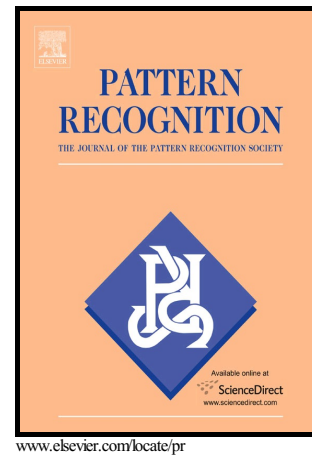


## Author's Accepted Manuscript

Are All Objects Equal? Deep Spatio-Temporal Importance Prediction in Driving Videos

Eshed Ohn-Bar, Mohan Manubhai Trivedi



PII: S0031-3203(16)30242-4  
DOI: <http://dx.doi.org/10.1016/j.patcog.2016.08.029>  
Reference: PR5858

To appear in: *Pattern Recognition*

Received date: 16 March 2016  
Revised date: 8 July 2016  
Accepted date: 24 August 2016

Cite this article as: Eshed Ohn-Bar and Mohan Manubhai Trivedi, Are All Objects Equal? Deep Spatio-Temporal Importance Prediction in Driving Videos *Pattern Recognition*, <http://dx.doi.org/10.1016/j.patcog.2016.08.029>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and a review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

# Are All Objects Equal?

## Deep Spatio-Temporal Importance Prediction in Driving Videos

Eshed Ohn-Bar\*, and Mohan Manubhai Trivedi

*Computer Vision and Robotics Research Lab  
University of California, San Diego  
CA, 92093, USA*

---

### Abstract

Understanding intent and relevance of surrounding agents from video is an essential task for many applications in robotics and computer vision. The modeling and evaluation of contextual, spatio-temporal situation awareness is particularly important in the domain of intelligent vehicles, where a robot is required to smoothly navigate in a complex environment while also interacting with humans. In this paper, we address these issues by studying the task of on-road object importance ranking from video. First, human-centric object importance annotations are employed in order to analyze the relevance of a variety of multi-modal cues for the importance prediction task. A deep convolutional neural network model is used for capturing video-based contextual spatial and temporal cues of scene type, driving task, and object properties related to intent. Second, the proposed importance annotations are used for producing novel analysis of error types in image-based object detectors. Specifically, we demonstrate how cost-sensitive training, informed by the object importance annotations, results in improved detection performance on objects of higher importance. This insight is essential for an application where navigation mistakes are safety-critical, and the quality of automation and human-robot interaction is key.

*Keywords:* Spatio-temporal object analysis, vision-based behavior analysis, intelligent and automated vehicles, human-centric artificial intelligence, contextual

---

\*Corresponding author  
Email address: eohnbar@ucsd.edu (Eshed Ohn-Bar)

Download English Version:

<https://daneshyari.com/en/article/4969926>

Download Persian Version:

<https://daneshyari.com/article/4969926>

[Daneshyari.com](https://daneshyari.com)