PoseFix: Correcting 3D Human Poses with Natural Language

- Ginger Delmas, Philippe Weinzaepfel, Francesc Moreno-Noguer, Gregory Rogez
- NAVER LABS Europe

- Prakash K Naikade

Problem

- **Correctional Text Generative Model**
  - **Given:** 3D Source pose and 3D Target pose
  - **Predict:** Correctional text generation
- **Text-based Pose Editing Model**
  - **Given:** 3D Target pose and Textual Prompt
  - **Predict:** 3D Target pose

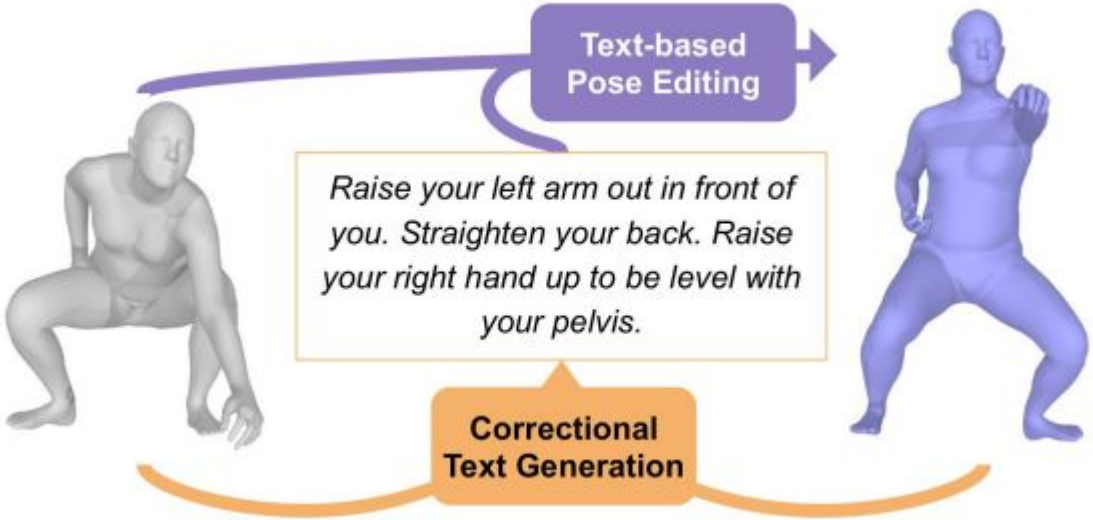
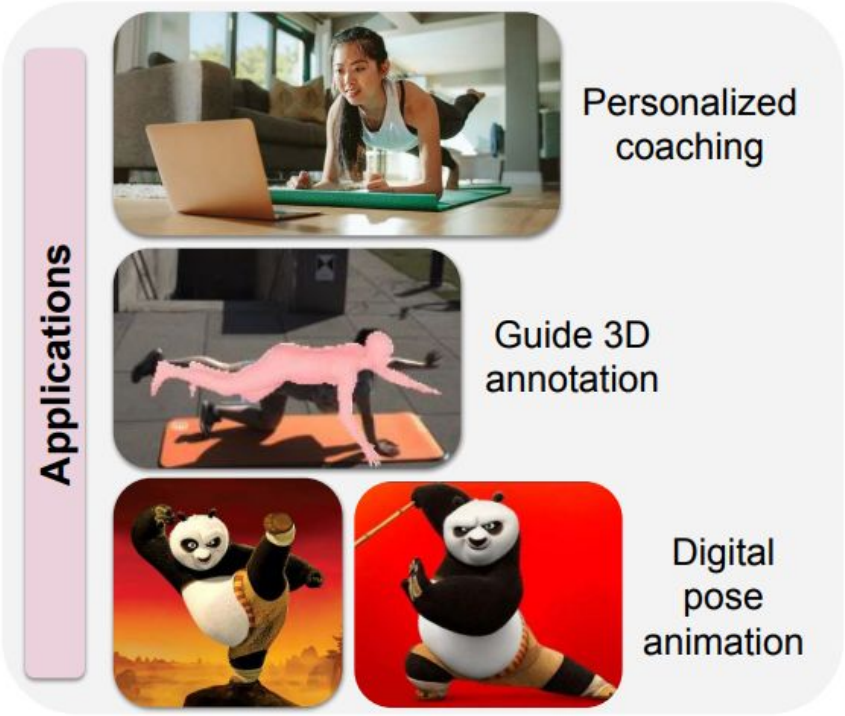


Fig 1: Problem description - Pose Fix

Motivation



- Sports Coaching
- Rehabilitation
- Labour Training
- Safety Instruction while handling dangerous machines
- Animations

Fig 2: Applications of PoseFix Models

Motivation

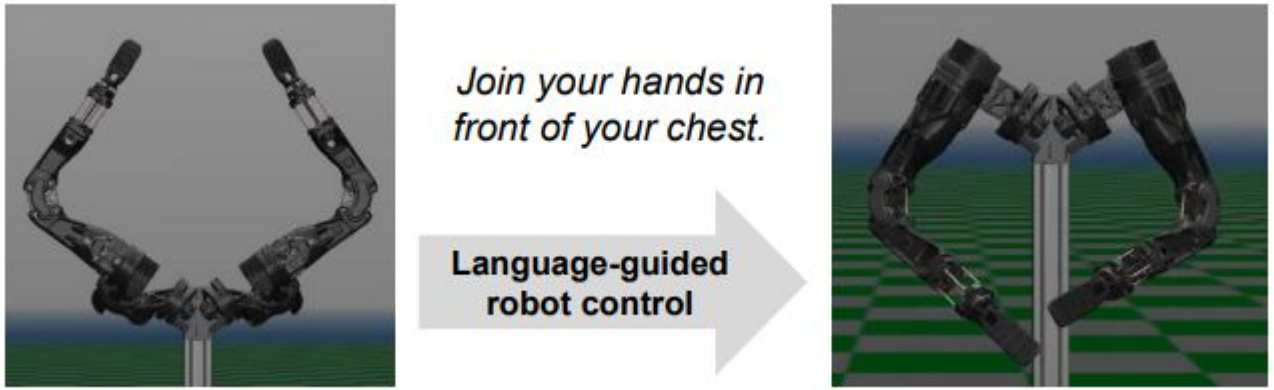


Fig 2a: Robot teaching application

Key Contribution

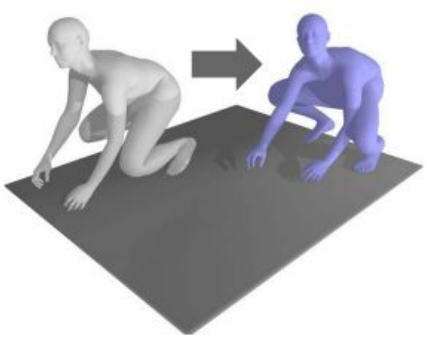
1. **PoseFix dataset:** 3D pose pairs with modifying instructions
2. **Text-based Pose Editing Model**
3. **Correctional Text Generative Model**

The PoseFix Dataset

- {pose A, pose B, text modifier}
- 135k pose pairs
- 3D human body poses were sampled from AMASS
- 6k human-written texts

Data collection on AMT

Stretch your thighs apart and project the knees forward so that they remain just along the elbows and then turn your face slightly to the left.



Automatic Comparative Pipeline

Your right thigh must be parallel to the floor while your right knee is bent to maximum, bring your right foot forward slightly, your right hand must be on the floor and your hands should be shoulder width apart.

Fig 3 - PoseFix Dataset

The PoseFix Dataset

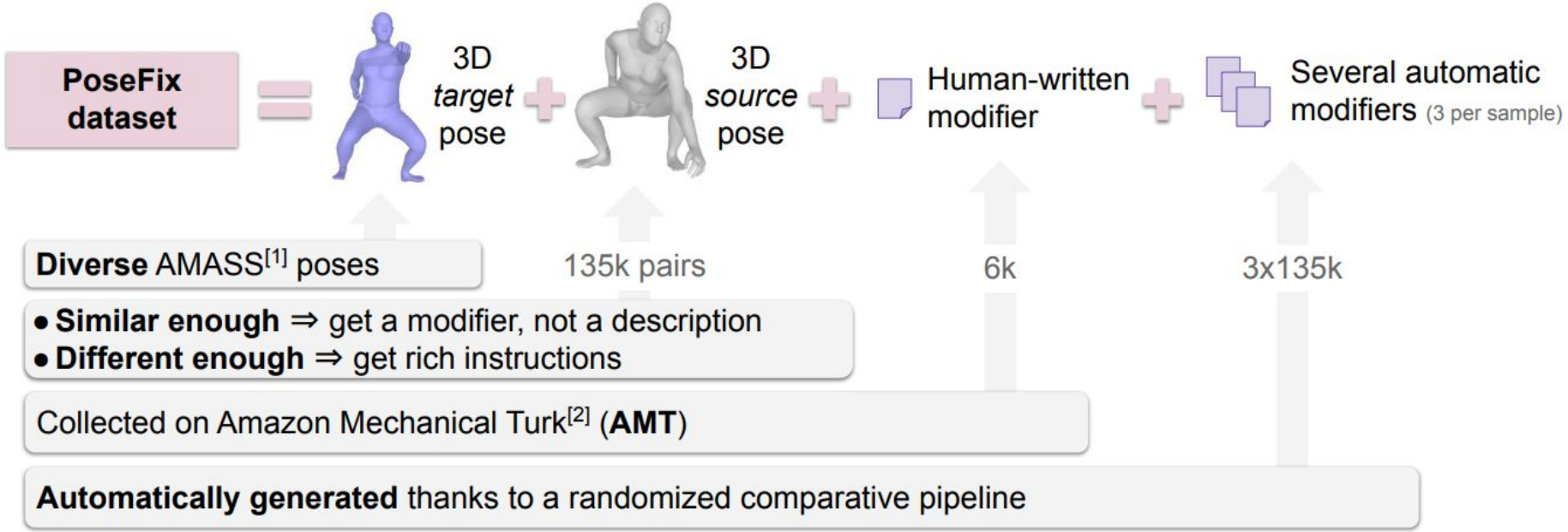


Fig 4 - PoseFix Dataset Generation

The PoseFix Dataset

- Randomized comparative pipeline for automatic generation of modifiers:
  - more training data at no cost: generate >10k modifiers in the time it takes to write 1!

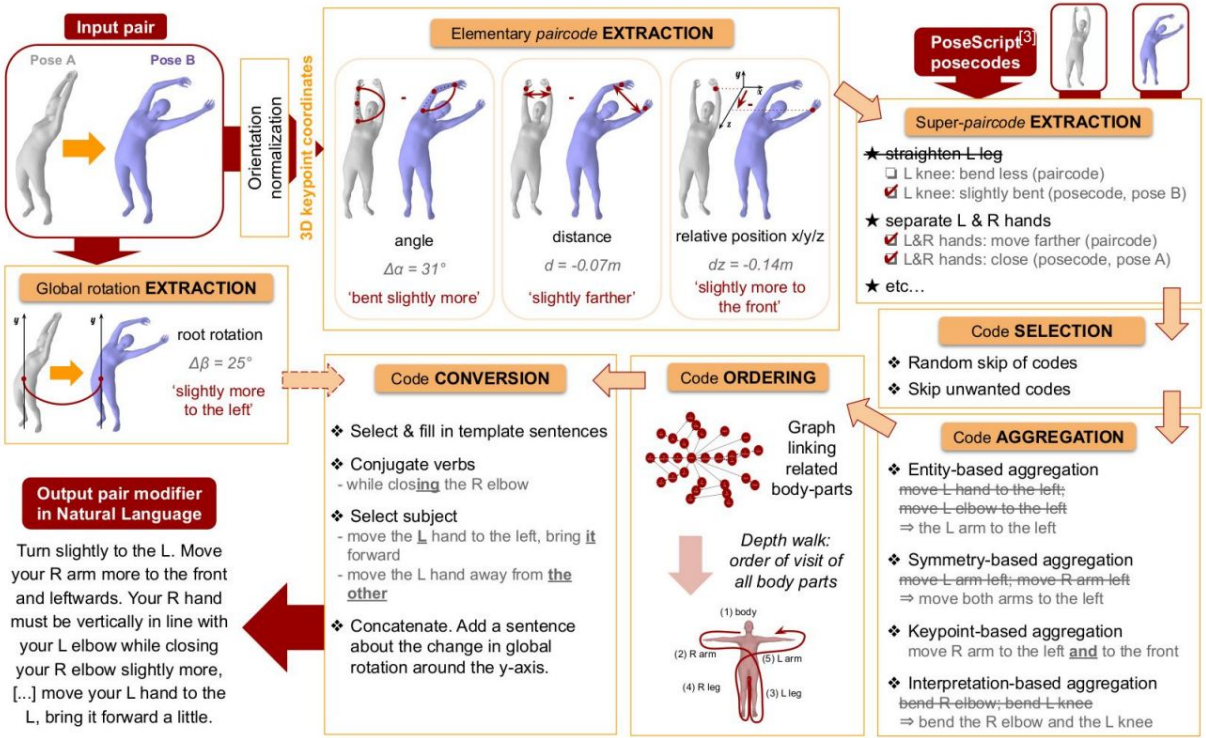


Fig 5 - PoseFix Dataset Generation - Randomized comparative pipeline

The PoseFix Dataset

- In-sequence pairs (IS)
- Out-of-sequence pairs (OOS)

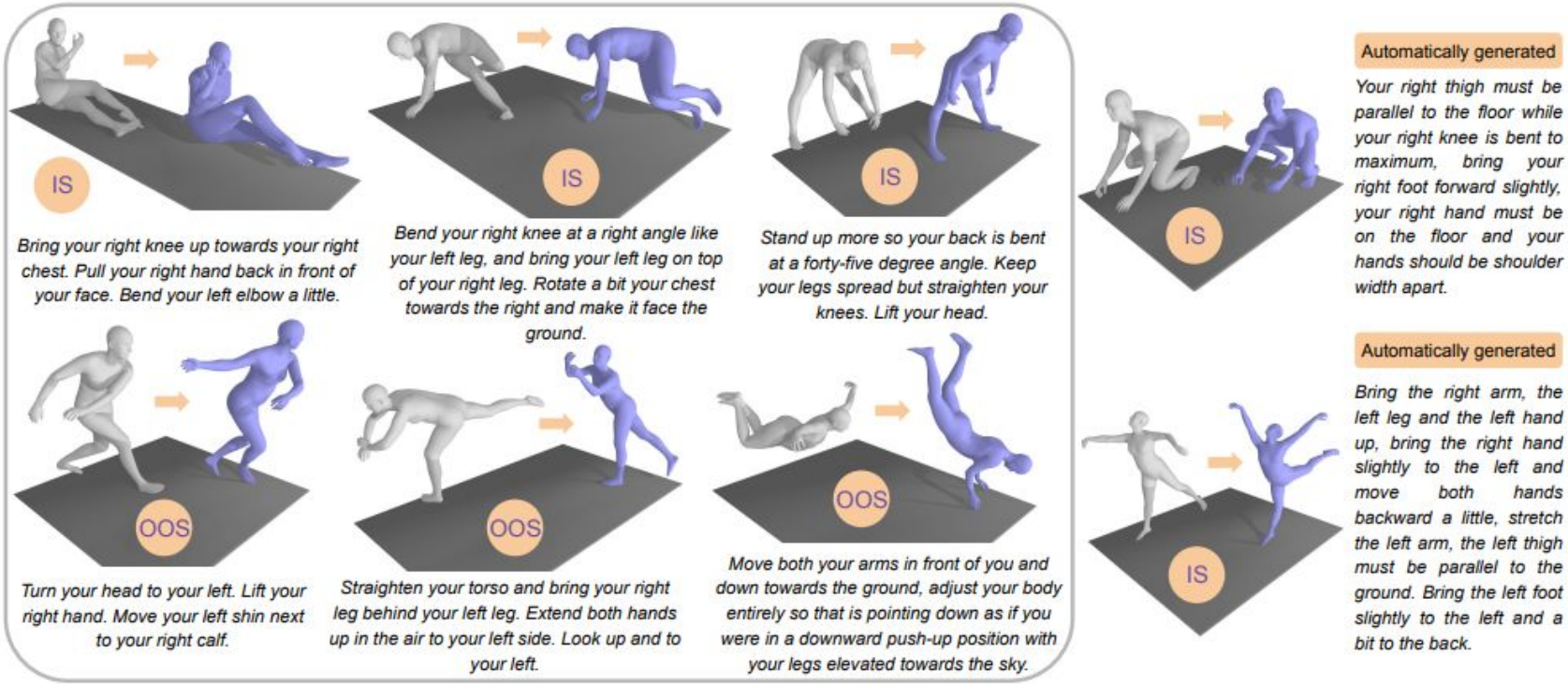
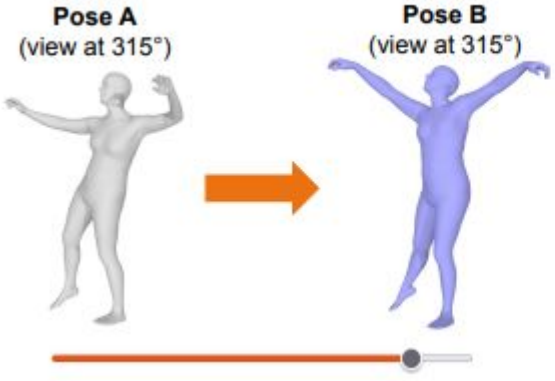


Fig 6 - Examples of pose pairs and their annotated modifier in PoseFix

The PoseFix Dataset



Left: Data presented to the annotators. The slider makes it possible to look at the poses under different viewpoints.

Right: word cloud of the PoseFix annotations.

Fig 7 - Annotators

Text-based Pose Editing

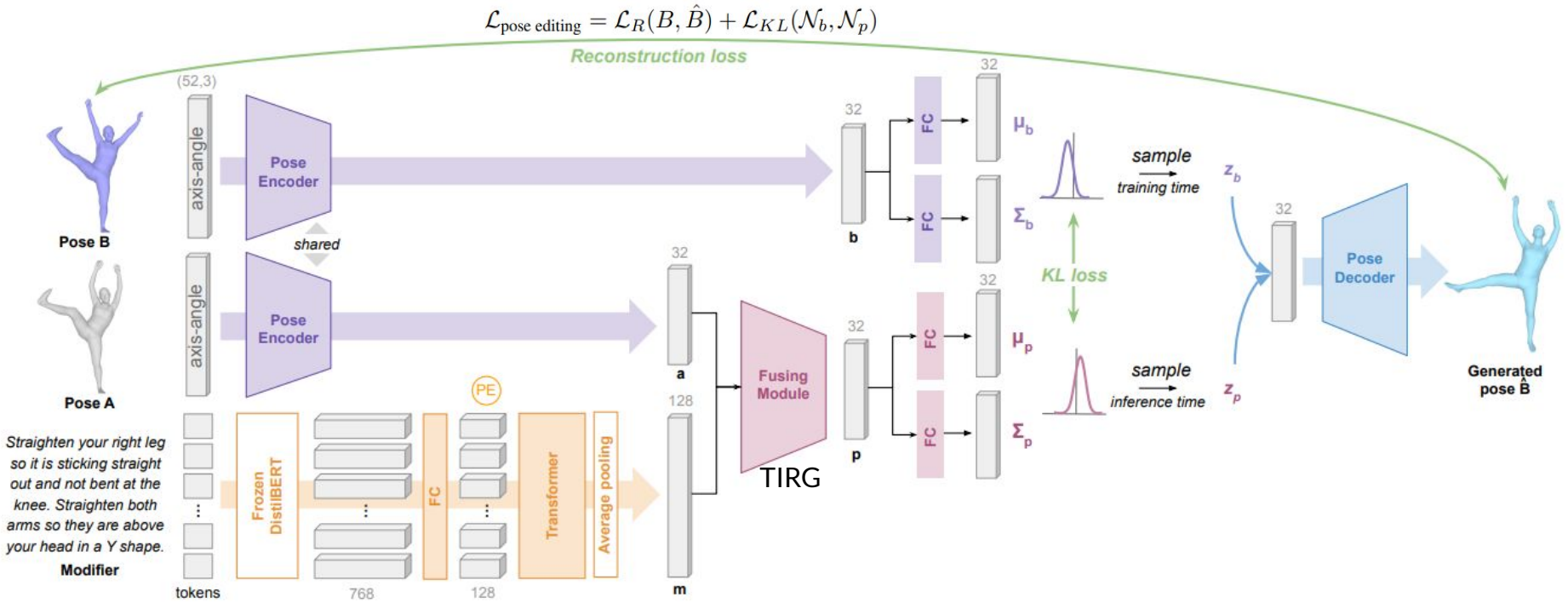
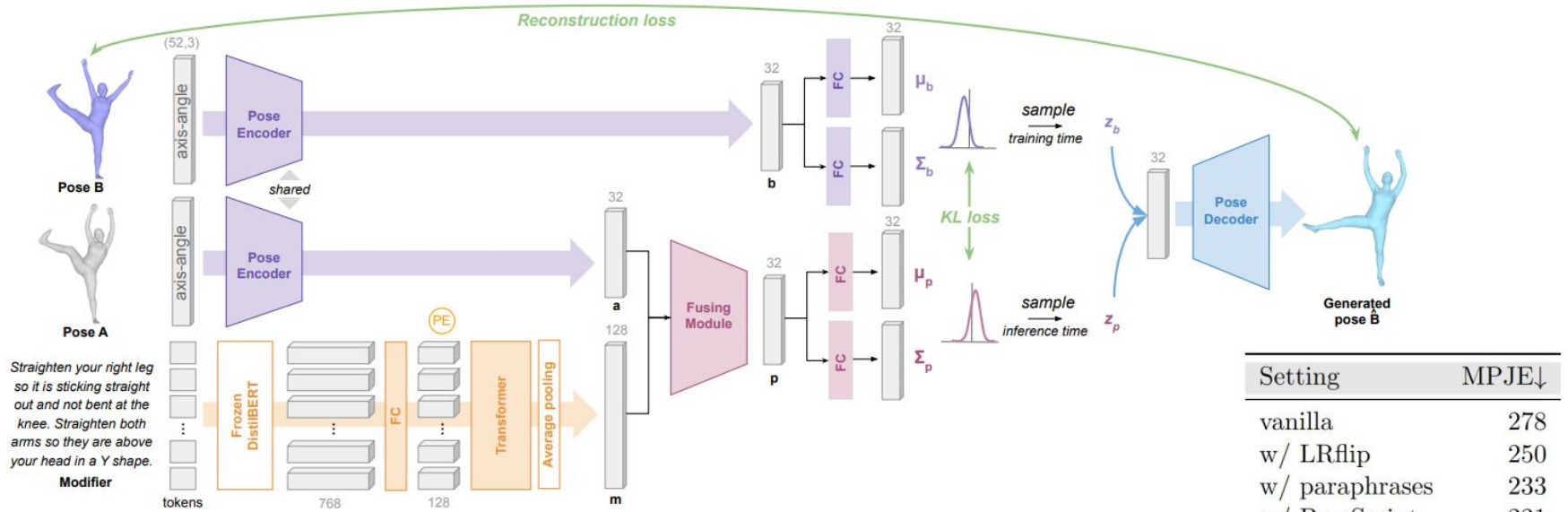


Fig 8: Overview of our text-based pose editing baseline

Text-based Pose Editing



Setting	MPJE↓
vanilla	278
w/ LRflip	250
w/ paraphrases	233
w/ PoseScript	221
w/ pretraining	199

Fig 8: Overview & Evaluation -The top part represents a standard VAE, where poses are encoded into a Gaussian distribution. At training time, a latent variable is sampled and decoded into a pose to learn pose reconstruction. The bottom left part represents the conditioning: the text is encoded using a frozen DistilBERT with a small transformer on top. It is combined with source pose features in the fusion module, from which we predict a Gaussian distribution. A KL loss ensures the alignment of the distributions from the standard VAE and the conditioning. At test time, we sample from the latter to predict the target pose.

Text-based Pose Editing

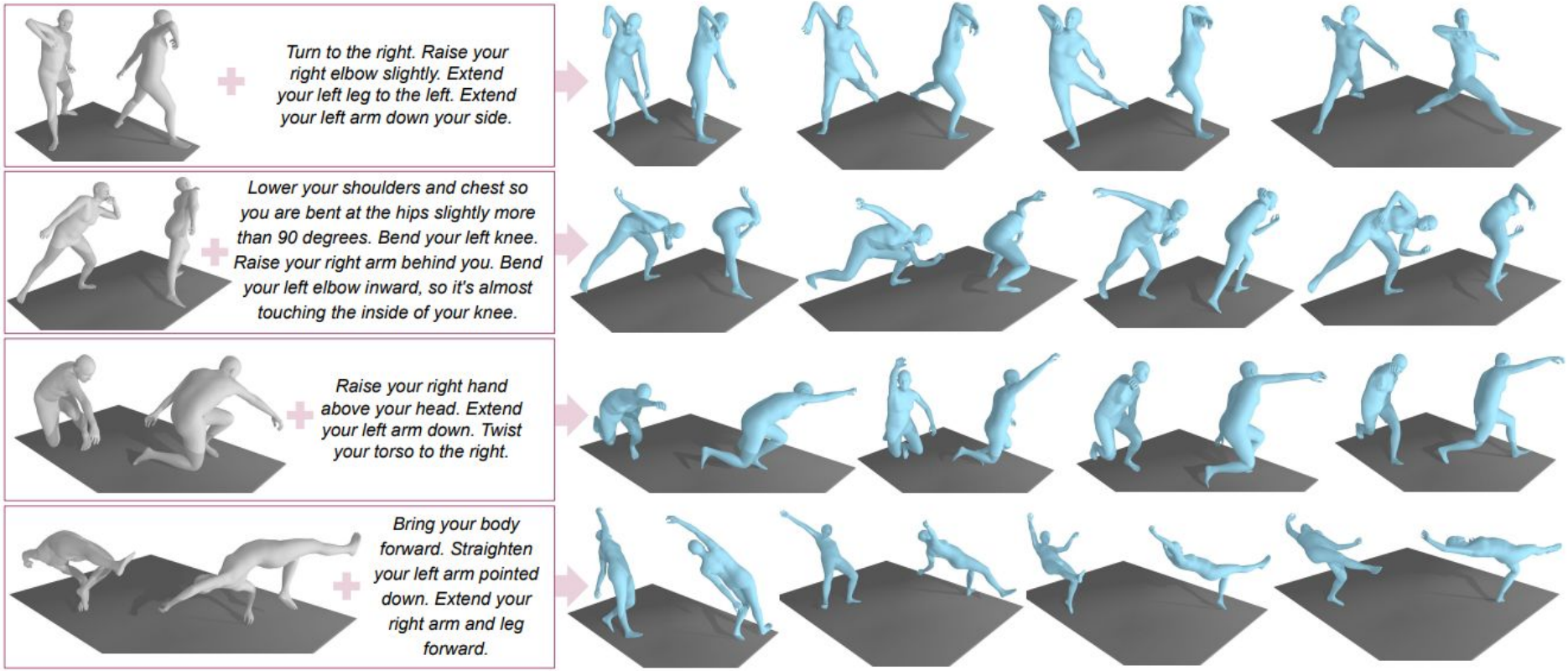


Fig 9: Generated poses for the text-based pose editing task

Correctional Text Generation

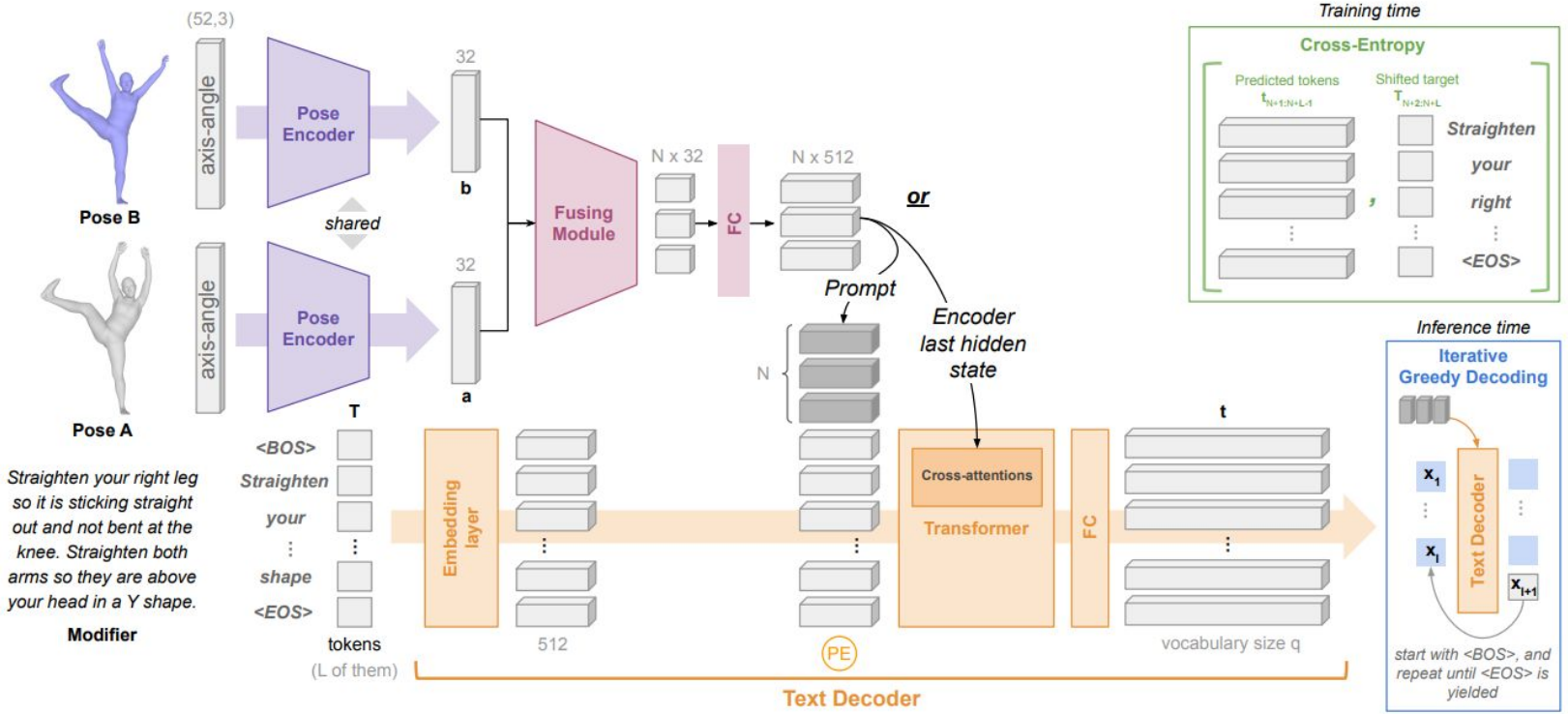
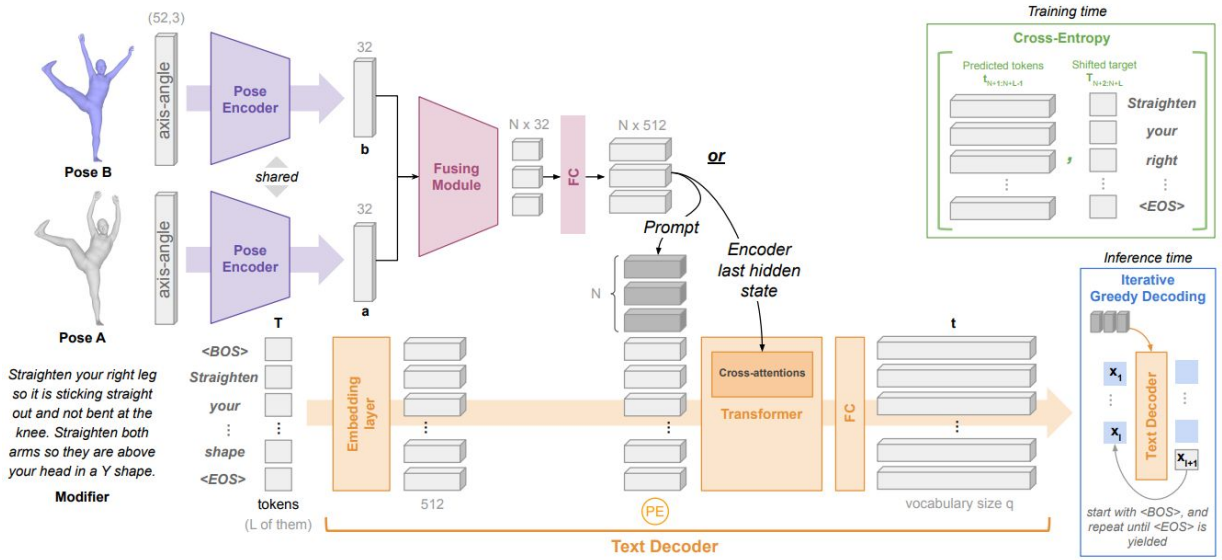


Fig 10: Overview of our baseline for correctional text generation

Correctional Text Generation



Setting	R@1-precision↑
random text	3.1
original text	62.7
vanilla	6.8
with pretraining	58.4
+ LRflip	60.7

Fig 11: Overview & Evaluation: The bottom part represents a standard auto-regressive transformer model: the next word is predicted from the previously generated tokens. The decoder outputs a distribution of probabilities over the vocabulary for each token. The top part represents the conditioning on the pose pair: the two pose embeddings are fused together into a set of “pose tokens”, further used for conditioning via prompting or via cross-attentions in the transformer. At inference, the modifier is generated iteratively using the greedy approach.

Correctional Text Generation

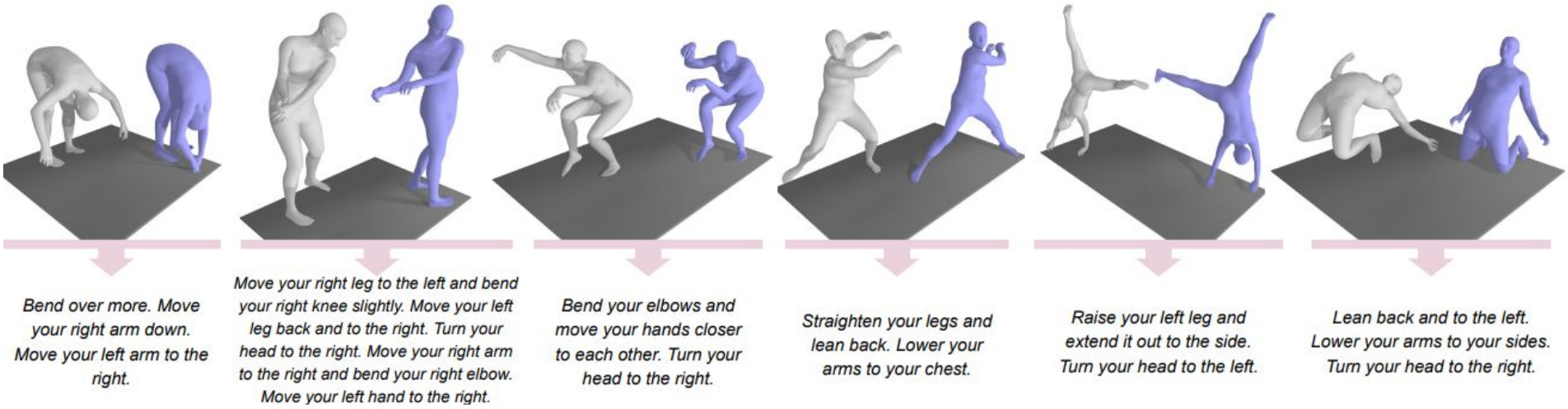
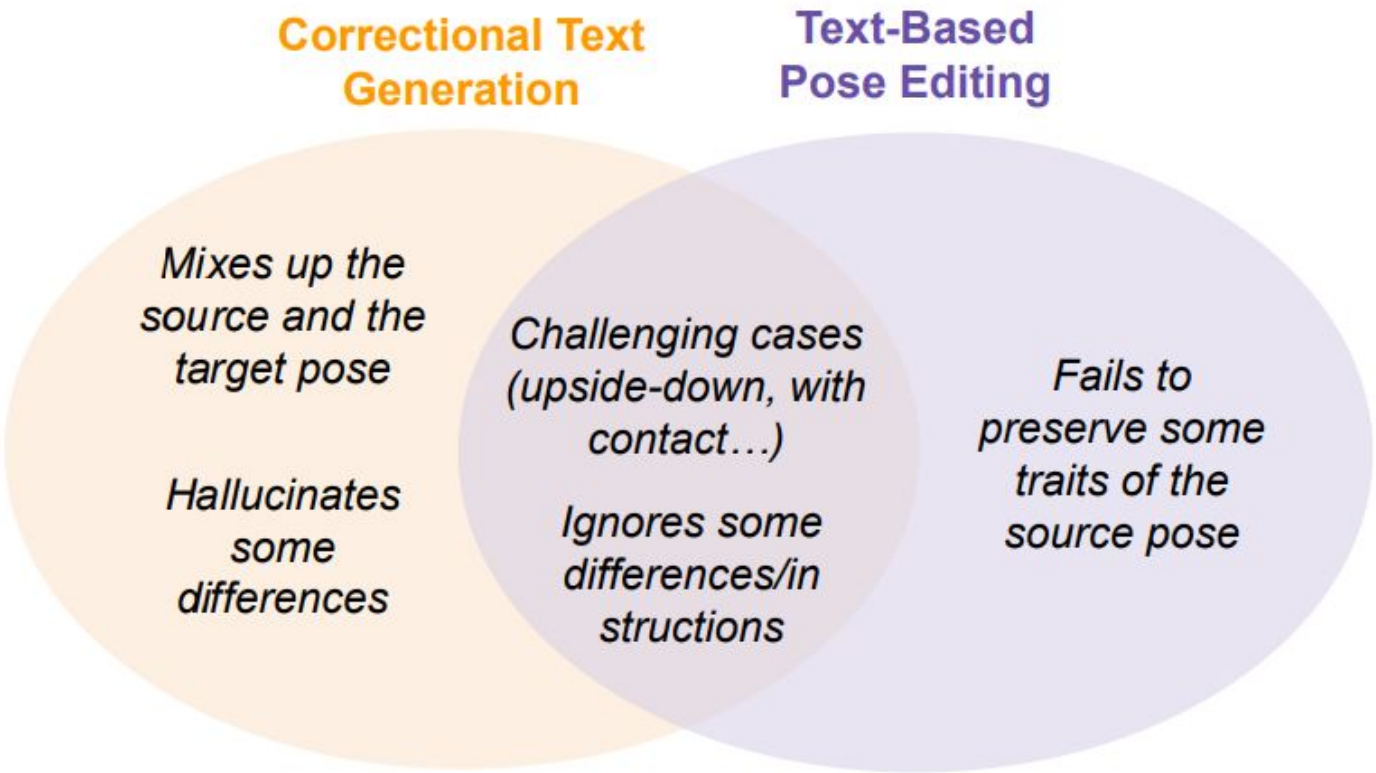
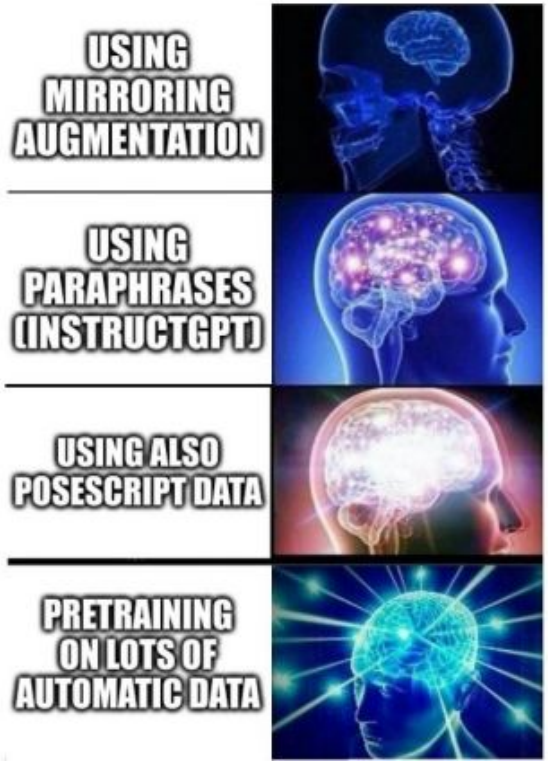


Fig 12: Generated correctional texts for PoseFix pose pairs

Challenges & Limitations



Conclusions



Thank You !

Questions Please !