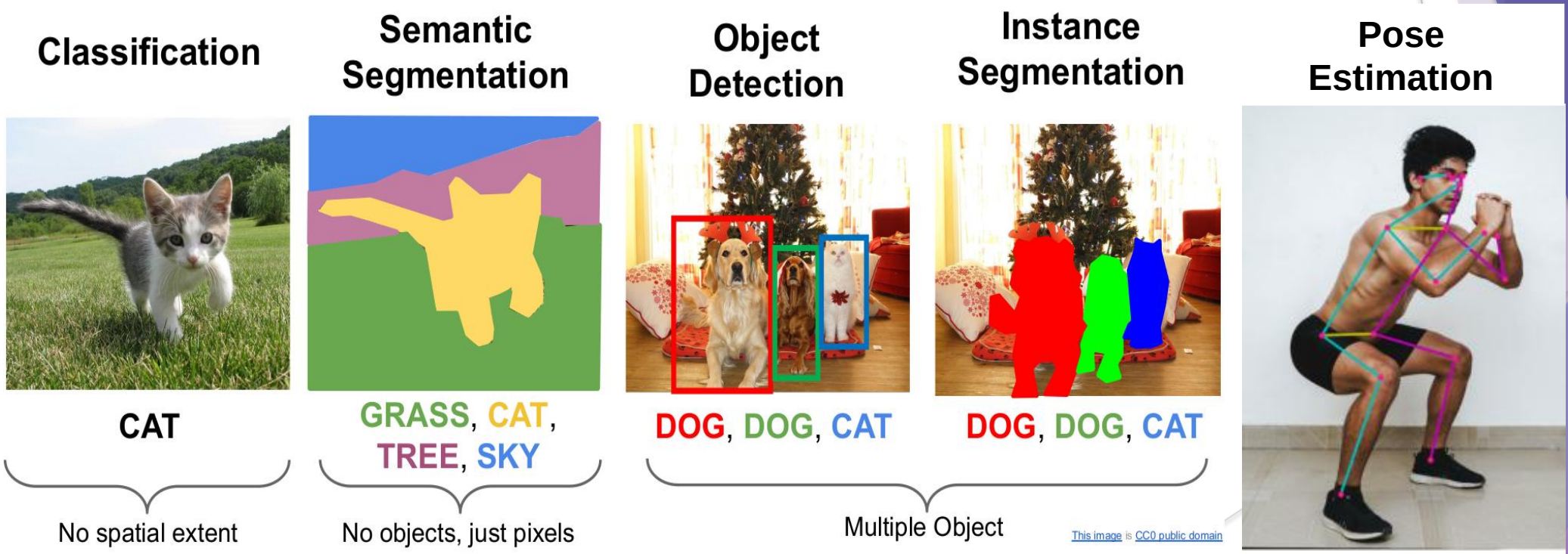
- **Image Classification:** A core task in Computer Vision.

(assume given a set of possible labels)
{dog, cat, truck, plane, ...}

cat

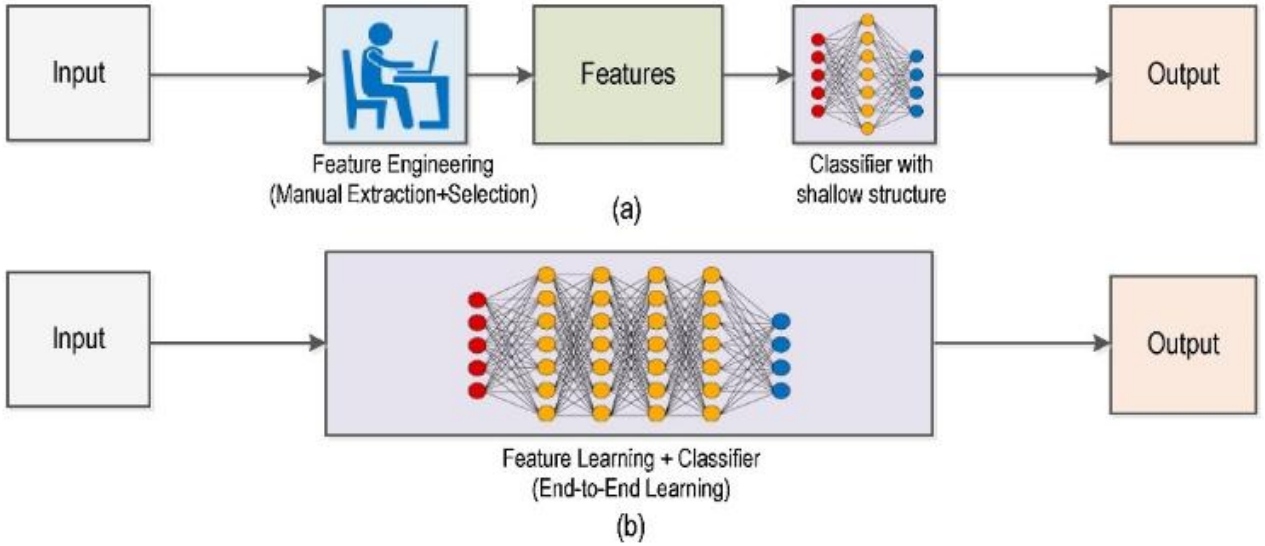
Classification, Object Detection, Segmentation and Pose Estimation

- **Computer Vision Tasks:**



Classification, Object Detection, Segmentation and Pose Estimation

- Traditional computer vision workflow (a) vs deep learning computer vision workflow (b):

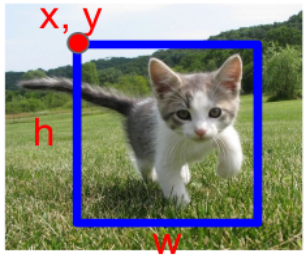


- Read more on: https://arxiv.org/abs/1910.13796

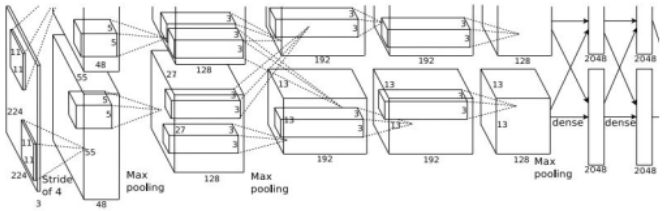
Classification, Object Detection, Segmentation and Pose Estimation

Object Detection: Single Object

(Classification + Localization)



This image is CC0 public domain



Fully Connected: 4096 to 1000

Class Scores
Cat: 0.9
Dog: 0.05
Car: 0.01
...

Correct label: Cat

Softmax Loss

Multitask Loss

Loss

Vector: 4096

Fully Connected: 4096 to 4

Box Coordinates (x, y, w, h)

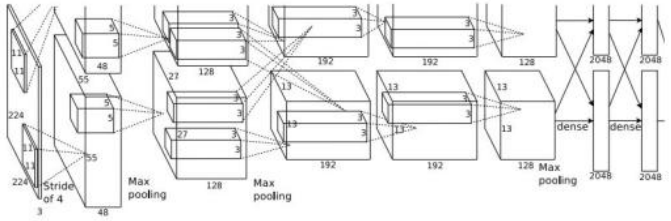
L2 Loss

Correct box: (x', y', w', h')

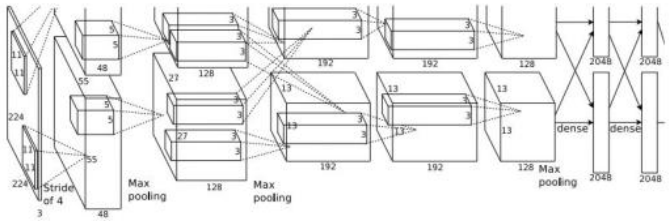
Treat localization as a regression problem!

Classification, Object Detection, Segmentation and Pose Estimation

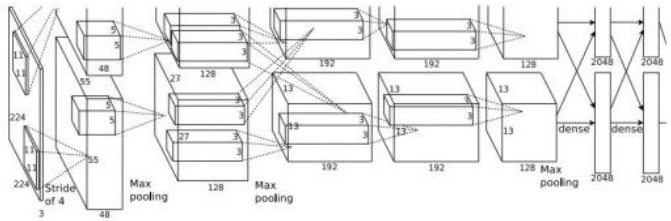
- Object Detection: Multiple Objects



CAT: (x, y, w, h)



DOG: (x, y, w, h)
DOG: (x, y, w, h)
CAT: (x, y, w, h)



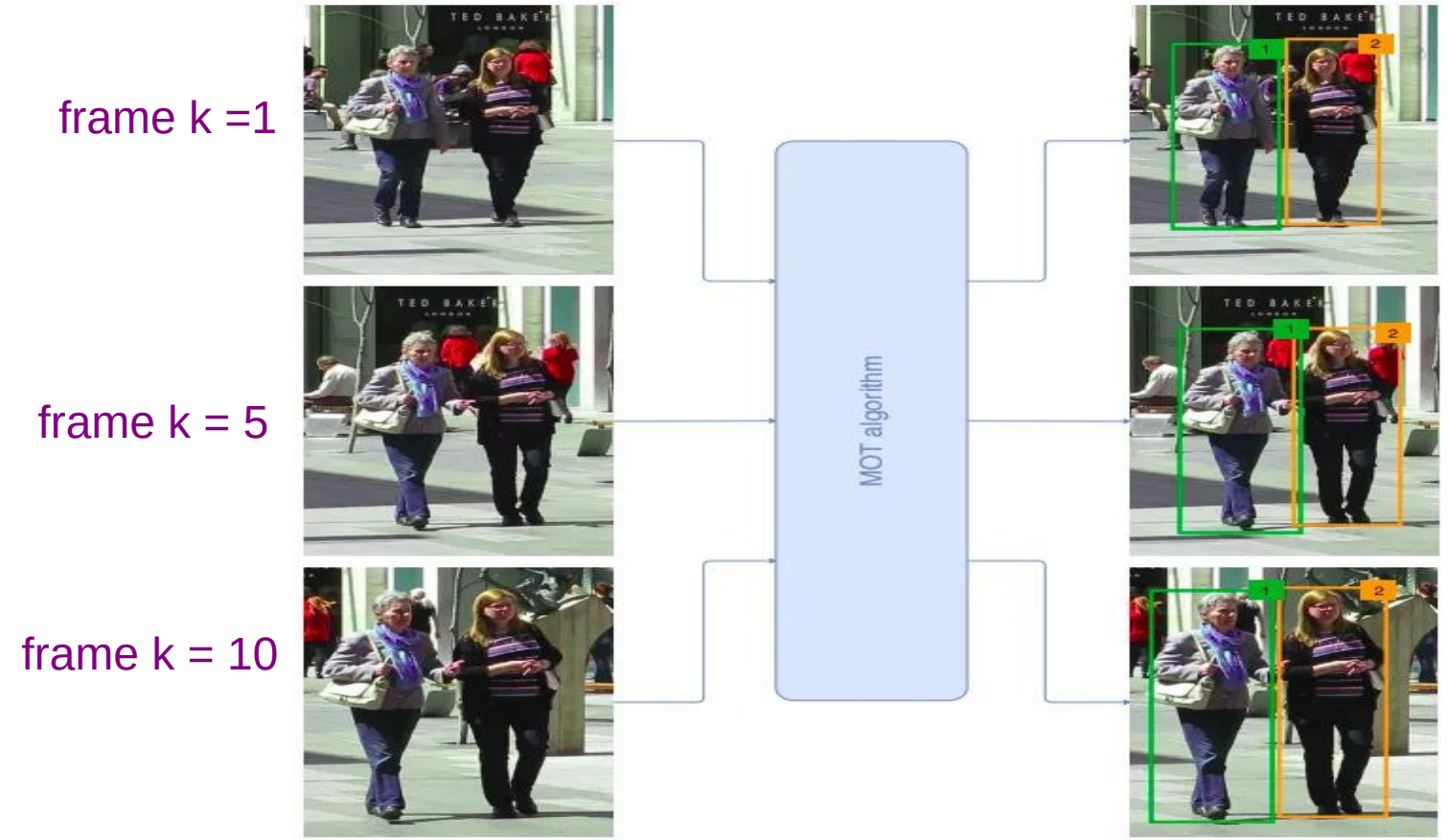
DUCK: (x, y, w, h)
DUCK: (x, y, w, h)
….

Classification, Object Detection, Segmentation and Pose Estimation

- Types of object detectors:
  - Two-stage object detectors
    - Region proposals generation + object classification and bounding box prediction.
    - E.g. Faster RCNN
  - Single-stage object detectors
    - Removes the region proposals generation stage.
    - E.g. YOLO family (YOLOv1, …, YOLOv8, …)

Visual Tracking

- Multi-Object Tracking (MOT):



Adapted from:
https://arxiv.org/pdf/1907.12740

Visual Tracking

- Key components of MOT:
  - Object detection
    - YOLO family, FRCNN, etc.
  - Motion prediction
    - Kalman filter
    - GM-PHD filter
  - Data association
    - Hungarian algorithm using:
      - Motion information i.e. Intersection over Union (IoU), Mahalanobis distance, etc. and/or
      - Appearance information

Visual Tracking

- Example of MOT:

- Figure from: Nathanael L. Baisa, "Occlusion-robust online multi-object visual tracking using a GM-PHD filter with CNN-based re-identification," Journal of Visual Communication and Image Representation, 2021.

Computer Vision Applications

- **Security** e.g. Face Recognition:
  - How does facial recognition work? (youtube.com)
- **Security** e.g. Surveillance:
  - How China is building an all-seeing surveillance state (youtube.com)
- **Computer Vision (CV) in Robotics:**
  - Computer vision applications in robotics (youtube.com)
- **Computer Vision in Self Driving Cars and Autonomous Vehicles:**
  - Self Driving Cars and Autonomous Vehicles Technology (youtube.com)
- **Computer Vision in HealthCare:**
  - What is Computer Vision in Healthcare Education (youtube.com)

Conclusion

- Image classification is a core task in computer vision.
- Feature detection and matching plays a crucial role in vision-based robot mapping, localization and visual SLAM.
- Understanding computer vision tasks is important to apply them to intelligent robot-based applications, security, healthcare, etc.

References

- R. Szeliski, 'Computer Vision: Algorithms and Applications', Springer, 2021. [https://szeliski.org/Book/]
- https://medium.com/data-breach/introduction-to-feature-detection-and-matching-65e27179885d
- OpenCV: https://opencv.org/
- Open3D: http://www.open3d.org/
- https://huggingface.co/learn/computer-vision-course/en/unit0/welcome/welcome
- https://github.com/polygon-software/python-visual-odometry
- https://github.com/luigifreda/pyslam