

基于卷积神经网络的人脸识别

骆广辉

华南师范大学数学科学学院 信息与计算科学

广州 510631

摘要：随着计算机科技与应用的发展，人脸识别作为一种常用的图像识别技术越来越受到关注。卷积神经网络的出现，大大提高了计算机对图像内容的分析和判断能力。通过结合卷积神经网络和人脸识别，形成一套以卷积神经网络为核心的人脸识别系统。

关键词：人脸识别；图像识别；卷积神经网络

中图法分类号：TP391.41

1 研究背景

随着计算机视觉的相关理论与应用研究的快速发展，计算机视觉技术在日常生活中应用的优越性也日益突显出来。用计算机对图像进行识别是计算机从相关的视频或图像序列中提取出相应的特征，从而让计算机“理解”图像的内容，并能正确分类的技术。安防意识的提升也让人们对于公共以及个人的安全需求不断攀升，使得计算机视觉在人脸识别、人脸检测等方面有了很高的应用价值。

人脸，与指纹、虹膜等一样，作为生物识别的一个重要特征，在档案管理系统、安全验证系统、公安系统的罪犯追踪、视频监控等方面有着很广阔的应用前景。虽然，相对于指纹识别、虹膜识别技术来说，人脸特征的唯一性要差一些，但是采集手段十分简单、方便、隐蔽，使用者也不会因为隐私等问题而产生抗拒心理。人脸识别技术具有以下优点：

（1）非强制性：用户不需要专门配合人脸采集设备，几乎可以在无意识的状态下被获取人脸图像，这样不会引起人的注意从而不会令人反感；

（2）非接触性：用户不需要和设备直接接触就能获取人脸图像，相对于指纹识别，更加安全卫生；

（3）直观性：当身份验证系统无法确定被识别者的身份或者无法对其正常完成识别时，工作人员一般会保留被识别者的信息进行后期人工核对，而人脸信息因为具有良好的直观特性，符合人的视觉特性，自然很容易对其进行辨别，但对指纹信息和虹膜信息，则无法识别；

（4）简易性：人脸识别系统应用摄像设备采集人脸信息进行识别，而对摄像设备的性能没有特殊要求，包括手机、摄像头在内的常见摄像设备都可使用，而且在不需要其他辅助设备的情况下就能完成识别。此外，摄像设备可安置在远处或者不易被人察觉的地方，避免被人恶意破坏。

目前，因为人脸特征的唯一性比较差，在对信息安全性的要求比较高的系统中，只能作为一般的辅助方法。然而在安全性要求相对较低的系统中，比如用于公安部门的罪犯追踪，普通的身份验证和鉴别系统等，人脸识别技术还是有用武之地的。人脸识别主要应用于以下几个方面：

（1）门禁系统：在需要受安全保护的地区通过人脸识别技术辨识试图进入者的身份，防止不可靠的人进入；

(2) 刑侦破案：工作人员通过一些途径获得某一嫌疑犯的相片或面部特征后，利用网络服务和人脸识别系统，在全国各地搜索逃犯，以便快速逮捕逃犯；

(3) 视频监控：在例如银行、机场、体育场、商场等公共场所对人群进行监视，防止恐怖分子的活动；

(4) 网络应用：利用人脸识别技术辅助信用卡进行网络支付，防止非信用卡的拥有者盗用信用卡；

(5) 人机交互：对个人计算机进行人脸识别开机，对手机进行人脸识别解锁，利用人脸识别进行真实感虚拟游戏等。

由此可见，人脸识别技术对于现实具有重大意义。虽然人脸识别技术已经发展了半个世纪，但是人脸识别技术依旧面临着姿态、表情、光照、遮挡等变化造成的影响的巨大的挑战。近年来，广泛应用于模式识别、图像处理邻域的卷积神经网络算法对这些影响具有一定程度的不变性，所以将卷积神经网络应用于人脸识别有很大的意义。

2 卷积神经网络

2.1 卷积神经网络模型

CNN 是一个多层的神经网络，由多个卷积层、池化层交替组成，而每层由多个独立神经元组成。神经网络模型如图 1 所示。

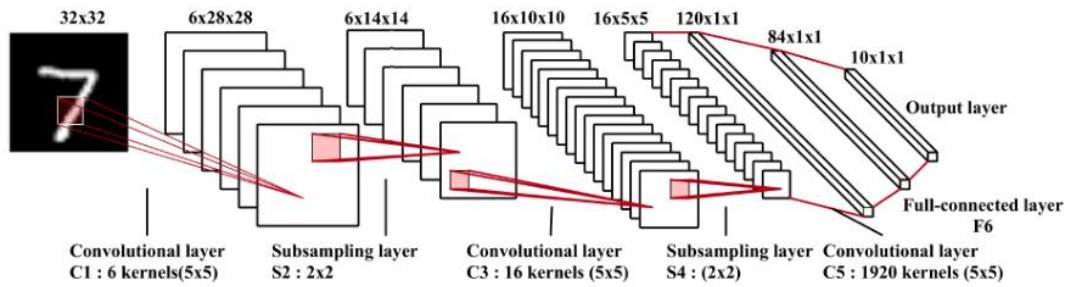


图 1 卷积神经网络模型

输入层：将 32×32 像素的 RGB 三通道图片作为输入 x ，CNN 的结构一般要根据输入图像的大小来进行设置。

卷积层 C_1 ：在输入层输入一个 32×32 像素的灰度图片，卷积层设计 6 个 5×5 大小的卷积核去卷积 32×32 大小的图片，得到 6 个 28×28 大小的二维特征图，然后通过 ReLU 激活函数激活。 C_1 层卷积操作的公式表达如式 (1)：

$$y_{c_1} = f(w_{c_1}x + b_{c_1}) \quad (1)$$

其中， w_{c_1} 表示卷积核， b_{c_1} 表示偏置项， $f(\bullet)$ 是 ReLU 激活函数。

池化层 S_2 ：使用 2×2 大小的 max_pool 滤波器，得到的结果为 6 个 14×14 的池化结果。计算过程为：

$$y_{s_2} = d(y_{c_1}) \quad (2)$$

其中, $d(\bullet)$ 表示 max_pool 函数。

卷积层 C_3 : 类似于 C_1 层, C_3 使用 16 个 5×5 的卷积核, 将获得 16 个 10×10 的特征图, 具体的卷积计算为:

$$y_{c_3} = f(w_{c_3}x + b_{c_3}) \quad (3)$$

池化层 S_4 : S_4 与 S_2 采用相同的 2×2 的滤波器, 提取的特征更加抽象, 最终得到 16 个大小为 5×5 的特征图。这些层的工作原理与前面是一样的, 只是随着深度的增加, 提取的特征更加抽象, 也更具有表达能力。 S_4 子采样计算过程为:

$$y_{s_4} = d(y_{c_3}) \quad (4)$$

全连接层: 把 S_4 的输出的二维特征图转换成一个一维向量。全连接层作为高度提纯了图像特征, 交给 Softmax 进行分类。而 Softmax 回归作为逻辑回归在多元分类问题上的推广, 它能够一次完成对多个类别的分类任务, 并且 Softmax 回归可以收敛到全局最小值, 避免了因局部收敛而导致的过拟合现象。Softmax 回归是一种有监督的单层神经网络的分类器, 目前它一般与深度学习的方法相结合从而用于分类识别问题。Softmax 的代价函数可以表示为:

$$p(i) = \frac{\exp(x_i)}{\sum_{k=1}^K \exp(x_k)} \quad (5)$$

其中, K 表示神经元的个数, x_k 为多个输入。通过此式每个输出都映射到了 0 到 1 区间, 可以看成是概率问题。

2.2 局部感知域

局部感知域, 普通神经网络会把输入层和隐藏层直接进行全连接的设计, 从整幅图像来计算相对较小的图像的特征是可行的, 但是, 如果是更大的图像, 这将会非常耗时。卷积层是解决这类问题的一种简便的方法, 是输入单元和隐藏单元间进行限制, 使得每个隐藏单元连接输入单元的一部分。对于输入图像, 每个隐藏单元可以连接输入图像的一部分相邻区域。

2.3 权值共享

权值共享是一种节省训练开销的策略。输入层是 32×32 的人脸灰度图像矩阵, 输出是其识别的结果, CNN 复合了多个卷积层和池化层对输入的图像矩阵进行处理, 而后再在全连接层进行实现与输出目标之间的映射。每个卷积层和池化层都包含着多个特征映射, 每一个特征映射一个平面, 平面由多个神经元构成, 通过一种卷积的滤波器用来提取输入的一种特征。如图 2, 输入图是 $7 \times 7 \times 3$ 的矩阵, 使用的过滤器是 $3 \times 3 \times 3$, 卷积步长为 2, 每个进行卷积的矩阵是 $3 \times 3 \times 3$, 分别平移, 通过滤波器后计算获得 $3 \times 3 \times 3$ 的输出图。

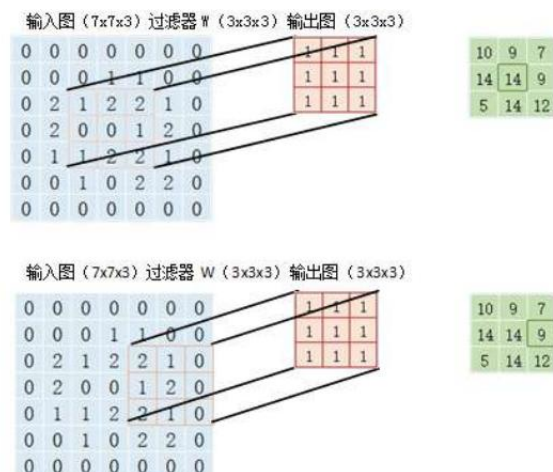


图 2 卷积的滤波器示例

2.4 池化

池化也称下采样，其目的是为减少特征映射，当在面临计算量非常大的时候，一个拥有过多特征输入的分类器不易形成，并且容易过拟合。由于卷积后的特征是一种静态属性，所以可能在两个不同图像区域的特征极可能一样。因此，描述大图像的时候可以对不同位置特征使用聚合统计。相对于卷积运算，池化常用的运算一般有：最大池化（Max Pooling）、均值池化（Mean Pooling）、高斯池化和可训练池化。最常见的池化规模 2×2 ，步长为 2，如图 3 所示，输入的每个 2×2 进行池化。每个 MAX 操作对四个数操作后分别获得一个值，池化操作的结果是使其深度大小不变。



图 3 最大池化示例

3 人脸识别系统

我们整个人脸识别系统总共分为 5 个部分：图像采集、人脸检测、数据整理、卷积神经网络的构建和训练、人脸实时识别。

3.1 图像采集

在卷积神经网络训练之前，首先得有数据。我们通过 opencv 调用电脑摄像头拍取约 10 个人的人脸照片，每人拍 600 张。为拍照的 10 个人分别建立一个文件夹，并将其所拍照片统一放置该文件夹中，文件夹以起名字拼音命名，最后将这 10 个文件夹统一放置于一个总文件夹中，并以“faceImages”命名。示意图

如下：

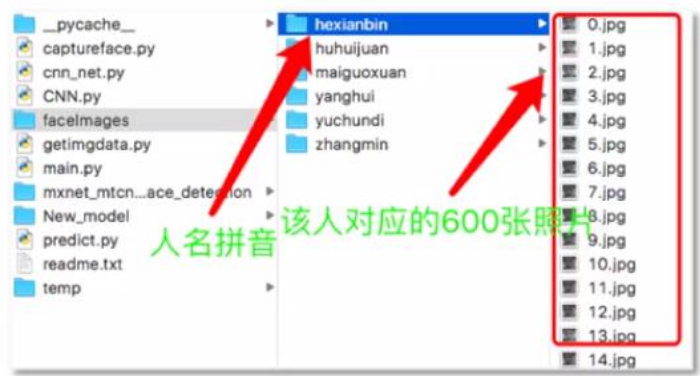


图 4 人脸数据示意

3.2 人脸检测

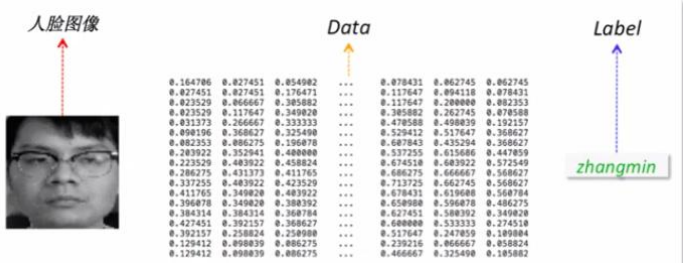
我们需要的只是人脸区域，对于背景无关的图像，我们应该去掉。因此我们通过 MTCNN 算法，框出人脸图像中的人脸部分。然后对检测结果进行灰度转换，最后统一图像大小 32×32 。灰度化是为了减少其他因素的干扰，方便卷积神经网络对人脸特征的提取，同时也减少了图像数据量的大小，减少了计算时间。



图 5 人脸检测

3.3 数据整理

该部分就是对人脸图像灰度图整理成 numpy 中的 ndarray 格式，并将整理后的数据按 8:2 划分成训练集和测试集。数据整理示意如下：



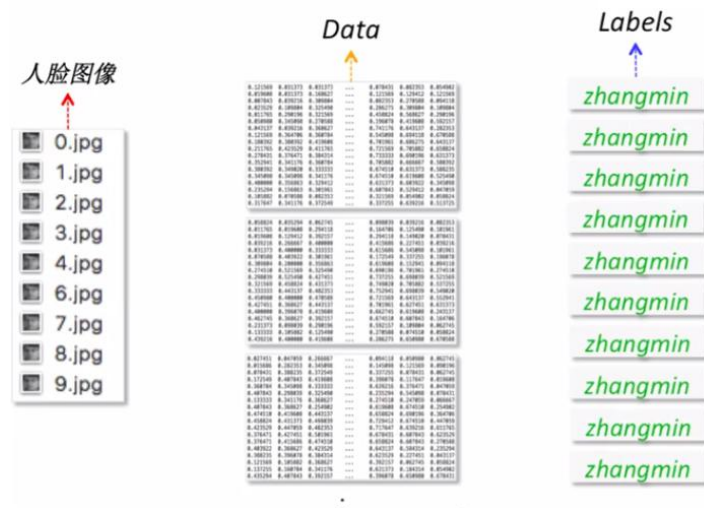


图 6 数据整理

3.4 卷积神经网络的构建和训练

我们构建的卷积神经网络采用了 2 层卷积层，2 层池化层，还有 1 层全连接层和输出层。

在经过参数调试和优化之后，我们在第 1 次卷积层设置 32 个 7×7 的卷积核，并设置步长为 1，并且设置补充图像边缘，经过 1 次卷积后得到的图像大小仍为 32×32 。第 1 次池化层采用 2×2 的 filter，步长为 2，得到的图像大小为 16×16 。第 2 次卷积操作的设置和第 1 次一样。第 2 次池化操作和第 1 次一样。

在完成了卷积和池化操作之后，我们将其提取出来的图像特征信息展开一个长度为 $8 \times 8 \times 64 \times 1024$ 的一维数组输入到全连接层中进行分类识别。

该过程使用的激活函数为 Relu 函数，分类函数使用 softmax 函数。

由于人脸识别是一个多分类问题，我们增加了对输出结果的 softmax 值与真实的 label 值进行交叉熵计算作为损失函数进行梯度下降的训练。为了提高训练的效率并且防止局部最优解的问题，我们采用了 mini-Batch Gradient Descent 的训练方法。除此之外为了避免过拟合，我们使用了 dropout 的训练方法。

3.5 人脸实时识别

将训练好的卷积神经网络进行保存，当需要对人脸图像进行识别时，只需将人脸图像进行人脸检测，然后将其输入到训练好的卷积神经网络中，神经网络就会根据输入数据进行计算并输出最终识别的结果。

4 实验验证与结果分析

4.1 数据来源

我们采用的人脸图像是周围同学的人脸拍照。一共 10 个人，每人 600 张，图像大小为 640×480 的彩色图，经过预处理后为 32×32 灰度图。按照每一个人 8:2 的比例进行训练集和测试集的划分，方式为随机，标签采用独热编码表示。

4.2 代码实现

我们使用 python 中的 TensorFlow 工具包对所设计的卷积神经网络进行编程实现。

4.3 实验结果分析

我们对卷积神经网络相关的参数进行各种调整，结果如表 1 所示。

Filter Size	Dropout Rate	Training Accuracy	Test Accuracy
5×5	1.0	1.00	0.99
5×5	0.5	1.00	0.99
7×7	1.0	1.00	0.99
7×7	0.5	1.00	0.99
4×4	1.0	1.00	0.92
3×3	1.0	1.00	0.90

从结果来分析，可以发现卷积核大小对结果有影响，这是因为卷积核太小，观察到的信息范围小，对整体把握不准确。当训练次数足够多的时候，dropout 的大小对结果影响不大，然后训练次数低的时候，越大的 dropout 结果会更准确。

通过上表我们选择了 7×7 的卷积核，dropout=1.0 的参数来训练卷积神经网络。

5 结 语

基于卷积神经网络的人脸识别算法已经具有很高的识别准确率了，能够在较少数据和较少网络层数的情况下达到 99% 的准确率。但是存在的问题是训练时间过长以及特殊样本的错误性，这些问题仍待改进。

参考文献：

- [1] 段萌. 基于卷积神经网络的图像识别方法研究. 郑州大学.
- [2] 蒋帅. 基于卷积神经网络的图像识别. 吉林大学.
- [3] 钱程. 基于深度学习的人脸识别技术研究. 西南交通大学.
- [4] 张少华. 基于深度卷积神经网络的人脸基准点定位研究. 华中科技大学.
- [5] 万士宁. 基于卷积神经网络的人脸识别研究与实现. 电子科技大学.