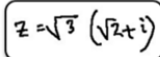
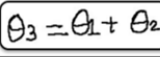
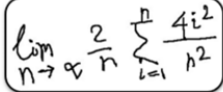
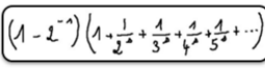


AI3607-深度学习及应用 大作业说明

任务：针对基于深度学习的手写数学表达式识别算法进行研究与探索。

背景：手写数学表达式识别（HMER）旨在从手写数学表达式图像生成相应的 LaTeX 序列，被广泛应用在自动评分、在线教育、公式图像搜索等领域。HMER 任务不仅需要识别 LaTeX 实体字符，还需理解实体符号间的结构关系去生成适合的 LaTeX 结构符号（例如“ \wedge ”、“ $-$ ”、“ $\{$ ”和“ $\}$ ”）。

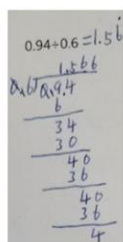
Input Image	Ground Truth
	$z = \sqrt{3} (\sqrt{2} + i)$
	$\theta_3 = \theta_1 + \theta_2$
	$\lim_{n \rightarrow \infty} \frac{2}{n} \sum_{i=1}^n \frac{4i^2}{n^2}$
	$(1 - 2^{-s}) (1 + \frac{1}{2^s} + \frac{1}{3^s} + \frac{1}{4^s} + \frac{1}{5^s} + \dots)$

现有的方法可以概括为基于序列和基于树的方法：

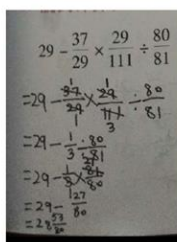
- 1) 基于序列的方法对公式图像看成是 Image-to-Sequence 任务，通过提取有助于字符识别的重要区域来顺序输出字符结果。
- 2) 基于树的方法对公式进行结构化建模，将符号之间的关系表示为树或分层结构，通过解码建模的结构来预测结果。

手写数学表达式识别（HMER）主要难点：

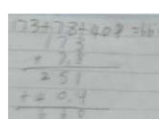
- 1) 手写风格多变，符号相似性（2/z, a/alpha, w/omega, ..., q/g, p/beta）
- 2) 公式结构复杂，多层嵌套
- 3) 多行文本



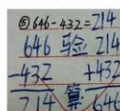
(a) Long Division with Multiple Line Structures



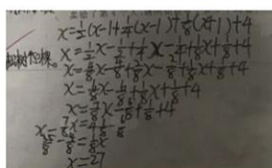
(b) Multi-line Equalities with Fraction Cancellation



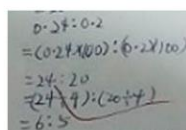
(c) Column Addition with Image Blur



(d) Two Column Calculations with Semantic Reading Order



(e) Equation Solving with Extremely Dense Text and Inadequate Illumination



(f) Simplification of Ratio Expression with Low Image Resolution

本课程的大作业共有 4 个任务可供选择，总结如下：

	任务简述	考核内容	满分
任务 A	复现已经开源的算法	1. 实验报告（50%） 2. Presentation（20%）	80%
任务 B	改进现有 HMER 算法	1. 源代码和训练模型（40%） 2. 算法报告（40%） 3. Presentation（20%）	100%
任务 C	开放性科研探索 【需要申请-审核】	1. 源代码和训练模型（40%） 2. 小论文（40%） 3. Presentation（20%）	100%

说明：

- 1) 最多两人一组，在最终报告中需注明贡献。
- 2) Task A 本身是 Task B 的一个步骤，可作为 TaskB 的保底分数，鼓励在有余力的基础上尝试 Task B/Task C
- 3) 有问题，可随时联系助教 官同坤 gtk0615@sjtu.edu.cn

任务 A：复现已开源的算法

A.1 任务简述

选择一个开源的 HMER 算法(推荐选择的算法 [CoMER](#), [SAN](#), [CAN](#), 和 [LAST](#))，仔细阅读其论文，深入理解其算法。在深度学习实验平台上配置环境，下载并运行代码。复现算法在若干数据集上的训练和测试结果。撰写至少 4 页的实验报告，包括配置环境、对算法的理解、训练测试结果、遇到的问题 and 解决方式等；准备一个 5 分钟的 pre（+3 分钟 QA），在 15-16 周课堂上进行一次展示。

A.2 主要步骤

1. 从 Github 开源仓库下载代码
2. 按照文档指示配置 PyTorch 环境
3. 选择进行实验 HMER 开源算法(推荐选择的算法 [CoMER](#), [SAN](#), [CAN](#), 和 [LAST](#))，仔细阅读论文并深入理解该算法

4. 运行实验，复现论文中的训练/测试结果
5. 撰写报告，准备展示

A.3 评分标准

实验报告 (60%): 实验过程描述清晰准确, 对算法的理解深刻、有独特见解。

Presentation (20%): 对实验、算法的描述清晰准确。

任务 B: 改进现有的手写数学表达式识别算法

B.1 任务简述:

阅读现有的 HMER 算法文章 (见[最后一页参考文献](#)), 仔细阅读论文并深入理解这些算法。首先在深度学习实验平台上配置 Pytorch 环境, 下载并运行代码 (推荐使用 Github 开源仓库的代码), 复现该算法在若干数据集上的训练和测试结果, 作为后续实验的参考。发挥你的聪明才智和探索精神, 针对前面提到的 HMER 难点, 修改算法模块、模型超参数等, 尝试提升算法的性能。

评估指标: ExpRate, Errors (Edit Distance)

数据集: CROHME14/16/19 数据集 [CROHME Home Page \(rit.edu\)](#)

鼓励选择基于树的方法作为 baseline model

鼓励探索更好的公式编码-解码方式

使用 Latex 模板, 撰写至少 4 页 (英文、双栏) 的算法报告, 包含背景介绍 (Introduction)、提出的方法 (Our Approach)、实验结果 (Experiment)、消融实验 (Ablation Study) 等部分; 准备一个 5 分钟的 pre (+3 分钟 QA), 在 15-16 周课堂上进行一次展示。

B.2 主要步骤

1. 阅读现有的 HMER 算法文章 (见[最后一页参考文献](#)), 仔细阅读论文并深入理解这些算法
2. 在 Github 开源仓库下载代码
3. 按照文档指示配置 PyTorch 环境, 运行实验, 复现论文中的训练/测试结果
4. 发挥你的聪明才智和探索精神, 修改算法模块、超参数等, 提升算法的性能
5. 使用 Latex 模板, 撰写至少 4 页 (英文、双栏) 的算法报告
6. 撰写报告, 准备展示

B.3 评分标准

源代码和训练模型 (40%): 代码清晰、注释丰富, 变量、函数等命名合理。汇报

的精度需要可以复现，无法复现的代码将会被惩罚。

算法报告 (40%): 算法创新程度、英语表达流利、行文逻辑清晰、对 HMER 模型有自己的洞察。文章图文并茂，实验结果清晰、消融实验合理。

Presentation (20%): 对实验、算法的描述清晰准确，能够准确地描述新算法。

任务 C: 开放性科研探索

C.1 任务简述:

作为一个深度学习落地的研究方向，文字识别领域还有许多有趣的研究方向等待大家探索，包括但不限于场景文本检测和识别、关键信息抽取、表结构识别、文本-图像匹配用于图像检索等。我们鼓励开放性的科研探索作为深度学习课程的大作业，选择该方向的同学需要在 **2024 年 5 月 1 日** 前向助教提交一份科研提案，篇幅不限，其中需要包括拟使用的技术、拟解决的问题、计划达成的效果。助教审核通过后即视为成功选择该任务。使用 Latex 模板，撰写至少 6 页（英文、双栏）的小论文，包含背景介绍（Introduction）、相关工作（Related Work）、提出的方法（Our Approach）、实验结果（Experiment）等部分；准备一个 5 分钟的 pre (+3 分钟 QA)，在 15-16 周课堂上进行一次展示。

C.2 主要步骤

1. 在 **2024 年 5 月 1 日** 前向助教提交一份科研提案，篇幅不限，其中需要包括拟使用的技术、拟解决的问题、计划达成的效果
2. 阅读现有算法介绍，仔细阅读论文并深入理解这些算法
3. 落实你的科研提案，开展理论和实验研究
4. 使用 Latex 模板，撰写至少 6 页（英文、双栏）的小论文
5. 撰写报告，准备展示

C.3 评分标准

源代码和训练模型 (40%): 代码清晰、注释丰富，变量、函数等命名合理。汇报的精度需要可以复现，无法复现的代码将会被惩罚。

小论文 (40%): 英语表达流利、行文逻辑清晰、技术新颖、对于图匹配等科学问题有自己的洞察。文章图文并茂，实验结果清晰、消融实验合理。我们按照计算机视觉领域 CCF-A 类会议审稿的标准评价每一篇论文。

Presentation (20%): 对实验、算法的描述清晰准确，能够准确地描述新算法、新技术。

LaTeX 模板参考：

https://cvpr2022.thecvf.com/sites/default/files/2021-10/cvpr2022-author_kit-v1_1-1.zip

代码规范参考：

https://mmlab.readthedocs.io/zh-cn/stable/community/code_style.html

参考文献：

- Read Ten Lines at One Glance: Line-Aware Semi-Autoregressive Transformer for Multi-Line Handwritten Mathematical Expression Recognition. ACM MM2023
- CoMER: Modeling Coverage for Transformer-based Handwritten Mathematical Expression Recognition. ECCV2022
- When Counting Meets HMER: Counting-Aware Network for Handwritten Mathematical Expression Recognition. ECCV2022
- A Tree-based Model with Branch Parallel Decoding for Handwritten Mathematical Expression Recognition. PR2024
- Improving Handwritten Mathematical Expression Recognition via Similar Symbol Distinguishing. TMM2023
- Syntax-Aware Network for Handwritten Mathematical Expression Recognition. CVPR2022
- TDv2: A Novel Tree-Structured Decoder for Offline Mathematical Expression Recognition. AAAI2022
- Language Model is Suitable for Correction of Handwritten Mathematical Expressions Recognition. EMNLP2023
- Semantic Graph Representation Learning for Handwritten Mathematical Expression Recognition. ICDAR2023
- Handwritten Mathematical Expression Recognition with Bidirectionally Trained Transformer. ICDAR2021