Knowledge Transfer for Scene-specific Motion Prediction

Lamberto Ballan[1,3], Francesco Castaldo[2], Alexandre Alahi[1], Francesco Palmieri[2], Silvio Savarese[1]

[1]*Stanford University*, [2]*Second Univ. of Naples*, [3]*University of Florence*

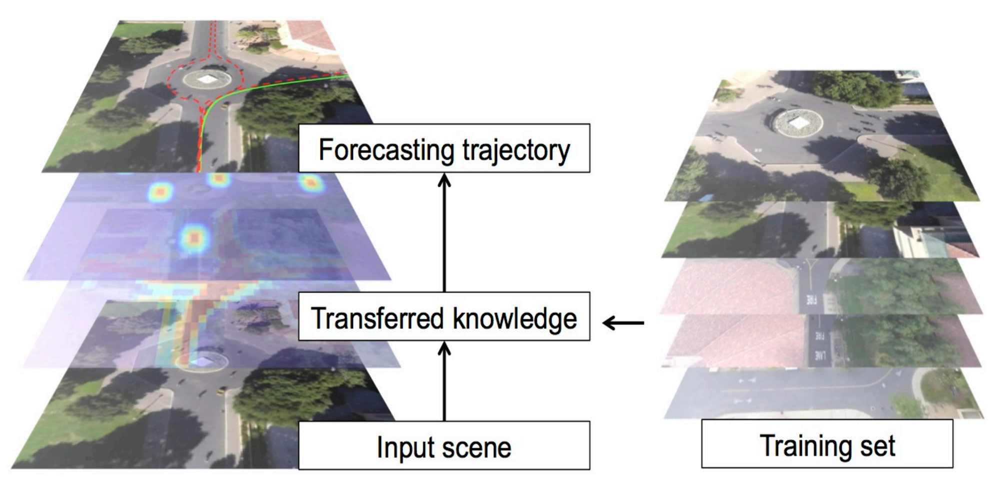
ECCV16

Motivation

When given a frame of a video, humans can not only interpret the scene, but also they are able to forecast the near future.

This ability is mostly driven by their rich prior knowledge about:

- *dynamics of moving agents*
- *semantic of the scene*



We exploit the interplay between these two key elements for trajectory prediction, and apply knowledge transfer to make predictions on a new scene.

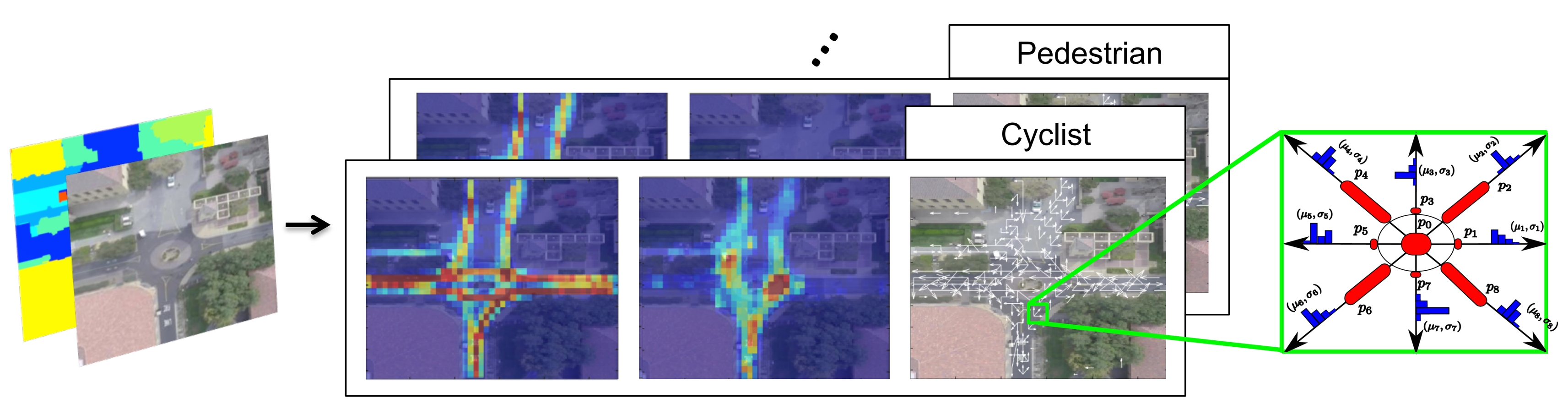
Our Model

Navigation Map

Given an input scene we overlay an uniform grid and build a map ***M*** which collects the navigation statistics for a given target class.

For each patch we encode four type of information:

- **Popularity score**: measures how many times a patch has been explored
- **Routing score**: measures the probability of changing behaviors
- **Histogram of Directions**
- **Histogram of Speeds**



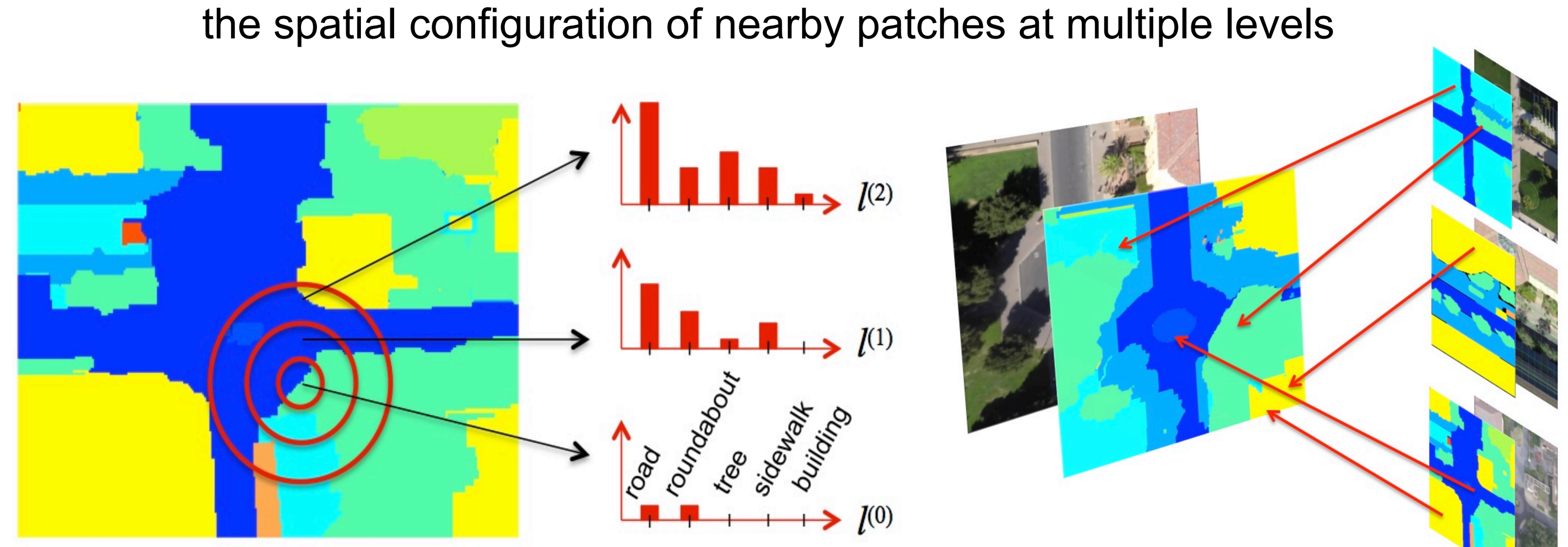
(a) Input scene (b) "Navigation map" (c) *HoD* and *HoS*

Prediction Model

- The target state is defined by its position and velocity: $\mathbf{X}_k = (\mathbf{P}_k, \mathbf{V}_k)^T$
- Starting from a given initial condition $\mathbf{X}_0$, our goal is to generate a sequence of future states $\mathbf{X}_1, \ldots, \mathbf{X}_T$, i.e. a path Ψ_T
- The dynamic process describing the target motion is defined by:
  - (1) $\mathbf{P}_{k+1} = \mathbf{P}_k + (\Omega_k \cos\Theta_k; \Omega_k \sin\Theta_k) + \mathbf{w}_k$ *(constant velocity model)*
  - (2) $\mathbf{V}_{k+1} = \Phi(\mathbf{P}_k, \mathbf{V}_k; \boldsymbol{M})$ *(this allows non-linear behaviors)*
- A Dynamic Bayesian Network exploits ***M*** for path prediction

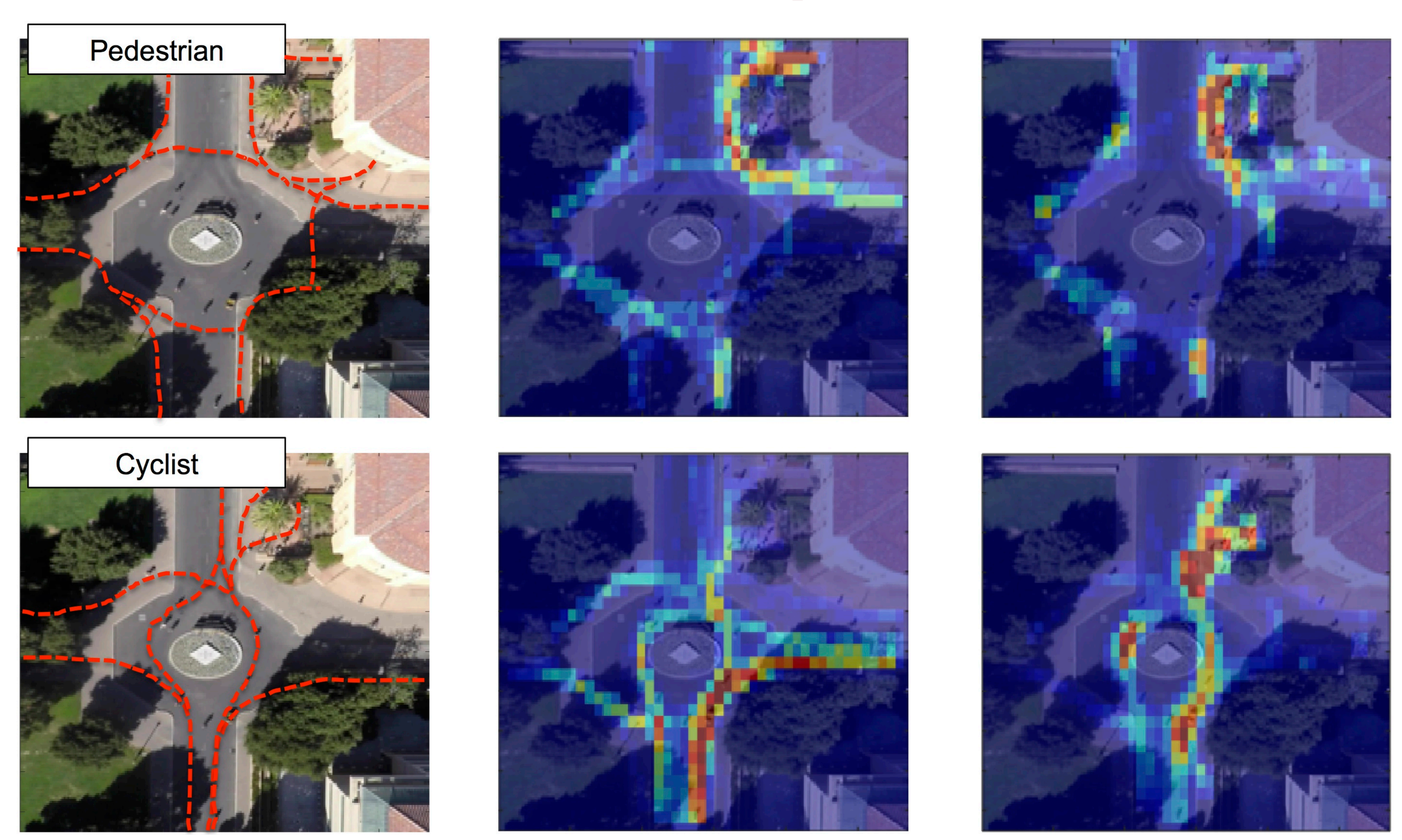
Knowledge Transfer

- Retrieval-based approach that uses scene similarity to transfer the functional properties that have been learned on the training set, to a new scene
- **Scene parsing**: we use the scene parsing algorithm in [37] (based on SIFT + LLC, GIST, color histograms and MRF inference to refine the labeling)
- **Semantic Context Descriptors**: each descriptor is a weighted concatenation of the *global* and *local semantic context* components: $\mathbf{p}_i = w\,\mathbf{g}_i + (1-w)\,\mathbf{l}_i$
  - (1) *global context*: C-dim vector of L2 distances between the centroid of the patch and the closest point in the full image labeled as c
  - (2) *local context*: this is a shape-context like representation which encodes the spatial configuration of nearby patches at multiple levels



(a) *local context* (3 levels) (b) Patch retrieval

Qualitative Examples



(a) (b) (c)

(a) most common paths for two classes (b,c) corresponding popularity and routing maps

Acknowledgements

This work is funded by Toyota (1186781-31-UDARO), ONR (1165419-10-TDAUZ), MURI (1186514-1-TBCJE). L. Ballan is supported by an EU Marie Curie Fellowship (623930).

Experiments

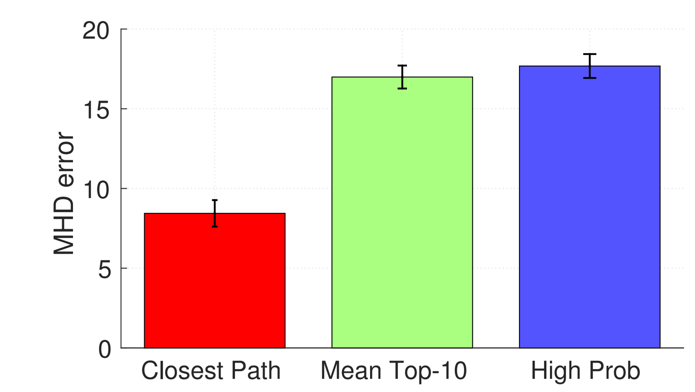
- **UCLA-courtyard dataset**: 6 annotated videos, 1 scene (2 views), single-class (*pedestrian*), scene labeled with 8 semantic classes
- **Stanford-UAV dataset** [28]: 21 video, 6 physical areas, 15 different scenes, multi-class (we use *pedestrian* and *cyclist*), 10 scene labels
- **Evaluation Metric**: *Modified Hausdorff Distance (MHD)*

Results: Path Prediction

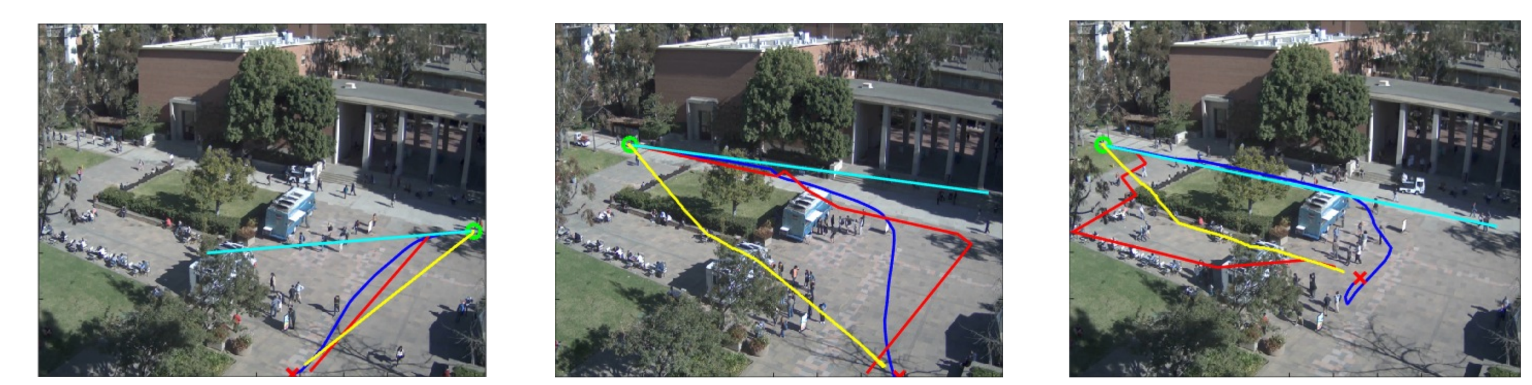
MHD error

	UCLA-courtyard	*Stanford-UAV*
LP	41.36±0.98	31.29±1.25
LP_{CA}	-	21.30±0.80
IOC [16]	14.47±0.77	14.02±1.13
SFM [43]	-	12.10±0.60
Ours	**10.32**±0.51	**8.44**±0.72

(a) MHD error for a given final destination



(b) Path generation strategies (ours)



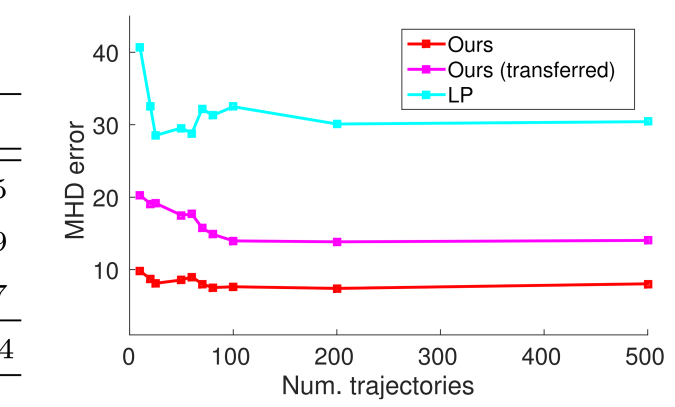
Qualitative results: *blue* is ground-truth, *cyan* is LP, *yellow* is IOC, *red* is our model

Results: Knowledge Transfer

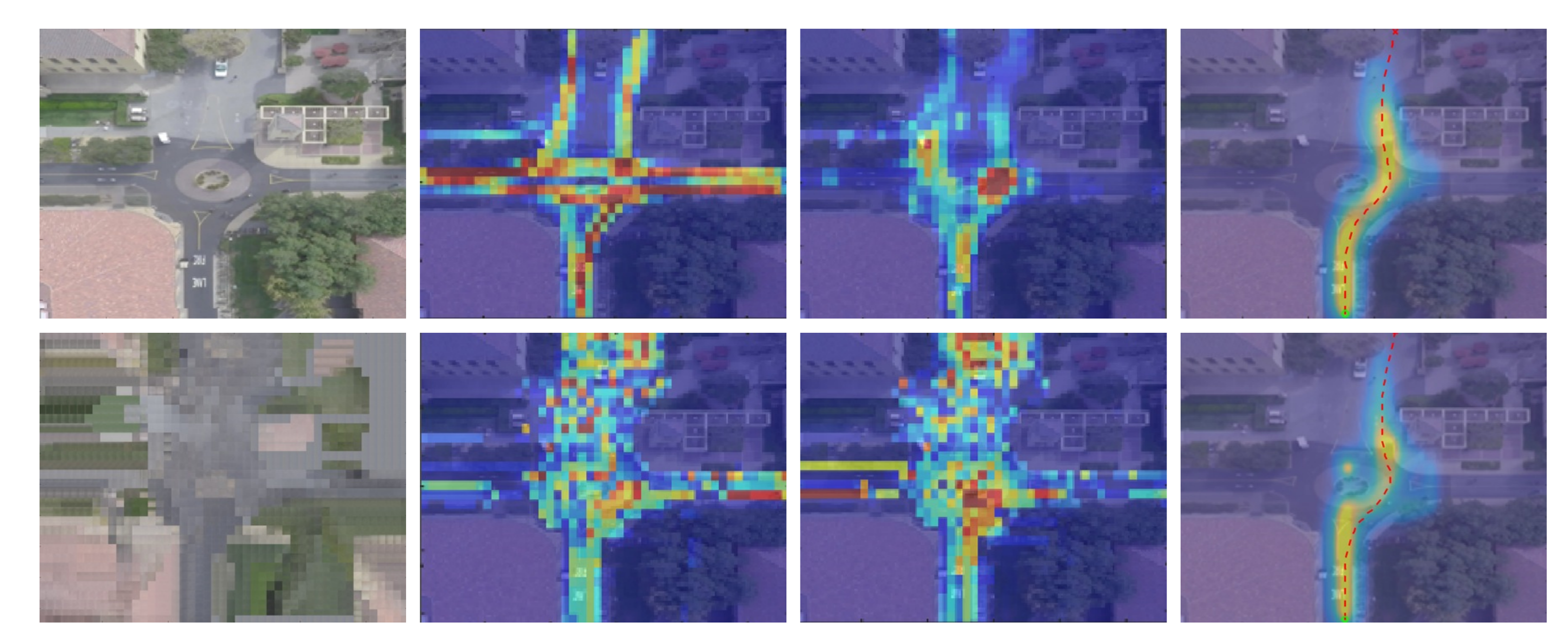
MHD error

	Pedestrian	*Cyclist*	Overall
LP	34.48	28.09	31.29±1.25
PM	22.75	20.58	21.67±1.19
IOC [16]	17.99	18.84	18.42±0.97
Ours	12.36	16.22	**14.29**±0.84

(a) Path prediction



(b) Impact of training data



Qualitative results: "standard" path prediction (1st row) *vs* knowledge transfer (2nd row)

[16] Kitani et al, "IOC", ECCV 2012 [28] Robicquet et al, ECCV 2016 [37] Yang et al, CVPR 2014 [43] Yamaguchi et al, "SFM", CVPR 2011

Stanford-UAV dataset: http://cvgl.stanford.edu/projects/uav_data