(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 4 月 4 日 (04.04.2024)

WIPO | PCT

(10) 国际公布号
WO 2024/066429 A1

(51) 国际专利分类号:
G06V 10/40 (2022.01) *G06V 10/762* (2022.01)

(21) 国际申请号: PCT/CN2023/097408

(22) 国际申请日: 2023 年 5 月 31 日 (31.05.2023)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202211176018.8 2022年9月26日 (26.09.2022) CN

(71) 申请人: 腾 讯 科 技 (深 圳) 有 限 公 司 **(TENCENT TECHNOLOGY (SHENZHEN) COMPANY LIMITED)** [CN/CN]; 中国广东省深圳市南山区高新区科技中一路腾讯大厦35层, Guangdong 518057 (CN)。

(72) 发明人:马星宇**(MA, Xingyu)**; 中国广东省深圳市南山区高新区科技中一路腾讯大厦35层, Guangdong 518057 (CN)。 郝少刚**(HAO, Shaogang)**; 中国广东省深圳市南山区高新区科技中一路腾讯大厦35层, Guangdong 518057 (CN)。 张胜誉**(ZHANG, Shengyu)**; 中国广东省深圳市南山区高新区科技中一路腾讯大厦35层, Guangdong 518057 (CN)。

(74) 代理人:北京三高永信知识产权代理有限责任公司**(BEIJING SAN GAO YONG XIN INTELLECTUAL PROPERTY AGENCY CO., LTD.)**; 中国北京市海淀区上地信息产业基地三街1号楼四层C段457, Beijing 100085 (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,

(54) **Title:** CHIP LAYOUT ENCODER TRAINING METHOD AND DEVICE, AND CHIP LAYOUT SCREENING METHOD AND DEVICE

(54) 发明名称: 用于芯片版图编码器的训练、芯片版图的筛选方法及装置

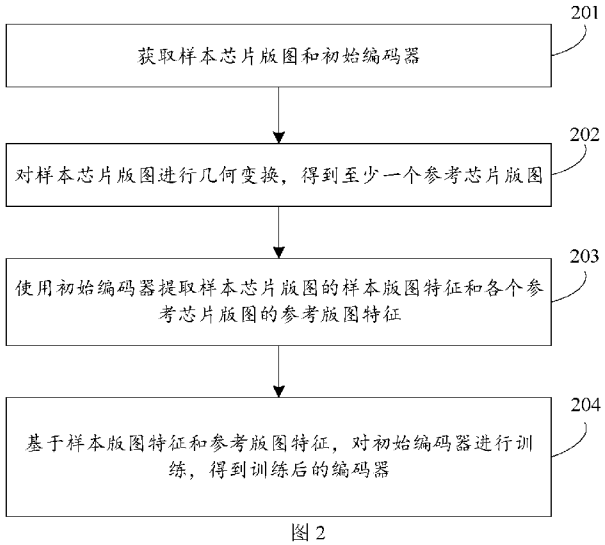


图 2

201 Acquire a sample chip layout and an initial encoder
202 Perform geometric transformation on the sample chip layout to obtain at least one reference chip layout
203 Extract sample layout features of the sample chip layout and reference layout features of the reference chip layouts by using the initial encoder
204 Train the initial encoder on the basis of the sample layout features and the reference layout features to obtain a trained encoder

(57) **Abstract:** A chip layout encoder training method and device, and a chip layout screening method and device, relating to the technical field of integrated circuits. The chip layout encoder training method comprises: acquiring a sample chip layout and an initial encoder (201); performing geometric transformation on the sample chip layout to obtain at least one reference chip layout (202); extracting sample layout features of the sample chip layout and reference layout features of the reference chip layouts by using the initial encoder (203); and training the initial encoder on the basis of the sample layout features and the reference layout features to obtain a trained encoder (204). By using the methods and the devices, for the chip layouts before and after geometric transformation, a chip layout encoder can output similar layout features, such that the chip layout encoder focuses on extracting a layout feature capable of distinguishing a chip layout type, thereby facilitating subsequent improvement of the accuracy of a clustering result when the chip layout is clustered on the basis of the layout feature.

WO 2024/066429 A1

[见续页]

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, MU, 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))。

(57) 摘要: 用于芯片版图编码器的训练、芯片版图的筛选方法及装置，属于集成电路技术领域。方法包括：获取样本芯片版图和初始编码器(201)；对样本芯片版图进行几何变换，得到至少一个参考芯片版图(202)；使用初始编码器提取样本芯片版图的样本版图特征和各个参考芯片版图的参考版图特征(203)；基于样本版图特征和参考版图特征，对初始编码器进行训练，得到训练后的编码器(204)。采用上述方法和装置，能够针对几何变换前后的芯片版图，芯片版图编码器能够输出相近的版图特征，因此，芯片版图编码器侧重于提取能区分芯片版图类型的版图特征，有利于后续基于版图特征对芯片版图进行聚类时提高聚类结果的准确性。

用于芯片版图编码器的训练、芯片版图的筛选方法及装置

本申请要求于 2022 年 09 月 26 日提交的申请号为 202211176018.8、发明名称为“芯片版图编码器的训练方法、芯片版图的筛选方法及装置”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本申请实施例涉及集成电路技术领域，特别涉及一种用于芯片版图编码器的训练、芯片版图的筛选方法及装置。

背景技术

随着集成电路（Integrated Circuit，IC）技术的不断发展，IC 版图（也称芯片版图）的特征尺寸也在不断缩小，其中，芯片版图的特征尺寸指的是半导体器件的最小尺寸。目前芯片版图的特征尺寸已达到纳米级别。

通常情况下，需要先调用光刻机利用光源将芯片版图曝光在感光胶上得到掩膜版图，再调用光刻机利用光源将掩膜版图曝光在晶圆片上得到成像版图。由于芯片版图的特征尺寸低于光刻工艺所使用的光源波长，导致干涉效应和衍射效应增强，致使将掩膜版图曝光在晶圆片上的成像版图发生扭曲模糊，进而降低了芯片的性能和良率。在这种情况下，需要对芯片版图进行光源掩膜优化，得到优化后的掩膜版图，从而提升芯片的性能和良率。

对芯片版图进行光源掩膜优化时，需要提取芯片版图的版图特征。因此，如何训练得到能提取芯片版图的版图特征的芯片版图编码器是亟需解决的问题。

发明内容

本申请提供了一种用于芯片版图编码器的训练、芯片版图的筛选方法及装置，以提取芯片版图的版图特征并对芯片版图进行筛选，所述技术方案包括如下内容：

一方面，提供了一种用于芯片版图编码器的训练方法，所述方法由电子设备执行，所述方法包括：

获取样本芯片版图和初始编码器；

对所述样本芯片版图进行几何变换，得到至少一个参考芯片版图；

使用所述初始编码器提取所述样本芯片版图的样本版图特征和各个参考芯片版图的参考版图特征；

基于所述样本芯片版图的版图特征和所述各个参考芯片版图的版图特征，对所述初始编码器进行训练，得到芯片版图编码器，所述芯片版图编码器用于提取目标芯片版图的版图特征。或者说，基于所述样本版图特征和所述参考版图特征，对所述初始编码器进行训练，得到训练后的编码器。

另一方面，提供了一种芯片版图的筛选方法，所述方法由电子设备执行，所述方法包括：

获取多个目标芯片版图和芯片版图编码器（即训练后的编码器），所述芯片版图编码器（即训练后的编码器）是上述所述的用于芯片版图编码器的训练方法训练得到的；

使用所述芯片版图编码器（即训练后的编码器）提取各个目标芯片版图的目标版图特征；

基于各个目标版图特征对所述多个目标芯片版图进行聚类处理，得到多个目标聚类簇，任一个目标聚类簇中包括至少一个目标芯片版图；

对于任一个目标聚类簇，从所述任一个目标聚类簇包括的各个目标芯片版图中筛选出关键芯片版图。

另一方面，提供了一种用于芯片版图编码器的训练装置，所述装置包括：

获取模块，用于获取样本芯片版图和初始编码器；

变换模块，用于对所述样本芯片版图进行几何变换，得到至少一个参考芯片版图；

提取模块，用于使用所述初始编码器提取所述样本芯片版图的样本版图特征和各个参考芯片版图的参考版图特征；

训练模块，用于基于所述样本芯片版图的版图特征和所述各个参考芯片版图的版图特征，对所述初始编码器进行训练，得到芯片版图编码器，所述芯片版图编码器用于提取目标芯片版图的版图特征。或者说，训练模块，用于基于所述样本版图特征和所述参考版图特征，对所述初始编码器进行训练，得到训练后的编码器。

另一方面，提供了一种芯片版图的筛选装置，所述装置包括：

获取模块，用于获取多个目标芯片版图和芯片版图编码器（即训练后的编码器），所述芯片版图编码器（即训练后的编码器）是按照上述所述的用于芯片版图编码器的训练方法训练得到的；

提取模块，用于利用所述芯片版图编码器（即训练后的编码器）提取各个目标芯片版图的目标版图特征；

聚类模块，用于基于各个目标版图特征对所述多个目标芯片版图进行聚类处理，得到多个目标聚类簇，任一个目标聚类簇中包括至少一个目标芯片版图；

筛选模块，用于对于任一个目标聚类簇，从所述任一个目标聚类簇包括的各个目标芯片版图中筛选出关键芯片版图。

另一方面，提供了一种电子设备，所述电子设备包括处理器和存储器，所述存储器中存储有至少一条计算机程序，所述至少一条计算机程序由所述处理器加载并执行，以使所述电子设备实现上述所述的用于芯片版图编码器的训练方法或者实现上述所述的芯片版图的筛选方法。

另一方面，还提供了一种非易失性计算机可读存储介质，所述非易失性计算机可读存储介质中存储有至少一条计算机程序，所述至少一条计算机程序由处理器加载并执行，以使电子设备实现上述所述的用于芯片版图编码器的训练方法或者实现上述所述的芯片版图的筛选方法。

另一方面，还提供了一种计算机程序，所述计算机程序为至少一条，至少一条计算机程序由处理器加载并执行，以使电子设备实现上述所述的用于芯片版图编码器的训练方法或者实现上述所述的芯片版图的筛选方法。

再一方面，还提供了一种计算机程序产品，所述计算机程序产品中存储有至少一条计算机程序，所述至少一条计算机程序由处理器加载并执行，以使电子设备实现上述所述的用于芯片版图编码器的训练方法或者实现上述所述的芯片版图的筛选方法。

本申请提供的技术方案对样本芯片版图进行几何变换得到参考芯片版图，利用样本芯片版图的版图特征和参考芯片版图的版图特征训练初始编码器得到芯片版图编码器，使得针对几何变换前后的芯片版图，芯片版图编码器能够输出相近的版图特征。由于几何变换前后的芯片版图属于同一种芯片版图类型，因此，芯片版图编码器侧重于提取能区分芯片版图类型的版图特征，有利于后续基于版图特征对芯片版图进行聚类处理和筛选处理时，提高聚类结果的准确性，降低筛选结果的冗余性。

附图说明

图 1 是本申请实施例提供的一种用于芯片版图编码器的训练方法或者芯片版图的筛选方法的实施环境示意图；

图 2 是本申请实施例提供的一种用于芯片版图编码器的训练方法的流程图；

图 3 是本申请实施例提供的一种对样本芯片版图进行几何变换的示意图；

图 4 是本申请实施例提供的一种 Unet 框架的示意图；

图 5 是本申请实施例提供的一种芯片版图的筛选方法的流程图；

图 6 是本申请实施例提供的一种芯片版图编码器以及基于该芯片版图编码器进行聚类筛选的示意图；

图 7 是本申请实施例提供的一种用于芯片版图编码器的训练装置的结构示意图；

图 8 是本申请实施例提供的一种芯片版图的筛选装置的结构示意图；

图 9 是本申请实施例提供的一种终端设备的结构示意图；

图 10 是本申请实施例提供的一种服务器的结构示意图。

具体实施方式

为使本申请的目的、技术方案和优点更加清楚，下面将结合附图对本申请实施方式作进一步地详细描述。

图 1 是本申请实施例提供的一种用于芯片版图编码器的训练方法或者芯片版图的筛选方法的实施环境示意图。如图 1 所示，该实施环境包括终端设备 101 和服务器 102。其中，本申请实施例中的用于芯片版图编码器的训练方法或者芯片版图的筛选方法可以由终端设备 101 执行，也可以由服务器 102 执行，或者由终端设备 101 和服务器 102 共同执行。

终端设备 101 可以是智能手机、游戏主机、台式计算机、平板电脑、膝上型便携计算机、智能电视、智能车载设备、智能语音交互设备、智能家电等。服务器 102 可以为一台服务器，或者为多台服务器组成的服务器集群，或者为云计算平台和虚拟化中心中的任意一种，本申请实施例对此不加以限定。服务器 102 可以与终端设备 101 通过有线网络或无线网络进行通信连接。服务器 102 可以具有数据处理、数据存储以及数据收发等功能，在本申请实施例中不加以限定。终端设备 101 和服务器 102 的数量不受限制，可以是一个或多个。

本申请实施例提供的用于芯片版图编码器的训练方法或者芯片版图的筛选方法可以基于人工智能技术实现。人工智能（Artificial Intelligence，AI）是利用数字计算机或者数字计算机控制的机器模拟、延伸和扩展人的智能，感知环境、获取知识并使用知识获得最佳结果的理论、方法、技术及应用系统。换句话说，人工智能是计算机科学的一个综合技术，它企图了解智能的实质，并生产出一种新的能以人类智能相似的方式做出反应的智能机器。人工智能也就是研究各种智能机器的设计原理与实现方法，使机器具有感知、推理与决策的功能。

人工智能技术是一门综合学科，涉及领域广泛，既有硬件层面的技术也有软件层面的技术。人工智能基础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、大数据处理技术、操作/交互系统、机电一体化等技术。人工智能软件技术主要包括计算机视觉技术、语音处理技术、自然语言处理技术以及机器学习/深度学习、自动驾驶、智慧交通等几大方向。

在集成电路技术领域中，集成电路版图是一种常见的版图，集成电路版图也称芯片版图。一般情况下，需要对芯片版图进行光源掩膜优化，以得到优化后的掩膜版图，使得调用光刻机将优化后的掩膜版图曝光在晶圆片上得到的成像版图与芯片版图之间的误差小，从而提高芯片性能，提升芯片良率。

在对芯片版图进行光源掩膜优化时，需要提取芯片版图的版图特征。因此，如何训练得到芯片版图编码器，以利用芯片版图编码器提取芯片版图的版图特征成为一个亟需解决的问题。

本申请实施例提供了一种用于芯片版图编码器的训练方法，该方法可应用于上述实施环境中，可以解决训练得到芯片版图编码器。以图 2 所示的本申请实施例提供的一种用于芯片版图编码器的训练方法的流程图为例，为便于描述，将执行本申请实施例中的用于芯片版图编码器的训练方法的终端设备 101 或者服务器 102 称为电子设备，该方法可以由电子设备来执行。如图 2 所示，该方法包括如下步骤。

步骤 201，获取样本芯片版图和初始编码器。

样本芯片版图是集成电路（Integrated Circuit，IC）版图或者 IC 版图中的子区域。通常情况下，样本芯片版图的数量为多个，每一个样本芯片版图的处理方式相同。

本申请实施例中，可以获取任意的 IC 版图，将该 IC 版图作为样本芯片版图。或者，可以将 IC 版图均匀的划分为若干个子区域，得到各个子区域，将任意的子区域作为样本芯片版图。可选地，可以人工从各个子区域中筛选出部分子区域（如筛选出不具有坏点的子区域，坏点是 IC 版图中不符合工艺窗口要求的部分），将这些子区域作为样本芯片版图，或者，对各个子区域进行分类，得到各个子区域对应的芯片版图类型，从各个芯片版图类型对应的子区域中均匀采样，以筛选得到样本芯片版图。其中，芯片版图类型包括但不限于接触孔类型、逻辑版图类型、密集线类型等。

可以理解的是，对 IC 版图进行均匀划分，可能会出现坏点正好处于划分线上的情况。在这种情况下，对 IC 版图进行划分时，会将一个坏点划分成至少两个部分，且不同部分位于不同子区域中。由于子区域包括坏点的一部分，是不完整的坏点，导致从子区域中难以识别出坏点，从而导致容易将包括坏点的子区域作为样本芯片版图，降低了样本芯片版图的质量。

在一种可能的实现方式中，可以先识别出 IC 版图上的各个坏点，本申请实施例不对坏点的识别方式做限定，示例性地，通过人工识别的方式识别 IC 版图上的各个坏点，或者，通过光刻仿真的方法识别 IC 版图上的各个坏点。接着，利用图像分割模型对 IC 版图进行智能划分，得到各个子区域。通过智能划分使坏点尽可能分布在子区域的中心，而不是分布在子区域的边缘，使得从各个子区域中筛选样本芯片版图时，可以快速准确地从子区域中识别出坏点，避免筛选出的样本芯片版图是具有坏点的子区域，提高样本芯片版图的质量，从而使得后续通过样本芯片版图训练芯片版图编码器时能够提高芯片版图编码器的训练效果。

本申请实施例中，样本芯片版图是图像数据，属于像素化的表征方式。通过对 IC 版图划分并筛选出样本芯片版图，使得接触孔类型、逻辑版图类型、密集线类型等各种芯片版图类型均可以使用像素表征。其中，样本芯片版图的像素尺寸需要大于或者等于光源波长/（4×数值孔径），也就是说，样本芯片版图的像素尺寸需要满足光刻机的最低分辨率。

样本芯片版图用于对初始编码器进行训练得到芯片版图编码器。本申请实施例不对初始编码器的结构和大小等做限定，示例性地，初始编码器为基于 U 网络的深度学习框架中的编码器，简称 Unet 框架中的编码器。Unet 框架中的编码器可以有效地对高维数据（比如图像数据）进行降维处理，从而提取出数据特征。Unet 框架是一种卷积自编码器，主要基于两个部分构成。第一个部分是编码器，由至少两个卷积块构成，该编码器可以作为本申请实施例中的初始编码器；第二个部分是解码器，由至少两个反卷积块构成，且卷积块的数量与反卷积块的数量相同。

步骤 202，对样本芯片版图进行几何变换，得到至少一个参考芯片版图。

几何变换包括翻转变换、旋转变换等至少一种变换。可以对样本芯片版图进行翻转变换、旋转变换等至少一种变换，得到参考芯片版图。其中，翻转变换就是下文提及的镜面翻转处理，旋转变换就是下文提及的旋转处理。

示例性地，步骤 202 包括：对样本芯片版图进行镜面翻转处理，得到对称芯片版图，将对称芯片版图作为至少一个参考芯片版图。可以理解的是，此处的对称芯片版图也可以称为镜面翻转处理后的样本芯片版图，简称翻转后的芯片版图。也就是说，对样本芯片版图进行镜面翻转，得到翻转后的芯片版图，至少一个参考芯片版图包括翻转后的芯片版图。下文描述有镜面翻转处理的实现方式，在此暂不赘述。

示例性地，步骤 202 包括：对样本芯片版图进行旋转处理，得到旋转芯片版图，将旋转芯片版图作为至少一个参考芯片版图。可以理解的是，此处的旋转芯片版图也可以称为旋转处理后的样本芯片版图，简称旋转后的芯片版图。也就是说，对样本芯片版图进行旋转，得到旋转后的芯片版图，至少一个参考芯片版图包括旋转后的芯片版图。下文描述有旋转处理的实现方式，在此暂不赘述。

在一种可能的实现方式中，步骤 202 包括：对样本芯片版图进行镜面翻转处理，得到对称芯片版图；对样本芯片版图和对称芯片版图进行旋转处理，得到旋转芯片版图；将对称芯片版图和旋转芯片版图作为至少一个参考芯片版图。可以理解的是，此处的对称芯片版图也可以称为镜面翻转处理后的样本芯片版图，简称翻转后的芯片版图。此处的旋转芯片版图也可以称为旋转处理后的样本芯片版图和旋转处理后的对称芯片版图，简称旋转后的芯片版图。也就是说，对样本芯片版图进行镜面翻转，得到翻转后的芯片版图；对样本芯片版图和翻转后的芯片版图进行旋转，得到旋转后的芯片版图，至少一个参考芯片版图包括翻转后的芯片版图和旋转后的芯片版图。

本申请实施例中，可以根据镜面信息对样本芯片版图进行镜面翻转处理，以实现对样本芯片版图进行翻转变换，得到对称芯片版图，并将对称芯片版图作为参考芯片版图。镜面信息可以表征镜面的位置和朝向等，其中，镜面的朝向与镜面的法线方向相同，镜面的位置由镜面中参考点（如中心点）的位置来表征。当镜面朝向水平方向时，表明镜面的法线方向为水平方向，这种情况下，可以实现对样本芯片版图进行左或右的镜面翻转处理；当镜面朝向垂直方向时，表明镜面的法线方向为垂直方向，这种情况下，可以实现对样本芯片版图进行上或下的镜面翻转处理。

可以理解的是，镜面信息为至少一个。对于任一个镜面信息，根据该镜面信息对样本芯片版图进行镜面翻转处理，得到一个对称芯片版图，并将该对称芯片版图作为参考芯片版图。通过这种方式，可以实现根据不同的镜面信息，确定不同的参考芯片版图，使得参考芯片版图的数量为至少一个。

可以按照旋转参数对样本芯片版图进行至少一次旋转处理，以实现对样本芯片版图进行旋转变换，得到各次旋转处理后得到的旋转芯片版图。也就是说，先按照旋转参数对样本芯片版图进行第一次旋转处理，将第一次旋转处理后得到的样本芯片版图作为第一次旋转处理后得到的旋转芯片版图。然后，按照旋转参数对第一次旋转处理后得到的旋转芯片版图进行第二次旋转处理，得到第二次旋转处理后得到的旋转芯片版图。接着，按照旋转参数对第二次旋转处理后得到的旋转芯片版图进行第三次旋转处理，得到第三次旋转处理后得到的旋转芯片版图。以此类推，实现对样本芯片版图进行至少一次旋转处理。

同样地，可以按照旋转参数对对称芯片版图进行至少一次旋转处理，以实现对对称芯片版图进行旋转变换，得到各次旋转处理后得到的旋转芯片版图。可以将各次旋转处理后得到的旋转芯片版图作为各个参考芯片版图。需要说明的是，旋转参数用于指示每次旋转处理需要转过的角度。例如，旋转参数为每次旋转处理需要转过目标角度，这种情况下，每目标角度旋转一次样本芯片版图，可以得到一张旋转芯片版图，每目标角度旋转一次对称芯片版图，也可以得到一张旋转芯片版图。

请参见图 3，图 3 是本申请实施例提供的一种对样本芯片版图进行几何变换的示意图。本申请实施例中，一方面可以每 45 度旋转一次样本芯片版图，得到两张旋转芯片版图，另一方面，可以先对样本芯片版图进行右的镜面翻转处理，得到对称芯片版图，再每 45 度旋转一次对称芯片版图，得到两张旋转芯片版图。

本申请实施例中，可以将对称芯片版图和旋转芯片版图作为参考芯片版图。可以理解的是，可以先对样本芯片版图进行旋转处理得到旋转芯片版图，再对样本芯片版图和旋转芯片版图进行镜面翻转处理，得到对称芯片版图，接着将对称芯片版图和旋转芯片版图作为参考芯片版图，且参考芯片版图的数量为至少一个。可以理解的是，此处的旋转芯片版图也可以称为旋转处理后的样本芯片版图，简称旋转后的芯片版图。此处的对称芯片版图也可以称为

镜面翻转处理后的样本芯片版图和镜面翻转处理后的旋转芯片版图，简称旋转后的芯片版图。也就是说，对样本芯片版图进行旋转，得到旋转后的芯片版图；对样本芯片版图和旋转后的芯片版图进行镜面翻转，得到翻转后的芯片版图，至少一个参考芯片版图包括旋转后的芯片版图和翻转后的芯片版图。在上文有描述旋转处理、镜面翻转处理的实现方式，在此暂不赘述。

本申请实施例中，对芯片版图（如上面提及的样本芯片版图、对称芯片版图、旋转芯片版图等）进行镜面翻转处理，指的是：根据镜面信息对芯片版图中各个像素点的位置信息进行调整，使得任一个像素点调整后的位置信息和该像素点的位置信息满足镜面对称。镜面翻转处理后得到的对称芯片版图包括各个像素点调整后的位置信息。

基于同样的原理，对芯片版图进行旋转处理，指的是：根据旋转参数对芯片版图中各个像素点的位置信息进行调整，使得任一个像素点调整后的位置信息和该像素点的位置信息满足旋转变换。旋转处理后得到的旋转芯片版图包括各个像素点调整后的位置信息。

由此可知，对样本芯片版图进行几何变换，就是调整样本芯片版图中各个像素点的位置信息，参考芯片版图包括各个像素点调整后的位置信息。可以理解的是，样本芯片版图中像素点之间的相邻关系和参考芯片版图中像素点之间的相邻关系相同，像素点之间的相邻关系指的是每两个像素点是否相邻的关系。也就是说，如果样本芯片版图中的某两个像素点相邻，则参考芯片版图中的这两个像素点也相邻。

本申请实施例中，可以对样本芯片版图进行镜面翻转处理、旋转处理，得到参考芯片版图，也可以对对称芯片版图进行旋转处理，得到参考芯片版图，还可以对旋转芯片版图进行镜面翻转处理，得到参考芯片版图。通过多种方式确定参考芯片版图，提高了参考芯片版图的多样性，使得通过参考芯片版图训练得到的芯片版图编码器具有较强的泛化能力，能够提高版图特征的表征能力，从而在基于版图特征对芯片版图进行聚类处理时能提高聚类结果的准确性。

步骤 203，使用初始编码器提取样本芯片版图的样本版图特征和各个参考芯片版图的参考版图特征。

可选地，样本版图特征可以称为样本芯片版图的版图特征，参考版图特征可以称为参考芯片版图的版图特征。使用初始编码器提取版图特征相当于通过初始编码器提取版图特征。这种情况下，步骤 203 可以表述为：通过初始编码器提取样本芯片版图的版图特征和各个参考芯片版图的版图特征。

可以将样本芯片版图输入初始编码器，由初始编码器对样本芯片版图进行特征提取，得到样本芯片版图的版图特征。同样地，针对任一个参考芯片版图，可以将该参考芯片版图输入初始编码器，由初始编码器对该参考芯片版图进行特征提取，得到该参考芯片版图的版图特征。下面着重介绍初始编码器对样本芯片版图进行特征提取的方式，而初始编码器对参考芯片版图进行特征提取的方式与初始编码器对样本芯片版图进行特征提取的方式相类似，下面不再赘述。

在一种可能的实现方式中，步骤 203 中的“使用初始编码器提取样本芯片版图的版图特征”包括：通过初始编码器（即使用初始编码器）对样本芯片版图进行多次下采样处理，得到各个下采样特征。各个下采样特征是各次下采样处理得到的下采样特征，样本芯片版图的版图特征也就是样本版图特征为最后一次下采样处理得到的下采样特征。

本申请实施例中，初始编码器包括至少两个卷积块，下一个卷积块的输入是上一个卷积块的输出。基于此，将样本芯片版图输入初始编码器，通过至少两个卷积块进行多次下采样处理，每一个卷积块的输出为一个下采样特征。可选地，每个卷积块的输出与该卷积块的输入相比，特征尺寸减少一半，通道数增大一倍。初始编码器最终的输出可以是一维特征矢量，例如，该一维特征矢量的维度为（1，1，1024）。该一维特征矢量即为样本芯片版图的版图特征。

在示例性实施例中，“使用初始编码器对样本芯片版图进行多次下采样处理，得到各个

下采样特征”，包括：使用初始编码器对样本芯片版图进行第一次下采样处理，得到第一个下采样特征；对于任一次下采样处理得到的任一个下采样特征，使用初始编码器对任一个下采样特征进行下一次下采样处理，得到下一次下采样处理得到的下采样特征。

也就是说，将样本芯片版图输入初始编码器后，初始编码器可以将样本芯片版图映射为初始特征，由至少两个卷积块中的第一个卷积块对样本芯片版图的初始特征进行下采样处理，得到第一个下采样特征。接着，由至少两个卷积块中的第二个卷积块对第一个下采样特征进行下采样处理，得到第二个下采样特征，以此类推。直至得到最后一个下采样特征，该最后一个下采样特征为样本芯片版图的版图特征。

也就是说，将样本芯片版图输入初始编码器后，由至少两个卷积块中的第一个卷积块对样本芯片版图映射成的初始特征进行下采样处理，得到第一个下采样特征。对于第 i 个卷积块（i 为大于 1 的正整数），由第 i 个卷积块对第 i-1 个下采样特征进行下采样处理，得到第 i 个下采样特征。假设共有 M 个卷积块，则第 M 个卷积块得到的第 M 个下采样特征为样本芯片版图的版图特征。

通过对样本芯片版图进行多次下采样处理，实现了不断地降低下采样特征的维度，提高下采样特征的表征能力，即提高样本芯片版图的版图特征的表征能力。

可选地，一个卷积块包括一个卷积层，卷积层由至少一个 3×3 的过滤器组成，且上一个卷积块包括的过滤器的数量少于下一个卷积块包括的过滤器的数量。其中，卷积层用于对样本芯片版图的初始特征或者下采样特征进行卷积处理，得到卷积特征。若卷积块仅包含卷积层，则该卷积处理相当于下采样处理，该卷积特征相当于下采样特征。

可选地，一个卷积块包括串联的卷积层和批归一化层。批归一化层用于对卷积特征进行批归一化处理，得到批归一化特征。若卷积块仅包含卷积层和批归一化层，则下采样处理包括卷积处理和批归一化处理，批归一化特征相当于下采样特征。

可选地，一个卷积块包括串联的卷积层、批归一化层和激活层。激活层用于对批归一化特征进行激活处理，得到激活特征。若卷积块仅包含卷积层、批归一化层和激活层，则下采样处理包括卷积处理、批归一化处理和激活处理，激活特征相当于下采样特征。其中，激活层可以采用线性整流函数（Rectified Linear Unit，ReLU），线性整流函数又被称为修正线性单元。

请参见图 4，图 4 是本申请实施例提供的一种 Unet 框架的示意图。Unet 框架包括编码器，该编码器为本申请实施例中的初始编码器。该编码器包括 8 个卷积块，每一个卷积块包括串联的卷积层、批归一化层和激活层，且激活层采用修正线性单元。这八个卷积块依次包括 8、16、32、64、128、256、512、1024 个 3×3 的过滤器。

在图 4 中示出了图 3 所示的样本芯片版图、对称芯片版图和旋转芯片版图，这些版图均会输入 Unet 框架，由 Unet 框架包括的编码器提取这些版图的版图特征。

以样本芯片版图为例，将样本芯片版图输入 Unet 框架，由 Unet 框架包括的编码器对样本芯片版图进行八次下采样处理，得到八个下采样特征。假设样本芯片版图的尺寸为（256，256，1），则这八个下采样特征的尺寸分别为（128，128，8）、（64，64，16）、（32，32，32）、（16，16，64）、（8，8，128）、（4，4，256）、（1，1，1024）。其中，最后一个下采样特征为样本芯片版图的版图特征，因此，样本芯片版图的版图特征的尺寸为（1，1，1024），是一个一维特征矢量。

步骤 204，基于样本版图特征和参考版图特征，对初始编码器进行训练，得到训练后的编码器。

可以理解的是，训练后的编码器也可以称为芯片版图编码器。基于此，步骤 204 可以表述为：基于样本芯片版图的版图特征和各个参考芯片版图的版图特征，对初始编码器进行训练，得到芯片版图编码器。

可选地，训练后的编码器用于生成芯片版图和/或提取芯片版图的特征。

本申请实施例中，可以基于样本芯片版图的版图特征和各个参考芯片版图的版图特征，

确定第一损失。可选地，针对任一个参考芯片版图，可以计算该参考芯片版图的版图特征和样本芯片版图的版图特征之间的距离，将该距离称为该参考芯片版图对应的距离，将各个参考芯片版图对应的距离进行加权计算（例如加权求和、加权求平均等），将加权计算结果作为第一损失。或者，可以对各个参考芯片版图的版图特征进行求平均计算，得到平均版图特征，计算平均版图特征和样本芯片版图的版图特征之间的距离，将该距离作为第一损失。

示例性地，可以采用如下所示的公式（1），对各个参考芯片版图的版图特征进行求平均计算，得到平均版图特征。

$$\overline{\mathrm{Encoder}\left(\widehat{\mathrm{T}}\mathrm{x}_{\mathrm{i}}\right)}=\frac{1}{K}\sum_{\mathrm{n}}^{\mathrm{K}}\mathrm{Encoder}(\widehat{\mathrm{T}_{\mathrm{n}}}\mathrm{x}_{\mathrm{i}}) \quad \text{公式（1）}$$

其中，x_{i}表征第 i 个样本芯片版图。$\widehat{\mathrm{T}_{\mathrm{n}}}\mathrm{x}_{\mathrm{i}}$表征对第 i 个样本芯片版图进行第 n 次几何变换后得到的参考芯片版图。K 表征几何变换的总次数，也表征参考芯片版图的数量。对于一个样本芯片版图，对该样本芯片版图、旋转芯片版图进行一次镜面翻转处理，均相当于对该样本芯片版图进行一次几何变换，得到对称芯片版图。对样本芯片版图、对称芯片版图、旋转芯片版图旋转目标角度，均相当于对样本芯片版图进行一次几何变换，得到旋转芯片版图。假如对一个样本芯片版图进行一次左和右的镜面翻转处理，得到两个对称芯片版图，再对这两个对称芯片版图每 45 度旋转一次，则旋转次数总共有 14 次，可以得到 14 个旋转芯片版图。由于对称芯片版图和旋转芯片版图均作为参考芯片版图，因此，参考芯片版图的数量为 16，则 K 的取值为 16。

此外，公式（1）中的 Encoder 表征特征提取操作，Σ 表征求和函数的符号。因此，$\overline{\mathrm{Encoder}\left(\widehat{\mathrm{T}}\mathrm{x}_{\mathrm{i}}\right)}$表征平均版图特征，$\mathrm{Encoder}(\widehat{\mathrm{T}_{\mathrm{n}}}\mathrm{x}_{\mathrm{i}})$表征对第 i 个样本芯片版图进行第 n 次几何变换后得到的参考芯片版图的版图特征。

在公式（1）的基础上，可以采用如下所示的公式（2）计算平均版图特征和样本芯片版图的版图特征之间的距离，得到第一损失。

$$\mathrm{Loss}_{\mathrm{rotation}}=\sum_{\mathrm{i}=0}^{\mathrm{N}}\left|\mathrm{Encoder}(\mathrm{x}_{\mathrm{i}})-\overline{\mathrm{Encoder}\left(\widehat{\mathrm{T}}\mathrm{x}_{\mathrm{i}}\right)}\right|^{2} \quad \text{公式（2）}$$

其中，$\mathrm{Loss}_{\mathrm{rotation}}$表征第一损失。$\mathrm{Encoder}(\mathrm{x}_{\mathrm{i}})$表征第 i 个样本芯片版图的版图特征。样本芯片版图的总数量为 N+1。Σ 表征求和函数的符号。

在计算出第一损失之后，可以将第一损失作为初始编码器的损失，通过梯度下降的方式，利用初始编码器的损失对初始编码器进行训练，得到训练后的初始编码器。对初始编码器进行训练就是调整初始编码器中神经元的权重和偏置，使得初始编码器的损失下降。若训练后的初始编码器满足训练结束条件，则将训练后的初始编码器作为芯片版图编码器（即训练后的编码器）。若训练后的初始编码器不满足训练结束条件，则将训练后的初始编码器作为下一次训练的初始编码器，并按照本申请实施例提供的用于芯片版图编码器的训练方法对该初始编码器进行下一次训练，直至得到芯片版图编码器。

本申请实施例不对满足训练结束条件做限定。示例性地，满足训练结束条件为训练次数达到目标次数（如 500 次）。或者，通过梯度下降的方式，利用初始编码器的损失对初始编码器进行训练时，需要确定初始编码器的损失的梯度，基于该梯度对初始编码器的参数进行调整，以使下一次训练确定的初始编码器的损失的梯度小于本次训练确定的初始编码器的损失的梯度。当下一次训练确定的初始编码器的损失的梯度与本次训练确定的初始编码器的损失的梯度之间的梯度差值小于差值阈值时，确定满足训练结束条件。也就是说，将本次训练确定的初始编码器的损失的梯度减去下一次训练确定的初始编码器的损失的梯度，得到梯度差值，如果梯度差值小于差值阈值，则满足训练结束条件。此时，满足训练结束条件相当于初始编码器的损失的梯度的下降范围在差值阈值的范围内。本申请实施例不对差值阈值的确定方式做限定，示例性地，差值阈值是根据人工经验确定的，或者，人工设定数据范围，差值阈值是随机生成的位于该数据范围内的数据。

需要说明的是，由于第一损失是基于样本芯片版图的版图特征和各个参考芯片版图的版图特征确定的，因此，利用第一损失训练初始编码器，可以使初始编码器提取到的样本芯片

版图的版图特征和参考芯片版图的版图特征相互靠近。由于参考芯片版图是对样本芯片版图进行几何变换得到的，因此，参考芯片版图和样本芯片版图属于同一类型的芯片版图。在训练初始编码器的过程中，可以使初始编码器提取的版图特征不断地倾向于表征芯片版图类别，从而使最终训练得到的芯片版图编码器侧重于提取能表征芯片版图类别的版图特征。芯片版图编码器提取的版图特征可以表征几何变换前的芯片版图和几何变换后的芯片版图，使得芯片版图编码器提取的版图特征具有几何变换前后不变的特性。因此，第一损失是可以表征几何变换前后不变的损失。

在一种可能的实现方式中，该用于芯片版图编码器的训练方法还包括步骤 205。其中，步骤 205 在步骤 203 之后执行。

步骤 205，基于样本芯片版图的版图特征，确定重构芯片版图。即基于样本版图特征，确定重构芯片版图，重构芯片版图是基于样本芯片版图的版图特征重构出的芯片版图。

本申请实施例中，可以将样本芯片版图的版图特征输入初始解码器，由初始解码器基于样本芯片版图的版图特征重构出新的芯片版图，得到重构芯片版图。本申请实施例不对初始解码器的结构和大小等做限定，示例性地，初始解码器 Unet 框架中的解码器。

在一种可能的实现方式中，步骤 205 包括：对样本芯片版图的版图特征进行多次上采样处理，得到各个上采样特征，各个上采样特征是各次上采样处理得到的上采样特征。或者说，对样本版图特征进行多次上采样处理，得到各次上采样处理得到的上采样特征。重构芯片版图为基于最后一次上采样处理得到的上采样特征得到的芯片版图。

本申请实施例中，初始解码器包括至少两个反卷积块，上一个反卷积块的输出为下一个反卷积块的输入。可选地，每一个反卷积块的输出与该反卷积块的输入相比，特征尺寸增大一倍，通道数降低一半。初始解码器最终输出的特征尺寸和初始编码器输入的样本芯片版图的尺寸一致。

也就是说，将样本芯片版图的版图特征输入初始解码器后，由至少两个反卷积块中的第一个反卷积块对样本芯片版图的版图特征进行上采样处理，得到第一个上采样特征。接着，由至少两个反卷积块中的第二个反卷积块对第一个上采样特征进行上采样处理，得到第二个上采样特征，以此类推。直至得到最后一个上采样特征，该最后一个上采样特征可以映射为一张芯片版图，该芯片版图为重构芯片版图。

也就是说，将样本芯片版图的版图特征输入初始解码器后，由至少两个反卷积块中的第一个反卷积块对样本芯片版图的版图特征进行上采样处理，得到第一个上采样特征。对于第 i 个反卷积块（i 为大于 1 的正整数），由第 i 个反卷积块对第 i-1 个上采样特征进行上采样处理，得到第 i 个上采样特征。假设共有 M 个反卷积块，则第 M 个反卷积块得到的第 M 个上采样特征可以映射为重构芯片版图。

可选地，一个反卷积块包括一个反卷积层，反卷积层由至少一个 3×3 的过滤器组成，且上一个反卷积块包括的过滤器的数量多于下一个反卷积块包括的过滤器的数量。其中，反卷积层用于对样本芯片版图的版图特征或者上采样特征进行反卷积处理，得到反卷积特征。若反卷积块仅包含反卷积层，则该反卷积处理相当于上采样处理，该反卷积特征相当于上采样特征。

可选地，一个反卷积块包括串联的反卷积层和批归一化层。批归一化层用于对反卷积特征进行批归一化处理，得到批归一化特征。若反卷积块仅包含反卷积层和批归一化层，则上采样处理包括反卷积处理和批归一化处理，批归一化特征相当于上采样特征。

可选地，一个反卷积块包括串联的反卷积层、批归一化层和激活层。激活层用于对批归一化特征进行激活处理，得到激活特征。若反卷积块仅包含反卷积层、批归一化层和激活层，则上采样处理包括反卷积处理、批归一化处理和激活处理，激活特征相当于上采样特征。其中，激活层可以采用 Leaky-ReLU 激活函数或者 sigmoid 激活函数。

可选地，对样本芯片版图的版图特征进行多次上采样处理，得到各个上采样特征，包括如下所示的步骤 A1 至步骤 A4。

步骤 A1，对样本芯片版图的版图特征（即样本版图特征）进行第一次上采样处理，得到第一个上采样特征，样本芯片版图的版图特征（即样本版图特征）是对样本芯片版图进行多次下采样处理得到的。

在步骤 203 已提及对样本芯片版图进行多次下采样处理得到样本芯片版图的版图特征，在此不再赘述。在本申请实施例中，将样本芯片版图的版图特征输入初始解码器后，由至少两个反卷积块中的第一个反卷积块对样本芯片版图的版图特征进行上采样处理，得到第一个上采样特征。

步骤 A2，对于任一次上采样处理得到的任一个上采样特征，获取任一个上采样特征对应的下采样特征，对样本芯片版图进行下采样处理得到任一个上采样特征对应的下采样特征时下采样处理的次数为第一次数，对样本芯片版图的版图特征（即样本版图特征）进行上采样处理得到任一个上采样特征时上采样处理的次数为第二次数，第一次数与第二次数之和为目标次数。

在本申请实施例中，初始编码器中卷积块的数量和初始解码器中反卷积块的数量相同，因此，初始编码器进行下采样处理的总次数和初始解码器进行上采样处理的总次数相同。目标次数为初始编码器进行下采样处理的总次数。比如，初始编码器共进行 8 次下采样处理，则目标次数为 8。

对于第 i 次上采样处理得到的第 i 个上采样特征（i 为大于等于 1 的正整数），第二次数为 i。由于第一次数与第二次数之和为目标次数，因此，如果目标次数为 M，则第一次数为 M-i。可以确定出第 i 个上采样特征对应的下采样特征为：对样本芯片版图进行第 M-i 次下采样处理得到的第 M-i 个下采样特征。

例如，目标次数为 8，第 1 个上采样特征对应的下采样特征为第 7 个下采样特征，第 2 个上采样特征对应的下采样特征为第 6 个下采样特征，以此类推。

步骤 A3，将任一个上采样特征与任一个上采样特征对应的下采样特征进行拼接，得到任一个上采样特征对应的拼接特征。

可以将第 i 个上采样特征与第 M-i 个下采样特征进行拼接，得到第 i 个上采样特征对应的拼接特征。

步骤 A4，对任一个上采样特征对应的拼接特征进行上采样处理，得到任一个上采样特征的下一个上采样特征。

对第 i 个上采样特征对应的拼接特征进行上采样处理，得到第 i+1 个上采样特征。

需要说明的是，当得到最后一个上采样特征时，可以将最后一个上采样特征映射成重构芯片版图，也可以将最后一个上采样特征和第 1 个下采样特征进行拼接，得到最后一个上采样特征对应的拼接特征，将最后一个上采样特征对应的拼接特征映射成重构芯片版图。

请参见图 4，图 4 所示的 Unet 框架包括解码器，该解码器为本申请实施例中的初始解码器。该解码器包括八个反卷积块，每一个反卷积块包括串联的反卷积层、批归一化层和激活层。这八个反卷积块依次包括 1024、512、256、128、64、32、16、1 个 3×3 的过滤器，其中，前 7 个反卷积块中的激活层均为 Leaky-ReLU 激活函数，第 8 个反卷积块中的激活层为 sigmoid 激活函数。上文已描述了 Unet 框架中的编码器，在此不再赘述。

在本申请实施例中，Unet 框架中的编码器可以对样本芯片版图进行八次下采样处理，得到八个下采样特征。对第八个下采样特征进行上采样处理，得到第一个上采样特征。基于第一个上采样特征与第七个下采样特征得到第二个上采样特征。基于第二个上采样特征与第六个下采样特征得到第三个上采样特征，以此类推。直至基于第七个上采样特征与第一个下采样特征得到第八个上采样特征为止。

其中，在基于第一个上采样特征与第七个下采样特征得到第二个上采样特征时，可以将第一个上采样特征与第七个下采样特征进行拼接，得到第一个上采样特征对应的拼接特征，对第一个上采样特征对应的拼接特征进行上采样处理，得到第二个上采样特征。第三个上采样特征至第八个上采样特征的确定方式与第二个上采样特征的确定方式相类似，在此不再赘

述。

假设样本芯片版图的版图特征的尺寸为（1，1，1024），是一个一维特征矢量，则八个上采样特征的尺寸依次为（2，2，1024）、（4，4，512）、（8，8，256）、（16，16，128）、（32，32，64）、（64，64，32）、（128，128，16）、（256，256，1）。基于第八个上采样特征得到的重构芯片版图的尺寸为（256，256，1），与样本芯片版图的尺寸一致。

本申请实施例中，基于任一个上采样特征与该上采样特征对应的下采样特征，确定该上采样特征的下一个上采样特征，本质上是将初始编码器中卷积块的输出域初始解码器中反卷积块的输出拼接到一起共同作为下一个反卷积块的输入，可以提高反卷积块进行上采样处理后得到的上采样特征的表达能力，从而提高重构芯片版图的准确性，使重构芯片版图接近样本芯片版图。

可选地，初始解码器输出的重构芯片版图为一个二值掩膜版图。二值掩膜版图包括多个像素点，每一个像素点的值表征该像素点透光或者不透光。可选地，如果一个像素点的值为第一值，表明该像素点透光；如果一个像素点的值为第二值，表明该像素点不透光。第一值为任意的数值，例如，第一值为 0。第二值为与第一值不同的任意数值，例如，第二值为 1。也就是说，如果第一值为 0，第二值为 1，则当二值掩膜版图上一个像素点的值为 1 时，表征该像素点不透光；当二值掩膜版图上一个像素点的值为 0 时，表征该像素点透光。

在可以得到重构芯片版图的情况下，步骤 204 包括：基于样本芯片版图、重构芯片版图、样本芯片版图的版图特征（即样本版图特征）和各个参考芯片版图的版图特征（即参考版图特征），对初始编码器进行训练，得到芯片版图编码器（即训练后的编码器）。

可以基于样本芯片版图、重构芯片版图、样本芯片版图的版图特征和各个参考芯片版图的版图特征，确定初始编码器的损失。利用初始编码器的损失对初始编码器进行训练，得到芯片版图编码器。

可选地，基于样本芯片版图、重构芯片版图、样本芯片版图的版图特征和各个参考芯片版图的版图特征，对初始编码器进行训练，得到芯片版图编码器，包括如下所示的步骤 B1 至步骤 B3。

步骤 B1，基于样本芯片版图的版图特征（即样本版图特征）和各个参考芯片版图的版图特征（即参考版图特征），确定第一损失。其中，在步骤 204 已描述了第一损失的确定过程，在此不再赘述。

步骤 B2，基于样本芯片版图和重构芯片版图确定第二损失。

本申请实施例中，可以在像素层面，对比样本芯片版图和重构芯片版图，得到样本芯片版图和重构芯片版图之间的距离，将该距离作为第二损失。

可选地，采用如下所示的公式（3）确定第二损失。

$$Loss_{rebuild} = \sum_{i=0}^{N} |x_i - x_{i(pred)}|^2 \qquad \text{公式（3）}$$

其中，$Loss_{rebuild}$表征第二损失。x_i表征第 i 个样本芯片版图。$x_{i(pred)}$表征第 i 个样本芯片版图对应的重构芯片版图。样本芯片版图的总数量为 N+1。Σ 表征求和函数的符号。

步骤 B3，基于第一损失和第二损失对初始编码器进行训练，得到芯片版图编码器（即训练后的编码器）。

本申请实施例中，可以对第一损失和第二损失进行求和、加权求和、求平均、加权求平均等运算处理，将运算处理结果确定为初始编码器的损失，例如，初始编码器的损失为$Loss = Loss_{rebuild} + Loss_{rotation}$。利用初始编码器的损失对初始编码器进行训练，得到芯片版图编码器。

由于第二损失是基于样本芯片版图和重构芯片版图确定的，因此，利用第二损失训练初始编码器，使得通过初始编码器提取到样本芯片版图的版图特征之后，通过初始解码器可以基于该样本芯片版图的版图特征重构出越来越接近样本芯片版图的重构芯片版图，从而保证样本芯片版图的版图特征可以准确的表征出样本芯片版图，提高样本芯片版图的版图特征的准确性。

需要说明的是，在利用初始编码器的损失训练初始编码器的过程中，还可以利用初始编码器的损失同步训练初始解码器，以提高训练效果。在应用时，还可以基于参考芯片版图的版图特征，确定重建芯片版图，重建芯片版图是基于参考芯片版图的版图特征重构出的芯片版图；利用参考芯片版图和重建芯片版图确定第三损失；利用第一损失、第二损失和第三损失中的至少一项确定初始编码器的损失，以利用初始编码器的损失对初始编码器进行训练，得到芯片版图编码器。其中，重建芯片版图的确定方式与重构芯片版图的确定方式相类似，第三损失的确定方式与第二损失的确定方式相类似，在此不再赘述。

需要说明的是，本申请所涉及的信息（包括但不限于用户设备信息、用户个人信息等）、数据（包括但不限于用于分析的数据、存储的数据、展示的数据等）以及信号，均为经用户授权或者经过各方充分授权的，且相关数据的收集、使用和处理需要遵守相关国家和地区的相关法律法规和标准。例如，本申请中涉及到的样本芯片版图等都是在充分授权的情况下获取的。

上述方法对样本芯片版图进行几何变换得到参考芯片版图，利用样本芯片版图的版图特征和参考芯片版图的版图特征训练初始编码器得到芯片版图编码器，使得针对几何变换前后的芯片版图，芯片版图编码器能够输出相近的版图特征。由于几何变换前后的芯片版图属于同一种芯片版图类型，因此，芯片版图编码器侧重于提取能区分芯片版图类型的版图特征，有利于后续基于版图特征对芯片版图进行聚类处理和筛选处理时，提高聚类结果的准确性，降低筛选结果的冗余性。

本申请实施例还提供了一种芯片版图的筛选方法，该方法可应用于上述实施环境中，可以从多个目标芯片版图中准确地筛选出关键芯片版图。以图 5 所示的本申请实施例提供的一种芯片版图的筛选方法的流程图为例，为便于描述，将执行本申请实施例中的芯片版图的筛选方法的终端设备 101 或者服务器 102 称为电子设备，该方法可以由电子设备来执行。如图 5 所示，该方法包括如下步骤。

步骤 501，获取多个目标芯片版图和训练后的编码器。

训练后的编码器也可以称为芯片版图编码器。步骤 501 可以表述为：获取多个目标芯片版图和芯片版图编码器。

本申请实施例中，任一个目标芯片版图是IC版图或者对IC版图进行划分得到的子区域，目标芯片版图的确定方式与样本芯片版图的确定方式相同，可以见步骤 201 的描述，在此不再赘述。芯片版图编码器是按照与图 2 相关的用于芯片版图编码器的训练方法训练得到的，可以见步骤 201 至步骤 205 的相关描述，在此不再赘述。

步骤 502，使用训练后的编码器提取各个目标芯片版图的目标版图特征。

目标版图特征也可以称为目标芯片版图的版图特征。步骤 502 也可以表述为：利用芯片版图编码器提取各个目标芯片版图的版图特征。

将任一个目标芯片版图输入芯片版图编码器，由芯片版图编码器对该目标芯片版图进行特征提取，得到该目标芯片版图的版图特征。其中，目标芯片版图的版图特征的确定方式与样本芯片版图的版图特征的特征方式相类似，可以见步骤 203 的相关描述，在此不再赘述。

步骤 503，基于各个目标版图特征对多个目标芯片版图进行聚类处理，得到多个目标聚类簇，任一个目标聚类簇中包括至少一个目标芯片版图。

目标芯片版图的版图特征可以表征目标芯片版图，通过对各个目标芯片版图的版图特征进行计算，可以实现将多个目标芯片版图聚在多个目标聚类簇中。

示例性地，对于任两个目标芯片版图，可以按照距离公式计算这两个目标芯片版图的版图特征之间的距离。若这两个目标芯片版图的版图特征之间的距离小于距离阈值，说明这两个目标芯片版图相似，则将这两个目标芯片版图聚在同一个初始聚类簇中；若这两个目标芯片版图的版图特征之间的距离不小于距离阈值，说明这两个目标芯片版图不相似，则将这两个目标芯片版图聚在不同的初始聚类簇中。通过这种方式，可以将多个目标芯片版图聚在多

个初始聚类簇中，每一个初始聚类簇中包括至少一个目标芯片版图。可选地，将各个初始聚类簇作为各个目标聚类簇，从而实现将多个目标芯片版图聚在多个目标聚类簇中。本申请实施例不对距离公式做限定，示例性地，距离公式是欧式空间距离的公式或者余弦距离的公式等。距离阈值的数值可以根据经验设定。

在一种可能的实现方式中，步骤 503 包括步骤 5031 至步骤 5034。

步骤 5031，获取多个第一版图特征，一个第一版图特征用于表征一个第一聚类簇的聚类中心。

本申请实施例不对电子设备获取第一版图特征的方式做限定。示例性地，可以在电子设备配置好多个第一版图特征。或者，用户可以将多个第一版图特征输入电子设备。或者，电子设备可以将多个目标芯片版图聚在多个初始聚类簇中，对于任一个初始聚类簇，对该初始聚类簇包括的各个目标芯片版图的版图特征进行求平均，得到的结果作为一个第一版图特征，这种情况下，第一版图特征的数量与初始聚类簇的数量相同。或者，按照步骤 5031 至步骤 5038 的方式，对多个目标芯片版图进行多次聚类，多次聚类中的最后一次聚类后得到多个聚类簇，将各个聚类簇包括的各个目标芯片版图的版图特征进行求平均，得到的结果作为各个第一版图特征。一个第一版图特征为一个第一聚类簇的期望聚类中心，可以表征该第一聚类簇的聚类中心。

步骤 5032，计算各个目标芯片版图的版图特征（即目标版图特征）和各个第一版图特征之间的距离。示例性地，可以按照欧式空间距离的公式或者余弦距离的公式等，计算任一个目标芯片版图的版图特征和任一个第一版图特征之间的距离。

步骤 5033，对于任一个目标芯片版图，从任一个目标芯片版图的版图特征（即目标版图特征）和各个第一版图特征之间的距离中选择最小的第一距离，将任一个目标芯片版图聚在第一距离对应的第一版图特征所对应的第一聚类簇中。

本申请实施例中，目标芯片版图的版图特征和第一版图特征之间的距离越小，表明目标芯片版图的版图特征与第一版图特征越相似，二者越能表征同一类芯片版图。基于这个原理，可以从任一个目标芯片版图的版图特征和各个第一版图特征之间的距离中选择最小的第一距离，该第一距离对应的第一版图特征是各个第一版图特征中与该目标芯片版图的版图特征最相似的第一版图特征。可以将该目标芯片版图聚在第一距离对应的第一版图特征所对应的第一聚类簇中。

通过这种方式，可以将多个目标芯片版图聚在多个第一聚类簇中。由于同一个第一聚类簇中的各个目标芯片版图对应同一类芯片版图，因此，通过这种方式可以实现对各个目标芯片版图按照芯片版图的类型进行聚类。

步骤 5034，若各个第一聚类簇满足聚类结束条件，则将各个第一聚类簇作为各个目标聚类簇。

本申请实施例不对任一个第一聚类簇满足聚类结束条件做限定。示例性地，任一个第一聚类簇满足聚类结束条件为：该第一聚类簇对应的聚类次数达到设定的次数，例如，一个第一聚类簇对应的聚类次数为 50 次，达到了设定的次数，则该第一聚类簇满足聚类结束条件。需要说明的是，下文还描述了任一个第一聚类簇满足聚类结束条件的其他实现方式，在此暂不赘述。

可选地，对于任一个第一聚类簇，可以基于该第一聚类簇中各个目标芯片版图的版图特征，确定该第一聚类簇的实际聚类中心。示例性地，按照如下所示的公式（4），对第 i 个第一聚类簇中各个目标芯片版图的版图特征进行求平均运算，得到第一聚类簇的实际聚类中心。

$$\mu_i = \frac{1}{|C_i|}\sum_{x \in C_i} x \qquad \text{公式（4）}$$

其中，μ_i表征第 i 个第一聚类簇的实际聚类中心。C_i表征第 i 个第一聚类簇中包括的目标芯片版图的数量。x 表征目标芯片版图的版图特征。Σ 为求和函数的函数符号。

在一种可能的实现方式中，对多个目标芯片版图进行第一聚类次数的聚类处理，第一聚

类次数的聚类处理中的最后一次聚类处理后得到各个第一聚类簇。此时，可以已完成对多个目标芯片版图进行第一聚类次数减 1 次的聚类处理，而第一聚类次数减 1 次的聚类处理中的最后一次聚类处理后得到各个聚类簇。也就是说，对多个目标芯片版图进行 M（M 为正整数）次聚类处理，将 M 次聚类处理中的最后一次聚类处理称为第 M 次聚类处理，则第 M 次聚类处理后得到各个第一聚类簇，第 M-1 次聚类处理后得到各个聚类簇。本申请实施例中，可以获取各个聚类簇的实际聚类中心。若各个聚类簇的实际聚类中心与各个第一聚类簇的实际聚类中心之间的误差在设定范围内，则确定各个第一聚类簇满足聚类结束条件。

在另一种可能的实现方式中，可以基于任一个第一聚类簇中各个目标芯片版图的版图特征和该第一聚类簇的实际聚类中心，确定该第一聚类簇的误差，通过确定各个第一聚类簇的误差之和，得到聚类总误差。可选地，按照如下所示的公式（5），计算任一个第一聚类簇中各个目标芯片版图的版图特征和该第一聚类簇的实际聚类中心之间的差值平方和，得到该第一聚类簇的误差，并将各个第一聚类簇的误差之和作为聚类总误差。

$$\text{Error} = \sum_{i=1}^{N} \sum_{x \in C_i} |x - \mu_i|^2 \qquad \text{公式（5）}$$

其中，Error 表征聚类总误差。N 表征第一聚类簇的数量。μ_i表征第 i 个第一聚类簇的实际聚类中心。C_i表征第 i 个第一聚类簇中包括的目标芯片版图的数量。x 表征目标芯片版图的版图特征。Σ 为求和函数的函数符号。

在一种可能的实现方式中，当各个第一聚类簇的聚类总误差均小于设定的总误差时，确定各个第一聚类簇满足聚类结束条件。或者，对多个目标芯片版图进行第一聚类次数的聚类处理，得到各个第一聚类簇。此时，可以获取对目标芯片版图进行第一聚类次数减 1 次的聚类处理后得到的各个聚类簇的聚类总误差。若各个聚类簇的聚类总误差与各个第一聚类簇的聚类总误差之间的差值在设定范围内，则确定各个第一聚类簇满足聚类结束条件。

当各个第一聚类簇满足聚类结束条件时，将各个第一聚类簇作为各个目标聚类簇。

可选地，在步骤 5033 之后还包括步骤 5035 至步骤 5038。

步骤 5035，若各个第一聚类簇不满足聚类结束条件，则对于任一个第一聚类簇，基于任一个第一聚类簇中各个目标芯片版图的版图特征，确定一个第二版图特征，一个第二版图特征用于表征一个第二聚类簇的聚类中心。

各个第一聚类簇不满足聚类结束条件，相当于多个第一聚类簇中存在不满足聚类结束条件的第一聚类簇。此时，可以基于任一个第一聚类簇中各个目标芯片版图的版图特征（即目标版图特征），确定该第一聚类簇的实际聚类中心，并将该第一聚类簇的实际聚类中心作为一个第二版图特征。这个第二版图特征为一个第二聚类簇的期望聚类中心，可以表征该第二聚类簇的聚类中心。

步骤 5036，计算各个目标芯片版图的版图特征（即目标版图特征）和各个第二版图特征之间的距离。示例性地，可以按照欧式空间距离的公式或者余弦距离的公式等，计算任一个目标芯片版图的版图特征和任一个第二版图特征之间的距离。

步骤 5037，对于任一个目标芯片版图，从任一个目标芯片版图的版图特征（即目标版图特征）和各个第二版图特征之间的距离中选择最小的第二距离，将任一个目标芯片版图聚在第二距离对应的第二版图特征所对应的第二聚类簇中。

目标芯片版图的版图特征和第二版图特征之间的距离越小，表明目标芯片版图的版图特征与第二版图特征越相似，二者越能表征同一类芯片版图。基于这个原理，可以从任一个目标芯片版图的版图特征和各个第二版图特征之间的距离中选择最小的第二距离，该第二距离对应的第二版图特征是各个第二版图特征中与该目标芯片版图的版图特征最相似的第二版图特征。可以将该目标芯片版图聚在第二距离对应的第二版图特征所对应的第二聚类簇中。

通过这种方式，可以将多个目标芯片版图聚在多个第二聚类簇中。由于同一个第二聚类簇中的各个目标芯片版图对应同一类芯片版图，因此，通过这种方式可以实现对各个目标芯片版图按照芯片版图的类型进行聚类。

步骤 5038，若各个第二聚类簇满足聚类结束条件，则将各个第二聚类簇作为各个目标聚

类簇。

当各个第二聚类簇满足聚类结束条件时，将各个第二聚类簇作为各个目标聚类簇。当各个第二聚类簇不满足聚类结束条件，则可以按照步骤 5035 至步骤 5038 的方式，对多个目标芯片版图再次进行聚类，直至得到各个目标聚类簇为止。

通过步骤 5031 至步骤 5038 的方式对多个目标芯片版图进行多次聚类，可以不断的改变聚类簇的聚类中心，使同一聚类簇中各个目标芯片版图越来越相似，而不同聚类簇中的目标芯片版图越来越不相似，直至满足聚类结束条件。满足聚类结束条件相当于聚类簇收敛，聚类簇的聚类中心不再波动。因此，通过步骤 5031 至步骤 5038 的方式对多个目标芯片版图进行多次聚类后得到的目标聚类簇的误差较小，准确性较高。

步骤 504，对于任一个目标聚类簇，从任一个目标聚类簇包括的各个目标芯片版图中筛选出关键芯片版图。

本申请实施例中，目标聚类簇中包括的各个目标芯片版图属于同一种芯片版图类型，因此，可以从任一个目标聚类簇包括的各个目标芯片版图中随机采样，得到关键芯片版图。关键芯片版图的数量为至少一个。通过对各个目标芯片版图进行聚类，得到各个目标聚类簇，并从各个目标聚类簇包括的目标芯片版图中筛选出关键芯片版图，可以避免从所有的目标芯片版图中手动选择关键芯片版图，防止人工经验、随机选择等对最终结果的干扰。

在一种可能的实现方式中，步骤 504 包括：获取任一个目标聚类簇中各个目标芯片版图的版图特征和任一个目标聚类簇的聚类中心之间的距离；基于任一个目标聚类簇中各个目标芯片版图的版图特征和任一个目标聚类簇的聚类中心之间的距离，对任一个目标聚类簇包括的各个目标芯片版图进行均匀采样，得到多个关键芯片版图。

本申请实施例中，任一个目标聚类簇的聚类中心可以为该目标聚类簇的期望聚类中心，也可以为该目标聚类簇的实际聚类中心。其中，目标聚类簇的期望聚类中心的计算方式可以见有关第二版图特征的确定内容，二者实现原理相类似。目标聚类簇的实际聚类中心的计算方式可以见有关第一聚类簇的实际聚类中心的确定内容，二者实现原理相类似。

对于任一个目标聚类簇，可以按照欧式空间距离的公式或者余弦距离的公式等，计算该目标聚类簇包括的任一个目标芯片版图的版图特征和该目标聚类簇的聚类中心之间的距离。通过这种方式，可以得到该目标聚类簇中各个目标芯片版图的版图特征和该目标聚类簇的聚类中心之间的距离。

可选地，从目标聚类簇中各个目标芯片版图的版图特征和该目标聚类簇的聚类中心之间的距离中确定最大距离，将最大距离除以采样数目，得到采样间距。从目标聚类簇中各个目标芯片版图的版图特征和该目标聚类簇的聚类中心之间的距离中筛选出距离满足采样间距的整数倍的采样距离。从该采样距离对应的目标芯片版图中采样出一个关键芯片版图。

示例性地，假设有 M 个目标聚类簇，采样数目为 L+1。对于任一个目标聚类簇，从该目标聚类簇中各个目标芯片版图的版图特征和该目标聚类簇的聚类中心之间的距离中确定最大距离，将最大距离除以 L+1，得到采样间距。从目标聚类簇中各个目标芯片版图的版图特征和该目标聚类簇的聚类中心之间的距离中筛选出距离为 0、1*采样间距、2*采样卷积、……（L+1）*采样间距的采样距离，从这些采样距离对应的目标芯片版图中共采样出 L+2 个目标芯片版图。其中，采样距离 0 对应的目标芯片版图的版图特征为目标聚类簇的聚类中心，采样距离“（L+1）*采样间距”对应的目标芯片版图的版图特征和目标聚类簇的聚类中心之间的距离为最大距离。由于一个目标聚类簇可以采样出 L+2 个目标芯片版图，因此，M 个目标聚类簇总共可以采样出 M×（L+2）个目标芯片版图。

在一种可能的实现方式中，步骤 504 之后还包括：对关键芯片版图进行光源掩膜优化，得到目标光源和关键芯片版图对应的掩膜版图；基于目标光源对其他芯片版图进行掩膜优化，得到其他芯片版图对应的掩膜版图，其他芯片版图是各个目标芯片版图中除关键芯片版图之外的目标芯片版图。

本申请实施例中，对关键芯片版图进行光源掩膜优化相当于对关键芯片版图进行光源和

掩膜的联合优化（Source and Mask Optimization，SMO），可以得到目标光源和关键芯片版图对应的掩膜版图。其中，SMO 是实现纳米（如 28 纳米甚至更小纳米）集成电路的关键分辨率增强技术。通过从目标芯片版图中筛选关键芯片版图并对关键芯片版图进行 SMO，可以使得关键芯片版图的数量小于目标芯片版图的数量，从而提高 SMO 的速度。此外，通过对目标芯片版图进行聚类，从各个聚类簇包括的目标芯片版图中筛选关键芯片版图，可以保证关键芯片版图覆盖各种芯片版图类型，且关键芯片版图的冗余现象较少，使得对关键芯片版图进行 SMO 得到的目标光源的准确性较高，提高了 SMO 的效率和效果。

SMO 属于一种光源优化（Source Optimization，SO）技术。SMO 技术通过改变光源的强度分布，可以调整入射光的强度和方向，从而使得利用光源将掩膜版图曝光在晶圆片上得到的成像版图的质量较高，有助于提高芯片良率。由于按照本申请实施例的方法对关键芯片版图进行 SMO 时，可以提高 SMO 的速度和效果，本申请实施例能够提高光刻分辨率，增大光刻工艺窗口。此外，本申请实施例中的目标芯片版图是用像素表征的版图，因此，普适性较高，可以适用于接触孔类型、逻辑版图类型、密集线类型等芯片版图类型。

在进行任一次 SMO 时，可以将任意的光源作为初始光源，调用光刻机基于该初始光源将关键芯片版图曝光在感光胶上，得到中间掩膜版图。再调用光刻机利用初始光源将中间掩膜版图曝光在晶圆片上，得到成像版图。通过对比成像版图和关键芯片版图，得到成像版图和关键芯片版图之间的误差。

可选地，若成像版图和关键芯片版图之间的误差满足优化条件，则将初始光源作为目标光源，将中间掩膜版图作为关键芯片版图对应的掩膜版图，关键芯片版图对应的掩膜版图用于通过调用光刻机利用目标光源将关键芯片版图对应的掩膜版图曝光在晶圆片上，得到关键芯片版图对应的成像版图；若成像版图和关键芯片版图之间的误差不满足优化条件，则基于成像版图和关键芯片版图之间的误差调整初始光源，得到调整后的初始光源。将调整后的初始光源作为下一次 SMO 的初始光源，基于该初始光源进行上述所示的至少一次 SMO，直至得到目标光源和关键芯片版图对应的掩膜版图为止。

可选地，满足优化条件为成像版图和关键芯片版图之间的误差在设定的误差范围内。或者，满足优化条件为成像版图和关键芯片版图之间的误差的梯度小于设定的梯度阈值，这种情况下，基于成像版图和关键芯片版图之间的误差调整初始光源时，可以基于成像版图和关键芯片版图之间的误差的梯度调整初始光源。

之后，可以调用光刻机基于该目标光源将其他芯片版图曝光在感光胶上，得到其他芯片版图对应的掩膜版图。其他芯片版图对应的掩膜版图用于通过调用光刻机利用目标光源将其他芯片版图对应的掩膜版图曝光在晶圆片上，得到其他芯片版图对应的成像版图。

可选地，若关键芯片版图和其他芯片版图是由一张芯片版图划分得到的，这种情况下，在对关键芯片版图进行光源掩膜优化得到目标光源后，可以调用光刻机基于该目标光源将这张芯片版图曝光在感光胶上，得到这张芯片版图对应的掩膜版图。这张芯片版图对应的掩膜版图用于通过调用光刻机利用目标光源将这张芯片版图对应的掩膜版图曝光在晶圆片上，得到这张芯片版图对应的成像版图。

需要说明的是，本申请所涉及的信息（包括但不限于用户设备信息、用户个人信息等）、数据（包括但不限于用于分析的数据、存储的数据、展示的数据等）以及信号，均为经用户授权或者经过各方充分授权的，且相关数据的收集、使用和处理需要遵守相关国家和地区的相关法律法规和标准。例如，本申请中涉及到的目标芯片版图等都是在充分授权的情况下获取的。

上述方法中的芯片版图编码器侧重于提取能区分芯片版图类型的版图特征。因此，利用芯片版图编码器提取各个目标芯片版图的版图特征，并基于各个目标芯片版图的版图特征对多个目标芯片版图进行聚类处理时，可以准确地将同一芯片版图类型的目标芯片版图聚在同一目标聚类簇中，将不同芯片版图类型的目标芯片版图聚在不同的目标聚类簇中，提高聚类结果的准确性。在从任一目标聚类簇包括的各个目标芯片版图中筛选关键芯片版图时，可以

降低关键芯片版图的冗余性，提高筛选质量。

上述从方法步骤的角度阐述了本申请实施例提供的用于芯片版图编码器的训练方法以及芯片版图的筛选方法，下面结合图 6 来进一步描述。图 6 是本申请实施例提供的一种芯片版图编码器以及基于该芯片版图编码器进行聚类筛选的示意图。

本申请实施例中，可以获取样本数据集，利用样本数据集按照与图 2 相关的用于芯片版图编码器的训练方法，对初始编码器进行训练得到芯片版图编码器。本申请实施例中的样本数据集包括两个开源的数据集。一个数据集包括 4877 个芯片版图，可以将这 4877 个芯片版图或者从这 4877 个芯片版图中提取出部分芯片版图作为样本芯片版图。另一个数据集包括 5394 个芯片版图，可以将这 5394 个芯片版图或者从这 5394 个芯片版图中提取出部分芯片版图作为样本芯片版图。样本芯片版图用于对初始编码器进行训练。此外，本申请实施例中的样本数据集还包括一些接触孔类型的芯片版图、逻辑版图类型的芯片版图以及密集线类型的芯片版图等。

在训练得到芯片版图编码器之后，可以利用芯片版图编码器从多个目标芯片版图中筛选出关键芯片版图。芯片版图编码器的结构可以见与图 4 相关的描述，在此不再赘述。

本申请实施例中，可以利用芯片版图编码器提取各个目标芯片版图的版图特征。接着，基于各个目标芯片版图的版图特征，对各个目标芯片版图进行聚类，得到多个目标聚类簇。任一个目标聚类簇中包括至少一个目标芯片版图。接着，从各个目标聚类簇中筛选出关键芯片版图，例如，对于任一个目标聚类簇，可以从该目标聚类簇包括的各个目标芯片版图中筛选出关键芯片版图。

由于上述芯片版图编码器侧重于提取能区分芯片版图类型的版图特征。因此，利用芯片版图编码器提取各个目标芯片版图的版图特征，并基于各个目标芯片版图的版图特征对多个目标芯片版图进行聚类处理时，可以提高聚类结果的准确性，从各个目标聚类簇中筛选出关键芯片版图时，可以降低关键芯片版图的冗余性，提高筛选质量。

图 7 所示为本申请实施例提供的一种用于芯片版图编码器的训练装置的结构示意图。如图 7 所示，该装置包括：

获取模块 701，用于获取样本芯片版图和初始编码器；

变换模块 702，用于对样本芯片版图进行几何变换，得到至少一个参考芯片版图；

提取模块 703，用于通过初始编码器提取样本芯片版图的版图特征和各个参考芯片版图的版图特征，或者说，提取模块 703，用于使用初始编码器提取样本芯片版图的样本版图特征和各个参考芯片版图的参考版图特征；

训练模块 704，用于基于样本芯片版图的版图特征（即样本版图特征）和各个参考芯片版图的版图特征（即参考版图特征），对初始编码器进行训练，得到芯片版图编码器（即训练后的编码器），芯片版图编码器用于提取目标芯片版图的版图特征。

在一种可能的实现方式中，变换模块 702，用于对样本芯片版图进行镜面翻转处理，得到对称芯片版图；对样本芯片版图和对称芯片版图进行旋转处理，得到旋转芯片版图；将对称芯片版图和旋转芯片版图作为至少一个参考芯片版图。或者说，变换模块 702，用于对样本芯片版图进行镜面翻转，得到翻转后的芯片版图；对样本芯片版图和翻转后的芯片版图进行旋转，得到旋转后的芯片版图，至少一个参考芯片版图包括翻转后的芯片版图和旋转后的芯片版图。

在一种可能的实现方式中，变换模块 702，用于对样本芯片版图进行镜面翻转处理，得到对称芯片版图，将对称芯片版图作为至少一个参考芯片版图；或者，对样本芯片版图进行旋转处理，得到旋转芯片版图，将旋转芯片版图作为至少一个参考芯片版图。或者说，变换模块 702，用于对样本芯片版图进行镜面翻转，得到翻转后的芯片版图，至少一个参考芯片版图包括翻转后的芯片版图；或者，对样本芯片版图进行旋转，得到旋转后的芯片版图，至少一

个参考芯片版图包括旋转后的芯片版图。

在一种可能的实现方式中，变换模块 702，用于对样本芯片版图进行旋转处理，得到旋转芯片版图；对样本芯片版图和旋转芯片版图进行镜面翻转处理，得到对称芯片版图；将对称芯片版图和旋转芯片版图作为至少一个参考芯片版图。或者说，变换模块 702，用于对样本芯片版图进行旋转，得到旋转后的芯片版图；对样本芯片版图和旋转后的芯片版图进行镜面翻转，得到翻转后的芯片版图，至少一个参考芯片版图包括旋转后的芯片版图和翻转后的芯片版图。

在一种可能的实现方式中，提取模块 703，用于通过（即使用）初始编码器对样本芯片版图进行多次下采样处理，得到各次下采样处理得到的下采样特征，样本芯片版图的版图特征为最后一次下采样处理得到的下采样特征。

在一种可能的实现方式中，装置还包括：

重构模块，用于基于样本芯片版图的版图特征，确定重构芯片版图，重构芯片版图是基于样本芯片版图的版图特征重构出的芯片版图；

训练模块 704，用于基于样本芯片版图、重构芯片版图、样本芯片版图的版图特征和各个参考芯片版图的版图特征，对初始编码器进行训练，得到芯片版图编码器。

在一种可能的实现方式中，重构模块，用于对样本芯片版图的版图特征进行多次上采样处理，得到各个上采样特征，重构芯片版图为基于最后一次上采样处理得到的上采样特征得到的芯片版图。

在一种可能的实现方式中，重构模块，用于对样本芯片版图的版图特征进行第一次上采样处理，得到第一个上采样特征，样本芯片版图的版图特征是对样本芯片版图进行多次下采样处理得到的；对于任一次上采样处理得到的任一个上采样特征，获取任一个上采样特征对应的下采样特征，对样本芯片版图进行下采样处理得到任一个上采样特征对应的下采样特征时下采样处理的次数为第一次数，对样本芯片版图的版图特征进行上采样处理得到任一个上采样特征时上采样处理的次数为第二次数，第一次数与第二次数之和为目标次数；将任一个上采样特征与任一个上采样特征对应的下采样特征进行拼接，得到任一个上采样特征对应的拼接特征；对任一个上采样特征对应的拼接特征进行上采样处理，得到任一个上采样特征的下一个上采样特征。

在一种可能的实现方式中，训练模块 704，用于基于样本芯片版图的版图特征和各个参考芯片版图的版图特征，确定第一损失；基于样本芯片版图和重构芯片版图确定第二损失；基于第一损失和第二损失对初始编码器进行训练，得到芯片版图编码器。

上述装置对样本芯片版图进行几何变换得到参考芯片版图，利用样本芯片版图的版图特征和参考芯片版图的版图特征训练初始编码器得到芯片版图编码器，使得针对几何变换前后的芯片版图，芯片版图编码器能够输出相近的版图特征。由于几何变换前后的芯片版图属于同一种芯片版图类型，因此，芯片版图编码器侧重于提取能区分芯片版图类型的版图特征，有利于后续基于版图特征对芯片版图进行聚类筛选，提高聚类准确性，降低筛选结果的冗余性。

应理解的是，上述图 7 提供的装置在实现其功能时，仅以上述各功能模块的划分进行举例说明，实际应用中，可以根据需要而将上述功能分配由不同的功能模块完成，即将设备的内部结构划分成不同的功能模块，以完成以上描述的全部或者部分功能。另外，上述实施例提供的装置与方法实施例属于同一构思，其具体实现过程详见方法实施例，这里不再赘述。

图 8 所示为本申请实施例提供的一种芯片版图的筛选装置的结构示意图，如图 8 所示，该装置包括：

获取模块 801，用于获取多个目标芯片版图和芯片版图编码器（即训练后的编码器），芯片版图编码器是按照上述的用于芯片版图编码器的训练方法训练得到的；

提取模块 802，用于利用芯片版图编码器提取各个目标芯片版图的版图特征（即目标版图特征）；

聚类模块 803，用于基于各个目标芯片版图的版图特征（即各个目标版图特征）对多个目标芯片版图进行聚类处理，得到多个目标聚类簇，任一个目标聚类簇中包括至少一个目标芯片版图；

筛选模块 804，用于对于任一个目标聚类簇，从任一个目标聚类簇包括的各个目标芯片版图中筛选出关键芯片版图。

在一种可能的实现方式中，聚类模块 803，用于获取多个第一版图特征，一个第一版图特征用于表征一个第一聚类簇的聚类中心；计算各个目标芯片版图的版图特征和各个第一版图特征之间的距离；对于任一个目标芯片版图，从任一个目标芯片版图的版图特征和各个第一版图特征之间的距离中选择最小的第一距离，将任一个目标芯片版图聚在第一距离对应的第一版图特征所表征的第一聚类簇中；若各个第一聚类簇满足聚类结束条件，则将各个第一聚类簇作为各个目标聚类簇。

在一种可能的实现方式中，聚类模块 803，还用于若各个第一聚类簇不满足聚类结束条件，则对于任一个第一聚类簇，基于任一个第一聚类簇中各个目标芯片版图的版图特征，确定一个第二版图特征，一个第二版图特征用于表征一个第二聚类簇的聚类中心；计算各个目标芯片版图的版图特征和各个第二版图特征之间的距离；对于任一个目标芯片版图，从任一个目标芯片版图的版图特征和各个第二版图特征之间的距离中选择最小的第二距离，将任一个目标芯片版图聚在第二距离对应的第二版图特征所表征的第二聚类簇中；若各个第二聚类簇满足聚类结束条件，则将各个第二聚类簇作为各个目标聚类簇。

在一种可能的实现方式中，筛选模块 804，用于获取任一个目标聚类簇中各个目标芯片版图的版图特征和任一个目标聚类簇的聚类中心之间的距离；基于任一个目标聚类簇中各个目标芯片版图的版图特征和任一个目标聚类簇的聚类中心之间的距离，对任一个目标聚类簇包括的各个目标芯片版图进行均匀采样，得到多个关键芯片版图。

在一种可能的实现方式中，装置还包括：

优化模块，用于对关键芯片版图进行光源掩膜优化，得到目标光源和关键芯片版图对应的掩膜版图；基于目标光源对其他芯片版图进行掩膜优化，得到其他芯片版图对应的掩膜版图，其他芯片版图是各个目标芯片版图中除关键芯片版图之外的目标芯片版图。

上述装置中的芯片版图编码器侧重于提取能区分芯片版图类型的版图特征。因此，利用芯片版图编码器提取各个目标芯片版图的版图特征，并基于各个目标芯片版图的版图特征对多个目标芯片版图进行聚类处理时，可以准确地将同一芯片版图类型的目标芯片版图聚在同一目标聚类簇中，将不同芯片版图类型的目标芯片版图聚在不同的目标聚类簇中，提高聚类结果的准确性。在从任一目标聚类簇包括的各个目标芯片版图中筛选关键芯片版图时，可以降低关键芯片版图的冗余性，提高筛选质量。

应理解的是，上述图 8 提供的装置在实现其功能时，仅以上述各功能模块的划分进行举例说明，实际应用中，可以根据需要而将上述功能分配由不同的功能模块完成，即将设备的内部结构划分成不同的功能模块，以完成以上描述的全部或者部分功能。另外，上述实施例提供的装置与方法实施例属于同一构思，其具体实现过程详见方法实施例，这里不再赘述。

图 9 示出了本申请一个示例性实施例提供的终端设备 900 的结构框图。该终端设备 900 包括有：处理器 901 和存储器 902。

处理器 901 可以包括一个或多个处理核心，比如 4 核心处理器、8 核心处理器等。处理器 901 可以采用 DSP（Digital Signal Processing，数字信号处理）、FPGA（Field-Programmable Gate Array，现场可编程门阵列）、PLA（Programmable Logic Array，可编程逻辑阵列）中的至少一种硬件形式来实现。处理器 901 也可以包括主处理器和协处理器，主处理器是用于对在唤醒状态下的数据进行处理的处理器，也称 CPU（Central Processing Unit，中央处理器）；协处理器是用于对在待机状态下的数据进行处理的低功耗处理器。在一些实施例中，处理器 901 可以集成有 GPU（Graphics Processing Unit，图像处理器），GPU 用于负责显示屏所需要显示的内容的渲染和绘制。一些实施例中，处理器 901 还可以包括 AI（Artificial Intelligence，

人工智能）处理器，该 AI 处理器用于处理有关机器学习的计算操作。

存储器 902 可以包括一个或多个计算机可读存储介质，该计算机可读存储介质可以是非暂态的。存储器 902 还可包括高速随机存取存储器，以及非易失性存储器，比如一个或多个磁盘存储设备、闪存存储设备。在一些实施例中，存储器 902 中的非暂态的计算机可读存储介质用于存储至少一个计算机程序，该至少一个计算机程序用于被处理器 901 所执行，以使终端设备 900 实现本申请中方法实施例提供的用于芯片版图编码器的训练方法或者芯片版图的筛选方法。

在一些实施例中，终端设备 900 还可选包括有：显示屏 905。

显示屏 905 用于显示 UI（User Interface，用户界面）。该 UI 可以包括图形、文本、图标、视频及其它们的任意组合。当显示屏 905 是触摸显示屏时，显示屏 905 还具有采集在显示屏 905 的表面或表面上方的触摸信号的能力。该触摸信号可以作为控制信号输入至处理器 901 进行处理。此时，显示屏 905 还可以用于提供虚拟按钮和/或虚拟键盘，也称软按钮和/或软键盘。在一些实施例中，显示屏 905 可以为一个，设置在终端设备 900 的前面板；在另一些实施例中，显示屏 905 可以为至少两个，分别设置在终端设备 900 的不同表面或呈折叠设计；在另一些实施例中，显示屏 905 可以是柔性显示屏，设置在终端设备 900 的弯曲表面上或折叠面上。甚至，显示屏 905 还可以设置成非矩形的不规则图形，也即异形屏。显示屏 905 可以采用 LCD（Liquid Crystal Display，液晶显示屏）、OLED（Organic Light-Emitting Diode，有机发光二极管）等材质制备。示例性地，样本芯片版图、旋转芯片版图、对称芯片版图等芯片版图通过显示屏 905 显示。

本领域技术人员可以理解，图 9 中示出的结构并不构成对终端设备 900 的限定，可以包括比图示更多或更少的组件，或者组合某些组件，或者采用不同的组件布置。

图 10 为本申请实施例提供的服务器的结构示意图，该服务器 1000 可因配置或性能不同而产生比较大的差异，可以包括一个或多个处理器 1001 和一个或多个的存储器 1002，其中，该一个或多个存储器 1002 中存储有至少一条计算机程序，该至少一条计算机程序由该一个或多个处理器 1001 加载并执行，以使服务器 1000 实现上述各个方法实施例提供的用于芯片版图编码器的训练方法或者芯片版图的筛选方法，示例性的，处理器 1001 为 CPU。当然，该服务器 1000 还可以具有有线或无线网络接口、键盘以及输入输出接口等部件，以便进行输入输出，该服务器 1000 还可以包括其他用于实现设备功能的部件，在此不做赘述。

在示例性实施例中，还提供了一种非易失性计算机可读存储介质，该非易失性计算机可读存储介质中存储有至少一条计算机程序，该至少一条计算机程序由处理器加载并执行，以使电子设备实现上述任一种用于芯片版图编码器的训练方法或者芯片版图的筛选方法。

可选地，上述计算机可读存储介质可以是只读存储器（Read-Only Memory，ROM）、随机存取存储器（Random Access Memory，RAM）、只读光盘（Compact Disc Read-Only Memory，CD-ROM）、磁带、软盘和光数据存储设备等。

在示例性实施例中，还提供了一种计算机程序，该计算机程序为至少一条，该至少一条计算机程序由处理器加载并执行，以使电子设备实现上述任一种用于芯片版图编码器的训练方法或者芯片版图的筛选方法。

在示例性实施例中，还提供了一种计算机程序产品，该计算机程序产品中存储有至少一条计算机程序，该至少一条计算机程序由处理器加载并执行，以使电子设备实现上述任一种用于芯片版图编码器的训练方法或者芯片版图的筛选方法。

应当理解的是，在本文中提及的“多个”是指两个或两个以上。“和/或”，描述关联对象的关联关系，表示可以存在三种关系，例如，A 和/或 B，可以表示：单独存在 A，同时存在 A 和 B，单独存在 B 这三种情况。字符“/”一般表示前后关联对象是一种“或”的关系。

上述本申请实施例序号仅仅为了描述，不代表实施例的优劣。

以上所述仅为本申请的示例性实施例，并不用以限制本申请，凡在本申请的原则之内，所作的任何修改、等同替换、改进等，均应包含在本申请的保护范围之内。

权 利 要 求 书

1.一种用于芯片版图编码器的训练方法，其中，所述方法由电子设备执行，所述方法包括：

获取样本芯片版图和初始编码器；

对所述样本芯片版图进行几何变换，得到至少一个参考芯片版图；

使用所述初始编码器提取所述样本芯片版图的样本版图特征和各个参考芯片版图的参考版图特征；

基于所述样本版图特征和所述参考版图特征，对所述初始编码器进行训练，得到训练后的编码器。

2.根据权利要求 1 所述的方法，其中，所述对所述样本芯片版图进行几何变换，得到至少一个参考芯片版图，包括：

对所述样本芯片版图进行镜面翻转，得到翻转后的芯片版图；

对所述样本芯片版图和所述翻转后的芯片版图进行旋转，得到旋转后的芯片版图，所述至少一个参考芯片版图包括所述翻转后的芯片版图和所述旋转后的芯片版图。

3.根据权利要求 1 所述的方法，其中，所述对所述样本芯片版图进行几何变换，得到至少一个参考芯片版图，包括：

对所述样本芯片版图进行镜面翻转，得到翻转后的芯片版图，所述至少一个参考芯片版图包括所述翻转后的芯片版图；

或者，对所述样本芯片版图进行旋转，得到旋转后的芯片版图，所述至少一个参考芯片版图包括所述旋转后的芯片版图。

4.根据权利要求 1 所述的方法，其中，所述对所述样本芯片版图进行几何变换，得到至少一个参考芯片版图，包括：

对所述样本芯片版图进行旋转，得到旋转后的芯片版图；

对所述样本芯片版图和所述旋转后的芯片版图进行镜面翻转，得到翻转后的芯片版图，所述至少一个参考芯片版图包括所述旋转后的芯片版图和所述翻转后的芯片版图。

5.根据权利要求 1 至 4 任一项所述的方法，其中，所述使用所述初始编码器提取所述样本芯片版图的样本版图特征，包括：

使用所述初始编码器对所述样本芯片版图进行多次下采样处理，得到各次下采样处理得到的下采样特征，所述样本版图特征为最后一次下采样处理得到的下采样特征。

6.根据权利要求 5 所述的方法，其中，所述使用所述初始编码器对所述样本芯片版图进行多次下采样处理，得到各次下采样处理得到的下采样特征，包括：

使用所述初始编码器对所述样本芯片版图进行第一次下采样处理，得到第一个下采样特征；

对于任一次下采样处理得到的任一个下采样特征，使用所述初始编码器对所述任一个下采样特征进行下一次下采样处理，得到所述下一次下采样处理得到的下采样特征。

7.根据权利要求 1 至 6 任一项所述的方法，其中，所述方法还包括：

基于所述样本版图特征，确定重构芯片版图；

所述基于所述样本版图特征和所述参考版图特征，对所述初始编码器进行训练，得到训练后的编码器，包括：

基于所述样本芯片版图、所述重构芯片版图、所述样本版图特征和所述参考版图特征，对所述初始编码器进行训练，得到训练后的编码器。

8.根据权利要求 7 所述的方法，其中，所述基于所述样本版图特征，确定重构芯片版图，包括：

对所述样本版图特征进行多次上采样处理，得到各次上采样处理得到的上采样特征，所述重构芯片版图为基于最后一次上采样处理得到的上采样特征得到的芯片版图。

9.根据权利要求 8 所述的方法，其中，所述对所述样本版图特征进行多次上采样处理，得

到各次上采样处理得到的上采样特征，包括：

对所述样本版图特征进行第一次上采样处理，得到第一个上采样特征，所述样本版图特征是对所述样本芯片版图进行多次下采样处理得到的；

对于任一次上采样处理得到的任一个上采样特征，获取所述任一个上采样特征对应的下采样特征，对所述样本芯片版图进行下采样处理得到所述任一个上采样特征对应的下采样特征时下采样处理的次数为第一次数，对所述样本版图特征进行上采样处理得到所述任一个上采样特征时上采样处理的次数为第二次数，所述第一次数与所述第二次数之和为目标次数；

将所述任一个上采样特征与所述任一个上采样特征对应的下采样特征进行拼接，得到所述任一个上采样特征对应的拼接特征；

对所述任一个上采样特征对应的拼接特征进行上采样处理，得到所述任一个上采样特征的下一个上采样特征。

10.根据权利要求 7 所述的方法，其中，所述基于所述样本芯片版图、所述重构芯片版图、所述样本版图特征和所述参考版图特征，对所述初始编码器进行训练，得到训练后的编码器，包括：

基于所述样本版图特征和所述参考版图特征，确定第一损失；

基于所述样本芯片版图和所述重构芯片版图确定第二损失；

基于所述第一损失和所述第二损失对所述初始编码器进行训练，得到训练后的编码器。

11.一种芯片版图的筛选方法，其中，所述方法由电子设备执行，所述方法包括：

获取多个目标芯片版图和训练后的编码器，所述训练后的编码器是按照权利要求 1 至 10 中任一项所述的用于芯片版图编码器的训练方法训练得到的；

使用所述训练后的编码器提取各个目标芯片版图的目标版图特征；

基于各个目标版图特征对所述多个目标芯片版图进行聚类处理，得到多个目标聚类簇，任一个目标聚类簇中包括至少一个目标芯片版图；

对于任一个目标聚类簇，从所述任一个目标聚类簇包括的各个目标芯片版图中筛选出关键芯片版图。

12.根据权利要求 11 所述的方法，其中，所述基于各个目标版图特征对所述多个目标芯片版图进行聚类处理，得到多个目标聚类簇，包括：

获取多个第一版图特征，一个第一版图特征用于表征一个第一聚类簇的聚类中心；

计算所述各个目标版图特征和各个第一版图特征之间的距离；

对于任一个目标芯片版图，从所述任一个目标芯片版图的目标版图特征和所述各个第一版图特征之间的距离中选择最小的第一距离，将所述任一个目标芯片版图聚在所述第一距离对应的第一版图特征所对应的第一聚类簇中；

若各个第一聚类簇满足聚类结束条件，则将所述各个第一聚类簇作为各个目标聚类簇。

13.根据权利要求 12 所述的方法，其中，所述方法还包括：

若所述各个第一聚类簇不满足所述聚类结束条件，则对于任一个第一聚类簇，基于所述任一个第一聚类簇中各个目标版图特征，确定一个第二版图特征，所述一个第二版图特征用于表征一个第二聚类簇的聚类中心；

计算所述各个目标版图特征和各个第二版图特征之间的距离；

对于任一个目标芯片版图，从所述任一个目标芯片版图的目标版图特征和所述各个第二版图特征之间的距离中选择最小的第二距离，将所述任一个目标芯片版图聚在所述第二距离对应的第二版图特征所对应的第二聚类簇中；

若各个第二聚类簇满足所述聚类结束条件，则将所述各个第二聚类簇作为所述各个目标聚类簇。

14.根据权利要求 11 至 13 任一项所述的方法，其中，所述从所述任一个目标聚类簇包括的各个目标芯片版图中筛选出关键芯片版图，包括：

获取所述任一个目标聚类簇中各个目标芯片版图的版图特征和所述任一个目标聚类簇的

聚类中心之间的距离；

基于所述任一个目标聚类簇中各个目标芯片版图的版图特征和所述任一个目标聚类簇的聚类中心之间的距离，对所述任一个目标聚类簇包括的各个目标芯片版图进行均匀采样，得到多个关键芯片版图。

15.根据权利要求 11 至 14 任一项所述的方法，其中，所述从所述任一个目标聚类簇包括的各个目标芯片版图中筛选出关键芯片版图之后，还包括：

对所述关键芯片版图进行光源掩膜优化，得到目标光源和所述关键芯片版图对应的掩膜版图；

基于所述目标光源对其他芯片版图进行掩膜优化，得到所述其他芯片版图对应的掩膜版图，所述其他芯片版图是所述各个目标芯片版图中除所述关键芯片版图之外的目标芯片版图。

16.一种用于芯片版图编码器的训练装置，其中，所述装置包括：

获取模块，用于获取样本芯片版图和初始编码器；

变换模块，用于对所述样本芯片版图进行几何变换，得到至少一个参考芯片版图；

提取模块，用于使用所述初始编码器提取所述样本芯片版图的样本版图特征和各个参考芯片版图的参考版图特征；

训练模块，用于基于所述样本版图特征和所述参考版图特征，对所述初始编码器进行训练，得到训练后的编码器。

17.一种芯片版图的筛选装置，其中，所述装置包括：

获取模块，用于获取多个目标芯片版图和训练后的编码器，所述训练后的编码器是按照权利要求 1 至 10 中任一项所述的用于芯片版图编码器的训练方法训练得到的；

提取模块，用于使用所述训练后的编码器提取各个目标芯片版图的目标版图特征；

聚类模块，用于基于各个目标版图特征对所述多个目标芯片版图进行聚类处理，得到多个目标聚类簇，任一个目标聚类簇中包括至少一个目标芯片版图；

筛选模块，用于对于任一个目标聚类簇，从所述任一个目标聚类簇包括的各个目标芯片版图中筛选出关键芯片版图。

18.一种电子设备，其中，所述电子设备包括处理器和存储器，所述存储器中存储有至少一条计算机程序，所述至少一条计算机程序由所述处理器加载并执行，以使所述电子设备实现如权利要求 1 至 10 任一所述的用于芯片版图编码器的训练方法或者实现如权利要求 11 至 15 任一所述的芯片版图的筛选方法。

19.一种非易失性计算机可读存储介质，其中，所述非易失性计算机可读存储介质中存储有至少一条计算机程序，所述至少一条计算机程序由处理器加载并执行，以使电子设备实现如权利要求 1 至 10 任一所述的用于芯片版图编码器的训练方法或者实现如权利要求 11 至 15 任一所述的芯片版图的筛选方法。

20.一种计算机程序产品，其中，所述计算机程序产品中存储有至少一条计算机程序，所述至少一条计算机程序由处理器加载并执行，以使电子设备实现如权利要 1 至 10 任一所述的用于芯片版图编码器的训练方法或者实现如权利要求 11 至 15 任一所述的芯片版图的筛选方法。

21.一种计算机程序，其中，所述计算机程序为至少一条，至少一条计算机程序由处理器加载并执行，以使电子设备实现如权利要 1 至 10 任一所述的用于芯片版图编码器的训练方法或者实现如权利要求 11 至 15 任一所述的芯片版图的筛选方法。

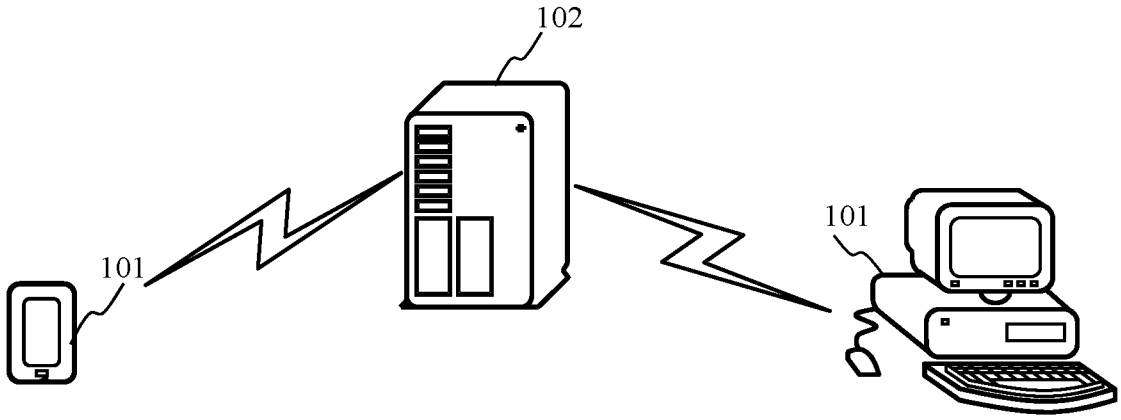

102
101
101

图 1

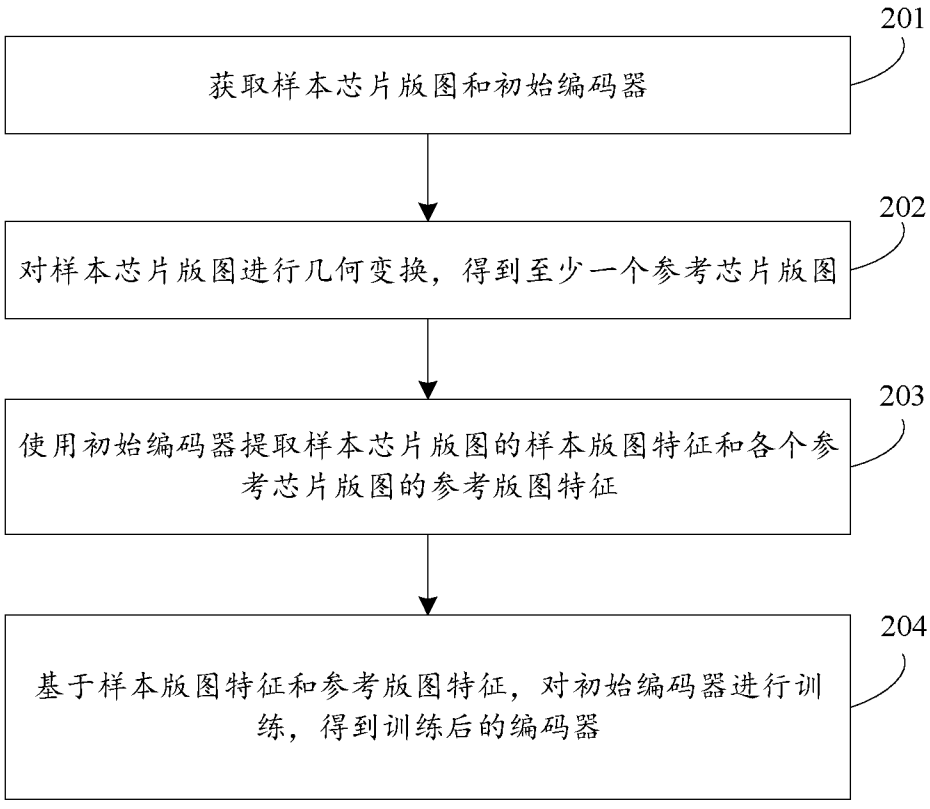

201
获取样本芯片版图和初始编码器
202
对样本芯片版图进行几何变换，得到至少一个参考芯片版图
203
使用初始编码器提取样本芯片版图的样本版图特征和各个参考芯片版图的参考版图特征
204
基于样本版图特征和参考版图特征，对初始编码器进行训练，得到训练后的编码器

图 2

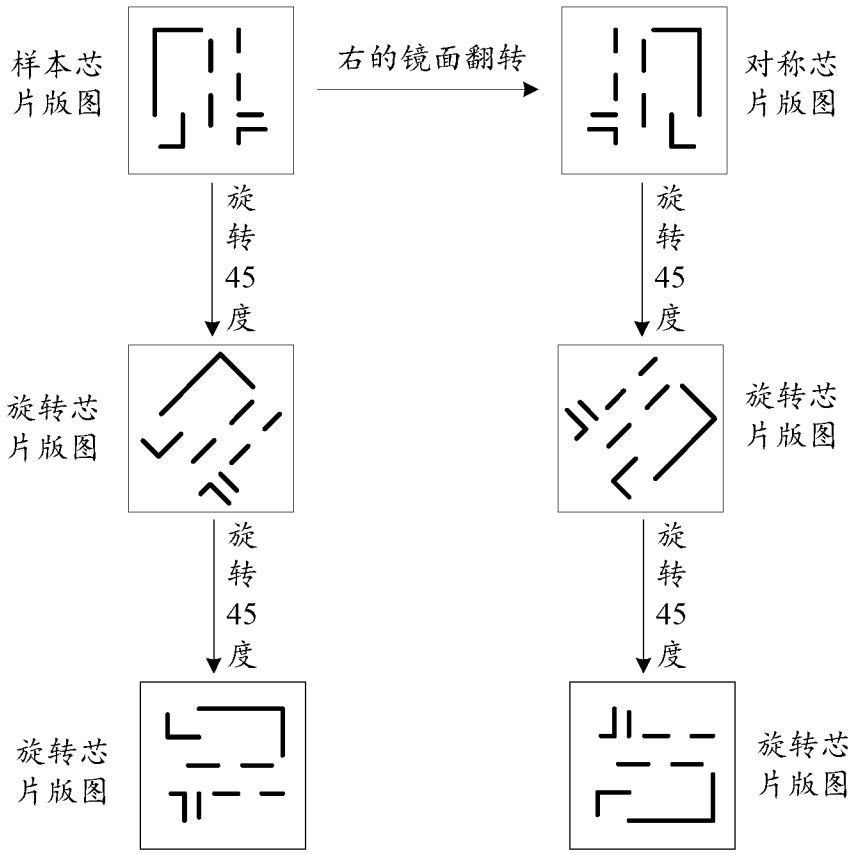

样本芯片版图
右的镜面翻转
对称芯片版图
旋转45度
旋转45度
旋转芯片版图
旋转芯片版图
旋转45度
旋转45度
旋转芯片版图
旋转芯片版图

图 3

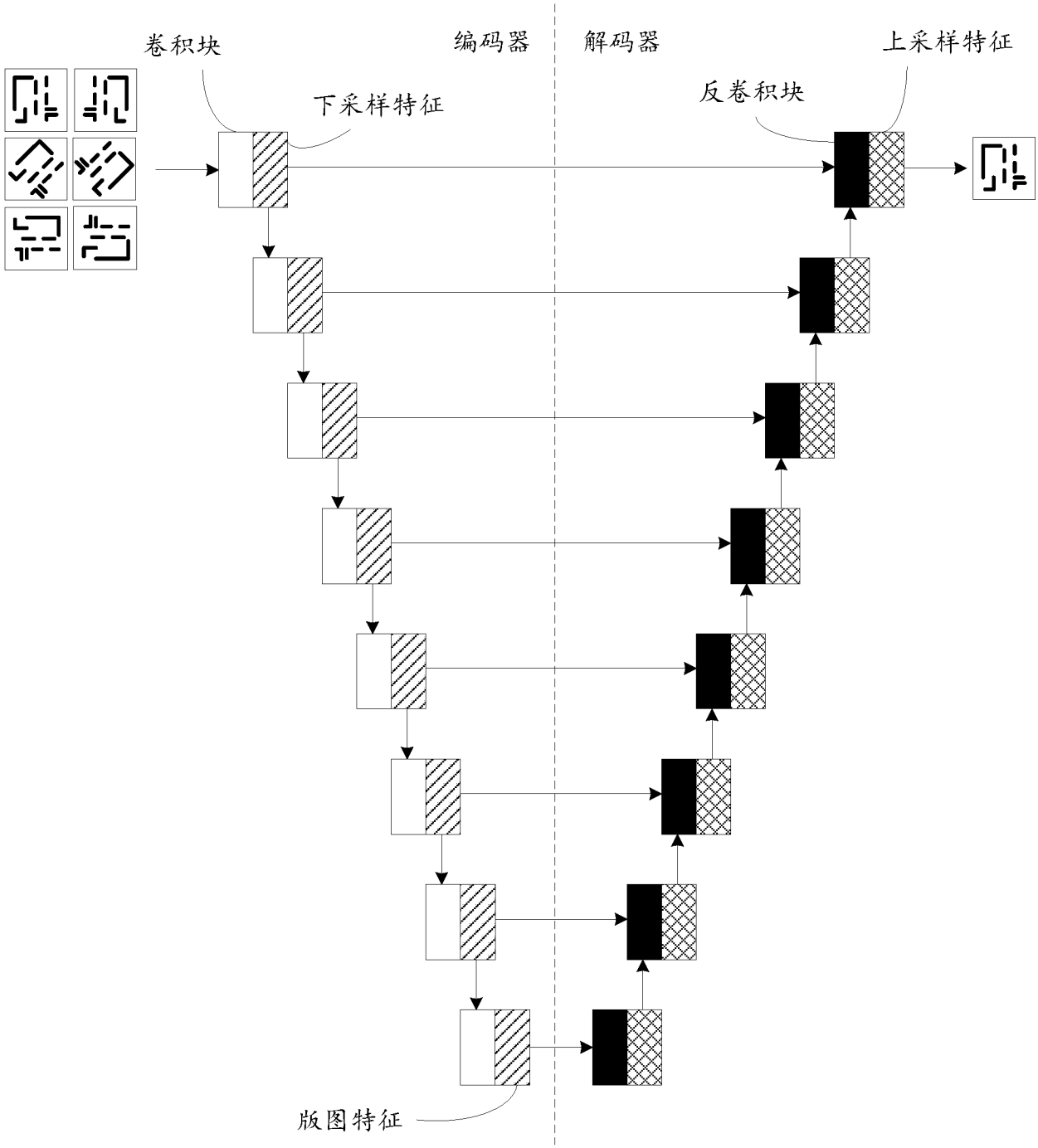


图 4

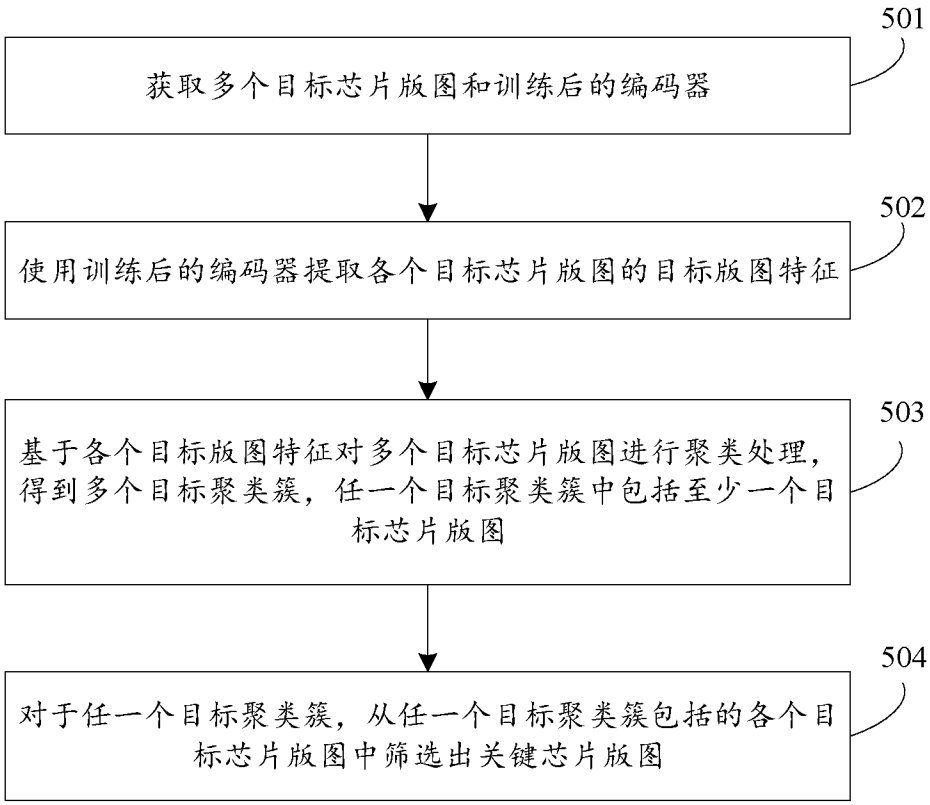

501
获取多个目标芯片版图和训练后的编码器
502
使用训练后的编码器提取各个目标芯片版图的目标版图特征
503
基于各个目标版图特征对多个目标芯片版图进行聚类处理，得到多个目标聚类簇，任一个目标聚类簇中包括至少一个目标芯片版图
504
对于任一个目标聚类簇，从任一个目标聚类簇包括的各个目标芯片版图中筛选出关键芯片版图

图 5

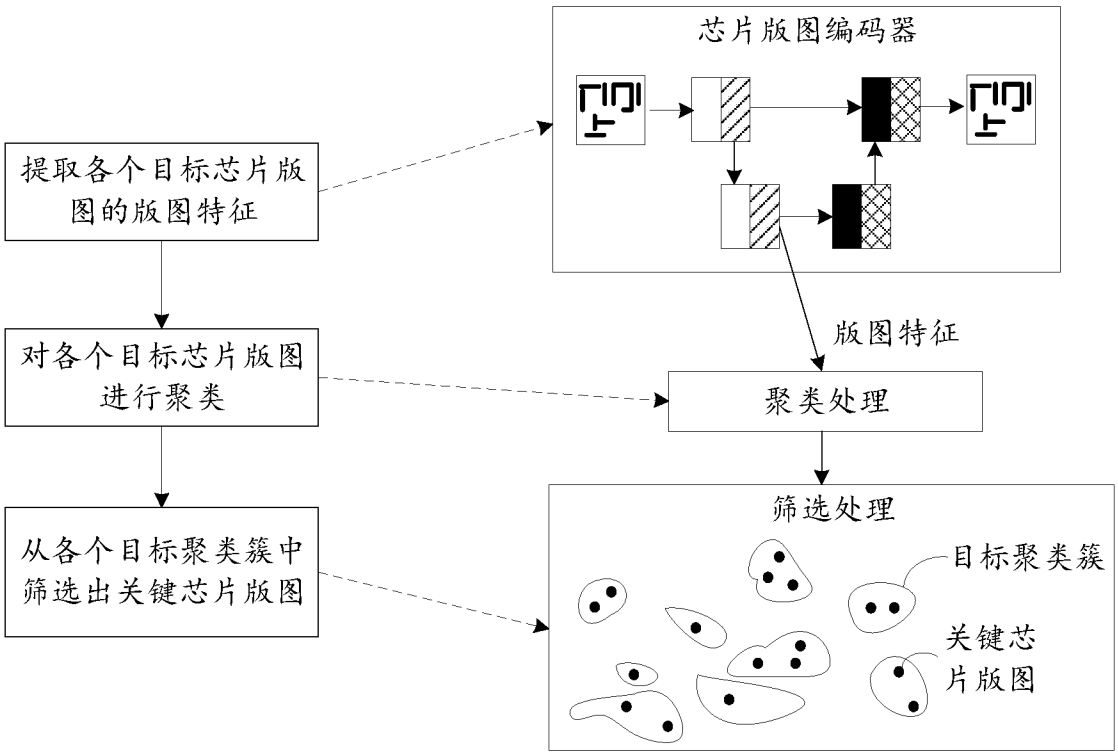

芯片版图编码器
提取各个目标芯片版图的版图特征
版图特征
对各个目标芯片版图进行聚类
聚类处理
筛选处理
从各个目标聚类簇中筛选出关键芯片版图
目标聚类簇
关键芯片版图

图 6

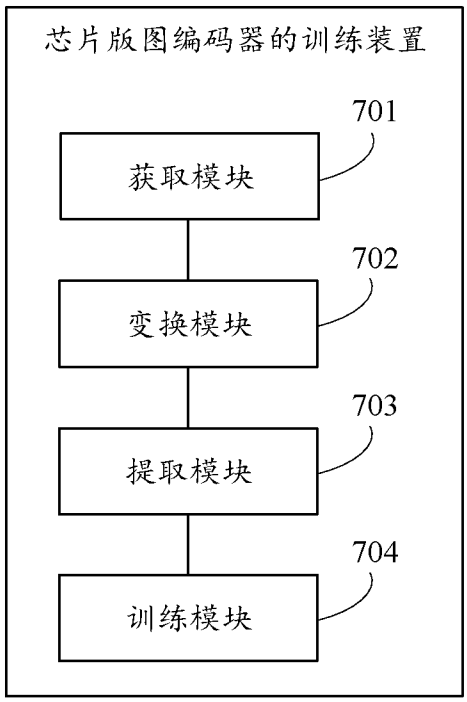

芯片版图编码器的训练装置
701
获取模块
702
变换模块
703
提取模块
704
训练模块

图 7

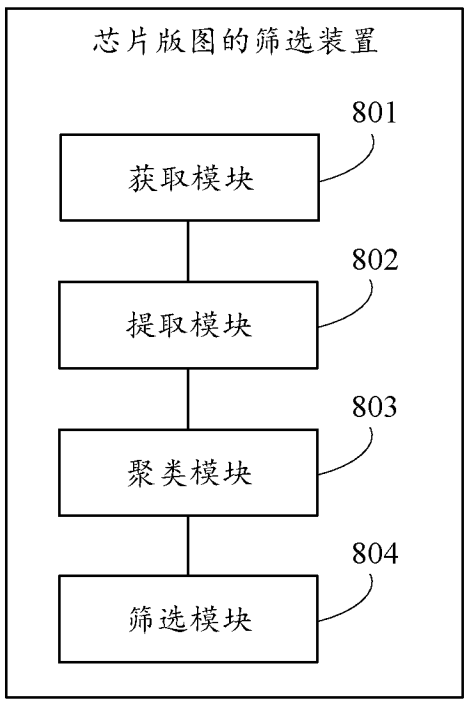

芯片版图的筛选装置
801
获取模块
802
提取模块
803
聚类模块
804
筛选模块

图 8

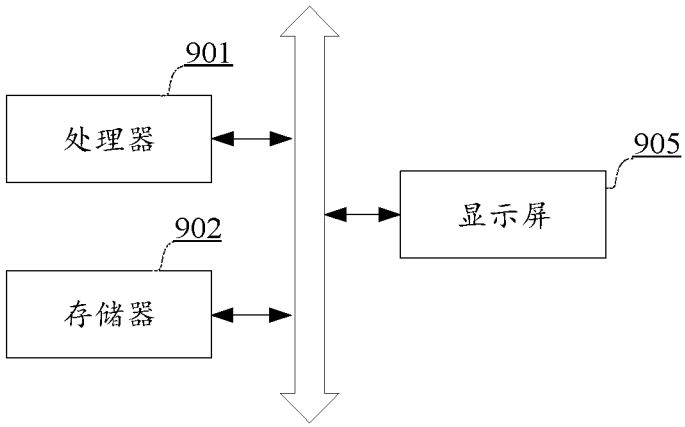


图 9

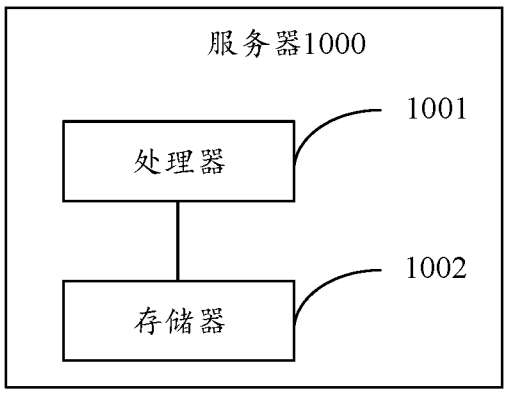


图 10

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2023/097408

A. CLASSIFICATION OF SUBJECT MATTER

G06V10/40(2022.01)i; G06V10/762(2022.01)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:G06V10/-; G06K9/-

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)

CNTXT, ENTXTC, VEN, ENTXT, IEEE, CJFD: 集成电路, 版图, 编码器, 旋转, 镜像, 翻转, 一致性, 损失函数, 特征, 提取, 图片, 图像, 芯片, 距离, 采样, 聚类, feature, extract+, flip+, rotat+, cluster+, distance, image, picture, loss function, consistency loss, sample+, u-net, encod+, integrated circuit layout

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 112418321 A (BEIJING WODONG TIANJUN INFORMATION TECHNOLOGY CO., LTD. et al.) 26 February 2021 (2021-02-26) description, paragraphs [0046]-[0094]	1-6, 16, 18-21
Y	CN 112418321 A (BEIJING WODONG TIANJUN INFORMATION TECHNOLOGY CO., LTD. et al.) 26 February 2021 (2021-02-26) description, paragraphs [0046]-[0094]	11-15, 17
Y	US 2021319227 A1 (INTERNATIONAL BUSINESS MACHINES CORPORATION) 14 October 2021 (2021-10-14) description, paragraphs [0018]-[0026]	11-15, 17
A	CN 108535952 A (BEIJING INSTITUTE OF TECHNOLOGY) 14 September 2018 (2018-09-14) entire document	1-21
A	CN 110187609 A (BEIJING INSTITUTE OF TECHNOLOGY) 30 August 2019 (2019-08-30) entire document	1-21

☑ 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	Date of mailing of the international search report
04 September 2023	**07 September 2023**

Name and mailing address of the ISA/CN	Authorized officer
China National Intellectual Property Administration (ISA/CN) **China No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing 100088**	Telephone No.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CN 111353411 A (SICHUAN YIFEISHI TECHNOLOGY CO., LTD.) 30 June 2020 (2020-06-30) entire document	1-21
A	CN 111445396 A (CHINA ELECTRONIC PRODUCT RELIABILITY AND ENVIRONMENTAL TESTING RESEARCH INSTITUTE ((FIFTH ELECTRONIC RESEARCH INSTITUTE OF MIIT) (CHINA CEPREI LABORATORY))) 24 July 2020 (2020-07-24) entire document	1-21
A	CN 111898619 A (SHANGHAI EYE CONTROL TECHNOLOGY CO., LTD.) 06 November 2020 (2020-11-06) entire document	1-21
A	US 2022156886 A1 (ADOBE INC.) 19 May 2022 (2022-05-19) entire document	1-21

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2023/097408

Patent document cited in search report	Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN 112418321 A	26 February 2021	None	
US 2021319227 A1	14 October 2021	WO 2021209835 A1	21 October 2021
		GB 2612707 A	10 May 2023
		DE 112021002291 T5	16 March 2023
		JP 2023520631 A	18 May 2023
		US 11164005 B1	02 November 2021
		CN 115280378 A	01 November 2022
		CN 115280378 B	22 August 2023
CN 108535952 A	14 September 2018	CN 108535952 B	11 August 2020
CN 110187609 A	30 August 2019	CN 110187609 B	21 August 2020
CN 111353411 A	30 June 2020	None	
CN 111445396 A	24 July 2020	None	
CN 111898619 A	06 November 2020	None	
US 2022156886 A1	19 May 2022	None	

国际检索报告

国际申请号
PCT/CN2023/097408

A. 主题的分类

G06V10/40(2022.01)i; G06V10/762(2022.01)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:G06V10/-; G06K9/-

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））

CNTXT,ENTXTC,VEN,ENTXT,IEEE,CJFD:集成电路,版图,编码器,旋转,镜像,翻转,一致性,损失函数,特征,提取,图片,图像,芯片,距离,采样,聚类, feature,extract+, flip+,rotat+,cluster+,distance,image,picture,loss function,consistency loss,sample+,u-net,encod+,integrated circuit layout

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	CN 112418321 A (北京沃东天骏信息技术有限公司 等) 2021年2月26日 (2021-02-26) 说明书[0046]-[0094]段	1-6,16,18-21
Y	CN 112418321 A (北京沃东天骏信息技术有限公司 等) 2021年2月26日 (2021-02-26) 说明书[0046]-[0094]段	11- 15,17
Y	US 2021319227 A1 (International Business Machines Corporation) 2021年10月14日 (2021-10-14) 说明书[0018]-[0026]段	11-15,17
A	CN 108535952 A (北京理工大学) 2018年9月14日 (2018-09-14) 全文	1-21
A	CN 110187609 A (北京理工大学) 2019年8月30日 (2019-08-30) 全文	1-21
A	CN 111353411 A (四川翼飞视科技有限公司) 2020年6月30日 (2020-06-30) 全文	1-21

☑ 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	国际检索报告邮寄日期
2023年9月4日	2023年9月7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	受权官员 邱川 电话号码 (+86) 028-62967639

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
A	CN 111445396 A（中国电子产品可靠性与环境试验研究所（（工业和信息化部电子第五研究所）（中国赛宝实验室））） 2020年7月24日 (2020 - 07 - 24) 全文	1-21
A	CN 111898619 A（上海眼控科技股份有限公司） 2020年11月6日 (2020 - 11 - 06) 全文	1-21
A	US 2022156886 A1 (ADOBE INC.) 2022年5月19日 (2022 - 05 - 19) 全文	1-21

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2023/097408

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)
CN	112418321	A	2021年2月26日	无			
US	2021319227	A1	2021年10月14日	WO	2021209835	A1	2021年10月21日
				GB	2612707	A	2023年5月10日
				DE	112021002291	T5	2023年3月16日
				JP	2023520631	A	2023年5月18日
				US	11164005	B1	2021年11月2日
				CN	115280378	A	2022年11月1日
				CN	115280378	B	2023年8月22日
CN	108535952	A	2018年9月14日	CN	108535952	B	2020年8月11日
CN	110187609	A	2019年8月30日	CN	110187609	B	2020年8月21日
CN	111353411	A	2020年6月30日	无			
CN	111445396	A	2020年7月24日	无			
CN	111898619	A	2020年11月6日	无			
US	2022156886	A1	2022年5月19日	无			