|  |  |
| --- | --- |
| SSRNet算法解析 | 文件类型：技术文档 |
| 文件编号： |
| 面向的部门： 研发中心 |
| 保密等级：高 |
| 作者： 刘凯 |
| 日期：2019-06-10 |
| 版本：1.0V |

SSRNet人脸年龄估计算法解析



目 录

[1 SSRNet简介 3](#_Toc11249020)

[1.1 网络结构 3](#_Toc11249021)

[1.2 Soft Stage Regression 4](#_Toc11249022)

[1.3 仿真结果 5](#_Toc11249023)

[2 详细网络解析 5](#_Toc11249024)

[参考文献 1](#_Toc11249025)

# SSRNet简介

* 1. 网络结构

SSR-Net(Soft Stagewise Regression Network)[1]是一个年龄预测模型，它的输入是人脸矩形框，输出是预测的年龄。SSR-Net采用coarse-to-fine的策略，采用多级分类来估计结果。SSR-Net的模型很小，只有0.32MB，却能达到sota(state of the art)的性能。

SSR的网络结构如图1所示，其中前向有两条heterogeneous stream，每条stream由3个stage级联。每个stage的基本块由一个3×3的卷积层、BN层、激活层和池化层组成。其中stream1的通道数为32，激活函数为Relu，池化采用平均池化；stream2的通道数为16，激活函数为Tanh，池化采用最大池化。两个streams里每个stage的输出结果进入一个融合模块(Fusion block)，得到一组输出，最后的年龄估计则由所有的Fusion block的输出计算得到。Fusion block的结构如图2所示，功能是融合两条stream每个stage的结果，得到概率分布，偏置矢量和缩放因子 (见1.2节)。

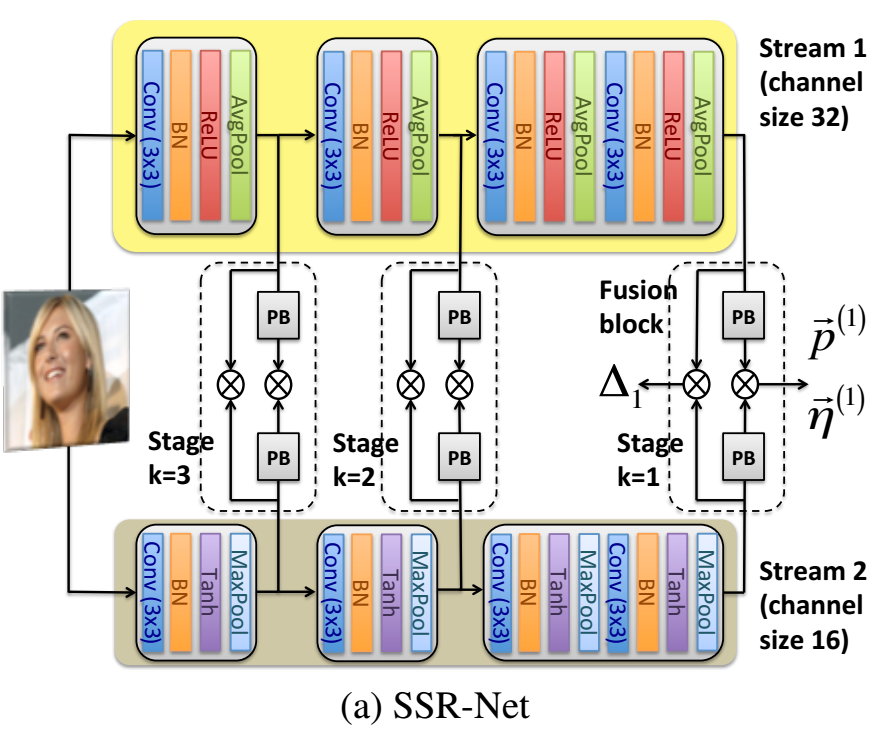


图1：SSRNet网络结构

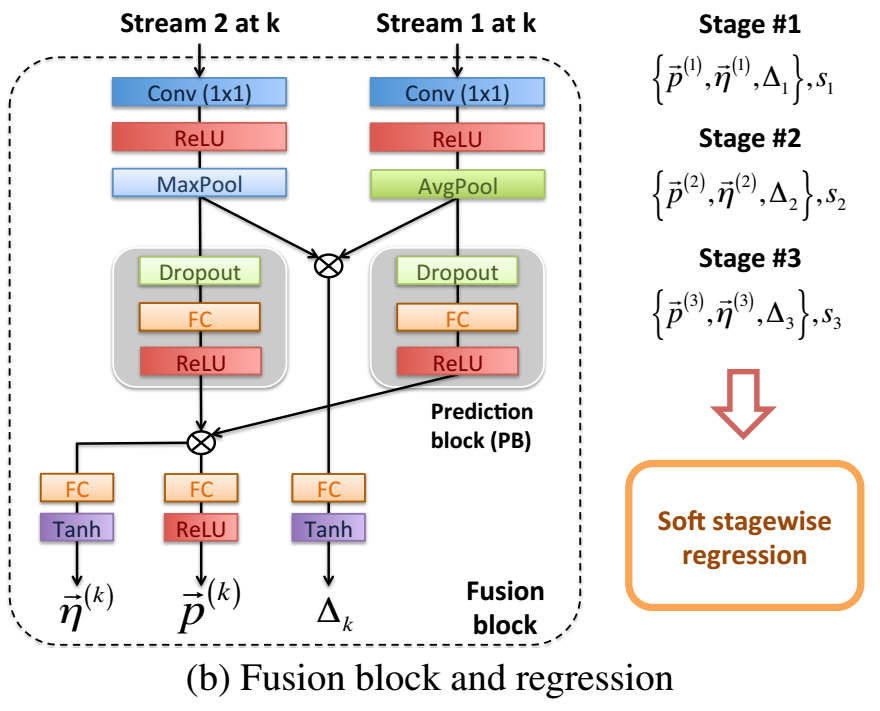
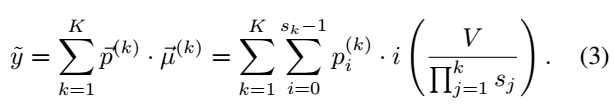


图2：Fusion block结构

* 1. Soft Stage Regression

设SSR有*K*个stage，第*k*个stage有*sk*个bins，且通过网络学习得到其概率分别为，则年龄估计值为：

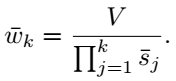


其中，*V*为总的年龄区间，则上式后面的括号表示每个bin的年龄宽度。一个例子，设年龄区间为1~90(*V*=90)，有K=2个stage，每个stage有*sk*=3个bins，则每个年龄bins的宽度为(*V*/(3\*3))=10。在stage1，分类器将年龄区间分为(1~30)，(31~60)和(61~90)；而stage2进一步分为(+0~10)，(+10~20)和(+20~30)。Stage1的粒度较粗，而stage2的粒度更细。采用分段回归的好处是每个阶段的回归器都很小，因此整个模型更紧致(compact)。

上面是将年龄区间分成不相交的bins。年龄模糊或年龄连续的情况会对结果有较大的影响，尤其是粗粒度的时候，情况更严重。因此作者将每个bins增加一个动态区间，使每个bins可以平移(shift)和缩放(scale)。首先是缩放：



其中，为缩放因子(scale factor)。相应的，每个bin的宽度为：



又引入偏移矢量，则bin index变为：



于是，给定融合模块每阶段的输出，估计的年龄为：



* 1. 仿真结果

SSRNet性能如图所示：



图3：SSRNet性能

# 详细网络解析

SSRNet的网络结构比较复杂，其总体架构如图4所示。输入data是维度为的人脸图片，输出output为估计的年龄(性别)。SSR分为*K*=3个阶段(Stages)，每个阶段会将年龄分为个bins，每个阶段有两条数据流(Stream)，每条数据流会各自经过一个**Filter**，其输出除了进入下一个stage外，还会作为 **FB**(Fusion Block)的输入。三个FB的输出进入**SSR**(Soft Stagewise Regression)模块，得到最终的输出。下面详细介绍SSRNet的Filter，FB和SSR模块

Stage 1和Stage 2的Filter模块相同，其结构如图5所示。Filter模块由一个的conv层，BN层，激活层和池化层组成，其中Stream 1的通道数为32，采用Relu激活和平均池化；Stream 2的通道数为16，采用Tanh激活和最大池化。Stage 3的Filter模块的结构如图6所示，除了上述四个模块外，还增加了Conv、BN和激活层，相当于深度增加一倍。

data

Filter1\_1

Filter1\_2

Filter2\_1

Filter3\_1

Filter3\_2

Filter2\_2

FB1

FB2

FB3

SSR

output

Stage1

Stage2

Stage3

Stream1

Stream2

图4：SSR年龄估计网络结构

conv

BN

Relu

AvgPool

conv

BN

Tanh

MaxPool

**Stream 1**

**Stream 2**

图5：Stage1(2)的Filter网络结构

conv

BN

Relu

AvgPool

conv

BN

Tanh

MaxPool

**Stream 1**

**Stream 2**

conv

BN

Relu

conv

BN

Tanh

图6：Stage3的Filter网络结构

Fusion Block的网络结构如图7所示。第*k*个Stage的输出作为FB模块的输入，输出结果为概率分布，偏置矢量和缩放因子。两条Stream会经过一个的卷积层，激活层和池化层，得到两个特征图。对于，上述特征图做元素相乘，再经过FC层和Tanh激活层，得到[-1, 1]的输出。对于和，上述特征图会再经过一组FC层和激活层，再做元素相乘，再分别经过两组FC层和激活层，分别得到和的输出。图7中的蓝色虚线框表示在Stage 3的FB模块在提取特征时，省略了激活和池化。

最后的SSR由1.2节的公式(1)得到。其中年龄的范围是[0, 100]，所以V=101。另外对于性别估计，一般女性为0，男性为1，所以最后V=1，输出的结果是性别为男性的概率。







conv

Relu

AvgPool

Dropout

FC

Relu

Flatten

conv

Relu

MaxPool

Dropout

FC

Relu

Flatten

FC

Relu

FC

Relu

FC

Tanh

FC

Tanh

stage3省略

图7：FB模块网络结构

参考文献

1. Tsun-Yi Yang, Yi-Hsuan Huang, Yen-Yu Lin, Pi-Cheng Hsiu, and Yung-Yu Chuang. Ssr-net: A compact soft stagewise regression network for age estimation. In IJCAI, 2018.