

基于虚拟染色的免疫组化图像生成

半决赛注意事项

一、比赛任务与适用范围

本赛道以单张 DAPI 图像为图像输入，生成同一视野、同一坐标下指定 marker 的图像。半决赛包含 colon（结肠）、liver（肝脏）、stomach（胃）三个数据集，每个数据集至少提交一个 marker。

术语说明：

- **DAPI**：显示细胞核的荧光染色，本赛道的输入图像。
- **Marker（标志物）**：生成目标，限定为 CD68、CD45RO、HLA-DR、Vimentin。
- **Train、Val、Test**：训练集、验证集、测试集，分别用于训练、选择方案和最终评价。
- **GT**：真实参考染色图像。
- **Checkpoint**：保存模型权重及必要固定状态的文件。
- **Seed（随机种子）**：控制随机过程的初始设置。
- **TTA（测试时增强）**：对同一评价图像进行变换、多次预测，再合并或选择结果。

本文“评价阶段”指用于选择模型或参数的验证集评价，以及正式测试/推理。验证成绩须按正式测试的输入、生成和保存要求计算，各候选分别计分。单模型、单张图像输入、单目标输出、禁止预测集成和后处理的要求适用于评价阶段。训练阶段允许数据增强、候选模型比较及合规的知识蒸馏，训练批量不要求为 1。各条适用阶段在标题中标明。

二、评分与违规处理

1. 图像评分

采用 SSIM（结构相似性）和 PSNR（峰值信噪比）评分，权重分别为 70% 和 30%。SSIM 限制在 0~1，PSNR 除以 50 后限制在 0~1；两项按权重加权后乘 100，得到单图分。

先平均同一 marker 的全部规定图像分，再平均该数据集已提交的 marker 分。未提交的 marker 不计零分、不参与平均、不扣分。

基础总分 = $10\% \times \text{colon 分} + 20\% \times \text{liver 分} + 70\% \times \text{stomach 分}$ 。

2. 违规处理类别

A 类【实质违规】：该队半决赛最终成绩记 0 分。

B 类【本次提交无效】：该次提交记 0 分，不影响其他有效提交；截止前可按平台规定重新提交。

C 类【轻度规范扣分】：根据违规的性质、影响范围及严重程度，每项扣 0.5~5 分，具体情形和分值见第四部分。C 类情形本身不使提交无效；违反不同 C 类情形的，分别扣分并累计，不设 C 类累计扣分上限。

最终成绩 = 最高有效基础成绩 - C 类累计扣分，最低为 0 分；发生 A 类违规时，最终成绩直接记 0 分。同一事实涉及多条规则的，只按最重一项处理。不同 C 类情形按第四部分的编号分别累计，同一编号不按图片、marker、文件数量或提交次数重复计扣。

任一已选 marker 出现导致提交无效的问题，整次提交均无效，不对剩余图片、marker 或数据集重新计分。参赛队须在提交前检查文件完整性；缺少文件、提交错误，后果自负。官方数据或平台自身问题不计为选手违规。

三、参赛规则

R01【全流程 | 数据使用范围】

仅可使用官方指定为训练用途的数据及明确授权可继续训练的历史数据。不得引入其他 **DAPI**、**marker**、标签或外部样本参与训练、微调或验证。公开预训练权重按 **R04** 执行。

违规处理：A 类。

R02 【全流程 | 数据划分】

不得改变官方 **Train/Test** 划分。可从 **Train** 中划分 **Val**；有官方患者、切片等同源分组时，应按组划分。

Val 不得参与对应候选模型或其教师模型的训练。使用多个教师时，须保证该 **Val** 未被其中任何教师用于训练。教师内部五折划分不自动构成学生模型的独立 **Val**。不得将 **Test** 及其衍生信息用于训练或模型选择。

违规处理：A 类。

R03 【训练阶段 | 数据增强与标签保护】

允许训练数据增强。旋转、翻转、平移、裁剪、缩放等几何变换须对配对 **DAPI** 和 **marker** 同步进行。亮度、对比度、噪声等非几何增强仅用于 **DAPI**。

不得通过人工或自动处理增删 **marker** 阳性区域或改变表达标签；配对几何变换的必要插值及固定数值表示转换除外。

违规处理：A 类。

R04 【训练阶段 | 预训练权重】

允许使用提交开放前已公开发布的预训练结构或权重作为初始化（需要说明）。不得额外使用其预训练数据继续训练，不得使用来自官方 **Test GT** 的权重。

违规处理：A 类。

R05 【评价阶段 | 单模型与训练蒸馏】

每个候选模型的验证成绩须由该候选的单一模型、单一 **checkpoint** 独立生成。正式测试时，三个数据集及所选 **marker** 须共用一套固定模型结构和一个最终 **checkpoint**。不得按器官、样本或 **marker** 切换独立模型，不得将多个独立模型打包后用于生成。

允许同一模型共享特征提取部分，并按目标 **marker** 选择输出头，每次只生成一张目标图像。

训练阶段允许知识蒸馏，即由教师模型提供辅助目标训练学生模型。可使用多个教师或五折交叉验证所得教师，并将其预测平均后用于训练数据上的辅助监督。五折交叉验证指将训练数据分成五组，轮换训练与验证。

教师和学生均须遵守数据、标签和 **seed** 要求。教师预测不得替换真实 **GT**，不得用 **Test** 图像或教师对 **Test** 的预测训练学生。使用蒸馏的方案在评价阶段只运行学生模型，教师不得参与生成、合并或择优。评价阶段不得下载额外模型或调用外部生成服务。

违规处理：评价阶段使用多模型、多 **checkpoint** 或教师参与生成，**A** 类。

R06 【评价阶段 | 单张 DAPI 输入】

每次仅输入一张完整的官方 **DAPI**，图像批量为 1；训练批量由队伍自行设置。不得附加其他 **DAPI**、真实或生成 **marker**、参考图像、标注、患者资料等输入。

允许以 **marker** 名称、固定编号（**marker_id**，如 0~3）、独热编码（对应类别位置为 1 的编码）或 **Embedding**（类别的数字向量表示）指定目标。条件表示只能包含目标 **marker** 类别，同一类别须使用固定表示，不得携带额外样本、器官、患者或 **Test GT** 信息。类别向量可在合法训练中学习，须属于同一最终模型并在评价前固定。

允许使用“**CD68**”等类别标识，不允许输入 **CD68** 真实染色图像。模型内部随机量按 **R12** 执行。

违规处理：使用不允许的额外输入，或评价图像批量不为 1，**A** 类。

R07【评价阶段 | 单张目标图像输出】

每个“样本+marker”为一次生成任务，只生成一张目标图像。同一 DAPI 可分别执行不同 marker 任务。

禁止一次生成多个 marker 后选取其中一个；禁止同一任务生成多个候选后平均、投票或择优；禁止先生成其他 marker 再辅助目标输出。允许固定模型内部的多步计算，但每项任务只能执行一条生成路径，不得重复采样择优。

违规处理：A 类。

R08【评价阶段 | 输入处理】

模型外仅允许读图、固定通道或数值表示转换、数据装载，以及使用 Train 数据或理论像素范围确定常数的归一化。

禁止对原始评价图像额外缩放、裁剪、切块、补边、拼接、旋转、翻转、配准、去噪或筛选区域；不得按单张或整套评价图像重新估计输入转换参数。

固定网络内部的正常特征运算及特征归一化可以使用，但不得更新跨样本持久状态。将禁用的输入处理写入模型函数，同样违规。训练数据处理按 R03 执行。

违规处理：A 类。

R09【评价阶段 | 禁止增强及集成】

禁止 TTA、多尺度或多裁剪测试、多模型、多 checkpoint、多 seed 结果平均，以及同一任务多次预测合并或择优。提前离线生成这些结果再提交，同样违规。复赛允许的 TTA 不适用于半决赛评价阶段；训练蒸馏按 R05 执行。

违规处理：A 类。

R10【评价阶段 | 模型固定】

须使用模型评估模式，不得训练、微调、自适应更新，或更新 BN（批归一化）运行统计量等持久状态。不得利用此前评价样本改变后续样本的生成方案。

违规处理：A 类。

R11【分阶段 | 训练辅助处理与预测后处理】

训练阶段允许对预测进行辅助处理，用于计算损失或辅助学习，不得违规修改真实 GT。

评价阶段禁止对生成结果进行额外修正。输出后仅允许必要的固定数值还原、合法范围限制、8 位整数转换和编码保存。禁止调色、亮度或对比度调整、阈值化、去噪、平滑、锐化、区域修正、人工修图及多次处理择优。

数值还原和保存方法须依据 Train/Val 预先确定，不得按当前评价图像的像素统计逐图调整。不得将训练辅助处理用于修正评价输出，或通过改名、封装进模型使用禁用处理。

违规处理：评价阶段使用非保存必需的后处理，A 类。

R12【全流程 | 随机种子】

涉及随机过程时，训练和评价的根 seed 均须为 2026，每个“样本+marker”评价任务从 seed=2026 开始。训练并行随机流可由 2026 按确定规则派生。

不得使用其他 seed 的候选参与选择；无随机过程时无需增加随机操作。

违规处理：仅 seed 设置或执行不符，按 C04 分级扣分：训练阶段 2 分，验证阶段 3 分，正式测试/推理阶段 5 分。评价阶段存在多 seed 结果合并或择优的，按 R09 的 A 类处理，同一事实不重复扣分。

R13【全流程 | 禁止 Test 调参】

模型、checkpoint、marker 子集、输入转换、保存及采样方案，须依据 Train/Val 确定，并在首次对正式 Test 执行模型推理前固定。

不得利用 Test GT、Test 统计信息、测试生成效果或正式排行榜反馈选择、调整方案。正式 Test 结果尚未提交评分，也不得用于调参。允许清点文件、检查公布的格式规格和整理目录，但不得改变生成方案。

违规处理：利用正式 Test 信息调参或选择输出，A 类。

R14 【结果提交 | 数据集与 marker 完整性】

必须同时提交 colon、liver、stomach 三个数据集，各自从四个指定 marker 中选择 1~4 个。三个数据集可选择不同子集，须按 R13 在 Test 推理前固定，不得依据 Test 效果或得分删减 marker。

每个已选 marker 须交齐官方输出清单中的全部预测图；未选 marker 不建空目录。

违规处理：缺少数据集、没有可评分 marker 或已选目录为空，B 类。未选 marker 不扣分。

R15 【结果提交 | 目录结构】

ZIP 解压后的第一层必须是 results，路径须为：

results/colon/test/<marker>/<预测图像>

results/liver/test/<marker>/<预测图像>

results/stomach/test/<marker>/<预测图像>

marker 目录名须按 CD68、CD45RO、HLA-DR、Vimentin 书写，大小写和连字符保持一致。不得增加外套目录、省略 results、分散存放三个数据集或提供多套结果。

示例（每个已选 marker 须包含全部规定图片）：

results/

colon/test/CD68/ROI000_00_00_fake.png

liver/test/CD45RO/ROI000_00_00_fake.png

stomach/test/Vimentin/ROI000_00_00_fake.png

违规处理：仅有额外一层包装目录且符合 C02 的，按 C02 扣分；仅 marker 目录书写不规范且符合 C03 的，按 C03 扣分；其他不符合本条的情况，B 类。

R16 【结果提交 | 文件命名与对应】

样本标识、大小写及相对路径须与官方输出清单一致。“原主文件名”指清单中不含扩展名的样本文件名。

文件名可用“原主文件名_fake”或“原主文件名”，后接 R17 允许的扩展名，_fake 使用小写。每个样本在每个已选 marker 下只能交一个文件，不得因命名或扩展名不同重复提交。

不得漏图、加图、错配或增加清单外目录，R15 列明的轻微目录违规按 C02、C03 处理。结果包只放预测图和必要目录，不放代码、模型等其他文件；系统文件.DS_Store、Thumbs.db、desktop.ini 可忽略。

违规处理：仅夹带普通说明文本且符合 C01 的，扣 0.5 分；预测文件缺失、多余、重复、命名或对应不符，以及其他夹带文件，B 类；伪造结果按 R19 处理。

R17 【结果提交 | 图像格式】

须为单帧、8-bit RGB 图像，即三个颜色通道、每通道整数值 0~255，格式限.png、.jpg、.jpeg。扩展名大小写不影响格式判定，文件名主体按 R16 对应。

尺寸须与官方输出清单一致，保持原视野、方向和坐标，不得额外裁剪、插值或拉伸。不得提交灰度、透明通道、16 位或多帧图像。

按提交文件解码后的像素评分，保存方式须事先固定。

违规处理：损坏、不可读取或格式、尺寸、空间对应不符，B 类；违规修正或 Test 调参按相应 A 类条款处理。

R18 【结果提交 | 压缩包要求】

须提交完整、可解压、未加密的 ZIP，不得包含越界路径（如../）、链接或重复压缩成员，不得恶意干扰评分服务。

提交上限：压缩包 2048 MiB、解压后 8192 MiB、成员总数 20000（含目录）、单文件 200 MiB。1 MiB=1,048,576 字节，上限包含等于该值的情况；如有变更，以提交开放前的官方公告为准。

违规处理：超过公布上限或压缩包本身不合法、无法处理，B 类；恶意干扰比赛，A 类。

R19 【评价结果及提交 | 结果真实性】

预测须由本队固定的最终模型按合规流程产生。禁止复制、拼贴或检索替换 Test GT，串用他队结果，按样本名查表返回预存答案，人工修图或用其他图片替代。不得依据评价反馈选择候选图。

违规处理：A 类。

四、C 类违规情形及扣分标准

C01 【夹带普通说明文件 | 扣 0.5 分】

压缩包根目录或 results 目录中夹带 README.txt 等与评分无关的普通说明文本，全部预测图片仍完整、可直接匹配和评分。

本项不包括 marker 预测目录内的辅助文件，也不包括额外图片、模型、代码或违禁数据。.DS_Store、Thumbs.db、desktop.ini 三种已列明的系统文件不扣分。

C02 【额外一层包装目录 | 扣 1 分】

仅在 results 外增加了一层包装目录，例如 submit/results/colon/test/...；包装目录内的 results 及三个数据集完整，结果路径唯一，其余结构符合 R15。

省略 results、数据集分散、多套结果或其他不能唯一对应的结构，不适用本项，按 B 类处理。

C03 【marker 目录名称书写不规范 | 扣 1 分】

marker 目录仅有大小写、连字符或下划线的书写偏差，仍能唯一对应 CD68、CD45RO、HLA-DR、Vimentin 中的一个，且全部预测完整、无重复或冲突。

例如 cd68 应写为 CD68，HLA_DR 应写为 HLA-DR。未知 marker、同一 marker 出现多个对应目录、样本文件名错误或对应不清，不适用本项，按 B 类处理。

C04 【随机种子不符合要求 | 按阶段扣 2~5 分】

训练阶段 seed 不符合 2026 要求：扣 2 分。

验证阶段 seed 或逐任务初始化不符合 2026 要求：扣 3 分。

正式测试/推理阶段 seed 或逐任务初始化不符合 2026 要求：扣 5 分。

同一 seed 问题涉及多个阶段时，按其中最高一档扣分，不将本项各档相加。评价阶段多 seed 预测合并或择优不属于轻度 seed 设置问题，按 R09 的 A 类处理。

以上 C 类只适用于各项明确列出的情形。不同编号对应的独立违规分别累计；出现 A 类或导致提交无效的 B 类问题时，按相应条款处理。