



改进ResNeSt网络的拓片甲骨文字识别

毛亚菲, 毕晓君

引用本文:

毛亚菲, 毕晓君. 改进ResNeSt网络的拓片甲骨文字识别[J]. 智能系统学报, 2023, 18(3): 450–458.

MAO Yafei, BI Xiaojun. Rubbing oracle bone character recognition based on improved ResNeSt network[J]. *CAAI Transactions on Intelligent Systems*, 2023, 18(3): 450–458.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202211041>

您可能感兴趣的其他文章

基于改进残差学习的东巴象形文字识别

Dongba pictographs recognition based on improved residual learning

智能系统学报. 2022, 17(1): 79–87 <https://dx.doi.org/10.11992/tis.202112009>

基于改进的Faster RCNN面部表情检测算法

Facial expression recognition based on improved Faster RCNN

智能系统学报. 2021, 16(2): 210–217 <https://dx.doi.org/10.11992/tis.201910020>

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>

注意力机制和Faster RCNN相结合的绝缘子识别

Insulator recognition based on attention mechanism and Faster RCNN

智能系统学报. 2020, 15(1): 92–98 <https://dx.doi.org/10.11992/tis.201907023>

层次化双注意力神经网络模型的情感分析研究

Hierarchical double-attention neural networks for sentiment classification

智能系统学报. 2020, 15(3): 460–467 <https://dx.doi.org/10.11992/tis.201812017>

基于卷积特征和贝叶斯分类器的人脸识别

Face recognition based on convolution feature and Bayes classifier

智能系统学报. 2018, 13(5): 769–775 <https://dx.doi.org/10.11992/tis.201706052>

DOI: 10.11992/tis.202211041

网络出版地址: <https://kns.cnki.net/kcms/detail/23.1538.tp.20230221.1647.002.html>

改进 ResNeSt 网络的拓片甲骨文字识别

毛亚菲¹, 毕晓君²

(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: 目前, 拓片甲骨文字的识别方法存在局部细节特征提取能力弱, 对部分高相似度的甲骨文字识别率较低的问题。为此, 本文提出了一种基于改进 ResNeSt 网络的甲骨文字识别方法, 通过设计跳转连接结构, 逐步将网络浅层特征向网络深层传递并进行融合; 同时结合甲骨文字“长条形”的特点, 引入坐标注意力机制模块, 从宽度和高度两个方向上对所得特征进行加权融合; 最后通过去掉网络最后一层的激活函数和全连接层以及对最后一个卷积层输出通道数的重新设置, 对网络分类器进行了有效优化。实验结果表明, 本文提出的改进拓片甲骨文字识别模型在 OBC306 数据集上识别准确率达到 93.53%, 取得了目前最好的识别效果。

关键词: ResNeSt 网络模型; 甲骨文字识别; 跳转连接; 坐标注意力机制; 分类器优化; OBC306; 深度学习; 神经网络

中图分类号: TP391 文献标志码: A 文章编号: 1673-4785(2023)03-0450-09

中文引用格式: 毛亚菲, 毕晓君. 改进 ResNeSt 网络的拓片甲骨文字识别 [J]. 智能系统学报, 2023, 18(3): 450-458.

英文引用格式: MAO Yafei, BI Xiaojun. Rubbing oracle bone character recognition based on improved ResNeSt network[J]. CAAI transactions on intelligent systems, 2023, 18(3): 450-458.

Rubbing oracle bone character recognition based on improved ResNeSt network

MAO Yafei¹, BI Xiaojun²

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

Abstract: At present, the methods for recognition of rubbing oracle bone characters have some problems, such as weak ability to extract local details and low recognition rate for some oracle characters with highly similarity. Therefore, an oracle bone character recognition method based on the improved ResNeSt network is proposed in this paper. By designing jump connection structure, the shallow network features are gradually transferred to the deep network and merged. At the same time, combined with the Strip characteristics of the oracle bone characters, the coordinate attention mechanism module is introduced to carry out the weighted fusion of the obtained features in both width and height directions. Finally, the network classifier is optimized effectively by removing the activation function of the last layer and the full connection layer, and resetting the number of output channels of the last convolution layer. The experimental results show that the recognition accuracy of the improved rubbing oracle bone character recognition model proposed in this paper reaches 93.53% on OBC306 data set, having achieved the best recognition effect at present.

Keywords: ResNeSt network model; oracle bone character recognition; jump connection; coordinate attention mechanism; optimization of classifier; OBC306; deep learning; neural network

甲骨文主要指殷墟甲骨文, 是殷商时代刻在龟甲兽骨上的文字, 它是中国已发现的古代文字中时代最早、体系较为完整的文字, 对中国文化

史的完善以及汉字形成和演变的研究有着极其重要的意义^[1]。拓片甲骨文字图像主要是由专家对出土的龟甲、兽骨等文字载体上进行拓片所获得的原始图像, 由于龟甲、兽骨埋藏于地下, 年代久远, 遭受侵蚀, 所以获得的拓片甲骨文字图像具有噪声严重、图像残缺、类内样本少、类间分布不均的

收稿日期: 2022-11-28. 网络出版日期: 2023-02-23.

基金项目: 国家自然科学基金重点项目 (62236011).

通信作者: 毕晓君. E-mail: bixiaojun@hrbeu.edu.cn.

问题^[2], 同时由于古代文字书写风格随意, 部分甲骨文字的相似度极高, 这些都增加了计算机识别甲骨文的难度, 从而导致识别准确率较低。

早期的甲骨文字识别方法主要通过分析甲骨文字的拓扑结构^[3], 利用支持向量机^[4]和分形几何^[5]等方法进行分类。这些方法不仅精度低, 而且泛化能力较差。近些年来, 深度学习技术发展迅速, 在图像分类和识别等方面取得了突出的成果。越来越多的研究尝试将深度卷积神经网络应用在拓片甲骨文字的识别工作当中^[6-9]。一些在图像领域取得显著效果的模型如 AlexNet^[10]、VGG^[11]、ResNet^[12]和 Inception-v4^[13]等都被应用到拓片甲骨文字识别任务当中, 取得了较好的效果。其中具有代表性的有: 针对样本数据分布不均造成的特征提取困难问题, 王浩彬^[13]引入了循环式生成对抗网络算法, 在没有配对图像的情况下, 使网络可以自行学习拓片甲骨文字的风格, 生成具有丰富多样性的字符图片。张颐康等^[14]提出了一种基于三元组损失的度量学习方法, 基于对抗训练的思路, 将拓片甲骨文字映射到与临摹甲骨文字相同的特征空间中, 并约束二者服从尽可能相似的分布, 从而隐式地解决甲骨文字识别中的类间不平衡问题。Liu 等^[15]提出了一种随机多边形覆盖算法, 在输入图像上随机遮蔽一个多边形区域, 以此来模拟字符图像的噪声及残缺问题, 从而使扩增出来的数据接近原本的数据风格。Li 等^[16]提出了一种混合数据增强的策略, 在训练过程中同一类别中的不同样本的特征进行随机混合, 从而得到新的样本特征, 使数据分布达到相对平衡的状态。上述方法虽然利用各种深度学习模型和数据增强的方法及策略, 在一定程度上提高了模型的特征提取能力, 使识别准确率获得了提升, 但是目前提出的甲骨文字识别模型对局部细节特征提取能力差, 且对于有效特征的利用能力较低, 因此对高相似性甲骨文字无法取得很好的识别效果。

目前, 在计算机视觉领域的图像分类任务中, ResNeSt^[17]达到了先进的分类水平, 其在保留了经典的残差结构的同时, 创造性地将通道注意力和多分支结构融合在特征提取单元中, 大大提高了模型对有效特征的感知能力。但是, 目前该网络尚未在甲骨文字识别中应用, 为此本文首次将该模型引入到甲骨文字识别研究当中, 并针对部分甲骨文字存在高度相似的问题, 为进一步提高局部细节特征提取能力, 进行了以下 3 个方面的改进。首先, 在原有模型的基础上设计了跳转连

接结构, 将网络提取到的浅层特征逐步引向网络深层, 并与深层图像特征进行融合, 由此获得更多的局部细节特征; 其次, 针对甲骨文字整体呈“长条形”, 且书写风格随意, 大小不一的特点, 引入了坐标注意力机制^[18], 沿着宽度方向和高度方向对通道进行编码, 将特征分别沿着两个方向进行加权融合。最后, 为了提高识别过程中模型对有效特征的利用, 对网络分类器进行了优化。去掉了网络最后一层的 ReLU 激活函数和全连接层, 根据甲骨文字类别数对最后一个卷积层的输出通道数进行重新设置, 经过平均池化层对获得的深度特征进行融合, 并对最终分类结果进行输出。

1 ResNeSt 分类网络模型

ResNeSt 保留了 ResNet^[19]独有的残差结构, 使网络层数可以增加得很深, 而不会出现深层网络退化的问题。同时, 为了从众多信息中选择出对当前任务目标更关键的信息^[20], 模型在设计特征提取单元的时候借鉴了 SENet (squeeze and excitation networks)^[21]和 SKNet (selective kernel network)^[22]2 种注意力机制思想。SENet 是最先被提出的通道注意力, 主要利用压缩和激励模块收集全局信息, 捕获通道间的关系, 从而自适应的对通道特征相响应进行了重新校准。SKNet 在前者的基础上, 通过具有不同大小卷积核的双分支卷积结构, 实现了感受野的动态调整。因此, ResNeSt 受 SENet 和 SKNet 核心思想的启发, 在 ResNet 的基础上, 设计出了特征提取单元 ResNeSt Block, 结构如图 1 所示。

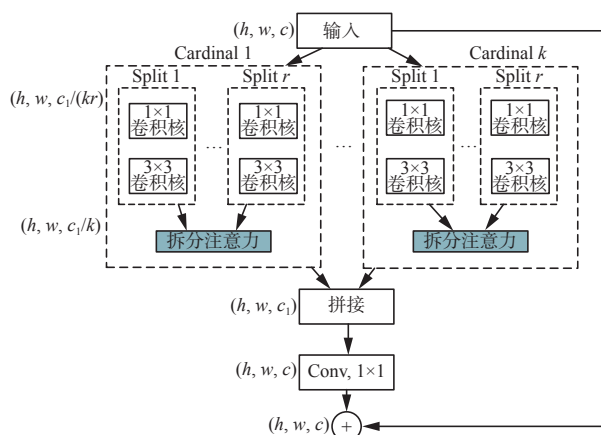


图 1 ResNeSt Block 结构

Fig. 1 Structure of ResNeSt Block

ResNeSt Block 中的多卷积分支结构, 可以在同一层中分别提取特征, 使网络提取的特征更为多样; 而其中的拆分注意力结构使模型对重要信

息的感知能力增强,从而有效提升了模型的特征提取能力。拆分注意力结构如图 2 所示。

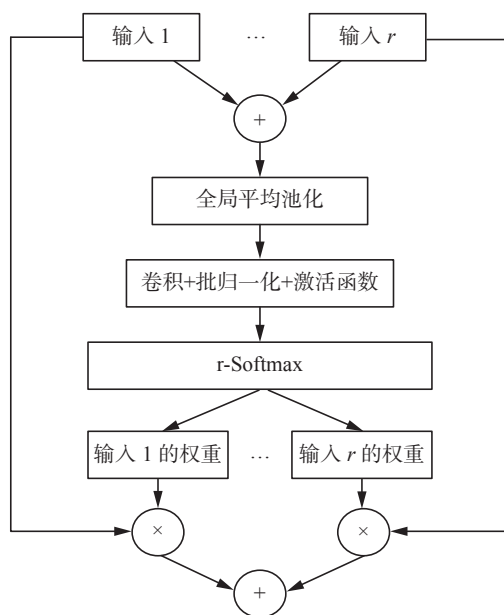


图 2 拆分注意力结构

Fig. 2 Structure of Split-Attention

输入图像在进入网络后,图像特征会被分为若干基数组(Cardinal k),基数组的数量由超参数 K 来确定。而在每个基数组内,特征将继续进行分裂,此时引入一个超参数 R ,来表示基数组内特征的分裂数量,因此得到总的特征组数为 $G=KR$ 。

在单个基数组内,分裂图像特征分别经过卷积层后,模型会对所得特征进行求和操作,然后通过全局平均池化操作,得到与分裂图像特征相同维度的特征向量。接着使用两组卷积核大小为 1 的卷积操作进行权重系数的分配,然后通过 r-softmax 对各个分裂图像特征的权重分别计算,得到各分裂图像特征对应的独立分布的权重,最后各分裂图像特征与其对应权重相乘再求和。在

各个基数组之间,将经过拆分注意力结构的特征进行聚合,然后以标准的残差结构将当前 ResNeSt Block 的输入与输出相连接,并作为新的特征输入到下一层网络中。

2 改进 ResNeSt 分类网络模型

ResNeSt 分类网络模型与其他经典网络模型相比,虽然在图像识别精度上有很大的性能提升,但是由于拓片甲骨文字部分文字相似度较高,有些释义不同的甲骨文字在轮廓方面近乎相同,仅仅在细节方面存在微小区别,辨识度较低,因此整体识别效果依旧不太理想。图 3 为字形相近的甲骨文字示例。

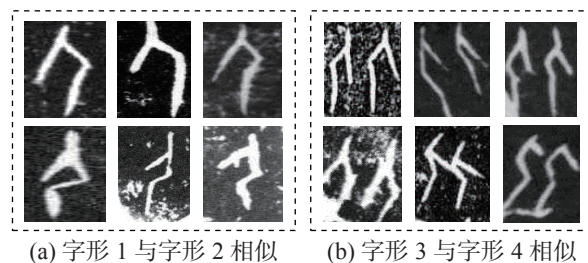


图 3 字形相近的文字示例

Fig. 3 Examples of near-form characters

为此,本文在 ResNeSt 网络模型的基础上,设计了跳转连接结构,将网络浅层特征向网络深层传递并进行融合,并在特征提取单元中加入了坐标注意力机制模块,从宽度和高度两个方向上对所得特征进行加权融合,最后去掉网络最后一层的激活函数和全连接层,并根据甲骨文字类别数对网络模型最后一层的卷积层输出通道数进行了重新设置,经过平均池化层融合学到的深度特征后,最终对分类结果进行输出。通过上述 3 个方面的改进,有效提升了算法的识别精度。本文方法的整体网络结构如图 4 所示。

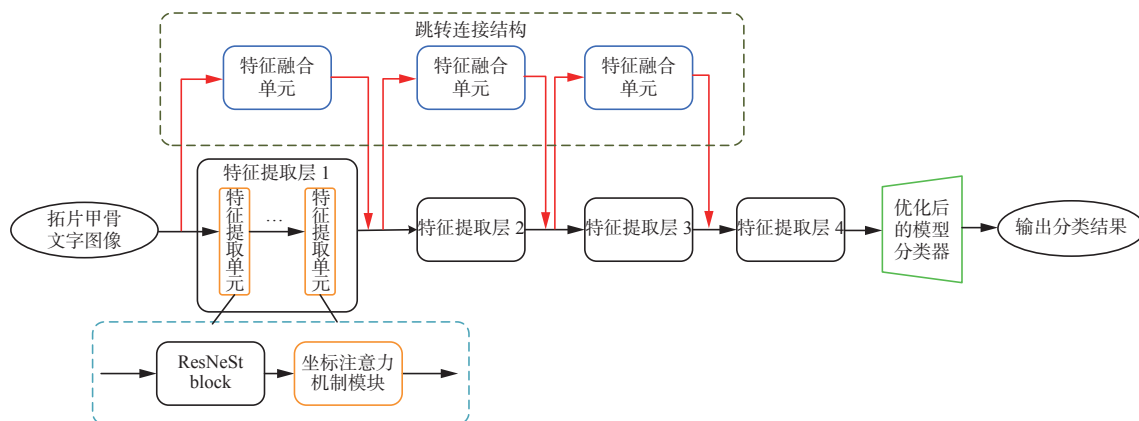


图 4 改进的 ResNeSt 网络结构

Fig. 4 Improved ResNeSt network structure

2.1 跳转连接的设计

ResNeSt 网络模型的特征提取单元通过步长大于 1 的卷积来降低特征图的大小, 因此, 特征信息从网络浅层传递到网络深层是一个特征图尺度逐渐缩小的过程。在这个过程中, 梯度和特征信息会变得不明确甚至消失, 最终得到的局部细节特征较少, 使得网络对高相似性甲骨文字产生误判, 最终导致识别率较低。

为解决这一问题, 本文借鉴 DenseNet^[23] 的跳转连接思想。在 DenseNet 的网络结构中, 靠近输入和靠近输出层之间包含更短的连接。对于每一层, 所有前一层的特征图作为输入, 而这一层的特征图作为后续所有层的输入, 有效地缓解了梯度消失问题, 并加强了特征的传播, 鼓励特征复用。因此, 本文在 ResNeSt 网络模型的特征提取结构中, 通过跳转连接逐步从网络浅层向网络深层引入图像特征, 并与深层图像特征进行融合, 从而较好解决了局部细节特征从网络浅层到深层传递过程的信息损失问题。

在 ResNeSt 的特征提取结构当中包含着 4 个特征提取层, 每个层中包含了不同数量的特征提取单元。随着网络层数的加深, 提取到的特征图尺寸在逐步减小。而本文提出的跳转连接结构是将特征提取层 i 的输入特征图与输出特征图进行通道级的连接, 然后作为特征提取层 $i+1$ 的输入特征图。此过程要求需要相加的特征图必须在维度、宽度以及高度保持一致, 因此, 本文设计了特征融合单元, 来对特征提取层 i 的输入特征图的维度、宽度以及高度进行调整, 此过程如图 5。

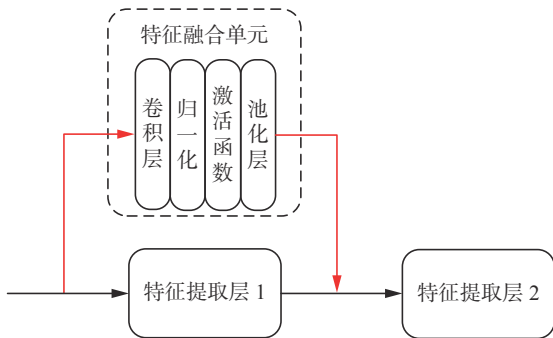


图 5 跳转连接结构

Fig. 5 Jump connection structure

假设特征提取层 1 的输入特征图信息 I_1 为 $C_1 H_1 W_1$, 输出特征图信息为 $C_2 H_2 W_2$, 则特征提取层 2 的输入特征图信息 I_2 为

$$I_2 = \sigma(T(C_1 H_1 W_1)) + C_2 H_2 W_2 \quad (1)$$

式中: σ 表示步长为 2 的最大池化操作; T 表示卷积核大小为 1 的卷积操作; C_1 、 C_2 代表特征提取层 1 输入输出特征图的通道数; H_1 和 W_1 分别为特征

提取层 1 输入特征图的高度和宽度; H_2 和 W_2 分别为特征提取层 1 输出特征图的高度和宽度。

从式 (1) 可以看出, 经过输入输出特征图的融合, 特征提取层 2 获得的输入特征图信息兼顾了浅层的局部细节信息和网络深层的全局语义信息, 因此信息内容更加丰富。通过逐层引入跳转连接, 将网络浅层提取出的特征信息与网络深层的特征信息进行融合, 提高了特征的利用率, 使网络深层包含更多的细节信息, 从而提高网络对相似特征的推理分辨能力, 有效提升其识别精度。

2.2 坐标注意力机制模块的引入

虽然跳转连接操作可以将大量浅层特征信息逐步传递至深层, 但是这些被传递的特征信息并非都是有用的信息。所以, 在网络的特征提取单元中引入注意力机制模块可以帮助筛选出对网络训练更加有利的特征^[24]。特别的, 由于古代人特定的书写方式, 甲骨文字整体呈长条形, 特征图在宽度方向和高度方向上所包含的有用信息量相差甚远。因此, 为了结合甲骨文字“长条形”的特点进行有利于网络训练的特征筛选, 本文在网络中引入坐标注意力机制 (coordinate attention, CA) 模块。

坐标注意力机制模块与其他经典的注意力机制模块 (如 SE 注意力模块) 相比, 放弃了原有的二维全局池化操作, 而是将其分解为两个一维特征编码过程, 分别沿着宽度和高度两个空间方向聚合特征, 得到输入特征图在宽度方向和高度方向的注意力权重, 最终与原始特征图通过乘法加权计算, 得到在宽度和高度方向上带有注意力权重的特征图, 更好地契合了甲骨文字的形状特点, 并且具有计算成本低, 适用性好等优点。坐标注意力机制实现过程如图 6 所示。

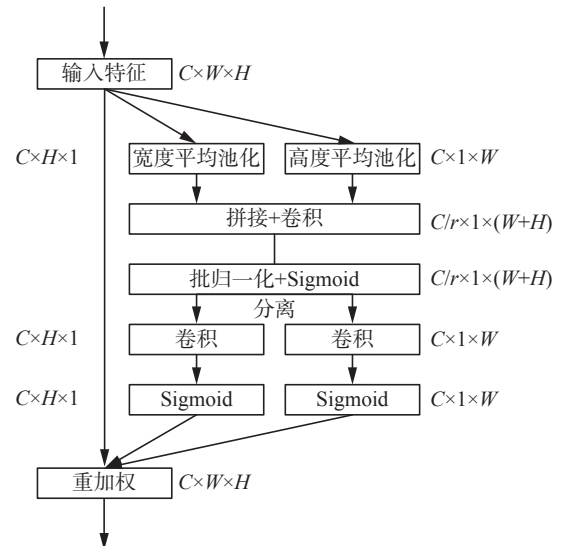


图 6 坐标注意力模块

Fig. 6 Coordinate attention module

1) 为了获取图像宽度和高度上的注意力并对精确位置信息进行编码, 坐标注意力模块利用池化核 $(H, 1)$ 和 $(1, W)$, 将输入特征图沿着宽度和高度两个方向分别集成特征, 最终获得在宽度和高度两个方向的特征图。宽度为 w 的第 c 通道的输出可以表示为

$$z_c^w(w) = \frac{1}{H} \sum_{0 \leq j \leq H} x_c(j, w)$$

同样的, 高度为 h 的第 c 通道的输出可以表示为

$$z_c^h(h) = \frac{1}{W} \sum_{0 \leq i \leq W} x_c(h, i)$$

2) 将获得全局感受野的宽度和高度两个方向的特征图拼接在一起, 将它们送入共享的卷积核为 1×1 的卷积模块, 将其维度降低为原来的 C/r , 最后将经过批量归一化处理的特征图送入 Sigmoid 激活函数, 得到形如 $1 \times (W+H) \times C/r$ 的特征图, 此阶段输出可以表示为

$$f = \delta(F_1([z^h, z^w]))$$

式中: δ 表示 Sigmoid 激活函数, F_1 表示共享的 1×1 卷积变换函数, $[\cdot, \cdot]$ 表示拼接操作, $f \in \mathbf{R}^{C/r \times (H+W)}$ 为包含宽度和高度空间信息的中间特征, r 为缩减因子。

3) 将得到的特征图 f 分为两个独立的特征图 f^w 和 f^h , 按照原来的宽度和高度进行卷积核为 1×1 的卷积, 分别得到通道数与原来一样的特征图, 经过 Sigmoid 激活函数后分别得到特征图在

宽度方向上的注意力权重 g^w 和高度方向上的注意力权重 g^h , 用公式表示为

$$g^w = \delta(F_w(f^w))$$

$$g^h = \delta(F_h(f^h))$$

式中: δ 表示 Sigmoid 激活函数, F_w 和 F_h 表示 1×1 的卷积变换。

4) 在原始特征图上通过乘法加权计算, 将得到最终在宽度和高度方向上带有注意力权重的特征图, 结果可以表示为

$$y_c(i, j) = x_c(i, j) \times g_c^h(i) \times g_c^w(j)$$

2.3 模型分类器结构的优化

原始的 ResNeSt 网络模型所设计的分类器部分包括卷积层、激活函数层、池化层和全连接层。输入图像后, 模型利用多层卷积层从复杂的训练数据中提取多维特征, 然后经过 ReLU 激活函数层提升模型的非线性表达能力, 最后通过池化层和全连接层对特征进行降维, 并且输出最终分类结果。在这个过程中, 包含特征信息的通道维数是在成倍增加的。本文所要识别的甲骨文字仅有 277 类, 而全连接层获得的通道维数却远超甲骨文字类别数, 多余的通道中必然夹杂着较多的冗余信息, 最终对识别精度产生了较大影响。因此, 针对本文实际识别任务需求, 对模型分类器结构进行了两个方面的优化, 删除网络最后一层的激活函数和全连接层, 并根据甲骨文字类别数对网络模型最后一层的卷积层输出通道数进行了重新设置, 最终设计的分类器结构如图 7。

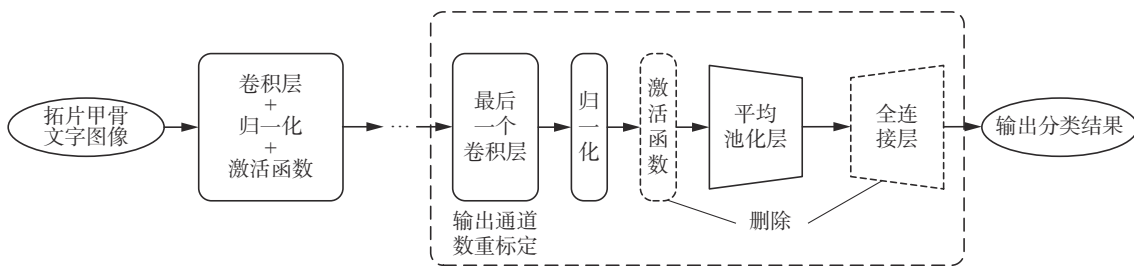


图 7 优化后的分类器结构

Fig. 7 Optimized classifier structure

这里选用 ReLU 激活函数来增加神经网络模型的非线性。与 Sigmoid 函数和 tanh 函数相比, ReLU 函数计算简单、更加高效且为非饱和函数, 可以有效克服梯度消失的问题, 并且加快了训练速度。ReLU 函数的公式为

$$\text{ReLU}(x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (2)$$

从式 (2) 中可以看到, 如果输入大于 0, 则直接返回作为输入提供的值; 如果输入是 0 或者负

数, 则返回值是 0。这就意味着输入负值后完全不激活, 部分关键信息将丢失, 最终降低了网络的分类精度。

同时, 在整个卷积神经网络中, 全连接层将前层 (卷积、池化等层) 计算得到的特征空间映射到样本标记空间。通过对输入输出通道数的设置, 达到线性降维的作用, 以适应类别的数量。由此可见, 全连接层此时在网络中是对前层所提取到的隐藏特征进行线性变换, 而 Zhu 等^[25] 提到卷积

层和池化层同样可以发挥类似的作用,因此全连接层在分类器中的作用是可以被卷积层和池化层替代的。更重要的是,根据 Zhu 等对全连接层输入的特征值的计算结果,关键特征值所占的比例超过了总特征值的 95%,而这些关键特征值集中在部分通道中。此时再利用全连接层进行线性降维,会大大降低网络对关键特征的利用能力,最终也将影响分类精度的提升。

基于上述两个方面的分析,本文去掉了网络最后一层中的 ReLU 激活函数和全连接层,将最后一个卷积层的输出通道数重新设置为甲骨文字类别数,经过平均池化层融合学到的深度特征后,最后对分类结果进行输出,完成了模型分类器结构的优化。

3 仿真实验及结果分析

为了验证本文创新工作的有效性和先进性,这里将进行 2 个方面的实验:1)分别验证引入的 ResNeSt 网络模型的有效性,并验证在 ResNeSt 网络模型基础上设计跳转连接结构、加入坐标注意力机制模块以及优化分类器结构的有效性;2)将本文提出的改进模型与目前识别性能较好的 3 种方法进行对比,验证本文提出改进模型的先进性。

3.1 数据集的选取与处理

本文使用的数据集 OBC306^[1]是目前拓片甲骨文字识别研究中唯一的大型公开数据集,其中包含 306 类拓片甲骨文字,共 309 551 个拓片甲骨文字符图像。少数高频字符(72 类,1 000 多幅图像)占图像总数的 83.74%,但仅占类别总数的 27.5%。此外,29 个类中只有一个图像。在本文的工作中,删除了上述提到的仅包含一个图像的所有类,保留了 277 个类,总共包含 309 552 个图像。除此之外,本文所用数据集中的图像因为分割的原因,分辨率各不相同,为增强数据的可靠性,将图像统一缩放到分辨率为 128 像素×128 像素。在实验过程中,将数据集按 4:1 的比例随机分为训练集和测试集,其中训练集包含 247 511 幅图像,测试集包含 62 011 幅图像。

3.2 实验条件

本文所有实验均在硬件配置为 NVIDIA GeForce GTX TITAN X, 16 GB 内存, Ubuntu 操作系统, Pytorch 深度学习框架的计算机上进行。本文算法训练时采用 Adam 优化函数,设置训练批次大小为 16,即神经网络每次取 16 个样本同时训练,迭代次数为 16 个 epoch,训练时间为 28 h。初始学习率设置为 0.001,每迭代 11 个 epoch 后衰减

为原来的 1/3。ResNeSt 中的超参数 K 设置为 1,超参数 R 设置为 2。

3.3 模型有效性验证

为了验证本文基于 ResNeSt 模型并设计跳转连接操作,引入坐标注意力机制模块和优化模型分类器结构的有效性,本文通过设计一系列实验,验证基于 ResNeSt 模型的拓片甲骨文字识别的有效性以及引入跳转连接操作的 ResNeSt-MS、引入坐标注意力机制的 ResNeSt-CA,优化了模型分类器结构的 ResNeSt-FCN 以及最终集合了上述 3 个改进点的 ResNeSt-Final 的有效性,并且在最后给出了改进之后识别正确,但在改进之前识别错误的示例。本文的评价指标选择的是整体准确率,计算公式为

$$A_{\text{Total}} = \frac{1}{H} \sum_{c=1}^C r_c$$

式中: H 代表了测试集的图像总数; C 代表了测试集图像的类别数; r_c 代表了测试集中模型识别正确的第 c 类的图像数量,数值越高代表模型的识别效果越好。

在进行模型有效性验证之前,本文首先对不同尺度图像的识别性能进行比较。如表 1 所示,分别对 64 像素×64 像素、128 像素×128 像素和 256 像素×256 像素这 3 种分辨率的拓片甲骨文图像进行了识别。从表中结果可以看出,图像分辨率是 64 像素×64 像素时,识别准确率比 128 像素×128 像素低了 0.44%,相差较多;而图像分辨率是 256 像素×256 像素时,识别准确率比 128 像素×128 像素仅仅高出 0.03%,但此时图像高度和宽度同时扩大,所带来的计算消耗将成倍增加,意义不大。因此,本文所设置的图像分辨率 128 像素×128 像素是从识别准确率和计算成本来看均比较合理的选择。

表 1 不同尺度图像的识别性能

Table 1 Recognition performance of images at different scales

图像分辨率/像素×像素	识别准确率/%
64×64	91.15
128×128	91.59
256×256	91.62

模型有效性验证结果如表 2 所示。从表 2 可以看出,ResNeSt 的识别准确率高达 91.59%,证明本文选取 ResNeSt 作为 baseline 的有效性。此外,引入了跳转连接操作之后,ResNeSt-MS 与 baseline 相比,识别准确率提高了 1.33%,证明本文所设计

的跳转连接结构能够提高模型对于局部细节特征的提取能力,此改进是有效的。

表 2 模型有效性验证
Table 2 Verification of model

模型	识别准确率/%
ResNeSt	91.59
ResNeSt-MS	92.92
ResNeSt-CA	91.86
ResNeSt-FCN	92.68
ResNeSt-Final	93.53

从表 2 中还可以看出,引入了坐标注意力机制之后,ResNeSt-CA 与 baseline 相比,识别准确率提升了 0.27%,证明本文在特征提取单元引入的坐标注意力机制模块能够有效提高模型在不同方向上对于有效特征的筛选能力。在对模型的分类器结构进行优化后,ResNeSt-FCN 与 baseline 相比,识别准确率提升了 1.09%,证明本文优化后的分类器可以有效地降低冗余信息的干扰,增强模型对有效特征的利用能力,从而提高模型的识别性能。

最后,将上述 3 个改进点同时引入到 baseline 后,ResNeSt-Final 与 baseline 相比,识别准确率提升了 1.94%,且相较于任意的单个改进点均有不同程度的提升,性能达到最优,从而证明本文提出的 3 个方面改进是有效的。

模型改进之前识别错误的示例如图 8 所示。

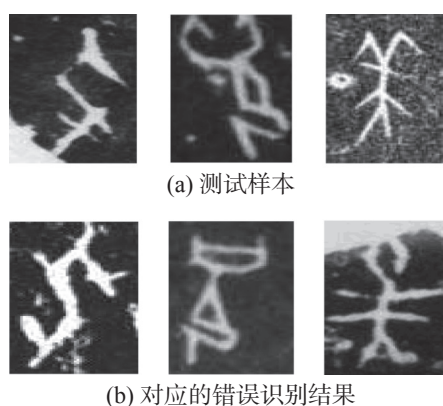


图 8 识别错误示例

Fig. 8 Examples of recognition effect comparison

图 8 中所示的是两行对应的 3 组字形相近的拓片甲骨文字。改进之前的模型将第一行甲骨文字错误的识别为了第二行对应位置的甲骨文字,在引入了上述 3 个方面的改进之后,图中相近的拓片甲骨文字均被正确识别,从而再一次证明本文提出的三个方面改进是有效的。

3.4 模型先进性验证

为验证本文改进算法的先进性,与目前最新的、综合识别性能较好的 3 种算法在公开数据集 OBC306 上进行了对比实验,对比结果如表 3 所示。

表 3 模型先进性验证
Table 3 Verification of model advancement

模型	识别类数	识别准确率/%
文献[14]	241	86.70
文献[15]	247	91.53
文献[16]	277	91.74
本文	277	93.53

由表 3 可以看出,在基于同一公开数据集的实验中,其他对比算法选取的识别类数最少为 241 类,最多为 277 类,本文选取的识别类数为 277 类,达到选取识别类数最多水平。在识别效果上,本文算法的识别率最高,并且提升幅度较大。与对比算法相比,本文提出的改进模型最小提升幅度达到 1.79%,最大提升幅度达到 6.83%。

综上所述,通过两个方面的仿真实验证明,本文提出的改进模型在识别准确率上达到目前最优的效果,改进具有有效性和先进性。

4 结束语

为了提升拓片甲骨文字的识别准确率,本文首次引入了目前较为先进的 ResNeSt 网络模型,并针对拓片甲骨文字中存在高相似度文字识别困难的问题进行了 3 个方面的改进。首先,设计了跳转连接操作,提高了网络模型捕捉局部细节信息的能力;其次,引入了坐标注意力机制模块,使模型更好地利用甲骨文字“长条形”的特点获得更多有用的特征信息;最后,对模型分类器进行了有效优化,提高了模型对于关键信息的利用能力。通过上述 3 个方面的改进,本文提出的算法在识别准确率上达到目前最优,为今后的拓片甲骨文字识别工作奠定了良好的基础。但是在目前研究工作当中,存在 29 类拓片甲骨文字仅有一张样本图像,无法用深度学习模型来进行训练和测试。同时由于原始兽骨、龟甲片年代久远,遭受损坏和侵蚀较为严重,目前拓片甲骨文字图像的噪声问题较为严重,影响了识别精度的提高。今后将进一步探索小样本学习方法,以及对先进的图像去噪方法进行研究,通过提高图像质量,使算法的识别效果进一步获得提升。

参考文献:

- [1] 谢乃和. 从殷墟走向世界的“绝学”甲骨文字研究: 韩国釜山“纪念甲骨文发现 120 周年国际学术研讨会”述评[J]. 管子学刊, 2020(3): 125–128.
XIE Naihe. Flourishing from Yin Ruins to the world—a review of “Busan, South Korea international symposium commemorating the 120th anniversary of oracle bone inscription discovery”[J]. Guanzi journal, 2020(3): 125–128.
- [2] HUANG Shuangping, WANG Haobin, LIU Yongge, et al. OBC306: a large-scale oracle bone character recognition dataset[C]//2019 International Conference on Document Analysis and Recognition. Sydney: IEEE, 2019: 681–688.
- [3] 顾绍通. 基于拓扑配准的甲骨文字形识别方法[J]. 计算机与数字工程, 2016, 44(10): 2001–2006.
GU Shaotong. Identification of oracle-bone script fonts based on topological registration[J]. Computer & digital engineering, 2016, 44(10): 2001–2006.
- [4] 刘永革, 刘国英. 基于 SVM 的甲骨文字识别[J]. 安阳师范学院学报, 2017(2): 54–56.
LIU Yongge, LIU Guoying. Oracle bone inscription recognition based on SVM[J]. Journal of Anyang Normal University, 2017(2): 54–56.
- [5] 顾绍通. 基于分形几何的甲骨文字形识别方法[J]. 中文信息学报, 2018, 32(10): 138–142.
GU Shaotong. Identification of oracle-bone script fonts based on fractal geometry[J]. Journal of Chinese information processing, 2018, 32(10): 138–142.
- [6] LIU Mengting, LIU Guoying, LIU Yongge, et al. Oracle bone inscriptions recognition based on deep convolutional neural network[J]. Journal of image and graphics, 2020, 8(4): 114–119.
- [7] YANG Zhen, FU Ting. Oracle detection and recognition based on improved Tiny-YOLOv4[C]//ICVIP 2020: 2020 The 4th International Conference on Video and Image Processing. New York: ACM, 2020: 128–133.
- [8] SUN Wenjie, ZHAI Guangtao, GAO Zhongpai, et al. Dual-view oracle bone script recognition system via temporal-spatial psychovisual modulation[C]//2020 IEEE Conference on Multimedia Information Processing and Retrieval. Shenzhen: IEEE, 2020: 193–198.
- [9] GAO Junheng, LIANG Xun. Distinguishing oracle variants based on the isomorphism and symmetry invariances of oracle-bone inscriptions[J]. IEEE access, 2020, 8: 152258–152275.
- [10] 刘梦婷. 基于深度卷积神经网络的甲骨文字识别研究[D]. 郑州: 郑州大学, 2020.
- LIU Mengting. Research on OBI recognition based on deep convolutional neural network[D]. Zhengzhou: Zhengzhou University, 2020.
- [11] 王琦琦. 基于深度卷积神经网络的甲骨文精确识别[D]. 南昌: 江西科技师范大学, 2020.
WANG Qiqi. Accurate recognition of oracle based on deep convolutional neural network[D]. Nanchang: Jiangxi Normal University of Science and Technology, 2020.
- [12] 高旭. 基于卷积神经网络的甲骨文识别研究与应用[D]. 长春: 吉林大学, 2021.
GAO Xu. Research and application of oracle bone inscriptions recognition based on convolution neural network[D]. Changchun: Jilin University, 2021.
- [13] 王浩彬. 基于深度学习的甲骨文检测与识别研究[D]. 广州: 华南理工大学, 2019.
WANG Haobin. The research of oracle bone script detection and recognition based on deep learning methods[D]. Guangzhou: South China University of Technology, 2019.
- [14] 张颐康, 张恒, 刘永革, 等. 基于跨模态深度度量学习的甲骨文字识别[J]. 自动化学报, 2021, 47(4): 791–800.
ZHANG Yikang, ZHANG Heng, LIU Yongge, et al. Oracle character recognition based on cross-modal deep metric learning[J]. Acta automatica sinica, 2021, 47(4): 791–800.
- [15] LIU Dazheng. Random polygon cover for oracle bone character recognition[C]//CSAI 2021: 2021 5th International Conference on Computer Science and Artificial Intelligence. New York: ACM, 2021: 138–142.
- [16] LI Jing, WANG Qiufeng, ZHANG Rui, et al. Mix-up augmentation for oracle character recognition with imbalanced data distribution[M]//Document Analysis and Recognition - ICDAR 2021. Cham: Springer International Publishing, 2021: 237–251.
- [17] ZHANG Hang, WU Chongruo, ZHANG Zhongyue, et al. ResNeSt: split-attention networks[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. New Orleans: IEEE, 2022: 2735–2745.
- [18] HOU Qibin, ZHOU Daquan, FENG Jiashi. Coordinate attention for efficient mobile network design[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville: IEEE, 2021: 13708–13717.
- [19] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016: 770–778.
- [20] GUO Menghao, XU Tianxing, LIU Jiangjiang, et al. Attention mechanisms in computer vision: a survey[J].

- [Computational visual media](#), 2022, 8(3): 331–368.
- [21] HU Jie, SHEN Li, SUN Gang. Squeeze-and-excitation networks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018: 7132–7141.
- [22] LI Xiang, WANG Wenhai, HU Xiaolin, et al. Selective kernel networks[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2020: 510–519.
- [23] HUANG Gao, LIU Zhuang, VAN DER MAATEN L, et al. Densely connected convolutional networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 2261–2269.
- [24] HUANG Zilong, WANG Xinggang, HUANG Lichao, et al. CCNet: criss-cross attention for semantic segmentation[C]//2019 IEEE/CVF International Conference on Computer Vision. Seoul: IEEE, 2020: 603–612.
- [25] ZHU Qiuyu, ZU Xuwen. Fully convolutional neural network structure and its loss function for image classification[J]. [IEEE access](#), 2022, 10: 35541–35549.

作者简介:



毛亚菲, 硕士研究生, 主要研究方向为甲骨文字识别、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、图像处理、机器学习。主持、参加国家和省部级重点科研项目 10 余项, 获省部级科学技术一等奖 1 项, 省部级科学技术二等奖 6 项, 发表学术论文 175 篇。