

PitcherNet: Powering the Moneyball Evolution in Baseball Video Analytics

Jerrin Bright, Bavesh Balaji, Yuhao Chen, David Clausi and John Zelek

Vision and Image Processing Lab

Department of Systems Design Engineering
University of Waterloo, Waterloo, ON, Canada

Date: Tuesday, June 18th, 2024

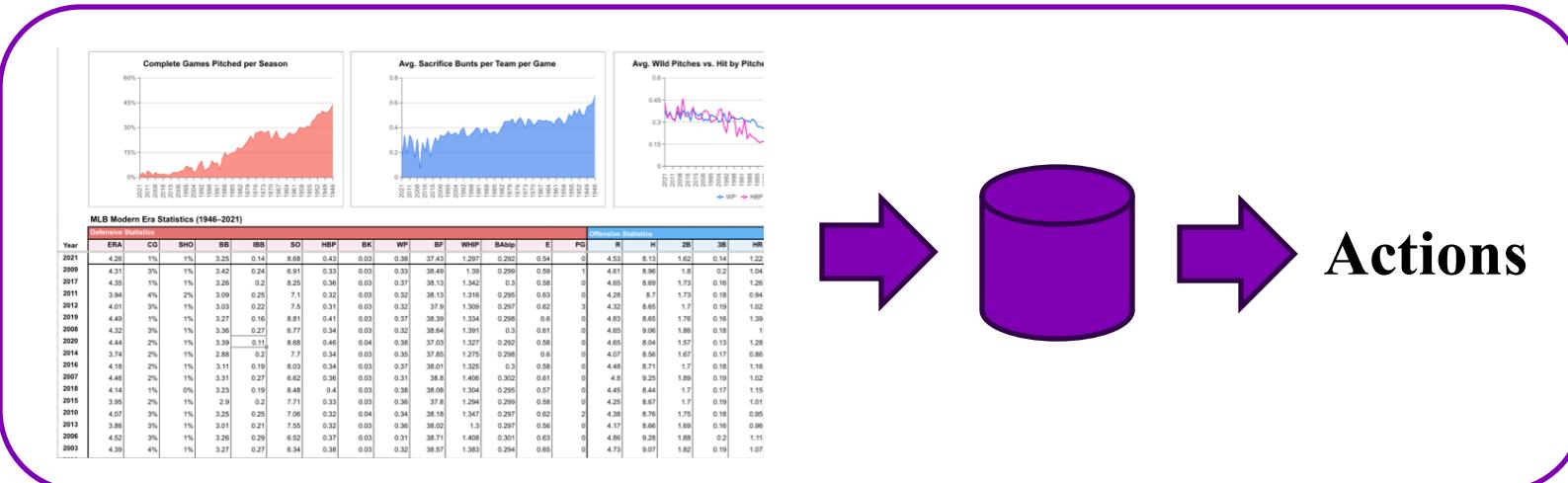


FACULTY OF
ENGINEERING



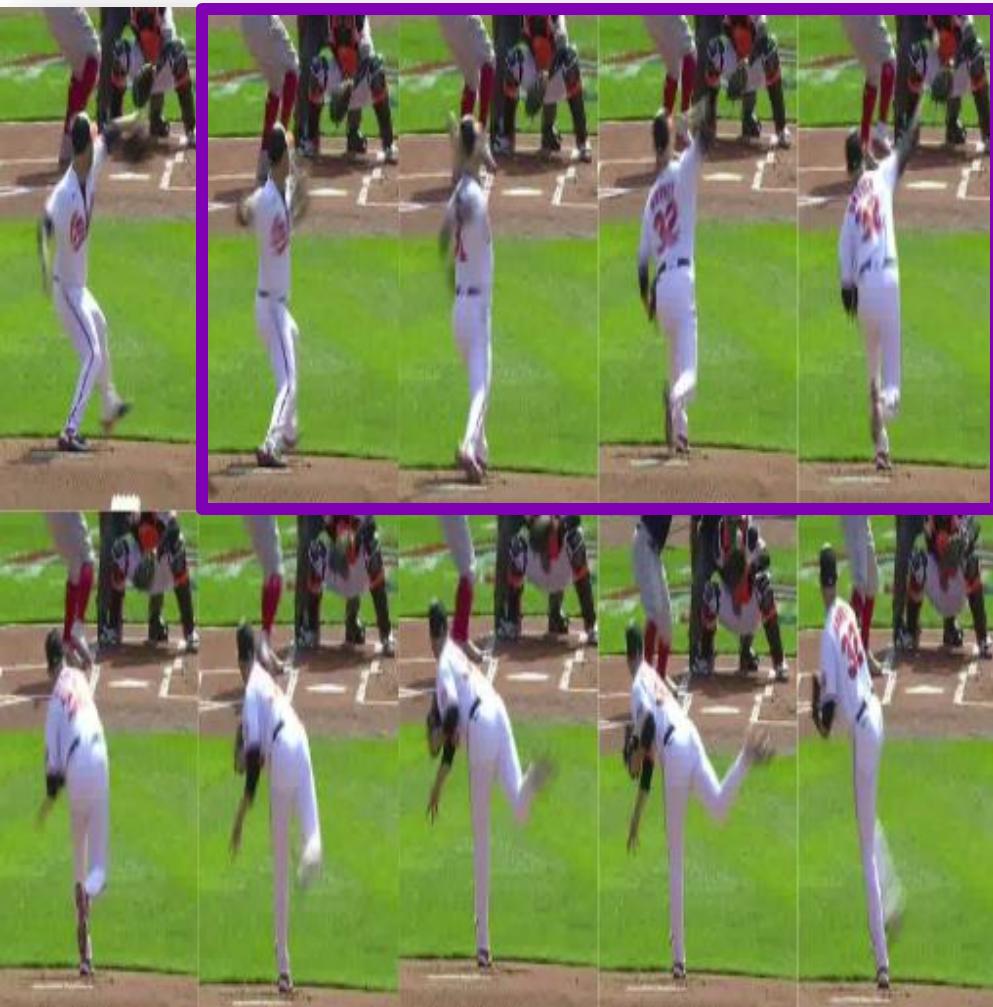
Prior Research on Baseball Analysis

- Pre-recorded baseball databases (Pitch f/x).
- Controlled environments (MoCap Systems).



Challenges with Video Inputs

Motion Blur



Self-Occlusion



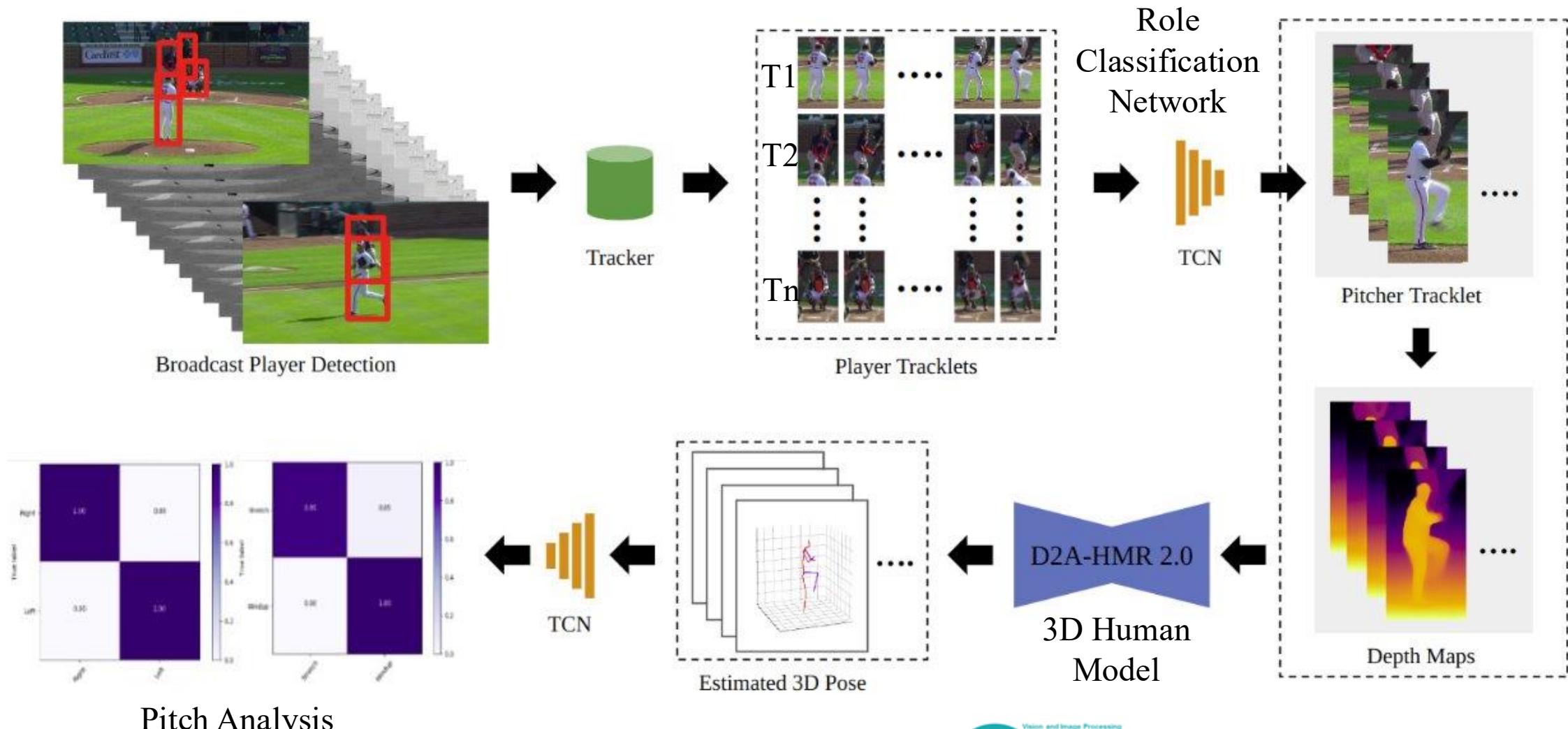
Out-of-distribution



Objective

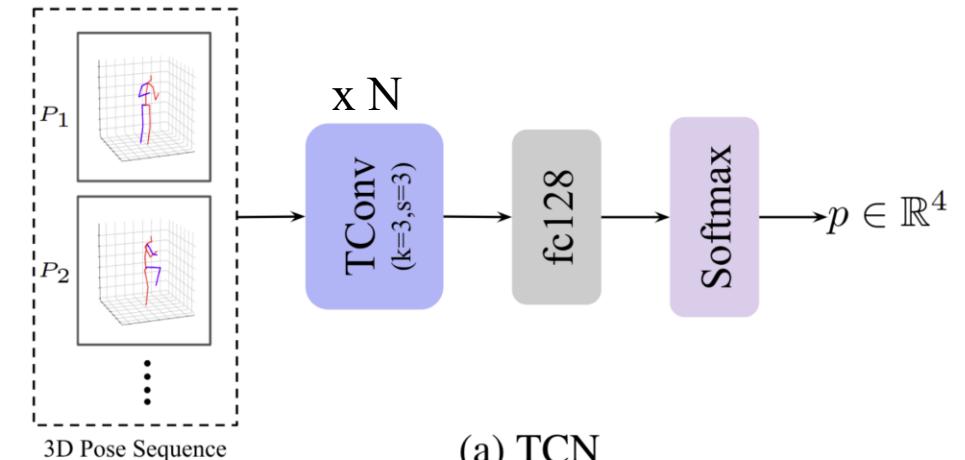
"Enable detailed analysis of pitcher dynamics from human models in 3D extracted solely from monocular broadcast feeds"

PitcherNet System

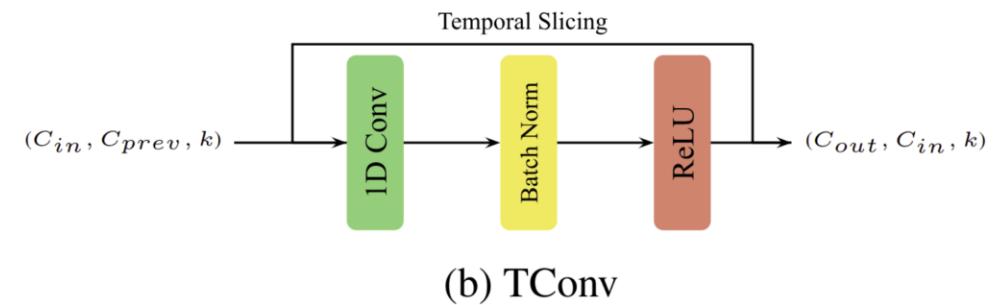


Role Classification

- Decouples action from player kinematics.
 - Input: Pseudo-pose from estimated tracklets.
 - Output: Player role.
- Invariant to viewpoint/ facial features/ player jersey numbers.



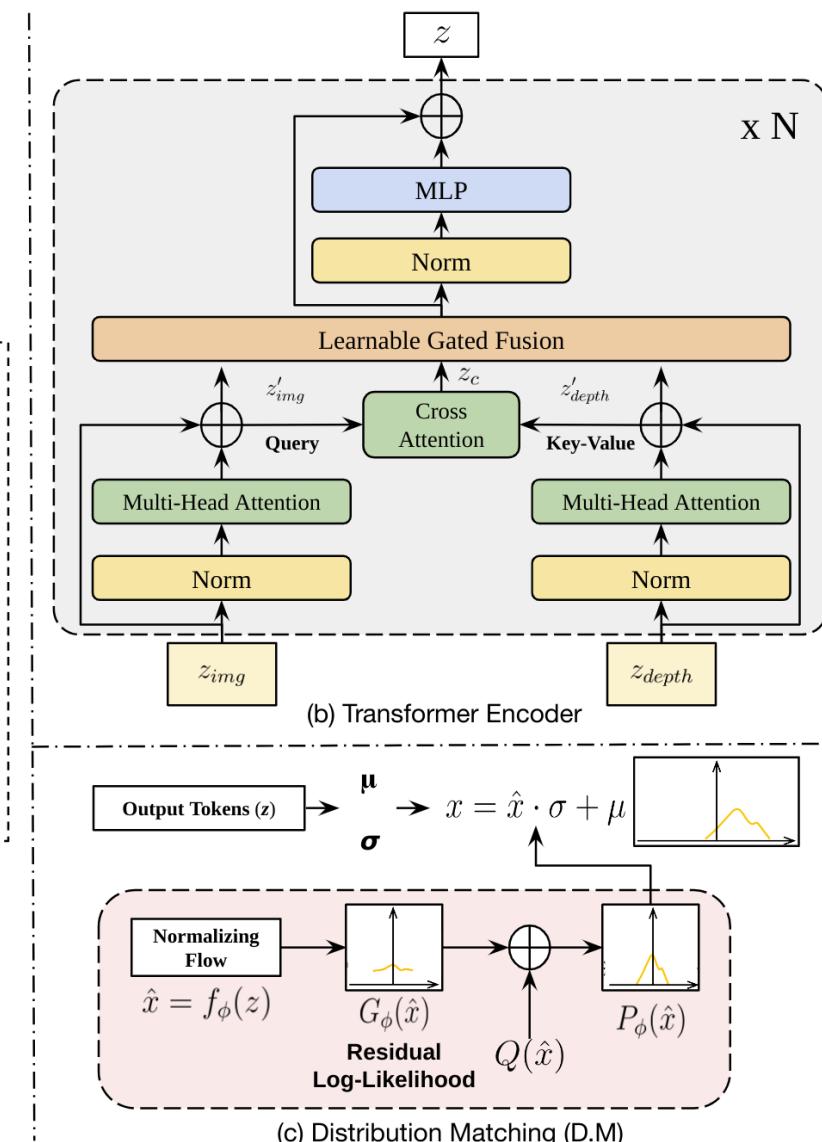
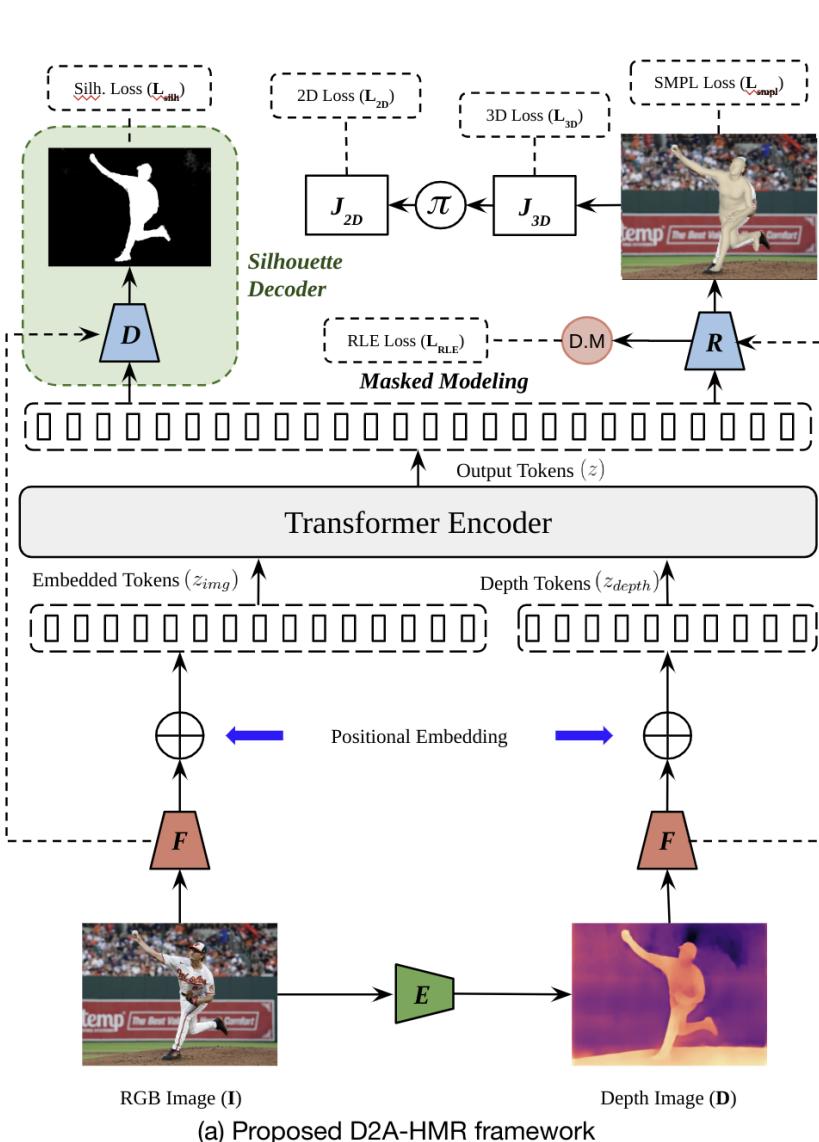
(a) TCN



(b) TConv

3D Human Model

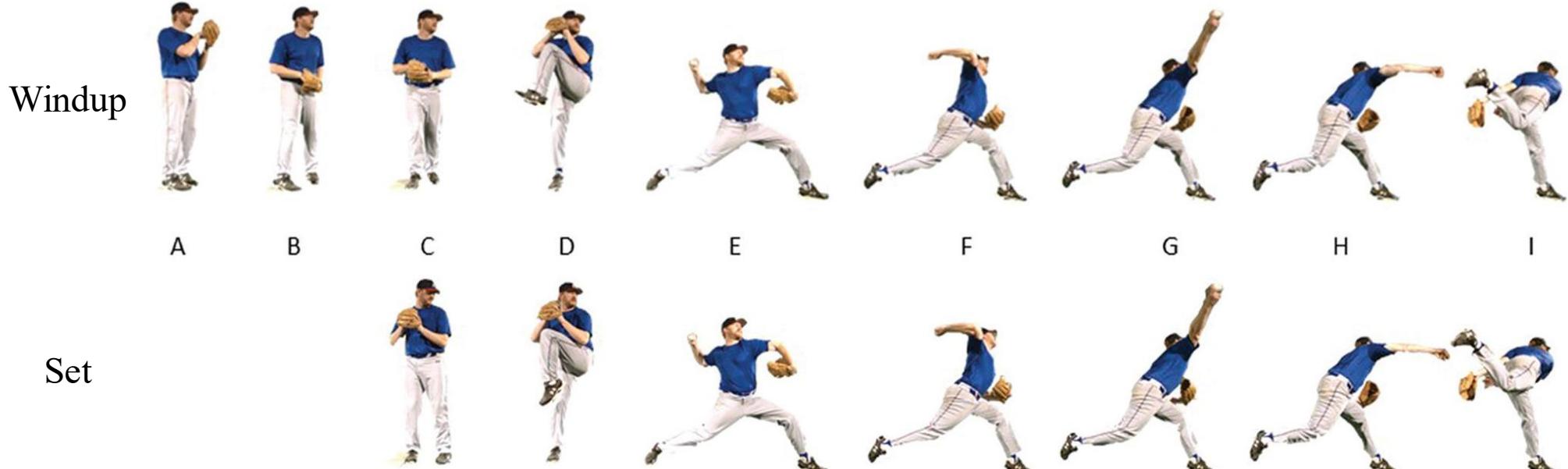
- Distribution and depth aware 3D modeling [2].
- Motion blur and in-the-wild data augmentation.
- Generalizable, reliable 3D human models.



Pitch Analysis

- Pitch Position

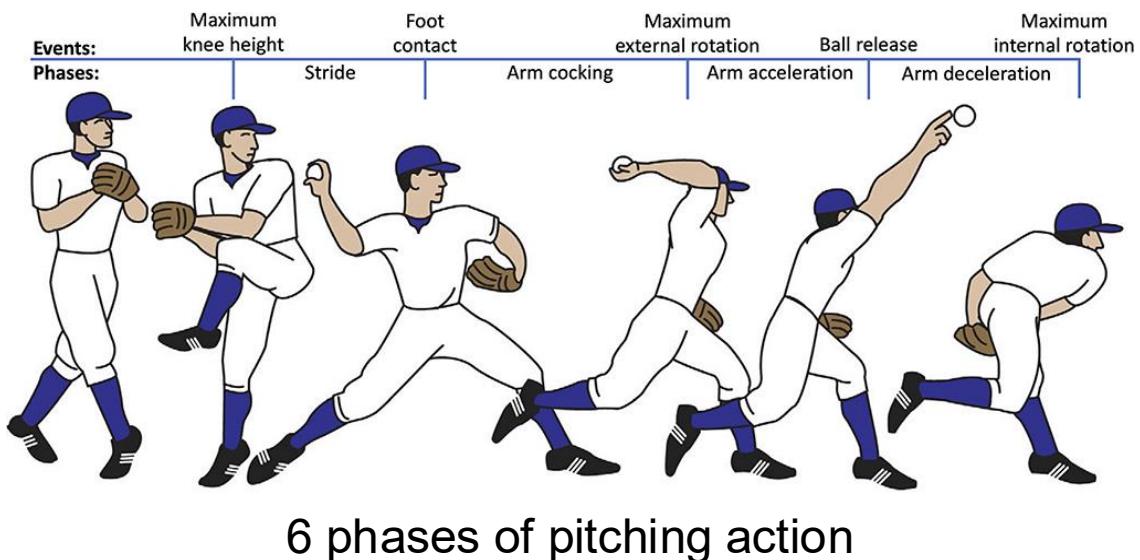
$$PP(windup, set) = \sigma(TCN(X))$$



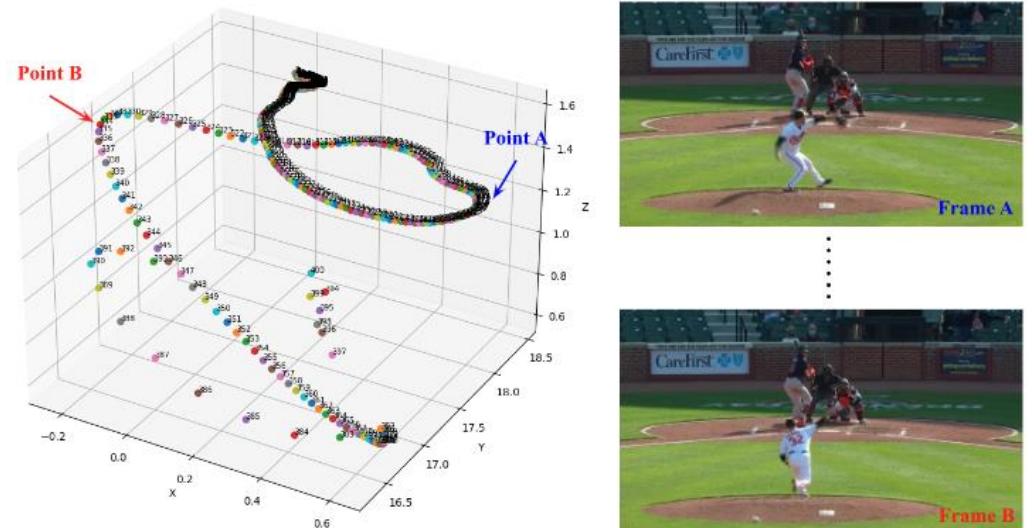
Pitch Analysis

- Release Point

$$P_{rel} = \text{argmax}(v(i) | i \in [P_b - n/2, P_b + n/2])$$



Point A- Arm Cocking
Point B- Arm Deceleration



Trajectory of the right wrist joint in 3D space

Pitch Analysis

- Pitch Velocity

$$v_p = \omega \times l = \{(\text{atan}(w_y^r, w_x^r) - \text{atan}(w_y^{r-1}, w_x^{r-1})) \times T\} \times l$$

- Release Extension

$$E_{rel} = \sqrt{(w_x - a_x)^2 + (w_y - a_y)^2 + (w_z - a_z)^2}$$

Quantitative Results of Role Classification

Table I. Comparison with baselines

Test Accuracy ↑	
LSTM	85.55
Transformer	91.11
Ours	96.66

Table II. Comparison with jersey identification techniques

	Test Accuracy ↑
Gerke <i>et al.</i> [21]	64.47
Li <i>et al.</i> [30]	88.29
Vats <i>et al.</i> [48]	89.46
Balaji <i>et al.</i> [2]	93.68
Balaji <i>et al.</i> [3]	94.70
Ours	96.82

Quantitative Results of Pitch Analysis

Table III. Performance of our pitch statistics modules

(a) Handedness			(b) Pitch Position								
	Accuracy ↑	F1 Score ↑	Precision ↑		Accuracy ↑	F1 Score ↑	Precision ↑				
LSTM	85.0	85.7	90.0	LSTM	81.3	82.5	85.0				
Ours (TCN)	100.0	100.0	100.0	Ours (TCN)	97.5	97.4	95.0				
(c) Release Point			(d) Pitch Velocity			(e) Release Extension					
	$A_1 \uparrow$	$A_2 \uparrow$	$A_5 \uparrow$		$A_{1\%} \uparrow$	$A_{2\%} \uparrow$	$A_{5\%} \uparrow$		$A_{5\%} \uparrow$	$A_{8\%} \uparrow$	$A_{10\%} \uparrow$
LSTM	31.3	46.4	63.5	LSTM	5.1	13.1	22.2	LSTM	4.0	7.1	11.1
TCN	43.4	51.5	77.6	TCN	10.1	18.1	48.4	TCN	14.1	19.1	25.2
Ours	80.8	85.8	97.9	Ours	43.4	68.6	94.9	Ours	24.2	31.3	37.3

Qualitative Results (Pitch Analysis)

	Pitch Hand	Pred: Left	GT: Left
	Pitch Position	Pred: Stretch	GT: Stretch
	Pitch Velocity	Pred: 90.48 Mph	GT: 87.58 Mph
	Release Point	Pred: 90	GT: 90
	Extension	Pred: 5.85 feet	GT: 6.13 feet
	Pitch Hand	Pred: Left	GT: Left
	Pitch Position	Pred: Windup	GT: Windup
	Pitch Velocity	Pred: 85.76 Mph	GT: 89.17 Mph
	Release Point	Pred: 88	GT: 89
	Extension	Pred: 6.01 feet	GT: 6.16 feet
	Pitch Hand	Pred: Right	GT: Right
	Pitch Position	Pred: Windup	GT: Windup
	Pitch Velocity	Pred: 85.46 Mph	GT: 85.65 Mph
	Release Point	Pred: 87	GT: 87
	Extension	Pred: 6.17 feet	GT: 6.11 feet

Summary

- Reliable pitch analysis driven by player kinematics and human model priors.
- Role classification aiming to classify players by decoupling actions.
- D2A-HMR v2 which improves 3D human modeling in degraded image quality.

PitcherNet: Powering the Moneyball Evolution in Baseball Video Analytics

Jerrin Bright Baveesh Balaji Yuhao Chen David A Clausi John S Zelek
{jerrin.bright, bbalaji, yuhao.chen, dclausi, jzelek}@uwaterloo.ca
University of Waterloo, Waterloo, ON, Canada, N2L 3G1



Figure 1. 3D player reconstruction and kinematic-driven pitch statistics from monocular video. We introduce PitcherNet, a pioneering deep learning system that tackles low-resolution video limitations through efficient 3D human modeling for robust player alignment (left) and reliable pitch statistics analysis from estimated kinematic data (right).

1. Introduction

In the high-stakes world of baseball, every moment of play matters. This holds the key to maximizing performance and minimizing runs. Traditional analysis methods often rely on pre-recorded offline numerical data, hindering their application in the dynamic environment of live games. Broadcast video analysis, while seemingly ideal, faces significant challenges due to motion blur and low resolution. To address these challenges, we introduce PitcherNet, an end-to-end automated system that analyzes pitcher kinematics directly from live broadcast video, thereby extracting valuable pitch statistics including velocity, release point, pitch position, and release extension. The system achieves this by (1) 3D Player Alignment and identification by decoupling actions from player kinematics; (2) Distribution and depth-aware 3D human modeling; and (3) Kinematic-driven pitch statistics. Experimental validation demonstrates that PitcherNet achieves robust analysis results with 96.82% accuracy in pitcher tracklet identification, reduced joint position error by 1.8mm and improved joint angle error by 1.1°, while enabling performance-critical kinematic analysis from broadcast video. PitcherNet paves the way for the future of baseball analytics by optimizing pitching strategies, preventing injuries, and unlocking a deeper understanding of pitcher mechanics, forever transforming the game.

3420



PITCHERNET: POWERING THE MONEYBALL EVOLUTION IN BASEBALL VIDEO ANALYTICS

JERRIN BRIGHT, BAVEESH BALAJI, YUHAO CHEN, DAVID CLAUSI, JOHN ZELEK
UNIVERSITY OF WATERLOO, WATERLOO, ONTARIO, CANADA

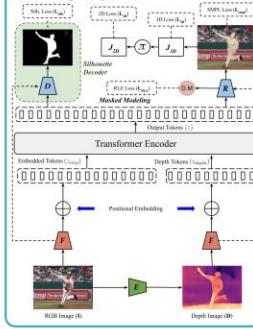


KEY CONTRIBUTIONS

- ❖ Accurate baseball pitch statistics prediction from low-quality videos (PitcherNet)
- ❖ Classifying players based on their role by decoupling actions from player kinematics;
- ❖ 3D human modeling that addresses motion blur.

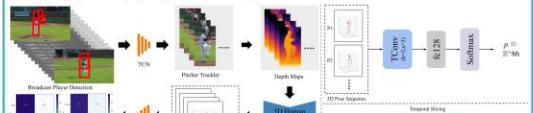
3D HUMAN MODELING

Designed to extract kinematic information about the pitcher from its corresponding tracked 3D skeleton.



METHODOLOGY

We introduce PitcherNet, a pioneering deep learning system that tackles low-resolution video limitations through efficient 3D human modeling for robust player alignment and reliable pitch statistics analysis from estimated kinematic data.



PITCH STATISTICS

Pitch Position (PP): $PP(wrist, set) = \sigma(TCN(X))$, where $X \in R^{100 \times 18 \times 3}$ (1)

$$\text{Release Point } (P_{rel}) = argmax(v(t)) | t \in [P_1 - n/2, P_1 + n/2] \quad (2)$$

$$\text{Pitch Velocity } (v_p): v_p = \omega \times l = \{(atan(w_y^{\prime}, w_x^{\prime}) - atan(w_y^{t-1}, w_x^{t-1})) \times T\} \times l \quad (3)$$

$$\text{Release Extension } (E_{rel}): E_{rel} = \sqrt{(w_x - a_x)^2 + (w_y - a_y)^2 + (w_z - a_z)^2} \quad (4)$$

LOSS FUNCTIONS

Overall objective function for the 3D model:

$$L_{model} = \lambda_{3D}\mathcal{L}_{3D} + \lambda_{NMP}\mathcal{L}_{NMP} + \lambda_{ID}\mathcal{L}_{ID} + \lambda_{A_{rel}}\mathcal{L}_{A_{rel}} + \lambda_{S_{pitch}}\mathcal{L}_{S_{pitch}} \quad (5)$$

Objective function for the TCN model:

$$L(p_i) = \sum_{t=1}^{T-1} a_t * (1 - p_t)^n + \log_2(p_t) \quad (6)$$

QUALITATIVE RESULTS



QUANTITATIVE RESULTS

- ❖ 3D Human Modeling

Method	Human3.6M		3DPW	
	mPIPE	Pa-mPIPE	mPIPE	Pa-mPIPE
HMMR'19	62.3	41.1	95.0	55.8
TCMR'21	65.6	41.4	92.0	56.5
VIBE'20	-	-	105.6	53.5
TCM'20	63.2	43.8	93.1	52.2
PARE'21	76.8	50.6	82.0	50.9
PoIMR'21	-	41.2	95.1	59.7
P2M'20	64.9	47.0	89.2	58.9
METRO'21	54.0	36.7	77.1	47.9
Ours	53.2	35.9	78.7	46.9

- ❖ Handedness and Pitch Position Prediction

Method	Handedness		Pitch Position			
	A ↑	F1 ↑	Precc. ↑	A ↑	F1 ↑	Precc. ↑
LSTM	85.0	85.7	90.0	81.3	82.5	85.0
Ours	100.0	100.0	100.0	97.5	97.4	95.0

- ❖ Release Point and Pitch Velocity Prediction

Method	Release Point		Pitch Velocity			
	A ₁ ↑	A ₂ ↑	A ₃ ↑	A ₄ ↑	A ₅ ↑	
LSTM	31.3	46.4	63.5	5.1	13.1	22.2
TCN	43.3	51.5	77.6	10.1	18.1	48.4
Ours	80.8	85.8	97.9	43.4	68.6	94.9

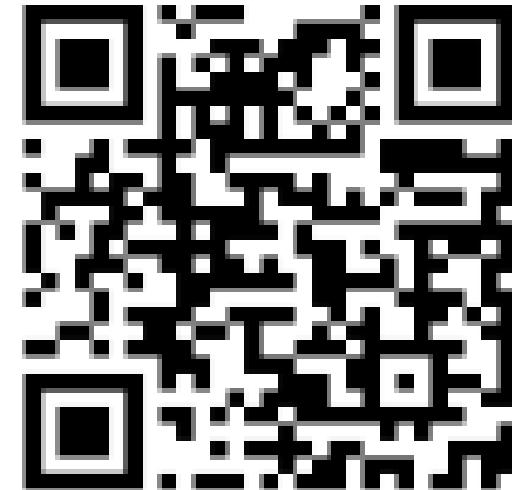
- ❖ Release Extension Prediction

Method	Release Extension		Ages ↑	
	A ₁₀ ↑	A ₁₅ ↑	A ₂₀ ↑	A ₂₅ ↑
LSTM	4.0	7.1	11.1	-
TCN	14.1	19.1	25.2	-
Ours	24.2	31.3	37.3	-

ACKNOWLEDGEMENT



Acknowledgment: This project is funded by the Baltimore Orioles of the Major League Baseball, whose generous support through the Mitacs Accelerate Program played a pivotal role in advancing this research. We also acknowledge the Digital Research Alliance of Canada for their invaluable hardware support.



FACULTY OF
ENGINEERING

Thank you!

Supported by:

