3D Face Reconstruction and Dense Alignment

段永杰

2020/9/25

背景

- 从单张图片进行人脸重建与人脸对齐
  - 需要先验信息
  - 3D deformable model (3DMM)[1]：shape manifold

1. V. Blanz and T. Vetter. “A morphable model for the synthesis of 3D faces”, ACM Conf. on Computer Graphics and Interactive Techniques (SIGGRAPH), 1999

背景

- 3DMM
  - 表示
    - 形状向量：$\mathrm{S_i} - (X_i, Y_i, Z_i)$
    - 纹理向量：$\mathrm{T_i} - (R_i, G_i, B_i)$

$$\mathbf{S}_{mod} = \sum_{i=1}^{m} a_i \mathbf{S}_i, \quad \mathbf{T}_{mod} = \sum_{i=1}^{m} b_i \mathbf{T}_i, \quad \sum_{i=1}^{m} a_i = \sum_{i=1}^{m} b_i = 1$$

  - 新的形状与纹理为上述的线性组合
  - An average component + a set of principal components
  - 类似PCA，计算线性组合参数

$$\mathbf{S}_{mod} = \bar{S} + \sum_{i=1}^{m-1} \alpha_i s_i, \quad \mathbf{T}_{mod} = \bar{T} + \sum_{i=1}^{m-1} \beta_i t_i$$

$$p(\vec{\alpha}) \sim \exp\left[-\frac{1}{2}\sum_{i=1}^{m-1}(\alpha_i/\sigma_i)^2\right]$$

背景

- 3DMM
  - 将统计模型与二维图片进行匹配，计算参数
  - Analysis-by-synthesis，病态问题
    - 根据现有模型参数产生带纹理的3D人脸
    - 根据现有相机等渲染参数渲染一张2D图片
    - 根据原始图片与渲染图片的残差更新参数
  - 优化过程复杂，计算量较大
    - 渲染参数：相机位置、相机角度、光照位置、光照角度…

背景

- 3DMM
  - 改进
    - 加入特征点信息进行约束
    - 利用神经网络直接预测参数
  - 限制
    - 受限于3D模型的模板质量——姿态变化大、人脸复杂表情
    - 3D模块之间的点集对应关系
    - 涉及到透视投影、3D TPS等操作，增加计算复杂度
    - 背景干扰以及遮挡会对精度产生影响
    - 对初始条件敏感

背景

- 端到端预测
  - 直接回归68个人脸特征点位置
    - 需要额外估计深度信息
    - 没有提供稠密对齐
  - 体积表示法
    - 3D体素表示人脸
    - 丢失了点的语义信息
    - 人脸在3D空间中只占小部分
    - 存在大量冗余计算与估计，重建分辨率较低

Face Alignment Across Large Poses: A 3D Solution

Xiangyu Zhu[1] Zhen Lei[1*] Xiaoming Liu[2] Hailin Shi[1] Stan Z. Li[1]
[1] Center for Biometrics and Security Research & National Laboratory of Pattern Recognition,
Institute of Automation, Chinese Academy of Sciences,
[2]Department of Computer Science and Engineering, Michigan State University
{xiangyu.zhu,zlei,hailin.shi,szli}@nlpr.ia.ac.cn liuxm@msu.edu

CVPR 2016

摘要

- 问题特点
  - Small-medium pose：转动角度小于等于45度
  - Large pose：转动角度大于45度（极端情况：只有侧脸信息）

  - 无法保证所有特征点可见
  - 无法保证不可见特征点的标记准确程度

方法

- 3DMM
  - 添加表情因素
  - A_{id}：形状主成分
  - A_{exp}：表情主成分

$$\mathbf{S} = \overline{\mathbf{S}} + \mathbf{A}_{id}\boldsymbol{\alpha}_{id} + \mathbf{A}_{exp}\boldsymbol{\alpha}_{exp}$$

方法

- 参数估计
  - 级联回归 + CNN
  - 输入：图像，当前参数下的PNCC的结果

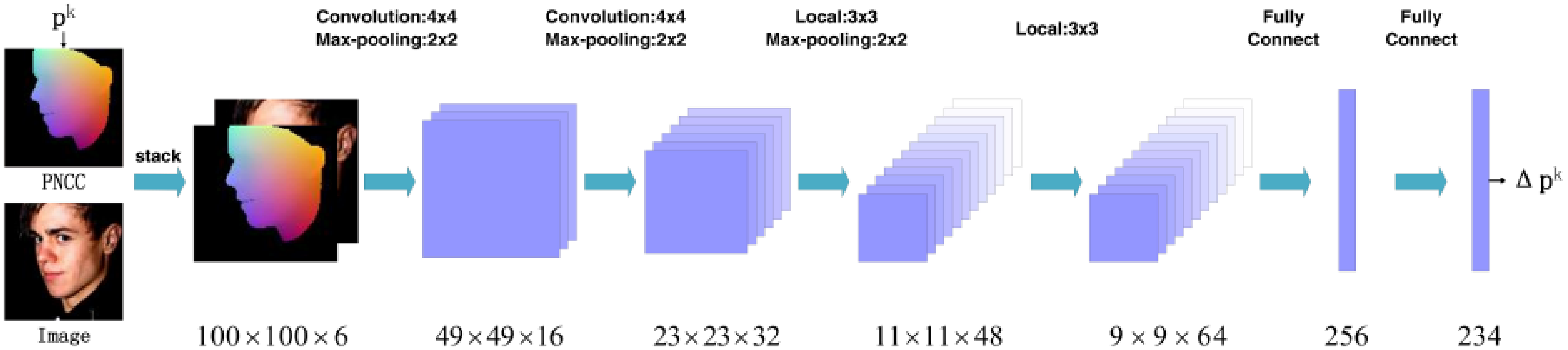
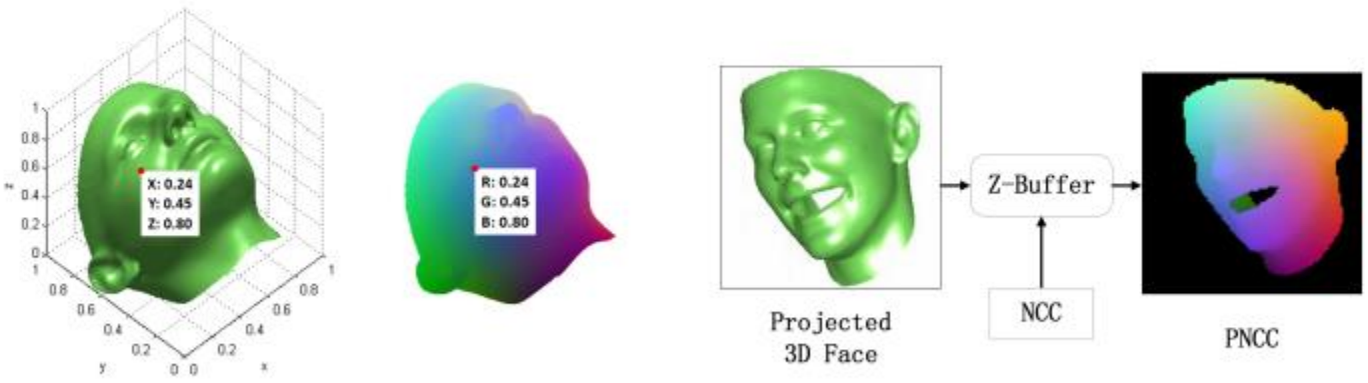


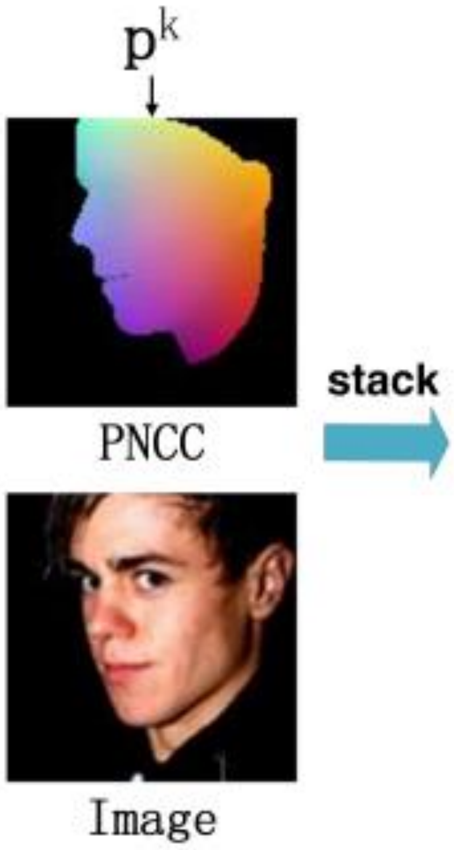
Figure 2. An overview of 3DDFA. At kth iteration, Net^k takes a medium parameter $\mathbf{p}^k$ as input, constructs the projected normalized coordinate code (PNCC), stacks it with the input image and sends it into CNN to predict the parameter update $\Delta\mathbf{p}^k$.

方法

- 参数估计
  - NCC：3D平均人脸的顶点坐标和 RGB 取值进行归一化[0,1]
  - PNCC：根据当前参数p，使用3DMM得到3D人脸模型，向2D平面投影并渲染



(a) NCC (b) PNCC



方法

- 损失函数
  - PDC：parameter distance cost，预测参数与真实参数之间的距离

$$E_{pdc} = \|\Delta p - (p^g - p^k)\|^2$$

  - VDC：vertex distance cost，预测模型点与真实模型点之间的距离

$$E_{vdc} = \|V(\mathbf{p}^0 + \Delta\mathbf{p}) - V(\mathbf{p}^g)\|^2$$

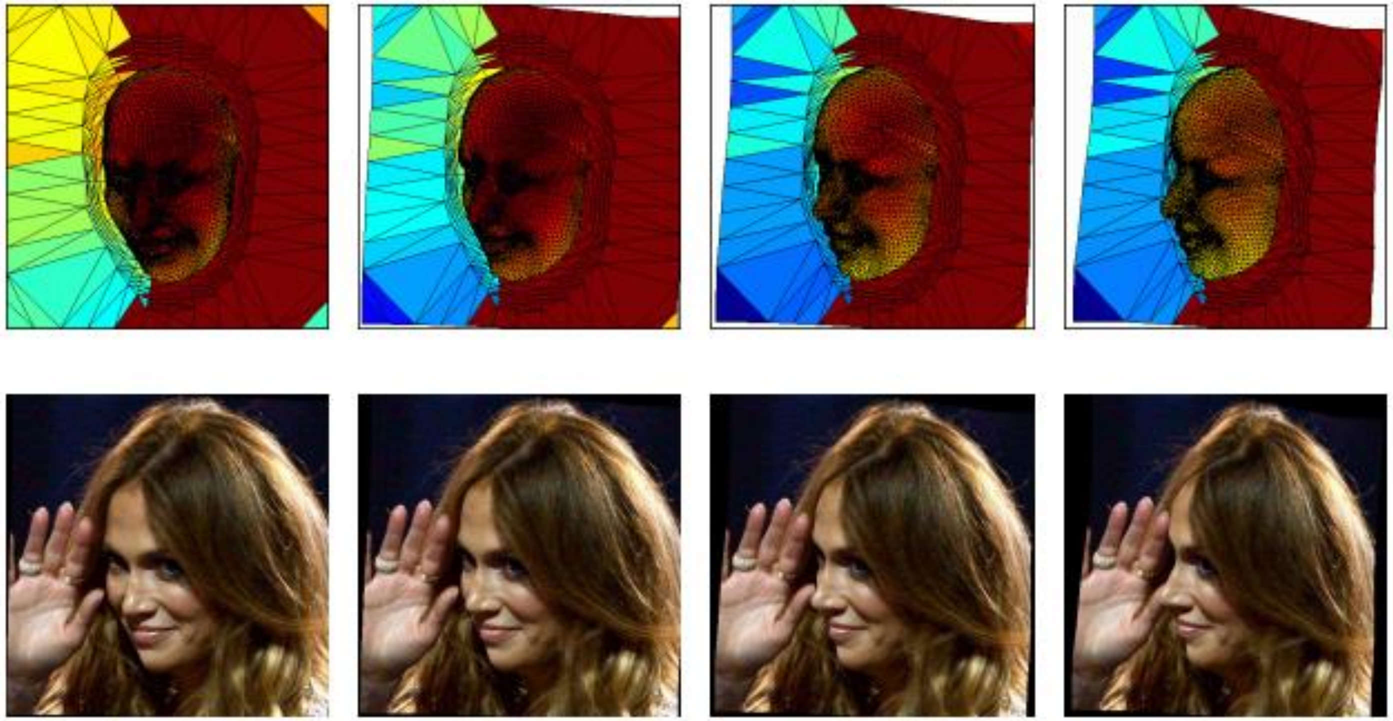
  - WPDC：weighted parameter distance cost，对不同的参数设计不同的权重

$$\begin{aligned}
&E_{wpdc} = (\Delta\mathbf{p} - (\mathbf{p}^g - \mathbf{p}^0))^T \mathbf{W} (\Delta\mathbf{p} - (\mathbf{p}^g - \mathbf{p}^0)) \\
&where \quad \mathbf{W} = diag(w_1, w_2, ..., w_n) \\
&w_i = \|V(\mathbf{p}^d(i)) - V(\mathbf{p}^g)\| / \sum w_i \\
&\mathbf{p}^d(i)_i = (\mathbf{p}^0 + \Delta\mathbf{p})_i \\
&\mathbf{p}^d(i)_j = \mathbf{p}^g_j, \quad j \in \{1, \ldots, i-1, i+1, \ldots, n\}
\end{aligned}$$

计算第i个参数带来的误差，此时除去第i个参数，其余的参数使用真实值

方法

- 数据集
  - 合成 large pose 的数据
  - 300W-LP
    - 从数据集300W 中合成
    - 300W由多个数据集组合而成
    - 合成 61225 张 large pose 的图片
    - 称为300W-LP



(a) (b) (c) (d)

Figure 5. 2D and 3D view of the image rotation. (a) The original yaw angle yaw_0. (b) $yaw_0+20°$. (c) $yaw_0+30°$. (d) $yaw_0+40°$.

实验

- 数据集
  - 300W-LP：含有68个关键点标注
    - 训练：97967
    - 测试：24483

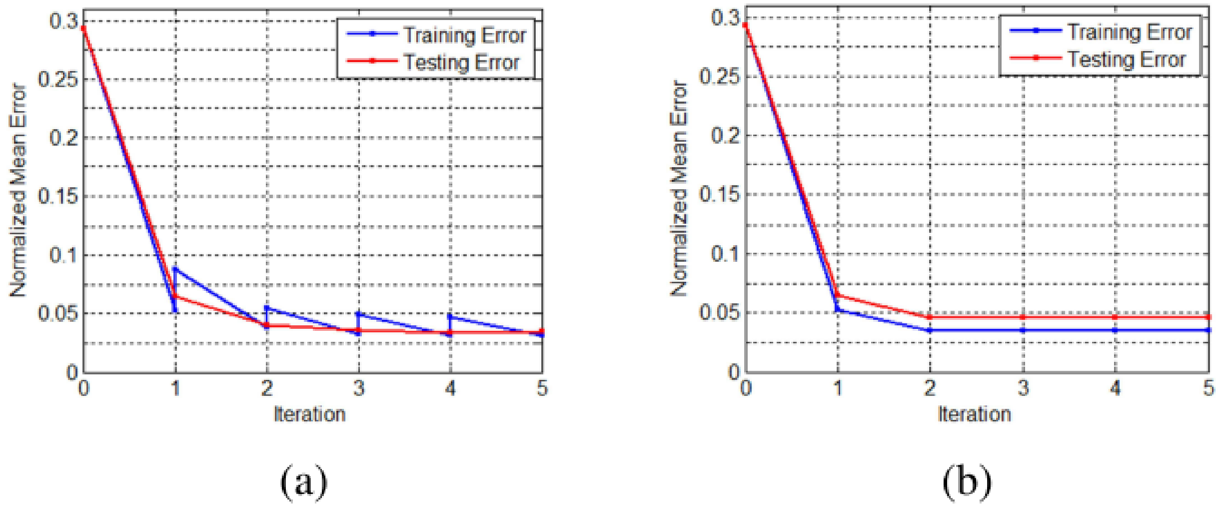


Figure 7. The training and testing errors with (a) and without (b) initialization regeneration.

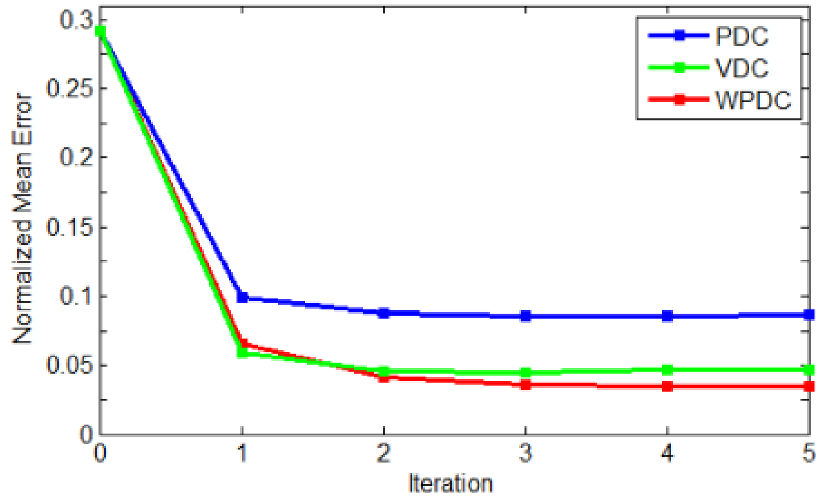


Figure 8. The testing errors with different cost function.

方法

- 数据集
  - AFLW：21080 大范围角度人脸图片 + 21 可见关键点
  - AFLW2000-3D：2000 张AFLW图片 + 3DMM系数 + 68个人脸特征点

Normalized Mean Error(NME)

Table 1. The NME(%) of face alignment results on AFLW and AFLW2000-3D with the first and the second best results highlighted. The bracket shows the training set. The results of provided alignment models are marked with their references.

	AFLW Dataset (21 pts)					AFLW2000-3D Dataset (68 pts)				
Method	[0, 30]	[30, 60]	[60, 90]	Mean	Std	[0, 30]	[30, 60]	[60, 90]	Mean	Std
CDM [49]	8.15	13.02	16.17	12.44	4.04	-	-	-	-	-
RCPR [7]	6.16	18.67	34.82	19.88	14.36	-	-	-	-	-
RCPR(300W)	5.40	9.80	20.61	11.94	7.83	4.16	9.88	22.58	12.21	9.43
RCPR(300W-LP)	5.43	6.58	11.53	7.85	3.24	4.26	5.96	13.18	7.80	4.74
ESR(300W)	5.58	10.62	20.02	12.07	7.33	4.38	10.47	20.31	11.72	8.04
ESR(300W-LP)	5.66	7.12	11.94	8.24	3.29	4.60	6.70	12.67	7.99	4.19
SDM(300W)	**4.67**	6.78	16.13	9.19	6.10	**3.56**	7.08	17.48	9.37	7.23
SDM(300W-LP)	**4.75**	5.55	9.34	6.55	2.45	3.67	4.94	9.76	6.12	3.21
3DDFA	5.00	**5.06**	**6.74**	**5.60**	**0.99**	3.78	**4.54**	**7.93**	**5.42**	**2.21**
3DDFA+SDM	**4.75**	**4.83**	**6.38**	**5.32**	**0.92**	**3.43**	**4.24**	**7.17**	**4.94**	**1.97**

SDM：得到特征到偏移量的映射，通过梯度下降计算

实验

- 耗时
  - PNCC：17.49ms (CPU)
  - CNN计算偏移量：7.75ms (TITAN GPU)

  - 一共3次迭代：25.24ms * 3

Joint 3D Face Reconstruction and Dense Alignment with Position Map Regression Network

Yao Feng[1 [0000−0002−9481−9783]], Fan Wu[2 [0000−0003−1970−3470]], Xiaohu Shao[3,4 [0000−0003−1141−6020]], Yanfeng Wang[1 [0000−0002−3196−2347]], and Xi Zhou[1,2 [0000−0003−2917−0436]]

[1] Cooperative Medianet Innovation Center, Shanghai Jiao Tong University
[2] CloudWalk Technology
[3] CIGIT, Chinese Academy of Sciences
[4] University of Chinese Academy of Sciences

ECCV 2018

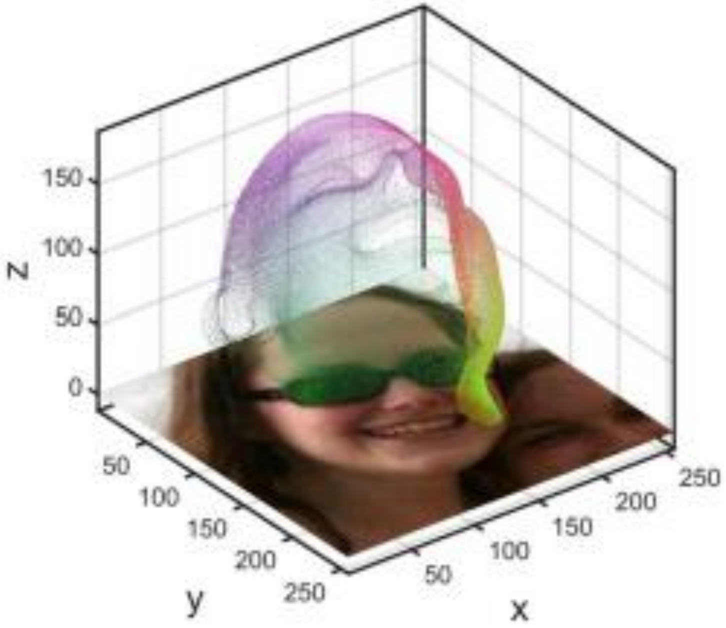
摘要

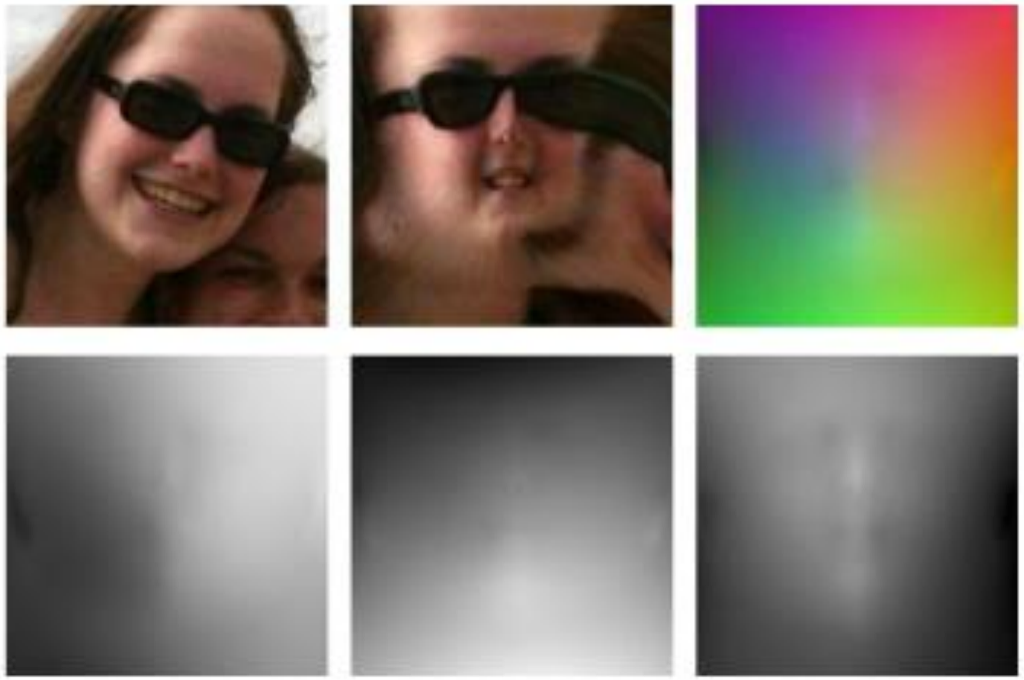
- 端到端
  - 同时完成稠密的人脸对齐与人脸重建
  - 速度快，单张人脸的估计任可以达到100FPS

方法

- 合理的模型表示
  - 保留模型点的语义信息以及相邻点之间的空间信息
  - 重建分辨率要高，计算量要小

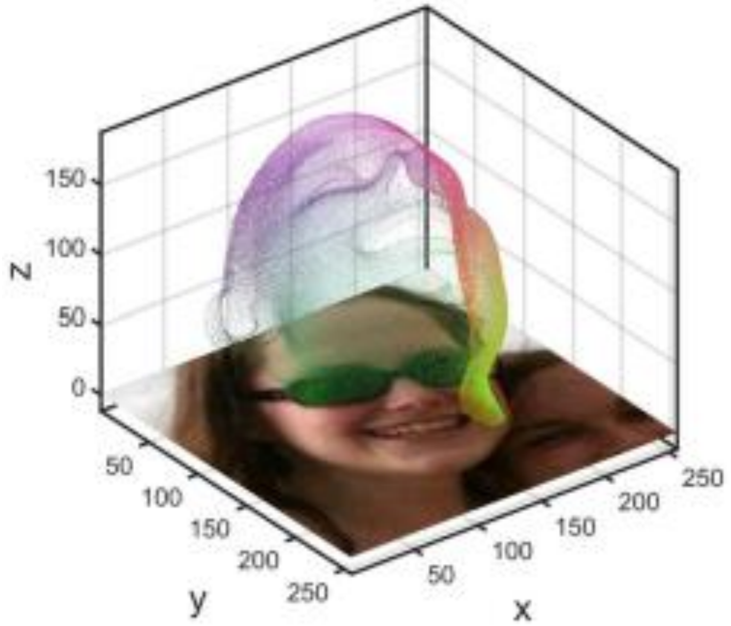
  - UV position map



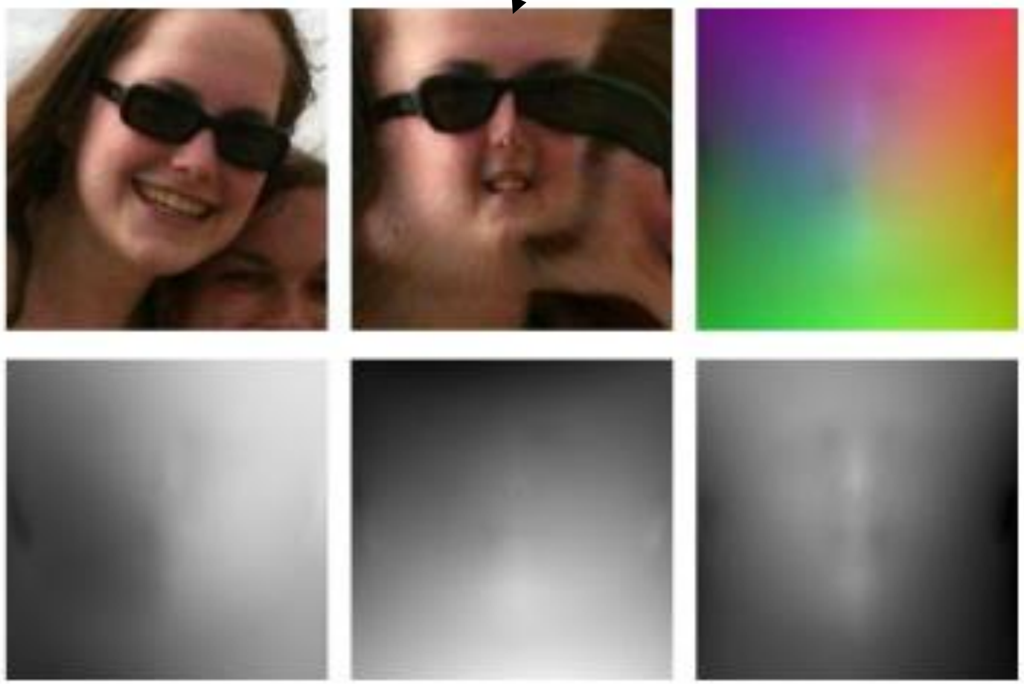


方法

- UV position map
  - 和3DMM类模型类比



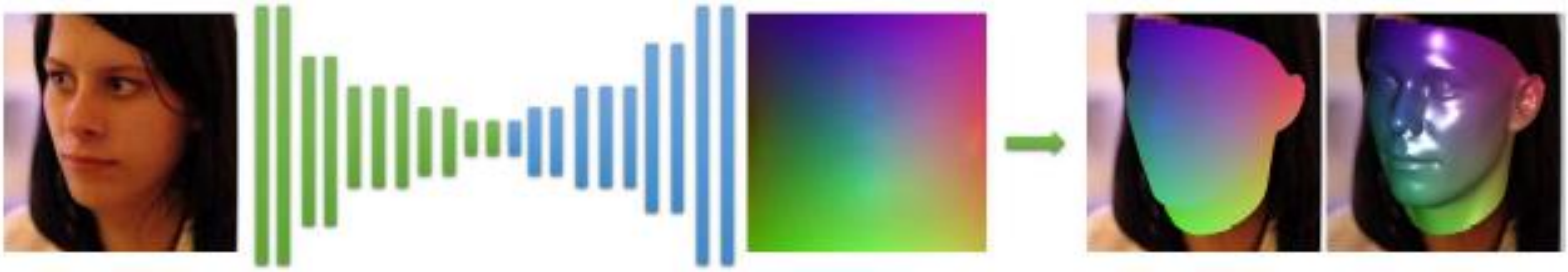
纹理信息



形状信息

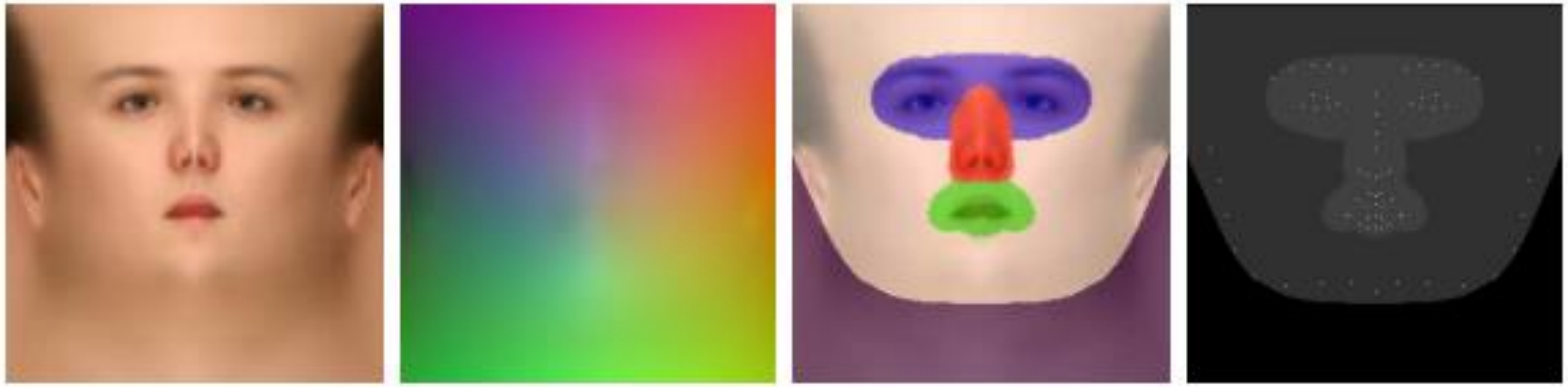
方法

- UV position map
  - 和3DMM类模型类比
  - 256 × 256，约65K个点表示人脸
  - 编解码 + 残差结构



方法

- UV position map
  - 和3DMM类模型类比
  - 256 × 256，约65K个点表示人脸
  - 编解码 + 残差结构
  - 分区域关注，损失函数权重——16:4:3:0
    - 68个人脸特征点
    - 鼻子、眼睛、嘴巴区域
    - 人脸其他部分，如脖子等区域



方法

- UV position map
  - 和3DMM类模型类比
  - 256 × 256，约65K个点表示人脸
  - 编解码 + 残差结构
  - 分区域关注，损失函数权重——16:4:3:0
    - 68个人脸特征点
    - 鼻子、眼睛、嘴巴区域
    - 人脸其他部分，如脖子等区域

$$Loss = \sum \| Pos(u,v) - \tilde{Pos}(u,v) \| \cdot W(u,v)$$

实验

- 数据集
  - 现有数据集只有3DMM的标注，没有 UV position map 的标注
  - 训练数据集：300W-LP
    - 包含不同角度的人脸图像以及3DMM系数
    - 根据3DMM系数合成UV position map
  - 数据增广
    - 旋转、平移、缩放、颜色通道缩放、模拟遮挡

实验

- 测试数据集
  - AFLW2000-3D
    - 2000 张AFLW图片 + 3DMM系数 + 68个人脸特征点，分析人脸重建及对齐性能

Fig. 6: Examples from AFLW2000-3D dataset show that our predictions are more accurate than ground truth in some cases. Green: predicted landmarks by our method. Red: ground truth from [67].

实验

- 测试数据集
  - AFLW2000-3D
    - 2000 张AFLW图片 + 3DMM系数 + 68个人脸特征点，分析人脸重建及对齐性能

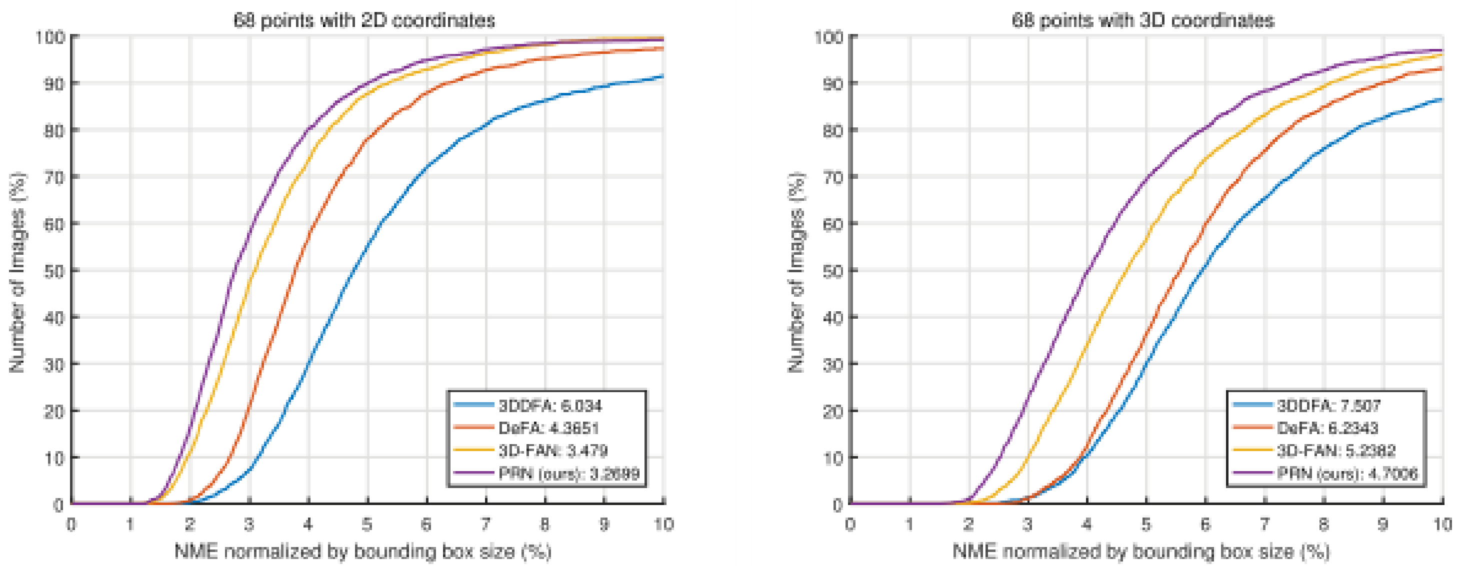


Fig. 5: Cumulative Errors Distribution (CED) curves on AFLW2000-3D. Evaluation is performed on 68 landmarks with both the 2D(left) and 3D(right) coordinates. Overall 2000 images from AFLW2000-3D dataset are used here. The mean NME% of each method is also showed in the legend.

关键点之间的误差

Normalized Mean Error(NME)

实验

- 测试数据集
  - AFLW2000-3D
    - 2000 张AFLW图片 + 3DMM系数 + 68个人脸特征点，分析人脸重建及对齐性能

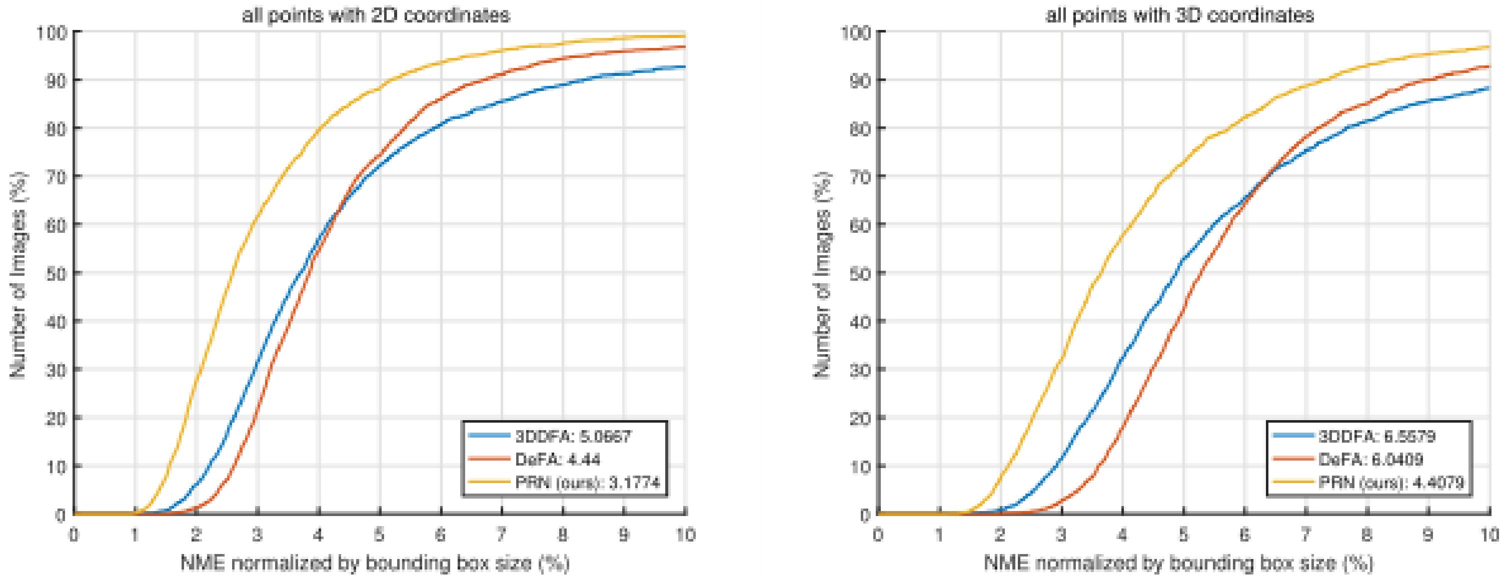


Fig. 7: CED curves on AFLW2000-3D. Evaluation is performed on all points with both the 2D (left) and 3D (right) coordinates. Overall 2000 images from AFLW2000-3D dataset are used here. The mean NME% is showed in the legend.

最大可见区域内人脸所有点之间的误差

Normalized Mean Error(NME)

实验

- 测试数据集
  - AFLW-LFPA
    - 1299 张AFLW图片 + 34个可见的人脸特征点，分析人脸对齐性能

Table 1: Performance comparison on AFLW2000-3D(68 landmarks) and AFLW-LFPA(34 visible landmarks). The NME (%) for faces with different yaw angles are reported. The first best result in each category is highlighted in bold, the lower is the better.

	AFLW2000-3D				AFLW-LFPA
Method	0 to 30	30 to 60	60 to 90	Mean	Mean
SDM[60]	3.67	4.94	9.67	6.12	-
3DDFA [67]	3.78	4.54	7.93	5.42	-
3DDFA + SDM [67]	3.43	4.24	7.17	4.94	-
PAWF[32]	-	-	-	-	4.72
Yu et al. [63]	3.62	6.06	9.56	-	-
3DSTN[4]	3.15	4.33	5.98	4.49	-
DeFA[40]	-	-	-	4.50	3.86
PRN (ours)	**2.75**	**3.51**	**4.61**	**3.62**	**2.93**

实验

- 测试数据集
  - Florence
    - 53个3D扫描人脸模型，在不同角度下渲染2D图片，分析人脸重建性能

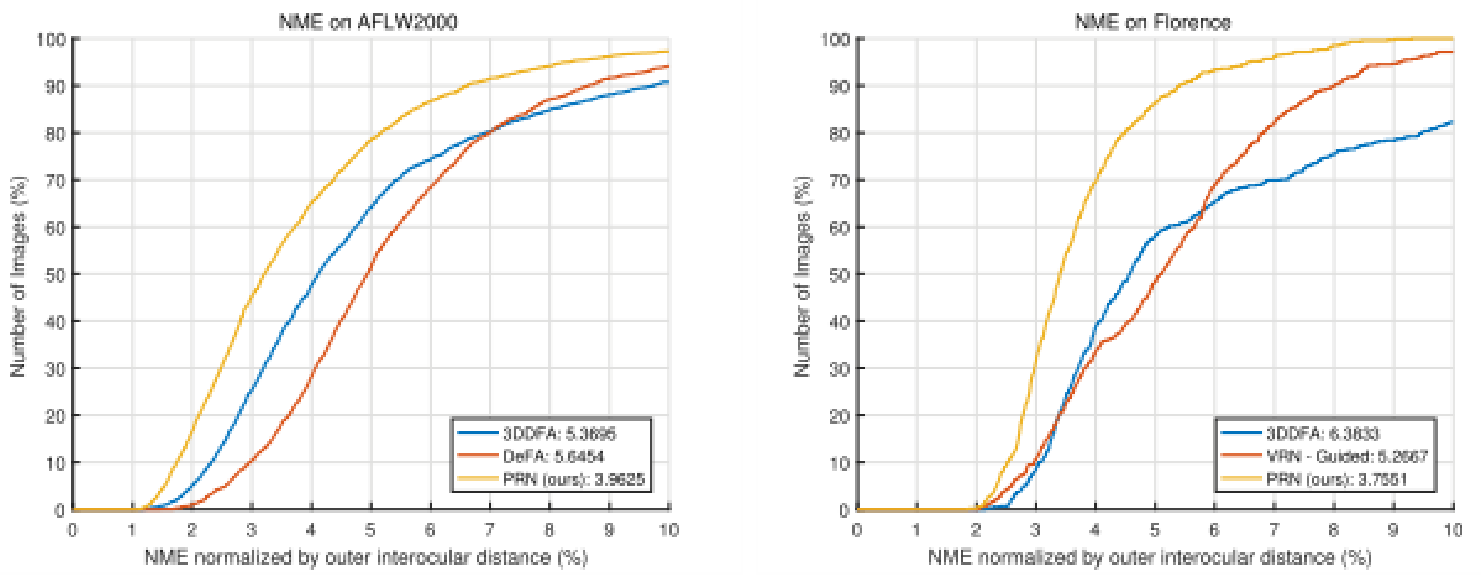


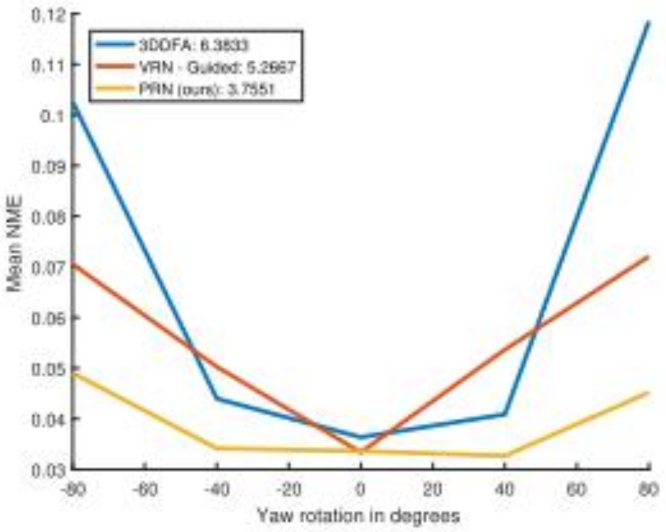
Fig. 8: 3D reconstruction performance(CED curves) on in-the-wild AFLW2000-3D dataset and Florence dataset. The mean NME% of each method is showed in the legend. On AFLW2000-3D, more than 45K points are used for evaluation. On Florence, about 19K points are used.

3D mesh点之间的误差

Normalized Mean Error(NME)

实验

- 测试数据集
  - Florence
    - 53个3D扫描人脸模型，在不同角度下渲染2D图片，分析人脸重建性能



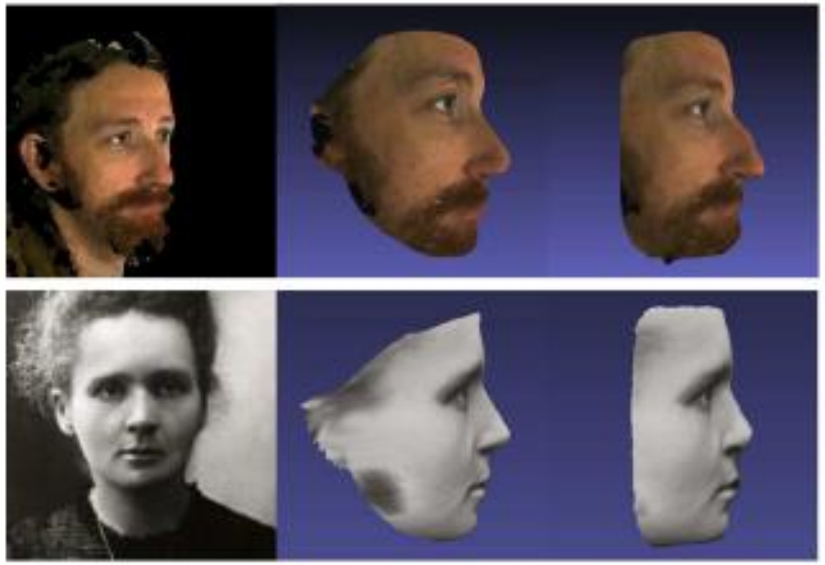
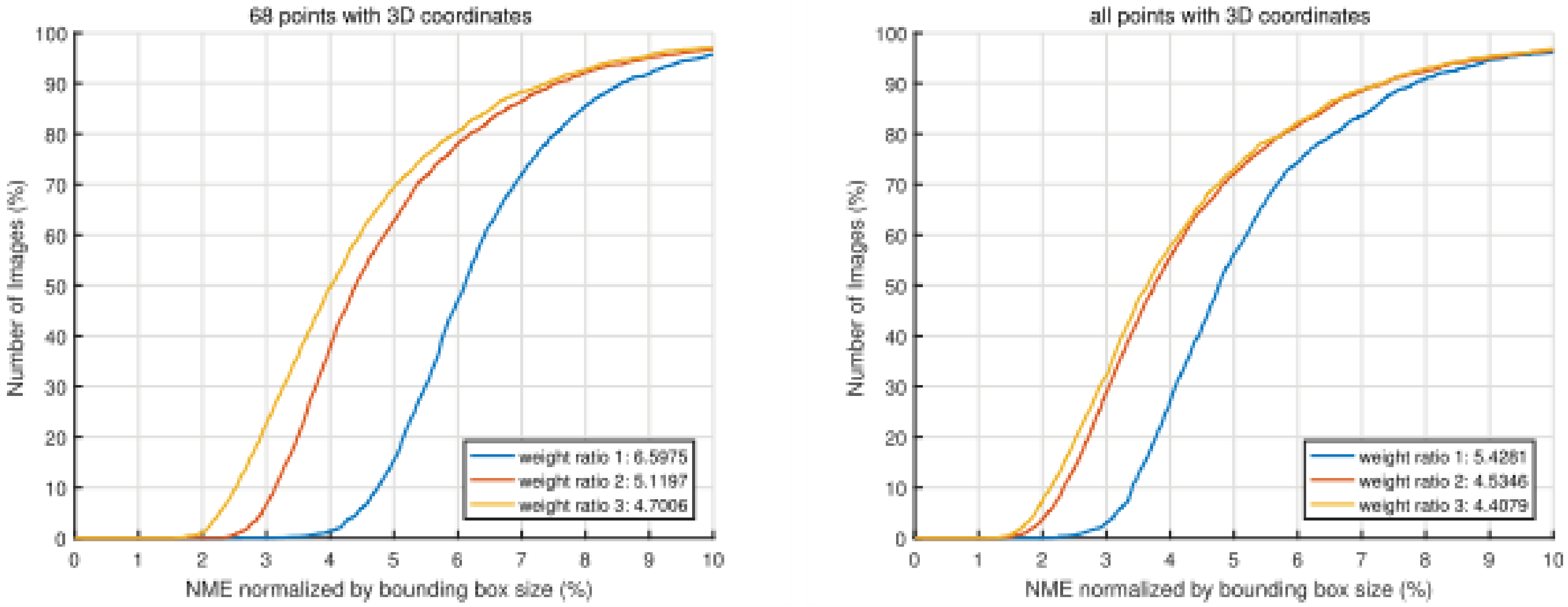


Fig. 9: Left: CED curves on Florence dataset with different yaw angles. Right: the qualitative comparison with VRN-Guided. The first column is the input images from Florence dataset and the Internet, the second column is the reconstructed face from our method, the third column is the results from VRN.

实验

- 消融实验
  - 主要探究不同区域的权重的性能的影响



Weight ratio 1： 1:1:1:1
Weight ratio 2： 1:1:1:0
Weight ratio 3： 16:4:3:0

Fig. 10: The effect of weight mask evaluated on AFLW2000-3D dataset with 68 landmarks(left) and all points(right).

实验

- 耗时

Table 2: Run time in Milliseconds per Image

3DDFA[67]	DeFA[40]	3D-FAN[9]	3DSTN[4]	VRN-Guided[28]	PRN (ours)
75.7	35.4	54.7	19.0	69.0	9.8