**基于CNN的色情图像识别**

摘要：本文主要利用深度学习中的卷积神经CNN对色情图像进行识别。利用在2012年ImageNet获得冠军的alexnet网络模型进行微调后对收集的数据进行训练，最终获得识别裸女，与半裸女的分类模型，最终获得84%的测试结果。

引言：

传统的色情图像识别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。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。

**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。

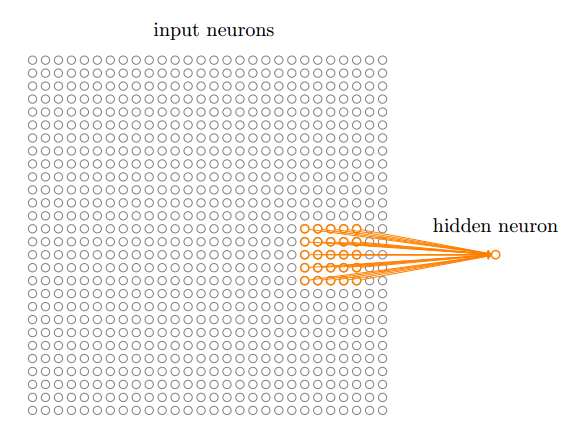
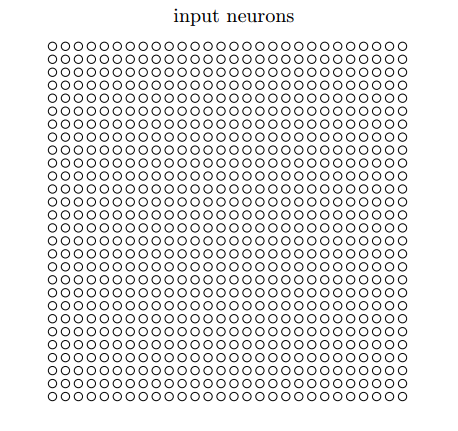


图 1.1：图像神经元示意图

* 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。

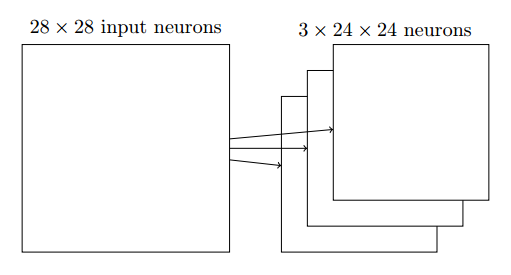


图 1.2 卷积层示意图

* 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：

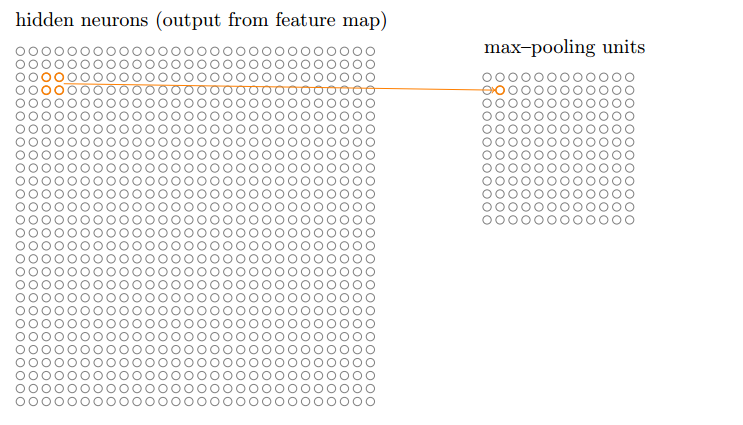


图 1.3 pooling示意图

1. **Alexnet网络模型**

Alexnet网络结构开启了数据集更大，更深的CNN时代，在传统的网络结构上进行了改进。包括：使用了ReLU非线性函数、局部响应归一化、重叠pooling等。

* 1. ReLU非线性函数

ReLU函数：f(x)=max(0,x) 被称为扭曲线性函数，是一种非线性的非饱和函数。在训练时间上，非饱和函数比饱和函数训练更快。这种扭曲线性函数能够保留非线性的表达能力，并且由于具有线性性质（如图2.1所示的正值部分），相比tanh和sigmoid函数在误差反向传递时不会出现由于非线性引起的梯度弥散现象。

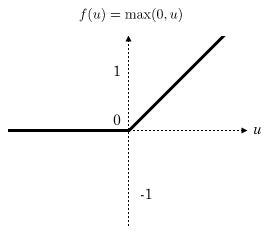


图 2.1 ReLU函数图

* 1. 局部响应归一化

即对每一组batch，在网络的每一层中，分feature对输入进行归一化处理。用表示点（x , y）处通过核计算出的神经元激活度，然后用ReLU非线性，响应归一化活性由公式=/ 给出。求和覆盖了n个相邻位于相同空间位置的核映射，N是该层中的核总数。核映射的顺序是任意的，且在训练开始前已经确定。这种响应归一化实现了一种侧向抑制，在使用不同核计算神经元输出的过程中创造了对大激活度的竞争。常数k，n，是超参数，数值需要用验证集确定，在模型中使用k=2，n=5，，。

* 1. 重叠pooling

CNN中的pooling层总结了同一种核映射中邻近神经元的输出。传统的pooling是不重合的，也就是stride==kernel size，在alexnet中stride<kernel size。

1. **实验**

3.1 实验网络模型

网络模型alexnet网络，该网络包括5个卷积层与3个全连接层，最后一个全连接层输出到2-way的softmax层以便分别裸女与半裸女。整体流程图如图3.1所示

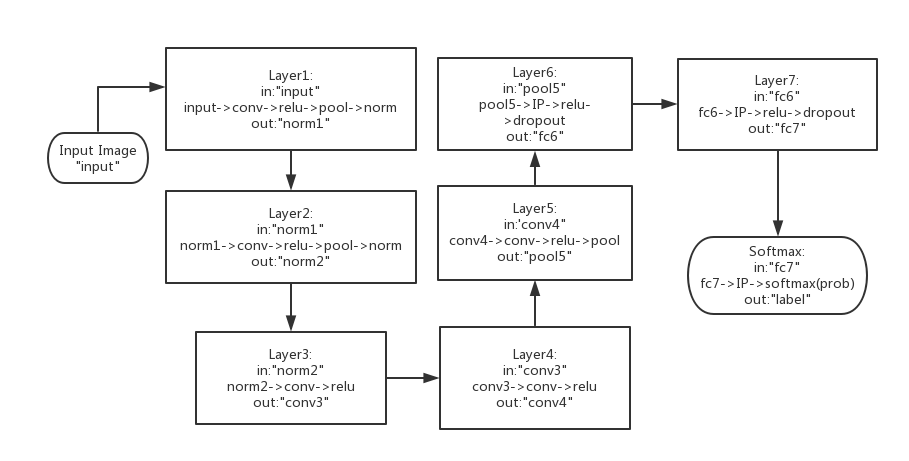


图 3.1：网络流程图

InputLayer即输入层，输入3\*227\*227大小的彩色图片，分RGB三个颜色维度输入。Layer1——Layer5是卷积层，对Layer1与Layer2的ReLU输出做了局部归一化处理，的其中Layer3与Layer的卷积结果并没有进行pooling，Layer6——Layer8是全连接层。

3.2 模型训练

共使用500张全裸与500张半裸的女性照片作为训练集，测试集全裸与半裸各50张，即训练集大小为1000，测试集的大小为100。由于数据集较少所以进行1200次迭代，由于使用CPU训练所以训练时间较长，差不多一个小时，测试准确度与迭代次数之间的关系如图3.2.1所示，可以发现当迭代次数达到800时基本饱和，最终能够达到84%的准确度。

测试loss与迭代次数的关系如图3.2.2所示：在迭代0-200区间下降较快，当迭代次数达到800时基本收敛，与达到最高准确率的迭代次数相同。

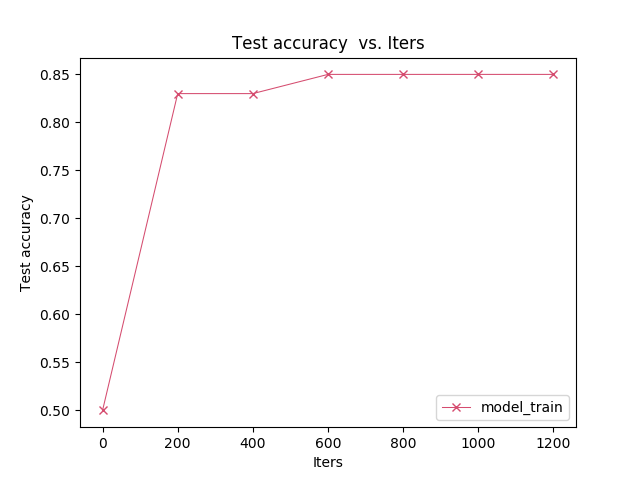


图 3.2.1 准确度与迭代次数关系图

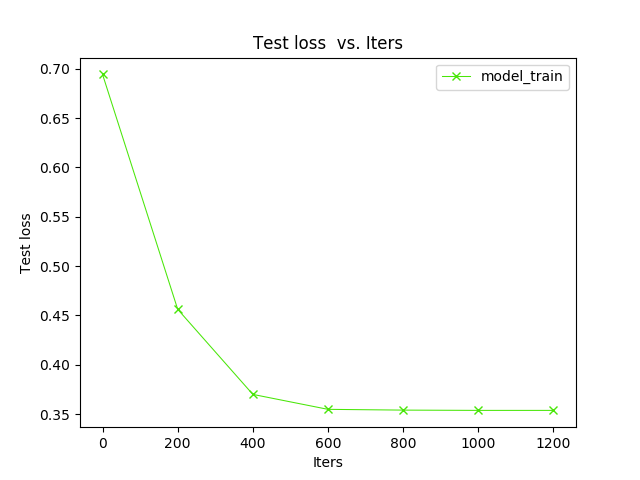


图 3.2.2 loss与迭代次数关系图

在测试一张全裸图片时可以观测网络中各层的输出，图3.2.3与图3.2.4左为第一层与第二层的卷积核图像，以及经过各卷积核后得到卷积的结果如图右所示，可以看到裸女的轮廓。

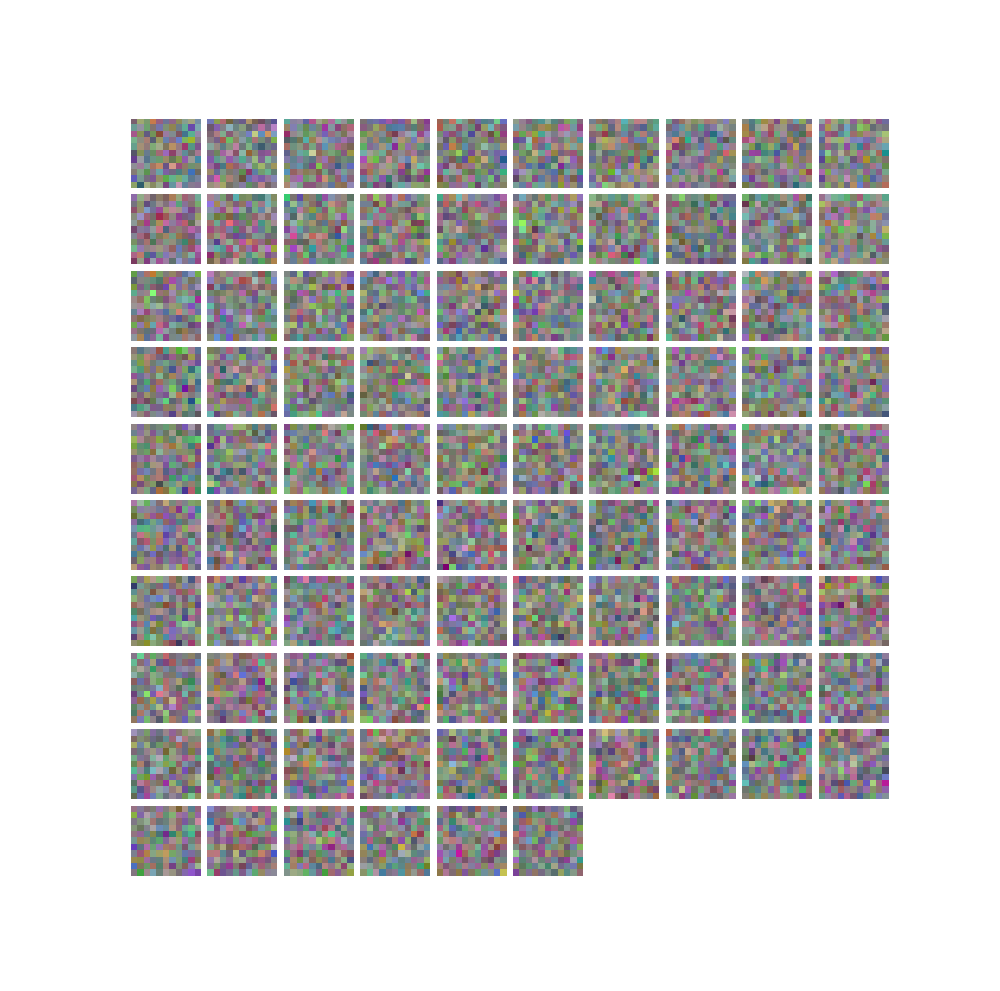
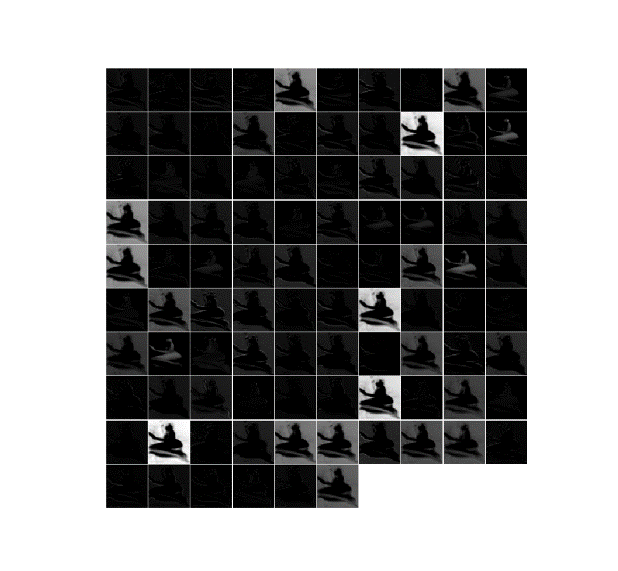
 

图 3.2.3 第一层卷积核与输出结果

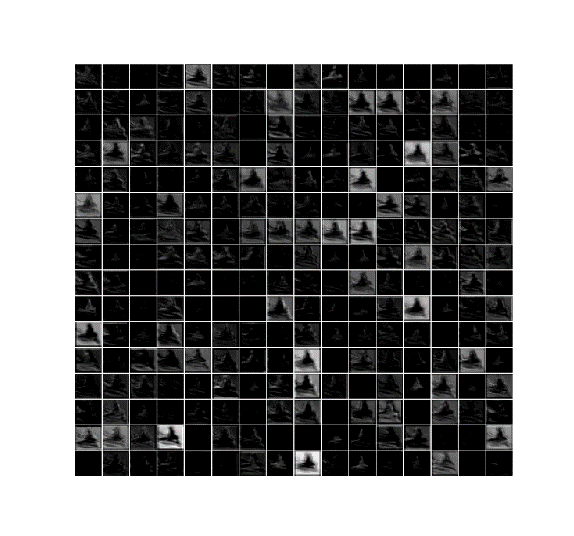
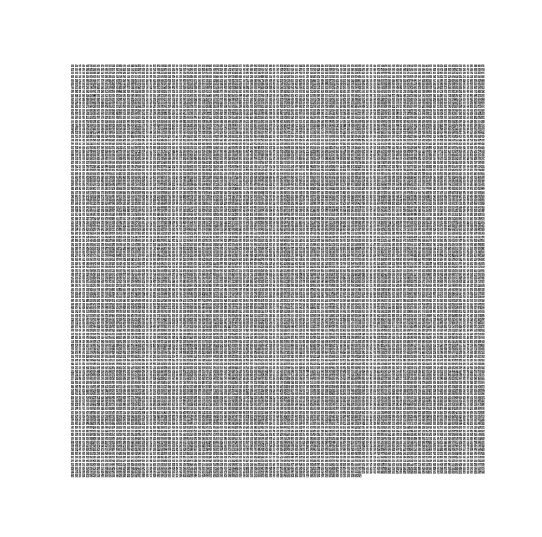


图 3.2.4 第二层卷积核与卷积结果输出

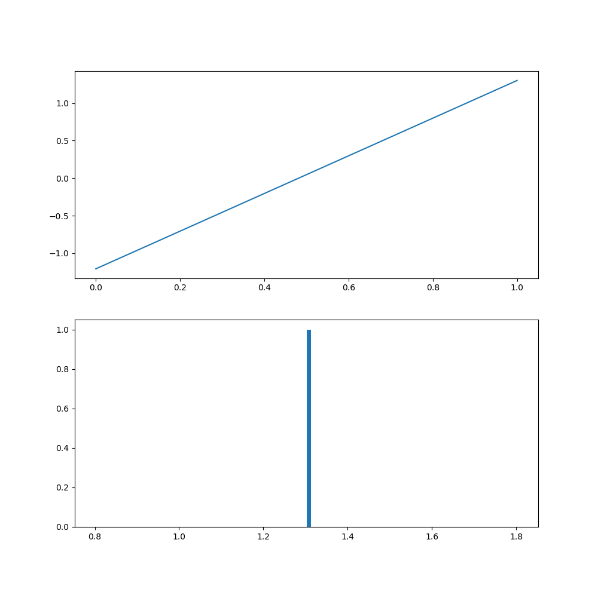
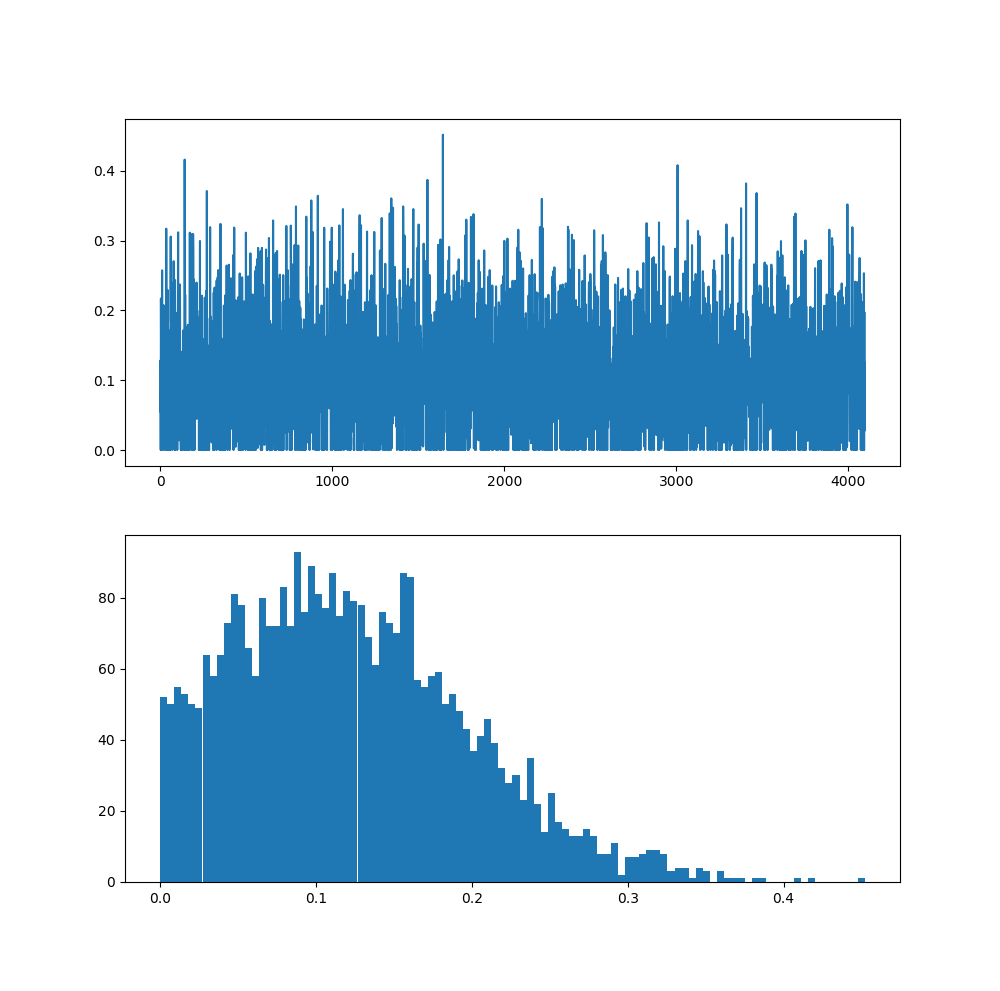


图 3.2.5 第七层与最后一层输出输出数值以及正样本直方图

3.3 结果分析

可以看出该模型可以有效的识别出全裸与半裸的女性，但是由于训练的数据集较少在某些场景下会出现误判的情况，以及只训练了女性的图像，男性图像并没有涉及。可以看出卷积神经网络在图像识别领域具有很强的识别率，仅仅使用1000张的训练集就可以达到84的训练结果，使用深度学习进行图像识别在模型训练时花费的时间较长，一旦模型训练完成使用现有的模型进行识别具有高效性，本文模型与使用肤色检测算法识别相同的图片花费1.7秒左右，而使用肤色检测算法用时2.5秒左右，具有高效性。

图

图