网络出版时间：2017-02-10 07:55:18
网络出版地址：http://www.cnki.net/kcms/detail/11.2127.TP.20170210.0755.004.html

融合多特征的深度学习标注方法

黄冬梅， 许琼琼， 贺 琪， 杜艳玲
HUANG Dongmei, XU Qiongqiong, HE Qi, DU Yanling
上海海洋大学 信息学院，上海 201306
Department of Information and Technology, Shanghai Ocean University, Shanghai 201306, China

HUANG Dongmei, XU Qiongqiong, HE Qi, DU Yanling. Multi-features fusion for image auto-annotation based on DBN model. Computer Engineering and Applications

Abstract：To bridge the semantic gap between low-level visual feature and high-level semantic concepts has been the subject of intensive investigation on big data management for years in order to improve the accuracy of image auto-annotation. Multi-features Fusion for Image Auto-Annotation Based on DBN Model is proposed, combine image visual features with different weights as inputs of DBN model and optimize parameters of DBN, achieve image automatic annotation on big image data. The experimental results based on Corel image database show that the proposed method, taking different features of images into considering, have good performance on annotation precision.

Key words：multi-features fusion; deep learning; restricted Boltzmann machine; image annotation

摘 要：缩小图像低层视觉特征与高层语义之间的鸿沟，以提高图像语义自动标注的精度，是研究大规模图像数据管理的关键。提出一种融合多特征的深度学习图像自动标注方法，将图像视觉特征以不同权重组合成词包，根据输入输出变量优化深度信念网络，完成大规模图像数据语义自动标注。在通用 Corel 图像数据集上的实验表明：融合多特征的深度学习图像自动标注方法，考虑图像不同特征的影响，提高了图像自动标注的精度。

关键词：多特征融合；深度学习；受限玻尔兹曼机；图像标注
doi:10.3778/j.issn.1002-8331.1607-0297 **文献标志码**：A **中图分类号**：TP391

1 引言

随着互联网突飞猛进的发展，移动终端和社交网络的兴起，网络中多媒体数据尤其是图像数据呈几何级增长。2009 年，Flichkr 拥有 40 亿张图像，到 2011 年为止上升到了 60 亿。Facebook 的图像以 150 亿/年的速度在增长[1]。图像数据的持续增多，劣质图像数据也随之而来，导致图像质量低劣，极大降低了图像数据的可用性。人们期望通过关键字检索，从大规模图像中快速获取目标图像，图像与标注词之间匹配质量是提高图像可用性的关键。针对大规模图像数据集，图像语义自动标注帮助判定和选择高质量图像数据源，增强图像数据的可用性。

图像语义的准确标注是提高大规模图像数据集可用性的关键。图像语义自动标注根据统计学习模型可以分为：基于生成模型的图像标注方法[2-5]、基于可判别分类模型的图像标注方法[6-8]。基于生成

基金项目：国家重点基础研究发展计划(2012CB316206)；国家自然科学基金(61402282, 61272098)；上海海洋大学科技发展专项基金项目(No. A2-0203-00-100210)。

作者简介：黄冬梅(1964-),女,教授,博士生导师，CCF 会员(E20-0006885S)，研究领域为大数据和智能信息处理；许琼琼(1991-),女，硕士研究生，研究领域为图像自动标注；贺琪(1979-),女,博士，讲师，研究领域为数据库和面向服务的计算;杜艳玲(1987-),女博士研究生，研究数据存储和数据迁移。E-mail：qihe@shou.edu.cn

模型的图像标注方法对视觉特征和语义概念之间的联合概率分布进行估计，从而对新图像标注。代表模型包括：LDA 模型[2]、MBRM[3]等。如 Lienou 等人[4]基于 LDA 模型进行学习，采用简单的视觉特征提取来描述图像内容，对未标注图像通过 LDA 模型分配概率，使用最大似然法将每个分块对应到语义概念。尚赵伟等人[5]基于日志数据增量挖掘的混合概率模型进行图像标注。基于生成模型的图像标注过程容易受到具有高视觉相似性而语义不同的图像的影响。基于可判别分类模型的图像标注方法将每个标注词看作一个独立的类，并为每个类学习相应的分类器。主要方法包括：基于 SVM 的方法[6]、基于组稀疏的多核学习方法[7]等。欧阳宁等人[8]利用颜色和纹理描述子提取图像视觉特征，为每个主题图像建立混合模型，实现基于决策融合的图像标注。但对复杂标注问题，其泛化能力受到制约，面临维数灾难和过学习问题等。Wei Wei 利用无线传感器网络帮助一个动态实时环境的信息查询和导航，提出一个多尺度梯度下降的方法来满足用户的无线传感器网络的要求[9]。

Hinton 等人提出深度信念网络（Deep Belief Networks, DBN）[10]，结合无监督学习和有监督学习的优点，对高维图像数据从低层到高层渐进地进行特征提取，挖掘其潜在规律，提高图像语义标注的精确性。DBN 已广泛应用于图像识别、语音识别、图像分类等各种领域。陈亮等人[11]基于 DBN 进行视频热度预测，为视频在上映前的投资和播放提供有价值的参考。吴财贵等人[12]基于深度学习方法对图片敏感文字的检测具有良好的鲁棒性，且检测速率和效率都得到了提高。杨阳等人[13]提出改进的深度学习模型，将图像的标注信息视为图像的类别信息，对图像特征关注不足。

基于此，提出一种大规模图像数据预处理的技术，融合多特征的深度学习图像自动标注方法，有效提高图像自动标注的精度，确保大规模图像数据的可用性。

2 多特征融合

图像语义自动标注是指让机器通过多实例学习，针对高级语义概念和图像视觉特征的关系进行自动建模，利用学习到的模型自动完成新图像语义的标注，能够帮助判定和选择高质量数据源，从而增强数据的可用性。

对图像进行实例表示，用 (X_i,Y_i) 表示图像集合，其中 $X_i=\{x_i^1,x_i^2...,x_i^m\}(1\le i\le n)$ 代表图像集合中第 i 幅图像，$Y_i=\{y_i^1,y_i^2...,y_i^l\}$ 代表该幅图像对应的标签集。m 表示图像的特征维数；l 表示图像的标注词个数，对于不同的图像，l 的个数是相等的；y_i^j 是一个二元变量，表示第 j 个标注词是否出现在第 i 幅图像中。利用图像的颜色、纹理、形状、空间关系等特征，提取高维图像特征。

图像中某一对象的颜色特征将不会随着图像的旋转、移动等而发生改变。采用一种基于 HSV 颜色空间的非等间隔量化方法，HSV 是一个颜色感知模型，能够较好地表示颜色的三个基本属性：色调、亮度和饱和度。颜色空间转换示意图如图 1。将得到的图像 RGB 空间转化到图像 HSV 空间，按照人对颜色感知，在 H、S、V 三个分量上进行非等间隔量化。图像 a 是原图，图像 aa 是转换为 HSV 空间的图像，图像 a'、图像 a''、图像 a''' 分别对应 H、S、V 三个分量。

(a)图像 a (b)图像 a' (c) 图像 a'' (d) 图像 a''' (e) 图像 aa

图 1 RGB 转换 HSV 举例

采用灰度共生矩阵法进行纹理特征的提取。假设一副二维图像 $f(x,y)$，大小 $M*N$，图像灰度级 N 级，则共生矩阵 $M_{d\theta}(i,j)$ 为 $M*N$ 矩阵，灰度值为 i，j 距离为 d 的两像素同时出现的概率分布为 $P_{d\theta}$。θ 选择四个离散方向：0°，45°，90°，135°，能量、熵、惯性矩、相关的均值和标准差等值。

基于尺度不变特征转换（SIFT）检测局部特征描述图像形状特征，先生成尺度空间，然后检测尺度空间的极值点，精确定位极值点，为每个关键点指定方向参数，最后生成点描述子。

所有特征以词包的形式存储，分别将颜色、纹理、形状特征以不同权重组合成一维向量，进行基于改进 DBN 模型的图像语义自动标注。

3 基于受限玻尔兹曼机的 DBN 图像标注

深度信念网络是由多层无监督的受限玻尔兹曼机网络和一层有监督的反向传播网络组成。通过受限玻尔兹曼机进行预训练，利用反向传播网络实现调优。

受限玻尔兹曼机是深度信念网络模型的核心，由两层神经元组成，一层是可见层，$\boldsymbol{v}\in\{0,1\}^I$ 表示可见层节点的状态，I 为可见层节点数目；一层是隐含层，$\boldsymbol{h}\in\{0,1\}^J$ 表示隐含层节点的状态，J 为隐含层节点数目，如图 2 所示。可见层和隐含层内部的节点都没有互连，只有层间的节点有对称的权连接 $\boldsymbol{W}$。当给定可见层节点的状态时，各个隐含层节点的激活状态之间是相互独立的，即：$P(\boldsymbol{h}\mid\boldsymbol{v})=\prod_{j=1}^{J}P(h_j\mid\boldsymbol{v})$。当给定隐含层节点的状态时，各个可见层节点的激活状态之间是相互独立的，即：$P(\boldsymbol{v}\mid\boldsymbol{h})=\prod_{i=1}^{I}P(v_i\mid\boldsymbol{h})$。

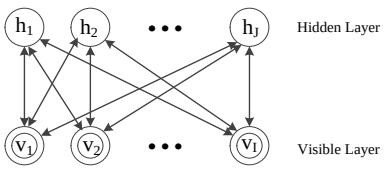

图 2 RBM 结构图

受限玻尔兹曼机的可见层输入为 $\mathbf{v}\in\{0,1\}^I$ 二元变量，假设每个神经元遵循伯努利分布，定义该模型的能量函数如下：

$$E(\boldsymbol{v},\boldsymbol{h}\mid\theta)=-\sum_i\sum_j w_{ij}v_ih_j-\sum_i b_iv_i-\sum_j a_jh_j$$

其中 $\theta=\{a,b,\boldsymbol{W}\}$ 是模型的参数。可得到 $(\boldsymbol{v},\boldsymbol{h})$ 的联合概率分布为

$$P(\boldsymbol{v},\boldsymbol{h})=e^{-E(\boldsymbol{v},\boldsymbol{h})}/Z$$

其中，Z 为归一化常数。

输入 $\boldsymbol{v}$，通过 $P(\boldsymbol{h}\mid\boldsymbol{v})$ 得到隐含层 $\boldsymbol{h}$，通过 $P(\boldsymbol{v}\mid\boldsymbol{h})$ 重构可视层 $\boldsymbol{v}_1$，目标是调整参数 θ，使得 $\boldsymbol{v}_1$ 和 $\boldsymbol{v}$ 尽可能一样，然后计算可视层的重构误差 ε，若 ε 较大，则要计算 $P(\boldsymbol{h}_1\mid\boldsymbol{v}_1)$ 求得新的隐含层 $\boldsymbol{h}_1$，继续重构可视层，反复执行，知道重构误差 ε 达到合理范围，或者重构步数达到要求。采用上述方式进行训练，学习出参数 $\theta=\{a,b,\mathbf{W}\}$ 的值。

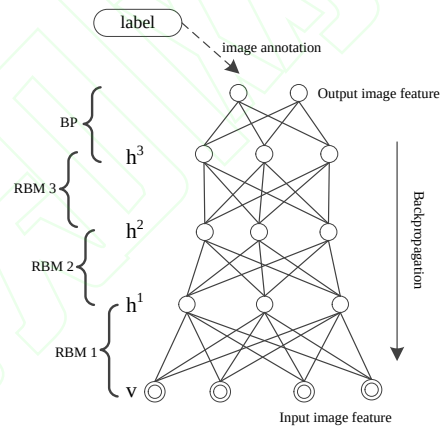

图 3 融合多特征的深度学习标注过程

融合多特征的深度学习标注方法分为两步，如图 3 所示。首先训练集中图像数据特征词包作为第一层输入，无监督地训练每一层 RBM 网络，逐层从底层特征转换成抽象的高层特征。训练中所有图像数据均已知类别，则图像标注词总数目可设定为 l，当前训练图像数据所属类别为 j，那么训练图像的输出层第 j 维为 1，其余维数为零。对输出层的结果做排序，排序靠前的类别为该神经网络对图像类别的预测结果。在反向传播网络中，通过计算训练集中图像数据的神经网络实际输出与图像标注词向量之间的差值衡量网络的收敛程度，当满足训练次数要求时停止训练。

4 图像标注改善

利用学习的模型自动完成新图像语义的标注，图像标注词汇之间存在着多种多样的语义层次和关系，包括相近、对立、包容等。比如，一副图像已被标注为“汽车”、“人”等词汇后，“道路”作为该图像标注词汇的概率就会相应提升。利用共生

词汇在同一副图像出现的相关性可以有效地提高词汇之间的语义相关信息，从而提高图像标注的准确率。根据文献提出的度量方式[13]，综合考虑基于共生关系的统计互相关度量及基于 WordNet 的语义相关度量方法，进行线性融合，作为词间关系的联合表示：

$$S=\varepsilon C_{TF}+(1-\varepsilon)C_{WN},0\le\varepsilon\le1$$

其中 C_{TF} 是基于共生关系对词汇间相似性的度量，C_{WN} 是基于 WordNet 的语义相关性度量。

本文提出融合多特征的深度学习图像自动标注方法，实现过程如图 4 所示。训练集图像对高层语义概念和图像低层视觉特征的关系进行自动建模，测试集图像对模型进行调优，利用学习到的标注模型自动完成新图像语义的标注。

主要任务有 2 个：（1）检验 DBN 结构对图像语义自动标注的影响；（2）验证融合多特征的深度学习标注方法适合图像语义自动标注。

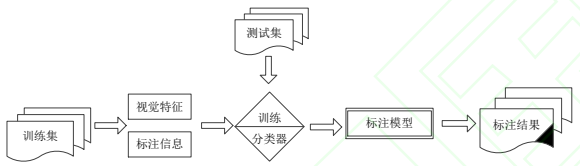

图 4 图像标注过程

图 5 Corel5K 数据库中图像示例

采用 10 次交叉验证方法评估隐含层节点数的影响，即将全部图像样本平均分成 10 份，每次使用其中的 9 份数据用于训练，剩余 1 份用于测试，保证每个标注至少出现一次。利用精度验证图像标注方法的性能，定义如下：

$$Precision=\frac{|W_c|}{|W_M|}$$

假设某一标注词 w 作为查询，在标注好的测试图像集上进行检索，假设标注正确的图像数为 $|W_c|$，可检索到的所有图像数为 $|W_M|$。

在任务（1）中，DBN 模型的层数与节点数是影响其标注效果的关键，过多的层数或隐含层节点数会导致过拟合现象。采用固定其他参数而变化一种参数的方法来讨论不同网络深度及隐含层节点数对标注效果的影响。

在任务（2）中，将训练集的图像特征分别用以训练 DBNF、DBN、LDA、SVM 的标注分类器，用训练完毕的模型对测试集图像数据进行测试。比较在相同测试集图像特征下，各个不同模型的语义标注分类器的效果。

5 实验与分析

本节主要讨论两方面内容，一是 DBN 结构对图像标注的影响，二是融合多特征的深度学习标注方法与其他方法的比较。

5.1 实验设置与评价

实验中使用 Corel5K 图像集，由 50 个 CD 组成，每个 CD 包含 100 张大小相等的图像，共 5000 幅图像，每张 CD 代表一个语义主题，例如有公共汽车、恐龙、海滩等，如下图 5 所示。每张图片被标注 3~5 个标注词，训练集中总共有 374 个标注词，在测试集中总共使用了 263 个标注词。

为了验证本文提出方法的有效性，并与其它方法进行比较，设计了 2 项任务：

（1）检验 DBN 结构对图像语义自动标注的影响；

（2）验证融合多特征的深度学习标注方法适合图像语义自动标注。

针对任务 1，设计了 4 组不同的实验进行图像标注。参数设置如下：RBM 模型可见层节点数等于所提取的输入图像样本的特征维数。隐含层层数分别取 1，2，3，4 层，每一隐含层的节点数分别

取 40，80，120，160 进行标注实验。

针对任务 2，设计了 3 组不同的实验。分别用本文提出的自动标注方法与文献[4][6][13]方法进行标注实验。SVM 是 Cortes 和 Vapnik 于 1995 年首先提出的，适合维度高数据的标注，本文采用径向核(Gaussian RBF)函数，利用 LIBSVM 软件包[14]进行基于 SVM 图像自动标注实验。实验中所有算法在 Matlab2013a 平台运行。

5.2 实验结果及分析

5.2.1 参数设置对该方法的影响

图 6 是在隐含层层数分别取 1，2，3，4 层，每一隐含层的节点数分别取 40，80，120，160 条件下标注结果的比较。在隐含层层数为 3 层时，隐含层节点数为 80，标注精度达到最优，随着隐含层节点数的增加，标注精度有所下降。主要是因为特征越多，用来训练的数据在每个特征上就会稀疏，会降低 DBN 的泛化性能，容易导致过拟合现象。

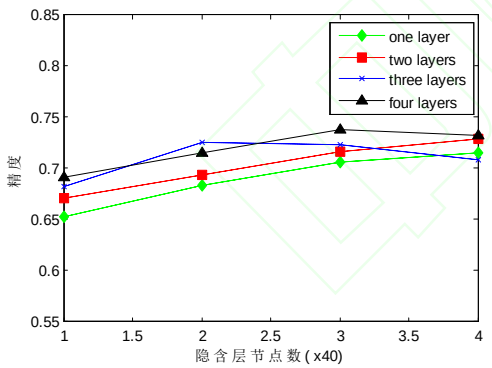

图 6 DBN 参数对标注结果的影响

隐含层层数分别取 1，2，4 层时，随着隐含层节点数的增多，标注精度越高。在每一隐含层的节点数为 120 时，达到最优。当隐含层节点数继续增加时标注精度缓慢提高。主要是因为过多隐含层节点数会过大时，神经元之间的连接较多，也会导致过拟合问题的出现。

说明本文融合多特征的深度学习标注方法中 DBN 层数最佳为 4 层，隐含层节点数最佳为 120。

5.2.2 相关方法的比较与分析

图 7 是本文标注方法、文献[4]基于 LDA 标注方法，文献[6]基于 SVM 标注方法与文献[13]基于 DBN 方法在提取不同特征维数的情况下的精度比较。

四种方法的精度都随着提取特征维数的提高而提升，尤其是在维数为 60 维以后，精度提升显著，在维数为 150 维时达到最优。

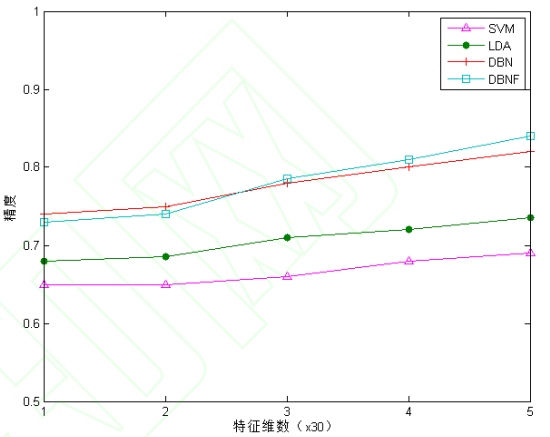

图 7 不同标注方法的比较

由上图可知，基于 SVM 标注方法与本文标注精度差距最明显，是因为基于 SVM 标注方法直接为每个标注词设计分类器，而本文通过逐层提取特征，形成抽象的高层特征进行标注。在 60 维特征之前，本文标注方法比基于 DBN 的标注方法精度要低，是因为对特征的融合，使得在特征维数较低情况下，精度较低。本文提出的方法在 60 维特征以后比文献[4]标注方法、文献[6]标注方法精度高，主要是因为本文考虑到了图像内容，融合图像的多语义特征，对标注性能的提高有更好的效果。

表 1 是四种方法在训练和测试阶段运行时间对比。

表 1 运行时间对比

分类器	训练时间/s	测试时间/s
SVM	4.135	5.252
LDA	5.212	3.252
DBN	47.028	0.019
DBNF	53.421	0.013

其中基于深度学习模型的标注方法训练时间均比 SVM、LDA 要长，但在测试阶段时间明显更短，说明合适的深度学习模型在处理大规模数据集过程中，可很少修正直接运行。在训练阶段本文标注方法比单纯使用深度学习的方法需要更多时间，

是因为本文融合了特征词包，使得测试阶段本文标注方法运行时间更短。

6 总结和展望

图像数据呈几何级增长，导致数据的可用性下降。本文提出一种融合多特征的深度学习标注方法，该方法对输入数据从底层到高层渐进地进行特征提取，进而提高标注的精度。实验表明，本文方法有效地减弱图像低层视觉特征与高层语义特征之间的“语义鸿沟”问题，在层数为 4 层，隐含层节点数为 120 时，标注精度较基于 SVM、LDA、DBN 的标注方法明显提高，明显提升了图像语义标注效果。将来的工作拟在以下几个方面展开：（1）将本文提出的方法在更大规模的图像数据集上测试，以验证方法的普适性；（2）深入研究低层视觉特征之间的权重、优先序，最优化选取 DBN 模型初始值。

参考文献:

[1] 邱泽宇, 方全, 桑基韬,等. 基于区域上下文感知的图像标注[J]. 计算机学报, 2014, 第 6 期:1390-1397.[doi:10.3724/SP.J.1016.2014.01390].

[2] Blei D M, Jordan M I. Modeling annotated data[C]//Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval. ACM, 2003: 127-134.

[3] Feng S L, Manmatha R, Lavrenko V. Multiple bernoulli relevance models for image and video annotation[C]//Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on. IEEE, 2004, 2: II-1002-II-1009 Vol. 2.

[4] Li énou M, Ma îre H, Datcu M. Semantic annotation of satellite images using latent dirichlet allocation[J]. Geoscience anLDAd Remote Sensing Letters, IEEE, 2010, 7(1): 28-32.

[5] 尚赵伟,李振华,张澜. 基于日志的协同图像自动标注[J]. 计算机工程与应用,2015,08:178-182+194.

[6] Verma Y, Jawahar C V. Exploring svm for image annotation in presence of confusing labels[C]//Proceedings of the 24th British Machine Vision Conference. 2013.

[7] 袁莹,邵健,吴飞,庄越挺.结合组稀疏效应和多核学习的图像标注.软件学报,2012,23(9):2500-2509

[8] 欧阳宁,罗晓燕,莫建文,张彤. 基于决策融合的图像自动标注方法[J]. 计算机工程与应用,2013,21:156-159.

[9] Wei Wei, Qi Yong. Information potential fields navigation in wireless Ad-Hoc sensor networks[J]. Sensors, 2011, 11(5): 4794-4807.

[10] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 313(5786): 504-507

[11] 陈亮,张俊池,王娜,李霞,陈宇环. 基于深度信念网络的在线视频热度预测[J]. 计算机工程与应用,,:1-10.

[12] 吴财贵,唐权华. 基于深度学习的图片敏感文字检测[J]. 计算机工程与应用,2015,14:203-206+230.

[13] 杨阳, 张文生. 基于深度学习的图像自动标注算法[J]. 数据采集与处理, 2015, 30:88-98.[doi:10.16337/j.1004-9037.2015.01.008]

[14] Chang C C, Lin C J. LIBSVM: a library for support vector machines[J]. ACM Transactions on Intelligent Systems and Technology (TIST), 2011, 2(3): 27.