

人脸图像检测算法研究与实现

密西根学院 宋世唯 5123709279

指导教师： 电子信息与电气工程学院 刘满华

摘要

人脸检测是在图像中确定人脸是否存在并找出其位置的过程。随着电子技术的广泛应用，人脸检测逐渐成为一门受到重视的研究课题。而人脸检测所使用的方法也越来越多。

在本文中，我们重点研究了两种人脸检测算法。肤色是人脸的一个显著特征，利用 YCrCb 颜色空间，我们实现了肤色区域分割；卷积神经网络是深度学习的一个分支，我们设计了一个网络训练其中的权重，以完成分类检测。

我们使用肤色分割和窗口扫描结合的方法，并在 Caltech 等图像库进行实验。我们分析比较了两种算法的问题，并提出了一些可以改进的地方。

关键词： 人脸检测，肤色分割，卷积神经网络

ABSTRACT

Face detection is the process of determining whether faces exist in an image and locating all existing faces. With the wide spread of electronic technology, face detection has become a hot topic. Also, more methods are designed in face detection.

In our research, we focus on two face detection algorithms. Skin color is one significant feature of human face. By transforming into YCrCb color space, we can do the color skin segmentation. Convolutional neural network is a branch of deep learning. We design a CNN and train its weight with some training samples to get the final network for detection.

We combine the skin color segmentation and window scan together. We experiment on the Caltech picture database and some other pictures. We find some problems and discuss probable improvement.

Keywords: face detection, skin color segmentation, convolutional neural network

1. 人脸检测概述

1.1 人脸检测简介

人脸检测是指在被检测的图像中确定所有人脸的位置、大小的过程^[1]。人脸检测的研究起源于对人脸识别的研究。人脸识别通过人眼睛，鼻子，嘴巴等特征识别不同的人。而人脸检测便是人脸识别的第一步。之后，人脸检测逐渐演变成一个独立的研究方向。目前，人脸检测在视觉监控、身份认证等方面有着广泛的应用。

1.2 常用人脸检测方法

1.2.1 基于人脸模式特征提取的方法

人脸图像包含众多特征，其中最常见有肤色特征，轮廓特征，五官特征等等。基于特征提取的方法试图提取某种或多种特征以进行人脸检测。我们之后将提到的基于 YCrCb 颜色空间的方法^[2]就是利用了肤色特征。

1.2.2 基于统计模型的方法

这种方法将检测人脸问题转化为检测人脸或非人脸的分类问题。然后通过大量带标签的人脸与非人脸样本进行有监督学习，以构造分类器。这种方法将问题简化为一个二分类问题，并利用模板特征进行分类。在实际应用中，这种方法被广泛地应用。常见的特征模板有 Haar-

like 特征^[3], LBP 特征^[4]等。

1.2.3 基于神经网络的方法

神经网络将模式的统计特性隐含在网络的结构和参数中。近年来,随着深度学习成为一个热门研究方向,基于神经网络的方法也被更多地研究和应用。本文中,我们也将详细研究一种基于卷积神经网络的人脸检测算法。

1.3 人脸检测评价指标

在人脸检测中,我们常用以下指标^[5]来评定一个检测系统的好坏

- 检测率: 正确检测到的人脸数与图像实际包含人脸数的比值。
- 误检率: 被错误检测为人脸的数目与图像数目的比值。
- 检测速度: 完成人脸检测的时间。
- 鲁棒性: 系统在各种条件下的适应性。

我们的目标是提高检测率,降低误检率。在这个基础上,我们希望尽量提升检测速度并且加强算法的鲁棒性以适应更多图片情况。

1.4 本文的研究概述

在本文中,我们主要研究并实现了两种算法。其中基于 YCrCb 颜色空间的肤色分割用于预处理,基于卷积神经网络的算法则用于窗口检测。我们将两种方法结合起来成为了一套人脸检测算法。

我们使用 Matlab 实现了算法并在人脸数据库和实际图像中进行了实验。我们将检测结果与另一种方法(基于 Adaboost 算法的人脸检测算法)进行了比较和分析,提出了一些问题和改进的方法。

2. 基于 YCrCb 颜色空间的肤色区域分割

肤色是人脸的一个明显特征,在人脸检测算法中,我们也可以利用肤色进行检测。然而,由于环境复杂,照片质量层次不齐,光照情况不确定等因素影响,要准确地使用肤色来完成人脸检测相当困难。因此,在我们的实践中,利用肤色完成一个肤色区域的选择和分割,为之后使用其他算法检测做预处理。

2.1 YCrCb 颜色空间

在肤色检测中,传统的 RGB 颜色空间由于上文中提到的种种原因,很难准确地将肤色区域与其他区域分割。而 YCrCb 颜色空间在肤色分割上则有不错的效果。

YCrCb 颜色空间中,Y 表示亮度,Cr 表示红色信号与信号亮度值的差异,Cb 表示蓝色信号与信号亮度值的差异。在这里,我们主要使用 Cr 和 Cb 作为区别肤色区域与其他颜色区域的工具。

2.2 确定阈值

我们使用了 50 张各种肤色,亮度,质量的人脸肤色区域作为训练样本。对所有样本区域计算平均值,得到了最终阈值:

$$\begin{aligned} 137 &\leq Cr \leq 163 \\ 100 &\leq Cb \leq 126 \end{aligned}$$

2.3 肤色区域分割

利用训练得到的阈值,我们可以对一张图片进行肤色区域分割。以下图为例



图 1 原始图像

- 将输入图片由 RGB 模式转化为 YCrCb 模式
- 遍历整张图片，将在阈值范围内的点标记为白色，反之标记为黑色。



图 2 肤色分割后二值化图像

- 框出白色区域，并记录这些区域作为之后检测的待检测区域。



图 3 肤色区域

3. 基于卷积神经网络的人脸检测算法

在这一章中，我们将介绍一种使用卷积神经网络的人脸检测算法。在这种算法中，我们使用卷积神经网络提取人脸特征，利用 BP 算法进行训练，我们可以不断调整网络权重，优化训练结果。

3.1 神经网络

3.1.1 人工神经网络

人工神经网络^[6] (Artificial Neural Network, ANN) 试图模拟人脑的神经网络结构。简单来说, 它是由神经元和神经元之间的连接构成的网络。如图, 是一个最简单的神经网络。

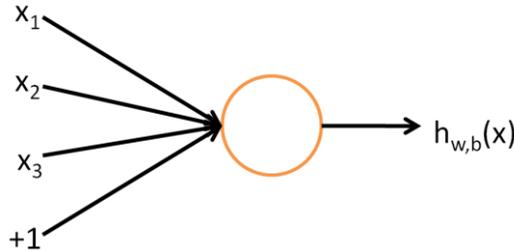


图 4 神经元

它包含 3 个输入 $\mathbf{x}=(x_1, x_2, x_3)$, 一个神经元, 一个输出 $h_{w,b}(\mathbf{x})$ 。我们可以通过调整连接的权重 $\mathbf{w}=(w_1, w_2, w_3)$ 和神经元的偏置 b 来获得我们希望得到的输出。该神经网络的输出为

$$h_{w,b}(\mathbf{x}) = f(\mathbf{w}^T \mathbf{x} + b) = f\left(\sum_{i=1}^3 w_i x_i + b\right)$$

其中 $f(z)$ 为激励函数。在神经网络中, 我们往往在神经元的输出加入激励函数以使输出改变线性的特性。我们选用的是常用的激励函数 Sigmoid 函数, 其表达式为

$$f(z) = \frac{1}{1 + \exp(-z)}$$

神经网络可能由多层, 大量神经元复杂连接构成。如图为一个三层神经网络, 它由输入层, 中间层和输出层构成。我们可以对网络中的每一个连接训练权重, 对每一个神经元训练偏置量以使网络满足我们的要求。

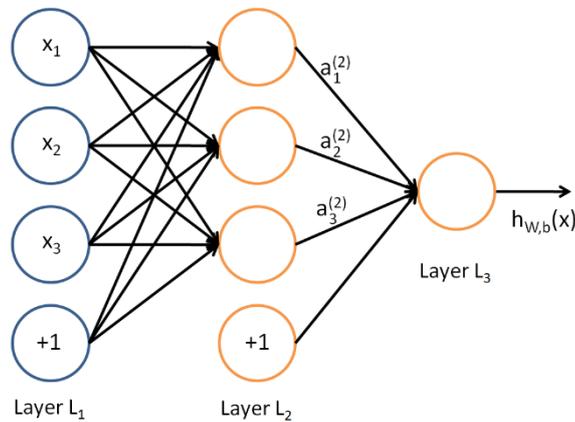


图 5 多层神经网络

3.1.2 卷积神经网络

卷积神经网络 (Convolutional Neural Network, CNN) 是多层神经网络的一种。它利用卷积运算的卷积核作为神经网络的权重。卷积神经网络的一个特点是它是局部连接的, 也就是说每个神经元只和上一层的部分神经元连接。卷积网络的另一个特点是权值共享, 也就是说每个神经元与上一层的连接权重是相同的。这两个特点极大降低了连接的维度, 同时确保了位置不变性。为了提取更多特征, 我们可以在每一层使用多个卷积核进行特征提取。每一个卷积核对输入图像卷积可以得到一个特征图。

在卷积神经网络中, 有两种层: 卷积层和下采样层。它们分别起到特征提取和特征集合

的作用。

在卷积层中，我们利用卷积运算提取特征。卷积核便是我们要训练的权重，我们将卷积核放置于不同位置，便能得到不同的连接。如图所示为一个简单的卷积层。我们的卷积核为 3*3，输入为 5*5，输出为 3*3。

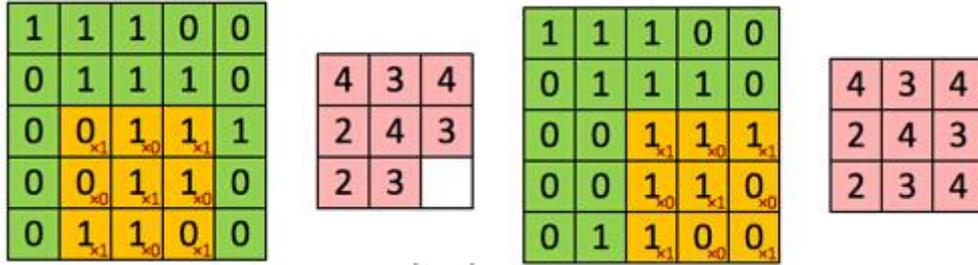


图 6 卷积核与卷积

下采样层通过在特征图中采样达到集合特征的目的。它的基本思想是整合多个输入节点得到一个输出。常用的采样方法有最大采样 (max-pooling) 和平均采样 (average-pooling)。分别使用提取范围内最大值和计算范围平均值的方法进行采样。在我们的实现中使用了平均采样。

3.2 卷积神经网络结构

在这一节中，我们将介绍我们所使用的卷积神经网络的整体结构。整个网络输入 24*24 大小的灰度图，输出为 2*1 的向量，当判断为人脸时输出 (1, 0)，非人脸时输出 (0, 1)。参考 LeNet-5 网络^[7]，我们的整个网络的结构如下图。

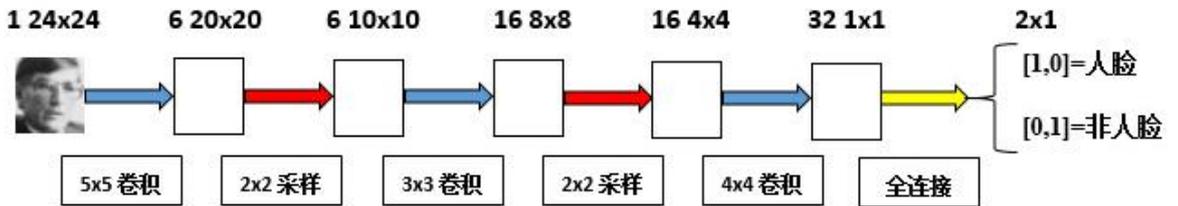


图 7 卷积神经网络结构

卷积层接下采样层接卷积层接下采样层接卷积层接全连接输出。全连接层中将 32 个节点与输出层的 2 个节点全连接得到输出。

3.3 训练卷积神经网络

卷积神经网络的训练即确定卷积核以及神经元偏置的过程。我们使用反向传播 (BP) 算法进行训练。

3.3.1 BP 算法

BP 算法是训练神经网络的一种常用算法。其基本思想为误差的反向传播。输入信号正向传播达到输出层。从输出层计算出误差后，误差从输出层反向传播，每一层根据误差调节权重。调节权重采用的是梯度降低方法。下面我们依据^[8]，推导出调节权重的公式。

我们首先定义几个值。第 1 层的输入，也就是第 l-1 层的输出为

$$x^l = f(u^l), \quad \text{且 } u^l = w^l x^{l-1} + b^l$$

其中 $f(z)$ 为激励函数， w 为权重， b 为偏置。

对于第 n 个样本，输出层误差为

$$E^n = \frac{1}{2} \sum_{k=1}^c (t_k^n - y_k^n)^2$$

其中 c 为输出层节点个数，对于我们的人脸检测问题， $c = 2$ 。 t 表示目标值， y 表示输出值。

我们定义反向传播的误差 δ 为

$$\frac{\partial E}{\partial b} = \frac{\partial E}{\partial u} \frac{\partial u}{\partial b} = \delta$$

输出层的误差为

$$\delta^L = f'(u^L) \cdot (y^n - t^n)$$

其他层的误差递推式为

$$\delta^l = (w^{l+1})\delta^{l+1} \cdot f'(u^l)$$

这里的 \cdot 运算符为元素分别相乘运算符。

我们要使用误差 δ 求出每一次迭代中 E 对 w 与 b 的偏导然后乘上学习率 α 就能得到 e 与 b 的改变量，即

$$\Delta w = -\alpha \frac{\partial E}{\partial w}, \quad \Delta b = -\alpha \frac{\partial E}{\partial b}$$

- 卷积层
误差为

$$\delta_j^l = \beta_j^{l+1} (f'(u_j^l) \cdot \text{up}(\delta_j^{l+1}))$$

其中 l 为层数， j 表示第 j 个特征图。 $\text{up}(x)$ 为上采样，它将 x 复制到对应的下采样区域内。

偏置的偏导为

$$\frac{\partial E}{\partial b_j} = \sum_{u,v} (\delta_j^l)_{uv}$$

卷积核的偏导为

$$\frac{\partial E}{\partial k_{ij}^l} = \sum_{u,v} (\delta_j^l)_{uv} (p_i^{l-1})_{uv}$$

其中 $(p_i^{l-1})_{uv}$ 表示在第 $l-1$ 层中被 k_{ij}^l 卷积的元素。

- 下采样层
误差为

$$\delta_j^l = f'(u_j^l) \cdot \text{conv}(\delta_j^{l+1}, \text{rot180}(k_j^{l+1}))$$

偏置的偏导为

$$\frac{\partial E}{\partial b_j} = \sum_{u,v} (\delta_j^l)_{uv}$$

系数的偏导为

$$\frac{\partial E}{\partial \beta_j} = \sum_{u,v} (\delta_j^l \cdot \text{down}(x_j^{l-1}))_{uv}$$

其中， $\text{down}(x_j^{l-1})$ 为下采样的值。

以上为通过 BP 算法调整权重的方法。

4.3.2 训练步骤

我们选择了 8500 个人脸样本以及大量非人脸样本。部分人脸样本如图。

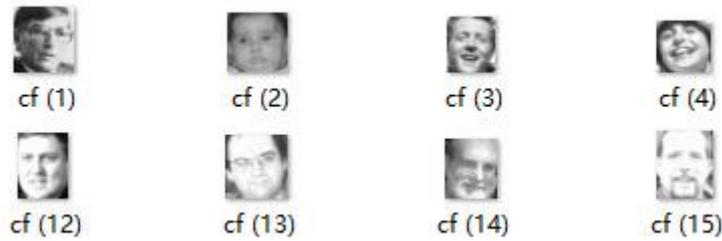


图 8 部分人脸训练样本

训练分成 6 轮，每一轮使用 8500 个人脸样本。非人脸样本则由两部分组成：一半由上一轮的非人脸样本中随机选择；另一半则是新加入的上一轮会误检的非人脸样本。

对于每一轮的训练，会将样本随机分成 50 个一组。每次将一组放入网络正向计算输出，然后反向传播修正权重。所有样本全部用完后，再次将样本分组反复进行训练。共迭代 50 次。

第一轮训练前，所有权重被初始化为略大于 0 的随机数。每一轮训练会在上一轮训练的网络权重基础上继续训练。另外，我们采用的学习率 $\alpha=1$ 。

4. 人脸检测

有了以上算法作为模块，我们可以拼接出一个完整的人脸检测算法了。算法的输入是一张任意大小包含任意人脸数量的彩色图片，输出是框出的人脸区域。

4.1 图像预处理

读取图片后，我们首先对过大的图片缩小一定比例。根据观察分析，人脸区域在图片中所占区域不会过小（最小约为 $30*30$ ）。同时，缩小图片对检测速度有很大的提升。

我们将得到的彩色图片输入到肤色分割以获得待检区域。同时对图像进行灰度化，为之后的检测做准备。

4.2 肤色分割

我们利用前面章节中提到的 YCrCb 颜色空间将肤色区域分割出来。此时，我们将得到多个待检测区域，我们可以将这些区域看作是一张张图，作为下一步的输入。

4.3 窗口扫描

在得到待检测区域后，我们需要对每一个区域进行扫描。由于人脸大小和位置是不确定的，我们必须不断调整图像大小扫描所有区域。检测窗口的大小由训练时的窗口大小有关，在我们的实现中，CNN 算法的检测窗口为 $24*24$ 。检测窗口以可选择步长在区域中扫描。完成扫描后，按可选择比例缩小图像，再进行扫描，直至区域小于窗口大小。这里值得注意的一点是在许多算法检测窗口时需要做一个归一化以获得更好的检测效果。而卷积神经网络由于其本身特性则不需要进行这一操作。我们记录下所有被检测为人脸的窗口大小和位置。

5.4 合并窗口

由于我们以不同大小和位置进行扫描，同一人脸可能被多次检测，因此我们需要合并窗口。合并窗口需要解决两个问题：是否需要合并和如何合并。

我们判断合并条件是：重合达到一定的程度。具体的判断条件是重合并且一个窗口的中心被另一个包含。这一条件可以有效地将同一人脸的窗口合并为一个，并且可以避免邻近的不同人脸被合并。

我们尝试使用的合并方法有两种，一是计算重合窗口的平均值，二是取重合窗口的最大值。用计算平均值的方法合并窗口可以得到更加准确的窗口，如图所示，均值法合并得到的窗口更小，更精确。而使用取最大值的方法可以避免人脸区域内的误检造成的不良影响，如图所示，嘴部被误检为人脸，若采用最大值法则可避免这一问题对最后结果的影响。最终，

我们使用了最大值法进行窗口合并。



图 9 合并前



图 10 最大值合并与均值合并结果

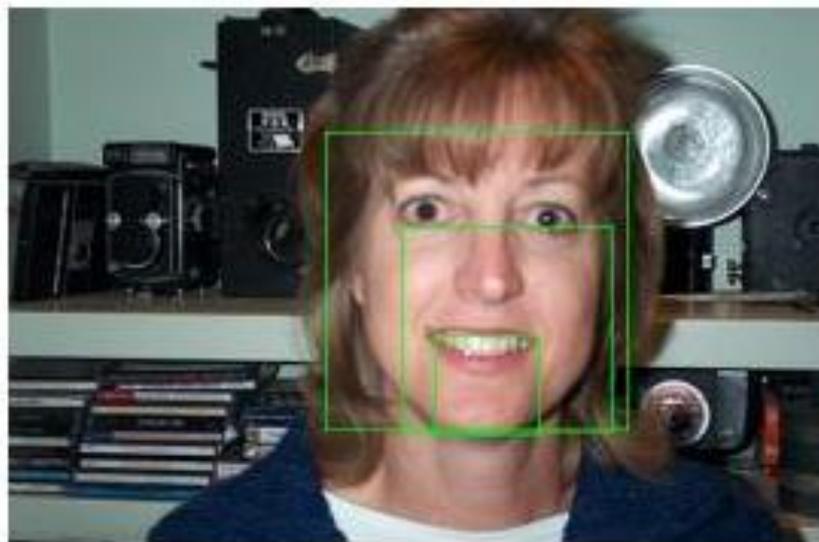


图 11 合并前

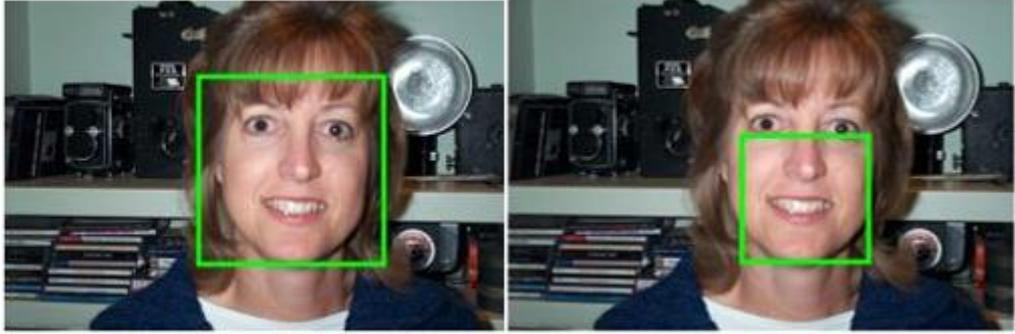


图 12 最大值合并与均值合并结果

5. 实验结果

5.1 结果比较

我们分别使用 Adaboost 和卷积神经网络算法检测。实验所用的数据库是 Caltech 人脸库，包含 450 张单人脸样本。

检测结果如下

算法	检测率	误检率	检测时间
Adaboost 算法	96.22%	12%	
卷积神经网络	98.44%	2.22%	

表 1

我们选取了部分不同情况下的检测效果图

- 部分遮挡

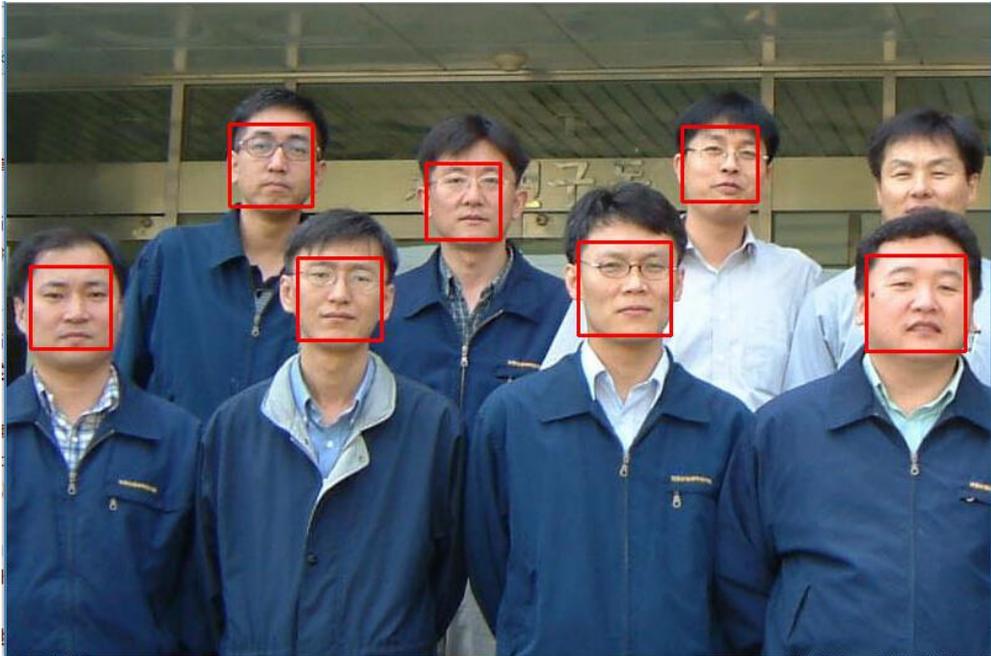


图 13 Adaboost 算法检测结果

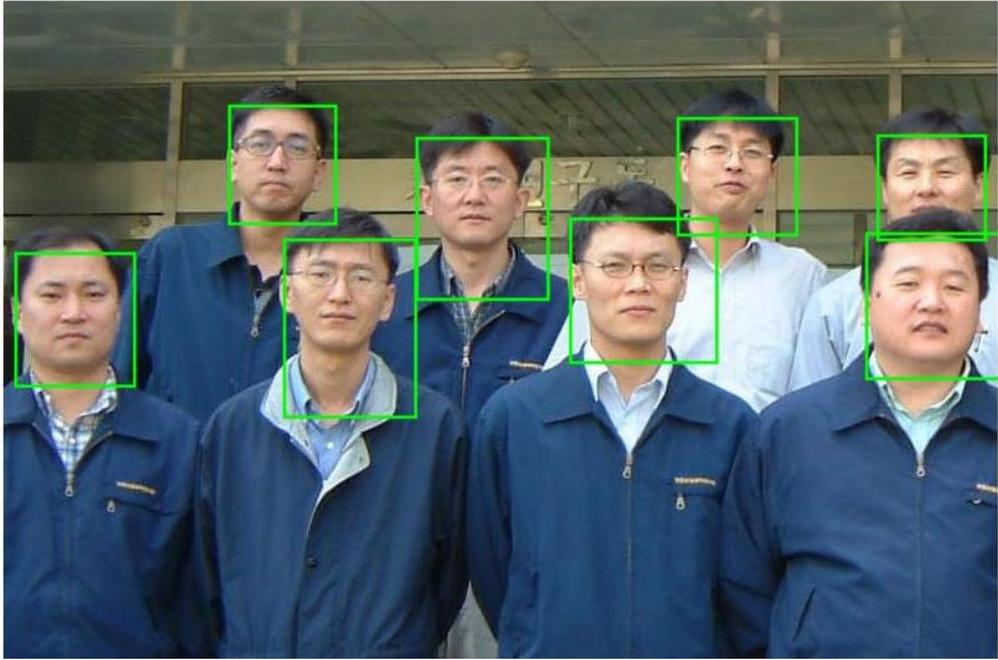


图 14 CNN 检测结果

- 大量人脸



图 15 Adaboost 算法检测结果



图 16 CNN 检测结果

- 光线昏暗



图 17 Adaboost 算法与 CNN 检测结果

从我们的实验结果来看，卷积神经网络算法无论在检测率还是误检率上都更优秀。这一结果可能存在一些疑问。首先，两种算法的检测窗口大小不同。第二，为了达到各自更好的训练效果，两种算法的训练样本（人脸样本）不同。第三，由于运算能力和时间的限制，两种算法的训练仍有改进余地。

在鲁棒性方面，卷积神经网络算法的优势则更为明显。从上面的检测结果看，在光照情况不良或环境复杂的情况下，卷积神经网络检测效果明显优于 Adaboost 算法。这一点从两种算法的基本原理便可得出。基于 Haar-like 特征的 Adaboost 算法尽管经过归一化处理在面对复杂环境光照时仍显得不利。而卷积神经网络自行提取特征，不仅避免归一化处理，在环境适应性上也更好。

5.2 问题

在训练和实验中，我们也发现了现存结果的一些问题

- 图片颜色异常造成的肤色分割错误
由于图片质量层次不齐，有些图片色温，色差有些奇怪，导致肤色分割时不准确，进而影响检测结果，如图



图 18 肤色分割后的二值化图像

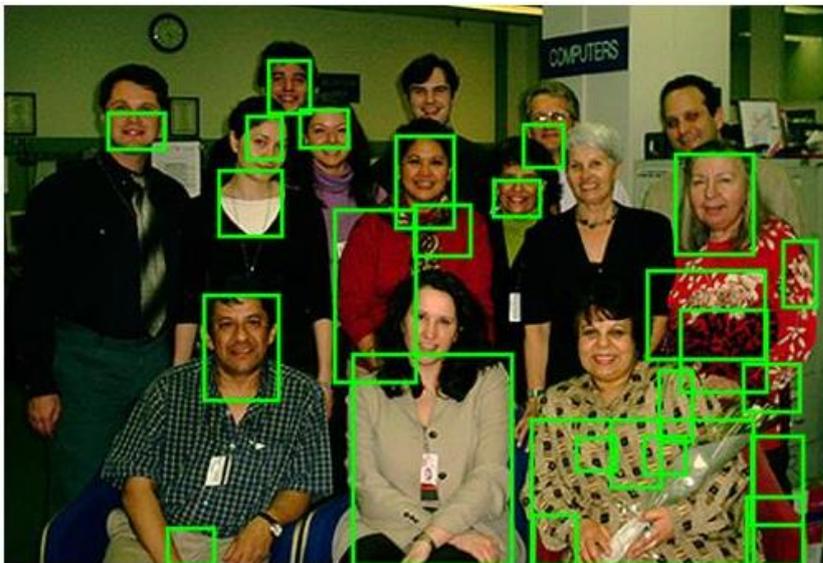


图 19 肤色区域

- 非正面人脸或倾斜人脸

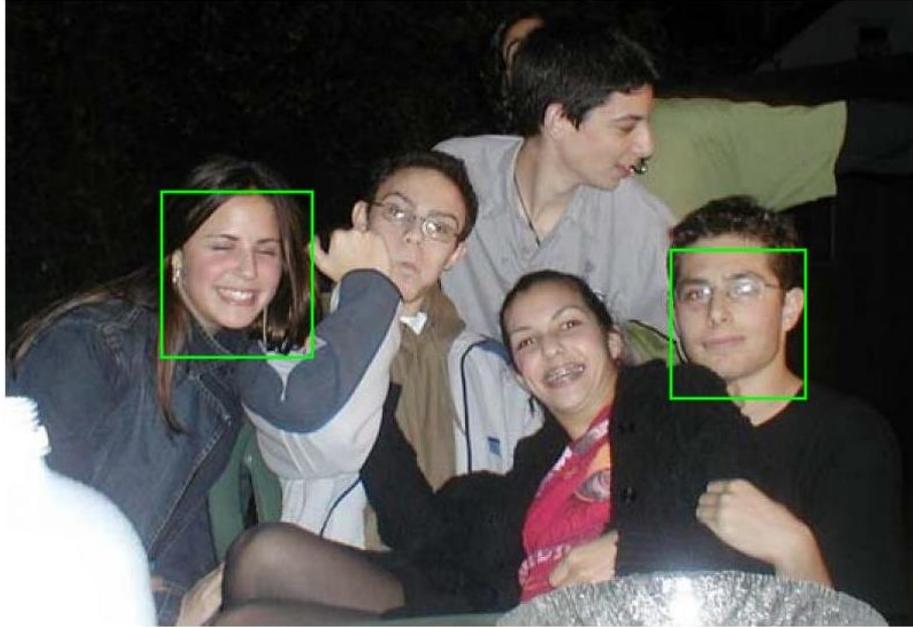


图 20 倾斜，侧面及遮挡人脸

该图中，三个未检测出的人脸为倾斜，侧面及遮挡。

- 多人脸，复杂环境



图 21 多人脸检测

此类图像容易发生漏检，误检情况，合并时也有可能出现问题。

6. 总结与展望

6.1 总结

在这次人脸图像检测算法研究与实现项目中，我们主要研究并实现了三种算法。其中基于 YCrCb 颜色空间的肤色分割用于预处理，基于 Adaboost 算法和卷积神经网络的算法则用于窗口检测。

我们在 Caltech 等人脸库进行实验。我们发现在我们的训练和实验条件下，基于卷积神经网络的算法在检测率和误检率上都优于基于 Adaboost 算法的人脸检测算法。尤其是在鲁

棒性方面，卷积神经网络体现出了明显更好的适应性。

6.2 展望

针对在训练检测中我们的一些经验体会以及遇到的问题，我们认为在未来的研究中可以做出以下的改进。

- 图像颜色预处理
在输入图片后，做一个颜色的预处理。针对一些色温，饱和度不正常的图片，这一操作十分有必要。
- 二值化图像的处理
在通过肤色阈值范围二值化图像后，可以对图像进行一些处理。比如腐蚀与膨胀，有利于更准确地定位肤色区域和去除更多杂色区域。
- 训练样本
在我们的实验中，两种算法的训练达到一定程度后，再训练的效果就变得不理想。这一点可能与训练样本有关。由于误检率已经相当低，导致获取训练样本变得困难，所得的样本质量也难以保证。这可能可以通过更多更好的训练样本改善。
- 检测速度
由于需要多大小全区域扫描，需要检测的窗口数量十分巨大，这也导致检测时间仍然较长。可能通过其他方式，加速检测。

参考文献

- [1] 梁路宏, 艾海舟, 徐光祐, 张钊. 人脸检测研究综述. 计算机学报, 2002, 25(5):449-458.
- [2] 王华清, 基于改进 YCrCb 色彩空间的人脸检测算法, [学位论文], 重庆交通大学, 2010.
- [3] P Viola, MJ Jones, Robust real-time face detection, International journal of computer vision, 2004, 57(2):137-154.
- [4] MA Rahim, MN Hossain, T Wahid, MS Azam, Face recognition using local binary patterns (LBP), Global journal of computer science and technology, 2013, 13(4).
- [5] 王超, 李东, 雷震, 人脸检测技术的研究进展, 电脑知识与技术, 2009, 5(9):7211-7213.
- [6] 朱大奇, 人工神经网络研究现状及其展望, 江南大学学报(自然科学版), 2004, 3(1):103-110.
- [7] Y LeCun, L Bottou, Y Bengio, P Haffner, Gradient-based learning applied to document recognition, Proceedings of the IEEE, November, 1998.
- [8] J Bouvrie, Notes on convolutional neural networks (unpublished), November, 2006.