一种基于 CLMF 的深度卷积神经网络模型

随婷婷 [1] 王晓峰 [1]

摘 要 针对传统人工特征提取模型难以满足复杂场景下目标识别的需求, 提出了一种基于 CLMF 的深度卷积神经网络 (Convolutional neural networks with candidate location and multi-feature fusion, CLMF-CNN). 该模型结合视觉显著性、多特征融合和 CNN 模型实现目标对象的识别. 首先, 利用加权 Itti 模型获取目标候选区; 然后, 利用 CNN 模型从颜色、亮度多特征角度提取目标对象的特征, 经过加权融合供目标识别; 最后, 与单一特征以及目前的流行算法进行对比实验, 结果表明本文模型不仅在同等条件下正确识别率得到了提高, 同时, 达到实时性要求.

关键词 图像识别, 深度学习, 卷积神经网络, 多特征融合

引用格式 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

DOI 10.16383/j.aas.2016.c150741

Convolutional Neural Networks with Candidate Location and Multi-feature Fusion

SUI Ting-Ting[1] WANG Xiao-Feng[1]

Abstract To solve the problem that the traditional manual feature extraction models are unable to satisfy object recognition in complex environment, an object recognition model based on convolutional neural networks with candidate location and multi-feature fusion (CLMF-CNN) model is proposed. The model combines the visual saliency, multi-feature fusion and CNN model to realize the object recognition. Firstly, the candidate objects are conformed via weighted Itti model. Consequently, color and intensity features are obtained via CNN model respectively. After the multi-feature fusion method, the features can be used for object recognition. Finally, the model is tested and compared with the single feature method and current popular algorithms. Experimental result in this paper proves that our method can not only get good performance in improving the accuracy of object recognition, but also satisfy real-time requirements.

Key words Image recognition, deep learning, convolutional neural networks (CNN), multi-feature fusion

Citation Sui Ting-Ting, Wang Xiao-Feng. Convolutional neural networks with candidate location and multi-feature fusion. *Acta Automatica Sinica*, 2016, **42**(6): 875−882

随着科学技术的飞速发展, 图像识别技术已从简单的理论融入到了大众的日常生活之中[1−2]. 从手机、电脑、打卡机等使用指纹身份识别, 到阿里巴巴发布的人脸识别支付技术, 都离不开图像识别. 然而, 在这个信息量爆炸的时代, 如何能够提高识别率意义重大, 直接关系到图像识别的实用性和安全性.

幸运的是, 深度学习的出现解决了如何自动学习出“优质特征”的问题[2−3]. 它通过模仿人脑分析学习的机制, 将分级信息处理过程引用到了特征表示上, 通过逐层特征变换, 将样本在原空间的特征表示变换到一个新特征空间, 从而使分类识别更加容易. 相比于人工构造特征的方法, 利用深度学习方法来学习特征, 能够更为丰富地刻画数据的内在信息[4].

深度卷积神经网络 (Convolutional neural networks, CNN) 作为深度学习的常用模型, 已成为众多科研领域的研究热点之一. 受到 Hubel-Wiesel 生物视觉模型的启发, LeCun 等于 1989 年首先提出了 CNN 模型, 解决了小规模的图像识别问题[5−6]. 但对于大规模的图像无法得到较好的效果. 直至 2012 年, Krizhevsky 等在传统的 CNN 模型上提出了深度的理念, 取得了不错的识别结果, 推进了图像识别技术[7]. 与传统识别算法相比, 它的输入不使用任何的人工特征, 避免了复杂繁琐的手动特征提取过程, 可实现自动特征学习, 在处理大规模的图像识别时同样具有优势. 目前, CNN 模型被广泛应用于图像识别领域之中[4, 7−9]. Ji 等通过延伸数据的空间维度, 提出一种 3D CNNs 模型[10], 用于人体运动

收稿日期 2015-11-03 录用日期 2016-03-24
Manuscript received November 3, 2015; accepted March 24, 2016
国家自然科学基金 (31170952), 国家海洋局项目 (201305026), 上海海事大学优秀博士学位论文培育项目 (2014bxlp005), 上海海事大学研究生创新基金项目 (2014ycx047) 资助
Supported by National Natural Science Foundation of China (31170952), Foundation of the National Bureau of Oceanography (201305026), Excellent Doctoral Dissertation Cultivation Foundation of Shanghai Maritime University (2014bxlp005), and Graduate Innovation Foundation of Shanghai Maritime University (2014ycx047)
本文责任编委 柯登峰
Recommended by Associate Editor KE Deng-Feng
1. 上海海事大学信息工程学院 上海 201306
1. College of Information Engineering, Shanghai Maritime University, Shanghai 201306

行为的识别之中, 取得了不错的识别效果. 2013 年, 徐姗姗等[11] 利用 CNN 模型对木材的缺陷进行识别, 降低时间消耗的同时, 获得了较高的缺陷识别精度. 2014 年, 贾世杰等将 CNN 模型用于商品图像分类中[12], 为电子商务软件提供了一种快捷、高效的分类过滤手段. 这无不说明 CNN 模型在图像识别方面的优势, 即高效特征抽取、权值共享、模型复杂度低的特点. 故本文采用 CNN 模型作为图像特征提取的基础模型.

然而, 在目标识别的初期阶段需要对目标对象进行定位 (Candidate location, CL), 这是 CNN 模型所忽略的. 近年来, 神经科学方面的研究者发现, 人类视觉系统具有快速定位兴趣目标的能力[13]. 显然, 将这种能力引入 CNN 模型, 无疑将提升目标识别的效率. 目前, 最具代表的是 Itti 模型[14−15], 它能模拟视觉注意机制, 利用颜色、亮度和朝向特征获取感兴趣区. 故采用 Itti 模型实现 CL 阶段.

同时, CNN 模型常采用灰度图像作为图像的输入, 缺失了对于颜色、亮度特征的理解. 而颜色特征对于图像的旋转、尺度变换和平移具有不错的稳定性[16]. 亮度是人类视觉系统较为敏感的图像特征. 若融合颜色、亮度特征, 能够更为完善地表达图像. 因此, 采用多特征融合的方法来表示图像具有一定的必要性.

综上所述, 为了能够使 CNN 模型更为快捷地实现 CL 阶段的目标定位, 多特征信息的互补, 本文以 CNN 模型为基础模型, 添加 Itti 模型以及多特征融合思想, 建立一种基于 CLMF 的深度卷积神经网络模型 (Convolutional neural networks with candidate location and multi-feature fusion, CLMF-CNN), 以便快速地获取目标区域, 提高目标识别效率和准确度.

1 深度卷积神经网络

深度卷积神经网络是第一个成功训练多层神经网络的学习算法. 由于该网络有效地避免了复杂的图像预处理, 可以自主学习图像特征, 所以得到了广泛的应用. CNN 模型通过对局部感受野卷积 (Local connections)、权值共享、下采样和多网络层[17], 实现 NN (Neural network) 结构的优化, 不但减少了神经元和权值的个数. 同时, 利用池化操作 (Pooling) 使特征具有位移、缩放和扭曲不变性[17].

典型的深度卷积网络结构如图 1 所示. 第一层为图像输入层, 然后由多个卷积层 (Convolution, C 层) 和下采样层 (Subsampling, S 层) 组成, 最后一层为全连接层.

1.1 C 层的学习

C 层主要是利用卷积核抽取特征, 实现对特征进行过滤和强化的效果. 在每个卷积层中, 将前一层输出的特征图与卷积核进行卷积操作[18], 然后通过激活函数, 即可输出该层的特征图 y_i^t, 如式 (1) 所示.

$$y_j^t = f\left(\sum_{i\in P_j} k_{i,j}^t * y_i^{t-1} + b_j^t\right) \tag{1}$$

其中, f 是激活函数, 本文选用 Sigmoid 函数. t 表示层数, $k_{i,j}$ 是卷积核, $*$ 表示 2D 卷积操作, b_j 是偏置, P_j 表示所选择的输入特征图的集合.

1.2 S 层的学习

S 层主要通过下采样减少 C 层的特征维数, 对 S 层中每个大小为 $n\times n$ 的池进行“池平均”或“池最大”操作[19], 以获取抽样特征, 如式 (2) 所示.

$$y_j^t = f(\text{down}(y_i^{t-1})\cdot w_j^t + b_j^t) \tag{2}$$

其中, w 为权重, down$(\cdot)$ 为下采样函数, 本文采用“池最大”操作. 通过池化操作, 不仅有效降低了 C 层的复杂度, 抑制了过拟合现象, 同时, 提升了特征对微小畸变、旋转的容忍能力, 增强了算法的性能和鲁棒性.

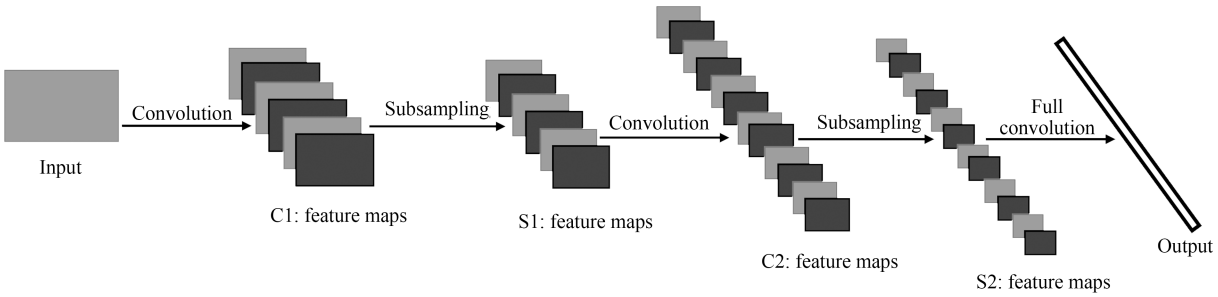


图 1 深度卷积神经网络的结构图

Fig. 1 The structure chart of CNN model

2　基于 CLMF 的深度卷积神经网络

为了使 CNN 模型能够在图像中快速搜索到目标对象, 模仿人脑视觉系统, 在 CL 阶段添加视觉注意模型, 旨在快速获取目标对象. 同时, 从特征融合的角度, 实现图像颜色、亮度的多特征表达. CLMF-CNN 的模型结构图如图 2 所示, 由候选目标区获取和多特征融合两模块组成.

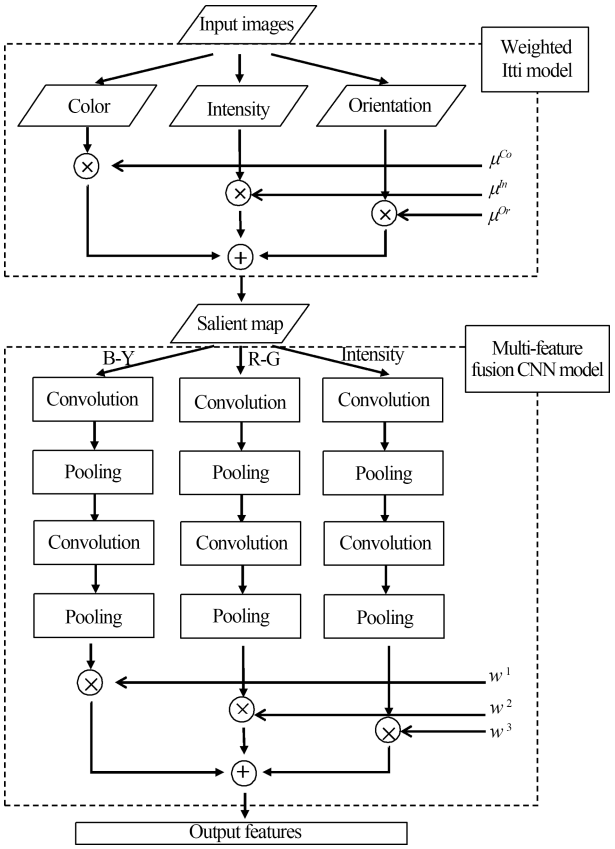


图 2　CLMF-CNN 模型结构图

Fig. 2　The structure chart of CLMF-CNN model

2.1　基于视觉显著性的候选目标获取

大量研究发现, 人类的视觉系统能够快速搜索到兴趣目标, 并进入视觉感知[20−21]. 受其启发, 若在目标识别的 CL 阶段采用视觉显著性获取候选目标, 能够有效地降低背景所带来的干扰. 目前最具代表性的是 Itti 等于 1998 年提出的选择注意模型, 该模型经过不断的改进, 已经可以较好地应用于目标识别之中. 其基本思想是采用自底向上的模式, 通过构建不同尺度的高斯金字塔, 并利用式 (3) ∼ 式 (5) 获取亮度、颜色、朝向特征[15]; 然后, 计算中央周边算子得到特征显著图; 最后, 通过归一化组合得到显著图, 从而模拟人类视觉系统选择出显著区域.

$$I = \frac{r+g+b}{3} \tag{3}$$

$$\begin{cases} RG(c,s) = |(R(c)-G(c)) \ominus (G(s)-R(s))| \\ BY(c,s) = |(B(c)-Y(c)) \ominus (Y(s)-B(s))| \end{cases} \tag{4}$$

$$O(c,s,\theta) = |O(c,\theta) \ominus O(s,\theta)| \tag{5}$$

其中, r、g、b 为三个颜色分量. $R = r-(g+b)/2$; $G = g-(r+g)/2$; $Y = (r+g)/2-|r-g|/2-b$; c、s 代表金字塔中央尺度和周边尺度. θ 为 Gabor 滤波器的方向; $\ominus$ 代表 “中央 – 周边” 算子.

然而, Itti 模型仅采用自底向上的机制, 缺失了高级认知的指导[14−15]. 特别地, 由其获取的显著图仅由各类特征叠加而成的, 这违背了视觉系统的选择机制. 在不同的环境下, 视觉系统搜索不同目标时, 对于各个特征的倚重应有所不同. 故综合考虑各类特征对于目标定位的贡献度, 赋予权重, 通过特征与权重的乘积和确定显著区, 如式 (6) 所示.

$$Sali = \sum_{j\in\{Co,In,Or\}} \beta^j \sum_{k=1}^{N} Sali_j^k \tag{6}$$

其中, β^j 为显著特征权重, 由式 (7) 获得. $Sali$ 代表显著值, $Sali_{Co}$ 为颜色显著值、$Sali_{In}$ 为亮度显著值、$Sali_{Or}$ 为朝向显著值, k 代表不同的尺度.

目前, 对于显著区域的提取多由目标知识驱动, 忽略了背景特征对于目标检测的抑制作用. 而神经物理学实验表明, 背景特征对于目标检测也具有重要意义[22]. 因此综合考虑目标和背景的概率知识, 利用式 (7) 确定显著特征权重 β^r.

$$\begin{aligned} \beta^r =& \frac{P(O|Fsali^r)}{P(O|Bsali^r)} = \\ & \frac{P(Fsali^r|O)P(O)}{P(Fsali^r)} \cdot \frac{P(Bsali^r)}{P(Bsali^r|O)P(O)} = \\ & \frac{P(Fsali^r|O)P(Bsali^r)}{P(Fsali^r)P(Bsali^r|O)} \end{aligned} \tag{7}$$

其中, β^r 表示显著特征权重, $P(O)$ 表示目标 O 出现的先验概率; $P(O|Fsali^r)$ 表示给定前景区的某一图像度量 $Fsali^r$ 时, 目标 O 出现的条件概率; $P(O|Bsali^r)$ 表示给定背景区某一图像度量 $Bsali^r$ 时, 目标 O 出现的条件概率; 图像度量包括颜色特征值 $Sali_{Co}$、亮度特征值 $Sali_{In}$ 和朝向特征值 $Sali_{Or}$.

2.2　多特征融合

由于 CNN 模型在特征提取过程中使用的特征单一, 忽略了颜色、亮度特征的影响, 如图 1 所示. 故本文在深度卷积神经网络的基础上, 添加颜色、亮

度特征提取的思想, 使用 B-Y 颜色通道、R-G 颜色通道以及亮度通道三通道对视觉图像进行特征提取. 其中, B-Y 和 R-G 颜色通道的图像表示可由式 (8) 和 (9) 获得.

$$P_{RG} = \frac{(r-g)}{\max(r,g,b)} \tag{8}$$

$$P_{BY} = \frac{b-\min(r,g)}{\max(r,g,b)} \tag{9}$$

因此, CLMF-CNN 模型不仅考虑了亮度特征, 同时考虑了对象的颜色特征, 使得特征向量更能表现目标对象的特性.

然而, 多特征的融合方法对于特征的表达能力具有一定的影响. 目前, 常用的多特征融合方法有简单叠加、串行连接等. 但这些方法不仅较难体现各种特征的差异性, 反而扩大了特征的维数, 增加了计算量. 因此, 引出权重的概念, 根据不同的特征在识别过程中的贡献度, 在 CNN 的全连接层后添加一层各类特征的权重计算层.

通常, 特征的识别效果采用误差率表示, 误差率越低则表示该类特征具有较强的区分能力. 受此启发, 从误差率的角度定义权重, 如式 (10) 所示.

$$w^n = \frac{\frac{1}{e^n}}{\sum\limits_{i=1}^{N}\left(\frac{1}{e^i}\right)} \tag{10}$$

其中, w^n 为特征 n 的权重, $0 \le w^n \le 1$ 且 $\sum_{n=1}^{N} w^n = 1$. e^n 表示特征 n 的误差率. 由此可以发现, e^n 越低的特征将获得越高的权重. 因此, 每个目标融合后的特征向量 $\boldsymbol{\mathcal{T}}$ 可表示为式 (11).

$$\boldsymbol{\mathcal{T}} = \sum_{n=1}^{N} w^n \boldsymbol{y}^n \tag{11}$$

其中, N 为特征类别数, $\boldsymbol{y}^n$ 表示特征 n 相应的特征向量.

2.3 算法流程

CLMF-CNN 模型由学习阶段以及目标识别阶段两部分组成. 具体步骤如下:

1) 学习阶段:

步骤 1. 根据学习样本, 采用样本统计分析法计算样本图像内目标对象与背景的条件概率 $P(O|Fsali^r)$ 和 $P(O|Bsali^r)$;

步骤 2. 根据式 (7) 确定 Itti 模型内的权重 β^j;

步骤 3. 利用 CNN 模型获取目标对象在 B-Y 颜色通道、R-G 颜色通道以及亮度通道三通道的特征向量;

步骤 4. 训练不同特征向量, 获取各类特征的误差率 e^n;

步骤 5. 根据误差率 e^n, 利用式 (10) 获取不同特征的权重.

2) 目标识别阶段:

步骤 1. 根据权重 β^j, 利用加权 Itti 模型获取测试图像相应的候选目标区域;

步骤 2. 利用 CNN 模型对候选目标进行 B-Y 颜色通道、R-G 颜色通道以及亮度通道三通道的特征提取;

步骤 3. 根据式 (11), 结合不同特征的权重 w^n 进行加权融合, 形成候选目标的特征表达;

步骤 4. 对候选目标进行识别, 输出测试图像的类别.

3 实验结果与分析

仿真实验平台配置为酷睿四核处理器 2.8 GHz, 8 GB 内存, 使用 Caltech 101 数据集, 该数据库包含 101 类, 每类大约包含 40 到 800 张彩色图片. 然而, CNN 模型需要建立在大量样本的基础上, 故选取其中样本量较大的 8 类: 飞机 (Airplanes)、人脸 (Faces)、钢琴 (Piano)、帆船 (Ketch)、摩托车 (Motor)、手枪 (Revolver)、手表 (Watch) 以及豹 (Leopards), 并利用 Google 对图库进行扩充, 每种类别选用 2 000 幅图像, 本文方法的参数设置如表 1 所示, 其中, 学习率初始值设为 0.1, 并在迭代过程中线性下降以寻找最优值. 同时, 为了评估识别效果, 采用十折交叉实验法进行验证, 并利用识别精度作为评价标准, 如式 (12) 所示.

$$PreVal_i = \frac{PT_i}{PT_i + FT_i} \tag{12}$$

其中, $PreVal_i$ 表示第 i 类图像的识别精度, PT_i 表示正确识别的样本数, FT_i 表示错误识别的样本数.

表 1 本文方法参数设置表

Table 1 Parameters setting of our method

层数	种类	特征图个数	卷积核大小
1	卷积层	100	7×7
2	下采样层	100	2×2
3	卷积层	150	4×4
4	下采样层	150	2×2
5	卷积层	250	4×4
6	下采样层	250	2×2
7	全连接层	300	1×1
8	全连接层	8	1×1
激活函数		Sigmoid	
损失函数		Mean square error	

3.1 CL 阶段提取候选目标的作用

由图 3 可以发现, 利用改进的 Itti 模型可以有效地在 CL 阶段提取目标候选区, 避免了背景的干扰, 便于后续 CLMF-CNN 模型的特征提取. 实验结果表明, 平均每幅图像的处理时间约为 62.76 ms. 显然, 在目标候选区的提取上消耗了一定的计算时间, 但是, 相应地减少了 30 %～50 % 的伪目标区域, 降低了识别干扰, 反而提高了识别效率. 从图 4 可以发现, 利用 Itti 模型改进的 CNN 模型的确提升了目标的识别精度.

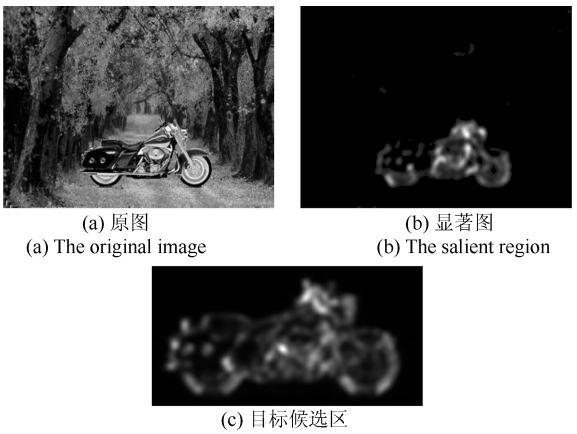


图 3　目标候选区域提取效果图

Fig. 3　The extraction of object candidate

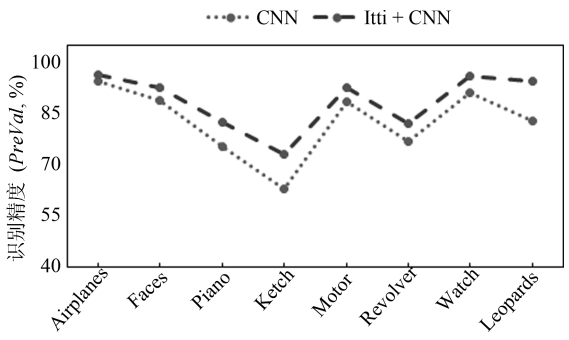


图 4　CNN 模型添加候选目标后的识别效果对比图

Fig. 4　The recognition performance of CNN model with candidate objects

为了进一步分析 CL 阶段目标定位的有效性, 选用覆盖率 (Overlap value, OV) 评价目标对象区界定的成功率, 如式 (13) 所示.

$$OV = \underset{i=1,2,\cdots,M}{\text{mean}}\left(\max_{j=1,2,\cdots,N}\left(\frac{prebox_{ij}\bigcap objbox_i}{prebox_{ij}\bigcup objbox_i}\right)\right) \tag{13}$$

其中, $prebox_{ij}$ 是图像 i 对应的第 j 个候选目标区域. $objbox_i$ 是图像 i 对应的目标区域.

由图 5 可以发现, 由于文献 [23] 利用固定窗口遍历搜索的方法, 所以对于脸、钢琴、手枪的定位效果较好. 然而, 对于飞机、帆船、豹等大小多变的目标对象, 界定的效果产生了一定的影响. 相反, 本文方法充分考虑了各项特征的贡献率, 能够较好地定位目标对象的区域, 为后期的目标识别提供了一定的保证.

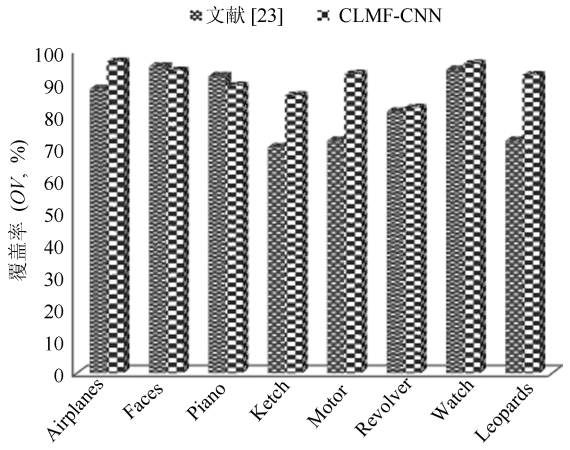


图 5　覆盖率对比图

Fig. 5　The comparison chat of OV

3.2 识别时间消耗对比

时间消耗无疑是对目标识别效果的一个重要评价指标. 图 6 从目标识别所需时耗的角度对比了文献 [23] 方法和 CLMF-CNN 模型. 由于文献 [23] 方法需要以固定大小的窗口遍历图像来实现目标的定位, 因此定位的时耗十分受滑动窗口大小以及图像大小的限制. 若以 30×30 的窗口遍历一幅 $N\times N$ 的图像时, 文献 [23] 方法在定位时将进行 $(N-29)^2$ 个操作. 若图像为 256×256, 则单幅图像在定位时的操作将超过 5 万次, 无疑增加了图像识别过程中的时间消耗. 相反, 由于 CLMF-CNN 模型采用视觉显著性定位的方法, 虽然在对单幅图像搜索目标时需要消耗时间用于定位显著区域, 但可以快速滤除图像中的伪目标区域, 大幅度地减少后期识别操作, 反而降低了目标识别的时间消耗, 十分有利于图像的快速识别.

3.3 特征融合的作用

在特征提取阶段, 采用了多特征融合方法, 利用各类特征的贡献度来分配权重. 为了验证权重的作用, 实验将本文的多特征融合方法与各类单一特征方法以及目前流行的多特征乘性融合方法[24]、多特征加性融合方法[25] 进行对比.

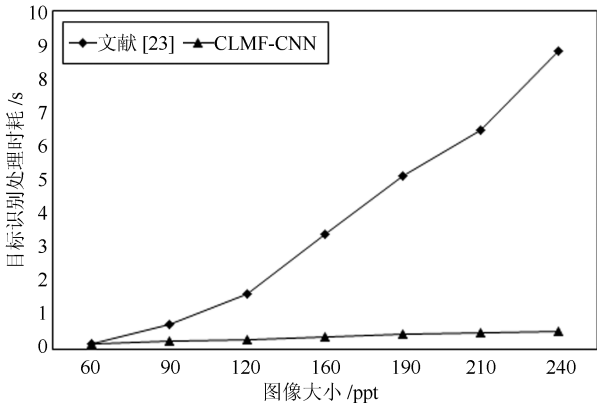


图 6 目标识别时耗对比图

Fig. 6 The comparison chat of time consumption on object recognition

从图 7 可以发现, 采用单一特征的 CNN 模型识别效果明显不佳, 且不稳定, 易受光照等外界因素的干扰. 说明需要通过特征融合, 使各类特征取长补短, 才能实现更好的识别效果. 文献 [24] 方法, 可实现各类特征的融合, 但该方法易放大噪声的影响, 导致融合结果对噪声较为敏感. 相反, 文献 [25] 在一定程度能够抑制噪声, 说明加性融合的确能较好地融合各类特征. 然而其识别效果仍不理想, 说明权重的分配对融合后特征向量的识别效果具有一定的影响. 本文的方法具有较好的识别结果, 原因在于: CLMF-CNN 模型充分考虑了各项特征对于识别效果的贡献度, 从误差率的角度分配各项权重, 降低了对于噪声的敏感度, 且提升了识别效果, 增强了识别方法的鲁棒性.

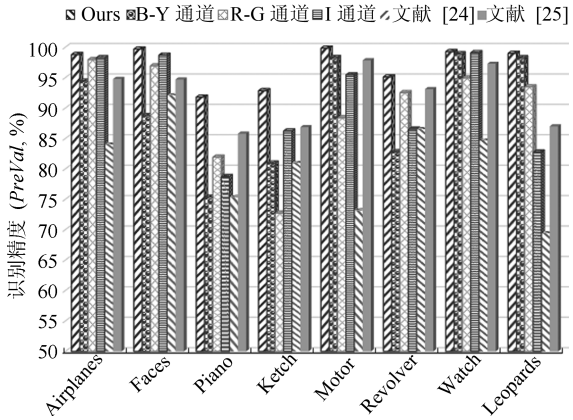


图 7 CNN 模型添加多特征后的识别效果对比图

Fig. 7 The recognition performance of CNN model with multi-features

3.4 识别效果对比

为了验证本文方法的有效性, 实验将 CLMF-CNN 模型和文献 [26−28] 的方法进行对比, 如图 8 所示. 其中, 对于人脸、摩托车和手表这些目标对象, CLMF-CNN 模型具有一定的优势. 原因在于, 这些目标较为显著, 对于 CLMF-CNN 模型更易找到目标对象区域. 而对于文献 [26−28] 方法, 由于过多的依赖固定窗口滑动搜索的方法, 导致对目标区域的定位有一定的偏差. 同时, 本文的多特征融合方法能够充分地考虑各类特征的贡献度, 合理地分配权重, 使得各类特征扬长避短, 更有效地表达目标对象. 由图 8 可以发现, CLMF-CNN 模型的识别效果基本优于其他方法, 为目标识别提供了一种较为有效的方法.

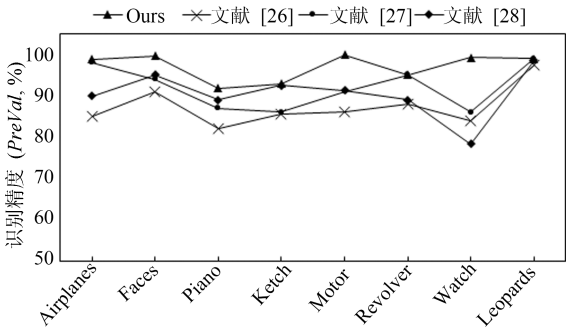


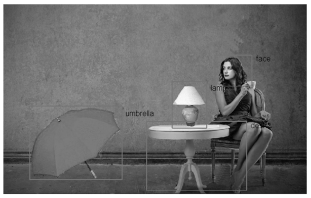
图 8 不同方法的分类效果对比图

Fig. 8 Recognition performance of different methods

同时, 为了进一步验证本文方法的识别效果, 实验将 CLMF-CNN 模型运用于图像标注中. 从表 2 可以发现, 本文方法基本可以标注出预先学习的目标对象, 说明 CLMF-CNN 模型可以较好地解决图像的自动标注问题.

表 2 CLMF-CNN 模型的图像标注效果

Table 2 The image annotation performance of CLMF-CNN

标识图像	标注信息
	飞机 船舶
	雨伞 台灯 书桌 人脸

4 结论

本文提出一种基于 CLMF 的卷积神经网络模

型, 并用于图像识别, 取得了较为满意的实验结果. 与现有方法相比, CLMF-CNN 具有以下几个突出的特点: 1) 模仿人脑视觉认知的过程添加了 CL 阶段的候选目标区选取模块, 确立了目标对象区, 减少了由于伪目标区域所造成的计算时间消耗和识别干扰. 2) 利用多特征的加权融合降低了由单一特征不充分所引起的歧义, 丰富了图像的特征表达.

然而, 图像质量对于目标识别具有一定影响. 下一步工作的重点将从图像融合技术文献 [29−30] 的角度提高图像质量, 进一步改善目标识别效果.

References

1 Sarikaya R, Hinton G E, Deoras A. Application of deep belief networks for natural language understanding. *IEEE/ACM Transactions on Audio, Speech, & Language Processing*, 2014, **22**(4): 778−784

2 Graves A, Mohamed A R, Hinton G. Speech recognition with deep recurrent neural networks. In: Proceedings of the 38th IEEE International Conference on Acoustics, Speech and Signal Processing. Vancouver, BC: IEEE, 2013. 6645−6649

3 Liu Jian-Wei, Liu Yuan, Luo Xiong-Lin. Research and development on deep learning. *Application Research of Computers*, 2014, **31**(7): 1921−1930
(刘建伟, 刘媛, 罗雄麟. 深度学习研究进展. 计算机应用研究, 2014, **31**(7): 1921−1930)

4 Najafabadi M M, Villanustre F, Khoshgoftaar T M, Seliya N, Wald R, Muharemagic E. Deep learning applications and challenges in big data analytics. *Journal of Big Data*, 2015, **2**: 1

5 LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 1998, **86**(11): 2278−2324

6 LeCun Y, Boser B, Denker J S, Henderson D, Howard R E, Hubbard W, Jackel L D. Backpropagation applied to handwritten zip code recognition. *Neural Computation*, 1989, **1**(4): 541−551

7 Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks. In: Proceedings of the Advances in Neural Information Processing Systems 25. Lake Tahoe, Nevada, USA: Curran Associates, Inc., 2012. 2012−2020

8 Wang Xin, Tang Jun, Wang Nian. Gait recognition based on double-layer convolutional neural networks. *Journal of Anhui University (Natural Science Edition)*, 2015, **39**(1): 32−36
(王欣, 唐俊, 王年. 基于双层卷积神经网络的步态识别算法. 安徽大学学报 (自然科学版), 2015, **39**(1): 32−36)

9 Ouyang W, Wang X. Joint deep learning for pedestrian detection. In: Proceedings of the 2013 IEEE International Conference on Computer Vision. Sydney, Australia: IEEE, 2013. 2056−2063

10 Ji S W, Xu W, Yang M, Yu K. 3D convolutional neural networks for human action recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, **35**(1): 221−231

11 Xu Shan-Shan, Liu Ying-An, Xu Sheng. Wood defects recognition based on the convolutional neural network. *Journal of Shandong University (Engineering Science)*, 2013, **43**(2): 23−28
(徐姗姗, 刘应安, 徐昇. 基于卷积神经网络的木材缺陷识别. 山东大学学报 (工学版), 2013, **43**(2): 23−28)

12 Jia Shi-Jie, Yang Dong-Po, Liu Jin-Huan. Product image fine-grained classification based on convolutional neural network. *Journal of Shandong University of Science and Technology (Natural Science)*, 2014, **33**(6): 91−96
(贾世杰, 杨东坡, 刘金环. 基于卷积神经网络的商品图像精细分类. 山东科技大学学报 (自然科学版), 2014, **33**(6): 91−96)

13 Unuma H, Hasegawa H. Visual attention and object perception: levels of visual features and perceptual representation. *Journal of Kawamura Gakuen Womans University*, 2007, **18**: 47−60

14 Serre T, Wolf L, Poggio T. Object recognition with features inspired by visual cortex. In: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR). San Diego, CA: IEEE, 2005. 994−1000

15 Itti L, Koch C, Niebur E. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 1998, **20**(11): 1254−1259

16 Yao Yuan-Qing, Li Feng, Zhou Shu-Ren. Target tracking based on color and the texture feature. *Computer Engineering & Science*, 2014, **36**(8): 1581−1587
(姚原青, 李峰, 周书仁. 基于颜色 – 纹理特征的目标跟踪. 计算机工程与科学, 2014, **36**(8): 1581−1587)

17 LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*, 2015, **521**(7553): 436−44

18 Huang F J, LeCun Y. Large-scale learning with SVM and convolutional for generic object categorization. In: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision & Pattern Recognition. New York, USA: IEEE, 2006. 284−291

19 Scherer D, Müller A, Behnke S. Evaluation of pooling operations in convolutional architectures for object recognition. In: Proceedings of the 20th International Conference on Artificial Neural Networks. Thessaloniki, Greece: Springer, 2010. 92−101

20 Serences J T, Yantis S. Selective visual attention and perceptual coherence. *Trends in Cognitive Sciences*, 2006, **10**(1): 38−45

21 Li Wan-Yi, Wang Peng, Qiao Hong. A survey of visual attention based methods for object tracking. *Acta Automatica Siinica*, 2014, **40**(4): 561−576
(黎万义, 王鹏, 乔红. 引入视觉注意机制的目标跟踪方法综述. 自动化学报, 2014, **40**(4): 561−576)

22 Maljkovic V, Nakayama K. Priming of pop-out: I. role of features. *Memory & Cognition*, 1994, **22**(6): 657−672

23 Roos M J, Wolmetz M, Chevillet M A. A hierarchical model of vision (HMAX) can also recognize speech. *BMC Neuroscience*, 2014, **15**(Suppl 1): 187

24 Li P H, Chaumette F. Image cues fusion for object tracking based on particle filter. In: Proceedings of the 3rd International Workshop on Articulated Motion and Deformable Objects. Palma de Mallorca, Spain: Springer, 2004. 99−110

25 Wang X, Tang Z M. Modified particle filter-based infrared pedestrian tracking. *Infrared Physics & Technology*, 2010, **53**(4): 280−287

26 Zhu Qing-Sheng, Zhang Min, Liu Feng. Hierarchical citrus canker recognition based on HMAX features. *Computer Science*, 2008, **35**(4): 231−232
(朱庆生, 张敏, 柳锋. 基于 HMAX 特征的层次式柑桔溃疡病识别方法. 计算机科学, 2008, **35**(4): 231−232)

27 Tang Yu-Jing. Classification and Recognition Research based on Human Visual Perception Mechanism [Master dissertation], Nanjing University of Science and Technology, China, 2009.
(汤毓婧. 基于人脑视觉感知机理的分类与识别研究 [硕士学位论文], 南京理工大学, 中国, 2009.)

28 Wang J, Yang J, Yu K, Lv F, Huang T, Gong Y. Locality-constrained linear coding for image classification. In: Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). San Francisco, CA: IEEE, 2010. 3360−3367

29 Zhang Xiao-Li, Li Xiong-Fei, Li Jun. Validation and correlation analysis of metrics for evaluating performance of image fusion. *Acta Automatica Sinica*, 2014, **40**(2): 306−315
(张小利, 李雄飞, 李军. 融合图像质量评价指标的相关性分析及性能评估. 自动化学报, 2014, **40**(2): 306−315)

30 Yang Bo, Jing Zhong-Liang. Image fusion algorithm based on the quincunx-sampled discrete wavelet frame. *Acta Automatica Sinica*, 2010, **36**(1): 12−22
(杨波, 敬忠良. 梅花形采样离散小波框架图像融合算法. 自动化学报, 2010, **36**(1): 12−22)

随婷婷 上海海事大学信息工程学院博士研究生. 2013 年获得上海海事大学信息工程学院硕士学位. 主要研究方向为深度学习, 人工智能, 数据挖掘与知识发现. 本文通信作者.
E-mail: suisui61@163.com
(**SUI Ting-Ting** Ph. D. candidate at the College of Information Engineering, Shanghai Maritime University. She received her master degree from the College of Information Engineering, Shanghai Maritime University in 2013. Her research interest covers deep learning, artificial intelligence, data mining and knowledge discovery. Corresponding author of this paper.)

王晓峰 上海海事大学教授, 博士. 主要研究方向为深度学习, 人工智能, 数据挖掘与知识发现.
E-mail: xfwang@shmtu.edu.cn
(**WANG Xiao-Feng** Ph. D., professor at Shanghai Maritime University. His research interest covers deep learning, artificial intelligence, data mining and knowledge discovery.)