

SMPL介绍

— . .



01

什么是SMPL?

— ...

SMPL定义

SMPL: A Skinned Multi-Person Linear Model

Matthew Loper^{*1,2} Naureen Mahmood^{†1} Javier Romero^{†1} Gerard Pons-Moll^{†1} Michael J. Black^{†1}

¹Max Planck Institute for Intelligent Systems, Tübingen, Germany

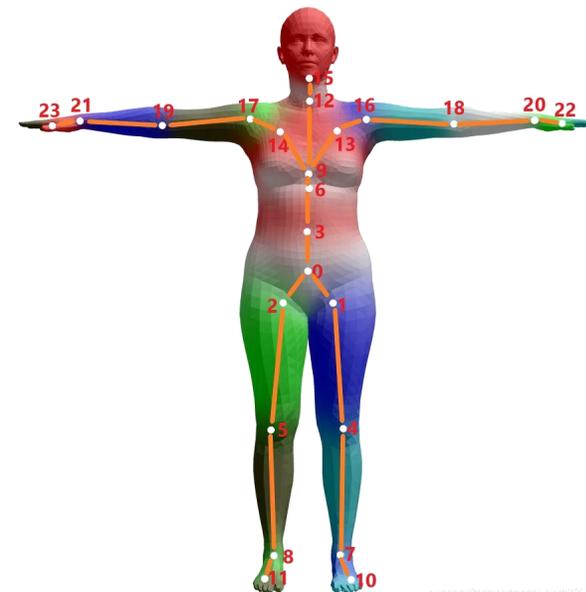
²Industrial Light and Magic, San Francisco, CA

SMPL是一种赤身的，基于顶点的人体三维模型，能够精确地表示人体的不同形状和姿态。

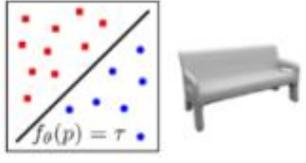
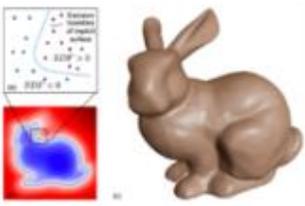
Skinned: 表示这个模型不仅仅是骨架点，是有蒙皮的，其蒙皮通过多边形网格表示。

Multi-person: 模型可以表示不同的人。

Linear: 表示人体的不同姿态或者不同形状变换是一个线性的过程



三维模型表示

表示方法	显式表示			隐式表示	
	Voxel	Point Cloud	Polygon Mesh	Occupancy Function	Signed Distance Function
表示图像					
表示原理	体素用规则的立方体表示3D物体，体素是数据在三维空间中的最小分割单位，类似于2D图像中的像素	点云将多面体表示为三维空间中点的集合，一般用激光雷达或深度相机扫描后得到点云数据	多边形网格将多面体表示为顶点与面片的集合，包含了物体表面的拓扑信息	occupancy function 将物体表示为一个占有函数，即空间中每个点是否在表面上	SDF 将物体表示为符号距离函数，即空间中每个点距离表面的距离
优缺点	<ul style="list-style-type: none"> + 规则表示，容易送入网络学习 + 可以处理任意拓扑结构 - 随着分辨率增加，内存呈立方级增长 - 物体表示不够精细 - 纹理不友好 	<ul style="list-style-type: none"> + 容易获取 + 可以处理任意拓扑结构 - 缺少点与点之间连接关系 - 物体表示不够精细 - 纹理不友好 	<ul style="list-style-type: none"> + 高质量描述3D几何结构 + 内存占有较少 + 纹理友好 - 不同物体类别需要不同的 mesh 模版 - 网络较难学习 	<ul style="list-style-type: none"> + 可以精细建模细节，理论上分辨率无穷 + 内存占有少 + 网络较易学习 - 需后处理得到显式几何结构 	<ul style="list-style-type: none"> + 可以精细建模细节，理论上分辨率无穷 + 内存占有少 + 网络较易学习 - 需后处理得到显式几何结构

最后输出的是一个什么样的模型

SMPL (Skinned Multi-Person Linear Model) 的输出就是一个三维网格模型 (3Dmesh)，这个模型中包含了顶点 (vertices)、边 (edges)、以及组成网格的三角形面 (triangles)。顶点是每个三角形顶点的坐标位置，边是相邻两个顶点间连线，面是由三个相邻的点构成的三角形面片。



模型输入输出概览

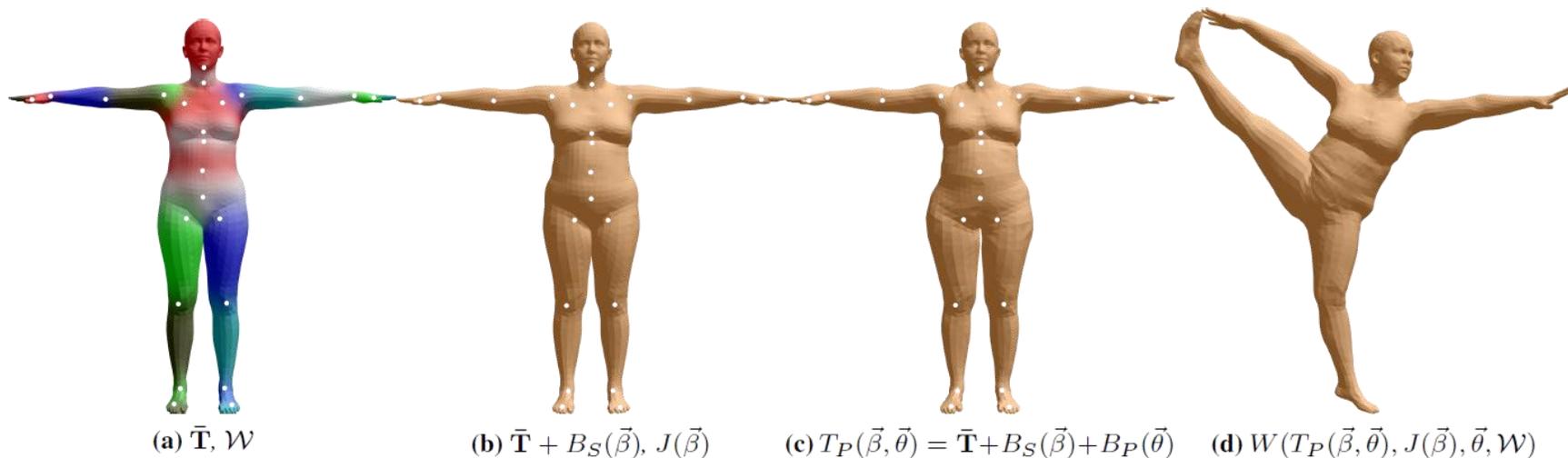


Figure 3: SMPL model. (a) Template mesh with blend weights indicated by color and joints shown in white. (b) With identity-driven blendshape contribution only; vertex and joint locations are linear in shape vector $\vec{\beta}$. (c) With the addition of of pose blend shapes in preparation for the split pose; note the expansion of the hips. (d) Deformed vertices reposed by dual quaternion skinning for the split pose.

β 和 θ 是其中的输入参数，其中 β 代表是个人体高矮胖瘦、头身比例等10个参数， θ 是代表人体整体24个关节之间的相对角度，共72维。

- Pose: 24 个关节点, 24*3 维旋转向量
- Shape: 10 维向量, 控制人体高矮胖瘦
- Mesh: 6890 顶点, 13776 面

有一个三维的球体，其中包含很多顶点。首先，我们需要定义一组形状参数和姿态参数，用于描述该球体的形状和姿态。然后，我们需要将每个顶点表示为由多个关节控制的线性组合。然后，根据每个关节在局部坐标系下的位置和旋转角度，以及每个关节对应的权重，可以计算出每个顶点在全局坐标系下的位置。通过这些信息，我们就可以用SMPL模型来描述该球体上的每个顶点了。



02

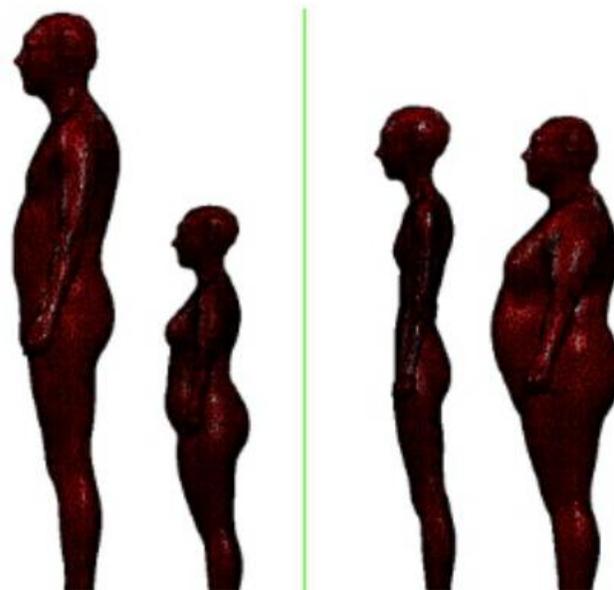
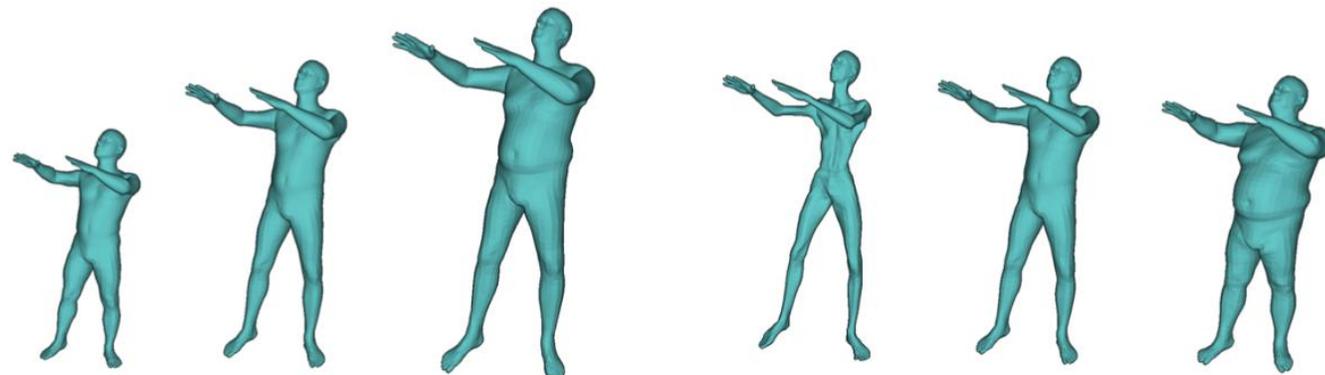
重要概念

— — —

Shape

shape: 一组形状参数有着10个维度的数值去描述一个人的形状，每一个维度的值都可以解释为人体形状的某个指标，比如高矮，胖瘦

shape (形状) 指的是人体的静态形状特征，如身高、体重、胸围等。在SMPL模型中，shape参数描述了一个人体的静态形状，如胖瘦、肌肉发达程度等。通过改变shape参数，可以得到不同形状的虚拟人体。



Shape参数及其影响

shape参数的变化会影响SMPL模型生成的人体顶点的位置和形状。例如，增加身高参数可能会导致生成的人体模型顶点向上延伸，使整个人体变得更高；增加体重参数可能会导致生成的人体模型顶点向外扩展，使整个人体变得更胖。

此外，shape参数的变化还会影响到SMPL模型输出的面片的形状和表面细节。例如，增加肢体比例参数可能会导致生成的人体模型的腿部或手臂变得更长或更短；增加体重参数可能会导致生成的人体模型的腹部或胸部变得更突出。

SMPL模型的shape参数是用来描述人体形状的系数，共有10个维度，分别是：

1. Body height: 身高，通过调整这个参数可以改变人体的整体高度。
2. Body weight: 体重，通过调整这个参数可以改变人体的整体体重。
3. Chest width: 胸部宽度，通过调整这个参数可以改变人体上半身的宽度。
4. Waist width: 腰部宽度，通过调整这个参数可以改变人体腰部的宽度。
5. Hip width: 臀部宽度，通过调整这个参数可以改变人体臀部的宽度。
6. Leg length: 腿长，通过调整这个参数可以改变人体腿部的长度。
7. Arm length: 臂长，通过调整这个参数可以改变人体手臂的长度。
8. Shoulder width: 肩宽，通过调整这个参数可以改变人体肩部的宽度。
9. Neck length: 颈长，通过调整这个参数可以改变人体颈部的长度。
10. Head size: 头部大小，通过调整这个参数可以改变人体头部的大小。

这些参数可以用来描述人体的形状特征，通过调整这些参数的值，可以得到不同形态的人体模型。需要注意的是，这些参数只能用来描述人体的形状，不能用来描述人体的姿态和动作。

Pose

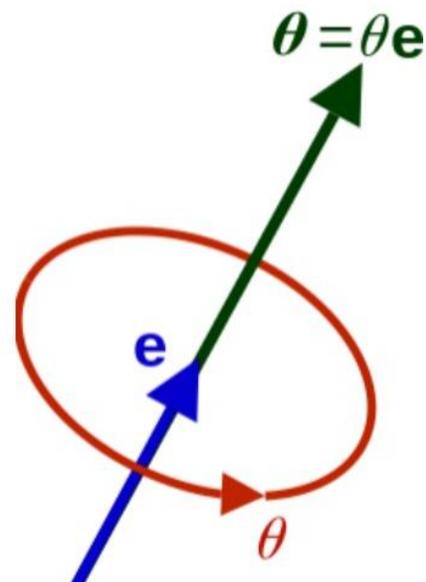
pose指的是人体在不同的动态状态下的姿势和动作，如行走、跑步、伸手等。在SMPL模型中，pose参数描述了一个人体的动态姿态，如膝盖弯曲的角度、手臂的伸直程度等。通过改变pose参数，可以得到不同姿态的虚拟人体。

一组姿态参数有着 24×3 维度的数字，去描述某个时刻人体的动作姿态，其中的24表示的是24个定义好的人体关节点，其中的3指的是该节点针对于其父级节点的旋转角度的轴角式表达。在轴角度旋转表示中，每个旋转都被编码为任意三维矢量



(c) $T_P(\vec{\beta}, \vec{\theta}) =$

$\mathcal{M}(\vec{\theta}, \mathcal{W})$



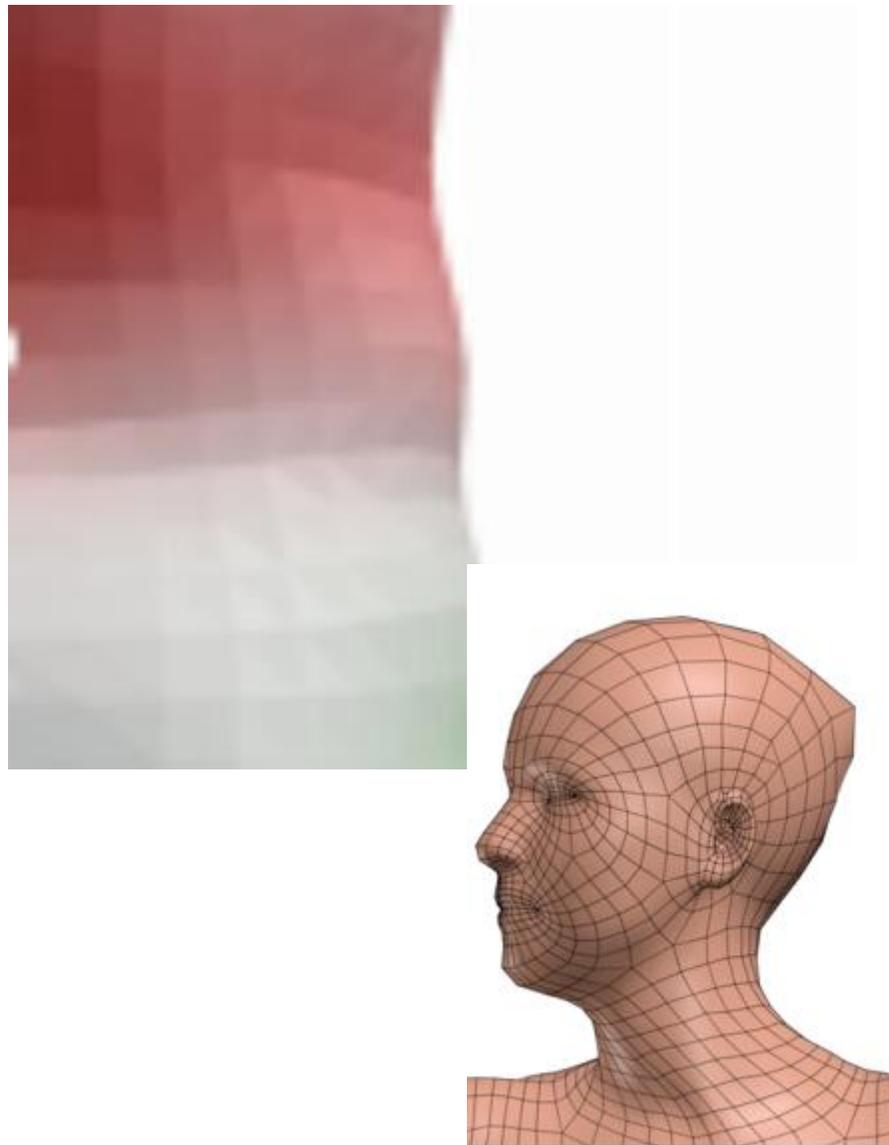
顶点

顶点:

如右图，SMPL的mesh模型可以看作由若干个小面片组成，每一个小面片的点就是SMPL的一个顶点。

SMPL模型中的顶点是模拟人体外形的一个3D网格模型中的位置点。顶点越多，模型越精细。

该模型将每个顶点表示为由多个关节控制的线性组合。每个关节都有一个权重。使用这些权重和每个关节在局部坐标系下的位置和旋转角度，可以计算出每个顶点在全局坐标系下的位置。



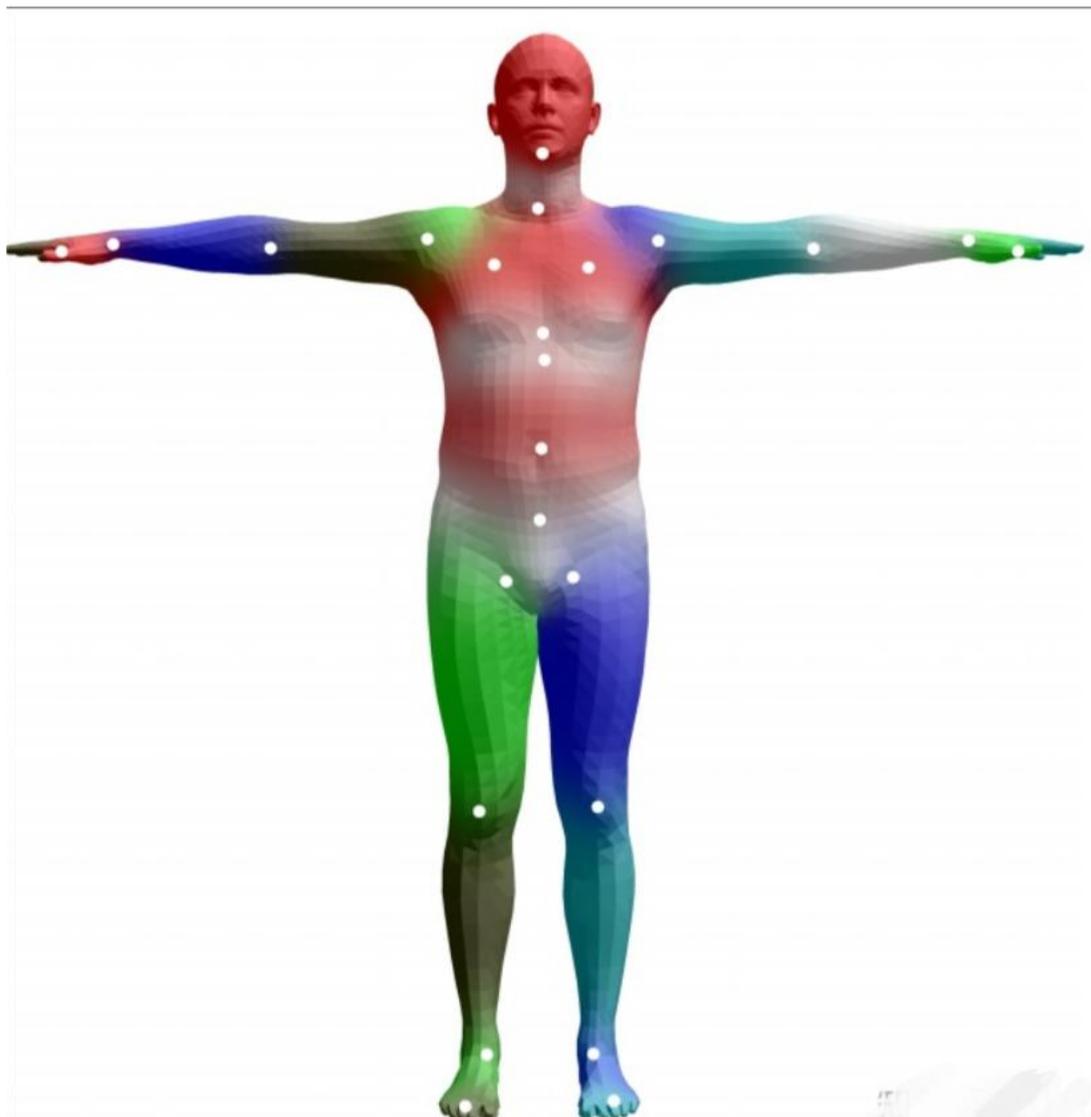
(b) SMPL parameterization

顶点与关节点

关节点：23个

顶点和关节点的关系：**变形时候，顶点由与之相关的关节点和对应的权重决定。**

关节和骨骼的关系：在SMPL模型中，人体的骨骼由24个关节 (joint) 组成。可以理解为关节是骨骼的端点。有的文章中，骨骼和关节互用，代表一个意思，即对顶点有影响的关节



Weight

SMPL的skinning weight是用来描述每个顶点对哪些关节起作用的参数。在SMPL模型中，**每个顶点都被分配到一个或多个关节，每个关节都有一个权重值，表示该关节对该顶点的影响程度。**这些权重值通常是通过**基于物理的方法**计算出来的，比如使用最近点法或加权最近点法等。

在计算SMPL模型的线性蒙皮变形时，需要使用这些权重值来计算每个顶点的最终位置。具体来说，对于一个给定的顶点，其最终位置可以通过将其与所有与之相关的关节的姿态矩阵相乘，得到一个相对于骨骼坐标系的位置，然后再将其乘以权重值，并将所有这些位置加起来，得到该顶点的最终位置。

Weight示例

现在有一个顶点，我们应该怎么计算出他在全局坐标系下的位置呢？

每个关节都有一个姿态矩阵和一个权重值，表示该关节对该顶点的影响程度。

假设左肩骨的权重值为0.3，右肩骨的权重值为0.5，脊柱骨的权重值为0.2。那么该顶点V的最终位置可以通过以下计算得到：

将该顶点V与左肩骨、右肩骨和脊柱骨的姿态矩阵相乘（姿态矩阵可以由轴角表达式变换而来），得到相对于骨骼坐标系的位置。

将上述位置分别乘以对应的权重值，得到三个加权位置。

将三个加权位置相加，得到该顶点V的最终位置。

$$\bar{\mathbf{t}}'_i = \sum_{k=1}^K w_{k,i} G'_k(\vec{\theta}, \mathbf{J}) \bar{\mathbf{t}}_i$$

↓ ↓ ↓
权重 姿态矩阵 顶点

$$\bar{\mathbf{t}}'_i = \sum_{k=1}^K w_{k,i} G'_k(\vec{\theta}, J(\vec{\beta})) (\bar{\mathbf{t}}_i + \mathbf{b}_{S,i}(\vec{\beta}) + \mathbf{b}_{P,i}(\vec{\theta})) \quad ($$

坐标系

在SMPL模型中，使用了两个坐标系：全局坐标系和局部坐标系。

全局坐标系

全局坐标系是SMPL模型最外层的坐标系，它是一个笛卡尔坐标系，原点位于人体的重心位置。在全局坐标系下，可以描述整个人体的姿态和位置（主要是关节、顶点的位置）。

局部坐标系

局部坐标系是SMPL模型中每个关节（Joint）所使用的坐标系。每个关节都有自己的局部坐标系，用于描述该关节对应的骨骼的姿态和位置。每个局部坐标系的原点位于该关节的位置，坐标系的方向和旋转轴也与该关节的骨骼相关。

需要注意的是，SMPL模型描述的是一个单人的人体模型，因此其坐标系只能用于描述单人的姿态和形状。如果需要处理多人的姿态估计，则需要使用其他方法和坐标系来描述。



03

建模过程

— · ·

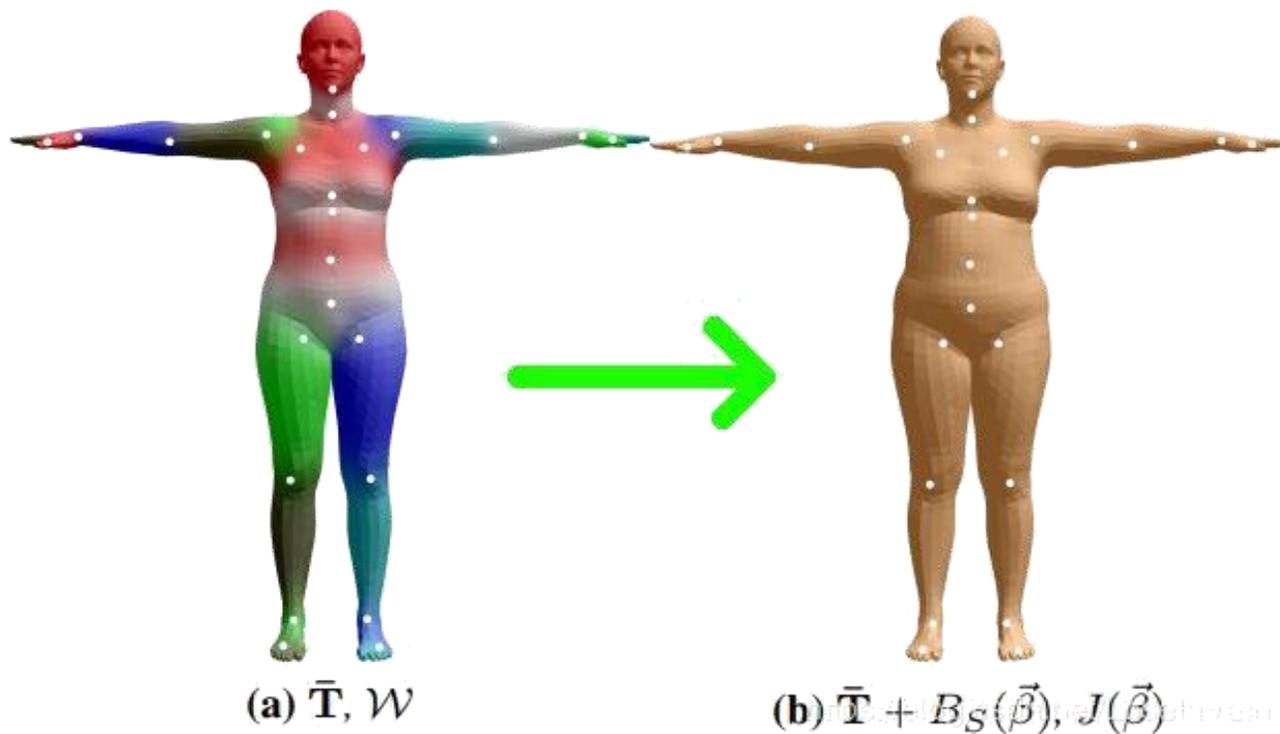
步骤一

- 基于形状的混合变形 (Shape Blend Shapes)

人体的网格顶点 (vertices) 会随着 shape 参数 β 变化而变化, 这个变化过程是在一个基模版 (或者称之为统计上的均值模版, mean template) 上线性叠加的。

此处得到的人体mesh的姿态称之为静默姿态(rest pose, 也可以称之为T-pose), 因为其并没有考虑姿态参数的影响。

变换公式如右, β 是输入的参数, S 是通过数据学习出来的, 它的维度为 (6890, 3, 10)。



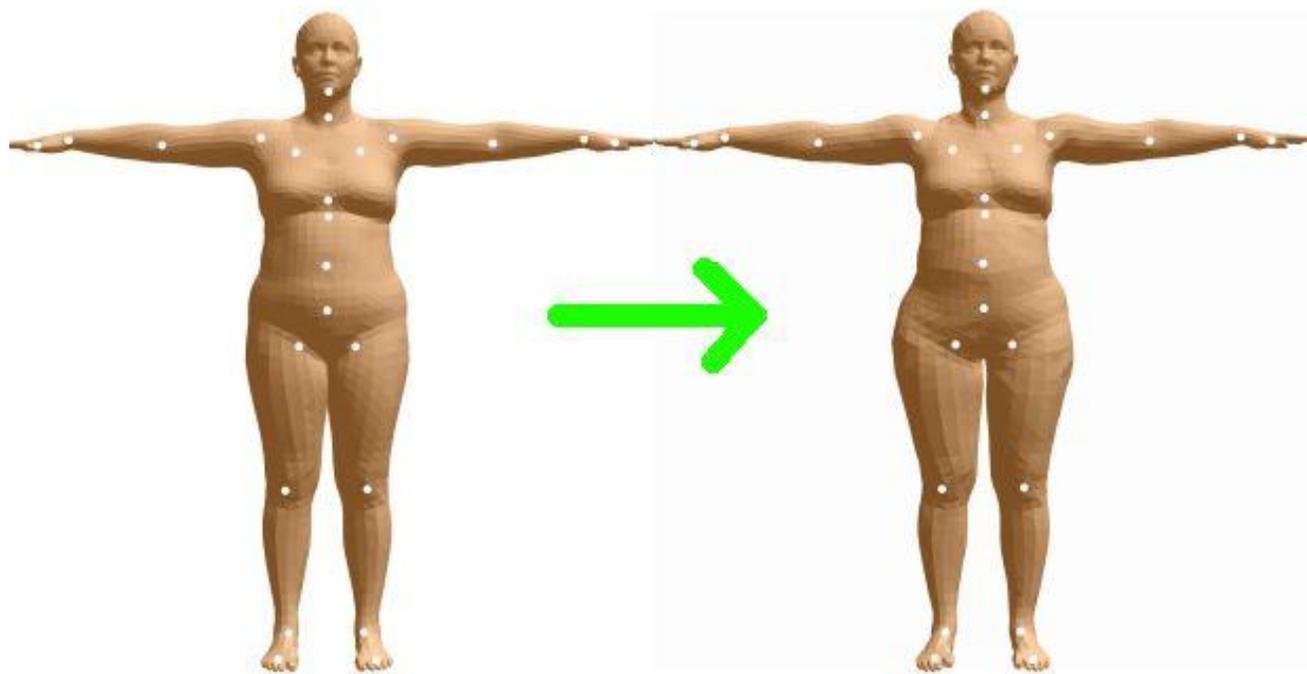
$$B_S(\vec{\beta}; \mathcal{S}) = \sum_{n=1}^{|\vec{\beta}|} \beta_n \mathbf{S}_n$$

步骤二

- 基于姿态的 blend shape (混合变形)

计算人体在不同 pose 参数 θ 下的影响, 根据指定的形状参数对人体 mesh 进行形状的指定后, 得到了一个具有特定胖瘦, 高矮的 mesh。

因为我们是计算相对 T-pose 状态下的线性叠加偏量, 所以人体的位姿应该也是要相对 T-pose 状态下进行变化 (括号内带 * 部分), 因此括号里减去了 T-pose 位姿的影响。每个 pose 参数都用旋转矩阵 R 表示, 所以是 $9K$ 。P 也是通过数据学习出来的, 它的维度为 $(6890, 3, 207)$, 其中 207 是因为 23×9 得到。



(b) $\bar{T} + B_S(\vec{\beta}), J(\vec{\beta})$

(c) $T_P(\vec{\beta}, \vec{\theta}) = \bar{T} + B_S(\vec{\beta}) + B_P(\vec{\theta})$

$$B_P(\vec{\theta}; P) = \sum_{n=1}^{9K} (R_n(\vec{\theta}) - R_n(\vec{\theta}^*)) P_n,$$

顶点偏差

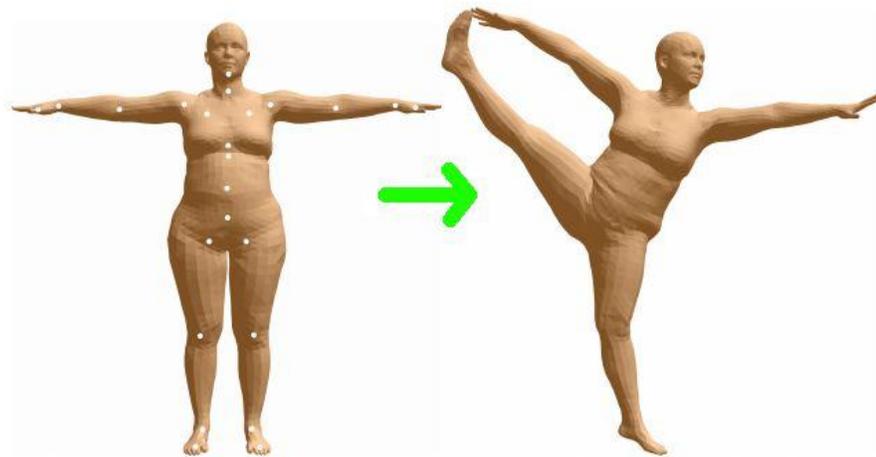
步骤三

- 蒙皮 (weight作用)

当人体关节运动时，由网格顶点组成的“皮肤”将会随着关节的运动而变化，这个过程称之为蒙皮。

具体来说，对于SMPL模型中的每个顶点，都会计算其与所有关节点之间的距离，根据距离的大小，为每个关节点分配一个权重。这些权重可以被理解为关节点对于某个顶点的影响程度。例如，如果一个顶点距离关节A比距离关节B更近，那么关节A的权重就会比关节B的权重更高。

然后，将这些权重与关节点的位置信息结合，就可以计算出每个顶点在不同姿态下的位置。这样，当模型被变换到不同的姿态时，每个顶点都会跟随其所绑定的关节点进行相应的变换，从而实现了蒙皮效果。



(c) $T_P(\vec{\beta}, \vec{\theta}) = \bar{\mathbf{T}} + B_S(\vec{\beta}) + B_P(\vec{\theta})$ (d) $W(T_P(\vec{\beta}, \vec{\theta}), J(\vec{\beta}), \vec{\theta}, W)_{\text{skin}}$

$$\bar{\mathbf{t}}'_i = \sum_{k=1}^K w_{k,i} G'_k(\vec{\theta}, J(\vec{\beta})) (\bar{\mathbf{t}}_i + \mathbf{b}_{S,i}(\vec{\beta}) + \mathbf{b}_{P,i}(\vec{\theta}))$$

转换后的坐标

$$\bar{\mathbf{t}}'_i = \sum_{k=1}^K w_{k,i} G'_k(\vec{\theta}, \mathbf{J}) \bar{\mathbf{t}}_i$$

与各个顶点加权

$$G'_k(\vec{\theta}, \mathbf{J}) = G_k(\vec{\theta}, \mathbf{J}) G_k(\vec{\theta}^*, \mathbf{J})^{-1}$$

$$G_k(\vec{\theta}, \mathbf{J}) = \prod_{j \in A(k)} \begin{bmatrix} \exp(\vec{\omega}_j) & \mathbf{j}_j \\ \mathbf{0} & 1 \end{bmatrix}$$

父关节的组合，迭代相乘



04

SMPL扩展

— — —

SMPL_X (人体表面细节改进)

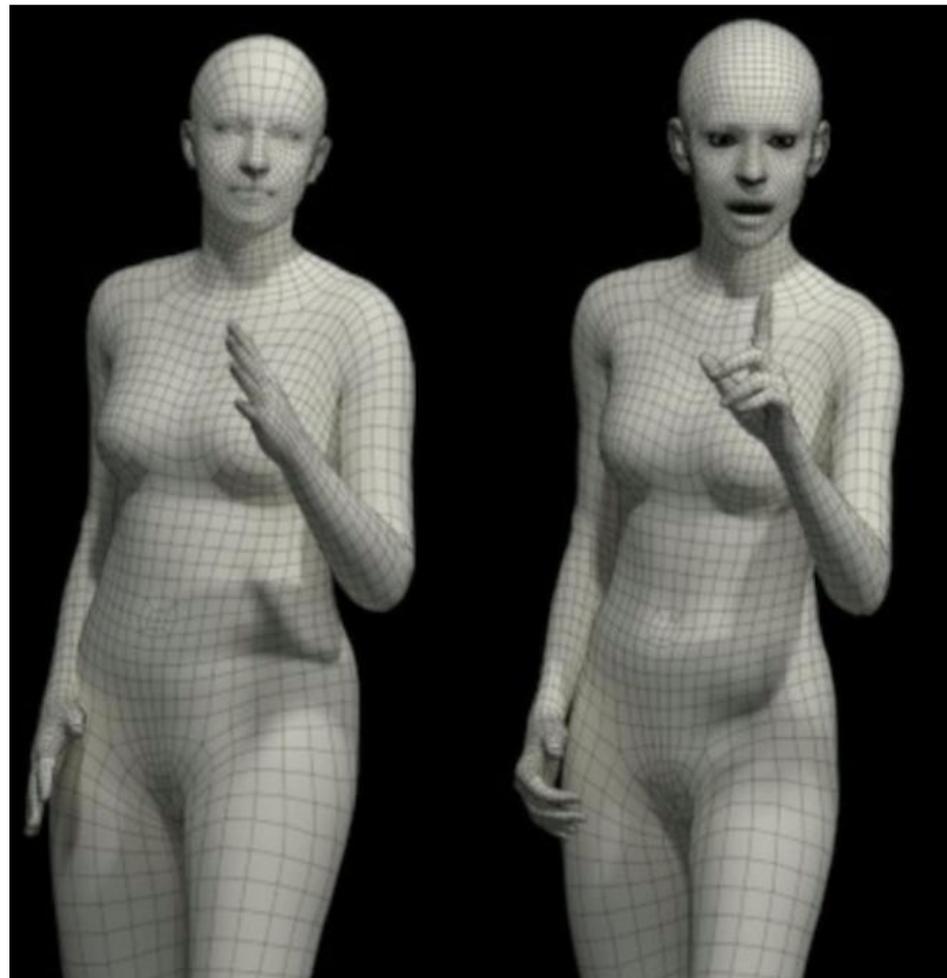
SMPLx具有更高分辨率和更精细的拓扑结构。与原始 SMPL 模型相比，SMPL-X 包括 10 个额外的姿势参数、10 个额外的形状参数和 23 个额外的关节。

模型精度：SMPL-X引入了更多的形状参数和姿态参数，以更好地捕捉人体的细微变化和多样性。

特征表达：SMPL-X在表示人体形状和姿态的参数化方法上进行了改进。它采用了全局和局部的形状特征表示方法，可以更准确地捕捉人体的细节和形状变化。

多样性建模：SMPL-X引入了多样性建模的概念，通过引入随机变量，可以生成不同个体之间的差异和多样性。

数据集和训练：SMPL-X使用了更大规模和多样化的数据集进行训练，以提高模型的泛化能力和适应性。此外，SMPL-X还引入了更多的约束和正则化项，以提高模型的稳定性和鲁棒性。



SMPL_h (手部姿势)

对 SMPL 的主要修改是**为手添加了详细的网格模型，并重新配置了模型的骨骼层次结构以包括 15 个手指和拇指的运动链**。此外，SMPL+H 包括一组预定义的抓握，有助于模拟手-对象交互。SMPL+H 模型提供了更真实、更详细的人体运动表现，尤其是复杂的手部运动。

手部关节：SMPL+H 将 3D 手部模型添加到 SMPL 身体模型中，从而实现逼真且细致的手部姿势。

关节数量增加：SMPL+H 将模型中的关节数量从 23 个增加到 46 个，从而使动作捕捉更加准确，动画更加灵活。

改进的手-物体交互：SMPL+H 模型包括一组预定义的抓握，这使其能够更准确地模拟手和物体之间的交互。



感谢观看





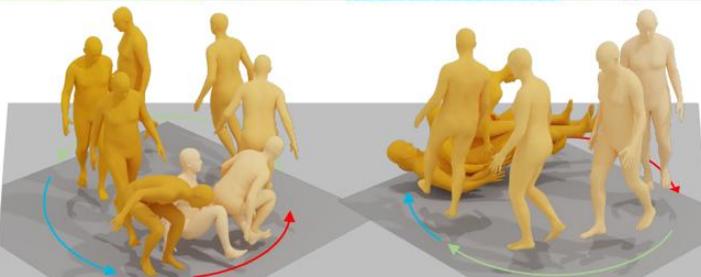
武漢大學

WUHAN UNIVERSITY

Text2motion

任务汇报

the person rises from a laying position and walks in a clockwise circle, and then lays back down the ground.



a person is careful while walking around the obstacles.



dribble with the basketball

run in a circle



1. The person is leaving at someone with his left hand.
2. A person shakes an item with his left hand.
3. A person waves his left hand repeatedly above his head.



1. A person doing jumping jacks and then running on the spot.
2. A person is doing jumping jacks, then starts jogging in place.
3. A person does four jumping jacks then three front lunges.

任务汇报

细节

数据

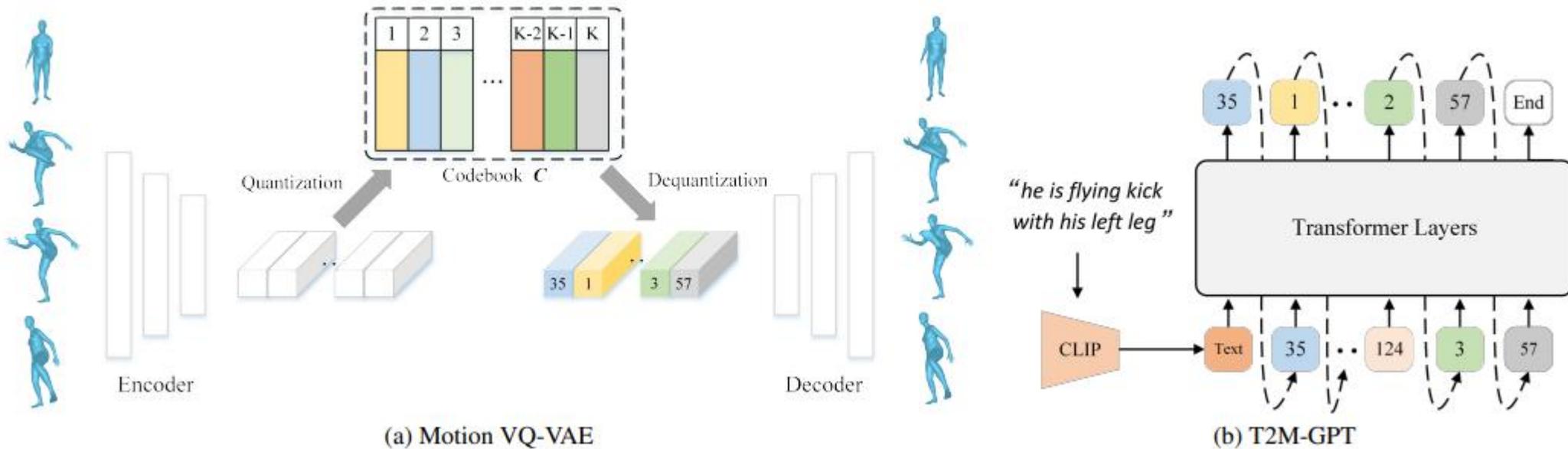
新输入

交互

Text2motion

- Generating diverse and natural 3d human motions from text. In CVPR2022.
- Actionconditioned 3d human motion synthesis with transformer VAE. In ICCV2021
- TEMOS generating diverse human motions from textual descriptions. In ECCV 2022
- Human motion diffusion model. In ICLR, 2023
- TM2T:stochastic and tokenized modeling for the reciprocal generation of 3d human motions and texts. In ECCV 2022.
- Motiondiffuse: Text-driven human motion generation with diffusion model. arXiv preprint arXiv2022
- Executing your commands via motion diffusion in latent space. In CVPR2023.
- **T2m-gpt**: Generating human motion from textual descriptions with discrete representations. In CVPR, 2023
- Motiongpt: Finetuned llms are general-purpose motion generators. arXiv preprint arXiv2023

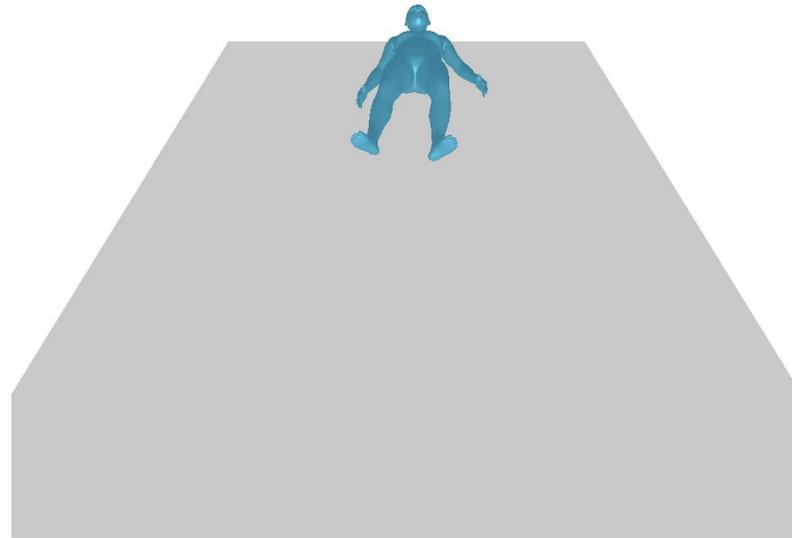
Text2motion



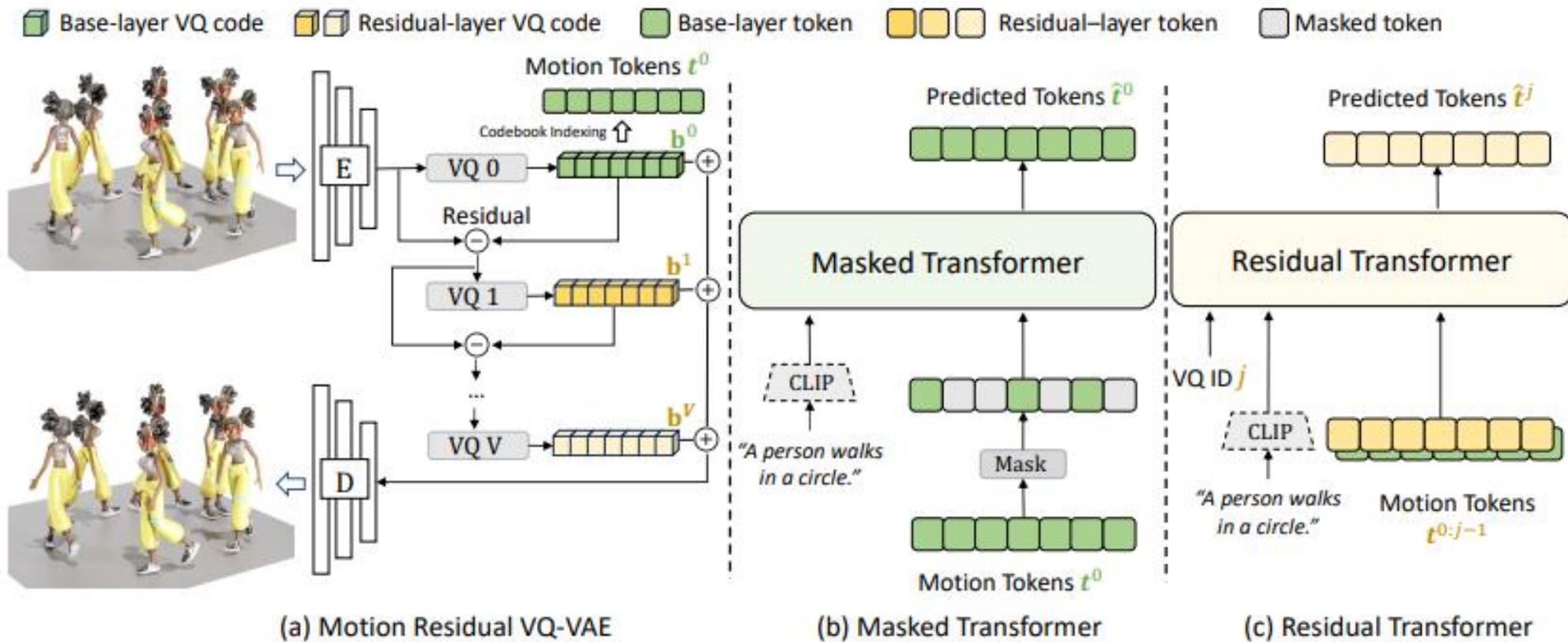
The naive training of VQ-VAE [68] suffers from code collapse. One effective solution is to leverage two standard recipes during the training: EMA and Code Reset.
The end token

Text2motion

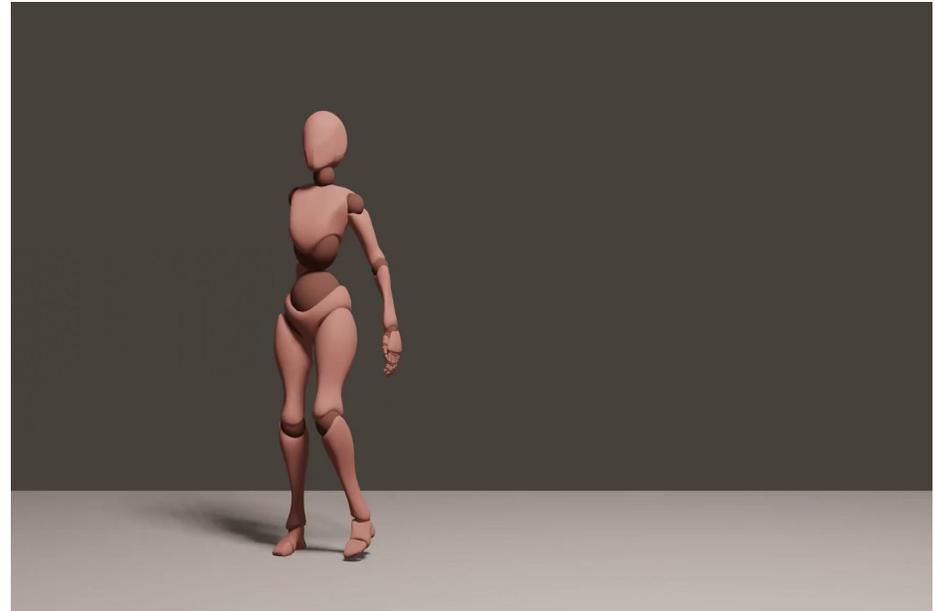
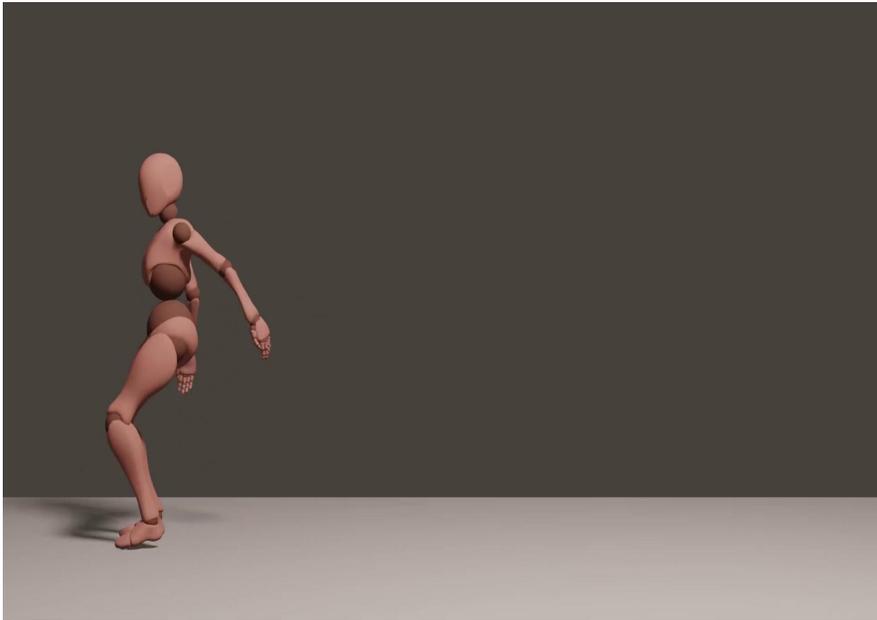
Text Prompting: A man rises from the ground, walks in a circle and sits back down on the ground.



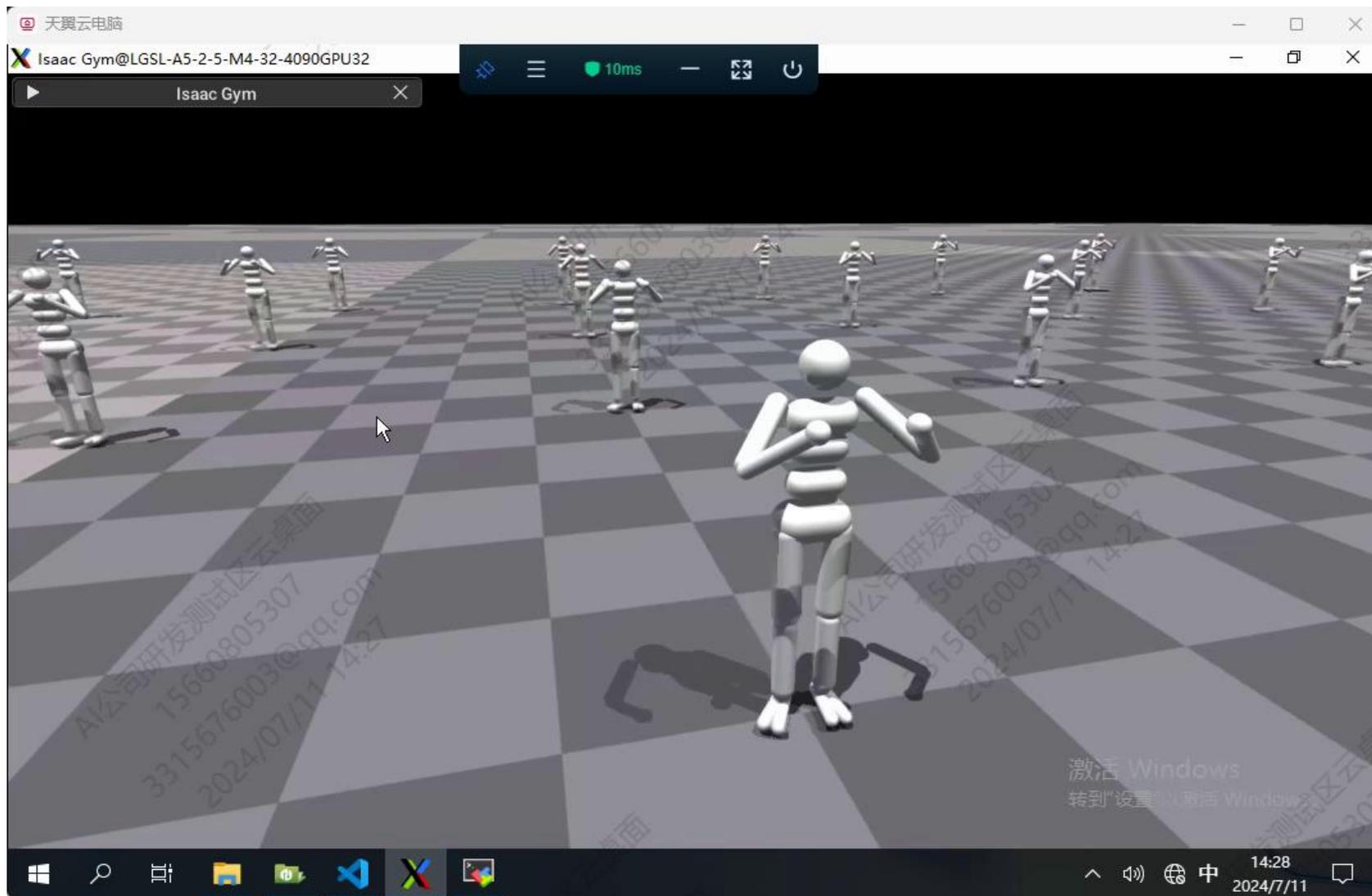
Text2motion



交互



Isaac Gym

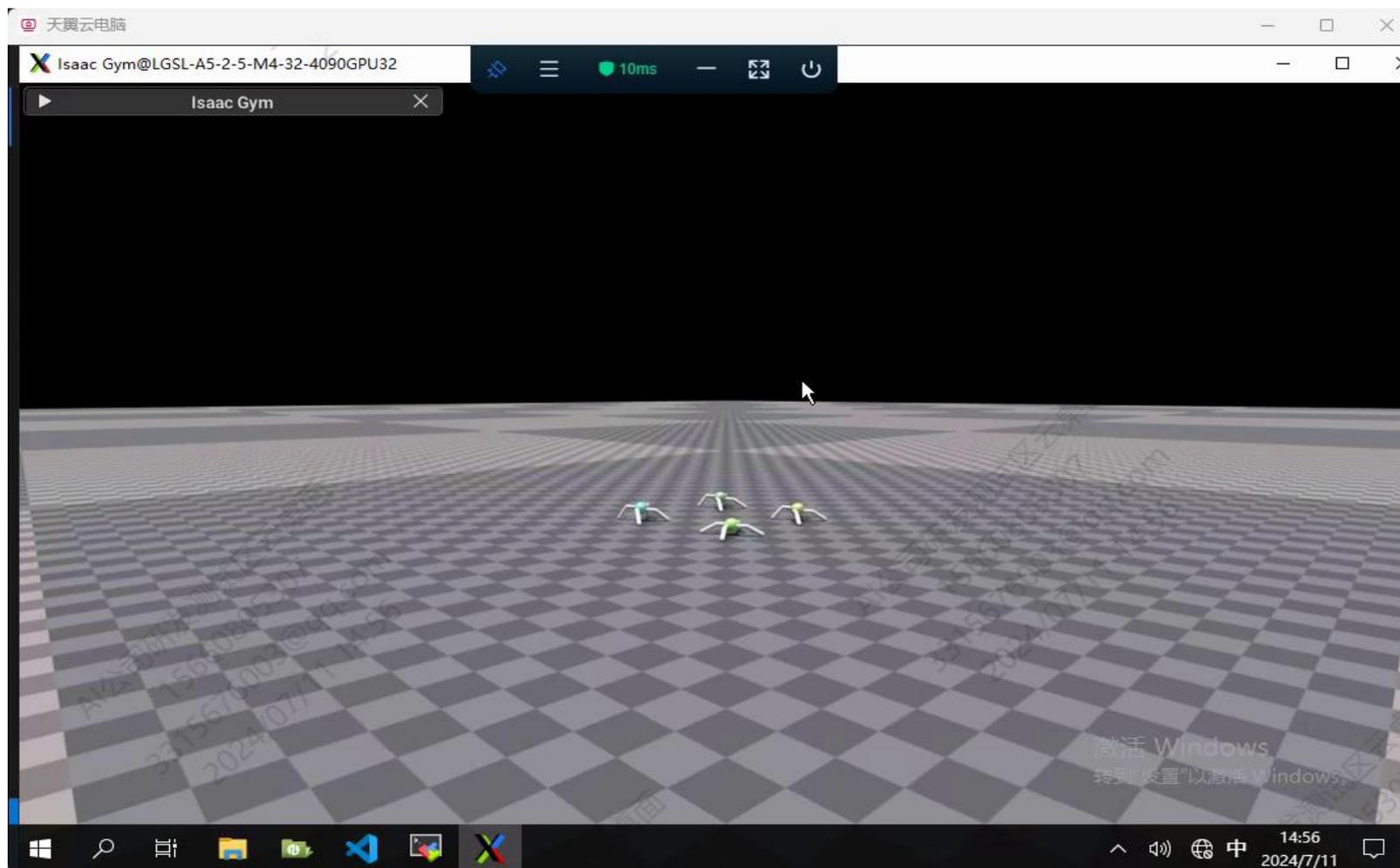


joint_monkey.py:

演示如何使用IsaacGym控制和模拟不同资产（机器人、物体等）的关节动画，包括如何初始化、更新和控制每个关节的运动范围和速度，同时通过视觉化工具（如绘制关节轴线）来辅助调试和理解。

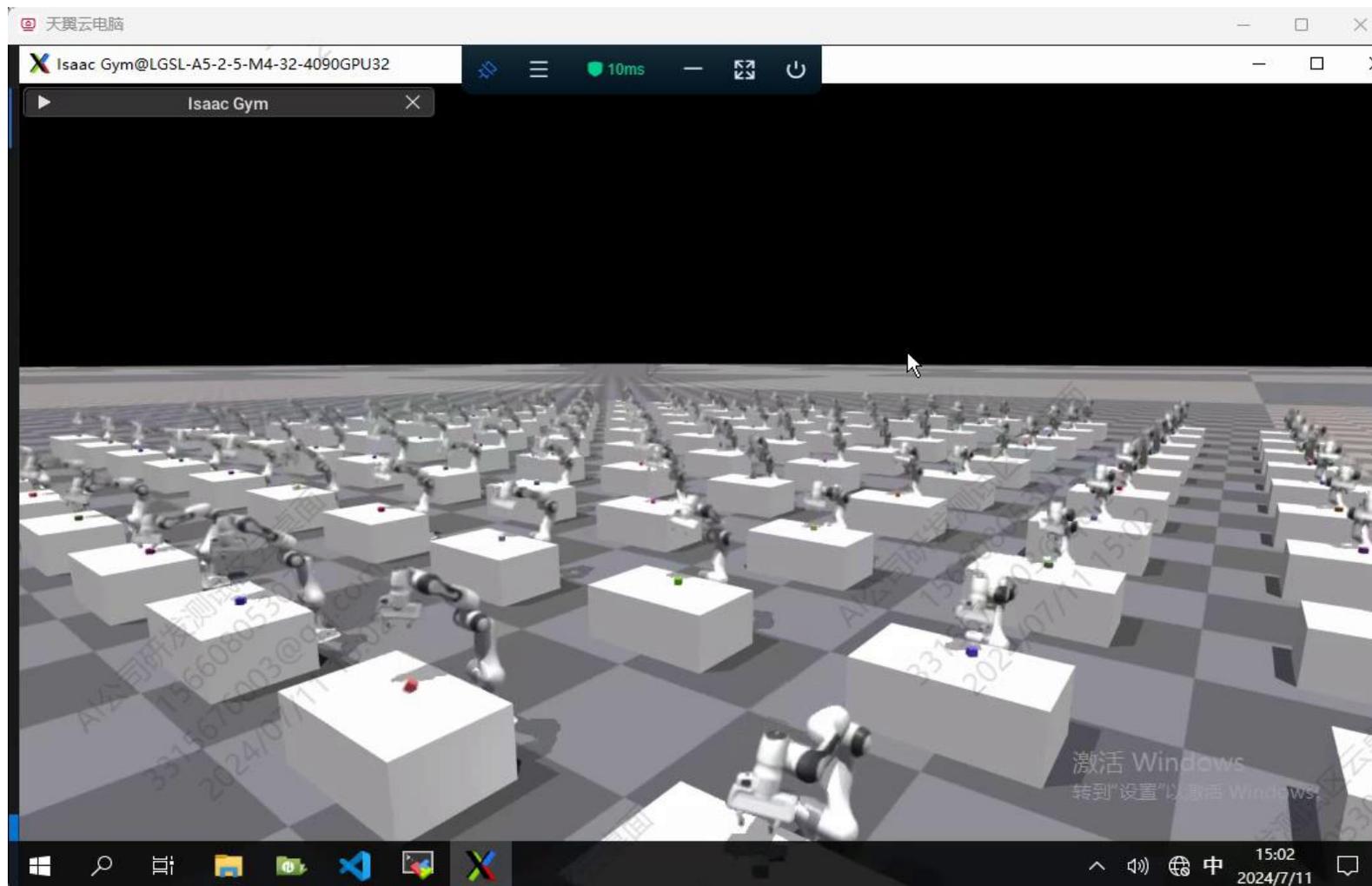
Isaac Gym

加入重力



Isaac Gym

跟环境的交互



交互

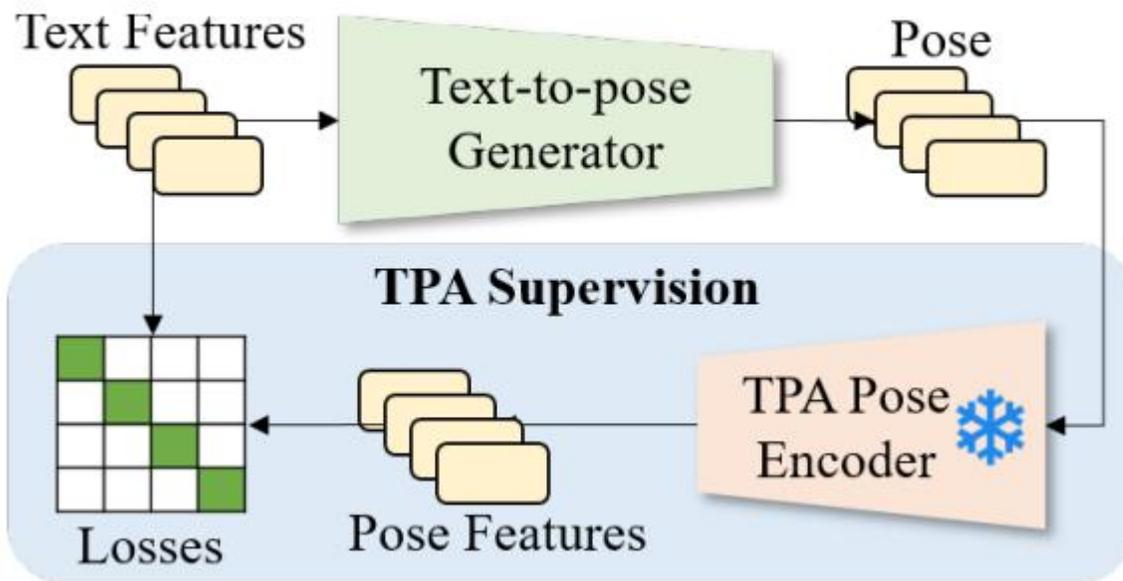
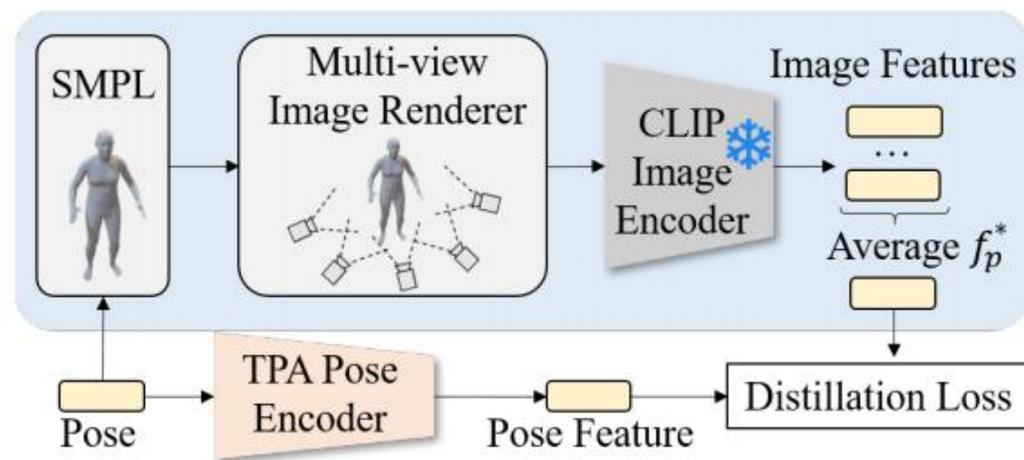
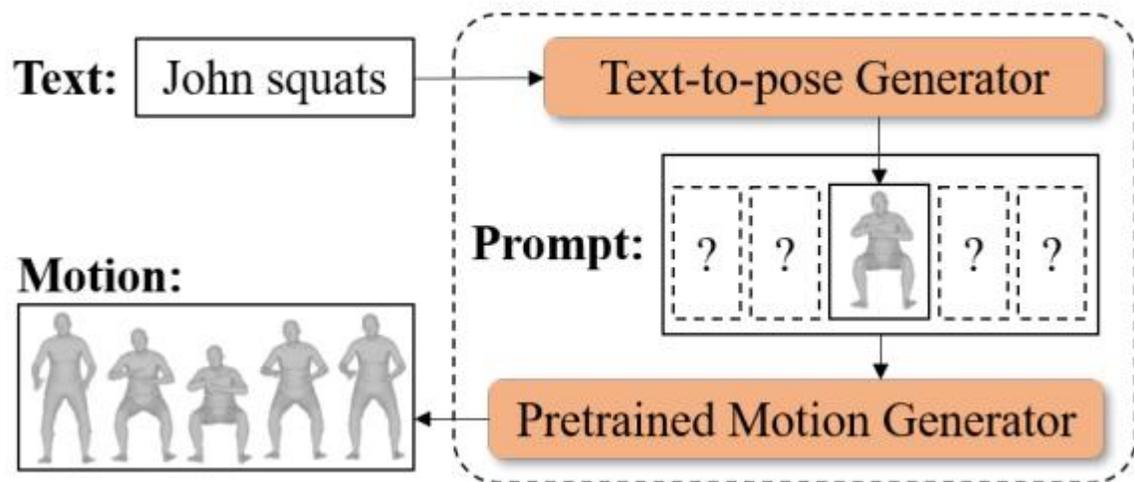


交互



Text Prompt	[Zhao et al. '22]	[Hassan et al. '19]	Ours (SV Only)	Ours
"sitting and playing piano, hands on piano keys, feet flat on the ground"				
"sitting and paddling a wooden boat, spreading arms"				
"walking on a bridge"				
"riding a recumbent bike, sitting"				

OOHMG



THANK YOU
