北京大学 PEKING UNIVERSITY

论文报告

- **《Joint Face Detection and Alignment using Multi-task Cascaded Convolutional Networks》**
  - 作者：Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, and Yu Qiao*
  - 单位：Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences
  - 发表：2016 SLP
- 1.Motivation
- 2.Novelty & Keypoint
- 3.Implementation

Ian Ren
ianren.ontheway@gmail.com

北京大学
PEKING UNIVERSITY

Motivation

- 1.前人工作没有考虑face detection与face alignment两任务的内在联系，即使考虑了，提出的算法结构也不够solid.
- 2.前人工作在hard sample mining的过程中，手工成分过大，不能做到end-to-end trainable.

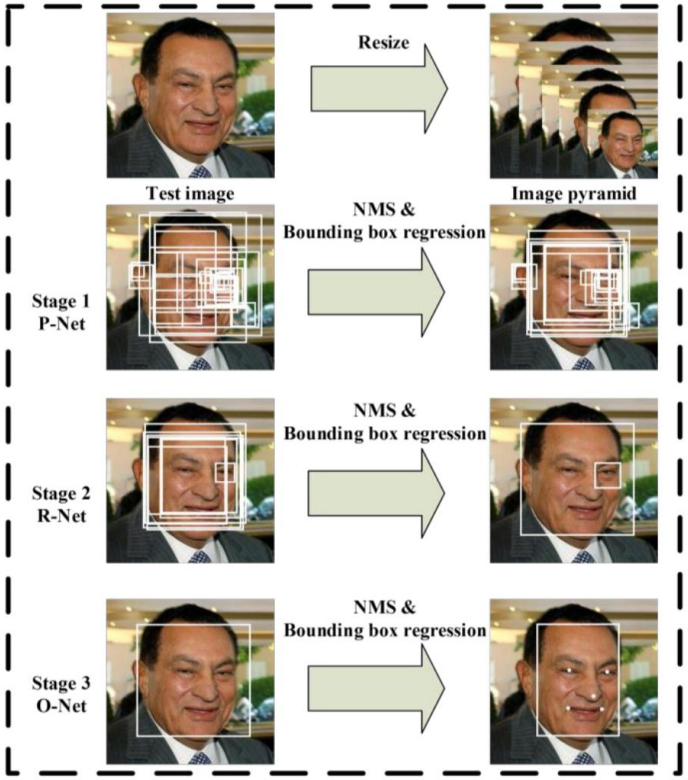
Ian Ren
ianren.ontheway@gmail.com

Novelty & Keypoint

- 1.提出一种级联CNN（MTCNN），联合实现face detection与face alignment, 并设计轻量级的CNN结构，保证网络预测的实时性。
  - 具体方法：
    - 提出3stage结构：
    - 1st stage: P-net(是一个FCN)，生成可能含有人脸的候选窗口，并利用bounding box regression校准候选窗口。最后通过非最大抑制（NMS）去除冗余窗口。

      （与faster rcnn中RPN的思路一致）
    - 2nd stage: R-net，继续执行bounding box regression与NMS过滤不含人脸的候选区域

      （与faster rcnn中ROI pooling要解决的问题一致）
    - 3rd stage: O-net，执行bounding box regression与NMS,并输出面部bounding box与5个面部关键点位置



Ian Ren
ianren.ontheway@gmail.com

Novelty & Keypoint

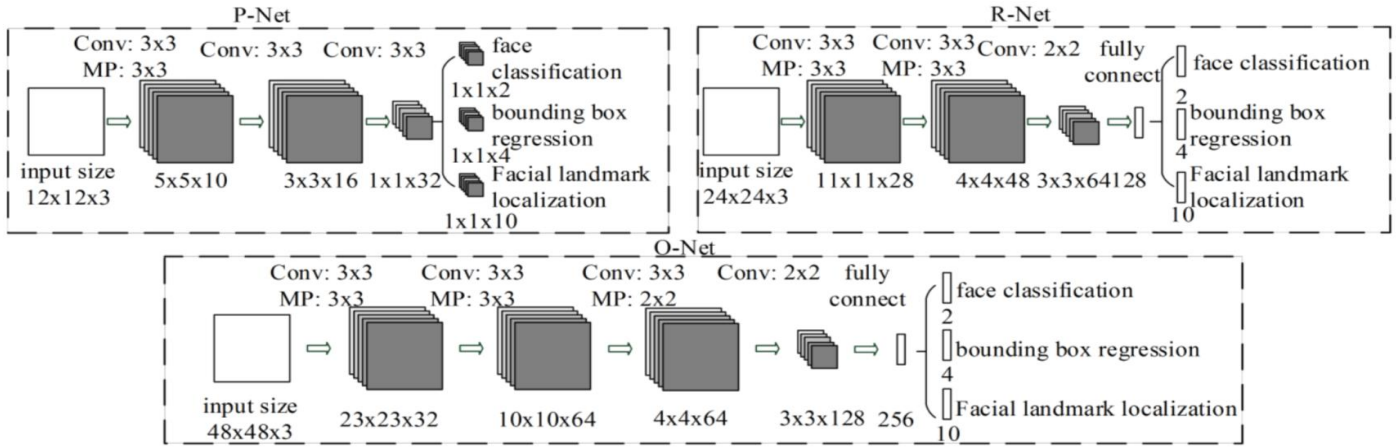
- 2.提出一种online hard sample mining（难例挖掘）方法，避免手工挑选困难样本，并忽略简单样本。
  - 具体方法：
    - 按loss排序，选择loss值最高的前70%的样本作为hard sample，在bp过程中只计算hard sample的梯度

- 3.采用轻量级CNN结构：网络中采用了更小的卷积核（3*3），减少模型参数，计算速度提升。

Ian Ren
ianren.ontheway@gmail.com

Implementation

- 实现想法：
  - 1.基本模型搭建
    - 根据论文提供的网络结构图，可以迅速搭建出基本模型



  - 2.关键算法实现
    - NMS(non-maximum suppression)
      - 在Faster RCNN里使用过，算法比较成熟
      - 计算各个box与其他box的重叠区域面积（IoU），若大于阈值，就将此box剔除

Ian Ren
ianren.ontheway@gmail.com

北京大学
PEKING UNIVERSITY

Implementation

- HSM(hard sample mining)
  - 可以参考《Training Region-based Object Detectors with Online Hard Example Mining》文中的方法：
  - 1. 首先构造一个HSM网络，将P-Net生成的所有候选框的feature map输入此网络，做一次前向传播，并得到对应的loss，此网络只用来做前向传播；
  - 2. 做一次非最大抑制(NMS)，去除冗余的候选框；
  - 3. 将剩余的候选框按照loss排序，选择loss值最高的70%的样本作为hard sample，并在之后的计算中，只将hard sample送入之后的网络（R-Net, O-Net）进行前向、后向传播。
- Generate_bbox
  - 根据feature map, 生成预测出的bounding box
  - 输入： 图像分类feature map、标注框回归feature map
  - 输出： bounding box坐标

Ian Ren
ianren.ontheway@gmail.com

Implementation

- Refine_bbox
  - 修正生成的bounding box，用于bounding box regression
  - 输入：生成的bounding box坐标、网络预测出的坐标
  - 输出：修正后的bounding box坐标

– 3.loss的实现

- 本文提出了4种loss函数，一种cross-entropy loss，两种Euclidean loss，一种Mutil-task training的loss，公式如下：
  - loss for P-Net: $$L_i^{det} = -(y_i^{det}\log(p_i) + (1 - y_i^{det})(1 - \log(p_i))) \quad (1)$$
  - loss for R-Net: $$L_i^{box} = \left\|\hat{y}_i^{box} - y_i^{box}\right\|_2^2 \quad (2)$$
  - loss for O-Net: $$L_i^{landmark} = \left\|\hat{y}_i^{landmark} - y_i^{landmark}\right\|_2^2 \quad (3)$$
  - loss for Mutil-task training: $$\min \sum_{i=1}^{N} \sum_{j\in\{det,box,landmark\}} \alpha_j \beta_i^j L_i^j \quad (4)$$

Ian Ren
ianren.ontheway@gmail.com

Implementation

- 4.preprocess data
  - 1.从网络结构图中看出，各个stage的网络的图像输入尺寸是不同的，因此需要将图像resize，得到image pyramid作为输入
  - 2.需要将图像分为positive、negative、part face、landmark face数据，来满足不同任务的要求

Ian Ren
ianren.ontheway@gmail.com

谢谢