基于异构特征聚合的局部视图扭曲型纸币识别

郭玉慧 梁 循
(中国人民大学信息学院 北京 100872)

摘 要 如何识别同一物体的不同结构的表现形式,对于机器而言,是一个比较困难的识别工作.本文以易变形的纸币为例,提出了一种基于异构特征聚合的局部视图扭曲型纸币识别方法.首先利用灰度梯度共生矩阵、Haishoku 算法和圆形 LBP 分别获得纹理风格、色谱风格和纹理,这些特征从不同的角度描述了局部纸币图像,然后通过 VGG-16、ResNet-18 和 DenseNet-121 网络学习这些不变形特征得到输出特征,将输出特征聚合后输入识别层 Softmax,达到三模型融合效果,进而识别局部视图扭曲型纸币.实验结果表明,多特征聚合和不同类型模型融合可以最大可能地捕获图像的语义,在准确率、精度、召回率和 F1 上优于基于单特征和双特征的识别,且优于单类模型和两类模型融合的识别性能,此外,在准确率和时间复杂度等评价标准下,与已有主流方法相比都取得了相对较好的效果.

关键词 纸币识别;局部视图扭曲;不变形特征;特征聚合;模型融合
中图法分类号 TP391 **DOI 号** 10.11897/SP.J.1016.2022.00098

Local View Distorted Banknote Recognition Based on Heterogeneous Feature Aggregation

GUO Yu-Hui LIANG Xun
(*School of Information, Renmin University of China, Beijing* 100872)

Abstract The same object may has different forms of manifestation for different structures in an unrestricted environment, how to recognize such objects is a relatively difficult recognition task for the machine. The banknote is a kind of object that can be easily distorted, as a result, in this paper, taking the distorted banknotes as an example, a local view distorted banknote recognition method based on heterogeneous feature aggregation was proposed. The texture style, color spectrum style and texture of local view distorted banknotes from multiple views were obtained, and these features describe local view distorted banknote images from different perspectives, so as to capture the semantics of local view distorted banknote images as much as possible. As a result, firstly, the gray gradient co-occurrence matrix was used to obtain texture style by carrying out secondary statistical calculation, the Haishoku algorithm was used to obtain color spectrum style, and the circular LBP was used to obtain texture. Then, considering that the multi-view invariant features describe the distorted banknote image in the local view, the VGG-16, ResNet-18 and DenseNet-121 networks were used for each type of feature respectively in the proposed method, and these three types deep models were fused, that is, these three types models did not recognize separately, but learn the invariant feature to get the output feature, which was normalized and aggregated with the other two output features. The VGG-16 network learned invariant feature of texture style to obtain output feature, the ResNet-18 network learned invariant feature of

收稿日期:2019-12-17;在线发布日期:2021-02-10.本课题得到国家自然科学基金(No. 62072463)、国家社会科学基金(No. 18ZDA309)、北京市自然科学基金(No. 4172032)资助.**郭玉慧**,博士研究生,主要研究领域为深度学习和图像处理. E-mail: yhguo@ruc.edu.cn.
梁 循(通信作者),教授,博士生导师,中国计算机学会(CCF)高级会员,主要研究领域为社会计算、神经网络、自然语言处理. E-mail: xliang@ruc.edu.cn.

color spectrum style to obtain output feature, and the DenseNet-121 network learned invariant feature of texture to obtain output feature. These output features were aggregated to obtain aggregated features. After aggregated, aggregated features were input into the recognition layer Softmax to achieve the fusion of three types models, and recognize local view distorted banknote images. We had carried out a lot of experiments to verify the effectiveness of the proposed method through three aspects (i.e. aggregation effect of output features based on invariant features, performance comparison of multiple models fusion, performance comparison with existing methods) and four evaluation indexes (i.e. accuracy, precision, recall and F1). Extensive experiment results showed that the recognition rate of this proposed method was higher and this method could be extended to other visual images, and aggregation of multiple classes features and fusion of different types of models could both capture the semantics of the local view distorted banknote images to the greatest extent, and the proposed method was universal, which could fuse different types of deep networks and apply them. Moreover, multiple classes features aggregation achieved higher recognition than the aggregation based on single and dual features in accuracy, precision, recall and F1, and multiple types models fusion obtained higher recognition than the recognition of the fusion based on single type and two types models in accuracy, precision, recall and F1. In addition, the proposed method had achieved relatively good results compared with the existing state-of-the-art methods under the evaluation criteria of accuracy, time complexity and so on.

Keywords　banknote recognition; local view distortion; invariant features; features aggregation; models fusion

1　引　言

由于深度卷积神经网络的快速发展,图像识别作为其基础的应用之一,已取得了优秀的成果,并且已广泛应用于许多实际场景中,例如银行业务、智能机械等.尽管图像识别算法的性能已得到改善,但其中大多数算法在没有人工配合的情况下,都无法在不受控制的环境中正确处理局部图像.如图 1 所示,在各种纸币存在形式的典型场景中,纸币可能是:被其他人的手部、钱包或文件等物体遮挡;以各种折叠形式存在.而且,在日常生活中,人们的衣食住行都离不开纸币的应用,例如:在无人售票的公交车上,司机在开车的同时需要关注人们的投币.但是行车过程中安全第一,投币基本是在公交车靠站后刚刚驶离站台的时候,存在严重的安全隐患,而且人们投币的时候,可能不会以平整的纸币投入,而是各种折叠等形式的纸币投入,这需要司机更多的关注.生活中还有许多诸如此类的纸币运用场景,没有对纸币的表现形式进行要求,如果能识别这样的纸币,就可以减少人工的工作量,并避免很多不必要的安全隐患.此外,现在大部分关于图像识别的研究从图像相似性进行研究,以局部视图扭曲的纸币为例,引进新的思想来研究视图扭曲而本质不变的物体识别.因此,从纸币的应用以及新的角度来分析和识别局部视图扭曲型纸币具有重要意义.

正如一般对象识别一样,局部视图扭曲的纸币识别的关键是提取可区分的特征.然而,作为特殊的物体识别任务,由于以下原因,局部视图扭曲的纸币识别尚未得到充分解决.首先,与一般物体识别不同,如图 1 所示,局部视图扭曲的纸币没有表现出独特的空间布局和结构.它们通常是非刚性的,不容易利用结构信息.因此,标准的对象识别方法可能在此任务上执行不佳;其次,局部纸币识别可以被认为是细粒度的识别.细粒度的对象识别通常是发现某些对象的固定语义部分.但是,常见的语义部分在许多局部视图扭曲的纸币图像中不存在.因此,很难通过现有的细粒度方法直接从扭曲型纸币图像中捕获语义信息进行识别;第三,图 1 所示的局部视图扭曲型纸币图像,要求纸币识别方法应具有几何不变性,能够可靠地识别这类纸币图像.现有的纸币识别方法通常使用卷积神经网络(Convolutional Neural Network,CNN)直接从整张纸币图像中提取视觉特征,并且当几何变化较大时可能会识别错误,因为 CNN

只能通过最大合并来处理具有小范围扭曲的图像.第四,在纸币识别方面虽已取得了进展,但对于诸如对象识别和场景识别之类的研究人员却没有给予足够的重视.没有像 ImageNet[1] 和 PlacesNet [2] 这样训练有素的模型,能够有利于推动其在计算机视觉社区的发展.

此外,局部视图扭曲的纸币图像通常不会表现出明显的空间排列,但我们可以探索不同尺度的不变形特征,提取这些特征进行学习并聚合,聚合后的特征包含来自区分图像区域的信息,而且可以对几何变形更加鲁棒.诸如[3]和[4]之类的一些工作已经验证了多尺度特征在分类和检索任务中的有效性.

考虑到这些因素,在本文中,我们提出了一种基于异构特征聚合的局部视图扭曲型纸币识别(Local View Distorted Banknote Recognition based on Heterogeneous Feature Aggregation, HFA-LVDBR)方法,其中局部视图扭曲意味着在人类视觉中表现出来的不规则性.由于局部视图扭曲的纸币图像通常不会表现出明显的空间排列方式,因此,通过获取不变形特征,对每种不变形特征进行不同粒度的分析,使最终的特征表示对几何变形更加鲁棒、稳定不变,能够更全面的反映局部视图扭曲型纸币图像的语义信息.特别地,HFA-LVDBR 中融合了不同类型的网络,即目前基本的流行的 CNN 体系结构 VGG[5]、ResNet[6] 和 DenseNet[7],将这些不变形特征经过融合网络学习,聚合为一个统一的表示形式,得到的聚合特征从不同的粒度描述了局部纸币图像,可以最大可能地捕获语义信息,从而精确地识别局部视图扭曲型纸币.

图 1 局部视图扭曲的纸币

本文的贡献可以总结如下:

(1)我们提出了 HFA-LVDBR 方法,该方法可以提取三种不变形的视觉特征,针对每种不变形特征,进行了细粒度分析和获取,从而产生更强大、更具区别性和全面的细粒度表示.

(2)HFA-LVDBR 不仅利用深度网络进行特征学习,而且将学习到的输出特征聚合为一个统一的表现形式.

(3)HFA-LVDBR 根据多尺度不变形特征进行了深度模型的融合,实现了对局部视图扭曲型纸币的识别.此外,HFA-LVDBR 在局部视图扭曲的图像识别方面的思想,可以扩展到其他计算机视觉问题上.

本文在第 2 节中,对与纸币识别相关的主要算法和模型进行了对比阐述;第 3 节介绍了 HFA-LVDBR 的实现细节;第 4 节为实验设置与结果分析内容,主要分析了特征聚合和模型融合在准确率、精度、召回率和 F1 方面的性能,以及与其它方法的性能比较;最后,在第 5 节中给出文章的主要结论以及未来的可能研究方向.

2 相关工作

我们的工作与两个研究领域密切相关:(1)纸币的识别,(2)面向局部视图的识别.

2.1 纸币的识别工作

2.1.1 整体纸币的识别工作

在图像处理和计算机视觉中,纸币识别的研究已取得一些成果.通过人工提取特征的纸币识别方法,其在识别污损严重的纸币图像时,准确率会降低.为了避免这些方法存在的问题,刘艳萍等人[8]将纸币图像划分为多个区域块,并计算平均值作为图像特征,输入训练好的 BP 神经网络,从而实现对纸币的识别.柳春华等人[9]提出基于统计特征算法解决了旧币分离问题,其利用数据集建立了学习向量,构建了神经网络模型,从而实现对纸币的识别.Tuyen 等人[10]提出了一种基于识别区域选择的钞票识别方法.该方法利用遗传算法进行掩模优化,然后利用主成分分析和基于最小欧氏距离的 K-means 质心匹配,对输入的纸币图像进行分类.Dittimi 等人[11]提出了基于梯度特征的多类支持向量机的钞票识别方法,该方法提取了纸币的梯度特征直方图,并在多类支持向量机中,利用交叉验证法实现对纸币的分类.随着 CNN 的发展,Liu 等人[12]提出了一种基于卷积神经网络的合格钞票识别方法.该方法从网络层和卷积核大小优化 Letnet-5 模型结构,确定了最佳的结构和性能参数,最后通过训练好的 Letnet-5 模型来识别合格和不合格的纸币.因为 CNN 对图像的视图扭曲、旋转等不敏感,而且对输入向量

要求具有确定的大小,所以它不能直接处理具有任意大小/比例的纸币图像.因此,盖杉等人[13]改进多层次空间金字塔算法,实现对不同尺寸的纸币特征的相同维度的输出表示,并经过 Softmax 层识别纸币图像.以上工作从不同角度实现了整体纸币的识别,没有针对局部纸币进行识别研究,但是目前针对纸币局部区域的识别工作也取得了一些显著的成果.

2.1.2　局部纸币的识别工作

在局部纸币识别的研究中,通过不同研究角度也已取得了一些成果.宋普庚等人[14]对纸币图像的深色斑块等进行了预处理,并提出新的算法来设计和训练神经网络,实现对纸币冠字号的识别.赵敏[15]通过对字符进行投影法分割,提出改进的 KNN 算法,对分割后的百元人民币冠字符号进行识别.Jang 等人[16]提出了一种基于卷积神经网络的印度卢比纸币序列号识别方法.首先采用二值化和仿射变换对钞票图像进行去歪斜,并基于钞票的纵横比提取序列号 ROI.然后,从序列号 ROI 中提取单个字符图像,利用 CNN 网络对字符进行识别.Tsai 等人[17]提出了一种基于规则的光学字符识别系统,用于识别人民币纸币上的序列号.根据英文字母和数字这两类字符的特征,设计了一些基于观察的规则来实现分类器,在人民币数据集上进行了分类测试.以上工作大部分是针对纸币的序列号进行了识别,没有针对如图 1 所示的局部视图扭曲型纸币进行识别,也没有从纸币的纹理等视觉不变形特征进行研究.

2.2　面向局部视图的识别

在图像识别领域,已有很多面向局部视图的图像识别方法.例如:为了提高 CNN 激活的不变性而不降低其判别能力,Gong 等人[3]提出了一种多尺度无序合并方案.该方案为多个规模级别的局部图像块提取 CNN 激活,分别在每个级别上对这些激活执行无序合并,并用于从图像分类到实例级检索的识别任务.Wu 等人[4]针对特定类别对象,通过局部 CNN 微调增强的判别聚类,将相似的对象聚合为元对象.并且,在多个空间尺度上,合并所有学习的元对象的特征响应图来构造场景图像表示.Song 等人[18]利用神经网络的判别性图像块表示,将语义流形建立在多尺度 CNN 上,并使用马尔可夫随机场制定了丰富的上下文模型.Herranz 等人[19]分析了多尺度 CNN 架构,并表明 ImageNet-CNN 和 Places-CNN 的多尺度组合可以提高场景识别性能.但是,CNN 缺少几何不变性,这限制了它们对高度可变图像的分类和匹配的鲁棒性.在本文工作中,将提取共有的不变形特征,通过分类模型学习并聚合,用于扭曲型纸币识别的任务,以处理几何变形.另外,我们的方法也与细粒度的图像识别相关.Peng 等人[20]提出了一种基于对象局部注意力模型的细粒度图像识别,通过对象层次的注意力定位图像的对象,以及区域级注意力选择对象的可区分部分,两者共同学习多视图和多尺度特征,并利用细微和局部差异来区分子类别.Branson 等人[21]提出了一种用于细粒度视觉分类的模型,首先将对对象姿态的估计值用于计算局部图像的特征,然后将这些特征用于分类.Fu 等人[22]提出一种改进的循环注意力卷积神经网络,该模型通过相互增强的方式学习基于区域的特征表示,并递归地学习判别性区域注意力,用于细粒度图像识别.Zhang 等人[23]通过采用区域关注机制,从原始图像生成关注区域来减轻背景影响,并在对象的序列化特征上,应用映射函数来隐式地学习序列表示,将区域参与网络和序列学习网络组合成一个统一的框架,利用该框架进行细粒度的图像识别.以上方法均从图像的特有区域、实体对象的独特结构和空间布局提取特征,然后将这些特征用于图像识别.然而,对未有独特的空间布局或者结构的扭曲型纸币来说,一些特有的语义部分是不存在的.因此,本文通过提取扭曲型纸币的共有特征,解决未表现出明显空间排列方式的图像识别问题.

3　局部视图扭曲的纸币识别

如图 1 所示,与一般图像不同,局部视图扭曲型纸币图像具有更复杂的表现形式,通常不会表现出独特的空间结构.因此,我们提出了 HFA-LVDBR 方法,对局部视图扭曲型纸币进行识别.首先,获取扭曲型纸币图像的不变形特征,即纹理风格、色谱风格、纹理.不同类型的特征从不同的粒度描述了局部视图扭曲型纸币,可以最大可能地捕获扭曲型纸币图像的语义信息.其次,对于每种不变形的特征,利用不同深度网络模型进行学习,将学习到的输出特征聚合,能够获得更鲁棒和有区别的表示,而且对几何变形不敏感.然后,将聚合特征输入不同网络模型融合的识别层 Softmax,实现对扭曲型纸币图像的识别.

如图 2 所示,给定输入图像,HFA-LVDBR 能够提取具有不同比例和不同粒度的不变形特征:纹理风格、色谱风格和纹理.为了提取这些特征,HFA-

LVDBR 中引入了灰度梯度共生矩阵、Haishoku 算法和圆形 LBP 三种方法. 通过灰度梯度共生矩阵，提取具有灰度和梯度综合信息排列的纹理风格；通过 Haishoku 算法获取纸币图像的色谱风格；通过圆形 LBP 获取图像的纹理. 这些特征均可以在细粒度和本地级别描述局部视图扭曲的纸币图像. 因此，与全局语义分布相比，这三种方法可以提取面向区域的不变形特征. 这些不变形特征利用不同类型的模型进行学习，比只用一类模型学习效果更好，所以在 HFA-LVDBR 中引入了三种基本的流行的深度神经网络，即 VGG-16、ResNet-18 和 DenseNet-121，将每一类特征分别输入对应的神经网络模型进行学习，然后将三种网络模型学习到的输出特征进行规范化，并将这些特征聚合到最终表示中. 最后，将聚合特征输入多模型融合的识别层 Softmax，进行了扭曲型纸币识别的训练和测试. 在以下各节中，我们将详细介绍 HFA-LVDBR 的主要部分.

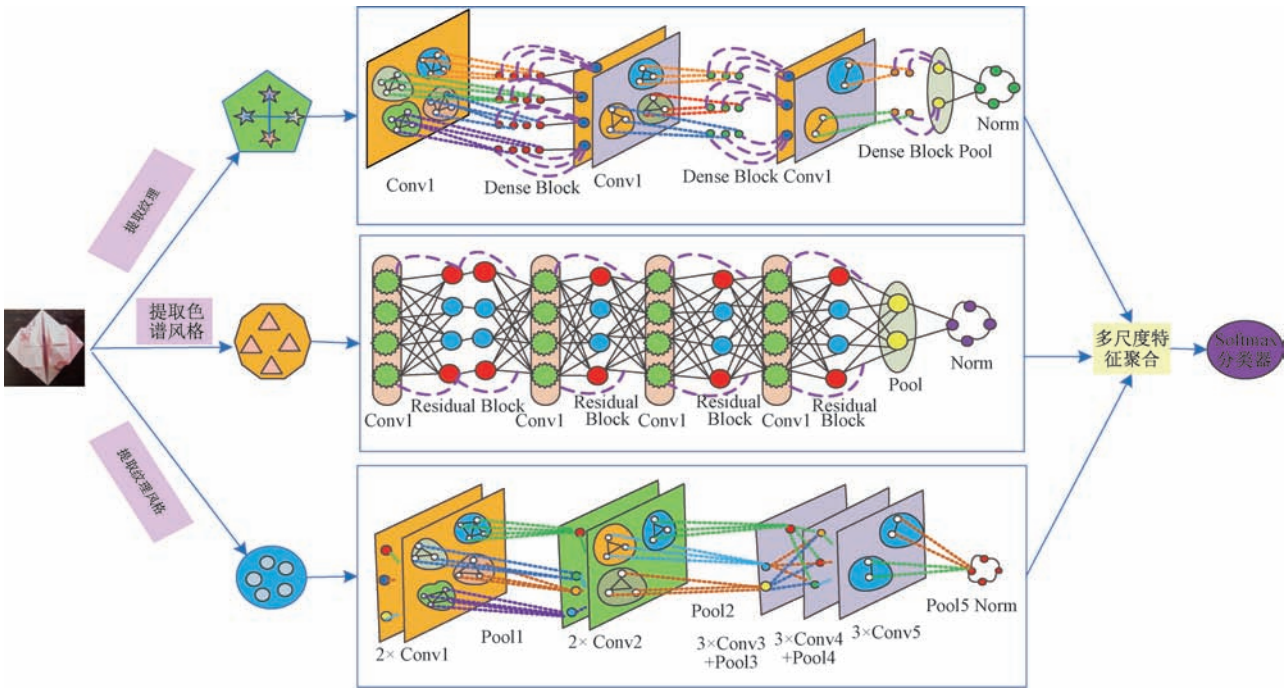

图 2 基于不变形特征的多模型融合框架

3.1 局部纸币风格的预处理

风格能够体现事物的本质特性. 纸币作为货币流通手段的价值符号，它也有不同于其他物体的独特风格，本节将从纸币的纹理和颜色这两种风格来表示纸币的不变形特征.

3.1.1 纹理风格的表示

纹理风格体现了图像表面的具有循环变化的灰度与梯度的排列特性.

图像局部纹理信息的循环变化既能够反映图像整体的纹理信息，也体现了图像或图像局部的纹理性质. 与色谱特征不同，纹理特征需要通过图像局部区域的多个像素点进行统计计算，而不只是依赖于像素点特征. 因此，在局部视图扭曲的图像识别中，纹理特征能够保持不变性. 在本节中，通过将图像的梯度信息加入到灰度共生矩阵形成灰度梯度共生矩阵，使其能够包含图像的纹理特征及其分布信息，从而利用灰度和梯度的综合信息对局部纸币图像的纹理风格进行表示.

基于归一化的灰度梯度共生矩阵，计算一系列的二次统计表示纹理风格，这样表示的原因是：(1)不仅反映了灰度之间的关系，还反映了梯度之间的关系. 灰度是构建图像色谱的基准，梯度是构建图像纹理的元素. (2)体现了图像的灰度与梯度的分布信息，反映了纹理像点与其周围区域像点的分布关系，而且梯度能够反映具有方向性的纹理，从而能够更好地描绘图像的纹理风格.

在灰度梯度共生矩阵中，其元素 $M(x,y)$ 的 M 表示元素数目，x 表示灰度值，y 表示梯度值. 对灰度梯度共生矩阵的元素进行归一化，其各元素之和为 1，变换公式如下：

$$M'(x,y)=\frac{M(x,y)}{\sum_{x=0}^{L_f-1}\sum_{y=0}^{L_g-1}M(x,y)} \tag{1}$$

如表 1 所示,其中不同的参数值表示计算灰度梯度共生矩阵的不同数字特征,代入公式(2)计算梯度优势和分布不均匀性.

$$T_1=\frac{\sum_2[\sum_1 M'(x,y)Z^m]^n}{\sum_{x=0}^{L_f-1}\sum_{y=0}^{L_g-1}M'(x,y)} \tag{2}$$

表 1　公式(2)参数表

参数	$\sum_1$	$\sum_2$	Z	n	m
小梯度优势	$\sum_{y=0}^{L_g-1}$	$\sum_{x=0}^{L_f-1}$	$\frac{1}{(y+1)^2}$	1	1
大梯度优势	$\sum_{y=0}^{L_g-1}$	$\sum_{x=0}^{L_f-1}$	y^2	1	1
灰度分布不均匀性	$\sum_{y=0}^{L_g-1}$	$\sum_{x=0}^{L_f-1}$	1	2	0
梯度分布不均匀性	$\sum_{x=0}^{L_f-1}$	$\sum_{y=0}^{L_g-1}$	1	2	0

选择它们表示局部纸币的纹理风格的原因如下:

(1)梯度能够反映图像灰度的变化情况,而梯度的大小表示图像等灰度线的密集程度,其中,小梯度优势和大梯度优势从不同角度表现了图像的灰度变化情况.小梯度优势大表示图像的灰度变化较为平缓;大梯度优势大表示图像的灰度变化较为剧烈.

(2)分布不均匀性包括灰度和梯度两方面的不均匀性,其中,灰度不均匀性反映图像像素的灰度分布情况,梯度不均匀性反映图像纹理的密集程度.

如表 2 所示,其中不同的参数值表示计算灰度梯度共生矩阵中其他的数字特征,代入公式(3)计算能量、灰度与梯度的平均性、均方差、相关性、熵值惯性和逆差距.

$$T_3=l\cdot\{\sum\nolimits_2 L^a[\sum\nolimits_1(M'(x,y))^nA]\}^b \tag{3}$$

选择它们作为其它表示局部纸币的纹理风格的原因如下:

(1)能量值能够反映图像局部的灰度分布情况和纹理粗细程度,其值越大表明图像局部的灰度均匀分布且纹理较细.

(2)灰度平均性反映图像整体的平均灰度水平;梯度平均性反映图像整体的平均梯度水平.

(3)灰度均方差反映图像灰度的离散程度;梯度均方差反映图像纹理分布的离散程度.

(4)自相关反映图像纹理灰度的相似程度,其值大小反映了图像局部区域的灰度相关性.

(5)灰度熵对图像灰度信息进行度量,反映图像灰度的复杂情况;梯度熵对图像梯度信息进行度量,反映图像梯度的复杂情况;混合熵反映图像整体灰度和梯度的复杂情况.

(6)惯性反映图像纹理密集程度之间的不变性.

(7)逆差距反映图像的纹理清晰度和分布信息,因此,其值越大表示纹理较清晰且具有规律性.

表 2　公式(3)参数表

参数	l	$\sum_1$	$\sum_2$	L	a	n	A	b
能量	1	$\sum_{y=0}^{L_g-1}$	$\sum_{x=0}^{L_f-1}$	1	0	2	1	1
平均灰度水平	1	$\sum_{y=0}^{L_g-1}$	$\sum_{x=0}^{L_f-1}$	x	1	1	1	1
平均梯度水平	1	$\sum_{x=0}^{L_f-1}$	$\sum_{y=0}^{L_g-1}$	y	1	1	1	1
灰度均方差	1	$\sum_{y=0}^{L_g-1}$	$\sum_{x=0}^{L_f-1}$	$x-T_{\text{灰度平均}}$	2	1	1	1/2
梯度均方差	1	$\sum_{x=0}^{L_f-1}$	$\sum_{y=0}^{L_g-1}$	$y-T_{\text{梯度平均}}$	2	1	1	1/2
自相关	1	$\sum_{y=0}^{L_g-1}$	$\sum_{x=0}^{L_f-1}$	1	0	1	$(x-T_{\text{灰度平均}})^2(y-T_{\text{梯度平均}})$	1
灰度熵	−1	$\sum_{y=0}^{L_g-1}$	$\sum_{x=0}^{L_f-1}$	1	0	1	$\log\sum_{y=0}^{L_g-1}M'(x,y)$	1

续表

参数	l	Σ_1	Σ_2	L	a	n	A	b
梯度熵	-1	$\sum_{x=0}^{L_f-1}$	$\sum_{y=0}^{L_g-1}$	1	0	1	$\log\sum_{x=0}^{L_f-1}M'(x,y)$	1
混合熵	-1	$\sum_{x=0}^{L_f-1}$	$\sum_{y=0}^{L_g-1}$	1	0	1	$\log M'(x,y)$	1
惯性	1	$\sum_{y=0}^{L_g-1}$	$\sum_{x=0}^{L_f-1}$	1	0	1	$(x,y)^2$	1
逆差距	1	$\sum_{x=0}^{L_f-1}$	$\sum_{y=0}^{L_g-1}$	1	0	1	$\frac{1}{1+(x-y)^2}$	1

通过公式(2)和(3)得到纹理风格的表示,其表示矩阵如下所示,其中 $GA_i(i=1,2)$分别表示小梯度优势和大梯度优势;$IOGD_i(i=1,2)$分别表示灰度分布不均匀性和梯度分布不均匀性;EG 表示能量;$GM_i(i=1,2)$分别表示灰度平均性和梯度平均性;$GMSD_i(i=1,2)$分别表示灰度均方差和梯度均方差;ACL 表示自相关;$GE_i(i=1,2,3)$分别表示灰度熵、梯度熵和混合熵;$TG_i(i=1,2)$分别表示惯性和逆差距.

$$\begin{bmatrix} GA_1 & GA_2 & IOGD_1 & IOGD_2 & EG \\ GM_1 & GM_2 & GMSD_1 & GMSD_2 & ACL \\ GE_1 & GE_2 & GE_3 & TG_1 & TG_2 \end{bmatrix}$$

3.1.2 色谱风格的表示

图像都具有其独特的色谱风格,本节通过 Haishoku 算法获取局部纸币图像的色谱,如图 3 所示,是不同局部纸币的色谱图,在每张色谱图中,每种颜色显示的色谱长短,表示该颜色在色谱中所占的比重,而同一种纸币的局部色谱是相近的,如图 4 所示,是同一张 100 元的两个局部色谱,因此,通过局部纸币的色谱能够反映整体纸币的色谱.

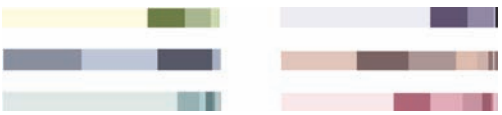

图 3 不同局部纸币的色谱图

图 4 同一种纸币的局部色谱

算法 1 Haishoku 算法

输入:图像的相对路径 image_path

输出:色谱图 palette

1. 设置颜色种类列表 palette_tmp []
2. 获得颜色 RGB 值 *colors*
3. 将 RGB 值 *colors* 加入 palette_tmp[]
4. 设置色谱列表 palette []
5. FOR *colors* ∈ palette_tmp:
6. 计算每一种颜色在色谱中的比重 p
7. 将比重 p 与相应的颜色 RGB 值的标量 c 表示为一个数组元素 pc
8. 将每一个元素 pc 放入 palette 列表
9. 设置色谱图列表 images []
10. FOR P_i ∈ palette:
11. 获得元素 P_i 的第一个值 w 和第二个值 r
12. 设置三元组('RGB',w,r)表示色谱中的某种颜色的信息框 color_box
13. 将所有 color_box 依次加入 images[]
14. 将 images 中的元素转化成色谱图并保存

3.2 局部纸币图像的纹理提取

局部二值模式(LBP)[24]是在图像的局部区域内,以中心像素的灰度值作为基准,通过与相邻区域相比,将获得的二进制编码表示局部纹理.原始的 LBP 通过特定区域的灰度值表示,当图像的局部视图发生扭曲时,原始的 LBP 编码将不能准确反映图像局部的纹理信息.为了处理不同视图的纹理特征,并能够保持旋转不变性,本节将原始 LBP 的正方形邻域变为圆形邻域,能够使任意多个像素点落在圆形 LBP 的区域内,且通过线性内插得到相邻区域的对角线像素值,从而保证所有对角线像素点获得整数坐标.如图 5 所示,图像旋转后 LBP 值会发生改变,在圆形邻域内,相对于中心等距分布的 P 个点的坐标可以表示为

$$(x_p,y_p)=(x_c+R\cos(2\pi p/P),y_c-R\sin(2\pi p/P)) \tag{4}$$

其中(x_c,y_c)为中心点坐标,(x_p,y_p)为圆形区域内第 p 个点坐标($p=0,1,\cdots,P-1$).

如图 6 所示,是 1 元人民币纸币的局部图像,根据圆形 LBP 对其进行纹理提取,如图 7、8 所示,采

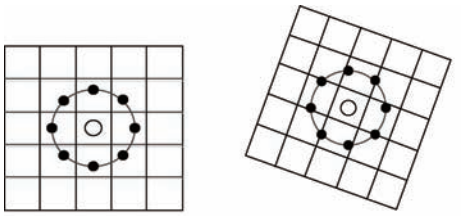

图 5　圆形 LBP 采样图

图 6　原始局部图像

样点数 P 和半径 R 不同,获得的局部纸币的纹理清晰度也不同.

(1)当采样点数 $P=8$ 时,设置半径 R 分别为 1,2,3,对原始局部图像提取纹理. 从图 7(a)、(b)、(c)可以看出,当 $R=1$ 时,提取的纹理清晰度和规则度最好.

(2)当半径 $R=1$ 时,设置采样点数 P 分别为 4,8,12,对原始局部图像提取纹理. 从图 8(a)、(b)、(c)可以看出,当采样点数 $P=8$ 时,提取的纹理清晰度最好.

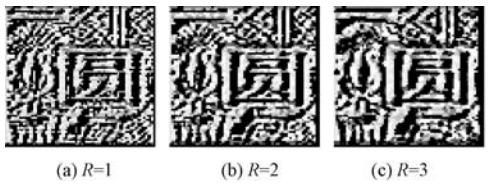

图 7　不同半径的纹理图像

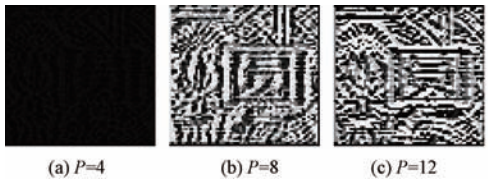

图 8　不同采样点数的纹理图像

因此,采用圆形 LBP 的圆形区域半径 $R=1$,采样点 $P=8$ 时提取的纹理最清晰.

3.3　深度学习模型的动态融合

3.3.1　改进的 VGG-16 模型结构

为了对局部纸币的纹理风格进行学习,并将学习到的特征与其他模型学习到的特征进行融合,本节将搭建改进的 VGG-16 模型,该模型结构如图 9 所示,主要部分的具体描述如下:

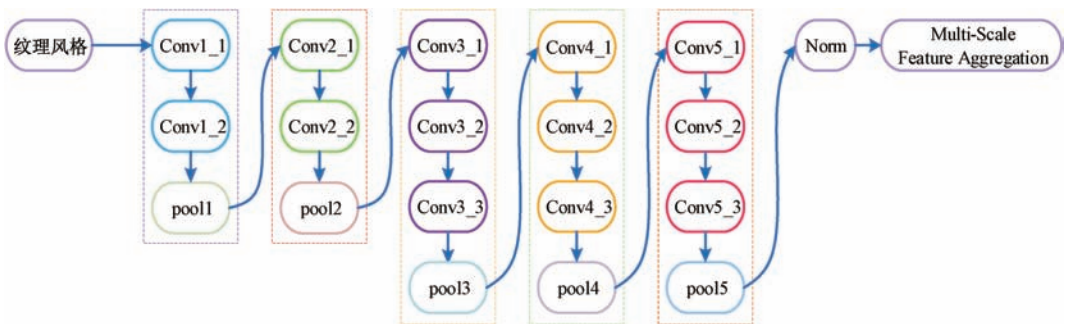

图 9　改进的 VGG-16 模型结构图

VGG-16 模型由 5 段卷积组成,其中,前 2 段含有 2 个卷积层,后 3 段含有 3 个卷积层,且在每段之后采用最大池化层来缩减图片尺寸. 为了与其它两个深度模型融合,在第 5 段卷积的池化层之后增加了规范化层 Norm 和特征聚合层 Multi-Scale Feature Aggregation.

第 5 段卷积的最大池化层的输出特征是改进的 VGG-16 模型的学习特征,将其经过 Norm 层进行规范化,并通过特征聚合层与其它两个深度模型的输出特征聚合,即改进的 VGG-16 模型用于学习局部纸币的纹理风格,并与其它两个深度模型在识别层进行融合.

3.3.2　改进的 ResNet-18 模型结构

为了对局部纸币的色谱风格进行学习,并将学习到的特征与其他模型学习到的特征进行聚合,本节将搭建改进的 ResNet-18 模型,该模型结构如图 10 所示,主要部分的具体描述如下:

在 ResNet-18 中,由两层卷积和一个跨层连接构建残差映射,且在相同维度的特征之间进行一样的特征映射,因此,可以将残差映射的输入和输出直接进行相加. 当网络延伸到特征维度不同时,利用 2 倍下采样将特征映射的维度减半,此时卷积核的数量增加一倍,因此,为了映射到与输出相同的维度,采用 1×1 大小的卷积操作. 残差映射如下式所示:

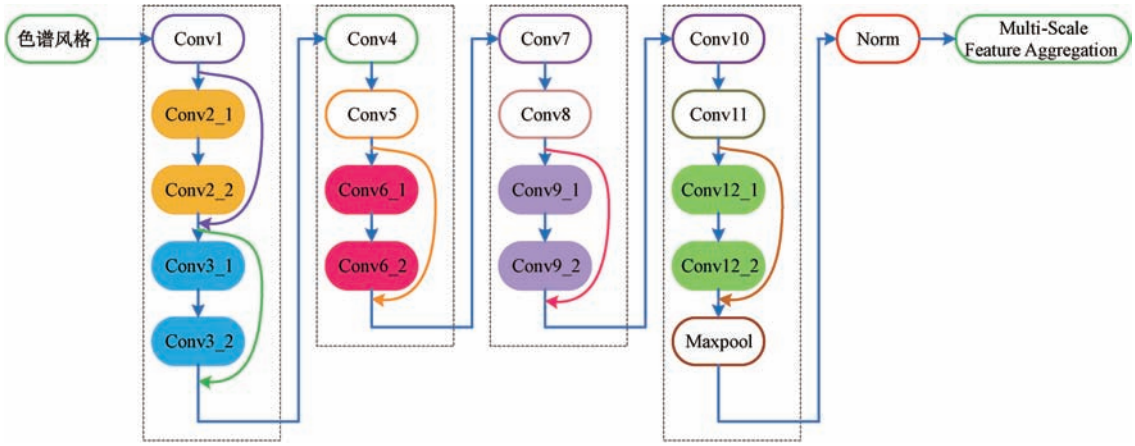

图 10 改进的 ResNet-18 模型结构图

$$y = F(x,\{W_j\}) \mid x \tag{5}$$

$$y = F(x,\{W_j\}) + G(x,\{W_s\}) \tag{6}$$

式中，$F(x,\{W_j\})$表示映射函数，x 表示映射输入，y 表示映射输出. 在式(5)中，表示相同维数 x 和 F 之间的映射，其跨层连接不会增加计算复杂度；在式(6)中，表示不同维数 x 和 F 之间的映射，为了跨层连接之后的维数匹配，采用了 1×1 卷积映射 $G(x,\{Ws\})$.

在最后一个残差块后连接最大池化层，并增加了规范化层 Norm 和特征聚合层 Multi-Scale Feature Aggregation，通过规范化层 Norm 对最大池化层的输出特征进行规范化，并通过特征聚合层与其它两个深度模型的输出特征聚合，即改进的 ResNet-18 模型用于学习局部纸币的色谱风格，并与其它两个深度模型在识别层进行融合.

3.3.3 改进的 DenseNet-121 模型结构

为了对局部纸币的纹理进行学习，并将学习到的特征与其他模型学习到的特征进行聚合，本节将搭建改进的 DenseNet-121 模型，该模型结构如图 11 所示，主要部分的具体描述如下：

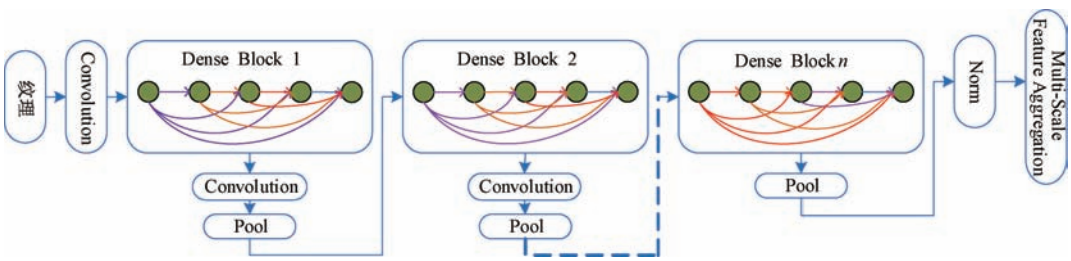

图 11 改进的 DenseNet-121 模型结构图

DenseNet-121 模型是由稠密连接块组成的深度卷积神经网络. 在该模型中，任意两层之间进行特征输入，且网络被划分为多个稠密连接的稠密块，即将每层的输出特征作为其后所有层的输入，使网络层之间越来越紧密，从而避免梯度消失的问题，其输出特征如式(7)所示：

$$F_l = H_l([F_0,F_1,\ldots,F_{l-1}]) \tag{7}$$

式中，F_l 表示第 l 层的输出，即其前面的所有层特征 $F_0,\ldots,F_{l-1}$ 作为输入，$[F_0,F_1,\ldots,F_{l-1}]$表示级联第 0 层到第 $l-1$ 层的输出特征，且每层都通过非线性的复合函数 $H_l(\cdot)$进行转换，即 $H_l(\cdot)$由批量归一化、整流线性单元及卷积等组成.

深度卷积网络的层输出维度，会随着网络层数的增加而快速增长，因此，DenseNet-121 通过在每个稠密模块后采用池化均衡，如图 11 所示，在池化操作前，利用卷积操作使特征维度变为输入维度的一半. 因此，不同于只依赖于前一层输出特征的深度网络，DenseNet-121 可以更好地利用低维特征.

在最后一个稠密块连接最大池化层，并增加了规范化层 Norm 和特征聚合层 Multi-Scale Feature Aggregation，通过规范化层 Norm 对最大池化层的输出特征进行规范化，并通过特征聚合层与其它两个深度模型的输出特征聚合，即改进的 DenseNet-121 模型用于学习局部纸币的纹理，并与其它两个深度模型在识别层进行融合.

3.3.4　深度模型的融合与识别

靠近输出层的深度神经网络层包含独立的类相关信息,这些信息不包含在输出层中[25].因此,对于不同的深度神经网络,均可以在靠近输出层的网络层提取深层的视觉特征 $\hat{h}=(\hat{h}^1,\dots,\hat{h}^d,\dots,\hat{h}^D)$,其中 D 表示特征维度的数量.

从图 9、10、11 可以看到,对于每一类特征,采用了不同的深度卷积网络模型进行学习,并且在网络模型得到深层视觉特征后,由于特征的维度不一样,采用最大池化操作后,对小维度特征进行上采样(比如最近邻插值等)操作,达到一致的维度.由于不同类型的特征元素具有不同的值范围,通过 $Norm(\cdot)$ 规范化每一种特征的元素值,然后进行聚合操作 $Agg(\cdot)$ 获得聚合特征表示:

$$F=Agg(Norm(\hat{V}),Norm(\hat{R}),Norm(\hat{D})) \quad (8)$$

其中,$Norm(\cdot)$ 表示规范化操作,例如 l_2 或 z-score. $Agg(\cdot)$ 表示简单的聚合操作,例如简单串联[21]和深层前馈网络[26].不失一般性,在本文中,采用 z-score 规范化操作和简单级联的聚合方法.

模型融合层 Softmax 能够获取输入样本的概率分布 Y,其覆盖所有可能的输出类别,从而预测最可能的结果,因此,将聚合特征输入融合层 Softmax 进行训练.

在测试阶段,输入测试的局部纸币图像,首先获得不变形特征:纹理风格、色谱风格、纹理,然后利用三类深度模型对其进行学习,将不同深度模型的学习特征聚合为最终特征.最后,将聚合特征输入模型融合层 Softmax 分类器以识别局部视图扭曲型纸币.

HFA-LVDBR 的优势从以上内容可以看出:首先,HFA-LVDBR 可以在无监督的信号下获得不同类型的不变形特征,从不同的视图和粒度来看,它们是互补的,且对几何变形不敏感.其次,HFA-LVDBR 可以将这些不变形特征通过深度模型得到深层视觉特征,并聚合为最终特征表示,最大可能地包含局部视图扭曲型纸币的判别信息.因此,HFA-LVDBR 方法即基于异构特征聚合的局部视图扭曲型纸币识别适用于图 1 所示的纸币图像.

4　实验分析

4.1　数据集的预处理

为了评估 HFA-LVDBR 的性能,在配置为(CPU:AMD Ryzen 7 3700X 8－Core 3.60GHz,内存:32.00GB,GPU:11GB,Windows 版本:Windows10 专业版)的计算机上进行了实验.在实验中,选择人民币图像与欧元图像构建局部纸币识别的实验数据集,不同币种与面值的纸币图像如图 12 所示,显示了 12 种面值的人民币和欧元纸币的正面图像,左边是人民币,与右边颜色相似的欧元形成对比.

图 12　欧元和人民币面值

根据纸币图像的尺寸,对其正反面裁剪成尺寸大小相同的局部纸币图像,如图 13 所示,获得不同币种与面值的局部纸币图像 768 张,然后从纹理风格、色谱风格、纹理这些特征获取数据集,数据集的细节描述如下:

图 13　部分纸币的局部图像

(1)纹理风格的数据集预处理

局部视图扭曲型纸币在光照强度不一样的环境中,图像的亮度也不一样.因此,为了获取局部纸币的纹理风格,在调节纸币图像亮度的同时,根据式

(1)—(3),共获取纹理风格表示矩阵 18432 个,其中不同币种不同面值均为 1536 个,且每个矩阵包含 15 个因素,本实验中将此数据集存在 CSV 文件中,便于改进的模型 VGG-16 进行学习.

(2)色谱风格的数据集预处理

为了获取纸币图像的色谱风格,与纹理风格处理相同,在调节纸币图像亮度的同时,根据 3.1.2 节的方法,共获得色谱风格图像 18432 张,其中不同币种不同面值均为 1536 张,且每一张色谱图像都显示了不同颜色的比例,反映了纸币图像的色谱风格.

(3)纹理图像的数据集预处理

由于光照强度不会影响纹理的清晰度和粗细程度,因此,在获取纸币图像的纹理时,不进行亮度调节,而是在局部纸币图像转为灰度图的基础上,如图 14 所示,以 30 度角为单位旋转图像,达到更全面反映现实场景中纸币的表现形态,共获得纹理图像 18432 张,其中不同币种不同面值均为 1536 张.

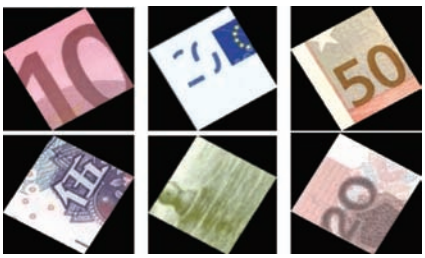

图 14 不同角度的局部纸币图像

通过以上步骤构建了纹理风格、色谱风格与纹理这三类特征的训练数据集.

为了验证本文所提方法的有效性,从百度图库收集如图 1 所示的图像 1200 张,不同币种的每一种面值的纸币数量均为 100 张,进行如训练数据集的预处理得到测试数据集:色谱风格图像 12000 张、纹理风格表示矩阵 12000 个、纹理图像 12000 张.

4.2 实验结果与比较

VGG-16,ResNet-18 和 DenseNet-121 是当前基本的 CNN 体系结构. 为了验证所提方法的有效性和鲁棒性,本实验运用这三种基本模型进行融合,每个模型的预训练参数如表 3 所示. 对于特征的聚合操作 $Agg(\cdot)$,虽然可选择其他的特征聚合方法,但在本文工作中,主要体现提出的 HFA-LVDBR 框架,解决未有明显空间排列方式的纸币识别问题,因此,只采用了简单的级联操作. 考虑到不同模型的学习特征具有不同的维度,对小维度特征进行最近邻插值的操作,由于特征元素在不同的值范围内,对每种特征的元素进行 z-score 规范化,然后聚合这些特征,输入融合层 Softmax 进行训练. 在训练过程中,通过 L2 正则化使权重衰减到更小的值,防止模型过拟合,并从召回率、精确度、准确率和 F1 四方面来验证不变形特征的聚合效果以及评估深度模型的融合性能.

表 3 模型的预训练参数

模型	learning_rate	batch_size	input_size	weight_decay
VGG-16	0.0001	16	224×224	0.0004
ResNet-18	0.01	16	56×56	0.0005
DenseNet-121	0.01	32	56×56	0.001

4.2.1 不变形特征的聚合效果

在本小节中,将图像的纹理风格、色谱风格、纹理这些特征聚合后,对纸币图像的识别进行比较和分析. 从图 15 到图 17 显示了不同特征组合的实验结果,可以看到:两种或三种类型的特征聚合后,纸币图像的识别在准确率、精度、召回率和 $F1$ 上高于基于单一类型特征的识别.

(1)基于单特征的识别

通过 VGG-16 模型对纹理风格进行识别,如图 15(a)所示,随着迭代轮数的增加,其测试的准确率、精度、召回率和 $F1$ 基本维持在 65%上下之间波动. 欧元和人民币属于不同的币种,不同币种的纹理风格不同,同一币种的纹理风格属于同类风格,而同类风格在不同面值之间的差异比较小,会对纸币图像的识别产生影响,因此,通过纹理风格识别纸币图像的准确率、精度、召回率和 $F1$ 比较低.

通过 ResNet-18 模型对色谱风格进行识别,如图 15(b)所示,随着迭代轮数的增加,其测试准确率、精度、召回率和 $F1$ 基本维持在 85%上下之间波动. 对于同一面值的纸币来说,其局部色谱风格相似,在色谱中,每一种颜色的比例不同,但颜色类型属于同一颜色系列,而欧元和人民币属于不同的币种,但欧元在色谱风格上与人民币相似,因此会导致在色谱风格的识别上准确率、精度、召回率和 $F1$ 降低.

通过 DenseNet-121 模型对纹理进行识别,如图 15(c)所示,随着迭代轮数的增加,其测试的准确率、精度、召回率和 $F1$ 基本维持在 87%. 不同面值的纸币,其纹理是不同的,但由于纸币局部扭曲的痕迹会对纹理产生一定的干扰,因此会导致在纹理的识别上准确率、精度、召回率和 $F1$ 降低.

(2)基于双特征聚合的识别

将纸币图像的纹理风格和色谱风格聚合后进行识别,如图 16(a)所示,随着迭代轮数的增加,

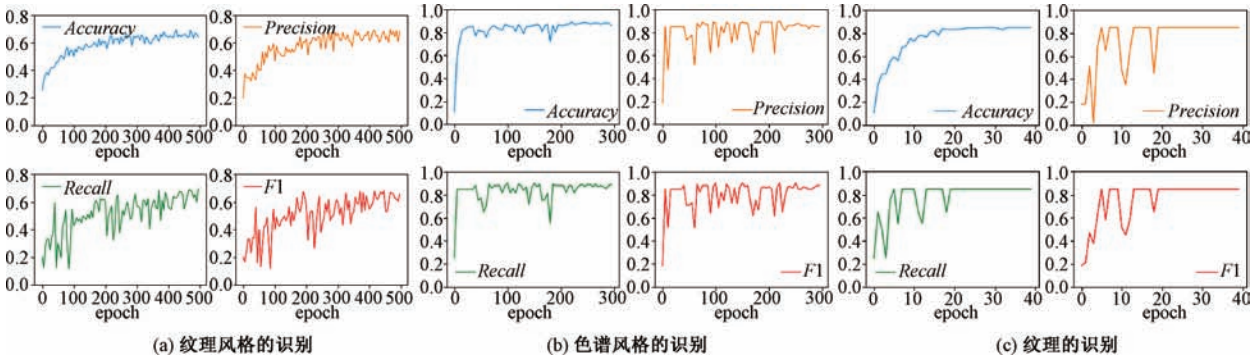

图 15　单特征的识别率变化图

其准确率、精度、召回率和 $F1$ 基本维持在 90% 上下之间波动. 同一币种的纹理风格相似,色谱风格可以减小纹理风格在同一币种间的相似度,欧元与人民币具有相似的色谱风格,不同币种间的纹理风格可以减小色谱风格的相似度,因此在准确率、精度、召回率和 $F1$ 上高于它们基于单特征的识别.

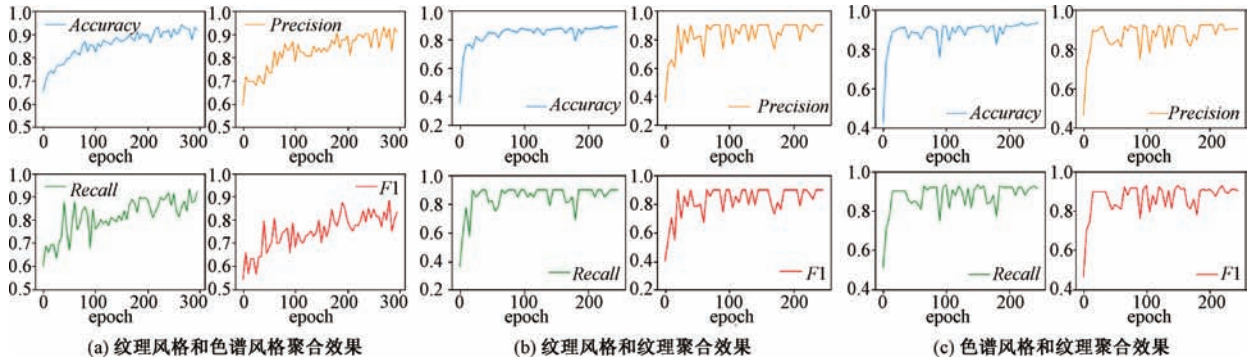

图 16　两种特征聚合的识别率变化图

将纸币图像的纹理风格和纹理聚合后进行识别,如图 16(b)所示,随着迭代轮数的增加,其测试的准确率、精度、召回率和 $F1$ 基本维持在 90%. 纹理可以减小纹理风格在同一币种间的相似度,纹理风格可以减小由于局部扭曲给纹理造成的影响,因此在准确率、精度、召回率和 $F1$ 上高于它们基于单特征的识别.

将纸币图像的色谱风格和纹理聚合后进行识别,如图 16(c)所示,随着迭代轮数的增加,其测试的准确率、精度、召回率和 $F1$ 基本维持在 91%上下之间波动. 欧元与人民币的色谱风格相似,纹理可以减小色谱风格在不同币种间的相似度,色谱风格可以减小由于局部扭曲对纹理产生的影响,因此在准确率、精度、召回率和 $F1$ 上高于它们基于单特征的识别.

(3)基于三特征聚合的识别

将纸币图像的纹理风格、色谱风格和纹理聚合后进行识别,如图 17 所示,随着迭代轮数的增加,其测试的准确率、精度、召回率和 $F1$ 基本维持在 94% 上下之间波动. 同一币种的纹理风格相似,色谱风格和纹理几乎可以消除其对纸币识别产生的影响,欧元和人民币的色谱风格相似,纹理风格和纹理几乎可以消除其对纸币识别产生的影响,纹理风格和色谱风格几乎可以忽略由于局部扭曲对纹理产生的影响,因此在准确率、精度、召回率和 $F1$ 上高于它们基于单特征和双特征的识别.

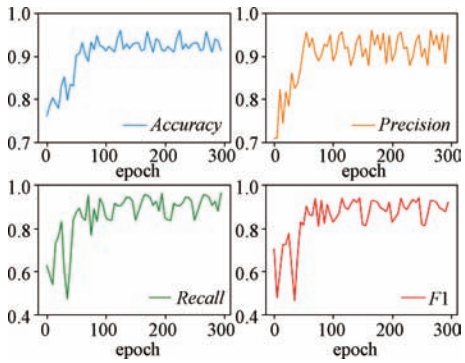

图 17　三种特征聚合的识别率变化图

如表 4 所示,是不同特征组合进行 10 折测试趋

于稳定的结果，其中 $F1$ 表示纹理风格，$F2$ 表示色谱风格，$F3$ 表示纹理. 从表中可以看出，两种或三种类型的特征聚合后，在准确率、精度、召回率和 $F1$ 上高于基于单一类型特征的识别.

通过以上内容的分析，可以得出结论：欧元和人民币的特征之间是互补的，可以最大可能地捕获图像的语义，消除同一币种内相似纹理风格产生的影响，减小不同币种间色谱风格的相似度，而且，这些特征的不变性，能够解决未表现出明显空间排列方式的图像识别问题.

表 4 不同特征组合的测试结果

特征	准确率(%)	精度(%)	召回率(%)	F1Score(%)
$F1$(VGG-16)	67.956±1.092	68.788±1.288	69.792±1.667	68.333±1.060
$F2$(ResNet-18)	88.000±1.000	88.889±1.111	88.000±0.500	87.000±1.000
$F3$(DenseNet-121)	84.900±0.100	85.000±0.000	85.000±0.000	85.000±0.000
$F1+F2$(VGG-16+ResNet-18)	89.762±1.290	89.424±1.591	88.563±1.167	88.833±1.378
$F1+F3$(VGG-16+DenseNet-121)	88.061±0.485	89.564±0.874	89.350±1.334	89.884±0.484
$F2+F3$(ResNet-18+DenseNet-121)	90.900±0.550	90.350±0.553	90.869±0.581	90.874±0.610
$F1+F2+F3$ (HFA-LVDBR)	95.632±0.450	94.546±0.331	94.100±0.884	92.484±0.485

4.2.2 多模型融合的性能比较

在本小节中，对不同类型的模型融合进行识别比较和分析，即 VGG-16、ResNet-18、DenseNet-121 之间的融合，其中每类模型符号前的数字表示在融合模型中，其被运用了几个. 例如：2VGG-16+ResNet-18 中，模型 VGG-16 前的数字 2，表示在三个模型融合中，运用了 2 个 VGG-16 和 1 个 ResNet-18. 从图 17 到图 19 显示了不同模型融合的实验测试变化趋势，可以看到：两种或三种类型的模型融合后，在准确率、精度、召回率和 $F1$ 上高于基于单一类型模型融合的识别.

(1)基于同类模型融合的识别比较

将获取的纹理风格、色谱风格和纹理依次输入 3VGG-16、3ResNet-18、3DenseNet-121 融合模型进行学习，然后将学习到的输出特征进行聚合并输入识别层 Softmax 进行识别. 如图 18(a)、(b)、(c)所示，随着迭代轮数的增加，其测试的准确率、精度、召回率和 $F1$ 分别维持在 82%、85%、88%上下之间波动. 因为均通过多特征学习，识别局部视图扭曲的纸币，比单网络的学习特征更全面反映局部视图扭曲的纸币，因此融合模型在准确率、精度、召回率和 $F1$ 上比单个网络的识别率高.

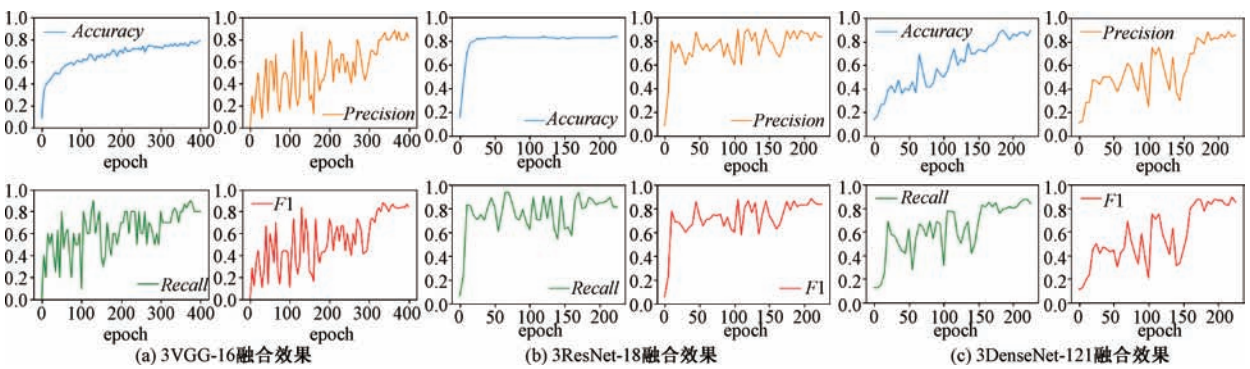

图 18 同类模型融合的测试变化图

(2)基于两类模型融合的识别比较

将获取的纹理风格、色谱风格和纹理依次输入 2VGG-16+ResNet-18、2VGG-16+ DenseNet-121、2ResNet-18 + VGG-16、2ResNet-18 + DenseNet-121、2DenseNet-121 + VGG-16、2DenseNet-121 + ResNet-18 融合模型进行学习，然后将学习到的输出特征输入识别层 Softmax 进行识别. 如图 19(a)、(b)、(c)、(d)、(e)、(f)所示，随着迭代轮数的增加，其测试的准确率、精度、召回率和 $F1$ 分别维持在 86%、88%、87%、89%、90%、91%上下之间波动. 因为均属于两类不同结构模型的融合，能够随着模型结构的不同，学习更多角度的输出特征，会比一类结构模型学习三类特征的角度更全面，因此在准确率、精度、召回率和 $F1$ 上比单类模型融合的识别率高.

如表 5 所示，是不同融合模型进行 10 折测试趋于稳定的结果，从表中可以看出，融合模型中的模型类型之间差距越大，其识别性能越高于基于单一或两类模型融合的性能. 可以得出结论，不同类型的模型得到的学习特征更具全面性，其融合模型能够从

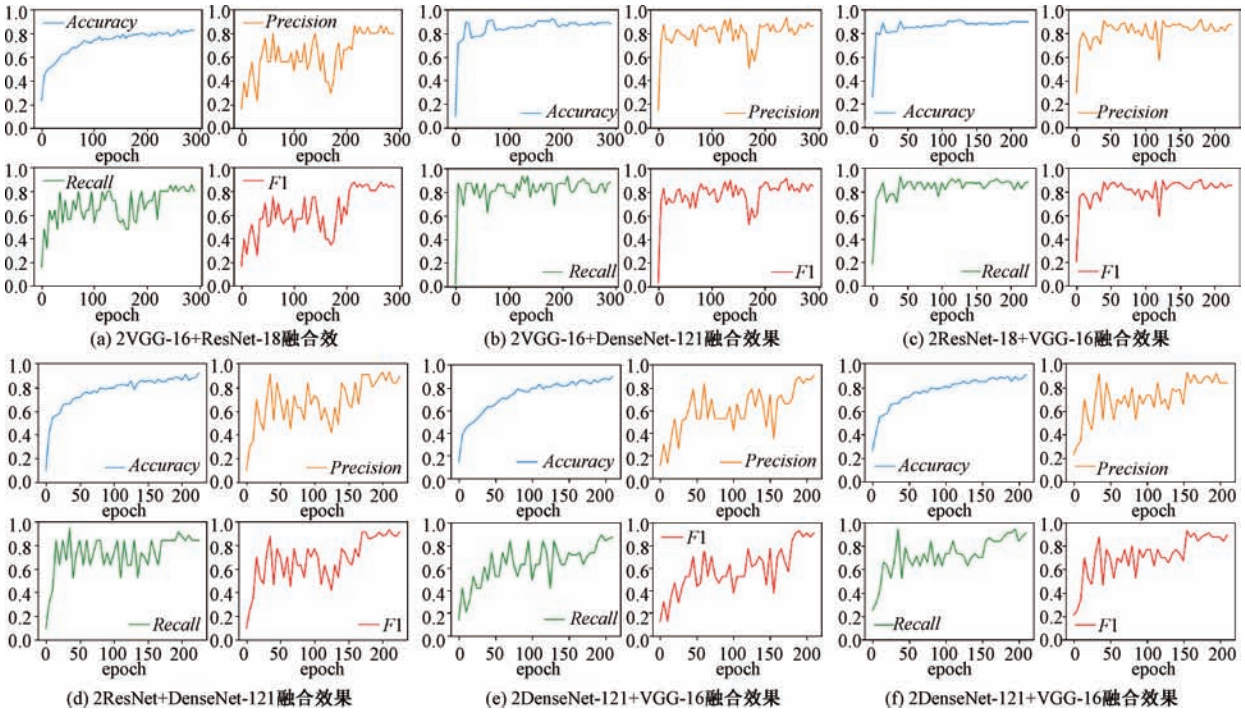

图 19　两类模型融合的测试变化图

表 5　不同模型融合的测试结果

Method	准确率(%)	精度(%)	召回率(%)	*F*1Score(%)
3VGG-16	79.313± 1.034	82.857±1.810	83.000±1.333	83.334±1.111
3ResNet-18	85.000±0.000	87.878±1.011	89.583±1.042	87.571±0.731
3DenseNet-121	89.526±0.768	88.750±1.250	87.500±0.833	88.889±0.540
2VGG-16+ResNet-18	85.595±0.817	86.917±1.245	85.778±1.213	88.546±1.034
2VGG-16+DenseNet-121	89.951± 0.082	88.857±0.810	87.667±1.083	88.334±1.143
2ResNet-18+VGG-16	89.869±0.068	87.778±0.635	88.350±0.465	86.667±0.834
2ResNet-18+DenseNet-121	91.553±0.668	91.534±0.467	89.162±0.561	91.599±379
2DenseNet-121+VGG-16	90.168±0.842	90.236±0.667	90.128±0.656	90.109±0.741
2DenseNet-121+ResNet-18	90.138±1.112	91.318±0.672	91.068±1.103	90.332±1.326
HFA-LVDBR (VGG-16+ResNet-18+DenseNet-121)	95.632±0.450	94.546±0.331	94.100±0.884	92.484±0.485

多角度抽取图像的本质特征,得到有效的识别区域,提升图像的识别率.

4.2.3　与现有方法的性能比较

因为本文方法既与纸币识别有关,又与面向局部视图的识别相关,所以选择方法[12]的纸币识别与方法[22]面向局部视图的细粒度识别,从识别性能和效率两方面与本文方法进行对比.

(1)各方法的识别性能比较

通过测试集,以不同币种(人民币 RMB、欧元 Euro)在 top1、top5 的识别准确率为评价指标,比较本文方法与文献[12]、[22]方法的识别性能. 如图 20 所示,是不同方法对不同币种的识别结果,从图中可以看出,本文方法在不同币种的识别准确率高于其它两个方法. 因为文献[12]方法直接从整张纸币图像中提取视觉特征,当几何结构变化较大时可能会识别错误,这是因为 CNN 只能通过最大合并来处理具有小范围扭曲的图像. 文献[22]方法主要针对特定的局部结构,在 CNN 网络上加入循环注意力机制进行增强识别,而扭曲型纸币没有特定的结构. 本文方法通过不变形特征对扭曲型纸币的几何变形不敏感,因此比文献[12]、[22]方法识别准确率更高.

(2)各方法的效率比较

在本文方法中,融合模型由三个基本卷积型模型融合. 其中,每个卷积层的时间复杂度为 $t \sim O(L^2 \cdot K^2 \cdot C_{in} \cdot C_{out})$,其中 L 表示输出特征的维度,K 表示卷积核的大小,C_{in} 表示输入通道数,C_{out} 表示当前卷积层的卷积核数,即输出通道数. 所以,本文融合模型的时间复杂度为 $T \sim O(\sum_{l=1}^{D} L_l^2 \cdot$

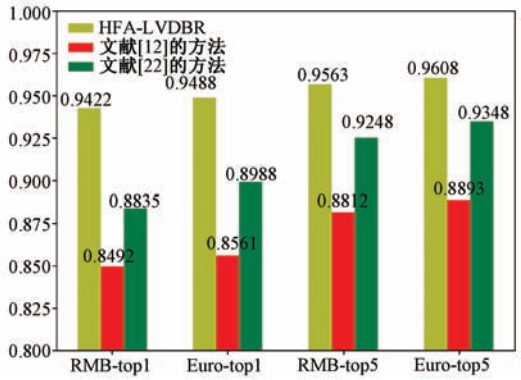

图 20 不同方案的识别结果

$K_l^2 \cdot C_{l-1} \cdot C_l$），其中 D 表示融合模型的最大卷积层数，C_{l-1} 表示输入第 l 个卷积层的通道数 C_{in}，C_l 表示输出第 l 个卷积层的通道数 C_{out}.

空间复杂度与总参数量、各层输出特征相关，与输入特征的大小无关. 总参数量只与卷积核的大小 K、通道数 C、平均层数 $\bar{D}=(D_1+D_2+D_3)/n$、基本模型个数 n 相关，其中 D_1、D_2、D_3 分别表示改进 VGG-16、ResNet-18、DenseNet-121 的层数. 输出特征空间是其特征维度的平方 L^2 连乘通道数 C 的大小. 因此，融合模型的空间复杂度为 $S\sim O\left(n\cdot\left(\sum_{l=1}^{\bar{D}}(K_l^2\cdot C_{l-1}\cdot C_l+L^2\cdot C_l)\right)\right)$.

如表 6 所示，从时间和空间复杂度对本文方法的融合模型与文献[12]、[22]方法的效率进行了对比. 本文融合模型的空间复杂度高于其它两个方法，是因为融合模型由三个基本卷积模型融合，而其它两个方法的模型是由单个卷积模型组成，但是三个方法的时间复杂度基本相同，因为本文融合模型是并行融合，所以在时间复杂度上只取决于网络层数.

表 6 不同方法的效率对比

方法	时间复杂度	空间复杂度
文献[12]方法	$T_1\sim O(\sum_{l=1}^{6}L_l^2\cdot K_l^2\cdot C_{l-1}\cdot C_l)$	$S_1\sim O(\sum_{l=1}^{6}(K_l^2\cdot C_{l-1}\cdot C_l+L^2\cdot C_l))$
文献[22]方法	$T_2\sim O(\sum_{l=1}^{20}L_l^2\cdot K_l^2\cdot C_{l-1}\cdot C_l)$	$S_2\sim O(\sum_{l=1}^{20}(K_l^2\cdot C_{l-1}\cdot C_l+L^2\cdot C_l))$
HFA-LVDBR	$T_3\sim O(\sum_{l=1}^{123}L_l^2\cdot K_l^2\cdot C_{l-1}\cdot C_l)$	$S_3\sim O\left(n\cdot\left(\sum_{l=1}^{53}(K_l^2\cdot C_{l-1}\cdot C_l+L^2\cdot C_l)\right)\right)$

5 结束语

在日常生活中，任何物体不可避免地出现像纸币这样被改变原有的结构，而物体的本质不变，针对识别这样的物体，我们提出了通过不变形特征来表示同一纸币不同结构的方法，即基于异构特征聚合的局部视图扭曲型纸币识别. HFA-LVDBR 将获取纹理风格、色谱风格和纹理这些不变形特征，分别通过改进的 VGG-16、ResNet-18、DenseNet-121 学习，将在每个网络最后的最大池化层得到输出特征，进行规范化并聚合，然后通过融合模型的识别层 Softmax 进行图像识别. HFA-LVDBR 可以生成更强大、更具区别性和综合性的特征表示，用于处理局部视图扭曲型纸币的视觉复杂性. 大量的实验结果表明，HFA-LVDBR 在识别局部视图扭曲型纸币图像上，其准确率、精度、召回率和 $F1$ 优于单特征和双特征聚合以及单类模型和两类模型融合的识别性能. 与现有方法相比，其在准确率和效率方面均表现出相对较好的结果. 可得出结论，特征之间是互补的，通过不同类型的模型学习，能够从多角度表示局部视图扭曲的物体，且不变形特征对物体的几何变形不敏感.

虽然本文对局部视图扭曲型纸币进行了识别，但是需要预先提取不变形特征，而且模型融合方式比较复杂，引入更多特征会增加其复杂度等，所以在未来的研究方向上，我们计划从以下几个方向来进行：(1)将注意力机制[26]引入到我们的方法中，以定位可辨别纹理区域，而不是预先提取不变形特征；(2)对于不同特征，可利用深度强化学习方法[27]，自动监测特征；(3)通过可变自动编码器[28]等方法，降低模型融合的复杂度.

致 谢 衷心感谢审稿专家在百忙之中提出的宝贵意见.

参 考 文 献

[1] Krizhevsky A, Sutskever I, Hinton G-E. Imagenet classification with deep convolutional neural networks//Proceedings of the 25th International Conference on Neural Information Processing Systems. Nevada, USA, 2012: 1097-1105

[2] Zhou B, Lapedriza A, Khosla A, et al. Places: A 10 million

image database for scene recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018,40(6):1452-1464

[3] Gong Y, Wang L, Guo R, et al. Multi-scale orderless pooling of deep convolutional activation features//Proceedings of the 2014 European Conference on Computer Vision. Zurich, Switzerland, 2014:392-407

[4] Wu R, Wang B, Wang W, et al. Harvesting discriminative meta objects with deep CNN features for scene classification//Proceedings of the 2015 IEEE International Conference on Computer Vision. Santiago, Chile, 2015:1287-1295

[5] Szegedy C, Liu W, Jia Y-Q, et al. Going deeper with convolutions//Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA, 2015:1-9

[6] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA, 2016:770-778

[7] Huang G, Liu Z, Maaten L-V-D, et al. Densely connected convolutional networks//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA, 2017:2261-2269

[8] Liu Y-P, Du Q-C, Zhang J-D. Recognising surface and direction of banknotes based on bp neural network. Computer Applications and Software, 2015,32(11):176-179(in Chinese)
(刘艳萍,杜秋晨,张进东. 基于 BP 神经网络的纸币面向识别方法. 计算机应用与软件,2015,32(11):176-179)

[9] Liu C-H,Yao J-F. An recognition algorithm based on statistical feature for damaged degree of paper money. Journal of Xinyang Normal University (Natural Science), 2016,29(4):617-620(in Chinese)
(柳春华,姚建峰. 基于图像统计特征的新旧纸币识别算法. 信阳师范学院学报(自然科学版),2016,29(4):617-620)

[10] Pham T-D, Kim K-W, Kang J-S, et al. Banknote recognition based on optimization of discriminative regions by genetic algorithm with one-dimensional visible-light line sensor. Pattern Recognition, 2017,72(01):27-43

[11] Dittimi T-V, Hmood A-K, Suen C-Y. Multi-class SVM based gradient feature for banknote recognition//Proceedings of the 2017 IEEE International Conference on Industrial Technology. Toronto, Canada, 2017: 1030-1035

[12] Liu Y-J, He J-B, Li M. New and old banknote recognition based on convolutional neural network//Proceedings of the 2018 International Conference on Mathematics and Statistics. Porto, Portugal, 2018:92-97

[13] Gai S, Bao Z-Y. Banknote recognition research based on improved deep convolutional neural network. Journal of Electronics & Information Technology, 2019,41(08):1992-2000(in Chinese)
(盖杉,鲍中运. 基于改进深度卷积神经网络的纸币识别研究. 电子与信息学报,2019,41(8):1992-2000)

[14] Song P-G, Li K-Y, Cheng W-P. Improved algorithm of paper currency number recognition rate. Techniques of Automation and Application, 2014,33(11):74-78(in Chinese)
(宋普庚,李开宇,程卫平. 纸币冠字号识别率改进算法. 自动化技术与应用,2014,33(11):74-78)

[15] Zhao M. Extraction, identification on crown word number of one hundred yuan RMB. Journal of Nanchang Hangkong University (Natural Sciences),2017,31(4):91-95(in Chinese)
(赵敏. 百元人民币冠字号码提取与识别. 南昌航空大学学报(自然科学版),2017,31(4):91-95)

[16] Jang U, Lee E-C. Convolutional neural network based serial number recognition method for indian rupee banknotes//Proceedings of the 2017 International Conference on Ubiquitous Information Technologies and Applications. Taichung, China, 2017:1445-1450

[17] Tsai Y-S, Hsieh Y-Y, Ho C-H, et al. Rule-based optical character recognition for serial number on renminbi banknote. Electronic Imaging, Image Processing: Algorithms and Systems XVI. Burlingame, USA, 2018: 308-1-308-6(6)

[18] Song X, Jiang S, Herranz L. Multi-scale multi-feature context modeling for scene recognition in the semantic manifold. IEEE Transactions on Image Processing, 2017,26(6):2721-2735

[19] Herranz L, Jiang S, Li X. Scene recognition with CNNs: Objects, scales and dataset bias//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA, 2016:571-579

[20] Peng Y, He X, Zhao J. Object-part attention model for fine-grained image classification. IEEE Transactions on Image Processing, 2018,27(3):1487-1500

[21] Branson S, Horn V-G, Belongie S, et al. Bird species categorization using pose normalized deep convolutional nets. arXiv:1406.2952v1, 2014:1-14

[22] Fu J, Zheng H, Mei T. Look closer to see better: Recurrent attention convolutional neural network for fine-grained image recognition//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA, 2017: 4476-4484

[23] Zhang P, Zhu X-Y, Cheng Z-Z, et al. REAPS: Towards better recognition of fine-grained images by region attending and part sequencing//Proceedings of the Chinese Conference on Pattern Recognition and Computer Vision. Xi'an, China, 2019:193-204

[24] Wang P-Z, Dong S, Zhang D-L, et al. Texture analysis of vitrinite macerals based on the uniform pattern of round LBP. Journal of Anhui University of Technology (Natural Science), 2014, 31(2):147-151(in Chinese)
(王培珍,董双,张代林,等. 基于圆形 LBP 均匀模式的煤镜质组显微组分纹理分析. 安徽工业大学学报(自然科学版),2014,31(2):147-151)

[25] Orhan E, Bengio S, Wallach H, et al. A simple cache model for image recognition//Proceedings of the Neural Information Processing Systems. Montreal, Canada, 2018:10128-10137

[26] Zheng H-L, Fu J-L, Mei T, et al. Learning multi-attention convolutional neural network for fine-grained image recognition//Proceedings of the 2017 IEEE International Conference on Computer Vision. Venice, Italy, 2017:5219-5227

[27] Tan L-T, Hu R-Q. Mobility-aware edge caching and computing in vehicle networks: A deep reinforcement learning. IEEE Transactions on Vehicular Technology, 2018,67(11): 10190-10203

[28] Tang L-M, Xue Y-X, Chen D, et al. Multi-entity dependence learning with rich context via conditional variational auto-encoder//Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence. New Orleans, USA, 2018: 824-832

GUO Yu-Hui, Ph. D. candidate. Her research interests include deep learning and image processing.

LIANG Xun, Ph. D., professor, Ph. D. supervisor. His research interests include neural networks, social computing and natural language processing.

Background

Banknote recognition is a new research direction in the field of computer science. Most of the existing methods directly extract deep visual features for banknotes recognition, which ignore how to recognize banknotes based on distorted local view deformation in human view. The banknote with a distorted local view does not show unique spatial layout and structure. Therefore, standard object recognition methods may not perform well on this task. Secondly, local banknote recognition can be considered as fine-grained recognition. The first step of fine-grained object recognition is usually to find the fixed semantic part of some objects. However, the common semantic part does not exist in many deformed local banknote images. Therefore, it is difficult to capture semantic information from the deformed local banknote image by the existing fine-grained method. Third, similar to object recognition, local banknote images also have various geometric changes, such as different views, rotation and scale. The banknote with distorted local view requires that the banknote recognition method should have geometric invariance and be able to recognize banknote images reliably. The existing banknote recognition methods usually use CNN to extract the visual features directly from the whole banknote image, and may recognize errors when the geometry changes greatly. This is because CNN can only process images with a small range of deformation through maximum merging. In this paper, we propose a new method, named Banknote Recognition based on Local View Distortion, to handle local view distorted banknote recognition. In the first stage, we extract the nondeformable feature to represent the local view distorted banknotes. In the second stage, we use VGG-16, ResNet-18, DenseNet-121 model to learn the nondeformable features, and aggregate the learned features. In the third stage, the three deep models are fused, and then the aggregation features are identified, so as to identify the distorted local view banknotes. The experimental results show that the recognition rate of this scheme is higher and more universal, which can be extended to other visual images. In addition, the HFA-LVDBR scheme is universal, which can fuse different types of deep networks and apply them. This scheme is superior to single and dual features in accuracy, precision, recall and F1 in different types of feature aggregation, and it is higher than that based on a single model and two kinds of models in different types of model fusion.

This work is supported by the National Social Science Foundation of China (18ZDA309), the National Natural Science Foundation of China (62072463), the Natural Science Foundation of Beijing (4172032), and the Opening Project of State Key Laboratory of Digital Publishing Technology of Founder Group (413217003).