

# Contents

<b>Q1. katago 项目中有没有存储历史经典对局？是否支持存储历史经典对局，并回放分析。一盘过往下过的棋局以什么格式存储？</b>	<b>1</b>
1) KataGo 项目里有没有自带“历史经典对局库”？	1
2) 是否支持存储历史对局，并回放分析？	1
3) 一盘棋通常以什么格式存储？	2
复盘分析结果怎么“存”？	2
<b>Q2. katrain 这个开源项目是否可以商用？</b>	<b>2</b>
1) KaTrain 本体：MIT License（允许商用）	2
2) KaTrain 常见会一起用到的组件，也基本允许商用	2
KataGo 引擎	2
KataGo 神经网络权重（你用于分析/下棋的模型文件）	2
3) 你在“商用”时最容易踩坑的点	2
<b>Q3. 对于一张经典棋谱的照片（里面是黑白棋完整的位置和顺序编号），我们是否可以把它转化成.sgf 格式？</b>	<b>3</b>
能不能做到、难点在哪	3
可行的实现路线（推荐）	3
路线 A：用现成工具/半自动（最快）	3
路线 B：自己做一个稳定的“照片 → SGF”流水线（最可靠）	3
你给我一张图，我能帮你到什么程度？	4
快速判断：什么照片最容易成功？	4
<b>Q4. 有没有开源项目收集了过往经典围棋对局？如果是像.sgf 这种标准格式最好。</b>	<b>4</b>
A) 可自由批量下载的 SGF 大集合（不保证都是经典名局）	4
1) OGS (Online-Go) 全站历史对局 SGF dump（社区整理）	4
2) KGS Archives: 按用户/日期下载 SGF（官方入口）	4
3) Kombilo 收集的 KGS 高段棋谱库（第三方资源）	5
B) GitHub 上“开源仓库里直接放 SGF”的（通常规模较小，但很干净）	5
1) GoGameGuru 的开源 SGF 集合（题目/部分棋谱资源）	5
2) 用 GitHub topic 直接“搜 SGF/Baduk/Weiqi 仓库”	5
C) 职业棋谱很全，但许可不一定开源/可商用的（给你备选）	5
1) PAGE（职业棋谱研究数据集）	5
2) GoGoDOnline（职业名局大库，但商业产品）	5
我建议你怎么选	5

## Q1. katago 项目中有没有存储历史经典对局？是否支持存储历史经典对局，并回放分析。一盘过往下过的棋局以什么格式存储？

整体结论：KataGo 本身更像“引擎”，不负责替你“收藏/管理经典棋谱库”，但完全支持你把历史对局保存下来、导入回放、并用 KataGo 做复盘分析。

### 1) KataGo 项目里有没有自带“历史经典对局库”？

一般没有。官方发布的 KataGo 核心就是引擎（命令行/GTP/分析引擎），不自带 GUI，也不自带“经典棋谱数据库”这类内容。（[Hugging Face](#)）所以如果你想看“经典对局”，通常是：

- 你自己收集/下载的棋谱（职业棋谱库、网站下载等）
- 或者 GUI/第三方软件自带/提供的棋谱资源（这不属于 KataGo 仓库本体）

### 2) 是否支持存储历史对局，并回放分析？

支持，但方式是“引擎 + 棋谱文件 + GUI/工具链”。

常见 workflow：

- 对局保存为 **SGF** 文件 (.sgf)
- 用 GUI (如 Lizzie / KaTrain / Sabaki 等) 打开 SGF 回放
- GUI 调用 KataGo 引擎, 对每一步给胜率、目差、推荐变化等, 完成复盘

另外, 如果你要批量分析很多盘棋, KataGo 还提供 **JSON 分析引擎**用于高效批处理 (更偏开发/工具侧)。(GitHub)

### 3) 一盘棋通常以什么格式存储?

围棋最主流的棋谱存储格式是 **SGF (Smart Game Format)**:

- 文本格式、树结构
- 支持主线、变化、注释、标记等 (也很适合“复盘标注”) (Wikipedia)

复盘分析结果怎么“存”?

常见有三种:

1. 只存原始 **SGF** (最通用): 棋谱回放时再现场跑 KataGo 得到分析结果
2. 把分析结果写回 **SGF** (非标准扩展): 一些脚本/工具会把 KataGo 的推荐变化、评估等写进 SGF 的自定义属性里, 生成 “analyzed.sgf”之类文件 (GitHub)
3. 另存为 **JSON/JSONL** (更适合程序处理/统计): 例如用封装脚本批量跑 SGF, 输出一行一局/一行一节点的 JSONL 结果 (GitHub)

---

如果你告诉我你现在用的是 **Lizzie / KaTrain / Sabaki / 纯命令行**哪种方式, 我可以按你的环境给一套“保存棋谱 + 回放 + 复盘结果落盘”的最省事方案 (含你该选 SGF / analyzed-SGF / JSONL 哪种更合适)。

## Q2. katrain 这个开源项目是否可以商用?

可以商用, 但要分清“KaTrain 本体代码”和“它打包/下载的 KataGo 引擎与神经网络权重”的授权边界。

### 1) KaTrain 本体: MIT License (允许商用)

KaTrain 仓库在 GitHub 上标注的许可证是 **MIT**, MIT 许可本身允许商业使用、修改、再发布 (包括闭源再发布), 通常只要求你在再发布时保留版权声明和许可文本, 并接受“无担保”。(GitHub)

### 2) KaTrain 常见会一起用到的组件, 也基本允许商用

**KataGo 引擎**

KaTrain 是基于 KataGo 来做分析/对弈的, 官方 KataGo 项目也属于开源可用范畴, 并且很多 GUI (含 KaTrain) 会把 KataGo 打包或引导用户下载使用。(GitHub) (如果你计划把 KataGo 一起打包进你的商业产品里, 务必把 KataGo 的许可证文本也一并遵守、附带。)

**KataGo 神经网络权重 (你用于分析/下棋的模型文件)**

KataGo 的官方训练站点对网络权重给出了许可条款, 文本里明确包含“use / copy / modify / merge / publish / distribute / sublicense / sell copies”等权限 (带条件), 也就是允许商业使用/分发的方向。(katagotraining.org) 另外 KaTrain 自身说明它会预置/可下载模型与引擎版本 (例如“预打包一个 KataGo(OpenCL) 与较老的 15-block 模型, 并可在设置里下载模型/引擎版本”)。(GitHub)

### 3) 你在“商用”时最容易踩坑的点

- 如果你只是内部使用/提供服务 (SaaS): 通常只要你合法获取并使用软件即可; 但若涉及向客户交付可执行文件/安装包, 建议把相关许可证文本一并附上。
- 如果你要“再分发”KaTrain 的安装包或二次开发版本:

- 至少要附带 **KaTrain** 的 **MIT License** (以及你打包进去的 KataGo、网络权重、以及其他依赖的许可证声明)。(GitHub)
- 注意 KaTrain 依赖的第三方库/资源 (UI、字体、图标等) 可能各自有许可证——商业分发时要一并合规 (通常是保留 NOTICE/License 文件)。

我不是律师；如果你要做“付费软件/硬件产品内置/大规模分发”，建议你把你准备打包的清单 (KaTrain 版本、是否内置 KataGo、使用哪个网络权重、是否带 GUI 资源) 列出来，我可以帮你做一份“需要附带哪些 License/Notice”的合规 checklist。

katagotraining.org - Neural Network License ”4: <https://github.com/sanderland/katrain> “GitHub - sanderland/katrain: Improve your Baduk skills by training with KataGo!”

### Q3. 对于一张经典棋谱的照片 (里面是黑白棋完整的位置和顺序编号)，我们是否可以把它转化成.sgf 格式?

可以，而且在工程上是可落地的：把“棋谱照片 (棋子位置 + 编号顺序)”转成 **SGF**，本质是一个图像到结构化棋谱的问题。

#### 能不能做到、难点在哪

- 能做到：只要照片里每个棋子上都有清晰的手数编号 (1、2、3...)，就可以恢复落子顺序，最终生成 .sgf。
- 主要难点：
  1. 棋盘网格定位 ( $19 \times 19$  或  $13 \times 13/9 \times 9$ )：需要识别棋盘四角/边框并做透视矫正
  2. 黑白棋子检测：识别每个交叉点是否有黑子/白子
  3. 编号识别 + 排序：从每个棋子上的数字读出手数，然后按手数排序生成落子序列
  4. 特殊情况：照片模糊、反光、编号被遮挡、同一格有多种标记 (圈、三角、注释)、死子标记、局面图不是“从第 1 手开始”而是“某个中盘局面”

#### 可行的实现路线 (推荐)

##### 路线 A：用现成工具/半自动 (最快)

很多围棋 GUI/工具社区里有“从图片/截图录入棋谱”的功能或插件思路，但效果强依赖图片质量，且不少需要你手工校正几处识别错误。如果你追求“可用就行”，半自动通常最省时间。

##### 路线 B：自己做一个稳定的“照片 → SGF”流水线 (最可靠)

一个实战可用的 pipeline:

1. 预处理
  - 旋转校正、裁剪棋盘区域、透视矫正 (把斜拍棋盘拉成正方形)
  - 提升对比度、去噪、锐化、二值化 (提高数字可读性)
2. 棋盘交叉点建模
  - 检测棋盘外框四角 → 透视变换
  - 在归一化后的正方形上按 19 等分采样出  $19 \times 19$  交叉点坐标
3. 棋子检测 (黑/白)
  - 在每个交叉点附近取一个小窗口，判断是否存在棋子 (圆形/亮度特征/边缘特征)
  - 判别黑白：亮度直方图 + 阈值、或一个小 CNN 分类器
4. 编号识别 (关键)
  - 对检测到的棋子区域做“数字区域分割”
  - OCR 识别数字 (1~361)
  - 更靠谱的做法：专门训练一个“棋子编号识别”小模型 (因为普通 OCR 对小号、弧面、反光很容易翻车)

## 5. 手数排序 → 生成落子序列

- 按编号排序得到 (move\_no, color, coord)
- 坐标转 SGF: 19 路坐标用 aa..ss (跳过 i? 注意: SGF 坐标不是跳 i 的, 字母连续; 但不同工具内部坐标要映射好)
- 输出 SGF: (;GM[1]FF[4]SZ[19]PB[...]PW[...]KM[...]B[pd];W[dd]...)

## 6. 一致性校验

- 颜色交替校验 (如果照片里编号只写在一色或有断点, 就要容错)
- 是否有重复编号、缺号
- 落子点是否冲突
- 可能需要手工修正少数识别错的数字

## 你给我一张图, 我能帮你到什么程度?

在这个聊天里, 如果你上传那张经典棋谱照片:

- 我可以先帮你判断它属于“可自动化程度高/需要手工校正多”的哪一类
- 并给你一套更具体的实现建议: 用什么检测方法、OCR 怎么做、怎么做容错与人工校正界面
- 如果你愿意走“工程落地”, 我也可以给你一个 **Python 脚本骨架** (OpenCV + 识别模块 + SGF 输出), 你在本地跑

## 快速判断: 什么照片最容易成功?

- 棋盘基本正视、畸变小
- 编号清晰、对比强、无遮挡
- 每个编号只出现一次, 且从 1 连到 N
- 没有大量手写注释/圈点覆盖棋子

---

如果你方便, 直接把那张棋谱照片上传。我会按你的图给一个“最短路径方案”: 是走半自动工具更快, 还是值得做一套自动识别流水线, 以及大概会卡在哪些地方。

## Q4. 有没有开源项目收集了过往经典围棋对局? 如果是像.sgf 这种标准格式最好。

有, 但要先把现实说清楚: “经典职业名局 (尤其含注释、官方赛事)” 的大规模 **SGF** 全量库, 很多都牵涉版权/整理成本, 所以真正又“大又全又完全开源许可”的并不多; 更常见的是:

- 开源/可自由下载的 **SGF** 大集合 (但主要是网络对局, 不一定“经典名局”)
- 小而精的开源 **SGF** 集合 (名局/讲解/题目)
- 商业或非商用许可的数据集 (职业棋谱很全, 但不一定能商用)

下面按你“最好是标准.sgf”的偏好给你几个靠谱入口。

### A) 可自由批量下载的 SGF 大集合 (不保证都是经典名局)

#### 1) OGS (Online-Go) 全站历史对局 SGF dump (社区整理)

OGS 论坛里有人把 **OGS 截至 2021** 的对局整理成可下载的 **SGF/JSON**, 并放到 Internet Archive (压缩后体量很大)。(Online Go Forum) 适合: 你要很多棋谱做训练/统计/检索; 缺点: 主要是线上对局 (职业名局占比未知)。

#### 2) KGS Archives: 按用户/日期下载 SGF (官方入口)

KGS 有官方的 Archives 页面, 可以查询并下载用户对局记录。(Go KGS) 另外 KGS 的 SGF 文件 URL 结构也很固定 (很多工具/脚本就是用这个规则批量拉)。(GitHub) 适合: 你想要高段玩家、特定账号历史棋谱; 缺点: 需要你“知道要抓谁”, 不是“经典名局库”。

### 3) Kombilo 收集的 KGS 高段棋谱库 (第三方资源)

Drago 的链接页提到: Kombilo 网站有一个 KGS 6d/7d 大量对局 SGF 集合。(Godrago) 适合: 想要“质量更高的人类对局”做研究/复盘库 (依旧不是“职业经典名局”定位)。

---

## B) GitHub 上“开源仓库里直接放 SGF”的 (通常规模较小, 但很干净)

### 1) GoGameGuru 的开源 SGF 集合 (题目/部分棋谱资源)

比如 gogameguru/go-problems 提供可下载的 SGF 题目集合, 仓库公开可拉取。(GitHub) 以及他们组织下还有其它与 SGF 相关的公开仓库入口。(GitHub) 适合: 你要“可复用、许可清晰、内容质量高”的 SGF (但不是海量经典对局)。

### 2) 用 GitHub topic 直接“搜 SGF/Baduk/Weiqi 仓库”

GitHub 的 baduk topic 可以快速浏览很多公开仓库 (有人会放名局 SGF、小型棋谱集、脚本爬虫等)。(GitHub) 适合: 你想找“开源 + 直接可 clone 的 SGF 资源”, 自己筛。

---

## C) 职业棋谱很全, 但许可不一定开源/可商用的 (给你备选)

### 1) PAGE (职业棋谱研究数据集)

PAGE 项目声称覆盖 1950–2021 的 98,525 盘职业对局, 但它的说明里明确提到数据集采用 CC BY-NC-SA 4.0 (非商用)。(GitHub) 适合: 研究/学习用途 (要商用就不合适)。

### 2) GoGoDOnline (职业名局大库, 但商业产品)

GoGoDOnline 提供下载, 包含大量 独立 SGF 文件; 但这是商业服务, 不是开源项目。(GoGoD)

---

## 我建议你怎么选

- 你要“经典名局 (职业) + SGF”: 现实里多半得靠 GoGoD/Go4Go/Waltheri/各种棋谱站 (不少是商业/半商业), 或者自己从公开站点逐步收集。
- 你要“开源许可清晰 + SGF”: 优先从 GitHub 上那些“直接放 SGF 文件”的小集合入手 (质量高、版权风险小), 再补 OGS/KGS 这种“海量对局库”。
- 你要“可商用”: 务必逐个看许可; PAGE 这种明确 NC 的就直接排除。(GitHub)

如果你告诉我你更偏向:

1. 职业经典名局 还是 高段网络实战, 以及
2. 是否需要 可商用, 我可以把上面这几类再精确收敛成“3 个最合适的来源 + 获取方式 (批量下载/按选手/按赛事)”。