



(12) 发明专利申请

(10) 申请公布号 CN 114463511 A

(43) 申请公布日 2022. 05. 10

(21) 申请号 202111543921.9 *G06V 10/82* (2022.01)

(22) 申请日 2021.12.16 *G06K 9/62* (2022.01)

(71) 申请人 浙江工业大学 *G06N 3/04* (2006.01)
地址 310014 浙江省杭州市下城区潮王路
18号 *G06N 3/08* (2006.01)

(72) 发明人 刘盛 邹思宇 姚璐 李超楠
陈胜勇

(74) 专利代理机构 杭州求是专利事务所有限公
司 33200
专利代理师 忻明年

(51) Int. Cl.
G06T 17/20 (2006.01)
G06T 5/00 (2006.01)
G06V 10/46 (2022.01)
G06V 10/80 (2022.01)

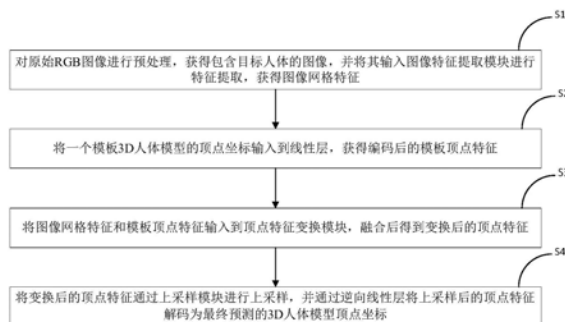
权利要求书2页 说明书6页 附图3页

(54) 发明名称

一种基于Transformer解码器的3D人体模型重建方法

(57) 摘要

本发明公开了一种基于Transformer解码器的3D人体模型重建方法,对原始RGB图像进行预处理,将其输入卷积神经网络进行特征提取,获得图像网格特征;将模板3D人体模型的顶点坐标输入一个线性层,获得编码后的模板顶点特征,将图像网格特征融入当前状态的顶点特征,并最终形成变换后的顶点特征,将变换后的顶点特征通过上采样层进行上采样,并通过逆向线性层将顶点特征解码为最终预测的3D人体模型顶点坐标。本发明更加充分地利用了图像网格特征中的局部信息,同时对顶点特征变换过程施加严格的语义一致性约束,帮助网络忽略不必要的语义转换,聚焦于预测更精准的人体模型。



1. 一种基于Transformer解码器的3D人体模型重建方法,其特征在于,所述基于Transformer解码器的3D人体模型重建方法,包括:

对原始RGB图像进行预处理,获得包含目标人体的图像,并将其输入图像特征提取模块进行特征提取,获得图像网格特征;

将一个模板3D人体模型的顶点坐标输入到线性层,获得编码后的模板顶点特征;

将图像网格特征和模板顶点特征输入到顶点特征变换模块,融合后得到变换后的顶点特征;

将变换后的顶点特征通过上采样模块进行上采样,并通过逆向线性层将上采样后的顶点特征解码为最终预测的3D人体模型顶点坐标。

2. 根据权利要求1所述的基于Transformer解码器的3D人体模型重建方法,其特征在于,所述图像特征提取模块使用ResNet网络。

3. 根据权利要求1所述的基于Transformer解码器的3D人体模型重建方法,其特征在于,所述模板3D人体模型使用无性别的、标准姿态的SMPL人体模型。

4. 根据权利要求3所述的基于Transformer解码器的3D人体模型重建方法,其特征在于,所述模板3D人体模型是经过网格下采样算法处理后的低分辨率的模板3D人体模型。

5. 根据权利要求1所述的基于Transformer解码器的3D人体模型重建方法,其特征在于,所述顶点特征变换模块包括预设数量的Transformer解码器,首层Transformer解码器的输入为图像网格特征和模板顶点特征,其他Transformer解码器的输入为图像网格特征和上一层Transformer解码器输出的顶点特征。

6. 根据权利要求5所述的基于Transformer解码器的3D人体模型重建方法,其特征在于,所述Transformer解码器包括依次连接的自注意力层、交叉注意力层与前馈网络层,所述自注意力层、交叉注意力层与前馈网络层前还分别设置有各自的归一化层,所述自注意力层、交叉注意力层与前馈网络层每一层的输入都与本层的输出进行累加后输入到下一层;所述自注意力层的输入是模板顶点特征或上一层Transformer解码器输出的顶点特征,所述交叉注意力层的输入还包括图像网格特征。

7. 根据权利要求1所述的基于Transformer解码器的3D人体模型重建方法,其特征在于,所述上采样层由两个相同的MLP网络组成,每个MLP由两个线性层与其间的非线性激活层组成,分阶段地将低分辨率的顶点特征转化为高分辨率的顶点特征。

8. 根据权利要求1所述的基于Transformer解码器的3D人体模型重建方法,其特征在于,所述线性层和逆向线性层严格互逆,所述线性层表示为:

$$z_0 = V_{3D}^T W + b;$$

其中, V_{3D}^T 表示模板3D人体模型的顶点坐标, z_0 表示模板顶点特征, W 和 b 分别表示线性层的权重矩阵与偏置向量;

所述逆向线性层表示为:

$$V_{3D} = (z_v - b) W^{-1};$$

其中, V_{3D} 表示逆向线性层输出的3D人体模型顶点坐标, z_v 表示顶点特征变换模块输出的顶点特征, W^{-1} 表示 W 的广义右逆矩阵, $W^{-1} = W^T (WW^T)^{-1}$ 。

9. 根据权利要求1所述的基于Transformer解码器的3D人体模型重建方法,其特征在

于,所述图像特征提取模块、线性层、顶点特征变换模块、上采样模块和逆向线性层构成网络模型,所述网络模型在训练时,采用的损失函数如下:

$$\mathcal{L} = \lambda_1 \mathcal{L}_V + \lambda_2 \mathcal{L}_{J,3D} + \lambda_3 \mathcal{L}_{J,2D}$$

其中, \mathcal{L}_V 表示最终预测的3D人体模型顶点坐标损失, $\mathcal{L}_{J,3D}$ 表示预测的3D人体关键点坐标损失, $\mathcal{L}_{J,2D}$ 表示预测的2D人体关键点坐标损失;所述预测的3D人体关键点坐标和预测的2D人体关键点坐标由最终预测的3D人体模型顶点坐标计算得到, λ_1 、 λ_2 、 λ_3 为超参数。

一种基于Transformer解码器的3D人体模型重建方法

技术领域

[0001] 本申请涉及计算机视觉领域,尤其涉及一种基于Transformer解码器的3D人体模型重建方法。

背景技术

[0002] 3D人体模型重建(Human Mesh Reconstruction,HMR)旨在通过视觉图像信号预测目标人体的3D人体网格模型,是计算机视觉的一个重要的任务。人体模型重建为许多下游任务提供目标人体丰富的姿态、形状等信息,如人体动作识别和虚拟试衣等,因此近年来受到越来越多的研究者的关注。同时,次领域中有许多挑战仍待解决,如人体关节的高度动态性、2D图像到3D预测的多义性、遮挡等。

[0003] 近年来,基于RGB图像的3D人体模型重建方法可主要分为两类:基于模型的(Model-based)与不基于模型的(Model-free)。基于模型的方法将例如SMPL的3D人体模型的参数视为预测的目标,而将预测参数输入SMPL获得最终的3D人体模型。其中,有的方法将模型参数从初始状态迭代式地优化至与2D视觉信息吻合,有的通过神经网络直接回归以获得模型参数。但由于2D图像与模型参数的表征差异过于巨大,为了获得精准的结果,这类方法一般需要大量的优化时间或训练数据。

[0004] 虽然基于模型与迭代优化的方法过于耗时,但这些方法离线地为许多没有3D标注的数据集创造了大量的3D人体模型伪标签。近期,在这些带有伪标签的数据集的帮助下,许多方法尝试不基于模型的方法来重建3D人体模型。这些方法固定了3D人体模型网格拓扑结构,并通过神经网络直接预测网格顶点坐标。这些仿佛普遍采用卷积神经网络、图卷积神经网络等局部建模能力强的网络结构,但无法充分表达人体长距离部件或顶点间的关系。

[0005] Transformer是一种基于注意力机制的网络结构,可以软式地对两个序列间的关系进行建模。由于其强大的非局部交互建模能力,Transformer在近期许多视觉任务上(如图像识别、目标检测等)获得了出色的结果。在人体模型重建领域中,有方法通过卷积神经网络提取图像特征向量,通过重复操作得到每个顶点的特征,并使用Transformer编码器回归逐顶点的3D坐标。但现有的方法只通过自注意力机制建模顶点与顶点间内部的关系,没有构建表达能力充足的图像特征到顶点特征的映射,对图像特征的利用不充分。

发明内容

[0006] 为解决上述问题,本申请提出一种基于Transformer解码器的3D人体模型重建方法,可充分利用RGB图片输入的信息,预测精准的目标3D人体模型。

[0007] 为了实现上述目的,本申请技术方案如下:

[0008] 一种基于Transformer解码器的3D人体模型重建方法,包括:

[0009] 对原始RGB图像进行预处理,获得包含目标人体的图像,并将其输入图像特征提取模块进行特征提取,获得图像网格特征;

[0010] 将一个模板3D人体模型的顶点坐标输入到线性层,获得编码后的模板顶点特征;

[0011] 将图像网格特征和模板顶点特征输入到顶点特征变换模块,融合后得到变换后的顶点特征;

[0012] 将变换后的顶点特征通过上采样模块进行上采样,并通过逆向线性层将上采样后的顶点特征解码为最终预测的3D人体模型顶点坐标。

[0013] 进一步的,所述图像特征提取模块使用ResNet网络。

[0014] 进一步的,所述模板3D人体模型使用无性别的、标准姿态的SMPL人体模型。

[0015] 进一步的,所述模板3D人体模型是经过网格下采样算法处理后的低分辨率的模板3D人体模型。

[0016] 进一步的,所述顶点特征变换模块包括预设数量的Transformer解码器,首层Transformer解码器的输入为图像网格特征和模板顶点特征,其他Transformer解码器的输入为图像网格特征和上一层Transformer解码器输出的顶点特征。

[0017] 进一步的,所述Transformer解码器包括依次连接的自注意力层、交叉注意力层与前馈网络层,所述自注意力层、交叉注意力层与前馈网络层前还分别设置有各自的归一化层,所述自注意力层、交叉注意力层与前馈网络层每一层的输入都与本层的输出进行累加后输入到下一层;所述自注意力层的输入是模板顶点特征或上一层Transformer解码器输出的顶点特征,所述交叉注意力层的输入还包括图像网格特征。

[0018] 进一步的,所述上采样层由两个相同的MLP网络组成,每个MLP由两个线性层与其间的非线性激活层组成,分阶段地将低分辨率的顶点特征转化为高分辨率的顶点特征。

[0019] 进一步的,所述线性层和逆向线性层严格互逆,所述线性层表示为:

$$[0020] \quad z_0 = V_{3D}^T W + b;$$

[0021] 其中, V_{3D}^T 表示模板3D人体模型的顶点坐标, z_0 表示模板顶点特征, W 和 b 分别表示线性层的权重矩阵与偏置向量;

[0022] 所述逆向线性层表示为:

$$[0023] \quad V_{3D} = (z_v - b) W^{-1};$$

[0024] 其中, V_{3D} 表示逆向线性层输出的3D人体模型顶点坐标, z_v 表示顶点特征变换模块输出的顶点特征, W^{-1} 表示 W 的广义右逆矩阵, $W^{-1} = W^T (WW^T)^{-1}$ 。

[0025] 进一步的,所述图像特征提取模块、线性层、顶点特征变换模块、上采样模块和逆向线性层构成网络模型,所述网络模型在训练时,采用的损失函数如下:

$$[0026] \quad \mathcal{L} = \lambda_1 \mathcal{L}_V + \lambda_2 \mathcal{L}_{J,3D} + \lambda_3 \mathcal{L}_{J,2D}$$

[0027] 其中, \mathcal{L}_V 表示最终预测的3D人体模型顶点坐标损失, $\mathcal{L}_{J,3D}$ 表示预测的3D人体关键点坐标损失, $\mathcal{L}_{J,2D}$ 表示预测的2D人体关键点坐标损失;所述预测的3D人体关键点坐标和预测的2D人体关键点坐标由最终预测的3D人体模型顶点坐标计算得到, λ_1 、 λ_2 、 λ_3 为超参数。

[0028] 本申请提出的一种基于Transformer解码器的3D人体模型重建方法,具有如下有益效果:

[0029] 1)、利用多阶段Transformer解码器的交叉注意力层构建图像网格特征到顶点特征的融合映射关系。相较于线性映射等简单的映射关系,本发明的方法可以更加充分地利用图像特征中的局部信息,提升预测的人体模型的精准度。

[0030] 2)、使用严格互逆的线性映射层处理顶点特征到顶点坐标之间的双向映射,同时将损失函数作用于顶点特征变换模块内部全部层的输出,保证了顶点特征变换模块内部及两侧的语义一致性,促使网络聚焦于学习更有价值的信息,而非关注不必要的语义转换。

[0031] 3)、提供了一种两阶段的训练方法,额外引入的模型平滑损失可以充分利用上采样模块本身的模型容量与表征能力,帮助网络输出更加平滑真实的3D人体模型。

[0032] 4)、提供的实施例采用较为轻量的网络设计,相较于领域内的其他方法,本申请的模型在保证预测精度的前提下容量更小,速度更快。

附图说明

[0033] 图1为本申请基于Transformer解码器的3D人体模型重建方法流程图;

[0034] 图2为本申请网络模型结构示意图;

[0035] 图3为本申请顶点特征变换模块示意图;

[0036] 图4为本申请实施例的上采样模块示意图;

[0037] 图5为本发明实施例的线性层与逆向线性层示意图。

具体实施方式

[0038] 以下结合附图及实施例对本发明做进一步详细说明。应强调的是,下述说明实施例仅用于解释本申请,而不是为了限制本发明的范围与应用。

[0039] 在一个实施例中,如图1所示,提供了一种基于Transformer解码器的3D人体模型重建方法,包括:

[0040] 步骤S1、对原始RGB图像进行预处理,获得包含目标人体的图像,并将其输入图像特征提取模块进行特征提取,获得图像网格特征。

[0041] 本步骤提取图像网格特征,先通过预处理模块进行预处理,然后再通过图像特征提取模块提取图像网格特征。

[0042] 如图2所示,预处理模块负责将输入的RGB图片进行裁剪与遮挡数据增广,以获得充分包含目标人像的图像。其中,上裁剪处理可使用人体检测方法,估计得到目标人像在原图像中被充分包含的矩形边界框,并据此进行图像裁剪。在一个实施例中,可使用YOLO v3等模型进行人体检测,检测得到的矩形边界框会根据最长的边进行扩张,以形成一个正方形边界框,据此裁剪原始RGB图片以得到一张正方形的预处理后图片,并调整尺寸统一为 224×224 的分辨率。裁剪区域超出原始图像边界时,超出的部分的像素强度置零。还可以使用其他方法来进行图像裁剪,这里不再赘述。

[0043] 上述遮挡数据增广可向裁剪后的图像上进一步叠加噪声,采用简单几何图形的噪声图案与取自分割数据集的实例物体图案进行随机组合,对目标图像的局部进行随机遮挡,为训练过程提供分布更广的训练样本。在本实例中,除了随机几何图案噪声外,还可选用Pascal VOC 2012数据集中的与人体无关的实例图案。使用已训练的模型对待处理的RGB图像进行3D人体模型重建时,仅进行裁剪即可。

[0044] 图像特征提取模块负责对预处理后的图像提取特征,获得图像网格特征。在本实例中,使用了ResNet50模型提取图像特征,也可使用ResNet101、HRNet-W64等模型提取图像特征。提取的图像网格特征分辨率缩小至原图的 $1/32$,为 7×7 ,同时其空间维度被拉平,形

成一个长度为 $7 \times 7 = 49$ 的图像网格特征序列,图像网格特征空间维度映射至512。

[0045] 步骤S2、将一个模板3D人体模型的顶点坐标输入到线性层,获得编码后的模板顶点特征。

[0046] 在本步骤中,线性层将模板3D人体模型的顶点坐标编码为对应的顶点特征。在本实例中,模板3D人体模型采用无性别的、标准姿态的SMPL人体模型,并预先通过网格模型下采样算法进行压缩,将原始的6980个顶点压缩为431个顶点的低分辨率的模板3D人体模型。

[0047] 事先对网格进行压缩可以降低Transformer处理的序列长度,从而减少内存使用量。本实例中,线性层将3维顶点坐标空间编码至512维顶点特征空间。

[0048] 需要说明的是,线性层和逆向线性层,常被用于特征空间维度的调整,是比较成熟的技术,这里不再赘述。

[0049] 步骤S3、将图像网格特征和模板顶点特征输入到顶点特征变换模块,融合后得到变换后的顶点特征。

[0050] 在本步骤中,顶点特征变换模块包括预设数量L层的Transformer解码器。如图3所示,首层Transformer解码器的输入为图像网格特征和模板顶点特征,其他Transformer解码器的输入为图像网格特征和上一层Transformer解码器输出的顶点特征。

[0051] 如图3所示,所述Transformer解码器包括依次连接的自注意力层、交叉注意力层与前馈网络层,所述自注意力层、交叉注意力层与前馈网络层前还分别设置有各自的归一化层,所述自注意力层、交叉注意力层与前馈网络层每一层的输入都与本层的输出进行累加后输入到下一层;所述自注意力层的输入是模板顶点特征或上一层Transformer解码器输出的顶点特征,所述交叉注意力层的输入还包括图像网格特征。交叉注意力层的输入不仅有图像网格特征,还包括自注意力层的输出与自注意力层的输入进行累加后的顶点特征。

[0052] 自注意力层可对顶点与顶点间的关联进行建模,交叉注意力层对图像网格特征与顶点特征间的关联进行建模。顶点特征变换模块的每个交叉注意力层中,图像网格特征与顶点特征会分别作为Key与Query计算一个交叉注意力图,从而据此软式地提取出Value(图像网格特征)中网络所关注的特征,并将其映射、叠加于顶点特征空间内,形成变换后的顶点特征。

[0053] 在整个过程中,输入的顶点特征序列长度自始至终保持不变,顶点特征维度与图像网格特征维度一致。

[0054] 本申请逐层地、内容相关地特征变换与融合方式可以更充分地利用图像网格特征中的局部信息。在本实例中,顶点特征序列的长度自始至终保持为下采样后的模型顶点数,即431。顶点特征变换模块中共包含结构完全相同的 $L=6$ 个Transformer解码器。

[0055] 步骤S4、将变换后的顶点特征通过上采样模块进行上采样,并通过逆向线性层将上采样后的顶点特征解码为最终预测的3D人体模型顶点坐标。

[0056] 本步骤主要是通过上采样模块与逆向线性层对变换后的顶点特征进行处理。

[0057] 上采样模块用于将变换后的顶点特征进行上采样操作,将低分辨率的顶点特征扩充为高分辨率的顶点特征。在本实例中,如图4所示,上采样模块由两个相同的MLP网络组成,每个MLP由两个线性层与其间的非线性激活层组成。非线性激活层使用ReLU激活函数。分阶段地将低分辨率的顶点特征转化为高分辨率的顶点特征。3D人体网格模型的拓扑结构

是完全固定的,解码后得到的3D人体模型顶点坐标也是按序排列,从而重建出3D人体模型。

[0058] 逆向线性层用于将上采样模块输出的高分辨率的顶点特征解码至高分辨率顶点坐标。为了保证顶点特征的语义一致性,此逆向线性层与步骤S2中的线性层严格互逆。在本实例中,如图5所示,步骤S2的线性层首先被定义,可表示为:

$$[0059] \quad z_0 = V_{3D}^T W + b;$$

[0060] 其中, V_{3D}^T 表示模板3D人体模型的顶点坐标, z_0 表示模板顶点特征, W 和 b 分别表示线性层的权重矩阵与偏置向量。而后,步骤S4的逆向线性层共享步骤S2中线性层的参数,可表示为:

$$[0061] \quad V_{3D} = (z_V - b) W^{-1};$$

[0062] 其中, V_{3D} 表示逆向线性层输出的3D人体模型顶点坐标, z_V 表示顶点特征变换模块输出的顶点特征, W^{-1} 表示 W 的广义右逆矩阵,可表示为:

$$[0063] \quad W^{-1} = W^T (W W^T)^{-1};$$

[0064] 由于 W^{-1} 与 W 互逆,上述两个线性层严格互逆,且整个方法的初始输入与最终输出都为三维顶点坐标,进而保证了Transformer解码器两侧的顶点特征的语义一致。在本实例中,这样的设计可避免网络关注不必要的语义转换,而聚焦于学习更有价值的信息。

[0065] 在一个具体的实施例中,本申请图像特征提取模块、线性层、顶点特征变换模块、上采样模块和逆向线性层,构成整个网络。在训练这个网络模型时,使用多种损失项的组合对网络输出的人体模型进行监督。

[0066] 本实施例,损失函数如下:

$$[0067] \quad \mathcal{L} = \lambda_1 \mathcal{L}_V + \lambda_2 \mathcal{L}_{J,3D} + \lambda_3 \mathcal{L}_{J,2D}$$

[0068] 其中, \mathcal{L}_V 表示最终预测的3D人体模型顶点坐标损失, $\mathcal{L}_{J,3D}$ 表示预测的3D人体关键点坐标损失, $\mathcal{L}_{J,2D}$ 表示预测的2D人体关键点坐标损失;所述预测的3D人体关键点坐标和预测的2D人体关键点坐标由最终预测的3D人体模型顶点坐标计算得到, λ_1 、 λ_2 、 λ_3 为超参数。

[0069] 具体的,对所述最终预测的3D人体模型顶点坐标计算与真值的L1距离损失 \mathcal{L}_V ,对预测的3D人体关键点坐标计算与真值的L1距离损失 $\mathcal{L}_{J,3D}$,对预测的2D人体关键点坐标计算与真值的L1距离损失 $\mathcal{L}_{J,2D}$ 。

[0070] 预测的3D人体关键点坐标 J_{3D} 由预测的3D模型顶点坐标 V_{3D} 与一个固定的关键点回归矩阵 G 相乘得到:

$$[0071] \quad J_{3D} = G V_{3D}$$

[0072] 预测的2D人体关键点坐标 J_{2D} 由 J_{3D} 经过弱透视投影得到:

$$[0073] \quad J_{2D} = (\Pi_{x,y} J_{3D} + t) s$$

[0074] 其中, t 与 s 是弱透视投影相机参数,分别表示偏移参数与缩放参数。所述损失项分别定义如下:

$$[0075] \quad \mathcal{L}_V = |V_{3D}^{GT} - V_{3D}|_1$$

$$[0076] \quad \mathcal{L}_{J,3D} = |J_{3D}^{GT} - J_{3D}|_1$$

$$[0077] \quad \mathcal{L}_{J,2D} = |J_{2D}^{GT} - J_{2D}|_1$$

[0078] 其中 V_{3D}^{GT} 、 J_{3D}^{GT} 、 J_{2D}^{GT} 分别代表3D人体模型顶点坐标、3D人体关键点坐标、2D人体关键点坐标的真值。

[0079] 在另一个实施例中,训练被分为了两个阶段。第一阶段采用上述损失函数,训练除了上采样模块之外的全部可训练网络参数。此阶段训练时,上采样模块由步骤S2所述的模型网格下采样算法生成的采样矩阵(即一个固定的线性插值模块)替代。第二阶段,在前述损失函数的基础上,引入额外的损失项,只对上采样模块进行训练,以提升其输出平滑的3D人体模型的能力。

[0080] 具体地,额外引入以下损失项:

$$[0081] \quad \mathcal{L}_{edge} = \sum_f \sum_{\{i,j\} \subset f} \left| \|v_i - v_j\|_2 - \|v_i^{GT} - v_j^{GT}\|_2 \right|$$

$$[0082] \quad \mathcal{L}_{norm} = \sum_f \sum_{\{i,j\} \subset f} \left| \left\langle \frac{v_i - v_j}{\|v_i - v_j\|_2}, n_f^{GT} \right\rangle \right|$$

[0083] 其中, f 为网格中任意面片的编号, i 、 j 为编号为 f 的面片中任意两个不相同顶点的编号, v 与 v^{GT} 分别代表某顶点的预测坐标与真值坐标, n^{GT} 代表某面片的法线方向向量。需要指出的是,本申请预测的人体模型与真值人体模型的拓扑结构完全相同,都固定为SMPL模型,同时拥有6890个固定数量的顶点与13776个固定数量的面片;步骤S4输出的顶点坐标是按照预设顶点编号有序排列的,因此预测模型与真值模型中一对相同编号的顶点、边或面片一一对应。上述两个额外的损失项中, \mathcal{L}_{edge} 衡量了预测的人体模型与真值人体模型中每一对相同编号的边的长度差异, \mathcal{L}_{norm} 衡量了预测的人体模型与真值人体模型中每一对相同编号的面片的法线方向差异。引入额外平滑项后的总损失函数为:

$$[0084] \quad \mathcal{L} = \lambda_1 \mathcal{L}_V + \lambda_2 \mathcal{L}_{J,3D} + \lambda_3 \mathcal{L}_{J,2D} + \lambda_4 \mathcal{L}_{edge} + \lambda_5 \mathcal{L}_{norm}$$

[0085] 其中, λ_4 、 λ_5 也为超参数。

[0086] 在上述两个实例中,所有损失项不仅作用于顶点特征变换模块的最后一层(即1=6)Transformer解码器的输出上,还作用在其他层($1 \in \{1, 2, 3, 4, 5\}$)的Transformer解码器的输出上。这样的训练策略强制让顶点特征变换模块内部的层输出的顶点特征也具有和最终输出相同的语义,进一步确保顶点特征的语义一致;宏观地,整个网络将模板人体模型顶点特征在图像网格特征的帮助下进行逐层、渐进地变形为输出的顶点特征,最终被解码为输出人体模型顶点坐标。

[0087] 以上所述实施例时结合具体的实施方式对本发明所做的进一步详细说明,但不能认定本发明的具体实施只局限于这些说明。应当指出的是,对于本发明所属技术领域的普通技术人员来说,在不脱离本发明构思的前提下,其还可以对这些已描述的实施方式做出若干替代或变型,而这些替代或变型方式都应当视为属于本发明的保护范围。因此,本申请专利的保护范围应以所附权利要求为准。

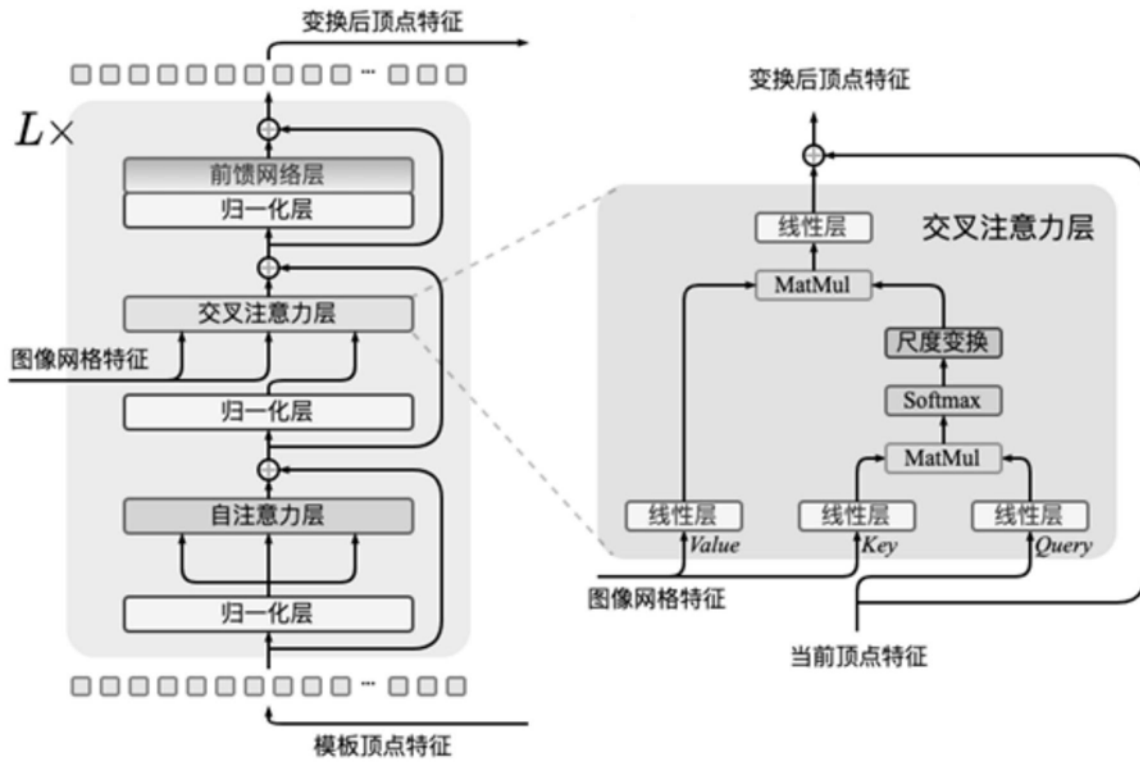


图3

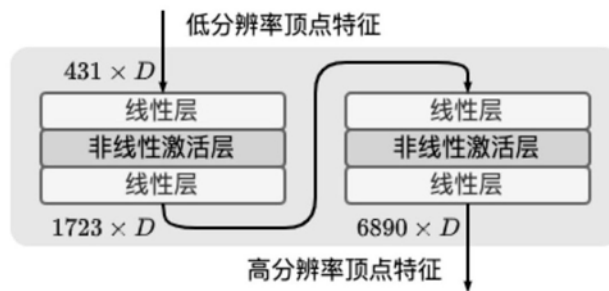


图4

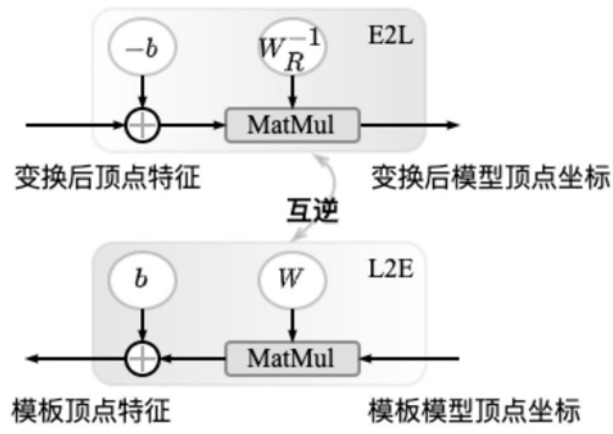


图5