

单位代码	10475
学号	104754180989
分类号	TP391

# 河南大學

# 硕士学位论文

(专业学位)

基于深度度量学习的战国简文字识别技术

专业学位领域：计算机技术  
专业学位类别：工程硕士  
申请人：田园  
指导教师：凡高娟 副教授

二〇二〇年八月

The Script Recognition Technology of the  
Bamboo Slips in the Warring States Period  
Based on Deep Metric Learning

A Dissertation Submitted to  
the Graduate School of Henan University  
in Partial Fulfillment of the Requirements  
for the Degree of  
Master of Engineering Science

By

*Tian Yuan*

Supervisor: Asso.Prof.Fan Gaojuan

August, 2020

## 关于学位论文独创声明和学术诚信承诺

本人向河南大学提出硕士学位申请。本人郑重声明：所呈交的学位论文是本人在导师的指导下独立完成的，对所研究的课题有新的见解。据我所知，除文中特别加以说明、标注和致谢的地方外，论文中不包括其他人已经发表或撰写过的研究成果，也不包括其他人为获得任何教育、科研机构的学位或证书而使用过的材料。与我一同工作的同事对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

在此本人郑重承诺：所呈交的学位论文不存在舞弊作伪行为，文责自负。

学位申请人（学位论文作者）签名： 田园

2020年8月30日

## 关于学位论文著作权使用授权书

本人经河南大学审核批准授予硕士学位。作为学位论文的作者，本人完全了解并同意河南大学有关保留、使用学位论文的要求，即河南大学有权向国家图书馆、科研信息机构、数据收集机构和本校图书馆等提供学位论文（纸质文本和电子文本）以供公众检索、查阅。本人授权河南大学出于宣扬、展览学校学术发展和进行学术交流等目的，可以采取影印、缩印、扫描和拷贝等复制手段保存、汇编学位论文（纸质文本和电子文本）。

（涉及保密内容的学位论文在解密后适用本授权书）

学位获得者（学位论文作者）签名： 田园

2020年8月30日

学位论文指导教师签名： 孔高娟

2020年8月30日

## 摘要

中国文字有着悠久的历史,而对语言文字学和历史文献学的研究也有了长期深入的研究,也经历了相当长的时代。随着计算机人工智能技术的高速发展,文字识别技术已融入我们生活中的各个领域,将信息处理技术用于古文字研究也越来越多的引起了广大学者的关注。目前,对于甲骨文智能识别的研究已取得了初步的进展,但对于战国简文字研究与计算机文字识别的交叉研究还较为少见。因此,本文将基于深度度量学习的小样本学习算法用于对战国简文字的识别。主要工作如下:

(1) 目前未发现有适合战国简文字识别所用的数据集,本文提出并整理了战国简数据集 **Bambooslips**。对于安徽大学初步整理的战国简文档,将每条竹简在保持高分辨率的前提下单独取出,并与汉字文档对应保存。对于目前未有字体通过“造字”实现。运用标注软件将每条竹简中的单个文字裁剪保存,并将同一字体的图片保存在一个文件夹中,构成 **Bambooslips** 数据集。用于后续战国简文字的识别。

(2) 考虑到战国简文字数据集 **Bambooslips** 属于小样本范畴,本文提出了基于孪生网络的小样本战国简文字识别算法。将算法对 **Bambooslips** 数据集整体识别准确率与单个类别的识别准确率依次输出。并在原始小数据中引入数据增强变换,增大数据集的容量。在此基础上对原数据与增强数据分别进行训练识别,对比识别准确率。在 **Bambooslips** 数据集上,基于孪生网络的小样本战国简文字识别算法的识别效果优于基于其他深度度量学习算法的战国简文字识别。

(3) 为了让算法有实际的应用,本文基于 **pycharm+PyQt5+pythonPyQt** 设计了战国简文字识别软件平台。包括图片的导入、对图片的放大、缩小、文字标注、裁剪保存单个字体、识别单个字体并输出竹简对应词义等功能。该软件可以将待识别的竹简图片输入,获取单个文字图片,进行算法识别后,再将对应的汉字依次输出。

**关键词:** 战国简数据集, 小样本学习, 孪生网络, 深度度量学习

## ABSTRACT

Chinese character has a long history, and there has also been a long-term and in-depth studies on linguistics and historical documentation. With the rapid development of computer artificial intelligence technology, character recognition technology has been integrated into every field of our lives, and the use of information processing technology for ancient character research has also attracted more and more scholar's attention. At present, preliminary research has been made on the intelligent recognition of Oracle inscriptions, but the cross-over study between the study of characters in the Warring States Period and the computer character recognition is still relatively rare. Therefore, in this academic paper, the few shot learning algorithm based on the siamese network is used to identify the character of the Warring States Period. Main tasks as follows:

(1) Based on the fact that there is no suitable data set for the identification of the character in the Warring States period, this paper proposes and collates the Bambooslips data set. For the documents of the Warring States Period, which were initially sorted out by Anhui University, each bamboo slip was taken out separately under the premise of maintaining high resolution, and stored in correspondence with the Chinese character file. For the currently unavailable character, it is realized by "making characters". Use the labeling software to cut and save the individual character in each bamboo slip, and save the pictures of the same character in a folder to form the Bambooslips data set. It is used for the recognition of the abbreviated character of the Warring States Period.

(2) Considering that the Bambooslips data set belongs to the category of small samples, this paper proposes a few-shot character recognition algorithm based on the siamese network for

bamboo slips. The overall recognition accuracy of the algorithm on the Bambooslips data set and the recognition accuracy of individual character are output in sequence. And introduce data enhancement transformation into the original small data to increase the capacity of the data set. On this basis, the original data and the enhanced data are trained and recognized separately, and the recognition accuracy is compared. On the Bambooslips data set, the recognition effect of the few-shot recognition algorithm based on the siamese network of the bamboo slips is superior to that of the other deep metric learning algorithms.

(3) In order to make the algorithm have practical application, this paper designs a character recognition software platform based on PyQt. It includes the functions of importing pictures, zooming in and out of pictures, labeling text, cutting and saving a single character, identifying a single character and outputting the corresponding meanings of bamboo slips. The software can input bamboo pictures to be recognized, obtain a single character picture, then output the corresponding Chinese characters in sequence by algorithm recognition.

**KEY WORDS:** Bambooslips dataset, Few-shot learning, siamese network, deep metric learning

# 目录

摘要.....	I
ABSTRACT.....	III
目录.....	V
第1章 绪论.....	1
1.1 研究背景及意义.....	1
1.2 国内外研究现状.....	3
1.3 本文研究的内容.....	4
1.4 文本组织结构.....	5
第2章 相关工作.....	7
2.1 卷积神经网络.....	7
2.1.1 卷积层.....	7
2.1.2 池化层.....	8
2.1.3 全连接层.....	9
2.1.4 网络训练.....	9
2.2 孪生网络.....	10
2.2.1 基本思想.....	10
2.2.2 孪生网络结构分类.....	11
2.2.3 孪生网络单样本实际应用.....	12
2.3 小样本学习方法.....	13
2.3.1 数据增强.....	13
2.3.2 迁移学习.....	14
2.3.3 元学习.....	14
2.3.4 小结.....	15
2.4 其它深度度量学习方法.....	16
2.4.1 SoftTriple.....	16
2.4.2 CosFace.....	17
2.4.3 最新度量算法.....	17
2.5 本章小结.....	18
第3章 战国简数据集.....	21
3.1 数据集提出的背景.....	21
3.1.1 战国简文字研究现状.....	21
3.1.2 战国简文字识别面临的困难.....	24
3.2 数据集描述.....	25
3.2.1 数据集标注方式及工具.....	25
3.2.2 文本识别数据集.....	27
3.3 本章小结.....	30

第 4 章 战国简文字识别算法.....	31
4.1 基于孪生网络的小样本战国简文字识别算法.....	31
4.1.1 SiameseWord 文字识别算法 .....	31
4.1.2 基于 Resnet50 网络的 SiameseWord 算法.....	34
4.2 基于其它深度度量学习算法的战国简文字识别.....	35
4.2.1 基于 SoftTriple 的战国简文字识别算法 .....	36
4.2.2 基于 CosFace 的战国简文字识别算法.....	37
4.3 实验设计.....	37
4.3.1 软硬件配置.....	37
4.3.2 评价指标.....	38
4.3.3 数据划分.....	39
4.3.4 数据准备.....	40
4.3.5 实验参数.....	42
4.4 实验结果分析.....	42
4.4.1 小样本学习方法对比分析.....	42
4.4.2 单类文字识别效果.....	44
4.4.3 度量学习方法对比分析.....	45
4.5 本章小结.....	47
第 5 章 战国简文字识别软件平台.....	49
5.1 开发环境.....	49
5.2 平台总体架构.....	49
5.3 软件平台模块的实现.....	50
5.4 平台的应用.....	52
5.5 本章小结.....	53
第 6 章 总结与展望.....	55
6.1 总结.....	55
6.2 展望.....	55
参考文献.....	57
致谢.....	61

## 第1章 绪论

在如今人工智能盛行的时代,科技的发展进一步促进了数据挖掘技术深度的研究。单一学科的研究已满足不了现如今人们的需求,因此交叉学科的出现填补了单学科研究的局限性。随着人类文明的不断进步,越来越多的人开始研究探讨中国古文字。文字的演变代表着一个民族发展的历程,也是语言发展的真实写照。在文字的演变过程中,战国简文字在春秋金文与秦汉篆隶之间起到了重要的枢纽作用,具有历史性意义。已有学者从理论上对战国文字进行了探讨与研究,了解掌握了战国文字的演变规律与形体结构。然而战国简文字的研究一般依赖于研究古文字的专业人士,每个文字的认识都需要要经过大量的检索,检索的过程严重增加了人工识别的工作量,且对非专业人员的学习造成困难。随着深度学习的发展,战国文字的研究越来越智能化。

### 1.1 研究背景及意义

外界信息的获取有众多方式,从文字图像中获取信息是目前最为广泛的一种途径。从20世纪50年代至今,随着诸如战国简与简帛资料等地下材料的不断出土发现,战国古文字的研究已愈演愈烈。不少学者对古文字形体结构与演变规律等理论上的探讨与研究已有成效,将信息处理技术用于古文字研究也越来越多的引起广大学者的关注。目前,对于甲骨文智能识别的研究已取得了初步的进展,但对于战国简文字研究与计算机文字识别的交叉研究还较为少见。战国简文字的认识面临着一些独特的问题:一、文字是人工刻在竹简上的,故同一字的不同字体可能形体不一,存在差异;二、随着字体的演变,许多出土文献整理的生僻字,在现在的词典中是不存在的;三、竹简底板颜色的不一致及保存过程中的损坏,导致部分文字的字体残缺;四、最重要的一点,是竹简文字的数据量问题。数据太少会造成大多深度学习方法的不适用。综上所述,战国简文字的认识面临着同类样本形态不一、部分文字不存在、数据量太少、单个样本数据量不均衡等问题。

与现有数据集的识别相比，战国简文字识别难度有所增加。

对于文字识别，首先想到的是深度学习。监督分类方法得益于深度学习技术的快速发展，但监督学习一般需要大量的训练样本做数据支撑，虽然现已取得了不错的研究成果，但也有一定的局限性。在现实场景中，许多的任务并没有充足的样本量供研究，很多任务可能需要对从未见过或数据量很少的样本进行识别，但因没有足够的训练样本，难以支持常规的分类与识别任务。在这种情况下，传统的深度学习训练效果可能就会不尽人意。因此，元学习、迁移学习等研究方法逐渐被提出用于小样本学习，不仅成效不错，且为后续人员的研究提供了方法。

小样本学习 (Few-Shot Learning, FSL)<sup>[1-3]</sup>主要是学习人类可以快速的将所学知识概括应用到新的任务场景的能力。例如：一个学过加法的孩子可以快速地转移他的知识，通过几个例子来掌握乘法，如  $2 \times 3 = 2 + 2 + 2$  和  $1 \times 3 = 1 + 1 + 1$ ；或者，给定一张陌生的照片，孩子可以轻松地从大量图片中识别出来。简而言之，小样本学习可以利用先验知识，在新任务中经由推广和类比方法，模拟实现人类从有限数据中学习及获取知识的能力。根据训练样本量的不同，小样本学习可以分为如下几例：1)当训练样本有且只有一个的时候，称这类问题为单样本学习；2)当训练样本存在量偏少时，称这类问题为小样本学习；3)当识别从未见过的样本类别，也就是待学习目标不存在标签样本时，我们称之为零样本学习。也有研究者将上述三种情况统称为小样本学习，只不过将前两种当作第三种的特殊情况。BPL(贝叶斯规划学习)框架<sup>[4]</sup>便是用来进行类人概念学习。它要求计算机对于给定的少量手写字符分解成落笔-提笔的每一个笔画，然后通过对这些笔画进行解析，利用解析出的小字符组合生成新字符，这种过程是一种像人类学习一样的方式<sup>[5]</sup>。当然，FSL 也推动了机器人<sup>[6]</sup>的发展，其目标是开发能复制人类动作的机器，从而在某些场景中代替人工劳动力。例如，单眼模拟<sup>[7]</sup>、多臂老虎机<sup>[7]</sup>、视觉导航<sup>[7]</sup>等。FSL 在图像方面的研究应用已颇有成效，包括：图像分类<sup>[8]</sup>、图像检索<sup>[9]</sup>、手势识别<sup>[10]</sup>、MRI 脑图像分析<sup>[11]</sup>、视频活动检测<sup>[12]</sup>等。在人脸识别方面，FSL 可以利用人脸训练的样本进行属性学习、特征提取、模型参数的更迭等，然后通过获取的知识与已知的样本，借助网络模型来进行小样本的人脸识别。FSL 在处理这些问题方面潜力无限。

目前,小样本学习相关的研究大多集中在图像分类<sup>[13-16]</sup>方面,在人脸识别方面也有涉猎,但在文字识别方面的研究相对较少。已有的小样本文字识别多是基于 omniglot 数据集与 MNIST 数据集的更改而实现的,汉字古文字体的小样本数据集目前未发现有公开存在的。因 omniglot 数据集与 MNIST 数据集是关于英文字母与阿拉伯数字的数据集,形态单一,笔画简单,故算法识别容易,准确率较高。但古文字体变化多端,笔画复杂,更因为人工刻写,每个字体或多或少均存在差异,对文字的识别增加了难度。基于以上原因,本文所研究的内容是构建一个战国简小样本数据集,使用基于深度度量学习的小样本学习算法对本文数据集进行识别。

## 1.2 国内外研究现状

在文字检测与识别任务中,对于常见的分类、识别等任务一般都需海量样本的数据集来算法,而小样本学习的目的在于从少量数据中获取足够的信息完成分类、识别任务。早期的 FSL 算法主要侧重于小样本图像识别的任务研究上,代表性数据集如 Omniglot 与 MiniImage。2003 年李飞飞教授首次提出了 One-shot Learning 问题并利用贝叶斯框架<sup>[17,18]</sup>进行学习。BMLake 等人<sup>[19]</sup>通过利用贝叶斯规划学习,结合一些先验知识完成类人概念学习的验证,证明了机器可以利用背景知识通过少量样本完成概率估计。Santoro 等人<sup>[20]</sup>基于 Meta Network,提出了使用 LSTM 等记忆增强的 RNN 的模型方法来解决 FSL 任务。Ravi 等人<sup>[21]</sup>研究了在少量数据下,学习一个模型参数的更新函数或更新规则,来解析基于梯度的优化算法失败的原因。Finn<sup>[22]</sup>提出的方法无需关心模型的形式,也不需要为元学习增加新的参数,用少量的迭代步骤就可以获得良好的泛化能力。

近年来,小样本学习已不仅只用于图像方面的研究,在自然领域方面的研究也颇有成效,已出现了适应自然语言处理领域的一些模型和代表性数据集。代表性数据集包括 ODIC 数据集、ARSC 数据集<sup>[23]</sup>与 FewRel 数据集<sup>[24]</sup>。模型使用情况为:Gao<sup>[25]</sup>等人提出了一种基于混合注意力的原型网络、Yu<sup>[23]</sup>等人提出使用多种度量方式融合来解决跨领域的 FSL 问题。同时,阿里巴巴小蜜北京团队基于 Metric Based 方法提出的 Encoder-Induction-Relation 三级框架打开了对话工厂平

台中小样本的研究与应用的大门。

现有的深度学习文字识别算法大都是基于海量的训练样本,对于战国简文字这种数据量较少的数据而言,训练效果不好,准确率不理想。小样本文字识别算法大多是针对背景相对纯净的英文或手写体的是识别。而战国简上的文字,很多因为竹简纹路、笔墨滴溅、年代久远、保存不当等因素造成字体残缺、高噪声等问题。基于此现状,本文提出整理了一个存在强噪声、高残缺、不均衡的战国简文字识别数据集;选择孪生网络结构模型与深度度量学习的方法来识别战国简文字,并得到了较好的识别准确率。

### 1.3 本文研究的内容

本文的主要内容是围绕战国简文字识别技术展开的。首先阐述了战国简文字的研究历史与现状,然后对小样本的研究现状进行了介绍与总结,并着重介绍了孪生网络模型和其他的一些深度度量学习方法;其次提出并整理了战国简文字识别数据集,阐述了其提出背景与整理过程及最终版竹简数据集的内容情况。接着介绍了基于孪生网络的小样本战国简文字识别算法与基于其他深度度量学习算法的战国简文字识别。最后将小样学习算法引用到竹简文字的识别平台上进行识别显示。本文旨在研究与以往数据集不同的小样本学习算法,为战国古文字的研究提供一个新的方向。

本文研究的内容主要分为三个方面:第一,标注整理了基于出土战国简的图像文本数据集,该部分主要介绍该数据集及其特征;第二,分析研究小样本学习算法,提出基于孪生网络的小样本战国简文字识别算法用于战国简文字识别,并在战国简数据集上使用与本文算法基础网络结相同的其他深度度量学习方法与本文算法的识别效果进行对比。分析对比了本文算法与现有小样本算法在 Omniglot 数据集与 Bambooslips 数据集上识别效果,及不同深度度量学习方法在 Bambooslips 数据集不同分类上的识别效果。第三,基于 pycharm+PyQt5+python 搭建了战国简文字识别软件平台,将待识别的竹简图片导入,经过已训练好的模型,输出其对应的文字。将战国简文字识别算法应用于实际工作中,可以减少战国简古文字识别对专业人士的依赖性。

## 1.4 文本组织结构

考虑到战国简文字数据集与目前公开数据集的不同特性,使用已有算法识别准确率不是很好,且未发现有战国简文字识别软件的公布。本文针对战国简文字识别技术与软件平台进行研究探讨,经过多次尝试,最终确定使用孪生神经网络作为本文战国简文字识别技术的基础网络,并使用深度度量学习的方法对竹简数据集进行评测。最后开发了一个软件,将战国简文字识别落于实处,方便他人对战国简文字的检索与查找。本文的组织结构如下:

第一章为绪论部分,首先分析了文字识别与战国古文字的现状,其次介绍了小样本学习的分类以及现有的模型结构及研究领域,最后介绍了本文的研究内容。

第二章为相关工作介绍,首先介绍了卷积神经网络定义,对卷积神经网络的各层结构及网络训练方面进行详细的阐述。随之引出对卷积孪生网络的介绍,为后续章节基于孪生网络的小样本文字识别算法做了铺垫。接着对目前小样本学习的几种方法做了详细介绍,最后对目前比较火热的两个深度度量学习方法进行了详细的阐述。

第三章为本文收集整理战国简数据集介绍,对竹简数据集提出的背景及战国简目前的研究现状做了详细介绍。分析了竹简文字识别存在的一些问题。其次介绍了在竹简数据集整理过程中遇到的困难及解决的办法,以及最终整理出的竹简数据集的内容情况。

第四章为战国简文字识别算法的介绍。首先对基于孪生网络的小样本文字识别算法进行了阐述。介绍了不同网络结构的孪生网络架构及数据输入流程等内容。并在多个算法上进行实验结果对比,不仅输出整体类别的准确率,而且输出单个例别的识别准确率。其次,将其他深度度量学习的方法应用于竹简文字识别中,对比其在战国简数据集上的识别性能,进行结果分析。

第五章为战国简文字识别软件平台。基于 `pycharm+PyQt5+python` 设计实现了一个战国简文字识别系统,可以快速、准确的满足竹简文字识别的需求。

第六章为总结与展望。总结本文的主要内容,并分析了论文的不足之处,以便后续深入研究。



## 第 2 章 相关工作

从 20 世纪 50 年代至今,随着诸如战国简与简帛资料等地下材料的不断出土发现,战国古文字的研究已愈演愈烈。但古文字的研究大多是专业人士对古文字的形体变化及其发展历程以及古文字材料的意义研究,将计算机识别技术用于古文字研究相对较少。战国简因其埋藏时间较久、抄传未必同一人,故竹条文字有所损坏且形态不一,再加上样本数量颇少,对战国简文字识别造成一定的困难。基于战国简数据集的特征,考虑使用单样本学习来研究本文战国简文字的识别。本章介绍了一些关键的相关知识。

### 2.1 卷积神经网络

卷积神经网络<sup>[26]</sup>是应用最广泛的神经网络之一,最早是由 Wiesel 与 Hubel 提出。卷积神经网络不仅适用于图像识别领域的研究,在语音识别与自然语言处理方面也有不错的表现。其结构有很多种,包括 LeNet、AlexNet、VGGNet、Google Inception Net 等,但基本组成是不变的。具体结构如图 2-1 所示。

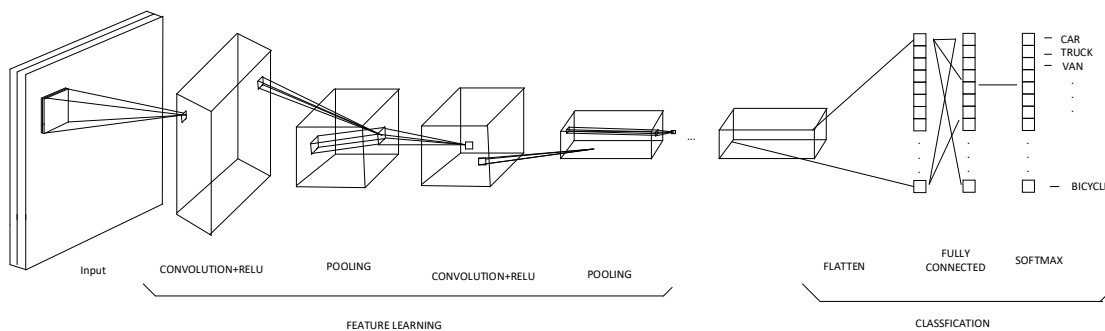


图 2-1 卷积神经网络结构

输入层即输入的图片等数据,常用尺寸一般为像素值的矩阵:[宽×高×深]。此处深度表示 R、G、B 通道。且通常输入层应该是 2 的整数倍。

#### 2.1.1 卷积层

当图像输入网络结构中时,卷积层是第一层关卡,目标是将输入数据的特征

提取出来，经过多次的卷积过程，可以得到深层次的特征信息。卷积核或过滤器（filter）是卷积层的重点之处，是指用来做卷积计算的部分，一般大小不限，深度不限，由自己设定。简单来说，就是滤波器在图像上滑动进行时，在每个位置上将该像素点及其领域像素点做内积，计算得到一个输出矩阵。其计算公式为：

$$W_{out} = (W_{in} - F + 2P) / S + 1 \quad (2-1)$$

其中， $W_{in}$  为输入矩阵的大小， $F$  是过滤器的大小。 $S$  为步长，表示过滤器在输入矩阵上每次滑动的单元数距离。 $P$  即 **Padding**，因为在卷积计算过程中，矩阵边缘元素在计算过程中容易被忽略，从而计算得到的输出矩阵会因边缘元素得不到计算而相对变小。因此会在矩阵的四周补零（即 **padding**），解决上述所说问题，其大小即为  $P$ 。例如当  $P=1$  时，原  $3*3$  的矩阵 **padding** 后的结果如图 2-2 所示。

0	0	0	0	0
0	2	1	0	0
0	1	3	4	0
0	2	4	1	0
0	0	0	0	0

图 2-2 Padding 过程

一般，都会将卷积层与激活函数同时使用，因为卷积层是线性的，而激活函数是非线性的，在卷积层时候加非线性激活函数可以保证网络机构适应复杂的非线性问题。

### 2.1.2 池化层

池化层的作用是用来特征降维，因此，通常会将池化层插入到卷积层之间，对卷积层提取到的特征信息维度进行降维，不仅减少了网络参数的个数，还减少了网络中的计算量。除了特征降维，池化层操作不会丢失图片原有的特征，也可在某种程度上防止过拟合。常见的池化操作有：平均池化（**Mean pooling**）、最大池化（**Max pooling**）与随机池化。平均池化是将图像划分为若干图像小区域，然后取这些小区域内的平均值，这样的操作可以很好的提取背景信息；最大池化

是对每个图像小区域中取最大值作为该区域的池化值，这样子会对图像纹理特征提取有优势；随机池化则是对特征区域元素按照其概率大小随机选择，一般在 Max pooling 与 Mean pooling 之间。通常情况下，Max pooling 是最为常见的。如图 2.3，对 3\*3 的矩阵进行最大池化操作，设定池化窗口为 2\*2，步长为 1，输出结果就是各个区域的最大值。

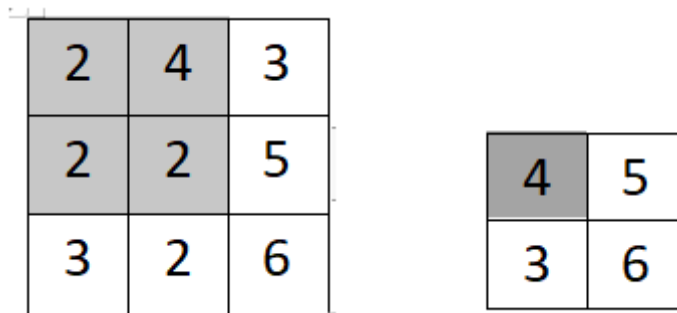


图 2-3 Max pooling 操作

### 2.1.3 全连接层

全连接层是建立在卷积层的基础上的，卷积层取的是局部特征，而全连接层就是把通过卷积层提取的局部特征重新通过权值矩阵拼成的一张完整的图，它相当于一个分类器。由于用到了所有的局部特征，所以被称作全连接。softmax 是卷积神经网络中常用的分类器之一，而 softmax 回归是将二分类回归应用到了多分类的问题上。

### 2.1.4 网络训练

卷积神经网络训练的目标是求解最小化代价函数对应的参数权重，一般分为前向传播与反向传播两个阶段。其过程如下：

- (1) 权值初始化，选取一个样本作为网络的输入；
- (2) 网络各层对输入的数据分别进行处理操作后，得到前向传播的最终结果值；
- (3) 计算求出 (2) 中输出值与实际目标值之间的误差；
- (4) 若误差不在容许范围内，计算网络层中神经元的误差，求梯度后更新权值；

(5) 不断进行迭代更新，直到找到最佳参数。

在整个过程中，最关键的在于梯度的求解。梯度的求解在反向传播阶段进行，其过程为先经由前向传播从输入计算得到输出，然后反向计算得到一个梯度积累的过程，通过各层的权值与偏置的梯度来更新参数，从而调整网络结构。

## 2.2 孪生网络

### 2.2.1 基本思想

孪生网络 (Siamese network) 可以称之为“连体的卷积神经网络”。为什么称之为连体呢？是因为孪生网络是由两个神经网络组成的。其作用就是通过相似性度量来判断两个输入的一致性程度。如图 2-4 中，孪生网络中 Network1 与 Network2 是相同的网络结构或同一个网络结构，因其权值共享，用于处理两个输入“比较类似”的情况。Network1 与 Network2 也可以是不同的网络结构，此时称之为伪孪生网络 (pseudo-siamese network) 用于处理两个输入“有一定差别”的情况。

孪生网络具有其独特的特点，它的训练样本作为输入数据是以图像对的方式进行输入。当将输入 feed 进神经网络后，每个神经网络都会将各自的输入经过一系列操作得到一个输出值。然后将两个网络的输出值进行计算得到损失值，将该损失值作为度量两个输入样本相似性程度的标准。因为孪生网络可以用于二分类，所以对于 loss 计算，softmax 可以选择使用，但却不是最优的。Contrastive Loss 一般常被用在 siamese network 中，其公式如下：

$$L = \frac{1}{2N} \sum_{n=1}^N yd^2 + (1 - y) \max(\text{margin} - d, 0)^2 \quad (2-2)$$

其中， $y$  是判断两个输入匹配与否的标签，当  $y=0$  时，表示两个输入样本不是同一类；当  $y=1$  时，表示两个输入样本是同一类。 $d = \|a_n - b_n\|_2$  是两个输入样本的欧式距离。 $\text{margin}$  代表两个输入样本不相似的最小欧氏距离的阈值，当  $d > \text{margin}$  时，loss 直接为 0。

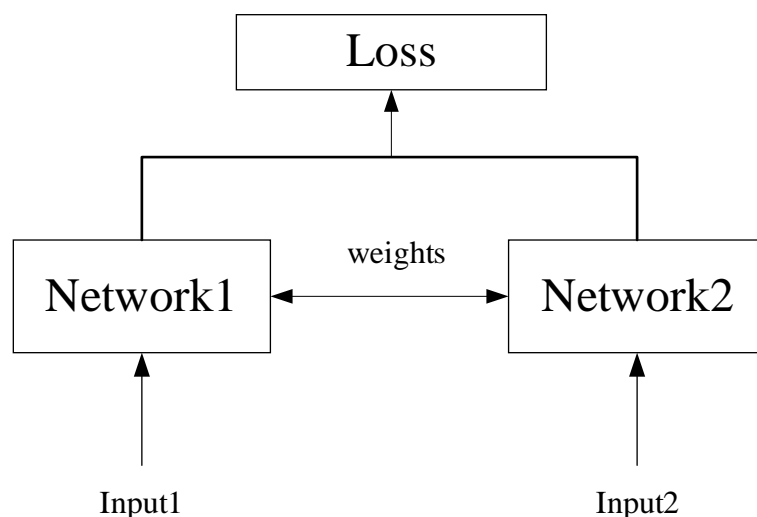


图 2-4 孪生网络结构

### 2.2.2 孪生网络结构分类

Siamese network 一般采用对称卷积网络,以两个大小相同的图像作为输入,输出为这两个图像的相似性度量,判断输入的两张图片是否为同一类。通常 Siamese network 的拓扑形式可以根据每个输入图像块被组合的位置分成 3 个基本的类型<sup>[27]</sup>,如图 2-5 所示。

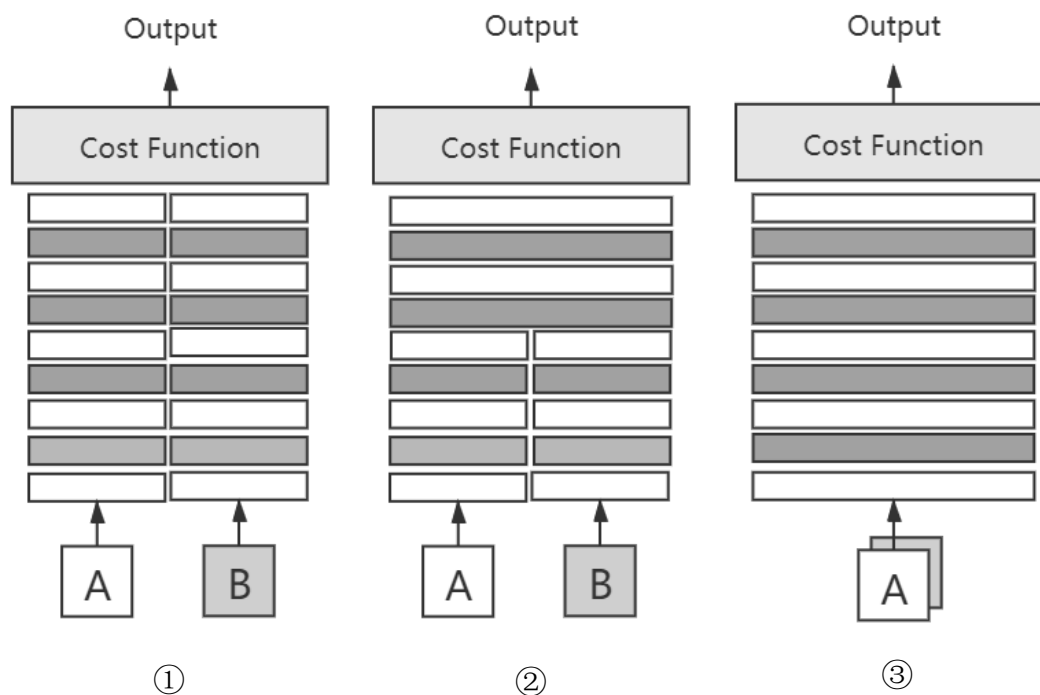


图 2-5 孪生网络拓扑结构

图中 A、B 表示待输入的图片。

结构①：输入 A 和 B 经过相同网络结构与参数的网络分支，对提取的特征计算它们的距离度量函数作为代价函数，以使得相同对象的距离接近，而不同对象的距离变大。旨在可以学会把不同的类别分开。

结构②：输入 A 和 B 先经过部分相同网络结构与参数的网络分支；然后将各分支的顶层级联起来，对获得的新特征添加多层卷积；最后使用标准的 softmax log-loss 函数作为代价函数。

结构③：对输入 A 与 B 叠加组合作为新网络的输入，通过网络滤波生成输出。

通常，前两种方法适用于分类任务，第三种方法一般适用于细节结构比较的任务。本文使用了结构①的拓扑结构。

### 2.2.3 孪生网络单样本实际应用

度量学习的基本思想是学习数据之间的相似性，而孪生网络属于度量学习，能很好的衡量两个数据数据是否为同一类别。目前，孪生网络的应用已取得了很大的进展。

单样本学习可以追溯到 2000 年初，Li 等人<sup>[28]</sup>开发的变分贝叶斯框架工作，该模型对于只有少量类可用的给定类别，可以利用之前学习过的类去对未知类进行预测来用于单样本图像分类。Lake 等人<sup>[29]</sup>从认知科学的方面提出层次贝叶斯程序学习（Hierarchical Bayesian Program Learning, HBPL）方法用来解决字符识别的单样本学习，模拟字符生成的过程让机器学习，然后让机器生成新的字体。然而，HBPL 因需要在数据集上学习笔画的生成模型，所以需要更加复杂的标注，因此对于图片上的汉字识别效果不是很好。

孪生网络目前在各个领域都有不错的发展。在半监督领域，Motiian 等<sup>[30]</sup>将对抗网络与孪生网络结合，在目标空间中，最大化不同类之间的差异，最小化相同类直接的差异。在目标跟踪领域，Li 等<sup>[31]</sup>为了用深度网络结构克服孪生网络的平移不变性，提出了基于层次的互相关操作特征聚集结构，有助于从多个层次特征中判断相似度。在小样本领域，Hsiao 等<sup>[32]</sup>通过孪生网络计算两个输入图片

的相似度来进行单样本图像识别；也有人利用孪生网络衡量两个输入样本的相似性特点，进行人脸识别；柳青林<sup>[33]</sup>基于结合孪生网络与特征模板匹配构建了一种基于小样本学习的目标匹配算法。而本文则探讨的市基于孪生网络的战国简文字识别技术。

## 2.3 小样本学习方法

### 2.3.1 数据增强

在如今信息爆炸的时代，大多深度学习算法都是基于海量数据而言的。然而事实上并不会每一类问题都有充足的数据量。因此小样本学习目前在各个领域的研究中占有很重要的研究价值。解决 FSL 问题，数据增强是一个行之有效的方法。该方法可以利用先验知识扩充数据，丰富样本信息，使数据足以满足算法所需。最常见的数据增强方法是采用手工规则方法来转换数据集中原始数据来扩充训练集。包括对图像反转、旋转、翻转、平移、缩放、反射、裁剪等操作实现训练集的扩充。

另一种方法是使用生成式对抗网络(GAN)<sup>[34]</sup>，GAN 可以通过扩充训练集样本量来达到增强数据的目的。其组成分为：生成网络(Generator)、判别网络(Discriminator)。生成网络(G(z))的作用是根据随机接收的噪声 z 来生成样本；判别网络的作用则是判断生成的样本是否是真实的，输出生成样本真实的概率。其数学公式为：

$$\min_G \max_D V(D, G) = E_{x \sim P_{data}(x)} [\log D(x)] + E_{z \sim P_{data}(z)} [\log(1 - D(G(z)))] \quad (2-3)$$

x 代表训练集中真实样本，z 代表 Generator 接收的随机噪声，G(z)代表 Generator 生成的样本。D(x)代表训练集中样本真实的概率，D(G(z))表示生成样本为真的概率。Han 等人<sup>[35]</sup>提出了边界和距离边界合成少数类过采样技术并生成模拟样本的 WGAN 方法，该方法通过用 Wasserstein 距离代替真实样本与生成样本之间的差异可以很好对模型进行训练。Giovanni Mariani 等人<sup>[36]</sup>提出的 BAGAN 方法完美的解决了不平衡数据集中数据量少的问题。

### 2.3.2 迁移学习

迁移学习 (Transfer learning) 就是将某一个任务或者领域上学习到的知识或训练好的模型参数应用到相关但不相同的问题或领域中来帮助训练。与其他的多任务学习不同, 迁移学习中的原目标与目标任务, 可以一致也可以不一致, 迁移学习的重心在目标任务上。在传统的多任务学习中, 训练样本与测试样本必须独立同分布, 而迁移不仅克服了这个弊端, 还有着其独特的特点。一、迁移学习可以很好的解决小数据问题, 不用事先给定上千上万个样本就能实现迁移。例如, 我们看见一只麻雀的照片后, 在以后生活中若再见到麻雀, 我们能很快的认识这是麻雀。二、迁移学习具有很好的可靠性。无论是在原先领域还是跨领域到周边领域, 迁移学习都能实现它的作用, 即使是将周边环境做了一定的改变。

迁移学习应用在在小样本学习中, 主要是从相关领域迁移标注数据或者知识结构, 来解决目标领域样本量不足的问题, 目前成果颇多。比如自学习、特征选择以及控制权重技术。领域自适应 (Domain adaptation)<sup>[37]</sup>也是迁移学习的一种, 它的特点是任务相同, 但是应用的领域不同。比如我们要做一个人物分析, 影剧的弹幕评论是原域数据, 手游的客户评论是目标域数据。因为小样本学习与零样本学习都需要将从其他领域或任务学到的先验知识迁移过来进行学习, 因此将小样本学习与零样本学习看作是迁移学习中的极端情况。迁移学习目前多被广泛用于目标检测、图像分类等领域中。

### 2.3.3 元学习

元学习 (meta learning) 通俗来讲就是 learning to learn—学会学习。通常对于机器学习算法, 让同一个模型去解决不同的问题, 一般效果会很差。然而, 人类就可以根据以往的经验快速的适应各种新的情况。因此, 希望机器能够像人类一样可以利用学过的经验去解决新的问题。元学习就是朝这个方向发展的关键一步。通过元学习, 让机器学会学习, 利用之前学习到的经验快速的适应新情况并指导新的学习任务。元学习的主旨在于如何能够建立一个模型去利用之前经验快速学习新任务, 而小样本学习的主旨在于如何建立具有良好泛化能力的模型, 因此, 利用元学习来解决小样本学习问题是一个不错的选择。

目前，元学习通常被用在探索好的网络结构、快速强化学习、优化神经网络和超参数与小样本图像识别等方面。已有的元学习方法包括：预测梯度<sup>[38]</sup>、预测损失<sup>[39]</sup>、预测参数、寻找任务共性等。与迁移学习相比，迁移学习通常研究的是一项任务到另一项任务的迁移，而元学习研究的是多任务中的共性。因此利用元学习进行的少样本学习大多侧重于图像分类问题，希望对于从未训练过的新任务，只需借助每类别中少量的样本就可以正确的分类。我们希望能够利用元学习进行小样本学习，实现快速学习能力与泛化能力的功能，让模型根据新任务进行调整，实现学会学习的能力。

### 2.3.4 小结

根据解决问题角度的不同，小样本学习方法可分为上述的三种：数据增强、迁移学习、元学习。数据增强是从提高数据集样本容量的角度解决因样本量不足而导致的过拟合或泛化能力不强的问题；迁移学习是从相关领域迁移标注数据或者知识结构，来解决目标领域样本量不足的问题；元学习解决的是学会如何学习的问题。其常用方法如表 2-1 所示。

表 2-1 小样本学习分类

	常用方法
数据增强	1、手工规则方法：反转、旋转、翻转、缩放、反射、裁剪、高斯噪声、高斯模糊等 2、使用 GAN 生成新样本
迁移学习	基于特征 <sup>[40]</sup> 、基于关系 <sup>[41]</sup> 、基于共享参数 <sup>[42]</sup>
元学习	基于度量 <sup>[43]</sup> 、基于模型 <sup>[44]</sup> 、基于优化 <sup>[45]</sup>

在元学习的小样本图像分类的常用方法中：基于度量的元学习旨在将图像映射到一个度量空间并使用某种度量方式计算不同图像样本的差异，借助了最近邻的思想完成分类；基于模型的元学习通过模型结构的设计快速在小样本上更新参数，建立输入和预测值的映射函数；基于优化的元学习目的是使网络具有一个好的初始化，通过调整优化方法来完成小样本任务分类。本文算法使用了基于度量的小样本学习算法进行战国简文字的识别。三个方法中各自常用结构如表 2-1 所

示。

表 2-2 各分类常用方法

	常用方法
基于度量的元学习	Siamese Network <sup>[46]</sup> 、Match Network <sup>[47]</sup> 、 Prototype Network <sup>[48]</sup>
基于模型的元学习	Meta Network <sup>[49]</sup> 、Deep Meta Learning <sup>[50]</sup>
基于优化的元学习	MAML <sup>[22]</sup>

## 2.4 其它深度度量学习方法

### 2.4.1 SoftTriple

SoftTriple Loss 是由文献<sup>[51]</sup>提出的对三元组损失 (Triple Loss) 与 softmax 的改进。Triple Loss 最初由文献<sup>[52]</sup>提出, 通常是在个体级别的细粒度识别上使用。Triple 是一个三元组  $\langle a, p, n \rangle$ ,  $a$  代表从训练集中随机选的样本,  $p$  代表随机选取与  $a$  属于同一类的样本,  $n$  代表与  $a$  不同类的样本。三元组的构成使 Triple Loss 可以学习一个好的嵌入 (embedding), 让相似图像在 embedding 空间里是相近的, 不同图像在 embedding 空间是远离的。公式化的表示是:

$$\|f(x_i^a) - f(x_i^p)\|_2^2 + \alpha < \|f(x_i^a) - f(x_i^n)\|_2^2 \quad (2-4)$$

式中  $f(x_i^a)$ 、 $f(x_i^p)$ 、 $f(x_i^n)$  表示三元组中每个元素训练一个参数共享或不共享的网络得到的特征表达。 $\alpha$  表示阈值, 控制正负样本的距离, 从而判断是不是属于同一个类别。

但是在使用 triple loss 训练时, 每次只有一个小批次的数据可以使用, triple 也只能在这个小批次中取样, 这会导致网络学习过于关注局部信息。Softmax loss 虽然可以学到有用的嵌入, 但 Softmax 假设每个类只有一个中心, 而在真实数据中, 一般都包含多了中心。因此为 soft loss 引入多个中心, 而这每一个类可以看作是只有一个中心的平滑 triple loss, 因此得出 SoftTriple Loss, 公式如式 2-5 所示。SoftTriple Loss 可以自适应地选择类别中心的个数; 可以在多个中心更好地捕获隐藏的数据分布, 因为其有助于减少类内的方差; 可以在不对 triple 进行采样的前提下进行优化。

$$l_{SoftTriple}(X_i) = -\log \frac{\exp(\varepsilon(S'_{i,y_i} - \delta))}{\exp(\varepsilon(S'_{i,y_i} - \delta)) + \sum_{j \neq y_i} \exp(\varepsilon S'_{i,j})} \quad (2-5)$$

### 2.4.2 CosFace

CosFace 由文献<sup>[53]</sup>提出的利用 large margin cosine loss (LMCL) 方法训练得到的方法。通过 L2 归一化特征和权重向量消除径向变化，将 softmax loss 重新改造为 consine loss。Softmax loss 公式如下：

$$L_s = \frac{1}{N} \sum_{i=1}^N -\log p_i = \frac{1}{N} \sum_{i=1}^N -\log \frac{e^{f_{y_i}}}{\sum_{j=1}^C e^{f_j}} \quad (2-6)$$

其中， $p_i$  为  $x_i$  正确识别的概率， $N$  为训练的样本数， $C$  为类别总数， $f_j$  表示当前类别权重  $W_j$  与特征  $x$  的乘积。

$$f_j = W_j^T x = \|W_j\| \|x\| \cos \theta_j \quad (2-7)$$

对  $W$  与  $x$  分别进行 L2 归一化操作，考虑到  $\|x\|$  对分数没有贡献，将其固定为  $s$ ,  $W$  定为 1，因此得到：

$$L_{ns} = \frac{1}{N} \sum_i -\log \frac{e^{s \cos(\theta_{y_i,i})}}{\sum_j e^{s \cos(\theta_{j,i})}} \quad (2-8)$$

最后引入度量的约束 consine margin，即  $\cos(\theta_1) - m > \cos(\theta_2)$ ，得到最终的损失函数 LMCL：

$$L_{lms} = \frac{1}{N} \sum_i -\log \frac{e^{s(\cos(\theta_{y_i,i}) - m)}}{e^{s(\cos(\theta_{y_i,i}) - m)} + \sum_{j \neq y_i} e^{s \cos(\theta_{j,i})}} \quad (2-9)$$

CosFace 就是最大化超球面特征空间中的决策裕度，实现类间差异的最大化和类内差异的最小化，当损失值收敛的时候，让类内更加聚拢。在训练阶段，LMCL 指导卷积网络对不同类图片学习得到较大余弦间距的特征；在测试阶段，测试数据输入到 CosFae 模型中提取特征，然后通过计算余弦相似性度量来进行验证与识别。

### 2.4.3 最新度量算法

Wang 等人<sup>[54]</sup>在研究过程中发现了“缓慢漂移”现象，即在整个训练过程中，

实例的嵌入实际上以一个相对缓慢的速度飘逸。基于此现象，论文提出了一个跨批处理内存（XBM）模块来记录和更新最近的小批处理的深层特性。这种方法能够用极小的代价提供巨量的样本对，为基于 pair-based 的深度度量学习方法取得巨大的效果提升。突破了加权与聚类的传统思想，很好的解决了深度度量学习的痛点。

对于困难样本的生成，一般使用生成对抗网络和自编码器实现。虽然达到一定的效果，但会导致超参数增多、优化难度增加、训练读取变慢等问题。基于此问题，Gu 等人<sup>[55]</sup>提出了对称生成（symmetrical synthesis）的生成困难样本的新方法。给定同一类中的两个原始特征点，首先会以彼此作为对称轴生成伪造点，其次在原始点和伪造点内执行难例挖掘去选择信息量大的负对去计算度量学习损失值。Ko 等人<sup>[56]</sup>受到查询扩展与数据库扩充的启发，提出了一种在嵌入空间内度量学习损失的扩充方法—embedding expansion。该方法首先通过特征点的组合生成包含增强信息的伪造点，然后使用困难负样本对挖掘学习具有更多信息的特征表示。

基于排序的指标是计算机视觉模式性能评估中使用最广泛的标准之一。但却存在直接优化困难的弊端。文献<sup>[57]</sup>提出了一种对于区分基于小批量梯度下降的排序度量标准的有效方法。讨论了由于使用基于排序度量作为优化目标而产生的优化不稳定性和监督信号的稀疏性，提高了标准图像检索数据集上性能。

枢纽点问题（Hubness Problem）是指在高维空间中，某些点会成为大多数点的最近邻点。针对这一问题，文献<sup>[58]</sup>等人采用余弦度量学习缓解了零样本学习中的 Hubness Problem，并首次提出了余弦是模型不可知的，即无论模型长什么样子，只要有一样的任务，一样的输入输出就可以了。

## 2.5 本章小结

本章中首先对卷积神经网络进行了一个简单的阐述，并对卷积神经网络各层结构及网络训练等方面进行了简单的介绍。然后通过对孪生网络的基本思想、结构分类、在单样本的应用以及小样本学习的三种方法做的详细介绍，为后续章节基于孪生网络的小样本文字识别算法做了铺垫。最后介绍了深度度量学习的两种

方法，为后续基于深度度量学习的战国简文字识别奠定了基础。



## 第3章 战国简数据集

### 3.1 数据集提出的背景

简牍是中国古代先民在纸张发明之前书写典籍、文书等文字载体的主要材料,是我国最古老的图书之一。简牍几乎与甲骨文、金文同时出现,春秋到东汉末年 是简牍最盛行的时期,纸张发明后,又与纸张并行几百年直至东晋末年简牍制度才结束。因此简牍被看作研究古文字形体发展与演变的重要依托工具。目前为止,已出土的战国竹简有郭店楚墓竹简<sup>[59]</sup>(简称“郭店简”)、清华大学入藏的楚简<sup>[60]</sup>(简称“清华简”)、上海博物馆入藏的楚简<sup>[61]</sup>(简称“上博简”)、安徽大学入藏的竹简<sup>[62]</sup>(简称“安大简”)等。但已有的战国简研究都是依托于对古文字形体结构、演变规律或内容结构等理论上的探讨研究,对古文字识别与人工智能文字识别的交叉研究目前还未发现。基于该背景下,本人在导师的指导带领下独自整理准备了基于“安大简”的战国简文字数据集,并使用基于孪生网络的小样本学习算法对其进行识别,减少战国简文字识别对专业人士的依赖。

#### 3.1.1 战国简文字研究现状

在导师的指导下,本人查阅大量文献与资料,发现对于已出土战国简的研究一直在进行,但大多是从文字、文章、书写形式三个方面为出发点。依据人工智能手段对战国简文字识别的研究较少。因此,创新性的提出将古文字研究与人工智能小样本文字识别相结合,为后续研究提供一种新的思绪。目前,具有代表性的出土竹简有“上博简”、“清华简”与“安大简”。

“上博简”是上海博物馆购入约 1200 余枚的一批战国楚竹书竹简。该批竹简书体为篆书的楚系文字,自 2001 年到 2012 年,已共出版九册,公布约 60 余篇。上博简着重研究的是书写形式,对当代书法艺术特点和艺术风格有很大的影响,促进了当代书法艺术的繁荣,扩大了战国简研究的影响。其代表图如图 3-1 所示。



图 3-1 上博藏楚竹书

“清华简”是 2008 年由清华大学收藏的一批楚地出土的战国简，包括断简残简在内一共 2388 枚，是现如今竹简出土数量最多的一批。这批竹简大多是以经、史类书籍为主，内容丰富，不仅对古文字研究提供了重要的参考价值，也对探索中国传统文化与历史有所帮助。其代表图如图 3-2 所示。

“安大简”是 2015 年初安徽大学与中国古代文明研究协同创新中心入藏的一批战国简。安大简共有 1167 个，长短不一，最长的约 21.3，最短的约 48.4，简宽 0.4~0.8，每简字数约 27~35 之间，完简较多，且字与字的间距比较大，字迹清晰。目前楚史资料中，所知系统最完整、所知时代最早的楚史资料就是安大简中的“楚史类”文献，其极大的丰富了楚史的研究内容。其代表性竹简如图 3-3 所示。图 3-4 展示了安大简的单条竹简示例。。



图 3-2 清华大学藏战国竹简尚书《金縢》



图 3-3 安徽大学藏战国竹简《周南》



图 3-4 安大简藏单独竹简图片

对比已整理的战国简，我们选取“安大简”作为本论文数据集制作的原型。一方面是因为，安大简是目前为止战国简研究的最新发现，研究内容较之以往更加完善；另一方面，安大简完简较多，字与字之间均有留白且间距较大，字迹清晰。最重要的一点是每条竹简均有简本与汉字的对照，方便对数据的标注与整理。

### 3.1.2 战国简文字识别面临的困难

因为竹简文字是古人写在竹片上的文字，其本质上来说是手写体。但这种竹简手写体文字识别与现有手写体识别又有很大的差别。其原因分析如下：

(1) 战国古文字的载体是由竹、木制成的简和牍。竹、木本身就有一定的纹路，对字体的识别造成影响，如图 3-5 (a) 所示。

(2) 竹简上的文字并非是用刀刻画上去的，而是依托人工使用毛笔书写上去的。因此，难免会在竹简上留下墨痕或书写过程中溅落的点滴墨迹，如图 3-5 (b) 所示。

(3) 战国时期距今已时代久远，许多竹简当初都是作为陪葬品被埋于地下，因时间久远，竹简腐化损坏比较严重，造成竹简上文字残缺，简背有刮痕，如图

3-5 (c) 所示。

(4) 竹简发现之后，要经过人为的修护与保存，经过一系列操作之后可能会造成文字纹路的干扰，如图 3-5 (d) 所示。

(5) 因竹简年代久远，文字形体的变化造成竹简中存有许多古文字字形与现如今文字对应不上的情况，也就是存在许多现如今不存在的字体，对数据集的标注造成了一定的困难。

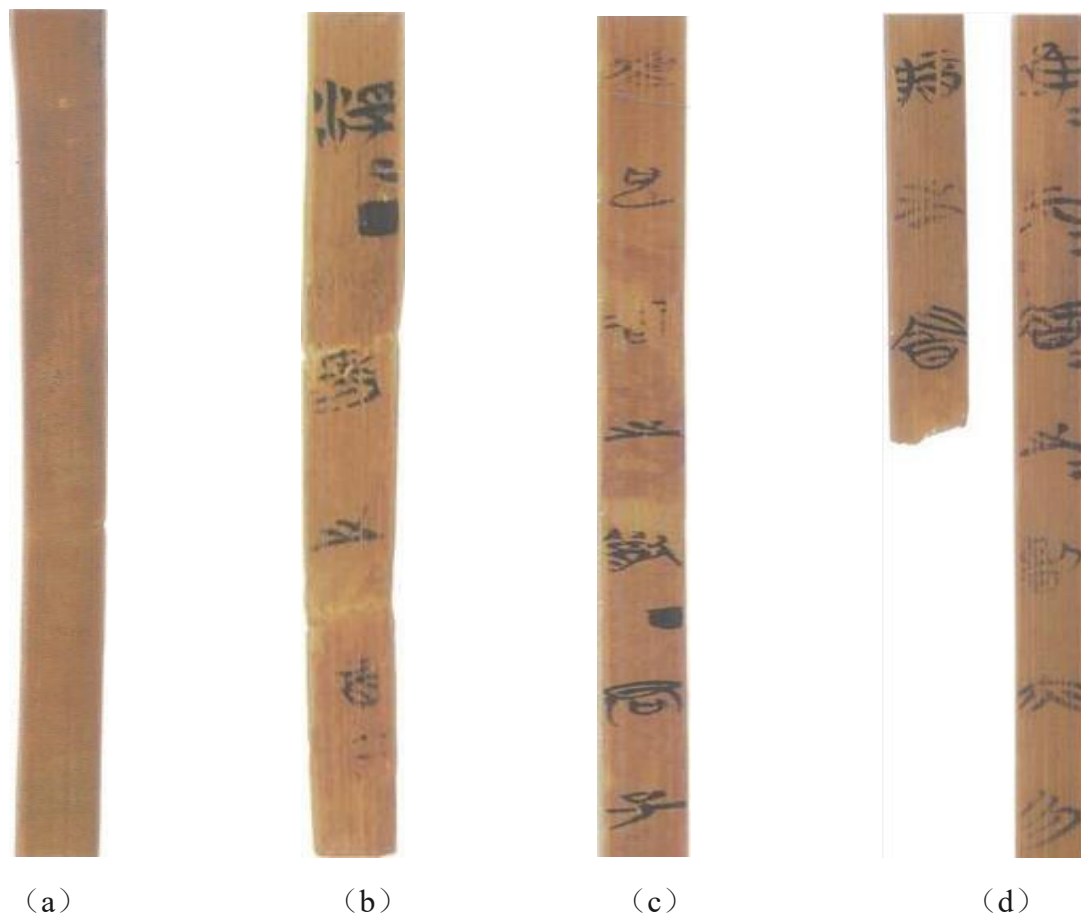


图 3-5 战国简牍示例

## 3.2 数据集描述

### 3.2.1 数据集标注方式及工具

本论文数据集图像来源于安徽大学藏战国简已出版的有关《诗经》部分的 pdf 文件，文件对六国国风竹简简本与汉字的对应及释文注释做了详细的介绍。包括《周南》十篇、《召南》十四篇、《秦》十篇、《侯》六篇、《邶》七篇、《魏》

十篇，共有单张竹简图片 347 张。在标注工作之前需要解决的首要问题是如何将竹简图片单独完整的保存下来以及竹简上部分古文字对应的今本不存在。对于如何保存竹简图片的问题，本人在对比若干裁剪工具后，使用系统自带的裁剪工具可以在保证图片分辨率的情况下将这 347 张图片从 pdf 文件中裁剪保存；对于竹简上部分古文字现如今不存在的问题，本人首先翻阅古文字字典去查找，对于现如今不存在的竹简字体采取了造字功能去实现。实现造字功能的软件使用 win10 系统自带的专用字符编辑程序。总体共造字 285 个，如图 3-6、3-7。随后，将竹简图片的 png 文件及其竹简图像中的简本对应的今本文字的 txt 文件以同名称的方式保存在不同的文件夹内。至此，标注前期工作完成，可以开始标注任务。标注工作及其准备工作是在计算机学院张重生教授的指导下本人独立完成。

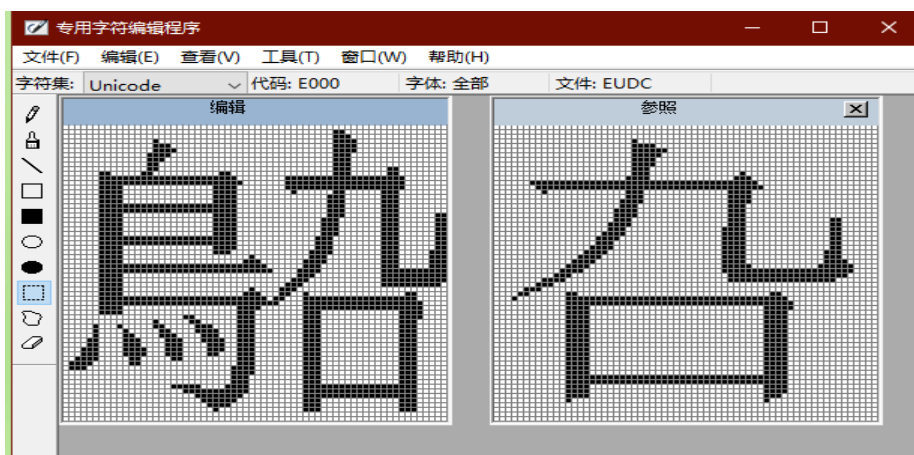


图 3-6 专用字符编辑软件



图 3-7 所造字文件

对于每一张图片采用单字标注法，将每个字体都单独对应一个标注框，记录其标注框对应的四个坐标及其今本文字，其内容格式为：左上 x，左上 y，右上 x，右下 y，“字”。标注功能实现借助了图片标注工具 LabelImg，如图 3-8 所示。最后，将其保存的 XML 文件用自己写的小脚本转换为 txt 文件。标注完成之后，将标注结果经过一遍非技术校对，尤其查看自己所造字体在文件里是否正常显示，或者图片与字体文件是否一致。

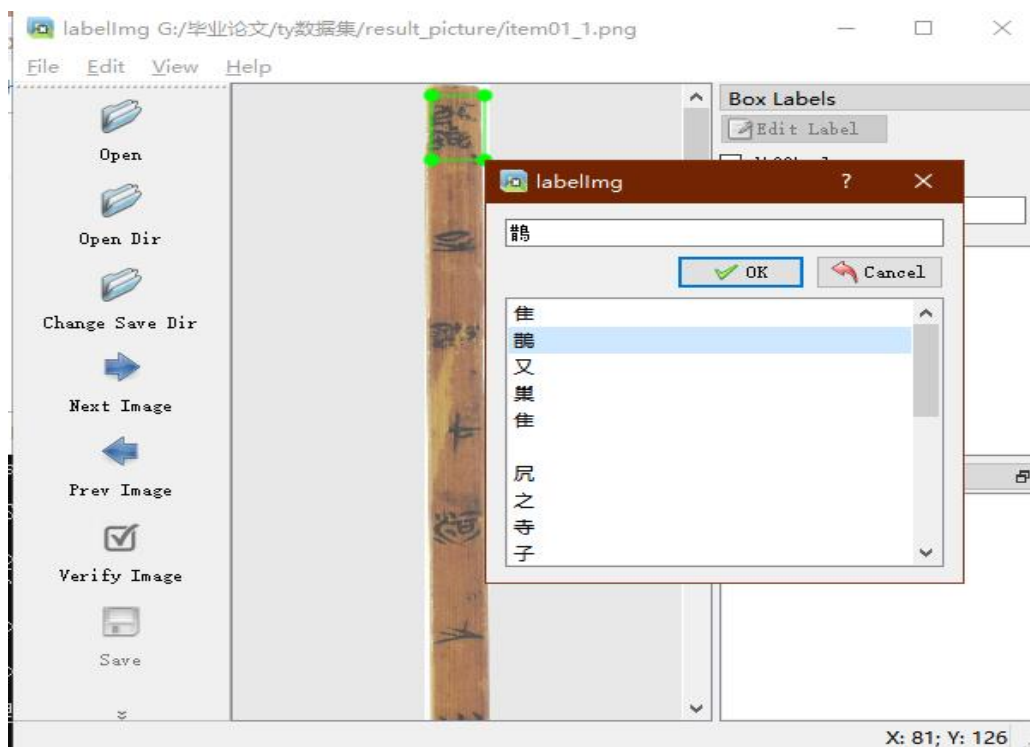


图 3-8 标注软件

### 3.2.2 文本识别数据集

标注任务结束后，根据标注结果将图片裁剪为一个个的单个字体图像，并将相同文字的图片归纳放在同一个文件夹内。筛选去除字体模糊或字体残缺等不能用字体图像，剩余字数 465 个。从标注结果中筛选适合小样本学习的单类别数制作成为战国简文字数据集（Bambooslips）。因为竹简图片内容对应的是诗经内容，无法保证每个字体出现的次数是一样的，因此部分汉字只出现一两次，部分却出现十次以上，这样的数据量正好符合小样本学习的数据要求。

但 Bambooslips 数据集存在数据类别不平衡问题，为了处理该问题，本文使

用了数据增广方法。查看对比 Omniglot 和 MiniImage 等小样本学习常用数据集发现每一个类别均包含 20 个样例左右。因此针对数据量在小于 20 的类别，本文将其通过翻转、旋转、缩放、高斯噪声、适当腐蚀、膨胀等数据增强方法将其数据量扩充为 20。与未通过数据增广前的数据集正好形成对比。

通过每个类别样本数出现次数及其数据增强后的类别，将竹简数据划分为不同的数据集：

- (1) 样本量 < 5;
- (2) 样本量在 5-9 之间;
- (3) 样本量  $\geq 10$ ;
- (4) 样本量在 5-9 之间的类别将单类别样本量数据增广为 20;
- (5) 样本量 > 10 的类别将单类别样本量数据增广为 20。
- (6) 将全部数据类别的样本量数据增广为 20。

以上数据集命名形式如表 3-1 所示。

表 3-1 数据集分类

(1)	(2)	(3)	(4)	(5)	(6)
Bambooslips1	Bambooslips2	Bambooslips3	Bambooslips2_ expanded	Bambooslips3_ expanded	Bambooslips4
[1-4]	[5-9]	[10, ]	20 each	20 each	20 each

因每个数据集类别个数都偏少，较少的数据集容易发生过拟合，所以考虑将上述几例数据集进行整合组成新的数据集。

**Bambooslips23:** 将第 (2) 类与第 (3) 类数据整合在一起组成，形成单类别样本量 > 5 的数据集；

**Bambooslips23\_expanded:** 将第 (4) 类与第 (5) 类数据整合在一起组成，形成样本量 > 5 的类别经由数据增广方法将数据量扩充为 20 的数据集。

上述八个数据集详细信息展示如表 3-2 所示。其中类别数代表该数据集共有多少类；样本总数表示该数据集总共的样本量；样本量最小类数代表数据集中样本数目达到最小值的类别总数；样本量最大类数代表数据集中样本数目达到最大值的类别总数；类间平均数表示该数据集中单个类的平均样本数目。

表 3-2 数据集详细信息

数据集	类别数	单类样本量	样本总数	样本量最小类数	样本量最大类数	平均每类样本量
Bambooslips1	332	1-4	634	163	41	1.91
Bambooslips2	81	5-9	545	22	11	6.73
Bambooslips2_expanded	81	20	1620	81	81	20
Bambooslips3	52	$\geq 10$	1331	8	20	25.60
Bambooslips3_expanded	52	20	1040	52	52	20
Bambooslips23	133	$\geq 1$	1876	163	20	14.11
Bambooslips23_expanded	133	20	2660	133	133	20
Bambooslips4	465	20	9300	465	465	20

从表中可以看出，数据集存在样本量不均衡的问题。在原数据上，Bambooslips3 是单类样本量最多的数据集；Bambooslips1 是单类样本量最少的数据集。为了探讨数据集中样本的其他相关特性，图 3-9 展示了从数据集中随机抽取的部分样本示例。

从图 3-9 可以明显看出，样本存在很多噪声及字迹不清晰等一系列问题，其造成该情况的原因已在 3.1.2 中做了详细的阐述。现将样本的相关特征介绍如下：

(1) 形态不一：因竹简上文字是用毛笔书写上去的，属于手写字的一种，即使同一个字，其字形也是各不相同，甚至有很大的差别。

(2) 数据量不均衡：因样本是从诗经中截取下来的，所以导致每一样本的数量不同，最少数量为 1，最多数据量为 53。

(3) 字体残缺：上图中“不”对应的图片 3、5、6 字体就残缺严重，甚至都看不清字体。如“息”字图片 3 所示，左半部分缺失，剩余部分反而会误认为市“李”字。

(4) 字迹清晰度不一致：如“才”图片中每张图片的清晰度都不一样。图片 8 就比较模糊，隐约才能看清字体。

(5) 其他噪声影响：竹简本身的纹路，或者因毛笔书写过程中笔墨的滴溅，

或者笔画与笔画之间的连接都对文字的识别带来一定的困难。

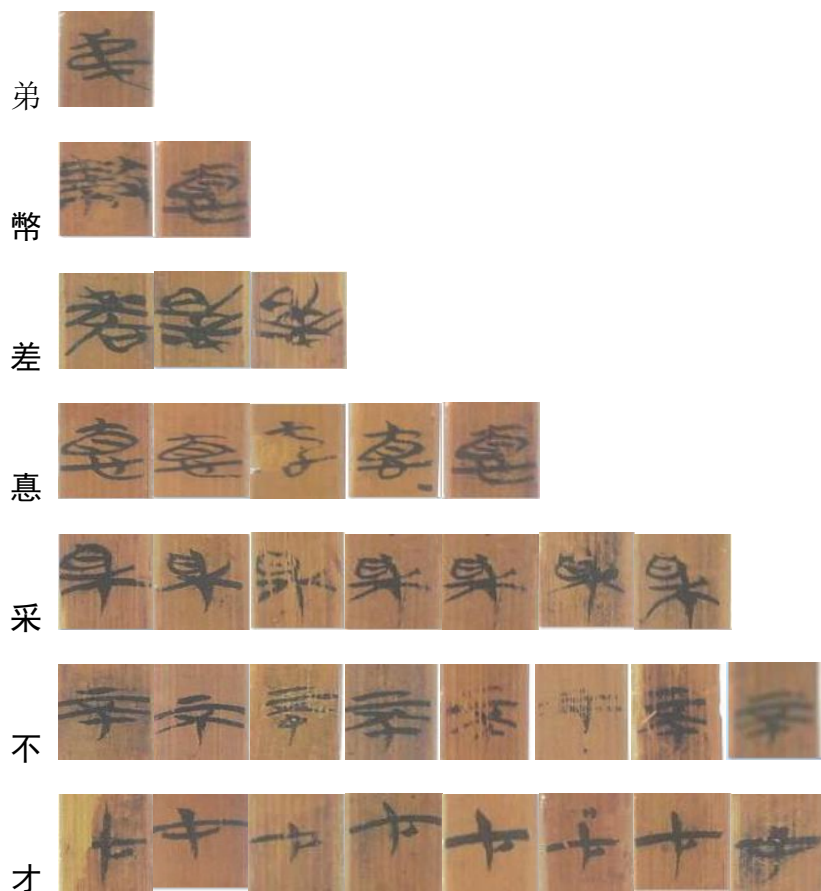


图 3-9 样本示例

### 3.3 本章小结

本章首先对 Bambooslips 数据集提出的背景做了详细的介绍，即为什么要提出这样的一个数据集：其中包括战国简目前研究的现状，以及造成竹简文字识别困难的原因。其次介绍了在数据集整理过程中遇到的困难及解决方法，标注过程所使用到的工具及最后整理出来的文字识别数据集分类和数据样例。

## 第 4 章 战国简文字识别算法

对于传统的文字识别算法，常用的模型是卷积神经网络。因为 CNN 具有一些自主学习等良好的性能，对于图像处理有很好的效果，在某些方面甚至能够超越人类能力。然而，CNN 一般需要海量的训练数据，因为要通过学习数据的特征，来获得合适的模型参数。所以在小样本情况下，CNN 的网络模型并不是很理想。战国简文字数据集中各个类别的图像数量都偏少，此时再使用 CNN 就不会得到较好的识别效果。因此，本章将卷积孪生网络应用于小样本学习中，用来进行战国简文字识别。在单个类别文字的样本量偏少的情况下，依然可以获得良好的识别准确率。

### 4.1 基于孪生网络的小样本战国简文字识别算法

面向于提高战国简文字识别准确率，本节基于孪生网络的框架，设计一个简单的网络模型 SiameseWord 与基于深度残差网络 Resnet50 的孪生网络 Siamese net (Resnet50)。同时，为了缓解单类别样本量不足缺陷，提出了数据生成算法，扩大了竹简数据集的样本量。最后，对算法所用的损失函数进行了介绍。

#### 4.1.1 SiameseWord 文字识别算法

##### (1) 神经网络结构与参数设置

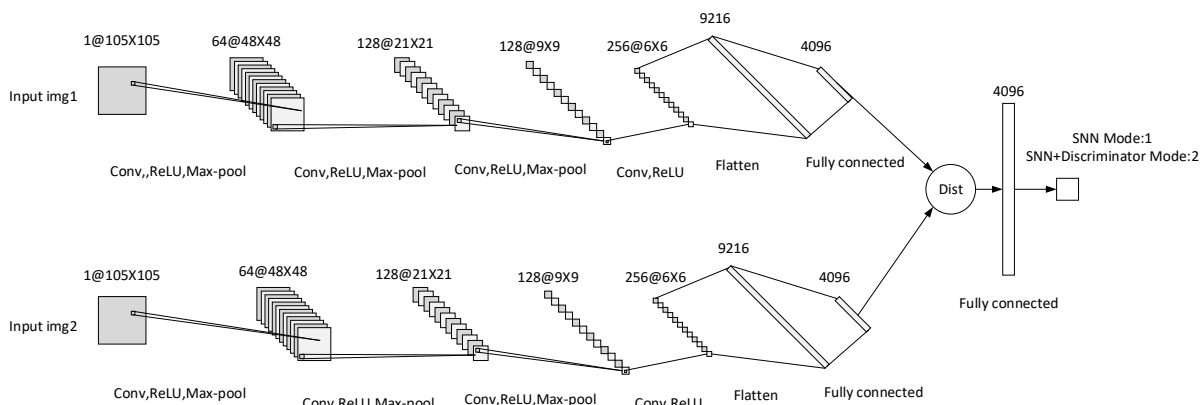


图 4-1 SiameseWord 网络模型

战国简文字识别所使用的网络模型是孪生网络模型结构，模型结构如图 4-1 所示。输入的数据经过模型后的输出结果为一列 4096 维的特征向量。图像输入尺寸为 105\*105，输入的每一对图像都会对应一个标签。标签为 0 代表图像对字体是同一字体，标签为 1 表示图像对字体不是同一个字体。该网络结构中，每一个卷积层之后是 ReLu 激活函数，池化操作采用最大池化。最后 1 两层是全连接层。网络的最终输出为一列 4096 的向量。网络模型的具体参数如表 4-1 所示。

表 4-1 SiameseWord 网络参数

卷积层	卷积核	步长	填充	输入	输出
Conv1	10*10	1	1	105*105	96*96*64
Pooling	2*2	2	0	96*96*64	48*48*64
Conv2	7*7	1	0	48*48*64	42*42*128
Pooling	2*2	2	0	42*42*128	21*21*128
Conv3	4*4	1	0	21*21*128	18*18*128
Pooling	2*2	2	0	18*18*128	9*9*128
Conv4	4*4	1	0	9*9*128	6*6*256
FC				6*6*256	4096

## (2) 距离计算

最近流行的深度度量学习使用最广泛的损失函数是对比损失和三元组损失，两者都对特征施加了欧几里得距离。在很多研究问题时常常需要估算不同样本之间的相似性度量，采用的方法就是计算样本间“距离”。欧氏距离是最常见的距离度量，衡量的是多维空间中各个点之间的绝对距离，能够体现个体数值特征的绝对差异，所以一般用于需要从维度的数值大小中体现差异的分析。

本文使用欧式距离来计算两个网络输出结果的相似性来判断输入的两个样本是否为同一类。欧式距离定义如式 (4.1) 所示。

$$D = \sqrt{\{G_w(X_1) - G_w(X_2)\}^2} \quad (4.1)$$

其中， $G_w(X_1) = (x_1^{(1)}, x_1^{(2)}, \dots, x_1^{(4096)})$ ， $G_w(X_2) = (x_2^{(1)}, x_2^{(2)}, \dots, x_2^{(4096)})$ 。因此，欧式距离公式也可以简化为如 (4.2) 所示公式。

$$D = \sqrt{\sum_{i=1}^{4096} (x_1^{(i)} - x_2^{(i)})^2} \quad (4.2)$$

一般情况，都会设定一个阈值  $m$ ，当  $D < m$  时，表示输入的图像字体对是同一个字；当  $D > m$  时，表示输入的图像字体对不是同一个字。

### (3) 损失函数

损失函数 (loss function) 是一种评估模型中真实与预测值差异的函数。常用的函数有 log 对数损失函数、指数损失函数、对比损失函数、Hinge 损失函数、平方损失函数等。在特征空间中使用对比损失函数可以对样本数据进行降维，相同的样本经过降维处理后其结果是相似的，但是不同的样本经过降维处理后结果具有一定的差异性。对比损失函数在样本匹配方面具有较好的判别能力，因此在孪生网络结构中该函数的使用最为广泛。

第  $i$  个样本对  $(x_1^{(i)}, x_2^{(i)})$  的对比损失值定义公式如式(4.3)所示。

$$H^{(i)} = f^{(i)} * E_w^{(i)} + (1 - f^{(i)}) * (m - E_w^{(i)}) \quad (4.3)$$

其中， $E_w^{(i)}$  表示第  $i$  个样本对  $(x_1^{(i)}, x_2^{(i)})$  最终输出特征的欧式距离。公式如图(4.4)所示。 $m$  表示边界值。 $f^{(i)} \in \{0,1\}$  表示第  $i$  个样本对的标签，当  $f^{(i)}=0$  时，表明该样本对不是同一个文字的图像，此时对比损失值  $H^{(i)} = m - E_w^{(i)}$ ；当  $f^{(i)}=1$  时，表示此时的样本对是同一个文字的图像，此时对比损失值  $H^{(i)} = E_w^{(i)}$ 。 $H^{(i)}$  越小，表示模型参数越合理。

$$E_w^{(i)} = \|G_w(X_1^{(i)}) - G_w(X_2^{(i)})\|^2 \quad (4.4)$$

其中， $G_w(X_1^{(i)})$  与  $G_w(X_2^{(i)})$  是输入图像对经由网络映射后输出的特征对。

为了对数据更有效的训练，对于输入数据，卷积神经网络一般使用最小批处理技术 (mini-batch) 将其分批处理。因此，损失函数公式的最终结果如下所示。

$$L(W) = \frac{1}{mb} \sum_{i=1}^{mb} H^{(i)} \quad (4.5)$$

其中， $mb$  表示每批次处理的样本数量。

### 4.1.2 基于 Resnet50 网络的 SiameseWord 算法

为了更好的进行对比，本文考虑使用较为深层次的模型——Resnet50<sup>[63]</sup>用于本文的孪生网络结构中。Resnet 已经被广泛运用于各种特征提取应用中，当深度学习网络层数越深时，理论上表达能力会更强。但是卷积神经网络达到一定的深度后，再加深，分类性能不会提高，反而会导致网络收敛缓慢，准确率也会降低。残差网络可以很好的解决这一问题。通常关于 Resnet50 的应用，一般使用 Resnet 官网提供的预训练模型。因为预训练模型及参数都是通过从大量数据训练得来的，可以加快训练，使得模型收敛更快，达到更好的效果。但为了与 4.1.1 节形成对比及 Bambooslips 数据集中的文字竹简图片的独特性，本文未使用官网提供的预训练模型，而使用 Resnet50 网络结构在 Bambooslips 数据集上从头训练。基于 Resnet50 的孪生网络模型如图 4-2 所示。

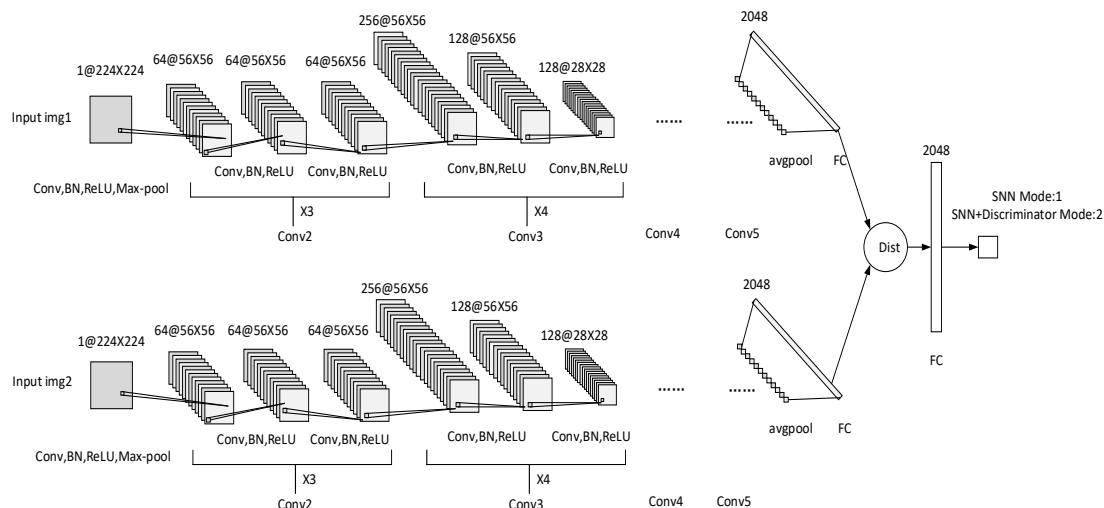


图 4-2 基于 Resnet 的孪生网络模型

图 4-2 所示模型中，一般 ResNet50 常用于 RGB（三通道）彩色图像的分类任务，如在 ImageNet 的训练；而在单通道图像的训练和测试较少。本文修改网络模型参数和数据读取，将 ResNet50 修改为适合本文数据集的单通道图像训练。输入的图片大小为 224\*224，先经过第一个卷积层，卷积核为 7\*7，填充为 3，步长为 2。接着包含 4 个残差块，在第一个残差块的第一层卷积层中，先通过 1\*1 的卷积对特征图进行降维，做一次 3\*3 的卷积操作，最后再通过 1\*1 卷积恢复维度。特征提取后是 BN 和 ReLU 函数；采用最大池化。经过后面的各个 Block，采用平均池化操作，用 softmax 实现回归（后面的 Block 结构类似，在此不再进行一

一叙述)。具体参数如表 4-2 所示。

表 4-2 ResNet50 网络参数

卷积层	卷积核	步长	填充	输入	输出
Conv1	7×7	2	3	224×224×1	112×112×64
maxpool	3×3	2	0	112×112×64	56×56×64
Conv2_x	$\begin{bmatrix} 1 \times 1, 1, 0 \\ 3 \times 3, 1, 1 \\ 1 \times 1, 1, 0 \end{bmatrix}$	×3		56×56×64	56×56×64
				56×56×64	56×56×64
				56×56×64	56×56×256
Conv3_x	$\begin{bmatrix} 1 \times 1, 1, 0 \\ 3 \times 3, 2, 1 \\ 1 \times 1, 1, 0 \end{bmatrix}$	×4		56×56×256	56×56×128
				56×56×128	28×28×128
				28×28×128	28×28×512
Conv4_x	$\begin{bmatrix} 1 \times 1, 1, 0 \\ 3 \times 3, 2, 1 \\ 1 \times 1, 1, 0 \end{bmatrix}$	×6		28×28×512	28×28×256
				28×28×256	28×28×256
				28×28×256	14×14×1024
Conv5	$\begin{bmatrix} 1 \times 1, 1, 0 \\ 3 \times 3, 2, 1 \\ 1 \times 1, 1, 0 \end{bmatrix}$	×3		14×14×1024	14×14×512
				14×14×512	14×14×512
				14×14×512	7×7×2048
avgpool	7×7	1	0	7×7×2048	1×1×2048
FC				1×1×2048	2048

注：表中[]中数据列数分别代表卷积核大小、步长、填充值。需要注意的是，Conv2\_2 的输入为 (56,56,256)，stride=1，后面都相同；Conv3\_2 的输入为 (28,28,512)，stride=1，其余结构都相同；同样的，Conv4\_2 的输入为(14,14,1024)，stride=1；Conv5\_2 的输入为 (14,14,1024)，stride=1；除此之外其余部分与各模块第一层结构都相同。

使用与 4.1.1 相同的距离函数与损失函数，在此不再进行赘述。

## 4.2 基于其它深度度量学习算法的战国简文字识别

度量学习(metric learning)研究如何在一个特定的任务上学习一个距离函数，使得该距离函数能够帮助近邻的算法取得较好的性能。深度度量学习 (deep

metric learning) 是度量学习的一种方法, 它的目标是学习一个从原始特征到低维稠密的向量空间的映射, 使得同类对象在嵌入空间上计算的距离较近, 而不同类的对象之间的距离较远。目前, 现有深度度量学习方法在车型识别数据集 Cars196、人脸数据集 LFW<sup>[65]</sup>等都取得了不错的效果。本节旨在探究小样本 Bambooslips 数据集在其他深度度量学习的方法上能否取得较好的识别效果。为了与 4.1 节中的算法形成对比, 本节基础网络仍使用 resnet50。

### 4.2.1 基于 SoftTriple 的战国简文字识别算法

SoftTriple 方法主要是对损失函数的改进。SoftTriple Loss 与 BN\_Inception 网络可以结合用于小样图片的识别。BN\_Inception 是以 GoogLeNet Inception V1 为基础, 引入了批次归一化 (Batch Normalization) 与将较大的二维卷积拆成几个较小维度卷积的做法, 使得梯度不会随着数据的变化而发生太大的变化, 不用担心因学习率太大而导致的模型发散问题, 并且在处理更多、更丰富的空间特征以及增加特征多样性等方面拆分的效果更好, 可以减少参数量, 减少计算过程。为了与 4.1 节中基于孪生网络的小样本战国简识别算法形成对比, 本节将 BN\_Inception 网络替换为 Resnet50 网络作为算法的基础网络。将 SoftTriple Loss 与 Resnet50 网络进行结合, 用于本文竹简数据集的识别。算法流程图如图 4-3 所示

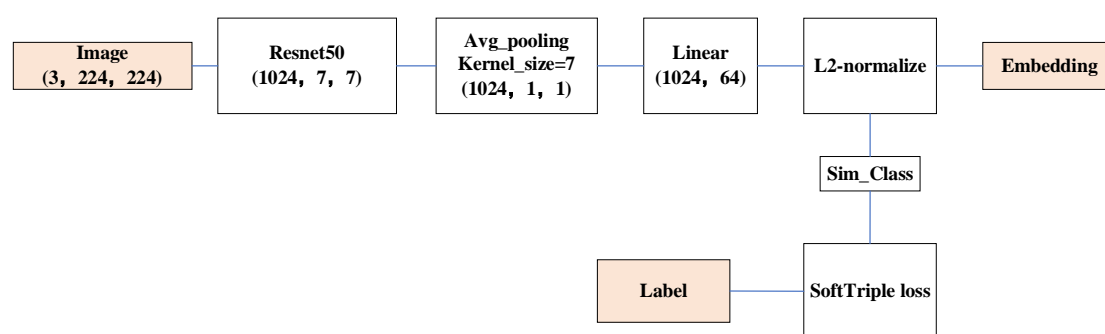


图 4-3 SoftTriple 网络结构

默认输入的图片大小是 224\*224, 首先将输入的图片 feed 进 Resnet50 网络结构中获得 7\*7\*1024 的输出特征向量; 再使用一个步长为 1, 池化窗口为 7\*7 的均值池化获得输出为 1024 的特征向量; 提高 FC 层的维度, 使得每个类别可以包含更多的中心; 应用 L2 归一化获取中心集合, 自适应的选取每个类别的中心

数目；最后通过相似度计算公式获取每个类别的相似度，最后计算 softmax loss。

损失函数使用了 2.4.1 节所介绍的 SoftTriple Loss 损失函数，在此不再进行赘述。

## 4.2.2 基于 CosFace 的战国简文字识别算法

CosFace 是在 SphereFace<sup>[66]</sup>基础上改进了对特征向量归一化和增添了余弦边界，通过归一化和余弦决策边界的最大化实现类间差异的最大化和类内差异的最小化。本文尝试将 CosFace 方法用于战国简文字的识别，判断其能否在本文竹简数据集上取得不错的识别效果。

文字识别的步骤：

- (1) 找文字：在图片中将文字框出来
- (2) 对齐字体：将文字笔画标出来，以此作为依据对齐文字
- (3) 识别：将对齐的字体进行识别，判断这个字是谁的

CosFace 主要用于第 (3) 步，解决如何能准确的识别这个文字是属于那个字的，其实也是一个分类问题。为了与上述算法形成对比，网络结构使用 Resnet50，损失函数使用 2.4.2 节中提及的 Large Margin Cosine Loss (LMCL) 损失函数。优化函数为 SGD，批次大小为 64，权重衰减取值 0.0005。在训练阶段，LMCL 引导 CNN 使用不同类别间的大边界来学习判别性的文字特征。在测试阶段，利用提取出的特征进行余弦相似度计算，得到角度度量，再进行验证与识别。

## 4.3 实验设计

在本节中展示算实验设计的相关细节。包括算法运行的软硬件环境配置、算法的性能度量、实验运行前期的数据准备工作以及影响实验结果的一些参数设置。

### 4.3.1 软硬件配置

硬件配置：

系统：Windows10 的 64 位操作系统

处理器：Intel(R) Core(TM) i5-1035G1 CPU @ 1.00GHz 1.19GHz

安装内存：16.0GB

显卡：1080Ti

软件配置：

Python 版本：python3.6

### 4.3.2 评价指标

为了评估机器学习算法在某任务中好坏，一般需要设计方法去度量性能。常用的度量方式有准确率、错误率、精确率、召回率等。为了更好的介绍这些度量方式，以二分类问题为例子，将样例根据其真实类别和学习期的预测类别组合为如下四种情形：

表 4-3 预测类别组合

真实情况	预测结果	
	正例	反例
正例	TP（真正例）	FN（假反例）
反例	FP（假正例）	TN（真反例）

其中：

TP：被模型预测为正的正样本；

FP：被模型预测为正的负样本；

FN：被模型预测为负的正样本；

TN：被模型预测为负的负样本。

准确率（Accuracy）：表示正确分类的测试实例的个数占测试实例总数的比例，反应了分类器对整个样本的判定能力。计算公式为：

$$Acc = \frac{TP + TN}{TP + FN + FP + TN} \quad (4.6)$$

错误率（Error Rate）：表示错误分类的测试实例的个数占测试实例总数的比例。

计算公式为：

$$ErrorRate = \frac{FP + FN}{TP + FN + FP + TN} \quad (4.7)$$

精确率（Precision）：也称为查准率，表示正确分类的正例个数占正例的实例个数的比例，计算公式为：

$$\text{Precision} = \frac{TP}{TP + FP} \quad (4.8)$$

召回率(Recall):也叫查全率,表示正确分类的正例个数占实际正例个数的比例,计算公式为:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (4.9)$$

不同的实际应用场景需要根据所需完成的任务选择性能度量方式。任务场景不同,度量手段也不同,如果性能度量选取不当,对于该任务来说可能会带来很很差的效果。考虑到本文算法的输入是一组组的图像对及其对应标签,通过对每组图片的相似性度量比较,预测其标签值。若预测标签与实际标签相等,则算法判断正确;反之,判断错误。因此选取准确率作为本文唯一的评价标准。公式表示如下:

$$\text{Acc} = \frac{\text{sum}(\text{predict\_label} == \text{test\_label})}{N} \quad (4.10)$$

其中, Acc 表示准确率, N 表示测试样本的总数。

### 4.3.3 数据划分

在使用数据集训练模型之前,一般需要将整个数据集分为训练集、验证集、测试集。训练集用来训练模型;验证集使用交叉验证来挑选最优的模型;测试集用来评估模型的性能。本文数据划分在 3.2.2 节中数据集分类的基础上,将数据集首先按单个类别样本量 8:2 的比例分为训练集(train)与测试集(test)。对于原样本量大于 5 的类别,可以很容易划分为 train 与 test。对于原样本量小于 5 的类别,按如下规则划分:

当原样本量=4, train/3 张, test/1 张;

当原样本量=3, train/2 张, test/1 张;

当原样本量=2, train/1 张, test/1 张;

当原样本量=1, test/1 张。

验证集是在训练集的基础上继续按 8:2 的比例进行划分,每迭代 1000 次用验证集测评一次。如果验证集准确率在 10000 迭代内没有提高就会停止训练。输出最佳验证集准确率。

#### 4.3.4 数据准备

SiameseWord 基本上遵循了孪生网络的算法思想：将不同的两张图片同时输入孪生神经网络模型中，输出的结果是两者之间的距离，我们用这个距离来衡量两者之间的相似性度量。若要计算两者之间的距离，A 到 B 与 B 到 A 的距离应该是一样的，所以要用到共享权值来控制距离的一致性。即使输入的 A、B 进行调换，距离不会发生改变。

利用少量的训练数据得到一个识别率较高的模型，现有的文字识别算法还是很难实现的。目前公开的文字样本库，大多是基于海量数据的，不符合小样本学习的样本量需求。故本文使用了小样本学习常用的数据集 Omniglot 以及本文自己制作的数据集 Bambooslips 进行对比分析。

(1) Omniglot 数据集<sup>[64]</sup>：包含来自 50 种不同语言的 1623 个不同手写字符，每个字符由 20 个不同人的在线绘制组成。选取其中 30 种语言作为训练集，剩下的 20 种作为验证集。每张图片的像素是 105\*105。且背景透明、字迹清晰、无任何干扰信息。部分字符图像如图 4.4 所示。

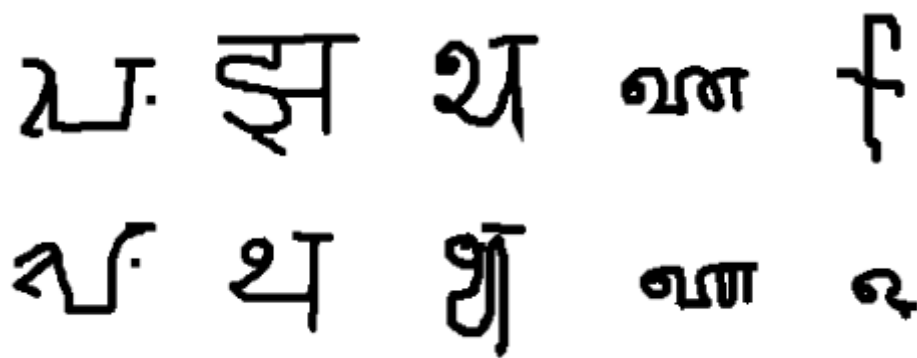


图 4-4 Omniglot 数据集部分字符图像

(2) Bambooslips 数据集：Bambooslips 数据集是本文自己制作的战国简文字数据集。该数据集由 465 个古竹简文字组成，原数据图片共约 2600 张，每张图片具有其唯一的标号与序号，大小为 105\*105，均为彩色图像。背景干扰纹路较多，部分字体存在残缺或字迹不清等情况。数据集中部分图像如图 4-5 所示。



图 4-5 Bambooslips 数据集部分字符图像

孪生网络模型的输入由两张图片与一个标签组成，因此在数据处理部分，要整理生成符合模型输入的训练数据，将同一个字的文字图像对标记为 1，不同文字的图像对标记为 0，按 8: 2 的比例生成训练集与测试集。因为战国简数据集每个类别的样本量偏少，在模型训练阶段对获取一个最佳的网络模型会造成一定影响。针对该问题，结合孪生网络的思想，使用训练数据生成算法，扩大了竹简数据集的样本量。具体生成过程算法如表 4-4 所示。

表 4-4 数据训练配对过程算法流程

训练数据的生成算法
<p>(1) 随机抽取两张战国简文字图像 <math>A_1</math> 与 <math>A_2</math> (<math>A_1 \in S, A_2 \in S</math>)，<math>S</math> 为竹简文字数据集；</p> <p>(2) 若 <math>A_1</math> 与 <math>A_2</math> 同一张图像，则继续 (1)；</p> <p>(3) 若 <math>A_1</math> 与 <math>A_2</math> 是同一个文字的不同图像，将标签记为 1；否则，标签记为 0，构成训练样本 <math>T</math>；</p> <p>(4) 记构造的训练样本集为 <math>S'</math>，若 <math>T \notin S'</math>，则将 <math>T</math> 加入到 <math>S'</math>；若 <math>T \in S'</math>，则继续步骤 (1)；</p> <p>(5) <math>S'</math> 的训练数据样本对达到设定值，结束。</p>

根据表 4-4 的训练数据生成流程，图 4-6 给出了对应的匹配对或不匹配对的结果。其中 A、B、C、D 表示不同文字的图片，生成的样本对是(A,B,0)、(A,C,0)、(A,D,0)；A、 $A_1$ 、 $A_2$ 、 $A_3$  表示同一个文字的图像，但其字体形态、清晰度、字体完整度等存在差异，生成的样本对是(A, $A_1$ ,1)、(A, $A_2$ ,1)、(A, $A_3$ ,1)。

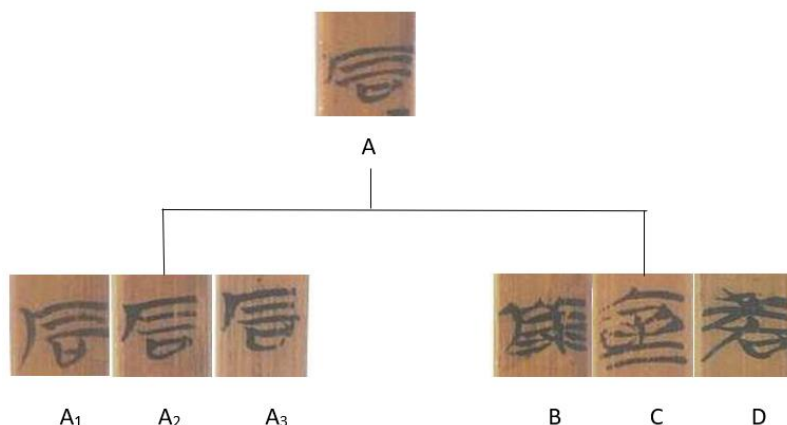


图 4-6 匹配对与不匹配对图像示例

### 4.3.5 实验参数

siameseWord 与 siameseWord (Resnet50) 算法中有多个参数，但大部分参数都可以在初始化的之后随着算法自动调整。批次大小 (batch\_size)、每次构建分类任务时从数据集抽取的类数 (N-way)、迭代次数 (max\_iter) 需要自己手动调试，找到最佳。具体设置如下：

batch\_size = 64

N-way = 20

在 siameseWord 中，因网络结构简单，数据量偏少，太多的训练迭代次数会造成算法的过拟合。经过多次对比尝试，选取迭代次数为 20000 时效果最佳。

siameseWord (Resnet50) 算法因网络结构复杂度增长，迭代次数取值为 60000 时性能最佳。

## 4.4 实验结果分析

### 4.4.1 小样本学习方法对比分析

为了证明基于孪生网络小样本学习算法的优越性与竹简数据集的适用性，本文在实验中采用多个算法进行分析。在 Bambooslips23\_expanded、Bambooslips4 数据集与 Omniglot 数据集上对比了现有的小样本学习算法，如不可知模型元学习、原型网络、匹配网络等算法的识别准确率，结果如表 4-5 所示。

表 4-5 Omniglot 与 Bambooslips23\_expanded 数据集上的实验结果

数据集 \ 算法	Omniglot	Bambooslips23_expanded	Bambooslips4
MAML <sup>[56]</sup>	95.8	46.25	60.62
Prototypical networks <sup>[57]</sup>	<b>99.0</b>	33.17	37.71
1-Nearest Neighbor	21.7	19.7	21.53
Matching nets <sup>[58]</sup>	93.8	37.26	52.64
SiameseWord	94.0	73.34	75.1
SiameseWord (Resnet50)	95.1	<b>75.76</b>	<b>76.77</b>

从上表可以看出，在 Omniglot 数据集上，SiameseWord 与 SiameseWord (Resnet50) 的识别准确率不是最高；但是在 Bambooslips 数据集上，SiameseWord 与 SiameseWord (Resnet50) 的识别效果优于其他算法。SiameseWord (Resnet50) 在 Bambooslips23\_expanded 数据集上比 SiameseWord 提高了 2.4% 左右，在 Bambooslips4 数据集上比 SiameseWord 提高了 1.6% 左右。

除此之外，本节在基于 SiameseWord 模型的孪生网络算法与基于 Resnet50 模型的孪生网络算法的基础上对 Bambooslips 数据集不同分类数据、原数据、增强数据分别进行实验，获得识别准确率并取平均值。结果如表 4-6 所示。

表 4-6 Bambooslips 各个数据集的结果对比

数据集 \ 算法	SiameseWord	SiameseWord(Resnet50)
Bambooslips2	70.23	68.45
Bambooslips2_expanded	72.59	64.32
Bambooslips3	71.15	66.15
Bambooslips3_expanded	73.17	68.8
Bambooslips23_expanded	73.34	75.76
Bambooslips4	75.1	76.77

如表 4-6 所示，通过反复尝试，找到模型训练的最佳迭代次数，防止训练太多反而造成算法的过拟合。从表中可以看出，数据量越多，训练效果越好。结果显示，数据类别数在 133 以下的数据集在算法 SiameseWord 的识别效果较

好；但数据类别数为 133 与 465 时，算法 SiameseWord (Resnet50) 的训练效果最佳。因为对于网络结构复杂的算法，参数太多，因样本数量不够支撑，可能结果并非最好。且算法在经由数据增广后的数据集上的识别效果比在原数据集的识别效果大多提高 2% 左右。

#### 4.4.2 单类文字识别效果

上一节所示结果均为不同方法在整体数据上的准确率，对于单个类别的识别准确率未有显示。本文针对 Bambooslips 数据集不同的类，将每个类别的识别准确率单独输出，作为一个更直观的结果显示。在训练阶段，验证集的单类识别率每迭代 1000 次输出一次结果，测试集单类准确率直接输出。训练过程中单类识别准确率的变化如图 4-7 所示。

从图 4-7 中可以看出，单个类别的准确率随着迭代次数的增加在不断的提高，当迭代次数达到 18k 的时候，验证集中单类别的识别准确率慢慢的趋于平缓。由于每个类别中样本图片的清晰度与残缺度存在差异，单个类别的识别准确率也相应的会有一定的误差，但总体效果良好。

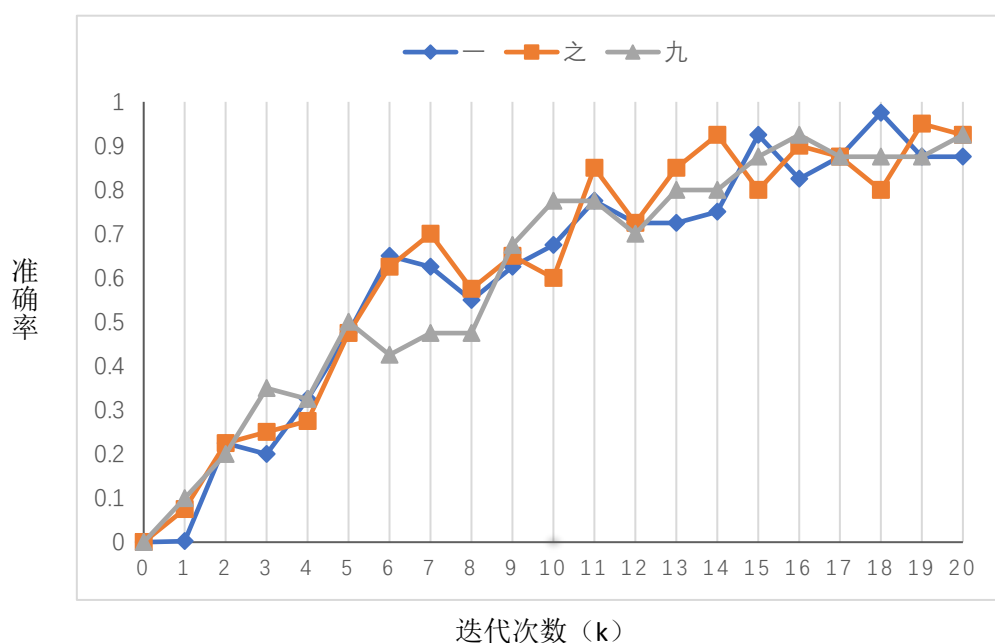


图 4-7(a) Bambooslips23\_expanded 数据集部分单类别训练识别准确率

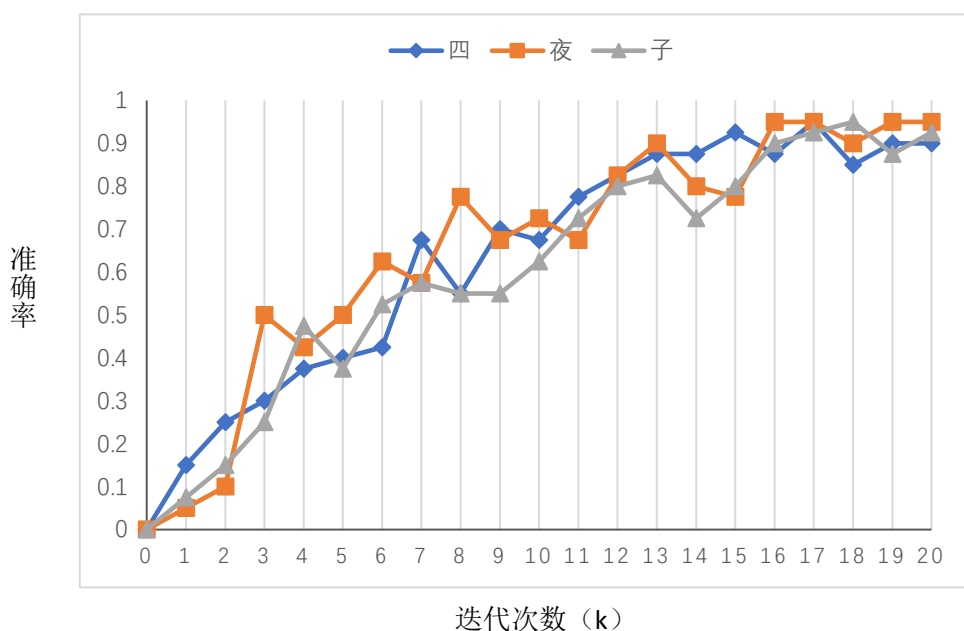


图 4-7(b) Bambooslips4 数据集部分单类别训练识别准确率

图中 k=1000。

### 4.4.3 度量学习方法对比分析

在 4.3.1 节中，主要对比了 Bambooslips 数据集在各个小样本学习算法上的识别性能。本节基于 SoftTriple 与 CosFace 两种深度度量学习方法，在竹简数据集上的识别效果进行了对比分析。结果如表 4-7、4-8 所示。

表 4-7 以 Resnet50 为基础的算法识别准确率

数据集 \ 算法	SoftTriple	CosFace
Bambooslis2	14.5	29.1
Bambooslis2_expanded	8.4	32.4
Bambooslis3	49.6	38.3
Bambooslis3_expanded	55.2	43.2
Bambooslis23_expanded	35.6	46.5
Bambooslis4	19.4	48.3

表 4-8 以 Inception 为基础的算法识别准确率

数据集 \ 算法	SoftTriple	CosFace
Bambooslis2	35.1	36.5
Bambooslis2_expanded	32.8	38.4
Bambooslis3	90.7	58.3
Bambooslis3_expanded	88.1	59.4
Bambooslis23_expanded	71.2	62.5
Bambooslis4	17.5	64.3

观察表中的数据，Bambooslis2 数据集不管是在原数据上的测试还是数据增强后的测试，识别准确率总是偏低，间接导致 Bambooslis23 数据集上的准确率低于 Bambooslis3 数据集上的 20%左右。证明单例样本量在 5~10 之间的数据集因字体残缺、字迹不清等原因存在影响最终识别效果的因素。且不管是在原算法上对竹简数据集的识别还是在修改网络结构为 resnet50 后对竹简数据集的识别，整体效果都不如本文提出的孪生网络算法，证明了本文算法的有效性。

孪生网络是一种端到端的度量学习方法。为了更好的对算法进行对比，现将本文提及的四个度量学习算法在 Bambooslis23\_expanded 的准确率、训练时间、测试时间、模型大小等进行分析，如表 4-8 所示。通过分别对比不同算法，我们发现，使用更复杂的网络训练出的模型虽然提高了战国简文字识别的准确率，但在空间与时间上的消耗也随之增长。并且从表中可以看出，使用孪生网络的算法会比 SoftTriple 与 CosFace 的识别准确率高，训练时间还短，证明了本文算法的有效性。

表 4-8 算法性能对比

算法 \ 性能	准确率	训练时间/ (小时)	测试时间 (FPS)	模型大小 (MB)
SiameseWord	73.34	2.56	3.56	56.2
SiameseWord(Resnet50)	75.76	7.68	4.01	94.3
SoftTriple	35.6	9.2	2.62	100
CosFace	42.1	7.8	3.22	114

## 4.5 本章小结

本文提出了基于孪生网络的战国简文字识别算法,采用图像对方式作为网络的输入,经过网络提取图像对中每张图片的特征向量,计算它们的相似性度量值,阐述了所使用的损失函数及数据训练的生成过程。同时,基于深度度量学习的方法对战国简数据集进行了识别性能方面的比较。通过实验结果的对比,证明基于孪生网络的小样本战国简文字识别算法的高效性以及竹简数据集的有效性与可用性。



## 第 5 章 战国简文字识别软件平台

为了让本文的算法有一个实际的应用,以及验证本文算法的实用性,本章设计实现了一个简单的战国简文字识别软件。大致包括战国简图片的导入、图片中文字图片的单独保存、战国简文字的识别三个部分。该系统实现了对导入到系统中的竹简图片进行文字标注后,将每个文字单独保存,与现有战国简数据集进行对比后,找出相似度最高的战国简文字图像,并输出显示其对应的汉字。

首先对该系统的开发环境与平台总体架构做了简要介绍,随后阐述了软件平台的各个模块,对各模块的功能做了详细介绍,并展示的相应的系统界面。最后,分析总结了整个系统平台的实际应用。

### 5.1 开发环境

本文中数据训练及算法都是基于 pytorch 1.0.1 框架,利用 python3.6 实现的。为了在软件平台中成功引用算法训练模型,因此该系统软件的界面设计与模块功能均是基于 pycharm+PyQt5+python 实现的。

### 5.2 平台总体架构

该系统是一个易于操作的战国简文字识别小软件。其主要功能就是对输入的战国简图片上的文字进行识别输出。但因为孪生网络输出的是图片对的相似性度量,因此需要准备两个数据集,一个存放待识别的竹简单字图片,由每条竹简的单字图片组成的文件夹构成;另一个数据集由现有竹简文字图像数据库的部分摘取组成,每个文字均保留一张字迹清晰的单字竹简图像。将待识别图片与现有图片一一组成输入对,找出相似性最高的图片,输出其图片对应的文件夹名称。该战国简文字识别系统的整体设计流程如图 5-1 所示。

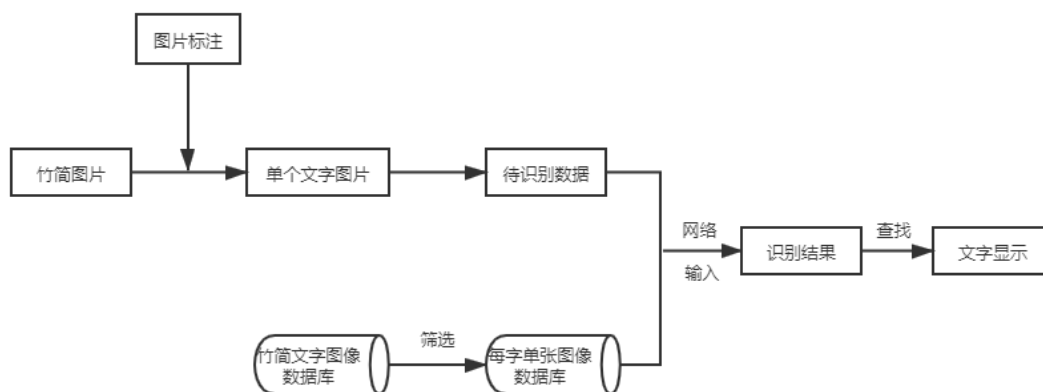


图 5-1 系统设计流程图

(1) 构建适用于软件平台所用的现有竹简文字单图片数据集。为了方便将已有图片与待识别图片组成输入对，在现有竹简文字图像数据库中，挑选每个文字图像字迹清晰的一张组成一个新的数据集。

(2) 导入待识别的竹筒图片，若导入为空，会有相应的提示；导入成功后，利用标注框对竹筒图片中的单个字体进行标记。

(3) 对于获取到的单张文字图像与现有竹简文字图像数据库中的图片组成输入对，利用训练好的模型完成相似性度量。

(4) 将待识别图片与已有图片的相似度进行排序，选择相似度最高的图片，输出其对应的文字。

### 5.3 软件平台模块的实现

前两节对战国简文字识别系统的开发环境与平台总体架构做了详细介绍，本章节将从系统的竹筒图片导入、单文字图片获取、文字识别三个模块进行详细介绍，并展示了其对应界面。

#### 1、竹筒图片导入模块

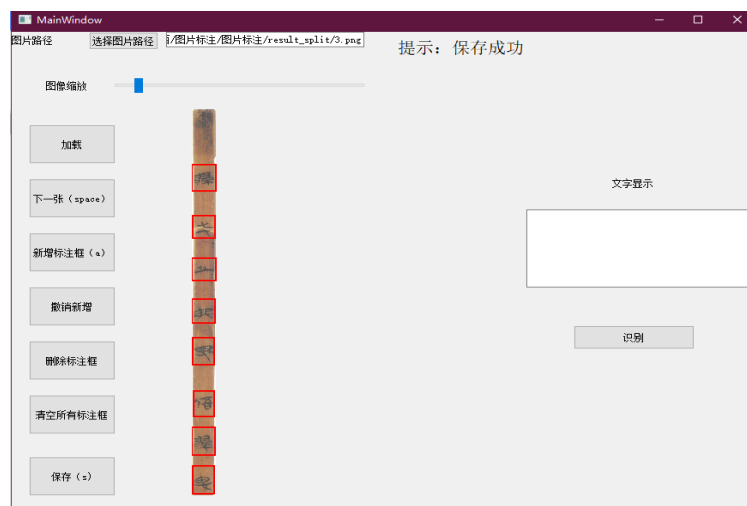
该系统的竹筒图片输入方式为直接从本机导入一张待识别的战国简图片，图片格式可以。若导入图片为空，会有相应的提示信息。界面如图 5-2 所示。



图 5-2 竹简图片导入界面

## 2、单文字图片获取模块

单文字图片获取模块主要用于得到单个文字图像，便于下一步处理。理想状态应该是通过文字检测算法自动识别定位文字图片，但因竹简图片噪声较大、竹简纹路众多、墨汁滴溅等因素造成文字检测定位会有众多干扰。故本模块由手工标注实现。将导入系统中的竹简图片根据大小可以进行缩放，使用四点坐标法将竹简中的每个文字进行标注，使用红色方框框出图片中需要识别的文字位置，错误的标注框可以删除重新标注。点击保存得到待处理的文字图片。将每条竹简对应的文字图片保存在一个文件夹内。界面如图 5-3 所示。



## 3、图 5-3 图片标注模块界面文字识别模块

文字识别模块需要有两个输入，一个是每条竹简对应的文字图像文件夹中的每个文字，一个是从现有竹简文字挑选出来的单字图片组成的文件夹。通过遍历，

将每张待识别图片与已有竹简单字图片，一一组成输入对输入到第四章已训练好的模型中进行相似性度量。按照相似度对图片进行排序，选出竹简文字数据库中与该输入文字图像相似度最高的图片；最后，输出该图片对应的文字。该模块流程图如图 5-4 所示。文字识别显示如图 5-5 所示。

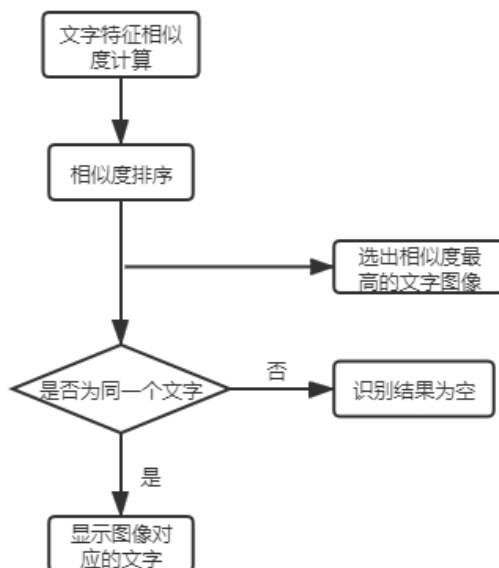


图 5-4 竹简文字识别流程图



图 5-5 文字显示界面

## 5.4 平台的应用

该软件平台通过导入待识别的竹简图片，快速的完成图片文字的对比识别，

输出竹简图片对应的战国文字。对于目前众多学者侧重的古文字形体结构与演变规律等的研究,本论文工作实现了战国简文字研究与计算机文字识别的交叉研究,对于非专业人士的战国简研究提供了便利。毕竟现在大多的竹简文字识别都依靠专业人士的识别判断,这个过程耗时耗力,大大的加重了研究过程的成本,该软件的使用,可以让非专业人士也可以快速的识别战国简文字。

## 5.5 本章小结

本章基于 `pycharm+PyQt5+python` 设计实现了战国简文字识别系统,可以快速、准确的满足竹简文字识别的需求,且也证明了本文竹简文字的小样本识别算法的可行性与实用性。



## 第6章 总结与展望

### 6.1 总结

古文字研究与计算机文字识别的交叉研究,目前已在生活中存在着广泛的需求。由于战国简文字图像的复杂性及较少的数据量导致现有的大多深度学习算法性能不佳。随着小样本学习的文符检测、人脸识别等算法的出现为战国简文字图像的识别开辟了新的方向。对于样本数量不足的研究,一般使用迁移学习。但方法众多,应用广泛,本文在此将基于度量学习的小样本学习算法应用于强噪声高残缺的战国简文字的识别。论文主要研究内容如下:

(1) 由于对出土的战国简的整理版本较少,导致目前暂未发现有公开的战国简文字数据集存在。因此,经过导师的指导,个人独立标注整理了一个战国简文字识别数据集。该数据集的详细介绍在第三章已进行阐述,在此不再赘述。

(2) 设计并实现了基于孪生网络的小样本学习算法,使其适用于战国简文字图像识别任务。同时为进一步提高算法的识别的精度,将基于简单 CNN 的孪生网络与基于 Resnet50 的孪生网络识别效果进行对比。随后,用多种小样本学习算法与本文算法在竹简数据集上的识别效果进行对比,证明本文算法对于战国简文字图像识别的精确性。最后,使用 SoftTriple 与 CosFace 两种方法去验证本文数据集的挑战性与可用性,经过实验结果的对比分析,进一步证明基于孪生网络的小样本战国简文字算法的有效性。

(3) 基于 pycharm+PyQt5+python 实现了战国简文字识别软件平台,将第四章的孪生网络算法模型导入,可以得出较好的文字识别显示结果,节省战国简文字识别的人力物力。

### 6.2 展望

本文将战国简文字识别的研究结果进行了对比分析,基于孪生网络的小样本战国简文字识别算法在战国简多个数据集上的识别效果都优于基于其他深度度

量学习算法。但战国简文字识别还存在一些待解决的问题。

(1) 孪生网络的小样本算法在噪声较大、干扰纹路较多、文字残缺的战国简文字识别准确率并非十分理想，且不同类别的单个类别识别准确率相差较大，仍需相关研究提高战国简文字数据集的识别准确率。

(2) 战国简文字识别软件应分为文字检测与文字识别两部分。但本文只实现了文字识别部分，对于文字检测部分因竹简图片噪声较大、竹简纹路众多、墨汁滴溅等因素造成文字检测定位会有众多干扰。故目前文字检测由手工标注实现，仍需相关研究实现竹简图片中文字的自动检测定位，美化完善战国简文字识别软件。

(3) 战国简文字数据集的类别数目相对较少，可能会对训练结果造成影响。虽然已通过数据增强方法将数据类别尽量扩充到了最大，但后续应该通过增加原数据手段，扩充战国简文字数据集的类别数和样本量，会获得更好的识别准确率与效果。

## 参考文献

- [1] Yaqing Wang, Quanming Yao. Few-shot Learning: A Survey[C].arXiv Preprint arXiv: 1904.05046, 2019.
- [2] Ziyang Wu, Yuwei Li, Lihua Guo, et al. PARN: Position-Aware Relation Networks for Few Shot Learning[C]. ICCV 2019: 6658-6666.
- [3] Oreshkin B N, Lacoste A, Rodriguez P. TADAM: Task dependent adaptive metric for improved few-shot learning[J]. NeurIPS 2018: 719-729.
- [4] Bodily P, Bay B, Dan V. Computational Creativity via Human-Level Concept Learning[C]. ICCV, 2017:57-64.
- [5] Lake B M, Ullman T D, Tenenbaum J B, et al. Building machines that learn and think like people[J]. Behavioral & Brain Sciences, 2016, 40: 1.
- [6] Porfiri M. Special Issue Editorial: Robotics: Mechanics and Control of Locomotion.[J]. Journal of Nonlinear Science, 2017, 27(4): 1089-1091.
- [7] Bonardi A, James S, Davison A J. Learning One-Shot Imitation From Humans Without Humans[J]. IEEE Robotics Autom. Lett, 2020, 5(2): 3533-3539.
- [8] Gao K, Liu B, Yu X, et al. Deep Relation Network for Hyperspectral Image Few-Shot Classification[J]. Remote Sens., 2020, 12(6): 923.
- [9] Gui L, Wang Y X, Hebert M. Few-Shot Hash Learning for Image Retrieval[C]. IEEE International Conference on Computer Vision Workshop (ICCVW), 2017:1228-1237.
- [10] Pfister T, Chaples J, Zisserman A. Domain-Adaptive Discriminative One-Shot Learning of Gestures[J]. ECCV, 2014, (6):814-829.
- [11] 许莹. Applications of FSL in the brain based on DTI [J]. 北京生物医学工程, 2017, 036(4): 420-426.
- [12] Fu Y, Wang C, FU Y, et al. Embodied One-Shot Video Recognition: Learning from Actions of a Virtual Embodied Agent[C]. ACM Multimedia, 2019: 411-419.
- [13] Rostami M, Kolouri S, Eaton E, et al. Deep Transfer Learning for Few-Shot SAR Image Classification[J]. Remote. Sens., 2019, 11(11): 1374.
- [14] Wen-Hsuan Chu, Yu-Jhe LI, Jing-Cheng Chang, et al. Spot and Learn: A Maximum-Entropy Patch Sampler for Few-Shot Image Classification[C].CVPR, 2019:6251-6260.
- [15] Bing Liu, Xuchu Yu, Anzhu Yu, Pengqiang, et al. Deep Few-Shot Learning for Hyperspectral Image Classification[J]. IEEE Trans. Geosci. Remote. Sens., 2019, 57(4): 2290-2304.

- [16] 郑欣悦. 基于深度学习的少样本图像分类方法[D].北京: 中国科学院大学, 2019.
- [17] Fei-Fei Li, Robert Fergus, Pietro Perona. A Bayesian Approach to Unsupervised One-Shot Learning of Object Categories[C].ICCV, 2003: 1134-1141.
- [18] Fei-Fei Li, Robert Fergus, Pietro Perona. One-shot learning of object categories[J] IEEE Trans.Pattern Anal. Intel., 2006, 28(4):594-611.
- [19] Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative Adversarial Networks[J]. NIPS, 2014, 3:2672-2680.
- [20] Santoro A, Bartunov S, Botvinick M, et al. One-shot Learning with Memory-Augmented Neural Networks[C]. Computer Research Repository, 2016.
- [21] Ravi S, Larochelle H. Optimization as a Model for Few-Shot Learning[C]. ICLR, 2017.
- [22] Finn C, Abbeel P, Levine S. Model-agnostic meta-learning for fast adaptation of deep networks[C]. Proceedings of the 34th International Conference on Machine Learning-Volume 70. JMLR. org, 2017: 1126-1135.
- [23] Yu M, Guo X, Yi J, et al. Diverse Few-Shot Text Classification with Multiple Metrics[C]. NAACL-HLT, 2018: 1206-1215.
- [24] Xu Han, Hao Zhu, Pengfei Yu, et al. FewRel: A Large-Scale Supervised Few-Shot Relation Classification Dataset with State-of-the-Art Evaluation[C]. EMNLP, 2018:4803-4809.
- [25] Gao T, Han X, Liu Z, et al. Hybrid Attention-Based Prototypical Networks for Noisy Few-Shot Relation Classification[C]. AAAI, 2019: 6407-6414.
- [26] 张亚倩. 卷积神经网络研究综述[J]. 信息通信, 2018, 191(11):32-34.
- [27] Leal-Taixe, Canton-Ferrer C, Schindler K. Learning by Tracking: Siamese CNN for Robust Target Association[C]. CVPR Workshops, 2016:418-425.
- [28] Fei-Fei Li, Robert Fergus, Pietro Perona. A Bayesian Approach to Unsupervised One-Shot Learning of Object Categories[C].ICCV, 2003: 1134-1141.
- [29] Lake B M, Salakhutdinov R, Gross J, et al. One shot learning of simple visual concepts[C].CogSci, 2011.
- [30] Motiian S, Jones Q, Iranmanesh S M, et al. Few-Shot Adversarial Domain Adaptation[J].NIPS, 2017: 6670-6680.
- [31] Li B, Wu W, Wang Q, et al. SiamRPN++: Evolution of Siamese Visual Tracking With Very Deep Networks[C]. CVPR, 2019: 4282-4291.
- [32] Hsiao S, Kao D, Liu Z, et al. Malware Image Classification Using One-Shot Learning with Siamese Networks[J]. Procedia Computer ence, 2019,159: 1863-1871.
- [33] 柳青林. 基于小样本学习的目标匹配研究[D]. 西安: 西安电子科技大学, 2018.
- [34] Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative Adversarial Networks[J].

- Advances in Neural Information Processing Systems, 2014, 3:2672-2680.
- [35] Han H, Wang W Y, Mao B H. Borderline-SMOTE: A New Over-Sampling Method in Imbalanced Data Sets Learning[C]. International Conference on Intelligent Computing, 2005:878-887.
- [36] MARIANI G, SCHEIDEGGER F, ISTRATE R, et al. BAGAN: Data Augmentation with Balancing GAN[C]. arXiv Preprint arXiv:1803.09655, 2018.
- [37] French G, Mackiewicz M, Fisher M. Self-ensembling for visual domain adaptation[C]. ICLR(Poster), 2017.
- [38] Andrychowicz M, Denil M, Gomez S, et al. Learning to learn by gradient descent by gradient descent[C]. NIPS, 2016:3981-3989.
- [39] Sung F, Zhang L, Xing T, et al. Learning to Learn: Meta-Critic Networks for Sample Efficient Learning[C]. arXiv Preprint arXiv:1706.09529 2017.
- [40] Hariharan B, Girshick R. Low-shot Visual Recognition by Shrinking and Hallucinating Features[C]. ICCV, 2017: 3037-3046.
- [41] Kimura A , Ghahramani Z , Takeuchi K , et al. Few-shot learning of neural networks from scratch by pseudo example optimization[J]. BMVC, 2018:105.
- [42] Maxime Oquab, Léon Bottou, Ivan Laptev, et al. Learning and Transferring Mid-Level Image Representations using Convolutional Neural Networks[C]. CVPR, 2014:1717-1724.
- [43] Careaga C, Hutchinson B, Hodas N, et al. Metric-Based Few-Shot Learning for Video Action Recognition[J]. arXiv Preprint arXiv:1909.09602, 2019.
- [44] Santoro A, Bartunov S, Botvinick M, et al. One-shot Learning with Memory-Augmented Neural Networks[J]. arXiv Preprint arXiv:1605.06065, 2016.
- [45] Ravi S, Larochelle H. Optimization as a Model for Few-Shot Learning[C].ICLR, 2017.
- [46] Wang B, Wang D. Plant Leaves Classification: A Few-Shot Learning Method Based on Siamese Network.[J]. IEEE Access, 2019,7: 151754-151763.
- [47] Oriol Vinyals, Charles Blundell, Tim Lillicrap, et al. Matching Networks for One Shot Learning[C]. NIPS, 2016: 3630-3638.
- [48] Liu L, Zhou T, Long G, et al. Prototype Propagation Networks (PPN) for Weakly-supervised Few-shot Learning on Category Graph[C]. IJCAI 2019, 2019:3015-3022.
- [49] Munkhdalai T, Yu H. Meta Networks.[C]. ICML 2017, 2017: 2554-2563.
- [50] Guangyi Chen, Tianren Zhang, Jiwen Lu, Jie Zhou. Deep Meta Metric Learning[C]. ICCV, 2019: 9546-9555.
- [51] Qian Q, Shang L, Sun B, et al. SoftTriple Loss: Deep Metric Learning Without Triplet Sampling[C]. ICCV, 2019: 6449-6457.

- [52] Schroff F, Kalenichenko D, Philbin J. FaceNet: A unified embedding for face recognition and clustering[C]. CVPR, 2015: 815-823.
- [53] Wang H, Wang Y, Zhou Z, et al. CosFace: Large Margin Cosine Loss for Deep Face Recognition[C]. CVPR, 2018:5265-5274.
- [54]Xun Wang, Haozhi Zhang, Weilin Huang, et al. Cross-Batch Memory for Embedding Learning. arXiv Preprint arXiv:1912.06798, 2019.
- [55]Geonmo Gu, ByungsOO Ko. Symmetrical Synthesis for Deep Metric Learning[C]. AAAI 2020, 2020: 10853-10860.
- [56] ByungSoo Ko, Geonmo Gu. Embedding Expansion: Augmentation in Embedding Space for Deep Metric Learning. arXiv Preprint arXiv: 2003.02546, 2020.
- [57] Michal R, Vit M, Anselm P, et al. Optimizing Rank-based Metrics with Blackbox Differentiation. arXiv Preprint arXiv: 1912.03500, 2019.
- [58] Jiayi Shen, Haochen Wang, Anran Zhang, et al. Model-Agnostic Metric for Zero-Shot Learning[C]. WACV 2020: 775-784.
- [59] 庞朴. 古墓新知—郭店楚简的价值[J]. 荆楚理工学院学报, 2003,18(2): 1-5.
- [60] 李学勤. 清华简九篇综述[J]. 文物, 2010(5): 51-57.
- [61] 马承源. 上海博物馆藏战国楚竹书.五[M]. 上海: 上海古籍出版社, 2002: 15-18.
- [62] 黄德宽. 安徽大学藏战国竹简概述[J]. 文物, 2017(09): 56-61.
- [63] Ioffe S, Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift[C]. ICML, 2015: 448-456.
- [64] Brenden M Lake,Ruslan Salakhutdinov,Joshua B Tenenbaum. The Omniglot challenge: a 3-year progress report[C]. arXiv:1902.03477. 2019.
- [65] Wolf L, Hassner T, Maoz I. Face recognition in unconstrained videos with matched background similarity[C]. CVPR, 2011: 529-534.
- [66] Liu W, Wen Y, Yu Z, et al. SphereFace: Deep Hypersphere Embedding for Face Recognition[C]. CVPR. 2017: 6738-6746.

## 致谢

时光在不经意中流逝，翻开自己的笔记本，字里行间透漏着实验室的枯燥与乏味，但这些字迹与数据却带着我将两年的生活回顾了一遍。我很庆幸加入了119这个大家庭，认识了优秀的你们，虽然实验是枯燥，但与你们一起学习、探讨、相处的时光却是愉悦的。

首先，我要对我的导师凡高娟老师说声谢谢，感谢您这两年的关心与照顾。也要把最真挚的感谢送给我的指导老师张重生老师。本文是在您的精心指导和悉心教诲下完成的，从选题确定、论证方法及基础资料的收集无不倾注着您的心血和智慧，使我把握了论文的方向，顺利地完成了论文的写作，在此对您表示最诚挚的感谢。您严谨的治学态度、科学的工作方法以及对工作的热忱无时无刻不在提醒着我，作为一名科研工作者应该怎么去对待自己的任务。

感谢实验室伙伴对我学习与生活的帮助，作为你们的实验室小伙伴，我觉得很幸运。在我实验出问题，理论分析不到位，科研遇到问题而陷入迷茫时，你们总能给我提出中肯的建议，你们的建议帮我解决了很多实验问题对你们不胜感激。正因为有你们的帮助，我的科研生涯才少了许多障碍。

在此感谢我生活学习了六年的母校—河南大学，母校给了我一个宽阔的学习平台，让我不断吸取新知，充实自己。

感谢我的父母一直在我身边给予鼓励与支持，他们在精神上物质上的无私支持坚定了我追求人生理想的信念。最后，感谢我的朋友们在这两年对我的关心与鼓励，因为有你们的相伴，我的生活才充满阳光。

2020年6月

民善  
新至  
德於  
明止