Deep Sensor Fusion for ADAS Applications

Vijay John, Seiichi Mita,
Smart Vehicle Research Center

TOYOTA TECHNOLOGICAL INSTITUTE 豊田工大

Title of Content

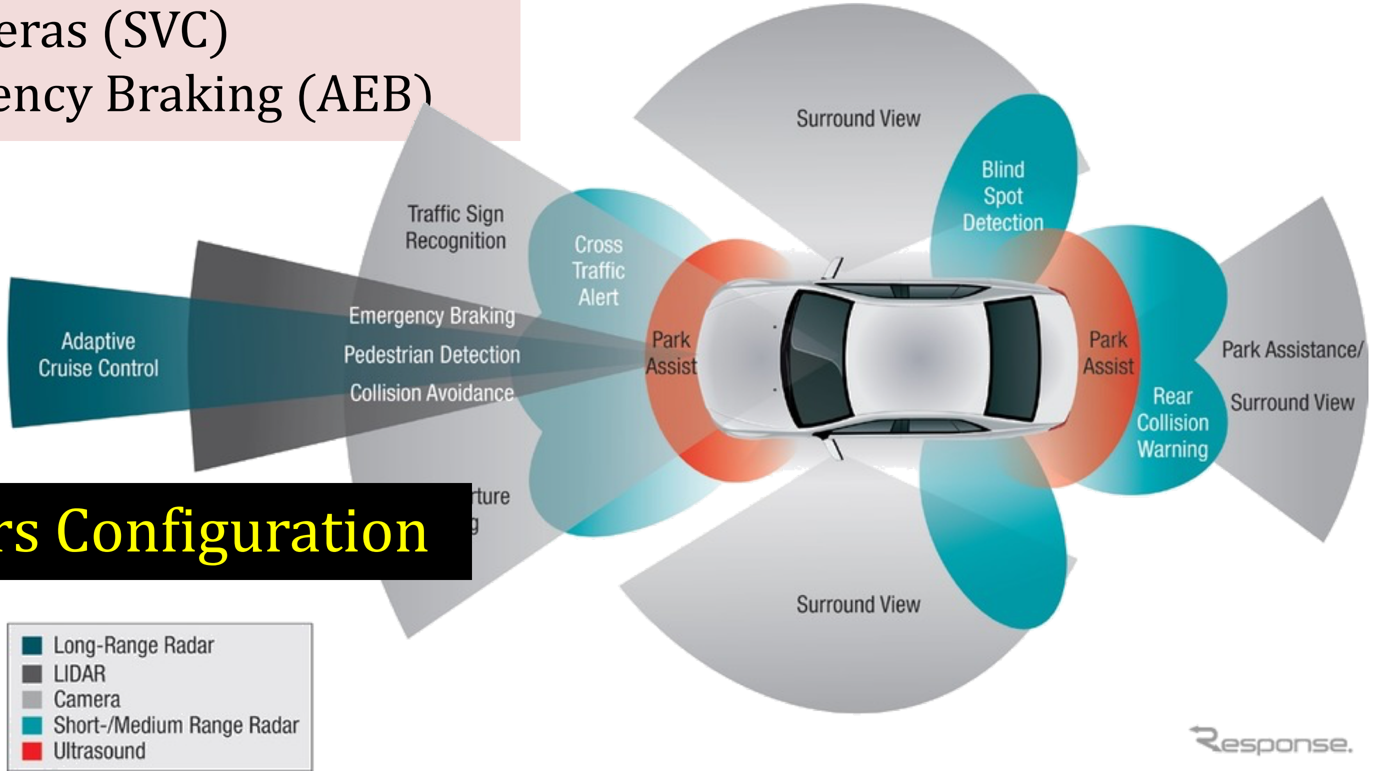
- Deep Learning-based ADAS Application
- Why Sensor Fusion (Depth + Texture) ?
- How to Fuse Depth & Intensity
- Proposed Deep Learning Model for Fusion
- Results and Evaluation

Popular ADAS Applications

ADAS Applications

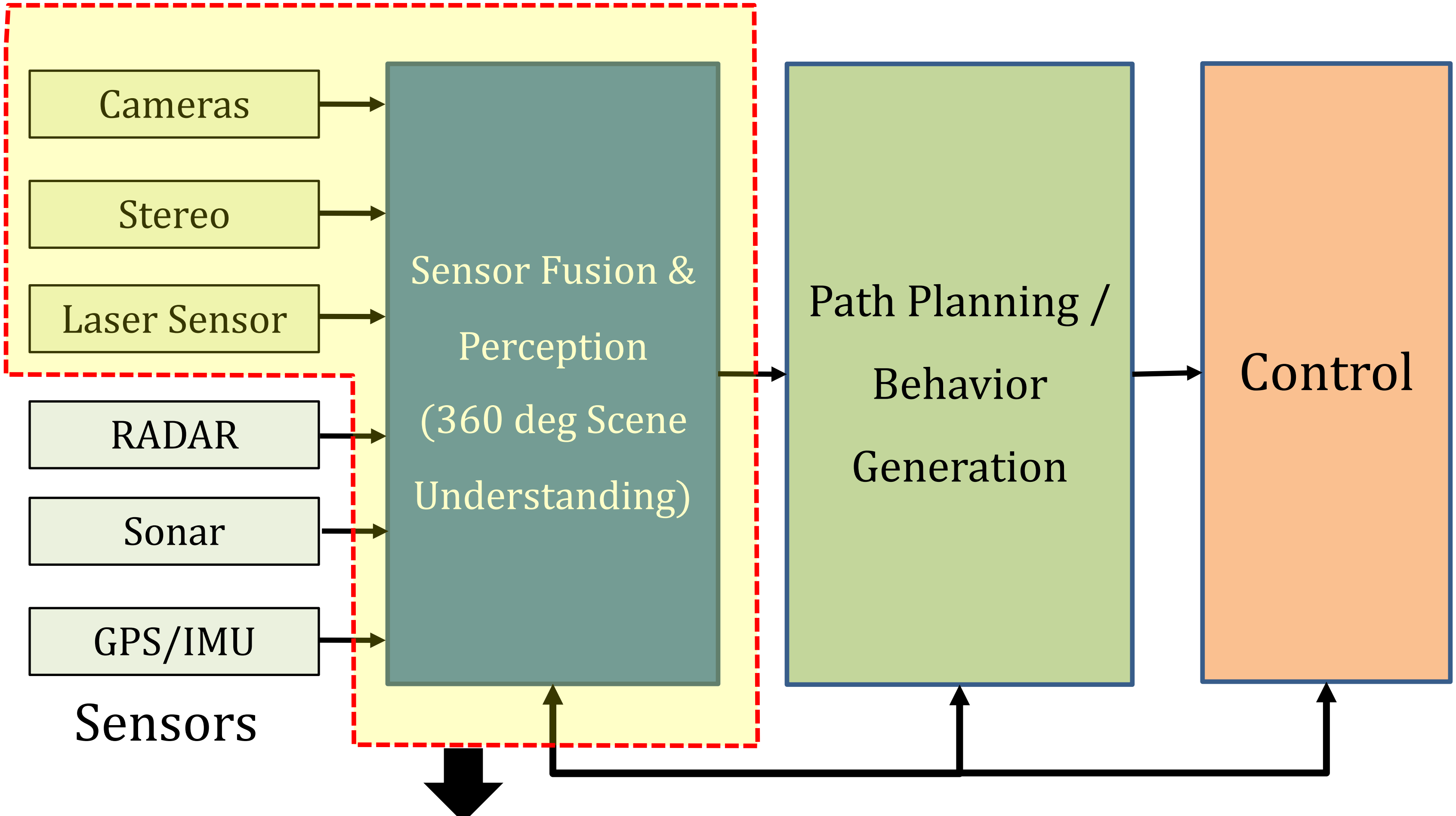
- Adaptive Cruise Control (ACC)
- Adaptive Front Lights (AFL)
- Driver Monitoring System (DMS)
- Forward Collision Warning (FCW)
- Intelligent Speed Adaptation (ISA)
- Lane Departure Warning (LDW)
- Pedestrian Detection System (PDS)
- Surround-View Cameras (SVC)
- Autonomous Emergency Braking (AEB)

TTI Research Vehicle



Standard Sensors Configuration

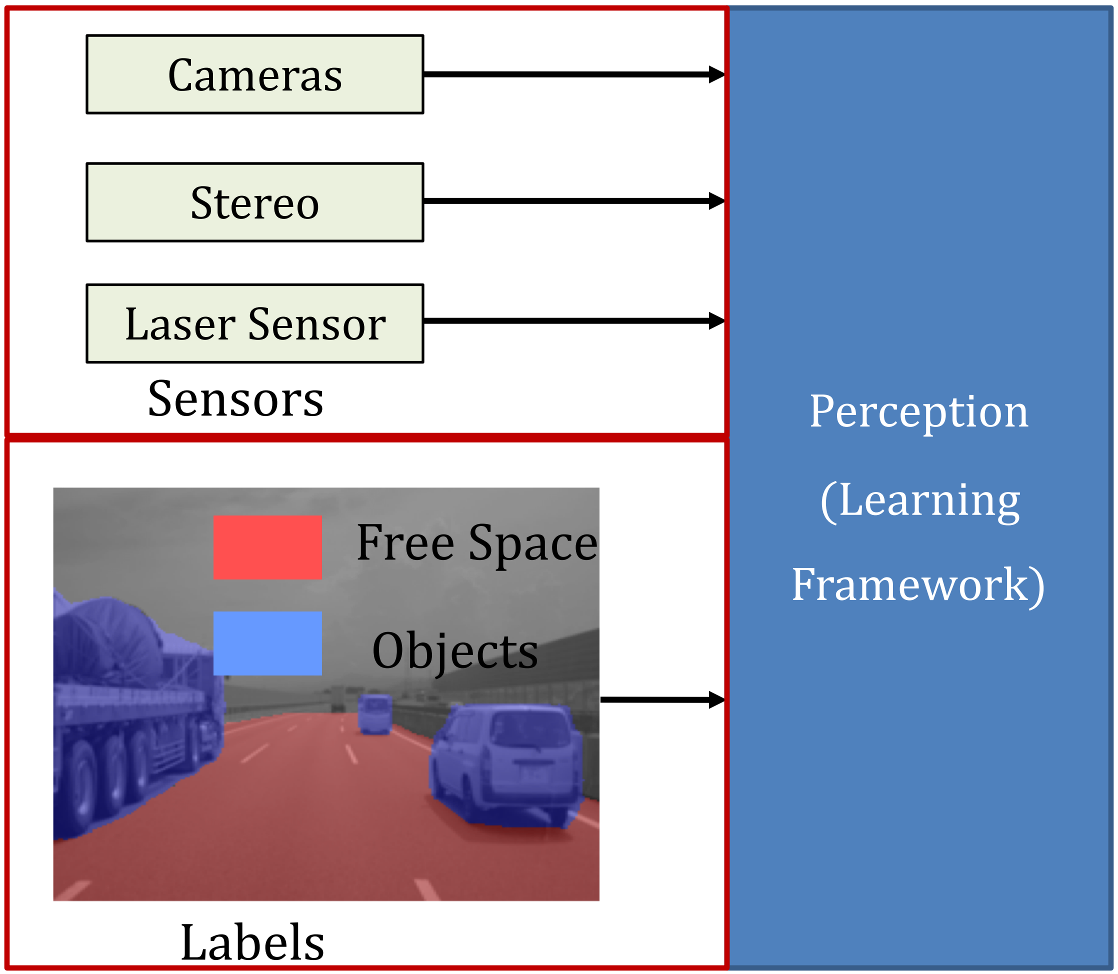
General Framework for ADAS and AD



Understanding the Environment

Learning Framework for Perception

Training a learning framework for perception tasks



Traditional Learning

Feature Extraction (HOG, DPM etc)

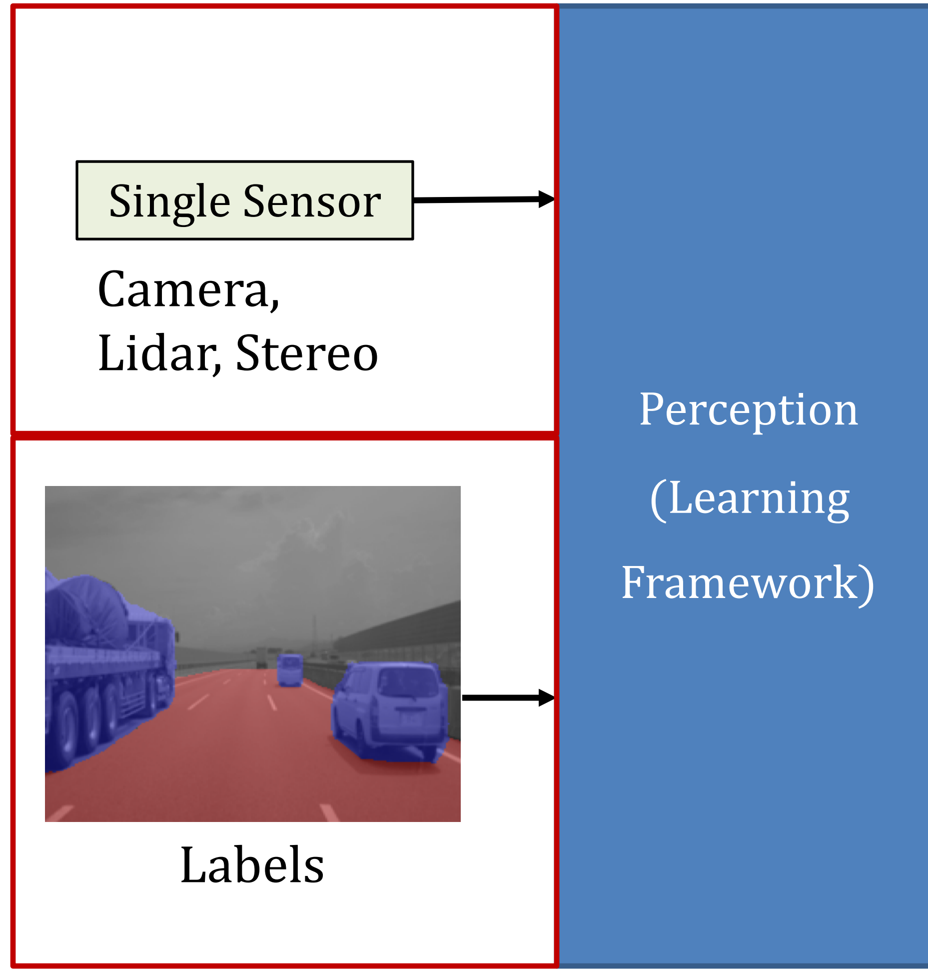
Feature Classification (SVM, Random Forest etc)

Deep learning

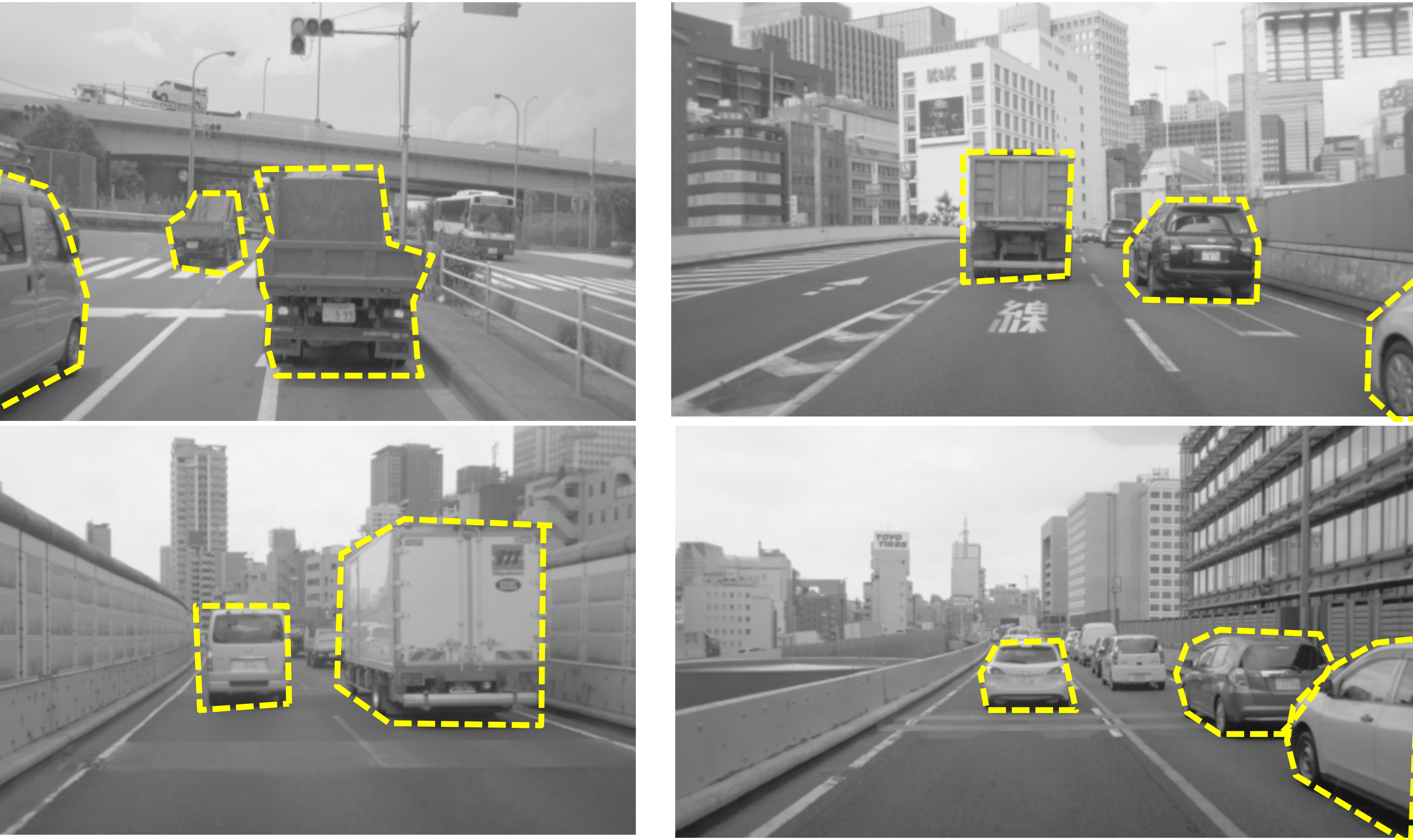
(**Feature Extraction + Feature Classification**)

Single Sensor-based learning

- Single sensor-based learning is not robust or descriptive enough
- **Challenges**
  - Environmental Variation (occlusion, illumination variation, etc.)
  - High Inter-Class and Intra-Class Variability



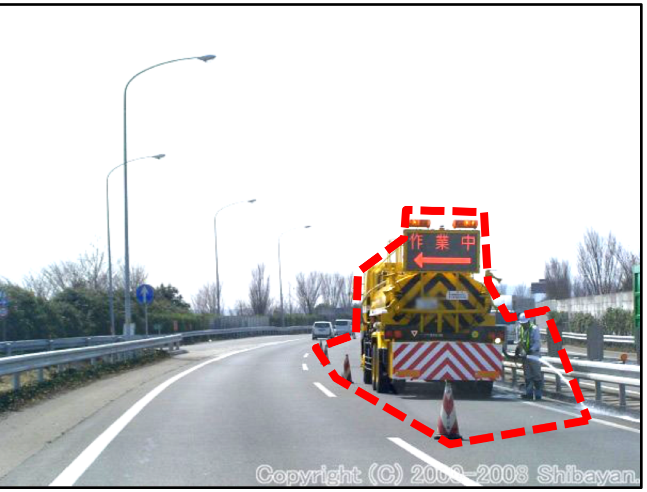
Intra-class Vehicle Variations

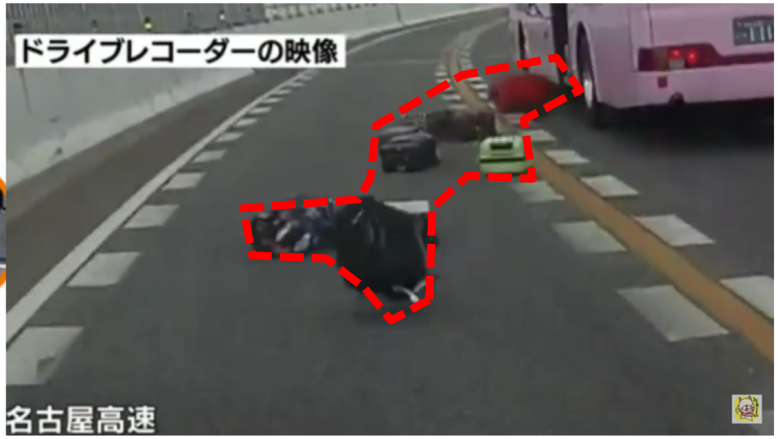


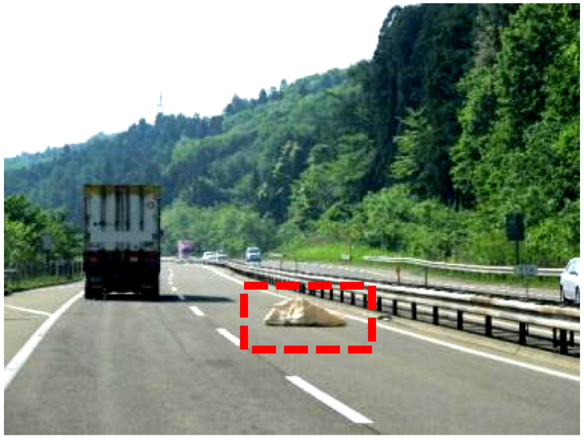
There are many vehicle varieties with different orientations

Intra-class On-Road Objects Variations

We have a large number of On-Road Objects



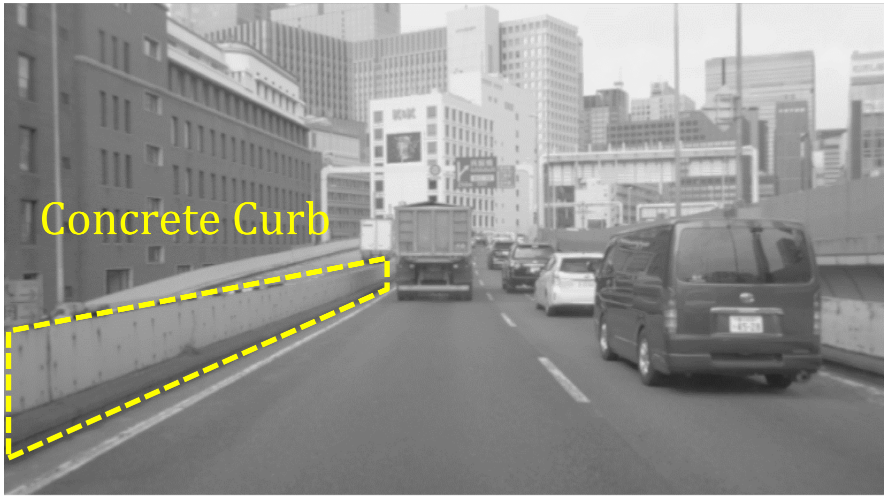


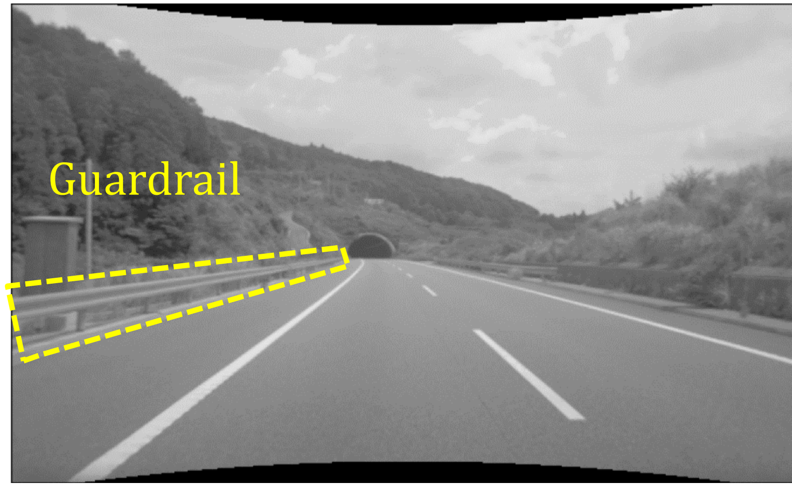


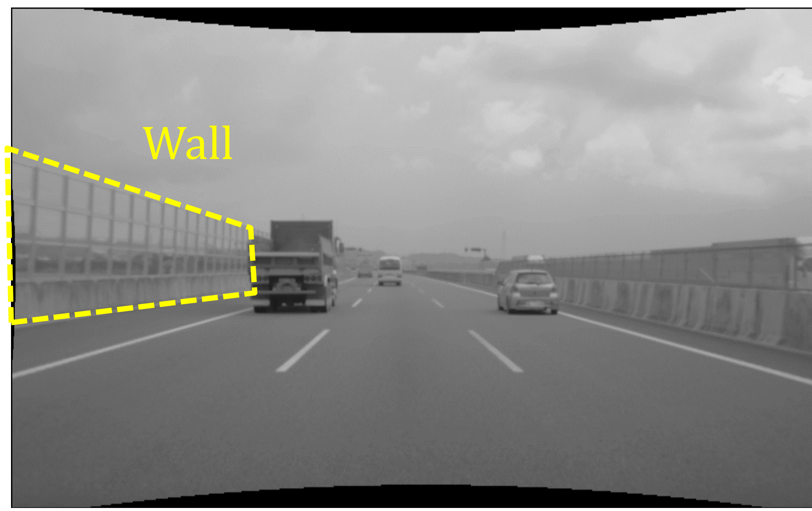
We have a lot of variety of on road objects!!!!

Intra-Class Free Space Boundary Variations

We have the different type of road boundaries







We have a lot of variety of Free Space Boundary!!!!

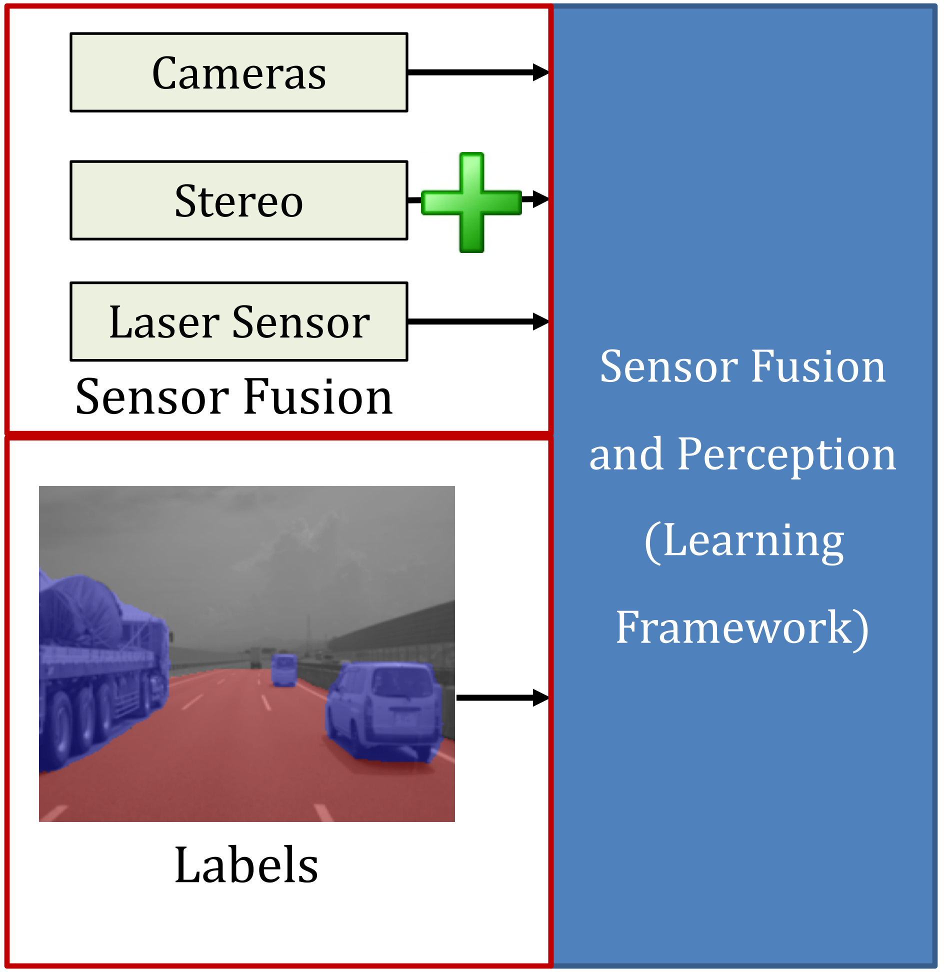
Environmental Variation

Illumination variation as observed by a monocular camera image with appearance features

Sensor fusion-based learning

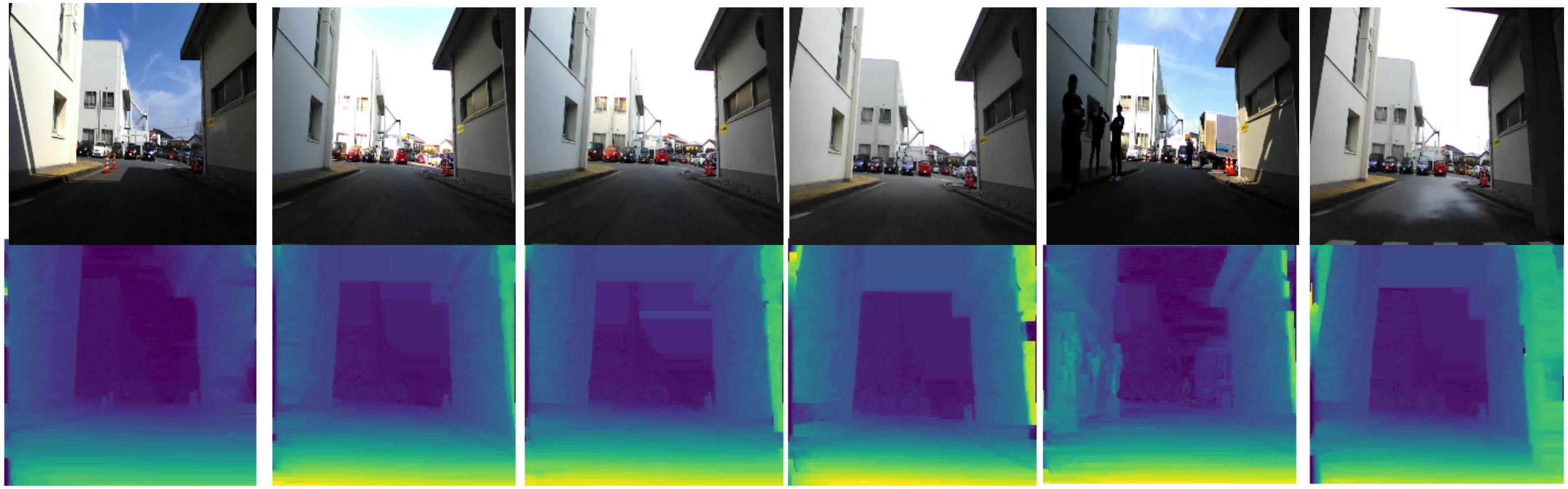
- Sensor Fusion-based learning with **complementary sensors** addresses these issues

- Monocular Camera appearance features and depth features are **complementary features**



Complementary sensors

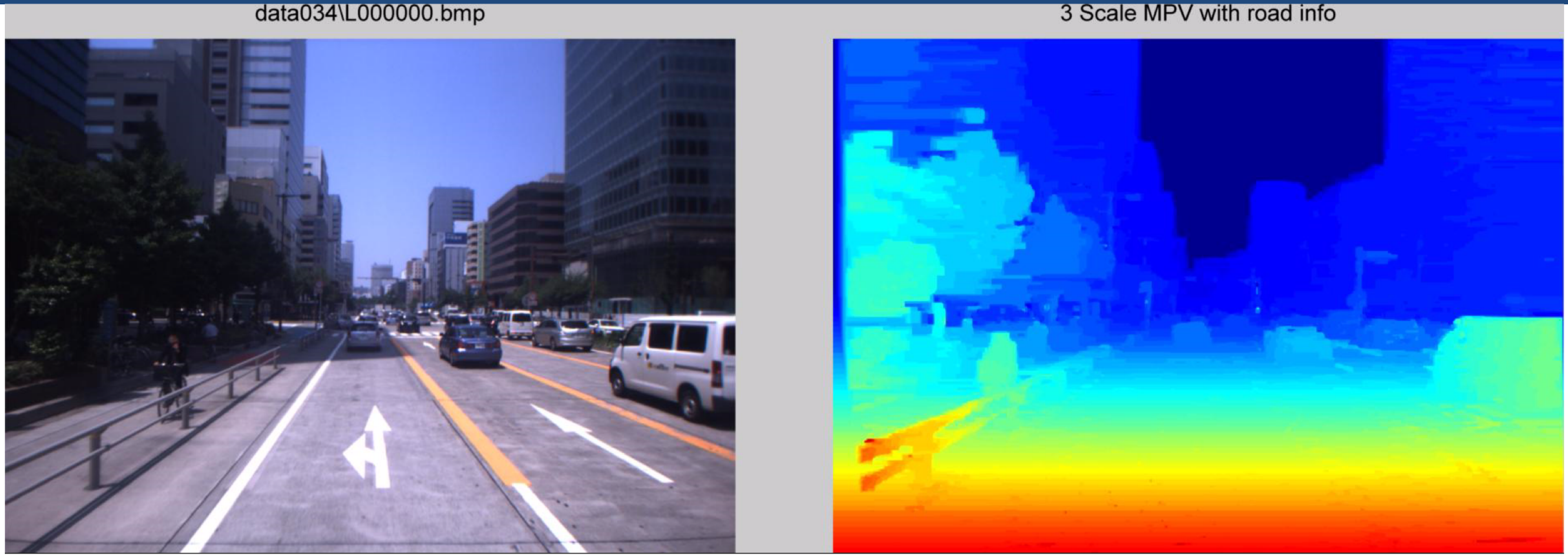
Monocular camera	Depth camera
Monocular camera provides rich appearance information	Depth camera provides depth information
Inexpensive	Stereo-based depth inexpensive
Illumination variation	Illumination invariant due to robust stereo algorithm [1]

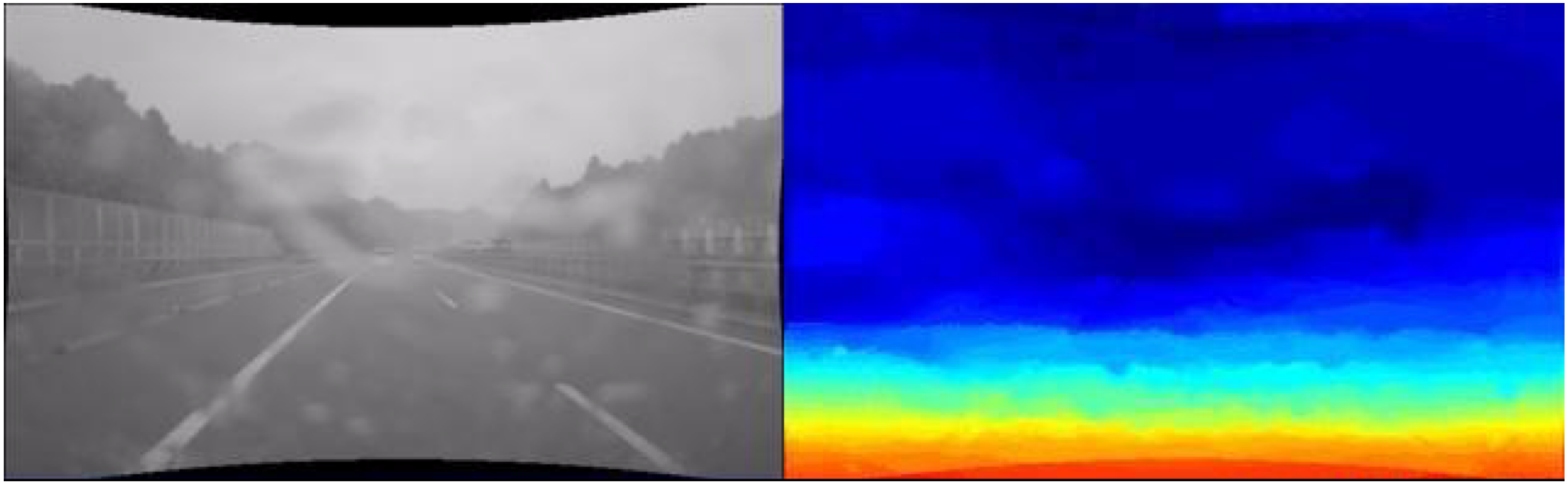


Depth information from stereo camera robust to illumination variatio

[1] Xu et al. Real-time Stereo Disparity Quality Improvement for Challenging Traffic Environments, IV 2018

Robust Stereo Vision



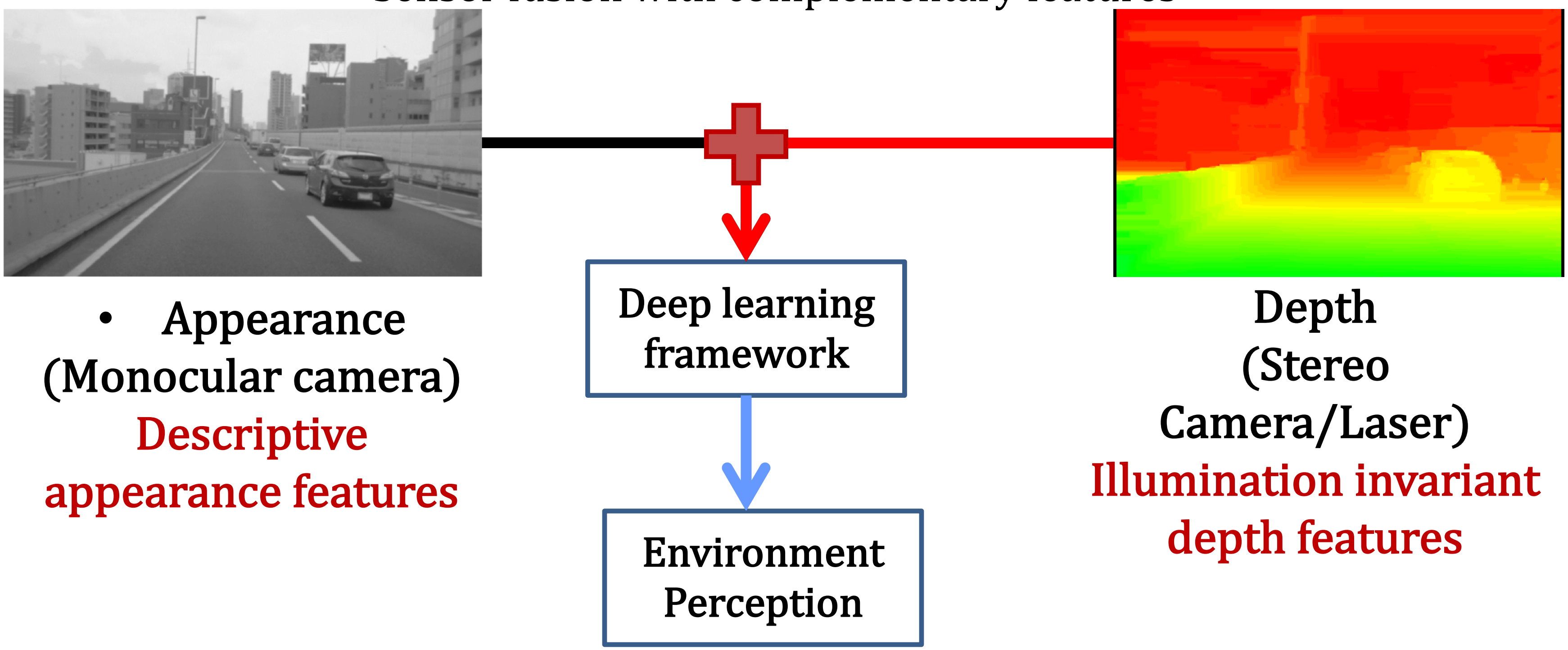


Stereo processing for 1280x960 image: 15ms/frame on Geforce GTX

Complementary Sensor Fusion for Deep Learning

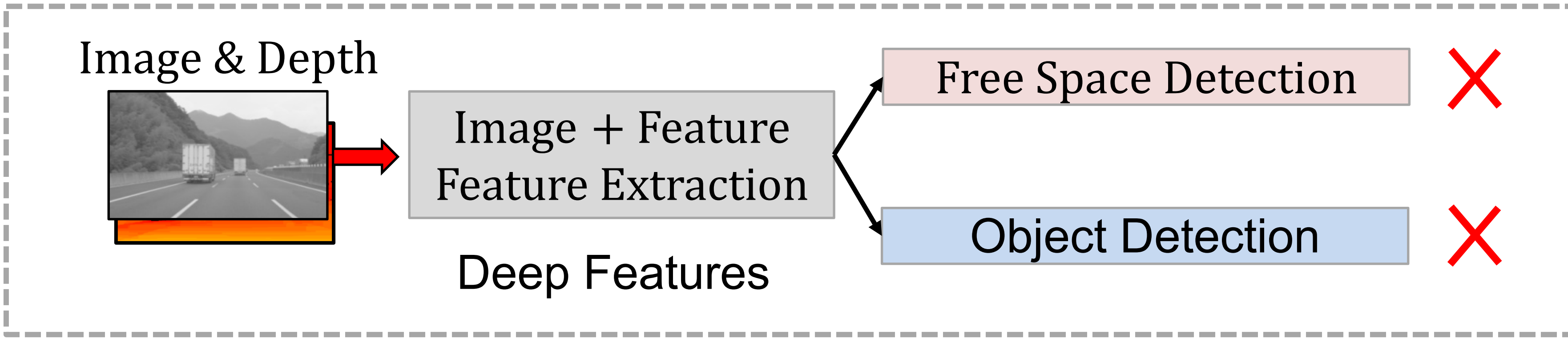
Appearance and Depth Features are Fused within a Deep learning Framework for Environment Perception

Sensor fusion with complementary features

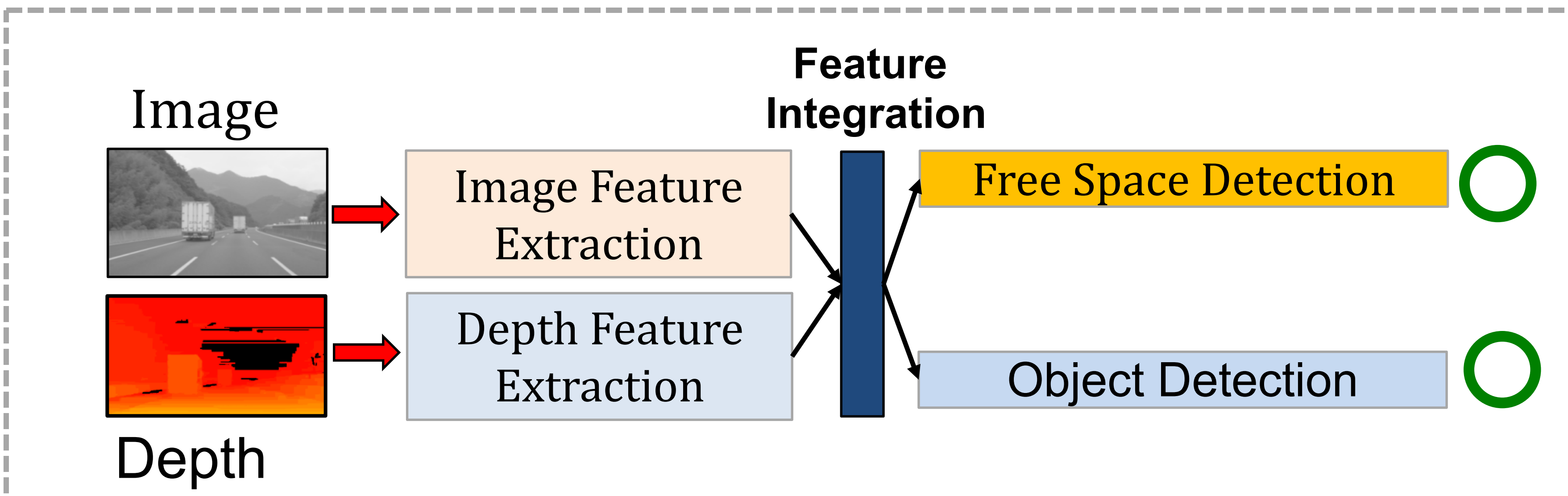


How to Fuse Sensors Data ??

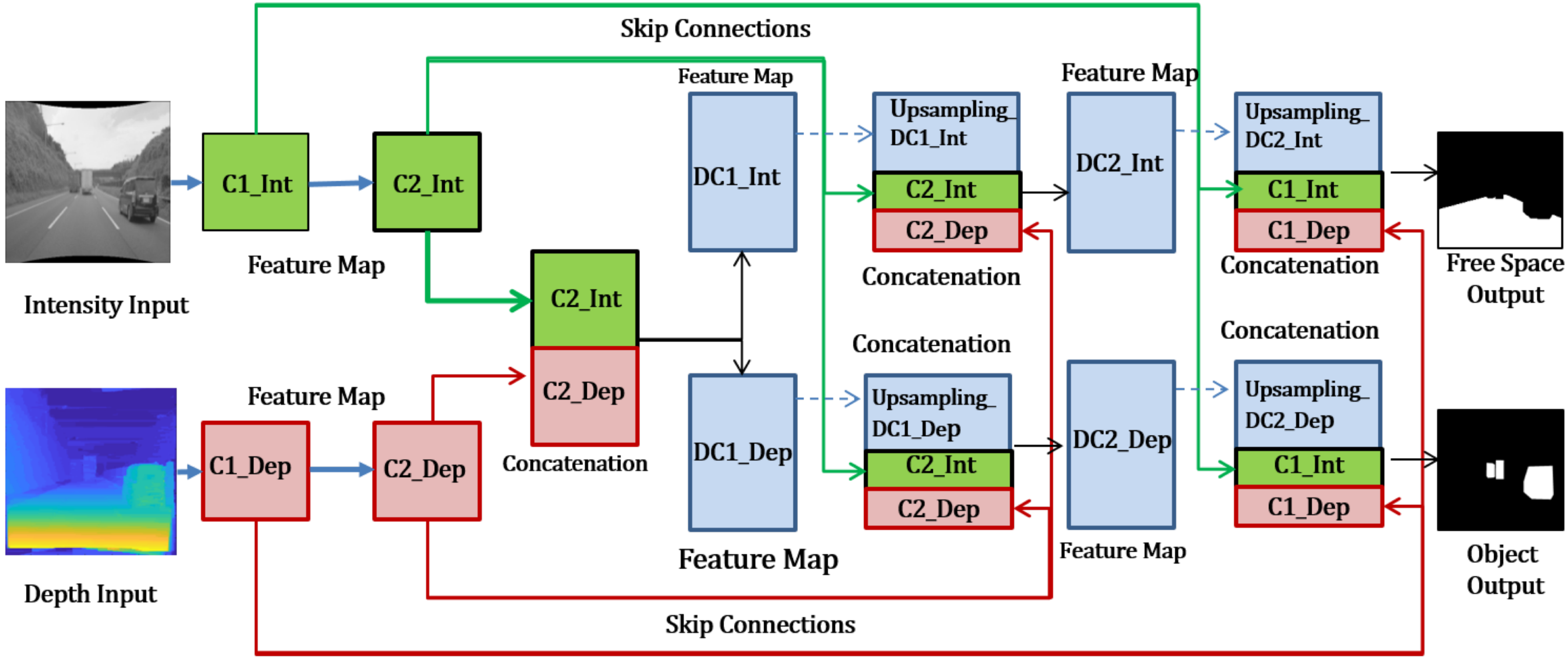
◻ Sensor Fusion : Raw Data Level Fusion



◻ Sensor Fusion : Feature Level Fusion



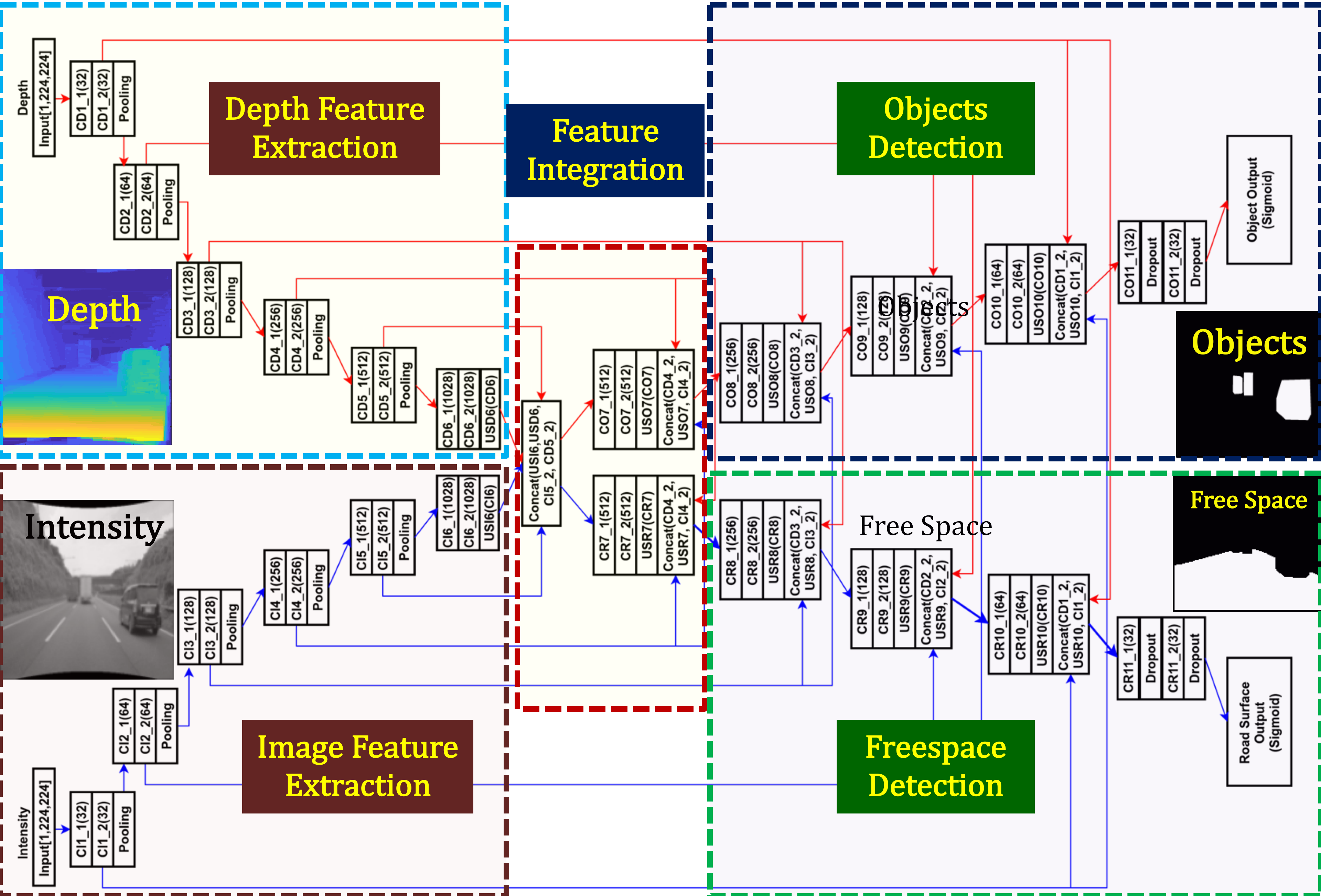
Proposed Model



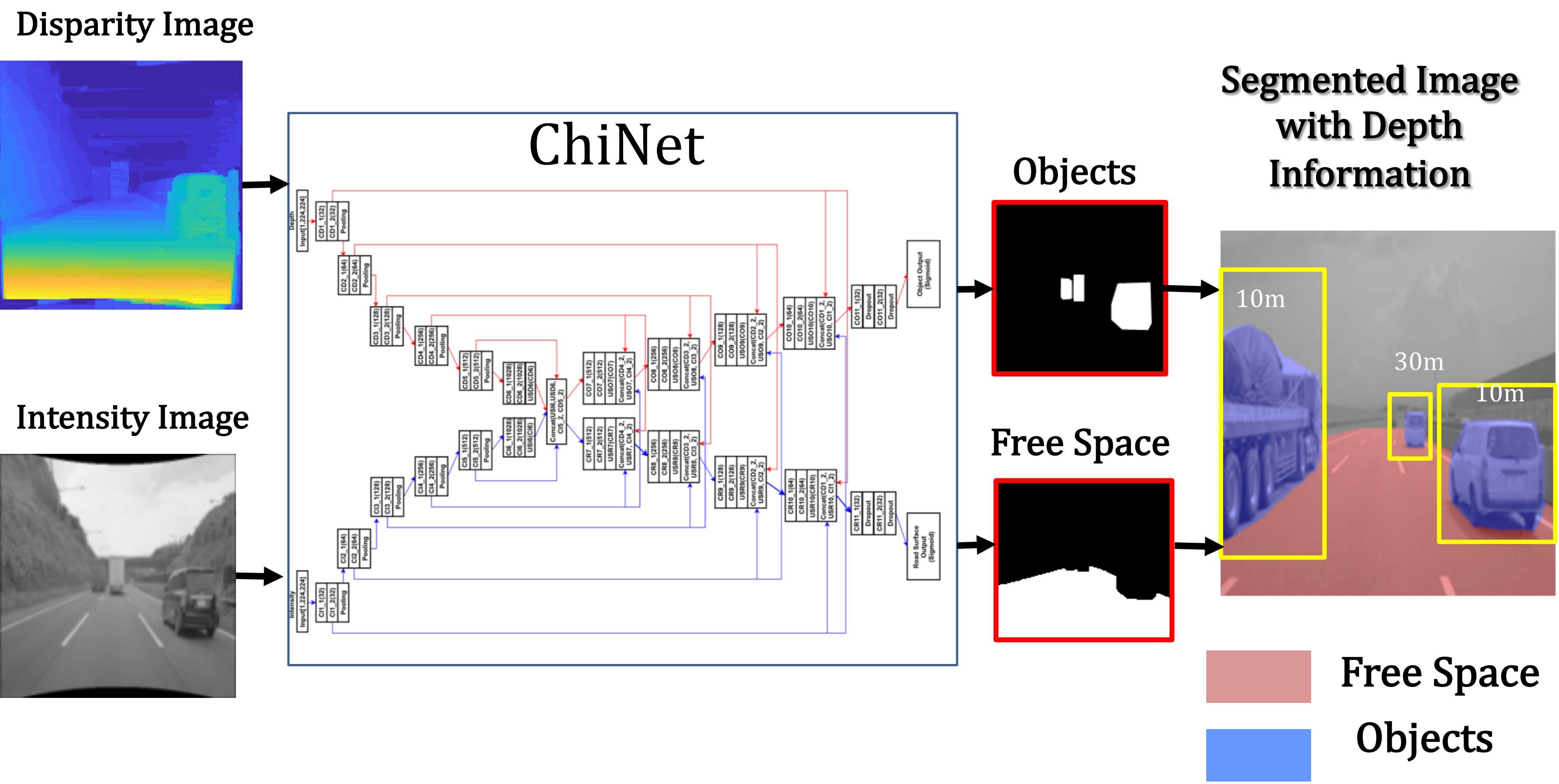
Skip Connections

- Entire Depth Encoder Feature Maps (*m,n,n*) are transferred to Free Space and Object Decoder Feature Maps (*o,n,n*) for Concatenation (*m+o,n,n*)
- Entire Intensity Encoder Feature Maps (*m,n,n*) are transferred to Free Space and Object Decoder Feature Maps (*o,n,n*) for Concatenation (*m+o,n,n*)

Final Proposed Architecture

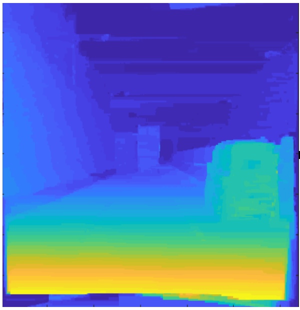


Proposed Method: ChiNet

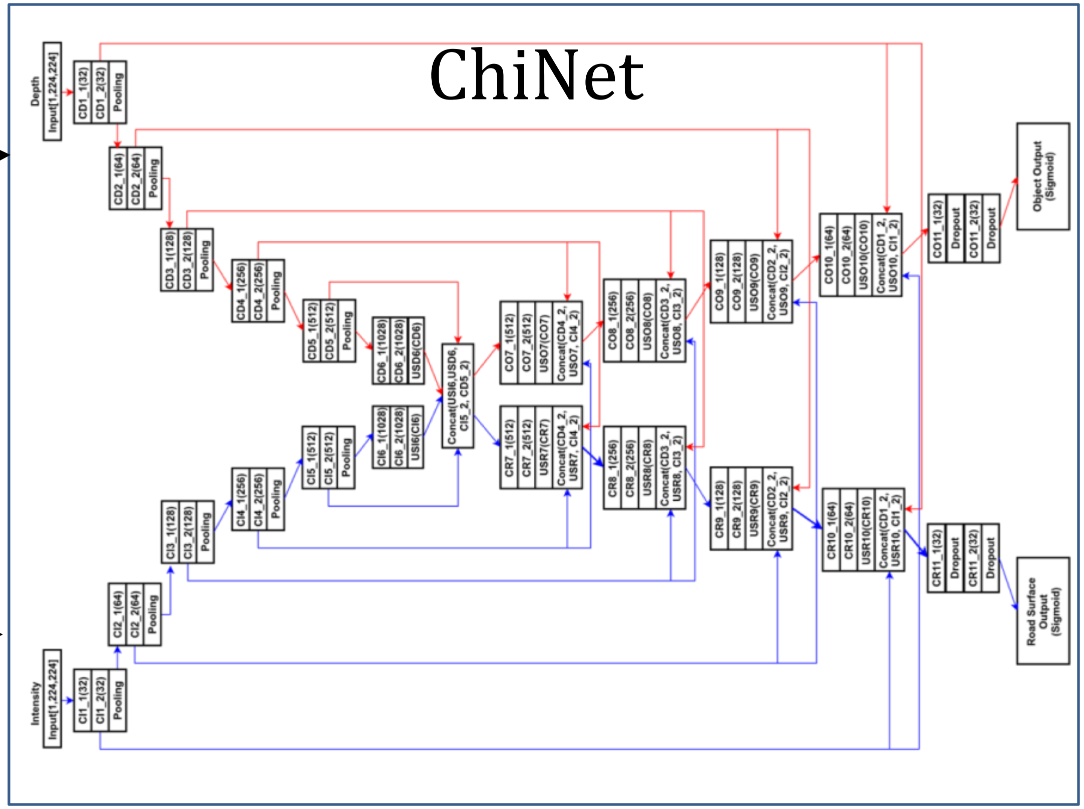


Learning ChiNet

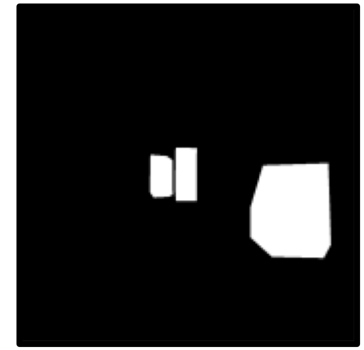
Disparity Image



Intensity Image



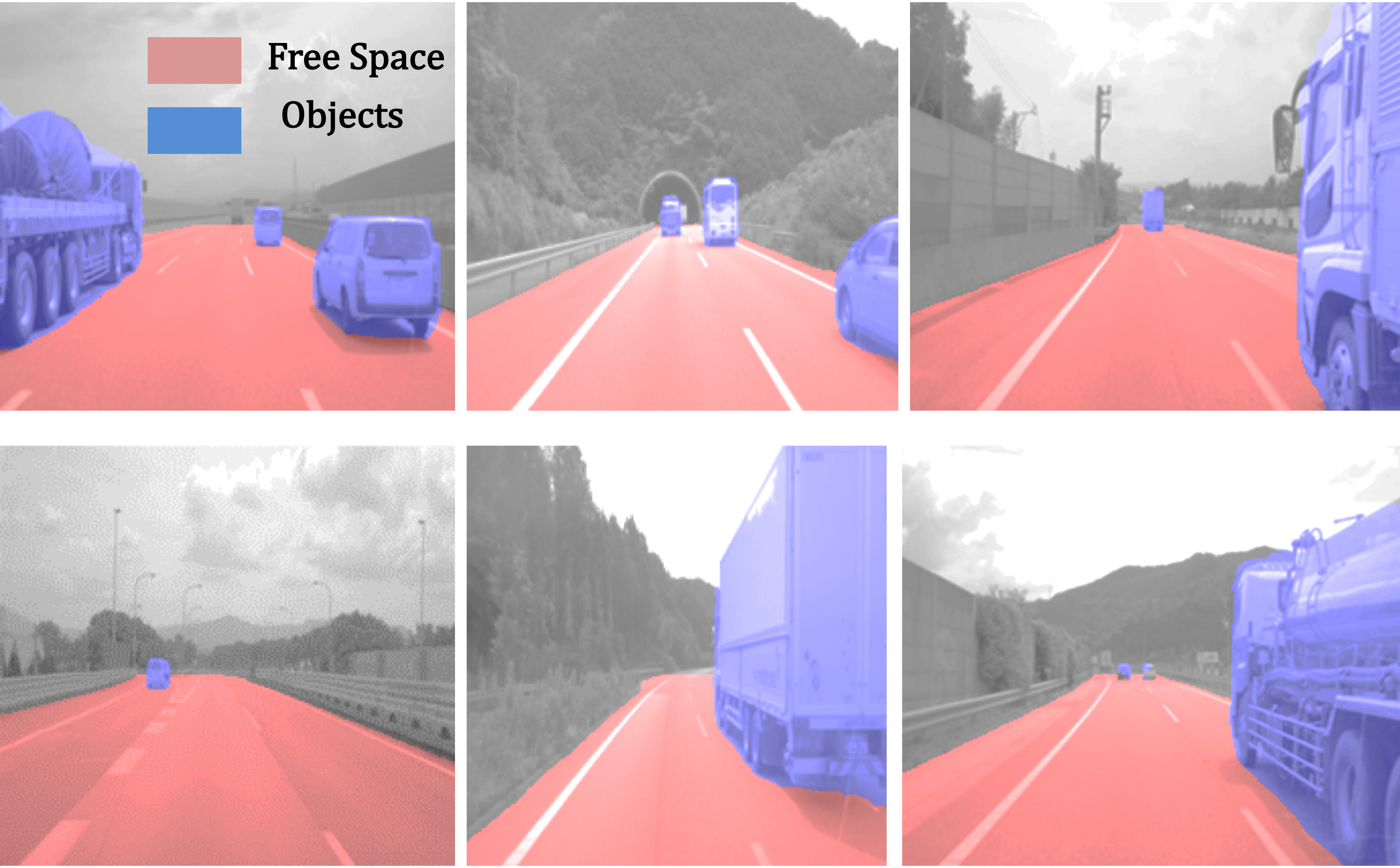
Objects



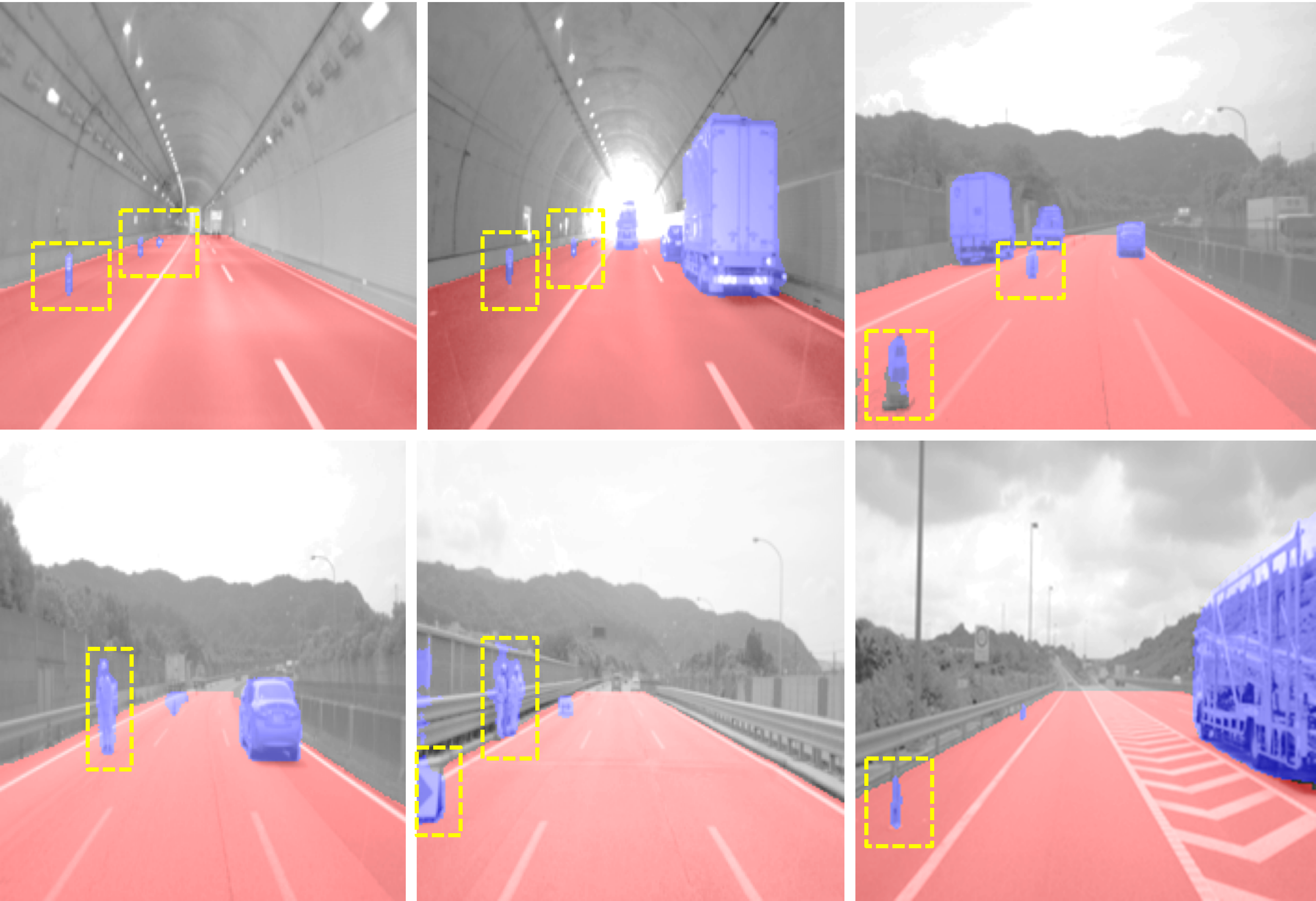
Free Space

- Trained with **9000 Samples** from Japanese Highway dataset
  - Manually annotated free space and objects
- Trained on **Keras with theano backend**
- Trained with **Nvidia Titan X GPU**

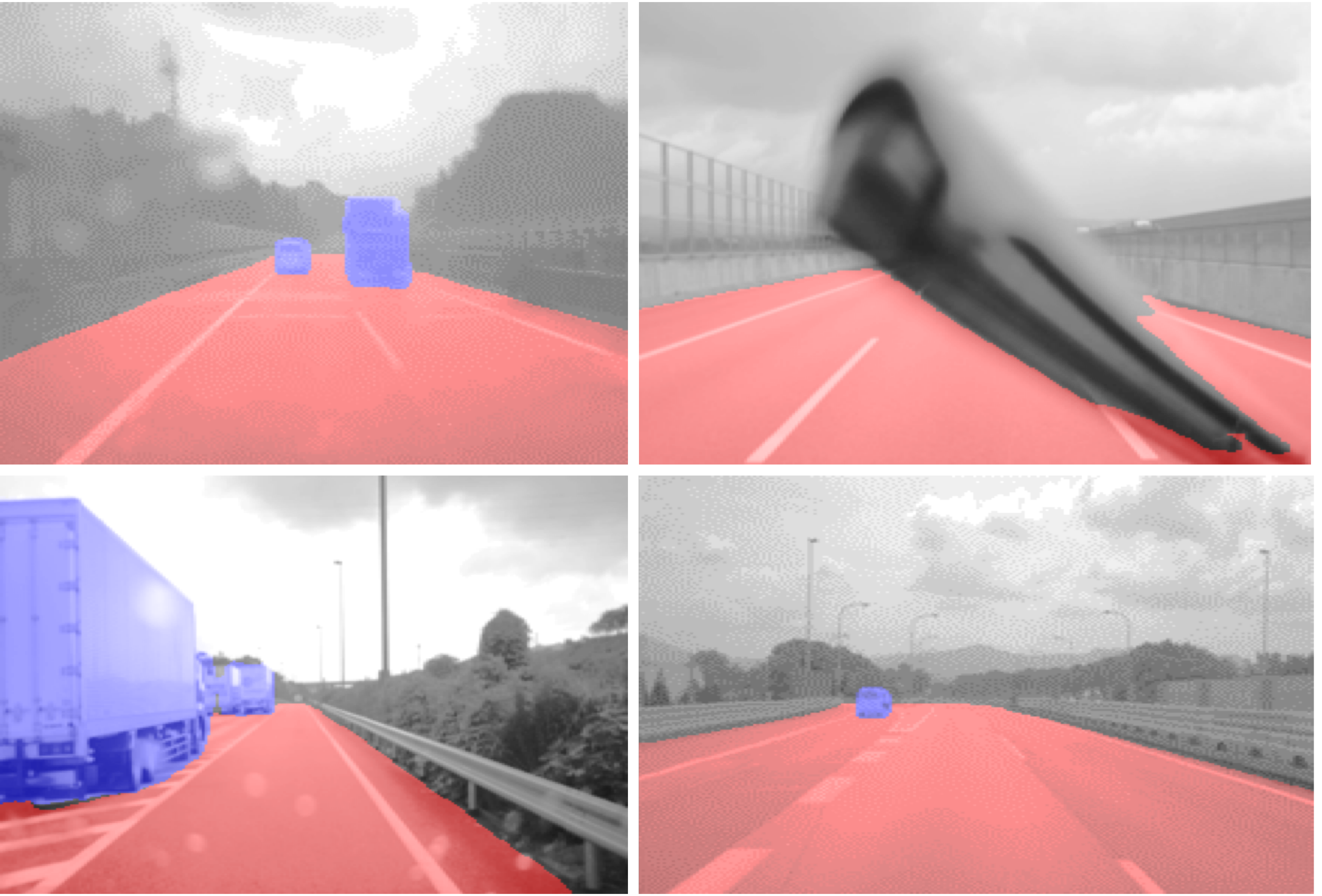
Results for Tomei Highway



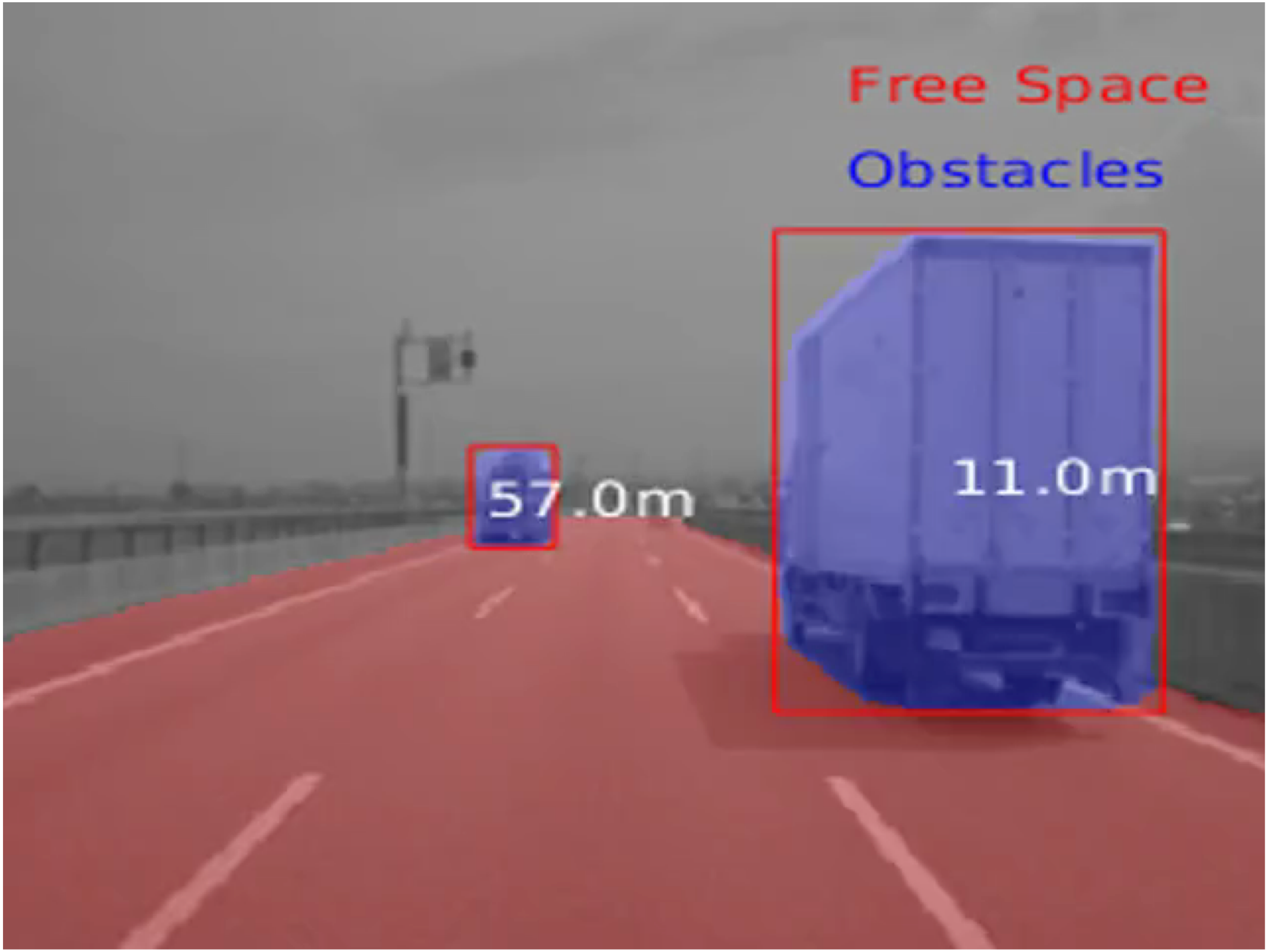
Small Objects Detection



Raining Weather



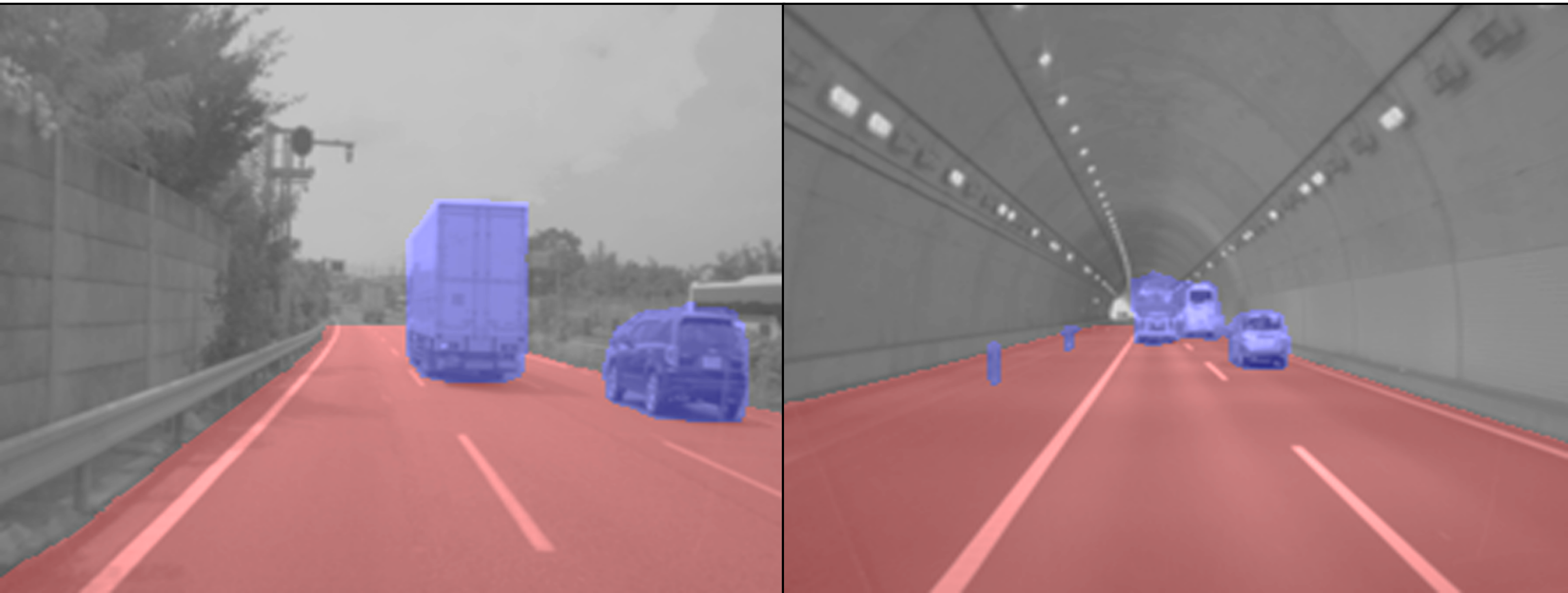
Sample Movie



ChiNet vs Baseline

Algorithm	Acc.	Time
ChiNet	97.35	192ms
U-Net [1]	94.2	82ms
FuseNet[2]	95.2	125ms

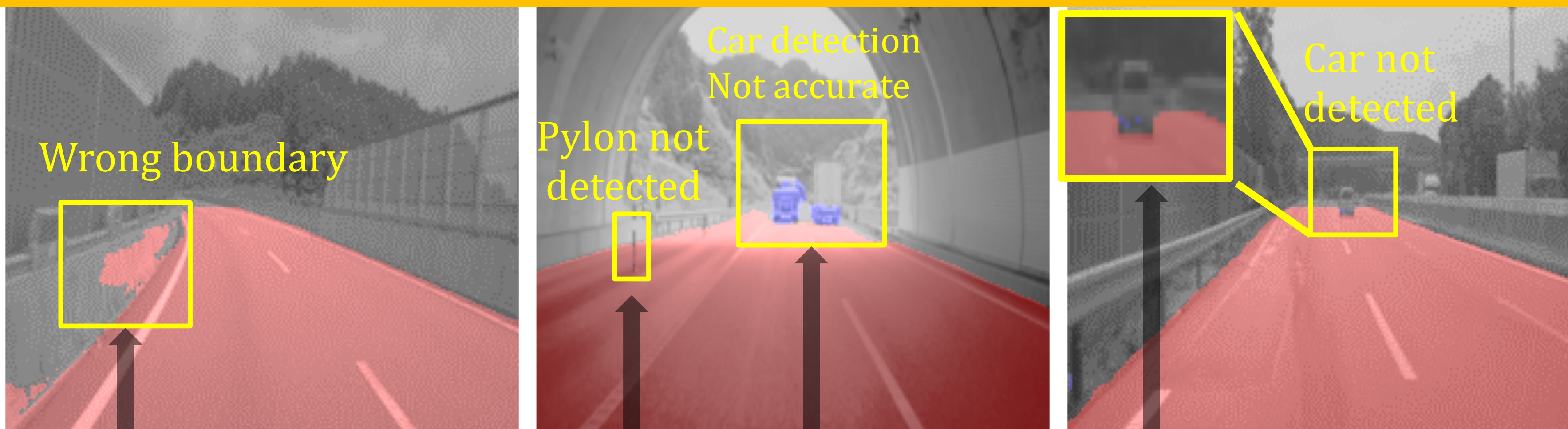
Implemented on **GeForce Titan X** using Keras with Theano backend



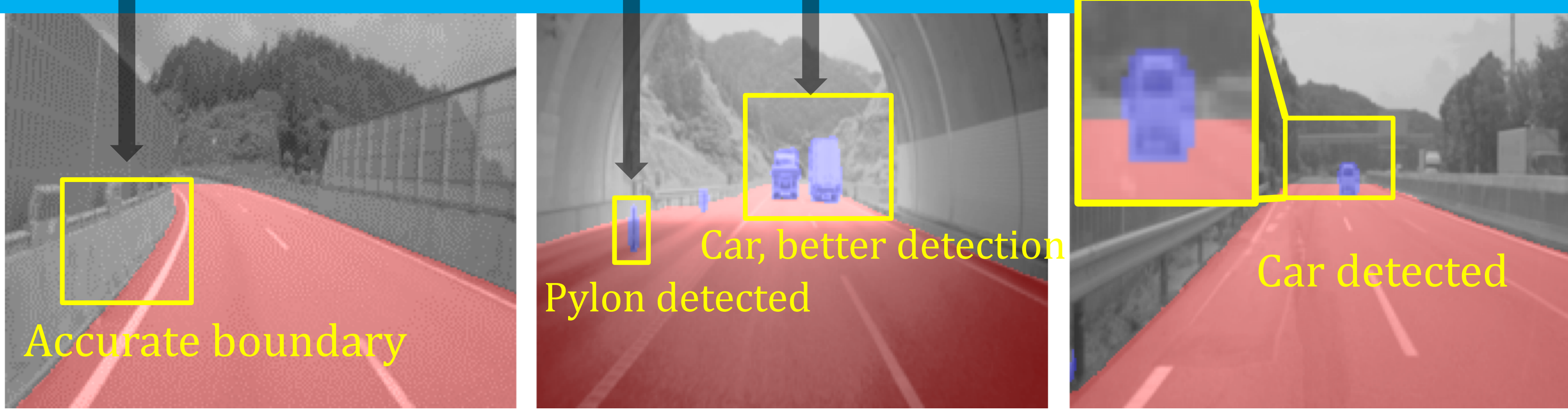
Evaluation Result

Comparison : "Intensity" vs "Intensity and Depth"

Intensity image only

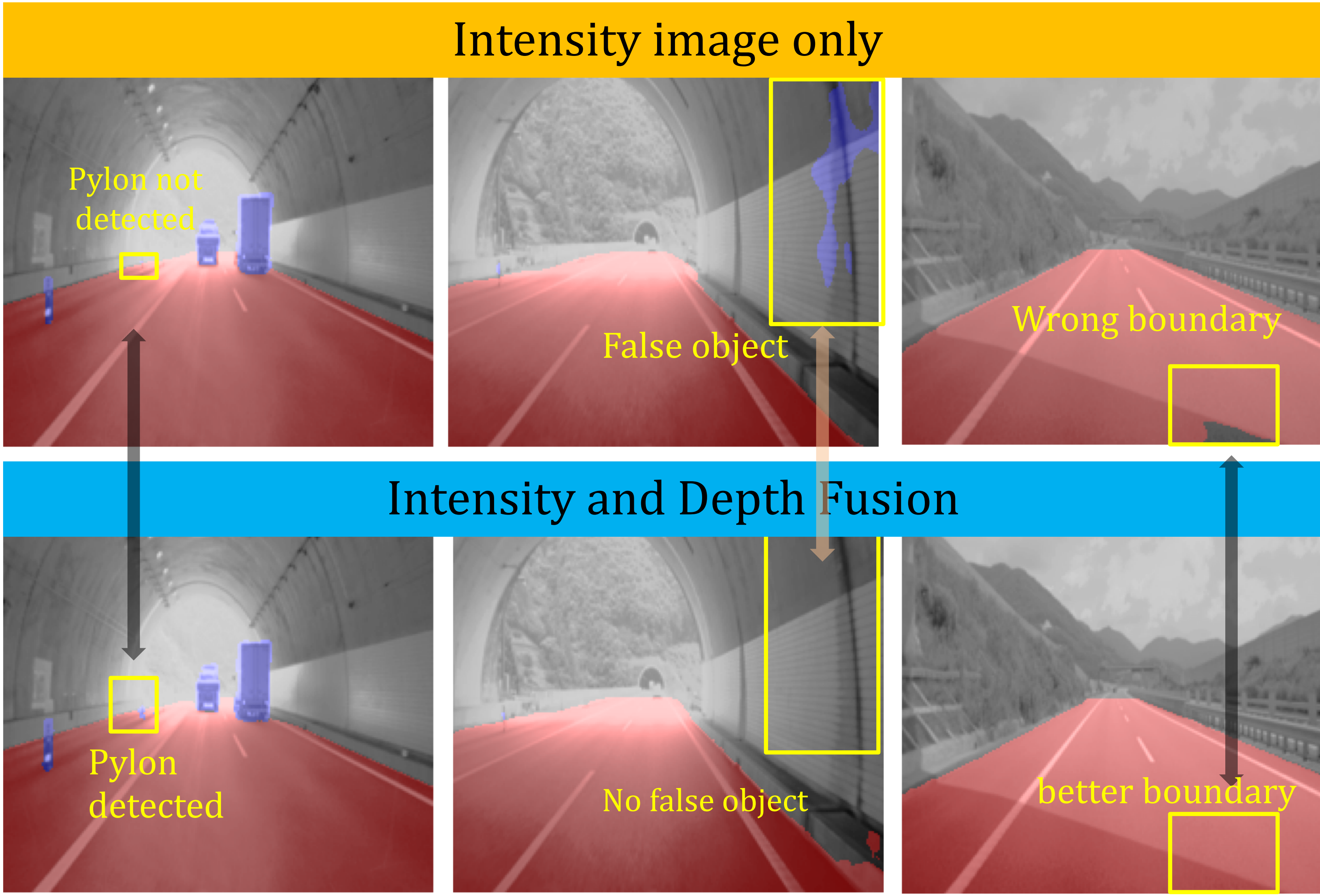


Intensity and Disparity fusion

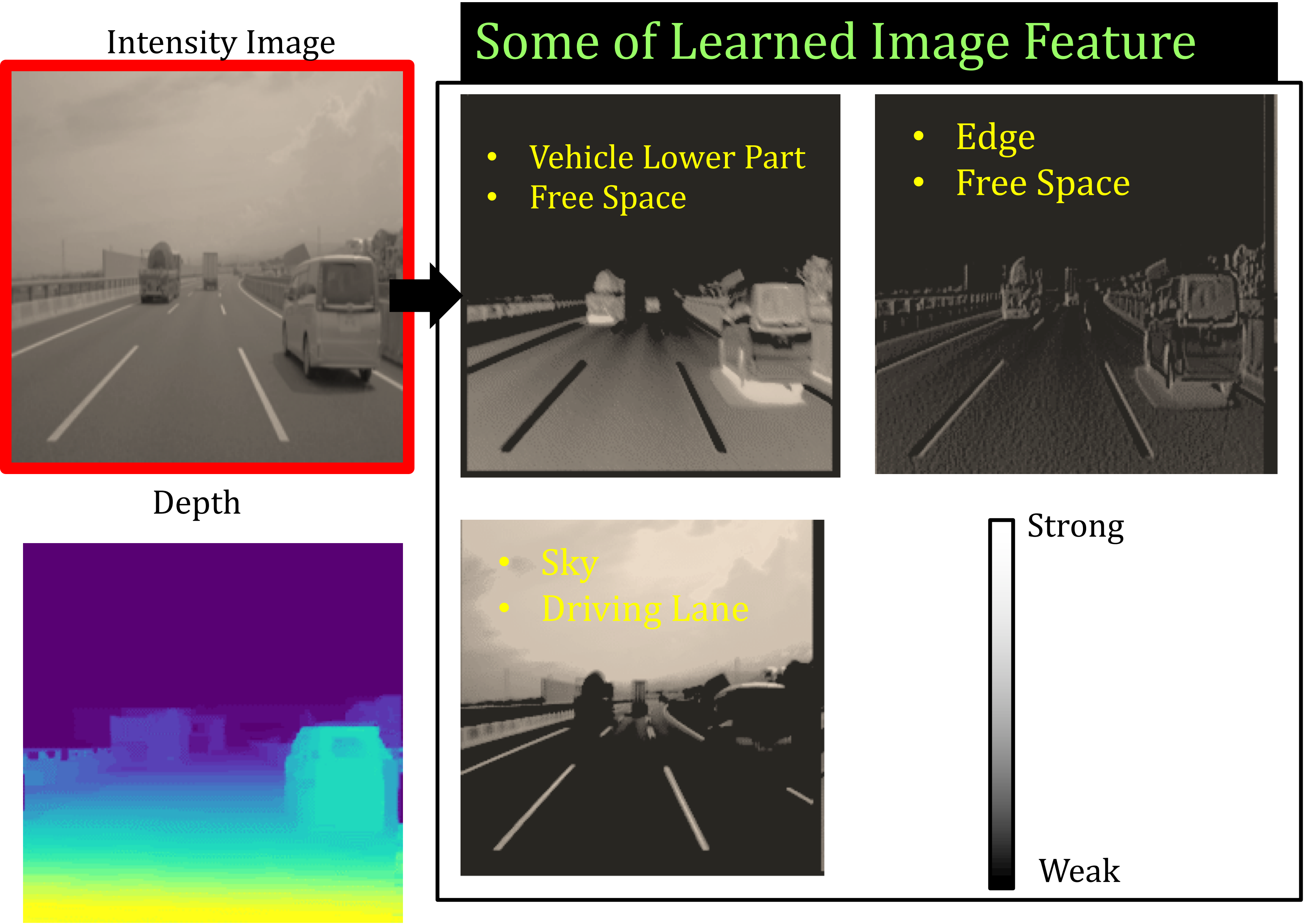


Evaluation Result

Comparison : "Intensity" vs "Intensity and Dept



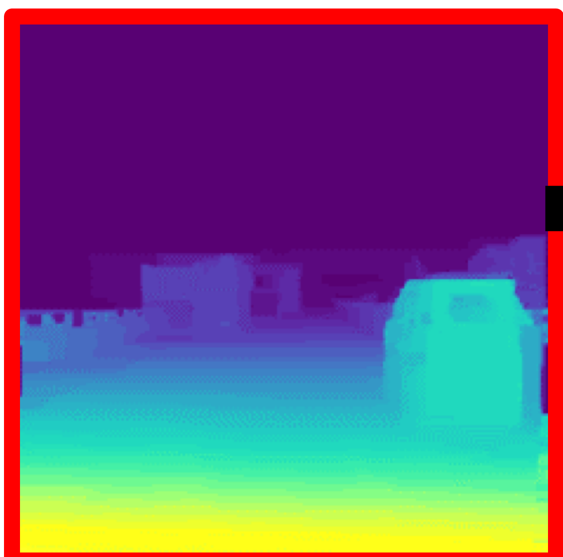
Learned Features by Chi-Net

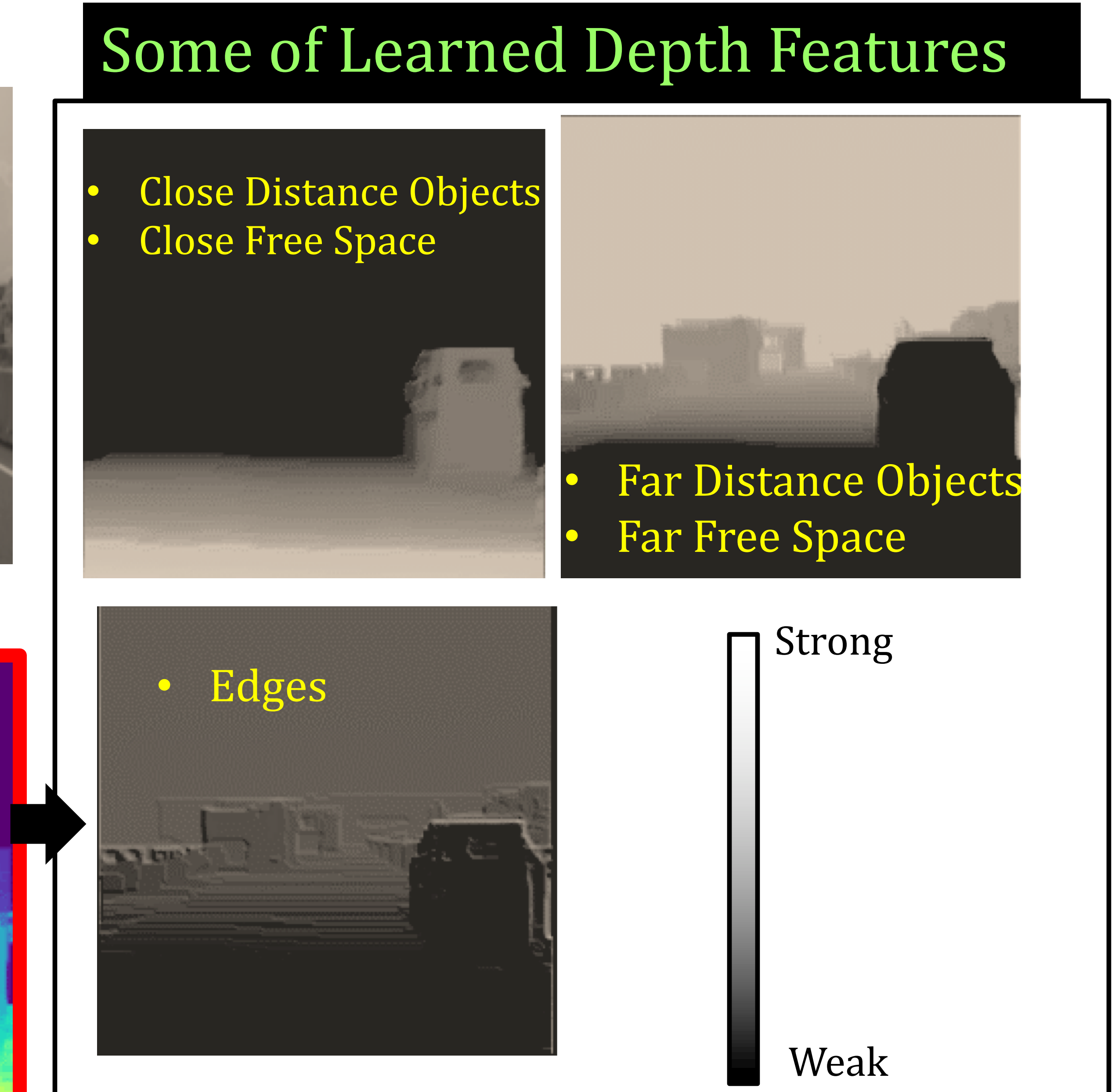


Learned Features by Chi-Net

Intensity Image

Depth





Conclusion

- Sensor fusion of appearance and depth features for environment perception
- Increased robustness and perception accuracy
- **ChiNet advantages**
  - Precise object boundary detection
  - Detection of small objects in the road
  - Detection of far-away objects
- Computational time
  - Reduction of computational time to ~50ms possible with optimized CUDA libraries and advances in GPU computing

THANKS AND QUESTIONS