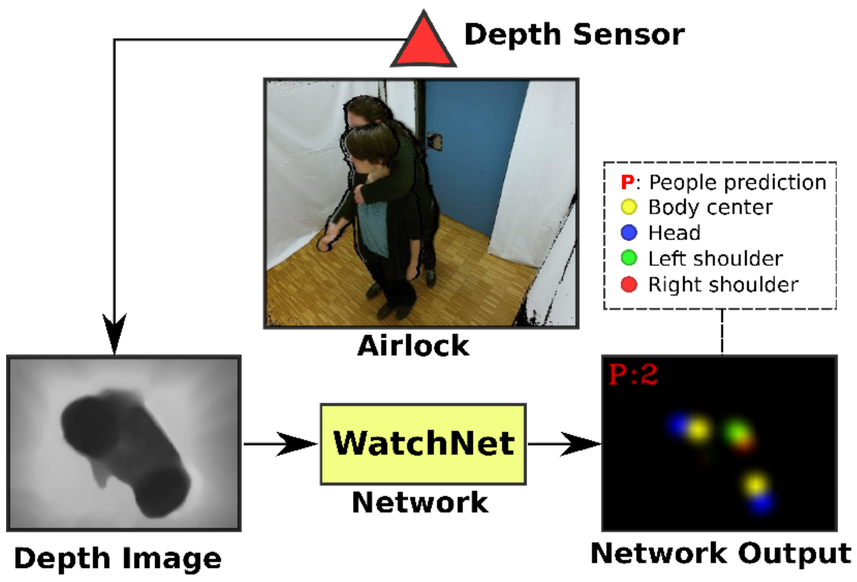
WatchNet: Efficient and Depth-based Network for People Detection in Video Surveillance Systems

M. Villamizar, A. Martinez-Gonzalez, O. Canevet and J-M. Odobez

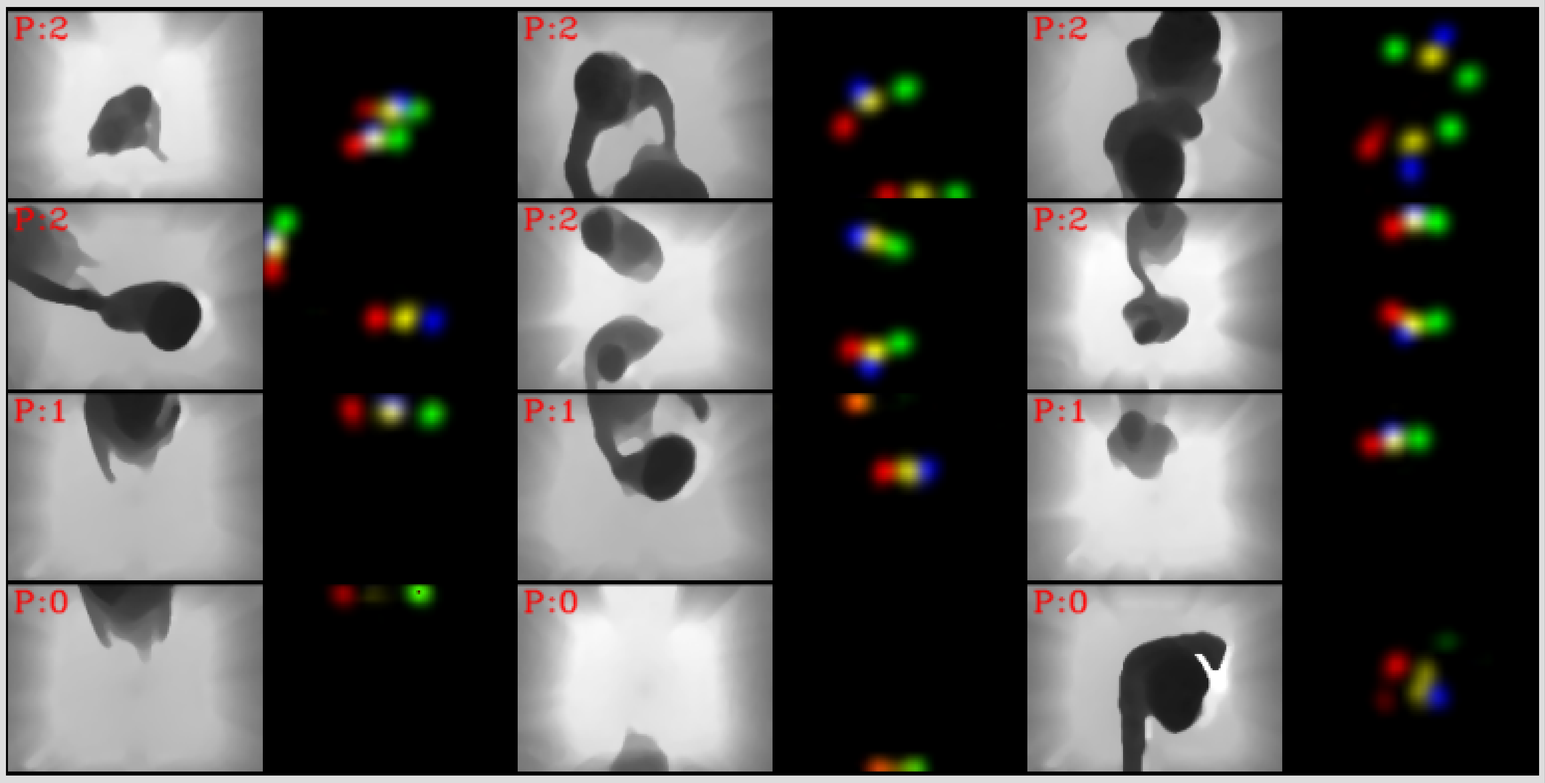
Idiap Research Institute - Switzerland

Goal: Detecting people attacks and intrusion inside security airlocks.



Proposed method: Convolutional network (*WatchNet*) to predict the number of people in airlocks by detecting body landmarks (head and shoulders) and body center.

Detection Results: Output of WatchNet for people detection in airlocks using top-view depth sensors.



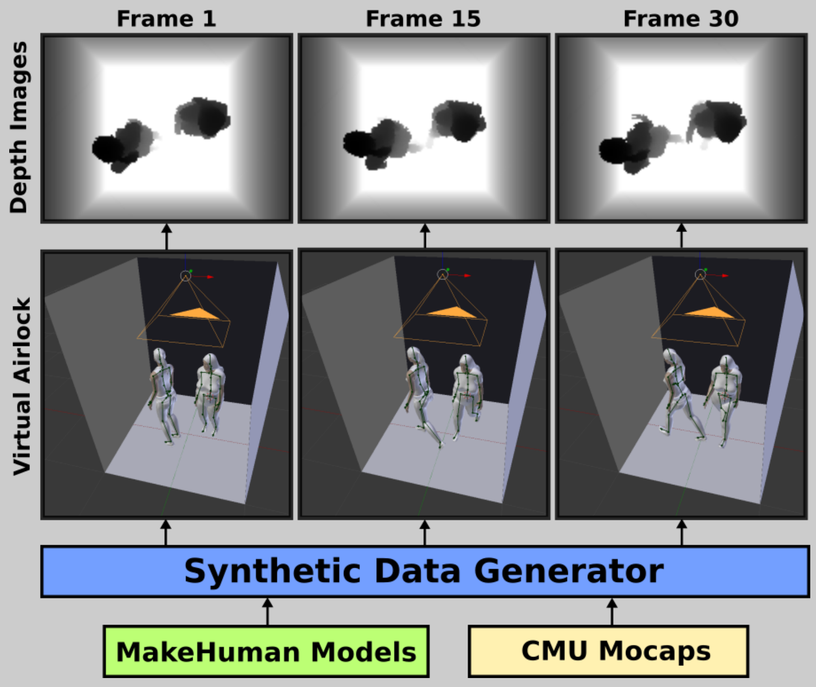
Depth-Image Datasets:

Synthetic Dataset:
- About 80k top-view depth images.
- Generated via Blender using MakeHuman models and CMU mocap sequences.
- Automatic body landmark annotations (head and shoulders).
- Used for training the network.

Real Dataset:
- UNICITY dataset [AVSSW 2018] .
- About 58k top-view depth images.
- Annotations made manually.
- Used for fine tuning and evaluation.

Synthetic Dataset:

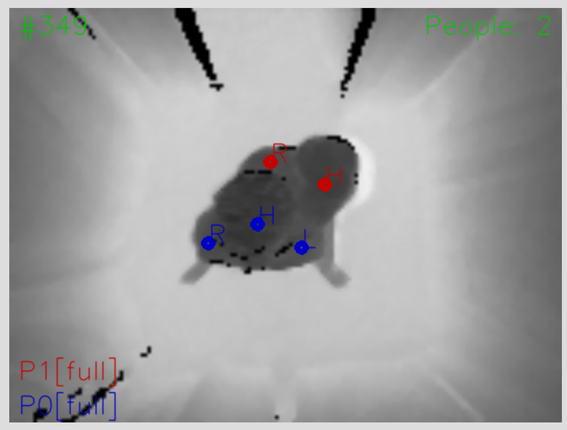
Pipeline to create synthetic depth images



Real Dataset: UNICITY Dataset

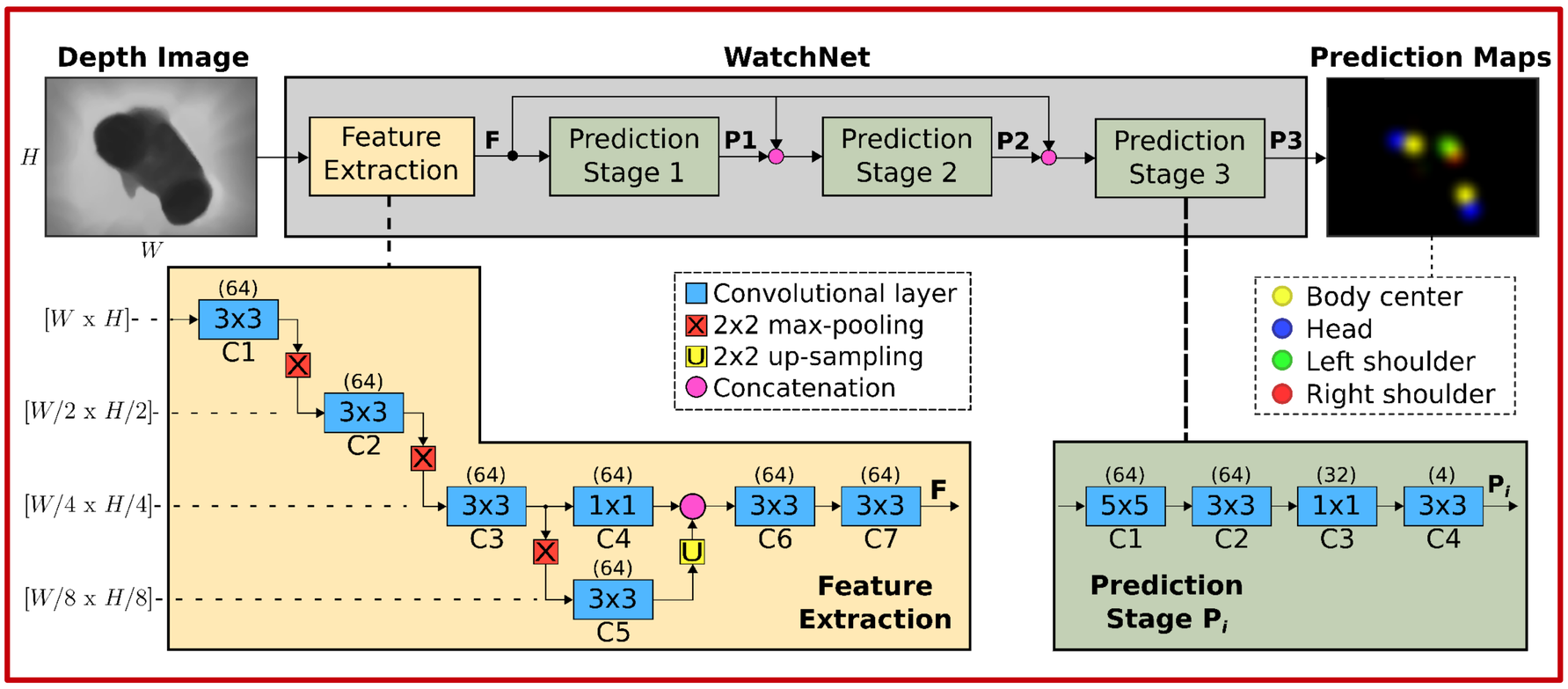
- Five levels of people visibility: *invisible*, *difficult*, *truncated*, *partial* and *full*.
- Four levels of evaluation:

Level 1: invisible + full
Level 2: invisible + partial + full
Level 3: invisible + truncated + partial + full
Level 4: invisible + difficult + truncated + partial + full



WatchNet:

- Fully convolutional network.
- WatchNet consists of a feature extraction module followed by three prediction stages.
- Feature extraction computes discriminative features that are shared to every prediction stage.
- Every prediction stage provides confidence maps with the location of head and shoulders and body center.
- WatchNet refines predictions sequentially adding context and previous predictions.



Exp 1: Synthetic and real data for training the network

	R	P	F	A	TP	TN	FN	FP	R	P	F	A	TP	TN	FN	FP
	Synthetic Data								Synthetic+Real Data							
Level 1	0.92	1.00	0.96	0.99	531	4098	49	0	0.99	1.00	1.00	1.00	576	4097	4	1
Level 2	0.83	0.98	0.90	0.95	1367	4867	274	34	0.96	1.00	0.98	0.99	1574	4894	67	7
Level 3	0.64	0.97	0.77	0.90	1512	6649	865	48	0.82	1.00	0.90	0.95	1953	6688	424	9
Level 4	0.48	0.97	0.64	0.85	1543	8083	1698	48	0.63	1.00	0.77	0.89	2050	8122	1191	9
Average	0.72	0.98	0.82	0.92	1238	5924	721	32	0.85	1.00	0.91	0.96	1538	5950	421	6

Exp 2: Comparing WatchNet against other approaches

	Baseline				FCN				WatchNet [Not multi-scale]				WatchNet			
	R	P	F	A	R	P	F	A	R	P	F	A	R	P	F	A
Level 1	0.97	0.55	0.70	0.90	0.92	0.99	0.96	0.99	0.95	0.99	0.97	0.99	0.99	1.00	1.00	1.00
Level 2	0.96	0.74	0.84	0.91	0.87	0.98	0.92	0.96	0.89	0.99	0.94	0.97	0.96	1.00	0.98	0.99
Level 3	0.88	0.79	0.83	0.91	0.74	0.98	0.84	0.93	0.78	0.99	0.87	0.94	0.82	1.00	0.90	0.95
Level 4	0.72	0.81	0.76	0.87	0.56	0.98	0.71	0.87	0.59	0.99	0.74	0.88	0.63	1.00	0.77	0.89
Average	**0.88**	0.72	0.78	0.90	0.77	0.99	0.86	0.94	0.80	0.99	0.88	0.95	0.85	**1.00**	**0.91**	**0.96**

Exp 3: Number of prediction stages

	1 Stage		3 Stages		5 Stages	
	F	A	F	A	F	A
Level 1	0.97	0.99	1.00	1.00	0.97	0.99
Level 2	0.95	0.98	0.98	0.99	0.96	0.98
Level 3	0.86	0.94	0.90	0.95	0.88	0.94
Level 4	0.72	0.87	0.77	0.89	0.74	0.88
Average	0.88	0.95	**0.91**	**0.96**	0.89	0.95

Exp 4: People counting method

	Body Center		Head		Head & Shld	
	F	A	F	A	F	A
Level 1	1.00	1.00	1.00	1.00	0.93	0.98
Level 2	0.98	0.99	0.93	0.97	0.96	0.98
Level 3	0.90	0.95	0.81	0.92	0.89	0.95
Level 4	0.77	0.89	0.67	0.86	0.76	0.89
Average	**0.91**	**0.96**	0.85	0.94	0.88	0.95