The Semantic Segmentation Problem

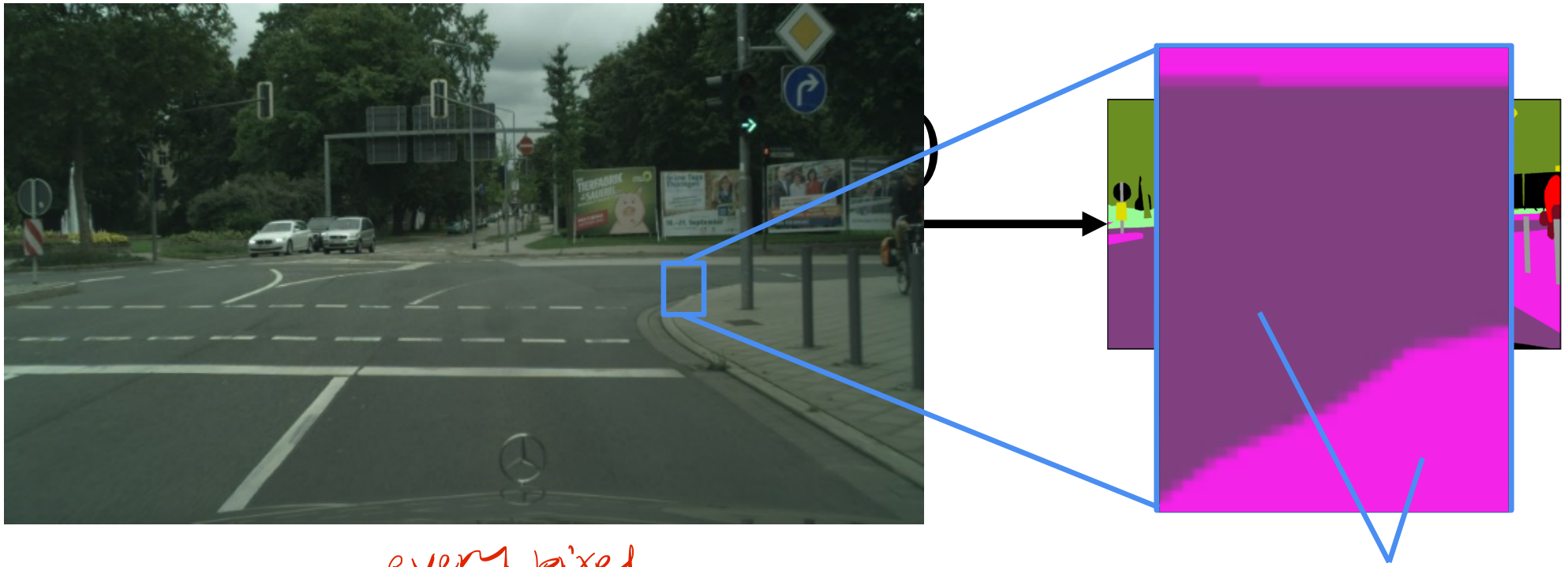
Course 3, Module 5, Lesson 1

UNIVERSITY OF TORONTO
FACULTY OF APPLIED SCIENCE & ENGINEERING

The Semantic Segmentation Problem

- Road
- Sidewalk
- Pole
- Traffic Light
- Traffic Signs
- Vegetation
- Terrain
- Sky
- Car
- Background

Mathematical Problem Formulation



every pixel

$$f(x;\theta) = [S_{class_1}, \cdots, S_{class_k}]$$

$S_{road} = 0.92$

$S_{sidewalk} = 0.02$

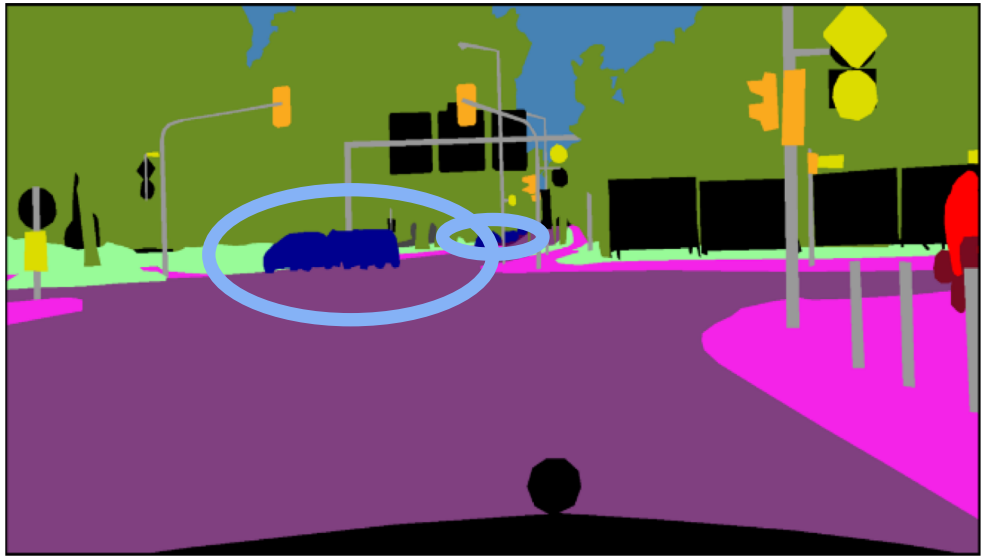
.

.

$S_{Background} = 0.02$

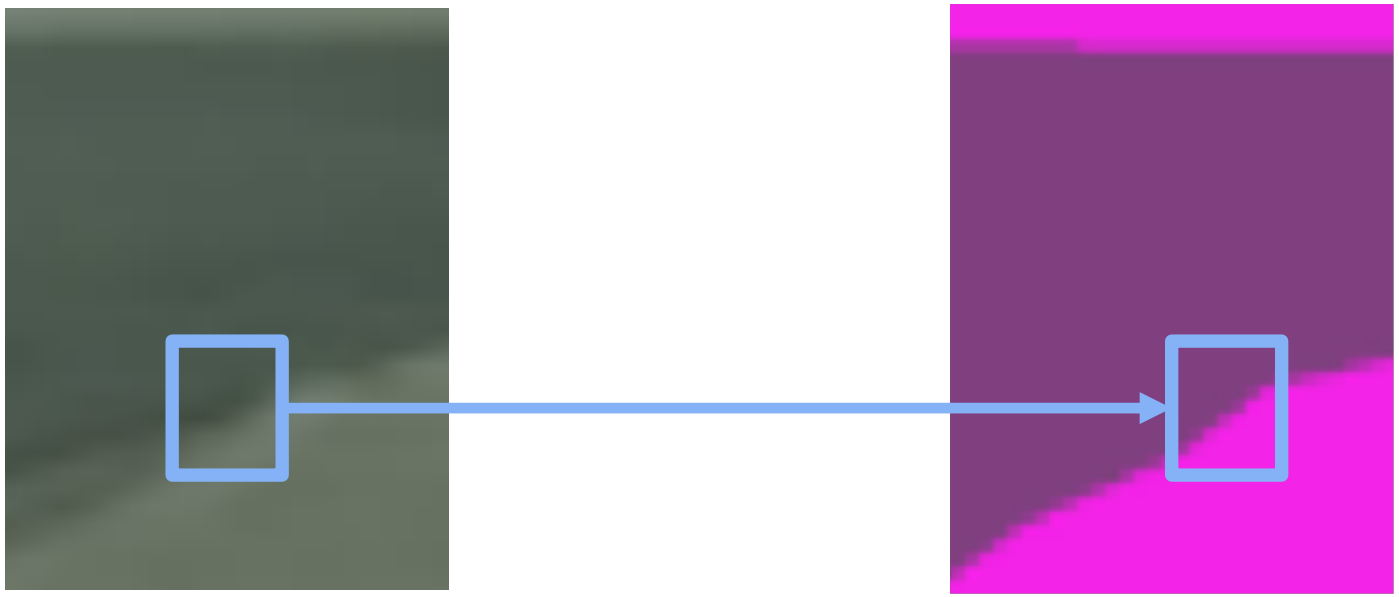
Semantic Segmentation is Not Trivial !

- Occlusion, truncation, scale, and illumination changes



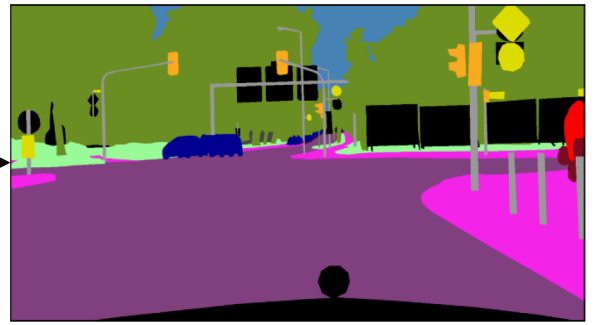
Semantic Segmentation is Not Trivial !

- Occlusion, truncation, scale, and illumination changes
- Smooth boundaries

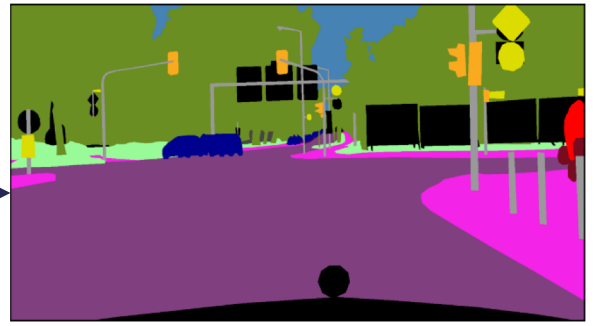


ConvNets For Semantic Segmentation

$$f(x;\theta)$$



ConvNet



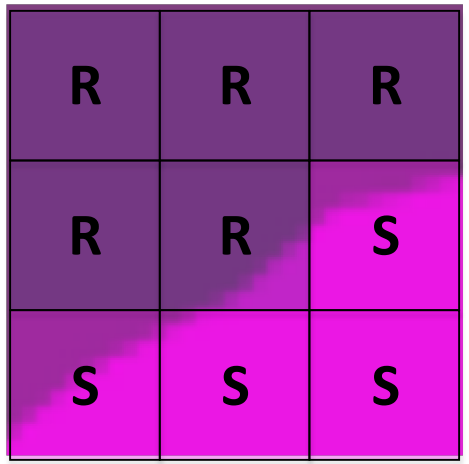
Evaluation Metrics

- **True Positive (TP):** The number of correctly classified pixels belonging to class X
- **False Positive (FP):** The number of pixels that **do not belong** to class X in ground truth but **are classified** that class by the algorithm
- **False Negative (FN):** The number of pixels that **do belong** to class X in ground truth, but **are not classified** as that class by the algorithm

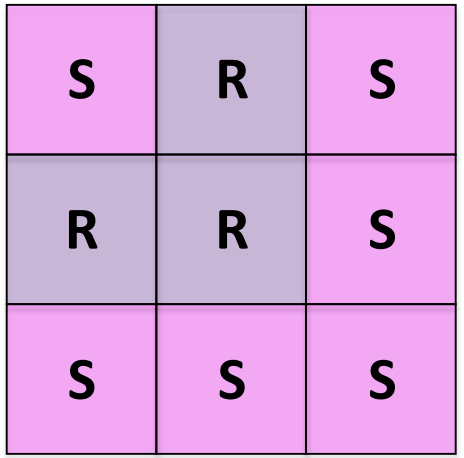
$$IOU_{class} = \frac{TP}{TP + FP + FN}$$

Evaluation Metrics

Ground Truth

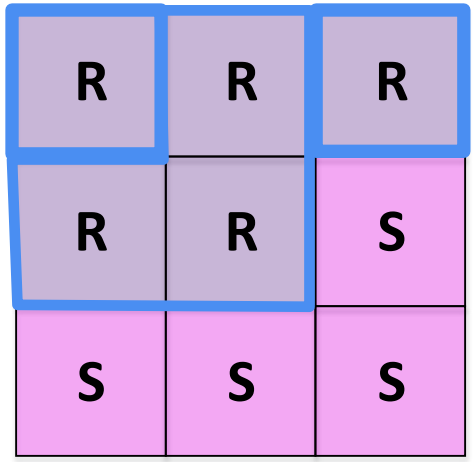


Prediction

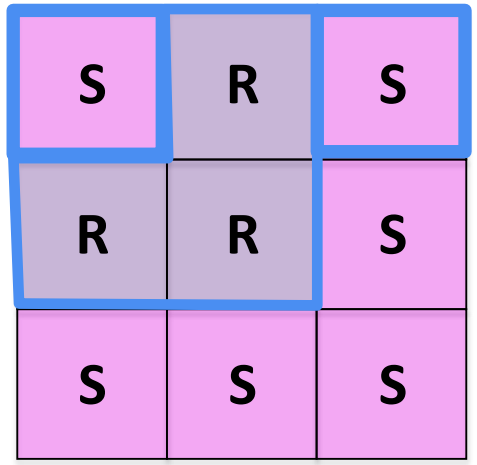


Evaluation Metrics

Ground Truth



Prediction



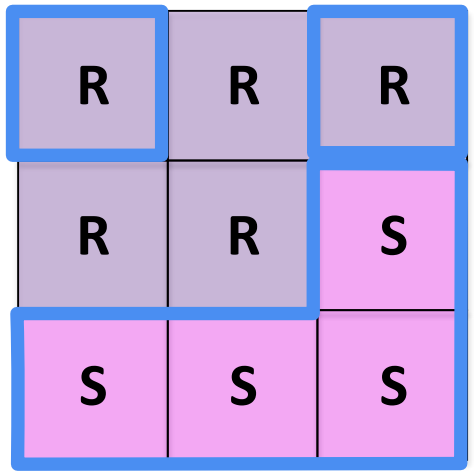
Class: Road

$$TP = 3$$
$$FP = 0$$
$$FN = 2$$

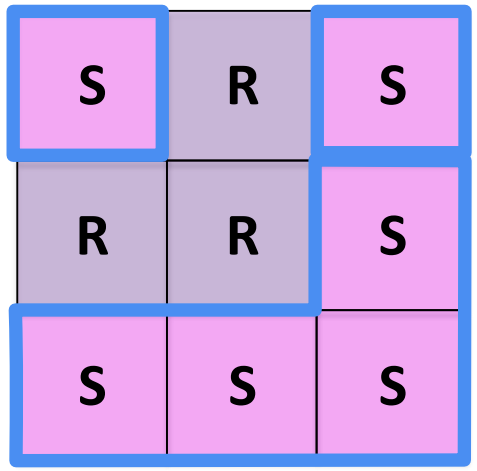
$$IOU_{Road} = \frac{3}{3+0+2} = \frac{3}{5}$$

Evaluation Metrics

Ground Truth



Prediction



Class: Sidewalk

$$TP = 4$$
$$FP = 2$$
$$FN = 0$$

$$IOU_{Road} = \frac{4}{4+2+0} = \frac{4}{6}$$

Evaluation Metrics

- **Class IOU** over all the data is calculated by computing the sum of TP, FP, FN for all images first
- Averaging the class IOU is usually not a very good idea!
- CityScapes Segmentation Dataset

benchmark

bias towards object instances that cover a large image area

Summary

- Semantic segmentation consists of providing a class label for every pixel in a 2D image
- Semantic segmentation models can be evaluated using class IOU

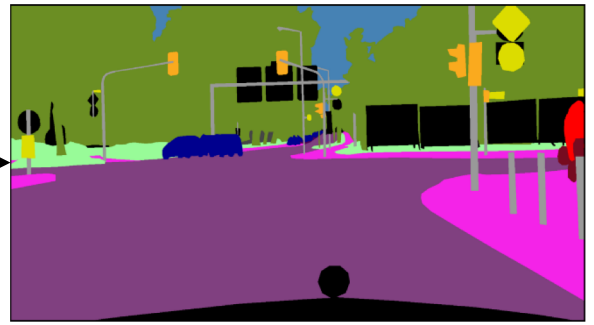
- **Next: ConvNets for Semantic Segmentation**

Convolutional Neural Networks For Semantic Segmentation

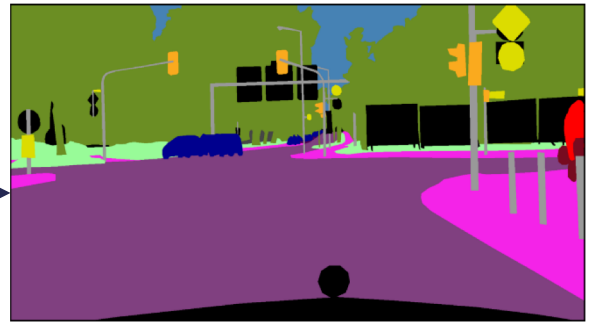
Course 3, Module 5, Lesson 2

ConvNets For Semantic Segmentation

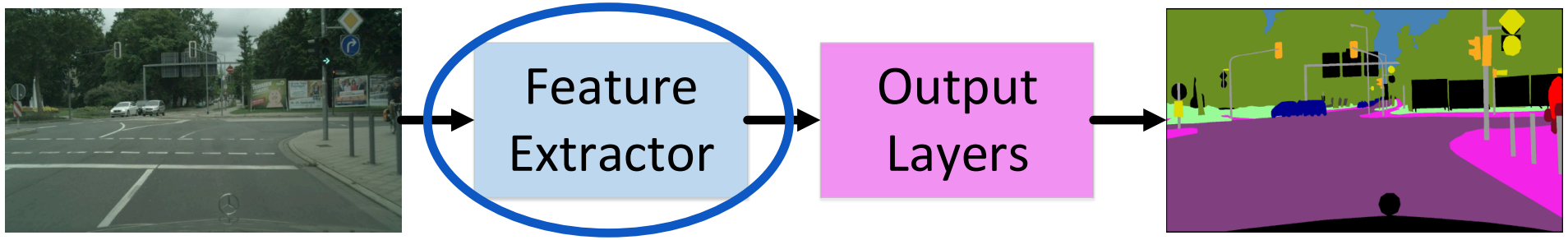
$$f(x;\theta)$$



ConvNet



ConvNets For Semantic Segmentation



The Feature Extractor

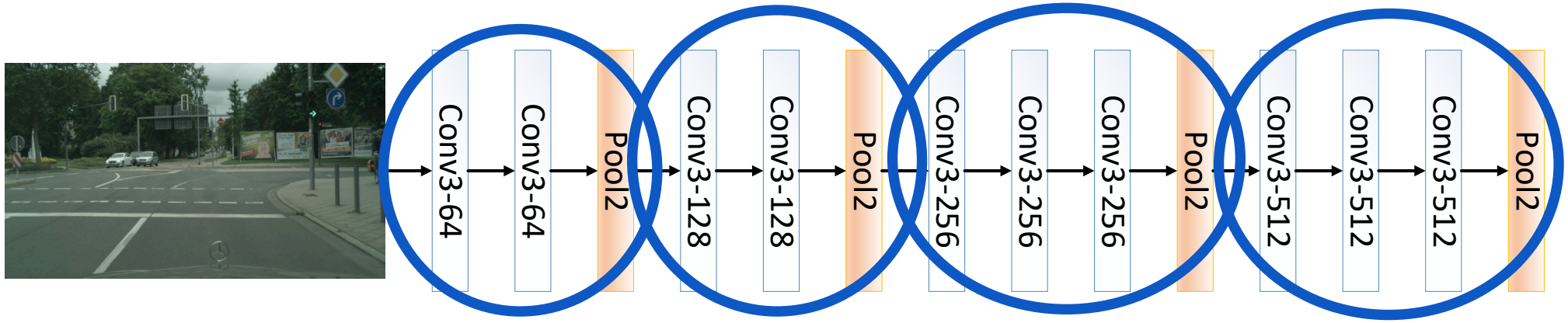
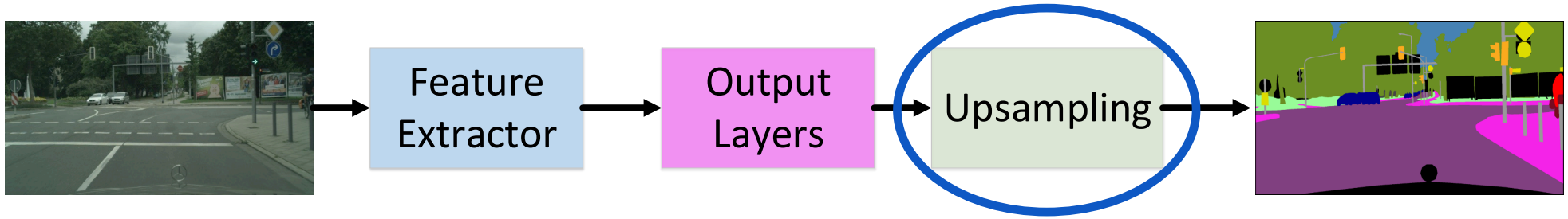


	Image	Conv1	Conv2	Conv3	Conv4
Width	M	M/2	M/4	M/8	M/16
Height	N	N/2	N/4	N/8	N/16
Depth	3	64	128	256	512

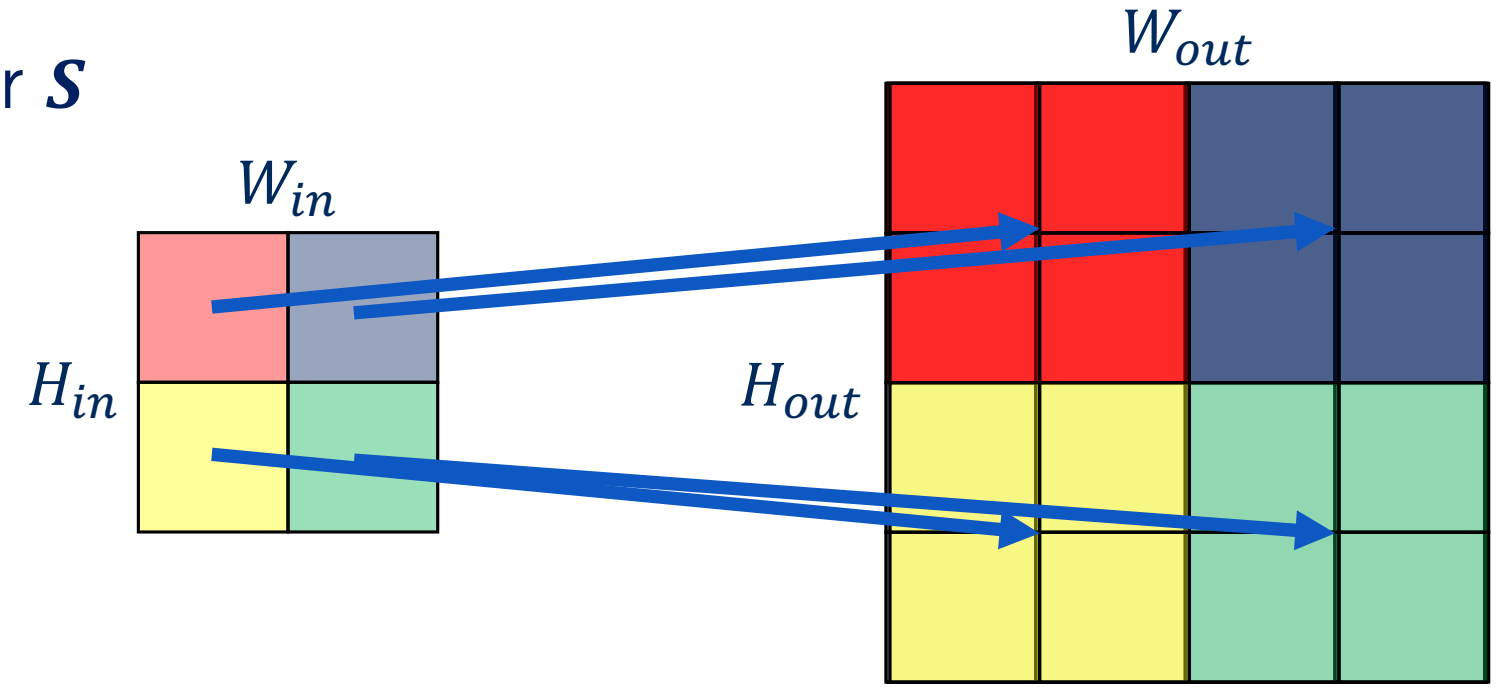
Upsampling the Output



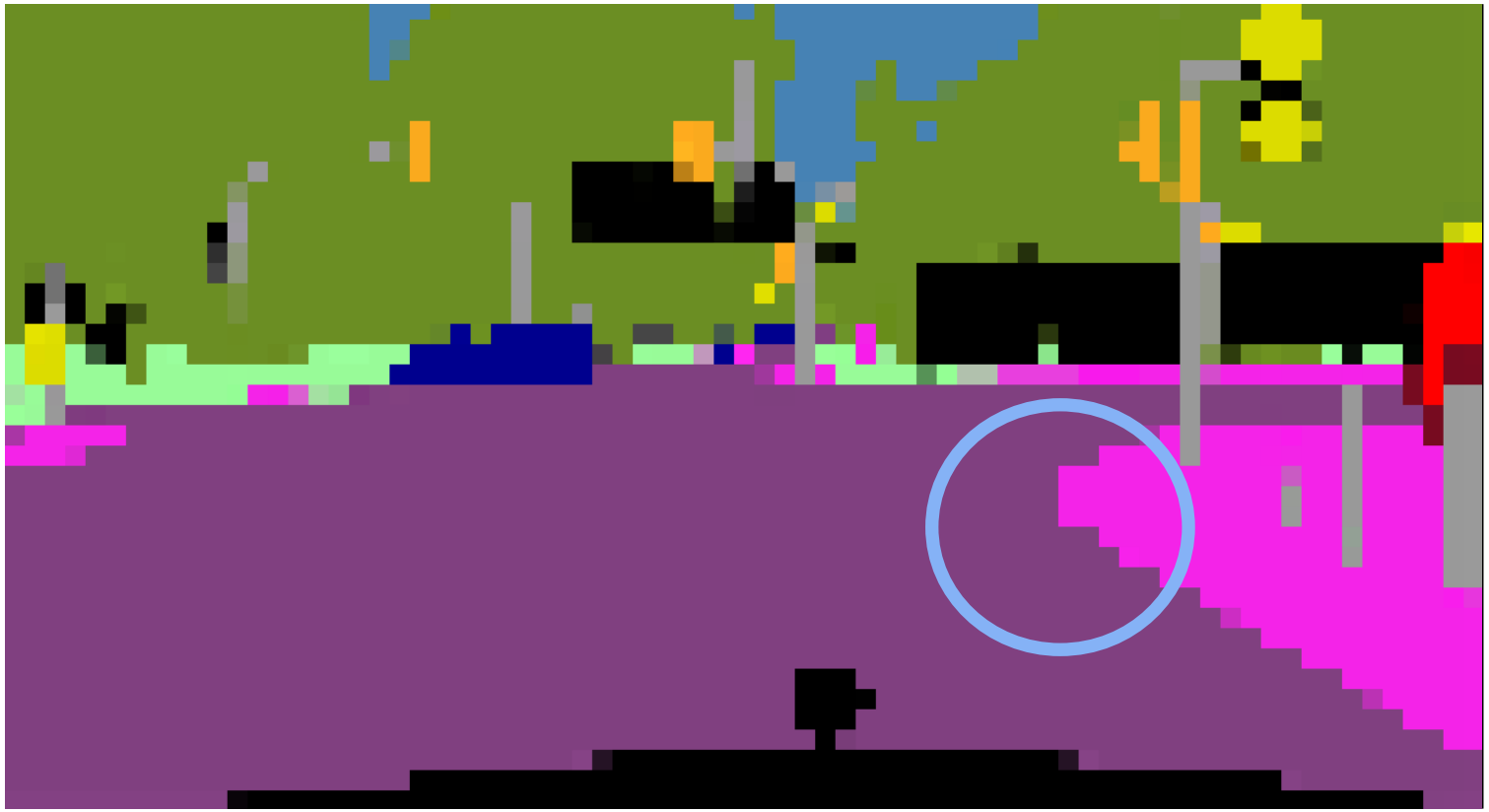
Upsampling Layer

- Upsampling Multiplier $\boldsymbol{S}$

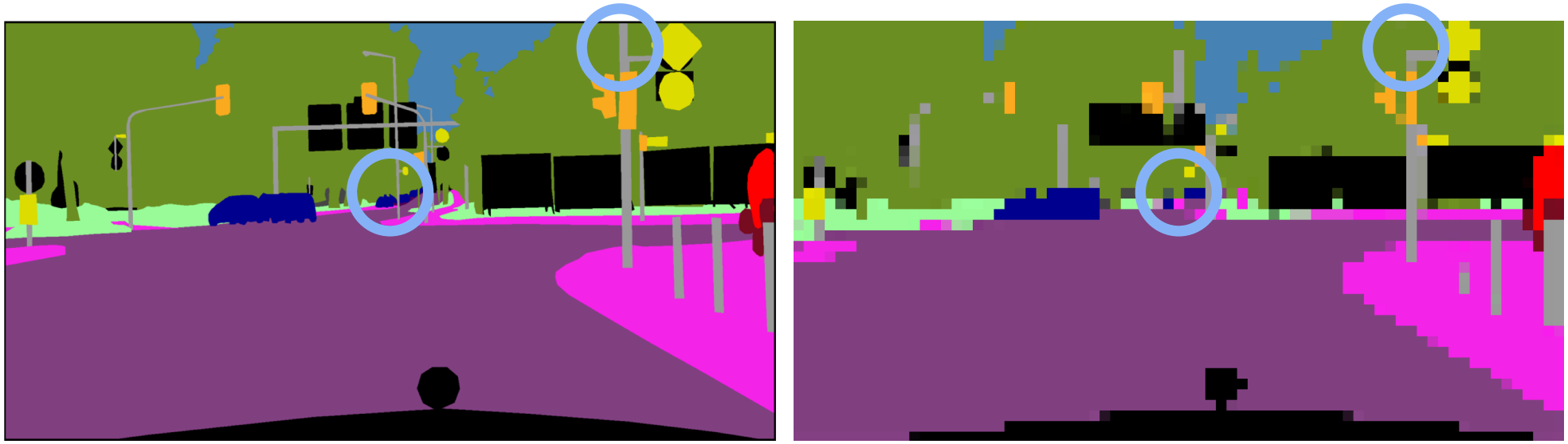
$$\boldsymbol{W_{out}} = S \times \boldsymbol{W_{in}}$$
$$\boldsymbol{H_{out}} = S \times \boldsymbol{H_{in}}$$
$$\boldsymbol{D_{out}} = \boldsymbol{D_{in}}$$



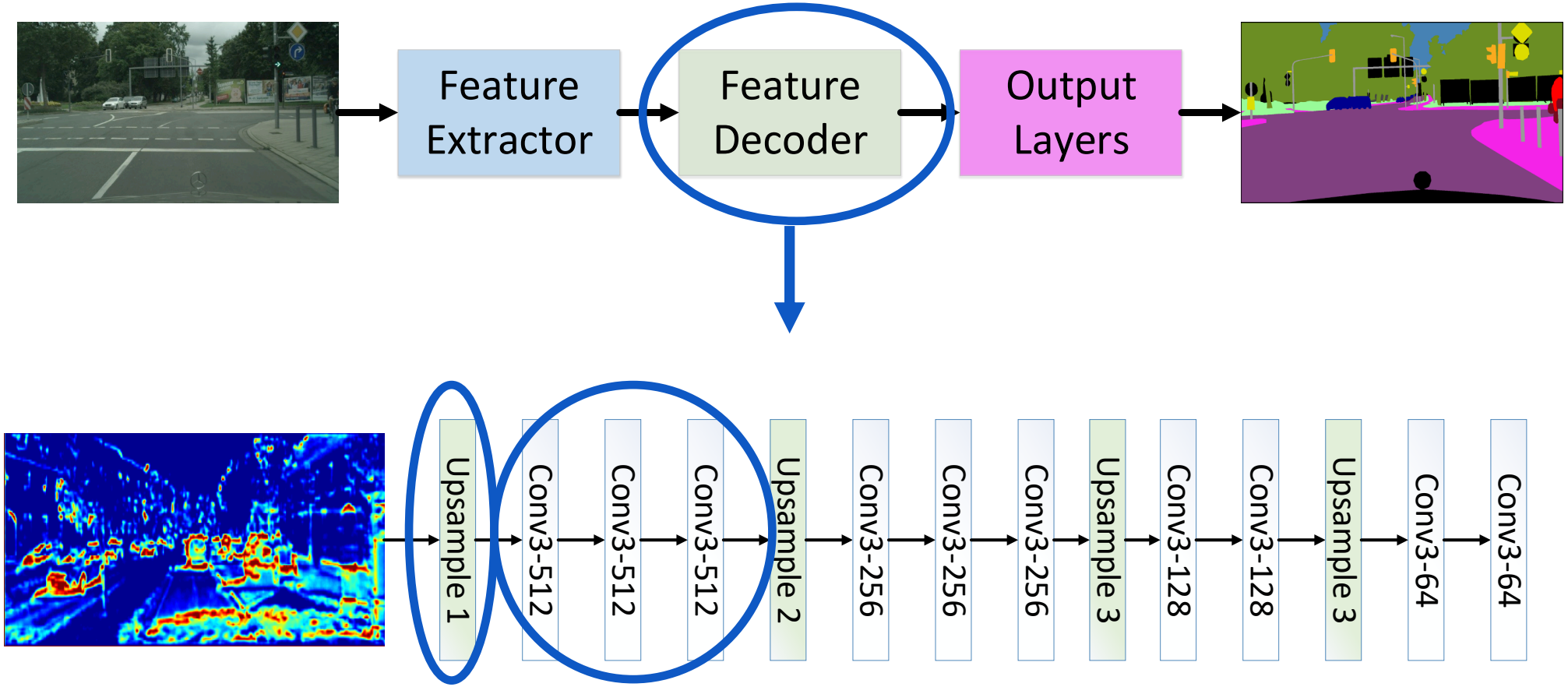
Upsampling The Output



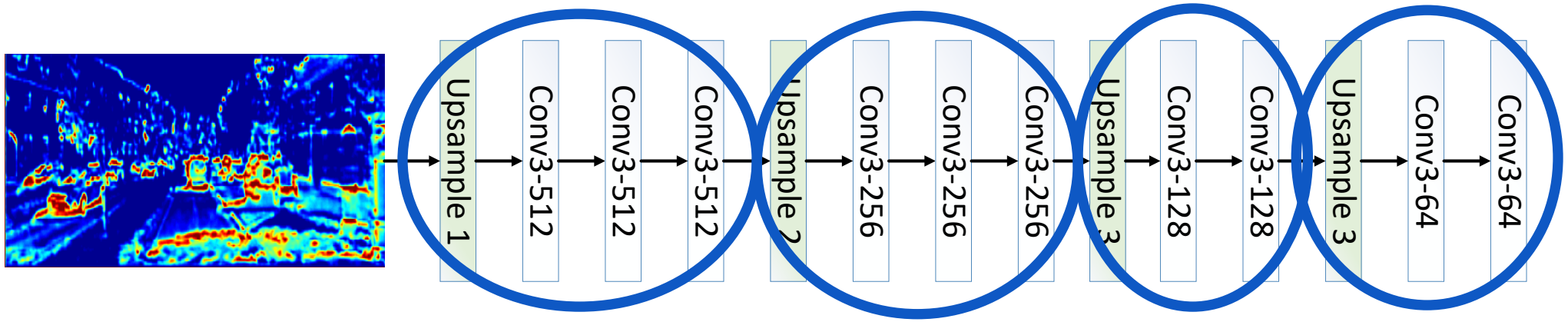
Upsampling The Output



Learning Same Resolution Feature Maps

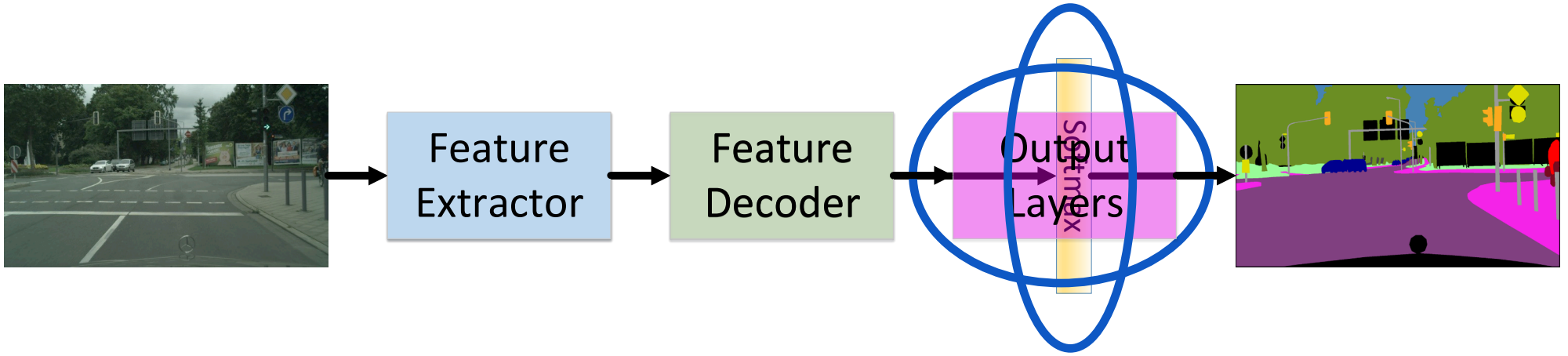


The Feature Decoder

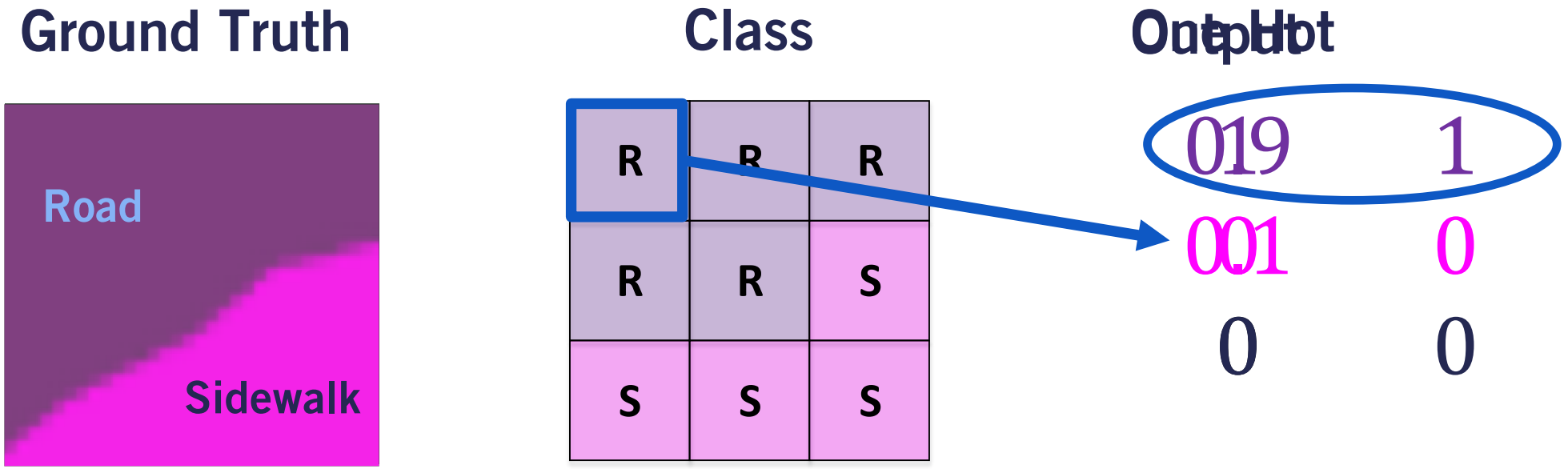


	Feature Map	Deconv1	Deconv2	Deconv3	Deconv4
Width	M/16	M/8	M/4	M/2	M
Height	N/16	N/8	N/4	N/2	N
Depth	512	512	256	128	64

Learning Same Resolution Feature Maps



Output Representation

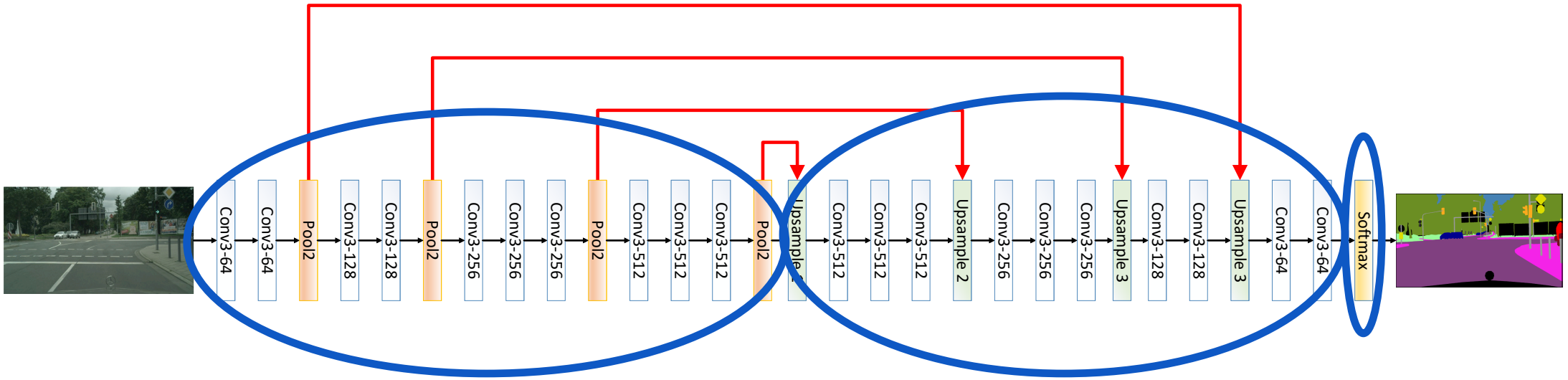


Classification Loss

$$L_{cls} = \frac{1}{N_{total}} \sum_{i} CrossEntropy(s_i^*, s_i)$$

- N_{total} is the number of pixels in all images of our **minibatch**
- s_i is the output of the neural network
- s_i^* is the ground truth classification

ConvNets For Semantic Segmentation



Summary

- Convolutional Neural Networks can be used to solve the semantic segmentation problem
- In a feature extractor and a feature decoder are required to provide the final output of semantic segmentation models

- **Next: Semantic Segmentation For Autonomous Driving**