

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2024 年 1 月 4 日 (04.01.2024)



(10) 国际公布号
WO 2024/001051 A1

(51) 国际专利分类号:
G16B 40/30 (2019.01)

(21) 国际申请号: PCT/CN2022/136698

(22) 国际申请日: 2022 年 12 月 5 日 (05.12.2022)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
PCT/CN2022/102505
2022年6月29日 (29.06.2022) CN
PCT/CN2022/102424
2022年6月29日 (29.06.2022) CN
PCT/CN2022/102360
2022年6月29日 (29.06.2022) CN

PCT/CN2022/102443
2022年6月29日 (29.06.2022) CN
PCT/CN2022/110779
2022年8月8日 (08.08.2022) CN

(71) 申请人: 深圳华大生命科学研究院 (BGI SHENZHEN) [CN/CN]; 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。

(72) 发明人: 李美 (LI, Mei); 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。刘焕林 (LIN, Huanlin); 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。李敏 (LI, Min); 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。岑伟轩 (CEN, Weixuan); 中国广东省深圳市盐田区北山

(54) Title: SPATIAL OMICS SINGLE CELL DATA ACQUISITION METHOD AND APPARATUS, AND ELECTRONIC DEVICE

(54) 发明名称: 空间组学单细胞数据获取方法及装置和电子设备

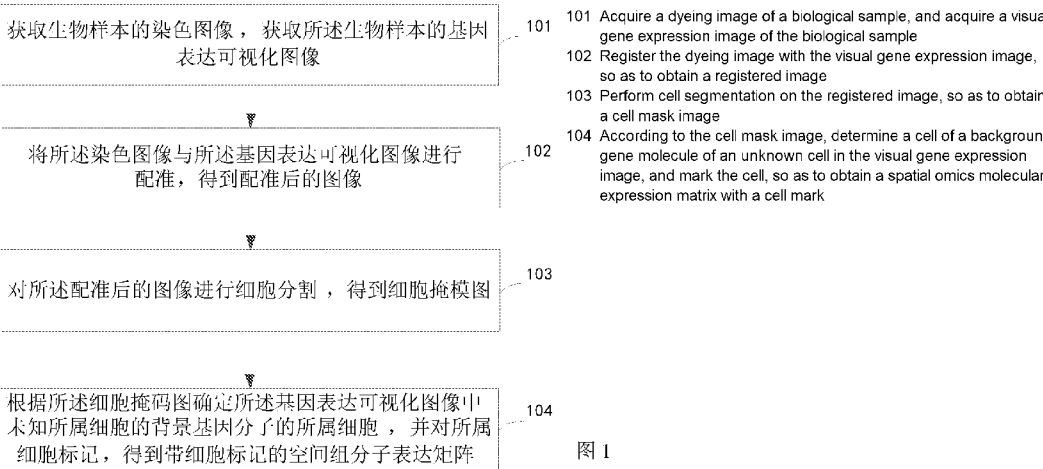


图 1

(57) Abstract: A spatial omics single cell data acquisition method and apparatus, and an electronic device and a storage medium. The method comprises: acquiring a dyeing image of a biological sample, and acquiring a visual gene expression image of the biological sample; registering the dyeing image with the visual gene expression image, so as to obtain a registered image; performing cell segmentation on the registered image, so as to obtain a cell mask image; and according to the cell mask image, determining a cell of a background gene molecule of an unknown cell in the visual gene expression image, and marking the cell, so as to obtain a spatial omics molecular expression matrix with a cell mark. Compared with the prior art, the present application relates to obtaining high-precision spatial single cell data by means of a gene expression matrix diagram, and a processing flow involving splicing, registration, segmentation and correction.

(57) 摘要: 一种空间组学单细胞数据获取方法及装置、电子设备和存储介质, 获取生物样本的染色图像, 获取所述生物样本的基因表达可视化图像; 将所述染色图像与所述基因表达可视化图像进行配准, 得到配准后的图像; 对所述配准后的图像进行细胞分割, 得到细胞掩模图; 根据所述细胞掩模图确定所述基因表达可视化图像中未知所属细胞的背景基因分子的所属细胞, 并对所属细胞标记, 得到带细胞标记的空间组分子表达矩阵。与相关技术相比, 本申请通过基因表达矩阵图、拼接、配准、分割和校正的处理流程, 以获得高精度的空间单细胞数据。

工业区综合楼, Guangdong 518083 (CN)。黄子睿(HUANG, Zirui); 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。邓忠寒(DENG, Zhonghan); 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。苏奎(SU, Kui); 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。李泽鹏(LI, Zepeng); 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。冯柠(FENG, Ning); 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。姜琪琛(JIANG, Qichen); 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。吴逸文(WU, Yiwen); 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。康强(KANG, Qiang); 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。黎宇翔(LI, Yuxiang); 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。张勇(ZHANG, Yong); 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。徐讯(XU, Xun); 中国广东省深圳市盐田区北山工业区综合楼, Guangdong 518083 (CN)。

(74) 代理人: 北京法胜知识产权代理有限公司 (FASHENG INTELLECTUAL PROPERTY COMPANY, LTD.); 中国北京市海淀区北洼路45号1号楼2层202室, Beijing 100142 (CN)。

(81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW。

(84) 指定国(除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, ME, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告(条约第21条(3))。

空间组学单细胞数据获取方法及装置和电子设备

相关技术的交叉引用

5 本申请基于申请号为 PCT/CN2022/102505、申请日为 2022 年 6 月 29 日的 PCT 专利申请提出、基于申请号为 PCT/CN2022/102424、申请日为 2022 年 6 月 29 日的 PCT 专利申请提出、基于申请号为 PCT/CN2022/102360、申请日为 2022 年 6 月 29 日的 PCT 专利申请提出、基于申请号为 PCT/CN2022/102443、申请日为 2022 年 6 月 29 日的 PCT 专利申请提出、基于申请号为 PCT/CN2022/110779、申请日为 2022 年 8 月 8 日的 PCT 专利申请提出，并要求上述 PCT 专利申请的优先权，上述 PCT 专利申请的全部内容在此引入本申请作为参考。

技术领域

15 本申请涉及数据处理技术领域，尤其涉及一种空间组学单细胞数据获取方法及装置和电子设备。

背景技术

20 空间组学技术可以揭示出不同时期组织结构内部空间的机能和组织结构之间的异质性，为深入研究生物体组织结构和细胞间通信创造了有利的条件。目前，部分空间转录组技术已经发展到可以捕获整个转录组，但是大部分情况下提供的表达信息矩阵还达不到单细胞层次的分辨率。因此如何实现对单细胞数据的获取是目前亟需解决的问题。

发明内容

25 本申请提供了一种空间组学单细胞数据获取方法、装置、电子设备和存储介质。根据本申请的第一方面，提供了一种空间组学单细胞数据获取方法，其中，包括：获取生物样本的染色图像，获取所述生物样本的基因表达可视化图像；将所述染色图像与所述基因表达可视化图像进行配准，得到配准后的图像；对所述配准后的图像进行细胞分割，得到细胞掩模图；根据所述细胞掩模图确定所述基因表达可视化图像中未知所属细胞的背景基因分子的所属细胞，并对所属细胞标记，得到带细胞标记的空间组分子表达矩阵。

30 可选的，所述获取生物样本的染色图像包括：通过显微镜获取所述生物样本的一个染色图像；或通过显微镜获取所述生物样本的由多个子染色图像拼接的拼接染色图像；或通过所述显微镜获取所述生物样本的多个子染色图像，对所述多个子染色图像进行拼接，得到拼接染色图像。

35 可选的，所述获取生物样本的染色图像包括：通过显微镜获取所述生物样本的多个子染色图像；对所述生物样本的多个子染色图像进行第一拼接，得到第一染色图像；根据所述第一染色图像的图像质量判断是否对所述生物样本的多个子染色图像进行第二拼接；

40 若否，将所述第一染色图像确定为所述染色图像；若是，对所述生物样本的多个子染色图像进行第二拼接，得到所述染色图像。可选的，所述通过所述显微镜获取所述生物样本的多个子染色图像，对所述多个子染色图像进行拼接，得到拼接染色图像包括：

45 分别确定多个子染色图像和模板图像的重叠区域；将所述多个子染色图像和所述模板图像的被确定为重叠区域的部分分别裁剪为多个染色图像重叠子图和数量相同的多个模板图像重叠子图，以配对为多个重叠子图对；

50 通过对所述多个重叠子图对进行频域计算，分别获得各子染色图像和所述模板图像之间的偏移量；以及

将所述各子染色图像分别朝向所述模板图像的方向平移所述偏移量，并与所述模板图像拼接在一起。

可选的，所述通过所述显微镜获取所述生物样本的多个子染色图像，对所述多个子染色图像进行拼接，得到拼接染色图像包括：

- 5 获取所述多个子染色图像以及关于所述生物样本的轨迹线或轨迹点的模板信息；
对所述多个子染色图像进行预拼接得到所述多个子染色图像的预拼接坐标；
根据轨迹线或轨迹点的特征从所述多个子染色图像中选取至少一个第一参考图像，根据所述生物样本的模板信息和所述第一参考图像的预拼接坐标推导全局模板；
以及

- 10 对于所述多个子染色图像中除了所述第一参考图像之外的第二参考图像，计算所述第二参考图像的预拼接坐标与所述第二参考图像在所述全局模板中的相应模板坐标之间的偏移量，并基于所述偏移量对所述第二参考图像进行坐标调整以拼接生成所述拼接染色图像。

- 15 可选的，所述将所述染色图像与所述基因表达可视化图像进行配准，得到配准后的图像包括：

获取所述染色图像中的轨迹线，基于所述轨迹线与水平方向的夹角旋转所述染色图像；其中，所述染色图像和所述基因表达可视化图像包括多条轨迹线和/或多个轨迹点，所述轨迹点是由两条所述轨迹线相交形成的，所述多条轨迹线中的每条轨迹线具有对应的索引号；

- 20 将所述染色图像进行放大或缩小得到第三参考图像，所述第三参考图像与所述基因表达可视化图像具有相同的比例；

基于所述基因表达可视化图像和所述第三参考图像确定目标偏移量；

基于所述目标偏移量移动所述第三参考图像，得到所述配准后的图像。

- 25 可选的，在将所述染色图像进行放大或缩小得到第三参考图像之后，所述方法还包括：

确定所述第三参考图像与基因表达可视化图像之间的相似度；

将相似度最大值对应的位置 作为粗配准结果。

可选的，将所述染色图像与所述基因表达可视化图像进行配准，得到配准后的图像包括：

- 30 获取染色图像和基因表达可视化图像，所述染色图像和所述基因表达可视化图像均包括多条轨迹线和多个轨迹点，所述轨迹点是由两条所述轨迹线相交形成的，每条所述轨迹线具有对应的索引号；

对所述染色图像进行预处理，获取候选图像集合；

确定所述候选图像集合对应的二值图集合；

- 35 基于所述二值图集合与所述基因表达可视化图像之间的相似度，确定所述二值图集合与所述基因表达可视化图像中具有对应关系的轨迹点对；

基于所述轨迹点对确定偏移信息；

基于所述偏移信息处理目标图像，以得到配准后的图像，所述目标图像为所述轨迹点对所对应的候选图像集合中的图像。

- 40 可选的，在对所述配准后的图像进行细胞分割，得到细胞掩模图之前，所述方法还包括：

对所述配准后的图像进行组织分割，得到组织掩模图。

可选的，所述对所述配准后的图像进行组织分割，得到组织掩模图包括：

45 获取所述配准后的图像 中待分割的目标组织的基因表达量并确定所述配准后的图像的基因表达矩阵，所述基因表达矩阵按照像素位置对所述基因表达量进行排列；

通过将所述基因表达矩阵中每一像素位置的基因表达量赋值于配准后的图像中对应的像素位置确定图像数据；

通过对所述图像数据进行图像处理以确定所述图像数据的前景像素点；

根据所述前景像素点对所述图像数据的连通区域进行标记得到所述组织掩模图。

- 50 可选的，在对所述配准后的图像进行组织分割，得到组织掩模图之后，所述方法还包括：

分别计算得到组织掩模图对应的灰度图像的外轮廓图像与内轮廓图像；

将所述外轮廓图像与所述内轮廓图像分别分为 N 段，并将每段所述外轮廓图像与每段所述内轮廓图像两两配对，得到 N 个图像组合，每个所述图像组合包括一段外轮廓图像与一段内轮廓图像；

5 分别计算每个所述图像组合中一段外轮廓图像与一段内轮廓图像的像素密度比值，得到 N 个所述像素密度比值，所述像素密度比值为每个所述图像组合中一段外轮廓图像与一段内轮廓图像的像素占比的比值，所述像素占比为像素点与总像素点的比值；

根据所述像素密度比值计算描述所述组织掩模图对应的灰度图像质量。

10 可选的，所述对所述配准后的图像进行细胞分割，得到细胞掩模图包括：

将所述配准后的图像输入所述图像分割模型，获得所述配准后的图像中各个像素点的分割结果，所述分割结果包括属于细胞、属于细胞轮廓或者图像背景；

将所述分割结果属于细胞或细胞轮廓的像素点确定属于细胞的像素点，输出所述配准后的图像的分割结果，得到所述细胞掩模图。

15 可选的，所述对所述配准后的图像进行细胞分割，得到细胞掩模图包括：

对所述基因表达可视化图像进行预处理，得到预处理图；

将所述预处理图进行二值化处理，得到初始掩模图；

根据所述初始掩模图，利用基于距离变换的分水岭算法对存在细胞粘连的连通域进行分割，得到分割后的所述细胞掩模图。

20 可选的，所述对所述配准后的图像进行细胞分割，得到细胞掩模图包括：

基于所述染色图像确定所述目标对象的各个组织类型对应的标注区域和细胞面积；

依据所述细胞面积确定每个所述标注区域包含的待处理子区域的面积，所述待处理子区域的面积与所述细胞面积正相关，所述待处理子区域用于表征对应的所述标注区域的最小处理单元；

25 依据每个所述待处理子区域的面积，将对应的所述标注区域进行网格化处理，得到每个所述标注区域对应的网格化坐标；

从所述基因表达可视化图像中提取所述网格化坐标对应的数据，并叠加至所述网格化坐标，得到所述目标对象的细胞处理结果。

30 可选的，所述根据所述细胞掩模图确定所述基因表达可视化图像中未知所属细胞的背景基因分子的所属细胞包括：

基于所述细胞掩模图，将所述基因表达可视化图像的每个基因表达特征分别确定为位于细胞内的细胞基因表达特征及位于细胞外的背景基因表达特征；

35 对每个所述细胞基因表达特征，通过空间位置和基因表达特征的信号值拟合出概率分布模型；

基于所述概率分布模型，确定出针对位于一目标细胞的预设范围内的目标背景基因表达特征属于所述目标细胞的概率值；

根据所述概率值输出校正结果以进行背景基因表达特征的校正。

40 可选的，所述根据所述细胞掩模图确定所述基因表达可视化图像中未知所属细胞的背景基因分子的所属细胞包括：

根据所述基因表达可视化图像，确定所属于目标细胞的初始基因分子以及未知所属细胞的背景基因分子；

获取所述初始基因分子的坐标信息；

根据所述坐标信息，确定所述目标细胞的中心点坐标；

45 根据所述中心点坐标和所述坐标信息，构建维诺图；

确定所述维诺图范围内的背景基因分子的所属细胞。

可选的，所述根据所述细胞掩模图确定所述基因表达可视化图像中未知所属细胞的背景基因分子的所属细胞包括：

50 确定所述基因表达可视化图像中所属于目标细胞的基因分子的初始分子数量以及所述所属于目标细胞的基因分子所占的目标区域的初始轮廓面积；

根据所述分子数量和所述初始轮廓面积，确定所述目标细胞的基因分子密度；

确定所述目标区域的预设区域范围内的背景基因分子的分子数量以及所述预设区域范围的区域面积；

根据所述背景基因分子的分子数量和区域面积，确定背景基因分子密度；

5 根据所述基因分子密度与所述背景基因分子密度确定所述背景基因分子的所属细胞。

可选的，所述根据所述细胞掩码图确定所述基因表达可视化图像中未知所属细胞的背景基因分子的所属细胞包括：

确定所述基因表达可视化图像中所属于目标细胞的基因分子所占的目标区域；

10 确定所述目标区域的预设校正范围内的未知所属细胞的背景基因分子的空间坐标及基因表达特征信号值；

将所述空间坐标和所述基因表达特征信号值输入概率分布模型，以得到所述背景基因分子所属于所述目标细胞的概率分数；所述概率分布模型根据所属于目标细胞的基因分子样本的空间坐标及基因表达特征信号值拟合得到；

根据所述概率分数确定所述背景基因分子的所属细胞。

15 可选的，所述对所属细胞标记包括：

以 ssDNA 染液和细胞边界荧光染色液对所属细胞进行染色，其中，所述 ssDNA 染液包括缓冲液和 ssDNA 试剂，所述细胞边界荧光染色液包括荧光染料。

可选的，所述方法还包括：

获取待测染色图像或基因表达可视化图像；以及

20 将所述待测染色图像或基因表达可视化图像输入有向目标检测网络，以获得目标检测结果，

其中所述有向目标检测网络包括：

主干 Backbone 层，被配置为对待测染色图像或基因表达可视化图像执行特征提取，以获得不同尺度的多个特征图；

25 颈部 Neck 层，被配置为从 Backbone 层接收所述多个特征图，并对所述多个特征图进行跨尺度融合，以获得多个融合特征图；以及

检测头 Head 层，被配置为对所述多个融合图像进行预测，以获得关于所述待测染色图像或基因表达可视化图像的目标检测结果，

30 其中所述目标检测结果包括：分类信息、目标框坐标预测信息、目标预测信息和角度预测信息。

可选的，所述染色图像所在的染色图像平面包括第一方向和与所述第一方向垂直的第二方向，所述染色图像包括横向于所述第二方向的多个第一直轨迹线和横向于所述第一方向的多个第二直轨迹线，

所述染色图像中的点坐标的确定方法包括：

35 采样步骤，包括：

应用第一采样条沿所述第一方向对所述染色图像进行第一方向等间隔采样以确定第一候选点集合，其中对于每次第一方向等间隔采样，执行以下子步骤：

40 将所述第一采样条在该次第一方向等间隔采样期间所选中的所述染色图像的采样区域内的像素按所述第一方向进行像素值累加求和以得到一系列第一像素值和；以及

45 确定所述一系列第一像素值和的极值点，并且对于确定的每个极值点：确定该次第一方向等间隔采样所获得的一个对应第一候选点，该对应第一候选点在所述第二方向上的坐标为该极值点所对应的第二方向位置在所述第二方向上的坐标，该对应第一候选点在所述第一方向上的坐标基于在该次第一方向等间隔采样期间所述第一采样条在所述第一方向上的位置确定，

其中，由每次第一方向等间隔采样所获得的第一候选点形成所述第一候选点集合；

应用第二采样条沿所述第二方向对所述染色图像进行第二方向等间隔采样以确定第二候选点集合，其中对于每次第二方向等间隔采样，执行以下子步骤：

50 将所述第二采样条在该次第二方向等间隔采样期间所选中的所述染色图像的采

样区域内的像素按所述第二方向进行像素值累加求和以得到一系列第二像素值和；以及

确定所述一系列第二像素值和的极值点，并且对于确定的每个极值点：确定该次第二方向等间隔采样所获得的一个对应第二候选点，该对应第二候选点在所述第一方向上的坐标为该极值点所对应的第一方向位置在所述第一方向上的坐标，该对应第二候选点在所述第二方向上的坐标基于在该次第二方向等间隔采样期间所述第二采样条在所述第二方向上的位置确定，

其中，由每次第二方向等间隔采样所获得的第二候选点形成所述第二候选点集合；

分组步骤，包括：

将所述第一候选点集合中的第一候选点分成多个第一候选点组；以及

将所述第二候选点集合中的第二候选点分成多个第二候选点组；

直线拟合步骤，包括：

对每个第一候选点组中的第一候选点进行直线拟合以得到一个相应的第一直线解析式；以及

对每个第二候选点组中的第二候选点进行直线拟合以得到一个相应的第二直线解析式；以及

坐标计算步骤，包括：

计算每个第一直线解析式与每个第二直线解析式的交点的坐标。

可选的，在对所属细胞标记，得到带细胞标记的空间组分子表达矩阵之后，所述方法还包括：

获取生物样本的第一图像，所述第一图像包括所述生物样本经染色的细胞边界；

获取所述生物样本的第二图像，所述第二图像包括所述生物样本经染色的遗传物质；

获取所述生物样本的基因表达量信息；

将所述第二图像与所述第一图像配准，得到第二图像与第一图像的配准图像；

将所述第二图像与所述基因表达量信息进行配准，得到第二图像与基因表达量信息的配准图像；

根据第二图像与第一图像的配准图像、第二图像与基因表达量信息的配准图像获取第一图像与基因表达量信息的配准图像；

根据第一图像与基因表达量信息的配准图像获取所述生物样本的细胞边界；

基于第一图像与基因表达量信息的配准图像根据所述细胞边界获取所述生物样本的细胞的基因表达信息。

第二方面，还提供一种空间组学单细胞数据获取装置，包括：

获取单元，用于获取生物样本的染色图像，获取所述生物样本的基因表达可视化图像；

配准单元，用于将所述染色图像与所述基因表达可视化图像进行配准，得到配准后的图像；

第一分割单元，用于对所述配准后的图像进行细胞分割，得到细胞掩模图；

修正单元，用于根据所述细胞掩模图确定所述基因表达可视化图像中未知所属细胞的背景基因分子的所属细胞；

标记单元，用于对所属细胞标记，得到带细胞标记的空间组分子表达矩阵。

可选的，所述获取单元，还用于：

通过显微镜获取所述生物样本的一个染色图像；或

通过显微镜获取所述生物样本的由多个子染色图像拼接的拼接染色图像；或

通过所述显微镜获取所述生物样本的多个子染色图像，对所述多个子染色图像进行拼接，得到拼接染色图像。

可选的，所述获取单元包括：

第一获取模块，用于通过显微镜获取所述生物样本的多个子染色图像；对所述生物样本的多个子染色图像进行第一拼接，得到第一染色图像；

判断模块，用于根据所述第一染色图像的图像质量判断是否对所述生物样本的多

个子染色图像进行第二拼接；

确定模块，用于若否，将所述第一染色图像确定为所述染色图像；

第二获取模块，用于若是，对所述生物样本的多个子染色图像进行第二拼接，得到所述染色图像。

5 可选的，所述第一获取模块，还用于：

分别确定多个子染色图像和模板图像的重叠区域；

将所述多个子染色图像和所述模板图像的被确定为重叠区域的部分分别裁剪为多个染色图像重叠子图和数量相同的多个模板图像重叠子图，以配对为多个重叠子图对；

10 通过对所述多个重叠子图对进行频域计算，分别获得各子染色图像和所述模板图像之间的偏移量；以及

将所述各子染色图像分别朝向所述模板图像的方向平移所述偏移量，并与所述模板图像拼接在一起。

可选的，所述第一获取模块，还用于：

15 获取所述多个子染色图像以及关于所述生物样本的轨迹线或轨迹点的模板信息；

对所述多个子染色图像进行预拼接得到所述多个子染色图像的预拼接坐标；

根据轨迹线或轨迹点的特征从所述多个子染色图像中选取至少一个第一参考图像，根据所述生物样本的模板信息和所述第一参考图像的预拼接坐标推导全局模板；以及

20 对于所述多个子染色图像中除了所述第一参考图像之外的第二参考图像，计算所述第二参考图像的预拼接坐标与所述第二参考图像在所述全局模板中的相应模板坐标之间的偏移量，并基于所述偏移量对所述第二参考图像进行坐标调整以拼接生成所述拼接染色图像。

可选的，所述配准单元包括：

25 获取模块，用于获取所述染色图像中的轨迹线，基于所述轨迹线与水平方向的夹角旋转所述染色图像；其中，所述染色图像和所述基因表达可视化图像包括多条轨迹线和/或多个轨迹点，所述轨迹点是由两条所述轨迹线相交形成的，所述多条轨迹线中的每条轨迹线具有对应的索引号；

30 处理模块，用于将所述染色图像进行放大或缩小得到第三参考图像，所述第三参考图像与所述基因表达可视化图像具有相同的比例；

第一确定模块，用于基于所述基因表达可视化图像和所述第三参考图像确定目标偏移量；

配准模块，用于基于所述目标偏移量移动所述第三参考图像，得到所述配准后的图像。

35 可选的，所述配准单元还包括：

第二确定模块，用于在所述处理模块将所述染色图像进行放大或缩小得到第三参考图像之后，确定所述第三参考图像与基因表达可视化图像之间的相似度；

粗配准模块，用于将相似度最大值对应的位置作为粗配准结果。

可选的，所述配准单元还用于：

40 获取染色图像和基因表达可视化图像，所述染色图像和所述基因表达可视化图像均包括多条轨迹线和多个轨迹点，所述轨迹点是由两条所述轨迹线相交形成的，每条所述轨迹线具有对应的索引号；

对所述染色图像进行预处理，获取候选图像集合；

确定所述候选图像集合对应的二值图集合；

45 基于所述二值图集合与所述基因表达可视化图像之间的相似度，确定所述二值图集合与所述基因表达可视化图像中具有对应关系的轨迹点对；

基于所述轨迹点对确定偏移信息；

基于所述偏移信息处理目标图像，以得到配准后的图像，所述目标图像为所述轨迹点对所对应的候选图像集合中的图像。

50 可选的，所述装置还包括：

第二分割单元，用于在所述第一分割单元对所述配准后的图像进行细胞分割，得

到细胞掩模图之前，对所述配准后的图像进行组织分割，得到组织掩模图。

可选的，所述第二分割单元还用于：

获取所述配准后的图像中待分割的目标组织的基因表达量并确定所述配准后的图像的基因表达矩阵，所述基因表达矩阵按照像素位置对所述基因表达量进行排列；

5 通过将所述基因表达矩阵中每一像素位置的基因表达量赋值于配准后的图像中对应的像素位置确定图像数据；

通过对所述图像数据进行图像处理以确定所述图像数据的前景像素点；

根据所述前景像素点对所述图像数据的连通区域进行标记得到所述组织掩模图。

可选的，所述装置还包括：

10 计算单元，用于在所述第二分割单元对所述配准后的图像进行组织分割，得到组织掩模图之后，分别计算得到组织掩模图对应的灰度图像的外轮廓图像与内轮廓图像；

配对单元，用于将所述外轮廓图像与所述内轮廓图像分别分为N段，并将每段所述外轮廓图像与每段所述内轮廓图像两两配对，得到N个图像组合，每个所述图像组合包括一段外轮廓图像与一段内轮廓图像；

15 所述计算单元，还用于分别计算每个所述图像组合中一段外轮廓图像与一段内轮廓图像的像素密度比值，得到N个所述像素密度比值，所述像素密度比值为每个所述图像组合中一段外轮廓图像与一段内轮廓图像的像素占比的比值，所述像素占比为像素点与总像素点的比值；

20 所述计算单元，还用于根据所述像素密度比值计算描述所述组织掩模图对应的灰度图像质量。

可选的，所述第一分割单元，还用于：

将所述配准后的图像输入所述图像分割模型，获得所述配准后的图像中各个像素点的分割结果，所述分割结果包括属于细胞、属于细胞轮廓或者图像背景；

25 将所述分割结果属于细胞或细胞轮廓的像素点确定属于细胞的像素点，输出所述配准后的图像的分割结果，得到所述细胞掩模图。

可选的，所述第一分割单元，还用于：

对所述基因表达可视化图像进行预处理，得到预处理图；

将所述预处理图进行二值化处理，得到初始掩模图；

30 根据所述初始掩模图，利用基于距离变换的分水岭算法对存在细胞粘连的连通域进行分割，得到分割后的所述细胞掩模图。

可选的，所述第一分割单元，还用于：

基于所述染色图像确定所述目标对象的各个组织类型对应的标注区域和细胞面积；

35 依据所述细胞面积确定每个所述标注区域包含的待处理子区域的面积，所述待处理子区域的面积与所述细胞面积正相关，所述待处理子区域用于表征对应的所述标注区域的最小处理单元；

依据每个所述待处理子区域的面积，将对应的所述标注区域进行网格化处理，得到每个所述标注区域对应的网格化坐标；

40 从所述基因表达可视化图像中提取所述网格化坐标对应的数据，并叠加至所述网格化坐标，得到所述目标对象的细胞处理结果。

可选的，所述修正单元，还用于：

基于所述细胞掩模图，将所述基因表达可视化图像的每个基因表达特征分别确定为位于细胞内的细胞基因表达特征及位于细胞外的背景基因表达特征；

45 对每个所述细胞基因表达特征，通过空间位置和基因表达特征的信号值拟合出概率分布模型；

基于所述概率分布模型，确定出针对位于一目标细胞的预设范围内的目标背景基因表达特征属于所述目标细胞的概率值；

根据所述概率值输出校正结果以进行背景基因表达特征的校正。

可选的，所述修正单元，还用于包括：

50 根据所述基因表达可视化图像，确定所属于目标细胞的初始基因分子以及未知所属细胞的背景基因分子；

获取所述初始基因分子的坐标信息；
根据所述坐标信息，确定所述目标细胞的中心点坐标；
根据所述中心点坐标和所述坐标信息，构建维诺图；
确定所述维诺图范围内的背景基因分子的所属细胞。

5 可选的，所述修正单元，还用于包括：

确定所述基因表达可视化图像中所属于目标细胞的基因分子的初始分子数量以及所述所属于目标细胞的基因分子所占的目标区域的初始轮廓面积；

根据所述分子数量和所述初始轮廓面积，确定所述目标细胞的基因分子密度；

10 确定所述目标区域的预设区域范围内的背景基因分子的分子数量以及所述预设区域范围的区域面积；

根据所述背景基因分子的分子数量和区域面积，确定背景基因分子密度；

根据所述基因分子密度与背景基因分子密度确定所述背景基因分子的所属细胞。

可选的，所述修正单元，还用于包括：

15 确定所述基因表达可视化图像中所属于目标细胞的基因分子所占的目标区域；

确定所述目标区域的预设校正范围内的未知所属细胞的背景基因分子的空间坐标及基因表达特征信号值；

20 将所述空间坐标和所述基因表达特征信号值输入概率分布模型，以得到所述背景基因分子所属于所述目标细胞的概率分数；所述概率分布模型根据所属于目标细胞的基因分子样本的空间坐标及基因表达特征信号值拟合得到；

根据所述概率分数确定所述背景基因分子的所属细胞。

可选的，所述标记单元，还用于：

25 以 ssDNA 染液和细胞边界荧光染色液对所属细胞进行染色，其中，所述 ssDNA 染液包括缓冲液和 ssDNA 试剂，所述细胞边界荧光染色液包括荧光染料。

可选的，所述装置还包括检测单元，包括：

获取待测染色图像或基因表达可视化图像；以及

30 将所述待测染色图像或基因表达可视化图像输入有向目标检测网络，以获得目标检测结果，

其中所述有向目标检测网络包括：

35 主干 Backbone 层，被配置为对待测染色图像或基因表达可视化图像执行特征提取，以获得不同尺度的多个特征图；

颈部 Neck 层，被配置为从 Backbone 层接收所述多个特征图，并对所述多个特征图进行跨尺度融合，以获得多个融合特征图；以及

40 检测头 Head 层，被配置为对所述多个融合图像进行预测，以获得关于所述待测染色图像或基因表达可视化图像的目标检测结果，

其中所述目标检测结果包括：分类信息、目标框坐标预测信息、目标预测信息和角度预测信息。

45 可选的，所述染色图像所在的染色图像平面包括第一方向和与所述第一方向垂直的第二方向，所述染色图像包括横向于所述第二方向的多个第一直轨迹线和横向于所述第一方向的多个第二直轨迹线，

所述染色图像中的点坐标的确定方法包括：

采样步骤，包括：

应用第一采样条沿所述第一方向对所述染色图像进行第一方向等间隔采样以确定第一候选点集合，其中对于每次第一方向等间隔采样，执行以下子步骤：

45 将所述第一采样条在该次第一方向等间隔采样期间所选中的所述染色图像的采样区域内的像素按所述第一方向进行像素值累加求和以得到一系列第一像素值和；以及

50 确定所述一系列第一像素值和的极值点，并且对于确定的每个极值点：确定该次第一方向等间隔采样所获得的一个对应第一候选点，该对应第一候选点在所述第二方向上的坐标为该极值点所对应的第二方向位置在所述第二方向上的坐标，该对应第一候选点在所述第一方向上的坐标基于在该次第一方向等间隔采样期间所述第一采样条

在所述第一方向上的位置确定，

其中，由每次第一方向等间隔采样所获得的第一候选点形成所述第一候选点集合；

5 应用第二采样条沿所述第二方向对所述染色图像进行第二方向等间隔采样以确定第二候选点集合，其中对于每次第二方向等间隔采样，执行以下子步骤：

将所述第二采样条在该次第二方向等间隔采样期间所选中的所述染色图像的采样区域内的像素按所述第二方向进行像素值累加求和以得到一系列第二像素值和；以及

10 确定所述一系列第二像素值和的极值点，并且对于确定的每个极值点：确定该次第二方向等间隔采样所获得的一个对应第二候选点，该对应第二候选点在所述第一方向上的坐标为该极值点所对应的第一方向位置在所述第一方向上的坐标，该对应第二候选点在所述第二方向上的坐标基于在该次第二方向等间隔采样期间所述第二采样条在所述第二方向上的位置确定，

15 其中，由每次第二方向等间隔采样所获得的第二候选点形成所述第二候选点集合；

分组步骤，包括：

将所述第一候选点集合中的第一候选点分成多个第一候选点组；以及

将所述第二候选点集合中的第二候选点分成多个第二候选点组；

直线拟合步骤，包括：

20 对每个第一候选点组中的第一候选点进行直线拟合以得到一个相应的第一直线解析式；以及

对每个第二候选点组中的第二候选点进行直线拟合以得到一个相应的第二直线解析式；以及

坐标计算步骤，包括：

25 计算每个第一直线解析式与每个第二直线解析式的交点的坐标。

可选的，所述装置还包括：分析单元，所述分析单元用于：

获取生物样本的第一图像，所述第一图像包括所述生物样本经染色的细胞边界；

30 获取所述生物样本的第二图像，所述第二图像包括所述生物样本经染色的遗传物质；

获取所述生物样本的基因表达量信息；

将所述第二图像与所述第一图像配准，得到第二图像与第一图像的配准图像；

将所述第二图像与所述基因表达量信息进行配准，得到第二图像与基因表达量信息的配准图像；

35 根据第二图像与第一图像的配准图像、第二图像与基因表达量信息的配准图像获取第一图像与基因表达量信息的配准图像；

根据第一图像与基因表达量信息的配准图像获取所述生物样本的细胞边界；

基于第一图像与基因表达量信息的配准图像根据所述细胞边界获取所述生物样本的细胞的基因表达信息。

根据本申请的第三方面，提供了一种电子设备，包括：

40 至少一个处理器；以及

与所述至少一个处理器通信连接的存储器；其中，

所述存储器存储有可被所述至少一个处理器执行的指令，所述指令被所述至少一个处理器执行，以使所述至少一个处理器能够执行前述第一方面所述的方法。

45 根据本申请的第四方面，提供了一种存储有计算机指令的非瞬态计算机可读存储介质，其中，所述计算机指令用于使所述计算机执行前述第一方面所述的方法。

根据本申请的第五方面，提供了一种计算机程序产品，包括计算机程序，所述计算机程序在被处理器执行时实现如前述第一方面所述的方法。

50 本申请提供的空间组学单细胞数据获取方法、装置、电子设备和存储介质，获取生物样本的染色图像，获取所述生物样本的基因表达可视化图像；将所述染色图像与所述基因表达可视化图像进行配准，得到配准后的图像；对所述配准后的图像进行细胞分割，得到细胞掩模图；根据所述细胞掩模图确定所述基因表达可视化图像中未知

所属细胞的背景基因分子的所属细胞，并对所属细胞标记，得到带细胞标记的空间组分子表达矩阵。与相关技术相比，本申请实施例通过基因表达矩阵图、拼接、配准、分割和校正的处理流程，以获得高精度的空间单细胞数据。

应当理解，本部分所描述的内容并非旨在标识本申请的实施例的关键或重要特征，也不用于限制本申请的范围。本申请的其它特征将通过以下的说明书而变得容易理解。

附图说明

附图用于更好地理解本方案，不构成对本申请的限定。其中：

图 1 为本申请实施例所提供的一种空间组学单细胞数据获取方法的流程示意图；

图 2 为本申请实施例提供的一种执行空间组学单细胞数据的获取示意图；

图 3 为本申请实施例提供的一种获取生物样本的染色图像方法的流程示意图；

图 4 为本申请实施例提供的一种对多个子染色图像进行拼接方法的流程示意图；

图 5 为本申请实施例提供的另一种对多个子染色图像进行拼接方法的流程示意图；

图 6 为本申请实施例提供的一种多个视场在空间上的分布的示意图；

图 7 为本申请实施例提供的一种图像拼接方法的预拼接结果示意图；

图 8a 为本申请实施例提供的一种预拼接时在水平方向上扫描得到的行向相邻的两个染色图像之间的位置关系示意图；

图 8b 为本申请实施例提供的一种预拼接时在竖直方向上扫描得到的列向相邻的两个染色图像之间的位置关系示意图；

图 9 为本申请实施例提供的一种用于拼接生物样本的单个视场模板的示意图；

图 10 为本申请实施例提供的一种用于拼接生物样本的单个视场模板的构造过程示意图；

图 11 为本申请实施例提供的一种将染色图像与基因表达可视化图像进行配准的方法流程示意图；

图 12 为本申请实施例提供的一种染色图像的示意图；

图 13 为本申请实施例提供的一种第三参考图像的示意图；

图 14a 为本申请实施例提供的一种具有多个周期的第三参考图像，

图 14b 为本申请实施例提供的一种多个周期的基因表达可视化图像；

图 15 为本申请实施例提供的一种以第一偏移量为中心所确定的预设周期的示意图；

图 16 为本申请实施例提供的另一种将染色图像与基因表达可视化图像进行配准的方法流程示意图；

图 17a 为本申请实施例提供的一种候选图像的示意图；

图 17b 为本申请实施例提供的一种初始二值图的示意图；

图 17c 为本申请实施例提供的一种目标二值图的示意图；

图 18 为本申请实施例提供的另一种将染色图像与基因表达可视化图像进行配准的方法流程示意图；

图 19 为本申请实施例提供的一种样本图像的组织分割方法的确定图像数据的流程图；

图 20 为本申请实施例提供的一种样本图像的组织分割方法的图像数据示意图；

图 21 为本申请实施例提供的一种样本图像的组织分割方法的 2d 卷积示意图；

图 22 为本申请实施例提供的一种样本图像的组织分割方法的低级特征示意图；

图 23 为本申请实施例提供的一种计算描述组织掩模图对应的灰度图像质量的方法流程示意图；

图 24 为本申请实施例提供的一种外部轮廓图像与内部轮廓图像的关系结构示意图；

图 25 为本申请实施例提供的另一种外部轮廓图像与内部轮廓图像的关系结构示意图；

图 26 为本申请实施例提供的一种对配准后的图像进行细胞分割的方法流程示意图

图；

图 27 为本申请实施例提供的一种图像分割模型在鼠脑图像上的分割效果；

图 28 为本申请实施例提供的另一种对配准后的图像进行细胞分割的方法流程示意图；

5 图 29 示出了本申请实施例提供的另一种细胞分割的处理方法的流程示意图；

图 30 示出了本申请实施例提供的基于本实施例方法的示例流程示意图；

图 31 示出了本申请实施例提供的基因表达图的效果示例图；

图 32 示出了本申请实施例提供的经锐化处理得到锐化图的效果示例图；

图 33 示出了本申请实施例提供的初始掩模图的效果示意图；

10 图 34 示出了本申请实施例提供的输出掩模图的效果示意图；

图 35 为本申请实施例提供的另一种对配准后的图像进行细胞分割的方法流程示意图；

图 36 为本申请实施例提供的一种确定基因表达可视化图像中未知所属细胞的背景基因分子的所属细胞的方法流程示意图；

15 图 37 为本申请实施例提供的一种 DNB 与细胞膜的对应关系示意图；

图 38 为本申请实施例提供的一种校正背景 DNB 的校正结果示意图；

图 39 为本申请实施例提供的另一种确定基因表达可视化图像中未知所属细胞的背景基因分子的所属细胞的方法流程示意图；

20 图 40 为本申请实施例提供的另一种确定基因表达可视化图像中未知所属细胞的背景基因分子的所属细胞的方法流程示意图；

图 41 为本申请实施例提供的一种目标细胞的中心点的示意图；

图 42 为本申请实施例提供的另一种确定基因表达可视化图像中未知所属细胞的背景基因分子的所属细胞的方法流程示意图；

25 图 43 为本申请实施例提供的一种染色图像中的点坐标的确定方法的流程示意图；

图 44 为本申请实施例提供的一种染色图像的示意图；

图 45 为本申请实施例提供的一种有向目标检测方法流程示意图；

图 46 为本申请实施例提供的一种生物样本的细胞基因表达信息获取方法流程示意图；

30 图 47 为本申请实施例提供的一种空间组学单细胞数据获取装置的结构示意图；

图 48 为本申请实施例提供的另一种空间组学单细胞数据获取装置的结构示意图；

图 49 为本申请实施例提供的示例电子设备的示意性框图。

具体实施方式

35 以下结合附图对本申请的示范性实施例做出说明，其中包括本申请实施例的各种细节以助于理解，应当将它们认为仅仅是示范性的。因此，本领域普通技术人员应当认识到，可以对这里描述的实施例做出各种改变和修改，而不会背离本申请的范围和精神。同样，为了清楚和简明，以下的描述中省略了对公知功能和结构的描述。

下面参考附图描述本申请实施例的空间组学单细胞数据获取方法、装置、电子设备和存储介质。

40 图 1 为本申请实施例所提供的一种空间组学单细胞数据获取方法的流程示意图。

如图 1 所示，该方法包含以下步骤：

步骤 101，获取生物样本的染色图像，获取所述生物样本的基因表达可视化图像；

依据空间组学技术 (Spatial resolved technology, SOP)，对生物样本进行染色拍图，由显微镜拍摄时空芯片上的生物样本而得到的，得到染色图像。

45 作为本申请实施例的一种可行方式，染色可以通过但不限于下述技术实现，例如：苏木素伊红染色 (Hematoxylin and Eosin Staining, HE)、ssDNA 及 DAPI，具体的，本申请实施例不做限定。

获取的显微镜拍照图(染色图像),一般为tif格式,包含但不限于16bit或8bit,彩色图和灰度图均可,具体的,本申请实施例不进行限定。

将生物样本进行空间测序,并通过测序序列比对分析等标准流程获取到生物样本的基因表达矩阵,并将基因表达矩阵转换成基因表达可视化图像。

5 时空芯片上有用于采集生物样本基因的规则排列的位点,每个位点可以捕获生物样本的基因序列,通过基因测序可以获得每个位点的基因序列或碱基的数量,然后将该位点的基因序列的数量作为基因表达矩阵中元素的值,从而使基因表达矩阵的元素与时空芯片的位点相对应(如一个元素对应一个位点,基因表达矩阵的第一行第一列的元素对应时空芯片左上角的位点)依次类推就可以得到基因表达矩阵,然后根
10 据基因表达矩阵绘制基因表达可视化图像。其中,基因表达可视化图像中的像素点对应基因表达矩阵的元素,像素点的像素值对应元素的值,所以基因表达可视化图像包含同一生物样本的空间信息。

有关测序序列比对分析的实现方法可参阅相关技术中的方法,本申请实施例在此不再进行一一赘述。

15 获取所述生物样本的基因表达可视化图像,包含但不限于以下方式:对所述生物样本进行空间组学测序序列处理,得到基因表达矩阵;将所述基因表达矩阵转换为对应的基因表达可视化图像。

步骤102,将所述染色图像与所述基因表达可视化图像进行配准,得到配准后的图像;

20 在一些实施例中,基于拼接染色图像中的线特征实现配准;

或者,在一些实施例中,通过图像上的轨迹点(track点),以及利用组织轮廓作为参照物,通过坐标计算实现在像素级别精度的准确自动配准;

或者,在一些实施例中,通过拼接染色图像的组织掩模实现配准。具体的,本申请实施例对配准方式不进行限定。

25 步骤103,对所述配准后的图像进行细胞分割,得到细胞掩模图;

将配准后的图像输入图像分割模型,以输出细胞掩模图,该细胞掩模图中包含细胞核。

作为本申请实施例的一种可行方式,在执行细胞分割之前,还可以先执行组织分割,做完组织分割后可以只关注组织区域不关注组织外,提升样本数据的信噪比。

30 步骤104,根据所述细胞掩模图确定所述基因表达可视化图像中未知所属细胞的背景基因分子的所属细胞,并对所属细胞标记,得到带细胞标记的空间组分子表达矩阵;

步骤103中在完成细胞分割后,至少包含细胞核,但是还存在一些背景基因分子未被标记为一个细胞的场景。

35 对未分配到细胞的DNA纳米球(DNB)位置上的背景基因进行重分配。以该位置为中心,圈选一定的区域(例如100*100pixel),对其中包含的细胞分别进行高斯混合模型拟合,假设,其中包含N个细胞,那对于该位置会有N个概率值,选取最大的概率的细胞作为该位置归属的细胞。每个未归属的有表达的位置都进行以上操作。通过二次修正,能够提升信噪比,缓解细胞分割,配准等的精度问题。

40 实际应用中,为了防止过分配,在选取最大的概率的细胞作为归属细胞时通过设定概率阈值的方式,实现细胞归属标记,当概率太低时,不对归属的细胞进行标记。具体的对概率阈值不做限定。

图1已说明空间组学单细胞数据获取方法,参阅图2,图2为本申请实施例提供的一种执行空间组学单细胞数据的获取示意图,其中,该图是以鼠脑切片为例进行的示意,其中A对应步骤101,C对应步骤102,D对应步骤103,E对应步骤104。

45 获取生物样本的染色图像,获取所述生物样本的基因表达可视化图像;将所述染色图像与所述基因表达可视化图像进行配准,得到配准后的图像;对所述配准后的图像进行细胞分割,得到细胞掩模图;根据所述细胞掩模图确定所述基因表达可视化图像中未知所属细胞的背景基因分子的所属细胞,并对所属细胞标记,得到带细胞标记
50 的空间组分子表达矩阵。与相关技术相比,本申请实施例通过基因表达矩阵图、拼接、配准、分割和校正的处理流程,以获得高精度的空间单细胞数据。

作为本申请实施例的一种可行方式，在步骤 101 中执行获取生物样本的染色图像时，可以通过但不局限于以下三种方式获取：

方式一：通过显微镜获取所述生物样本的一个染色图像；

方式二：通过显微镜获取所述生物样本的由多个子染色图像拼接的拼接染色图像；

方式三：通过所述显微镜获取所述生物样本的多个子染色图像，对所述多个子染色图像进行拼接，得到拼接染色图像。

其中方式一和方式二是显微镜下执行拼接的结果，方式一是直接通过显微镜获取到一个大图（染色图像），方式二是在获取多个小图（子染色图像）后，采用显微镜对多个子染色图像进行拼接，得到染色图像；方式三是在显微镜拼接的结果不符合需求时，通过预设算法进行拼接所得到的染色图像的拼接结果。在具体本申请实施例对染色图像的获取方式不进行限定。

在一些实施例，如图 3 所示，图 3 为本申请实施例所提供的获取生物样本的染色图像方法的流程示意图。包括：

步骤 1011，通过显微镜获取所述生物样本的多个子染色图像；对所述生物样本的多个子染色图像进行第一拼接，得到第一染色图像。

可以采用但不局限于以下方式执行拼接，如图 4 所示，包括：

步骤 10111，分别确定多个子染色图像和模板图像的重叠区域；

在一些示例中，确定子染色图像和模板图像的重叠区域可以包括：确定子染色图像与模板图像的匹配方向；以及在所确定的匹配方向上，在子染色图像和模板图像中选择相邻的相同尺寸的区域来作为重叠区域。

例如，可以基于在拍摄该两个图像时拍摄设备的移动方向来确定匹配方向（或称为拼接方向）。如子染色图像和模板图像是拍摄设备在水平移动时顺序拍摄的两个图像，则该两个图像之间的匹配方向可以确定为水平方向。当然，也可以采用其他可用的方式来确定该两个图像之间的相对位置关系，并从而确定匹配方向，在此不再赘述。

在一些示例中，重叠区域被选择为使得子染色图像和模板图像中的实际重叠部分在重叠区域中的比率大于预定阈值。

在一些示例中，该预定阈值被确定为 10%，或 30%等。需要注意的是，该值仅是被用作示例的值。在其他一些示例中，预定阈值也可以取更大或更小的其他值。

请继续参阅图 2，该步骤对应图 2 中的步骤 B。

步骤 10112，将所述多个子染色图像和所述模板图像的被确定为重叠区域的部分分别裁剪为多个染色图像重叠子图和数量相同的多个模板图像重叠子图，以配对为多个重叠子图对；

在一些示例中，该操作可以包括：将子染色图像和模板图像的被确定为重叠区域的部分分别进行降噪和/或加窗处理；在垂直于匹配方向的方向上，将经过降噪和/或加窗处理的子染色图像和模板图像进行合并，并以相同的图像裁剪大小对子染色图像和模板图像的进行降噪和/或加窗处理的部分分别进行重采样，以获得该多个子染色图像重叠子图和该多个模板图像重叠子图；以及在垂直于匹配方向的方向上，将该多个子染色图像重叠子图和该多个模板图像重叠子图顺序地一一配对，以形成该多个重叠子图对。

上述操作中的降噪处理可通过例如高斯滤波来进行，但图像处理领域中所使用的其他降噪方式也可用于该处理。加窗处理可获得抑制频谱泄露的图像。所使用的窗口包括但不限于 Hanning 窗口、Hamming 窗口等。

在本申请的一些示例中，可采用滑窗的方式来执行上述的重采样。滑窗方式可设定一定的重叠区域，多个堆叠的小子图对应保证源图和模板图的裁剪区域位于重叠区域且位于同一坐标系位置。可使用任何可行的方式来实现重采样，例如利用目标检测方式裁剪特征较多区域等，而不同于上述的滑窗方式。

在获得多个重叠子图对之后，图 1 所示的方法还可以可选地包括从该多个重叠子图对中去特征少于特定标准和/或噪声大于特定标准的重叠子图对。例如，可用利用梯度检测来过滤特征较少和/或噪声较多的子图对。所使用的梯度检测方法包括但不限于 Sobel、Canny 等方法。此外，其他任何能够用于检测图像中的特征数量和/或噪声

的方法都能够用在此处，而不限于梯度检测方法。

步骤 10113，通过对所述多个重叠子图对进行频域计算，分别获得各子染色图像和所述模板图像之间的偏移量；以及

5 可选地，在此进行频域的叠加共振的重叠子图对是上述可选步骤中去除了特征较少和/或噪声较多的子图对之后得到的重叠子图对。

10 在一些示例中，该步骤可以采用狄拉克函数来执行。例如，包括：对该多个重叠子图对进行互功率频谱的堆叠，并利用狄拉克函数来从堆叠的互功率频谱中获得偏移量。具体地，在一些示例中，可以将该多个重叠子图对中的每个重叠子图从位置坐标系变换到频域。该变换例如可以通过二维图像离散傅里叶变换来完成，或通过能实现域变换的任何其他方式来完成，在此不再赘述。然后，可以计算该多个重叠子图对中的每个重叠子图对所对应的互功率频谱；对所获得的每个重叠子图对所对应的互功率频谱执行离散傅里叶逆变换，以获得重叠区域的狄拉克函数表示；以及将获得的所述重叠区域中的狄拉克函数表示的最大峰所在的坐标确定为偏移量。

15 在一些示例中，在将所获得的每个重叠子图对所对应的互功率频谱进行累加之前，针对所获得的每个重叠子图对，还可以改善该重叠子图对中的特征，以使其更加明显。例如，可以将该重叠子图对变换到位置坐标系，并对变换到位置坐标系的重叠子图进行加窗处理，以增强该重叠子图中的特征。这里所使用的窗口函数可以是汉宁窗 Hanning 窗或汉明窗 Hamming 窗，或者是可增强图像中的特征的任何其他函数。增强重叠子图中的特征可以有效提高特征不明显图像的图像配准的成功率。在此，所使用的增强图像特征的方式可以不限于加窗处理，也可以是能够增强图像特征的任何其他方法。

20 在一些实施例，对所获得的每个重叠子图对所对应的互功率频谱执行离散傅里叶逆变换，以获得所述重叠区域的狄拉克函数表示可以包括：对所获得的每个重叠子图对所对应的互功率频谱执行离散傅里叶逆变换以获得堆叠的狄拉克函数表示；以及将所获得的堆叠的狄拉克函数表示进行加权累加，以获得重叠区域的狄拉克函数表示。在该方案中，具有更明显特征的重叠子图对可被赋予更大的加权系数。例如，每个重叠子图对的加权系数可以取决于该重叠子图对的能量频谱（例如，狄拉克函数的最大峰的能量强度）和/或该重叠子图对的特征点的匹配数和/或该重叠子图对的图像清晰度，或者可基于能够表征重叠子图对是否具有更明显特征的其他任何度量。

30 步骤 10114，将所述各子染色图像分别朝向所述模板图像的方向平移所述偏移量，并与所述模板图像拼接在一起。

步骤 1012，根据所述第一染色图像的图像质量判断是否对所述生物样本的多个子染色图像进行第二拼接；

35 在一些实施例，第一染色图像的图像质量需要根据空间组学单细胞数据获取需求进行设定，本申请实施例不进行限定。

若否，则执行步骤 1013；若是，则执行步骤 1014。

步骤 1013，将所述第一染色图像确定为所述染色图像；

步骤 1014，对所述生物样本的多个子染色图像进行第二拼接，得到所述染色图像。

第二拼接方式的具体实现过程可参阅图 4 的相关说明，故在此不再进行赘述。

40 如图 5 所示，图 5 为本申请实施例提供的另一种对多个子染色图像进行拼接，得到拼接染色图像的方法流程图，包括：

步骤 10115，获取所述多个子染色图像以及关于所述生物样本的轨迹线或轨迹点的模板信息；

45 在测序过程中，对单个生物样本的局部可以依次拍摄若干幅图像。每个视野的图像的水平方向与竖直方向上可以各分布有若干条轨迹线，轨迹线（track line）的交点为轨迹点（track cross，或者称“轨迹交叉点”）。轨迹线和轨迹点可以相互转换；根据交叉的两个轨迹线可以得到轨迹点；根据共线的轨迹点可以得到轨迹线。所述模板信息可以包括所述生物样本上的多条轨迹线的索引号。如本领域技术人员已知的，生物样本上的轨迹线的分布规律在生物样本设计之初就已经被规定好，举例来说，假设生物样本的竖直方向上有 10 条轨迹线，它们之间的间距可以分别为例如 1、2、3、4、5、4、3、2、1。假设以第一条轨迹线所在的位置为起始点（坐标为 0），则上述 10

条轨迹线的横轴（x 轴）坐标分别是 0、1、3、6、10、15、19、22、24、25。生物样本的模板信息中的索引号可以与生物样本中的轨迹线的排布相关联，上述 10 条轨迹线的索引号可以人为规定为分别是 0、1、2、3、4、5、6、7、8、0。

此外，所获取的生物样本的模板信息还可以包括周期性，例如对于上述生物样本的
5 的竖直方向上的 10 条轨迹线，它们可以向左和向右分别扩展一个或多个周期。例如，左右扩展各一个周期所形成的轨迹线之间的间距从左到右依次为 1、2、3、4、5、4、3、2、1、1、2、3、4、5、4、3、2、1，由此可以得到扩展后的 28 条轨迹线的横轴（x 轴）坐标分别是 -25、-24、-22、-19、-15、-10、-6、-3、-1、0、1、3、6、10、15、19、22、24、25、26、28、31、35、40、44、47、49、
10 50，它们对应的索引号分别为：0、1、2、3、4、5、6、7、8、0、1、2、3、4、5、6、7、8、0、1、2、3、4、5、6、7、8、0、1、2、3、4、5、6、7、8、0。

值得注意的是，一般而言，某条轨迹线对应的生物样本的索引号是惟一的，而生物样本上的某个索引号会对应多条轨迹线，这是生物样本上的轨迹线设计时的周期性决定的。

15 显微镜可以例如通过对生物样本进行逐行扫描或逐列扫描来拍摄生物样本的图像。扫描后多个图像在空间上的分布可以如图 6 所示，所述多个图像可以共有例如 $n+1$ 列， $m+1$ 行，包括从第一行第一列的图像 (r_0, c_0) 到最后一行最后一列的图像 (r_m, c_n) 。对于如图 3 中所例示的图像 (r_{m-1}, c_{n-1}) ，该图像最多可以有 8 个相邻的图像。另外，显微镜在拍摄图像时也可以记录下拍摄时的相机位置信息，后续可以根据任一图像记录的拍摄时的相机位置信息，确定该图像的对应位置并且为该图像确定其相邻的
20 图像。

图 7 是根据本发明的一个实施方案的图像拼接方法的预拼接结果示意图。

步骤 10116，对所述多个子染色图像进行预拼接得到所述多个子染色图像的预拼接坐标；

25 可以通过对多个染色图像进行扫描，根据扫描顺序或者图像拍摄时的相机位置信息采用例如 2-近邻的方式确定当前染色图像的左邻居和右邻居的图像，并进而确定染色图像和左右相邻图像之间的偏移量和重叠量。在扫描过程中，结合生物样本的模板信息可以优选地采用快速傅里叶变换（FFT）或者尺度不变特征变换（SIFT）获取相邻染色图像之间的偏离量以及重叠量。基于扫描图像过程中获取的相邻染色图像之间的
30 偏离量和重叠量，可以得到所述多个染色图像的预拼接坐标。

图 8a 是预拼接时在水平方向上扫描得到的行向相邻的两个染色图像之间的位置关系示意图；图 8b 是预拼接时在竖直方向上扫描得到的列向相邻的两个染色图像之间的位置关系示意图。

具体地，图 8a 和 8b 分别示出了在水平方向和竖直方向对染色图像进行扫描后得到的相邻染色图像之间的偏离量和重叠量。对于水平方向扫描来说（参考图 8a），所得到的行向相邻的两个图像在水平方向重叠的距离被称为水平方向上的重叠量（overlap），在竖直方向偏离的距离被称为竖直方向上的偏离量（offset）。类似的，对于竖直方向扫描来说（参考图 8b），所得到的列向相邻的两个染色图像在竖直方向重叠的距离被称为竖直方向上的重叠量（overlap），在水平方向偏离的距离被称为水平方向上的偏离量（offset）。水平方向和竖直方向扫描分别完成后可以得到所述多个染色图像在对应方向上的偏离量矩阵 A 及重叠量矩阵 B。利用偏离量矩阵 A 和重叠量矩阵 B，可映射出各个染色图像在空间上的坐标，从而完成如图 7 所示出的多个染色图像的预拼接图像。

但是，仅对染色图像进行预拼接处理并不能达到像素级别的精度，因此还需要在此基础上在 S10117 和 S10118 步骤中继续对染色图像的坐标进行进一步精确调整。

步骤 10117，根据轨迹线或轨迹点的特征从所述多个子染色图像中选取至少一个第一参考图像，根据所述生物样本的模板信息和所述第一参考图像的预拼接坐标推导全局模板；以及

50 所选取的所述第一参考图像（或者说基准图像）可以是所述多个染色图像中轨迹线特征最明显的图像或者轨迹点最多的图像。所述第一参考图像例如是图 6 和图 7 中所突出显示的图像 (r_{m-1}, c_{n-1}) ，然后对所选取的该图像 (r_{m-1}, c_{n-1}) 进行轨迹

线检测，并将检测到的相邻轨迹线之间的间距与上面提到的芯片设计时所规定好的轨迹线之间的间距进行匹配，进而确定所选取的该第一参考图像中的轨迹线对应的索引号。也就是说，根据所述第一参考图像中的相邻轨迹线之间的距离可以确定所述第一参考图像中的各个轨迹线所对应的索引号。然后，可以基于所述第一参考图像的所述索引号以及所述第一参考图像的预拼接坐标推导出全局模板。

在一个实施例中，由于生物样本上的组织切片会覆盖轨迹线，而部分轨迹点可以在组织空腔的位置显露出来，因此在一个图像内，完整轨迹点相较于完整轨迹线出现的概率会更高，因此也可以优选地选择轨迹点最多的图像作为第一参考图像。

在一个实施例中，根据所述第一参考图像中的相邻轨迹线之间的距离确定所述第一参考图像中的各个轨迹线所对应的索引号，并且基于所述第一参考图像的所述索引号以及所述第一参考图像的预拼接坐标推导全局模板。

图 9 是示出根据本发明的一个实施方案的用于拼接生物样本的全局模板的示意图；图 10 是根据本发明的一个实施方案的用于拼接生物样本的全局模板的构造过程示意图。

参考图 10 所示出的全局模板的构造示例，根据得到的所选取的第一参考图像 (rm-1, cn-1) 在预拼接阶段的位置坐标以及轨迹线的索引号，可以通过沿水平方向进行左右双向周期扩展以及在竖直方向进行上下双向周期扩展得到以第一参考图像 (rm-1, cn-1) 为基准的全局模板（或者称全局参考系）。

对于除所选取的第一参考图像 (rm-1, cn-1) 之外的第二参考图像的坐标的调整，可以分为两类，即对于包含轨迹点的第二参考图像进行坐标调整以及对于不包含轨迹点的第二参考图像进行坐标调整。

对于包含轨迹点的第二参考图像进行坐标调整是通过将每个所述包含轨迹点的第二参考图像的预拼接坐标与该包含轨迹点的第二参考图像的各自的唯一偏移量相加。所述第二参考图像的所述偏移量是该第二参考图像上的轨迹点指向所述全局模板中的与该第二参考图像上的轨迹点对应的模板点所形成的矢量。可参考图 9 在局部放大的示意图中用箭头所示出的偏移量，该偏移量表示从生物样本上的轨迹点指向推导出的全局模板上的与该轨迹点对应的模板线轨迹点的矢量。

步骤 10118，对于所述多个子染色图像中除了所述第一参考图像之外的第二参考图像，计算所述第二参考图像的预拼接坐标与所述第二参考图像在所述全局模板中的相应模板坐标之间的偏移量，并基于所述偏移量对所述第二参考图像进行坐标调整以拼接生成所述拼接染色图像。

获取每个所述包含轨迹点的第二参考图像中的每个轨迹点的各自的偏移量，对于每个所述第二参考图像选取所述偏移量中的具有中值角度或中值距离的偏移量作为该第二参考图像的所述唯一偏移量；或者，

获取每个所述包含轨迹点的第二参考图像中的每个轨迹点的各自的偏移量，根据所述偏移量的矢量长度对所述偏移量进行排序，选择预设排序位置的偏移量作为该第二参考图像的所述唯一偏移量。

将每个所述包含轨迹点的第二参考图像的唯一偏移量与所述包含轨迹点的第二参考图像的各自的唯一偏移量相加，可以得到包含轨迹点的第二参考图像的调整后的坐标。

对于不包含轨迹点的第二参考图像进行坐标调整可以采用近邻调整法，近邻调整法包括：对于每个所述不包含轨迹点的第二参考图像，根据距离该不包含轨迹点的第二参考图像最近的包含轨迹点的第二参考图像的唯一偏移量以及该不包含轨迹点的第二参考图像与所述最近的包含轨迹点的第二参考图像之间的偏离量和重叠量，对该不包含轨迹点的第二参考图像进行坐标调整。

具体而言，由于组织覆盖、芯片特征被污染以及显微相机的拍摄质量等问题，存在一些其上并没有轨迹点的图像，在这种情况下，可以基于不包含轨迹点的第二参考图像与和它相邻或者距离最近的调整后的包含轨迹点的第二参考图像之间的偏离量和重叠量对齐该不包含轨迹点的第二参考图像。需要说明的是，该最近距离的度量为曼哈顿距离，而非欧式距离。

所述对第二参考图像进行坐标调整还包括对相邻的第二参考图像之间进行接缝

融合处理，例如，可以采用线性加权融合方法等。

在一些实施例中，如图 11 所示，图 11 为本申请实施例提供的一种将染色图像与基因表达可视化图像进行配准的方法流程示意图，包括：

5 步骤 1021，获取所述染色图像中的轨迹线，基于所述轨迹线与水平方向的夹角旋转所述染色图像；其中，所述染色图像和所述基因表达可视化图像包括多条轨迹线和/或多个轨迹点，所述轨迹点是由两条所述轨迹线相交形成的，所述多条轨迹线中的每条轨迹线具有对应的索引号；

10 由于染色图像是由显微镜图像拍摄得到的，所以显微镜图像中的轨迹线并不一定处于水平方向和垂直方向，所以可以选取显微镜图像中的第一轨迹线作为基准，计算第一轨迹线与水平方向的夹角，将显微镜图像旋转上述夹角后位于水平方向。参见图 12，展示了一种染色图像的示意图。由图 12 可知，染色图像没有位于水平方向，第一轨迹线与水平方向存在一个夹角。所以可以计算出所述夹角后将染色图像旋转至水平方向。

15 由于基因表达可视化图像是由基因表达矩阵生成的，表现为许多离散的点，对表达聚集特征或生物样本的边缘特征的能力较弱。在一种优选的实现方式中，可以对基因表达可视化图像进行组织分割以获取基因表达展现出的生物样本轮廓，对样本轮廓所在的大致区域进行卷积增强处理，增强基因表达可视化图像的边缘特征，方便后续将显微镜图像与基因表达可视化图像进行配准。

20 步骤 1022，将所述染色图像进行放大或缩小得到第三参考图像，所述第三参考图像与所述基因表达可视化图像具有相同的比例；

当染色图像是由显微镜拍摄的图像时，由于基因表达可视化图像可以作为基准图像，而显微镜拍摄的染色图像与基因表达可视化图像的缩放比例可能存在差异。为了实现将染色图像与基因表达可视化图像的配准，所以需要将染色图像进行放大或缩小后得到与基因表达可视化图像大小相同的第三参考图像。

25 在一种可能的实现方式中，可以根据染色图像与基因表达可视化图像中对应索引号的轨迹线之间的间距确定染色图像的缩放比例。具体地，获取染色图像中的第二轨迹线，基于染色图像的第一轨迹线和第二轨迹线确定第一间距；获取基因表达可视化图像中的第三轨迹线和第四轨迹线，其中，第三轨迹线的索引号与第一轨迹线的索引号相同，第四轨迹线的索引号与第二轨迹线的索引号相同，基于第三轨迹线和第四轨迹线确定第二间距。然后基于第一间距和第二间距确定染色图像和基因表达可视化图像之间的缩放比例值，将染色图像按照上述比例值进行放大或缩小，得到与基因表达可视化图像相同比例的第三参考图像。其中，第一轨迹线和第二轨迹线平行；第三轨迹线和第四轨迹线平行。

30 当基于第一轨迹线和第二轨迹线确定第一间距时，可以利用第一轨迹线上的轨迹点与第二轨迹线上的轨迹点计算，当两个轨迹点具有相同的横坐标或者纵坐标时，对应的纵坐标的差值或者横坐标的差值即可作为第一间距。同理，可以基于相同的原理确定第三轨迹线和第四轨迹线之间的第二间距。需要说明的是，上述实施例所提供的计算染色图像和基因表达可视化图像的缩放比例的方法仅为示例性的说明，并非仅限于上述实现方式。

40 步骤 1023，基于所述基因表达可视化图像和所述第三参考图像确定目标偏移量；

当染色图像缩放至第三参考图像，具有与基因表达可视化图像相同的比例后，后续即可以根据基因表达可视化图像和第三参考图像确定目标偏移量。通过上述实施例可知，将原始染色图像旋转至水平方向后，染色图像与基因表达可视化图像之间可能存在 90 度倍角的差异，为了进一步提高图像处理的准确性，在将染色图像缩放至与基因表达可视化图像相同比例的第三参考图像后，可以将第三参考图像按照预设旋转角度进行旋转，在每个旋转角度下确定对应的目标偏移量。下面将结合一种具体应用场景进行说明。

50 如图 13 所示，S 表示第三参考图像，V 表示基因表达可视化图像，黑色点表示第三参考图像的重心，灰色点表示基因表达可视化图像的重心，图 13 中四个区域的图分别表示了原始第三参考图像以及将第三参考图像顺时针旋转 90 度、180 度、270 度后的图像。在将第三参考图像进行旋转后，可以根据第三参考图像的原点坐标和基因表

达可视化图像的原点坐标进行对齐，然后根据原点对齐后的第三参考图像和基因表达可视化图像确定目标偏移量。

针对任一预设旋转角度，为了方便、更准确地处理图像，本实施例提供一种优选的实现方式，即首先可以根据第三参考图像与基因表达可视化图像的重心坐标，将第三参考图像进行初步定位到与基因表达可视化图像大致相同的位置，然后再根据移动后的第三参考图像计算目标偏移量，进行精准定位。具体实现时，将染色图像进行放大或缩小得到第五参考图像，其中第五参考图像与基因表达可视化图像具有相同的比例。然后分别计算第五参考图像的第一重心坐标和基因表达可视化图像的第二重心坐标，基于第一重心坐标和第二重心坐标确定重心偏移量，并基于该重心偏移量移动第五参考图像，得到第四参考图像，即得到初步定位的第四参考图像。其中，在计算第五参考图像和基因表达可视化图像的重心坐标时，可以利用基于像素灰度值的密度质心算法实现，本实施例对此并不做限定。

需要说明的是，关于中间过程中的图像移动，可以在复制染色图像的临时图像上实现的处理过程，即并非将染色图像进行移动配准，当计算完成得到最终的目标偏移量之后，再一次性移动染色图像，与基因表达可视化图像进行配准。

在得到初步定为的第四参考图像后，可以根据第四参考图像和基因表达可视化图像确定目标偏移量。在实际应用中，时空芯片上自带的轨迹线可能具有多个周期，因此基于时空芯片得到的显微镜图像和基因表达可视化图像也可能具有多个周期，即第四参考图像和基因表达可视化图像具有多个周期。每个周期内具有一个对应的子图像，每个子图像中的轨迹线和轨迹点数目相同，并且子图像之间轨迹线的索引号一一对应。如图 14a 和图 14b 所示，图 14a 展示了具有多个周期的第三参考图像，图 14b 展示了多个周期的基因表达可视化图像。在图 14a 中，仅表示了第一周期和第二周期的子图像，可以得知每个周期的子图像中，具有相同数目的索引线和索引点，每条索引线均对应一个索引号。图 14b 中也表示了第一周期和第二周期对应的子图像。上述实施例所提供的第三参考图像和基因表达可视化图像的示意图仅为一种可能的实现方式，并非对图像的具体形式做任何限定。所以基于第四参考图像和基因表达可视化图像确定目标偏移量时，需要考虑周期的影响。

具体实现时，首先可以根据第四参考图像和基因表达可视化图像确定第一偏移量，该第一偏移量为第四参考图像与基因表达可视化图像之间的待定偏移量，然后以第一偏移量为中心，确定第一偏移量在预设周期内所对应的多个第二偏移量，即在预设周期内的每个周期对应一个第二偏移量。针对任意一个第二偏移量，基于该第二偏移量移动第四参考图像，得到第六参考图像，然后计算第六参考图像与基因表达可视化图像之间的第一相似度，从而得到在预设周期内的每个周期所对应的第一相似度。由于上述确定相似度是在任意一个旋转角度下所得到的，所以需要在其他预设旋转角度下，重新执行上述步骤，得到每个旋转角度下所对应的多个第一相似度。基于所有旋转角度和所有周期所确定的多个第一相似度，确定第一最大相似度，将该第一最大相似度所对应的第六参考图像确定为第三参考图像。然后可以确定第三参考图像所对应的第二偏移量为目标偏移量。

其中，根据第四参考图像和基因表达可视化图像确定第一偏移量时，一种可能的实现方式为，首先获取多组轨迹点集合，其中每组轨迹点集合包括第一轨迹点和第二轨迹点，第一轨迹点位于第四参考图像，第二轨迹点位于基因表达可视化图像，第一轨迹点与第二轨迹点具有相同的索引标识。由于轨迹点是由两条相交的轨迹线形成的，每条轨迹线都具有索引号，所以轨迹点的索引标识可以由该轨迹点的两条相交轨迹线的索引号组成。在本实施例中，由于在第四参考图像和基因表达可视化图像中均具有多个轨迹点，所以需要获取具有对应关系的轨迹点才可以作为一组轨迹点集合，即相同索引号的轨迹点。针对多组轨迹点集合中的任一组轨迹点集合，可以根据第一轨迹点的坐标和第二轨迹点的坐标确定第四偏移量，然后基于多个第四偏移量确定第一偏移量。

为便于更清楚地理解本方案，在介绍第三偏移量的实现方式之前，首先对第四偏移量的实现方式进行说明。

具体实现时，由于基因表达可视化图像存在多个周期，即具有多个子图像，由此

所获取的具有相同索引标识的第一轨迹点和第二轨迹点可能不属于对应的周期，所以所确定的第二轨迹点可能有多个。在基于第一轨迹点的坐标和第二轨迹点的坐标确定第四偏移量时，可以首先选取第四参考图像中的任一第一轨迹点，根据第一轨迹点的索引标识确定具有相同标识的多个第二轨迹点。为了获取与第一轨迹点对应的第二轨迹点，可以分别计算第一轨迹点与每一个第二轨迹点的欧式距离，在计算得到多个欧式距离后，获取欧式距离最小时所对应的第二轨迹点，即为与第一轨迹点对应的第二轨迹点。然后根据最小欧式距离所对应的第二轨迹点以及第一轨迹点的坐标差值确定为第四偏移量。由于在第四参考图像中具有多个轨迹点，即对应多组轨迹点集合，所以需要遍历所有的轨迹点集合，按照上述确定基本偏移量的方法，确定与每一组轨迹点集合相对应的第四偏移量。在一种可能的实现方式中，在确定的多个第四偏移量中，可以首先排除超出预设范围的第四偏移量，然后根据剩余的第四偏移量确定目标偏移量。例如，可以选择多个第四偏移量中的中位数作为目标偏移量，或者计算所有第四偏移量的平均值作为目标偏移量，以尽可能地提高确定偏移量的准确性。需要说明的是，本实施例对根据第四偏移量确定目标偏移量的具体方式不做限定。

下面将结合一种具体应用场景对预设周期内的图像移动进行说明。

参见图 15，图 15 为以第一偏移量为中心所确定的预设周期的示意图。

在一种可能的实现方式，以基因表达可视化图像作为基准，由图 15 可知，中间灰色区域表示第二轨迹点所在的周期，根据第一轨迹点和第二轨迹点确定第一偏移量后，以该第一偏移量（对应的第二轨迹点所属的周期）为中心选择附近 5×5 个周期，确定每个周期所对应的第二偏移量，相邻周期所对应的第二偏移量的差值为子图像的宽度，以此类推，可以确定每个周期所对应的第二偏移量。

然后可以针对任一周期所对应的第二偏移量移动第四参考图像得到第六参考图像，并计算第六参考图像与基因表达可视化图像之间的相似度。在计算相似度时，本实施例提供一种可能的实现方式，即对第六参考图像进行频域变换得到第一频域图像，对基因表达可视化图像进行频域变换得到第二频域图像，计算第一频域图像和第二频域图像之间的汉明距离作为相似度。当汉明距离越小时，表明第一频域图像与第二频域图像的相似度越大。在确定所有的预设旋转角度和所有预设周期所对应的汉明距离之后，将最小汉明距离作为第一最大相似度。

其中，可以利用离散余弦变换获取频域图像，即对移动后的第三参考图像进行离散余弦变换得到第一频域图像，对基因表达可视化图像进行离散余弦变换得到第二频域图像，然后从第一频域图像中截取低频频域图像，以低频频域图像为基准，计算低频频域图像的像素均值，在低频频域图像中将像素值高于像素均值的像素点的像素值置为 1，将像素值低于像素均值的像素点的像素值置为 0，将像素值等于像素均值的像素点的像素值置为 0 或 1，然后根据重新置值后的低频频域图像获取哈希指纹。按照同样的执行步骤，获取第二频域图像的哈希指纹。然后基于第一频域图像的哈希指纹和第二频域图像的哈希指纹确定两个哈希指纹对比有多少位不同的值，即作为第一频域图像和第二频域图像的汉明距离。需要说明的是，上述实施例中计算相似度的方法仅为示例性的实现方式，并非仅限于上述实现方式，其他可实现的计算图像相似度的方法也在本申请的保护范围内。

在本申请的上述实施例中，可以首先将第四参考图像按照与基因表达可视化图像的重心偏移量进行初步移动定位，在另一种可能的实现方式中，在确定第三参考图像时，并不对第四参考图像进行初定位，针对任一预设旋转角度，将第四参考图像分别按照预设周期内的每个周期进行移动，得到每个周期所对应的第七参考图像。例如，可以先计算第四参考图像的第三重心坐标，以第三重心坐标所在的周期为中心，分别按照预设周期内的每个周期移动第四参考图像，得到第七参考图像。然后计算第七参考图像与基因表达可视化图像之间的第二相似度，从而得到预设周期内每个周期所对应的第二相似度。然后根据其他的预设旋转角度，重复执行上述步骤，获取在每个预设旋转角度以及每个周期内所对应的第二相似度，并基于上述多个第二相似度确定第二最大相似度。将第二最大相似度所对应的第七参考图像作为第三参考图像，然后基于该第三参考图像和基因表达可视化图像确定目标偏移量。

具体地，首先可以根据基因表达可视化图像和第三参考图像确定多个第三偏移

量，基于该多个第三偏移量确定目标偏移量。其中，该第三偏移量的计算原理与上述实施例中计算第四偏移量的原理相同。即，获取多组轨迹点集合，其中，每组轨迹点集合包括第三轨迹点和第四轨迹点，第三轨迹点位于第三参考图像，第四轨迹点位于基因表达可视化图像，并且第三轨迹点与第四轨迹点具有相同的索引标识。针对任一

5 组轨迹点集合，基于第三轨迹点的坐标和第四轨迹点的坐标确定第三偏移量。其中，基于第三轨迹点的坐标和第四轨迹点的坐标确定第三偏移量的具体实现方式可以参见基于第一轨迹点的坐标和第二轨迹点的坐标确定第四偏移量的具体实现方式，在此不再赘述。

同样地，在确定多组轨迹点集合所对应的第三偏移量之后，可以首先排除超出预设范围的第三偏移量，然后在满足预设范围的多个第三偏移量中选择中位数作为目标偏移量，或者计算多个第三偏移量的平均值作为目标偏移量，以尽可能地提高确定偏移量的准确性。

步骤 1024，基于所述目标偏移量移动所述第三参考图像，得到所述配准后的图像。

基于上述实施例提供的方法确定目标偏移量之后，确定该目标偏移量所对应的预设旋转角度，基于该预设旋转角度和目标偏移量移动第三参考图像。

需要说明的是，关于中间过程中的图像移动，均为在复制染色图像的临时图像上实现的处理过程，即并非将染色图像进行移动配准，当计算完成得到最终的目标偏移量之后，再一次性移动染色图像，与基因表达可视化图像进行配准。

本申请实施例所提供的图像处理方法中，可以基于像素级精度的图像坐标，将染色图像与基因表达可视化图像进行配准，提高图像处理的准确性，便于后续利用配准的图像进行研发和分析。

当第三参考图像和基因表达可视化图像具有多个周期时，由于预设周期对应的偏移量是由计算得到的偏移量利用理论周期宽度推理得到的，在实际图像中可能存在误差。为了修正理论推理的误差，进一步提高图像处理的准确性，在一种优选的实现方式中，当基于目标偏移量移动第三参考图像之后，基于移动后的第三参考图像以及基因表达可视化图像，可以重新利用该第三参考图像和基因表达可视化图像确定修正偏移量作为最终偏移量，并基于所确定的最终偏移量继续移动该第三参考图像，将该第三参考图像与基因表达可视化图像进行配准，提高配准的准确性。例如，基于移动后的第三参考图像以及基因表达可视化图像，获取多组轨迹点集合，其中，每组轨迹点集合包括第五轨迹点和第六轨迹点，第五轨迹点位于该第三参考图像，第六轨迹点位于基因表达可视化图像，并且第五轨迹点与第六轨迹点具有相同的索引标识。针对任一

25 组轨迹点集合，基于第五轨迹点的坐标和第六轨迹点的坐标确定第五偏移量。在确定多组轨迹点集合中每组轨迹点集合所对应的第五偏移量之后，基于多个第五偏移量确定最终偏移量。其中，确定第五偏移量以及基于多个第五偏移量确定最终偏移量的实现方式可以参见上述实施例，在此不再赘述。

作为本申请实施例的一种可行方式，在步骤将所述染色图像进行放大或缩小得到第三参考图像之后，还包括：确定所述第三参考图像与基因表达可视化图像之间的相似度；将相似度最大值对应的位置作为粗配准结果。该相似度作为粗配准，图 11 所示的方法作为精细配准。

在一些实施例中，如图 16 所示，图 16 为本申请实施例提供的一种将染色图像与基因表达可视化图像进行配准的方法流程示意图，包括：

步骤 1025，对所述染色图像进行预处理，获取候选图像集合；

在将染色图像与基因表达可视化图像进行配准之前，由于染色图像为显微镜实际拍摄的图像，受外界等因素影响，与基因表达可视化图像存在一定的差异，例如，染色图像未处于水平位置，所以为了方便后续的图像配准，可以首先对染色图像进行预处理。

具体地，获取染色图像的第一轨迹线和第二轨迹线，该第一轨迹线和第二轨迹线可以为随机确定的轨迹线。然后基于第一轨迹线与水平方向的夹角旋转染色图像，也就是，将染色图像旋转至水平位置。并基于第一轨迹线和第二轨迹线确定两条轨迹线之间的第一间距。获取基因表达可视化图像中的第三轨迹线和第四轨迹线，其中，第三轨迹线的索引号与第一轨迹线的索引号相同，第四轨迹线的索引号与第二轨迹线的

索引号相同。并基于第三轨迹线和第四轨迹线两条轨迹线之间的确定第二间距，基于第一间距和第二间距的比值调整旋转后的染色图像，获取候选图像。也就是，染色图像和基因表达可视化图像的轨迹线间距可能存在缩放比例差异，为了实现准确地图像配准，可以将染色图像缩放至与基因表达可视化图像相同的比例。其中，当基于第一

5 轨迹线和第二轨迹线确定两条轨迹线之间的第一间距时，可以获取第一轨迹线和第二轨迹线中具有相同横坐标的两个轨迹点，则两个轨迹线的纵坐标的差值即为第一间距。或者，获取第一轨迹线和第二轨迹线中具有相同纵坐标的两个轨迹点，则两个轨迹线的横坐标的差值即为第一间距。

在获取候选图像后，由于此时候选图像为处于水平位置的图像，但是可能与基因

10 表达可视化图像存在 90° 倍角的差异，所以可以确定多个旋转角度，基于每个旋转角度旋转候选图像，从而将每个旋转角度所对应的旋转后的图像组成候选图像集合。例如，该多个旋转角度可以为 90° 、 180° 以及 270° 。经过旋转后，该候选图像集合中的任一图像与染色图像仍具有相同的索引号。

步骤 1026，确定所述候选图像集合对应的二值图集合；

15 二值图即为图像二值化，就是将图像上每个像素点的灰度值设置为 0 或 255，使整个图像呈现出明显的黑白效果，方便后续的图像配准。可选地，可以通过以下方式确定二值图集合，针对候选图像集合中的任一第四图像，基于第四图像的组织轮廓进行分割，获取组织分割图像。其中，进行组织分割时，可以基于组织分割算法实现，也可以使用组织分割工具手动圈出组织轮廓获得。然后对组织分割图像进行二值化处

20 理，获取初始二值图。针对初始二值图的多个边界中的任一边界，确定组织轮廓中距离边界最近的边界点，并在与边界平行的轨迹线中，确定距离边界点最近的边界轨迹线，并且保证边界轨迹线不经过组织轮廓，从而基于每个边界对应的边界轨迹线确定目标二值图。候选图像集合中的多个第四图像所对应的目标二值图组成二值图集合。也就是，目标二值图为包括整个生物组织的最小二值图，并且目标二值图的边界由轨

25 迹线组成。参见图 17a 所示，为本申请实施例提供的一种第四图像的示意图，对第四图像的组织轮廓进行分割处理后，得到组织分割图像，然后对组织分割图像进行二值化处理，得到的初始二值图如图 17b 所示。然后基于初始二值图确定包括所有组织的最小二值图，即目标二值图如图 17c 所示。

步骤 1027，基于所述二值图集合与所述基因表达可视化图像之间的相似度，确定

30 所述二值图集合与所述基因表达可视化图像中具有对应关系的轨迹点对；

在实际应用中，基因表达可视化图像可能具有多个周期，每个周期的基因表达可

35 视化图像为相同的，也就是，基因表达可视化图像的索引号可能具有多个周期。基于此，在确定轨迹点对时，针对二值图集合中的任一二值图，基于该二值图的顶点所对应的坐标和基因表达可视化图像的轨迹点所对应的坐标，在基因表达可视化图像中确定与二值图所对应的多个区域。其中，所确定的多个区域与多个周期一一对应。二值图的顶点即为二值图中的轨迹点，顶点所对应的坐标可以由相交形成顶点的两条轨迹线对应的索引号确定。针对上述多个区域中的任一区域，确定二值图与该区域之间的相似度指标，从而可以确定多个二值图与多个区域两两之间的多个相似度指标。基于多个相似度指标，确定二值图集合与基因表达可视化图像中具有对应关系的轨迹点对。

40 其中，该相似度指标可以表示二值图所对应的图像与区域对应的图像之间的相似程度，相似度指标越大，表明相似程度越高。

其中，在确定二值图与区域之间的相似度指标时，可以通过以下方式实现，确定二值图对应的二值图矩阵，由于二值图的像素点的灰度值只能为 0 或 255，可选地，可以将灰度值为 0 的像素点在二值图矩阵中的元素设置为 0，将灰度值为 255 的像素点在

45 二值图矩阵中的元素设置为 1，也就是，二值图为由 0 和 1 组成的矩阵。当基因表达可视化图像为基因表达可视化图像时，可以确定该基因表达可视化图像对应的基因表达矩阵，基因表达可视化图像的每个像素点在基因表达矩阵中对应一个基因表达量。然后利用该二值图矩阵和该区域对应的基因表达矩阵确定相似度指标。例如，可以将二值图矩阵与基因表达矩阵进行点乘，即二值图矩阵与基因表达矩阵中对应位置的元素

50 相乘，然后对点乘得到的矩阵中的所有元素进行求和，得到相似度指标。从而可以根据二值图集合与基因表达可视化图像的多个区域之间的相似度指标，确定具有对应关

系的轨迹点对。

根据上述实施例可知，染色图像可以为利用轨迹线定位算法对第一原始图像进行处理得到的，也可以为将第一原始图像按照预设比例进行缩放后，再利用轨迹线定位算法进行处理得到的。下面将针对上述两种获取染色图像和基因表达可视化图像的方式，分别介绍基于多个相似度指标确定轨迹点对的方式。

当利用轨迹线定位算法处理第一原始图像和第二原始图像，获取染色图像和基因表达可视化图像时，在获取二值图集合中的多个二值图与多个区域之间的多个相似度指标之后，可以获取多个相似度指标中的最大值，确定该最大值所对应的二值图为第一目标图像，确定该最大值所对应的区域为第一目标区域。然后可以获取第一目标图像中的第一顶点以及第一目标区域中与第一顶点的坐标相同的第一索引点，并确定第一顶点与第一索引点为具有对应关系的轨迹点对。其中，该第一顶点可以为第一目标图像中的任一顶点。

当将第一原始图像和第二原始图像按照预设比例进行缩放后，利用轨迹线定位算法处理缩放后的第一原始图像和缩放后的第二原始图像，获取染色图像和基因表达可视化图像时，在获取二值图集合中的多个二值图与多个区域之间的多个相似度指标之后，可以确定多个相似度指标的排序结果，并基于该排序结果获取预设数量的相似度指标，并确定预设数量的相似度指标所对应的多个二值图，以及所对应的多个区域。例如，可以选择数值最大的前预设数量的相似度指标。其中，本申请实施例并不限定预设数量的具体数值，可以结合实际应用场景进行确定。由于利用缩放后的图像进行配准时可能存在误差，为了进一步提高图像处理的准确性，可以获取预设数量的相似度指标对应的二值图缩放前的图像。也就是，确定该预设数量的相似度指标所对应的多个二值图之后，将多个二值图按照缩放预设比例的倒数进行放大，确定多个第二目标图像，即缩放前的图像。然后基于多个第二目标图像对应的二值图矩阵和多个区域对应的基因表达矩阵，确定多个候选相似度指标。获取多个候选相似度指标中的最大值，并确定该最大值所对应的二值图为第三目标图像，以及该最大值所对应的区域为第三目标区域。获取第三目标图像中的第二顶点以及第三目标区域中与第二顶点的坐标相同的第二索引点，确定第二顶点与第二索引点为具有对应关系的轨迹点对。其中，该第二顶点可以为第三目标图像中的任一顶点。

步骤 1028，基于所述轨迹点对确定偏移信息；

可选地，可以获取规矩点对的两个轨迹点对应的坐标，确定横坐标的差值和纵坐标的差值为偏移信息，即偏移量。

步骤 1029，基于所述偏移信息处理目标图像，以得到配准后的图像，所述目标图像为所述轨迹点对所对应的候选图像集合中的图像。

在确定偏移信息之后，可以基于偏移信息移动目标图像，也就是，轨迹点对中属于候选图像集合中的轨迹点对应的图像，即可实现图像的配准。

需要说明的是，关于中间过程中的图像处理以及移动，均为在复制染色图像的临时图像上实现的处理过程，即并非将染色图像进行移动配准，当计算完成得到最终的偏移信息之后，可以根据目标图像的旋转方向以及偏移信息，再移动染色图像，与基因表达可视化图像进行配准。

通过本申请所提供的图像处理方法，无需通过人工配准图像，可以实现在像素级精度的图像处理，从而提高图像处理的准确性。

在对所述配准后的图像进行细胞分割，得到细胞掩模图之前，所述方法还包括：步骤 105，对所述配准后的图像进行组织分割，得到组织掩模图。

在一些实施例中，如图 18 所示，图 18 为本申请实施例提供的一种对配准后的图像进行组织分割的方法流程示意图，包括：

步骤 1051，获取所述配准后的图像中待分割的目标组织的基因表达量并确定所述配准后的图像的基因表达矩阵，所述基因表达矩阵按照像素位置对所述基因表达量进行排列；

在一个实施例中，目标组织可以包括但不限于样本组织、生物组织、肌肉组织、骨骼组织、细胞组织中任一项或多项。基因表达矩阵表征基因在不同位置（对应基因表达矩阵的行列位置）的表达。每一条基因表达量数据包括基因标识符、基因坐标和

基因表达量，表征基因在一个位置的表达量。以细胞组织为例，可以通过对细胞组织的 DNA 进行转录处理以确定细胞组织的基因表达量。此外，基因表达量通常为 utf-8 格式。

本实施中，通过获取样本图像中每一个像素位置的基因表达量，并按照像素位置的行坐标、列坐标对基因表达量进行排列以得到基因表达矩阵，基因表达矩阵与样本图像的大小一致，当每个像素位置的基因表达量为 0 时表征在该像素位置不存在目标组织的基因。基因表达矩阵中每个像素位置的基因表达量可以表征一个或多个基因的基因表达量。若基因表达矩阵与样本图像的大小不一致，可以对两个图像或其中的一个图像进行缩放。

步骤 1052，通过将所述基因表达矩阵中每一像素位置的基因表达量赋值于配准后的图像中对应的像素位置确定图像数据；

确定与所述基因表达矩阵相同大小的图像矩阵；

将所述基因表达矩阵中每一像素位置的基因表达量赋值于图像矩阵中对应的像素位置并对所有像素位置进行归一化处理得到样本图像的图像数据。

参见图 19，将基因表达矩阵中每一像素位置的基因表达量以每个像素位置的行坐标与列坐标为键、基因表达量为值存储至字典中。创建一个与基因表达矩阵相同大小的全 0 图像矩阵，并按照字典中存储的基因表达量的行坐标与列坐标的位置赋值于图像矩阵中的相同位置，即图像矩阵中相同像素位置的基因表达量与基因表达矩阵中相同像素位置的基因表达量相同。对赋值后的图像矩阵进行归一化，使每一像素位置的像素值在 [0, 255] 区间内，得到相对样本图像更为准确的图像数据，并减少对于后续组织分割图像的影响。图像数据参见图 20 所示，图像数据的格式优选为 TIFF 格式，也可以为 JPG 格式等，并不局限于某一具体表现形式，可根据实际情况进行选择。

在一个实施例中，通过获取确定每一像素位置的基因表达量生成基因表达矩阵，并将基因表达矩阵转换为图像数据，而减少通过原始显微镜对组织图像进行反复拍摄和校正的过程，提高分割效率，并使得图像数据相较于初始的样本图像对于目标组织的展示更为准确，且对于目标组织的分割效果更好。

步骤 1053，通过对所述图像数据进行图像处理以确定所述图像数据的前景像素点；

在一个实施例中，步骤 S1053 之前，组织分割方法还包括：参见图 21，对图像数据进行卷积，以提取图像数据的低频特征。其中，卷积方法优选采用 2d 卷积，也可以采用空洞卷积、深度可分离卷积等卷积方式来提升计算虚度和扩大感受视野，卷积后的图像数据参见图 22 所示。

通过对进行卷积后的图像数据图像处理以确定图像数据的前景像素点。

步骤 1054，根据所述前景像素点对所述图像数据的连通区域进行标记得到所述组织掩模图。

在一个实施例中，连通区域指图像数据中具有相同像素值且位置相邻的前景像素点组成的图像区域，即目标组织的分割区域。

在一些实施例中，在组织分割得到组织掩模图后，如图 23 所示，图 23 为本申请实施例提供的一种计算描述组织掩模图对应的灰度图像质量的方法流程示意图，包括：

步骤 1061，分别计算得到组织掩模图对应的灰度图像的外轮廓图像与内轮廓图像；

可基于矩阵运算分别得到所述目标区域组织灰度图像的外轮廓图像与内轮廓图像，在本实施例中所述外部轮廓图像与所述内轮廓图像为带状环形轮廓，所述外轮廓图像可包围住所述内轮廓图像。

步骤 1062，将所述外轮廓图像与所述内轮廓图像分别分为 N 段，并将每段所述外轮廓图像与每段所述内轮廓图像两两配对，得到 N 个图像组合，每个所述图像组合包括一段外轮廓图像与一段内轮廓图像；

为了更直观的展示所述外轮廓图像与所述内轮廓图像的关系，图 24 为本公开实施例提供的一种外部轮廓图像与内部轮廓图像的关系结构示意图，图 25 为本申请实施例提供的另一种外部轮廓图像与内部轮廓图像的关系结构示意图，如图 24 所示，所述将所述外轮廓图像与所述内轮廓图像分别分为 N 段即将所述外轮廓图像均匀分割成 N

段，将所述内轮廓图像均匀分割成 N 段，在所述 N 段外轮廓图像中取一段与所述 N 段内轮廓图像中的一段进行配对，形成图像组合，配对完成后便可以得到 N 个图像组合。

步骤 1063，分别计算每个所述图像组合中一段外轮廓图像与一段内轮廓图像的像素密度比值，得到 N 个所述像素密度比值，所述像素密度比值为每个所述图像组合中一段外轮廓图像与一段内轮廓图像的像素占比的比值，所述像素占比为像素点与总像素点的比值；

在计算所述像素密度比值之前，需要先计算每段所述外轮廓图像与每段所述内轮廓图像的像素占比，计算一段所述外轮廓图像的像素占比包括使用该段外轮廓图像内的像素点的数量除以该段外轮廓图像内的像素点的总数量，在本实施例中所述像素点为灰度值大于 127 的像素点，计算一段内轮廓图像的像素占比与计算一段所述外轮廓图像的像素占比的方法一致，即使用该段内轮廓图像内的像素点的数量除以该段内轮廓图像内的像素点的总数量，以此类推最终计算完成所有段的外轮廓图像的像素占比和所有段的内轮廓图像的像素占比。

在完成上述像素占比的计算之后，还需计算一段外轮廓图像的像素占比与一段内轮廓图像的像素占比的比值，所述比值即为像素密度比值，所述一段外轮廓图像与所述一段内轮廓图像存在于一个图像组合，即二者之间相互配对，完成所有图像组合中的一段外轮廓图像的像素占比与一段内轮廓图像的像素占比的比值的计算，最终得到 N 个像素密度比值。

步骤 1064，根据所述像素密度比值计算描述所述组织掩模图对应的灰度图像质量。

所述根据所述像素密度比值计算描述所述目标区域组织灰度图像质量包括但不限于以下实现方式，例如：将所述像素密度比值的平均值代入预设函数计算出描述所述目标区域组织灰度图像质量的数值，所述预设函数可以为线性函数也可为非线性函数。

为了便于理解上述细化的内容，本实施例提供了一种结合公式的示例性说明，例如：所述像素密度比值的平均值为 x，当所述 x 大于 0 小于 200 时，可代入公式 (1) 得到描述所述目标区域组织灰度图像质量的数值，公式 (1) 如下所示：

$$F(x)=100 * x^{0.7} \quad (0 < x < 200) \quad \text{公式 (1)}$$

$$F(x)=0 \quad (x \leq 0) \quad \text{公式 (2)}$$

$$F(x)=100 \quad (x \geq 200) \quad \text{公式 (3)}$$

上述公式仅为示例性说明，其可以被任意函数替换，本申请实施例对公式不进行限定。

在一些实施例中，如图 26 所示，图 26 为本申请实施例提供的一种对配准后的图像进行细胞分割的方法流程示意图，包括：

步骤 1031，将所述配准后的图像输入所述图像分割模型，获得所述配准后的图像中各个像素点的分割结果，所述分割结果包括属于细胞的像素点、属于细胞轮廓的像素点或者图像背景的像素点。

将配准后的图像输入上述图像分割模型，可以得到配准后的图像中各个像素点属于细胞、细胞轮廓或者图像背景的概率值，根据概率值是否大于阈值确定配准后的图像中各个像素点的初始分割结果，即确定配准后的图像中各个像素点属于细胞的像素点、细胞轮廓的像素点或者图像背景的像素点。在实际应用中，可以将配准后的图像扩展为三通道（如 RGB 三通道）图像，输入图像分割模型，得到每一通道图像中各个像素点的初始分割结果，即每一通道图像中各个像素点属于目标对象、目标对象轮廓或者图像背景，再综合各通道图像中各个像素点的初始分割结果，确定配准后的图像中各个像素点的初始分割结果。

步骤 1032，将所述分割结果属于细胞的像素点或细胞轮廓的像素点确定属于细胞的像素点，输出所述配准后的图像的分割结果，得到所述细胞掩模图。

将属于细胞或者细胞轮廓的像素点进行融合，实现细胞的分割，即将细胞对象或者细胞轮廓的像素点确定为属于细胞的像素点，进行图像分割，输出配准后的图像的分割结果。

在本申请实施例中，还可以利用分水岭算法过滤配准后的图像的分割结果中面积小于阈值的目标对象，和/或，对不完整的目标对象进行修复。

通过过滤面积小于阈值的目标对象和/或修复不完整的目标对象，实现对目标对象的修正，完成配准后的图像的分割结果的后处理，提高图像分割的准确性。例如，
5 目标对象为细胞时，通过细胞过滤和修正等算法，过滤非常小的细胞，以及修正不完整的细胞，使细胞分割更为准确。

本申请实施例使用的图像分割模型，在编码器模块中采用了金字塔挤压注意力模块，提高了编码器模块的特征提取能力，使图像分割模型更加关注具有显著性特征的目标对象，从而提高了图像分割模型的分割准确性。

10 将本申请实施例应用于时空组学中的细胞分割，参见图 27 所示，展示了本申请实施例的图像分割模型在鼠脑图像上的分割效果。截取了上图鼠脑图像的一部分作为展示，从左下图中可以看出，鼠脑图像中存细胞黏连情况，部分细胞边界严重模糊，右下图利用本申请实施例提出的图像分割模型可以从黏连细胞中分割出独立细胞，并提供比较可靠的边界。

15 在一些实施例中，如图 28 所示，图 28 为本申请实施例提供的一种对配准后的图像进行细胞分割的方法流程示意图，包括：

步骤 1033，对所述基因表达可视化图像进行预处理，得到预处理图；

由于基因表达可视化图像为散点图，不便于进行细胞分割，因此需要进行预处理，将细胞的基因表达可视化图像处理得到边界效果增强的预处理图。

20 步骤 1034，将所述预处理图进行二值化处理，得到初始掩模图；

本实施例可使用大津法对预处理图进行二值化处理，得到初始掩模图。

步骤 1035，根据所述初始掩模图，利用基于距离变换的分水岭算法对存在细胞粘连的连通域进行分割，得到分割后的所述细胞掩模图。

25 分水岭算法(Watershed Algorithm)，是根据分水岭的构成来考虑图像的分割。分水岭算法(Watershed Algorithm)，是根据分水岭的构成来考虑图像的分割。

进一步的，作为上述实施例的细化和扩展，为了完整说明本实施例方法的具体实现过程，本实施例提供了如图 29 所示的具体方法，该方法包括：

步骤 201、获取包含空间位置的基因表达矩阵。

30 从细胞的基因表达数据中获取包含空间位置的基因表达矩阵。基因表达数据可以包括多个基因的基因标识符、坐标位置及相应坐标位置的总基因表达量。

步骤 202、基于基因表达矩阵，生成细胞的基因表达可视化图像。

作为一种可选方式，步骤 202 具体可包括：首先获取基因表达矩阵中的表达基因的坐标位置及相应坐标位置的总基因表达量；再根据表达基因的坐标位置及相应坐标位置的总基因表达量，生成基因表达可视化图像，其中，基因表达可视化图像为灰度图，基因表达
35 可视化图像中像素点的灰度值为与像素点对应坐标位置的总基因表达量。通过这种可选方式，可准确生成得到细胞的基因表达可视化图像。

例如，在具体的细胞分割使用中，输入含空间位置的基因表达矩阵，根据表达基因的坐标位置及相应位置的总基因表达量生成表达图像，该图像具体形式为灰度图，坐标点的灰度值为该坐标上基因表达的总量。

40 作为另一种可选方式，步骤 202 具体还可包括：根据基因表达矩阵和分割的掩模图，绘制细胞的基因表达可视化图像，其中，基因表达矩阵中的空间位置与分割的掩模图中的空间位置相对应。例如，根据基因表达矩阵和分割的掩模图绘制基于细胞的表达量图，该基因表达矩阵中的空间位置与分割的掩模图中的空间位置可以相对应。通过这种可选方式，可准确生成得到细胞的基因表达可视化图像。

45 由于生成的基因表达可视化图像为散点图，不便于分割，需要进行处理，具体可执行步骤 203 所示的过程。

步骤 203、对基因表达可视化图像进行预处理，得到中值图，并将中值图进行锐化处

理, 得到锐化图。

由于基因表达可视化图像为散点图, 不便于进行细胞分割, 因此需要进行预处理, 在本实施例中, 首先将细胞的基因表达可视化图像处理成中值图。然后可使用拉普拉斯算子 (laplacian 算子) 对中值图进行锐化处理, 以增强中值图的边界效果, 进而得到锐化图。

5 可选的, 对基因表达可视化图像进行预处理, 得到中值图, 具体可包括: 首先利用预设尺寸 (如 13×13) 的卷积核对基因表达可视化图像进行卷积操作, 使得基因表达可视化图像中的散点粘连, 得到第一卷积图; 再根据图像二维灰度峰值检测第一卷积图的局部最大值点; 获取局部最大值点中的第 p 百分位数, 其中, p 为预设数值, 如第 p 百分位数可为 98% 分位值、或 99% 分位值等; 若第 p 百分位数在预设范围内, 则利用第一中值滤波器对
10 第一卷积图进行中值过滤, 得到中值图, 其中, 第一中值滤波器的滤波器大小是根据预设尺寸确定得到的。

进一步可选的, 在获取局部最大值点中的第 p 百分位数之后, 本实施例方法还可包括: 若第 p 百分位数在预设范围外, 则根据第 p 百分位数和预设尺寸, 确定卷积核的新尺寸; 利用新尺寸的卷积核对基因表达可视化图像进行卷积操作, 使得基因表达可视化图像中的
15 散点粘连, 得到第二卷积图; 利用第二中值滤波器对第二卷积图进行中值过滤, 得到中值图, 其中, 第二中值滤波器的滤波器大小是根据新尺寸确定得到的。

示例性的, 根据第 p 百分位数和预设尺寸, 确定卷积核的新尺寸, 具体可包括: 按照公式 $K=N*(N/R)$, 计算得到卷积核的新尺寸, 其中, $K \times K$ 表示卷积核的新尺寸, $N \times N$ 表示预设尺寸, R 表示第 p 百分位数。

20 例如, 首先使用大小 13×13 (经验数值) 的卷积核, 对基因表达可视化图像的原图进行卷积操作使散点图粘连, 得到第一卷积图; 再根据图像二维灰度峰值检测第一卷积图的局部最大值点, 取出所得所有局部最大值点中的 99% 分位值 R , 如 R 值与一个设定好的经验阈值差距过大 (过高或过低, 均会影响后续处理), 则认为 13×13 卷积核不适用于原图, 此时计算 $K=13*(13/R)$, 将原本 13×13 的卷积核改为 $K \times K$ 的卷积核, 重新处理原图, 得到第二
25 卷积图; 如 R 值在允许范围内, 则沿用第一卷积图。卷积图表现为不同灰度的方块层叠的形式, 灰度跨度较剧烈, 且部分表达图像存在聚集点簇中心空洞的现象, 为消除这一问题, 使用 $13 \times 2.7=35$ (2.7 可为经验数值, 小数向下取整) 为滤波器大小 (如需要计算 K , 则此处核大小为 $K \times 2.7$) 对第二卷积图进行中值滤波, 得到中值图。

30 为了填充点簇空洞, 上述所用中值滤波器较大, 中值图的灰度边界可能较为模糊, 使用 laplacian 算子对中值图进行锐化以增强灰度边界, 得到锐化图, 完成预处理, 下面执行细胞分割的过程, 具体可执行步骤 204 至 207 所示的过程。

步骤 204、将锐化图进行二值化处理, 得到初始掩模图。

使用大津法对上一步得到的锐化图进行二值化, 得到初始掩模图。

步骤 205、过滤初始掩模图中面积不符合预设条件的连通域, 得到过滤掩模图。

35 可选的, 步骤 205 具体可包括: 将初始掩模图中面积大于第一预设阈值的连通域、或面积小于第二预设阈值的连通域进行过滤, 得到过滤掩模图, 其中, 第一预设阈值大于第二预设阈值。例如, 使用经验阈值对初始掩模中面积过小或过大的连通域进行过滤, 得到过滤掩模图。

40 步骤 206、遍历过滤掩模图中的每个连通域, 基于连通域的最小外接矩形提取连通域所在区域, 利用分水岭算法对存在细胞粘连的连通域进行分割, 得到分割后的掩模图。

可选的, 步骤 206 具体可包括: 将过滤掩模图中的每个连通域内的像素点的灰度值设置为第一数值, 并将在每个连通域外的像素点的灰度值设置为第二数值; 针对连通域内每个灰度值为第一数值的目标像素点, 将目标像素点的灰度值重新映射为目标像素点到距其最近的灰度值为第二数值的像素点的距离, 得到连通域的距离图谱; 对连通域的距离图谱
45 进行二值化处理, 获得连通域中距离灰度值为第二数值的像素点最远的预设个数像素点; 将预设个数像素点作为分水岭算法的注水点, 利用分水岭函数对过滤掩模图中连通域的原

始掩模进行分水岭分割，得到分割后的目标连通域，并将目标连通域覆盖到过滤掩模图，得到分割后的掩模图。

例如，遍历过滤掩模图中的每个连通域，基于连通域的最小外接矩形提取连通域所在区域，使用基于距离变换的分水岭算法对可能存在细胞粘连的连通域进行进一步分割，具体步骤如下：

步骤 a、对每个提取出的连通域区域，目标连通域内的点灰度值设置为 1，连通域外的点灰度值设置为 0（包括背景和非目标连通域），对每个值为 1 的点进行距离变换，将其灰度值重新映射为该点到距其最近的值为 0 的点的距离（相邻点距离为 1），得到该连通域的距离图谱。

步骤 b、使用一个经验阈值对距离图谱进行二值化，得到连通域中距离值为 0 的点最远的数个点（由于每个连通域情况不同，各自数量不定）。

步骤 c、将上一步得到的数个点作为分水岭算法的注水点，使用 opencv 中的分水岭函数对连通域的原始掩模进行分水岭分割，得到分割后的目标连通域，将该结果覆盖至过滤掩模图。

步骤 d、对遍历的每个连通域进行上述步骤，得到分割后的掩模图。

步骤 207、对分割后的掩模图进行闭包操作，得到细胞分割结果的掩模图。

为使细胞边界更规则，对分割后的掩模进行闭包操作，得到最终掩模图。最后可输出并保存最终掩模图，用于后续结合原始表达矩阵进行细胞层面的生物学分析。

例如，如图 30 所示，为基于本实施例方法的示例流程示意图。首先可输入基因表达矩阵，基于矩阵可生成基因表达可视化图像，如图 31 所示。然后利用 13×13 卷积核对基因表达可视化图像进行卷积操作，使得基因表达可视化图像中的散点粘连，得到第一卷积图。此时需要通过阈值进行判断，即根据图像二维灰度峰值检测第一卷积图的局部最大值点，取出所得所有局部最大值点中的 99% 分位值 R，如 R 值与一个设定好的经验阈值差距过大（过高或过低，均会影响后续处理），则认为 13×13 卷积核不适用于原图，使用比例计算新的卷积核大小，然后使用新的卷积核处理原图得到第二卷积图，使用这一比例计算中值过滤器大小，然后用这个大小的中值过滤器处理第二卷积图得到中值图。而如果 R 值在允许范围内，则沿用第一卷积图，使用大小为 35 的中值过滤器处理第一卷积图得到中值图。

在得到中值图之后，使用 laplacian 算子对中值图进行锐化处理，得到锐化图，如图 32 所示。采用大律法对锐化图进行二值化处理得到初始掩模图，如图 33 所示，然后通过面积过滤和分水岭算法分割粘连细胞。最后输出并保存细胞掩模图像，如图 34 所示。

与目前现有的细胞分割方式相比，本实施例提供一种直接基于基因表达可视化图像进行细胞分割的方案，使用多种图像处理方法相结合，可提供较为可靠的细胞分割结果。细胞分割不依赖影像图，不需要额外引入将影像图与基因表达可视化图像进行配准的技术，排除了引入的额外误差，同时节省了整体操作时间和技术成本，可提高细胞分割处理的效率和准确性。

在一些实施例中，如图 35 所示，图 35 为本申请实施例提供的一种对配准后的图像进行细胞分割的方法流程示意图，包括：

步骤 1036，基于所述染色图像确定所述目标对象的各个组织类型对应的标注区域和细胞面积；

具体地，可以结合染色图提供的组织信息和目标对象相关先验知识，基于染色图确定目标对象所包含的各个组织类型对应的标注区域。其中，目标对象相关先验知识可以是目标对象包含哪些细胞组织，也可以是这些组织的细胞形态。在一种可选的情况下，可以通过特定的手动工具画出目标对象各个细胞组织的组织轮廓，并且可以通过不同颜色对不同组织对应的区域进行标注。其中，特定的手动工具可以是开源工具 3Dslicer，支持手动标注，还可以结合开源工具 plant-seg 网络进行组织区域的标注。

本步骤中的细胞面积可以是预先为目标对象的组织类型设定的参考细胞面积。也就是说，预先将各个组织类型的细胞面积进行设定，该设定好的细胞面积可以根据经验设定，

也可以根据不同对象的同一类型的组织细胞的平均值来设定。所以在执行到本步骤时，可以直接获取该预先设定好的细胞面积，将其作为标注区域对应的细胞面积，进而执行后续的处理。

5 步骤 1037，依据所述细胞面积确定每个所述标注区域包含的待处理子区域的面积，所述待处理子区域的面积与所述细胞面积正相关，所述待处理子区域用于表征对应的所述标注区域的最小处理单元；

其中，标注区域对应的待处理子区域可以作为该标注区域的最小处理单元进行组织细胞的处理和分析，在利用基因芯片提取目标对象内细胞的位置信息时，将一个待处理子区域作为一个单元来进行提取，以待处理子区域之间的位置关系来表征该标注区域内细胞间的位置关系。因为待处理子区域的面积与细胞面积呈正相关，也就是说，上一步骤中确定出的标注区域的细胞面积越大，该标注区域被划分出的待处理子区域就越大。所以能够更好的表达出不同组织类型所对应的标注区域内细胞间的位置关系。

步骤 1038，依据每个所述待处理子区域的面积，将对应的所述标注区域进行网格化处理，得到每个所述标注区域对应的网格化坐标；

15 具体地，对标注区域进行网格化处理的一种可选方式可以通过在每个标注区域中分别划分面积大小为对应的待处理子区域面积的网格，并得到网格化坐标。

可选地，还可以将网格的中心点坐标作为该网格的网格化坐标，其中，网格中心点坐标可以通过在捕获单元建立的坐标中获取网格所在的坐标区间，计算坐标区间的中心点坐标，确定为该网格的中心点坐标。

20 步骤 1039，从所述基因表达可视化图像中提取所述网格化坐标对应的数据，并叠加至所述网格化坐标，得到所述目标对象的细胞处理结果。

具体地，在基因表达数据中提取各个网格化坐标对应的数据，得到目标对象的组织细胞处理数据，其中，包含目标对象的不同基因在各个组织类型的细胞组织中的表达水平，进一步结合对应的网格化坐标，生成目标对象的组织细胞处理结果。

25 在一种可能的实现方式中，基因表达数据包括基因标识符 ID、基因坐标数据 x 、 y ，以及基因表达量 MIDCount。由于此时已经确定网格化坐标，则可通过网格化坐标，确定每个网格的坐标范围。然后对比每个数据的基因坐标数据 x 、 y 与每个网格的坐标范围，如果基因坐标数据 x 、 y 落在了哪个网格的坐标范围内，则意味着该数据要叠加至这个网格，此时，可以在表达数据中为该数据设置其对应的网格的 ID，即 Cell ID，在后续进行数据叠加时，根据添加的 Cell ID，就可以明确每个数据落在哪个网格中，也就是需要将数据叠加至哪个网格的中心点坐标上。

在一些实施例中，如图 36 所示，图 36 为本申请实施例提供的一种确定基因表达可视化图像中未知所属细胞的背景基因分子的所属细胞的方法流程示意图，包括：

35 步骤 1041，基于所述细胞掩模图，将所述基因表达可视化图像的每个基因表达特征分别确定为位于细胞内的细胞基因表达特征及位于细胞外的背景基因表达特征；

在本步骤中，基于所述细胞掩模图，将基因图像的每个基因表达特征分别确定为位于细胞内的细胞基因表达特征及位于细胞外的背景基因表达特征。

在本实施例中，基因表达特征为 DNB，参考图 37 所示，图 37 为 DNB 与细胞膜的对应关系示意图，其中， x 和 y 分别为 DNB 坐标，geneID、MIDCount 分别表征基因信息，label 表征细胞标识。将 Stereo-seq 空间转录组数据的每个 DNB 通过对应关系，得到 DNB 所属的细胞信息，其余不在细胞内的 DNB 作为背景 DNB。

步骤 1042，对每个所述细胞基因表达特征，通过空间位置和基因表达特征的信号值拟合出概率分布模型；

45 在本步骤中，对每个所述细胞基因表达特征，通过空间位置和基因表达特征的信号值拟合出概率分布模型。

作为优选实施方式，在本实施例中，概率分布模型可以为高斯混合模型，但可以采用通过不同的概率分布模型，以及对背景信号所属细胞判断中，可通过贝叶斯模型进行所属细胞判断，可根据实际需求进行相应的选择及调整。

50 在本步骤中，对每个细胞内的 DNB，通过坐标 (x, y) 以及 DNB 信号值拟合出高斯混合模型。

步骤 1043，基于所述概率分布模型，确定出针对位于一目标细胞的预设范围内的目

标背景基因表达特征属于所述目标细胞的概率值；

在本步骤中，基于概率分布模型，确定出针对位于一目标细胞的预设范围内的目标背景基因表达特征属于该细胞的概率值。其中，目标细胞可以是生物样本包含的所有细胞，也可以是生物样本包含的部分细胞。

5 具体地，作为可选实施方式，在本步骤中，以目标细胞为细胞 a 为例，对于细胞 a，通过高斯混合模型计算以细胞 a 的中心范围 50x50 像素点内的背景 DNB 属于该细胞 a 的概率值。

步骤 1044，根据所述概率值输出校正结果以进行背景基因表达特征的校正。

基于概率值进行背景基因表达特征的校正。

10 具体地，在本步骤中，判断概率值是否大于或等于预设概率值，若是，将目标背景基因表达特征作为位于目标细胞内的细胞基因表达特征，以进行背景基因表达特征的校正，若否，将目标背景基因表达特征继续作为背景基因表达特征。

参考图 38 所示，图 32 校正背景 DNB 的校正结果示意图，对于背景 DNB，将小于预设概率值的作为背景 DNB，大于或等于预设概率值的作为细胞的 DNB。

15 在本实施例中，并不具体限定预设概率值，可根据实际需求进行相应的选择及调整。

作为可选实施方式，在本步骤中，响应于目标背景基因表达特征同时位于至少两个目标细胞外，分别确定出属于每一个对应的目标细胞的概率值；从确定出的至少两个概率值中选取最高的概率值作为输出校正结果的概率值，并且将最高的概率值对应的目标细胞作为目标背景基因表达特征所属的细胞。

20 以目标细胞为细胞 a 和细胞 b 为例，假设对于背景基因表达特征 A 同时在细胞 a 和细胞 b 的中心范围 50x50 像素点内，计算背景基因表达特征 A 所属于细胞 a 的概率值为 x_1 ，背景基因表达特征 A 所属于细胞 b 的概率值为 x_2 ，若 x_1 大于 x_2 ，则确定背景基因表达特征 A 为细胞 a 的基因表达特征。

25 在一些实施例中，如图 39 所示，图 39 为本申请实施例提供的一种确定基因表达可视化图像中未知所属细胞的背景基因分子的所属细胞的方法流程示意图，包括：

步骤 1045，根据所述基因表达可视化图像，确定所属于目标细胞的初始基因分子以及未知所属细胞的背景基因分子；

步骤 1046，获取所述初始基因分子的坐标信息；

30 将显微镜图像和基因图像按照坐标进行对齐，通过显微镜图的细胞分割边界，即可以获取细胞内初始基因分子的坐标信息。

步骤 1047，根据所述坐标信息，确定所述目标细胞的中心点坐标；

步骤 1048，根据所述中心点坐标和所述坐标信息，构建维诺图；

35 其中，维诺图（Voronoi 图，又叫泰森多边形或 Dirichlet 图）是由一组由连接两邻点直线的垂直平分线组成的连续多边形组成。N 个在平面上有区别的点，按照最邻近原则划分平面；每个点与它的最近邻区域相关联。

步骤 1049，确定所述维诺图范围内的背景基因分子的所属细胞。

根据对维诺图范围内的背景基因分子设置的标签，可以将维诺图范围内的背景基因分子确定完全属于目标细胞，也可以确定部分属于目标细胞，也可以确定完全不属于目标细胞。

40 在本实施例中，通过所属于目标细胞的初始基因分子的坐标信息，确定目标细胞的中心点坐标，再根据目标细胞的中心点坐标构建维诺图，确定维诺图范围内的背景基因分子的所属细胞，提高了校正效率，同时也提高了校正的精确度。另外，本实施例只要能确定目标细胞内的初始基因分子即可，无需了解目标细胞的轮廓图，因此，降低了对细胞分割高精度的要求，减少对目标细胞形态影像图的依赖程度，提高了校正方案的稳健性。

45 在一些实施例中，所述根据基因图像，确定所属于目标细胞的初始基因分子以及未知所属细胞的背景基因分子的步骤包括：

获取生物样本的显微镜图像及基因图像；

对生物样本的显微镜图像及基因图像进行图像配准；

对所述显微镜图像进行细胞分割以获取细胞分割结果；

50 基于所述细胞分割结果和所述基因图像的图像配准结果，确定目标细胞的基因图像；根据所述目标细胞的基因图像，确定所属于目标细胞的初始基因分子以及未知所属细

胞的背景基因分子。

在一些实施例中，所述根据所述坐标信息，确定所述目标细胞的中心点坐标的步骤包括：

统计所述初始基因分子的分子数量；

5 计算所述初始基因分子的横坐标数值总和以及纵坐标数值总和；

将第一比值确定为所述中心点坐标的横坐标值；所述第一比值为所述初始基因分子的横坐标数值总和与所述分子数量的比值；

将第二比值确定为所述中心点坐标的纵坐标值；所述第二比值为所述初始基因分子的纵坐标数值总和与所述分子数量的比值。

10 在一些实施例中，所述根据所述坐标信息，确定所述目标细胞的中心点坐标的步骤包括：

获取所述初始基因分子中最大的横坐标值、最小的横坐标值、最大的纵坐标值以及最小的纵坐标值；

15 将第三比值确定为所述中心点坐标的横坐标值；所述第三比值为最大的横坐标值与最小的横坐标值之差的二分之一；

将第四比值确定为所述中心点坐标的纵坐标值；所述第四比值为最大的纵坐标值与最小的纵坐标值之差的二分之一。

在一些实施例中，所述基因图像数据校正方法还包括：

20 若所述背景基因分子在不同的维诺图范围内，则将所述背景基因分子输入高斯混合模型，分别确定所述背景基因分子属于距离预设范围内的目标细胞的概率分数；

将所述背景基因分子校正为所属于所述概率分数高的目标细胞的基因分子。

在一些实施例中，所述确定所述维诺图范围内的背景基因分子的所属细胞的步骤后包括：

根据校正后所属于所述目标细胞的基因分子的结果生成目标细胞基因表达信息。

25 在一些实施例中，所述根据所述中心点坐标和所述坐标信息，构建维诺图的步骤包括：对所有所述中心点坐标构建多边形网；

根据所述多边形网，确定维诺图的边；

根据所述维诺图的边，构建维诺图。

30 在一些实施例中，如图 40 所示，图 40 为本申请实施例提供的一种确定基因表达可视化图像中未知所属细胞的背景基因分子的所属细胞的方法流程示意图，包括：

步骤 10410，确定所述基因表达可视化图像中所属于目标细胞的基因分子的初始分子数量以及所述所属于目标细胞的基因分子所占的目标区域的初始轮廓面积；

35 采用凸包算法，对细胞分割结果进行运算确定所属于目标细胞的基因分子所占的目标区域的初始轮廓面积。其中，凸包算法（Convex Hull Algorithm）即根据细胞分割结果，在目标区域内寻找 N 个点的凸包，这 N 个点之间的连接包围的就是初始轮廓面积。

步骤 10411，根据所述分子数量和所述初始轮廓面积，确定所述目标细胞的基因分子密度；

在一个可选的实施方式中，将分子数量和初始轮廓面积的比值确定为目标细胞的基因分子密度。

40 步骤 10412，确定所述目标区域的预设区域范围内的背景基因分子的分子数量以及所述预设区域范围的区域面积；

其中，背景基因分子为根据基因图像不能确定所属细胞的基因分子，预设区域范围位于目标区域外，但不能超过所属于邻近细胞的基因分子所占的区域。可选区域面积为 5 像素*5 像素作为预设区域范围。

45 在一个可选的实施方式中，预设区域范围可以是以根据目标细胞的中心点为中心确定的区域。目标细胞的中心点由基因图像确定，目标细胞的中心点的示意图参见图 41。

步骤 10413，根据所述背景基因分子的分子数量和区域面积，确定背景基因分子密度；

在一个可选的实施方式中，将背景基因分子的分子数量和区域面积的比值确定为背景基因分子密度。

50 步骤 10414，根据所述基因分子密度与所述背景基因分子密度确定所述背景基因分子的所属细胞。

在本实施例中，利用上述细胞特性，根据基因图像中所属于目标细胞的基因分子的初始分子数量与目标细胞初始轮廓面积，确定目标细胞的基因分子密度。比较目标细胞的基因分子密度和背景基因分子密度，根据基因分子密度与背景基因分子密度确定背景基因分子的所属细胞，大幅度提高了校正目标细胞内的基因分子的效率，同时也增加了校正目标细胞内的基因分子的精确度。

在一些实施例中，如图 42 所示，图 42 为本申请实施例提供的一种确定基因表达可视化图像中未知所属细胞的背景基因分子的所属细胞的方法流程示意图，包括：

步骤 10415，确定所述基因表达可视化图像中所属于目标细胞的基因分子所占的目标区域；

步骤 10416，确定所述目标区域的预设校正范围内的未知所属细胞的背景基因分子的空间坐标及基因表达特征信号值；

其中，本实施例的基因表达特征为 DNB，将 Stereo-seq 空间转录组数据的每个 DNB 通过对应关系，可以得到 DNB 所属的细胞信息，其余不在细胞内的 DNB 作为背景 DNB。背景基因分子的空间坐标可以用背景 DNB 的坐标表示。

步骤 10417，将所述空间坐标和所述基因表达特征信号值输入概率分布模型，以得到所述背景基因分子所属于所述目标细胞的概率分数；所述概率分布模型根据所属于目标细胞的基因分子样本的空间坐标及基因表达特征信号值拟合得到；

拟合 GMM 模型（高斯混合模型）时，自变量有三个：空间坐标（x，y）和信号值。因变量即输出是每个基因分子属于目标细胞的概率分数。具体实现过程为：引入了 Scikit-learn（一种机器学习库）库中的 Gaussian Mixture（高斯混合），利用输入数据自动拟合 GMM 模型的参数信息，得到最终的概率模型。

其中，概率分布模型根据所属于目标细胞的基因分子样本的空间坐标及基因表达特征信号值样本拟合得到。作为优选实施方式，在本实施例中，概率分布模型可以为高斯混合模型。

步骤 10418，根据所述概率分数确定所述背景基因分子的所属细胞。

比较概率分数是否大于或者等于概率阈值。

若概率分数大于或者等于概率阈值，则执行将背景基因分子校正为所属于目标细胞的基因分子。若概率分数小于概率阈值，背景基因分子仍然是未知所属细胞的背景分子。

在一个可选的实施方式中，将第一概率分数与第二概率分数之间的第三四分位数确定为概率阈值。

其中，第一概率分数为概率分数中分数值最高的概率分数；第二概率分数为概率分数中分数值最低的概率分数。

在本实施例中，在大致了解目标细胞的目标区域前提下，计算预设校正范围内的未知所属细胞的背景基因分子的所属于目标细胞的概率分数，并根据概率分数确定背景基因分子的所属细胞。减少了因基因分子扩散的原因导致基因图像数据误差大的情况，提高了校正效率。同时不依赖于高精度的细胞分割结果，增强了校正方案的稳健性。

进一步的，本申请还提供芯片上的 track cross 的检测，主要是为了后续的拼接修正和配准。拼接和配准都有用 track 的检测的结果。

图 43 示出了根据本发明的一个实施例的染色图像中的点坐标的确定方法 100 的流程图。所述方法 100 可以包括采样步骤 1001、分组步骤 1002、直线拟合步骤 1003 以及坐标计算步骤 1004。

为了便于理解本发明的确定方法，现在参考图 44 示例性说明本文所涉及的染色图像。图 38 示出了本发明适用的一个示例染色图像。该染色图像可以通过摄像机、染色图像传感器等获取。该染色图像可以是具有直轨迹线的例如机械结构、载体、载片或生物芯片的染色图像。如图 38 中所示出的，该染色图像所在的染色图像平面可包括第一方向 X 和与第一方向 X 垂直的第二方向 Y，该染色图像包括横向于第二方向 Y 的多个第一直轨迹线（即，该染色图像中的大体上平行于第一方向 X 的多个暗线）和横向于第一方向 X 的多个第二直轨迹线（即，该染色图像中的大体上平行于第二方向 Y 的多个暗线）。具体地，在该示例染色图像中存在 6 个第一直轨迹线和 8 个第二直轨迹线。由于每个第一直轨迹线与每个第二直轨迹线相交，因此在该示例染色图像中存在共 $6 \times 8 = 48$ 个交点。如图 38 中所示出的，该

染色图像可以具有像素长度 L 和像素宽度 W 。在图 38 中还示出了示例第二采样条。作为示例而非限制，该采样条的像素长度与该示例染色图像的像素长度相同，即为 L ，并且该采样条的像素宽度小于该示例染色图像的像素宽度，为 W_2 。在本文中，一直轨迹线横向于任一方向是指该直轨迹线平行于垂直于该任一方向的方向或与垂直于该任一方向的方向的夹角小于一预设值。所述预设值的范围可以包括 1 到 10 度、10 到 20 度、20 到 30 度、30 到 40 度或 40 到 50 度等，例如所述预设值可以是 1 度、2 度、3 度、4 度、5 度、6 度、7 度、8 度、9 度、10 度。

采样步骤 1001 包括第一方向采样步骤和第二方向采样步骤。

在第一方向采样步骤中，应用第一采样条沿所述第一方向对所述染色图像进行第一方向等间隔采样以确定第一候选点集合，其中对于每次第一方向等间隔采样，执行以下第一方向采样子步骤一和第一方向采样子步骤二。第一方向采样子步骤一包括：将所述第一采样条在该次第一方向等间隔采样期间所选中的所述染色图像的采样区域内的像素按所述第一方向进行像素值累加求和以得到一系列第一像素值和，其中所述一系列第一像素值和和所述第二方向上的任一第二方向位置处的值等于所述采样区域内的、在该第二方向位置处沿所述第一方向对准的各像素的像素值之和。第一方向采样子步骤二包括：确定所述一系列第一像素值和的极值点，并且对于确定的每个极值点：确定该次第一方向等间隔采样所获得的一个对应第一候选点，该对应第一候选点在所述第二方向上的坐标为该极值点所对应的第二方向位置在所述第二方向上的坐标，该对应第一候选点在所述第一方向上的坐标基于在该次第一方向等间隔采样期间所述第一采样条在所述第一方向上的位置确定。如此，由每次第一方向等间隔采样所获得的第一候选点形成所述第一候选点集合。

在第二方向采样步骤中，应用第二采样条沿所述第二方向对所述染色图像进行第二方向等间隔采样以确定第二候选点集合，其中对于每次第二方向等间隔采样，执行以下第二方向采样子步骤一和第二方向采样子步骤二。第二方向采样子步骤一包括：将所述第二采样条在该次第二方向等间隔采样期间所选中的所述染色图像的采样区域内的像素按所述第二方向进行像素值累加求和以得到一系列第二像素值和，其中所述一系列第二像素值和和所述第一方向上的任一第一方向位置处的值等于所述采样区域内的、在该第一方向位置处沿所述第二方向对准的各像素的像素值之和。第二方向采样子步骤二包括：确定所述一系列第二像素值和的极值点，并且对于确定的每个极值点：确定该次第二方向等间隔采样所获得的一个对应第二候选点，该对应第二候选点在所述第一方向上的坐标为该极值点所对应的第一方向位置在所述第一方向上的坐标，该对应第二候选点在所述第二方向上的坐标基于在该次第二方向等间隔采样期间所述第二采样条在所述第二方向上的位置确定。如此，由每次第二方向等间隔采样所获得的第二候选点形成所述第二候选点集合。

在采样步骤 1001 中所用的所有极值点可以同为极大值点或同为极小值点。例如，若所述直轨迹线为暗线，则所用的所有极值点可同为极小值点；若所述直轨迹线为亮线，则所用的所有极值点可同为极大值点。

在分组步骤 1002 中，将所述第一候选点集合中的第一候选点分成多个第一候选点组；以及将所述第二候选点集合中的第二候选点分成多个第二候选点组。

在一个实施例中，以第一分组方式将所述第一候选点集合中的第一候选点分成多个第一候选点组，其中所述第一分组方式被配置成使得每个第一候选点组中的各第一候选点基本上呈横向于所述第二方向的一直线排列。并且，以第二分组方式将所述第二候选点集合中的第二候选点分成多个第二候选点组，其中所述第二分组方式被配置成使得每个第二候选点组中的各第二候选点基本上呈横向于所述第一方向的一直线排列。

在直线拟合步骤 1003 中，对每个第一候选点组中的第一候选点进行直线拟合以得到一个相应的第一直线解析式；以及对每个第二候选点组中的第二候选点进行直线拟合以得到一个相应的第二直线解析式。

基于已知坐标的多个点进行直线拟合以得到直线解析式是本领域已知的，在此不作赘述。常用的直线拟合方法包括：最小二乘法；梯度下降法；高斯牛顿、列-马法。

在坐标计算步骤 1004 中，计算得到的每个第一直线解析式与每个第二直线解析式的交点的坐标。对于已知的两个直线解析式求其交点的坐标的方法是本领域已知的，在此不作赘述。

图 45 为本申请实施例提供的一种有向目标检测方法，包括：

步骤 1005, 获取待测染色图像或基因表达可视化图像。

步骤 1006, 将所述待测染色图像或基因表达可视化图像输入有向目标检测网络, 以获得目标检测结果。

具体地, 待测染色图像或基因表达可视化图像输入到训练后的基于 YOLOV6 的有向目标检测模型中, 利用有向目标检测模型中训练好的多级卷积神经网络对待检测图像进行卷积操作、特征图的融合, 输出不同尺度的特征融合图; 对于不同尺度的特征融合图, 分别采用对应的预测器来预测类别置信度、目标框定位置信度、目标检测置信度和角度分类置信度。

其中, 所述 Backbone 层采用 RepVGG 架构, 并且所述 Neck 层采用 FPN-PAN 架构。

所述 Head 层采用解耦头, 使得经由分类分支执行类别分析以获得所述分类信息, 并经由回归分支执行回归分析以分别获得所述目标框坐标预测信息、所述目标预测信息和所述角度预测信息。用于表示类别分析结果的分析信息、用于确定目标检测框位置的目标框坐标预测信息、用于确定目标检测框内是否存在对应目标的目标预测信息和用于确定目标检测框的角度的角度预测信息, 对应目标可以包括 track 线或 track cross (也称轨迹交叉点)。本申请实施例提供一种用于训练基于 YOLOV6 的有向目标检测网络的方法, 所述方法包括: 获取多个训练图像和所述多个训练图像中包括标签的标签信息; 以及将所述多个训练图像和所述标签信息输入所述基于 YOLOV6 的有向目标检测网络以通过训练获得所述基于 YOLOV6 的有向目标检测网络的设置参数; 其中, 所述基于 YOLOV6 的有向目标检测网络包括: 主干 Backbone 层, 被配置为对训练图像执行特征提取, 以获得不同尺度的多个特征图; 颈部 Neck 层, 被配置为从 Backbone 层接收所述多个特征图, 并对所述多个特征图进行跨尺度融合, 以获得多个融合特征图; 以及检测头 Head 层, 被配置为对所述多个融合特征图进行预测, 以获得关于所述训练图像的预测结果, 其中所述预测结果包括: 分类信息、目标框坐标预测信息、目标预测信息和角度预测信息。

在一些实施例中, 所述 Backbone 层采用 RepVGG 架构, 并且所述 Neck 层采用 FPN-PAN 架构。

在一些实施例中, 所述 Head 层采用解耦头, 使得经由分类分支执行类别分析以获得所述分类信息, 并经由回归分支执行回归分析以分别获得所述目标框坐标预测信息、所述目标预测信息和所述角度预测信息。

在一些实施例中, 将所述多个训练图像和所述标签信息输入所述基于 YOLOV6 的有向目标检测网络以通过训练获得所述基于 YOLOV6 的有向目标检测网络的设置参数包括: 在每次训练中,

获取针对当前训练的相应训练图像和相应标签信息;

使用所述基于 YOLOV6 的有向目标检测网络对所述相应训练图像中的目标进行预测, 以获得相应预测结果;

根据相应标签信息和相应预测结果, 计算所述基于 YOLOV6 的有向目标检测网络的损失函数; 以及

基于所述损失函数的计算结果, 调整所述基于 YOLOV6 的有向目标检测网络的设置参数, 直到所述损失函数的计算结果收敛为止。

在一些实施例中, 还包括: 对所获取的具有第一数据形式的标签信息进行预处理, 以将所述标签信息处理为 $[x, y, w, h, \theta, \text{classID}]$ 形式的数据, 并将经处理的数据输入所述基于 YOLOV6 的有向目标检测网络,

其中, (x, y) 表示目标检测框的中心的位置, w 表示所述目标检测框的长边宽度, h 表示所述目标检测框的短边高度, θ 表示所述目标检测框的长边相对于水平轴旋转的角度, 且 classID 表示所述目标检测框的类别。

在一些实施例中, 所述标签信息的第一数据形式为 $[x1, y1, x2, y2, x3, y3, x4, y4, \text{classID}]$, 其中, $(x1, y1)$ 、 $(x2, y2)$ 、 $(x3, y3)$ 和 $(x4, y4)$ 分别定义目标检测框的四个顶点的位置。

在一些实施例中, 还包括: 对所获得的 $[x, y, w, h, \theta, \text{conf}, \text{classID}]$ 形式的预测结果进行后处理, 以将所述预测结果处理为具有第二数据形式的数据, 并将所述具有第二数据形式的数据呈现给用户,

其中, (x, y) 表示目标检测框的中心的位置, w 表示所述目标检测框的长边宽度, h

表示所述目标检测框的短边高度， θ 表示所述目标检测框的长边相对于水平轴旋转的角度， conf 表示所述目标检测框具有相应目标的置信度，且 classID 表示所述目标检测框的类别。

在一些实施例中，所述第二数据形式为 $[x1, y1, x2, y2, x3, y3, x4, y4, \text{conf}, \text{classID}]$ ，其中， $(x1, y1)$ 、 $(x2, y2)$ 、 $(x3, y3)$ 和 $(x4, y4)$ 分别定义所述目标检测框的四个顶点的位置。

进一步的，如图 46 所示，图 46 为本申请实施例提供的一种生物样本的细胞基因表达信息获取方法包括：

步骤 1101，获取生物样本的第一图像，所述第一图像包括所述生物样本经染色的细胞边界；

步骤 1102，获取所述生物样本的第二图像，所述第二图像包括所述生物样本经染色的遗传物质；

步骤 1103，获取所述生物样本的基因表达量信息；

步骤 1104，将所述第二图像与所述第一图像配准，得到第二图像与第一图像的配准图像；

步骤 1105，将所述第二图像与所述基因表达量信息进行配准，得到第二图像与基因表达量信息的配准图像；

步骤 1106，根据第二图像与第一图像的配准图像、第二图像与基因表达量信息的配准图像获取第一图像与基因表达量信息的配准图像；

步骤 1107，根据第一图像与基因表达量信息的配准图像获取所述生物样本的细胞边界；

步骤 1108，基于第一图像与基因表达量信息的配准图像根据所述细胞边界获取所述生物样本的细胞的基因表达信息。

在一些实施例中，以 ssDNA 染液和细胞边界荧光染色液对生物样本的细胞进行染色，其中，所述 ssDNA 染液包括缓冲液和 ssDNA 试剂，所述细胞边界荧光染色液包括荧光染料。ssDNA 染液可以对细胞的遗传物质进行染色，细胞边界荧光染色液可以对细胞的边界进行染色。

在一些实施例中，所述细胞边界包括细胞壁或细胞膜。

在一些实施例中，所述组织样本的细胞的基因表达信息包括载有所述组织样本的芯片上每个位点的位置信息、基因表达量和细胞类型。

在一些实施例中，将所述第二图像与所述基因表达量信息进行配准，得到第二图像与基因表达量信息的配准图像包括：

根据所述第二图像中的 track 线和所述基因表达量信息的 track 线对所述第二图像和所述基因表达量信息进行配准，得到所述第二图像与基因表达量信息的配准图像。

在一些实施例中，通过细胞染色试剂对组织样本进行染色后进行呈像，所述呈像通过 FITC 通道获取所述组织样本的第二图像，通过 DAPI 通道获取所述组织样本的第一图像，所述细胞染色试剂包括：ssDNA 染液和细胞边界荧光染色液；所述 ssDNA 染液包括缓冲液和 ssDNA 试剂；所述细胞边界荧光染色液包括荧光染料。

需要说明的是，上述序号的顺序不代表具体的执行步骤。

与上述的空间组学单细胞数据获取方法相对应，本发明还提出一种空间组学单细胞数据获取装置。由于本发明的装置实施例与上述的方法实施例相对应，对于装置实施例中未披露的细节可参照上述的方法实施例，本发明中不再赘述。

图 47 为本申请实施例提供的一种空间组学单细胞数据获取装置的结构示意图，如图 47 所示，包括：

获取单元 21, 用于获取生物样本的染色图像, 获取所述生物样本的基因表达可视化图像;

配准单元 22, 用于将所述染色图像与所述基因表达可视化图像进行配准, 得到配准后的图像;

5 第一分割单元 23, 用于对所述配准后的图像进行细胞分割, 得到细胞掩模图;

修正单元 24, 用于根据所述细胞掩模图确定所述基因表达可视化图像中未知所属细胞的背景基因分子的所属细胞;

标记单元 25, 用于对所属细胞标记, 得到带细胞标记的空间组分子表达矩阵。

10 获取生物样本的染色图像, 获取所述生物样本的基因表达可视化图像; 将所述染色图像与所述基因表达可视化图像进行配准, 得到配准后的图像; 对所述配准后的图像进行细胞分割, 得到细胞掩模图; 根据所述细胞掩模图确定所述基因表达可视化图像中未知所属细胞的背景基因分子的所属细胞, 并对所属细胞标记, 得到带细胞标记的空间组分子表达矩阵。与相关技术相比, 本申请实施例通过基因表达矩阵图、拼接、配准、分割和校正的处理流程, 以获得高精度的空间单细胞数据。

15 进一步地, 在本实施例一种可能的实现方式中, 如图 48 所示, 所述获取单元 21, 还用于:

通过显微镜获取所述生物样本的一个染色图像; 或

通过显微镜获取所述生物样本的由多个子染色图像拼接的拼接染色图像; 或

20 通过所述显微镜获取所述生物样本的多个子染色图像, 对所述多个子染色图像进行拼接, 得到拼接染色图像。

进一步地, 在本实施例一种可能的实现方式中, 如图 48 所示, 所述获取单元 21 包括:

第一获取模块, 用于通过显微镜获取所述生物样本的多个子染色图像; 对所述生物样本的多个子染色图像进行第一拼接, 得到第一染色图像;

25 判断模块, 用于根据所述第一染色图像的图像质量判断是否对所述生物样本的多个子染色图像进行第二拼接;

确定模块, 用于若否, 将所述第一染色图像确定为所述染色图像;

第二获取模块, 用于若是, 对所述生物样本的多个子染色图像进行第二拼接, 得到所述染色图像。

30 进一步地, 在本实施例一种可能的实现方式中, 如图 48 所示, 所述第一获取模块, 还用于:

分别确定多个子染色图像和模板图像的重叠区域;

35 将所述多个子染色图像和所述模板图像的被确定为重叠区域的部分分别裁剪为多个染色图像重叠子图和数量相同的多个模板图像重叠子图, 以配对为多个重叠子图对;

通过对所述多个重叠子图对进行频域计算, 分别获得各子染色图像和所述模板图像之间的偏移量; 以及

将所述各子染色图像分别朝向所述模板图像的方向平移所述偏移量, 并与所述模板图像拼接在一起。

40 进一步地, 在本实施例一种可能的实现方式中, 如图 48 所示, 所述第一获取模块, 还用于:

获取所述多个子染色图像以及关于所述生物样本的轨迹线或轨迹点的模板信息;

对所述多个子染色图像进行预拼接得到所述多个子染色图像的预拼接坐标;

45 根据轨迹线或轨迹点的特征从所述多个子染色图像中选取至少一个第一参考图像, 根据所述生物样本的模板信息和所述第一参考图像的预拼接坐标推导全局模板; 以及

对于所述多个子染色图像中除了所述第一参考图像之外的第二参考图像, 计算所述第二参考图像的预拼接坐标与所述第二参考图像在所述全局模板中的相应模板坐标之间的偏移量, 并基于所述偏移量对所述第二参考图像进行坐标调整以拼接生成所述拼接染色图像。

进一步地, 在本实施例一种可能的实现方式中, 如图 48 所示, 所述配准单元 22

包括：

获取模块，用于获取所述染色图像中的轨迹线，基于所述轨迹线与水平方向的夹角旋转所述染色图像；其中，所述染色图像和所述基因表达可视化图像包括多条轨迹线和/或多个轨迹点，所述轨迹点是由两条所述轨迹线相交形成的，所述多条轨迹线中的

5 每条轨迹线具有对应的索引号；

处理模块，用于将所述染色图像进行放大或缩小得到第三参考图像，所述第三参考图像与所述基因表达可视化图像具有相同的比例；

第一确定模块，用于基于所述基因表达可视化图像和所述第三参考图像确定目标偏移量；

10 配准模块，用于基于所述目标偏移量移动所述第三参考图像，得到所述配准后的图像。

进一步地，在本实施例一种可能的实现方式中，如图 48 所示，所述配准单元 22 还包括：

15 第二确定模块，用于在所述处理模块将所述染色图像进行放大或缩小得到第三参考图像之后，确定所述第三参考图像与基因表达可视化图像之间的相似度；

粗配准模块，用于将相似度最大值对应的位置作为粗配准结果。

进一步地，在本实施例一种可能的实现方式中，如图 48 所示，所述配准单元 22 还用于：

20 获取染色图像和基因表达可视化图像，所述染色图像和所述基因表达可视化图像均包括多条轨迹线和多个轨迹点，所述轨迹点是由两条所述轨迹线相交形成的，每条所述轨迹线具有对应的索引号；

对所述染色图像进行预处理，获取候选图像集合；

确定所述候选图像集合对应的二值图集合；

25 基于所述二值图集合与所述基因表达可视化图像之间的相似度，确定所述二值图集合与所述基因表达可视化图像中具有对应关系的轨迹点对；

基于所述轨迹点对确定偏移信息；

基于所述偏移信息处理目标图像，以得到配准后的图像，所述目标图像为所述轨迹点对所对应的候选图像集合中的图像。

30 进一步地，在本实施例一种可能的实现方式中，如图 48 所示，所述装置还包括：

第二分割单元 26，用于在所述第一分割单元对所述配准后的图像进行细胞分割，

得到细胞掩模图之前，对所述配准后的图像进行组织分割，得到组织掩模图。

进一步地，在本实施例一种可能的实现方式中，如图 48 所示，所述第二分割单元

26 还用于：

35 获取所述配准后的图像中待分割的目标组织的基因表达量并确定所述配准后的图像的基因表达矩阵，所述基因表达矩阵按照像素位置对所述基因表达量进行排列；

通过将所述基因表达矩阵中每一像素位置的基因表达量赋值于配准后的图像中对应的像素位置确定图像数据；

通过对所述图像数据进行图像处理以确定所述图像数据的前景像素点；

根据所述前景像素点对所述图像数据的连通区域进行标记得到所述组织掩模图。

40 进一步地，在本实施例一种可能的实现方式中，如图 48 所示，所述装置还包括：

计算单元 27，用于在所述第二分割单元对所述配准后的图像进行组织分割，得到组织掩模图之后，分别计算得到组织掩模图对应的灰度图像的外轮廓图像与内轮廓图像；

45 配对单元 28，用于将所述外轮廓图像与所述内轮廓图像分别分为 N 段，并将每段所述外轮廓图像与每段所述内轮廓图像两两配对，得到 N 个图像组合，每个所述图像组合包括一段外轮廓图像与一段内轮廓图像；

所述计算单元 27，还用于分别计算每个所述图像组合中一段外轮廓图像与一段内轮廓图像的像素密度比值，得到 N 个所述像素密度比值，所述像素密度比值为每个所述图像组合中一段外轮廓图像与一段内轮廓图像的像素占比的比值，所述像素占比为

50 像素点与总像素点的比值；

所述计算单元，还用于根据所述像素密度比值计算描述所述组织掩模图对应的灰

度图像质量。

进一步地，在本实施例一种可能的实现方式中，如图 48 所示，所述第一分割单元 23，还用于：

5 将所述配准后的图像输入所述图像分割模型，获得所述配准后的图像中各个像素点的分割结果，所述分割结果包括属于细胞、属于细胞轮廓或者图像背景；

将所述分割结果属于细胞或细胞轮廓的像素点确定属于细胞的像素点，输出所述配准后的图像的分割结果，得到所述细胞掩模图。

进一步地，在本实施例一种可能的实现方式中，如图 48 所示，所述第一分割单元 23，还用于：

10 对所述基因表达可视化图像进行预处理，得到预处理图；

将所述预处理图进行二值化处理，得到初始掩模图；

根据所述初始掩模图，利用基于距离变换的分水岭算法对存在细胞粘连的连通域进行分割，得到分割后的所述细胞掩模图。

15 进一步地，在本实施例一种可能的实现方式中，如图 48 所示，所述第一分割单元 23，还用于：

基于所述染色图像确定所述目标对象的各个组织类型对应的标注区域和细胞面积；

20 依据所述细胞面积确定每个所述标注区域包含的待处理子区域的面积，所述待处理子区域的面积与所述细胞面积正相关，所述待处理子区域用于表征对应的所述标注区域的最小处理单元；

依据每个所述待处理子区域的面积，将对应的所述标注区域进行网格化处理，得到每个所述标注区域对应的网格化坐标；

从所述基因表达可视化图像中提取所述网格化坐标对应的数据，并叠加至所述网格化坐标，得到所述目标对象的细胞处理结果。

25 进一步地，在本实施例一种可能的实现方式中，如图 48 所示，所述修正单元 24，还用于：

基于所述细胞掩模图，将所述基因表达可视化图像的每个基因表达特征分别确定为位于细胞内的细胞基因表达特征及位于细胞外的背景基因表达特征；

30 对每个所述细胞基因表达特征，通过空间位置和基因表达特征的信号值拟合出概率分布模型；

基于所述概率分布模型，确定出针对位于一目标细胞的预设范围内的目标背景基因表达特征属于所述目标细胞的概率值；

根据所述概率值输出校正结果以进行背景基因表达特征的校正。

35 进一步地，在本实施例一种可能的实现方式中，如图 48 所示，所述修正单元 24，还用于包括：

根据所述基因表达可视化图像，确定所属于目标细胞的初始基因分子以及未知所属细胞的背景基因分子；

获取所述初始基因分子的坐标信息；

根据所述坐标信息，确定所述目标细胞的中心点坐标；

40 根据所述中心点坐标和所述坐标信息，构建维诺图；

确定所述维诺图范围内的背景基因分子的所属细胞。

进一步地，在本实施例一种可能的实现方式中，如图 48 所示，所述修正单元 24，还用于包括：

45 确定所述基因表达可视化图像中所属于目标细胞的基因分子的初始分子数量以及所述所属于目标细胞的基因分子所占的目标区域的初始轮廓面积；

根据所述分子数量和所述初始轮廓面积，确定所述目标细胞的基因分子密度；

确定所述目标区域的预设区域范围内的背景基因分子的分子数量以及所述预设区域范围的区域面积；

根据所述背景基因分子的分子数量和区域面积，确定背景基因分子密度；

50 根据所述基因分子密度与所述背景基因分子密度确定所述背景基因分子的所属细胞。

进一步地，在本实施例一种可能的实现方式中，如图 48 所示，所述修正单元 24，还用于包括：

确定所述基因表达可视化图像中所属于目标细胞的基因分子所占的目标区域；

确定所述目标区域的预设校正范围内的未知所属细胞的背景基因分子的空间坐标及基因表达特征信号值；

将所述空间坐标和所述基因表达特征信号值输入概率分布模型，以得到所述背景基因分子所属于所述目标细胞的概率分数；所述概率分布模型根据所属于目标细胞的基因分子样本的空间坐标及基因表达特征信号值拟合得到；

根据所述概率分数确定所述背景基因分子的所属细胞。

进一步地，在本实施例一种可能的实现方式中，如图 48 所示，所述标记单元 25，还用于：

以 ssDNA 染液和细胞边界荧光染色液对所属细胞进行染色，其中，所述 ssDNA 染液包括缓冲液和 ssDNA 试剂，所述细胞边界荧光染色液包括荧光染料。

进一步地，在本实施例一种可能的实现方式中，如图 48 所示，所述装置还包括检测单元 29，包括：

获取待测染色图像或基因表达可视化图像；以及

将所述待测染色图像或基因表达可视化图像输入有向目标检测网络，以获得目标检测结果，

其中所述有向目标检测网络包括：

主干 Backbone 层，被配置为对待测染色图像或基因表达可视化图像执行特征提取，以获得不同尺度的多个特征图；

颈部 Neck 层，被配置为从 Backbone 层接收所述多个特征图，并对所述多个特征图进行跨尺度融合，以获得多个融合特征图；以及

检测头 Head 层，被配置为对所述多个融合图像进行预测，以获得关于所述待测染色图像或基因表达可视化图像的目标检测结果，

其中所述目标检测结果包括：分类信息、目标框坐标预测信息、目标预测信息和角度预测信息。

可选的，还包括确定单元 210，确定单元 210 用于：

所述染色图像所在的染色图像平面包括第一方向和与所述第一方向垂直的第二方向，所述染色图像包括横向于所述第二方向的多个第一直轨迹线和横向于所述第一方向的多个第二直轨迹线，

采样步骤，包括：

应用第一采样条沿所述第一方向对所述染色图像进行第一方向等间隔采样以确定第一候选点集合，其中对于每次第一方向等间隔采样，执行以下子步骤：

将所述第一采样条在该次第一方向等间隔采样期间所选中的所述染色图像的采样区域内的像素按所述第一方向进行像素值累加求和以得到一系列第一像素值和；以及

确定所述一系列第一像素值和的极值点，并且对于确定的每个极值点：确定该次第一方向等间隔采样所获得的一个对应第一候选点，该对应第一候选点在所述第二方向上的坐标为该极值点所对应的第二方向位置在所述第二方向上的坐标，该对应第一候选点在所述第一方向上的坐标基于在该次第一方向等间隔采样期间所述第一采样条在所述第一方向上的位置确定，

其中，由每次第一方向等间隔采样所获得的第一候选点形成所述第一候选点集合；

应用第二采样条沿所述第二方向对所述染色图像进行第二方向等间隔采样以确定第二候选点集合，其中对于每次第二方向等间隔采样，执行以下子步骤：

将所述第二采样条在该次第二方向等间隔采样期间所选中的所述染色图像的采样区域内的像素按所述第二方向进行像素值累加求和以得到一系列第二像素值和；以及

确定所述一系列第二像素值和的极值点，并且对于确定的每个极值点：确定该次第二方向等间隔采样所获得的一个对应第二候选点，该对应第二候选点在所述第一方

向上的坐标为该极值点所对应的第一方向位置在所述第一方向上的坐标，该对应第二候选点在所述第二方向上的坐标基于在该次第二方向等间隔采样期间所述第二采样条在所述第二方向上的位置确定，

其中，由每次第二方向等间隔采样所获得的第二候选点形成所述第二候选点集合；

分组步骤，包括：

将所述第一候选点集合中的第一候选点分成多个第一候选点组；以及

将所述第二候选点集合中的第二候选点分成多个第二候选点组；

直线拟合步骤，包括：

对每个第一候选点组中的第一候选点进行直线拟合以得到一个相应的第一直线解析式；以及

对每个第二候选点组中的第二候选点进行直线拟合以得到一个相应的第二直线解析式；以及

坐标计算步骤，包括：

计算每个第一直线解析式与每个第二直线解析式的交点的坐标。

进一步地，在本实施例一种可能的实现方式中，如图 48 所示，所述装置还包括：分析单元 211，所述分析单元用于：

获取生物样本的第一图像，所述第一图像包括所述生物样本经染色的细胞边界；

获取所述生物样本的第二图像，所述第二图像包括所述生物样本经染色的遗传物质；

获取所述生物样本的基因表达量信息；

将所述第二图像与所述第一图像配准，得到第二图像与第一图像的配准图像；

将所述第二图像与所述基因表达量信息进行配准，得到第二图像与基因表达量信息的配准图像；

根据第二图像与第一图像的配准图像、第二图像与基因表达量信息的配准图像获取第一图像与基因表达量信息的配准图像；

根据第一图像与基因表达量信息的配准图像获取所述生物样本的细胞边界；

基于第一图像与基因表达量信息的配准图像根据所述细胞边界获取所述生物样本的细胞的基因表达信息。

根据本申请的实施例，本申请还提供了一种电子设备、一种可读存储介质和一种计算机程序产品。

图 49 示出了可以用来实施本申请的实施例的示例电子设备 300 的示意性框图。电子设备旨在表示各种形式的数字计算机，诸如，膝上型计算机、台式计算机、工作台、个人数字助理、服务器、刀片式服务器、大型计算机、和其它适合的计算机。电子设备还可以表示各种形式的移动装置，诸如，个人数字处理、蜂窝电话、智能电话、可穿戴设备和其它类似的计算装置。本文所示的部件、它们的连接和关系、以及它们的功能仅作为示例，并且不意在限制本文中描述的和/或者要求的本申请的实现。

如图 49 所示，设备 300 包括计算单元 301，其可以根据存储在 ROM (Read-Only Memory, 只读存储器) 302 中的计算机程序或者从存储单元 308 加载到 RAM (Random Access Memory, 随机访问/存取存储器) 303 中的计算机程序，来执行各种适当的动作和处理。在 RAM 303 中，还可存储设备 300 操作所需的各种程序和数据。计算单元 301、ROM 302 以及 RAM 303 通过总线 304 彼此相连。I/O (Input/Output, 输入/输出) 接口 305 也连接至总线 304。

设备 300 中的多个部件连接至 I/O 接口 305，包括：输入单元 306，例如键盘、鼠标等；输出单元 307，例如各种类型的显示器、扬声器等；存储单元 308，例如磁盘、光盘等；以及通信单元 309，例如网卡、调制解调器、无线通信收发机等。通信单元 309 允许设备 300 通过诸如因特网的计算机网络和/或各种电信网络与其他设备交换信息/数据。

计算单元 301 可以是各种具有处理和计算能力的通用和/或专用处理组件。计算单元 301 的一些示例包括但不限于 CPU (Central Processing Unit, 中央处理单元)、GPU (Graphic Processing Units, 图形处理单元)、各种专用的 AI (Artificial

Intelligence, 人工智能) 计算芯片、各种运行机器学习模型算法的计算单元、DSP (Digital Signal Processor, 数字信号处理器)、以及任何适当的处理器、控制器、微控制器等。计算单元 301 执行上文所描述的各个方法和处理, 例如空间组学单细胞数据获取方法。例如, 在一些实施例中, 空间组学单细胞数据获取方法可被实现为计算机程序, 其被有形地包含于机器可读介质, 例如存储单元 308。在一些实施例中, 计算机程序的部分或者全部可以经由 ROM 302 和/或通信单元 309 而被载入和/或安装到设备 300 上。当计算机程序加载到 RAM 303 并由计算单元 301 执行时, 可以执行上文描述的方法的一个或多个步骤。备选地, 在其他实施例中, 计算单元 301 可以通过其他任何适当的方式 (例如, 借助于固件) 而被配置为执行前述空间组学单细胞数据获取方法。

本文中以上描述的系统和技术各种实施方式可以在数字电子电路系统、集成电路系统、FPGA (Field Programmable Gate Array, 现场可编程门阵列)、ASIC (Application-Specific Integrated Circuit, 专用集成电路)、ASSP (Application Specific Standard Product, 专用标准产品)、SOC (System On Chip, 芯片上系统的系统)、CPLD (Complex Programmable Logic Device, 复杂可编程逻辑设备)、计算机硬件、固件、软件、和/或它们的组合中实现。这些各种实施方式可以包括: 实施在一个或者多个计算机程序中, 该一个或者多个计算机程序可在包括至少一个可编程处理器的可编程系统上执行和/或解释, 该可编程处理器可以是专用或者通用可编程处理器, 可以从存储系统、至少一个输入装置、和至少一个输出装置接收数据和指令, 并且将数据和指令传输至该存储系统、该至少一个输入装置、和该至少一个输出装置。

用于实施本申请的方法的程序代码可以采用一个或多个编程语言的任何组合来编写。这些程序代码可以提供给通用计算机、专用计算机或其他可编程数据处理装置的处理器或控制器, 使得程序代码当由处理器或控制器执行时使流程图和/或框图中所规定的功能/操作被实施。程序代码可以完全在机器上执行、部分地在机器上执行, 作为独立软件包部分地在机器上执行且部分地在远程机器上执行或完全在远程机器或服务上执行。

在本申请的上下文中, 机器可读介质可以是有形的介质, 其可以包含或存储以供指令执行系统、装置或设备使用或与指令执行系统、装置或设备结合地使用的程序。机器可读介质可以是机器可读信号介质或机器可读储存介质。机器可读介质可以包括但不限于电子的、磁性的、光学的、电磁的、红外的、或半导体系统、装置或设备, 或者上述内容的任何合适组合。机器可读存储介质的更具体示例会包括基于一个或多个线的电气连接、便携式计算机盘、硬盘、RAM、ROM、EPROM (Electrically Programmable Read-Only Memory, 可擦除可编程只读存储器) 或快闪存储器、光纤、CD-ROM (Compact Disc Read-Only Memory, 便捷式紧凑盘只读存储器)、光学储存设备、磁储存设备、或上述内容的任何合适组合。

为了提供与用户的交互, 可以在计算机上实施此处描述的系统和技术, 该计算机具有: 用于向用户显示信息的显示装置 (例如, CRT (Cathode-Ray Tube, 阴极射线管) 或者 LCD (Liquid Crystal Display, 液晶显示器) 监视器); 以及键盘和指向装置 (例如, 鼠标或者轨迹球), 用户可以通过该键盘和该指向装置来将输入提供给计算机。其它种类的装置还可以用于提供与用户的交互; 例如, 提供给用户的反馈可以是任何形式的传感反馈 (例如, 视觉反馈、听觉反馈、或者触觉反馈); 并且可以用任何形式 (包括声输入、语音输入或者、触觉输入) 来接收来自用户的输入。

可以将此处描述的系统和技术实施在包括后台部件的计算系统 (例如, 作为数据服务器)、或者包括中间件部件的计算系统 (例如, 应用服务器)、或者包括前端部件的计算系统 (例如, 具有图形用户界面或者网络浏览器的用户计算机, 用户可以通过该图形用户界面或者该网络浏览器来与此处描述的系统和技术实施方式交互)、或者包括这种后台部件、中间件部件、或者前端部件的任何组合的计算系统中。可以通过任何形式或者介质的数字数据通信 (例如, 通信网络) 来将系统的部件相互连接。通信网络的示例包括: LAN (Local Area Network, 局域网)、WAN (Wide Area Network, 广域网)、互联网和区块链网络。

计算机系统可以包括客户端和服务端。客户端和服务端一般远离彼此并且通常通

过通信网络进行交互。通过在相应的计算机上运行并且彼此具有客户端-服务器关系的计算机程序来产生客户端和服务器的关系。服务器可以是云服务器，又称为云计算服务器或云主机，是云计算服务体系中的一项主机产品，以解决了传统物理主机与 VPS 服务（“Virtual Private Server”，或简称“VPS”）中，存在的管理难度大，业务扩展性弱的缺陷。服务器也可以为分布式系统的服务器，或者是结合了区块链的服务器。

其中，需要说明的是，人工智能是研究使计算机来模拟人的某些思维过程和智能行为（如学习、推理、思考、规划等）的学科，既有硬件层面的技术也有软件层面的技术。人工智能硬件技术一般包括如传感器、专用人工智能芯片、云计算、分布式存储、大数据处理等技术；人工智能软件技术主要包括计算机视觉技术、语音识别技术、自然语言处理技术以及机器学习/深度学习、大数据处理技术、知识图谱技术等几大方向。

应该理解，可以使用上面所示的各种形式的流程，重新排序、增加或删除步骤。例如，本申请中记载的各步骤可以并行地执行也可以顺序地执行也可以不同的次序执行，只要能够实现本申请公开的技术方案所期望的结果，本文在此不进行限制。

上述具体实施方式，并不构成对本申请保护范围的限制。本领域技术人员应该明白的是，根据设计要求和因素，可以进行各种修改、组合、子组合和替代。任何在本申请的精神和原则之内所作的修改、等同替换和改进等，均应包含在本申请保护范围之内。

权 利 要 求 书

1. 一种空间组学单细胞数据获取方法，其特征在于，包括：

获取生物样本的染色图像，获取所述生物样本的基因表达可视化图像；

5 将所述染色图像与所述基因表达可视化图像进行配准，得到配准后的图像；

对所述配准后的图像进行细胞分割，得到细胞掩模图；

10 根据所述细胞掩模图确定所述基因表达可视化图像中未知所属细胞的背景基因分子的所属细胞，并对所属细胞标记，得到带细胞标记的空间组分子表达矩阵。

2. 根据权利要求 1 所述的方法，其特征在于，所述获取生物样本的染色图像包括：

通过显微镜获取所述生物样本的一个染色图像；或

15 通过显微镜获取所述生物样本的由多个子染色图像拼接的拼接染色图像；或

通过所述显微镜获取所述生物样本的多个子染色图像，对所述多个子染色图像进行拼接，得到拼接染色图像。

3. 根据权利要求 2 所述的方法，其特征在于，所述获取生物样本的染色图像包括：

20 通过显微镜获取所述生物样本的多个子染色图像；对所述生物样本的多个子染色图像进行第一拼接，得到第一染色图像；

根据所述第一染色图像的图像质量判断是否对所述生物样本的多个子染色图像进行第二拼接；

若否，将所述第一染色图像确定为所述染色图像；

25 若是，对所述生物样本的多个子染色图像进行第二拼接，得到所述染色图像。

4. 根据权利要求 3 所述的方法，其特征在于，所述通过所述显微镜获取所述生物样本的多个子染色图像，对所述多个子染色图像进行拼接，得到拼接染色图像包括：

30 分别确定多个子染色图像和模板图像的重叠区域；

将所述多个子染色图像和所述模板图像的被确定为重叠区域的部分分别裁剪为多个染色图像重叠子图和数量相同的多个模板图像重叠子图，以配对为多个重叠子图对；

35 通过对所述多个重叠子图对进行频域计算，分别获得各子染色图像和所述模板图像之间的偏移量；以及

将所述各子染色图像分别朝向所述模板图像的方向平移所述偏移量，并

与所述模板图像拼接在一起。

5. 根据权利要求 3 所述的方法，其特征在于，所述通过所述显微镜获取所述生物样本的多个子染色图像，对所述多个子染色图像进行拼接，得到拼接染色图像包括：

5 获取所述多个子染色图像以及关于所述生物样本的轨迹线或轨迹点的模板信息；

对所述多个子染色图像进行预拼接得到所述多个子染色图像的预拼接坐标；

10 根据轨迹线或轨迹点的特征从所述多个子染色图像中选取至少一个第一参考图像，根据所述生物样本的模板信息和所述第一参考图像的预拼接坐标推导全局模板；以及

15 对于所述多个子染色图像中除了所述第一参考图像之外的第二参考图像，计算所述第二参考图像的预拼接坐标与所述第二参考图像在所述全局模板中的相应模板坐标之间的偏移量，并基于所述偏移量对所述第二参考图像进行坐标调整以拼接生成所述拼接染色图像。

6. 根据权利要求 1 所述的方法，其特征在于，所述将所述染色图像与所述基因表达可视化图像进行配准，得到配准后的图像包括：

20 获取所述染色图像中的轨迹线，基于所述轨迹线与水平方向的夹角旋转所述染色图像；其中，所述染色图像和所述基因表达可视化图像包括多条轨迹线和/或多个轨迹点，所述轨迹点是由两条所述轨迹线相交形成的，所述多条轨迹线中的每条轨迹线具有对应的索引号；

将所述染色图像进行放大或缩小得到第三参考图像，所述第三参考图像与所述基因表达可视化图像具有相同的比例；

基于所述基因表达可视化图像和所述第三参考图像确定目标偏移量；

25 基于所述目标偏移量移动所述第三参考图像，得到所述配准后的图像。

7. 根据权利要求 6 所述的方法，其特征在于，在将所述染色图像进行放大或缩小得到第三参考图像之后，所述方法还包括：

确定所述第三参考图像与基因表达可视化图像之间的相似度；

将相似度最大值对应的位置作为粗配准结果。

30 8. 根据权利要求 1 所述的方法，其特征在于，将所述染色图像与所述基因表达可视化图像进行配准，得到配准后的图像包括：

对所述染色图像进行预处理，获取候选图像集合；

确定所述候选图像集合对应的二值图集合；

35 基于所述二值图集合与所述基因表达可视化图像之间的相似度，确定所述二值图集合与所述基因表达可视化图像中具有对应关系的轨迹点对；

基于所述轨迹点对确定偏移信息；

基于所述偏移信息处理目标图像，以得到配准后的图像，所述目标图像为所述轨迹点对所对应的候选图像集合中的图像。

9. 根据权利要求 1 所述的方法，其特征在于，在对所述配准后的图像进行细胞分割，得到细胞掩模图之前，所述方法还包括：

5 对所述配准后的图像进行组织分割，得到组织掩模图。

10. 根据权利要求 9 所述的方法，其特征在于，所述对所述配准后的图像进行组织分割，得到组织掩模图包括：

10 获取所述配准后的图像中待分割的目标组织的基因表达量并确定所述配准后的图像的基因表达矩阵，所述基因表达矩阵按照像素位置对所述基因表达量进行排列；

通过将所述基因表达矩阵中每一像素位置的基因表达量赋值于配准后的图像中对应的像素位置确定图像数据；

通过对所述图像数据进行图像处理以确定所述图像数据的前景像素点；

15 根据所述前景像素点对所述图像数据的连通区域进行标记得到所述组织掩模图。

11. 根据权利要求 10 所述的方法，其特征在于，在对所述配准后的图像进行组织分割，得到组织掩模图之后，所述方法还包括：

分别计算得到组织掩模图对应的灰度图像的外轮廓图像与内轮廓图像；

20 将所述外轮廓图像与所述内轮廓图像分别分为 N 段，并将每段所述外轮廓图像与每段所述内轮廓图像两两配对，得到 N 个图像组合，每个所述图像组合包括一段外轮廓图像与一段内轮廓图像；

25 分别计算每个所述图像组合中一段外轮廓图像与一段内轮廓图像的像素密度比值，得到 N 个所述像素密度比值，所述像素密度比值为每个所述图像组合中一段外轮廓图像与一段内轮廓图像的像素占比的比值，所述像素占比为像素点与总像素点的比值；

根据所述像素密度比值计算描述所述组织掩模图对应的灰度图像质量。

12. 根据权利要求 1 所述的方法，其特征在于，所述对所述配准后的图像进行细胞分割，得到细胞掩模图包括：

30 将所述配准后的图像输入所述图像分割模型，获得所述配准后的图像中各个像素点的分割结果，所述分割结果包括属于细胞的像素点；

将所述分割结果属于细胞的像素点或细胞轮廓的像素点确定属于细胞的像素点，输出所述配准后的图像的分割结果，得到所述细胞掩模图。

13. 根据权利要求 1 所述的方法，其特征在于，所述对所述配准后的图像进行细胞分割，得到细胞掩模图包括：

35 对所述基因表达可视化图像进行预处理，得到预处理图；

将所述预处理图进行二值化处理，得到初始掩模图；

根据所述初始掩模图，利用基于距离变换的分水岭算法对存在细胞粘连的连通域进行分割，得到分割后的所述细胞掩模图。

14. 根据权利要求 1 所述的方法，其特征在于，所述对所述配准后的图像进行细胞分割，得到细胞掩模图包括：

5 基于所述染色图像确定所述目标对象的各个组织类型对应的标注区域和细胞面积；

依据所述细胞面积确定每个所述标注区域包含的待处理子区域的面积，所述待处理子区域的面积与所述细胞面积正相关，所述待处理子区域用于表征对应的所述标注区域的最小处理单元；

10 依据每个所述待处理子区域的面积，将对应的所述标注区域进行网格化处理，得到每个所述标注区域对应的网格化坐标；

从所述基因表达可视化图像中提取所述网格化坐标对应的数据，并叠加至所述网格化坐标，得到所述目标对象的细胞处理结果。

15 15. 根据权利要求 1 所述的方法，其特征在于，所述根据所述细胞掩模图确定所述基因表达可视化图像中未知所属细胞的背景基因分子的所属细胞包括：

基于所述细胞掩模图，将所述基因表达可视化图像的每个基因表达特征分别确定为位于细胞内的细胞基因表达特征及位于细胞外的背景基因表达特征；

20 对每个所述细胞基因表达特征，通过空间位置和基因表达特征的信号值拟合出概率分布模型；

基于所述概率分布模型，确定出针对位于一目标细胞的预设范围内的目标背景基因表达特征属于所述目标细胞的概率值；

根据所述概率值输出校正结果以进行背景基因表达特征的校正。

25 16. 根据权利要求 1 所述的方法，其特征在于，所述根据所述细胞掩模图确定所述基因表达可视化图像中未知所属细胞的背景基因分子的所属细胞包括：

根据所述基因表达可视化图像，确定所属于目标细胞的初始基因分子以及未知所属细胞的背景基因分子；

30 获取所述初始基因分子的坐标信息；

根据所述坐标信息，确定所述目标细胞的中心点坐标；

根据所述中心点坐标和所述坐标信息，构建维诺图；

确定所述维诺图范围内的背景基因分子的所属细胞。

35 17. 根据权利要求 1 所述的方法，其特征在于，所述根据所述细胞掩模图确定所述基因表达可视化图像中未知所属细胞的背景基因分子的所属细胞包括：

确定所述基因表达可视化图像中所属于目标细胞的基因分子的初始分子数量以及所述所属于目标细胞的基因分子所占的目标区域的初始轮廓面积；

根据所述分子数量和所述初始轮廓面积，确定所述目标细胞的基因分子密度；

确定所述目标区域的预设区域范围内的背景基因分子的分子数量以及所述预设区域范围的区域面积；

根据所述背景基因分子的分子数量和区域面积，确定背景基因分子密度；

根据所述基因分子密度与所述背景基因分子密度确定所述背景基因分子的所属细胞。

18. 根据权利要求 1 所述的方法，其特征在于，所述根据所述细胞掩码图确定所述基因表达可视化图像中未知所属细胞的背景基因分子的所属细胞包括：

确定所述基因表达可视化图像中所属于目标细胞的基因分子所占的目标区域；

确定所述目标区域的预设校正范围内的未知所属细胞的背景基因分子的空间坐标及基因表达特征信号值；

将所述空间坐标和所述基因表达特征信号值输入概率分布模型，以得到所述背景基因分子所属于所述目标细胞的概率分数；所述概率分布模型根据所属于目标细胞的基因分子样本的空间坐标及基因表达特征信号值拟合得到；

根据所述概率分数确定所述背景基因分子的所属细胞。

19. 根据权利要求 1 所述的方法，其特征在于，所述对所属细胞标记包括：

以 ssDNA 染液和细胞边界荧光染色液对所属细胞进行染色，其中，所述 ssDNA 染液包括缓冲液和 ssDNA 试剂，所述细胞边界荧光染色液包括荧光染料。

20. 根据权利要求 5 或 6 所述的方法，其特征在于，所述方法还包括：

获取待测染色图像或基因表达可视化图像；以及

将所述待测染色图像或基因表达可视化图像输入有向目标检测网络，以获得目标检测结果，

其中所述有向目标检测网络包括：

主干 Backbone 层，被配置为对待测染色图像或基因表达可视化图像执行特征提取，以获得不同尺度的多个特征图；

颈部 Neck 层，被配置为从 Backbone 层接收所述多个特征图，并对所述

多个特征图进行跨尺度融合，以获得多个融合特征图；以及

检测头 Head 层，被配置为对所述多个融合图像进行预测，以获得关于所述待测染色图像或基因表达可视化图像的目标检测结果，

其中所述目标检测结果包括：目标框坐标预测信息、目标预测信息和角度预测信息。

21. 根据权利要求 5 或 6 所述的方法，其特征在于，所述染色图像所在的染色图像平面包括第一方向和与所述第一方向垂直的第二方向，所述染色图像包括横向于所述第二方向的多个第一直轨迹线和横向于所述第一方向的多个第二直轨迹线，

所述染色图像中的点坐标的确定方法包括：

采样步骤，包括：

应用第一采样条沿所述第一方向对所述染色图像进行第一方向等间隔采样以确定第一候选点集合，其中对于每次第一方向等间隔采样，执行以下子步骤：

将所述第一采样条在该次第一方向等间隔采样期间所选中的所述染色图像的采样区域内的像素按所述第一方向进行像素值累加求和以得到一系列第一像素值和；以及

确定所述一系列第一像素值和的极值点，并且对于确定的每个极值点：确定该次第一方向等间隔采样所获得的一个对应第一候选点，该对应第一候选点在所述第二方向上的坐标为该极值点所对应的第二方向位置在所述第二方向上的坐标，该对应第一候选点在所述第一方向上的坐标基于在该次第一方向等间隔采样期间所述第一采样条在所述第一方向上的位置确定，

其中，由每次第一方向等间隔采样所获得的第一候选点形成所述第一候选点集合；

应用第二采样条沿所述第二方向对所述染色图像进行第二方向等间隔采样以确定第二候选点集合，其中对于每次第二方向等间隔采样，执行以下子步骤：

将所述第二采样条在该次第二方向等间隔采样期间所选中的所述染色图像的采样区域内的像素按所述第二方向进行像素值累加求和以得到一系列第二像素值和；以及

确定所述一系列第二像素值和的极值点，并且对于确定的每个极值点：确定该次第二方向等间隔采样所获得的一个对应第二候选点，该对应第二候选点在所述第一方向上的坐标为该极值点所对应的第一方向位置在所述第一方向上的坐标，该对应第二候选点在所述第二方向上的坐标基于在该次第二方向等间隔采样期间所述第二采样条在所述第二方向上的位置确定，

其中，由每次第二方向等间隔采样所获得的第二候选点形成所述第二候

选点集合；

分组步骤，包括：

将所述第一候选点集合中的第一候选点分成多个第一候选点组；以及

将所述第二候选点集合中的第二候选点分成多个第二候选点组；

5 直线拟合步骤，包括：

对每个第一候选点组中的第一候选点进行直线拟合以得到一个相应的第一直线解析式；以及

对每个第二候选点组中的第二候选点进行直线拟合以得到一个相应的第二直线解析式；以及

10 坐标计算步骤，包括：

计算每个第一直线解析式与每个第二直线解析式的交点的坐标。

22. 根据权利要求 19 所述的方法，其特征在于，在对所属细胞标记，得到带细胞标记的空间组分子表达矩阵之后，所述方法还包括：

15 获取生物样本的第一图像，所述第一图像包括所述生物样本经染色的细胞边界；

获取所述生物样本的第二图像，所述第二图像包括所述生物样本经染色的遗传物质；

获取所述生物样本的基因表达量信息；

20 将所述第二图像与所述第一图像配准，得到第二图像与第一图像的配准图像；

将所述第二图像与所述基因表达量信息进行配准，得到第二图像与第基因表达量信息的配准图像；

根据第二图像与第一图像的配准图像、第二图像与基因表达量信息的配准图像获取第一图像与基因表达量信息的配准图像；

25 根据第一图像与基因表达量信息的配准图像获取所述生物样本的细胞边界；

基于第一图像与基因表达量信息的配准图像根据所述细胞边界获取所述生物样本的细胞的基因表达信息。

23. 一种空间组学单细胞数据获取装置，其特征在于，包括：

30 获取单元，用于获取生物样本的染色图像，获取所述生物样本的基因表达可视化图像；

配准单元，用于将所述染色图像与所述基因表达可视化图像进行配准，得到配准后的图像；

第一分割单元，用于对所述配准后的图像进行细胞分割，得到细胞掩模图；

35 修正单元，用于根据所述细胞掩模图确定所述基因表达可视化图像中未知所属细胞的背景基因分子的所属细胞；

标记单元，用于对所属细胞标记，得到带细胞标记的空间组分子表达矩阵。

24. 一种电子设备，其特征在于，包括：

至少一个处理器；以及

与所述至少一个处理器通信连接的存储器；其中，

5 所述存储器存储有可被所述至少一个处理器执行的指令，所述指令被所述至少一个处理器执行，以使所述至少一个处理器能够执行权利要求 1-22 中任一项所述的方法。

25. 一种存储有计算机指令的非瞬时计算机可读存储介质，其特征在于，所述计算机指令用于使所述计算机执行根据权利要求 1-22 中任一项所述的方法。

10 26. 一种计算机程序产品，其特征在于，包括计算机程序，所述计算机程序在被处理器执行时实现根据权利要求 1-22 中任一项所述的方法。

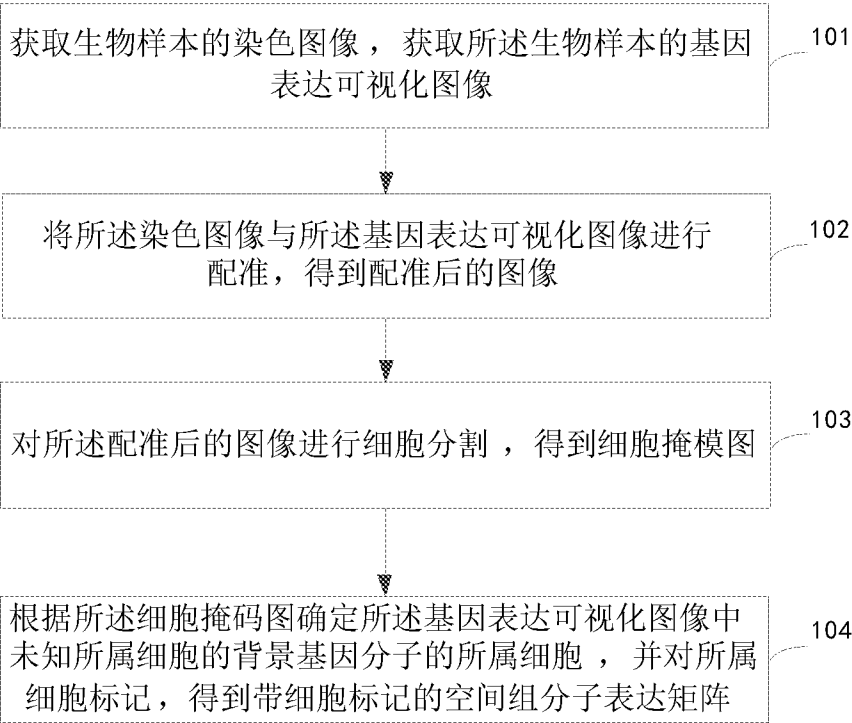


图 1

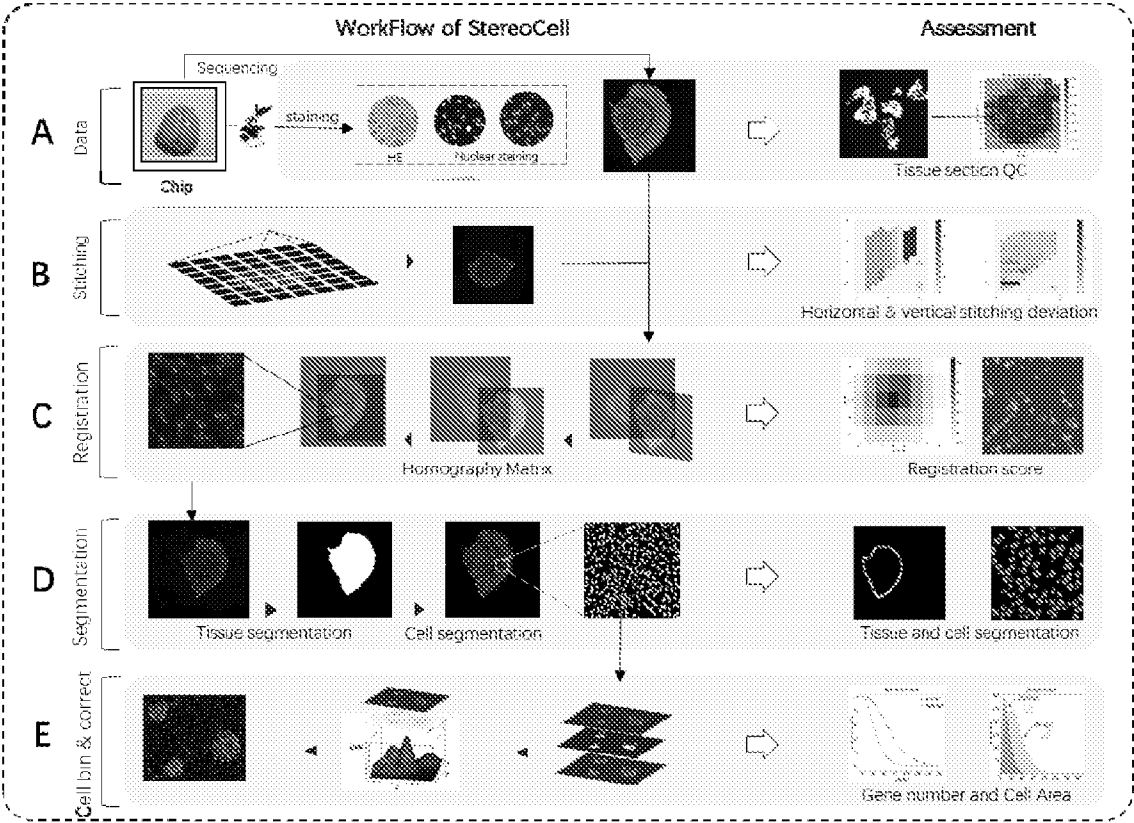


图 2

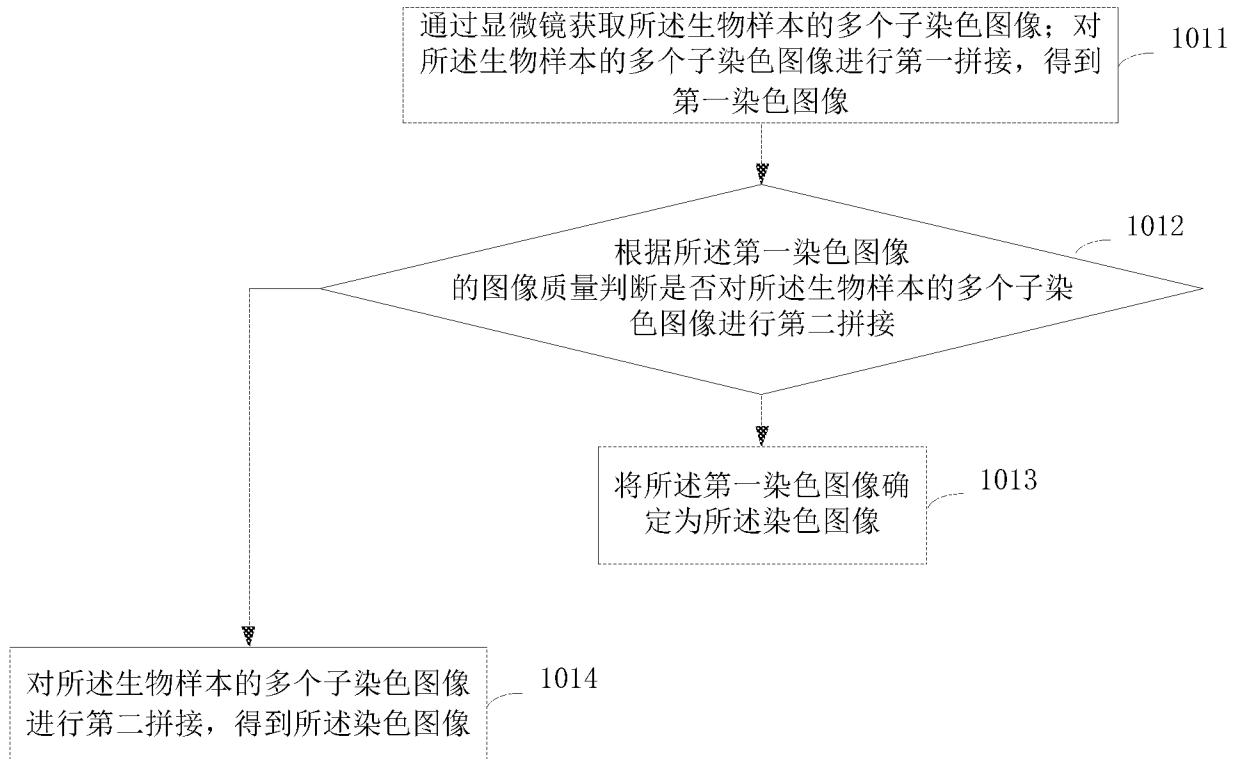


图 3

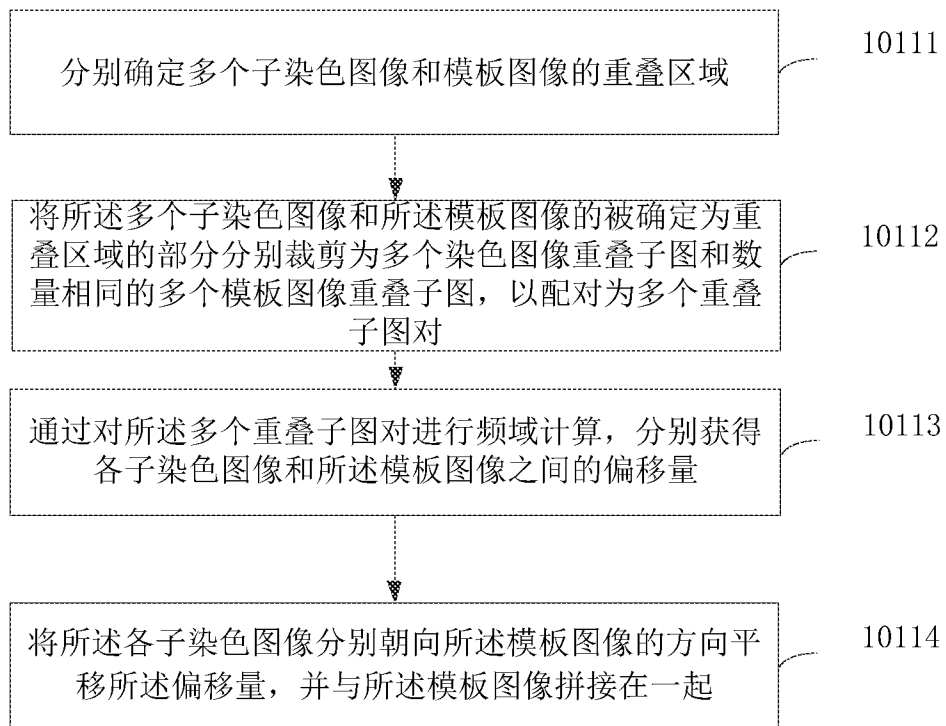


图 4

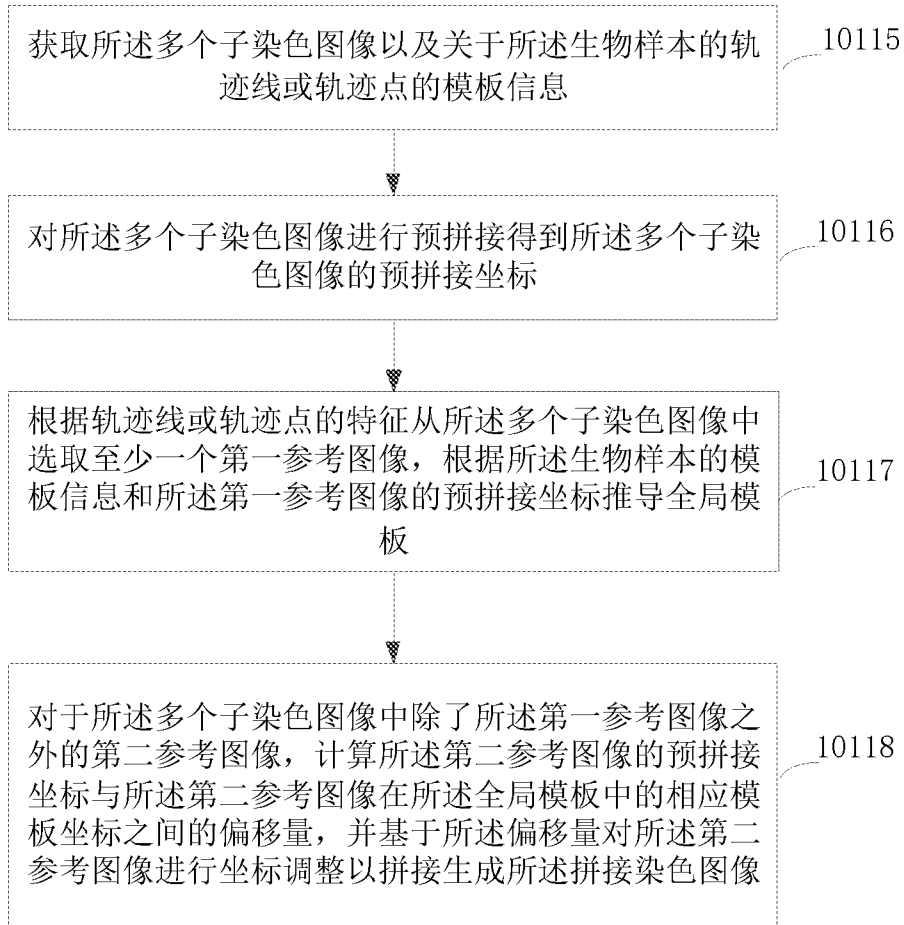


图 5

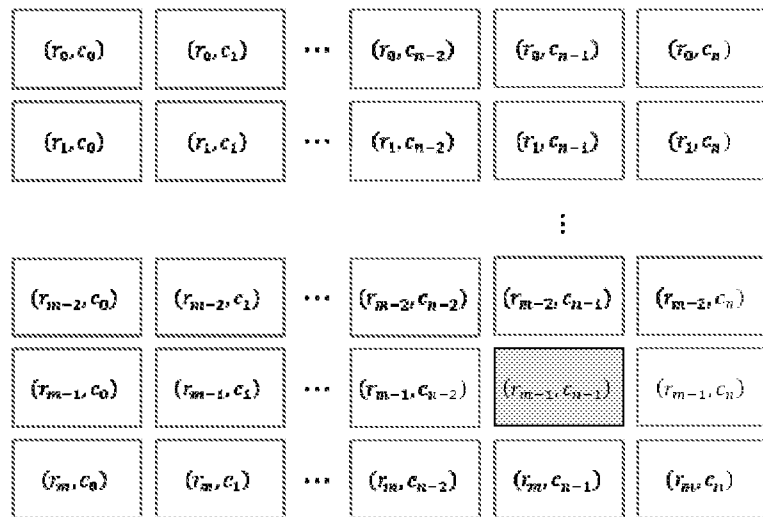


图 6

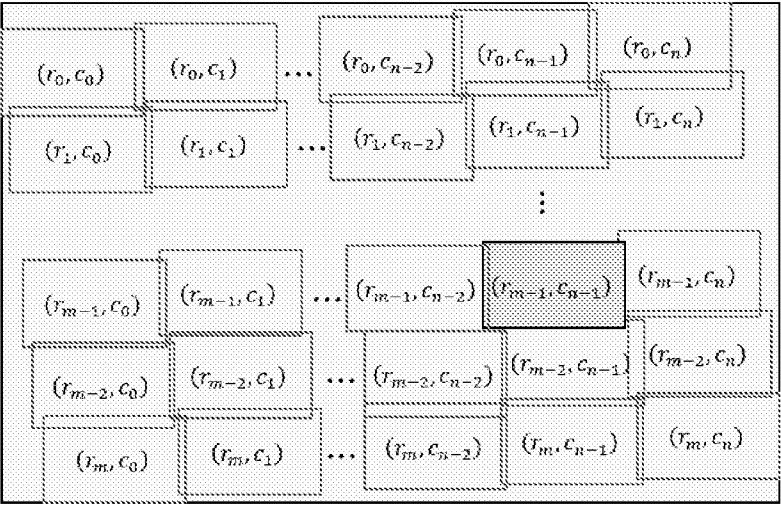


图 7

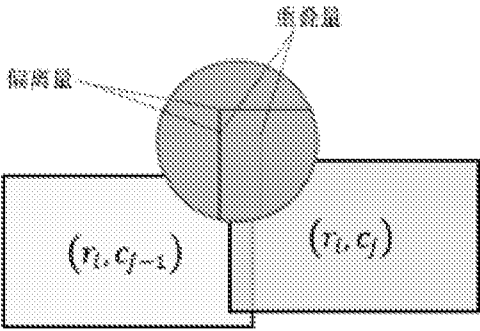


图 8a

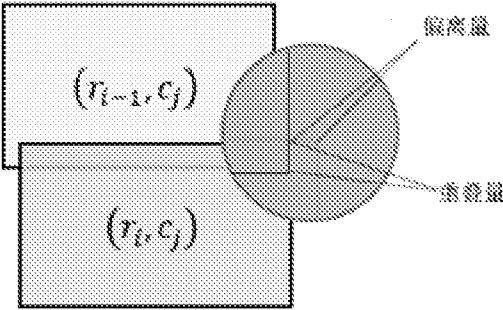


图 8b

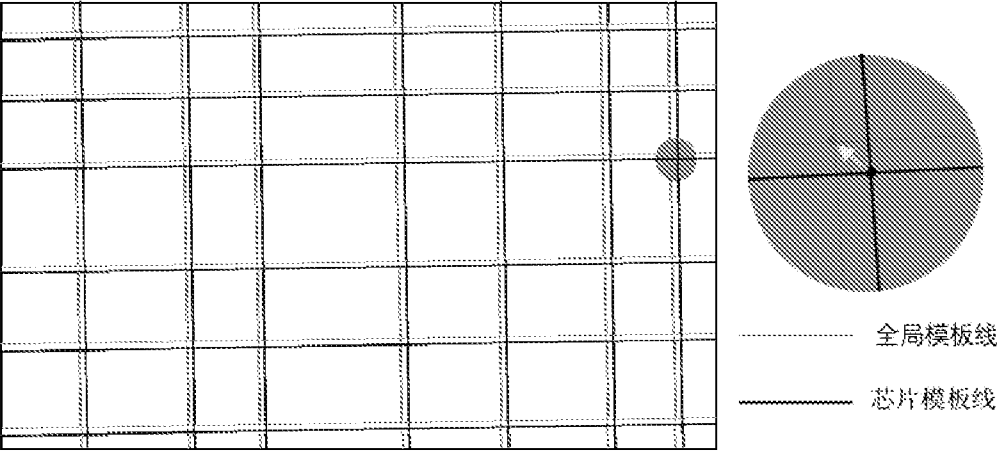


图 9

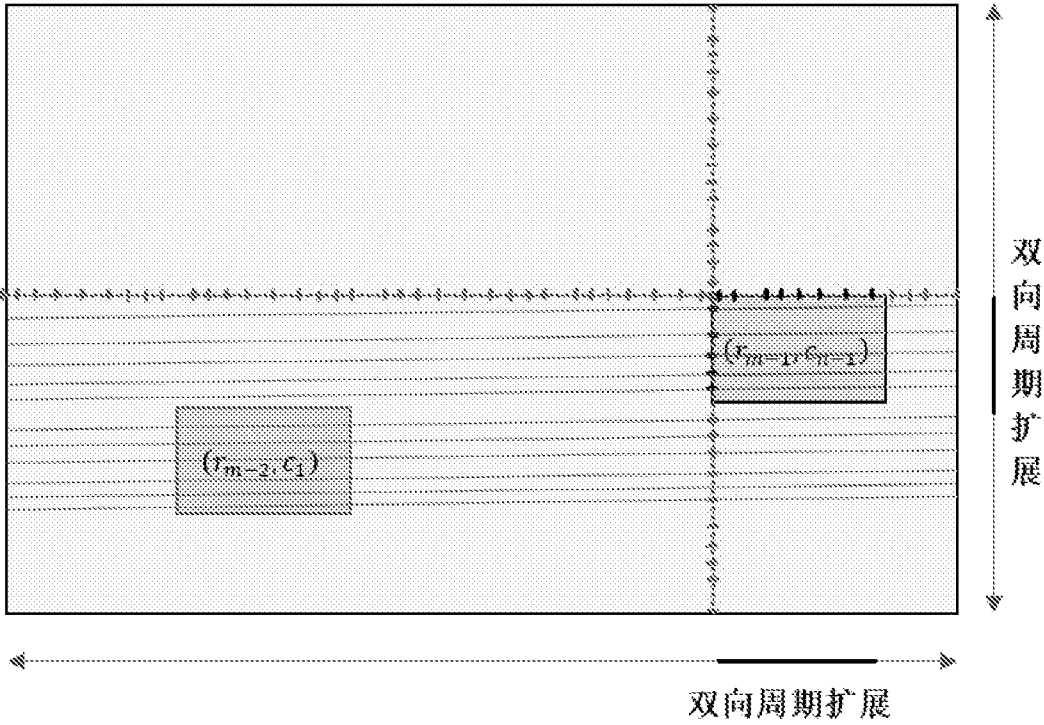


图 10

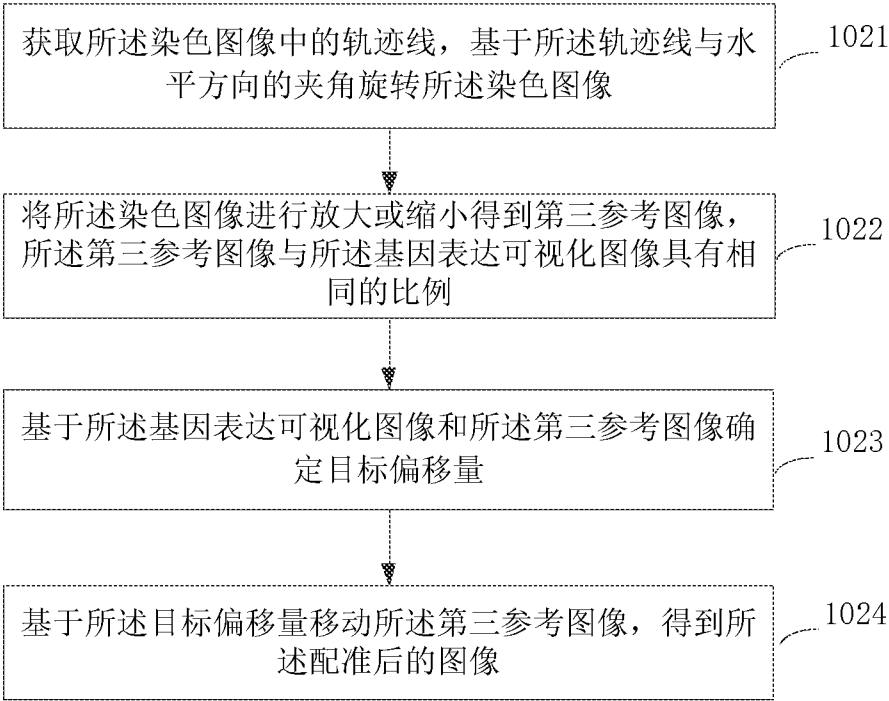


图 11

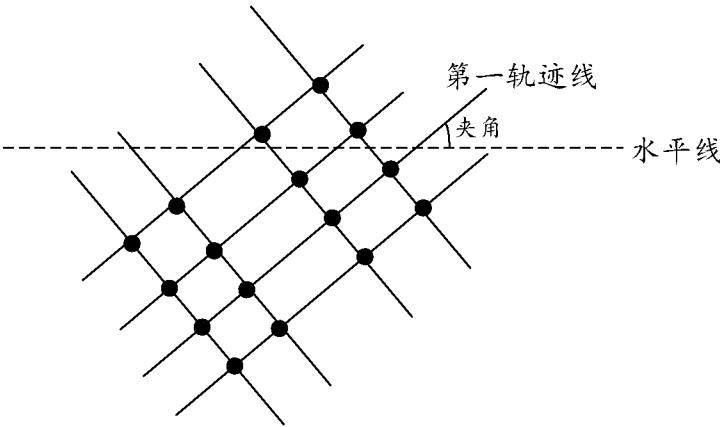


图 12

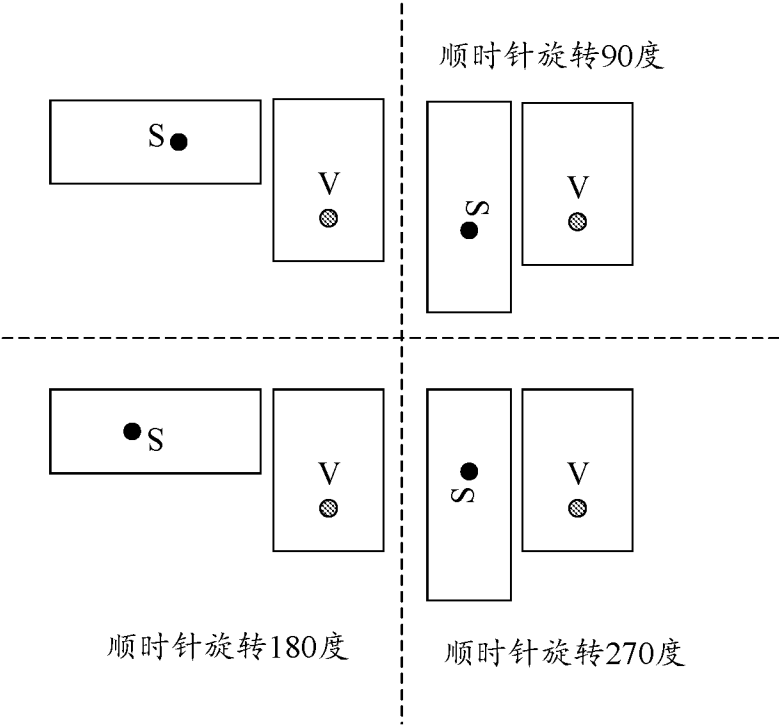


图 13

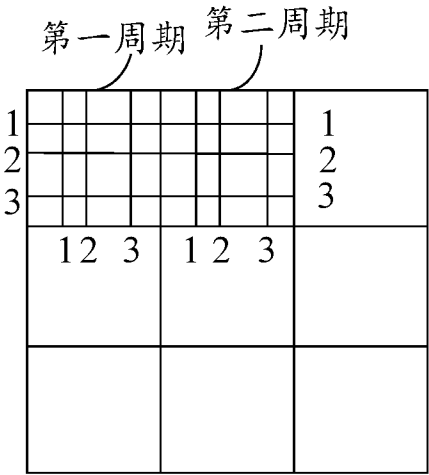


图 14a

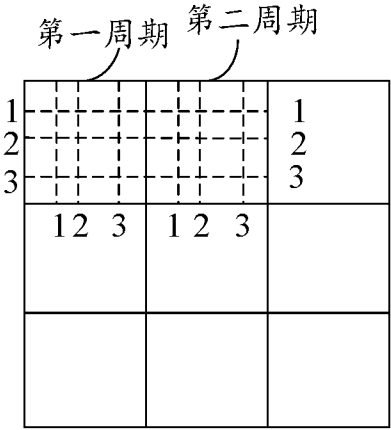


图 14b

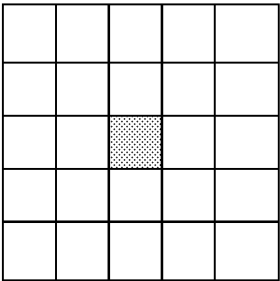


图 15

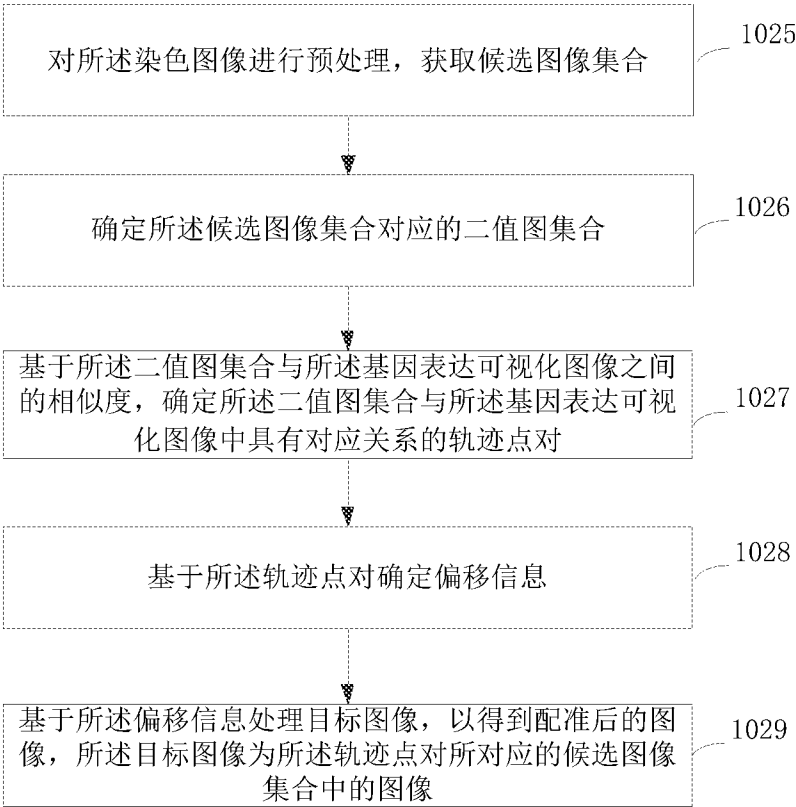


图 16

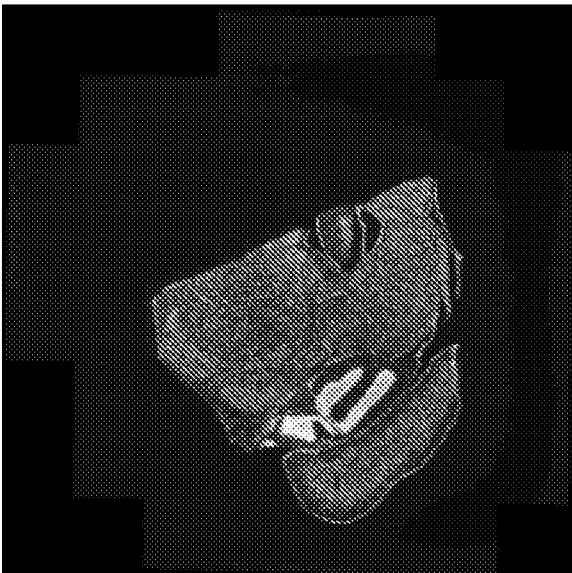


图 17a

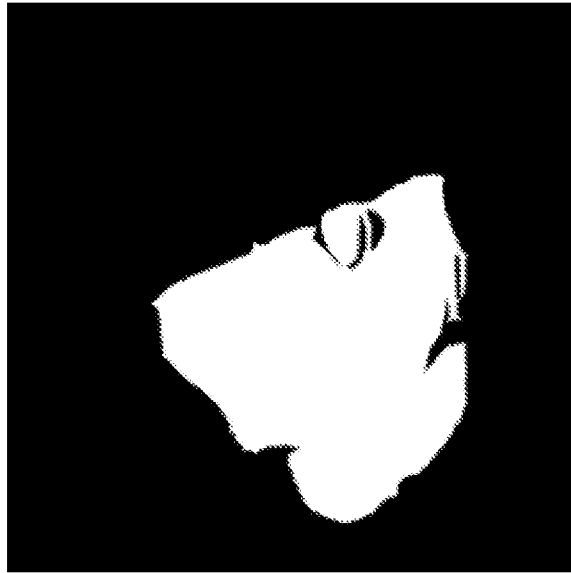


图 17b

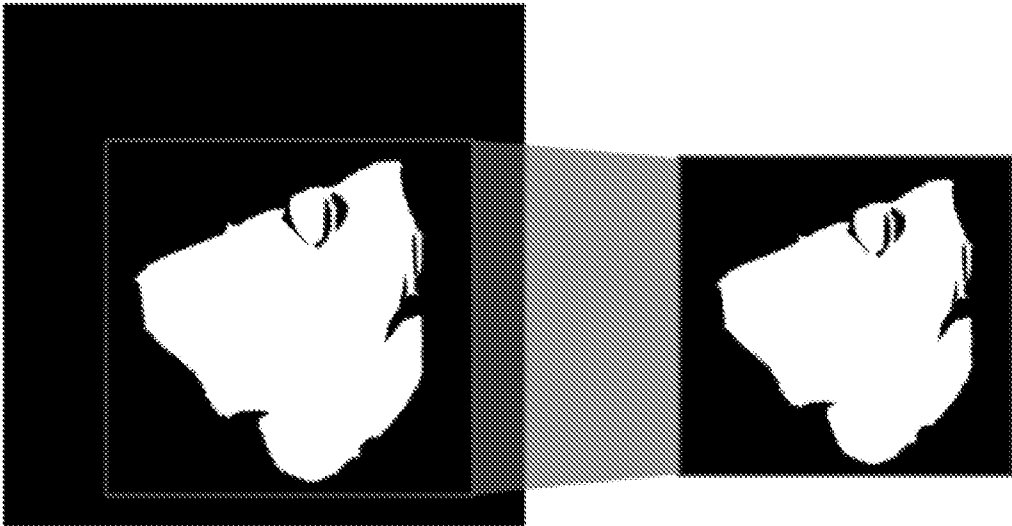


图 17c

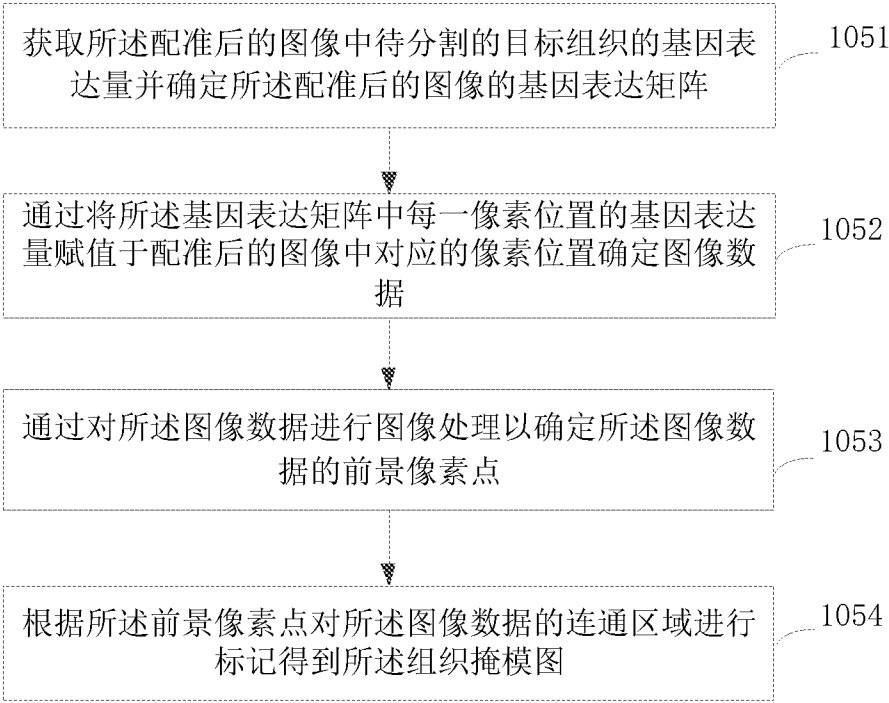


图 18

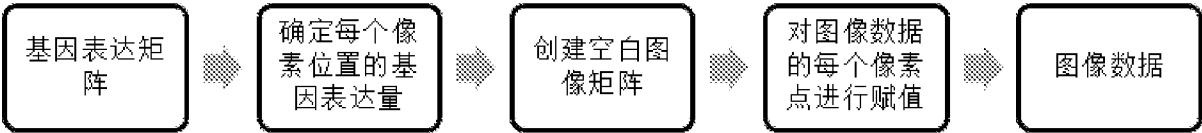


图 19

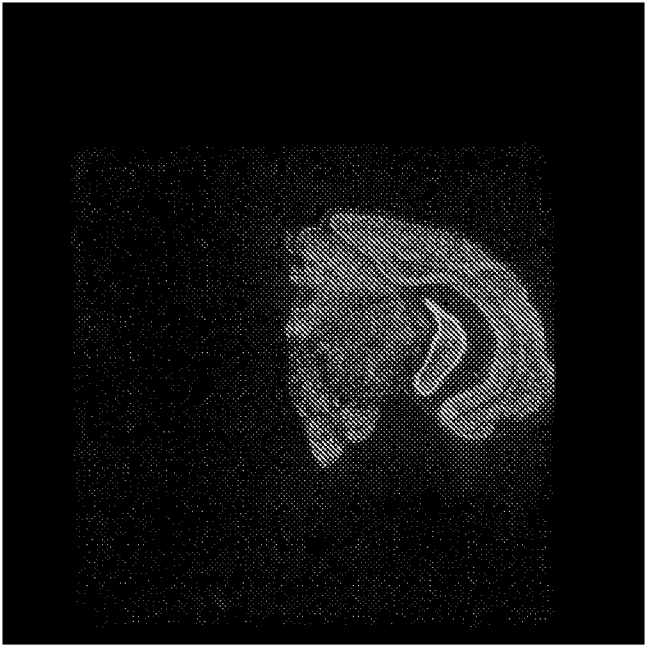


图 20

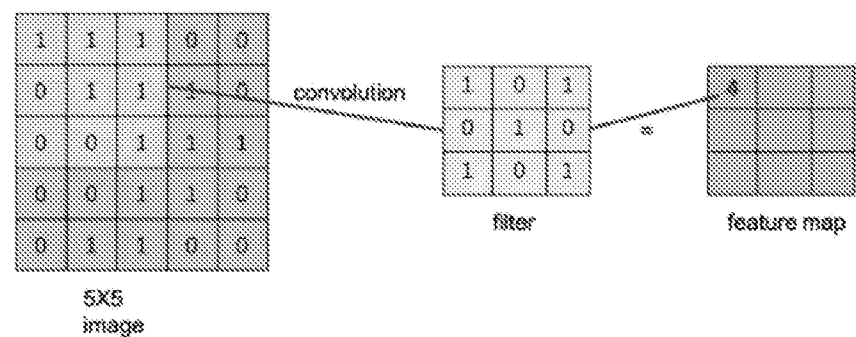


图 21

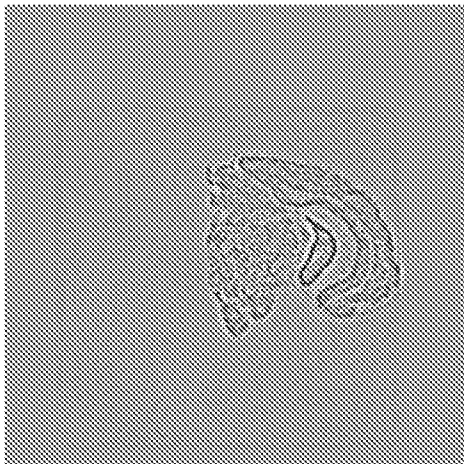


图 22

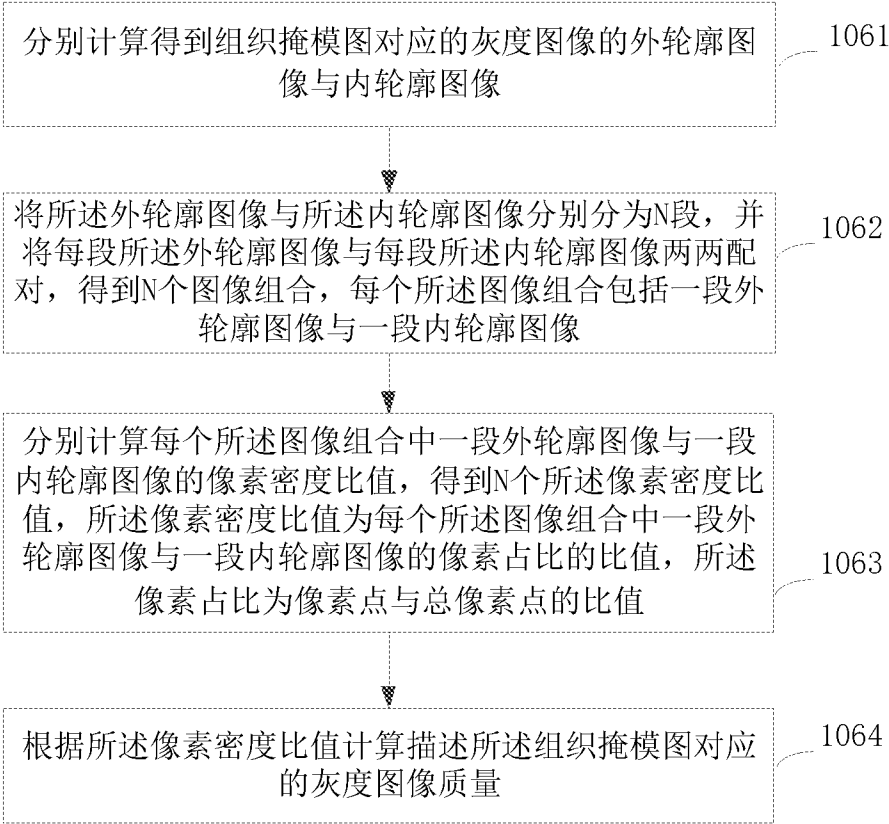


图 23

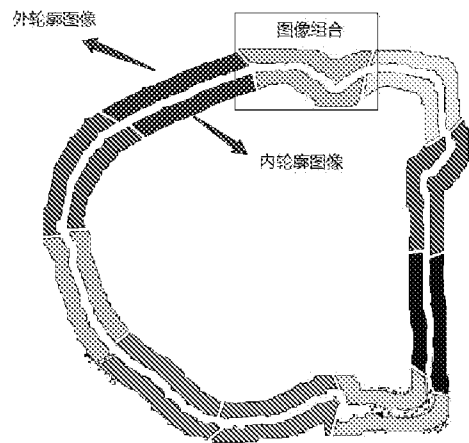


图 24

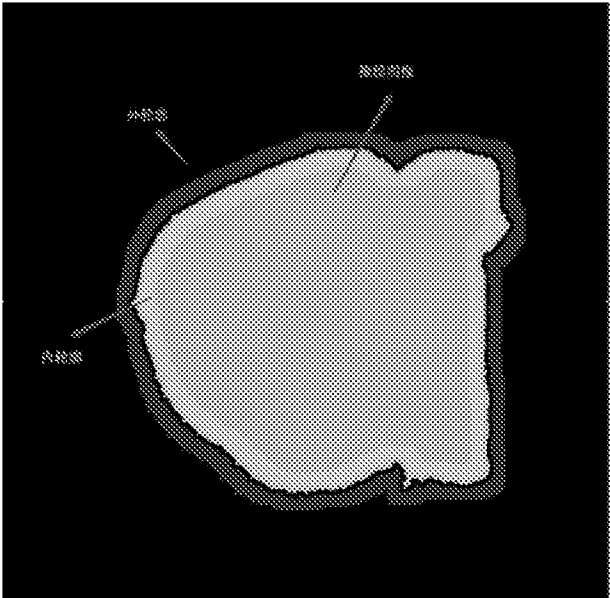


图 25

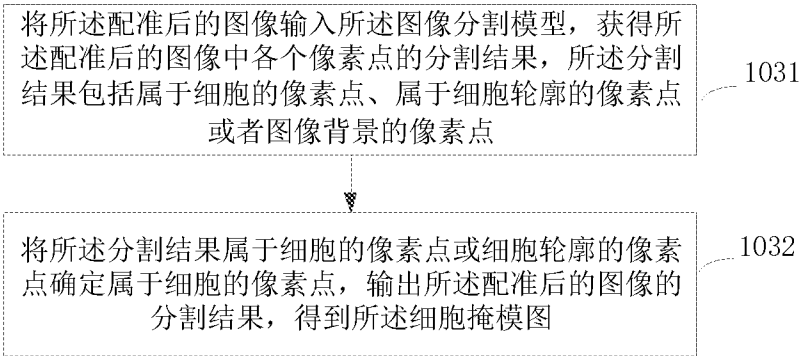


图 26

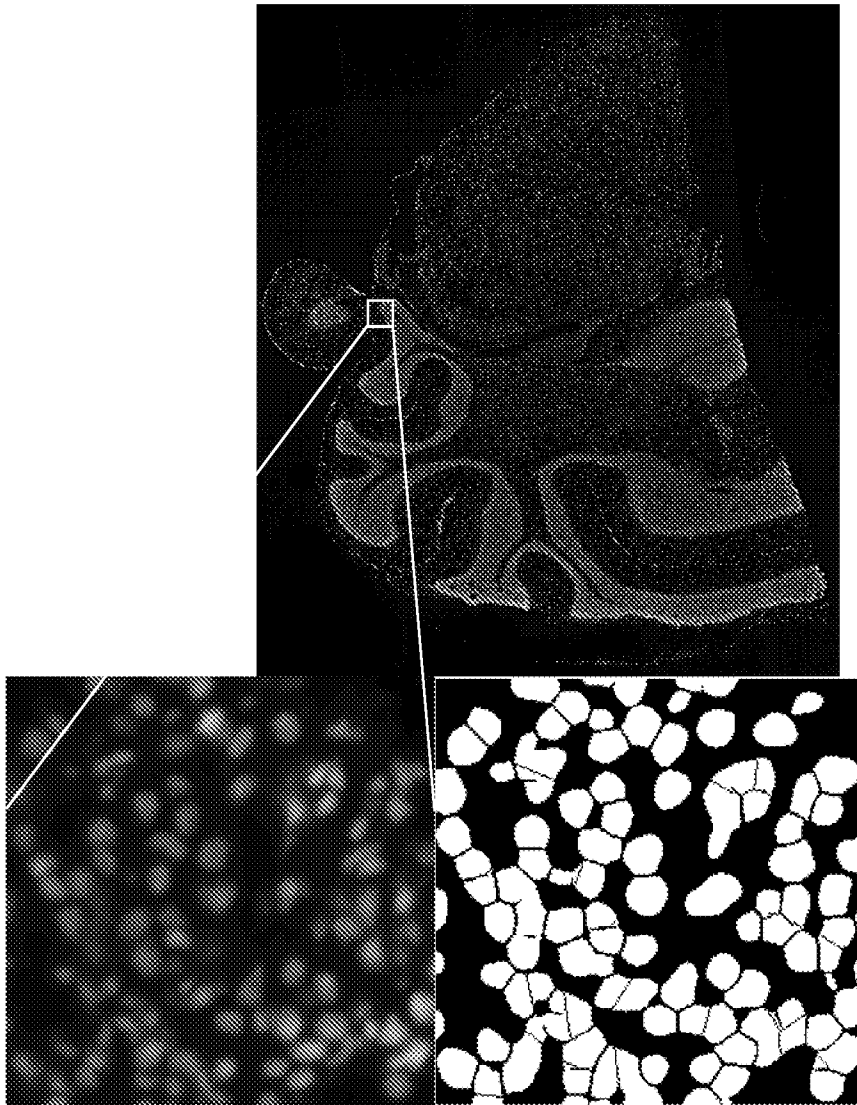


图 27

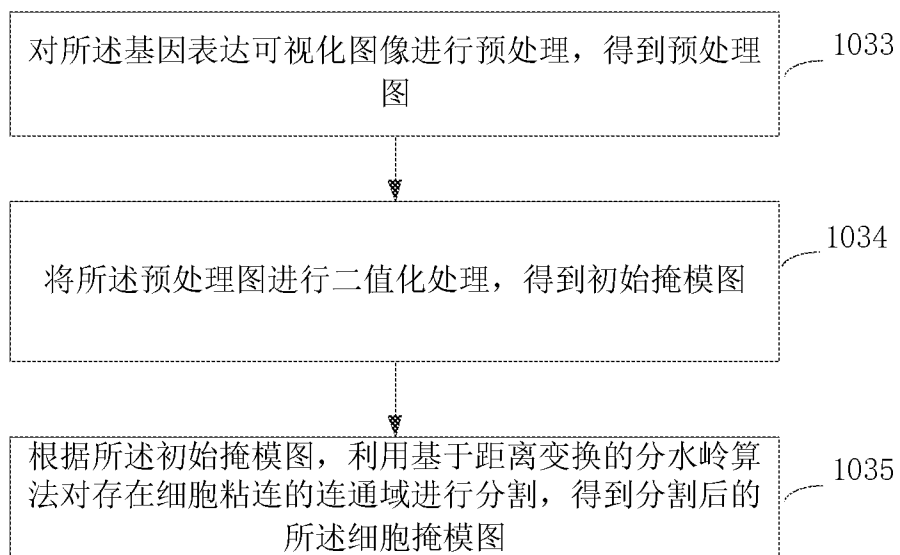


图 28

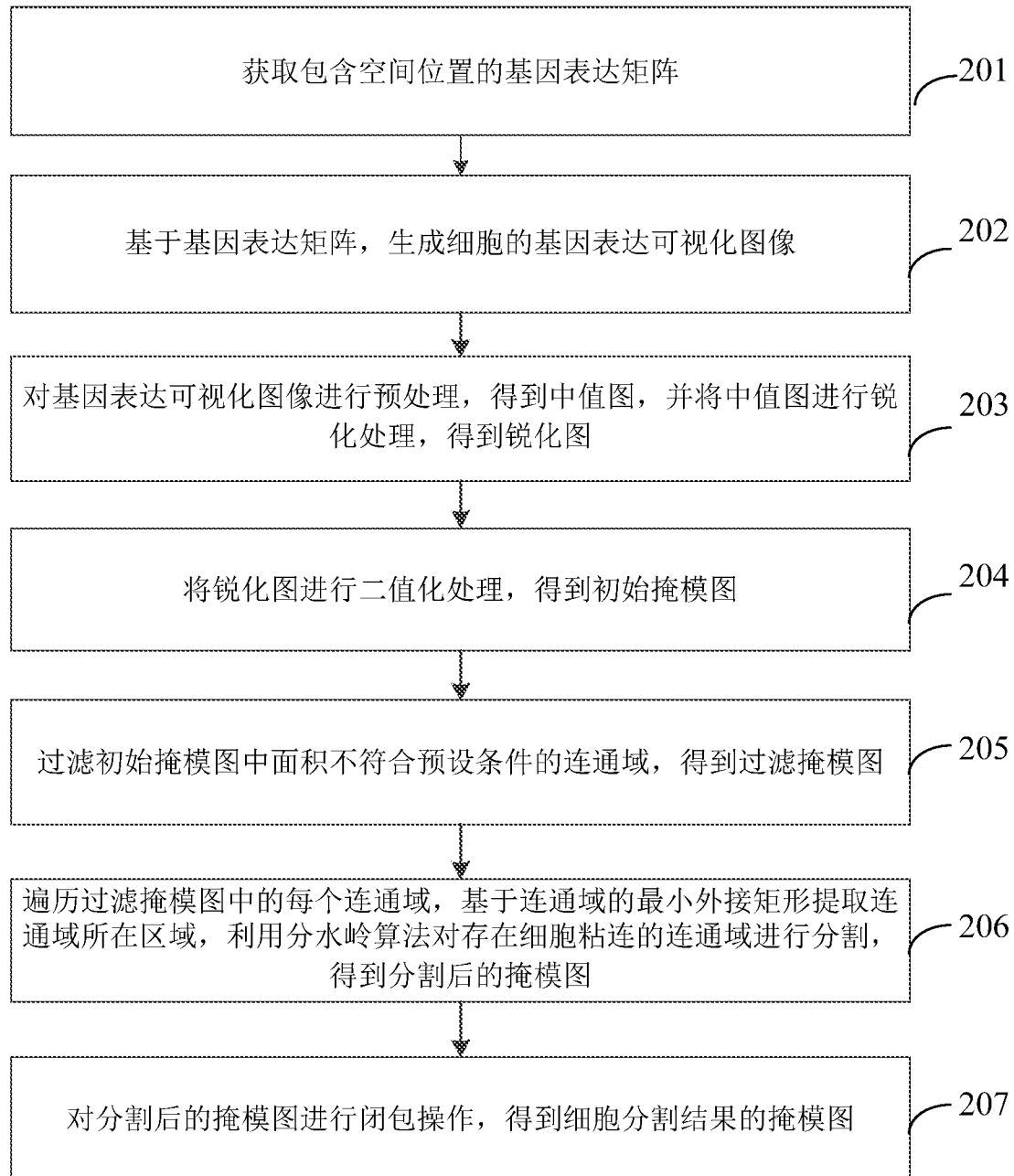


图 29

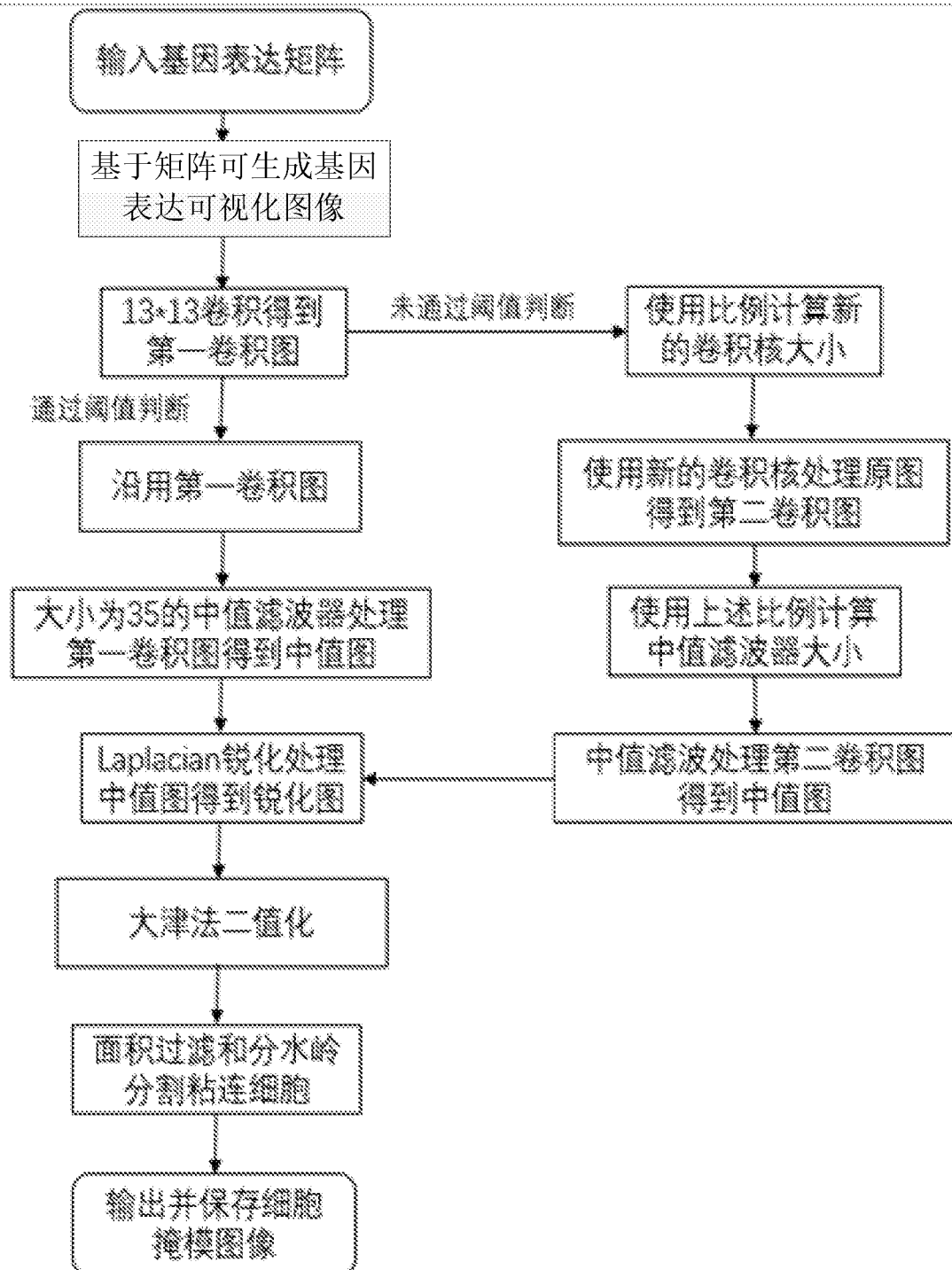


图 30

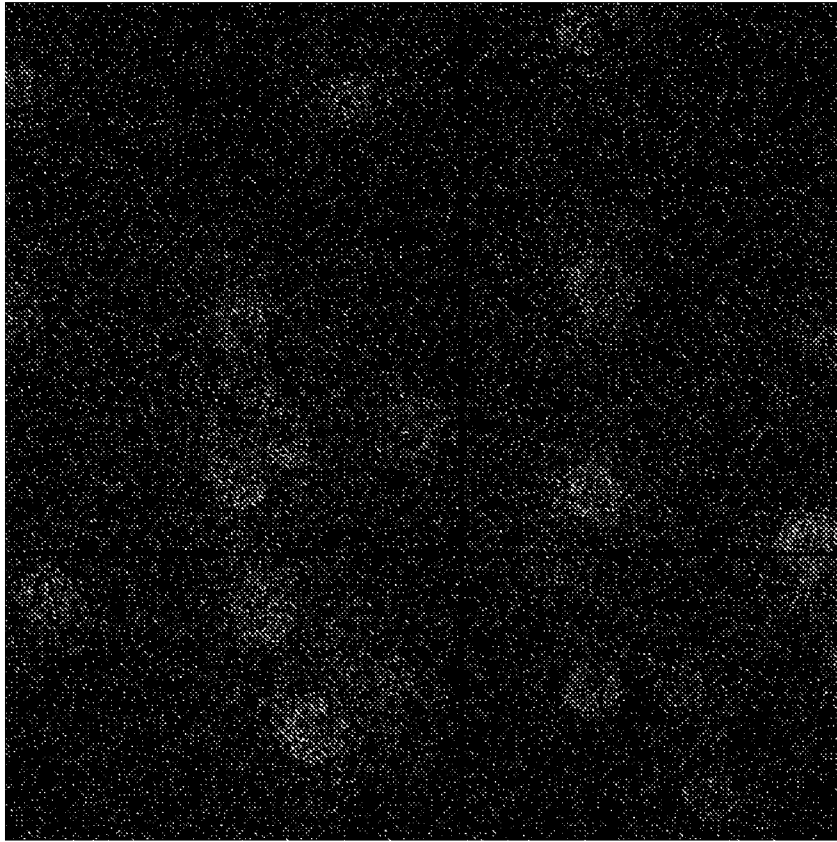


图 31

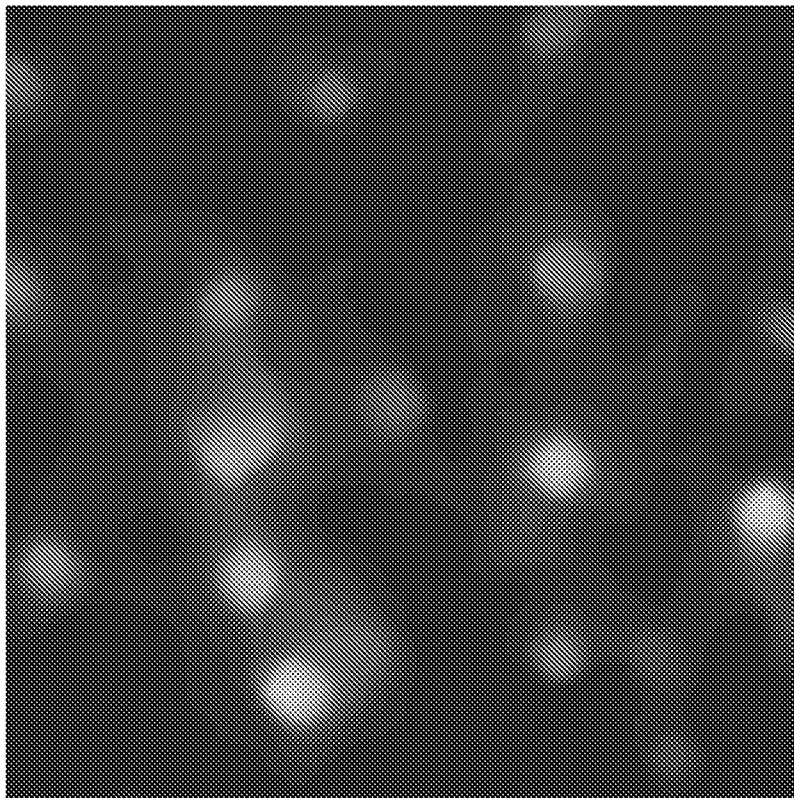


图 32

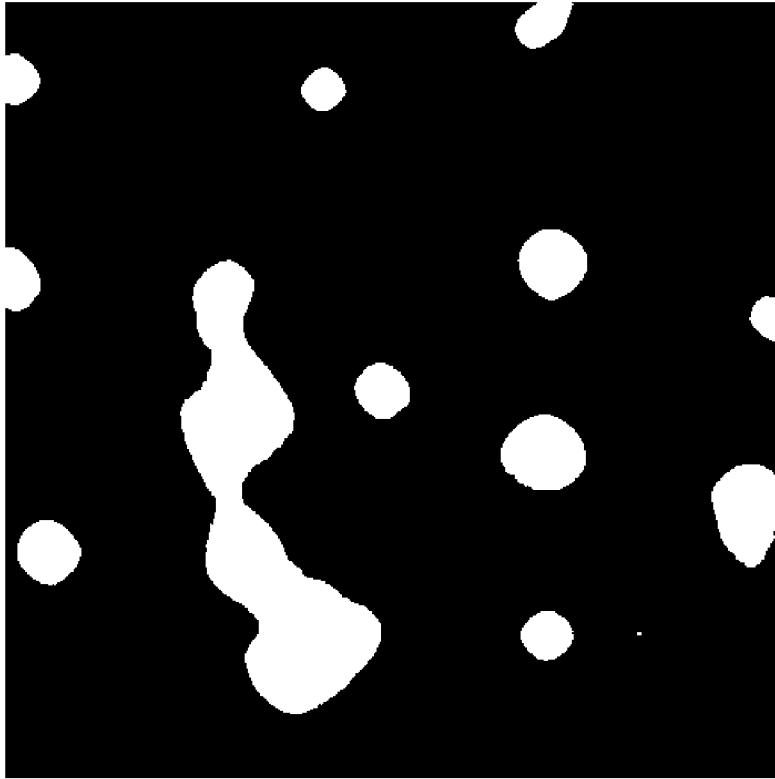


图 33

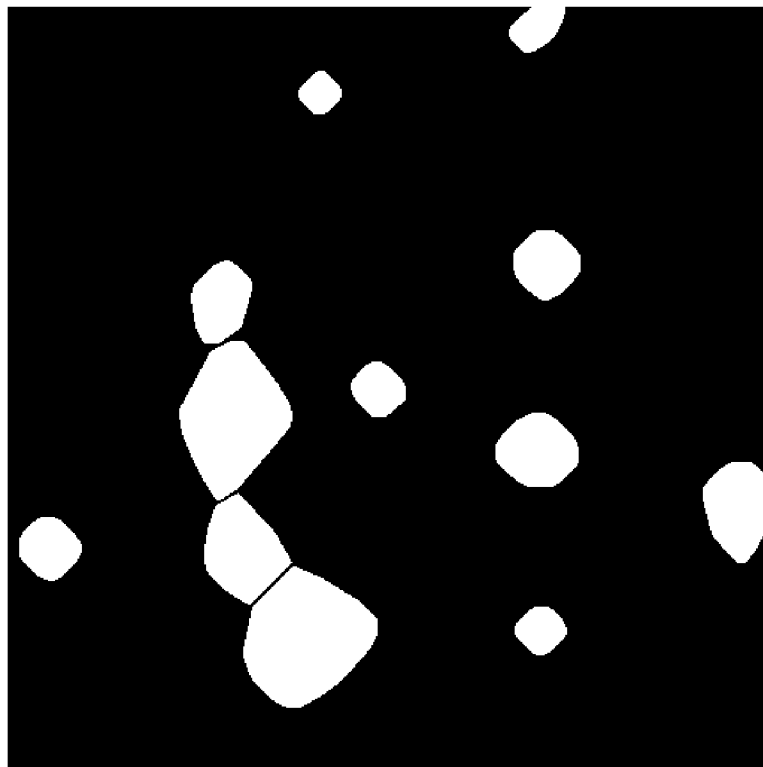


图 34

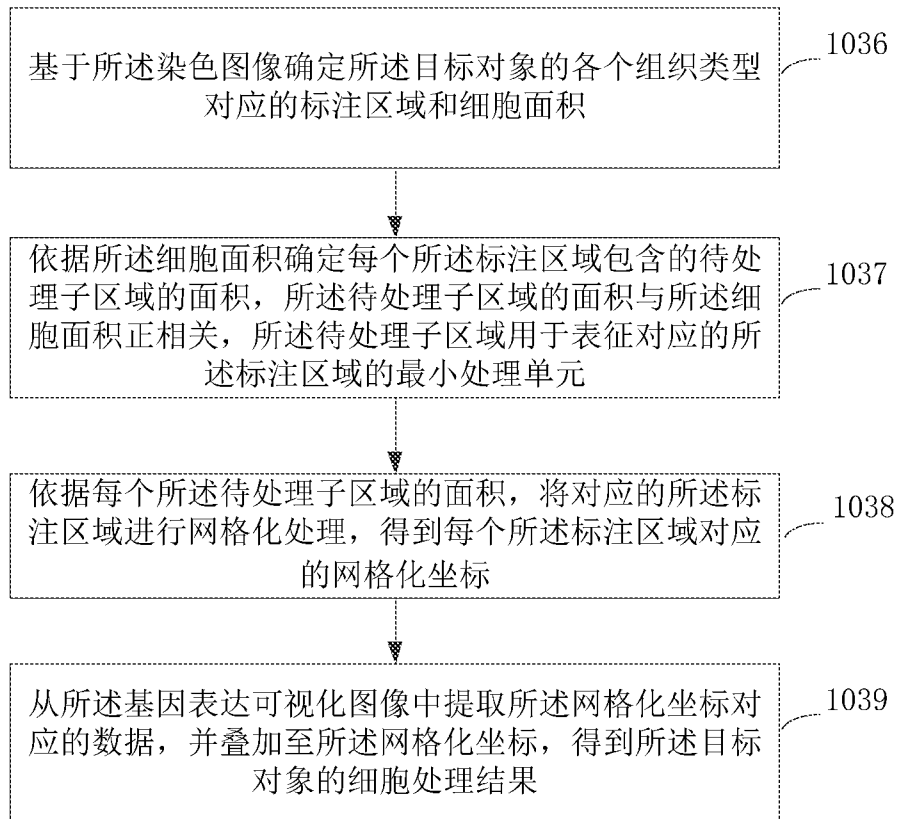


图 35

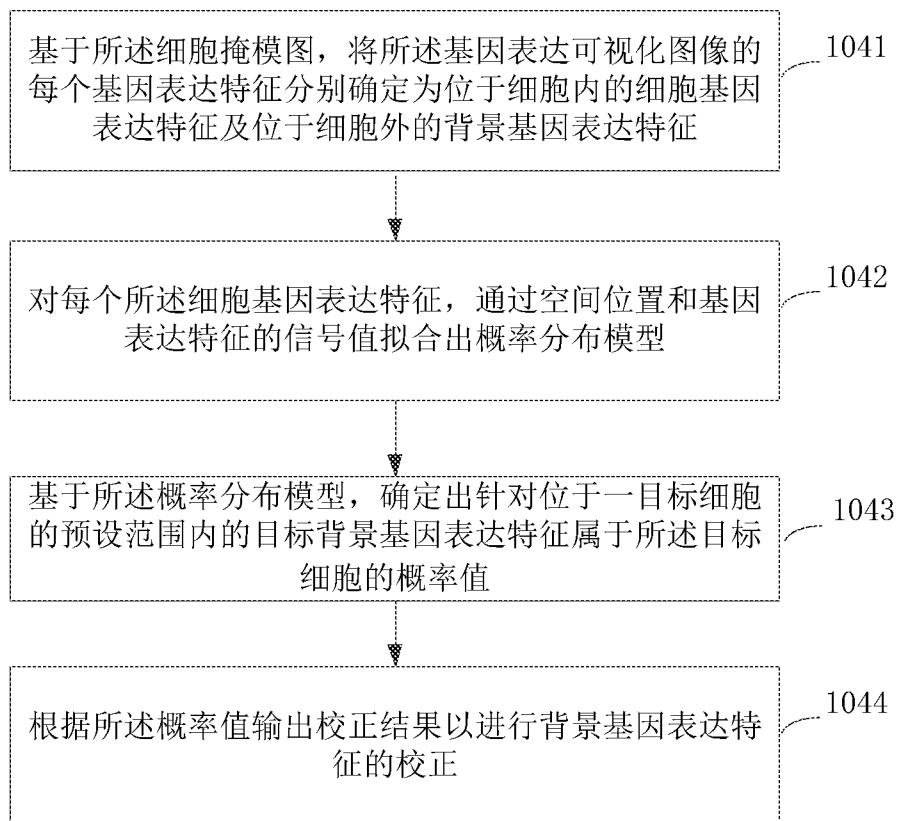


图 36

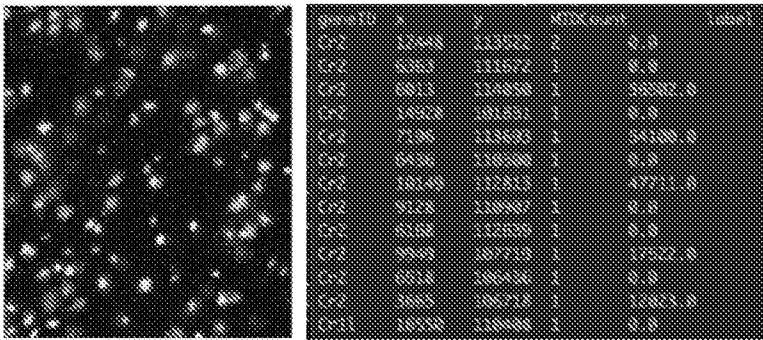


图 37

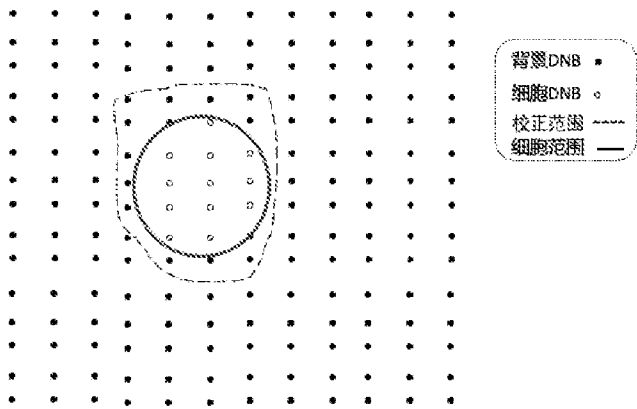


图 38

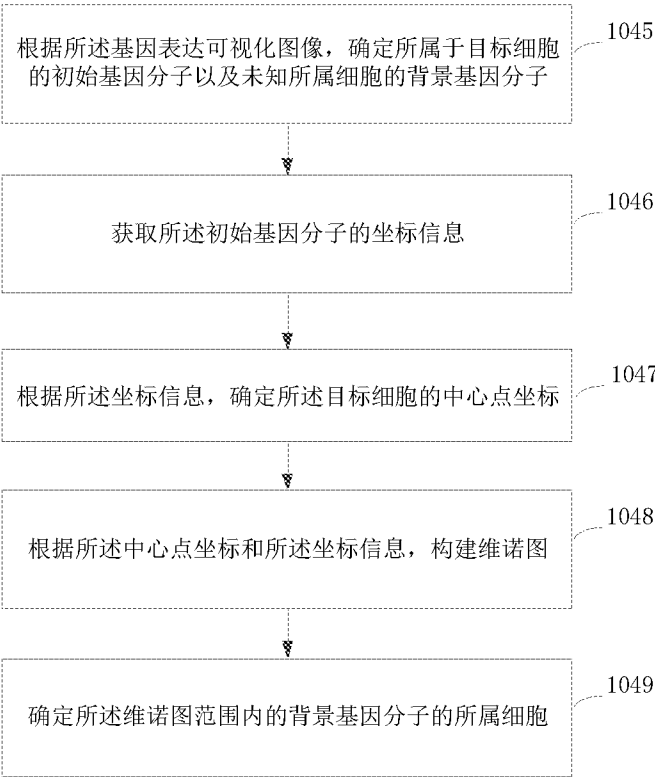


图 39

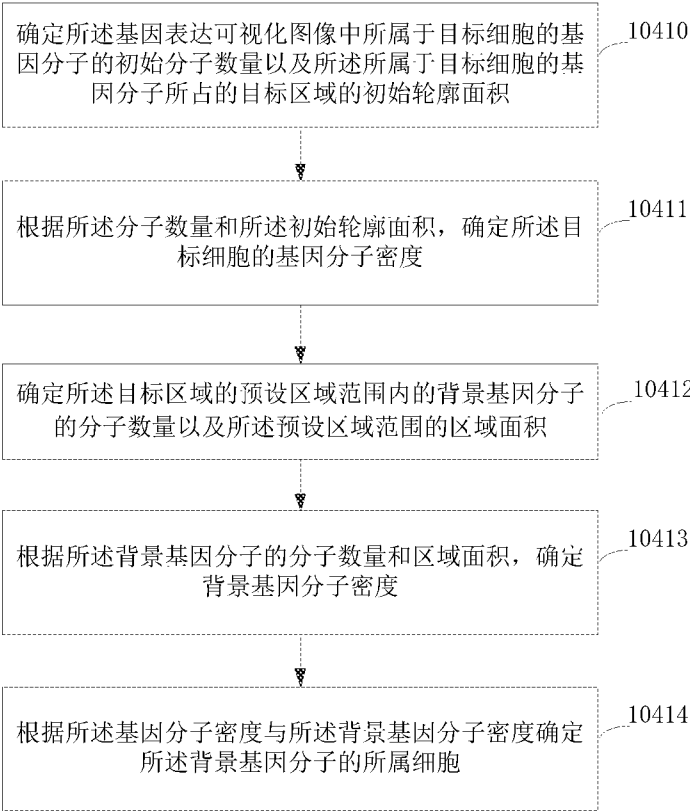


图 40

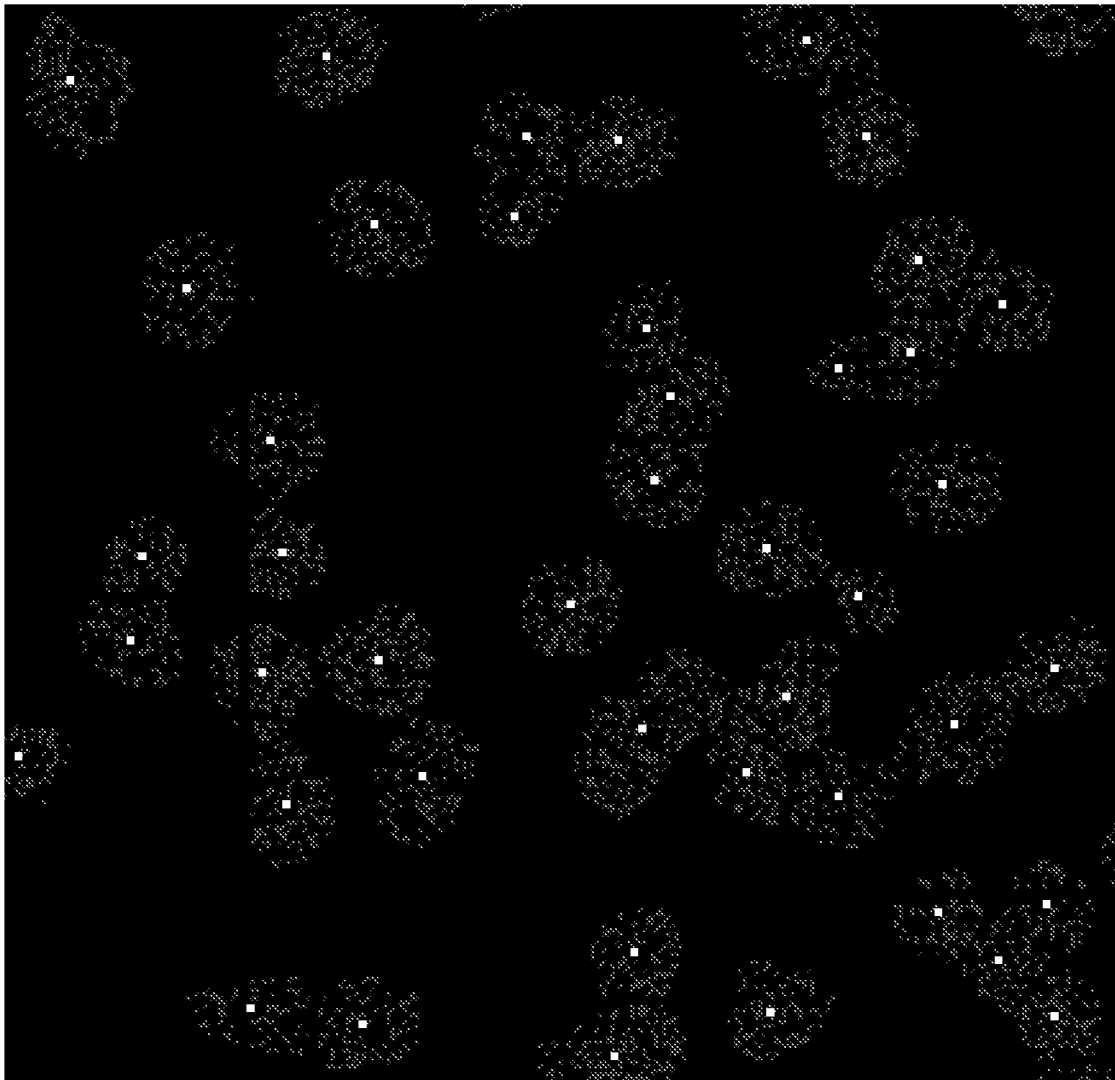


图 41

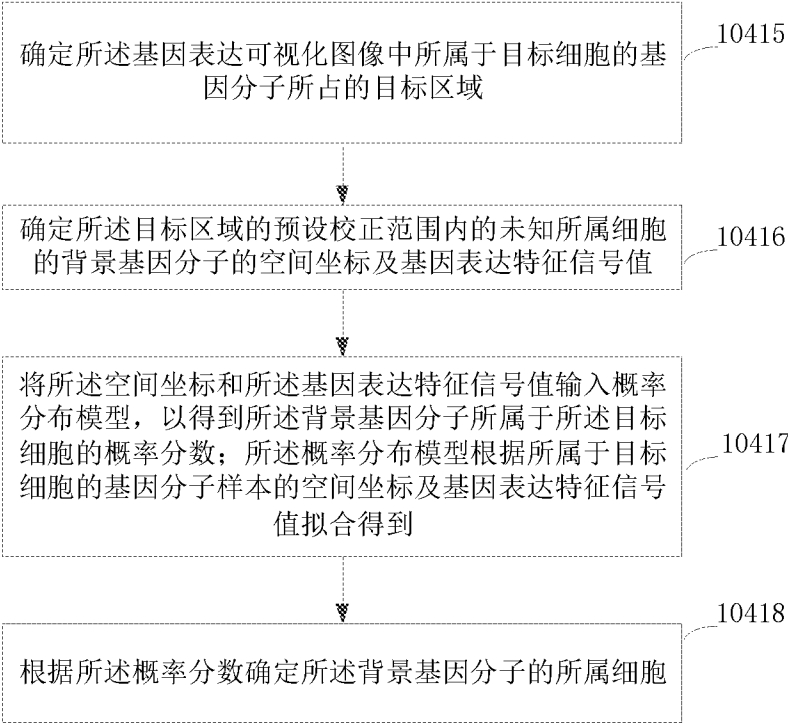


图 42

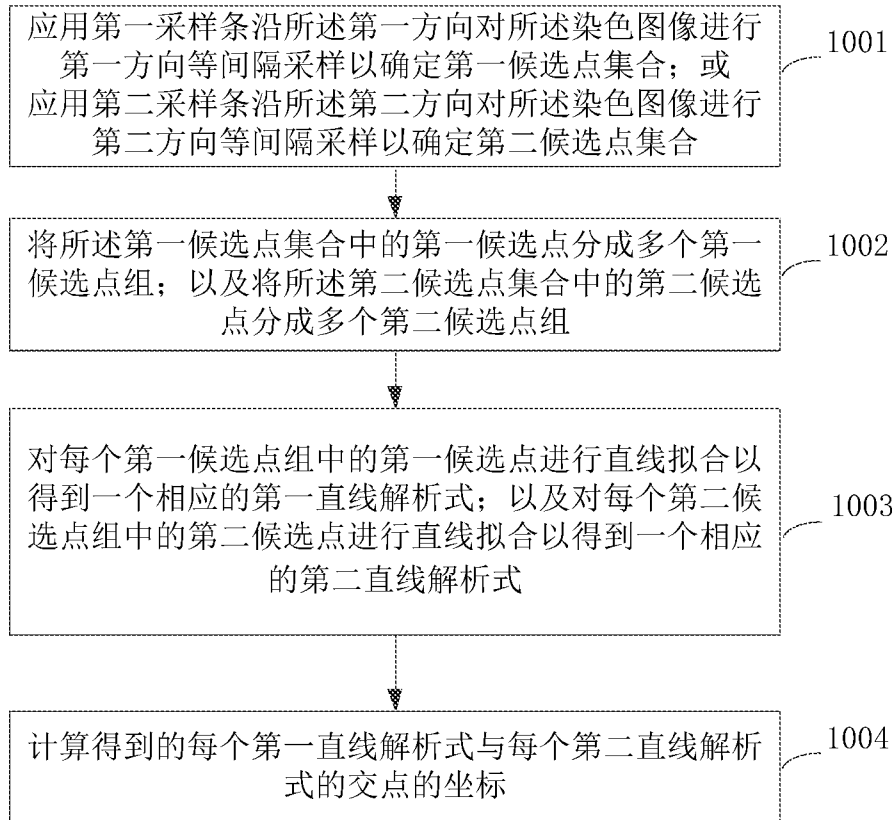


图 43

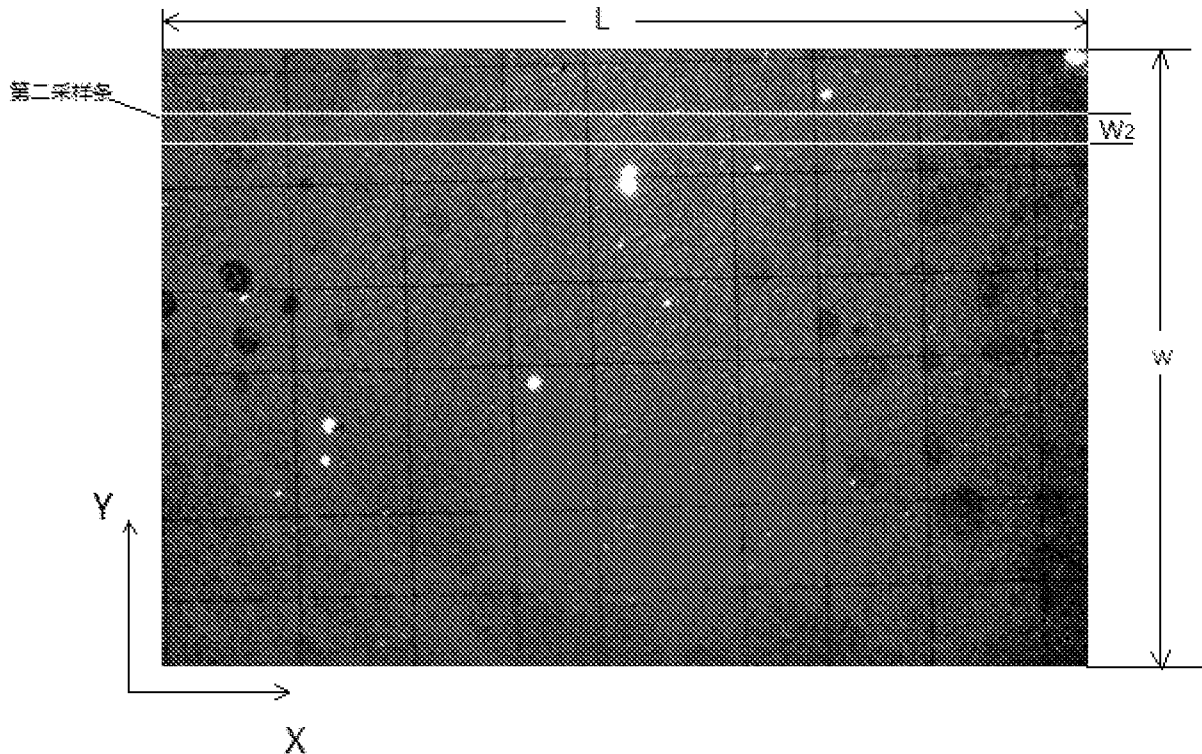


图 44

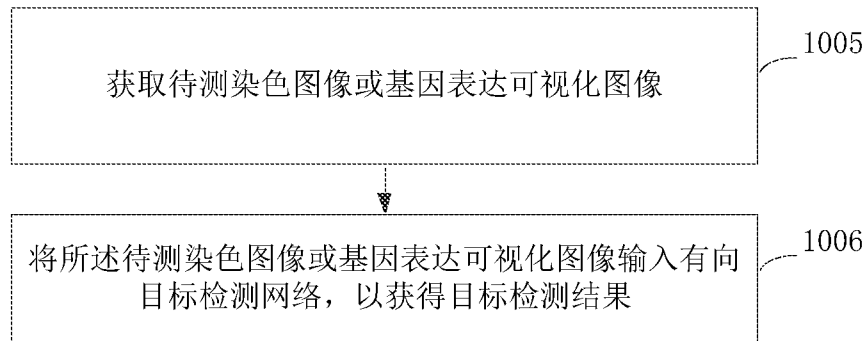


图 45

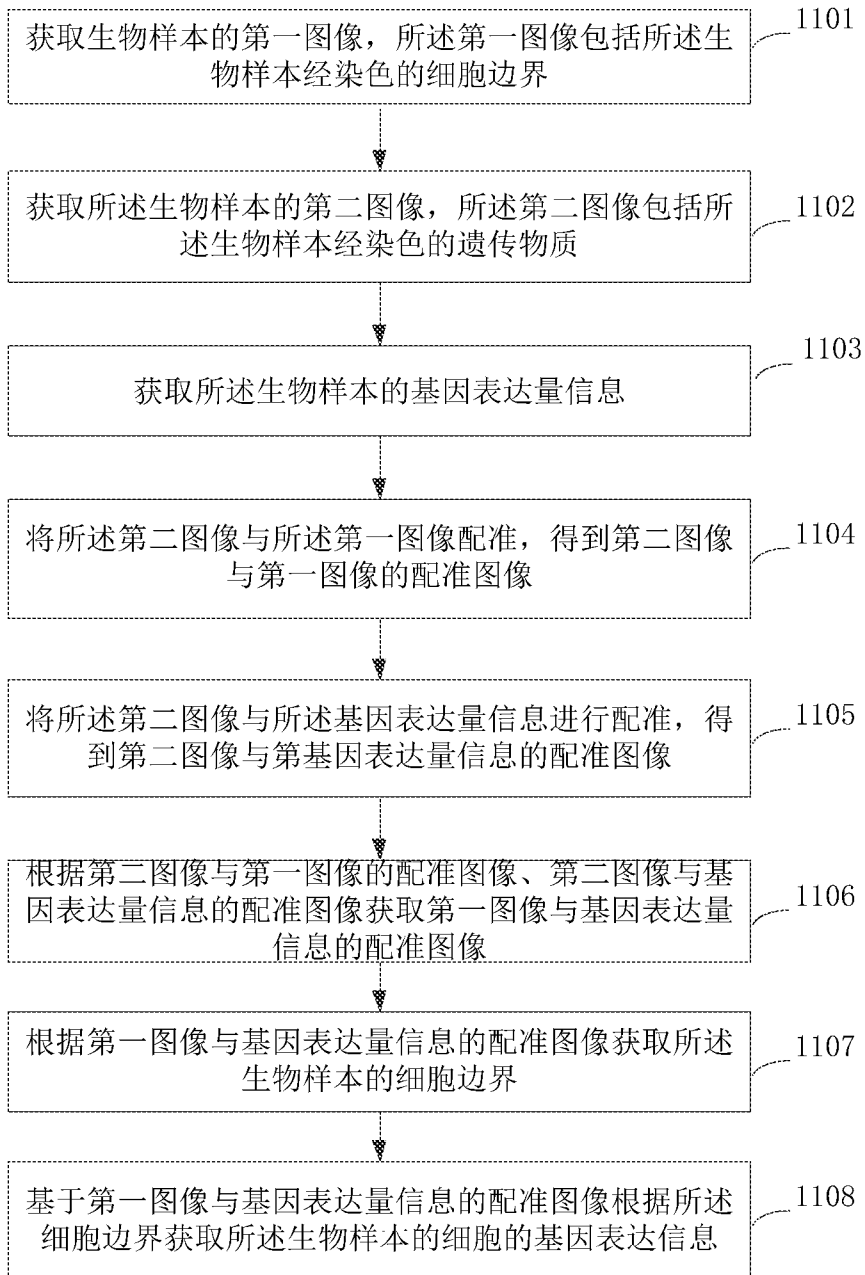


图 46

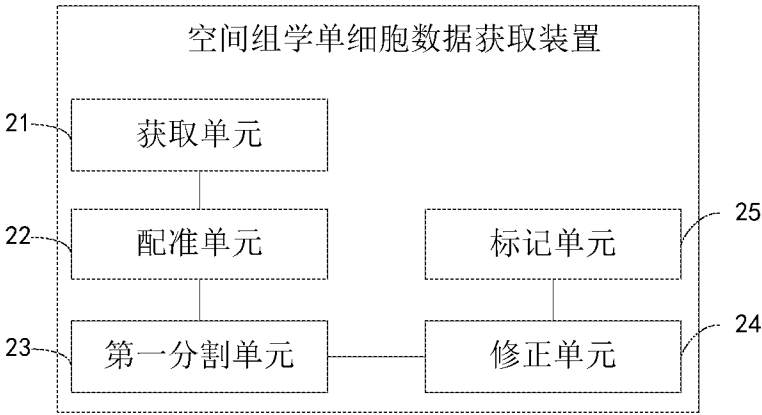


图 47

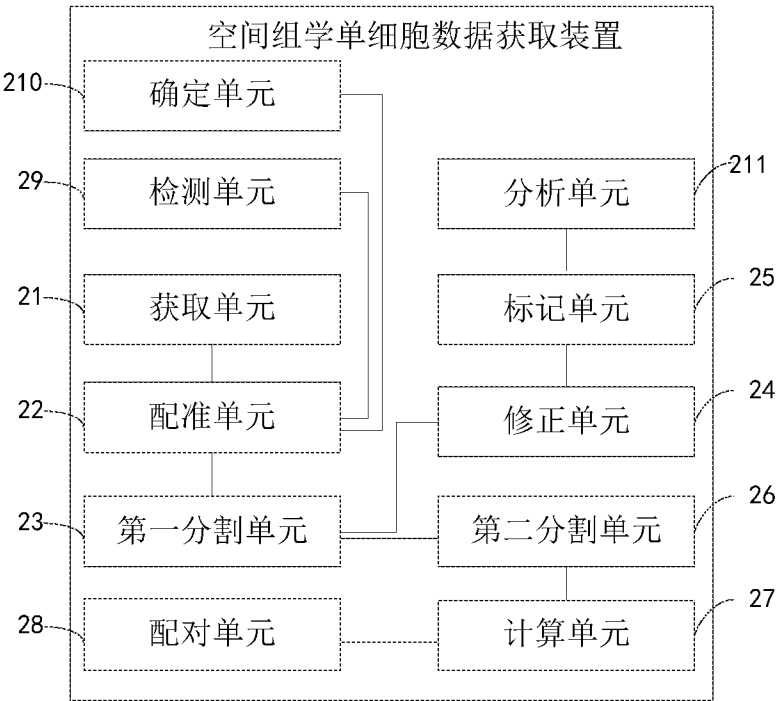


图 48

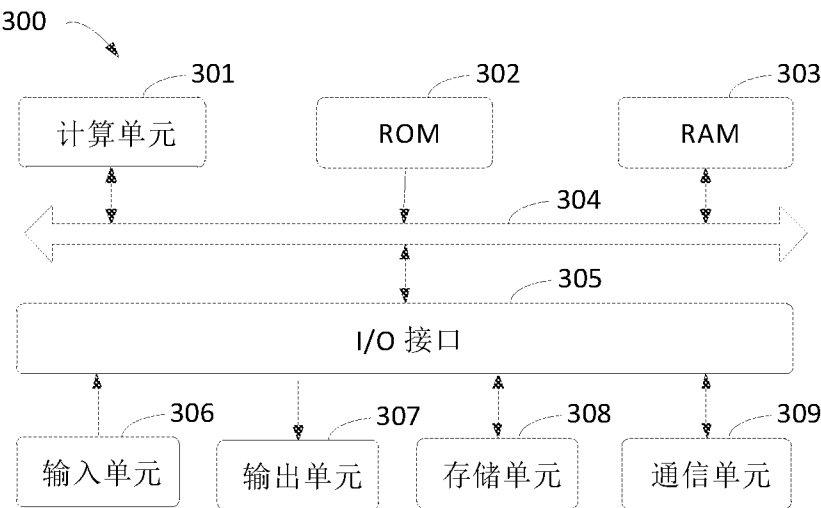


图 49

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2022/136698

A. CLASSIFICATION OF SUBJECT MATTER

G16B 40/30(2019.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC:G16B

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

USTXT; VEN; CNABS; CNTXT; EPTXT; WOTXT; CNKI; IEEE: 单细胞, 染色, 基因, 图像, 配准, 对齐, 分割, 掩模, 标记, 标注, 拼接, 重叠, 偏移, 轨迹, 模板, 坐标, single-cell, stain, gene, image, register, align, segment, mask, label, splice, overlap, offset, track, template, coordinate

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 2020155043 A1 (SHENZHEN BGI LIFE SCIENCES RESEARCH INSTITUTE) 06 August 2020 (2020-08-06) description, pages 8-18	21, 24-26
X	US 2022068438 A1 (THE BROAD INSTITUTE INC. et al.) 03 March 2022 (2022-03-03) description, paragraphs 26-193	1-20, 22-26
Y	US 2022068438 A1 (THE BROAD INSTITUTE INC. et al.) 03 March 2022 (2022-03-03) description, paragraphs 26-193	21, 24-26
Y	WO 2019196019 A1 (SHENZHEN MGI TECHNOLOGY CO., LTD.) 17 October 2019 (2019-10-17) description, pages 8-19	21, 24-26
A	CN 110047107 A (SHENZHEN MGI TECHNOLOGY CO., LTD.) 23 July 2019 (2019-07-23) entire document	1-26
A	CN 113781530 A (SHANGHAI HISTO MEDICAL PATHOLOGY DIAGNOSIS CENTER CO., LTD.) 10 December 2021 (2021-12-10) entire document	1-26



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“D” document cited by the applicant in the international application

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

12 March 2023

Date of mailing of the international search report

22 March 2023

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/
CN)
China No. 6, Xitucheng Road, Jimenqiao, Haidian District,
Beijing 100088

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2022/136698

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CHENG, Mengnan et al. "A Cellular Resolution Spatial Transcriptomic Landscape of the Medial Structures in Postnatal Mouse Brain" <i>Frontiers in Cell and Developmental Biology</i> , Vol. 10, 17 May 2022 (2022-05-17), pages 1-4 and 8	1-20, 22-26
Y	CHENG, Mengnan et al. "A Cellular Resolution Spatial Transcriptomic Landscape of the Medial Structures in Postnatal Mouse Brain" <i>Frontiers in Cell and Developmental Biology</i> , Vol. 10, 17 May 2022 (2022-05-17), pages 1-4 and 8	21, 24-26
A	ZUBAIR, Asif et al. "Cell type identification in spatial transcriptomics data can be improved by leveraging cell-type-informative paired tissue images using a Bayesian probabilistic model" <i>Nucleic Acids Research</i> , Vol. 50, No. 14, 10 May 2022 (2022-05-10), pages 1-14	1-26

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2022/136698

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)		Publication date (day/month/year)	
WO	2020155043	A1	06 August 2020	US	2022108462	A1	07 April 2022
				CN	112204615	A	08 January 2021
				HK	40035019	A0	07 May 2021
US	2022068438	A1	03 March 2022	None			
WO	2019196019	A1	17 October 2019	RU	2749893	C1	18 June 2021
				KR	20200142040	A	21 December 2020
				BR	112020020949	A2	02 March 2021
				EP	3779868	A1	17 February 2021
				CA	3096723	A1	17 October 2019
				US	2021019895	A1	21 January 2021
				JP	2021517701	A	26 July 2021
				JP	7072081	B2	19 May 2022
				AU	2018418609	A1	12 November 2020
				AU	2018418609	B2	07 October 2021
				CN	111971711	A	20 November 2020
				SG	11202010040	A1	27 November 2020
				IN	202017044815	A	22 January 2021
				HK	40033773	A0	16 April 2021
CN	110047107	A	23 July 2019	CN	110047107	B	13 April 2021
				HK	40010290	A0	03 July 2020
CN	113781530	A	10 December 2021	None			

国际检索报告

国际申请号

PCT/CN2022/136698

A. 主题的分类

G16B 40/30 (2019.01) i

按照国际专利分类 (IPC) 或者同时按照国家分类和 IPC 两种分类

B. 检索领域

检索的最低限度文献 (标明分类系统和分类号)

IPC: G16B

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库 (数据库的名称, 和使用的检索词 (如使用))

USTXT; VEN; CNABS; CNTXT; EPTXT; WOTXT; CNKI; IEEE: 单细胞, 染色, 基因, 图像, 配准, 对齐, 分割, 掩模, 标记, 标注, 拼接, 重叠, 偏移, 轨迹, 模板, 坐标, single-cell, stain, gene, image, register, align, segment, mask, label, splice, overlap, offset, track, template, coordinate

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
Y	WO 2020155043 A1 (深圳华大生命科学研究院) 2020年8月6日 (2020 - 08 - 06) 说明第8-18页	21、24-26
X	US 2022068438 A1 (THE BROAD INSTITUTE INC 等) 2022年3月3日 (2022 - 03 - 03) 说明书第26-193段	1-20、22-26
Y	US 2022068438 A1 (THE BROAD INSTITUTE INC 等) 2022年3月3日 (2022 - 03 - 03) 说明书第26-193段	21、24-26
Y	WO 2019196019 A1 (深圳华大智造科技有限公司) 2019年10月17日 (2019 - 10 - 17) 说明书第8-19页	21、24-26
A	CN 110047107 A (深圳华大智造科技有限公司) 2019年7月23日 (2019 - 07 - 23) 全文	1-26
A	CN 113781530 A (上海衡道医学病理诊断中心有限公司) 2021年12月10日 (2021 - 12 - 10) 全文	1-26



其余文件在C栏的续页中列出。



见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“D” 申请人在国际申请中引证的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

2023年3月12日

国际检索报告邮寄日期

2023年3月22日

ISA/CN的名称和邮寄地址

中国国家知识产权局

中国北京市海淀区蓟门桥西土城路6号 100088

传真号 (86-10)62019451

授权官员

李富贵

电话号码 (+86) 020-28958943

C. 相关文件		
类 型*	引用文件，必要时，指明相关段落	相关的权利要求
X	CHENG, Mengnan 等. "A Cellular Resolution Spatial Transcriptomic Landscape of the Medial Structures in Postnatal Mouse Brain" Frontiers in Cell and Developmental Biology, 第10卷, 2022年5月17日 (2022 - 05 - 17), 第1-4、8页	1-20、22-26
Y	CHENG, Mengnan 等. "A Cellular Resolution Spatial Transcriptomic Landscape of the Medial Structures in Postnatal Mouse Brain" Frontiers in Cell and Developmental Biology, 第10卷, 2022年5月17日 (2022 - 05 - 17), 第1-4、8页	21、24-26
A	ZUBAIR, Asif 等. "Cell type identification in spatial transcriptomics data can be improved by leveraging cell-type-informative paired tissue images using a Bayesian probabilistic model" Nucleic Acids Research, 第50卷, 第14期, 2022年5月10日 (2022 - 05 - 10), 第1-14页	1-26

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2022/136698

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
WO	2020155043	A1	2020年8月6日	US	2022108462	A1	2022年4月7日
				CN	112204615	A	2021年1月8日
				HK	40035019	A0	2021年5月7日
US	2022068438	A1	2022年3月3日	无			
WO	2019196019	A1	2019年10月17日	RU	2749893	C1	2021年6月18日
				KR	20200142040	A	2020年12月21日
				BR	112020020949	A2	2021年3月2日
				EP	3779868	A1	2021年2月17日
				CA	3096723	A1	2019年10月17日
				US	2021019895	A1	2021年1月21日
				JP	2021517701	A	2021年7月26日
				JP	7072081	B2	2022年5月19日
				AU	2018418609	A1	2020年11月12日
				AU	2018418609	B2	2021年10月7日
				CN	111971711	A	2020年11月20日
				SG	11202010040	A1	2020年11月27日
				IN	202017044815	A	2021年1月22日
				HK	40033773	A0	2021年4月16日
CN	110047107	A	2019年7月23日	CN	110047107	B	2021年4月13日
				HK	40010290	A0	2020年7月3日
CN	113781530	A	2021年12月10日	无			