



Improving 2D Human Pose Estimation in Rare Camera Views with Synthetic Data

Miroslav Purkrabek and Jiri Matas

Visual Recognition Group, Faculty of Electrical Engineering, Czech Technical University in Prague

Introduction

Methods and datasets for human pose estimation focus on side- and front-views.

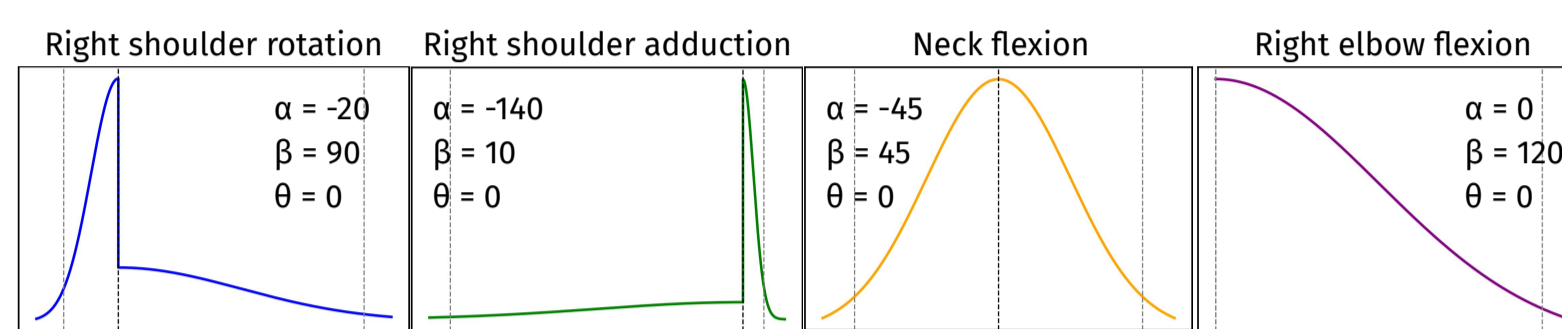
We propose RePoGen (Rare POses GENERator), an SMPL-based method for generating synthetic humans with comprehensive control of pose and view.

Experiments on real images with diverse poses show that, e.g., ViTPose, trained on COCO with added RePoGen data, outperforms prior art in top- and bottom-view pose estimation without performance loss on common views.

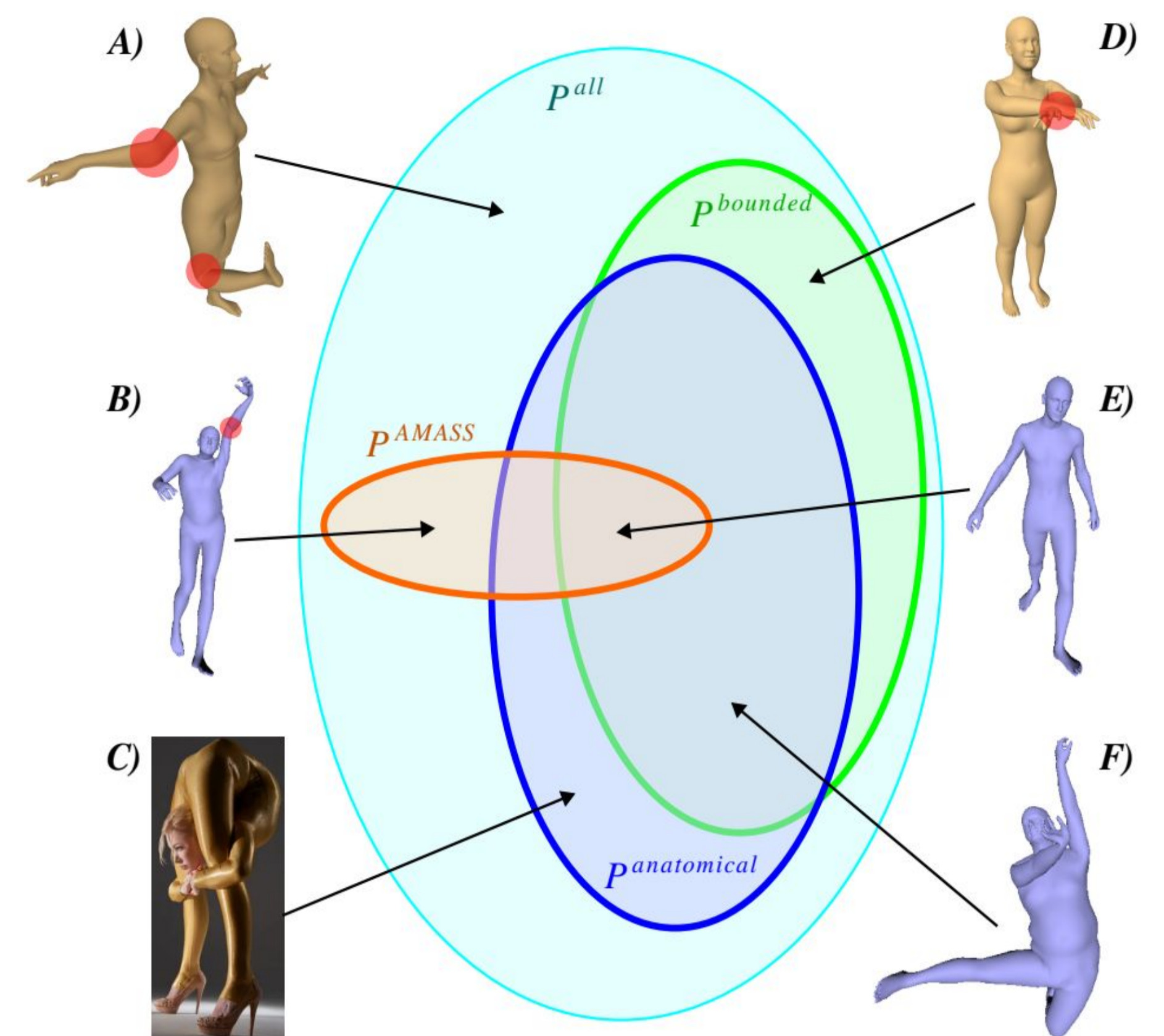
Ablation shows that anatomical plausibility, a focus of prior research, is not a prerequisite for effective performance.

RePoGen – a novel rare pose generation method

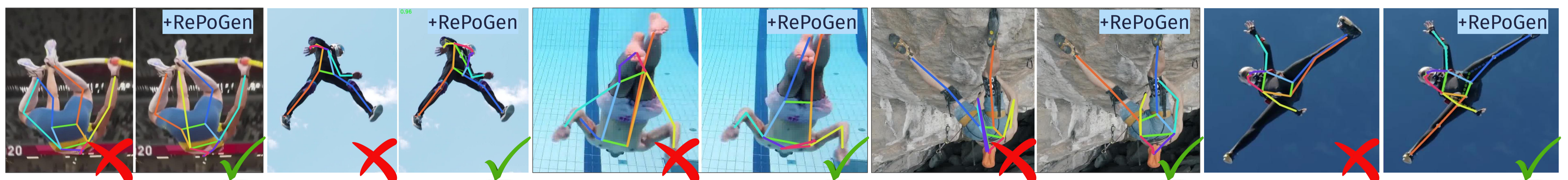
- each joint sampled independently
- ranges from physio textbook
- tradeoff between anatomical feasibility and novel poses that are hard to capture using MoCap systems
- a single distribution with parameters α , β and θ models all joints



- p_{all} - All possible with SMPL
- $p_{anatomical}$ - One human can achieve
- p_{AMASS} - Captured in AMASS
- $p_{bounded}$ - Generated by RePoGen



Results

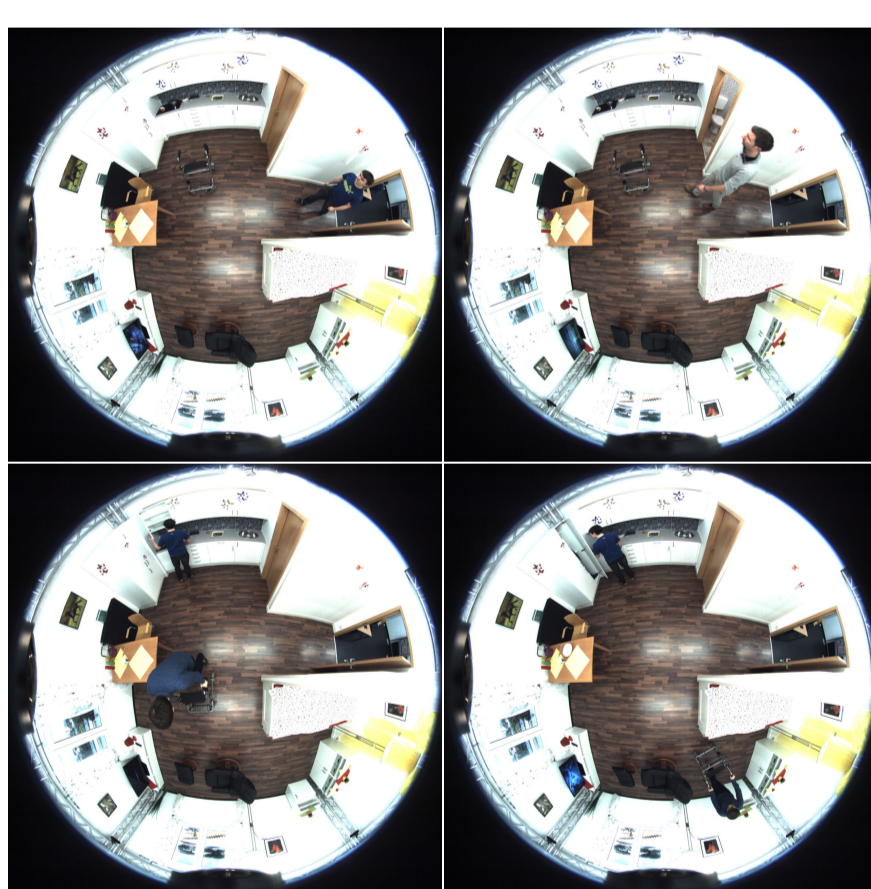


Datasets

RePo includes high percentage of top- and bottom-view images of rare poses despite having fewer images than previous datasets.

Existing Datasets: only PoseFES and WEPDToF-Pose available, both capturing common poses from ceiling-mounted fisheye cameras.

PoseFES 1



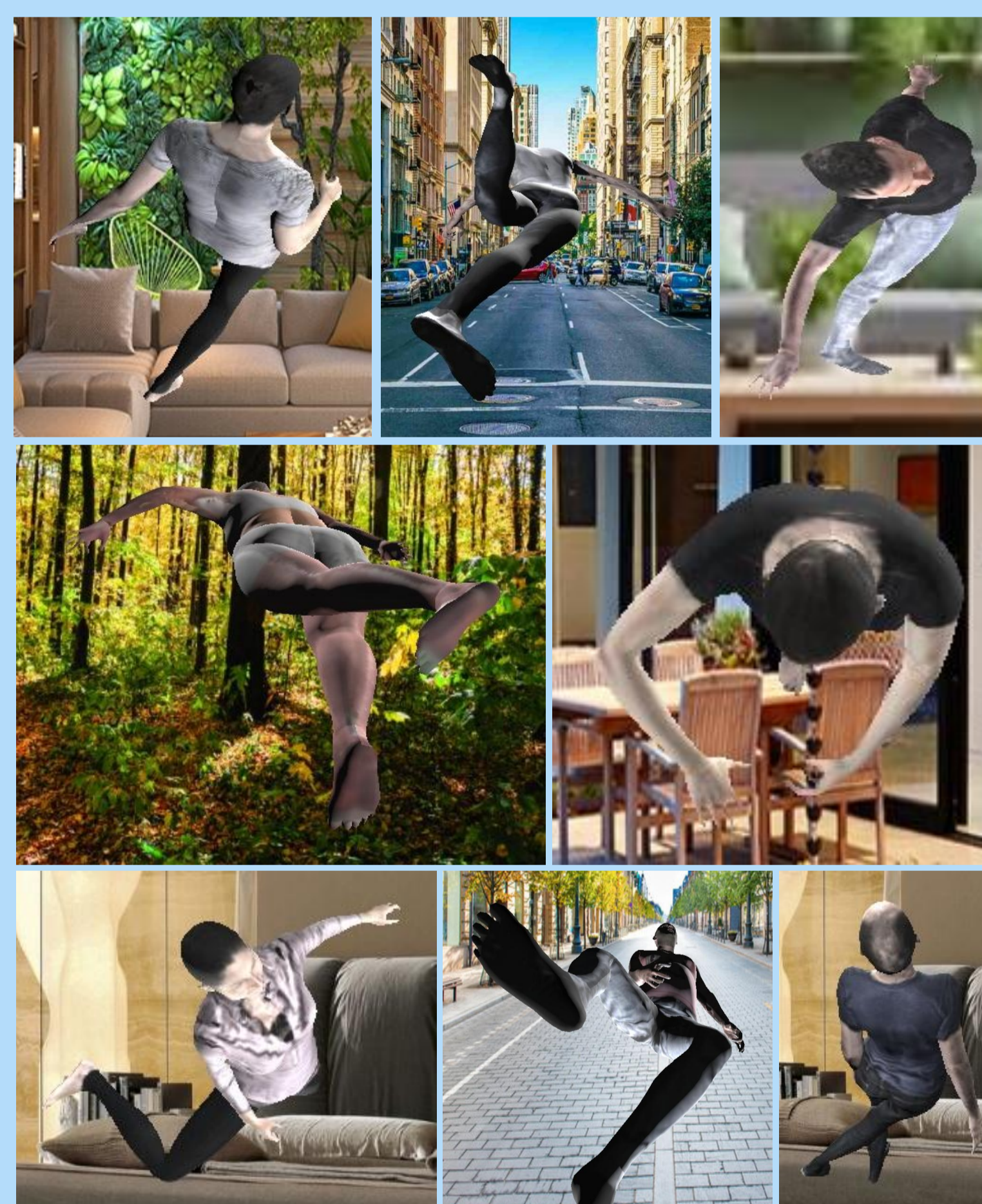
736 annotations
76 top-view

WEPDToF-Pose



6749 annotations
67 top-view

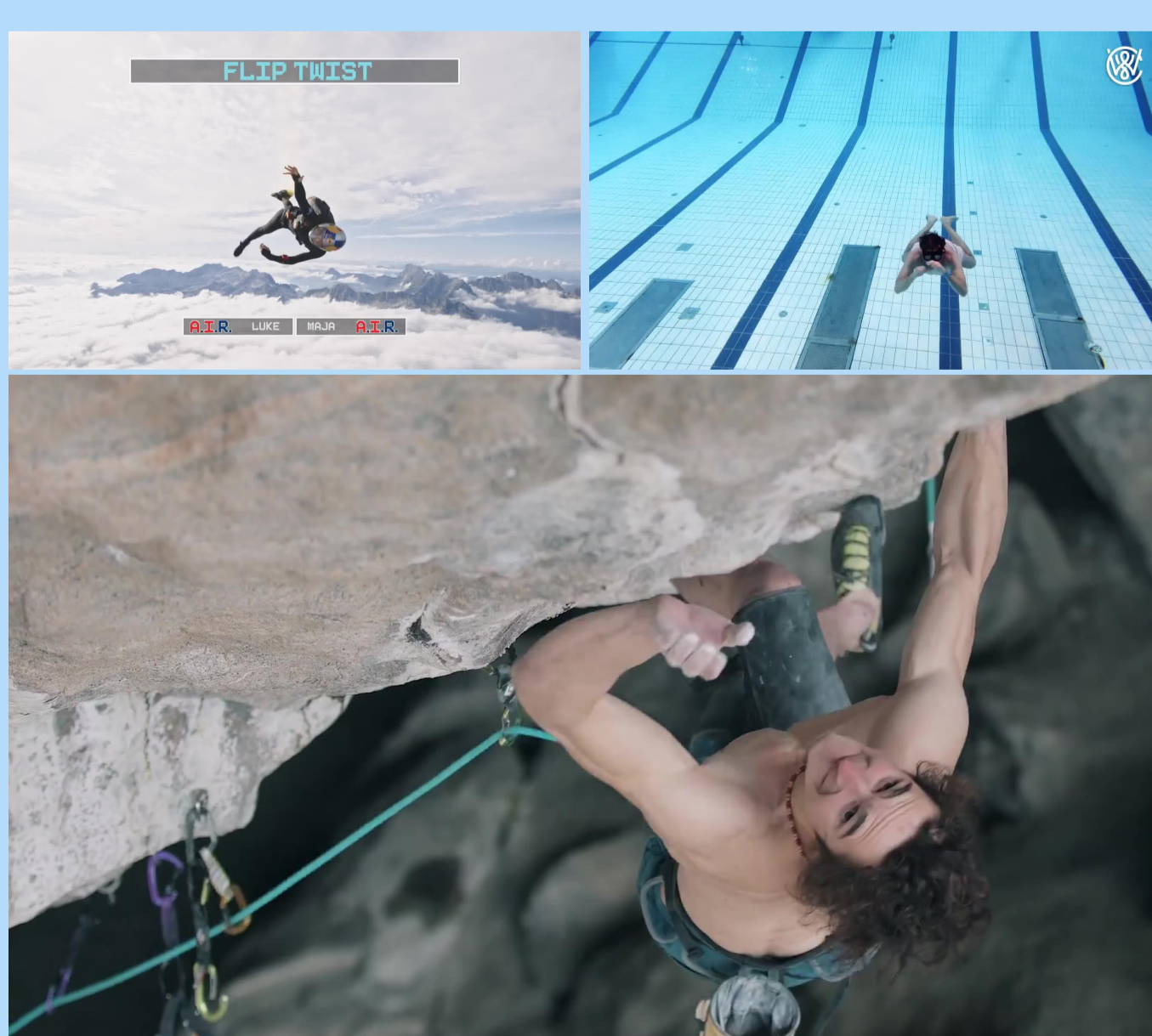
RePoGen (synthetic)



3000 annotations
all top or bottom view

RePo

Part: Top



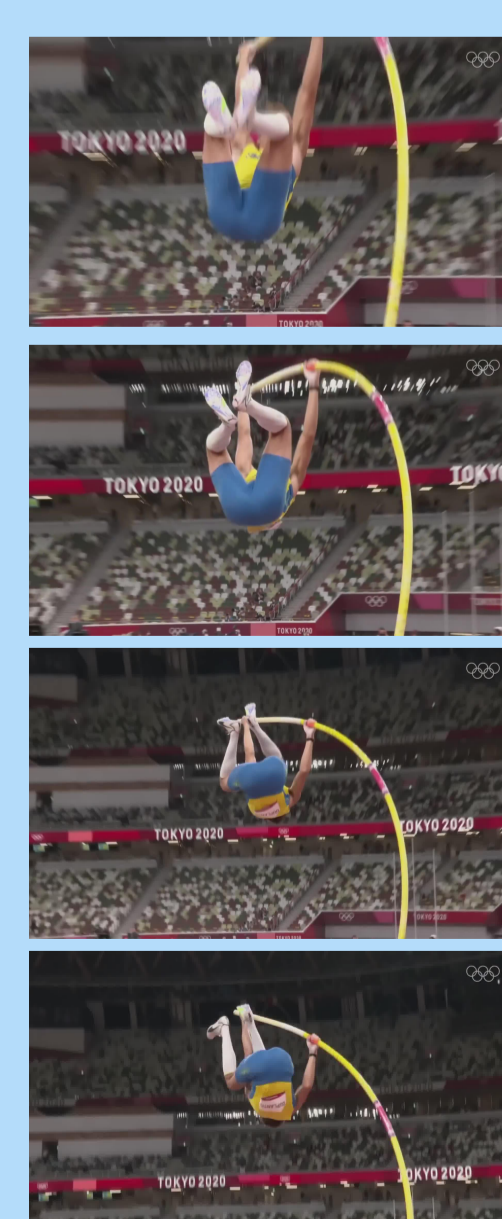
91 annotations,
all top-view

Part: Bottom



125 annotations
all bottom-view

Part: Seq

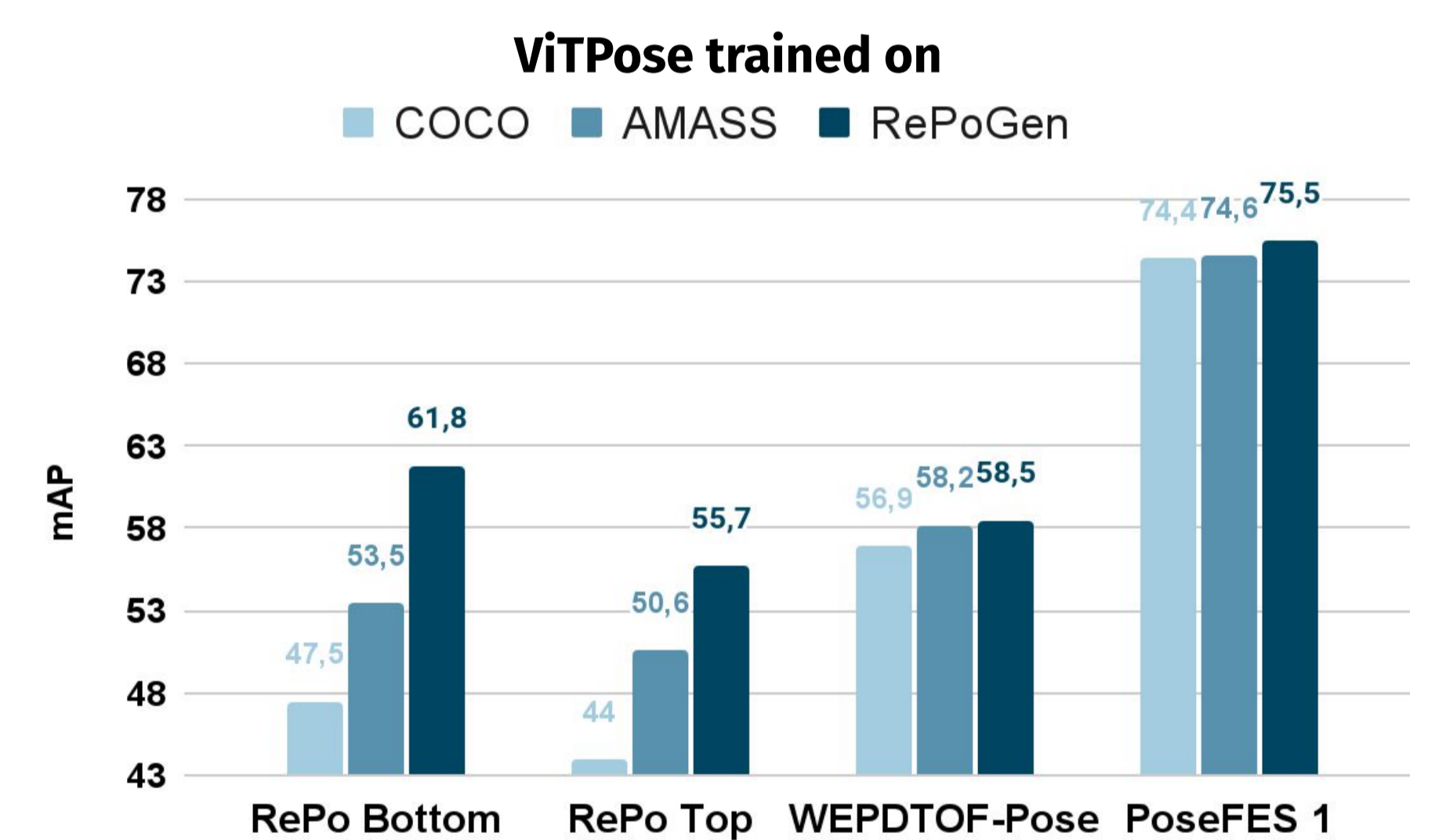


62 annotations
all views

Performance Evaluation

ViTPose trained on RePoGen performs:

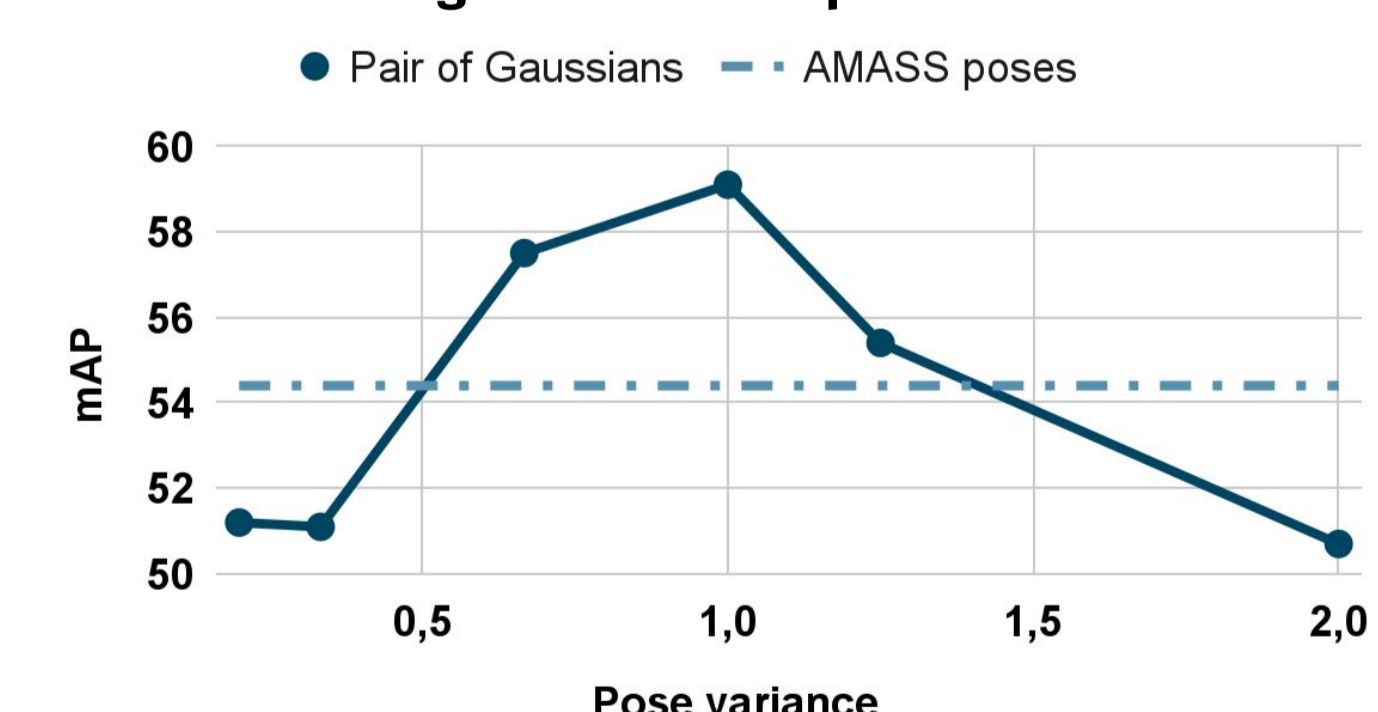
- Better than THEODORE+ and AMASS Poses
- The difference is more evident on RePo
- Keeps SOTA performance on orbital views



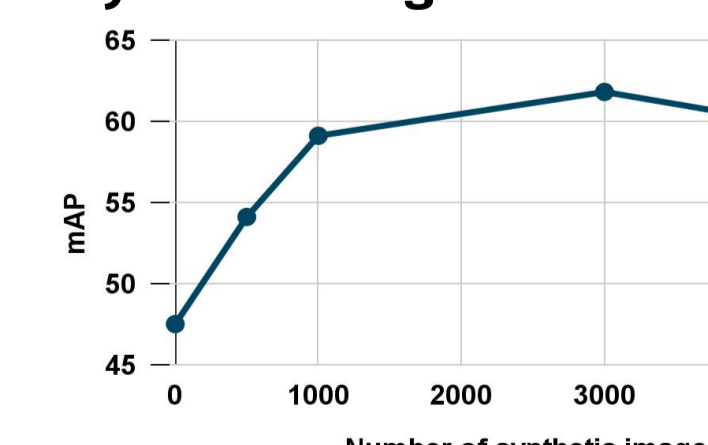
Ablation study

- Both too "unrealistic" and too many common poses hinder training.
- Synthetic images create domain gap
- Visual realism \leq data diversity.

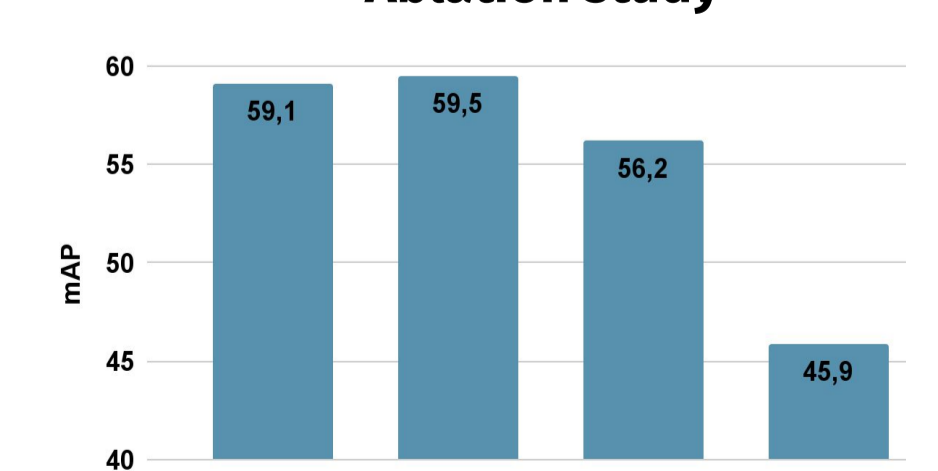
Training with different pose variance



Training with different amount of synthetic images



Ablation study



Code, models, datasets:

MiraPurkrabek.github.io/RePoGen-paper/

