# LIDAR-Camera Fusion for Road Detection Using Fully Convolutional Neural Networks

Luca Caltagirone, Mauro Bellone, Lennart Svensson, Mattias Wahde

Jingyuan Dai

- Motivation and Contribution
- Data
- Methodology and Result
- Conclusion

# Motivation and Contribution

- Current approaches for road detection use either cameras(image) or LIDAR sensors(point cloud). However, using the two devices solely set limitations for road detection.
- Camera – provide denser information; affected by lighting conditions
- LIDAR – provide distance measurements, less affected by the external lighting condition; sparse data

This paper proposes

- A deep learning approach to fuse LIDAR point clouds and camera images for road detection. Early, late, and cross fusion are tested using KITTI road benchmark dataset.
- A sub dataset(33 images used for validation) of visually challenging scenes extracted from KITTI driving sequences to carry out road segmentation, and now become part of the KITTI road detection dataset

# Data

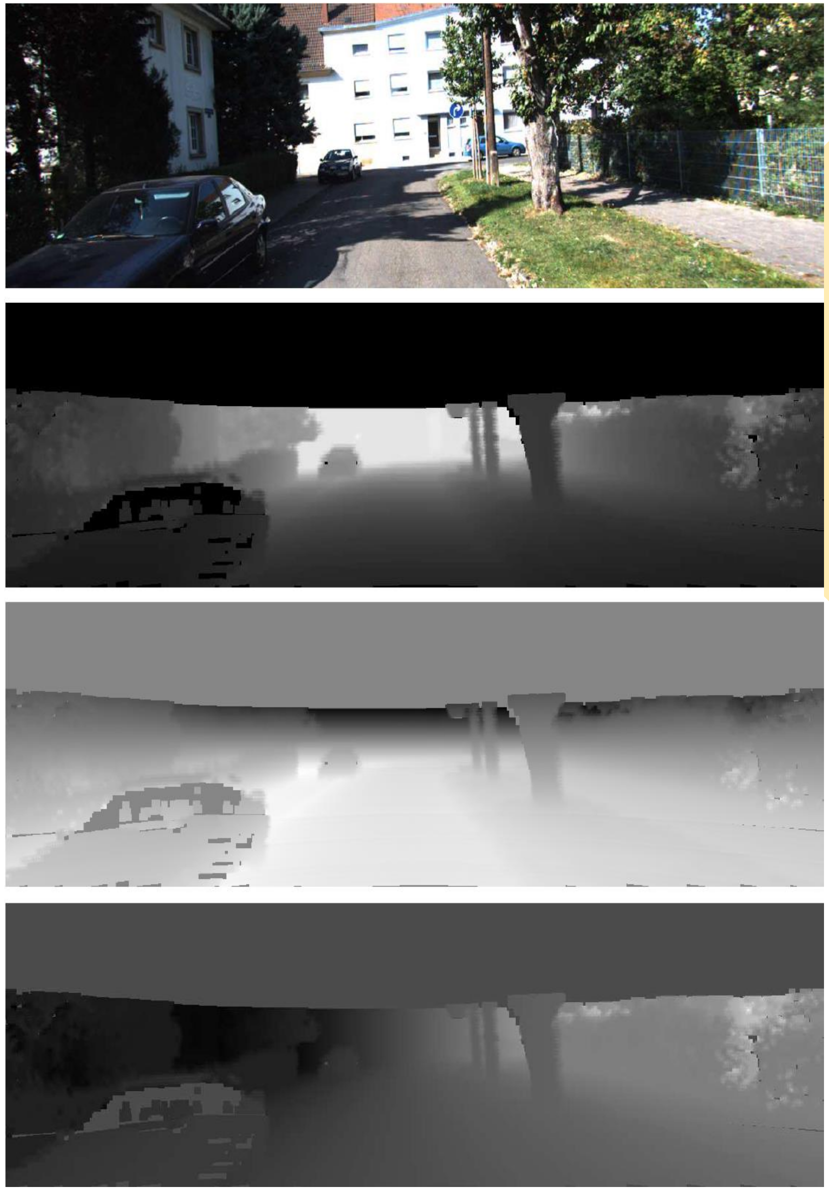
- Data obtained from the KITTI road detection benchmark(http://www.cvlibs.net/datasets/kitti/eval_road.php)

# Methodology

- First, 3D point cloud data is projected into image plane, then upsample those images
- Given the LIDAR-camera transformation matrix **T**, the rectification matrix **R**, and the camera projection matrix **P**, calculate the column position ***u***, and the row position ***v***

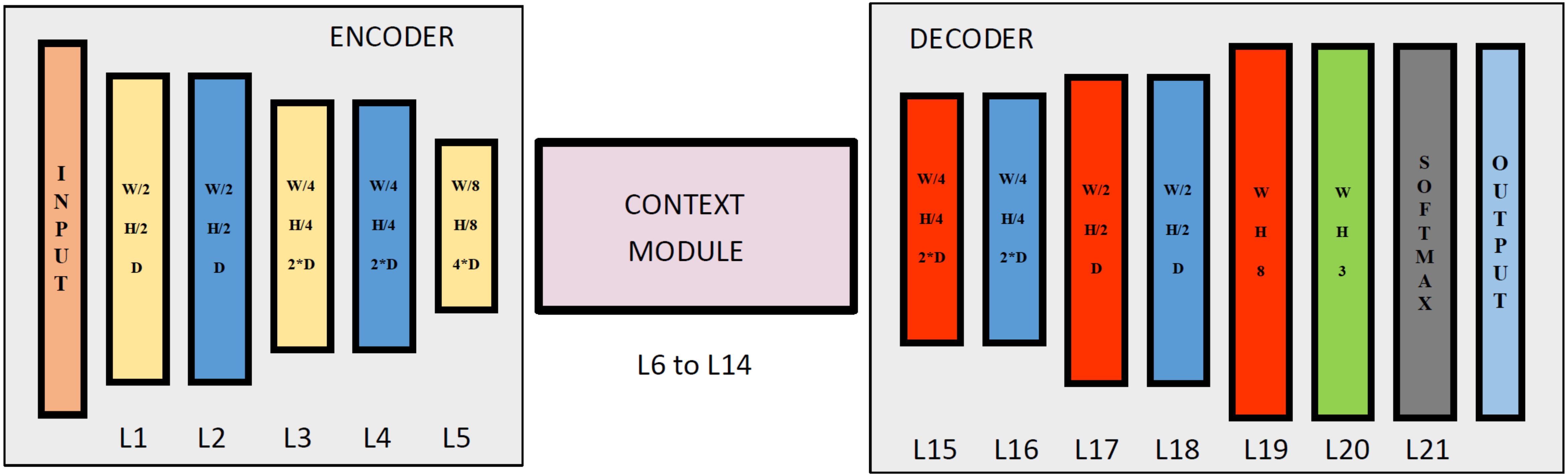
$$\lambda\,[u, v, 1]^{\mathrm{T}} = \mathbf{P}\,\mathbf{R}\,\mathbf{T}\,p$$

- X, Y, and Z where each pixel contains the x, y, and z coordinates of the 3D point that is projected into it



# Methodology

- Second, a fusion FCN (based) is proposed.




Convolution 3x3, stride 1, zero-padding 1+ ELU

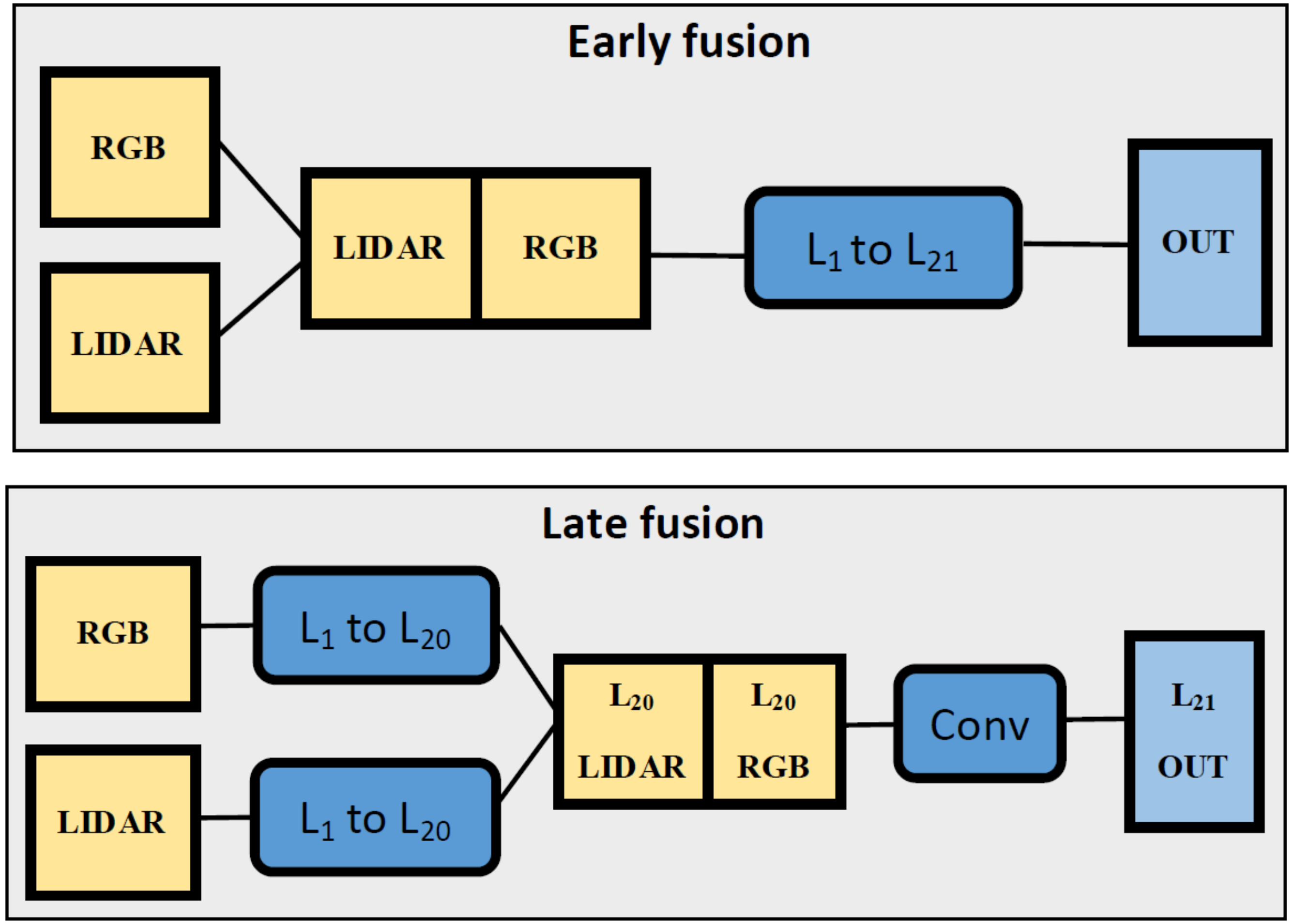
Convolution 4x4, stride 2, zero-padding 1+ ELU

Deconvolution 4x4, stride 2, zero-padding 1+ ELU

Convolution 3x3, stride 1, zero-padding 1

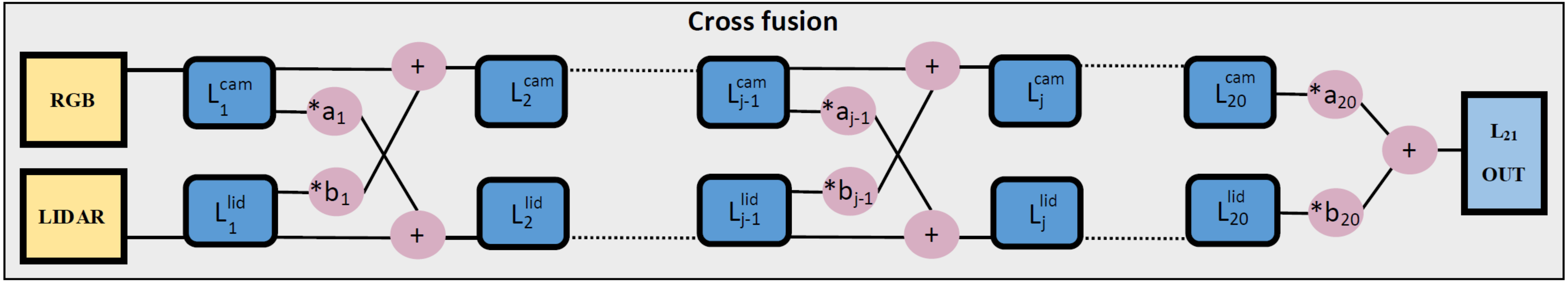
# Methodology

- Third, five different fusion methods are tested - ZYX and RGB image
- Early fusion and Late fusion



# Methodology

- Cross fusion




$$I_j^{\text{Lid}} = L_{j-1}^{\text{Lid}} + a_{j-1} L_{j-1}^{\text{Cam}}$$

$$I_j^{\text{Cam}} = L_{j-1}^{\text{Cam}} + b_{j-1} L_{j-1}^{\text{Lid}}$$

# Result

TABLE III

PERFORMANCE COMPARISON OF SINGLE MODALITY AND FUSION FCNs EVALUATED ON THE VALIDATION SET.

| Fusion strategy | # param. | MaxF [%] | PRE [%] | REC [%] |
|---|---|---|---|---|
| ZYX | 1623395 | 94.96 | 94.05 | 95.89 |
| RGB | 1623395 | 96.00 | 96.16 | 95.84 |
| Early fusion | 1624931 | 95.41 | 94.62 | 96.21 |
| Late fusion | 3246787 | 96.06 | 95.97 | 96.15 |
| Cross fusion | 3246830 | **96.25** | **96.17** | **96.34** |

# Result

TABLE IV

PERFORMANCE COMPARISON OF SINGLE MODALITY AND FUSION FCNS EVALUATED ON THE CHALLENGING SET.

| Fusion strategy | # params | MaxF [%] | PRE [%] | REC [%] |
|---|---|---|---|---|
| ZYX | 1623395 | 95.21 | 93.40 | 97.09 |
| RGB | 1623395 | 91.81 | 89.18 | 94.61 |
| Early fusion | 1624931 | 95.44 | 93.54 | 97.42 |
| Late fusion | 3246787 | 95.24 | 92.73 | 97.09 |
| Cross fusion | 3246830 | **96.02** | **94.39** | **97.70** |

# Result

TABLE V

KITTI ROAD BENCHMARK RESULTS (IN %) ON THE URBAN ROAD CATEGORY. ONLY RESULTS OF PUBLISHED METHODS ARE REPORTED.

| Method | MaxF | AP | PRE | REC | Time (s) |
|---|---|---|---|---|---|
| **LidCamNet** (our) | **96.03** | **93.93** | **96.23** | 95.83 | 0.15 |
| RBNet [8] | 94.97 | 91.49 | 94.94 | 95.01 | 0.18 |
| StixelNet II [31] | 94.88 | 87.75 | 92.97 | **96.87** | 1.2 |
| MultiNet [7] | 94.88 | 93.71 | 94.84 | 94.91 | 0.17 |
| LoDNN [9] | 94.07 | 92.03 | 92.81 | 95.37 | **0.018** |
| DEEP-DIG [32] | 93.98 | 93.65 | 94.26 | 93.69 | 0.14 |
| Up-Conv-Poly [28] | 93.83 | 90.47 | 94.00 | 93.67 | 0.08 |

http://www.cvlibs.net/datasets/kitti/eval_road.php

# Result

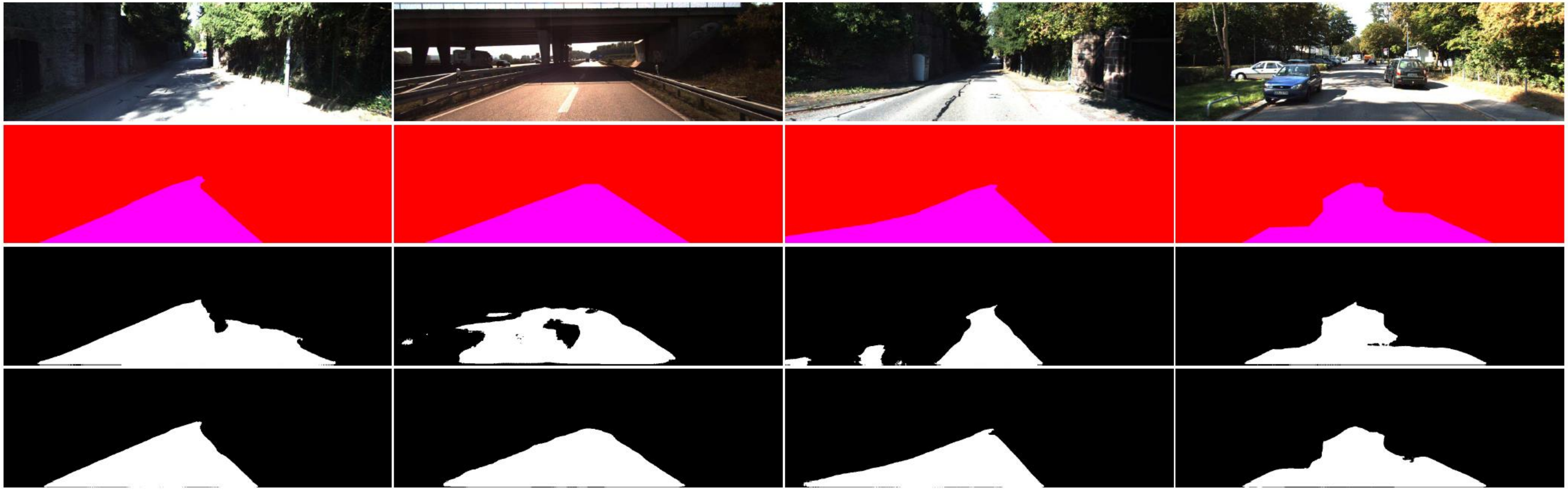


Fig. 4. (Top row) Some examples of camera images captured in difficult lighting conditions and included in the challenging set. (Second row) Corresponding ground truth annotations: The road is depicted as violet, whereas not-road is red. (Third row) Road segmentations obtained by the RGB-FCN. (Fourth row) Road segmentations generated by the cross fusion FCN.

# Conclusion

- ✓ A novel fusion FCN integrating camera images and LIDAR point clouds for carrying out road detection. Results from KITTI road detection dataset achieve top-performing algorithms.
- ✓ Robust system compared with single sensors (either camera or LIDAR) under a wider range of external conditions

- ○ Higher computation cost due to additional dimension of information added
- ○ Performance depending on devices. Eg. a 16-beam LIDAR vs a 64-beam LIDAR
- ○ Have not tested on augmented images, so performance may drop

**Thanks!**