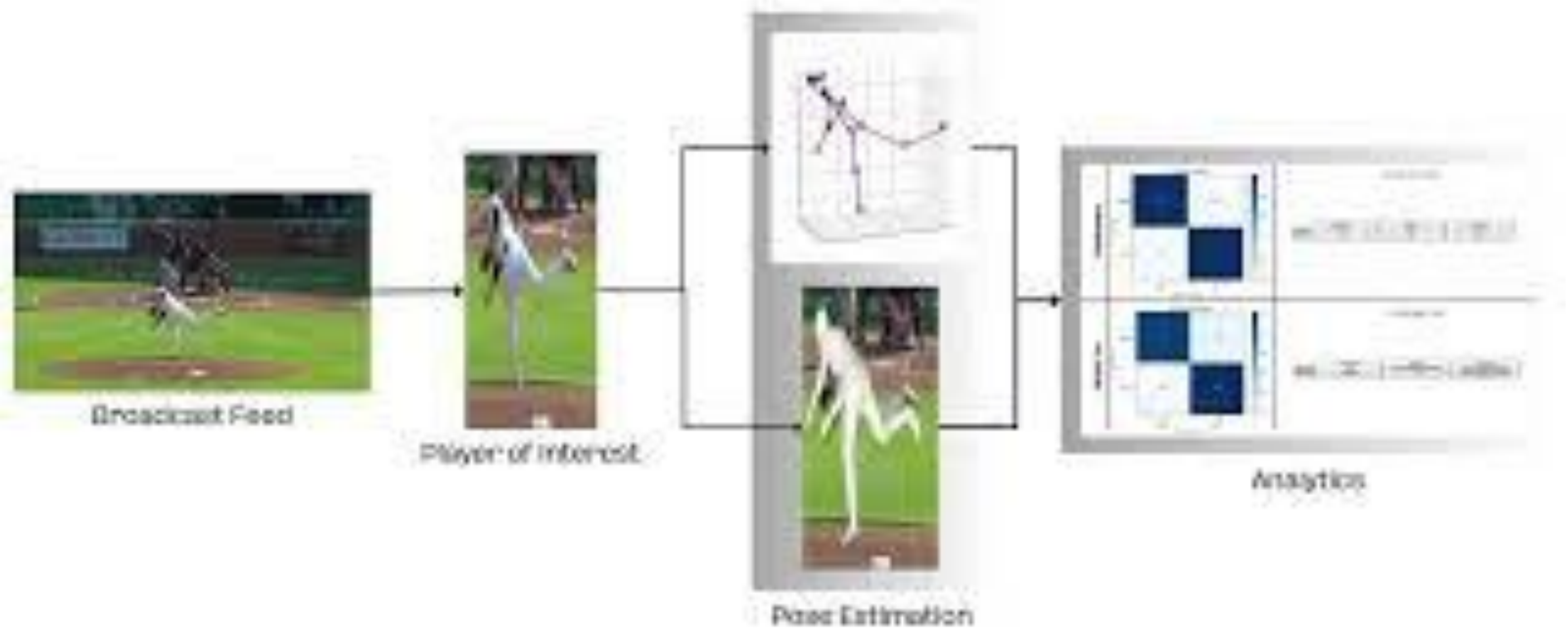
Mitigating Motion Blur for Robust 3D Baseball Player Pose Modeling for Pitch Analysis

Jerrin Bright, Yuhao Chen, John Zelek
Vision and Image Processing Lab (VIP)
Systems Design Engineering Department
University of Waterloo
Waterloo, Ontario, Canada

Motivation

- Quantitative performance indication of baseball pitchers.
- Early identification of deceptive patterns in pitching.

Analytics Framework

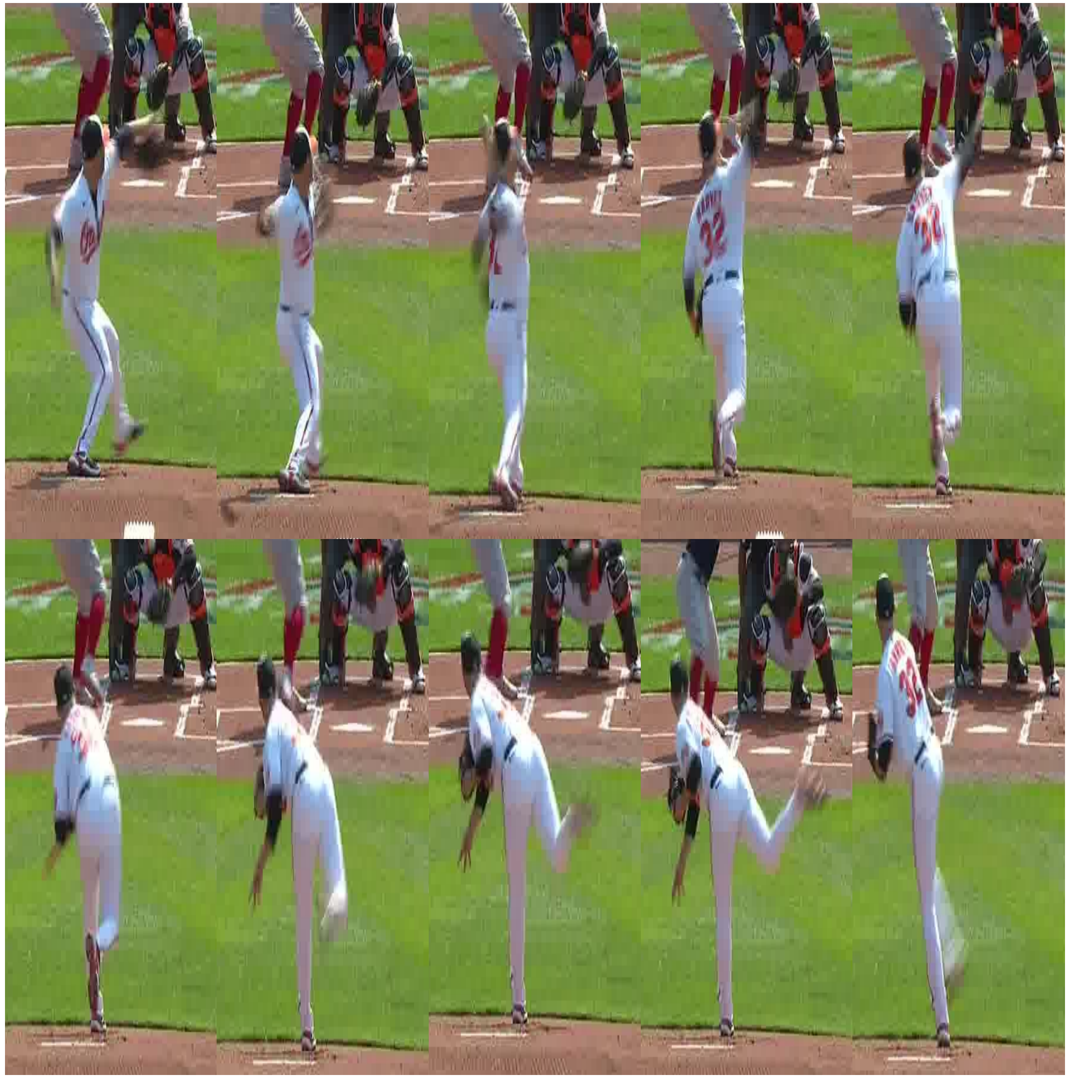


https://www.youtube.com/watch?v=ciWA41xPG4k

Challenges

1. Motion Blur
2. Self-Occlusion
3. Out-of-distribution pose

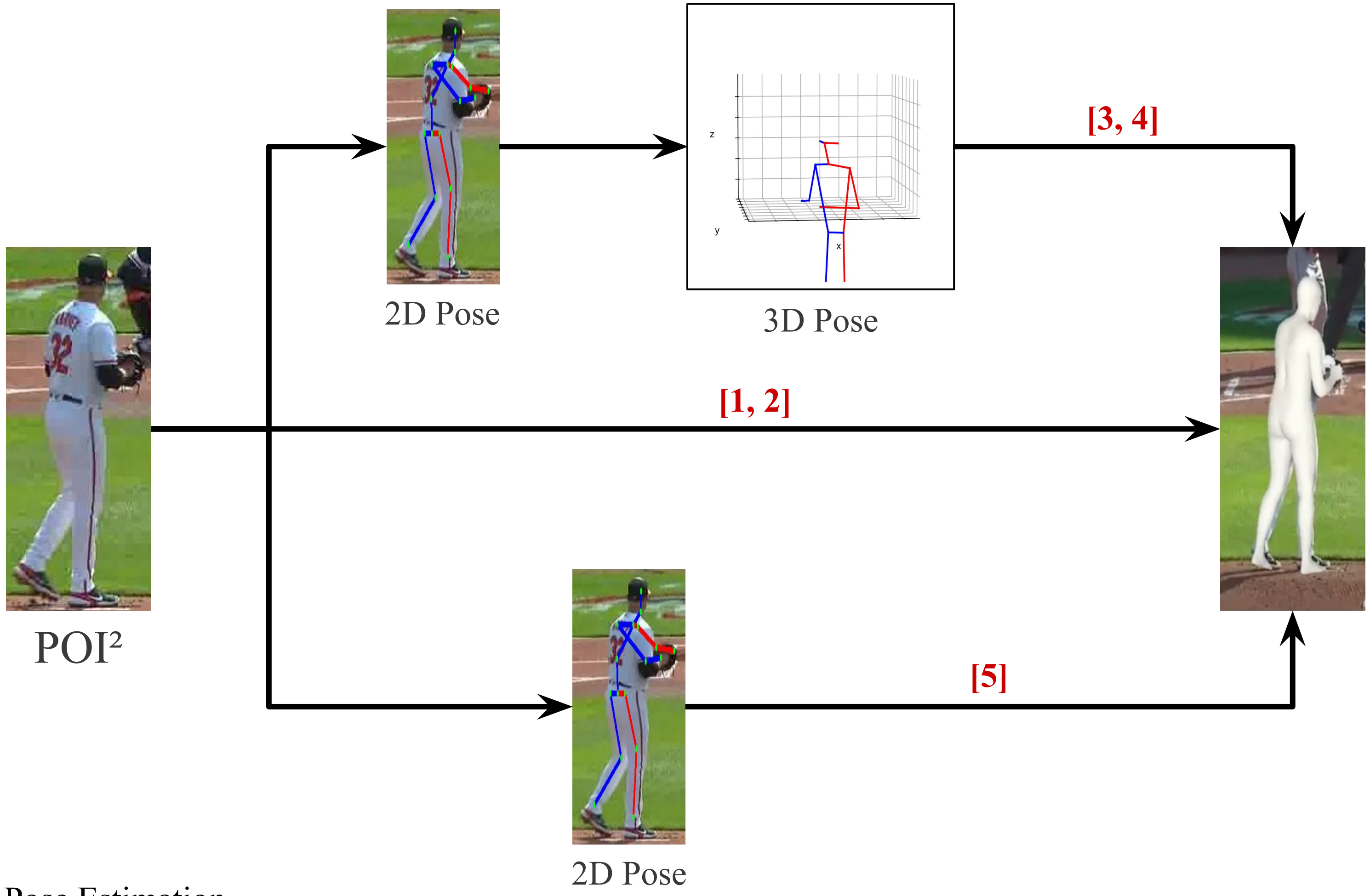
1



3

2

Existing HPE[1] Approaches



[1] Human Pose Estimation
[2] Player of Interest

[1] Lin, Kevin, Lijuan Wang, and Zicheng Liu. "End-to-end human pose and mesh reconstruction with transformers." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2021.

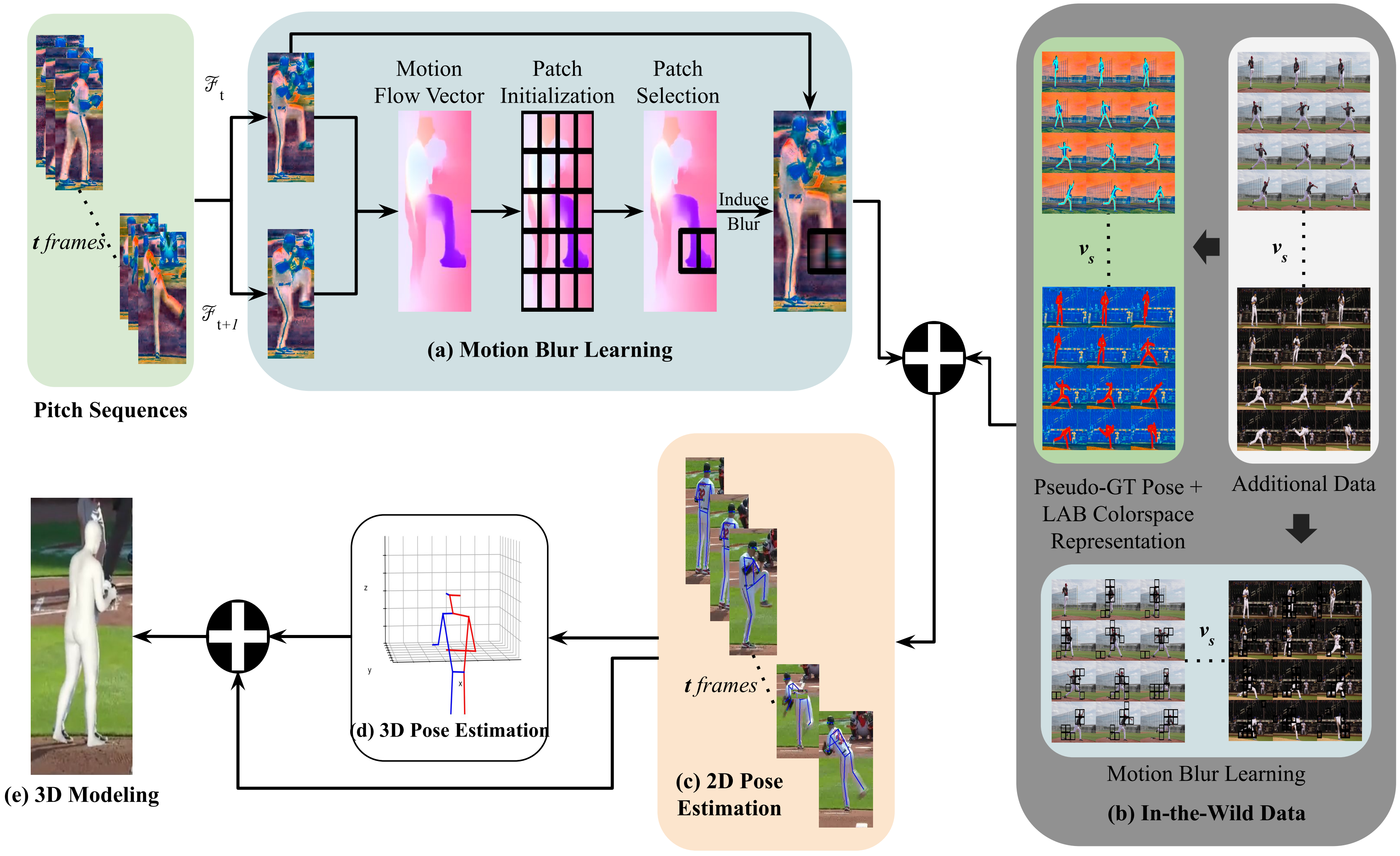
[2] Lin, Jing, et al. "One-Stage 3D Whole-Body Mesh Recovery with Component Aware Transformer." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2023.

[3] Choi, Hongsuk, Gyeongsik Moon, and Kyoung Mu Lee. "Pose2mesh: Graph convolutional network for 3d human pose and mesh recovery from a 2d human pose." *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VII 16*. Springer International Publishing, 2020.

[4] Moon, Gyeongsik, and Kyoung Mu Lee. "I2l-meshnet: Image-to-lixel prediction network for accurate 3d human pose and mesh estimation from a single rgb image." *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VII 16*. Springer International Publishing, 2020.

[5] Bogo, Federica, et al. "Keep it SMPL: Automatic estimation of 3D human pose and shape from a single image." *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part V 14*. Springer International Publishing, 2016.

Our Work



The proposed approach comprises several key steps:

1. Data Representation: Each pitch sequence is represented as $\hat{\mathcal{P}} = \{\mathcal{F}_t : \mathcal{F}_t \in \mathbb{R}^{H\times W\times 3}\}_{t=1}^{t_n}$

2. Motion Blur Augmentation: Motion flow (MF) between consecutive frames is analyzed by dividing each frame into k patches. The top N patches with the highest MF are selected as target regions for inducing blur

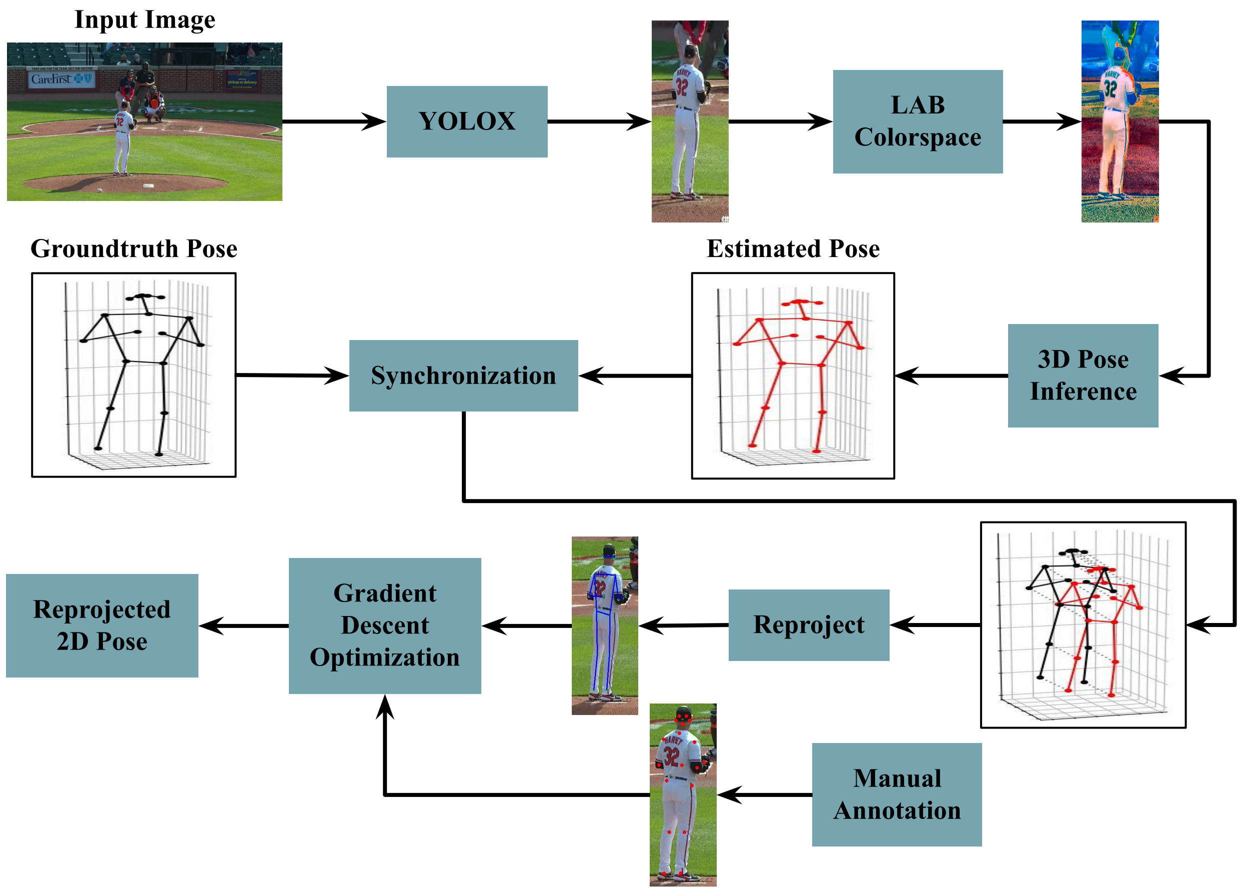
3. 2D Pose Estimation: In each frame F_t, the 2D pose of the pitcher is estimated, resulting in $\mathcal{P}_{2D}^{(t)} \in \mathbb{R}^{\mathcal{J}\times 2}$

4. 3D Pose Estimation: Utilizing a receptive field of s consecutive 2D pose, $\mathcal{P}_{2D} \in \mathbb{R}^{s\times\mathcal{J}\times 2}$ the 3D pose of the pitcher is estimated, producing $\mathcal{P}_{\text{concat}}^{(t)} \in \mathbb{R}^{1\times\mathcal{J}\times 5}$

5. Concatenation: The 2D and 3D poses are concatenated represented by $\mathcal{P}_{\text{concat}}^{(t)} \in \mathbb{R}^{1\times\mathcal{J}\times 5}$

6. Human Mesh Recovery: The 3D body mesh represented by $\mathcal{H}_{3D} \in \mathbb{R}^{\mathcal{V}\times 3}$ is then modeled using spectral convolutional networks

Dataset



1. Synchronization

$$\mathcal{G} = g_s \left(\frac{1}{\mathcal{J}} \sum_{i=1}^{\mathcal{J}} (kp_{gt}^{(i)} - kp_{pred}^{(i)})^2 \right) + g_t \left(1 - \frac{\sum_{i=1}^{\mathcal{J}} kp_{gt}^{(i)} \cdot kp_{pred}^{(i)}}{\sqrt{\sum_{i=1}^{\mathcal{J}} (kp_{gt}^{(i)})^2} \cdot \sqrt{\sum_{i=1}^{\mathcal{J}} (kp_{pred}^{(i)})^2}} \right)$$

2. Camera parameters

$$\hat{f} = f_i - \alpha \Delta L(f_i)$$

Results

Table 1. Performance of different SOTA 2D pose estimation approaches with the proposed motion blur learning module.

Method	Type	MB	Loss
Xu et al.	Heatmap		1.37
Ke et al.	Heatmap		1.46
Panteleris et al.	Regressor		1.15
Li et al.	Heatmap		1.83
Mao et al.	Regression		1.26
Xu et al.	Heatmap	✓	1.17 (+0.20)
Ke et al.	Heatmap	✓	1.21 (+0.25)
Panteleris et al.	Regressor	✓	0.55 (+0.60)
Li et al.	Heatmap	✓	1.46 (+0.37)
Mao et al.	Regressor	✓	0.61 (+0.65)

Results

Table 2. Results of the estimated pose with different modules for training.

Base Model	ItW	MB	2D Loss	3D Loss
✓			1.05	1.93
✓	✓		0.88	1.61
✓		✓	0.55	1.47
✓	✓	✓	**0.48**	**1.23**

Ablation Study

Table 3. Ablation study on varying number of filters for motion blur effect.

Filters	Loss
0	1.15
1	0.68
2	**0.55**
3	1.43
4	2.28
5	3.44

Table 5. Comparison with different patch types

Patch Type	Loss
None	1.15
Binary Mask	2.12
Inpainting	1.57
Gaussian Blur	0.99
Motion Blur	**0.55**

Table 4. Ablation study on the region size and frequency of motion blur effect

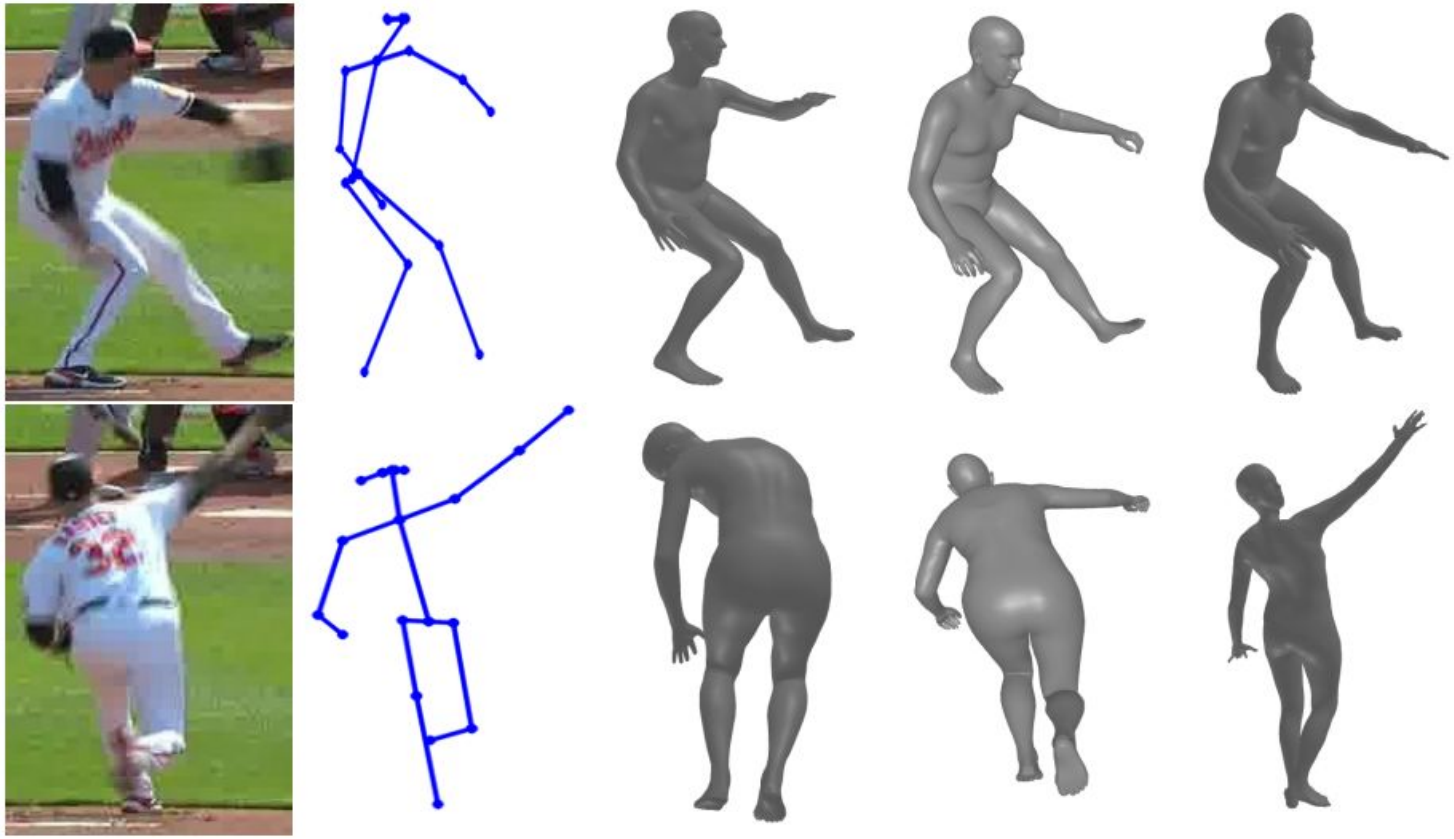
s_{patch} \ $\mathcal{N}$	1	3	5	7	9
10	0.83	0.74	0.66	0.64	0.67
20	0.71	0.57	0.62	0.60	0.62
30	0.68	**0.55**	0.61	0.639	0.59
40	0.74	0.63	0.68	0.75	0.78
50	0.77	0.75	0.71	0.83	0.97

Conclusion

1. **Innovative Augmentation for Motion Blur:** The research introduces a unique technique to strategically induce motion blur, improving the network's ability to handle this challenge during pose estimation.

2. **In-the-Wild Video Data Integration:** Incorporating in-the-wild video data, along with pseudo-groundtruth pose information, improves the network's performance under varying lighting and camera conditions.

3. **Significant Accuracy Improvement:** Substantial increase in SOTA pose estimation accuracy, particularly during pitching actions, underscores the importance of thoughtful augmentation to address motion blur.



(a) Input frame (b) GT 2D pose (c) ICON (d) OSX (e) Ours

Acknowledgement

Thank you!

Open to any questions