Deep Computer Vision

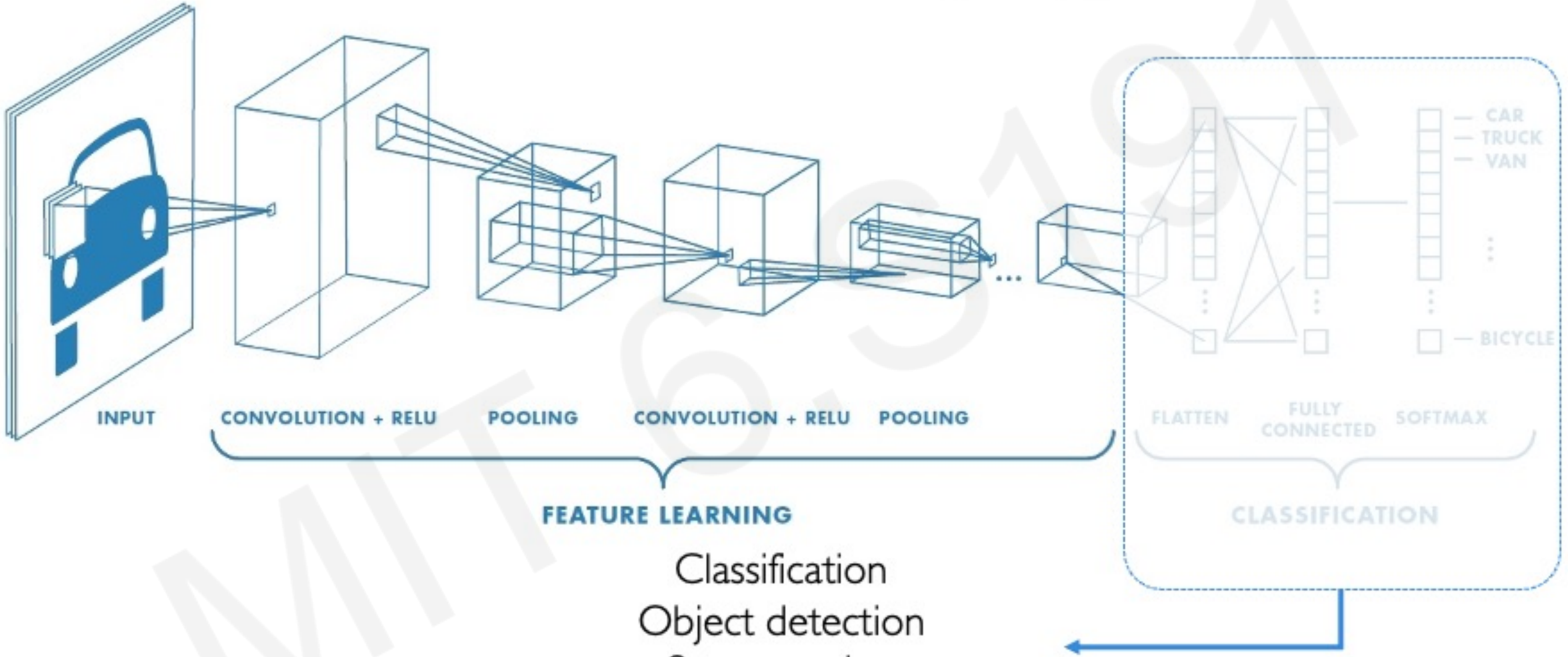
Alexander Amini

MIT Introduction to Deep Learning

January 7, 2025

An Architecture for Many Applications

An Architecture for Many Applications

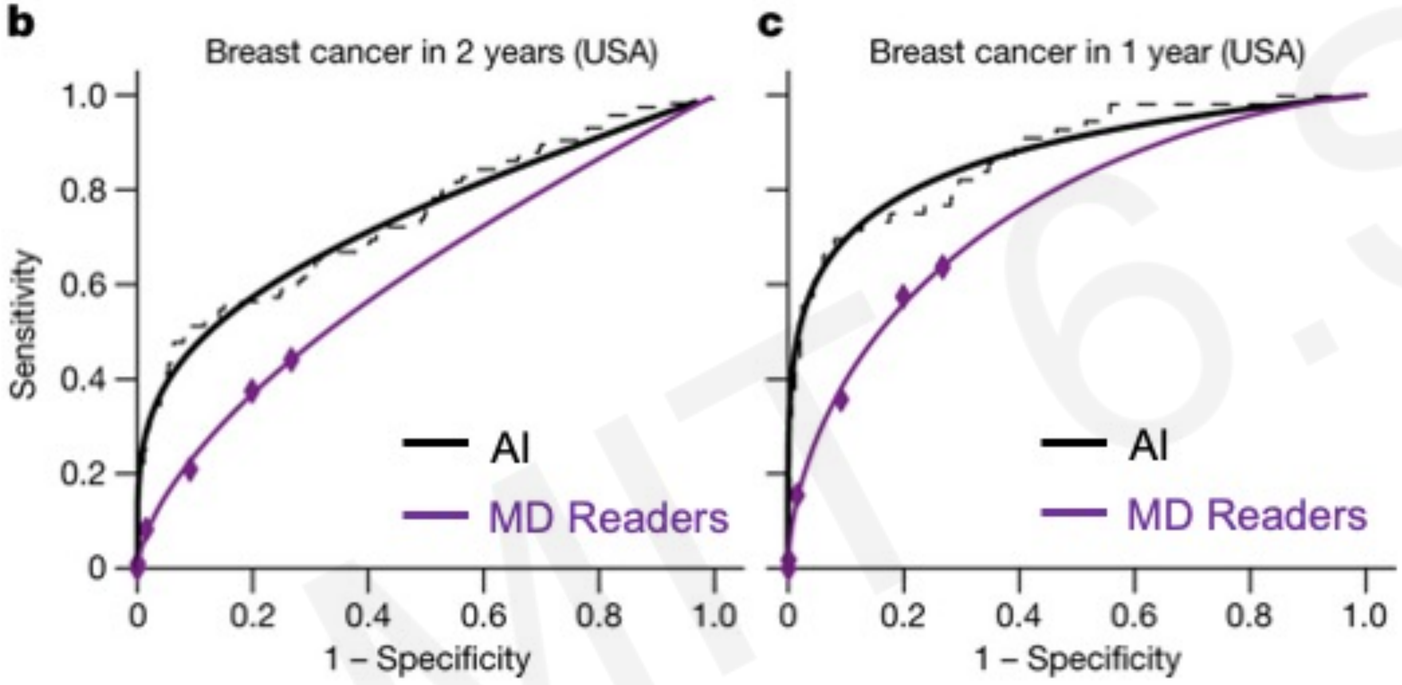


Classification
Object detection
Segmentation
Probabilistic control

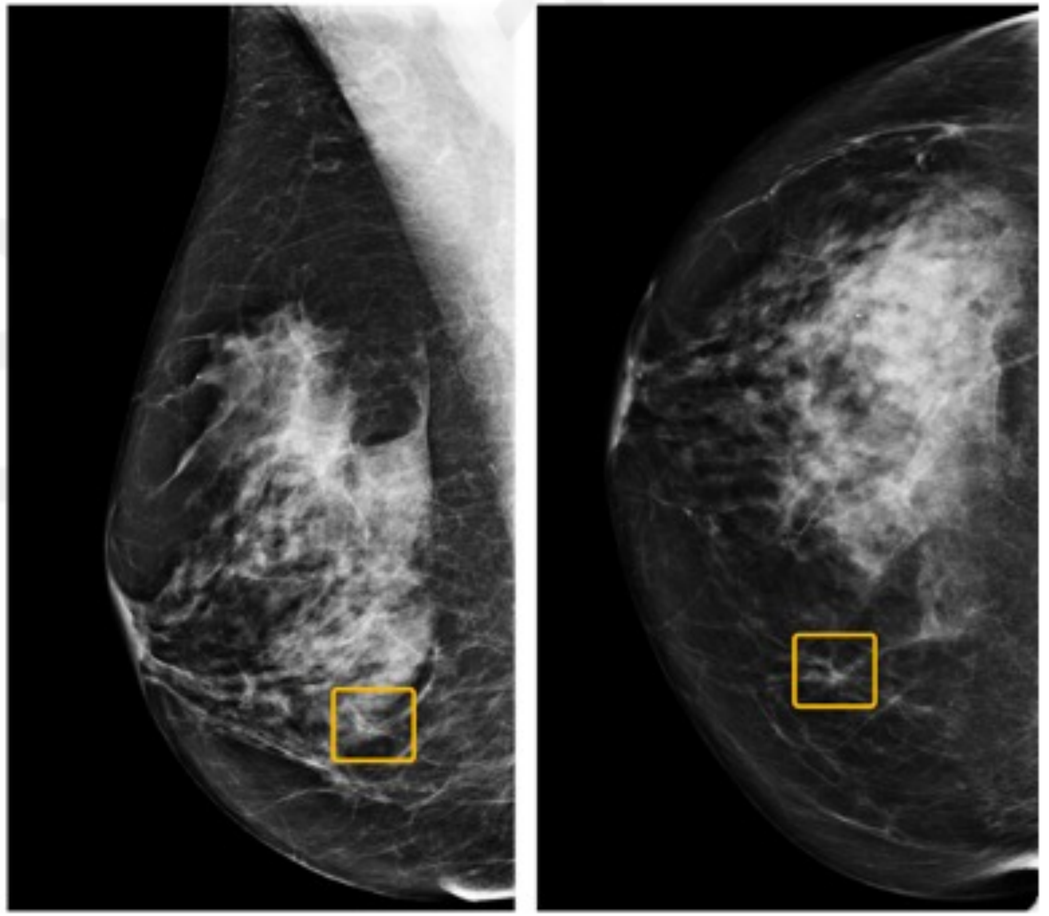
Classification: Breast Cancer Screening

International evaluation of an AI system for breast cancer screening

nature



CNN-based system outperformed expert radiologists at detecting breast cancer from mammograms



Breast cancer case missed by radiologist but detected by AI

Object Detection

Image x

CNN

Taxi

Class label y

Image x

CNN

Label (x, y, w, h)

Object Detection

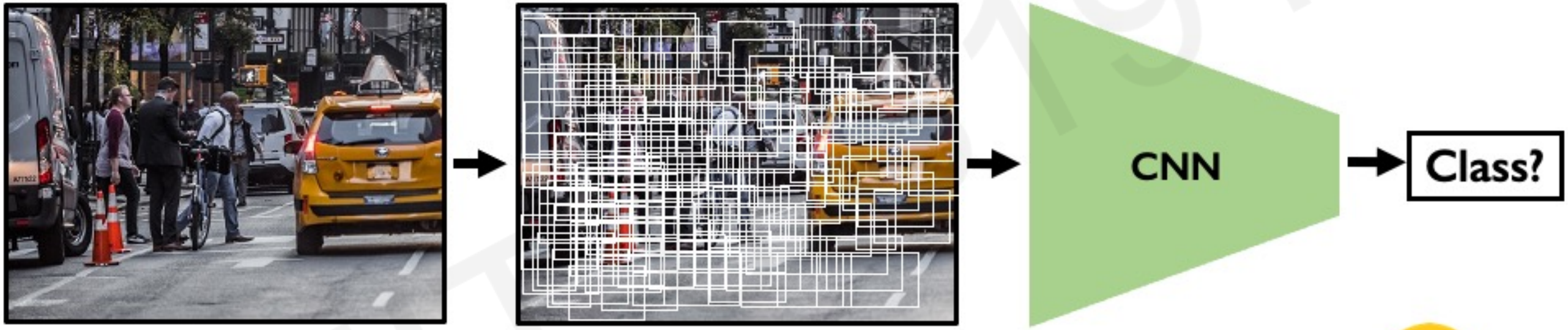
Image x

Output:
taxi: (x1, y1, w1, h1)

Image x

Output:
taxi: (x1, y1, w1, h1)
person: (x2, y2, w2, h2)
person: (x3, y3, w3, h3)
....

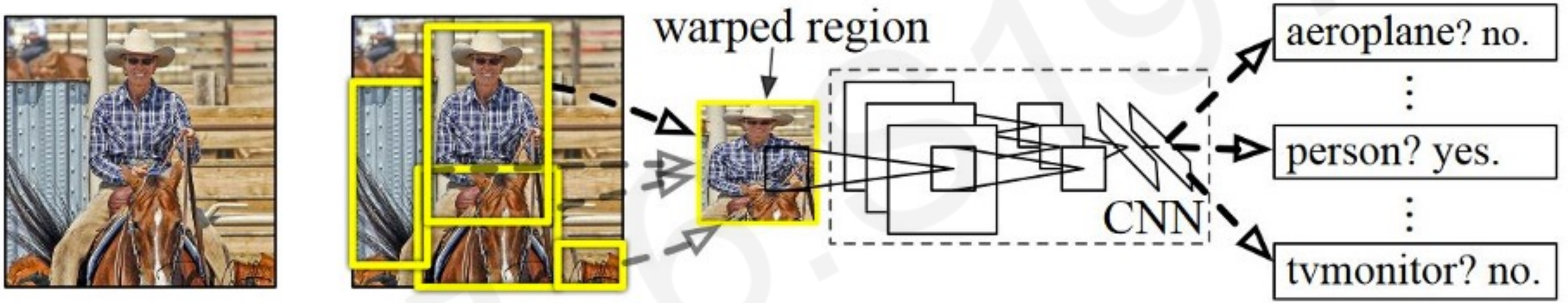
Naïve Solution to Object Detection



Problem: Way too many inputs! This results in too many scales, positions, sizes!

Object Detection with R-CNNs

R-CNN algorithm: Find regions that we think have objects. Use CNN to classify.



1. Input image

2. Extract region proposals (~2k)

3. Compute CNN features

4. Classify regions

Problems: 1) Slow! Many regions; time intensive inference.
2) Brittle! Manually defined region proposals.

Faster R-CNN Learns Region Proposals

Classification of regions → object detection

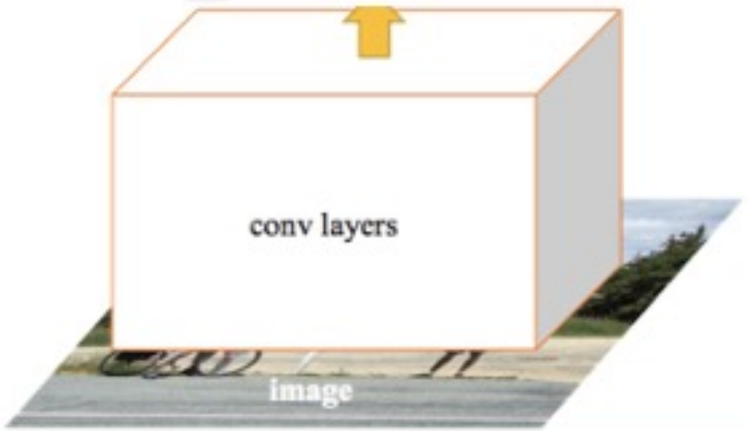
Feature extraction over proposed regions

Region proposal network to learn candidate regions

Learned, data-driven

Image input directly into convolutional feature extractor

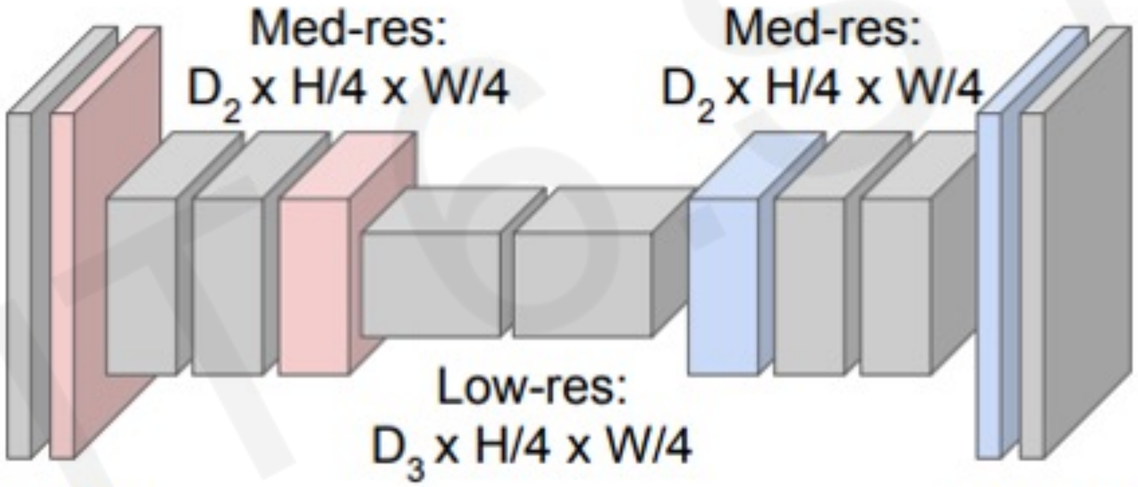
Fast! Only input image once!



Semantic Segmentation: Fully Convolutional Networks

FCN: Fully Convolutional Network.
Network designed with all convolutional layers,
with **downsampling** and **upsampling** operations

Input:
3 x H x W



High-res:
D_1 x H/2 x W/2

High-res:
D_1 x H/2 x W/2

Predictions:
H x W

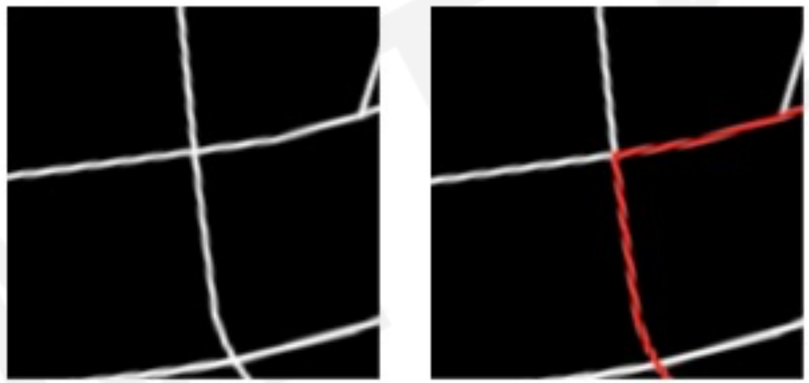
`tf.keras.layers.Conv2DTranspose`

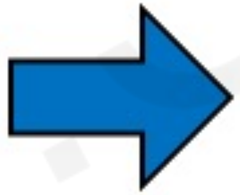
`torch.nn.ConvTranspose2d`

Continuous Control: Navigation from Vision

Raw Perception
I
(ex. camera)

Coarse Maps
M
(ex. GPS)

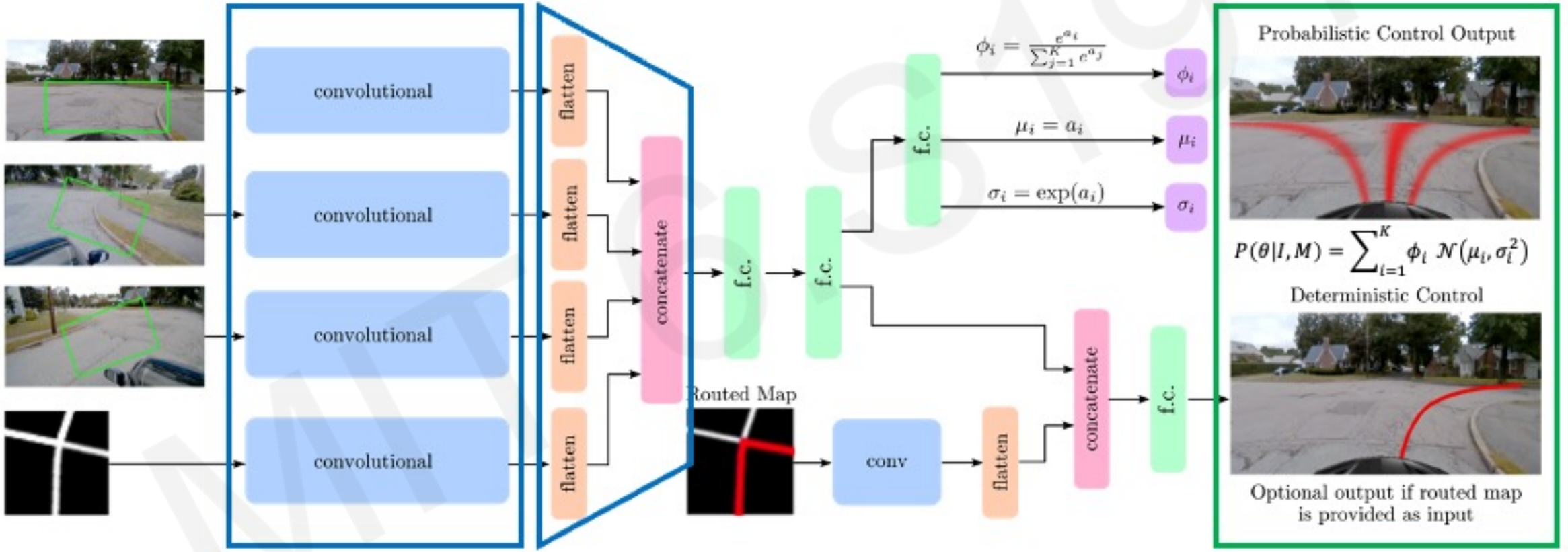




Possible Control Commands

End-to-End Framework for Autonomous Navigation

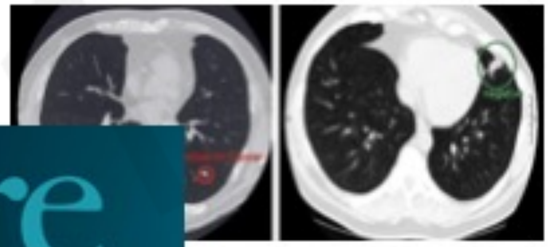
Entire model is trained end-to-end **without any human labelling or annotations**

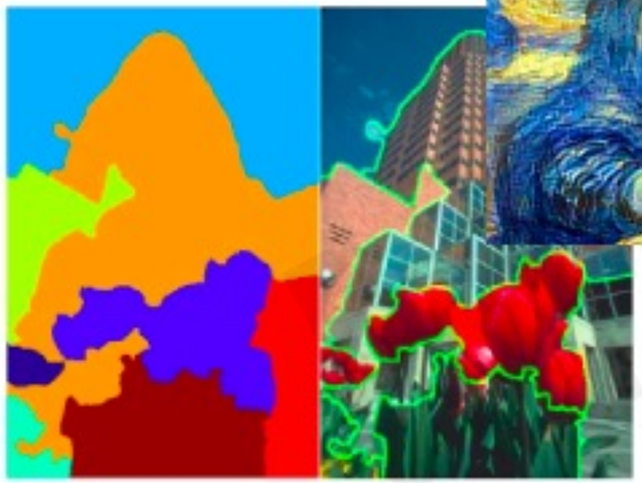


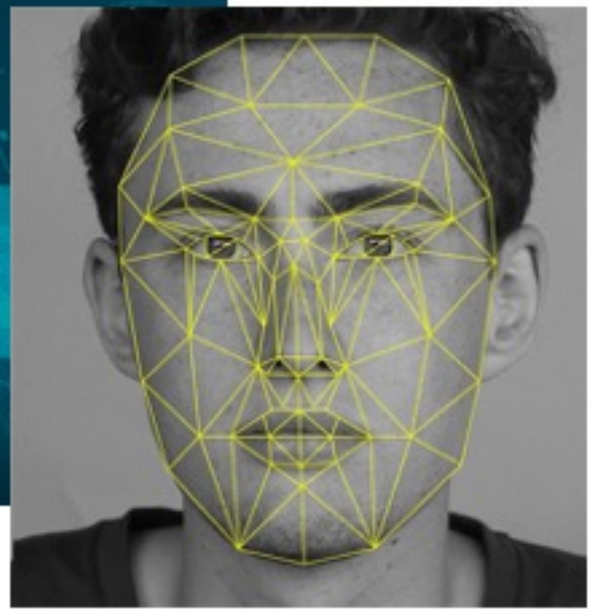
$$L = -\log(P(\theta|I,M))$$


Auto ON
Navigation and Localization
Amini+ ICRA 2019.

Deep Learning for Computer Vision: Impact







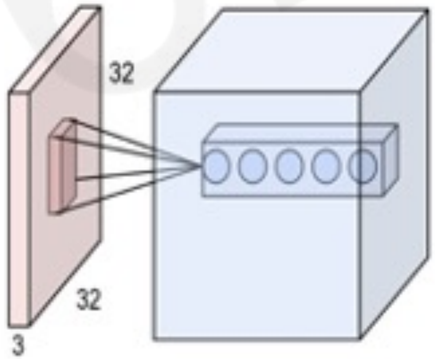
Deep Learning for Computer Vision: Summary

Foundations

- Why computer vision?
- Representing images
- Convolutions for feature extraction

CNNs

- CNN architecture
- Application to classification
- ImageNet



Applications

- Segmentation, image captioning, control
- Security, medicine, robotics