中文长文档的版面恢复

双栏版式在学术期刊中极为常见，几何方法需要正确处理栏间空白与通栏元素。

本样本在右栏刻意加入更长的段落，用于验证行带聚类对块高不均的鲁棒性。

阅读顺序还原的目标，是让输出顺序与人类阅读顺序一致，而不是与内部对象顺序一致。

当块高差异超过两倍时，简单阈值会把同一行带拆成两半，需要按列局部计算中位数。

表格与插图的锚定关系也需要被保留，否则图文对应关系会丢失。

此外，栏沟宽度阈值应随版心宽度缩放，绝对像素阈值在不同纸张尺寸下不可移植。