

融合全局与局部特征的国画图像分类算法

李大湘^{1,2}, 张玥¹

(1. 西安邮电大学通信与信息工程学院, 西安 710121; 2. 电子信息现场勘验应用技术公安部重点实验室, 西安 710121)

摘要:

国画图像分类是数字化艺术作品管理系统的重要组成部分。为了获取国画的整体风格和局部笔触特征,提出一种融合全局与局部特征的国画图像分类算法。首先提出提取国画全局特征和局部特征的方法,通过学习卷积神经网络(CNN)和长短时记忆神经网络(LSTM)并行组成的网络模型,为图像返回描述其整体风格和局部笔触特点的特征。然后,提出一种并联分类策略,融合两种特征,将其送入 Softmax 分类器实现分类。在文中所构建的包含 1015 幅国画作品的数据集上的实验结果表明,所提方法在五类图像上取得很好的分类效果,分类精度达到 96%,比单一使用卷积神经网络和长短时记忆神经网络实现分类精度分别高 5%和 8%。

关键词:

图像分类; 卷积神经网络; 长短时记忆神经网络

基金项目:

陕西省科技创新创业双导师制项目(No.2019JM-604);陕西省国际合作交流项目(No.2017KW-013)

0 引言

国画,主要指的是画在绢、宣纸、帛上并加以装裱的卷轴画,是中国传统绘画形式之一,其作为中国文化艺术的瑰宝,受到了极大的重视,越来越多的书画作品出现在网络中供人欣赏,博物馆积极建设数字化国画档案,以便人们查阅,国画的数字化管理成为现阶段不可阻挡的研究趋势。因此,对数字化国画的分类具有重要研究意义。

目前对国画分类的相关算法可分为两种:基于浅层特征和基于深度学习的分类方法。文献[1]基于颜色的统计计算提取图像全局特征,并通过分隔绘画对象提取图像结构特征,实现分类。文献[2]采用蒙特卡洛凸壳特征选择模型来整合基础特征描述子,再使用支持向量机对不同艺术家的作品实现分类。文献[3]提出融合 SIFT 特征检测子和边缘检测得到国画关键区域,对关键区域视觉特征及内部领域差异性进行描述得到图像特征,采用融合不同维度特征,级联分类策略实现分类。文献[4]分析国画图像的多维低阶特征与高阶语义之间的相关性,采用支持向量机实现语义分类。文

献[5]设计了对中国画分类的通用框架,用混合二维多分辨率马尔科夫模型(MHMM)来表示不同艺术家的笔画属性实现分类。文献[6]提出在小波域内利用不同分辨率及频带的图像结构所展现的艺术风格的不同表现形式来获得国画艺术深度信息的方法,再利用 3 种不同的分类器进行分类。这类方法通常是利用人工经验或特征转换来抽取特征,对算法性能带来了限制,基于深度学习的方法克服了此问题,文献[7]提出构造国画图像的多尺度金字塔,再学习 SH-CNN 模型返回多个标签,采用自适应融合方法实现分类。文献[8]提出基于稀疏编码混合深度学习神经网络的方法从中国水墨画的笔划中提取不同画家作画风格特征从而实现分类。目前用深度学习方法对基于内容的国画图像分类研究较少。

针对上述问题,本文提出了一种融合全局和局部特征的国画图像分类算法,利用卷积神经网络提取图像全局特征,描述其整体风格,利用长短时记忆神经网络提取图像局部特征,再融合两类特征,将其送入 Softmax 分类器实现分类。

1 国画全局特征提取方法

本文采用卷积神经网络来获取国画图像的整体风格描述, CNN 是一种带有卷积结构的深层神经网络, 能够提取图像不同层次的特征, Xception 网络^[9]是一种经典的 CNN 网络, 它拥有深层次的结构和丰富的参数, 能够学习到输入图像的不同层次的特征, 因此被广泛地运用在计算机视觉领域。网络包含输入层、卷积层、池化层和全连接层和输出层。

输入层: 国画图像作为输入, I 张图像 $x_i(i=1, 2, \dots, I)$ 为 $w \times h \times n_c$ 的三维向量, w 为宽度, h 为高度, n_c 为通道数, 此处 $n_c=3$ 。

卷积层: 通过卷积操作提取输入图像的不同特征, 卷积层中每一个神经元的权值矩阵称为卷积核 ω , 大小为 $f \times f \times n_c$ 。 n 个卷积核与输入层进行卷积操作得到 n 个特征图, 默认步长为 1, 不填充, 则过程可表示为:

$(w \times h \times n_c) * (f \times f \times n_c) \rightarrow (w - f + 1 \times h - f + 1 \times n)$ 卷积后再通过激活函数得到特征图, 可用公式表示为: $h_{conv} = g(\omega * x_i + b)$, 其中, b 为卷积核的偏置值, g 为线性修正单元 (ReLU) 激活函数。

池化层: 对卷积后的特征图 h_{conv} 二次处理, 提取其局部最大值 c_j , 再将这些值拼接在一起得到新的特征图 h_{pool} 。

$$c_j = \max\{h_{conv}\}$$

$$h_{pool} = c_1 \oplus c_2 \oplus \dots \oplus c_j, j=1, 2, \dots, n$$

图 1 为最大池化操作过程。

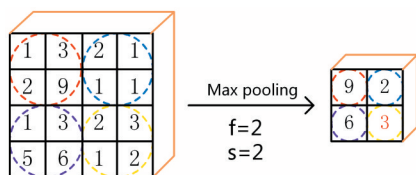


图1 最大池化

全连接层: 经池化操作后的特征输入给全连接层, 该层整合具有类别区分性的局部信息, 并输出成一个值。不难发现, 全连接层相当于特征的空间变换, 同时也将原始特征映射到语义节点上。经全连接运算过程后得到 $1 \times 1 \times 2048$ 维的特征向量 h_c 。

2 国画局部特征提取方法

为了获取国画图像的局部笔触细节特征, 引入循环神经网络 (Recurrent Neural Network, RNN) 来提取其

局部特征。传统的 RNN 存在梯度爆炸和梯度消失的问题, 不能记忆经较长时间前后的序列关系, 而长短期记忆网络能够得到长距离长时间下的序列信息, 能更好地表达出局部特征的相关性。因此我们用 LSTM 代替 RNN 实现局部特征的提取。过程为首先将原始的国画图像灰度化, 再分块, 分块后的图像作为 LSTM 网络的输入。

LSTM 改变了 RNN 的隐藏层, 增加了控制门和记忆单元 $c^{<i>}$; 控制门为输入门 $i^{<i>}$, 遗忘门 $f^{<i>}$ 和输出门 $o^{<i>}$, 分别控制是否输入、遗忘和输出当前的状态, 用 sigmoid 函数表达。LSTM 单元隐藏层的可视化如图 2 所示。

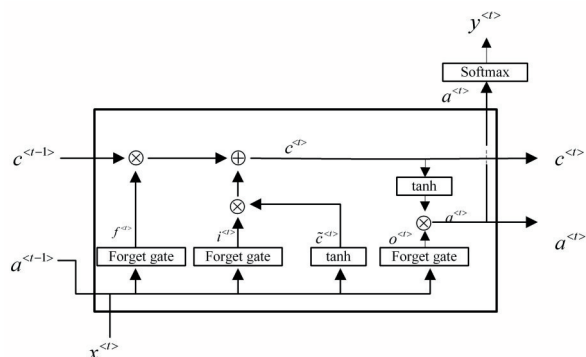


图2 LSTM内部结构图

具体公式为:

$$\tilde{c}^{<i>} = \tanh(W_{cx}x^{<i>} + W_{ca}a^{<i-1>} + b_c)$$

$$i^{<i>} = \sigma(W_{ix}x^{<i>} + W_{ia}a^{<i-1>} + b_i)$$

$$f^{<i>} = \sigma(W_{fx}x^{<i>} + W_{fa}a^{<i-1>} + b_f)$$

$$o^{<i>} = \sigma(W_{ox}x^{<i>} + W_{oa}a^{<i-1>} + b_o)$$

$$c^{<i>} = i^{<i>} * \tilde{c}^{<i>} + f^{<i>} * \tilde{c}^{<i-1>}$$

$$a^{<i>} = o^{<i>} * c^{<i>}$$

其中, $W_{cx}, W_{ca}, W_{ix}, W_{ia}, W_{fx}, W_{fa}, W_{ox}, W_{oa}$ 分别是记忆细胞、输入门、遗忘门、输出门的权重矩阵, b_c, b_i, b_f, b_o 为对应的偏置。

假设所需时间步长为 T , 则经过两层 LSTM 网络处理后, 得到最终的特征向量 h_l 。

$$h_l = LSTM(a^{<T>})$$

3 融合全局与局部特征的国画图像分类算法

本文提出的模型结构如图 3 所示, 左边是输入的

国画图像,中间是并行的 CNN 和 LSTM 网络,右边是特征融合和 Softmax 分类层。

在 CNN 分支上,最终得到了具有全局特征的卷积块 h_c ,其大小为 2048 维,包含国画图像的整体风格特征。另一个分支上,将分块的图像通过两个 LSTM 层,每个层的输出都是 2048 维,取第二个层的输出向量 h_l ,它拥有图像的局部细节特征。采用逐元素乘法合并来融合两个分支得到的特征,形成最终的特征向量 $h_{finally}$,实现特征融合。

$$h_{finally} = h_c + h_l$$

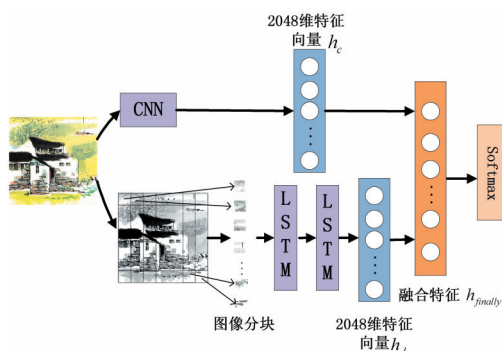


图3 混合神经网络模型

分类层使用 dropout 技术防止过拟合,假设类别数为 N ,输入为融合特征,通过一个 Softmax 函数得到分类的概率分数,使用交叉熵函数作为损失函数。

$$\hat{y}_i = \text{softmax}(Wh_{finally} + b)$$

$$L = -\sum_{i=1}^N y_i \log \hat{y}_i, i = 1, 2, \dots, N.$$

其中 \hat{y}_i 为预测图像属于类别 i 的概率分数, L 为损失函数, y_i 为指示向量,如果图像实际类别与预测类别一致时为 1,否则为 0。

4 实验结果与分析

4.1 实验数据与方法

为了验证本文提出的算法性能,我们搜集了由古树(GS)、人物(RW)、花鸟(HN)、江南水乡(JNSX)和水墨画(SM)五类国画组成的一共 1015 张图片。其中,训练集包含 711 张,测试集包含 304 张图片。

所提方法中,提取全局特征的 CNN 方法采用 Xception^[9]的网络结构,该网络规定输入图像大小为 $299 \times 299 \times 3$,网络参数采用在 ImageNet 大规模单标签数据集上预训练的参数。从网络最后一个全连接层中

提取大小为 2048 维的全局特征向量。提取局部特征的 LSTM 部分则首先将三维的彩色国画图像转换为灰度图像,再把每张图像分割成大小为 23×23 ,数量为 169 的小块,将小块按行输入 LSTM 网络中,使其成为具有 23 个时间步长,每个步长输入为 3887 的序列。然后,经过两个 LSTM 网络处理后,得到局部特征向量。最后,将两个分支的特征向量逐元素乘法合并,得到融合特征向量,并送入由 Softmax 激活函数组成的分类层中实现分类。

网络训练由两个阶段构成,在第一阶段,冻结 CNN 的所有层,使用 RMSProp 优化器对 CNN 网络最后一层和 RNN 进行训练。第二阶段,使用 Adam 优化器对所有层进行训练,学习率设置为 0.0001。

实验使用四个评价指标来衡量模型的效果:准确率(Accuracy)、精确率(Precision)、召回率(Recall)、F1 值,其计算公式如下:

$$\text{Accuracy} = \frac{\#True_positive + \#True_negative}{\#Positive + \#Negative},$$

$$\text{Precision} = \frac{\#True_positive}{\#True_positive + \#False_positive},$$

$$\text{Recall} = \frac{\#True_positive}{\#True_positive + \#False_negative},$$

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}.$$

4.2 实验结果与分析

(1)所提算法在国画图像数据上的分类效果

本节主要分析所提算法在五类国画图像上的分类效果,图 4 展示了该算法在五类国画图像分类问题上的混淆矩阵(Confusion Matrix)。从图中可以看出,五个类别的图像可以很好地被分开。

(2)所提算法与其他方法的分类效果对比

将所提算法与相关的方法在上述的 5 个指标上进行对比,表 1 中前 5 行为基于浅层特征的非深度学习算法,包括 SIFT-NDD^[3]、SIFT-Bow^[13]、CECC^[14]、CTCP^[15]、KSRC^[16]。后 4 行为基于深度学习提取特征的方法,分别是仅使用卷积神经网络模型实现分类的 CNN、将图像转换成序列输入长短期记忆网络模型的 LSTM、SH-CNN^[8]以及本文所提算法 CNN-LSTM。

这些算法所得到的实验结果中,可以发现传统方法在准确率的评价指标上最好效果为 89.96%,未达到 90%。采用结合 CNN 的深度学习方法精确度都达到 90%以上,实验表明本文并行使用 CNN 和 LSTM 网络

模型在国画数据集上的分类算法准确率最高,达到96%,超过仅使用CNN的网络约2%,超过仅使用LSTM网络模型约8%,验证了实验的有效性。

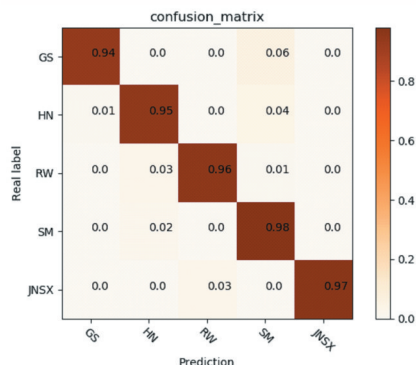


图4 所提算法在五类国画图像分类问题上的混淆矩阵

5 结语

本文提出了一种融合卷积神经网络和长短时记忆神经网络提取全局与局部特征的国画图像分类模型,图像首先通过由大规模数据集 ImageNet 预训练的

CNN 和 LSTM 并行组成的网络学习到分类模型,再从模型中提取出具有全局与局部特征的特征向量,采用逐元素相乘的方法融合两种特征向量,最后将其送入到 Softmax 分类层实现分类。在五类国画图像数据集上进行分类实验,分别与非深度学习的传统分类算法和基于深度学习的部分分类算法进行对比,实验证明,这种混合神经网络的分类效果最好,这种融合全局与局部特征的分类算法不仅能捕获图像的整体风格特征,也能把握图像局部笔触细节特征,优于其他算法得到的特征,使分类精度达到最高。

表1 相关方法在国画图像数据集上的实验对比

	方法	Accuracy	Precision	Recall	F1
非深度学习方法	SIFT-DNN ^[2]	0.899	0.837	0.878	0.896
	SIFT-Bow ^[24]	0.896	0.880	0.873	0.877
	CEC ^[24]	0.662	0.685	0.680	0.678
	CTCP ^[25]	0.667	0.710	0.683	0.672
	KSRC ^[26]	0.830	0.782	0.843	0.804
	mSRC ^[27]	0.867	0.834	0.913	0.862
深度学习方法	CNN	0.955	0.941	0.951	0.958
	LSTM	0.887	0.875	0.882	0.869
	SH-CNN ^[6]	0.958	0.951	0.949	0.960
	CNN-LSTM	0.962	0.957	0.950	0.963

参考文献:

- [1]Lee,SG.,Cha,et al. Style Classification and Visualization of Art Painting's Genre Using Self-Organizing Maps[J/OL]. Human-Centric Computing and Information Sciences, 2016, 6(7):1-11[2016-07-01]. <https://doi.org/10.1186/s13673-016-0063-4>.
- [2]SUN M.J., ZANG D., WANG Z., et al. Monte Carlo Convex Hull Model for Classification of Traditional Chinese Paintings[J/OL]. Neuro-computing, 2016, 171(1):788-797[2016-01-01]. <https://doi.org/10.1016/j.neucom.2015.08.013>.
- [3]高峰, 聂婕, 黄磊, 段凌宇, 等. 基于表现手法的国画分类方法研究[J/OL]. 计算机学报, 2017, 40(12):2871-2882[2017-12-01]. <http://dx.chinadai.cn/10.11897/SP.J.1016.2017.02871>.
- [4]陈俊杰, 杜雅娟, 李海芳. 中国画的特征提取及分类[J/OL]. 计算机工程与应用, 2008, 44(15):166-169[2008-02-21]. <http://dx.chinadai.cn/10.3778/j.issn.1002-831.2008.15.052>.
- [5]LI J., WANG J.Z.. Studing Digital Imagery of Ancient Paintings by Mixtures of Stochastic Odels[J/OL]. IEEE Transactions on Image Processing, 2004, 13(3):340-353[2004-07-01]. <https://doi.org/10.1109/TIP.2003.821349>.
- [6]盛家川. 基于小波变换的国画特征提取及分类[J/OL]. 计算机科学, 2014, 41(2):317-319[2014-07-01]. <http://dx.chinadai.cn/10.3969/j.issn.1002-137X.2014.02.069>.
- [7]Kevin Alfianto Jangtik, Trang-Thi Ho, Mei-Chen Yeh, et al. A CNN-LSTM Framework for Authorship Classification of Paintings [C/OL]//2017 IEEE International Conference on Image Processing. Beijing:China, 2017:2866-2870[2018-2-2]. <https://doi.org/10.1109/ICIP.2017.8296806>.
- [8]SUN M.J., ZHANG D., REN J.C., et al. Brushstroke Based Sparse Hybrid Convolutional Neural Networks for Author Classification of Chinese Ink-Wash Paintings[C/OL]//2015 IEEE International Conference on Image Processing (ICIP). Quebec, Canada, 2015:626-630 [2015-09-27]. <https://doi.org/10.1109/ICIP.2015.7350874>.
- [9]François Chollet.Xception:Deep Learning with Depthwise Separable Convolutions[C/OL]//2017 Computer Vision and Pattern Recognition. Honolulu, HI, USA, 2017:1251-1268[2017-0921]. <https://doi.org/10.1109/CVPR.2017.195>.
- [10]JIANG S.Q., HUANG Q.M., YE Q.X., et al. An Effective Method to Detect and Categorize Digitized Traditional Chinese Paintings

- [J/OL]. Pattern Recognition Letters, 2006, 27(7):734–746[2006–05–01]. <https://doi.org/10.1016/j.patrec.2005.10.017>.
- [11] SHENG J.C., JIANG J.M.. Recognition of Chinese Artists Via Windowed and Entropy Balanced Fusion in Classification of Their Authored Ink and Wash Paintings (IWPs)[J/OL]. Pattern Recognition, 2014, 47(2):612–622[2014–07–01]. <https://doi.org/10.1016/j.patcog.2013.08.017>.
- [12] QI H.C., Taeb A., Hughes S.M.. Visual Stylometry Using Background Selection and Wavelet–HMT–Based Fisher Information Distances for Attribution and Dating of Impressionist Paintings[J/OL]. Signal Processing, 2013, 93(3):541–553[2013–03–01]. <https://doi.org/10.1016/j.sigpro.2012.09.025>.
- [13] Csúrká G., Dance C., Fan L., et al. Visual Categorization with Bags of Keypoints[J/OL]. Proceedings of Workshop on Statistical Learning in Computer Vision, 2004:1–22[2004–01–01].
- [14] YAO L., LI J., WANG J.Z. Characterizing Elegance of Curves Computationally for Distinguishing Morissey Paintings and the Imitations[C/OL]. Proceedings of 16th IEEE International Conference on Image Processing. Piscataway, USA, 2009:73–76[2009–07–01]. <https://doi.org/10.1109/ICIP.2009.5414102>.
- [15] JIANG S., HUANG Q., YE Q., et al. An Effective Method to Detect and Categorize Digitized Traditional Chinese Paintings[J/OL]. Pattern Recognition Letters, 2007, 27(7):734–746[2007–07–01]. <https://doi.org/10.1016/j.patrec.2005.10.017>.
- [16] ZHANG L., ZHOU W.D., CHANG P.C., et al. Kernel Sparse Representation–Based Classifier[J/OL]. IEEE Transactions on Signal Processing, 2012, 60(4):1684–1695[2012–08–01]. <https://doi.org/10.1109/TSP.2011.2179539>.
- [17] SHI Y., GAO Y., YANG Y., et al. Multimodal Sparse Representation Based Classification for Lung Needle Biopsy Images[J/OL]. IEEE Transactions on Biomedical Engineering, 2013, 60(10):2675–2785[2013–07–01]. <https://doi.org/10.1109/TBME.2013.2262099>.

作者简介:

李大湘(1974–),男,博士,副教授,研究方向为刑侦图像分析

张玥(1996–),女,硕士研究生,研究方向为国画图像处理

收稿日期:2019–10–22 修稿日期:2019–11–21

Chinese Painting Image Classification Algorithm Combining Global and Local Features

LI Da-xiang^{1,2}, ZHANG Yue¹

(1. School of Communication and information Engineering, Xi'an University of Posts and Telecommunications, Xi'an 710121;

2. Key Laboratory of Electronic Information Application Technology for Scene Investigation, Ministry of Public Security, Xi'an 710121)

Abstract:

Chinese painting image classification is an important part of the digital art management system. In order to obtain the overall stroke characteristics of traditional Chinese painting, proposes a Chinese image classification algorithm combining global and local features. Firstly, proposes the method of extracting the global features and local features of Chinese painting by learning Convolutional Neural Network (CNN) and uses network model consisting of Long Short-Term Memory (LSTM) Neural Network to return the characteristics of the overall style and local stroke characteristics. Then, proposes a parallel classification strategy, which combines two characteristics. It is sent to the Soft-max classifier to realize the classification. The experimental results on the data set containing 1015 Chinese paintings constructed show that the proposed method achieves a good classification effect on five types of images, and the classification accuracy reaches 96%. Compared with single use convolutional neural network and long-term memory neural network, the classification accuracy is 5% and 8% higher, respectively.

Keywords:

Image Classification; Convolutional Neural Network (CNN); Long Short-Term Memory (LSTM) Neural Network