

Single View Metrology in the Wild

Rui Zhu¹, Xingyi Yang¹, Yannick Hold-Geoffroy², Federico Perazzi², Jonathan Eisenmann²,
Kalyan Sunkavalli², Manmohan Chandraker¹

University of California San Diego¹, Adobe Research²



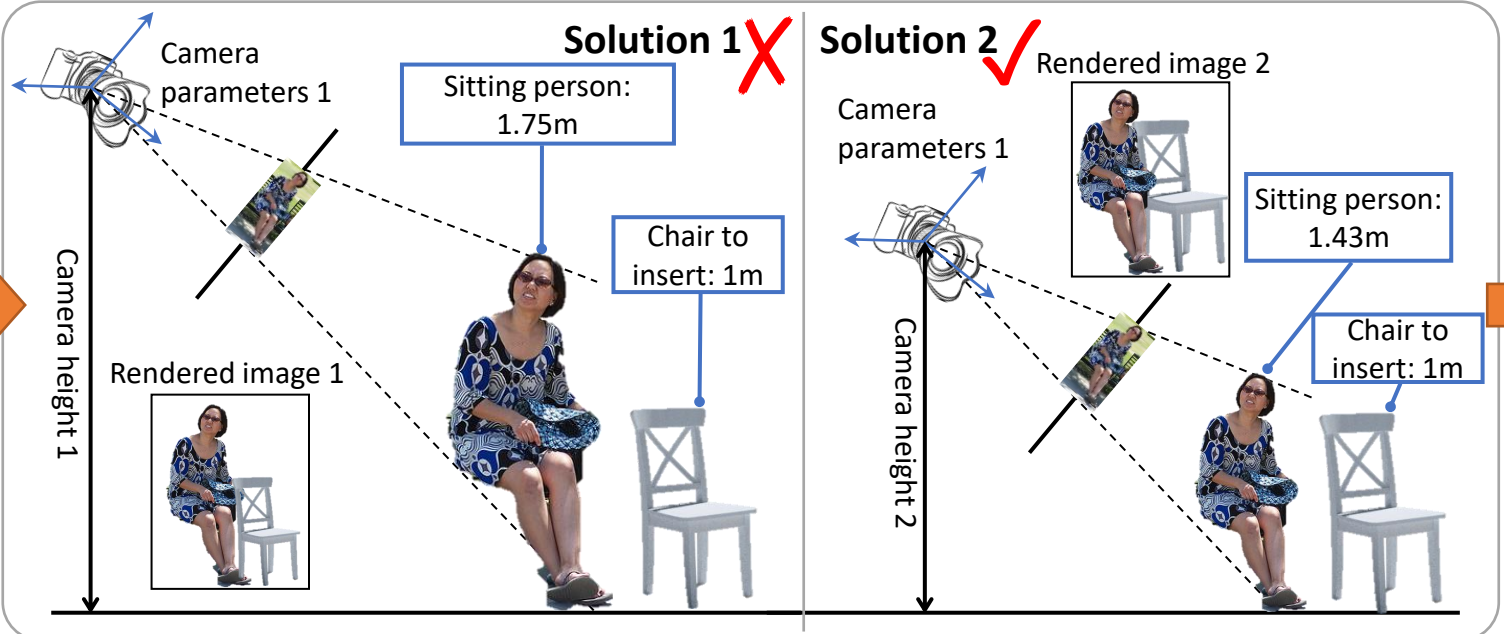
UC San Diego
JACOBS SCHOOL OF ENGINEERING



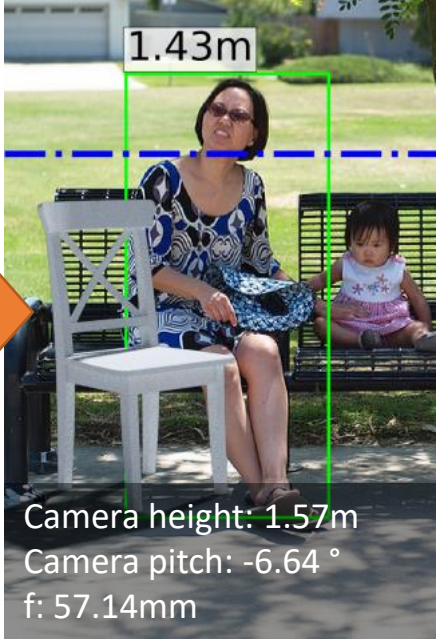
Motivation: Single View Metrology with *Absolute* Scale Estimation



Input image



Geometric camera calibration with scale estimation

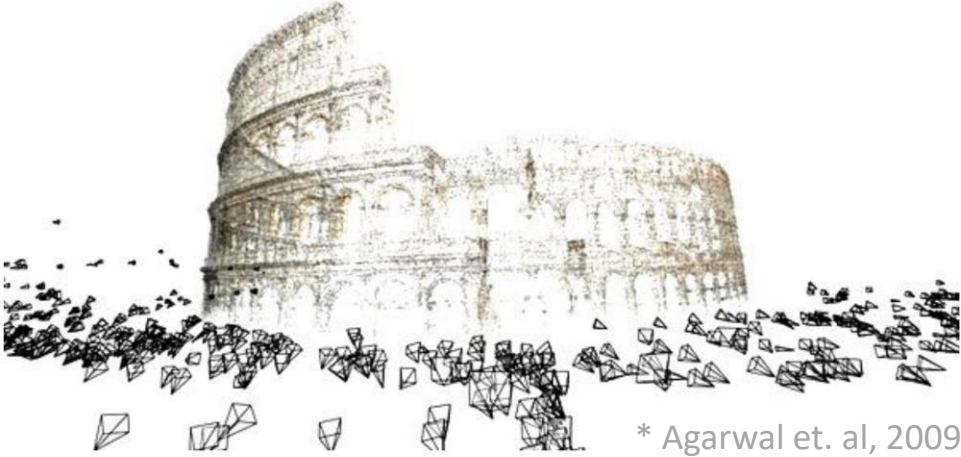


Virtual object insertion



Motivation: Single View Scale Ambiguity

Structure-from-motion (SfM) from uncalibrated cameras



Toy 1: "Toy Story Roarin' Rex", height = 27cm
Toy 2: "Toy Story Rex Figure", height = 12cm



Monocular camera calibration in the wild

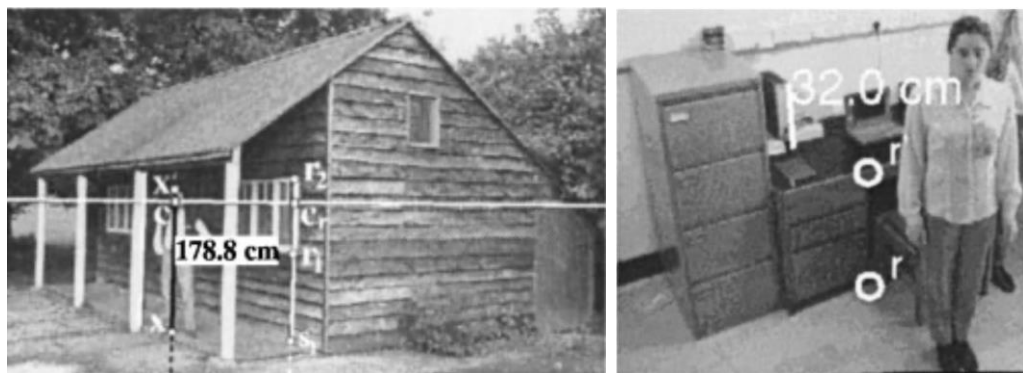


Single image depth estimation



Motivation: Single View Metrology

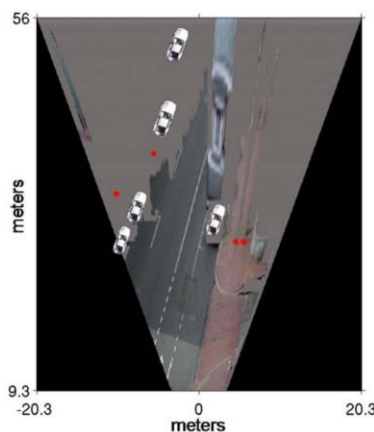
3D estimation from 2D measurements via perspective geometry



* Criminisi et. al, 2000



* Kar et. al, 2015



* Hoiem et. al, 2008



Motivation: Single View Metrology *in the Wild*



Input Image in the Wild

Camera Calibration

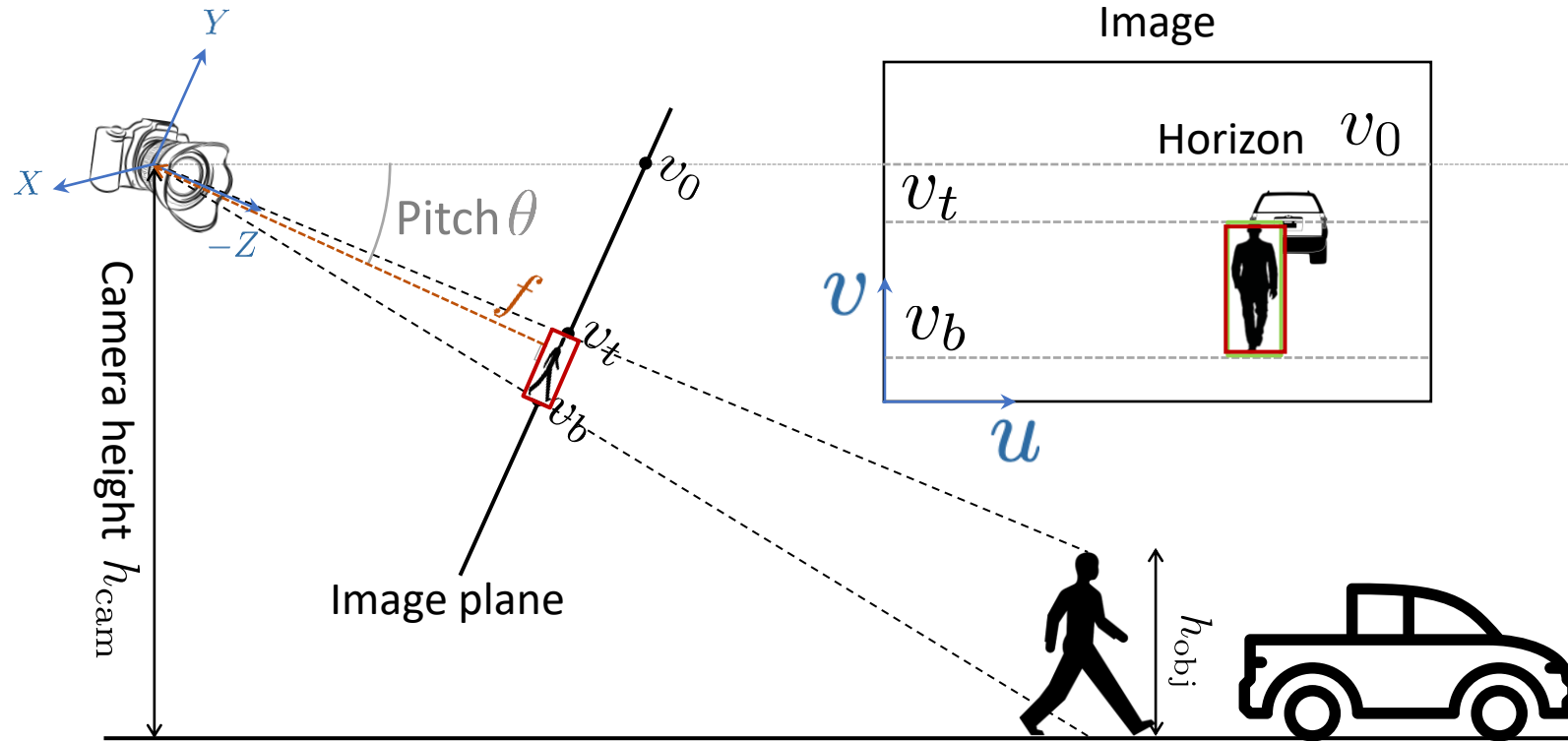
- Orientation
- Field of View

Camera Height Estimation

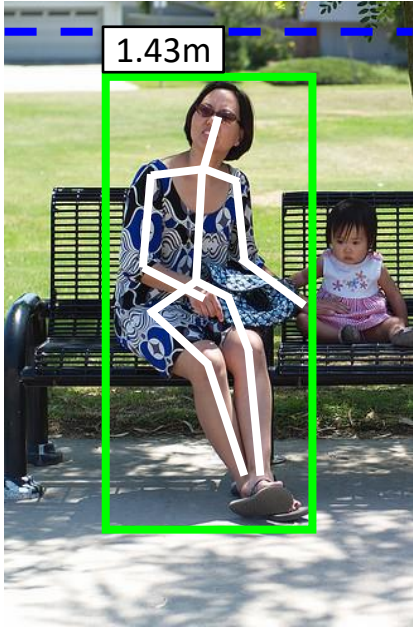
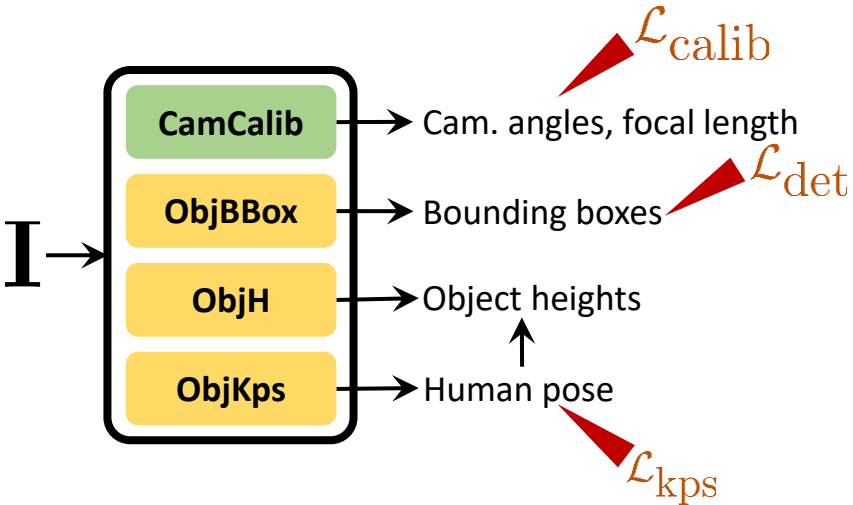
Object Heights Estimation



Scene Model



Pipeline



- Cam. H: 1.17m
- Cam. Pitch: -6.64°
- Focal Length: 57.14mm

Camera estimation heads

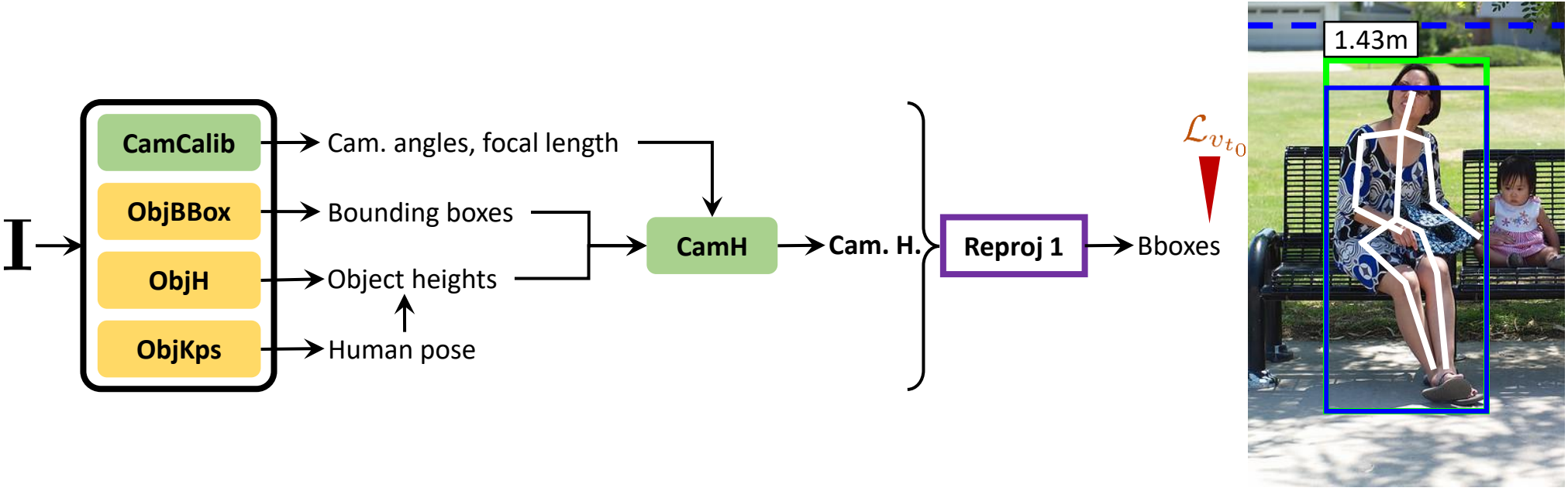
Object estimation heads

Reproj 1 BBox reprojection function

L Loss



Pipeline



- Cam. H: 1.17m
- Cam. Pitch: -6.64°
- Focal Length: 57.14mm

Camera estimation heads

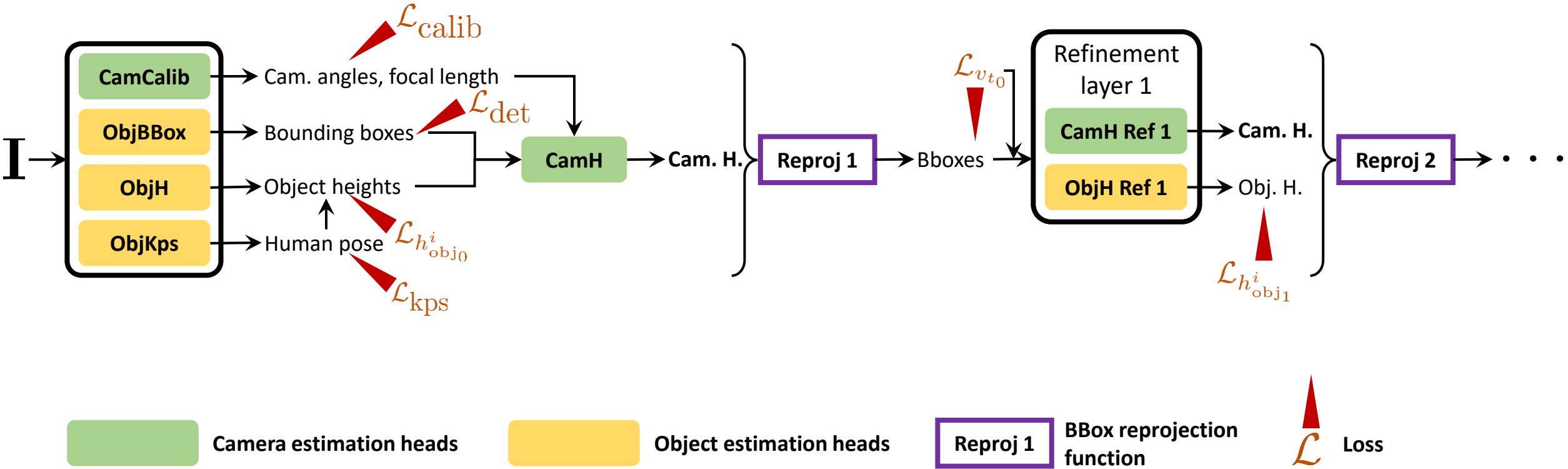
Object estimation heads

Reproj 1 BBox reprojection function

∇ \mathcal{L} Loss

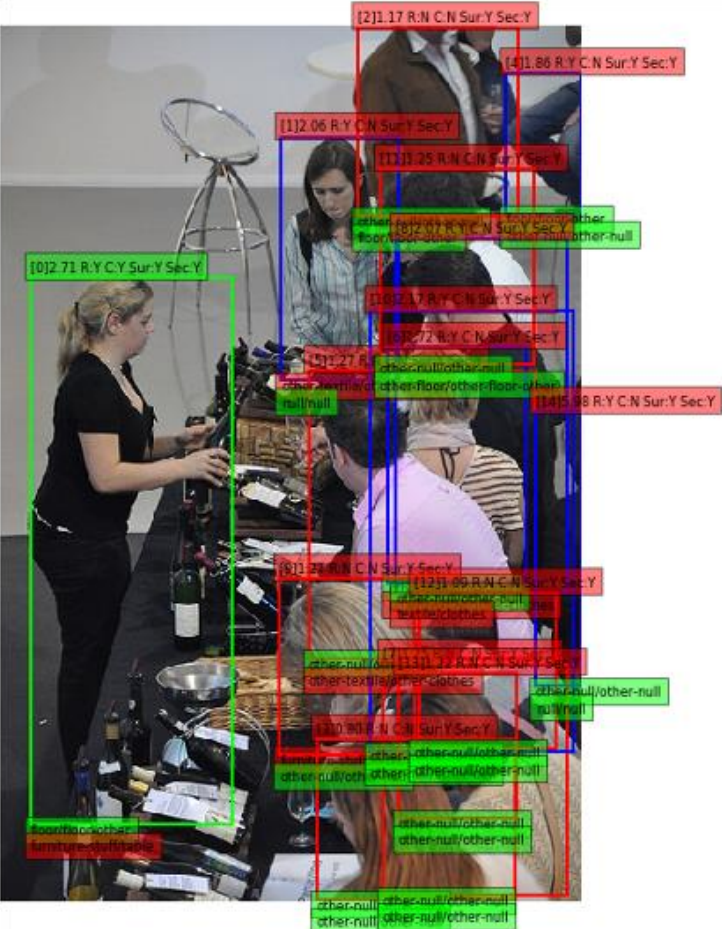
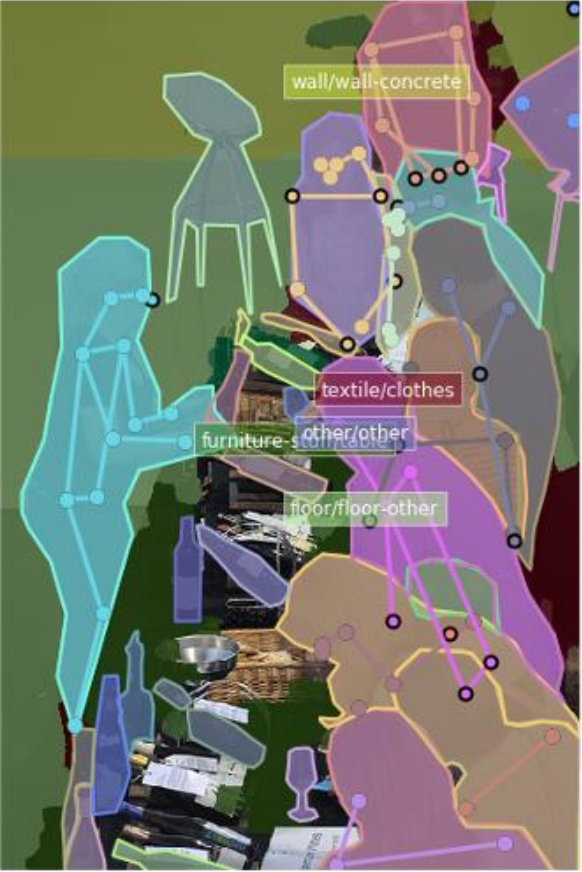


Pipeline

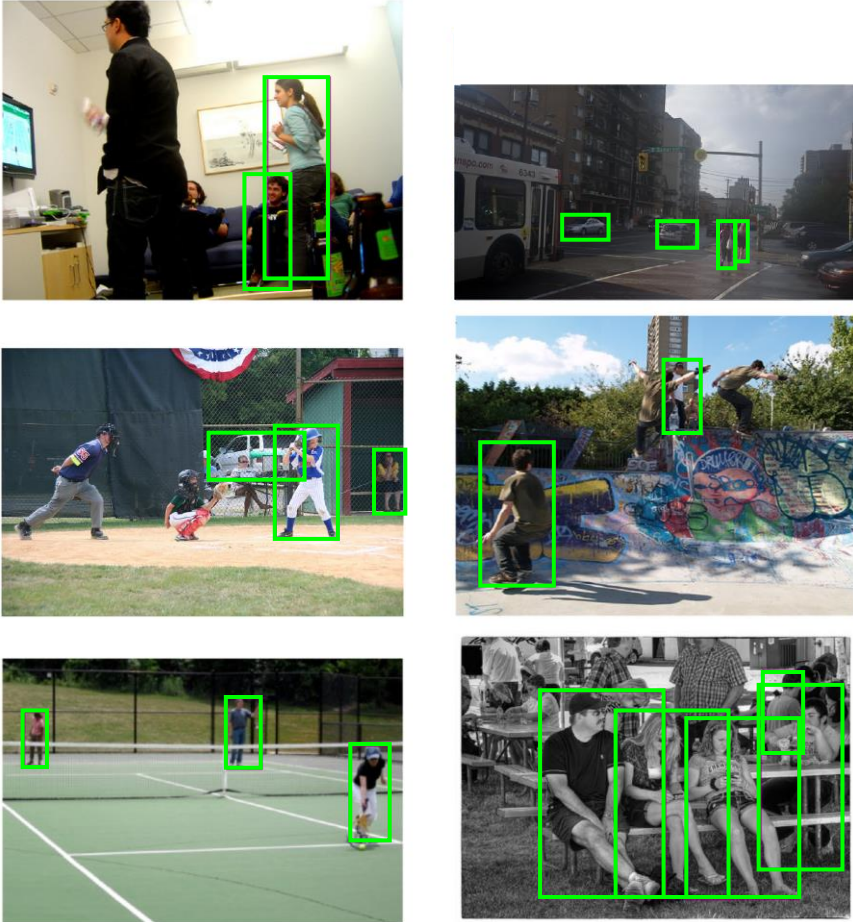


Evaluation

- Training data: *COCO-Scale*



Pruning from original COCO dataset (green is valid)



Sample images from COCO-Scale



Evaluation: on *COCO-Scale*

	Ours (1 layer)	Ours (3 layers)
Preference (↑)	43.8%	54.6%
	Ours (3 layers)	Ours (3 layers w/ keypoints)
Preference (↑)	42.8%	57.2%
	Our best	PGM*
Preference (↑)	59.5%	40.5%



PGM*

Ours w/o refinements

Ours w/ refinements

* Hoiem et. al, 2008

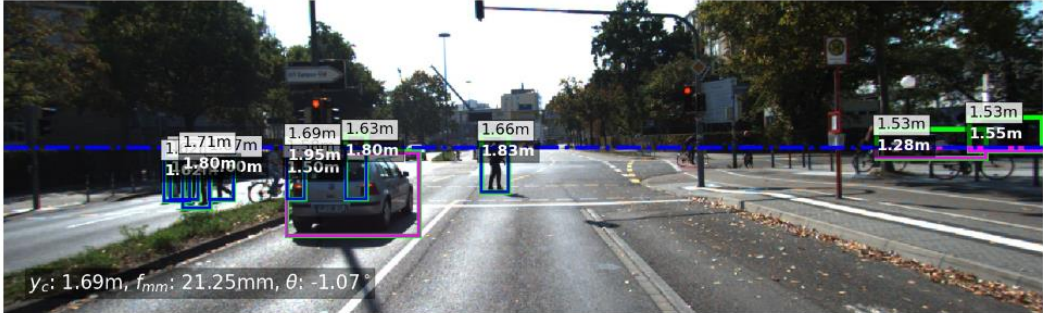
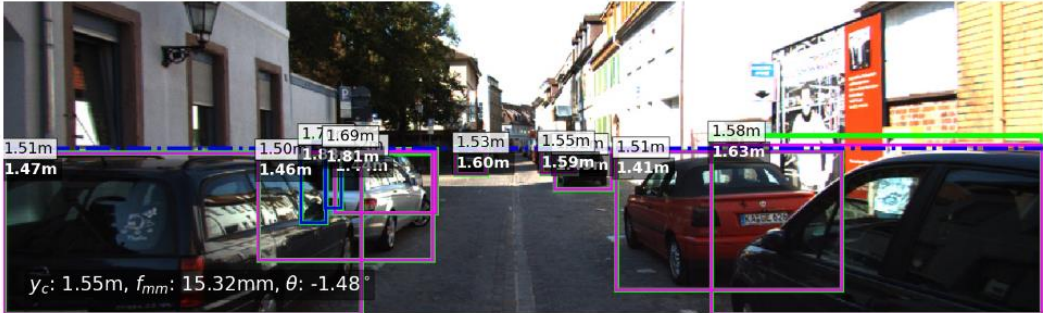
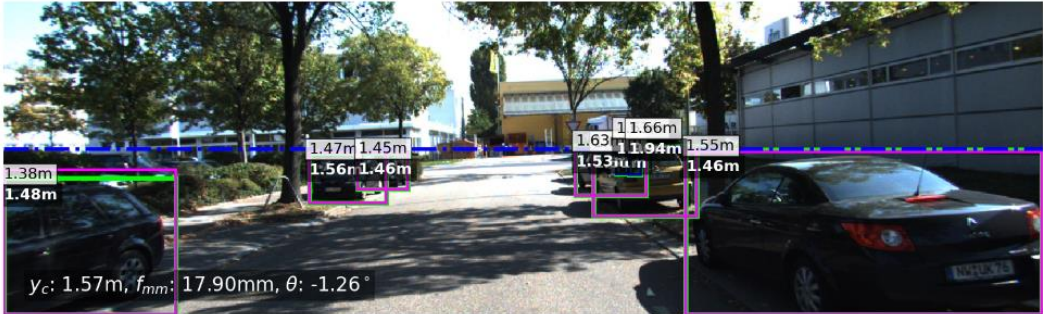


Evaluation

- Test data:
 - *KITTI*: With ground truth camera height & object heights



Evaluation: on *KITTI* & *IMDB-23K*



KITTI



IMDB-23K celebrity dataset



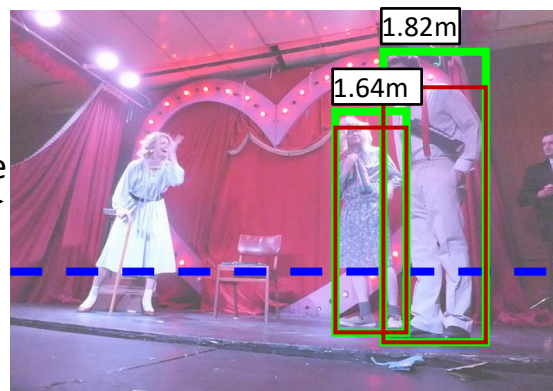
Analysis: Cascade Refinements

Camera height refinement



Camera height = 1.13m

refine
→



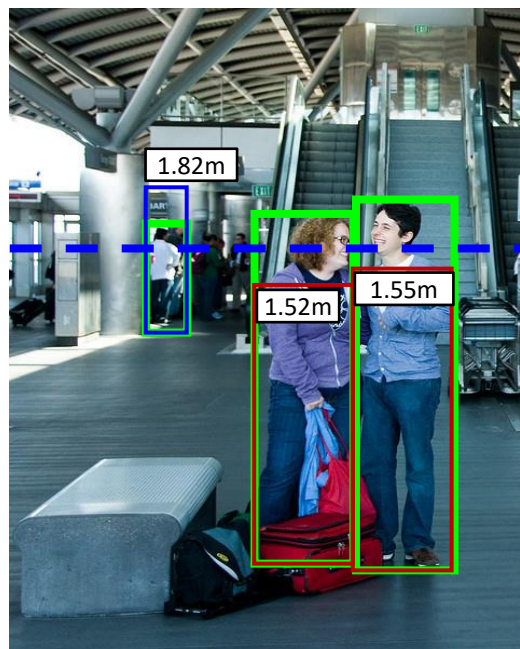
Camera height = 1.13m - 0.46m

refine
→



Camera height = 1.13m - 0.46m - 0.18m

Object height refinement



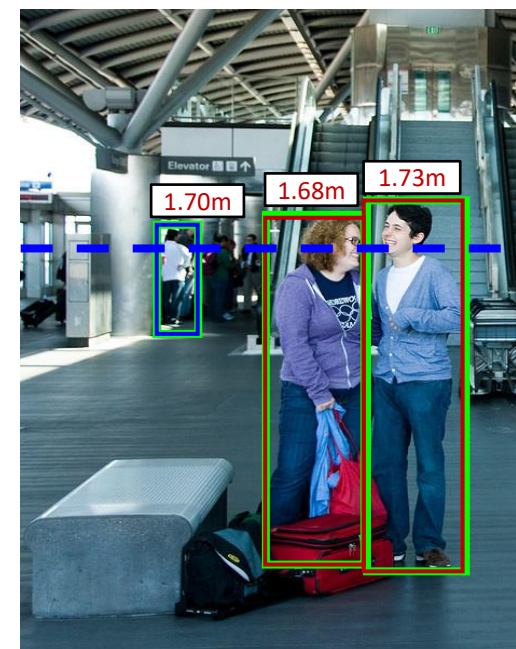
Camera height = 1.64m

refine
→



Camera height = 1.64m

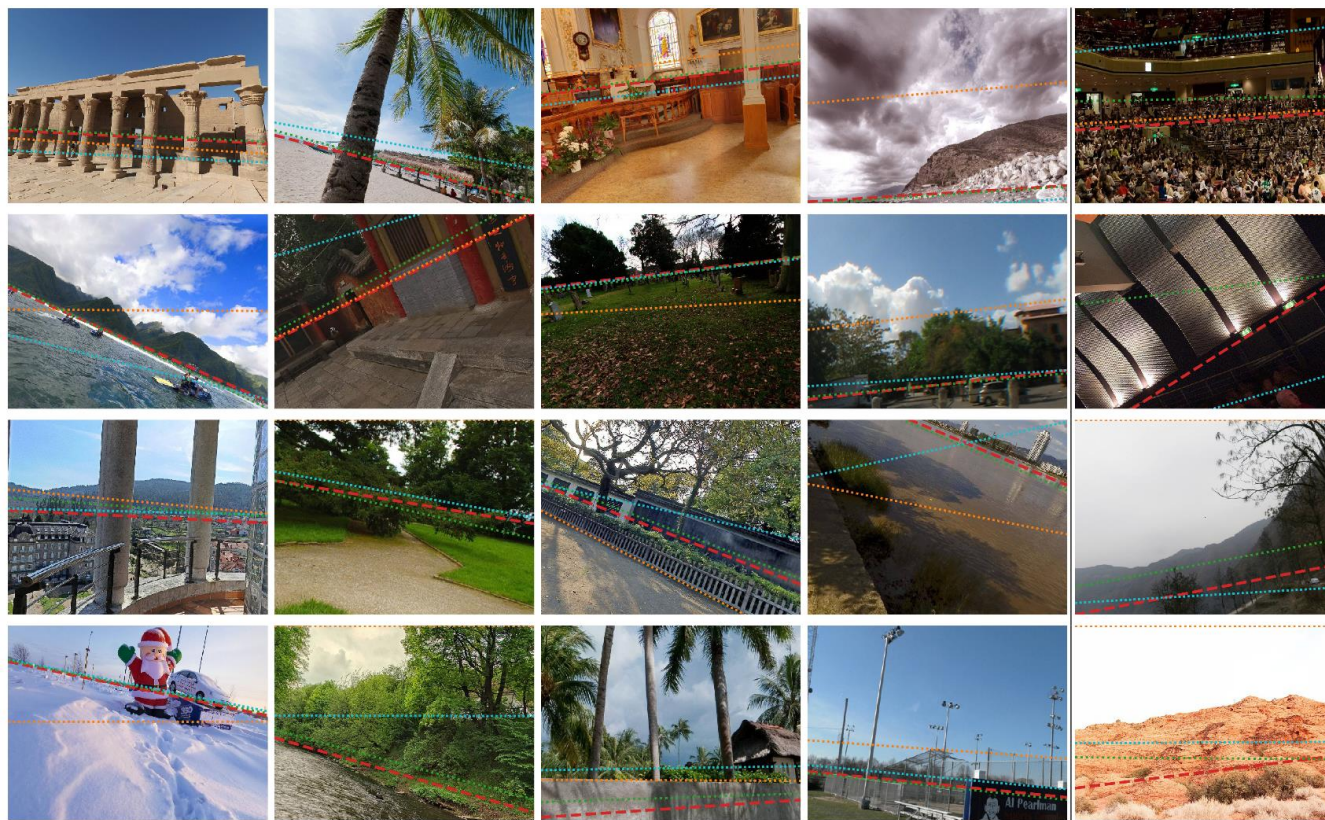
refine
→



Camera height = 1.64m



Analysis: Camera Calibration



(a)

(b)

— ground truth ··· DEEPHORIZON - - - Upright ··· ours

Calib dataset*

	pitch (°)	roll (°)	field of view (°)
Hold-Geoffroy et al, 18	2.11±3.10	1.19±1.89	4.39±3.67
Ours (3 layers)	1.83±2.64	1.02±1.46	3.61±3.21
Ours (3 layers w/ keypoints)	1.82±2.62	1.05±1.94	3.63±3.22

Summary

Single View Metrology with **absolute** scale



Input Image in the Wild

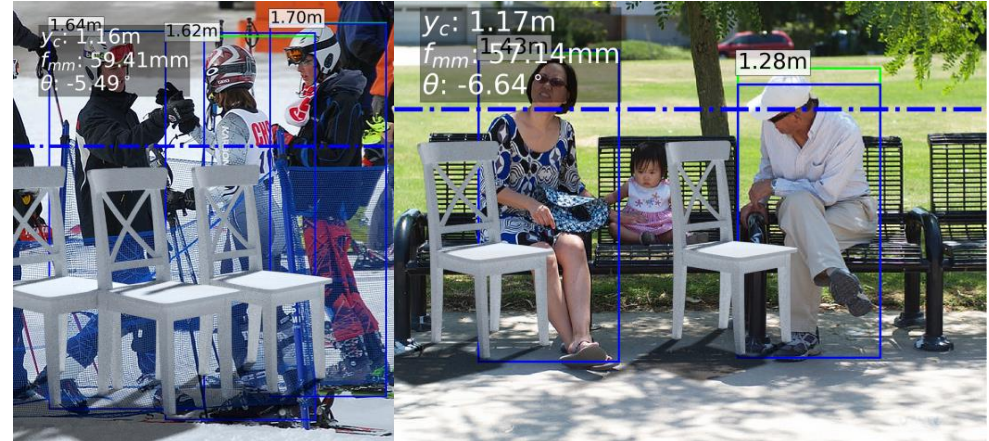
Camera Calibration

- Orientation
- Field of View

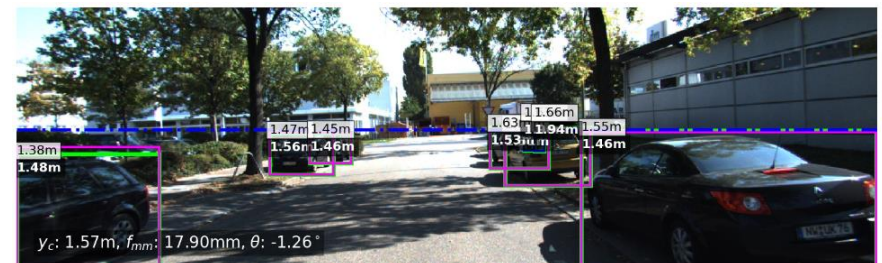
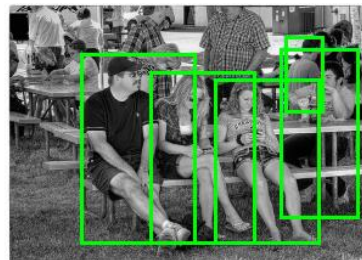
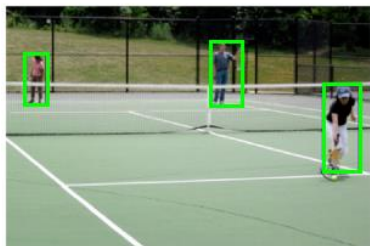
Camera Height Estimation

Object Heights Estimation

Scale-consistent virtual object insertion in the wild



Datasets and benchmarks



Single View Metrology in the Wild

Rui Zhu¹, Xingyi Yang¹, Yannick Hold-Geoffroy², Federico Perazzi², Jonathan Eisenmann²,
Kalyan Sunkavalli², Manmohan Chandraker¹

University of California San Diego¹, Adobe Research²



UC San Diego
JACOBS SCHOOL OF ENGINEERING

