

# 控制与决策

Control and Decision

联合SENet异构层特征融合与集成学习的材质图像识别

张红斌, 熊其鹏, 蒋子良, 石焱炜, 周娟, 武晋鹏

引用本文:

张红斌, 熊其鹏, 蒋子良, 石焱炜, 周娟, 武晋鹏. 联合SENet异构层特征融合与集成学习的材质图像识别[J]. 控制与决策, 2022, 37(6): 1632–1642.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2020.1559>

---

## 您可能感兴趣的其他文章

Articles you may be interested in

### 一种基于多层语义特征的图像理解方法

An image understanding method based on multi-level semantic features

控制与决策. 2021, 36(12): 2881–2890 <https://doi.org/10.13195/j.kzyjc.2020.0927>

### 基于卷积神经网络的云雾遮挡舰船目标识别

Obscured ship target recognition based on convolutional neural network

控制与决策. 2021, 36(3): 661–668 <https://doi.org/10.13195/j.kzyjc.2019.0781>

### 基于FRC框架的红外与可见光图像融合方法

Infrared and visible image fusion based on FRC algorithm

控制与决策. 2021, 36(11): 2690–2698 <https://doi.org/10.13195/j.kzyjc.2020.0669>

### 基于联合知识表示学习的多模态实体对齐

Multi-modal entity alignment based on joint knowledge representation learning

控制与决策. 2020, 35(12): 2855–2864 <https://doi.org/10.13195/j.kzyjc.2019.0331>

### 融合稀疏编码与深度学习的草图特征表示

A feature representation of sketch based on fusion of sparse coding and deep learning

控制与决策. 2021, 36(3): 699–704 <https://doi.org/10.13195/j.kzyjc.2019.0941>

# 联合 SENet 异构层特征融合与集成学习的材质图像识别

张红斌<sup>†</sup>, 熊其鹏, 蒋子良, 石皞炜, 周娟, 武晋鹏

(华东交通大学 软件学院, 南昌 330013)

**摘要:** 材质图像识别具备广阔的应用前景, 如衣物识别、机器人拾取、工业检测等. 受光照强度和拍摄角度等影响, 材质图像易发生变化, 而挖掘鲁棒、高效的图像特征是应对该变化的关键. 对此, 提出 SECF<sup>2</sup> 模型: 抽取 SENet 中具有良好互补性的异构层特征; 改进聚类典型相关性分析模型, 实现异构层特征融合, 生成刻画材质图像的深层视觉语义, 它是一种判别性更强且鲁棒的新特征; 采用深层视觉语义训练分类模型并执行集成学习, 完成材质图像识别. 实验表明: SECF<sup>2</sup> 模型在两个材质图像数据集上都有效, 其中 Fabric 上的识别精准度较最强基线提升 8.85%; SECF<sup>2</sup> 模型还具备较强通用性, 在图像情感分析基准数据集上取得优异的表现. 此外, SECF<sup>2</sup> 仅需两个特征和一次融合, 模型复杂度降低且实时效率优良.

**关键词:** 材质图像识别; 深层视觉语义; 特征融合; SENet; cluster-CCA; 集成学习

中图分类号: TP391

文献标志码: A

DOI: 10.13195/j.kzyjc.2020.1559

开放科学(资源服务)标识码(OSID):



**引用格式:** 张红斌, 熊其鹏, 蒋子良, 等. 联合 SENet 异构层特征融合与集成学习的材质图像识别 [J]. 控制与决策, 2022, 37(6): 1632-1642.

## Material image recognition combining heterogeneous-layer feature fusion of SENet and ensemble learning

ZHANG Hong-bin<sup>†</sup>, XIONG Qi-peng, JIANG Zi-liang, SHI Hao-wei, ZHOU Juan, WU Jin-peng

(School of Software, East China Jiaotong University, Nanchang 330013, China)

**Abstract:** Material image recognition has broad application prospects, such as clothing recognition, automatic picking by a robot, industrial detection, etc. Owing to the influence of light intensity and camera angle, material image is easy to change. Mining robust and effective image features is one of the most important factors to handle this change. To address this problem, an improved model called SECF<sup>2</sup> is proposed. Heterogeneous-layer features in the SENet model are extracted respectively, which can complement each other well. Then the traditional cluster canonical correlation analysis (cluster-CCA) model is modified to complete the feature fusion of these heterogeneous-layer features. As a significant result, more discriminant and robust deep-level visual semantics is obtained after feature fusion to better characterize the material image. Finally, the deep-level visual semantics is employed to train classification models, and a voting-based ensemble learning strategy is designed to further boost the final classification performance. Experimental results demonstrate that the SECF<sup>2</sup> model is effective on two material image datasets. Especially, compared with the best baseline, it obtains 8.85 % performance improvement in the Fabric dataset. In addition, the SECF<sup>2</sup> model has strong versatility and it has achieved excellent performance in the image sentiment analysis task. Moreover, the proposed ECF<sup>2</sup> model only needs two features and one time feature fusion, hence the model obtains better real-time efficiency.

**Keywords:** material image recognition; deep-level visual semantics; feature fusion; SENet; cluster-CCA; ensemble learning

## 0 引言

材质图像识别具备广阔的应用前景, 如场景识别<sup>[1]</sup>、工业检测<sup>[2]</sup>、医学图像识别<sup>[3]</sup>、机器人视觉<sup>[4]</sup>等. 材质图像识别<sup>[5-6]</sup>已成为机器视觉领域的热点研

究问题. 材质是一种非常重要的视觉线索, 它广泛存在于自然界物体表面, 在视觉上具备丰富的多样性, 而人眼难以区分它们的细微差别. 材质图像识别受诸多外部因素, 如光线、角度和拍摄距离等影响, 而

收稿日期: 2020-11-11; 录用日期: 2021-04-07.

基金项目: 国家自然科学基金项目(61762038, 61861016); 教育部人文社会科学研究规划基金项目(20YJAZH142).

责任编辑: 邹长亮.

<sup>†</sup>通讯作者. E-mail: zhanghongbin@whu.edu.cn.

且材质图像类内差异大而类间差异小. 因此, 材质图像容易发生变化, 而挖掘鲁棒、高效的图像特征是应对这一变化的关键环节.

近年来, GS-XGBoost 模型<sup>[7]</sup>, 根据特征在不同类别材质图像上的有效区域, 计算其 ERGS (effective range based gene selection) 权重, 该方法需设计特征组合并优选分类模型, 以获取更好的识别效果. HMF<sup>2</sup> 模型<sup>[8]</sup> 改进了 MDCA (multiset discriminant correlation analysis) 算法<sup>[9]</sup> 以完成多层次特征融合. 该方法需挑选一组特征并执行多次融合. 然而, 上述工作<sup>[7-8]</sup> 加入了人工干预, 需提取大量特征且特征融合策略较为复杂, 模型实用性受到影响.

本文结合主流深度学习框架 SENet (squeeze-and-excitation network)<sup>[10]</sup> 并改进聚类典型相关分析 (cluster canonical correlation analysis, cluster-CCA) 模型<sup>[11]</sup>, 提出 SECF<sup>2</sup> (SE 来自 SENet, C 来自 cluster, F<sup>2</sup> 即 feature fusion) 模型, 旨在充分挖掘 SENet 异构层特征间的深层视觉语义, 仅用少量特征准确刻画材质图像, 最终提升材质图像识别精度. SENet<sup>[10]</sup> 的 SE (squeeze-and-excitation) 模块学习不同通道中特征的重要性, 进而提升特征表现力; 其次, SENet 异构层特征蕴涵不同的视觉语义, 它们是材质图像识别的重要前提, 同时, 异构层特征指向相同或相似的材质语义, 故特征间存在互补性<sup>[8]</sup>, 充分利用该互补性可生成更加高效、鲁棒的新特征, 即深层视觉语义, 最终改善材质图像识别性能. 此外, 通过设计集成学习策略组合多个分类模型的预测结果, 进一步提升识别精度. 本文主要贡献如下:

1) 提出 SECF<sup>2</sup> 模型, 聚焦 SENet 异构层特征的深度融合, 生成判别性更强且鲁棒的深层视觉语义, 以准确刻画材质图像, 继而引入集成学习策略继续改善识别精度.

2) 提出基于 SECF<sup>2</sup> 模型的实时材质识别系统并部署至服务器, 自动完成材质图像识别, 这有助于缩小理论研究与实际应用间的“鸿沟”.

SECF<sup>2</sup> 模型的主要特点如下:

1) 它仅需 SE-ResNet50、SE-ResNet101 等深度学习特征, 且只执行一次特征融合. 相比于 HMF<sup>2</sup><sup>[8]</sup>、GS-XGBoost<sup>[7]</sup> 等模型, SECF<sup>2</sup> 更简单、易于实现, 仅需少量特征和算力便能获取更优的识别精度.

2) 它是集早期融合 (特征融合)、后期融合 (集成学习) 于一体的材质图像识别模型, 具有较强通用性, 可复用到其他图像识别任务, 以解决更多实际问题.

## 1 相关工作

### 1.1 材质图像识别

材质图像识别是机器视觉领域中非常重要的研究方向. 如 Lai 等<sup>[12]</sup> 通过双向反射分布函数完成材质识别, 然而获取光谱反射十分耗时且易出错. Chen 等<sup>[13]</sup> 提出韦伯定律描述符构造局部特征. Liu 等<sup>[14]</sup> 引入基于反射的边缘特征. Zhang 等提出 GS-XGBoost<sup>[7]</sup> 和 HMF<sup>2</sup><sup>[8]</sup> 模型, 从特征融合角度改善识别性能. 显然, 图像特征学习是提高材质图像识别精度的关键. 近年来, 深度学习模型在材质图像识别中发挥了重要作用. Cimpoi 等<sup>[15]</sup> 基于卷积神经网络计算 Fisher 向量并获得较好性能. Bell 等<sup>[5]</sup> 训练目标识别网络以实现像素级材质标记. Schwartz 等<sup>[6]</sup> 基于卷积神经网络完成局部材质的识别. 综上, 现有研究<sup>[5-8]</sup> 多使用卷积神经网络中某一层特征完成材质图像识别, 未考虑同构网络中异构层特征在识别中的互补作用. 受 MDCA<sup>[9]</sup>、HMF<sup>2</sup><sup>[8]</sup> 等模型启发, 本文深入挖掘 SENet 中异构层特征间的互补信息, 能更准确地刻画材质图像并改善识别精度.

### 1.2 特征融合

为充分利用不同特征间的互补性, 各类特征融合方法应运而生, 可分为早期融合、中期融合和后期融合. 早期融合在特征层融合特征, 例如将异构特征映射到同一语义空间, 采用求和或拼接操作完成特征融合. 相关模型为深度典型相关分析 (deep canonical correlation analysis, DCCA)<sup>[16]</sup>. 早期融合实现简单、操作方便, DenseNet (densely connected convolutional networks)<sup>[17]</sup>、ResNet<sup>[18]</sup> 模型均借鉴该思想. 中期融合对特征线性加权, 训练一个更具判别性的分类模型, ERGS<sup>[7]</sup> 属中期融合. 后期融合集成多个预测结果以构建一个更强的分类模型, 例如 XGBoost<sup>[19]</sup>、CatBoost<sup>[20]</sup> 均执行后期融合. 综上, 本文设计一种集特征早期融合、后期融合于一体的材质图像识别模型, 充分发挥不同融合方法的优点.

## 2 SECF<sup>2</sup> 模型

### 2.1 模型框架

SECF<sup>2</sup> 模型框架如图 1 所示, 它包括: 异构层特征提取 (选用 SE-ResNet50、SE-ResNet101、SE-ResNet152、SE-ResNeXt50 及 SE-ResNeXt101 五种 SENet 网络的预训练模型进行迁移学习, 抽取 2048 维平均池化层特征, 池化层保留网络关键特征, 减少参数, 防止过拟合并提高特征泛化能力), 生成深层视觉语义 (基于鲁棒、高效且异构的 SENet 特征, 基于改进的 Cluster-CCA 模型分析异构特征间隐含的语义

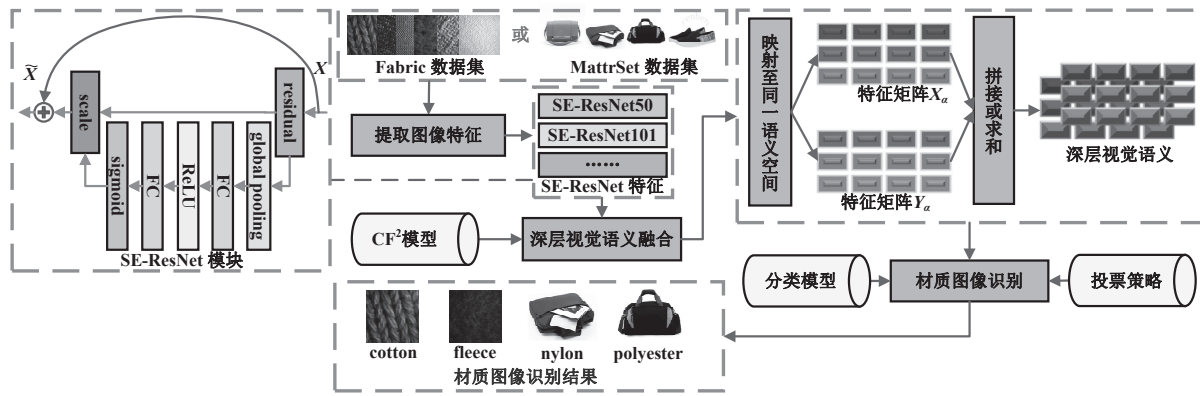


图1 SECF<sup>2</sup>模型框架

关联,即映射到同构子空间中的一组聚类典型相关性,基于相关性从互补的角度完成早期融合,更准确、高效地刻画材质视觉特性)以及集成学习(从互补角度选择多组分类模型,基于投票策略进一步提升识别准确率)。

综上,SECF<sup>2</sup>模型执行步骤为:1)提取一组异构层特征;2)采用CF<sup>2</sup>(cluster-CCA feature fusion)模型挖掘异构层特征间蕴含的视觉语义,融合生成深层视觉语义;3)设计集成学习方法,进一步改善识别性能。

### 2.2 改进的cluster-CCA: CF<sup>2</sup>模型

改进cluster-CCA为CF<sup>2</sup>模型,挖掘异构层特征间的深层视觉语义。cluster-CCA模型将多模态信息映射到同一语义空间完成相关任务,如跨模态检索<sup>[21]</sup>,仅利用了单模态映射后的相关性,并未发挥多模态映射信息的互补作用。因此,受MDCA<sup>[9]</sup>、HMF<sup>2</sup><sup>[8]</sup>等模型启发,本文在特征映射基础上设计融合策略,充分利用SENet异构层特征间的聚类典型相关性,生成更有效的深层视觉语义,最终训练出优质的材质图像分类模型。此外,CF<sup>2</sup>模型能自动处理多模态信息,易于部署至服务器,运行效率良好(参见3.6节)。综上,CF<sup>2</sup>模型如算法1所示。

#### 算法1 CF<sup>2</sup>模型。

输入: 材质图像数据集  $I \in \{(x_1, y_1), \dots, (x_n, y_n)\}$ ,  $x$  和  $y$  分别为图像及其所对应标签,  $n$  为材质图像总数;

输出: 聚类典型相关性融合特征。

1) 将特征命名为  $X, X^*, Y$  及  $Y^*$ ,  $X$  和  $Y$  是基于训练集的特征,  $X^*$  和  $Y^*$  是基于测试集的特征。特征均有  $C$  类, 则  $T_x = \{X_1, \dots, X_C\}, T_y = \{Y_1, \dots, Y_C\}$ , 其中  $X_c = \{X_1^c, \dots, X_{|X_c|}^c\}, Y_c = \{Y_1^c, \dots, Y_{|Y_c|}^c\}$  分别表示在第  $C$  组中  $X$  和  $Y$  的数据。

2)  $w, v$  分别为  $X, Y$  所对应的投影向量。  $X, Y$  的相关系数  $\rho$  为

$$\rho = \max_{w,v} \frac{w' C_{XY} v}{\sqrt{w' C_{XX} w} \sqrt{v' C_{YY} v}}$$

3) 协方差矩阵定义如下:

$$C_{XY} = \frac{1}{S} \sum_{c=1}^C \sum_{i=1}^{|X_c|} \sum_{j=1}^{|Y_c|} x_i^c y_j^c,$$

$$C_{XX} = \frac{1}{S} \sum_{c=1}^C \sum_{i=1}^{|X_c|} |Y_c| x_i^c x_i^c,$$

$$C_{YY} = \frac{1}{S} \sum_{c=1}^C \sum_{j=1}^{|Y_c|} |X_c| y_j^c y_j^c,$$

其中  $S = \sum_{c=1}^C |X_c| |Y_c|$  为  $X$  和  $Y$  的成对关系总对数。

4) 最大化  $\rho$ , 得到特征  $X$  和  $Y$  所对应的投影向量  $w, v$ 。

5) 基于  $w$  和  $v$  变换  $X, X^*, Y$  及  $Y^*$ , 生成映射特征矩阵  $X_\alpha, X_\alpha^*, Y_\alpha$  和  $Y_\alpha^*$ , 并构造训练集和测试集特征  $U$  和  $V$ , 有

$$\begin{cases} X_\alpha = w' X, X_\alpha^* = w' X^*; \\ Y_\alpha = v' Y, Y_\alpha^* = v' Y^*. \end{cases}$$

$$U = \begin{pmatrix} X_\alpha \\ X_\alpha^* \end{pmatrix}, V = \begin{pmatrix} Y_\alpha \\ Y_\alpha^* \end{pmatrix}.$$

6) 借鉴MDCA、HMF<sup>2</sup>等模型,对  $U$  和  $V$  执行早期融合,生成聚类典型相关性融合特征  $F_1, F_2$ , 即

$$F_1 = \begin{pmatrix} U \\ V \end{pmatrix} = \begin{pmatrix} X_\alpha & Y_\alpha \\ X_\alpha^* & Y_\alpha^* \end{pmatrix},$$

$$F_2 = U + V = \begin{pmatrix} X_\alpha + X_\alpha^* \\ Y_\alpha + Y_\alpha^* \end{pmatrix}.$$

### 2.3 SECF<sup>2</sup>模型

CF<sup>2</sup>模型输出的聚类典型相关性特征融合结果包括:基于SENet特征的CF<sub>S</sub><sup>2</sup>、基于ResNet特征的CF<sub>R</sub><sup>2</sup>(实验对比用)、基于SENet和ResNet特征的CF<sub>SR</sub><sup>2</sup>

(实验对比用). 在此基础上可以训练分类模型, 完成材质图像识别. MDCA、GS-XGBoost、HMF<sup>2</sup>等模型需大量异构特征并执行多次融合, 模型较为复杂. 而 SECF<sup>2</sup> 模型仅需一类深度学习网络的异构层特征, 且只执行一次特征融合, 使模型复杂度大大降低, 便于实现和部署. SECF<sup>2</sup> 模型的执行步骤为:

1) 选取主流深度学习网络 SENet, 自适应地增强特征表征能力<sup>[10]</sup>;

2) 基于 SENet 抽取异构层特征;

3) 基于 CF<sup>2</sup> 模型分析异构层特征间的聚类典型相关性, 充分利用异构层特征间互补性, 输出深层视觉语义;

4) 采用集成学习方法对分类结果增强, 提升材质图像识别性能.

综上, SECF<sup>2</sup> 模型如算法2所示.

**算法2** SECF<sup>2</sup> 模型.

输入: 材质图像数据集  $I \in \{(x_1, y_1), \dots, (x_n, y_n)\}$ ,  $x$  和  $y$  分别为图像及其所对应标签,  $n$  为材质图像总数;

输出: 材质图像识别结果.

1) 提取 5 类异构层特征: SE-ResNet50、SE-ResNet101 等, 得到 Features =  $\{f_1, f_2, f_3, f_4, f_5\}$ .

2) 对 Features 中的特征两两组合, 一共有 10 种组合, 即

$$D = \{f\_comb_1, \dots, f\_comb_k\}, k \in \{1, 2, \dots, 10\}.$$

3) repeat.

选取  $D$  中  $f\_comb_k$  输入 CF<sup>2</sup> 模型中生成 CF<sub>S</sub><sup>2</sup>, 直到生成全部特征组合所对应的 CF<sub>S</sub><sup>2</sup>.

4) repeat.

将 CF<sub>S</sub><sup>2</sup> 输入 6 类分类模型中, 分别计算识别性能, 直到全部 CF<sub>S</sub><sup>2</sup> 均执行完分类.

5) 根据识别性能对分类模型排序, 生成 voting\_classifier =  $\{top_2, top_3, \dots, top_i\}$ , 例如 top<sub>2</sub> 表示排名前二的两个分类模型.

6) repeat.

选取 4) 中表现最优的 CF<sub>S</sub><sup>2</sup>, 使用 top<sub>i</sub> 分类模型执行投票策略, 直到 voting\_classifier 中分类模型执行完毕.

7) 输出材质图像识别结果.

## 3 实验结果

### 3.1 数据集及基线

#### 3.1.1 Fabric 和 MattrSet 数据集

Fabric<sup>[22]</sup> 数据集是由光度立体传感器采集实体店 2 000 多件衣物样本, 进而构建的细粒度数据集,

共有 9 种材质, 即 cotton(棉)、wool(毛毡)、terrycloth(厚绒布)、nylon(尼龙)、silk(丝绸)、denim(牛仔布)、viscose(纤维)、fleece(羊毛) 和 polyester(涤纶). 数据集共有 1 266 个样本, 各样本以不同视角、不同光照为条件获取 4 张图像, 共计 5 064 张图像, 约 1.7 GB 数据.

MattrSet<sup>[7]</sup> 数据集来自 Web 上的真实商品图像 (<https://www.made-in-china.com/>), 是材料学专家指导构建的粗粒度数据集. 数据集包含 bags 和 shoes 两大类商品: bags 类分为 pu(皮革)、canvas(帆布)、nylon(尼龙) 和 polyester(涤纶) 4 种材质; shoes 类仅包含 pu 和 canvas 两种材质. 因此, 数据集由 6 类、11 021 张图像组成, 约 3.2 GB 数据.

本文在这两个基准数据集上验证 SECF<sup>2</sup> 模型的有效性和鲁棒性. 为与基线进行公平比较, 所有实验均以四折交叉验证方式完成.

#### 3.1.2 基线

SECF<sup>2</sup> 使用 6 类分类模型: 逻辑回归 (logistic regression, LR)、支持向量机 (support vector machines, SVM)、 $K$  最近邻 ( $K$ -nearest neighbor, KNN)、决策树 (decision tree, DT)、自适应提升 (adaptive boosting, AdaBoost) 和极端梯度提升 (eXtreme gradient boosting, XGBoost). 故 SECF<sup>2</sup> 模型包含一组变体: SECF<sub>KNN</sub><sup>2</sup>、SECF<sub>LR</sub><sup>2</sup>、SECF<sub>DT</sub><sup>2</sup>、SECF<sub>SVM</sub><sup>2</sup>、SECF<sub>Ada</sub><sup>2</sup> 和 SECF<sub>XGB</sub><sup>2</sup>. 对比基线如下:

1) 单一模型. 主流深度学习模型: DenseNet169<sup>[17]</sup>、Kampouris<sup>[22]</sup> 模型、ResNeXt WSL<sup>[23]</sup> 和 SENet<sup>[10]</sup>.

2) 融合类模型. 特征融合模型: GS-XGBoost<sup>[7]</sup>、HMF<sup>2</sup><sup>[8]</sup>、改进的 ERGS<sup>[24]</sup> 和 MDCA<sup>[9]</sup>; SECF<sup>2</sup> 模型的变种, 即基于 CF<sub>R</sub><sup>2</sup>、CF<sub>SR</sub><sup>2</sup> 的模型: SECF<sub>α</sub><sup>2</sup> 与 SECF<sub>β</sub><sup>2</sup>; 集成学习 (后期融合) 类模型: AdaBoost、XGBoost、CatBoost<sup>[20]</sup> 和 LightGBM<sup>[25]</sup>.

#### 3.2 评估指标

采用精准度 (accuracy) 评估模型. 采用平均精准度 (average accuracy, AA) AA<sub>F</sub> 综合评价特征优劣, 即

$$AA_F = \frac{\sum_{m=1}^{N_{\text{classifier}}} (\text{accuracy}_m)}{N_{\text{classifier}}}. \quad (1)$$

采用平均精准度 AA<sub>C</sub> 综合评价分类模型优劣, 即

$$AA_C = \frac{\sum_{n=1}^{N_{\text{feature}}} (\text{accuracy}_n)}{N_{\text{feature}}}. \quad (2)$$

采用均值平均精度 (mean average accuracy, MAA) 全面评估模型性能, 即

$$MAA = \frac{\sum_{n=1}^{N_{\text{classifier}}} \sum_{m=1}^{N_{\text{feature}}} (\text{accuracy}_{(m)(n)})}{N_{\text{classifier}} \times N_{\text{feature}}}. \quad (3)$$

3.3 实验结果与分析

3.3.1 基于SECF<sup>2</sup>模型的实验结果

所有实验均在如下配置完成: Intel Core i7-7700 CPU (2.8 GHz)、16 G RAM 及 GTX 1050 GPU, 属于普通类计算资源.

表1为SECF<sup>2</sup>模型生成的深层视觉语义CF<sub>S</sub><sup>2</sup>的识别精准度. 其中: SR101-SRxt101表示深层视觉语

义, SR是SE-ResNet的缩写, SRxt是SE-ResNeXt的缩写, 而

$$\Delta\text{accuracy}_1 = AA_{F-\text{SENet}} - AA_{F-\text{ResNet}} \quad (4)$$

为SENet较ResNet的AA<sub>F</sub>指标提升幅度,

$$\Delta\text{accuracy}_2 = AA_{F-\text{CF}_S^2} - AA_{F-\text{SENet}} \quad (5)$$

为CF<sub>S</sub><sup>2</sup>较SENet的AA<sub>F</sub>指标提升幅度, Δaccuracy<sub>1</sub> + Δaccuracy<sub>2</sub>为CF<sub>S</sub><sup>2</sup>特征较ResNet特征的AA<sub>F</sub>指标提升幅度.

表2为ResNet、SE-ResNet和CF<sub>S</sub><sup>2</sup>的AA<sub>F</sub>与MAA两大指标. 表1和表2中粗体为最优值.

表1 基于SECF<sup>2</sup>模型的识别精准度

%

数据集	CF <sub>S</sub> <sup>2</sup>	各SECF <sup>2</sup> 模型的识别精准度						
		KNN	LR	DT	SVM	Ada	XGB	Avg
Fabric	SR101-SRxt101	89.08	91.13	69.83	<b>86.30</b>	87.46	87.16	<b>85.16</b>
	SRxt101-SR152	89.44	91.05	<b>71.01</b>	85.07	86.95	86.85	85.06
	SRxt50-SR152	<b>89.55</b>	90.80	67.73	85.76	88.19	87.54	84.93
	SR50-SRxt101	87.99	<b>91.23</b>	69.14	83.31	<b>88.45</b>	<b>88.19</b>	84.72
	Δaccuracy <sub>1</sub>	47.79	35.54	32.86	34.50	37.98	36.62	37.55
	Δaccuracy <sub>2</sub>	2.79	3.59	6.21	4.61	1.59	1.61	3.40
MatrSet	SR50-SR152	<b>66.23</b>	<b>68.94</b>	57.16	69.97	67.26	68.68	<b>66.37</b>
	SR50-SR101	65.48	68.72	<b>57.86</b>	69.70	67.49	68.40	66.27
	SR50-SRxt101	64.91	68.88	57.01	<b>70.30</b>	67.46	<b>69.02</b>	66.26
	SRxt101-SR152	65.15	68.73	56.88	69.61	<b>67.79</b>	68.89	66.18
	Δaccuracy <sub>1</sub>	16.12	11.07	19.84	43.81	18.60	14.40	20.63
	Δaccuracy <sub>2</sub>	1.19	1.11	9.37	1.10	4.39	2.90	3.35

表2 各特征的AA<sub>F</sub>、MAA及最优值(MAX)指标

%

数据集	特征	AA <sub>F</sub>					MAA	MAX
		KNN	LR	DT	SVM	Ada		
Fabric	ResNet	38.38	<b>51.84</b>	30.34	46.25	48.15	49.34	53.79
	SE-ResNet	86.18	<b>87.37</b>	63.19	80.75	86.12	85.96	89.38
	CF <sub>S</sub> <sup>2</sup>	88.97	<b>90.97</b>	69.41	85.36	87.71	87.57	91.23
MatrSet	ResNet	48.14	<b>56.52</b>	27.93	24.89	44.43	51.33	58.76
	SE-ResNet	64.26	67.59	47.77	<b>68.70</b>	63.03	65.73	69.34
	CF <sub>S</sub> <sup>2</sup>	65.45	68.70	57.15	<b>69.80</b>	67.42	68.63	70.30

综合表1和表2可得如下重要结论:

1) 由Δaccuracy<sub>1</sub>可知: 对于Fabric数据集, SE-ResNet特征较ResNet的最大提升幅度为47.79%, Δaccuracy<sub>1</sub>均值达到37.54%, 这说明, SE模块确实能增强模型对材质视觉特性的刻画, 这种改变是鲁棒的, 适用于不同的分类模型; MatrSet数据集亦是如此.

2) 由Δaccuracy<sub>2</sub>可知: 对于Fabric数据集, CF<sub>S</sub><sup>2</sup>较SE模块的最大提升为6.21%, Δaccuracy<sub>2</sub>均值达到3.40%; MatrSet数据集所对应的最大提升为9.37%; Δaccuracy<sub>2</sub>均值达到3.35%. 这说明, 无论是细粒度数据集, 还是粗粒度数据集, SECF<sup>2</sup>模型都能有效改善材质图像识别性能, SECF<sup>2</sup>模型是有效且鲁棒的.

3) 由表 1 和表 2 不难发现: SR50-SRxt101 在 Fabric 和 MattrSet 数据集上均取得峰值, 较 SE-ResNet 最优指标分别提升 1.85% (SR50-SRxt101+LR 模型) 和 0.96% (SR50-SRxt101+SVM).  $CF^2$  模型是有效和鲁棒的: 它通过挖掘 SENet 异构层特征间的有效信息, 逐步改善识别性能, 并在两种粒度的数据集上均表现优异, 故新特征是鲁棒的. 值得注意的是,  $SECF^2$  模型在细粒度数据集上的表现优于粗粒度数据集, 因为细粒度图像中蕴含更多有价值的材质视觉细节, 而粗粒度图像来自 Web, 包含一定噪声, 会干扰识别模型.

4) 由表 2 中  $AA_F$  可知: 无论是 Fabric 还是 MattrSet 数据集, SE-ResNet 较 ResNet 均有大幅提升,  $CF^2_S$  较 SE-ResNet 亦有提升. 由 MAA 指标知: Fabric 数据集中, SE-ResNet 的综合识别性能较 ResNet 提升 37.55%,  $CF^2_S$  的综合识别性能较 SE-ResNet 提升 3.4%; MattrSet 数据集所对应的值分别为 20.63% 和 3.35%. 这充分说明: 无论细粒度还是粗粒度数据集, SE-ResNet 均能有效改善模型对材质视觉属性的刻画能力, 它适用于不同的分类模型, 具备较好的鲁棒性;  $CF^2$  模型挖掘异构层特征间的相关性, 使特征间的互补信息得到充分利用, 进而更准确地刻画材质属性, 并优化识别精度. 此外, 经过特征融合, 即使 LR、SVM 等传统模型也能获取较好的性能.

### 3.3.2 基于投票策略的实验结果

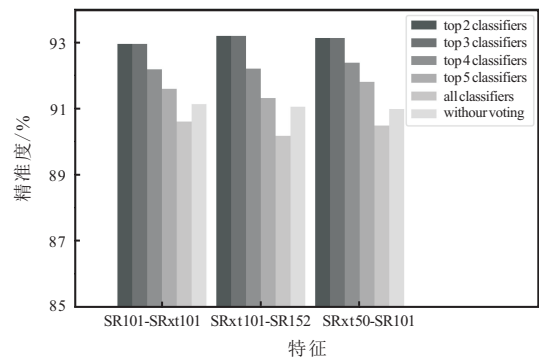
图 2 给出了采用投票策略后,  $CF^2_S$  中  $AA_C$  指标 top 3 的特征组合的实验结果.

在图 2(a) 中, 基于软投票机制, SRxt101-SR152 特征结合 top 2 (或 top 3) 的分类模型, 在 Fabric 数据集上获得最佳识别精准度 (93.19%), 较硬投票最优值提升 1.68%, 较 3.3.1 节中最高值 (表 2 的 91.23%) 提升 1.96%. 因此, 投票策略在 Fabric 数据集上有效.

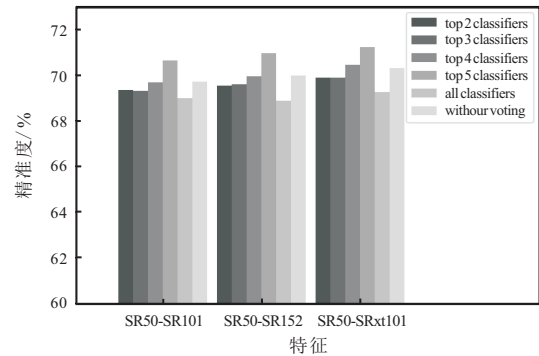
由图 2(b) 知, 在 MattrSet 数据集上, 相比于软投票策略 (0.92%), 硬投票策略性能改进 (0.51%) 不够显著. 这进一步说明, 粗粒度材质图像中的噪声会影响识别性能. 总体而言, 投票策略在 MattrSet 数据集上有效. 而  $HMF^2$  模型<sup>[8]</sup> 在粗粒度材质图像数据集上表现更优, 故  $SECF^2$  模型是对  $HMF^2$  的有力补充. 相比于  $HMF^2$ ,  $SECF^2$  模型仅需两种异构层特征, 这大大降低了模型复杂度和训练难度, 为缩小理论研究与实际应用间的“鸿沟”奠定了重要基础.

### 3.3.3 与主流基线比较

选取多个主流基线与  $SECF^2$  模型进行比较, 结果如表 3 所示 (其中:  $SECF^2$ (soft) 表示基于软投票策



(a) Fabric 数据集: 软投票



(b) MattrSet 数据集: 软投票

图 2 各数据集的软投票策略识别精准度

略的  $SECF^2$  模型, S (single)、F (fusion) 和 M (main) 分别代表单一模型, 融合类模型及主流材质图像识别模型). 在 Fabric 数据集中,  $SECF^2$  优于全部基线, 它较最强的  $HMF^2$  模型提升 8.85%. 在 MattrSet 数据集中,  $SECF^2$  性能逼近最强基线. 此外,  $SECF^2$  的变种模型表现也不错. 例如, 在 Fabric 数据集上,  $SECF^2_\beta$  较最强基线  $HMF^2$  提升 4.58%. 这进一步表明了异构层特征融合的必要性及  $SECF^2$  模型的鲁棒性.

与单一模型相比: Fabric 数据集上,  $SECF^2$  模型较最强基线 SENet 提升 3.81%; 在 MattrSet 数据集上, 较最优基线 SENet 提升 1.88%. 这表明充分利用模型间的互补性能有效改善材质图像识别性能. 不同于单一模型,  $SECF^2$  还利用了异构层特征间的互补性. 因此, 早期融合与后期融合形成合力, 使  $SECF^2$  表现优异.

相比于 GS-XGBoost<sup>[7]</sup>、 $HMF^2$ <sup>[8]</sup> 等融合类模型,  $SECF^2$  在 Fabric 数据集上有 8.85% 的性能提升, 在 MattrSet 数据集上亦逼近最优值, 故  $SECF^2$  模型是有效的. 不同于上述融合类模型,  $SECF^2$  仅需两种 SENet 异构层特征, 只执行一次特征融合, 故模型复杂度更低. 此外, 本文使用多折交叉验证完成实验, 进一步支撑了  $SECF^2$  模型的鲁棒性. 综上,  $SECF^2$  模型是有效和鲁棒的, 它易于实现, 便于对接实际应用.

除精准度外, 本文对  $SECF^2$  模型进行实时效率评

表3 与主流基线对比

%

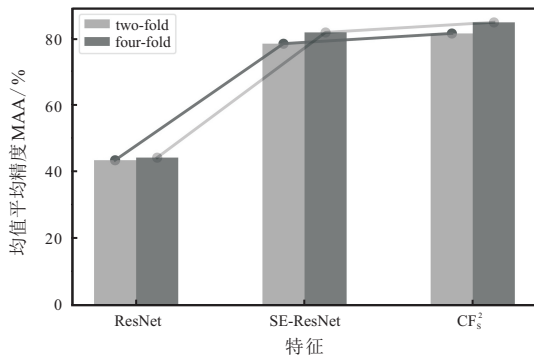
数据集	模型	模型类别	accuracy	数据集	模型	模型类别	accuracy
Fabric	DCA-XGBoost <sup>[9]</sup>	F	65.36	MatrSet	DCA-XGBoost <sup>[9]</sup>	F	49.84
	GS-XGBoost-SG <sup>[7]</sup>	F&M	84.20		GS-XGBoost-SGL <sup>[7]</sup>	F&M	67.67
	Densenet169 <sup>[17]</sup>	S	46.29		Densenet169 <sup>[17]</sup>	S	59.77
	HMF <sup>2[8]</sup>	F&M	84.34		HMF <sup>2[8]</sup>	F&M	<b>72.20</b>
	Albe+Norm <sup>[22]</sup>	S&M	79.60		LR-R152	S	58.76
	ResNeXt WSL <sup>[23]</sup>	S	59.66		ResNeXt WSL <sup>[23]</sup>	S	67.87
	SENet <sup>[10]</sup>	S	89.38		SENet <sup>[10]</sup>	S	69.34
	CatBoost <sup>[20]</sup>	F	88.47		CatBoost <sup>[20]</sup>	F	68.68
	LightGBM <sup>[25]</sup>	F	88.03		LightGBM <sup>[25]</sup>	F	68.72
	VGG&Sift_RBF <sup>[24]</sup>	F&M	79.09		VGG-16&S.G.L_RBF <sup>[24]</sup>	F&M	63.11
	SECF <sup>2<sub>α</sub></sup>	F	58.63		SECF <sup>2<sub>α</sub></sup>	F	62.61
	SECF <sup>2<sub>β</sub></sup>	F	88.92		SECF <sup>2<sub>β</sub></sup>	F	68.67
	SECF <sup>2<sub>LR</sub></sup>	F	91.23		SECF <sup>2<sub>LR</sub></sup>	F	70.29
	SECF <sup>2</sup> (hard)	F	91.61		SECF <sup>2</sup> (hard)	F	70.81
	SECF <sup>2</sup> (soft)	F	<b>93.19</b>		SECF <sup>2</sup> (soft)	F	71.22

估(参见3.6节的表5)。SECF<sup>2</sup>模型识别一张图像需5.89s,较最优基线HMF<sup>2</sup>少用11s。这说明SECF<sup>2</sup>模型的实时效率较好。综上,SECF<sup>2</sup>模型合理且有效,具备较好的运行效率。此外,模型具有较强的通用性,详见3.5节。

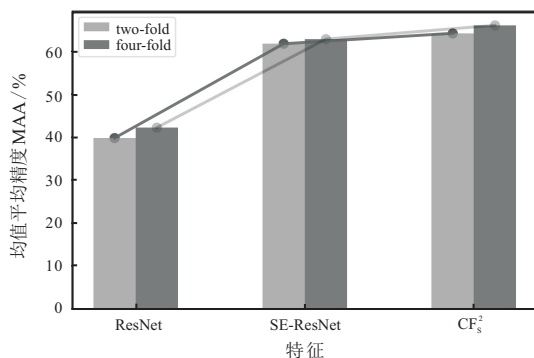
### 3.4 参数调制

#### 3.4.1 交叉验证分析

图3分别展示了在两折、四折交叉验证下,基



(a) Fabric数据集



(b) MatrSet数据集

图3 二折交叉验证与四折交叉验证的材质识别性能对比

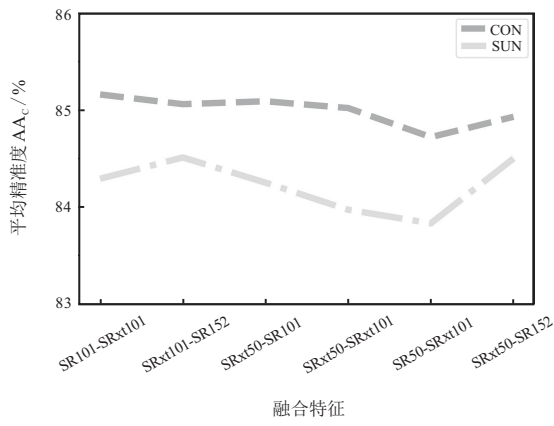
于ResNet、SE-ResNet和CF<sub>s</sub><sup>2</sup>特征的材质图像识别性能,使用更具统计意义的MAA为评价指标。如图3所示,伴随SE模块、CF<sup>2</sup>模型的使用,识别性能呈上升趋势。这说明SE模块能生成判别性更强的SE-ResNet特征,选取SENet异构层特征执行早期融合是合理的(参见表2)。其次,深层视觉语义是有效的:一方面,SENet异构层特征从不同角度刻画了材质视觉特性;另一方面,CF<sup>2</sup>模型充分挖掘异构层特征间的聚类典型相关性,生成了判别性更强且鲁棒的深层视觉语义CF<sub>s</sub><sup>2</sup>。再次,四折验证性能优于两折,进一步说明SECF<sup>2</sup>模型是鲁棒的,这有助于增强模型实用性。Fabric数据集较MatrSet的识别性能更优,原因同上。综上,以四折交叉验证完成了本文实验。

#### 3.4.2 特征融合方式比较

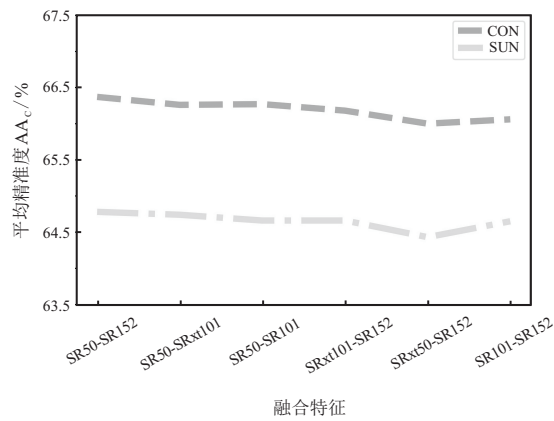
SECF<sup>2</sup>模型使用两种特征融合方式,即求和(SUM)和拼接(CON)。为更全面地评估这两种融合方式,使用表1中CF<sub>s</sub><sup>2</sup>特征完成相关评估。本实验使用6类分类模型以四折交叉验证执行材质图像识别,并选用AA<sub>C</sub>指标绘制折线图形,结果如图4所示。可见,在两大基准数据集中,拼接融合均优于求和融合。求和操作降低了特征维度,进而稀释了特征中关键判别信息,而拼接操作生成的维度较合适,它极大地保留了来自SENet异构层特征中的聚类典型相关信息。因此,SECF<sup>2</sup>模型使用拼接操作完成特征融合。

### 3.5 模型通用性实证

为了验证SECF<sup>2</sup>模型的通用性,本文选取Twitter<sup>[26]</sup>和FI<sup>[27]</sup>(Flickr and Instagram)数据集完成图像情感分析任务<sup>[28-33]</sup>。



(a) Fabric数据集



(b) MatrxSet数据集

图4 特征融合方式比较

Twitter1 数据集来自社交网站,由正面 (Pos, positive) 和负面 (Neg, negative) 两类共计 1269 张图像组成,共 57.5 MB,可在 <https://www.cs.rochester.edu/u/q-you/DeepSent/deepsentiment.html> 下载。FI 数据集分 8 类情绪关键词,共计 23 308 张图像,约 1.5 G,

可从 <https://www.cs.rochester.edu/u/qyou/deepemotion/index.html> 下载。两数据集均随机选取 70% 的数据作为训练集,剩余 30% 数据作为测试集。

为了验证 SECF<sup>2</sup> 模型通用性,选取多个主流基线进行比较(参见表 4。其中:D(deep)、M(main)和 F(fusion)分别代表深度学习、主流图像情感分析及融合类模型;在全部基线中,D5 表示使用 D5 数据进行评估,其余使用 D3 数据,因 5 名标注者给出一致标签,故 D5 数据的质量优于 D3)。基线描述如下:

1) 深度学习模型: VGG16<sup>[34]</sup>、 AlexNet<sup>[35]</sup>、 ResNet<sup>[18]</sup>、 SENet<sup>[10]</sup> 和 CAM-Res101<sup>[36]</sup>。

2) 图像情感分析模型: Yang 等模型<sup>[32]</sup>、 WILDCAT(weakly supervised learning of deep convNets)<sup>[29]</sup>、 WSCNet(weakly supervised coupled networks)<sup>[30]</sup>、 Sun 等模型<sup>[31]</sup>、 ME<sup>2</sup>M(M)<sup>[33]</sup> 和 SR-w-DCA<sup>[33]</sup>。

3) 融合类模型: 稀疏核典型相关分析 (gradient kernel canonical correlation analysis, gradKCCA)<sup>[37]</sup>、 MDCA<sup>[9]</sup> 和 GS-XGBoost<sup>[7]</sup> 模型。

由表 4 知,无论是 Twitter1 还是 FI 数据集,SECF<sup>2</sup> 模型均优于全部对比基线: 1) 与深度学习模型相比,SECF<sup>2</sup> 的识别性能有较大幅度提升。如 Twitter1 数据集上,SECF<sup>2</sup>(hard) 相比于最强基线提升 2.80%,FI 数据集提升幅度为 3.07%。这是因为深度学习网络过于复杂,且缺乏高质量样本,导致出现过拟合。相反,SECF<sup>2</sup> 从增强特征判别性的角度改善其认知图像情感的能力,而集成学习策略能进一步提升其识别能力。此外,SECF<sup>2</sup> 模型复杂度远低于深度学习网

表4 SECF<sup>2</sup> 模型与主流图像情感分析基线比较

数据集	模型	模型类别	accuracy	数据集	模型	模型类别	accuracy
Twitter1	GS-XGBoost <sup>[7]</sup>	F	73.21	FI	GS-XGBoost <sup>[7]</sup>	F	43.32
	AlexNet <sup>[35]</sup>	D	73.24		AlexNet <sup>[35]</sup>	D	58.13
	VGG16 <sup>[34]</sup>	D	76.75		VGG16 <sup>[34]</sup>	D	63.75
	CAM-Res101 <sup>[36]</sup>	D	82.67		CAM-Res101 <sup>[36]</sup>	D	68.54
	gradKCCA(D5) <sup>[37]</sup>	F	77.07		gradKCCA(D5) <sup>[37]</sup>	F	61.84
	ResNet(D5) <sup>[9]</sup>	D	66.96		ResNet(D5) <sup>[9]</sup>	D	44.36
	SENet(D5) <sup>[12]</sup>	D	87.35		SENet(D5) <sup>[12]</sup>	D	77.21
	MDCA(D5) <sup>[9]</sup>	F	86.47		MDCA(D5) <sup>[9]</sup>	F	73.71
	WILDCAT <sup>[29]</sup>	M	79.53		WILDCAT <sup>[29]</sup>	M	67.03
	Sun 等 (D5) <sup>[31]</sup>	M	88.94		Yang 等 (D5) <sup>[32]</sup>	M	67.48
	WSCNet <sup>[30]</sup>	M	84.25		WSCNet <sup>[30]</sup>	M	70.07
	ME <sup>2</sup> M(M) <sup>[33]</sup>	M&F	87.15		SR-w-DCA <sup>[33]</sup>	M&F	75.72
	SECF <sup>2</sup> (hard)(D5)	F	<b>90.15</b>		SECF <sup>2</sup> (hard)(D5)	F	79.13
	SECF <sup>2</sup> (soft)(D5)	F	89.77		SECF <sup>2</sup> (soft)(D5)	F	<b>80.28</b>

络,这有助于减少资源消耗并加速复现,从而提升模型实用价值. 2)与特征融合类模型相比, Twitter1数据集上, SECF<sup>2</sup>模型较最强基线 DCA(D5)提升3.68%, FI数据集提升幅度是6.57%. 虽然均采用特征融合思想,但不同于 MDCA、GS-XGBoost等模型, SECF<sup>2</sup>仅需两类特征及一次融合便能获得非常不错的识别性能,这进一步提升了模型的实用价值. 3)与主流图像情感分析模型相比, SECF<sup>2</sup>模型同样表现优异. Twitter1数据集上, SECF<sup>2</sup>(hard)较最强基线 Sun等(D5)<sup>[31]</sup>提升1.21%, FI数据集提升幅度为4.56%. 相比于 ME<sup>2</sup>M(M)、SR-w-DCA等模型, SECF<sup>2</sup>无需执行额外的数据增强. 相比于 SPN、WILDCAT等模型依赖额外情感区域标注, SECF<sup>2</sup>无需人工干预,实用性更高.

综上, SECF<sup>2</sup>模型是有效且鲁棒的. 它在材质图像识别、图像情感分析两类任务中均表现优异. 同时,它易于实现、部署简单、无需人工干预,这均有助于缩小理论研究与实际应用之间的“鸿沟”.

### 3.6 实时材质识别系统

为验证 SECF<sup>2</sup>模型实用性,设计基于 SECF<sup>2</sup>模型的实时材质识别系统,提供材质图像在线识别和移动识别两大功能:用户上传本地图像或使用移动IP摄像头拍摄实时材质图像,系统基于 SECF<sup>2</sup>模型展开

特征提取、深层视觉语义挖掘、集成学习,最终完成材质识别. 部署系统仅需配置 Python 深度学习环境并安装 Matlab, 识别操作均在前台网页执行,无需人工干预,可提升用户交互体验,并为后续配置云服务端奠定重要基础.

表5为计算实时材质识别系统的运行效率(计算机配置参见3.3.1节)并与基线进行比较,可见,无论是特征提取与融合,还是图像识别, SECF<sup>2</sup>模型的运行效率均优于基线. 若采用高配置计算机,则实时效率将进一步提升. 不同于 HMF<sup>2</sup>、GS-XGBoost等模型需提取大量特征并执行多次融合操作, SECF<sup>2</sup>仅需两类异构层特征及一次特征融合,模型复杂度大大降低,它更易部署且实时效率大幅提升. 因此,综合识别精度和运行效率,该实时材质识别系统能缩小理论研究与实际应用(场景识别、工业检测、机器人视觉、即时购物等)之间的“鸿沟”. 例如:在工业检测领域,本系统可辅助纺织企业更高效地完成织物疵点分类;在商场购物时,顾客只需一部手机便可快速、准确地鉴别待购商品材质,以辅助其购买决策. 该系统还可与智能机器人(如 NAO<sup>[38]</sup>)相结合,完成实时衣物识别与抓取,让机器人更好地辅助人们的工作、生活. 综上, SECF<sup>2</sup>模型能与多个日常应用无缝对接,以满足人们的实际需求.

表5 实时运行效率比较

单位: s

特征	模型	特征提取与融合时间/图像	识别时间/图像	总时间/图像
Fabric	HMF <sup>2</sup> <sup>[8]</sup>	16.72	0.14	16.86
	MDCA <sup>[9]</sup>	19.98	0.01	19.99
	GS-XGBoost <sup>[7]</sup>	10.45	1.15	11.60
	SECF <sup>2</sup>	5.69	0.20	5.89
MatrSet	HMF <sup>2</sup> <sup>[8]</sup>	18.43	0.26	18.69
	MDCA <sup>[9]</sup>	19.67	0.18	19.85
	GS-XGBoost <sup>[7]</sup>	14.87	1.94	16.82
	SECF <sup>2</sup>	5.61	0.36	5.98

## 4 结论

材质图像识别在生产、生活等领域都具有重要的潜在研究价值. 本文提出了 SECF<sup>2</sup>模型:提取 SENet框架中的异构层特征,充分挖掘异构层特征间的聚类典型相关性,生成判别性更强且鲁棒的深层视觉语义,准确刻画多样、易变的材质图像. 在两个材质数据集上验证了 SECF<sup>2</sup>模型的有效性. 综上, SECF<sup>2</sup>模型的主要优点是:它仅需两类特征和一次早期融合即可生成深层视觉语义,故其复杂度大大降低,易于部署至服务器,完成实时材质识别. 此外, SECF<sup>2</sup>是

一个集早期融合、后期融合(集成学习)于一体的新模型,它具有较强的通用性,可运用到其他图像识别(如图像情感分析)任务中,以解决更多实际问题. 未来研究将从以下几个方面开展工作:

- 1)采用 APE-GAN<sup>[39]</sup>完成数据增强,解决材质数据不平衡,为材质图像识别建立更好的数据基础;
- 2)将 SECF<sup>2</sup>模型运用到肿瘤图像识别、新冠肺炎检测、图像场景识别等领域,进一步验证模型的有效性和通用性;
- 3)开发材质识别 APP 并进行推广,努力缩小理

论研究与实际应用间的“鸿沟”。

### 参考文献(References)

- [1] Zheng C X, Chen J Y, Kong J, et al. Scene recognition via semi-supervised multi-feature regression[J]. *IEEE Access*, 2019, 7: 121612-121628.
- [2] Rivas-Lopez M, Sergiyenko O, Flores-Fuentes W, et al. Optoelectronics in machine vision-based theories and applications[M]. Hershey: IGI Global, 2019: 233-254.
- [3] Liu Y, Chen X, Ward R K, et al. Medical image fusion via convolutional sparsity based morphological component analysis[J]. *IEEE Signal Processing Letters*, 2019, 26(3): 485-489.
- [4] Bednarek J, Bednarek M, Kicki P, et al. Robotic touch: Classification of materials for manipulation and walking[C]. *The 2nd IEEE International Conference on Soft Robotics (RoboSoft)*. Korea, 2019: 527-533.
- [5] Bell S, Upchurch P, Snaveley N, et al. Material recognition in the wild with the Materials in Context Database[C]. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Boston, 2015: 3479-3487.
- [6] Schwartz G, Nishino K. Recognizing material properties from images[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 42(8): 1981-1995.
- [7] 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.
- [8] Zhang H B, Jiang Z L, Xiong Q P, et al. Gathering effective information for real-time material recognition[J]. *IEEE Access*, 2020, 8: 159511-159529.
- [9] Haghighat M, Abdel-Mottaleb M, Alhalabi W. Discriminant correlation analysis for feature level fusion with application to multimodal biometrics[C]. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Shanghai, 2016: 1866-1870.
- [10] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, 2018: 7132-7141.
- [11] Rasiwasia N, Mahajan D, Mahadevan V, et al. Cluster canonical correlation analysis[C]. *Proceedings of the 17th International Conference on Artificial Intelligence and Statistics (AISTATS)*. Reykjavik, 2014: 823-831.
- [12] Lai Q, Liu B, Zhao J, et al. BRDF characteristics of different textured fabrics in visible and near-infrared band[J]. *Optics Express*, 2020, 28(3): 3561-3575.
- [13] Chen J, Shan S, He C, et al. WLD: A robust local image descriptor[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, 32(9): 1705-1720.
- [14] Liu C, Sharan L, Adelson E H, et al. Exploring features in a Bayesian framework for material recognition[C]. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. San Francisco, 2010: 239-246.
- [15] Cimpoi M, Maji S, Kokkinos I, et al. Deep filter banks for texture recognition, description, and segmentation[J]. *International Journal of Computer Vision*, 2016, 118(1): 65-94.
- [16] Andrew G. Arora R, Bilmes J, et al. Deep canonical correlation analysis[C]. *Proceedings of International Conference on Machine Learning (ICML)*. Atlanta, 2013: 1247-1255.
- [17] Huang G, Liu Z, van der Maaten L, et al. Densely connected convolutional networks[C]. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Honolulu, 2017: 2261-2269.
- [18] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, 2016: 770-778.
- [19] Chen T Q, Guestrin C. XGBoost: A scalable tree boosting system[C]. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Francisco, 2016: 785-794.
- [20] Prokhorenkova L, Gusev G, Vorobev A, et al. CatBoost: unbiased boosting with categorical features[J/OL]. 2017, arXiv: 1706.09516.
- [21] 张凯, 刘华平, 邓晓燕, 等. 面向室外移动机器人的雷达-图像跨模态检索[J]. *控制理论与应用*, 2018, 35(12): 1759-1764.  
(Zhang K, Liu H P, Deng X Y, et al. Radar-image cross-modal retrieval for outdoor mobile robots[J]. *Control Theory & Applications*, 2018, 35(12): 1759-1764.)
- [22] Kampouris C, Zafeiriou S, Ghosh A, et al. Fine-grained material classification using micro-geometry and reflectance[C]. *Computer Vision—ECCV 2016*. Amsterdam, 2016: 778-792.
- [23] Mahajan D, Girshick R, Ramanathan V, et al. Exploring the limits of weakly supervised pretraining[C]. *Computer Vision—ECCV 2018*. Munich, 2018: 185-201.
- [24] 张红斌, 蒋子良, 熊其鹏, 等. 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[J]. *电子学报*, 2020, 48(4): 790-799.  
(Zhang H B, Jiang Z L, Xiong Q P, et al. Image attribute annotation via a modified effective range based gene selection and cross-modal semantics mining[J]. *Acta Electronica Sinica*, 2020, 48(4): 790-799.)

- [25] Ke G, Meng Q, Finley T, et al. LightGBM: A highly efficient gradient boosting decision tree[C]. Proceedings of Neural Information Processing Systems (NIPS). Foundation, 2017: 3149-3157.
- [26] You Q, Luo J, Jin H, et al. Robust image sentiment analysis using progressively trained and domain transferred deep networks[C]. Proceedings of AAAI Conference on Artificial Intelligence (AAAI). Austin, 2015: 381-388.
- [27] You Q, Luo J, Jin H, et al. Building a large-scale dataset for image emotion recognition: The fine print and the benchmark[C]. Proceedings of AAAI Conference on Artificial Intelligence (AAAI). Phoenix, 2016: 308-314.
- [28] 晁浩, 刘永利, 连卫芳. EEG情感识别中基于集成深度学习模型的多分析域特征融合[J]. 控制与决策, 2020, 35(7): 1674-1680.  
(Chao H, Liu Y L, Lian W F. Multi-analysis domain feature fusion of EEG emotion recognition based on integrated deep learning model[J]. Control and Decision, 2020, 35(7): 1674-1680.)
- [29] Durand T, Mordan T, Thome N, et al. WILDCAT: Weakly supervised learning of deep ConvNets for image classification, pointwise localization and segmentation[C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, 2017: 5957-5966.
- [30] She D Y, Yang J F, Cheng M M, et al. WSCNet: Weakly supervised coupled networks for visual sentiment classification and detection[J]. IEEE Transactions on Multimedia, 2020, 22(5): 1358-1371.
- [31] Sun M, Yang J F, Wang K, et al. Discovering affective regions in deep convolutional neural networks for visual sentiment prediction[C]. IEEE International Conference on Multimedia and Expo (ICME). Seattle, 2016: 1-6.
- [32] Yang J F, She D Y, Sun M. Joint image emotion classification and distribution learning via deep convolutional neural network[C]. Proceedings of the 26th International Joint Conference on Artificial Intelligence. Melbourne, 2017: 3266-3272.
- [33] Zhang H B, Wu J P, Shi H W, et al. Multidimensional extra evidence mining for image sentiment analysis[J]. IEEE Access, 2020, 8: 103619-103634.
- [34] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[C]. Proceedings of International Conference on Learning Representation (ICLR). San Diego, 2015: 1409-1556.
- [35] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.
- [36] Zhou B L, Khosla A, Lapedriza A, et al. Learning deep features for discriminative localization[C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, 2016: 2921-2929.
- [37] Uurtio V, Bhadra S, Rousu J. Large-scale sparse kernel canonical correlation analysis[C]. Proceedings of International Conference on Machine Learning (ICML). Long Beach, 2019: 6383-6391.
- [38] Liu Q Y, Zhang C J, Song Y, et al. Real-time object recognition based on NAO humanoid robot[C]. IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData). Halifax, 2018: 644-650.
- [39] Jin G Q, Shen S W, Zhang D M, et al. APE-GAN: Adversarial perturbation elimination with GAN[C]. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Brighton, 2019: 3842-3846.

### 作者简介

张红斌(1979—), 男, 副教授, 博士, 从事计算机视觉、自然语言处理、推荐系统等研究, E-mail: zhanghongbin@whu.edu.cn;

熊其鹏(1997—), 男, 硕士生, 从事计算机视觉、材质图像识别的研究, E-mail: danicaghost@163.com;

蒋子良(1997—), 男, 硕士生, 从事机器学习、材质图像识别的研究, E-mail: 923886217@qq.com;

石皞炜(1996—), 男, 硕士生, 从事图像情感分析、深度学习的研究, E-mail: 18007001885@163.com;

周娟(1977—), 女, 讲师, 硕士, 从事计算机算法设计与分析、智能计算等研究, E-mail: 422879727@qq.com;

武晋鹏(1992—), 男, 硕士生, 从事机器学习、图像情感分析的研究, E-mail: wujinpeng920421@163.com.

(责任编辑: 李君玲)