



第十四届全国大学生软件创新大赛

文档编号: SWC2021-T20210533-花生队

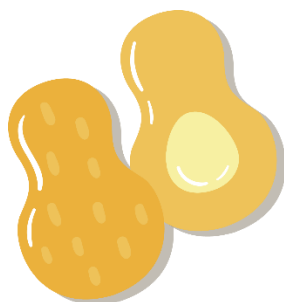


落笔云烟

Magpie's Pen

技术研究报告

Version: V1.3.1



花生队

2021.01.05

All Rights Reserved

# 目录

<b>1</b>	<b>问题聚焦</b>	<b>1</b>
1.1	问题描述	1
1.2	问题抽象	1
1.3	问题定位	2
1.4	问题评估	3
1.5	问题分解	6
<b>2</b>	<b>相关工作</b>	<b>8</b>
<b>3</b>	<b>技术方案</b>	<b>11</b>
3.1	技术方向	11
3.2	技术选择	11
3.3	结果期望	20
<b>4</b>	<b>技术实践</b>	<b>21</b>
4.1	使用的开发框架及依赖的库	21
4.2	技术实践过程	21
<b>5</b>	<b>结果验证</b>	<b>22</b>

文档修订历史

序号	修订原因	版本号	作者	修订日期	备注
1	创建技术研究报告	V1.0.0	A	2020.11.15	
2	创建问题聚焦模块	V1.1.0	C	2020.11.15	
3	更新问题聚焦模块	V1.1.1	C	2020.11.18	
4	创建相关工作模块	V1.2.0	B	2020.11.19	
5	创建技术方案模块	V1.3.0	C	2020.12.10	
6	更新技术方案模块	V1.3.1	C	2020.12.20	

# 1 问题聚焦

## 1.1 问题描述

“落笔云烟”是一个使用深度学习算法辅助用户练习硬笔书法的软件。项目按功能可以分成三个模块：书写图像识别、单字分析与评价、用户个性化服务。以下将从这三个模块对所涉及到的流程与问题简要描述：

1. 书写图像识别：用户上传书写练习的图像（包括但不限于字帖、作业等），系统对上传的图像进行处理与分析，并将处理后的结果保存于系统内，方便系统其他功能直接使用。同时，系统会对用户上传的练习图像进行整体分析，并将评价结果反馈给用户。
2. 单字分析与评价：该功能的输入是用户书写的单个字体的图像，图像可以由用户手动上传，也可以由系统的其他功能模块处理后得到。该模块可以对用户书写的单个字体进行打分，并且识别出可能导致评价下降的字体结构，并将结果反馈给用户。同时，系统会将用户单个字体处理后的结果保存于系统内，方便其他功能使用。
3. 用户个性化服务：该功能模块主要为用户提供个性化的服务，包括以下几个方面：书写推荐——根据用户薄弱字体分析的结果为用户推荐同样结构的字体进行练习；学习追踪——依据历史数据，项目可为用户提供具体字体风格变化情况以及评分变化曲线，记录用户的成长；娱乐模块——为解决现有练字所不得不面对的费用及枯燥，通过设定字体闯关等方式，为用户设定成长系统，变被动练习为主动训练。

## 1.2 问题抽象

根据 [1.1](#) 所列出的三大模块，本节将详细阐述各个模块包含的具体问题所对应的技术问题。

### ❖ 书写图像识别模块：

书写图像识别需要对用户上传的图像进行分析，识别出图像中的字，并向用户反馈总体的书写情况，因此涉及到以下技术问题：

- 图像的处理，去除图像中的噪声，并做一定程度的图像增强。
- 图像中单字的检测，检测出图像中手写的字，忽略非手写的字，例如图像中包含的印刷字等。
- 整体书写情况评价，结合处理后的图像与识别出的单个字体进行多元化的总体情况反馈。
- 书写报告生成，根据之前生成的整体书写评价，使用自然语言处理技术生成用户易于理解的书写报告。

### ❖ 单字分析与评价模块：

该功能需要对输入的每个字的图像进行分析，给出单个字体的评分，并给用户反馈该字体书写的缺陷，因此涉及到以下技术问题：

- 单个字体的图像处理，去除字体图像的噪声，并做一定的图像增强，方便后续流程的处理。
- 字体的评分，对用户书写的字体给出合理的评价，用户书写的字体可能有自己的风格，需要一个灵活的模型给出合理的评分。
- 总体书写缺陷，需要识别出用户书写的字体中的结构问题与总体问题，并以易于理解的方式反馈给用户。
- 局部书写缺陷，需要识别出用户书写的字体中可能存在的局部书写缺陷，例如部分笔画的问题，并将结果以易于理解的方式反馈给用户。
- ❖ 用户个性化服务：
 

该功能对用户的书写历史进行记录并建模分析，绘制出用户的总体评分曲线，并根据用户的书写历史推荐相似结构或易错的单字供用户练习；同时，根据用户的书写特点推荐最适合用户练习的字体、根据用户的字体特征生成用户的专属字体文件。因此涉及到以下技术问题：

  - 根据书写历史推荐练习内容及相似字体，需要根据用户之间以及单字的相似度来动态计算推荐群及信任子群，同时需要考虑解决冷启动问题。
  - 根据书写特征生成用户专属字体文件，以用户的历史书写的字体风格数据为基础，迁移生成个人字体库。

### 1.3 问题定位

对 1.2 节中所列出的技术问题总结如下：

技术问题	模块	业务领域	技术领域	备注
图像降噪、图像增强	书写图像识别、单字分析与评价	后台算法	数字图像处理	后台完成，不将结果反馈给用户
手写字检测	书写图像识别	后台算法	目标识别，深度学习	
整体书写评价	书写图像识别	后台算法	数字图像处理，机器学习	可以得到不同维度的原始评价信息，需要经过处理后可以生成用户可以理解的信息
书写报告生成	书写图像识别	前端算法，后台算法	可视化技术，自然语言处理	根据整体书写评价得到的信息生成用户可以理解的书写报告
单字评分	单字分析与评价	后台算法	计算机视觉，深度学习	根据用户书写的字体生成评分
总体缺陷分析	单字分析与评价	后台算法	计算机视觉，	分析单个字存

	价		深度学习	在的总 体缺陷，包括但不限于结构，笔画等
局部缺陷分析	单字分析与评价	后台算法	目标识别，深度学习	识别出单个字中存在的局部缺陷，与总体缺陷的不同之处在于局部缺陷可以通过目标识别等方式更直观得展示给用户

## 1.4 问题评估

技术问题	技术性	普适性	研究热度	问题热度
图像降噪	技术选取范围较广，技术难度不大。但考虑到书写图像环境因素的不确定因素较大，因此需要选择鲁棒性高的算法。	普适性高，针对书写图像问题优化后的具有高鲁棒性的图像降噪算法可以适应大部分的书写环境（指田字格、练习纸等含有图案或打印字体的书写环境）	OCR 问题的图像降噪技术在 70 年代便已被提出，在深度学习方法出现后效果又得到了显著提升。目前相关算法已经可以实现商业化大规模应用，近几年来研究热度有所下降。	广泛应用于各类视觉，图像处理任务中，热度较高。
手写字检测	手写字检测可以看作目标检测问题。目前主流的目标检测算法皆已成熟，技术难度较低。	主流的目标检测算法即可完成手写字识别任务，对大部分手写图像均可准确识别，普适性高。	Faster-RCNN 与 YOLO 模型的提出标志着目标检测问题的成熟化解决方案。现阶段大部分的目标检测任务均可使用这两类模型很好地完成，因此近几年的研究热度有所下降。	手写字的目标检测任务被广泛应用于 OCR 等任务当中，热度较高。

整体书写评价	整体书写评价涉及到文字提取与整齐度计算两个部分,文字提取可以使用目标检测方式完成,技术性一般,整齐度计算主要使用数字图像处理的方法完成,技术原理较为简单,具体实现时可以针对问题作出相应优化。	普适性高,目标检测模型与整齐度计算算法在单独问题下均可取得理想效果,两者结合可以处理大部分书写图像的评价问题。	手写图像检测问题见“手写检测”部分。整齐度计算算法现阶段相关研究较少,但可以利用自研究的整齐度模型进行计算。	整体书写评价在当下主要还是靠人工的方式来评价,对于该问题的数字化解决方案的需求很大,问题热度目前一般。
书写报告生成	根据整体书写评价功能给出的各个维度的评价,使用NLP与可视化技术给出用户方便理解的报告。该问题只需要使用到简单的NLP方法,技术性较低。可视化展示方面可以从多个维度进行探究,技术性中等。	普适性高,由于“整体书写评价”部分已经给出具体的评价数据,只需要将数据合理地展示给用户即可,结合NLP与可视化技术,大部分的情况下均可取得良好效果。	对于给定数据生成自然语言的问题属于NLG领域,NLG领域的技术仍然在不断发展,但是本项目中仅需要使用Slot Filling相关的技术即可,此方面的研究已经相对成熟,近些年来研究热度不高。	报告生成是当下众多APP都广泛应用的一项技术,包括但不限于学习报告、诊断报告等。即将数字化的信息以人方便接受的方式传递给用户,当下的应用热度很高。
单字评分	汉字的单个字体的评分涉及到的问题较多,工整、结构、笔画都是影响一个字美观的因素,这些因素很难用固定的标准来量化。因此我们选择使用深度学习相关的方法,通过大量输入人工评分的汉字,让模型来学习评价字是否美观的特征。由于美观这一属性的变化很大,所以在训练时需要特别注意其他和美	普适性较高,通过使用深度学习模型来学习评价一个字是否美观的特征,深度神经网络可以学习到很多有效的特征,因此该方法有较高的普适性。	单独针对汉字评分的模型到现在还没有相关研究。但对于其他一些由有限字母组成的语言,例如英语和阿拉伯语,有书写风格方面的研究。但针对书写美观领域的研究一直以来都处于空白状态,我们的工作也是第一次	在经过市场调研后我们发现有个别针对书法练习的APP也有评分功能,不过评分方式过于机械,且受限于部分汉字,实用性不高。总体而言,在相关技术均已成熟的现在,汉字的美观评价方面的需

	观无关的因素,并防止这些因素对最后的结果产生影响,这要求模型结构的合理设计与目标函数的合适选择,因此该任务技术难度较大。		使用深度学习来解决汉字美观度评价的问题。该领域总体研究热度较低。	求会越来越来大,这方面的热度也会越来越高。
总体缺陷分析	该技术期望找到影响汉字美观性的总体结构,并将结果反馈给用户,让用户有针对性地练习,纠正自己的错误。由于影响汉字书写美观的因素过多且难以量化,我们使用深度学习模型学习到的影响汉字美观的特征并将梯度反向传播找到原始书写图像中对汉字美观度影响最大的像素并将其高亮反馈给用户。需要注意的是该任务的效果非常依赖于方法的选择与评价模型的表现,因此如何选择合理的方法以及将结果合理地反馈给用户仍然是一项非常有技术挑战性的工作,该问题的技术难度高。	普适性高,总体缺陷分析的方法依赖汉字美观评价模型学习到的影响美观性的特征来进行分析与反馈,因此对输入有很高的普适性。	汉字总体缺陷分析的任务到目前为止并没有相关的研究。不过如何找深度学习网络中影响结果的输入特征是模型可解释性方面的一个很重要的研究方向,我们的方法也大致采用这种思路实现,该领域的研究热度很高。	市面上可以反馈书法缺陷的 APP 很少,实现了该功能的 APP 也只能根据机械化的几个评价指标给用户反馈,几乎不具有实用性。但缺陷分析是书法指导的重要一环,虽然当下该领域受制于技术问题热度很低,该领域值很有研究价值。
局部缺陷分析	汉字书写的局部缺陷分析任务主要是为了弥补总体缺陷分析的不足之处,给用户更具体的练习指导。总体缺陷分析反馈的是像素级信息,包含更多的是汉字书写的结构上的反馈,但用户书写问题可能会反映在部	普适性高,基于目标检测方式的局部缺陷检测方式可以适应大部分基于像素的汉字图片输入,因此普适性高。	汉字局部缺陷分析目标没有相关研究,不过汉字书写正确性的检测有相关的研究,但研究热度一直不是很高。目标检测任务是几年前热度很高的计算机	局部缺陷检测几乎没有 APP 有类似实现,该问题受制于技术因素热度也不是很高,但对于书法纠正而言这项技术很有意义。



	分笔画、部分局部结构上,因此本功能旨在将局部缺陷反馈给用户,提供更精细化的指导方向。该技术可以使用基于目标检测方法的模型来完成,但需要针对汉字缺陷分析任务做调整,因此技术难度较高。		视觉任务,我们也是首个将目标检测应用于汉字书写缺陷的工作。	
--	--	--	-------------------------------	--

## 1.5 问题分解

技术问题	子问题	描述	难度	依赖关系
图像降噪	①图像降噪	输入一张图片,去除图片中的噪声	中等	
手写字检测	②手写字检测	输入降噪后的手写字的图片,检测出图片中所有手写的字,返回 Bounding Box 和对应标签	简单	依赖①
整体书写评价	③单字评价	在多字图片中提取单字重心及 Bounding Box	中等	依赖②
	④整齐度计算	计算整体重心偏移量方差以及距离方差,并进行合理评估打分	简单	依赖③
书写报告生成	⑤参数预处理	根据整体书写评价得到的各项参数做相关预处理	简单	依赖③④
	⑥自然语言生成	根据得到的各项评价参数生成用户可以理解的书写报告	中等	依赖⑤
单字评分	⑦单字评价模型	使用深度学习	高	依赖①

		网络得到单字的评分		
全局缺陷分析	⑧ Attention 生成显著性区域	使用注意力机制生成缺陷部分的显著性图	高	依赖⑦
	⑨ Deconvolutional Network 显示最大刺激像素	使用反向卷积神经网络找到输入图片中对评分影响最大的像素	高	依赖⑦
局部缺陷分析	⑩检测缺陷区域	目标检测网络检测缺陷区域	中等	

## 2 相关工作

### 图像去噪

图像去噪声是对图像做预处理,使得处理后的图像更适合用于文字识别与目标检测等任务。常用的图像去噪方法包括二值化并设置阈值 (binarization)、模糊逻辑 (fuzzy logic)、图像直方图 (histogram) 与使用仿真算法和遗传算法为基础的方法。Farahmand 与 Ganchimeg 等人 [1][2] 总结了使用传统数字图像处理与启发式算法进行图像降噪的方法。Liu 等人 [3] 使用噪声水平函数 (Noise Level Function, NLF) 与高斯条件随机场 (Gaussian Conditional Random Field, GCRF) 提出了一种可以自动估计图像中的色彩噪声并产生去噪后清晰图片的方法。Sobia 等人 [4] 提出了去除乌尔语文档图片中的孔洞与噪声的方法。这些方法大多基于传统的数字图像处理技术,并没有使用深度学习领域的方法。

CycleGAN [5] 在图像-图像翻译领域取得了很好的效果,Sharma 等人 [6] 将 CycleGAN 方法应用到图像去噪领域,在多个数据集上均取得了良好的效果。本项目中将主要采用基于 CycleGAN 的图像降噪方法。

### 手写字检测

手写字检测本质上是目标检测问题,目标检测任务的模型可以分为基于 R-CNN 的模型和 YOLO 系列模型。Ren 等人 [7] 提出了 Faster R-CNN 模型,该模型解决了传统 R-CNN 模型无法端到端训练问题,并且使用 RPN 网络大大减少了提出区域的时间,在效率和准确度上均取得了很好的成果,这也意味着目标检测领域算法的成熟。Redmon 等人 [8] 提出了 YOLO 模型,该模型是目标检测算法的另一大家族,YOLO 算法可以在保持高准确率的同时完成实时的目标检测,很适用于实际应用场景。本项目中主要采用 Faster R-CNN 作为手写字检测的模型。

### 单字评分

单字评分的模型由我们独立设计,图像的全局特征提取部分使用 He 等人 [9] 提出的 ResNet 模型,之后通过全连接层得到单字的特征表示,提取特征后我们分别计算评分的回归 Loss 与汉字的分类 Loss,分类 Loss 使用 Tong 等人 [10] 提出的 RSS,最后模型的 Loss 是这两者的加权和。

### 全局缺陷检测

书写汉字的全局书写检测基于之前我们提出的单字评分模型,通过解释评分模型中可能导致评分下降的特征,将该刺激通过网络反向传播到输入的图像像素层得到影响评分的图像。我们通过使用 Zeiler 等人 [11][12][13] 提出的反向卷积神经网络的基本架构,结合 Springenberg [14] 等人提出的根据梯度寻找刺激像素的方法,完成书写汉字的全局缺陷检测。

另外,除了根据梯度反向研究输入像素的刺激外,我们还使用视觉注意力机制 [15][16][17] 在正向训练过程中寻找图像中影响评分的显著性区域,从而得到可能的全局书写缺陷的显著性图像。

## 局部缺陷检测

局部缺陷检测需要在输入的图像中检测出影响汉字美观性的局部区域,这个问题以目标检测模型为基本模型,使用 Faster R-CNN 架构 [7],使用我们手动标注的缺陷区域图像作为训练集,由于不存在分类的误差,所以可以直接使用 Faster R-CNN 架构中的 RPN 架构,并对提取到的区域使用 Non-maximum Suppression (NMS) [18] 算法做后处理,去除重复区域。

## 参考文献

- [1] Farahmand, Atena, Hossein Sarrafzadeh, and Jamshid Shanbehzadeh. "Document image noises and removal methods." (2013).
- [2] Ganchimeg, Ganbold. "History document image background noise and removal methods." *International Journal of Knowledge Content Development & Technology* 5.2 (2015): 11-24.
- [3] Liu, Ce, et al. "Automatic estimation and removal of noise from a single image." *IEEE transactions on pattern analysis and machine intelligence* 30.2 (2007): 299-314.
- [4] Javed, Sobia Tariq, et al. "Background and punch-hole noise removal from handwritten urdu text." *2017 International Multi-topic Conference (INMIC)*. IEEE, 2017.
- [5] Zhu, Jun-Yan, et al. "Unpaired image-to-image translation using cycle-consistent adversarial networks." *Proceedings of the IEEE international conference on computer vision*. 2017.
- [6] Sharma, Monika, Abhishek Verma, and Lovekesh Vig. "Learning to clean: A GAN perspective." *Asian Conference on Computer Vision*. Springer, Cham, 2018.
- [7] Ren, Shaoqing, et al. "Faster r-cnn: Towards real-time object detection with region proposal networks." *IEEE transactions on pattern analysis and machine intelligence* 39.6 (2016): 1137-1149.
- [8] Redmon, Joseph, et al. "You only look once: Unified, real-time object detection." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
- [9] He, Kaiming, et al. "Deep residual learning for image recognition." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.
- [10] Xiao, Tong, et al. "End-to-end deep learning for person search." *arXiv preprint arXiv:1604.01850* 2.2 (2016).
- [11] Zeiler, Matthew D., et al. "Deconvolutional networks." *2010 IEEE Computer Society Conference on computer vision and pattern recognition*. IEEE,

2010.

[12] Zeiler, Matthew D., Graham W. Taylor, and Rob Fergus. "Adaptive deconvolutional networks for mid and high level feature learning." *2011 International Conference on Computer Vision*. IEEE, 2011.

[13] Zeiler, Matthew D., and Rob Fergus. "Visualizing and understanding convolutional networks." *European conference on computer vision*. Springer, Cham, 2014.

[14] Springenberg, Jost Tobias, et al. "Striving for simplicity: The all convolutional net." *arXiv preprint arXiv:1412.6806* (2014).

[15] Anderson, Peter, et al. "Bottom-up and top-down attention for image captioning and visual question answering." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018.

[16] Jetley, Saumya, et al. "Learn to pay attention." *arXiv preprint arXiv:1804.02391* (2018).

[17] Lu, Jiasen, et al. "Knowing when to look: Adaptive attention via a visual sentinel for image captioning." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.

[18] Neubeck, Alexander, and Luc Van Gool. "Efficient non-maximum suppression." *18th International Conference on Pattern Recognition (ICPR'06)*. Vol. 3. IEEE, 2006.

## 3 技术方案

### 3.1 技术方向

功能	技术问题	技术方向
书写图像识别	图像降噪	数字图像处理, 计算机视觉, 深度学习
	手写字检测	计算机视觉, 深度学习
	整体书写评价	数字图像处理
	书写报告生成	自然语言处理
单字分析与识别	图像降噪	数字图像处理, 计算机视觉, 深度学习
	单字评分	计算机视觉, 深度学习
	全局缺陷分析	计算机视觉, 深度学习
	局部缺陷分析	计算机视觉, 深度学习
用户个性化服务	练习推荐	推荐系统, 信息检索
	用户字库生成	计算机视觉, 深度学习

### 3.2 技术选择

#### 图像降噪

本项目选择使用 CycleGAN 作为图像降噪模型。对抗生成网络 (Generative Adversarial Networks, GANs) 是基于深度学习的生成模型。GAN 模型体系结构涉及两个子模型: 一个用于生成新示例的生成器模型, 以及一个用于对所生成的示例是真实的还是由生成器模型伪造的图片进行分类的鉴别器模型。GAN 模型的结构如下图所示:

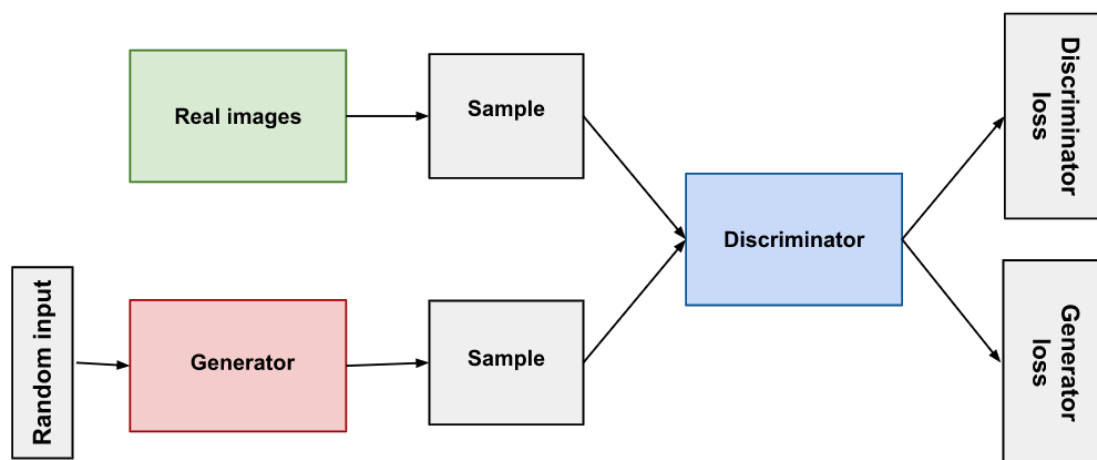


图 3.2.1 CycleGAN Architecture

CycleGAN 是 OCR 领域中出色的图像预处理模型, CycleGAN 的一个优点就是它不需要图像对作为训练数据。尽管仍然需要输入和输出数据,但它们不需要彼此直接对应。由于在大多数领域都很难找到图像对作为训练数据,因此 CycleGAN 的无监督训练功能确实非常有用。CycleGAN 的体系结构包括两对生成器和鉴别器。每个生成器都有一个对应的鉴别器,该鉴别器试图从真实图像中评估其合成图像。与任何 GAN 一样,生成器和鉴别器在对抗中学习。每个生成器都试图“欺骗”相应的鉴别器,而鉴别器则学会了不被“欺骗”。为了使生成器保留原始输入文本的文字信息, CycleGAN 模型计算了循环一致性误差 (Cycle Consistency Loss), 该损失评估了往返于产生空间的图像在多少程度上与其原始版本相似。下面是 CycleGAN 模型的整体流程:

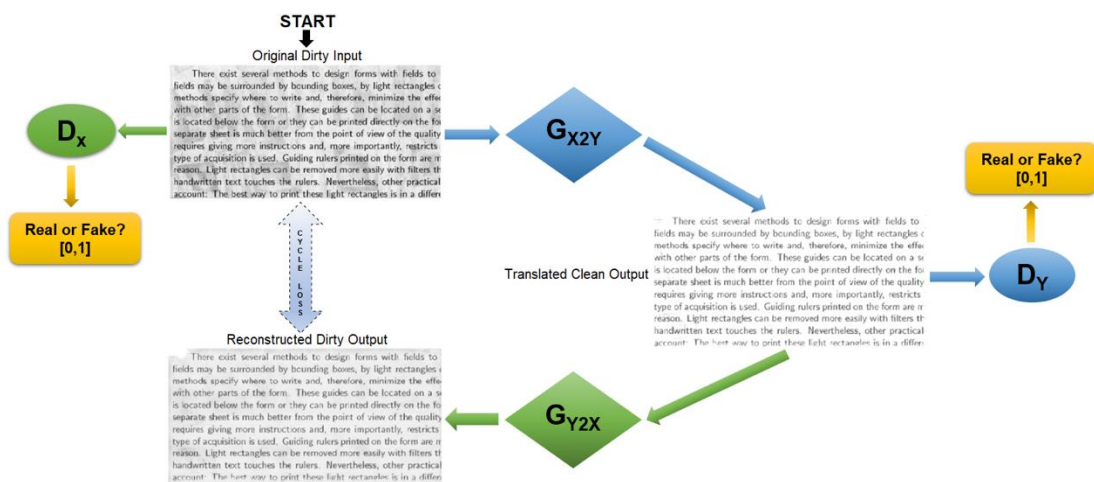


图 3.2.1 First Generator

第一个生成器  $G_{\{X2Y\}}$  将原始图片输入转换为清洗后的输出。鉴别器  $D_Y$  将尝试评估转换后的输出是真实图像还是生成器生成的图像。然后,鉴别器将提供所评估的图像是真实图像的可能性。

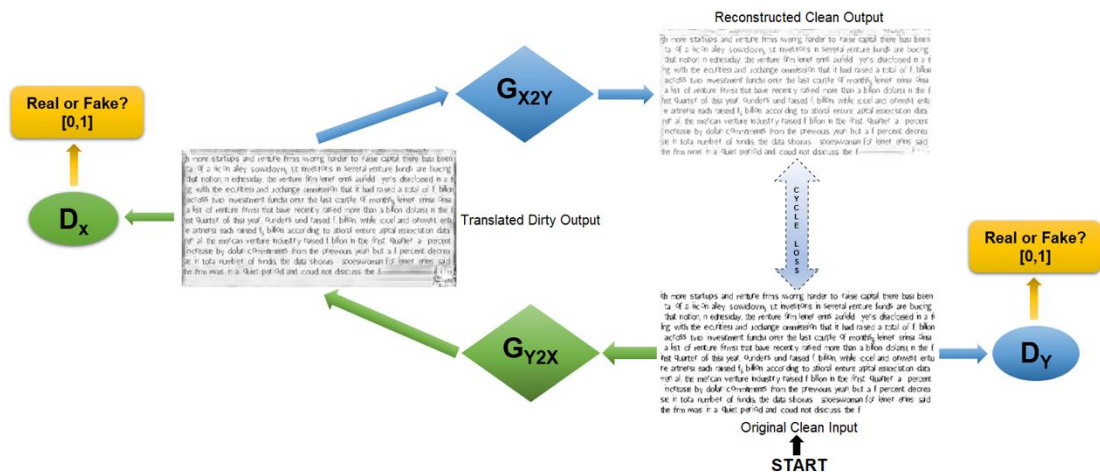


图 3.2.3 Second Generator

第二个生成器  $G_{\{Y2X\}}$  将原始的清洗后输入转换为转换后的有噪声的输出。鉴别器  $D_X$  将尝试从生成的图像中分辨出真实的图像。创建的模型将在两个方向上进行训练,分别带有一组有噪声的图像和一组无噪声图像。

## 手写字检测

手写字检测基于 Faster R-CNN 模型, Faster R-CNN 模型是 R-CNN 模型的第三个迭代版本, 其实现了端到端的检测, 因此可以有效利用 GPU 运算的并行性, 大大提升了目标检测的效率, 同时端到端的学习也提升了模型的特征学习效果, 使得模型的准确性也有所提升。Faster R-CNN 的整体架构如下图所示:

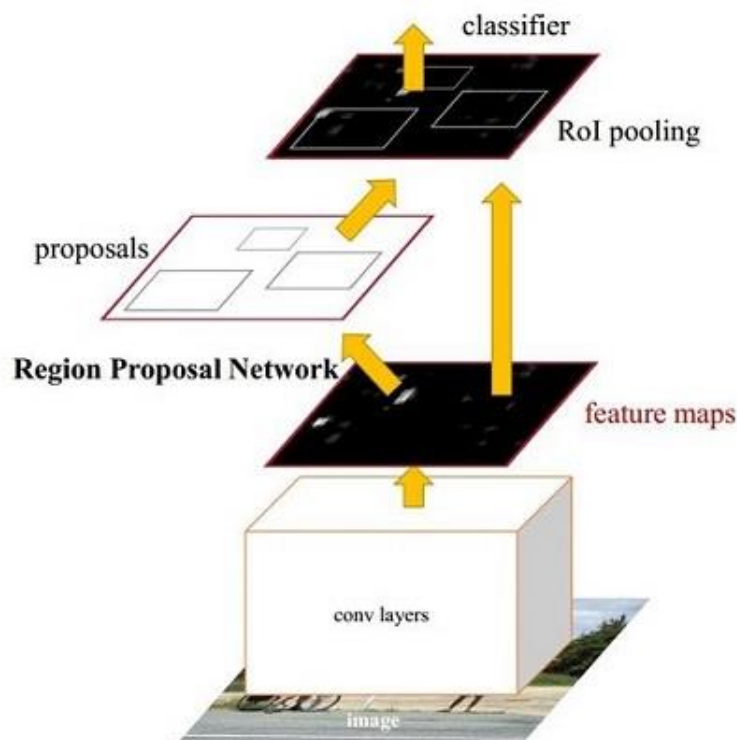


图 3.2.2 Region Proposal Network

Faster R-CNN 的核心部分是区域提出网络 (Region Proposal Network, RPN), RPN 可以单独从 Faster R-CNN 架构中抽出, 作为独立网络结构训练, RPN 的作用是针对给定的输入图片, 提取出可能包含目标的区域, RPN 的作用原理是将 Feature Maps 中的每一个点都作为 Anchor, 计算该点是否包含目标, 并且生成一个以这个点为基准的 Bounding Box。RPN 设置中预先设置了几个比例值, 对每个 Anchor 生成这几个比例值的 Bounding Box, 然后对每个 Bounding Box 计算其修正后的框的位置, 其包含四个值共包含四个值  $\{\Delta cx, \Delta cy, \Delta h, \Delta w\}$  代表 Bounding box 的中心的偏移量和 Bounding Box 的宽高的修正值。



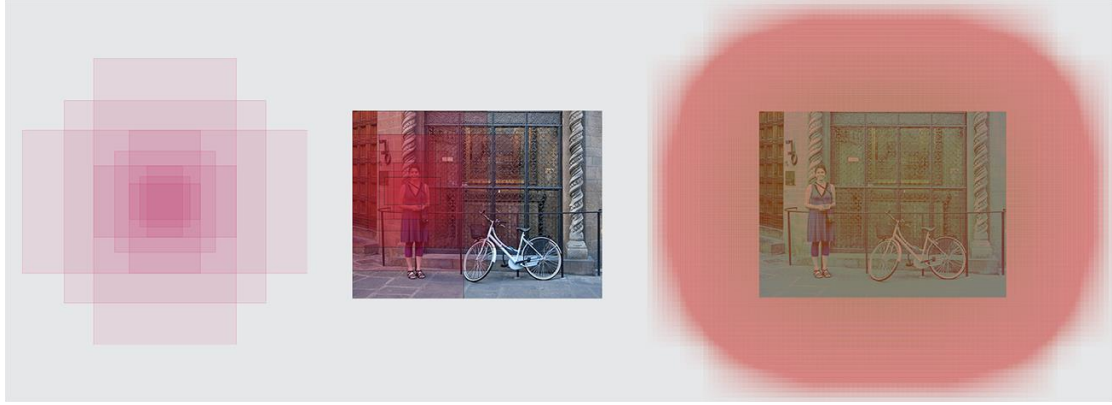


图 3.2.3 Different Ratio of Anchors

RPN 同时需要预测对应的 Bounding Box 中是否含有目标，这里并不是要检测出目标具体是什么，只需要区分 Anchor 所在的这个区域中是目标还是背景即可，所以其实可以看作一个二分类任务。RPN 识别出所有的 Bounding Box 后根据图片的 Bounding Box 与真实图片的 IoU 判断其是否包含目标。RPN 的 Loss 由两部分组成，一部分是分类的 Loss，另一部分是 Bounding Box 的四个值的回归 Loss，这两部分误差经反向传播后更新模型的参数。手写字检测任务只需要检测出当前字是否是手写字即可，所以其只需要两个标签，判断是否是手写字，因此可以直接使用 RPN 模型作为手写字识别模型。

### 整体书写评价

取图像分割模型中的单字分割结果序列作为输入，记为  $P = \{p_1, p_2, \dots, p_n\}$ ，其中  $p_i = \{x_1, y_1, x_2, y_2\}$  代表单字的包围盒坐标。据此可以计算得到每一行字的重心方差  $S_g^2$ 、字距方差  $S_{space}^2$ 、单字面积方差  $S_{square}^2$ 、底部位置方差  $S_{bottom}^2$  以及平均字距与平均单字面积比值  $E(P)$ 。最终文章的章法评价函数可以采用

$$E(Article) = S_g^2 + S_{space}^2 + S_{bottom}^2 + \frac{E(P)}{E_{std}(P)},$$

其中  $E_{std}(P)$  为预训练样本中得到的标准比值。

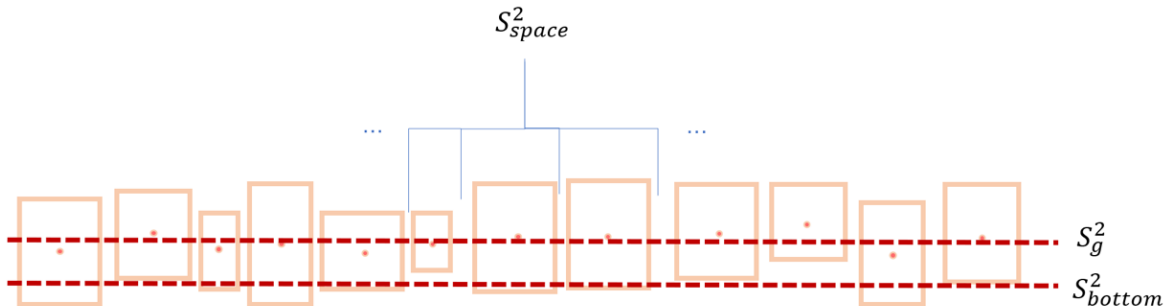


图 3.2.6 整体书写评价模型图

### 单字评分

单字评分使用端到端的深度学习模型，输入一张图片，输出图片中书写的汉字的评分。本项目中使用到的模型由我们独立设计与训练，模型需要提取图片的特征，生成对应的 Feature Maps，这部分我们使用 ResNet-50 模型来完成。ResNet 提取到的 Feature Maps 经过全连接层的映射可以得到图片的特征向量，这里我们需要根据特征向量计算出汉字的评分，这里可以将其看作回归问题，使用全连接层直接计算汉字的评分，并根据 Loss 函数更新模型的参数，但存在的问题是当某个特定汉字的训练图片数量较少且其评分分布不均匀时，模型可能会认为某个固定的结构会导致评分降低，这种现象在训练集较少且分布不均匀时尤其明显。所以我们增加一个分支使用 Softmax 层来计算汉字的分类，从而得到分类的误差  $Loss_{id}$ ，这里需要注意的是汉字的分类非常多，所以若是直接使用反向传播来优化效果并不会很好，所以这里使用随机采样误差 (Random Sampling Loss, RSS) 来计算分类的误差，RSS 的做法如下：

假设目标的类别为 1 到 C，共有 C 个类别，假设训练集中的数据可以表示为  $\{x, t\}$ ，其中  $x \in R^C$  是 Softmax 层的输出值，代表汉字属于各个类别的概率， $t$  是一个 C 维的 one hot 向量表示汉字正确的标签，那么原本的 Softmax Loss 可以表示为：

$$L = - \sum_{i=1}^C t_i \log y_i$$

RSS 则会随机从 C 个类别中选择 K 个类别 (K 远小于 C) 来计算误差与梯度，假设选择的类别为  $i_1, i_2, \dots, i_K$ ，则分类器的概率分数与目标标签可以表示为  $\tilde{x} = (x_{i_1}, x_{i_2}, \dots, x_{i_K})^T$ ， $\tilde{t} = (t_{i_1}, t_{i_2}, \dots, t_{i_K})^T$ ，RSS 可以用以下公式表示：

$$\tilde{L} = - \sum_{i=1}^K \tilde{t}_i \log \tilde{y}_i$$

其中类别的选择可以按如下规则：假设  $x$  的真实类别为  $c$ ，则设置  $k_1 = c$ ，之后从  $\frac{\{1, \dots, C\}}{c}$  中随机选择 K-1 个类别。

计算汉字评分的误差函数我们选择使用 Smooth L1 Loss，记为  $L_{reg} = SmoothL1(x, y)$ ，评价模型的误差是回归误差与分类误差的和：

$$Loss = L_{id} + L_{reg}$$

模型的总体结构如下图所示

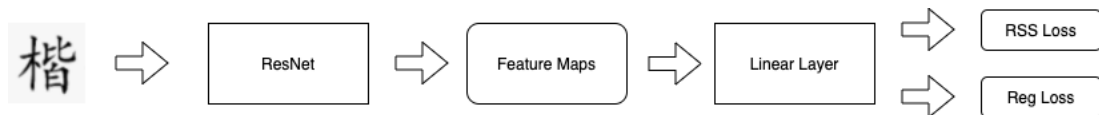


图 3.2.7 Score Model Architecture

### 全局缺陷分析

汉字的全局缺陷分析主要使用 Zeiler 等人提出的反向卷积神经网络

(Deconvolutional Networks), 其架构如下图所示。反向卷积神经网络是可视化 CNN 隐层的比较流行的架构, 其可以可视化 CNN 的各个 Kernel 学习到的不变特征, 并将刺激反向传递到输入像素, 显示出输入像素中对该 Kernel 生成的 Feature Map 刺激最大的一系列像素点。

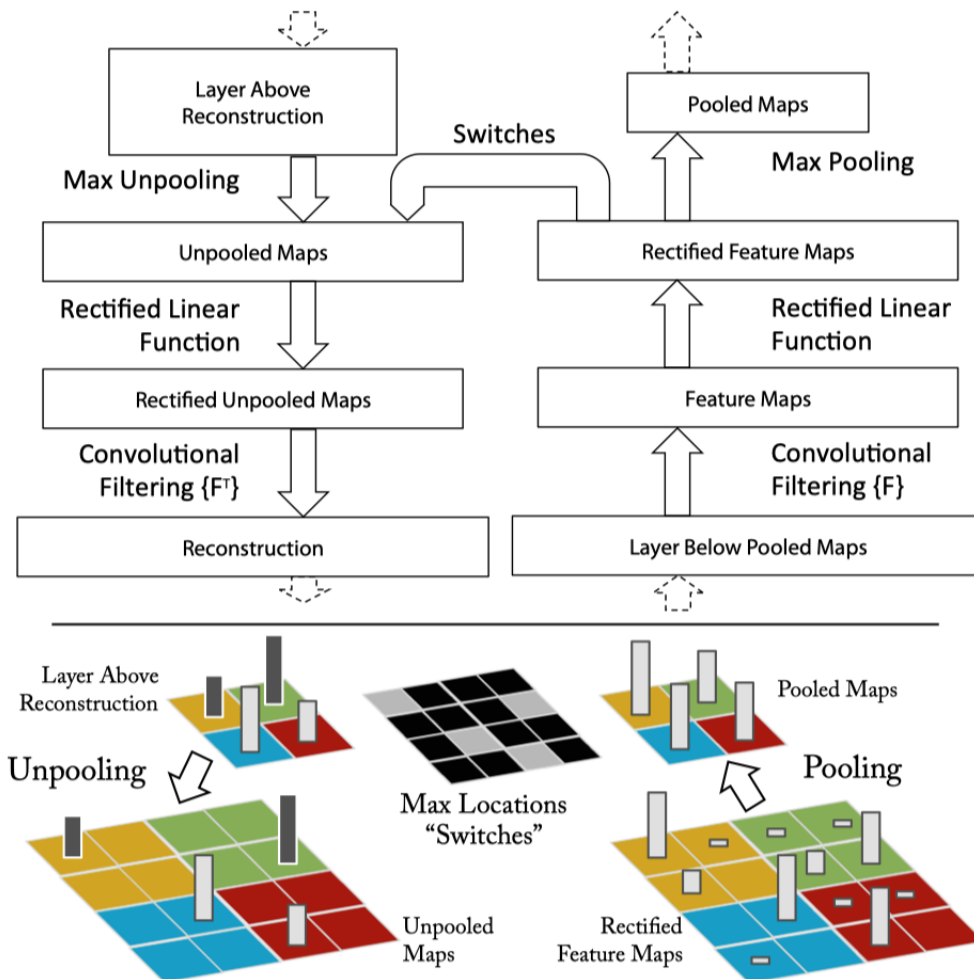


图 3.2.8 Deconvolutional Networks

Springenberg 等人在提出的 All Convolutional Network 中改进了 Zeiler 等人的架构, 使用梯度指导的传递, 让刺激信号在反向传播时可以获得更鲁棒的效果, 他们的模型的实际效果如下图所示。在我们的项目中, 对于输入的汉字图片, 我们需要找到其输入像素中最影响评分的部分像素, 这可能是由于部分笔画的缺陷, 或者书写的汉字的结构缺陷引起的。因此根据我们之前设计的评分模型, 我们将最后输出层的分数输出设置为 0, 再将这部分梯度反向传播至输入层, 过程中将保持特定 Kernel 的有效性, 其余 Kernel 均设为 0, 便可得到书写汉字的缺陷像素图。

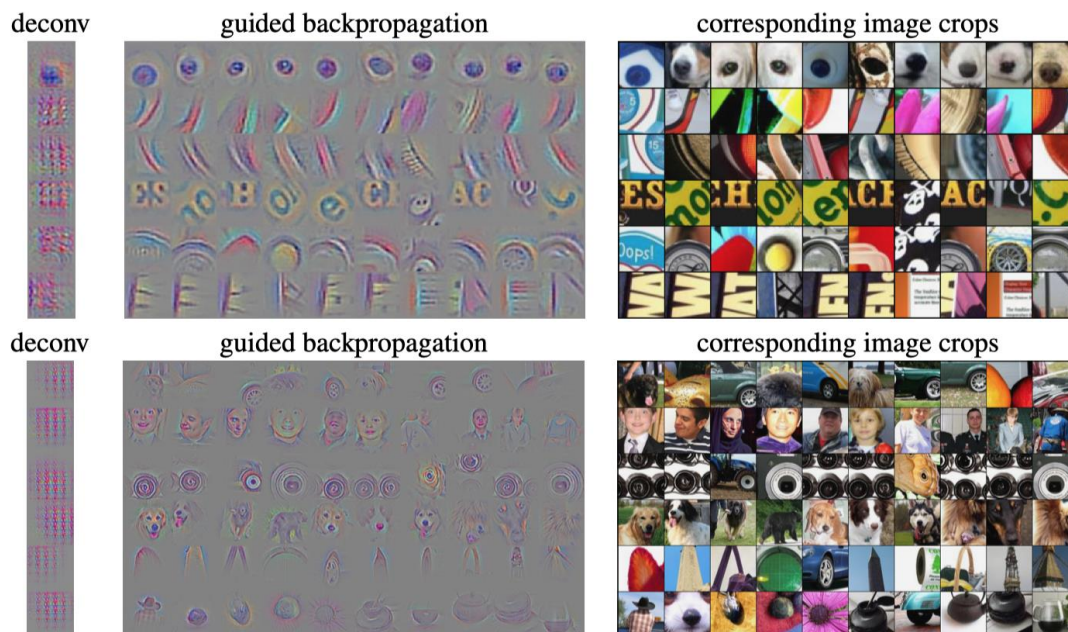


图 3.2.9 Guided Backpropagation

### 局部缺陷检测

局部缺陷检测可以视作目标检测问题，这里我们仍然使用 Faster R-CNN 作为检测模型，根据我们手动标注的缺陷部分数据集检测汉字的缺陷部分。这里主要注意的是缺陷部分需要尽量精准地表示出汉字的缺陷，否则与全局缺陷的功能便有所冲突，我们只需要找到汉字出问题的局部，如下图所示：



图 3.2.10 文字汉字局部示例

因此 RPN 网络给出的大量的 Bounding Box 我们需要做后处理，传统的后处理方法是 Non-Maximum Suppression (NMS) 算法，但该算法在处理有部分重合的 BBox 时会遇到阈值问题，比如下图中的三匹马，仅仅是 0.9 的马会被保留，其余的 Bbox 均会被舍弃

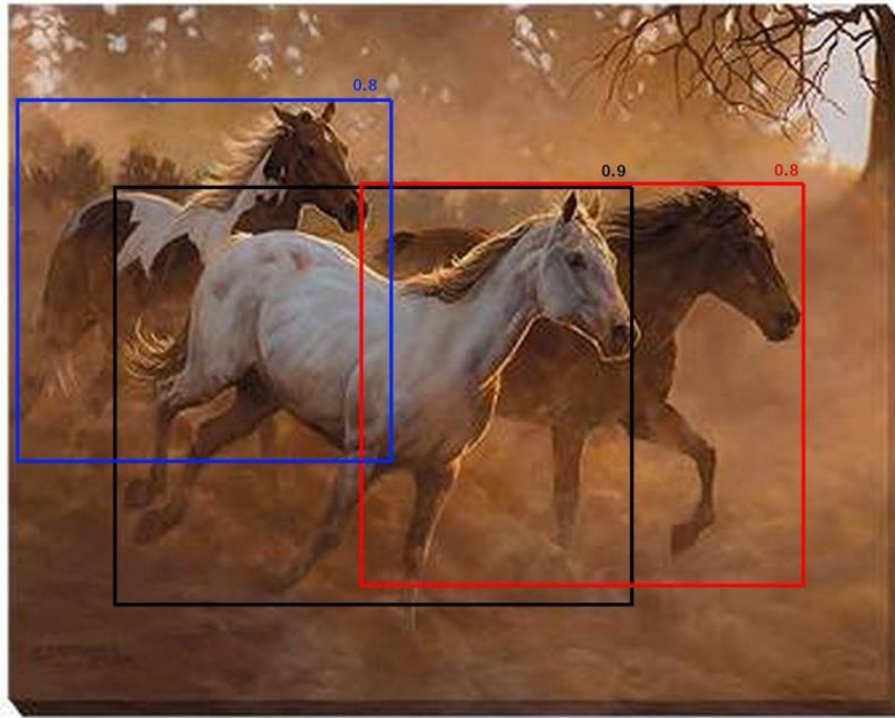


图 3.2.11 Example of Bounding Boxes

因此我们选择了 NMS 算法的变种 Soft-NMS 算法作为我们的后处理算法，其计算公式如下：

$$s_i = \begin{cases} s_i, & \text{iou}(\mathcal{M}, b_i) < N_t \\ s_i(1 - \text{iou}(\mathcal{M}, b_i)), & \text{iou}(\mathcal{M}, b_i) \geq N_t \end{cases},$$

该算法不是完全删除具有较高 IOU 和高置信度的区域，而是降低与 IOU 值成比例的区域置信度来计算，因此更适合我们的项目。

### 书法练习推荐

通过不确定近邻的协同过滤算法实现针对性推荐。由于易错字相似的用户群体在错字选择直接存在较强的相关度，因此可以根据基于用户之间易错字的相似性计算，来适当的选择预测目标的近邻作为推荐群，同时计算推荐群中对预测目标的信任子群，在充分结合推荐群和信任子群的基础上，通过不确定近邻因子分析度量来计算预测目标的推荐结果。该算法相比较传统的基于用户以及基于产品的协同过滤推荐算法，可以有效平衡用户群以及产品群推荐结果所带来不确定的影响，有效缓解用户评分数据极端稀疏情况使用传统性度量方法带来的问题，并显著提高推荐系统的推荐质量。算法包括以下五个步骤：

#### 1. 构建评分矩阵。

在推荐系统中，用户所有历史单字评分的数据库中包含  $s$  个用户的集合  $U = \{U_1, U_2, \dots, U_s\}$  和  $t$  个产品的集合  $I = \{I_1, I_2, \dots, I_t\}$ ，用户评分数据集可用一个  $s \times t$  阶矩阵表示， $s$  代表  $s$  个用户， $t$  列代表  $t$  个产品（见表 1）。假设某一用户  $U_a$  历史中  $I_j$  的评分为  $R_{a,j}$ ，则该评分体现了用户对单字的  $I_j$  的



掌握程度。

	$I_1$	$\cdots$	$I_j$	$\cdots$	$I_t$
$U_1$	$R_{1,1}$	$\cdots$	$R_{1,j}$	$\cdots$	$R_{1,t}$
$\cdots$	$R_{a,1}$	$\cdots$	$R_{s,j} - 7$	$\cdots$	$R_{a,t}$
$\cdots$	$R_{s,1}$	$\cdots$	$R_{ij}$	$\cdots$	$R_{s,t}$

图 3.2.10 用户-单字评分矩阵  $R(s \times t)$

## 2. 相似性计算

可以是用户之间的相似性计算，也可以是字形相似度间的计算。相关计算方法有夹角余弦、修正的夹角余弦、Pearson 相关系数、jaccard 相似度、巴氏系数等。

下列为采用修正的余弦相似度计算用户之间相似性，通过减去单字评分平均值将整体数据移动到空间原点。选取用户  $U_a$  和  $U_b$  打分的交集 ( $I_{U_a} \cap I_{U_b}$ )，定义为

$$\text{sim}(U_a, U_b) = \frac{\sum_{i \in I'} (R_{a,k} - R_a) \times (R_{b,k} - R_b)}{\sqrt{\sum_{i_k \in I'} (R_{a,k} - R_a)^2} \times \sqrt{\sum_{i_k \in I'} (R_{b,k} - R_b)^2}}$$

其中  $R_a$  是用户对于  $a$  字得分的平均值，计算结果  $\text{sim}(U_a, U_b)$  的值落在  $[0, 1]$  区间中， $\text{sim}(U_a, U_b)$  值越大，则表示用户  $U_a$  和  $U_b$  之间的相似性越高。基于产品的相似度计算过程与之类似。

## 3. 动态选择目标的推荐对象群

在进行邻近对象选择之前，需要界定预测目标的推荐对象应该如何选取，通过定义两个相似度计算的阈值，只考虑选择与目标较为接近的作为推荐对象，定义

$$S(U_a) = \{U_x | \text{sim}'(U_a, U_x) > \mu, a \neq x\}$$

$$S(I_j) = \{I_y | \text{Sim}'(I_j, I_y) > \nu, j \neq y\}$$

## 4. 在推荐对象中选择信任子群

针对目标进行推荐对象选择过程中，相似度计算成了主要的衡量指标，但是，在实际的推荐系统中，往往用户的相似度计算，可能仅仅来源于对少数几个字的得分，甚至可能只有一个共同评分的字，这样的相似度计算，存在较大的偶然因素。因此，除了要考虑相似度，也需要考虑两者之间共同评价单字的个数。计算共同打分个数大于设定的阈值的用户推荐群，定义为  $S'(U_a)$ ，计算目标项目推荐准确度较高的信任因子，定义为  $S'(I_j)$ 。

$$S'(U_a) = \{U_x | \text{Sim}'(U_a, U_x) > \mu \& |I_{U_a} \cap I_{U_x}| > \varepsilon, a \neq x\}$$

$$S'(I_j) = \{I_y | \text{sim}'(I_j, I_y) > \nu \& |U_{I_j} \cap U_{I_y}| > \gamma, j \neq y\}$$

计算两个信任子群的对象个数，分别计算  $|S'(U_a)| = m'$  和  $|S'(I_j)| = n'$

## 5. 不确定近邻的协同过滤算法

对于目标的在线用户  $U_a$  以及其他未浏览过的单字  $I_j$ ，同时结合用户的最近邻集和单字的最近邻集对用户单字上的得分进行预测，推荐公式为：

$$R_{a,j} = \lambda \times \left( R_a + \frac{\sum_{x \in S(U_a)} \text{sim}'(U_a, U_x) \times (R_{x,j} - R_x)}{\sum_{U_x \in S(U_a)} \text{Sim}'(U_a, U_x)} \right) +$$

$$(1 - \lambda) \times \left( R_i + \frac{\sum_{l_y \in S(s_j)} \text{sim}'(I_j, I_y) \times (R_{a,y} - \bar{R}_y)}{\sum_{l_y \in S(I_j)} \text{sim}'(I_j, I_y)} \right)$$

其中  $R_a$ ,  $R_b$  分别表示用户  $U_a$ ,  $U_b$ , 对其他字所有得分的均值,  $R_j$ ,  $R_y$  表示字  $I_j$ ,  $I_y$  已知所有用户得分的均值, 公式中根据用户  $U_a$  和字  $I_j$  的不确定近邻群进行推荐, 假如用户  $U_a$  的近邻群为空, 则完全按照字  $I_j$  的近邻群进行协同过滤, 若  $I_j$  的近邻群为空, 则完全按照用户  $U_a$  的近邻群进行协同过滤。

### 3.3 结果期望

功能	技术问题	结果期望
书写图像识别	图像降噪	可以明显地去除原始图片中的噪声, 因光线引起的像素值变化, 多余的线条、污点等。
	手写字检测	可以检测出所有手写的字, 并精确地生成包含单个字的边框。
	整体书写评价	可以生成多个维度的评价分数, 该分数应该是可解释的。
	书写报告生成	生成用户可以理解的一段话, 报告书写的好坏情况。
单字分析与识别	图像降噪	同“书写图像识别”部分
	单字评分	可以给出一个汉字美观性的合理评分, 尤其是两极化比较严重的汉字(很规范与很差), 需要给出合理的分数。
	全局缺陷分析	可以给出大致反映汉字结构问题的显著性图像。
	局部缺陷分析	可以给出可以精确指示汉字局部缺陷部分的边界框。

## **4 技术实践**

### **4.1 使用的开发框架及依赖的库**

### **4.2 技术实践过程**



## 5 结果验证