Animate-X: Universal Character Image Animation with Enhanced Motion Representation

Shuai Tan[1], Biao Gong[1], Xiang Wang[2], Shiwei Zhang[2],
Dandan Zheng[1], Ruobing Zheng[1], Kecheng Zheng[1], Jingdong Chen[1], Ming Yang[1]
[1] Ant Group, [2] Tongyi Lab

ANT GROUP
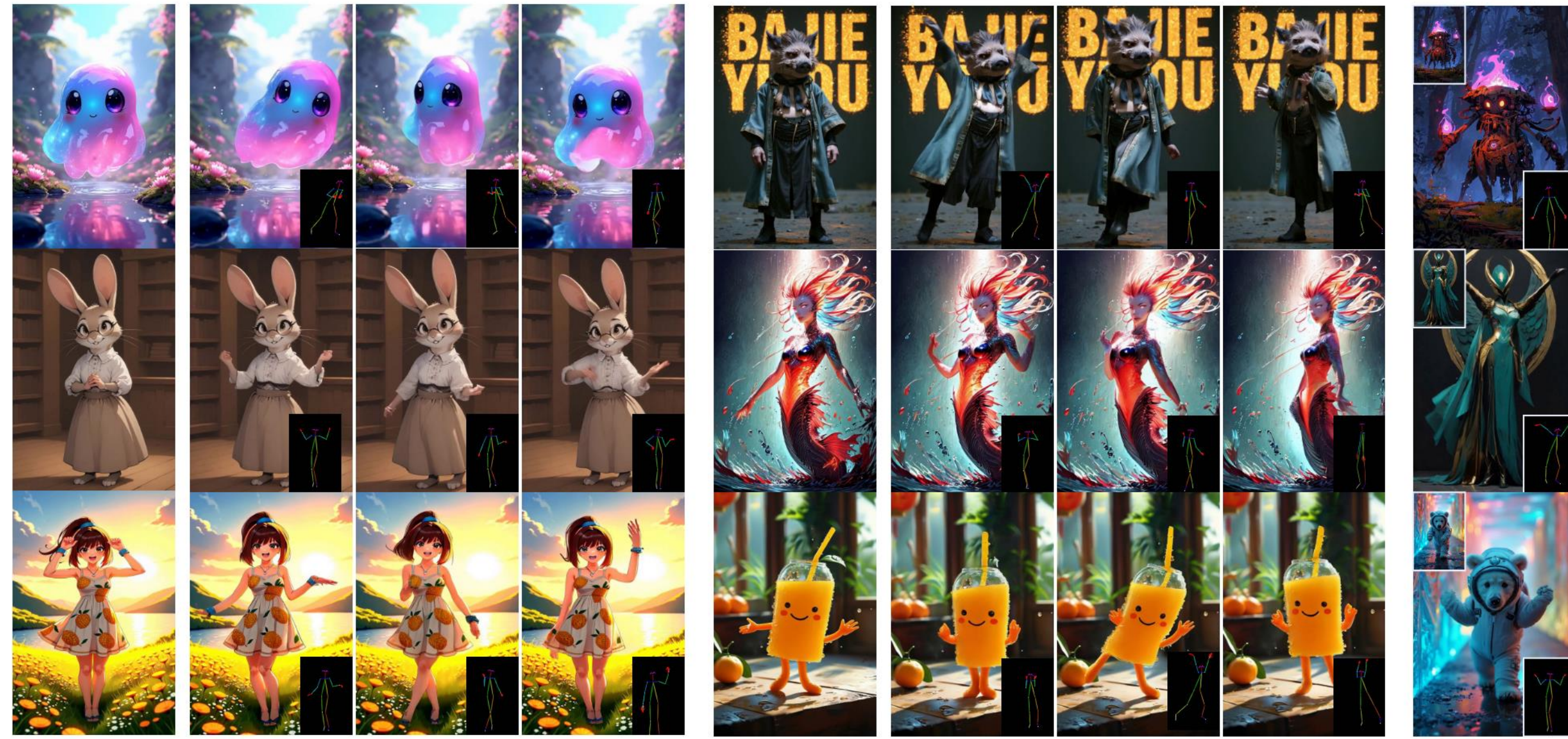
Animate-X Page

ICLR

Introduction:

➤ **Key issue to address**: Given a reference image and target pose sequence, **Animate-X** synthesizes animated videos which extends beyond human to **anthropomorphic characters** with various body structures, e.g., without limbs, from games, animations, and posters

➤ **Motivation:**

➤ Insufficient modeling of motion of previous methods

➤ Simple 2D pose skeletons lack of **image-level details**

➤ Self-driven reconstruction **strategy** ignoring shape differences

➤ Animate-X **enhances motion representation** in Implicit and explicit way

➤ **Contribution:**

➤ We present **Animate-X**, which facilitate pose-guided video generation with high generalizability, particularly for attractive **anthropomorphic characters**.

➤ The rethinking about the motion inspire us to propose Pose Indicator, which extracts motion representation suitable for anthropomorphic characters in both **implicit and explicit manner**, enhancing the **robustness** of Animate-X.

➤ we present a new **A²Bench**, specifically for evaluating performance on anthropomorphic characters. Extensive experiments demonstrate that our Animate-X **surpasses** the competing methods in both quantitative and qualitative evaluation.

Methodology:

➤ **Animate-X - Overall Framework:**
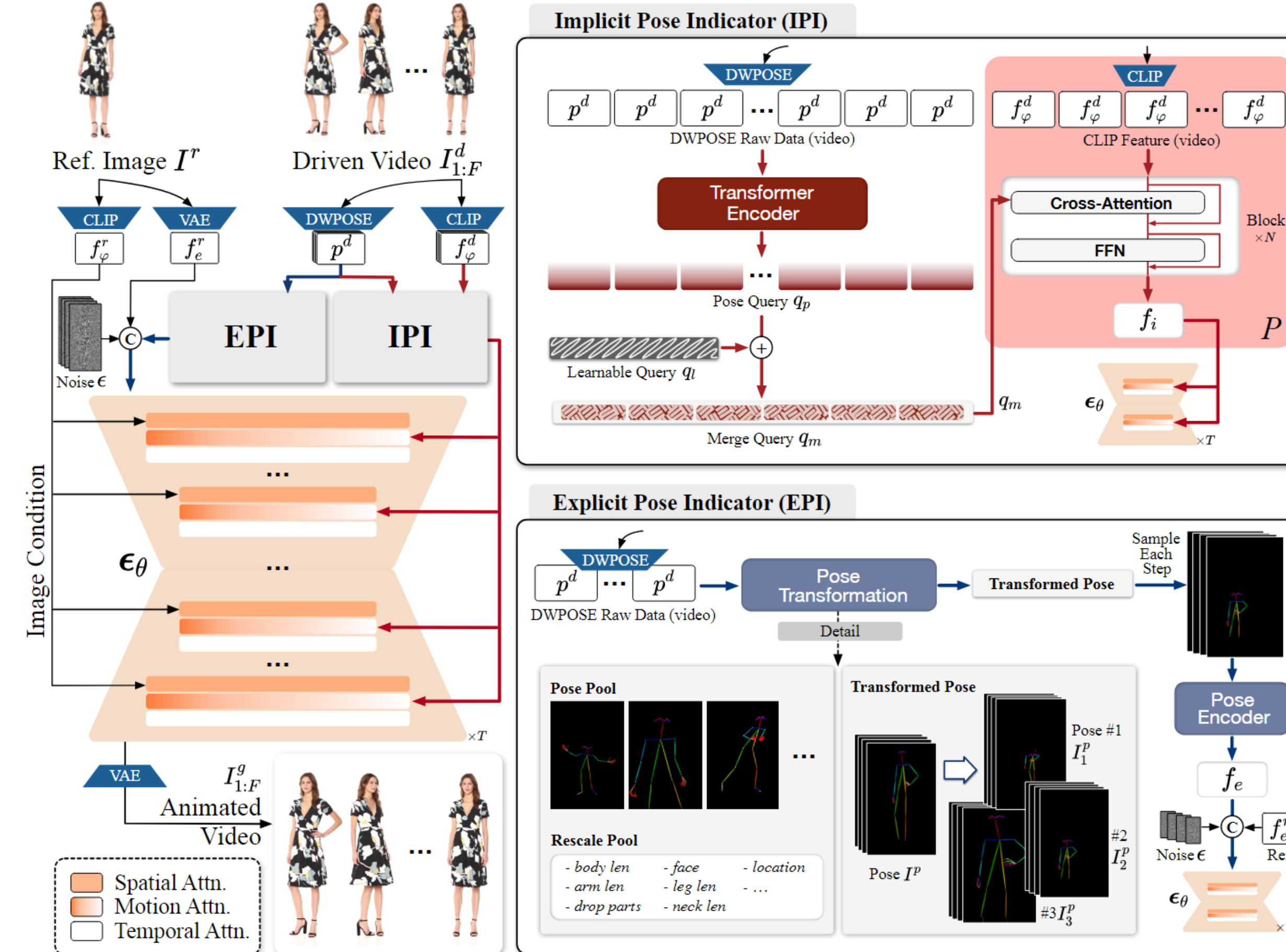
➤ **Animate-X Framework :**

➤ **Key component**: Pose Indicator, which consists of **Implicit Pose Indicator (IPI)** and **Explicit Pose Indicator (EPI)**

➤ **Implicit Pose Indicator:**

➤ Resort to the **CLIP image feature**

➤ Lightweight extractor extracts **motion patterns** helpful to animation generation

➤ **Explicit Pose Indicator:**

➤ **Pose Realignment → Pose Pool**

➤ **Pose Rescale → Rescale Pool**

➤ **Simulating misalignments** between reference image and pose images during training

A²Bench:

Experiment Results:

➤ **Qualitative comparison with SOTA emotional talking face methods**

Ref. Image | Ref. Pose | LIA [ICLR22] | Moore-AA [CVPR24] | MusePose [ArXiv24] | Unianimate [ArXiv24] | MimicMotion [ArXiv24] | ControlNeXt [ArXiv24] | Ours