论文

基于卷积神经网络与Bayesian决策的图像识别与分类记忆建模

姜英, 王延江*, 林青, 刘伟锋

中国石油大学(华东)信息与控制工程学院, 青岛 266580

* E-mail: yjwang@upc.edu.cn

收稿日期: 2016-11-21; 接受日期: 2017-07-20; 网络出版日期: 2017-09-07

国家自然科学基金项目(批准号: 61271407, 61671480)和中央高校基本科研业务费专项资金项目(编号: 14CX06066A)资助

摘要 当前热门图像分类方法大多侧重在分类能力, 忽视识别新事物, 然而人类认识事物时侧重认识, 只在细小之处重视分类, 这一点与人类记忆机制密切相关. 尽管目前有许多记忆建模理论被相继提出, 但大多以单词列表的形式学习, 对自然图像列表的研究有限. 基于此, 本文提出了基于卷积神经网络与Bayesian决策的图像识别分类记忆建模方法, 首先利用卷积神经网络提取图像特征, 并采用二进制形式存储特征向量; 然后进行视觉图像的表达, 存储与提取记忆建模, 将测试图像特征向量与所有已存储特征向量进行匹配对比, 计算似然率值; 最后在所有似然率基础上计算测试图像是新类别的几率, 若该几率大于某个阈值则判别其为新类别; 反之, 利用Bayesian决策规则进行图像分类. 在Caltech-101与Caltech-256数据库上的实验表明所提方法能很好地应用于图像识别分类任务中, 其击中率比目前代表性的稀疏表达分类(SRC)以及极限学习机(ELM)方法高, 且虚报率比其他两种方法低的多.

关键词 卷积神经网络, Bayesian决策规则, 记忆建模, 图像识别与分类

1 引言

计算机视觉是人工智能非常重要的应用领域, 其融合了计算机图形学、图像处理、模式识别、生理学、人工神经网络等多种应用理论. 计算机视觉研究的主要目标是使得计算机能够像人类那样轻易地对周围环境进行感知、理解并快速反应. 作为计算机视觉研究的一个非常活跃的方向, 视觉图像分类与识别研究已有50多年的历史, 其在很多领域中应用广泛, 如银行系统的人脸识别、防御系统的行人检测与跟

踪、交通系统的车牌检测与车辆跟踪等. 图像分类识别的理论与算法层出不穷, 如最近邻分类器^[1]、神经网络分类器^[2]、支持向量机分类器(SVM)^[3]、稀疏表达分类器(SRC)^[4]、协同表示分类器(CRC)^[5]、线性回归分类器(LRC)^[6]以及极限学习机(ELM)^[7]等. 其中, 作为非常受欢迎的分类器之一, 支持向量机分类器在高维模式识别中表现出许多特有的优势, 近几年研究人员还提出了最小二乘支持向量机(LS-SVM)^[8]以及近似支持向量机(PSVM)^[9]等分类器算法. 而极限学习机将最小二乘支持向量机与近似支持向量机等统一到

引用格式: 姜英, 王延江, 林青, 等. 基于卷积神经网络与Bayesian决策的图像识别与分类记忆建模. 中国科学: 技术科学, 2017, 47: 977-984
Jiang Y, Wang Y J, Lin Q, et al. A memory model for image recognition and classification based on convolutional neural network and Bayesian decision (in Chinese). Sci Sin Tech, 2017, 47: 977-984, doi: [10.1360/N092016-00405](https://doi.org/10.1360/N092016-00405)

一个总的框架中. 最近, 深度学习作为机器学习的一个新兴领域, 因其在模式识别中的优秀表现, 吸引了越来越多的关注, 主要包括自编码网络^[10]、受限玻尔兹曼机^[11]、深度置信网络^[12]、循环神经网络^[13]和卷积神经网络^[14]等.

尽管当前热门图像分类方法能展示很好的分类性能, 但是大多都侧重在分类能力上, 缺少识别能力. 它们大多侧重于“区别”, 即一类物体与有限类已学过物体的区分. 然而人类认识事物的过程是对一类物体与无限类未知物体的区分, 侧重于“认识”, 只在细小之处重视“区别”(如区别桌子与椅子或者鱼与海豚等). 传统的模式识别方法忽略了“认识”, 会将首次遇到的从未学习过的物体误认为是某一个已学习过的物体. 人类对首次遇到的新物体的直接反应是从未见过或者不认识该物体, 而不是直接判断其属于哪一类已经学习过的物体.

众所周知, 人类能对周围环境进行感知, 并能轻而易举地对视觉图像进行识别与分类, 与人类的认知、注意与记忆机制都是密不可分的. 其中记忆机制尤为重要, 包括前文提及的人类先侧重“认识”再“区别”都与人类记忆机制有关. 人类所看到与经历过的所有事情都需要经过人类记忆系统进行处理. 人类记忆机制在认知心理学、认知神经科学以及认知信息学等领域得到广泛研究. Murdock等人^[15]指出任何记忆模型至少要说明4个问题: 如何表达信息, 被存储与提取的信息种类, 存储与提取运算的本质以及信息存储的格. 围绕这些问题, 自20世纪80年代以来, 许多记忆建模理论被相继提出, 主要包括情景记忆建模、语义记忆建模及神经计算建模, 如SAM模型^[16], MINERVA2模型, SLiM模型, BCDMEM模型, TCM模型^[17], REM模型^[18], Rumellart模型^[19]及CLS模型^[20]等. 其中REM模型可以解释情景记忆研究中许多现象, 例如列表长度效应、列表强度效应、词频效应与镜面效应等, 并且因其坚实的数学基础表现突出, 吸引了众多研究者对其进一步研究. 然而上述记忆模型包括REM模型大多以单词列表的形式学习, 为单词列表的情景记忆现象提供解释, 针对自然图像分类的记忆研究有限.

基于此, 本文将认知心理学中的人类记忆建模方法引入到计算机视觉处理中, 提出了基于卷积神经网络与Bayesian决策的图像识别与分类方法: (1) 利用卷积神经网络提取图像特征向量, 并将所有已学习图像

特征向量在记忆空间中以二进制形式存储, 且所存储的特征以一定的概率被正确复制; (2) 在测试图像与所有已学习图像的匹配结果基础上计算二者之间的似然率值, 从而得到测试图像与每类图像之间的类别似然率值; (3) 计算测试图像是旧而非新图像类别的几率, 并根据Bayesian决策判断测试图像所属图像类别.

2 人类认知记忆REM模型

学习与记忆研究的主要目标是理解人类如何存储与提取信息, 计算机的出现为人类记忆模型的理论实现提供了平台. 众所周知, 认知心理学家根据记忆信息在大脑中维持时间的长短将记忆分为瞬时记忆、短时记忆与长时记忆. 根据提取阶段是否需要意识参与, 又将人类长时记忆分为外显记忆与内隐记忆. 而外显记忆又可以被分为两大类: 情景记忆与语义记忆. 自20世纪80年代以来, 许多记忆建模理论被陆续提出, 包括情景记忆建模、语义记忆建模以及神经计算建模等, 如SAM模型, TCM模型, REM模型, Rumellart模型及CLS模型等.

REM模型由Shiffrin等人^[18]提出, 其可以预测情景记忆研究中的许多现象, 如列表长度、列表强度与词频效应等. REM模型假设人类记忆由图像构成, 每幅图像被表示为一个特征向量. 被存储在记忆中的向量是对已学习向量的一个不完整且容易储存的复制. 学习某一个单词时, 对每一个特征位置存储信息的概率是 u^* , 且该值从已学习向量中正确复制的概率为 c . 还有 $1-c$ 的概率根据几何分布随机取值, 并且允许偶然取到正确值. 如果某个特征位置上已经存储过某个值, 那么该值不再发生改变.

给定测试单词, 将其与列表中的所有已学习单词向量进行匹配, 匹配结果记为 $D = \{D_j\}_{j=1, \dots, N}$, 其中 D_j 代表测试单词向量与列表中的第 j 个单词特征向量的对比匹配结果. 对有信息存储的位置标记特征值匹配与不匹配的区域, 忽略没有信息的特征位置.

在观测到的匹配结果 $D = \{D_j\}_{j=1, \dots, N}$ 的基础上分别计算似然率 λ_j , 似然率的计算在REM记忆模型中是非常关键的部分. 然后利用Bayesian决策来计算测试单词是旧而非新的几率 ϕ . 计算所得的几率用于反映对某个项目的熟悉度感觉, 并应用于生成识别决策. 若 $\phi > 1$, 测试单词被认为是已学习过的, 且其与

最大 λ_j 值对应的单词匹配;反之,认为测试单词是新单词,并对其进行存储。

3 基于记忆模型的视觉图像表达、存储与提取

3.1 卷积神经网络特征表达

卷积神经网络(convolutional neural network, CNN)^[14]是深度监督学习下的机器学习模型,已经成为众多科学领域的研究热点之一,其可以从原始图像中识别视觉模式,避免了对图像的复杂的预处理工作。卷积神经网络可以挖掘图像数据的局部特征,提取全局训练特征并进行分类,在模式识别的许多领域有着优秀的成果,已成功地应用到人脸识别、行人检测与机器人导航等。

卷积神经网络是一个前馈神经网络,其从图像中提取拓扑结构,采用反向传播算法优化网络结构,并求解网络中的未知参数。CNN通过局部感受野、权值共享与池化,以达到简化网络参数并获得一定的平移、尺度与形状不变性。一个完整的卷积神经网络大致包括输入层、卷积层、下采样层、全连接层与输出层,基本框架如图1所示。

在图像处理中,通常将图像像素值作为输入数据,卷积神经网络能直接处理最初的输入数据。卷积层,也被称为特征提取层,该层中有多个不同的特征图,每个特征图提取一种特征,因此在卷积层中能提取多种不同特征。提取特征时,同一特征图的权值是共享的,即使用相同的卷积核,不同的特征图使用不同的卷积核。卷积层将不同的局部特征保存下来,使得所提取的特征具有旋转与平移不变性。下采样层,也被称为特征映射层,作用是负责将上一层卷积层获得的特征

进行子采样,使得所提取的特征具有缩放不变性。卷积神经网络中的卷积层紧跟着一个特征映射层,这种特有的两次特征提取结构使得网络对输入图像样本具有较高的畸变容忍能力。卷积神经网络包含多个全连接层,全连接层中的每一个神经元都与上一层的所有神经元相连接。在CNN的末层输出节点个数是分类目标数目,网络训练的目的就是使CNN的输出尽可能与原始的标签相近。

本文采用已经训练好的VGGNet-16模型^[21]对图像进行特征提取,该模型由Simonyan和Zisserman^[21]提出,他们建立了一个从Alex-net发展而来的深度网络。给定图像集合:

$$I = \{I_{11}, \dots, I_{1n_1}, I_{21}, \dots, I_{2n_2}, \dots, I_{M1}, \dots, I_{Mn_M}\},$$

其中 I_{ij} 代表属于第 i 个图像类别的第 j 幅图像,该图像集合由 M 个不同类别的图像构成,每个类别 ω_k 包含 n_k 幅图像,利用VGGNet-16模型提取每幅图像的特征向量,得到所有图像的特征集合:

$$V^0 = \{V_{11}^0, \dots, V_{1n_1}^0, V_{21}^0, \dots, V_{2n_2}^0, \dots, V_{M1}^0, \dots, V_{Mn_M}^0\},$$

其中 $V_{ij}^0 \in R^{1 \times K}$ 代表图像 I_{ij} 的卷积神经网络特征向量, K 为特征维数。

3.2 视觉特征存储

正如在REM模型所描述的,我们认为在人脑记忆存储过程中,特征向量的存储并不是完全正确的,是不完整的复制过程。对特征向量上的每个特征位置,该位置上有某个值存储的概率是 u^* 。注意到此时所存储的值可能是正确复制的特征值,也可能是随机取值。正确复制到特征值的概率是 c ,通过指数分布随机取值的概率是 $1-c$ 。

图1 (网络版彩图)卷积神经网络基本框架

$$g[V = x] = \begin{cases} ge^{-gx}, & x > 0, \\ 0, & \text{其他}, \end{cases} \quad (1)$$

式中, g 是指数分布的参数. 经过上述不完整复制过程之后, 之前所提取到的每一个特征向量 V_{ij}^0 的存储结果用 V_{ij}^1 表示, 那么此时所得到的所有图像特征集合表示为 $V^1 = \{V_{ij}^1\}_{i=1, \dots, M, j=1, \dots, n_i}$. 事实上, 利用卷积神经网络提取图像特征时, 已经有部分信息被丢失, 因此我们假设特征提取过程隐含上述不完整复制过程, 即 V_{ij}^0 被认为是已经过不完整复制过程之后的特征集合. 此时我们得到的特征集合中的特征值有 0, 也有非零浮点数. 众所周知, 人类大脑是一个由约 140 亿个神经元组成的繁复的神经网络, 当人类在存储信息时, 仅有很少部分的神经元处于活跃状态. 因此所得到的特征向量越稀疏越便于模拟人脑记忆过程. 为此, 存储图像特征向量时, 将特征值大于某个给定阈值的特征位置用数值 1 代替, 意味着该特征位置处于激活状态, 其余位置用数值 0 代替. 最终存储的特征集合用 $V = \{V_{11}, \dots, V_{Mn_M}\}$ 来表示, 是一个只含有 0 与 1 元素的简单矩阵, 便于在计算机中进行存储.

3.3 提取

给定测试图像 I_{test} , 其可能属于某个已学习图像类别, 也可能是一个新的图像, 采用卷积神经网络提取其特征向量 $V_{\text{test}}^0 \in R^{1 \times K}$. 接着对该特征向量进行预处理, 将特征值小于给定阈值的位置用 0 取代, 其余特征值不变, 得到测试图像的特征向量 V_{test} , 将其与已存储特征集合 $V = \{V_{11}, \dots, V_{Mn_M}\}$ 中的每个特征向量进行匹配对比. 正如之前提到的, 已学习特征向量中特征值为 1 意味着该特征位置被激活, 特征值为 0 意味着该位置未被激活, 定义测试特征向量 V_{test} 与已学习特征向量 V_{ij} 之间的匹配规则如下:

$$D_{ij}(k) = \begin{cases} V_{\text{test}}(k), & \text{若 } V_{\text{test}}(k) \neq 0 \text{ 且 } V_{ij}(k) = 1, \\ 1 - c, & \text{若 } V_{\text{test}}(k) \neq 0 \text{ 且 } V_{ij}(k) = 0, \\ 1, & \text{其他}, \end{cases} \quad (2)$$

式中, $1 \leq k \leq K$, D_{ij} 表示测试特征向量 V_{test} 与已学习特征向量 V_{ij} 之间的匹配结果, 那么测试特征向量与列表中的所有已学习特征向量的整体匹配结果用 $D = \{D_{11}, \dots, D_{Mn_M}\}$ 表示. 为便于理解, 用 s-图像表示与测试图像的相同的已存储图像, d-图像表示除测试

图像之外的其他存储图像, 图像 I_{ij} 是 s-图像与 d-图像的事件分别记为 S_{ij} 与 N_{ij} . 在观测到的匹配结果 D_{ij} 基础上计算测试图像 I_{test} 与已学习图像 I_{ij} 之间的似然率 λ_{ij} , 即计算图像 I_{ij} 是 s-图像的概率除以其是 d-图像的概率:

$$\begin{aligned} \lambda_{ij} &= \frac{P(S_{ij} | D_{ij})}{P(N_{ij} | D_{ij})} = \frac{P(D_{ij} | S_{ij})}{P(D_{ij} | N_{ij})} \\ &= \prod_{k=1}^K \frac{P(V_{k,ij} | S_{ij}, V_k)}{P(V_{k,ij} | N_{ij}, V_k)} \\ &= \prod_{k \in Q} \frac{P(V_{k,ij} | S_{ij}, V_k)}{P(V_{k,ij} | N_{ij}, V_k)} \prod_{k \in T} \frac{P(V_{k,ij} | S_{ij}, V_k)}{P(V_{k,ij} | N_{ij}, V_k)}, \end{aligned} \quad (3)$$

式中, V_k 指测试特征向量 V_{test} 的第 k 个特征值, $V_{k,ij}$ 则指已学习特征向量 V_{ij} 的第 k 个特征值. T 和 Q 分别代表匹配与不匹配的非零特征的序号集. 当 $k \in Q$ 时,

$$P(V_{k,ij} | S_{ij}, V_k) = (1 - c)P(V_{k,ij} | N_{ij}, V_k), \quad (4)$$

用 $n_{ij,q}$ 标记已学习图像 I_{ij} 在匹配过程中的所有不匹配的非零特征数目, 将式(1)与(4)代入到式(3)中得到

$$\begin{aligned} \lambda_{ij} &= (1 - c)^{n_{ij,q}} \prod_{k \in T} \frac{P(V_{k,ij} | S_{ij}, V_k)}{P(V_{k,ij} | N_{ij}, V_k)} \\ &= (1 - c)^{n_{ij,q}} \prod_{k \in T} \frac{c + (1 - c)g(V_{k,ij})}{g(V_{k,ij})} \\ &= (1 - c)^{n_{ij,q}} \prod_{k \in T} \frac{c + (1 - c)ge^{-gV_{k,ij}}}{ge^{-gV_{k,ij}}}, \end{aligned} \quad (5)$$

那么测试图像 I_{test} 与所有已学习图像

$$I = \{I_{ij}\}_{i=1, \dots, M, j=1, \dots, n_i}$$

之间的似然率标记为:

$$\lambda = \{\lambda_{ij}\}_{i=1, \dots, M, j=1, \dots, n_i}.$$

4 基于Bayesian决策的图像识别分类规则

4.1 识别

首先要确定探测图像 I_{test} 属于旧类别还是新类别. 令 Φ 表示测试图像 I_{test} 是旧类别的概率除以其是新类别的概率, 假设先验概率相同, 那么有

$$\begin{aligned}
\Phi &= \frac{P(O|D)}{P(N|D)} = \frac{P(D|O)}{P(D|N)} = \sum_{i=1}^M \sum_{j=1}^{n_i} \frac{P(D|S_{ij})P(S_{ij})}{P(D|N)} \\
&= \sum_{i=1}^M \sum_{j=1}^{n_i} \frac{1}{M} \frac{1}{n_i} \frac{P(D|S_{ij})}{P(D|N)} = \frac{1}{M} \sum_{i=1}^M \frac{1}{n_i} \left(\sum_{j=1}^{n_i} \frac{P(D|S_{ij})}{P(D|N)} \right) \\
&= \frac{1}{M} \sum_{i=1}^M \frac{1}{n_i} \left(\sum_{j=1}^{n_i} \frac{P(D_{ij}|S_{ij}) \prod_{i' \neq i, j' \neq j} P(D_{ij'}|N_{ij'})}{P(D_{ij}|N_{ij}) \prod_{i' \neq i, j' \neq j} P(D_{ij'}|N_{ij'})} \right) \\
&= \frac{1}{M} \sum_{i=1}^M \frac{1}{n_i} \left(\sum_{j=1}^{n_i} \frac{P(D_{ij}|S_{ij})}{P(D_{ij}|N_{ij})} \right) \\
&= \frac{1}{M} \sum_{i=1}^M \sum_{j=1}^{n_i} \frac{1}{n_i} \lambda_{ij}. \tag{6}
\end{aligned}$$

若 $\Phi > \lambda_{\text{given}}$, λ_{given} 为给定阈值, 那么测试图像 I_{test} 被认为属于已学习类别; 反之, 则认为测试图像是从未学习过, 属于新类别。

4.2 分类

当 $\Phi > \lambda_{\text{given}}$ 时, 需要进一步确定测试图像 I_{test} 所属的已学习图像类别。为便于理解, 假设只有两个不同类别, ω_u 与 ω_v , $1 \leq u, v \leq M$ 。记给定 I_{test} 下类别 ω_u 和 ω_v 的后验概率分别为 $P(\omega_u|I_{\text{test}})$ 与 $P(\omega_v|I_{\text{test}})$, 那么给出基于最小误差的Bayesian决策规则如下:

如果 $P(\omega_u|I_{\text{test}}) > P(\omega_v|I_{\text{test}})$, 那么测试图像 I_{test} 属于图像类别 ω_u ; 反之测试图像属于类别 ω_v 。

令 $\alpha = P(\omega_u|I_{\text{test}}) / P(\omega_v|I_{\text{test}})$, 那么决策规则简化为若 $\alpha > 1$, 则 $I_{\text{test}} \in \omega_u$; 反之, $I_{\text{test}} \in \omega_v$ 。

已知属于类别 ω_u 与 ω_v 的已学习图像集合分别为 $\{I_{uj}\}_{j=1 \dots n_u}$ 与 $\{I_{vj}\}_{j=1 \dots n_v}$, 则计算 α :

$$\begin{aligned}
\alpha &= \frac{P(\omega_u|I_{\text{test}})}{P(\omega_v|I_{\text{test}})} \propto \frac{P(D|S_u)}{P(D|S_v)} \\
&= \frac{\sum_{j=1}^{n_u} P(D|S_{uj})P(S_{uj})}{\sum_{j=1}^{n_v} P(D|S_{vj})P(S_{vj})} = \frac{\frac{1}{M} \cdot \frac{1}{n_u} \sum_{j=1}^{n_u} P(D|S_{uj})}{\frac{1}{M} \cdot \frac{1}{n_v} \sum_{j=1}^{n_v} P(D|S_{vj})} \\
&\propto \frac{\frac{1}{n_u} \sum_{j=1}^{n_u} P(D_{uj}|S_{uj})}{\frac{1}{n_v} \sum_{j=1}^{n_v} P(D_{vj}|S_{vj})} \propto \frac{\frac{1}{n_u} \sum_{j=1}^{n_u} \lambda_{uj}}{\frac{1}{n_v} \sum_{j=1}^{n_v} \lambda_{vj}}. \tag{7}
\end{aligned}$$

那么图像分类的Bayesian决策规则定义如下:

$$I_{\text{test}} \in \begin{cases} \omega_u, & \text{若 } \frac{1}{n_u} \sum_{j=1}^{n_u} \lambda_{uj} > \frac{1}{n_v} \sum_{j=1}^{n_v} \lambda_{vj}. \\ \omega_v, & \text{其他.} \end{cases} \tag{8}$$

类似地, 对于多类别分类任务, 已学习图像类别集合 $\omega = \{\omega_1, \omega_2, \dots, \omega_M\}$, 给出多类别的Bayesian分类决策规则如下:

$$I_{\text{test}} \in \omega_u,$$

若

$$\frac{1}{n_u} \sum_{j=1}^{n_u} \lambda_{uj} > \frac{1}{n_v} \sum_{j=1}^{n_v} \lambda_{vj},$$

对所有的 $v \neq u, 1 \leq u, v \leq M$ 。

5 结果与分析

为了验证所提方法的有效性, 实验以MATLAB R2012a为仿真平台, 在Intel(R) Pentium(R) G620 @2.60 GHz的CPU, 4G内存, win7操作系统平台实现。在Caltech-101^[22]与Caltech-256^[23]数据库上分别进行实验, 同时将本文方法与目前比较有代表性的两种算法SRC^[4]和ELM^[7]进行对比实验并给出识别分类性能对比结果。

5.1 Caltech-101数据库

Caltech-101数据库由101个图像类别组成, 共有8677幅图像用于分类测试, 每一类中包含大概40~800幅不同图像。图2展示了该数据库上的部分图像。

训练时, 从数据库中随机选取20个类别, 每类随机选取20幅图像, 共有400幅训练图像。然后从这20个类别中每类随机选取10幅图像(训练图像除外)共200幅图像进行测试, 同时随机选取20个新类别(每类随机选取10幅图像)参与测试, 即测试图像集合包含200幅已学习过类别图像与200幅新类别图像。实验中所提取的图像特征是利用VGGNet-16模型提取的卷积神经网络特征, 其维数大小为1×4096, 实验参数选择 $c=0.01$, $g=0.4$, $\lambda_{\text{given}}=200$ 。表1给出了本文方法与SRC, ELM方法在该数据库上的识别分类性能对比实验结果。

从表1可以看出, 对于本判别为“已学习过”的训练图像, 三种方法均提供了在该数据库上的良好分类性能, 但本文方法的击中率要比SRC与ELM方法都要

高;同时对于那些属于新类别的训练图像,本文方法的虚报率(29.5%)明显远远低于其他两种方法(71.8%, 100%). 本文方法在良好分类性能的基础上还有着比传统模式识别方法更低的虚报率,这正是其区别于其他传统模式识别方法的关键之处.

5.2 Caltech-256数据库

Caltech-256数据库包含256个图像类别的共30607幅图像,在该数据库中每个类别包含不少于80幅属于该类别的不同图像. Caltech-256数据库上的部分示例图像如图3所示.

实验中,从数据库中随机选取20个类别,每个类别选取20幅不同图像进行训练,共有400幅训练图像. 测试图像集包含40个类别,其中20个类别与训练类别相同,还有20个随机选取的新类别,每个类别随机选取10幅图像进行测试,共有400幅测试图像. 实验中所提取的图像特征是利用 VGGNet-16 模型提取的卷积神

图2 (网络版附图) Caltech-101数据库示例

表1 Caltech-101数据库上三种方法识别分类性能对比(%)^{a)}

方法	$P(H)$	$P(F)$
本文	80.2	29.5
SRC	77.9	71.8
ELM	75.8	100

a) 击中率 $P(H)$ 指当目标出现时认为其是“已学习过”的概率;虚报率 $P(F)$ 指当干扰项出现时认为其是“已学习过”的概率

图3 (网络版附图) Caltech-256数据库示例

表2 Caltech-256数据库上三种方法识别分类性能对比(%)^{a)}

方法	$P(H)$	$P(F)$
本文	75.6	30.5
SRC	73.7	73.7
ELM	72.7	100

a) $P(H)$ 与 $P(F)$ 分别代表击中率与虚报率

经网络特征,其维数大小为 1×4096 ,实验参数选择 $c=0.01, g=0.4, \lambda_{\text{given}}=280$. 将本文方法与SRC, ELM方法在Caltech-256数据库上进行识别分类对比实验,实验结果如表2所示.

很显然,本文方法的击中率高于其他两种方法,且其虚报率(30.5%)远远低于其他两种方法(73.7%, 100%),其在虚报率方面比传统模式识别方法有着更好的表现.

6 结论与讨论

受人类认知记忆模型启发,本文提出了一种基于卷积神经网络与Bayesian决策的图像识别分类记忆建模方法. 该方法首先利用卷积神经网络提取图像特征,并采用二进制形式进行存储特征向量,便于计算机存储;然后在记忆模型基础上进行视觉图像的表达、存储与提取建模,将测试图像特征向量与所有已存储特征向量进行匹配对比,计算似然率值;最后给出在Bayesian决策下的视觉图像的分类与识别规则,先对测试图像进行识别判断,若其为已学习类别再根据Bayesian分类决策规则判断测试图像所属类别. 在Caltech-101与Caltech-256数据库上分别进行识别分类实验,并构造与SRC, ELM方法的对比实验. 实验结果表明本文方法能很好地应用于图像识别分类任务中,并且具有比SRC, ELM方法高的击中率,同时其虚报率比其他两种方法低得多. 这也从人类记忆模型的角度上解释了视觉图像的识别与分类过程,即先“认识”再“区别”. 尽管本文从人类认知记忆角度上提供了视觉图像分类识别的一个新思路,但还有其他问题需要解释与解决,如如何快速提取与回忆记忆信息,以及如何实现人类记忆中的遗忘过程等.

参考文献

- 1 Djouadi A, Bouktache E. A fast algorithm for the nearest-neighbor classifier. In: *Intelligence IEEE Transactions on Pattern Analysis & Machine*, vol 19. 1997. 277–282
- 2 Denoeux T. A neural network classifier based on Dempster-Shafer theory. *IEEE Trans Syst Man Cybern A*, 2010, 30: 131–150
- 3 Chang C C, Lin C J. LIBSVM: A library for support vector machines. *ACM Trans Intell Syst Technol*, 2011, 2: 1–27
- 4 Wright J, Ma Y, Mairal J, et al. Sparse representation for computer vision and pattern recognition. *Proc IEEE*, 2010, 98: 1031–1044
- 5 Lei Z, Meng Y, Feng X. Sparse representation or collaborative representation: Which helps face recognition? In: *IEEE International Conference on Computer Vision*, Barcelona. 2012. 471–478
- 6 Naseem I, Togneri R, Bennamoun M. Linear regression for face recognition. In: *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol 32. 2010. 2106–2112
- 7 Huang G B. What are extreme learning machines? Filling the gap between Frank Rosenblatt’s dream and John von Neumann’s puzzle. *Cogn Comput*, 2015, 7: 263–278
- 8 Abe D S. Support vector machines for pattern classification. In: *International Joint Conference on Neural Networks*, vol 2. 2010. 1449–1454
- 9 Fung G, Mangasarian O L. Proximal support vector machine classifiers. In: *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM. 2001. 77–86
- 10 Chicco D, Sadowski P, Baldi P. Deep autoencoder neural networks for gene ontology annotation predictions. In: *ACM Conference on Bioinformatics, Computational Biology, and Health Informatics*, ACM. 2014. 533–540
- 11 Larochelle H, Mandel M, Pascanu R, et al. Learning algorithms for the classification restricted Boltzmann machine. *J Machine Learn Res*, 2012, 13: 643–669
- 12 Mohamed A, Dahl G E, Hinton G. Acoustic modeling using deep belief networks. *IEEE Trans Audio Speech Lang Process*, 2012, 20: 14–22
- 13 Gregor K, Danihelka I, Graves A, et al. DRAW: A recurrent neural network for image generation. In: *Proceedings of the 32nd International Conference on International Conference on Machine Learning*. 2015, 37: 1462–1471
- 14 Karpathy A, Toderici G, Shetty S, et al. Large-Scale Video Classification with Convolutional Neural Networks. In: *IEEE Conference on Computer Vision and Pattern Recognition*. 2014. 1725–1732
- 15 Murdock B B. A theory of the storage and retrieval of item and associative information. *Psychological Review*, 1982. 89(6): 609–626
- 16 Raaijmakers J G W, Shiffrin R M. SAM: A Theory of Probabilistic Search of Associative Memory. *Psychol Learn Motiv*, 1980, 14: 207–262
- 17 Howard M W, Kahana M J. A distributed representation of temporal context. *J Math Psychol*, 2001, 46: 269–299
- 18 Shiffrin R M, Steyvers M. A model for recognition memory: REM-retrieving effectively from memory. *Psychon Bull Rev*, 1997, 4: 145–166
- 19 Rogers T T, McClelland J L. *Precis of semantic cognition: A parallel distributed processing approach*. *Behav Brain Sci*, 2010, 31: 689–714
- 20 McClelland J L, McNaughton B L, O’Reilly R C. Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychol Rev*, 1995, 102: 419–457
- 21 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *Computer Sci*, 2014, .
- 22 Li F F, Fergus R, Perona P. Learning generative visual models from few training examples: An incremental Bayesian approach tested on 101 object categories. *Computer vision and Image understanding*, 2007, 106: 59–70
- 23 Griffin G, Holub A, Perona P. Caltech-256 Object Category Dataset. California Institute of Technology Report. 2007

A memory model for image recognition and classification based on convolutional neural network and Bayesian decision

JIANG Ying, WANG YanJiang, LIN Qing & LIU WeiFeng

College of Information and Control Engineering, China University of Petroleum, Qingdao 266580, China

Most popular image classification methods mainly focus on classification ability rather than recognizing new things. However, human lay emphasis on cognition first and then classification, which is closely related to human memory system. Though many memory models have been proposed, they are studied in word list whereas the reports about natural images are still limited. This paper proposes a memory model for image recognition and classification based on convolutional neural network and Bayesian decision. First the image feature is extracted by convolutional neural network and stored in binary form. Then the representation, storage and retrieval processes of visual images are modeled. The test image feature vector is matched in parallel to the studied image vectors, and the likelihood values are calculated. Finally, the odd that the test image belongs to a new class is computed based on all likelihood values. If the odd value is greater than a certain threshold, the test image is regarded as new; otherwise, the Bayesian decision rule for image classification is performed. Experimental results on Caltech-101 and Caltech-256 datasets show that the proposed method can perform well in image recognition and classification tasks. And the hit probability of the method is higher than two typical methods, SRC and ELM, at present while the false alarm rate is far lower than them.

convolutional neural network, Bayesian decision rule, memory modeling, image recognition and classification

doi: 10.1360/N092016-00405