PFLD

笔记本：	Papers		
创建时间：	2019/3/3 17:10	**更新时间：**	2019/3/3 17:17
作者：	WarmIce		
标签：	cs:CV		

提出了一个优良的人脸关键点检测算法的理想目标是：精准、处理速度、模型大小，即accurate、efficient、compact。
为了达到这样的目标，眼下有4个Challenge：
#1 Local Variation：expression、illumination、occlusion、makeup
#2 Global Variation：pose、imaging quality（blur）
#3 Data Imbalance：
#4 Model Efficiency
Contributions：

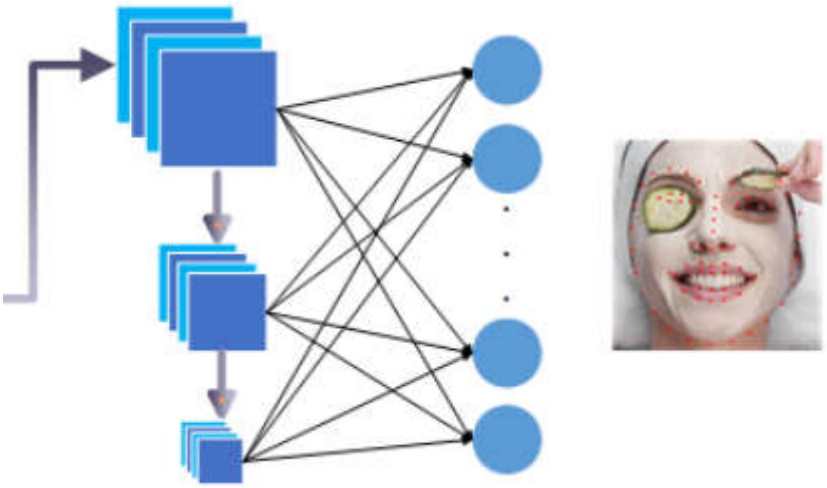
#1 A novel loss：concerns geometric constraint and data imbalance, which is penalizing more on large pose(large yaw, pitch and roll angle) and rare training samples.
So need to get the geometric information(which is yaw, pitch, and roll angle) by an auxiliary network and the fraction of different kinds of training samples.

$$\mathcal{L} := \frac{1}{M}\sum_{m=1}^{M}\sum_{n=1}^{N}\left(\sum_{c=1}^{C}\omega_n^c\sum_{k=1}^{K}(1-\cos\theta_n^k)\right)\|\mathbf{d}_n^m\|_2^2$$

Notice:
1. **theta^k** is the angles of deviation between the **ground-truth** and **estimated** yaw, pitch, and roll angles.
Use the auxiliary network to estimate the yaw, pitch, and roll angles, the auxiliary network loss comes from the target Euler angles and predicted angles.
Target Euler angles come from the caculatation in section 2.3
2. categorize a sample into one or multiple attribute classes including profile-face, frontal-face, head-up, head-down, expression, and occlusion(**6** categories).
这里的weight^c的计算方法是，我们一次送入网络batchsize个sample，记为samples，对于category c来说，这批样本中属于这一类别的样本数量为n，则weight^c=batchsize/n。这样之所以有效的原因在于，更多的数据都是frontal-face，而非其余5种类别，因此计算出来的大多数weight^c求和都是1，但是一旦样本中出现了侧脸、头部上下、表情以及遮挡，其数量也不会大到所取的整个batchsize都是侧脸，都是头朝上，都是头朝下，都是夸张表情，或者都有遮挡，即在出现rare

training sample的时候，weight^c很难退化到1，这就足以保证loss函数中处理data imbalance的效果。

#2 A **multi-scale fully-connected layer** is added to precisely localize the landmarks



#3 Use MobileNet blocks as the backbone network

The network can be compressed by adjusting the width parameter of MobileNets according to demand from users, for making the model smaller and faster.[**quantization techniques** are totally compatible with ShuffleNet and MobileNet, which means the size of our model can be further reduced by quantization]

TODO:

- [ ] Read the comprehensive survey of facial landmark localization (**classic** previous arts)
- [ ] What is the MobileNet?
- [ ] Learn the 300W and AFLW dataset
- [ ] Face detector: Joint face detection and alignment using multitask cascaded convolutional networks
- [ ] A weak perspective model: Pose-invariant 3d face alignment