COMPRESSED VOLUMETRIC HEATMAPS FOR MULTI-PERSON 3D POSE ESTIMATION

Matteo Fabbri[1] Fabio Lanzi[1] Simone Calderara[1] Stefano Alletto[2] Rita Cucchiara[1]

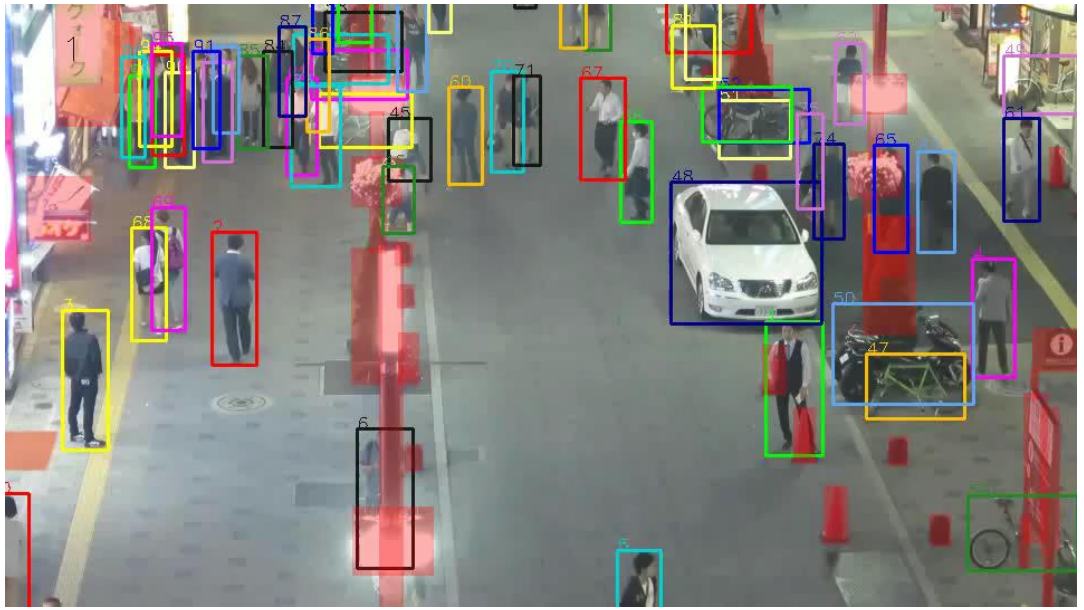
[1]University of Modena and Reggio Emilia
{name.surname}@unimore.it

[2]Panasonic R&D Company of America
{name.surname}@us.panasonic.com

CVPR SEATTLE WASHINGTON JUNE 16-18 2020

PEOPLE TRACKING

- It is defined as the problem of estimating the trajectory of multiple people in an image plane as it moves around a scene.
- We are addressing a 3D problem using 2D data
- It’s hard to annotate 3D data, especially in surveillance scenarios



Anton Milan, Laura Leal-Taixe, Ian Reid, Stefan Roth, Konrad Schindler.
MOT16: A Benchmark for Multi-Object Tracking, arXiv 2017

JTA DATASET

AImage Lab

- Let's utilize a videogame:
  - Free Annotation
  - No errors
  - Photorealism

- 512 Full HD videos 30s long
- 10M people with 3D annotation

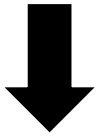
Matteo Fabbri, Fabio Lanzi, Simone Calderara, Andrea Palazzi, Roberto Vezzani, Rita Cucchiara.
Learning to Detect and Track Visible and Occluded Body Joints in a Virtual World, ECCV 2018

3D POSE ESTIMATION

AImage Lab

- **Single person** -> Coordinate regression

- **Multi person** -> People Detection + Coordinate regression



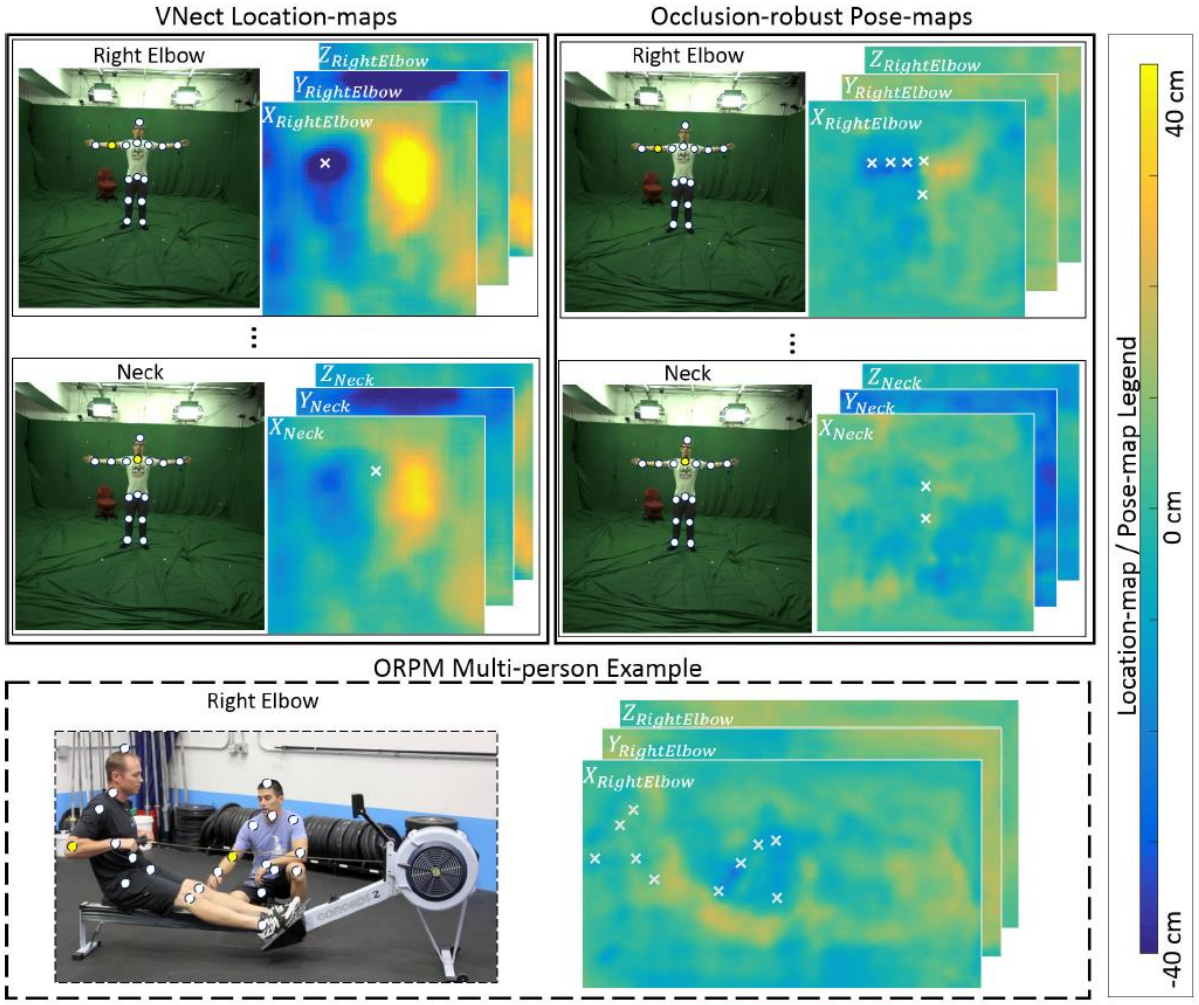
This approach does not scale with the number of people!

We need a bottom-up approach.

BOTTOM-UP 3D POSE ESTIMATION

AImage Lab

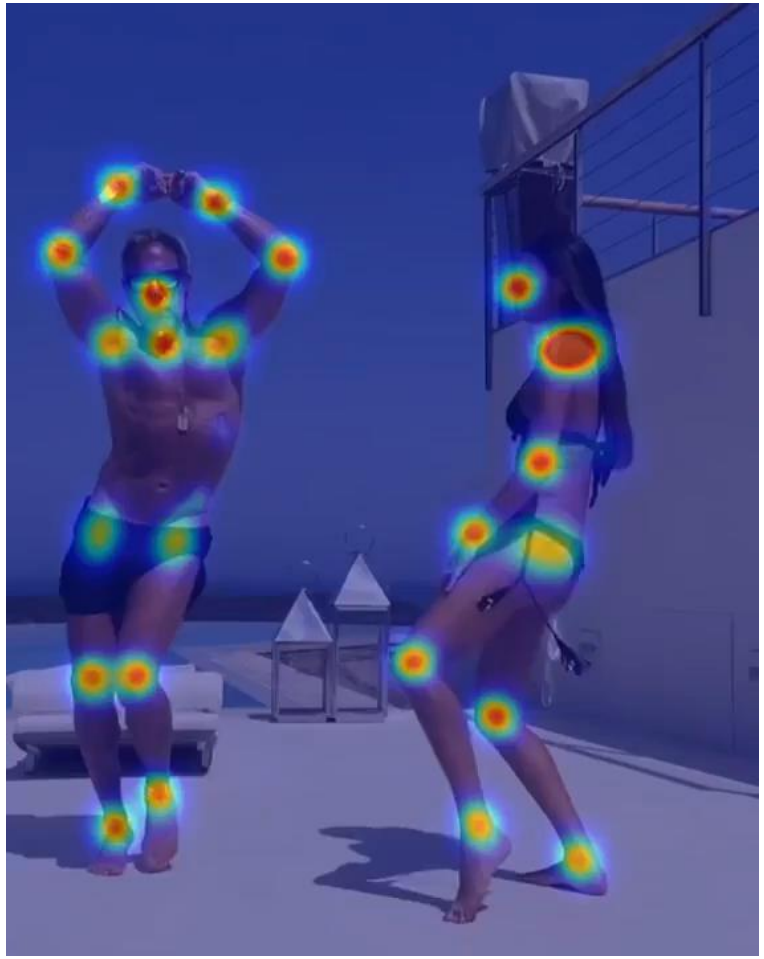
- Few methods in literature
- Location Maps are not suitable in surveillance
- What about 2D pose estimation?



2D POSE ESTIMATION

AImage Lab

- For 2D pose estimation Heatmaps are the best choice
- What about 3D heatmaps?



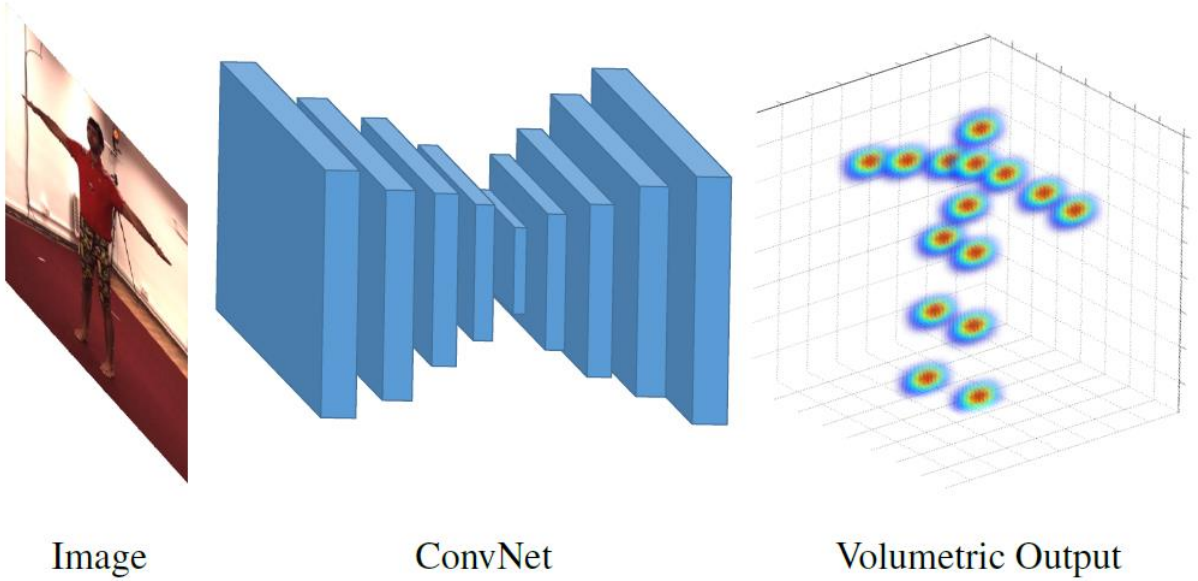
VOLUMETRIC HEATMAPS

AImage Lab

- Only for single person

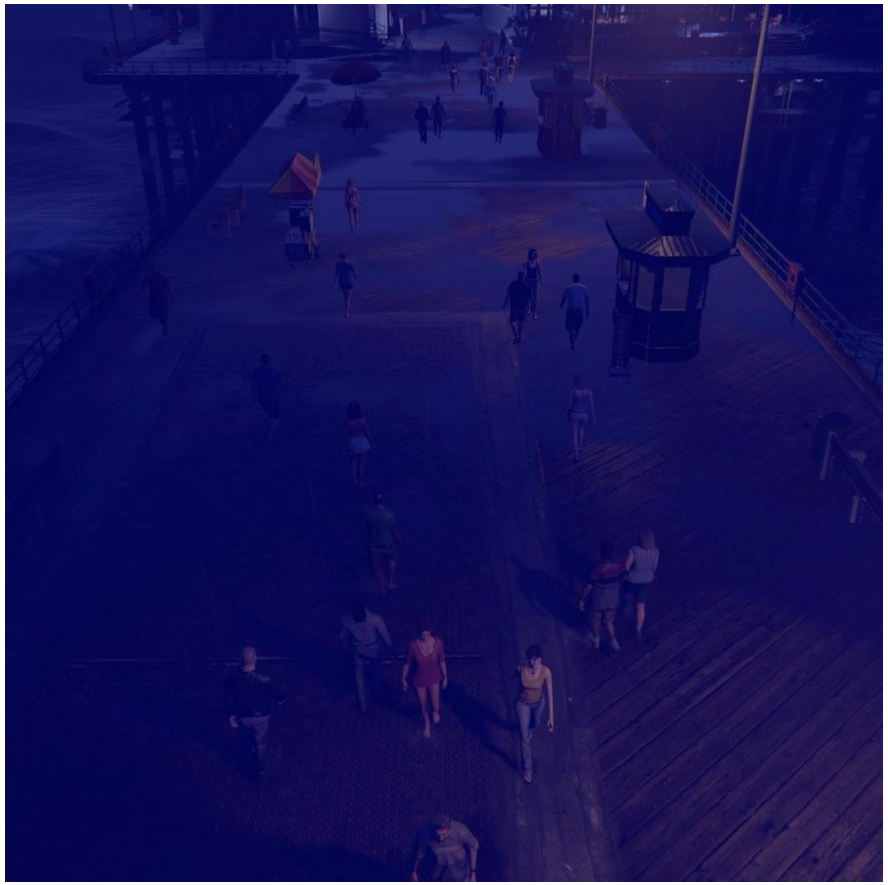
- **Memory and computational demanding**
- **Sparse signal**

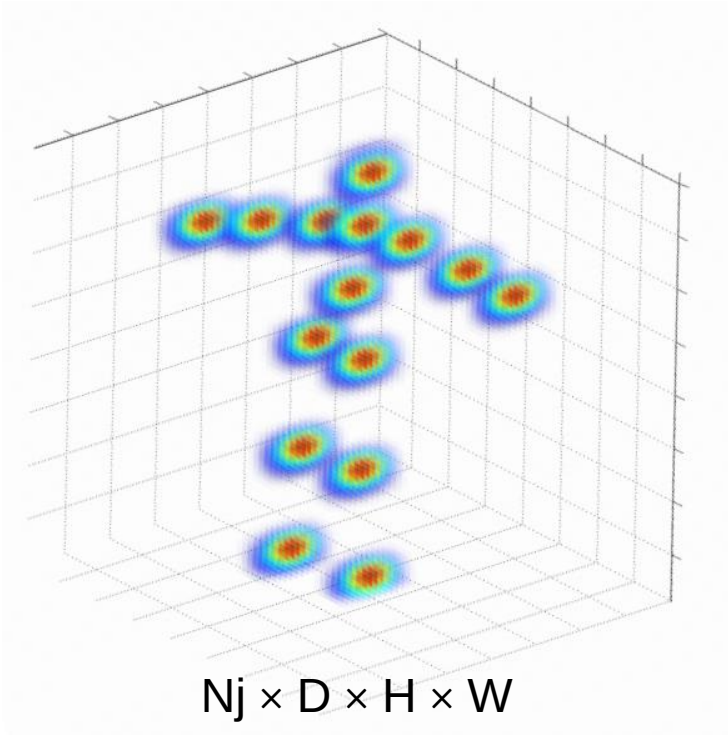
- Some compromise:
  - low resolution heatmaps that introduce quantization errors
  - complex training strategies that involve coarse-to-fine predictions



VOLUMETRIC HEATMAPS FOR MULTI PERSON

AImage Lab





$Nj \times D \times H \times W$

Full HD images with stride 8 and 10cm quantization

$14 \times 1000 \times 128 \times 240$

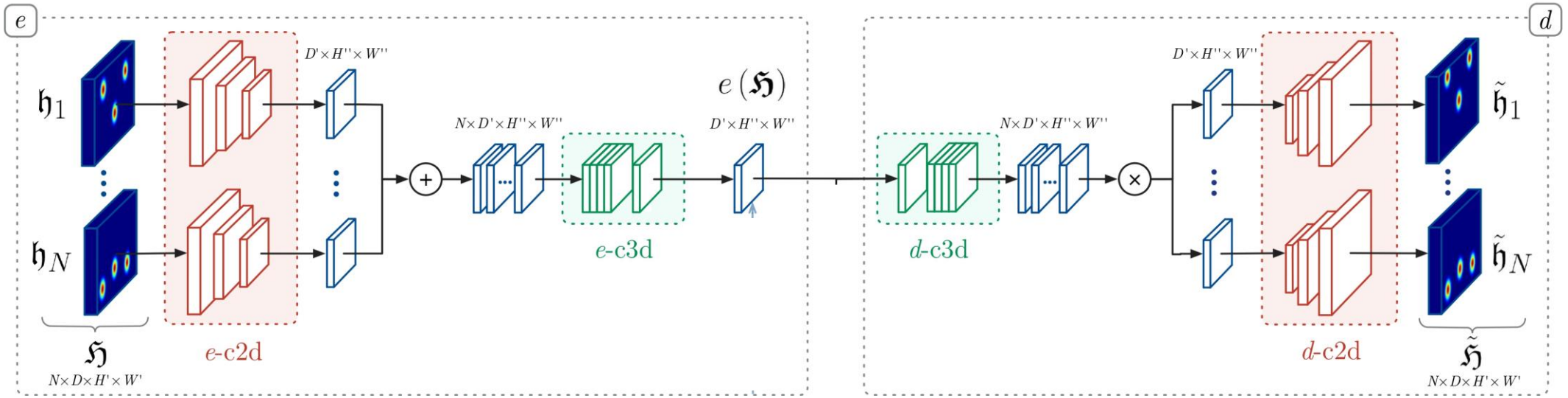
COMPRESSION

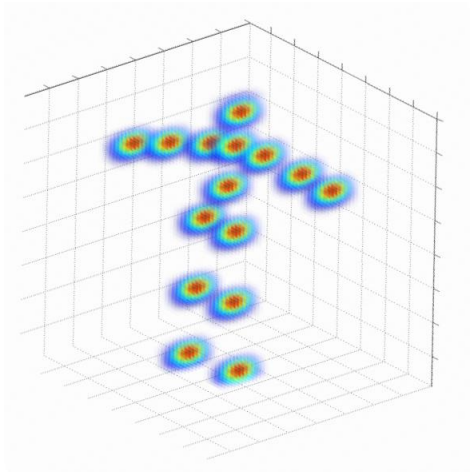
AImage Lab

- Let’s compress the Volumetric Heatmaps!
- Sparse data representation -> easy to compress
- RLE?
- Deep compression?

AUTOENCODER

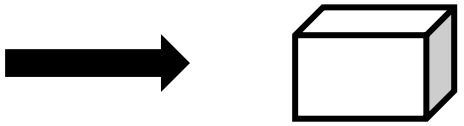
AImage Lab





14 × 316 × 128 × 240

~7000 smaller



1 × 39 × 17 × 30

OUR SOLUTION: THE **LOCO** APPROACH

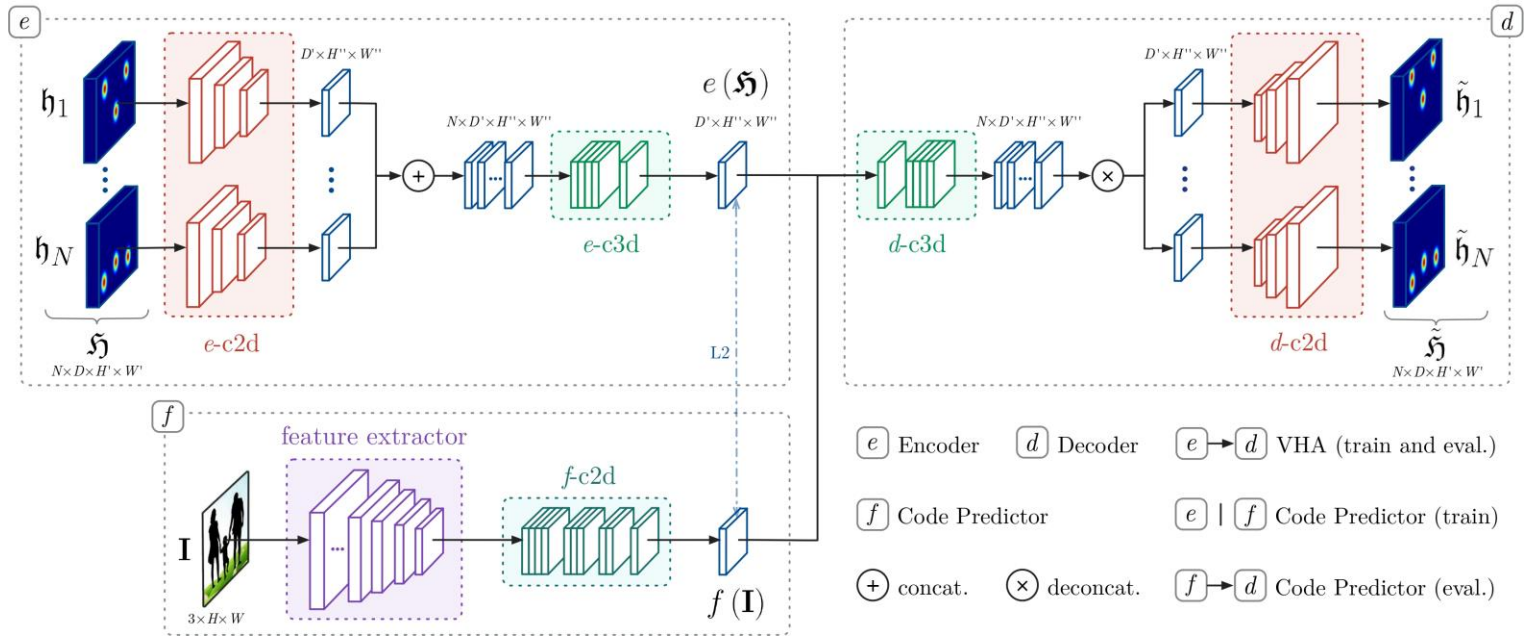
AImage Lab

We propose a simple method that maps high-resolution volumetric heatmaps to a **compact** and **more tractable** representation.

To this end, we use an **Autoencoder** to compress those **Volumetric Heatmaps** into a **Code** that must:

- **preserve** the gaussian **peaks** at every joint location
- be **smaller** w.r.t. the full resolution heatmaps

This representation enables, for the **very first** time, the use of volumetric heatmaps to tackle 3D Human Pose Estimation in a bottom-up fashion.



COMPRESSION LEVELS

AImage Lab

model	bottleneck size	F1 on JTA @0vx	F1 on JTA @1vx	F1 on JTA @2vx
VHA$^{(1)}$	$\frac{D}{2} \times \frac{H'}{2} \times \frac{W'}{2}$	97.1	98.4	98.5
VHA$^{(2)}$	$\frac{D}{4} \times \frac{H'}{4} \times \frac{W'}{4}$	92.5	97.0	97.1
VHA$^{(3)}$	$\frac{D}{8} \times \frac{H'}{8} \times \frac{W'}{8}$	56.5	90.3	92.9

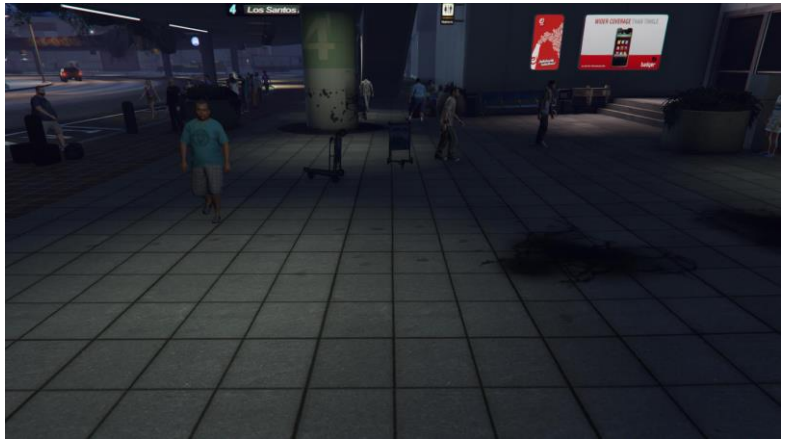
QUANTITATIVE RESULTS

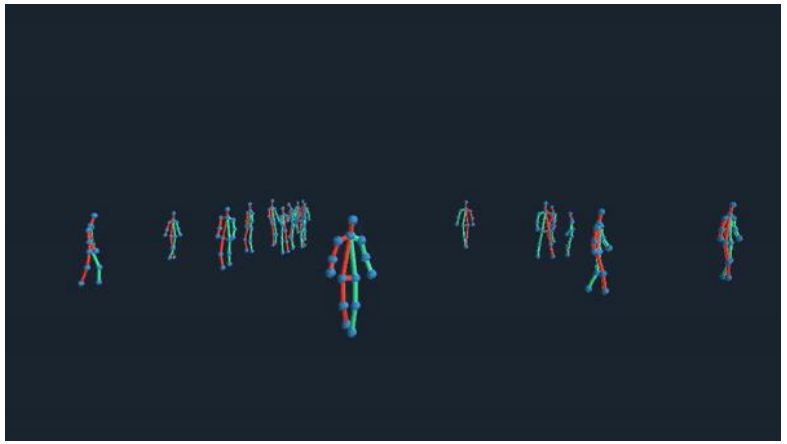
AImage Lab

	PR	RE	F1	PR	RE	F1	PR	RE	F1
	@0.4 m			@0.8 m			@1.2 m		
Location Maps [21, 22]	5.80	5.33	5.42	24.06	21.65	22.29	41.43	36.96	38.26
Location Maps [21, 22] + ref.	5.82	5.89	5.77	23.28	23.51	23.08	38.85	39.17	38.49
[33] + [19]	**75.88**	28.36	39.14	**92.85**	34.17	47.38	**96.33**	35.33	49.03
Uncompr. Volumetric Heatmaps	25.37	24.40	24.47	45.40	43.11	43.51	55.55	52.44	53.08
LoCO$^{(1)}$	48.10	42.73	44.76	65.63	58.58	61.24	72.44	64.84	67.70
LoCO$^{(1)}$+.	49.37	43.45	45.73	66.87	59.02	62.02	73.54	65.07	68.29
LoCO$^{(2)}$	54.76	46.94	50.13	70.67	60.48	64.62	77.00	65.92	70.40
LoCO$^{(2)}$+.	55.37	**47.84**	**50.82**	70.63	**60.94**	**64.76**	76.81	**66.31**	**70.44**
LoCO$^{(3)}$	48.18	41.97	44.49	66.96	58.22	61.77	74.43	64.71	68.65
LoCO$^{(3)}$+.	49.15	42.84	45.36	67.16	58.45	61.92	74.39	64.76	68.57
GT Location Maps [21, 22]	76.07	64.83	69.59	76.07	64.83	69.59	76.07	64.83	69.59
GT Volumetric Heatmaps	99.96	99.96	99.96	99.99	99.99	99.99	99.99	99.99	99.99

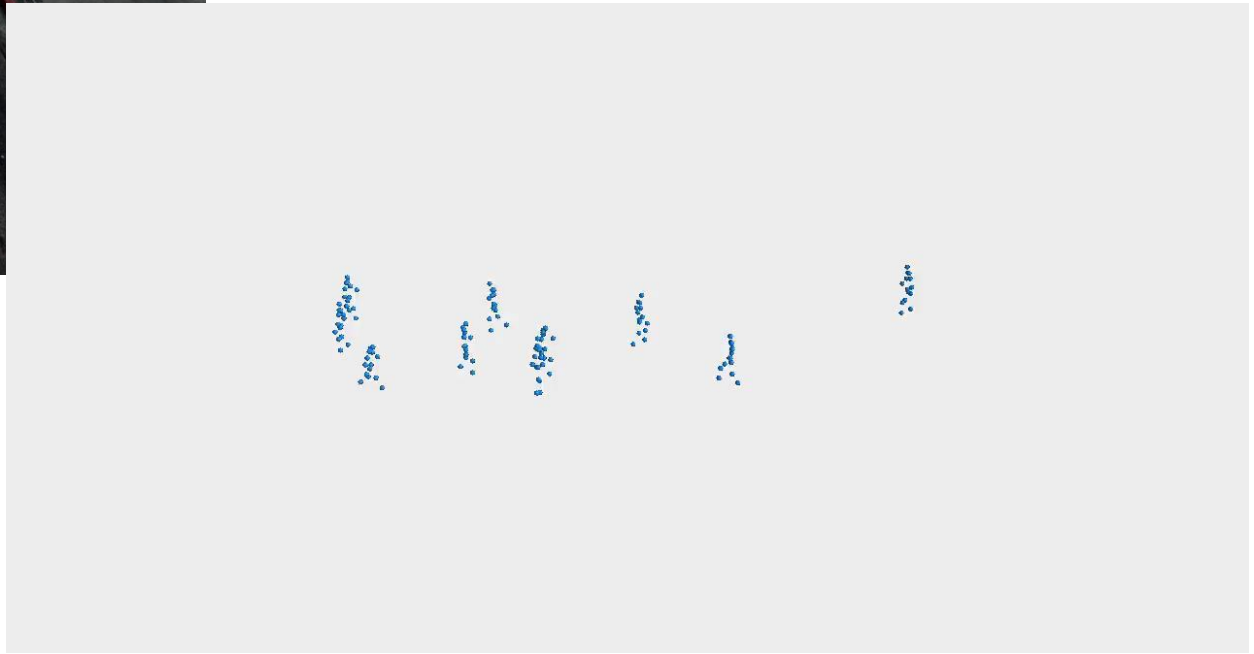
QUALITATIVE RESULTS

AImage Lab

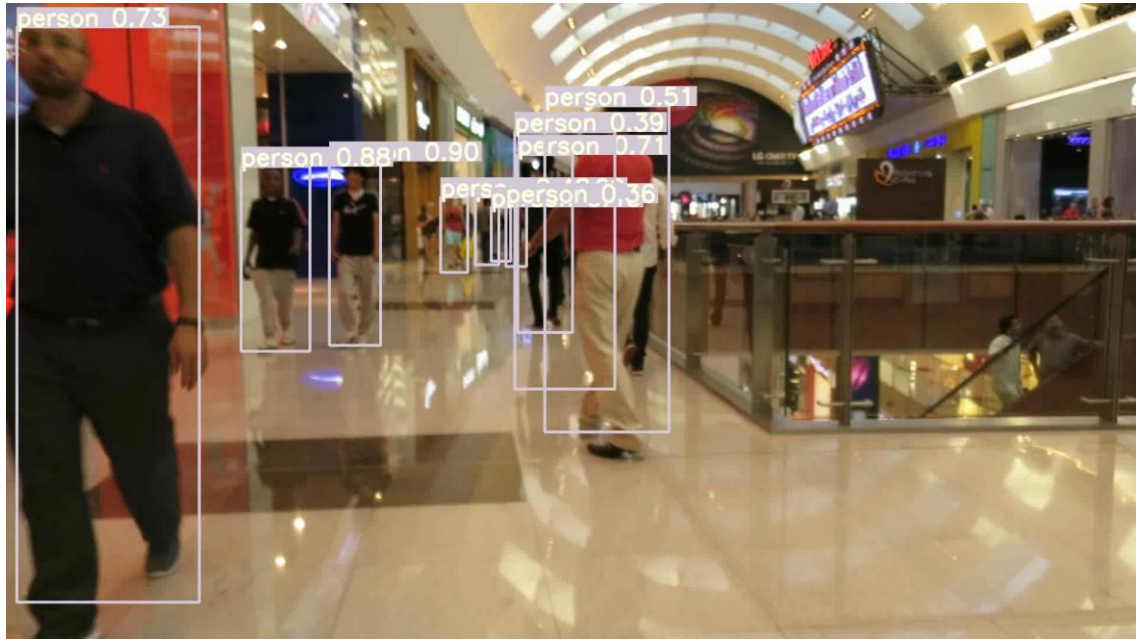




QUALITATIVE RESULTS



DOES GTA GENERALIZE TO REAL WORLD?

AImage Lab



For people detection, YES!

THANK YOU!

If you have questions, please contact me

matteo.fabbri@unimore.it

Matteo Fabbri[1]

Fabio Lanzi[1]

Simone Calderara[1]

Stefano Alletto[2]

Rita Cucchiara[1]

[1] AImagelab, AI Research & Innovation Center, University of Modena and Reggio Emilia, Italy

[2] Panasonic R&D Company of America