



RenderScope

使用说明

2026 年 07 月

法律事项

本文档中呈现的信息属于格兰菲智能科技股份有限公司和/或其附属公司（以下统称为“格兰菲”）所有，且格兰菲保留不经通知随时对本文档信息或对本文档所述任何产品和服务做出修改的权利。格兰菲不担保信息、文本、图案、链接或本文档内所含其他项目的准确性或完整性。本文档内容仅供参考，格兰菲不对本文档所述产品的可销售性、所有权、不侵犯知识产权、准确性、完整性、稳定性或特定用途适用性做任何暗示担保、保证。格兰菲可不经通知随时对本文档或本文档所述产品或服务做出更改，但不承诺因此更新本文档。

非经格兰菲事先书面许可，任何实体或个人均不得获得本文档的副本，且无权以任何方式处理本文档，包括但不限于使用、复制、修改、合并、出版、发行、销售或传播本文档的部分或全部。在任何情况下，格兰菲均不对因本文档引起的、由本文档造成的、或与之相关的任何索赔、损害或其他责任负责。格兰菲保留自行决定随时更改、修改、添加或删除本文档的部分或全部的权利。

格兰菲对本文档享有最终解释权。

目录

法律事项.....	2
修订记录.....	4
1 Pixel Rate 像素填充率测试.....	5
1.1 测量指标定义.....	5
1.2 测量原理与执行流程.....	5
1.3 结果指标说明.....	6
2 Texel Rate 纹理填充率测试.....	6
2.1 测量指标定义.....	6
2.2 测量原理与执行流程.....	6
2.3 结果指标说明.....	7
3 FP32 单精度/FP16 半精度浮点计算吞吐量（GFLOPS）测试.....	8
3.1 测量指标定义.....	8
3.2 测量原理与执行流程.....	8
3.3 结果指标说明.....	8

修订记录

版本	日期	作者	描述
A0	2026/07/13	格兰菲	初次发布

格兰菲智能科技有限公司

1 Pixel Rate 像素填充率测试

1.1 测量指标定义

像素填充率 (Pixel Fill Rate) 是评估图形处理器 (GPU) 核心光栅化性能的关键指标之一，其物理含义为：在单位时间内，GPU 的光栅化单元 (ROPs) 能够处理并写入帧缓冲区 (显存) 的像素总量。该数值直接反映了 GPU 在显存带宽利用率和像素级操作方面的处理上限。

本测试在设计上剔除了纹理采样、复杂着色器运算及几何变换等干扰变量，仅采用基础的全屏三角形作为绘制图元，并禁用深度测试与混合功能。这种实现方式被称为极限测试，旨在将测试负载聚焦于光栅化阶段与显存之间的数据传输通道，最大限度地排除 GPU 其他计算单元的并发干扰，从而测得硬件在理想条件下的理论峰值像素填充带宽。

1.2 测量原理与执行流程

为获得准确的峰值填充率，软件采用分阶段的负载探测与定量测量机制，具体流程如下：

第一阶段：动态负载校准与饱和点探测

在测量周期的初始阶段，软件通过闭环反馈算法动态调整每帧的渲染负载。具体而言，系统会根据上一帧的渲染耗时，实时增减当前帧的绘制指令 (Draw Command) 数量。此过程的目的是使 GPU 的核心单元达到满负荷工作状态，同时将帧率稳定在预设的目标区间 (约 30 FPS)。该阶段结束后，系统将锁定一个关键参数——使 GPU 刚好处于性能饱和临界点的最佳每帧绘制指令数。

第二阶段：定量采集与持续渲染测试

锁定前述校准参数后，软件随即进入正式的数据采集阶段。在接下来的固定时间窗口 (30 秒) 内，GPU 将以此饱和负载稳定运行，持续执行全屏覆盖的像素渲染操作，软件会运行多次测试，以最高分作为最终的峰值结果输出。系统在此阶段仅进行精确计时与渲染帧数累加，不介入任何额外干扰，以确保采集到的数据严格反映 GPU 的持续稳态性能。

1.3 结果指标说明

测试完成后，软件输出包含以下四项核心指标的报告，各指标定义如下：

Frame Count (帧数总计)：在整个测量时间窗口内，GPU 完成渲染并提交至显示缓冲区的画面总帧数。

Draw Count Per Frame (单帧绘制批次)：由第一阶段校准得出的饱和负载值，表示 GPU 在满负荷状态下，每渲染一帧画面所需执行的绘制命令批次数量。

Used Time (测量耗时)：定量采集阶段的实际运行时长。

Score (像素填充率)：该值为本次测试的最终量化结果。其计算方式为测量期间总处理像素数（帧数 × 单帧像素总量）与测量耗时的比值，单位为 **GPixels/s**（每秒十亿像素）。该数值越高，表示 GPU 的像素填充能力越强。

2 Texel Rate 纹理填充率测试

2.1 测量指标定义

纹理填充率（Texel Fill Rate）是评估图形处理器（GPU）纹理映射与采样性能的核心量化指标之一。其物理含义为：在单位时间内，GPU 的纹理单元能够从纹理内存中读取、过滤并输出至帧缓冲区的纹理像素总量。该数值直接反映了 GPU 在纹理带宽利用率和多纹理混合处理方面的硬件上限。

本测试在设计上模拟了复杂图形应用中的多纹理叠加场景，通过同时混合 8 张全分辨率纹理并将其渲染至输出画布，使纹理单元与显存通道进入持续饱和状态。这种实现方式旨在将测试负载高度集中于纹理读取与过滤流水线，最大限度地暴露 GPU 在纹理带宽和缓存方面的性能瓶颈，从而测得硬件在理想重负载条件下的纹理填充带宽峰值。

2.2 测量原理与执行流程

为获得准确且可重复的峰值填充率，软件采用分阶段的负载探测与定量测量机制，

具体流程如下：

第一阶段：动态负载校准与饱和点探测

在测量周期的初始阶段，软件通过闭环反馈算法动态调整每帧的渲染负载。系统会实时监测上一帧的渲染耗时，并根据反馈结果以 5 个绘制批次为步长动态增减当前帧的绘制命令（Draw Command）数量，直至将帧率稳定在预设的 30 FPS 目标窗口内。此过程旨在使 GPU 的纹理单元与显存控制器达到满负荷工作状态，同时避免 CPU 或垂直同步机制引入的伪瓶颈。该阶段结束后，系统将锁定使 GPU 恰好处于纹理带宽饱和临界点的最佳每帧绘制命令数。

第二阶段：定量采集与持续渲染测试

锁定校准参数后，软件随即进入正式的数据采集阶段。在接下来的固定时间窗口（30 秒）内，GPU 将以此饱和负载稳定运行，持续执行全屏覆盖的纹理采样与混合操作，软件会运行多次测试，以最高分作为最终的峰值结果输出。系统在此阶段仅进行精确计时与渲染帧数累加，不介入任何额外干扰，以确保采集到的数据严格反映 GPU 在持续稳态纹理负载下的真实吞吐能力。

2.3 结果指标说明

测试完成后，软件输出包含以下四项核心指标的报告，各指标定义如下：

Frame Count（帧数总计）：在整个测量时间窗口内，GPU 完成渲染并提交至显示缓冲区的画面总帧数。

Draw Count Per Frame（单帧绘制批次）：由第一阶段校准得出的饱和负载值，表示 GPU 在满负荷状态下，每渲染一帧画面所需执行的绘制命令批次数。

Used Time（测量耗时）：定量采集阶段的精确运行时长。

Score（纹理填充率）：该值为本次测试的最终量化结果。其计算方式为：

纹理填充率 = (纹理采样层数 × 总绘制次数 × 画布宽度 × 画布高度) ÷ 测量耗时

其中纹理采样层数固定为 8。结果为 GTexels/s（每秒十亿纹理像素），该数值越高，表示 GPU 的纹理映射与采样能力越强。

3 FP32 单精度/FP16 半精度浮点计算吞吐量（GFLOPS）测试

3.1 测量指标定义

FP32/FP16 计算吞吐量（GFLOPS）是评估 GPU 核心算术逻辑单元（ALU）浮点计算性能的关键指标之一，其物理含义为：在单位时间内，GPU 的计算核心能够完成的 32 位单精度/16 位半精度浮点运算总量，单位为 GFLOPS（每秒十亿次浮点运算）。该数值直接反映了 GPU 在通用计算（GPGPU）场景下的理论峰值算力上限。

本测试在设计上剔除了访存延迟、分支预测、纹理采样及渲染管线干扰，仅采用高度紧凑的向量乘加运算（FMA）作为计算负载，并确保所有操作数均驻留在寄存器或高速缓存中。这种实现方式被称为纯算术极限测试，旨在将测试负载聚焦于 ALU 单元的发射与执行效率，最大限度地排除显存带宽及调度开销的干扰，从而测得硬件在理想条件下的理论峰值浮点计算吞吐量。

3.2 测量原理与执行流程

为获得准确的峰值 GFLOPS 值，Kernel 采用大量 FMA 计算。

FMA（融合乘加）：执行 $a = a * b + c$ 操作，在支持该指令的架构上，一次指令可完成两次浮点运算（乘和加），效率最高。同时，软件采用 float4/half4，以充分利用 SIMD 单元的并行发射能力。

测试过程中对 kernel 执行固定次数，并保留 kernel 最短执行时间。系统在此阶段仅进行精确计时与总浮点操作数累加，不介入任何额外干扰（如内存读写、条件分支），以确保采集到的数据严格反映 GPU 计算单元的持续稳态算力。

3.3 结果指标说明

测试完成后，软件输出包含以下四项核心指标的报告，各指标定义如下：

Flops Per Workitem(单个workitem浮点运算量): GPU在满负荷状态下,单个worklitem

所执行的浮点运算总次数。

Total Flops（单个 kernel 浮点运算量）：单个 workitem 浮点运算量*global_size。

Used Time（测量耗时）：选取执行时间最短的一次。

Score（浮点计算吞吐量）：该值为本次测试的最终量化结果。其计算方式为 Total Flops/Used Time，单位为 GFLOPS（每秒十亿次浮点运算）。该数值越高，表示 GPU 的 FP32/FP16 浮点计算能力越强。