Learning (Multi-)Human Optical Flow

Anurag Ranjan [*,♠], David T. Hoffmann [*,♠], Dimitrios Tzionas [♠], Siyu Tang [♠], Javier Romero [⧗] and Michael J. Black [♠]
[*] Contributed equally   [♠] Perceiving Systems, MPI for Intelligent Systems   [⧗] Amazon Inc.

Goal

Design an Optical Flow method that
- performs well on human motions
- is fast
- is compact

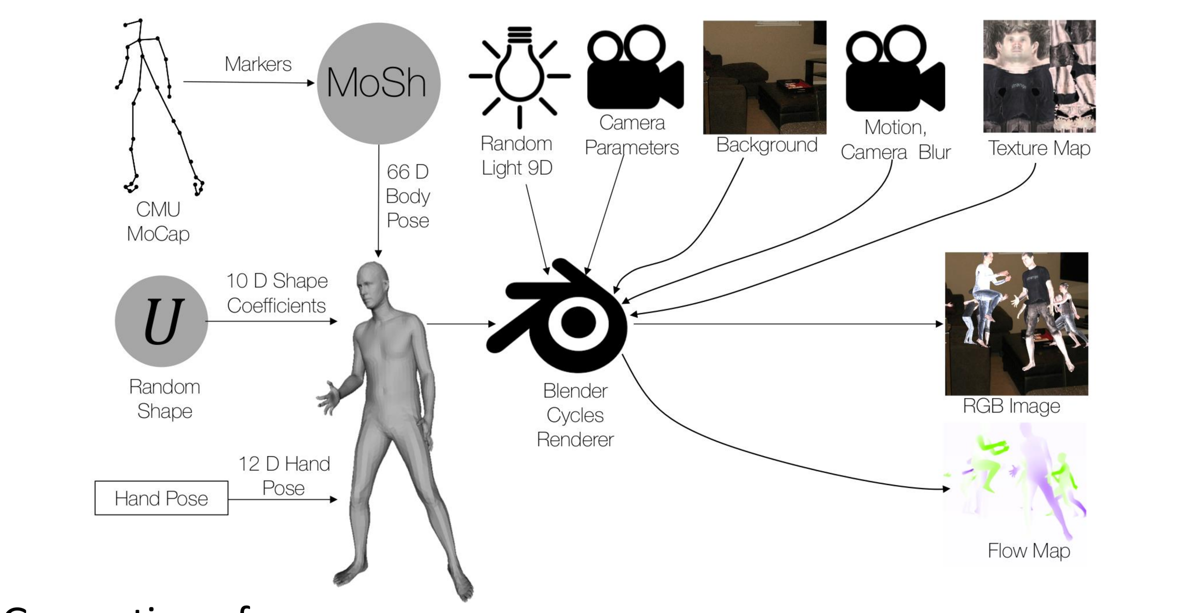
Problem

- Classical methods work but are slow
- Fast and compact deep networks do not generalize well to human motions
- No ground truth human optical flow data for training

Idea

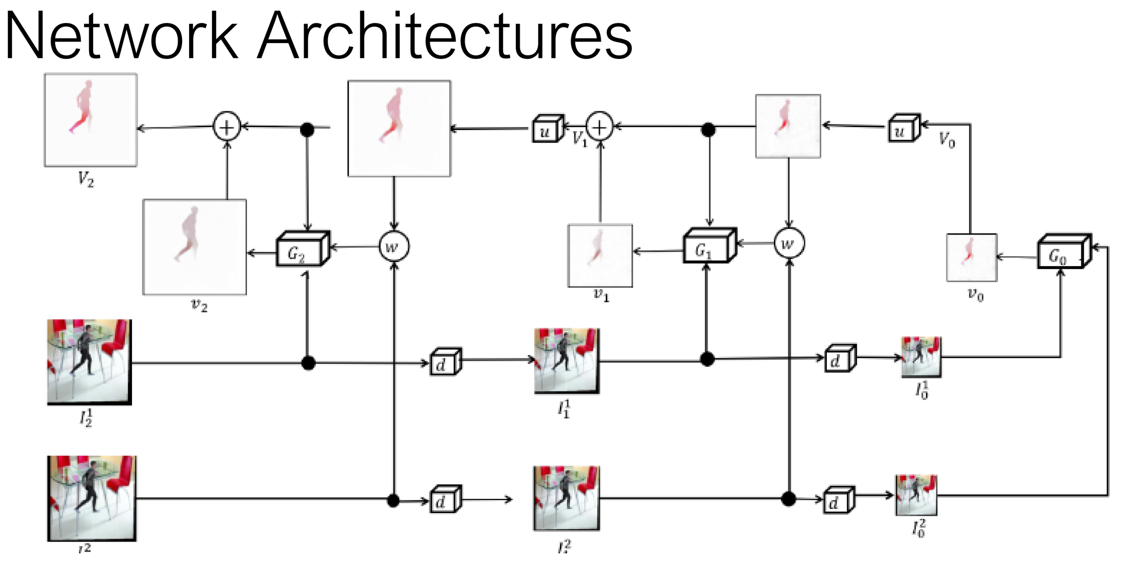
- Generate synthetic data of humans in motion
- Use SMPL [1], a realistic human body model and captured human MoCap sequences
- Use this data for training small neural networks

Data Generation Single- and Multi-Human

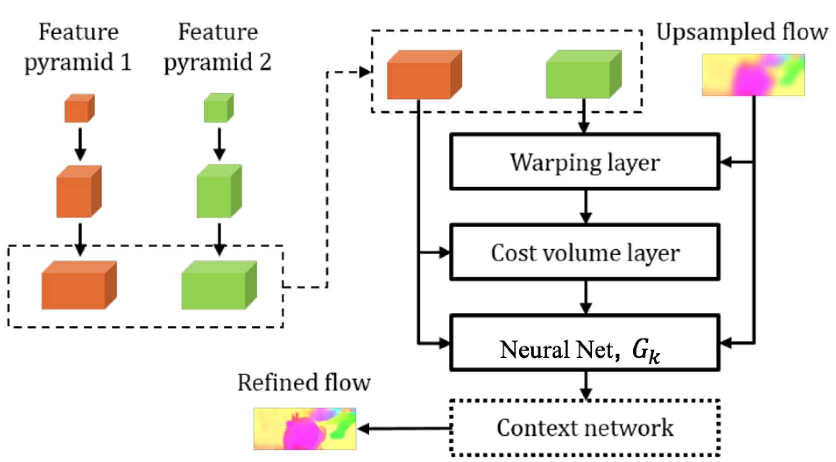


Generation of
- single-human optical flow dataset (SHOF)
- and multi-human optical flow dataset (MHOF)

Network Architectures

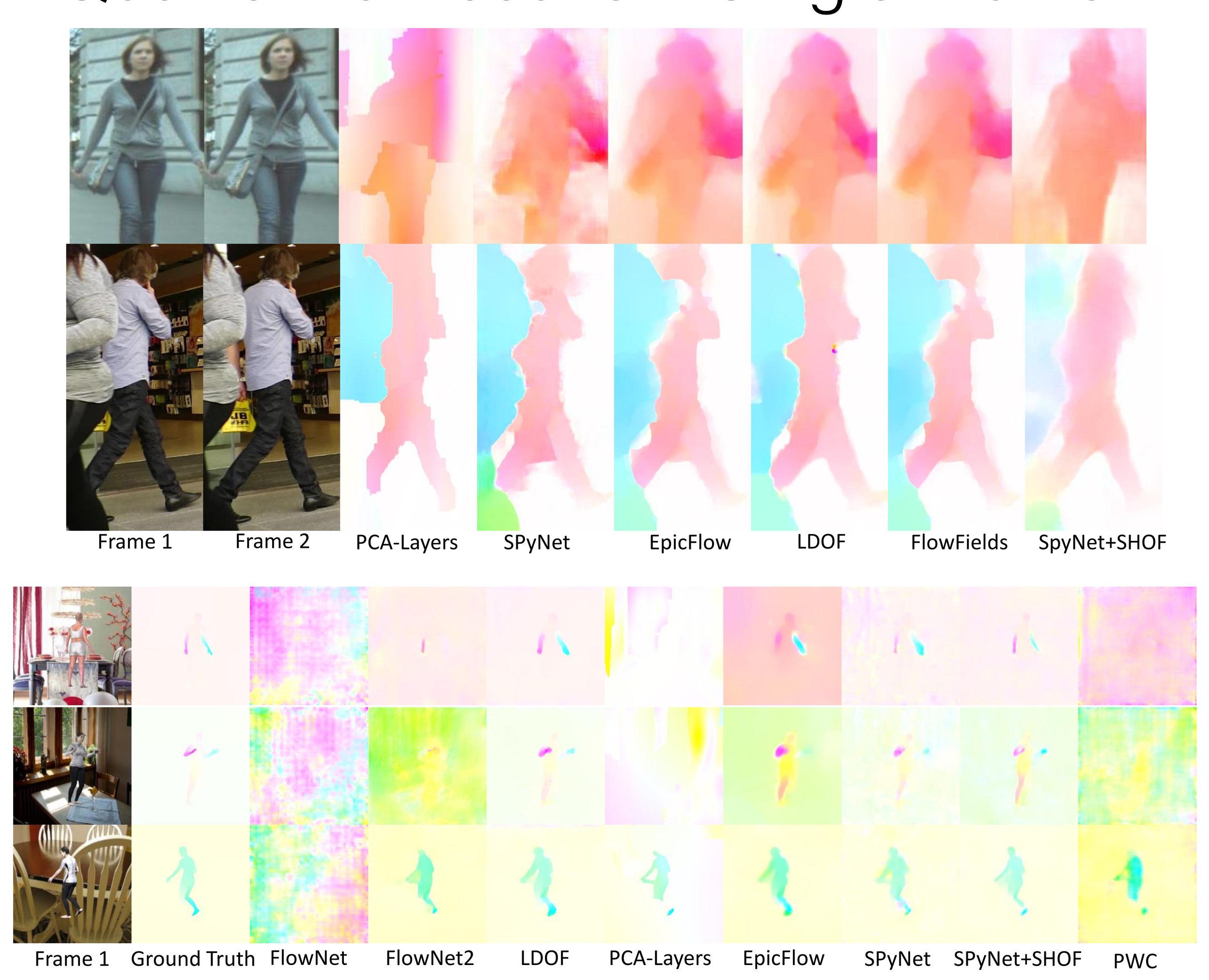


SPyNet [2]



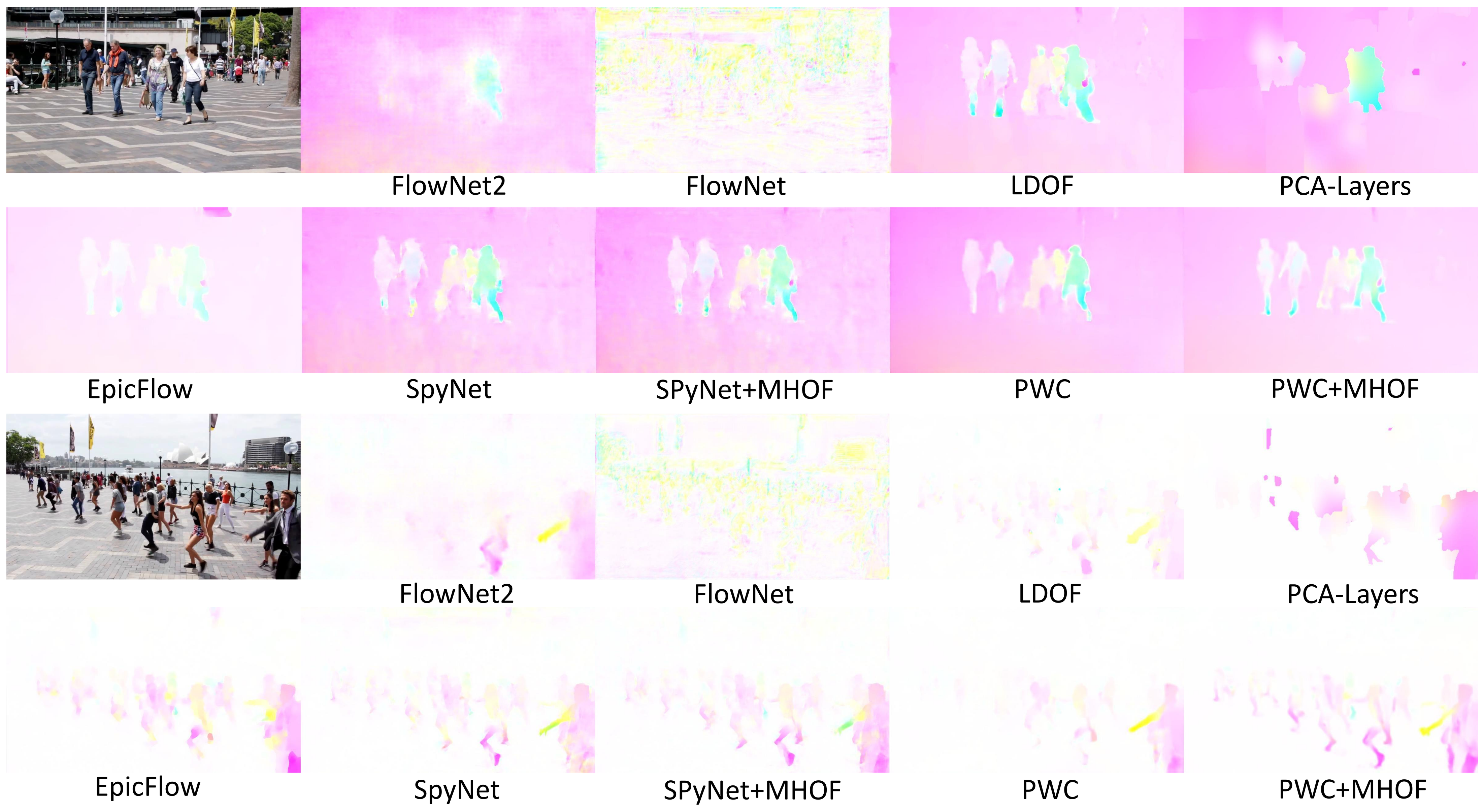
PWC-Net [3]

Qualitative Results – Single-Human



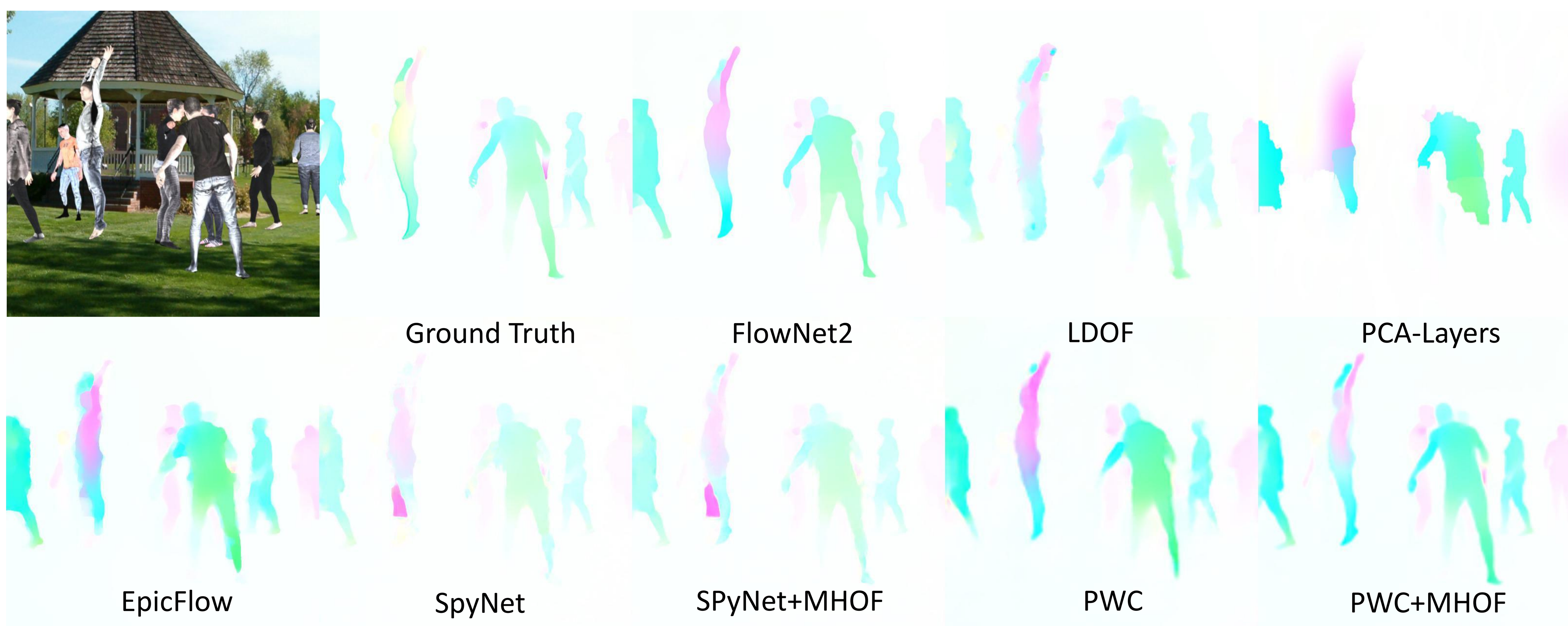
Qualitative Results – Multi-Human

Real Sequences



- Networks seem to generalize to real data
- Training on MHOF improves estimated flow maps, especially for human regions

Synthetic Sequences



- Methods trained on MHOF predict more detailed optical flow for humans with sharper edges

Quantitative Results

Single-Human

Method	AEPE	Time(s)	Learned	Fine-tuned on SHOF
Zero	0.6611	-	-	
FlowNet	0.5846	0.080	✓	✗
PCA Layers	0.3652	10.357	✗	✗
PWC-Net	0.2158	0.024	✓	✗
PWC+SHOF	0.2158	0.024	✓	✓
SPyNet	0.2066	**0.022**	✓	✗
Epic Flow	0.1940	1.863	✗	✗
LDOF	0.1881	8.620	✗	✗
FlowNet2	0.1895	0.127	✓	✗
Flow Fields	0.1709	4.204	✗	✗
SPyNet+SHOF	**0.1164**	**0.022**	✓	✓

Multi-Human

Method	Average EPE	Average EPE on *body pixels*	Fine-tuned on MHOF
FlowNet	0.808	2.574	✗
PCA Layers	0.556	2.691	✗
Epic Flow	0.488	1.982	✗
SPyNet	0.429	1.977	✗
SPyNet+MHOF	0.391	1.803	✓
PWC-Net	0.369	2.056	✗
LDOF	0.360	1.719	✗
FlowNet2	0.310	1.863	✗
PWC+MHOF	**0.306**	**1.620**	✓

- SPyNet trained on SHOF outperforms all generic methods
- PWC trained on MHOF outperforms all generic methods
- Training on MHOF improves results on body pixels

Conclusion

- Training on the human flow datasets improves optical flow estimation
- Our models improve over generic state-of-the-art flow prediction methods while being substantially smaller and faster
- Improvements over state-of-the-art methods are strongest for human regions of the image
- Qualitative results suggest generalization from synthetic to real data

References

1. Loper, Matthew et al. SMPL: A skinned multi-person linear model. ACM TOG 2015.
2. Anurag Ranjan and Michael J. Black, Optical flow estimation using a spatial pyramid network. CVPR 2017.
3. Deqing Sun, et al. PWCNet: CNNs for optical flow using pyramid, warping, and cost volume. In *CVPR*, 2018