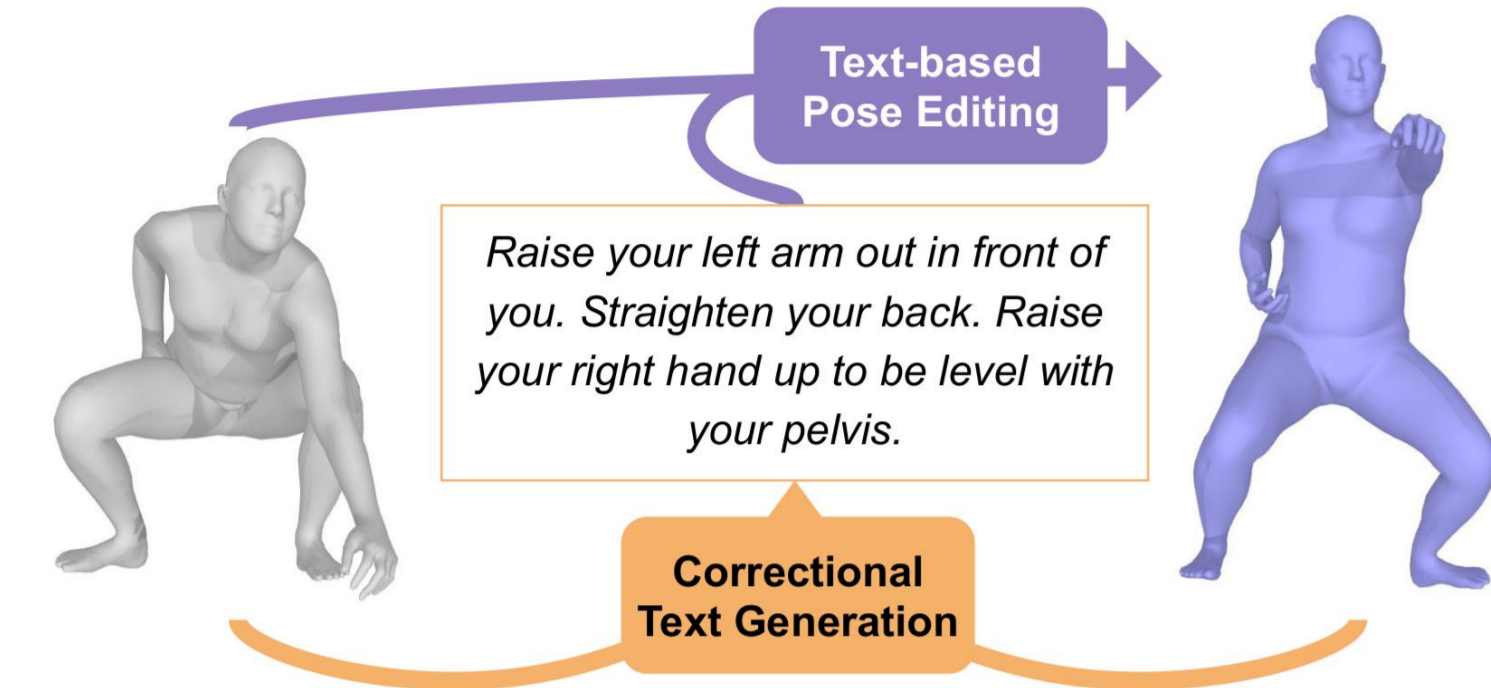




Motivation & Contributions

Using text to learn fine-grained 3D body pose semantics.



Applications

- Personalized coaching
- Guide 3D annotation
- Digital pose animation

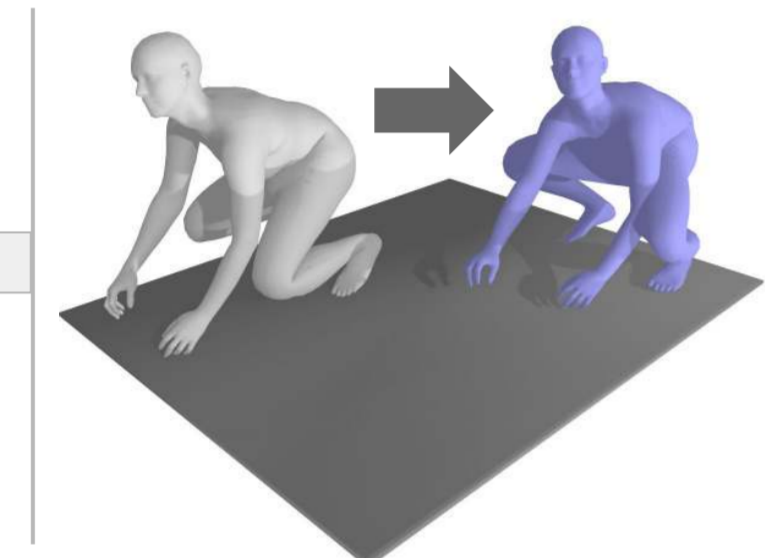
- ❖ **PoseFix dataset**: 3D pose pairs with modifying instructions
- ❖ **Text-based Pose Editing** model
- ❖ **Correctional Text Generative** model

The PoseFix dataset

135k pose pairs, 6k human-written texts

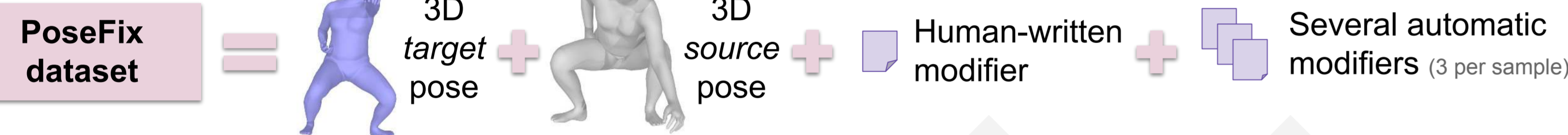
Data collection on AMT

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.



Diverse AMASS^[1] poses

135k pairs

6k

3x135k

- **Similar enough** ⇒ get a modifier, not a description
- **Different enough** ⇒ get rich instructions

Collected on Amazon Mechanical Turk^[2] (AMT)

Automatically generated thanks to a randomized comparative pipeline

Get more training data at no cost: generate >10k modifiers in the time it takes to write 1!

Input pair (Pose A, Pose B) → **Orientation normalization** → **3D keypoint coordinates**

Elementary paircode EXTRACTION

- angle: $\Delta\alpha = 31^\circ$ → 'bent slightly more'
- distance: $d = -0.07m$ → 'slightly farther'
- relative position x/y/z: $dz = -0.14m$ → 'slightly more to the front'

Global rotation EXTRACTION

- root rotation: $\Delta\beta = 25^\circ$ → 'slightly more to the left'

Output pair modifier in Natural Language

Turn slightly to the L. Move your R arm more to the front and leftwards. Your R hand must be vertically in line with your L elbow while closing your R elbow slightly more, [...] move your L hand to the L, bring it forward a little.

Code SELECTION

- ❖ Random skip of codes
- ❖ Skip unwanted codes

Code ORDERING

- ❖ Graph linking related body-parts
- ❖ Depth walk: order of visit of all body parts

Code CONVERSION

- ❖ Select & fill in template sentences
- ❖ Conjugate verbs - while closing the R elbow
- ❖ Select subject - move the L hand to the left, bring it forward - move the L hand away from the other
- ❖ Concatenate. Add a sentence about the change in global rotation around the y-axis.

Code AGGREGATION

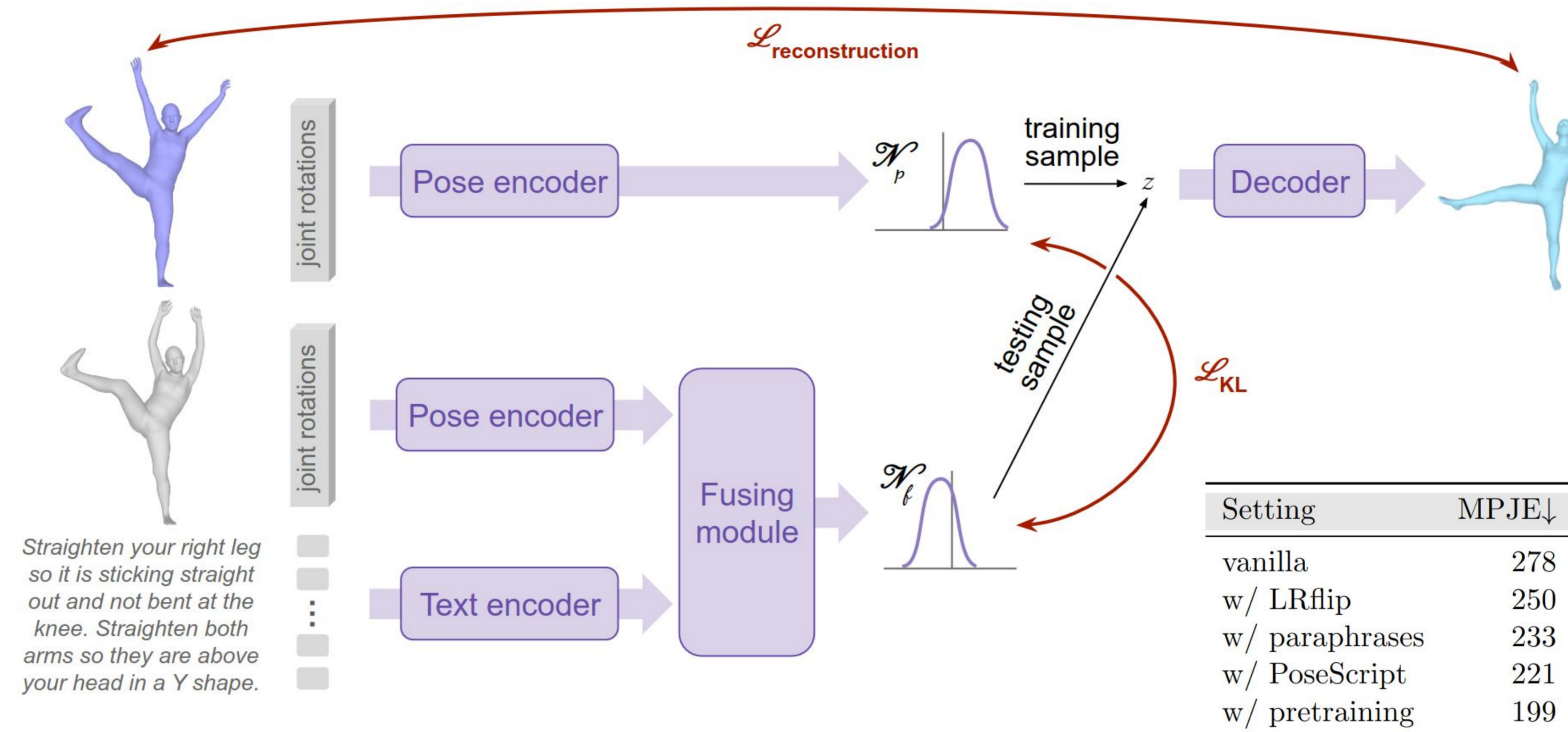
- ❖ Entity-based aggregation: move L hand to the left; move L elbow to the left ⇒ the L arm to the left
- ❖ Symmetry-based aggregation: move L arm-left; move R arm-left ⇒ move both arms to the left
- ❖ Keypoint-based aggregation: move R arm to the left and to the front
- ❖ Interpretation-based aggregation: bend R elbow; bend L knee ⇒ bend the R elbow and the L knee

PoseScript^[3] posecodes

- ★ straighten L leg
- ❑ L knee: bend less (paircode)
- ☑ L knee: slightly bent (posecode, pose B)
- ★ separate L & R hands
- ☑ L&R hands: move farther (paircode)
- ☑ L&R hands: close (posecode, pose A)
- ★ etc...

Text-based Pose Editing

⇒ Conditional VAE^[4]



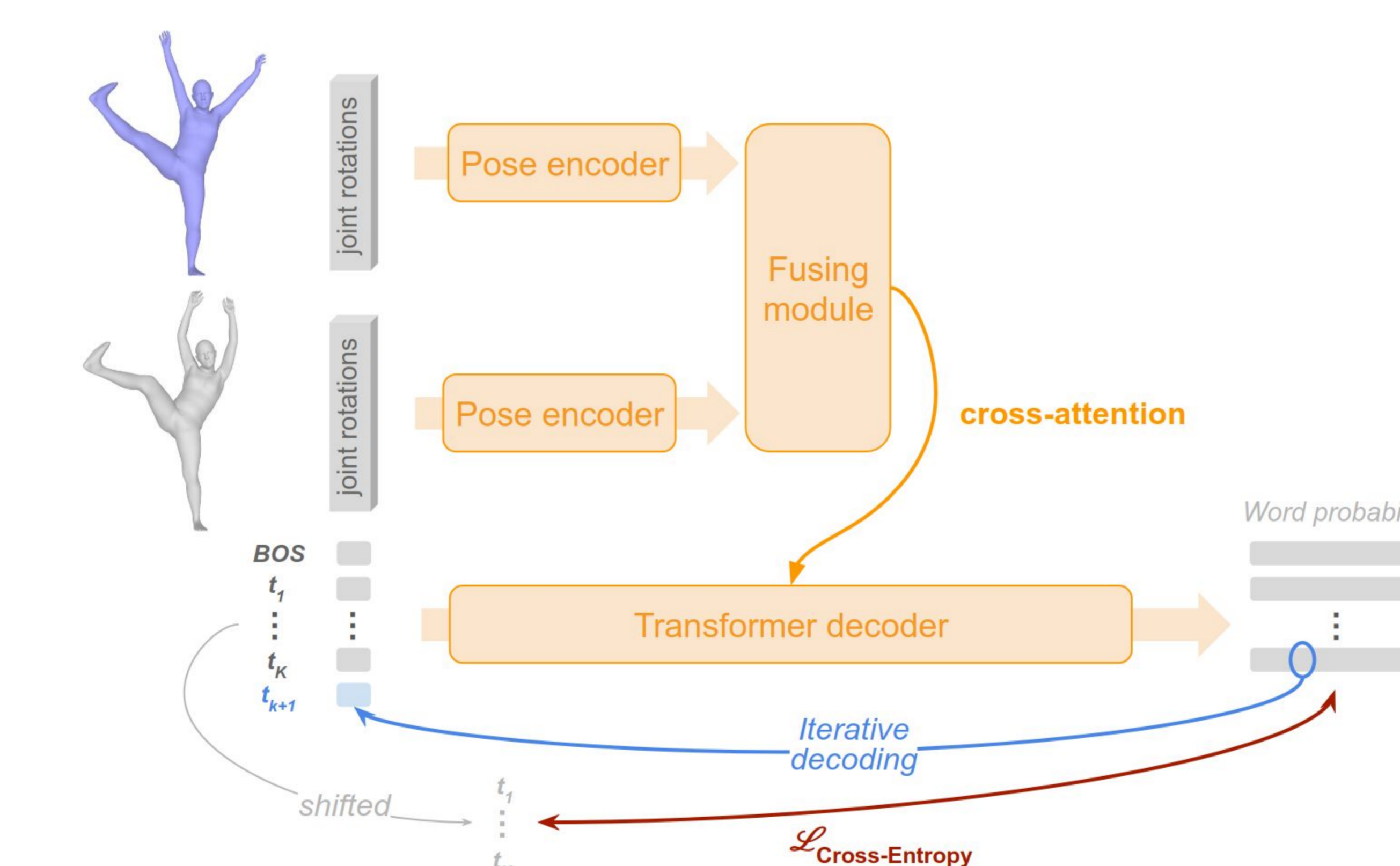
Setting	MPJE↓
vanilla	278
w/ LRflip	250
w/ paraphrases	233
w/ PoseScript	221
w/ pretraining	199

Text-based Pose Editing Examples:

- Move your left hand to the right. Extend your right arm behind you. Turn your head slightly to the right.
- Lower your shoulders and chest so you are bent at the hips slightly more than 90 degrees. Bend your left knee. Raise your right arm behind you. Bend your left elbow inward, so it's almost touching the inside of your knee.

Correctional Text Generation

⇒ Auto-regressive model

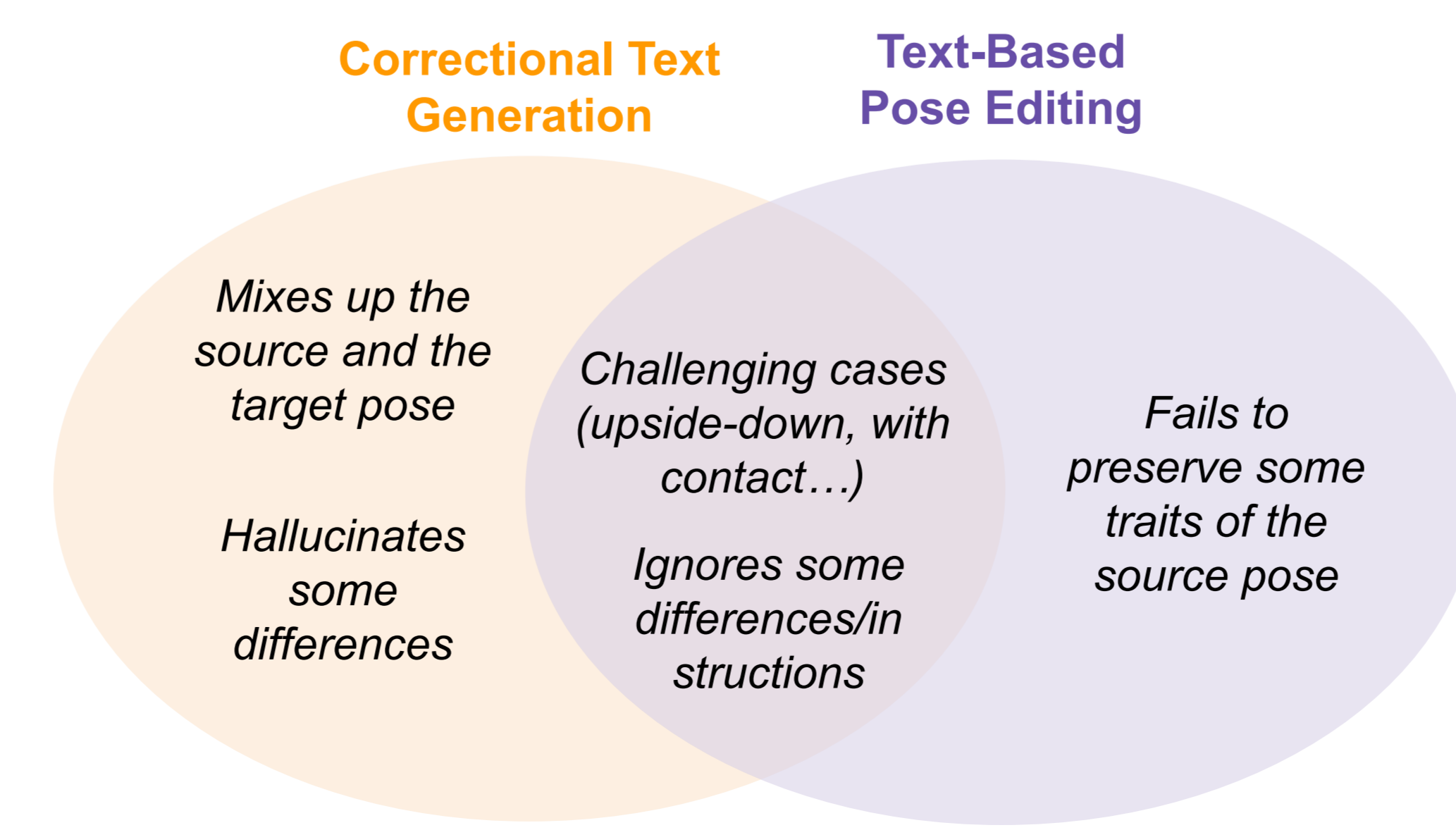


Correctional Text Generation Examples:

- Bend over more. Move your right arm down. Move your left arm to the right.
- Straighten your legs and lean back. Lower your arms to your chest.
- Bend your elbows and move your hands closer to each other. Turn your head to the right.
- Lean back and to the left. Lower your arms to your sides. Turn your head to the right.

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

Challenges & limitations



Some conclusions

[text-based pose generation]

- USING MIRRORING AUGMENTATION
- USING PARAPHRASES (INSTRUCTGPT)
- USING ALSO POSESCRIPT DATA
- PRETRAINING ON LOTS OF AUTOMATIC DATA

Training with several texts for each pose

Training with only 1 text/pose but more poses

References

- [1] AMASS: Archive of Motion Capture as Surface Shapes, Mahmood et al., ICCV 2019
- [2] <https://www.mturk.com/>
- [3] PoseScript: 3D Human Poses from Natural Language, Delmas et al., ECCV 2022
- [4] Auto-encoding variational bayes, Kingma and Welling, ICLR, 2014
- [5] Training language models to follow instructions with human feedback, Ouyang et al., NeurIPS 2022