南昌大学 NANCHANG UNIVERSITY

第十六届全国高校材料学科实验教学研讨会

金相大赛AI评分系统的研发与思考

刘越、李样生、黎鹰

2026年10月24日





南昌大学
NANCHANG UNIVERSITY

报告内容

CONTENTS

01 研发背景

02 研发过程

03 近期成效

04 未来规划



南昌大学
NANCHANG UNIVERSITY

报告内容

CONTENTS

01 研发背景

02 研发过程

03 近期成效

04 未来规划

金相 AI 评分系统解决的问题

南昌大学 NANCHANG UNIVERSITY

传统人工评分面临的三大现实痛点

赛事规模持续扩大，“浏览完单人全套 8 张试样图后快速估总分”难以兼顾阅片效率与细项判罚精度。

痛点一：易受“第一印象”主导

细项难逐一清点：8 张图/人、单图数百颗晶粒，肉眼难逐一数划痕、算面积，底色白亮易掩盖高倍细划痕。

痛点二：视觉疲劳与跨组尺度不一

给分随顺序浮动：连续数小时阅片敏感度下降；缺少固定同屏参照，给分松紧随阅片顺序浮动。

痛点三：单一总分缺乏教学反馈

无法定位失分细项：赛后仅知总分，师生无法定位具体扣分细项（划痕、假象还是磨斜失焦）。

核心痛点：几秒目测难以完成各评分项内微观细项的清点与面积折算，亟需客观稳定的定量标尺。

系统建设定位——做“算得清、看得见”的客观标尺

南昌大学 NANCHANG UNIVERSITY

将国赛 80 分制文字细则转化为全视场逐晶粒定量计算，服务赛场复核与日常教学。

1 对齐 80 分制三维度

三维分项量化：拆解为 **组织清晰度（40分）+ 划痕控制（20分）+ 表面假象（20分）**。

2 全视场逐晶粒统计

全图无死角测量：高倍百颗、低倍千颗晶粒统一测量，不漏看、不漂移。

3 可视化诊断输出

扣分有据可查：同步输出组织分割、晶界清晰度、划痕分级与假象定位图层，每一分扣得明明白白。

建设目标：以统一量化标准辅助赛场客观复核，以分项诊断支撑日常教学反馈。



南昌大学
NANCHANG UNIVERSITY

报告内容

CONTENTS

01 研发背景

02 研发过程

03 近期成效

04 未来规划

系统版本更替与多模协同演进

南昌大学 NANCHANG UNIVERSITY

围绕业务需求的四阶段研发演进

系统随大赛评分需求逐步深入，完成了从“单材料粗分档”到“三大材料高低倍精细量化”的跨越。

01 第一阶段：单材料粗分档

- **目标与材料**：粗略区分样品“好/中/差”，仅覆盖工业纯铁（500× 高倍）。
- **核心突破**：基础晶粒分割 + 全图平均清晰度。

02 第二阶段：落地 80 分制三分项

- **目标与材料**：落地国赛 80 分制三大分项，从纯铁拓展至 20 钢（高倍）。
- **核心突破**：晶粒四级划痕分类 + 高分辨率缺陷检测。

03 第三阶段：三大材料与高低倍全覆盖

- **目标与材料**：覆盖 **纯铁 + 20钢 + 球墨铸铁**（100× + 500×）。
- **核心突破**：**16类统一分割模型 + 低倍大图分块拼接。**

04 第四阶段：攻克专家关注的四大难点

- **目标与材料**：三大材料全覆盖 + 国赛典型试样实测。
- **核心突破**：**深度学习模型 + 光学特征分析协同评判。**

评分规则细化——从“阶梯跳变”到“80分制连续量化”

南昌大学 NANCHANG UNIVERSITY

告别早期“多一个瑕疵跳扣 5 分”的生硬分档，实现三大维度 12 项细则指标的连续平滑打分。

1. 组织清晰度（40分）

- **核心量化指标**：模糊晶界占比、晶界对比度与锐度、双晶界比例、分区对焦落差。
- **重点解决难点**：区分**欠腐蚀浅晶界**、**过腐蚀双晶界**与**磨斜两侧失焦**。

2. 划痕控制（20分）

- **核心量化指标**：长划痕条数、多方向交叉点数、四级划痕晶粒面积占比。
- **重点解决难点**：兼顾**跨晶粒长划痕（含交叉锁死上限）**与**晶内密集细划痕**。

3. 表面假象（20分）

- **核心量化指标**：坑洞/夹杂物个数、水渍/锈蚀/污渍面积比、铁素体灰白块占比。
- **重点解决难点**：剥离第二相暗区干扰，精准量化**灰白块严重度**与晶间反差。

规则升级：将定性文字细则转化为可解释、连续平滑的 12 项定量指标。

材料与倍率扩展——三大材料 × 高低倍评判适配

南昌大学 NANCHANG UNIVERSITY

针对不同材料的物相构成与高低倍视场差异，定制专属分割类别与干扰剥离策略。

工业纯铁(单相铁素体晶粒)

- **低倍（100×）重点**：全视场磨面平整度、大面积灰白块、粗长划痕。
- **高倍（500×）与专属算法**：**低倍大图分块拼接**分割数千微小晶粒；高倍精检**双晶界**与晶内细划痕。

20钢（铁素体 + 珠光体）

- **低倍（100×）重点**：铁素体与珠光体分布均匀性、基体灰白块、表面污染。
- **高倍（500×）与专属算法**：识别**珠光体**片层清晰度；算灰白块时**自动排除深色珠光体**防误判。

球墨铸铁（石墨球 + 铁素体 + 渗碳体）

- **低倍（100×）重点**：石墨球分布、石墨剥落假象坑洞（占33.5%）、边缘失焦。
- **高倍（500×）与专属算法**：细分 **5 种石墨球制样状态**与**渗碳体**；增设 **24 宫格两侧磨斜失焦检测。**

多材料适配：先剥离正常第二相（珠光体/石墨球），再对铁素体基体与边缘失焦做精准定量。

数据集工程——16类精细标签体系与 20 万级标注库

南昌大学 NANCHANG UNIVERSITY

将“晶粒划痕等级、多材料物相、球墨抛光状态、微观假象”解耦编码，建成大规模金相标注库。

16 类全要素高倍标签体系

- **晶粒划痕四级（4类）**：**无划痕**、**较少划痕**、**较多划痕**、**很多划痕**
- **第二相组织（2类）**：**珠光体**、**渗碳体**
- **球墨制样状态（5类）**：**抛干净石墨球**、**较干净**、**未抛干净**、**脱落石墨球**、**非常规球墨**
- **表面假象与缺陷（5类）**：**坑洞**、**夹杂物**、**水渍**、**锈迹**、**出胶物**

四大核心数据集规模（主库 + 专项派生）

- **1. 16类高倍全要素主库**：**1,782 张**高倍图，**201,095 个**多边形实例（晶粒/第二相/微缺陷）
- **2. 2560px 微缺陷专集**：从主库抽取 **1,649 张**含缺陷原图、**26,740 个**坑洞与夹杂物实例
- **3. 高低倍长划痕专集**：**2,208 张**图（高倍 757 + 低倍 1,451），**10,514 个**长划痕实例
- **4. 低倍表面假象统计集**：**1,245 份**低倍样本（发现纯铁 29.9% 含灰白块、球铁 77.1% 含细划痕）

数据基石：共标注 3,233 张高低倍金相图（211,609 个多边形实例），用于建立上述数据集。

模型与算法规划——“三专模协同 + 光学特征分析”双轨架构

南昌大学 NANCHANG UNIVERSITY

用三个不同分辨率的深度学习模型兼顾大晶粒与微缺陷，再用光学特征分析测量连续明暗渐变。

LENS 01

深度学习层（识别边界明确目标 · mAP 为综合识别精度，满分 1.0）

- **1. 组织与晶粒分割主模型（8类 2560px / 16类 2048px）**：纯铁与20钢 8类专模（珠光体精度 **mAP 0.835**）+ 球铁 16类五态球墨专模（球墨 **0.909** / 渗碳体 **0.475**）
- **2. 高分辨率微缺陷专模（2560 × 2560）**：原图高分辨率精检微小坑洞与夹杂物（微小坑洞识别精度 **mAP 0.707**）
- **3. 长划痕分割专模（高倍 1280px / 低倍 1920px）**：专职提取跨晶粒细长划痕与交叉点（细线像素级 **mAP** 高倍 **0.365** / 低倍 **0.244**）

LENS 02

光学特征计算层（测量连续明暗渐变与浅划痕）

- **1. 细浅长划痕光学特征协同检测**：结合多方向光学线特征与晶界形态区分，避免将平直晶界误判为划痕，同时完整捕获细浅抛光划痕
- **2. 晶界清晰度、双晶界与灰白块测量**：与专家盲测**名次吻合度**（满分 **1.0**）清晰度达 **0.930**、灰白块达 **0.965**；20钢/球铁先剥离第二相再算色差，**避免将珠光体/球墨误判为灰白块**
- **3. 24宫格磨面倾斜失焦检测**：对比左右边缘与中心区域清晰度落差，量化磨斜导致的边缘失焦模糊



南昌大学
NANCHANG UNIVERSITY

报告内容

CONTENTS

01 研发背景

02 研发过程

03 近期成效

04 未来规划

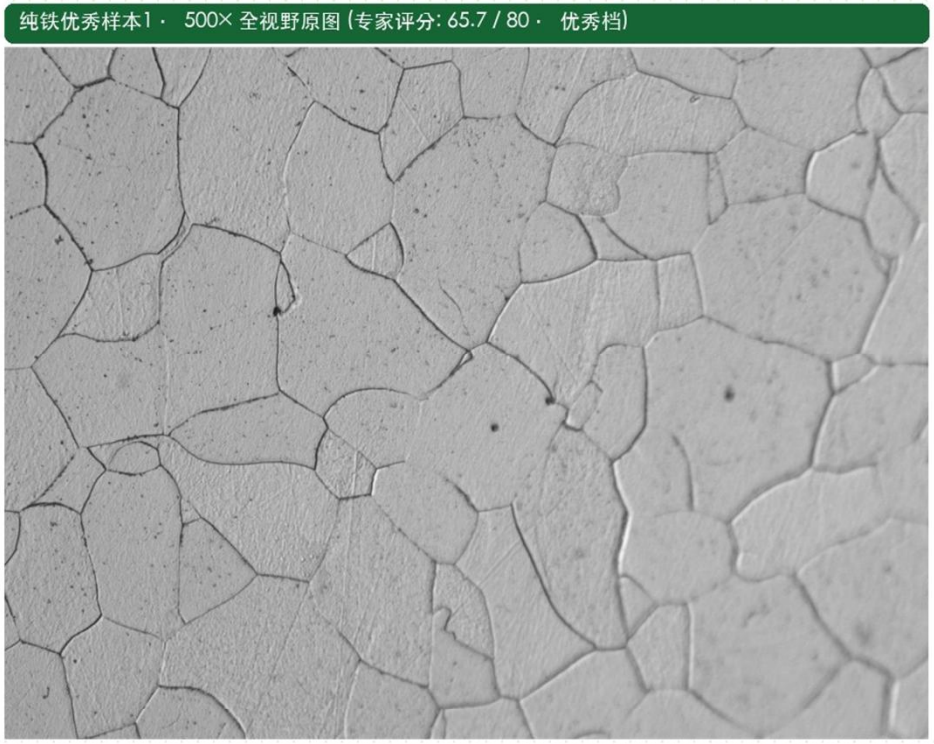
国赛实测样本总览（45套典型试样 / 360张高低倍图像）

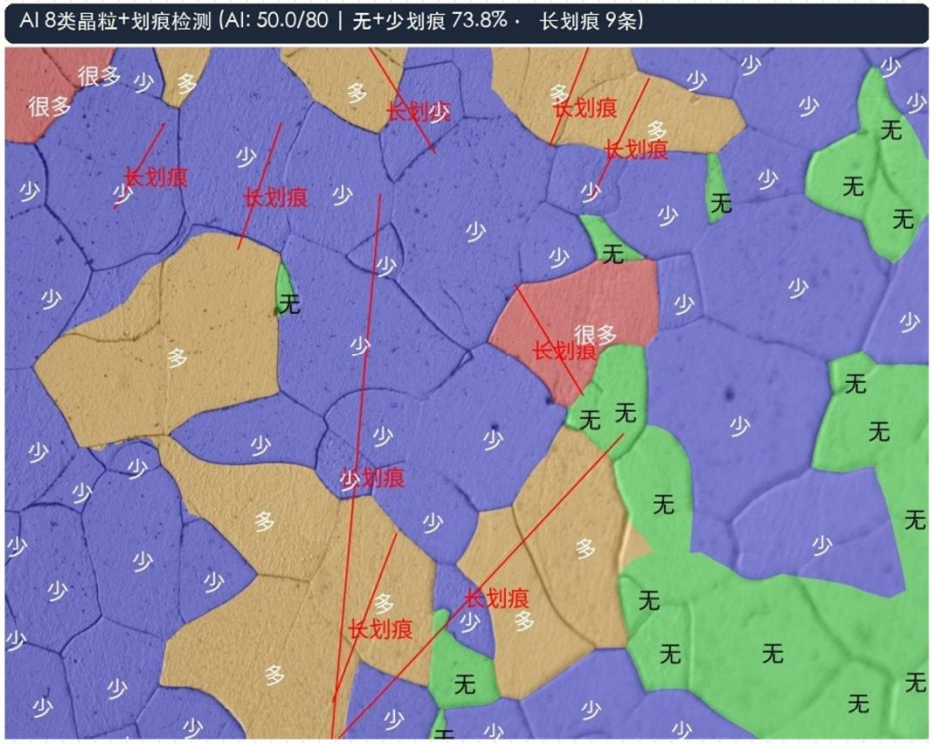
南昌大学 NANCHANG UNIVERSITY

覆盖 2025 与 2026 两届国赛、全部三大材料，按“好/中/差直观对比 + 典型试样完整数据”逐项复核

国赛批次与材料	优秀档 (3套/24图)	中等档 (9套/72图)	较差档 (3套/24图)	本次复核发现的核心结论
26年 · 工业纯铁	专家均分 **65.0**	专家均分 **50.2**	专家均分 **26.3**	**晶粒四级划痕分级 + 过腐蚀宽槽与浅划痕检出**：准确区分晶界与浅划痕，优/中/差代表作 AI 评分 **50.0 → 40.0 → 29.3** 单调递减，全样本严格对齐
25年 · 20钢	专家均分 **65.9**	专家均分 **49.4**	专家均分 **25.7**	**剥离珠光体第二相 + 粗磨痕与蚀斑联合检出**：精准分割 48~96 处珠光体，优/中/差代表作 AI 评分 **66.2 → 53.6 → 20.3**，按档位严格递减
25年 · 球墨铸铁	专家均分 **69.1**	专家均分 **51.0**	专家均分 **29.7**	**五态石墨球识别 + 24宫格磨斜失焦量化**：精准识别 37~73 颗石墨球与渗碳体，优/中/差代表作 AI 评分 **61.7 → 51.7 → 40.8**，全样本严格单调对齐

材料一（26年工业纯铁·优秀样本1）：全视野二联图与人机评分对比





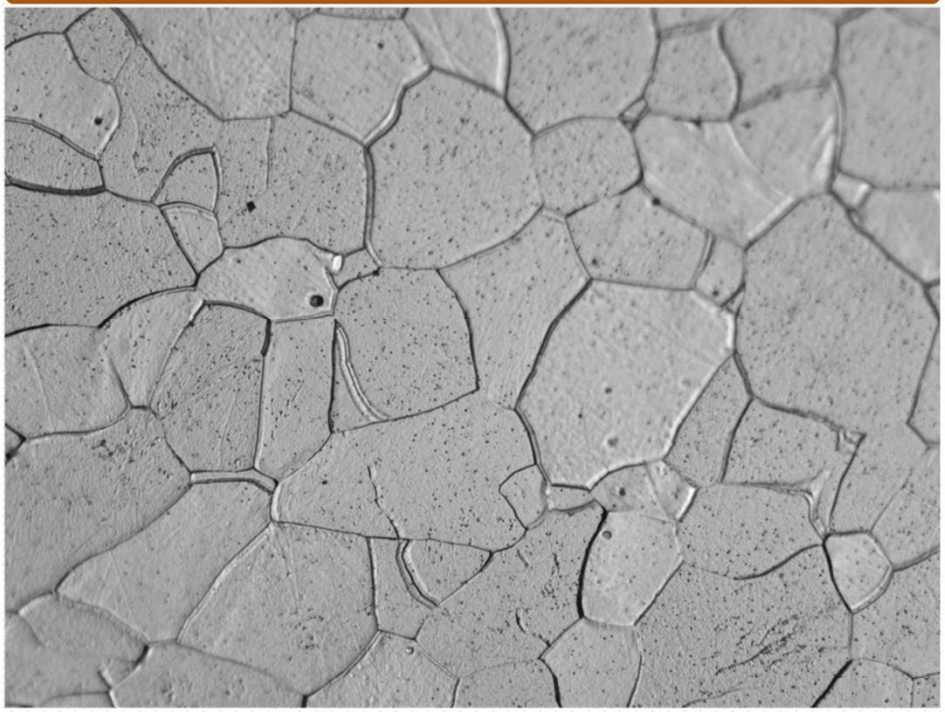
专家人工评分：65.7 / 80（优秀档）

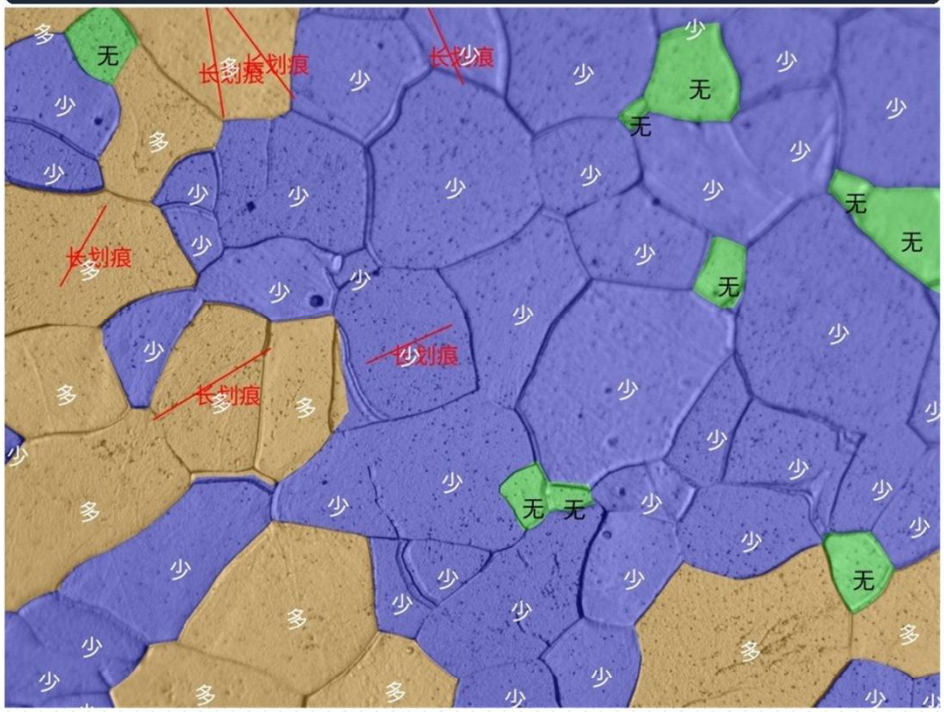
高低倍整体规整洁净：该选手 100× 低倍视野无划痕晶粒达**100.0%**（低倍 AI 达 **62.0 / 80**），500× 高倍铁素体晶界清晰闭合、右侧与下方大面积晶粒洁净无痕（高倍AI **50.0 / 80** 同样位居纯铁高倍榜首），人机分档一致。

同视野 AI 评分：50.0 / 80（清晰度 22.4 | 划痕 8.8 | 假象 18.8）

精准区分晶界与浅划痕+ 逐晶粒洁净度统计：针对上方**同一张 500× 高倍视野（高倍1.jpg）**，系统在不把深色晶界误判为划痕的前提下，**精准检出左上与中下局部的9 条真实细长划痕（1 处交叉）**；同时测出右侧与下方**无划痕洁净晶粒占15.0%**（无+少划痕合计 **73.8%**），清晰度与划痕综合位列纯铁高倍第一。

材料一（26年工业纯铁·中等样本1）：全视野二联图与人机评分对比





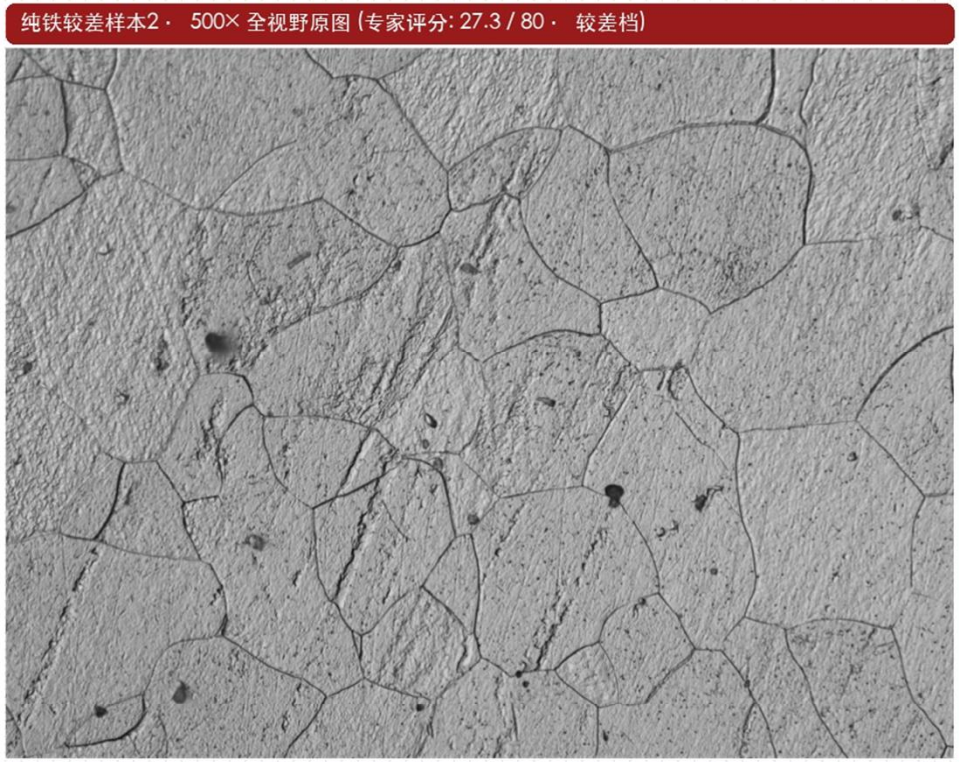
专家人工评分：50.7 / 80（中等档）

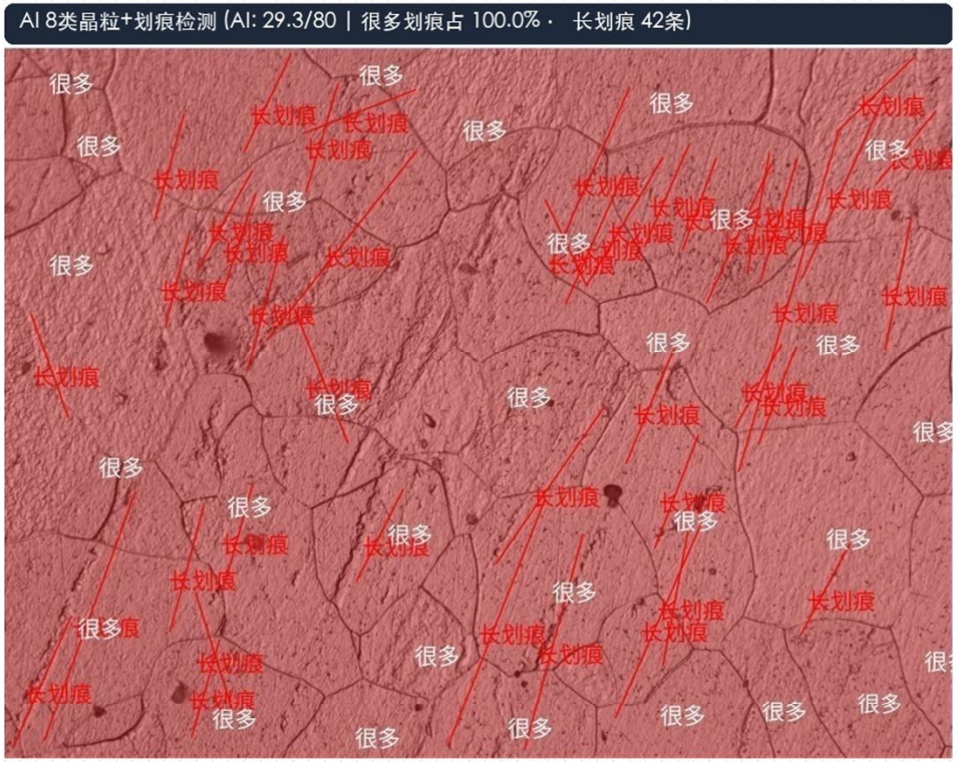
严重过腐蚀宽黑晶界 + 全视场细碎划痕：高倍晶界严重过腐蚀呈粗黑宽槽，晶粒内部遍布细碎抛光划痕与微蚀坑，且 100× 低倍存在大面积灰白块（占 **47.3%**），专家与 AI 均评为中等档。

同视野 AI 评分：40.0 / 80（清晰度 16.0 | 划痕 5.6 | 假象 18.4）

检出 28.7% 过腐蚀宽晶界 + 无划痕晶粒骤降至 4.4%：系统准确检出高倍 **28.7% 的严重过腐蚀双晶界/宽黑槽**（清晰度由 22.4 降至 **16.0 分**）；同时**无划痕晶粒从 15.0% 骤降至 4.4%**（较少划痕 **67.3%**、较多划痕 **28.3%**）并检出 **6 条明显长划痕**（划痕分 **5.6 分**）。

材料一（26年工业纯铁·较差样本2）：全视野二联图与人机评分对比





专家人工评分：27.3 / 80（较差档）

全视场密集粗磨痕与低倍重度灰白块：高倍磨面遗留全视场密集平行粗磨痕（晶界本身未明显过腐蚀），低倍伴随 **58.3%** 深色灰白块，专家综合全套表现评为较差档。

同视野 AI 评分：29.3 / 80（清晰度 10.5 | 划痕 0.0 | 假象 18.8）

全视场 100% 晶粒判为"很多划痕" + 检出 42 条平行粗磨痕：系统直接检出全视场 **42 条平行粗磨痕（7 处交叉）**，并将 **100.0% 的晶粒判为"很多划痕"（深红色掩码）**，划痕单项直接扣至 **0.0 分**；同时准确判别其过腐蚀并不严重（双晶界仅 **8.7%**），但因 **38.4%** 晶界被密集粗磨痕割裂遮蔽，清晰度压至 **10.5 分**，人机分档一致。

材料一（26年工业纯铁）：典型试样完整专家 vs AI 对比数据

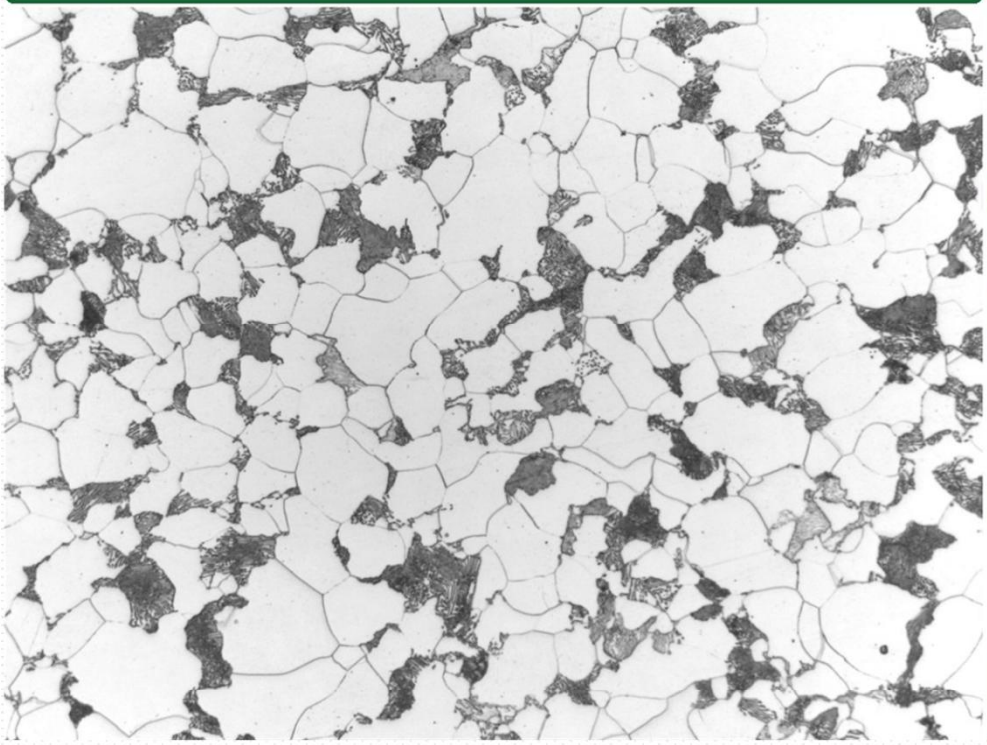
南昌大学 NANCHANG UNIVERSITY

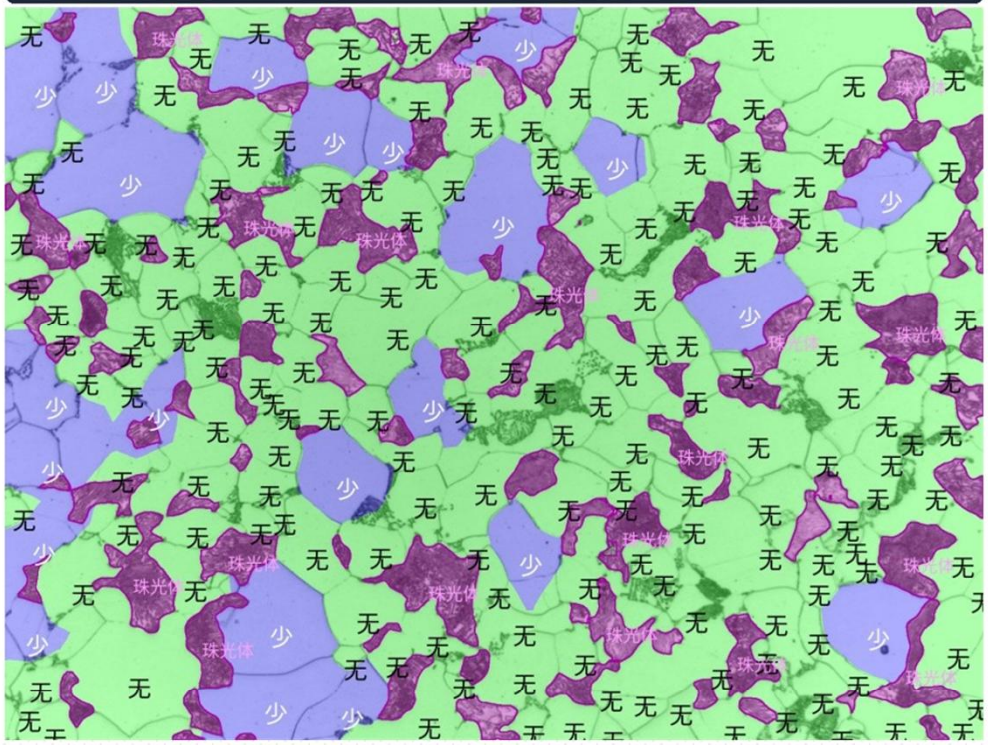
综合高倍晶界清晰度、过腐蚀宽槽与晶内洁净度，工业纯铁全部 7 套代表试样按档位严格单调递减 (50.0/47.5/45.7 → 43.5/40.0 → 29.3/26.9)

典型试样（匿名）	档位	专家全套分 (/80)	高倍单图 AI (/80)	清晰度 (/40)	划痕 (/20)	假象 (/20)	检出长划痕条数	无+少划痕占比	多+很多划痕占比	过腐蚀双晶界	100×低倍AI分 (灰白块)
纯铁优秀样本 1	优秀	**65.7**	**50.0**	**22.4**	**8.8**	18.8	**9条 (1交叉)**	73.8% (**无 15.0%**)	**26.2%**	11.1%	**62.0分** (25.7%)
纯铁优秀样本 2	优秀	**65.0**	**47.5**	**22.9**	5.8	18.8	**15条 (3交叉)**	80.3% (无 0.9%)	**19.7%**	15.5%	**58.5分** (24.4%)
纯铁优秀样本 3	优秀	**64.3**	**45.7**	**20.4**	6.5	18.8	**8条 (1交叉)**	59.8% (无 0.2%)	**40.1%**	**5.1% (浅晶界)**	**57.5分** (18.4%)
纯铁中等样本 1	中等	**50.7**	**40.0**	16.0	5.6	18.4	**6条 (1交叉)**	71.7% (无 4.4%)	**28.3%**	**28.7% (重过腐蚀)**	低倍灰白块 **47.3%**
纯铁中等样本 2	中等	**51.6**	**43.5**	17.4	7.7	18.4	**8条**	84.2% (无 21.0%)	**15.9%**	15.2%	低倍灰白块 29.8%
纯铁较差样本 1	较差	**23.7**	**26.9**	**7.7**	**0.4**	18.8	**20条 (6交叉)**	5.2% (无0.0%)	**94.8% (很多)**	**24.4% (局部黑斑)**	低倍灰白块 **50.7%**
纯铁较差样本 2	较差	**27.3**	**29.3**	**10.5**	**0.0**	18.8	**42条 (7交叉)**	**0.0%**	**100.0% (很多)**	**8.7% (磨痕遮晶界)**	低倍灰白块 **58.3%**

关键发现：系统精准区分浅腐蚀细晶界（优秀样本3仅5.1%）与严重过腐蚀宽黑槽（中等样本1达28.7%），使纯铁全部 7 套样本的 AI 评分与专家排序 100% 单调一致。

材料二（25年 20钢·优秀样本1）：全视野二联图与人机评分对比





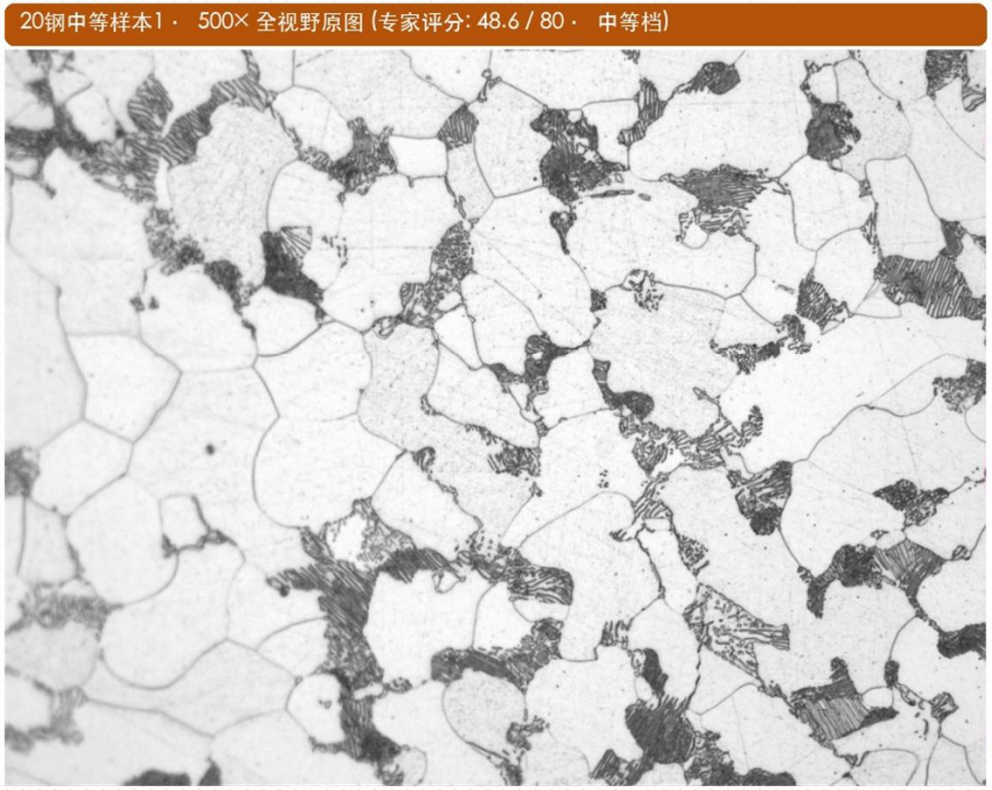
专家人工评分：65.0 / 80（优秀档）

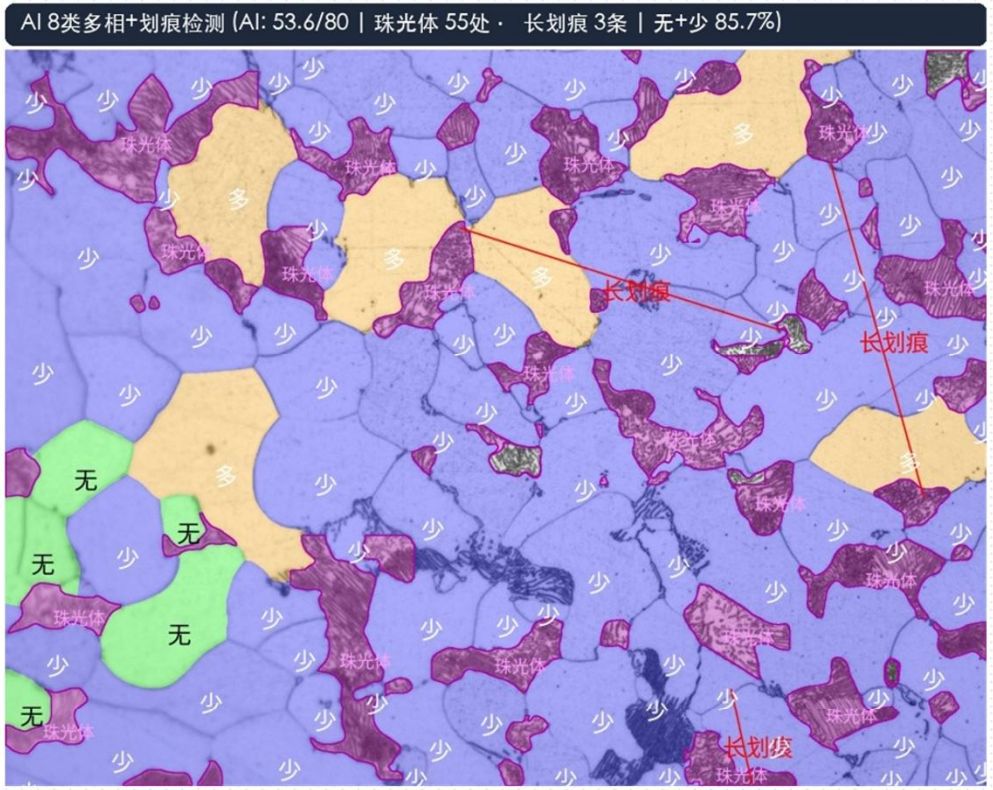
组织清晰且极少划痕：铁素体晶界清晰完整，珠光体（暗区）分布均匀、片层反差适中，高倍视场几乎无粗划痕与脏污，人机分档一致。

同视野 AI 评分：66.2 / 80（清晰度 31.0 | 划痕 17.2 | 假象 18.0）

精准剥离 96 处珠光体 + 100% 洁净基体：主模型精准分割 **96 处珠光体岛（紫红色掩码）** 并完整保留铁素体晶界，铁素体基体 **无划痕晶粒占 76.6%、较少划痕占 23.4%（无+少合计 100.0%）**；全视场 **0 条跨晶长划痕**，划痕得 **17.2 分**，清晰度达 **31.0 分**。

材料二（25年 20钢·中等样本1）：全视野二联图与人机评分对比





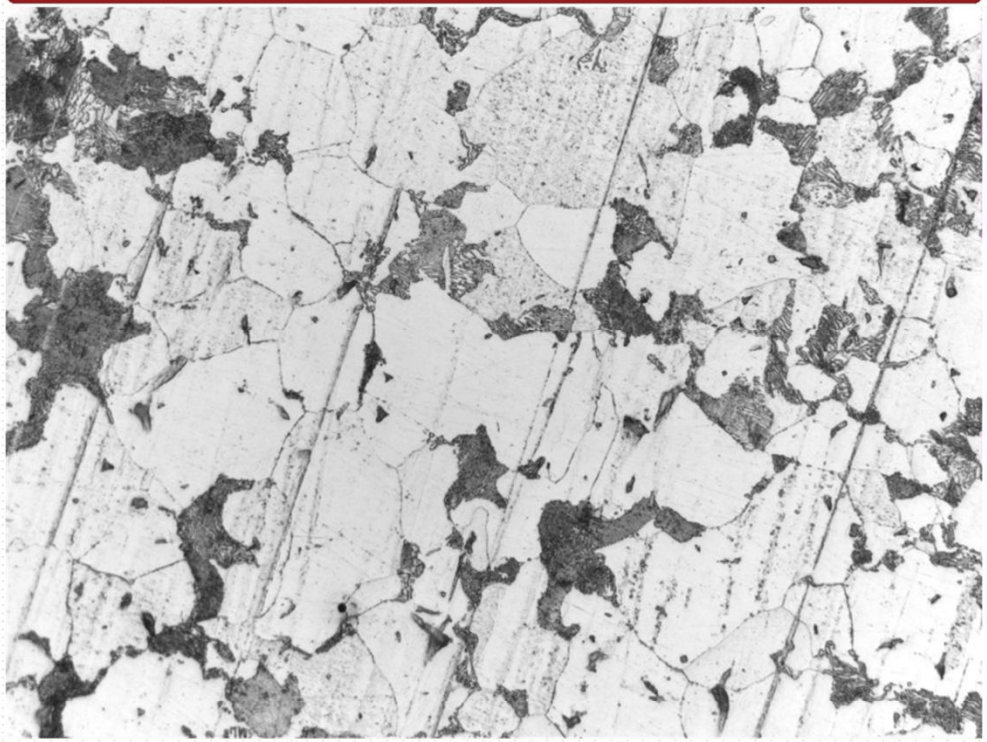
专家人工评分：48.6 / 80（中等档）

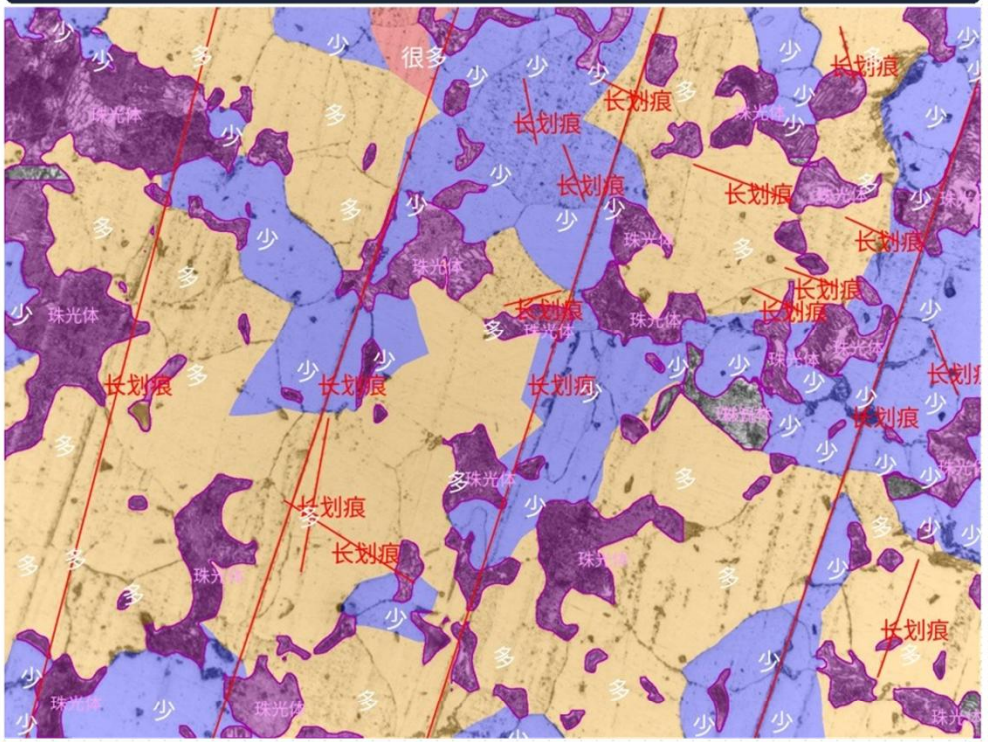
局部浅晶界与多条抛光划痕：腐蚀偏浅导致部分铁素体晶界断续，且视场中部及右侧残留 4 条斜向长划痕与普遍晶内细碎划痕，专家与 AI 均评为中等档。

同视野 AI 评分：53.6 / 80（清晰度 26.3 | 划痕 7.7 | 假象 19.6）

8类专模剥离 55 处珠光体 + 双路线检出 3 条真实长划痕：排除珠光体片层暗线干扰后，**双路线检出右侧与中部 3 条真实长划痕**；**8类主模型**测出无划痕晶粒从 76.6% 降至 **4.8%**（较少划痕 **80.9%**、较多划痕 **14.3%**），划痕分降至 **7.7 分**，磨面洁净无蚀斑（假象 **19.6 分**）。

材料二（25年 20钢·较差样本1）：全视野二联图与人机评分对比





专家人工评分：24.6 / 80（较差档）

粗磨划痕与基体蚀斑假象双重失误：粗磨痕未抛除即行腐蚀，全视场 4 根贯穿上下的粗磨划痕及多条斜向划痕交错，且铁素体基体残留大面积灰黑腐蚀脏污与黑斑坑洞，专家与 AI 均评为较差档。

同视野 AI 评分：20.3 / 80（清晰度 9.8 | 划痕 0.4 | 假象 10.1）

17 条粗磨划痕（含 4 根贯穿全图大划痕）+ 63.0% 多划痕晶粒 + 基体蚀斑：主模型独立分割 **78 处珠光体**并将 **63.0% 的铁素体晶粒判为较多/很多划痕（无划痕 0.0%）**；完整检出 **4 根贯穿上下的粗磨大划痕及共 17 条长划痕（3 处交叉）**，划痕分扣至 **0.4 分**；同时检出 **19 个坑洞/黑斑**与铁素体基体灰黑蚀斑（水渍脏污 **4.1%**），假象分降至 **10.1 分**、清晰度压至 **9.8 分**，与中等样本拉开 **33.3 分**差距。

材料二（25年 20钢）：典型试样完整专家 vs AI 对比数据

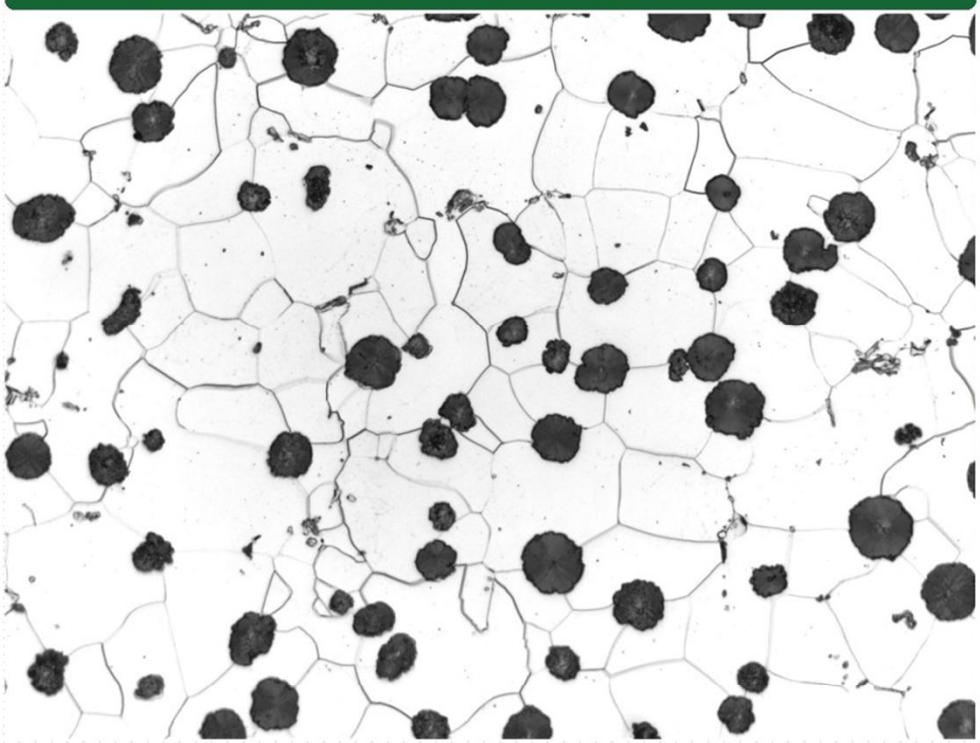
南昌大学 NANCHANG UNIVERSITY

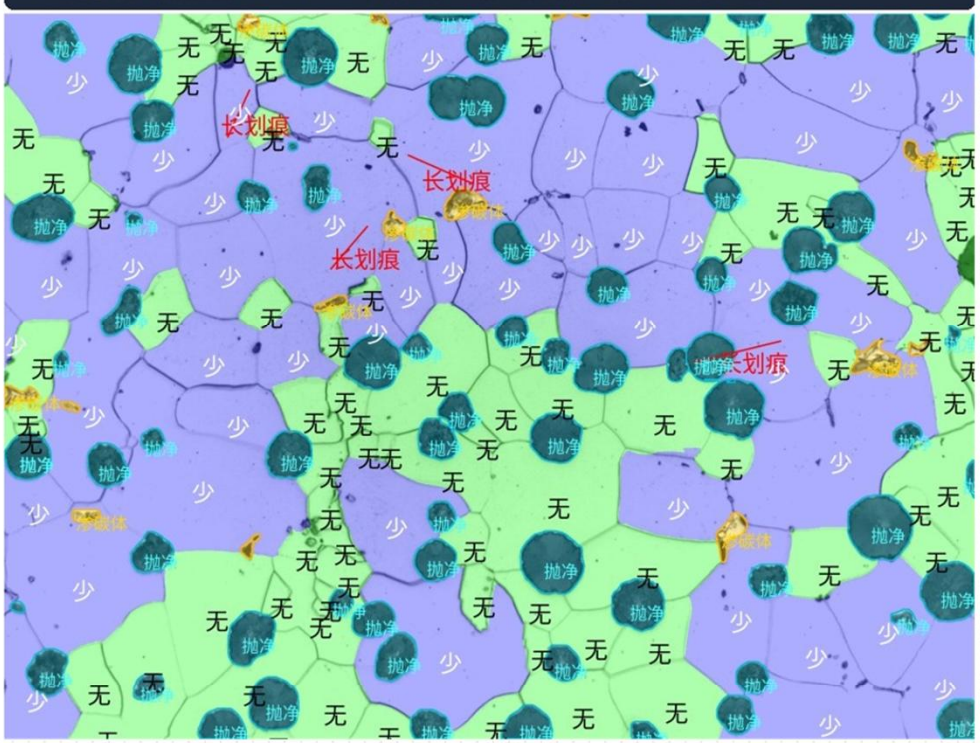
精准剥离珠光体第二相 + 跨晶长划痕与铁素体蚀斑检测，20 钢全部样本按档位严格单调递减（66.2/63.2 → 54.8/53.6 → 27.9/20.3）

典型试样（匿名）	档位	专家评分 (/80)	AI 综合分 (/80)	清晰度 (/40)	划痕 (/20)	假象 (/20)	珠光体数量	检出长划痕与微缺陷	无划痕晶粒占比	无+少划痕占比	多+很多划痕占比
20钢优秀样本 1	优秀	65.0	66.2	31.0	17.2	18.0	96处	0条	76.6%	100.0%	0.0%
20钢优秀样本 2	优秀	65.0	63.2	29.6	16.0	17.6	52处	0条	66.5%	100.0%	0.0%
20钢中等样本 1	中等	48.6	53.6	26.3	7.7	19.6	55处	3条	4.8%	85.7%	14.3%
20钢中等样本 2	中等	50.4	54.8	27.9	10.1	16.8	82处	1条 + 3坑洞	5.2%	79.5%	20.5%
20钢较差样本 1	较差	24.6	20.3	9.8	0.4	10.1	78处	17条 (3交叉) + 19坑洞	0.0%	37.0%	63.0%
20钢较差样本 2	较差	31.8	27.9	10.7	8.6	8.6	48处	1条 + 13坑洞 + 灰白块	0.2%	71.6%	28.5%

关键发现：系统将深色珠光体与铁素体基体精准分离，避免把正常珠光体误判为灰白块或模糊区，同时准确捕获粗磨划痕与基体蚀斑，使 20 钢全部代表试样按优、中、差严格单调排序。

材料三（25年球墨铸铁·优秀样本1）：全视野二联图与人机评分对比





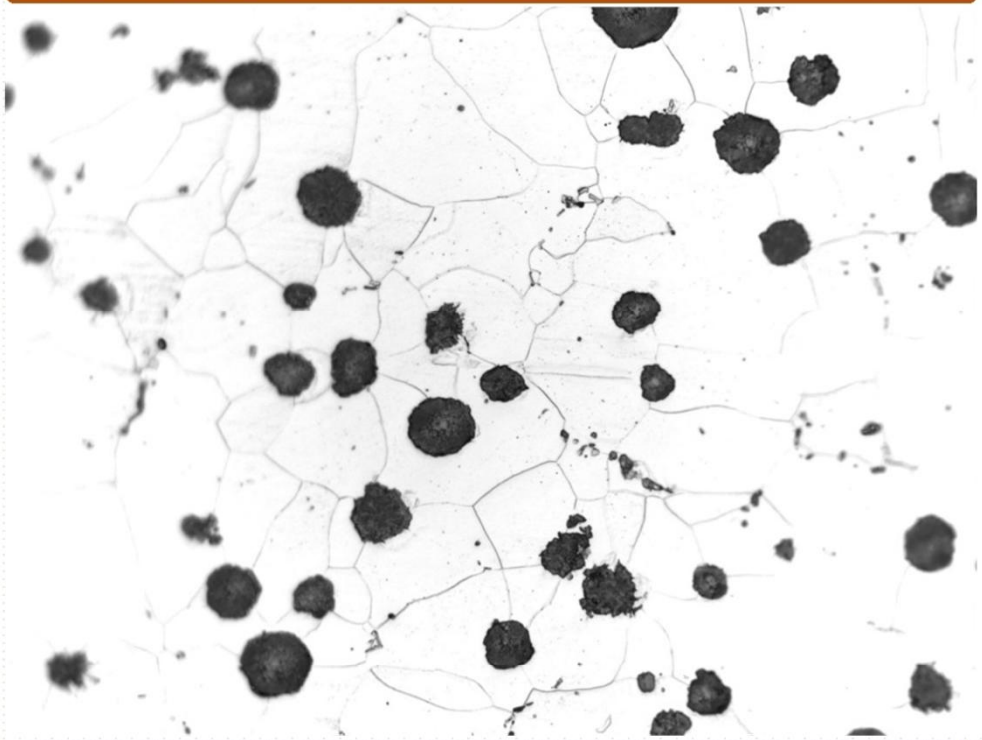
专家人工评分：69.0 / 80（优秀档）

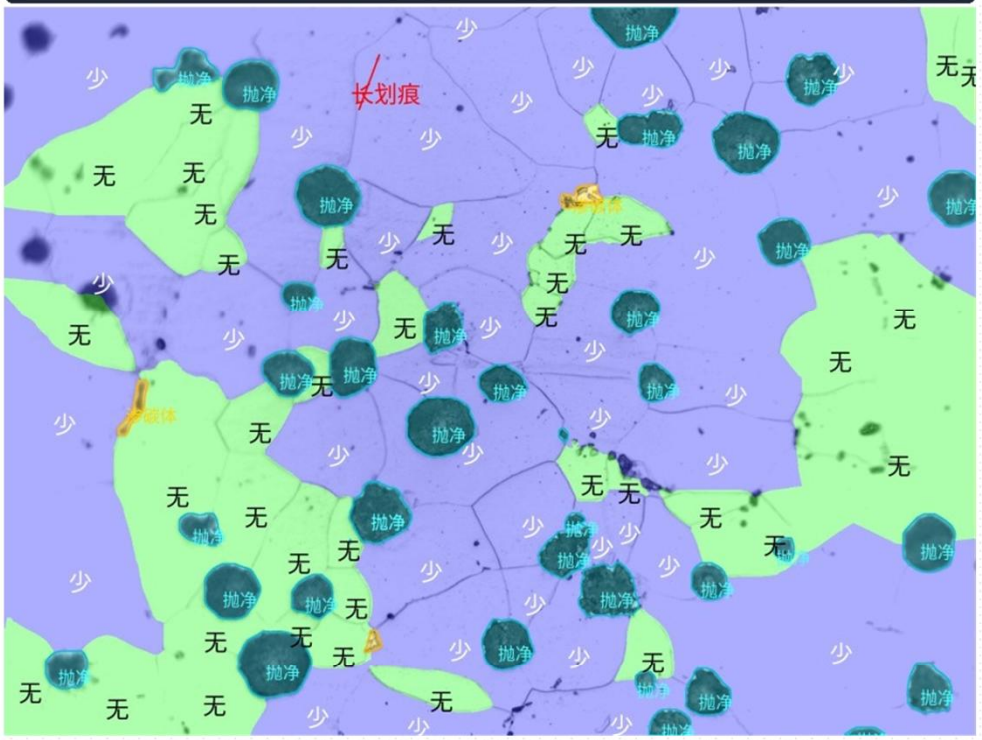
全视场合焦且球墨抛光洁净：石墨球圆整饱满、无拖尾剥落，铁素体基体晶界全视场清晰合焦，抛光面整体洁净，人机分档一致。

同视野 AI 评分：61.7 / 80（清晰度 37.4 | 划痕 8.3 | 假象 16.0）

精准识别 73 颗“抛净”石墨球与 15 处渗碳体：系统准确识别 **73 颗“抛净”石墨球（青绿色掩码）** 与 **15 处渗碳体（金黄色掩码）**，剥离第二相后铁素体晶界清晰度达 **37.4 / 40**；仅检出 **4 条局部浅划痕**，**无+少划痕晶粒达 100.0%**。

材料三（25年球墨铸铁·中等样本1）：全视野二联图与人机评分对比





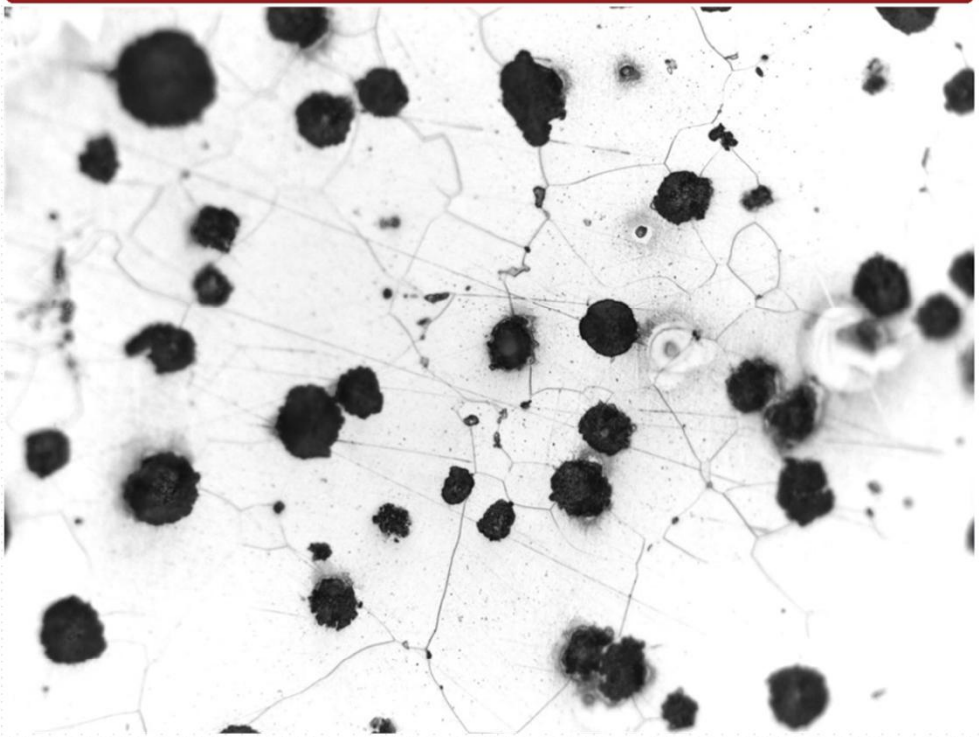
专家人工评分：51.4 / 80（中等档）

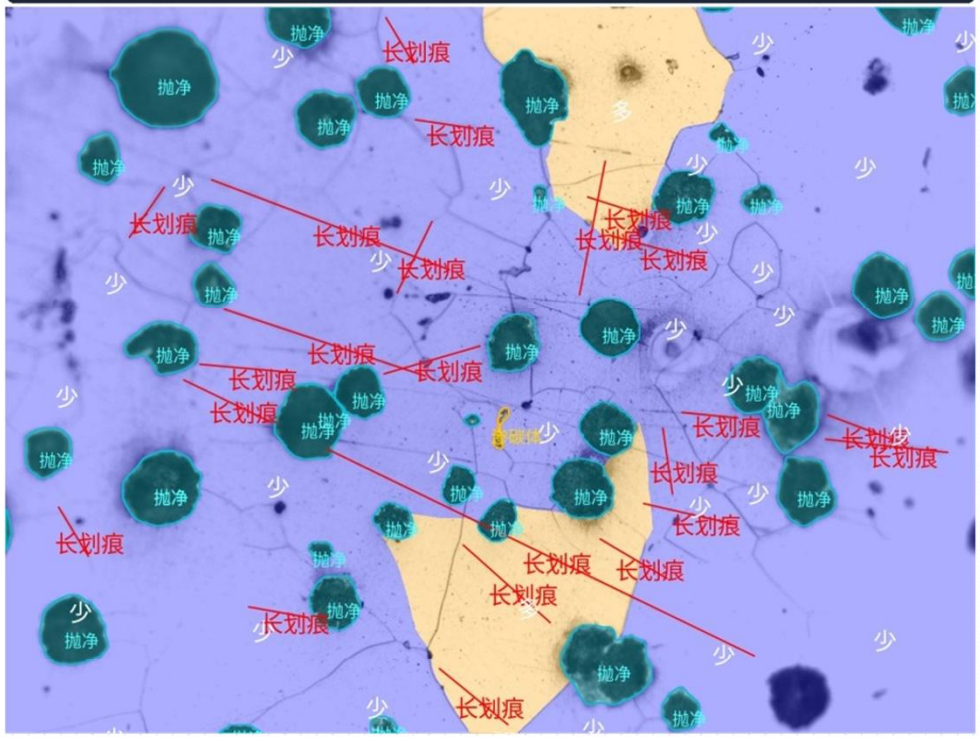
“中间凸、左右塌”磨斜失焦：制样磨面倾斜导致 500× 全视野**仅中间竖条能看清晶界，左右两侧严重发虚失焦**，专家与 AI 均将其评为中等档。

同视野 AI 评分：51.7 / 80（清晰度 24.9 | 划痕 10.8 | 假象 16.0）

37 颗石墨球 + 3 处渗碳体精分与两侧失焦量化：系统准确分割 **37 颗“抛净”石墨球**与 **3 处渗碳体**，仅检出 **1 条短划痕**；24 宫格对焦检测发现**左右两侧磨斜区域清晰度较中心骤降 73.9%**（计入两侧失焦发虚区域后全图模糊晶界占比达 **32.2%**，清晰度由 37.4 降至 **24.9 分**）。

材料三（25年球墨铸铁·较差样本1）：全视野二联图与人机评分对比





专家人工评分：35.0 / 80（较差档）

边缘失焦叠加密集交叉长划痕：不仅存在边缘磨斜失焦与石墨球周边拖尾污染，而且全视场布满粗大交叉长划痕，磨面质量差，专家评为 35.0 分。

同视野 AI 评分：40.8 / 80（清晰度 24.8 | 划痕 0.0 | 假象 16.0）

分割 45 颗石墨球 + 检出 23 条交叉长划痕：系统识别出 **45 颗石墨球**与 **2 处渗碳体**，无划痕晶粒 **0.0%**（较多划痕占 **11.9%**）；全视场检出 **23 条跨晶长划痕（3 处交叉）**，划痕单项直接扣至 **0.0 分**；模糊晶界占比达 **30.8%**。

材料三（25年球墨铸铁）：典型试样完整专家 vs AI 对比数据

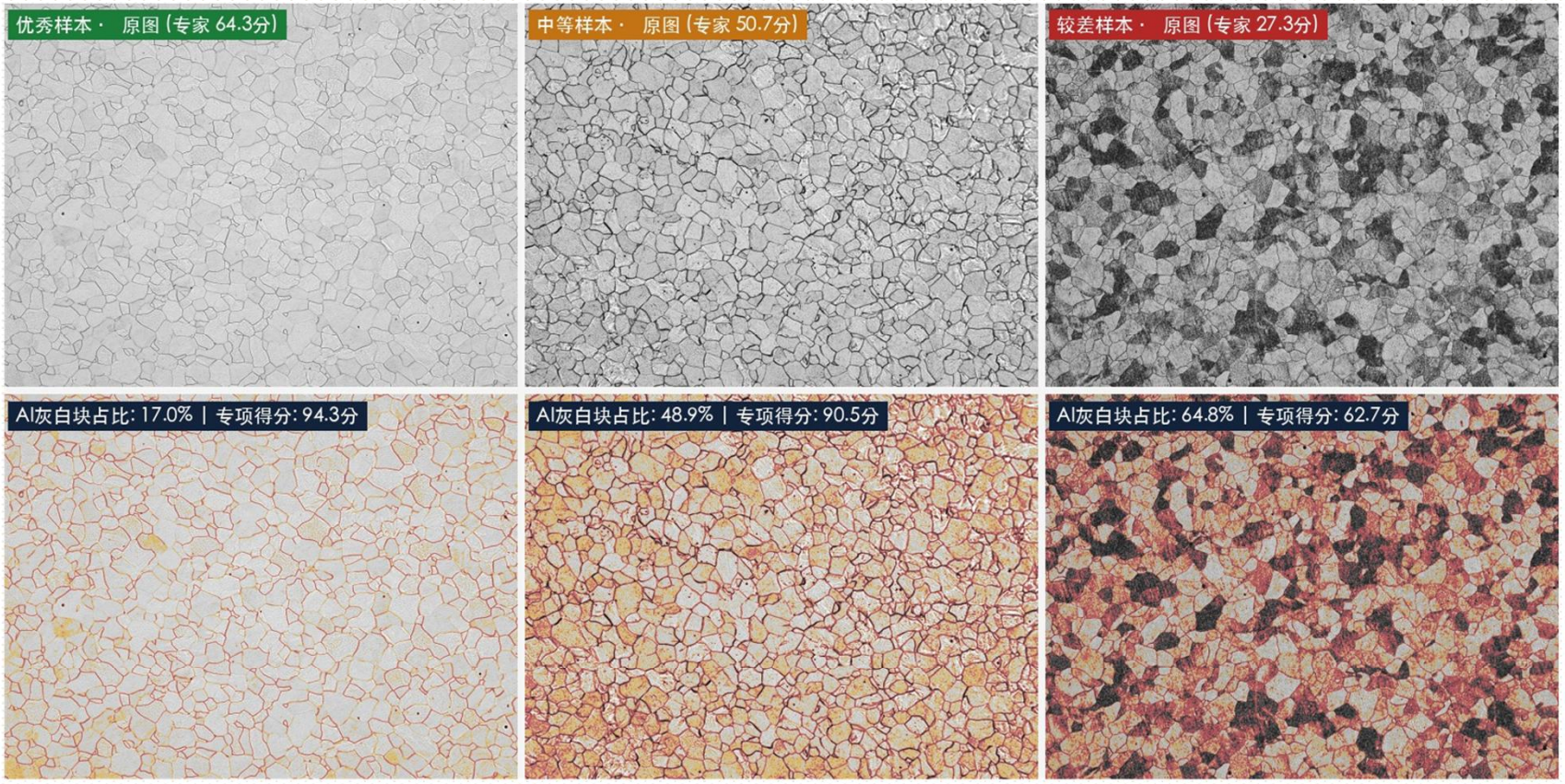
南昌大学 NANCHANG UNIVERSITY

五态石墨球精准分割 + 24 宫格磨斜失焦检测 + 长划痕量化，球铁全部样本严格单调递减 (61.7/58.1 → 51.7/50.4 → 40.8/34.8)

典型试样（匿名）	档位	专家评分 (/80)	AI 综合分 (/80)	清晰度 (/40)	划痕 (/20)	假象 (/20)	石墨球 / 渗碳体	检出长划痕条数	无+少划痕占比	模糊晶界占比 (均值)	两侧失焦最大衰减
球铁优秀样本 1	优秀	**69.0**	**61.7**	**37.4**	**8.3**	16.0	**73颗 / 15处**	**4条**	**100.0%** (无 36.0%)	**10.1% (99.6)**	29.2% (全图合焦)
球铁优秀样本 2	优秀	**68.2**	**58.1**	**28.7**	**13.5**	**15.9**	**40颗 / 7处**	**1条**	**100.0%** (无 52.7%)	25.9% (62.8)	33.3% (全图合焦)
球铁中等样本 1	中等	**51.4**	**51.7**	24.9	10.8	16.0	**37颗 / 3处**	1条	**100.0%** (无 29.9%)	**32.2%** (49.1)	**73.6% (双侧失焦)**
球铁中等样本 2	中等	**49.8**	**50.4**	24.2	10.3	15.9	**49颗 / 1处**	3条	**100.0%** (无 44.5%)	**40.0%** (53.7)	**80.3% (单侧磨斜)**
球铁较差样本 1	较差	**35.0**	**40.8**	**24.8**	**0.0**	16.0	**45颗 / 2处**	**23条 (3交叉)**	88.1% (多 11.9%)	**30.8%** (49.6)	**59.0% (双侧失焦)**
球铁较差样本 2	较差	**29.2**	**34.8**	**24.7**	**2.2**	**7.9**	**46颗 / 3处**	**28条 (2交叉)**	100.0% (无 18.4%)	**28.9%** (49.1)	36.2% (基体污染)

关键发现：系统精准区分“抛净”石墨球、渗碳体与铁素体基体，配合 24 宫格磨斜失焦量化与跨晶长划痕检测，使球铁 6 个代表试样按优、中、差实现 100% 严格单调排序。

专项难点验证一：低倍“灰白块”腐蚀假象 2×3 分级对比



左列：优秀样本（专家 64.3分）

晶面亮白洁净（灰白块 17.0% · 专项 94.3分）：仅零星晶粒呈浅黄色轻微衬度差，无深色腐蚀斑，磨面均匀干净。

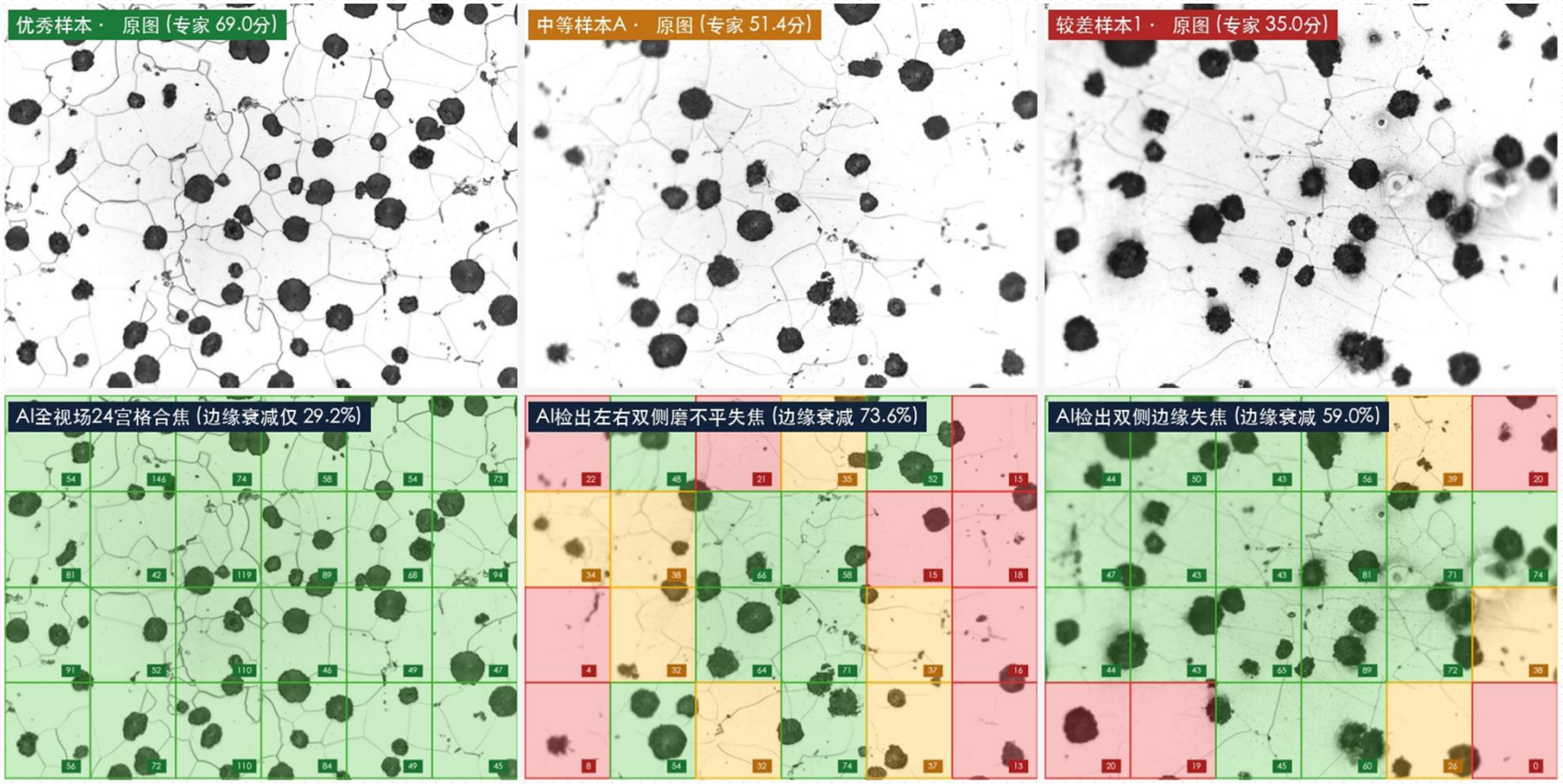
中列：中等样本（专家 50.7分）

普遍轻度发灰（灰白块 48.9% · 专项 90.5分）：近半数晶粒呈橙黄色轻度发灰，但无重度黑斑，专项仅作温和扣减。

右列：较差样本（专家 27.3分）

重度黑灰腐蚀块（灰白块 64.8% · 专项 62.7分）：大面积深红色重度黑斑覆盖——**揭示需在规则中区分“正常晶粒取向轻度发灰”与“扰动层未除尽的重度黑灰块”。**

专项难点验证二：球墨铸铁“磨样不平两侧失焦” 2×3 直观对比



左列：优秀样本（专家 69.0分）

全视场 24 宫格 100% 绿格合焦：各格清晰度 42 ~ 146，边缘衰减仅 29.2%，磨面平整、全视场同焦。

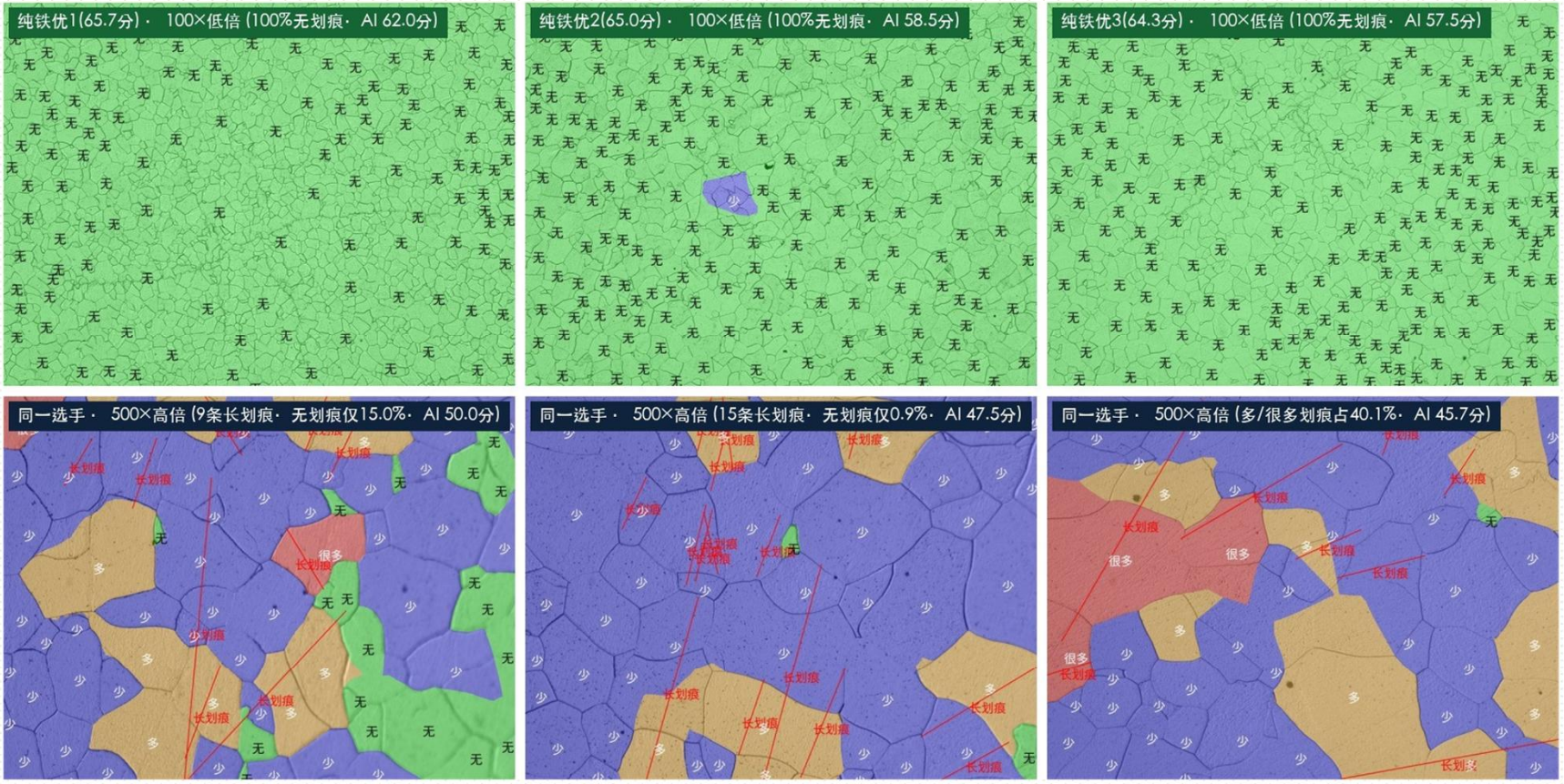
中列：中等样本A（专家 51.4分）

双侧严重磨不平（16格失焦 · 衰减 73.6%）：失焦面积反超右列，但因**中心两列合焦且无粗划痕**，人工仍给 51.4 分。

右列：较差样本1（专家 35.0分）

失焦较中列轻（8格失焦 · 衰减 59.0%）：主要因叠加 **23 条粗长划痕**拿低分——**揭示现行评分未将“磨面不平”作为关键扣分项，需组委会确认是否优化评分规则。**

专项难点验证三：纯铁优秀样本“低倍光环掩盖高倍细划痕” 2×3 直观对比



左列：纯铁优秀样本 1（专家 65.7分）

100×全绿无痕（62.0分）vs 500×检出 9 条长划痕（50.0分）：高倍无划痕晶粒仅 15.0%，低倍与高倍得分相差 12.0分。

中列：纯铁优秀样本 2（专家 65.0分）

100×全绿无痕（58.5分）vs 500×检出 15 条长划痕（47.5分）：高倍含 3 处交叉长划痕、无划痕晶粒仅 0.9%，但全套仍获 65.0 分。

右列：纯铁优秀样本 3（专家 64.3分）

100×全绿无痕（57.5分）vs 500×多划痕占 40.1%（45.7分）：揭示全套评分受“低倍好图光环”主导，需组委会明确高低倍占分权重。



南昌大学
NANCHANG UNIVERSITY

报告内容

CONTENTS

01 研发背景

02 研发过程

03 近期成效

04 未来规划

未来规划起点——AI 校准需要什么样的“人类专家黄金基准”？

传统“看完单人全套 8 张图后孤立估分”缺乏单图单维度标准分，难以直接作为校准 AI 的精细标尺。

局限一：全套 8 图混评易受“光环效应”主导

低倍好图掩盖高倍瑕疵：看完 4 低倍 + 4 高倍只给 1 套总分，低倍洁净易掩盖高倍瑕疵，缺乏单图单维度真值。

局限二：缺少选手间同屏参照，且无 8 张图各自的单图分值

受出场顺序干扰且跨图折算难：逐套孤立阅片易受前后反差影响；无 8 张图各自单图分，好坏不均时难心算折合。

局限三：跨组尺度不一

组间给分松紧漂移：不同评审组的“65 分”对应的实际样品质量存在偏差。

核心命题：要让 AI 算得准，首先必须通过评审机制创新获得单张图、单维度的高一致性“专家黄金真值”。

评判方式改革——从“8图合并孤立估分”到“单图同屏两两比对”

南昌大学 NANCHANG UNIVERSITY

将人类专家高负荷的“看 8 张图合并估绝对分值”转化为最敏锐的“单张图、单维度同屏相对优劣判断”。

传统模式：看完单人 8 张试样图后合并估分（难且波动大）

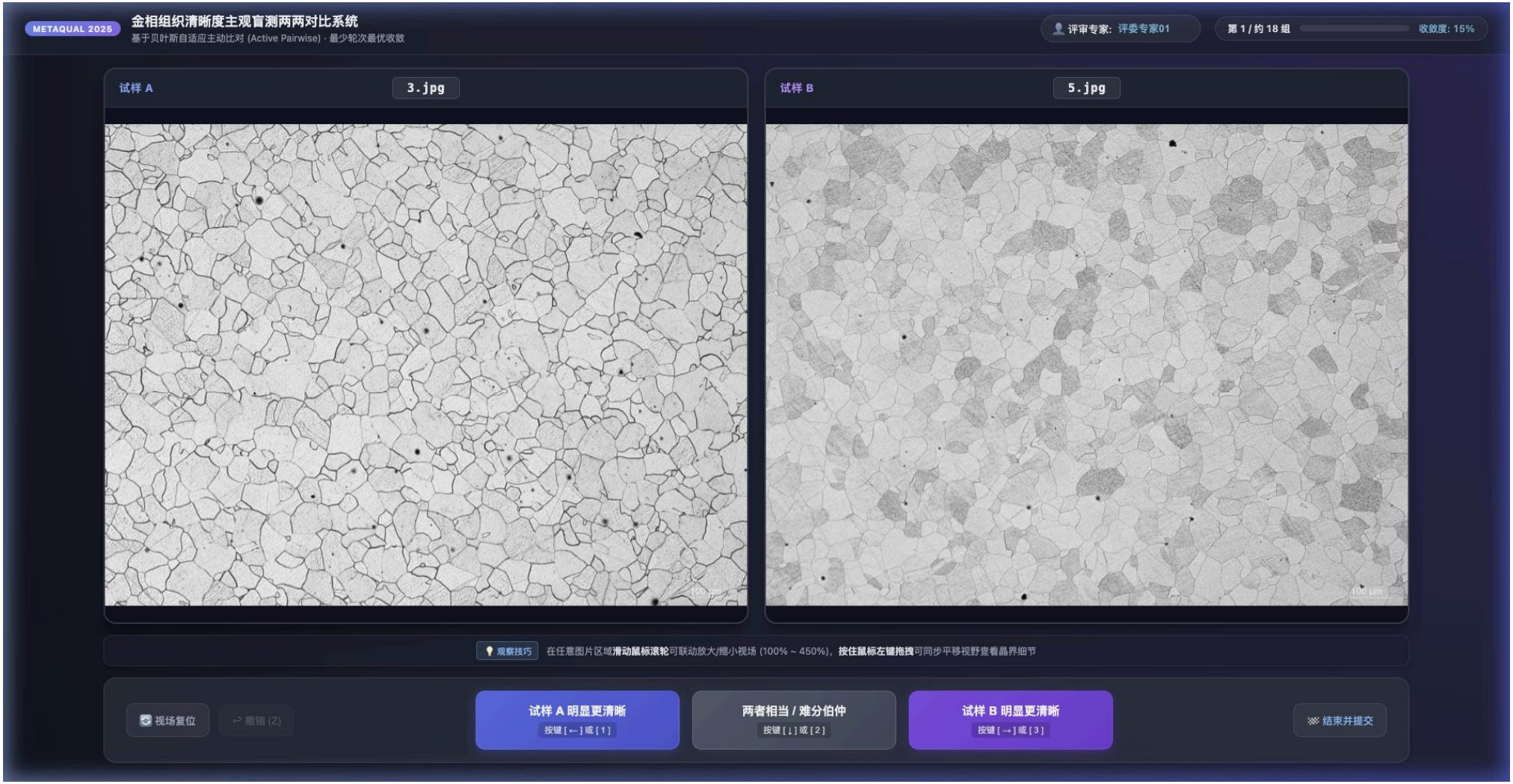
8图冷热不均难折合、孤立估分波动大：连续看完 8 张高低倍图后凭印象在脑中折算整套图的分项分值（如清晰度该给 32 分还是 36 分？），既易受低倍好图光环主导，折算为百分制标准分后不同专家打分标准差也达 **± 5.5 分**。

创新模式：单图单维度同屏两两比对（易且极准）

- **三选一极速比高低**：A、B 两张单图并排同屏展示，每次只问单一维度（如“哪张晶界更清晰？”或“哪张灰白块更轻？”），百分制标准分下判断误差骤降至 **± 1.8 分**。
- **已上线在线盲测工具**：支持**组织清晰度**与**灰白块程度**在线比对，专家每次点击仅需 **3~5 秒**。

范式转换：变“看8图孤立估绝对分”为“单图单维同屏比高低”，直接产出每张图的精准专家标准分。

评委端双盲比对单手按键交互界面



高效配对与全局算分——赛前定标尺、赛中减 90% 比对量

通过“AI 梯度初筛 + 瑞士轮同水平配对”，赛前 2 分钟即可定标官方基准库，赛中无需全员互比。

1. 赛前怎么挑基准样？瑞士轮怎么省力？

- **12 件官方基准样（AI 按梯度初筛 + 专家定标）**：赛前由 **AI 从往届国赛数千张图中按 60~95 分均匀梯度挑出 12 件覆盖优/中/差的代表样**；瑞士轮**只让水平接近的两张图相比**，专家仅需比对 **18~22 次（约 2 分钟）** 即可排定官方基准库。
- **赛场单组（如 40 人 × 8 图）无需另建基准库**：320 张图拍完保存后，AI 对齐官方基准库在后台数秒内完成全组初排，仅抽取**前 15% 争冠区与分差胶着试样**供专家同屏比对终审（减少 **90% 以上**工作量）。

2. 全局统计排序算分（跨组共用同一标尺）

- **胜负关系转化为客观标准分**：汇总多位专家的两两比对胜负记录，**通过统计模型（Bradley-Terry）自动解算出每张图的客观标准分**，彻底消除不同评委手松手严差异。
- **跨组分数尺度完全拉齐**：全赛场**所有分组共用同一套“官方黄金基准库”锚定分值**，避免因各组选手水平高低不同而产生组间给分漂移。

核心机制：赛前用往届样建“全国统一官方基准库”（2分钟定标），赛中每组 40 人由 AI 对齐基准初排、专家仅比争冠与临界样。

仿真实验与实测盲测验证——专家共识与 AI 吻合度双高

南昌大学 NANCHANG UNIVERSITY

在同等阅片耗时下，两两比对将一等奖错判率从 18.4% 降至 3.1%，且 AI 算法与专家盲测名次高度吻合（满分 1.0）

核心评估指标（满分 1.0 代表名次或意见 100% 完全一致）	传统8图合并孤立估分	双盲两两比对 + 统计求解	验证结论与实际效果
1. 评委排出的名次与真实质量名次的吻合度 *(仿真)*	0.892	**0.986**	**排名误差降低 87%（逼近满分 1.0）**
2. 相对真实分数的平均赋分误差 *(百分制标准分)*	± 3.84 分	**± 1.12 分**	**赋分误差缩小 70.8%，消除顺序干扰**
3. 前 10% 顶尖奖项（一等奖）错判率 *(仿真)*	**18.4%**	**3.1%**	**一等奖边界错判风险降低 83.2%**
4. 专家之间的意见一致度 *(12件纯铁盲测实测)*	0.68 ~ 0.74	**清晰度 0.946 / 灰白块 0.965**	**消除个人手松手严差异，达成极高专家共识**
5. AI 算法与专家盲测共识的名次吻合度 *(实测)*	缺乏单图真值	**清晰度 0.930 / 灰白块 0.965**	**AI 清晰度前4名与后4名同专家 100% 重合**

双重验证：两两比对既让专家间意见一致度达 0.946~0.965，又使校准后的 AI 与专家盲测名次吻合度达 0.930~0.965。

2026—2028 人机协同路线图与组委会审议建议

南昌大学 NANCHANG UNIVERSITY

以 25/26 年国赛试样建成**跨届统一的“官方黄金基准库”**，比赛现场无需每组另建基准库。

第一阶段（2026赛后~2027省赛）：建成全国统一的“官方黄金基准库”

行动建议一（赛前定标尺）：由 AI 从 25/26 年国赛图库按梯度初筛各材料 12~15 件典型样，邀请专家在线两两比对（每人 10~15 分钟），建成**后续各届、各组共用的官方黄金基准库**并完成 AI 标定。

第二阶段（2027全国大赛）：每组拍完后 AI 对齐官方基准初排 + 异常复核

赛场协同机制（无需每组另建基准库）：单组（如 40 人 × 8 图 = 320 图）拍完保存后，AI 直接按**官方黄金基准库**统一标尺完成全组评分与初排，自动预警“人机分差过大”或“高分多划痕”样品供仲裁组同屏复核。

第三阶段（2028及未来）：AI 全量初评 + 专家聚焦前 15% 争冠区同屏终审

行动建议二（赛教融合）：省赛与实训开放 AI 诊断报告；国赛每组 40 人由 AI 完成常规初评，系统自动提取该组**前 15% 争冠区与分界线胶着试样（同屏对照官方基准样）**交由专家两两比对终审。

协同闭环：赛前用往届国赛图建“全国统一官方黄金基准库”，赛中每组 40 人拍完后由 AI 对齐基准库初排、专家仅对争冠与临界样做同屏比对终审。

南昌大学
NANCHANG UNIVERSITY

感谢各位老师

请批评指正！

