

文献分享_Mesh 合成

潘治宇 20200709

空间物体模型的三维表示形式

用voxel或者点云在三维空间表示一个物体往往比较粗糙，而且很难看出模型表面的细节特征，用mesh表示物体会更清晰。

Voxel,



point cloud,



mesh



人体的参数化模型 以SMPL为例

Brief Introduction to SMPL (Skinned Multi-Person Linear) model

Pose parameters:

$$\vec{\theta} = [\vec{\omega}_0^T, \dots, \vec{\omega}_K^T]^T \text{ where } \vec{\omega}_K^T \in \mathbb{R}^3$$

SMPL model has 23 joints, it also needs one angle to represent the orientation, Therefore $\vec{\theta} \in \mathbb{R}^{3 \times 24}$

Shape parameters:

$$\vec{\beta} \in \mathbb{R}^{10}$$

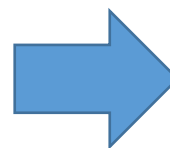
Stand for the PCA (principal component analysis) representation of low dimensional shape coefficients learned from registered meshes. Generally, using the first 10 main coefficients.

$$M(\beta, \theta) = W(T(\beta, \theta), J(\beta), \theta, \mathbf{W})$$

$$T(\beta, \theta) = \mathbf{T}_\mu + B_s(\beta) + B_p(\theta)$$



6890 vertexes coordinates;
23 joints' coordinate in 3D



Mesh generation

□ Model-based

通过学习或优化隐参数 例如：学习SMPL 姿态向量 θ and 形状向量 β 来得到人体模型

方法：HMR, SMPLify, SPIN, VIBE 等

优：生成的模型更容易像个人样，便于解决单视角下scale或 depth ambiguity

缺：shape表示不灵活，且缺少 Image \rightarrow mesh 的紧密关联性 (dense correspondence)

□ Model-free (free-form) (Voxel, point cloud, mesh...)

直接对mesh的vertex进行表示/操作，不采用（或部分采用）模型的隐参数

（另起炉灶；or 添砖加瓦）

优：可以进行更细节的形状表示，且更容易建立起 **image pixel \rightarrow mesh vertex 的关联性**

缺：训练难度较model-based较大，缺少相关数据集，方向仍在萌芽期

分享文献

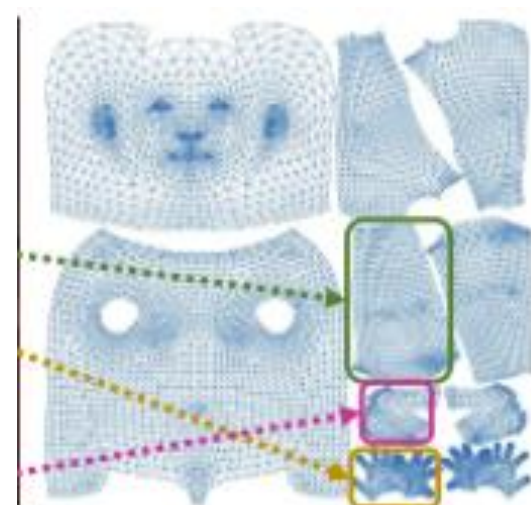
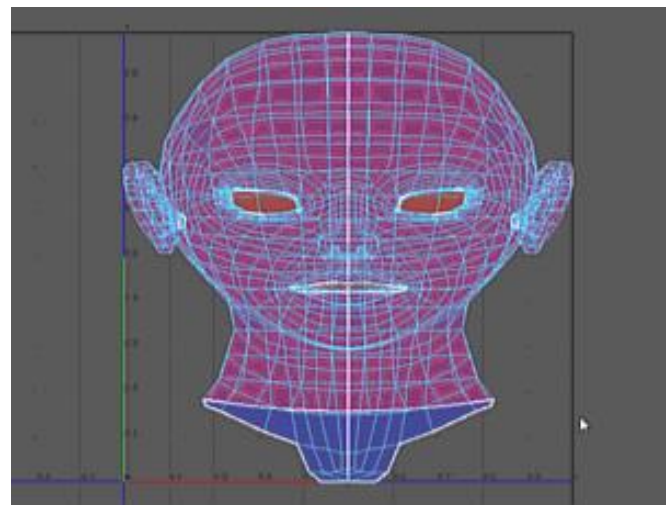
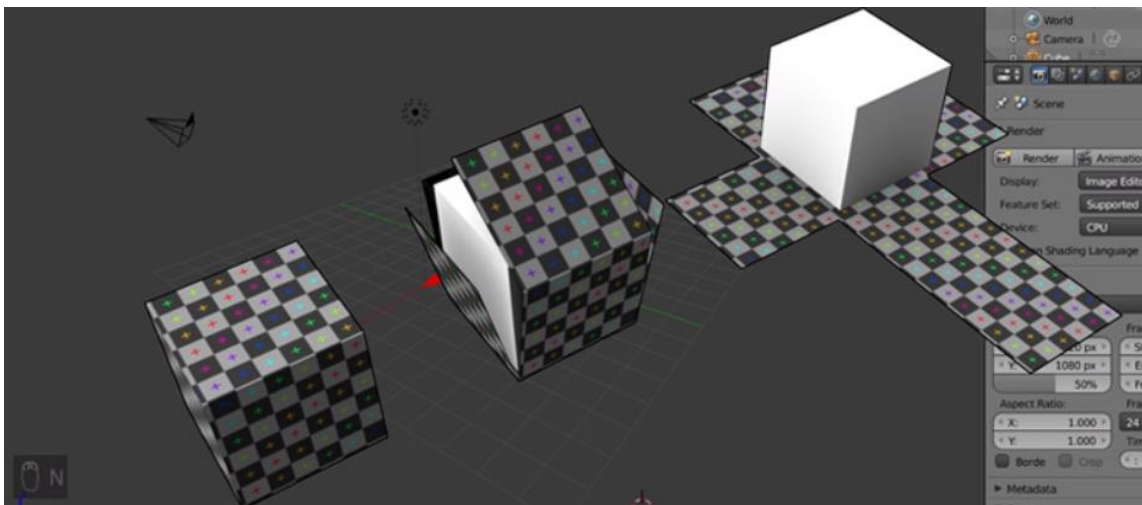
- ❑ Densepose: Dense human pose estimation in the wild (CVPR 2018) (IUV 图的诞生)
[图片→IUV image]
- ❑ Tex2shape: Detailed full human body geometry from a single image (ICCV 2019)
(UV 空间存储模型的形状细节信息) [图片 → UV 空间 法线图 偏移图]
- ❑ 3D Human Mesh Regression with Dense Correspondence (CVPR 2020) (UV 空间存储模型的姿态信息) [图片 → UV 空间的顶点坐标图]
- ❑ Learning to Transfer Texture from Clothing Images to 3D Humans (CVPR 2020) (UV map 作为 inverse IUV map 的作用) [图片 → UV 空间 图片像素坐标图]
- ❑ Implicit Functions in Feature Space for 3D Shape Reconstruction and Completion (CVPR 2020) (隐函数对Mesh的精确表示) [体素, 点云 → mesh]

UV map

UV map: 将三维物体模型的**标准模板**的mesh 展到 2D 平面（可以展开成一个部分，也可以展开成多个部分），展开方式不唯一

→ 2D 平面（UV 空间）上的pixel 点与3D 空间中mesh的顶点具有对应关系。

→ 3D mesh 在2D UV 空间中的表示，UV空间中pixel的值可以储存相关顶点的相关信息（贴图）。



UV map --- 贴图

UV 空间中表示原模型的纹理，可直接用于模型的纹理贴图



brown_eye



eyebrow004



eyelashes04



female_sportsui
t01_diffuse



female_sportsui
t01_normal



ponytail01_diff
use



young_lightskin
ned_female_diff
use2

UV map

- UV 空间（2维）到 mesh空间（3维）建立关系
- → 便于神经网络找到从原始2维图片XY空间（2维）到 UV 空间（2维）的学习关系
- 但是：XY空间 和 UV 空间不是对齐的，具有不同的意义
- 从 XY-空间 学习特征直接预测 UV 空间的信息 仍然缺少image 到 mesh 的关联性，XY 空间和 UV空间 他们不是对齐的。
- 需要某种方式 建立起从 XY 空间 到 UV 空间 的联系
- → Densepose

Densepose CVPR 2018

- Create the dense correspondence between image \rightarrow mesh



Ref: Alp Güler, Rıza, Natalia Neverova, and Iasonas Kokkinos. "Densepose: Dense human pose estimation in the wild." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018.

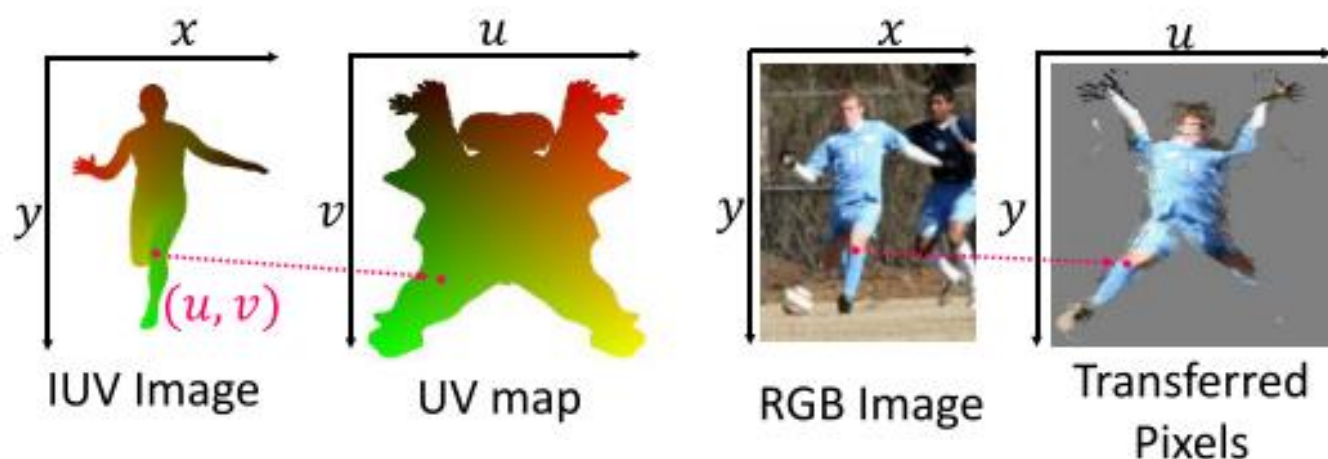
2020/7/13

15

IUV image

- IUV: 图片每个像素点的值对应该点在UV空间的UV坐标,
- IUV image的作用是作为一种转换关系 \rightarrow 从 XY 像素空间 映射到 UV 空间

IUV 作用在原图的纹理转换

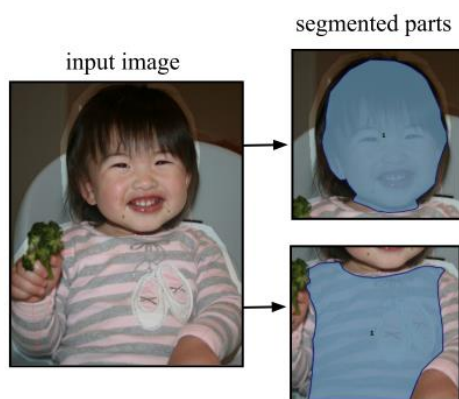
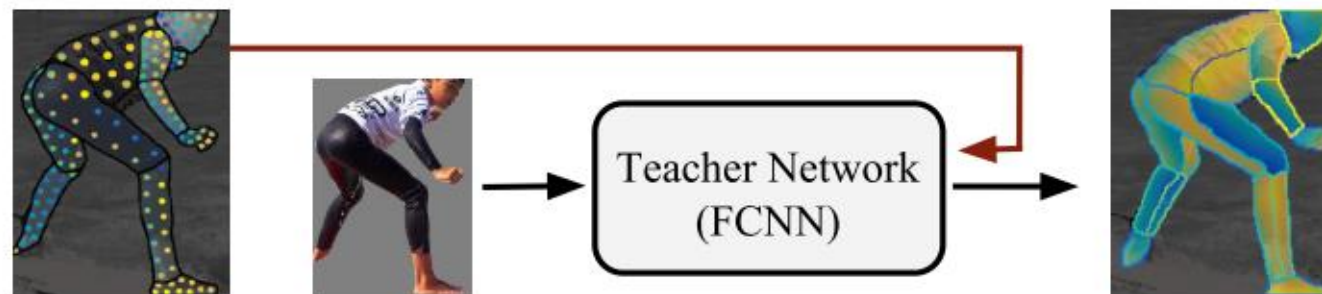
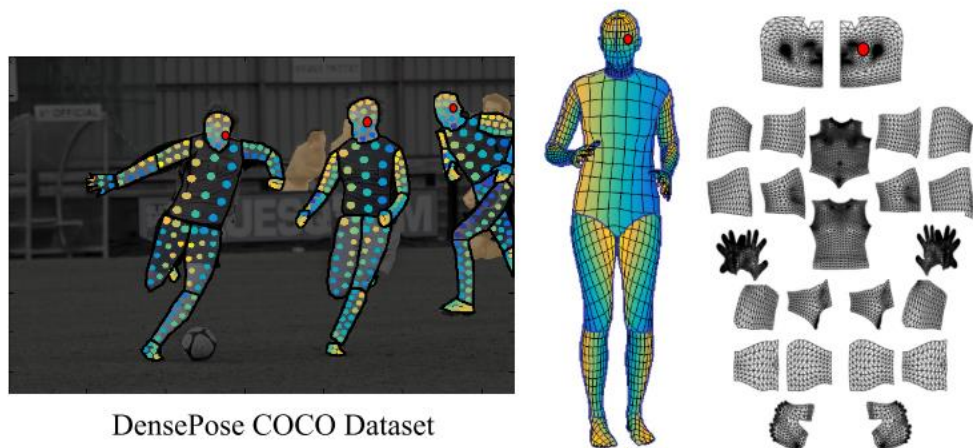


IUV image 代表的对应关系，让 mask 前景的每一个像素点的值代表 UV 空间的唯一坐标。

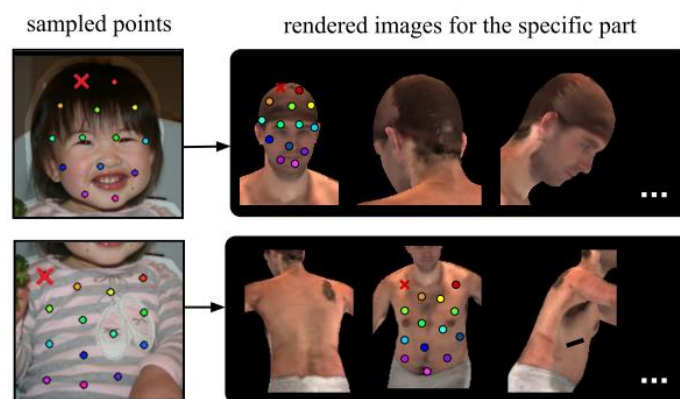
数据准备

□ COCO-DensePose Dataset + Teacher Network

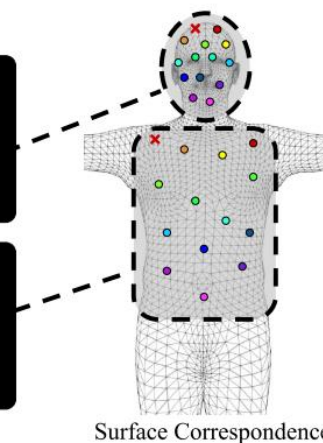
Densepose的UV 展开图
将人体分割成24个部分



TASK 1: Part Segmentation



TASK 2: Marking Correspondences

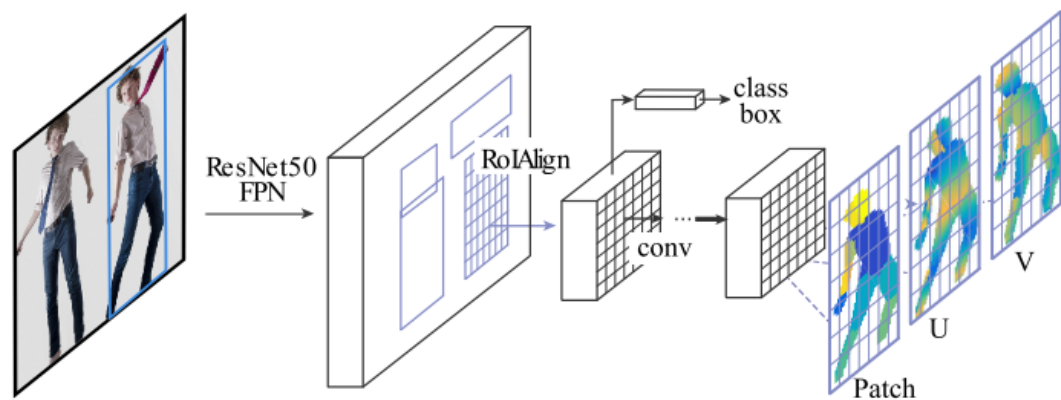


Surface Correspondence

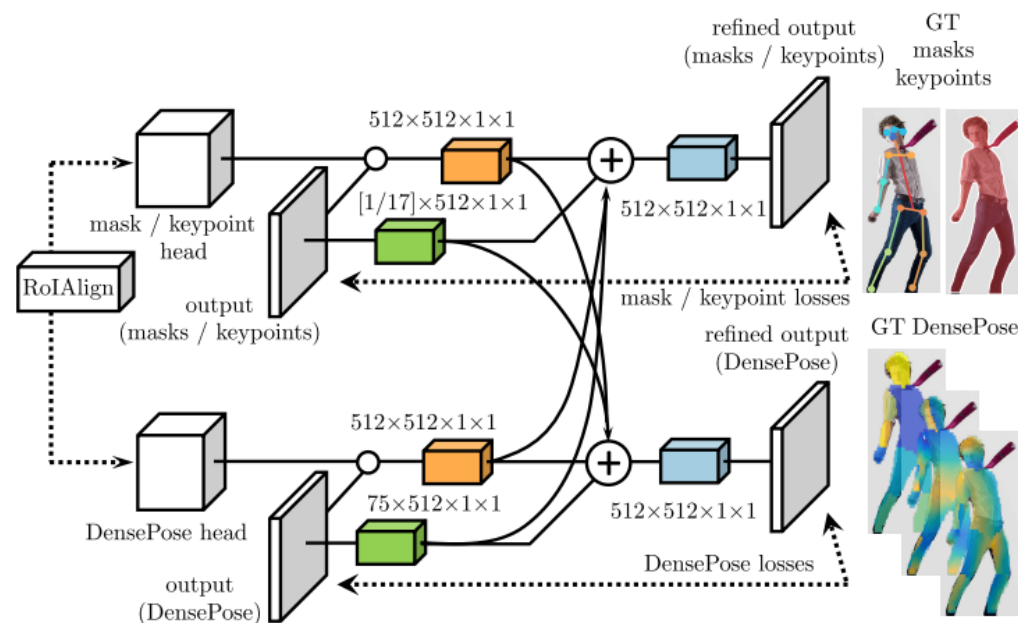
2020/7/13

17

网络结构



训练目标：获得24个part的分割图以及各自在UV空间的坐标值（IUUV imae）



结果

□ 特殊的2D pose estimation 并且能够反应遮挡关系

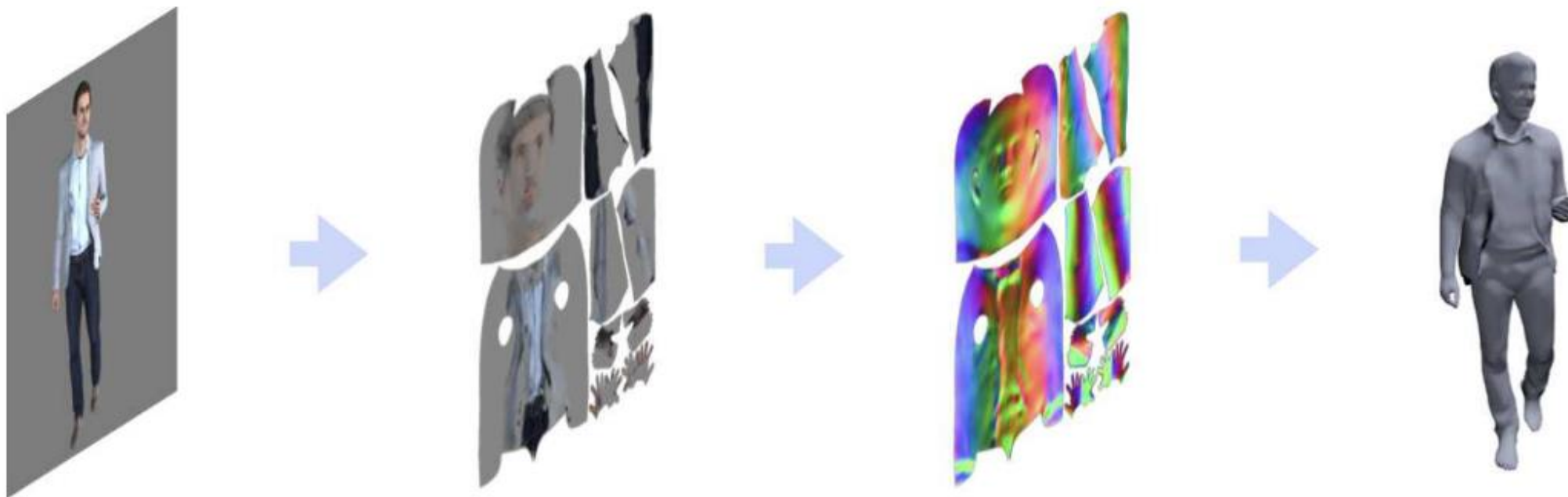


分析

- UV 概念的出现使 image 到 UV map 建立关系，便于网络在UV 空间进行学习
- UV map 不是仅仅用来贴图的， UV map 不仅仅可以保存贴图(纹理)信息

Tex2shape (ICCV 2019)

完全的pose-independent, 仅优化人物模型的shape



Ref: Alldieck, Thiemo, et al. "Tex2shape: Detailed full human body geometry from a single image." *Proceedings of the IEEE International Conference on Computer Vision*. 2019.

创新点

□ Extended SMPL template

$N = 6890$ vertices and $F = 13776$ faces $\rightarrow N = 27554$ vertices and $F = 55104$ faces

□ 思路：预测出基于UV 空间的 normal map 和 displacement map 来补充细节

\rightarrow 将一个困难的**3D模型形状细节重构**的问题转换成从**图片到图片的转换**问题。

\rightarrow The displacement maps encode the free-form offsets, that are not part of SMPL;

\rightarrow The normal maps contain surface normals in canonical T-pose.

□ 网络预测 两个 map 一个 vector β

数据准备

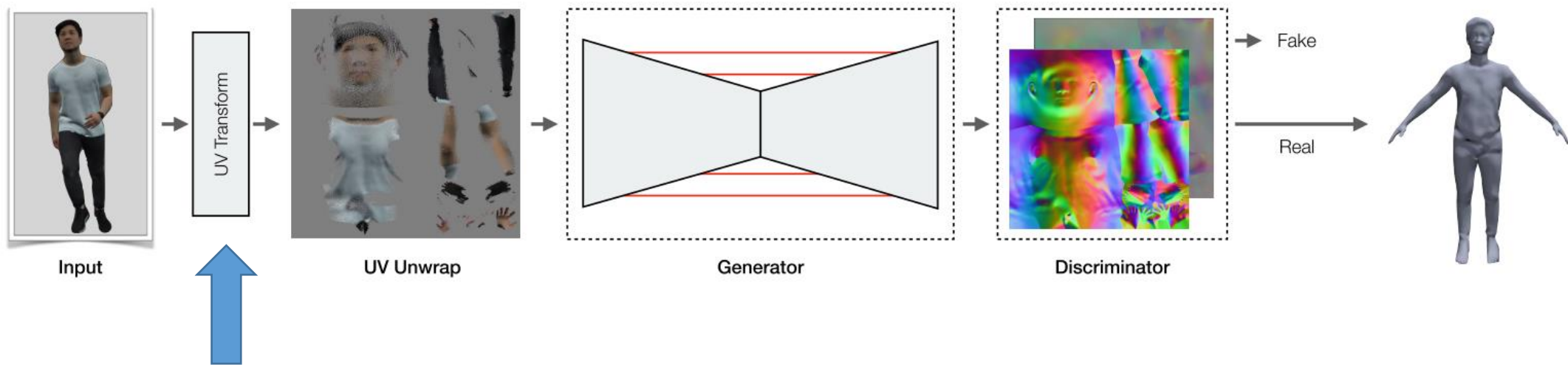
□ Difficulty : Dataset

□ Synthetic dataset : 共 2043 meshes 验证: 20 meshes 测试: 55 meshes

□ 构造: 购买2000+的人体扫描mesh, 用extended SMPL template做配准, 计算normal / displacement (SMPL 模型匹配; 光照渲染得到2D image; UV map信息获取)

□ 问题: 数据集数量较少, 均为合成, 且男多女少

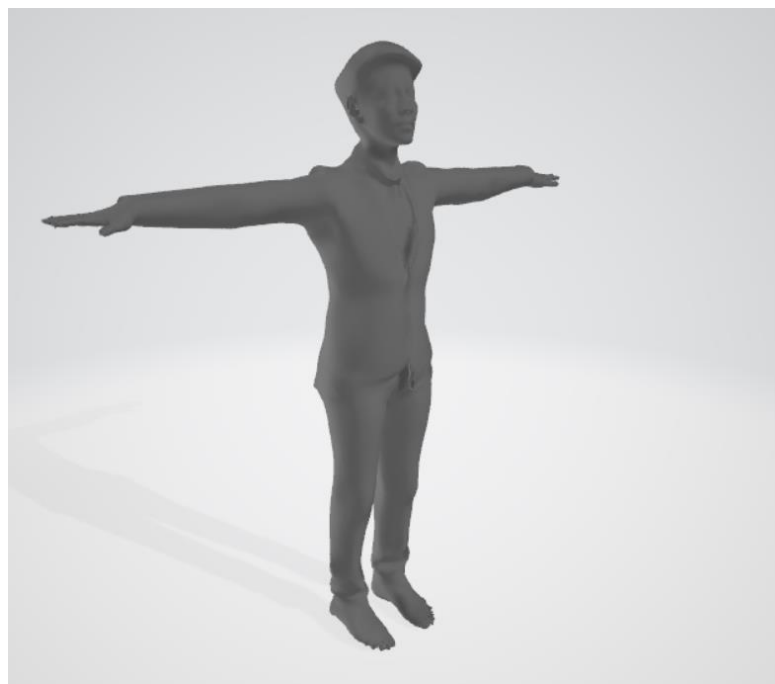
网络结构



利用Densepose得到的IUV image 将
原XY 空间的RGB图映射到UV 空间

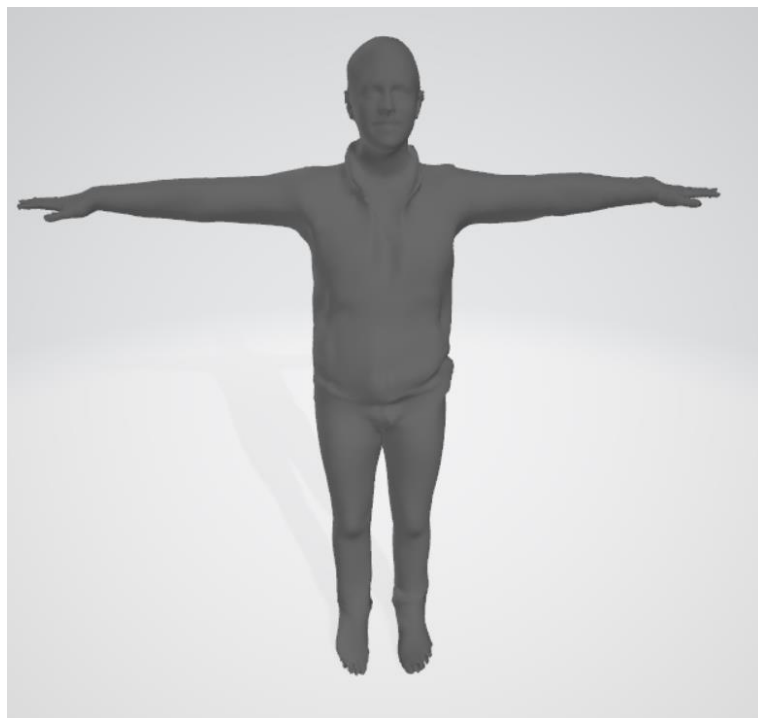
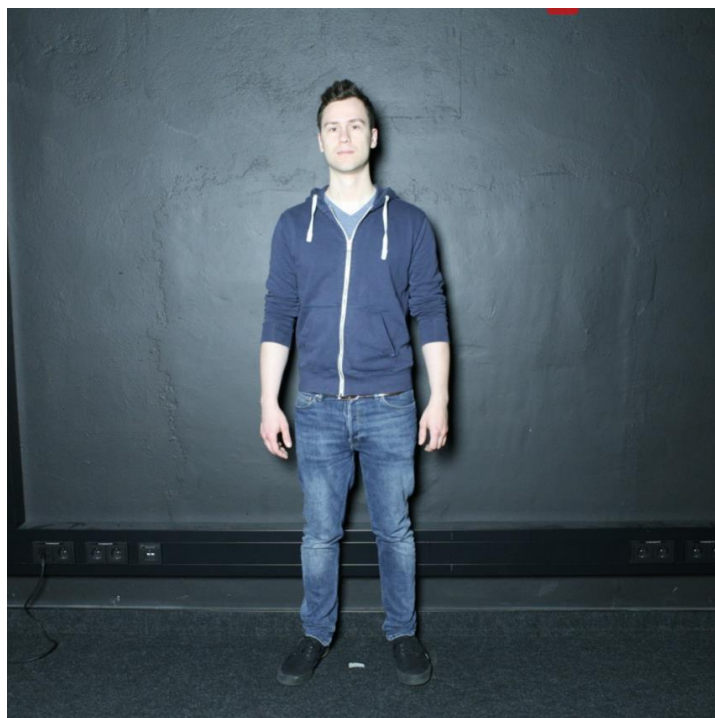
复现结果展示

□ 代码复现的结果：（仿真数据图片输入） 因为该文章只优化shape，所以pose采用标准T-pose



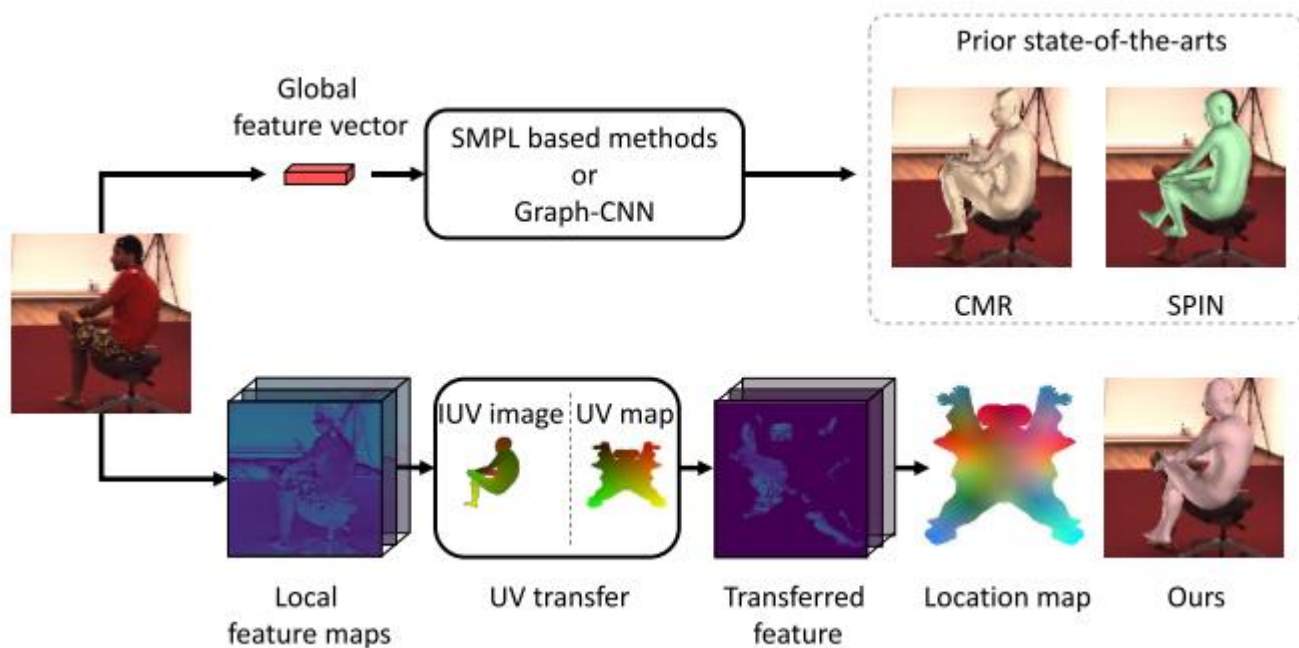
复现结果展示

□ 代码复现的结果：（现实场景图片输入）



3D Human Mesh Regression with Dense Correspondence (CVPR 2020)

- UV map 包含 vertex 位置信息 → 可以决定模型的pose



Ref: Zeng, Wang, et al. "3D Human Mesh Regression with Dense Correspondence." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020.

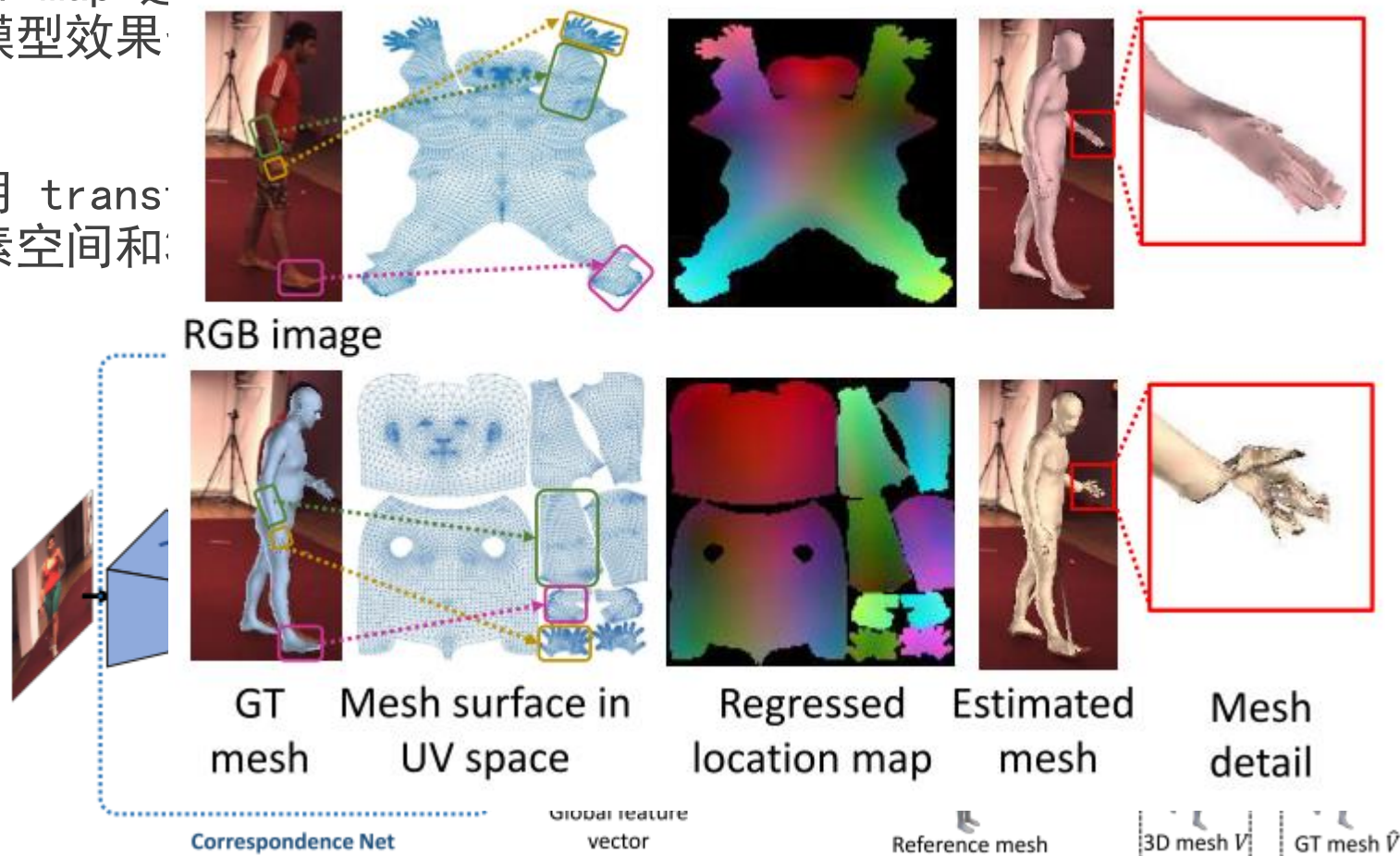
创新点

□ UV map 建模效果

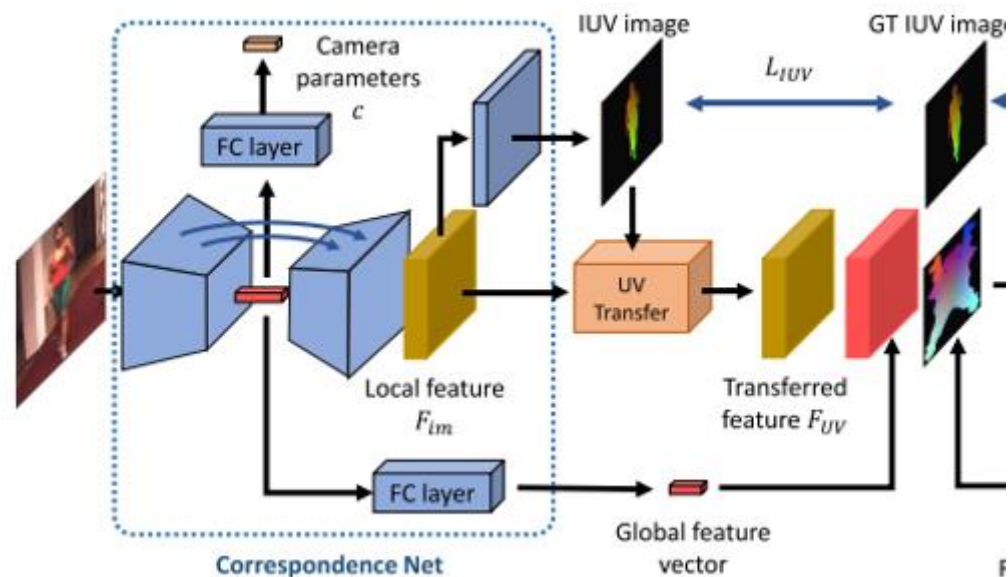
□ 用 trans 素空间和

录的坐标信息合成的

tex 的**强**关联性（像



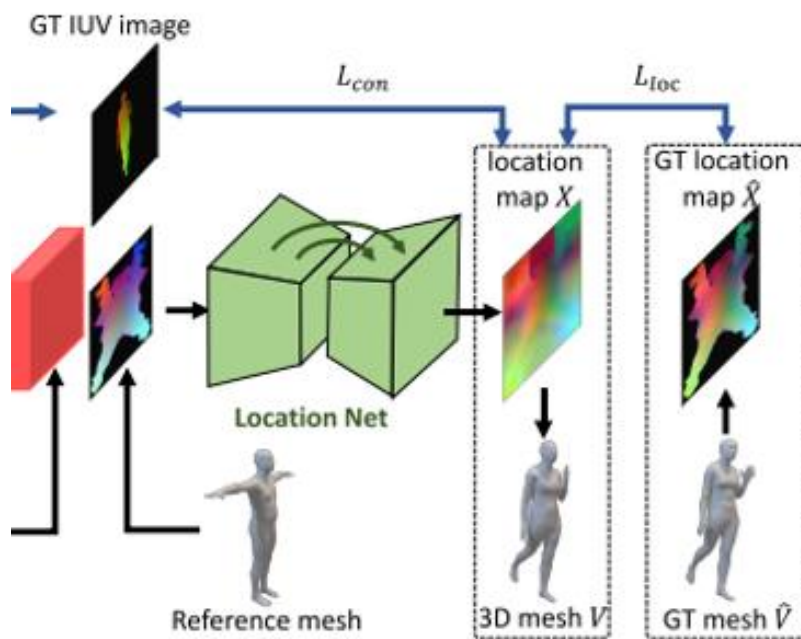
网络结构



$$\mathcal{L}_{IUUV} = \lambda_c \mathcal{L}_c + \lambda_r \mathcal{L}_r,$$

\mathcal{L}_c is a dense binary cross-entropy loss for classifying each pixel as 'fore' or 'back', \mathcal{L}_r is an l1 dense regression loss for predicting the exact UV coordinates

网络结构



$$\mathcal{L}_{loc} = \mathcal{L}_{map} + \mathcal{L}_J^{3D} + \mathcal{L}_J^{2D}.$$

$$\mathcal{L}_{map} = \sum_u \sum_v W(u, v) \cdot \left\| X(u, v) - \hat{X}(u, v) \right\|_1.$$

\mathcal{L}_{map} 为预测的location map 与GT 的L1 loss（带权重）；

\mathcal{L}_{3D} 与 \mathcal{L}_{2D} 是由模型计算的关节点位置与GT的L1 loss

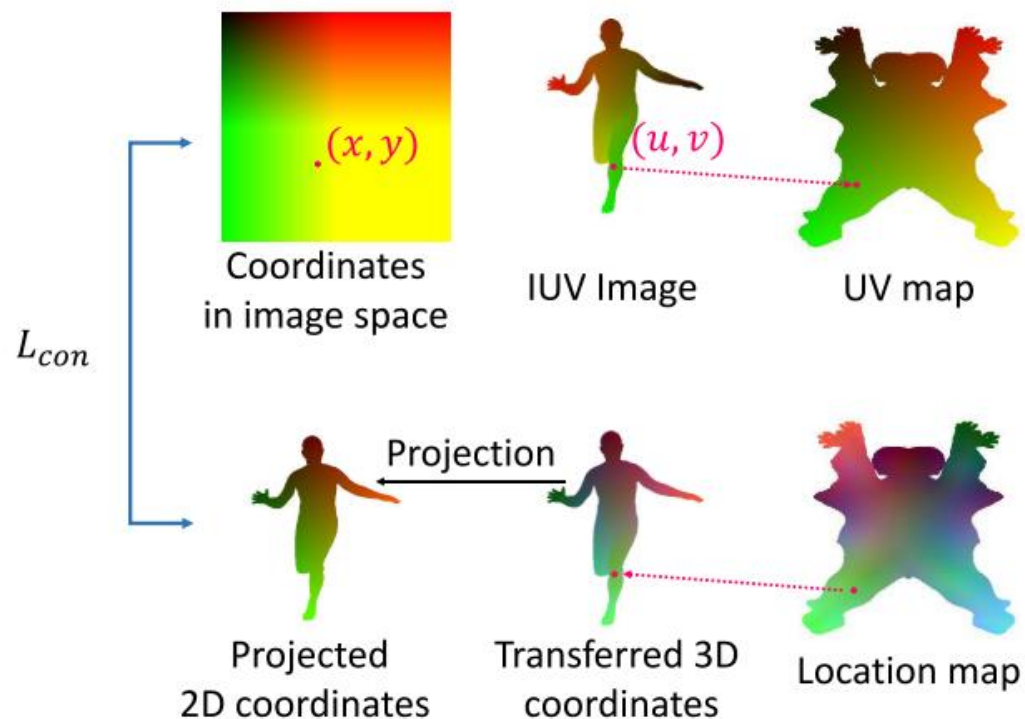
损失函数

$$\mathcal{L} = \mathcal{L}_{IUV} + \mathcal{L}_{Loc} + \lambda_{con} \mathcal{L}_{con}.$$

$$\mathcal{L}_{IUV} = \lambda_c \mathcal{L}_c + \lambda_r \mathcal{L}_r,$$

$$\mathcal{L}_{loc} = \mathcal{L}_{map} + \mathcal{L}_J^{3D} + \mathcal{L}_J^{2D}.$$

$$\mathcal{L}_{con} = \sum_{(x,y)} \|(x,y) - \pi(X(u,v), c)\|_2^2,$$

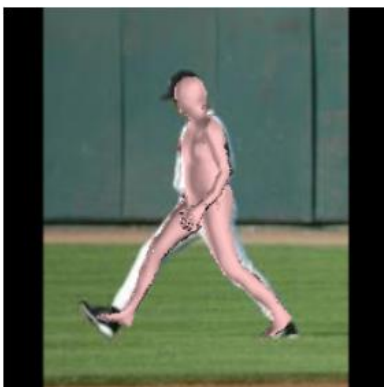


Methods	MPJPE-PA
Lassner <i>etc.</i> [22]	93.9
SMPLify [4]	82.3
Pavlakos <i>etc.</i> [26]	75.9
HMR[18]	56.8
NBF[25]	59.9
CMR[21]	50.1
DenseRaC[36]	48.0
SPIN[20]	41.1
Ours	39.3

Methods	Surface Error
SMPLify++ [22]	75.3
Tunget <i>al.</i> [32]	74.5
BodyNet[33]	73.6
Ours	56.5

复现结果展示

代码复现结果：



2020/7/13

35

Learning to Transfer Texture from Clothing Images to 3D Humans (CVPR 2020)

- 衣服正背面图片 → 贴纹理的衣服模型 → 找到 image空间 到 UV 空间的对应关系
- RGB 信息输入? Mask + 相关坐标信息输入 得到 image XY 到 UV 空间的联系 ✓



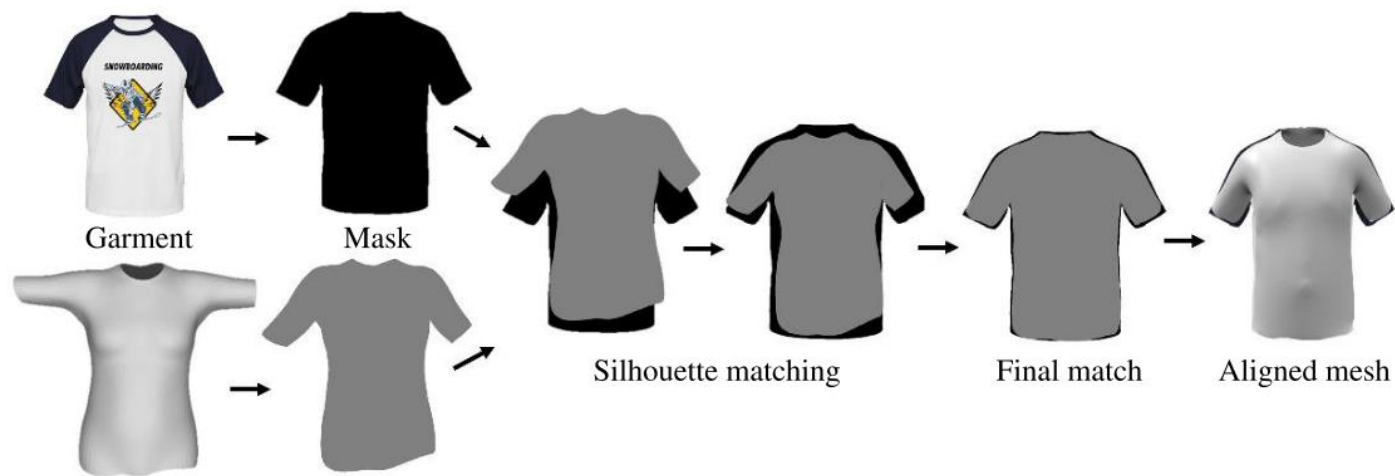
Ref: Mir, Aymen, Thiemo Alldieck, and Gerard Pons-Moll. "Learning to transfer texture from clothing images to 3d humans." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020.

创新点

- ❑ 方法：SMPL 模型为基础，截取衣服部分的mesh作为garment mesh，UV 空间展开分为 mesh-front UV 展开和mesh-back UV 展开，网络分别预测前部mesh 和 背部 mesh
- ❑ 特殊性：图片pixel到UV pixel的关系基本是 一一 对应的，UV 空间并没有表示整体全部的Mesh.
- ❑ 网络目的：要建立UV 空间坐标到image空间坐标的对应关系。
- ❑ 预测类似于Inverse IUV image 的图

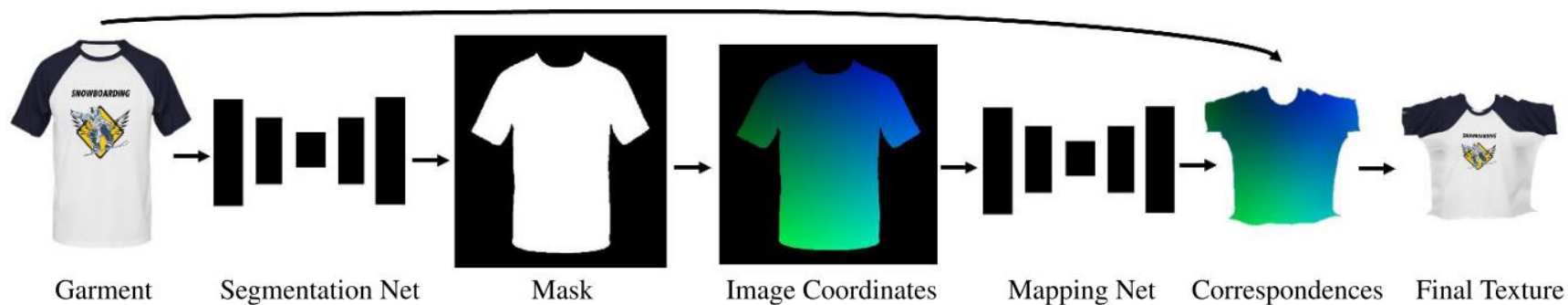
数据准备以及网络结构

I. Data collection by non-rigid 3D mesh to image alignment



训练数据的建立

II. Pix2Surf Network



Pix2surf 网络结构

除此之外 (From voxel or point cloud to mesh)

□ Mesh 还可以被implicit functions 表示

$$f(\mathbf{z}, \mathbf{p}): \mathcal{Z} \times \mathbb{R}^3 \mapsto [0, 1],$$

\mathcal{Z} 是将相关模型编码的相关的shape vector (global vector), \mathbf{p} 是 3D 空间的坐标值。

$f(\cdot)$ 输入是 \mathbf{z} 和 \mathbf{p} , 输出为一个 $0 \sim 1$ 的值, 0表示该点在mesh外, 1表示该点在mesh内。

→ 可以进行连续的mesh预测, 可以恢复mesh的一些细节, 对一些刚体预测效果较好。

缺点: 复杂形状物体, 例如人类恢复较差; 容易漏掉原始3D 输入的一些细节。

Implicit Functions in Feature Space for 3D Shape Reconstruction and Completion (CVPR 2020)

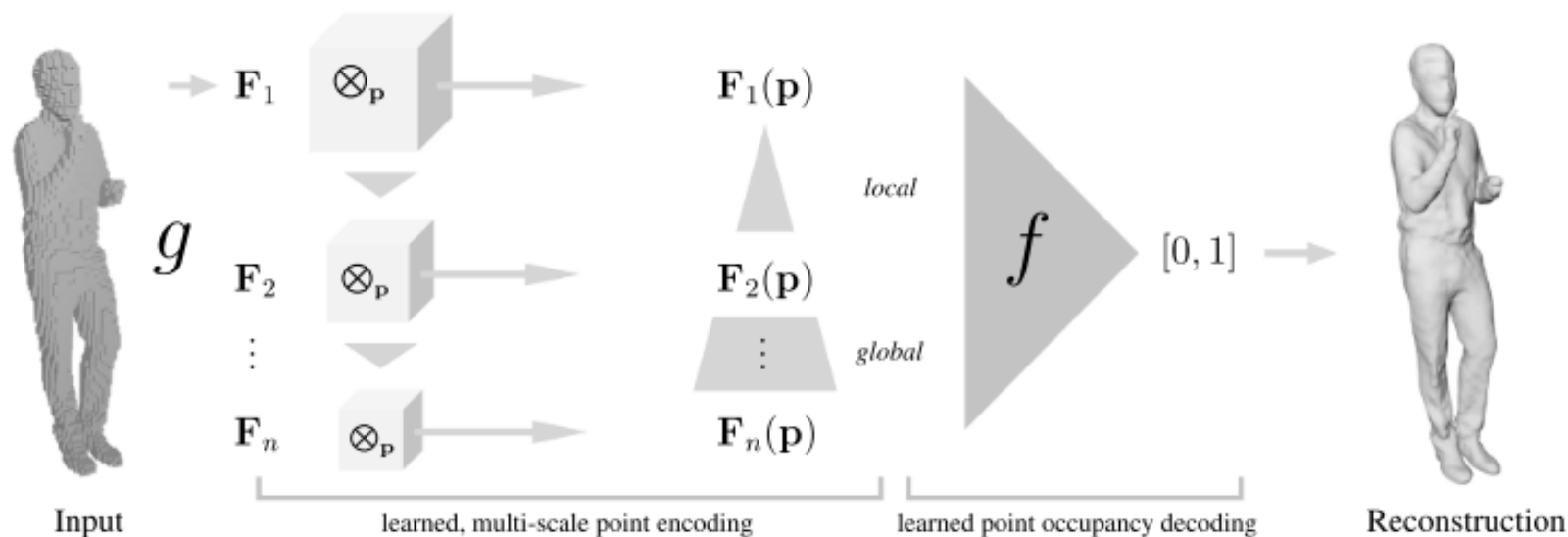
- ❑ 以点的3D 坐标作为输入 会让网络忽略学习整体形状框架，反而更多注意几个典型坐标点上。
- ❑ 用一个global vector代表原物体特征 会损失原始输入的一些细节。
- ❑ → 用3D 卷积获得 多尺度 全局和局部特征 表示输入的形状框架

Ref: Chibane, Julian, Thiemo Alldieck, and Gerard Pons-Moll. "Implicit functions in feature space for 3d shape reconstruction and completion." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020.

2020/7/13

40

网络结构



编码器部分

$$g(\mathbf{X}) := \mathbf{F}_1, \dots, \mathbf{F}_n .$$

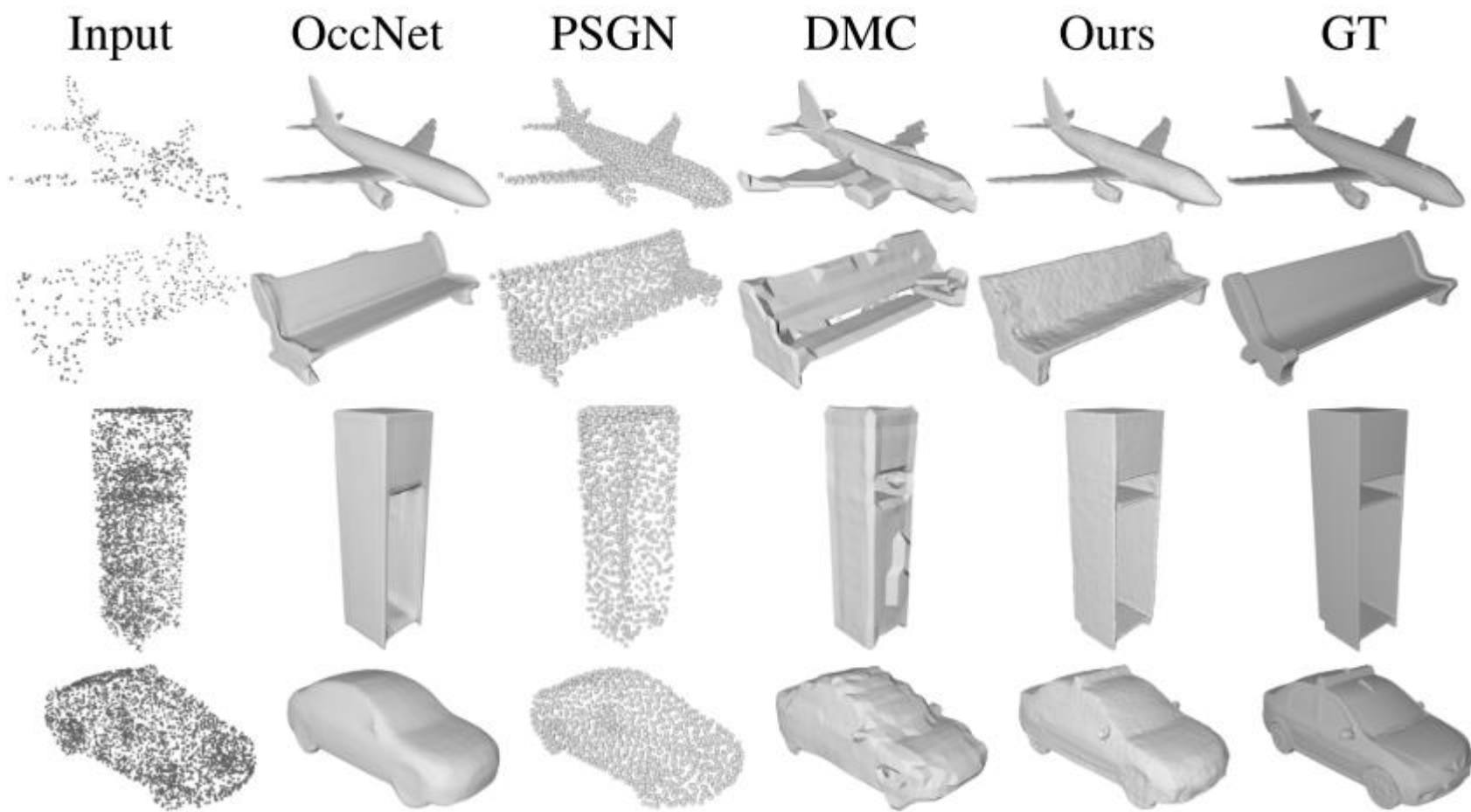
编码器获得多尺度特征 $\mathbf{F}_1, \dots, \mathbf{F}_n$

解码器部分

$$f(\mathbf{F}_1(\mathbf{p}), \dots, \mathbf{F}_n(\mathbf{p})) : \mathcal{F}_1 \times \dots \times \mathcal{F}_n \mapsto [0, 1]$$

$$\{\mathbf{p} + a \cdot \mathbf{e}_i \cdot d \in \mathbb{R}^3 | a \in \{1, 0, -1\}, i \in \{1, 2, 3\}\},$$

结果



2020/7/13

42

结果

