

TL;DR

We introduce Gen4D, an automated pipeline for synthesizing diverse, realistic 4D human animations, and use it to build SportPAL, a large-scale synthetic sports dataset for human-centric vision tasks.

MOTIVATION

Collecting diverse, high-quality human motion data in real-world sports is costly and logically difficult. While synthetic datasets offer a promising alternative, most rely on **fixed 3D assets, and repetitive animations**, resulting in limited diversity across appearance, motion, and viewpoint; ultimately restricting generalization to in-the-wild settings.

PROBLEM STATEMENT

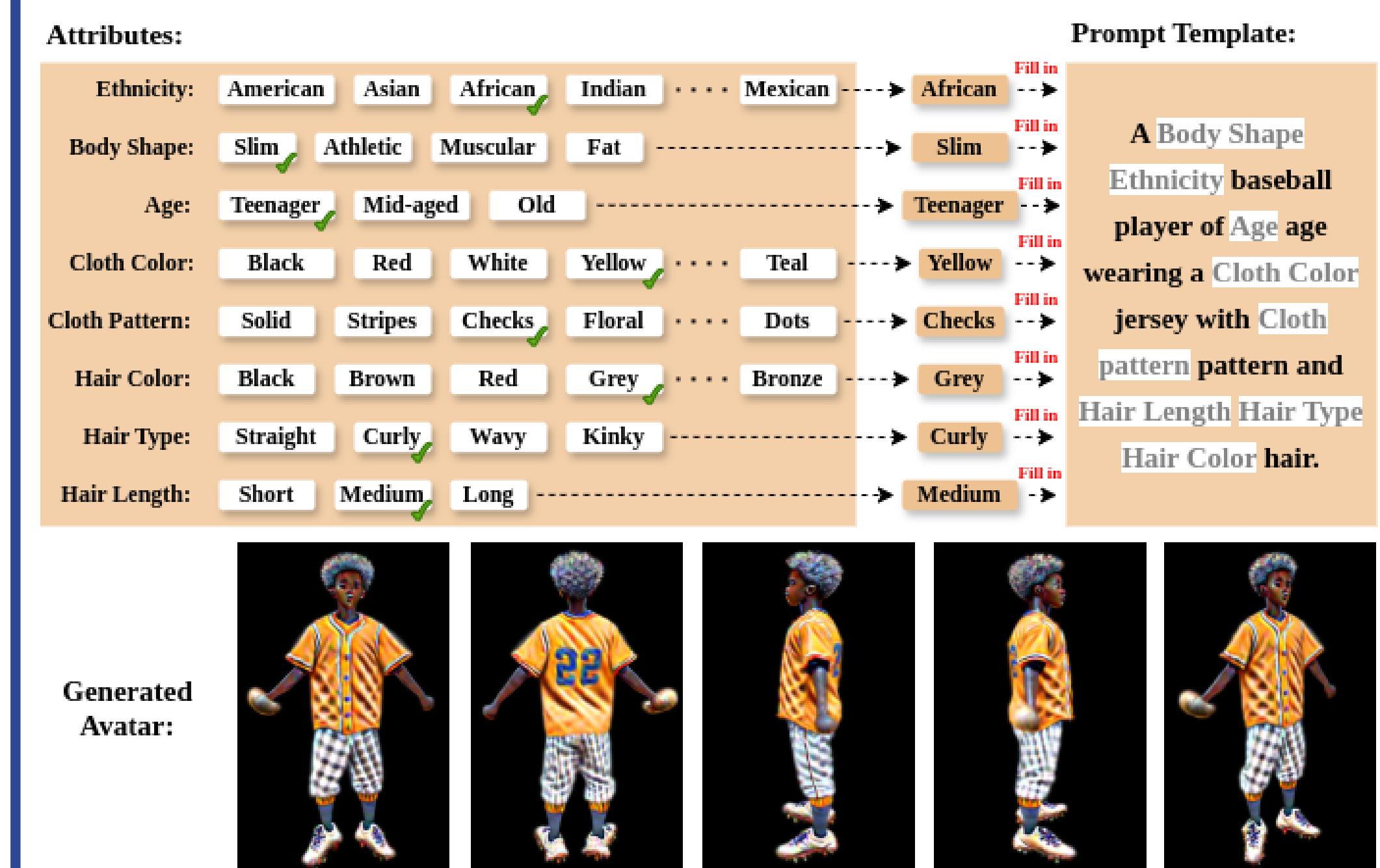
Can we fully automate the synthesis of lifelike human animation from raw motion to scene synthesis without any manual 3D modeling?

KEY CONTRIBUTIONS

- ❖ Gen4D: A fully automated pipeline for synthesizing lifelike human avatars with realistic animations.
- ❖ SportPAL: A large-scale, richly annotated synthetic dataset spanning baseball, ice hockey, and soccer, designed for human-centric vision tasks.

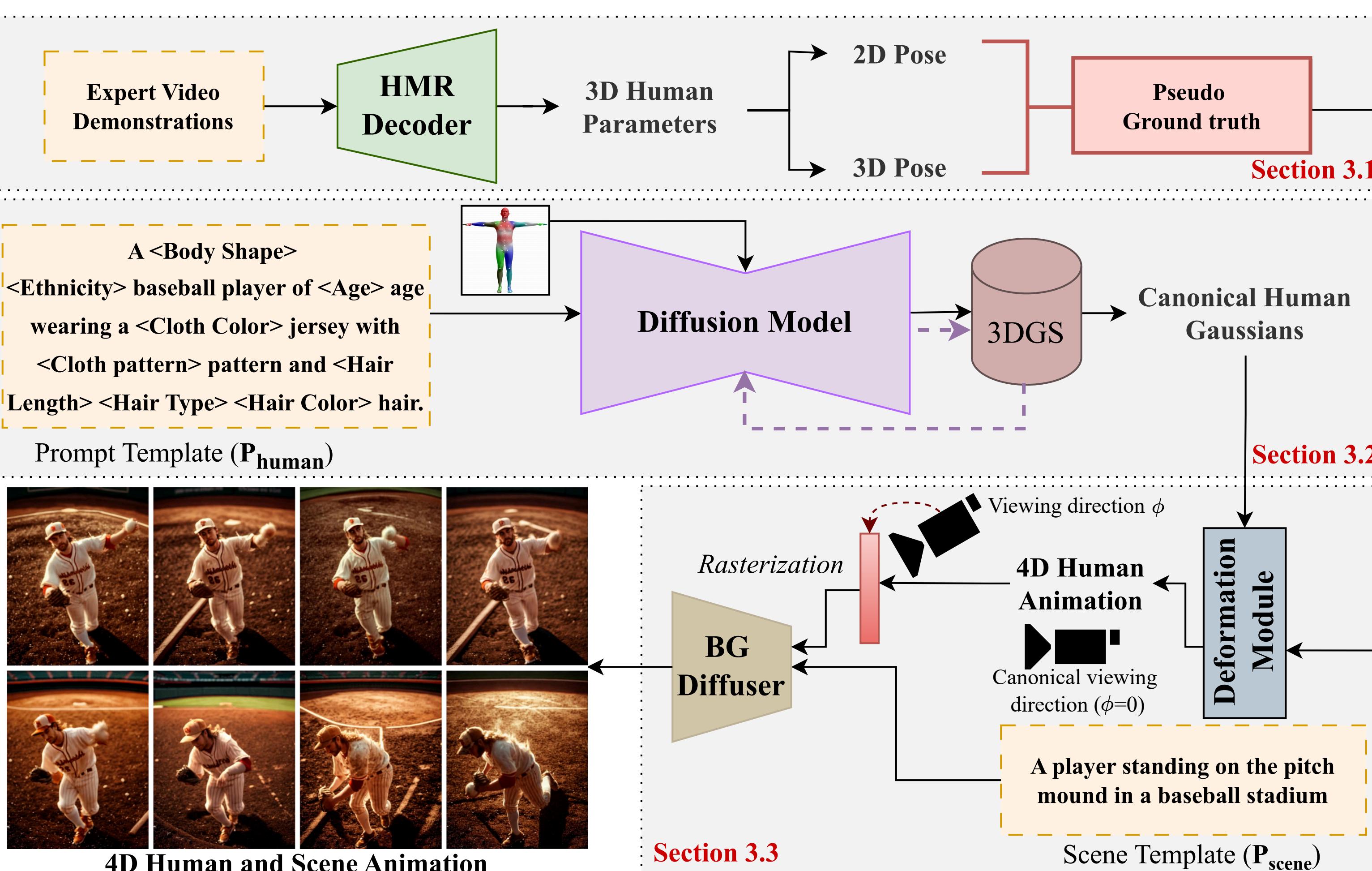
PROMPT MODELING

- ❖ Uses text templates to guide avatar generation via diffusion models.
- ❖ Captures diversity in appearance attributes like ethnicity, body type, age, hair, and clothing.
- ❖ Attributes are combined programmatically to avoid repetition and ensure balanced sampling.
- ❖ Enables generation of highly varied and realistic human avatars without manual asset design.



THE PROPOSED METHOD: GEN4D

1. **Motion Extraction:** obtain motion representation from internet videos.
2. **Canonical Human Gaussians:** generates diverse human avatars via text-guided diffusion and Gaussian splatting in canonical space.
3. **Scene Composition:** Animate avatars with motion, render from multiple viewpoints, and synthesize human-aware backgrounds using diffusion-based scene generation.



QUALITATIVE RESULTS



SPORTPAL

- ❖ Includes 580K+ synthetic frames across baseball, ice hockey, and soccer.
- ❖ Rich annotations: 2D/3D poses, SMPLX parameters, bounding boxes, and action labels.
- ❖ Built from 50 unique subjects with varied ethnicity, body types, clothing, and viewpoints.

SportPAL dataset split

Sport	Split	#Subjects	#Clips	#Frames
Baseball	Train	15	1,000	253,869
	Valid	15	304	80,810
	Test	5	300	71,875
Icehockey	Train	10	195	75,468
	Valid	10	50	18,867
	Test	5	12	7,487
Soccer	Train	10	116	57,110
	Valid	10	30	14,277
	Test	5	5	3639
Total	-	50	2,012	583,403

QUANTITATIVE RESULTS

Impact of fine-tuning with cross-domain sports

Sport	Method	AP ⁵ ↑	AP ¹⁰ ↑	AP ¹⁵ ↑
Icehockey	w/o finetuning	62.75	96.92	99.91
	w/ finetuning	63.47	98.10	99.98
Soccer	w/o finetuning	67.28	92.56	98.51
	w/ finetuning	71.46	94.68	99.13

ACKNOWLEDGEMENT



REFERENCES

- [1] Yanjie Li, Shoukui Zhang, Zhicheng Wang, and et al. Tokenpose: Learning keypoint tokens for human pose estimation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 11313–11322, 2021.