武汉大学学报(信息科学版)
Geomatics and Information Science of Wuhan University
ISSN 1671-8860,CN 42-1676/TN

《武汉大学学报(信息科学版)》网络首发论文

题目：自适应多级特征融合的场景古汉字识别
作者：涂超虎，易尧华，王凯丽，彭继兵，尹爱国
DOI：10.13203/j.whugis20230176
收稿日期：2023-10-26
网络首发日期：2023-12-09
引用格式：涂超虎，易尧华，王凯丽，彭继兵，尹爱国．自适应多级特征融合的场景古汉字识别[J/OL]．武汉大学学报(信息科学版).
https://doi.org/10.13203/j.whugis20230176

Doi: 10.13203/j.whugis20230176
引用格式：
涂超虎，易尧华，王凯丽，等. 自适应多级特征融合的场景古汉字识别[J]. 武汉大学学报（信息科学版），2023，Doi:10.13203/j.whugis20230176.（TU Chaohu，YI Yaohua，WANG Kaili，et al. Adaptive Multi-level Feature Fusion Based Scene Ancient Chinese Text Recognition[J] *Geomatics and Information Science of Wuhan University*, 2023，Doi:10.13203/j.whugis20230176.）

自适应多级特征融合的场景古汉字识别

涂超虎[1,2] 易尧华[1,2] 王凯丽[3] 彭继兵[1,2] 尹爱国[2,4]

1 武汉大学遥感信息工程学院，湖北 武汉，430079
2 武汉大学数字成像与智能感知研究中心，湖北 武汉，430079
3 武汉工程大学计算机科学与工程学院，湖北 武汉，430205
4 珠海奔图电子有限公司，广东 珠海，519060

摘 要：自然场景中的古汉字图像具有背景复杂、字符数量庞大、书体形式多样的特点，其字符与书体形式众多导致了文本结构复杂度不同，现有研究方法未针对性解决复杂结构古汉字的识别难题。针对这一问题，本文提出一种自适应多级特征融合网络，首先根据古汉字的结构复杂度，自适应选择融合古汉字浅层细节信息和高层语义信息，获取古汉字的高区分度特征，提高模型对古汉字的识别能力。然后使用最大边界余弦损失，增大古汉字的类间间距，提高模型对相似结构古汉字特征的判别能力。实验结果表明，本文方法在多场景古汉字数据集上 Top-1 识别准确率为79.58%，与目前最优方法相比提高了 3.27%，提高了场景古汉字的识别准确率。

关键词：场景古汉字识别；自适应多级特征融合；最大边界余弦损失

Adaptive Multi-level Feature Fusion for Scene Ancient Chinese Text Recognition

TU Chaohu[1,2] *YI Yaohua*[1,2] *WANG Kaili*[3] *PENG Jibing*[1,2] *YIN Aiguo*[2,4]

1 School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China
2 Digital Imaging and Intelligent Perception Research Center, Wuhan University, Wuhan 430079, China
3 School of Computer Science and Engineering, Wuhan Institute of Technology, Wuhan 430205, China
4 Zhuhai Pantum Electronics Co. , Ltd, Zhuhai 519060, China

Abstract: Objectives: Ancient Chinese text are widely distributed in inscriptions, couplets, stone engravings and other scenes, which have the characteristics of complex background, large number of characters, and diverse writing forms. The large number of characters and writing forms directly lead to the difference in text structure complexity. **Methods:** To solve the difficulty of recognizing ancient Chinese text with complex structures, we propose an adaptive multilevel feature fusion network. First, ResNet152 is the main backbone network, and its deeper network and residual structure can fit more parameters to learn the features of ancient Chinese text and avoid the degradation of the model. Second, according to the structural complexity of ancient Chinese text, the importance of each feature map is automatically obtained through learning, so that the model adaptively selects and merges the shallow detail information and high-level

收稿日期：2023-10-26
项目资助：国家重点研发计划(2021YFB2206200)。
第一作者：涂超虎，硕士生，主要从事古汉字识别相关研究。chaohu.tu@whu.edu.cn
通讯作者：易尧华，博士，教授。yyh@whu.edu.cn

semantic information of ancient Chinese text, obtains the high discrimination features of ancient Chinese text and improves the recognition ability of the model. Finally, the maximum boundary cosine loss is used to minimize the cosine similarity between different ancient Chinese text, increase the inter-class distance of ancient Chinese text, and reduce the intra-class distance between similar Chinese text. Combined with the cross entropy loss function as a loss function, the model can improve the discrimination ability of ancient Chinese text with similar structures. **Results:** The experimental results show that when the multistage feature fusion module is added to the proposed method, the Top-1 accuracy rate is increased by 1.59%, and when the maximum boundary cosine loss function is added, the Top-1 accuracy rate is increased by 1.09%. The best effect of Top-1 identification accuracy rate on the multi-scene ancient Chinese character dataset is 79.58%. Compared with the current optimal method, it improves the recognition accuracy of scene ancient Chinese text by 3.27%. **Conclusions:** In this paper, a multistage feature fusion network is designed to improve the feature extraction ability of the model, and the maximum boundary cosine loss is introduced to increase the distance between ancient Chinese text and narrow the distance within ancient Chinese text.
Key words: multi-scene ancient Chinese text; adaptive multi-level feature fusion; large margin cosine loss

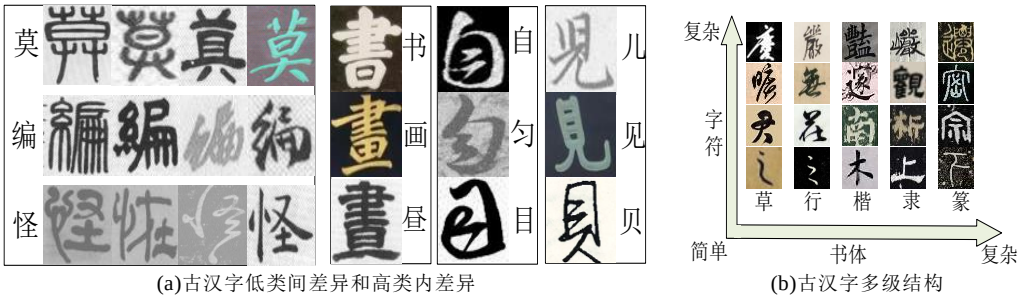
文本识别技术已逐渐成熟，被应用于图像检索、书籍文档数字化、自动驾驶等众多场景中[1-3]。与英文、现代汉字相比，古汉字作为文本识别的一种特殊研究对象，具有象形甲骨文、草书、隶书等多种书体形式，其字体形式多变，字符数量庞大；与背景简单的文档古汉字相比，分布场景复杂，广泛分布在碑文、古籍、对联等场景中，具有笔画缺失、形状不规则等特点[4-5]。针对场景古汉字进行研究，能够帮助人们识别自然场景中的古汉字，帮助研究人员进行古籍整理、古籍数字化等工作，因此场景古汉字识别具有重大意义。

场景古汉字是指分布在碑文、牌匾、石刻等多个场景中的篆书、隶书、行书等汉字，具有以下特点：自然场景古汉字图像背景复杂、书体众多、字符数量庞大，多书体、字体导致同一汉字类内差异大，二维“簇块状”结构的庞大字符量导致类间相似性高，如图 2(a)所示。众多书体、字符下古汉字的结构存在层次性，复杂度不一。如图 2(b)所示，与其他书体相比，篆书结构复杂度最高，同一书体中，某些字符如“木”、“上”等结构简单，“严”、“艳”等字符结构复杂。

针对古汉字识别，国内外学者提出了基于传统技术的方法和基于深度学习的方法。基于传统技术的古汉字识别方法依赖于传统图形处理技术，赵若晴等人[6]提出了一种基于方向梯度直方图和灰度共生矩阵的金文识别方法，融合文字整体特征和局部特征的方向梯度直方图和灰度共生矩阵，使用支持向量机进行分类；陈丹等人[7]提出了用于识别联机手写古汉字的方法，通过笔型特征、笔画交叉点和字元相对位置特征与模板进行匹配，对古汉字进行识别。

图 1 多书体多场景古汉字

Fig. 1 Multi-style and Multi-scene Ancient Chinese Text



(a)古汉字低类间差异和高类内差异　　(b)古汉字多级结构

图 2 场景古汉字特点

Fig. 2 Characteristics of Ancient Chinese Text in Scene

这些传统方法结合古汉字的笔画、部首等局部特征，对古汉字结构进行拆解，能对简单古汉字进行拆分，但不易实现结构复杂古汉字的拆分，拆分得到的笔画较多且不规则，无法建立合适的模板。

随着深度学习发展，Inception[8]、ResNet[9]和 DenseNet[10]等网络被广泛用于文本识别中，基于深度学习的古汉字识别方法通过反复学习古汉字的结构特征，了解每个汉字的规律，对其准确识别，性能优于传统方法。Nguyen 等人[11]提出了 CAGAN 网络，使用 GAN 网络[12]对古籍中缺失的汉字部件进行重构，修复汉字图像，并采用 ResNet 进行识别；Ma 等人[13]通过版面布局信息辅助汉字特征提取；陈娅娅等人[14]使用 ResNet 为特征提取网络，为了使模型学习更多的特征，引入迁移学习进行数据增强，实现古印章识别。Tang 等人[15]提出了一种基于注意力机制的轻量级人工神经网络模型 ShuiNet-A，该模型结合通道和空间维度提取关键特征，用于水书(一种象形文字)的识别。这些方法仅以特定场景的古汉字为研究对象，缺乏对全类别、全场景的古汉字的研究。由于全类别古汉字类别较多，同一个字的不同书体结构不同，较大的类内差异会给识别任务带来困难。Wang 等人[16]提出多模型融合方法，通过多个模型的特征提取能力对古汉字进行识别，但缺少对古汉字样本特征的具体分析；Wang 等人[17]提出基于域适应与交叉域融合的多场景古汉字识别方法，但未对古汉字多级复杂特征进行具体研究。

改进特征融合方式是提高模型性能的重要途经，比如 Shi 等人[18]利用 SENet 融合注意力特征，Yang 等人[19]利用多级特征融合 Transformer 和 CNN，融合全局信息和特征信息，Liu 等人[20]也融合了多层次和多尺度特征，提高图像分类的性能。

因此，为解决场景古汉字的多级复杂结构和相似结构给识别任务带来的困难，本文提出自适应多级特征融合(Adaptive Multilevel Feature Fusion，AMFF)的场景古汉字识别方法，首先提取古汉字的多级特征，然后使用自适应融合方法(Adaptive Fusion，AF)获得古汉字中更具判别力的综合特征，最后使用最大边界余弦损失函数增大古汉字的类间差异，优化模型特征提取能力，提高场景古汉字识别准确率。实验结果表明，在多场景古汉字数据集中，本文算法 Top-1 识别准确率为 79.58%，与最先进的方法相比提高了 3.27%，达到场景古汉字识别任务的最好结果。

1 本文算法

为了准确提取古汉字的多级复杂结构，区分相似结构古汉字，本文设计了自适应多级特征融合网络，以 ResNet152 为主干网络，设计多级特征融合方法(Multilevel Feature Fusion，MFF)提取古汉字的多级特征，使用自适应特征融合方法选取关键特征信息，并引入最大边界余弦损失函数，提高模型对相似古汉字的判断能力，如图 3 所示。

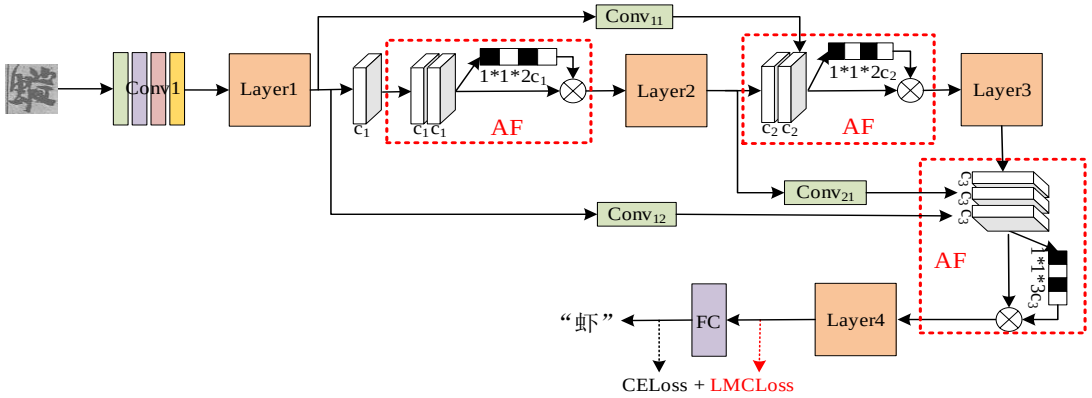


图 3 自适应多级特征融合网络

Fig. 3 Adaptive Multi-level Feature Fusion Network

1.1 自适应多级特征融合网络

古汉字笔画较多、结构复杂，现有方法容易忽略了对细节信息的提取，且对结构较为复杂的汉字识别能力较弱，故本文提出多级特征融合方法，融合浅层细节信息和深层语义信息，一定程度上提高了古汉字的识别效果。

本文选取 ResNet152 网络作为主干网络，该网络较深，能够学习到更多的参数来拟合古汉字复杂的笔画特征；并采用 Bottleneck 残差结构，使用跳跃连接在输出结果中加入原始特征映射，具有更好的古汉字特征提取效果。深度学习网络中，浅层网络可以提取更好的细节信息，如古汉字的笔画，帮助辨别形近字，但语义信息较少；深层特征具有丰富的语义信息，但细节信息较少。故为了有效结合两者的信息，本文提出了两种多级特征融合的策略。

第一种策略如图 4 所示，为了使对应特征图的大小维度相同，对浅层 Layer 的输出做卷积处理，并将每层 Layer 前面所有层的输出直接相加作为深层 Layer 的输入，结合浅层 Layer 的细节信息和深层 Layer 的语义信息，有助于融合对古汉字的浅层、深层信息特征，从而提高对古汉字识别的准确率。

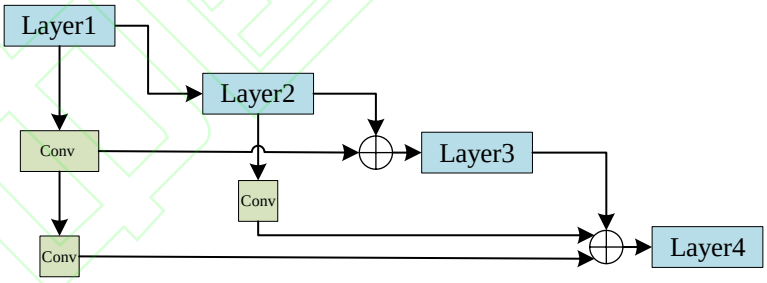


图 4 多级特征融合

Fig. 4 Multi-level Feature Fusion

由于浅层特征中含有较多的噪声，会给识别带来一定的影响，而高层语义信息对于识别较为重要，第一种策略仅将不同层级的特征图直接相加，给予两类特征图相同的权重，会增大细节中的噪声对结果的影响，由于不同古汉字结构复杂度不同，细节信息的重要程度也不相同，无法直接对不同层级的特征图进行固定的权重分配。故本文引入第二种策略——自适应特征融合方法，通过学习的方式自动获取到对每个特征图的重要程度，使模型自适应调整不同分支特征图的权重[21]，原理如图 5 所示。

由于自适应融合方法所处位置不同，故需要对特征图分别预处理，如式(1)所示，其中，x_0、x_1、x_2、x_3 为输入的特征图，n 为输入特征图的数量，n' 为加权融合特征图的数量，Y_1、Y_2、Y_3 为预处理后的特征图，$Conv$、$Conv'$为卷积映射，$Conv_{ij}$ 为第 i 个 Layer 的第 j 个卷积映射，以下以 $n=2$ 为例。

$$\begin{cases} n'=2, Y_1=Conv(x_0), \quad Y_2=Conv'(x_0) & n=1 \\ n'=2, Y_1=Conv_{11}(x_1), \quad Y_2=x_1 & n=2 \\ n'=3, Y_1=Conv_{12}(x_1), \quad Y_2=Conv_{21}(x_2), \quad Y_3=x_3 & n=3 \end{cases} \tag{1}$$

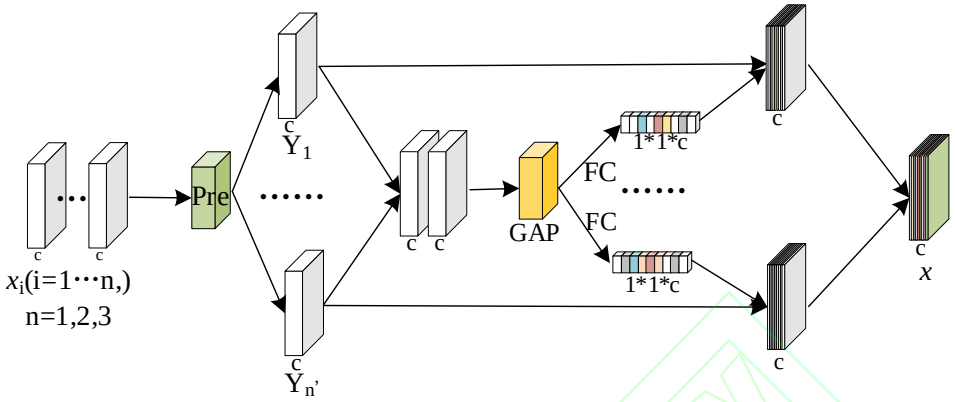


图 5 自适应融合方法
Fig. 5 Adaptive Fusion Method

为使模型自适应调整不同层级特征图的重要程度，需要将两特征图 Y_1、Y_2 相加，生成 Y'，通过全局平均池化生成尺寸为 $1\times1\times C$ 的特征图(C 为通道数)，然后由全连接层得到紧凑特征 m，使其能进行精确的自适应特征选择，最后通过全连接层和 Softmax 函数生成不同向量 a、b，作为施加给 Y_1 和 Y_2 的权重。公式如下：

$$m = f_{fc}(f_{gp}(Y')) \tag{2}$$

$$a=\frac{e^{Am}}{e^{Am}+e^{Bm}},\ b=\frac{e^{Bm}}{e^{Am}+e^{Bm}} \tag{3}$$

其中，f_{gp}、f_{fc} 为全局平均池化、全连接层的映射函数，A、B 为全连接层的权重。

为提升对有用特征的重视程度，并抑制对识别结果贡献较小的特征，需对不同层级的特征图进行加权融合，公式如下：

$$x=a\cdot Y_1'+b\cdot Y_2',\ a+b=1 \tag{4}$$

结合自适应融合方法，得到自适应多级特征融合模型如图 6。

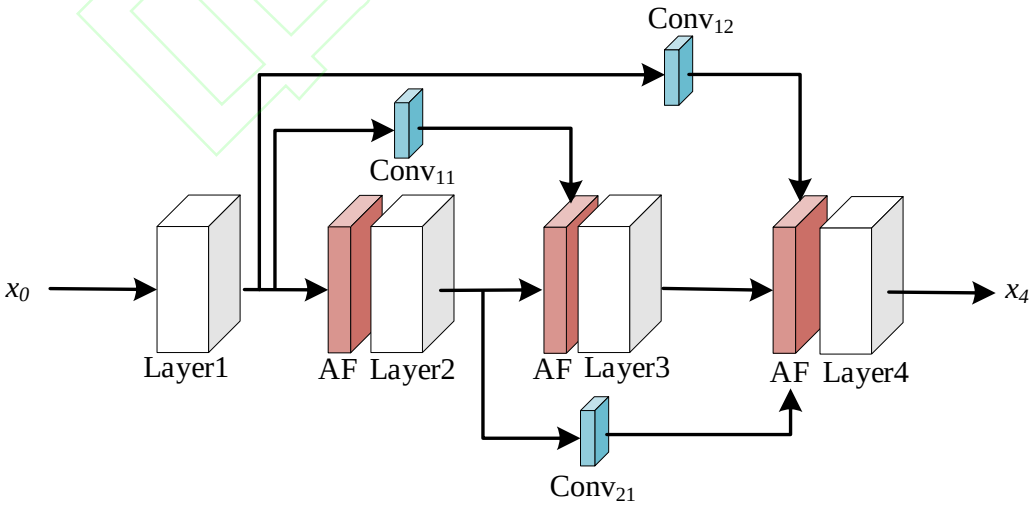


图 6 自适应多级特征融合模型
Fig. 6 Adaptive Multi-level Feature Fusion Model

1.2 损失函数

为增大古汉字的类间间距，对结构相似的汉字进行准确识别，本文使用最大边界余弦损失函数，并结合交叉熵损失对模

型进行监督，损失函数如下：

$$L = L_{CE} + \alpha \cdot L_{LMC} \tag{5}$$

其中，L_{CE}、L_{LMC} 分别为交叉熵损失函数、最大边界余弦损失函数；α 为最大边界余弦损失函数的系数，本文设置为0.2。

1.2.1 交叉熵损失函数

交叉熵损失函数公式如下：

$$L_{CE} = \frac{1}{N}\sum_{i} L_i = -\frac{1}{N}\sum_{i}\sum_{c=1}^{M} y_{ic}\log(p_{ic}) \tag{6}$$

其中，M 为类别的数量，在本文中 M =3755(GB2312 字符的数量)；y_{ic} 为真实标签，如果图像 i 的类别为 c 时 y_{ic}=1，否则为 0；p_{ic} 为模型预测图像 i 属于类别 c 的概率。

实际训练中，当特征提取不足、模型泛化能力较弱时，模型容易过拟合。为了缓解这个问题，本文使用标签平滑[22]的正则化方法，软化真实标签的编码方式，则公式(14)中的 L_i 变为：

$$L_i = \begin{cases} (1-\beta)\cdot\sum_{c=1}^{M} y_{ic}\log(p_{ic}), & i = c \\ \beta\cdot\sum_{c=1}^{M} y_{ic}\log(p_{ic}), & i \neq c \end{cases} \tag{7}$$

其中，β 为超参数，本文设置为 0.1。

1.2.2 最大边界余弦损失函数

为了提升模型对相似结构古汉字的区分能力，本文使用最大边界余弦损失函数[23]，使用余弦相似度作为度量指标，对于相同汉字的不同书体形式，选择最大的余弦相似度作为损失函数的一部分；对于不同古汉字，选择最小的余弦相似度作为损失函数的另一部分。通过最小化不同古汉字之间的余弦相似度，最大边界余弦损失可以训练模型更好地捕捉到不同古汉字的特征，提高模型在场景古汉字识别中的性能。如公式(8)所示：

$$L_{LMC} = -\frac{1}{N}\sum_{i=1}\log\frac{e^{\|x\|\cos(\theta_{p_i}-m)}}{e^{\|x\|\cos(\theta_{p_i}-m)} + \sum_{j\neq p_i} e^{\|x\|\cos\theta_j}} \tag{8}$$

其中，N 为训练集数量，p_i 为输入图像 i 的标签，x 为特征向量，θ_j 为全连接层和 x 的夹角，m 为一个正值，是对余弦函数施加的约束。

2 实验结果与分析

2.1 数据集与评价指标

2.1.1 数据集

本文使用的数据集为多场景古汉字（multi-scene ancient Chinese text，MACT）数据集[17]，该数据集包含篆、隶、楷、行、草等多种书体和古籍、石碑、墓碑、雕刻等多种场景的样本，因保存不完整或受到环境侵蚀的影响，多数样本存在模糊、失真等现象，识别难度较高。

该数据集为单字数据集，共有 138935 张训练集图像和 14318 张测试集图像，其中训练集通过人工生成，共有 5 种书体、37 种字体，测试集为真实场景文本图像。图像尺寸统一为 64×64×3。因 MACT 数据集是目前场景古汉字识别领域较为完善的数据集，因此本文测试结果对场景古汉字识别具有较强的参考意义。

2.1.2 评价指标

文字识别领域的评价指标为识别准确率，本文将 Top1-Top5 识别准确率作为评价指标，公式表示为：

$$P_{Top\,i} = \frac{\sum_{j=1}^{M} R(j,\max_i)}{M},\quad i\in[1,5] \tag{9}$$

$$R(j,\max_i) = \begin{cases} 1, & c\in\max_i(p_j) \\ 0, & c\notin\max_i(p_j) \end{cases} \tag{10}$$

其中，M 为测试集所有样本的数量，max_i 为数值最大的前 i 个置信度对应的标签，R 为判断函数，若预测的标签 p_j 内含有真实标签 c 为 1，反之为 0。

2.2 实验细节

本实验的训练与测试是基于 Ubuntu20.04 系统的 Pytorch 平台，使用了 NVIDIA RTX3090 的服务器和 NVIDIA RTX3060 的工作站。本实验使用的优化函数为随机梯度下降法，学习率为 0.01，动量为 0.9，学习率调整方式为按需调整学习率，学习率调整倍数为 0.1，批次为 16，图片大小为 224×224，最大边界余弦损失函数系数 α 为 0.2。本实验对训练集进行了随机裁剪、随机翻转和数据归一化等基础数据增强处理。

2.3 实验结果及分析

为了验证算法的先进性，本文在

MACT 数据集上与其他方法进行对比，如表 1 所示。MME 通过改进特征提取方法，融合多模型特征提高识别结果，与 MME 方法相比，AMFF 的 Top-1 准确率提高了 6.22%，说明 AMFF 在改进特征提取的基础上使用最大边界余弦损失函数，增大了古汉字的类间间距，增强了模型对相似结构古汉字的区分能力，提高了古汉字识别准确率；CA-CF 方法在古汉字类别特征分布方面进行改进，与 CA-CF 方法相比，AMFF 的 Top-1 准确率提高了 3.27%，说明 AMFF 中自适应多级特征融合网络能有效融合不同结构复杂度古汉字的细节信息和语义信息，对其具有较强的特征提取能力，大幅度提高了古汉字识别准确率。部分识别结果如图 7 所示，实验结果表明 AMFF 能有效提高场景古汉字的识别准确率。

表1 MACT数据集实验结果

Tab. 1 Experimental Results on the MACT Dataset

方法	准确率/%				
	Top-1	Top-2	Top-3	Top-4	Top-5
人类主观视觉识别[17]	52.98	-	-	-	-
Alexnet_V2[17]	41.59	50.58	55.29	58.33	41.67
OverFeat[17]	53.79	64.65	69.56	72.99	75.10
Inception_v1[17]	66.88	75.78	79.65	81.90	83.24
Inception_v4[17]	66.94	75.30	78.90	81.06	82.61
MME[17]	73.36	81.55	84.94	86.76	87.91
CA-CF[18]	76.31	83.05	86.16	87.72	88.73
Baseline	75.69	82.75	85.72	87.32	88.39
AMFF(Ours)	**79.58**	**85.00**	**87.28**	**88.52**	**89.42**

注：粗体表示最优值

图 7 识别正确样本举例(下侧标签为“预测结果-置信度”)

Fig.7 Examples of Recognizing Correct Samples (Labeled "Predicted Result - Confidence" on the Lower Side)

表 2 消融实验对比

Tab. 2 Comparison of Ablation Experiments

方法	模型	损失函数	准确率/%				
			Top-1	Top-2	Top-3	Top-4	Top-5
1	Baseline	L_{CE}	75.69	82.75	85.72	87.32	88.39
2	Baseline +MFF	L_{CE}	76.28	83.38	85.91	87.47	88.72
3	Baseline +AF	L_{CE}	76.83	83.87	86.57	88.03	89.01
4	Baseline +MFF+AF	L_{CE}	77.28	84.77	86.82	88.18	88.98
5	Baseline	$L_{CE}+L_{LMC}$	76.78	83.65	86.37	88.09	89.15
6	Baseline +MFF	$L_{CE}+L_{LMC}$	77.27	84.45	86.97	88.45	89.37
7	Baseline +AF	$L_{CE}+L_{LMC}$	77.34	84.19	86.70	88.26	89.25
8	Baseline +MFF+AF	$L_{CE}+L_{LMC}$	**79.58**	**85.00**	**87.28**	**88.52**	**89.42**

注：本文将 ResNet152 作为 Baseline，粗体表示最优值

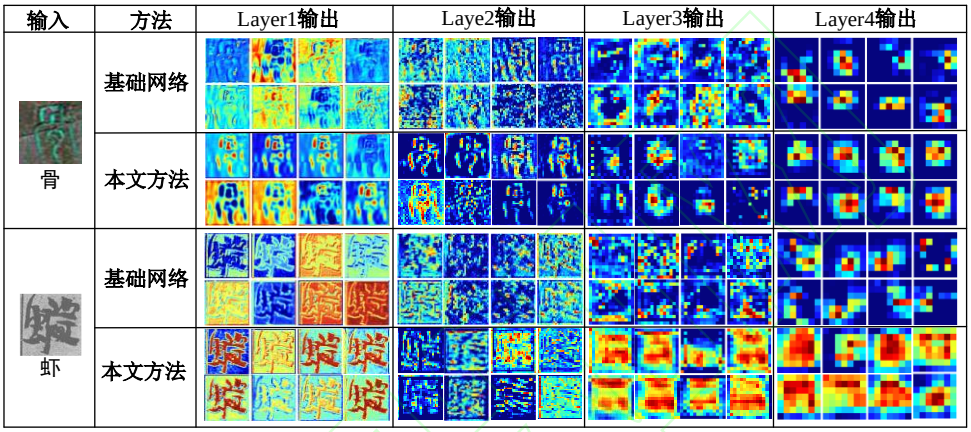


图 8 本文方法与基础网络中各 Layer 输出的热力图

Fig. 8 Partial Heat Maps of Output at Each Layer in the Method Presented in this Paper and Baseline

2.4 消融实验

本文实验由三部分组成：多级特征融合(MFF)、自适应融合(AF)和最大边界余弦损失函数(L_{LMC})。本文增加了几组消融实验来验证各部分对实验结果的影响，消融实验结果如表 2 所示。

表 1 中方法 1 和方法 2 的结果表明，加入 MFF 后，Top-1 准确率提高了 0.59%，说明结合细节信息和语义信息可以提高模型对古汉字不同复杂度特征的提取效果。方法 2 和方法 4 的结果表明，加入 AF 后，Top-1 准确率提高了 1.00%，说明自适应融合方法可以使模型自动选择对识别任务有用的古汉字特征，提高了模型对不同复杂度古汉字的判断能力。方法 1 和方法 5 的结果表明，加入 L_{LMC} 后，Top-1 准确率提高了 1.09%，说明最大边界余弦损失能优化模型对相似结构古汉字的辨别能力，提高场景古汉字的识别准确率。

如图 8 所示，与基础网络相比，本文方法在 Layer1 层提取的古汉字浅层轮廓特征更加明显，在 Layer2 层聚合的汉字构形特征更清晰，在 Layer3、Layer4 层提取的深层特征信息更多，表明本文方法对汉字特征的提取聚合能力更好，对不同复杂度古汉字的特征区分能力更强，有效地提高了场景古汉字识别的准确率。

2.5 识别错误结果分析

虽然AMFF在MACT数据集上取得了很好的效果，能识别出结构比较复杂和因书体字体不同导致类内差异较大的古汉字，但某些古汉字未能识别成功，如图9所示。总结原因可知，这些汉字的笔画、部首或部件非常相似，如“苟-苛”，仅在笔画层级上多了一个“丿”；“凉-谅”，仅部首有所区别，在书写后“冫”“讠”极为相似；如“阁-闽”，前者中心部件为“各”，后者为“虫”，书写后较为相似，不易区分。

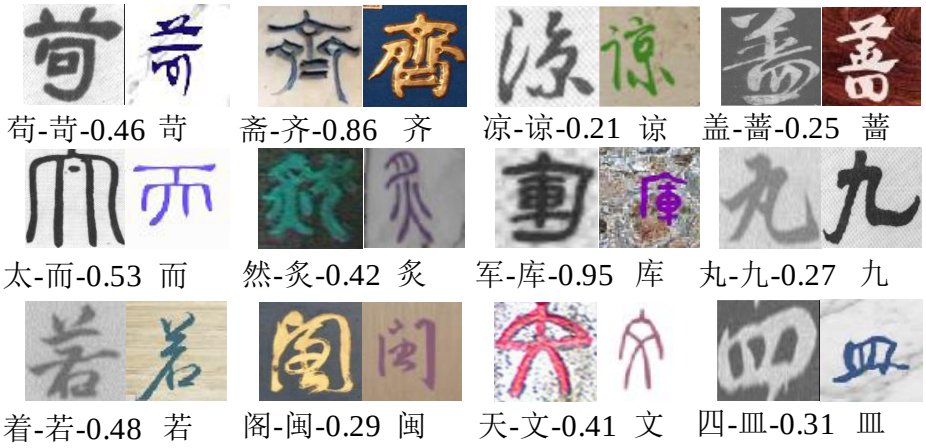


图 9 识别错误样本举例(下侧标签为”标签-Top-1 结果-置信度”，右图为 Top-1 结果对应的图像)
Fig. 9 Examples of Recognition Errors (the Label on the Lower Side is "Label-Top-1 Result-Confidence", and the Image Corresponding to Top-1 Result is Shown on the Right)

3 结 语

本文提出了一种自适应多级特征融合的场景古汉字识别方法。针对结构复杂度不同的古汉字，提出自适应多级特征融合方法，融合古汉字的细节特征和语义信息；采用最大边界余弦损失函数，增大模型对汉字类间差异的区分度，提高场景古汉字识别的准确率。实验结果表明，该方法在 MACT 数据集上的 Top-1 识别准确率超过了目前最优方法，实现了场景古汉字的准确识别。本文仅以整个汉字为研究对象，但古汉字拥有较为复杂的二维拓扑结构，由汉字部件构成，后续拟结合古汉字的部件特征，使用部件推理汉字，并采用古汉字整体与局部特征相结合的方法进行古汉字识别。

参 考 文 献

[1] Mishra A, Alahari K, Jawahar C V. Image retrieval using textual cues[C]//Proceedings of the IEEE international conference on computer vision. 2013: 3040-3047.

[2] Wang B, Ma Y W, Hu H T. Hybrid model for Chinese character recognition based on Tesseract-OCR[J]. *International Journal of Internet Protocol Technology*, 2020, 13(2): 102-108.

[3] Gómez L, Mafla A, Rusinol M, et al. Single shot scene text retrieval[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 700-715.

[4] Yang L, Ma J, Xu T, et al. Fast searching Chinese calligraphic information based on character recognition[C]// IEEE International Conference on Applied System Invention (ICASI). IEEE, 2018: 358-361.

[5] Huang J, Cheng G, Zhang J, et al. Recognition method for stone carved calligraphy characters based on a convolutional neural network[J]. *Neural Computing and Applications*, 2023, 35(12): 8723-8732.

[6] Zhao R Q, Wang H Q, Wang K, et al. Recognition of bronze inscriptions image based on mixed features of histogram of oriented gradient and gray level co-occurrence matrix[J]. *Laser & Optoelectronics Progress*, 2020, 57(12): 121003.（赵若晴, 王慧琴, 王可,等. 基于方向梯度直方图和灰度共生矩阵混合特征的金文图像识别[J]. 激光与光电子学进展, 2020, 57(12): 121003.）

[7] Chen D, Li N, Li L et al. Online recognition of ancient characters[J]. *Joumal of Beijing Institute of Machinery*(4) :32-37.(陈丹,李宁,李亮. 古文字的联机手写识别研究[J]. 北京机械工业学院学报(4):32-37.)

[8] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 1-9.

[9] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[10] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 4700-4708.

[11] Nguyen K C, Nguyen C T, Hotta S, et al. A character attention generative adversarial network for degraded historical document restoration[C]//2019 International Conference on Document Analysis and Recognition (ICDAR). IEEE, 2019: 420-425.

[12] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[J]. *Advances in neural information processing systems*, 2014, 27.

[13] Ma W, Zhang H, Jin L, et al. Joint layout analysis, character detection and recognition for historical document digitization[C]//2020 17th International Conference on Frontiers in Handwriting Recognition (ICFHR). IEEE, 2020: 31-36.

[14] Chen Y Y, Liu Q X, Wang K L, Yi Y H, et al. Historical Chinese Seal Text Recognition based on ResNet and Transfer Learning[J]. *Computer Engineering and Applications*,2022,58 (10) :125-131.(陈娅娅,刘全香,王凯丽,等.基于 ResNet 和迁移学习的古印章文本识别[J].计算机工程与应用,2022,58(10):125-131.)

[15] Tang M, Xie S, Liu X. Ancient Character Recognition: A Novel Image Dataset of Shui Manuscript Characters and Classification Model[J]. *Chinese Journal of Electronics*, 2023, 32(1): 64-75.
[16] Wang K, Yi Y, Liu J, et al. Multi-scene ancient chinese text recognition[J]. *Neurocomputing*, 2020, 377: 64-72.
[17] Wang K, Yi Y, Tang Z, et al. Multi-scene ancient Chinese text recognition with deep coupled alignments[J]. *Applied Soft Computing*, 2021, 108: 107475.
[18] Shi Y X, Zhou W X, Shao Z F. Multi-view remote sensing image scene classification by fusing multi-scale attention [J].*Geomatics and Information Science of Wuhan University*, 2023, DOI: 10.13203/j.whugis20220737(时永欣，周维勋，邵振峰. 融合多尺度注意力的多视角遥感影像场景分类[J]. 武汉大学学报（信息科学版），2023, DOI:10.13203/j.whugis20220737)
[19] Yang H, Yu H, Zheng K, et al. Hyperspectral Image Classification Based on Interactive Transformer and CNN with Multilevel Feature Fusion Network[J]. *IEEE Geoscience and Remote Sensing Letters*, 2023.
[20] Liu S, Zhang Q, Huang L. Graphic image classification method based on an attention mechanism and fusion of multilevel and multiscale deep features[J]. *Computer Communications*, 2023, 209: 230-238.
[21] Li X, Wang W, Hu X, et al. Selective kernel networks[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 510-519.
[22] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 2818-2826.
[23] Wang H, Wang Y, Zhou Z, et al. Cosface: Large margin cosine loss for deep face recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 5265-5274.

网络首发：
标题：自适应多级特征融合的场景古汉字识别
作者：涂超虎，易尧华，王凯丽，彭继兵，尹爱国
Doi: 10.13203/j.whugis20230176
收稿日期：2023-10-26
引用格式：
涂超虎，易尧华，王凯丽，等. 自适应多级特征融合的场景古汉字识别[J]. 武汉大学学报（信息科学版），2023，Doi:10.13203/j.whugis20230176.（TU Chaohu，YI Yaohua，WANG Kaili，et al. Adaptive Multi-level Feature Fusion Based Scene Ancient Chinese Text Recognition[J] *Geomatics and Information Science of Wuhan University*, 2023，Doi:10.13203/j.whugis20230176.）

您感兴趣的其他相关论文：

一种结合低级视觉特征和 PAPCNN 的 NSST 域遥感影像融合方法
侯昭阳，吕开云，龚循强，支君豪，王楠
武汉大学学报(信息科学版), 2023, 48(6): 960-969.
http://ch.whu.edu.cn/cn/article/doi/10.13203/j.whugis20220168

融合多尺度注意力的多视角遥感影像场景分类
时永欣，周维勋，邵振峰
武汉大学学报(信息科学版). doi: 10.13203/j.whugis20220737
http://ch.whu.edu.cn/cn/article/doi/10.13203/j.whugis20220737

面向多源数据地物提取的遥感知识感知与多尺度特征融合网络
龚健雅，张展，贾浩巍，周桓，赵元昕，熊汉江
武汉大学学报(信息科学版), 2022, 47(10): 1546-1554.
http://ch.whu.edu.cn/cn/article/doi/10.13203/j.whugis20220580