# 基于深度学习混合模型迁移学习的图像分类

石祥滨[1,2,3]，房雪键[3]，张德园[1]，郭忠强[3]
(1. 沈阳航空航天大学计算机学院，沈阳 110136；
2. 沈阳航空航天大学辽宁通用航空重点实验室，沈阳 110136；3. 辽宁大学信息学院，沈阳 110036)

**摘要：**为提高深度模型迁移学习的特征识别力，提出一种基于受限玻尔兹曼机与卷积神经网络混合模型迁移学习的图像分类方法。该方法融合了 2 种模型特征的学习能力，提取图像的结构性高阶统计特征进行主题分类。*该方法在迁移预训练的卷积神经网络模型到小目标集时，使用受限玻尔兹曼机代替卷积神经网络模型中的全连接层，在目标集上重新训练受限玻尔兹曼机层和 Softmax 层，并使用 BP 算法进行参数调整。*加入的受限玻尔兹曼机层不仅全连接所有特征 maps，还从最大对数似然的角度学习目标集特有的统计特征，消除了数据集间内容差异对迁移学习特征识别力的影响。在 Pascal VOC2007 和 Caltech101 数据集上的实验结果表明，该方法具有较高的分类准确率。

**关键词：**图像分类；卷积神经网络；受限玻尔兹曼机；迁移学习；Softmax
中图分类号：TP391.9 文献标识码：A 文章编号：1004-731X (2016) 01-0167-08
DOI:10.16182/j.cnki.joss.2016.01.023

## Image Classification Based on Mixed Deep Learning Model Transfer Learning

*Shi Xiangbin*[1,2,3], *Fang Xuejian*[3], *Zhang Deyuan*[1], *Guo Zhongqiang*[3]
(1. Department of Computer, Shenyang Aerospace University, Shenyang 110136, China;
2.Liaoning General Aviation Key Laboratory, Shenyang Aerospace University, Shenyang 110136, China;
3. College of Information, Liaoning University, Shenyang 110036, China)

**Abstract:** In order to obtain high discrimination image representations in limited amount of datasets, the method based on mixed deep transfer learning model was proposed. *When trained CNNs transferred to the target datasets, fully-connected layers were replaced by RBM layers. The method retrained the RBM layers and Softmax classifier, then fine-tuned the mixed model with backpropagation algorithm.* The RBM layers not only fully connected whole feature maps, but also learned the target datasets' statistical features in the view of the biggest logarithmic likelihood, to eliminate the effects caused by the content differences between datasets. The experimental results show that the method has improved the accuracy of image classification, outperforming other methods on Pascal VOC2007 and Caltech101 datasets.

**Keywords:** image classification; CNN(Convolutional Neural Networks); RBM(Restricted Boltzmann Machines); transfer learning; softmax

## 引言

图像分类是指利用计算机模拟人类对图像的理解和认知，自动把图像划分到不同的语义类别，在信息搜索、安全监控、医疗信息、航空航天等领域有着广泛的应用。图像分类问题面临着如何用计算机语言表示图像、如何选择合适的分类算法以及如何按人类的理解对图像归类等方面的挑战，图像分类问题至今没有达到人们期望的智能、高效、精确的目标。

收稿日期：2015-06-09 修回日期：2015-07-30；
基金项目：国家自然科学基金(61170185)；航空科学基金(2013ZC54011)；辽宁省博士启动基金(20121034)；辽宁省教育厅资助项目(L2014070)；
作者简介：石祥滨(1963-)，男，辽宁，博士，教授，研究方向为虚拟现实、图像处理，网络游戏。



?1994-2016 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

目前, 国内外对图像分类的研究主要是 2 个方向：(1) 分类算法[1]的研究，如决策树、贝叶斯、神经网络，尤其是支持向量机[2]的出现，极大提高了分类的准确率；(2) 图像特征提取，出现了 SIFT 特征、BOW 模型、SPM 模型[3]、FV 模型[4]以及稀疏编码 SC[5]等方法，大大增加了图像的特征识别力。但这些方法获得的都是图像底层特征，其与图像高级语义之间还存在很大的差异[6]，而深度学习的优越性就在于其能提取图像的高级语义特征。

卷积神经网络(Convolutional Neural Networks, CNN)[7-8]通过逐层抽取图像，获得能代表一幅图像高级语义的结构化特征。模型从低层到高层的特征表示越来越抽象，越来越能表现图像具体主题，从而存在的不确定特征就越少, 在分类中的识别力就越高。受限玻尔兹曼机(Restricted Boltzmann Machines，RBM)[9-10]具有强大的无监督[11]特征学习能力, 其从最大化对数似然的角度学习输入数据的复杂规律，重构出图像高识别力的统计特征。

卷积神经网络对图像逐层抽象过程中, 每层通过多个数字滤波器提取输入数据的显著特征, 最终得到的结构性特征与图像的高级语义相吻合。文献[12]阐述了 CNN 卷积层输出的特征 maps 的意义，第五卷积层的每个卷积核激活一种特征, 如第一个卷积核激活图像中“方形”特征，第二个卷积核激活图中“圆形”特征等。之后的全连接层把各种显著特征加权组合，构成图像的完整结构性语义特征。

目前对 CNN 模型在图像识别中的研究，一般是通过调整网络的结构或者改进对各层特征的处理方式来提高模型的特征学习能力。文献[13]提出改进传统激活函数为非饱和非线性的最大化函数，并加入了特征标准化操作, 对同层相邻节点的响应进行局部归一化，提高了特征识别力。文献[14]提出更改传统网络结构为多列网络结构, 并联合多个网络模型做平均化处理。文献[12]为使固定网络结构可以处理任意大小的图像, 在卷积层和全连接层之间加入了空间金字塔池化过程, 把卷积层得到的任意大小特征 maps 处理成固定维数的特征向量。文献[15]提出一种快速、全 GPU 部署的 CNN 计算框架，不仅提高了训练速度，而且模型参数的调整具有很大的灵活性。文献[16]对卷积神经网络产生的特征进行反卷积，重构出对应的输入刺激，通过分析重构模型, 探究图片中的哪一部分刺激网络产生了具体特征，从而有针对性的进行模型调优。

CNN 模型的训练需要估计上百万个参数，因此 CNN 训练时需要使用大量的标记样本，而在小样本集上一般是直接使用预训练好的模型进行特征提取。为了解决源数据集和目标数据集类别不同的情况, 文献[17]提出了一种称为迁移学习的方法，将从大数据集上学习的 CNN 当做目标集的底层和中层特征提取器, 并修改最后一全连接层为自适应特征层。在训练的过程中，只训练自适应特征层。但在迁移学习时会因为两数据集间内容差异而降低特征识别力, 因此本文使用受限玻尔兹曼机层代替全连接层来解决这一问题。

## 1 主要思想

CNN 模型迁移学习时，由于数据集间内容差异，导致直接提取的特征识别力受到影响。为解决这一问题, 本文提出在迁移 CNN 模型到目标集时，改传统卷积神经网络模型的全连接层为受限玻尔兹曼机层，不仅全连接所有特征 maps 组合成整体结构性特征，还能从输入的特征 maps 中学习目标集特有的统计特征, 从而提取出图像的结构性高阶统计特征，提高图像分类准确率。

如图 1 为本文提出的基于 RBM 与 CNN 混合模型迁移学习的结构图。首先在大数据集上预训练 CNN 模型, 得到卷积层 C1-C5 和全连接层 FC6-FC8 的参数；之后迁移该 CNN 模型到小目标集，使用 C1-C5 层参数提取图像的卷积层特征 maps，将每幅图像的所有特征 maps 串联成一个特征 map；然后用 RBM 模型全连接输入的特征 map，依次无监督学习 R6，R7 层参数，并使用 BP 算法[18]有监督微调 Softmax 回归及 R6，R7 层参数，得到训练好的混合模型分类器；最后对于目标集中的待分类图像，使用该混合模型计算出图像类别。



?1994-2016 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

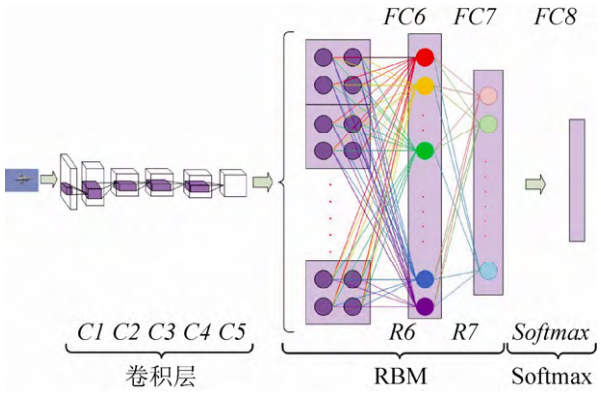


图 1　基于 RBM 与 CNN 混合模型迁移学习的结构图

## 2　混合模型迁移学习

迁移学习是指迁移预训练的 CNN 模型到其它数据集，并重新学习目标集特征。本文提出的混合模型迁移学习，在预训练模型迁移到目标集时，融合受限玻尔兹曼机模型，加入目标集特有高阶统计特征，消除数据集间内容差异对目标集特征识别力的影响。

### 2.1　预训练卷积神经网络模型

迁移学习首先在大数据集上预训练 CNN 模型。CNN 模型一般由卷积层和全连接层组成，最后一全连接层为 Softmax 分类器。

CNN 的预训练分为正向传播和反向调参 2 个过程[3]。假设该网络处理 $K$ 类 $m$ 个训练样本，单个输入样本为$(x^{(i)}, y^{(i)})$，其中 $x^{(i)}$为 $n$ 维输入向量，$y^{(i)}$为该样本标记好的真正所属类别。用 $l$ 表示当前层，$l$ 层输入特征向量为 $x^{l-1}$，输出特征向量为 $x^l$，该层某个滑动卷积滤波器权值 $w^l$ 和偏置 $b^l$。则前向传播在每层对于输入特征如下计算：

$$x^l = f(u^l), u^l = w^l x^{l-1} + b^l \tag{1}$$

其中 $f(\cdot)$一般为 sigmoid 函数。对于含有 $m$ 个样本的样本集$\{(x^{(1)},y^{(1)}),\ldots,(x^{(m)},y^{(m)})\}$，CNN 模型的整体代价函数为：

$$J(w,b) = \frac{1}{m}\sum_{i=1}^{m}\left(\frac{1}{2}\| h_{w,b}(x^{(i)}) - y^{(i)} \|^2\right) + \frac{\lambda}{2}\sum_{l=1}^{n_l-1}\sum_{i=1}^{s_l}\sum_{j=1}^{s_{l+1}}(w_{ji}^{(l)})^2 \tag{2}$$

其中 $\lambda$ 为权重衰减参数，$n_l$ 为网络总层数，$s_l$ 为网络第 $l$ 层节点数。使用批量梯度下降法[8]调整参数，使得整体代价函数最小化。如下公式对每层参数 $w_{ij}^{(l)}$和 $b_i^{(l)}$进行更新：

$$w_{ij}^{(l)} = w_{ij}^{(l)} - \alpha\frac{\partial}{\partial w_{ij}^{(l)}}J(w,b)$$
$$b_i^{(l)} = b_i^{(l)} - \alpha\frac{\partial}{\partial b_i^{(l)}}J(w,b) \tag{3}$$

其中，$\alpha$ 为学习速率。上式参数关于代价函数偏导的求解需要计算每层的残差，具体如文献[8]。当整个网络结构的代价误差最小时，就得到预训练好的卷积神经网络模型。

卷积神经网络模型最后一全连接层 FC8 实质为一个 Softmax 回归分类器，其类标签为多值向量 $y^{(i)}\in\{1,2,...,K\}$。对于给定的测试输入 $x$，使用训练好的 Softmax 分类器估计出 $x$ 属于每一种分类结果 $k$ 的概率 $p(y=k|x)$。

本文使用 Krizhevsky 等在文献[13]中提出的 CNN 模型，该模型由 5 个连续的卷积层 C1-C5 和 3 个全连接层 FC6-FC8 组成，用于处理 224×224 大小的 RGB 图像。卷积层 C1 使用 96 个 11×11×3 的卷积核滑动处理 224×224×3 的输入图像，滑动步长为 4 像素。卷积层 C2 使用 256 个 5×5×96 的卷积核处理 C1 输出的 96 个特征 maps。卷积层 C3-C5 依次使用 384 个 3×3×256，384 个 3×3×384，256 个 3×3×384 的卷积核，全连接层 FC6-FC8 神经元数依次为 4 096，4 096，1 000。所有层激活函数使用非饱和非线性函数 $f(x)=\max(0, x)$，C1 和 C2 层都对得到的特征 maps 进行最大池化处理和标准化处理，C5 层也对得到的特征 maps 进行最大池化处理。

### 2.2　模型迁移学习

将预训练的 CNN 模型迁移到小目标集上，重新训练 RBM 与 CNN 混合模型，并使用 BP 算法微调 RBM 和 Softmax 分类器层参数。如图 1 在混合模型重训练时，去掉预训练 CNN 模型中的 FC6-FC8 全连接层，改为受限玻尔兹曼机 R6-R7



?1994-2016 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

层和新 Softmax 层，RBM 层提取的特征输出到 Softmax 分类器。为能组合卷积层激活的所有特征，使 RBM 层起到全连接的作用，本文将每幅训练图像在 C5 卷积层输出的 256 个 6×6 特征 maps 串联成一个大小为(256×6)×6 的特征 map。RBM 模型 R6 可视层有 1 536×6 个节点, 隐层有 5 000 个节点，R7 层有 10 000 个隐节点。假设训练集有 K 类，则 Softmax 层输入 10000 维特征向量, 输出 $K$ 维向量。

### 2.2.1 受限玻尔兹曼机模型重训练

本文模型迁移时加入受限玻尔兹曼机层, 一是起到全连接的作用，二是为了从输入特征 maps 中学习目标集特有的统计特征。RBM 是一种具有两层结构、对称连接且无自反馈的网络模型，层间全连接，层内无连接。RBM 重构的可视层 $v^1$ 如果与原可视层 $v$ 近似一样，则得到的隐藏层 $h$ 就是从可视层的复杂数据规律中学习到的高识别力统计特征。

假设所有可见单元和隐单元均为二值单元, 即 $v_i\in\{0,1\}$, $h_j\in\{0,1\}$，1 表示该单元被激活状态，0 表示未激活状态。对于一个有 $n$ 个可见单元 $m$ 个隐单元的受限玻尔兹曼机[9]，定义该系统所具备的能量为：

$$E(v,h\mid\theta)=-\sum_{i=1}^{n}a_iv_i-\sum_{j=1}^{m}b_jh_j-\sum_{i=1}^{n}\sum_{j=1}^{m}v_iW_{ij}h_j \tag{4}$$

其中，$\theta=\{W_{ij},a_i,b_j\}$各参数分别表示可见单元 $i$ 与隐单元 $j$ 之间的连接权重、可见单元的偏置及隐单元的偏置。对于参数确定的 RBM 模型，可视层 $v$ 和隐层 $h$ 被激活的概率分别为：

$$P(h_j=1\mid v,\theta)=\sigma(b_j+\sum_i v_iW_{ij}) \tag{5}$$

$$P(v_i=1\mid h,\theta)=\sigma(a_i+\sum_j W_{ij}h_j) \tag{6}$$

其中 $\sigma(\cdot)$为 sigmoid 函数。RBM 模型训练的过程实质是最大对数似然下的参数估计, 也是模型能量最小化问题，即：

$$\begin{aligned}\theta^*&=\arg\max L(\theta)=\arg\max_{\theta}\sum_{t=1}^{T}\log P(v^{(t)}\mid\theta)\\ P(v\mid\theta)&=\frac{1}{\sum_{v,h}e^{-E(v,h|\theta)}}\sum_h e^{-E(v,h|\theta)}\end{aligned} \tag{7}$$

其中，$P(v|\theta)$为似然函数，$v^{(t)}$表示第 $t$ 个训练样本。在文献[19]中提出的对比散度快速学习法，通过随机梯度上升法求解(7)式，各参数更新如下：

$$\begin{aligned}&-W\leftarrow W+\varepsilon(P(h=1\mid v)v^T-P(h^1=1\mid v^1)v^{1T})\\&-a\leftarrow a+\varepsilon(v-v^1)\\&-b\leftarrow b+\varepsilon(P(h=1\mid v)-P(h^1=1\mid v^1))\end{aligned} \tag{8}$$

其中，$\varepsilon$ 是学习率。当重构的可视层 $v^l$ 与原可视层 $v$ 误差 $E=\|v-v^l\|$小于某阈值时，终止训练。

严格意义上说，RBM 并不是真正的深度学习模型，它只是很多深度学习模型的基本功能模块。为获得输入数据的高阶统计特征，本文使用两层 RBM 来逐层无监督特征学习。

### 2.2.2 BP 算法调整参数

对于混合模型中 RBM 层输出的高阶统计特征, 需要有监督训练一个 Softmax 分类器对本数据集图像进行分类。为提高混合模型的分类准确率，在训练 Softmax 分类器的同时, 继续将训练残差反向往前传, 使用 BP 算法有监督调整 *R6-R7-Softmax* 层参数。为保持 RBM 提取统计特征的性质，参数调整时只有反向传播是 BP 训练，而正向传播依旧按 RBM 的似然函数形式计算。

BP 算法反向调参最主要的是计算每层每个输出节点的残差 $\delta_i^l$, 残差表示该节点对最终输出值的偏差产生了多大的影响。Softmax 分类器每个输出单元 $i$ 的残差定义为：

$$\delta_i^{n_l}=-(y_i-x_i^{(n_l)})\cdot f'(u_i^{n_l}) \tag{9}$$

后面层残差反向往前传播，则 R6 和 R7 层每个隐单元 $i$ 的残差定义为：

$$\delta_i^l=(\sum_{j=1}^{S_{l+1}}w_{ji}\delta_j^{l+1})\cdot p'(h_i^l=1\mid v,\theta) \tag{10}$$

结合公式(3)，R6，R7 及 Softmax 层如下调整参数：



?1994-2016 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

$$\frac{\partial}{\partial w_{ij}^{(l)}}J(w,b;x,y)=\delta_i^{(l)}v_j^{(l-1)}$$
$$\frac{\partial}{\partial b_i^{(l)}}J(w,b;x,y)=\delta_i^{(l)} \tag{11}$$

$$\frac{\partial}{\partial w_{ij}^{(l)}}J(w,b)=\left[\frac{1}{m}\sum_{i=1}^{m}\frac{\partial}{\partial w_{ij}^{(l)}}J(w,b;x^{(i)},y^{(i)})\right]+\lambda w_{ij}^{(l)}$$
$$\frac{\partial}{\partial b_i^{(l)}}J(w,b)=\left[\frac{1}{m}\sum_{i=1}^{m}\frac{\partial}{\partial b_i^{(l)}}J(w,b;x^{(i)},y^{(i)})\right] \tag{12}$$

上述反向调参及正向传播过程中, 可视层的偏置 $a_i$ 并未参与运算, 因此不考虑对其更新。每次 BP 反向传播到 R6 层即停止, 之后正向传播时, R6 和 R7 层按公式(5)计算, Softmax 层按公式(1)计算, 迭代训练直至混合模型代价误差最小。

对于本文混合深度学习模型的训练, 分为三个阶段：大数据集上的 CNN 预训练、目标数据集上的 RBM 重训练以及 R6-R7-Softmax 层参数有监督调整, 具体训练如算法 1：

**算法 1**：混合模型训练

**输入**：大数据集 $I_1$, 小目标集 $I_2$, 误差阈值 threshold1, threshold2

**输出**：混合深度学习模型 C-R-Softmax

**训练**：

1：初始化 $I_1$, $I_2$ 为 224×224 大小；

2：在 $I_1$ 上预训练 CNN 模型 C-FC(公式(1)~(3))；

3：For $i$=1,2,...,$m$ (转移 C-FC 到 $I_2$ 的 $m$ 幅图像)

4： 用 C1-C5 层参数提取图像 $i$ 的 256 个 6×6 特征 maps (公式(1))；

5： 图像 $i$ 的所有 maps 串联为一个(256×6)×6 维特征 map $F_i$；

6：EndFor

7：把上层输出的 m 个特征平均划分为 n 批, 并初始化本层参数；

8：Repeat

9： 样本总误差 Error=0；

10： For $j$=1,2,…,$n$

11： 按批数据无监督训练 R6 层参数(公式(4)~(8))；

12： 计算重构误差 $E_j$, Error=Error+$E_j$；

13：调整后的参数作为下批次数据的初始参数；

14： EndFor

15：最后一批次数据调整后的参数作为下次迭代的初始参数；

16：Until Error/$m$ ≤ threshold1

17：计算样本 R6 层的特征输出 FR6(公式(5))；

18：输入 FR6, 重复步骤 7-16 训练 R7 层, 并计算输出特征 FR7；

19：使用 FR7 初步训练 10 次 Softmax 分类器(公式(1)~(3))；

20：计算此时 Softmax 的代价函数 $J$(公式(2))；

21：Repeat

22：BP 反向调整 Softmax, R7 及 R6 层参数(公式(2), (3)及(9)~(12))；

23：正向计算 R6, R7 层输出(公式(5))、Softmax 层输出(公式(1))及代价函数 $J$(公式(2))；

24：Until $J$ ≤ threshold2

训练完混合深度学习模型 C-R-Softmax 后, 对目标集图像进行分类。待分类图像输入混合模型, 根据公式(1)和(5)逐层特征提取, 最终层输出一 $K$ 维向量。

为弥补图像分类时裁剪图像造成的内容丢失, 本文首先对每幅图像从 10 个不同角度分别裁剪 224×224 大小的图像, 然后使用训练好的混合模型分类器处理每幅图像的每个视角块 $I_j$, 最终使用文献[17]的方法计算图像 $I$ 属于类别 $k$ 的概率：

$$P(k)=\frac{1}{10}\sum_{j=1}^{10}p(k\mid I_j) \tag{13}$$

其中, $p(k|I_j)$表示图像 $I$ 的第 $j$ 视角块 $I_j$ 属于类别 $k$ 的概率, 图像 $I$ 最终属于概率值 $P(k)$最大的类别。



?1994-2016 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

## 3 图像分类实验与分析

为了验证本文提出的混合模型迁移学习的有效性，将在 Pascal VOC2007 数据集和 Caltech101 数据集上分别测试其分类准确率, 并与其他同类方法进行比较。Pascal VOC2007 数据集有 20 种类别, 共 9 963 张图像，其中 5 011 张作为训练集，剩余为测试集。Caltech101 数据集有 102 类，共 9 144 张图像, 每类随机抽取 50 和 30 张图像分别作为训练集和测试集。本文采用平均准确率(mAP)作为图像分类准确率的评价标准。

卷积神经网络模型在大数据集上的训练是比较耗时耗资源的, 为减少训练开销，本文直接使用 MatConvNet 函数库中提供的在 ILSVRC2012 大数据集预训练好的 CNN 模型 imagenet-caffe-alex。之后使用该预训练模型 C1-C5 层参数提取小目标集的特征 maps，作为后面迁移学习的输入。本文所有实验使用机器的配置均为 16 GB 内存，不使用 GPU。混合模型的训练如算法 1，其中每层 RBM 重构误差 threshold1 设为 0.1，BP 调参时整体代价误差 threshold2 设为 0.001。

表 1 为基于 Pascal VOC2007 数据集的各分类方法平均识别率比较，其中 CNN 方法是指直接使用 imagenet-caffe-alex 模型来提取 Pascal VOC2007 数据集的 FC7 层特征, 之后使用 Softmax 分类器进行分类。从表 1 可看出，与模型迁移学习方法[17]相比，本文算法的分类准确率提高了 1.7%。相对于直接迁移 imagenet-caffe-alex 模型的 CNN 方法, 本文混合模型使分类平均准确率提高了 4.2%。

表 1 基于 Pascal VOC2007 数据集的各方法平均准确率(%)比较

| 模型 | plane | bike | bird | boat | btl | bus | car | cat | chair | cow | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| NUS-PSL[20] | 82.5 | 79.6 | 64.8 | 73.4 | 54.2 | 75.0 | 77.5 | 79.2 | 46.2 | 62.7 | |
| CNN | 84.3 | 80.2 | 79.8 | 83.1 | 51.2 | 72.9 | 81.1 | 88.6 | 56.7 | 66.5 | |
| PRE-1000C[17] | 88.5 | 81.5 | 87.9 | 82.0 | 47.5 | 75.5 | 90.1 | 87.2 | 61.6 | 75.7 | |
| 本文方法 | 87.2 | 83.3 | 83.9 | 85.0 | 57.1 | 79.6 | 86.4 | 86.8 | 70.5 | 71.4 | |
| 模型 | table | dog | horse | moto | pers | plant | sheep | sofa | train | tv | mAP |
| NUS-PSL[20] | 41.4 | 74.6 | 85.0 | 76.8 | 91.1 | 53.9 | 61.0 | 67.5 | 83.6 | 70.6 | 70.5 |
| CNN | 69.4 | 77.2 | 86.9 | 82.1 | 93.9 | 66.3 | 76.2 | 60.3 | 87.1 | 74.2 | 75.2 |
| PRE-1000C[17] | 67.3 | 85.5 | 83.5 | 80.0 | 95.6 | 60.8 | 76.8 | 58.0 | 90.4 | 77.9 | 77.7 |
| 本文方法 | 65.2 | 88.4 | 87.3 | 83.8 | 93.5 | 66.1 | 79.9 | 68.3 | 85.9 | 78.4 | 79.4 |

本文模型和 CNN 方法、文献[17]中方法都转移预训练模型到其他目标集，但 CNN 方法是直接在目标集上使用预训练模型进行特征提取，[17]方法只添加了有监督训练最后的自适应特征层, 而本文使用 RBM 层代替全连接层，加入了目标集特有的统计特征，之后还对无监督学习的模型使用 BP 算法调整，因此图像分类正确率有了提高。

为验证加入 RBM 层能使提取的特征识别力提高，本实验在 Pascal VOC2007 数据集上，分别提取 CNN 方法的 FC6，FC7 层输出的特征向量以及本文混合模型在 R6，R7 层输出的特征向量，并分别训练 Softmax 分类器, 然后在测试集上进行分类。结果如表 2，可知 RBM 层提取的特征向量比全连接层提取的特征向量更具有识别力, 因此迁移学习时使用 RBM 模型整合卷积层输出的特征 maps 比直接使用全连接层更好。

表 2 RBM 和全连接层提取特征的平均准确率(%)比较

| 模型层 | FC6 | R6 | FC7 | R7 |
|---|---|---|---|---|
| mAP | 73.7 | 75.5 | 75.2 | 79.4 |

表 3 为基于 Caltech101 数据集的各方法平均识别率比较, 与已有的方法相比本算法的准确度提高了 1.5%，与直接转移模型 imagenet-caffe-alex 到目标集提取特征分类相比, 本文混合模型使分类准确率提高了 3.2%。另外，由表 3 可知，与传统的空



?1994-2016 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

间金字塔、Fisher 向量等方法相比，深度学习方法使图像分类识别率有了极大的提高。

表 3　基于 Caltech101 数据集的各方法平均准确率(%)比较

| 模型 | mAP |
|---|---|
| SPM[3] | 64.6 |
| ScSPM[21] | 73.2 |
| FK[4] | 77.78 |
| CNN | 84.8 |
| Zeiler[16] | 86.5 |
| 本文方法 | 88.0 |

本文在 Caltech101 数据集上还验证了本文混合模型比单一 RBM 模型的有效性，如图 2。单一 RBM 模型是在 *R*6 层直接无监督学习原始图像，之后 R7 层再无监督学习上层输出，最后 BP 算法调整 R6-R7-Softmax，每层 RBM 重构误差 threshold1 设为 0.1。随着 BP 算法调整次数的增加，两种方法都会使分类识别率增加，但本文混合模型比单一 RBM 模型的分类准确率平均高 10.17%，证明了使用 RBM 模型从卷积层特征 maps 中无监督学习的结构性高阶统计特征，优于直接使用 RBM 模型从原图像中学习的特征。

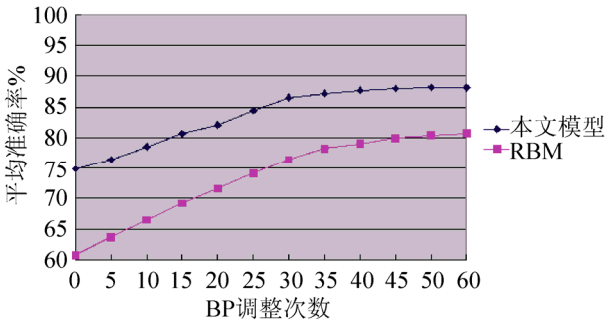


图 2　本文模型和 RBM 模型平均识别率比较

## 4　结论

本文提出基于 RBM 与 CNN 混合模型迁移学习的图像分类方法，在迁移大数据集上预训练好的卷积神经网络模型到小目标集时，使用受限玻尔兹曼机代替传统卷积神经网络模型中的全连接层进行特征提取，并对受限玻尔兹曼机模型和 Softmax 分类器进行有监督调整。本文方法结合了 2 种模型的优点，不仅能全连接卷积层特征 maps 获得结构性语义特征，还能提取目标数据集特有的高阶统计特征。本文混合模型消除了在模型转移时，因数据集间内容差异而导致的目标集特征识别力降低的问题，提高了图像分类准确率。

## 参考文献：

[1] Fernández-Delgado M, Cernadas E, Barro S, et al. Do we Need Hundreds of Classifiers to Solve Real World Classification Problems? [J]. Journal of Machine Learning Research (S1532-4435), 2014, 15(1): 3133-3181.

[2] Joachims T. Making Large-scale Support Vector Machine Learning Practical [C]// Advances in kernel methods. USA: MIT Press, 1999: 169-184.

[3] Harada T, Ushiku Y, Yamashita Y, et al. Discriminative Spatial Pyramid [C]// Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. USA: IEEE, 2011: 1617-1624.

[4] Sánchez J, Perronnin F, Mensink T, et al. Image Classification with the Fisher Vector: Theory and Practice [J]. International Journal of Computer Vision (S0920-5691), 2013, 105(3): 222-245.

[5] Zhang C, Liu J, Tian Q, et al. Image Classification by Non-negative Sparse Coding, Low-rank and Sparse Decomposition [C]// Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. USA: IEEE, 2011: 1673-1680.

[6] Boureau Y L, Bach F, LeCun Y, et al. Learning mid-level features for recognition[C]//Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. USA:IEEE, 2010: 2559-2566.

[7] Lecun Y, Kavukcuoglu K, Farabet C. Convolutional Networks and Applications in Vision [C]// Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium on. USA: IEEE, 2010: 253-256.

[8] Bouvrie J. Notes on Convolutional Neural Networks [R]// MIT-CBCL Technical Reports. Germany: Springer International, 2006: 38-44.

[9] Fischer A, Igel C. Training Restricted Boltzmann Machines: An Introduction [J]. Pattern Recognition (S0031-3203), 2014, 47(1): 25-39.

[10] Tang Y, Salakhutdinov R, Hinton G. Robust Boltzmann Machines for Recognition and Denoising [C]// Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. USA: IEEE, 2012: 2264-2271.

（下转第 182 页）



?1994-2016 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

$$f_2(x)=\frac{\int_{\{\phi<0\}}K_\sigma(x-y)I(y)dy}{\int_{\{\phi<0\}}K_\sigma(x-y)dy}=$$

$$(a\int_{\{\phi<0\}\cap\omega}K_\sigma(x-y)I(y)dy+$$

$$b\int_{\{\phi<0\}\setminus(\{\phi<0\}\cap\omega)}K_\sigma(x-y)I(y)dy)\times$$

$$(\int_{\{\phi<0\}}K_\sigma(x-y)dy)^{-1}=$$

$$\frac{a(p-q)+b[(P-Q)-(p-q)]}{P-Q}=$$

$$\frac{(a-b)(p-q)+b(P-Q)}{P-Q}$$

在 $\Omega\setminus\omega$ 内

$$F^L(\phi)=b-f_1(x)H_\varepsilon(\phi)-f_2(x)(1-H_\varepsilon(\phi))=$$

$$\frac{b-a}{Q(P-Q)}\times[q(P-Q)H_\varepsilon(\phi)+$$

$$Q(p-q)(1-H_\varepsilon(\phi))]$$

有如下事实：

$$q(P-Q)H_\varepsilon(\phi)+Q(p-q)(1-H_\varepsilon(\phi))=$$

$$H_\varepsilon(\phi)\int_{\{\phi<0\}}K_\sigma(x-y)dy\int_{\{\phi>0\}\cap\omega}K_\sigma(x-y)dy+$$

$$(1-H_\varepsilon(\phi))\int_{\{\phi>0\}}K_\sigma(x-y)dy\int_{\{\phi<0\}\cap\omega}K_\sigma(x-y)dy>$$

$$H_\varepsilon(\phi)\int_{\{\phi<0\}}Sdy\bullet\left(\int_{\{\phi>0\}\cap\omega}Sdy\right)+$$

$$(1-H_\varepsilon(\phi))\int_{\{\phi>0\}}Sdy\left(\int_{\{\phi<0\}\cap\omega}Sdy\right)=$$

$$S[n(M-N)H_\varepsilon(\phi)+N(m-n)\times$$

$$(1-H_\varepsilon(\phi))>S$$

其中，$S=\min\{K_\sigma(x-y),y\in\Omega\}$。

故有：$\text{sign}(F^L(\phi))=-\text{sign}(a-b),\quad \text{in } \Omega\setminus\omega$。

同理可证：$\text{sign}(F^L(\phi))=+\text{sign}(a-b),\quad \text{in } \omega$。

所以: $\text{sign}(F^L(\phi))=\begin{cases}+\text{sign}(a-b), & \text{in}\,\omega\\ -\text{sign}(a-b), & \text{in}\,\Omega\setminus\omega\end{cases}$

性质 2 证毕.

(上接第 173 页)

[11] Le Q V, Ranzato M, Monga R, *et al*. Building High-level Features using Large Scale Unsupervised Learning [C]// Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on. USA: IEEE, 2011: 8595 - 8598.

[12] He K, Zhang X, Ren S, *et al*. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition [M]// Computer Vision – ECCV 2014. Germany: Springer International, 2014: 346-361.

[13] Krizhevsky A, Sutskever I, Hinton G E. Imagenet Classification with Deep Convolutional Neural Networks [C]// Advances in neural information processing systems. USA: The MIT Press, 2012: 1097-1105.

[14] Ciresan D, Meier U, Schmidhuber J. Multi-column deep neural networks for image classification[C]//Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. USA:IEEE, 2012: 3642-3649.

[15] Ciresan D C, Meier U, Masci J, *et al*. Flexible, high performance convolutional neural networks for image classification [C]// IJCAI Proceedings-International Joint Conference on Artificial Intelligence. USA: Morgan Kaufmann , 2011, 22(1): 1237.

[16] Zeiler M D, Fergus R. Visualizing and understanding convolutional networks[M]//Computer Vision – ECCV 2014. Germany: Springer International Publishing, 2014: 818-833.

[17] Oquab M, Bottou L, Laptev I, *et al*. Learning and Transferring Mid-level Image Representations Using Convolutional Neural Networks [C]// 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). USA: IEEE Computer Society, 2014: 1717-1724.

[18] Rumelhart D E, Hinton G E, Williams R J. Learning representations by back-propagating errors [J]. Nature (S0028-0836), 1986, 323(6088): 533-536.

[19] Hinton G E. Training Products of Experts by Minimizing Contrastive Divergence [J]. Neural Computation (S0899-7667), 2002, 14(8): 1771-1800.

[20] Song Z, Chen Q, Huang Z, *et al*. Contextualizing Object Detection and Classification[C]// Proceedings / CVPR, IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society Conference on Computer Vision and Pattern Recognition. USA: 2011: 1585-1592.

[21] Yang J, Yu K, Gong Y, et al. Linear spatial pyramid matching using sparse coding for image classification[C]//Computer Vision and Pattern Recognition(CVPR), 2009 IEEE Conference on. USA:IEEE, 2009: 1794-1801.



?1994-2016 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net