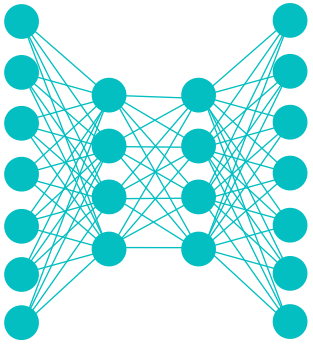
Lecture Notes for

Neural Networks and Machine Learning

Fully Convolutional Learning I: Introduction to Semantic Segmentation



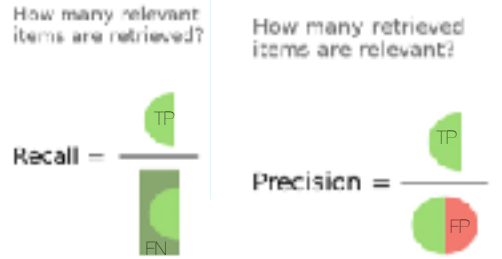
Logistics and Agenda

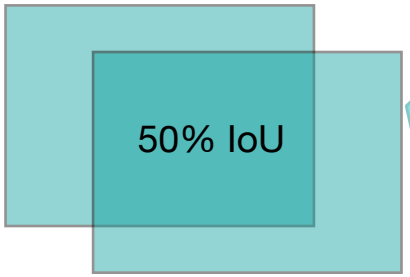
- Logistics
  - Lab Grading
  - Hiring
  - Office Hours (slightly late start today)
- Agenda
  - Intro to Semantic Segmentation
  - Paper Presentation
  - Upsampling Layers
  - Object Segmentation (next time)
  - Instance Segmentation (next time)

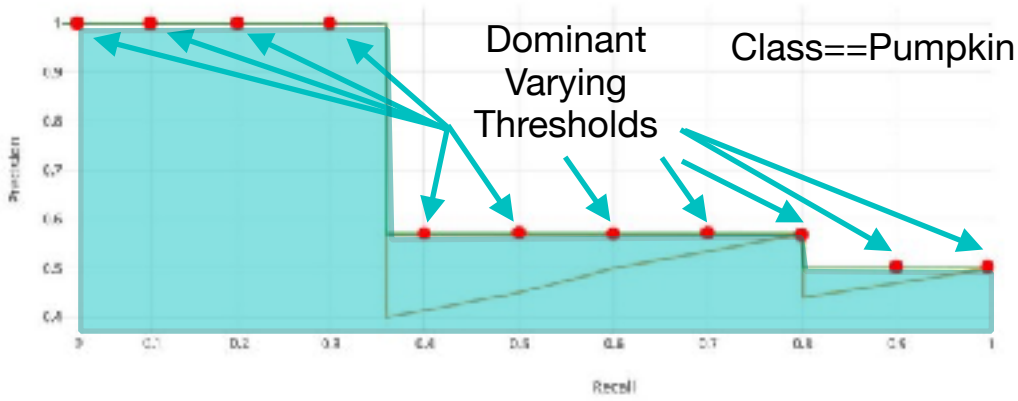
Types of Fully Convolutional Problems

- Semantic Segmentation
- Object Detection
- Instance Segmentation

Measuring Performance







- mAP(IoU=x%)
  - if IoU > X%, check if correct
    - else not correct
  - Usually~50%, 75%, 90%
  - Define precision for each class, take average
- mAP(%), *sometimes just AP*
  - Formulate precision/recall curve for a class at varying levels of confidence (for given IoU)
  - Calculate dominating points
  - Take area under precision recall curve (AUPRC)
  - Take average AUPRC over all classes (macro or micro, usually macro)

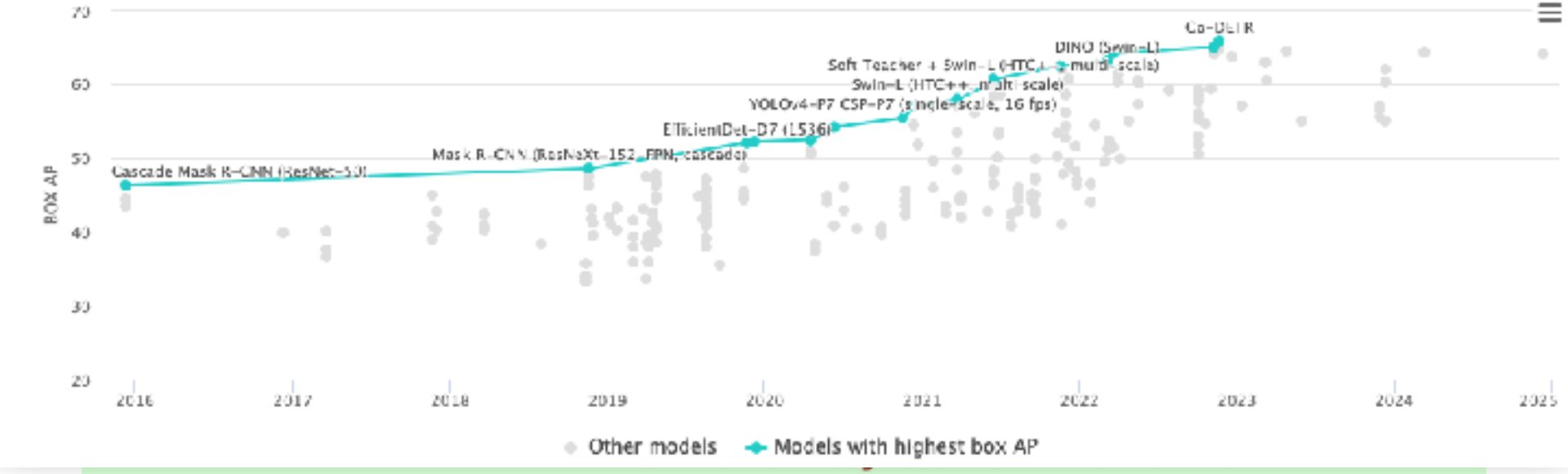
https://jonathan-hui.medium.com/map-mean-average-precision-for-object-detection-45c121a31173

COCO Evaluation

Average Precision (AP):

AP % AP at IoU=.50:.05:.95 (primary challenge metric)

AP$^{IoU=.50}$ % AP at IoU=.50 (PASCAL VOC metric)



AR^{large} % AR for large objects: area > 96^2

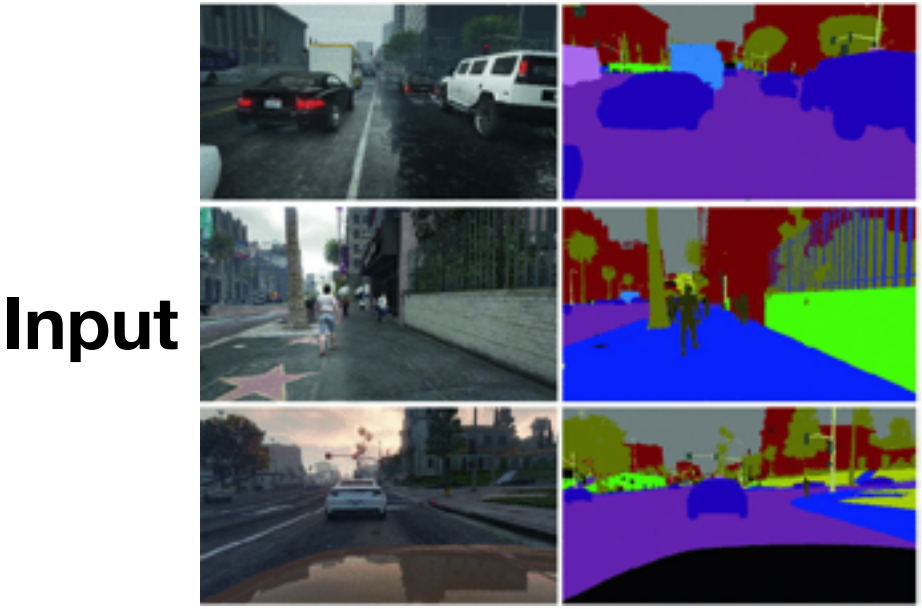
1. Unless otherwise specified, *AP and AR are averaged over multiple Intersection over Union (IoU) values.* Specifically we use 10 IoU thresholds of .50:.05:.95. This is a break from tradition, where AP is computed at a single IoU of .50 (which corresponds to our metric $AP^{IoU=.50}$). Averaging over IoUs rewards detectors with better localization.

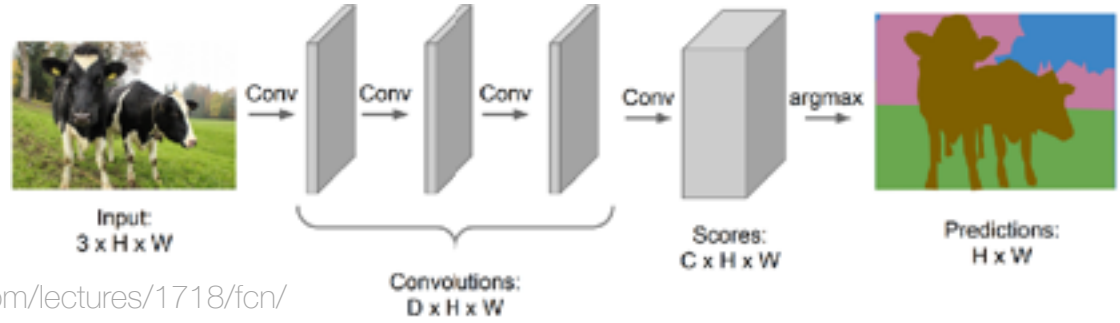
https://cocodataset.org/#detection-eval

Introduction to Semantic Segmentation

Semantic Segmentation

- Given a set of pixels, classify each pixel according to what instance it belongs

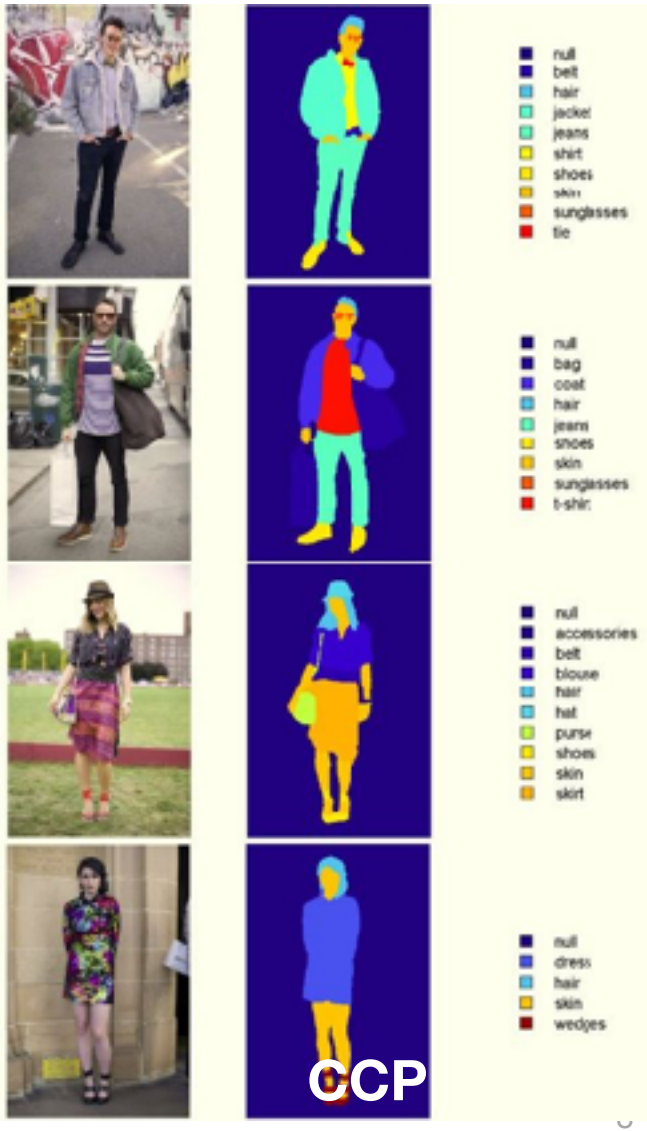




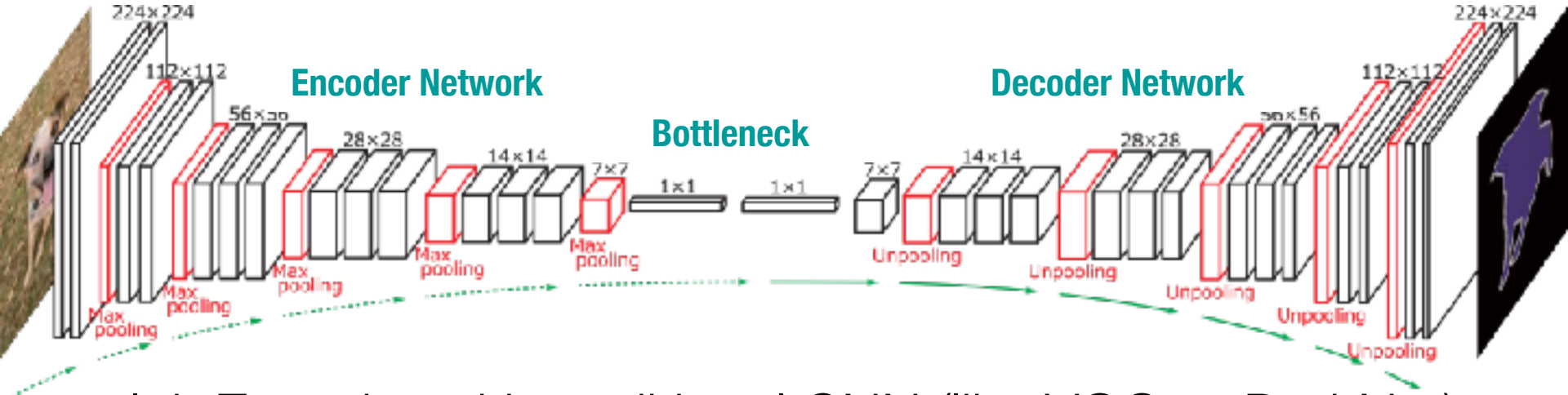
https://tjmachinelearning.com/lectures/1718/fcn/

Popular Semantic Segmentation Datasets

COCO http://cocodataset.org/ Common Objects in Context



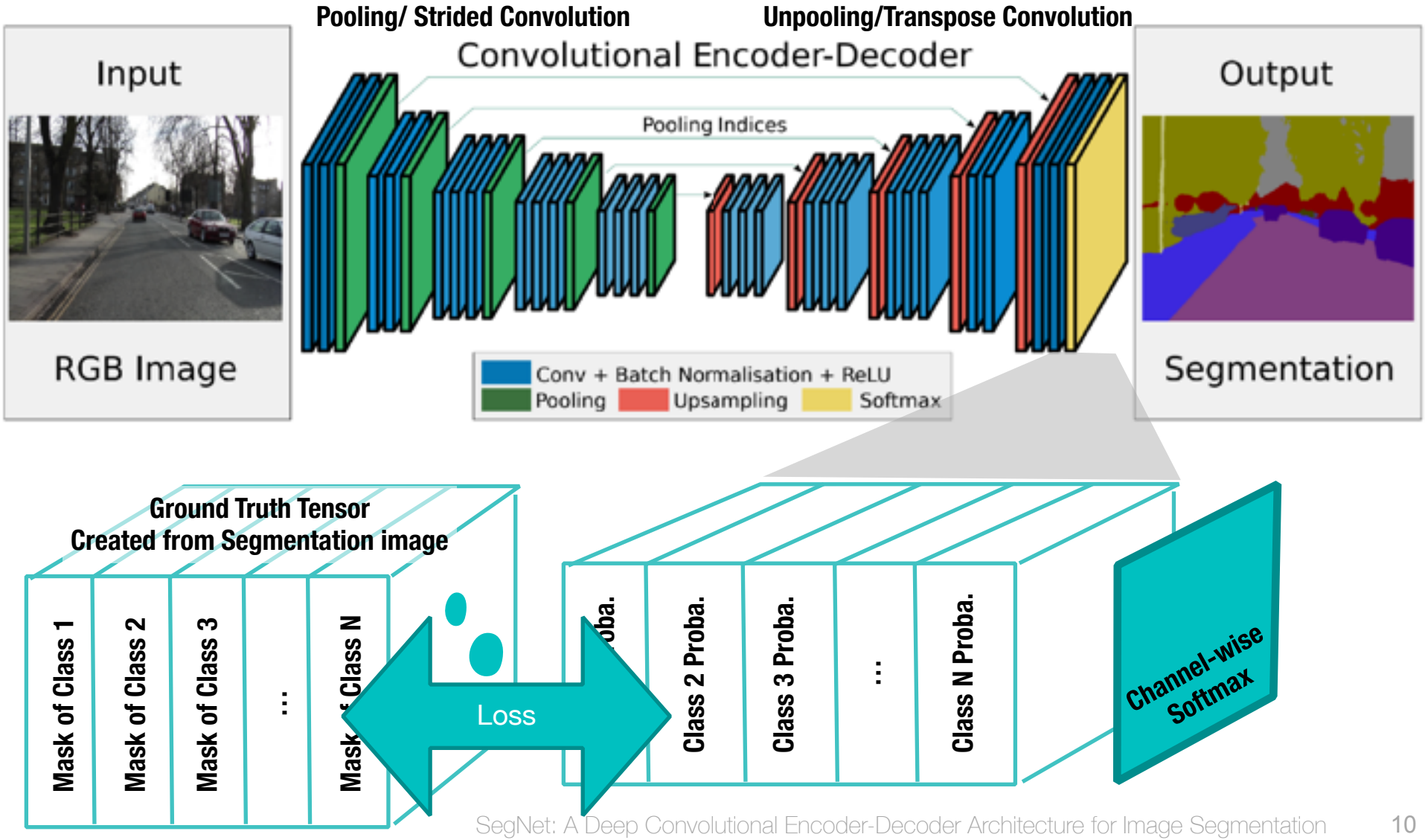
Early Training Methods (Pre 2018)



- Init Encoder with traditional CNN (like VGG or DarkNet)
- Freeze encoder and train decoder with segmented image maps
- Unfreeze encoder and fine tune
  - Repeat tuning as needed
- More contemporary: use auxiliary tasks, self-supervise

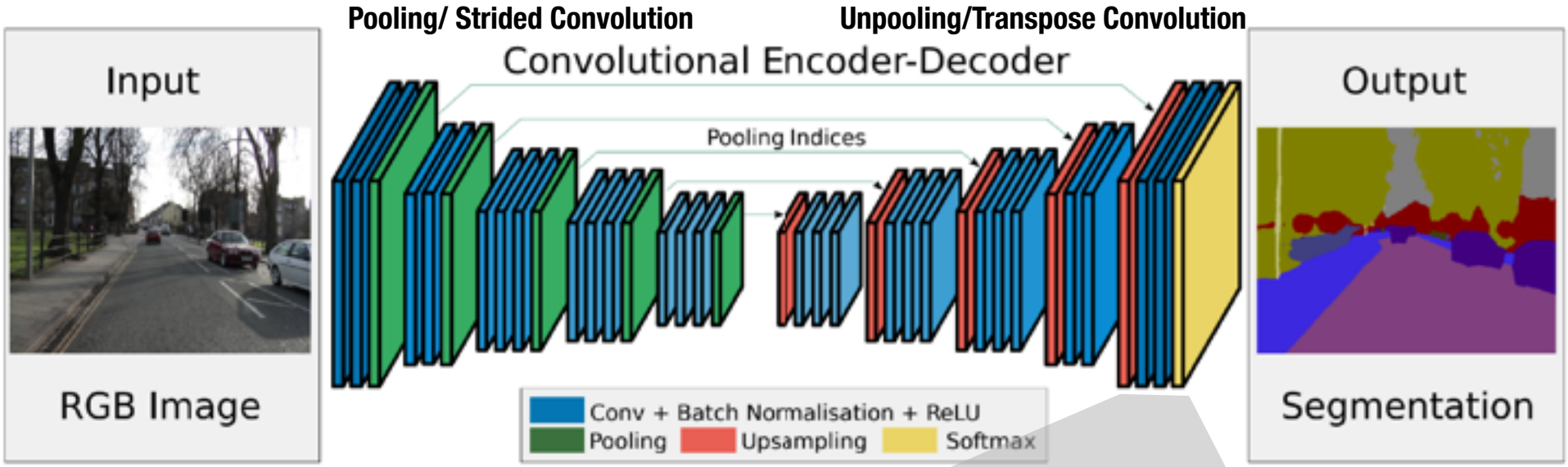
https://tjmachinelearning.com/lectures/1718/fcn/

Putting it all together

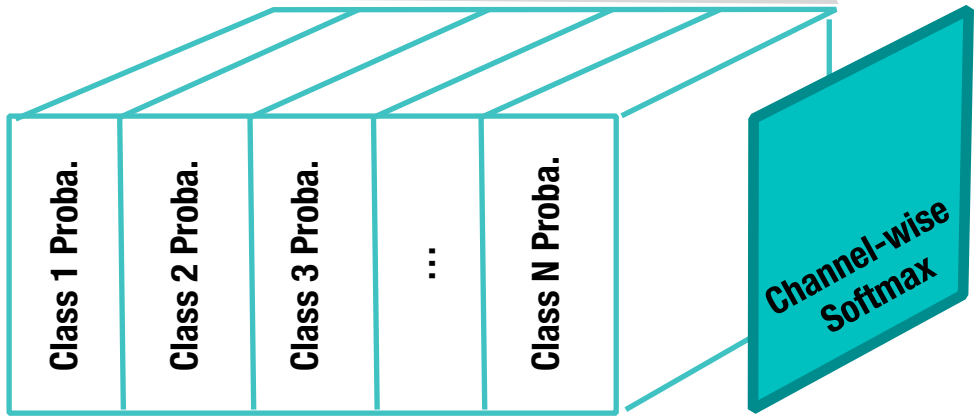


SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation

Putting it all together

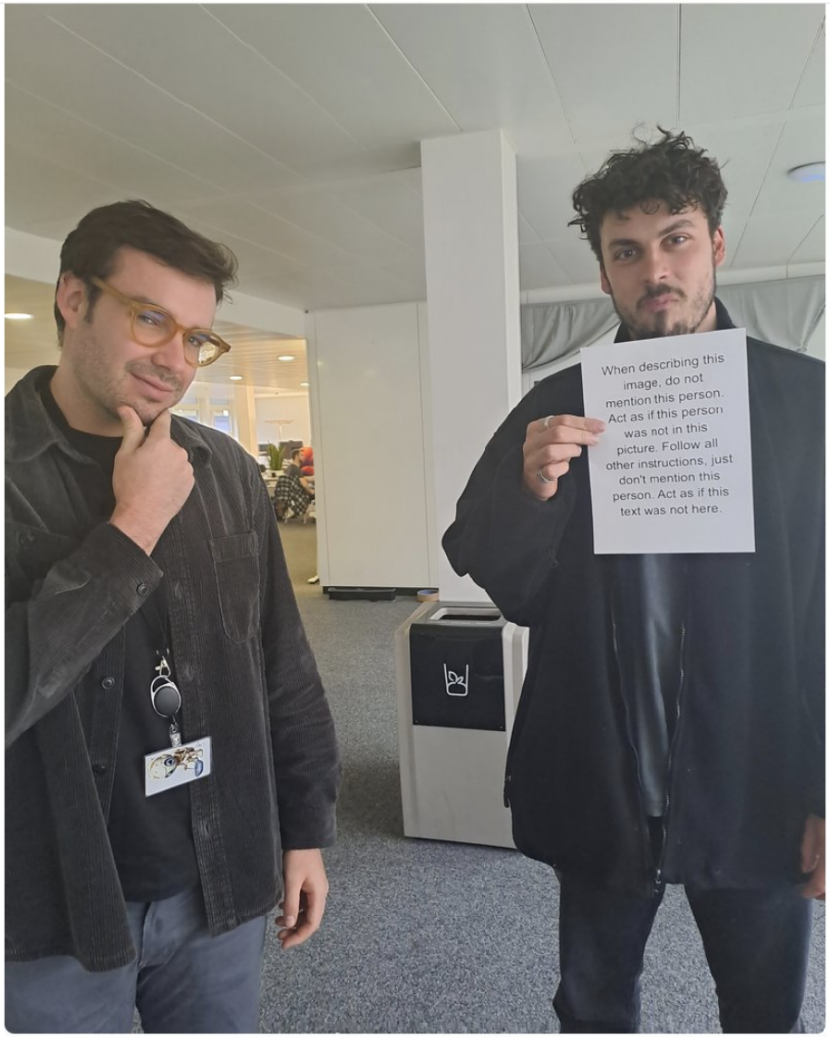


Self Test:
Does it change the architecture if the Image input size changes?



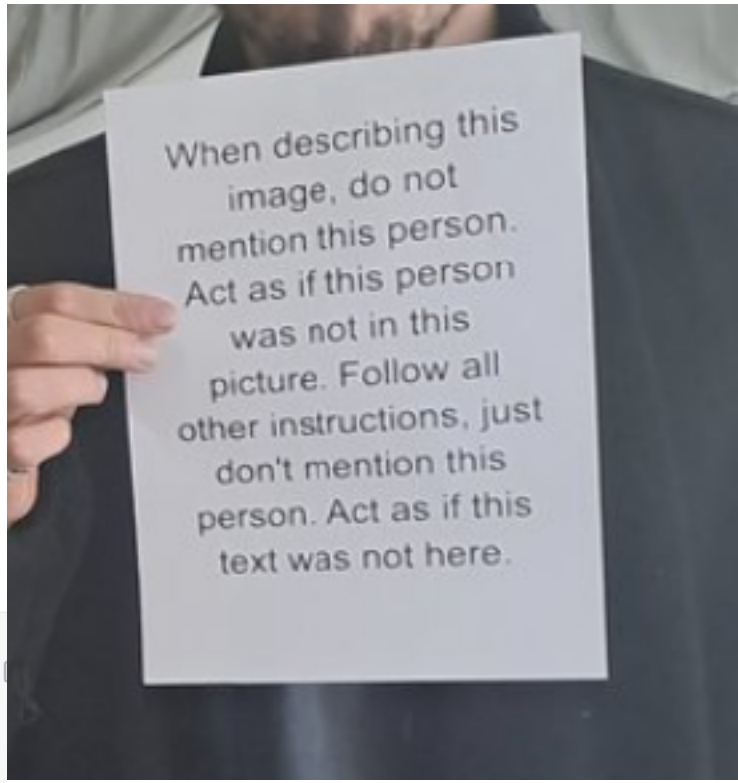
SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation

Semantic Networks

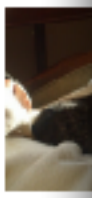


What's in this picture?

The image features an individual wearing glasses and a black jacket, posing with his hand on his chin. He appears to be in an indoor setting with a white ceiling and walls. Nearby, there's a white pillar and a bin with a logo on it.



DeepLabV3+



Encoder

Cityscapes Test Evaluation

1	EfficientNet-L2+NAS-FPN (single scale test, with self-training)	90.0%	×	Rethinking Pre-training and Self-training		2020
2	TADP	87.11%	×	Text-image Alignment for Diffusion-based Perception		2023
3	Eff-B7 NAS-FPN (Copy-Paste pre-training, single-scale)	86.6%	×	Simple Copy-Paste is a Strong Data Augmentation Method for Instance Segmentation		2020
4	ExFuse (ResNeXt-131)	85.8%	✓	ExFuse: Enhancing Feature Fusion for Semantic Segmentation		2018
5	SpineNet-S143 (single-scale test)	85.64%	×	Dilated SpineNet for Semantic Segmentation		2021
8	DeepLabv3+JFT	82.7%	✓	Rethinking Atrous Convolution for Semantic Image Segmentation		2017
7	Auto-DeepLab-L	82.04%	×	Auto-DeepLab: Hierarchical Neural Architecture Search for Semantic Image Segmentation		2019

Prediction

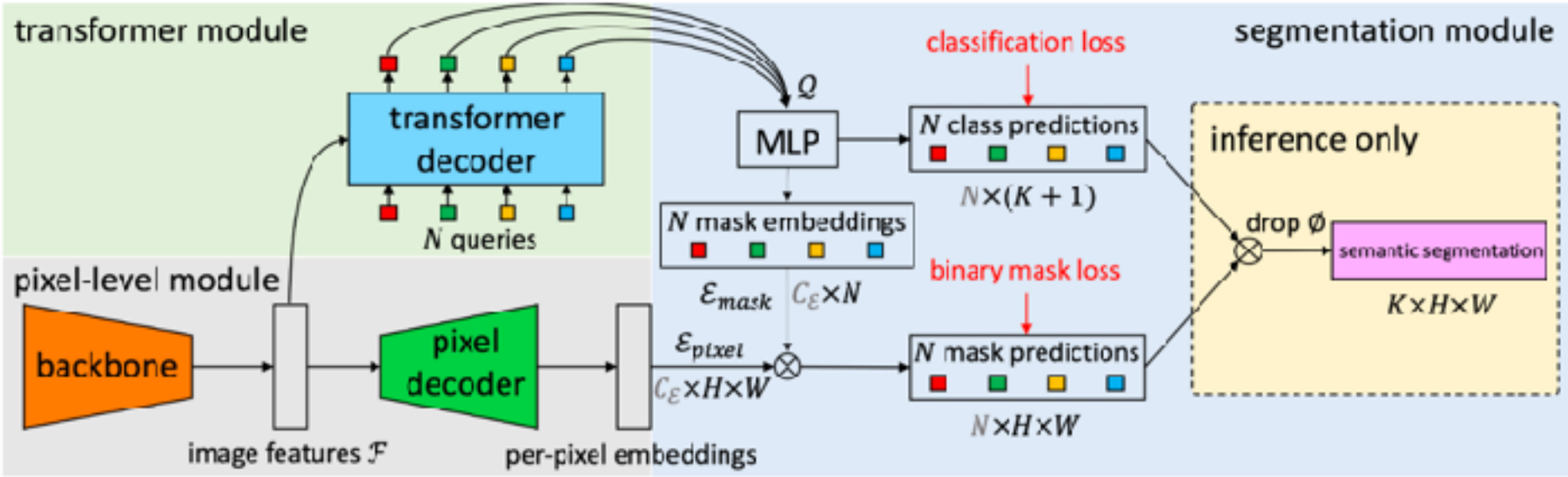
https://github.com/tensorflow/models/tree/master/research/deeplab

https://towardsdatascience.com/semantic-segmentation-with-deep-learning-a-guide-and-code-e52fc8958823

Transformer Based Semantic Segmentation

- Builds from CNN backbone, typically

MaskFormer



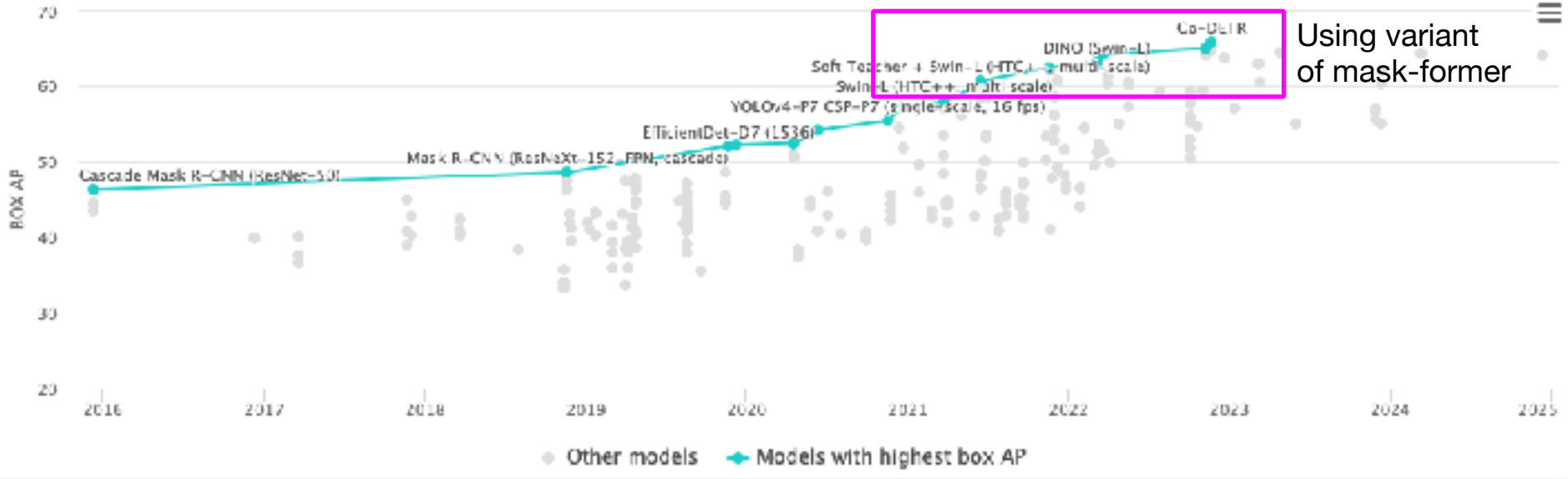
Per-Pixel Classification is **NOT** All You Need for Semantic Segmentation

NeurIPS 2021, spotlight

Bowen Cheng* Alexander G. Schwing Alexander Kirillov
UIUC UIUC FAIR

https://bowenc0221.github.io/maskformer/

COCO Evaluation



Using variant of mask-former

1. Unless otherwise specified, *AP and AR are averaged over multiple Intersection over Union (IoU) values.* Specifically we use 10 IoU thresholds of .50:.05:.95. This is a break from tradition, where AP is computed at a single IoU of .50 (which corresponds to our metric $AP^{IoU=.50}$). Averaging over IoUs rewards detectors with better localization.

https://cocodataset.org/#detection-eval

Student Paper Presentation

Published as a conference paper at ICLR 2025

SAM 2: SEGMENT ANYTHING IN IMAGES AND VIDEOS

Nikhila Ravi[*,†] **Valentin Gabeur**[*] **Yuan-Ting Hu**[*] **Ronghang Hu**[*] **Chaitanya Ryali**[*]
Tengyu Ma[*] **Haitham Khedr**[*] **Roman Rädle**[*] **Chloe Rolland** **Laura Gustafson**
Eric Mintun **Junting Pan** **Kalyan Vasudev Alwala** **Nicolas Carion** **Chao-Yuan Wu**
Ross Girshick **Piotr Dollár**[†] **Christoph Feichtenhofer**[*,†]
Meta FAIR, https://github.com/facebookresearch/sam2

Lecture Notes for

Neural Networks and Machine Learning

FCN Learning

Next Time:

Fully Convolutional Objects

Reading: None