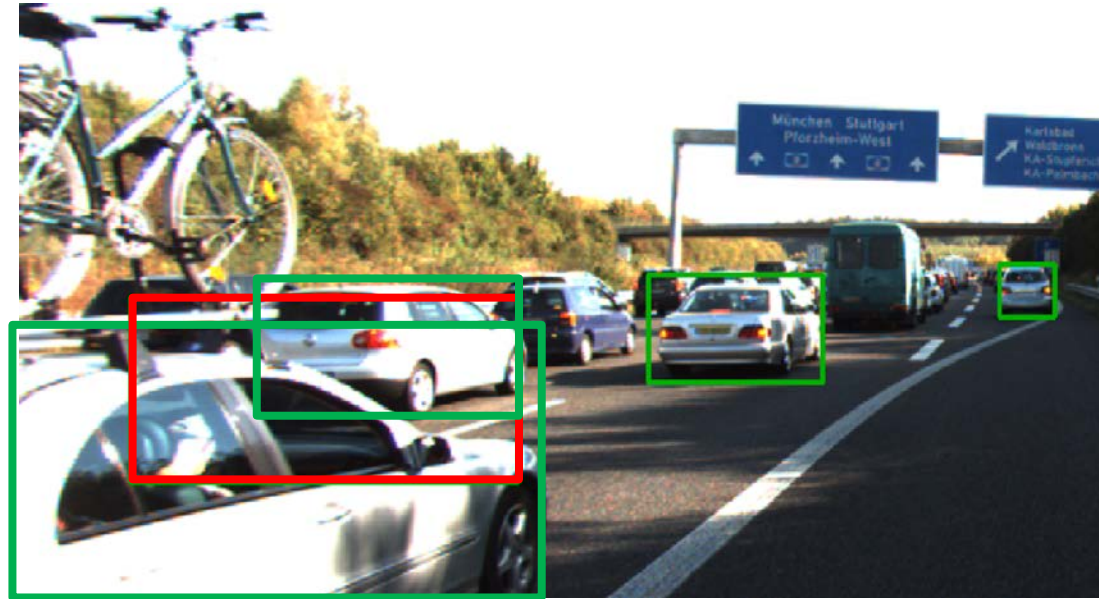


Lab Course Cognitive Systems: Depth & Bounding Box Prediction

Interim Presentation
WS 17/18

Problem Definition

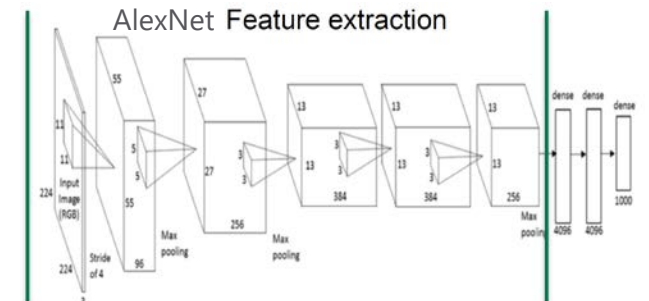
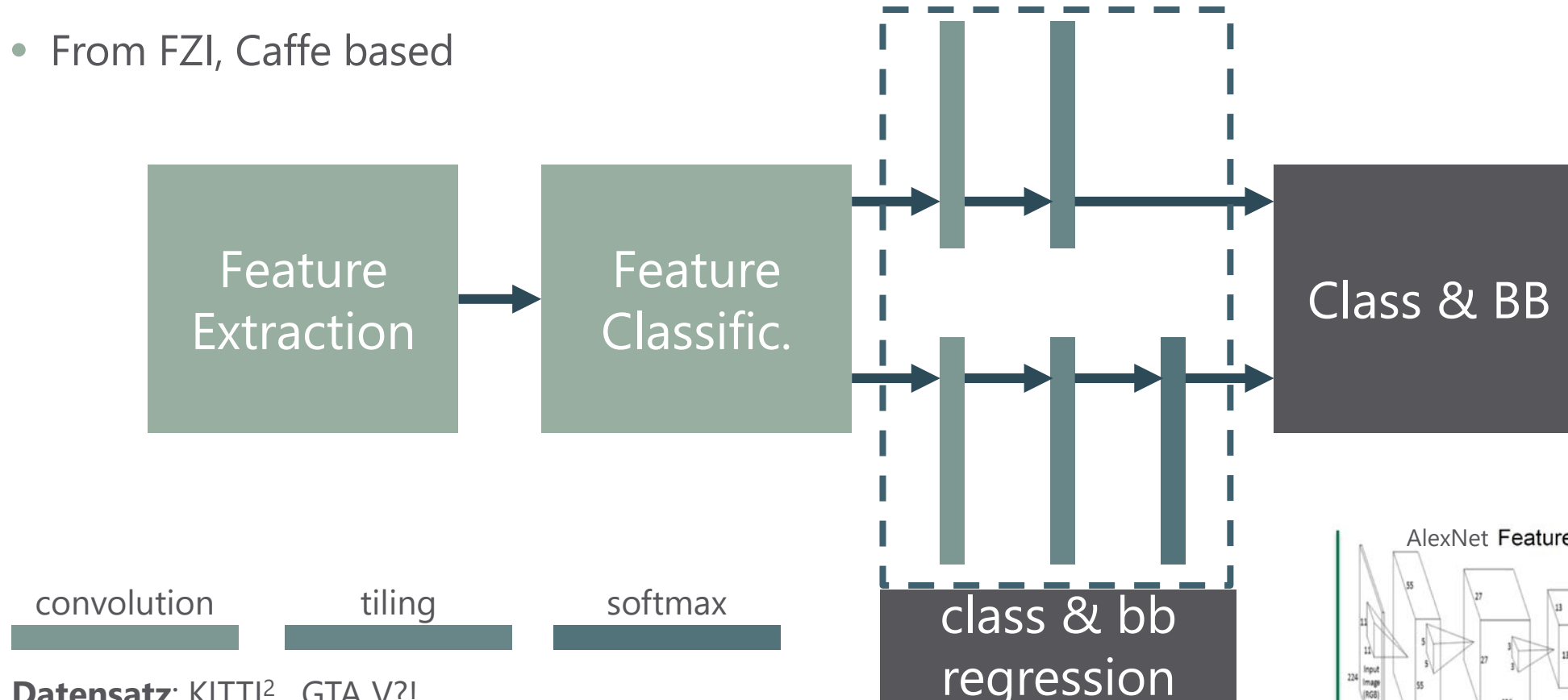
- Vehicle detection with CNN
- Issues: poor performance for largely overlapping vehicles
- Possible solution: using depth information



1. Concept

Learning from depth images for monocular object detection with convolutional neural networks, MA A. Lesi

- DeepTLR¹ based object detection - class and bounding box regression
- From FZI, Caffe based



Datensatz: KITTI², GTA V?!

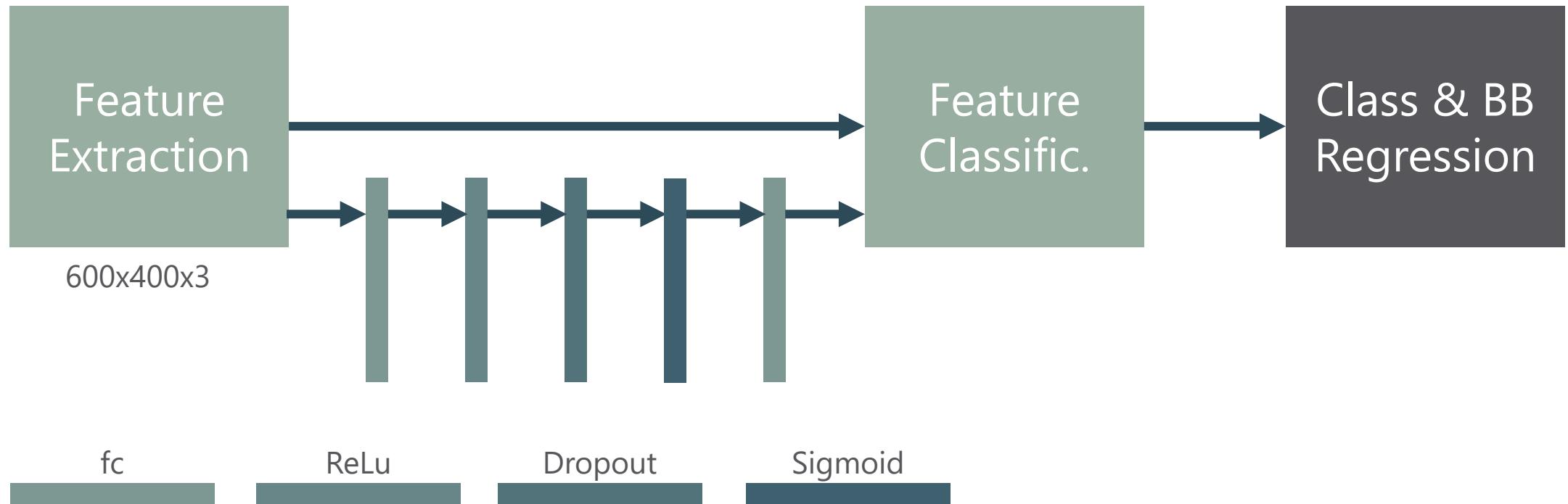
¹<http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7535408>

²<http://www.cvlbbs.net/publications/Geiger2012CVPR.pdf>

1. Concept

Learning from depth images for monocular object detection with convolutional neural networks, MA A. Lesi

- Joint convolution depth prediction and object detection
- Against using of superpixels



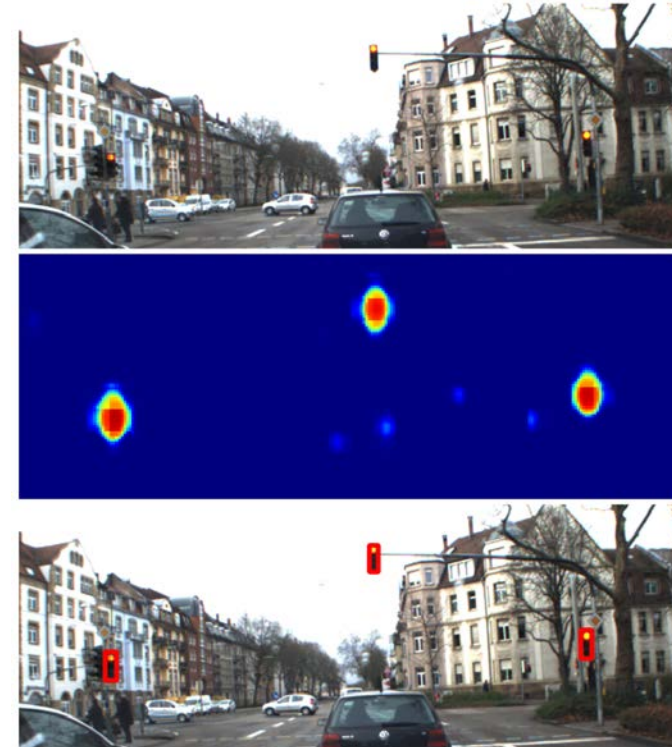
1. Concept

Learning from depth images for monocular object detection with convolutional neural networks, MA A. Lesi

- Current State:

- ✓ Caffe running
 - ✓ Data processing & augmentation
 - ✓ LMDB creation
 - ✗ First training on KITTI
 - ✓ Evaluation routines
-
- Using code from M. Weber and A. Lesi
 - No modifications on framework planned

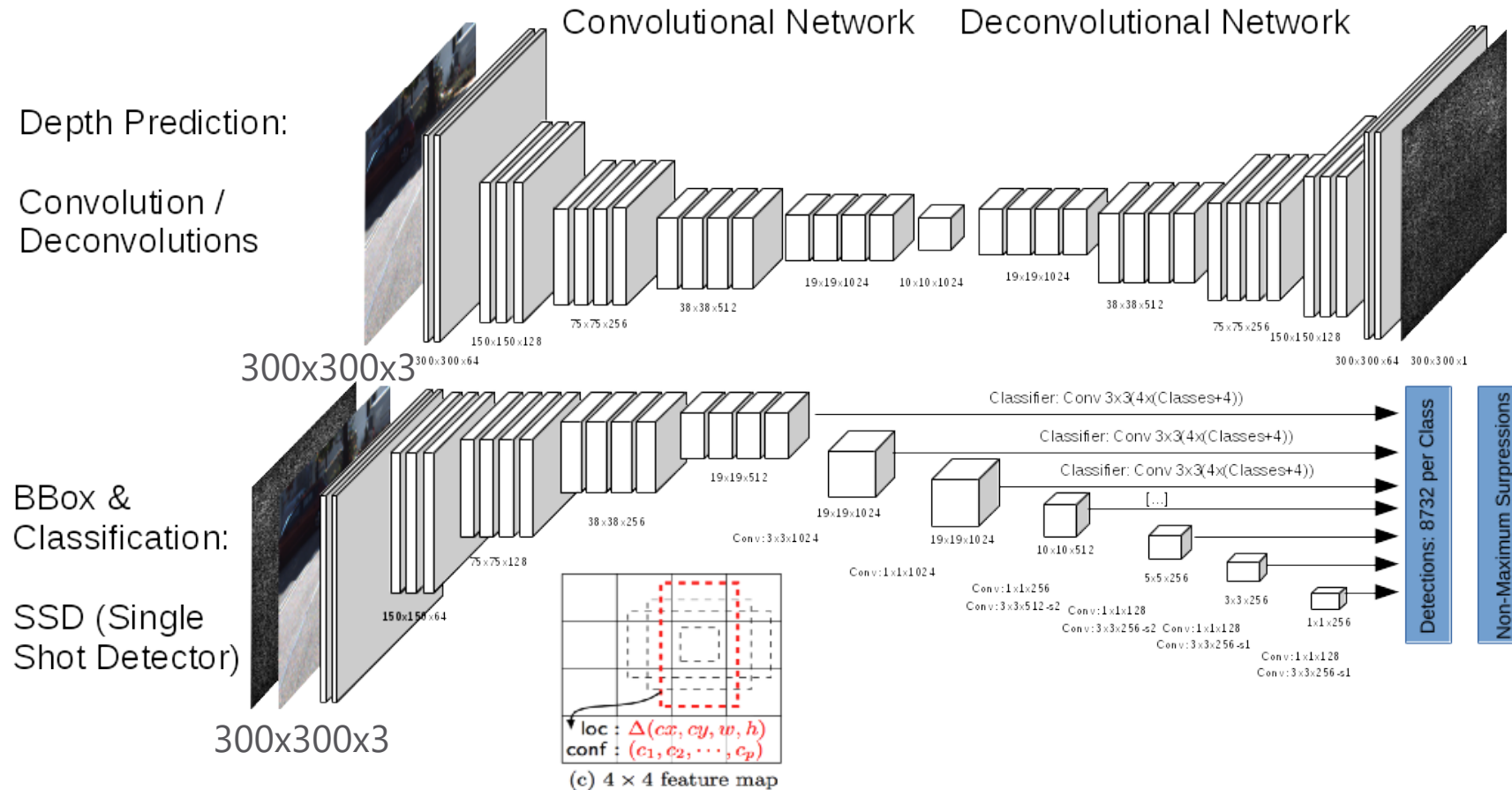
Caffe



2. Concept

Deeper Depth Prediction with Fully Convolutional Residual Networks, Laina et al.
+ SSD: Single Shot MultiBox Detector, Liu et al.

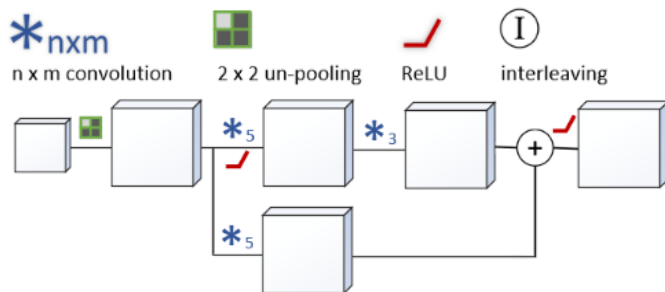
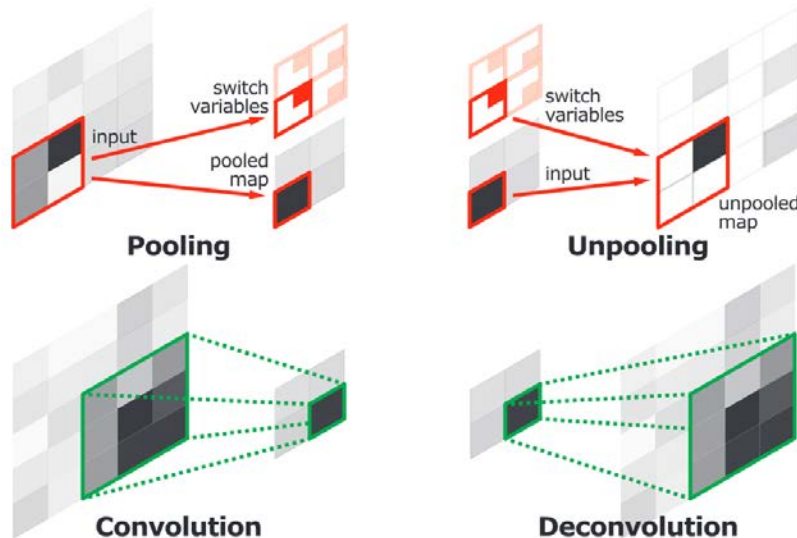
- 2 Stages: DeconvNet with ResNet50 layers for depth prediction + SSD detector



2. Concept

Deeper Depth Prediction with Fully Convolutional Residual Networks, Laina et al.
+ SSD: Single Shot MultiBox Detector, Liu et al.

- Using pooling and unpooling layers



Pooling: is designed to filter noisy activations in a lower layer by abstracting activations in a receptive field with a single representative value

- spatial information within a receptive field is lost during pooling, which may be critical for precise localization

Unpooling: performs the reverse operation of pooling and reconstruct the original size of activations

- increases the spatial resolution of feature maps

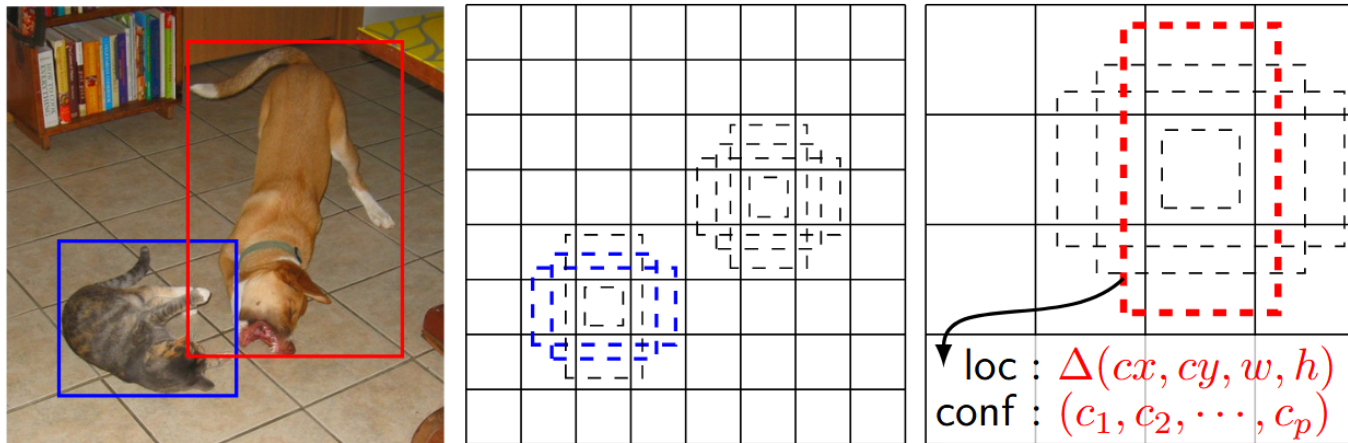
Up-projection: adds 3×3 convolution after the unpooling and a projection (=res-blocks) connection from the lower resolution feature map to the result

- allows high-level information to be more efficiently passed forward in the network while progressively increasing feature map sizes

2. Concept

Deeper Depth Prediction with Fully Convolutional Residual Networks, Laina et al.
+ SSD: Single Shot MultiBox Detector, Liu et al.

- Object detection based on VGG16
- Uses predefined anchor boxes with different sizes and ratios
- For every box it will be calculated if object is in box or near it, the object class and the bb



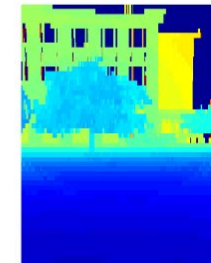
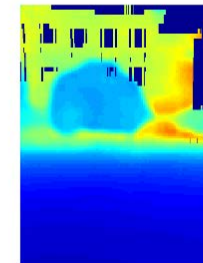
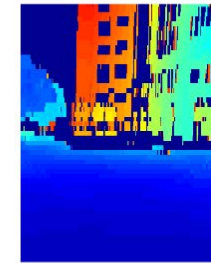
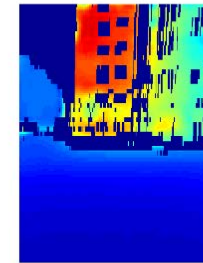
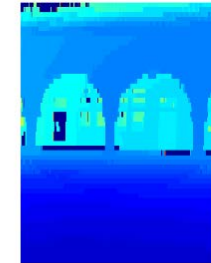
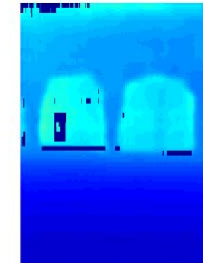
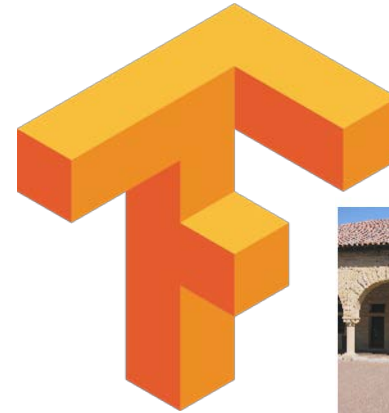
- Unused boxes will be deleted and non-maximum suppression is used to obtain final boxes
- Faster than YOLO!

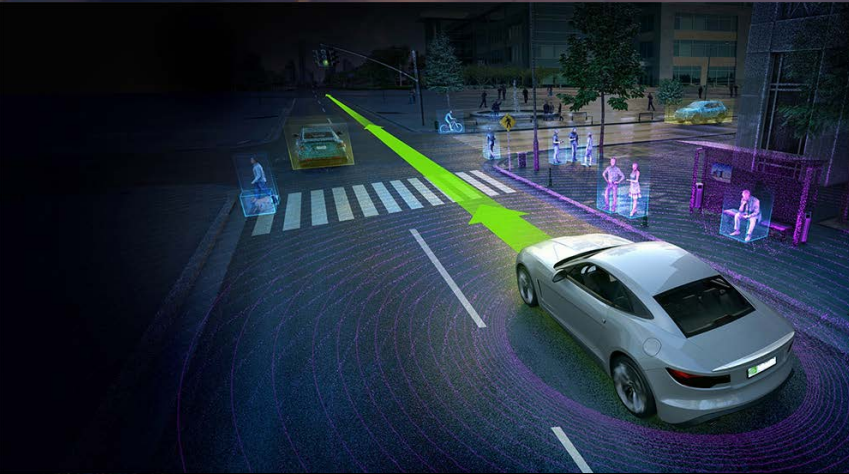
2. Concept

Deeper Depth Prediction with Fully Convolutional Residual Networks, Laina et al.
+ SSD: Single Shot MultiBox Detector, Liu et al.

- Current State:

- ✓ Up and running isolated frameworks
 - ✗ When and where combine outputs – inputs?!
 - ✓ Data processing & augmentation
 - ✗ First training on KITTI
 - ✓ Evaluation routines
-
- Both networks use VGG for feature extraction
 - Both are implemented in Tensorflow 🤖





Questions?