# Human Pose Estimation in Trampoline Gymnastics: How to Improve Performance on Extreme Poses

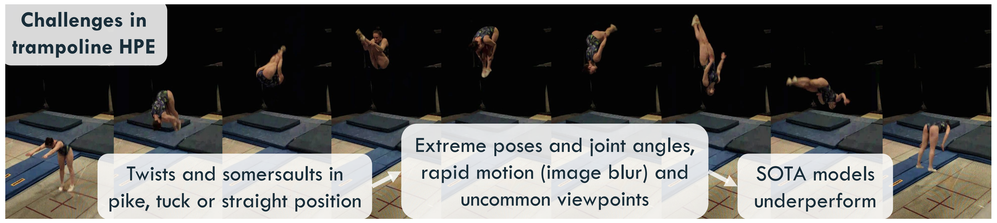
Léa Drolet-Roy[1,2], Victor Nogues[1], Bérenger Chedal-Anglay[1], Sylvain Gaudet[2], Eve Charbonneau[3], Mickaël Begon[4] & Lama Séoud[1]

[1] Polytechnique Montréal, [2] Institut national du sport du Québec, [3] Université de Sherbrooke, [4] Université de Montréal

ECCV 2026 | MALMÖ | SEPT 8–12

## Problem & Objective

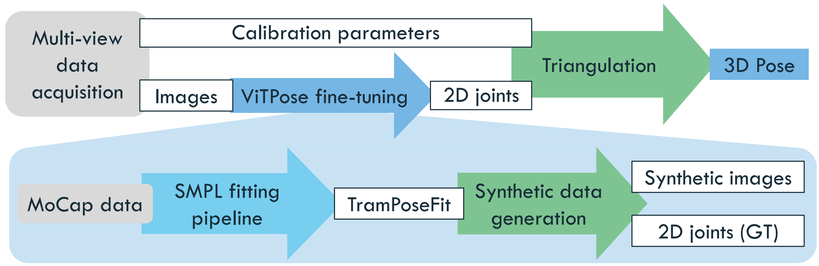
In competitive trampoline gymnastics and other high-level sports, 3D human pose estimation (HPE) could enable quantitative performance analysis for training-load monitoring and objective scoring.




**Objective: Improve HPE on extreme acrobatic poses**

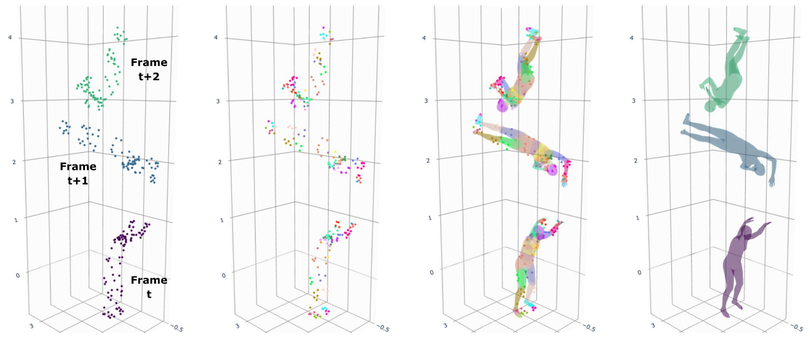
1. Fit SMPL[1] models to trampoline MoCap data and generate a dataset of synthetic images with 2D pose annotations
2. Fine-tune a HPE model, ViTPose[2], on a combination of real extreme poses (LSPe) and domain-specific synthetic poses (STP)
3. Evaluate the 3D pose triangulation accuracy on multi-view real trampoline images (MRT)

## Method Overview




## SMPL fitting pipeline

We have MoCap acquisitions of 15 jumps executed by 3 athletes. Up to 95 marker positions are captured using 18 Vicon optoelectronic cameras. Due to large accelerations, markers detach during MoCap, leading to incomplete or diverging marker positions. Our goal is to represent smooth, plausible and complex motion, creating the dataset TramPoseFit.




**Filtering**

- Markers with invalid labels or high errors
- Non-moving markers
- Markers deviating from the dominant jump pattern
- 59 ± 19 markers remain

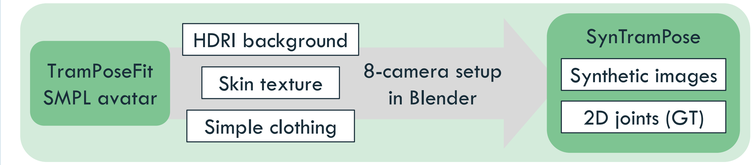
**Fitting in 2 steps**

1. Synthetic marker to vertex associations on a few frames
2. Synthetic to real marker associations over all frames with a distance loss, temporal smoothing and GMM prior

**Results**

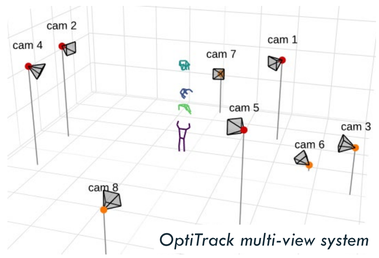
- Smooth and realistic SMPL sequence
- 44% closer to the markers than SOMA[3] and running 2.6x faster

## Synthetic Data Generation




## Triangulation

- 8 OptiTrack cameras synchronized and calibrated
- 3D poses triangulated with Pose2Sim[4] iterative optimization, trying all possible combinations of $N \geq 3$ cameras until $RMSE < 15$ px




*OptiTrack multi-view system*

## ViTPose fine-tuning : Datasets


| Dataset | N | Split | Description |
|---|---|---|---|
| **SynTramPose (STP)** | 3,624 | Train | Synthetic images generated from TramPoseFit (9 jumps, 2 athletes) |
| **Leeds Sports Pose[5] (LSPe)** | 10,000 | Train | Real images of extreme sports poses |
| **Multiview Real TramPose (MRT)** | 849 | Val | Video frames of trampoline training sessions (13 jumps, 3 athletes) |
| | 556 | Test | |

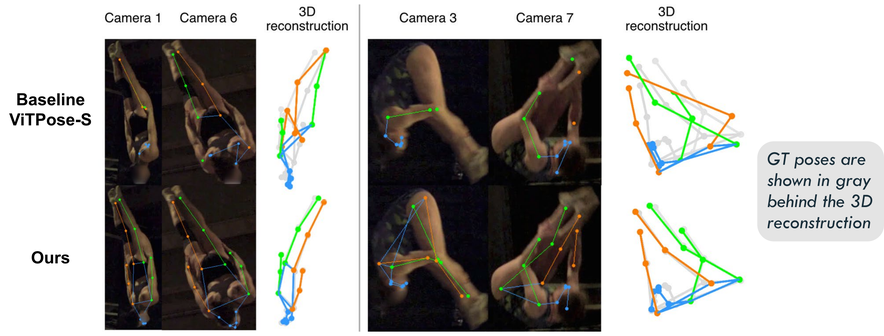
## ViTPose fine-tuning : Results

*Performance on MRT – fine-tuning data includes 5,000 COCO images and * refers to the provided weights*

| Model | Fine-tuning data | N images | AP | AP 50 | AR | AR 50 |
|---|---|---|---|---|---|---|
| RTMPose-S | None *(Baseline)* | - | 40.1 | 69.8 | 47.9 | 77.4 |
| ViTPose-S | None *(Baseline)* | - | 50.5 | 72.9 | 55.6 | 77.9 |
| | STP | 3,624 | 59.3 | 84.6 | 64.0 | 88.0 |
| | LSPe | 10,000 | 64.1 | 87.4 | 70.2 | 90.8 |
| | LSPe + STP *(Ours)* | 13,624 | **68.4** | **88.0** | **74.5** | **91.7** |
| | RePoGen * | - | 59.1 | 80.9 | 70.0 | 88.7 |

**2D HPE performance**

- LSPe and STP each improve accuracy, but best performance with LSPe + STP
- Our viewpoint-agnostic model outperforms RePoGen[6], a ViTPose fine-tuned on viewpoint-specific extreme poses

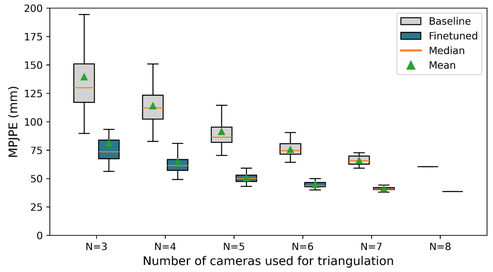



*GT poses are shown in gray behind the 3D reconstruction*

**Multi-view 3D reconstruction**

- MPJPE reduced by 46 mm (43%)
- Many views do not mitigate 2D HPE errors: triangulation can propagate them instead
- Our model yields better accuracy with 3 views than the baseline model with 8 views

*MPJPE distribution among combinations of N cameras, using baseline or fine-tuned model 2D predictions for triangulation*




## Discussion

- Ongoing work to test generalization to other athletes, routines or sports
- STP underperforming LSPe suggests that STP's size or diversity is insufficient or that the domain gap between real and synthetic data remains too important
- End-to-end toolchain for generating realistic SMPL models and training data of extreme poses
- **Proof of concept showing strong in-domain performance and reliable multi-view 3D pose reconstruction in real-world settings**

[1] Loper, M. et al. SMPL: a skinned multi-person linear model. ACM 2015.
[2] Su, Y. et al. ViTPose: simple vision transformer baselines for human pose estimation. NeurIPS 2022.
[3] Ghorbani, N. et al. SOMA: Solving optical marker-based MoCap automatically. ICCV 2021.
[4] Pagnon. D. Pose2Sim: An open-source Python package for multiview markerless kinematics. JOSS 2022.
[5] Johnson, S. & Everingham, M. Learning effective human pose estimation from inaccurate annotation. CVPR 2001.
[6] Purkrabek, M. et al. Improving 2D Human Pose Estimation in Rare Camera Views with Synthetic Data. FG 2024.