Contextual Visual Object Recognition and Behavior Modeling for Human-Robot Interactivity

Defense Presentation by

Eshed Ohn-Bar

Chair: Prof. Mohan M. Trivedi

Prof. Serge Belongie

Prof. Garrison W. Cottrell

Prof. Bhaskar Rao

Prof. Nuno Vasconcelos

Outline

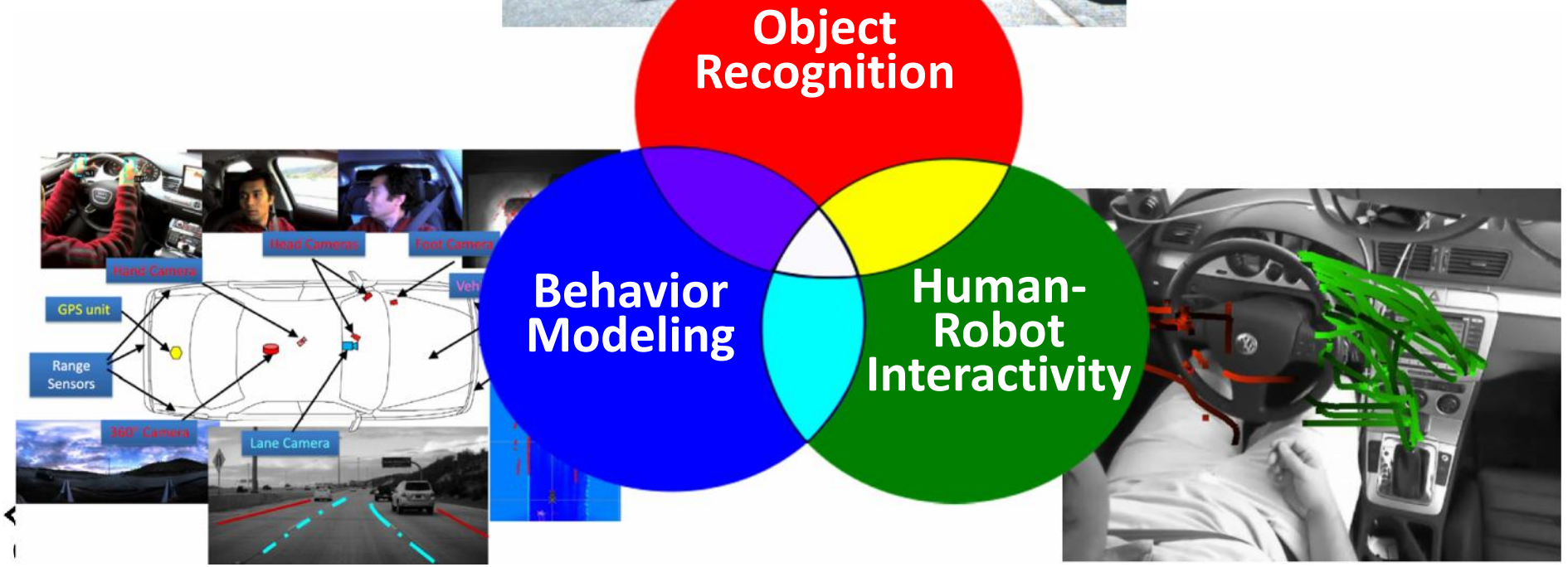
1) Dissertation introduction and **contributions**

2) **Contextual** visual object **detection** and **localization**

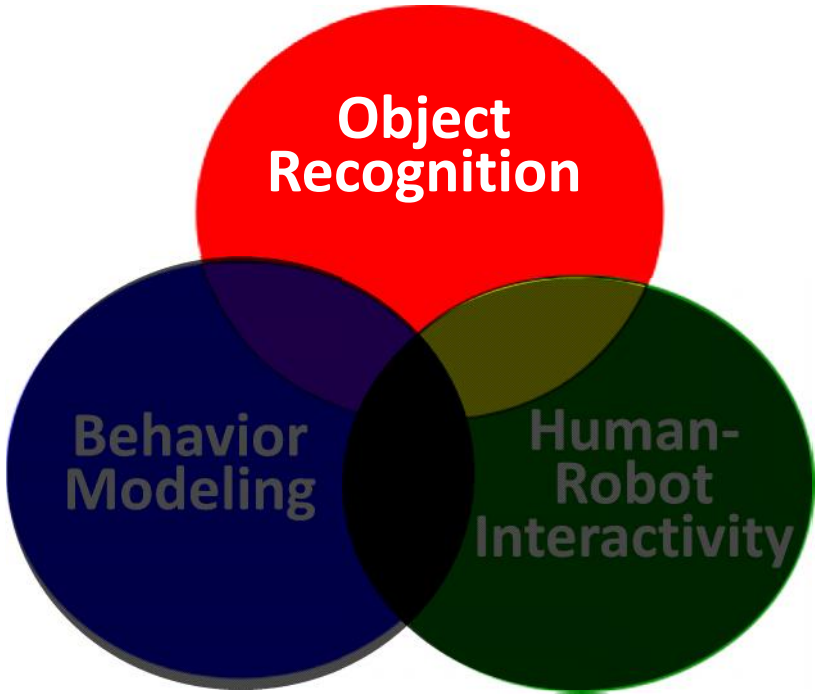
3) Learning spatio-temporal dynamics for **behavior modeling**

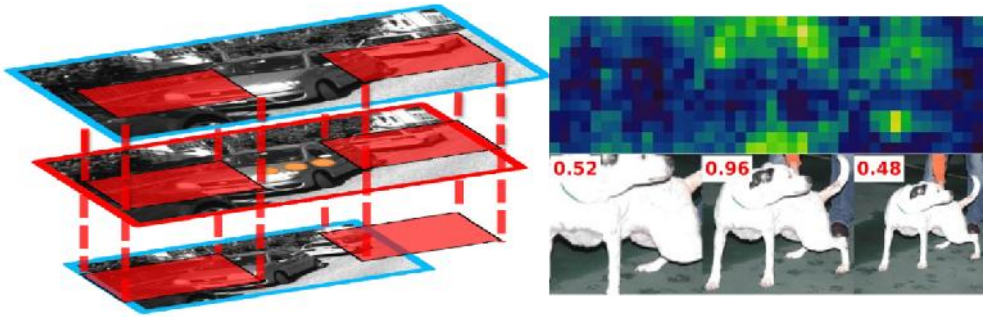
4) Situational awareness and **human-centric recognition** in driving videos

Contextual Visual Object Recognition and Behavior Modeling for Human-Robot Interactivity

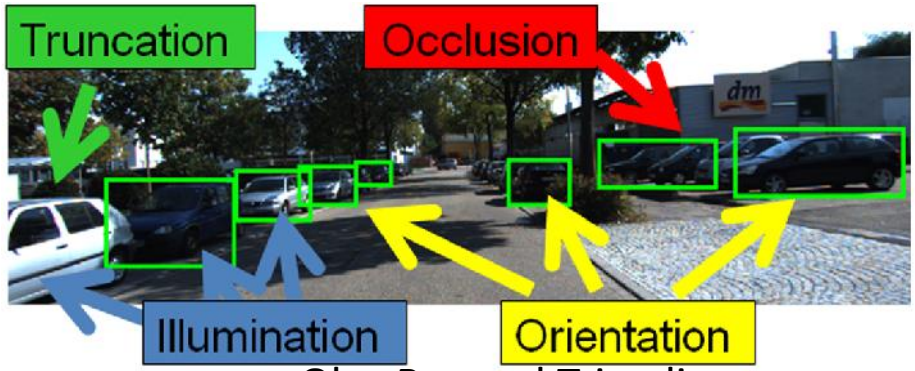


Object Recognition



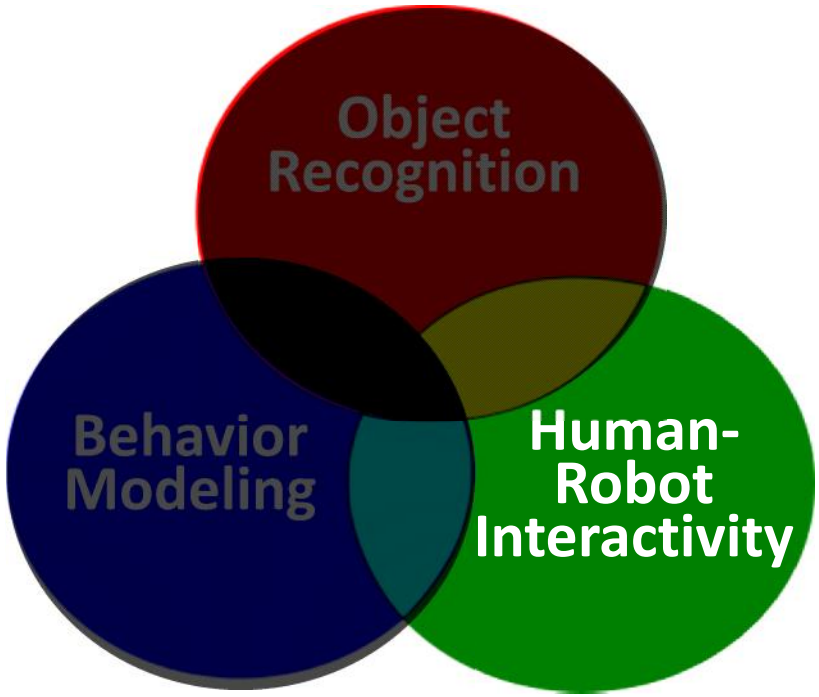


Ohn-Bar and Trivedi, ICPR, Pattern Recognition, 2016

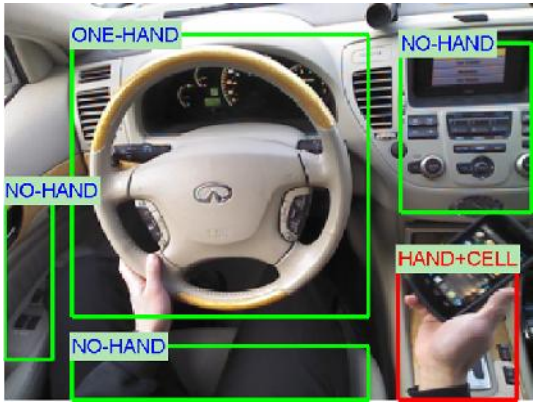


Ohn-Bar and Trivedi, IEEE T-ITS, 2015

Hand Gestures for Interactivity

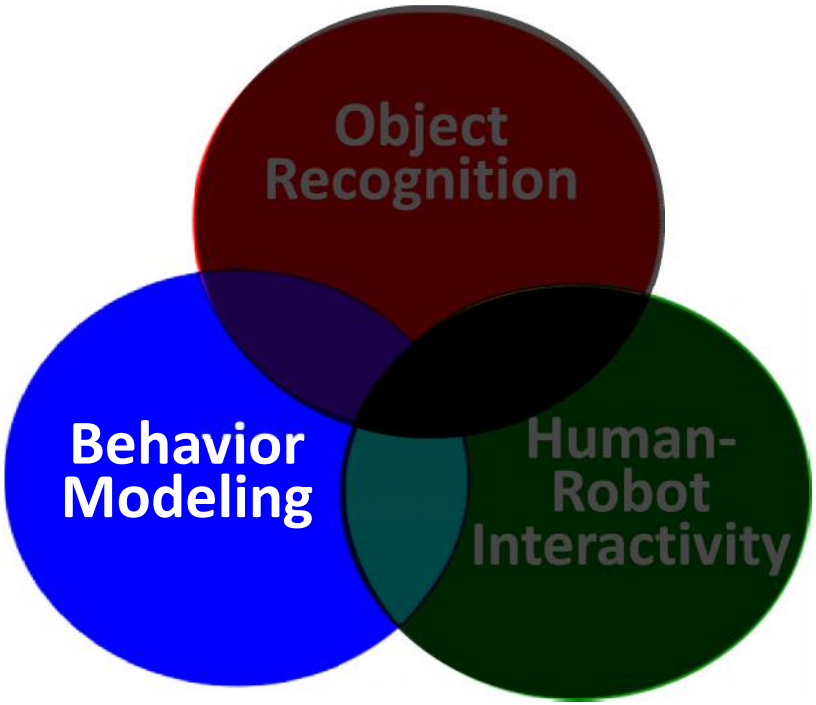


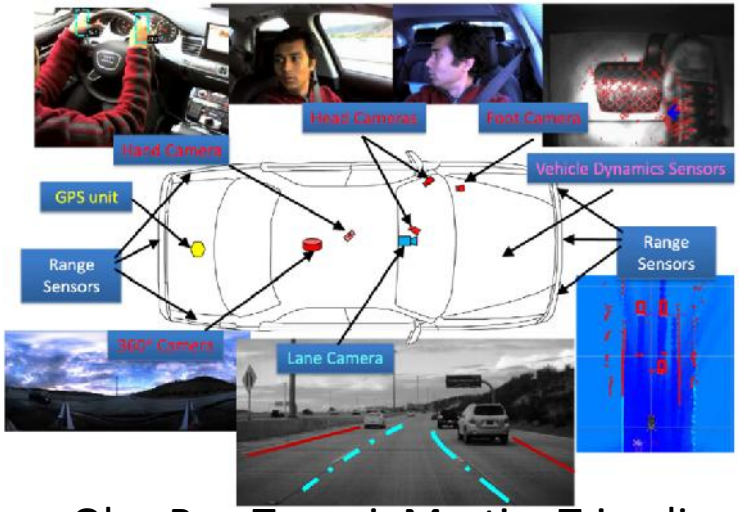
Ohn-Bar and Trivedi, ITSC, T-ITS, 2014



Ohn-Bar and Trivedi, CVPRW, 2013

Multi-Cue Behavior Modeling



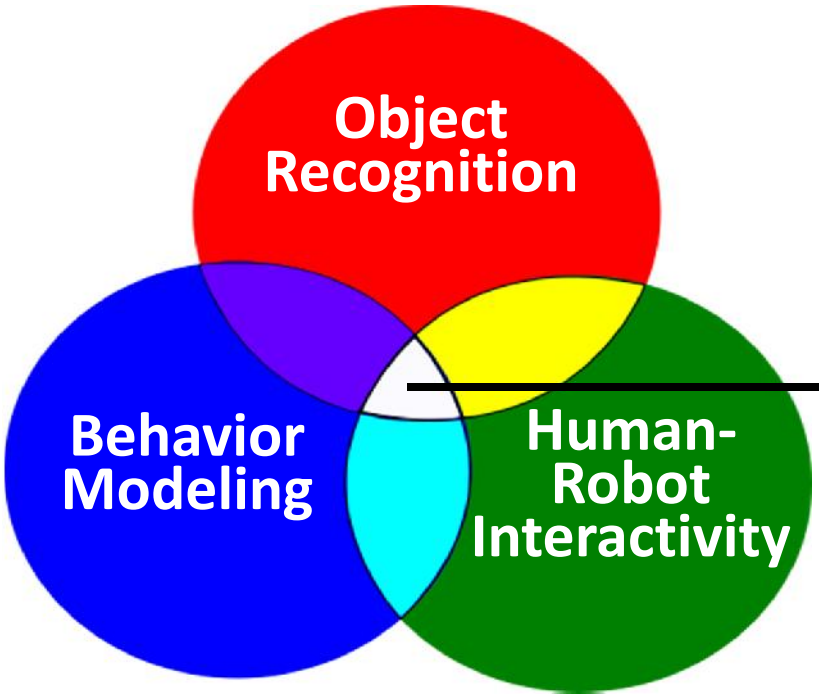


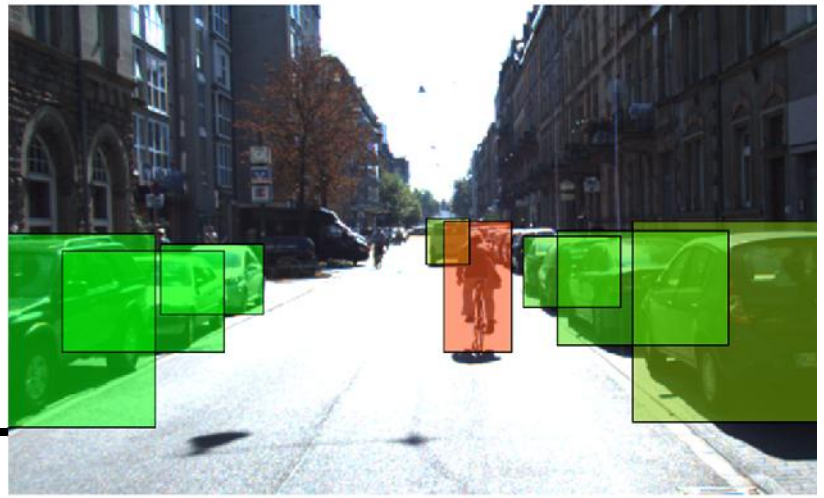
Ohn-Bar, Tawari, Martin, Trivedi, IV, CVIU, 2015

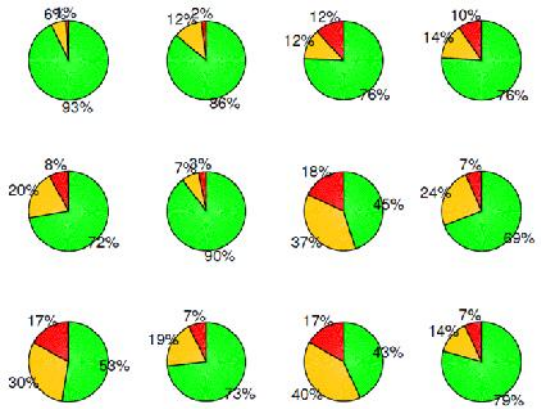
Ohn-Bar, Martin, Tawari, Trivedi, ICPR, 2014

CVRR LISA UCSD LABORATORY FOR INTELLIGENT & SAFE AUTOMOBILES

Situational Awareness and Human-Centric Recognition in Video



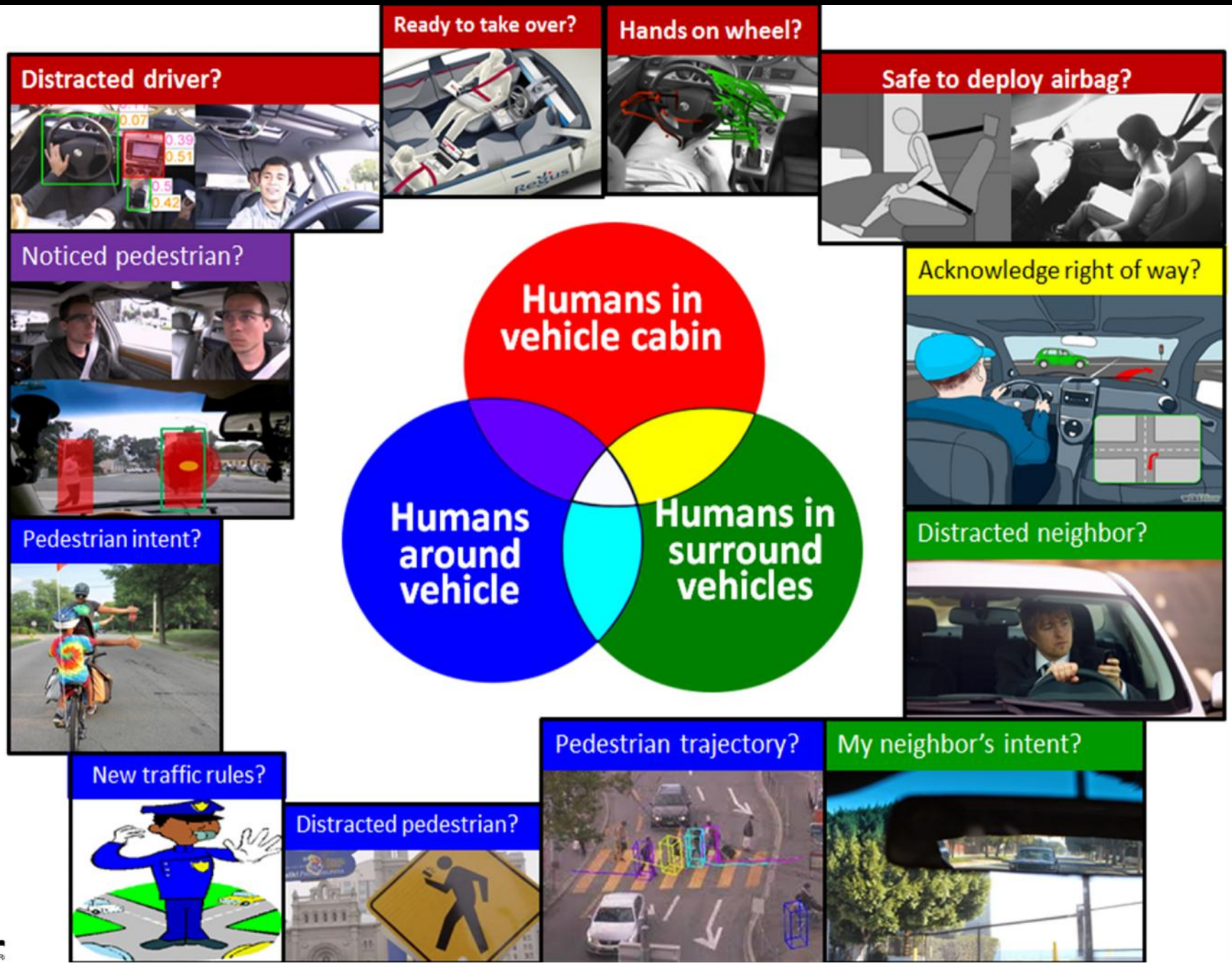




Ohn-Bar and Trivedi, ICPR, Pattern Recognition 2016

CVRR LISA UCSD LABORATORY FOR INTELLIGENT & SAFE AUTOMOBILES

Looking at ~~Drivers~~ Humans for Intelligent Vehicles



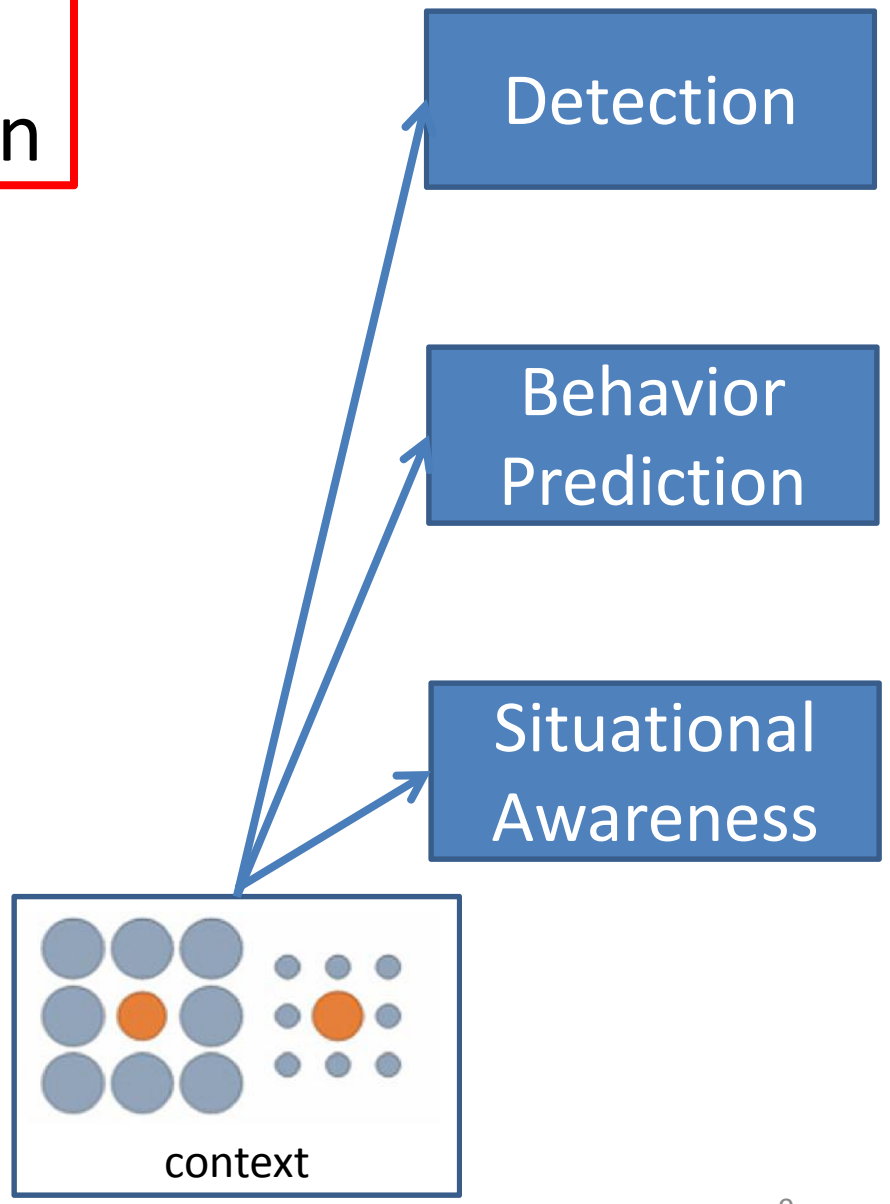
CVRR LABOR

Research Objectives

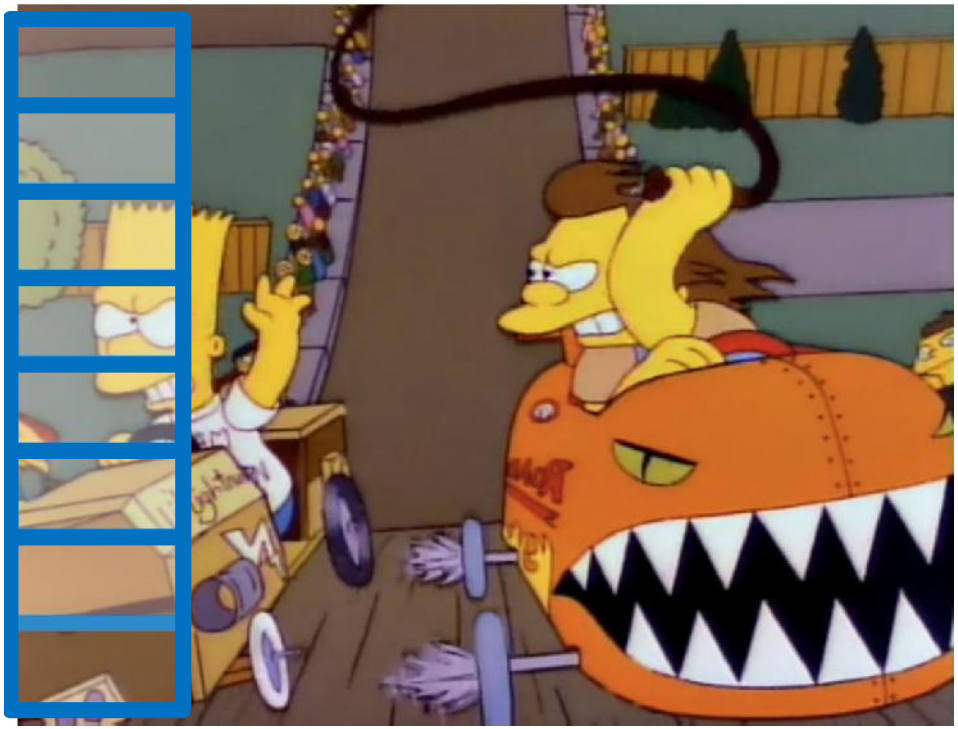
1) Contextual visual object **detection** and localization

2) Learning spatio-temporal dynamics for **behavior** prediction

3) **Situational awareness** and **human-centric** recognition in videos



Single-scale Models for Detection



Single-scale Models for Detection

- Viola-Jones
- HOG-SVM/DPM
- Overfeat
- R-CNN/SPP variants

Single-scale
Model Training

Multi-scale Testing

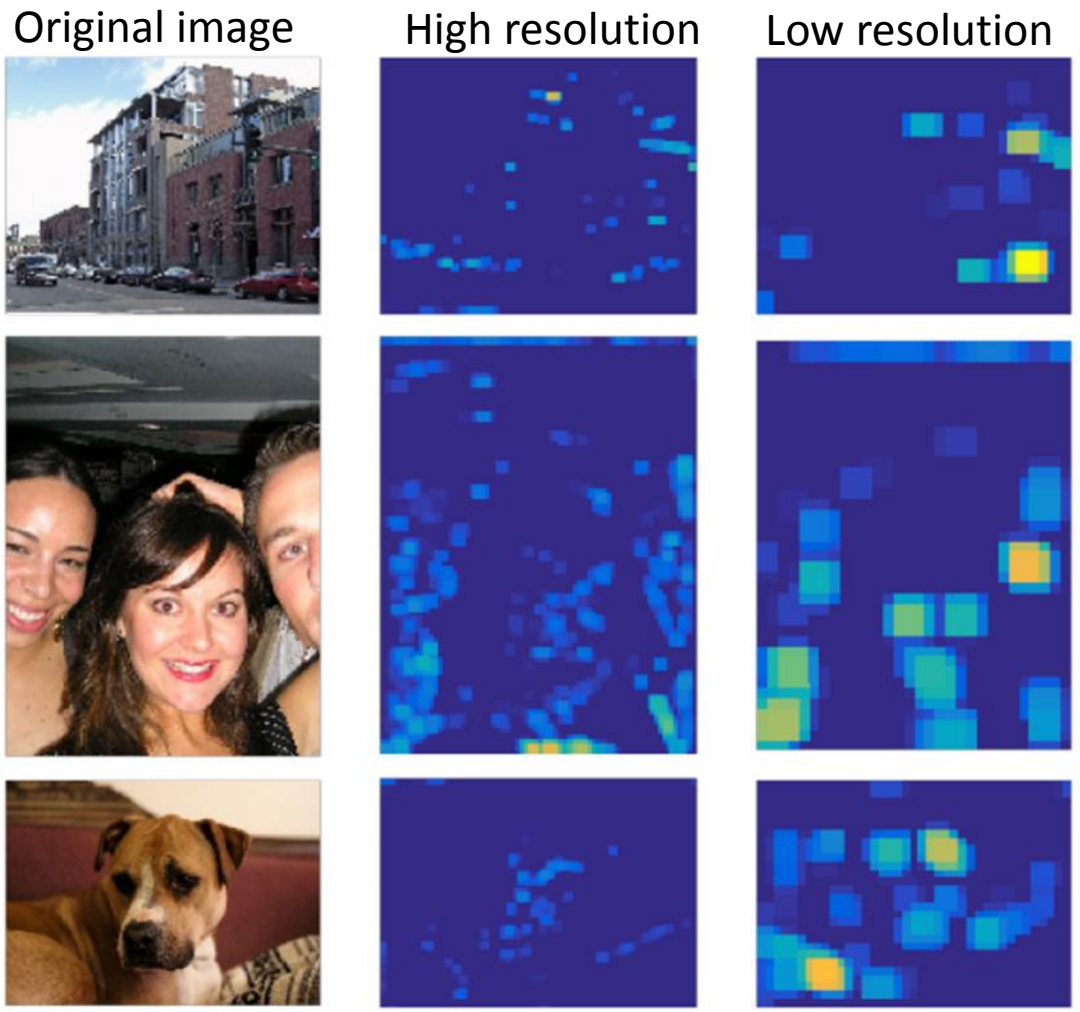
CVRR LISA UCSD
LABORATORY FOR INTELLIGENT & SAFE AUTOMOBILES

Features at Different Scales

- Convolutional feature responses at **different image scales** of two octaves apart.

- The responses are **scale-selective**, capturing different levels of **contextual** information.

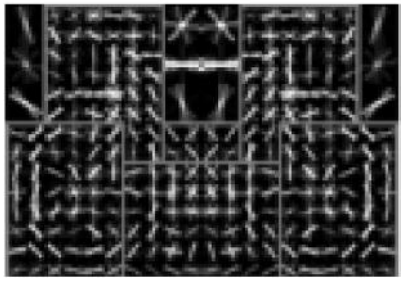
- We study this phenomenon using **scale volumes**. Modeling produces **improved** object **detection** and **localization** performance.



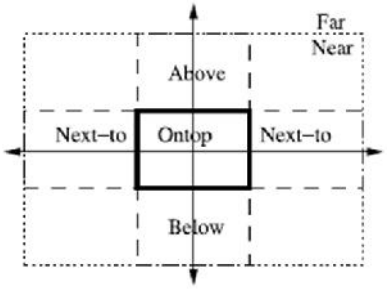
CVRR LISA UCSD LABORATORY FOR INTELLIGENT & SAFE AUTOMOBILES

Related Research Studies

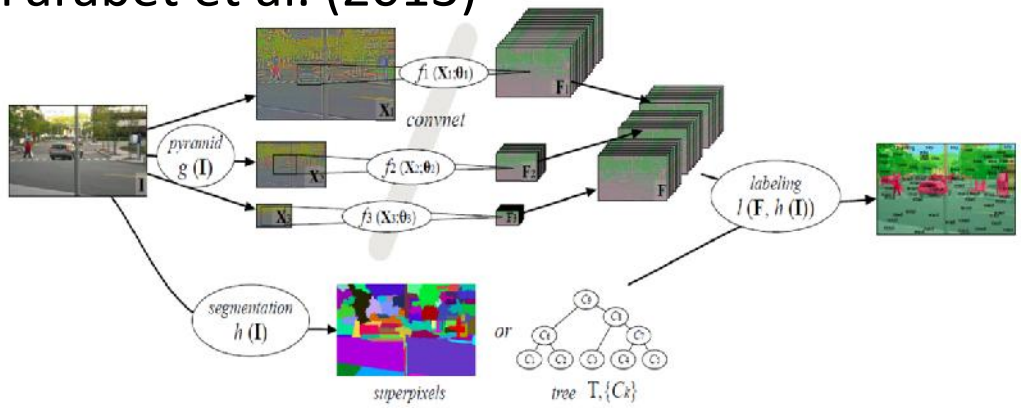
Felzenszwalb et al. (2010)



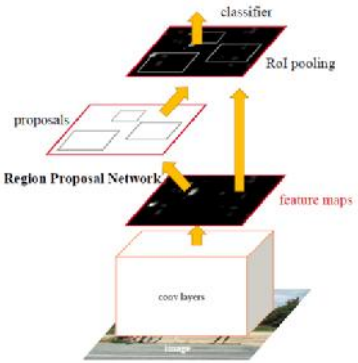
Desai and Ramanan (2011)



Farabet et al. (2013)



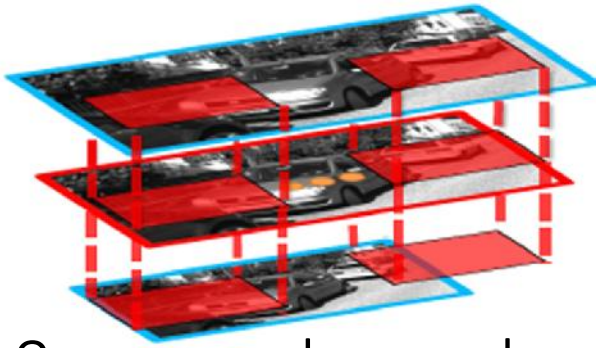
Girshick (2015), Ren et al. (2016)



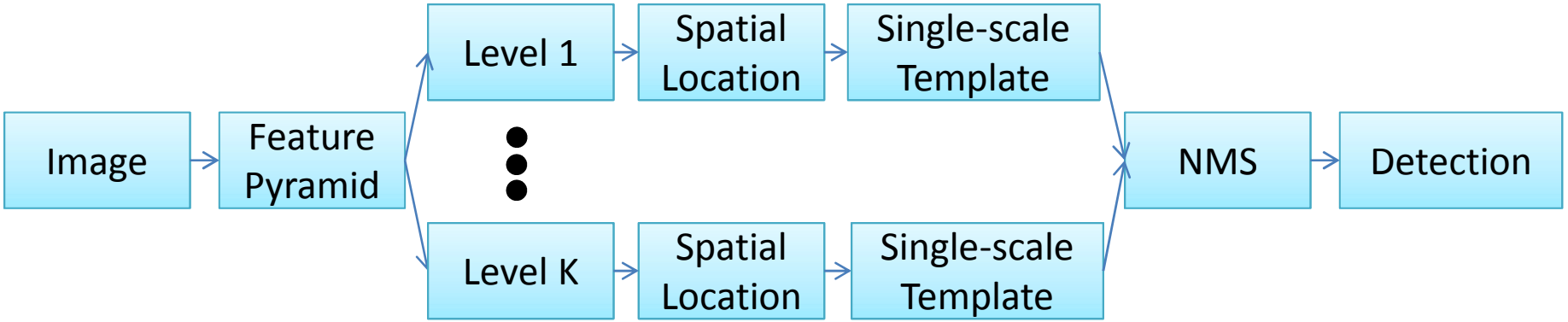
- **Scale and context** as interleaved fundamental tasks in computer vision.
- Contextual **post-processor** vs. **joint** detection and localization
- Segmentation/boundary vs. efficient and accurate **localization**
- Single or two-scale model vs. scale-specific, **scale volumes**
- Sliding window over scales vs. appropriate **label space**

Traditionally…

Traditional, single-scale approaches



Our proposed approach



$$f(p_s) = w \cdot \phi(p_s)$$

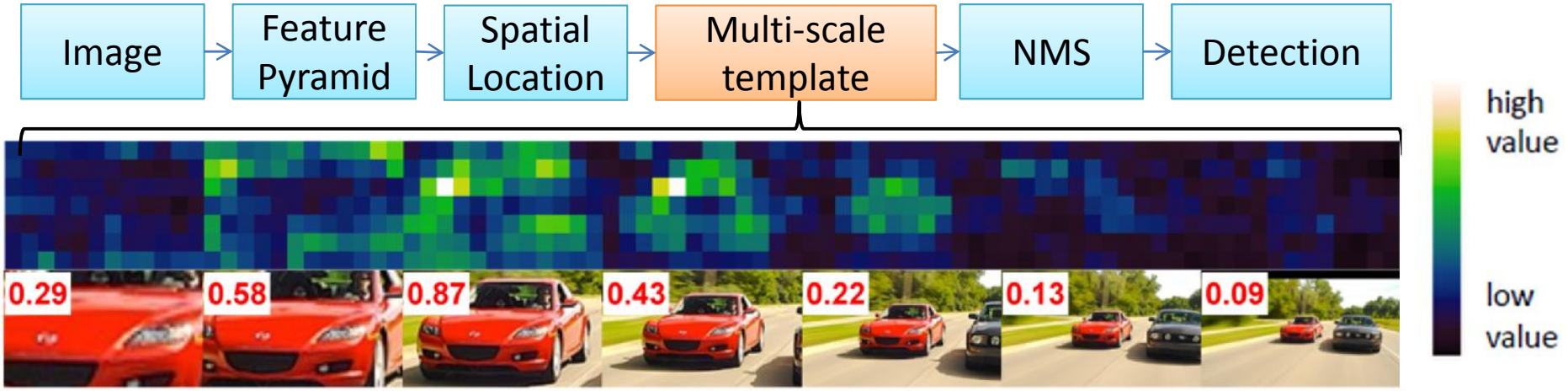
$\in \{-1, 1\}$

= (x,y,s), s-th level of a feature pyramid

Multi-Scale Structure Approach - Inference

$$f(p_s) = w \cdot \phi(p_s)$$

$\in \{-1, 1\}$ = (x,y,s), s-th level of a feature pyramid



$$\psi(p) = (\phi(p_1), \ldots, \phi(p_S)) \in \mathbb{R}^{d \times S}$$

$$y = (y^l, y^b, y^s) \in \mathcal{Y}$$

$$y^l \in \{-1, 1\}$$

$$y^b \in \mathbb{R}^4$$

$$f(p) = \max_{s \in \{1,\ldots,K\}} w_s \cdot \psi(p)$$

Generalizes single-scale

Train with Structural SVM and an overlap loss

Multi-Scale Structure Approach - Training

Training with Structural SVM

$$\min_{w,\xi} \frac{1}{2}||w||^2 + C\sum_{i=1}^{n} \xi_i$$

$$\text{s.t. for } \forall i, \bar{y} \in \mathcal{Y} \setminus y_i$$

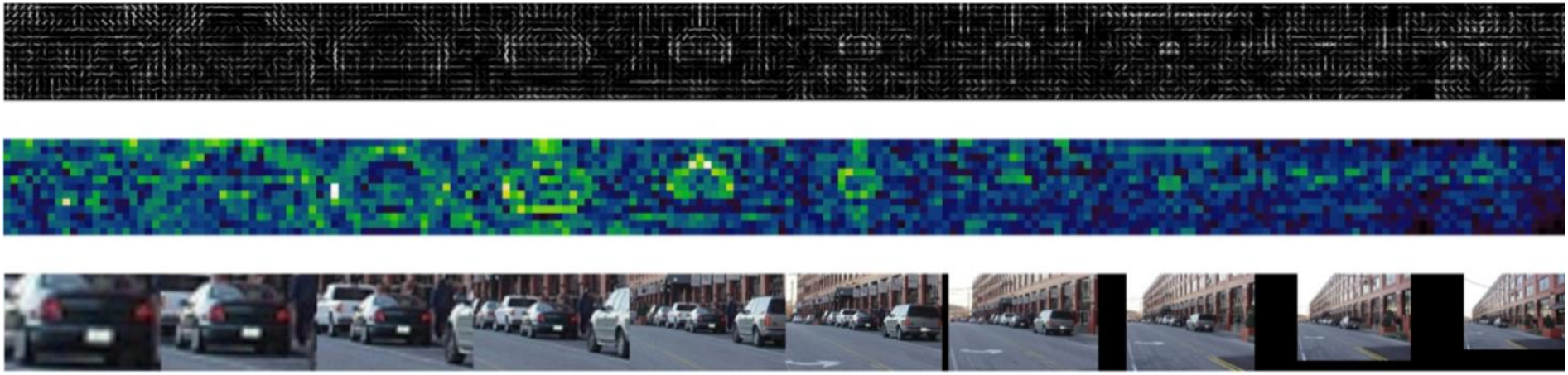
$$w \cdot (\Phi(p^i, y_i) - \Phi(p^i, \bar{y})) \geq L(y_i, \bar{y}) - \xi_i$$

$$L(y, \hat{y}) = \begin{cases} 0 & \text{if } y^l = \hat{y}^l = -1 \text{ or} \\ & \max_{i \in \{1,\ldots,N\}} \text{ov}(y^b, \hat{y}_i^b) < 0.6 \\ 1 & \text{otherwise} \end{cases}$$

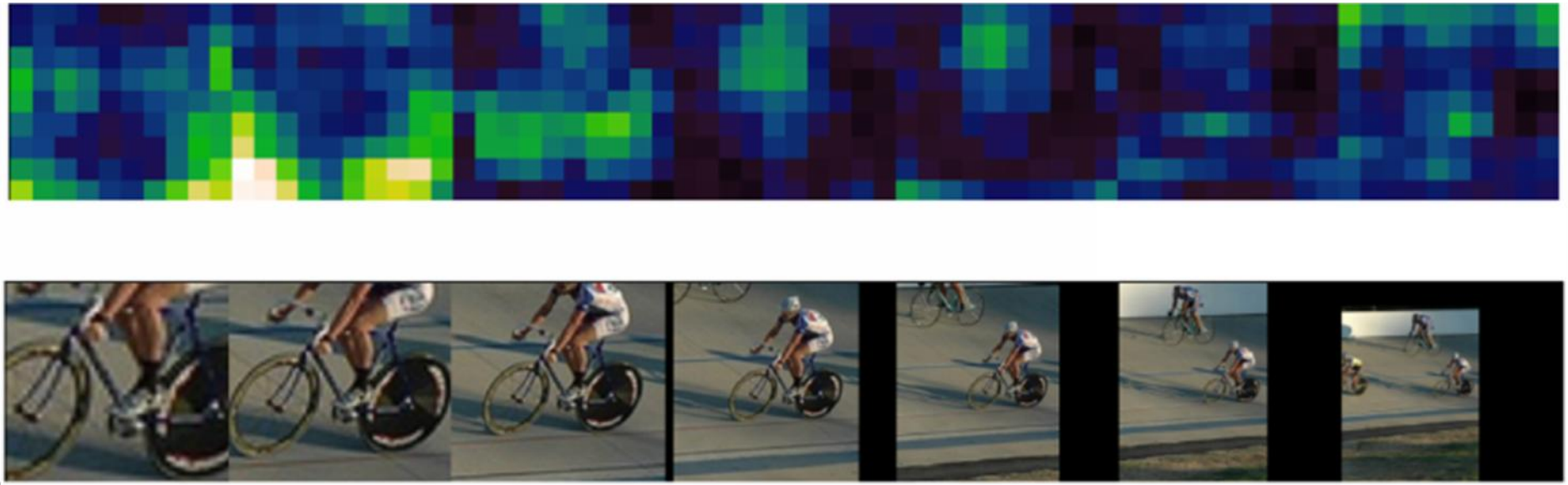
CVRR LISA UCSD LABORATORY FOR INTELLIGENT & SAFE AUTOMOBILES

Multi-Scale Models - Evaluation

- HOG and CNN-AlexNet
- PASCAL VOC 2007, **improves** mAP by **7.74%** and **4.48%**, respectively

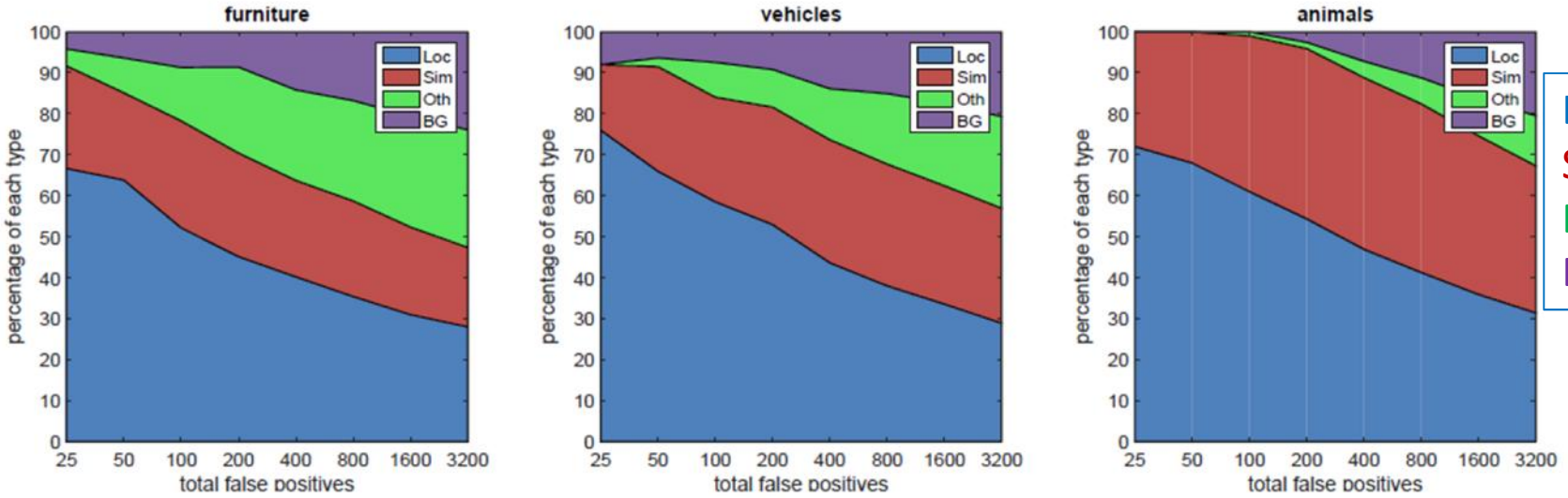


Car-HOG

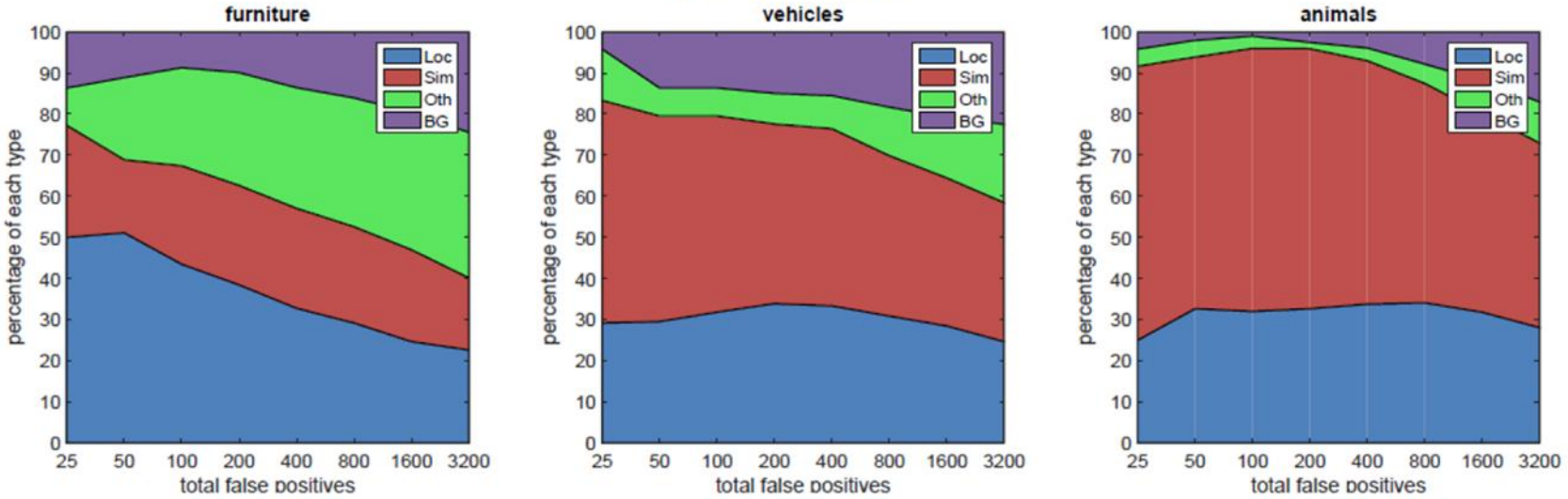


Bicycle-CNN

Multi-Scale Models Reduce Localization Errors



(a) CNN-baseline



(b) CNN-MSS

Legend

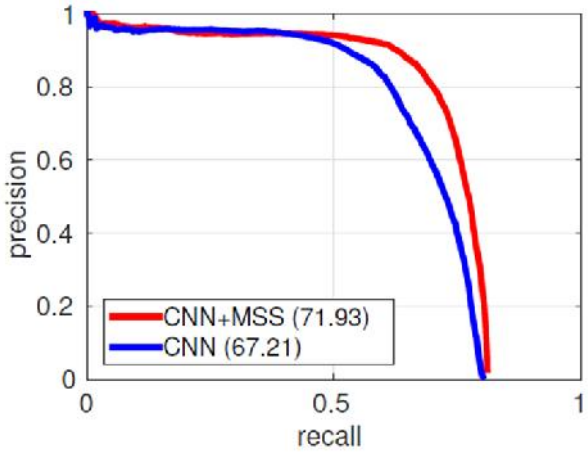
Localization
Similar
Dissimilar
Background

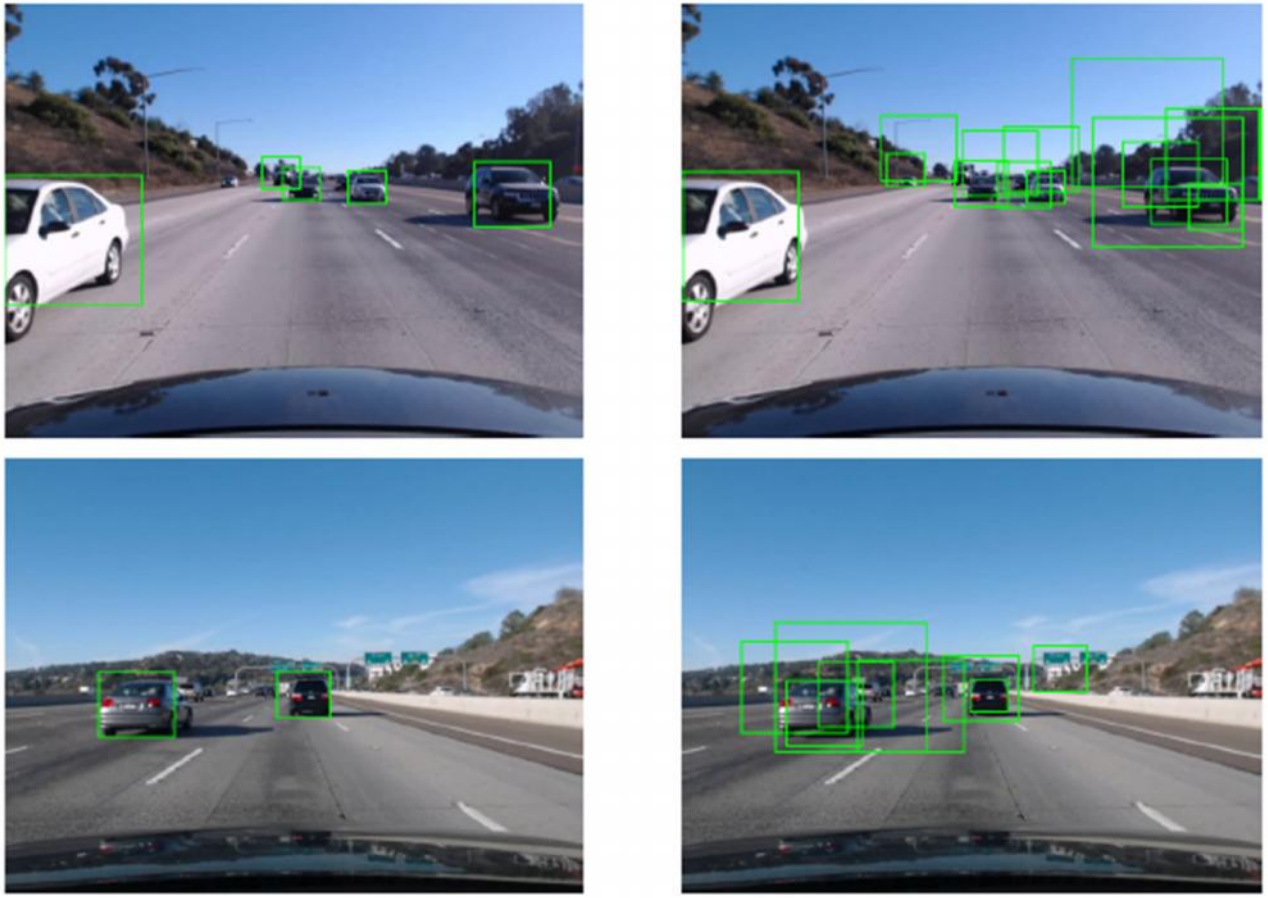
CVRR LISA UCSD LABORATORY FOR INTELLIGENT & SAFE AUTOMOBILES

Hoiem et al. 2012

Multi-Scale Models - Results

- LISA Highway Rear+Front view dataset
- Vehicle detection task, large variation in scale
- AP increase from 67.21% to 71.93% (4.72%)





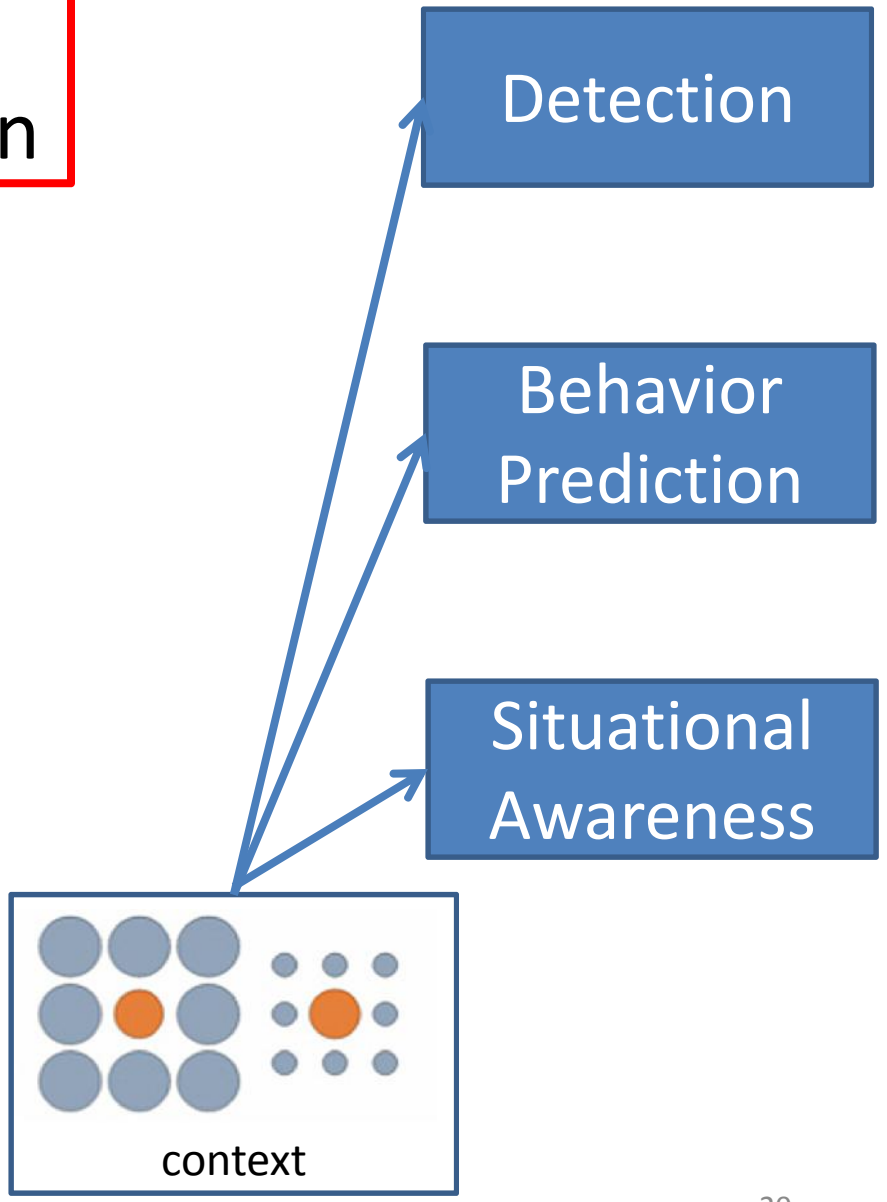
CNN-MSS (proposed) CNN (baseline)

Research Objectives

1) Contextual visual object **detection** and localization

2) Learning spatio-temporal dynamics for **behavior** prediction

3) **Situational awareness** and **human-centric** recognition in videos

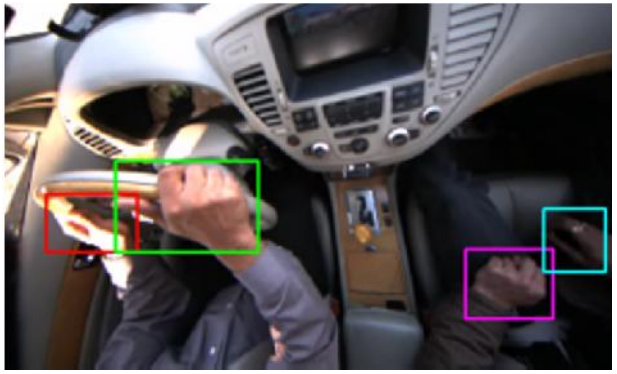


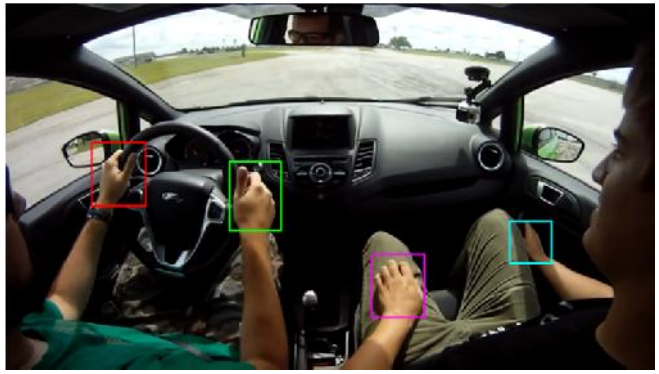
Looking at Hands

52 videos, 44 vehicles, 7 viewpoints

Datasets and current top results
cvrr.ucsd.edu/vivachallenge

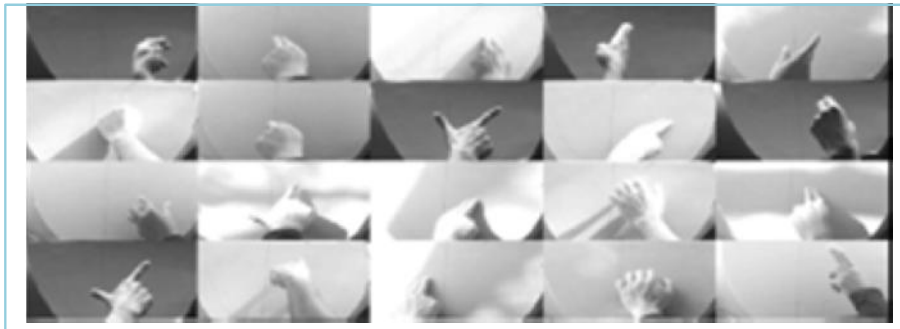
- **Detection:**





- **Tracking:**

- **Dynamic Gestures:**



CVPR-HANDS,
IV 2015, 2016

CVRR LISA UCSD LABORATORY FOR INTELLIGENT & SAFE AUTOMOBILES

Looking at Hands

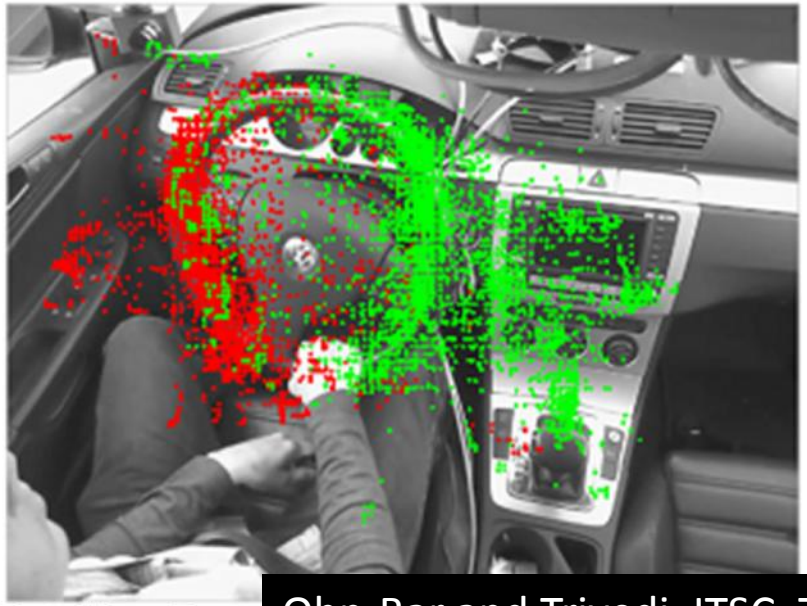
Legend:
Left Hand
Right Hand

Ohn-Bar and Trivedi, ITSC, T-ITS, 2014-2015

Looking at Hands

Left Hand

Right Hand

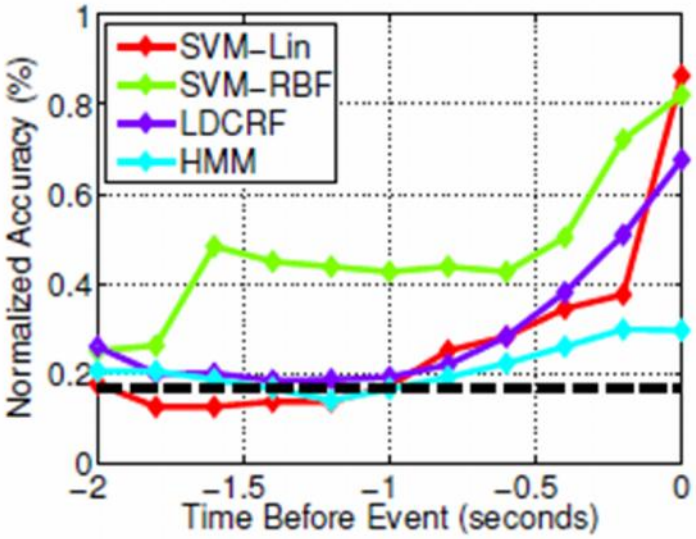


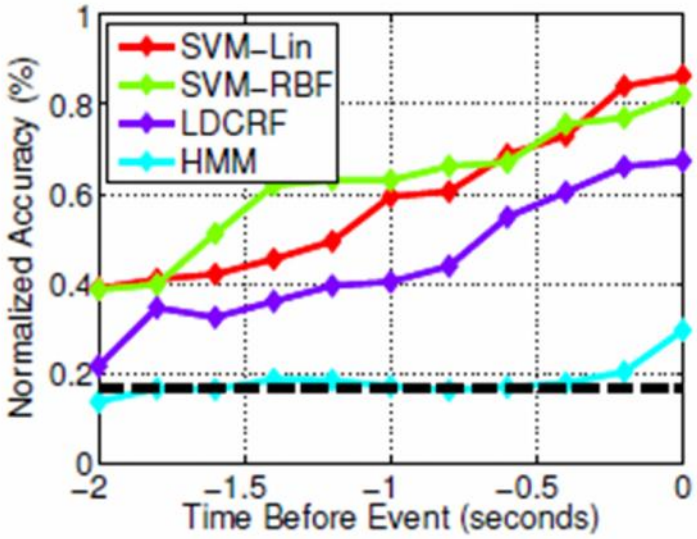
Ohn-Bar and Trivedi, ITSC, T-ITS, 2014-2015

CVRR LISA UCSD LABORATORY FOR INTELLIGENT & SAFE AUTOMOBILES

Reaching and Retracting Gestures

Prediction of about 200 ms





CVRR LISA UCSD LABORATORY FOR INTELLIGENT & SAFE AUTOMOBILES

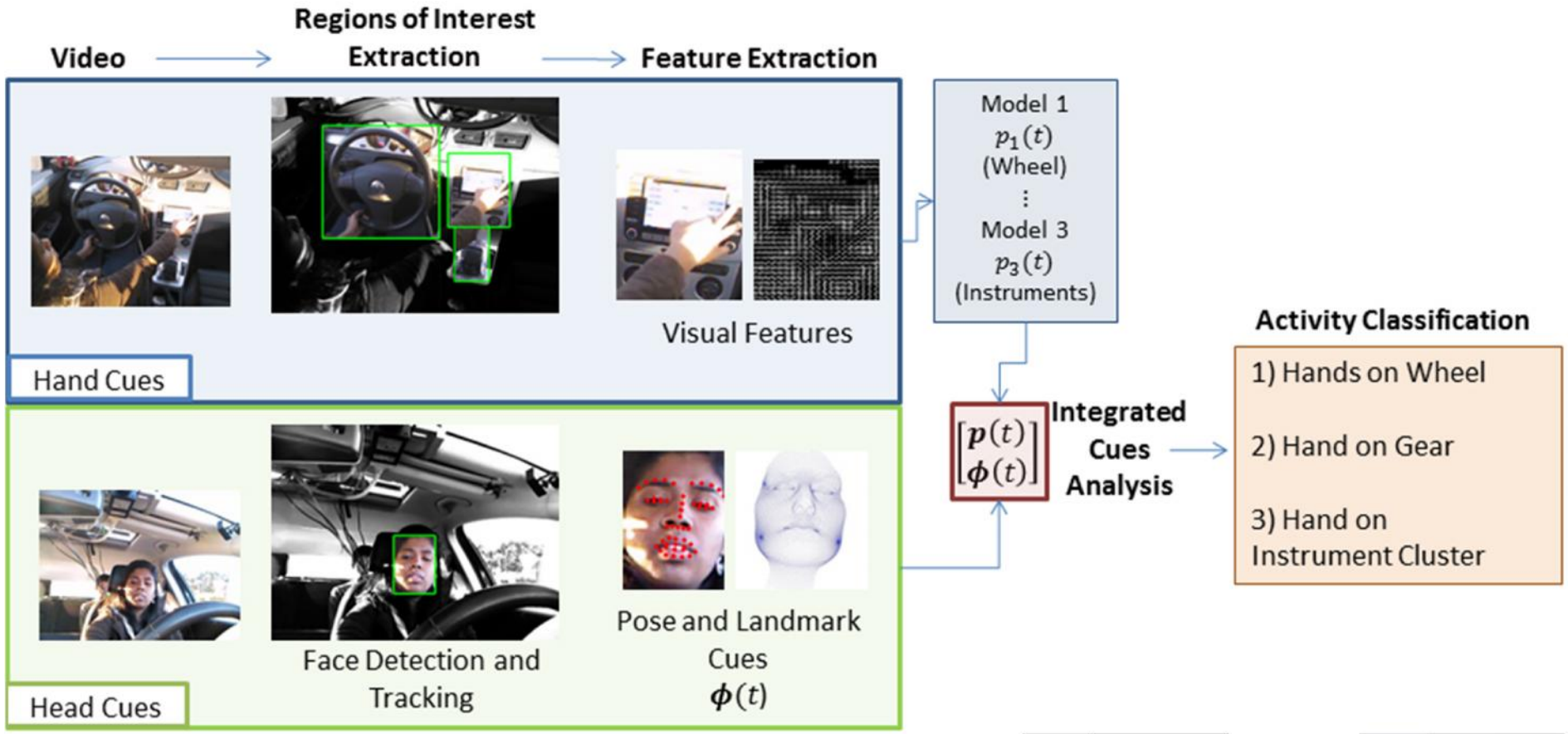
Ohn-Bar and Trivedi, ITSC, T-ITS, 2014-2015

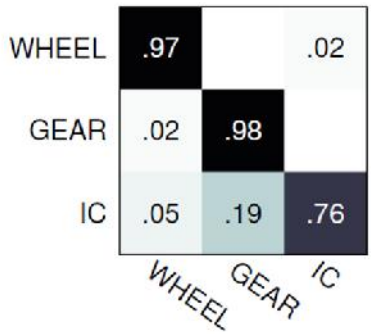
Multi-Cue, Temporal Context for Driver Behavior Modeling

CVRR LISA UCSD LABORATORY FOR INTELLIGENT & SAFE AUTOMOBILES

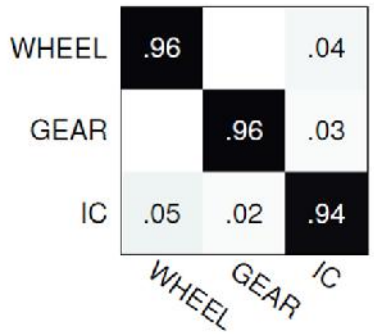
Ohn-Bar, Martin, Tawari, Trivedi, ICPR, 2014

Multi-Cue, Temporal Context for Driver Behavior Modeling





(a) Hand Only (**90%**)

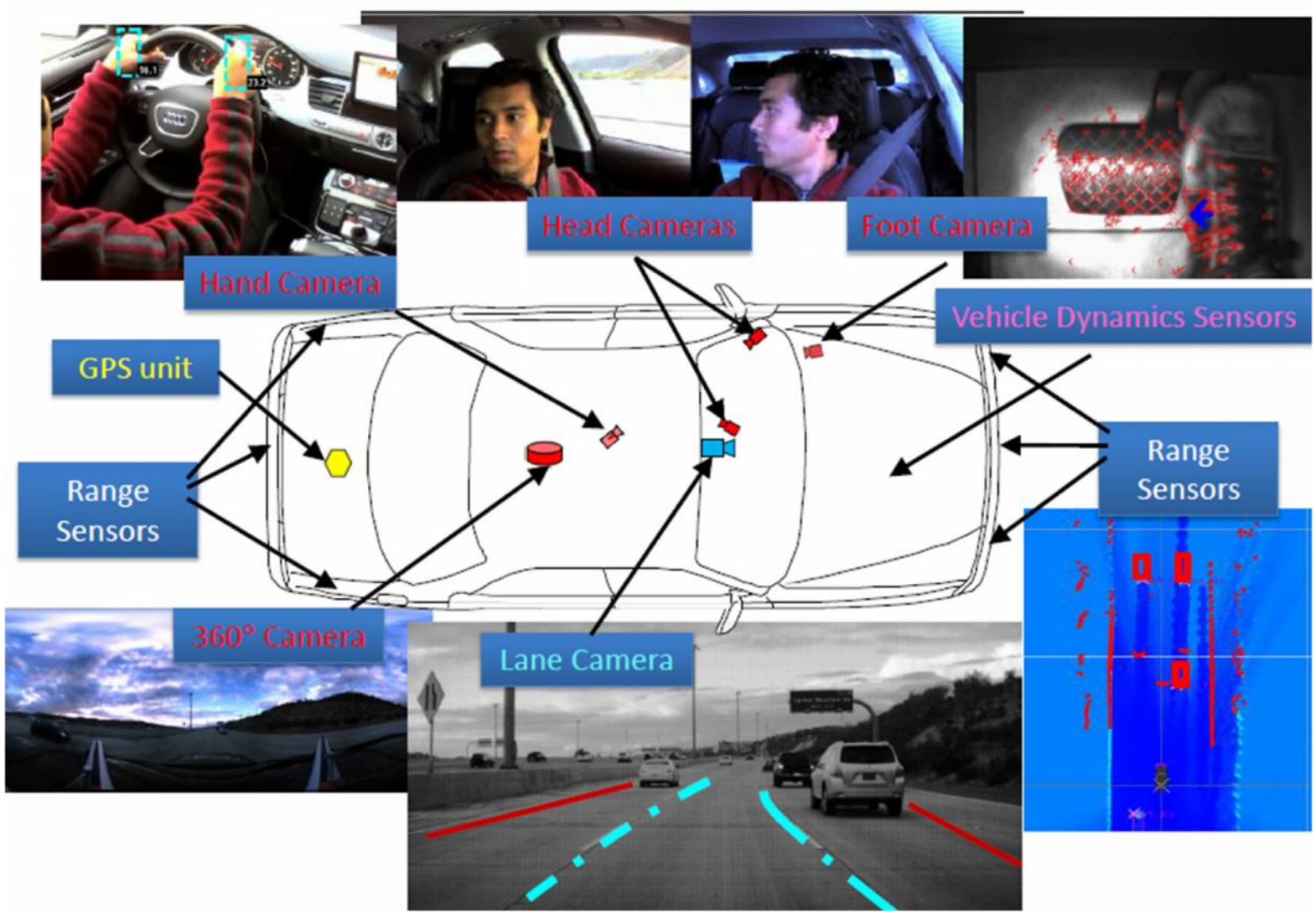


(b) Hand+Head (**94%**)

Ohn-Bar, Martin, Tawari, Trivedi, ICPR, 2014

CVRR LISA UCSD LABORATORY FOR INTELLIGENT & SAFE AUTOMOBILES

Multi-Cue, Multi-Modal Temporal Context for Driver Behavior Modeling



CVRR LISA UCSD LABORATORY FOR INTELLIGENT & SAFE AUTOMOBILES

Ohn-Bar, Tawari, Martin, Trivedi, CVPRW 2014, CVIU, 2015

Overtake or Brake?

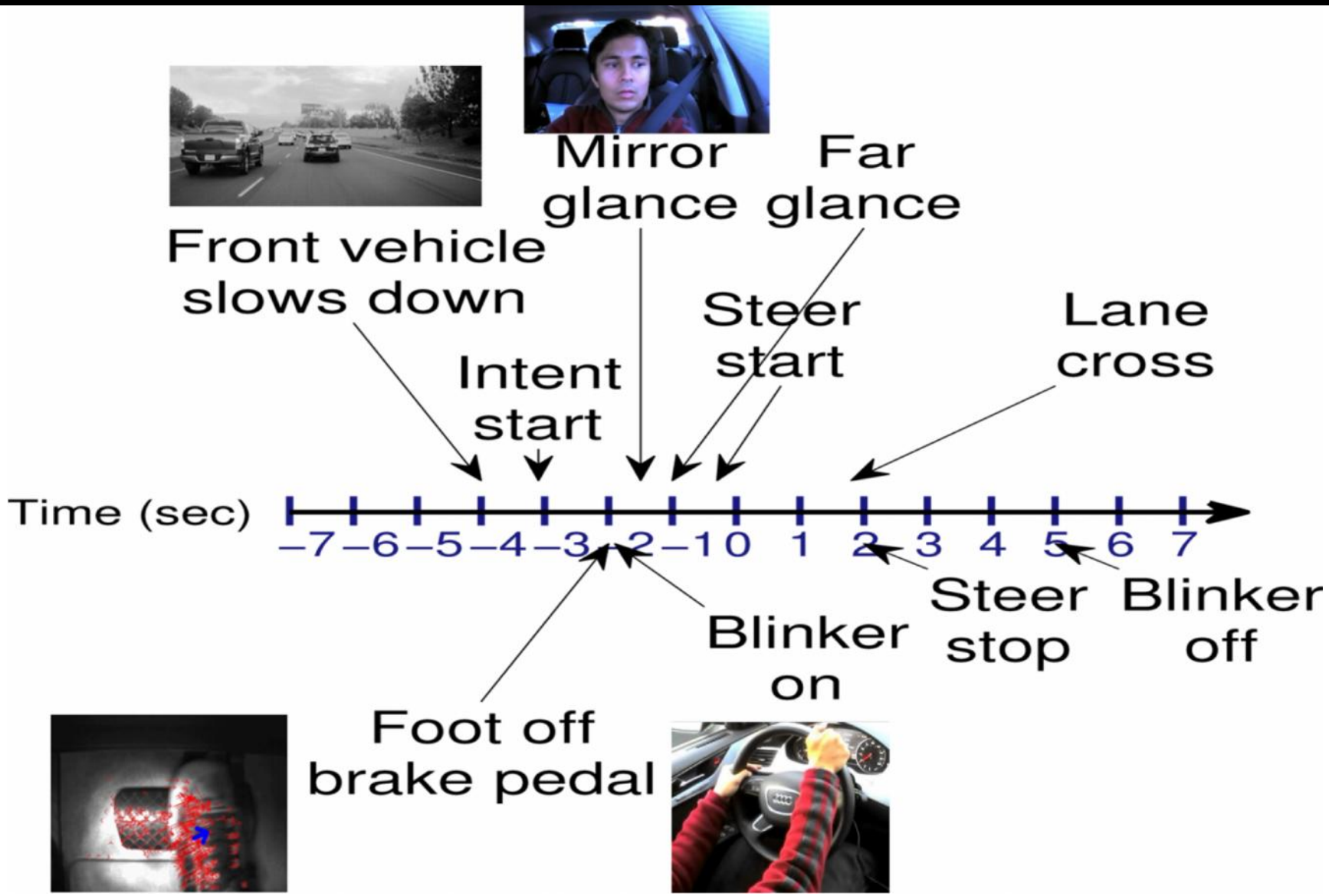
Ohn-Bar, Tawari, Martin, Trivedi, CVPRW 2014, CVIU, 2015

Driver Maneuver Analysis – Related Studies

Study	Maneuvers	Inputs
Doshi et al.	Lane Change	Vehicle, Head, Lane, Radar
Tran et al.	Brake	Foot
Cheng et al.	Turns	Vehicle, Head, Hand
Pugeault and Bowden	Brake, acceleration, steering	Pre-attentive
Mori et al.	Lane change	Radar, Gaze
Liebner et al.	Intersection turns and stop	GPS
Berndt and Dietmayer	Lane change	Vehicle, Lane, GPS, Map
Our work	Overtake or Brake	Vehicle, Head, Hand, Lane, Radar, Foot, Pre-attentive, Gaze

Ohn-Bar, Tawari, Martin, Trivedi, CVPRW 2014, CVIU, 2015

An Example Over-Take Maneuver



Ohn-Bar, Tawari, Martin, Trivedi, CVPRW 2014, CVIU, 2015

Testbed and Signals

- 2011 Audi A8

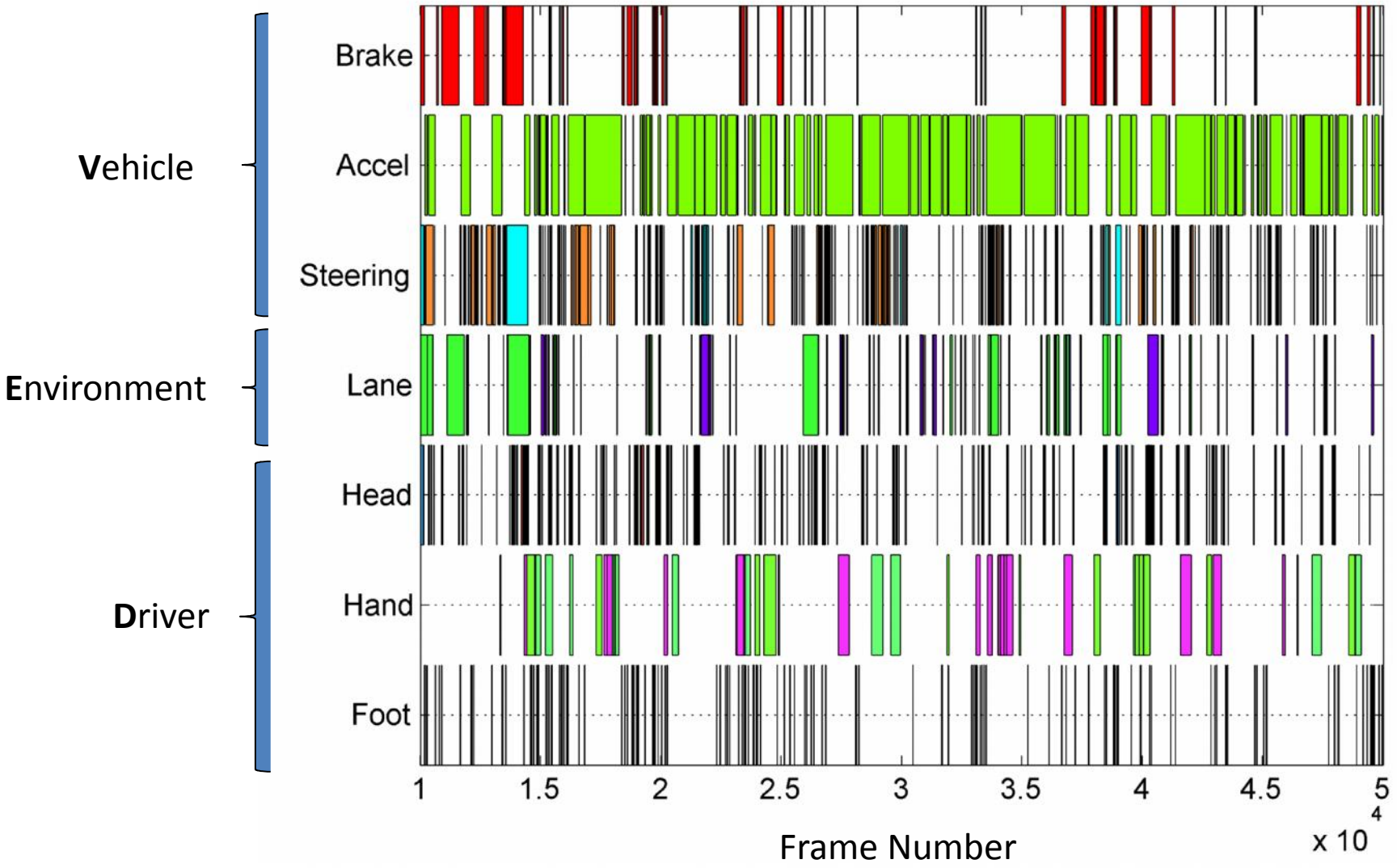
Vehicle Parameters.
- Acceleration, brake, steering

Vision inside the vehicle:
- Two cameras for head pose tracking under large head movement
- Hand camera
- Foot camera

Vision outside the vehicle:
- Lane camera
- Two lidar and two radar for surround perception

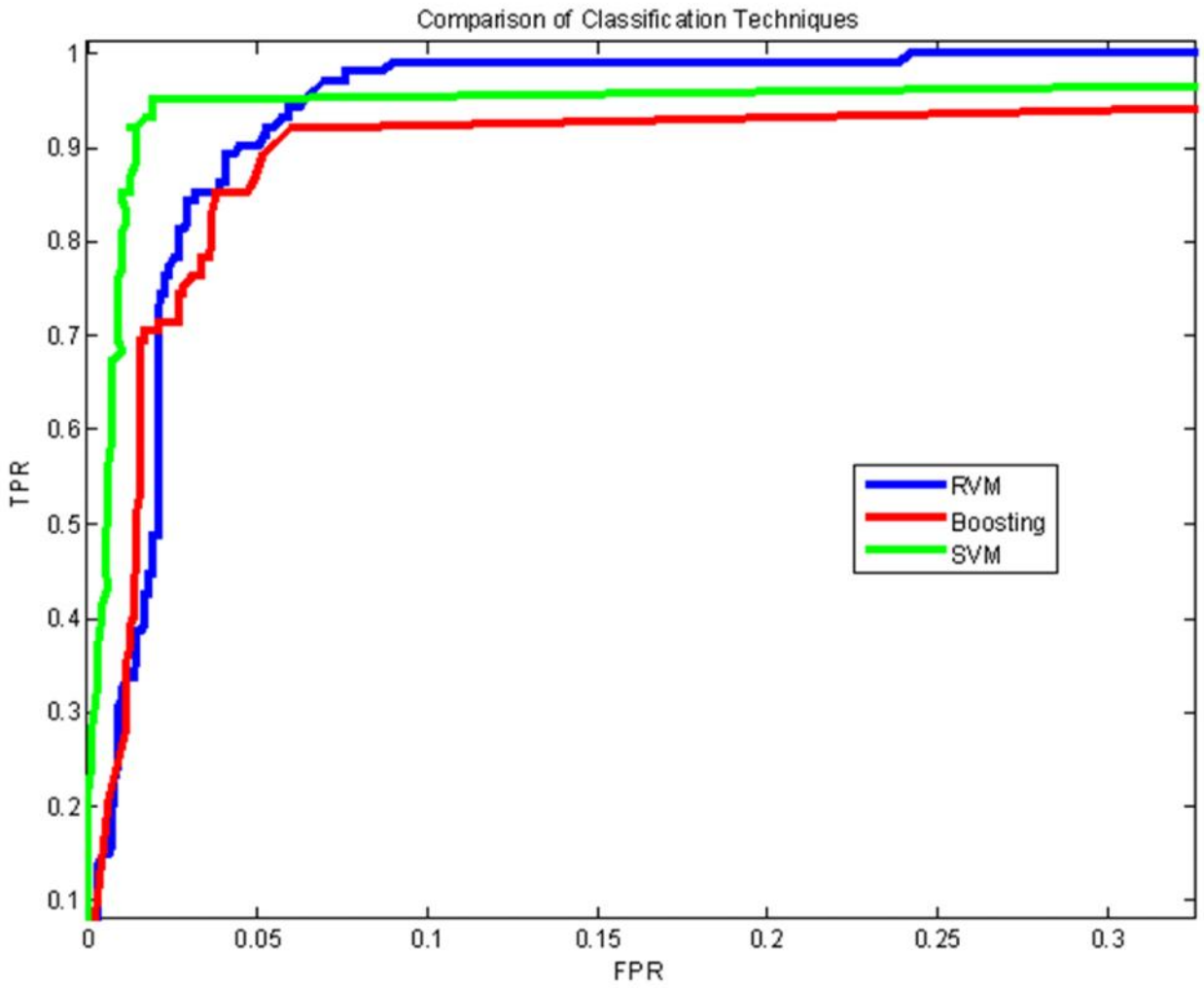
Ohn-Bar, Tawari, Martin, Trivedi, CVPRW 2014, CVIU, 2015

Overview of the Cues



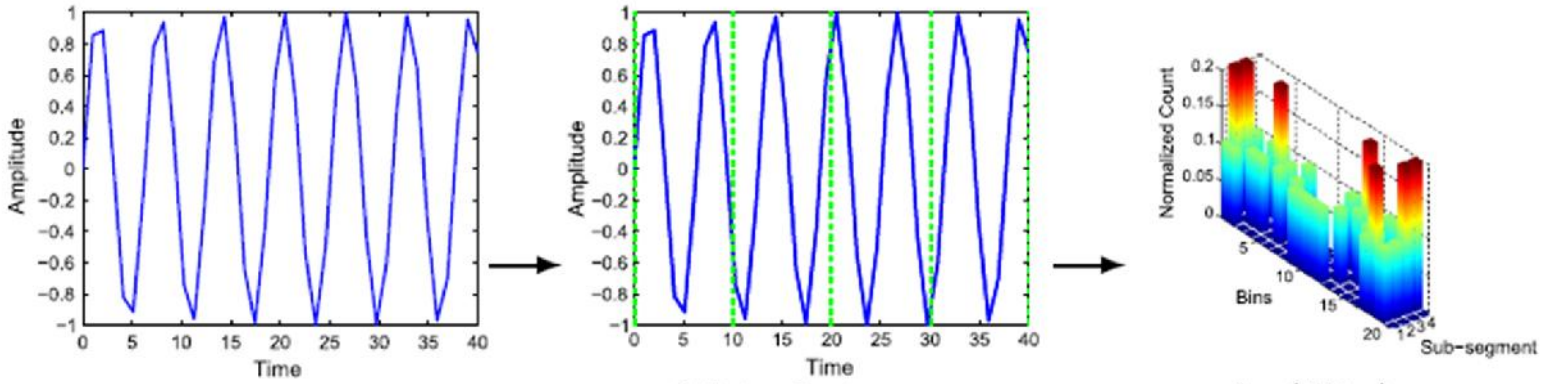
Modeling On-Road Maneuvers

- Comparison for lane change prediction
- Used RVM previously
- No fusion modeling
- No temporal state model



Doshi, Morris, and Trivedi, 2012

Multiple Kernel Learning for Cue-Fusion



Latent-Dynamic Conditional Random Field

MKL

$$\mathbf{K}(\mathbf{p}) = \sum_{l=1}^{s} p^l \mathbf{K}^{c_l}$$

$$\mathbf{p} = (p^1, \ldots, p^s), \text{ with } p^l \in \mathbb{R}_+ \text{ and } \mathbf{p}^T \mathbf{1} = 1$$

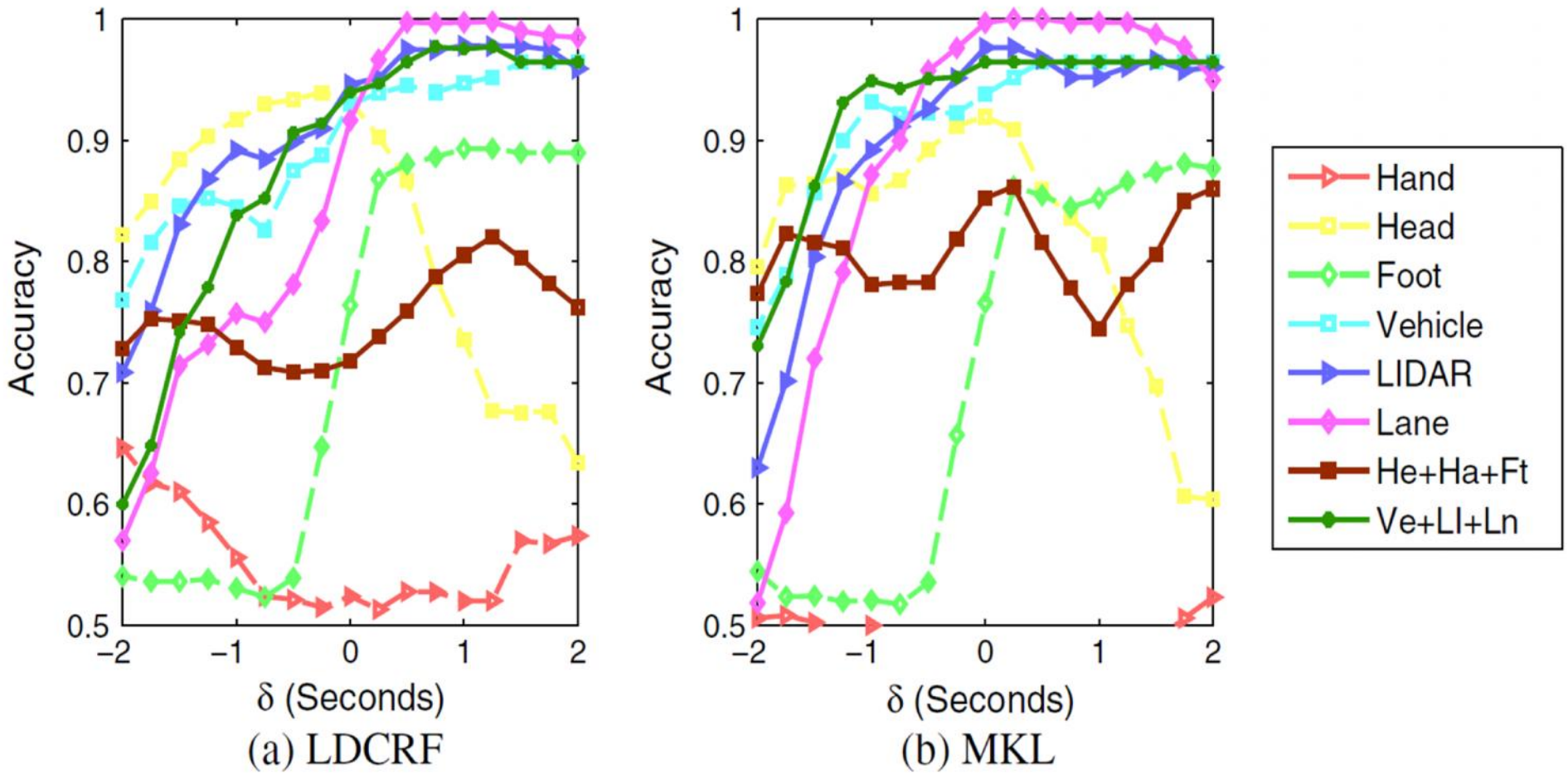
$$\{\mathbf{K}^{c_l} \in \mathbb{R}^n \times \mathbb{R}^n, l = 1, \ldots, s\}$$

$$K_{ij}^{c_l} = \kappa(\mathbf{x}_i, \mathbf{x}_j) = \exp(-||\mathbf{x}_i - \mathbf{x}_j||/\gamma)$$

F. R. Bach, G. R. G. Lanckriet, and M. I. Jordan, ICML 2004

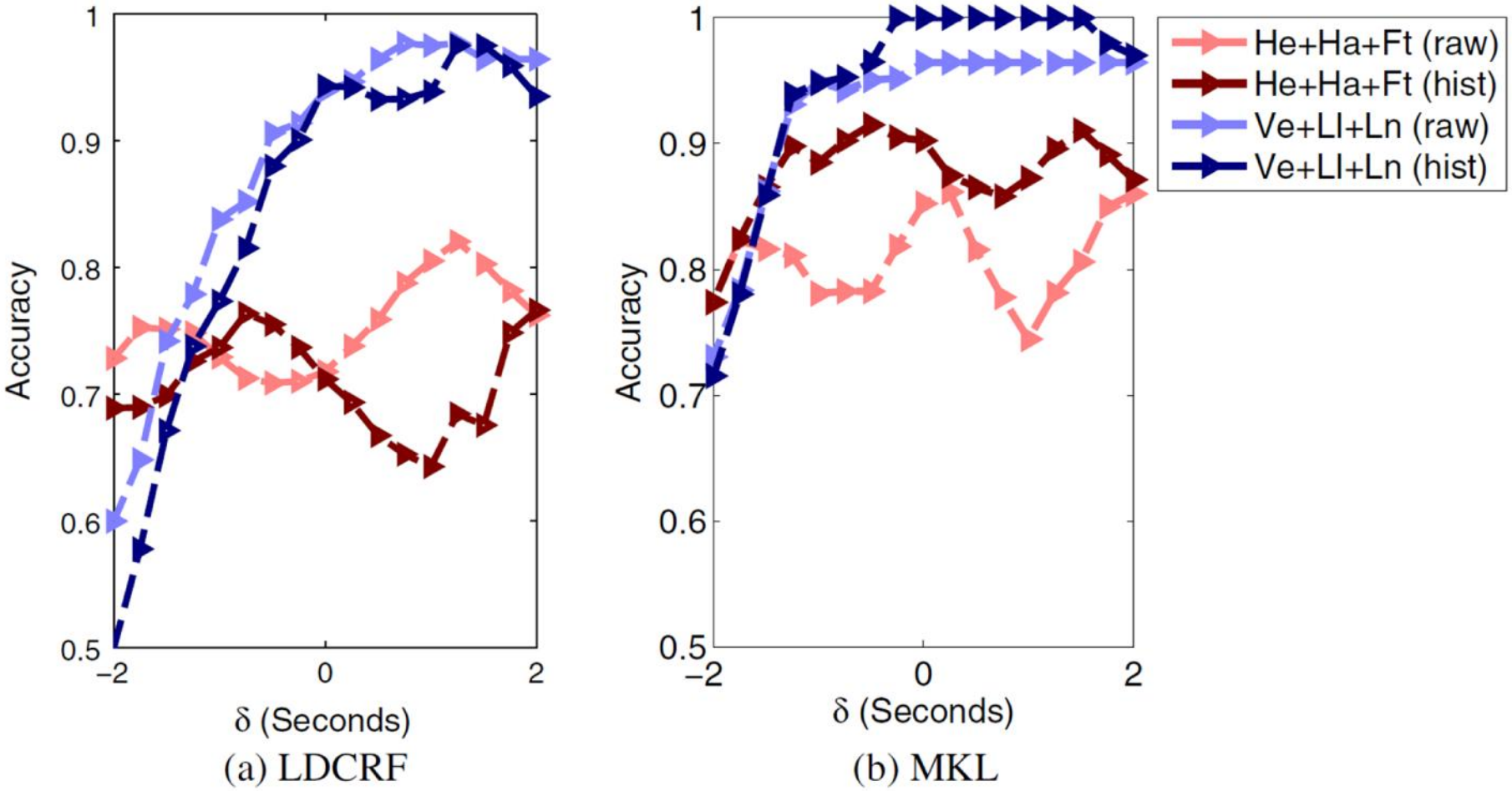
L. P. Morency, A. Quattoni, and T. Darrell,, CVPR 2007

Evaluation – Cue Analysis (Overtake vs. Brake)



- **Driver** cues are useful for **early prediction**
- MKL **fusion** significantly improves performance

Evaluation – Feature and Model Comparison



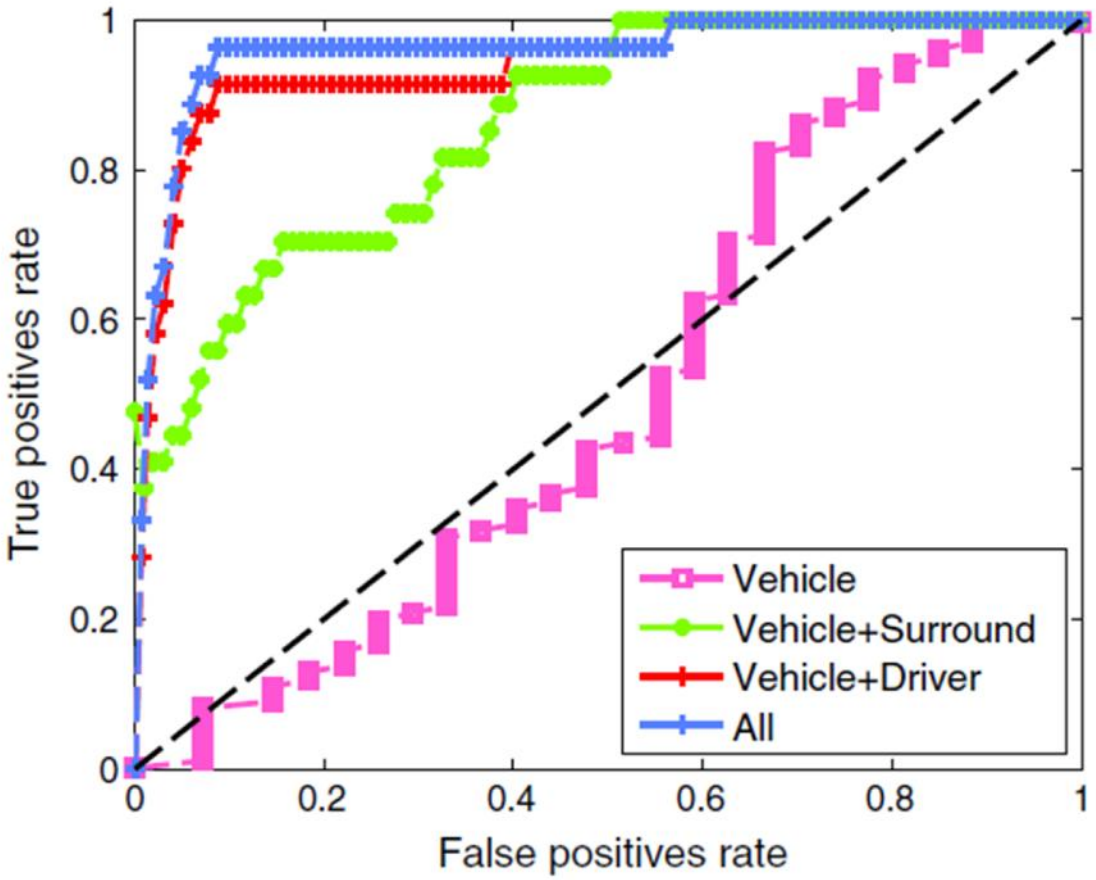
- Histogram features improved MKL model

Evaluation – Cue Analysis

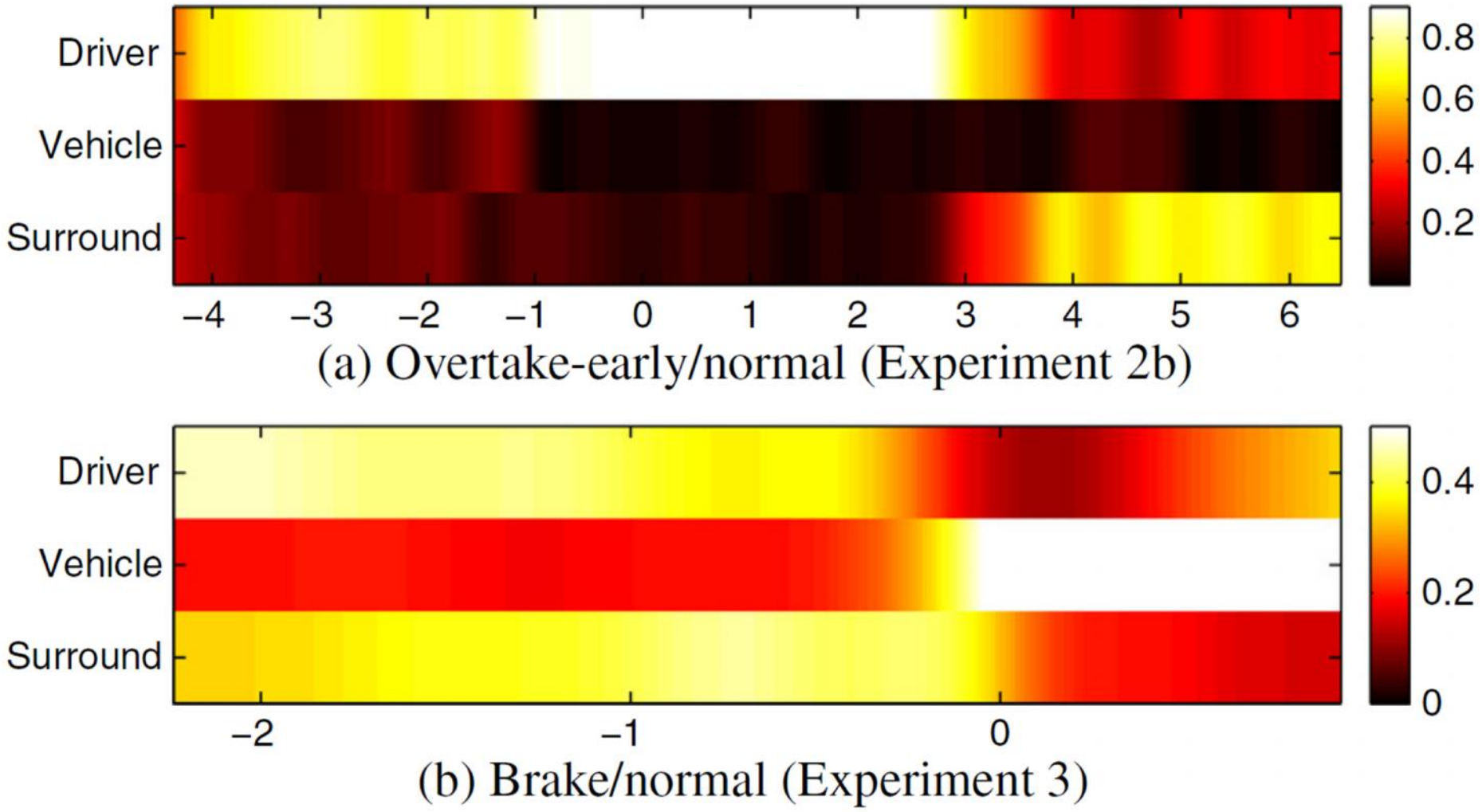
- For a fixed prediction time of **-2 seconds**, we show the marginal rate of return starting with vehicle dynamics

- Driver cues are head, hand, and foot

- Surround cues utilize surround vehicles and lane position



Evaluation – Cue Analysis



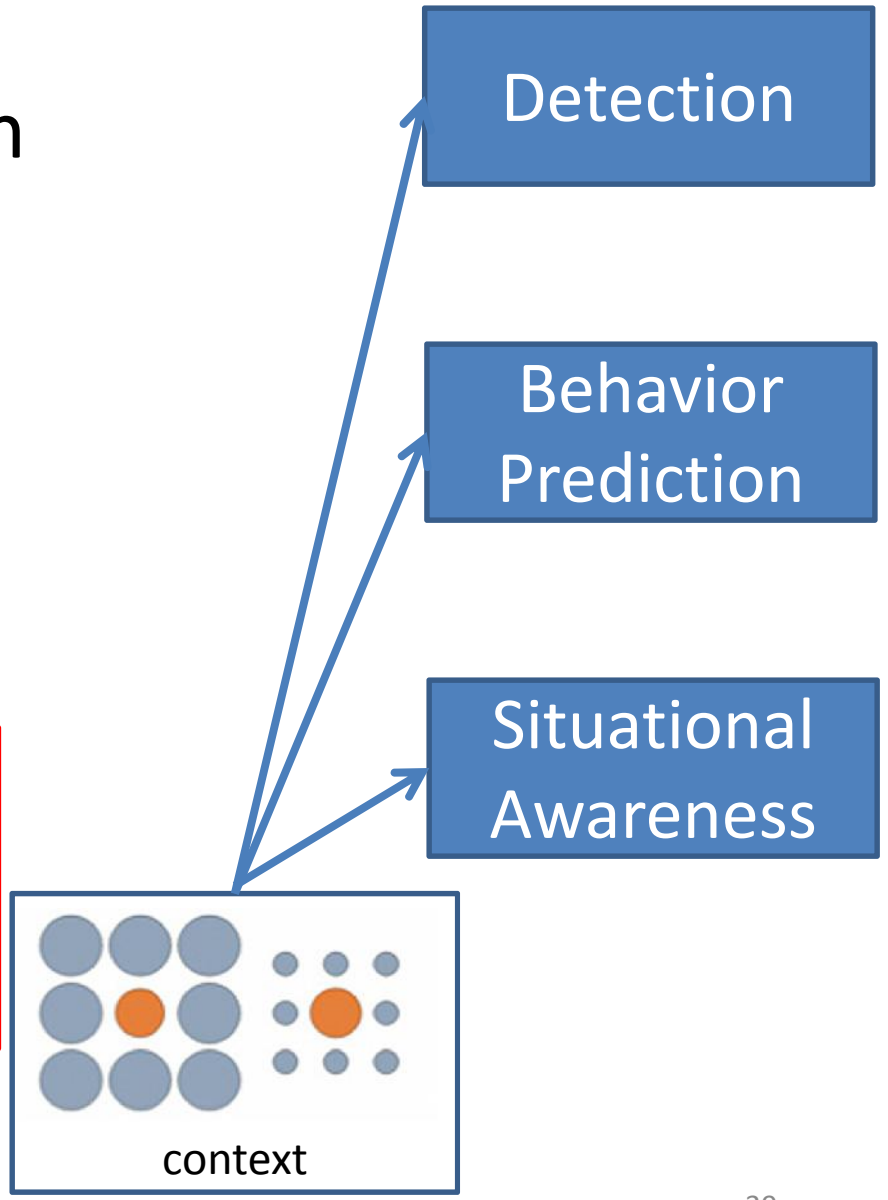
Learned kernel weight for each cue category (each column sums up to 1)

Research Objectives

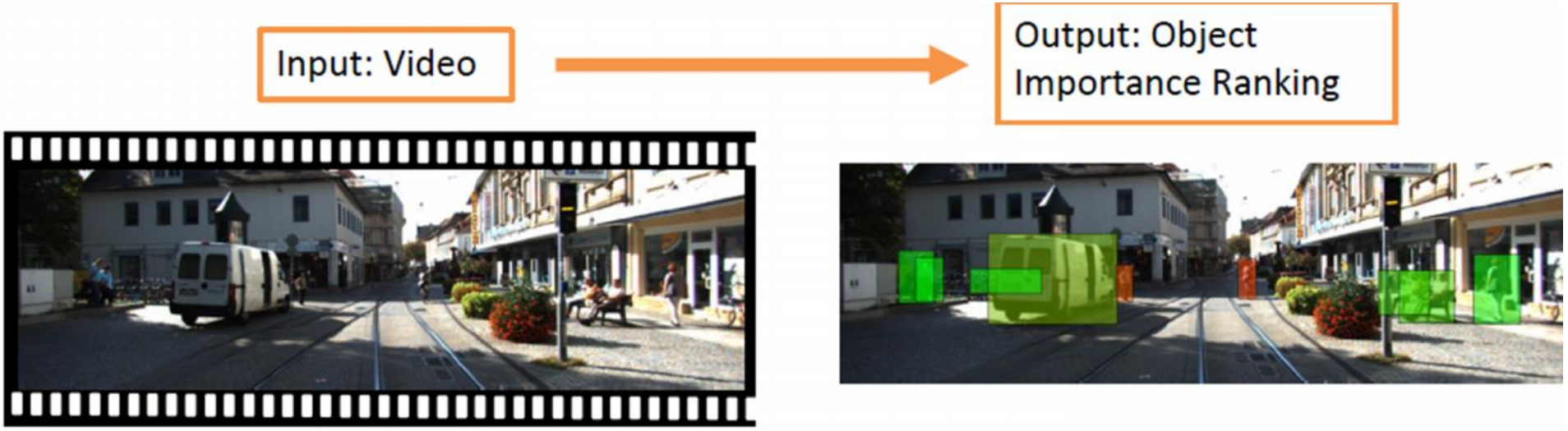
1) Contextual visual object **detection** and localization

2) Learning spatio-temporal dynamics for **behavior** prediction

3) **Situational awareness** and **human-centric** recognition in videos



Which of the Surrounding Agents Are Most Important?



- Analyze **subject variation**
- Various spatio-temporal and **contextual** cues
- Employ human-centric recognition metrics to study **dataset bias** and better develop models suitable for **safety-critical** applications

E. Ohn-Bar and M. Trivedi, ICPR 2016, Pattern Recognition 2017

Motivation Video

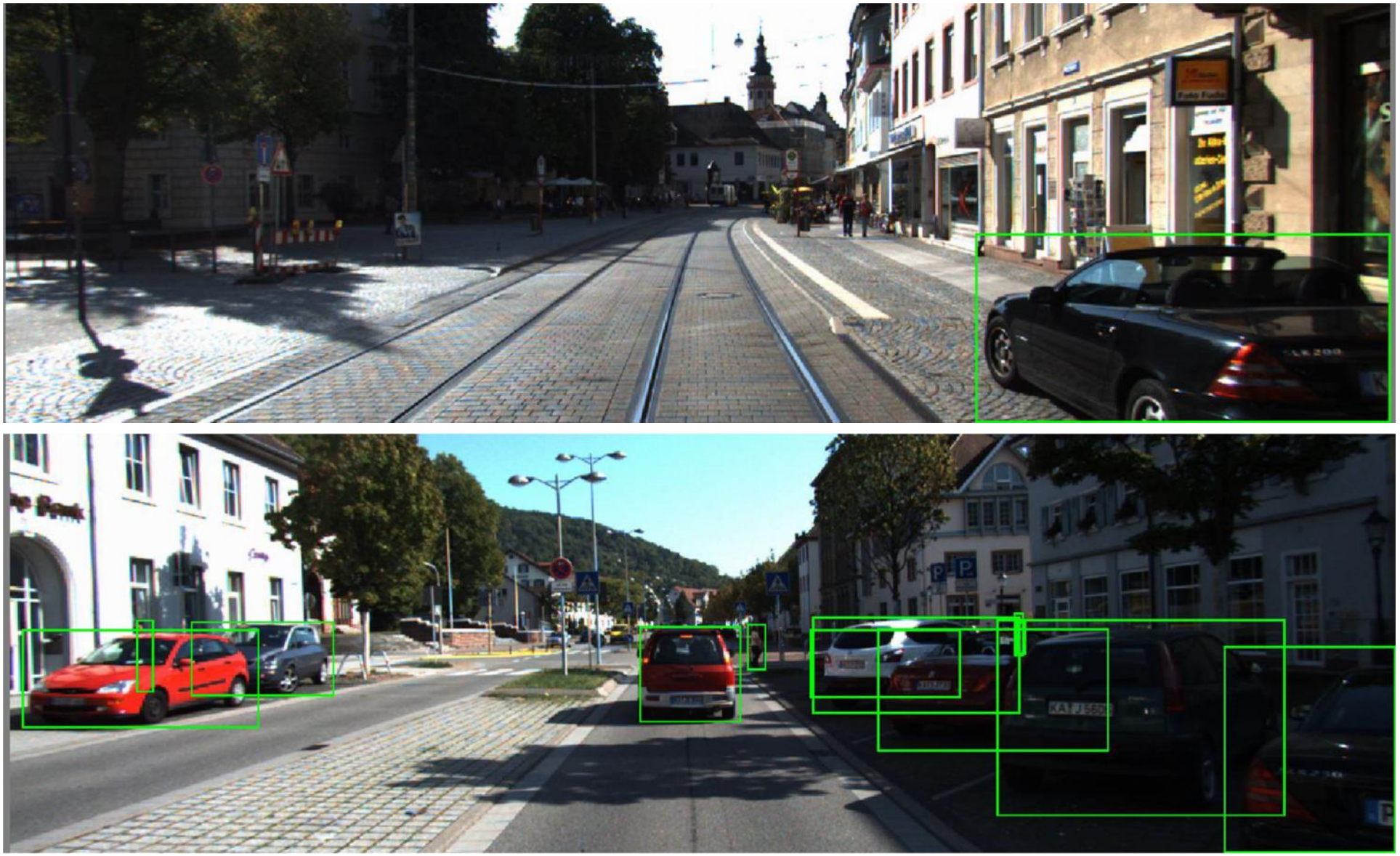
- **Safe** and **smooth** navigation in an intricate and uncertain environment (performed by human drivers continuously)

- **Complex reasoning** over the current driving task,
object properties, scene context, intent, and possible future actions.

- Robot and human are required to cooperate and trust. **Errors are costly.**

E. Ohn-Bar and M. Trivedi, ICPR 2016, Pattern Recognition 2017

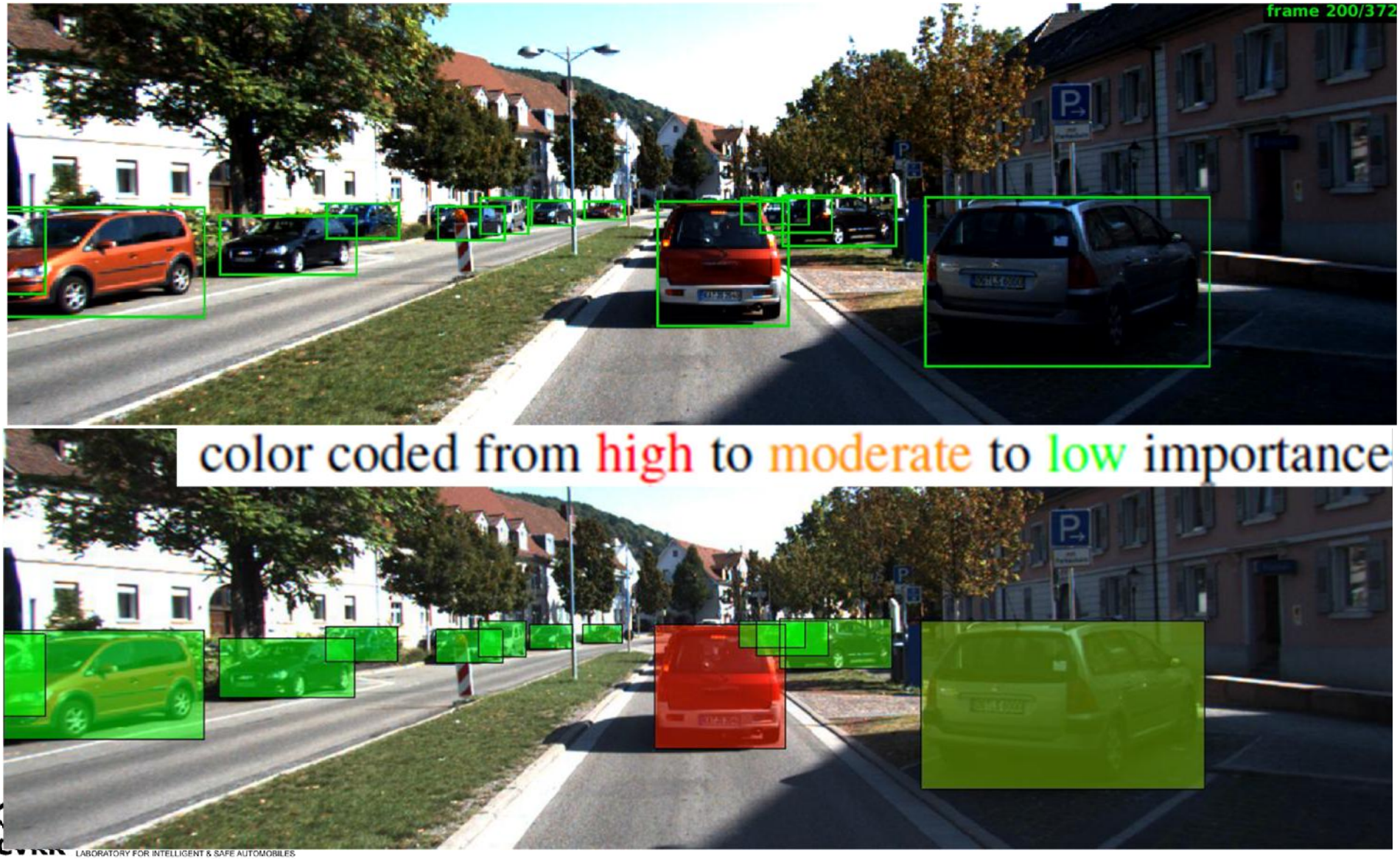
The Need for Useful Scene Representation



E. Ohn-Bar and M. Trivedi, ICPR 2016, Pattern Recognition 2017

Which of the surrounding agents are most relevant to driving task?

Which of the surrounding agents are most relevant to driving task?



Which of the surrounding agents are most important to you, as a driver?

Semantic Level

Spatio-Temporal Object Modeling

color coded from high to moderate to low importance

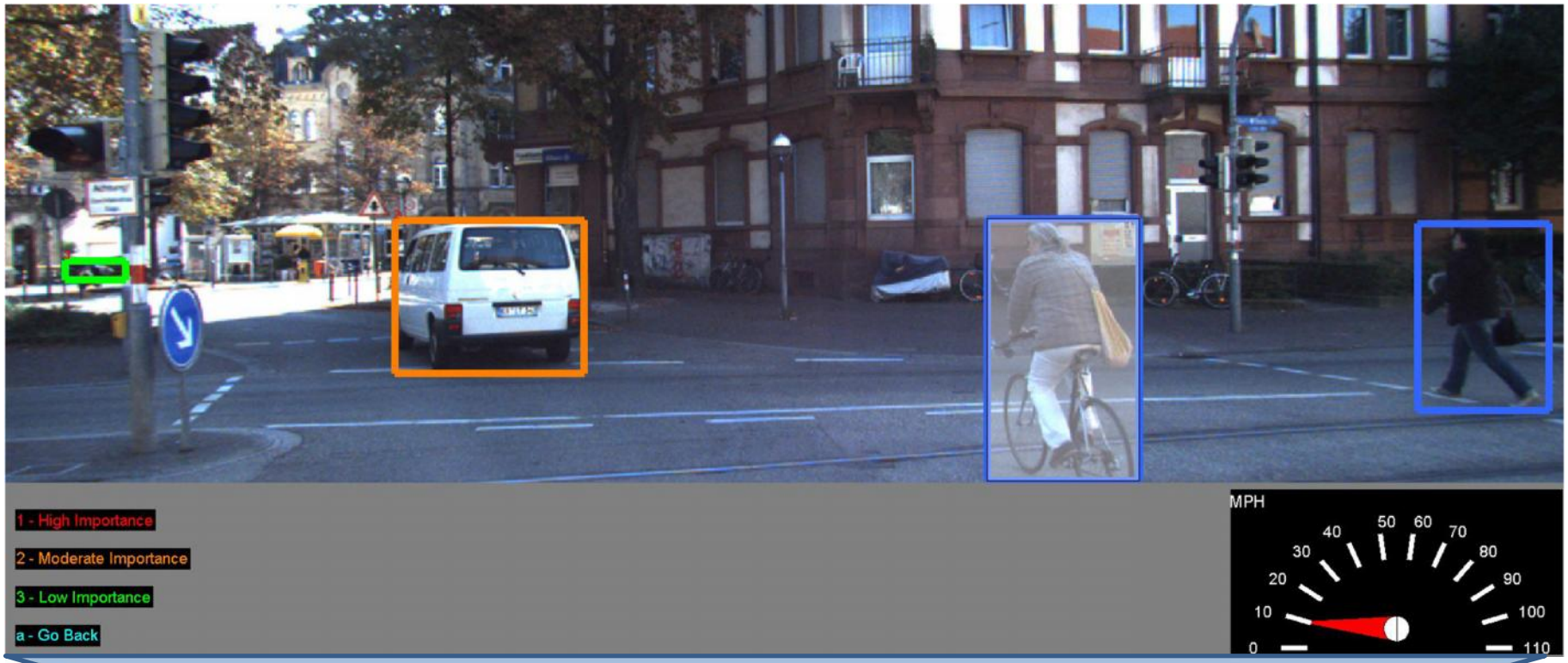
Object Detection

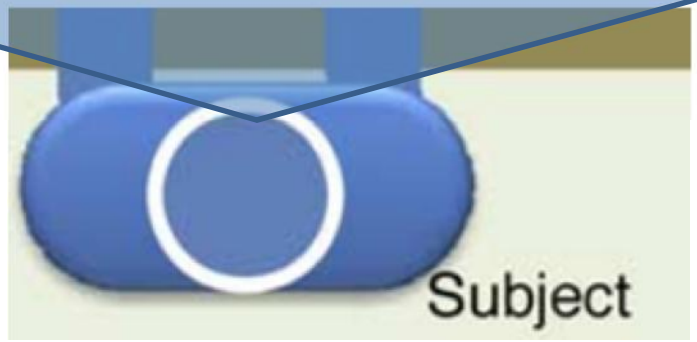
- Relevant agents for navigation
- Risk/threat estimation
- Surround behavior analysis
- Perception modeling
- Driver-specific modeling
- Agent activity and intent
- Event Prediction
- Situational Awareness

Spatio-temporal context will be shown to be crucial!

E. Ohn-Bar and M. Trivedi, ICPR 2016, Pattern Recognition 2017

Annotation Process

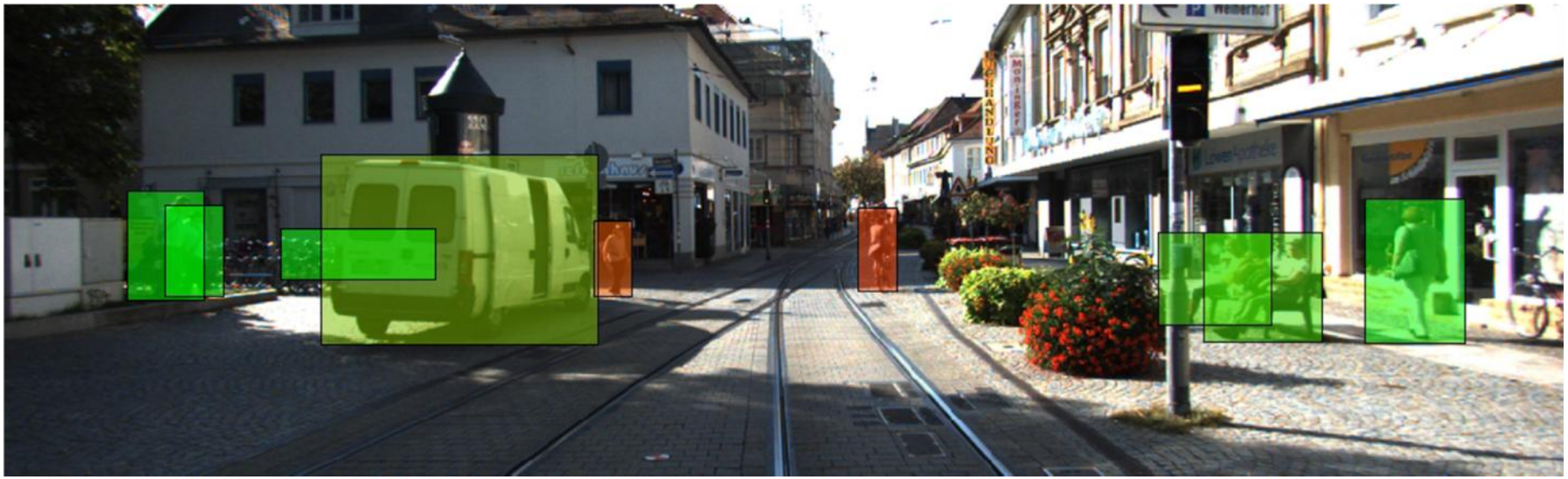


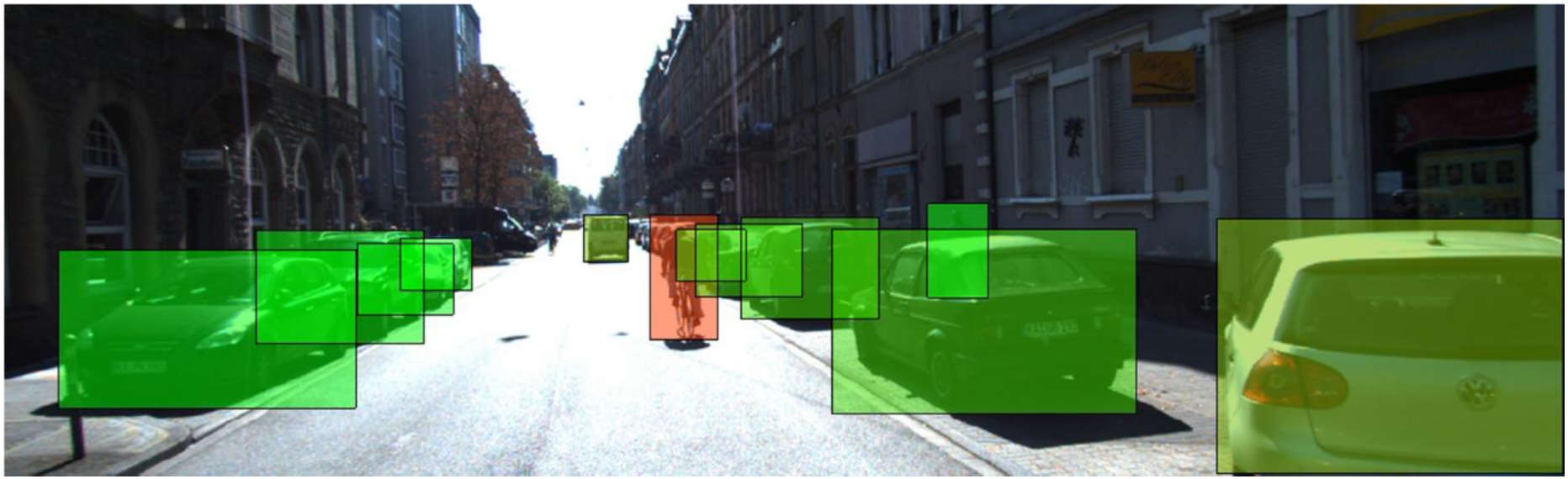


(18 subjects)

CVRR LISA UCSD LABORATORY FOR INTELLIGENT & SAFE AUTOMOBILES

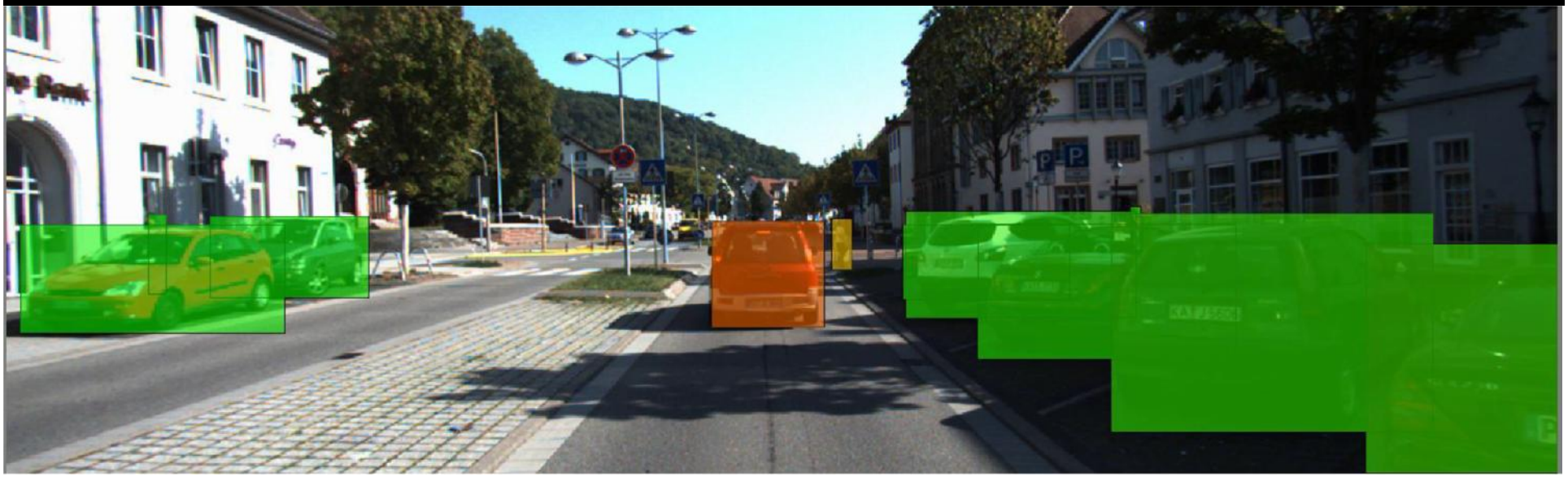
Annotation Visualization





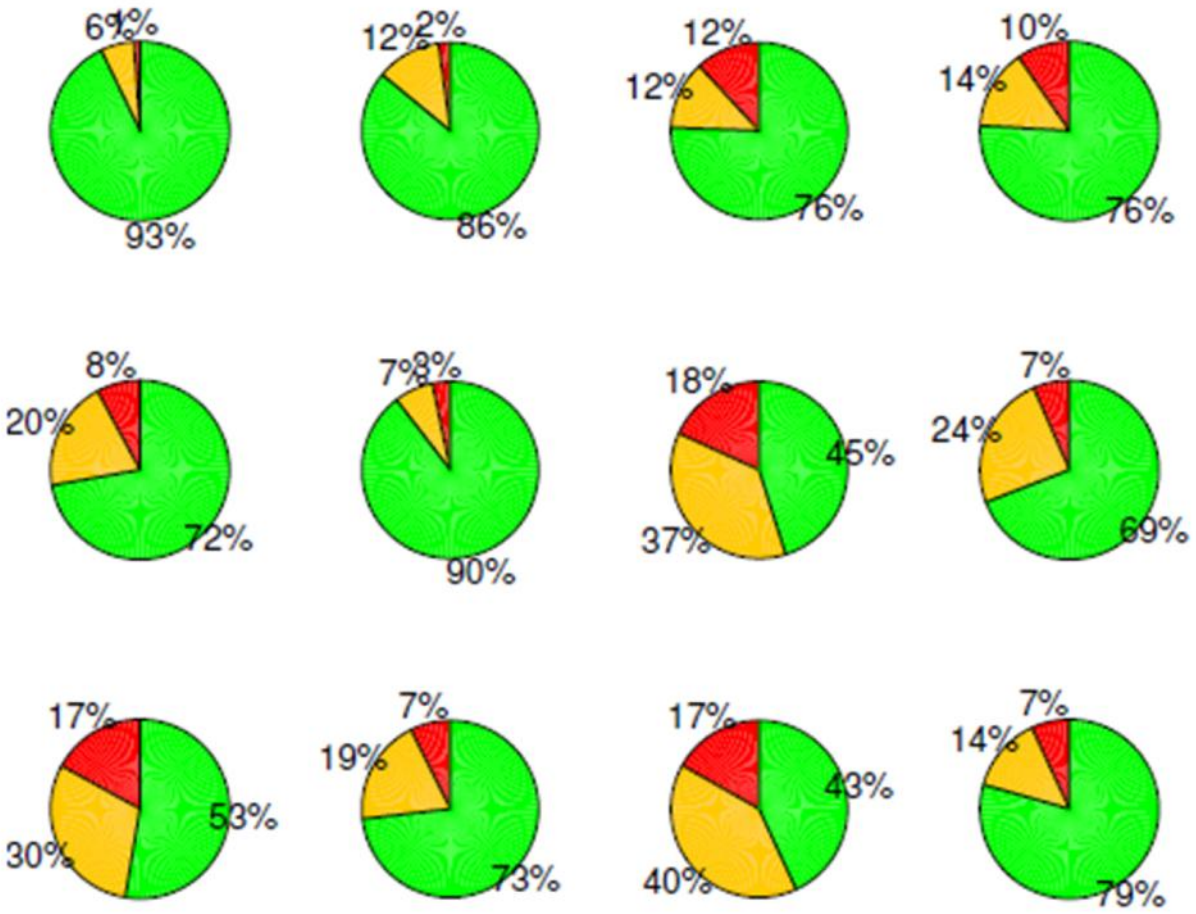
Annotation Visualization

Annotation Results



Annotation Distribution by Subject

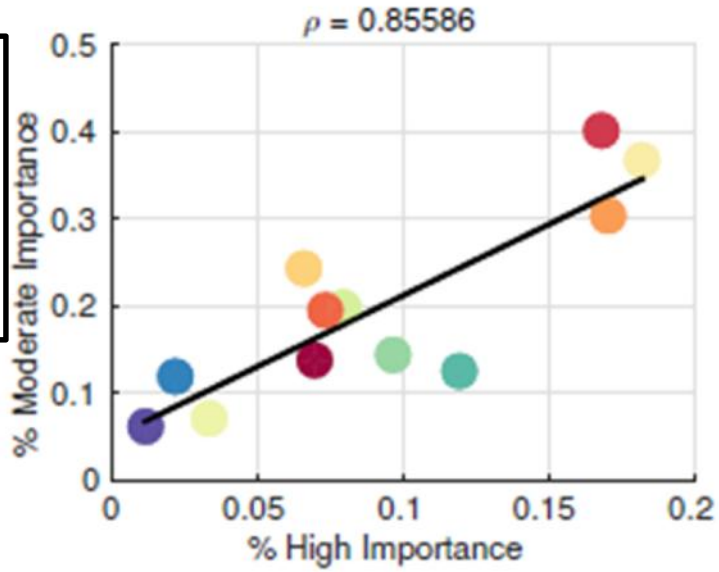
- **Each subject** is represented by a pie chart.
- **Majority** of the agents fall under `**low** importance' class
- Variation among subjects in high and moderate importance classes

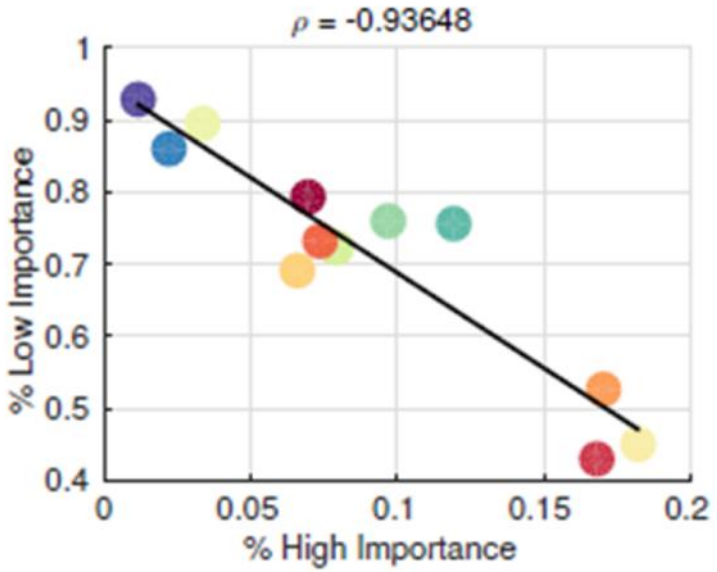


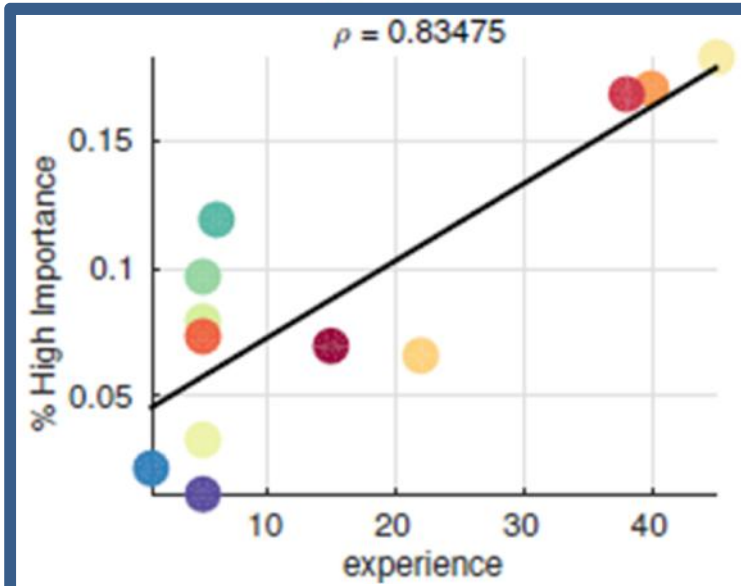
E. Ohn-Bar and M. Trivedi, ICPR 2016, Pattern Recognition 2017

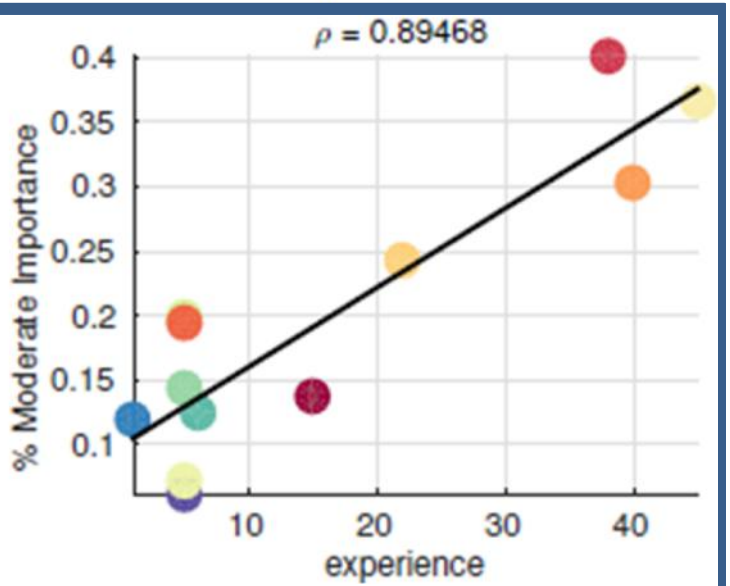
Annotation Distribution by Subject

Each subject is represented by a differently colored dot



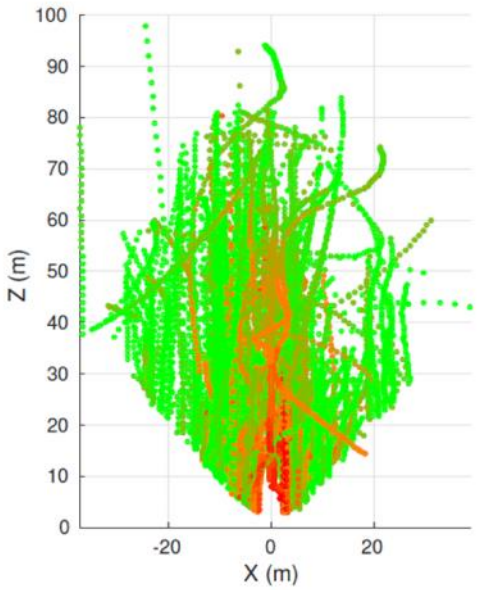




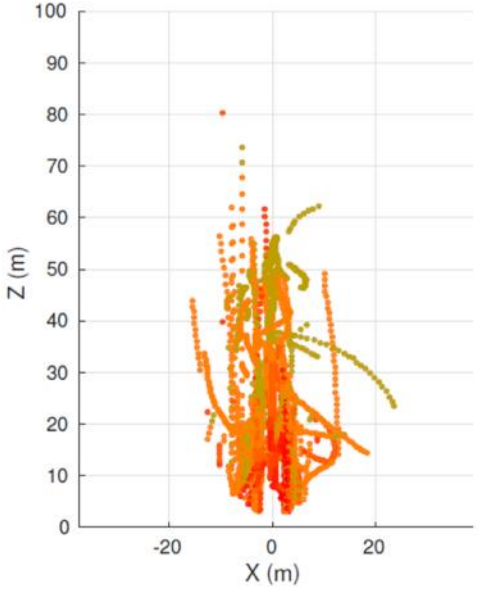


(experience in years, since driver license)

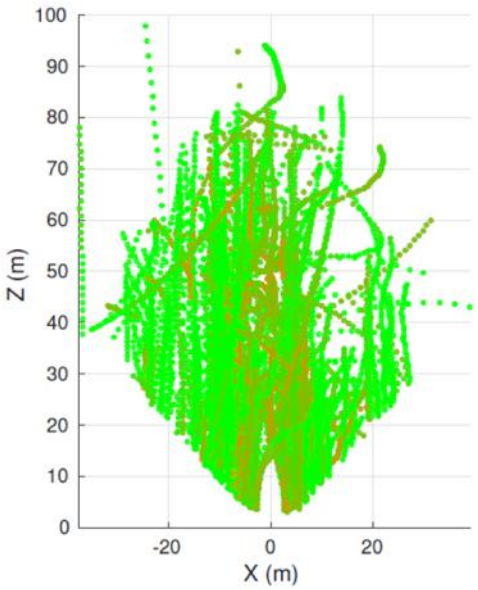
Annotation Distribution by Distance



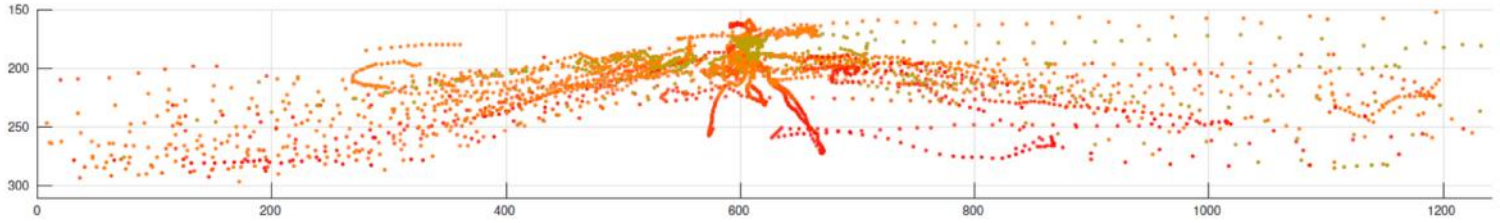
(a) All objects



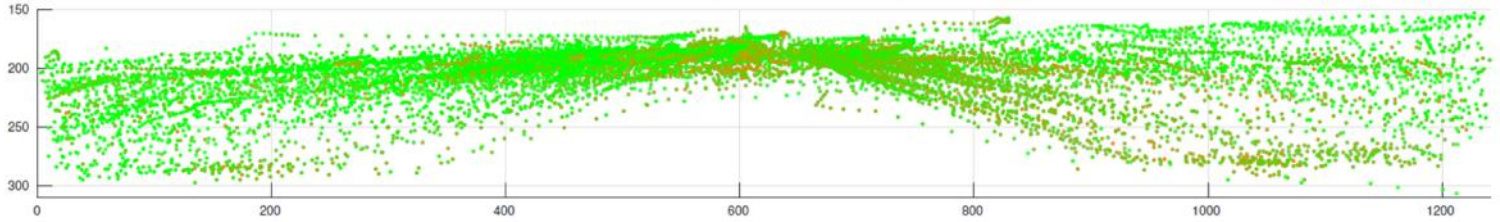
(b) High importance objects



(c) Low importance objects



(e) High importance objects



(f) Low importance objects

Explaining Attributes for Object Importance

$$M_{attributes}(s) = \boldsymbol{w}_{c,2D-obj}^{T}\phi_{2D-obj}(s) + \boldsymbol{w}_{c,3D-obj}^{T}\phi_{3D-obj}(s) + \boldsymbol{w}_{c,ego}^{T}\phi_{ego}(s) + \boldsymbol{w}_{c,temporal}^{T}\phi_{temporal}(s)$$

$\phi_{2D-obj} \in \mathbb{R}^4$ - > Height, aspect ratio, occlusion state, truncation %

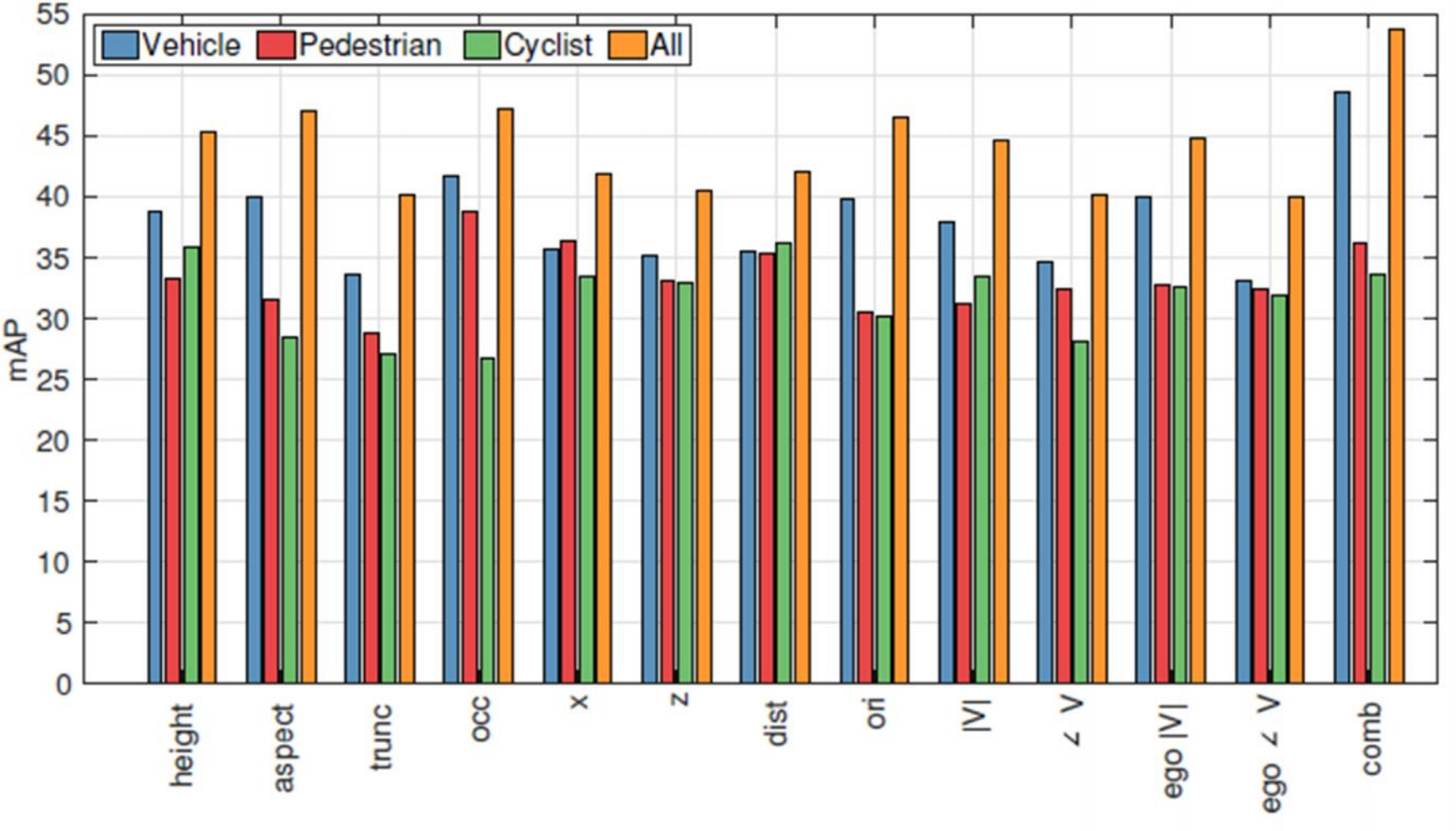
$\phi_{3D-obj} \in \mathbb{R}^6$ - > (lat.,long.), distance, orientation, object velocity

$\in \mathbb{R}$ - > Ego-vehicle velocity magnitude and orientation

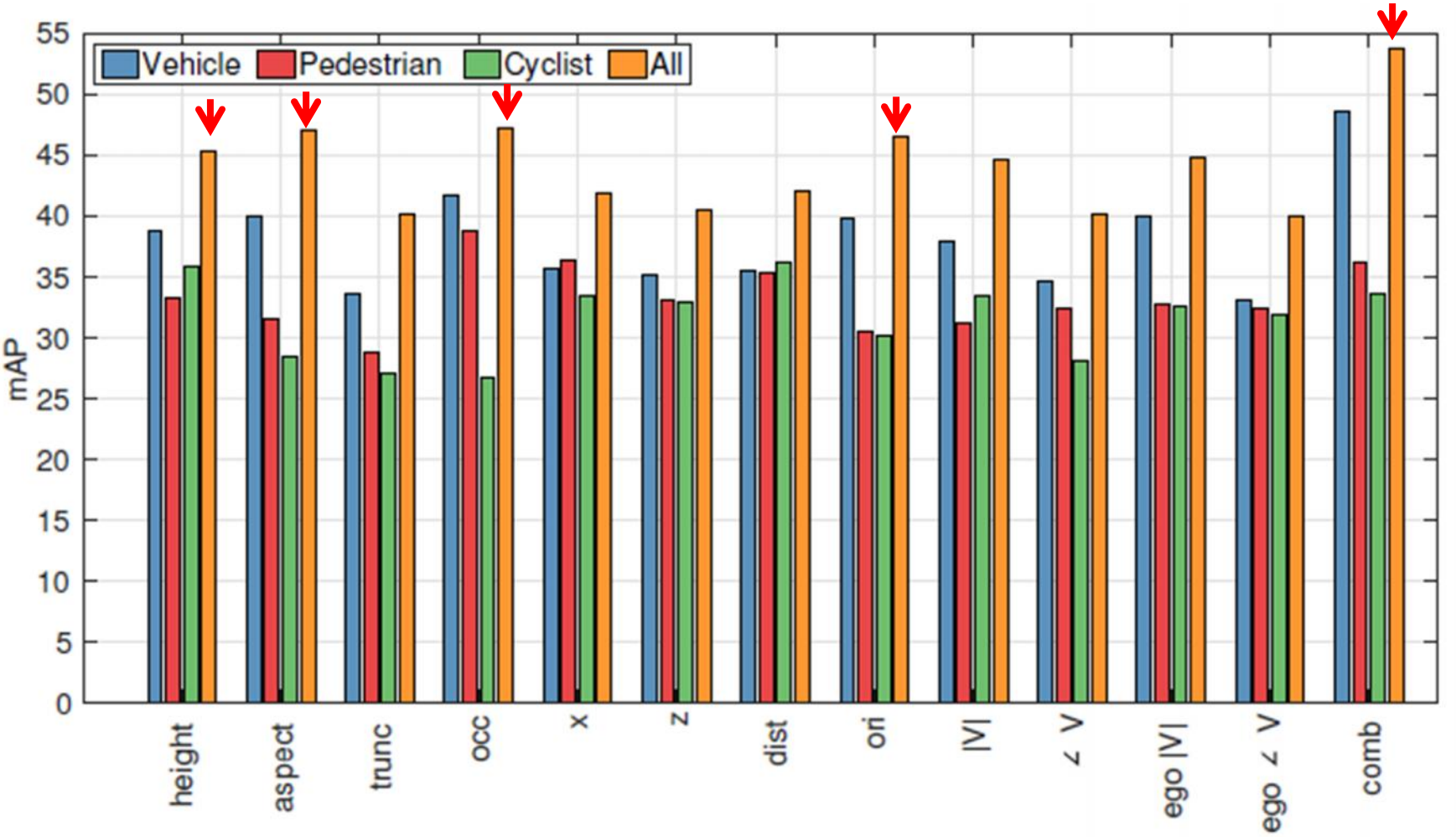
- > Concatenation, max-pooling, Discrete Cosine coefficients

Weights learned with logistic regression

Explaining Object Importance

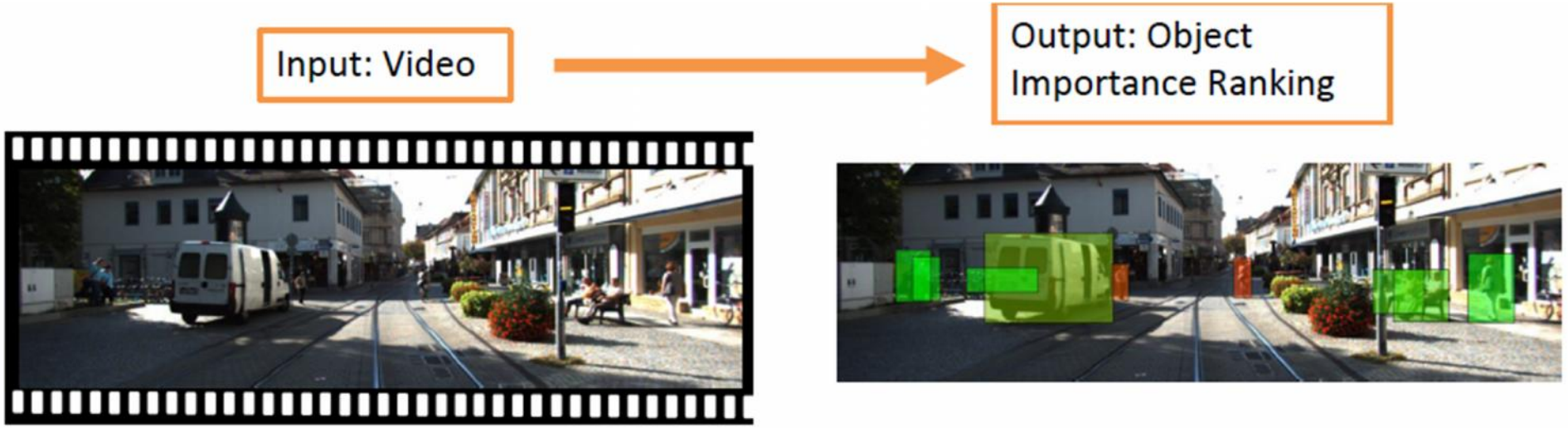


Explaining Object Importance



- **Occlusion**, **orientation**, and **size** are strong predictors.
- **Combination** of all the cues provides significant **improvement**

Visual Prediction Model



1) Overarching goal is to perform **visual prediction** – more challenging, no annotation/LIDAR

2) May include additional **object**- and **scene**-level **cues**

Visual Prediction Model

Input: Video

Output: Object Importance Ranking

$$M_{visual}(s) = \boldsymbol{w}_{c,obj}^T \phi_{obj}(s) + \boldsymbol{w}_{c,spatial}^T \phi_{spatial}(s) + \boldsymbol{w}_{c,temporal}^T \phi_{temporal}(s)$$

- > Local object features

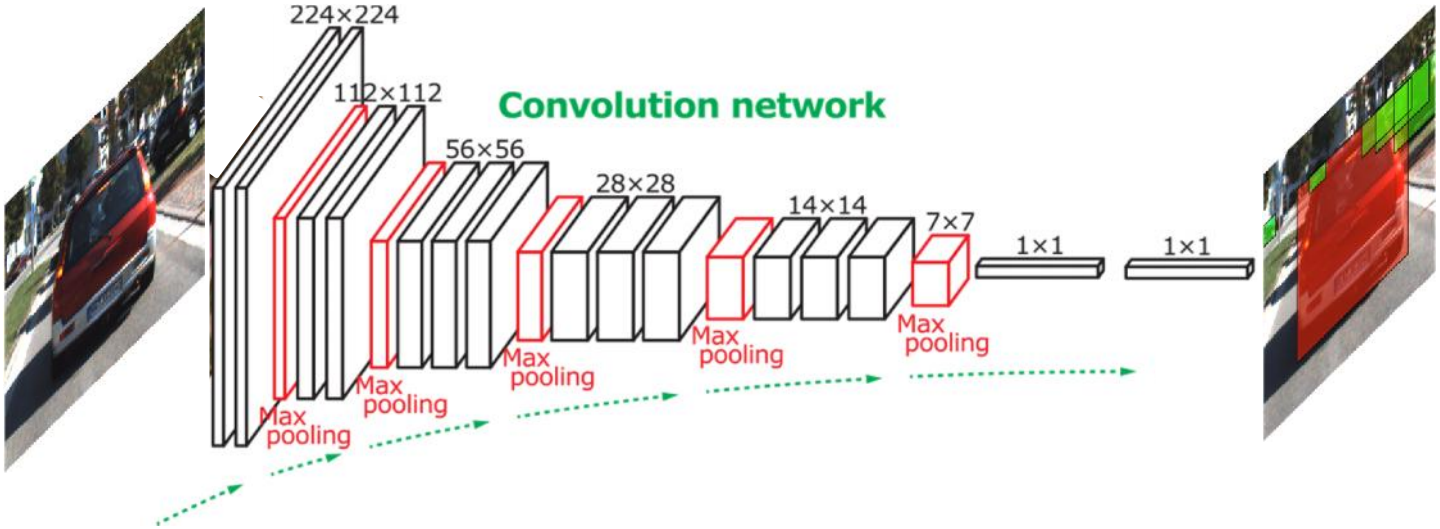
- > Spatial context features

- > Temporal context features (concatenation, max-pooled over time)

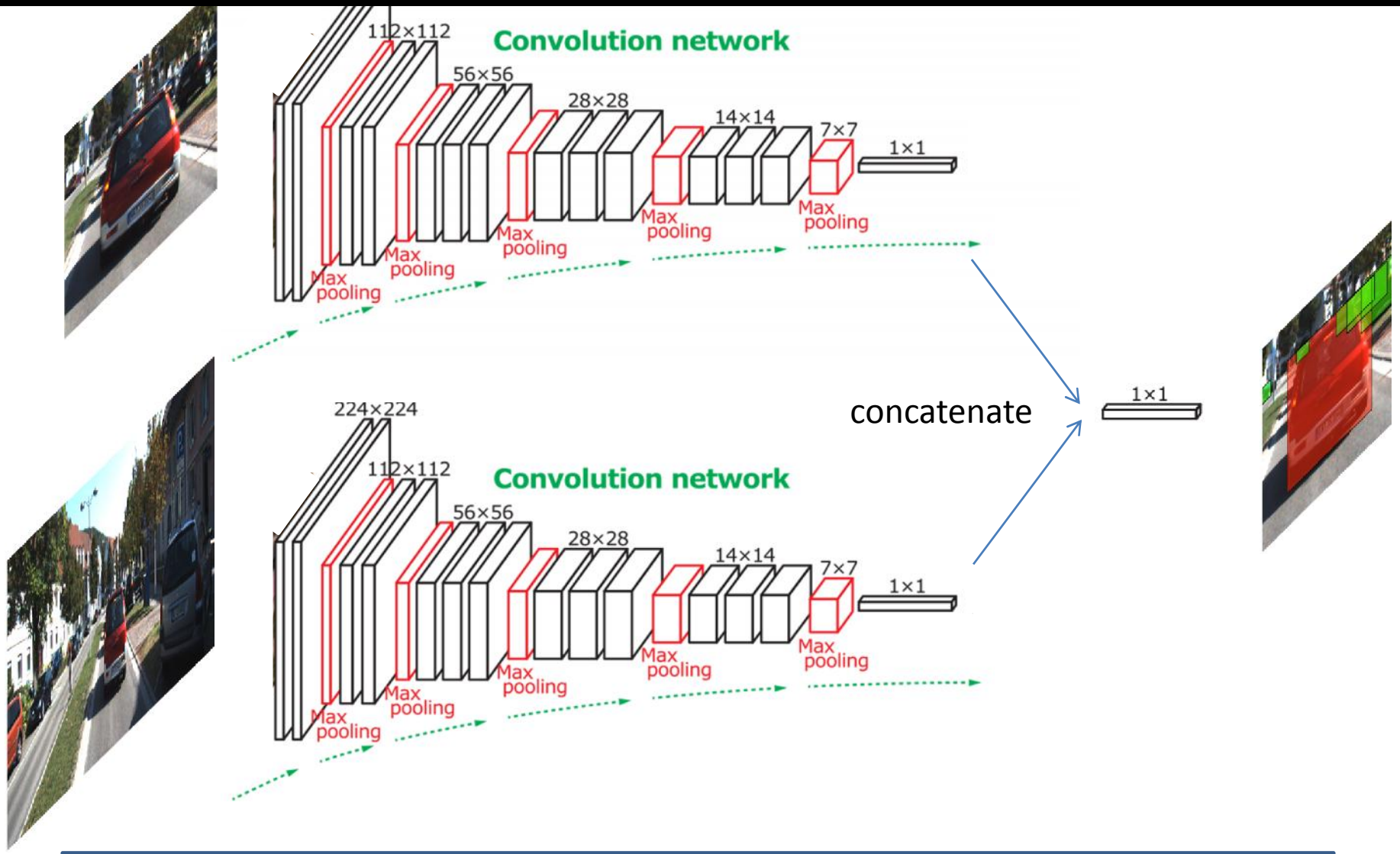
CVRR LISA UCSD LABORATORY FOR INTELLIGENT & SAFE AUTOMOBILES

Object-level Cues Term

- > Local object features model

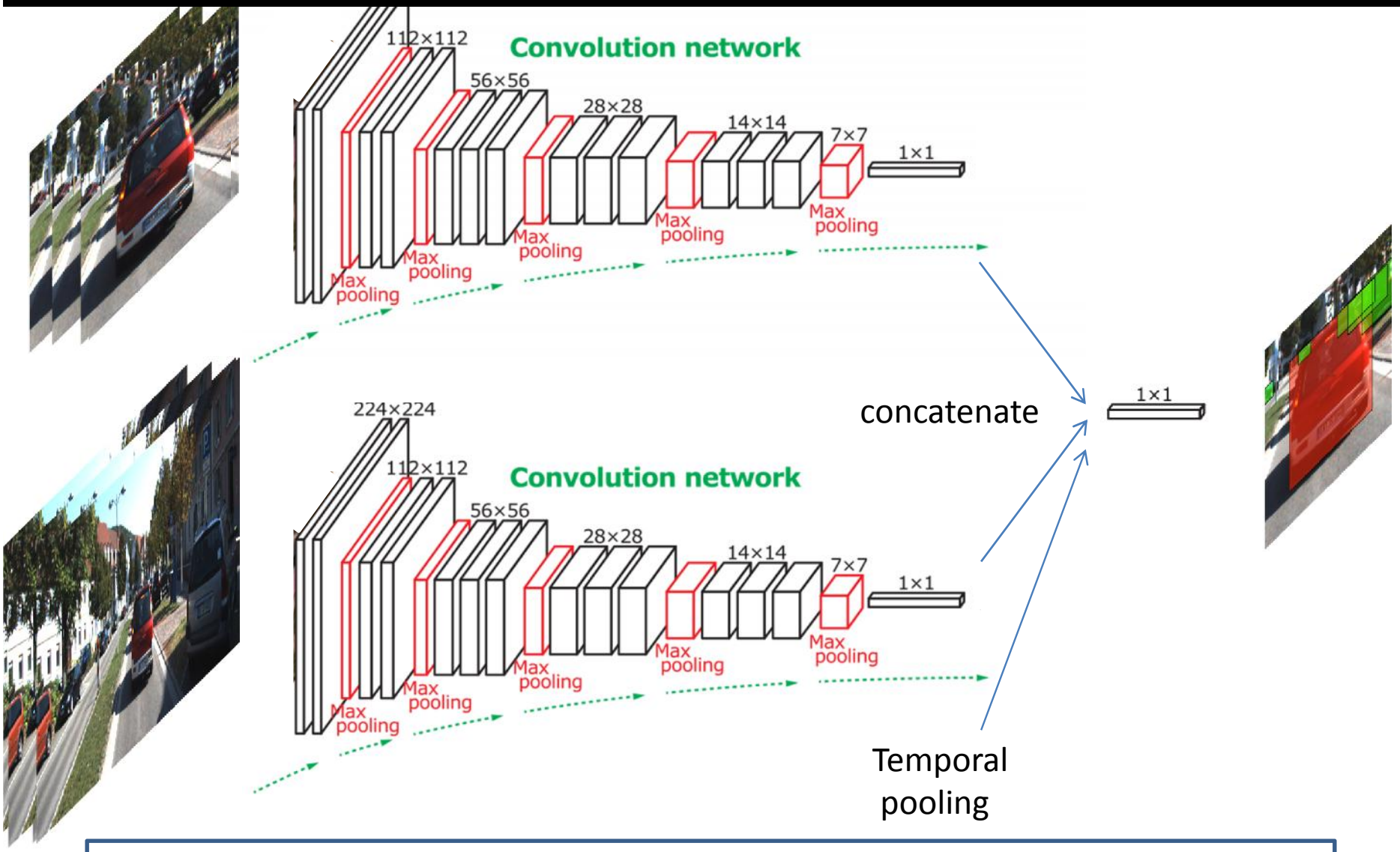


Model with a Spatial Context Term



\+ Model with spatial context

Full Spatio-Temporal Visual Prediction Model



\+ \+ Model with spatio-temporal context

Cue Analysis

Model	mAP (%)	MAE	$\text{MAE}_{\gamma=2.25}$
$M_{visual}(\phi_{obj})$	51.06	0.2648	0.5392
$M_{visual}(\phi_{obj} + \phi_{spatial})$	55.53	0.2611	0.5007
$M_{visual}(\phi_{obj} + \phi_{temporal})$	53.30	0.2507	0.4765
$M_{visual}(\phi_{obj} + \phi_{spatial} + \phi_{temporal})$	56.34	0.2447	0.4625

Cue Analysis

Model	mAP (%)	MAE	$MAE_{\gamma=2.25}$
$M_{visual}(\phi_{obj})$	51.06	0.2648	0.5392
$M_{visual}(\phi_{obj} + \phi_{spatial})$	55.53	0.2611	0.5007
$M_{visual}(\phi_{obj} + \phi_{temporal})$	53.30	0.2507	0.4765
$M_{visual}(\phi_{obj} + \phi_{spatial} + \phi_{temporal})$	56.34	0.2447	0.4625

Cue Analysis

Model	mAP (%)	MAE	$\text{MAE}_{\gamma=2.25}$
$M_{visual}(\phi_{obj})$	51.06	0.2648	0.5392
$M_{visual}(\phi_{obj} + \phi_{spatial})$	55.53	0.2611	0.5007
$M_{visual}(\phi_{obj} + \phi_{temporal})$	53.30	0.2507	0.4765
$M_{visual}(\phi_{obj} + \phi_{spatial} + \phi_{temporal})$	56.34	0.2447	0.4625

$M_{attributes}$ (without $\phi_{temporal}$)	53.70	0.2440	0.3853
$M_{attributes}$ (with $\phi_{temporal}$)	**60.35**	**0.2148**	**0.2914**

Cue Analysis

Model	mAP (%)	MAE	$MAE_{\gamma=2.25}$
$M_{visual}(\phi_{obj})$	51.06	0.2648	0.5392
$M_{visual}(\phi_{obj} + \phi_{spatial})$	55.53	0.2611	0.5007
$M_{visual}(\phi_{obj} + \phi_{temporal})$	53.30	0.2507	0.4765
$M_{visual}(\phi_{obj} + \phi_{spatial} + \phi_{temporal})$	56.34	0.2447	0.4625

Temporal cues are **essential** for importance prediction, in particular on objects of higher importance

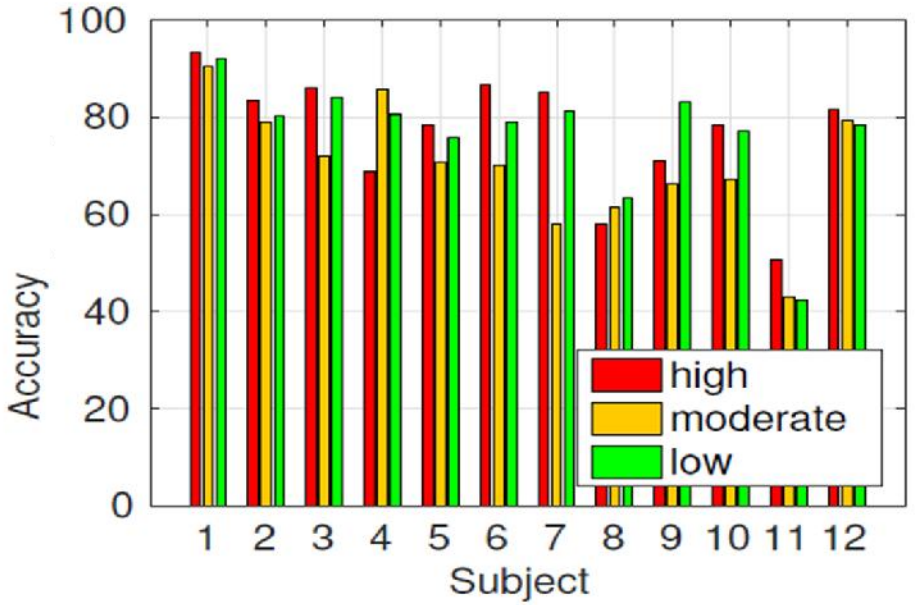
$M_{attributes}$ (without $\phi_{temporal}$)	53.70	0.2440	0.3853
$M_{attributes}$ (with $\phi_{temporal}$)	**60.35**	**0.2148**	**0.2914**

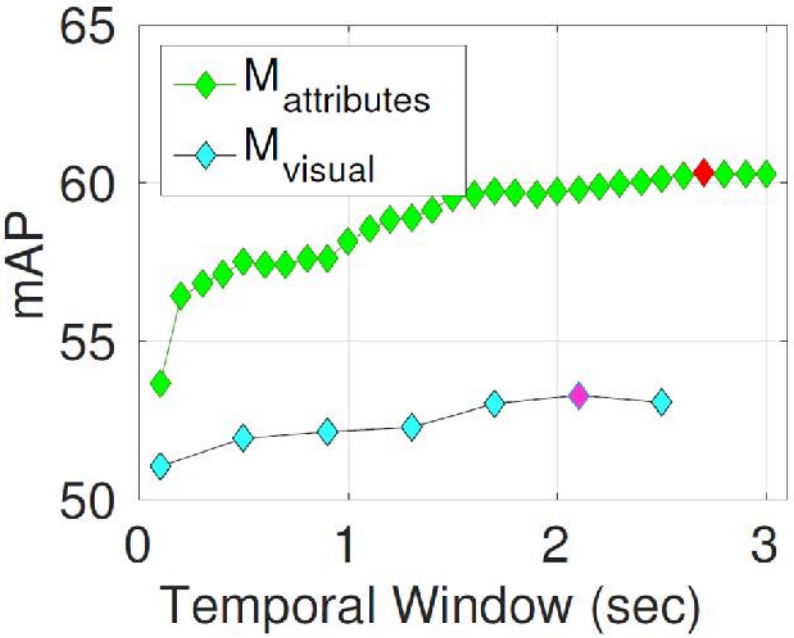
Higher is better

Lower is better

Subject-specific Attributes Model

- Training/testing within each subject for each importance class
- Measures **predictability** within the subject's annotations
- Temporal window of **2.7 seconds** is best



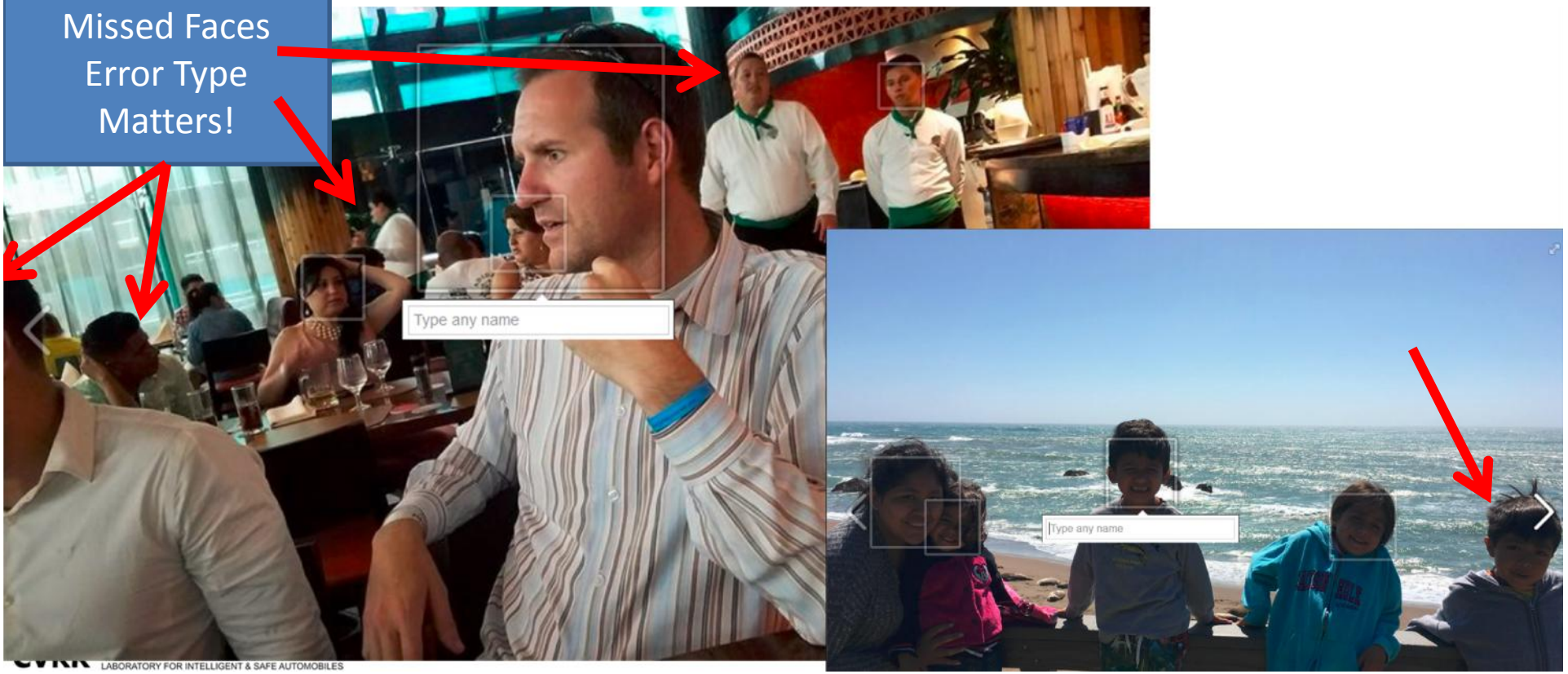


Let's Talk Metrics!

- Are all objects **equal**?

PASCAL, Caltech, KITTI **exclude** training/testing
difficult objects of small size, high truncation, heavy occlusion

- **Appropriate evaluation**? Under safety-critical events, these may be most relevant.
- **Dataset bias** may hinder insights, and impact training



Importance-Guided Training and Evaluation of Object Detectors

Can we train models better at detecting objects of high importance?

Cost-sensitive training loss function:

$$L(p, u, \gamma, t^u, v) = L_{cls}^{IG}(p, u, \gamma) + \lambda_{loc}[u \geq 1]L_{loc}(t^u, v)$$

Importance-Guided Training and Evaluation of Object Detectors

Can we train models better at detecting objects of high importance?

Importance Guided (IG) Classification Log Loss

Cost-sensitive training loss function:

$$L(p, u, \gamma, t^u, v) = L_{cls}^{IG}(p, u, \gamma) + \lambda_{loc}[u \geq 1] L_{loc}(t^u, v)$$

Localization loss

Importance-Guided Training and Evaluation of Object Detectors

Can we train models better at detecting objects of high importance?

Cost-sensitive training loss function:

Importance Guided (IG) Classification Log Loss

$$L(p, u, \gamma, t^u, v) = L_{cls}^{IG}(p, u, \gamma) + \lambda_{loc}[u \geq 1] L_{loc}(t^u, v)$$

Localization loss

$$L_{cls}^{IG}(p, u, \gamma) = -\alpha_\gamma \log p_u$$

Sample's average importance ranking annotation score

$$\alpha_\gamma = \begin{cases} \lambda & \gamma \leq 2.25 \\ 1/\lambda & \text{otherwise} \end{cases}$$

Importance-Guided Training and Evaluation of Object Detectors

Can we train models better at detecting objects of high importance?

Cost-sensitive training loss function:

Importance Guided (IG) Classification Log Loss

$$L(p, u, \gamma, t^u, v) = L_{cls}^{IG}(p, u, \gamma) + \lambda_{loc}[u \geq 1] L_{loc}(t^u, v)$$

Localization loss

$$L_{cls}^{IG}(p, u, \gamma) = -\alpha_\gamma \log p_u$$

Sample's average importance ranking annotation score

$$\alpha_\gamma = \begin{cases} \lambda & \gamma \leq 2.25 \\ 1/\lambda & \text{otherwise} \end{cases}$$

Zieler and Fergus (ZF) Network, ECCV 2014
Simonyan and Zisserman (VGG) Network, ICLR 2014

Importance-Guided Training and Evaluation of Object Detectors

1) Traditional test settings

easy -> min height 40 pixels, no occlusion, 15% truncation
moderate -> min height 25 pixels, partial occlusion, 30% truncation
hard -> min height 25 pixels, heavy occlusion, 50% truncation

2) Importance Guided (**IG**) training –
significant impact with importance test settings

	Traditional Test Settings			**Importance Test Settings**	
Method	Easy	Moderate	Hard	High	Low
FRCN-ZF	89.26	79.70	64.96	66.89	58.85
FRCN-ZF-IG	91.09	80.86	66.18	73.00	59.90
ΔAP	1.83	1.16	1.22	**6.11**	1.05

CVRR LISA UCSD LABORATORY FOR INTELLIGENT & SAFE AUTOMOBILES

Importance-Guided Training and Evaluation of Object Detectors

	Traditional Test Settings			Importance Test Settings	
Method	Easy	Moderate	Hard	High	Low
FRCN-ZF	89.26	79.70	64.96	66.89	58.85
FRCN-ZF-IG	91.09	80.86	66.18	73.00	59.90
ΔAP	1.83	1.16	1.22	**6.11**	1.05

	Traditional Test Settings			Importance Test Settings	
Method	Easy	Moderate	Hard	High	Low
FRCN-VGG	95.63	88.98	74.65	81.73	69.54
FRCN-VGG-IG	94.54	88.71	74.01	85.13	69.09
ΔAP	-1.09	-0.27	-0.64	**3.40**	-0.45

Summary and Future Work

- Contextual, human-centric object recognition framework
- Behavior analysis with a multi-cue, multi-modal framework.

Future work:

- Late vs. early fusion for activity prediction
- Subject-specific situational awareness modeling on US Highway
- Better temporal modeling of visual cues

Journal Publications

- Eshed Ohn-Bar and Mohan M. Trivedi, “Are all objects equal? Deep Spatio-Temporal Importance Prediction in Driving Videos," Pattern Recognition, 2017.
- Eshed Ohn-Bar and Mohan M. Trivedi, "Looking at Humans in the Age of Self-Driving and Highly Automated Vehicles," IEEE Transactions on Intelligent Vehicles, 2016.
- Eshed Ohn-Bar and Mohan M. Trivedi, "Multi-scale Volumes for Deep Object Detection and Localization," Pattern Recognition, 2016.
- Eshed Ohn-Bar and Mohan M. Trivedi, "Learning to Detect Vehicles by Clustering Appearance Patterns” IEEE Transactions on Intelligent Transportation Systems, 2015.
- Eshed Ohn-Bar, Ashish Tawari, Sujitha Martin and Mohan M. Trivedi, "On Surveillance for Safety Critical Events: In-Vehicle Video Networks for Predictive Driver Assistance Systems,” Computer Vision and Image Understanding, vol. 134, pp. 130-140, 2015.
- Eshed Ohn-Bar and Mohan M. Trivedi, "Hand Gesture Recognition in Real-Time for Automotive Interfaces: A Multimodal Vision-based Approach and Evaluations,” IEEE Transactions on Intelligent Transportation Systems, Dec 2014.
- Eshed Ohn-Bar, Sujitha Martin and Mohan M. Trivedi, "Driver Hand Activity Analysis in Naturalistic Driving Studies: Issues, Algorithms and Experimental Studies,” Journal of Electronic Imaging: special section on Video Surveillance and Transportation Imaging Applications, Vol. 22, No. 4, 2013.

Conference Publications

- Eshed Ohn-Bar and Mohan M. Trivedi, What Makes an On-road Object Important? International Conference on Pattern Recognition (ICPR), 2016.
- Eshed Ohn-Bar and Mohan M. Trivedi, Detection and Localization with Multi-scale Models, International Conference on Pattern Recognition (ICPR), 2016.
- Eshed Ohn-Bar and Mohan M. Trivedi, To Boost or Not to Boost? On the Limits of Boosted Trees for Object Detection, International Conference on Pattern Recognition (ICPR), 2016
- Nikhil Das, Eshed Ohn-Bar and Mohan M. Trivedi, "On Performance Evaluation of Driver Hand Detection Algorithms: Challenges, Dataset, and Metrics," IEEE Conference on Intelligent Transportation Systems, September, 2015.
- Eshed Ohn-Bar and Mohan M. Trivedi, "A Comparative Study of Color and Depth Features for Hand Gesture Recognition in Naturalistic Driving Settings," IEEE Intelligent Vehicles Symposium, June, 2015.
- Eshed Ohn-Bar and Mohan M. Trivedi, "Beyond Just Keeping Hands on the Wheel: Towards Visual Interpretation of Driver Hand Motion Patterns," IEEE Intelligent Transportation Systems Conference, (ITSC2014), Oct. 2014.
- Eshed Ohn-Bar, Sujitha Martin, Ashish Tawari, and Mohan M. Trivedi, "Head, Eye, and Hand Patterns for Driver Activity Recognition," International Conference on Pattern Recognition (ICPR2014), August 2014.
- Eshed Ohn-Bar and Mohan M. Trivedi, "Joint Angles Similarities and HOG² for Action Recognition," IEEE Computer Vision and Pattern Recognition Workshop on Human Activity Understanding from 3D Data (HAU3D), June 2013.
- Eshed Ohn-Bar and Mohan M. Trivedi, "The Power is in Your Hands: 3D Analysis of Hand Gestures in Naturalistic Video," IEEE Computer Vision and Pattern Recognition Workshop on Analysis and Modeling of Faces and Gestures (AMFG), June 2013.
- Eshed Ohn-Bar and Mohan M. Trivedi, "In-Vehicle Hand Activity Recognition Using Integration of Regions," IEEE Intelligent Vehicles Symposium, June 2013.

Acknowledgment

Committee

Professor Trivedi, for valuable guidance and mentorship

Professors Belongie, Cottrell, Rao, and Vasconcelos

Colleagues

Cuong, Sayanan, Ashish, Sujitha, Larry, Kevan, Rakesh, Sean, Frankie, Ravi, Andreas, Miklas, Jacob, Alfredo, Nikhil, Akshay, Borhan, Aida, Sourabh, Nachiket, Grady, Mengying, Alice, Jesse, Mo, Gabrielle, Crystal, Karen, Kacy, Shana, Sam, Todd, Martha,

Sponsors: Toyota-CSRC, KETI, Audi, UC Discovery, VW-ERL

Family and Friends