Computer vision, wearable computing and the future of transportation

Amnon Shashua

Hebrew University, Mobileye, OrCam

Computer Vision that will Change Transportation

Amnon Shashua

Mobileye

Computer Vision

- Making “Computer See and Understand What they See”
- Major branch of A.I. goes together with “Machine Learning”.
- Major progress in the last decade. Human level perception is achievable in some narrow domains (face recognition, object detection).
- Camera: lowest cost sensor with highest information density.

Avoiding Collisions: Under the Hood

Technology: Machine Perception & System-on-Chip

- Lane Detection
  - Lane Departure Warning
  - Lane Keeping and Support
- Vehicle Detection
  - Forward Collision Warning
  - Adaptive Cruise Control
  - Traffic Jam Assistant
  - Emergency Braking
- Pedestrian Detection
  - Collision Warning
  - Emergency Braking
- Traffic Sign Recognition
- Intelligent High Beam Control

Autonomous Driving

- Free-space Estimation
- Environmental Model
- Holistic Path Planning
- General Object Detection
- Road Profile Reconstruction
- Traffic Light Detection
- Surround Vision (Hyper-AVM)
- Multi-focal configurations
- 360 awareness

Computer Vision Disruption

2007	2008	2010-11	2013	2015
EyeQ1 180nm 1X	EyeQ1 180nm	EyeQ2 90nm 6X	EyeQ2 90nm	EyeQ3 40nm 48X

2007

First Camera/Radar Fusion

2008

First Bundling of LDW, IHC, TSR

2010-11

First Pedestrian AEB

First Camera Only FCW

2013

First Camera Only ACC and Traffic Jam Assistant

First Camera Only AEB (partial braking)

2015

First Camera Only Full Auto Braking (AEB)

The Camera Disruption

- The functional territory taken by the camera is rapidly increasing:
  - 2011: warning against collisions
  - 2013: ACC, partial brake AEB, TJA
  - 2015: full brake AEB

WHY?

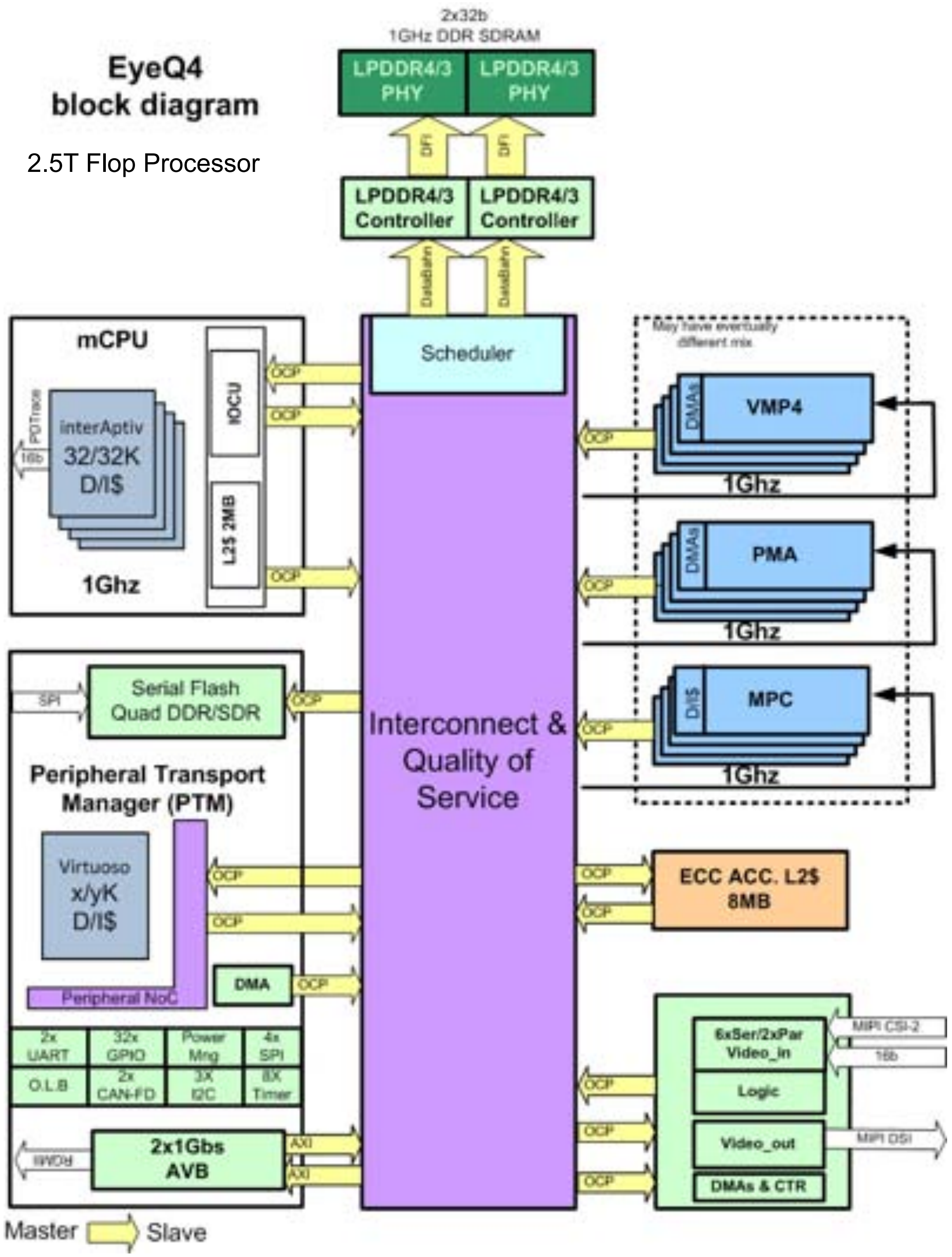
- Richest source of raw data about the scene - only sensor that can reflect the true complexity of the scene.

- The lowest cost sensor - nothing can beat it, not today and not in the future.

- Cameras are getting better - higher dynamic range, higher resolution

Radars/Lidar/Ultrasonic: for redundancy, robustness

EyeQx Vision Application Processor

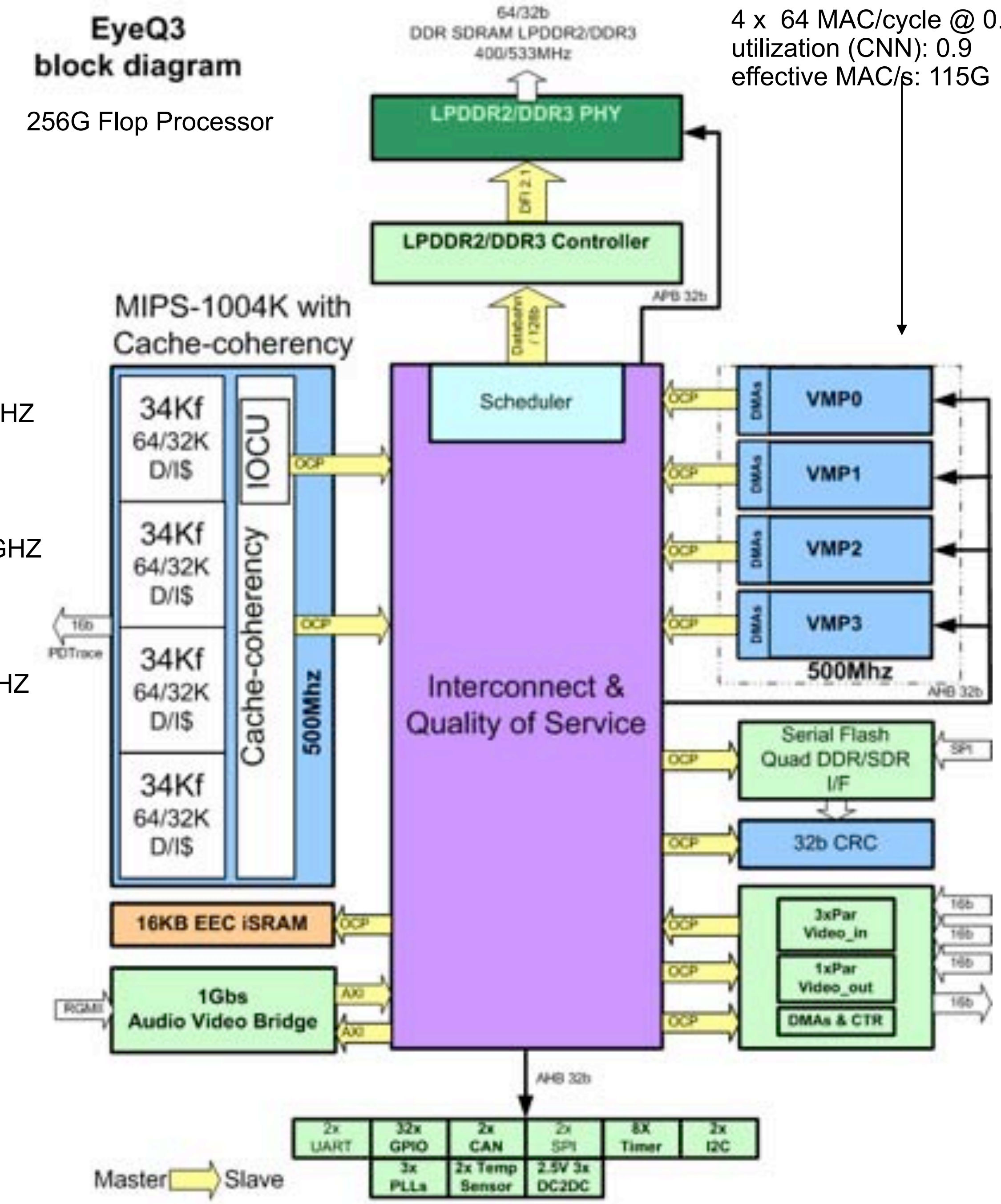


6 x 76 MAC/cycle @ 1GHZ

2 x 384 MAC/cycle @ 1GHZ

2 x 8 MAC/cycle @ 1GHZ

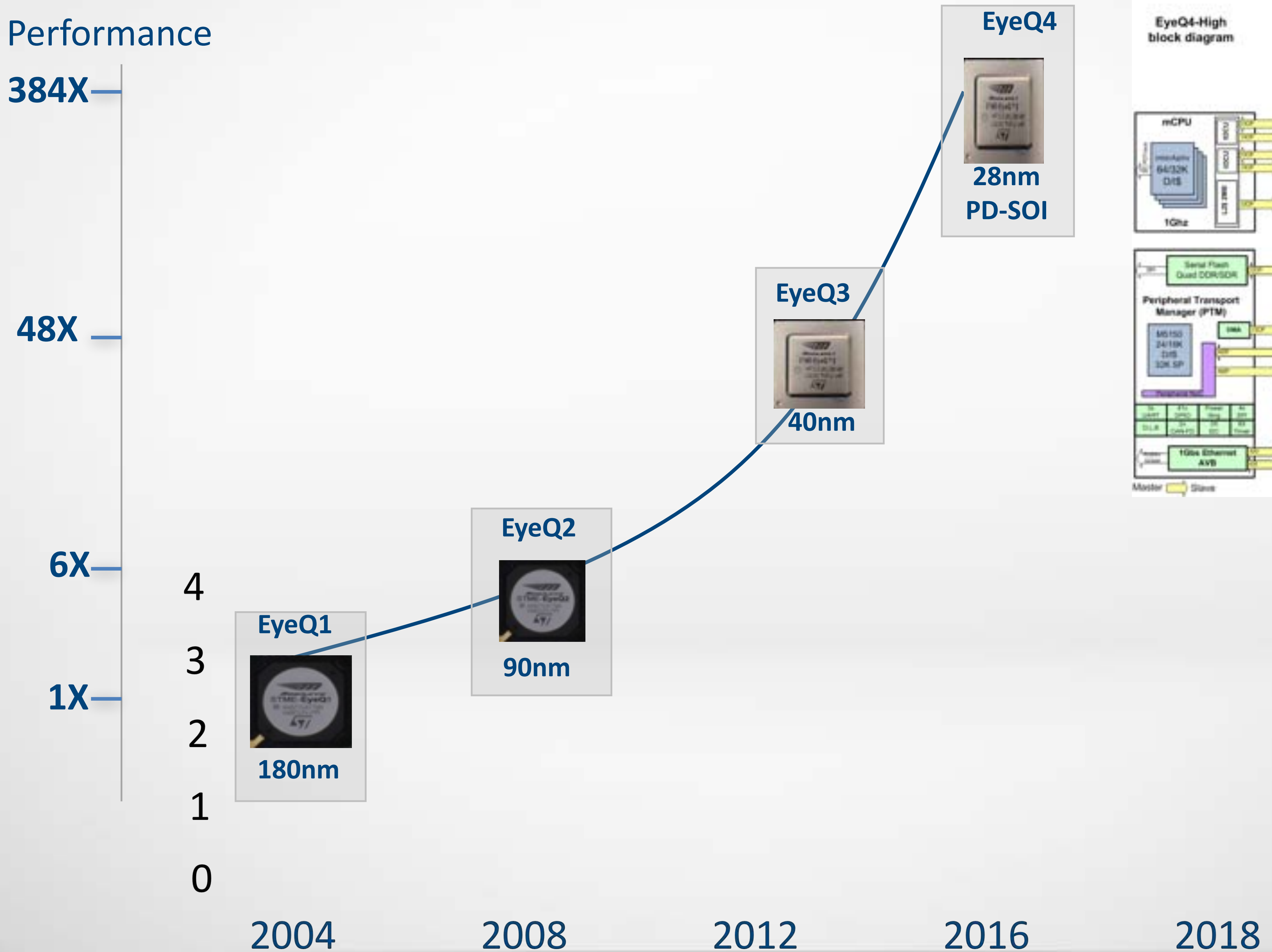
4 x 64 MAC/cycle @ 0.5GHZ
utilization (CNN): 0.9
effective MAC/s: 115G

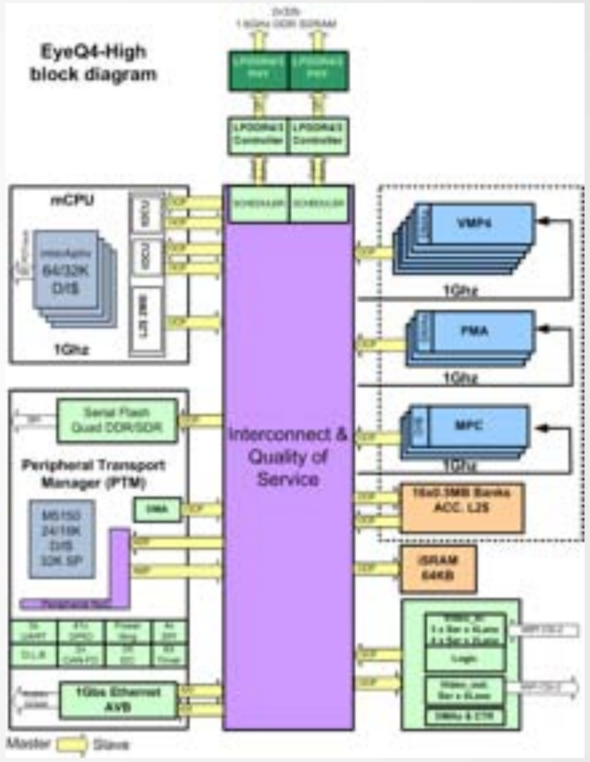


PMA-Program Macro Array
VMP-Vector Microcode Processor
MPC-Multi-Thread Processor Cluster

EyeQ System on Chip Roadmap

Performance of EyeQ Chip has Increased Rapidly over Time





Market Drivers

Two Major Trends

Evolution

Revolution

New Safety Rating Regulations

Autonomous Driving Megatrend

TEST RESULTS

EURO NCAP

Nissan Qashqai

Nissan Qashqai 1.5dCi Acenta, LHD

2014 ★★★★★

88% ADULT OCCUPANT	83% CHILD OCCUPANT
69% PEDESTRIAN	79% SAFETY ASSIST

DETAILS OF TESTED CAR

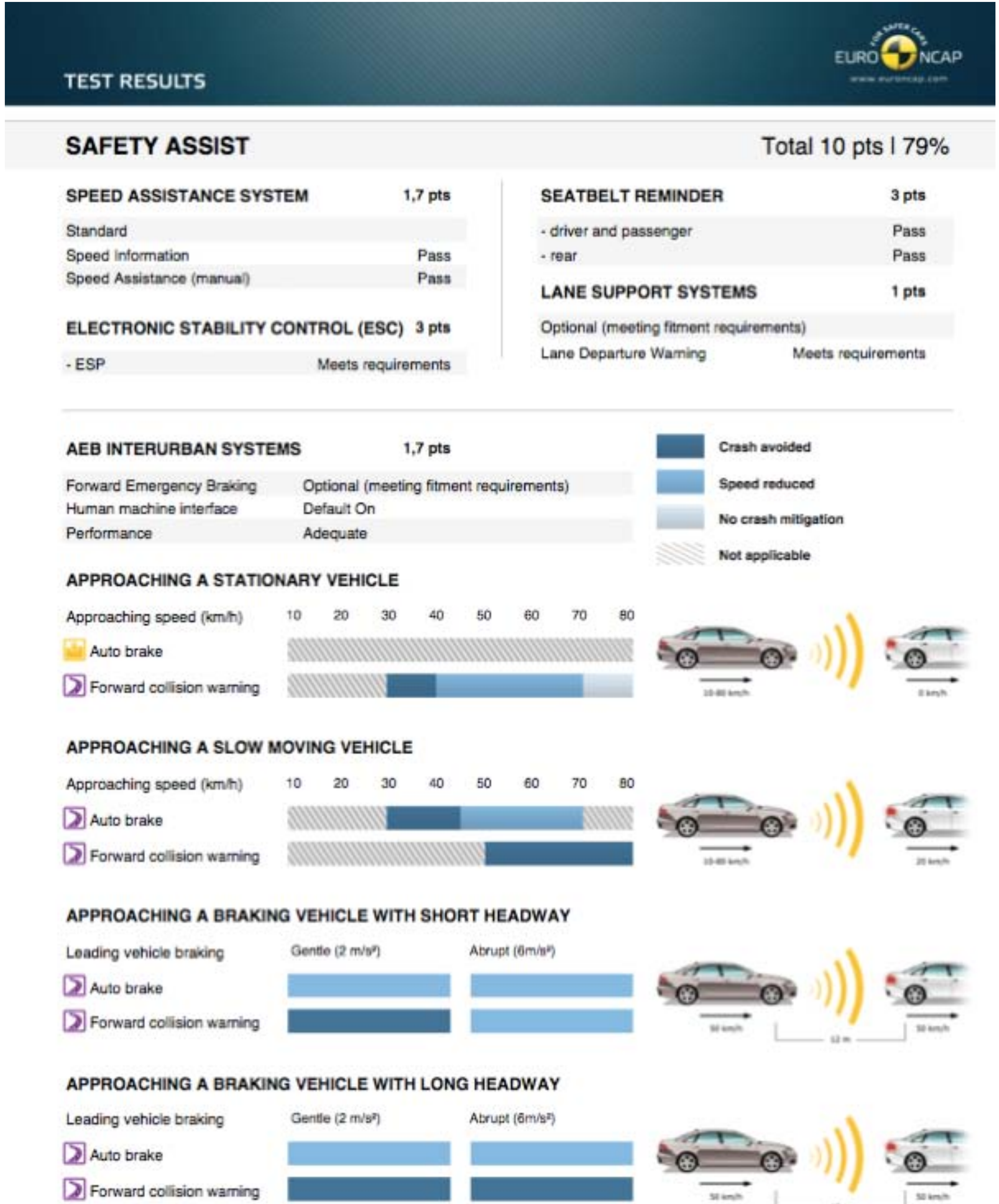
SPECIFICATIONS

Tested model	Nissan Qashqai 1.5dCi Acenta, LHD
Body type	5 door hatchback
Year of publication	2014
Kerb weight	1388kg
VIN from which rating applies	applies to all Nissan Qashqais of the specification tested

SAFETY EQUIPMENT

Frontal airbags	Driver (single stage), Passenger (single stage)
Pre-tensioners	Driver (dual), Passenger (single)
Load-limiters	Driver, Passenger
Knee airbags	None
Side airbags	Head (front and rear), Thorax
Front head restraints	Passive
Passenger airbag switch	Manual switch
ISOFIX anchorages	Rear outboard seats
Integrated child restraint	None
Active Pedestrian Protection	None,
Seatbelt Reminder	Driver, Passenger, Rear
Electronic Stability Control	ESP, Standard, Manual Switch
Speed Assistance Systems	Driver-set speed limitation, Standard
Lane Support	Lane Departure Warning, Optional (meeting fitment requirements)
Autonomous Braking	Forward Emergency Braking, City and Inter-Urban (auto-brake and forward collision warning) system, Optional (meeting fitment requirements)
Other	Not applicable

Safety equipment is standard across the model range unless stated otherwise


TEST RESULTS

EURO NCAP

SAFETY ASSIST

Total 10 pts | 79%

SPEED ASSISTANCE SYSTEM	1,7 pts
Standard	
Speed Information	Pass
Speed Assistance (manual)	Pass

ELECTRONIC STABILITY CONTROL (ESC)	3 pts
- ESP	Meets requirements

SEATBELT REMINDER	3 pts
- driver and passenger	Pass
- rear	Pass

LANE SUPPORT SYSTEMS	1 pts
Optional (meeting fitment requirements)	
Lane Departure Warning	Meets requirements

AEB INTERURBAN SYSTEMS	1,7 pts
Forward Emergency Braking	Optional (meeting fitment requirements)
Human machine interface	Default On
Performance	Adequate

Our Vision. Your Safety.™

Mobileye in Numbers

2007-2012: 1,000,000 EyeQ
2013: 1,300,000 EyeQ
2014: 2,700,000 EyeQ
H1 2015: ~2.5M EyeQ

2010: 36 car models, 7 auto-makers
2014: 160 car models, 18 auto-makers
2016: 240 car models, 25 auto-makers

Increasing Awareness: Hyundai Super Bowl 2014 Commercial

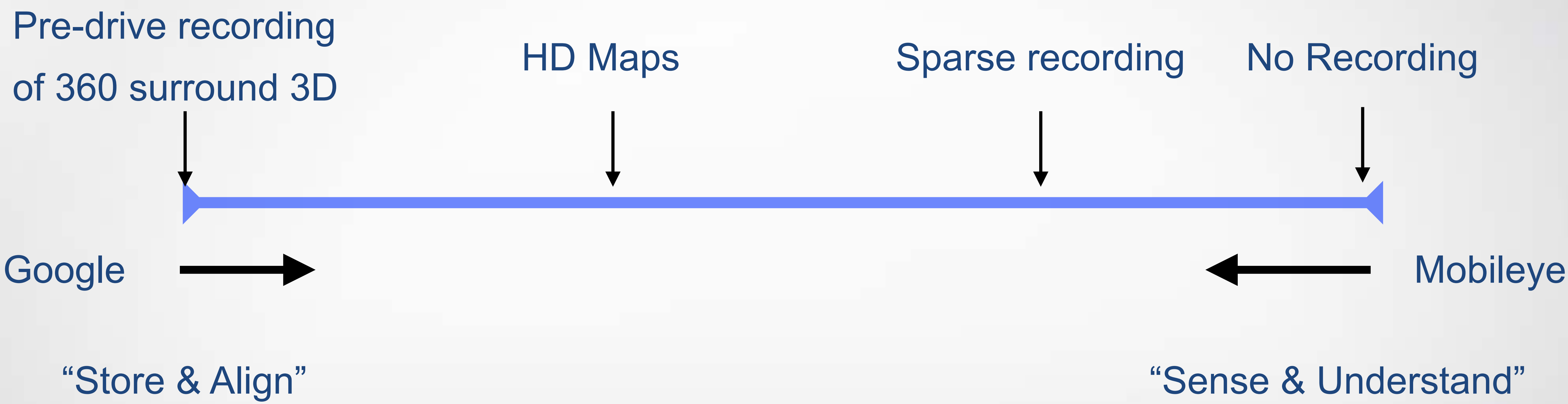
Still running on Times Square (as of 5/2015)

Autonomous Emergency Braking

Volvo S60 - launched 5/2010 - tests by Polish "warriors"

ADAS 2016-2020

Two Paradigms for Achieving Autonomous Driving



Leap I: Human Level Perception

Human Level Perception is possible - already achieved in narrow domains.

- ADAS -> HLP requires:
  - Extending list of objects (cars at all angles, general objects, ~1000 traffic signs, traffic lights,…)
  - Using context to predict path (“holistic path planning”)
  - Detailed Road interpretation: free-space, curbs, barriers, guard rails, construction, highway exits,…
  - Deep Layered Networks is the tool required for the leap.

The Need for Context (the rise of Deep Layered Networks)

- Path planning: fuse all the information available from the image, not only lane marks.

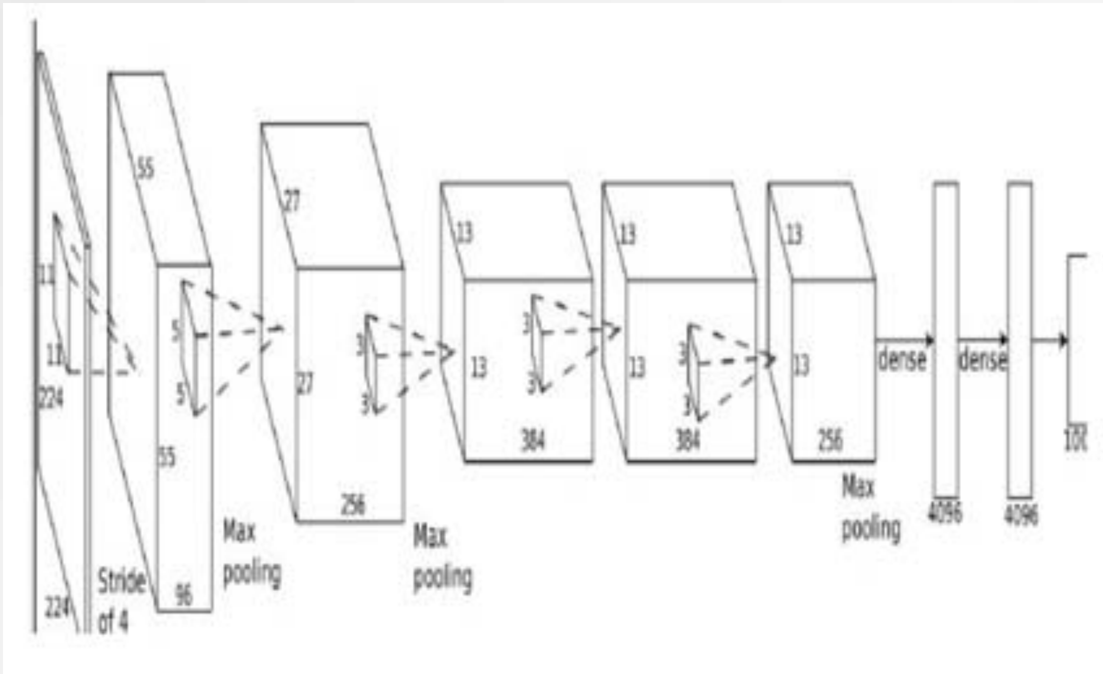
- Environmental Model: ultimately a category label for every pixel in the image

- “3D” Model for Vehicles (VD at any angle, Viewed from any angle).

- “Scene Recognition”: Stop-line, Bumps, Road Surface…

Deep Networks

Convolutional Neural Network



Krizhevsky, A., Sutskever, I., & Hinton, G. (2012)

60M parameters

832M MAC ops

Visualizing and Understanding Convolutional Networks

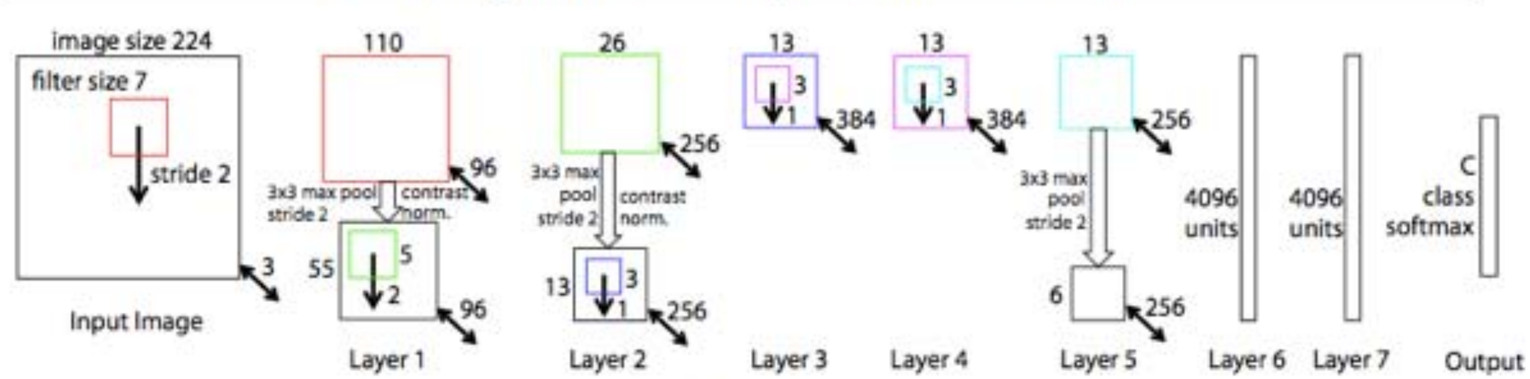


Figure 3. Architecture of our 8 layer convnet model. A 224 by 224 crop of an image (with 3 color planes) is presented as the input. This is convolved with 96 different 1st layer filters (red), each of size 7 by 7, using a stride of 2 in both x and y. The resulting feature maps are then: (i) passed through a rectified linear function (not shown), (ii) pooled (max within 3x3 regions, using stride 2) and (iii) contrast normalized across feature maps to give 96 different 55 by 55 element feature maps. Similar operations are repeated in layers 2,3,4,5. The last two layers are fully connected, taking features from the top convolutional layer as input in vector form ($6 \cdot 6 \cdot 256 = 9216$ dimensions). The final layer is a C-way softmax function, C being the number of classes. All filters and feature maps are square in shape.

Zeiler & Fergus, 2013

Breakthroughs in object recognition

Image credit Deng09

Imagenet: 1000 classes, 1.2M images

Top 5 err. 2011	Top 5 err, 2012
25.8% .	1.Krizhevsky-et-al 16.4% 2. ISI 26.2%

Deep Nets in computer vision: Follow-up

Imagenet object recognition competitions

Number of Deep Net approaches / Total	Top 5 error (%)	Winning Team, Year
1 / 6	***16.4***	Supervision Kizhevsky et al, 2012
17 / 24	***11.7***	Clarifai Zeiler & Fergus, 2013
31 / 32	***6.66***	GoogLe Net Szegedy et al, 2014

Recent results:

Top 5 err. (%)	Team/Company
6.8	VGG,Simonyan'14
5.98	Baidu, Wu'15
4.94	Microsoft, He'15
4.82	Google, Ioffe'15

Human: 5.1% (estim)

Note: Error was 25.8% in 2011!

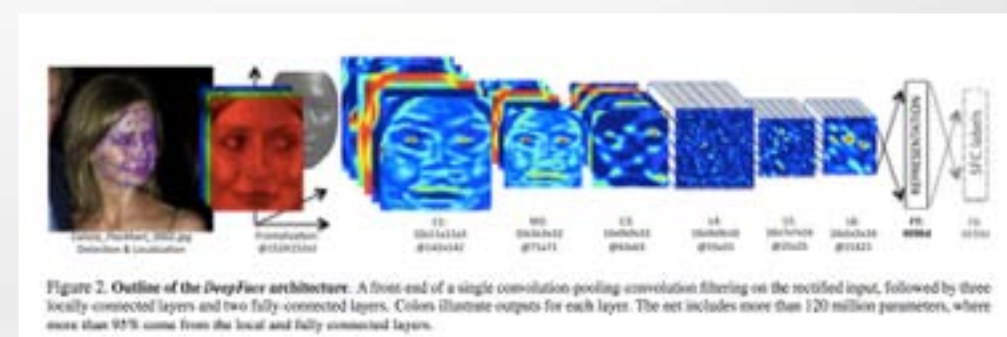
All subsequent years : DNN solutions
Wide adoption in industry: Google, Microsoft, Baidu, Apple, Nuance, Mobileye, etc integrate deep network solutions

‘Human Level’ Face Recognition

Labeled Faces in the Wild
LFW benchmark

99.70% 99.62%	NUS-LV* Baidu*
99.50%	Face++, Megvii
99.47%	DeepId2+, CUHK
97.35%	DeepFace, Facebook, 2014
91.37%	LBP/SVM

Deep
Nets



Human performance ~97.5%

*Newest results, no publications yet

Breakthrough's in Speech and NLP

Deep Speech: Scaling up end-to-end speech recognition

Awni Hannun[*], Carl Case, Jared Casper, Bryan Catanzaro, Greg Diamos, Erich Elsen, Ryan Prenger, Sanjeev Satheesh, Shubho Sengupta, Adam Coates, Andrew Y. Ng

Baidu Research – Silicon Valley AI Lab

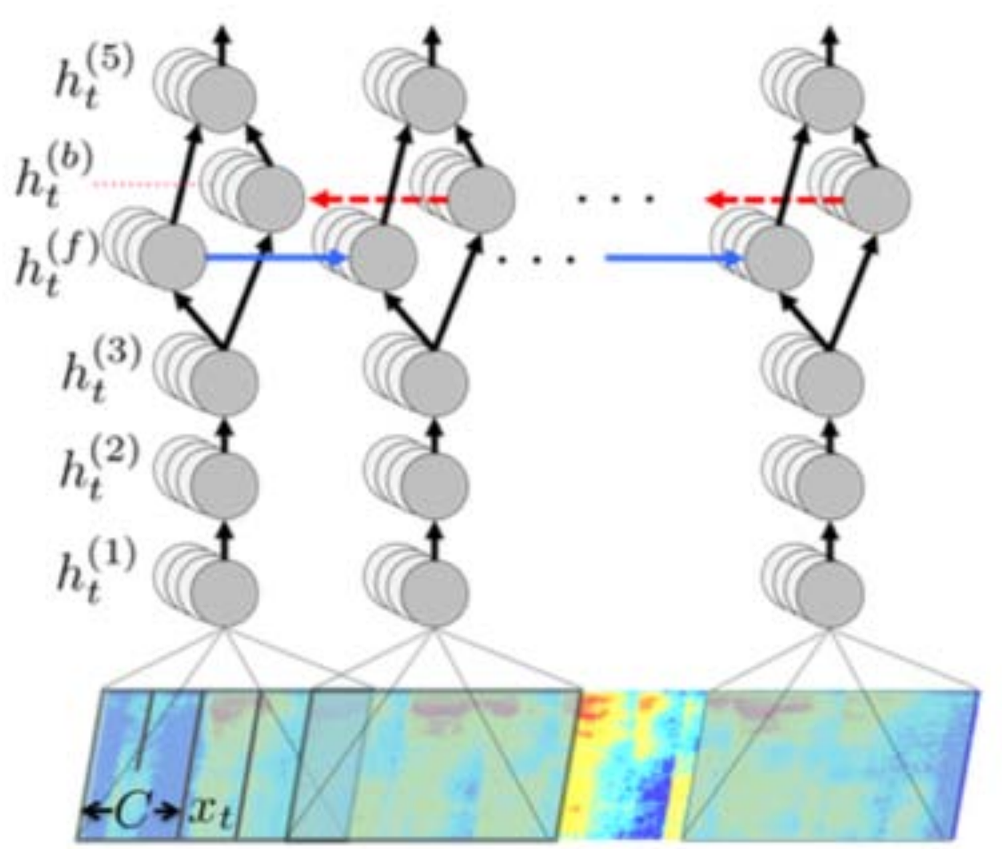


Figure 1: Structure of our RNN model and notation.

Potential Impact of DNN for Automotive

- Networks are at their best for multi-class problems - enables a rich vocabulary of objects (vehicles, pedestrians, types-of, traffic signs, etc.)
- Networks are very good at using "context" - holistic perception. Case in point: Path Planning.
- Network design is ideal for "pixel-level" labeling - objects that do not fit into a bounding-box. Examples, barriers, curbs, guard-rails,… Case in point: Semantic Free-space.
- Networks can be used for Sensor Integration and Control decisions. The classical "control point" can be determined using a holistic process.

Challenges for using DNN for Automotive

- Networks are very large ~1.5B parameters
- Require huge training sets
- Not real-time driven
- Success for “easy” problems: Object detection. Academic research on higher-level perception (like pixel-level labeling) are sketchy.

DNNs at Mobileye

The Need for Context (the rise of Deep Layered Networks)

- Path planning: fuse all the information available from the image, not only lane marks.

- Environmental Model: ultimately a category label for every pixel in the image

- “3D” Model for Vehicles (VD at any angle, Viewed from any angle).

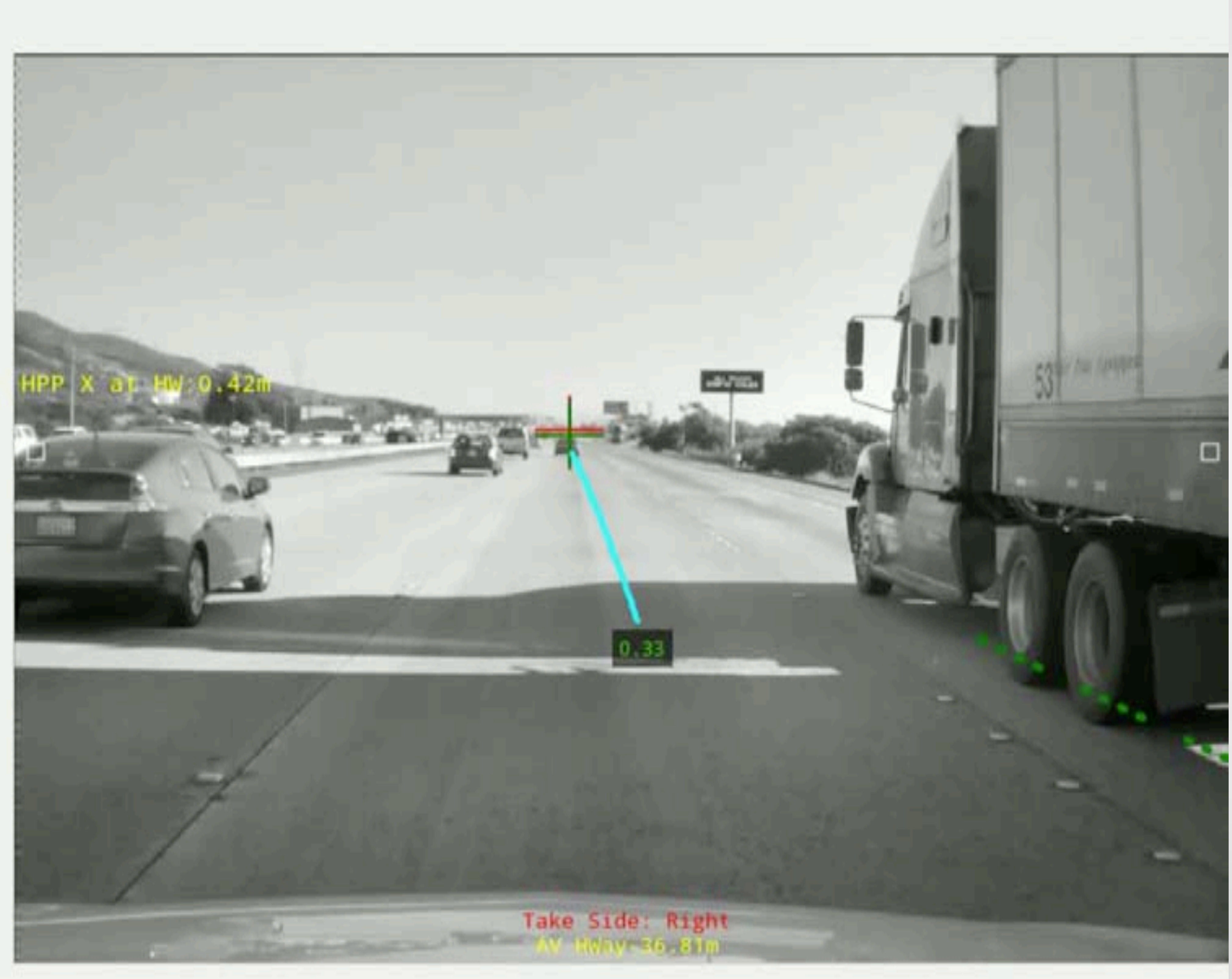
- “Scene Recognition”: Stop-line, Bumps, Road Surface…

Holistic Path Planning (HPP)

Path Planning using Holistic Cues

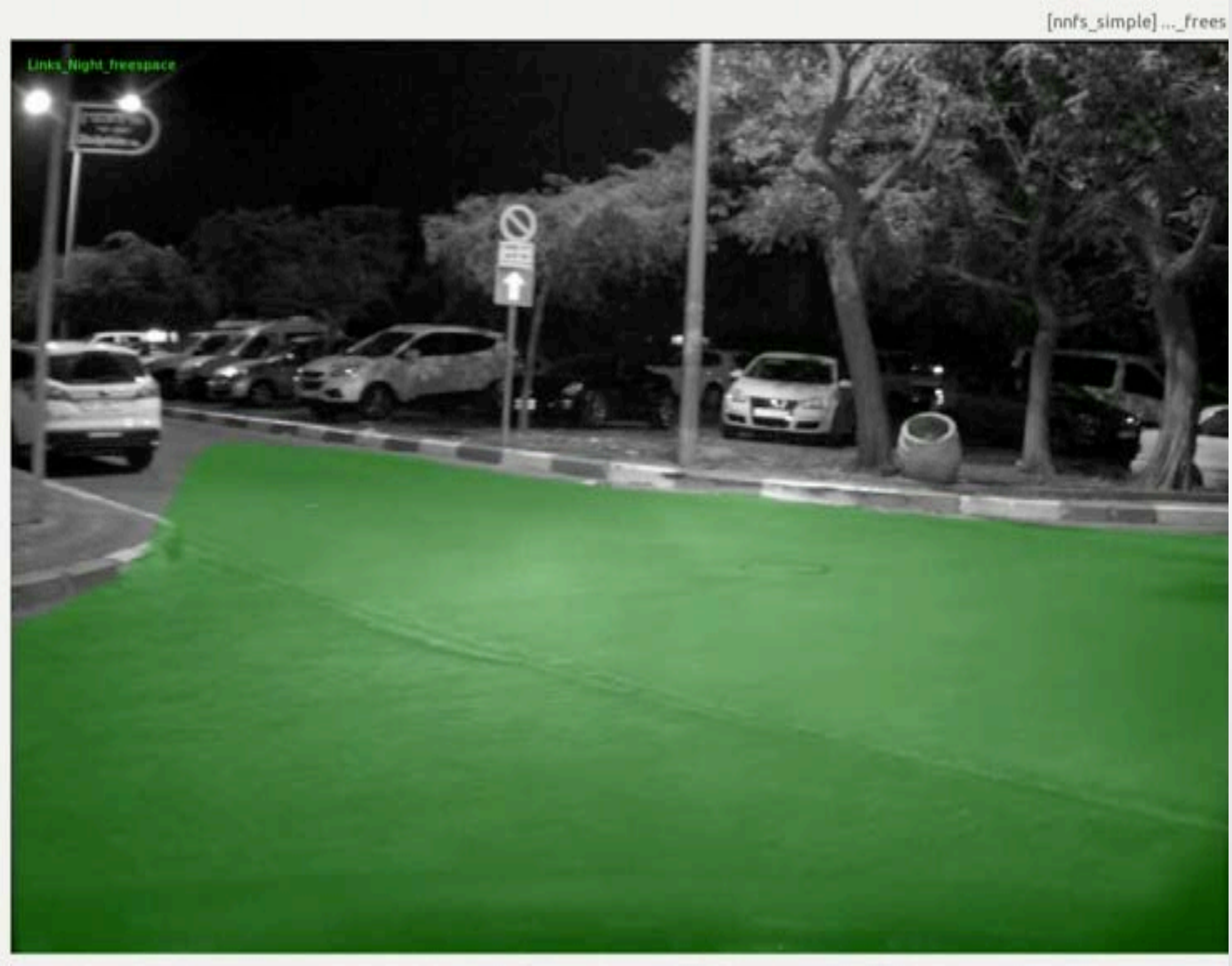
Path Planning using Holistic Cues

HPP: Increasing Availability of Road



Semantic Free Space (SFS)

Free Space through Pixel Labeling

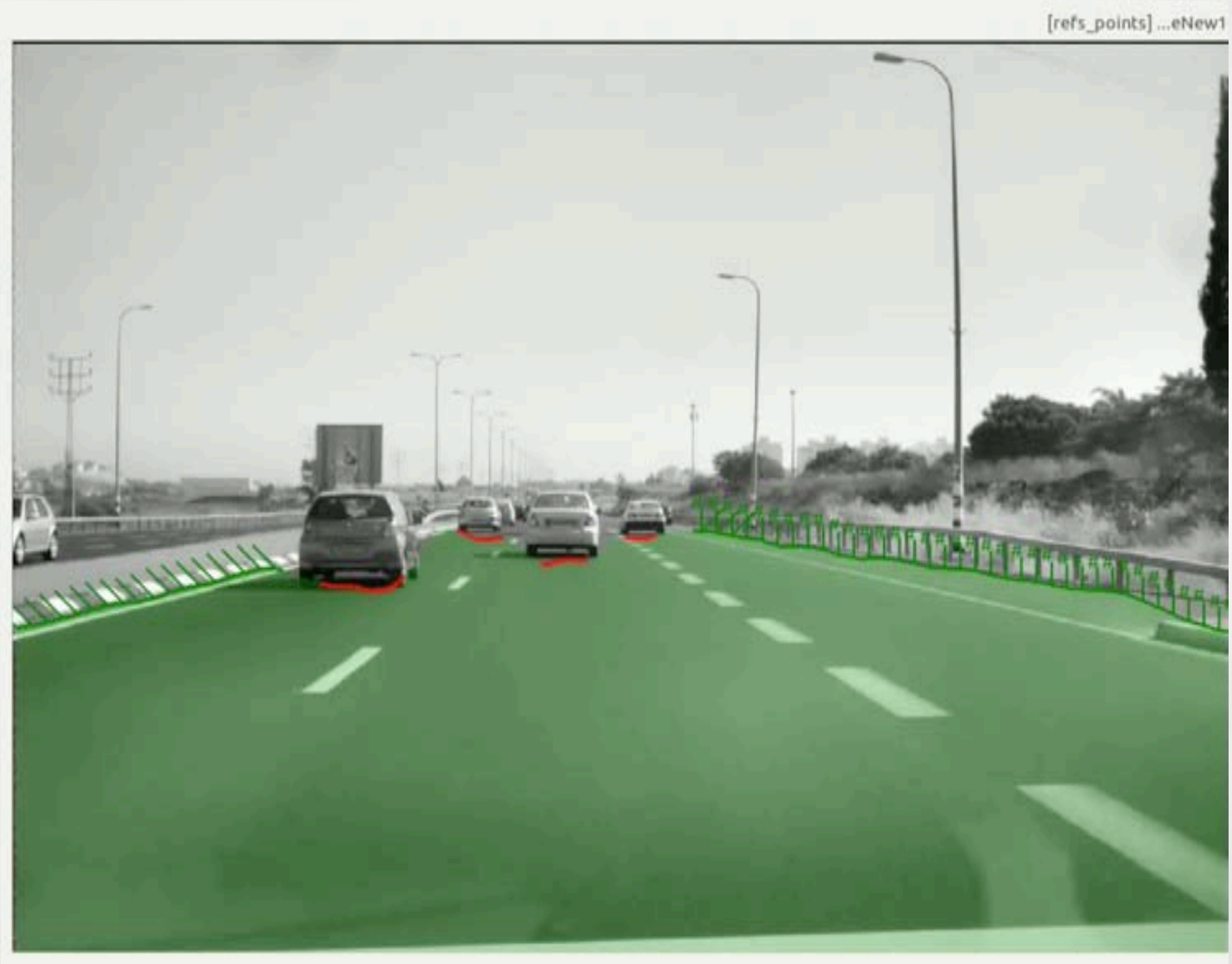


Free Space through Pixel Labeling

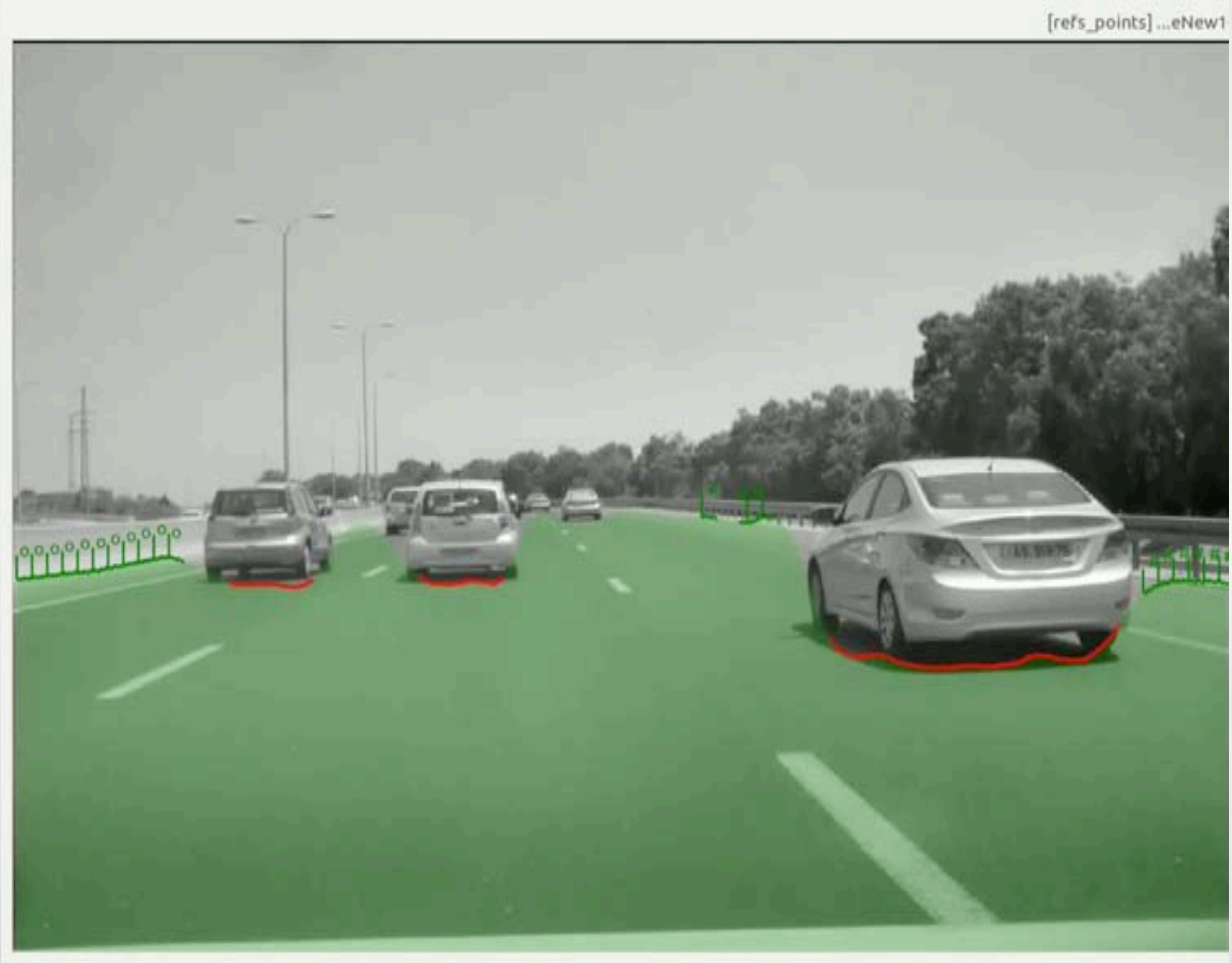
Free Space through Pixel Labeling

Free Space through Pixel Labeling

Free Space through Pixel Labeling



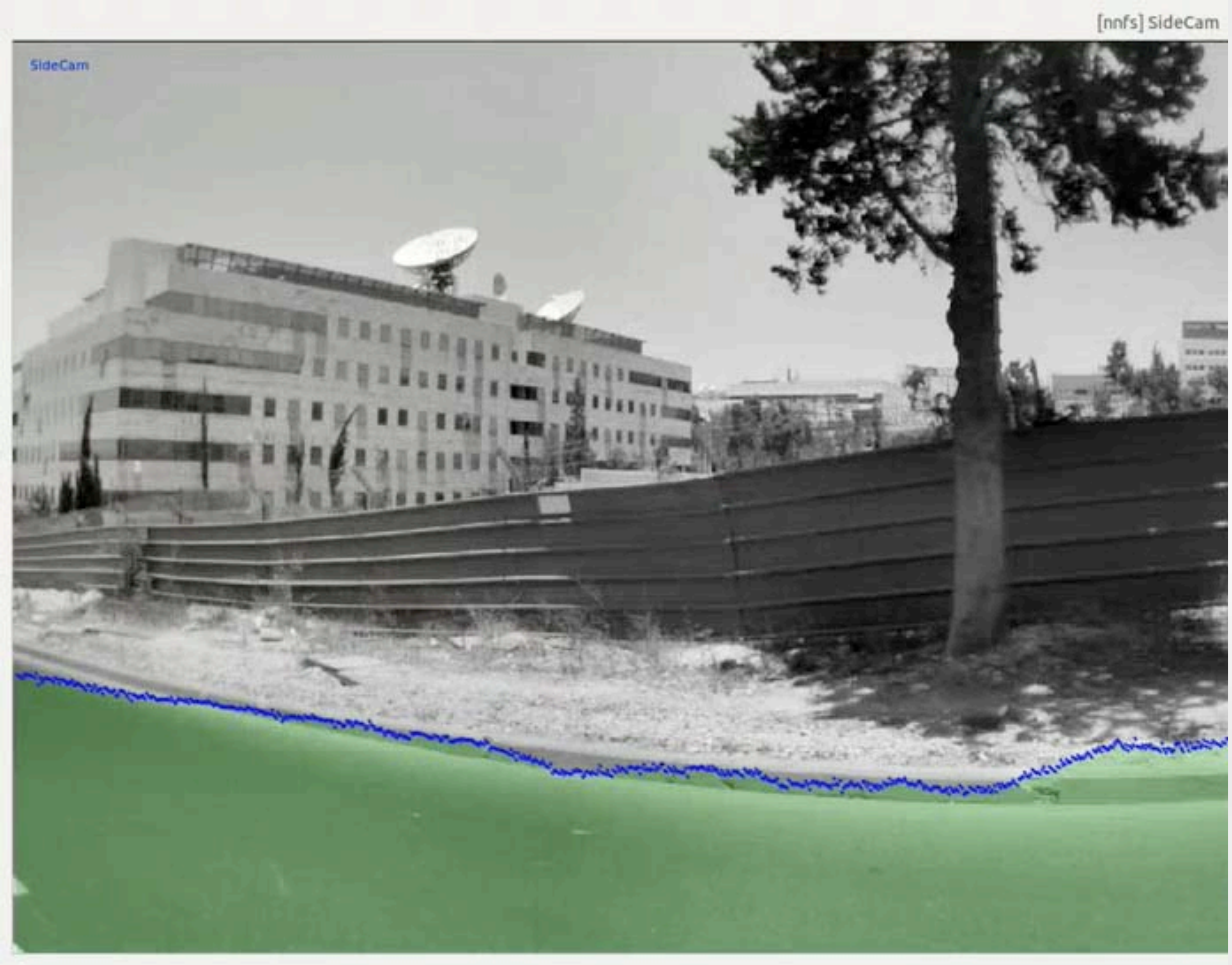
Free Space through Pixel Labeling



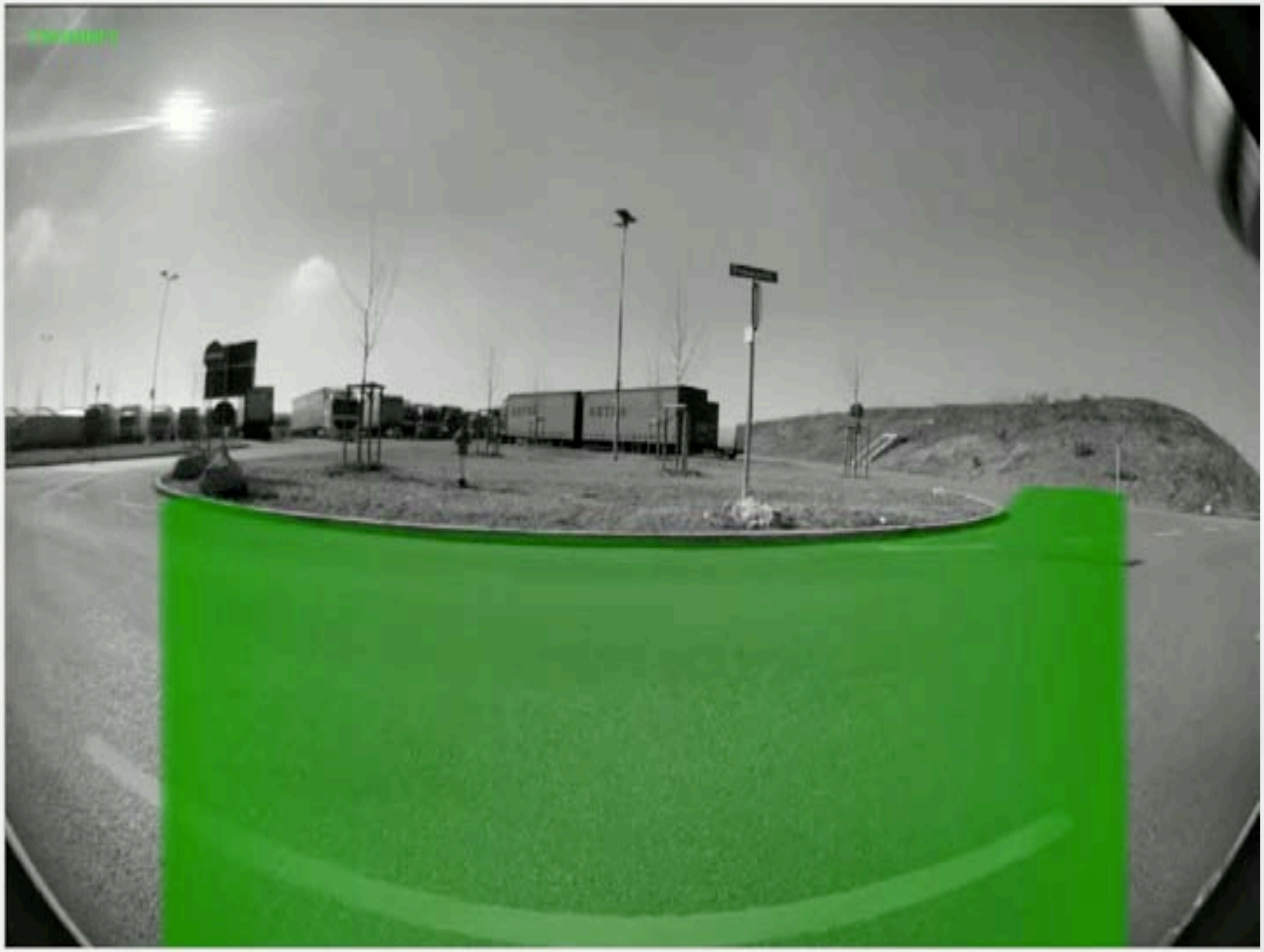
Free Space through Pixel Labeling

SFS from Various Viewpoints and Fields of View

SFS from Various Angles



SFS from Various Angles



SFS from Various Angles

SFS from Various Angles

3D Modeling of Vehicles (3DVD)

3DVD

3DVD

Scene Recognition

Bump Detection (non-geometric)

Bump Detection (non-geometric)

Long Range Stop Line

Lane Assignment

Road Surface Recognition

Traffic Light Detection

TFL: main building blocks

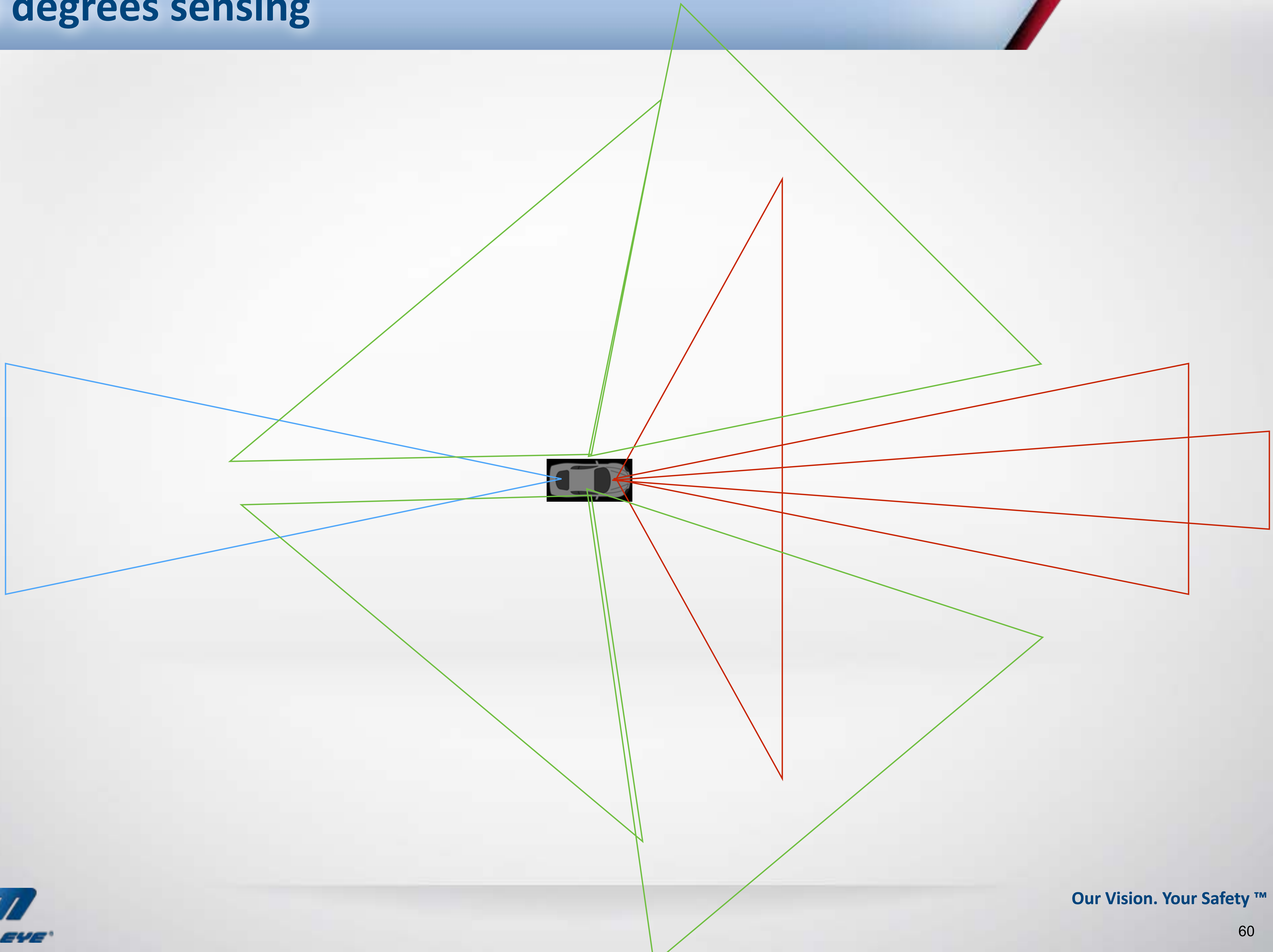
- Detect Traffic Lights (some are country specific)
- Decide “Relevancy” for each TFL in a Junction
- Detect Stop Line
- Detect Road Markings
- Decide on “Lane Assignment”

Scene Recognition: detection junctions in general (as a Prior)

TFL: main building blocks

Multiple Cameras

360 degrees sensing



Hardware Architecture

Automated Driving

Impact of Autonomous Driving

- Hands-free on Highways (no lane change) - Now on Tesla, 2016 GM, Audi,… Driver has primary responsibility (and Alert)
- Highway to Highway: on and off-ramps executed autonomously. Early 2016. Driver has primary responsibility (and Alert)
- ~2018-2020: Driver responsible but not alert. Driver is “attendant” (transition from “primary responsibility” to Monitoring - like in Aviation). The beginnings of disruption.
- ~2020-2022: Driverless cars without passengers. Big disruption.
- ~2025-2030: No driver. Transformative.

Automated Driving