人脸识别算法

商业级项目实战系列

知识点介绍

- 人脸识别算法简介
- MTCNN人脸侦测
- 人脸特征提取与对比
- 代码实现

人脸识别3个步骤

- ▶ 人脸检测

- ▶ 特征提取

- ▶ 人脸对比

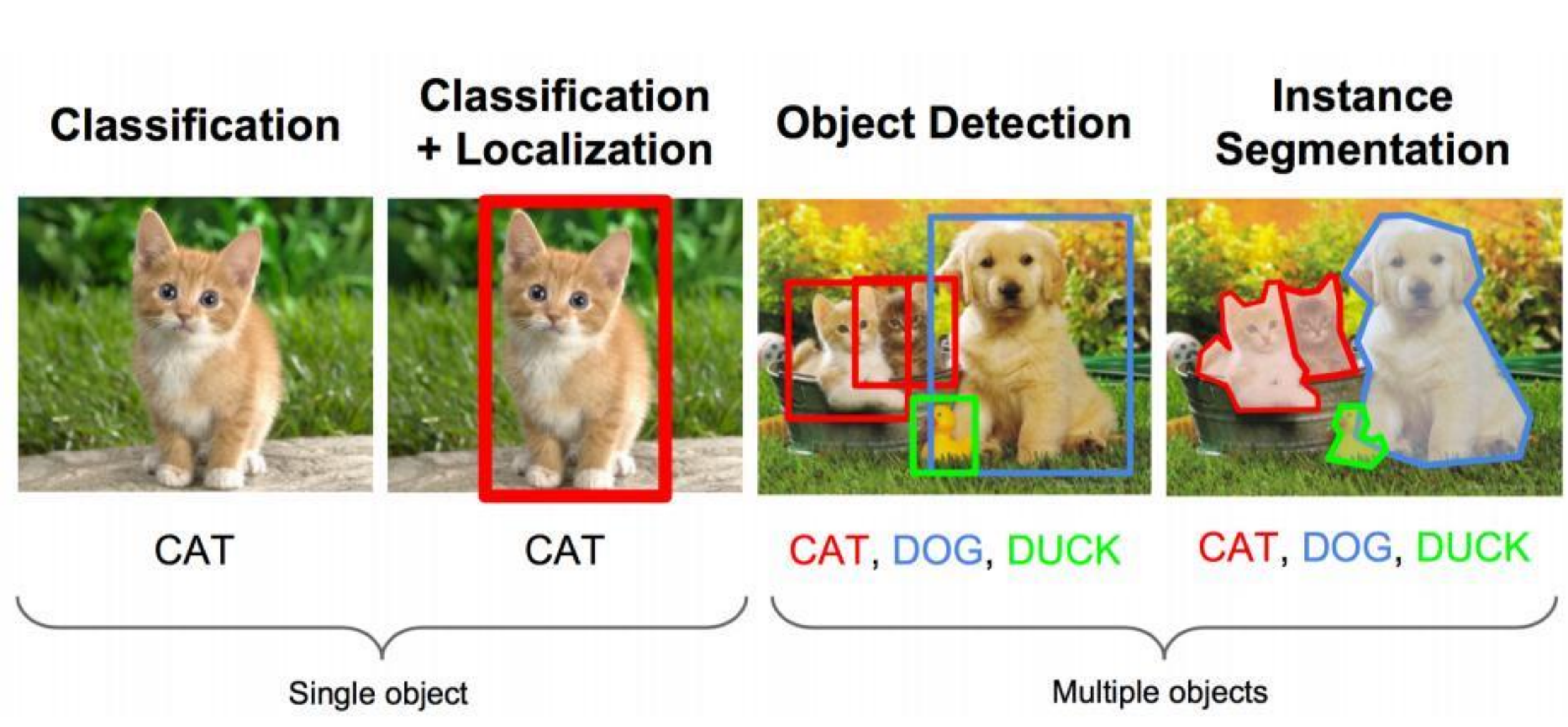
MTCNN简介

猎维科技
LIEWEI . AI

《Joint Face Detection and Alignment using Multi-task Cascaded Convolutional Networks》

来自于中国科学院深圳先进技术研究院，乔宇老师组

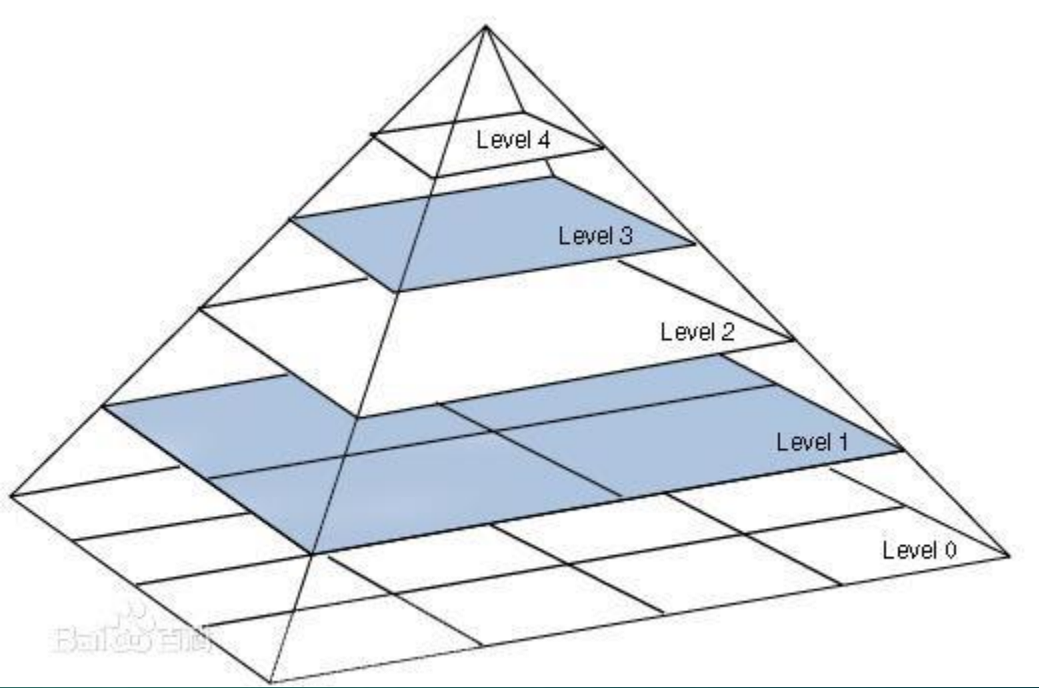
图像跟踪算法思路


Classification
Classification + Localization
Object Detection
Instance Segmentation
CAT
CAT
CAT, DOG, DUCK
CAT, DOG, DUCK
Single object
Multiple objects

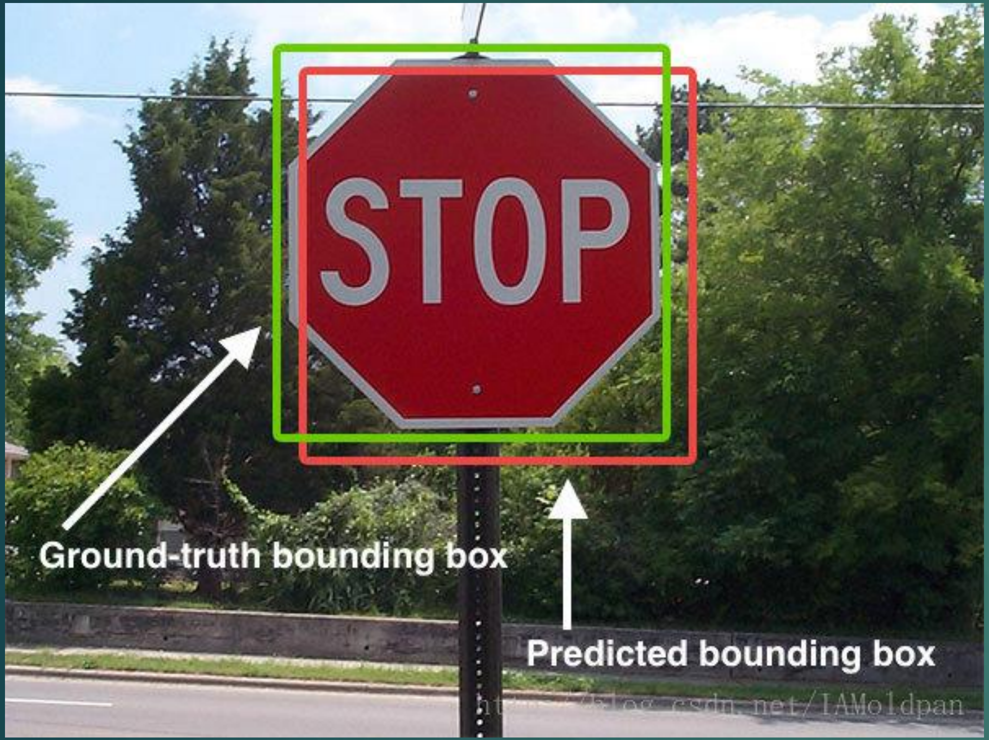
图像金字塔

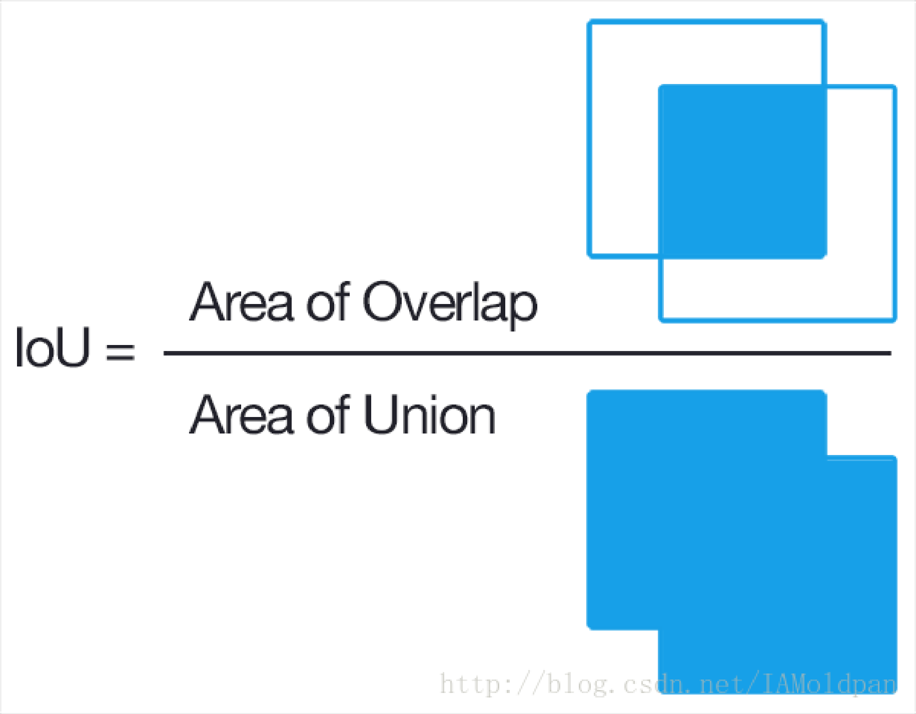
猎维科技 LIEWEI.AI

图像金字塔是图像多尺度表达的一种，是一种以多分辨率来解释图像的有效但概念简单的结构。一幅图像的金字塔是一系列以金字塔形状排列的分辨率逐步降低，且来源于同一张原始图的图像集合。其通过梯次向下采样获得，直到达到某个终止条件才停止采样。我们将一层一层的图像比喻成金字塔，层级越高，则图像越小，分辨率越低。

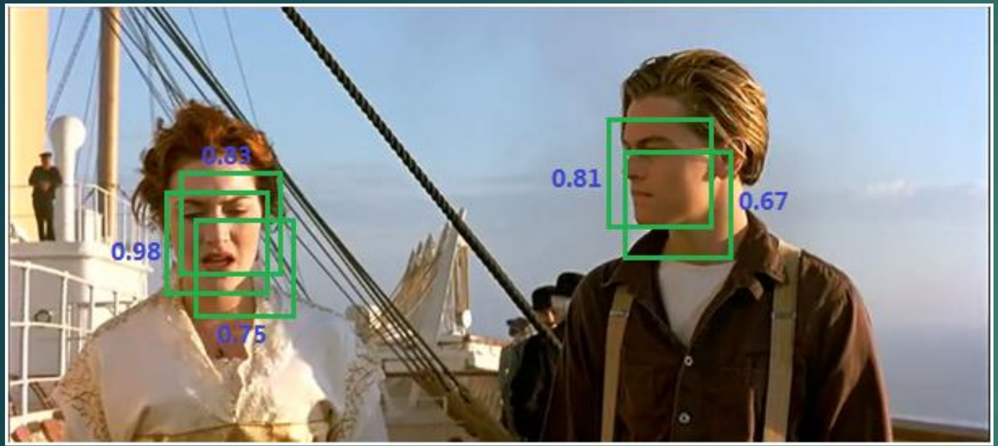


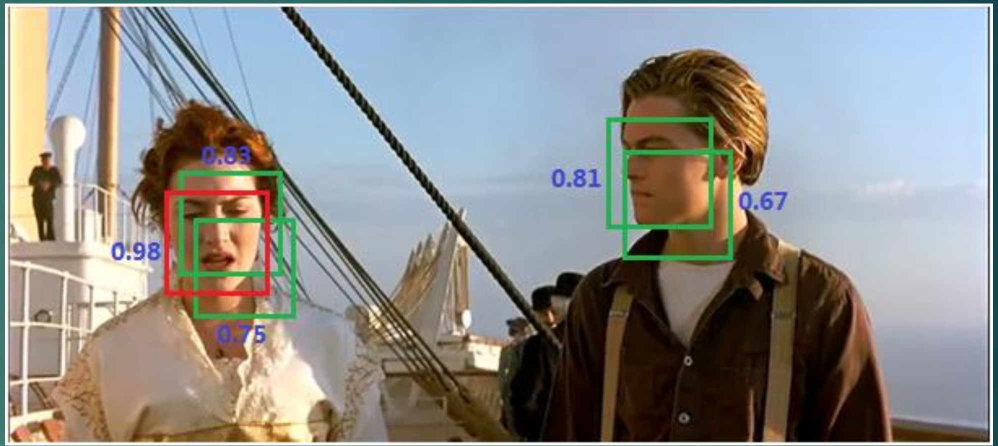
重叠度IOU

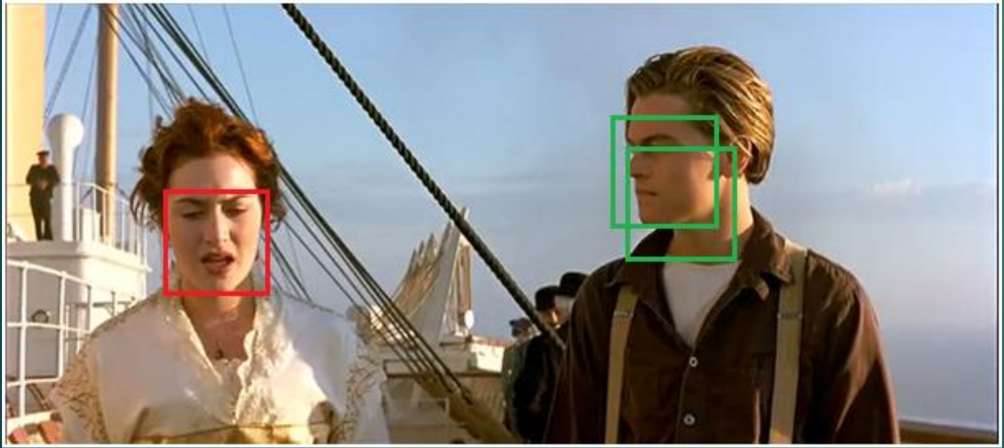

STOP
Ground-truth bounding box
Predicted bounding box
http://blog.csdn.net/IAMoldpan

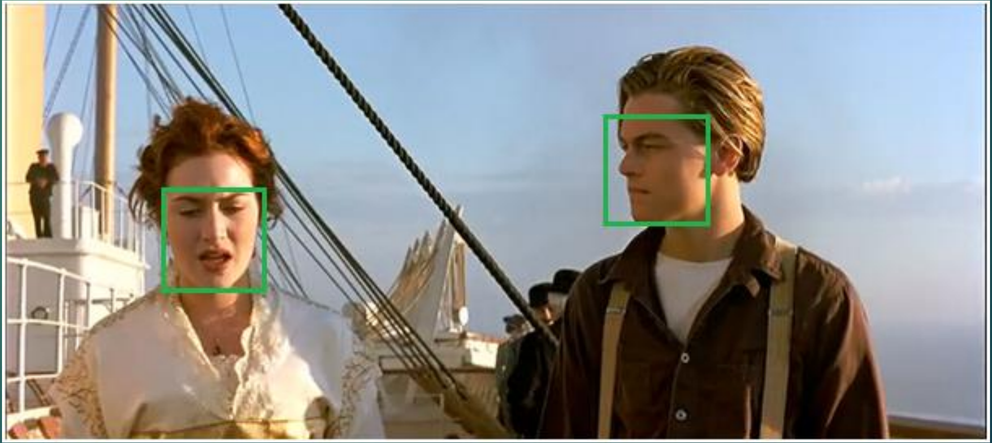

IoU =
Area of Overlap
Area of Union
http://blog.csdn.net/IAMoldpan

NSM非极大值抑制









特征层到原图的映射

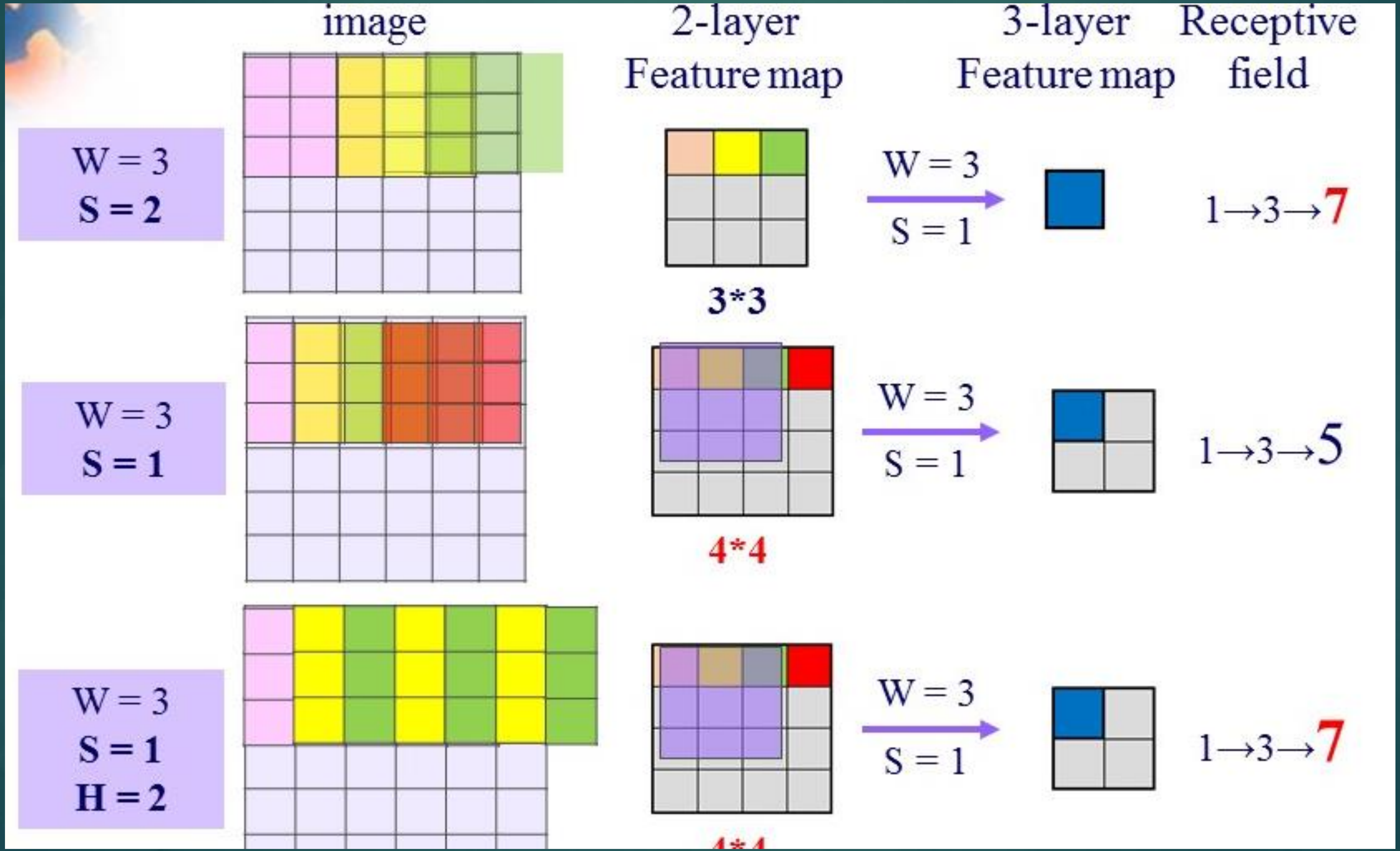
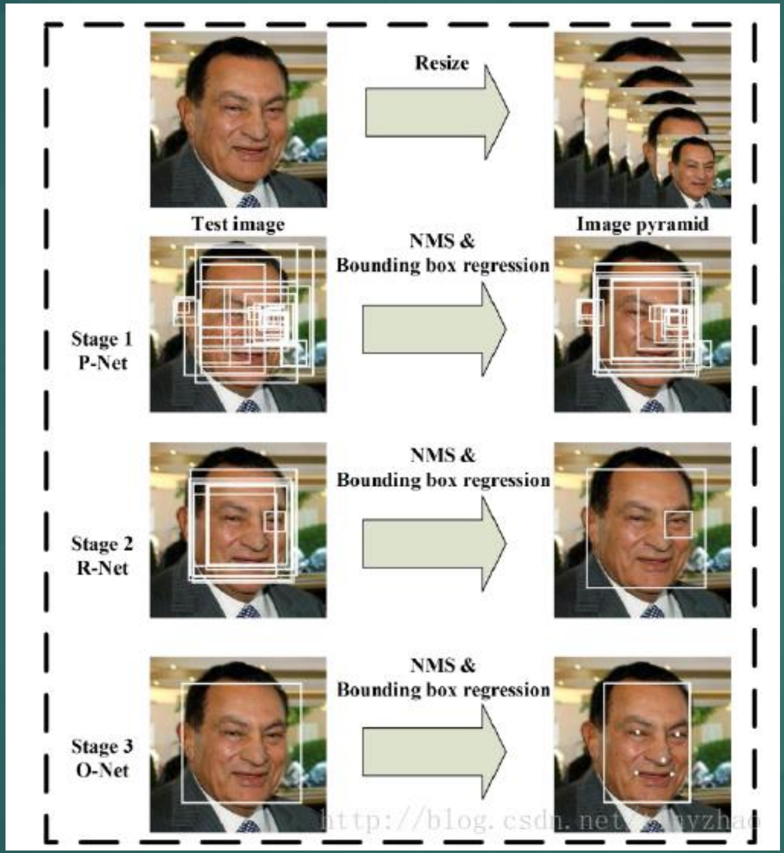
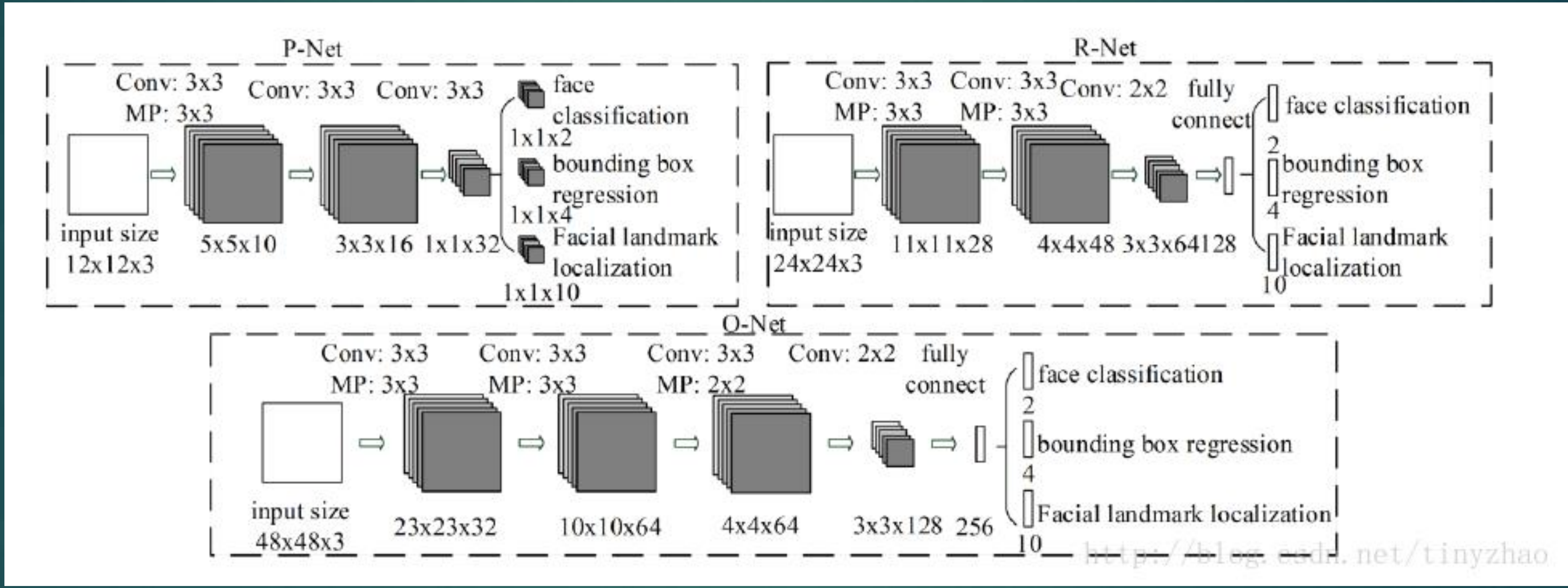


image
2-layer Feature map
3-layer Feature map
Receptive field
W = 3
S = 2
3*3
W = 3
S = 1
1→3→7
W = 3
S = 1
4*4
W = 3
S = 1
1→3→5
W = 3
S = 1
H = 2
W = 3
S = 1
1→3→7

MTCNN算法流程


Resize
Test image
Image pyramid
NMS &
Bounding box regression
Stage 1
P-Net
NMS &
Bounding box regression
Stage 2
R-Net
NMS &
Bounding box regression
Stage 3
O-Net

MTCNN网络结构


P-Net
Conv: 3x3
MP: 3x3
Conv: 3x3
Conv: 3x3
face classification
1x1x2
bounding box regression
1x1x4
Facial landmark localization
1x1x10
input size 12x12x3
5x5x10
3x3x16
1x1x32
R-Net
Conv: 3x3
MP: 3x3
Conv: 3x3
MP: 3x3
Conv: 2x2
fully connect
face classification
2
bounding box regression
4
Facial landmark localization
10
input size 24x24x3
11x11x28
4x4x48
3x3x64
128
O-Net
Conv: 3x3
MP: 3x3
Conv: 3x3
MP: 3x3
Conv: 3x3
MP: 2x2
Conv: 2x2
fully connect
face classification
2
bounding box regression
4
Facial landmark localization
10
input size 48x48x3
23x23x32
10x10x64
4x4x64
3x3x128
256
http://blog.csdn.net/tinyzhao

MTCNN训练样本

训练数据集：Wider_face 和 CelebA

0-0.3：非人脸

0.65-1.00：人脸

0.4-0.65：Part人脸

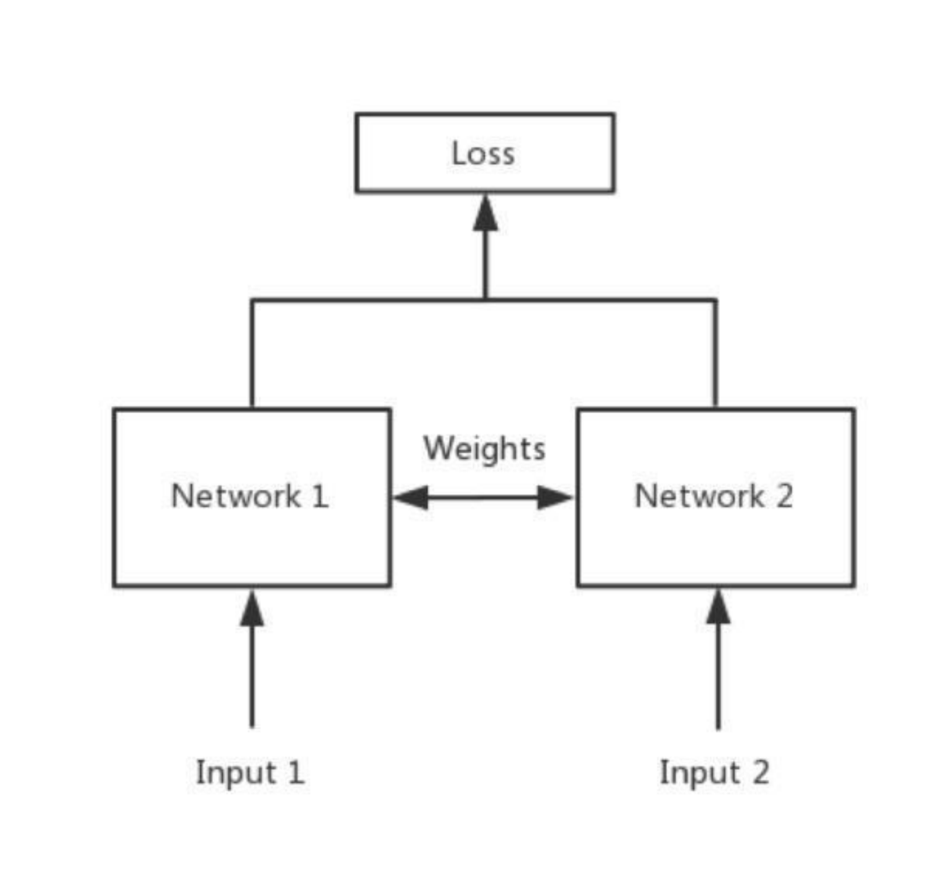
0.3-0.4：地标

训练样本的比例，负样本:正样本:part样本:地标=3:1:1:2

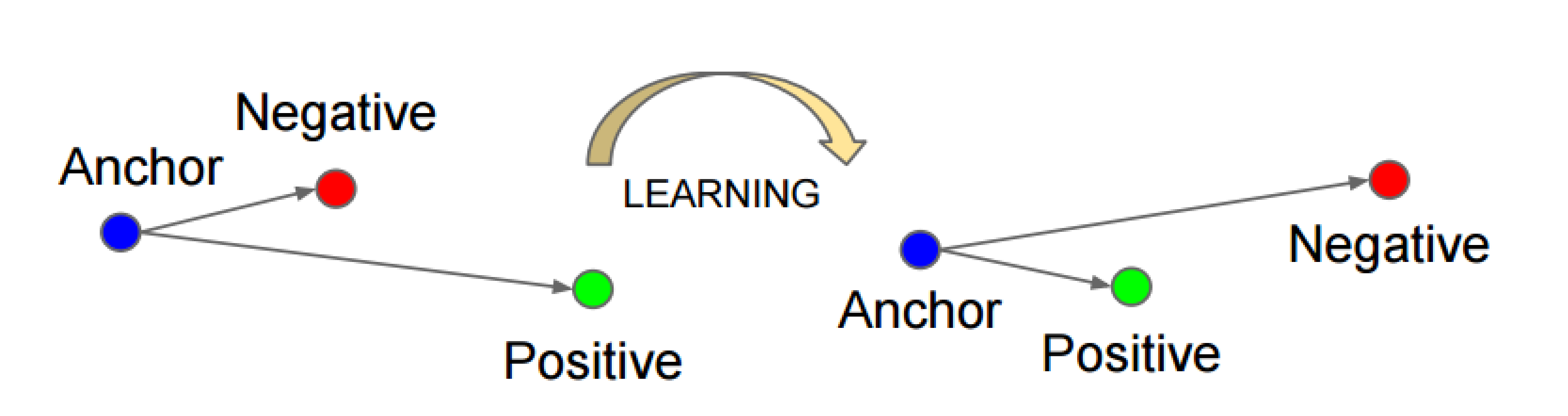
人脸识别的难点

- 它不是一个分类问题
- 脸与脸之间相似度很高，有时候人类都难以区分

Siamese network 孪生神经网络

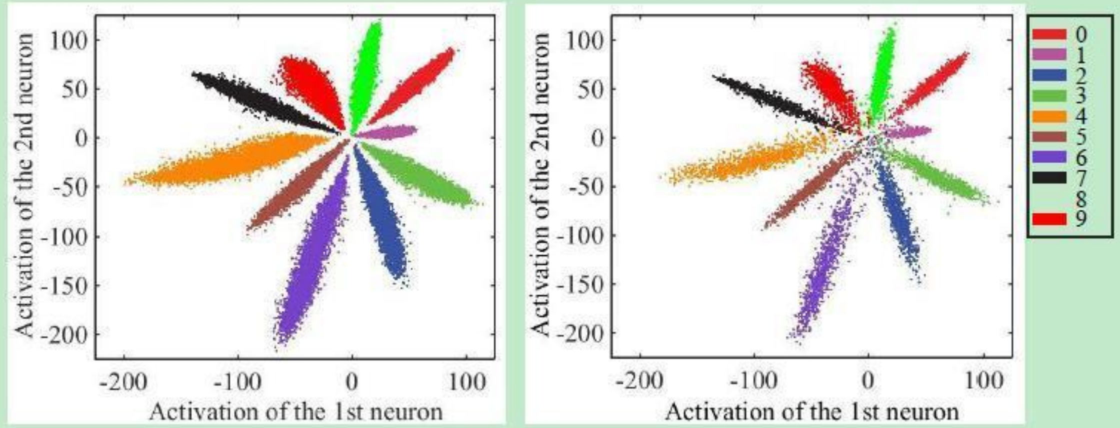


Triplet Loss



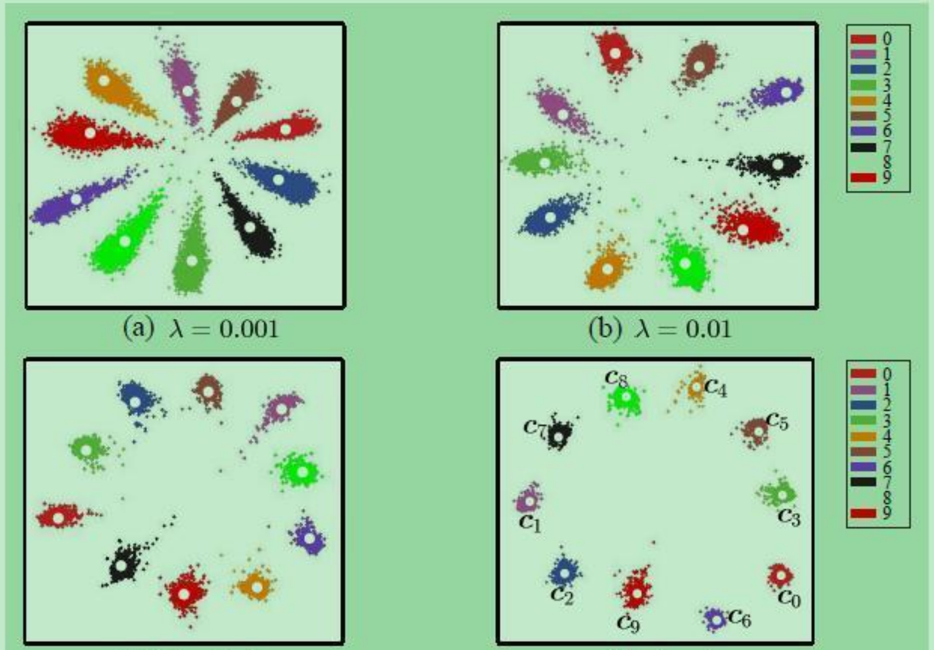
$$\sum_{i}^{N} \left[\|f(x_i^a) - f(x_i^p)\|_2^2 - \|f(x_i^a) - f(x_i^n)\|_2^2 + \alpha \right]_+$$

Center Loss



(a) (b)

Fig. 2. The distribution of deeply learned features in (a) training set (b) testing set, both under the supervision of softmax loss, where we use 50K/10K train/test splits. The points with different colors denote features from different classes. **Best viewed in color.**



(a) $\lambda = 0.001$ (b) $\lambda = 0.01$

(c) $\lambda = 0.1$ (d) $\lambda = 1$

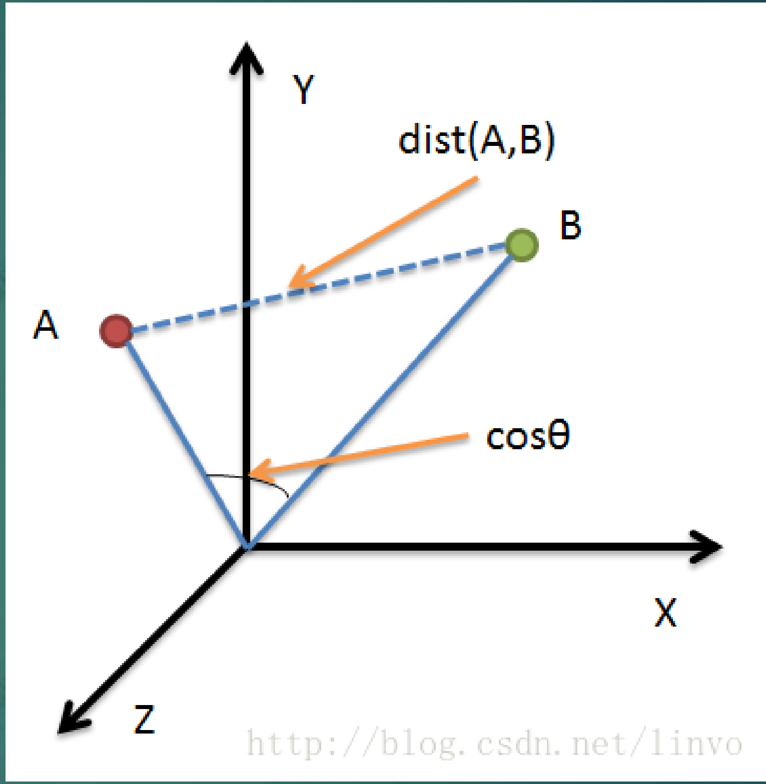
Fig. 3. The distribution of deeply learned features under the joint supervision of softmax loss and center loss. The points with different colors denote features from different classes. Different λ lead to different deep feature distributions ($\alpha = 0.5$). The white dots (c_0, c_1,...,c_9) denote 10 class centers of deep features. **Best viewed in color.**

$$\mathcal{L}_C = \frac{1}{2}\sum_{i=1}^{m} \|\boldsymbol{x}_i - \boldsymbol{c}_{y_i}\|_2^2$$

距离衡量

$$dist(X,Y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

$$\cos\theta = \frac{\sum_{i=1}^{n}(A_i \times B_i)}{\sqrt{\sum_{i=1}^{n}(A_i)^2} \times \sqrt{\sum_{i=1}^{n}(B_i)^2}} = \frac{A^T \cdot B}{\|A\| \times \|B\|}$$



余弦相似度和欧氏距离等价性

For ℓ^2-normalized vectors $\mathbf{x}, \mathbf{y}$,

$$||\mathbf{x}||_2 = ||\mathbf{y}||_2 = 1,$$

we have that the *squared* Euclidean distance is proportional to the cosine distance,

$$\begin{aligned} ||\mathbf{x} - \mathbf{y}||_2^2 &= (\mathbf{x} - \mathbf{y})^\top (\mathbf{x} - \mathbf{y}) \\ &= \mathbf{x}^\top \mathbf{x} - 2\mathbf{x}^\top \mathbf{y} + \mathbf{y}^\top \mathbf{y} \\ &= 2 - 2\mathbf{x}^\top \mathbf{y} \\ &= 2 - 2\cos\angle(\mathbf{x}, \mathbf{y}) \end{aligned}$$

重要结论：cosine similarity is identical to l2-normalized euclidean distance someway.

A-SoftmaxLoss

$$L_{\text{ang}} = \frac{1}{N}\sum_i -\log\left(\frac{e^{\|\boldsymbol{x}_i\|\cos(m\theta_{y_i,i})}}{e^{\|\boldsymbol{x}_i\|\cos(m\theta_{y_i,i})} + \sum_{j\neq y_i} e^{\|\boldsymbol{x}_i\|\cos(\theta_{j,i})}}\right) \quad (6)$$

$$L_6 = -\frac{1}{m}\sum_{i=1}^{m}\log\frac{e^{s(\cos(\theta_{y_i})-m)}}{e^{s(\cos(\theta_{y_i})-m)} + \sum_{j=1,j\neq y_i}^{n} e^{s\cos\theta_j}}.$$

$$L_7 = -\frac{1}{m}\sum_{i=1}^{m}\log\frac{e^{s(\cos(\theta_{y_i}+m))}}{e^{s(\cos(\theta_{y_i}+m))} + \sum_{j=1,j\neq y_i}^{n} e^{s\cos\theta_j}}, \quad (9)$$

人脸识别数据集

1.Labeled Faces in the Wild Home (LFW)

2.CASIA-FaceV5

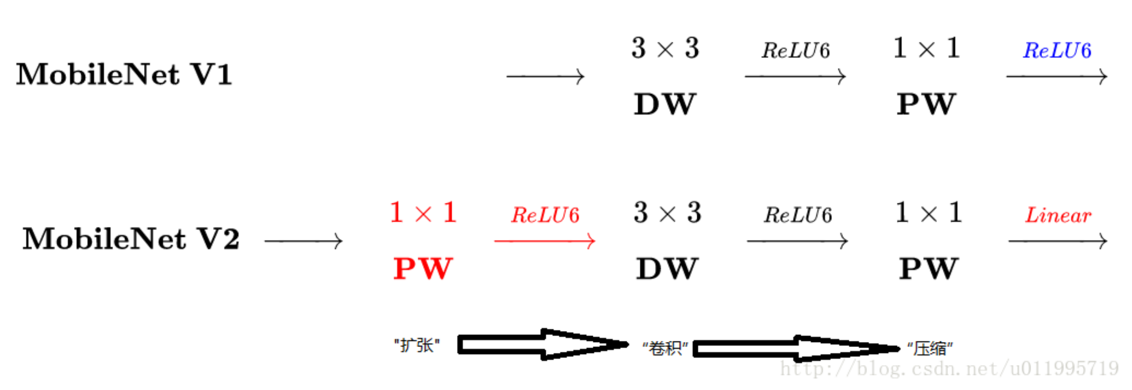
3.CASIA-3D FaceV1

4.CASIA-WebFace

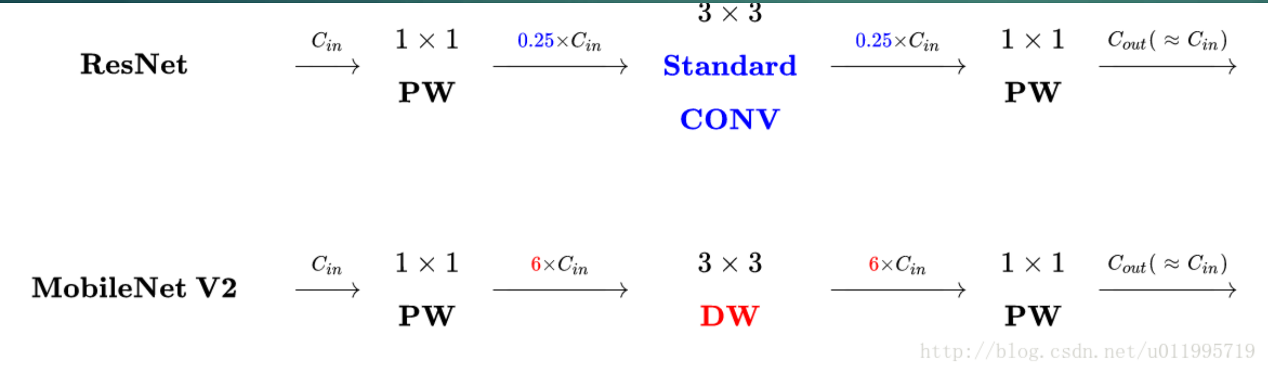
5.VGGFace2

6.mscebel1m

mobilenet



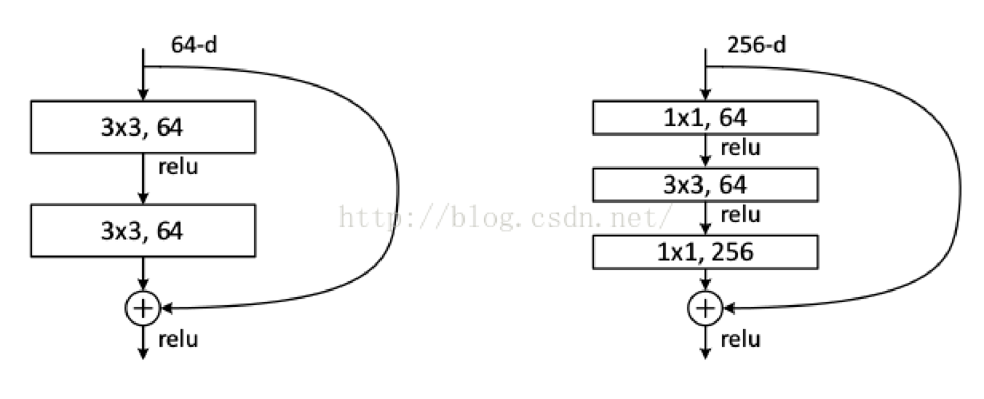
MobileNet V1
3 × 3
DW
ReLU6
1 × 1
PW
ReLU6
MobileNet V2
1 × 1
PW
ReLU6
3 × 3
DW
ReLU6
1 × 1
PW
Linear
"扩张"
"卷积"
"压缩"
http://blog.csdn.net/u011995719

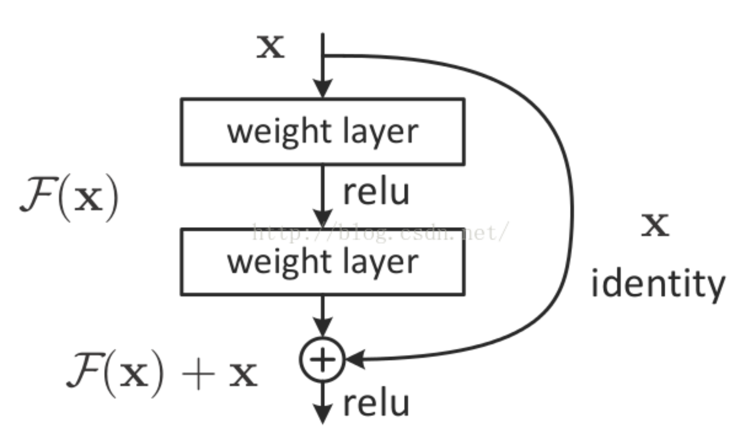


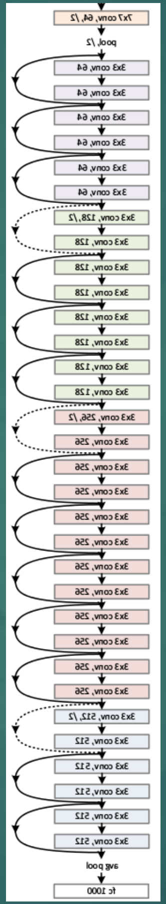
ResNet
C_{in}
1 × 1
PW
$0.25 \times C_{in}$
3 × 3
Standard
CONV
$0.25 \times C_{in}$
1 × 1
PW
$C_{out}(\approx C_{in})$
MobileNet V2
C_{in}
1 × 1
PW
$6 \times C_{in}$
3 × 3
DW
$6 \times C_{in}$
1 × 1
PW
$C_{out}(\approx C_{in})$
http://blog.csdn.net/u011995719

RESNET

猎维科技
LIEWEI.AI

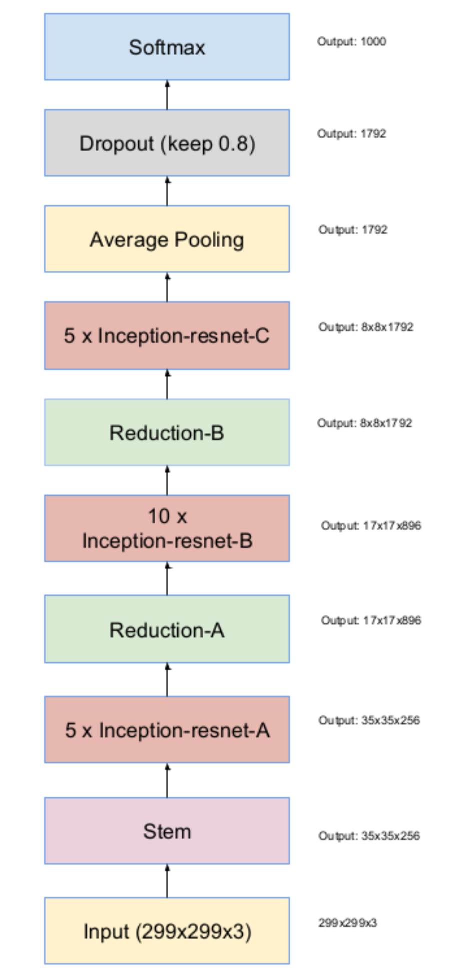






INCPTION - RESNET V2

https://arxiv.org/pdf/1602.07261.pdf



谢 谢