

利用 MLP 和 FFT 实现快速脸部检测^{*}

罗小燕, 侯晓微, 胡海峰, 熊银根
(中山大学电子与通信工程系, 广东 广州 510275)

摘要: 提出了一种对静止图像进行快速脸部检测的方法。通过将多层感知器隐层神经元视为滤波器, 利用快速傅立叶变换的性质, 把时域卷积转换为频域乘积形式, 在保持较高检测率的情况下提高了检测速度。实验证明, 利用本文提出的算法可以获得较好的结果。

关键词: FFT; MLP; 训练样本; 脸部检测

中图分类号: TN919.8 **文献标识码:** A **文章编号:** 0529-6579 (2003) S2-0116-04

当前, 随着多媒体及信息技术的不断发展, 如何通过自动脸部检测技术实现人机交互和进行身份识别成为一个热门话题。脸部检测在这些问题的实现中起了重要的作用。所谓脸部检测, 就是要确定图像中是否存在包含人脸的子图像, 并找出其所在位置。

人脸是一个非刚性物体, 加上摄像时内、外条件的变化, 使我们从摄像头得到的人脸图像有很大的自由度。这些变化主要表现在: ①外部条件: 摄像头位置的改变或其内部参数以及外界光照强度的变化都会影响所拍摄的图像质量; ②内部条件: 对个体而言, 是否佩戴眼镜、长有胡须等可有可无的脸部特征以及脸部表情的变化都给检测带来了困难^[1,2]。

脸部检测实质上可认为属于模式识别的范畴, 它将要识别的物体分为两大类: 脸部和非脸部图像。人工神经网络是完成模式识别的有利工具, 因而也是目前普遍应用的一种脸部检测方法, 它通过对样本的学习不断修正神经元之间的连接权值, 以完成识别功能^[3,4]。

1 人工神经网络的结构和训练

人工神经网络的识别功能很大程度上取决于学习样本, 对神经网络用不同的样本进行训练, 它将具有不同的识别功能。所以, 第一个要解决的问题就是如何获得更具有代表性的训练样本^[5]。

1.1 训练样本的获取

要利用神经网络对脸部/非脸部图像进行识别, 训练样本库中必然包含这两种图像。脸部图像库的

收集相对简单, 但应满足以下要求:

(1) 脸部库中的训练样本图像有固定大小 (一般 $20 \times 20 \sim 30 \times 30$ 像素), 且包含所有重要的脸部特征信息, 如图 1 所示。



图 1 脸部库训练样本

Fig. 1 Training example

(2) 要使获取的脸部图像库保持一定的性别比例, 并尽可能包含人脸所有的可能特征, 比如是否有胡须、是否佩戴眼镜等, 还要将人脸各种不同的表情变化也纳入其中。只有这样得到的人脸训练样本才最具有代表性。

非脸部图像的获取要相对复杂一些, 因为自然界中任何不包含人脸的图像都可归入非脸部图像库, 数量庞大, 而实际上我们也不可能获得一个完整的非脸部图像库来对神经网络进行训练, 但我们仍然希望所获得的非脸部图像库具有一定的代表性, 为此, 我们采用了 bootstrap 的方法来构建训练样本中的非脸部图像库: 即先任意选取一部分自然/人工景物图像构成最初的非脸部图像库, 与所获得的脸部图像库一起对神经网络进行训练, 训练完成之后, 用该神经网络对任意图像进行检测, 当发生误检测 (即将非脸部图像认为是脸部图像) 时, 就将该图像也加到非脸部库中, 然后, 再用新收集到的脸部/非脸部图像库对该神经网络进行再次训练, 如此循环几次后, 则可获得一个检测精度较高

* 收稿日期: 2003-04-14

基金项目: 广东省自然科学基金资助项目 (001172)

作者简介: 罗小燕 (1978 年生), 女, 硕士研究生; 通讯联系人: 熊银根, E-mail: isde34@zsu.edu.cn

的神经网络^[9]。

1.2 神经网络的结构

本文中的神经网络采用了多层感知器的结构 (MLP)，共有 3 层，如图 2 所示。输出层用一个神经元实现脸部/非脸部图像的分类，若输出为“1”表示检测结果为脸部图像，否则输出为“0”；输入层神经元的个数取决于训练样本的选取，而隐层神经元数一般根据经验确定。用 BP 学习算法对神经网络的连接权值进行调整。

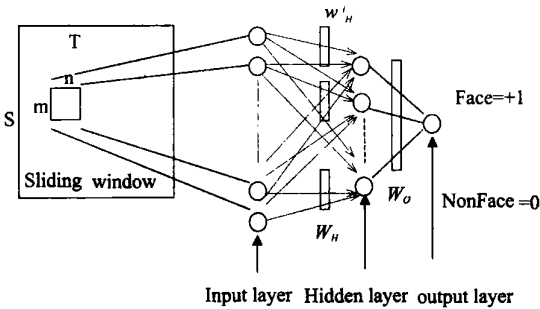


图 2 神经网络的结构
Fig 2 Classical MLP architecture

设 W_H 、 W_o 为隐层和输出层的权值矩阵， w_{ij} 表示第 i 个神经元与第 j 个输入之间的连接权。而权向量 w_H^i 为与第 i 个神经元相连接的所有权值，是由隐层和输出层各神经元的阈值组成的列向量，有以下公式：

第 i 个神经元的净输入值为：

$$\text{net}_i = \sum_{j=1}^n w_{ij} x_j \quad j = 1, 2, \dots, m \tag{1}$$

隐层输出为：

$$H = g(W_H X + B_H) \tag{2}$$

X 为输入向量。

神经网络的输出为：

$$O = f(W_o H + B_o) \tag{3}$$

2 用人工神经网络实现脸部检测

2.1 用人工神经网络实现脸部检测的传统方法

用神经网络实现脸部检测的经典方法一般是这样做的 (如图 2 所示)：人们事先训练好一个神经网络，使它能检测出一幅固定大小 (20×20~30×30 像素) 的图像是否是脸部图像。然后，让滑动窗口在整幅输入图像中逐个像素移动，窗口移动到某处，则将所有包括在窗口内的图像信息作为输入向量送给神经网络进行判断；然后改变图像分辨率进行同样的检测，如此来完成脸部检测的任务，这样的搜索，理论上可以达到较高的检测率，但是比较费时，不利于实时应用。

为了提高检测速度，本文将 MLP 的隐层神经元视为一个滤波器，利用 FFT 的性质，把时域卷积转换为频域乘积。具体做法如下所述。

2.2 利用 FFT 和 MLP 实现快速脸部检测

我们对人工神经网络进行训练的数据是脸部/非脸部库中图像的灰度信息。设滑动窗口的大小为 19×19 像素，相应的，脸部/非脸部库中图像的大小也为 19×19 像素。所以，确定神经网络输入层神经元数为 381，隐含层神经元个数为 19。为了消除光照等外部条件对检测结果的影响，首先对训练样本进行标准化处理如下：(为滑动窗口位于输入图像处时子图像的灰度信息，以下 $n=25$)

$$\begin{aligned} \|X_i\| &= 1 \text{ and } \text{mean} X_i = 0 \\ [X]_{rc} &= \frac{[X]_x - \tilde{x}_{rc} 1_{nn}}{\|[X]_x - \tilde{x}_{rc} 1_{nn}\|} \end{aligned} \tag{4}$$

X_{rc} 为输入到神经网络的子图像的灰度平均值。

用人工神经网络实现对输入图像的快速脸部检测的原理如图 2 所示。让滑动窗口在整幅输入图像中逐个像素的移动，当滑动窗口位于图像某处，就将该窗口所包含的所有像素的灰度信息经过标准化处理后作为输入向量，送入神经网络进行检测。一般滑动窗口的大小取为 $n \times n$ 的方阵。

若将大小为 n^2 的输入向量写为 $n \times n$ 的矩阵 I ， $I_{(i,j)}$ 为图像点 (i, j) 处的灰度值；将隐层神经元 i 的权向量 w_H^i (大小为 n^2) 用 $n \times n$ 的矩阵 Φ_i 表示，则在二维空间中，隐含层神经元 i 的输出值为：

$$h(i) = g\left[\sum_{j=1}^n \sum_{k=1}^n \Phi_i(j, k) I(j, k) + b_H(i)\right] \tag{5}$$

公式 (5) 描述了隐含层神经元 i 对某个具体子窗口的作用，我们把它推广到整个输入图像。当滑动窗口的中心位于输入图像 (r, c) 处，隐含层神经元 i 对应于该子窗口的输出值为：

$$\begin{aligned} h_i(r, c) &= g\left(\sum_{j=-n/2}^{n/2} \sum_{k=-n/2}^{-n/2} \Phi_i(j, k) \cdot \right. \\ &\quad \left. I(r+j, c+k) + b_H(i)\right) \end{aligned} \tag{6}$$

式 (6) 是一个相关运算，经数学化简后得到：

$$H_i = g(\Phi \otimes I + B_H^i) \tag{7}$$

一个相关运算，可用两次傅立叶变换和一次傅立叶反变换来实现，对于相关运算有以下性质：

$$\Phi \otimes I = F^{-1}(F(\Phi) \cdot F^*(\Phi_i)) \tag{8}$$

因此，矩阵描述了隐含层神经元 i 对于整幅输入图像的作用，表示当滑动窗口中心位于输入图像处时，神经元 i 的输出值。由此，得到神经网络最后的输出为：

$$O=f(\sum_{i=1}^q w_o(i) H_i+B_o) \tag{9}$$

输出矩阵描述了神经网络对整幅输入图像的作用,表示当滑动窗口中心位于输入图像处时,神经网络的最终检测值。由于隐层神经元的傅立叶变换可以离线计算,减少了实时运算量。而用神经网络进行脸部识别的传统方法是滑动窗口每移动一个像素,就要进行一次公式(5)的运算,运算量很大,检测效率不高。两种方法的检测结果是相同。

当要改变输入图像的分辨率进行再检测时,无需再如上所述进行傅立叶变化。因为对于傅立叶变换有如下性质(设表示原始图像的灰度值函数,子抽样后得到) $g(x,y)$:

$$\left. \begin{aligned} g(x,y) &= f(2x,2y) \\ F(u,v) &= FT(f(x,y)) \\ FT(g(x,y)) &= G(u,v) = \frac{1}{4}F(\frac{u}{2},\frac{v}{2}) \end{aligned} \right\} \tag{10}$$

由于训练神经网络时,采用的是经过标准化处理的样本图像,在实际检测的时候,理论上也要求对所有检测窗口进行相应的处理,经过数学运算得到如下结果:

$$[X]_n \otimes \Phi = \Phi [X]_n \tag{11}$$

即对输入向量的标准化等价于对神经网络隐含层权值的标准化。所以,实际上,在学习得到神经网络的权值后,即对其进行标准化处理;那么,以后在对实际图像进行检测时就无需对输入到神经网络的子图像进行标准化处理。由于这个标准化处理也是离线计算的,也大大节省了实时计算的时间。

由公式(7)、(8)看到,我们将隐层神经元的输出表示为输入图像和一系列滤波器 $\Phi_i(i=1,2,\cdots,m)$ 的相关的运算,并利用FFT的性质,把时域卷积转换到频域计算。对一幅 $N \times N$ 大小的输入图像 I 进行二维FFT需 $O(N^2 \log_2 N^2)$ 次运算;而滤波器 Φ_i 的二维FT可以离线计算(因为神经网络一旦训练好后,在实时应用时权值保持不变)。所以,实时检测时,共需进行 $(m+1)$ 次FT,即 $O((m+1)N^2 \log_2 N^2)$ 次运算。而用人工神经网络实现脸部检测的传统方法需要的进行 $O((m+1)N^2 \log_2 N^2 m)$ 运算(设神经网络隐层神经元数为 m ,滑动窗口大小为 $n \times n$)。那么,总体速度提高了

$$O\left(\frac{(N-n+1)^2 n^2}{(m+1)N^2 \log_2 N^2}\right)。$$

实质上,我们将神经网络视作一个滤波器,利用傅立叶变换的性质,通过一次运算完成了对整幅输入图像的检测。该方法和传统的在输入图像中逐个像素移动滑动窗口的方法其实质是一样的,但却

大大提高了检测率。

3 试验结果和分析

为了检测输入图像中所有的脸部子图像,对输入图像进行多分辨率检测。如图3中图(a)和图(c)为经过多分辨率检测的结果,而最终的检测结果是对不同分辨率下得到的结果的一个分析。由于非脸部图像在不同多分辨率检测下出现误检测(即把非脸部图像认为是脸部图像)的次数不会太高,而且一般来说会大大低于脸部图像在多分辨率情况下被正确检测到的总次数,所以用检测到的次数做限制,即可将脸部图像找出来。图3(b)和图(d)为检测的结果。

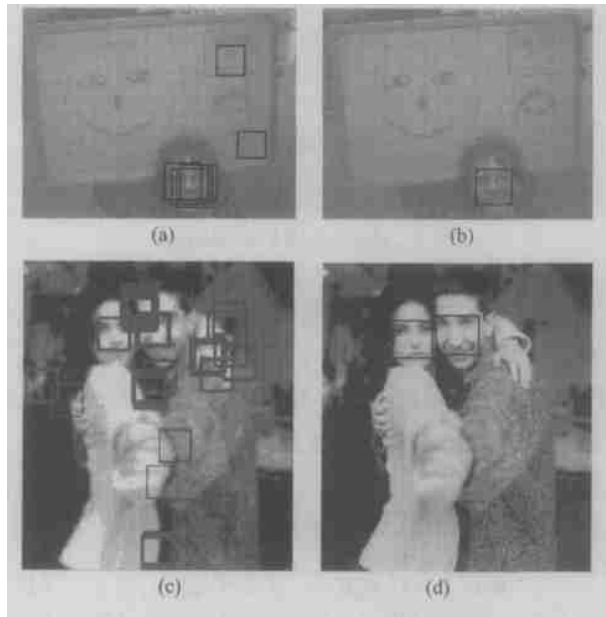


图3 多分辨率脸部检测实例
Fig 3 Multiscale face detection

为了对该脸部检测算法进行检验,我们对不同的脸部库进行了检测,试验结果证明这种方法能够在保持一定的检测率的情况下,能获得较高的检测速度。对CMU的脸部库图像的检测也证实了这一点(图3给出了对CMU中部分图像的检测结果)。如果我们能够增加脸部库训练样本数,并利用bootstrap扩充非脸部库训练样本,则能得到更高的检测率。试验也证明了这一点。

参考文献:

[1] CHELLAPPA R, WILSON C L, SIROHEY S. Human and machine recognition of faces: a survey[J]. Proc IEEE, 1995, 83 (5): 705—740.
[2] YANG M H, KRIEGMAN D J, AHUJA N. Detecting faces in images: a survey[J]. IEEE Transactions on Pattern Analysis

and Machine Intelligence, 2002, 24(9) .

[3] YOW K C, CIPOLLA R. Feature-based human face detection [J] . Image and Vision Computing, 1997, 15(9) : 713—735.

[4] LANITIS A, TAYOR C J, COOTES T F. An automatic face i dentification system using flexible appearance models[J] . Image and Vision Computing, 1995, 13(5): 393—401.

[5] ROWLEY H, BALUJA S, KANADE T. Neural network-based face detection[J] . IEEE Trans, Pattern Analysis and Machine Intelligence, 1998, 20(1) : 23—28.

[6] SUNG K K, POGGIO T. Exampled-based learning for view-based human face detection[J] . Technical Report AI Memo 1521, Massachusetts Inst of Technology AI Lab, 1994.

Face Detecting Using MLP and FFT

LUO Xiao yan, HOU Xiao wei, HU Hai feng, XIONG Yin gen

(Department of Electronics and Communication Engineering, Sun Yat senUniversity, Guangzhou 510275, China)

Abstract: New technique that speeds up the time needed by a trained network (MLP) to detect a face in a large input image is presented. The hidden units of the MLP are taken in terms of filter convolution enabling the using of fast Fourier transform for an efficient computation of neural activities, and experiments show that it's effective.

Key words: FFT; MLP; training database; face detection