Master in
Computer Vision
Barcelona

# Week 1:
# Semantic and Instance Segmentation 1
# Visual Recognition

Issam Laradji
issam.laradji@servicenow.com

March 8, 2023

# About me

- **Issam Laradji**, a Research Scientist at **ServiceNow Research**
  - Montreal, Canada
- Adjunct Professor at the **University of British Columbia**
  - Vancouver, Canada

**History**

- Postdoc at McGill and MILA (2020-2021)
- PhD at University of British Columbia (2014-2020)


# About me

## Research interests

- **Low data learning methods,**
  - Few-shot learning, Active learning, Semi-supervised learning
- **Computer vision**
  - Segmentation, detection, counting, VQA
- **Natural language processing**
  - Chat Dialogues, Slot Filling, and Summarization

servicenow®

# About me

**Volunteered to teach two lectures for this class**

- Semantic and Instance Segmentation 1 (March 8th)
- Semantic and Instance Segmentation 2 (March 15th)
- Exam Questions (May 8th)

**Master/PhD Thesis supervision and internships**

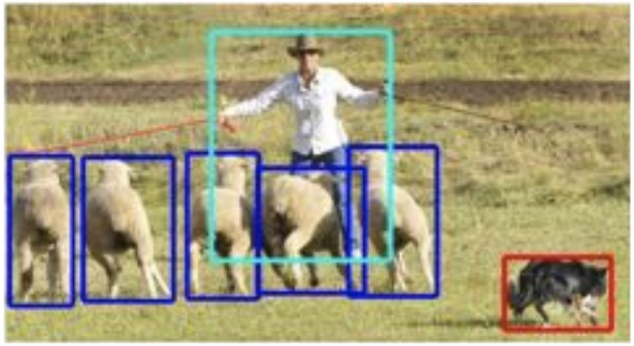
  - Contact me in case you are interested

servicenow®

# Spanish Researchers in Canada

# Computer Vision Tasks


(a) image classification

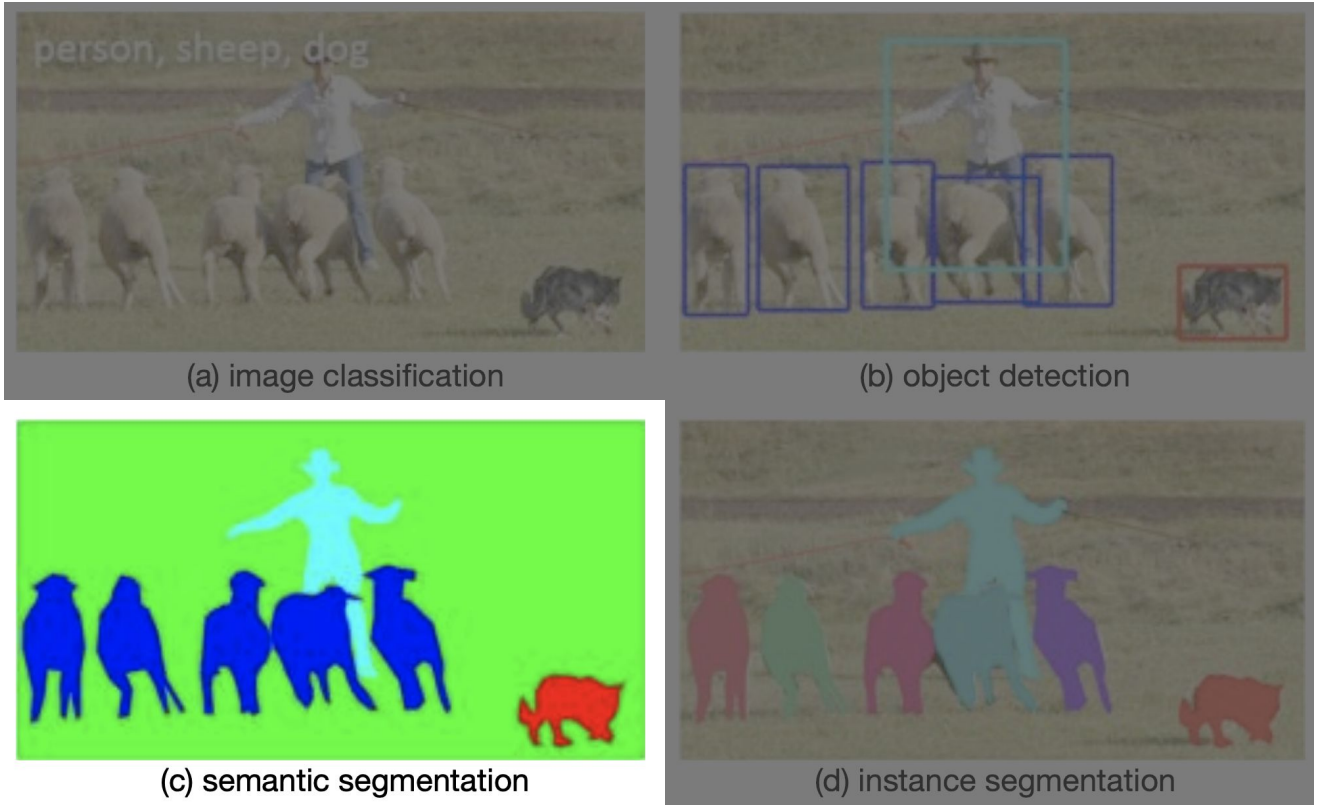


(b) object detection

(c) semantic segmentation

(d) instance segmentation

# Semantic Segmentation

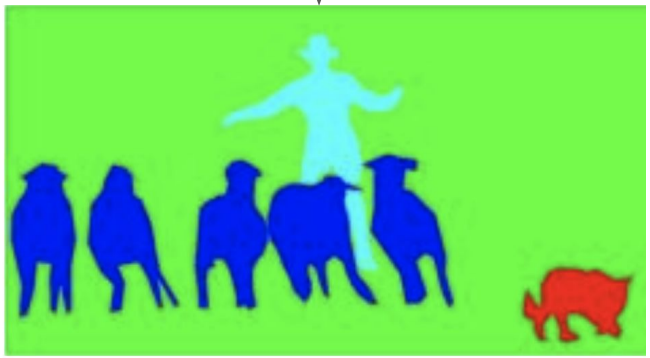


# Semantic segmentation: Problem statement

- Give a semantic label to every pixel in an image

- Given a set of training examples, can we predict the segmentation labels of an unseen example?

Training Set

Test Set

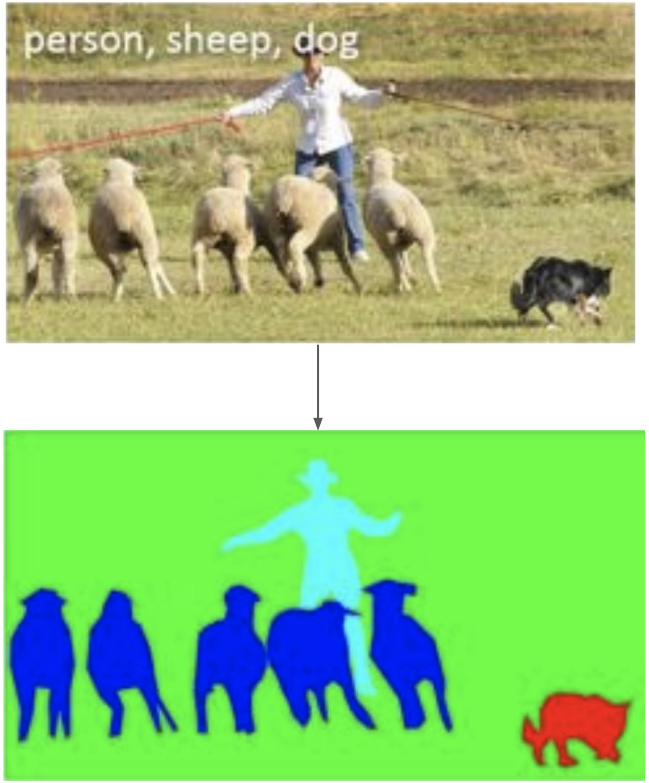
?

# Semantic segmentation: Problem statement

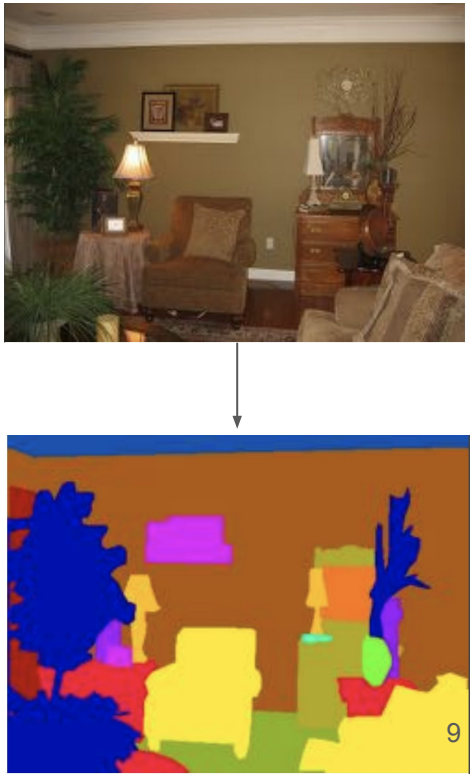
- Give a semantic label to every pixel in an image

- Given a set of training examples, can we predict the segmentation labels of an unseen example?

Training Set




Test Set



# Semantic segmentation: Problem statement

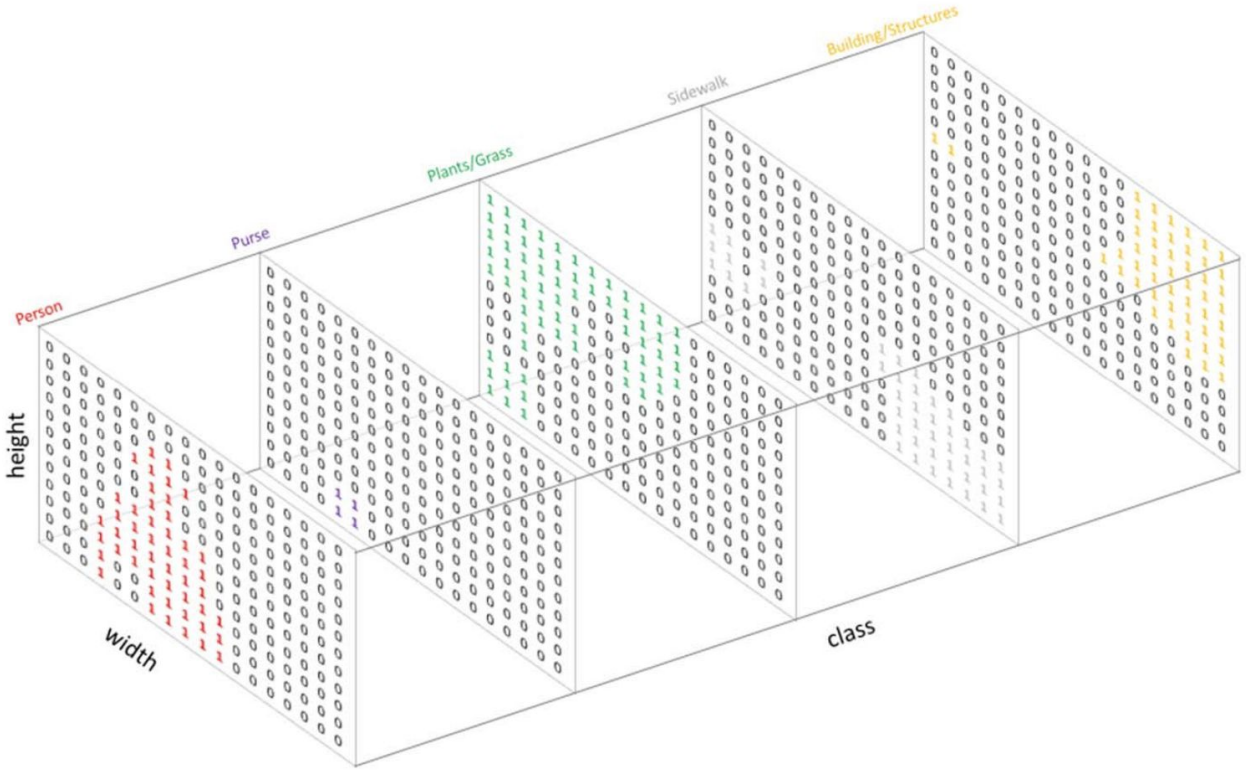
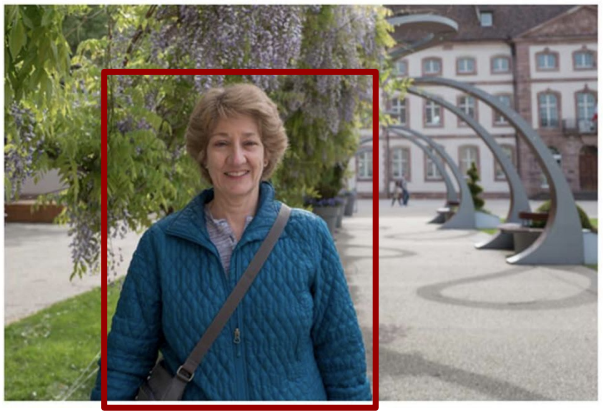



*Image source: https://www.jeremyjordan.me/semantic-segmentation/*

# Semantic segmentation: Problem statement



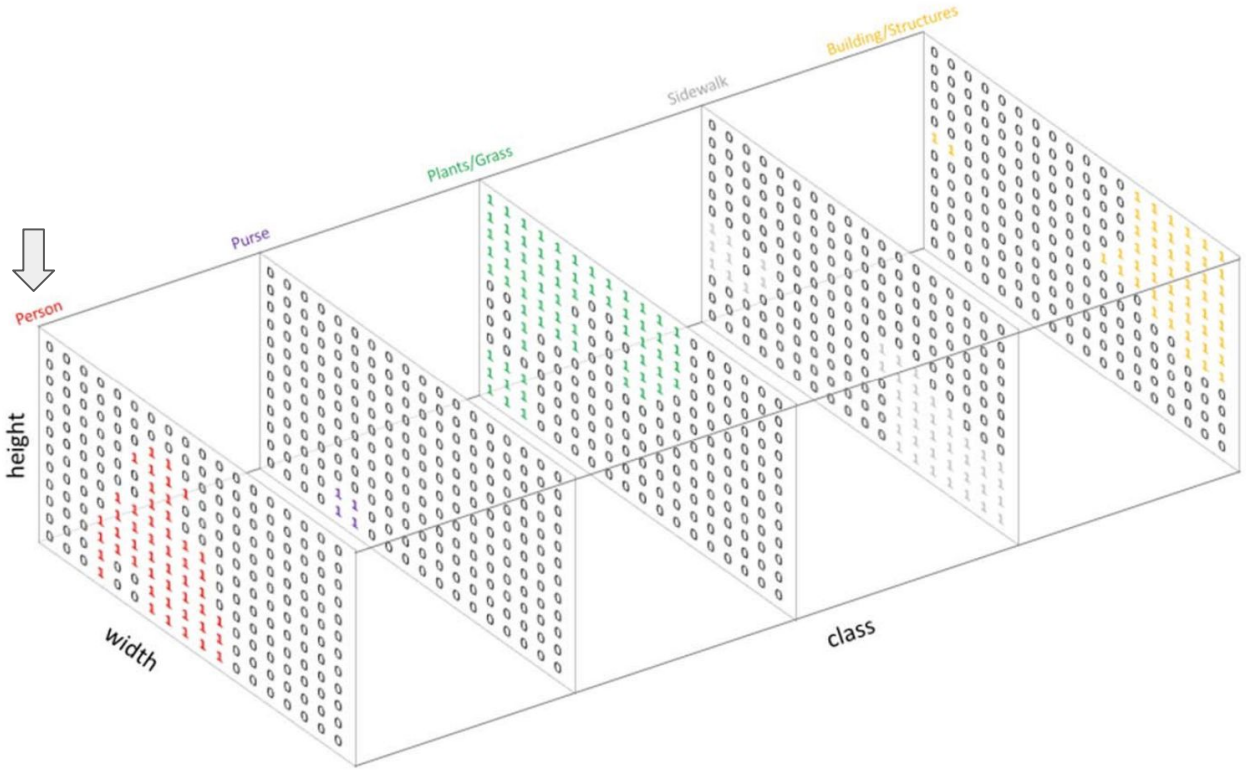



*Image source: https://www.jeremyjordan.me/semantic-segmentation/*

# Semantic segmentation: Problem statement

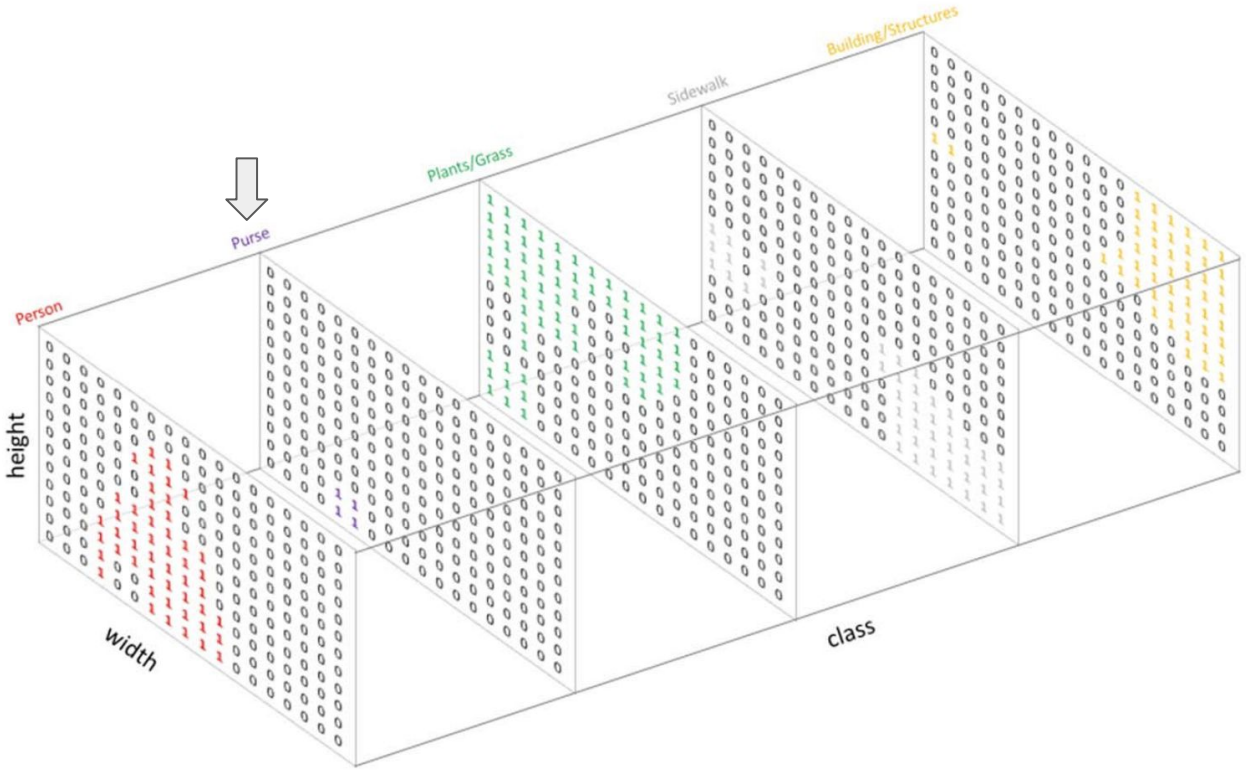
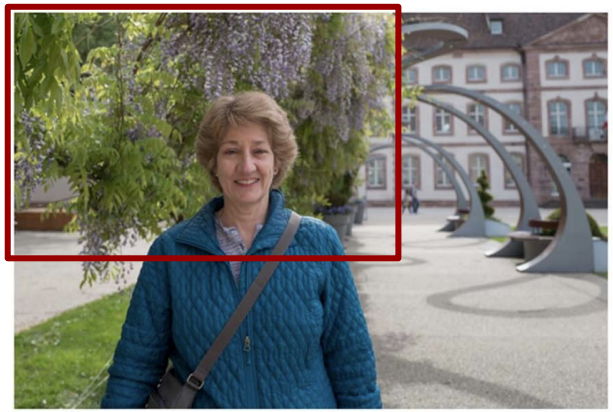



*Image source: https://www.jeremyjordan.me/semantic-segmentation/*

# Semantic segmentation: Problem statement



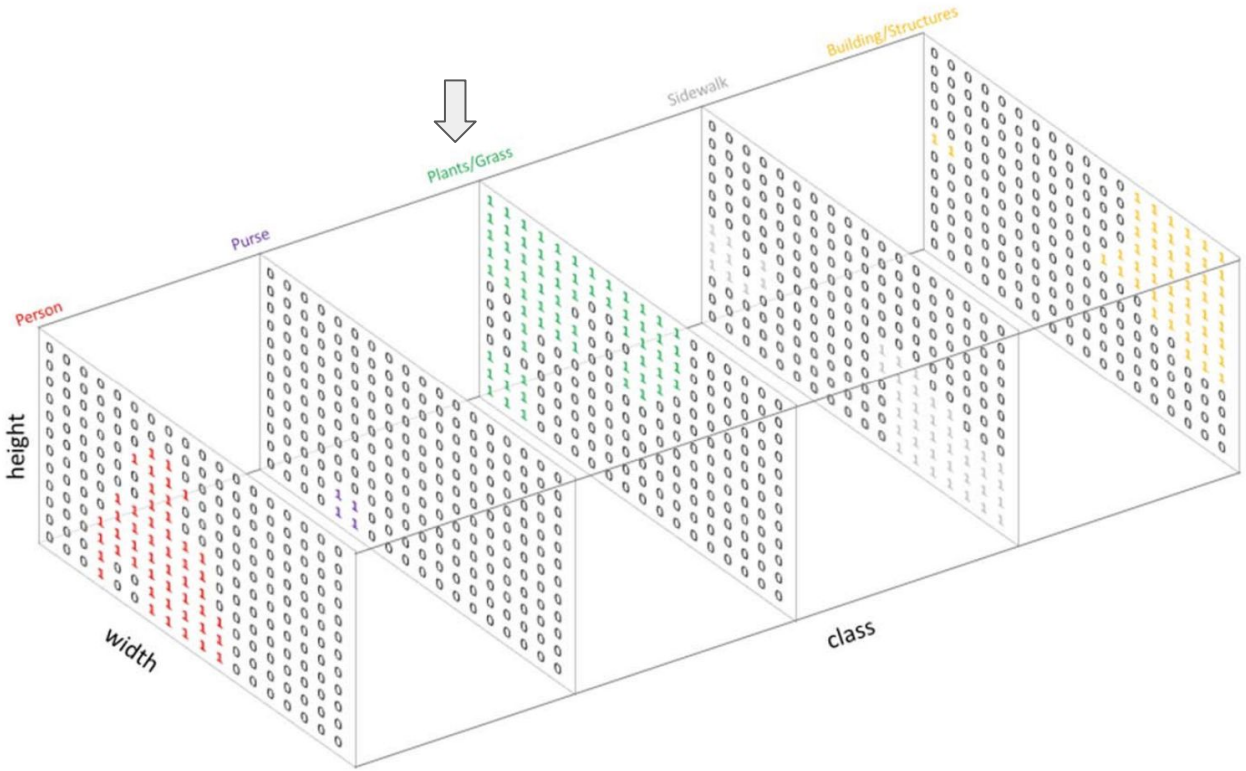



*Image source: https://www.jeremyjordan.me/semantic-segmentation/*

# Semantic segmentation: Problem statement

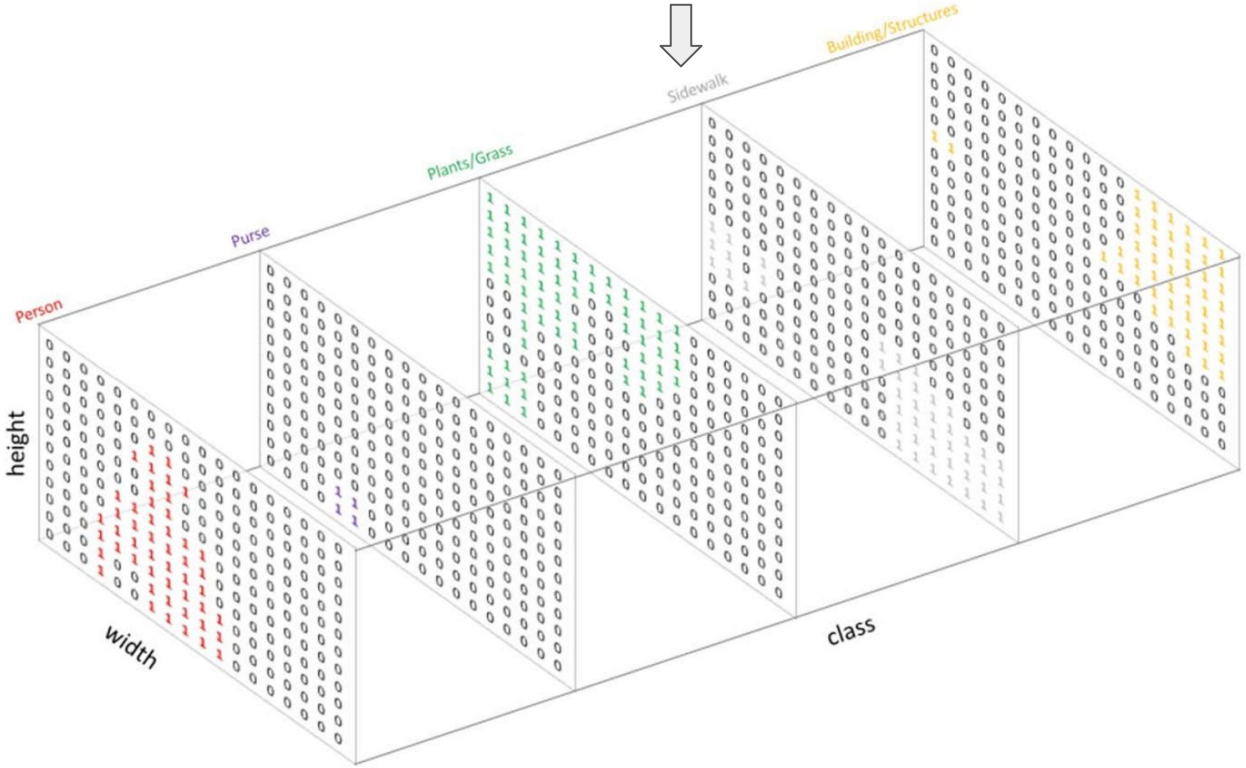



*Image source: https://www.jeremyjordan.me/semantic-segmentation/*

# Semantic segmentation: Problem statement

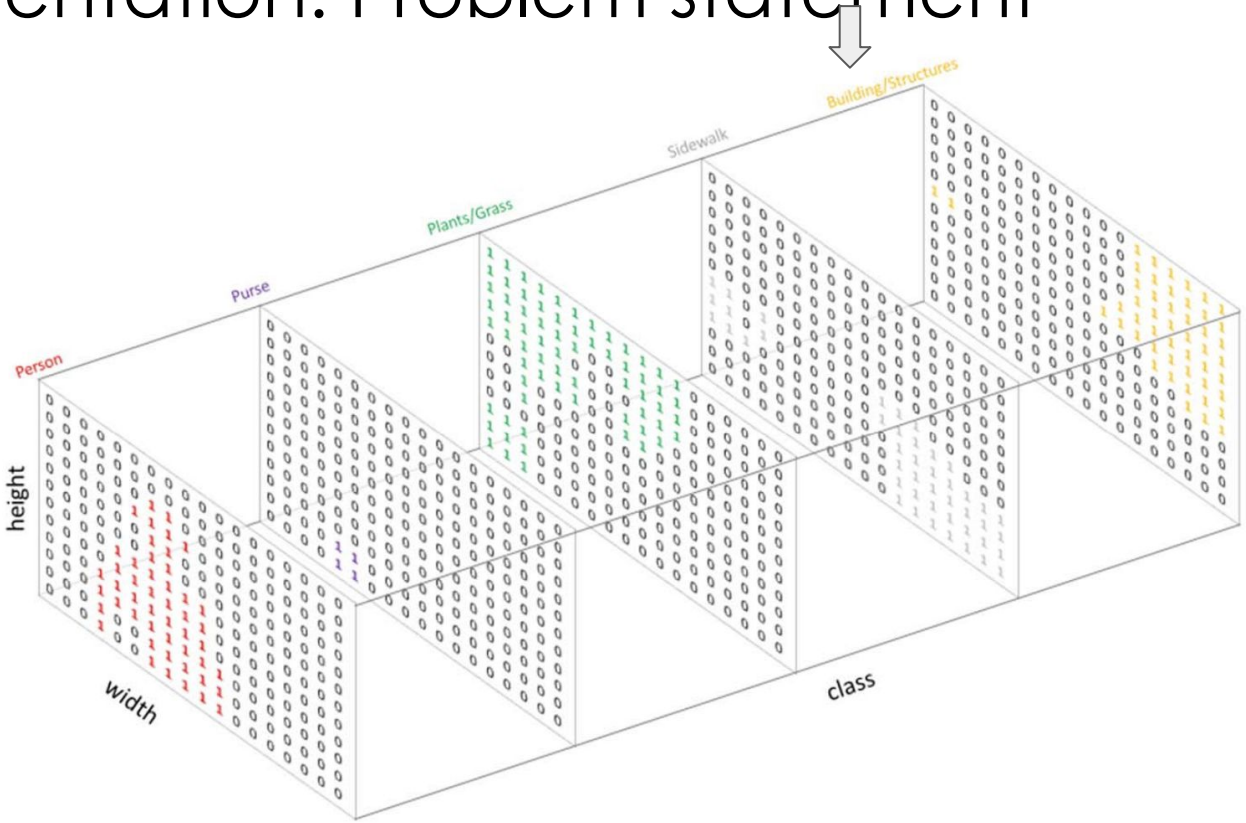



*Image source: https://www.jeremyjordan.me/semantic-segmentation/*

# Semantic segmentation: Problem statement

segmented

1: Person
2: Purse
3: Plants/Grass
4: Sidewalk
5: Building/Structures

| | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 5 | 5 | 5 | 5 | 5 | 5 |
| 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 5 | 5 | 5 | 5 | 5 | 5 |
| 3 | 3 | 3 | 3 | 3 | 3 | 1 | 1 | 3 | 3 | 3 | 3 | 5 | 5 | 5 | 5 | 5 | 5 |
| 3 | 3 | 3 | 3 | 3 | 1 | 1 | 1 | 1 | 3 | 3 | 3 | 5 | 5 | 5 | 5 | 5 | 5 |
| 3 | 3 | 3 | 3 | 3 | 3 | 1 | 1 | 3 | 3 | 3 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |
| 5 | 5 | 3 | 3 | 3 | 3 | 1 | 1 | 3 | 3 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |
| 4 | 4 | 3 | 4 | 1 | 1 | 1 | 1 | 1 | 1 | 4 | 4 | 4 | 5 | 5 | 5 | 5 | 5 |
| 4 | 4 | 3 | 4 | 1 | 1 | 1 | 1 | 1 | 1 | 4 | 4 | 4 | 4 | 4 | 5 | 5 | 5 |
| 4 | 4 | 4 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| 3 | 3 | 3 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| 3 | 3 | 3 | 1 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| 3 | 3 | 3 | 1 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |

Input

Semantic Labels

*Image source: https://www.jeremyjordan.me/semantic-segmentation/*

# Semantic segmentation: Applications

- Autonomous driving

*Image source: https://medium.com/intro-to-artificial-intelligence/semantic-segmentation-udaitys-self-driving-car-engineer-nanodegree-c01eb6eaf9d*

# Semantic segmentation: Applications

- Medical imaging

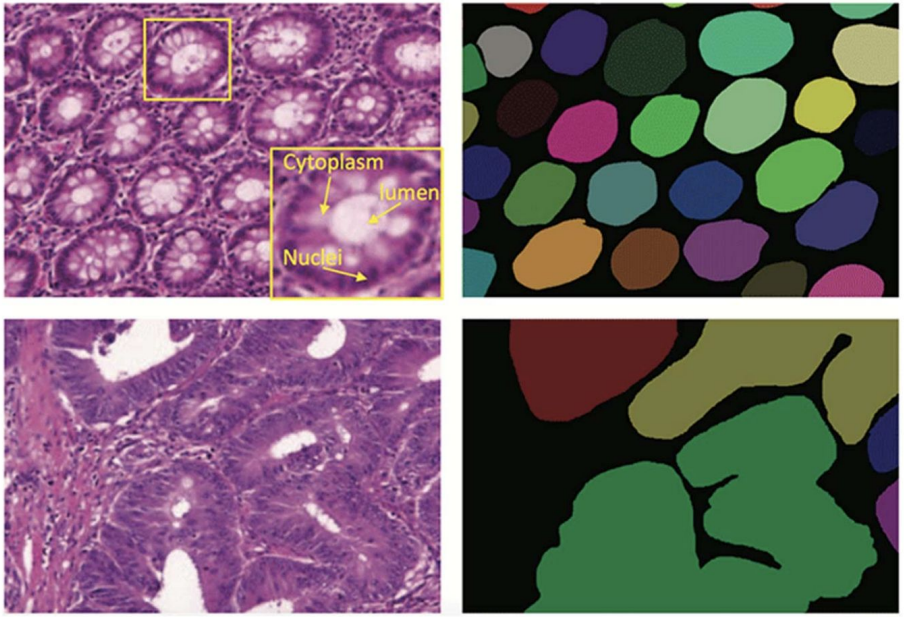



*Image source: NVIDIA Blog: GPU-Powered Deep Learning Being Used to Speed Colon Cancer Diagnosis*

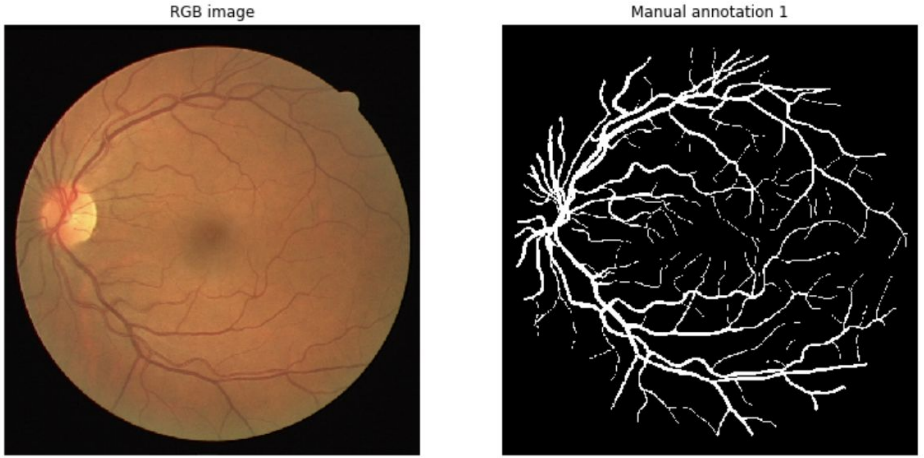
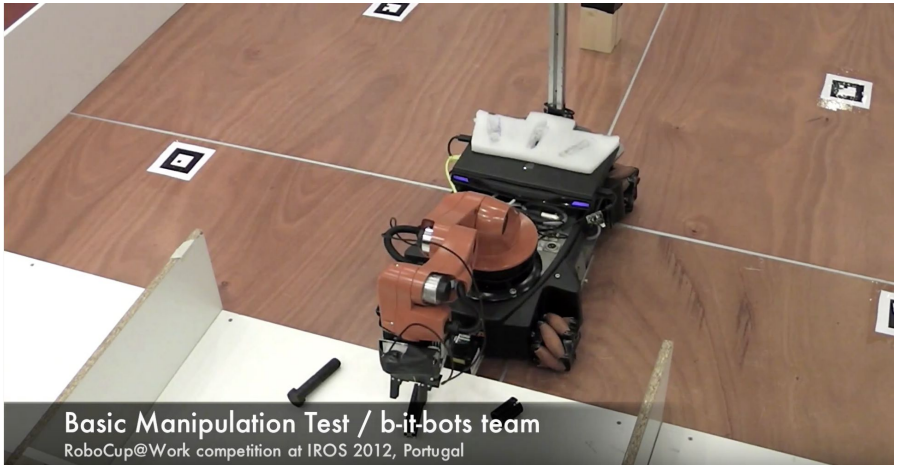
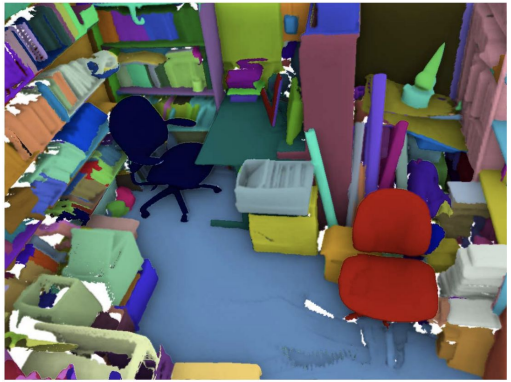



Image source: DRIVE Digital Retinal Image Vessel Extraction

# Semantic segmentation: Applications

- Robotic applications & Scene Understanding



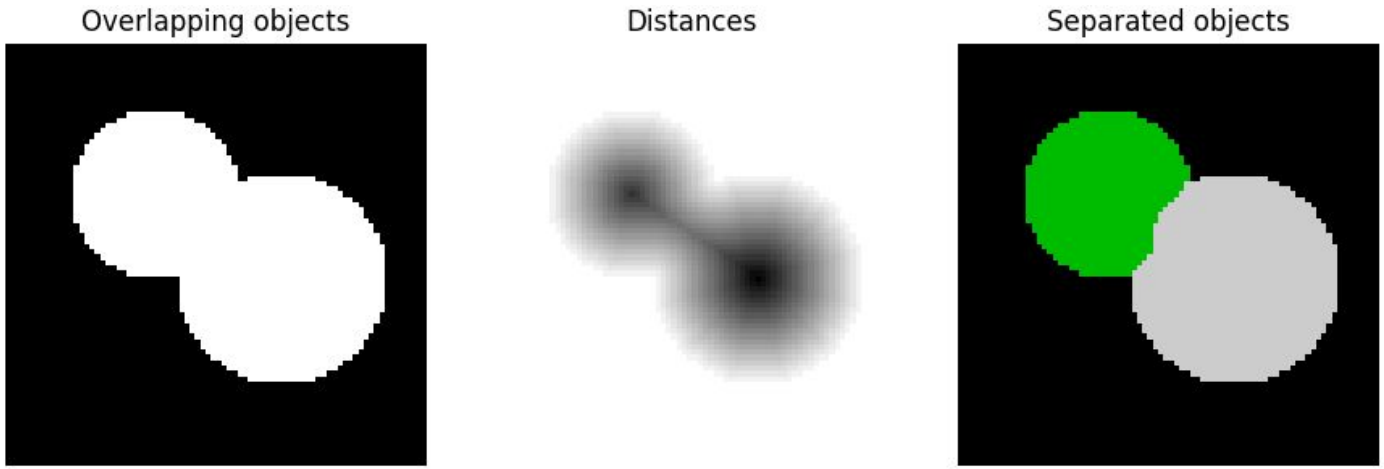



# Semantic segmentation: Traditional Approaches

- Thresholding

# Semantic segmentation: Traditional Approaches

- Region growing methods like the Watershed Algorithm

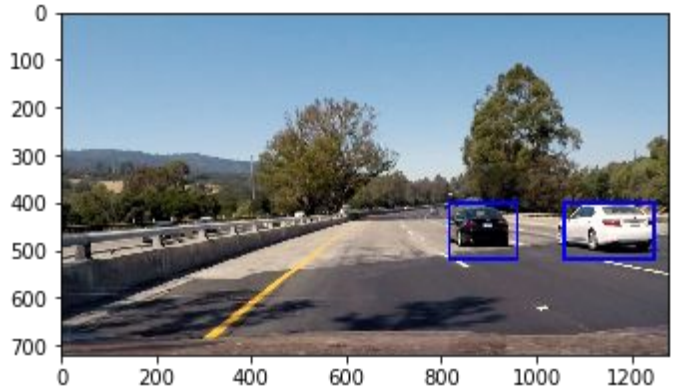


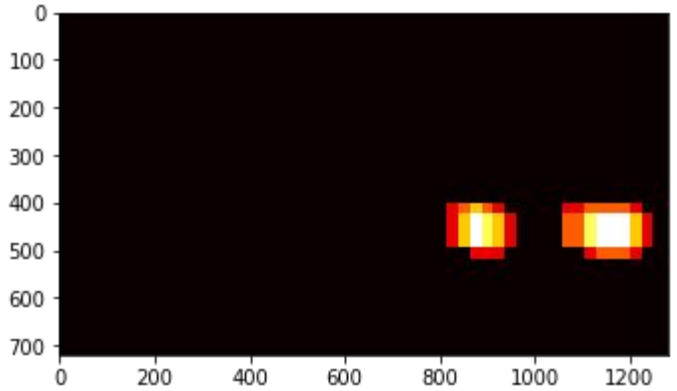
# Semantic segmentation: Superpixel segmentation

- SLIC - K-Means based image segmentation
  - Generates superpixels by clustering
    - color similarity and proximity in the image plane.

# Semantic segmentation: Superpixel segmentation

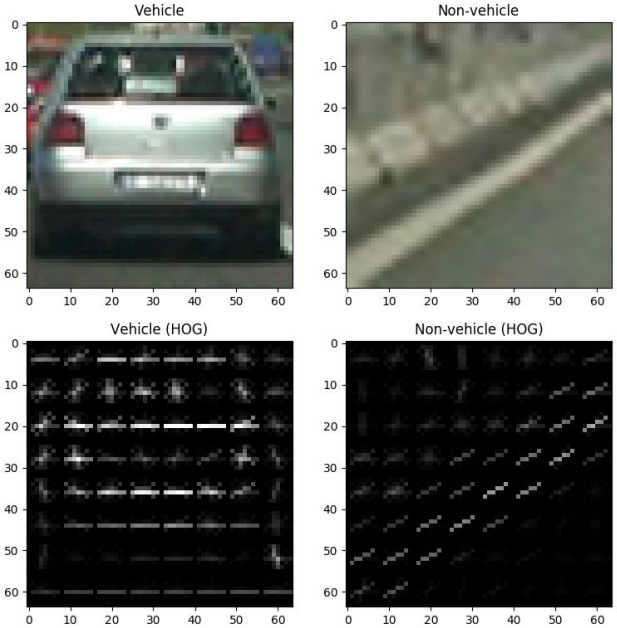
- Hog Features + SVM



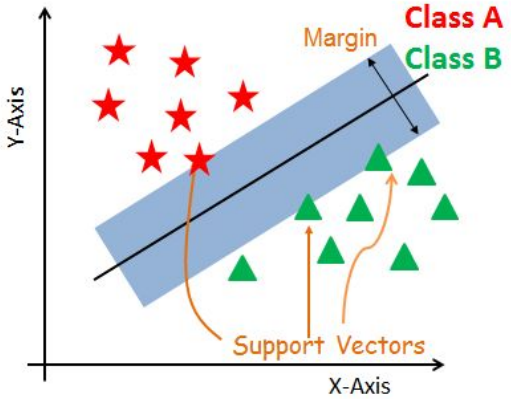



# Semantic segmentation: Superpixel segmentation

- Hog Features + SVM




Hog Features




SVM

# Semantic segmentation: Traditional Approaches

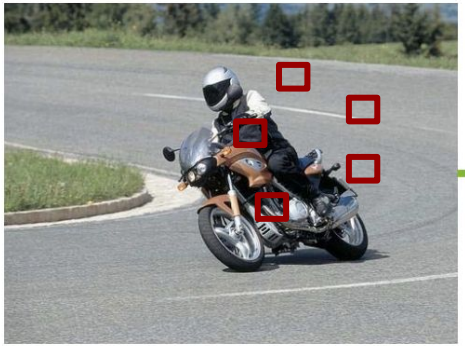
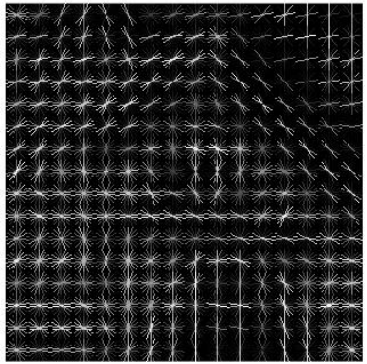
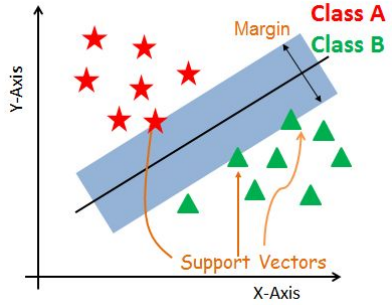


*Image source: PASCAL VOC 2012*

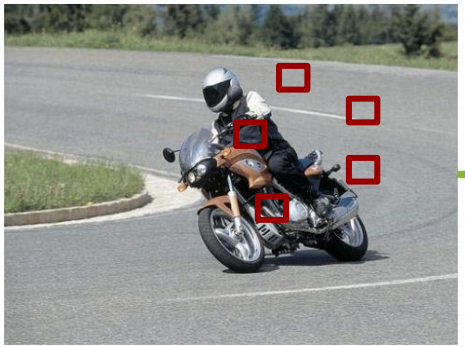
Pre-segmentation
+
Hand-crafted Descriptors



Local Classification
+
Global Context



# Semantic segmentation: Traditional Approaches



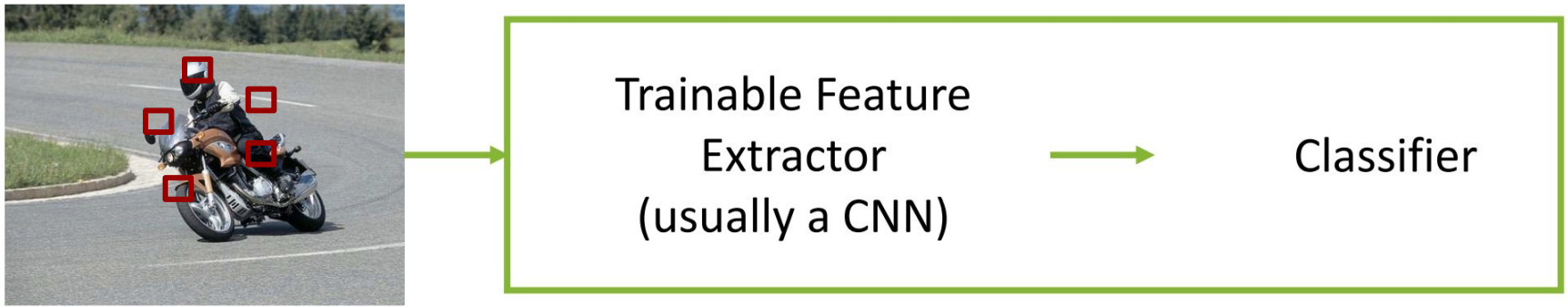
Pre-segmentation
+
Hand-crafted Descriptors

Local Classification
+
Global Context

*Image source: PASCAL VOC 2012*

- Drawbacks:
  - Pre-segmentation methods are not perfect and are hard to tune
  - Hand-crafted descriptors
  - Global context integration (e.g. CRF) is computationally expensive

# Semantic segmentation: Traditional approaches




- Pre-segmentation and description are replaced by a trainable feature extractor
- The classifier is trained jointly with the previous stage
- **… but how do we move from classification to segmentation?**

# Semantic segmentation: traditional approaches

- Intermediate solution: Patch-based + Sliding Window approach

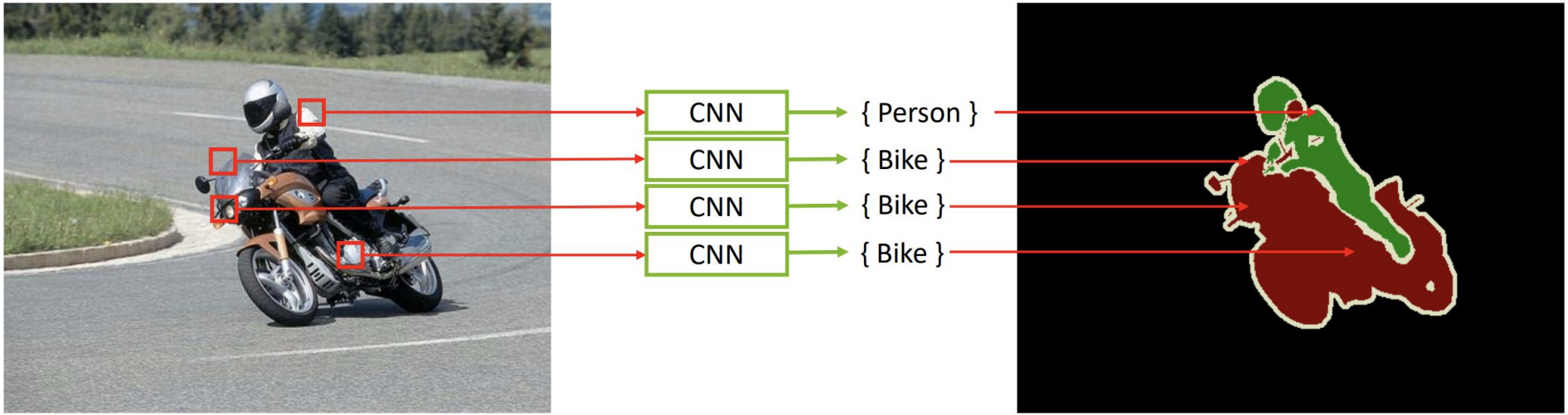
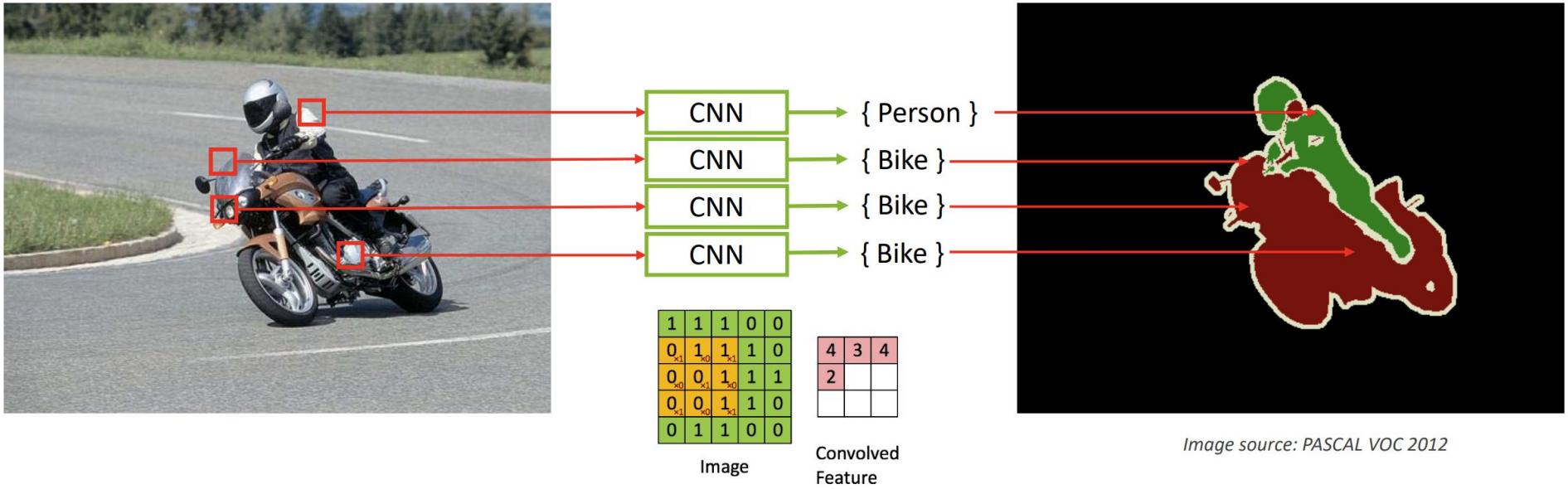



*Image source: PASCAL VOC 2012*

# Semantic segmentation: traditional approaches

- Intermediate solution: Patch-based + Sliding Window approach



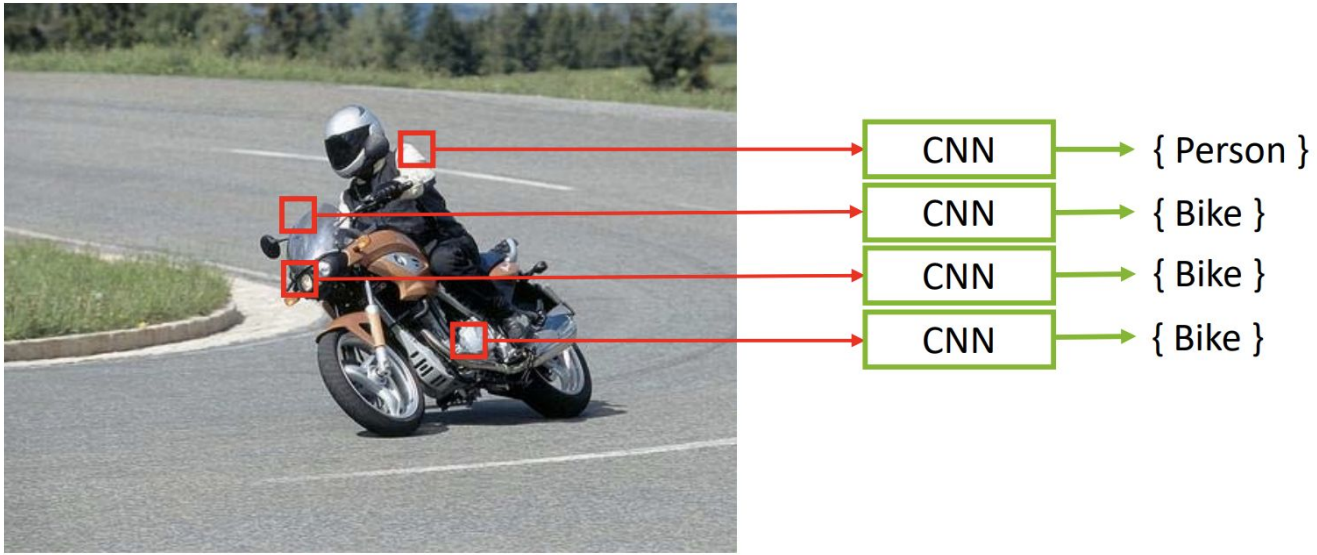
# Semantic segmentation: Traditional Approaches

- Intermediate solution: Patch-based + Sliding Window approach

Basic and intuitive

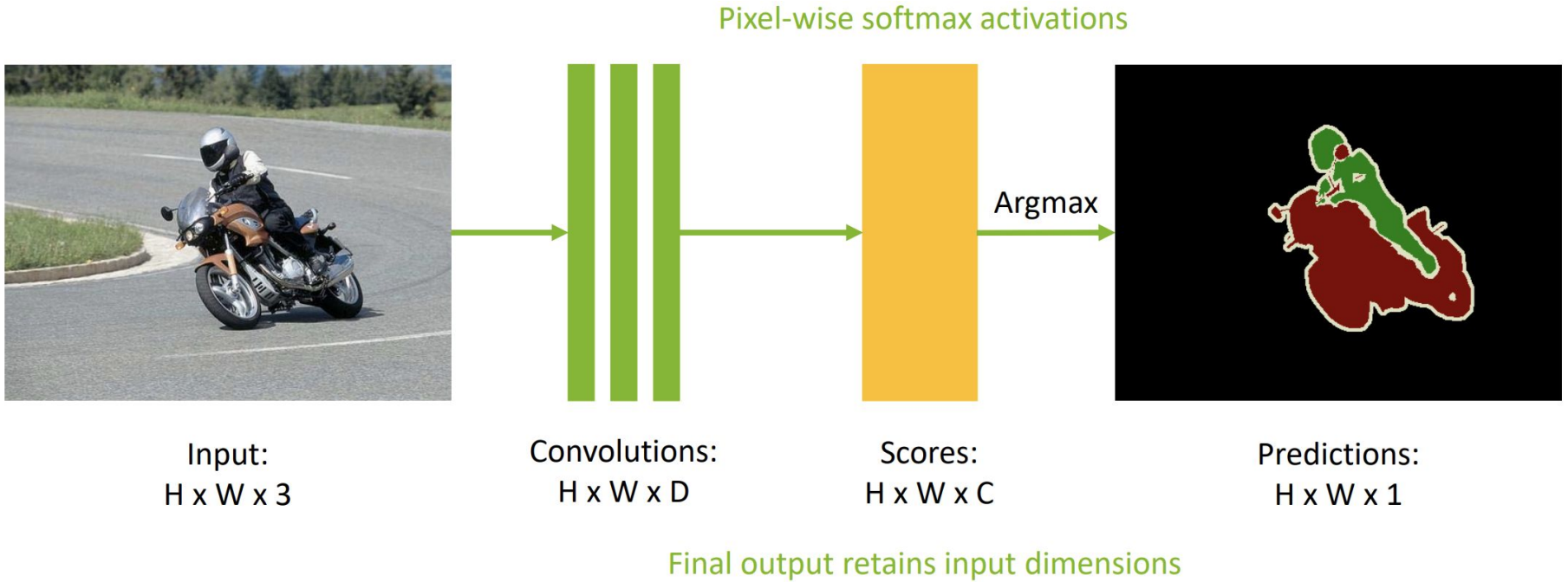
Cheap (memory)

Inefficient training does not reuse shared features between overlapping patches!



# Semantic segmentation: Traditional Approaches

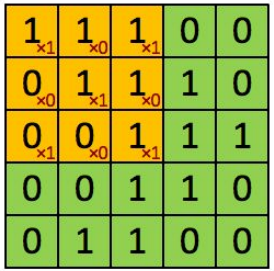
- Intermediate solution: Fully Convolutional



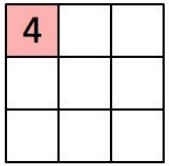
# Semantic segmentation: Traditional Approaches

- Intermediate solution: Fully Convolutional

Input:
H x W x 3




Image




Convolved
Feature

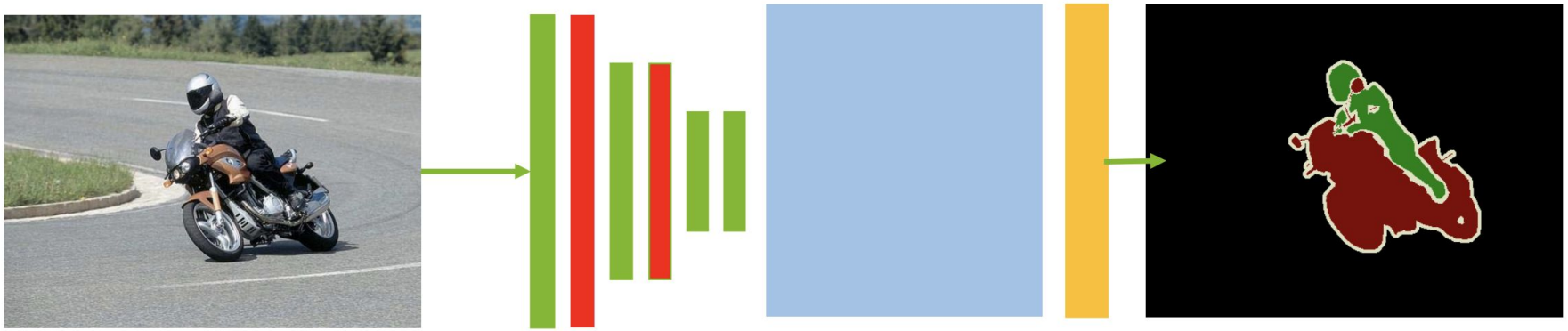
# Semantic segmentation: Traditional Approaches

- Intermediate solution: Fully Convolutional


Convolutions at original image resolution are expensive and impractical in most cases!

# Semantic segmentation: Traditional Approaches

- Intermediate solution: Fully Convolutional




Downsampling with pooling and convolutions renders the network more efficient!

But how do we implement upsampling?

# Semantic segmentation: Traditional Approaches

- Intermediate solution: Fully Convolutional
  - Bilinear interpolation

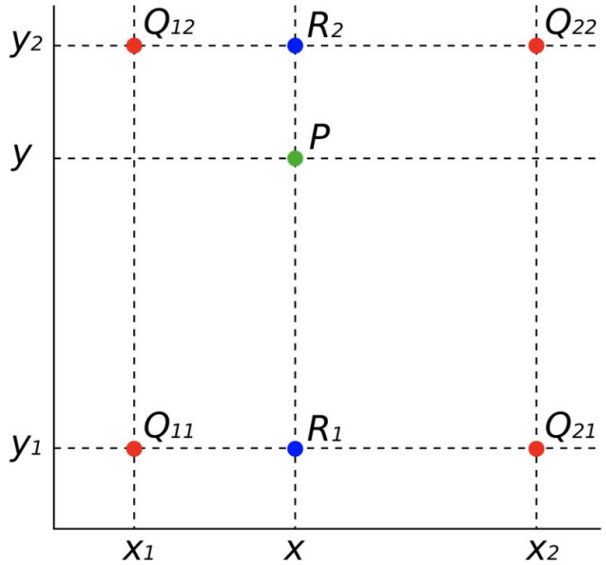



*Image source: Wikipedia*

# Semantic segmentation: Traditional Approaches

- Intermediate solution: Fully Convolutional
  - Unpooling

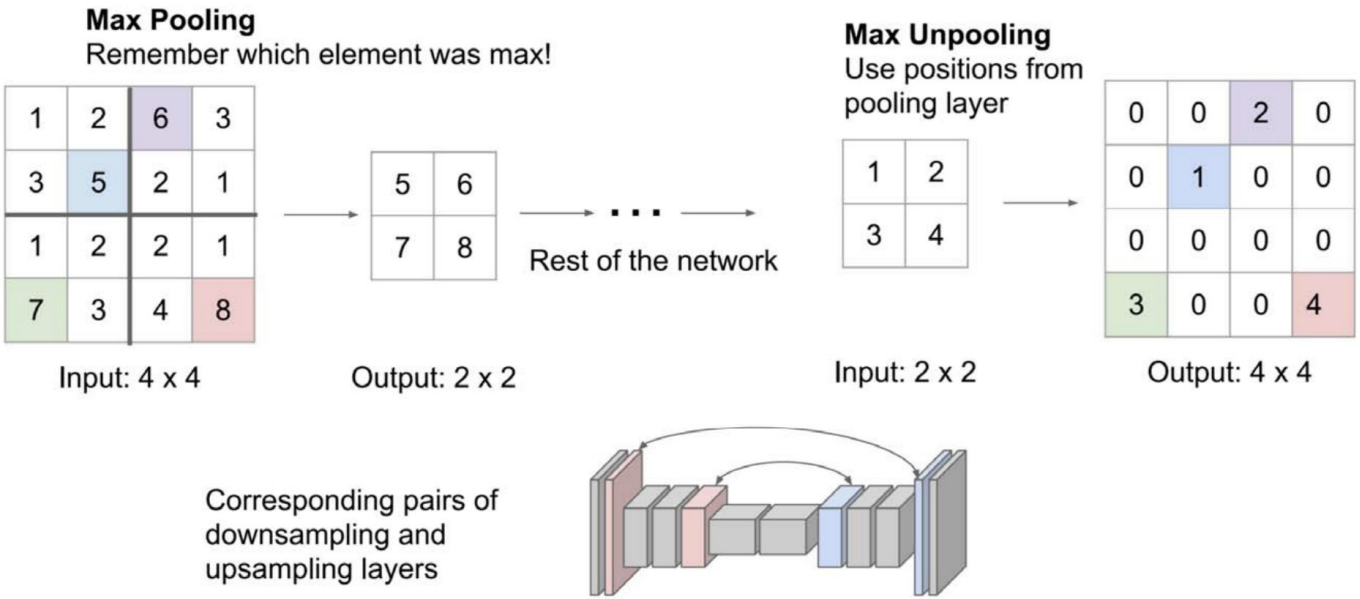
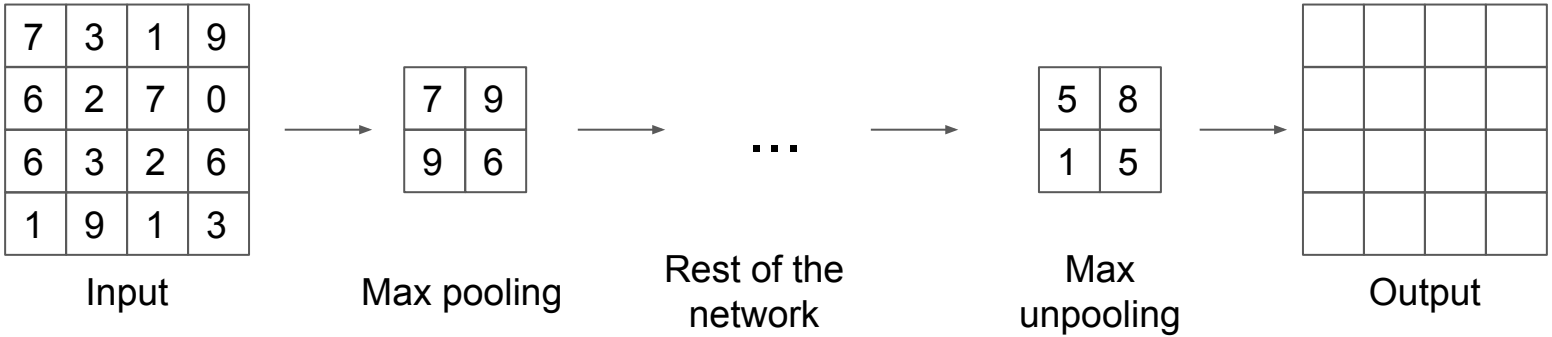
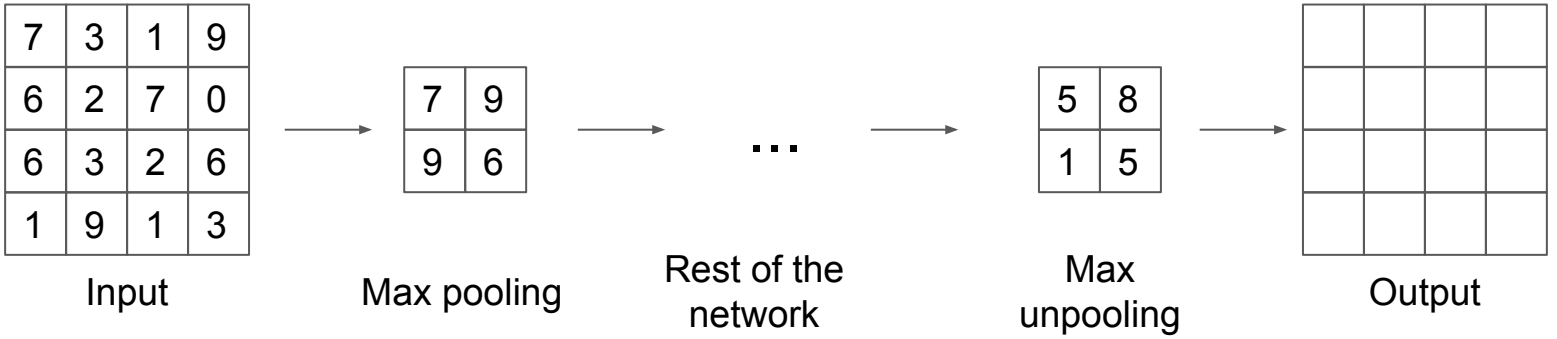


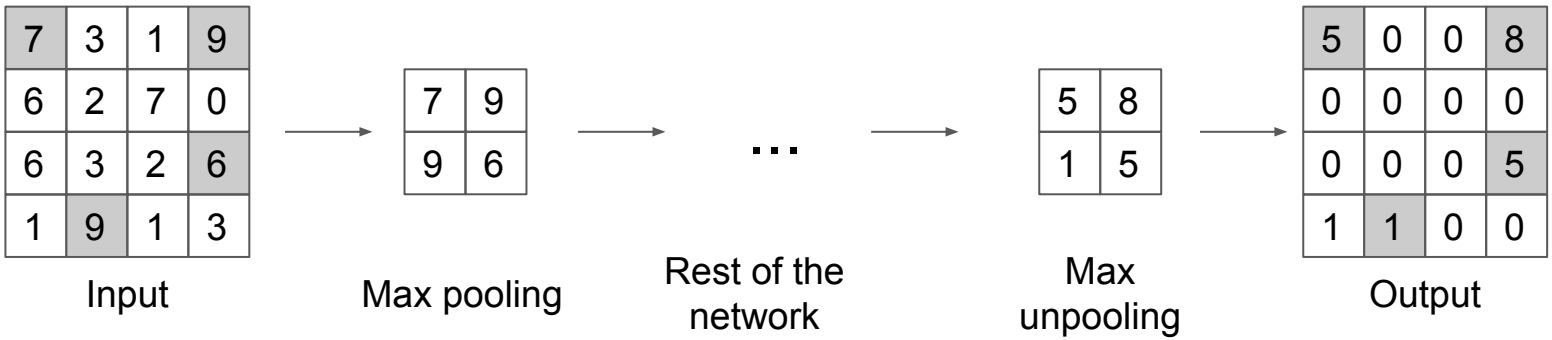

*Image source: Stanford CS231 Lecture 11*

# Exercise 2021

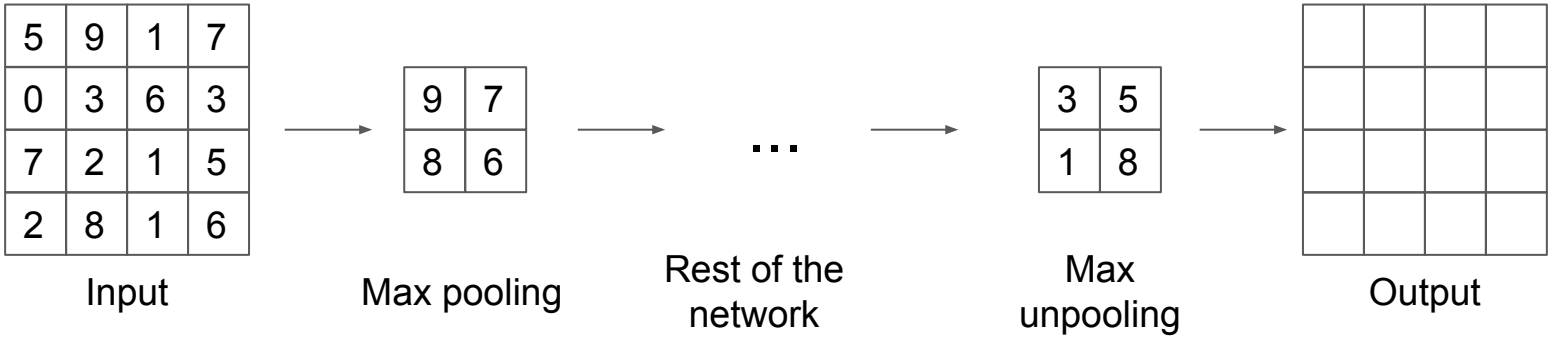


# Exercise 2021

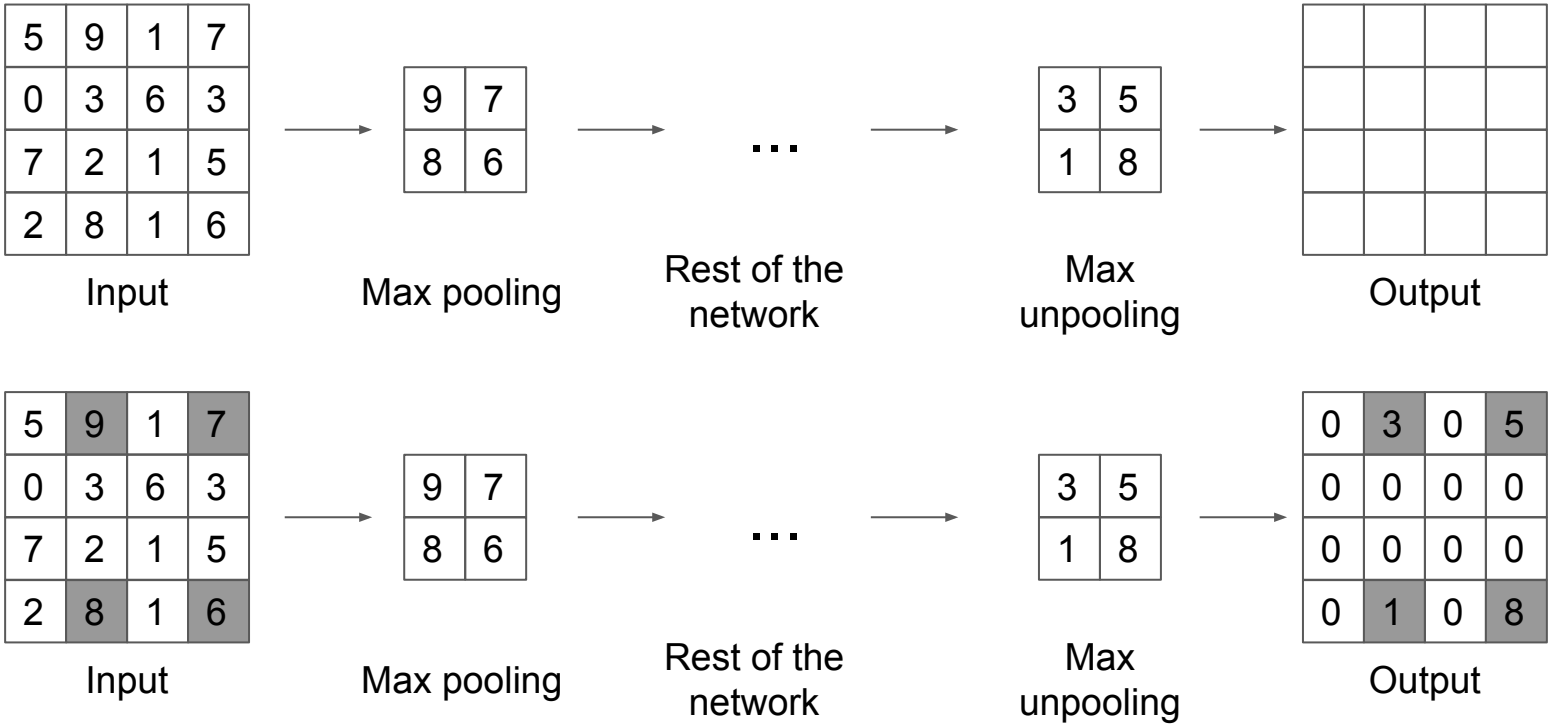


# Exercise 2022



# Exercise 2022

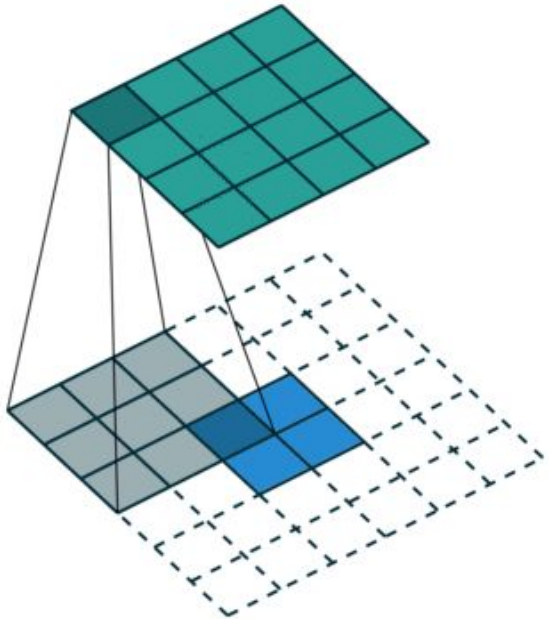


# Semantic segmentation: Traditional Approaches

**Transposed Convolution**

- Convolution with a 3x3 kernel on a 2x2 input with padding

- Outputs a 4x4 output

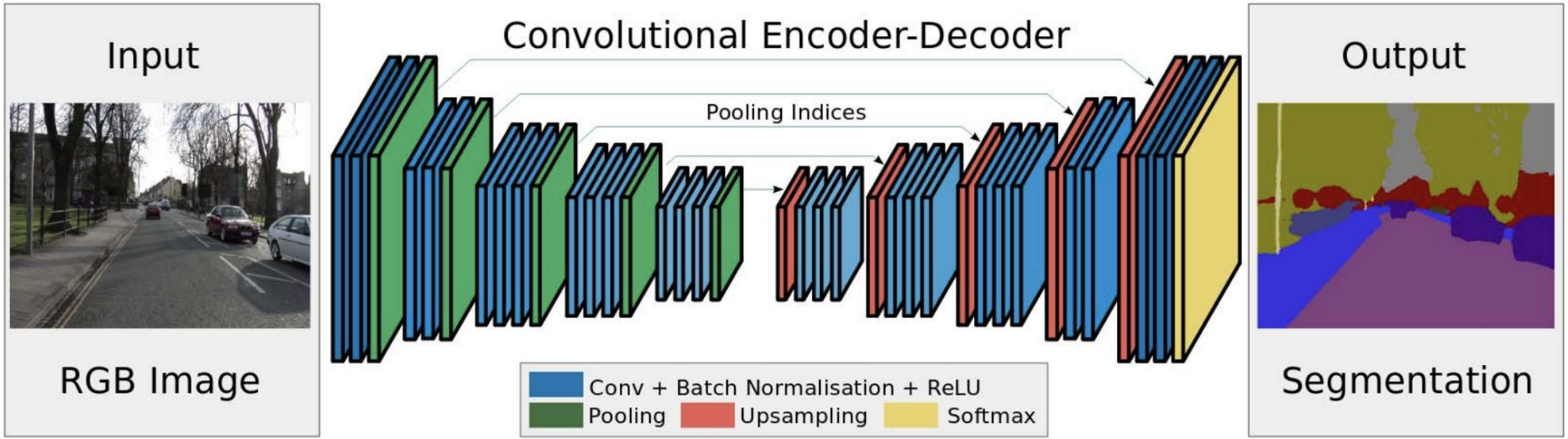


# Semantic segmentation: Modern Approaches

- Deep learning based approaches

# Semantic segmentation: modern approaches

- SegNet [arxiv2015, PAMI2017] (>10K citations)




SegNet encoder/decoder followed by softmax for pixel-wise classification

Image sourced from Badrinarayanan, Vijay, Alex Kendall, and Roberto Cipolla. "Segnet: A deep convolutional encoder-decoder architecture for image segmentation."

# Semantic segmentation: modern approaches

- SegNet [arxiv2015, PAMI2017]
  - PAMI 2017: https://ieeexplore.ieee.org/abstract/document/7803544
  - arxiv 2015: https://arxiv.org/abs/1505.07293

# Semantic segmentation: modern approaches

- Fully Convolutional Network (FCN) [CVPR2015, PAMI2016] (>25K citations)

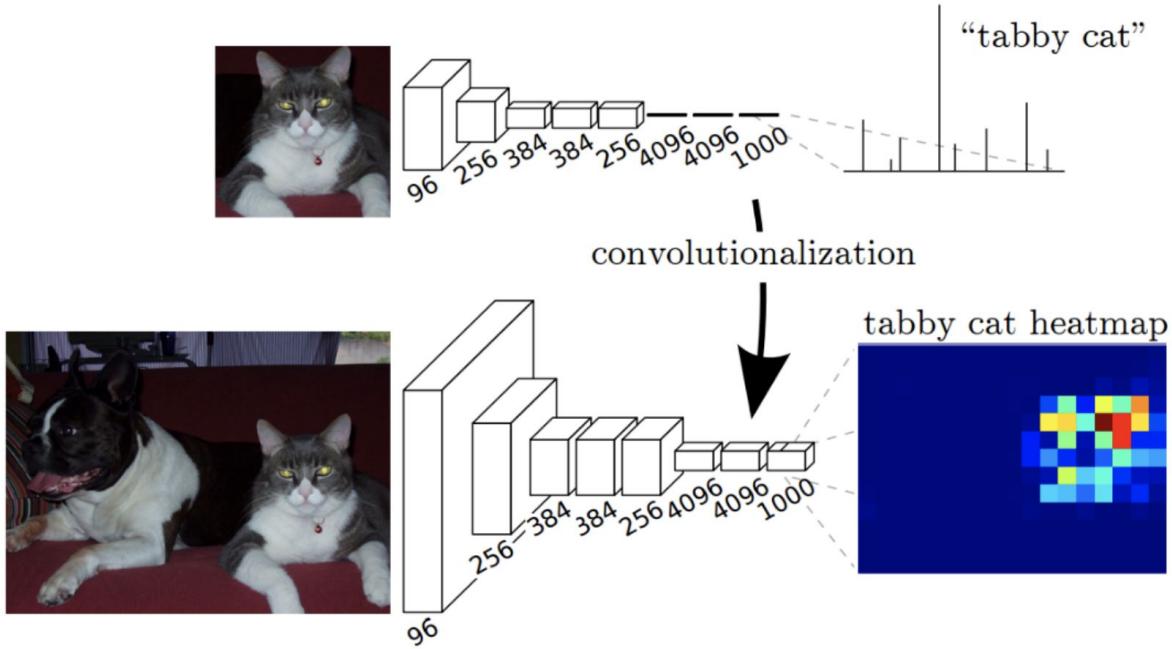



Image sourced from Long, Jonathan, Evan Shelhamer, and Trevor Darrell. "Fully convolutional networks for semantic segmentation."

# Semantic segmentation: modern approaches

- Fully Convolutional Network (FCN) [CVPR2015, PAMI2016] (>25K citations)

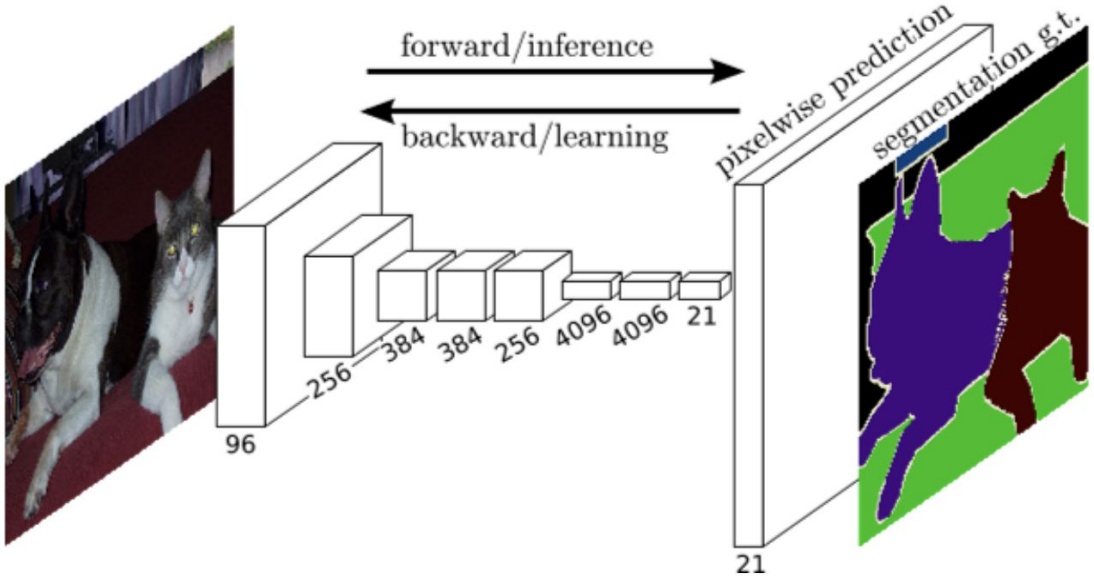
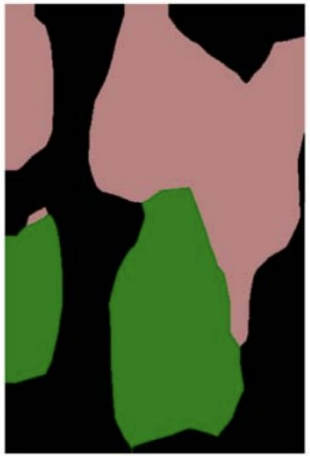



Image sourced from Long, Jonathan, Evan Shelhamer, and Trevor Darrell. "Fully convolutional networks for semantic segmentation."

# Semantic segmentation: modern approaches

- Fully Convolutional Network (FCN) [CVPR2015, PAMI2016] (>25K citat

FCN-32s



Encoder reduces input resolution by a factor or 32x!!!

Decoder struggles to produce good segmentations!

How can we improve that?

Ground truth

Image sourced from Long, Jonathan, Evan Shelhamer, and Trevor Darrell. "Fully convolutional networks for semantic segmentation."

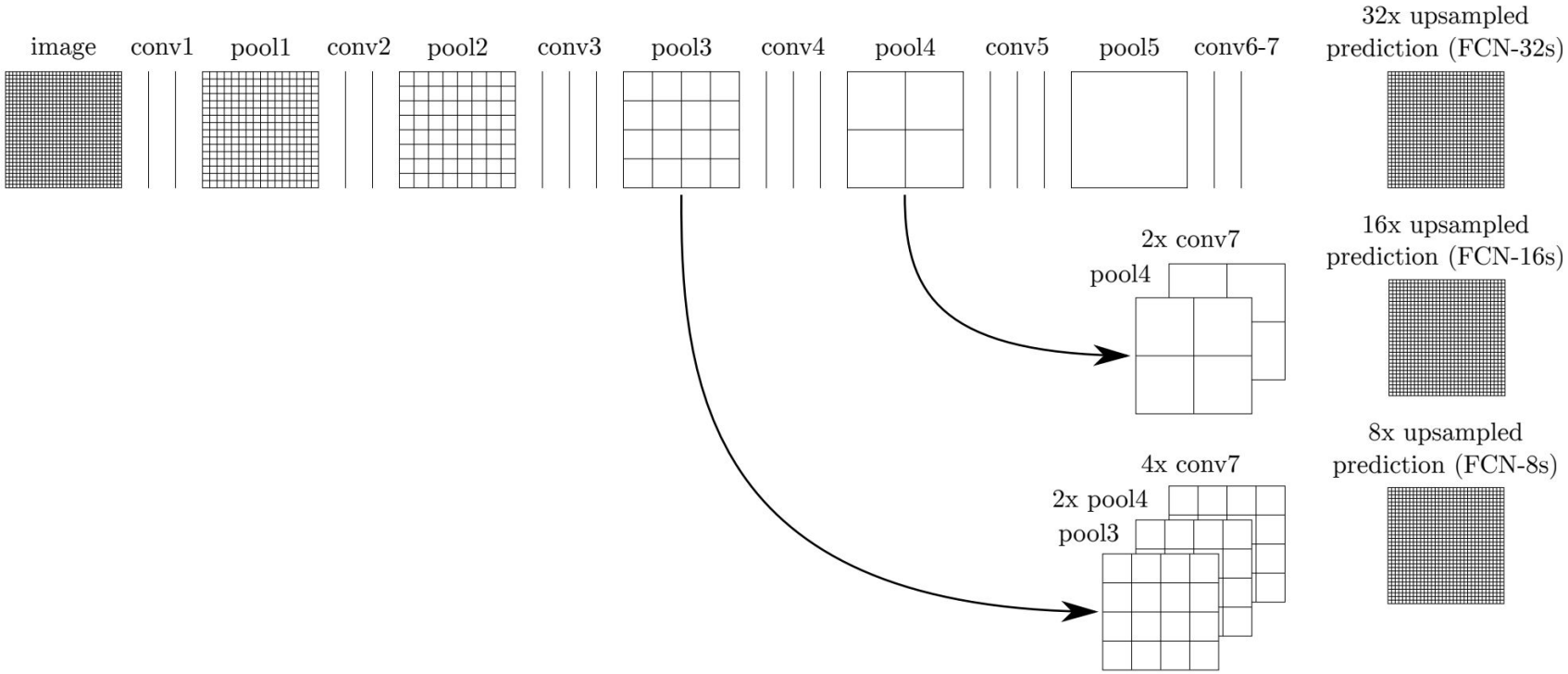
# Semantic segmentation: modern approaches

- Fully Convolutional Network (FCN) [CVPR2015, PAMI2016] (>25K citations)

"Semantic segmentation faces an inherent tension between semantics and location: global information resolves **what** while local information resolves **where**... Combining fine layers and coarse layers lets the model make local predictions that respect global structure."

# Semantic segmentation: modern approaches

- Fully Convolutional Network (FCN) [CVPR2015, PAMI2016] (>25K



# Semantic segmentation: modern approaches

- Fully Convolutional Network (FCN) [CVPR2015, PAMI2016] (>25K cit

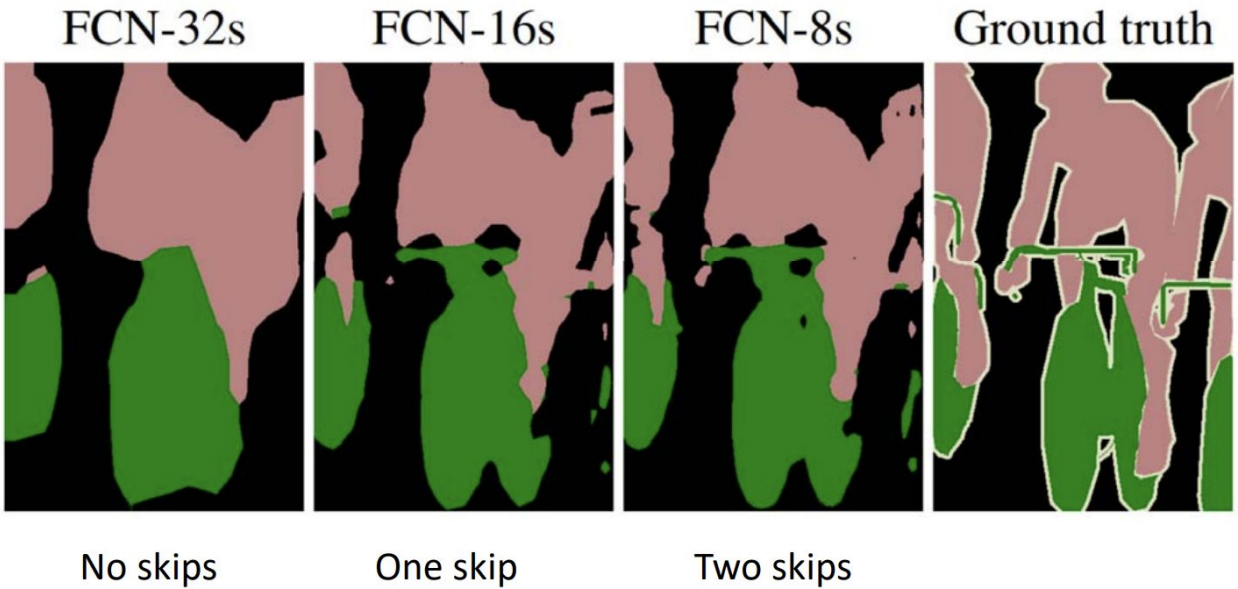


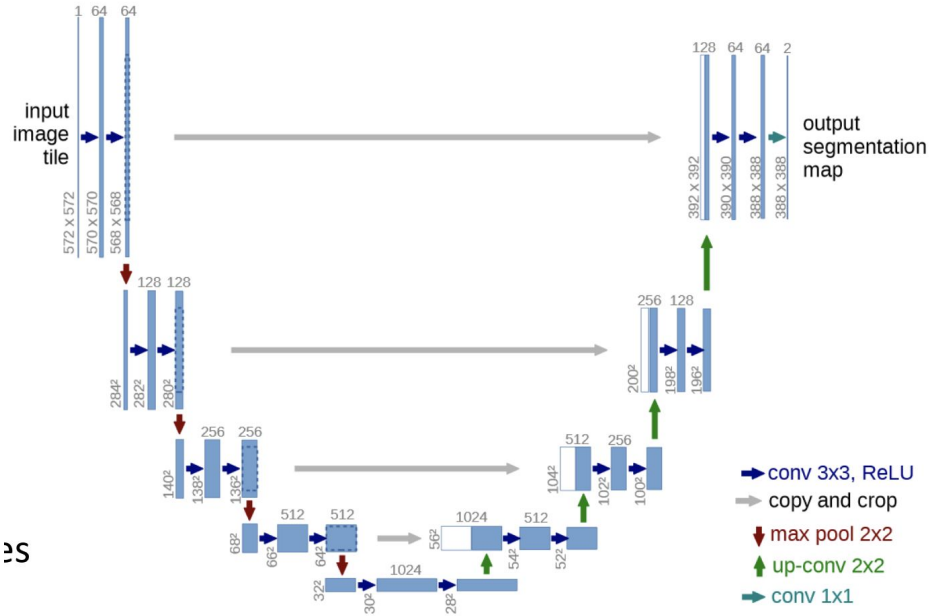

Image sourced from Long, Jonathan, Evan Shelhamer, and Trevor Darrell. "Fully convolutional networks for semantic segmentation."

# Semantic segmentation: modern approaches

- Fully Convolutional Network (FCN) [CVPR2015, PAMI2016] (>25K citations)
  - CVPR 2015: https://www.cv-foundation.org/openaccess/content_cvpr_2015/papers/Long_Fully_Convolutional_Networks_2015_CVPR_paper.pdf
  - PAMI 2016: https://ieeexplore.ieee.org/document/7478072

# Semantic segmentation: modern approaches

- U-Net [MICCAI2015, PAMI2016] (>38K citations)
  - Improves upon FCN
  - Higher decoder capacity
  - Contractive path (context)
  - Expanding path (precision)
  - Heavy data augmentation
  - Used widely for medical images




age sourced from Ronneberger, O., Fischer, P., & Brox, T. "U-net: Convolutional networks for biomedical image gmentation"

# Semantic segmentation: modern approaches

- U-Net [MICCAI2015, PAMI2016] (>38K citations)

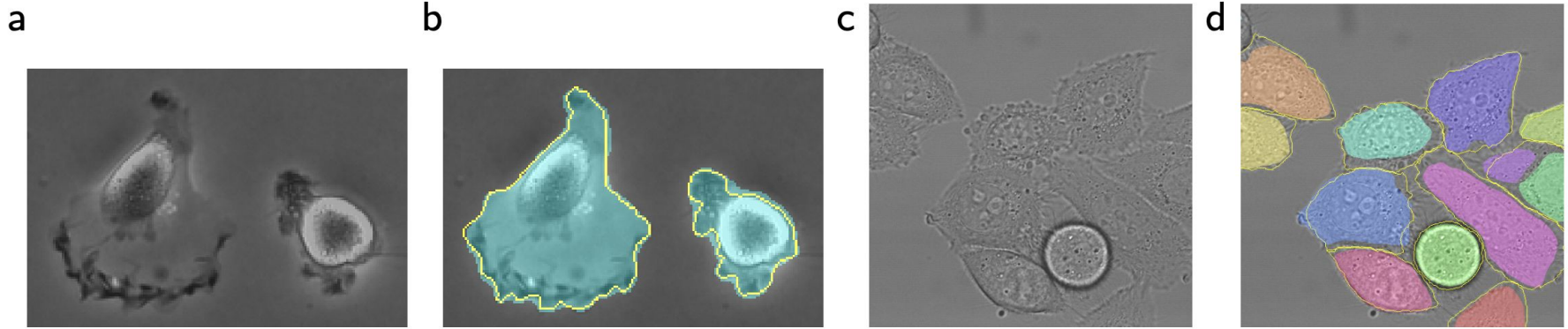



**Fig. 4.** Result on the ISBI cell tracking challenge. (**a**) part of an input image of the "PhC-U373" data set. (**b**) Segmentation result (cyan mask) with manual ground truth (yellow border) (**c**) input image of the "DIC-HeLa" data set. (**d**) Segmentation result (random colored masks) with manual ground truth (yellow border).

# Semantic segmentation: modern approaches

- U-Net [MICCAI2015, PAMI2016] (>38K citations)
  - MICCAI 2015: https://link.springer.com/chapter/10.1007/978-3-319-24574-4_28
  - arxiv version: https://arxiv.org/abs/1505.04597

# Semantic segmentation: modern approaches

- DeepLab [ICLR2015, PAMI2017] (>10K citations)

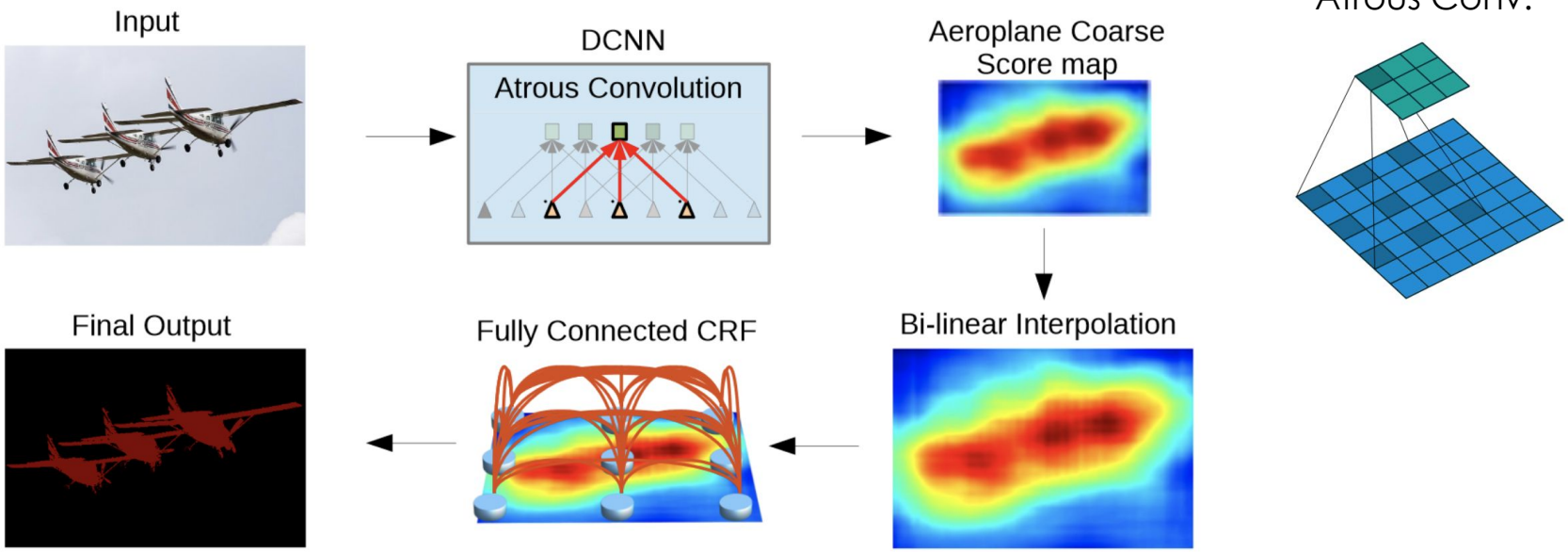



Image sourced from Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L.. "DeepLab: Semantic Image Segmentation withDeep Convolutional Nets, Atrous Convolution,and Fully Connected CRFs"

# Semantic segmentation: modern approaches

- DeepLab [ICLR2015, PAMI2017] (>10K citations)

Pooling loses spatial information

But if we don't pool the receptive field becomes too small

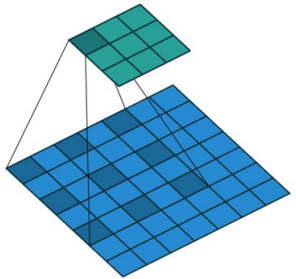
That leads to bad performance

Use dilated convolutions instead!

Widen the receptive field

While avoiding spatial resolution coarsening!

Atrous Conv.



# Semantic segmentation: modern approaches

- DeepLab [ICLR2015, PAMI2017] (>10K citations)
- Conditional Random Fields
  - CRF is a probabilistic framework for labeling and segmenting structured data
  - CRF is still applied nowadays as a post-processing technique
  - Basic ideas:
    - Nearby pixels more likely to have same label
    - Pixels with similar color/texture/… more likely to have same label
    - Pixels surrounded by "river" label more likely to be a boat than a car
    - Refine results by iterations

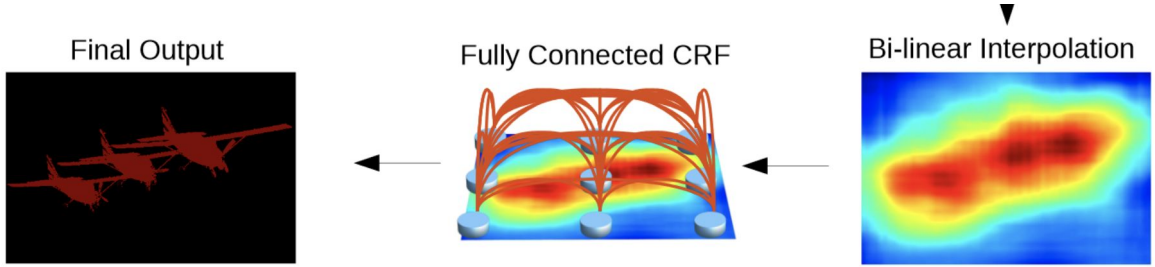
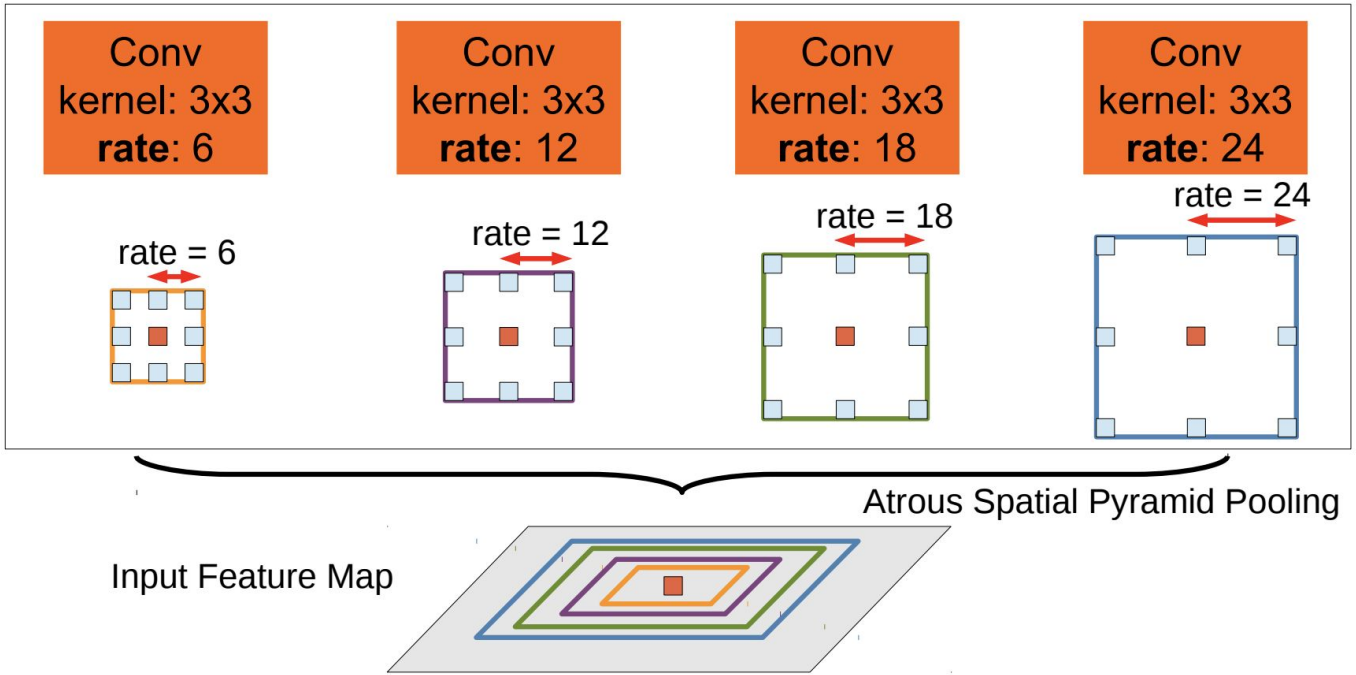



Image sourced from Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L.. "DeepLab: Semantic Image Segmentation withDeep Convolutional Nets, Atrous Convolution,and Fully Connected CRFs"

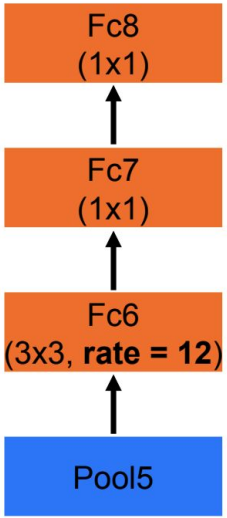
# Semantic segmentation: modern approaches

- DeepLab [ICLR2015, PAMI2017] (>10K citations)

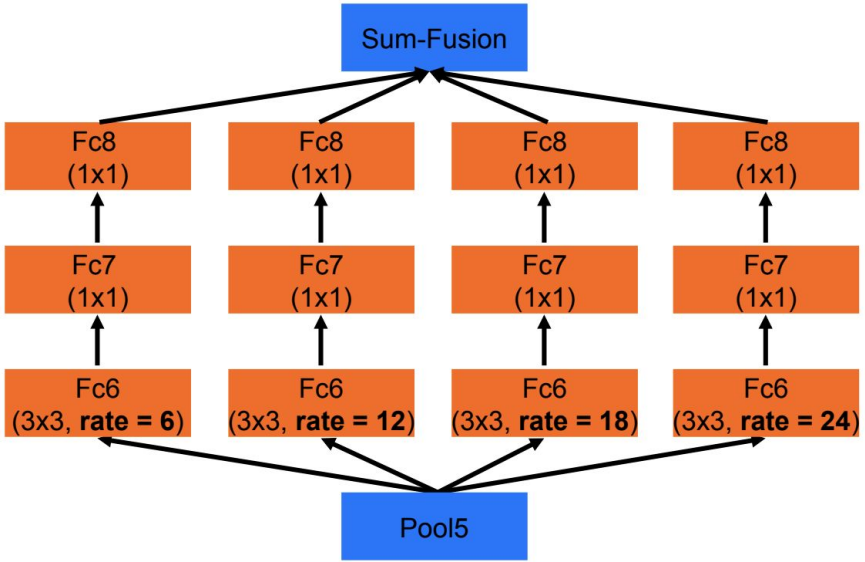


# Semantic segmentation: modern approaches

- DeepLab [ICLR2015, PAMI2017] (>10K citations)



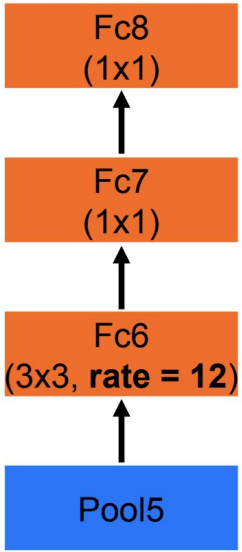

(a) DeepLab-LargeFOV




(b) DeepLab-ASPP

# Semantic segmentation: modern approaches

- DeepLab [ICLR2015, PAMI2017] (>10K citations)




DeepLab-LargeFOV

| Kernel | Rate | FOV | Params | Speed | bef/aft CRF |
|---|---|---|---|---|---|
| $7\times7$ | 4 | 224 | 134.3M | 1.44 | 64.38 / 67.64 |
| $4\times4$ | 4 | 128 | 65.1M | 2.90 | 59.80 / 63.74 |
| $4\times4$ | 8 | 224 | 65.1M | 2.90 | 63.41 / 67.14 |
| $3\times3$ | 12 | 224 | 20.5M | 4.84 | 62.25 / 67.64 |

TABLE 1: Effect of Field-Of-View by adjusting the kernel size and atrous sampling rate $r$ at 'fc6' layer. We show number of model parameters, training speed (img/sec), and *val* set mean IOU before and after CRF. DeepLab-LargeFOV (kernel size $3\times3$, $r = 12$) strikes the best balance.

# Semantic segmentation: modern approaches

- DeepLab [ICLR2015, PAMI2017] (>10K citations)

| MSC | COCO | Aug | LargeFOV | ASPP | CRF | mIOU |
|---|---|---|---|---|---|---|
| | | | | | | 68.72 |
| ✓ | | | | | | 71.27 |
| ✓ | ✓ | | | | | 73.28 |
| ✓ | ✓ | ✓ | | | | 74.87 |
| ✓ | ✓ | ✓ | ✓ | | | 75.54 |
| ✓ | ✓ | ✓ | | ✓ | | 76.35 |
| ✓ | ✓ | ✓ | | ✓ | ✓ | 77.69 |

TABLE 4: Employing ResNet-101 for DeepLab on PASCAL VOC 2012 *val* set. **MSC**: Employing mutli-scale inputs with max fusion. **COCO**: Models pretrained on MS-COCO. **Aug**: Data augmentation by randomly rescaling inputs.

# Semantic segmentation: Modern Approaches

- DeepLab [ICLR2015, PAMI2017] (>10K citations)

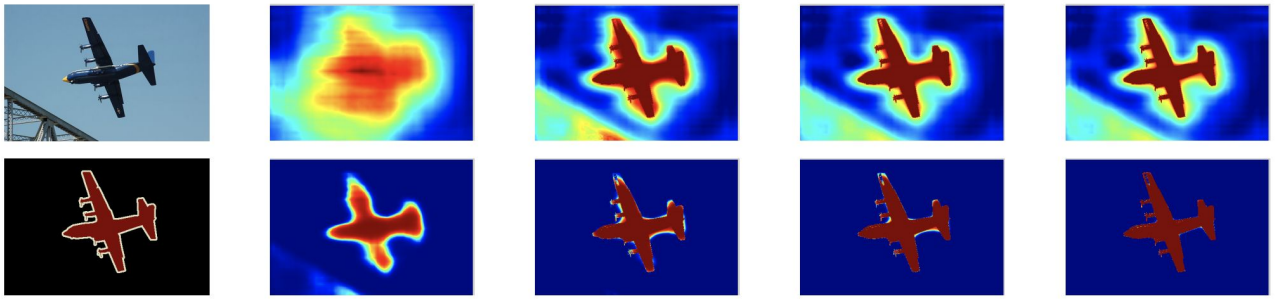


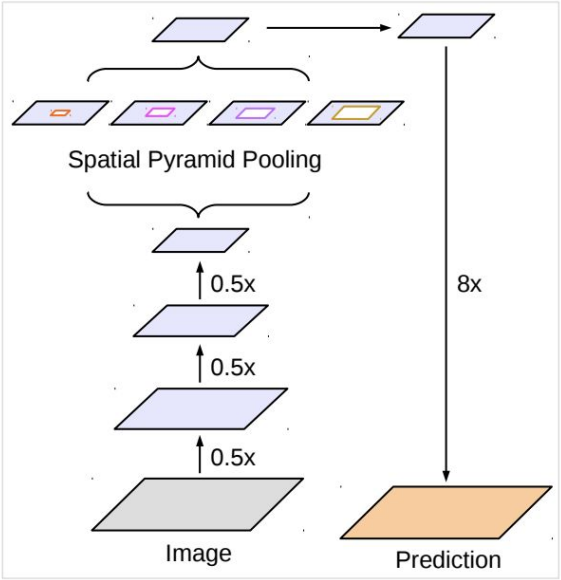

Fig. 5: Score map (input before softmax function) and belief map (output of softmax function) for Aeroplane. We show the score (1st row) and belief (2nd row) maps after each mean field iteration. The output of last DCNN layer is used as input to the mean field inference.

# Semantic segmentation: modern approaches

- DeepLab [ICLR2015, PAMI2017] (>10K citations)
  - PAMI 2017: https://ieeexplore.ieee.org/abstract/document/7913730
  - arxiv version: https://arxiv.org/pdf/1606.00915.pdf

# Semantic segmentation: modern approaches

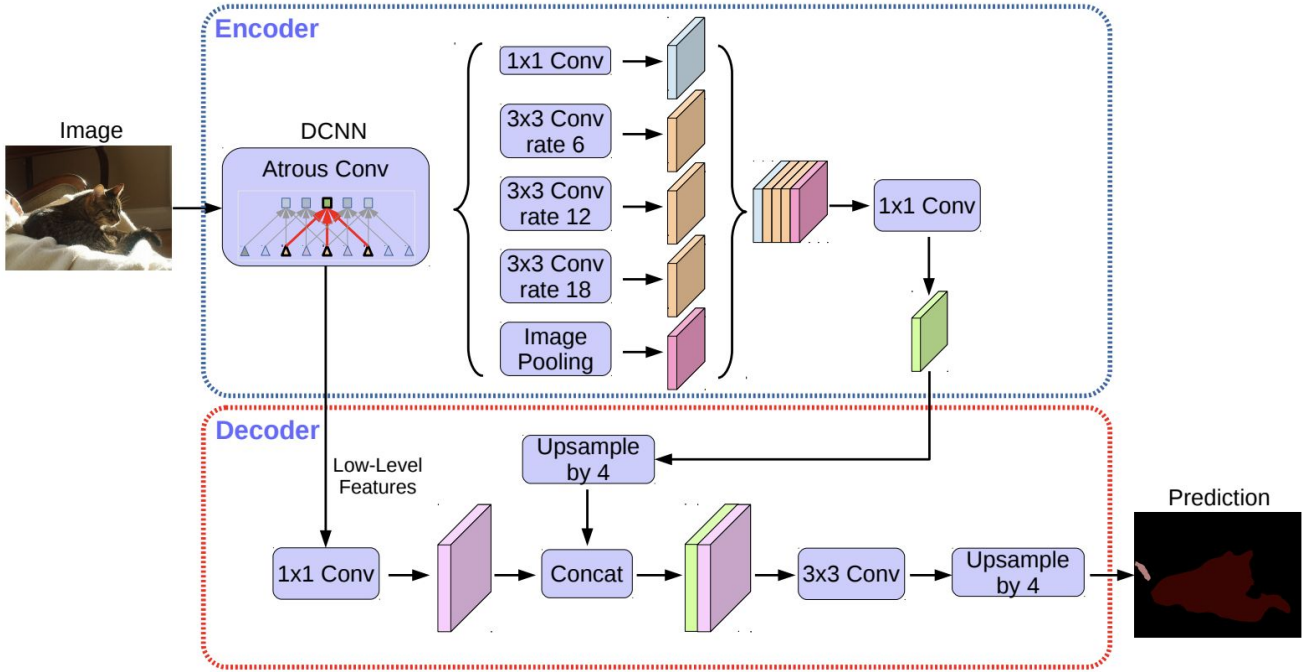
- DeepLab v3+ [ECCV2018] (>6K citations)
- Uses a more powerful backbone network than DeepLab
- Includes a feature pyramid network (FPN) for multi-scale feature fusion
- Includes an atrous spatial pyramid pooling (ASPP) layer for capturing features at different scales




(a) Spatial Pyramid Pooling

# Semantic segmentation: modern approaches

- DeepLab v3+ [ECCV2018] (>6K citations)

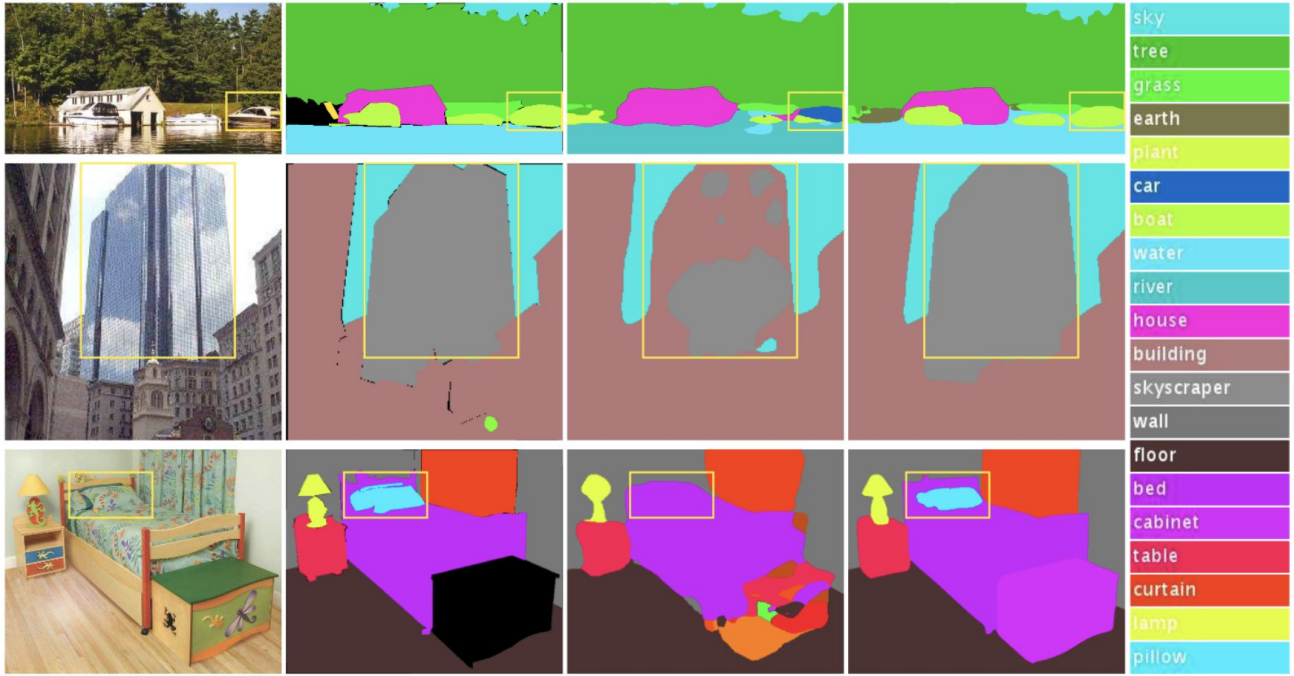


# Semantic segmentation: modern approaches

- DeepLab v3+ [ECCV2018] (>6K citations)
  - ECCV 2018: http://openaccess.thecvf.com/content_ECCV_2018/papers/Liang-Chieh_Chen_Encoder-Decoder_with_Atrous_ECCV_2018_paper.pdf

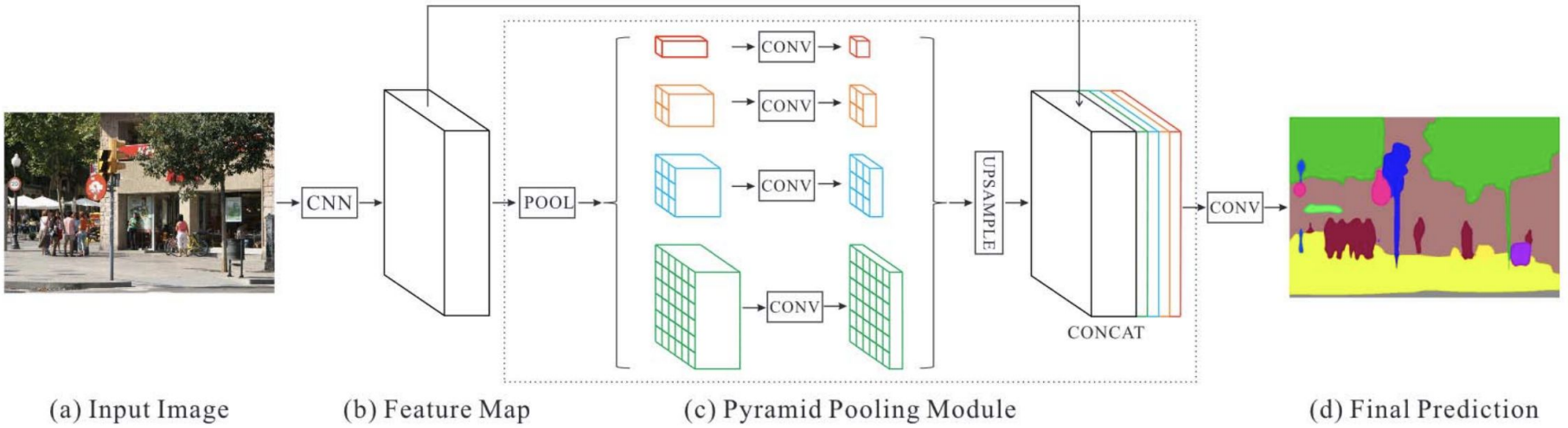
# Semantic segmentation: modern approaches

- Pyramid Scene Parsing Network (PSPNet) [CVPR2017] (>6K citations)



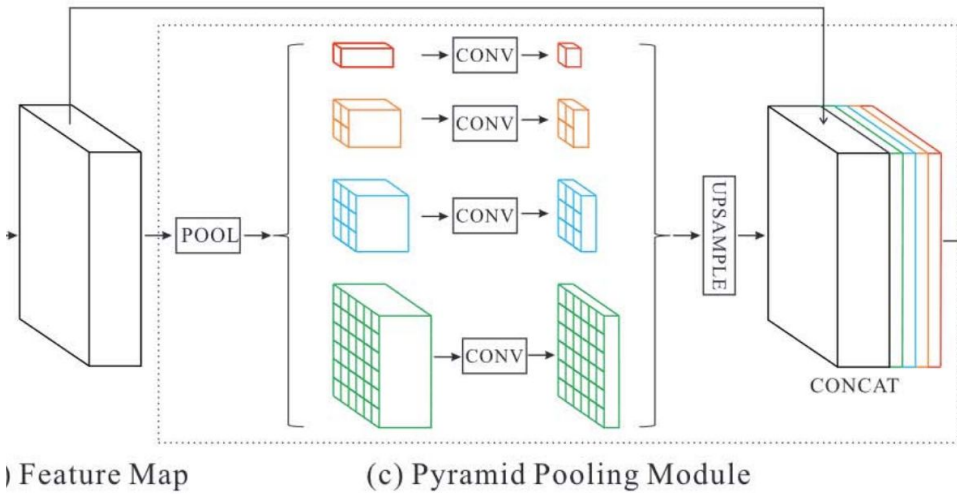
# Semantic segmentation: modern approaches

- Pyramid Scene Parsing Network (PSPNet) [CVPR2017] (>6K citations)



# Semantic segmentation: modern approaches

- Pyramid Scene Parsing Network (PSPNet) [CVPR2017] (>6K citations)
  - Patterns are sometimes relationships between small and large objects
  - Hard to account for both with classical architectures
  - Pyramid pooling module provides global-scene priors over 4 different scales:
    - over whole input
    - over 2x2 areas of input
    - over 3x3 areas of input
    - over 6x6 areas of input



# Semantic segmentation: modern approaches

- Pyramid Scene Parsing Network (PSPNet) [CVPR2017] (>6K citations)

| Method | Mean IoU(%) | Pixel Acc.(%) |
|---|---|---|
| ResNet50-Baseline | 37.23 | 78.01 |
| ResNet50+B1+MAX | 39.94 | 79.46 |
| ResNet50+B1+AVE | 40.07 | 79.52 |
| ResNet50+B1236+MAX | 40.18 | 79.45 |
| ResNet50+B1236+AVE | 41.07 | 79.97 |
| ResNet50+B1236+MAX+DR | 40.87 | 79.61 |
| ResNet50+B1236+AVE+DR | **41.68** | **80.04** |

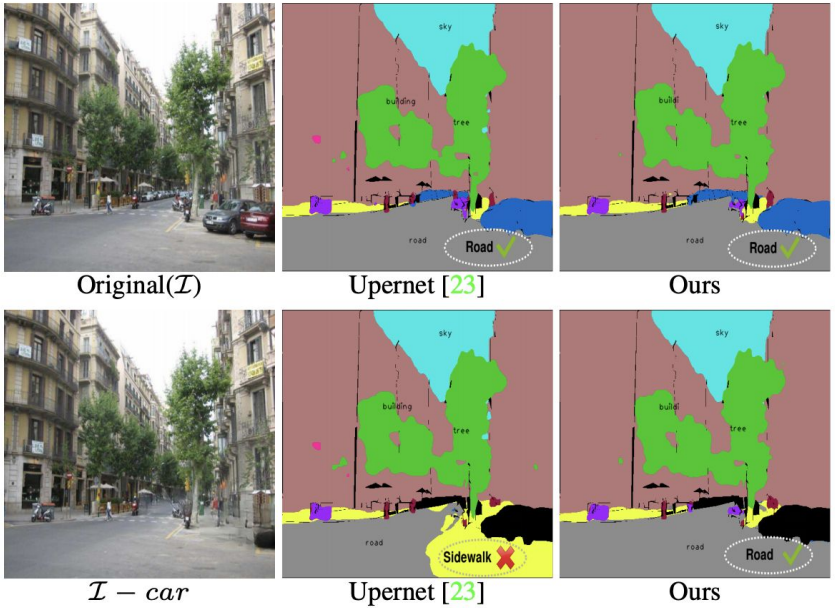
Table 1. Investigation of PSPNet with different settings. Baseline is ResNet50-based FCN with dilated network. 'B1' and 'B1236' denote pooled feature maps of bin sizes $\{1 \times 1\}$ and $\{1 \times 1, 2 \times 2, 3 \times 3, 6 \times 6\}$ respectively. 'MAX' and 'AVE' represent max pooling and average pooling operations individually. 'DR' means that dimension reduction is taken after pooling. The results are tested on the validation set with the single-scale input.

# Semantic segmentation: modern approaches

- Pyramid Scene Parsing Network (PSPNet) [CVPR2017] (>6K citations)
  - CVPR 2017: http://openaccess.thecvf.com/content_cvpr_2017/papers/Zhao_Pyramid_Scene_Parsing_CVPR_2017_paper.pdf

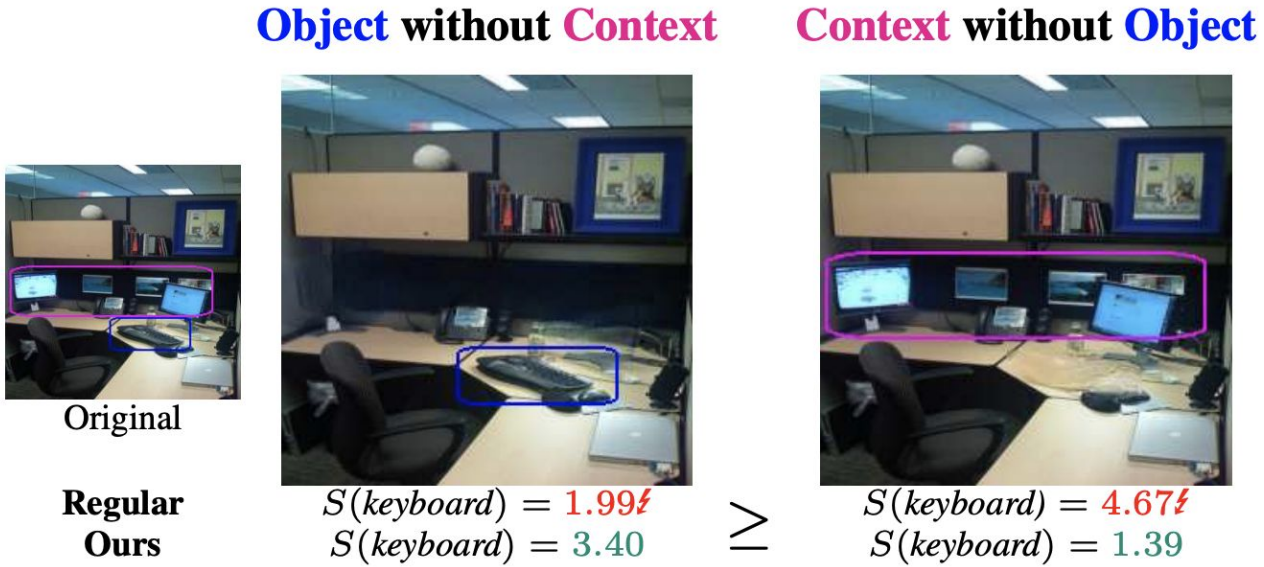
# Semantic segmentation: modern approaches

- Not Using the Car to See the Sidewalk — Quantifying and Controlling the Effects of Context in Classification and Segmentation [CVPR2019]
  - Interesting paper about the **influence of context**



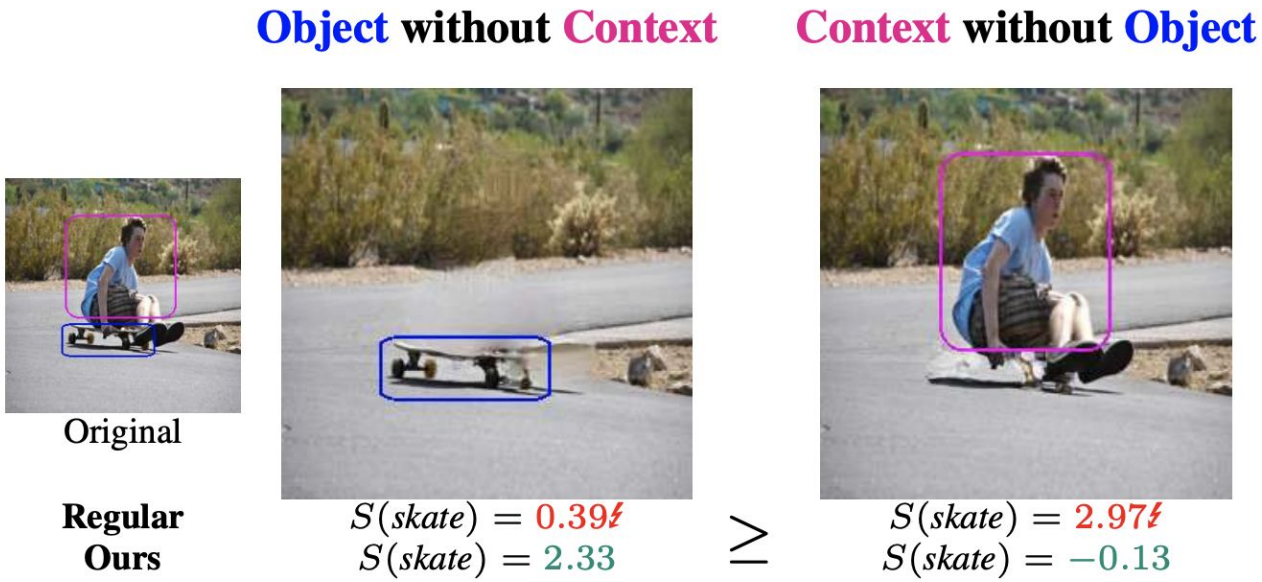
# Semantic segmentation: modern approaches

- Not Using the Car to See the Sidewalk — Quantifying and Controlling the Effects of Context in Classification and Segmentation [CVPR2019]



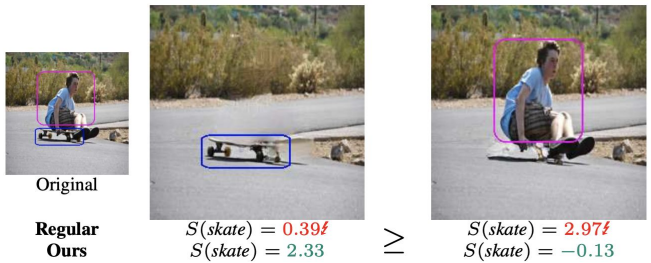
# Semantic segmentation: modern approaches

- Not Using the Car to See the Sidewalk — Quantifying and Controlling the Effects of Context in Classification and Segmentation [CVPR2019]



# Semantic segmentation: modern approaches

- Not Using the Car to See the Sidewalk — Quantifying and Controlling the Effects of Context in Classification and Segmentation [CVPR2019]
  - Object removal based data augmentation
    - Mitigate dependency between objects and context
    - Increase robustness of classification and segmentation models to contextual variations
  - Results obtained
    - Improve performance in out-of-context scenarios
    - Preserve performance on regular data

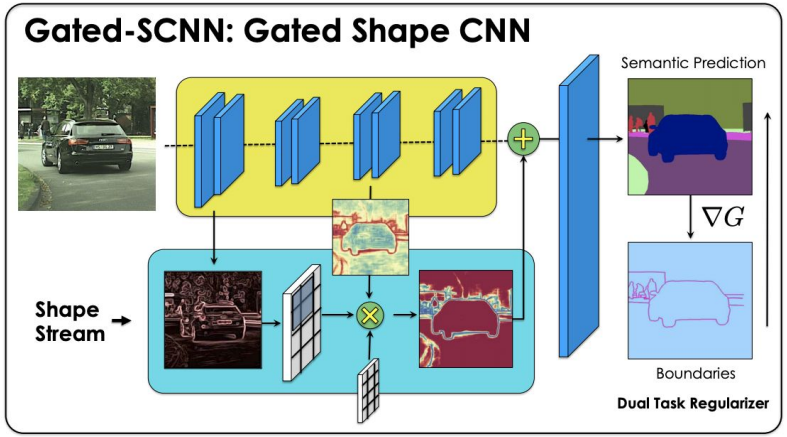


# Semantic segmentation: modern approaches

- Not Using the Car to See the Sidewalk — Quantifying and Controlling the Effects of Context in Classification and Segmentation [CVPR2019]
  - http://openaccess.thecvf.com/content_CVPR_2019/papers/Shetty_Not_Using_the_Car_to_See_the_Sidewalk_--_Quantifying_CVPR_2019_paper.pdf

# Semantic segmentation: modern approaches

- Gated-SCNN: Gated Shape CNNs for Semantic Segmentation [ICCV2019]
  - Two streams: appearance features and shape features
  - The shape stream selectively activates or suppress the shape features, based on their relevance to the segmentation task.
  - This helps prevent the shape features from interfering with the appearance features or vice versa.



# Semantic segmentation: modern approaches

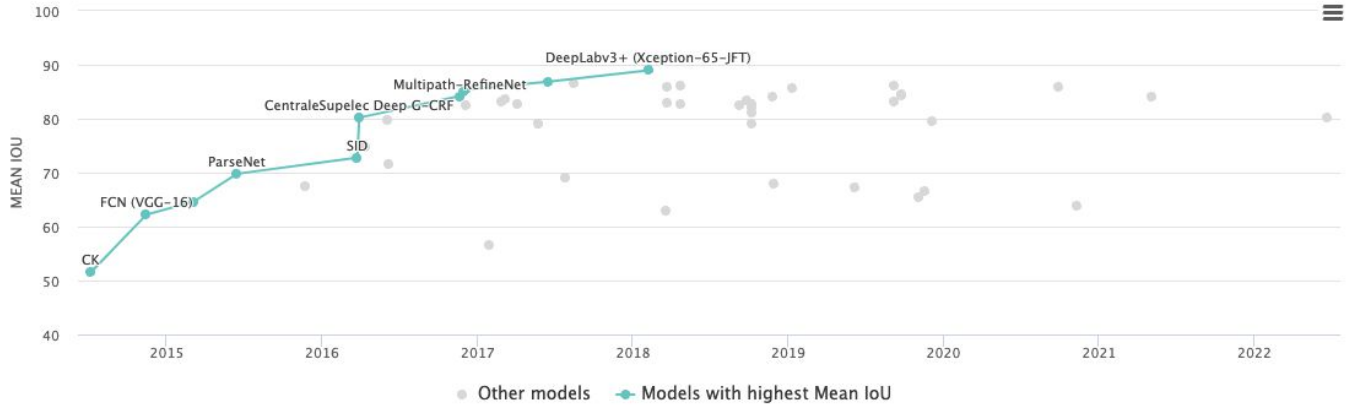
- Gated-SCNN: Gated Shape CNNs for Semantic Segmentation [ICCV2019]
  - http://openaccess.thecvf.com/content_ICCV_2019/papers/Takikawa_Gated-SCNN_Gated_Shape_CNNs_for_Semantic_Segmentation_ICCV_2019_paper.pdf

paperswithcode.com/sota/semantic-segmentation-on-pascal-voc-2012

# Semantic Segmentation on PASCAL VOC 2012 test

Leaderboard Dataset

View Mean IoU by Date for All models




Filter: ResNet CNN Transformer untagged

Edit Leaderboard

| Rank | Model | Mean IoU | FLOPS | Params | Extra Training Data | Paper | Code | Result | Year | Tags |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | **DeepLabv3+** (Xception-65-JFT) | 89.0% | | | ✓ | Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation | | | 2018 | |
| 2 | **DeepLabv3+** | 89.0% | | | ✓ | Encoder-Decoder with Atrous Separable Convolution for Semantic Image | | | 2018 | |

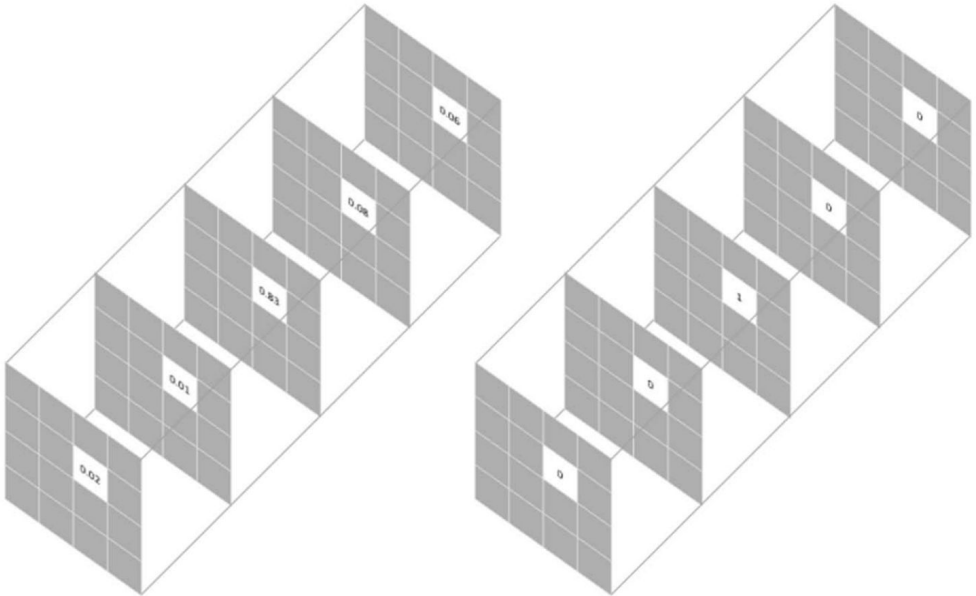
# Semantic segmentation: loss

- Pixel-wise cross entropy loss

Pixel-wise loss is calculated as the log loss, summed over all possible classes

$$-\sum_{classes} y_{true} \log\left(y_{pred}\right)$$

This scoring is repeated over all **pixels** and averaged



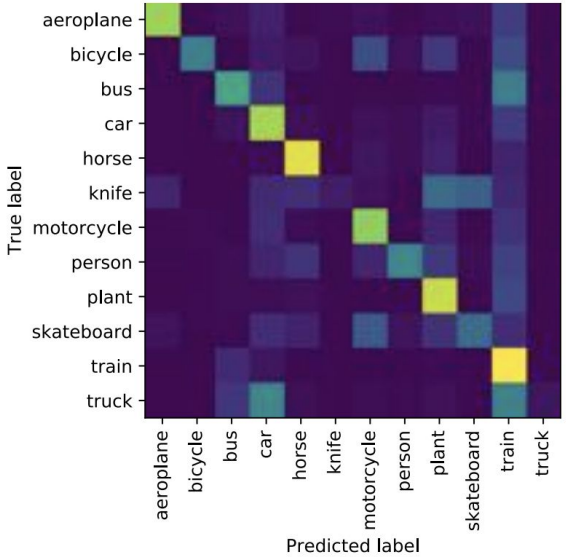

Prediction for a selected pixel

Target for the corresponding pixel



*Image source: https://www.jeremyjordan.me/semantic-segmentation/*

# Semantic segmentation: metrics




- Confusion matrix (same as in classification)
- Global accuracy
  - # correct pixels / total pixels
- Average per-class accuracy
  - avg. # correct pixels class(i) / total pixels class(i), for i=1..
  - suitable for dataset with unlabeled categories
- Mean Intersection over Union (mIoU) (a.k.a. Jaccard)
  - Avg. IoU(i) = TP / (TP + FP + FN) for all classes i=1...C
  - More strict than average per-class accuracy
  - Penalizes false positive predictions
  - Favors region smoothness and does not evaluate boundary accuracy

# Semantic segmentation: datasets

- MS COCO
  - Common Objects in COntext
  - 80 categories
  - ~300K RGB images
  - Dense pixel-wise annotations for semantic segmentation (on superpixels)
  - Class and instance segmentation
  - http://cocodataset.org/#home

# Semantic segmentation: datasets

- PASCAL Context
  - “everyday” images
  - 540 categories
  - ~10k RGB images (train)
  - Dense pixel-wise annotations for semantic segmentation
  - Class segmentation
  - https://cs.stanford.edu/~roozbeh/pascal-context/

# Semantic segmentation: datasets

- ADE20k
  - Diverse set of scenes, objects and object parts
  - Large and unrestricted open vocabulary
  - ~20k RGB images (train)
  - ~2700 categories
  - Dense pixel-wise annotations for semantic segmentation
  - Class segmentation
  - https://groups.csail.mit.edu/vision/datasets/ADE20K/

# Semantic segmentation: datasets

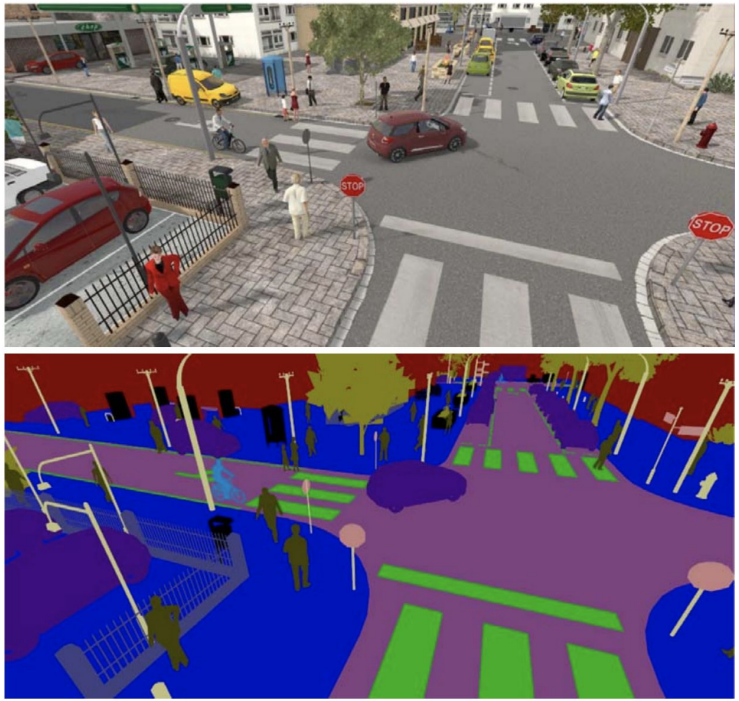
- Cityscapes
  - Urban driving scenarios
  - 30 categories
  - ~25k images with dense annotations
  - ~5k images with dense pixel-wise annotations for semantic segmentation
  - Class and instance segmentation
  - https://www.cityscapes-dataset.com/

# Semantic segmentation: dataset

- Mapillary Vistas Dataset
  - Urban driving scenarios
  - 152 categories
  - ~25k images with dense pixel-wise annotations for semantic segmentation
  - Class and instance segmentation
  - Variety of weather, season, time of day, camera, and viewpoint
  - https://www.mapillary.com/dataset/vistas/

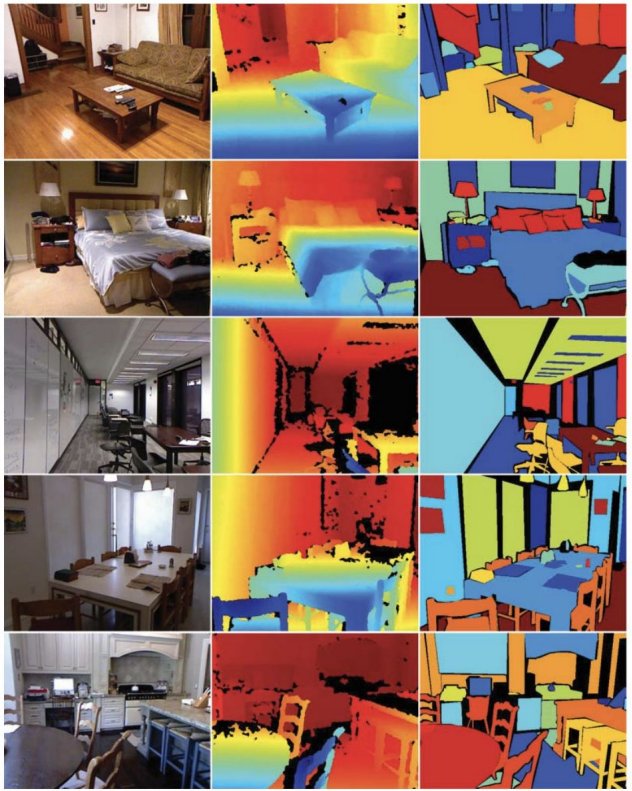

# Semantic segmentation: datasets

- Synthia (from CVC!)
  - Synthetic driving scenarios
  - 13 categories
  - ~500k images with dense pixel-wise annotations for semantic segmentation
  - video streams
  - Class and instance segmentation
  - https://synthia-dataset.net/



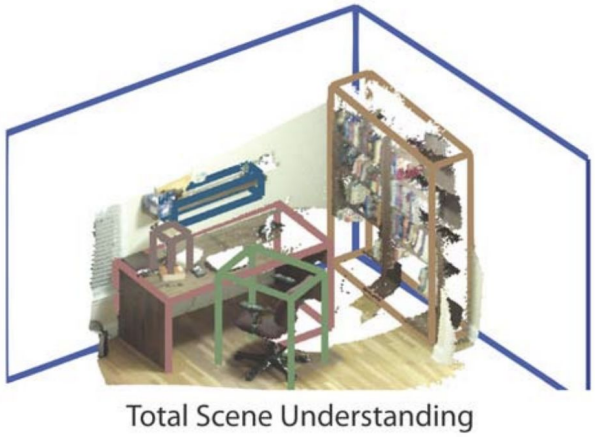
# Semantic segmentation: datasets

- NYUD-V2
  - Real indoor scenes
  - 464 scenes
  - ~1500 RGB-D images
  - Dense pixel-wise annotations for semantic segmentation
  - Class and instance segmentation
  - https://cs.nyu.edu/~silberman/datasets/nyu_depth_v2.html



# Semantic segmentation: datasets

- SUN RGB-D
  - Real indoor scenes
  - ~10k RGB-D images
  - Dense pixel-wise annotations for semantic segmentation
  - Class and instance segmentation
  - http://rgbd.cs.princeton.edu/

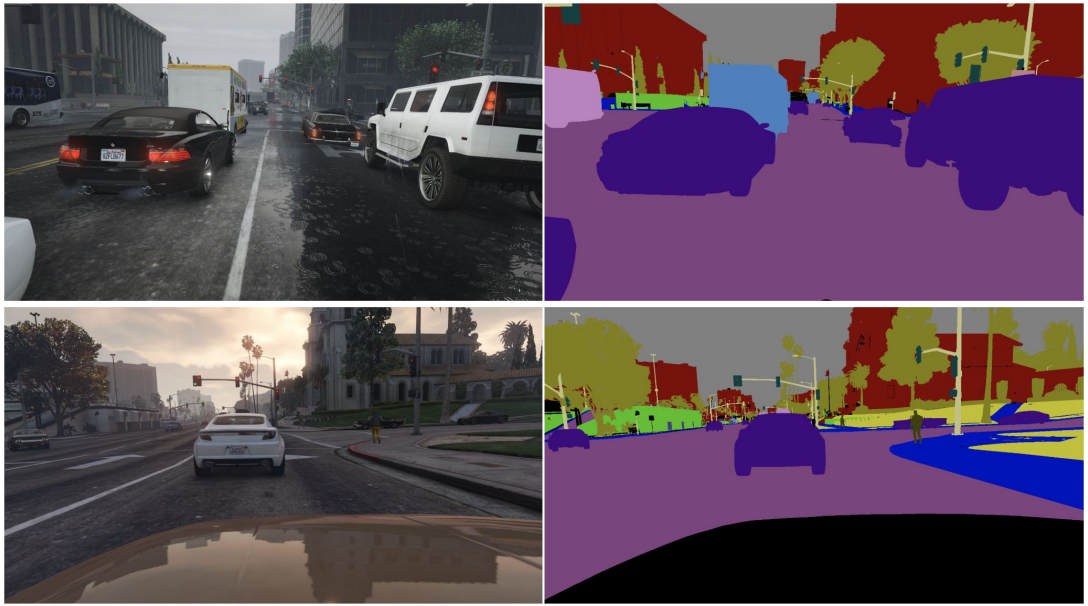


# Semantic segmentation: datasets

- SceneNet RGB-D
  - Synthetic indoor scenes
  - ~5M RGB-D images
  - Dense pixel-wise annotations for semantic segmentation
  - Class and instance segmentation

# Semantic segmentation: datasets

- … and many more:
  - KITTI
  - Virtual KITTI
  - CamVid
  - SynthCity
  - LabelMe
  - GTA-V
  - Pascal Semantic Part
  - The Robotrix
  - ...



# Semantic segmentation: datasets