

基于改进的有效区域基因选择与跨模态语义挖掘的图像属性标注

张红斌¹, 蒋子良¹, 熊其鹏¹, 武晋鹏¹, 邬任重¹, 袁 天², 姬东鸿³

(1. 华东交通大学软件学院, 江西南昌 330013; 2. 华东交通大学信息工程学院, 江西南昌 330013;
3. 武汉大学国家网络安全学院, 湖北武汉 430072)

摘 要: 图像属性标注是一种更细化的图像标注, 它能缩小认知与特征间“语义鸿沟”。现有研究多基于单特征且未挖掘属性蕴含的深层语义, 故无法准确刻画图像内容。改进有效区域基因选择算法融合图像特征, 并设计迁移学习策略, 实现材质属性标注; 基于判别相关分析挖掘特征间跨模态语义, 以改进相对属性模型, 标注材质属性蕴含的深层语义-实用属性。实验表明: 材质属性标注精准度达 63.11%, 较最强基线提升 1.97%; 实用属性标注精准度达 59.15%, 较最强基线提升 2.85%; 层次化的标注结果能全面刻画图像内容。

关键词: 图像标注; 有效区域基因选择; 相对属性; 迁移学习; 跨模态语义; 判别相关分析

中图分类号: TP 391 **文献标识码:** A **文章编号:** 0372-2112 (2020)04-0790-10

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2020.04.021

Image Attribute Annotation via a Modified Effective Range Based Gene Selection and Cross-Modal Semantics Mining

ZHANG Hong-bin¹, JIANG Zi-liang¹, XIONG Qi-peng¹, WU Jin-peng¹,
WU Ren-zhong¹, YUAN Tian², JI Dong-hong³

(1. School of Software, East China Jiaotong University, Nanchang, Jiangxi 330013, China;

2. School of Information Engineering, East China Jiaotong University, Nanchang, Jiangxi 330013, China;

3. School of Cyber Science and Engineering, Wuhan University, Wuhan, Hubei 430072, China)

Abstract: Image attribute annotation is a refined method of image annotation. It can narrow the “semantic gap” between cognition and features. However, a single feature is used to characterize images and the deep-level semantics are not fully explored. So annotations cannot depict images comprehensively. The traditional effective range based gene selection algorithm is modified to complete feature fusion. And transfer learning strategy is designed to complete material annotation. The cross-modal semantics among features are mined by the discriminant correlation analysis algorithm. So the relative attribute model is optimized to complete deep-level semantics (practical attributes) annotation. Experimental results demonstrate: Material attributes annotation accuracy reaches 63.11%, which is improved by 1.97% compared with baseline. Practical attributes annotation accuracy reaches 59.15%, which is improved by 2.85% compared with baseline. The proposed hierarchical annotation mechanism can more comprehensively depict images.

Key words: image annotation; effective range based gene selection; relative attribute; transfer learning; cross-modal semantics; discriminant correlation analysis

1 引言

图像属性标注是一种更细化的图像标注, 它更贴

近人类客观认知, 能细致、全面地描绘图像内容。

近年来, 图像属性标注已成为机器视觉领域的研究热点。如基于鼻、眼、口等属性的人脸识别^[1,2]、

收稿日期: 2018-07-02; 修回日期: 2019-12-20; 责任编辑: 孙瑶

基金项目: 国家自然科学基金(No. 61762038, No. 61861016); 教育部人文社会科学研究规划基金项目(No. 17YJAZH117); 江西省自然科学基金(No. 20171BAB202023); 江西省科技厅重点研发计划(No. 20171BBG70093, No. 20192BBE50071); 江西省教育厅科技项目(No. GJJ190323)

基于属性标注实现迁移学习^[3,4]、基于属性标注的零样本学习^[5,6]、面向用户需求的图像检索^[7]、基于属性标注的时尚趋势检索^[8]等. 上述工作^[1-8]仅提取单一特征,且多聚焦二元属性(Binary Attribute, BA)标注,即图像中是否包含相关属性(“1”表示包含,“0”表示不包含). 相对属性(Relative Attribute, RA)^[9-13]是对传统二元属性的改进,它着力度量属性程度. 因此,相对属性蕴含更丰富的语义信息,能改善人机交互体验.

不同于现有工作^[1-13],本文关注图像材质属性及其所蕴含的实用属性(防水性、透气性等. 它们来自生活,故称实用属性)标注:建立材质属性数据集 MattrSet 和相对属性数据集 Mattr_RA,构建材质属性与相对属性间的映射关系;提取局部二值模式(Local Binary Patterns, LBP)、尺度不变特征变换(Scale-Invariant Feature Transform, SIFT)、Gist、卷积神经网络(Convolutional Neural Networks)^[14]等特征,多角度刻画图像;设计新的特征融合算法^[15,16],完成图像材质属性标注;挖掘特征间跨模态语义,以改进相对属性模型,标注实用属性,向人们提供高质量的人机交互体验.

2 研究方法

2.1 图像材质属性标注方案

模型包括:构建新数据集、图像特征学习、特征中融合、实用属性标注及迁移学习等. 设计如图 1 所示图像材质属性标注方案. 从网站 <http://www.made-in-china.com/> 抓取图像,在材料学专家指导下对数据做去噪、筛选、标注,建立图像材质属性数据集. 从形状、纹理等角度提取传统图像特征:SIFT、LBP、Gist. 基于 VGG(Visual Germany Group)^[17]模型提取图像深度学习特征:VGG-16(维度 1000)、VGG-19(维度 1000). 改进有效区域基因选择(Effective Range based Gene Selection, ERGS)^[15,16]算法,完成特征中融合,生成材质属性标注模型. 基于判别相关分析(Discriminant Correlation Analysis, DCA)^[18]模型挖掘异构特征间的跨模态语义,综合属性映射表与跨模态语义改进 RA 模型^[10],完成实用属性标注. 构造两类迁移学习策略:迁移异类数据中的有效样本,扩展训练集;迁移训练好的中融合模型. 对比两类迁移学习策略的优劣.

2.2 改进 ERGS 模型实现特征中融合

ERGS 模型^[15,16]利用区间估计评估每个特征在每一类样本上的有效范围:根据特征在异类样本间的重叠区域,计算特征区分能力,生成特征权重. 原 ERGS

模型^[18]选取权重最大的特征而放弃其它特征,未充分利用特征间互补性,会影响标注性能. 本文对 ERGS 模型做改进:在特征评估基础上,设计一个决策函数,由所有特征共同决定最终标注. 数据集 $\text{Data} = \{(x_i, y_i), x_i \in \mathbb{R}^n, y_i \in Y, i = 1, 2, \dots, m\}$. n 为特征数, m 为样本数, $Y = \{c_1, c_2, \dots, c_l\}$ 为类标签集合,特征集 $F = \{f_1, f_2, \dots, f_n\}$. 图像特征 $f_i \in F$, 在 c_k 类样本上有效范围 ER_{ik} 如下:

$$ER_{ik} = [r_{ik}^-, r_{ik}^+] \\ = [\mu_{ik} - (1 - p_k)\gamma\sigma_{ik}, \mu_{ik} + (1 - p_k)\gamma\sigma_{ik}] \quad (1)$$

r_{ik}^+ 和 r_{ik}^- 分别表示特征 f_i 在 c_k 类样本上有效范围的上界和下界; μ_{ik} 和 σ_{ik} 分别表示特征 f_i 在 c_k 类样本上的平均值和标准偏差; p_k 为 c_k 类样本的先验概率 ($0 < p_k < 1$), $(1 - p_k)$ 因子可降低标准偏差 σ_{ik} 对特征有效范围上下边界的影响; γ 是通过切比雪夫不等式(式(2))推导的常量,即 1.732.

$$P(|F - \mu_{ik}| \geq \gamma\sigma_{ik}) \leq \frac{1}{\gamma^2} \quad (2)$$

基于改进的 ERGS 模型的特征中融合如下:

第 1 步 计算 f_i 在 c_k 类样本上有效范围 ER_{ik} .

第 2 步 计算 f_i 的重叠区域 OA_i :

$$OA_i = \sum_{j=1}^{l-1} \sum_{k=j+1}^l \varphi_i(j, k) \quad (3)$$

$$\varphi_i(j, k) = \begin{cases} r_{ik}^+ - r_{ik}^-, & r_{ik}^+ > r_{ik}^- \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

第 3 步 计算 f_i 的重叠区域系数 AC_i :

$$AC_i = \frac{OA_i}{\text{Max}_k(r_{ik}^+) - \text{Min}_k(r_{ik}^-)} \quad (5)$$

第 4 步 计算特征的 ERGS 权重 w_i :

$$w_i = \frac{AC_i}{\text{sum}(AC_j; 1 \leq j \leq n)} \quad (6)$$

第 5 步 设计决策函数. 基于 SVM 模型输出各类别预估概率,根据 ERGS 权重 w_i 对特征的预估概率加权,取最大值生成融合结果. 决策函数如下:

$$\max(p_i) = \max\{\text{sum}(p_{ij} * w_j; 1 \leq j \leq n)\} \quad (7)$$

2.3 基于跨模态语义的实用属性标注

基于 DCA 模型挖掘特征间蕴含的跨模态语义,改进传统相对属性模型^[10],完成实用属性标注. 相对属性是从材质属性中挖掘出的深层语义:防水性、透气性、柔软性、可水洗性及耐磨性. 它们贴近人类客观认知,实用价值更大. 首先,对实用属性程度进行划分,进而构建材质属性与实用属性间的映射关系,如表 1 所示.

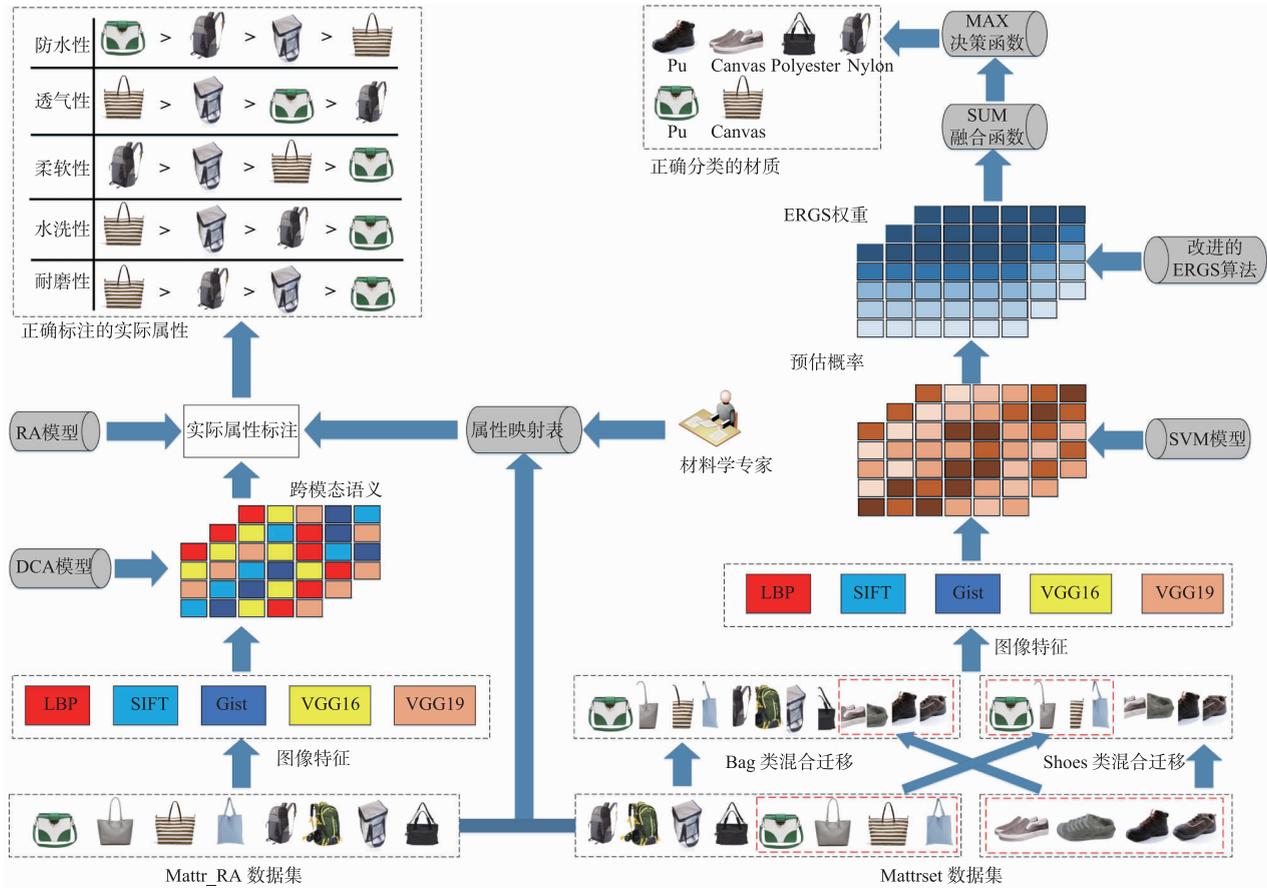


图1 本文所提的图像属性标注方案

表1 材质属性与实用属性之间的映射关系

材质属性	二元映射关系					相对映射关系				
	防水性	透气性	柔软性	水洗性	耐磨性	防水性	透气性	柔软性	水洗性	耐磨性
Pu(皮革)	1	0	0	0	0	4	2	1	1	1
Canvas(帆布)	0	1	0	1	1	1	4	2	4	4
Polyester(涤纶)	1	0	1	1	0	3	1	4	2	3
Nylon(尼龙)	0	0	1	1	0	2	3	3	3	2

如表1所示:在二元映射关系中,属性描述为0或1.例如,Pu防水性标记为1,表示Pu具有防水性.在相对映射关系中,针对某材质的实用属性,标记数值越大,则该材质对应的实用属性程度越高,最高程度值为4.如Canvas的透气性标记为4,即它的透气性最好.综上,无论是二元映射关系还是相对映射关系,它们刻画材质属性与实用属性间的语义关联,有助于人们更全面地认知材质属性.在进行相对属性排序时,先对每一个属性 $a_m, m \in (1, 2, \dots, M)$, M 即属性总数,构造一个相对属性集 O_m 和一个相似属性集 S_m .若 (i, j) 属于 O_m ,则 a_m 在 i 中比 j 有更强表现,即属性程度更高.若 (i, j) 属于 S_m ,则 a_m 在 i 和 j 中表现差不多,即属性程度相似.对于每个 a_m ,学习一个排序函数,如下所示:

$$rm(x_i) = \omega_m^T cm_i \quad (8)$$

$$\forall (i, j) \in O_m: \omega_m^T cm_i > \omega_m^T cm_j \quad (9)$$

$$\forall (i, j) \in S_m: \omega_m^T cm_i = \omega_m^T cm_j \quad (10)$$

ω_m 为学习到的权重, cm_i 是本文挖掘的特征间跨模态语义.得到排序函数后,对每一个新图像,基于权重 ω_m 计算图像关于这 m 个属性的程度表现,即相对属性.优化排序权重的目标函数如式(11)所示:

$$\min \left(\frac{1}{2} \|\omega_m^T\|_2^2 + C \left(\sum \varepsilon_{ij}^2 + \sum \gamma_{ij}^2 \right) \right) \quad (11)$$

$$\text{subject to} \begin{cases} \omega_m^T (cm_i - cm_j) \geq 1 - \varepsilon_{ij}, \forall (i, j) \in O_m \\ |\omega_m^T (cm_i - cm_j)| \leq \gamma_{ij}, \forall (i, j) \in S_m \\ \varepsilon_{ij} \geq 0; \gamma_{ij} \geq 0 \end{cases} \quad (12)$$

ε_{ij} 和 γ_{ij} 是松弛变量, C 是惩罚系数,该框架最大化有序对之间的边距.如式(12)所示.学习到每个属性的权重 ω_m 后,就可计算样本关于属性 a_m 的相对大小,基于该值完成属性程度排序.

3 实验结果

3.1 数据集

建立 MattrSet 数据集,它有 11021 张图像,包含:Pu、Canvas、Nylon、Polyester 四类材质属性. Pu 和 Canvas 属性包括 Bag、Shoes 两类商品. Bag 材质属性分布:Pu 1982 张、Canvas 1948 张、Nylon 1764 张、Polyester 1715 张. Shoes 材质属性分布:Pu 1757 张、Canvas 1855 张(Nylon、Polyester 非常稀缺). Mattr_RA 是 MattrSet 数据集的子集,以完成相对属性标注:选取 800 张 Bag 类图像, Pu、Canvas、Nylon、Polyester 四类材质属性各 200 张,为材质属性进行实

用属性程度排序,即根据防水性、透气性、耐磨性、柔软性及可水洗性的程度标注图像对.改进的 RA 模型基于映射关系(表 1)完成零样本学习,它将 Mattr_RA 中的数据作为测试集.二元属性模型采用 Mattr_RA 做测试集,剩余 10221 张图像作为训练集.本实验基于以下指标^[19]完成评估:Accuracy、Precision、AP(Average Precision).

3.2 实验结果

3.2.1 基于改进的 ERGS 模型的标注结果

表 2 为改进 ERGS 模型后的标注结果, VGG 表示 VGG-16、VGG-19 两种特征融合, S. G. L 表示 SIFT、Gist 及 LBP 三种特征融合.如 VGG&SIFT_Linear 表示 VGG-16、VGG-19、SIFT 三种特征融合,并基于线性核完成标注,其它模型命名同.

如表 2 所示.

表 2 ERGS 融合后的标注结果(每列最优值如 68.07 等所示,单位:%,“↑”表示 ERGS 中融合后性能提升)

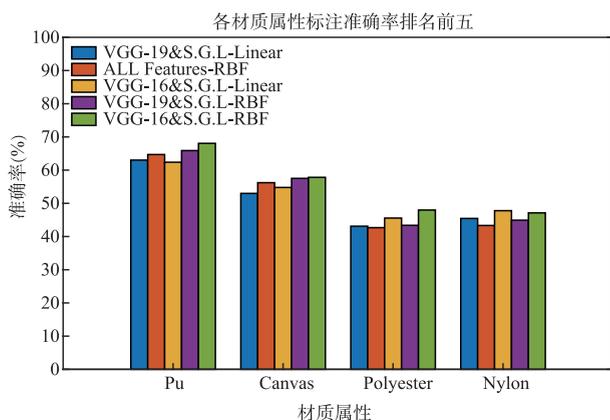
中融合特征	准确率				AP _{material}	Accuracy
	Pu	Canvas	Polyester	Nylon		
VGG&SIFT_Linear	60.40	50.04 ↑	42.91 ↑	44.89	49.56	51.43
VGG&Gist_Linear	55.31 ↑	48.05 ↑	41.87 ↑	42.16 ↑	46.85	48.57
VGG&LBP_Linear	56.44	48.55	42.99 ↑	43.70 ↑	47.92	49.49
VGG-16&S. G. L_Linear	62.38	54.78 ↑	45.56 ↑	47.79 ↑	52.63	54.62
VGG-19&S. G. L_Linear	63.02	52.99 ↑	43.11 ↑	45.45	51.14	53.46
All Features_Linear	61.72	52.41 ↑	43.81 ↑	45.11	50.76	53.00
VGG&SIFT_RBF	61.14	51.84 ↑	39.93 ↑	41.49	48.60	50.65
VGG&Gist_RBF	58.92	50.22	40.50 ↑	40.49	47.53	49.19
VGG&LBP_RBF	58.08	50.91	40.14	40.57	47.42	49.27
VGG-16&S. G. L_RBF	68.07	57.83 ↑	47.98 ↑	47.12 ↑	55.25	57.29
VGG-19&S. G. L_RBF	65.89	57.52 ↑	43.38 ↑	44.92	52.93	55.32
All Features_RBF	64.70	56.22	42.68	43.33	51.73	54.02
VGG&SIFT_Sigmoid	60.73	38.01	32.96	48.80 ↑	45.13	46.25
VGG&Gist_Sigmoid	44.56	32.75	29.44	50.34 ↑	39.27	37.23
VGG&LBP_Sigmoid	36.97	27.87 ↑	0.00	0.00	16.21	31.32
VGG-16&S. G. L_Sigmoid	66.36	44.33	37.42 ↑	47.28 ↑	48.85	50.86
VGG-19&S. G. L_Sigmoid	66.33	44.22	36.94 ↑	46.90 ↑	48.60	50.76
All Features_Sigmoid	63.27	39.69	36.92 ↑	51.19 ↑	47.77	48.25
VGG&SIFT_Polynomial	42.99 ↑	34.98 ↑	22.11 ↑	33.88	33.49	38.42
VGG&Gist_Polynomial	48.34	35.45 ↑	25.93 ↑	34.51 ↑	36.06	39.98
VGG&LBP_Polynomial	45.08 ↑	37.44	27.28 ↑	37.14	36.73	40.60
VGG-16&S. G. L_Polynomial	50.46	37.00	16.67	35.83	34.99	40.77
VGG-19&S. G. L_Polynomial	53.58	35.94	28.44 ↑	37.57	38.88	41.41
All Features_Polynomial	50.43	36.75	22.73	39.00	37.23	41.82

(1) 共计 42.71% 的 ERGS 指标标注性能提升, 传统特征与深度学习特征间有较强互补性.

(2) 标注性能最佳的核函数是 RBF, 最差的是 Polynomial, 这是由于材质属性在低维空间中存在大量线性不可分情况, RBF 相对 Polynomial 能将特征由低维空间映射到高维空间, 此时样本线性可分, 对应标注性能更优. 4 种核函数 MAP 值降序排列: RBF > Linear > Sigmoid > Polynomial.

(3) VGG-16&S. G. L_RBF 表现最佳, 其标注精准度达 57.29%, 较次优指标 VGG-19&S. G. L_RBF 提升 1.97%. 相反, 全部特征融合性能并非最优, 即执行 ERGS 模型时应合理组合图像特征, 以获取最佳性能.

(4) 改进的 ERGS 模型最易于改善 Polyester 材质



(a) 材质属性与特征之间的对应关系

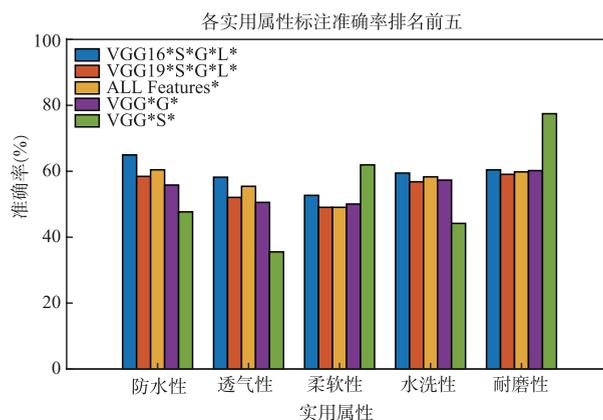
标注性能, 有 $17/24 = 70.83\%$ 的 ERGS 指标性能有提升. 相反, 该算法最难改善 Pu 材质标注性能, 仅有 $3/24 = 12.5\%$ 的指标性能提升.

(5) VGG-16 较 VGG-19 能获取更优标注性能, 其所参与的特征融合结果普遍更好.

为更好地展示材质属性与图像特征间的关系, 根据表 2 绘制图 2(a) 所示对应关系图.

(1) VGG16&S. G. L-RBF 在四种材质属性上均表现良好. 除 Nylon 材质, Pu、Canvas、Polyester 三属性达最佳.

(2) RBF 核整体标注性能最佳, Linear 核次之. RBF 核能更准确地标注四种材质属性, 且标注性能由大到小排序: Pu > Canvas > Polyester > Nylon.



(b) 实用属性与特征之间的对应关系

图2 图像属性与特征之间的对应关系

3.2.2 迁移学习实验结果

设计了两类迁移策略: 样本迁移和模型迁移. 每类策略又包含两种迁移方向: ①基于 Bag 类的迁移, 将 Shoes 类中图像迁移到 Bag 类中 (“Shoes → Bag”); ②基于 Shoes 类的迁移, 将 Bag 类中图像迁移到 Shoes 中 (“Bag → Shoes”). 考虑到 Bag 类中有四种材质属性, 而 Shoes 类中只有两种, 故选择共有属性完成迁移. 图 3 为选择 Linear、Polynomial、RBF 数及 Sigmoid 等核函数后, 实现第一类策略、第一种迁移方向的实验结果. 蓝色柱体和红色柱体分别表示迁移学习前、后的标注性能. 图 4 为选择 Linear、Polynomial、RBF 及 Sigmoid 等核函数后, 实现第一类策略、第二种迁移方向的实验结果. 蓝色柱体、红色主体分别表示表示迁移学习前、后的标注性能. 图 5 是执行模型迁移后的实验结果, 我们选取在单 Bag 类、单 Shoes 类上标注性能最好的 RBF 核完成模型迁移. “Bag-Shoes” 表示迁移基于 Bag 数据集训练好的 ERGS 模型, “Shoes-Bag” 表示迁移基于 Shoes 数据集训练好的 ERGS 模型.

如图 3(a) 所示: 迁移学习后, 8 组特征性能有提升, 平均提升幅度 3.73%, 迁移学习效果显著. 其中,

VGG-19 提升最显著, 这表明采用 Linear 核时, VGG-19 相比 VGG-16 在 Bag、Shoes 两类数据间拥有更好“鲁棒性”. 迁移后, VGG-16&S. G. L 性能最佳. 如图 3(b) 所示: 迁移学习后, 8 组特征性能有提升, 平均提升幅度 10.99%, 迁移学习效果更显著. VGG-16 提升最显著, 这表明采用 Polynomial 核时, VGG-16 相比 VGG-19 在 Bag、Shoes 两类数据间拥有更好“鲁棒性”. 迁移后, VGG&LBP 性能最佳. 如图 3(c) 所示: 迁移学习后, 8 组特征性能均有提升, 平均提升幅度 6.91%, 迁移学习效果较显著. VGG-19 提升最显著. 迁移后, VGG-16&S. G. L 性能最佳, 且优于图 3 其它最优结果. 如图 3(d) 所示: 迁移学习后, 8 组特征中仅 5 组性能有提升, 迁移学习效果不显著. 显然 Sigmoid 核不利于材质属性标注的迁移学习. 综上, 迁移学习后精准度平均提升幅度降序排列: Polynomial > RBF > Linear > Sigmoid, Polynomial 核对数据有更好“包容性”; 迁移学习后, 采用 RBF 核的 VGG-16&S. G. L 特征标注性能最佳, 其精准度达 63.11%. 加入深度学习特征后, 共计 $29/32 \approx 90.63\%$ 指标的性能得到提升, 这进一步说明: 深度学习特征能

较好地适应迁移学习,并体现较强鲁棒性。

如图 4 所示:迁移学习后,所有特征性能均有下降,其中 Polynomial 核的下降幅度最小,而 Sigmoid 核迁移学习效果最差.这表明:在实现第一类策略、第二种迁移方向(“Bag \rightarrow Shoes”)后,模型未有效学到关键的

判别信息.相反,Bag 类中的形态、光照等噪声在材质识别中起到负面作用.此外,执行 ERGS 特征中融合后,无论是迁移前还是迁移后,标注精准度都有不同程度提升.这表明:改进后的 ERGS 模型确实能利用特征间的互补信息提升标注性能。

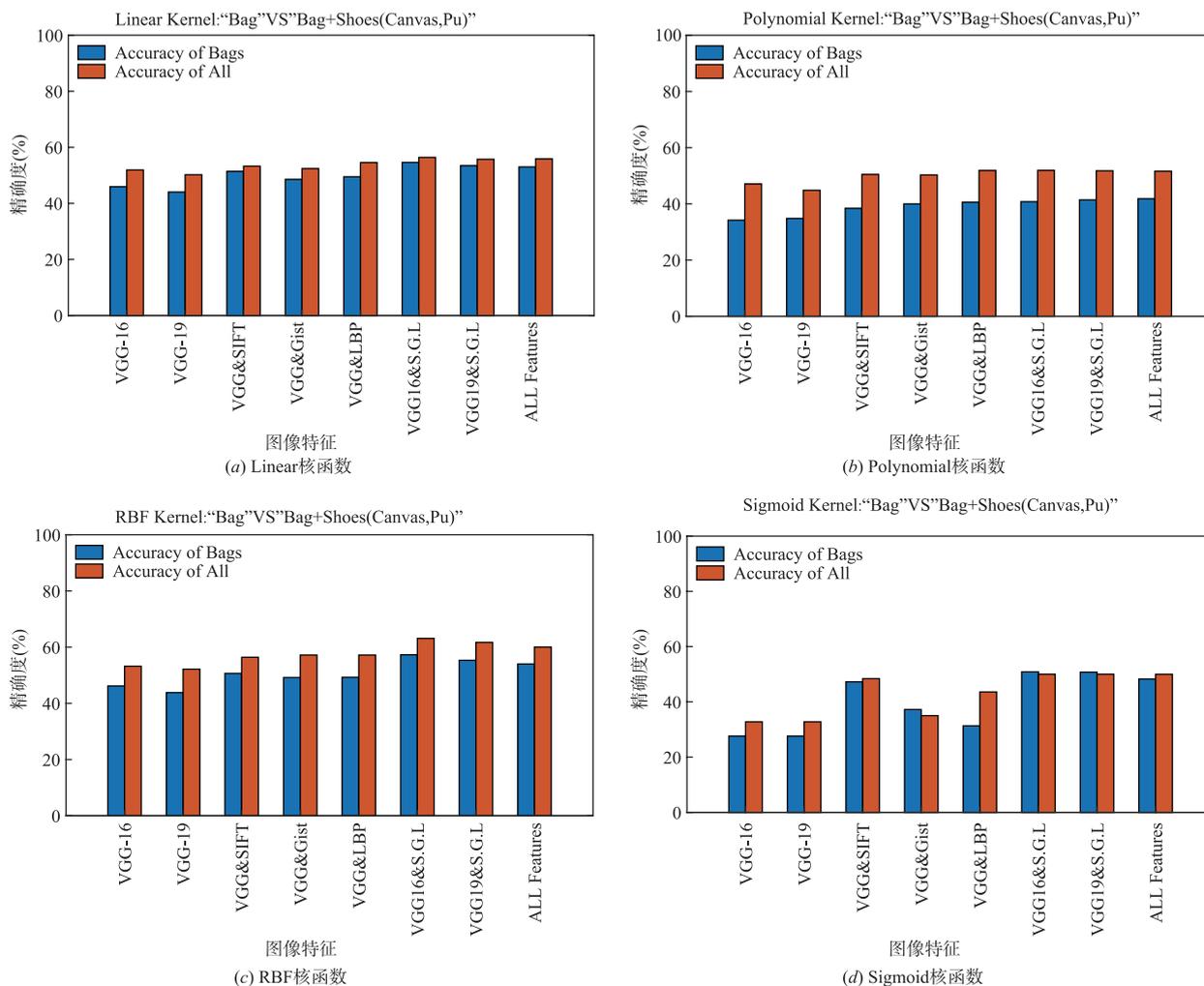


图3 实现第一类策略、第一种迁移方向后的标注精准度对比

如图 5 所示:在模型迁移过程中,“Bag \rightarrow Shoes”的材质分类性能低于“Shoes \rightarrow Bag”,尤其是特征中融合后,性能衰减严重.这说明:Shoes 类中的两种材质数据能较好地刻画 Bag 类中对应材质.相反,Bag 数据的噪声很大,模型迁移后,识别效果不显著。

综上,对比图 3、图 4、图 5,可以发现:样本迁移优于模型迁移,模型迁移的难度较大。

3.2.3 材质属性标注综合比较

实现第一类策略、第二种迁移方向以及第二类迁移策略均未较好地涵盖数据集中四种材质属性,故本文仅挑选第一类策略、第一种迁移方向完成学习,并与主流基线进行比较.基线包括:①单模型,决策树(Deci-

sion Tree, DT)、K 近邻(K-Nearest Neighbor, KNN)、逻辑回归(Logistic Regression, LR)、朴素贝叶斯(Naive Bayesian, NB)、随机森林(Random Forest, RF);②Boosting 类模型,梯度提升决策树(Gradient Boosting Decision Tree, GBDT)^[20]、自适应提升(Adaptive Boosting, AdaBoost)^[20]、极端梯度提升(eXtreme Gradient Boosting, XGBoost)^[21];③ERGS 类模型^[19],GS-DT、GS-KNN、GS-LR、GS-NB、GS-RF 等,即基于改进的 ERGS 模型执行特征中融合;④现有工作, Farhadi^[22]的 SVM 模型.对于单模型、Boosting 类模型、Farhadi 的工作,分别结合四种特征进行材质属性标注,选取迁移学习性能最优的结果.图 6 是本文最佳结果(图 3(c)中 Our Model: VGG-

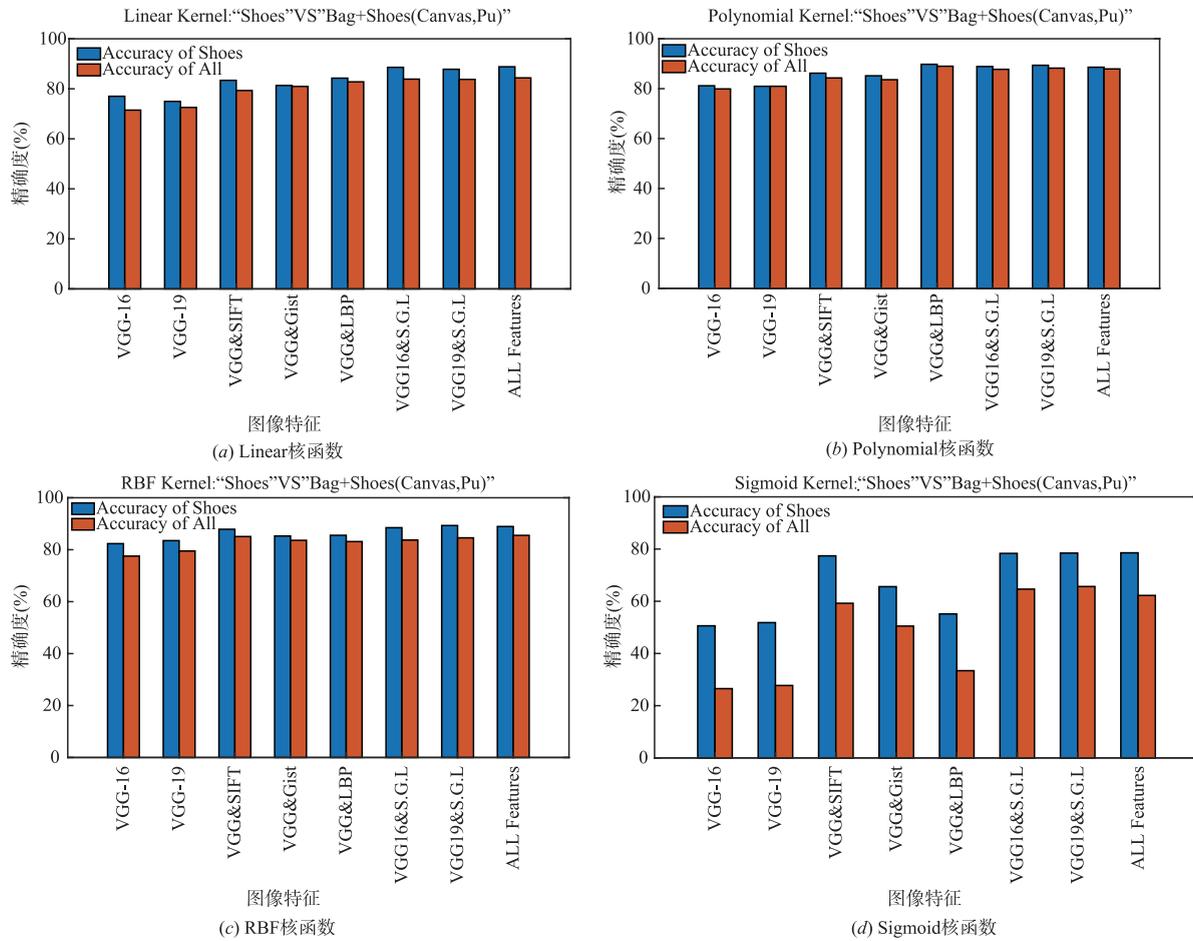


图4 实现第一类策略、第(2)种迁移方向后的标注精准度对比

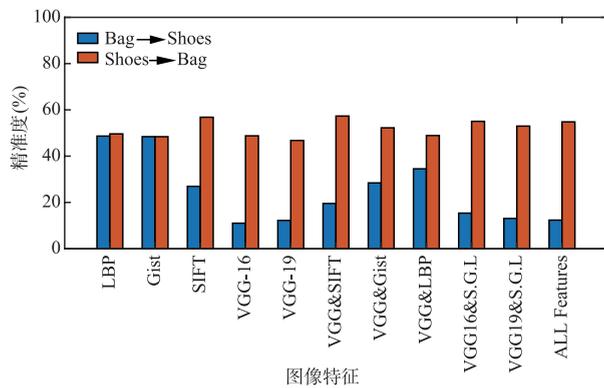


图5 基于RBF核实现第二类迁移学习后的标注精准度对比

16&S. G. L的标注性能:63.11%)与主要基线的比较。

图7中,“DT-G”表示DT模型选取Gist特征时性能最优.对于ERGS类模型,执行表2中全部特征融合,并选取迁移学习性能最优结果.如“GS-DT-S. G. L”表示ERGS类新模型“GS-DT”,它选取三种传统特征S、G和L做中融合时性能最优.由图7可见,本文模型优于基线模型,其标注精准度较最强基线提升1.97%.

3.2.4 相对属性标注结果

基于DCA模型挖掘特征间的跨模态语义,以改进RA模型,完成实用属性标注.计算特征平均精度^[19]、属性平均精度^[19],得到如图6所示实验结果.VGG*表示VGG-16、VGG-19两种特征进行跨模态语义挖掘,S*L*表示对SIFT、Gist及LBP三种特征进行跨模态语义挖掘.其它特征组合命名同理.

图6(a)(b)展示不同特征在二元属性标注和相对属性标注中性能差异.图6(c)(d)展示跨模态语义挖掘前后二元属性标注和相对属性标注的结果.本文对比Zhang^[19]的工作,在图6(c)(d)中Zhang的结果与本文结果分别用蓝、红色表示.

如图6(a)所示,SIFT在二元属性标注中最优,DCA融合后的特征普遍不佳,尤其是“VGG*L*”特征,即它们之间的互补性较低.如图6(b)所示,在挖掘跨模态语义并执行特征融合后,相关特征表现优异,尤其是“VGG16*S*G*L*”特征,其相对属性标注精度达59.15%,较最强基线提升2.85%.如图6(c)所示,二元属性标注精度普遍不高,其中“透气性”最易识别,

“水洗性”最难识别. 特征融合后,除“水洗性”之外,其它实用属性的性能急速衰减,该结果与图 6(a)吻合. 显然,二元属性不能处理低维度 DCA 特征. 如图 6(d)所示,在挖掘跨模态语义并执行特征融合后,除“柔软性”之外,其它实用属性的性能均有提升,且大幅优于二元属性标注,其中“透气性”提升幅度最大,该结果与图 6

(b)吻合. 主要原因:二元属性模型对属性的判断主要基于材质属性,它会忽视实用属性之间的差异,且材质属性间的混淆性也会干扰其完成实用属性标注. 相反,相对属性模型定量评估实用属性的程度并做标注,它基于零样本学习,故整体优势更为突出. 而经过跨模态语义挖掘之后,实用属性标注性能进一步提升.

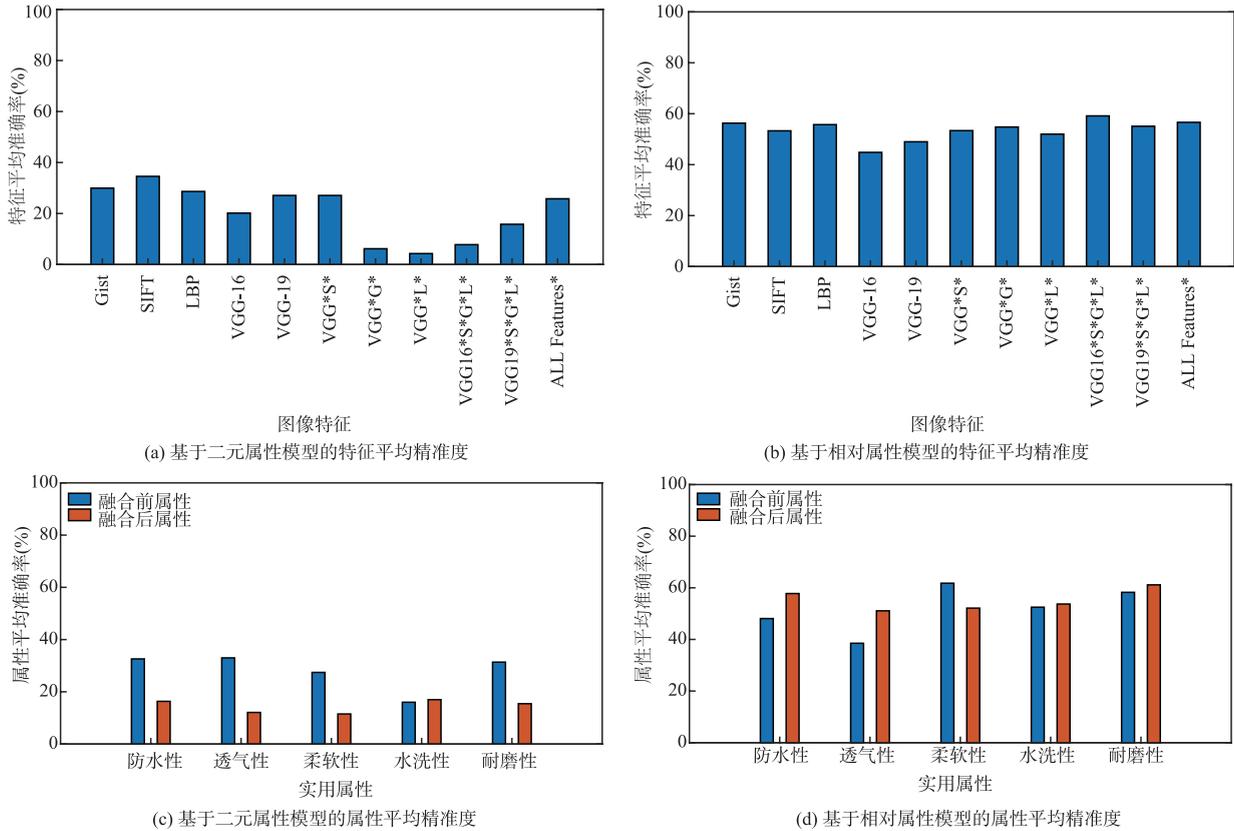


图6 相对属性标注结果对比

图 6(b)展示了实用属性与图像特征间的关系. 如图 6(b)所示,除 VGG * S * 外,其它融合后特征的标注性能较均衡;这说明:经过 DCA 融合之后,不同特征间的语义信息得以充分利用,融合后的特征更具鲁棒性. 此外,VGG16 * S * G * L * 整体标注性能最优(3 个最

优,2 个次优). 这表明:在跨模态语义挖掘中,VGG-16 特征扮演重要角色. 相反,VGG * S * 特征表现不稳定(2 个最优,3 个最差).

综上,实用属性标注有助于更准确、全面地刻画材质属性,并建立友好的人机交互(详见图 1,针对每个实用属性进行其程度值降序排列).

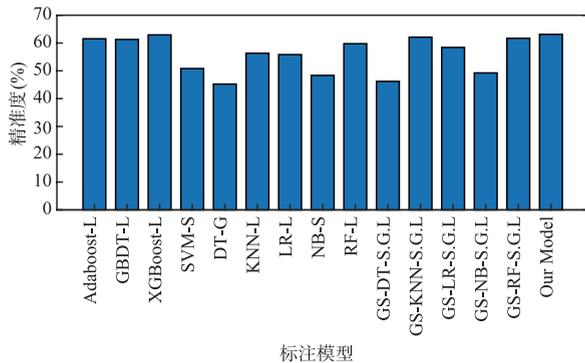


图7 本文结果与主要基线的标注精度对比

4 结论及展望

图像属性标注从中间语义层刻画视觉内容,其语义信息准确、丰富,故它在机器视觉领域中具有广阔的应用前景. 本文聚焦材质属性及其所蕴含的实用属性标注:在传统图像特征、深度学习特征抽取基础上,改进 ERGS 模型以融合不同特征,完成材质属性标注. 其中,引入迁移学习策略,有效扩展训练数据;挖掘不同特征间的跨模态语义,以改进相对属性模型,准确度量并标注材质属性蕴含的深层语义-实用属性. 结果表明:

改进的 ERGS 模型能充分利用特征间互补性,以改善材质属性标注性能;迁移学习策略扩充出了高质量样本,为训练鲁棒的材质识别模型奠定了重要基础;特征间的跨模态语义有助于提升实用属性标注性能. 综上,本文工作为人们提供了一个层次化的材质属性知识体系,以辅助其对材质的“去伪存真”,并更全面地认知材质. 此外,材质属性知识体系在商品检索、机器人操控、原材料自动分拣等领域中将发挥作用.

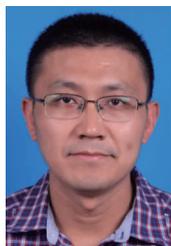
未来拟基于 Attention 机制^[23,24] 聚焦图像中的局部视觉区域,降低来自图像背景的噪声干扰;采用 PG-GAN^[25]、BigGAN^[26] 等模型完成更高质量的数据增强;引入 IRGAN^[27] 模型,继续改进相对属性模型的 Rank 函数,以提升实用属性标注性能.

参考文献

- [1] 李倩玉,蒋建国,齐美彬,等. 基于改进深层网络的人脸识别算法[J]. 电子学报,2017,45(3):619-625.
LI Qian-yu,JIANG Jian-guo,QI Mei-bin et al. Face recognition algorithm based on improved deep networks[J]. Acta Electronica Sinica, 2017, 45 (3) : 619 - 625. (in Chinese)
- [2] ZHU X Y, LIU H, LEI Z, et al. Large-scale bisample learning on id versus spot face recognition [J]. International Journal of Computer Vision, 2019, 127(6-7):684-700.
- [3] 周家琪,刘丽,崔晓萍,等. 基于相似性迁移学习的图像标注[J]. 山东师范大学学报(自然科学版),2016,31(2):22-26.
ZHOU Jia-qi,LIU Li,CUI Xiao-ping, et al. An image annotation method based on similarity transfer learning[J]. Journal of ShanDong Normal University (Natural Science), 2016, 31(2) : 22 - 26. (in Chinese)
- [4] Chen C Y, GRAUMAN K. Inferring analogous attributes [A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) [C]. Piscataway, NJ:IEEE,2014. 200-207.
- [5] JAYARAMAN D, GRAUMAN K. Zero-shot recognition with unreliable attributes [A]. Proceedings of the Advances in Neural Information Processing Systems (NIPS) [C]. New York;Curran Associates,2014. 3464-3472.
- [6] 乔雪,彭晨,段贺,张钰尧. 基于共享特征相对属性的零样本图像标注[J]. 电子与信息学报,2017,39(7):1563-1570.
QIAO Xue, PENG Chen, DUAN He, ZHANG Yu-yao. Shared features based relative attributes for zero-shot image classification [J]. Journal of Electronics & Information Technology, 2017, 39(7) : 1563 - 1570. (in Chinese)
- [7] 张峰,钟宝江. 基于兴趣目标的图像检索[J]. 电子学报, 2018, 44(8) : 1915 - 1923.
ZHANG Feng, ZHONG Bao-jiang. Image retrieval based on interested objects [J]. Acta Electronica Sinica, 2018, 44 (8) : 1915 - 1923. (in Chinese)
- [8] ZHAO B, FENG J, WU X, et al. Memory-augmented attribute manipulation networks for interactive fashion search [A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) [C]. Piscataway, NJ:IEEE,2017. 1520-1528.
- [9] YU A, GRAUMAN K. Thinking outside the pool: active training image creation for relative attributes [A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) [C]. Piscataway, NJ:IEEE, 2019. 708-718.
- [10] KOVASHKA A, PARIKH D, GRAUMAN K. WhittleSearch; interactive image search with relative attribute feedback [J]. International Journal on Computer Vision, 2015, 115(2) : 185 - 210.
- [11] YU A, GRAUMAN K. Just noticeable differences in visual attributes [A]. Proceedings of the IEEE International Conference on Computer Vision (ICCV) [C]. Piscataway, NJ:IEEE,2015. 2416-2424.
- [12] SINGH K K, LEE Y J. End-to-end localization and ranking for relative attributes [A]. Proceedings of the European Conference on Computer Vision (ECCV) [C]. Charm, Switzerland: Springer, 2016. 753-769.
- [13] DUBEY A, AGARWAL S. Modeling image virality with pairwise spatial transformer networks [A]. Proceedings of the 25th ACM International Conference on Multimedia (MM) [C]. New York;ACM,2017. 663-671.
- [14] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [A]. Proceedings of the Advances in Neural Information Processing Systems (NIPS) [C]. New York:Curran Associates,2012. 1097-1105.
- [15] CHANDRA B, GUPTA M. An efficient statistical feature selection approach for classification of gene expression data [J]. Journal of Biomedical Informatics, 2011, 44(4) : 529-535.
- [16] 王晓梅,林晓惠,黄鑫. 基于特征有效范围的前向特征选择及融合分类算法[J]. 小型微型计算机系统, 2016, 37(6) : 1159-1163.
WANG Xiao-mei, LIN Xiao-hui, HUANG Xin. Algorithm of forward feature selection and aggregation of classifiers based on feature effective range [J]. Journal of Chinese Computer Systems, 2016, 37(6) : 1159 - 1163. (in Chinese)
- [17] SIMONYAN K, ZISSERMAN A. Very Deep Convolutional Networks for Large-Scale Image Recognition [OL]. <https://arxiv.org/abs/1409.1556>, 2014/2019-12-18.

- [18] HAGHIGHAT M, ABDEL-MOTTALEB M, ALHALABI W. Discriminant correlation analysis; real-time feature level fusion for multimodal biometric recognition [J]. IEEE Transactions on Information Forensics and Security, 2016, 11(9):1984 – 1996.
- [19] ZHANG H B, QIU D D, WU R Z, et al. Novel framework for image attribute annotation with gene selection XGBoost algorithm and relative attribute model [J]. Applied Soft Computing, 2019, 80:57 – 79.
- [20] 李航. 统计学习方法 [M]. 北京: 清华大学出版社, 2012. 151 – 152.
- [21] Chen T Q, Guestrin C. XGBoost: a scalable tree boosting system [A]. Proceedings of ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD) [C]. New York: ACM, 2016. 785 – 794.
- [22] FARHADI A, ENDRES I, HOIEM D, et al. Describing objects by their attributes [A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) [C]. Piscataway, NJ: IEEE, 2009. 1778 – 1785.
- [23] HAMKER F H. Predictions of a model of spatial attention using sum-and max-pooling functions [J]. Neurocomputing, 2018, 56(1):329 – 343.
- [24] WU L, WANG Y, LI X, et al. Deep attention-based spatially recursive networks for fine-grained visual recognition [J]. IEEE Transactions on Cybernetics, 2018, 49(5):1791 – 1802.
- [25] KARRAS T, AILA T, LAINE S, et al. Progressive Growing of Gans for Improved Quality, Stability, and Variation [OL]. <https://arxiv.org/abs/1710.10196>, 2017/2019-12-18.
- [26] BROCK A, DONAHUE J, SIMONYAN K. Large Scale Gan Training for High Fidelity Natural Image Synthesis [OL]. <https://arxiv.org/abs/1809.11096>, 2018/2019-12-18.
- [27] WANG J, YU L, ZHANG W, et al. IRGAN: a minimax game for unifying generative and discriminative information retrieval models [A]. In Proceedings of ACM SIGIR [C]. New York: ACM, 2017. 515 – 524.

作者简介



张红斌 男, 1979 年 10 月生, 江苏如皋人. 华东交通大学软件学院副教授、硕士生导师. 主要研究方向: 计算机视觉、自然语言处理、推荐系统.
E-mail: zhanghongbin@whu.edu.cn



蒋子良 男, 1997 年 4 月生, 江西婺源人. 华东交通大学软件学院硕士研究生. 主要研究方向: 机器学习、图像材质识别.
E-mail: ziliangjiang@163.com