光学 精密工程
OPTICS AND PRECISION ENGINEERING
核心期刊
主管：中国科学院
主办：中国科学院长春光学精密机械与物理研究所
中国仪器仪表学会
主编：曹健林

多标签分类的传统民族服饰纹样图像语义理解

赵海英，周伟，侯小刚，齐光磊

引用本文:
赵海英, 周伟, 侯小刚, 等. 多标签分类的传统民族服饰纹样图像语义理解[J]. 光学 精密工程, 2020, 28(3): 695-703.
ZHAO Hai-ying, ZHOU Wei, HOU Xiao-gang, et al. Multi-label classification of traditional national costume pattern image semantic understanding[J]. *Optics and Precision Engineering*, 2020, 28(3): 695-703.

在线阅读 View online: https://doi.org/10.3788/OPE.20202803.0695

文章编号 1004-924X(2020)03-0695-09

多标签分类的传统民族服饰纹样图像语义理解

赵海英[1*]，周 伟[2]，侯小刚[3]，齐光磊[4]
(1. 北京邮电大学 计算机学院，北京 100876；
2. 北京邮电大学 数字媒体与设计艺术学院，北京 100876；
3. 北京邮电大学 网络技术研究院，北京 100876；
4. 北京邮电大学 世纪学院，北京 102101)

摘要：针对当前图像多标签分类方法只关注图像本体类别信息(本体)，而忽略图像深层次语义信息(隐义)的问题，本文提出了一种“本体-隐义”融合学习的图像多标签分类模型。该模型首先利用 CNN 中间层和较高层分别学习图像的本体信息和隐义信息，然后利用本体信息与隐义信息之间的依赖关系设计了融合学习模型，同时对提出模型的不同中间层特征和模型的不同结构进行了深入研究，最终实现了对图像中多类别以及各类别蕴含的隐义信息分类。在传统民族服饰纹样图像数据集上进行实验，得到图像本体多标签分类和隐义多标签分类的 mAP 分别为 0.88 和 0.82；在 Scene 数据集上进行对比实验，本文模型在 Hamming loss，One-error 以及 Average precision 指标上分别优于其他最好方法 0.103，0.091 和 0.083，实验结果证明了本文方法的有效性和优越性。

关 键 词：多标签分类；融合学习；传统民族服饰；语义理解
中图分类号：TP391 **文献标识码**：A **doi**：10.3788/OPE.20202803.0695

Multi-label classification of traditional national costume pattern image semantic understanding

ZHAO Hai-ying[1*], ZHOU Wei[2], HOU Xiao-gang[3], QI Guang-lei[4]

(1. *School of Computer Science, Beijing University of Posts and Telecommunications, Beijing* 100876, *China*;
2. *School of Digital Media and Design Art, Beijing University of Posts and Telecommunications, Beijing* 100876, *China*;
3. *Research Institute of Network Technology, Beijing University of Posts and Telecommunications, Beijing* 100876, *China*;
4. *Century College, Beijing University of Posts and Telecommunications, Beijing* 102101, *China*)
* *Corresponding author, E-mail: zhaohaiying@bupt.edu.cn*

Abstract: Since current image multi-label classification methods only focus on the category information of image ontology (ontology) and ignore the deep semantic information of the image (implicit), this study proposed an image multi-label classification model of “ontology-implicit” fusion learning. The

收稿日期：2019-09-16；**修订日期**：2019-11-26.
基金项目：中央文化产业发展专项资金申报项目资助(No. GSSKS-2015-035)；国家自然科学基金资助项目(No. 61163044)；国家社会科学基金重大研究专项资助项目(No. 18VDL001)；北京市科技计划课题资助项目(No. D171100003717003)

model first used the middle and higher layers of CNN to learn the image ontology information and implicit information, respectively, and then it used the dependency relationship between the ontology information and implicit information to design the fusion learning model. Meanwhile, the different characteristics of the middle layer and different structures of the model were studied in-depth, to realize the classification of implicit information contained in multiple image categories. Experiments conducted on the traditional national costume pattern image datasets show that the mAP of image ontology multi-label classification and implicit multi-label classification are 0.88 and 0.82, respectively. Comparative experiments conducted on the Scene dataset show that the model is superior to other methods in Hamming loss, one error, and average precision indices, with values of 0.103, 0.091, and 0.083, respectively. Therefore, the experimental results prove the effectiveness and superiority of this method.

Key words: multi-label classification; fusion learning; traditional national costumes; semantic understanding

1 引　言

传统民族服饰纹样记载着一个民族从建立到发展过程的历史文化演变,在对服饰纹样进行解读时,不仅需要明确纹样的类别(本体),更需要诠释各纹样所具有的深层文化语义信息(隐义)。例如,传统民族服饰中纹样本体“龙”是古代皇帝的象征,隐义是“权势、高贵”;纹样本体“牡丹”被誉为花王,隐义是“富贵、美满”;纹样本体“桃”具有图腾、生殖崇拜的原始信仰,隐义是“长寿、健康”。因此,在对传统民族服饰纹样进行多标签分类时,从“本体”和“隐义”两个层面分类,可以更全面地阐述传统民族服饰纹样所蕴含的文化语义信息。

近年来,基于深度学习的方法在图像分割[1]、识别[2]和检索[3]等一系列计算机视觉任务中取得了巨大的成功。与此同时,基于深度学习的图像多标签分类方法越来越受欢迎。一方面,由于卷积神经网络(Convolutional Neural Networks, CNN)在单标签分类任务的成功,很多研究者直接将 CNN 应用到多标签分类任务上[4-8]。例如 WEI 等[4]以任意数量的对象片段假设作为输入,将共享的 CNN 与每个假设相连,最后将不同假设的 CNN 输出结果用最大池化进行聚合,得到最终的多标签预测。Yu 等[6]将图像的全局先验信息和局部实例信息相结合构建了一个新的双流网络,可以自动定位触发标签的关键图像模式。另一方面,由于循环神经网络(Recurrent Neural Network, RNN)在机器翻译、图像描述以及视觉问题回答等任务的成功应用,一些学者将图像多分类看作是序列生成问题,同时利用 RNN 建立标签之间的依赖关系[9-12]。例如,JIN 等[9]将图像标注任务作为一个序列生成问题,提出 RIA 模型能够根据图像内容对标签的长度进行原生预测,并考虑训练标注序列输入到 LSTM 顺序的影响。WANG 等[10]构建一个端到端的 CNN-RNN 框架,学习图像标签嵌入的方法来表征隐义标签依赖关系和图像标签相关性。

然而,上述方法只能对图像中多个物体类别(本体)进行分类,而不能对同一张图像中各类别所蕴含的深层次语义信息(隐义)进行分类。为解决上述问题,本文提出了一个“本体-隐义”融合学习的图像多标签分类模型,该模型首先利用 CNN 中间层学习图像的本体信息,利用 CNN 较高层学习图像的隐义信息,然后利用本体信息和隐义信息之间的依赖关系设计融合学习模型,实现对图像中多类别以及各类别蕴含的深层语义信息分类。此外,在传统民族服饰纹样图像数据集和 Scene 数据集上进行对比实验,实验结果证明了本文方法在图像多标签分类任务上的有效性和优越性。

2 “本体-隐义”融合学习的多标签分类模型

图像多标签分类模型学习将多个标签分配给一张图像的问题。假设训练集中图像 x_i 对应的本体标签向量为 $y_i=\{0,1\}^m$,对应的隐义标签向量为 $s_i=\{0,1\}^n$,其中 $y_i^j=1$ 表示图像 x_i 的第 j

个本体标签出现，$y_i^j=0$ 表示本体标签缺失；$s_i^j=1$ 表示图像 x_i 的第 j 个隐义标签出现，$s_i^j=0$ 表示隐义标签缺失。其中，m 为本体标签的类别数，n 为隐义标签的类别数（m 和 n 不一定相等），N 为图像的数量。本文构建一种“本体-隐义”融合学习的图像多标签分类模型，即从训练集 $\{(x_i,y_i,s_i)|1\leqslant i\leqslant N\}$ 学习一个映射函数 $f:x\rightarrow(y,s)$，从而可以同时预测图像的本体标签和隐义标签，构建的模型纵览图如图 1 所示。

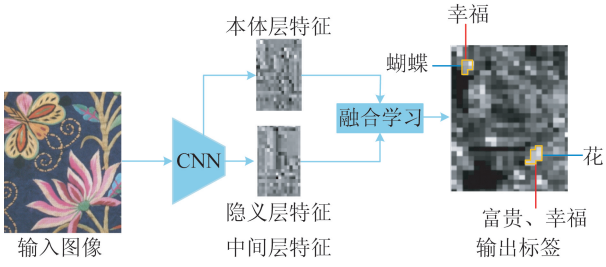

图 1　构建模型纵览图

Fig. 1　Overview of the present model

从“本体”和“隐义”两个层面对传统民族服饰纹样进行分类，如表 1 所示，可以更全面地阐述传统民族服饰纹样所蕴含的文化语义信息。

2.1　模型结构

CNNs 可以学习丰富的图像层次特征，例如 AlexNet 模型前两层学习的是颜色、边缘等低层特征，第 3 层学习的是较复杂的纹理特征，第 4 层学习的是特定类别的局部特征，第 5 层学习的是具有辨别性的完整特征[13]，即网络的低层特征包含更多的图像结构信息，中间层因卷积核感受野小且个数多，更容易学习图像的区域或局部特征，而高层特征更关注图像的语义信息。

表 1　图像的本体标签和隐义标签多分类

Tab. 1　Multi-classification of ontology labelsand implicit labels for images

传统民族服饰纹样图像数据集	本体标签	隐义标签
	花	富贵 幸福
	龙 蝙蝠 云	权势 幸福 吉祥

在本文中，本体信息描述图像中存在的物体类别，而隐义信息诠释图像所蕴含的深层次文化语义，与本体信息相比较，隐义信息需要考虑图像中存在的物体类别、组合规则等信息，从而需要更高层的特征进行表征。因此，在同一个网络中，可以利用网络的中间层学习图像的本体信息，高层学习图像的隐义信息。然后，将本体信息和隐义信息分别利用损失函数进行训练后，采用融合学习的方式更新网络参数，可以捕获图像的本体信息与隐义信息之间的关联性，从而实现了对同一张图像的本体标签分类和隐义标签分类。本文选取 Inception-V3[14] 作为基准模型，构建了一种“本体-隐义”融合学习的图像多标签分类模型，如图 2 所示（彩图见期刊电子版）。为简单直观地表述，图中红色虚线模块为重复模块，“x3”表示此模块按先后顺序重复 3 次，“x4”同理。（注：为便于下文表述，图中红色箭头为对应节点名称，其中 mixed_2 和 mixed_7 分别是最后一个重复模块的节点名称）

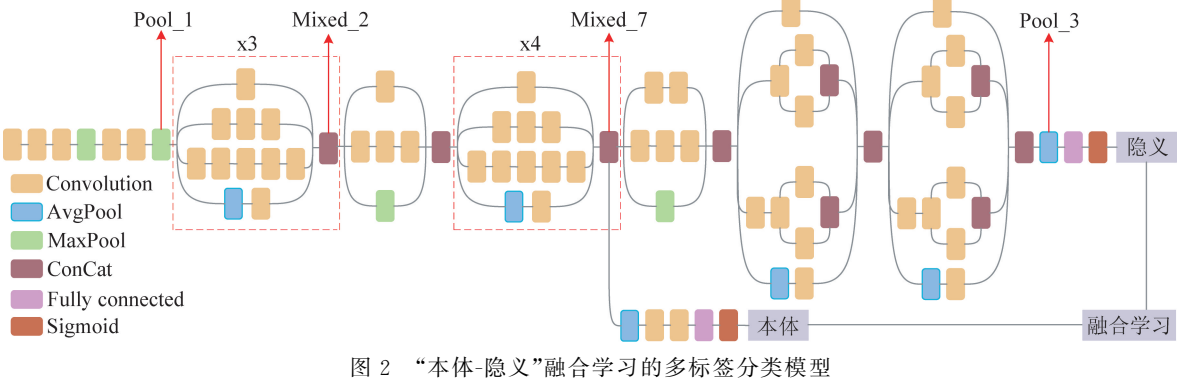

图 2　“本体-隐义”融合学习的多标签分类模型

Fig. 2　Multi-label classification model based on “ontology-implicit” fusion learning

由于 Inception-V3 模型网络层次较深，通常需要在非常大的数据集上经过数百万次训练才能达到较好的效果。因此，本文利用 ImageNet 上预训练的 Inception-V3 在多标签分类任务上微调。本文采用的多标签损失函数为交叉熵损失。其中，采用 Sigmoid 作为激活函数得到输出$h_\theta(\boldsymbol{x})$：

$$h_\theta(\boldsymbol{x})=\frac{1}{1+\exp(-\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{x})}. \tag{1}$$

预测值和真实值概率之间的交叉熵如下：

$$J(\boldsymbol{\theta})=-\frac{1}{N}\Big[\sum_{i=1}^{N}y_i\log h_{\boldsymbol{\theta}}(x_i)+(1-y_i)\log(1-h_{\boldsymbol{\theta}}(x_i))\Big], \tag{2}$$

其中 $y_i\in\{0,1\}$。记本体标签的损失函数为 $J_{\mathrm{y}}(\boldsymbol{\theta})$，隐义标签的损失函数为 $J_{\mathrm{s}}(\boldsymbol{\theta})$，将两者加权求和进行融合学习，得到模型总损失函数为 $J_{\mathrm{sum}}(\boldsymbol{\theta})$，最后进行反向传播降低误差，以获得更好的分类效果：

$$J_{\mathrm{sum}}(\boldsymbol{\theta})=\lambda J_{\mathrm{y}}(\boldsymbol{\theta})+J_{\mathrm{s}}(\boldsymbol{\theta}), \tag{3}$$

其中 λ 为本体标签的损失函数 $J_{\mathrm{y}}(\boldsymbol{\theta})$的权重。

2.2　模型不同中间层特征影响研究

Inception-V3 网络模型深度较深，模型中包含许多卷积操作和池化操作。为了研究 Inception-V3 模型不同位置获取的中间层图像特征对本体标签多分类结果的影响，如图 3 所示，本文对 Inception-V3 模型的 3 个不同位置做出以下探讨。

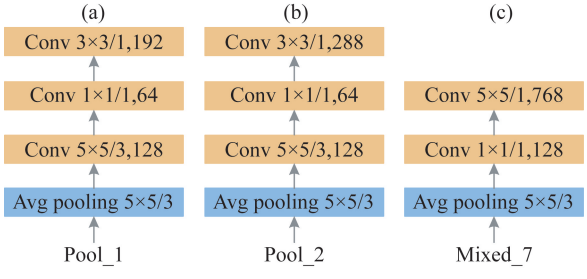

图 3　卷积和池化操作

Fig. 3　Convolution and pooling operations

在图 3(a)中，获取 Inception-V3 网络模型的非 Inception Module 的第 2 个池化层(即 Pool_1)输出特征，在其后面接 1 个平均池化层和 3 个卷积层，然后再接全连接层。在图 3(b)中，在 Inception-V3 模型中的第 1 个 Inception 模块组的输出(即 Mixed_2)后面，连接 1 个平均池化层和 3 个卷积层，然后再接全连接层。章节 2.1 中间层如图 3(c)所示，即在 Inception-V3 模型中的第 2 个 Inception 模块组的输出(即 Mixed_7)后面，连接一个平均池化层和两个卷积层，然后再接全连接层。下文将对不同中间层特征进行实验，对比效果差异。

3　模型不同结构对比研究

在 2.1 节中介绍的“本体-隐义”融合学习的图像多标签分类 CNN 模型基础上，对模型结构进行修改，提出另外 3 种“本体-隐义”图像多标签分类模型，下面对 3 个对比模型进行详细阐述。

3.1　单流网络

如图 4 所示，单流网络将“本体-隐义”标签分类看作是一个简单的图像多标签分类任务，即将本体标签与隐义标签混合为一个整体进行训练。本文将 Inception-V3 作为基线模型，将模型最终预测 1 000 个类别的全连接层替换为预测“本体-隐义”的 $m+n$ 个类别，采用交叉熵损失作为损失函数，记单流网络的损失函数为 $J_{\mathrm{a}}^{\mathrm{y-s}}(\boldsymbol{\theta})$。

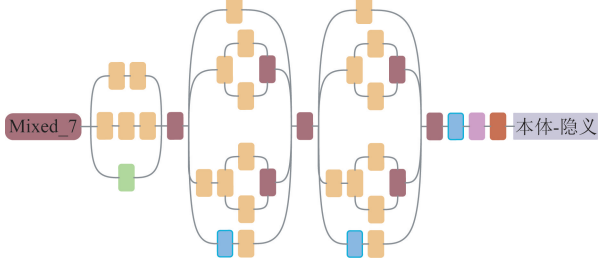

图 4　单流网络

Fig. 4　Single stream network

3.2　分流网络

如图 5 所示，分流网络共享 Inception-V3 模型早期特征，在 Pool_3 处将本体和隐义拆分为两个单独的处理流，分别学习本体和隐义的多标签分类器。分流网络使用交叉熵损失作为损失函数，记本体标签损失函数为 $J_{\mathrm{b}}^{\mathrm{y}}(\boldsymbol{\theta})$，隐义标签的损失函数为 $J_{\mathrm{b}}^{\mathrm{s}}(\boldsymbol{\theta})$。

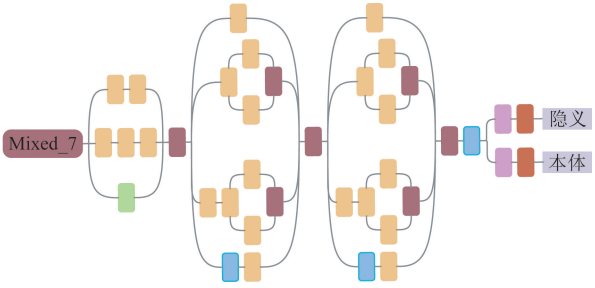

图 5　分流网络

Fig. 5　Shunt network

3.3　中间层分流网络

如图 6 所示，中间层分流网络利用中间层特征来表征图像本体特征信息，即在第 2 个 Inception 模块组的输出（Mixed_7 处）后面，进行池化和卷积操作，将其输出特征作为本体标签的图像特征，然后连接全连接层，使输出特征为维度 m 的特征向量。由于第 3 个 Inception 模块组中包含许多卷积和池化操作，得到的图像特征更加抽象，故将后面连接的池化层输出（Pool_3 处）取出，作为隐义标签的图像特征，然后连接全连接层，使输出特征为维度 n 的特征向量。最后分别使用交叉熵损失作为损失函数进行训练，记该网络的本体标签损失函数为 $J_c^y(\boldsymbol{\theta})$，隐义标签的损失函数为 $J_c^s(\boldsymbol{\theta})$。中间层分流网络与 2.1 节中提出的网络模型的差别在于本体标签损失函数与隐义标签损失函数没有进行融合学习。

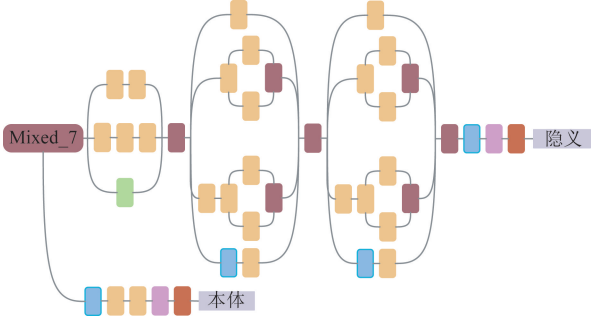

图 6　中间层分流网络

Fig. 6　Intermediate shunt network

4　实验结果与分析

4.1　实验数据与评价指标

本文在传统民族服饰纹样图像数据集和 Scene 数据集上进行多标签分类的对比实验。传统民族服饰纹样图像数据集由本实验室构建，共有 3 000 张图像，每张图像均含有本体和隐义两层标签。本体标签包含 8 个不同类别，即花、桃、鸟、龙、蝴蝶、蝙蝠、祥云以及人物；隐义标签包含有 7 个不同类别，即富贵、喜庆、权势、幸福、典故、吉祥以及长寿。另外，Scene 数据集[15]共有 2 000 张图像，包含沙漠、山脉、海洋、日落和树木类等 5 类自然场景。另外，将两个数据集分别按 6∶2∶2 的比例划分为训练集、验证集和测试集。两个数据集的统计信息如表 2 所示，其中“>1 标签”表示同时含有多个标签的图像在数据集中所占比例大小。

表 2　数据集统计信息

Tab. 2　Statistics of data sets

数据集名称	标签类别数	>1 标签/%	数据集大小
传统民族服饰纹样图像	本体标签：8 隐义标签：7	37	3 000
Scene	5	22	2 000

在传统民族服饰纹样图像数据集上，使用和文献[9]同样的评价指标，利用公式(4)计算总体的精确率、召回率和 F1 值（O-P、O-R、O-F1），以及每个类别的精确率、召回率和 F1 值（C-P、C-R、C-F1）。同时参考文献[5]，本文对每个类别也采用平均精度（Average Precision，AP），对于总体也采用平均精度均值（mAP）。

$$\begin{aligned} OP &= \frac{\sum_i N_i^c}{\sum_i N_i^p}, OR = \frac{\sum_i N_i^c}{\sum_i N_i^g}, \\ OF1 &= \frac{2 \times OP \times OR}{OP + OR}, \\ CP &= \frac{1}{C}\sum_i \frac{N_i^c}{N_i^p}, CR = \frac{1}{C}\sum_i \frac{N_i^c}{N_i^g}, \\ CF1 &= \frac{2 \times CP \times CR}{CP + CR}, \end{aligned} \tag{4}$$

其中：C 是标签的总数，N_i^c 是正确预测 i 个标签的图像数量，N_i^p 是第 i 个标签的预测图像的数量，N_i^g 是第 i 个真实图像的数量。

在 Scene 数据集上，实验采用与 NGRM-ℓ1 方法[20]同样的评价指标，即包括汉明损失（Hamming loss）、1-错误率（One-error）、覆盖率（Coverage）、排序损失（Rank loss）以及平均精度（AP）。

在实验中，本文使用 TensorFlow 深度学习框架，利用在 ImageNet2012 分类挑战数据集上预训练的 Inception-v3 作为基础模型。模型训练的初始学习率为 0.001，共训练 15 000 步，在第 10 000 步时学习率降低为之前的 1/10，动量率为 0.9，权重衰减率为 0.000 5，使用随机梯度下降法(Stochastic Gradient Descent，SGD)进行优化，batch_size 设置为 100，同时将原始图像大小调整为 299×299 作为模型的输入。

为标记简洁，将单流网络、分流网络、中间层分流网络以及 2.1 章节中提出的模型依次记为 M1，M2，M3 以及 M4，同时考虑 Inception-v3 中间层的影响，将 Pool_1，Mixed_2 和 Mixed_7 输出依次记为 A，B 和 C，并将 Pool_3 的输出记为 D。在 Scene 数据集上进行对比实验时，遵循 NGRM-ℓ1 方法实验设置，即使用标准 5-折交叉验证进行评估，报告 5 次实验的平均性能，此外对比方法还有 MLR[16]，MIMLfast[17]，KISAR[18] 和 MIMLcaus[19]。

4.2 实验结果

4.2.1 模型不同中间层特征对比结果

通过对“本体-隐义”融合学习的图像多标签分类模型(M4)的不同中间层特征进行研究，在传统民族服饰纹样图像数据集上的实验结果如表 3 所示，同时 M4-A、M4-B 和 M4-C 的参数量(单位：$\times 10^5$)分别是 26.32，26.68 和 28.15。从表中可以看出，在 AP 的大多数指标上，M4-C 的结果优于 M4-A 和 M4-B，并且在 mAP 指标上获得了最好的结果，尽管 M4-C 模型参数量相较于 M4-A 和 M4-B 有所增加，但性能的提升是可观的。因此，在本文的后续实验中，将采用 Mixed_7 作为中间层输出。

表 3 模型 M4 不同中间层特征的 AP 和 mAP 对比结果

Tab. 3 Comparison of AP and mAP with different intermediate layer characteristics in M4 model

方法	花	鸟	龙	蝴蝶	蝙蝠	人物	祥云	桃	mAP
M4-A	0.58	**0.82**	**0.96**	**0.81**	0.74	0.75	0.83	0.81	0.79
M4-B	0.74	0.76	0.92	0.83	0.79	0.86	0.76	0.87	0.82
M4-C	**0.81**	0.79	0.94	0.75	**0.84**	**0.87**	**0.91**	**0.92**	**0.85**

注：黑色加粗为单列指标最好结果。

4.2.2 模型融合学习参数选择

由公式(3)可知，λ 为模型 M4 中本体标签损失函数的权重值。将模型 M4 在传统民族服饰纹样图像数据集进行实验，并且 λ 在[0.4，1.6]内取值，实验结果如图 7 所示(彩图见期刊电子版)。从图中可以看出，在 λ 的取值范围内本体标签分类的 mAP 值(红线)显著高于隐义标签分类的 mAP 值(绿线)，另外当 λ 取 0.8 或 1.2 时，本体和隐义的 mAP 值之和取得最大值(蓝线)。考虑到隐义标签分类的 mAP 值较小，在误差反向传播时应赋予较大的权重，因此 λ 取值为 0.8。

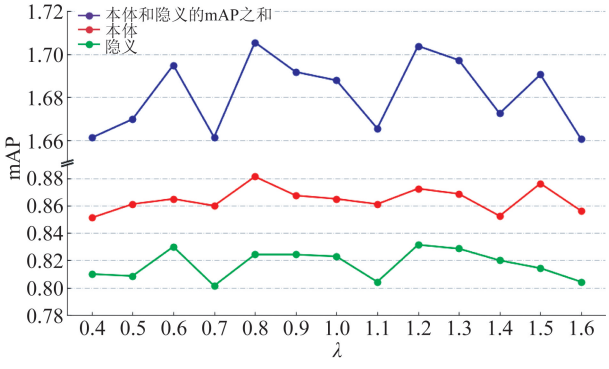

图 7 不同 λ 对应的本体标签和隐义标签分类的 mAP

Fig. 7 mAP of ontology label and implicit label classification corresponding to different λ

4.2.3 模型不同结构对比结果

本文比较了提出的 4 种模型结构，在传统民族服饰纹样图像数据集上的实验结果如表 4 所示(由于 M1 是将本体标签和隐义标签看作一个整体进行训练，得到的是整体 C-P，C-R，C-F1，O-P，O-R，O-F1 以及 mAP 值，为方便直观与其它网络结果进行比较，将其看作是本体标签分类和隐义标签分类的结果)。从表中可以看出，无论是本体标签分类结果还是隐义标签分类结果，M4 在大多数指标上获得了最好的结果，并且在 mAP 指标上明显优于其它方法。另外，比较 4 个模型的参数量(单位：$\times 10^5$)，模型 M4 的参数量相较于模型 M1 和模型 M2 有所增加，但模型性能的提高是巨大的，如在本体标签分类的指标 mAP 上分别提高 10% 和 5%，同时比较模型 M4 和 M3 的参数量，可以发现参数量几乎相等(数据只保留两位小数)，但采用融合学习使得在隐义标签分类的指标 mAP 上提高 1%。因此，将本体信息和隐义信息进行融合学习，可以很好地捕获两种之间的关联性，更好地提高模型分类的性能。

表 4　4 种模型结构对比结果

Tab. 4　Structural comparison of four models

方法	参数量/$\times 10^5$	本体标签分类							隐义标签分类						
		C-P	C-R	C-F1	O-P	O-R	O-F1	mAP	C-P	C-R	C-F1	O-P	O-R	O-F1	mAP
M1	25.60	0.47	**0.77**	0.59	0.45	**0.82**	0.58	0.78	0.47	0.77	0.59	0.45	**0.82**	0.58	0.78
M2	25.63	0.50	0.64	0.56	0.51	0.69	0.59	0.83	0.53	0.72	0.61	0.49	0.76	0.60	0.79
M3	28.15	0.55	**0.77**	0.64	0.55	0.80	0.65	0.85	0.58	**0.80**	**0.67**	0.52	0.80	0.63	0.81
M4	28.15	**0.64**	0.75	**0.69**	**0.64**	0.76	**0.70**	**0.88**	**0.59**	0.77	**0.67**	**0.55**	0.78	**0.65**	**0.82**

注：黑色加粗为单列指标最好结果。

4.2.4　Scene 数据集实验对比结果

为验证本文提出模型的图像多标签分类效果，在公开的 Scene 数据集上与其他方法进行性能比较。实验对比结果如表 5 所示，其中“↓”表示“越小越好”，“↑”表示“越大越好”。可以看出，本文方法 M4-D 在 Hamming loss、One-error 和 Average precision 指标上分别优于其他最好方法 0.103，0.096 和 0.083。在 Coverage 和 Rank loss 指标上与 NGRM-ℓ1(SVM)方法性能接近，可以表明本文方法的优越性。同时，本文 M4-C 方法在 Hamming loss 和 Average precision 指标上优于 NGRM-ℓ1(3NN)方法，验证了 CNN 中间层的有效性，即 CNN 中间层能够有效学习图像特征。

表 5　图像多标签分类方法的性能比(mean±std)

Tab. 5　Performance ratio of multi-label image classification method(mean±std)

方法	Hamming loss ↓	One-error ↓	Coverage ↓	Rank loss ↓	Average precision ↑
MLR[16]	0.268±0.012	0.377±0.038	1.137±0.076	0.210±0.018	0.750±0.021
MIMLfast[17]	0.199±0.013	0.369±0.032	1.119±0.079	0.209±0.017	0.754±0.017
KISAR[18]	0.187±0.011	0.365±0.035	1.069±0.101	0.198±0.023	0.761±0.022
MIMLcaus[19]	0.178±0.010	0.333±0.039	0.989±0.077	0.179±0.018	0.783±0.021
NGRM-ℓ1(3NN)[20]	0.187±0.011	0.350±0.017	0.995±0.012	0.182±0.014	0.795±0.026
NGRM-ℓ1(SVM)[20]	0.175±0.004	0.355±0.016	**0.980±0.026**	**0.174±0.007**	0.794±0.011
M4-A	0.154±0.002	0.540±0.019	1.606±0.191	0.331±0.050	0.745±0.024
M4-B	0.113±0.005	0.448±0.030	1.450±0.072	0.296±0.021	0.797±0.015
M4-C	0.093±0.004	0.360±0.007	1.276±0.091	0.238±0.020	0.839±0.012
M4-D	**0.072±0.002**	**0.259±0.007**	0.991±0.039	0.180±0.005	**0.877±0.003**

注：黑色加粗为单列指标最好结果。

5　结　论

本文提出了“本体-隐义”融合学习的图像多标签分类模型。该模型能够模仿人类的方式观察图像，既能对图像中物体类别信息(本体)进行分类，又能识别各个物体类别所蕴含的深层次语义信息(隐义)。该模型首先利用 CNN 中间层和较高层分别学习图像的本体信息和隐义信息，然后利用本体信息与隐义信息之间的依赖关系设计融合学习模型，从而实现对图像中多类别以及各类别蕴含的深层语义信息分类。在传统民族服饰纹样图像数据集进行实验，结果表明模型的中间层特征能够有效表示图像的本体信息，利用融合学

习可进一步提高分类的准确性;在 Scene 数据集上进行实验,结果表明本文方法在指标 Hamming loss、One-error 和 Average precision 上大幅度优于其他方法。在后续的工作中,将尝试利用两个 CNN 网络分别学习本体信息和隐义信息再进行融合学习。

参考文献:

[1] 张坤华,谭志恒,李斌. 结合粒子群优化和综合评价的脉冲耦合神经网络图像自动分割 [J]. 光学 精密工程, 2018, 26(4): 962-970.
ZHANG K H, TAN ZH H, LI B. Automated image segmentation based on pulse coupled neural network with partide swarm optimization and comprehensive evaluation [J]. *Opt. Precision Eng.*, 2018, 26(4):962-970. (in Chinese)

[2] 刘智,黄江涛,冯欣. 构建多尺度深度卷积神经网络行为识别模型 [J]. 光学 精密工程, 2017, 25(3): 799-805.
LIU ZH, HUANG J T, FENG X. Action recognition model construction based on multi-scale deep convolution neural network [J]. *Opt. Precision Eng.*, 2017, 25(3): 799-805. (in Chinese)

[3] 李宇,刘雪莹,张洪群,等. 基于卷积神经网络的光学遥感图像检索 [J]. 光学 精密工程, 2018, 26(1): 200-207.
LI Y, LIU X Y, ZHANG H Q, *et al*.. Optical remote sensing image retrieval based on convolutional neural networks [J]. *Opt. Precision Eng.*, 2018, 26(1): 200-207. (in Chinese)

[4] WEI Y, XIA W, LIN M, *et al*.. HCP: A flexible CNN framework for multi-label image classification [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 38(9): 1901-1907.

[5] WANG Z, CHEN T, LI G, *et al*.. Multi-label image recognition by recurrently discovering attentional regions [C]. *Proceedings of the IEEE International Conference on Computer Vision (CVPR)*, 2017: 464-472.

[6] YU W J, CHEN ZH D, LUO X, *et al*.. DELTA: A deep dual-stream network for multi-label image classification [J]. *Pattern Recognition*, 2019, 91: 322-331.

[7] YAN Z, LIU W W, WEN SH P, *et al*.. Multi-label image classification by feature attention network [J]. *IEEE Access*, 2019, 7: 98005-98013.

[8] SONG P, JING L P, *et al*.. Exploiting label relationships in multi-label classification with neural networks [J]. *Journal of Computer Research and Development*. 2018, 55(8): 1751-1759.

[9] JIN J, NAKAYAMA H. Annotation order matters: Recurrent image annotator for arbitrary length image tagging [C]. 2016 *23rd International Conference on Pattern Recognition (ICPR)*, IEEE, 2016: 2452-2457.

[10] WANG J, YANG Y, MAO J, *et al*.. Cnn-rnn: A unified framework for multi-label image classification [C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016: 2285-2294.

[11] ZHAO B, LI X, LU X, *et al*.. A CNN-RNN architecture for multi-label weather recognition [J]. *Neurocomputing*, 2018, 322: 47-57.

[12] LYU F, HU F, SHENG V S, *et al*.. Coarse to fine: multi-label image classification with global/local attention [C]. 2018 *IEEE International Smart Cities Conference (ISC2)*, IEEE, 2018: 1-7.

[13] ZEILER M D, FERGUS R. Visualizing and understanding convolutional networks [C]. *European Conference on Computer Vision. Springer, Cham*, 2014: 818-833.

[14] SZEGEDY C, VANHOUCKE V, IOFFE S, *et al*.. Rethinking the inception architecture for computer vision [C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016: 2818-2826.

[15] ZHOU ZH H, ZHANG M L. Multi-instance multi-label learning with application to scene classification [C]. *Advances in Neural Information Processing Systems*, 2007: 1609-1616.

[16] DONG M, PANG K, WU Y, *et al*.. Transferring CNNS to multi-instance multi-label classification on small datasets [C]. 2017 *IEEE International Conference on Image Processing (ICIP)*, IEEE, 2017: 1332-1336.

[17] LI Y F, HU J H, JIANG Y, *et al*.. Towards discovering what patterns trigger what labels [C]. *Twenty-Sixth AAAI Conference on Artificial Intelligence (AAAI)*, 2012: 1012-1018.

[18] HUANG S J, GAN W, ZHOU ZH H. Fast multi-instance multi-label learning [C]. *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence (AAAI)*, 2014: 1868-1874.

[19] WANG T Z, HUANG S J, ZHOU ZH H. Towards identifying causal relation between instances and labels [C]. *Proceedings of the* 2019 *SIAM International Conference on Data Mining*. *Society for Industrial and Applied Mathematics*, 2019: 289-297.

[20] LIU K, WANG H, NIE F P, *et al*.. Learning multi-instance enriched image representations via non-greedy ratio maximization of the L1-norm distances [C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), 2018: 7727-7735.

作者简介：

赵海英(1972－)女，山东烟台人，博士，副教授，2012 年于北京科技大学获得博士学位，主要从事方向为多媒体数据挖掘、图像处理和文化计算理论体系等研究。E-mail: zhaohaiying @ bupt.edu.cn