半决赛评测规则与模型使用公告

为统一半决赛评测标准，现将评分方式及模型使用规定公告如下。

一、评测图片

半决赛评测图片编号为 test_169 至 test_318，共 150 张。

二、评分规则

总分由实例分割得分和铁素体平均面积得分组成，各占 50 分：

> *总分 = 实例分割得分 × 50 + 铁素体面积得分 × 50*

（一）实例分割得分

对每张图片中的每个 GT 实例，评测程序在尚未匹配的预测实例中，寻找类别一致且 IoU 最高的实例。预测实例最多只能匹配一个 GT。

类别一致且 IoU 大于或等于 0.5 时，记入该 GT 的 IoU；否则该 GT 记为 0。单图得分为所有 GT 对应分数的平均值：

$$S_i=\frac{\sum_{g\in GT_i} score_g}{N_i},\quad score_g=\begin{cases} IoU_g, & \text{匹配成功} \\ 0, & \text{未匹配或类别不一致} \end{cases}$$

其中，N_i 为该图片的 GT 实例总数。最终实例分割得分为所有评测图片单图得分的算术平均值。

未匹配的预测实例不直接计入实例分割项的扣分；类别判错或漏检会使对应 GT 记为 0。铁素体面积项仍按下述规则独立计算。

（二）铁素体平均面积得分

分别计算 GT 和预测结果中铁素体实例的平均面积，并按相对误差计算得分：

$$A_i = \max\left(0, 1 - \frac{|\bar{a}_{pred} - \bar{a}_{GT}|}{\bar{a}_{GT}}\right)$$

该项按现有规则统计所有铁素体实例，不要求实例先与 GT 匹配；最终面积得分为所有评测图片单图面积得分的算术平均值。

三、模型使用规定

初赛、复赛阶段对 SAM 等规模较大的视觉分割模型作出限制，主要是为了探索传统无监督方法和通用分割算法在工业场景中的适用性，并关注工业边缘侧算力有限等实际条件。

半决赛阶段将重点探索前沿技术在工业场景中的应用效果与性能上限，因此调整模型使用规定如下：

- 计算机视觉领域的基础模型全面开放，包括 SAM、SAM2、SAM3，DINOV3 等视觉分割模型，不设置额外限制。
- 仅限制多模态大模型及大语言模型：不得调用此类模型参与参赛预测结果的生成，包括分割、分类、推理及预测结果后处理。

请各参赛队伍按照以上规则准备并提交结果。