

# 毕设课题：基于强化学习的图像到标记识别

LaTeX 数学公式 OCR

---

方盛俊

# 课题背景及意义

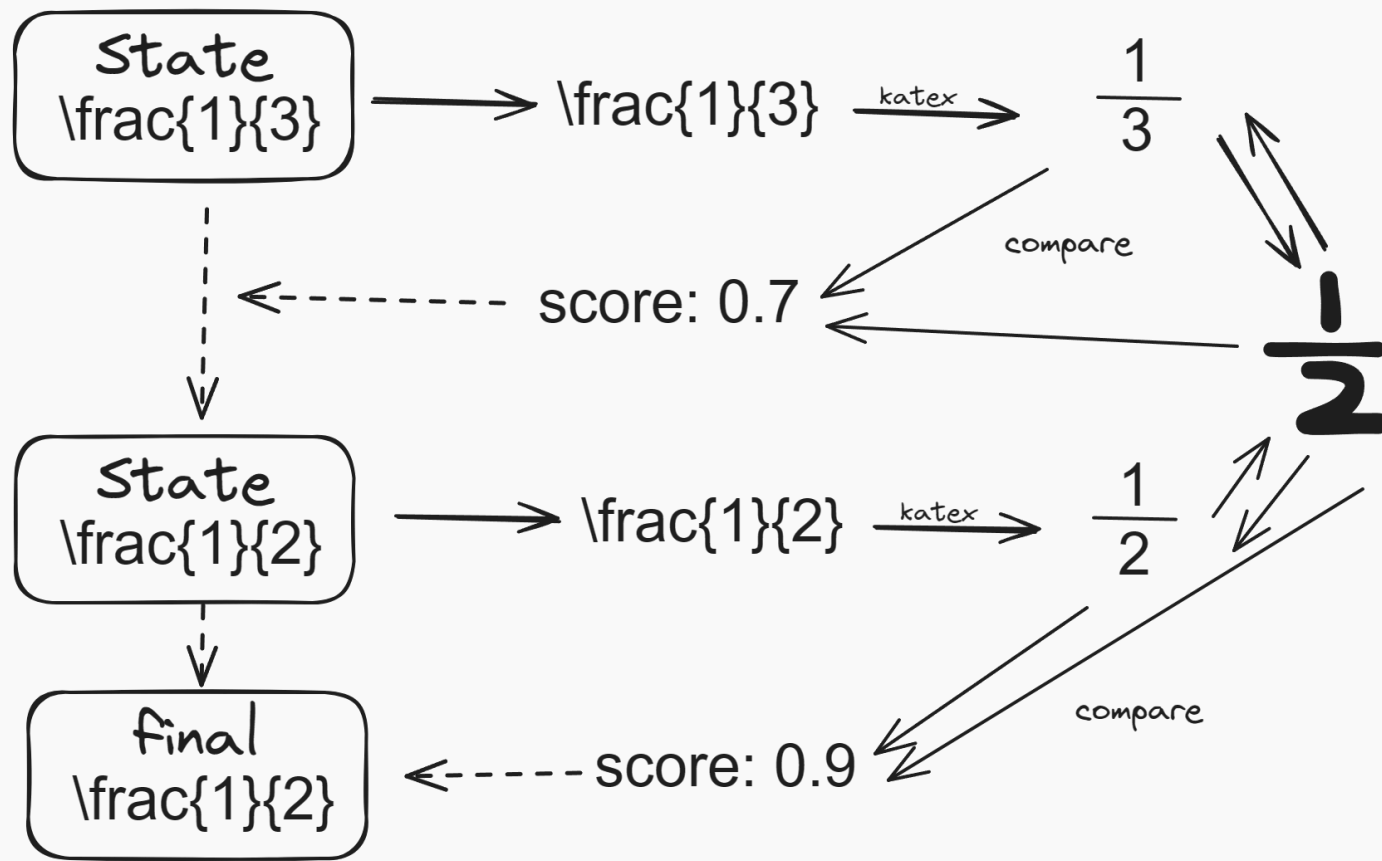
LaTeX 数学公式识别是一种在文档自动化<sup>1</sup>、笔记软件等领域都有迫切需求的 OCR 任务，有广泛的应用场景。目前的 LaTeX 公式识别基本都是基于图像识别的有监督学习<sup>2</sup>，它们只使用了已有的训练集。这种方法忽略了 LaTeX 公式识别任务本身是一种状态完全可观测的、有模型的、可以写出完美的真实环境模拟器的特殊任务的事实。我们能够使用 KaTeX 等工具实时渲染当前状态对应的数学公式图像，并与目标图像进行对比评分，进而进行实时的强化学习训练和评估。

$$\text{Tr}(\mathbf{M} + \mathbf{M}^-) = 0,$$

<sup>1</sup><https://mathpix.com/>

<sup>2</sup><https://github.com/lukas-blecher/LaTeX-OCR>

# 主要技术路线



核心在于如何训练一个评分网络，并内嵌到强化学习的训练和评估流程中，以及如何设计 MDP。

首先需要的网络和 MDP：

1. **评分网络**：能够得到两个 LaTeX 公式图像的相似度评分，**输入**是两张图片，**输出**这两张图片的相似度评分。评分网络用于辅助强化学习训练和评估的决策。可以使用已有的数据集进行有监督学习训练，也可以自主生成训练集进行训练。也许还可以使用一点 GAN 的思想。
2. **策略网络**：能够根据当前的 LaTeX 序列、当前评分与目标图像，生成一个新的 LaTeX 序列作为新状态，或者调用 KaTeX 与评分网络对当前状态进行评分，后者开销较大。
3. **State**：「目标图像特征向量」+「上一次生成图像特征向量」+「上一次评分网络评分」+「上一次序列」+「当前序列」
4. **Action**：
  - a. 生成一个新的 LaTeX 序列 / 对 LaTeX 语法树进行增删改，Reward (-1).
  - b. 调用 KaTeX 和评分网络对当前状态评分，Reward (-1000, 待定，即开销较大).

5. **终止**：终止条件可以是 **最大步数**、**最大可用时间** 或 **评分网络的评分阈值**。评估时终止，当前状态作为最后输出。如果是训练状态，则使用训练集真实序列与最后生成的 LaTeX 序列进行打分（例如 BLEU 文本相似度评分）作为 Reward，并使用策略网络梯度下降更新网络参数。
6. **优点**：由于强化学习是用状态转换进行渐进式地对当前状态进行增强，所以可能还可以支持**渐进式 OCR**，理论上只需要更改当前状态的「目标图像特征向量」即可，这点对手写笔记软件可能有很大帮助。

$$\mathcal{X} \xRightarrow{\text{ocr}} X \xRightarrow{\text{write}} X + y \xRightarrow{\text{ocr}} X + Y \xRightarrow{\text{write}} \frac{X + Y}{Z} \xRightarrow{\text{ocr}} \frac{X + Y}{Z}$$

1. 基于 Attention 的图像识别，提出了基于图像识别技术的 LaTeX 数学公式 OCR 方法 Image-to-Markup Generation with Coarse-to-Fine Attention<sup>3</sup>
2. AlphaGo Zero 不再需要从人类棋谱中学习，而是通过自我对弈的方式，模仿学习 MCTS。这可以视为一种有模型学习，其依据是基于当前策略网络的 MCTS 一般总是优于当前策略网络，因此可以不断地自我对弈强化学习。
3. 基于强化学习的自动生成 SQL 语句，其没有简单地模仿人类专家，而是在真实数据库实际执行 SQL 语句进行验证 Seq2SQL: Generating Structured Queries from Natural Language using Reinforcement Learning<sup>4</sup>

---

<sup>3</sup><https://arxiv.org/abs/1609.04938v2>

<sup>4</sup><https://arxiv.org/abs/1709.00103>

# 需要解决的问题

1. **动作空间较为庞大与复杂**：根据当前 LaTeX 序列，生成一个完整的新序列，这是一个 seq2seq 的结构，动作空间较大，难以进行训练。是否能想办法在神经网络表示一个 **LaTeX 语法树结构** 作为状态，这样动作空间就不再是生成完整序列，而是一个树操作：新增叶节点、删除某个分支或新增某个分支，这样也可能和树搜索算法进行结合。
2. **怎么融合评分网络和策略网络**：能否在训练过程中让评分网络和策略网络一起训练，进而增强两者？评分网络是否要拟合文本序列相似度指标，例如 BLEU？
3. **使用什么强化学习方法进行训练**：策略学习，使用策略梯度方法。
4. **怎么处理字号、字体和手写体**：考虑如何增强对字号大小变化的稳健性；字体和手写体需要想办法生成更多的训练数据，尤其是生成手写数学公式数据的方法。
5. **如何进行预训练**：为了减少拟合所需的时间，需要进行预训练。也许可以考虑随意删除 LaTeX 序列内部的部分文本，生成一批半完整的训练集，先进行有监督学习。
6. **是否能迁移到其他任务**：可能迁移到的任务有流程图生成、Logo 生成、SVG 矢量图像生成、前端 UI 代码生成、PDF 文件识别等，这些都是比较典型的识别图像到标记任务。