S9195-Deep Sensor Fusion for Visible Stereo and Thermal Stereo for Autonomous Driving

Vijay John, Yuquan Xu , Seiichi Mita,
Smart Vehicle Research Center

Hossein Tehrani, Tomoyoki Oishi,
Masataka Konishi, Hakusho Chin
Advanced Mobility Development

DENSO
Crafting the Core

Kazuhisa Ishimaru, Sakiko Nishino
Research Division 2.0

SOKEN

Title of Content

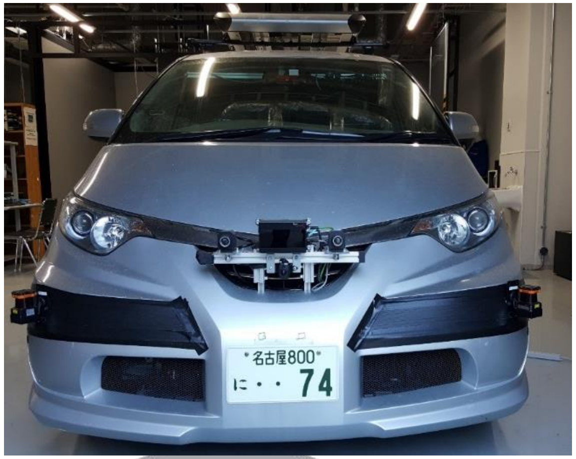
- ADAS and Automated Driving
- World 3D Reconstruction
- 3D Deep Sensor Fusion
- Future Plan
- Conclusion

Popular ADAS Applications

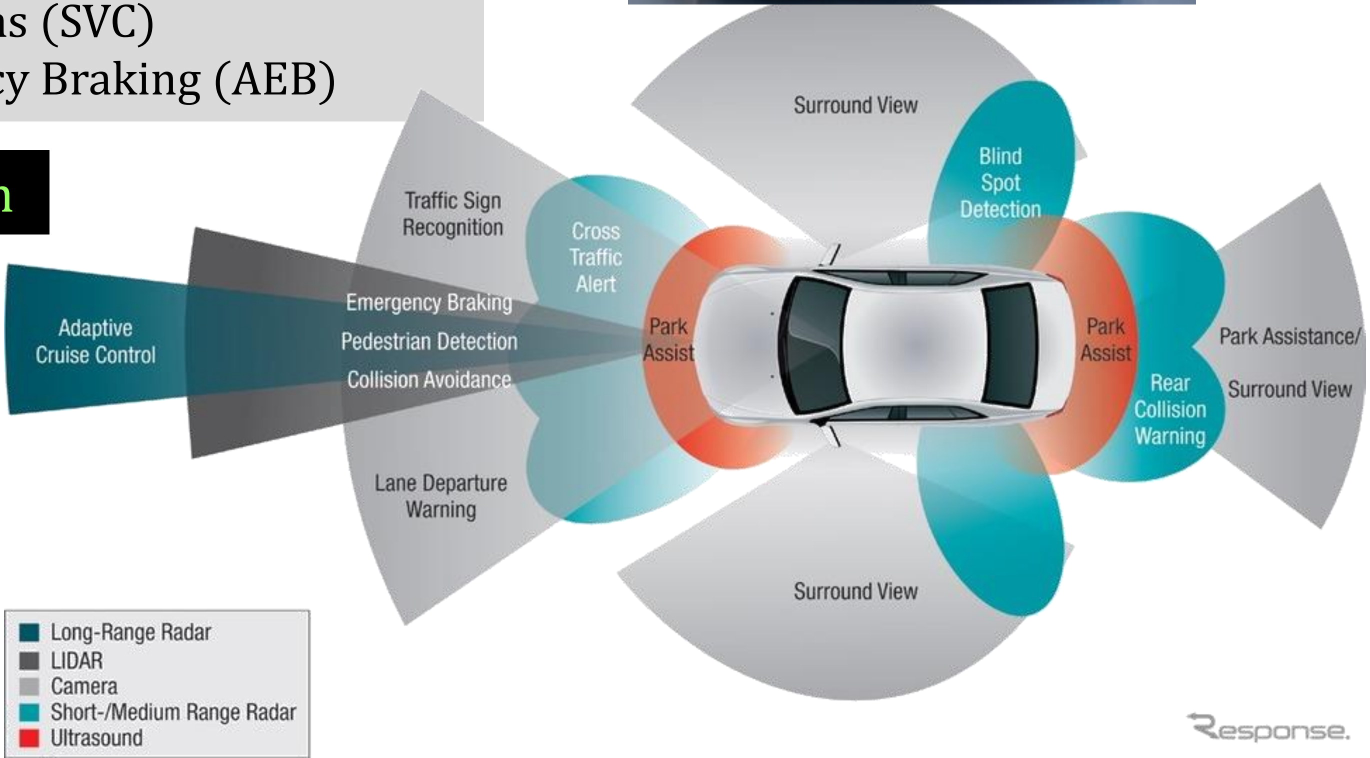
ADAS Applications are booming

- 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)

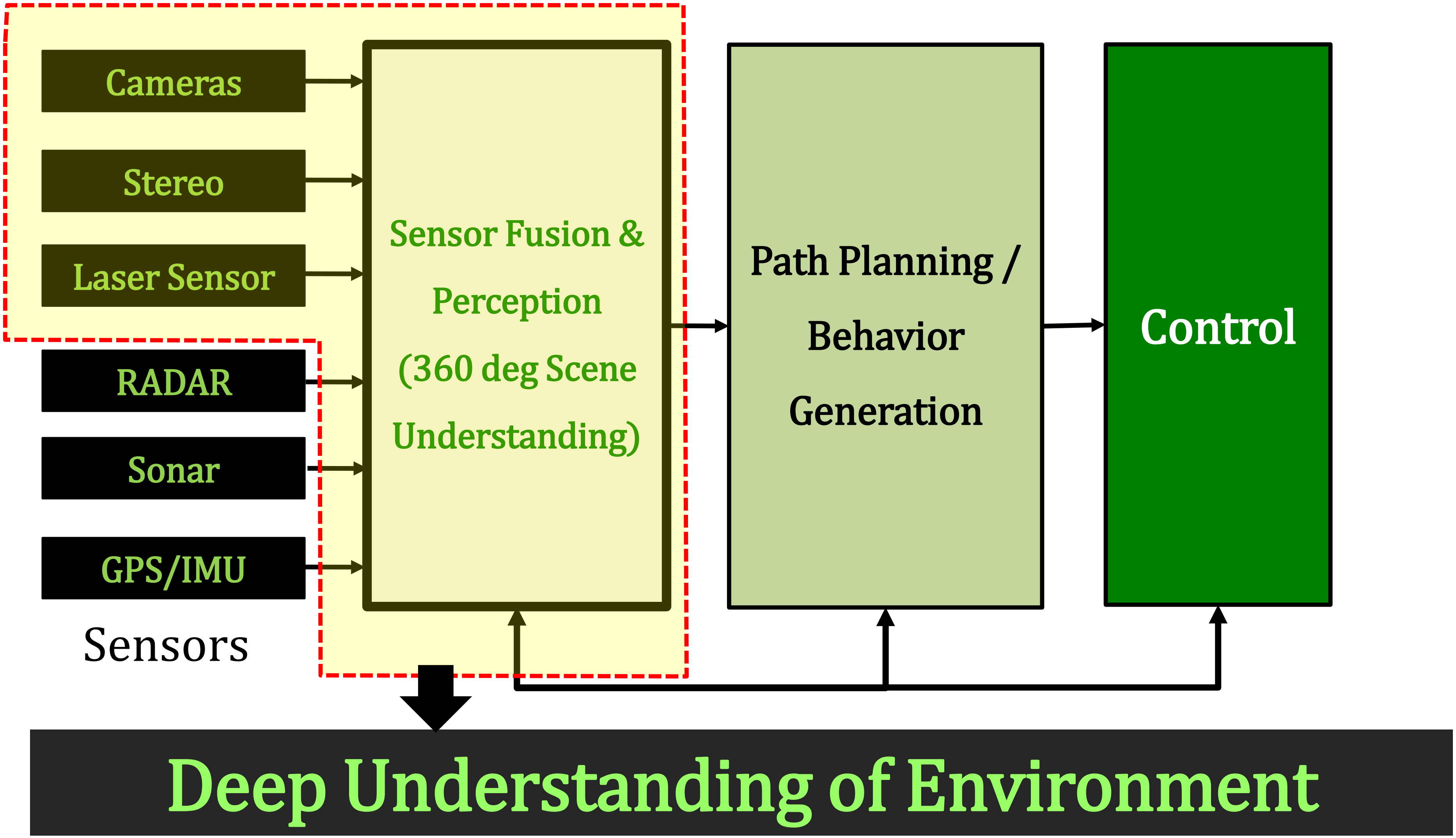
Vehicle Platform



Sensors Configuration



General Framework for ADAS and AD

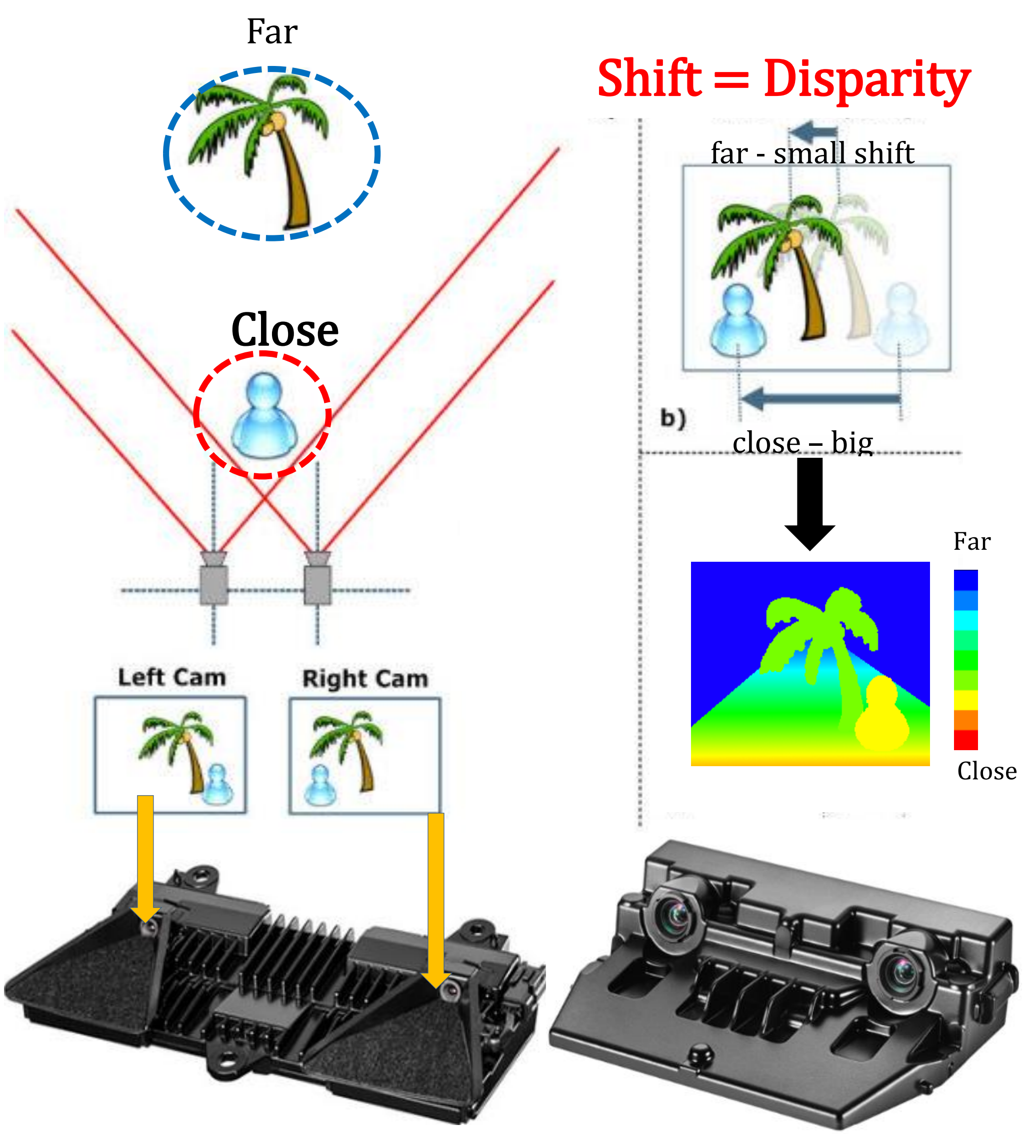


World 3D Reconstruction

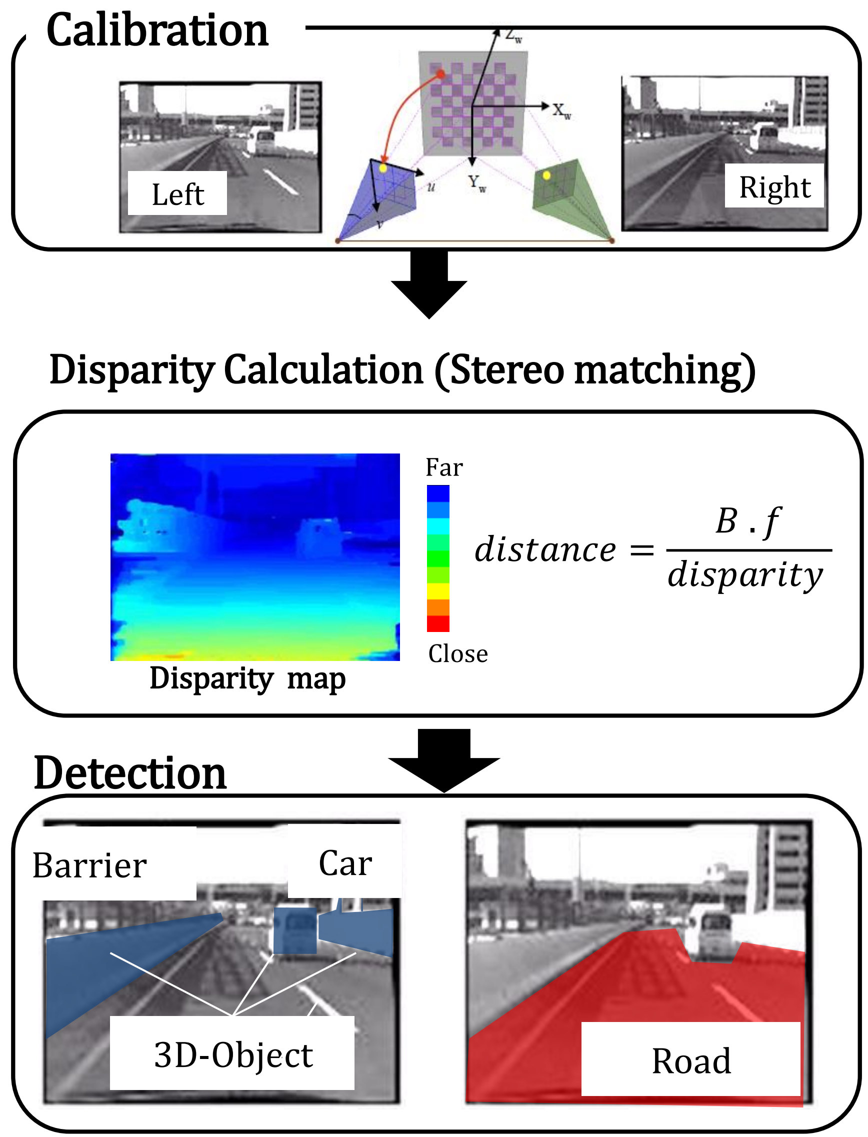
Dense Stereo Sensor

Stereo Vision Concept

<Stereo Vision>

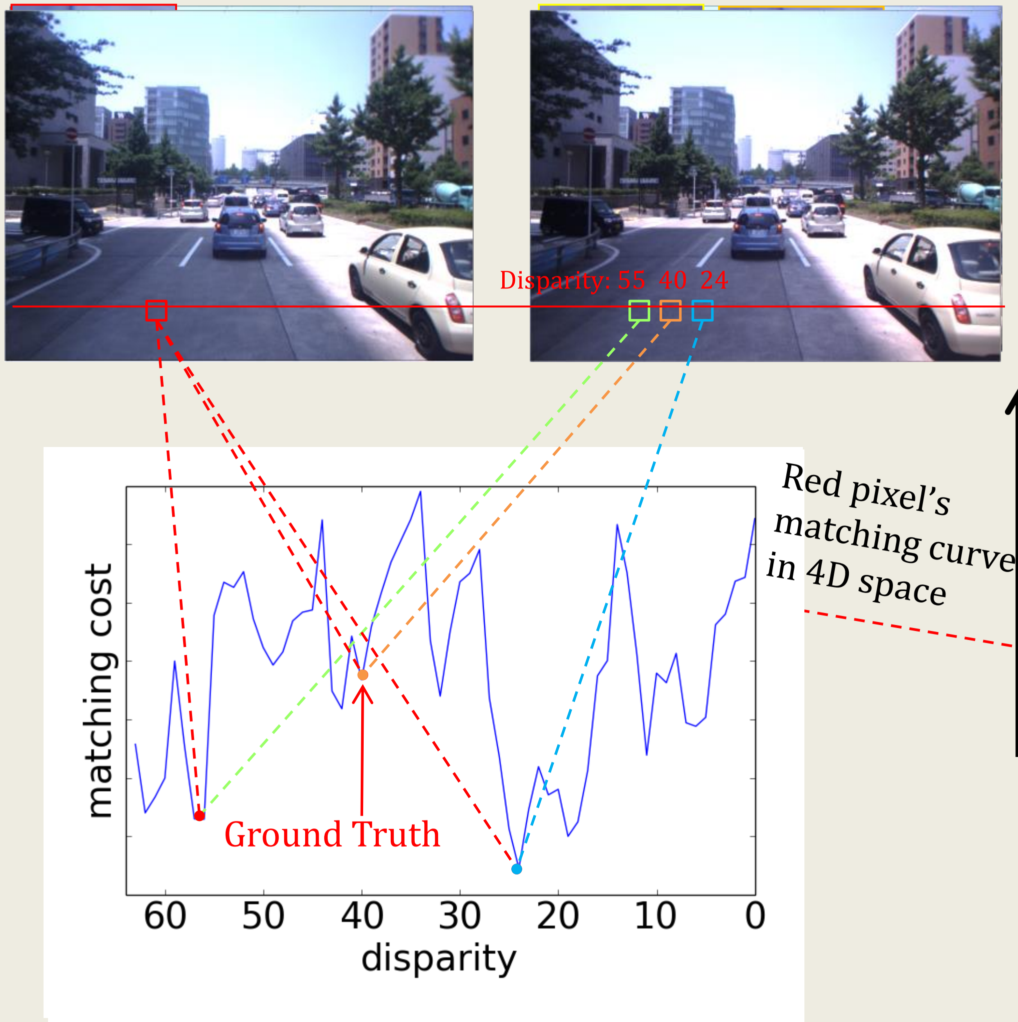


<Process>



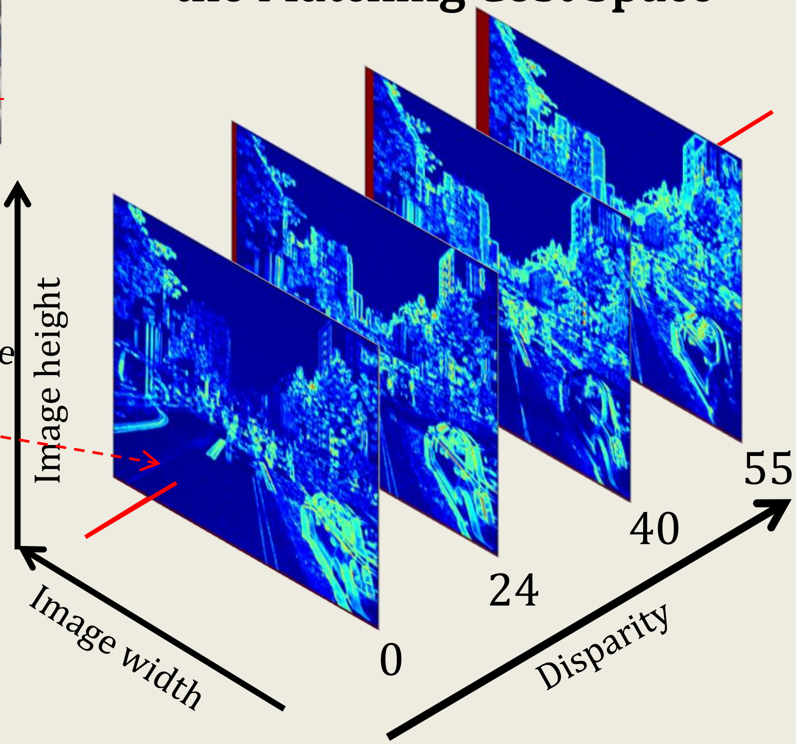
Matching Cost

< Matching Cost >



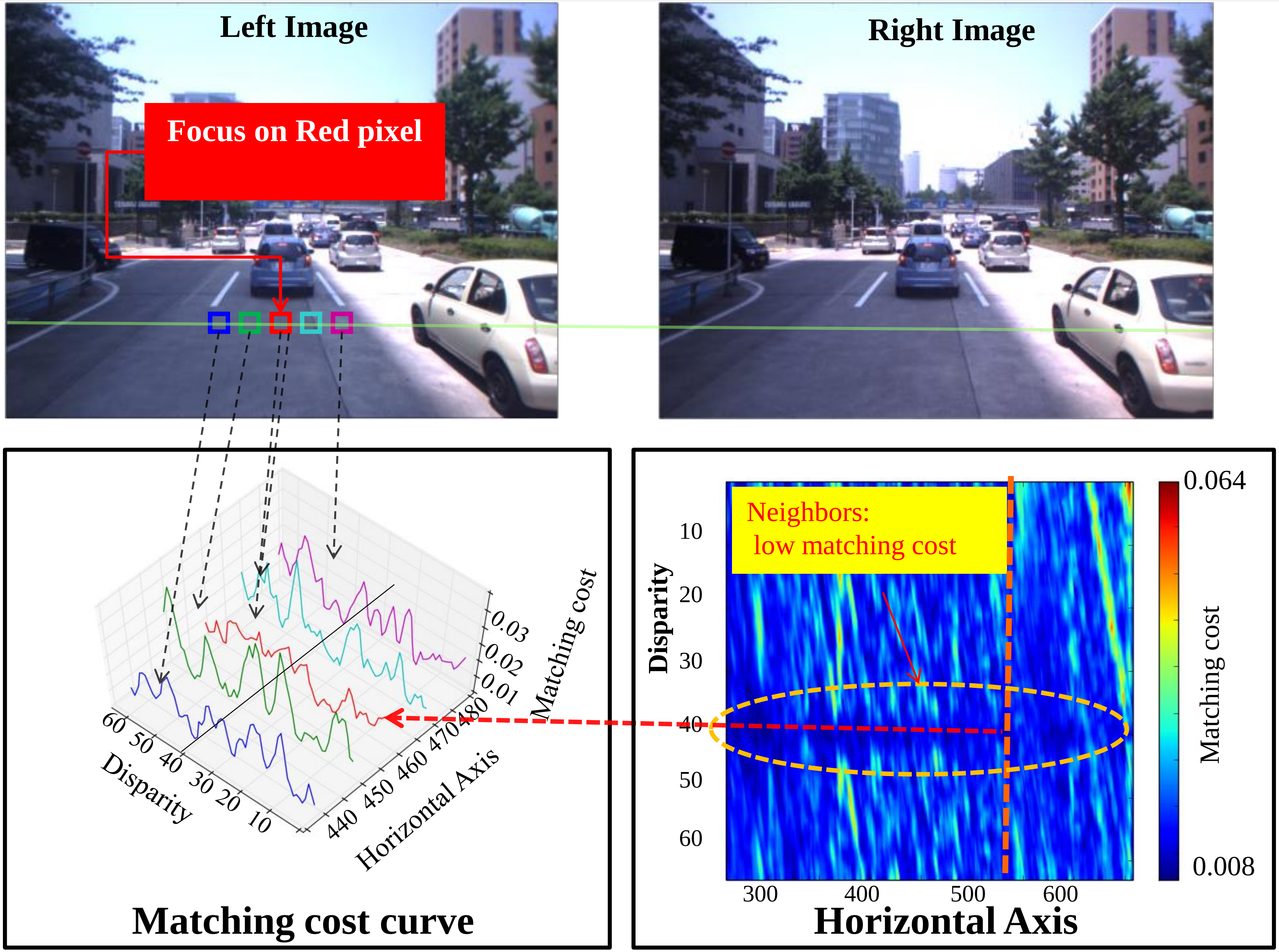
<Cost Space >

Each pixel has such matching cost curve, which constructs the Matching Cost Space



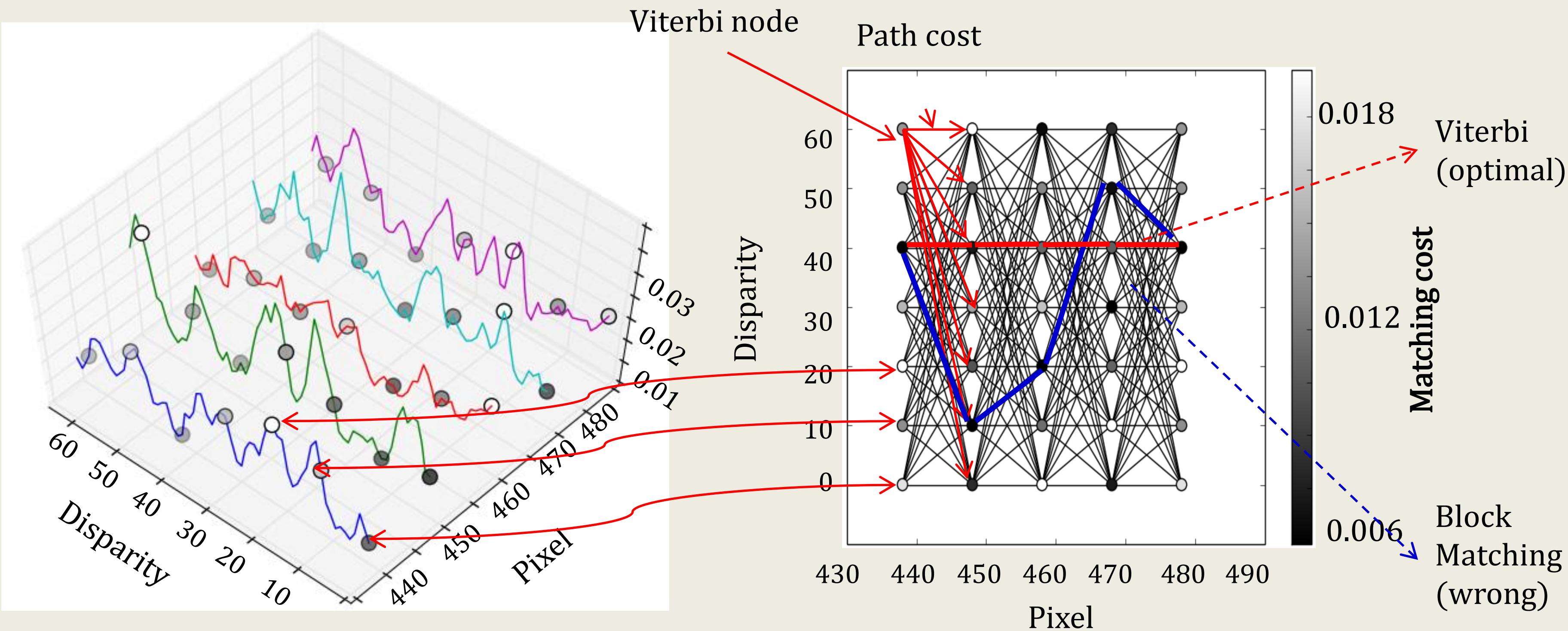
Finding true disparity value for every pixel from Matching Cost Space

Cost Calculation Method



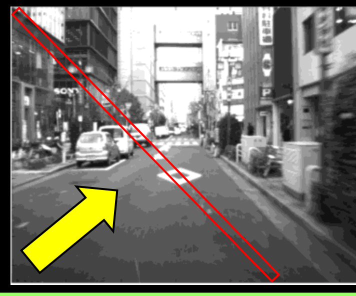
Optimization: Multi Viterbi Path

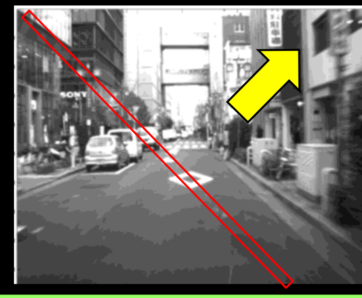
Exploiting the neighbors' matching cost can be translated into Mathematical Optimization about the **Shortest Path Problem**

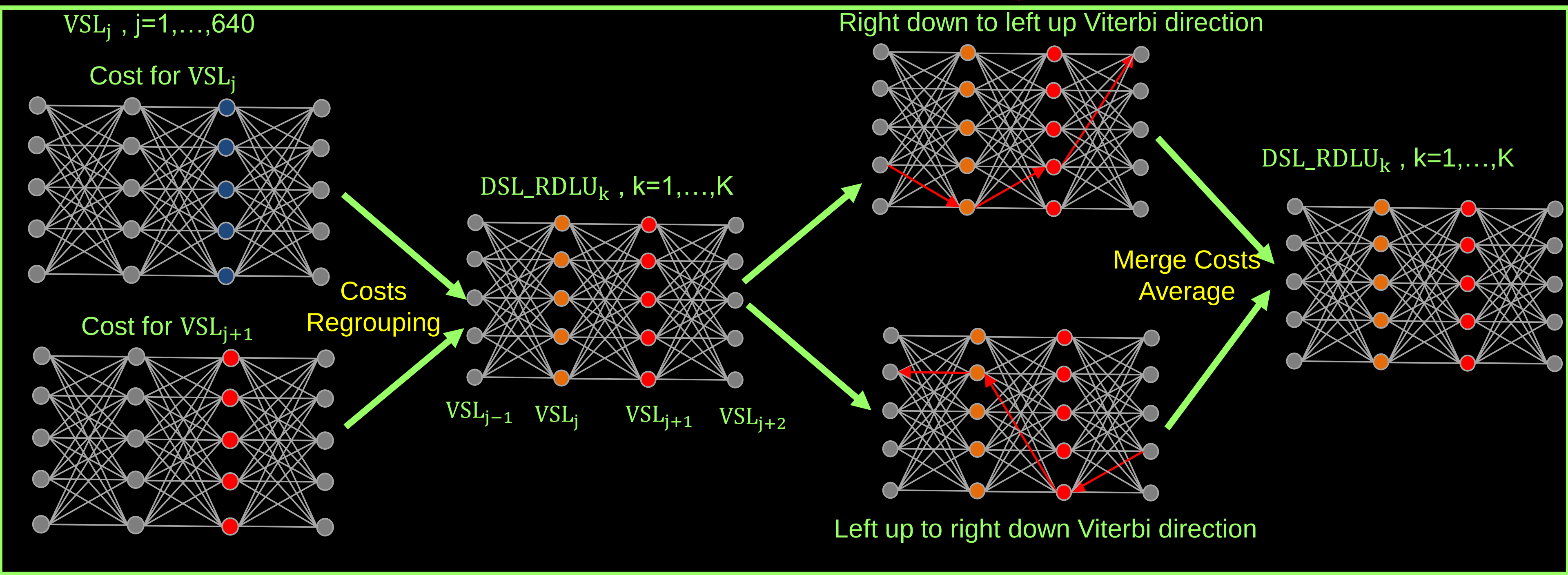


Viterbi algorithm can find the global optimum

Fast Implementation of Multi Path Viterbi

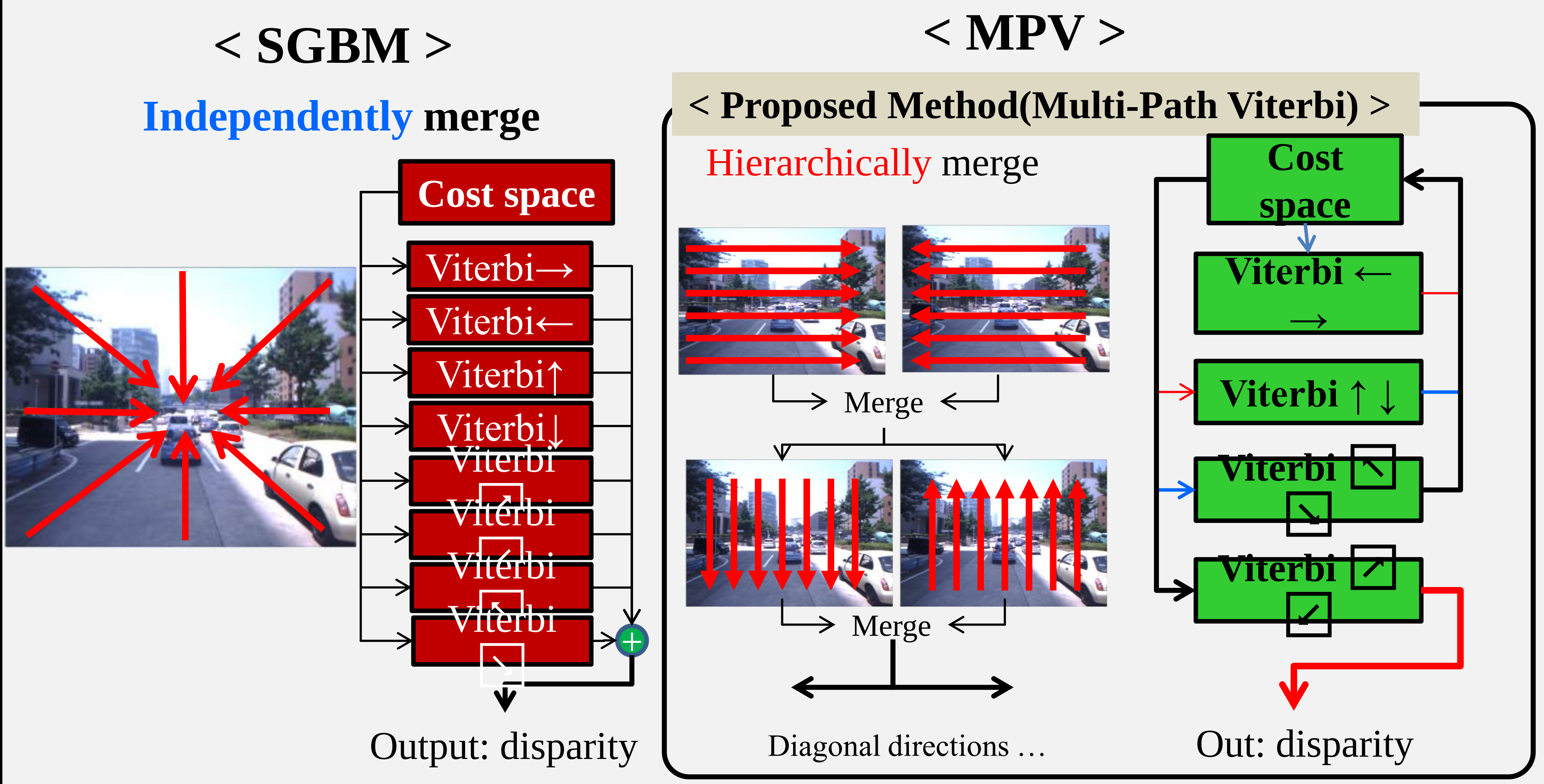






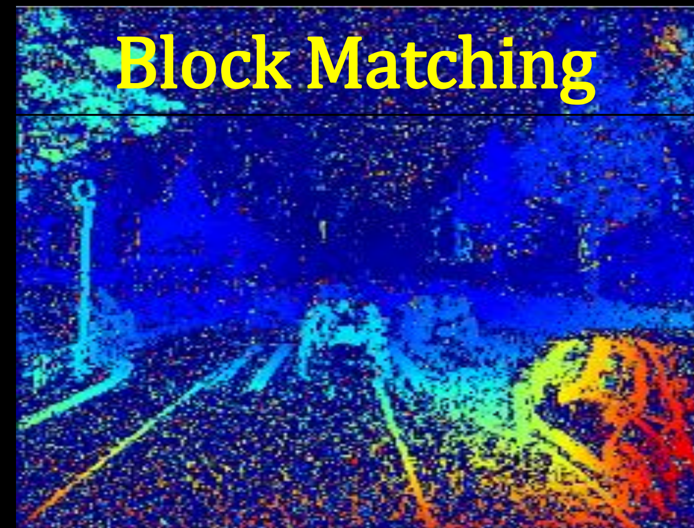
Huge Networks with Parallel Optimization

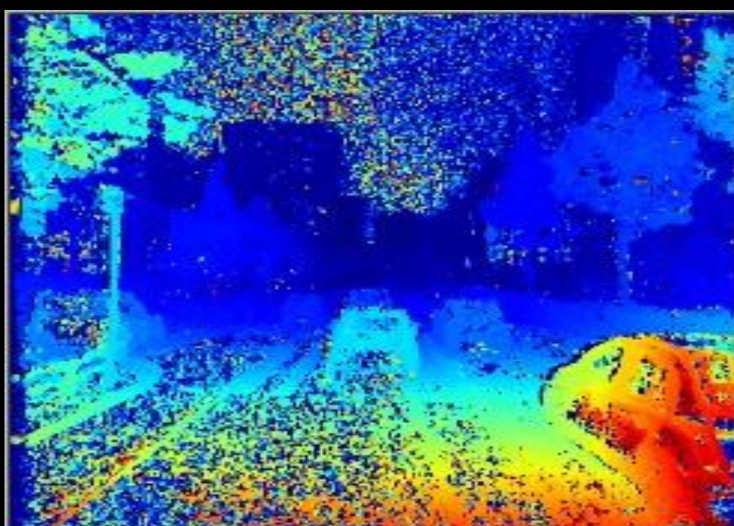
Proposed Method: Multi Path Viterbi



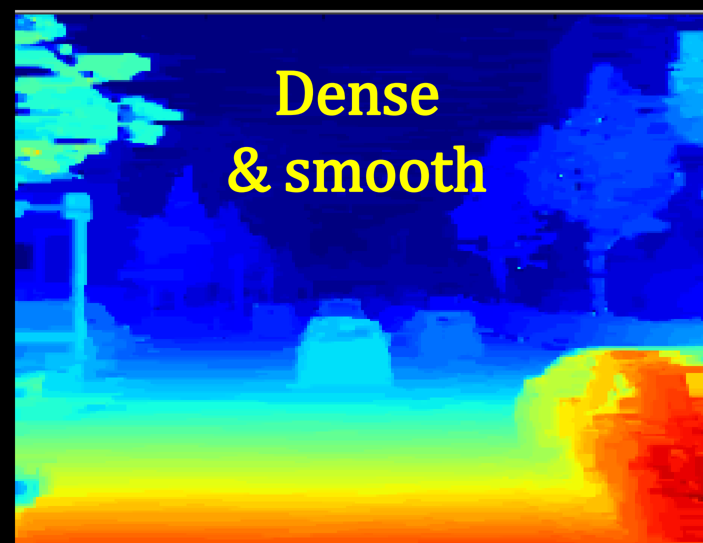
MPV Optimize the Accumulates Information Step by Step

Benchmark: Street Data





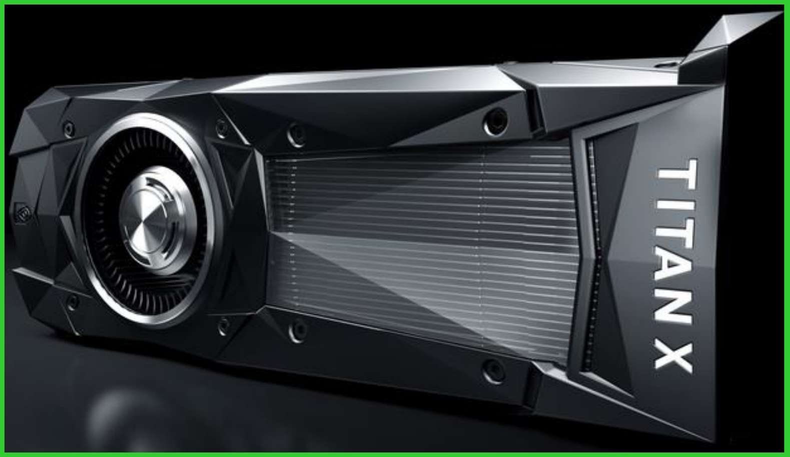
Conventional method: SGBM



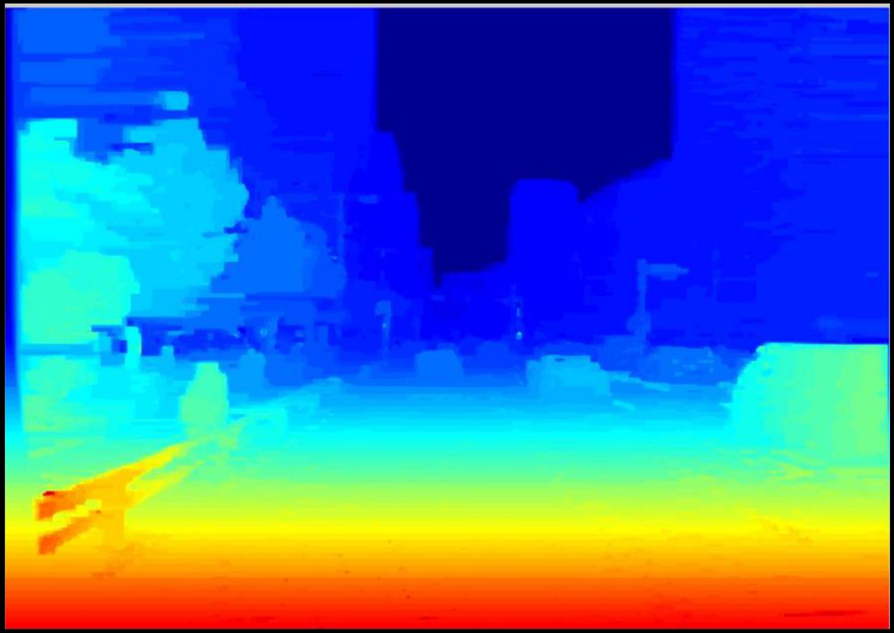
Multi Path Viterbi

Stereo Vision –Results (Urban)

Image Size　　　:　１２８０×９６０
Calculation Time :　　15ms／Frame
GPU　　:　　GeFORCE GTX 1080



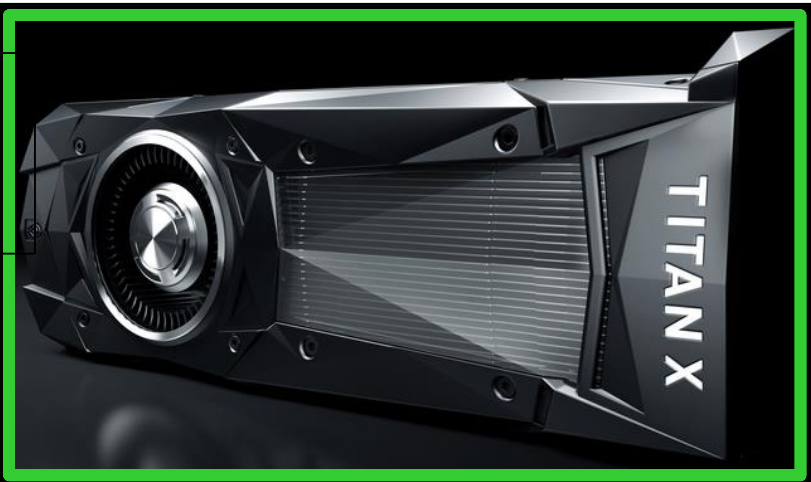
Nagoya Urban Road

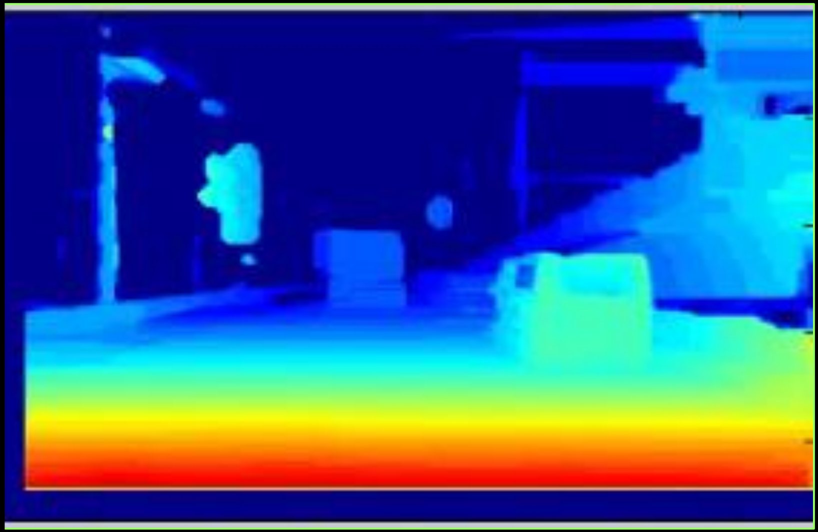


Stereo Vision –Results (Highway)

Calculation Time :15ms／Frame

GPU : GeFORCE GTX 1080

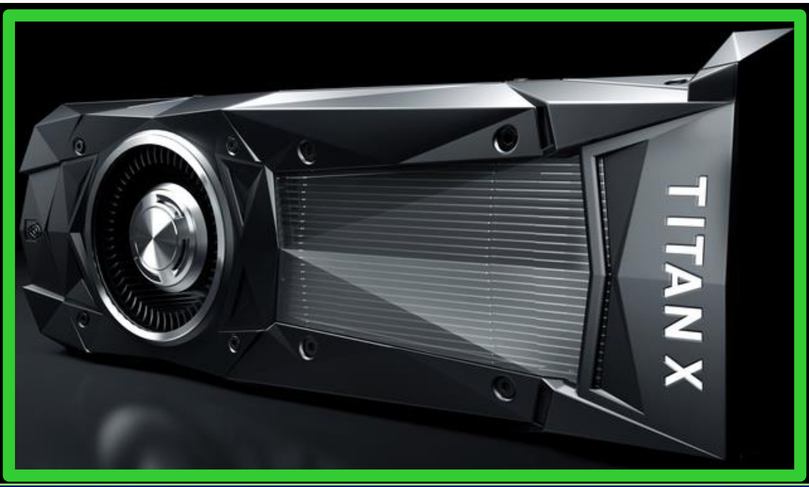


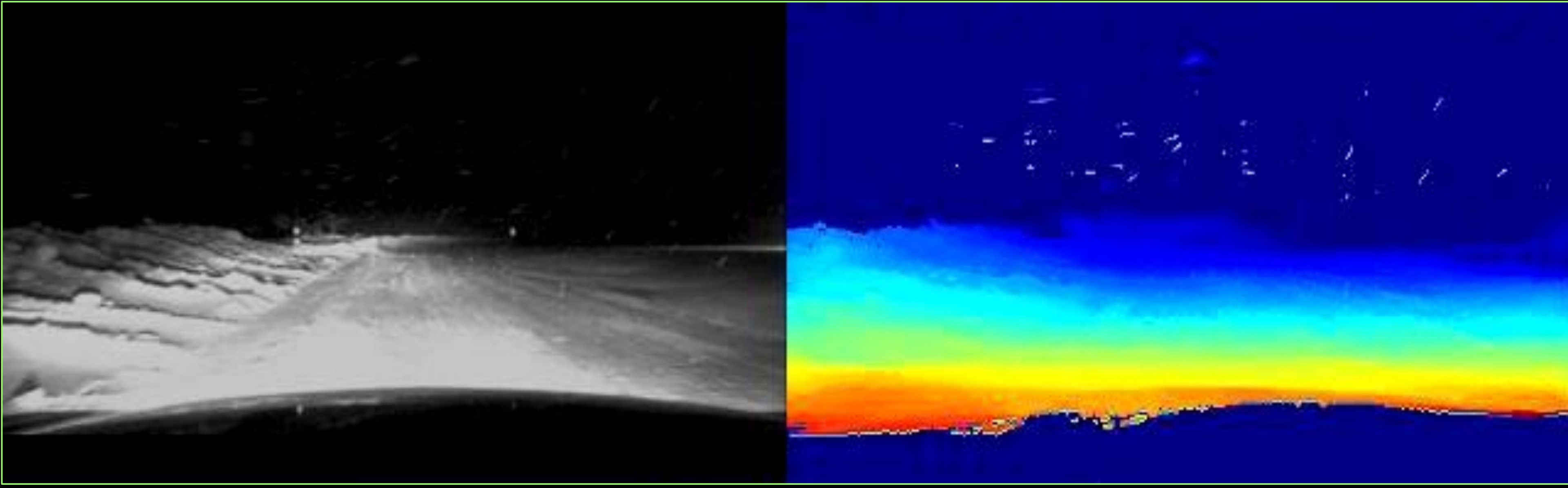


Tokyo Metropolitan Highway

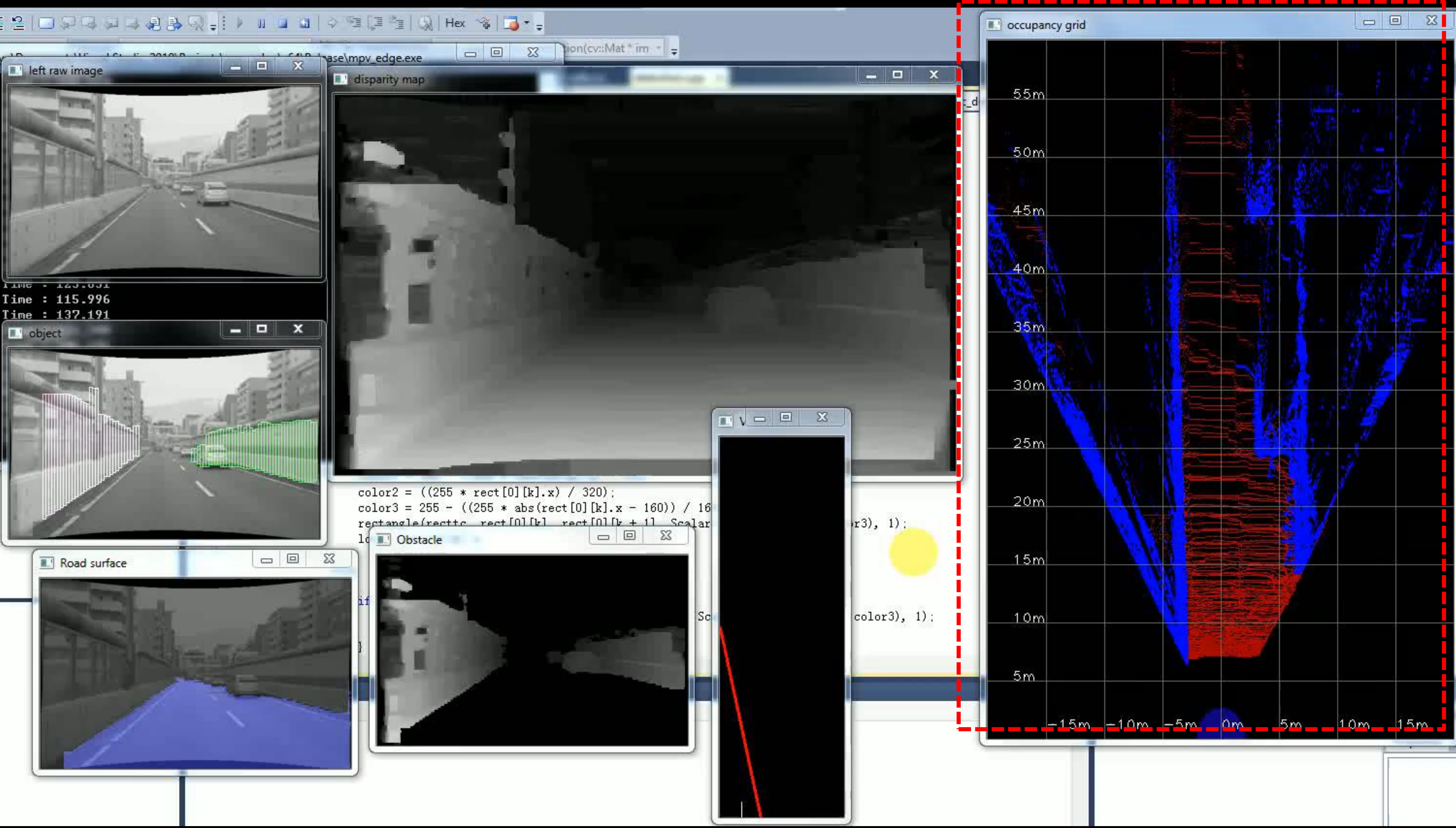
Stereo Vision –Snowy and Dark Weather

Calculation Time :　　15ms／Frame
GPU　　:　　GeFORCE GTX 1080





World 3D Representation

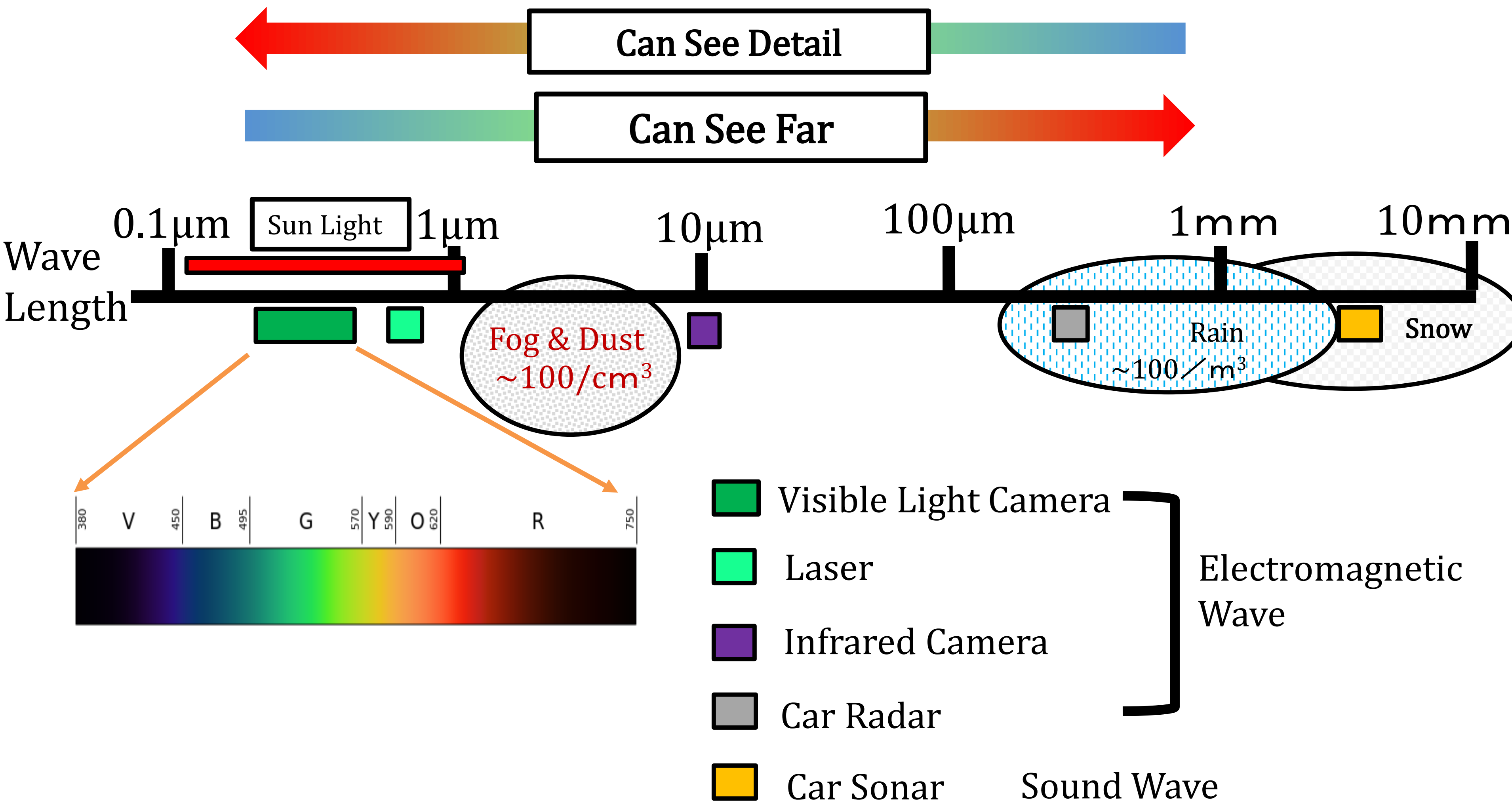


3D Deep Sensor Fusion

Sensors Properties

Electromagnetic Range Wave: Camera ,Laser, Radar

Sensor Fusion is necessary to cover Detail and Far Detection



Learning Framework for Perception

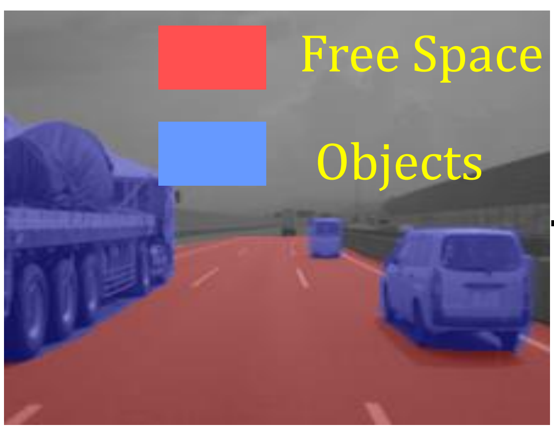
Training a learning framework for perception tasks

Sensors

Cameras

Stereo

Laser Sensor



Perception (Learning Framework)

Traditional Learning

Feature Extraction (HOG, DPM etc)

Feature Classification (SVM, Random Forest etc)

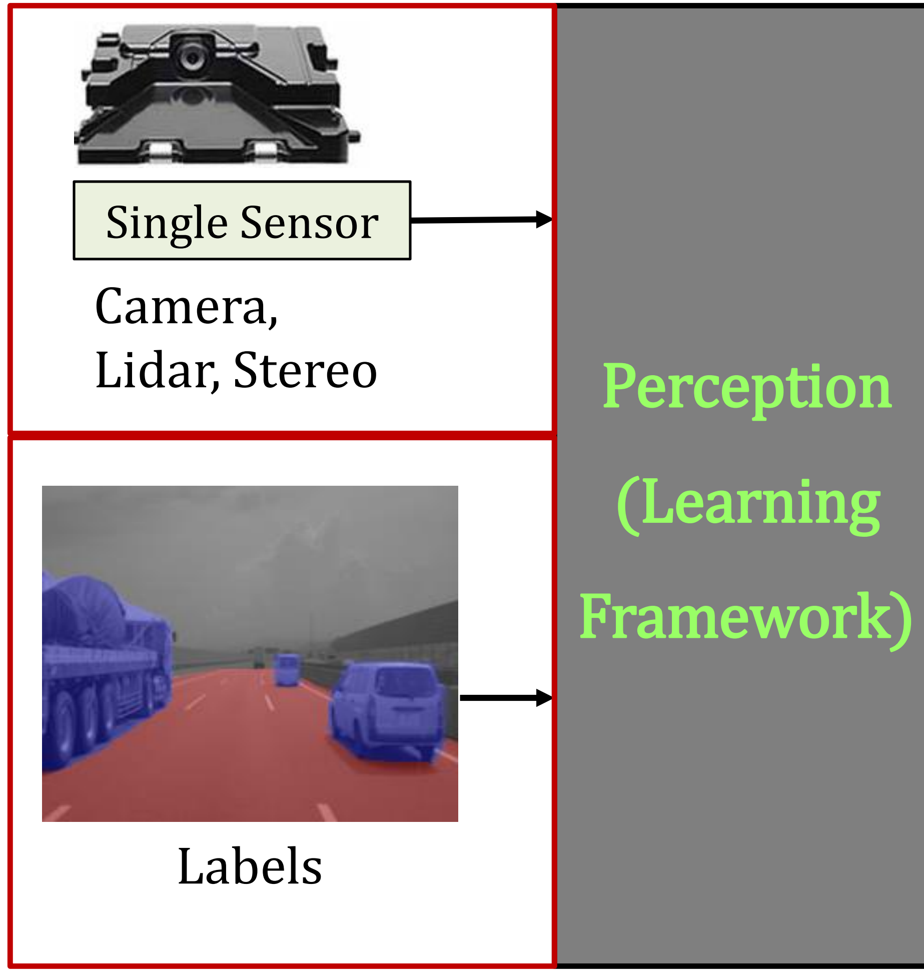
Deep learning

(**Feature Extraction + Feature Classification**)

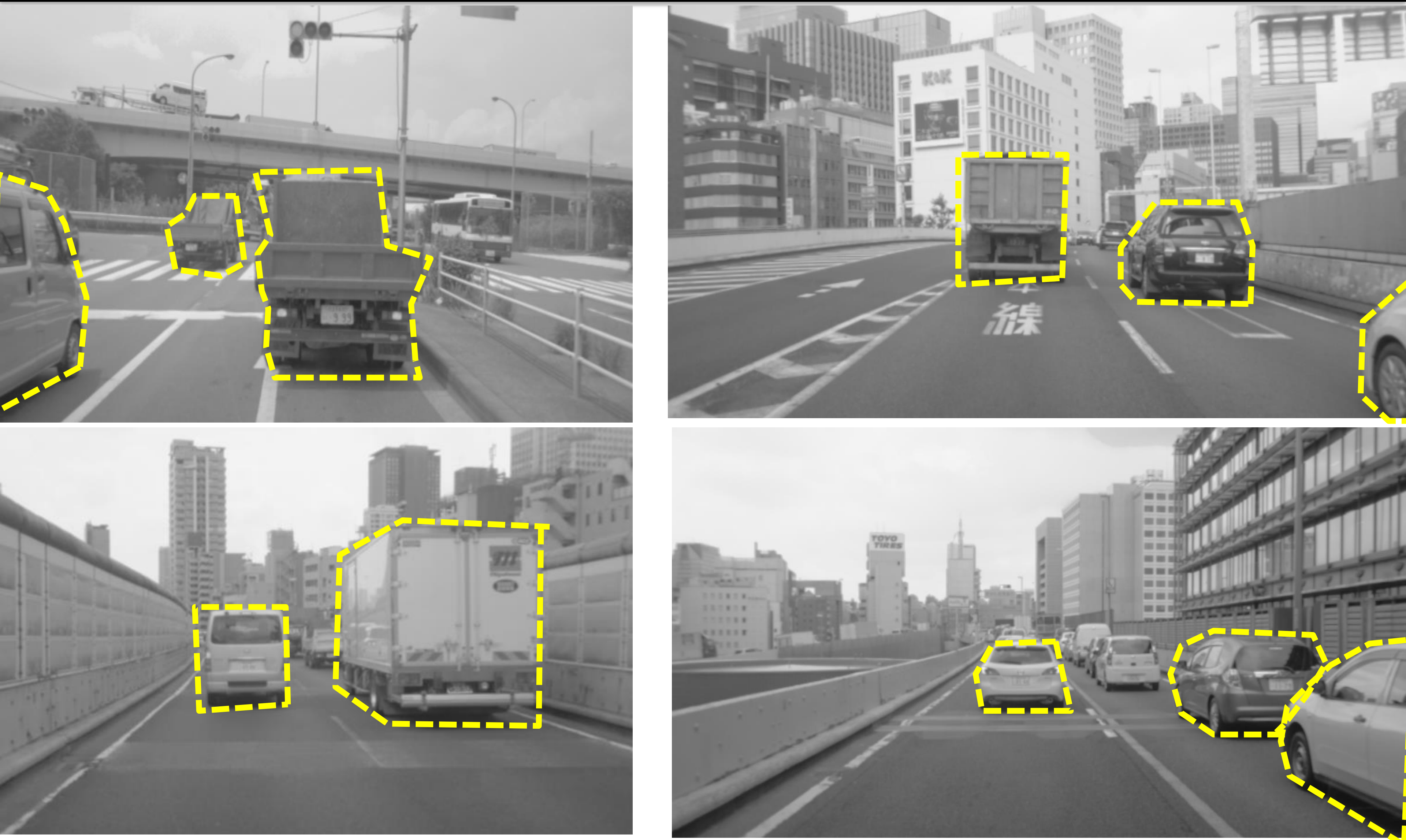
Single Sensor-based learning

- Single sensor-based learning is not robust or descriptive enough

- **Challenges**
  - Environmental Variation (occlusion, illumination variation, etc.)
  - High Inter-Class and Intra-Class Variability



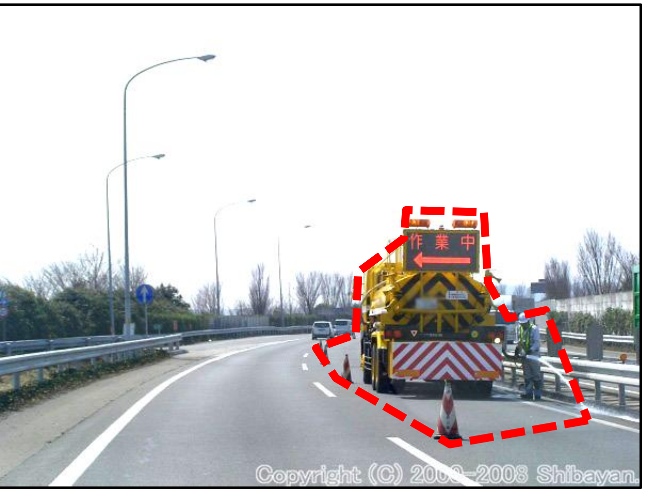
Intra-class Vehicle Variations

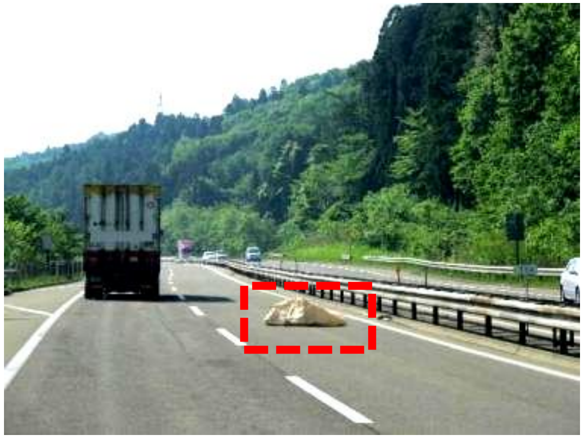


There are many vehicle varieties with different orientations

Intra-class On-Road Objects Variations

We have a large number of On-Road Objects

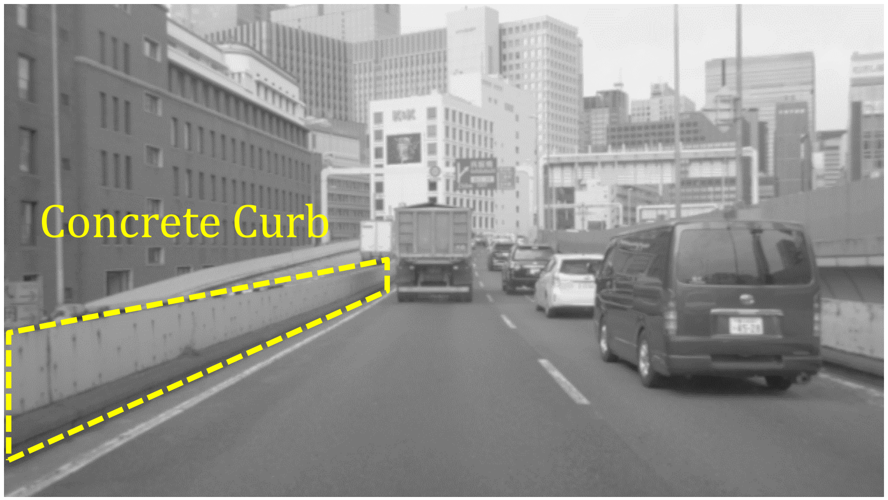


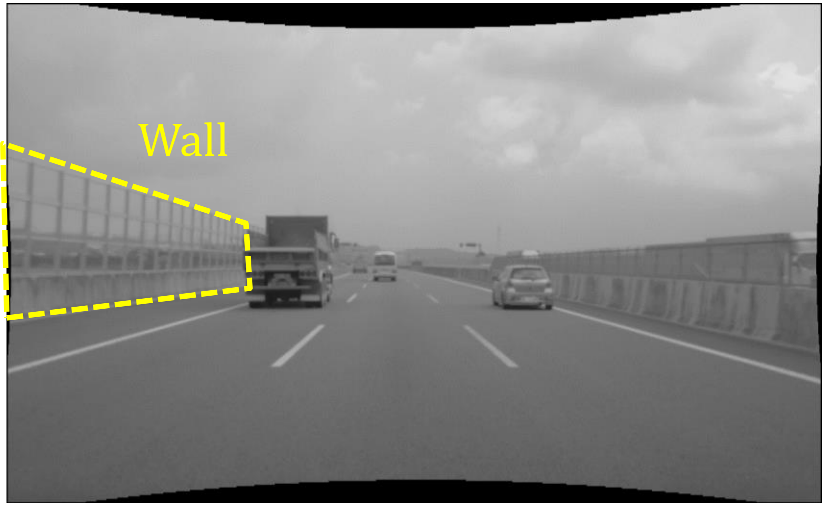


We have a lot of variety of on road objects!!!!

Intra-Class Free Space Boundary Variations

We have the different type of road boundaries





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

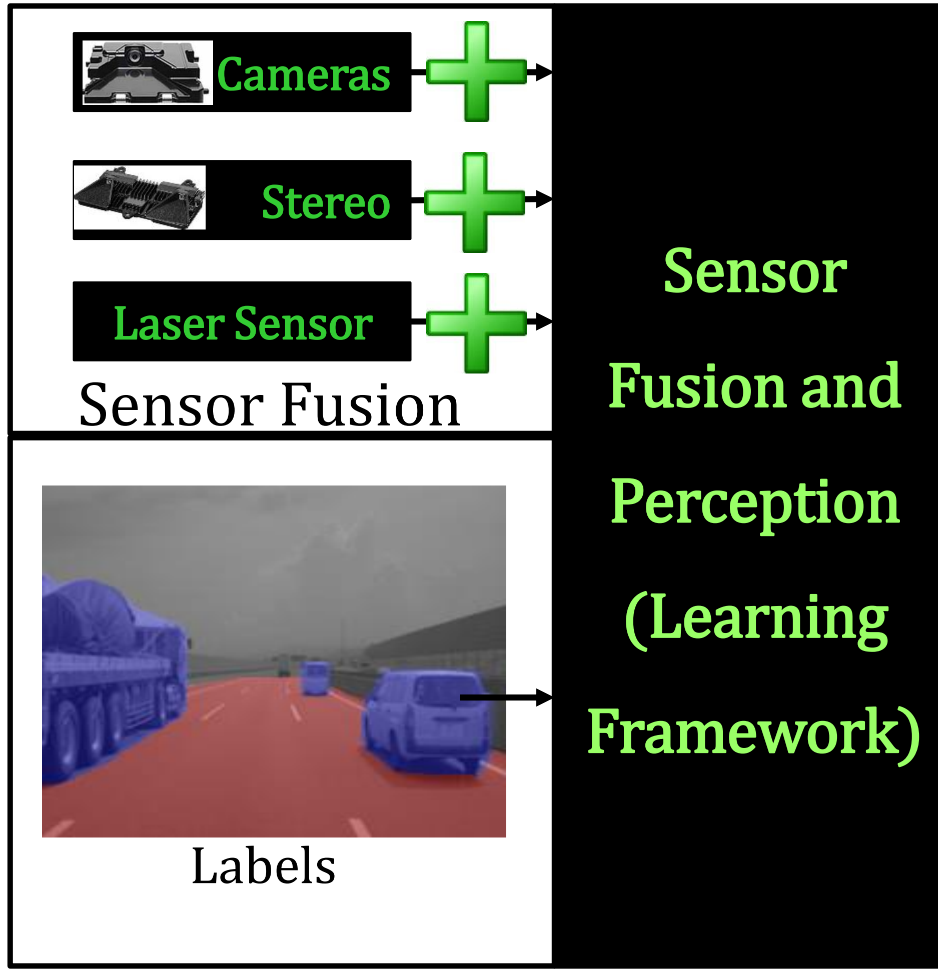
Environmental Variation

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

Sensor Fusion-based learning

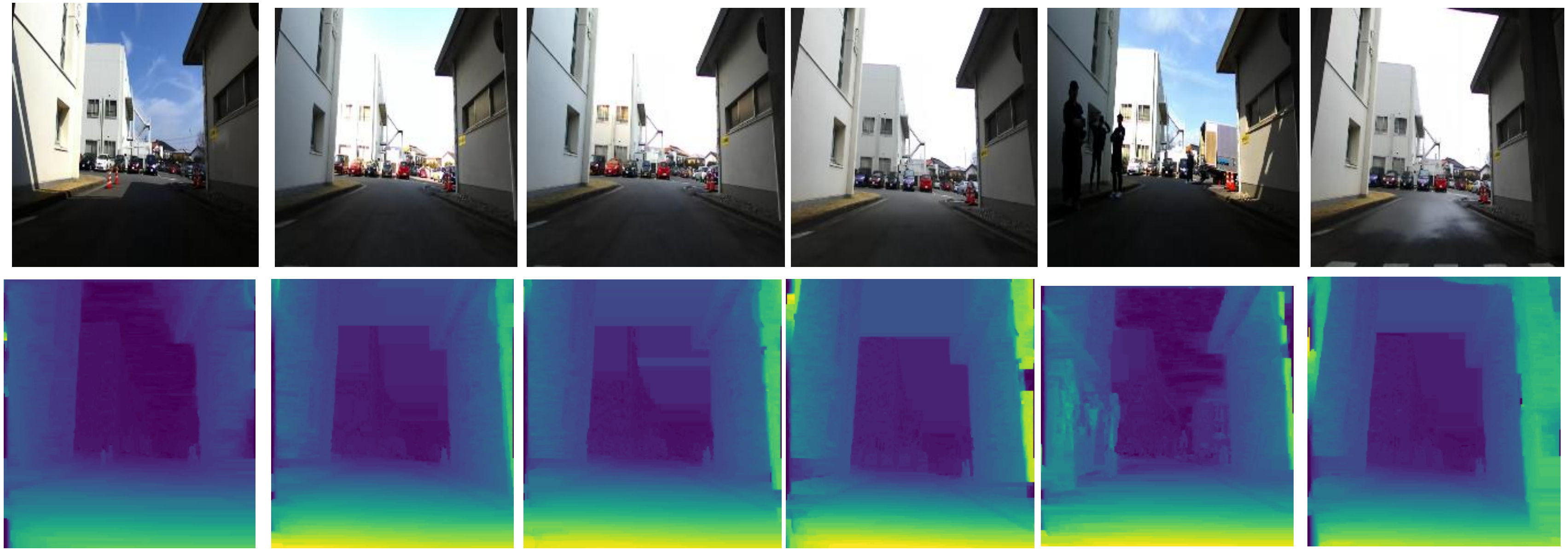
- Sensor Fusion-based learning with **Complementary Sensors** addresses these issues

- Monocular Camera appearance features and depth features are **Complementary Features**



Complementary sensors

Monocular Camera	Depth Camera
Monocular Camera ⇒ Rich Appearance Information	Depth Camera ⇒ Depth Information (3D Data)
Inexpensive	**Stereo-based Depth Inexpensive**
Illumination Variation	**Illumination Invariant** due to robust stereo algorithm [1]



Depth information from stereo camera robust to illumination variation

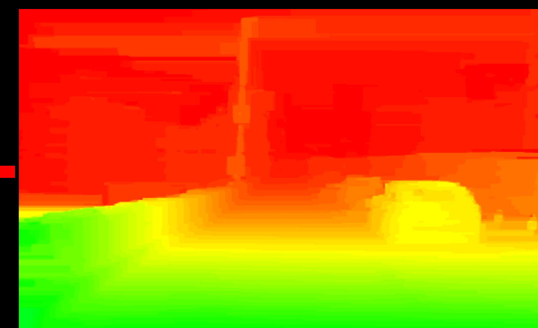
[1] Xu et al. Real-time Stereo Disparity Quality Improvement for Challenging Traffic Environments, IV 2018

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

Appearance
(Monocular camera)
Descriptive
Appearance Features



Depth
(Stereo Camera/Laser)
Illumination invariant
depth features

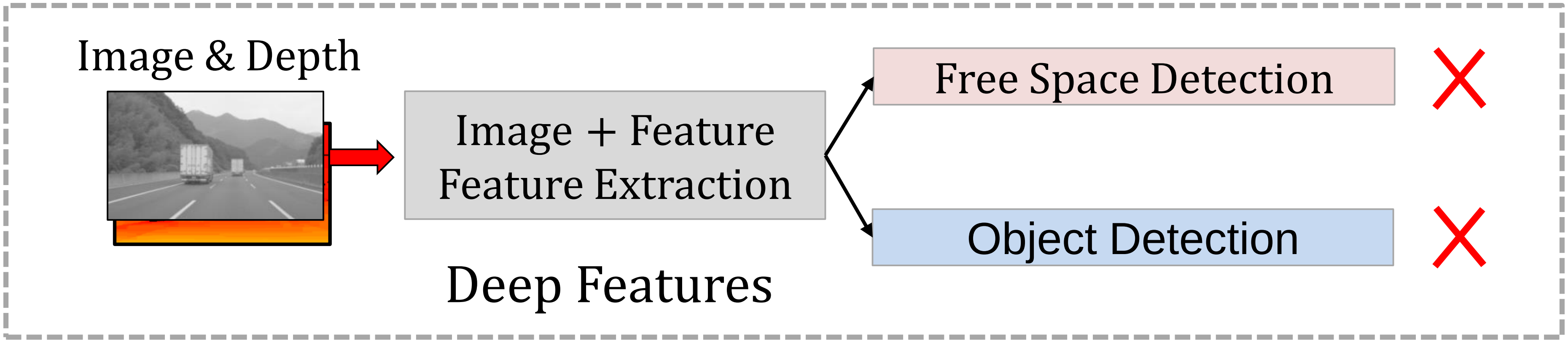
Deep learning framework

3D Environment Perception

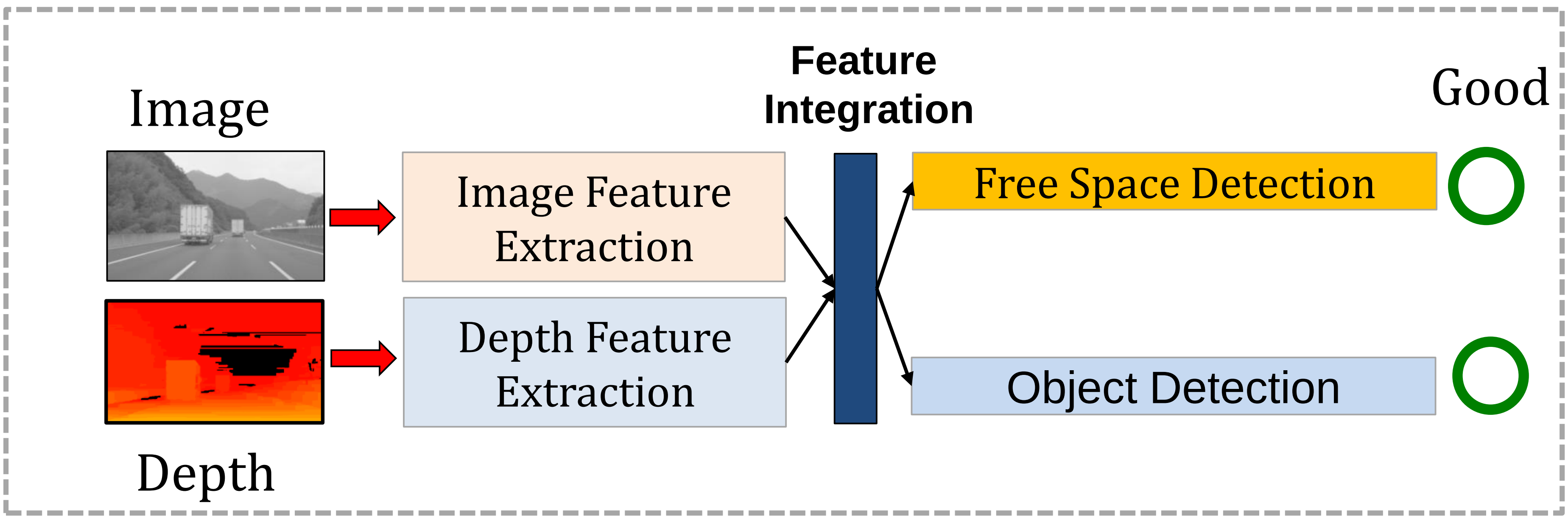
How to Fuse Sensors Data ??

☐ Sensor Fusion : Raw Data Level Fusion

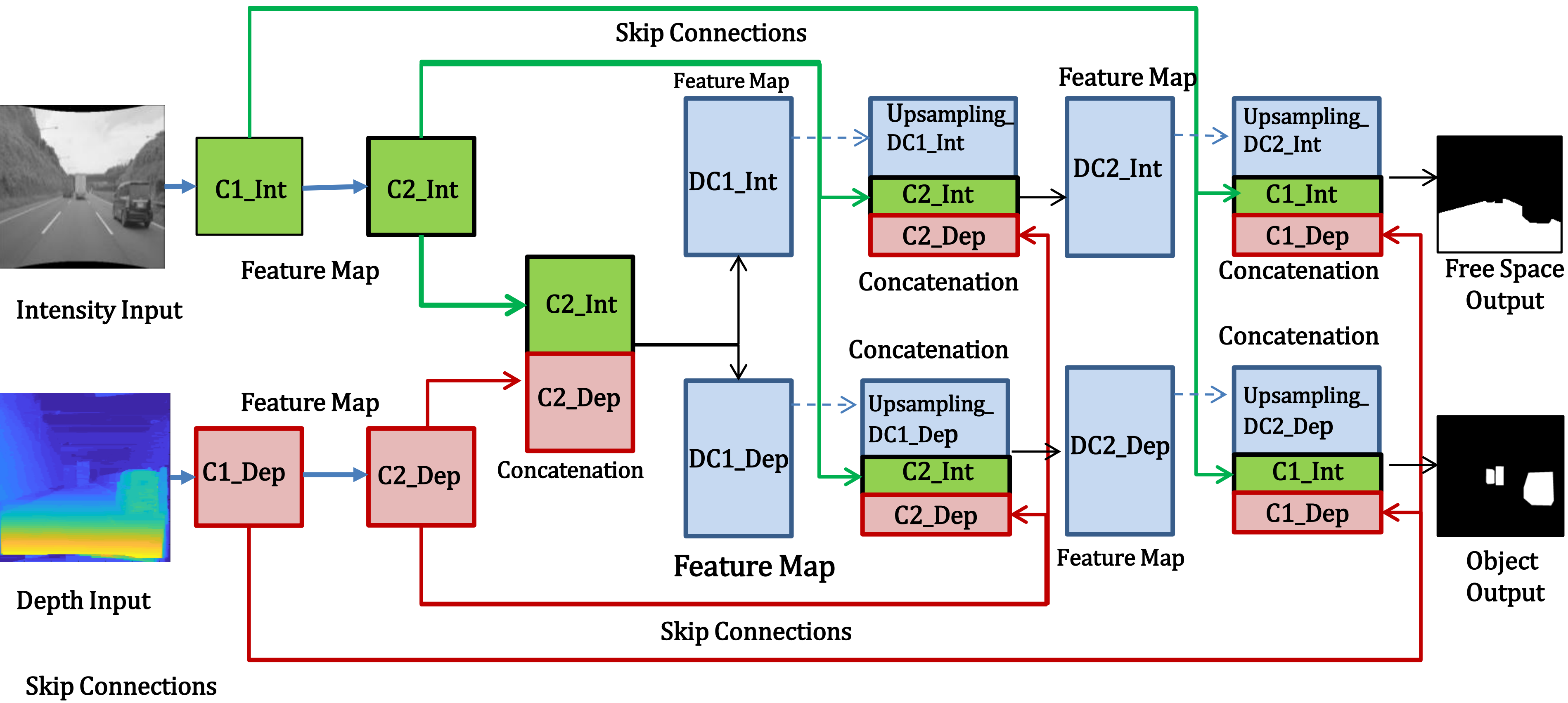
Bad



☐ Sensor Fusion : Feature Level Fusion



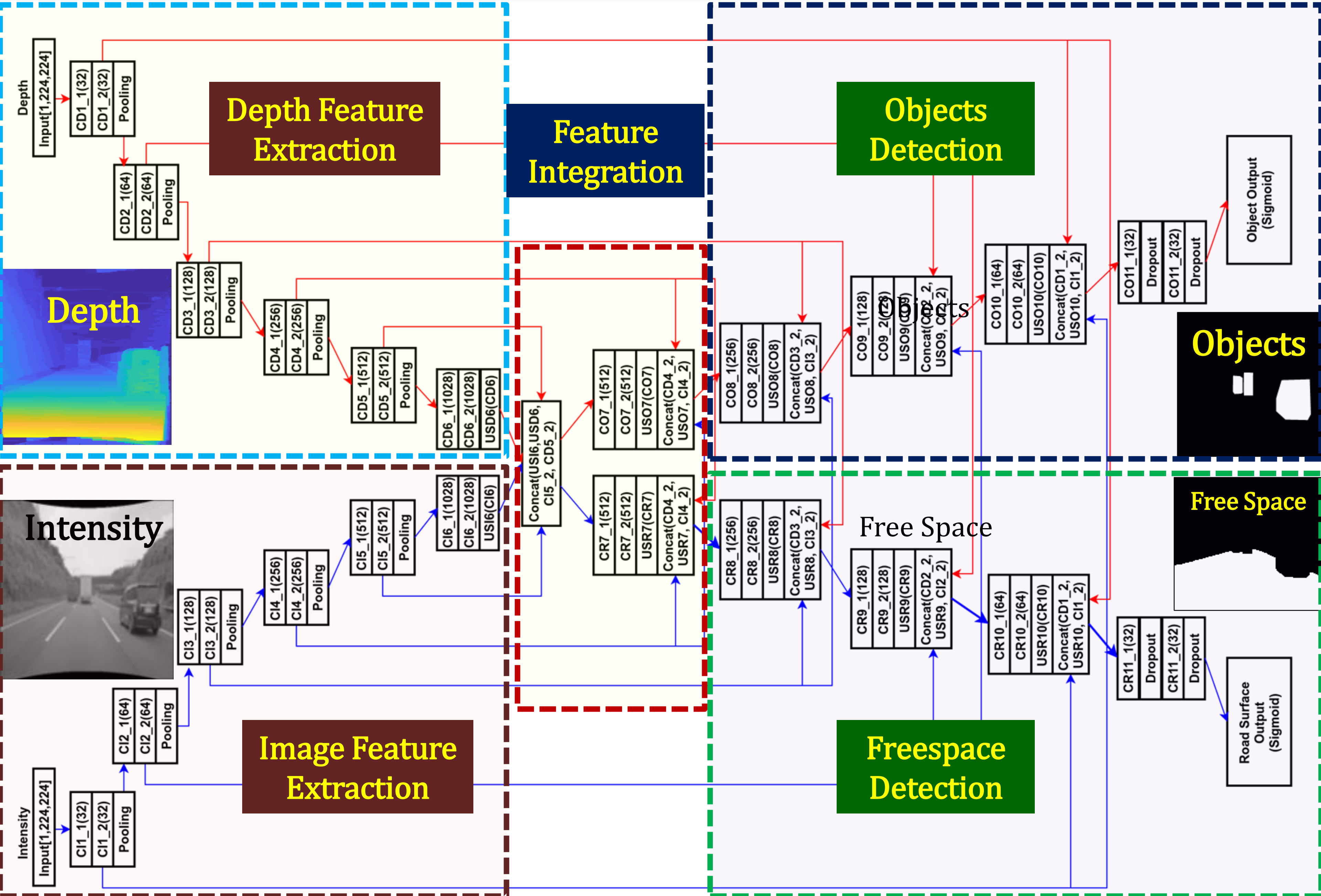
Proposed Model



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

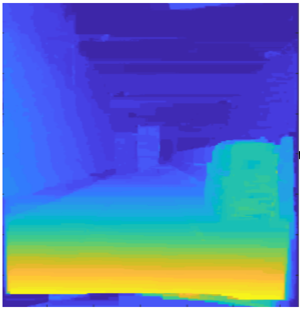
Entire Intensity Encoder Feature Maps (m,n,n) are transferred to Free Space and Object Decoder Feature Maps (o,n,n) for Concatenation ($m+o,n,n$)

Final Proposed Architecture

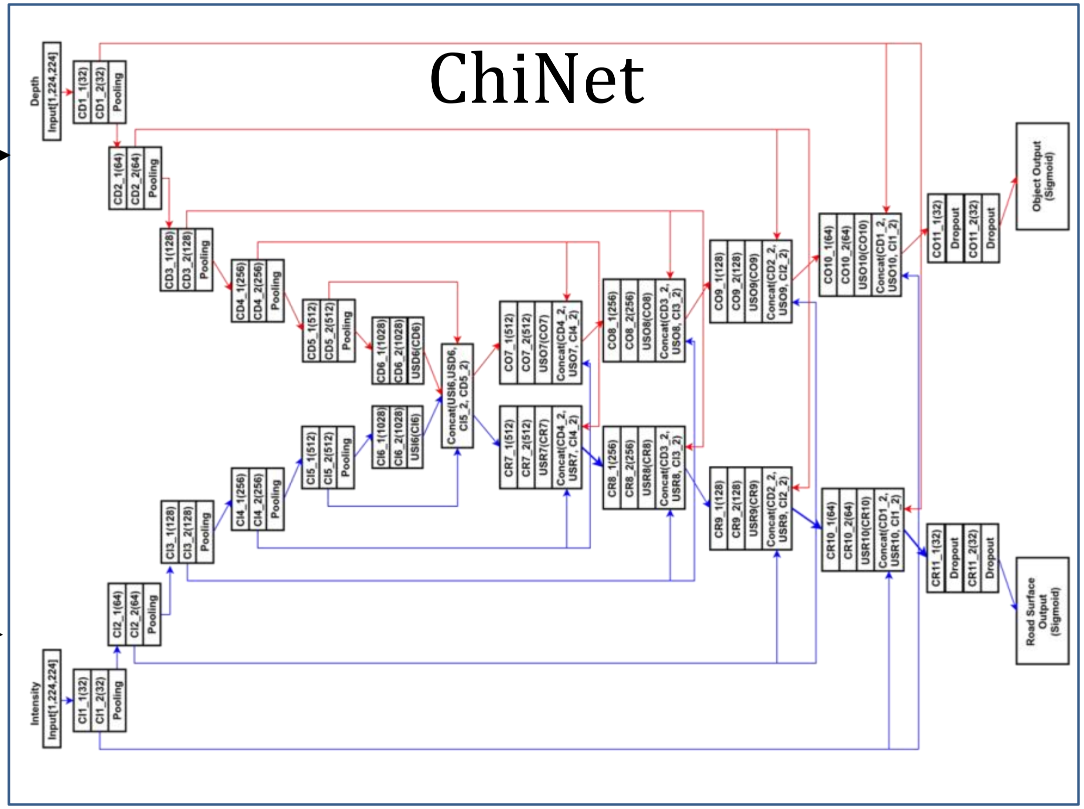


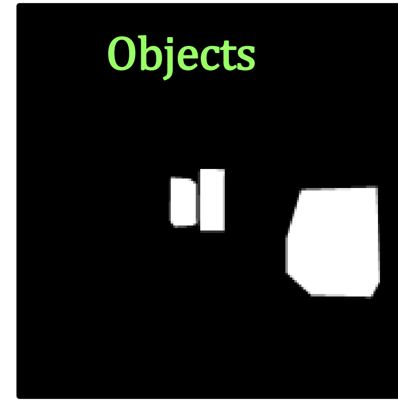
Learning Chi-Net

Disparity Image



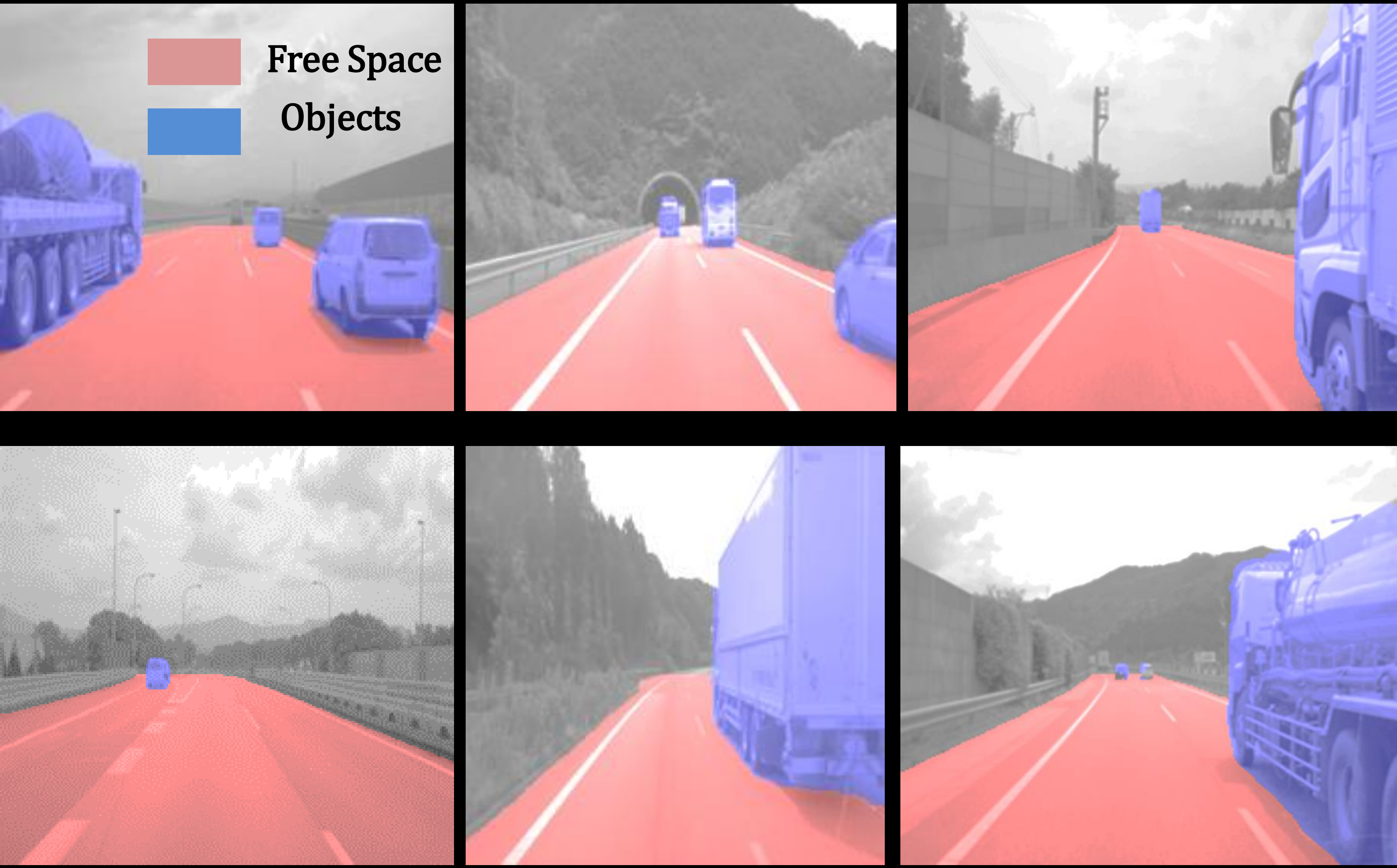
Intensity Image



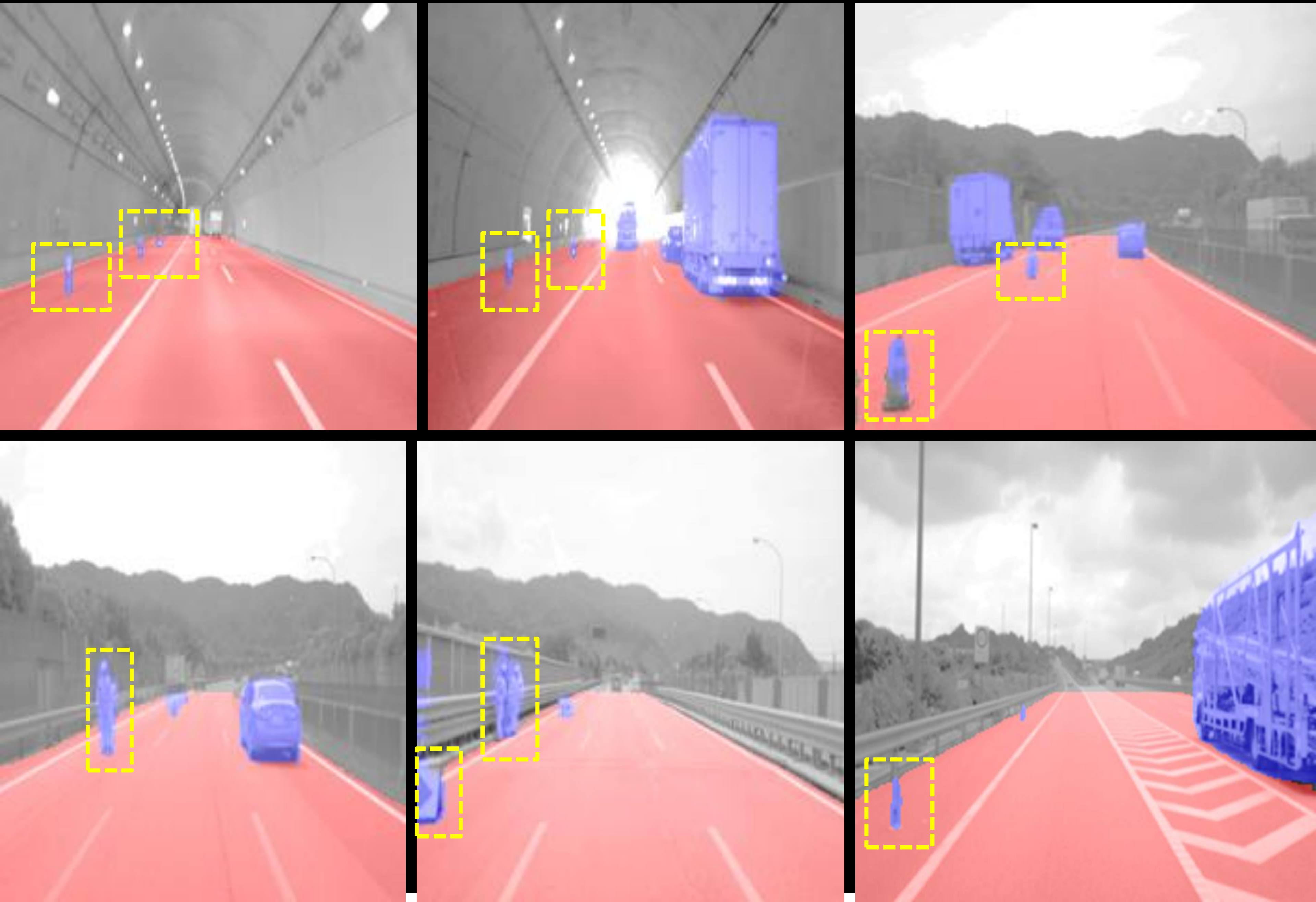


- Trained with **9000 Samples** from Japanese Highway dataset
  - Manually annotated free space and objects
- Trained on **Keras with theano backend**
- Trained with **Nvidia Titan X GPU**

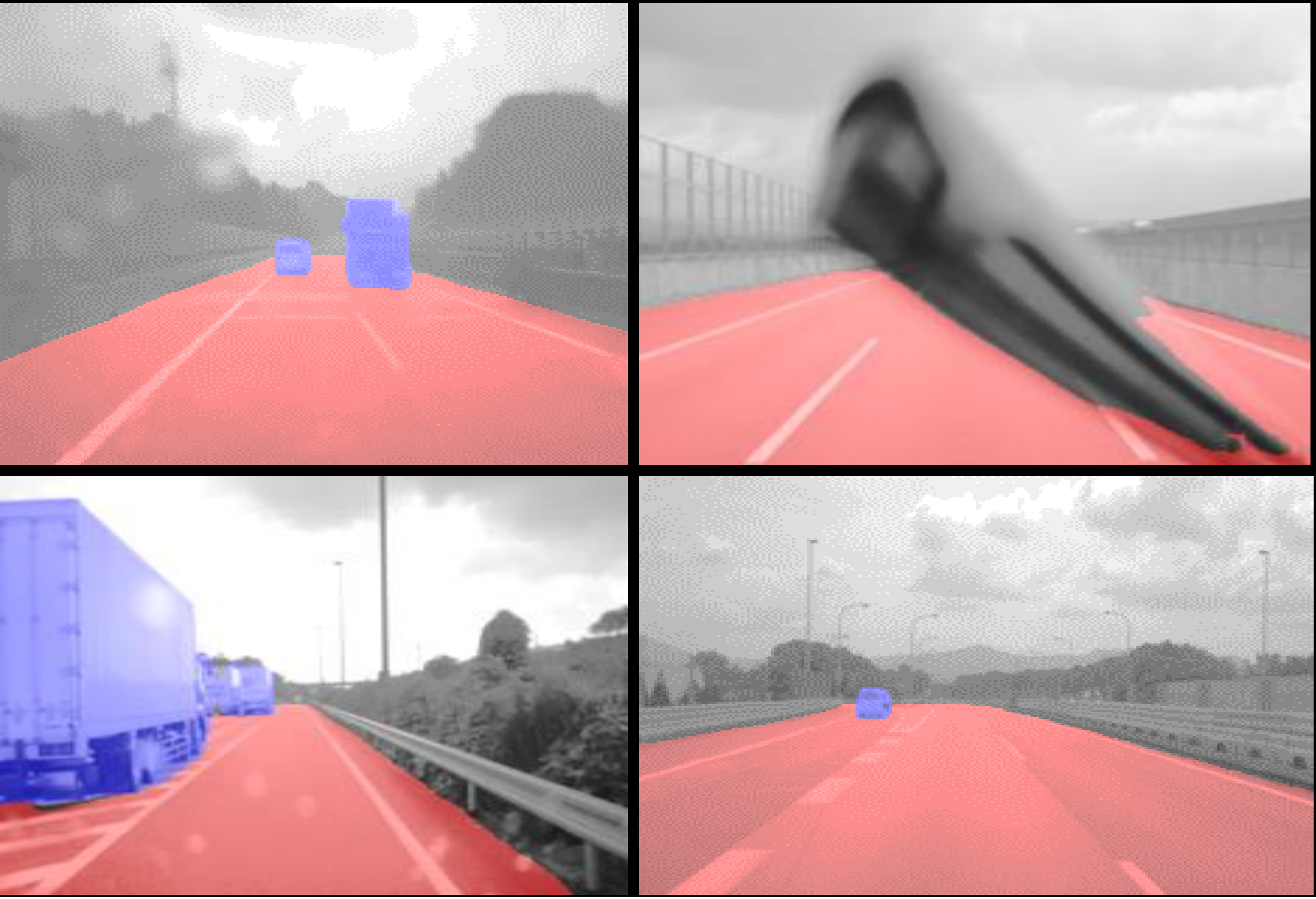
Results for Japan Highway



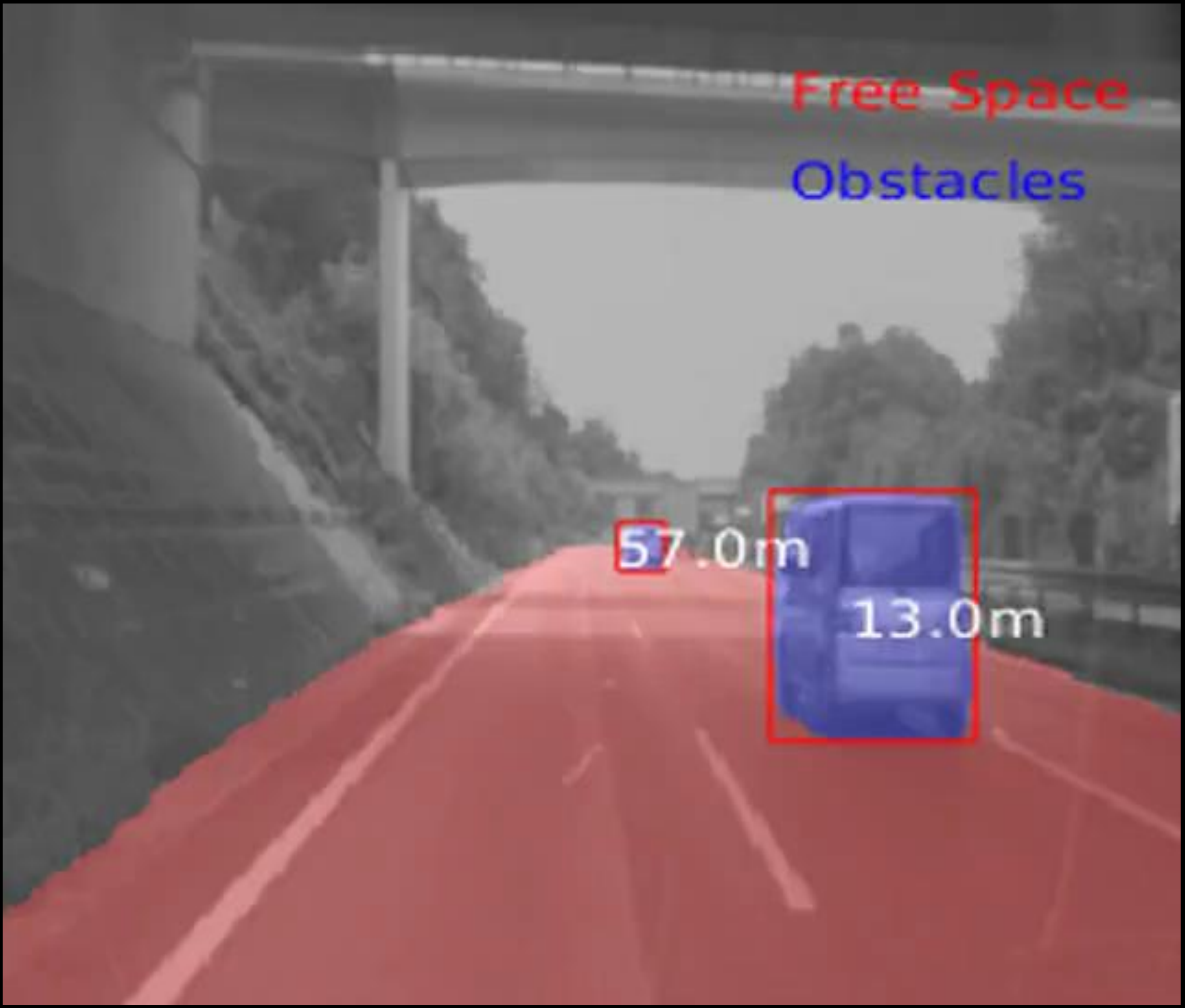
Small Objects Detection



Raining Weather



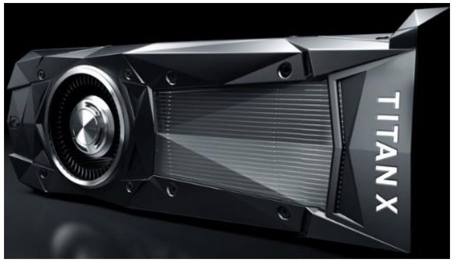
Movie: Tokyo Expressway

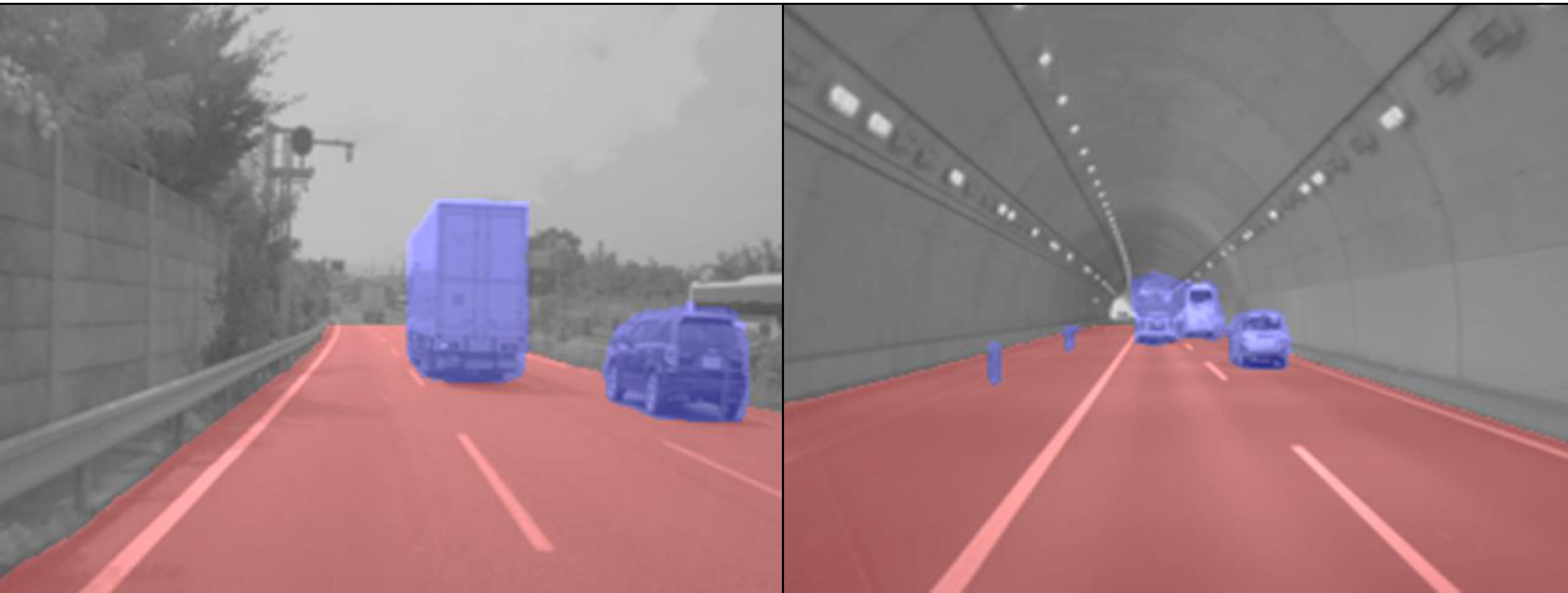


ChiNet vs Baseline

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

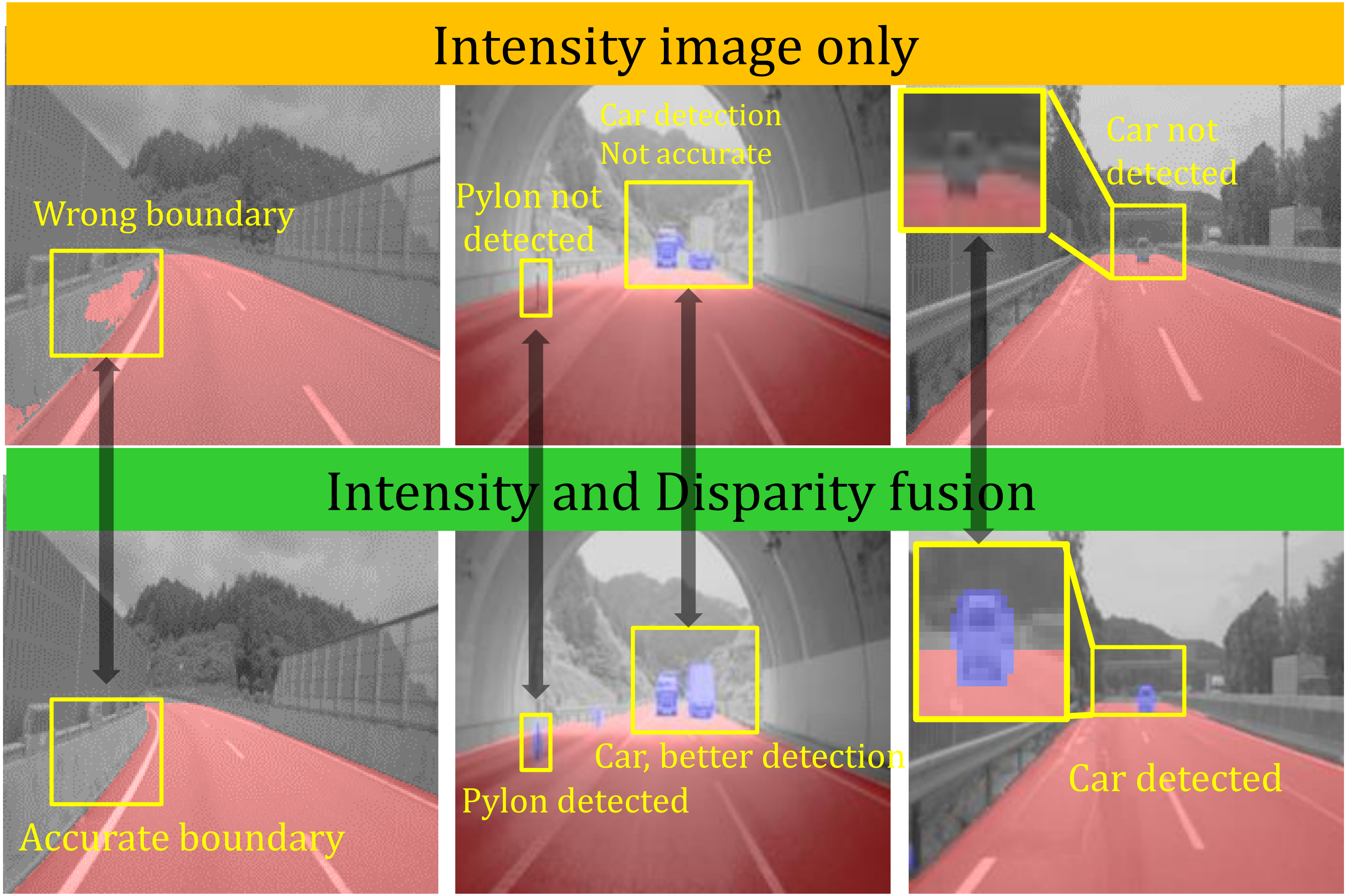
Implemented on **GeForce Titan X** using Keras with Theano backend





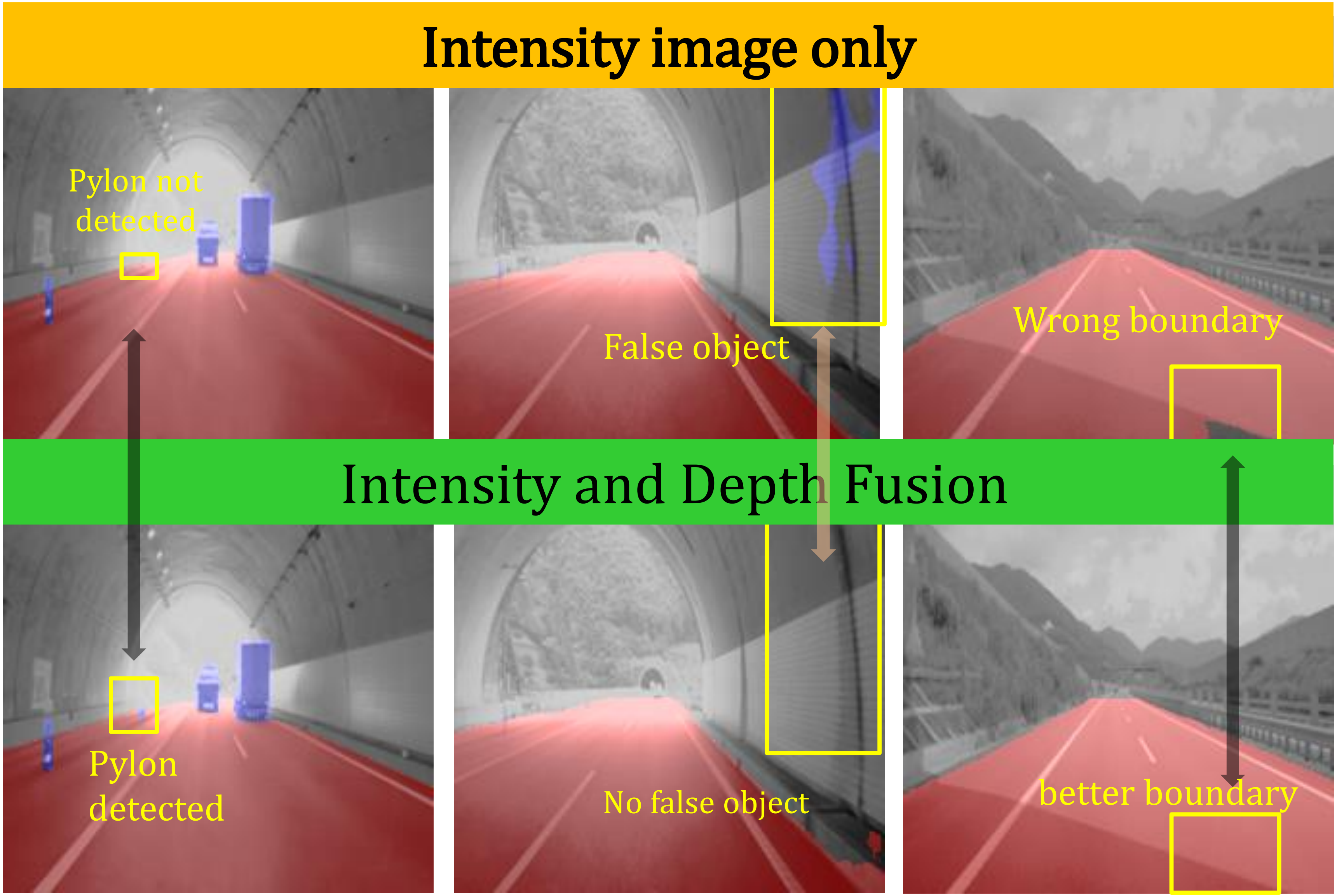
Evaluation Result

Comparison : “Intensity” vs “Intensity and Depth”



Evaluation Result

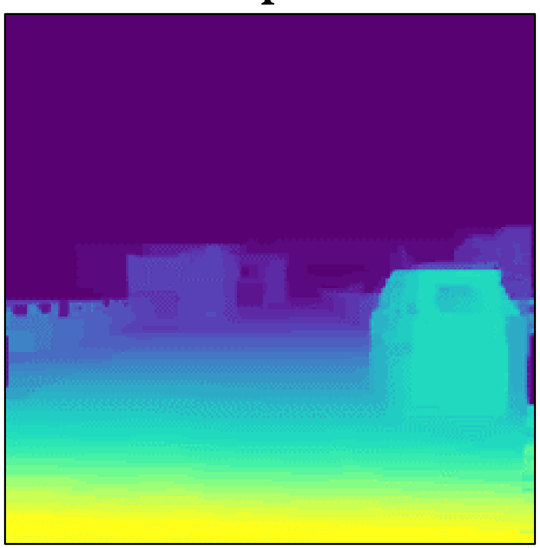
Comparison : "Intensity" vs "Intensity and Dept



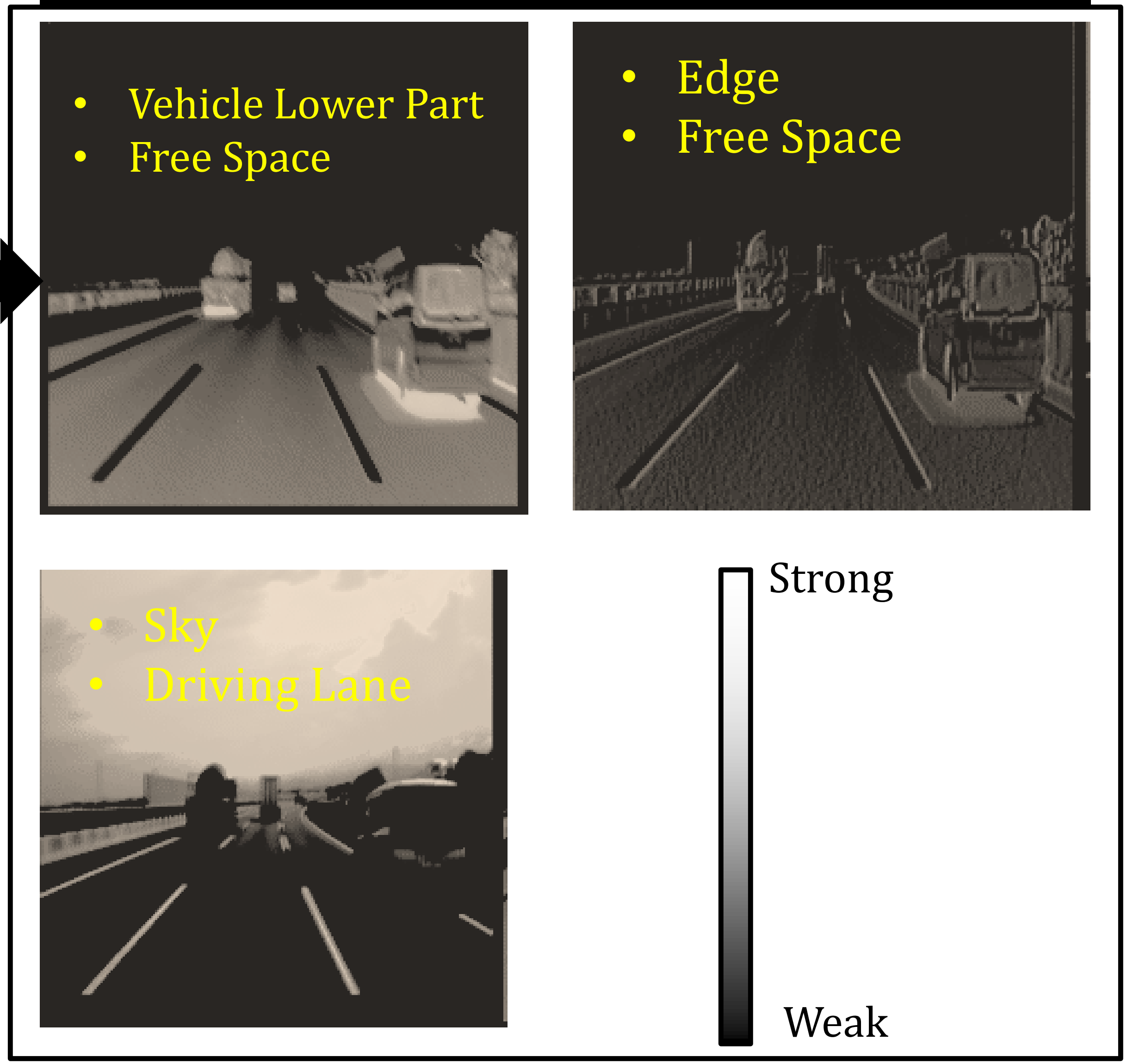
Learned Features by Chi-Net

Intensity Image

Depth



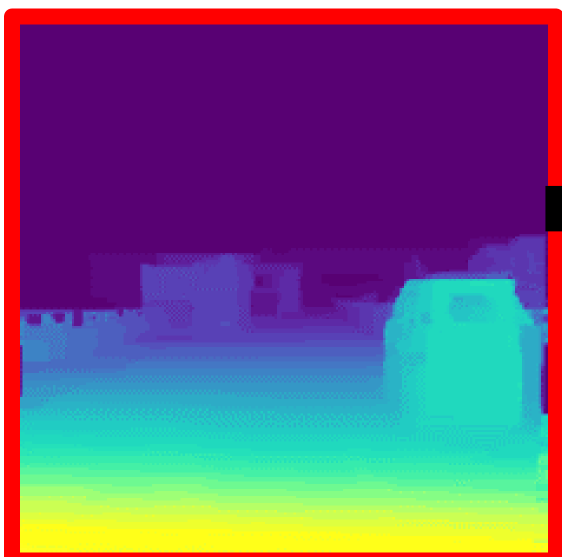
Some of Learned Image Feature



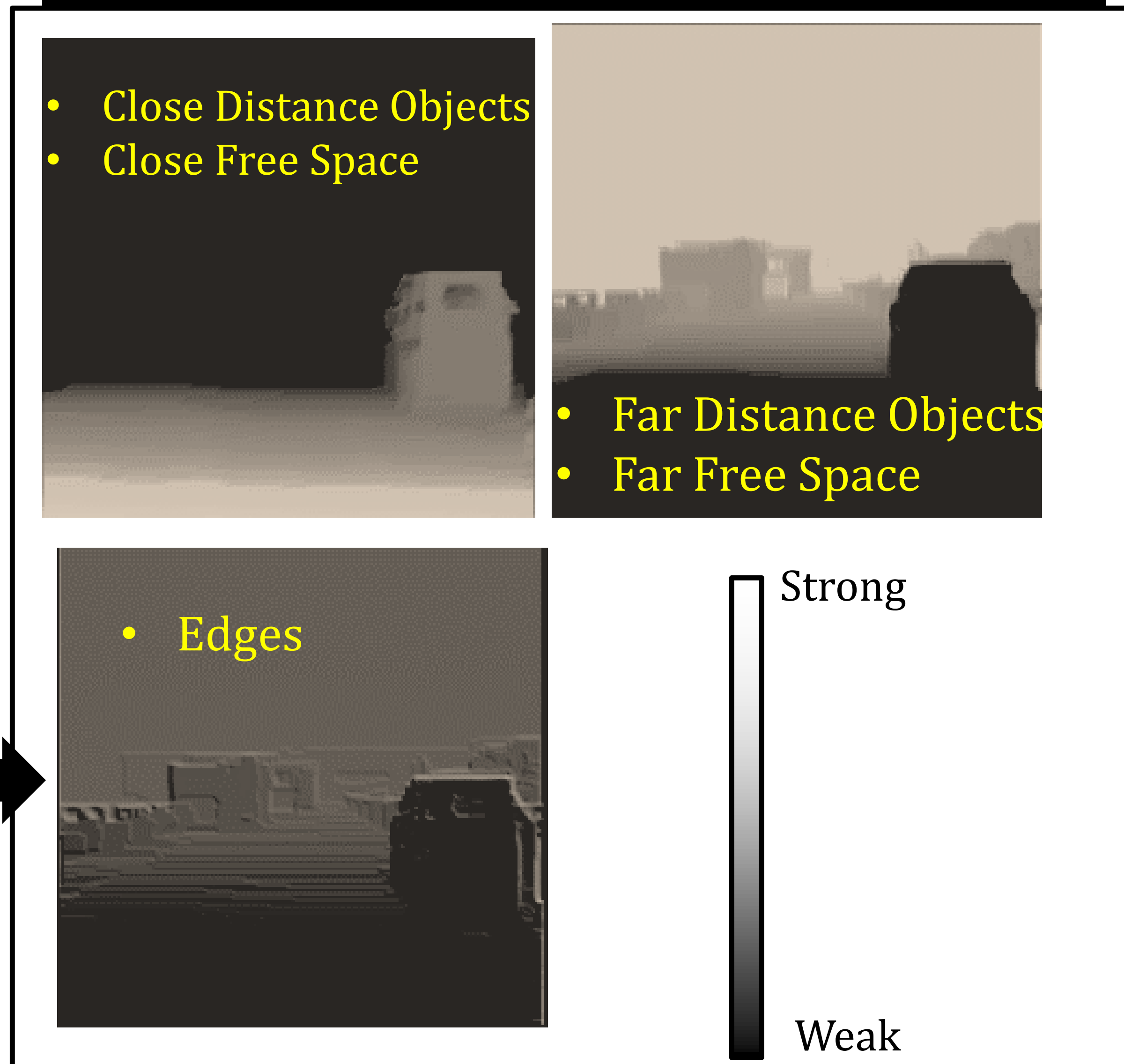
Learned Features by Chi-Net

Intensity Image

Depth



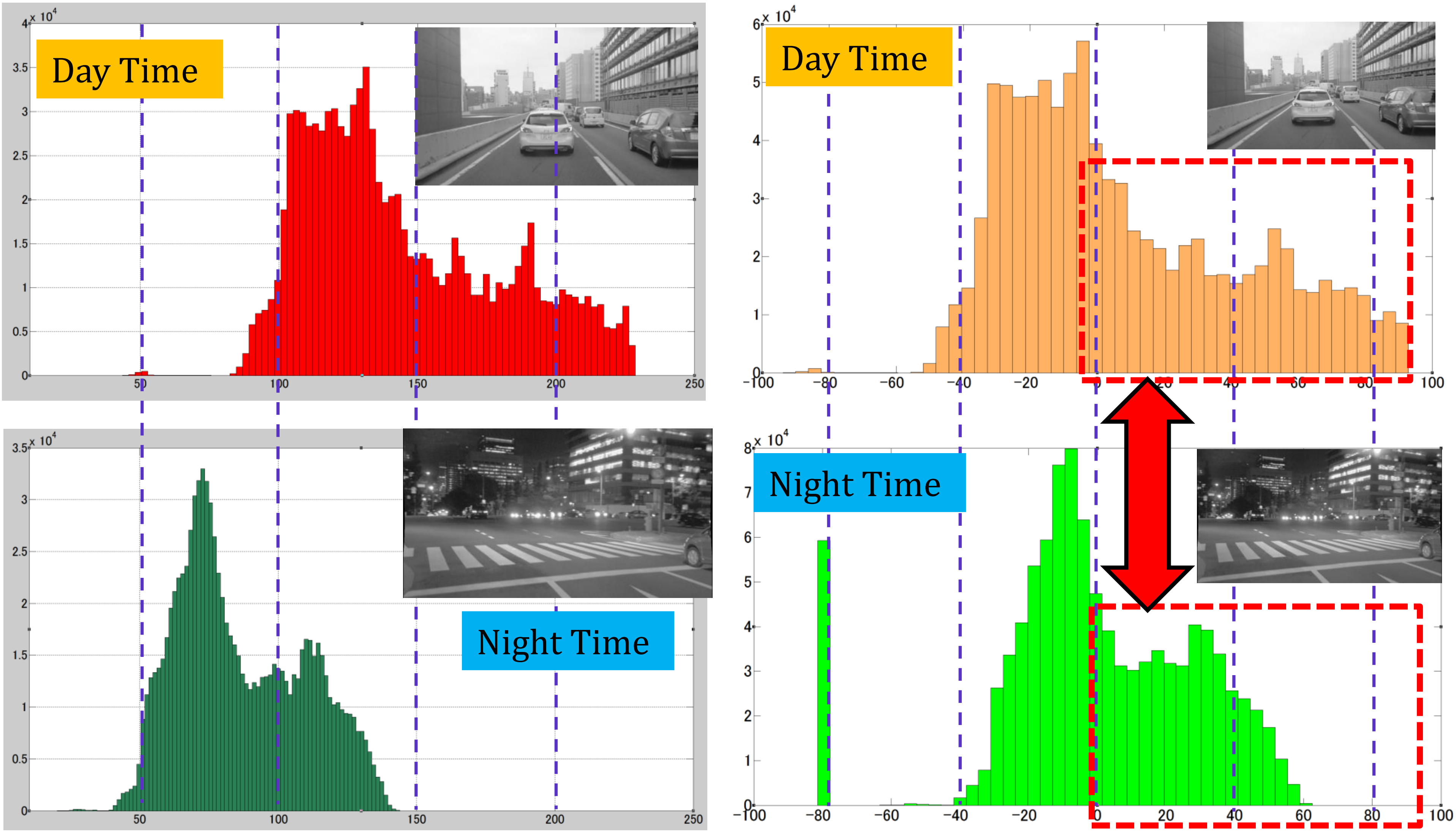
Some of Learned Depth Features



Future Plan

Distribution of Camera Data

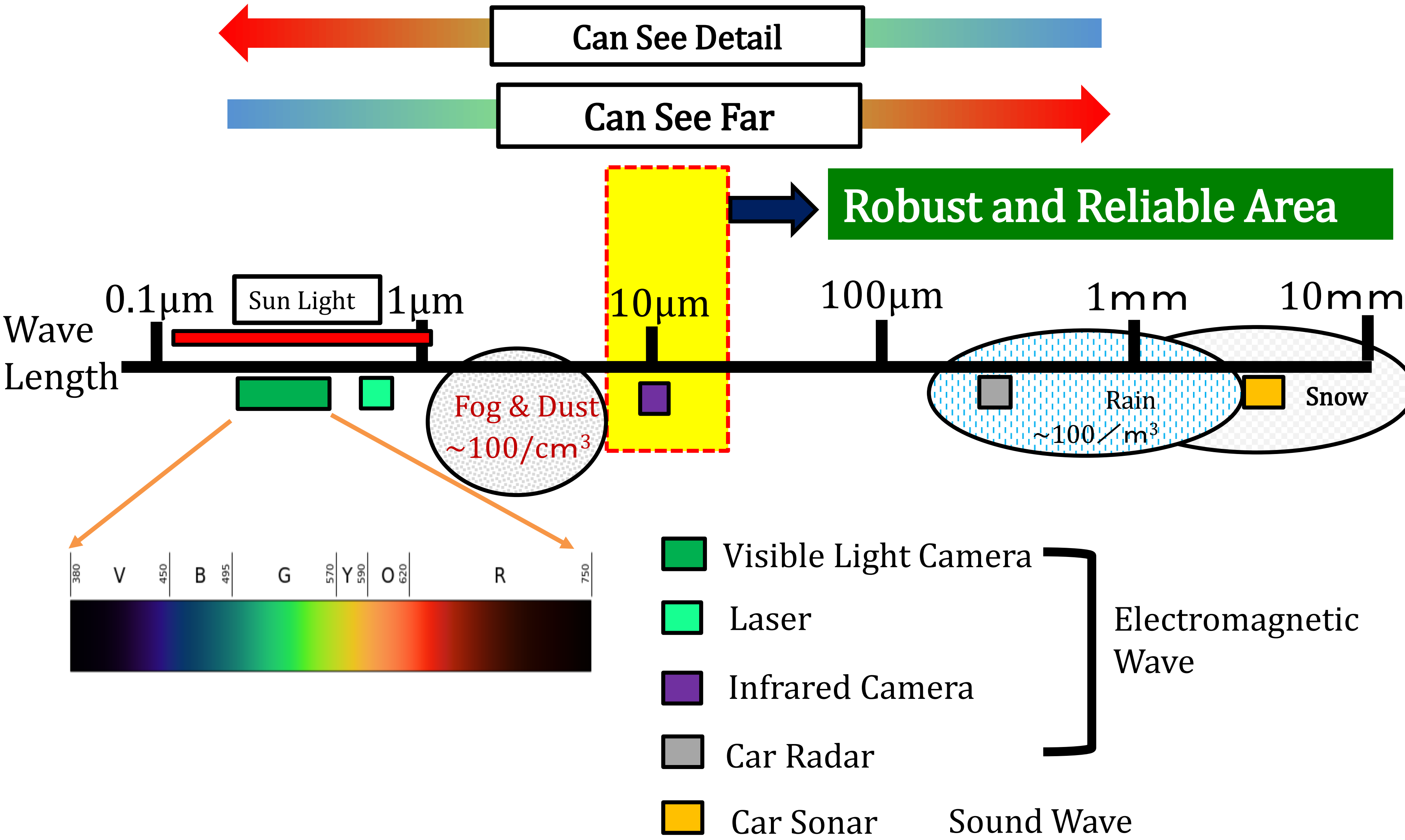
Pixel Data | After Mean Centering



We have different distribution even after mean centering

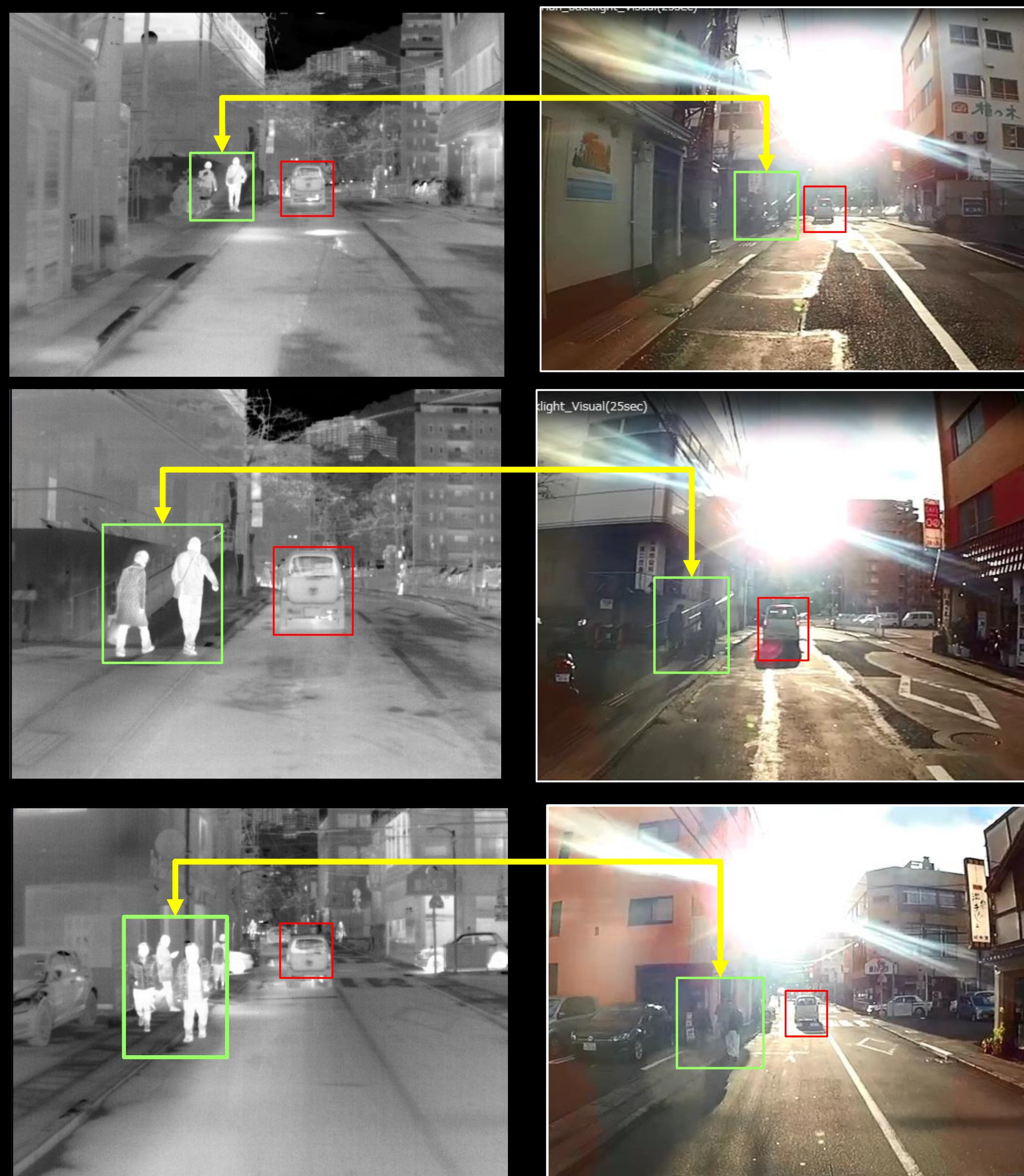
Sensors Properties

Electromagnetic Range Wave: Camera ,Laser, Radar



Challenges of Camera for different Environment

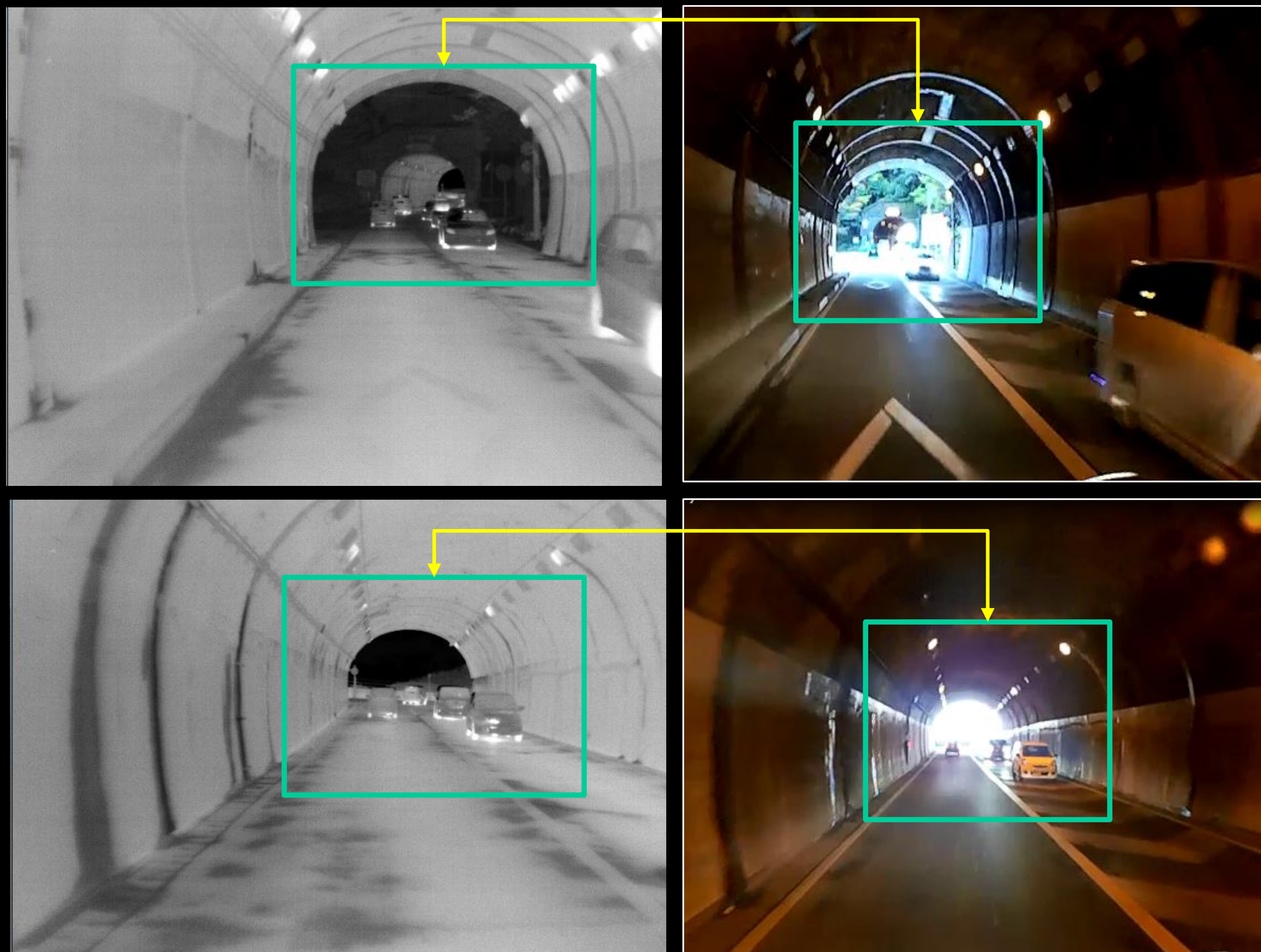
Thermal Camera



Normal Camera

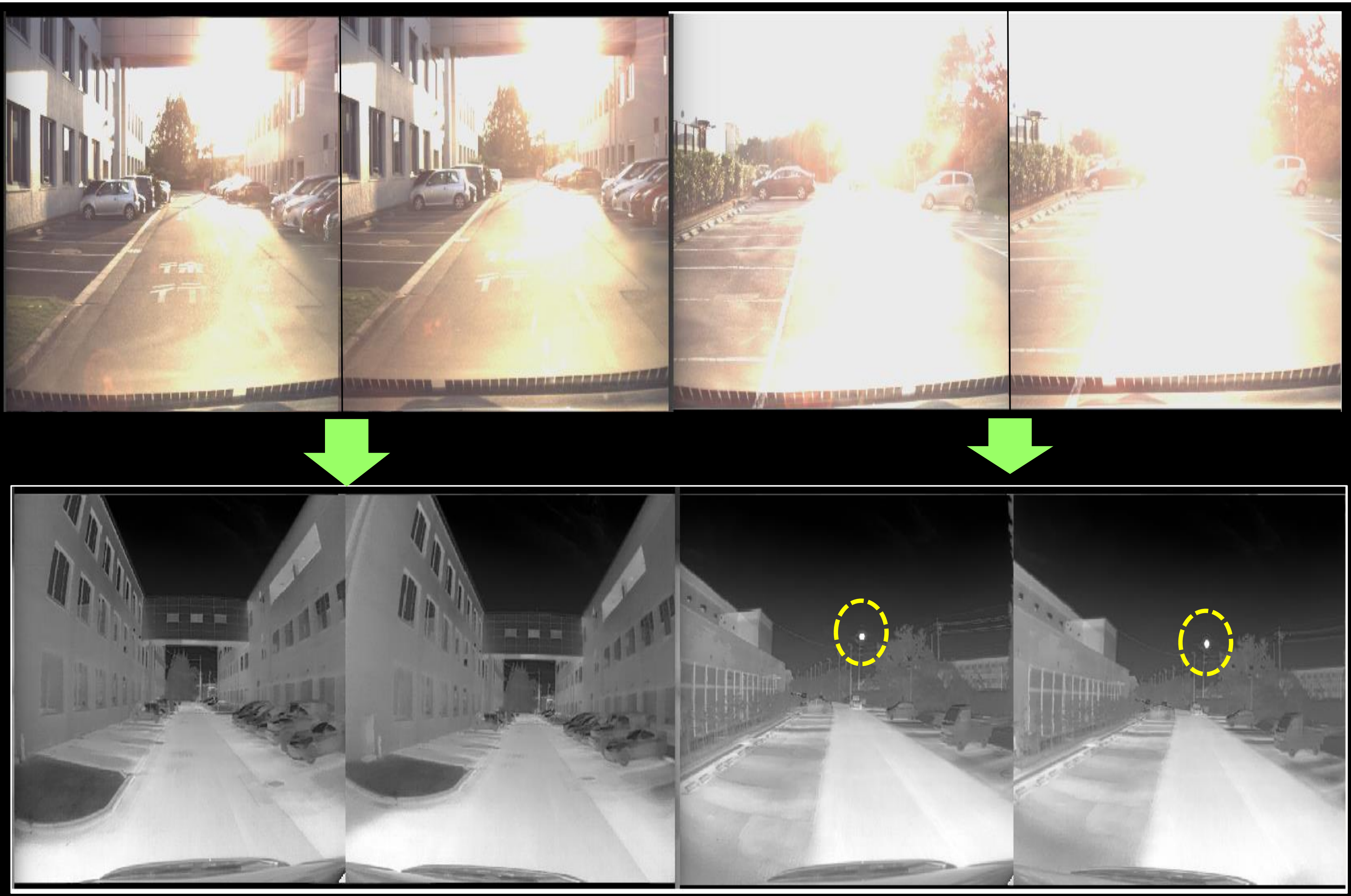
Challenges of Camera for different Environment

Thermal Camera | **Normal Camera**



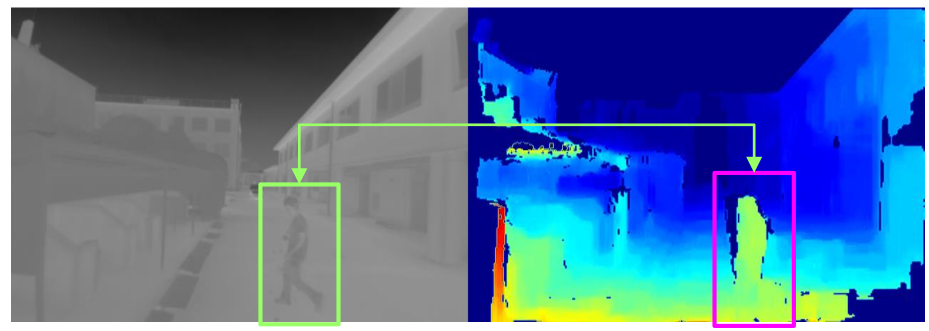
Challenges of Camera for different Environment

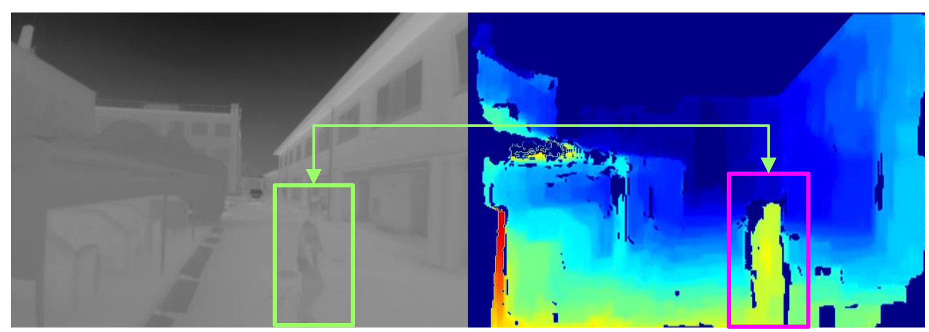
Thermal Camera Images

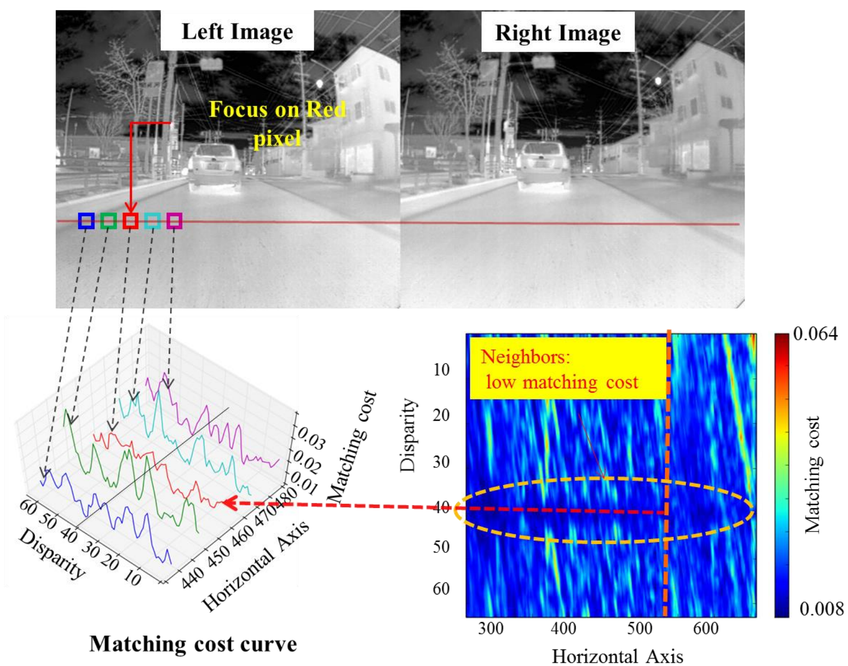


Stability against a variety of light conditions

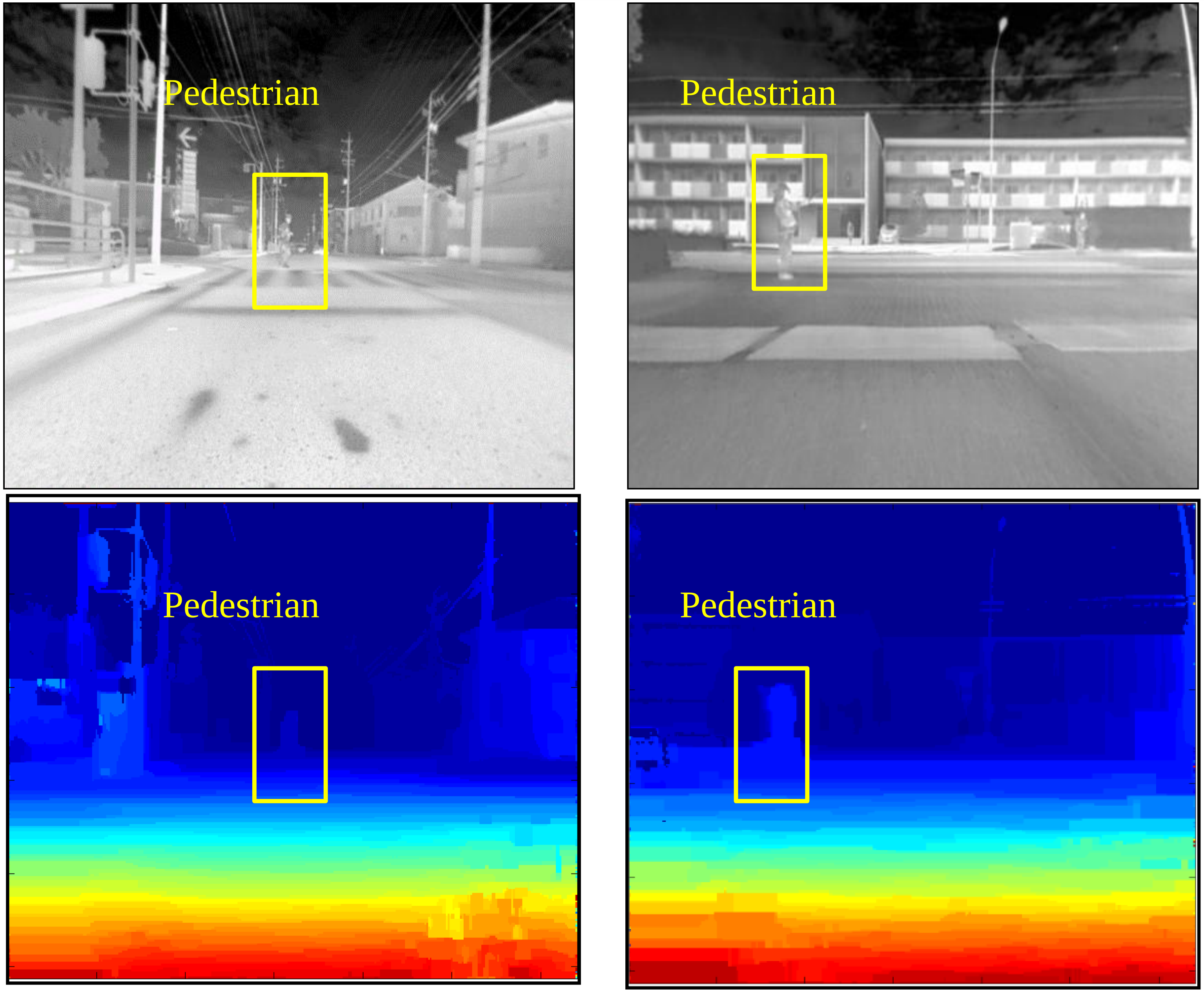
Thermal Stereo



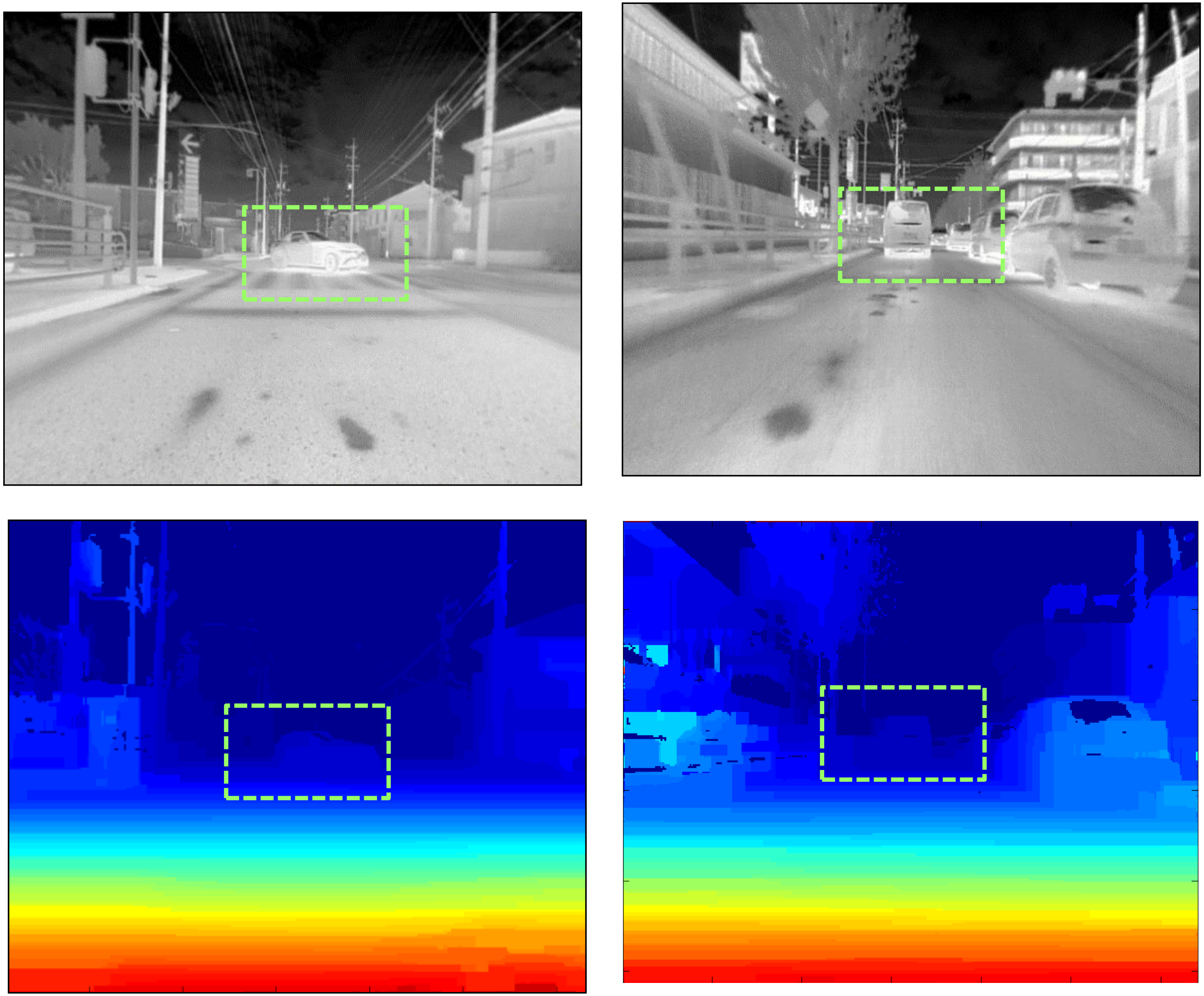




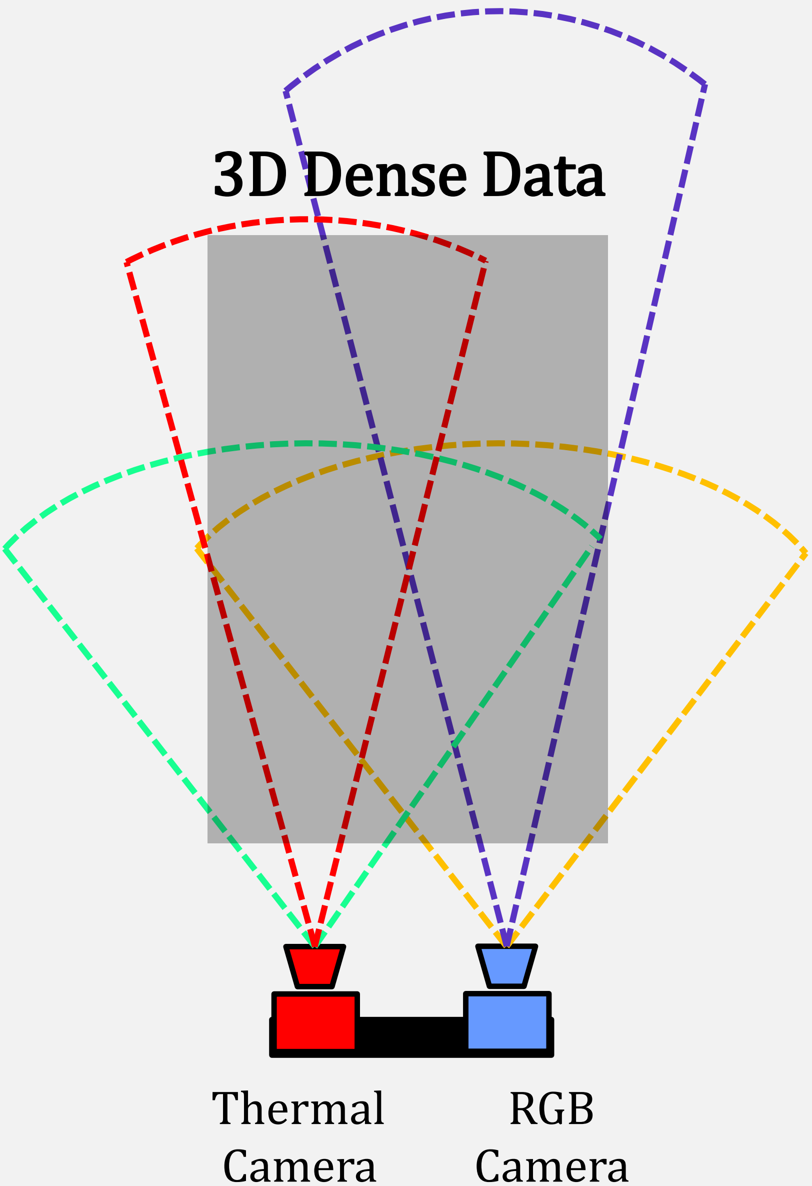
Thermal Stereo Results for Night

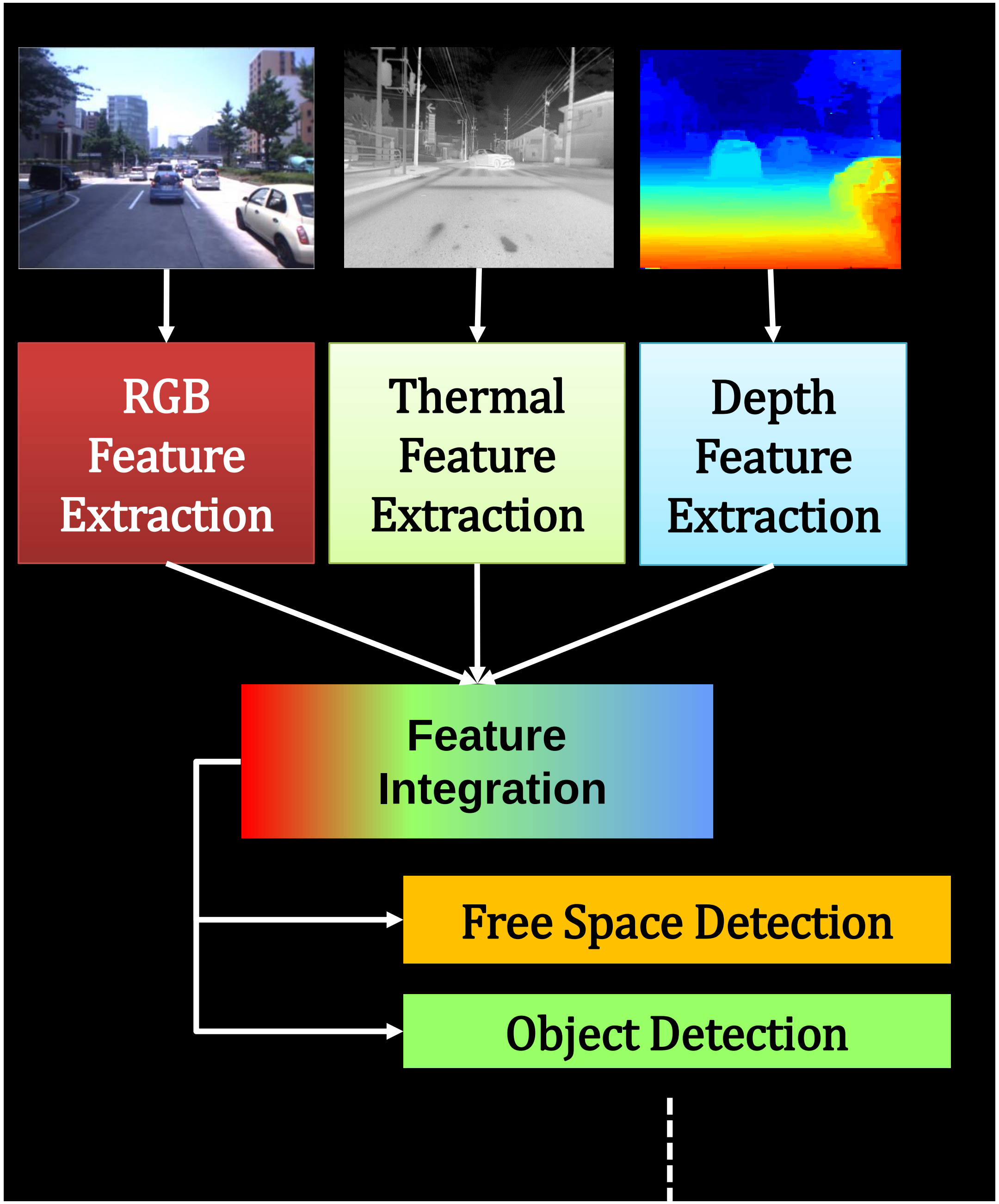


Thermal Stereo Results for Night



Sensor for Future ADAS and Automated Driving

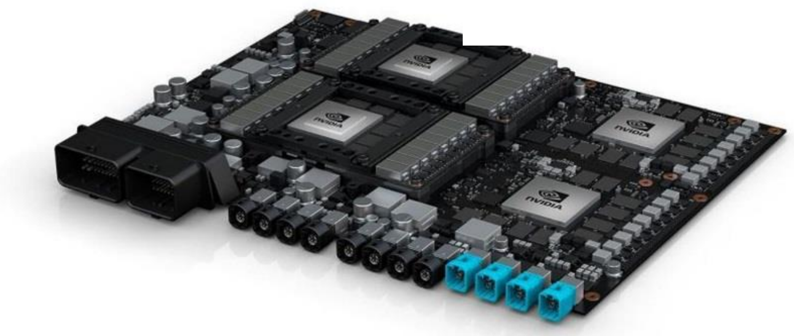




Automated Driving Computer and Sensors

Process and Integration

NVIDIA



320TOPS

Automated Driving Unit

DRIVE PX PEGASUS

Support the High Speed and Processing Requirement for Lev. 5

- Laser , Stereo, Sonar
- Far Infrared Camera, Visible Camera
- Milliwave Radar
- IMU & GNSS & Map

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 ~15 ms possible with optimized CUDA libraries and advances in GPU computing

Question ?