

## 三维人脸形状的分离表示学习

CVPR 2019 中国科学技术大学 Zi-Hang Jiang, Qianyi Wu, Keyu Chen, Juyong Zhang

一、

问题：3D 人脸形状由 identity part and expression part 组成，非线性方式编码

针对以上问题，提出了一个 3D 面部网格的属性分解框架

本文试图建立一个 3D 人脸形状的分离参数空间，存在两个问题：

如何设计学习方法并不明确；

如何充分利用已有的 identity and expression 标签来进行数据集的训练也是一大问题

二、

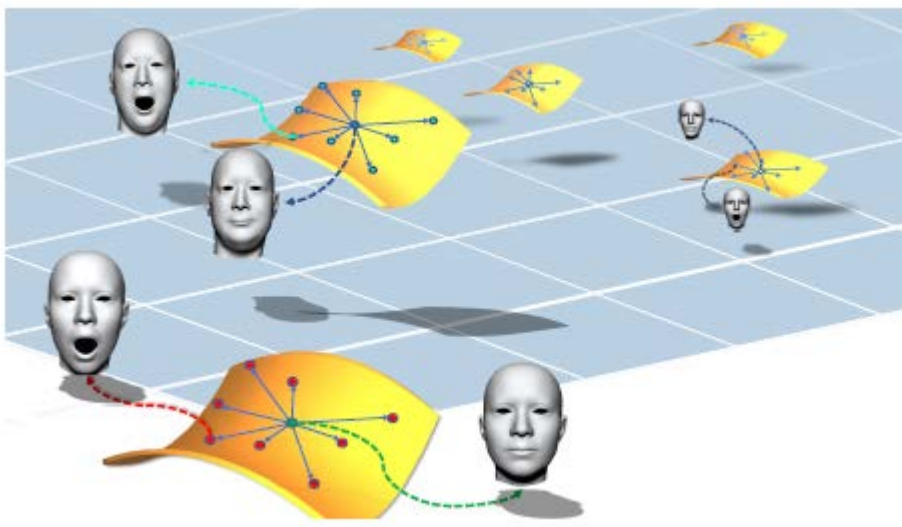
论文贡献：

- 1) learn a disentangled latent space for 3D face shape （3D 人脸的分离潜在空间）允许在 identity and expression domains 进行编辑
- 2) novel framework for the disentangling task defined on 3D face mesh; Vertex-based deformation representation better;
- 3) much better results in disentangling identity and expression, applications like expression transfer

三、

Disentangled 3D Face Representation

人类表达位于高维流形中, 原点：中性表达；identity attribute expression on mean face represents her/his expression attribute



disentangled 3D face representation 包括两部分：分解和融合网络。分解网络通过将输入面网

格解耦为 identity mesh and expression mesh 来分解属性。融合网络从 identity mesh and expression mesh 中恢复原始面部网格。

### 谱图卷积

采用图卷积算子，谱图卷积，用于在网格上提取有用的顶点特征。

### 变形表示

谱图卷积是拉普拉斯平滑的一种特殊形式。由于不同面网格之间的主要差异主要是由非刚性变形引起的，采用变形表示（DR）来模拟 3D 网格。选择平均面的中性表达作为参考网格，其他被视为变形网格。

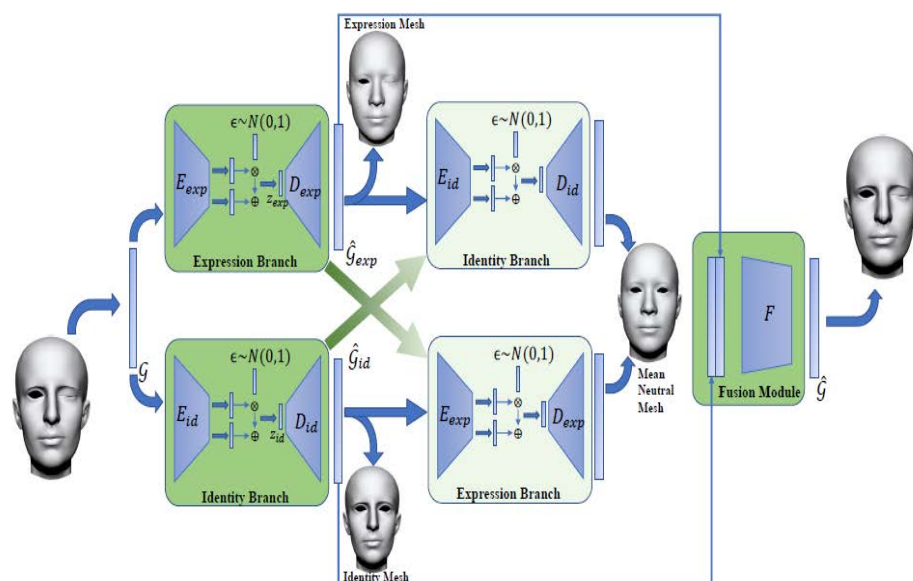
### 分解网络

输入是三维人脸网格的变形表示特征  $G$ ，目标是将其分解为 identity and expression attributes。它相当于将输入 meshM 映射到另外两个三元组元素（Mid; Mexp）。

分解部分包括两个结构相同的并行网络，一个用于提取表达式网格 Mexp，另一个用于提取身份 meshMid。

输入将通过若干光谱图卷积层进行网格特征提取，完全连接层的瓶颈架构作为编码器 - 解码器结构。应用该结构以获得潜在的身份表示。输出应接近 Mid 的 DR 功能。

瓶颈层用作相应属性的新紧凑参数空间



### Fusion Network

从分解的 identity and expression attributes 重建原始输入

融合模块的目标是将具有  $R^{2d}$  中的顶点特征的这个新图转换为具有  $R^d$ （原始输入）中的顶点特征的同构图。还将光谱图卷积与激活层一起应用以实现此目标。

### 训练过程

以端到端的策略训练整个网络。

数据增强:

FaceWare-House dataset: 150 identities and 47 expressions for each identity

Identities 数量很小, 有过拟合风险——新方法: 插值

#### 四、实验:

##### 1、实施细节:

##### 2、评估指标:

1) Reconstruction Measurement : 平均顶点距离; 感知错误

2) Decomposition Measurement : 为了测量三维人脸形状的分离表示, 从具有相同 identity 和不同表情的模型中提出了重构 identity 网格的度量, 并且从具有不同 identity 和相同表情的模型中表达网格

##### 3、Experiments on FaceWareHouse:

140 训练, 10 个测试

1) Baseline Comparison: 双线性模型; FLAME; MeshAE

Method	$E_{avd}$		$E_{sed}$		$E_{id}$		$E_{exp}$	
	Mean Error	Median	Mean Error	Median	Mean Error	Median	Mean Error	Median
Bilinear [13]	0.993	0.998	0.0243	0.0183	0.477	0.472	0.527	0.484
FLAME [25]	0.882	0.905	0.0144	0.0074	0.329	0.328	0.711	0.630
MeshAE [31]	0.825	0.811	0.0151	0.0777	-	-	-	-
Ours w/o DR & Fusion	0.981	1.292	0.177	0.0938	0.395	0.380	0.170	0.160
Ours w/o DR	0.939	0.836	0.447	0.388	0.446	0.463	0.0992	0.0750
Ours w/o Fusion	0.661	0.579	<b>0.00283</b>	<b>0.0000</b>	0.183	0.178	0.0582	0.0494
Ours	<b>0.472</b>	<b>0.381</b>	0.00333	<b>0.0000</b>	<b>0.121</b>	<b>0.121</b>	<b>0.0388</b>	<b>0.0267</b>

Table 1. Quantitative results on Facewarehouse. All number were in millimeters. DR: deformation representation; Fusion: fusion module.

##### 2) Ablation Study (消融研究):

Tab 1

新的基于顶点的变形表示用于 3D 人脸形状 vs 直接使用欧几里德坐标方法

4、Experiment on COMA Dataset: COMA 数据集用于构建非线性 3D 人脸表示, 同时将整个 3D 人脸形状编码并解码为潜在空间中的一个向量而不考虑身份和表情属性

表情外推实验中的方法与 FLAME 进行比较, 并在表 2 中的所有 12 个交叉验证实验中报告平均顶点距离。可以观察到, 与现有技术的 FLAME 方法相比, 我们的方法获得了更好的泛化结果。关于外推实验, 所有 12 个表达式外推交叉验证实验都在补充中给出。

Average error	Mean Error	Median Error
FLAME [25]	2.001	1.615
Ours	<b>1.643</b>	<b>1.536</b>

Table 2. Extrapolation results on COMA dataset. All results are in millimeters.

#### 5、Discussion on Larger Dataset:

我们通过非刚性配准将 Bosphorus 转换为网格并与 FaceWareHouse (FWH) 结合, 创建更大的数据集。

三种不同的训练数据集上评估我们的方法: 原始的 FaceWareHouse, FWH 和 Bosphorus 的组合, 以及 DR 增强的 FWH。

Dataset	$E_{avd}$	$E_{sed}$	$E_{id}$	$E_{exp}$
Original FWH	18.3/18.0	0.05/0.03	1.4/1.4	0.5/0.3
Combination	16.9/16.6	0.06/0.03	1.6/1.6	0.5/0.4
DR-augmented	4.7/3.8	0.03/0.00	1.2/1.2	0.4/0.3

Table 3. More quantitative results. Table gives our results on different datasets: original FWH, combination of Bosphorus and FWH (Combination) and our DR-augmented FWH. All number in 0.1 millimeters.

## 五、应用

### 1、Expression Transfer:

对于双线性模型, 我们首先求解参考模型的身份和表情参数, 然后将表情参数从源面转移到目标面。在我们的方法中, 我们直接将源面的潜在代码应用于目标面。

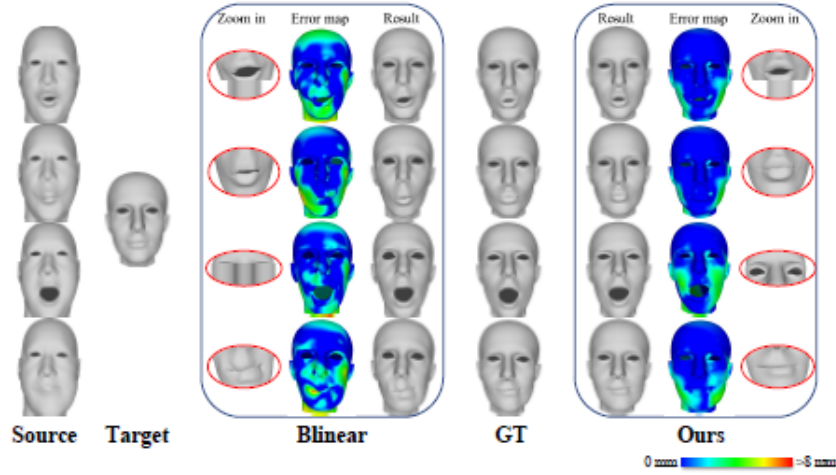


Figure 5. Expression transfer application. Comparing to the bi-linear model, our method achieves more natural and stable visual results.

## 2、Latent space interpolation (潜在空间插值)

插值操作应用于潜码，并且使用训练的解码器从生成的代码恢复模型。

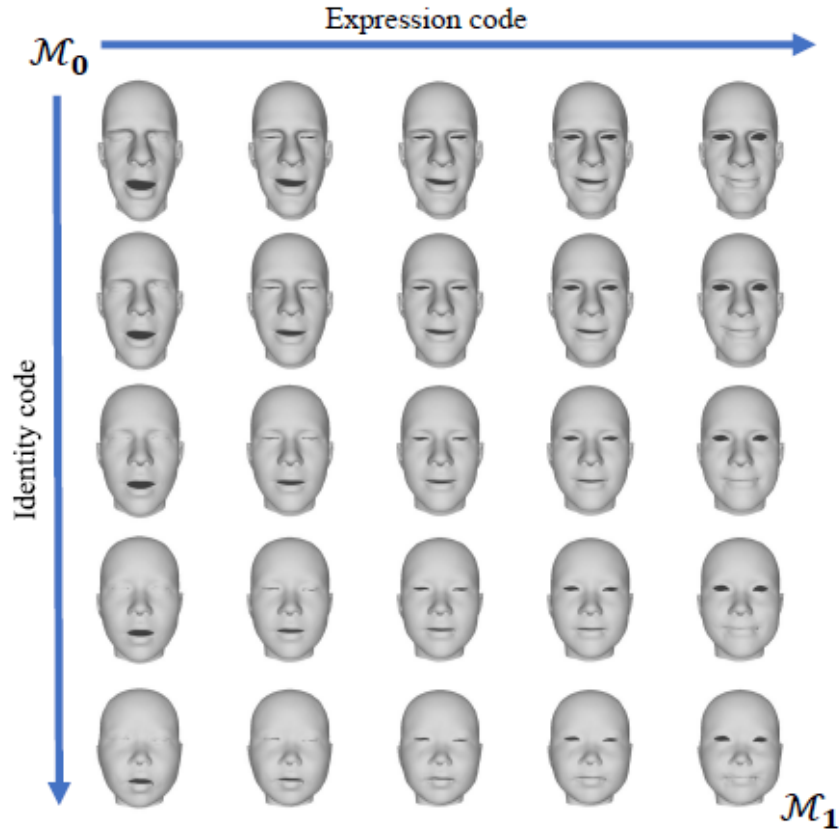


Figure 4. Exploring interpolation results on latent space. Based on our method, we can obtain identity and expression code for two 3D face model  $\mathcal{M}_0$  and  $\mathcal{M}_1$ , and we interpolate latent identity and expression vectors individually, in stride of 0.25.

## 六、结论

给定的三维人脸形状可以准确地分解为身份部分和表情部分；提出了一个精心设计的框架来训练分解网络和融合网络。

为了更好地表示非刚性变形空间，输入面形状表示为基于顶点的变形表示而不是欧几里德坐标。

我们通过消融研究和广泛的定量和定性实验证明了所提方法的有效性。

与传统方法相比，基于我们解开表示的表达转移等应用表现出更自然，更准确的结果。