

## Risky Region Localization

### Goals

- Localize risky regions in images
- Estimate risk value



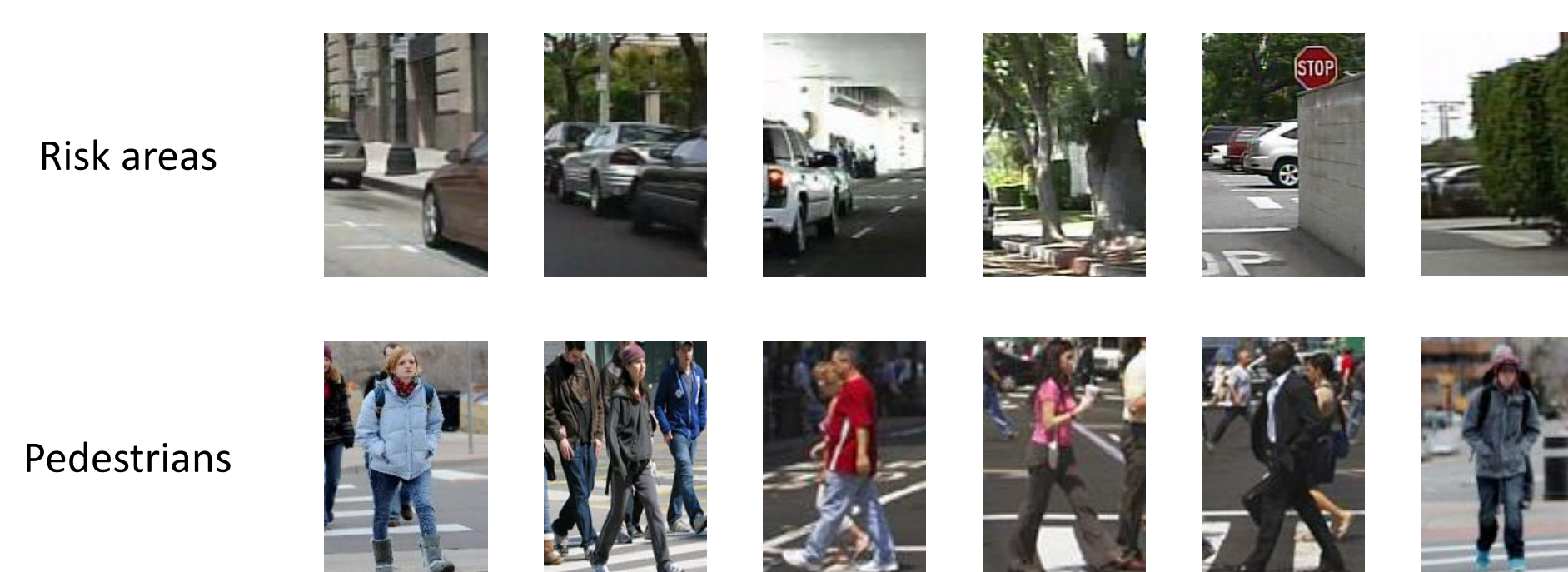
### Motivations

- Early control for risk avoidance
- Safe trajectory generation and comfort speed control

## Challenges

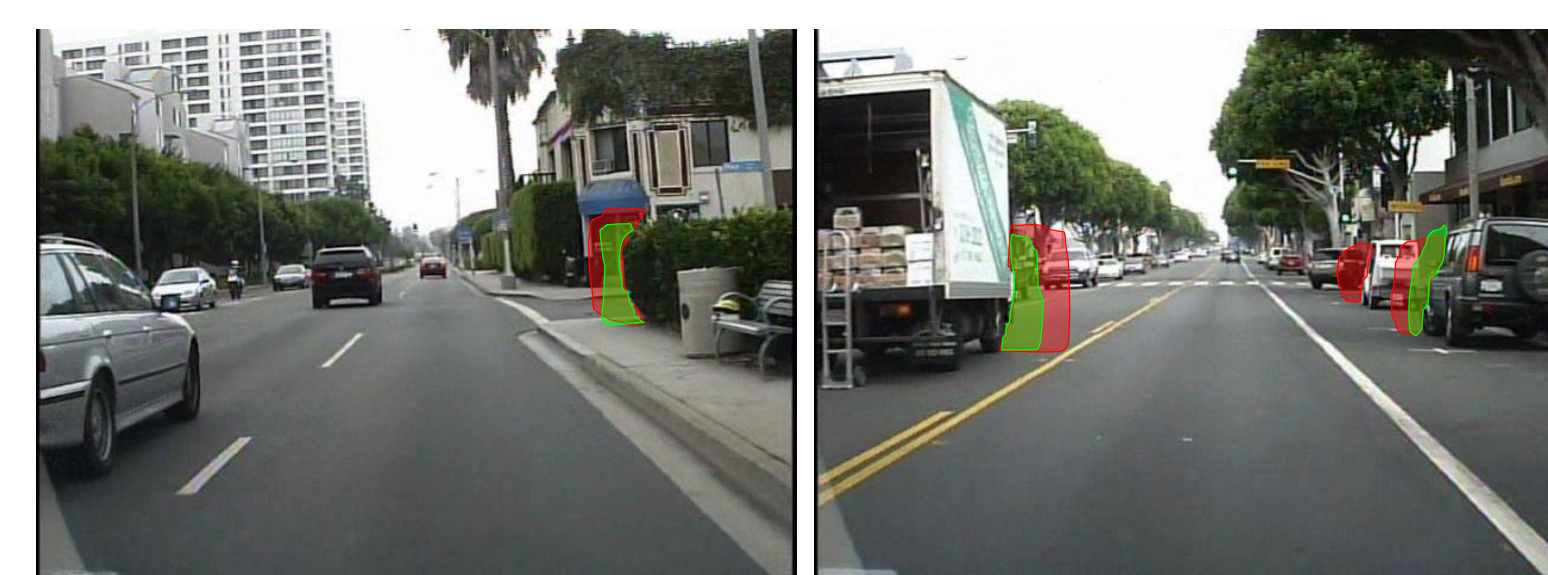
### Challenges

- Risky regions exhibit large visual variations
- Risk areas are not easily characterized by appearances of single objects



### Related works

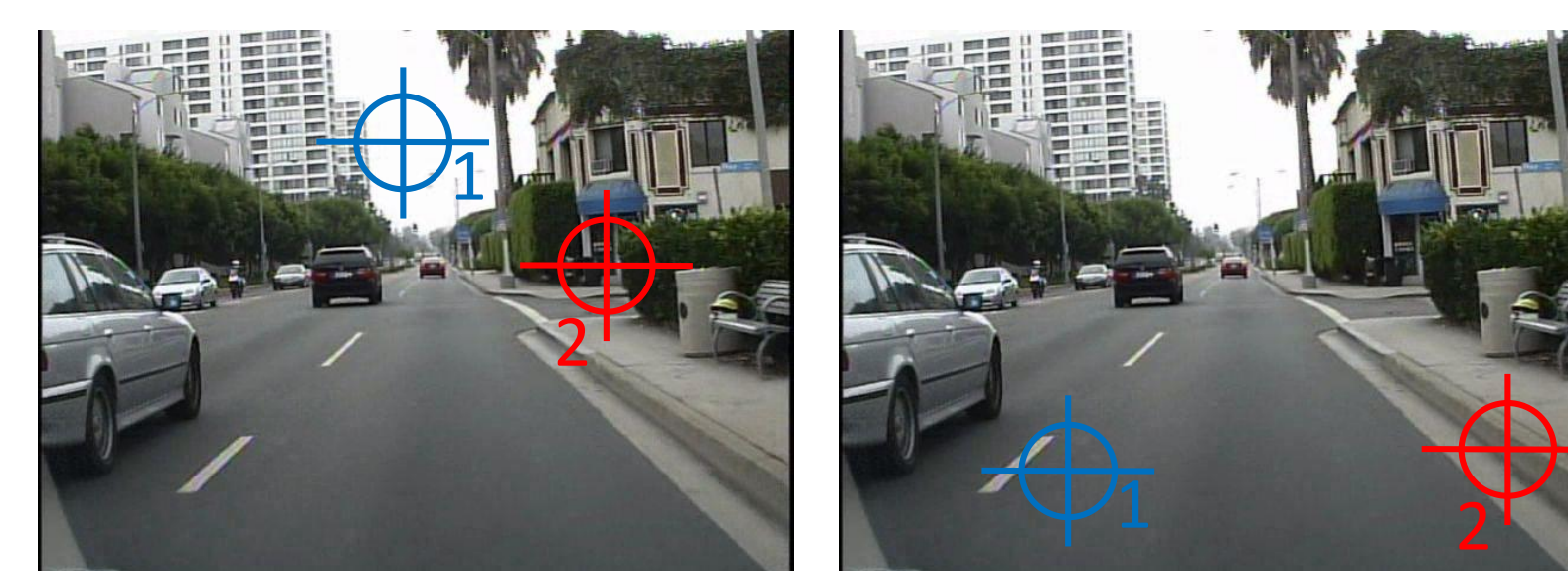
- Segmentation: hard to obtain accurate per-pixel masks



Green: Per-pixel mask of worker1, Red: Per-pixel mask of worker2

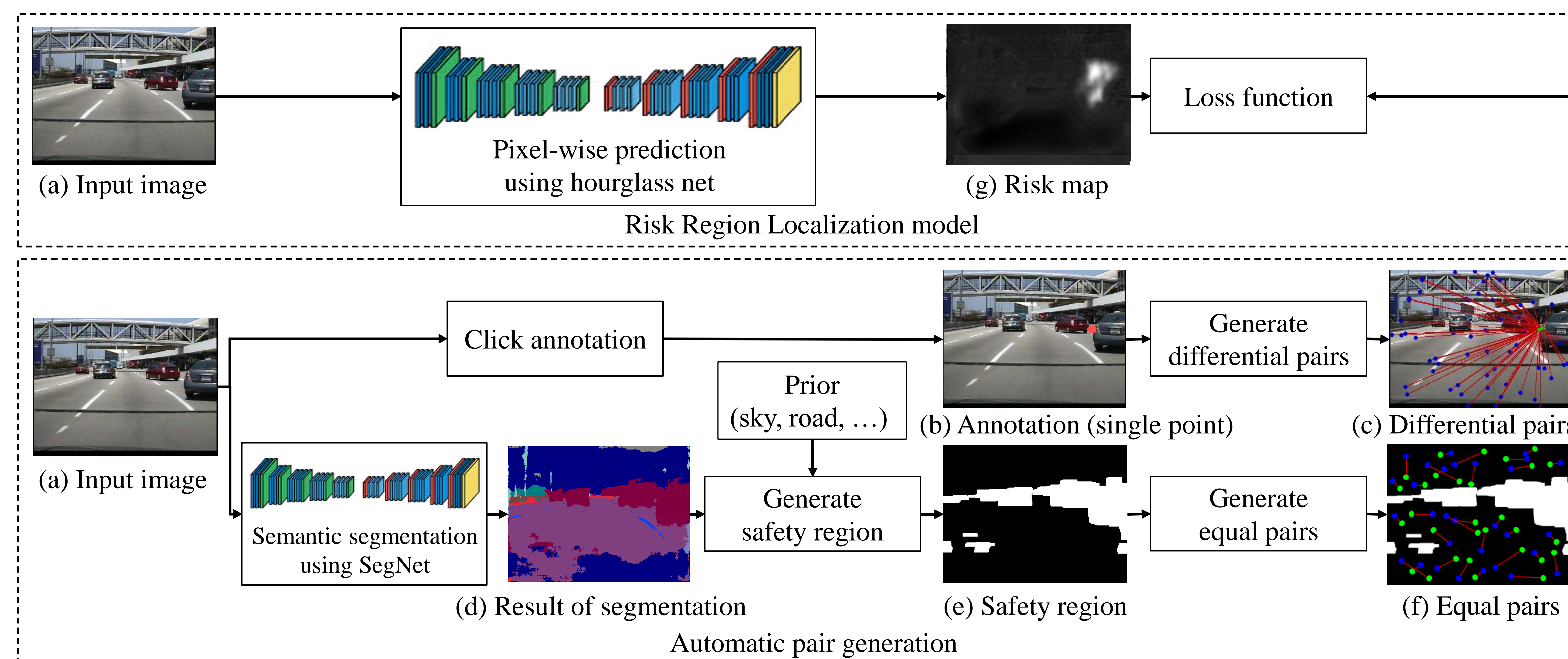
Two workers only have a coincidence of 54% on average (over 3,438 images from the Caltech pedestrian dataset)

- Single Image Depth Estimation: hard to collect pairwise annotations



Risk of point 1 < Risk of point 2      Risk of point 1 = Risk of point 2

## Our Framework



### Risk Localization Network

- Supervision: pairwise ordinal risk relationship
- Input: image, Output: risk map

### Loss Function

- Differential pairs: ordinal relationships
- Equal pairs: risk values of safety areas → equal, small

### Single Point Annotation and Automatic pair generation

- Single Point Annotation: no need to input the region boundary
- Automatic pair generation: reduce annotation cost
  - (1) Generate differential risk pairs using the annotated point and random points
  - (2) Utilize prior about safety areas: generate same risk pairs

$$L(I, R, z) = \sum_{k=1}^K \psi_k(I, p_{1k}, p_{2k}, r, z)$$

$$\psi_k(I, p_{1k}, p_{2k}, r, z) = \begin{cases} \log(1 + \exp(-z_{p_{1k}} + z_{p_{2k}})) & (r_k = 1) \\ (z_{p_{1k}} - z_{p_{2k}})^2 + z_{p_{1k}}^2 + z_{p_{2k}}^2 & (r_k = 0) \end{cases}$$

## Experiments

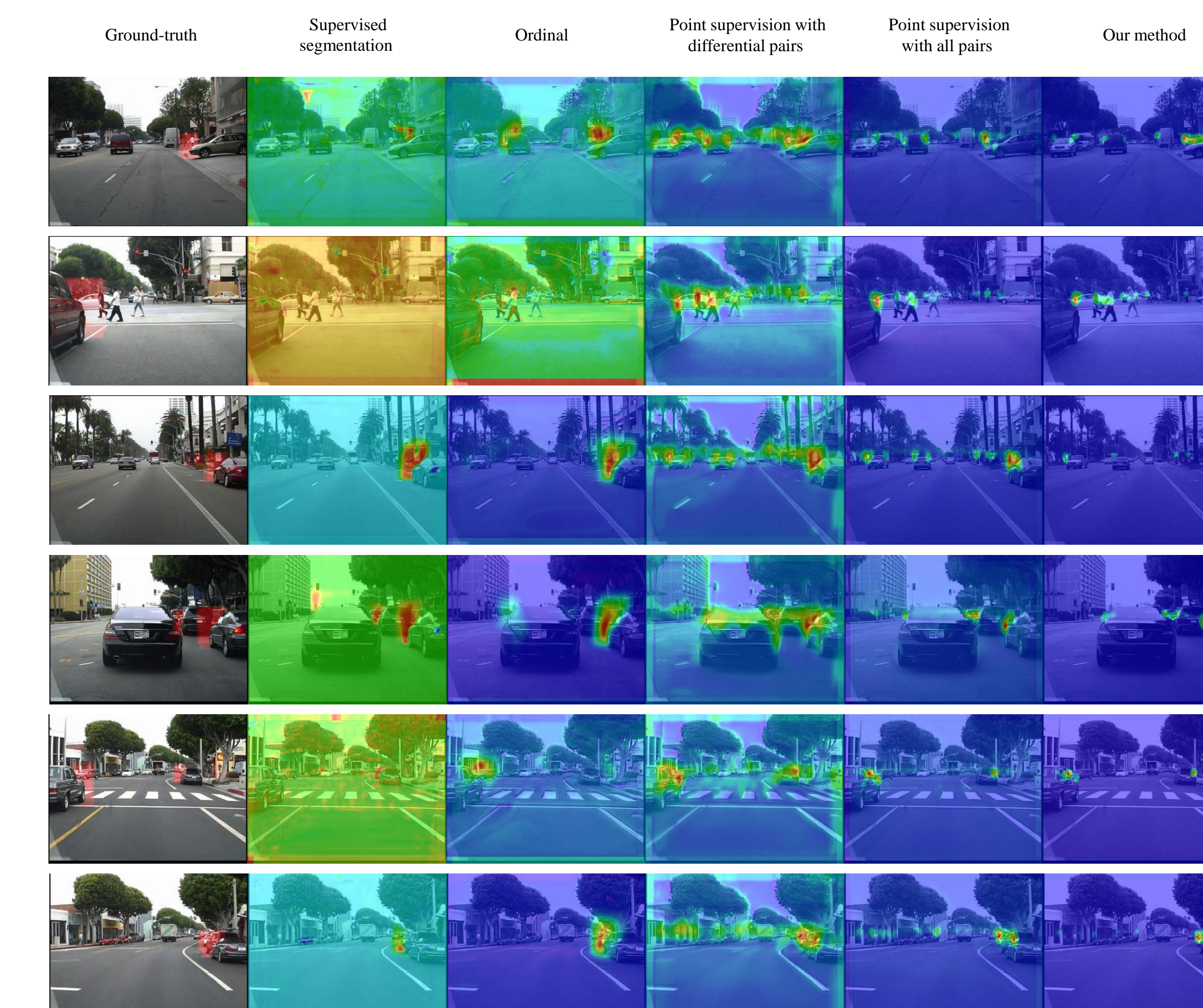
### Dataset

- Over 4,500 images from Caltech pedestrian dataset (train: 2260, validation: 1178, test: 1193)
- risky region labels from 2 workers for training set, 10 workers for test set

### Evaluations

- Annotation cost: time/image
- Risk localization: Recall, Precision, F1 from binary map (risk output > 0.5)
- Risk value: Root Mean Square Error

Method (supervision)	Annotation Time (sec/img)	Risk Localization			Risk Value		
		Recall (%)	Precision (%)	F1 score	RMSE (All-pixel)	RMSE (low-risk)	RMSE (high-risk)
Supervised segmentation [1]	32.2	77.3	29.7	0.429	0.216	0.215	<b>0.239</b>
Ordinal [2]	50.6	72.5	27.6	0.400	0.128	0.125	0.280
Point supervision with differential pairs	<b>3.4</b>	68.5	38.9	0.496	0.085	0.075	0.627
Point supervision with all pairs	<b>3.4</b>	75.1	38.3	0.507	0.093	0.083	0.599
<b>Our method</b>	<b>3.4</b>	<b>77.4</b>	<b>43.7</b>	<b>0.558</b>	<b>0.072</b>	<b>0.061</b>	0.625



[1] V. Badrinarayanan, A. Kendall, and R. Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. arXiv:1511.00561, 2015.

[2] W. Chen, Z. Fu, D. Yang, and J. Deng. Single-image depth perception in the wild. In NIPS, 2016.