

## 基于跨连卷积神经网络的性别分类模型

张婷<sup>1</sup> 李玉鑑<sup>1</sup> 胡海鹤<sup>1</sup> 张亚红<sup>1</sup>

**摘要** 为提高性别分类准确率, 在传统卷积神经网络 (Convolutional neural network, CNN) 的基础上, 提出一个跨连卷积神经网络 (Cross-connected CNN, CCNN) 模型. 该模型是一个 9 层的网络结构, 包含输入层、6 个由卷积层和池化层交错构成的隐含层、全连接层和输出层, 其中允许第 2 个池化层跨过两个层直接与全连接层相连接. 在 10 个人脸数据集上的性别分类实验结果表明, 跨连卷积网络的准确率均不低于传统卷积网络.

**关键词** 性别分类, 卷积神经网络, 跨连卷积神经网络, 跨层连接

**引用格式** 张婷, 李玉鑑, 胡海鹤, 张亚红. 基于跨连卷积神经网络的性别分类模型. 自动化学报, 2016, 42(6): 858–865

**DOI** 10.16383/j.aas.2016.c150658

### A Gender Classification Model Based on Cross-connected Convolutional Neural Networks

ZHANG Ting<sup>1</sup> LI Yu-Jian<sup>1</sup> HU Hai-He<sup>1</sup> ZHANG Ya-Hong<sup>1</sup>

**Abstract** To improve gender classification accuracy, we propose a cross-connected convolutional neural network (CCNN) based on traditional convolutional neural networks (CNN). The proposed model is a 9-layer structure composed of an input layer, six hidden layers (i.e., three convolutional layers alternating with three pooling layers), a fully-connected layer and an output layer, where the second pooling layer is allowed to directly connect to the fully-connected layer across two layers. Experimental results in ten face datasets show that our model can achieve gender classification accuracies not lower than those of the convolutional neural networks.

**Key words** Gender classification, convolutional neural network (CNN), cross-connected convolutional neural network (CCNN), cross-layer connection

**Citation** Zhang Ting, Li Yu-Jian, Hu Hai-He, Zhang Ya-Hong. A gender classification model based on cross-connected convolutional neural networks. *Acta Automatica Sinica*, 2016, 42(6): 858–865

人脸性别分类是一个二分类问题, 通过人脸图像信息自动发现和分析人脸性别属性. 它被广泛应用于以客户为中心的广告、视觉监控、智能用户界面和人口统计等领域.

作为一种传统方法, 全连接网络在多年前已被应用于人脸性别分类. Golomb 等<sup>[1]</sup> 通过训练一个两层的全连接网络 SEX-NET (Sex network), 对大小为  $30 \times 30$  的人脸图像进行性别分类. 这是第一次将神经网络应用于性别分类. Brunelli 等<sup>[2]</sup> 使用 HyperBF 网络进行性别分类. 首先使用两个径向基函数 (Radial basis function, RBF) 网络提取人脸

特征, 分别用于提取男性和女性人脸特征, 然后采用提取的特征进行性别分类. Tamura 等<sup>[3]</sup> 探索了神经网络在不同分辨率图像上的性别分类, 通过采用一个三层的反向传播网络对不同分辨率 (分辨率的范围从  $32 \times 32$  到  $8 \times 8$ ) 的人脸图像进行性别分类. 该网络在 30 幅大小为  $8 \times 8$  的低分辨率测试人脸图像上获得了 93% 的准确率. Jiao 等<sup>[4]</sup> 提出使用极限学习机进行人脸性别分类, 取得了较好的分类准确率. 以上使用全连接神经网络进行人脸性别分类的方法没有考虑到图像的二维信息, 忽略了图像中像素间的二维相关性, 因此, 分类准确率也不够高. 2014 年, Verma 等<sup>[5]</sup> 使用一个 6 层的卷积神经网络 (Convolutional neural network, CNN) 进行人脸性别分类, 获得了更高的准确率. 自深度学习出现以来<sup>[6]</sup>, 相比传统的手工特征提取方法<sup>[7]</sup>, CNN 表现出优越的特征提取能力<sup>[8]</sup>, 因而在字符识别<sup>[9]</sup>、人脸识别<sup>[10]</sup>、人脸跟踪<sup>[11]</sup>、交通标志识别<sup>[12]</sup>、手写体数字分类<sup>[13]</sup>、目标检测<sup>[14–16]</sup>、目标识别<sup>[17]</sup> 等领域都获得了广泛应用.

生物学研究表明, 在一幅给定的自然图像中, 人类视觉系统能够从整幅图像中找到最感兴趣的区域

收稿日期 2015-10-16 录用日期 2016-04-01  
Manuscript received October 16, 2015; accepted April 1, 2016  
国家自然科学基金 (61175004), 高等学校博士学科点专项科研基金 (20121103110029), 北京市博士后工作资助项目 (2015ZZ-24: Q6007011201501) 资助  
Supported by National Natural Science Foundation of China (61175004), Specialized Research Fund for the Doctoral Program of Higher Education of China (20121103110029), and Project Funding of Postdoctor in Beijing (2015ZZ-24: Q6007011201501)  
本文责任编辑 柯登峰  
Recommended by Associate Editor KE Deng-Feng  
1. 北京工业大学计算机学院 北京 100124  
1. Computer School, Beijing University of Technology, Beijing 100124

和物体, 并进而利用细节特征进行目标识别<sup>[18]</sup>. 这说明, 综合利用高层次特征和低层次特征, 对改善视觉系统的识别性能是有利的<sup>[19]</sup>. 以此为出发点, 本文在传统卷积神经网络的基础上, 通过引入跨层连接的思想, 提出一个具有 9 层结构的跨连卷积神经网络模型, 目的是为了有效地融合低层次特征和高层次特征, 以构造性能更好的分类器.

将跨连卷积神经网络应用于人脸性别分类的优点在于, 在对人脸图像进行分类时, 既考虑了高层次特征, 又考虑了低层次特征, 从而可能获得比传统卷积神经网络更高的准确率. 在本文的跨连卷积神经网络中, 虽然跨层连接的思想借鉴了 DeepID 网络<sup>[20]</sup>, 但二者并不完全相同. 在 DeepID 网络中, 跨层是将最后一个池化层与全连接层进行连接, 跨层连接的权值是可以学习和调整的; 而在本文的网络中, 跨层是将某一隐含层直接拷贝连接到全连接层的部分输入, 跨层连接的权值是固定的, 无需调整和学习.

## 1 跨连卷积神经网络

### 1.1 卷积神经网络

卷积神经网络最初是受视觉神经机制的启发针对二维形状的识别而设计的一种多层感知器, 在平移情况下具有高度不变性. 1962 年, Hubel 等通过对猫视觉皮层细胞的研究, 提出了感受野 (Receptive field) 的概念<sup>[21]</sup>, 1983 年, 日本学者 Fukushima 等基于感受野概念提出的神经认知机 (Neocognitron) 模型<sup>[22]</sup>, 被认为是第一个实现的卷积神经网络.

实际上, 传统卷积神经网络是一种特殊的深层前向神经网络模型, 其结构一般由输入层、多个交替的卷积层和池化层、全连接层及输出层组成<sup>[23]</sup>. 输入层通常是一个矩阵, 比如一幅图像. 从前馈网络的角度来看, 卷积层和池化层可以看作是特殊的隐含层, 而除输出层外的全连接层都是普通的隐含层.

在卷积神经网络中, 有四种基本的运算, 依次定义为: 内卷积、外卷积、下采样和上采样.

假设  $A$  和  $B$  为矩阵, 大小分别为  $M \times N$  和  $m \times n$ , 且  $M \geq m$ ,  $N \geq n$ , 则它们的内卷积  $C = A \tilde{*} B$  的所有元素定义为:

$$c_{ij} = \sum_{s=1}^m \sum_{t=1}^n a_{i+m-s, j+n-t} \cdot b_{st} \quad (1)$$

式中,  $1 \leq i \leq M - m + 1, 1 \leq j \leq N - n + 1$ .

它们的外卷积定义为:

$$A \widehat{*} B = \widehat{A}_B \tilde{*} B \quad (2)$$

式中,  $\widehat{A}_B = (\widehat{a}_{ij})$  是一个通过用 0 对  $A$  进行扩充得到的矩阵, 大小为  $(M + 2m - 2) \times (N + 2n - 2)$ ,

且

$$\widehat{a}_{ij} = \begin{cases} a_{i-m+1, j-n+1}, & \text{若 } m \leq i \leq M + m - 1, \\ & n \leq j \leq N + n - 1 \\ 0, & \text{其他} \end{cases} \quad (3)$$

如果对矩阵  $A$  进行不重叠的分块, 设每块的大小恰为  $\lambda \times \tau$ , 第  $ij$  块用矩阵  $G_{\lambda, \tau}^A(i, j)$  来表示, 则其构造如下:

$$G_{\lambda, \tau}^A(i, j) = (a_{st})_{\lambda \times \tau} \quad (4)$$

式中,  $(i-1) \times \lambda + 1 \leq s \leq i \times \lambda, (j-1) \times \tau + 1 \leq t \leq j \times \tau$ .  $G_{\lambda, \tau}^A(i, j)$  的下采样定义为:

$$\text{down}(G_{\lambda, \tau}^A(i, j)) = \frac{1}{\lambda \times \tau} \sum_{s=(i-1) \times \lambda + 1}^{i \times \lambda} \sum_{t=(j-1) \times \tau + 1}^{j \times \tau} a_{st} \quad (5)$$

用大小为  $\lambda \times \tau$  的不重叠块对矩阵  $A$  的下采样定义为:

$$\text{down}_{\lambda, \tau}(A) = \text{down}(G_{\lambda, \tau}^A(i, j)) \quad (6)$$

对矩阵  $A$  进行倍数为  $\lambda \times \tau$  的不重叠上采样定义为:

$$\text{up}_{\lambda \times \tau}(A) = A \otimes 1_{\lambda \times \tau} \quad (7)$$

式中  $1_{\lambda \times \tau}$  是一个元素全为 1 的矩阵,  $\otimes$  代表克罗内克积.

### 1.2 跨连卷积神经网络

传统卷积神经网络的一个缺点是难以有效地利用低层次特征与高层次特征进行融合去构造性能更好的分类器. 针对该问题, 本文通过引入跨层连接的思想, 提出了一个跨连卷积神经网络 (Cross-connected convolutional neural network, CCNN), 用于建立人脸性别分类模型. 该模型包含 1 个输入层  $x$ 、3 个卷积层 ( $h_1, h_3, h_5$ )、3 个池化层 ( $h_2, h_4, h_6$ )、1 个全连接层  $h_7$  和 1 个输出层  $o$ , 如图 1 所示. 注意, 该网络首先以一幅图像为输入, 然后使用 3 个交错的卷积层和池化层提取图像的特征, 最后将 2 个池化层 ( $h_4$  和  $h_6$ ) 提取的特征直接传递给全连接层进行融合处理并进行分类. 输出层有 2 个节点, 代表输入图像所属的类别. 不难看出, 在该网络中, 从池化层  $h_4$  到全连接层  $h_7$  之间的连接跨过了 2 层, 称为跨层连接. 此时, 全连接层的节点数是第 4 个隐含层和第 6 个隐含层节点数之和, 即  $n_{h_7} = n_{h_4} + n_{h_6}$ , 其中  $n_{h_4}$ 、 $n_{h_6}$  和  $n_{h_7}$  分别表示第 4、6、7 个隐含层的节点数. 表 1 给出了 CCNN 模型具体描述, 包含每层的类型 (Type)、卷积核大小

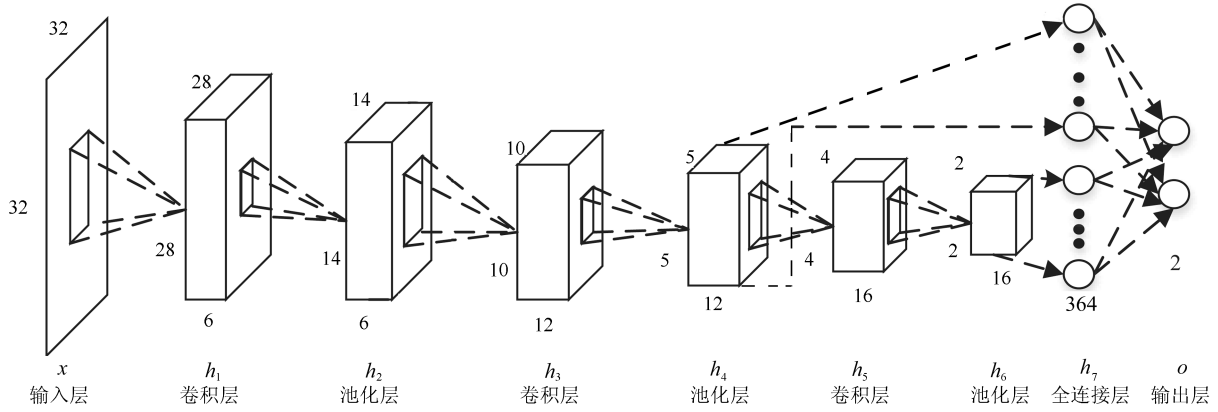


图 1 跨连卷积神经网络结构示意图

Fig. 1 The cross-connected convolutional neural network

表 1 CCNN 的网络描述

Table 1 Description of the CCNN

Layer	Type	Patch size	Stride	Output size
$x$	Input			$32 \times 32$
$h_1$	Convolution	$5 \times 5$	1	$28 \times 28 \times 6$
$h_2$	Mean pooling	$2 \times 2$	2	$14 \times 14 \times 6$
$h_3$	Convolution	$5 \times 5$	1	$10 \times 10 \times 12$
$h_4$	Mean pooling	$2 \times 2$	2	$5 \times 5 \times 12$
$h_5$	Convolution	$2 \times 2$	1	$4 \times 4 \times 16$
$h_6$	Mean pooling	$2 \times 2$	2	$2 \times 2 \times 16$
$h_7$	Fully-connected			364
$o$	Output			2

(Patch size)、步长 (Stride) 和每层输出的大小 (Output size)。

### 1.3 跨连卷积神经网络的学习过程

对于第  $l$  个样本  $x^l$ , CCNN 的计算过程为:

$$\begin{cases}
 h_{1,j}^l = \sigma(u_{1,j}^l) = \sigma(x^l * W^{1,j} + b^{1,j}), 1 \leq j \leq 6 \\
 h_{2,j}^l = g(u_{2,j}^l) = g(\text{down}_{\lambda,\tau}(h_{1,j}^l)), 1 \leq j \leq 6 \\
 h_{3,j}^l = \sigma(u_{3,j}^l) = \sigma\left(\sum_{i=1}^6 h_{2,i}^l * W^{3,j} + b^{3,j}\right), \\
 \quad 1 \leq j \leq 12 \\
 h_{4,j}^l = g(u_{4,j}^l) = g(\text{down}_{\lambda,\tau}(h_{3,j}^l)), 1 \leq j \leq 12 \\
 h_{5,j}^l = \sigma(u_{5,j}^l) = \sigma\left(\sum_{i=1}^{12} h_{4,i}^l * W^{5,j} + b^{5,j}\right), \\
 \quad 1 \leq j \leq 16 \\
 h_{6,j}^l = g(u_{6,j}^l) = g(\text{down}_{\lambda,\tau}(h_{5,j}^l)), 1 \leq j \leq 16 \\
 \mathbf{H}_7^l = (h_{4,(1,2,\dots,12)}^l \quad h_{6,(1,2,\dots,16)}^l) \\
 \mathbf{o}_8^l = \sigma(\mathbf{u}_8^l) = \sigma(W^8 \mathbf{H}_7^l + \mathbf{b}^8)
 \end{cases}
 \quad (8)$$

式中, 函数  $\sigma(x) = 1/(1 + e^{-x})$ ,  $g(x) = x$ ,  $\text{down}_{\lambda,\tau}(\cdot)$  表示下采样函数, 滑动窗口的大小为  $\lambda \times \tau$ , 同式 (6).  $W^{1,j}$  和  $b^{1,j}$  分别表示第 1 个隐含层的第  $j$  个卷积核和偏置,  $h_{1,j}^l$  表示第 1 个隐含层的第  $j$  个卷积面 ( $1 \leq j \leq 6$ );  $h_{2,j}^l$  表示第 2 个隐含层的第  $j$  个池化面 ( $1 \leq j \leq 6$ );  $W^{3,j}$  和  $b^{3,j}$  分别表示第 3 个隐含层的第  $j$  个卷积核和偏置,  $h_{3,j}^l$  表示第 3 个隐含层的第  $j$  个卷积面 ( $1 \leq j \leq 12$ );  $h_{4,j}^l$  表示第 4 个隐含层的第  $j$  个池化面 ( $1 \leq j \leq 12$ );  $W^{5,j}$  和  $b^{5,j}$  分别表示第 5 个隐含层的第  $j$  个卷积核和偏置,  $h_{5,j}^l$  表示第 5 个隐含层的第  $j$  个卷积面 ( $1 \leq j \leq 16$ );  $h_{6,j}^l$  表示第 6 个隐含层的第  $j$  个池化面 ( $1 \leq j \leq 16$ );  $\mathbf{H}_7^l$  表示第 7 个隐含层的输出值;  $W^8$  和  $\mathbf{b}^8$  分别表示输出层的权值和偏置,  $\mathbf{o}_8^l$  表示输出层的输出值。

参照 CNN 的学习过程, 本文提出了训练 CCNN 的反向传播算法 (Cross-connected convolutional neural network-Back propagation, CCNN-BP) (详见算法 1), 用于学习网络参数. 假设共有  $N$  个训练样本  $(x^l, y^l)$  ( $1 \leq l \leq N$ ),

网络的输入是  $x^l = (x_1^l, x_2^l, \dots, x_m^l)^T$ , 期望输出是  $y^l = (y_1^l, y_2^l, \dots, y_c^l)^T$ , 实际输出是  $o^l = (o_1^l, o_2^l, \dots, o_c^l)^T$ , 优化的目标函数是全部样本的均方误差:

$$L_N = \frac{1}{2} \sum_{l=1}^N \sum_{n=1}^c (o_n^l - y_n^l)^2 \quad (9)$$

CCNN-BP 首先需要计算出网络中每一层的反馈传递误差  $\delta_k$  ( $1 \leq k \leq 8$ ). 这里需要特别注意的是第 4 个隐含层的反馈传递误差  $\delta_{4,j}$  ( $1 \leq j \leq 12$ ), 因为该层同时与第 5 个隐含层  $h_5$  和第 7 个隐含层  $h_7$  相连, 因此该层的反馈传递误差为两部分之和. 其余各层的反馈传递误差的计算方式与 CNN 的计算方式相同, 因此, CCNN 中每一层的反馈传递误差矩阵为:

$$\begin{cases} \delta_8^l = (o^l - y^l) \circ \sigma'(u_8^l) \\ \delta_7^l = [(W^8)^T \delta_8^l] \circ H_7^l \circ (1 - H_7^l) \\ \delta_7^l = (\delta_{41,1}^l, \delta_{41,2}^l, \dots, \delta_{41,12}^l, \delta_{6,1}^l, \delta_{6,2}^l, \dots, \delta_{6,16}^l) \\ \delta_{5,j}^l = \frac{1}{\lambda \times \tau} (\sigma'(u_{5,j}^l) \circ up_{\lambda \times \tau}(\delta_{6,j}^l)), 1 \leq j \leq 16 \\ \delta_{42,j}^l = [\delta_{5,j}^l \tilde{*} W^{5,j}] \circ g'(u_{4,j}^l), 1 \leq j \leq 12 \\ \delta_{4,j} = (\delta_{41,j} + \delta_{42,j}), 1 \leq j \leq 12 \\ \delta_{3,j}^l = \frac{1}{\lambda \times \tau} (\sigma'(u_{3,j}^l) \circ up_{\lambda \times \tau}(\delta_{4,j}^l)), 1 \leq j \leq 12 \\ \delta_{2,j}^l = [\delta_{3,j}^l \tilde{*} W^{3,j}] \circ g'(u_{2,j}^l), 1 \leq j \leq 6 \\ \delta_{1,j}^l = \frac{1}{\lambda \times \tau} (\sigma'(u_{1,j}^l) \circ up_{\lambda \times \tau}(\delta_{2,j}^l)), 1 \leq j \leq 6 \end{cases} \quad (10)$$

式中, 符号  $\circ$  表示矩阵 (或者向量) 对应元素相乘,  $up_{\lambda \times \tau}(\cdot)$  表示上采样函数, 同式 (7).  $\delta_8^l$  和  $\delta_7^l$  分别表示输出层和第 7 个隐含层的反馈传递误差,  $\delta_{6,j}^l$  ( $1 \leq j \leq 16$ ) 表示第 6 个隐含层第  $j$  个池化面的反馈传递误差,  $\delta_{5,j}^l$  ( $1 \leq j \leq 16$ ) 表示第 5 个隐含层第  $j$  个卷积面的反馈传递误差,  $\delta_{41,j}^l$  ( $1 \leq j \leq 12$ ) 表示第 7 个隐含层传给第 4 个隐含层第  $j$  个池化面的反馈传递误差,  $\delta_{42,j}^l$  ( $1 \leq j \leq 12$ ) 表示第 5 个隐含层传给第 4 个隐含层第  $j$  个池化面的反馈传递误差,  $\delta_{4,j}^l$  ( $1 \leq j \leq 12$ ) 表示第 4 个隐含层第  $j$  个池化面总的反馈传递误差,  $\delta_{3,j}^l$  ( $1 \leq j \leq 12$ ) 表示第 3 个隐含层第  $j$  个卷积面的反馈传递误差,  $\delta_{2,j}^l$  ( $1 \leq j \leq 6$ ) 表示第 2 个隐含层第  $j$  个池化面的反馈传递误差,  $\delta_{1,j}^l$  ( $1 \leq j \leq 6$ ) 表示第 1 个隐含层第  $j$  个卷积面的反馈传递误差.

在获得每一层的反馈传递误差之后, CCNN-BP

就可以计算关于权值和偏置的偏导数, 如下:

$$\begin{cases} \frac{\partial L_N}{\partial W^8} = \sum_{l=1}^N \delta_8^l (H_7^l)^T, \frac{\partial L_N}{\partial b^8} = \sum_{l=1}^N \delta_8^l \\ \frac{\partial L_N}{\partial W^{5,j}} = \sum_{l=1}^N h_{4,j}^l \tilde{*} \delta_{5,j}^l, \frac{\partial L_N}{\partial b^{5,j}} = \sum_{l=1}^N \delta_{5,j}^l, 1 \leq j \leq 16 \\ \frac{\partial L_N}{\partial W^{3,j}} = \sum_{l=1}^N h_{2,j}^l \tilde{*} \delta_{3,j}^l, \frac{\partial L_N}{\partial b^{3,j}} = \sum_{l=1}^N \delta_{3,j}^l, 1 \leq j \leq 12 \\ \frac{\partial L_N}{\partial W^{1,j}} = \sum_{l=1}^N x^l \tilde{*} \delta_{1,j}^l, \frac{\partial L_N}{\partial b^{1,j}} = \sum_{l=1}^N \delta_{1,j}^l, 1 \leq j \leq 6 \end{cases} \quad (11)$$

式中,  $\frac{\partial L_N}{\partial W^8}$  和  $\frac{\partial L_N}{\partial b^8}$  分别表示目标函数对权值  $W^8$  和偏置  $b^8$  的偏导数;  $\frac{\partial L_N}{\partial W^{5,j}}$  和  $\frac{\partial L_N}{\partial b^{5,j}}$  分别表示目标函数对权值  $W^{5,j}$  和偏置  $b^{5,j}$  ( $1 \leq j \leq 16$ ) 的偏导数;  $\frac{\partial L_N}{\partial W^{3,j}}$  和  $\frac{\partial L_N}{\partial b^{3,j}}$  分别表示目标函数对权值  $W^{3,j}$  和偏置  $b^{3,j}$  ( $1 \leq j \leq 12$ ) 的偏导数;  $\frac{\partial L_N}{\partial W^{1,j}}$  和  $\frac{\partial L_N}{\partial b^{1,j}}$  分别表示目标函数对权值  $W^{1,j}$  和偏置  $b^{1,j}$  ( $1 \leq j \leq 6$ ) 的偏导数.

最后, CCNN-BP 利用梯度下降法更新网络的权值和偏置.

#### 算法 1. CCNN-BP

**Input:** 训练集  $S = \{(x^l, y^l), 1 \leq l \leq N\}$ , 网络结构, 最大迭代次数.

- 1: 随机初始化所有的权值和偏置;
- 2: **for** epoch = 1 to *maxepoch* **do**;
- 3: **for**  $l = 1$  to  $N$  **do**;
- 4: 计算每个样本的实际输出;
- 5: 计算每一层的反馈传递误差  $\delta_8^l, \delta_7^l, \delta_{5,j}^l$  ( $1 \leq j \leq 16$ ),  $\delta_{4,j}^l$  ( $1 \leq j \leq 12$ ),  $\delta_{3,j}^l$  ( $1 \leq j \leq 12$ ),  $\delta_{2,j}^l$  ( $1 \leq j \leq 6$ ),  $\delta_{1,j}^l$  ( $1 \leq j \leq 6$ );
- 6: 计算网络中权值和偏置的偏导数;
- 7: 使用梯度下降法更新所有权值和偏置;
- 8: **End for**
- 9: **End for**

**Output:** 网络的权值和偏置.

## 2 实验结果与分析

为验证 CCNN 的性能, 本文将其与 CNN 的性别分类性能进行了比较. 实验共采用 10 个人脸数据集, 分别为: UMIST (University of Manchester Institute of Science and Technology) 数据集<sup>[24]</sup>、ORL (Olivetti Research Laboratory) 数据集、Georgia Tech 数据集<sup>[25]</sup>、FERET (Face recognition technology) 数据集、Extended Yale B 数据集<sup>[26]</sup>、AR (Aleix Martinez and Robert Benavente) 数据集<sup>[27]</sup>、Faces94 数据集、LFW (Labeled faces in the wild) 数据集<sup>[28]</sup>、MORPH (Cranio facial

表 2 实验数据集的训练集和测试集信息描述

Table 2 Number of training samples and testing samples of the experiments

数据集	训练集			测试集		
	男	女	混合	男	女	混合
UMIST	209	57	266	95	19	114
ORL	320	30	350	40	10	50
Georgia Tech	450	75	525	195	30	225
FERET	658	532	1 190	105	105	210
Extended Yale B	1 280	384	1 664	576	192	768
AR	910	910	1 820	390	390	780
Faces94	2 000	400	2 400	660	20	680
LFW	8 000	1 900	9 900	2 000	800	2 800
MORPH	40 997	7 102	48 099	3 000	1 000	4 000
CelebFaces+	27 887	37 113	65 000	2 500	2 500	5 000

longitudinal morphological face) 数据集和 CelebFaces+ 数据集<sup>[29]</sup>. 所有实验均在 i7-3770 CPU 3.40 GHz 8.00 GB 内存的 Dell 台式机上使用 Matlab R2010a 运行完成.

CNN 的层数及参数设置与 CCNN 相同, 即, 3 个卷积层 ( $h_1$ ,  $h_3$ ,  $h_5$ ) 的卷积核的大小分别为  $5 \times 5$ 、 $5 \times 5$  和  $2 \times 2$ , 池化因子的大小为  $2 \times 2$ . CNN 和 CCNN 的学习率都设置为 0.1. 此外, CNN 和 CCNN 在 7 个数据集 (UMIST 数据集、ORL 数据集、Georgia Tech 数据集、FERET 数据集、Extended Yale B 数据集、AR 数据集、Faces94 数据集) 上的最大迭代次数为 30, 在 3 个数据集 (LFW 数据集、MORPH 数据集、CelebFaces+ 数据集) 上的最大迭代次数为 200.

## 2.1 实验数据集简介

在实验采用的 10 个数据集中, UMIST 数据集、ORL 数据集、FERET 数据集、Extended Yale B 数据集和 AR 数据集的图像为灰度图像, Faces94 数据集、Georgia Tech 数据集、LFW 数据集、MORPH 数据集、CelebFaces+ 数据集的图像为彩色图像. 为了便于处理, 本文将彩色图像均转变为灰度图像, 并统一把图像大小归一化为  $32 \times 32$ , 把像素值归一化到  $[0, 1]$ .

图 2 给出了各个数据集的示例图像. 表 2 给出了这些数据集的训练集及测试集的情况, 其中 7 个数据集 (UMIST、ORL、Georgia Tech、Extended Yale B、AR、Faces94、MORPH) 使用数据集中的全部图片, 3 个数据集 (FERET、LFW、CelebFaces+) 使用数据集中的部分图片. 此外, 本文在实验中用到的训练集和测试集都没有交集, 即同一个人的人脸图像不能同时出现在训练集和测试集中.



图 2 10 个数据集的示例人脸图像

Fig. 2 Examples of face images in ten datasets

## 2.2 CCNN 与 CNN 的人脸性别分类比较

本文采用 CCNN 与 CNN 模型在 10 个人脸数据集上的性别分类实验结果总结在表 3 中. 从中不难看出: CCNN 在 UMIST、Georgia Tech、FERET、Extended Yale B、Faces94、LFW、MORPH 和 CeleFaces+ 这 8 个数据集上的分类准确率均高于 CNN, 而在 ORL 和 AR 上与 CNN 相同. 与 CNN 相比, CCNN 在 UMIST 和 LFW 数据集上的分类准确率分别提高了 2.71% 和 0.86%. 这可能是因为 CCNN 可以直接使用两个池化层提取到的特征去分类, 而 CNN 直接使用一个池化层提取到的特征去分类, 导致 CCNN 在融合不同尺度的信息时能够更加准确地反映图像信息, 从而获得了更好的分类性能.

表 3 CNN 和 CCNN 在 10 个数据集上的分类准确率 (%)

Table 3 Classification accuracies of CNN and CCNN in ten datasets (%)

数据集	CNN	CCNN
UMIST	96.49	99.20
ORL	98.00	98.00
Georgia Tech	97.60	97.78
FERET	94.77	96.44
Extended Yale B	98.53	98.82
AR	98.71	98.71
Faces94	96.46	97.35
LFW	87.00	87.86
MORPH	92.73	94.56
CelebFaces+	85.18	88.70

## 2.3 CCNN 与 CNN 在男性和女性图像集中的分类比较

为比较 CCNN 和 CNN 分别在男性和女性图像集中的分类效果, 本文在 Georgia Tech、Extended Yale B、Faces94 和 LFW 这 4 个数据集上进行了实验, 结果如表 4 所示. 从表 4 可以看出, CCNN 对男性、女性和双性混合的分类准确率均不低于 CNN. CCNN 在 Faces94 数据集上对男性、女性和双性混合的分类准确率分别为 100%、94.70% 和 97.5%, 与 CNN 相比分别提高了 1.52%、0.26% 和 0.89%. 此外, CCNN 和 CNN 对男性的分类准确率高于女性. 在 Georgia Tech 数据集上, CNN 和 CCNN 对男性的分类准确率都为 99.49%, 而对女性的分类准确率分别为 95.71% 和 96.07%. 这可能是因为数据集中女性人脸图像明显少于男性人脸图像, 导致女性人脸图像没有得到充分的训练造成的.

表 4 CNN 和 CCNN 在 4 个数据集上的分类准确率 (%)

Table 4 Classification accuracies of CNN and CCNN in four datasets (%)

数据集	CNN			CCNN		
	男	女	混合	男	女	混合
Georgia Tech	99.49	95.71	97.60	99.49	96.07	97.78
Extended Yale B	100	97.06	98.53	100	97.64	98.82
Faces94	98.48	94.44	96.46	100	94.70	97.35
LFW	95.00	79.00	87.00	96.80	78.92	87.86

## 2.4 不同跨连方式对 CCNN 性能的影响

为探索不同的跨连方式对 CCNN 的性能产生的影响, 本文在 Georgia Tech、AR、Faces94、LFW 和 MORPH 五个数据集上进行了实验, 实验结果如表 5 所示. 其中,  $h_2 - h_7$  表示使用第 2 个隐含层与全连接层连接时的分类准确率,  $h_3 - h_7$  表示使用第 3 个隐含层与全连接层连接时的分类结果,  $h_4 - h_7$  表示使用第 4 个隐含层与全连接层连接时的分类结果,  $h_5 - h_7$  表示使用第 5 个隐含层与全连接层的连接时的分类结果.

表 5 CCNN 在不同跨连方式的分类准确率 (%)

Table 5 Classification accuracies of the CCNN with different cross-connections (%)

数据集	$h_2 - h_7$	$h_3 - h_7$	$h_4 - h_7$	$h_5 - h_7$
Georgia Tech	97.96	97.84	97.78	97.33
AR	98.85	98.85	98.71	98.59
Faces94	97.50	97.35	97.35	97.35
LFW	88.13	88.04	87.86	87.86
MORPH	94.63	94.63	94.56	94.45

从表 5 可以看出, 不同的跨连方式对 CCNN 的性能具有一定的影响. 整体来看, 使用低层的隐含层与全连接层连接时, CCNN 的分类准确率不低于其之后的隐含层与全连接层连接时的分类准确率. 例如, CCNN 在 Georgia Tech 数据集上使用 4 种不同的跨连方式获得的分类准确率分别为 97.96%、97.84%、97.78% 和 97.33%, 使用第 2 个隐含层 ( $h_2$ ) 与全连接层连接时的分类准确率 (97.96%) 不低于其之后的隐含层 ( $h_3, h_4, h_5$ ) 与全连接时的分类准确率 (分别为 97.84%, 97.78%, 97.33%).

## 3 结论

本文在传统卷积神经网络的基础上, 提出了一个用于性别分类的跨连卷积神经网络模型 CCNN. 该模型能够通过跨层连接融合两个隐含层的图像特征, 从而更加准确地反映图像信息, 并获得更好的性别

分类性能. 实验结果表明, 其分类准确率在 10 个人脸数据集上的分类准确率不低于 CNN. 此外, 不同的跨连方式对 CCNN 的性能具有一定的影响, 通常使用低层的隐含层与全连接层跨层连接时, CCNN 的分类准确率不低于其之后的隐含层与全连接层连接时的分类准确率.

## References

- Golomb B A, Lawrence D T, Sejnowski T J. SEXNET: a neural network identifies sex from human faces. In: Proceedings of Advances in Neural Information Processing Systems (NIPS). Colorado, USA: Morgan Kaufmann Publishers Inc., 1991. 572–579
- Brunelli R, Poggio T. HyberBF networks for gender classification. In: Proceedings of the 1992 DARPA Image Understanding Workshop. Detroit, USA, 1992. 311–314
- Tamura S, Kawai H, Mitsumoto H. Male/female identification from 8×6 very low resolution face images by neural network. *Pattern Recognition*, 1996, **29**(2): 331–335
- Jiao Y B, Yang J C, Fang Z J, Xie S J, Park D S. Comparing studies of learning methods for human face gender recognition. In: Proceedings of the 7th Chinese Conference on Biometric Recognition (CCBR). Guangzhou, China: Springer, 2012. 67–74
- Verma A, Vig L. Using convolutional neural networks to discover cognitively validated features for gender classification. In: Proceedings of the 2014 International Conference on Soft Computing and Machine Intelligence (SCMI). New Delhi, India: IEEE, 2014. 33–37
- Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks. *Science*, 2006, **313**(5786): 504–507
- Wei Wei-Bo, Hong Dan-Feng, Pan Zhen-Kuan, Wu Xin. Blurred palmprint recognition algorithm based on region feature map. *Acta Automatica Sinica*, 2015, **41**(2): 386–395 (魏伟波, 洪丹枫, 潘振宽, 吴鑫. 基于区域特征映射的模糊掌纹识别方法. *自动化学报*, 2015, **41**(2): 386–395)
- Tivive F H C, Bouzerdoum A. A gender recognition system using shunting inhibitory convolutional neural networks. In: Proceedings of the 2006 IEEE International Joint Conference on Neural Networks (IJCNN). Vancouver, BC: IEEE, 2006. 5336–5341
- Ciresan D C, Meier U, Gambardella L M, Schmidhuber J. Convolutional neural network committees for handwritten character classification. In: Proceedings of the 2011 International Conference on Document Analysis and Recognition. Beijing, China: IEEE, 2011. 1135–1139
- Khalajzadeh H, Mansouri M, Teshnehlab M. Face recognition using convolutional neural network and simple logistic classifier. In: Proceedings of the 17th Online World Conference on Soft Computing in Industrial Applications. Switzerland: Springer International Publishing, 2014. 197–207
- Fan J, Xu W, Wu Y, Gong Y H. Human tracking using convolutional neural networks. *IEEE Transactions on Neural Networks*, 2010, **21**(10): 1610–1623
- Jin J Q, Fu K, Zhang C S. Traffic sign recognition with hinge loss trained convolutional neural networks. *IEEE Transactions on Intelligent Transportation Systems*, 2014, **15**(5): 1991–2000
- Xu C Y, Lu C Y, Liang X D, Gao J B, Zheng W, Wang T J, Yan S C. Multi-loss regularized deep neural network. *IEEE Transactions on Circuits and Systems for Video Technology*, DOI: 10.1109/TCSVT.2015.2477937
- Krizhevsky A, Sutshever I, Hinton G E. ImageNet classification with deep convolutional neural networks. In: Proceedings of Advances in Neural Information Processing Systems (NIPS). Lake Tahoe, USA: Curran Associates, Inc., 2012. 4–13
- Szegedy C, Liu W, Jia Y Q, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. Going deeper with convolutions. arXiv: 1409.4842, 2014.
- He K M, Zhang X Y, Ren S Q, Sun J. Deep residual learning for image recognition. arXiv: 1512.03385, 2015.
- Sui Ting-Ting, Wang Xiao-Feng. Convolutional neural network with candidate location and multi-feature fusion. *Acta Automatica Sinica*, 2016, **42**(6): 875–882 (随婷婷, 王晓峰. 一种基于 CLMF 的深度卷积神经网络模型. *自动化学报*, 2016, **42**(6): 875–882)
- Zhong S H, Liu Y, Ren F F, Zhang J H, Ren T W. Video saliency detection via dynamic consistent spatio-temporal attention modelling. In: Proceedings of the 2013 AAAI Conference on Artificial Intelligence. Bellevue, USA: AAAI, 2013. 1063–1069
- Qi Mei-Bin, Tan Sheng-Shun, Wang Yun-Xia, Liu Hao, Jiang Jian-Guo. Multi-feature subspace and kernel learning for person re-identification. *Acta Automatica Sinica*, 2016, **42**(2): 299–308 (齐美彬, 檀胜顺, 王运侠, 刘皓, 蒋建国. 基于多特征子空间与核学习的行人再识别. *自动化学报*, 2016, **42**(2): 299–308)
- Sun Y, Wang X G, Tang X O. Deep learning face representation from predicting 10 000 classes. In: Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Columbus, OH: IEEE. 2014. 1891–1898
- Hubel D H, Wiesel T N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *Journal of Physiology*, 1962, **160**(1): 106–154
- Fukushima K, Miyake S, Ito T. Neocognitron: a neural network model for a mechanism of visual pattern recognition. *IEEE Transactions on Systems, Man, and Cybernetics*, 1983, **SMC-13**(5): 826–834
- LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 1998, **86**(11): 2278–2324

- 24 Graham D B, Allinson N M. Characterising virtual eigensignatures for general purpose face recognition. *Face Recognition: from Theory to Applications*. Berlin Heidelberg: Springer-Verlag, 1998. 446–456
- 25 Chen L, Man H, Nefian A V. Face recognition based on multi-class mapping of Fisher scores. *Pattern Recognition*, 2005, **38**(6): 799–811
- 26 Lee K C, Ho J, Kriegman D J. Acquiring linear subspaces for face recognition under variable lighting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, **27**(5): 684–698
- 27 Maetinez A M, Kak A C. PCA versus LDA. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2001, **23**(2): 228–233
- 28 Huang G B, Ramesh M, Berg T, Learned-Miller E. Labeled Faces in the Wild: a Database for Studying Face Recognition in Unconstrained Environment, Technical Report 07-49, University of Massachusetts, Amherst, USA, 2007.
- 29 Liu Z W, Luo P, Wang X G, Tang X O. Deep learning face attributes in the wild. In: Proceedings of the 2015 International Conference on Computer Vision (ICCV). Santiago, Chile: IEEE, 2015. 3730–3738



张婷 北京工业大学计算机学院博士研究生. 主要研究方向为模式识别, 深度学习, 大数据分析. 本文通信作者.  
E-mail: zhangting08@emails.bjut.edu.cn  
(ZHANG Ting Ph. D. candidate at the Computer School, Beijing University of Technology. Her research inter-

est covers pattern recognition, deep learning, and big data analysis. Corresponding author of this paper.)



李玉鑑 北京工业大学计算机学院教授. 主要研究方向为模式识别, 图像处理, 机器学习, 数据挖掘.

E-mail: liyujian@bjut.edu.cn

(LI Yu-Jian Professor at the Computer School, Beijing University of Technology. His research interest covers pattern recognition, image processing, machine learning, and data mining.)



胡海鹤 北京工业大学计算机学院博士后. 主要研究方向为模式识别, 机器学习, 红外技术.

E-mail: huhaihe@bjut.edu.cn

(HU Hai-He Postdoctor at the Computer School, Beijing University of Technology. Her research interest covers pattern recognition, machine learning, and infrared technology.)



张亚红 北京工业大学计算机学院博士研究生. 主要研究方向为模式识别, 数据挖掘, 大数据分析.

E-mail: plahpu@163.com

(ZHANG Ya-Hong Ph. D. candidate at the Computer School, Beijing University of Technology. Her research interest covers pattern recognition, data mining, and big data analysis.)