

DOI:10.19651/j.cnki.emt.1802302

# 基于卷积神经网络和语义特征的眼型分类<sup>\*</sup>

尚姝睿 卜凡亮

(中国人民公安大学 信息技术与网络安全学院 北京 100038)

**摘要:** 眼睛是人脸特征最重要的构成部分,研究眼睛的特征点定位和形状分类问题,对扩充智能模拟画像系统中的眼睛库起着至关重要的作用。提出一种基于级联卷积神经网络与语义特征的人眼分类方法,采用三级的级联卷积神经网络,检测并由粗略到细致地定位出106个特征点,根据定位眼睛的20个特征点对眼睛进行形状建模并定义3个确定眼型的形状参数,对这3个参数进行分段处理,通过每一区间对应不同的语义描述来达到眼型分类的目的。实验结果表明,此方法定位准确,能达到良好的眼睛分类效果。

**关键词:** 级联卷积神经网络;人脸特征点定位;语义特征;眼型分类

**中图分类号:** TP183    **文献标识码:** A    **国家标准学科分类代码:** 520.60

## Classification of eye shape based on convolutional neural network and semantic feature

Shang Yaorui Bu Fanliang

(College of Information Technology &amp; Network Security, People's Public Security University of China, Beijing 100038, China)

**Abstract:** The eye is the most important part of the facial features. Studying the feature point location and shape classification of the eye plays an important role in face recognition. This paper proposes a human eye classification method based on cascaded convolutional neural network and semantic features. Three levels of cascaded convolutional neural networks were used to detect and locate 106 feature points from rough to meticulous, among which 20 feature points can accurately locate the eyes. Based on these 20 feature points, the shape of the eye is modeled and three shape parameters defining the eye shape are defined. These three parameters are segmented to achieve the purpose of eye classification according to different semantic descriptions of each interval. Experimental results show that this method is accurate and can achieve good eye classification effect.

**Keywords:** cascaded convolutional neural network; the facial feature points location; semantic feature; classification of eye shape

## 0 引言

眼睛作为人脸重要的五官部件,具有独特性和遗传性,可为人脸识别和亲缘关系识别等研究提供关键信息,成为研究人脸的新的突破点。针对不同眼睛形状进行眼型分类,建立不同眼型的数据库,为后续利用计算机智能模拟人脸画像<sup>[1]</sup>时抽取不同类型的眼睛提供便利。同时,将在海量的人脸数据库中识别人脸缩小范围至在眼睛数据库内,有助于提高识别率和识别速度,降低识别算法的复杂度。

目前已有一些对眼型分类的研究。文献[2]使用眼型美学的分类方法,根据眼睛大小、眼睑、眼角的形态变化,将眼睛分为了18种类型,这种分类方法符合人类的日常习惯

但却不利于计算辨别;文献[3]利用眼睑曲线的不同形状,提出一种基于眼睑轮廓曲线相似度与眼型指数相结合的人眼分类方法,使用该方法对标准眼、圆眼、眯缝眼、细长眼等4种眼型,具有良好的分类效果,但不适用于更多类型。文献[4-5]建立眼睛的几何模型,对其进行语义化的描述和分类,该方法通过人脸分割、边缘检测等方法确定眼睛的几何特征点,并依此设定不同的形状特征参数,建立语义化的眼睛特征数据库,具有快速、高效、实用的特点。但是通过该方法确定的特征点数量少且不够精确,将直接影响后续的分类效果。

由此可以看出,面部特征点的检测与定位在人脸部件的分类中扮演着重要角色。检测人脸特征点的方法大致分

收稿日期:2018-11-20

<sup>\*</sup> 基金项目:中国铁道总公司重大课题(2017X001-A)项目资助

为3种,分别是基于主动形状模型(active shape model, ASM)<sup>[6-7]</sup>和主动外观模型(active appearance model, AAM)<sup>[8]</sup>的传统方法;基于级联形状回归的方法<sup>[9]</sup>;基于深度学习的方法<sup>[10-15]</sup>。目前,应用最广泛、精度最高、效果最好的是基于深度学习的方法。Sun等<sup>[10]</sup>首次将卷积神经网络(convolutional neural network, CNN)应用到人脸特征点检测,提出了一个基于回归方法的CNN框架,直接将特征点的定位看作一个回归任务,用全局回归器来计算特征点的坐标,不仅对局部图像损坏具有强大的适用性,而且借助于CNN强大的特征提取能力,获得更为精准的特征点。这种方法的不足之处是,当需要大量的特征点时,单一模型的方法可能会失败,另一方面,为每个点使用单独的系统会大大增加计算时间。为了应对这一挑战,Zhou等<sup>[11]</sup>精心设计了一种由粗到细的多级卷积网络级联,充分利用面部组件全局分布的几何约束以及组件内各点之间的相互作用,对于大量的面部特征点定位问题十分有效,同时提高了准确性和鲁棒性。Zhang等<sup>[14]</sup>提出一种多任务级联卷积神经网络用以同时处理人脸检测和人脸特征点定位问题,但该方法仍然只定位出5个特征点,不利于后续对部件的分类。

综合上述方法的优缺点,本次研究使用由粗到细的三级级联卷积神经网络定位人脸特征点,再结合语义化的方法对眼型分类,易于实现且能达到更精准的分类效果。

### 1 三级级联卷积神经网络检测人脸特征点

#### 1.1 多层次定位脸部特征点

该卷积网络用全脸作为输入,在第1级卷积网络中充分利用上下文信息,从深度构架的高层中提取了全局高级特征,即使局部的低级特征不明确或在具有挑战性的图像中被损坏,也能在第1级有效地预测关键点的位置。另一方面,由于该卷积网络被训练来同时预测所有关键点,关键点之间的几何约束被隐性地编码。因此,在第1级提取全局特征可以避免由于遮挡、姿态、照明等原因造成图像本身难以处理的缺陷和局部化方法中的数据损坏问题。下面的两级卷积网络结构较浅,用于对第1级的初始预测进行局部校正,以进一步提高定位的准确率,所以他们的输入被限制在初始位置周围的小局部区域。

本文用多模型来定位不同位置的点,如图1所示。用内部点来表示定位眼睛、眉毛、嘴巴和鼻子的73个特征点,用轮廓点来表示轮廓上的33个特征点。估计内部点和轮廓点的子系统与第1级网络分开。在第1级中,训练两个神经网络F1和F2,以获得分别包含所有内部点(F1)和轮廓点(F2)的最小包围盒(x-y坐标的最大值和最小值),以此来缩小接下来的搜索范围。对于内部点,另外训练两级卷积神经网络,将第1级给定的包围盒的范围裁剪出来后,由第2级初步估计73个内部特征点的位置;第3级进一步细化6个部件的特征点位置(即左右眉毛、眼睛、嘴巴

和鼻子),利用第2级网络预测到的73个点,把6个部件裁剪出来,对裁剪的部件分开进行训练、预测,因此本层需要有4个CNN模型,每个模型用于精确各部件特征点的位置。对于轮廓点,在第1级给定覆盖脸颊的包围盒后,第2级将裁剪后的图像作为输入并计算来自原始像素的33个轮廓点的准确坐标。

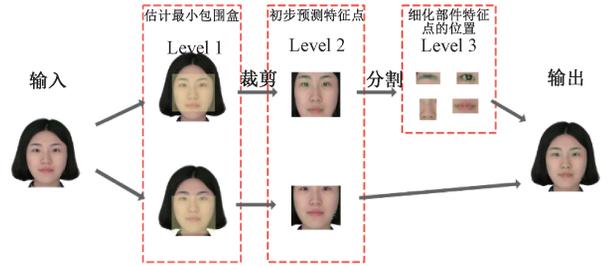


图1 多层次定位脸部特征点的系统框架

#### 1.2 典型的神经网络结构

具体每一级CNN的设计如图2所示。

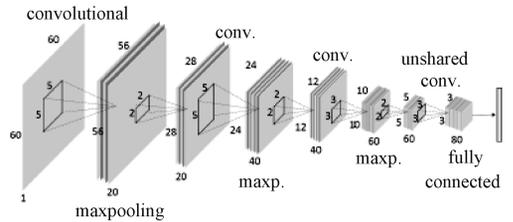


图2 系统中典型的网络结构

该系统中网络的典型结构与传统的CNN模型差不多,由卷积层、非共享卷积层和完全连接层组成。输入节点之后堆叠3个卷积层。每个卷积层将多个滤波器应用于多通道输入图像和输出响应。设第t卷积层的输入为I<sup>t</sup>,则输出为:

$$C_{i,j,k}^t = \left| \tanh \left( \sum_{x=0}^{h_t-1} \sum_{y=0}^{w_t-1} \sum_{z=0}^{c_t-1} I_{i-x,j-y,z}^{t-1} \cdot F_{x,y,k,z}^{t-1} + B_k \right) \right| \quad (1)$$

式中:I表示卷积层的输入;F和B是可调参数。

每个卷积层后跟着一个最大值池化层(maxpooling),其中的卷积公式为:

$$I_{i,j,k}^t = \max_{0 \leq x < d, 0 \leq y < d} (C_{i+d+x,j+d+y,k}^t) \quad (2)$$

3个卷积层后是不同于普通CNN模型的非共享卷积层(unshared conv),原因在于平时所遇到的CNN都是完全共享权重(global sharing weights),但是对于特征点的预测,五官的高层特征差别比较大,因此在卷积网络的高层采用非共享权重(unshared weight)的方式,即在不同位置采用不同权值,即不同位置所应用的过滤器并不相同,所以该层严格来说是一个局部的感知器,而非传统的卷积层。

$$C_{i,j,k} = \left| \tanh \left( \sum_{x=0}^{h-1} \sum_{y=0}^{w-1} \sum_{z=0}^{c-1} I_{i-x,j-y,z} \cdot F_{i,j,x,y,k,z} + B_{i,j,k} \right) \right| \quad (3)$$

最终的预测是由 1 个或 2 个完全连接层产生的。最后的输出公式如下。

$$\sum_{l^r} \left| \text{layer}_m \circ \text{layer}_{m-1} \circ \dots \circ \text{layer}_1(I^0) - \text{label}(I^0) \right|_2^2 \quad (4)$$

## 2 眼睛形状分类

当利用深度卷积神经网络准确定位全脸的特征点后，即知道了各个器官部件的特征点对应的坐标位置，然后按照固定比例扩展到一定的矩形区域，将部件抠取出来，同时分解出每个部件的正面形状。

眼睛作为人脸最重要的组成成分，具有较强的形状特征，不同人的眼睛有明显的区别，人们习惯将眼睛分为丹凤眼、圆眼、细长眼等类型，这种按照人类习惯的分类方法没有统一的分类标准，计算机根据这些称呼辨别不同眼型是非常复杂的，因此，本文采用语义化的方法进行形状分类，即将形状特征进行语义化，并建立相应的语义分类标准。

本文第 1 节已经用深度卷积神经网络来准确定位了 20 个眼睛特征点(左右眼分别 10 个)，接下来，根据这些特征点所确定的形状特征对眼睛进行形状建模，定义考虑到眼睛的长度、高度、角度等关键信息的分类数据标准。眼睛形状模型如图 3 所示。

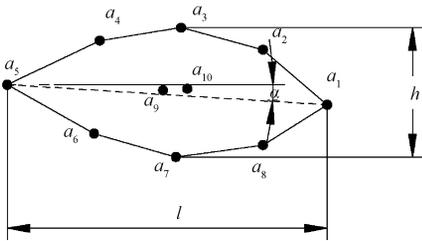


图 3 眼睛形状模型

由于双眼具有对称性，因此在建模时以左眼为例进行分析。图 3 中的 8 个特征点分别采用  $a_1 \sim a_{10}$  表示，根据这 10 个特征点，定义出如下 4 个眼睛的形状参数：

$$l = x_{a_1} - x_{a_5} \quad (5)$$

$$h = \max(y_{a_2}, y_{a_3}, y_{a_4}) - \min(y_{a_6}, y_{a_7}, y_{a_8}) \quad (6)$$

$$d = \frac{h}{l} \quad (7)$$

$$\alpha = \arctan[(y_{a_5} - y_{a_1}) / (x_{a_1} - x_{a_5})] \quad (8)$$

式中： $x_{a_i}$  代表第  $i$  个表示眼睛的特征点  $a_i$  的 X 坐标； $y_{a_i}$  代表第  $i$  个表示眼睛的特征点  $a_i$  的 Y 坐标； $l$  是  $a_1 \sim a_5$  之间的横向距离，表示眼睛的长度； $h$  是眼睛的最大纵向距离，表示眼睛的高度； $d$  是眼睛的高度和长度比值，表示眼睛的椭圆程度； $\alpha$  为眼睛的倾斜角度。

根据对样本的统计分析，本文对 3 个形状参数的值进行了分段处理，得到了 6 个分段阈值和 6 个参数区间，每个区间对应不同的语义描述，如表 1 所示。

表 1 眼睛特征参数语义化描述

形状参数	语义描述		
$l$	$l < s_1$	$s_1 \leq l < s_2$	$l \geq s_2$
	短眼睛	中眼睛	长眼睛
$d$	$d < s_3$	$s_3 \leq d < s_4$	$d \geq s_4$
	尖叶形	椭圆形	圆形
$\alpha$	$\alpha < s_5$	$s_5 \leq \alpha < s_6$	$\alpha \geq s_6$
	下倾斜(垂眼)	水平	上倾斜(吊眼)

## 3 实验分析

本文使用的图片像素大小为  $1704 \times 2272$ ，内容为人脸的正面照片，均无后期处理，共 366 张。在实验阶段，首先调用 Face++ 的 API 对特征点定位的准确性进行验证，再利用眼睛的特征点对眼睛形状的分类进行分析。

由于没有大量的数据供神经网络训练，而 Face++ 的云服务平台提供了人脸检测的接口，可以直接调用了其 API 得到人脸的 106 个特征点，部分左眼的特征点如图 4 所示，可以看到，级联卷积神经网络即使针对不同类型的眼睛，都能够准确地定位其特征点。



图 4 部分左眼的特征点

根据这些特征点可以得到 366 个眼睛所对应的形状模型，结合式(5)~(8)，通过 Python 计算出每个眼睛的形状参数  $l, d, \alpha$ ，图 4 中 20 个眼睛对应的参数值如表 2 所示。同时，人工判断这些眼睛属于什么类别也标注在表 2 里。

通过对比分析实验中 3 个参数的值和它本身应属的类别，得到左眼的最佳模型参数： $s_1 = 0.04, s_2 = 0.09, s_3 = 0.43, s_4 = 0.48, s_5 = 0.31, s_6 = 0.38$ 。

最后，将  $s_1 \sim s_6$  的值代入到表 1，得到对眼睛特征参数语义化描述分类的 9 个区间，如表 3 所示。

表2 图4中眼睛所对应的形状参数和语义描述

图片编号	$l$	$d$	$a$	$l$ 对应的 语义描述	$d$ 对应的 语义描述	$a$ 对应的 语义描述
4-1	0.101 103	0.442 308	0.384 058	长	椭圆	上斜
4-2	0.056 543	0.470 414	0.327 044	中	椭圆	下斜
4-3	0.093 254	0.426 380	0.323 741	中	尖	水平
4-4	0.022 384	0.422 713	0.305 970	短	尖	下斜
4-5	0.062 419	0.433 735	0.319 444	中	椭圆	水平
4-6	0.060 532	0.413 793	0.340 909	中	椭圆	水平
4-7	0.097 253	0.425 606	0.382 114	长	尖	上斜
4-8	0.078 578	0.424 749	0.354 331	中	椭圆	水平
4-9	0.037 576	0.430 421	0.390 977	中	圆	上斜
4-10	0.086 117	0.407 625	0.294 964	中	椭圆	下斜
4-11	0.021 579	0.482 639	0.359 712	短	圆	水平
4-12	0.093 742	0.428 571	0.418 803	长	尖	上斜
4-13	0.086 738	0.467 797	0.369 565	中	椭圆	水平
4-14	0.078 410	0.484 429	0.321 429	中	圆	水平
4-15	0.065 599	0.458 194	0.350 365	中	椭圆	水平
4-16	0.099 669	0.443 686	0.300 000	长	椭圆	下斜
4-17	0.083 141	0.415 225	0.316 667	中	尖	水平
4-18	0.049 141	0.423 611	0.344 262	中	椭圆	水平
4-19	0.024 385	0.434 629	0.341 463	中	椭圆	水平
4-20	0.068 270	0.417 857	0.393 162	短	尖	水平

表3 眼睛特征参数语义化描述分类

形状参数	语义描述		
$l$	$l < 0.04$	$0.04 \leq l < 0.09$	$l \geq 0.09$
	短眼睛	中眼睛	长眼睛
$d$	$d < 0.43$	$0.43 \leq d < 0.48$	$d \geq 0.48$
	尖叶形	椭圆形	圆形
$a$	$a < 0.31$	$0.31 \leq a < 0.38$	$a \geq 0.38$
	下倾斜(垂眼)	水平	上倾斜(吊眼)

4 结 论

本文提出了在级联卷积神经网络定位人脸特征点的前提下,利用语义特征针对眼睛分类的方法,对不同眼型有良好的分类效果,且容易实现。可以把本文方法扩展到眉毛、鼻子、嘴巴等其他人脸部件,在人脸识别、亲缘关系识别中,通过对不同部件的形状进行细致分类后再进行分析,进而提高识别的效率和准确率。该方法的不足之处是对照片的分辨率要求较高,否则会导致特征点定位不准确而使分类效果有所下降。未来研究方向为尝试利用支持向量机和神经网络的方法对4种显著的眼型进行分类,与语义分类的效果进行比对。

参考文献

[1] 袁梦琪. 智能人脸模拟画像系统与实现[D]. 北京:中

国人民公安大学,2016.

[2] 张威. 东方人的眼形分类与美容术[J]. 医学美学美容, 1995(5):38-39.

[3] 孙劲光, 荣文钊. 曲线相似度眼型分类[J]. 计算机科学与探索, 2017, 11(8):1305-1313.

[4] 段萍. 基于眼部特征语义化人脸图像检索[D]. 大庆:东北石油大学,2013.

[5] 彭明超. 基于语义特征的人脸匹配算法研究与应用[D]. 成都:电子科技大学,2016.

[6] COOTES T F, TAYLOR C J, COOPER D H, et al. Active shape models-their training and application[J]. Computer Vision & Image Understanding, 1995, 61(1):38-59.

[7] 白中浩, 刘浏, 焦英豪, 等. 基于ASM的多特征融合驾驶员疲劳检测方法[J]. 电子测量与仪器学报, 2016, 30(12):1877-1883.

[8] COOTES T F, EDWARDS G J, TAYLOR C J. Active appearance models[C]. European Conference on Computer Vision. Springer Berlin Heidelberg, 1998:484-498.

[9] DOLLÁR P, WELINDER P, PERONA P. Cascaded pose regression[J]. IEEE, 2010, 238(6):1078-1085.

[10] SUN Y, WANG X, TANG X. Deep convolutional network cascade for facial point detection [C].

- Computer Vision and Pattern Recognition, IEEE, 2013:3476-3483.
- [11] ZHOU E, FAN H, CAO Z, et al. Extensive facial landmark localization with coarse-to-fine convolutional network cascade[C]. IEEE International Conference on Computer Vision Workshops, 2014:386-391.
- [12] ZHANG Z, LUO P, CHEN C L, et al. Facial landmark detection by deep multi-task learning[C]. European Conference on Computer Vision, 2014: 94-108.
- [13] WU Y, HASSNER T, KIM K, et al. Facial landmark detection with tweaked convolutional neural networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018,40(12):3067-3074.
- [14] ZHANG K, ZHANG Z, LI Z, et al. Joint face detection and alignment using multitask cascaded convolutional networks[J]. IEEE Signal Processing Letters, 2016, 23(10):1499-1503.
- [15] KOWALSKI M, NARUNIEC J, TRZCINSKI T. Deep alignment network: A convolutional neural network for robust face alignment[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2017:2034-2043.

### 作者简介

尚焱睿,1994 年出生,硕士研究生,主要研究方向为信号与信息处理。

E-mail:510381123@qq.com

卜凡亮,1965 年出生,工学博士、教授,主要研究方向为信号与信息处理。

E-mail:bufanliang@sina.com