



## 基于改进残差学习的东巴象形文字识别

骆彦龙, 毕晓君, 吴立成, 李霞丽

引用本文:

骆彦龙, 毕晓君, 吴立成, 等. 基于改进残差学习的东巴象形文字识别[J]. *智能系统学报*, 2022, 17(1): 79–87.

LUO Yanlong, BI Xiaojun, WU Licheng, et al. Dongba pictographs recognition based on improved residual learning[J]. *CAAI Transactions on Intelligent Systems*, 2022, 17(1): 79–87.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202112009>

## 您可能感兴趣的其他文章

### 基于数据增广和复制的中文语法错误纠正方法

Chinese grammatical error correction method based on data augmentation and copy mechanism  
*智能系统学报*. 2020, 15(1): 99–106 <https://dx.doi.org/10.11992/tis.202001014>

### 卷积神经网络的贴片电阻识别应用

Chip resistance recognition based on convolution neural network  
*智能系统学报*. 2019, 14(2): 263–272 <https://dx.doi.org/10.11992/tis.201710005>

### 基于Capsule网络的甲骨文构件识别方法

Recognition of Oracle Radical based on the Capsule network  
*智能系统学报*. 2020, 15(2): 243–254 <https://dx.doi.org/10.11992/tis.201904069>

### 一种基于2D时空信息提取的行为识别算法

A behavioral recognition algorithm based on 2D spatiotemporal information extraction  
*智能系统学报*. 2020, 15(5): 900–909 <https://dx.doi.org/10.11992/tis.201906054>

### 一种改进的深度学习的路标交通标识识别算法

An improved deep learning algorithm for road traffic identification  
*智能系统学报*. 2020, 15(6): 1121–1130 <https://dx.doi.org/10.11992/tis.201811009>

### 基于增强AlexNet的音乐流派识别研究

Music genre recognition research based on enhanced AlexNet  
*智能系统学报*. 2020, 15(4): 750–757 <https://dx.doi.org/10.11992/tis.201909032>

 微信公众平台



关注微信公众号, 获取更多资讯信息

DOI: 10.11992/tis.202112009

网络出版地址: <http://kns.cnki.net/kcms/detail/23.1538.TP.20211224.0916.002.html>.

## 基于改进残差学习的东巴象形文字识别

骆彦龙<sup>1</sup>, 毕晓君<sup>2</sup>, 吴立成<sup>2</sup>, 李霞丽<sup>2</sup>

(1. 哈尔滨工程大学信息与通信工程学院, 黑龙江哈尔滨 150001; 2. 中央民族大学信息工程学院, 北京 100081)

**摘要:** 基于深度学习模型的东巴象形文字识别效果明显优于传统算法, 但目前仍存在识别字数少、识别准确率低等问题。为此本文建立了包含 1387 个东巴象形文字、图片总量达到 22 万余张的数据集, 大幅度增加了可识别字数, 并辅助提高了东巴象形文字的识别准确率。同时, 本文根据东巴象形文字相似度高、手写随意性大的特点, 选择 ResNet 模型作为改进的网络结构, 设计了残差跳跃连接方式和卷积层的数量, 并通过加入最大池化层实现了下采样的改进。实验结果表明, 在本文建立的东巴象形文字数据集上, 改进的 ResNet 模型实现了东巴象形文字识别字数多且识别准确率高的最好效果, 识别准确率可达到 98.65%。

**关键词:** 深度学习; 东巴象形文字; 图像识别; 数据集建立; ResNet 模型; 残差跳跃连接; 下采样改进; 识别准确率  
**中图分类号:** TP18   **文献标志码:** A   **文章编号:** 1673-4785(2022)01-0079-09

中文引用格式: 骆彦龙, 毕晓君, 吴立成, 等. 基于改进残差学习的东巴象形文字识别 [J]. 智能系统学报, 2022, 17(1): 79-87.  
英文引用格式: LUO Yanlong, BI Xiaojun, WU Licheng, et al. Dongba pictographs recognition based on improved residual learning[J]. CAAI transactions on intelligent systems, 2022, 17(1): 79-87.

## Dongba pictographs recognition based on improved residual learning

LUO Yanlong<sup>1</sup>, BI Xiaojun<sup>2</sup>, WU Licheng<sup>2</sup>, LI Xiali<sup>2</sup>

(1. College of Information and Communication Engineering, Harbin Engineering University, Harbin 150001, China; 2. School of Information Engineering, Minzu University of China, Beijing 100081, China)

**Abstract:** Dongba pictographs recognition based on deep learning model has better recognition effect than that of traditional algorithms. However, these methods have disadvantages such as small number of recognizable Dongba pictographs and low recognition accuracy. Aiming at these problems, in this study, we build a novel dataset of Dongba pictographs that contains 1387 Dongba pictographs and more than 220 thousand images. Therefore, the number of recognizable Dongba pictographs is greatly increased and the Dongba pictographs recognition accuracy is improved. Since Dongba pictographs are characterized by high similarity and random writing, ResNet is adopted as an improved network structure. Moreover, we design a residual shortcut connection and the number of convolutional layers and introduce the max-pooling into the ResNet to improve down-sampling. The experimental results demonstrate that the improved ResNet model can recognize more Dongba characters, and has achieved the highest recognition accuracy 98.65% in our dataset.

**Keywords:** deep learning; Dongba pictographs; image recognition; build dataset; ResNet model; residual shortcut connection; improved down-sampling; recognition accuracy

东巴象形文字由纳西族祖先创造, 至今已有两千多年的历史。2003 年, 东巴古籍文献被联合国教科文组织列为“世界记忆遗产”名录, 成为人类共同拥有的宝贵财富。东巴象形文字的识别一

直是研究的热点和重点。早期的东巴象形文字识别研究一般采用传统算法提取东巴象形文字特征进行识别, 关键步骤一般包括图像去噪、特征提取和分类器识别 3 个步骤。常用的图像去噪方法有中值去噪、自适应去噪和小波去噪<sup>[1]</sup>; 在特征提取方面, 方向元素、粗网格<sup>[2-3]</sup>等统计特征比分析东巴象形文字的结构、笔画等结构特征取得的效

收稿日期: 2021-12-05. 网络出版日期: 2021-12-24.

基金项目: 国家社科基金重大项目 (20&ZD279).

通信作者: 毕晓君. E-mail: [bixiaojun@hrbeu.edu.cn](mailto:bixiaojun@hrbeu.edu.cn).

果更好;常用的分类器模型包括支持向量机<sup>[4]</sup>、随机森林<sup>[5]</sup>等。代表性的研究有 2017 年徐小力等采用拓扑特征法和投影法相结合的特征提取方法,取得了 84.4% 的识别准确率<sup>[6]</sup>。2019 年杨玉婷等通过结合东巴象形文字的结构和形态,提出了基于网格分辨率的东巴象形文字相似度测量算法,能够检索和识别不同形状的东巴象形文字<sup>[7]</sup>。上述研究虽然取得了一定的成果,但实现过程复杂且效率较低,算法的识别准确率有待提高。直到 2019 年,随着人工智能技术的发展,国内外开始出现基于深度学习的东巴文识别文章,2019 年张泽晖建立了包含 30 592 张图片的东巴象形文字数据集,设计了孪生网络并协同进行文字语义识别,对 956 个东巴象形文字测试,取得了 85.6% 识别准确率<sup>[8]</sup>;同年, Wu<sup>[9]</sup> 在训练集图像 3 800 张,测试集图像 200 张的条件下,使用 VGGNet 取得了 95.8% 的识别准确率;2021 年谢裕睿等提出了基于 ResNet 网络的东巴象形文字识别方法,建立了包含 536 个东巴象形文字的数据集,并对 94 个东巴象形文字测试,取得了 93.58% 的识别准确率<sup>[10]</sup>。

以上研究对东巴象形文字识别做出较大的贡献,但目前还存在一些问题:1) 东巴象形文字大多包含多个异体字,且在东巴经典中广泛存在;但现有的数据集都没有涉及异体字,导致东巴经典中的大量文字不能识别;2) 现有的东巴象形文字数据集规模较小,影响了算法识别的准确率;3) 所采用的深度学习模型较为初级,无法适应东巴象形文字的随机性和手写不确定性,识别准确率有待进一步提高。

为了解决上述问题,本文主要做了以下两个方面的工作:

1) 根据东巴象形文字字典<sup>[11-12]</sup>,采用人工仿写的方法建立了 1 387 个东巴象形文字(包括异体字)、图像规模达 22 万余张的东巴象形文字数据集,有效解决了异体字问题,大幅增加了可识别东巴象形文字的数量,并有效扩充了数据集的规模。2) 根据东巴象形文字的图像特点,选择应用效果最好的 ResNet 模型作为改进的网络结构,设计了残差跳跃连接方式和卷积层的数量,并通过加入最大池化层实现了下采样的改进,有效提高了算法识别的准确率。

## 1 东巴象形文字数据集建立

### 1.1 东巴象形文字图像获取

深度学习模型能够取得较好的识别效果,其前提是需要大量带标注的训练数据。为此本文首

先研究如何建立大规模的东巴象形文字数据集,来保证识别的东巴象形文字更多,并可辅助提高算法识别的准确率。

东巴象形文字的特点可总结如下。1) 内容广泛、字数多。按照属性可分为天文、地理、建筑等十八大类,共有 2 000 余字(包括异体字)<sup>[11]</sup>。2) 相似度高。结构相似的东巴象形文字因其细节部分不同,其字义亦不同。3) 书写随意性较大。不同人书写的东巴象形文字都会有不规则的形变。4) 异体字多。大多数东巴象形文字都有多个异体字。

上述特点增加了东巴象形文字的识别难度,因此为了获得更好的识别效果,数据集中每个东巴象形文字大约需要 150 张图像,才能满足训练的要求。但是仅通过东巴古籍来获取远远不能达到数量的要求,常用的数据增强方法主要是几何变换<sup>[13-15]</sup>,但由于东巴文本本身象形字的图画特点,相近的形状可表达不同的含义,通过几何变换可能变成其他文字,所以这种数据增强的方法难以适用东巴象形文字。

为此本文根据东巴象形文字手写或刀刻的书写习惯,提出采用人工仿写东巴象形文字字典的方法建立大规模数据集,再通过图像预处理方法提高数据集的图片质量,这样可以保证数据集中东巴象形文字的数量足够多,既可以最大程度地增加算法可识别的东巴象形文字字数,又可以辅助提高算法识别的准确率。本文建立的东巴象形文字数据集示例如图 1 所示,其中每一行的 5 幅图片同属异体字,共有相同的释义,第一列为统一的文字释义,从中可以看出异体字之间的差别较大。

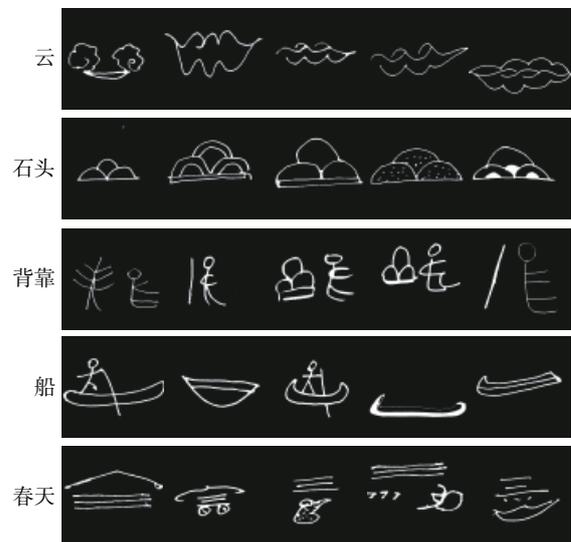


图 1 东巴象形文字数据集示例

Fig. 1 Samples of Dongba pictographs datasets

### 1.2 东巴象形文字图像预处理

人工仿写的东巴象形文字受光照以及拍照设备等的影 响, 往往会产生极大的噪声, 影响东巴象形文字数据集的质量, 因此必须对其进行一系列的图像预处理。图 2 给出了本文建立东巴象形文字数据集的技术路线, 具体步骤如下。

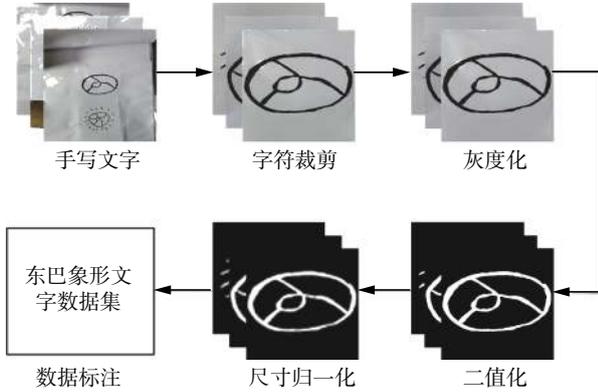


图 2 东巴象形文字数据集建立技术路线

Fig. 2 Technical route for Dongba pictographs dataset establishment

1) 字符裁剪。对人工仿写的原始图像进行字符裁剪, 使得每张图像中仅包含一个东巴象形文字。具体过程如算法 1 所示。

#### 算法 1 符裁剪算法

输入 未裁剪的手写东巴象形文字图像  $X$ ;

输出 仅包含一个东巴象形文字的图像  $Y$ 。

- ①  $X_h \leftarrow$  图像  $X$  的高度;
- ②  $X_w \leftarrow$  图像  $X$  的宽度;
- ③  $Y_h \leftarrow 1/3X_h - 2/3X_h$ ;
- ④  $Y_w \leftarrow 1/3X_w - 2/3X_w$ ;
- ⑤  $Y \leftarrow Y_h - Y_w$ 。

2) 灰度化。黑白两种颜色反差较大, 可提高东巴象形文字识别的效果。为此, 使用加权平均值法进行图像灰度化, 去除图像的颜色信息, 将三通道的彩色图像转换成单通道的灰度图像。灰度化公式如式 (1) 所示:

$$\text{Gray}_{i,j} = 0.299R_{i,j} + 0.587G_{i,j} + 0.114B_{i,j} \quad (1)$$

式中:  $R_{i,j}$ 、 $G_{i,j}$ 、 $B_{i,j}$  分别代表图像在  $(i,j)$  处的红、绿、蓝 3 种颜色分量像素值;  $\text{Gray}_{i,j}$  代表图像在  $(i,j)$  处的灰度值。

3) 二值化。为了极大程度减少图像数据量, 通过全局阈值二值化减少图像无关像素信息, 并使整个图像呈现出明显的黑白效果, 凸显东巴象形文字轮廓, 图像二值化公式如式 (2) 所示:

$$b_{i,j} = \begin{cases} 0, & \text{Gray}_{i,j} \geq 125 \\ 255, & \text{Gray}_{i,j} < 125 \end{cases} \quad (2)$$

式中  $b_{i,j}$  表示图像二值化后图像在  $(i,j)$  处的像素值。

4) 尺寸归一化。常用的图像尺寸归一化方法

是双线性插值法, 但是当原图像与尺寸归一化图像尺寸相差过大时, 尺寸归一化后的图像纹理特征易损坏, 不利于深度学习模型识别。而像素区域关系重采样法能够保留完整图像信息的条件下, 将输入图像尺寸最大程度减小, 大幅度减少图像像素数以及数据量, 在保证深度学习模型识别准确率不变的前提下, 加快模型的训练速度。根据其他数据集图像尺寸大小设置的经验以及多次对比实验验证, 我们发现当图像尺寸归一化为  $64 \times 64$  时, 可以取得最好的识别效果, 并且模型训练速度快。本文对像素区域关系重采样法和双线性插值法在东巴文字图像上的效果进行了简单的实验对比, 分别将图像尺寸归一化为  $64 \times 64$ 。图 3 给出了实验结果。



(a) 双线性插值法 (b) 像素区域关系重采样法

图 3 两种尺寸归一化方法示例

Fig. 3 Samples of two size normalization methods

从图 3 中可以看出, 双线性插值法后的东巴象形文字纹理特征有残缺, 而像素区域关系重采样可获得更好的尺寸归一化效果。

因此本文选择像素区域关系重采样法进行尺寸归一化操作, 其公式如式 (3) 所示:

$$A = B \odot \alpha$$

$$f_{i,j} = \frac{\sum_{i,j} a_{i,j}}{\sum_{i,j} \alpha_{i,j}} \quad (3)$$

式中:  $B$  表示图像某区域内像素值矩阵;  $\alpha$  是与  $B$  相对应的像素值系数矩阵, 其取值取决于原图像与尺寸归一化图像的尺寸大小关系;  $\odot$  表示 Hadamard 积;  $f_{i,j}$  表示图像  $B$  区域通过尺寸归一化后的像素值。

5) 数据标注。通过数据编码标注, 将第  $i$  个东巴象形文字的所有图像  $I$  统一编码为  $i$ , 使计算机将图像和编码相互对应, 如式 (4) 所示:

$$F(I) = i \quad (4)$$

式中  $F(\cdot)$  表示编码标注算法, 具体过程如算法 2 所示。

#### 算法 2 编码标注算法

输入 train, test (其中有命名为  $i$  (包含图像  $I$ ) 的文件夹)

输出 图像  $I$  与其编码  $i$  相互对应的 txt 文档

- ① for  $i \in \text{train, test}$ ;
- ② for  $I \in i$ ;

- ③将  $I$  的绝对地址和  $i$  写入 txt 文档;
- ④换行;
- ⑤重复迭代 2)~4);
- ⑥返回图像  $I$  与其编码  $i$  相互对应的 txt 文档

通过上述一系列的图像预处理, 本文建立了东巴象形文字数据集, 该数据集包含 1387 个东巴象形文字(包括异体字), 每个东巴象形文字对应 160 余张书写各异的图片, 数据集图片总量为 223 050 张。

## 2 改进残差学习神经网络

近年来, 深度学习成功应用于图像识别<sup>[16-19]</sup>领域, 提出了一系列性能优异的网络模型, 其中 ResNet 模型首次提出残差跳跃连接(residual short-

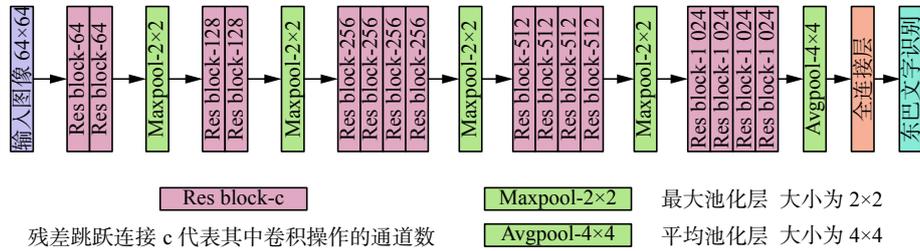


图 4 东巴象形文字识别网络结构

Fig. 4 Network structure of Dongba pictographs recognition

### 2.1 残差跳跃连接

残差跳跃连接可以解决神经网络随着深度增加出现性能退化的问题。深层神经网络难以拟合的原因是恒等映射  $H(x) = x$  的学习比较困难, 但当把网络设计为  $H(x) = F(x) + x$  时, 可以把学习恒等映射转化为更加容易学习的残差映射  $F(x) = H(x) - x$ , 并且  $F(x)$  对输出变化更加敏感, 参数的调整幅度更大, 从而可以加快学习速度, 提高网络的优化性能。残差跳跃连接的一般定义如式 (5) 所示:

$$y = F(x, \{W_i\}) + W_s x \quad (5)$$

其中  $W_s$  主要是用  $1 \times 1$  卷积<sup>[21]</sup> 来匹配残差跳跃连接输入  $x$  和输出  $y$  的通道维度。  $F(x, \{W_i\})$  为网络需要学习的残差映射。而当残差跳跃连接输入和输出维度相同时, 可将其定义如式 (6) 所示:

$$y = F(x, \{W_i\}) + x \quad (6)$$

文献 [22] 证明了越是接近当前卷积层的前层输出对当前层的特征提取效果影响越大, 可以使网络更容易训练。为此本文仅将相邻堆叠的卷积层组成残差跳跃连接结构, 在有效重复利用特征图的同时, 降低网络参数量和复杂度。本文残差跳跃连接结构如图 5 所示。图 5 中残差映射  $F(x)$  如式 (7) 所示, 输出  $y$  和输入  $x$  的关系式如式 (8) 所示:

$$F(x) = W_2 \sigma(W_1 x) \quad (7)$$

cut connection) 结构<sup>[20]</sup>, 解决了网络加深带来的梯度消失问题以及神经网络深度与识别准确度之间的矛盾, 可有效提取更多的图像细节特征, 目前已成为图像识别的主流深度学习模型。

考虑到东巴象形文字识别的具体问题, 不仅字数多、书写随意性较大, 而且有些字形较为相似, 因此需要提取细节特征能力强的网络结构, 为此本文选择 ResNet 模型作为本文改进的网络结构, 设计了残差跳跃连接方式和卷积层的数量, 并通过加入最大池化层实现了对下采样的改进, 更好地提取了东巴象形文字的纹理分布特征。本文设计的网络主要框架如图 4 所示。下面将详细介绍设计思路和改进方法。

$$y = \sigma(W_2 \sigma(W_1 x) + x) \quad (8)$$

式中:  $\sigma$  均表示 ReLU 激活函数,  $W_1$  和  $W_2$  分别表示卷积层学习的参数。

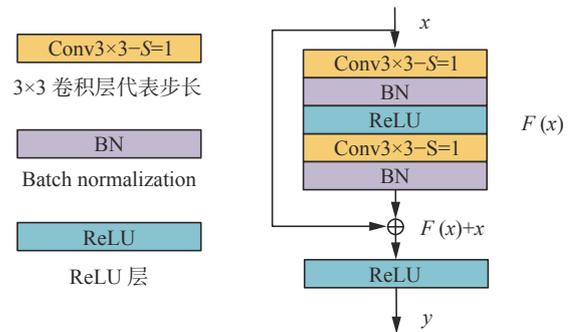


图 5 本文残差跳跃连接模块

Fig. 5 Residual skip connection module of this paper

### 2.2 卷积层的选择

网络模型卷积核的选择与网络计算量大小密切相关。尽管大的卷积核可以直接增大感受野, 但是会带来网络计算量的暴增, 而多个  $3 \times 3$  卷积核可以在降低计算量的前提下实现  $5 \times 5$  或  $7 \times 7$  乃至更大卷积核的效果。由式 (9) 可以发现, 3 个  $3 \times 3$  卷积核实现  $7 \times 7$  卷积核效果时, 其参数量可以减少到 55%。

$$\eta = \frac{3 \times 3 \times 3 \times C^2}{7 \times 7 \times C^2} \times 100\% \approx 55\% \quad (9)$$

其中  $C$  指输入和输出的通道数。

除了能够降低网络模型的计算量, 单个 $3 \times 3$ 卷积核还可以捕获特征图像素四周的信息, 多个 $3 \times 3$ 卷积核的叠加还可以直接增加网络深度<sup>[23]</sup>, 使网络模型的特征提取能力更强, 从而取得更好的识别效果。

东巴象形文字具有字数多、字形相似等特点, 因此需要提取特征能力强的网络结构, 而多个卷积层的叠加能够在参数量最少的前提下实现最好的特征提取能力。因此本文设计了 32 层 $3 \times 3$ 的卷积层, 再加 1 层全连接层, 构成 33 层网络模型, 用以获得东巴象形文字最好的识别效果。

同时, 为了防止网络过拟合, 加快网络训练速度, 本文对每一个卷积层执行批量归一化 (batch normalization)<sup>[24]</sup>操作。然后再使用修正线性单元 ReLU<sup>[25]</sup>(rectified linear units) $f(x) = \max(0, x)$ 作为激活函数, 增强网络的非线性表达能力, 在 $x > 0$ 时保持梯度不衰减, 从而缓解网络出现的梯度消失问题。

### 2.3 下采样改进

下采样可以降低特征图维度, 保留图像主要特征的同时减少网络模型的参数量, 防止过拟合现象的发生。在 ResNet 模型中通常采用令卷积步长 $\text{Stride} = 2$ 来实现下采样的效果, 但是由于本文建立的东巴象形文字数据集经过灰度归一化后, 其前景像素值远远大于背景像素值, 用这种方法实现下采样获得的东巴象形文字纹理特征不够丰富, 影响了识别效果, 因此有必要对下采样进行改进。

最大池化层通过提取特征图局部区域内的像素最大值, 可以最大程度降低特征图背景的无关信息, 使网络模型提取更多有用的前景特征, 降低背景特征干扰。因此, 本文对 ResNet 模型中的下采样方式进行了改进, 通过采用最大池化层来获得丰富的纹理特征。最大池化层的公式如式 (10) 所示:

$$\text{pooling}_{\max} = \max a_i, \quad i \in r_k \quad (10)$$

其中:  $r_k (k = 1, 2, \dots, K)$ 为特征图所划分的多个区域,  $a_i$ 表示第  $i$  区域内的像素值。

而网络深层的平均池化层通过提取特征图的像素加权值, 可以保留更加完整的特征图信息。并且通过平均池化将特征图下采样为 $1 \times 1$ 后再与全连接层相连接, 可以减少网络参数。池化层的池化区域为特征图中的连续区域, 对小的形态改变具有不变性, 不仅能够逐步减少特征图的空间大小、参数数量、内存占用和计算量, 而且拥有更大的感受野, 可有效控制过拟合现象的发生。

## 3 实验结果及分析

为验证本文创新工作的有效性与先进性, 实验部分主要做了 3 个方面的工作: 1) 本文建立的东巴象形文字数据集对比实验及分析; 2) 本文提出的东巴象形文字识别方法对比实验及分析, 包括网络改进前后的对比实验; 3) 结合实验结果, 分析归纳了目前仍存在的问题。

### 3.1 实验条件

实验中所有对比实验均在表 1 所示的实验平台上运行。

表 1 实验环境配置

| 配件   | 参数  |
|------|---|
| 操作系统 | Ubuntu 16.04  |
| 处理器  | Intel <sup>R</sup> Core <sup>TM</sup> i7-7700CPU@3.60 GHz |
| 显卡   | GeForce GTX 1070 Ti                                       |
| 内存   | 16 GB   |
| 开发工具 | Pytorch1.0.0 Python3.6.2                                  |

实验 epoch 设置为 80, 初始学习率设置为 0.001, 每 50 个 epoch 将学习率降低为原来的三分之一, 直到运行结束所有 epoch。

本文梯度优化函数选择 Adam 函数, 损失函数使用交叉熵函数, 交叉熵函数定义如式 (11) 所示:

$$L = -\log \left( \frac{\exp(x_j)}{\sum_i \exp(x_i)} \right) = -x_j + \log \sum_i \exp(x_i) \quad (11)$$

其中 $x_j$ 代表全连接层第  $j$  个网络节点输出值。

本文改进的 ResNet 模型具体参数如表 2 所示。

### 3.2 本文建立的数据集验证

目前关于东巴象形文字的数据集较少, 文献 [8-10] 是目前已知的 3 个东巴象形文字数据集, 因此将本文的数据集与上述 3 种数据集都进行了对比实验。

#### 3.2.1 数据集有效性验证

这里选取在图像识别领域表现优异的 ResNet18、ResNet34、VGGNet 以及本文的改进网络模型在本文建立的东巴象形文字数据集上进行识别效果对比。在数据集中随机选取 5 000 张图像计算其均值和方差, 然后将图像归一化处理输入网络。随机选取数据集图片总数的 80% 作为训练集, 即 178 223 张图片, 其余 44 827 张图片作为测试集。在训练集上训练网络模型后, 在测试集上对 1 387 个东巴象形文字 (包括异体字) 进行识别准确率测试。其实验结果如表 3 所示。

表 2 本文网络参数设计  
Table 2 Network configurations of this paper

| 网络层          | 卷积核   | 输出图像<br>尺寸大小/<br>像素×像素 | 输出通道数 |
|--------------|---|------------------------|-------|
| 输入层          | /   | 64×64                  | 1     |
| 卷积层          | $\begin{bmatrix} 3 \times 3 \text{步长}=1 \\ 3 \times 3 \text{步长}=1 \end{bmatrix} \times 2$ | 64×64                  | 64    |
| 最大池化         | 2×2步长=2   | 32×32                  | 64    |
| 卷积层          | $\begin{bmatrix} 3 \times 3 \text{步长}=1 \\ 3 \times 3 \text{步长}=1 \end{bmatrix} \times 2$ | 32×32                  | 128   |
| 最大池化         | 2×2步长=2   | 16×16                  | 128   |
| 卷积层          | $\begin{bmatrix} 3 \times 3 \text{步长}=1 \\ 3 \times 3 \text{步长}=1 \end{bmatrix} \times 4$ | 16×16                  | 256   |
| 最大池化         | 2×2步长=2   | 8×8                    | 256   |
| 卷积层          | $\begin{bmatrix} 3 \times 3 \text{步长}=1 \\ 3 \times 3 \text{步长}=1 \end{bmatrix} \times 4$ | 8×8                    | 512   |
| 最大池化         | 2×2步长=2   | 4×4                    | 512   |
| 卷积层          | $\begin{bmatrix} 3 \times 3 \text{步长}=1 \\ 3 \times 3 \text{步长}=1 \end{bmatrix} \times 4$ | 4×4                    | 1024  |
| 平均池化         | 4×4   | 1×1                    | 1024  |
| 全连接层 softmax |   |                        |       |

表 3 数据集有效性验证实验  
Table 3 Experiment of dataset validity verification

| 测试模型        | 识别准确率/%      |
|-------------|--------------|
| ResNet18    | 98.22        |
| ResNet34    | 98.01        |
| VGGNet      | 98.34        |
| <b>本文算法</b> | <b>98.65</b> |

从表 3 中可以看出,对于不同的网络模型,本文建立的东巴象形文字数据集都获得了高于 98% 的识别准确率,最高可达 98.65%,这说明本文建立的东巴象形文字数据集是有效的,每个东巴象形文字多达 160 多张书写各异的图片,其数据规模完全满足具体识别的要求。

### 3.2.2 数据集先进性验证

文献 [8-10] 分别给出了 3 种东巴象形文字识别方法和与之对应的 3 个东巴象形文字数据集,这里采用这 3 种识别方法在本文提出的数据集上分别进行了识别准确率方面的对比实验。表 4 给出了各个数据集能够识别的字数和不同模型在数据集上进行识别的准确率。

从表 4 中可以看出,首先本文建立的数据集能够识别的东巴象形文字最多;其次,相同的网

络模型在不同的东巴象形文字数据集上取得的识别效果不同,相较于其他 3 个文献所建立的数据集,本文建立的数据集采用 3 种相对应的网络模型都取得了最高的识别准确率,说明本文建立的数据集在数据规模和数据质量上都是目前最好的,也说明优秀的数据集可辅助提高深度学习模型的性能。

表 4 数据集先进性验证实验  
Table 4 Experiment of dataset advancement verification

| 实验数据集     | 可识别东巴文字数 | 测试模型     | 识别准确率/% |
|-----------|----------|----------|---------|
| 文献[8]数据集  | 956      | ResNet18 | 85.6    |
| 文献[9]数据集  | 200      | VGGNet   | 95.8    |
| 文献[10]数据集 | 536      | ResNet20 | 93.58   |
|           |          | ResNet18 | 98.22   |
| 本文数据集     | 1387     | VGGNet   | 98.34   |
|           |          | ResNet20 | 97.7    |

### 3.3 本文识别算法的实验验证

根据东巴象形文字识别的特点,本文对 ResNet 模型进行了改进,提高了东巴象形文字的识别准确率。这里将验证本文网络模型改进的有效性。通过将其与采用残差跳跃连接加传统池化方式以及无残差跳跃连接加最大池化方式的网络模型进行消融实验。同时,将本文改进的网络模型与文献 [8-10] 中取得识别准确率最高的网络模型以及 ResNet34 进行对比实验,以验证其先进性。所有实验在本文建立的数据集上进行。

#### 3.3.1 算法的有效性验证

为了验证本文改进 ResNet 模型的有效性,这里进行了改进前后的对比实验。将本文改进的网络模型(残差+最大池化)与残差加传统池化、无残差加最大池化 3 种网络模型进行识别效果对比,实验结果如表 5 所示。

表 5 算法有效性验证实验  
Table 5 Experiment of algorithm validity verification

| 测试模型           | 识别准确率/%      |
|----------------|--------------|
| 残差+传统池化        | 98.11        |
| 无残差+最大池化       | 97.64        |
| <b>残差+最大池化</b> | <b>98.65</b> |

由表 5 可以看出,本文改进的残差跳跃连接加最大池化下采样网络模型取得了最高的识别准确率,相较于残差跳跃连接加传统池化的网络模型提高了 0.54%;相较于无残差跳跃连接加最大池化下采样的网络模型提高了 1.01%,从而验证

了本文改进残差跳跃连接加最大池化网络模型的有效性。

### 3.3.2 算法的先进性验证

为了验证本文改进网络模型的先进性, 在相同的实验环境下, 本文分别与文献 [8] 采用的 ResNet18 网络模型、文献 [9] 采用的 VGGNet 网络模型以及文献 [10] 采用的 20 层 ResNet 网络模型进行了对比实验, 实验结果如表 6 所示。

表 6 算法先进性验证实验

Table 6 Experiment of model advancement verification

| 测试模型           | 识别准确率/% |
|----------------|---------|
| ResNet18文献[8]  | 98.22   |
| VGGNet文献[9]    | 98.34   |
| ResNet20文献[10] | 97.7    |
| ResNet34       | 98.01   |
| 本文算法           | 98.65   |

从表 6 中可以看出, 本文改进的网络模型识别准确率最高, 相较于文献 [8] 的方法提高了 0.43%; 相较于文献 [9] 的方法提高了 0.31%; 相较于文献 [10] 的方法提高了 0.95%。充分验证了本文改进网络模型的先进性。

同时, 本文又与层数有所增加的 ResNet34 网络进行了对比性实验。从表 6 中可以看出, 34 层网络模型的识别准确率不仅低于本文的 33 层网络模型, 而且也低于 18 层的网络模型, 这说明网络层数的简单叠加在具体的东巴象形文字识别中不一定获得更好的识别效果。

### 3.4 存在的问题

虽然本文取得了 98.65% 的识别准确率, 但对于误识别问题我们又进行了深入分析, 通过观察多次实验结果, 发现错误识别的东巴象形文字都有一个共同的特点, 那就是都有与之非常相似的东巴象形文字, 图 6 给出了部分相似文字的示例。



图 6 相似东巴象形文字示例

Fig. 6 Samples of similar Dongba pictographs

从图 6 可以看出, “水槽”和“水洞”, “侧视之人”和“左”或“爬”等字的区别仅仅体现在线条的弯曲程度不同; “腰”和“爬”更多体现在它们之间大小有所差异; “神山山脚”和“神山山腰”, “中”和“矛”主要体现在图像上部分所画的高度不同; “尾巴”和“树倒”的差异体现在右下角线条的长度和弯曲程度; “臂膀”和“手”则几乎相同。

可见, 东巴象形文字中有很多相似乃至接近“相同”的文字, 又因为东巴象形文字的手工书写形式, 随意性较大, 这些相似的东巴象形文字在书写过程中极易导致差异性变小、辨识度下降, 这是影响东巴象形文字识别准确率的主要原因。

## 4 结束语

针对现有东巴象形文字识别方法存在的识别文字数量少、识别准确率较低等问题, 本文首先建立了包含 1387 个东巴象形文字 (包括异体字)、图片总量达到 22 万余张的东巴象形文字数据集, 可识别的东巴象形文字大幅增加。通过扩大数据集的规模, 辅助提高了算法识别的准确率; 更为重要的是本文选择 ResNet 模型作为改进的网络结构, 设计了残差跳跃连接方式和卷积层的数量, 并通过加入最大池化层实现了对下采样的改进, 更好地提取了东巴象形文字的纹理分布特征。通过对 1387 个东巴象形文字 (包括异体字) 分别进行测试, 实验结果表明, 本文提出的改进 ResNet 模型识别准确率平均达到 98.65%, 取得了当前识别字数最多、识别准确率最高的效果。

未来将继续扩大东巴象形文字数据集的文字数量, 力争包含现存的所有东巴象形文字。针对其中相似度极高的文字, 将研究设计专门的网络模型来有效将它们区别开来, 从而进一步提高东巴象形文字识别的准确率。

## 参考文献:

- [1] GUO Hai, ZHAO Jingying, LI Xiaoniu. Preprocessing method for NaXi pictographs character recognition using wavelet transform[J]. *International journal of digital content technology and its applications*, 2010, 4(3): 117-131.
- [2] GUO Hai, ZHAO Jingying. Research on feature extraction for character recognition of NaXi pictograph[J]. *Journal of computers*, 2011, 6(5): 947-954.
- [3] DA Mingjun, ZHAO Jingying, SUO Guojie, et al. Online handwritten Naxi pictograph digits recognition system us-

- ing coarse grid[C]//Computer science for environmental engineering and EcoInformatics. Berlin, German: Springer, 2011: 390–396.
- [4] HSU C W, LIN C J. A comparison of methods for multi-class support vector machines[J]. *IEEE transactions on neural networks*, 2002, 13(2): 415–425.
- [5] BREIMAN L. Random forests[J]. *Machine learning*, 2001, 45(1): 5–32.
- [6] 徐小力, 蒋章雷, 吴国新, 等. 基于拓扑特征和投影法的东巴象形文识别方法研究 [J]. *电子测量与仪器学报*, 2017, 31(1): 150–154.
- XU Xiaoli, JIANG Zhanglei, WU Guoxin, et al. Identification method of Dongba pictograph based on topological characteristic and projection method[J]. *Journal of electronic measurement and instrumentation*, 2017, 31(1): 150–154.
- [7] 杨玉婷, 康厚良. 东巴象形文字特征曲线提取算法研究 [J]. *图学学报*, 2019, 40(3): 591–599.
- YANG Yuting, KANG Houliang. Research on the extracting algorithm of dongba hieroglyphic feature curves[J]. *Journal of graphics*, 2019, 40(3): 591–599.
- [8] 张泽晖. 基于卷积神经网络的东巴文字分类与识别 [D]. 昆明: 云南大学, 2019.
- ZHANG Zehui. Classification and recognition of dongba characters based on convolutional neural network[D]. Kunming: Yunnan University, 2019.
- [9] WU Guoxin, LIU Xiuli, JIANG Zhanglei, et al. Dongba classical ancient books image classification method based on ReN-Softplus convolution residual neural network [C]//2019 14th IEEE International Conference on Electronic Measurement & Instruments. New York, USA: IEEE, 2019: 398–404.
- [10] 谢裕睿, 董建娥. 基于 ResNet 网络的东巴象形文字识别研究 [J]. *计算机时代*, 2021(1): 6–10.
- XIE Yurui, DONG Jian'e. Research on Dongba hieroglyph recognition using ResNet network[J]. *Computer era*, 2021(1): 6–10.
- [11] 李霖灿. 纳西族象形标音文字字典 [M]. 昆明: 云南民族出版社, 2001.
- LI Lincan. *Naxi Pictographs and Transcription Characters Dictionary*[M]. Kunming: Yunnan Nationalities Publishing House, 2001.
- [12] 方国瑜, 和志武. 纳西象形文字谱 [M]. 昆明: 云南人民出版社, 1981.
- FANG Guoyu, HE Zhiwu. *Naxi pictograph character chart*[M]. Kunming: Yunnan People's Publishing House, 1981.
- [13] BUSLAEV A, PARINOV A, KHVEDCHENYA E, et al. Alumentations: fast and flexible image augmentations[EB/OL]. (2018-09-18) [2021-11-30].<https://arxiv.org/abs/1809.06839>.
- [14] ZHANG Hongyi, MOUSTAPHA Cisseet, YANN N Dauphin, et al. Mixup: beyond empirical risk minimization[EB/OL]. (2017-10-25) [2021-11-30].<https://arxiv.org/abs/1710.09412>.
- [15] DAN Hendrycks, NORMAN Mu, EKIN D Cubuk, et al. AugMix: A simple data processing method to improve robustness and uncertainty[EB/OL]. (2019-12-05) [2021-11-30].<https://arxiv.org/abs/1912.02781>.
- [16] HE Tong, ZHANG Zhi, ZHANG Hang, et al. Bag of tricks for image classification with convolutional neural networks[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York, USA: IEEE, 2019: 558–567.
- [17] KHAN A, SOHAIL A, ZAHORA U, et al. A survey of the recent architectures of deep convolutional neural networks[J]. *Artificial intelligence review*, 2020, 53(8): 5455–5516.
- [18] XIE Saining, GIRSHICK R, DOLLÁR P, et al. Aggregated residual transformations for deep neural networks [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. New York, USA: IEEE, 2017: 5987–5995.
- [19] TAN Mingxing, LE Quoc V. EfficientNet: rethinking model scaling for convolutional neural networks [EB/OL]. (2019-05-28) [2021-11-30].<https://arxiv.org/abs/1905.11946>.
- [20] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. New York, USA: IEEE, 2016: 770–778.
- [21] LIN Min, CHEN Qiang, YAN Shuicheng. Network in network[EB/OL]. (2013-12-16) [2021-11-30].<https://arxiv.org/abs/1312.4400>.
- [22] HUANG Gao, LIU Shichen, MAATEN L V D, et al. CondenseNet: an efficient DenseNet using learned group convolutions[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York, USA: IEEE, 2018: 2752–2761.
- [23] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [EB/OL]. (2014-09-04) [2021-11-30].<https://arxiv.org/abs/1409.1556v3>.
- [24] IOFFE Sergey, SZEGEDY Christian. Batch normalization: accelerating deep network training by reducing internal covariate shift [C]//2015 International Conference

on Machine Learning. New York, USA: ACM, 2015: 448-456.

- [25] VINOD Nair, GEOFFREY E Hinton. Rectified Linear Units Improve Restricted Boltzmann Machines[C]//2010 International Conference on Machine Learning. New York, USA: ACM, 2010: 807-814.

#### 作者简介:



骆彦龙, 博士研究生, 主要研究方向为图像识别、深度学习。



毕晓君, 教授, 博士生导师, 主要研究方向为智能信息处理技术、数字图像处理、机器学习。主持国家重点研发计划项目、国家社科基金重大项目等国家级、省部级项目 6 项。获得高等学校科学技术进步一等奖 1 项、省部级科学技术奖 7 项。发表学术论文 170 余篇。



吴立成, 教授, 博士生导师, 主要研究方向为智能机器人、人工智能。主持国家自然科学基金、军口 863 项目等国家级、省部级项目十余项。获教育部科技进步奖、江苏省科技进步奖各 1 项。发表学术论文 80 余篇。

## 第五届中国模式识别与计算机视觉大会 The 5th Chinese Conference on Pattern Recognition and Computer Vision

第五届中国模式识别与计算机视觉大会 (The 5th Chinese Conference on Pattern Recognition and Computer Vision, PRCV 2022) 将于 2022 年 10 月 14 日至 10 月 17 日在深圳举办。PRCV 2022 由中国人工智能学会 (CAAI)、中国计算机学会 (CCF)、中国自动化学会 (CAA) 和中国图象图形学学会 (CSIG) 联合主办; 由南方科技大学和深圳职业技术学院共同承办; 并由香港浸会大学、香港中文大学 (深圳)、哈尔滨工业大学 (深圳)、中国科学院深圳先进技术研究院和中山大学联合承办, 是国内顶级的模式识别和计算机视觉领域学术盛会。

本届会议旨在汇聚国内国外模式识别和计算机视觉理论与应用研究的广大科研工作者及工业界同行, 共同分享我国模式识别与计算机视觉领域的最新理论和技术成果。现向广大科技工作者公开征集高质量、原创性的优秀英文学术论文。录用的论文将在大会上展示, 会议论文集将由 Springer 出版社出版, 并被 EI 和 ISTP 检索。

#### 征稿范围 (包括但不限于):

模式分类与聚类分析; 结构模式识别; 机器学习; 神经网络与深度学习; 特征提取与特征选择; 计算机视觉基础理论; 底层视觉理解、图像处理; 三维视觉与重构; 计算摄像学、传感与显示技术; 文档分析与识别; 字符识别; 人脸识别与姿态识别; 性能评测和基准数据库; 目标检测、跟踪与识别; 行为识别; 多媒体分析与推理; 医学图像处理与分析; 生物特征识别; 遥感影像解译; 优化及学习方法; 多模态信息处理; 视频分析与理解; 视觉应用与系统; 机器人、自动驾驶中的视觉问题。

#### 重要日期:

投稿截止日期: 2022 年 4 月 15 日

录用通知日期: 2022 年 6 月 15 日

终稿截止日期: 2022 年 7 月 5 日

联系咨询: prcv2022@163.com