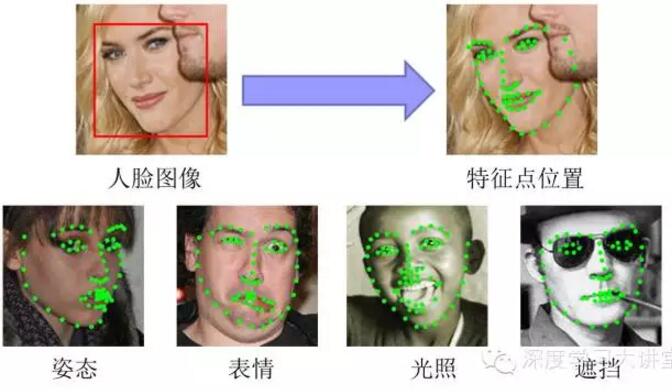
一、人脸对齐概述

人脸对齐任务即根据输入的人脸图像，自动定位出面部关键特征点，如眼睛、鼻尖、嘴角点、眉毛以及人脸各部件轮廓点等，如下图所示。



人脸对齐任务可拆分为三个子问题：

1. 如何对人脸表观图像（输入）建模
2. 如何对人脸形状（输出）建模
3. 如何建立人脸表观图像（模型）与人脸形状（模型）的关联

人脸形状建模典型的方法有可变形模板（Deformable Template）、点分布模型（主动形状模型Active Shape Model）、图模型等。

人脸表观建模又可分为全局表观建模和局部表观建模。全局表观建模简单的说就是考虑如何建模整张人脸的表观信息，典型的方法有主动表观模型Active Appearance Model（产生式模型）和Boosted Appearance Model（判别式模型）。对应的局部表观建模则是对局部区域的表观信息建模，包括颜色模型、投影模型、侧剖线模型等。

二、人脸对齐采用的方法

1、级联线性回归模型

级联形状回归模型在特征点定位任务上取得了重大突破，该方法使用回归模型，直接学习从人脸表观到人脸形状（或者人脸形状模型的参数）的映射函数，进而建立从表观到形状的对应关系。此类方法不需要复杂的人脸形状和表观建模，简单高效，在可控场景（实验室条件下采集的人脸）和非可控场景（网络人脸图像等）均取得不错的定位效果。

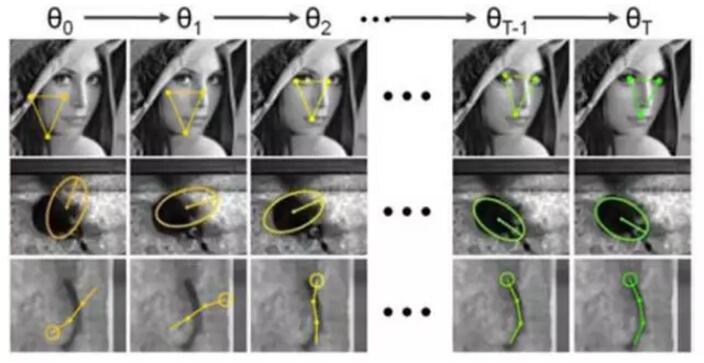
人脸对齐问题可以看作是学习一个回归函数F，以图象I作为输入，输出θ为特征点的位置（人脸形状）：θ = F（I）。

级联回归模型可以统一为以下框架：

学习多个回归函数{ ,…, , }来逼近函数F：

所谓的级联，即当前函数的输入依赖于上一级的输出，而每一个的学习目标都是逼近特征点的真实位置θ，为初始形状。通常不是直接回归真实位置θ，而是回归当前形状与真实位置θ之间的差：

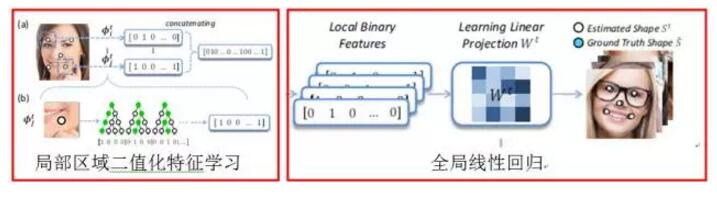
下面介绍几种典型的形状回归方法，根据不同点在于函数的设计不同及输入特征不同：

* 级联形状回归模型CascadedPose Regression（CPR），使用于预测物体的形状，如下图所示，给定初始形状（通常为平均形状），根据初始形状提取特征（两个像素点的差值）作为的输入。每个函数建模成Random Fern回归器（用多个特征组合来表达对象，博客地址: <http://blog.sina.com.cn/s/blog_8ff949cf0100yfof.html> ），预测当前形状与目标的差，并根据预测结果更新当前形状，作为下级函数的输入。

缺点：对于初始化的比较敏感，使用不同的初始化参数融合多次预测结果可以一定程度上缓解初始化对于算法的影像，但并不能完全解决该问题，当目标物体被遮挡，性能也会差。

* Robust Cascaded Pose Regression（RCPR）方法，为了解决遮挡问题，提出同时预测人脸形状和特征点是否被遮挡的状态，即的输出包含和每个特征点是否被遮挡的状态：

当某些特征点被遮挡时，则不选取该特征点所在区域的特征作为输入，从而避免遮挡对定位的干扰。此外，提出智能重启技术来解决形状初始化敏感的问题：随机初始化一组形状，运行 ,…, , }的前10%的函数，统计形状预测的方差，如果方差小于一定阈值，说明这组初始化不错，则跑完剩下的90%的级联函数，得到最终的预测结果；如果方差大于一定阈值，则说明初始化不理想，选择重新初始化一组形状。

* Supervised Descent Method（SDM）方法，使用监督梯度下降的方法来求解非线性最小二乘问题，并成功地应用在人脸对齐任务上。与CPR和RCPR不同的地方在于：建模成了线性回归模型；的输入为与人脸形状相关的SIFT特征。该特征的提取也很简单，即在当前人脸形状的每个特征点上提取一个128维的SIFT特征，并将所有SIFT特征串联到一起作为的输入。
* Discriminative Response Map Fitting（DRMF）方法，使用支持向量回归SVR来建模回归函数，并使用形状相关的HOG特征（提取方式与形状相关的SIFT类似）作为输入，来级联预测人脸形状。与SDM最大的不同在于，DRMF对于人脸形状做了参数化的建模。的目标变为预测这些形状参数而不再是直接的人脸形状。由于人脸形状参数化模型很难完美地刻画所有形状变化，SDM的实测效果要优于DRMF。
* Regressing Local Binary Features（LBF）方法，和SDM类似，也是建模成线性回归模型；不同的地方在于，SDM直接使用SIFT特征，LBF则基于随机森林回归模型在局部区域学习稀疏二值化特征。通过学习稀疏二值化特征，大大减少了运算开销，比CRP、RCPR、SDM、DRMF等方法具有更高的运行效率。并且在IBUG公开评测集上取得优于SDM、RCPR的性能。

级联形状回归模型成功的关键在于：

1) 使用了形状相关特征，即函数的输入和当前的人脸形状紧密相关；

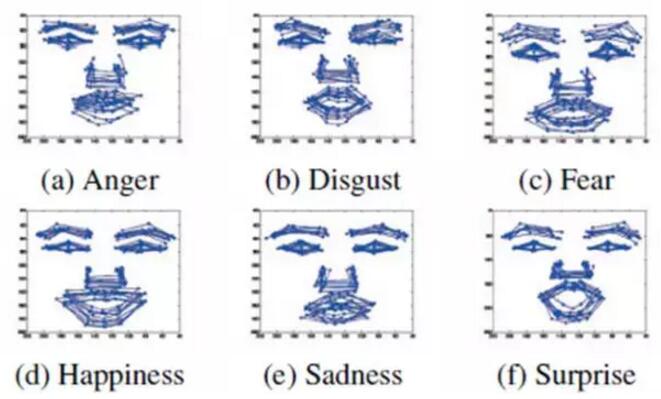
2) 函数的目标也与当前的人脸形状相关，即的优化目标为当前形状与真实位置之间的差。

此类方法在可控和非可控的场景下均取得良好的定位效果，且具有很好的实时性。

2、深度学习模型

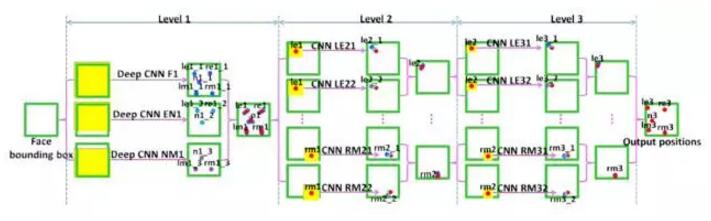
级联形状回归方法每一个回归函数都是浅层模型（线性回归模型、Random Fern等）。深度网络模型，比如卷积神经网络（CNN）、深度自编码器（DAE）和受限玻尔兹曼机（RBM）在计算机视觉的诸多问题，如场景分类，目标跟踪，图像分割等任务中有着广泛的应用，当然也包括特征定位问题。具体的方法分为两大类：使用深度模型建模人脸形状和表观的变化和基于深度网络学习从人脸表观到形状的非线性映射函数。

主动形状模型ASM和主动表观模型AAM使用主成分分析（PCA）来建模人脸形状的变化。由于姿态表情等因素的影响，线性PCA模型很难完美地刻画不同表情和姿态下的人脸形状变化。深度置信网络（DBN方法，来刻画不同表情下人脸形状的复杂非线性变化。处理不同姿态的特征点定位问题，使用RBM网络建模从正面到非正面的人脸形状变化。

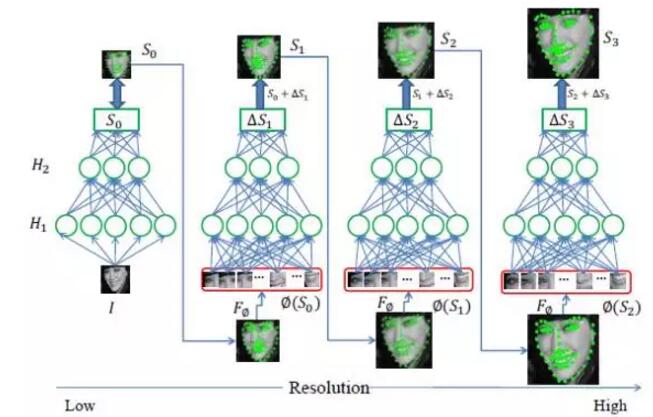


深度置信网络（DBN）：建模不同表情下的人脸形状变化的示意图。

* 级联（3级）卷积神经网络DCNN来实现人脸对齐的方法，统一在级联形状回归模型的大框架下，和CPR、RCPR、SDM、LBF等方法不一样的是，DCNN使用深度模型-卷积神经网络，来实现。第一级使用人脸图像的三块不同区域（整张人脸，眼睛和鼻子区域，鼻子和嘴唇区域）作为输入，分别训练3个卷积神经网络来预测特征点的位置，网络结构包含4个卷积层，3个Pooling层和2个全连接层，并融合三个网络的预测来得到更加稳定的定位结果。后面两级, 在每个特征点附近抽取特征，针对每个特征点单独训练一个卷积神经网络（2个卷积层，2个Pooling层和1个全连接层）来修正定位的结果。该方法在LFPW数据集上取得当时最好的定位结果。



* CFAN，一种由粗到精的自编码器网络（CFAN）来描述从人脸表观到人脸形状的复杂非线性映射过程。该方法级联了多个栈式自编码器网络，每一个刻画从人脸表观到人脸形状的部分非线性映射。具体来说，输入一个低分辨率的人脸图像I，第一层自编码器网络可以快速地估计大致的人脸形状，记作基于全局特征的栈式自编码网络。网络包含三个隐层，隐层节点数分别为1600,900,400。然后提高人脸图像的分辨率，并根据得到的初始人脸形状，抽取联合局部特征，输入到下一层自编码器网络来同时优化、调整所有特征点的位置，记作基于局部特征的栈式自编码网络。该方法级联了3个局部栈式自编码网络{ , , }直到在训练集上收敛。每一个局部栈式自编码网络包含三个隐层，隐层节点数分别为1296,784,400。得益于深度模型强大的非线性刻画能力，该方法在XM2VTS，LFPW，HELEN数据集上取得比DRMF、SDM更好的结果。此外，CFAN可以实时地完成人脸人脸对齐（在I7的台式机上达到23毫秒/张），比DCNN（120毫秒/张）具有更快的处理速度。



基于由粗到精自编码器网络的实时人脸对齐方法的示意图

三、人脸对齐的应用

(1)人脸器官定位、器官跟踪。通过人脸对齐，能够定位到人脸的每个部件，提取相应的部件特征。

(2)人脸器官定位、器官跟踪。通过人脸对齐，我们能够定位到人脸的每个部件，提取相应的部件特征。

(3)人脸漫画/素描图像生成。通过人脸对齐后，我们能够进行人脸漫画和素描生成。如：魔漫相机

(4)虚拟现实和增强现实。通过人脸对齐后，能够做出很多好玩的应用。

(5)人脸老化、年轻化、年龄推断。特征融合/图像增强。通过人脸对齐后，我们能够有效提取人脸特征，并分析人脸年龄、人脸老化等。

(6)纹理过渡。如：换脸

(7)性别鉴别。通过人脸对齐，能够对人脸进行性别识别，男女之间的人脸形状有一定的差异性。

(8)3D卡通。通过人脸对齐能够进行3D卡通模拟。

目前有很多的人脸对齐算法，比较传统的有ASM、AAM、CLM和一些列改进算法，而目前比较流行的有ESR、3D-ESR、SPR、LBF、SDM、CFSS等。

ASM算法相对容易，其中STASM是目前正面脸当中比较好的算法，原作者和CLM比较过。但是STASM速度较慢，大概10frame/s左右。ASM对齐在精度上不如AAM，AAM由于使用全局纹理信息，因此精度较高，但是遇到光照和多姿态时，对初始化Shape要求很高，不然容易陷入局部优化。CLM 分别继承了ASM和AAM的一些特征，效果得到了提升。对局部器官特征的概率假设和优化算法的选择，是CLM各种算法的本质区别。