基于多尺度显著区域特征学习的场景识别

李彦冬，雷 航，郝宗波，唐雪飞

(电子科技大学信息与软件工程学院 成都 610054)

【摘要】场景识别是图像高层语义信息理解的重点和难点领域。如何寻找场景中有效信息的位置是场景识别领域中非常困难的问题。该文提出了一种基于多尺度显著区域特征学习的场景识别方法。首先，提取一个场景中在多尺度下的显著区域；然后，通过卷积神经网络的迁移学习，利用学习到的特征在多尺度的显著区域内对场景进行识别。基于两个公共场景识别数据库上的实验证明了该方法的有效性和良好的泛化能力。实验结果表明，该方法相对于传统的场景识别方法能取得更好的场景识别准确度。

关 键 词 深度学习； 特征学习； 场景分析； 场景识别； 迁移学习

中图分类号 TP391.4 文献标志码 A doi:10.3969/j.issn.1001-0548.2017.03.020

Scene Recognition Based on Feature Learning from Multi-Scale Salient Regions

LI Yan-dong, LEI Hang, HAO Zong-bo, and TANG Xue-fei

(School of Information and Software Engineering, University of Electronic Science and Technology of China Chengdu 610054)

Abstract Scene recognition is an important and challenging topic in the research filed of high level image understanding. Traditional researches of scene recognition focused on handcrafted features, which result in limited discriminative and generalization ability. In addition, finding regions in a scene with rich information is always very challenging. This paper presents an effective method for scene recognition based on learned features from multi-scale salient regions. The method first finds multi-scale salient regions in a scene and then extracts the features from the regions via transfer learning using convolutional neural networks (ConvNets). Experiments on two popular scene recognition datasets show that our proposed method is effective and has good generalization ability for scene recognition, compared with the benchmarks on both of the datasets.

Key words deep learning; feature learning; scene analysis; scene recognition; transfer learning

场景识别的目标是让计算机能够自动提取出图像的高层语义信息，从而对图像所属的场景进行识别。场景识别是最终实现计算机能在高层语义层面“理解”一幅图像的关键技术，是计算机视觉领域中的一个重要而困难的研究课题[1]。

空间金字塔匹配[2](spatial pyramid matching, SPM)是一种典型的传统场景识别方法。SPM将一幅场景图像按照空间金字塔结构划分成固定的栅格区域，以栅格区域为单位提取特征，然后将这些区域特征组合起来构成整个场景的特征。SPM方法主要存在两个缺陷：1) 空间金字塔的区域定义不够灵活，影响了算法的泛化能力；2) SPM使用了传统的人工设计特征，如GIST[3]、SIFT[4]等，这些特征的判别性能和泛化性能都有一定的局限性。

针对场景识别中的区域选择问题，文献[5]提出了一种方向金字塔匹配(orientational pyramid matching, OPM)的方法。OPM利用场景中物体的3D方向特征构建金字塔区域，弥补了SPM仅仅运用场景空间信息的局限。文献[6]运用在目标检测领域取得了良好效果的DPM[7](deformable part-based model)，通过定位场景中的物体对象来寻找场景中含有丰富信息的区域。文献[8]利用了无监督的聚类算法来评价场景中不同区域对于场景类别判断的贡献。

针对场景识别中的特征提取问题，文献[9]对一些传统的特征(GIST、SIFT、HOG、LBP等)在场景识别中的应用进行了分析，取得的效果并不理想。近年来，深度学习[10]的兴起使得“特征学习”逐渐取代了传统的手工设计特征，成为计算机视觉领域

收稿日期：2015－12－28；修回日期：2016 －05－24
基金项目：广东省产学研项目(M17010601CXY2011057)；国家科技支撑计划(2012BAH44F02).
作者简介：李彦冬(1984－)，男，博士生，主要从事机器学习及计算机视觉方面的研究.

的一个新的研究热点。研究表明，学习特征的判别性能远远超过了传统的人工设计特征[11-12]。并且，在特定领域学习到的特征可以通过迁移学习应用到更为广泛的领域中[13-14]。迁移学习的定义是[15]：运用已存有的知识对不同但相关领域问题进行求解的一种机器学习方法。迁移学习的思想使得特征学习得以实现跨领域的应用。

受到近期的相关研究成果启发，针对传统场景识别方法中存在的缺陷，本文提出一种基于多尺度显著区域特征学习的场景识别方法。相比于传统的场景识别策略，该方法利用了相比于人工设计特征具有更好判别性能和泛化性能的深度学习特征。另外，本文提出了一种寻找多尺度显著区域的方法。实验表明，基于多尺度显著区域的特征提取相比于单一尺度的特征提取更有助于提高场景识别的准确度。

1　显著区域提取

显著区域是指一个场景中含有丰富的语义信息，并且能够在一定程度上代表场景特征的区域范围。通过场景中的显著区域，能够提取出更加具有判别性的场景特征，从而提高场景识别准确度。

1.1　场景中的区域划分

针对场景中区域的划分，有一些非常具有代表性的分割方法[16-18]。这些分割方法的基本思想是利用图像的低层特征信息(如：色彩、纹理等)，针对图像的像素点进行分割。这些传统方法虽然能够有效地将一幅图像划分为不同的区域，但是这些区域对于场景的重要性程度，传统的图像分割方法并没有给出评价，因此对于场景识别并不十分适用。

本文的显著区域提取方法在传统的区域分割方法上，进一步针对场景中能够提供具有判别性的场景信息区域进行提取，以适应场景识别的应用需求。与传统的基于低层特征的区域划分方法不同，本文针对场景识别应用的需求，更加关注场景中物体的分布，如一个活动室里面的人、台球桌以及吊灯等，如图1所示。对于一个场景中物体分布更为密集的区域，本文认为这个区域对于场景的特点能够具有更好的代表性。

1.2　场景中的显著区域提取

文献[19]提出了一种基于图像低层特征来提取一个场景中潜在物体框集合$\boldsymbol{L}$的方法。目标检测实验表明，基于低层图像特征而产生的潜在物体框集合$\boldsymbol{L}$具有反映一个场景中目标物体潜在分布的能力。因此，将这些潜在目标物体的分布，作为本文场景显著区域提取的一个基本因素。本文提取场景显著区域的方法如图1所示。

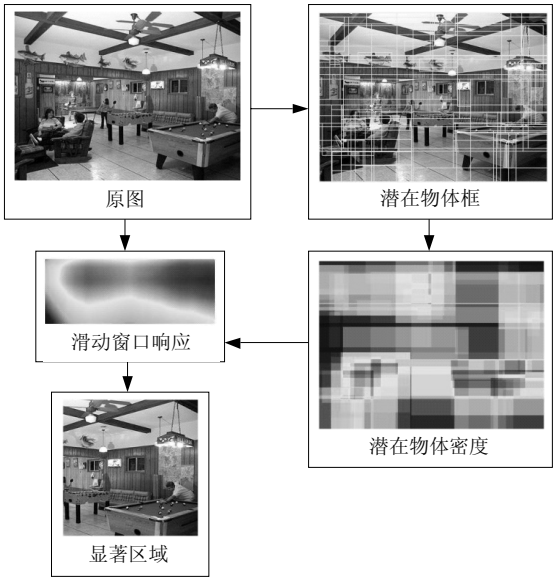

图1　显著区域提取方法

对于一幅场景图像$\boldsymbol{X}$，根据潜在物体框$\boldsymbol{L}$的分布，计算出场景中每个位置的潜在物体密度为：

$$\boldsymbol{M}(i,j)=\frac{\sum_{t=1}^{\mathrm{num}(\boldsymbol{L})} g(\boldsymbol{X}(i,j),\boldsymbol{L}(t))}{\mathrm{num}(\boldsymbol{L})} \tag{1}$$

式中，i和j分别代表场景$\boldsymbol{X}$中每个像素点的横坐标和纵坐标的索引；t是关于场景中潜在物体框集合$\boldsymbol{L}$的索引。$g(\boldsymbol{X}(i,j),\boldsymbol{L}(t))$的定义如下：

$$g(\boldsymbol{X}(i,j),\boldsymbol{L}(t))=\begin{cases}1 & \boldsymbol{X}(i,j)\in\boldsymbol{L}(t)\\ 0 & \boldsymbol{X}(i,j)\notin\boldsymbol{L}(t)\end{cases} \tag{2}$$

针对场景中的潜在物体密度，本文利用滑动窗口$\boldsymbol{B}$计算目标场景中在窗口区域内的物体密度。最终物体密度最高的区域被提取出来作为显著区域：

$$a_{\max}=\underset{a\in[1,\mathrm{num}(\boldsymbol{B})]}{\arg\max}\,\Psi(\boldsymbol{M},\boldsymbol{B}(a)) \tag{3}$$

式中，a是针对滑动窗口的索引；$\Psi(\boldsymbol{M},\boldsymbol{B}(a))$函数用于计算滑动窗口内的物体密度总和。潜在物体密度最高的滑动窗口位置被选为最终的显著区域位置$\boldsymbol{B}(a_{\max})$。

$\boldsymbol{B}(a_{\max})$反映了在以滑动窗口大小为尺度的条件下，一个场景中包含物体信息最为丰富的区域。通过对划分的显著区域进行特征提取，以达到提高场景识别准确度的目的。

2　特征学习

特征设计曾经是一个研究热点，最初的计算机视

觉技术都是基于一些特别设计的特征或者是一些特征的组合。但是，随着研究的深入，传统人工设计特征的缺陷逐渐显现出来，主要有以下两点：

1) 传统设计特征对于应用的针对性比较强，往往针对不同的应用需要设计不同的特征提取方式才能取得理想的结果，缺乏泛化能力；

2) 传统设计特征的判别能力较弱，单一的特征很难取得良好的实验效果。因此，研究中采用多种特征结合的方式以获取较为理想的结果。但是，多种特征结合的方式也随之带来如何选择特征，选择多少特征以及特征维度过高等一系列问题。

近年来兴起的深度学习技术在一定程度上解决了传统人工设计特征的缺陷，逐渐取代了传统设计特征成为了当前的主流特征提取方法。另外，特征的迁移学习进一步拓展了特征学习的应用领域。

2.1 卷积神经网络

卷积神经网络[20]是深度学习领域中的一个重要分支，特别是在图像识别领域，卷积神经网络已经成为一个重要的研究方向。

典型的卷积神经网络包含卷积层、下采样层和全连接层3种基本结构。卷积层的卷积核在输入图像(特征图)上滑动，通过权值共享提取一幅图像(特征图)上各个区域的特征信息：

$$\boldsymbol{H}_m = f(\boldsymbol{W}_m \boldsymbol{H}_{m-1} + \boldsymbol{b}_m) \tag{4}$$

式中，$\boldsymbol{H}_m$表示第m层的特征图($\boldsymbol{H}_0$为输入图像)；$\boldsymbol{W}$和$\boldsymbol{b}$是可训练的参数；$f(\cdot)$是激励函数(如：sigmoid、hyperbolic tangent、rectified linear unit等)。下采样层的作用是对特征图进行降维，并且提供一定程度的尺度不变特性。全连接层通常在整个卷积神经网络的末端，并且通过一个softmax层输出针对输入图像所属类别的一个概率分布。卷积神经网络的训练通常采用随机梯度下降的方法(stochastic gradient descent, SGD)，训练过程中$\boldsymbol{W}$和$\boldsymbol{b}$会被更新，更新的幅度由学习速率η控制。为了减轻网络的过拟合，“weight decay”参数λ通常会被加入网络的损失函数中以控制整个网络的过拟合强度。

2.2 卷积神经网络的迁移学习

卷积神经网络的训练对于训练数据集的数量要求很高，因此大型图像分类数据集(如ImageNet[11])对卷积神经网络的成功起着非常重要的推动作用。而本文希望将卷积神经网络在图像分类领域的成功扩展到场景识别领域，迁移学习是采用的一个主要思路。图2是关于本文利用卷积神经网络进行迁移学习的模型，卷积神经网络将大型图像分类数据集作为先验知识进行预训练，训练好的模型作为通用的特征提取器应用场景识别任务。

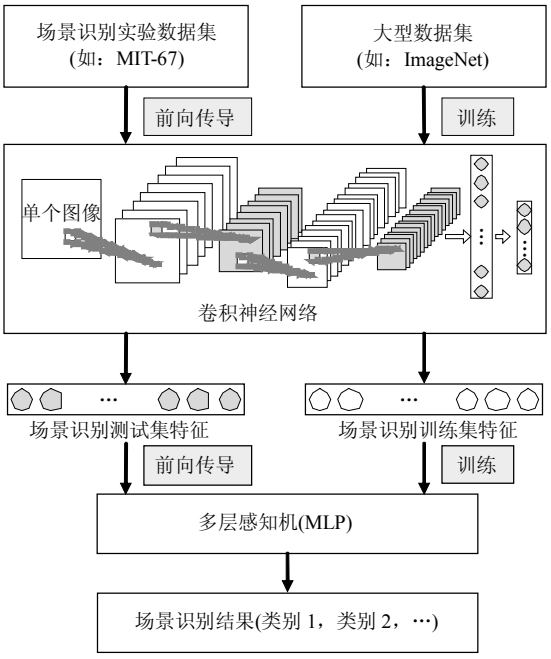

图2 卷积神经网络的迁移学习模型

按照本文提出的场景识别方法的需求，将卷积神经网络的迁移学习过程分为以下4个步骤：

1) 在大型数据集(如：ImageNet[11])上完成卷积神经网络的训练，确定模型中的可训练参数(如：$\boldsymbol{W}$和$\boldsymbol{b}$)；

2) 对于场景识别相关的实验数据集(如：MIT-67[21])，将数据集中的训练集和测试集都通过卷积神经网络进行前向传导，获取到各自的特征向量，而不对卷积神经网络中的训练参数进行更新；

3) 利用场景识别训练集的特征向量训练多层感知机(MLP)；

4) 将MLP用于场景识别测试集的分类，完成场景的识别任务。

通过卷积神经网络的迁移学习，在特定数据集上完成训练的网络模型成为了一个通用的特征提取器。

3 优化的场景识别策略

基于显著区域提取和卷积神经网络的迁移学习方法，本文提出了一个优化的场景识别策略，目标是能够准确地完成场景识别任务。如图3所示，本文提出的场景识别策略主要由多尺度的显著区域提取和基于显著区域的特征学习两大部分组成。

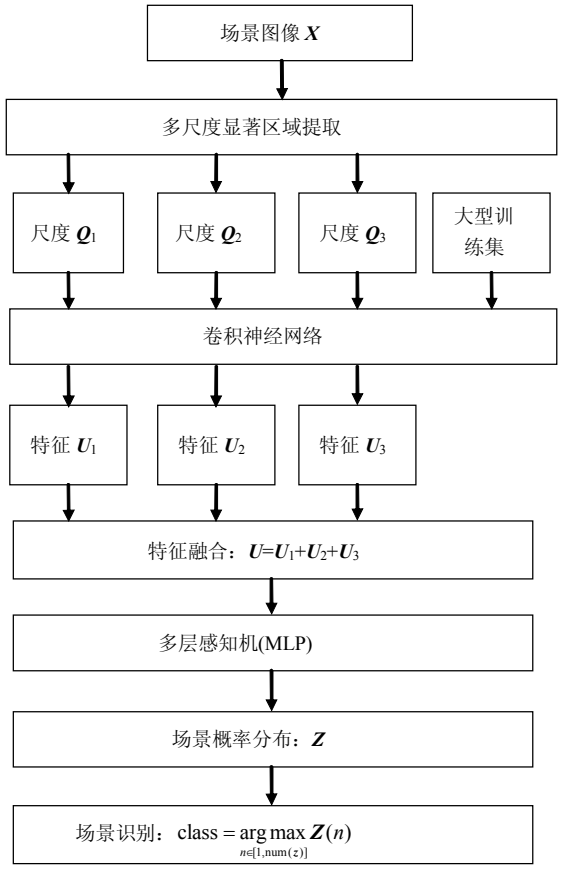

图3　优化的场景识别策略

3.1　多尺度显著区域的提取

由于视距的不确定性，一个未知场景中的物体大小也是无法确定的。不确定的物体大小会影响到目标或者场景识别的准确度。本文利用了多尺度的显著区域提取策略来减轻这一问题的影响。

针对一幅场景图像$\boldsymbol{X}$，除了场景本身($\boldsymbol{Q}_1$=$\boldsymbol{X}$)以外，利用1.2节中提出的方法分别提取两个不同尺度的显著区域$\boldsymbol{Q}_2$和$\boldsymbol{Q}_3$。$\boldsymbol{Q}_1$、$\boldsymbol{Q}_2$和$\boldsymbol{Q}_3$组合成为了原本场景的一个多尺度显著区域的表达。实验表明，基于多尺度的显著区域特征提取相比于单一的尺度显著区域特征提取能取得更好的场景识别准确度。

3.2　显著区域的特征学习

对于提取得到的显著区域$\boldsymbol{Q}_1$、$\boldsymbol{Q}_2$和$\boldsymbol{Q}_3$，本文利用了已经在大型数据库上完成训练的卷积神经网络对其进行前向传导，分别提取特征$\boldsymbol{U}_1$、$\boldsymbol{U}_2$和$\boldsymbol{U}_3$：

$$\boldsymbol{U}(k)=g(\boldsymbol{Q}(k);(\boldsymbol{W},\boldsymbol{b})) \tag{5}$$

式中，$\boldsymbol{W}$和$\boldsymbol{b}$通过在大型图像数据库上训练得到；$g(\boldsymbol{Q}(k);(\boldsymbol{W},\boldsymbol{b}))$表示对输入$\boldsymbol{Q}(k)$进行前向传导，而不更新网络的训练参数($\boldsymbol{W}$和$\boldsymbol{b}$)。对3个尺度下显著区域的特征进行相加融合：

$$\boldsymbol{U}=\boldsymbol{U}_1+\boldsymbol{U}_2+\boldsymbol{U}_3 \tag{6}$$

基于融合后的特征$\boldsymbol{U}$，本文训练了一个多层感知机(MLP)用于预测场景类别：$\boldsymbol{U}\xrightarrow{\text{MLP}}\boldsymbol{Z}$，$\boldsymbol{Z}$是MLP的输出，即为针对场景类别的一个概率分布。MLP的损失函数定义为：

$$E=-\sum_{k=1}^{\text{num}(\boldsymbol{Z})}\log\boldsymbol{Z}(k)+\frac{\lambda}{2}\boldsymbol{W}_{\text{mlp}}^{\text{T}}\boldsymbol{W}_{\text{mlp}} \tag{7}$$

式中，$\boldsymbol{W}_{\text{mlp}}$是MLP的可训练参数。MLP的训练对象是场景识别训练集通过卷积神经网络前向传导后得到的特征向量$\boldsymbol{U}$。训练过程采用常用的随机梯度下降方法，通过残差的反向传导，更新$\boldsymbol{W}_{\text{mlp}}$的值，以降低整个网络的“损失”($E$)。场景最终的类别由预测概率最大的类别确定：

$$\text{class}=\arg\max_{n\in[1,\text{num}(z)]}\boldsymbol{Z}(n) \tag{8}$$

4　实验分析

为了验证本文提出的场景识别方法的有效性，选择在场景识别领域的基准测试数据库MIT-67[21]和SUN397[1]作为本文的测试实验数据集。

4.1　实验配置

为了验证本文方法的泛化能力，本文在两个数据集上的实验都使用了同样的一套参数和模型：

1) 在显著区域的尺度上，选择原图($\boldsymbol{Q}_1$)以及原图较短边长的90%($\boldsymbol{Q}_2$)和80%($\boldsymbol{Q}_3$)共3个尺度。并且，在实验中针对单一尺度和多尺度融合的场景识别效果进行比较。

2) 运用在ImageNet[11]和Places[22]两个大型图像数据库上训练得到的卷积神经网络：HBCNN (Hybrid-CNN)[22]。HBCNN采用了经典的AlexNet[11]网络结构，由5个卷积层和3个全连接层组成。与AlexNet在ImageNet上训练不同，HBCNN在训练集的选择上，融合了ImageNet和Places两个数据集的图片。ImageNet属于物体识别数据集，包含1 000种物体类别共150万张图片。Places属于场景识别数据集，包含476个场景类别和700万张图片。在与场景相关的大型训练集上的训练保证了HBCNN在后续相关实验数据集(MIT-67、SUN397)的迁移学习能力。

3) 在特征提取过程中，输入HBCNN的图像均通过双线性插值缩放到HBCNN的输入大小(227×227×3)。输出方面，采用了HBCNN的“fc8”层的输出特征，特征的维度是1 183。

4) 多层感知机(MLP)总共3层，包含输入层、隐含层和输出层。输入层的神经元数量跟提取特征的维度相同，共1 183个。MLP的隐含层包含512个神经元。MLP的输出层维度与相应测试数据集的场景

类别数量一致，针对MIT-67数据集的MLP输出层包含67个神经元，而针对SUN397数据集的MLP输出层包含397个神经元。

5) 将学习速率(η)和weight decay(λ)的值分别设置为1×10^{-5}和5×10^{-4}。对于η，从1×10^{-1}开始进行多次试验，每次η的取值都是上一次的十分之一，最终选择了实验结果最好的1×10^{-5}。而λ则是基于经验值，并未进行特别的调试。

4.2 MIT-67室内场景数据集测试

MIT-67数据集包含了67种室内场景，每种场景100张图片，共6 700张场景图片的大小并不完全相同。实验采用MIT-67数据集提供的标准训练集和测试集划分。总共包含6 700张图片的数据集中，80%的图片被划分为训练集，而测试集包含了数据集中剩余20%的图片。

如表1所示，本文提出的基于显著区域特征学习的场景识别方法相比于传统的场景识别算法在场景识别的准确度上具有竞争力。另外，相比于单一尺度的特征提取，多尺度的显著区域的特征融合能够有效地提高场景识别的准确度。

表1 MIT-67场景识别数据上的实验结果对比

方法	准确度/%
可变形部件模型[6]	30.4
判别性图像块查找[8]	38.1
基于滤波器的特征学习[9]	52.2
空间金字塔匹配(SPM)[5]	61.2
方向金字塔匹配(OPM)[5]	51.5
SPM + OPM[5]	63.5
尺度$\boldsymbol{Q}_1$	61.6
尺度$\boldsymbol{Q}_2$	60.1
尺度$\boldsymbol{Q}_3$	58.2
多尺度：$\boldsymbol{Q}_1+\boldsymbol{Q}_2+\boldsymbol{Q}_3$	65.6

4.3 SUN397大型场景识别数据库基准测试

相比于MIT-67，SUN397是一个更加大型和完善的场景识别数据集。SUN397总共包含了397种场景，场景类型涵盖了室内和室外的各种环境。每种场景的图片数量与MIT-67一样是100张图片，总共39 700张图片的大小也并不完全相同。实验依据SUN397提供的10组训练集和对应测试集的划分，每组训练集和测试集的图片均是通过等分整个数据集的图片而得到。最终的结果通过对10组测试集上的实验结果取平均值获得。

SUN397数据集上的实验结果如图4所示。GIST、LBP、SIFT、Texton和HOG是基于单一传统人工设计特征的场景识别准确率。“all”项表示的是包含了以上单一特征的一系列传统特征叠加后取得的结果。从实验结果看，本文提出的基于显著区域的特征学习方法相比于传统的人工设计特征在场景识别的准确度上有明显的提高。另外，多尺度的显著区域特征学习($\boldsymbol{Q}_1$-$\boldsymbol{Q}_3$)相对于基于单一尺度的特征学习($\boldsymbol{Q}_1$, $\boldsymbol{Q}_2$, $\boldsymbol{Q}_3$)在场景识别的准确度上取得了更好结果。这一特点与本文在MIT-67数据集上获得的实验结论一致。

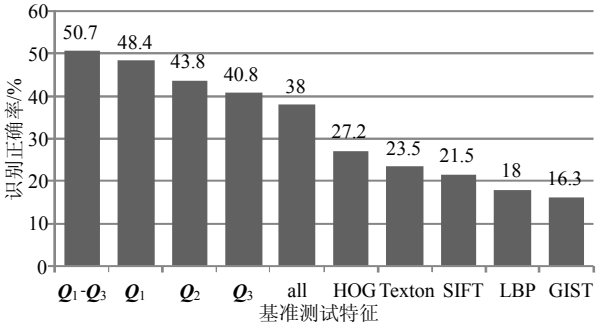

图4 SUN397数据集基准测试结果

5 讨 论

本文提出的方法虽然在MIT-67和SUN397两个常用的场景识别数据集上均取得了良好的实验效果，证实了方法的有效性。但是，本文的方法仍然存在一定的改善空间，一些改善思路如下：

1) 采用更好的学习特征。近期，文献[23]利用global average pooling方法[24]结合GoogLeNet[12]提取场景特征，在MIT-67和SUN397数据集上分别取得了66.6%和51.7%的场景识别准确度，略高于本文中的65.6%和50.7%的实验结果。相比于本文中HBCNN使用的AlexNet结构，GoogLeNet的网络结构更加优化，并且在ImageNet的数据集测试中，GoogLeNet的准确度(93.3%)远高于AlexNet(83.6%)，体现出更强的特征提取能力。另外，global average pooling的方法取消了卷积网络中的全连接层，直接对特征图进行下采样，有效地解决了全连接层的过拟合问题，提高了提取特征的判别和泛化能力。利用判别性能更强的网络，或者针对HBCNN在目标训练集上进行fine-tuning，均有助于进一步提高网络提取特征的判别性能，是改进本文方法的一个途径。

2) 多个显著区域特征提取。根据式(3)，本文的方法只提取出一个尺度下最为显著的单个区域$\boldsymbol{B}(a_{\max})$的特征。但是，针对一些较为复杂的场景条件，其显著区域并不止一处。设计一种有效的方式来改进显著区域的评价标准，提取场景中可能存在的多个显著区域是改善本文方法的一个思路。

3) 特征融合方式的改进。针对多尺度显著区域的特征，本文采用了简单的相加融合方式。针对多特征的融合，对各种特征进行带权值的相加，或者通过特征拼接后降维，以获得更具判别性能的特征，都是进一步改善实验结果的潜在方法。

6 结 束 语

本文提出了一种基于多尺度显著区域特征学习的场景识别方法。该方法通过在多尺度条件下提取一个场景的显著区域，并且利用卷积神经网络的迁移学习来提取这些区域的特征信息，能够有效地完成场景识别的任务。基于场景识别数据库的基准测试表明，本文提出的方法相比于现有的典型场景识别方法对于场景识别的准确度有较为明显的提高。

参 考 文 献

[1] XIAO J, HAYS J, EHINGER K A, et al. Sun database: Large-scale scene recognition from abbey to zoo[C]//CVPR. San Francisco, USA: IEEE, 2010: 3485-3492.

[2] LAZEBNIK S, SCHIMID C, PONCE J. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories[C]//CVPR. New York, USA: IEEE, 2006: 2169-2178.

[3] OLIVA A, TORRALBA A. Modeling the shape of the scene: a holistic representation of the spatial envelope[J]. International Journal of Computer Vision, 2001, 42(3): 145-175.

[4] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91-110.

[5] XIE L, WANG J, GUO B, et al. Orientational pyramid matching for recognizing indoor scenes[C]//CVPR. Columbus, USA: IEEE, 2014: 3734-3741.

[6] PANDY M, LAZEBNIK S. Scene recognition and weakly supervised object localization with deformable part-based models[C]//ICCV. Barcelona, Spain: IEEE, 2011: 1307-1314.

[7] FELZENSZWALB P F, GIRSHICK R B, MCALLESTER D, et al. Object detection with discriminatively trained part based models[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(9): 1627-1645.

[8] SINGH S, GUPTA A, EFROS A. Unsupervised discovery of mid-level discriminative patches[C]//ECCV. Florence, Italy: Springer, 2012: 73-86.

[9] ZUO Z, WANG G, SHUAI B, et al. Learning discriminative and shareable features for scene classification[C]//ECCV. Zurich, Switzerland: Springer, 2014: 552-568.

[10] LECUN Y, BENGIO Y, HINTON G E. Deep learning[J]. Nature, 2015, 521: 436-444.

[11] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[C]//NIPS. Lake Tahoe, USA: MIT Press, 2012: 1106-1114.

[12] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[C]//CVPR. Boston, USA: IEEE, 2015:1-9.

[13] RAZAVIAN A S, AZIZPOUR H, SULLIVAN J, et al. CNN features off-the-shelf: an astounding baseline for recognition[C]//CVPR Workshops. Columbus, USA: IEEE, 2014: 512-519.

[14] ZEILER M D, FERGUS R. Visualizing and understanding convolutional networks[C]//ECCV. Zurich, Switzerland: Springer, 2014: 818-833.

[15] 庄福振, 罗平, 何清, 等. 迁移学习研究进展[J]. 软件学报, 2015, 26(1): 26-39.
ZHUANG Fu-zhen, LUO Ping, HE Qing, et al. Survey on transfer learning research[J]. Journal of Software, 2015, 26(1): 26-39.

[16] ARBELAEZ P, MAIRE M, FOWLKES C, et al. Contour detection and hierarchical image segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(5): 898-916.

[17] ACHANTA R, SHAJI A, SMITH K, et al. SLIC superpixels compared to state-of-the-art superpixel methods[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(11): 2274-2282.

[18] FELZENSZWALB P F, HUTTENLOCHER D P. Efficient graph based image segmentation[J]. International Journal of Computer Vision, 2004, 59: 167-181.

[19] UIJLINGS J, SANDE K, GEVERS T, et al. Selective search for object recognition[J]. International Journal of Computer Vision, 2013, 104(2): 154-171.

[20] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

[21] QUATTONI A, TORRALBA A. Recognizing indoor scenes[C]//CVPR. Miami, USA: IEEE, 2009: 413-420.

[22] ZHOU B, LAPEDRIZA A, XIAO J, et al. Learning deep features for scene recognition using places database[C]//NIPS. Montreal, Canada: MIT Press, 2014: 487-495.

[23] ZHOU B, KHOSLA A, LAPEDRIZA A, et al. Learning deep features for discriminative localization[C]//CVPR. Las Vegas, USA: IEEE, 2016: 2921-2929.

[24] LIN M, CHEN Q, YAN S. Network in network[EB/OL]. [2016-12-14]. http://arxiv.org/pdf/1312.4400v3.pdf.

编　辑　税　红