

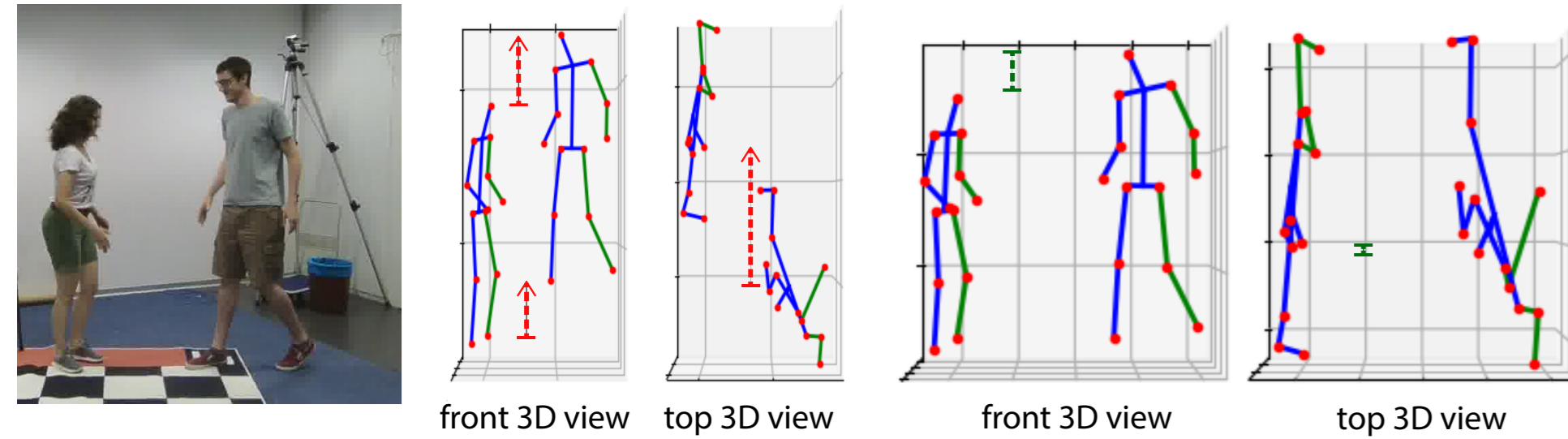
Semih Günel¹, Helge Rhodin^{1,2} and Pascal Fua¹¹Computer Vision Lab, EPFL, Lausanne, Switzerland²UBC, Vancouver, Canada

Problem

- 3D Pose estimation methods not do not recover the **correct scales** (marked in red). Correct scales are shown in green.

Input image

LCRNet output



- Current datasets do have enough subject variability with known height to generalize on height (or scale) prediction task.

Dataset	H3.6M	HumanEva	MPII-INF-3DHP	IMDB-100K (Ours)
# Subjects	11	4	8	12,104

- The field needs a better dataset with more subjects with known heights. **We provide IMDB-273K dataset to the community.**

Dataset Collection (IMDB)

- Identity matching using SOTA face recognition using IMDB profile image. Matching propagates the height information, which is taken from IMDB.

Match:

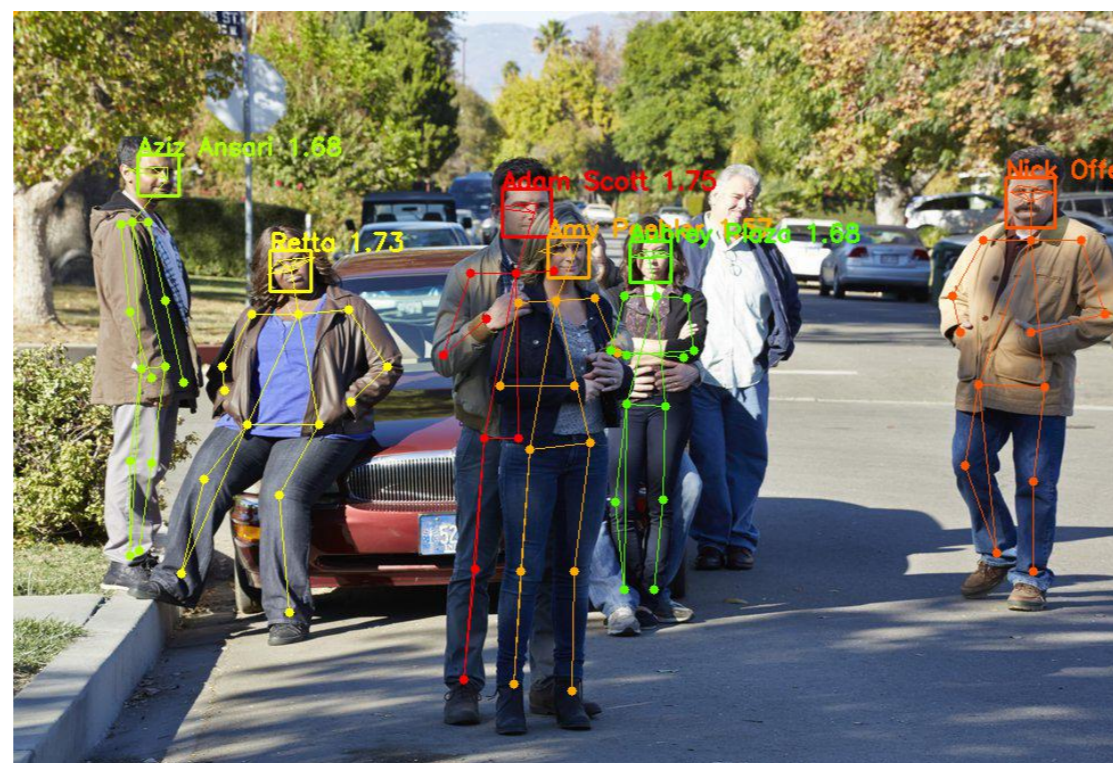
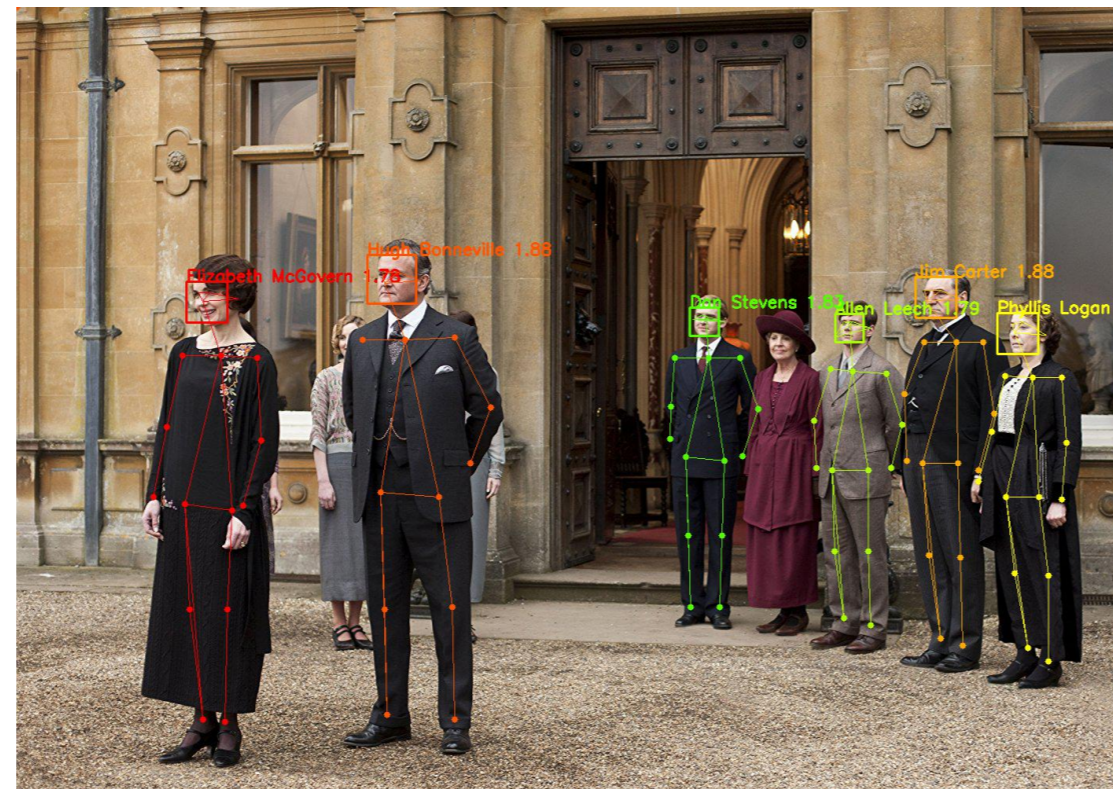
Source(IMDB):



Height: 6' 1" (1.86 m)

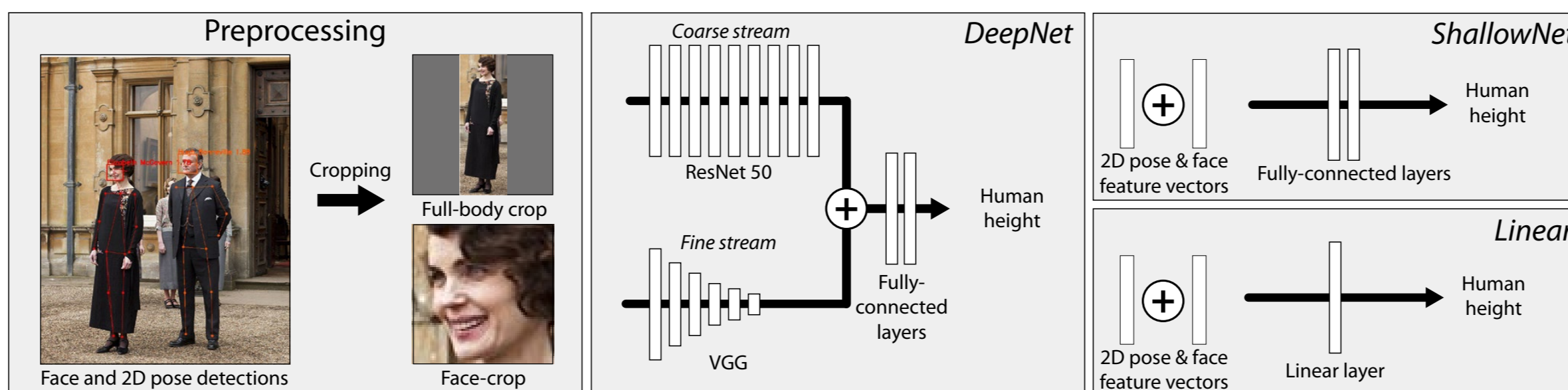


Dataset



- Includes **12,104 subjects** with known height. This is **three magnitudes larger** than current datasets!
- 273K images** with at least a person with known height!
- Calculated pose, identities and bounding boxes for **1,000,000 images**.

Network



Evaluation (ground-truth/prediction)



Results

- Larger datasets improve the results** (Figure 2).
- However, **scale-depth ambiguity continues in the age of deep learning** (Figures 1&3).
- Future work: combine shape and segmentation tasks into height prediction.

Method	IMDB-100K			Lab-test
	all	women	men	all
ConstantMean	8.25	7.46	9.22	11.0
GenderPred	6.61	6.28	7.12	9.26
PoseNet [20]	-	-	-	10.65
DeepNet (ours)	6.14	5.88	6.40	9.13
GenderMean	5.91	5.63	6.23	8.66
DeepNet (gender-specific)	5.56	5.23	6.03	8.53

(a)

Input features	Regression type		
	Linear	ShallowNet	DeepNet
Body crop only	7.56 / 11.10	7.10 / 10.40	6.40 / 9.43
Face crop only	6.49 / 10.25	6.31 / 9.99	6.25 / 8.87
Body and Face	6.40 / 10.2	6.29 / 9.92	6.14 / 9.13

(b)

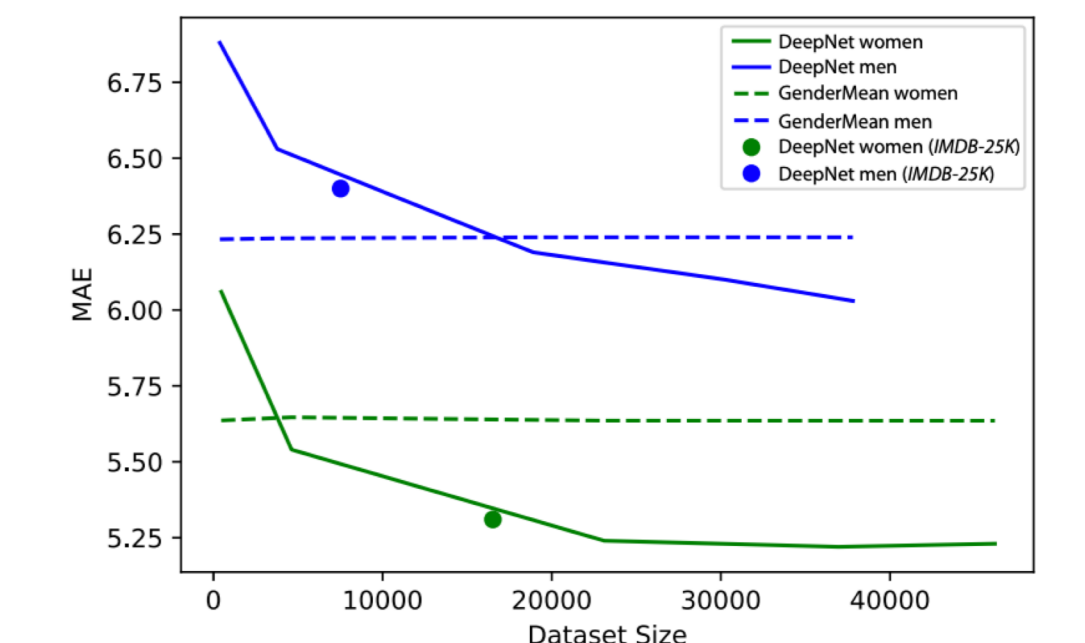


Figure 2

Figure 1

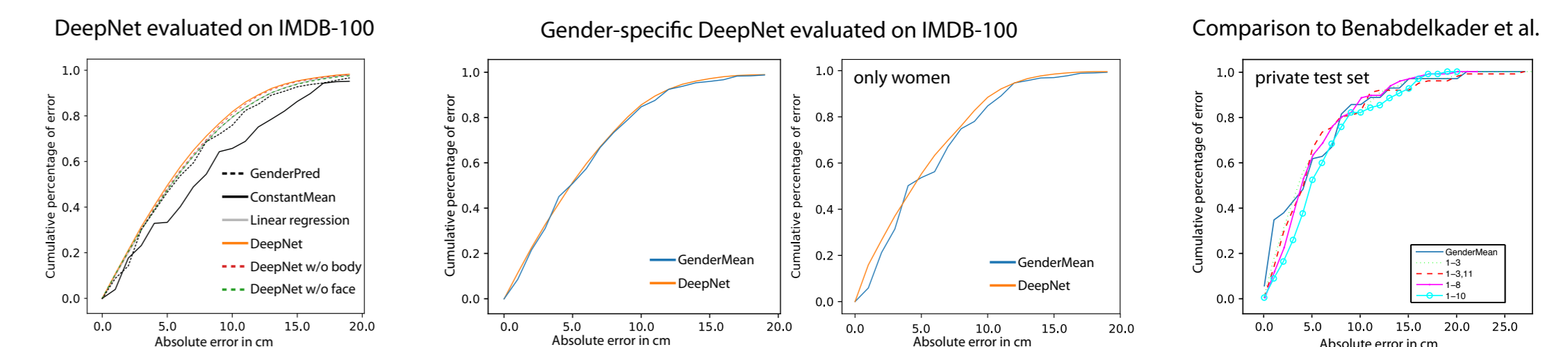


Figure 3