这个是香港中文大学的特征点定位的论文，现在可以说是最棒的特征点定位技术啦！！！

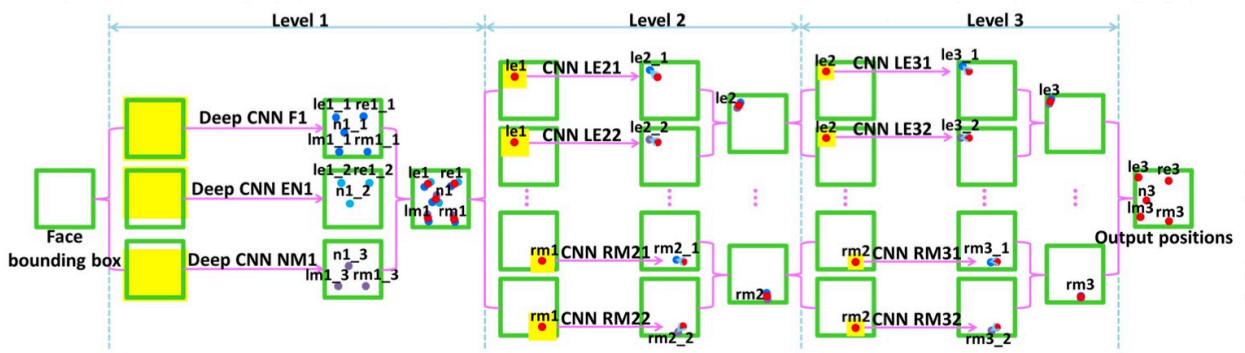


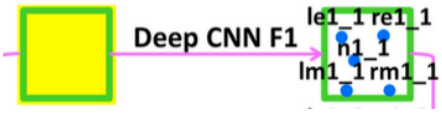

Figure 2: Three-level cascaded convolutional networks. The input is the face region returned by a face detector. The three networks at level 1 are denoted as F1, EN1, and NM1. Networks at level 2 are denoted as LE21, LE22, RE21, RE22, N21, N22, LM21, LM22, RM21, and RM22. Both LE21 and LE22 predict the left eye center, and so forth. Networks at level 3 are denoted as LE31, LE32, RE31, RE32, N31, N32, LM31, LM32, RM31, and RM32. Green square is the face bounding box given by the face detector. Yellow shaded areas are the input regions of networks. Red dots are the final predictions at each level. Dots in other colors are predictions given by individual networks.

他这里面第一层输入会有三种不同的输入
一个是输入全部的脸F，一个是眼鼻子NE，嘴鼻子NM

| | layer 0 | layer 1 | layer 2 | layer 3 | layer 4 | layer 5 | layer 6 | layer 7 | layer 8 | layer 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| S0 | I(39,39) | CR(4,20,2,2) | P(2) | CR(3,40,2,2) | P(2) | CR(3,60,3,3) | P(2) | CR(2,80,2,2) | F(120) | F(10) |
| S1 | I(31,39) | CR(4,20,1,1) | P(2) | CR(3,40,2,2) | P(2) | CR(3,60,2,3) | P(2) | CR(2,80,1,2) | F(100) | F(6) |
| S2 | I(15,15) | CR(4,20,1,1) | P(2) | CR(3,40,1,1) | P(2) | F(60) | F(2) | | | |
| S3 | I(39,39) | CR(4,20,2,2) | P(2) | CR(3,40,2,2) | P(2) | CR(3,60,3,3) | P(2) | F(120) | F(10) | |
| S4 | I(39,39) | CR(4,20,2,2) | P(2) | CR(3,40,2,2) | P(2) | F(120) | F(10) | | | |
| S5 | I(39,39) | CR(4,20,2,2) | P(2) | F(120) | F(10) | | | | | |
| S6 | I(39,39) | C(4,20,2,2) | P(2) | C(3,40,2,2) | P(2) | C(3,60,3,3) | P(2) | C(2,80,2,2) | F(120) | F(10) |
| S7 | I(39,39) | CR(4,20,1,1) | P(2) | CR(3,40,1,1) | P(2) | CR(3,60,1,1) | P(2) | CR(2,80,1,1) | F(120) | F(10) |

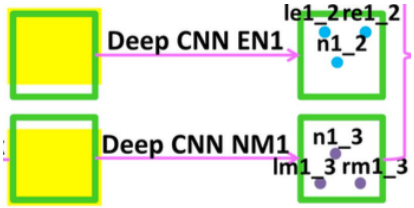
Table 1: Summary of network structures. F1 adopts S0. Both EN1 and NM1 adopt S1. All the networks at the second and third levels share S2. To investigate different designs of network structures, we also compare different structures S3-S7 for F1 in experiments.

这里黄色标记的那句话：All the networks... 注意⚠️
是这个意思👇：



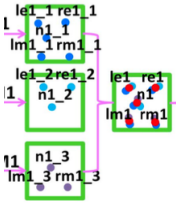
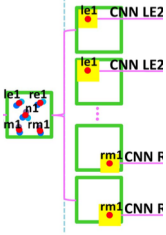

这个是F按照S0的方式计算，对应下面10层

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| S0 | I(39,39) | CR(4,20,2,2) | P(2) | CR(3,40,2,2) | P(2) | CR(3,60,3,3) | P(2) | CR(2,80,2,2) | F(120) | F(10) |

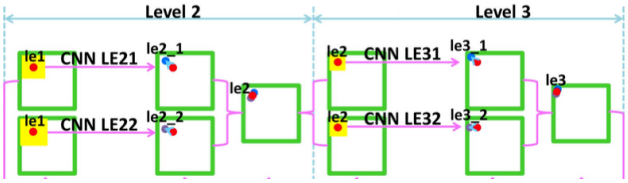



这个是NE，NM按照S1的方式计算，对于下面十层

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| S1 | I(31,39) | CR(4,20,1,1) | P(2) | CR(3,40,2,2) | P(2) | CR(3,60,2,3) | P(2) | CR(2,80,1,2) | F(100) | F(6) |

把上面三个合并在一起，这里相当于左右眼睛、左右嘴角分别有两个估算，鼻子有三个估算，就像这个一样  ，然后把这个几个估算（蓝色）合成一个（红色），像这个 

然后后面的输入是：对于5个部位的15*15的小框框，👇按照S2的方法来，每个点生成两个估算




合并的方法👉：

$$x = \frac{x_1^{(1)} + \cdots + x_{l_1}^{(1)}}{l_1} + \sum_{i=2}^{n} \frac{\Delta x_1^{(i)} + \cdots + \Delta x_{l_i}^{(i)}}{l_i} \quad (1)$$

for an n-level cascade with $l_i$ predictions at level $i$. Note that predictions at the first level are absolute positions while predictions at the following levels are adjustments.

举个🌰：这个论文里面的是3级联，
对于鼻子有三个估计点，那么
第一层的预测就是这个 X1(nose)=(X1+X2+X3)/3

第二层的预测就是X2(nose) = (X21 + X22)/2 + {[X21-X1(nose) ] + [X22 - X1(nose)]}/2

第三层的预测就是X3(nose) = (X1 + X2)/2 + {[X31-X2(nose) ] + [X32 - X2(nose)]}/2
+ {[X21-X1(nose) ] + [X22 - X1(nose)]}/2