

《人工智能实验》期末作业说明

一、任务简介

出土竹简是了解我国古代历史、思想、制度、文化等的重要资料。但由于多种因素影响，竹简在埋藏过程中常常散乱断裂，给解读工作带来了较大的挑战。缀合就是将这些竹简残片依据形状、文字内容、断口痕迹等信息重新拼接起来。只有成功缀合，才能恢复竹简原本的顺序和内容的完整性，从而获取准确的历史信息。然而，缀合工作常常需要人类专家人工完成，工作量较大。因此，使用人工智能技术实现竹简残片的缀合技术研究正日益成为研究的焦点。

本次任务从《里耶秦简（二）》中摘取已缀合和未缀合的竹简残片图片作为数据集，要求使用包括但不限于本课程所学的知识，实现缀合技术。

二、任务描述

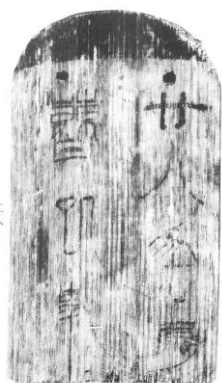
给定单根简的图片和已知的可缀合组，从图片中找到可以两两拼缀的竹简组合。具体要求：

（一）判断竹简是否为完整简，挑选出其中的残断简

完整简如：



下端削尖



上端弧形



四边平整

残断简如：



上端残断



上下残断



上下左三边残



四边均残

（二）提取残断简的各种信息

残简拼缀工作一般需要获取以下信息：

1. 测量简的宽度；
2. 判断断口的位置（上下左右皆有可能）；
3. 提取断口的形状；
4. 提取竹简的纹路；
5. 提取断口附近残字的墨迹（不一定有）；

（三）判断是否可以拼缀

两两比对残简的断口形状、纹路、与墨迹的衔接情况。对于可拼缀的组合而言，断口应比较吻合、竹简纹路比较相似。如果断口处有残字，拼合后墨迹连贯，可复原为完整的字。且上下拼缀的简一般宽度也较为一致。如：



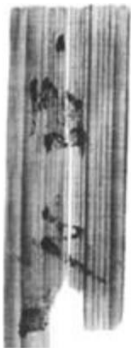
二、9-172+9-1267+9-1404



六、9-605+9-861



一七、9-2679++
9-2949



9-3094+9-3093

三、任务文件说明

（一）文件说明

- weizhuihe.zip: 无需缀合的完整简和学界未缀合的残简图片
- yizhuihe.zip: 学界已缀合的残简图片，每个子文件夹内的残简图片为一个缀合组
- zhuihetongji.xlsx: 里耶秦简二缀合统计表，学界已知的拼缀组合信息

（二）图片命名格式

一般来说，对于出土的每一支简（无论是完整的简还是残简），都会为其赋予一个唯一的编号，即简号。简的图片有红外和彩色两种，所有简都有红外图

片，少数还有彩色图片。有的筒只有一面的图片，有的两面都有字，因此有正背两面的图片，少数侧面也有字，故有侧面图片。有的筒长度较长，故在文件中分成两段，两段图片有重叠部分。缀合组 76（简 526+527）还有缀合后的图片。命名规则如下：

- （1）一支筒只有一张图片，文件名为简号，简号前大部分有加前导零。
- （2）一支筒具有彩版和红外版，彩版命名为“简号_c”（color），不加后缀为红外版。
- （3）一支筒具有正背面图像，正面命名为“简号_a”，背面命名为“简号_b”，只有一面的不加后缀。一支筒如果具有正侧面图像，正面命名为“简号_a”，侧面命名为“简号_s”（side），只有一面图像的理论上加不加后缀，但实际上有的文件名也加了后缀_a。
- （4）一支筒在文档中分上下两段，上段命名为“简号_1”，下段命名为“简号_2”，只有一段的不加后缀。
- （5）原书中筒的图片默认原大（即图片大小和筒的实际大小相同），但有的缩小了比例，则命名为“简号_缩小至比例”。原图不加后缀。因此，建议利用好图片大小的信息减少多余的缀合尝试。
- （6）具备以上多种情况的，按照“简号_缩小到比例-正背侧-上下-彩色/红外”的顺序描述，如 1 号筒具有彩色正面上段照，并缩小了 30%，则命名为“0001_70%-a-1-c”。

（三）里耶秦简二缀合统计表格式

统计表包含多个分表，主要使用“上下”和“左右”拼两个表，两表包含了所有已知可两两拼合的竹简对。

57	55 9-249
⏪ ⏩ ⏴ ⏵	总表 上下拼 左右拼 上下+左右拼 遥缀 特殊 缀合出处统计 +

表格中 A 列为缀合组号，与文件夹的名称对应。B 至 F 列为简号，按照拼合后从上到下或从左到右的顺序排列。一般来说竹筒的正面和竹筒的正面相拼，但少数整理者的正背标注有误，因此是正面与背面相拼。

	A	B	C	D	E	F	G	H	I
1	缀合组	简号	简号	简号	简号	简号	缀合情况	备注	
2	文件夹名	从左到右是缀合后简由上至下的顺序。					正常拼合的标“0”，正面与背面缀合的标“1”，断口非完全吻合的拼合标“2”，如果有其他情况在备注说明。	文章原文有说明拼合情况为不完全吻合或依据实物的，在此说明。	
3		5 9-1897	9-1848					0	

四、任务文件获取

在校园网内，访问 FTP 服务下载：

- 节点：172.18.166.91
- 用户：ftp_user
- 密码：ftp_user

推荐使用 FileZilla Client 访问并下载。因文件较大，请注意预留一定大小的流量和磁盘空间

五、任务要求

（一）实验环节

- 综合运用数据预处理、数字图像处理、深度学习等技术，在已缀合数据集内（至少是上下缀合组的数据）设计模型实现自动缀合，并采用某种评估指标对结果进行评估。
- （可选）使用在已缀合数据集上训练好的模型，在未缀合数据集内尝试进行缀合。可使用完整的未缀合数据集或选取其子集。

（二）实验报告内容

因为本次实验为开放性任务，对实现方式、模型架构、数据处理方式等均无统一要求，最终效果明显即可。**要求着重体现在完成实验任务过程中的探索、优化和创新过程，并将此作为评分的重要依据。**

实验报告要求包含以下内容，除了文字内容外，还需要提供关键代码辅以说明：

- 数据处理：描述筛选数据、图像处理、数据集划分等涉及到数据处理的实验操作。
- 评估方式：描述自己拟采用的模型评估方式。
- 模型框架以及设计思路：描述自己设计模型的思路，并且对模型框架的细节进行详细阐述。
- 模型优化和创新过程：详细描述自己的探索、优化和创新过程。需要对每一步的探索过程带来的性能提升或者降低对进行比较。
- 实验结果：对自己的实验结果进行评估。

六、完成形式及提交

- 完成形式：单人独立完成
- 提交网站：超算习堂
- 截止日期：2025 年 7 月 6 日

七、注意事项

1. 请同学们自行独立完成，允许互相交流学习，但禁止照搬或抄袭
2. 本任务所涉及各项材料仅限本课程期末考核所用，禁止外传

八、参考资料

1. 孪生神经网络: <https://zhuanlan.zhihu.com/p/35040994>
2. <https://arxiv.org/abs/2312.08704>
3. <https://www.nature.com/articles/s40494-025-01651-9>