

基于深度学习的形状建模方法

张 娟 汪西莉 杨建功

(陕西师范大学计算机科学学院 西安 710119)

摘 要 在目标检测、图像分割、图像修复等应用中,一个好的目标形状模型有着非常重要的作用.与灰度、纹理、边缘等底层的视觉特征相比,作为一种高层视觉特征的形状更有利于描述目标的全局视觉信息.一般在加入全局形状信息的情况下算法的性能将有所提高.近年来,由于其具有卓越的学习数据中包含的内部结构的能力,深度学习模型吸引了越来越多的关注.包含多层隐层单元的深度学习模型用于模拟人类大脑的认知机制,可以提取多层数据特征和表示复杂的数据分布,学习到的多层抽象表示更有助于了解形状等数据的信息.针对复杂而又多变的目标形状,该文基于深度信念网络和深度玻尔兹曼机构造形状模型,给出了各自的模型构造方法、模型训练方法以及基于模型生成目标形状的方法.这两种模型是深层概率模型,能够充分利用底层特征和多层高层特征(越高层越复杂)拟合关于训练集形状的概率分布,所以它们能够有效地建模形状.这类模型的一个关键特性是使用贪心逐层训练快速地找到一组好的模型参数.该文的形状模型不仅能够很好地表达出训练集中的形状,还能够生成不同于训练集中样本的形状.另外,当训练集中的形状是多类别时,此时涉及到的形状变化比较大,文中模型同样能够定义形状所属的多峰分布.深度学习模型具有较强的形状表达能力,可以应用于生成形状、形状修复和去噪等多种任务.该文在 Weizmann Horse 和 Caltech101 Silhouettes 数据集上进行了实验,结果表明,和浅层的受限玻尔兹曼机模型相比,深度学习模型能够更好地表达训练集形状,从包含多层非线性处理的深层结构中生成的形状看起来更符合实际,且深度玻尔兹曼机模型生成的目标形状要比深度信念网络模型更为清晰.实验中以形状图形化结果和形状相似性度量结果说明深度学习模型对训练形状的拟合能力较强,且具有较好的泛化能力.

关键词 形状建模;深度学习;受限玻尔兹曼机;深度信念网络;深度玻尔兹曼机

中图法分类号 TP18 **DOI号** 10.11897/SP.J.1016.2018.00132

Shape Modeling Method Based on Deep Learning

ZHANG Juan WANG Xi-Li YANG Jian-Gong

(School of Computer Science, Shaanxi Normal University, Xi'an 710119)

Abstract Models of the object shape play an important role in many image applications such as object detection, image segmentation and inpainting. Compared with the gray, texture, edge and other low-level visual features of the image, shape as high-level visual feature helps to describe the global visual information of an object better. Generally, the performance of algorithm will be improved after global shape information (expressed by a good model) joined. In recent years, deep learning models have attracted more and more attentions because of their excellent learning ability on modeling data with intrinsic structure. Deep learning models contain multiple layers and hidden perception units, which are used to simulate the cognitive mechanism of human brain. They can extract multi-level data features and represent complicated data distributions. They form high-level data abstract representation through combining the lower layer features. A deep

收稿日期:2016-06-15;在线出版日期:2016-11-21. 本课题得到国家自然科学基金(41171338,41471280,61401265)资助. 张娟,女,1992年生,硕士研究生,主要研究方向为机器学习、图像处理. E-mail: chenmzhang@foxmail.com. 汪西莉(通信作者),女,1969年生,博士,教授,博士生导师,中国计算机学会(CCF)会员,主要研究领域为智能信息处理、模式识别、图像处理. E-mail: wangxili@snnu.edu.cn. 杨建功,男,1974年生,博士研究生,讲师,主要研究方向为机器学习、图像处理.

learning model learning the representation and abstraction of multiple levels is more helpful to make sense of data, such as shape images. In view of the complicated and changeable object shape, we propose two kinds of shape models based on deep belief network and deep Boltzmann machine, and the methods of two models constructing, training and shape generating in this paper. Both of these two models are deep, hierarchical probabilistic models that make heavy use of hidden units to modal high-order dependencies between the observed units, which can make full use of low-level features and multi-layer high-level features of object shape to describe the probability distribution of training set. They have the potential of learning internal representation that become progressively complex at higher layers. The key feature of these two models is greedy layer-by-layer training that can be repeated several times in order to find a good set of modal parameters quickly, even for modal contains many layers and parameters. Thus they can effectively model the shape, this kind of model not only can express the shape of the training set, but also can generate shape that differ from samples in the training set. The shape change is relatively large when the shape of the training set is multiple categories, the modal can also define the multi-model distribution of the shape set. A deep learning model has strong ability of shape expression and can be applied in many kinds of tasks such as shape generating, shape completion, and denoising. We give some of the experimental results on the Weizmann Horse data set and Caltech101 Silhouettes data set. The results show that compared with the restricted Boltzmann machine model, the proposed two models based on deep learning can express the shape of training set better. The samples generated from the deep models involving structure of multi-layer nonlinear processing are more realistic, and the samples from the deep Boltzmann machine model are much more clear than deep belief network model. Experimental results in the graphical shape and similarity measure illustrate that deep learning model is fit for the training shape, and the model has better generalization ability.

Keywords shape modeling; deep learning; restricted Boltzmann machine; deep belief network; deep Boltzmann machine

1 引言

在目标检测、图像分割和图像修复等应用^[1-2]中,目标的形状信息起到至关重要的作用,形状作为一种高层的视觉特征,与图像中的灰度、纹理、边缘等底层视觉特征相比,它描述的视觉信息有利于完整地表达目标物体。若图像中存在噪声、背景和特征比较相似、部分目标被遮挡等情况,形状作为先验知识有助于得到较为完整且正确的目标,从而提高目标检测或分割等的正确率。此外,形状在图像的生成过程中也扮演着重要的角色^[3-4]。一般来说,上述这些应用的效果能够通过使用更好的目标形状模型得以改善。

目前,已经有很多方法来表示二维形状。用标记

点集表示形状是一种最简单的方法,它基于标记点分布模型,先选取一批已标记的训练形状,用形状边缘位置、角点等一些关键点作为标记点描述形状,然后使用主成分分析法等统计分析获得全局的形状模型^[5],但其形状表示过于被动,需要对训练集中所有的形状进行标注,要求从形状中选出的主要特征点有相同的数量和位置,并且在求解过程中容易陷入局部极值。用来表示形状属性常用的模型还有网格结构的马尔科夫随机场或条件随机场^[6],网格结构的马尔科夫随机场模型可以表示为由一元项和二元项组成的能量函数形式,其中定义的二元能量项表示相邻像素之间的依赖关系,也即是施加给相邻像素之间局部区域限制强调形状的局部特征(平滑性或连通性),但无法捕捉到凸性或曲率等更复杂的形状属性,同时网格结构的马尔科夫随机场在表达全

局形状时受到一定的限制. 另外一种方法是使用水平集^[7]表示形状, 其基本思想是施加足够的约束使形状是有效的、真实的, 同时允许捕获的可能的形状要具有一定的灵活性. 在水平集方法中采用符号距离函数表示形状, 在一个区域上确定一个点到区域边界的距离并对其符号进行定义, 点位于边界内部为正, 位于边界外部为负, 位于边界上为 0, 这种方法捕获目标形状的全局特征和关于目标形状的全局变化(比如形状纵横比的变化). 但是这类方法的缺点是在曲线的演化过程中, 符号距离函数会发生震荡, 导致最终的曲线演化形状偏离真实的目标形状轮廓. 另外, 当训练集中存在不同种类时, 这种方法不能有效地为形状建模, 即不能拟合多峰分布, 这使得形状只能在一定范围内变化.

近年来备受瞩目的深度学习模型^[8]能够通过组合底层的特征形成高层的数据抽象特征进而模拟人类大脑认知机制, 且能够自动发现隐藏在数据中的模式, 具有强大的学习数据中包含的结构的能力, 可以使用这类模型来学习形状复杂的全局特征进而表征形状. 玻尔兹曼机(Boltzmann Machine, BM)^[9]是一种根植于统计力学的随机神经网络模型, 基于能量函数建模, 其样本的分布服从玻尔兹曼分布. 模型由可见层和隐层组成, 可见层由可见单元组成, 隐层由隐单元组成, 其单元之间是全连接的, 具有强大的无监督学习能力, 能够从复杂数据中学习隐藏在数据中的高阶特征, 但是其学习时间很长, 且不能够精确计算出 BM 所表示的分布, 同时从模型中得到服从其定义的分布的样本也很困难. Smolensky^[10]提出受限的玻尔兹曼机(Restricted Boltzmann Machine, RBM)模型, 模型包括一个可见层和一个隐层, 它是 BM 的一种特殊结构. 相比于 BM, RBM 模型去除了同层单元之间的连接, 即可见层与隐层单元之间是全连接, 可见层和隐层的层内单元之间无连接, 这使得高效快速的模型训练算法的出现成为可能, 有效地缩短了训练时间. 尽管仍然无法有效计算 RBM 所表示的分布, 但可以通过采样方法得到服从模型所表示分布的样本. 但是在训练样本和计算单元有限的情况下, 两层结构的 RBM 模型对复杂分布的表示能力有限. 2006 年 Hinton 等人^[11]提出深度信念网络(Deep Belief Network, DBN)模型, 通过构建具有多个隐层的结构来学习更有用的高层特征. 该模型是一种生成模型, 包括一个可见层和多个隐层, 模型可以使用极大似然估计方法来训练得到参数.

DBN 模型的最上面两层单元之间的连接是无向的, 构成一个无向二部图, 较低的其它层单元之间的连接是有向的, 构成一个有向的 sigmoid 置信网, 其较高层单元表示较低层单元之间的高阶关系. 全局优化 DBN 往往是困难的、耗时的, 为加快模型的训练速度和获得更好的优化性能, Hinton 给出了模型训练的一个高效算法——在全局训练整个模型之前, 先使用逐层训练算法用于预训练模型, 即先训练较低层模块后训练较高层模块, 进行的是无监督训练. 文献[12]表明添加隐层在严格意义上能够提高模型的建模能力, 即深层模型相比于浅层模型有更好的非线性函数表示能力. 深度玻尔兹曼机(Deep Boltzmann Machine, DBM)模型是 Salakhutdinov 等人^[13]于 2009 年提出的另外一种深度学习模型, 模型包括一个可见层和多个隐层, 可以看作由多个 RBM 堆叠而成. 该模型与 DBN 模型一样能够学习到数据的多层次抽象特征, 但 DBM 与 DBN 不同的是 DBM 是一个无向图模型, 即层与层之间的连接均是无向的. 由于 DBM 模型是深层结构, 全局优化 DBM 模型也是很困难的, 所以 Salakhutdinov 等人提出了类似于 DBN 模型的训练算法, 将训练分成了两个阶段: 第 1 个阶段是预训练阶段, 分别训练各个子模块; 第 2 个阶段是微调阶段, 即在预训练的基础上全局训练整个模型. 同样, 模型的训练也是无监督的. 目前已有研究将 RBM 作为全局形状先验成功地应用于图像分割问题, 文献[14]将 RBM 模型作为形状模型添加到条件随机场图像标记模型中提供全局的形状约束, 这种混合的模型比缺少形状先验信息的传统模型减少了错误标记像素的个数, 提高了图像标记的正确率. 文献[15]是结合 DBM 的一种变形模型^[16]对先验形状建模方法提出的一种变分分割模型, 该方法是利用深度学习方法解决先验形状的估计并作为先验形状加入到变分分割模型中用以解决自然图像的分割问题, 形状模型由一个可见层和两个隐层组成, 可见层与第 1 个隐层之间是局部连接, 第 1 个隐层与第 2 个隐层之间是全连接, 但是由于模型可见层与隐层间的局部连接, 在使用模型生成形状时目标形状的图像会出现接缝.

针对复杂而又多变的目标形状, 本文提出两种基于深度学习的形状表示模型, 它们是由 RBM 模型扩展得到的 DBN 和 DBM 模型, 模型各层之间均是全连接, 这两种模型通过训练样本训练得到模型参数, 用以表示形状和生成形状. 文中根据不同模型

的结构特点,给出了各自的模型构造方法、模型的训练方法,以及当用该类模型为形状进行建模时提出了基于模型生成目标形状的方法. 与传统的形状模型相比,这类模型不仅能够很好的表达出训练集中的形状,还能够生成不同于训练集中样本的形状,且形状符合实际. 当训练集中的形状是多类别时,此时涉及到的形状变化比较大,该类模型同样能够定义形状所属的多峰分布. 这样,在图像检测、图像分割等应用中,即使在图像存在遮挡或噪声等情况下,形状的全局特性也会使得目标结果符合实际. 本文在 Weizmann Horse 和 Caltech101 Silhouettes 数据集上进行了实验,结果表明深度学习模型具有较强的形状表达能力,可以非常灵活地应用于生成形状、完成部分缺失或被遮挡形状的修复、噪声情况下的去噪等多种任务.

2 深度学习模型理论

2.1 受限玻尔兹曼机模型

RBM 是包含一个可见层和一个隐层的两层结构且对称连接的概率图模型,如图 1 所示,层间全连接、层内无连接. 可见层包括 m 个可见单元 $v = (v_1, v_2, \dots, v_m)$, 表示观测数据,其中 v_i 表示第 i 个可见单元的状态;隐层包括 n 个隐单元 $h = (h_1, h_2, \dots, h_n)$, 可看作是从可见层数据提取的抽象特征,其中 h_j 表示第 j 个隐单元的状态.

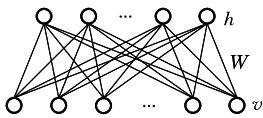


图 1 RBM 模型结构

给定一组状态 (v, h) , RBM 模型定义的联合概率分布为

$$P(v, h; \theta) = \frac{e^{-E(v, h; \theta)}}{Z(\theta)}, \quad Z(\theta) = \sum_{v, h} e^{-E(v, h; \theta)} \quad (1)$$

其中: $Z(\theta)$ 称为划分函数或者归一化常量; $E(v, h; \theta)$ 是状态 (v, h) 所对应的能量函数, 定义为

$$\begin{aligned} E(v, h; \theta) &= -\mathbf{v}^T \mathbf{W} \mathbf{h} - \mathbf{a}^T \mathbf{v} - \mathbf{b}^T \mathbf{h} \\ &= -\sum_{i=1}^m \sum_{j=1}^n w_{ij} v_i h_j - \sum_{i=1}^m a_i v_i - \sum_{j=1}^n b_j h_j \end{aligned} \quad (2)$$

其中: $\theta = \{W, a, b\}$ 是模型参数, 对于所有的 $i \in \{1, \dots, m\}$ 和 $j \in \{1, \dots, n\}$, w_{ij} 是可见单元 v_i 和隐单

元 h_j 间连接边的权重, a_i 是第 i 个可见单元的偏置, b_j 是第 j 个隐单元的偏置. 关于可见单元 v 的概率分布可通过求 $P(v, h; \theta)$ 的边缘概率分布得到.

训练 RBM^[17] 模型的任务是求出模型参数的值以拟合训练数据. 给定一组训练数据 $S = \{x_1, \dots, x_k, \dots, x_l\}$, $k = 1, \dots, l$, 假定这些训练数据是独立同分布的, 常用的方法是使用最大似然估计方法估计模型参数, 即

$$\begin{aligned} \theta^* &= \arg \max_{\theta} \ln(L(\theta; S)) \\ &= \arg \max_{\theta} \ln \prod_{k=1}^l p(x_k; \theta) \\ &= \arg \max_{\theta} \sum_{k=1}^l \ln p(x_k; \theta) \end{aligned} \quad (3)$$

为获得较好的模型参数值, 使用随机梯度上升法求 RBM 在训练集上的对数似然函数 $\ln(L(\theta; S))$ 取最大值时对应的参数. 其参数更新公式为

$$W = W + \Delta W, \quad \Delta W = \langle v_i h_j \rangle_{\text{data}} - \langle v_i h_j \rangle_{\text{model}} \quad (4)$$

$$a = a + \Delta a, \quad \Delta a = \langle v_i \rangle_{\text{data}} - \langle v_i \rangle_{\text{model}} \quad (5)$$

$$b = b + \Delta b, \quad \Delta b = \langle h_j \rangle_{\text{data}} - \langle h_j \rangle_{\text{model}} \quad (6)$$

其中: $\langle \cdot \rangle_{\text{data}}$ 表示经验分布下的期望; $\langle \cdot \rangle_{\text{model}}$ 表示 RBM 模型所定义分布下的期望. 精确的最大似然估计学习是很困难的, 因为想要精确地计算第 2 项期望的时间复杂度是 $O(2^{\min(m, n)})$. 所以, Hinton^[18] 于 2002 年提出了一个快速学习算法——对比散度 (Contrastive Divergence, CD) 算法, 该算法已成为训练 RBM 模型的一个标准算法.

由于 RBM 模型的结构是层间全连接、层内无连接, 这也就意味着在给定可见变量状态的情况下隐变量是相互独立的, 同时在给定隐变量状态的情况下可见变量是相互独立的. 尽管 RBM 模型所定义的分分布仍然无法精确计算, 但通过从边缘概率分布中采样可以得到服从 RBM 模型所表示分布的样本.

2.2 深度信念网络模型

DBN 是包含一个可见层和多个隐层的概率图模型, 可见层由可见单元组成, 隐层由隐单元组成, 每一个隐层都对应捕获其下面一层特征的更高阶的抽象特征, 能够更好地反映输入数据的内在结构信息. DBN 是由两种不同的网络混合而成, 最上面两层形成一个无向网, 其它较低层形成一个有向的 sigmoid 信念网络, 构成一个概率生成模型, 如图 2 (a) 所示, 模型参数为单元之间连接的权重以及各单元所对应的偏置. 最底层表示观测数据, 即模型的输

入. 预训练时 DBN 模型可以被视为由多个简单的无监督模块(即 RBM)组成, 其中每个子模块的隐层可以作为下一个子模块的可见层输入, 如图 2(b) 所示. 这种结构特点也促使了一种快速的、逐层无监督训练算法的提出, 有效地避免了传统的深层网络训练时导致的时间复杂度过高问题, 每个子模型的训练使用 CD 算法. 相比于只有一个隐层的 RBM 模型而言, 具有多个隐层的 DBN 模型能够获取更多高层的抽象特征来描述输入数据, 另外理论结果表明深层结构模型在表示训练数据所服从复杂分布的函数时要比浅层模型更有效^[19].

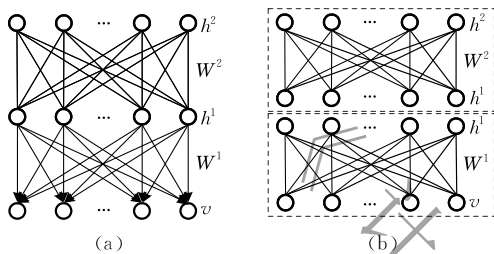


图 2 DBN 模型结构及其对应的预训练模块

训练一个包含一层可见层单元与两层隐层单元的 DBN 模型, 由模型定义的联合概率分布为

$$P(v, h^1, h^2; \theta) = P(v | h^1; \theta^1) P(h^1, h^2; \theta^2) \quad (7)$$

其中: $\theta = \{\theta^1, \theta^2\}$, θ^1 是第 1 层 RBM 的参数, θ^2 是第 2 层 RBM 的参数; v 是可见单元, $P(v | h^1; \theta^1)$ 表示在给定隐单元状态的条件上获得可见单元状态的条件分布, $P(h^1, h^2; \theta^2)$ 是模型最顶层 RBM 的联合分布. 条件分布 $P(v | h^1; \theta^1)$ 和联合分布 $P(h^1, h^2; \theta^2)$ 共同构成生成模型 DBN. 由于同层单元间没有连接, 即层内单元之间是相互独立的, 所以有:

$$P(v | h^1; \theta^1) = \prod_i p(v_i | h^1; \theta^1) \quad (8)$$

$$P(h^1, h^2; \theta^2) = \frac{1}{Z(\theta^2)} e^{-E(h^1, h^2; \theta^2)} \quad (9)$$

其中 $E(h^1, h^2; \theta^2)$ 是顶层 RBM 定义的能量函数.

Hinton^[11] 于 2006 年提出了 DBN 模型的一种快速、无监督的训练算法, 其训练过程分为两个阶段: 预训练阶段和微调阶段. 在预训练阶段的一个关键特性是逐层训练, 将 DBN 模型看作由多个 RBM 模块堆叠而成, 通过自下而上的顺序训练各组成模块来学习深层模型. 同时, 在模型的训练过程中该算法也为执行近似推理提供了一个有效的方式, 它利用自下而上的传递推断出较高层隐单元的状态. 其预训练步骤为:

(1) 以训练样本作为可见层数据 v , 随机初始化

模型参数 θ^1 . 采用 2.1 节中介绍的 RBM 训练方式训练第一个 RBM.

(2) 固定第一个 RBM 的模型参数 θ^1 不变, 将上步中训练得到的隐层单元状态 h^1 作为下一个 RBM 的可见层数据, 随机初始化模型参数 θ^2 . 构造较高一层的 RBM 并训练该 RBM.

(3) 将训练得到的各 RBM 模块的参数赋值给 DBN 模型.

用一次只训练一个模块的方式训练 DBN 模型是有效的, 但不是最优的. 所以, 在微调阶段使用基于梯度算法优化策略的对比 wake-sleep 算法^[20] 对整个模型的参数进行全局调优, 得到最终的模型参数, 以预训练得到的模型参数作为该阶段的初始值. DBN 的子模型除了最顶层的 RBM 外, 其它 RBM 模块的权重分为向上的认知权重和向下的生成权重, 向下的生成权重表示 DBN 是一个生成模型, 向上的认知权重不属于模型的一部分, 只用来执行自底向上的近似推断. 在 wake 阶段, 通过底层的特征和认知权重确定每一层的抽象特征表示, 使用梯度上升法更新生成权重; 在 sleep 阶段, 通过高层特征和生成权重产生底层状态, 同时更新认知权重. 本文使用上述算法训练 DBN 模型.

Hinton 提出的快速无监督训练算法通过逐层训练各模块能够有效地加快深层结构模型的训练且得到较好的模型参数值. 许多现有的机器学习算法使用浅层结构(如支持向量机、RBM 等), 但是理论结果表明, 使用这样的模型不能有效地从输入数据中提取复杂的结构信息, 而 DBN 等深层结构模型结合有效且快速的训练算法能够利用大量的无标签数据提取更多高层的抽象特征, 更好地描述隐藏在输入数据内部的结构信息.

2.3 深度玻尔兹曼机模型

DBM 是不同于 DBN 的另外一种深度学习模型, 可以看作是 RBM 模型的扩展, 包含多个隐层. 它是一个无向图, 所有层与层之间的连接都是无向的, 如图 3(a) 所示. 由于 DBN 模型的结构特点, 训

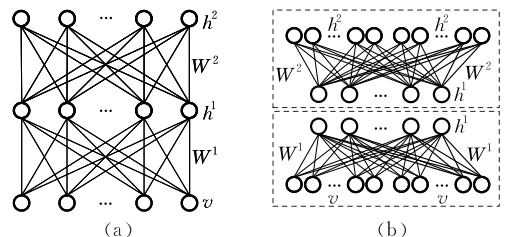


图 3 DBM 模型结构及其对应的预训练模块

练时它是基于一种自下而上的近似推理方法,在推断过程中忽略自上而下反馈的结果,因而不能充分考虑数据的不确定性,而 DBM 能够结合自下而上的传递和自上而下的反馈来解决不确定性问题,从而能够生成更好地依赖于数据和高阶特征的样本。

训练一个具有两个隐层的 DBM, v 是可见层单元的状态, h^1 和 h^2 是各个隐层对应的隐层单元,关于 (v, h^1, h^2) 对应的能量定义为

$$E(v, h^1, h^2; \theta) = -v^T W^1 h^1 - h^{1T} W^2 h^2 - a^T v - b^T h^1 - c^T h^2 \quad (10)$$

其中: $\theta = \{W^1, W^2, a, b, c\}$ 是模型参数, W^1 是可见层 v 与隐层 h^1 之间连接的权重, W^2 是隐层 h^1 与隐层 h^2 之间连接的权重, a, b, c 分别是可见层 v 、隐层 h^1 、隐层 h^2 所对应的偏置。那么模型所定义的关于 v 的边缘概率分布为

$$P(v; \theta) = \frac{1}{Z(\theta)} \sum_{h^1, h^2} e^{-E(v, h^1, h^2; \theta)} \quad (11)$$

这里 $Z(\theta)$ 是一个常数, 定义为

$$Z(\theta) = \sum_v \sum_{h^1, h^2} e^{-E(v, h^1, h^2; \theta)} \quad (12)$$

给定一批训练数据, DBM 的学习就是确定式(10)中的权重矩阵和偏置。本文使用的 DBM 模型的训练方法是 Salakhutdinov 等人^[21] 根据训练 DBN 模型的逐层预训练算法思想提出的训练算法, 其训练过程也包括预训练阶段和微调阶段, 预训练阶段为在微调阶段训练整个模型参数提供一个较好的初始训练值, 但不同的是预训练时的各模块不同、微调时采用的算法不同。使用逐层预训练算法训练 DBM 的思想是训练一批结构上有微小改变的 RBM 模型, 然后将其堆叠成一个 DBM 模型, 其结构如图 3(b) 所示。对最底层的 RBM 模型的改动是用增加一倍自下而上的传递弥补缺少的自上而下的反馈确定隐层单元的状态, 即可见层单元的个数翻了一番。相反, 对最顶层 RBM 模型的改动是用增加一倍自上而下的反馈弥补缺少的自下而上的传递确定可见层单元的状态, 即隐层单元的个数翻了一番。训练得到的参数为 DBM 模型提供一个合理的参数值, 便于后续的联合训练整个模型。在微调阶段, 使用预训练阶段得到的各模块参数作为训练整个 DBM 模型时的初始值, 即使用第一个训练的 RBM 参数初始化 DBM 模型的较低层参数 (W^1 和 a), 使用第 2 个训练的 RBM 模型参数作为模型的较高层参数的初始值 (W^2 和 c)。其后, 使用结合均值场推理和马尔

科夫链蒙特卡罗采样的随机梯度上升法可以得到关于 DBM 模型参数的一个很好的近似, 从而更好地拟合训练数据。其预训练步骤为:

(1) 采用 2.1 节介绍的 RBM 训练方法训练结构经过改动的底层 RBM 模型, 训练样本作为模型的可见层数据。连接权重为 W^1 。

(2) 固定底层 RBM 模型的参数不变, 并使用从 $P(h^1 | v; 2W^1)$ 中采样得到的数据作为上层 RBM 模型的训练数据, 训练上层 RBM 模型。

(3) 将上述训练得到的各模块参数赋值给 DBM 模型, 作为微调阶段时的模型参数初始值。

目前 DBM 模型受到关注有很多原因。首先, 它保留了 DBN 模型很多好的性质: 能够得到输入数据的多种抽象表示; 具有一个高效的贪心逐层训练的预训练算法; 使用无标签数据训练, 是无监督的。其次, 它又与 DBN 等深层结构模型不同, DBM 的近似推理过程包括一个自下而上的传递和自上而下的反馈, 使得 DBM 能够更好地应对数据的不确定性问题, 这也是不同于 DBN 模型与自动编码器等深层模型的最大特点。最后, 通过模型能量函数使用随机梯度上升法联合优化所有层的参数, 这极大地促进了训练得到更好的生成模型。

3 基于 DBN 和 DBM 的形状模型

DBN、DBM 深度学习模型能够获取输入数据的高阶特征, 可以用来为形状建模, 并用于表示和生成形状。给定一批训练形状, 用其训练上述模型得到模型参数。模型训练时以形状像素点的二值特征作为输入, 每个像素点为可见层的一个可见单元, 其值是 0 或 1。在确定模型参数之后, 如何从模型中生成形状是一个重要的问题, 通常的方法是使用采样得到服从模型所定义分布的形状样本。MCMC 方法^[22] 是由 Metropolis 于 1954 年提出的从一个特定概率分布采样的算法, 最早应用于物理领域, 其理论基础是在满足平衡方程的条件下可以通过状态转移到达稳态。后经 Hastings^[23] 改进得到 M-H 算法, M-H 算法是 MCMC 的基础方法。MCMC 方法中最常用的是 Gibbs 采样^[24], 它是 M-H 算法的一个特例。Gibbs 采样可以应用于从变量的联合概率分布和边缘概率分布中采样, 其基本思想是对高维或复杂总体采样时, 构造一个不可约的马尔科夫链, 得到的样本可以近似的作为模型所定义的概率分布的样

本, Gibbs 采样的步骤为:

(1) 随机初始化 $x^{(0)} = \{x_i^{(0)}; i=1, \dots, t\}$

(2) 对 $j=0, 1, 2, \dots$ 依次采样

$$x_1^{(j+1)} \sim p(x_1 | x_2^{(j)}, x_3^{(j)}, \dots, x_t^{(j)})$$

$$x_2^{(j+1)} \sim p(x_2^{(j+1)} | x_3^{(j)}, \dots, x_t^{(j)})$$

...

$$x_t^{(j+1)} \sim p(x_t | x_1^{(j+1)}, x_2^{(j+1)}, x_3^{(j+1)}, \dots, x_{t-1}^{(j+1)})$$

按照上面的步骤, 经过多步迭代采样直至收敛, 则可以得到 $x^{(1)}, x^{(2)}, \dots$, 即近似服从概率分布 $p(x_1, x_2, \dots, x_t)$ 的样本。

以形状作为训练数据, 在训练好模型之后, 可以执行多步 Gibbs 采样得到服从模型定义分布的形状。图 4 是 DBN、DBM 与 RBM 模型的采样过程对比。在 RBM 模型中执行 Gibbs 采样的过程是用一个形状作为可见层的初始化状态, 交替进行采样, 得到近似服从 RBM 模型所定义分布的形状, 如图 4(a) 所示。为生成近似服从 DBN 模型定义分布的形状, 通过在模型的最顶层 RBM 模块执行多步 Gibbs 采样得到第 $L-1$ 层 (L 是 DBN 模型的层数) 的状态, 然后通过自上而下的 sigmoid 置信网络依次随机激活每层单元, 从而得到可见层状态 (即样本), 如图 4(b) 所示。DBM 模型的采样过程如图 4(c) 所示, 依次采样得到可见层、隐层的状态, 并且隐层 h^1 条件依赖于可见层 v 和隐层 h^2 。从 DBN、DBM 采样过程可以看出其不同之处在于确定隐层 h^1 状态时的条件依赖不同, 在 DBN 模型中推导 h^1 时只依赖于 h^2 , 而在 DBM 模型中依赖于 v 和 h^2 两个方向信息的影响, 即既受到底层特征的影响也受到高层抽象特征的影响, 从而在生成可见层单元时会综合考虑输入

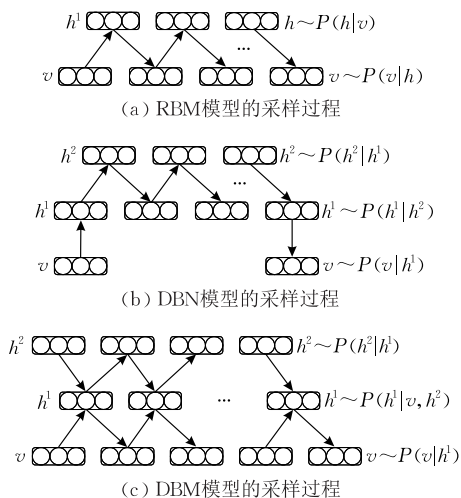


图 4 DBN、DBM 模型与 RBM 模型的采样过程

数据和模型所表达形状的特征, 其生成的形状数据会相对比较确定。

4 实验与分析

4.1 实验数据与参数设置

实验环境为 Windows10 64 位操作系统, 内存为 16 GB, 处理器为 Intel(R) Xeon(R) v3 1.90 GHz. 编程环境为 MATLAB R2014a. 为验证各模型的建模能力, 以 Weizmann Horse (figure-ground) 数据集和 Caltech101 Silhouettes 数据集作为实验对象, 图像均是二值的。从 Weizmann Horse 数据集中选择 200 幅图像形状作为训练样本, 其余 128 幅图像形状作为测试样本, 并将所有形状图像归一化至 32×32 像素大小, 所有图像中目标 (马) 均是面向左方且具有不同的姿势。该数据集是具有挑战性的, 原因是除了它们整体的目标形状不同之外, 马的头、尾和腿的姿态与位置也各不相同。从 Caltech101 Silhouettes 数据集中选择 Airplanes Side、Faces、Motorbikes 类的数据作为实验数据集, 其中 Airplanes Side 类 798 幅图像、Faces 类有 870 幅图像、Motorbikes 有 798 幅图像, 每幅图像的大小为 28×28 , 共有 2466 幅图像。选择其中的 2200 幅图像作为训练样本、266 幅图像作为测试样本。该实验数据集包括 3 种不同类别的形状数据, 且形状各异。

本文中各模型的主要参数和具体数值是根据文献[25]和具体试验选取的。模型的结构及其每层的单元个数设置如下: 对于 Weizmann Horse (figure-ground) 数据集, RBM、DBN 和 DBM 模型的可见层单元个数均设置为 1024、第 1 隐层的单元个数均设置为 500, DBN 与 DBM 模型中涉及到的第 2 隐层单元个数设置为 500; 对于 Caltech101 Silhouettes 数据集, RBM、DBN 和 DBM 模型的可见层单元个数均设置为 784、第 1 隐层的单元个数均设置为 2000, DBN 与 DBM 模型中涉及到的第 2 隐层单元个数设置为 1000。训练过程中涉及到的参数设置为: 对于 Weizmann Horse (figure-ground) 数据集, 迭代次数设置为 1500, 学习率为 0.01, CD 的步数设置为 1; 对于 Caltech101 Silhouettes 数据集, 训练过程中涉及到的参数及其设置为: 迭代次数设置为 3000, 学习率为 0.005, CD 的步数设置为 1。

4.2 模型建模能力比较与分析

为评估 RBM 与 DBN、DBM 深度学习模型对形

状的建模能力,从每个模型中使用各模型的采样方法得到一批形状,以训练形状和测试形状作为模型采样初始时的输入. 对于 Weizmann Horse 数据集的采样迭代次数设置为 100, Caltech101 Silhouettes 实验数据集的采样迭代次数设置为 300,此时模型的采样结果收敛. 从模型中采样生成的形状如图 5 所示,图 5(A)是以训练数据作为输入的结果,图 5(B)是以测试数据作为输入的结果,其中(a)列是输入形状数据,(b)列是从 RBM 模型中采样得到的形状,(c)列是从 DBN 模型中采样得到的形状,(d)列是从 DBM 模型中采样得到的形状. 从图 5 中各模型采样得出的样本对比可以看出,RBM 模型在建模形状时丢失了大部分的细节形状,比如采样得到的形状图像大部分缺少马的头、尾、腿或飞机的机翼等的形状信息,形状的边缘比较模糊;DBN 模型在建模形状时能够得到整体姿势变化的形状,整体结构完整,但目标形状比较模糊;DBM 模型能够生成各种不同姿势的形状且能够观察到生成的形状更为清晰和确定.

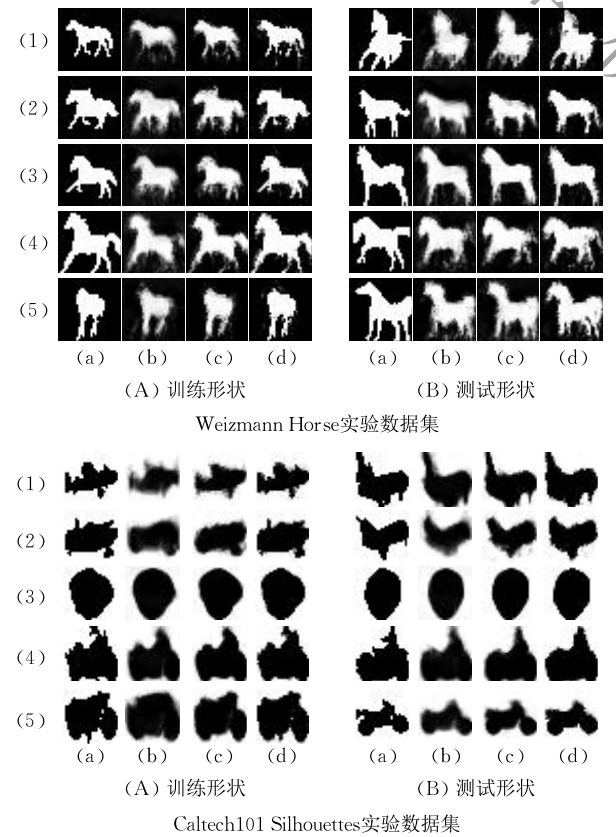


图 5 从各模型中采样生成的形状((a)列是输入的形状数据,(b)列是从 RBM 模型中采样得到的形状,(c)列是从 DBN 模型中采样得到的形状,(d)列是从 DBM 模型中采样得到的形状)

为定量分析模型对形状的建模能力,以形状作为输入,使用欧氏距离度量各模型采样输出与输入之间的相似性. 表 1 展示的是图 5 中 Weizmann Horse 实验数据集的结果对比,表 2 展示的是图 5 中 Caltech101 Silhouettes 实验数据集的结果对比,表中的各行与图 5 中的各行对应. 从表 1 和表 2 中的数据可以看出,不管是以训练形状作为输入,还是以测试形状作为输入,使用深度学习模型采样得到的形状与输入更为相似. 这说明使用深度学习模型对形状进行建模能够很好的拟合训练形状,而且对不存在于训练集中的测试形状有很好的泛化能力.

表 1 Weizmann Horse 实验数据集各模型采样输入与输出的欧氏距离对比

	训练形状			测试形状		
	RBM	DBN	DBM	RBM	DBN	DBM
(1)	2.9409	2.0273	1.4929	7.0077	5.7850	4.8580
(2)	8.3623	5.3784	3.8203	8.5591	6.3248	4.9093
(3)	5.6794	3.7824	2.4826	6.8093	5.6830	4.3737
(4)	5.8087	4.0506	3.4129	9.4532	7.3871	6.6293
(5)	8.4750	5.3821	3.3566	9.2414	8.2455	7.7785

表 2 Caltech101 Silhouettes 实验数据集各模型采样输入与输出的欧氏距离对比

	训练形状			测试形状		
	RBM	DBN	DBM	RBM	DBN	DBM
(1)	7.2408	5.1785	2.5925	8.8310	6.1231	5.8971
(2)	3.8234	2.8699	2.0163	6.1590	5.4701	4.5739
(3)	7.2730	5.7200	2.7203	5.2438	4.9308	4.4797
(4)	6.9246	5.1209	3.1096	3.5990	3.2472	2.7927
(5)	4.4929	3.7117	2.8088	6.4578	5.6874	4.7095

图 6 是从形状修复方面对比 RBM 与 DBN、DBM 深度学习模型的建模能力. 形状修复的目标是在给定输入形状部分缺失的情况下,使用模型生成形状缺失部分可能的形状姿态. 由于经过训练之后的模型能够为在 V_o 的条件下关于 V_u 的值指定一个确定的概率分布(o 表示可确定的形状区域, u 表示缺失的形状区域),所以在形状修复过程中可以通过在未丢失形状 V_o 的条件下采样得到丢失的或未观测到的其余形状 V_u ,这一过程可通过使用 Gibbs 采样方法从条件分布中采样得到,即执行一条马尔科夫链,从可见层、隐层单元对应的条件分布中得到各自的状态. 注意在每次更新可见层单元状态时,确定的形状部分 V_o 不作变化,只对未观测到的形状部分 V_u 更新. Weizmann Horse 数据集的迭代次数为 100 次, Caltech101 Silhouettes 实验数据集的迭代次数为 300 次. 图 6 对应的是形状修复的实验结果,其中(a)列对应的是从训练集中选取的完整形状,(b)列

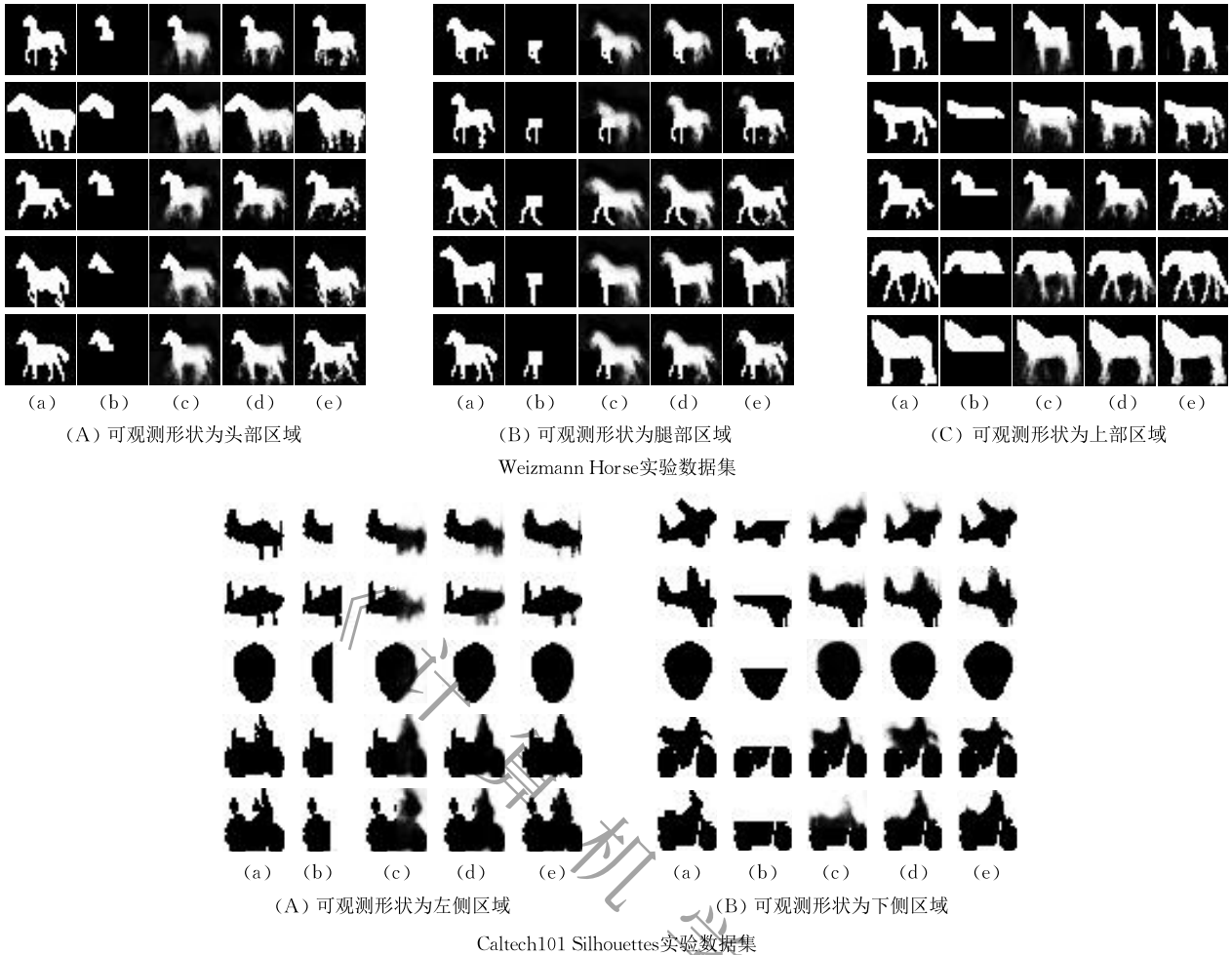


图 6 各模型图像修复的结果((a)列是训练集形状,(b)列是输入的部分缺失形状数据,(c)列是利用 RBM 模型修复得到的形状,(d)列是利用 DBN 模型修复得到的形状,(e)列是利用 DBM 模型修复得到的形状)

对应的是缺失部分形状区域作为输入形状数据,(c)列是使用 RBM 模型修复的结果,(d)列是使用 DBN 模型修复的结果,(e)列是使用 DBM 模型修复的结果.对于 Weizmann Horse 实验数据集,(A)组对应的是可观测的形状是头部区域,其它部分缺失;(B)组对应的是可观测的形状是部分腿部区域,其它部分缺失;(C)组对应的是缺失腿部形状区域;对于 Caltech101 Silhouettes 实验数据集,(A)组对应的是可观测的形状是左侧区域,右侧区域缺失;(B)组对应的是可观测的形状是下侧区域,上侧区域缺失.从实验结果中可以看出,RBM 模型能够根据确定的区域采样出大体形状,但经过修复之后的形状还是缺失形状的细节信息,生成的形状缺失部分模糊;相对于 RBM 模型,DBN 模型能够获得更为清楚的形状姿势和轮廓,整体结构完整但目标形状还是较为模糊;DBM 模型则能够获得目标形状更为清晰的形状.尽管缺失形状部分的大小不确定和

形状姿势不同,但经过形状修复之后的结果看起来还是较为符合实际的.

图 7 给出的是不同初始化状态作为模型采样的初始化输入对从模型中采样得到形状的影响,Weizmann Horse 实验数据集的迭代次数设置为 1000 次,Caltech101 Silhouettes 实验数据集的迭代次数设置为 3000 次.图 7 中给出的采样结果是重复采样 5 次的结果,其采样得到的形状有些不同.对于 Weizmann Horse 实验数据集,(A)组是用一个随机初始化的数据作为各模型采样的初始状态,然后经过多次状态转移得到近似服从模型所定义的分

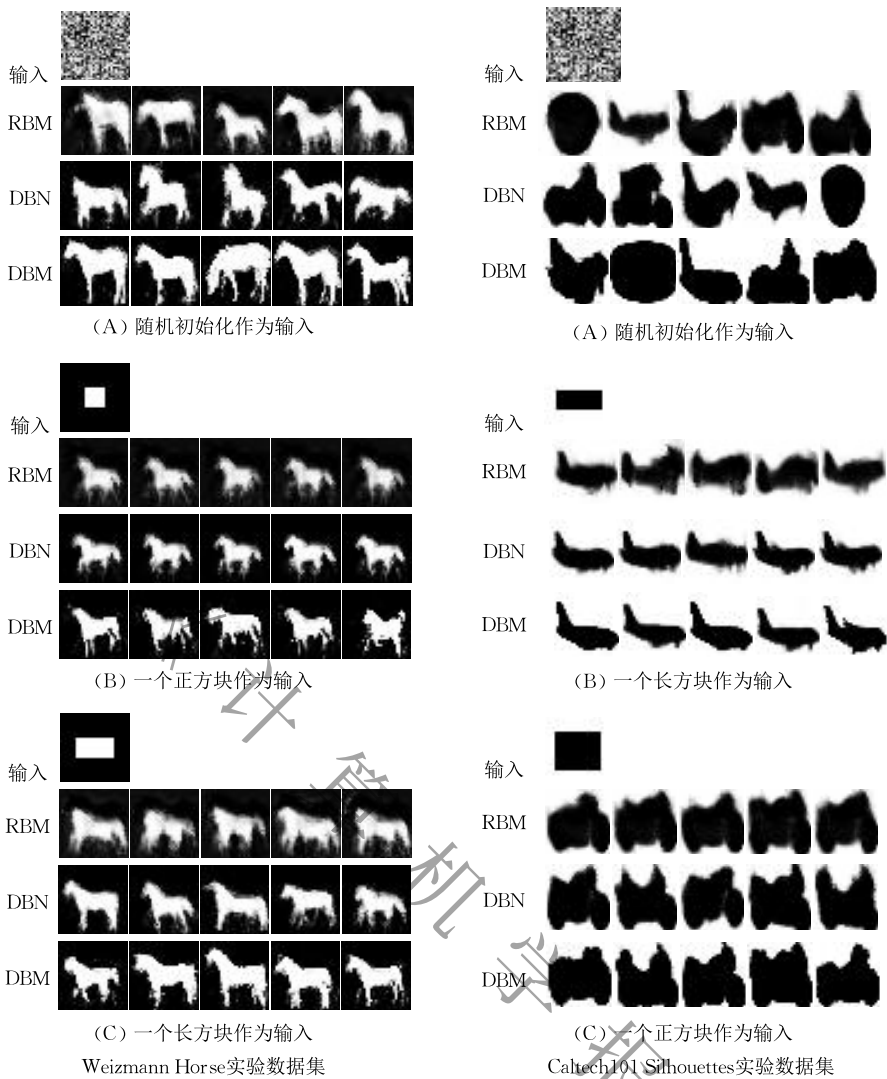
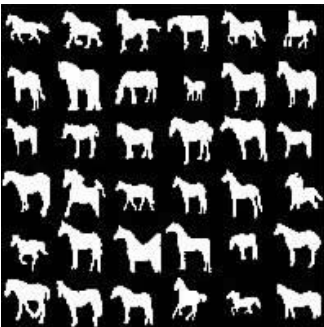


图 7 不同的输入对采样得到样本的影响

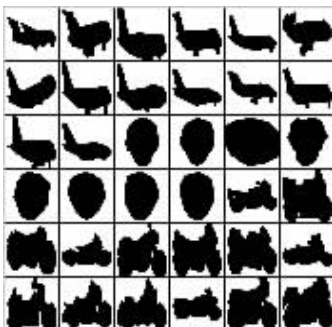
采样的初始状态得到的形状；(C)组是使用一个大的正方形作为模型采样的初始状态产生的形状. 从实验结果中可以看出, 初始化状态的不同能够影响由 RBM、DBN、DBM 模型生成的样本的形状, 它能够决定形状的大体姿势和位置. 并且能够从结果中看出由 DBN、DBM 深度学习模型生成的样本要比 RBM 模型更符合目标, 即形状的边缘比较清晰、四肢等形状比较分明, 形状数据更为确定, DBN 模型生成的样本目标完整但有些模糊, 而 DBM 模型生成的形状的清晰度明显提高. 这在实际应用中, 可以用初始形状作为输入数据, 根据模型采样的过程生成一个既受到初始输入形状的影响也受到模型约束的目标形状, 该形状可以作为一个有效的形状模板为其它应用提供形状先验约束.

图 8 是 RBM 与 DBN、DBM 模型以含有噪声的图像作为模型采样过程的输入, 迭代一次后的模型

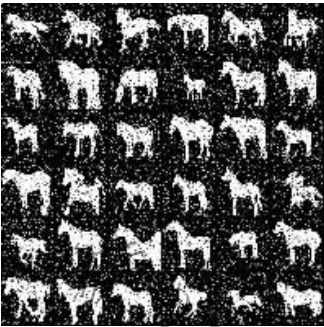
输出结果. 结果表明模型在输入的形状含噪声的情况下, 输出能够有效地去除噪声对目标形状的干扰, 模型具有较强的抑制噪声的能力. 图 8(A)是从测试集中选取的 36 幅原始形状图像, 图 8(B)是在图 8(A)的基础上, Weizmann Horse 实验数据集添加 20%、Caltech101 Silhouettes 实验数据集添加 15% 的服从 $(0, 1)$ 均匀分布的噪声图像, 图 8(C)是 RBM 模型产生的结果, 图 8(D)是 DBN 模型产生的结果, 图 8(E)是 DBM 模型产生的结果. 从图 8 中可以看出以噪声作为输入经各模型采样得到的结果有效地去除了噪声, 而且还保留了原图像中的许多边缘. RBM 模型在采样过程中形状的边缘轮廓部分丢失严重且变得模糊; 而 DBN 和 DBM 模型的整体效果更好, 减少了形状边缘轮廓和形状弱细节的丢失, DBM 模型的形状整体相对来说更为清晰.



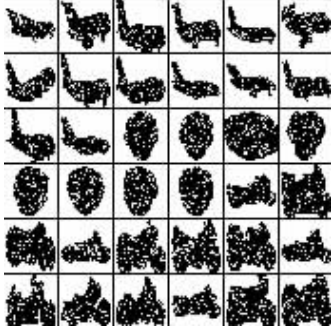
(A) 未添加噪声的图像



(A) 未添加噪声的图像



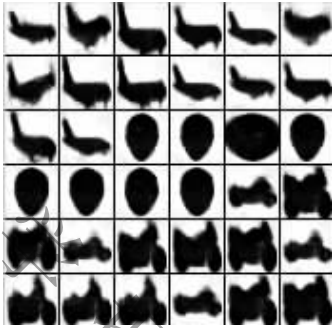
(B) 添加噪声后的图像



(B) 添加噪声后的图像



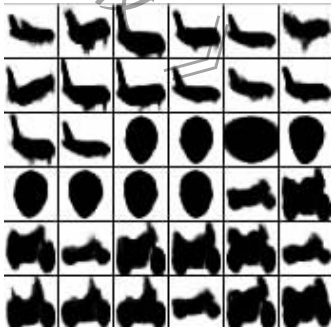
(C) RBM输出图像



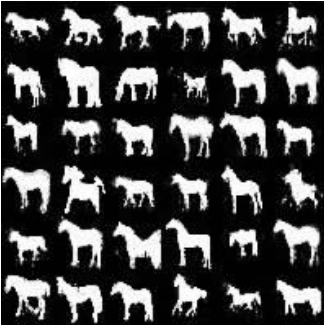
(C) RBM输出图像



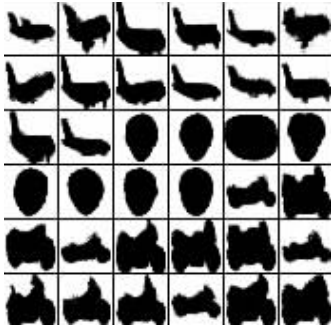
(D) DBN输出图像



(D) DBN输出图像



(E) DBM输出图像



(E) DBM输出图像

图 8 各模型抑制噪声的结果比较

5 总 结

作为统计建模方法中的一种,由 RBM 扩展的 DBN、DBM 深度学习模型在图像处理中有着广泛的应用。本文提出一类基于 DBN、DBM 深度学习模型的形状建模方法,这类模型关注的是全局形状特征,能够更好地体现形状的内在特性,对目标形状更准确地建模。实验从多个方面对 DBN、DBM 深度学习模型和 RBM 形状模型进行比较,结果表明 DBN、DBM 这两种深度形状表示模型比 RBM 模型的建模能力强,生成的形状更为符合现实。未来的研究包括考虑形状对齐以及模型层数、节点数等参数的自适应确定,将这类模型作为形状先验进一步扩展到图像分割、图像检测应用中等。

参 考 文 献

- [1] Alexe B, Deselaers T, Ferrari V. ClassCut for unsupervised class segmentation//Proceedings of the 11th European Conference on Computer Vision. Crete, Greece, 2010: 380-393
- [2] Eslami S M, Williams C. Factored shapes and appearances for parts-based object understanding//Proceedings of the 22nd British Machine Vision Conference. Dundee, Scotland, 2011: 1-12
- [3] Williams C K I, Titsias M K. Greedy learning of multiple objects in images using robust statistics and factorial learning. *Neural Computation*, 2004, 16(5): 1039-1062
- [4] Le Roux N, Heess N, Shotton J, et al. Learning a generative model of images by factoring appearance and shape. *Neural Computation*, 2011, 3(3): 593-650
- [5] Cootes T F, Taylor C J, Cooper D H, et al. Active shape models-their training and application. *Computer Vision and Image Understanding*, 1995, 61(1): 38-59
- [6] Boykov Y Y, Jolly M P. Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images//Proceedings of the 8th IEEE International Conference on Computer Vision. Vancouver, Canada, 2001: 105-112
- [7] Rousson M, Paragios N. Prior know ledge level set represent & visual grouping. *International Journal of Computer Vision*, 2008, 76(3): 231-243
- [8] LeCun Y, Bengio Y, Hinton G E. Deep learning. *Nature*, 2015, 521(7553): 436-444
- [9] Hinton G E, Sejnowski T J. Learning and relearning in Boltzmann machines. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, 1986, 1: 282-317
- [10] Smolensky P. Information processing in dynamical systems: Foundations of harmony theory. University of Colorado, Boulder: Technical Report CU-CS-321-86, 1986
- [11] Hinton G E, Osinder S, Teh Y. A fast learning algorithm for deep belief nets. *Neural Computation*, 2006, 18(7): 1527-1554
- [12] Roux N L, Bengio Y. Representational power of restricted Boltzmann machines and deep belief networks. *Neural Computation*, 2008, 20(6): 1631-1649.
- [13] Salakhutdinov R, Hinton G E. Deep Boltzmann machines//Proceedings of the 12th International Conference on Artificial Intelligence and Statistics. Florida, USA, 2009: 448-455
- [14] Kae A, Sohn K, Lee H, et al. Augmenting CRFs with Boltzmann machine shape priors for image labeling//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Portland, USA, 2013: 2019-2026
- [15] Chen F, Yu H, Hu R, et al. Deep learning shape priors for object segmentation//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Portland, USA, 2013: 1870-1877
- [16] Eslami S M A, Heess N, Williams C K I, et al. The shape Boltzmann machine: A strong model of object shape. *International Journal of Computer Vision*, 2014, 107(2): 155-176
- [17] Fischer A, Igel C. Training restricted Boltzmann machines: An introduction. *Pattern Recognition*, 2014, 47(1): 25-39
- [18] Hinton G E. Training products of experts by minimizing contrastive divergence. *Neural Computation*, 2002, 14(8): 1771-1800
- [19] Bengio Y, LeCun Y. Scaling learning algorithms towards AI. *Large-Scale Kernel Machines*, 2007, 34(5): 321-359
- [20] Hinton G E, Dayan P, Frey B, et al. The wake-sleep algorithm for unsupervised neural networks. *Science*, 1995, 268(5214): 1158-1161
- [21] Salakhutdinov R, Hinton G E. An efficient learning procedure for deep Boltzmann machines. *Neural Computation*, 2012, 24(8): 1967-2006
- [22] Zhu S C, Dellaert F, Tu Z W. Markov chain Monte Carlo basics. *Advances in Neural Information Processing Systems*, 2005, 5: 497-537
- [23] Hastings W K. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 1970, 57(1): 97-109
- [24] Walsh B. Markov chain Monte Carlo and Gibbs sampling. *Notes*, 2004, 91(8): 497-537
- [25] Hinton G E. A practical guide to training restricted Boltzmann machines. *Momentum*, 2010, 9(1): 599-619



ZHANG Juan, born in 1992, M. S. candidate. Her research interests include machine learning and image processing.

WANG Xi-Li, born in 1969, Ph. D. , professor, Ph. D. supervisor. Her major research interests include intelligent information processing, pattern recognition and image processing.

YANG Jian-Gong, born in 1974, Ph.D. candidate, lecturer. His major research interests include machine learning and image processing.

Background

Models of the shape of an object play a crucial role in many applications such as object detection, image segmentation and inpainting. Compared with the gray, texture, edge and other low-level visual features of the image, shape as high level visual feature helps to describe the global visual information of an object better. In general, the algorithm's performance will be improved after global shape information (expressed by good model) joined. At present, there are many methods to modeling 2-dimensional shape. The shape expressed in marked point set is one of the easiest way. It is based on marked point distribution model. But such representation is too passive. The points need to mark manually and it is error-prone. The most commonly used models are Markov random fields or conditional random fields. In such models, the pairwise potentials connecting neighboring pixels impose very local constraints, thus they are unable to capture more complex properties. Level set can also be used to represent shape. In level set method, signed distance function is used to represent shape. The shortcomings of this kind of method is that during the curve evolution process, signed distance function will oscillate, and then makes the final shape of the curve evolution deviate from the real shape of the object's contour. In addition, this method cannot fit multi-modal

distribution.

Recently, deep learning models have attracted more and more attentions because of their excellent learning ability on modeling data with intrinsic structure. Deep learning models contain multiple levels and hidden perception units. They can extract multi-level features and represent complicated data distribution. They form high-level data abstract representation through combining the lower layer features. In 2012, Eslami et al. proposed a probabilistic model for the task of modeling binary object shapes. However, due to the partial connection between visible layer and hidden layer, the object shape will appear seams when the model is used to generate shape. In this paper, we propose two shape models based on deep models to represent the complicated global shape features. The samples generated from the deep models look more realistic and completed, which indicate the proposed shape models are effective in shape representing and generating.

This work is supported by the National Natural Science Foundation of China under Grant Nos. 41171338, 41471280, and 61401265. They are focus on object detection in high-resolution remote sensing image. We will apply the proposed shape models to object detection in further study.