**基于深度学习的直播敏感内容方法研究**

摘要：本文使用人脸检测确定图像中人脸位置，根据人脸位置与大小推算躯干的位置以及大小提取ROI区，然后将该ROI区域送入yahoo开源的NSFW模型，最终获得直播内容的色情度。

引言：（语言需要更）

2016年被称为“中国网络直播元年”，从斗鱼开始到熊猫、虎牙等等各大直播平台相继出现，直播内容多样，对于直播的监管问题急需解决，很多主播在利益的驱使下色情直播屡禁不止。对于直播色情的检测迫在眉睫。传统主要利用肤色检测的方法，对皮肤的暴露程度对色情图像进行判断，然而肤色检测模型很容易受到光照影响，而且肤色检测需要对每一个像素点进行肤色判断，最终进行肤色分割，最终获得肤色检测的结果，时间复杂度较高不适合直播这种实时性较高的场景。而是用深度学习在训练模型时花费时间比较大外，使用训练好的模型对图像进行检测可以在很短的时间内获得较高的检测结果。本文使用的Yahoo的NSFW模型finetuned来源于ImageNet 1000 class dataset，使用的框架是resnet50 1 by2 (深度残差学习)具有很高的准确度。由于在直播过程中主播一般正面观众，使用人脸检测获得人体躯干排除背景的干扰具可以提高识别的准确性。

1 卷积神经网络

卷积神经网络是人工神经网络的一种，在语音分析和图像识别领域都有不俗的表现。目前Google的ImageNet模型大多是在CNN的基础上进行设计。卷积神经网络采用了三种基本概念：局部感受野（local receptive fields）、共享权值（shared weights）和混合（pooling）。

* 1. 局部感受野

使用局部感受野的原因有三点：

* 人眼在识别图像时，往往是从局部到全局
* 局部与局部之间联系往往不太紧密
* 可以减少网络需要学习的参数数量

对于一张图片，每一个像素相当于一个神经元。如图1.1（左）所示，局部感受野即图1.1（右）中5\*5区域。

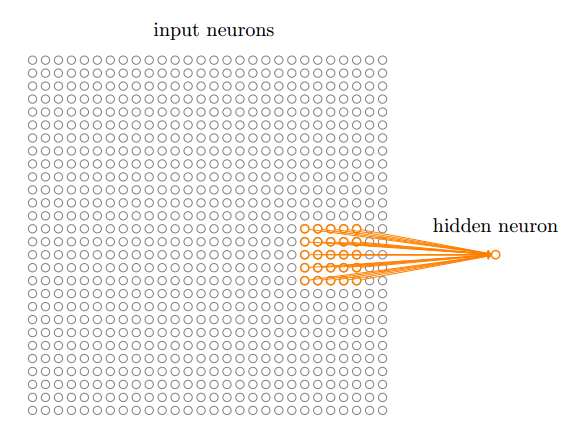
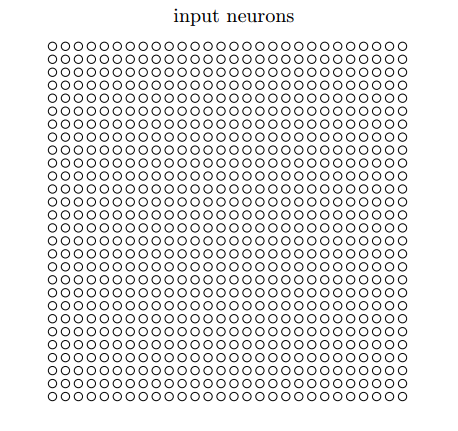


图 1.1：图像神经元示意图

* 1. 共享权值

即每一个局部感受野都有一个与之对应大小的权值以及偏置，在局部感受野移动过程该感受野的权值以及偏置的数值不变。共享权值和偏置被称为一个卷积核或者滤波器。每一个卷积核对应一个特征映射，一个完整的卷积层由几个不同的特征映射组成。共享权值与偏置可以大大减少卷积网络的参数。如图1.2所示一个28\*28的输入神经元，使用3\*5\*5的卷积核，输出3\*24\*24的特征映射。

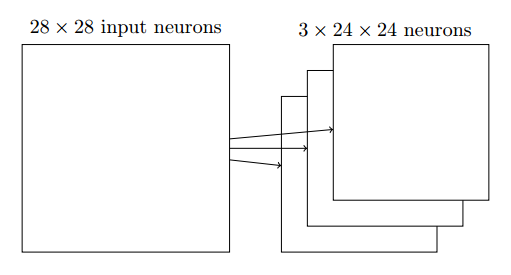


图 1.2 卷积层示意图

* 1. 混合

卷积网络的卷积层后是一个混合层，目的是简化从卷积层的输出信息。一个混合层取得从卷积层输出的每一个特征映射并且准备一个凝缩的特征映射，常见的混合有最大混合（max-pooling）、均值混合（mean-pooling）、L2混合即取激活值得平方和的平方根，比如一个特征映射的大小为x\*x，pooling的大小为y\*y，那么输出一个（x/y）\*（x/y）大小的数据，如图1.3所示：

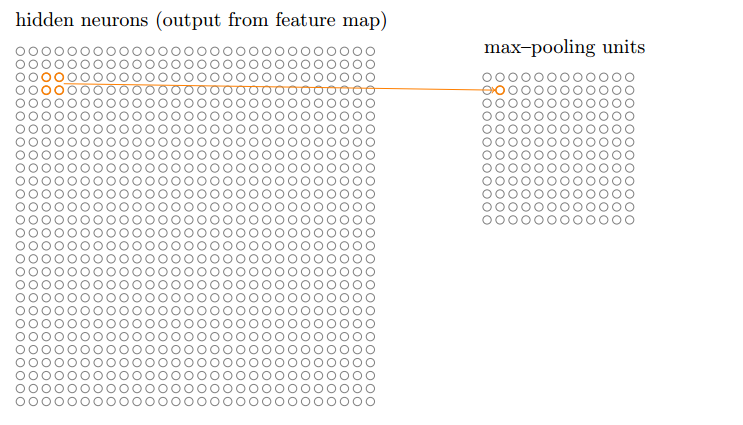


图 1.3 pooling示意图