文章编号:1671-4598(2017)12-0259-04 DOI:10.16526/j.cnki.11—4762/tp.2017.12.067 中图分类号:TP391 文献标识码:A

基于深度残差网络的脱机手写汉字识别研究

张 帆[1],张 良[1],刘 星[1],张 宇[2]
(1.湖北大学 资源环境学院,武汉 430062;2.重庆大学 建设管理与房地产学院,重庆 400045)

摘要:手写汉字识别是模式识别与机器学习的重要研究方向和应用领域;近年来,随着深度学习理论方法的完善、新技术的层出不穷,深度神经网络在图像识别分类、图像生成等典型应用中取得了突破性的进展,其中,深度残差网络作为最新的研究成果,已成功应用于手写数字识别、图片识别分类等多个领域;将研究深度残差网络在脱机孤立手写汉字识别中的应用方法,通过改进残差学习模块的单元结构,优化深度残差网络性能,同时通过对训练集的预处理,从数据层面实现训练生成模型性能的提升,最后设计实验,验证深度残差网络、End—to—End模式在脱机手写汉字识别中的可行性,分析、总结存在的问题及今后的研究方向。

关键词:手写汉字识别;深度学习;深度残差网络;End—to—End;卷积神经网络

Recognition of Off—line Handwritten Chinese Character Based on Deep Residual Network

Zhang Fan[1], Zhang Liang[1], Liu Xing[1], Zhang Yu[2]
(1. Faculty of Resources and Environmental Science, Hubei University, Wuhan 430062, China;
2. Faculty of Construction Management and Real Estate, Chongqing University, Chongqing 400044, China)

Abstract: Handwritten Chinese character recognition is an important research direction and application field of pattern recognition and machine learning. In recent years, with the development of the theory and the new technology, deep neural network have made a breakthrough in the field of image recognition and image generation. Specialty, Deep Residual Networks as the latest method, has been successfully applied to handwritten numeral recognition, image recognition classification and other fields. In this paper, we study the application of Deep Residual Networks in off—line isolated handwritten Chinese character recognition, and optimize the performance of Deep Residual Networks by improving the unit structure of residual learning module. At the same time, we improve the model performance by preprocessing the training set. Then, the experiment is designed to verify the feasibility of the Deep Residual Networks and End—to—End mode in off—line handwritten Chinese character recognition. And finally we analyze and summarize the existing problems and future research directions.

Keywords: handwritten Chinese character recognition; deep learning; deep residual networks; end—to—end; convolutional neural network

0 引言

手写体汉字识别(handwritten Chinese character recognition, HCCR)是计算机视觉和光学字符识别(optical character recognition, OCR)领域最具挑战性的问题之一。它涉及机器学习、模式识别、数字信号处理、自然语言处理、统计学、信息论等多门学科。手写体汉字识别根据数据采集方式不同可以划分为脱机手写体汉字识别和联机手写体汉字识别两大类[1]。脱机手写汉字识别识别的对象为手写汉字的图片,通过对图片上汉字的分析处理来识别汉字;联机手写汉字识别通过鼠标、触摸屏、手写板等硬件设备实时采集书写者的手写汉字,除了采集汉字的形体特征,同时将收集书写者在书写汉字时的笔划轨迹的信息。因此,从理论上来看,由于联机状态时将收集到更多的有用信息,手写汉字的正确识别率将高于脱机状态。

在脱机手写汉字识别方面,传统的方法包括数据预处理、特征提取、分类识别三个步骤。数据预处理部分需要完成原始样本图像的去噪、归一化、数据增强等处理;特征提取完成对孤立汉字在二维像素空间中的特征提取,需要提取具备高区分度的统计特征,包括Gabor特征[2]、Gradient特征[3]等;分类识别步骤通过选用合适的分类器对提取的手写汉字特征识别分类,目前较常用分类器有欧氏距离分类器、改进的二次判别函数分类器、贝叶斯分类器及支持向量机分类器等[4]。

另一方面,随着近年来深度学习研究热潮的兴起,特别是卷积神经网络(convolutional neural network, CNN)和递归神经网络(recurrent neural network, RNN)的引入,国内外研究学者在图像识别和分类领域中取得了一系列振奋人心的研究成果。牛津大学计算机视觉组(Visual Geometry Group)和Google DeepMind团队于2014年研究的深度卷积神经网络VGGNet[5],在ImageNet数据集上实现了7.3%的top—5错误率;Google公司于2016年改进的深度卷积神经网络模型Inception[6]及其后续版本,将ImageNet数据集的top—5识别率降低至4.8%;微软研究院的Kaiming He等人提出的ResNet[7]模型,进一步的在ImageNet数据上的top—5识别率降低至3.57%。这些近年来在图像识别分类领域的先进方法与

收稿日期:2017-09-18; **修回日期**:2017-10-17。
基金项目:国家自然科学基金资助项目(41301516);区域开发与环境响应湖北省重点实验室基金(2016B003)。
作者简介:张 帆(1981-),男,湖北武汉人,博士,讲师,主要从事机器学习、语音信号处理方向的研究。

技术，也为研究脱机手写汉字识别提供了基础与借鉴。

本文将研究深度 0 残差网络（deep residual networks，DRNs）模型在脱机手写汉字识别中的应用方法，并针对手写汉字图像的特征，构建、训练基于脱机手写汉子识别的深度残差网络，测试、分析最终的实验结果。

1 深度残差网络

在传统的卷积神经网络中，随着神经网络深度的不断增加，分类精度会逐步达到饱和，继续加深神经网络的深度反而会令分类精度下降[7-9]。这种情况来源于训练过程中存在的梯度消失现象。深度残差网络模型正是为了解决深度神经网络的以上问题而提出的。

1.1 残差学习模块

DRNs 模型的核心在于残差学习模块，其基本思想为：通过在卷积神经网络单元训练过程中，保存部分原始输入信息，从而避免由于卷积层数过多引起的分类精度饱和问题；同时，残差模块（residual module）不需要学习完整的输出，只需学习输入、输出差别的部分，简化了学习目标和难度。

设 x 为输入，经过卷积层运算后输出为 $F(x,W)$，激活函数采用 Sigmoid 或 ReLU[10]，激活函数变换用 f 表示。因此，学习模块单元的最终输出 y 可定义为：

$$y = f(F(x,W_i)) \tag{1}$$

其中：W_i 表示卷积神经网络第 i 层所要学习的权重参数。

则最终输出可定义为：

$$y = F(x,W_i) + x \tag{2}$$

在以上传统残差学习模块的基础上，采用线性变换构建新的 $h(x)$ 函数，增加模型的学习性能。重新构建的 $h(x)$ 函数可描述为：

$$h(x) = a_k x \tag{3}$$

其中：k 为所有残差学习模块中，卷积层的层数。每个学习模块间，具有不同的线性变换参数需要通过训练来学得。所构建的模块最终输出 y 定义为：

$$y = F(x,W_i) + a_i x \tag{4}$$

此外，在所设计的残差学习模块中，还将采用批归一化（batch normalization，BN）处理技术[12]，加快整个训练网络的训练速度，并提高分类性能；同时，借鉴 VGG－19 的结构，采用 3×3 的卷积核大小构建，每个残差学习模块包含 2 个卷积层。完整结构如图 1 所示。

1.2 深度残差网络

深度残差网络是由残差学习模块重复堆积而成的完整神经网络，与普通卷积神经网络的最大不同之处在于学习模块的内部结构。因此，通过残差学习模块可以构建不同架构的卷积神经网络，如在文献［13］中，即将残差学习模块与 Inception 模型结构结合，搭建了 Inception－ResNet 卷积神经网络模型，在 ImageNet 数据集的分类问题上，取得了优异的性能。

由于每个残差学习模块内部包括 2 个卷积层，所设计的整个残差神经网络中共包含 18 个残差学习模块，也即 36 个卷积层（不包括接收初始输入的 7×7 卷积层）。在这 18 个卷积层中，初始输入为 x_1，第 n 层输入为 $x_n(1 \leqslant n \leqslant 18)$，第 18 层（最有一个卷积层）输出为 y。由于第 i 层输出等于第 $i+1$ 层输入，其中（$1 \leqslant i \leqslant 18$），因此有：

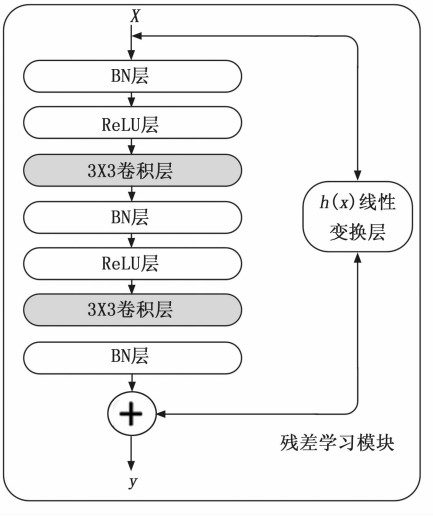

图 1 改进的的残差学习模块完整结构

$$x_{i+1} = F(x_i,W_i) + a_i x_i \tag{5}$$

每一层的输入依次传递至最后一个卷积层，则输出 y 可定义为：

$$y = \sum_{i=1}^{18}\left(\prod_{j=i+1}^{18} a_j\right)F(x_i,W_i) + \left(\prod_{i=1}^{18} a_i\right)x_i \tag{6}$$

设 ε 为损失函数，则有：

$$\frac{\partial \varepsilon}{\partial x_i} = \frac{\partial \varepsilon}{\partial y}\cdot\frac{\partial y}{\partial x_i} =$$

$$\frac{\partial \varepsilon}{\partial y}\left(\frac{\partial}{\partial x_i}\left(\left(\sum_{i=1}^{18}\left(\prod_{j=i+1}^{18} a_j\right)F(x_i,W_i)\right)\right)\right) + \prod_{i=1}^{18} a_i \tag{7}$$

由文献［11］中的分析可知，当常数项 $\prod_{i=1}^{18} a_i$ 大于 1 或小于 1 时，会出现梯度爆炸或梯度消失的问题，不利于整个神经网络的训练。也正因为如此，文献［7］和文献［11］中均采用了 $h(x) = x$ 恒等变换的方式。

构建的神经网络采用公式（3）所示的线性变换，来直接限制保留上层数据的比例。为避免梯度爆炸和梯度消失，需额外限制 a_i，将 $\left|\prod_{i=1}^{18} a_i - 1\right|$ 加入总的损失函数中迭代训练。

2 End－to－End 模式的手写汉字识别

端对端（End－to－End）模式最早起源于系统设计领域[14]。近年来，随着深度神经网络模型性能的不断提升，优秀的神经网络模型已经能够实现高维度、多参数的大规模训练，这也为 End－to－End 模式在机器学习领域的运用提供了基础：在进行训练、分类任务前，不需要进行繁琐的数据处理、特征提取等步骤，只需建立合适的深度训练模型，直接完成从最原始的输入信息训练，到最终的分类结果输出，也即 End－to－End。采用这种模式，不仅能够减少大量的数据处理工作量，提升训练效率；同时，由计算机对特征自动抽象、识别、分类，能够有效提升分类精度。

2.1 训练集预处理

CASIA－HWDB（V1.1）数据集为未经过处理的原始样本。为了提高训练模型的准确率，需对样本进行筛选处理；同时为保证后续对训练学得模型的测试，因此样本集筛选处理仅针对训练集。

为尽量保存训练集原始样本，减少构建模型过拟合风险，本文所述方法仅对训练集中 3 类样本做相应处理，如图 2 所示。

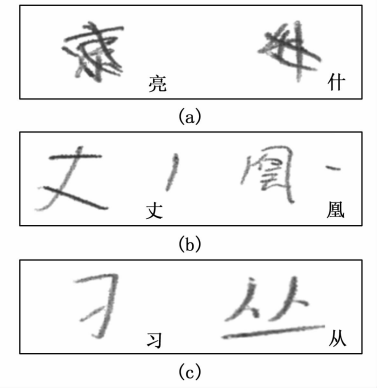

图 2　需预处理训练集样本示例图

其中，(a) 图中所示为训练集上书写错误，已在数据采集阶段被划除的手写汉字数据，此类样本直接从训练集中剔除；

(b) 类数据在手写汉字主体区域外，有额外的笔划，此类样本将剪裁图片，剔除额外笔划部分；

(c) 类数据为标记错误样本，此类样本将调整样本标签至正确分类。

另外，训练集还存在一类样本，在孤立汉字情形下无法识别属于哪类汉字，此种情形往往出现在字形相近的汉字之间。此类样本将保存原始标签信息，不予调整。

2.2　数据增强

数据增强 (Data Augmentation) 的目的在于将每个汉字的手写体样本训练集扩大，降低模型学习的过拟合风险。

因此，将借鉴文献 [15] 所述方法，首先将训练集原始样本随机翻转，按照 1：2 的比例扩充训练集样本；其次，将原始样本随机放大 0.5－2 倍，生成新的训练集样本；最后，在训练集样本中随机取 30%的样本，进行主成分分析 (Principal Components Analysis，PCA)，对主成分随机加入标准差为 0.01－0.1 的高斯扰动，增加噪声。

新的训练样本集将限定为原样本集大小的 4－6 倍，保证在训练集上，每个手写汉字类别有 1000 张以上的样本图片。

2.3　End－to－End 训练

在 End－to－End 训练过程中，不需要额外的指定提取特征，取而代之由整个深度卷积神经网络自动化提取、分类。因此，只需将准备好的数据，转换成复合神经网络的输入，并制定学习规则，即可开始模型的训练。

2.3.1　样本导入

与传统图像识别不同，手写汉字识别的训练集需要保存字体的全部信息，因此不能对图片样本剪裁，而需做拉伸、填充处理。为匹配训练集中最大的样本图片，需要将样本统一调整为 108×108 像素大小，对于小于此尺寸的样本图片，进行“0 填充”处理。

同时，将训练数据样本按照 1：10 的比例划分为训练集和和验证集，以提高最终训练生成模型的泛化能力。

2.3.2　模型评估

模型评估通过损失函数和训练结束后的泛化测试来实现。

损失函数包括常规的 Softmax 回归损失、辅助分类损失和正则化损失。总的损失将三者求和，用于梯度下降训练，以及模型性能的评估。

分类精度测试在训练完成后，在训练集和测试集上分别来计算模型对手写汉字的正确识别率。

3　实验与分析

如前所述，本研究采用谷歌公司的开源机器学习框架 Tensorflow (V1.2) 的 Slim 模块完成深度残差网络训练模型的搭建，同时使用 GPU 计算加速训练过程，硬件运行环境选择为 TitanX 显卡、16G 内存。

训练结束后生成的各类损失示意图如图 3 所示。从图中可见，对于迭代训练了 30 万步后的生成的模型，Softmax 回归损失、辅助分类损失已逐步趋向于稳定；对于正则化损失仍有下降趋势。由于正则化损失在总的损失中所占比例较小，因此对总损失的影响较小，模型的分类精度趋向于稳定。

同时，由图 3 (d) 总损失示意图可知，损失随迭代步数的变化并不平滑，而是始终小幅震荡。因此，对于某点（如迭代 30 万步时）的损失值的计算，取其邻域内的平均值，记为：

$$Loss_s = \frac{\left(\sum_{n=s-i}^{s+i} Loss_n\right)}{(2i)} \qquad (i \in N) \tag{8}$$

其中：$Loss_n$ 为实际训练过程中记录的第 n 步的总损失。

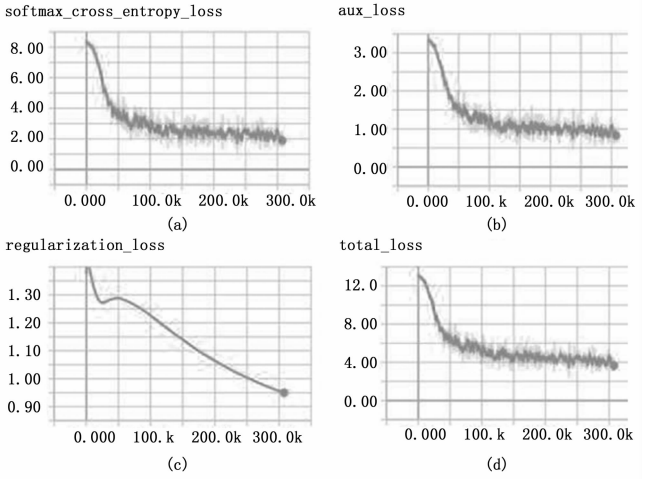

图 3　模型训练损失示意图

与损失的计算方法类似，精度同样存在小幅波动的情况，因此用某点领域内的平均值来表征该点的精度，记为：

$$Ac_s = \frac{\left(\sum_{n=s-i}^{s+i} Ac_n\right)}{(2i)} \qquad (i \in N) \tag{9}$$

其中：Ac_n 为实际评测得到的第 n 步的精度。

对比所设计的改进深度残差网络模型与传统的 Inception_4 模型、ResNet 模型，以及文献 [4]、文献 [17] 中的方法进行对比实验。实验的对比结果如表 1 所示。

从表 1 的统计结果可以看出，在采用了新的改进深度残差网络进行训练的手写汉字分类模型，在测试集上 top－1 精度与 top－5 精度均有一定的提升，相比于传统的深度残差网络 ResNet，top－1 错误率降低了 24.63，top－5 错误率降低

了 28.50%。

表 1 不同训练模型的分类精度统计

训练模型	模型评估(s=600,000)	
	top-1	top-5
Inception	92.37%	98.01%
ResNet	92.45%	98.07%
仿射传播聚类[4]	88.53%	—
DLQDF[8]	92.08%	—
改进残差模型	94.31%	98.62%

3.1 实验结果分析

从实验的结果可以看出，基于所设计的深度残差网络在脱机孤立手写汉字识别中，表现出较为良好的识别精度。通过对训练集的预处理与数据增强操作，能够有效提高模型在训练集和测试集上的分类精度。另一方面，相比训练集上的分类精度，在测试集上精度下降较为明显，说明虽然对于训练集的预处理与数据增加等操作，能够总体上提升模型性能，但对于模型的泛化能力，依然产生了较为明显的影响。同时，经过改进的深度残差网络训练的生成模型，相比于原始的残差学习模型有着较为明显的性能提升；同时，在于近年来部分其它相关文献中所采用的手写汉字识别方法，也有一定的优势。

4 结语

实验基于中科院自动化研究所模式识别国家重点实验室发布的手写汉字数据集 CASIA-HWDB (V1.1)，研究孤立手写汉字的脱机识别方法。

以 ResNet 为代表的深度残差网络在传统的图像识别分类领域，取得良好的分类精度。结合 ResNet 的残差学习模型和 VGG-19 模型的卷积核结构，提出了一种新的残差学习模块，以及由其构建的深度神经网络，并将其应用于脱机手写汉字的识别。从实验结果来看，新的深度残差网络的训练生成模型有较好的分类性能；同时深度残差网络在手写汉字识别中也具有一定的研究价值与应用潜力。另一方面，对于原始数据集的预处理，也能够在一定程度上提升最终模型的分类精度。

最后，设计的手写汉字识别方法，采用 End-to-End 模式构建。End-to-End 模式在普通图像的识别、合成领域取得了优异的性能。然而，手写体汉字则具有更为明显的结构特征。所研究设计的残差神经网络，能否和一些非 End-to-End 模式下构建的手写汉字分类模型结合起来，如 GoogleNet[16]、Multi-CNN Voting[17] 等，从而再次优化模型性能，值得进一步研究。

参考文献：

[1] 金连文，钟卓耀，杨 钊，等. 深度学习在手写汉字识别中的应用综述 [J]. 自动化学报，2016，42 (8)：1125-1141.

[2] Ge Y，Huo Q，Feng Z. Offline recognition of handwritten Chinese characters using Gabor features，CDHMM modeling and MCE training [A]. ICASSP 2002：2002 IEEE International Conference on Acoustics，Speech and Signal Processing [C]. Orlando：IEEE，2002：1053-1056.

[3] Liu C L. Normalization-cooperated gradient feature extraction for handwritten character recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，2007，29 (8)：1465-1469.

[4] 杨 怡，王江晴，朱宗晓. 基于仿射传播聚类的自适应手写字符识别 [J]. 计算机应用，2015，35 (3)：807-810.

[5] Karen S，Andrew Z. Very deep convolutional networks for Large-Scale image recognition [A]. ICLR 2015：International Conference on Learning Representations 2015 [C]. San Diego：arXiv，2015：1409-1556.

[6] Szegedy C，Liu W，Jia Y Q. Going deeper with convolutions [A] CVPR 2015：2015 IEEE Conference on Computer Vision and Pattern Recognition [C]. Boston：IEEE，2015：1-9.

[7] He K M，Zhang X Y，REN S Q，et al. Deep residual learning for image recognition [A]. CVPR 2016：IEEE Conference on Computer Vision and Pattern Recognition [C]. Las Vegas：IEEE，2016：770-778.

[8] Liu C L，Yin F，Wang D H，et al. Online and offline handwritten Chinese character recognition：Benchmarking on new databases [J]. Pattern Recognition，2013，46 (1)：155-162.

[9] He K，Sun J. Convolutional neural networks at constrained time cost [A]. CVPR 2015：2015 IEEE Conference on Computer Vision and Pattern Recognition [C]. Boston：IEEE，2015：5353-5360.

[10] Nair V，Hinton G E. Rectified linear units improve restricted boltzmann machines [A]. ICML 2010：The 27th International Conference on Machine Learning [C]. Haifa：Omnipress，2010：807-814.

[11] He K M，Zhang X Y，Ren S Q，et al. Identity mappings in deep residual networks [A]. ECCV 2016：The 14th European Conference on Computer Vision [C]. Amsterdam：Springer，2016：630-645.

[12] Loffe S，Szegedy C. Batch normalization：Accelerating deep network training by reducing internal covariate shift [A]. ICML 2015：The 32th International Conference on Machine Learning [C]. Lille：JMLR. org，2015：448-456.

[13] Szegedy C，Loffe S，Vanhoucke V，et al. Inception-v4，Inception-ResNet and the impact of residual connections on learning [A]. AAAI-17：The Thirty-First AAAI Conference on Artificial Intelligence [C]. San Francisco：AAAI Press，2017：4278-4284.

[14] Saltzer J H，Reed D P，Clark D D. End-to-end arguments in system design [J]. ACM Transactions on Computer Systems，1984，2 (4)：277-288.

[15] Krizhevsky A，Sutskever I，Hinton G E. ImageNet classification with deep convolutional neural networks [A]. NIPS 2012：The Twenty-sixth Annual Conference on Neural Information Processing Systems [C]. Lake Tahoe：ACM，2012：84-90.

[16] Zhong Z Y，Jin L W，Xie Z C. High performance offline handwritten Chinese character recognition using GoogLeNet and directional feature maps [A]. ICDAR 2015：The 13th International Conference on Document Analysis and Recognition [C]. Tunis：IEEE，2015：846-850.

[17] Chen L，Wang S，Fan W，et al. Beyond human recognition：a CNN-based framework for handwritten character recognition [A]. IAPR 2015：The 3rd IAPR Asian Conference on Pattern Recognition [C]. Kuala Lumpur，Malaysia：IEEE，2015：695-699.