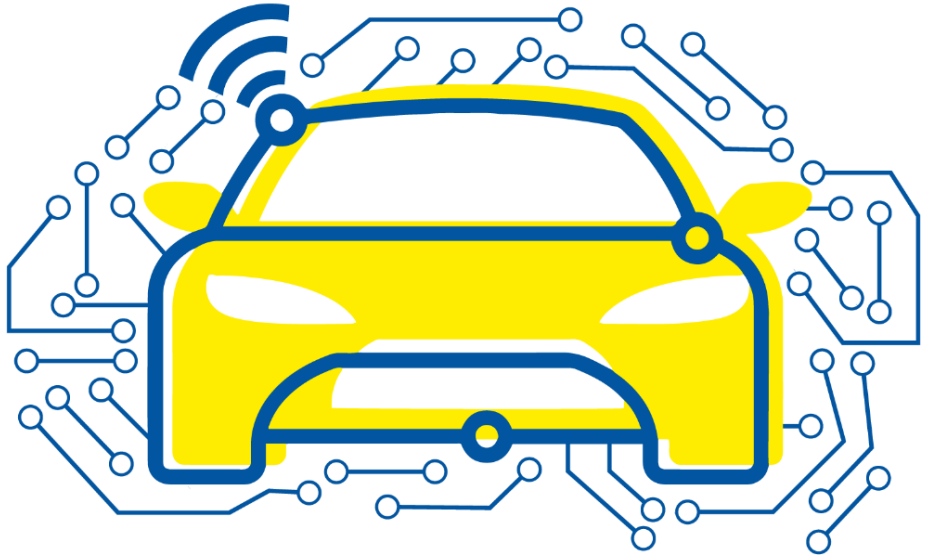
Automated and Connected Driving Challenges

Section 2 – Sensor Data Processing

Semantic Image Segmentation Training



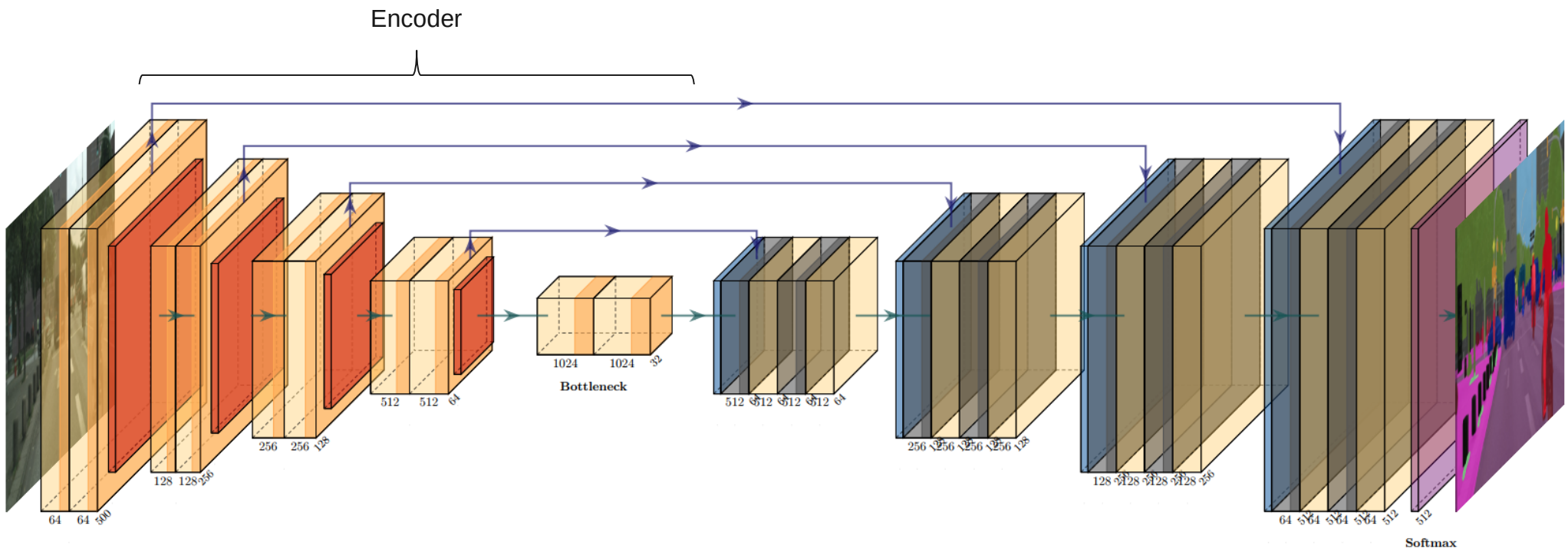
Bastian Lampe

Institute for Automotive Engineering

Semantic Image Segmentation – Training

Network Architecture Details

- **Downsampling (Encoder)**
  - Convolutions with stride and padding
  - Pooling Operations



Source: ika

Semantic Image Segmentation – Training

Network Architecture Details

- **Upsampling (Decoder)**
  - Unpooling Operations
  - Transpose Convolutions with stride

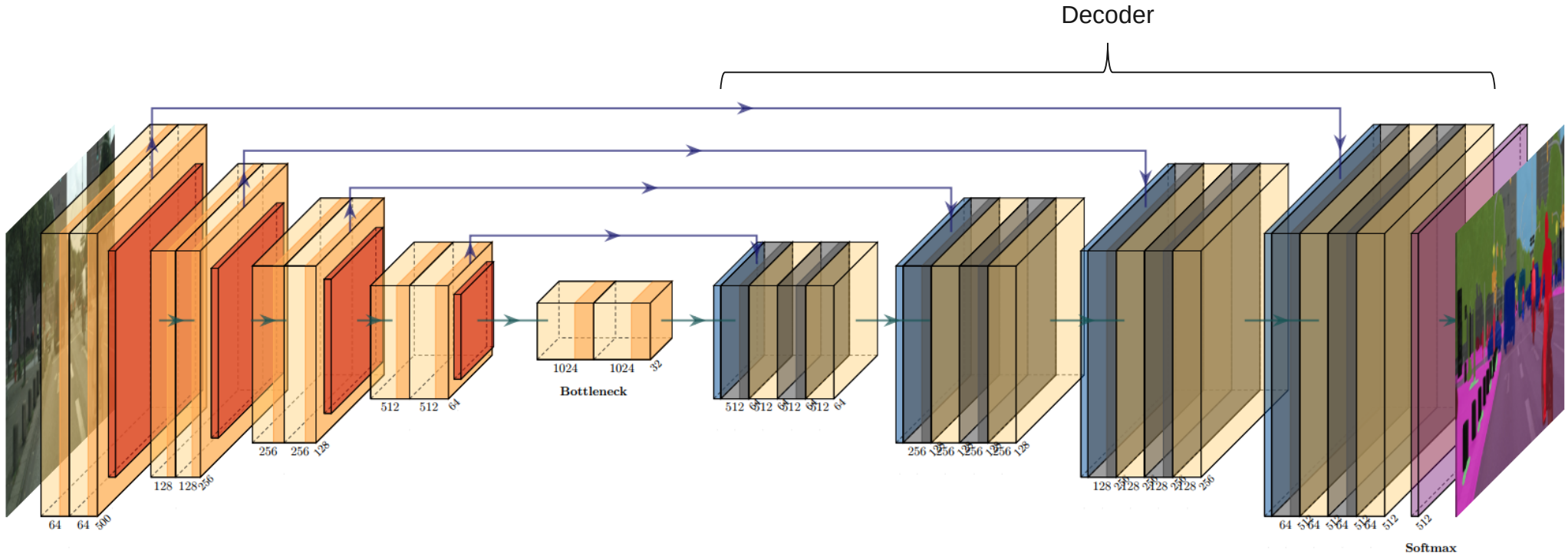


Image: ika, PlotNeuralNet

Semantic Image Segmentation – Training

Network Architecture Details

- **Skip Connections**
  - Copies the intermediate representations from encoder layers to decoder layers
  - Downsampling loses high-resolution information
  - Use skip connections to preserve this higher-resolution information

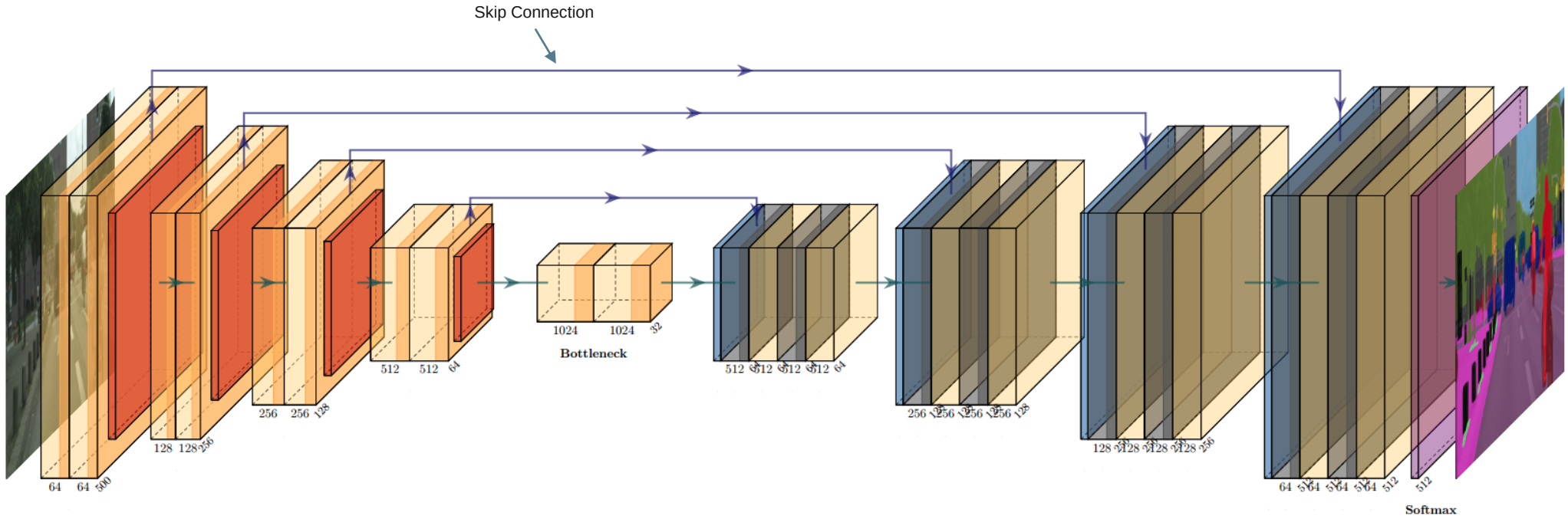


Image: ika, PlotNeuralNet

Semantic Image Segmentation – Training

Softmax

- **Softmax Activation Function**
  - Final prediction layer
  - Input: Computed „logits“ from the network
  - Computes the class probabilities for each pixel
  - Output Shape: [Height, Width, Number of Classes]

$$\mathbf{Softmax}(x_i) = \frac{\exp(x_i)}{\sum_j \exp(x_j)}$$

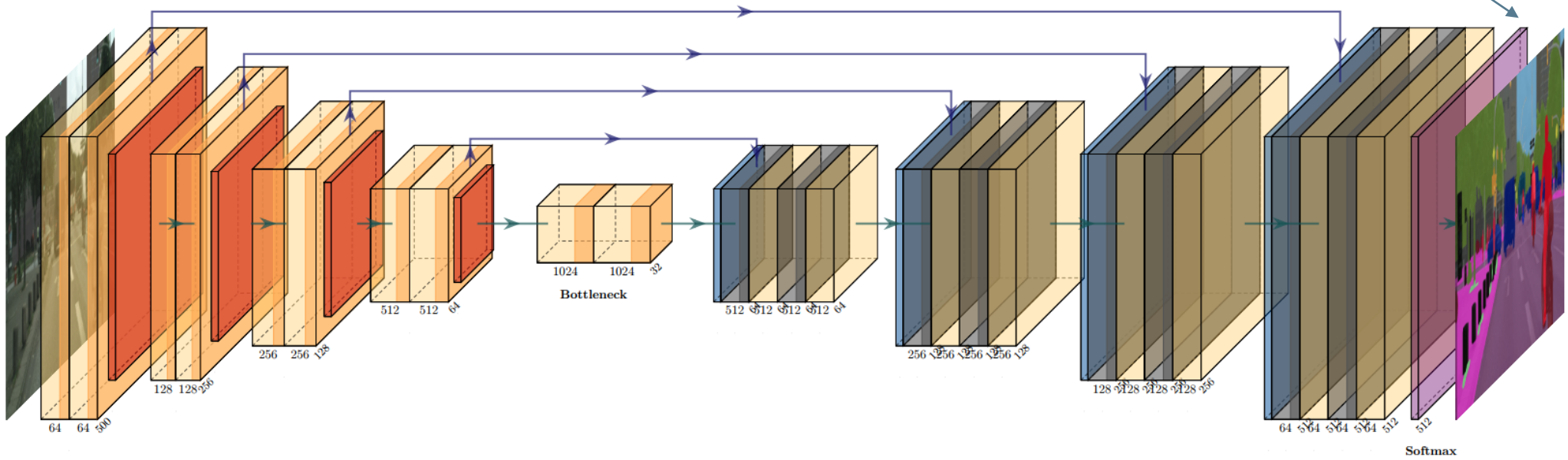


Image: ika, PlotNeuralNet

Semantic Image Segmentation – Training

Loss Function

- **Loss Function**
  - Computes the **error** between model output and ground truth label
  - Multi-Class **Categorical Cross-Entropy**
  - Input: **Probabilities from the Softmax activation** and **Ground Truth label**

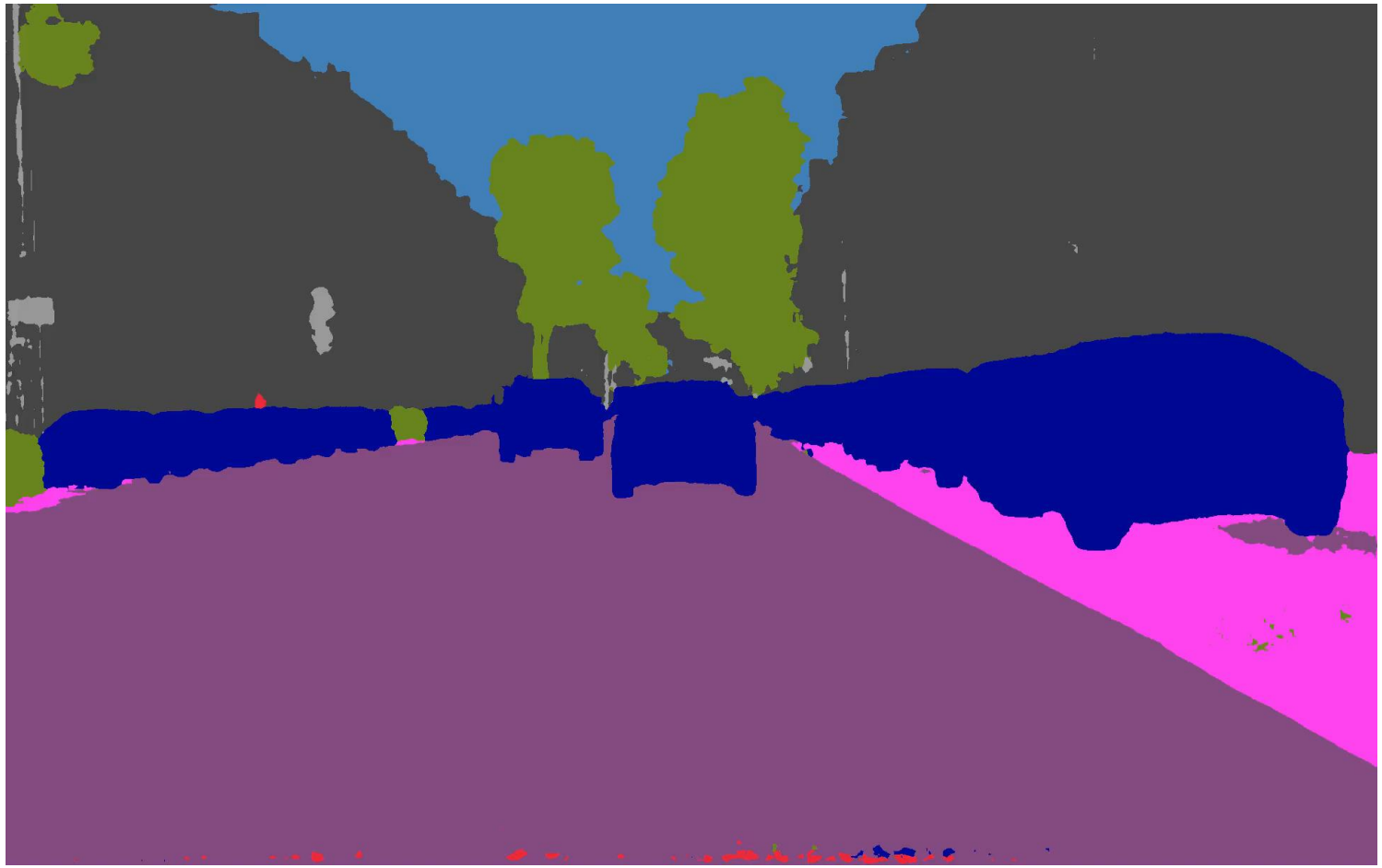
$$CE = -\sum_{i}^{C} t_i \log(p_i)$$

  - with t_i as the ground truth label (as one-hot encoding) and p_i as probability for each class i in C (set of all classes)
  - Compute this error for all pixels
  - Use this error and **propagate it back** to the network
  - Determine the gradients and use gradient descent to **train the network**

Semantic Image Segmentation – Training

Training

- **Hyperparameters**
  - Batch size
  - Epochs
  - Number of filters
  - Learning rate
  - Input image size
  - ...
- **Dataset**
  - E.g. Cityscapes
    - 3000 Training Samples
    - 500 Validation Samples
  - Test Video: Aachen downtown
- **Inference**
  - E.g. 78 % MIoU on the validation data with a pretrained Xception network

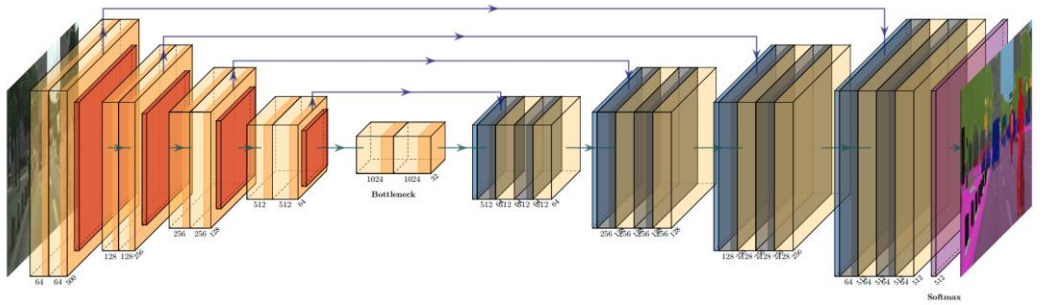


Video: ika

Semantic Image Segmentation – Training

Summary

- Network architecture
  - **Encoder**: downsampling of the input
  - **Decoder**: upsampling of the intermediate representations
  - **Skip Connections**: Preserve higher-resolution information
  - **Softmax Activation**: Compute class probabilities for each pixel



Source: ika

- Loss Function
  - **Categorical Cross-Entropy:** Compute the classification error

$$CE = -\sum_{i}^{C} t_i \log(p_i)$$

- Training
  - The network is trained with **Gradient Descent**
  - Many different **hyperparameters** can be tuned