

中图法分类号: TN391.41 文献标志码: A 文章编号: 1006-8961(2010)10-1521-09

索引信息: 钱乐乐,高隽,谢昭. 一种融合神经稀疏编码机制的层次目标识别算法[J]. 中国图象图形学报,2010,15(10):1521-1529

一种融合神经稀疏编码机制的层次目标识别算法

钱乐乐, 高隽, 谢昭

(合肥工业大学图像信息处理研究室, 合肥 230009)

摘要: 针对基本层次化目标识别计算模型缺乏明确学习概念和有效学习方法的问题,利用神经稀疏编码的学习规则,生成原型向量集合。通过模拟复杂细胞的感受野特性,实现层次化的稀疏编码过程,提出基于神经稀疏编码的层次目标特征提取算法。并利用简化的分类器设计,完成复杂场景下的广义目标识别问题。在 Caltech-101 数据库上进行实验对比,结果表明本文算法相对 Serre 计算模型在识别正确率上有较大提高,时间复杂度增加并不明显,且更加符合生物视觉机理。

关键词: 生物视觉系统; 层次化计算模型; 广义目标识别; 神经稀疏编码

An improved hierarchical generic object recognition algorithm based on neural sparse coding

QIAN Lele, GAO Jun, XIE Zhao

(Laboratory of Image Information Processing, Hefei University of Technology, Hefei 230009)

Abstract: To address the lack of explicit concepts and effective methods with learning in most hierarchical visual computational models, we propose a novel hierarchical generic object recognition sketch using neural sparse coding. Firstly, the learning strategies are embedded in a hierarchical system for generating several prototypes to model the characteristic of complex cell receptive fields. Secondly, based on a hierarchical sparse coding process, we present a hierarchical feature extraction method for generic object recognition. Finally, a simplified classifier is designed to achieve our goals in complex scenes according to the extracted robust features. Experiments on Caltech-101 demonstrate the effectiveness in our method and the more preferable results comparing with Serre's show greater consistency in biological vision.

Keywords: biological vision system; hierarchical computational model; generic object recognition; neural sparse coding

0 引言

理解生物视觉皮层如何实现目标识别是视觉神经科学的一个关键问题。Hubel 和 Wiesel 通过神经生理方法开创了对视皮层细胞的研究^[1],自此关于生物视觉系统的研究成果层出不穷,大量生物视觉

计算模型涌现出来。这些模型大多以生物视觉系统的层次结构为基础,首先建立分层的具有变换不变性的特征提取机制,而后采用判别模型算法,完成目标分类和识别过程^[2-7]。

最早的前馈层次目标识别系统是 Fukushima 提出的“神经认知机”^[2],它是建立在逐层的局部模板匹配和空间汇聚操作基础上的层次化系统,但在将

基金项目: 国家高技术研究发展计划(863)项目(2006AA704332); 国家自然科学基金面上项目(60875012); 安徽省优秀青年科技基金(08040106901)。

收稿日期: 2009-02-10; **改回日期:** 2009-05-22

第一作者简介: 钱乐乐(1981—),男。合肥工业大学计算机与信息学院信号与信息处理专业博士研究生,主要研究方向为图像分析与理解、模式识别、神经计算。E-mail: qianlele@gmail.com。

其应用到手写字符识别时,需要大量手写训练样本,合适的手写样本选择成为其性能优劣的一个关键因素。Perrett 和 Oram 拓展了汇聚的思想,通过对相应特征选择细胞的输入进行合适的汇聚操作^[3],实现了目标由平移不变性到关于任何图像平面变换的不变性。Poggio 和 Edelman 提出一种以 Gaussian 径向基函数为基础的网络结构^[4],利用 Gaussian 函数的调节特性将细胞调节到特定的目标视角上,并进一步汇聚这些细胞响应,实现视角不变的目标识别。Riesenhuber 和 Poggio 进一步强调了有效的汇聚操作的重要性,认为层次化网络结构可以避免细胞组合过程中的组合爆炸问题,提出了一种类似于“神经认知机”的层次化计算视觉框架,称之为 HMAX 模型^[5],其最大特点在于采用非线性操作(SUM 和 MAX 操作)实现前馈信息的传递。多层卷积网络通过反向传播算法对特征进行优化学习,其中平移不变性的约束可以通过权值共享策略实现^[6],但其多层次的梯度学习算法对于大样本和大网络尺寸耗时巨大。Wallis 和 Rolls 提出一种 4 层 VisNet 网络结构^[7],在跟踪规则^[8]的基础上,选择合适的信息测度,利用分布式的信息表达,采用逐层无监督的学习方式,实现兼顾目标不变性和选择性的识别过程。Serre 在 HMAX 基础上,更进一步细化了各个层次的功能,并引入了特征字典的学习过程,使得网络能够计算各种复杂场景下目标的鲁棒性特征^[8]。

上述层次化计算模型均是以前馈信息传递为基础,强调对视觉信息的多层分解,较少涉及神经系统的学习过程,虽然其中也采用了类似于 RBF(radial basis function)基函数^[5]的学习机制,但对 Gaussian 函数中心采取的是随机固定中心的学习方式,其选择的合适程度无法得到保障;另一方面,生理学研究发现从 V1(初级视觉皮层的第一视觉区域)到 IT(颞下视皮层区域)区的细胞主要是以自组织学习为主。大量证据表明^[9],哺乳动物视皮层细胞对自然场景的学习采用神经稀疏编码的过程,即生物细胞在感受视觉刺激时,某一时间段内只有少量神经元处于激活状态,形成对外界视觉刺激的稀疏表达方式,能够有效提取场景中的关键信息,提高对自然场景的编码和表征效率,神经稀疏编码贯穿于初级视觉皮层信息处理的多个阶段中。因此,它也就成为了层次化计算模型中一个必不可少的环节。

以上观点成为了本文研究内容的出发点和关键问题,即如何形成基于视皮层稀疏表达的目标识别

模型。针对 Serre 提出的视觉层次计算模型^[10],融入神经稀疏编码的思想,在其 S2(第 2 层简单细胞)层中引入基于神经稀疏编码的视觉字典学习方法,生成用于编码的特征字典或原型向量;进一步在其 C2(第 2 层复杂细胞)层中对稀疏编码系数计算简单的低维统计特征,生成用于识别的特征向量;最后利用 SVM(支持向量机)分类器实现复杂背景下的广义目标识别。

1 视皮层目标识别模型

研究表明视觉目标描述主要分为两大类模型:一类是基于物体的结构化描述模型,另一类则是基于外观的特定视角特征集合描述模型。前者强调对物体的结构分解和组合,以产生具有视角不变性的目标描述;后者则强调不同视角条件下(不同视点、旋转、光照等)所产生的外观特征的不同。因而,对于复杂背景下的目标识别而言,后一种目标表示方法更具实际意义,并且也符合早期生物视觉处理的一般过程^[10]。HMAX 模型^[5]是 Poggio 等人提出的基于上述外观描述的层次化前馈结构模型,它主要包括两个标准模块:模块 1 表现出与视角调节的 IT 神经元类似的尺度和位置不变性,该模块通过如同纹区复杂细胞的操作方式来实现,即由简单细胞到复杂细胞的最大化(MAX)操作,其中最强的输入决定了当前细胞的输出;模块 2 描述了对多个同类样本的学习过程,其关键思想是通过类似 Gaussian RBF 的网络结构获得同类样本不同视角实例的插值和泛化能力。最大化操作提供了层次结构的不变性,RBF 学习提供了对具体目标的特定描述。

Serre 扩展了 HMAX 模型^[10],提出较为明确的 4 层前馈计算框架,分别由对应于简单细胞的 S 单元与对应于复杂细胞的 C 单元交替组成。S 单元采用 TUNING 操作,用于增加目标的选择性,C 单元具有更大的感受野范围,它们通过 MAX 操作汇聚 S 单元的输出,从而引入了对目标尺度和平移的不变性。该模型可以定量模拟灵长类动物视皮层腹侧视觉流的前馈通路的计算机制,利用基于外观的描述算子权衡不变性与选择性之间的关系。如图 1 所示,其具体的分层结构及其功能为:S1 层计算单元通过一组多尺度多方向的 Gabor 滤波器作用于输入图像,生成分布在 4 个方向和 8 个尺度波段(即每个尺度波段包含两个相邻尺度)上的 64 幅响应图像;C1 层

计算单元分别在 4 个方向 8 个尺度波段上,计算每个尺度波段内相邻两幅 S1 层响应图像对应位置处的最大值,并对获得的图像进行邻域最大化以及降采样操作;在 C1 层响应图像的各个尺度对应 4 个方向上提取 4 种不同尺寸(尺寸为 $4 \times 4, 8 \times 8, 12 \times 12$ 和 16×16)图像块,将得到的这些图像块作为原型基向量。在 C1 层所有位置处测度每个图像块与原型向量之间的高斯距离,进而计算得到 S2 层的响应图像;取 S2 层图像在所有位置和尺度下的最大值,获得平移和尺度不变的 C2 层特征。模型具体参数根据简单和复杂细胞的电生理学特性确定。

如上所述,该模型通过随机抽取大量不同尺寸的各个尺度和方向上的 C1 层图像块,作为 S2 层的原型或中心向量,实现类似于 RBF 基函数的参数拟合。但这一学习过程仅仅是对 C1 层特征的随机抽样,反映了特征块的低阶统计平均特性。然而所抽取的块集合中所包含的高阶统计特性使得块样本集仍存在较大冗余性,不能有效表示 C1 层的关键特征,影响了视觉字典表示的鲁棒性,也未能真正体现学习的含义所在,同时与生物视觉系统的运作机制不相适应。因此,需要引入明确的层次学习概念和方式,采取对 C1 层特征块的有效学习方法,将生物视觉系统的学习感知机理纳入到该计算框架中。

2 神经稀疏编码

2.1 稀疏编码的概念及含义

神经稀疏编码是一种模拟哺乳动物视觉系统主视觉皮层 V1 区细胞感受野特性的神经计算方法。实验研究表明,自然场景的超高斯统计特性与神经元的稀疏编码过程具有高度一致性。

从数学的观点看,神经稀疏编码可以看作是对图像数据的一种分解方法,假设输入图像 $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)^T$ 为 n 维随机向量,经过线性转换矩阵 \mathbf{A} 后,生成 m 维向量 $\mathbf{S} = (s_1, s_2, \dots, s_m)^T$,即

$$\mathbf{S} = \mathbf{A}\mathbf{X} \quad (1)$$

式中, \mathbf{A} 为稀疏变换矩阵,稀疏分量 \mathbf{S} 满足稀疏(超高斯)分布,且其各分量 s_1, s_2, \dots, s_m 尽可能相互独立。“稀疏性”是指随机向量 s_i 中大部分元素为 0,极少元素不为 0,即对应于当视觉系统接收一幅图像时,只有较少的神经元具有较强响应,大部分神经

元则处于弱激活或 0 激活状态。

2.2 稀疏编码的算法实现

神经稀疏编码存在多种实现方法^[11-14],其中非负矩阵分解方法(NMF)方法分解速度快,且认为整体特征是局部特征的非负线性组合,局部特征在构成整体特征时不会产生正负抵消的情况,这种非负分解特性更合乎人类视觉感知的直观体验;Hoyer 将稀疏性约束引入到 NMF 框架中^[13],建立了具有稀疏性约束的非负稀疏编码模型(以下简称 NMFSC),该方法通过稀疏性测度控制稀疏分解过程,使得基向量和系数向量满足规定的稀疏性要求。该方法使用基于 L_1 范数与 L_2 范数之间关系作为稀疏性测度:

$$\text{sparseness}(\mathbf{x}) = \frac{\sqrt{n} - (\sum |x_i|) / \sqrt{\sum x_i^2}}{\sqrt{n} - 1} \quad (2)$$

式中, n 为向量 \mathbf{x} 的维数。当且仅当 \mathbf{x} 仅包含一个非零元素时,该函数值为 1;当且仅当 \mathbf{x} 的所有元素都相等时,该函数值为 0。

算法目标是给定期望的基向量 \mathbf{w}_i 和系数向量 \mathbf{h}_i ,稀疏性程度,约束 NMF 以寻找最优解。

在给定稀疏性约束的条件下:

$$\begin{aligned} \text{sparseness}(\mathbf{w}_i) &= S_w, \quad \forall i \\ \text{sparseness}(\mathbf{h}_i) &= S_h, \quad \forall i \end{aligned} \quad (3)$$

构建目标函数:

$$E(\mathbf{W}, \mathbf{H}) = \|\mathbf{V} - \mathbf{WH}\|^2 \quad (4)$$

式中 \mathbf{V} 表示非负矩阵, \mathbf{W}, \mathbf{H} 分别表示其分解后的非负因子。

对上式采用投影梯度下降算法对代价函数进行优化。

神经稀疏编码机制是生物初期视觉感知所采用的基本视觉响应方式,它能有效地描述场景中包含的高阶统计特性,尤其对于复杂场景而言,目标受背景的影响较为强烈,稀疏编码所产生的基分量在一定程度上可以实现目标和背景在某些高阶统计量上的分离,更有利于识别过程的执行。我们可以看出在 Hoyer 提出的算法中,基向量和系数的稀疏性程度可以得到灵活控制;其分解速度较快,适用于快速的视觉字典(原型向量)训练与生成;该模型考虑了视皮层分别以 ON-Channel 和 OFF-Channel 接收数据,符合主视皮层 V1 区的电生理特性。但另一方面直接对图像进行非负矩阵分解则无法考虑神经元对图像数据在多尺度、多方向上的感受野响应特性。

因此,需将层次化分解思想与非负稀疏编码模型整合起来形成具有复杂细胞感受野特性的稀疏编码方法。

3 层次稀疏化目标特征提取算法

3.1 框架结构及其实现

为了赋予 S2 层以明确的学习含义,将神经稀疏编码的概念引入进来。对 C1 层响应图像进行局部感受野的稀疏编码,随机抽取大量 C1 层的图像块,利用具有稀疏性约束的非负稀疏编码方法进行学习,提取具有一定稀疏性的基向量作为 S2 层的中心或原型向量。一方面,学习所得的原型向量具有可

控的稀疏性,通过调节稀疏性参数可以得到不同的基向量形式,也说明了复杂细胞所具有的稀疏的响应特性;另一方面,这些原型向量能够反映图像中各类目标的局部基元信息,呈现出对训练样本中同类目标局部结构的良好描述,也充分说明了复杂细胞的局部响应特性。

所提出的模型也分为 4 层结构,如图 1 所示,其 S1 和 C1 层在结构和功能上与 Serre 模型相同,根据简单细胞感受野特性,S1 层单元采用 Gabor 滤波器对图像进行多尺度和方向分解;C1 层单元则模拟复杂细胞感受野特性对同一尺度波段内相邻两个尺度的 S1 层响应图像进行汇聚操作,并按照不同尺寸的邻域窗口对其进行局部最大化和降采样操作。

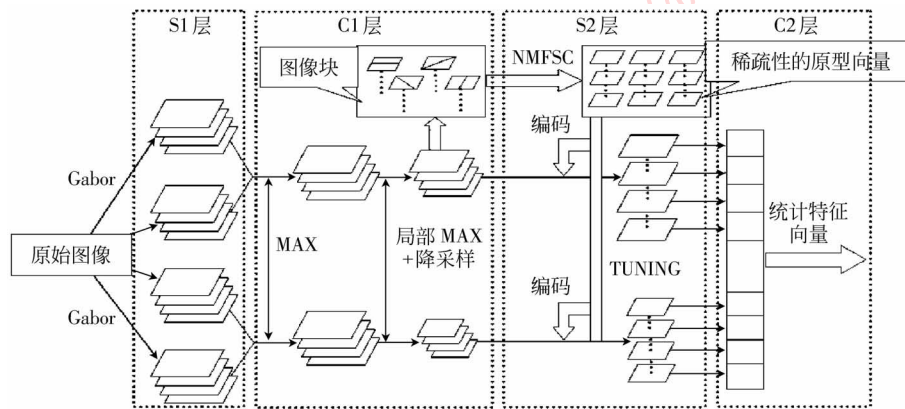


图 1 模型框图

Fig. 1 Block diagram of the proposed model

模型中的 S2 层,我们称之为稀疏编码层,其作用是对所抽取的来自训练样本集中 C1 层图像块进行非负稀疏编码,学习生成用于编码的原型向量,或者说形成复杂细胞的感受野。首先对经白化和低通滤波预处理^[10]后的 C1 层响应图像取块,这里采用大量随机取块方式,而并未采用基于检测算子的局部兴趣区域提取方法,其原因是随机采样使得各图像块之间存在重叠,重叠关系恰好可以为稀疏分解提供更可靠的相关信息,更有利于提高分解的稳定性。之后对取得的图像块数据分别在 ON 和 OFF 通道内实现离线的 NMFSFC 过程,将得到的两组基向量相减,获得 S2 层的特征编码字典,即具有稀疏性的原型向量,并利用它们与 C1 层输出进行卷积操作,生成编码系数。这里需要强调的是我们采用对每类图像分别生成原型向量的方法,即从每类图像集合中随机选择若干幅图像用于 NMFSFC 的基向量学习。

C2 层的操作相对简单,即对每幅图像的编码系数进行统计特征提取,如均值、方差、直方图等简单特征,最终形成描述某幅图像的特征向量。图 2 为某幅图像经过各层细胞模型分解后的响应示意图。

3.2 功能解释

从层次化结构模型看,通过 S1 和 C1 层的分解,所得到的响应图像已经呈现出一定的稀疏特性,并且保持了原始图像中场景的边缘、轮廓、区域形状等结构信息,但这种稀疏特性只是对某一幅图像在多个尺度和方向上直接分解的结果。而对于图像来说,其大部分重要特征与像素间的高阶统计特性有着密切联系,高阶统计特性往往包含了更重要的图像结构和相位特征。神经稀疏编码作为视觉皮层对自然场景的一种编码和表达方式,可以有效提取视觉刺激中的高阶统计特征,这不仅体现在简单细胞层次上,更能体现在复杂细胞层次上^[14]。C1 层复杂细胞的响应特征经过稀疏编码之后,能够将尺度

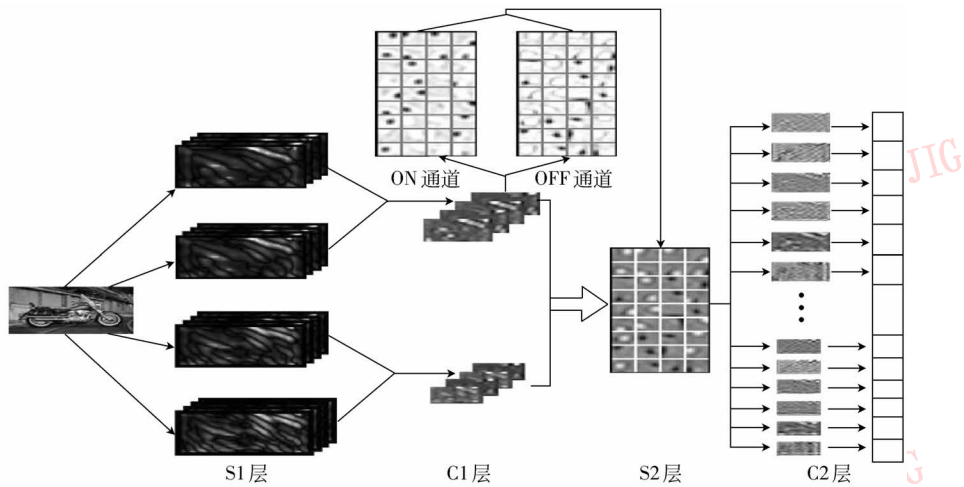


图 2 各层响应示意图

Fig. 2 Response schematic diagram for each layer

和位置的不变性特征融合到所形成的视觉编码字典中,通过具有稀疏性约束的非负稀疏编码,学习复杂细胞的感受野特性,更好地描述目标在场景中的各种变换。可以看出,在经过 S2 层的稀疏编码之后,特征字典描述更多的则是角点、条纹、块等基元信息,这也正体现了复杂细胞对训练样本集的高阶统计特性以及微观稀疏特性的描述。

从稀疏编码的观点看,所提出的模型也可以认为是扩展了多层稀疏编码^[14]的概念,即稀疏编码过程不仅模拟了初级视觉皮层简单细胞的感受野特性,通过引入层次化的多尺度和方向的空间滤波以及 MAX 操作,使得它也适用于学习复杂细胞的感受野特性。

4 实验及其分析

从 CalTech-101 数据库中选择飞机、汽车、人脸、豹子 4 类目标作为正的测试数据集,从每类图像集中随机选取 100 幅待分类的正图像,划分成各 50 幅正的训练和测试图像;同理,从 CalTech-101 数据库中选择背景 (BACKGROUND_Google) 作为负的杂乱背景数据集,划分得到各 50 幅负的训练和测试图像,如图 3 所示。选取任意 30 幅正的训练图像分别随机抽取 4 种不同尺寸 (4 × 4、8 × 8、12 × 12、16 × 16) 的 1000 个 C1 层图像块 (本文只选取了 C1 层第一个尺度),对其在各个尺寸下进行 NMFSC 学习,生成 S2 层编码字典 (稀疏原型向量);对每幅图像

在各尺度下的编码及其特征提取过程详见第 3 节;最后,利用线性 SVM 分类器对获得的特征向量进行训练、测试,得到最终的分类结果。以下所有实验均按照随机生成的训练和测试样本组合重复进行 10 次获得交叉验证后的性能估计,算法分类性能均以 ROC 曲线下的面积表示。通过改变稀疏编码块的尺寸,将本文算法与 Serre 视觉层次计算模型^[9] (非稀疏的 C2 特征) 进行性能比较。

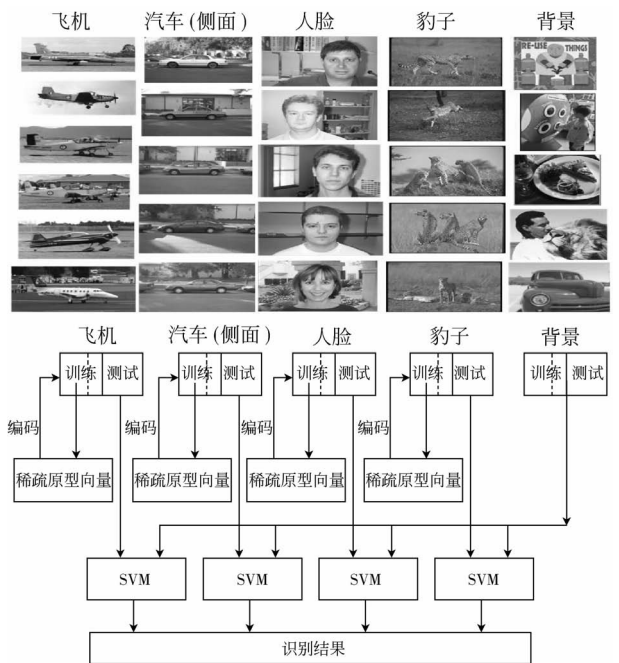


图 3 CalTech101 数据库及其训练测试

Fig. 3 Caltech-101 object database and the training and testing on it

1) 首先对 50 幅正的训练图像分别提取维数为 1 000 的原始 C2 层特征向量,改变特征维数,随机抽取其中的 9、18、36、72、144、288、500、1 000 维特征;其次利用经过稀疏编码后的 C2 层特征向量进行目标分类,在进行稀疏编码时,选取的基向量个数与特征维数相对应,即设置基向量个数分别为 9、18、36、72、144、288,对不同块尺寸(4×4、8×8、12×12、16×16)分别进行特征提取。观察各种不同特征数目下的 Serre 模型与本文提出方法的目标分类性能,如图 4 所示。从图 4 中可以看出,本文提出的方法较 Serre 模型在分类性能上有较大提高,尤其是在特征数较少的情况下,对 S2 层采用稀疏编码学习方式能够提高模型的泛化能力,使用较少的特征即能得到较高的分类正确率;另一方面,不同的块尺寸差别也较大,4×4 分块尺寸较小,对于图像中尺度较大的目标(如飞机和人脸)呈现的分类稳定性较差,因此,中等尺寸分块是一个较为合理的选择(如表 1)。表 1 描述了特征数目为 288 时,各目标类的

分类正确率,并比较了算法对各目标类的平均时间复杂度。本文算法的时间耗费分为前后两部分,分别是前期的离线学习时间和后期的在线特征提取时间。离线学习时间指采用 NMFSC 算法学习生成稀疏原型向量的时间,在线特征提取时间指利用生成的稀疏原型向量完成所有特征提取过程的时间。这两部分时间耗费均随着编码块尺寸的增大而增加。从表 1 可以看出,虽然学习过程需要花费一定时间,但与 Serre 模型中对 4 种块尺寸的联合表示不同,本文算法仅采用一种编码块尺寸作为特征描述,因此其时间耗费增加并不明显。

2) 改变正的训练样本的数目,分别为 1、3、6、15、30、40,观察两种模型不同的目标分类效果,如图 5 所示。从图 5 中可以看出,随着正样本数的增加,本文方法的分类精度要逐渐优于 Serre 模型,其主要原因是样本数的增加给稀疏编码原型向量提供了更加可靠的关于目标变化的信息,增强了其泛化性能,提高了分类正确率。

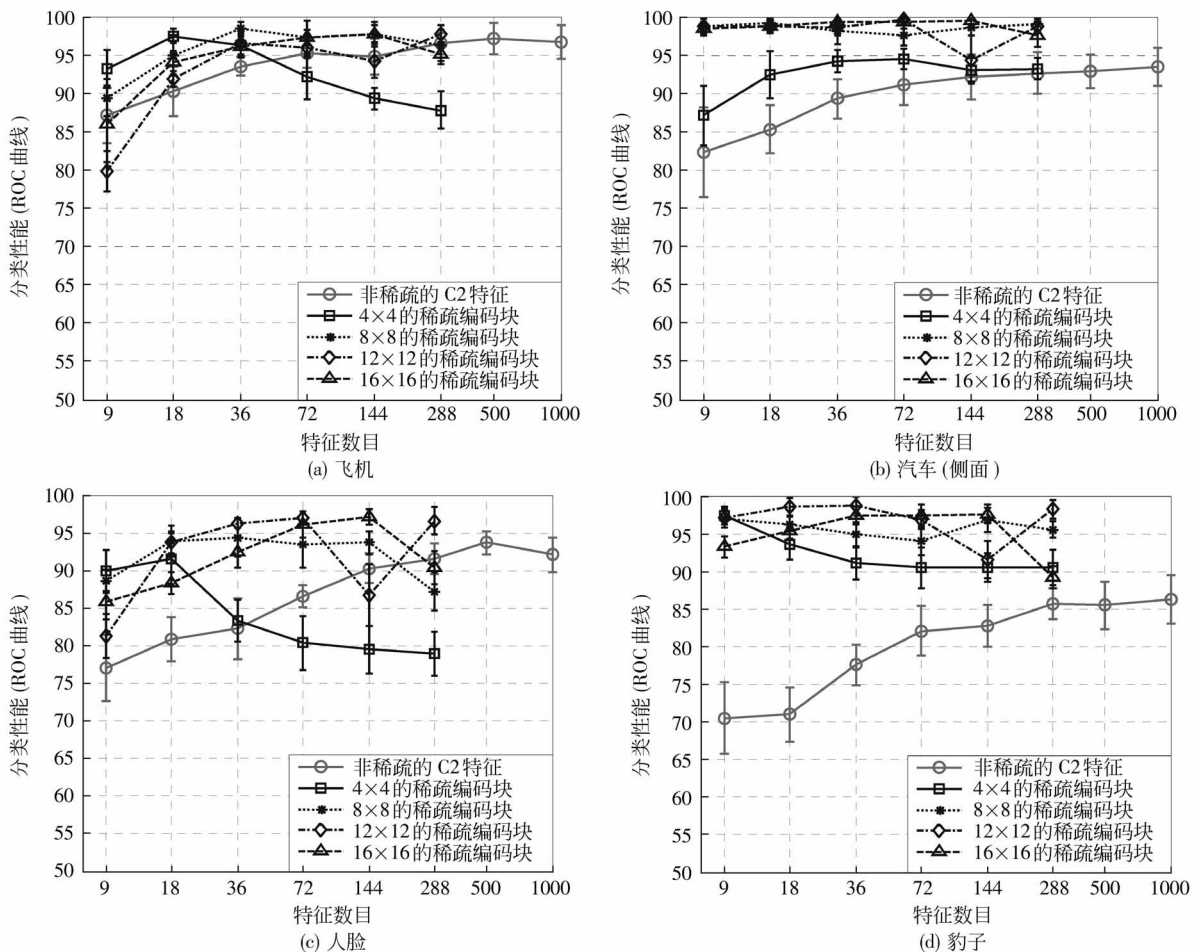


图 4 特征数目对分类性能的影响

Fig. 4 Influence of number of features on the performance

表 1 分类正确率比较 (特征数目为 288)
Tab. 1 Comparison of classification accuracy (number of feature is 288)

目标类	非稀疏 C2 特征	稀疏编码块尺寸			
		4 × 4	8 × 8	12 × 12	16 × 16
飞机	95 ± 2.666 7	86.7 ± 3.164	96.1 ± 2.7264	97.9 ± 1.5951	95.6 ± 2.503 3
汽车	94 ± 1.247 2	93.7 ± 2.213 6	99 ± 0.6667	98.9 ± 1.100 5	96.8 ± 1.398 4
人脸	89.7 ± 2.1108	77.4 ± 3.687 8	90.9 ± 4.508	95.4 ± 2.118 7	88.6 ± 4.599 5
豹子	85.7 ± 3.301 5	89.3 ± 1.4181	96.7 ± 1.337 5	98.4 ± 0.843 2	88.2 ± 3.011 1
平均算法时间/s	3068.288	454.3 + 606.485	846.5 + 670.292	1680.8 + 759.500	2686.5 + 877.197

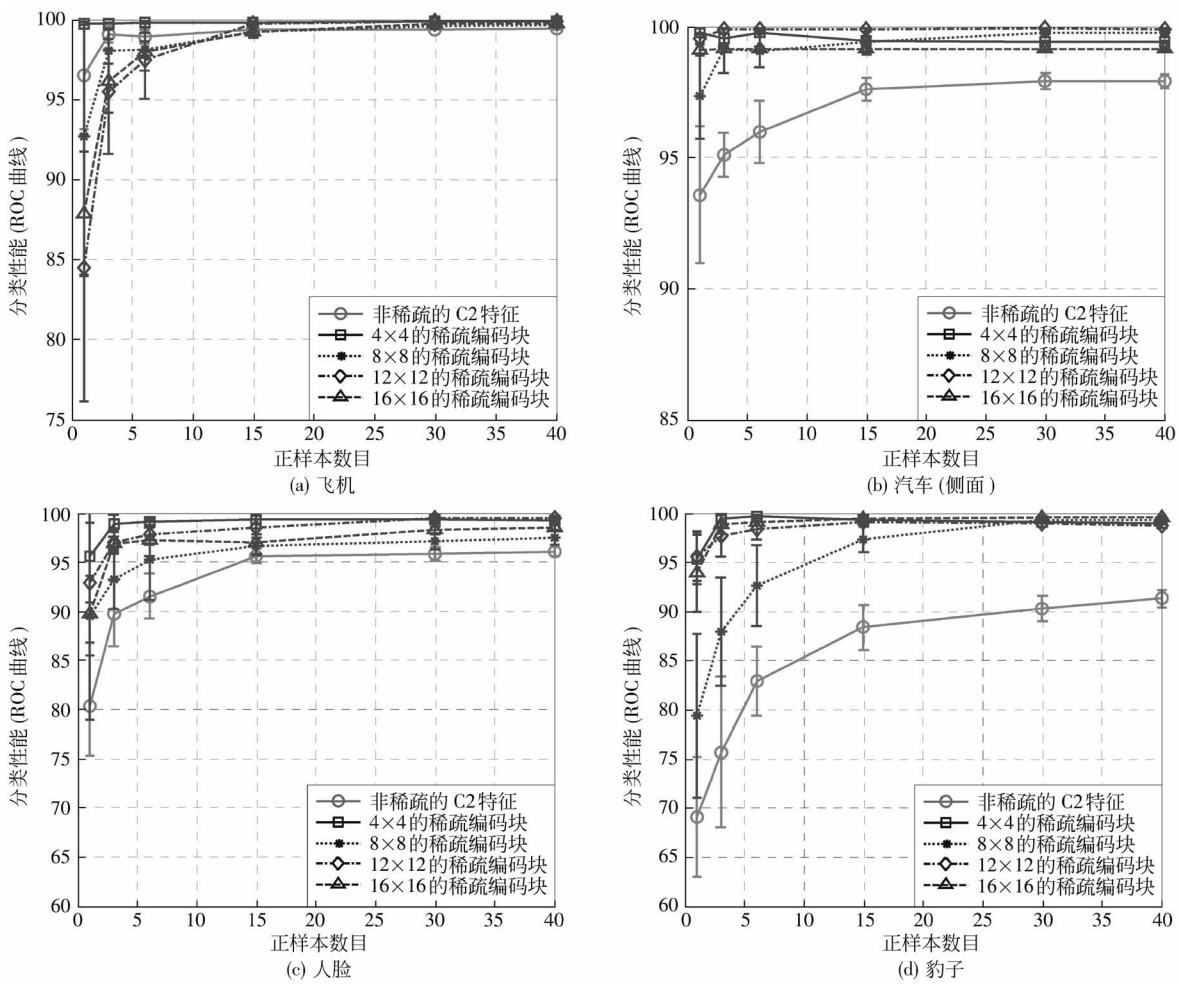


图 5 样本数对分类性能的影响

Fig. 5 Influence of number of training examples on the performance

3) 改变权值稀疏性参数, 分别设置为 0.2、0.5、0.8, 即权值稀疏性由弱到强变化, 这里选择人脸数据集, 设置其基向量个数为 36, 生成的编码原型向量如图 6 所示, 可以看出, 随着稀疏性参数的增加, 稀疏编码算法对人脸目标的局部化描述程度越来越

强; 观察参数变化对分类性能的影响, 如图 7 所示, 从图中可以看出, 无论稀疏性参数设置为何值, 其总体分类性能总是优于 Serre 模型。尽管稀疏性参数变化较大, 但不同参数设置下生成的特征分类性能的差别并不明显。

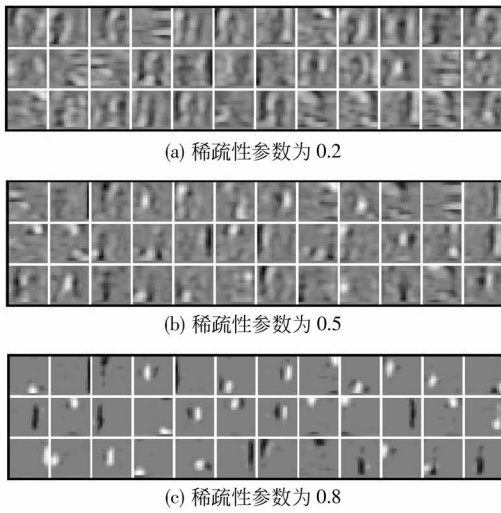


图 6 稀疏性参数变化下的编码原型向量

Fig. 6 Representation of prototypes induced by different sparseness constraints

5 结 论

利用神经稀疏编码规则对 Serre 模型的复杂细胞 C1 层学习,实现 S2 层稀疏的原型向量提取,对不同图像的 C1 层响应进行稀疏编码,进一步对得到的编码系数计算简单的统计特征。实验结果表明,本文算法在对复杂场景进行分类时较 Serre 模型有较大的性能提高,采用中等尺寸的图像块能够使得分类性能更加稳定;稍增加样本数目即可使分类性能有显著提高;且该算法受到稀疏性参数的影响较小。

本文算法侧重于神经元对视觉刺激的稀疏性响应特性的模拟,将层次化计算模型与这一特性有机融合,针对广义目标识别取得了较好的性能。文中采用了固定尺寸的编码块,如何根据图像中目标大小自适应调整块尺寸是值得探讨的问题。

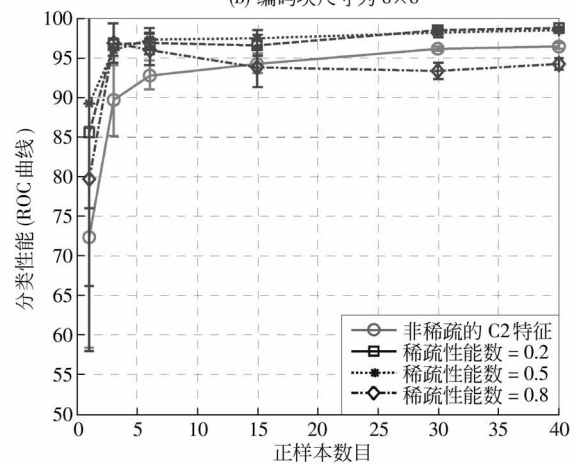
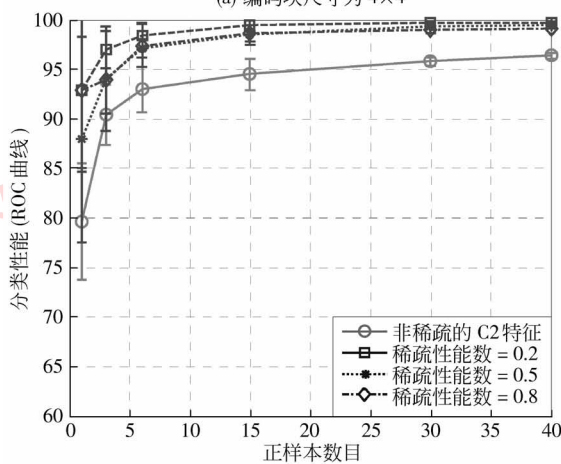
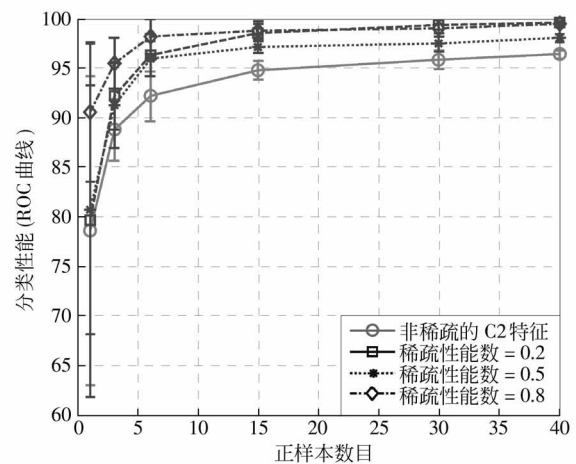
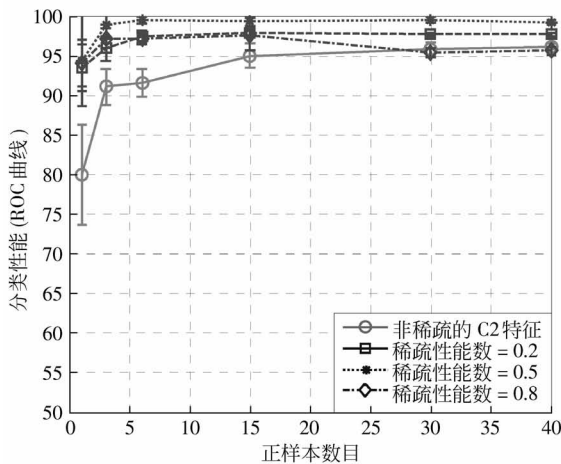


图 7 不同块尺寸下稀疏性参数变化对分类性能的影响

Fig. 7 Influence of different sparseness constraints on the performance in different sizes of coding patches

参考文献 (References)

- [1] Hubel D H, Wiesel T N. Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat [J]. *Journal of Neurophysiology*, 1965, 28(2): 229-289.
- [2] Fukushima K. Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position [J]. *Biological Cybernetics*, 1980, 36(4): 193-202.
- [3] Perrett D, Oram M. Neurophysiology of shape processing [J]. *Image and Vision Computation*, 1993, 11(6): 317-333.
- [4] Poggio T, Edelman S. A network that learns to recognize 3D objects [J]. *Nature*, 1990, 343(6255): 263-266.
- [5] Riesenhuber M, Poggio T. Are cortical models really bound by the "binding problem" [J]. *Neuron*, 1999a, 24(1): 87-93.
- [6] 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.
- [7] Wallis G, Rolls E T. A model of invariant object recognition in the visual system [J]. *Progress in Neurobiology*, 1997, 51: 167-194.
- [8] Földiák P. Learning invariance from transformation sequences [J]. *Neural Computation*, 1991, 3(2): 194-200.
- [9] Olshausen B A, Field David J. Emergence of simple-cell receptive field properties by learning a sparse code for natural images [J]. *Nature*, 1996, 381: 607-609.
- [10] Thomas Serre, Lior Wolf, Stanley Bileschi, Maximilian and Tomaso Poggio. Robust object recognition with cortex-like mechanisms [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007, 29(3): 411-426.
- [11] Tenenbaum J B. and Freeman William T. Separating style and content with bilinear models [J]. *Neural Computation*, 2000, 12(6): 1247-1283.
- [12] Hyvärinen A, Oja E, Hoyer P. et al. Image feature extraction by sparse coding and independent component analysis [C] // *Proceedings of the 14th International Conference on Pattern Recognition*. Washington DC, USA: IEEE Computer Society, 1998, 2: 1268-1273.
- [13] Hoyer P O. Non-negative matrix factorization with sparseness constraints [J]. *Journal of Machine Learning Research*, 2004, 5: 1457-1469.
- [14] Hyvärinen A, Hoyer P O. A two-layer sparse coding model learns simple and complex cell receptive fields and topography from natural images [J]. *Vision Research*, 2001, 41(18): 2413-2423.