

## 基于结构相似度的稀疏编码模型<sup>\*</sup>

李志清<sup>1,2,3+</sup>, 施智平<sup>1</sup>, 李志欣<sup>1,2</sup>, 史忠植<sup>1</sup>

<sup>1</sup>(中国科学院 计算技术研究所 智能信息处理重点实验室,北京 100190)

<sup>2</sup>(中国科学院 研究生院,北京 100049)

<sup>3</sup>(湘潭大学 信息工程学院,湖南 湘潭 411105)

### Sparse Coding Model Based on Structural Similarity

LI Zhi-Qing<sup>1,2,3+</sup>, SHI Zhi-Ping<sup>1</sup>, LI Zhi-Xin<sup>1,2</sup>, SHI Zhong-Zhi<sup>1</sup>

<sup>1</sup>(Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, The Chinese Academy of Sciences, Beijing 100190, China)

<sup>2</sup>(Graduate University, The Chinese Academy of Sciences, Beijing 100049, China)

<sup>3</sup>(College of Information Engineering, Xiangtan University, Xiangtan 411105, China)

+ Corresponding author: E-mail: lizhiqingchina@gmail.com, http://www.intsci.ac.cn

Li ZQ, Shi ZP, Li ZX, Shi ZZ. Sparse coding model based on structural similarity. *Journal of Software*, 2010,21(10):2410–2419. <http://www.jos.org.cn/1000-9825/3675.htm>

**Abstract:** Current existing sparse coding models employ the mean square of the error between the actual image and the reconstructed image to measure how well the code describes the image. Under the assumption that human visual perception is highly adapted for extracting structural information from a scene or a video, an alternative measure for information preservation assessment, based on the structural similarity, is introduced. After minimizing the cost function, the improved model attains a complete family of localized, oriented, and bandpass receptive fields, similar to those found in the primary visual cortex. The experimental results show that the improved sparse coding model is more consistent in human visual system.

**Key words:** sparse coding; natural image; structural similarity; biological visual system; computational model

**摘要:** 已有的稀疏编码模型采用误差的平方和作为信息保持的客观评价标准,但最近的研究表明,人眼视觉系统的主要功能是从视觉区域提取图像和视频中的结构化信息.引入结构相似度来衡量信息保持的程度,通过对改进的目标函数进行优化,获得与初级视皮层中具有局部性、朝向性和带通性的感受野相类似的基函数集.实验结果表明,改进后的稀疏编码模型更符合人眼视觉系统特性.

**关键词:** 稀疏编码;自然图像;结构相似度;生物视觉系统;计算模型

中图法分类号: TP18 文献标识码: A

\* Supported by the National Natural Science Foundation of China under Grant No.60805041 (国家自然科学基金); the National Basic Research Program of China under Grant No.2007CB311004 (国家重点基础研究发展计划(973)); the National High-Tech Research and Development Plan of China under Grant No.2007AA01Z132 (国家高技术研究发展计划(863)); the National Science and Technology Support Plan of China under Grant No.2006BAC08B06 (国家科技支撑计划)

Received 2008-12-01; Revised 2009-04-03; Accepted 2009-07-07

大脑从外部世界接收到的大量信息中,绝大部分是通过视觉系统进行加工处理的.视觉信息的加工机制一直是心理学、神经科学、计算机科学等学科研究的重大课题之一<sup>[1]</sup>.众所周知,生物视觉系统是在长期的进化和不断地适应自然环境中发展起来的.但是,生物视觉系统与外界环境之间到底有什么样的联系?生物视觉系统是如何对外界环境的刺激模式做出响应的?成为各个领域的科学家们长期探索的热点问题.

Barlow<sup>[2]</sup>最先提出有效编码假设,认为生物视觉系统在初级阶段的一个重要功能就是尽可能地去除输入刺激中的统计冗余.他在研究中发现,初级视皮层的视神经细胞对外界环境的响应满足稀疏编码的特性.所谓稀疏编码是指在尽可能完整地表示输入刺激模式的基础上,同时使得被激活的神经细胞的个数尽可能地少.

另外有研究者们利用自然图像的统计特性,通过建立数学模型来模拟早期视觉处理系统的加工机制<sup>[3]</sup>,然后再与初级视皮层中神经细胞的响应特性(神经生理学)进行比较.Field 和 Olshausen<sup>[4]</sup>提出了著名的稀疏编码模型,该模型通过基函数线性叠加表示输入刺激图像,在最小均方差意义下使得重构图像尽可能地与原图像相似,同时要求“响应”尽量稀疏化.通过训练学习,该模型获得的基函数具有视皮层简单细胞表现出来的性质,即局部性、朝向性和带通性,这与神经生理学实验结果非常吻合.在此基础上,很多的研究者在稀疏编码模型的理论和应用方面做了大量的工作,取得了丰硕的成果,也提出了许多改进的稀疏编码模型.为了叙述方便,在本文中我们称 Olshausen 提出的稀疏编码模型为标准稀疏编码模型.

从 20 世纪 70 年代开始人们就已经发现,误差的平方和没有充分地考虑人眼的视觉特性,不能很好地和视觉质量相吻合,因此它不能很好地评价重构图像与原图像之间的相似程度,即信息保持程度.本文引入结构相似度作为信息保持程度的重要衡量指标,提出了基于结构相似度的稀疏编码模型,通过训练学习获得了与标准稀疏编码模型相似的基函数集合.自然图像的编码实验结果表明,基于结构相似度的稀疏编码模型比标准稀疏编码模型更好地保持了原图像中的结构信息,“响应”结果更加符合生物视觉系统特性.

## 1 稀疏编码模型的相关研究

Field 提出稀疏编码也称为信息熵最小化编码的方法<sup>[5,6]</sup>,将自然图像刺激对应于神经网络的输入,通过训练学习使得被激活的神经细胞的个数尽可能地少.另外,他们通过对 Gabor 滤波器的参数进行最优化,发现优化后的滤波器与视皮层中简单细胞的响应特性非常相似,表现为位置选择性、方向选择性和频率选择性.

Olshausen 等人使用线性叠加的方法进一步研究了简单细胞感受野性质与稀疏编码之间的关系,如图 1 所示.利用基函数的线性叠加和表示输入刺激图像,在最小均方差意义下使得线性叠加和的结果尽可能地与原图像相似.同时,尽可能地使表示的特征稀疏化.换句话说,使得基函数的权值  $a_i$  尽可能地多地为 0 或者接近 0.

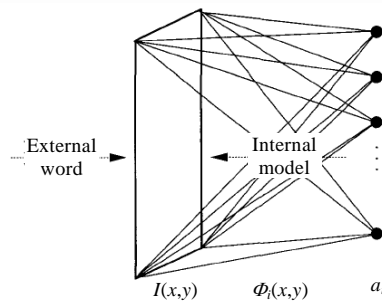


Fig.1 Linear superposition model of image

图 1 图像线性叠加和模型

图像的线性叠加和可以表示为

$$X=AS \quad (1)$$

向量  $X$  表示输入刺激图像,由若干个基函数线性叠加而成,这些基函数分别构成了基函数矩阵  $A$  的列;同时,在线性叠加时,每个基函数的权值分别对应于向量  $S$  的一个分量.从神经生物学的角度考虑,图像线性叠加模型

可以被看成是视皮层神经细胞的响应过程,而  $S$  表示初级神经细胞对外界图像刺激的响应值.在标准稀疏编码模型中,Olshausen 提出的优化准则可以表示为

$$E(a, \Phi) = \sum_{x,y} \left[ I(x, y) - \sum_i a_i \Phi_i(x, y) \right]^2 + \lambda \sum_i S \left( \frac{a_i}{\sigma_i} \right) \quad (2)$$

这里,  $I(x, y)$  表示原图像块  $X$  中的像素值,  $\Phi_i(x, y)$  表示基函数矩阵  $A$  的第  $i$  个列向量,  $a_i$  表示向量  $S$  的第  $i$  个响应值.公式(2)中的第 1 项为重构图像的信息保持程度,用原图像与重构图像之间的误差平方和表示;第 2 项反映了编码即响应的稀疏程度.由于在线性叠加和模型中直接处理的对象是神经细胞感受野大小的图像块,因此我们需要先对原图像进行窗口采样,再将每个图像块按列排列成一个  $N$  维列向量  $X$ .

根据标准稀疏编码模型,从自然图像中随机选取若干图像块构成一个训练集合,经过几百次训练学习后,基函数具有与视皮层简单细胞类似的性质,如图 2 所示.

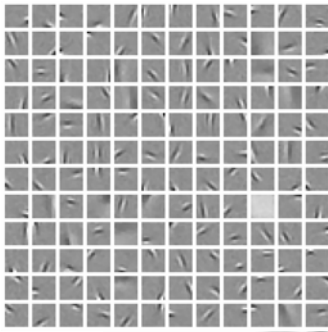


Fig.2 Basis functions of standard sparse coding model

图 2 标准稀疏编码模型的基函数

Hyvarinen<sup>[7]</sup>提出了一个两层的稀疏编码模型,通过训练学习后,获得了类似于复杂细胞响应特性的基函数,该基函数集合表现为有规律的拓扑结构.因此,稀疏编码模型适合模拟视觉系统高级区域的信息处理机制.Tenenbaum 和 Rao 等人<sup>[8,9]</sup>对视觉不变性问题进行研究,提出了双线性稀疏编码模型.Hoyer<sup>[10]</sup>通过非负稀疏编码模型更合理地解释了神经细胞“响应”的非负性.Simoncelli<sup>[11,12]</sup>提出非线性稀疏编码模型,对“响应”结果进行非线性变换,从而消除了线性编码结果的高阶相关性.Hoyer<sup>[13]</sup>根据 V1 区复杂细胞“响应”的统计特性,提出多层次稀疏编码模型,获得了自然图像的轮廓编码.

更多的学者在稀疏编码方面进行了大量的研究<sup>[14-18]</sup>,由于篇幅所限,这里就不一一列举了.在国内,尽管稀疏编码的研究起步较晚,但也取得了不错的成果.李清勇等人<sup>[19]</sup>提出了基于模式分类任务的稀疏编码模型;廖灵芝等人<sup>[20,21]</sup>对稀疏编码中的自适应问题作了大量的研究;结合视觉注意机制,李清勇等人提出基于注意的稀疏编码模型<sup>[22]</sup>.同时,在图像压缩、去噪等信号处理方面,国内的学者也进行了广泛研究<sup>[23-27]</sup>,获得了很好的效果.

上面提到的稀疏编码模型在计算原图像和重构图像的相似度(或者称为信息保持程度)时,广泛采用误差的平方和作为客观评价标准,原因在于其计算复杂度低、物理意义清晰.但是,自 20 世纪 70 年代开始人们就已经发现,误差的平方和不能很好地评价图像的视觉质量,因为它没有充分考虑人眼的视觉特性,这对图像这种最终信宿为人眼的信号而言是非常不合理的.自然图像信号具有高度的结构化,它们的像素点之间具有强烈的相关性,特别是当这些像素点在空间位置临近时,这些相关性携带了重要的视觉物体的结构性信息,而误差的平方和仅仅考虑了图像逐像素点之间的不同,并未考虑到图像像素间的这种强烈相关性和结构对视觉物体的重要性.因此,本文在标准稀疏编码模型的基础上,引入一种新的图像质量评价方法,对公式(2)中的信息保持项进行改进,提出了基于结构相似度的稀疏编码模型.

## 2 结构相似性度量

最近几年,人们对人眼视觉系统的研究有了新的突破,Wang 等人<sup>[28,29]</sup>认为,人眼视觉系统的主要功能是从视觉区域提取图像和视频中的结构化信息,提出了结构相似度(structural similarity,简称 SSIM)的图像质量评价方法,并且通过仿真实验证明了它优于 PSNR(peak signal to noise ratio)方法.如图 3 所示,图 3(a)是原图像,图 3(b)~图 3(f)是图 3(a)失真后的图像,失真的误差平方和都为 210,但是与原图像之间的 SSIM 值各不相同.图 3(b)、图 3(c)的 SSIM 值很高,视觉质量也较好;图 3(d)~图 3(f)的 SSIM 值较低,视觉质量也较差.显然,根据与主观感知的一致性,SSIM 比误差平方和更适合于评价图像的感知质量.

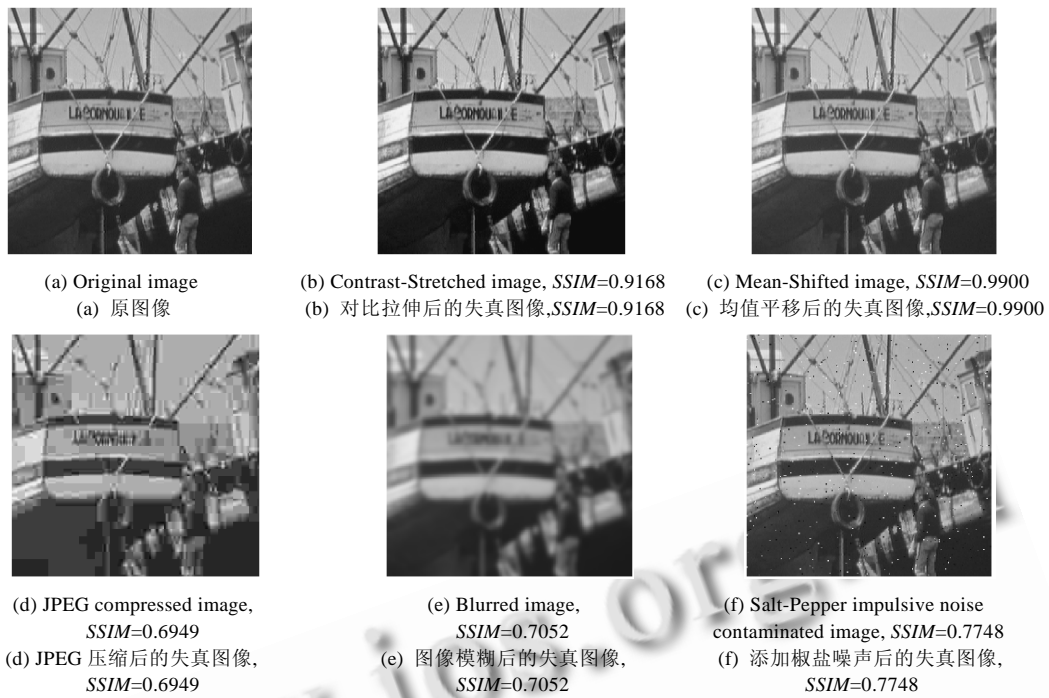


Fig.3 Comparison of "Boat" images with different types of distortions

图3 误差平方和相同而感知质量不一样的 Boat 图像

SSIM 包括亮度变化信息、对比度变化信息和结构变化信息.假设原图像块为  $x$ , 重构图像块为  $y$ , 均为  $N$  维列向量.首先,分别提取原图像块和重构图像块的亮度变化信息,然后提取图像的对比度变化信息,在此基础上再提取图像的结构变化信息,并对以上提取的 3 种变化信息分别进行相似度比较.最后综合比较结果,从而得到一种相似度度量.

为了表述方便,我们使用  $\mu_x$  和  $\mu_y$  分别表示  $x$  和  $y$  的均值:

$$\mu_x = \bar{x} = \frac{1}{N} \sum_{i=1}^N x_i, \quad \mu_y = \bar{y} = \frac{1}{N} \sum_{i=1}^N y_i \quad (3)$$

$\sigma_x$  和  $\sigma_y$  分别表示  $x$  和  $y$  的标准差:

$$\sigma_x = \left( \frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)^2 \right)^{1/2}, \quad \sigma_y = \left( \frac{1}{N-1} \sum_{i=1}^N (y_i - \mu_y)^2 \right)^{1/2} \quad (4)$$

$\sigma_{xy}$  为  $x$  和  $y$  的协方差:

$$\sigma_{xy} = \frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)(y_i - \mu_y) \quad (5)$$

$l(x,y)$  表示  $x$  和  $y$  之间的亮度相关函数:

$$l(x,y) = l(\mu_x, \mu_y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \quad (6)$$

$c(x,y)$  表示  $x$  和  $y$  之间的对比度相关函数:

$$c(x,y) = c(\sigma_x, \sigma_y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \quad (7)$$

$s(x,y)$  表示  $x$  和  $y$  之间的结构相关函数:

$$s(x, y) = s\left(\frac{x - \mu_x}{\sigma_x}, \frac{y - \mu_y}{\sigma_y}\right) = \frac{\sigma_{xy} + C_3}{\sigma_x \sigma_y + C_3} \quad (8)$$

为了避免公式(6)~公式(8)中的分母为0而特设3个极小的正常数  $0 < C_1, C_2, C_3 \ll 1$ .  $x$  和  $y$  之间的结构相似度表示为

$$SSIM(x, y) = f(l(x, y), c(x, y), s(x, y)) = [l(x, y)]^\alpha [c(x, y)]^\beta [s(x, y)]^\gamma \quad (9)$$

这里,参数  $\alpha, \beta, \gamma > 0$  用来调整3部分的相对权重.根据文献[29],我们通常取  $\alpha = \beta = \gamma = 1, C_3 = C_2/2$ , 则结构相似度为

$$SSIM(x, y) = \frac{(2\mu_x \mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (10)$$

结构相似度的值越高,说明原图像块  $x$  与重构图像块  $y$  之间越相似,重构后的结构信息保持程度越好.为了使得  $l(x, y), c(x, y)$  和  $s(x, y)$  介于0和1之间,我们分别对  $\mu_x, \mu_y$  和  $\sigma_{xy}$  取绝对值.

### 3 基于结构相似度的稀疏编码

如前所述,误差的平方和没有充分考虑人眼的视觉特性,所以我们引入结构相似度到标准稀疏编码模型中,使得重构图像块尽量保持原图像块中的结构信息.

为了叙述方便,我们将原图像块按列排列成一个  $N$  维列向量,用  $I$  表示,  $I_i (i=1, \dots, N)$  表示每一个像素点;基函数矩阵用  $A$  表示,大小为  $N \times M$ , 其中,每一个  $N$  维列向量用  $\Phi_k (k=1, \dots, M)$  表示,  $A$  中的每个元素用  $\phi_{i,j} (i=1, \dots, N, j=1, \dots, M)$  表示;用  $M$  维列向量  $S$  表示原图像块对应的“响应”,  $a_i (i=1, \dots, M)$  表示每一个响应值;重构后的图像块用  $N$  维列向量  $Y$  表示,那么  $Y = AS = \sum_{k=1}^M a_k \Phi_k$ ,  $Y_i (i=1, \dots, N)$  表示重构图像块的每一个像素点.我们引入结构相似度到优化的目标函数中,改进后的目标函数如下所示:

$$E(A, S) = \lambda_1 \times \sum_{i=1}^N (I_i - Y_i)^2 + \lambda_2 \times (1 - SSIM(I, Y)) + \lambda_3 \times \sum_{i=1}^M S \left( \frac{a_i}{\sigma_i} \right) \quad (11)$$

其中:  $\lambda_1, \lambda_2, \lambda_3 \geq 0$  分别表示公式(11)右边3个分项的权重系数;  $S(a_i/\sigma_i)$  是一个稀疏惩罚函数,“响应”越稀疏,该函数值越小.第1项和第3项与标准稀疏编码模型是一致的,第1项表示重构图像块与原图像块的误差平方和,第3项表示“响应”的稀疏性.我们同时要求重构图像块尽量保持原图像块中的结构信息,于是添加了第2项约束.当重构图像块与原图像块的结构相似度越大时,第2项的值越接近于0.当误差平方和越小,信息保持程度越高,稀疏性越大时,  $E(A, S)$  就越小.根据有效编码理论,在使得重构图像块与原图像块之间的误差平方和尽量小、结构相似度尽量大的基础上,使得被激活的神经细胞的个数尽可能地少,即“响应”尽量稀疏.因此,稀疏编码的关键就是最优化  $E(A, S)$ .

我们把公式(10)代入到公式(11)中,可以得到:

$$E(A, S) = \lambda_1 \times \sum_{i=1}^N (I_i - Y_i)^2 + \lambda_2 \times \left( 1 - \frac{(2\mu_l \mu_y + C_1)(2\sigma_{ly} + C_2)}{(\mu_l^2 + \mu_y^2 + C_1)(\sigma_l^2 + \sigma_y^2 + C_2)} \right) + \lambda_3 \times \sum_{i=1}^M S \left( \frac{a_i}{\sigma_i} \right) \quad (12)$$

在优化公式(12)时,每个循环分两步进行:第1步固定  $A$ , 采用共轭梯度算法得到  $S$ ; 第2步固定  $S$ , 采用简单梯度算法得到  $A$ . 令  $B_1 = \sum_{i=1}^N (I_i - Y_i)^2$ ,  $B_{21} = 2\mu_l \mu_y + C_1$ ,  $B_{22} = 2\sigma_{ly} + C_2$ ,  $B_{23} = \mu_l^2 + \mu_y^2 + C_1$ ,  $B_{24} = \sigma_l^2 + \sigma_y^2 + C_2$ ,  $B_3 = \sum_{i=1}^M S \left( \frac{a_i}{\sigma_i} \right)$ , 那么梯度计算如公式(13)和公式(14)所示:

$$\vec{\nabla}_{a_i} E(A, S) = \lambda_1 \times \vec{\nabla}_{a_i} B_1 - \lambda_2 \times \frac{B_{21} \times B_{22}}{B_{23} \times B_{24}} \times \left( \frac{\vec{\nabla}_{a_i} B_{21}}{B_{21}} + \frac{\vec{\nabla}_{a_i} B_{22}}{B_{22}} - \frac{\vec{\nabla}_{a_i} B_{23}}{B_{23}} - \frac{\vec{\nabla}_{a_i} B_{24}}{B_{24}} \right) + \lambda_3 \times \vec{\nabla}_{a_i} B_3 \quad (13)$$

$$\vec{\nabla}_{\phi_{i,j}} E(A, S) = \lambda_1 \times \vec{\nabla}_{\phi_{i,j}} B_1 - \lambda_2 \times \frac{B_{21} \times B_{22}}{B_{23} \times B_{24}} \times \left( \frac{\vec{\nabla}_{\phi_{i,j}} B_{21}}{B_{21}} + \frac{\vec{\nabla}_{\phi_{i,j}} B_{22}}{B_{22}} - \frac{\vec{\nabla}_{\phi_{i,j}} B_{23}}{B_{23}} - \frac{\vec{\nabla}_{\phi_{i,j}} B_{24}}{B_{24}} \right) \quad (14)$$

其中,

$$\begin{aligned}\vec{v}_{a_i} B_1 &= (-2) \times \sum_{k=1}^N (I_k - Y_i) \phi_{k,i}, & \vec{v}_{\phi_{i,j}} B_1 &= (-2) \times (I_i - Y_i) \times a_j, \\ \vec{v}_{a_i} B_{21} &= (2/N) \times \mu_I \times \sum_{k=1}^N \phi_{k,i}, & \vec{v}_{\phi_{i,j}} B_{21} &= (2/N) \times \mu_I \times a_j, \\ \vec{v}_{a_i} B_{22} &= (2/(N-1)) \times \sum_{k=1}^N ((I_k - \mu_I) \times \phi_{k,i}), & \vec{v}_{\phi_{i,j}} B_{22} &= (2/(N-1)) \times (I_i - \mu_I) \times a_j, \\ \vec{v}_{a_i} B_{23} &= (2/N) \times \mu_Y \times \sum_{k=1}^N \phi_{k,i}, & \vec{v}_{\phi_{i,j}} B_{23} &= (2/N) \times \mu_Y \times a_j, \\ \vec{v}_{a_i} B_{24} &= (2/(N-1)) \times \sum_{k=1}^N ((Y_k - \mu_Y) \times \phi_{k,i}), & \vec{v}_{\phi_{i,j}} B_{24} &= (2/(N-1)) \times (Y_i - \mu_Y) \times a_j.\end{aligned}$$

因此,通过训练学习获取基函数集合的算法如下所示:

**算法 1.** 获取基函数集合的算法.

Step 1. 对训练数据集进行白化低通滤波预处理;随机初始化基函数矩阵  $A$ ;

Step 2. 固定  $A$ ,使用共轭梯度下降法和公式(13)优化  $E(A,S)$ ,得到“响应” $S$ ;

Step 3. 固定  $S$ ,使用简单梯度下降法和公式(14)优化  $E(A,S)$ ,得到  $A$ ;

Step 4. 若训练次数达到设定值,得到训练好的基函数矩阵  $A$ ;否则,转入 Step 2 继续进行优化.

从网址 <http://www.cis.hut.fi/projects/ica/data/images/> 下载得到在稀疏编码研究领域广泛使用的自然图像集,然后对自然图像随机地选取 250 000 个  $12 \times 12$  图像块,组成训练图像块集合.经过 5 000 次的训练学习后获得 144 个基函数,每个基函数具有与视皮层简单细胞类似的性质,表现为局部性、朝向性和带通性,如图 4 所示.虽然这些基函数与标准稀疏编码模型的结果相类似,但是由于在优化目标函数中引入了结构相似度,使得本文的稀疏编码模型具有更好的结构信息保持能力,更加符合视觉感知特性.第 4 节的实验结果证实了这一点.

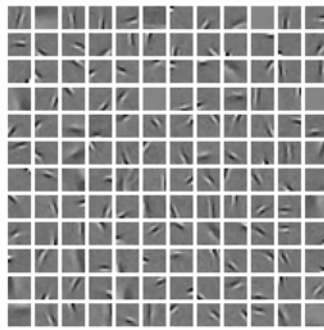


Fig.4 Basis functions of sparse coding model proposed in this paper

图 4 本文稀疏编码模型的基函数

## 4 实验结果

对于某一个图像块,利用已经训练学习好的基函数集合,通过优化后得到相应的“响应”,即稀疏编码.在基函数集合  $A$  已经通过训练取得的前提下,求“响应”的算法如下:

**算法 2.** 求“响应”的算法.

Step 1. 对原始数据进行白化低通滤波预处理;已知训练好的基函数矩阵  $A$ ;

Step 2. 固定  $A$ ,使用共轭梯度下降法和公式(13)优化  $E(A,S)$ ,即可得到“响应” $S$ .

图 5 显示了一个图像块的原始像素值、应用标准稀疏编码模型得到的“响应”和应用本文稀疏编码模型得到的“响应”.无论是应用标准稀疏编码模型得到的“响应”,还是应用本文模型得到的“响应”,都比图像块原始像素值的分布要明显地稀疏得多.同时,两种方法得到的“响应”稀疏程度差不多,分别为 8.433 8 和 8.323 1.图中 SC 表示标准稀疏编码模型,SS\_SC 表示基于结构相似度的稀疏编码模型.

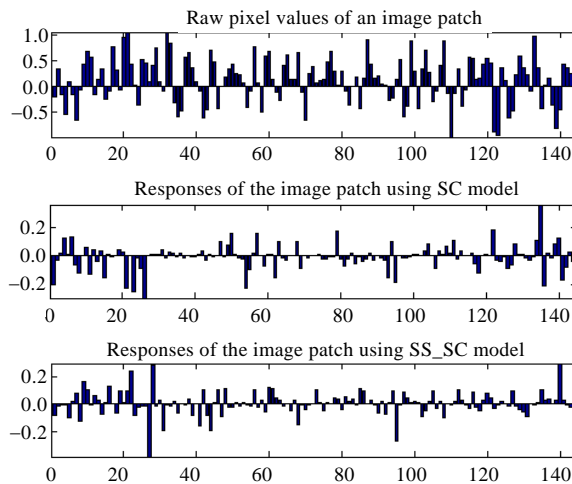


Fig.5 Raw pixel values of an image patch and responses of the image patch using different models

图 5 一个图像块的原始像素值和两种“响应”

本节比较标准稀疏编码模型和本文模型的编码重构性能和响应稀疏性,验证本文模型的编码性能.同时,为了考察不同的图像块数目是否影响各自的性能,我们分别在前面提到的自然图像集中随机地选取不同数目的  $12 \times 12$  图像块组成不同的实验图像块集.实验中,我们选取 4 个图像块数目分别为 100,500,2 000 和 10 000 的实验图像块集.根据各自已经训练好的基函数集,两个模型分别得到对应的“响应”集,再计算出它们的平均结构相似度、平均误差平方和以及平均稀疏性,见表 1(其中, *Avg\_SSim* 表示平均结构相似度, *Avg\_Error* 表示平均误差平方和, *Avg\_Sparseness* 表示平均稀疏性).

Table 1 Comparison of different set of image patches using different models

表 1 两种模型、不同实验图像块集的比较

Experiments	<i>Avg_SSim</i>	<i>Avg_Error</i>	<i>Avg_Sparseness</i>
(SC,100)	0.866 7	0.079 3	8.006 1
(SS_SC,100)	0.931 7	0.067 5	8.192 2
(SC,500)	0.856 3	0.084 9	7.946 7
(SS_SC,500)	0.923 6	0.073 3	8.113 8
(SC,2000)	0.857 4	0.083 6	7.998 8
(SS_SC,2000)	0.924 9	0.072 3	8.152 6
(SC,10000)	0.854 2	0.084 2	8.015 3
(SS_SC,10000)	0.923 1	0.075 3	8.160 0

如表 1 所示,两种模型的性能在不同图像块数目的实验图像块集上基本稳定.标准稀疏编码模型的平均结构相似度在 0.85~0.86 之间,而结构相似度稀疏编码模型的平均结构相似度在 0.92~0.93 之间,大约提高了 8% 左右;标准稀疏编码模型的平均误差平方和在 0.080~0.085 之间,而结构相似度稀疏编码模型的平均误差平方和在 0.070~0.075 之间,大约下降了 12% 左右;标准稀疏编码模型的平均稀疏代价函数值在 8.00 左右,而结构相似度稀疏编码模型的平均稀疏代价函数值在 8.15 左右,增加了 1.8%,变化不大.因此,应用本文的模型与标准稀疏编码模型相比,得到的“响应”稀疏程度基本相似,但是重构图像块与原图像块之间的误差平方和明显地小于标准稀疏编码模型.这说明,在相似的“响应”稀疏程度下,或者说使用基本相同的“响应”个数,基于结构相似度的稀疏编码模型比标准稀疏编码模型更好地保持了原图像块的信息.同时我们还发现,重构图像块与原图像块之间的结构相似度明显大于原来的方法.这说明本文的模型比标准稀疏编码模型更好地保持了原图像块的结构信息,“响应”的结果更加符合生物视觉系统特性.

更为重要的是,应用本文提出的模型,重构图像块与原图像块之间的结构相似度变化范围明显比较小,性能稳定;而原来的标准稀疏编码模型则变化范围比较大,性能不够稳定.对于一定数量图像块的集合,在原来的模型条件下,有些图像块的“响应”经过重构后,重构图像块与原图像块之间的结构相似度很小,重构图像块质量很

差;而使用本文提出的模型,则信息保持程度很高、很稳定.图像块数目为 100 和 200 的情况如图 6 和图 7 所示.

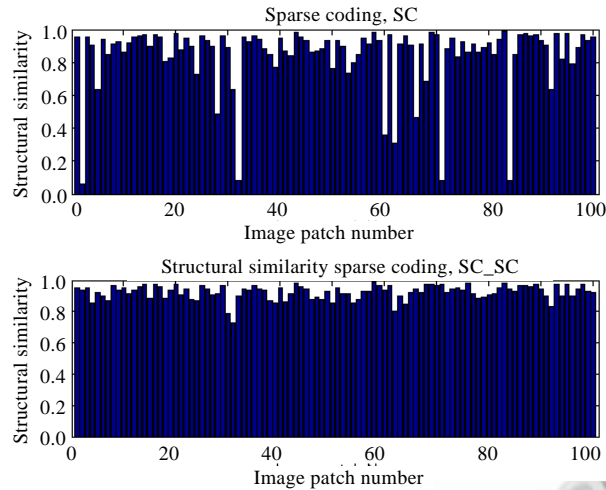


Fig.6 100 pieces of image patches

图 6 图像块数为 100

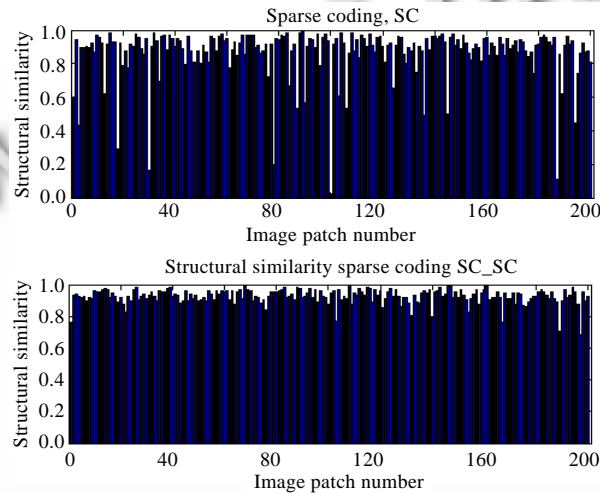


Fig.7 200 pieces of image patches

图 7 图像块数为 200

对一幅自然图像,在提取稀疏编码特征时需要先对图像进行子窗口取样,获得感受野大小的图像块集,然后分别对每个图像块提取稀疏编码特征.另外,为了在重构整幅图像时能够准确地找到每个图像块对应的位置,子窗口取样时需要记录下各图像块在原图像中的位置.由于每个像素可能出现在不同的图像块中,所以在图像重构时需要对每个像素的多个重构值取平均,用平均值作为该位置上的像素灰度值.

图像重构实验中采用了各种类型的自然图像,包括人工建筑的图像、自然环境的图像、结构性强的图像和纹理比例高的图像等.在“响应”稀疏程度基本相同的情况下,分别使用 SC 和 SS\_SC 对原图像提取图像特征,然后进行图像重构.重构图像和原图像部分如图 8 所示(第 1 行为使用标准稀疏编码“响应”后得到的重构图像,第 2 行为使用本文稀疏编码“响应”后得到的重构图像,第 3 行为原

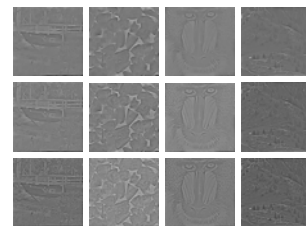


Fig.8 Paradigms of the reconstructed images and original images

图 8 重构图像与原图像示例



图像).显然,使用本文模型的重构图像能够保持原图像中更多的信息,尤其是结构信息,更完整地表示了原图像,这与 Barlow 提出的有效编码假设是完全一致的.

## 5 结 论

稀疏编码利用心理学和神经科学等方面的研究成果,运用信息学理论,在自然环境统计特性与生物(或者机器)视觉系统功能(响应)之间建立了一种数量联系,模拟生物视觉系统神经细胞的编码特性和功能,给设计和研究机器视觉系统提供了理论指导.原有的稀疏编码模型采用误差的平方和作为信息保持的判定标准,没有充分考虑人眼的视觉特性.我们加入结构相似度作为判定标准,更好地反映了视觉系统的特性.通过训练学习后得到的基函数具有与视皮层简单细胞类似的性质,表现为局部性、朝向性和带通性.

通过实验对比发现,引入结构相似度的稀疏编码模型对自然图像的“响应”稀疏程度与标准稀疏编码模型得到的“响应”稀疏程度基本相似,但是应用本文的模型,信息保持程度得到显著的提高.应用本文的模型,重构图像块与原图像块之间的结构相似度明显大于原来的方法,提高了 8%左右;同时,误差平方和下降了 12%左右.更为重要的是,应用本文的模型,重构图像块与原图像块之间的结构相似度的波动范围明显要小于原来的方法.同时,在使用基本相同的“响应”个数的前提下,基于结构相似度的稀疏编码模型比标准稀疏编码模型更好地保持了原图像块的结构信息.因此,本文提出的稀疏编码模型更好地反映了生物视觉系统特性.

## References:

- [1] Marr D. Vision: A Computational Investigation into the Human Representation and Processing of Visual Information. New York: W. H. Freeman and Company, 1982.
- [2] Barlow HB. Possible principles underlying the transformation of sensory messages. In: Rosenblith WA, ed. Proc. of the Sensory Communication. Cambridge: MIT Press, 1961. 217–234.
- [3] Atick JJ, Redlich AN. What does the retina know about natural scenes? *Neural Computation*, 1992,4(2):196–210. [doi: 10.1162/neco.1992.4.2.196]
- [4] Olshausen BA, Field DJ. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 1996,381(6583):607–609. [doi: 10.1038/381607a0]
- [5] Field DJ. What is the goal of sensory coding? *Neural Computation*, 1994,6(44):559–601. [doi: 10.1162/neco.1994.6.4.559]
- [6] Field DJ. Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A: Optics, Image Science, and Vision*, 1987,4(12):2379–2394. [doi:10.1364/JOSAA.4.002379]
- [7] Hyvarinen A, Hoyer PO. A two-layer sparse coding model learns simple and complex cell receptive fields and topography from natural images. *Vision Research*, 2001,41(18):2413–2423. [doi:10.1016/S0042-6989]
- [8] Tenenbaum JB, Freeman WT. Separating style and content with bilinear models. *Neural Computation*, 2000,12(6):1247–1283. [doi: 10.1162/089976600300015349]
- [9] Grimes DB, Rao RPN. Bilinear sparse coding for invariant vision. *Neural Computation*, 2005,17(1):47–73. [doi: 10.1162/0899766052530893]
- [10] Hoyer PO. Modeling receptive fields with non-negative sparse coding. *Neurocomputing*, 2003,52-54(1):547–552. [doi: 10.1016/S0925-2312(02)00782-8]
- [11] Malo J, Epifanio I, Navarro R, Simoncelli EP. Non-Linear image representation for efficient perceptual coding. *IEEE Trans. on Image Processing*, 2006,15(1):68–80. [doi: 10.1109/TIP.2005.860325]
- [12] Schwartz O, Simoncelli EP. Natural signal statistics and sensory gain control. *Nature Neuroscience*, 2001,4(8):819–825. [doi: 10.1038/90526]
- [13] Hoyer PO, Hyvarinen A. A multi-layer sparse coding network learns contour coding from natural images. *Vision Research*, 2002, 42(12):1593–1605. [doi:10.1016/S0042-6989(02)00017-2]
- [14] Teh YW, Welling M, Osindero S, Hinton GE. Energy-Based models for sparse overcomplete representations. *Journal of Machine Learning Research*, 2003,4(9):1235–1260. [doi: 10.1162/jmlr.2003.4.7-8.1235]

- [15] Olshausen BA, Field DJ. Sparse coding of sensory inputs. *Current Opinion in Neurobiology*, 2004,14(4):481–487. [doi: 10.1162/jmlr.2003.4.7-8.1235]
- [16] Doi E, Lewicki MS. Sparse coding of natural images using an overcomplete set of limited capacity units. In: Saul L, Weiss Y, Bottou L, eds. *Advances in Neural Information Processing Systems 17*. Cambridge: The MIT Press, 2005. 377–384.
- [17] Lee H, Battle A, Raina R, Ng AY. Efficient sparse coding algorithms. In: Scholkopf B, Platt J, Hofmann T, eds. *Advances in Neural Information Processing Systems 19*. Cambridge: The MIT Press, 2007. 801–808.
- [18] Hyvarinen A, Koster U. Complex cell pooling and the statistics of natural images. *Network: Computation in Neural Systems*, 2007, 18(2):81–100. [doi: 10.1080/09548980701418942]
- [19] Li QY, Lin DC, Shi ZZ. Task-Oriented sparse coding model for pattern classification. In: Wang L, Chen K, Ong YS, eds. *Proc. of the ICNC 2005*. LNCS 3610, Heidelberg: Springer-Verlag, 2005. 903–914.
- [20] Liao LZ, Luo SW, Tian M. “Whitenedfaces” recognition with PCA and ICA. *IEEE Signal Processing Letters*, 2007,14(12): 1008–1011. [doi:10.1109/LSP.2007.904704]
- [21] Liao LZ, Luo SW, Tian M, Zhao LW. Fast and adaptive low-pass whitening filters for natural images. In: King I, *et al.*, *Proc. of the ICONIP 2006, Part II*. LNCS 4233, Heidelberg: Springer-Verlag, 2006. 343–352.
- [22] Li QY, Shi J, Shi ZZ. A model of attention-guided visual sparse coding. In: Kinsner W, Zhang D, Wang Y, Tsai J, eds. *Proc. of the 4th IEEE Int’l Conf. on Cognitive Informatics*. Irvine: IEEE Computer Society, 2005. 120–125.
- [23] Sun J, Wang WY, Zhuo Q. Sparse coding-based global face feature extraction algorithm. *Journal of Tsinghua University (Science and Technology)*, 2002,42(3):411–413 (in Chinese with English abstract).
- [24] Liu WX, Zheng NN. Learning sparse features for classification by mixture models. *Pattern Recognition Letters*, 2004,25(2): 155–161. [doi: 10.1016/j.patrec.2003.09.007]
- [25] Shang L, Huang DS, Zheng CH, Sun ZL. Noise removal using a novel non-negative sparse coding shrinkage technique. *Neurocomputing*, 2006,69(7-9):874–877. [doi: 10.1016/j.neucom.2005.07.004]
- [26] Liu WX, Zheng NN, You QB. Nonnegative matrix factorization and its applications in pattern recognition. *Chinese Science Bulletin*. 2006,51(1):7–18. [doi: 10.1007/s11434-005-1109-6]
- [27] Shang L, Cao FW, Zhao ZQ, Chen J, Zhang Y. Palmprint recognition using a novel sparse coding technique. In: Liu D, *et al.*, *Proc. of the ISNN 2007, Part II*. LNCS 4492, Heidelberg: Springer-Verlag, 2007. 810–818.
- [28] Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. on Image Processing*, 2004,13(4):600–612. [doi: 10.1109/TIP.2003.819861]
- [29] Wang Z, Simoncelli EP. Stimulus synthesis for efficient evaluation and refinement of perceptual image quality metrics. In: Rogowitz BE, Pappas TN, eds. *Proc. of the Society of Photo-Optical Instrumentation Engineers Conf. on Human Vision and Electronic Imaging*. Bellingham: SPIE, 2004. 99–108.

#### 附中文参考文献:

- [23] 孙俊,王文渊,卓晴.基于稀疏编码的提取人脸整体特征算法.清华大学学报(自然科学版),2002,42(3):411–413.



李志清(1975—),男,湖南娄底人,博士生,讲师,CCF 学生会会员,主要研究领域为计算机视觉,机器学习,人工智能.



李志欣(1971—),男,博士生,讲师,主要研究领域为基于内容的视觉信息检索,图像理解,机器学习.



施智平(1974—),男,博士,助理研究员,CCF 会员,主要研究领域为机器学习,图像理解,视觉信息挖掘.



史忠植(1941—),男,研究员,博士生导师,CCF 高级会员,主要研究领域为人工智能,机器学习,神经计算,认知科学.