Introduction to Convolutional Neural Networks (CNNs) and Object Detection

CET 513 Autumn 2025

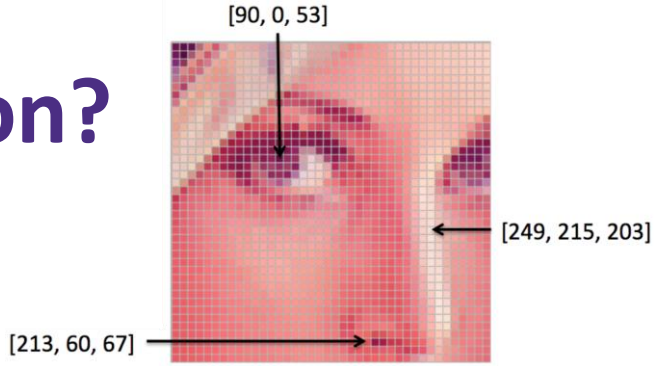
Soheil Keshavarz -- Nov 24, 2025

Intelligent Urban Transportation Systems
UNIVERSITY of WASHINGTON

Outline

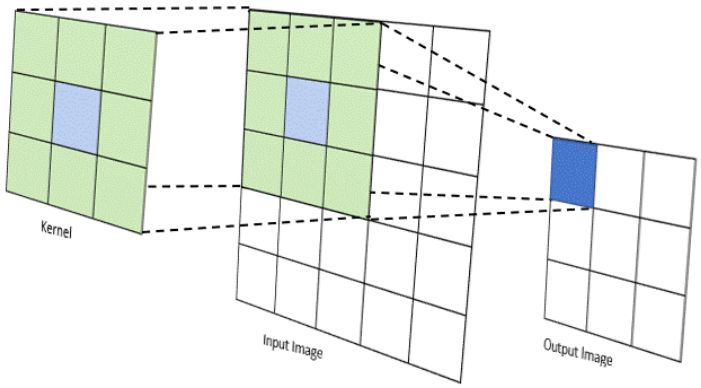
- Convolution in digital images
- Neural Networks Refresher
- Convolutional neural networks
- R-CNN
- YOLO

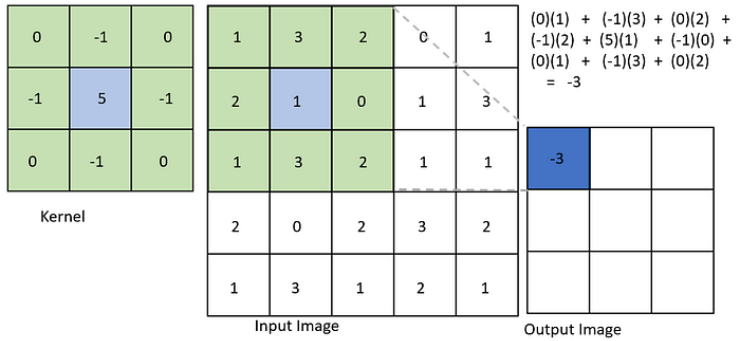
1- What is Image Convolution?



https://ai.stanford.edu

- Digital images are represented by pixels in RGB,
  1 byte for each channel
- Convolution:
  - an operation that slides a small kernel (aka filter, feature detector) across the image and computes weighted sums at each position.
  - The kernel size and values determine the effect the kernel has on the image.

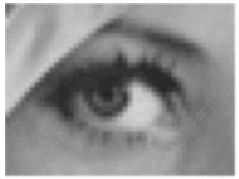




https://medium.com

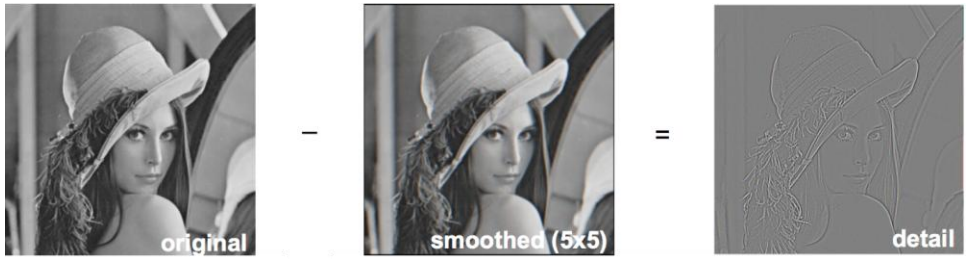
1- Example of Convolution

- Convolution extracts local spatial features from images (e.g., edges, corners, textures)
- Examples:
  - Smoothing/Blurring


Original

$$\frac{1}{9}\begin{bmatrix}1 & 1 & 1\\ 1 & 1 & 1\\ 1 & 1 & 1\end{bmatrix} = \; ?$$

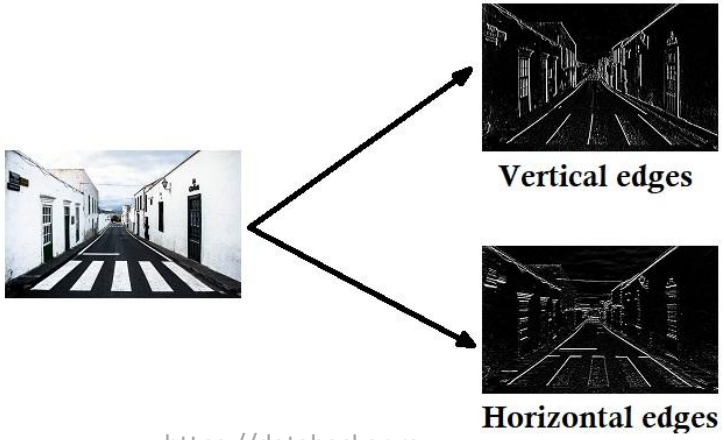
  - Finding details



$$\begin{bmatrix}0 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 0\end{bmatrix} - \frac{1}{9}\begin{bmatrix}1 & 1 & 1\\ 1 & 1 & 1\\ 1 & 1 & 1\end{bmatrix}$$

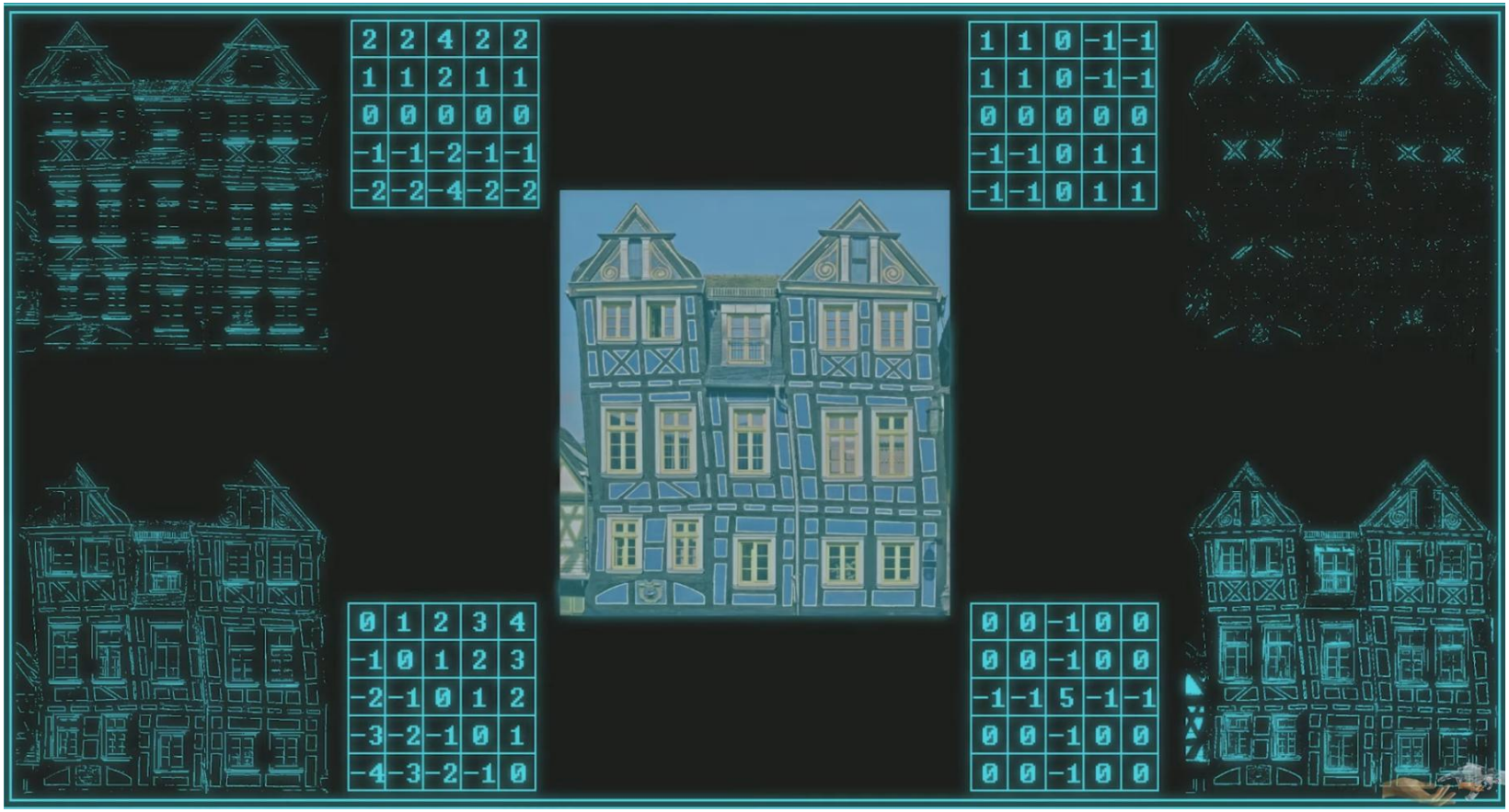
https://ai.stanford.edu

  -- Edge detection



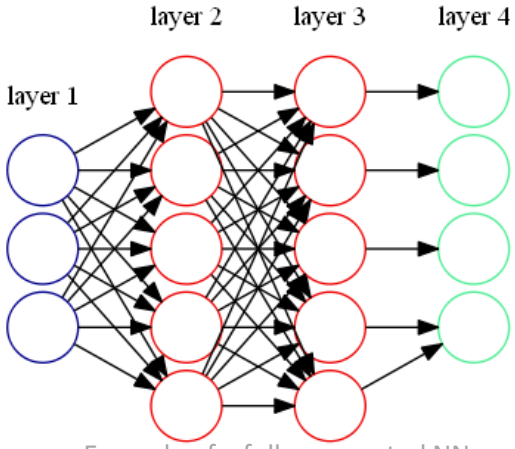
https://datahacker.rs

1- Example of Convolution



https://www.youtube.com/watch?v=pj9-rr1wDhM

2- Neural Networks



Example of a fully connected NN
https://www.geeksforgeeks.org/

- A multi-layered network of artificial neurons
- An artificial **neuron**:
  - receives inputs, performs **a linear operation** on these inputs, and passes the result through a **non-linear activation function**.
- **Network Structure:**
  - Input Layer, Hidden Layers, Output Layer
- **Backpropagation**:
  - Forward Pass, Compute Loss, Backward Pass, Update Weights

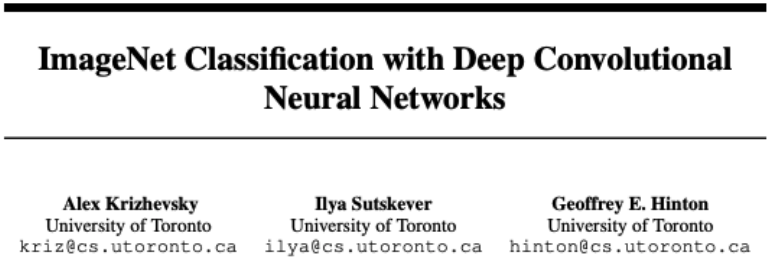
$$a = \sigma(Wx + b) \qquad L = \text{Loss}(y, \hat{y}) \qquad \frac{\partial L}{\partial W_i} = \frac{\partial L}{\partial a_i} \cdot \frac{\partial a_i}{\partial z_i} \cdot \frac{\partial z_i}{\partial W_i} \qquad W_i \leftarrow W_i - \eta \frac{\partial L}{\partial W_i}$$

- Visualizations:
  - https://mlu-explain.github.io/neural-networks/
  - https://playground.tensorflow.org/

3- Convolutional Neural Networks (CNNs)

> The standard deep learning architecture for computer vision and image processing

> Popularity started from AlexNet (2012)



ImageNet Classification with Deep Convolutional Neural Networks

Alex Krizhevsky
University of Toronto
kriz@cs.utoronto.ca

Ilya Sutskever
University of Toronto
ilya@cs.utoronto.ca

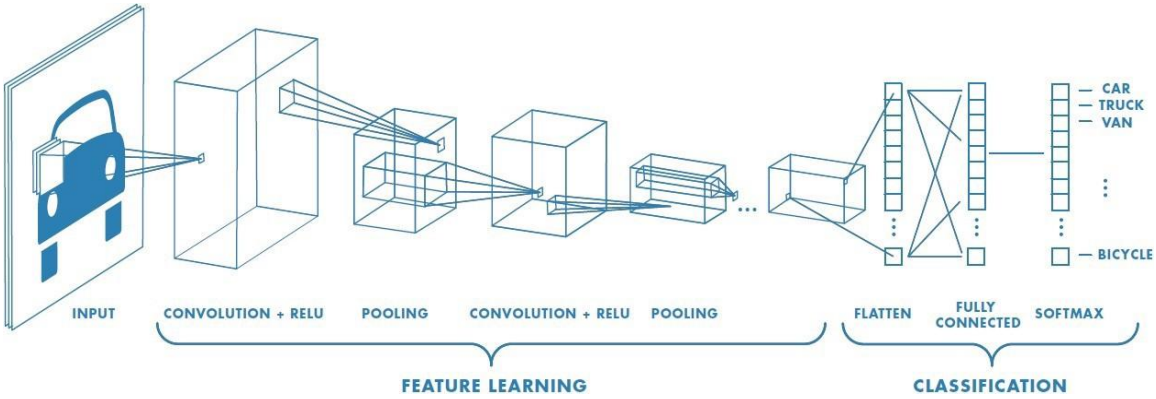
Geoffrey E. Hinton
University of Toronto
hinton@cs.utoronto.ca

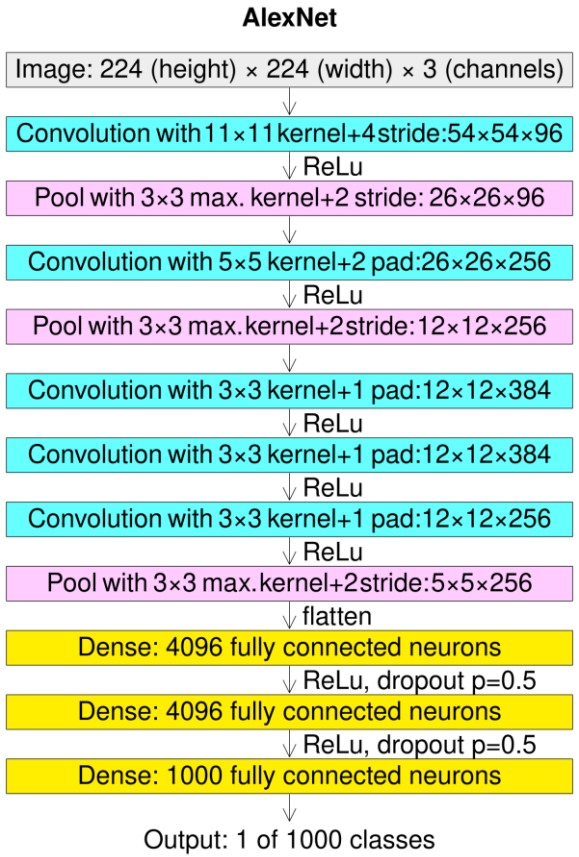
- Ignited the deep learning revolution
- Extended (and deepened) Yann LeCun's earlier CNN models which used back-propagation to learn convolution kernels
- Trained with GPUs on ImageNet dataset (Fei-Fei Li)

3- CNN architecture

- Goal: to predict a label, given an image

- Comprised of
  1. Convolutional layers to learn features (why?)
  2. Pooling layers to reduce dimensionality and preserve spatial invariance
  3. A classifier (features → detection labels)



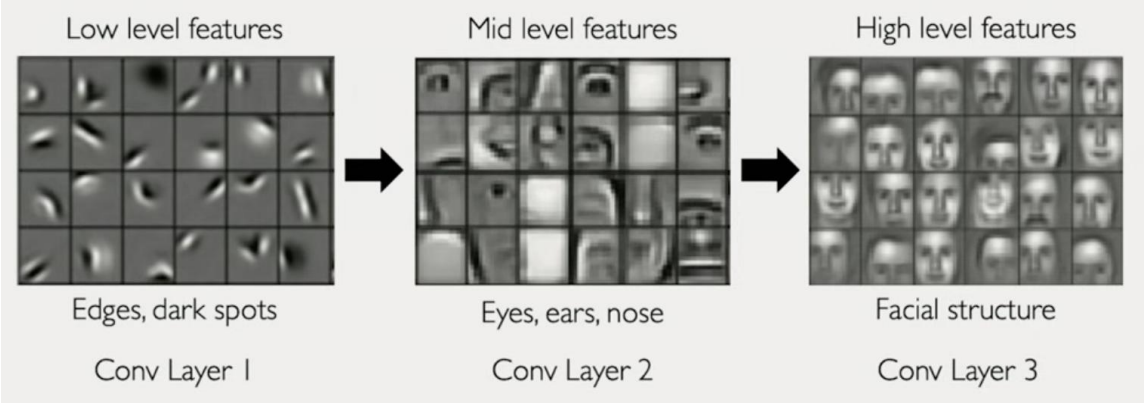


AlexNet layers (Wikipedia)

3- CNN feature extraction

- Kernels are learned from training

- Early layers detect basic features
  - Ex: edges in various angels, corners, simple shapes
- Deeper layers detect high-level features
  - Ex: faces, textures, or uninterpretable features



3- CNNs – More resources

- YouTube!
  - A brief history of AlexNet and interpretation of its layers
  - A general explanation of CNNs and its components
  - a more detailed video
  - a great one-hour MIT lecture

- Chenxi Liu two one-hour workshops in C2SMART Learning Hub:
  - "How to Build a Toy Computer Vision (CV) Model"
  - "Computer Vision: Object Detection & Tracking"

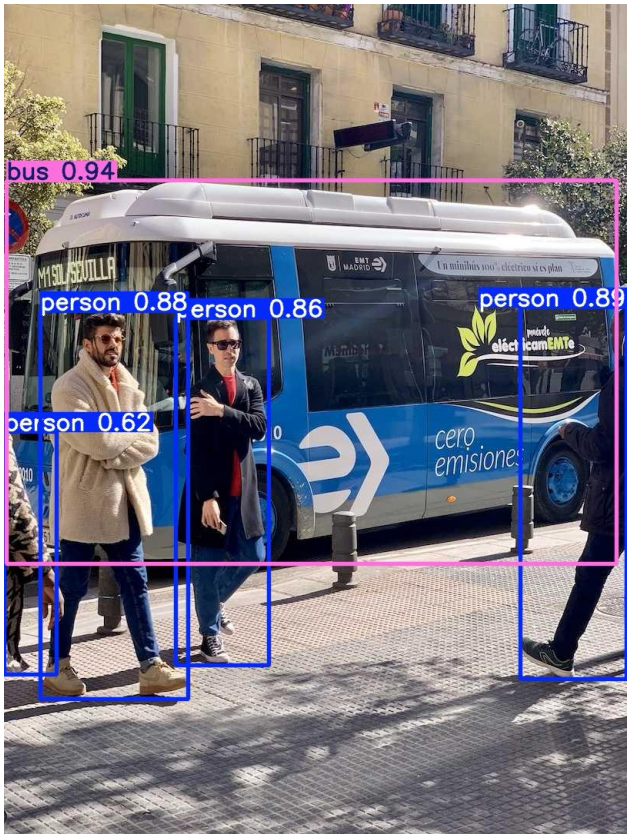
- Visualization:
  - https://adamharley.com/nn_vis/cnn/2d.html
  - https://poloclub.github.io/cnn-explainer/

Classification → object detection

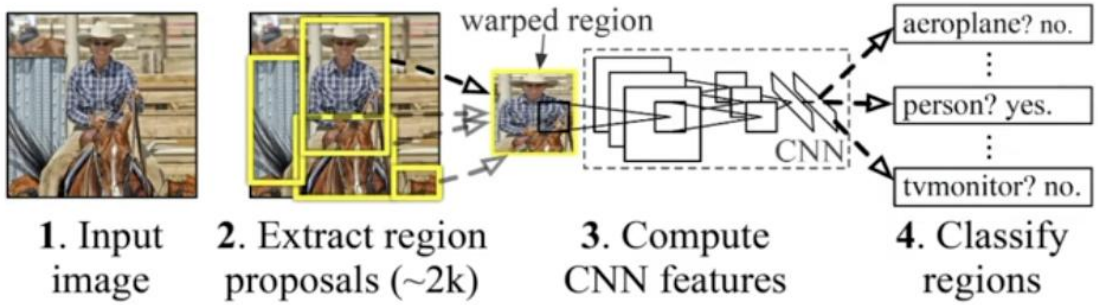
- Good for one object.

- How to find multiple objects?
- How to find a good bounding box?
  - Feed all random boxes to the model?

MIT lecture



4 – R-CNN, Faster R-CNN



Three models to run:

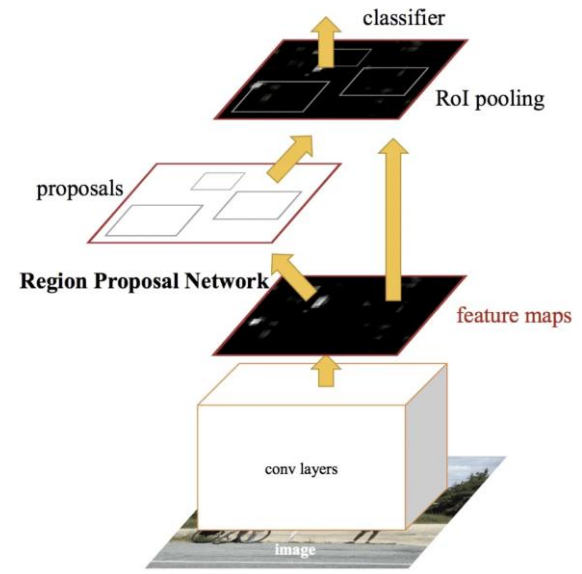
1- Extract 2k region proposals (heuristic)

2- Use CNN on each proposal to extract features and classify it

- Very slow and CNN is used a lot

Faster R-CNN:

removes Selective Search and uses a Region Proposal Network (RPN), allowing the model to learn proposals directly from the shared convolutional features.



Faster R-CNN Model Architecture.Taken from: Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, 2016.

5 – YOLO *(You only look once)*

- RCNN models:
  - Two steps for object detection:
    extract region proposals + CNN feature extraction

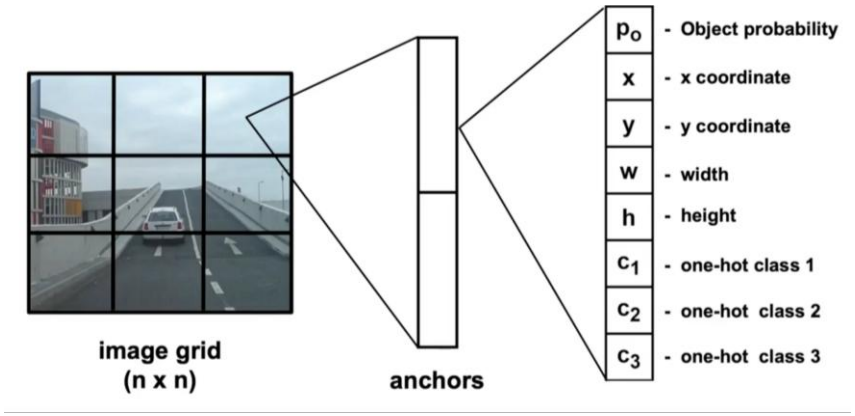
- YOLO runs the entire detection in ONE forward pass
  - Predicts bounding boxes and classes at the same time
  - Feasible for real-time application

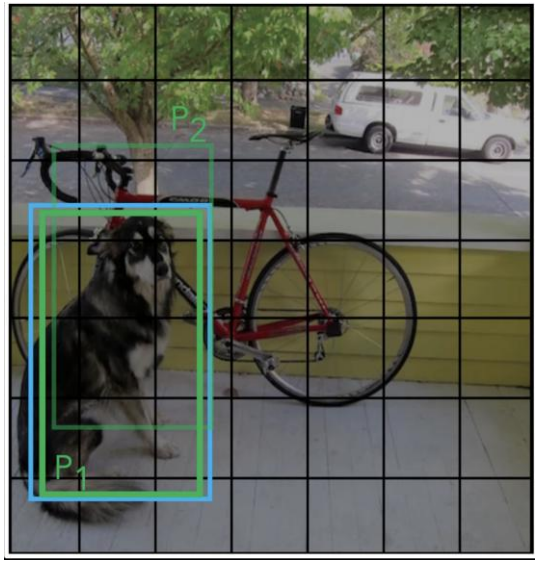
  ☑ Pro: faster
  ✗ Con: traditionally lower accuracy

- detailed video
- short video

5 – YOLO *(Cont'd)*

- YOLO:
  - Image → nxn grid → for each cell: k anchors generated → remove low probability anchors + apply non-max suppression, which is based on IOU (intersection over union)





5 – Want to build a YOLO model?

A simple workflow without dealing with model details:

1. Select a dataset or create one in Roboflow. Used for
   - For dataset creation/annotation
   - Data augmentation
   - Train/test/val split

2. Select a pre-trained YOLO model from Ultralytics
   - Hosts image segmentation and object detection models

3. Train the model (use Google Colab if GPU is needed)
4. Evaluate your model
   - Does it detect objects correctly?
   - Do you need more training data?

5- YOLO demo

- We will show:
  - A real-time object detection
  - Training a model
  - vehicle and pedestrian classification given a fisheye video

- Codes: GitHub repo

Thank you!

Other references:

- https://chatgpt.com/ all contents reviewed and verified.
- https://ai.stanford.edu/~syyeung/cvweb/tutorial1.html
- https://medium.com/@timothy_terati/image-convolution-filtering-a54dce7c786b
- https://datahacker.rs/edge-detection/
- https://www.geeksforgeeks.org/deep-learning/how-to-visualize-a-neural-network-in-python-using-graphviz/