E-mail: jig@aircas.ac.cn
Website: www.cjig.cn
Tel: 010-58887035

中国图象图形学报
JOURNAL OF IMAGE AND GRAPHICS
©中国图象图形学报版权所有

中图法分类号:TP399　　**文献标识码**: A　　**文章编号**: 1006-8961(2023)08-2370-12

论文引用格式:Cheng W Y, Zhou Y, Tao C Y, Liu L, Li Z G and Qiu T R. 2023. Multi-loss siamese convolutional neural network for Chinese calligraphy font and style classification. Journal of Image and Graphics, 28(08):2370-2381(程文炎, 周勇, 陶承英, 刘丽, 李志刚, 邱桃荣. 2023. 多损失融合网络的中国书法字体与风格分类. 中国图象图形学报, 28(08):2370-2381)[DOI:10.11834/jig.220252]

多损失融合网络的中国书法字体与风格分类

程文炎[1],周勇[1],陶承英[2],刘丽[1*],李志刚[1],邱桃荣[1]

1. 南昌大学数学与计算机学院,南昌　330031; 2. 北京师范大学实验小学,北京　100875

摘　要:**目的**　中国书法博大精深,是中国文化很重要的组成部分。书法字体与风格分类是书法领域的研究热点。目前书法字体和书法风格两个概念混淆,并且书法风格分类准确率不高,针对上述问题,本文将两个概念进行区分,并提出了一个融合多损失的孪生卷积神经网络,能同时解决中国书法字体以及风格分类问题。**方法**　提出的网络包含两个共享权重的分支,每个分支用于提取输入图像的特征。为了获得不同尺度下的特征表示,将Haar小波分解嵌入到每个网络分支中。与传统孪生神经网络不同的是,将网络的每个分支扩展为一个分类网络。网络训练时融合了两类不同的损失,即对比损失和分类损失,进而从两个角度同时对网络训练进行监督。具体来说,为了使来自同一类的两幅输入图像特征之间的距离尽可能小、使来自不同类的两幅输入图像特征之间的距离尽可能大,网络采用对比损失作为损失函数。此外,为了充分利用每幅输入图像的类别信息,在网络每个分支上采用交叉熵作为分类损失。**结果**　实验结果表明,本文方法在两个中国书法字体数据集和两个中国书法风格数据集上的分类准确率分别达到了99.90%、94.09%、99.38%和93.28%,高于对比方法。两种损失起到了良好的互补作用,Haar小波分解的引入在4个数据集上均提升了分类准确率,在风格数据集的提升效果更为明显。**结论**　本文方法在中国书法字体以及风格分类两个任务中取得了令人满意的效果,为书法领域研究工作提供了新思路。

关键词:中国书法;风格分类;字体分类;多损失融合孪生卷积神经网络;对比损失;交叉熵损失

Multi-loss siamese convolutional neural network for Chinese calligraphy font and style classification

Cheng Wenyan[1], Zhou Yong[1], Tao Chengying[2], Liu Li[1*], Li Zhigang[1], Qiu Taorong[1]

1. School of Mathematics and Computer Sciences, Nanchang University, Nanchang　330031, China;

2. Experimental Primary School, Beijing Normal University, Beijing　100875, China

Abstract: **Objective** Chinese calligraphy can be seen as one of the symbolized icons in Chinese culture. Nowadays, Machine learning and pattern recognition techniques-derived calligraphy artworks are required for digitalization and preservation intensively. Our research is mainly focused on Chinese calligraphy classification in related to such font and style classification. However, the difference between calligraphy font and calligraphy style is often distorted. To resolve the problem of style classification, first, we distinguish the difference between font and style. Then, we illustrate a novel multi-loss siamese convolutional neural network to cope with the two mentioned problems simultaneously. **Method** The difference

收稿日期:2022-03-14;**修回日期**:2022-07-12;**预印本日期**:2022-07-19

*通信作者:刘丽　liuli_033@163.com

基金项目:国家自然科学基金项目(61603256,62006150);上海青年科技英才扬帆计划项目(19YF1418400)

Supported by: National Natural Science Foundation of China (61603256, 62006150); Shanghai Young Science and Technology Talents Sailing Program (19YF1418400)

can be summarized between calligraphy font and style as follows: Calligraphy font refers to a broad taxonomy of scripts. For example, the popular Chinese calligraphy fonts are composed of standard, seal, clerical, cursive and semi-cursive fonts. Calligraphy style is closely related to the calligraphers to a certain extent. Each calligrapher has its own unique style. Compared to calligraphy font classification, calligraphy style classification is more challenging due to the subtle difference among different styles. Current researches are dedicated to Chinese calligraphy font classification. Yet, there are only a few literatures are concerned with Chinese style classification. Our network is proposed and composed of two weights-shared streams. Each stream of the network can be used to extract features from the input image using convolutional neural network (CNN). In detail, the CNN has involved of five convolutional layers, while each layer is followed via a max pooling layer. Batch normalization is used to speed up training as well. The ReLU is used as the activation function. Afterwards, the global average pooling is used to aggregate the feature maps into a compactable feature vector. To get a multi-resolution representation of the image, the Haar wavelet decomposition is embedded into each stream. To optimize traditional siamese network, each stream of the proposed siamese network is extended as a classification network. In this way, image-related features extraction is then fed to a fully-connected layer for classification. The cross-entropy loss is employed for each stream. So, the supervised information of each individual image can be fully exploited. The contrastive loss can be used for feature constraints: 1) features-between distance of the two input images from the same category will be reduced; 2) features-between distance of the two input images from different categories will be enlarged. Overall, the proposed network is trained to optimize the two types of loss jointly: contrastive loss and cross-entropy loss. A weight parameter is used to balance the contribution of the two parts as well. **Result** We carried out extensive experiments to validate the effectiveness of the proposed network. Since there are no public datasets for Chinese calligraphy font and style classification, we have collected four sort of datasets in related to viz. CNCalliFont, CNCalliNoisyFont, CNCalliStyle and CNCalliNoisyStyle. The CNCalliFont dataset is composed of 30 000 images and melted into five different fonts, called clerical, cursive, seal, semi-cursive and standard fonts for each. Each font has final 6 000 images. The CNCalliNoisyFont dataset can be used to extend CNCalliFont dataset in terms of the added Gaussian noise. The CNCalliStyle dataset consists of 12 000 images, which represent four styles from four popular Chinese ancient calligraphers in related to viz., Ouyang Xun, Yan Zhenqing, Zhao Mengfu and Liu Gongquan. Therefore, each style of that has 3 000 images more. All the images are related to grayscale and kept in JPEG format. Likewise, the CNCalliNoisyStyle dataset can be focused on CNCalliStyle dataset extension via adding Gaussian noise. Each dataset is split into training set, validation set and test set further with a ratio of 6∶2∶2. The training set is used to learn the parameters in the proposed network. Different configuration of the hyper-parameters is compared on the validation set with the best configuration selected, and it is applied to the test set. Ten sort of random splits and the average classification accuracy are melted into as the evaluation metric. The four datasets-related experiments demonstrate that the performance is increased by embedding the Haar wavelet decomposition in each stream of the network. It is mutual-benefited on CNCalliStyle and CNCalliNoisyStyle datasets, indicating that the subtle difference among different styles can be captured better in terms of the Haar wavelet decomposition. The performance of the proposed network is compared with them when the cross-entropy loss is employed only. The result shows that the performance is decreased when the cross-entropy loss is employed only. So, the two types of loss can be mutual-benefited as well. Moreover, we compare the proposed network with such popular methods in the context of manual feature-based, CNN-based and vision transformer-based. For the handcrafted feature-based methods, features like local binary pattern (LBP), Gabor and histogram of oriented gradient (HOG) are first employed, and the suppert vector machine (SVM) classifier is then used. For CNN-based methods, recent four sorts of methods for Chinese calligraphy font and style classification are listed. Additionally, we compare our proposed network with four sort of popular CNNs, which involves AlexNet, Visual Geometry Group (VGG-16), residual neural network (ResNet-50) and Xception. The performances of all the methods are still decreased on the CNCalliStyle dataset. Specifically, a very sharp decrease is observed for the feature-handcrafted methods, the four popular CNNs, as well as the vision transformer-based methods. It indicates that these methods cannot capture the subtle difference among different styles. Each of the accuracy can be reached to 99.90%, 94.09%, 99.38% and 93.28% on the four datasets. **Conclusion** The proposed multi-loss siamese CNN can be dealt with the two tasks in relevance to viz. Chinese calligraphy font and calligraphy style classification simultaneously. Two sorts of task can be jointly

optimized based on two types of loss as well.

Key words: Chinese calligraphy; style classification; font classification; multi-loss siamese convolutional neural network; contrastive loss; cross-entropy loss

0 引 言

书法是中国传统文化的精粹，是中华民族珍贵的文化遗产。中国历史上涌现出许多著名的书法家，他们的作品享誉全球，对世界书法艺术产生了深远影响。随着数字化时代的到来，越来越多的书法作品扫描成数字图像格式存储，避免了因保存不当造成的破损及老化等问题。同时，吸引了学者们尝试用机器学习以及模式识别等方法辅助书法图像的分类、识别以及检索等相关问题的研究。

本文研究目标为中国书法分类，具体来说包括书法字体和书法风格分类。目前大部分研究通常将书法字体和书法风格两个概念混淆（Zhang 等，2019）。著名书画家、教育家启功先生在《古代字体论稿》中提到可以从字体和风格两个不同角度来描述书法文字。书法字体指汉字本身的不同体式、体裁，主要包括5类，即篆书、隶书、草书、楷书以及行书。书法中每一种字体在形态特征和总体风貌方面都有各自明显的特点。例如，篆书是早期的汉字，不易辨识，笔画圆润、瘦劲，粗细变化较少，结体多纵长，常呈左右对称状；隶书由篆书演变而来，笔画多波挑，一个明显特征是“蚕头雁尾”，形体多呈宽扁之势；草书点画连绵，形态自由草率，不易识别；楷书是现在通行的字体，笔画类型多样，结构接近长方形，中规中矩，有如楷模，故称“楷书”。行书介于楷书和草书之间，字中有一定的牵连笔画，灵动活泼，有如行云流水，容易辨识，便于书写，是人们日常手写体最常用的字体。图1展示了不同书法字体的示例。

相比较书法字体而言，书法风格主要指由有代表性的著名书法家形成的独特风格，更侧重于书法家个体在其书迹中展现的个性鲜明的风格特色，书法家书法风格的形成受天赋、阅历以及师承等多方面因素的影响。例如，同为楷书字体，“楷书四大家”（欧阳询、颜真卿、柳公权、赵孟頫）的书法风格各异，欧体楷书笔画多用方笔，结构严谨；颜体楷书用笔丰腴、圆润，体态宽博，给人以大气磅礴之感；柳体楷书点画有如斧劈刀削，富于骨力，体势开张挺拔；赵体楷书笔法秀媚，结体活泼，于庄重中见潇洒。图2给出了上述4位书法家的书法风格示例。



(a) 草书 (b) 楷书 (c) 隶书 (d) 篆书 (e) 行书

图1 不同书法字体示例

Fig. 1 Samples of different calligraphy fonts ((a)cursive; (b)standard; (c)clerical; (d)seal; (e)semi-cursive)



(a) 柳公权 (b) 欧阳询 (c) 颜真卿 (d) 赵孟頫

图2 不同书法风格示例

Fig. 2 Samples of different calligraphy styles ((a) Liu Gongquan; (b) Ouyang Xun; (c) Yan Zhenqing; (d) Zhao Mengfu)

书法字体和书法风格分类在书法图像组织存储、识别、检索、书法鉴赏、书法教育，乃至书法文物鉴定等领域，均起到了非常重要的作用。此外，书法字体和书法风格分类能为名家书法作品的鉴定提供重要参考，对提高文博部门的工作效率和准确性很有裨益。然而，由于较大的类内差异及较小的类间差异使得这两个问题均具有较大的挑战性。相比较书法字体分类，书法风格分类难度更大。主要原因在于不同书法风格之间的差异通常来说非常细微，对于未经专业训练的普通人来说很难觉察，给分类带来困难。因此，书法风格分类问题属于细分类问题。由于经过数字化后的书法图像样本通常数量较多，依靠专业人士进行人工分类不仅存在成本问题，而且人工作业易疲劳进而导致误分类。因此研究一种自动的书法字体和风格分类方法具有非常重要的意义。

本文提出了一个端到端融合多损失的孪生卷积神经网络，能同时解决中国书法字体和风格分类问题。网络由两个共享权重的分支构成，与传统孪生神经网络不同的是，本文将网络中每个分支扩展为

一个分类网络。主要贡献如下：1）网络训练时融合不同类型的损失，可以从不同角度对网络进行监督，起到良好的互补作用。具体来说，为了使来自同一类的两幅输入图像特征之间的距离尽可能小，使来自不同类的两幅输入图像特征之间的距离尽可能大，采用对比损失作为损失函数。此外，在网络每个分支上利用交叉熵作为损失函数，可以充分利用每幅输入图像的类别信息。2）每个分支利用卷积神经神经网络提取输入图像特征，并将Haar小波嵌入到每个分支中，获得不同尺度下的图像特征表示，能较好地捕捉不同类别图像之间的细微差异。3）与对比方法相比，本文方法在中国字体与风格分类两个任务上均取得了令人满意的效果，优于对比方法。

1 相关工作

1.1 中国书法字体分类

目前，中国书法字体分类方法大致可以分为两类，即基于传统手工特征的方法以及基于深度学习的方法。基于传统手工特征的方法主要包括特征提取以及分类器设计两个步骤。例如，Ding等人（2007）首先提取图像中的小波特征，然后用修正二次判别函数（modified quadratic discriminative function，mqdf）进行分类。Zhang等人（2013）首先检测图像中的Harris角点，并在每个角点上提取其Gabor特征，然后利用支持向量机（support vector machine，SVM）分类器进行分类。Song等人（2015）利用多尺度稀疏表示来解决字体分类问题。然而手工特征提取不仅烦琐复杂，具有较强的主观性，而且特征泛化能力较弱。

与基于传统手工特征的方法不同的是，基于深度学习的方法能自动学习图像表示（夏梦 等，2017）。Tao等人（2016）将字体分类看成一个序列分类问题，提出一个名为PC-2DLSTM（principal component-2D long short-term memory）的网络解决该问题。Huang等人（2018）提出IFN（inception font network）网络，整合了一个改进的Inception模块、CCCP（cascaded cross-channel parametric pooling）层以及全局平均池化，并提出一种新的数据增强方法DropRegion且能有效地嵌入到上述IFN网络中。Dai等人（2018）提出3种不同的卷积神经网络模型用于解决字体分类问题，包括局部卷积神经网络（local convolutional neural network，LCNN）、全局卷积神经网络（global convolutional neural network，GCNN）和双通道卷积神经网络（two pathway convolutional neural network，TPCNN）。Zhang等人（2019）同样利用卷积神经网络解决字体分类问题，并在网络中嵌入注意力机制SE（squeeze-and-excitation）模块（Hou 等，2021）以及Haar小波分解（Cao等，2020）。

1.2 中国书法风格分类

相比较中国书法字体分类，目前针对中国书法风格分类的研究较少，现有方法均为基于深度学习的方法。上述Dai等人（2018）以及Zhang等人（2019）将提出方法同时用于解决风格分类问题。此外，Zhang等人（2019）提出的方法得到进一步改进，具体来说，利用CBAM（convolutional block attention module）模块（Woo等，2018）替换网络中原来的SE模块（Zhang等，2021），主要原因在于SE模块仅在通道上应用了注意力机制，而CBAM模块在通道和空间上均应用了注意力机制。Liu等人（2021）提出一种基于孪生神经网络的书法风格分类方法，在实验中将所提出方法用于解决4种不同书法风格的分类问题。

1.3 孪生神经网络

传统孪生神经网络一般由两个共享权重的分支构成，主要用于衡量输入图像之间的相似性。将两个输入样本用相同的方法映射成两个特征向量，进而通过判断二者特征是否相似来衡量输入图像的匹配度（王德港 等，2021）。孪生神经网络通常采用对比损失函数来优化网络，在图像检索（Dubey，2022）、目标跟踪（Sun等，2022）等领域具有广泛应用。

与传统孪生神经网络不同的是，本文将孪生神经网络每个分支扩展为一个分类网络。网络训练时融合了对比损失和分类损失，不仅可以利用两个输入图像的相似性关系，而且能充分利用每幅输入图像的类别信息。

2 本文方法

本文采用基于深度学习的方法，提出了一个融合多损失的孪生卷积神经网络，能同时解决中国书法与风格的分类问题。具体网络结构如图3所示，包括两个共享权重的分支。网络训练时同时优化对比损失以及分类损失。

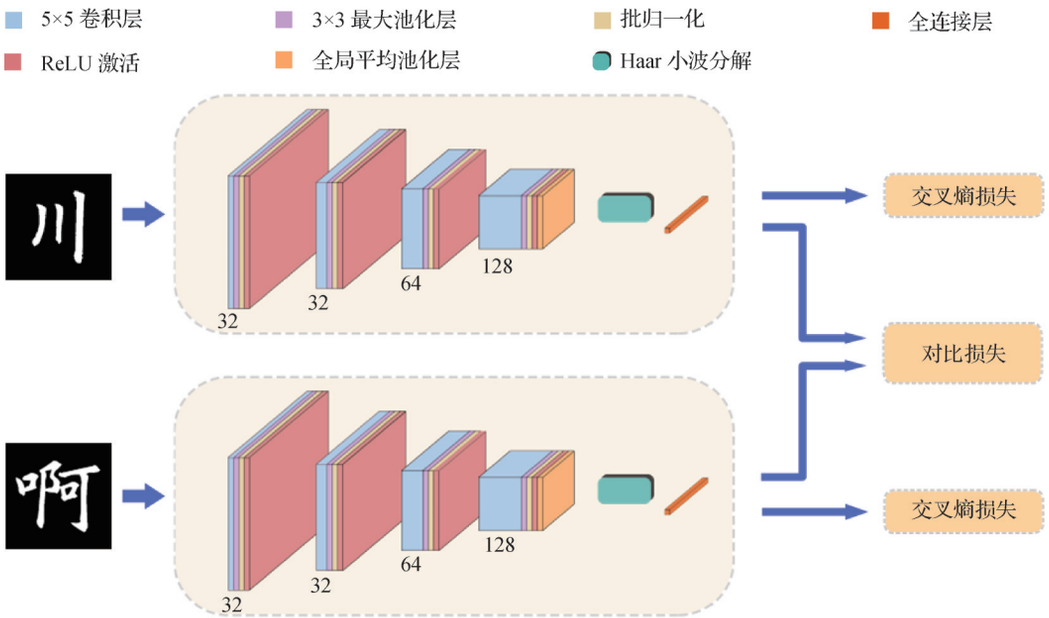


图3　本文网络架构示意图

Fig. 3　Architecture of the proposed network

2.1　图像特征提取

网络中每个分支采用卷积神经网络提取图像特征。受Zhang等人(2019)的启发,本文采用的卷积神经网络包含5个卷积层,其中每个卷积层后面紧跟一个最大池化层,在每个最大池化层后面引入批归一化(batch normalization, BN)层(蔡哲栋等,2021),从而加快网络训练的收敛速度。每层均采用ReLU(rectified linear unit)作为激活函数,利用全局平均池化将特征聚合,具体如图3所示。

由于小波具有良好的时频局部化、多分辨率分析等特性,在很多领域具有广泛应用,例如图像降噪(Liu等,2020)、图像分类(Khatami等,2020)等。考虑到Haar小波具有计算简单的特点,而且在许多任务中取得了令人满意的结果(Cao等,2020),本文利用Haar小波将上述提取的特征进一步进行多分辨率分解,获得不同尺度下的特征表示。具体来说,给定特征向量$\boldsymbol{F}$,假设其维数为T,对其进行K层Haar小波分解的算法表示为:

while $K>0$;

$T \leftarrow T/2$;

for $j \leftarrow 0$ to $j \leftarrow T-1$;

$F'[j] \leftarrow (F[2\times j]+F[2\times j+1])/2$;

$F'[j+1] \leftarrow (F[2\times j]-F[2\times j+1])/2$;

end for;

$\boldsymbol{F} \leftarrow \boldsymbol{F}'$;

$K \leftarrow K-1$;

end while。

其中,$\boldsymbol{F}'$为经过K层Haar小波分解的特征。本文实验中验证了将Haar小波分解嵌入到网络每个分支中的有效性,具体见3.4节。

2.2　多损失融合

本文网络融合了两类不同的损失,即分类损失以及对比损失。这两类损失分别从不同角度对网络训练进行监督,起到了良好的互补作用。

2.2.1　分类损失

本文网络的每个分支扩展为一个分类网络。具体来说,将上述提取的图像特征输入到一个包含N个神经元的全连接层,其中N为图像类别数目。每个分支采用常用的交叉熵作为损失函数,其定义为

$$L_{\mathrm{CE}}=-\frac{1}{M}\sum_{i=1}^{M}\sum_{k=1}^{N}y_{ik}\log p_{ik} \tag{1}$$

式中,M为样本个数,p_{ik}代表第i个训练样本属于第$k(k=1, 2, \cdots, N)$类的预测概率;而y_{ik}为真实概率,若第i个训练样本属于第k类,则$y_{ik}=1$;若第i个训练样本不属于第k类,则$y_{ik}=0$。

2.2.2　对比损失

本文所提出孪生网络输入为两幅图像,为了使来自同一类的两个图像特征之间的距离尽可能小,而来自不同类的两个图像特征之间的距离尽可能大,采用对比损失作为孪生神经网络的损失函数,具

体定义为

$$L_{CT}=\frac{1}{2M}\sum_{i=1}^{M}yd^2+(1-y)\times\max(m-d,0)^2 \quad (2)$$

式中，M为样本个数，$y=1$代表两幅输入图像来自同一类，即正样本；$y=0$代表两幅输入图像来自不同类，即负样本；超参数m为人工设定阈值，d为两个输入图像特征$\boldsymbol{F}_1$和$\boldsymbol{F}_2$之间的欧氏距离，即

$$d=\left\|\boldsymbol{F}_1-\boldsymbol{F}_2\right\| \quad (3)$$

本文比较了采用上述网络分支中不同层作为图像特征时的性能。实验结果表明，当选取全连接层作为输入图像特征时效果最好。

2.2.3　融合两类损失

网络训练时同时优化交叉熵损失以及对比损失。具体来说，网络损失函数定义为

$$L_{total}=L_{CE_1}+L_{CE_2}+\lambda\times L_{CT} \quad (4)$$

式中，L_{CE_1}和L_{CE_2}分别代表两个分支的分类损失，L_{CT}为对比损失，参数λ为权重，通过调节该参数可以实现两类不同损失之间的权衡。

网络训练好后，可以将测试图像输入网络的任意一个分支进行分类。

3　实　验

本节介绍实验数据集以及实现细节，通过消融实验验证所提出网络中各个模块的有效性，并将本文方法与现有中国书法字体和风格分类方法进行比较。

3.1　数据集

由于目前没有公开的用于中国书法字体和风格分类的数据集，本文采用自建的数据集，每个数据集的信息如表1所示。

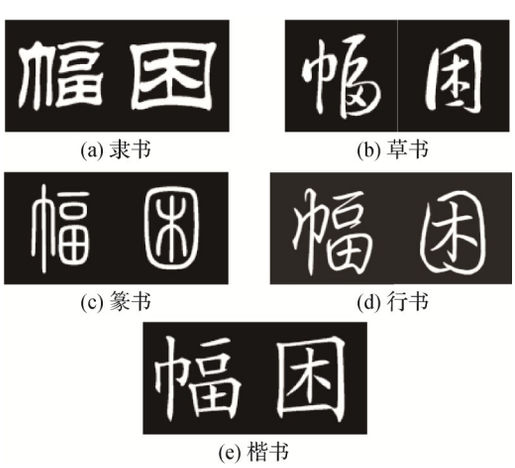
表1　本文自建数据集信息

Table 1　Details of the datasets built in our study

数据集	每类图像数/幅	类别数目/个
CNCalliFont	6 000	5
CNCalliNoisyFont	12 000	5
CNCalliStyle	3 000	4
CNCalliNoisyStyle	6 000	4

1）CNCalliFont（Chinese calligraphy font）数据集。该数据集为字体数据集，包含5种中国书法字体：隶书，草书，篆书，行书和楷书。每类字体的图像数目为6 000幅，并且混合了多种书法风格。其中，隶书字体由4位隶书书法家赵孟頫、席夔、黄葆戊和刘炳森的隶书组成，每位书法家的样本数目为1 500幅；草书字体由4位草书名家智永、怀素、赵孟頫和孙过庭的草书组成，每位书法家的样本数目为1 500幅；篆书字体由4位篆书书法家李阳冰、赵孟頫、吴叡和邓石如的篆书组成，每位书法家的样本数目为1 500幅；行书字体由4位行书大家欧阳修、赵孟頫、文征明和启功的行书组成，每位书法家的样本数目为1 500幅；楷书字体由楷书四大家欧阳询、颜真卿、柳公权和赵孟頫的楷书组成，每位书法家的样本数目为1 500幅。

因此，该数据集共包含30 000幅图像，图像尺寸为96×96像素。所有图像均为灰度图，图像格式为JPEG。图4给出了该数据集中不同字体的示例。



(a) 隶书　(b) 草书　(c) 篆书　(d) 行书　(e) 楷书

图4　CNCalliFont数据集上来自不同类的图像示例

Fig. 4　Sample images from each category on CNCalliFont dataset ((a) clerical; (b) cursive; (c) seal; (d) semi-cursive; (e) standard)

2）CNCalliNoisyFont数据集。为了使数据集中的图像更接近现实世界中的书法图像，对上述CNCalliFont数据集进行扩展，即针对每幅图像添加均值为0、方差为0.01的高斯噪声，具体如图5所示。故每类字体图像数目为12 000，扩展后的数据集命名为CNCalliNoisyFont。

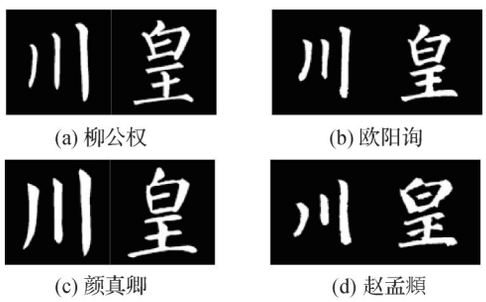
3）CNCalliStyle数据集。CNCalliStyle为风格数据集，本文主要考虑楷书四大家欧阳询、颜真卿、柳公权和赵孟頫的书法风格。每种风格3 000幅图像，因此该数据集共包含12 000幅图像。图像尺寸为96×96像素。所有图像均为灰度图，图像格式为

JPEG。图6中给出该数据集上来自4种不同风格的图像示例。



图5　CNCalliNoisyFont数据集上添加高斯噪声的图像示例

Fig. 5　Sample images contaminated with Gaussian noise from CNCalliNoisyFont dataset



(a) 柳公权　(b) 欧阳询　(c) 颜真卿　(d) 赵孟頫

图6　CNCalliStyle数据集上来自不同书法家的图像示例

Fig. 6　Sample images from each calligrapher on CNCalliStyle dataset ((a) Liu Gongquan; (b) Ouyang Xun; (c) Yan Zhenqing; (d) Zhao Mengfu)

4)CNCalliNoisyStyle数据集。对CNCalliStyle数据集进行扩展,将每幅图像添加均值为0、方差为0.01的高斯噪声,如图7所示。因此该数据集中每种风格有6 000幅图像,扩展后数据集命名为CNCalliNoisyStyle。



图7　CNCalliNoisyStyle数据集上添加高斯噪声的图像示例

Fig. 7　Sample images contaminated with Gaussian noise from CNCalliNoisyStyle dataset

3.2　评价指标

在中国书法字体数据集CNCalliFont与CNCalliNoisyFont上分别随机采用其中两位书法家的样本作为训练集,另外一位书法家的样本作为验证集;一位书法家的样本作为测试集。其中训练集用于学习网络中的参数,验证集用于调试网络中的超参数,最后将在验证集上获得最好性能的超参数应用于测试集。在训练集和测试集上采用不同书法家的样本,可以有效验证本文提出方法的拓展性。在中国书法风格数据集CNCalliStyle与CNCalliNoisyStyle上,分别按6∶2∶2划分训练集、验证集和测试集。本文将数据集进行随机10次划分,并且利用平均分类准确率作为评价指标。

3.3　实现细节

如2.1节所示,本文所提出孪生网络每个分支中包括卷积层、池化层等。表2给出了具体每一层的参数,例如卷积核大小、个数以及步长等。这些参数的取值主要通过在验证集上的实验而得,具体来说,比较了参数取不同值时网络在验证集上的性能,最终选取性能最优的一组参数。其中,对于卷积层,$W \times H \times S$表示卷积核的宽度为W,高度为H,卷积核的个数为S。对于池化层,池化窗口大小用$W' \times H'$表示。针对网络分支中的其他层,括号中给出了神经元个数。如果是针对中国书法字体分类这一任务,则最后一层有5个神经元;如果是针对中国书法风格分类这一任务,最后一层有4个神经元。

表2　本文网络中每一层具体的参数

Table 2　The parameters in each stream of the proposed network

网络层	参数	步长
Convolutional layer	5 × 5 × 32	1
MaxPooling+BatchNormalization+ReLU	3 × 3	2
Convolutional layer	5 × 5 × 32	1
MaxPooling+BatchNormalization+ReLU	3 × 3	2
Convolutional layer	5 × 5 × 64	1
MaxPooling+BatchNormalization+ReLU	3 × 3	2
Convolutional layer	5 × 5 × 128	1
MaxPooling+BatchNormalization+ReLU	3 × 3	2
Global average pooling layer	(128)	–
Haar wavelet decomposition	(128)	–
Fully-connected layer	(4/5)	–

注:参数步长中括号数据代表神经元个数。

网络训练时采用随机梯度下降算法,其中,学习率初始设置为0.000 1,当验证集上的损失不在下降时,将学习率降低为原来的1/10。动量设置为0.9,minibatch的大小为64。在验证集上比较了式(2)中参数m取不同值时的网络性能。实验结果表明,在CNCalliFont和CNCalliNoisyFont数据集上m设置为1.5,在CNCalliStyle和CNCalliNoisyStyle数据集上m设置为0.4时,取得了最高的准确率。式(4)中的权

重λ在4个数据集上均设置为0.5时，在验证集上取得了最高的准确率，因此在后续实验中将λ设置为0.5。网络共训练100个epoch。

3.4 消融实验

为了验证本文网络中各个模块的有效性，进行了消融实验，以每个分支上的卷积神经网络作为基准网络，网络训练时仅采用交叉熵损失。表3中对比了如下几个网络的性能：1）基准网络；2）基准网络 + Haar小波分解；3）基准网络 + 对比损失；4）基准网络 + Haar小波分解 + 对比损失，即本文网络。

对比表3中的1）和2）可以看出，在基准网络中加入Haar小波分解在4个数据集上均提高了分类准确率，尤其是在CNCalliStyle和CNCalliNoisyStyle数据集上，嵌入Haar小波分解相较基准网络的分类准确率分别提高了16.54%与10.79%，验证了在网络中嵌入小波分解的有效性。从表3中1）和3）的对比可以看出，融合了对比损失后4个数据集上的准确率分别提高了2.44%、4.99%、17.42%和20.58%，证明了两类损失起到了良好的互补作用。网络4）即本文网络融合了交叉熵损失与对比损失，并且在网络每个分支上嵌入Haar小波分解，在两个数据集上的性能较基准网络均得到了较大的提升。

表3 四个数据集上的消融实验结果分类准确率

Table 3 Accuracy results of ablation experiments on four datasets

/%

网络	CNCalliFont	CNCalliNoisyFont	CNCalliStyle	CNCalliNoisyStyle
1）	95.01	87.88	72.60	69.27
2）	97.17	90.01	89.14	80.06
3）	97.45	92.87	90.02	89.85
4）	**99.90**	**94.09**	**99.38**	**93.28**

注：加粗字体表示各列最优结果。

3.5 与现有方法比较

表4展示了本文方法与现有中国书法字体和风格分类方法的比较结果。具体来说，一方面将本文方法与传统基于手工特征的方向梯度直方图（histogram of oriented gradient，HOG）（Omid-Zohoor等，2018）、局部二值模式（local binary pattern，LBP）（Liu等，2016）以及Gabor（Zhang等，2013）的方法进行比较。针对所提取的特征，比较了采用不同分类器时的性能，实验结果表明，当采用SVM分类器时上述3种手工特征均取得了最高的分类性能，故本文实验中采用SVM分类器。另一方面，将本文方法与基于卷积神经网络的书法字体和风格分类方法（Dai等，2018；Zhang等，2019，2021；Liu等，2021）进行比较，由于本文解决的中国书法字体与风格分类问题本身也是一个分类问题，将其与目前常用于图像分类的经典卷积神经网络模型AlexNet（Krizhevsky等，2012）、VGG-16（Visual Geometry Group）（Simonyan和Zisserman，2015）、ResNet-50（residual neural network）（He等，2016）以及Xception（Chollet，2017）进行了比较。此外，将本文方法与两种基于ViT（vision transformer）的图像分类方法（Dosovitskiy等，2021；Chen等，2021）进行了比较。

从表4可以得出如下结论：1）所有方法在CNCalliFont和CNCalliNoisyFont数据集上的性能优于在CNCalliStyle和CNCalliNoisyStyle数据集上的性能，主要原因在于相比较书法字体而言，不同书法风格之间的差异非常细微，具有更大的挑战性。2）在CNCalliFont数据集上，基于卷积神经网络以及基于ViT的方法均优于传统基于手工特征的方法。在CNCalliNoisyFont数据集上，所有方法的性能较在CNCalliFont数据集上均有所下降。虽然训练集和测试集上样本来自不同的书法家，本文方法在CNCalliFont和CNCalliNoisyFont数据集上分别取得了99.90%以及94.09%的准确率，说明其具有良好的拓展性。3）传统基于手工特征的方法在CNCalliStyle数据集上的性能大幅降低。此外，4个经典卷积神经网络模型以及基于ViT的方法在CNCalliStyle数据集上的分类准确率同样大幅降低，说明这些方法对于不同书法风格之间的细微差别具有较低的区分度。相比之下，Dai等人（2018）、Zhang等人

(2019, 2021)和Liu等人(2021)方法在CNCalliStyle数据集上的性能均具有小幅下降。具体来说,这4种方法的准确率分别为98.01%、98.38%、97.63%和98.08%,较其在CNCalliFont数据集上的准确率分别下降了1.00%、0.74%、1.40%和1.34%。在CNCalliNoisyStyle数据集上,所有方法的准确率均低于在CNCalliStyle上的准确率。4)本文提出的融合多损失的孪生卷积神经网络在4个数据集上分别取得了99.90%、94.09%、99.38%和93.28%的准确率,高于对比方法。

表4　本文方法与现有方法在不同数据集上准确率性能比较

Table 4　Accuracy comparison between proposed approach and other methods in the literature on different datasets

/%

类别	方法	数据集			
		CNCalliFont	CNCalliNoisyFont	CNCalliStyle	CNCalliNoisyStyle
基于传统手工特征的方法	Omid-Zohoor等人(2018)	97.04	79.92	78.87	69.17
	Liu等人(2016)	96.89	79.01	75.87	70.12
	Zhang等人(2013)	97.00	80.08	72.37	70.07
基于ViT的方法	Dosovitskiy等人(2021)	98.09	87.19	89.91	82.03
	Chen等人(2021)	97.89	84.31	87.23	81.12
基于卷积神经网络的方法	Krizhevsky等人(2012)	99.01	83.96	75.50	70.12
	Simonyan和Zisserman(2015)	99.12	85.91	84.87	79.23
	He等人(2016)	98.75	87.29	74.50	68.12
	Chollet(2017)	97.89	87.03	69.88	62.12
	Dai等人(2018)	99.01	88.92	98.01	88.57
	Zhang等人(2019)	99.12	91.29	98.38	86.81
	Zhang等人(2021)	99.03	91.21	97.63	88.30
	Liu等人(2021)	99.42	91.02	98.08	89.08
	本文	**99.90**	**94.09**	**99.38**	**93.28**

注:加粗字体表示各列最优结果。

3.6　讨论

对其他工作在表4中CNCalliFont数据集和CNCalliNoisyFont数据集上的结果进一步分析可知,在中国书法字体分类任务中,易将楷体和行书两类字体混淆,主要原因在于行书是在楷书的基础上发展起源的,仅在笔画上增加了节奏感,使笔画活跃了。以CNCalliFont数据集为例,图8中给出了其他工作分类错误的例子,图中的行书字体均被误分为楷书这一类。



图8　其他工作在CNCalliFont数据集上分类错误样本示例

Fig. 8　Sample images misclassified by the other methods on CNCalliFont dataset

对其他工作在表4中CNCalliStyle数据集和CNCalliNoisyStyle数据集上的结果进一步分析可知,在中国书法风格分类任务中,易将柳公权和颜真卿两位书法家的书法风格混淆。这是因为柳公权书法曾取经于颜真卿书法,二者具有较高的相似性。图9中给出了其他工作在CNCalliStyle数据集上分类错误的例子,其中图9(a)中的样本被误分为颜真卿书法风格,而图9(b)中的样本被误分为柳公权书法风格。

从上述分析可知,其他工作难以区分不同类别图像之间的细微差异,针对这一问题,本文提出的融合多损失的孪生卷积神经网络模型在训练时由两种不同的损失共同监督,除了常用的分类损失,采用对



(a) 真实类别：柳公权　　(b) 真实类别：颜真卿

图9　其他工作在CNCalliStyle数据集上分类错误样本示例

Fig. 9　Sample images misclassified by the other methods on CNCalliStyle dataset ((a) ground-truth label: Liu Gongquan; (b) ground-truth label: Yan Zhenqing)

比损失可以使来自同一类的两个输入图像特征之间的距离尽可能小，而来自不同类的两个输入图像特征之间的距离尽可能大，并且由于网络每个分支中采用Haar小波对特征进行分解，所提取的图像特征具有更高的区分度，能有效捕捉图像之间的细微差异。为了更好地说明这一点，针对本文网络所提取的图像特征利用t-SNE(t-distributed stochastic neighbor embedding)方法(Sun等，2021)进行降维并且可视化。图10展示了在CNCalliStyle数据集上的可视化结果，可以看出，所提取的特征能较好地实现将同一风格的图像进行聚类，并且能有效地区分不同风格的图像。

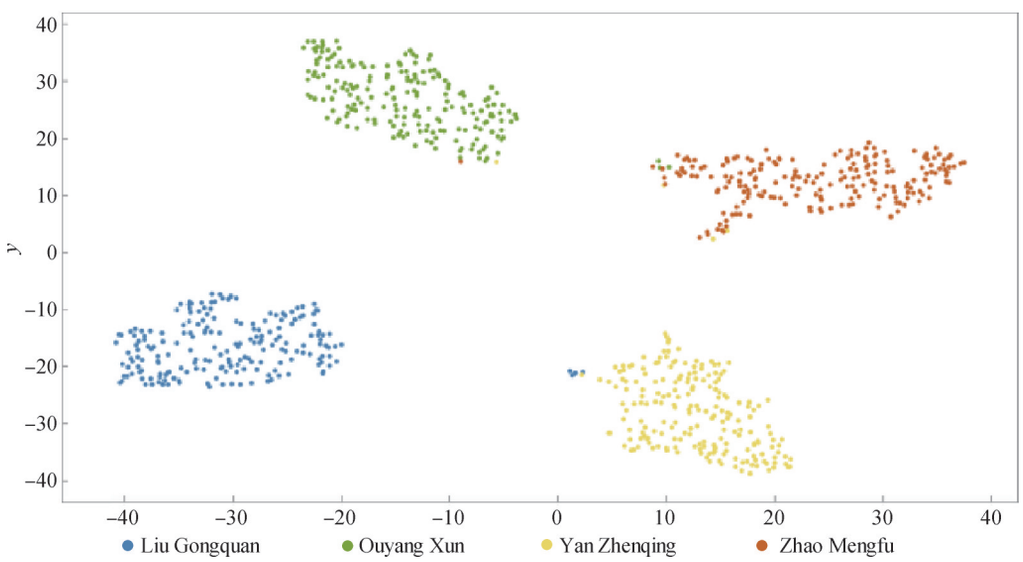


图10　CNCalliStyle数据集上的特征可视化

Fig. 10　Feature visualization on CNCalliStyle dataset

4　结　论

本文主要研究了中国书法分类问题，具体包括中国书法字体和风格分类。首先将书法字体和风格两个概念进行了区分，并且提出了一个融合多损失的孪生神经网络，能同时解决这两个问题。网络包括两个分支，每个分支利用卷积神经网络提取图像特征，并且嵌入了Haar小波分解对提取的图像特征进行分解。与传统孪生神经网络不同的是，本文将孪生网络每个分支扩展为一个分类网络。网络训练时同时优化对比损失以及分类损失，可以从不同角度对网络训练进行监督。具体来说，通过采用对比损失可以保证来自同一类的两个输入图像特征之间的距离尽可能小，而来自不同类的两个输入图像特征之间的距离尽可能大。在网络每个分支上利用交叉熵作为分类损失，可以充分利用每幅输入图像的类别信息。

为了验证本文方法的有效性，在两个中国书法字体数据集CNCalliFont、CNCalliNoisyFont和两个中国书法风格数据集CNCalliStyle、CNCalliNoisyStyle上进行了大量的实验。实验结果表明，本文网络采用的两类损失起到了良好的互补作用，在网络分支中嵌入Haar小波可以有效捕捉不同类别图像之间的细微差异，提升了网络的分类性能，尤其是在书法风格数据集上。本文方法在4个数据集上的分类准确率分别为99.90%、94.09%、99.38%以及93.28%。可以看出，在书法风格数据集CNCalliStyle以及CNCalliNoisyStyle上的性能低于在书法字体数据集CNCalliFont以及CNCalliNoisyFont上的性能，主要原因在于相比较书法字体而言，不同书法风格之间的差异非常细微，给分类带来挑战。此外，将本文方法

与现有中国书法字体和风格分类方法进行了比较，结果显示本文方法取得了最高的分类准确率。

下一步工作，尤其是在中国书法风格分类这一任务中，将继续提高本文方法的性能。本文方法目前同等地对待书法文字的各个组成部分，然而文字的起笔和收笔部分在区分不同书法风格方面起到了非常重要的作用，在后续工作中，将引入注意力机制重点关注这些部分。此外，本文采用的4个数据集包含的样本数目较小，下一步将对数据集进行扩展，并探索本文提出方法在扩展后数据集上的性能。

参考文献（References）

Cai Z D, Ying N, Guo C S, Guo R and Yang P. 2021. Research on multiperson pose estimation combined with YOLOv3 pruning model. Journal of Image and Graphics, 26(4): 837-846 (蔡哲栋, 应娜, 郭春生, 郭锐, 杨鹏. 2021. YOLOv3剪枝模型的多人姿态估计. 中国图象图形学报, 26(4): 837-846) [DOI: 10.11834/jig.200138]

Cao L C, Li H Q, Zhang Y J, Zhang L and Xu L. 2020. Hierarchical method for cataract grading based on retinal images using improved Haar wavelet. Information Fusion, 53: 196-208 [DOI: 10.1016/j.inffus.2019.06.022]

Chen C F R, Fan Q F and Panda R. 2021. CrossViT: cross-attention multi-scale vision transformer for image classification//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 347-356 [DOI: 10.1109/ICCV48922.2021.00041]

Chollet F. 2017. Xception: deep learning with depthwise separable convolutions//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 1800-1807 [DOI: 10.1109/CVPR.2017.195]

Dai F R, Tang C W and Lyu J C. 2018. Classification of calligraphy style based on convolutional neural network//Proceedings of the 25th International Conference on Neural Information Processing. Siem Reap, Cambodia: Springer: 359-370 [DOI: 10.1007/978-3-030-04212-7_31]

Ding X Q, Chen L and Wu T. 2007. Character independent font recognition on a single Chinese character. IEEE Transactions on Pattern Analysis and Machine Intelligence, 29(2): 195-204 [DOI: 10.1109/TPAMI.2007.26]

Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X H, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, Uszkoreit J and Houlsby N. 2021. An image is worth 16 × 16 words: transformers for image recognition at scale [EB/OL]. [2021-03-22]. https://arxiv.org/pdf/2010.11929.pdf

Dubey S R. 2022. A decade survey of content based image retrieval using deep learning. IEEE Transactions on Circuits and Systems for Video Technology, 32(5): 2687-2704 [DOI: 10.1109/TCSVT.2021.3080920]

He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]

Hou Q B, Zhou D Q and Feng J S. 2021. Coordinate attention for efficient mobile network design//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 13708-13717 [DOI: 10.1109/CVPR46437.2021.01350]

Huang S P, Zhong Z Y, Jin L W, Zhang S Y and Wang H B. 2018. DropRegion training of inception font network for high-performance Chinese font recognition. Pattern Recognition, 77: 395-411 [DOI: 10.1016/j.patcog.2017.10.018]

Khatami A, Nazari A, Beheshti A, Nguyen T T, Nahavandi S and Zieba J. 2020. Convolutional neural network for medical image classification using wavelet features//Proceedings of 2020 International Joint Conference on Neural Networks. Glasgow, UK: IEEE: 1-8 [DOI: 10.1109/IJCNN48605.2020.9206791]

Krizhevsky A, Sutskever I and Hinton G E. 2012. ImageNet classification with deep convolutional neural networks//Proceedings of the 25th International Conference on Neural Information Processing Systems. Lake Tahoe, Nevada: Curran Associates Inc.: 1097-1105

Liu L, Lao S Y, Fieguth P W, Guo Y L, Wang X G and Pietikäinen M. 2016. Median robust extended local binary pattern for texture classification. IEEE Transactions on Image Processing, 25(3): 1368-1381 [DOI: 10.1109/TIP.2016.2522378]

Liu L, Cheng W Y, Qiu T R, Tao C Y, Chen Q, Lu Y and Suen C Y. 2021. Multi-loss Siamese convolutional neural network for Chinese calligraphy style classification//Proceedings of the 28th International Conference on Neural Information Processing. Sanur, Indonesia: Springer: 425-432 [DOI: 10.1007/978-3-030-92310-5_49]

Liu W, Yan Q and Zhao Y Z. 2020. Densely self-guided wavelet network for image denoising//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Seattle, USA: IEEE: 1742-1750 [DOI: 10.1109/CVPRW50498.2020.00224]

Omid-Zohoor A, Young C, Ta D and Murmann B. 2018. Toward always-on mobile object detection: energy versus performance tradeoffs for embedded HOG feature extraction. IEEE Transactions on Circuits and Systems for Video Technology, 28(5): 1102-1115 [DOI: 10.1109/TCSVT.2017.2653187]

Simonyan K and Zisserman A. 2015. Very deep convolutional networks for large-scale image recognition [EB/OL]. [2021-03-22]. https://arxiv.org/pdf/1409.1556.pdf

Song W K, Lian Z H, Tang Y M and Xiao J G. 2015. Content-independent font recognition on a single Chinese character using

sparse representation//Proceedings of the 13th International Conference on Document Analysis and Recognition. Tunis, Tunisia: IEEE: 376-380 [DOI: 10.1109/ICDAR.2015.7333787]

Sun X L, Han G L, Guo L H, Yang H, Wu X T and Li Q Q. 2022. Two-stage aware attentional Siamese network for visual tracking. Pattern Recognition, 124: #108502 [DOI: 10.1016/j. patcog. 2021. 108502]

Sun Z S, Li J, Liu P, Cao W J, Yu T and Gu X F. 2021. SAR image classification using greedy hierarchical learning with unsupervised stacked CAEs. IEEE Transactions on Geoscience and Remote Sensing, 59(7): 5721-5739 [DOI: 10.1109/TGRS.2020.302319]

Tao D P, Lin X, Jin L W and Li X L. 2016. Principal component 2-D long short-term memory for font recognition on single Chinese characters. IEEE Transactions on Cybernetics, 46(3): 756-765 [DOI: 10.1109/TCYB.2015.2414920]

Wang D G, Rao W Q, Sun X, Qu Y, Liu X M and Gao L R. 2021. Siamese network with pixel-pair for hyperspectral image anomaly detection. Journal of Image and Graphics, 26(8): 1860-1870 (王德港, 饶伟强, 孙旭, 渠瀛, 刘雪梅, 高连如. 2021. 结合孪生网络和像素配对的高光谱图像异常检测. 中国图象图形学报, 26(8): 1860-1870) [DOI: 10.11834/jig.210073]

Woo S, Park J, Lee J Y and Kweon I S. 2018. CBAM: convolutional block attention module//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer: 3-19 [DOI: 10.1007/978-3-030-01234-2_1]

Xia M, Cao G, Wang G Y and Shang Y F. 2017. Remote sensing image classification based on deep learning and conditional random fields. Journal of Image and Graphics, 22(9): 1289-1301 (夏梦, 曹国, 汪光亚, 尚岩峰. 2017. 结合深度学习与条件随机场的遥感图像分类. 中国图象图形学报, 22(9): 1289-1301) [DOI: 10.11834/jig.170122]

Zhang J L, Guo M T and Fan J P. 2019. A novel CNN structure for fine-grained classification of Chinese calligraphy styles. International Journal on Document Analysis and Recognition, 22(2): 177-188 [DOI: 10.1007/s10032-019-00324-1]

Zhang J L, Yu W H, Wang Z X, Li J H and Pan Z G. 2021. Attention-enhanced CNN for Chinese calligraphy style classification//Proceedings of the 7th IEEE International Conference on Virtual Reality. Foshan, China: IEEE: 352-358 [DOI: 10.1109/ICVR51878.2021. 9483820]

Zhang S Y, Jin L W, Tao D P and Yang Z. 2013. A faster method for Chinese font recognition based on Harris corner//Proceedings of 2013 IEEE International Conference on Systems, Man, and Cybernetics. Manchester, UK: IEEE: 4271-4275 [DOI: 10.1109/SMC. 2013.728]

作者简介

程文炎，男，硕士研究生，主要研究方向为模式识别与机器学习、书法风格分类。E-mail: 2385162408@qq.com

刘丽，通信作者，女，副教授，主要研究方向为文本图像分析与识别、机器视觉和深度学习。E-mail: liuli_033@163.com

周勇，男，硕士研究生，主要研究方向为机器学习与模式识别。E-mail: 401030920001@email.ncu.edu.cn

陶承英，男，硕士研究生，一级教师，主要研究方向为书法教学。E-mail: 24112013005@bnu.edu.cn

李志刚，男，硕士研究生，主要研究方向为图像处理与深度学习。E-mail: chnlzg@126.com

邱桃荣，男，教授，主要研究方向为模式识别与人工智能、机器学习、粗糙集、粒计算、知识发现和数据挖掘、脑电信号处理和应用。E-mail: qiutaorong@ncu.edu.cn