

# 融合卷积神经网络与主题模型的图像标注

张蕾\*, 蔡明

江南大学物联网工程学院, 无锡 江苏 214122

**摘要** 为降低图像文本数据的稀疏性和传统图像特征的局限性,提出一种融合卷积神经网络(CNN)和主题模型的图像标注算法。利用狄利克雷主题模型对图像训练集的文本数据进行建模,生成文本主题分布和文本主题标注词分布,以降低图像文本数据的维度和稀疏性。考虑到图像文本主题的稀疏分布,利用 CNN 提取图像的高层视觉特征,同时改进损失函数以重构 CNN。利用图像的高层视觉特征和对应的多个文本主题构建多分类器,进行图像文本主题多标签分类学习,并获得图像的文本主题分布。最后,将该文本主题分布和主题模型生成的文本主题标注词分布融合计算出图像的标注词概率。由 Corel5K 和 IAPR TC-12 图像标注数据集的对比实验可知,本文方法有效提高了图像的标注性能。

**关键词** 图像处理; 卷积神经网络; 主题模型; 图像标注; 损失函数; 多标签分类学习

中图分类号 TP391

文献标识码 A

doi: 10.3788/LOP56.201004

## Image Annotation Based on Convolutional Neural Network and Topic Model

Zhang Lei\*, Cai Ming

School of Internet of Things Engineering, Jiangnan University, Wuxi, Jiangsu 214122, China

**Abstract** To address the issue of the sparsity of image text data and the limitation of traditional image features, this study proposes an image annotation algorithm that combines a convolutional neural network (CNN) and a topic model. Herein, a Dirichlet topic model is used to model text data on image training sets and generate text topic distribution and text topic label distribution, which reduces the dimension and sparsity of image text data. Considering the sparse distribution of image text topic, the CNN is used to extract high-level visual image features, and the loss function is improved to reconstruct the CNN. Multiple classifiers are constructed based on the high-level visual image features and corresponding multi-text topics to perform multi-label classification learning on image text topics and obtain the text-topic distribution of image. Finally, the text-topic distribution and text-topic label distribution are combined to calculate the probability of the image label. Based on the contrast experiment on Corel5K and IAPR TC-12 image annotation datasets, the proposed algorithm effectively improves the performance of image annotation.

**Key words** image processing; convolutional neural network; topic model; image annotation; loss function; multi-label classification learning

**OCIS codes** 100.2000; 100.4996; 100.3008

## 1 引言

随着网络社交平台及 APP 的普及,互联网中无标签图像数据的规模日益增长。由于大数据技术的高速发展,网络中的无标签图像展现出巨大的商业价值,如何快速筛选并使用这些无标签图

像数据成为一个极具研究价值的问题。图像检索<sup>[1]</sup>可以高效检索出需要的图像,而其在无标签图像数据的检索效果依赖于图像标注方法。由于图像的视觉特征与文本标签之间存在“语义鸿沟”<sup>[2]</sup>,研究图像标注方法依旧是一份十分具有挑战性的工作。

收稿日期: 2019-03-20; 修回日期: 2019-04-23; 录用日期: 2019-04-26

\* E-mail: 289253808@qq.com

当前图像标注方法主要分为4类:第1类是判别式图像标注<sup>[3-4]</sup>,将图像自动标注问题看作是一个有监督的分类问题,通过计算视觉特征的相似性将一幅图分类为一个或多个语义类别;第2类是生成式图像标注模型<sup>[5-6]</sup>,计算出视觉特征和语义文本之间的联合分布概率,然后使用分布模型估计新图像的语义标注文本概率;第3类是最邻近图像标注模型<sup>[7-8]</sup>,依据图像特征找出与测试图像最相似的多张图像,并利用这些图像的标注词为测试图像进行标注;第4类是基于深度学习的图像标注模型。由于深度学习在多个领域取得了较好的成绩,特别是卷积神经网络<sup>[9]</sup>(CNN)在计算机视觉领域(如图像分类<sup>[10]</sup>和目标识别<sup>[11]</sup>等)表现出优秀的图像处理能力,因此基于深度学习的图像标注效果一般优于其他模型。例如,基于线性回归的CNN-R(CNN-Regression)模型,该方法通过优化模型参数,提高了图像标注的性能<sup>[12]</sup>;高耀东等<sup>[13]</sup>提出的CNN-MSE(CNN-Mean Squared Error)方法,通过将CNN的损失函数改进为均方差误差损失函数,极大地提高了图像的标注性能。虽然基于深度学习的图像标注性能较好,但忽略了图像集文本数据的稀疏分布及分布不平衡对图像标注的影响。为解决该问题,本文提出融合CNN和主题模型的图像标注方法。

本文充分利用狄利克雷(LDA)模型对图像文本信息处理和CNN对图像视觉特征提取的优势。利用LDA主题模型对图像训练集的文本数据建模,生成图像训练集的潜在文本主题分布和文本主题标注词分布,对图像训练集文本数据的处理弥补了CNN分类训练集中文本数据维度大、分布稀疏的问题;CNN提取图像的高层视觉特征弥补

了传统图像特征提取复杂和传递图像视觉信息有限的问题。通过结合CNN和主题模型对图像视觉数据和文本数据的处理,改进CNN中的损失函数,使其可以计算图像对应多个文本主题分类学习的损失和提高低频文本主题分类的准确率,从而提高图像标注的性能。

## 2 主题模型

LDA模型是一个生成式主题模型,其概率图模型如图1所示。模型中用到的符号及其所代表意义如表1所示。

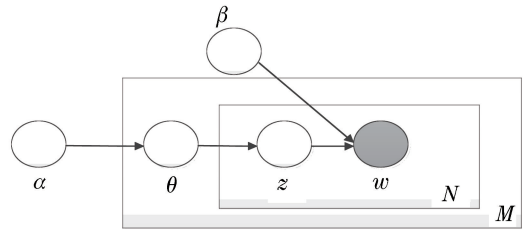


图1 LDA的图模型

Fig. 1 Graphical model of LDA

在LDA模型下文档的具体生成过程为

- 1) 抽取一个主题比例  $\theta \sim P_{\text{dir}}(\alpha)$ ;
- 2) 对于文档  $d_i$  的每个单词  $w_n, n \in (1, \dots, N), n$  为当前单词数;
- 3) 抽取主题分派  $z_n | \theta \sim \text{Mult}(\theta), z_n$  为当前单词对应的主题;
- 4) 生成单词  $w_n | z_n \sim \text{Mult}(\beta_{z_n})$ 。

潜在变量和可观测变量的联合分布为

$$P(w, \theta, z | \alpha, \beta) = P(\theta | \alpha) \left( \prod_{n=1}^N A \right) \quad (1)$$

式中,  $A$  为文档中单个单词在主题比例中出现的概率,  $A = P(z_n | \theta) P(w_n | z_n, \beta)$ 。

表1 符号及其意义

Table 1 Symbols and their meaning

Symbol	Meaning of symbol	Symbol	Meaning of symbol
$M$	Size of training set	$N$	Number of vocabulary
$K$	Number of topics	$w$	Vocabulary
$z$	Potential topic	$\theta$	Proportion of topic
$\alpha$	Parameter of model	$\beta$	Parameter of model
$\gamma$	Variational parameter $\alpha$	$\varphi$	Variational parameter $\beta$
$P_{\text{dir}}$	Dirichlet distribution	$\text{Mult}(\cdot)$	Polynomial distribution

利用LDA主题模型进行图像标注,需要得到测试集的图像主题分布  $P(T_k | d_{\text{test}})$ ,然后根据该分布选择图像的主题  $T$ ,最后计算出图像的标注词概率。图像的标注词概率为

$$P(v | d_{\text{test}}) = \sum_{k=1}^K P(v | T_k) P(T_k | d_{\text{test}}), \quad (2)$$

式中,  $v$  为测试集图像中文本标注词,  $d_{\text{test}}$  为测试集图像,  $k$  为当前主题数,  $T_k$  为第  $k$  个图像主题。

### 3 改进的卷积神经网络

#### 3.1 基于迁移学习的卷积神经网络

用来进行图像特征学习和图像文本主题多标签分类学习的模型是 AlexNet<sup>[14]</sup> 模型,它曾是 ImageNet<sup>[15]</sup> 图像分类比赛中第一名。AlexNet 模型的网络结构有 5 个卷积层、3 个池化层、3 个全连接层,所有网络共计 11 层。CNN 需要大量的图像数据进行训练以得到较好的模型参数,但是这种大规模数据训练对实验环境的要求高,耗费时间长,而采用小规模的数据集进行训练又易导致网络过拟合。因此采用迁移学习<sup>[16]</sup> 的思想,即获得 CNN 在大规模数据集 ImageNet 上的模型参数,将其迁移到 CNN 上进行初始化,然后利用指定图像数据集对最后一层网络的模型参数进行微调,同时为适应对指定实验数据集的文本主题进行多标签分类的任务,需要修改网络输出层的损失函数。最后,使用完成微调的 CNN 模型提取图像数据集的高层特征同时进行图像文本主题多标签分类。基于迁移学习的 CNN 模型结构如图 2 所示。

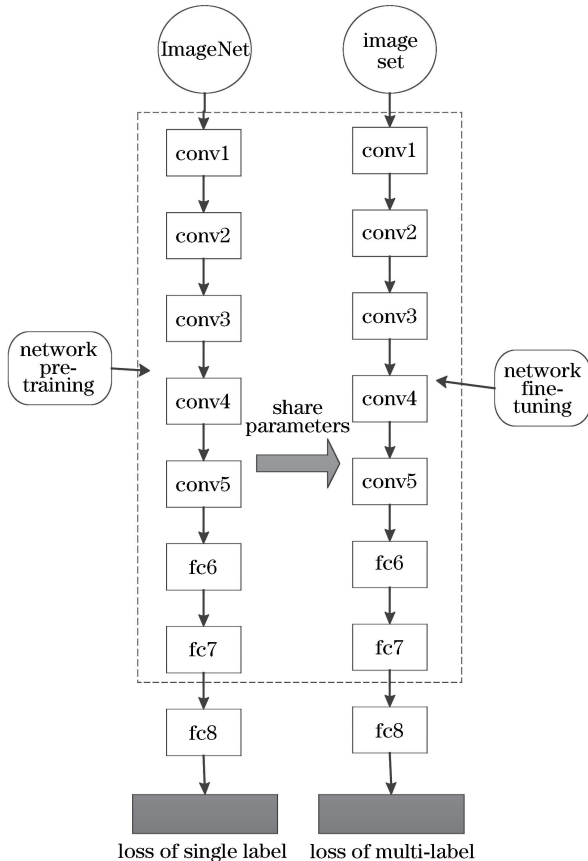


图 2 基于迁移学习的卷积神经网络结构

Fig. 2 Structure of CNN based on transfer learning

图像在输入 CNN 之前需要裁剪为指定大小  $227 \times 227$ 。在图 2 中,conv $i$  表示卷积层  $i$ ,fc $i$  表示全连接层  $i$ , $i$  表示数字。CNN 中所有网络层的参数设置如表 2 所示,其中  $F$ 、 $S$ 、 $P$  分别表示卷积池化窗的大小、窗口滑动的步长、图像的边界填充, $K_f$  表示在当前网络层中卷积池化窗的个数, $D_f$  表示当前网络层输出特征的维度。当网络层为全连接层时,全连接 fc6 和 fc7 输出的特征维度为 4096,而最后一个全连接层 fc8 输出的特征维度为  $N_{df}$ 。CNN 的初始化参数如下:动量设置为 0.5,参数衰减量为 0.0002,学习率为 0.001。由于 CNN 模型的图像文本主题多标签分类学习任务,该网络输出特征维度  $N_{df}$  由实验指定的图像数据集 Corel5K 和 IAPR TC-12 的文本主题的个数决定。

表 2 CNN 各层参数设置

Table 2 Parameters of different layers of CNN

Type of network layer	$K_f$	$F$	$S$	$P$	$D_f$
conv1	96	$11 \times 11$	4	0	$55 \times 55 \times 96$
Max-Pooling1	—	$3 \times 3$	2	0	$27 \times 27 \times 96$
conv2	256	$5 \times 5$	1	2	$27 \times 27 \times 256$
Max-Pooling2	—	$3 \times 3$	2	0	$13 \times 13 \times 256$
conv3	384	$3 \times 3$	1	1	$13 \times 13 \times 384$
conv4	384	$3 \times 3$	1	1	$13 \times 13 \times 384$
conv5	256	$3 \times 3$	1	1	$13 \times 13 \times 256$
Max-Pooling5	—	$3 \times 3$	2	0	$6 \times 6 \times 256$

#### 3.2 基于多标签分类的损失函数

CNN 中损失层的损失函数主要适用于单标签分类,但本文在 CNN 中需要进行图像文本主题多标签分类。因此需要将 CNN 的分类训练集由图像的高层视觉特征对应单个标签改为对应多个文本主题,并对原有的损失函数进行改进。CNN 在进行单标签分类时,选用 softmax 交叉熵函数作为其损失函数,其公式为

$$p_i = \frac{\exp(a_i)}{\sum_{n=1}^{N_t} \exp(a_n)}, \quad (3)$$

$$L = - \sum_{i=1}^{N_t} y_i \ln(p_i). \quad (4)$$

为使 CNN 适应多标签分类,其损失函数需要计算训练集中图像的多个文本主题的全部损失,即对一张图像对应的多个潜在文本主题的损失进行累加。改进的损失函数为

$$L = - \sum_{j=1}^{K_m} \sum_{i=1}^{N_t} y_{ji} \ln(p_i), \quad (5)$$

式中:  $N_i$  为图像文本主题的个数;  $a_n$  为图像训练样本  $x$  对应的高层特征向量的特征值;  $a_i$  为图像训练样本  $x$  对应的高层特征向量第  $i$  维的特征值;  $p_i$  为 CNN 预测训练样本  $x$  属于第  $i$  个文本主题的概率;  $L$  为交叉损失函数;  $y_i \in \{0, 1\}$  为训练样本  $x$  对应一个标签时的标签值;  $K_m$  为训练样本  $x$  对应多个标签时标签的个数;  $y_{ji}$  为训练样本  $x$  对应多个标签时第  $j$  个标签的标签值。

图像的文本主题分布基于图像训练集文本标注词生成, 因此图像训练集的文本标注词分布不平衡会影响图像生成的对应文本主题, 使得该文本主题分布不平衡。由于图像训练集的文本主题分布不平衡, CNN 在进行图像文本主题多标签分类学习时, 低频文本主题分类学习的准确性要远低于高频文本主题。为提高分类学习中低频文本主题的准确率, 在图像训练集中的高频标签中加入噪声进行平滑处理<sup>[17]</sup>, 在损失函数中加入频率系数, 改进(5)式, 表示为

$$L = - \sum_{j=1}^{K_m} \sum_{i=1}^{N_t} \alpha_k y_{ji} \ln(p_i), \quad (6)$$

式中:  $\alpha_k$  表示图像对应的多个文本主题中第  $k$  个主题的相关频率系数。  $\alpha_k = \frac{1}{m} \sum_{i=1}^m \frac{n_i}{M}$ ,  $n_i$  表示在对应的文本主题中第  $i$  个标注词在图像训练集中出现的次数,  $m$  表示对应的文本主题中标注词的个数,  $M$  表示图像训练集的样本数。当  $\alpha_k$  的值较大时, 依据

$$y'_{ji} = \alpha_k y_{ji} + \frac{(1 - \alpha_k)}{N_t} s_{\text{sign}}(y_{ji}), \quad (7)$$

对原有的文本主题  $y_{ji}$  进行替换, 式中  $s_{\text{sign}}$  为符号函数。图像对应的高频原始文本主题由其对应的频率系数来调整权重, 即在高频文本主题中加入噪声。

### 4 图像标注框架

在 CNN 和 LDA 主题模型的基础上, 提出融合 CNN 和主题模型的图像标注方法。利用 LDA 主题模型对图像训练集的文本标注词进行建模, 生成图像训练集的文本主题分布和文本主题的标注词分布。该处理充分利用主题模型的优势, 降低图像文本数据的维度和稀疏性。考虑到判别式分类算法在图像稀疏标签分类中的优势, 利用 CNN 的分类器进行图像文本主题多标签分类, 可获取比生成式主题模型中更加准确的潜在文本主题分布。由于传统特征的表达能力弱和提取过程复杂, 这里采用 CNN 模型提取图像的高层视觉特征。在 CNN 完

成图像特征学习后, 改进 CNN 的分类器使其适应多标签分类, 利用图像的高层特征及其对应的多个文本主题构造分类器, 进行图像的文本主题多标签学习。图像标注的框架如图 3 所示。

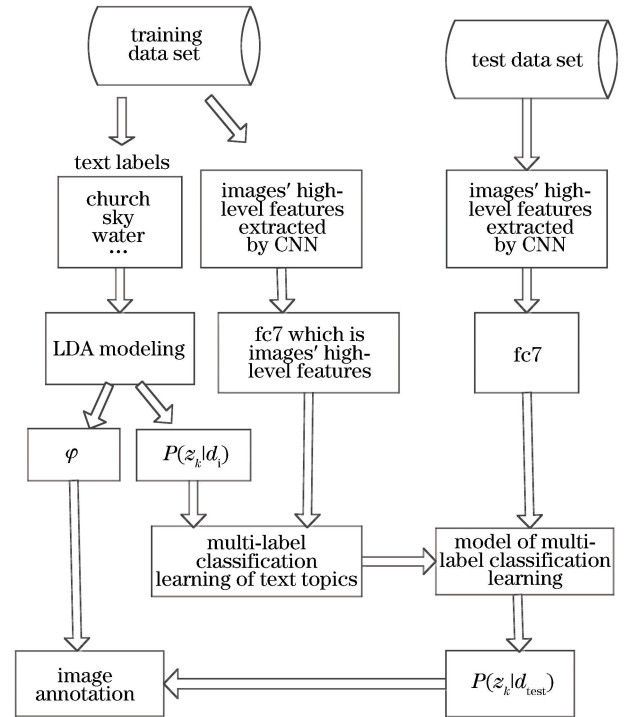


图 3 融合 CNN 和主题模型的图像标注框架

Fig. 3 Framework of image annotation that combines CNN and topic model

由图 3 可看出, 图像语义标注方法的训练过程可分为 2 部分:

1) 采用 LDA 主题模型为图像训练集的文本标注词建模, 生成图像文本模态的主题分布  $P(z_k | d_i)$  和文本主题标注词分布  $P(v_n | z_k)$  即  $\varphi$ 。文本主题标注词分布对非训练集的图像数据同样有效, 而图像文本模态的主题分布  $P(z_k | d_i)$  只是针对训练集中对应图像的文本标注词。在 LDA 主题模型上完成图像训练集的文本标注词建模后, 可由图像对应生成的文本主题代表其文本标注词。

2) 利用完成微调的 CNN 提取图像的高层视觉特征, 此特征表达的图像视觉信息比传统手工特征更加全面。由于图像训练集的文本主题分布稀疏且判别式分类处理图像标签的稀疏分布更具有优势, 利用训练集中图像的高层语义特征及其文本主题构造分类器, 其中每一类代表一个图像文本模态的主题。同时, 每张训练图像可由一个图像的高层特征向量表示, 其文本模态对应多个文本主题, 将 CNN 模型分类器改为多标签分类器, 在此过程中也需

注意改善训练集图像的文本主题分布不平衡的问题。

在图像标注阶段,即标注测试图像的过程可分为2步:

1) 利用完成微调的 CNN 提取测试图像的高层特征。然后利用完成训练的 CNN 的分类器对测试图像的高层特征进行分类,得到该测试图像的文本主题分布。

2) 将测试图像的文本主题分布和由主题模型生成的文本主题标注词分布按照主题模型的图像标注(2)式进行计算,以得到测试图像的文本标注词概率,选取概率最大的5个标注词作为测试图像的标注结果。

## 5 实验结果分析

### 5.1 实验设置

为验证本文标注方法的有效性,分别在两个图像标注数据集 Corel5K 与 IAPR TC-12 上进行实验。Corel5K 图像数据集共有 4999 张彩色图像,图像大小为  $192 \times 168$ ,每张图像有 1~5 个标签。其中,4500 张图像用于训练,剩余的 499 张用于测试。IAPR TC-12 数据集共有 19627 张  $480 \times 360$  大小的彩色图像,平均每张图像有 5.7 个标注。在此数据集中用于训练的图像个数为 17665,用于测试的图像个数为 1962。另外,验证图片包含于训练图像集。实验结果的评价指标是平均查准率(AP)、平均查全率(AR)和综合评价指标 F1<sup>[18]</sup>。

### 5.2 实验参数设置

实验中的 CNN 是在 Caffe 深度学习框架的基础上运行的,并且 CNN 在运行过程中使用 NVIDIA K620 GPU 和 cuDNN 进行计算和加速<sup>[17]</sup>。在图像标注方法中,图像训练集中文本主题的个数  $K$  影响图像标注效果。为找到文本主题数  $K$ ,使图像标注方法的标注性能达到最佳,需在图像数据集上进行多次实验。在图像数据集 Corel5K 中,当文本主题数  $K=80$  时,图像自动标注实验的标注性能达到最佳;在图像数据集 IAPR TC-12 中,当文本主题数  $K=80$  时,图像标注方法的标注性能最好。

### 5.3 实验结果

为验证本文图像标注算法的有效性,在图像数据集上将本文图像标注方法与当前一些传统的图像标注方法进行比较,如 MBRM<sup>[7]</sup>模型、JEC<sup>[8]</sup>模型、及改进的 TagProp-ML 方法<sup>[11]</sup>、2PKNN 方法<sup>[12]</sup>。

同时,与使用深度 CNN 的图像标注方法进行比较,如 CNN-R 模型<sup>[12]</sup>、CNN-MSE 模型<sup>[13]</sup>。为验证图像高层视觉特征对图像标注的积极影响,在 Corel5k 数据集上,取不同图像视觉特征(传统特征和高层特征)的传统标注模型 PLSA-WORDS<sup>[6]</sup>和 HGMD<sup>[17]</sup>与本文图像标注方法进行对比。

表 3 为传统标注方法 PLSA-WORDS 和 HGMD 使用传统图像特征(TVS)和 CNN 提取的图像高层特征 fc7 时,与本文图像标注模型在 AR、AP 及综合评价指标 F1 中的对比。在同一模型中对比不同图像特征的标注性能,可以看出图像高层特征明显提升对传统图像标注模型的标注性能。本文融合 CNN 和主题模型的图像标注方法与使用图像高层特征的传统图像标注方法相比,其召回率、准确率和 F1 值都有了大幅度的提升。

表 3 模型在 Corel5K 上的标注结果

Table 3 Annotation results of different models on Corel5K

Model	Visual feature	AP	AR	F1
PLSA-WORDS	TVS	0.121	0.221	0.191
	fc7	0.217	0.275	0.269
HGDM	TVS	0.293	0.321	0.263
	fc7	0.305	0.364	0.297
Proposed model	fc7	0.380	0.490	0.420

通过表 4 可以看出,在 Corel5K 和 IAPR TC-12 图像集上,本文图像标注模型较传统图像语义标注方法均有较大提升。在 Corel5K 图像数据集中,本文标注方法的平均查全率较 MBRM 提高了 24%,较 JEC 提高了 17%;对比改进的传统方法,本文图像标注方法的平均查准率低于 2PKNN,但平均查全率却提升了 9%,并且对应 F1 值亦有所提高,这表明本文图像标注模型的有效性;与 CNN-R 相比,图像标注的平均查全率提升了 8%,且平均查准率提升了 6%。在 IAPR TC-12 图像数据集中,

表 4 通用数据集上所有图像标注方法的标注结果

Table 4 Annotation results of all image annotation models on common datasets

Model	Corel5K			IAPR TC-12		
	AR	AP	F1	AR	AP	F1
MBRM	0.25	0.24	0.25	0.23	0.24	0.24
JEC	0.32	0.27	0.29	0.29	0.28	0.29
TagProp-ML	0.37	0.31	0.34	0.25	0.48	0.33
2PKNN	0.40	0.39	0.40	0.32	0.49	0.39
CNN-R	0.41	0.32	0.37	0.31	0.49	0.37
CNN-MSE	0.35	0.41	0.38	0.35	0.40	0.37
Proposed model	0.49	0.38	0.43	0.40	0.44	0.42

本文图像标注方法与改进的传统标注方法及其其他的基于 CNN 的图像标注方法进行比较,在平均查全率上有较好的表现,其平均查准率与部分图像标注方法如 CNN-R 等存在一些差距,但综合两者考虑即综合指标 F1 依旧有提升。

## 6 结 论

为有效进行图像标注,提出融合卷积神经网络与主题模型的图像标注方法。对比其他标注方法,本文充分利用了主题模型在文本处理的优势,降低了图像文本数据的维度和稀疏性,运用卷积神经网络提取图像的高层视觉特征和进行图像文本主题多标签学习,获得更多的图像信息和更准确的文本主题分布。实验表明,本文方法显著提升了图像标注的性能。

### 参 考 文 献

- [1] Peng Y F, Song X N, Zi L L, *et al.* Remote sensing image retrieval based on convolutional neural network and modified fuzzy C-means [J]. *Laser & Optoelectronics Progress*, 2018, 55(9): 091008.  
彭晏飞, 宋晓男, 訾玲玲, 等. 基于卷积神经网络和改进模糊 C 均值的遥感图像检索[J]. *激光与光电子学进展*, 2018, 55(9): 091008.
- [2] Liu Y, Zhang D S, Lu G J, *et al.* A survey of content-based image retrieval with high-level semantics[J]. *Pattern Recognition*, 2007, 40(1): 262-282.
- [3] Li J, Wang J Z. Automatic linguistic indexing of pictures by a statistical modeling approach[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2003, 25(9): 1075-1088.
- [4] Cusano C, Ciocca G, Schettini R. Image annotation using SVM[J]. *Proceedings of SPIE*, 2003, 5301: 330-338.
- [5] Blei D M, Jordan M I. Modeling annotated data[C]// *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval*, July 28-August 1, 2003, Toronto, Canada. New York: ACM, 2003: 127-134.
- [6] Monay F, Gatica-Perez D. Modeling semantic aspects for cross-media image indexing [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007, 29(10): 1802-1817.
- [7] Guillaumin M, Mensink T, Verbeek J, *et al.* TagProp: discriminative metric learning in nearest neighbor models for image auto-annotation[C]// *2009 IEEE 12th International Conference on Computer Vision*, September 29-October 2, 2009, Kyoto, Japan. New York: IEEE, 2009: 309-316.
- [8] Makadia A, Pavlovic V, Kumar S. Baselines for image annotation [J]. *International Journal of Computer Vision*, 2010, 90(1): 88-105.
- [9] Guo C C, Yu F Q, Chen Y. Image semantic segmentation based on convolutional neural network feature and improved superpixel matching[J]. *Laser & Optoelectronics Progress*, 2018, 55(8): 081005.  
郭呈呈, 于凤芹, 陈莹. 基于卷积神经网络特征和改进超像素匹配的图像语义分割[J]. *激光与光电子学进展*, 2018, 55(8): 081005.
- [10] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks [J]. *Communications of the ACM*, 2017, 60(6): 84-90.
- [11] He K M, Zhang X Y, Ren S Q, *et al.* Deep residual learning for image recognition [C]// *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 770-778.
- [12] Murthy V N, Maji S, Manmatha R. Automatic image annotation using deep learning representations[C]// *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval*, June 23-26, 2015, Shanghai, China. New York: ACM, 2015: 603-606.
- [13] Gao Y D, Hou L Y, Yang D L. Automatic image annotation method using multi-label learning convolutional neural network [J]. *Journal of Computer Applications*, 2017, 37(1): 228-232.  
高耀东, 侯凌燕, 杨大利. 基于多标签学习的卷积神经网络的图像标注方法[J]. *计算机应用*, 2017, 37(1): 228-232.
- [14] Ma Y J, Li X Y, Song X F. Traffic sign recognition based on improved deep convolution neural network[J]. *Laser & Optoelectronics Progress*, 2018, 55(12): 121009.  
马永杰, 李雪燕, 宋晓凤. 基于改进深度卷积神经网络的交通标志识别[J]. *激光与光电子学进展*, 2018, 55(12): 121009.
- [15] Russakovsky O, Deng J, Su H, *et al.* ImageNet large scale visual recognition challenge [J]. *International Journal of Computer Vision*, 2015, 115(3): 211-252.
- [16] Zhuang F Z, Luo P, He Q, *et al.* Survey on transfer

- learning research [J]. Journal of Software, 2015, 26(1): 26-29.
- 庄福振, 罗平, 何清, 等. 迁移学习研究进展[J]. 软件学报, 2015, 26(1): 26-29.
- [17] Li Z X, Zheng Y Z, Zhang C L, *et al.* Combining deep feature and multi-label classification for semantic image annotation [J]. Journal of Computer-Aided Design & Computer Graphics, 2018, 30(2): 318-326.
- 李志欣, 郑永哲, 张灿龙, 等. 结合深度特征与多标记分类的图像语义标注[J]. 计算机辅助设计与图形学学报, 2018, 30(2): 318-326.
- [18] Wang P, Zhang A F, Wang L Q, *et al.* Image automatic annotation based on transfer learning and multi-label smoothing strategy [J]. Journal of Computer Applications, 2018, 38(11): 3199-3203, 3210.
- 汪鹏, 张奥帆, 王利琴, 等. 基于迁移学习与多标签平滑策略的图像自动标注[J]. 计算机应用, 2018, 38(11): 3199-3203, 3210.