

# 基于跨模态学习的甲骨文识别系统的研发

指导老师：刘江 章晓庆

组员：桂宸麓 孙清扬

## 一、 研究背景

甲骨文是早在商周时期就已出现的文字，是世界上最古老的文字之一，也是中国及东亚已知最早的成体系文字。甲骨文的研究不仅对了解中国文字起源与演化十分重要，而且对中国乃至全世界文字文化的溯源与传承有非比寻常的意义。甲骨文的发现不但促进了世界对于人类早期文字的研究，也促进了对甲骨文研究的“甲骨学”学科的诞生。刻有甲骨文的龟甲于 1899 年被清末官员、金石学家王懿荣首次发现，这一重大发现随即引起中外学术界轰动。百年间，曾有数十次以甲骨文相关历史材料为目标的考古发掘活动。迄今为止，作为甲骨文主要发掘地的河南省安阳市殷墟已经出土了超过 15 万片的字甲骨<sup>[1]</sup>。2017 年 12 月 26 日，甲骨文成功入选“世界记忆名录”。甲骨文的重要性已经得到中国乃至全世界的公认。



图 1 博物馆甲骨展品

尽管甲骨文自被发现之后，其重要性得到广泛承认，其衍生的甲骨学发展迅速。然而作为距今 3000 年的文字体系，甲骨文字符的含

义仍然较难破译。在出土的 15 万余片甲骨中，大约包含 4500 个不同的甲骨字符，然而仅有 2000 个左右被成功破译解读<sup>[2]</sup>，破译量不足二分之一。目前甲骨文识别主要依赖于学科专家的识别工作，往往需要专家有丰富的甲骨文有关知识与经验。这种专家识别方式的成本较高且效率较低。而对于大众来说，目前大部分普通人对甲骨文认识较为浅显，仅仅停留在它作为文字符号的层面上，很少了解文字涵义。这种现象的主要原因是大众很少受到甲骨知识熏陶。

## 关于征集评选甲骨文释读优秀成果的奖励公告

作者：

《光明日报》（2016年10月28日 06版）

经全国哲学社会科学规划办公室同意，为充分调动海内外甲骨文研究者的积极性、创造性，解决甲骨文释读目前面临的瓶颈问题，推动甲骨文释读取得新的实质性进展，中国文字博物馆自即日起组织实施甲骨文释读成果专项奖励计划，面向海内外公开征集优秀成果并予以奖励。现公告如下：

### 划重点！

### 三、奖励标准

1. 对破译未释读甲骨文并经专家委员会鉴定通过的研究成果，单字奖励10万元；对存争议甲骨文作出新的释读并经专家委员会鉴定通过的研究成果，单字奖励5万元。

图 2 中国文字博物馆面向社会面征集甲骨文破译成果的公告

近年来，人工智能技术已经成功应用于很多领域，如图像识别和自然语言处理。本项目拟与考古专家合作，基于多模态甲骨文图像和利用人工智能技术研发一个先进的跨模态甲骨文智能识别系统，加快考古专家识别未破译甲骨文效率并普及甲骨文的大众教育，具有较高的研究价值和应用意义。

## 二、 文献调研

经过调研，发现多篇使用人工智能辅助汉字识别的研究。基于这些研究使用的技术与方法，可大致分为利用机器学习进行文字识别与利用深度学习进行文字识别两类。

就机器学习进行文字识别的工作而言，刘永革等人《基于 SVM 的甲骨文字识别》<sup>[3]</sup> 颇具代表性。该研究选择使用支持向量机 SVM 完成汉字分类任务，训练集包含 1300 张图片，15 个单字字符，其识别流程如图 3。然而，这项研究的局限性也比较明显。其一，其数据集的样本量较少，且均来自甲骨摹本，是一个单模态的数据集；其二，其分类任务较简单，实验结果不具有复杂情况下的普适性；其三，该研究未使用深度学习方法，深度学习方法实现简单，识别精度高，目前在文字识别领域被广泛使用。

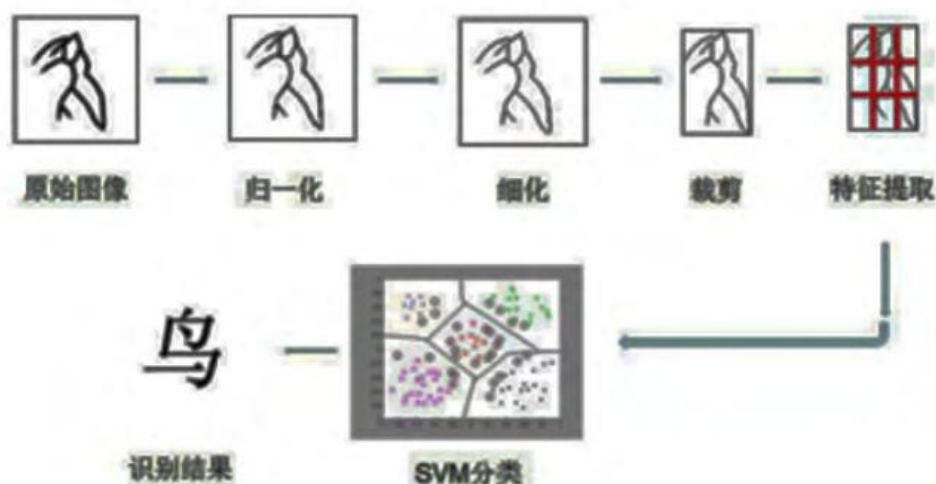


图 3 《基于 SVM 的甲骨文字识别》识别流程图

对于使用深度学习方法进行文字识别的工作，刘芳等人的《基于 Mask R-CNN 的甲骨文拓片的自动检测与识别研究》<sup>[4]</sup>较为新颖。该研究选取约 3000 张甲骨文拓片建立数据集。为了丰富数据集样本容量，该研究应用了一些简单图像变换以对原图像处理，得到新的训练与测试样本。该研究基于 Mask R-CNN 建立识别模型，并应用 Triplet Loss 等算法对识别效果进行优化，其识别流程如图 4。然而该研究也有一定局限性。其一，其同样存在数据集样本量不够大的问题；其二，数据集样本不均衡，常用字出现频率较高，而生僻字仅包含个别样本；其三，其数据集均采用甲骨拓片来构建，因此，该研究建立的是单模态的神经网络。

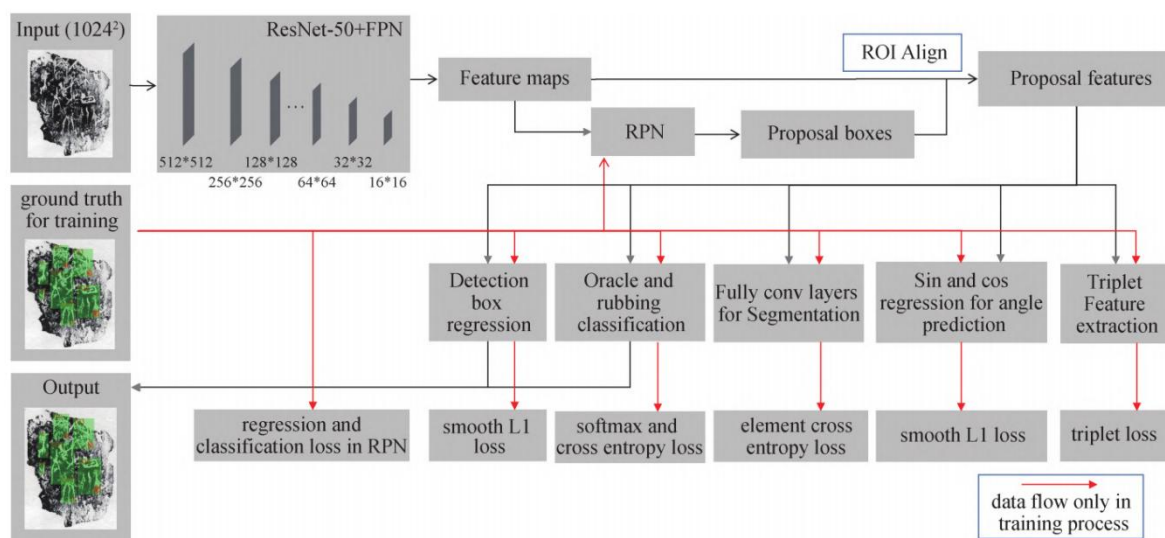


图 4 《基于 Mask R-CNN 的甲骨文拓片的自动检测与识别研究》识别流程图

## 三、 项目计划

### (一) 研究内容

本项目拟基于甲骨拓片与甲骨摹本图像, 并利用人工智能技术研发一种跨模态学习的甲骨文智能识别系统, 用于普及甲骨文教育和辅助考古专家识别尚未破译的甲骨文。

拟基于以下三个研究内容来构建基于跨模态学习的甲骨文智能识别系统:

#### 1. 跨模态甲骨文数据库构建

目前甲骨文字图像可以分为甲骨摹本图像与甲骨拓片图像两类。甲骨拓片指使用宣纸与墨汁, 将龟甲、兽骨上的纹路、图样拓印下来的一种技术。拓片作为我国一项古老的传统技艺, 是千百年来传承我国文化艺术的一种载体。

甲骨摹本指甲骨学者手工临摹甲骨纹路得到的摹本。这种保存形式较甲骨照片更为清晰, 修复了甲骨文图像的残缺与噪声。然而甲骨摹本需要较多人力物力得到摹本, 同时对参与临摹的学者专业水平要求较高。

由于缺少训练样本, 仅使用甲骨拓片图像对模型进行训练很难取得较高的识别精度<sup>[5]</sup>。同时, 由于甲骨摹本具备一些甲骨拓片文字类别没有的训练样本, 甲骨文字的增量识别也是辅助甲骨文字专家进行语言研究的重要手段。因此, 本研究构建了一个基于甲骨摹本与甲骨拓片的跨模态甲骨文数据库, 从而提高模型的识别精度。



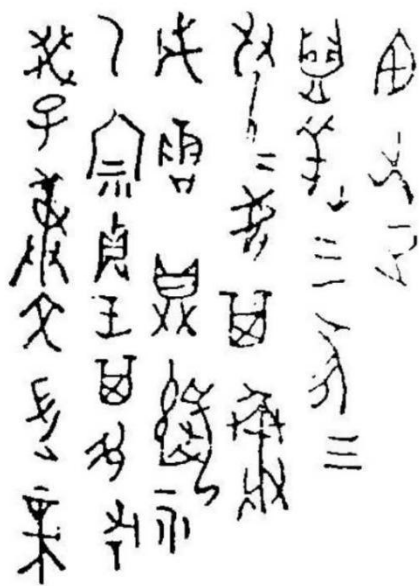


图 5 甲骨摹本（左）与甲骨拓片（右）

## 2. 智能甲骨文识别框架

根据常识，汉字的偏旁一定程度反映了汉字的含义，且人类在认识汉字时根据偏旁进行分类，如“芷”、“芬”、“芳”等字具有相同的偏旁，且都与花草相关。同样地，面对不认识的汉字，人类可根据偏旁推断含义，如“芑”。虽然该字较为生僻，但仍可根据其偏旁推知其与植物有一定相关性。人类在观察汉字时先注意到较重要的偏旁，以此辅助识别汉字，这启发了本研究在建立识别模型时使用注意力机制。

人类能够自然而有效地发现复杂场景中的显著区域，如图 6。基于这一观察，注意力机制被引入计算机视觉，目的是模仿人类视觉系统的这方面，将注意力转移到图像最重要区域而忽略无关部分的方法称为注意机制。注意力机制基于数据域可以分为通道注意、时间注意、空间注意、分支注意，空间时间注意、空间通道注意<sup>[6]</sup>，如图 7。

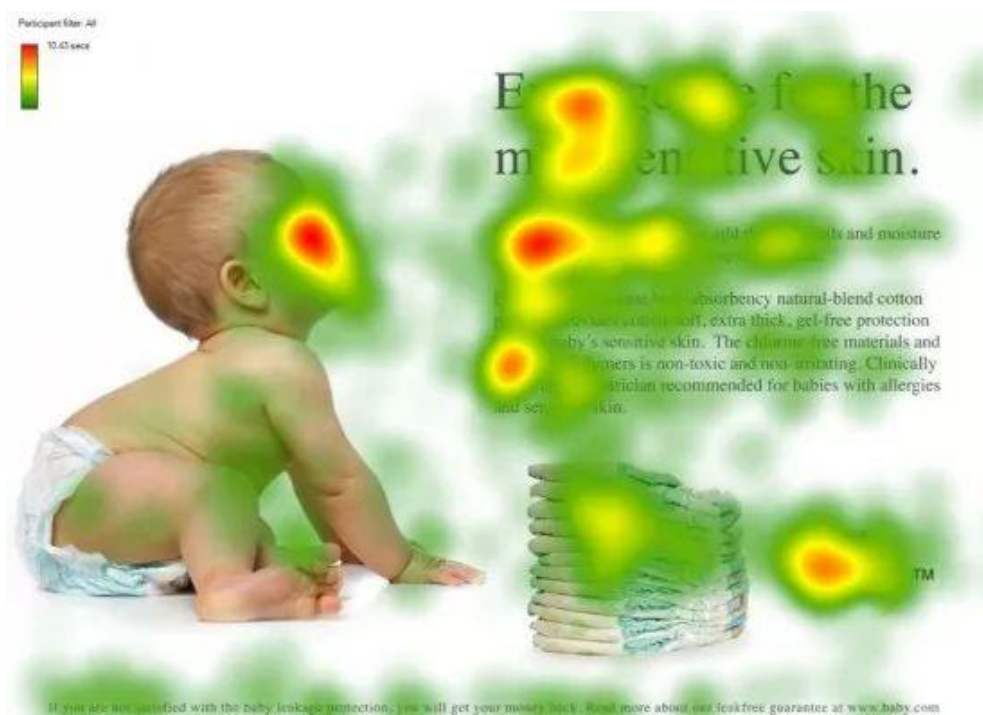


图 6 人类的视觉注意力

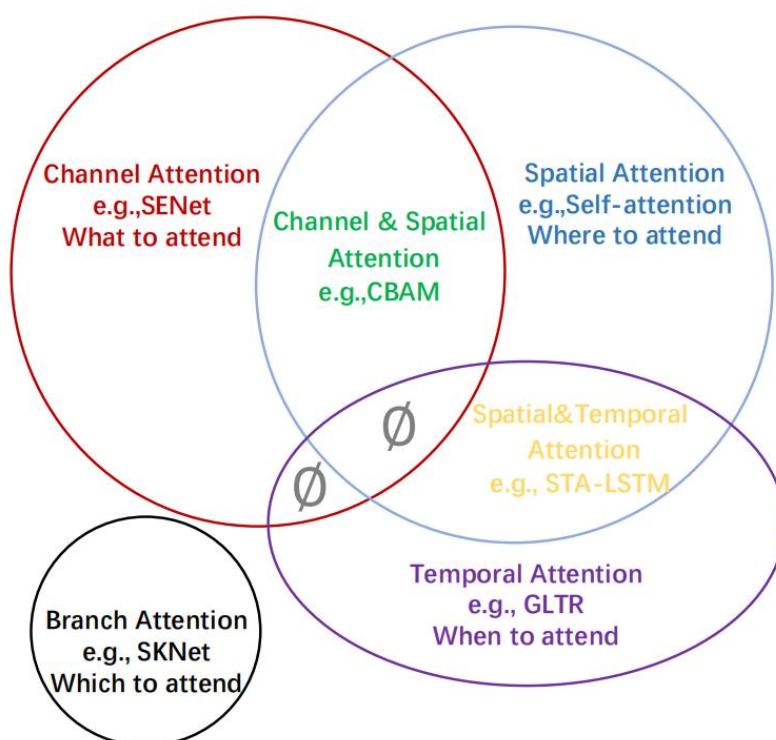


图 7 注意力机制基于数据域的分类

### 3. 智能甲骨文识别系统

图 8 是甲骨文识别系统的实现流程。识别系统主要由 4 部分组成：

- (1) 图像上传模块：用户将待识别的甲骨文图像进行上传。



- (2) 图像处理模块：将图片去噪及尺寸归一化。
- (3) 图像识别模块：利用已经训练好的分类器对甲骨文图像进行识别。
- (4) 结果输出模块：将预测结果概率分布最大的 3 张图片及对应简体字输出到前端界面。

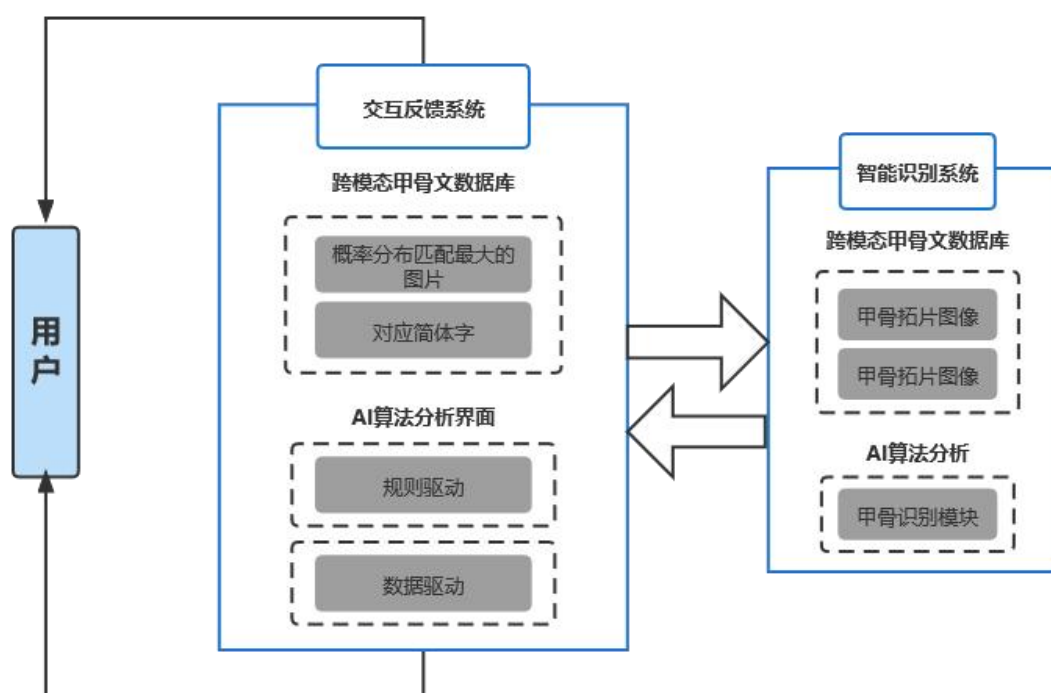


图 8 甲骨文识别系统实现流程

## (二) 项目阶段时间规划

### 1. 2.15 - 3.31 背景调研，知识学习与建立甲骨数据集

- (1) 与考古专家讨论研究方法和收集相关资料；
- (2) 学习深度学习及相关代码库知识；
- (3) 甲骨数据采集与建立数据集。

## 2. 4.1-4.30 进一步扩充数据集与训练识别模型

- (1) 进一步扩充数据集，形成足够大的训练样本；
- (2) 建立甲骨文识别模型并提高识别效果。

## 3. 5.1-final 提高模型识别效果与设计前端展示界面

- (1) 不断改进模型，提高模型的识别准确率并减少推断时间；
- (2) 设计前端展示网页并不断美化。

# 四、 项目总结

## （一）跨模态甲骨文数据库构建

### （1）甲骨拓片的爬取

由于目前几乎没有甲骨拓片的公开数据集，因此本研究需自行创建甲骨拓片数据集作后续研究使用。选取甲骨文拓片书籍《甲骨金文拓本精选释译》，将书籍中拓片图像手动爬取并进行数据标注。

### （2）甲骨拓片的标注

本研究对爬取的拓片进行了手动标注，标注拓片约 900 张。该标注任务难度较大，首先，《甲骨金文拓本精选释译》中的字符标注为手写体汉字（如图 9）。其次，甲骨拓片图像与字符标注需自行对应。

由于部分甲骨拓片存在破损严重的情况，因此在标签中标注了字符的句间位置，即该字符位于第几篇文章中的第几个字，以便后续可

能进行的使用 NLP 的方法根据上下文对字符推断的研究。

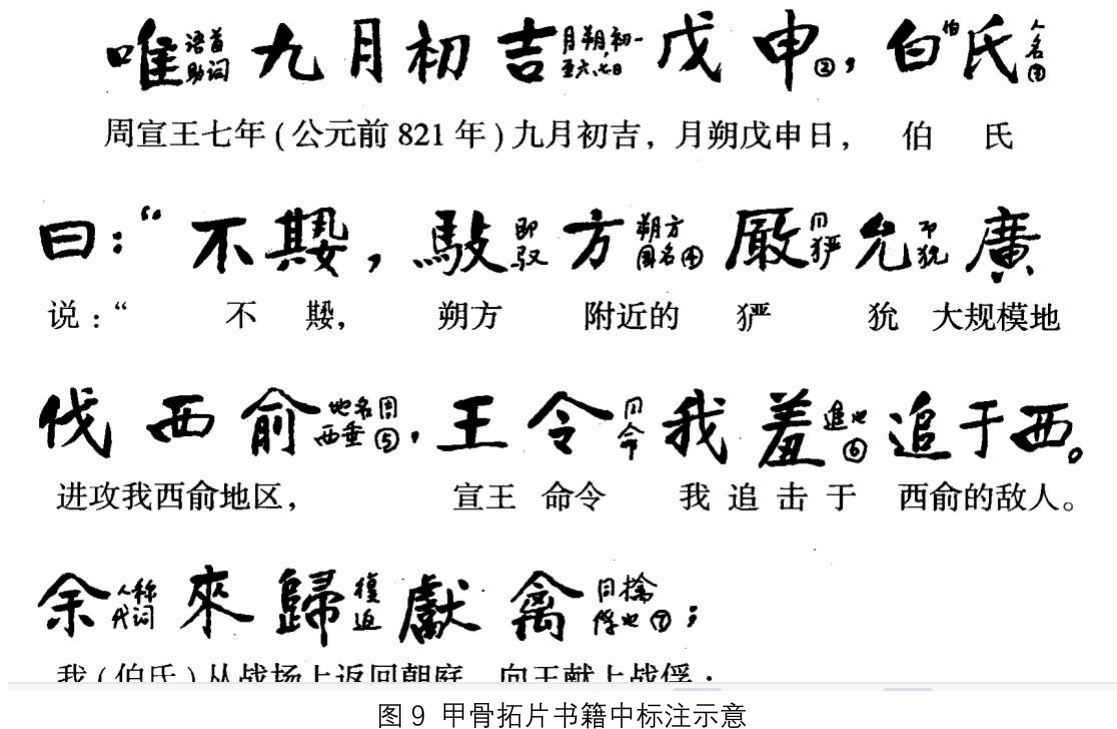


图 9 甲骨拓片书籍中标注示意

值得注意的是, 部分甲骨文字并没有简体字对应的字符, 因此采取“字海”网站 (一个收录了《康熙字典》全部字符并编号的网站) 的编号作为标注。此外, 还存在一个甲骨字符对应多个简体字的情况, 暂时使用多个简体字做标注。

### (3) 甲骨拓片预处理

甲骨拓片往往噪声大, 边缘不够清晰。为提高甲骨拓片的训练效率与识别准确率, 本研究需对甲骨拓片进行数据预处理, 以清洗甲骨拓片图像的噪音, 并对甲骨拓片的大小归一化, 并且保留其形态学信息。

相比将图像中的像素点传入神经网络进行训练, 本研究尝试先对甲骨拓片进行边缘检测、特征提取, 然后以图像的特征进行训练。因此本研究尝试了基于甲骨拓片的图像特征提取的创新探究。手写文字

图像的边缘特征多为简单的直线和曲线，因此适合用贝塞尔曲线拟合。使用贝塞尔曲线拟合可以降低图像特征的参数量，拓片图像有几百个像素点，而三阶贝塞尔曲线约具有 10 个参数。

因此本研究先采用 Canny 算法进行图像边缘检测，通过“拐点”将边缘进行分割为多条曲线，然后使用贝塞尔曲线分别进行拟合。本研究还尝试了使用多线程加速单张图片处理的过程。然而在图片处理中，出现了分辨率过低与图像拐点密集问题。为了更好地进行边缘分割，本研究还做了放大光滑实验，探究了超分辨率算法。

在进行超分辨率算法时，对于放大倍数、模糊核等可选取的超参数存在一个多参数多目标优化的问题。本研究还遍历了可选取参数的解，找到了一组极具鲁棒性的超参数，可以使得各张图片的形态学特征提取效果最好。

在设计边缘分割算法时，需要使得边缘像素点序列化。为了探索具有鲁棒性的算法，固定左上一点为起点，做修改后的广度优先探索。目前，能将闭合边缘，宽度不超过 2 像素的边缘序列化后使前后两点彼此相邻。然而，更为复杂的边缘分割，如汉字中存在分支的字形，需要更为复杂的算法，目前算法的时间复杂度为  $O(n)$ 。

为了更好地以贝塞尔曲线拟合边缘。将序列分为多段，以用更低阶的贝塞尔曲线拟合。线性拟合的  $R^2$  会出现震荡的问题，因此不能用二分搜索解决，于是增加限制，要求每个点的 xy 误差值不大于 3，然后使用二分法。

贝塞尔四阶曲线由四个点与变量  $t \in [0,1]$  控制，该多项式方程可

以通过最小二乘法的线性拟合求出系数，然后利用系数解出贝塞尔曲线的起终点与控制点。

将提取到的甲骨拓片信息送入神经网络学习，识别准确率约为 2%，误差大于 2 万，均不理想。这是因为深度学习的方法对“位置”或者“等位性”更加敏感，而对边缘特征不敏感。后改进网络，使用共享参数的方式，识别准确率达到 4%。由于该分类任务为 558 类，因此随机分类的识别准确率约为 0.2%，可见使用边缘特征让网络进行学习的方式还是有一定潜力的，未来我们将继续探索。

## （二）智能甲骨文识别框架

### （1）Vision Permutator

基于对汉字的认知，汉字存在一定空间结构特征，例如：上下结构、左右结构、半包围结构等。这些汉字的空间结构可以辅助汉字识别，以一些左右结构的字为例，仅凭借字的左半边或右半边信息即可推断其具体字形。并且，不同区域对于汉字识别的重要性可能不同。因此，我们联想到在识别汉字时可以将汉字切成“小块”处理。

这种处理汉字的方法与 Vision Permutator (ViP)<sup>[7]</sup>的核心思想相近。其处理图像识别任务时通过将图片“切片”以进行编码，这种处理图像的方式可以充分利用汉字的空间结构特征；此外，ViP 是一种类 MLP 的神经网络结构，其参数量较少，因此其训练与推断更为高效。

Permutator block 是 ViP 的核心结构，这正是 ViP 与其他类 MLP



结构，如 Mlp-mixer<sup>[8]</sup>的区别的核心之处。在处理 token 时，Mlp-mixer 在空间维度将 token 展平，然后在空间维度线性投影以编码空间信息，其输入是二维的，导致了部分空间信息的丢失。与之不同，ViP 保持了原来的空间维度，以三维向量作为输入，分别沿着宽和高两个维度编码空间信息，保证 token 具有特定方向的信息，避免了空间信息的丢失。

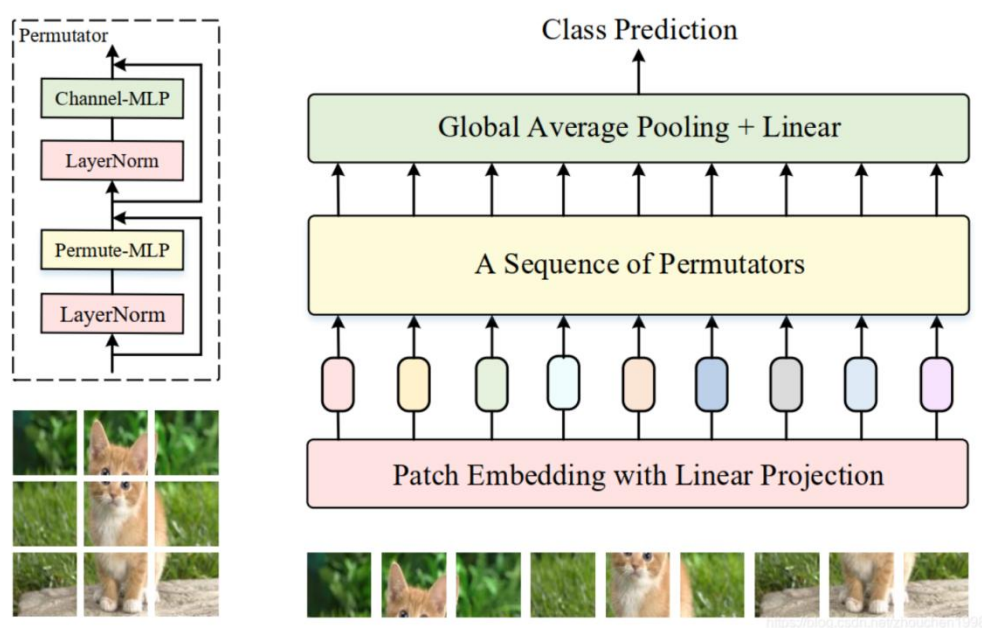


图 10 ViP 结构示意图

经过一定的数据预处理工作后，根据现有甲骨文数据集中的文字种类进行分类，统计各类字符的样本数量情况。

经过运行相关代码进行统计，该数据集包含 18000 张图像，558 种单字，统计结果如图 11。然后，基于该甲骨摹本数据集，本研究设计相关程序将数据集划分为训练集与测试集。本研究还编码了甲骨文字母表，用于划定推断范围。

size of dataset: 18689  
 {'一': 2, '元': 12, '天': 9, '上': 7, '帝': 113, '旁': 1, '下': 18, '示': 7, '祥': 1, '福': 174, '祭': 37, '祀': 35, '祖': 1, '初': 2, '祔': 4, '祝': 43, '祈': 1, '三': 10, '王': 270, '玉': 13, '中': 34, '屯': 57, '每': 70, '兹': 1, '若': 100, '咎': 42, '折': 2, '雀': 2, '蒿': 1, '莫': 36, '葬': 4, '小': 19, '少': 8, '八': 11, '分': 1, '曾': 3, '介': 6, '公': 9, '余': 78, '牛': 97, '牡': 60, '牝': 69, '牢': 134, '物': 30, '告': 147, '口': 19, '吮': 1, '吹': 3, '名': 3, '君': 2, '召': 46, '问': 1, '唯': 1, '咸': 31, '呈': 4, '吉': 101, '周': 41, '唐': 83, '吝': 1, '各': 16, '单': 4, '止': 13, '归': 64, '登': 3, '爰': 3, '步': 91, '岁': 193, '此': 2, '正': 84, '征': 1, '逆': 20, '遭': 37, '追': 26, '逐': 55, '复': 21, '住': 1, '得': 35, '行': 144, '齿': 44, '龋': 3, '品': 9, '禽': 13, '酥': 2, '册': 42, '囿': 1, '舌': 16, '干': 38, '商': 130, '古': 40, '十': 6, '千': 8, '廿': 15, '言': 21, '竞': 10, '妾': 7, '对': 3, '兵': 3, '异': 21, '兴': 17, '高': 4, '虞': 1, '为': 9, '书': 14, '斗': 40, '又': 147, '父': 70, '尹': 53, '及': 47, '秉': 4, '取': 110, '友': 10, '史': 77, '丰': 4, '臣': 48, '爰': 3, '寻': 2, '专': 20, '启': 89, '敏': 13, '攸': 4, '败': 2, '败': 3, '牧': 27, '教': 10, '卜': 14, '贞': 140, '占': 72, '用': 117, '甫': 22, '爻': 8, '目': 5, '眚': 2, '相': 6, '眉': 5, '省': 83, '自': 169, '鲁': 13, '百': 48, '鼻': 1, '习': 4, '羽': 73, '翊': 27, '隹': 215, '只': 174, '雀': 68, '雉': 12, '鸡': 10, '雁': 6, '霍': 43, '旧': 12, '蔑': 12, '羊': 98, '美': 4, '羌': 215, '龔': 1, '集': 3, '鸟': 23, '鸣': 20, '弃': 2, '聿': 72, '幼': 1, '幽': 3, '瘞': 22, '爰': 7, '受': 151, '争': 133, '死': 42, '膏': 4, '刀': 3, '利': 15, '初': 2, '刚': 13, '刃': 1, '楷': 14, '角': 9, '解': 1, '箕': 179, '典': 12, '奠': 39, '工': 29, '甘': 3, '曰': 133, '曹': 2, '乃': 8, '宁': 39, '可': 4, '兮': 11, '乎': 107, '于': 198, '旨': 28, '喜': 34, '烹': 49, '彭': 25, '鼓': 17, '豆': 3, '豐': 8, '丰': 9, '虎': 45, '皿': 9, '孟': 78, '户': 2, '尽': 9, '盥': 4, '血': 14, '丹': 3, '井': 26, '即': 36, '既': 42, '鬯': 53, '爵': 18, '食': 28, '合': 2, '今': 185, '会': 1, '仓': 1, '入': 94, '内': 16, '缶': 30, '矢': 18, '射': 90, '高': 60, '毫': 7, '央': 8, '京': 17, '厚': 1, '良': 11, '膏': 21, '来': 180, '麦': 14, '秋': 4, '舞': 1, '书': 46, '乘': 37, '木': 1, '杏': 13, '柳': 3, '杞': 2, '朱': 1, '果': 4, '枚': 1, '橘': 3, '杲': 3, '棋': 1, '杳': 2, '乐': 3, '采': 6, '析': 12, '休': 9, '东': 81, '林': 3, '楚': 4, '麓': 7, '森': 1, '才': 107, '秦': 5, '之': 265, '出': 201, '南': 108, '生': 52, '束': 8, '囿': 1, '困': 2, '困': 7, '员': 4, '贝': 12, '贮': 25, '责': 6, '买': 6, '邑': 61, '邦': 5, '乡': 1, '日': 231, '晋': 1, '昼': 10, '晨': 13, '昏': 4, '昔': 8, '旦': 6, '旋': 4, '旅': 125, '族': 21, '晶': 2, '星': 5, '月': 65, '明': 19, '夕': 156, '夙': 3, '多': 140, '母': 18, '自': 12, '栗': 2, '齐': 5, '束': 23, '鼎': 18, '克': 34, '木': 39, '年': 122, '秦': 6, '杯': 1, '黍': 79, '米': 11, '春': 1, '家': 13, '宅': 12, '室': 28, '宜': 3, '向': 27, '安': 16, '宝': 2, '辛': 3, '宜': 2, '宿': 15, '宕': 1, '宋': 3, '宗': 103, '宫': 60, '梦': 55, '疾': 4, '同': 3, '网': 5, '帝': 206, '帛': 1, '白': 65, '敝': 1, '藩': 2, '人': 130, '保': 24, '金': 6, '伊': 30, '何': 51, '井': 11, '依': 3, '传': 1, '伐': 160, '化': 25, '匕': 42, '比': 11, '北': 50, '丘': 25, '壬': 14, '監': 2, '衣': 62, '卒': 3, '裘': 1, '老': 13, '畫': 1, '尸': 4, '舟': 15, '朕': 25, '服': 1, '方': 291, '儿': 49, '允': 91, '兄': 115, '先': 91, '见': 70, '盜': 1, '无': 3, '页': 3, '首': 4, '虎': 1, '面': 4, '须': 3, '文': 21, '后': 28, '司': 1, '令': 98, '卯': 63, '辟': 1, '旬': 63, '鬼': 23, '丑': 1, '山': 12, '岩': 2, '庞': 11, '危': 32, '石': 8, '礲': 7, '长': 33, '勿': 113, '易': 13, '而': 6, '豕': 93, '豕': 20, '彘': 59, '豚': 10, '豹': 8, '兕': 49, '易': 73, '象': 17, '马': 84, '駁': 2, '鹿': 30, '鹿': 61, '麋': 46, '麋': 2, '麋': 12, '兔': 5, '犬': 97, '龙': 1, '臭': 1, '猋': 1, '犹': 24, '火': 19, '熹': 3, '焚': 11, '光': 15, '炎': 1, '焱': 3, '赤': 3, '大': 246, '夷': 6, '亦': 69, '天': 2, '壺': 5, '执': 36, '圉': 10, '奚': 8, '夫': 13, '立': 32, '汝': 5, '淮': 3, '涿': 1, '泚': 69, '祀': 1, '淵': 57, '澆': 3, '酒': 1, '涉': 31, '泉': 10, '永': 36, '谷': 2, '冬': 13, '雨': 270, '雷': 21, '霤': 8, '霖': 2, '云': 19, '鱼': 20, '渔': 37, '燕': 42, '龙': 42, '乳': 1, '不': 240, '至': 107, '西': 82, '户': 1, '门': 32, '耳': 4, '声': 1, '闻': 16, '聃': 3, '承': 1, '扞': 1, '女': 34, '姜': 2, '妇': 2, '妊': 1, '母': 91, '妹': 11, '侄': 3, '奴': 26, '好': 70, '嫫': 16, '如': 10, '嫫': 2, '爰': 12, '义': 1, '弗': 75, '戈': 9, '戎': 1, '戎': 1, '武': 17, '戈': 2, '戎': 46, '戚': 1, '我': 142, '义': 3, '直': 1, '亡': 155, '乍': 59, '区': 5, '匸': 38, '匸': 6, '曲': 2, '甌': 12, '弓': 7, '系': 29, '孙': 3, '丝': 2, '率': 22, '虫': 2, '虹': 3, '蟲': 17, '风': 1, '它': 89, '龟': 20, '鼃': 1, '鼃': 2, '二': 18, '巫': 2, '士': 110, '基': 1, '封': 4, '奎': 4, '尧': 1, '田': 173, '畜': 1, '黄': 41, '男': 3, '荔': 84, '斧': 1, '新': 32, '车': 13, '官': 3, '阳': 1, '陟': 17, '降': 48, '四': 16, '亚': 61, '五': 46, '六': 30, '七': 10, '九': 30, '万': 5, '甲': 50, '乙': 91, '丙': 32, '丁': 141, '戊': 30, '成': 44, '己': 13, '巴': 22, '庚': 65, '辛': 17, '癸': 4, '子': 63, '孕': 1, '季': 5, '孽': 1, '藿': 6, '寅': 13, '辰': 31, '巳': 24, '以': 37, '午': 9, '未': 42, '申': 11, '酉': 43, '酒': 121, '尊': 16, '戌': 53, '亥': 16}

type of different chars: 558

图 11 现有数据集样本类型统计

本研究尝试以多种网络搭建识别框架，其中 VGGNet 在测试集上的准确率为 87.731%，在测试集上准确率为 80.105%(200epoch)。ResNet 的准确率分别为 88.539%与 80.476%(200epoch)，ResNet-SE: 的准确率分别为 90.801%/81.891%(200epoch)，如图 12。

而以 ViP 搭建识别框架时遇到了一定的问题。该网络在训练集表现很好，识别准确率有 94.83%。然而在测试集上表现不佳，仅有 50%左右。经过探索，我们发现 ViP 的训练依赖于预训练模型。

于是我们从 ViP 处理图片的方式中汲取了思路，依旧采用将汉字分块处理的方式，并使用注意力机制中的协调注意力（Coordination Attention）<sup>[9]</sup>。

```
1 10.20.82.129 2 10.20.82.129
Epoch: 192
[===== 263/263 =====>] Step: 107ms | Tot: 31s254ms | Loss: 0.696 | Acc: 90.658% (15256/1682)
[===== 30/30 =====>..] Step: 26ms | Tot: 1s794ms | Loss: 1.044 | Acc: 81.300% (1513/1861)
oracle

Epoch: 193
[===== 263/263 =====>] Step: 108ms | Tot: 31s314ms | Loss: 0.697 | Acc: 90.652% (15255/1682)
[===== 30/30 =====>..] Step: 22ms | Tot: 1s804ms | Loss: 0.979 | Acc: 81.354% (1514/1861)
oracle

Epoch: 194
[===== 263/263 =====>] Step: 109ms | Tot: 31s98ms | Loss: 0.687 | Acc: 90.605% (15247/16828)
[===== 30/30 =====>..] Step: 31ms | Tot: 1s887ms | Loss: 0.986 | Acc: 81.730% (1521/1861)
oracle

Epoch: 195
[===== 263/263 =====>] Step: 112ms | Tot: 30s888ms | Loss: 0.695 | Acc: 90.563% (15240/1682)
[===== 30/30 =====>..] Step: 29ms | Tot: 1s848ms | Loss: 1.022 | Acc: 81.730% (1521/1861)
oracle

Epoch: 196
[===== 263/263 =====>] Step: 108ms | Tot: 31s146ms | Loss: 0.682 | Acc: 90.641% (15253/1682)
[===== 30/30 =====>..] Step: 29ms | Tot: 1s842ms | Loss: 0.990 | Acc: 81.838% (1523/1861)
Saving..
oracle

Epoch: 197
[===== 263/263 =====>] Step: 109ms | Tot: 31s209ms | Loss: 0.690 | Acc: 90.801% (15280/1682)
[===== 30/30 =====>..] Step: 29ms | Tot: 1s832ms | Loss: 0.980 | Acc: 81.891% (1524/1861)
Saving..
oracle

Epoch: 198
[===== 263/263 =====>] Step: 109ms | Tot: 31s279ms | Loss: 0.694 | Acc: 90.795% (15279/1682)
[===== 30/30 =====>..] Step: 36ms | Tot: 1s856ms | Loss: 1.009 | Acc: 81.193% (1511/1861)
oracle

Epoch: 199
[===== 263/263 =====>] Step: 108ms | Tot: 31s114ms | Loss: 0.693 | Acc: 90.575% (15242/1682)
[===== 30/30 =====>..] Step: 32ms | Tot: 1s750ms | Loss: 0.986 | Acc: 81.677% (1520/1861)
(mtan) cseadmin@liuj001:/data/sun_qing_yang/pytorch-cifar$
```

图 12 ResNet-SE 的训练过程示意

于是我们从 ViP 处理图片的方式中汲取了思路，依旧采用将汉字分块处理的方式，并使用注意力机制中的协调注意力（Coordination Attention）<sup>[9]</sup>。

以 SE（Squeeze and excitation）、CBAM（Convolutional Block Attention Module）与 CA（Coordination Attention）做对比，如图 13。SE 只考虑了通道上的注意力，空间上的信息被 GAP 直接池化为  $1 * 1$  的值，损失了空间信息。而 CBAM 虽然同时考虑了通道注意力与空间注意力，然而其感受野为  $7 * 7$ ，不足够感受全局信息。而 CA 通过双分支编码空间信息，融合了水平和垂直方向的大感受野，空间信息得以保留，同时融合了空间的注意力和通道的注意力，因此可以更有效编码空间信息。

我们以将汉字切块处理的思路，并基于 CA 改进了甲骨文识别框架，如图 14。CA 中编码空间信息使用两个分支，分别沿着长与宽，

将空间信息压缩为“一条线”，而改进后的甲骨文识别框架分别沿着图片上半部分与下半部分编码空间信息。这样能更好地利用汉字特有的空间信息。

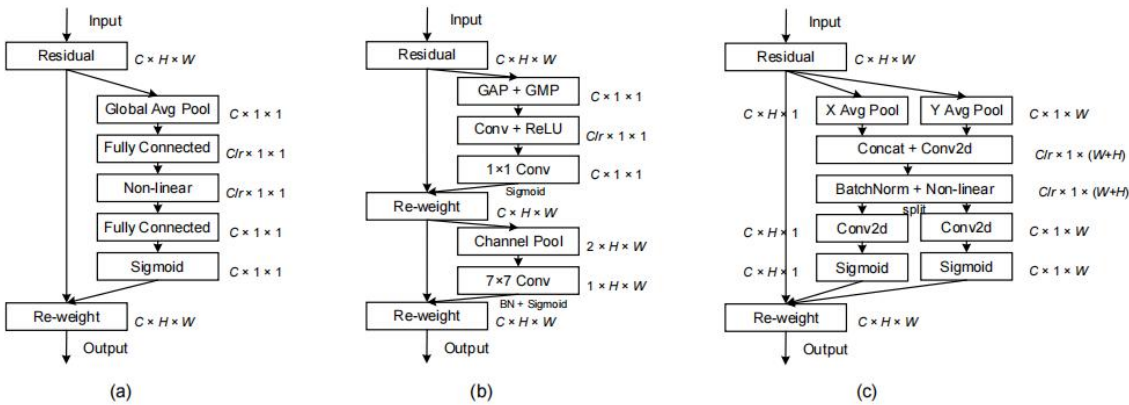


图 13 SE、CBAM 与 CA 的对比

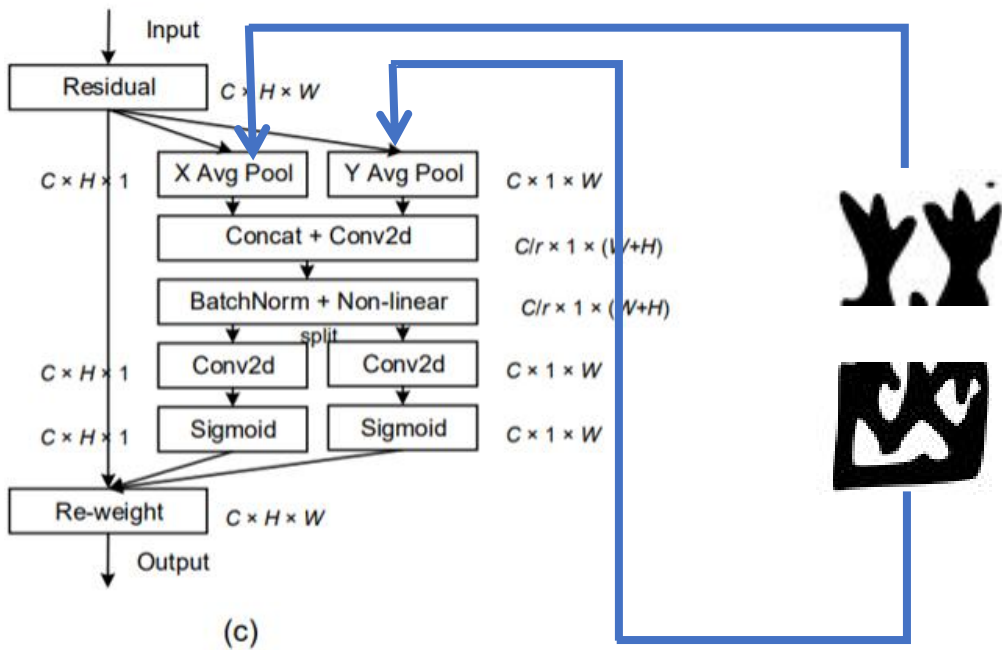


图 14 基于 CA 的甲骨文识别框架

这种改进后的甲骨文识别框架在训练集上达到 94.967%的识别准确率，在测试集上准确度达到 84.358%（图 15）。可以看出，这个结果是优于 RESNet-SE 的。

```

Epoch: 138
[===== 116/116 =====>] Step: 235ms | Tot: 30s577ms | Loss: 0.516 | Acc: 94.961% (14081/1483)
[===== 29/29 =====>..] Step: 136ms | Tot: 3s988ms | Loss: 1.073 | Acc: 83.415% (3092/3707)
oracle

Epoch: 139
[===== 116/116 =====>] Step: 237ms | Tot: 30s559ms | Loss: 0.522 | Acc: 94.927% (14088/1483)
[===== 29/29 =====>..] Step: 129ms | Tot: 3s997ms | Loss: 1.014 | Acc: 84.358% (3127/3707)
Saving..
oracle

Epoch: 140
[===== 116/116 =====>] Step: 239ms | Tot: 30s566ms | Loss: 0.516 | Acc: 94.947% (14088/1483)
[===== 29/29 =====>..] Step: 108ms | Tot: 3s857ms | Loss: 1.243 | Acc: 80.853% (2997/3707)
oracle

Epoch: 141
[===== 116/116 =====>] Step: 234ms | Tot: 30s461ms | Loss: 0.516 | Acc: 94.967% (14091/1483)
[===== 29/29 =====>..] Step: 134ms | Tot: 3s947ms | Loss: 1.027 | Acc: 83.764% (3105/3707)
oracle

```

图 15 基于 CA 的甲骨文识别框架的识别效果

### (三) 智能甲骨文识别系统

#### (1) 甲骨文化展示界面

为科普甲骨文化，本研究制作了甲骨文科普网页。该网页包含甲骨文科普短视频、甲骨文科普文字、甲骨文科普音频等多种展示形式。

此外，该网页还科普了甲骨文简单字形，旨在提升公众对甲骨文化的兴趣，如图 16。



图 16 甲骨文字形科普

#### (2) 甲骨文识别窗口

该识别窗口基于 nodejs 实现。首先在后台运行 node 服务器，服务器将会接收前端传入的图片，并上传到前端展示。后端的甲骨文识



别程序同时会读入图片，并且进行识别，并将识别的结果反馈到前端，前端将在指定位置展示识别结果与简体字图片，便于公众对比字形变化，演示如图 17。

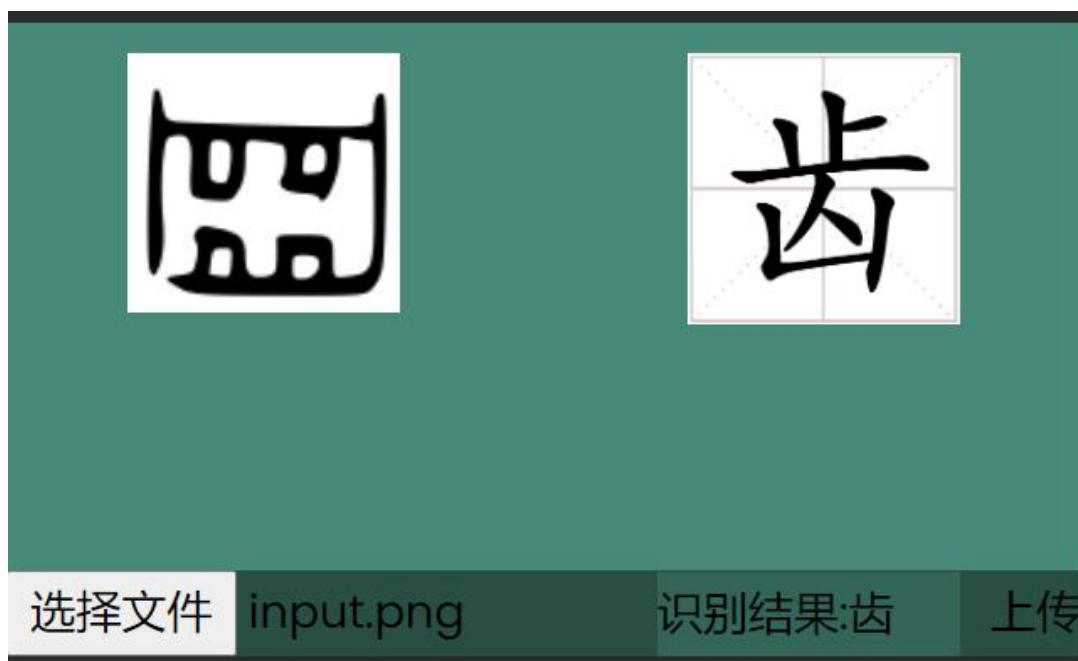


图 17 甲骨文识别窗口演示

## 五、 下一阶段目标展望

1. 尝试使用拓片特征信息；
2. 继续改进识别模型；
3. 继续美化识别平台

## 参考文献

- [1]胡厚宣.八十五年来甲骨文材料之再统计[J].史学月刊,1984(05):17-24.
- [2]Cheung C. The Chinese History That Is Written in Bone: The bones of 3,000-year-old sacrificial victims in China are revealing unexpected new twists[J]. American Scientist, 2018, 106(3): 133-135.
- [3]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(02):54-56.DOI:10.16140/j.cnki.1671-5330.2017.02.012.
- [4]刘芳,李华飙,马晋,闫升,金沛然.基于 Mask R-CNN 的甲骨文拓片的自动检测与识别研究[J].数据分析与知识发现,2021,5(12):88-97.
- [5]Huang S, Wang H, Liu Y, et al. OBC306: A large-scale Oracle Bone character recognition dataset[C]//2019 International Conference on Document Analysis and Recognition (ICDAR). IEEE, 2019: 681-688.
- [6]Guo M H, Xu T X, Liu J J, et al. Attention mechanisms in computer vision: A survey[J]. Computational Visual Media, 2022: 1-38.
- [7]Hou Q, Jiang Z, Yuan L, et al. Vision permutator: A permutable mlp-like architecture for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022.
- [8]Tolstikhin I O, Houlsby N, Kolesnikov A, et al. Mlp-mixer: An all-mlp architecture for vision[J]. Advances in Neural Information Processing Systems, 2021, 34.