人脸重建相关论文整理

Xinyuan

October 15, 2021

1 单视角 Model Base

1.1 3DDFA

2 单视角 Model Free

2.1 Unsupervised Learning of Probably Symmetric Deformable 3D Objects from Images in the Wild

本文精彩之处在于利用人脸等三维物体的对称性进行三维估计。我认为作者的初衷是认为脱离了人脸几何先验（33DMM）之后，需要对不同的样本找到对齐的方式（**特别是对于无监督的方法，如果需要用到神经网络统计的方式对输入图片进行分析，需要引入一定的先验信息，为不同的样本找到一致性。这一点也可以在监督的网络里面由 gt 引入。**）作者本着为不同样本找到一致性的对齐方式的原则，希望可以重建出正面'canonical view' 的结果，故而引入对称性这一种先验信息。

1、基本信息预估正面视角结果，包括 depth,albedo,viewpoint,lighting，且用不同的网络结构分别估计。

注：在一般的实验中，albedo 和 lighting 往往混合在一起，如果将 albedo 假设为对称的，则 albedo 的不对称性可以通过 lighting 引起，从而将两个分离开来。除此之外，还应该将 shadeing 和 shape 分开估计，shading 不代表完全的 shape，可能也混有 lighting 的信息。

2、对称性的利用以及 loss 的设计：在本文中为约束对称性，显示将预估得到的 depth,albedo 进行水平翻转。通过渲染投影将得到的结果和水平翻转过的结果重投影回输入图片的视角，计算重投影误差。

$$L = \frac{1}{Z}\sum_{uv} ln\frac{exp(-\frac{\sqrt{2}l_{uv}}{\delta_{uv}})}{\delta_{uv}}$$

这里 Z 是归一化因子，LOSS 计算对水平翻转和未水平翻转过的结果计算两次，所以 l_uv 分别表示两次计算的结果，除此之外，δ_{uv} 在不同的计算中也有不同的意思，在正常的估计结果中，表示输入图片信息对正面视角预估出来的结果的置信度。在翻转的结果中，表示预估的部分是否是对称的置信度。loss 还包括 perceptual loss，同样也对翻转和未翻转的结果进行计算。

3、一些小细节：

- 网络在估计 depth 和 albedo 的时候，用到了 encoder 和 decoder 结构，但是并没有用到常用的 skip 方法，是因为输入图片和输出结果并不是对齐的。这一点在后面论文中有其他的解决办法。

- 对于深度的尺度不一致问题, 作者对预估出来的深度先进行归一化处理，归一化之后再通过 tanh 激活函数，再对估计得到的深度进行缩放到统一的尺度上。

3 多视角 Model Base

3.1 Deep Facial Non-Rigid Multi-View Stereo

作者基于多视角人脸图片，实现端到端的训练神经网络得到对应每一张输入图片的人脸三维重建模型。

1、人脸模型的描述方便：3DMM+ 非线性部分（网络）

同一组输入图片有相同的非线性基，作者利用这部分非线性基来弥补 3DMM 模型表达能力不足的问题。利用当前重建 MESH 的结果，到输入图片上提取特征，并得到 UVmap，以及 position MAP, 得到各组输入图片的基。

2、**借助多视角之间的几何一致性融合多视角信息**。（在算法的正向计算中利用多视角信息）

由于 3dmm 可以作为不同视角下人脸的对齐桥梁（理解为语义对齐），投影到各个视角下，能够实现**语义上的一致性**，并将这个一致性 loss 作为后面的能量函数。

3、不同于用网络回归得到外参以及人脸模型的参数，作者参考传统 MVS 利用几何信息的方法，借助梯度下降对当前参数进行更行。**同时利用学习的方法，得到数据集中的先验信息**，可以调整学习率，以更少的迭代次数，达到最优点。

基于 2 中的能量函数进行随机梯度下降算法，更新参数（外参，人脸模型参数），这里步长是 mlp 学习得到的。

通过实验可以知道，在有限的迭代次数只能，优化的效果要优于 ADAM 优化器或者固定学习率的随机梯度下降。

4、训练 LOSS

监督训练，包含三维点 location 的误差，法向量 + 边距的平滑项 +landmark 点。

5、量化结果：三维误差:1.11mm(BU3DFE)，大尺度上的人脸对齐效果不好, 但是好于单视角。（our dataset）

3.2 Self-Supervised Monocular 3D Face Reconstruction by Occlusion-Aware Multi-view Geometry Consistency

这篇文章主要运用多视角信息解决姿态较大的人脸重建问题，自监督

1、人脸模型：3DMM

先天限制，3DMM 表述能力。

2、多视角几何一致性

作者对每一张输入图片进行建模，利用不同视角之间的相对外参进行一致性比较。比如说，A，B 两个视角，分别可以预估外参和模型参数，计算 AB 视角之间的相对外参，以 A 视角的三维模型，AB 的相对外参，找到 B **视角对应点的特征以及深度**，可以进行特征一致性的比较以及深度一致性的比较。注意这里还考虑到了不同视角尺度的变化，故而有一定的缩放。

但是这里将几何一致性的约束放在了无监督上面，所以在算法的正向计算中，并没有引入这一点。

3、极线约束

作者创新型引入**极线约束**，由于 landmark 的对应点已知，故而可以利用 landmark 的对应关系，建立极线约束关系，使得对应点在极线附近，这样可以有助于大尺度外参的估计。（后来的实验中发现极线约束与三维重建的重投影是一个物理意义，都是为了重建的两条射线共面。）

4、共同可见部分

在多视角中由三维模型投影到不同视角下需要计算不同视角的共同可见点。作者提出了实现方式：**对于 target 视角中，三维 P 点可见，则包含 P 点的所有三角面片，在 source 视角中同样可见。**

5、实验结果：重建细节不明显，大尺度人脸对齐效果优于单视角，三维误差：1.55mm（BU3DFE）

4 多视角 Model Free

4.1 Learning to Aggregate and Personalize 3D Face from In-the-Wild Photo Collection

该论文可以看作是 Unsuper3d 的多视角版本，同样估计“正面视角”的结果。所以该论文重点在于如何融合多视角信息。主要分为两个步骤：

STEP1: 预估出人脸基础的几何结构和有身份特征的 texture;

STEP2: 对预估出的结果进行微调，包括表情，几何细节等。对于多视角信息融合的点在于：

1、Aggregation Network: 首先对不同的输入图片进行估计 depth,albedo,pose,lighting, 对于每一个视角的 depth 和 albedo 的 encoder 结果 x_i^a 进行融合。利用 adptive aggregation 的方法，强调 encoder 的结果的每一个特征维度对确定 id 的重要性是不一致的，故而对齐进行加权 w_i^a 求和。

$$x_c^a = \sum_{i=1}^{N} w_i^a * x_i^a$$

2、学习各个输入图片独特的特征：

- 首先对多视角中每一张图片提取特征，与得到的有 id 信息的 depth 和 albedo 进行对比。但由于不对齐的原因，没有办法进行直接的 concat。故而只将 input 图片中比较低维的特征进行 concat。
- 除此之外还参考了人脸编辑中的方法，对输出的有特征点的图片应该有一部分和原有的结果一致，一部分和原有的不一致，例如表情等。所以提出 attention mask 结构。

  $Mask = con(feature_e, feature_d) feature_d = con(Mask * feature_e + feature_d)$

  这里 $feature_d$ 代表的是更加高分辨率的结果是从低分辨率的结果加上 mask 的 encoder 的结果的来的。

3、loss: loss 的设计同 super3d，所以也需要估计置信度。还有一点训练技巧在预计 id 一致性的部分，有对不同区域的人脸进行加权。

4、在训练方法上采用了 curriculum learning , 也就是先在简单的数据集上进行训练，再迁移到复杂的数据集。这训练的主要难点在于 wild 的多视角数据集再光照，背景等方面都比较复杂，所以现在用 GAN 生成的较为简单的数据集上进行训练。GAN 生成的数据集人脸识别 ARCFACE，loss 在 0.6 以下即可。

5、一些思考

文章主要有启发性的地方在于多视角信息的利用上面，以及在网络结构的设计上要注意对齐问题。感觉在训练的技巧上比较多，值得积累。