Sparse and Privacy-enhanced Representation for Human Pose Estimation

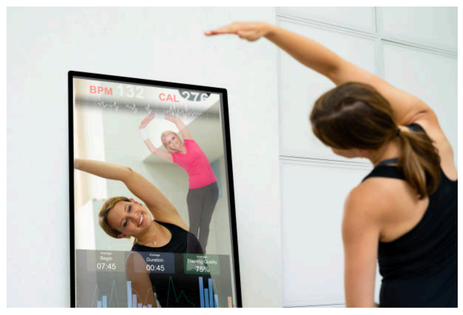
Ting-Ying Lin[1*], Lin-Yung Hsieh[1*], Fu-En Wang[1], Wen-Shen Wuen[2], Min Sun[1]

[1]National Tsing Hua University [2]Novatek Microelectronics Corp.

BMVC 2023

Motivation

- Human pose estimation (HPE) has great potential in
  - personal fitness
  - human activity recognition in nursing homes, hospitals, etc



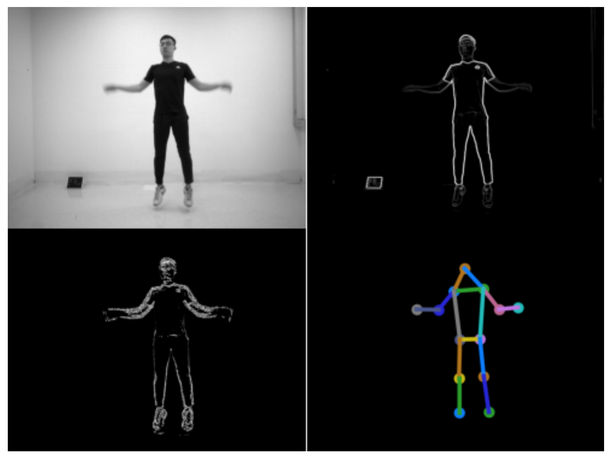
Challenges

- Edge devices have limited computational resources.
- Privacy must be maintained in healthcare and fitness facilities.

Contribution

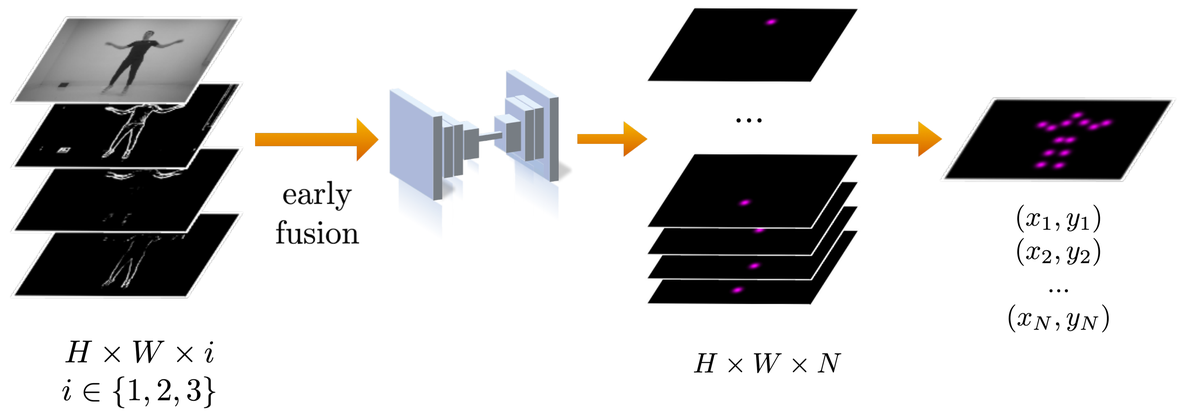
- Introduce the Sparse and Privacy-enhanced Dataset for Human Pose Estimation (**SPHP**)
- 13x acceleration in inference time and decrease FLOPs by 96% via sparse convolution

SPHP Dataset

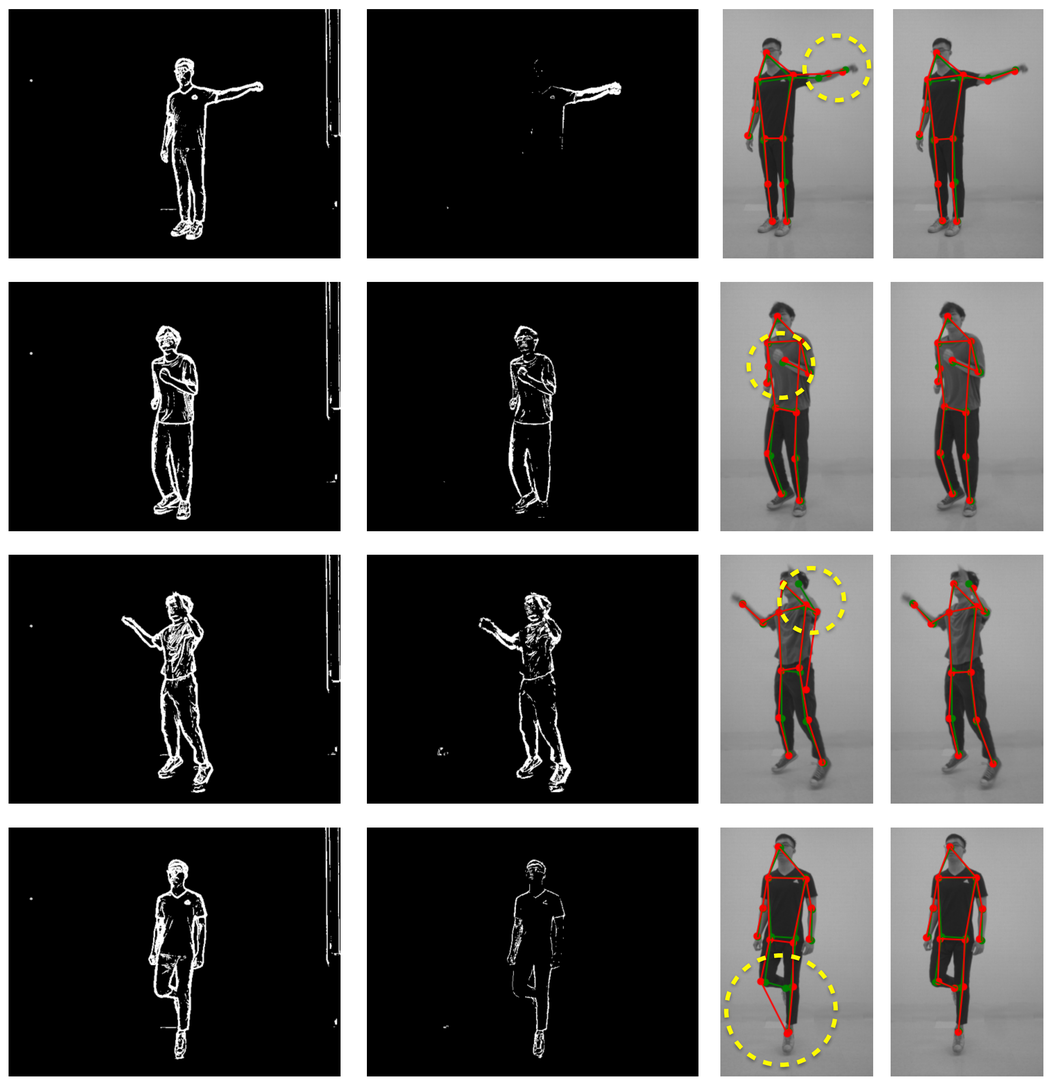


C1	C2	C3	C4
1. Arm abduction	5. Leg knee lift	8. Squat	12. Elbow-to-knee
2. Arm bicep curl	6. Leg abduction	9. Walk in place	13. Jump in place
3. Wave hello	7. Leg pulling	10. Standing side bend	14. Jumping jack
4. Punch up forward		11. Roll wrists & ankles	15. Hop on one foot
			16. Jog in place

Pipeline Overview



Qualitative Results



Acknowledgement

NSTC NARLabs National Applied Research Laboratories

Experiments

- We conduct the human pose estimation experiments on three datasets: SPHP, MMHPSD and HumanEva-I.
- We also evaluate the computation efficiency on Intel Core i9-7940 3.1GHz CPU.
- Face recognition experiments are conducted on CelebA dataset.

SPHP

$$\text{MPJPE} = \frac{1}{N}\sum_i \|y_i - \hat{y}_i\|$$

Backbone	# of Params	Input	C					SC				
			C1	C2	C3	C4	Mean	C1	C2	C3	C4	Mean
DHP19 [5]	218K	GR	2.62	3.08	3.33	3.62	3.20	-	-	-	-	-
		MV	16.96	6.50	6.43	5.11	8.66	56.96	27.12	25.86	14.12	30.20
		ED	3.14	3.64	3.71	4.03	3.65	5.18	6.47	5.94	6.76	6.10
		FS	3.36	3.32	3.56	3.88	**3.56**	5.00	6.52	6.10	6.45	**6.01**
U-Net-Small	1.9M	GR	1.82	2.08	2.13	2.48	2.15	-	-	-	-	-
		MV	19.56	5.41	5.69	3.82	8.52	54.29	20.93	19.25	8.11	24.85
		ED	3.20	3.49	3.19	3.49	3.35	3.35	3.78	3.48	3.95	3.65
		FS	3.32	2.91	2.79	3.18	**3.07**	3.42	3.61	3.41	3.69	**3.54**
U-Net-Large	7.7M	GR	1.73	2.14	2.00	2.33	2.06	-	-	-	-	-
		MV	18.76	5.27	5.54	3.79	8.25	52.08	20.00	18.65	7.77	23.86
		ED	2.95	3.08	2.86	3.27	3.05	3.32	3.70	3.47	3.86	3.60
		FS	2.68	2.94	2.79	3.14	**2.90**	3.32	3.64	3.41	3.64	**3.50**

MMHPSD

	Edge	Event	MV	Event + Edge	MV + Edge
Conv.	3.50	4.66	3.30	3.26	**2.95**
Sparse Conv.	3.84	9.75	6.91	3.28	**3.06**

HumanEVA-I

Gray	Edge	MV	Fusion
4.03	4.78	9.58	**4.42**
-	5.70	20.01	**5.37**

Computation efficiency

Backbone	Params	Conv.	GFLOPs	FPS
DHP19 [5]	218K	C	275.45	26.89
		SC	**33.25** (↓87%)	**38.88** (1.5x)
U-Net-Small	1.9M	C	1135	11.82
		SC	**46.74** (↓96%)	**36.13** (3x)
U-Net-Large	7.7M	C	4510	1.07
		SC	**186.80** (↓96%)	**13.89** (13x)

Fusion comparison

Model	Params#		FPS		MPJPE	
	EF	LF	EF	LF	EF	LF
DHP19	218K	655K	38.88	9.32	3.56	3.89
U-Net-Small	1.9M	5.8M	36.13	7.22	3.07	2.83
U-Net-Large	7.7M	23.1M	13.89	7.09	2.90	2.82

Face recognition

Input	Acc.	Drop	Recall	Drop
RGB	88.9	-	84.1	-
Grayscale	88.7	0.2	84.2	-0.1
Edge	**84.8**	4.1	**73.9**	10.2

Image from pixabay