

# Automated and Connected Driving Challenges

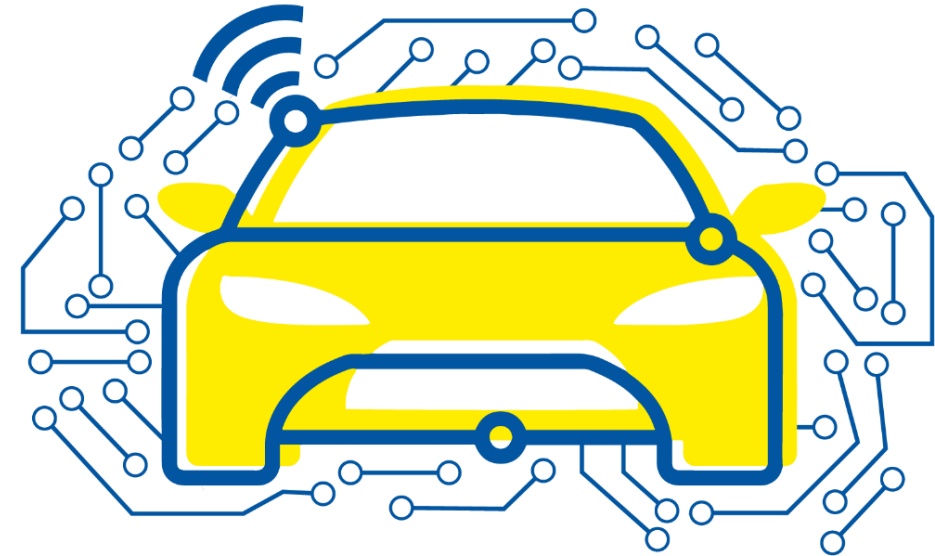
Section 2 – Sensor Data Processing

Semantic Image Segmentation

Deep Learning

Bastian Lampe

Institute for Automotive Engineering





# Semantic Image Segmentation – Deep Learning

## Overview

- Modern **Deep Learning** methods for image segmentation outcompete classical machine learning methods
  - large amounts of training data, powerful hardware, efficient algorithms
- Supervised learning
  - Datasets containing many samples (input + label)
- Inputs and outputs
  - Input: RGB camera image [Height, Width, 3]
  - Output: Segmentation Map [Height, Width, 1]
- Convolutional Neural Networks
  - Encoder-Decoder Architectures with skip connections

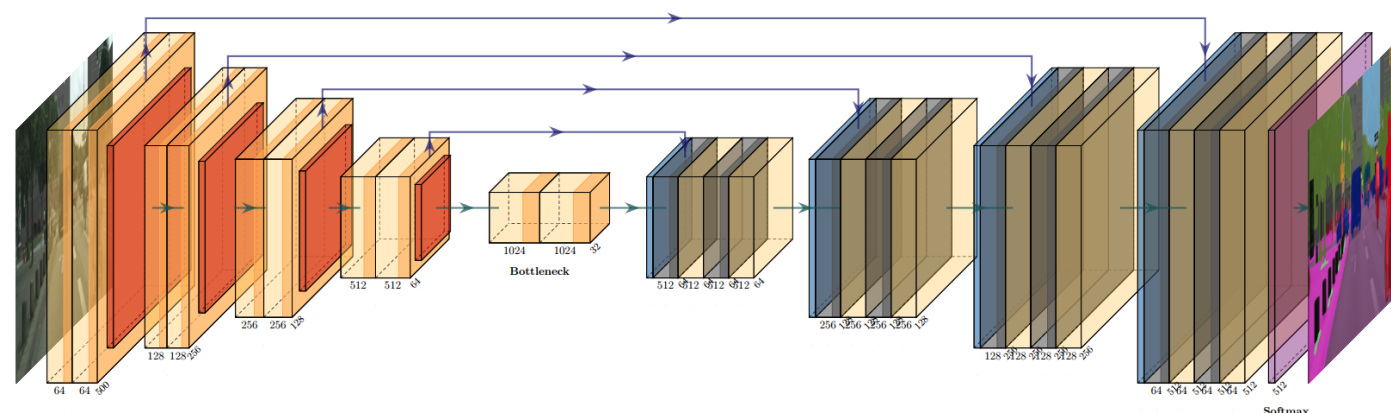


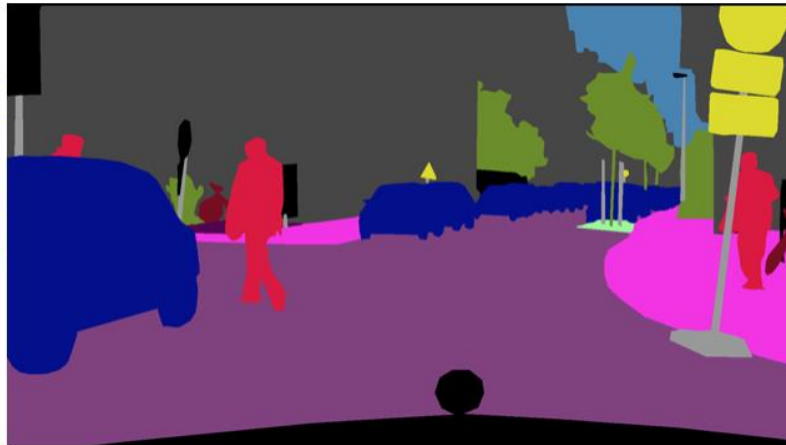
Image: ika, [PlotNeuralNet](#)



# Semantic Image Segmentation – Deep Learning

## *Datasets*

- Cityscapes

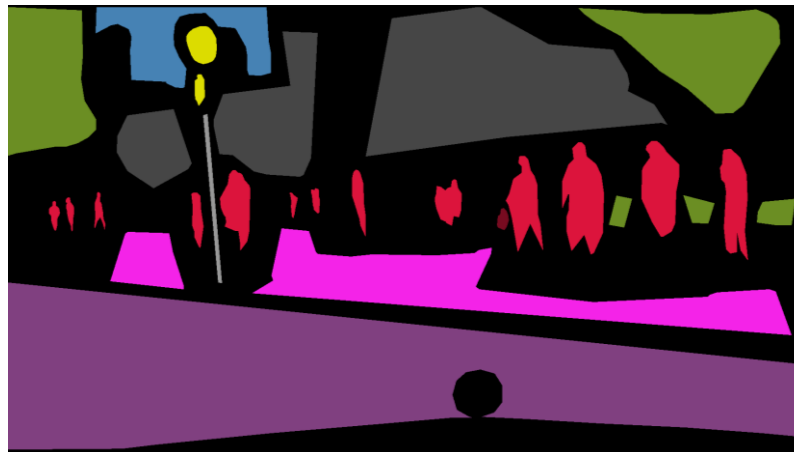
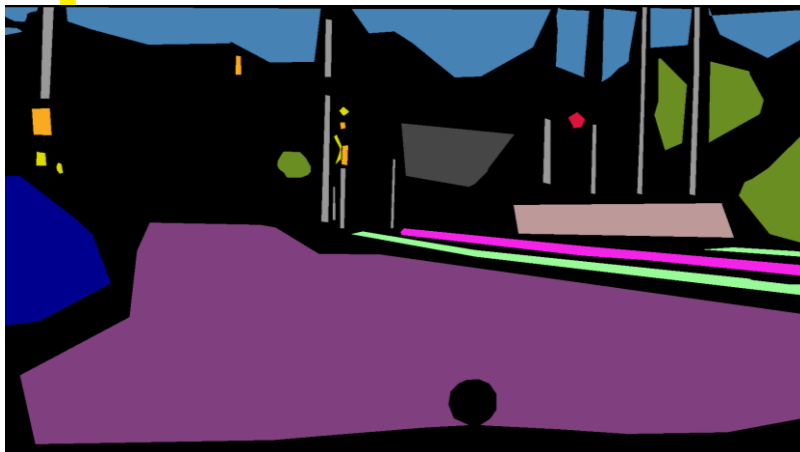




# Semantic Image Segmentation – Deep Learning

## *Datasets*

- Cityscapes Coarse



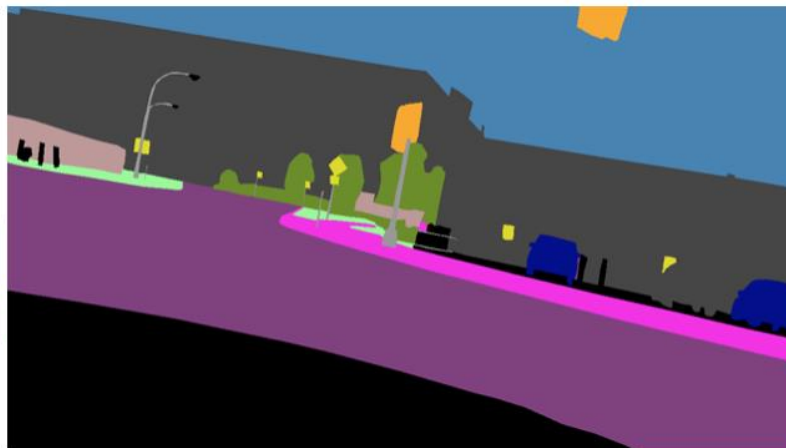
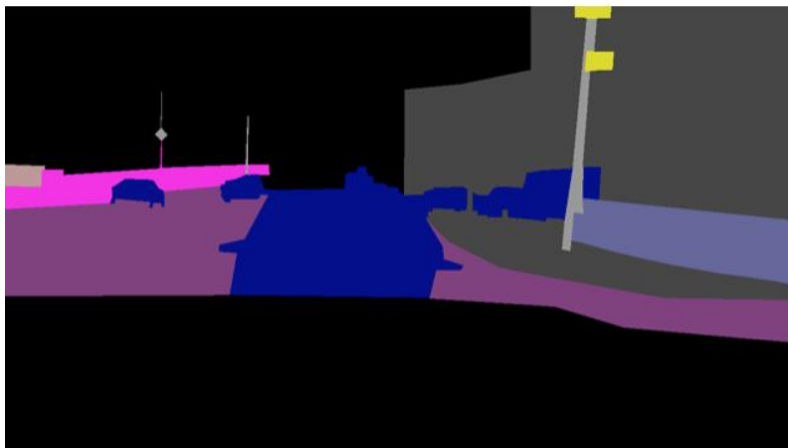




# Semantic Image Segmentation – Deep Learning

## *Datasets*

- Berkeley Deep Drive 100k





# Semantic Image Segmentation – Deep Learning

## *Datasets*

- Mapillary Vista



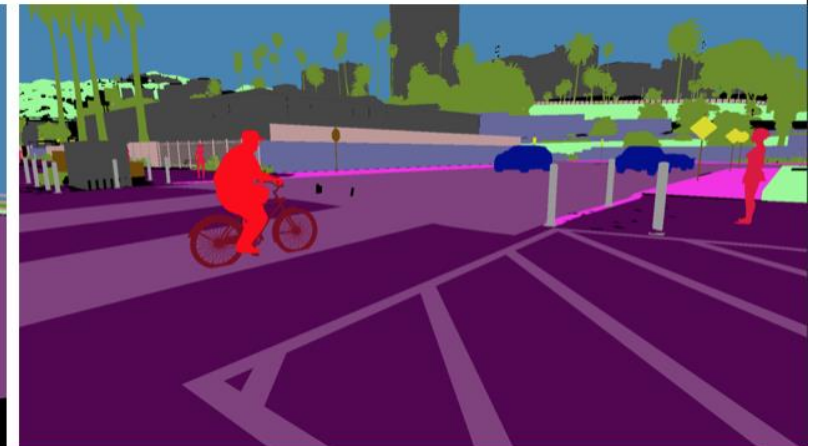




# Semantic Image Segmentation – Deep Learning

## *Datasets*

- GTA





# Semantic Image Segmentation – Deep Learning

## Datasets

### Example Datasets

Dataset	Year	Samples	Places	Scene	Weather	Illumination	Road view	Classes	Resolution
CamVid [BRO08]	2008	700	UK	city	constant	day time	yes	32	960×720
Synthia [ROS16]	2016	213,400	synthetic	city,residential,highway	various	varies	yes	13	960×720
GTA [RIC16]	2016	24,966	synthetic	city,residential,highway	various	varies	partially	30	1914×1052
City [COR16]	2016	2,975/500	Germany	city	constant	day time	yes	29	2048×1024
City Coarse [COR16]	2016	20,000	Germany	city	constant	day time	yes	29	2048×1024
Mapillary [NEU17]	2017	18,000/2,000	Worldwide	city,residential,highway	various	varies	partially	66	3418×2481 †
KITTI [ALH18]	2018	200/200	Germany	city,residential	constant	day time	yes	29	1242×375
BDD100k [YU18b]	2018	7,000/1,000	US	city,residential,highway	various	varies	yes	37	1280×720
ApolloScapes [HUA18]	2018	146,997	China	city,residential,highway	various	day time	yes	25	3384×2710

Source: ika





# Semantic Image Segmentation – Deep Learning

## *Input and Label Representation*

- **Input**
  - RGB Camera Image
  - Shape: [Height, Width, 3]
  - 3 corresponds to the three RGB channels
  - RGB colour representation:
    - As **uint8** with values in [0, 255]
    - As **float16** / **float32** with values in [0, 1]
- **Label**
  - Color Encoding
  - Segmentation Map
  - One-Hot Encoding



Source: [Cityscapes](#)



Source: [Cityscapes](#)

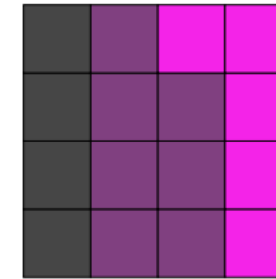


# Semantic Image Segmentation – Deep Learning

## *Input and Label Representation*

- How to represent the label of semantic segmentation ?
  - **Color Encoding**
    - Each class has a specific **RGB Value**
    - Data format: **3 x uint8**
    - Shape: [Height, Width, 3]
  - **Segmentation Map**
    - Each class has a specific **class ID**
    - Data format: **1 x uint8**
    - Shape: [Height, Width, 1]
  - **One-Hot Encoding**
    - A pixel is represented as a **one-hot vector**
    - The  $i$ th value is set to 1 all other values are set to 0. The index  $i$  corresponds to the  $i$ th class
    - Data format: **Number of classes x Boolean**
    - Shape: [Height, Width, Number of Classes]

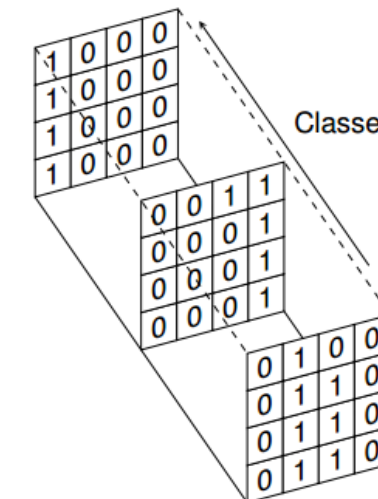
Example with 3 Classes:



Color Encoding

2	0	1	1
2	0	0	1
2	0	0	1
2	0	0	1

Segmentation Map



One-Hot-Encoding



# Semantic Image Segmentation – Deep Learning

## *Network Architecture*

- **Fully-Convolutional Network**
- Network as a sequence of convolutional layers
- Make class predictions for all pixels **at once**

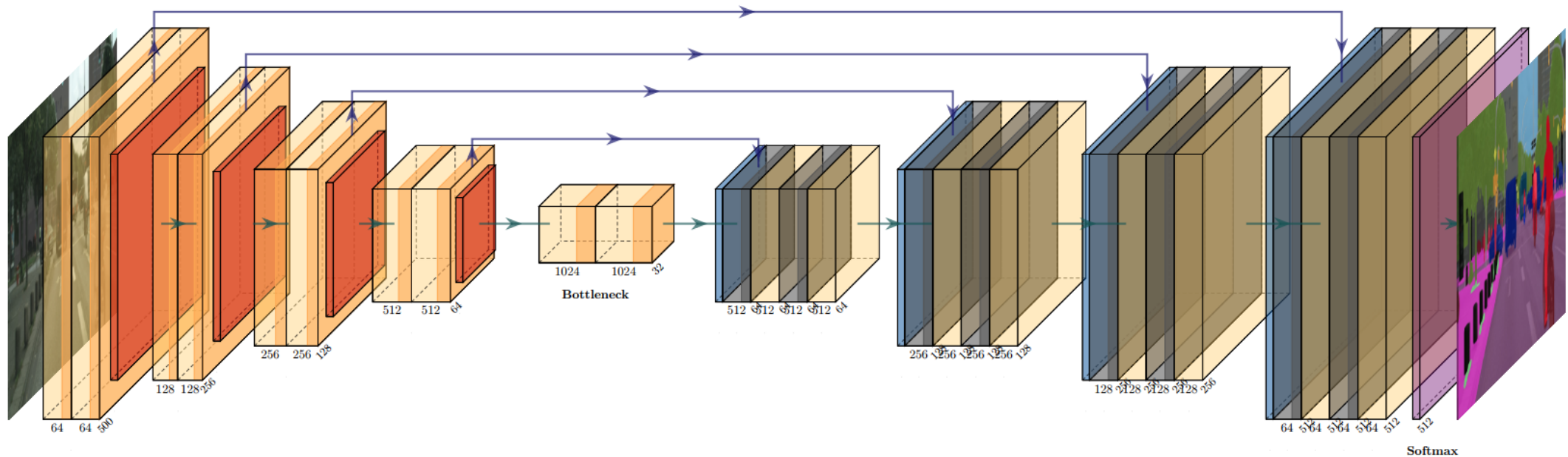


Image: ika, [PlotNeuralNet](#)





# Semantic Image Segmentation – Deep Learning

## Summary

- Several different publicly available **datasets**
- Datasets consists of **camera image** and **label** pairs
- Label can be represented as
  - **RGB encoding**
  - **Segmentation Map**
  - **One-Hot Encoding**
- **Fully-Convolutional Network** architectures are used to learn representations from these datasets



Images: Mapillary Vista

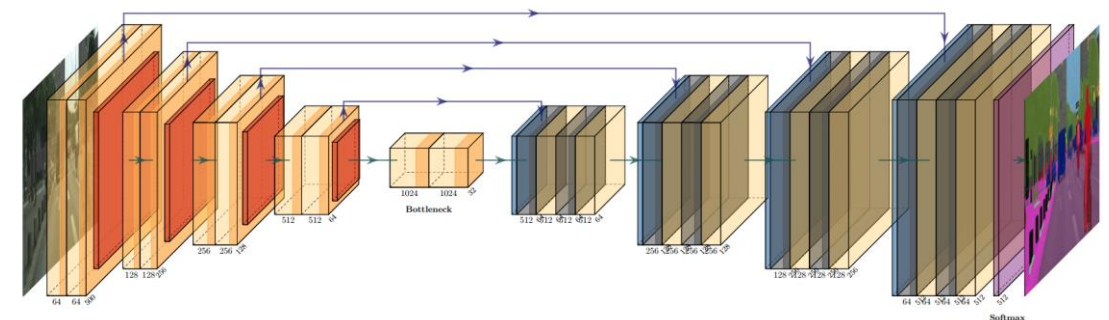


Image: ika, [PlotNeuralNet](#)