

图像去重处理方法、装置、存储介质和电子设备

本申请要求于 2023 年 12 月 26 日提交中华人民共和国国家知识产权局、申请号为 202311803404.X、申请名称为“图像去重处理方法、装置、存储介质和电子设备”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。

5 技术领域

本申请涉及图像处理领域，更具体地涉及一种图像去重处理方法、图像去重处理装置、存储介质和电子设备。

背景技术

10 随着计算机处理能力的提高，计算机视觉技术越来越多地应用到各个场景中。但是，一方面，图像的数据量大，计算复杂；另一方面，在实际的应用场景中，可能采集到大量的图像，信息重复度高。为此，需要进行图像去重处理。该图像去重处理可以从大量图像中筛选出一些期望图像，用于后续应用。

例如，目前在晶圆等高精尖的技术领域中，对所生产的产品的良率要求极高。通常可以利用诸如神经网络模型等学习模型来自动检测晶圆的良率。当为了训练学习模型进行样本图像收集时，所收集到的样本图像往往数据过于丰富、信息重复度较高。因此，需要对所收集到的样本图像进行图像去重处理。

现有的图像去重处理方法，多是直接进行各个图像之间的无差别比较，从而基于比较结果筛选掉重复度较高的图像。这种图像去重方式计算量大、实时性差。

20

发明内容

考虑到上述问题而提出了本申请。

本申请第一方面提供了一种像去重处理方法。所述方法包括：多个样本图像中确定模板图像；

25 针对多个样本图像中的每个样本图像，获取待处理图像的频谱图，其中，该待处理图像包括该样本图像和/或该样本图像的特征图像；基于该待处理图像的频谱图和模板图像的频谱图，确定该待处理图像与模板图像的频率相似度；基于频率相似度，确定该样本图像和模板图像的总相似度；

根据多个样本图像中每个样本图像分别与模板图像的总相似度，在多个样本图像
30 中确定预设数量的目标图像。

在一种可能的实施方式中，从多个样本图像中确定模板图像包括：计算多个样本图像中不同样本图像之间的余弦相似度；

针对多个样本图像中的每个样本图像，统计第一样本图像的个数，其中第一样本图像与该样本图像的余弦相似度大于相似度阈值；

5 比较针对每个样本图像所统计的第一样本图像的个数，并将针对其所统计的第一样本图像的个数最多的样本图像确定为模板图像。

在一种可能的实施方式中，频率相似度包括高频相似度和/或低频相似度；获取该待处理图像的频谱图包括：对该待处理图像执行傅里叶变换，以获得该待处理图像的频谱图；

10 基于该待处理图像的频谱图和模板图像的频谱图，确定与该待处理图像与模板图像的频率相似度，包括：确定该待处理图像的频谱图中的高通滤波区域和/或低通滤波区域，并且确定模板图像的频谱图中的高通滤波区域和/或低通滤波区域；对所确定的高通滤波区域和/或低通滤波区域进行傅里叶逆变换，以获得对应的高通滤波图像和/或低通滤波图像；

15 基于该待处理图像对应的高通滤波图像和模板图像对应的高通滤波图像，计算该待处理图像与模板图像的高频相似度，和/或，基于该待处理图像对应的低通滤波图像和模板图像对应的低通滤波图像，计算该待处理图像与模板图像的低频相似度。

在一种可能的实施方式中，确定该待处理图像的频谱图中的高通滤波区域和/或低通滤波区域，并且确定模板图像的频谱图中的高通滤波区域和/或低通滤波区域包括：

20 对于该待处理图像的频谱图或模板图像的频谱图，在该频谱图中，确定基准点，其中，基准点是该频谱图的中心点；在该频谱图中，确定以基准点为中心点的矩形区域为低通滤波区域，所确定的低通滤波区域以外区域为高通滤波区域。

在一种可能的实施方式中，基于该待处理图像对应的高通滤波图像和模板图像对应的高通滤波图像，计算该待处理图像与模板图像的高频相似度包括：

25 计算该待处理图像对应的高通滤波图像与模板图像对应的高通滤波图像的结构相似性，以作为该待处理图像与模板图像的高频相似度；和/或基于该待处理图像对应的低通滤波图像和模板图像对应的低通滤波图像，计算该待处理图像与模板图像的低频相似度，包括：计算该待处理图像对应的低通滤波图像与模板图像对应的低通滤波图像的结构相似性，以作为该待处理图像与模板图像的低频相似度。

30 在一种可能的实施方式中，频率相似度包括高频相似度和低频相似度；基于频率

相似度，确定该样本图像和模板图像的总相似度包括：计算该样本图像的特征图像与模板图像的特征图像的高频相似度和该样本图像与模板图像的高频相似度的乘积，以作为第一相似度；计算该样本图像的特征图像与模板图像的特征图像的低频相似度和该样本图像与模板图像的低频相似度的乘积，以作为第二相似度；以及计算第一相似度与第二相似度的和，以作为该样本图像和模板图像的总相似度。

在一种可能的实施方式中，根据多个样本图像中每个样本图像分别与模板图像的总相似度，在多个样本图像中确定预设数量的目标图像包括：对多个样本图像分别与模板图像的总相似度进行聚类，其中，所聚类别为预设数量；根据聚类结果，在每个所聚类别的样本图像中分别选择一个目标图像。

本申请第二方面还提供了一种图像去重处理装置，该装置包括：

选取模块，用于从多个样本图像中确定模板图像；

获取模块，用于针对多个样本图像中的每个样本图像，获取待处理图像的频谱图，其中，该待处理图像包括该样本图像和/或该样本图像的特征图像；

确定模块，用于基于该待处理图像的频谱图和模板图像的频谱图，确定该待处理图像与模板图像的频率相似度；

对比模块，用于基于频率相似度，确定该样本图像和模板图像的总相似度；

去重模块，用于根据多个样本图像中每个样本图像分别与模板图像的总相似度，在多个样本图像中确定预设数量的目标图像。

本申请第三方面还提供了一种存储介质，在存储介质上存储了程序指令，程序指令在运行时用于执行上述图像去重处理方法。

本申请第四方面还提供了一种电子设备，该电子设备包括处理器和存储器，存储器中存储有计算机程序指令，计算机程序指令被处理器运行时用于执行上述图像去重处理方法。

在上述技术方案中，首先，从多个样本图像中确定模板图像；然后利用模板图像基于样本图像所对应的频率相似度对样本图像集进行去重。这不仅保证了图像去重处理的准确性，而且保证了处理速度，提高了实时性。

上述说明仅是本申请技术方案的概述，为了能够更清楚了解本申请的技术手段，而可依照说明书的内容予以实施，并且为了让本申请的上述和其它目的、特征和优点能够更明显易懂，以下特举本申请的具体实施方式。

附图说明

通过结合附图对本申请实施例进行更详细的描述，本申请的上述以及其它目的、特征和优势将变得更加明显。附图用来提供对本申请实施例的进一步理解，并且构成说明书的一部分，与本申请实施例一起用于解释本申请，并不构成对本申请的限制。

在附图中，相同的参考标号通常代表相同部件或步骤。

图 1 示出了根据本申请的一个实施例的图像去重处理方法的示意性流程图；

图 2 示出了根据本申请的一个实施例的确定模板图像的示意性流程图；

图 3 示出了根据本申请的一个实施例的图像去重处理装置的示意性框图；

图 4 示出了根据本申请的一个实施例的电子设备的示意性框图。

具体实施方式

为了使得本申请的目的、技术方案和优点更为明显，下面将参照附图详细描述根据本申请的示例实施例。显然，所描述的实施例仅仅是本申请的一部分实施例，而不是本申请的全部实施例，应理解，本申请不受这里描述的示例实施例的限制。基于本申请中描述的本申请实施例，本领域技术人员在没有付出创造性劳动的情况下所得到的所有其它实施例都应落入本申请的保护范围之内。

为了至少部分地解决上述技术问题，根据本申请的一个方面，提供了一种图像去重处理方法。

图 1 示出了根据本申请的一个实施例的图像去重处理方法的示意性流程图。如图 1 所示，该方法包括步骤 S110、步骤 S120、步骤 S130、步骤 S140 和步骤 S150。

步骤 S110：从多个样本图像中确定模板图像。

样本图像是待去重的图像。例如，在晶圆生产领域，可以采集大量的晶圆图像，用于对人工智能模型进行训练。因为大量的晶圆图像中存在重复信息，可以对这些晶圆图像进行去重处理，以筛选出期望图像，用于后续对人工智能模型的训练。

样本图像可以是 RGB 图像或灰度图像。样本图像可以是任意合适尺寸、合适分辨率的图像。样本图像可以是图像采集装置直接采集到的原始图像，也可以是对原始图像进行预处理操作后的图像。该预处理操作可以包括为了改善样本图像的视觉效果，提高其清晰度，或是突出图像中的某些特征（例如上述晶圆）的所有操作。示例性而非限制性地，预处理操作可以包括对原始图像的数字化的、几何变换、归一化、滤波等操作。

可以从多个样本图像中确定一个模板图像。可选地，可以从多个样本图像中随机指定一个模板图像。又可选地，可以根据多个样本图像之间的相似度，从多个样本图像中确定与其他图像相似度最高的图像为模板图像。模板图像可以用于与多个样本中的每个样本图像进行相似性的比较。

步骤 S120: 针对多个样本图像中的每个样本图像, 获取待处理图像的频谱图, 其中, 该待处理图像包括该样本图像和/或该样本图像的特征图像。

样本图像的特征图像可以通过卷积神经网络 (CNN) 的卷积操作提取的特征图。卷积操作通过滤波器的移动和逐个元素相乘实现了对图像的空间信息的提取和处理, 并且可以学习到不同特征的权重和偏置, 使得 CNN 能够有效地识别和分类图像。具体地, 可以通过 resnet50 等特征提取网络来进行特征提取。

在步骤 S120 中, 将每个样本图像和/或样本图像的特征图像, 转换为频率域表示的频谱图。当然, 模板图像是样本图像之中的一个, 同样也获得与模板图像相对应的频谱图。

可以理解, 将待处理图像转换为其频谱图可以采用任何现有的或未来研发的算法具体实现。

步骤 S130: 针对每个待处理图像, 基于该待处理图像的频谱图和模板图像的频谱图, 确定该待处理图像与模板图像的频率相似度。

待处理图像包括样本图像和/或该样本图像的特征图像, 模板图像包括模板图像和/或该模板图像的特征图像。可以将样本图像的频谱图和模板图像的频谱图进行相似度计算, 或者, 将样本图像的特征图像的频谱图和模板图像的特征图像的频谱图进行相似度计算。由此, 可以得到与每个待处理图像相对应的频率相似度。可以理解, 根据与每个待处理图像相对应的频率相似度, 可以确定与每个样本图像相对应的频率相似度。

可以理解, 可以基于待处理图像的频谱图和模板图像频谱图, 计算该待处理图像与模板图像在不同频段的频率相似度。示例性地, 对于每个待处理图像, 分别计算其与模板图像高频相似度和/或低频相似度。可以根据频率阈值来确定高频相似度和低频相似度。其中高频相似度表示高于频率阈值的频率的相似度, 低频相似度表示低于频率阈值的频率的相似度。

步骤 S140: 对于每个待处理图像, 基于频率相似度, 确定该样本图像和模板图像的总相似度。

示例性地, 在步骤 S130 中, 可以确定了待处理图像与模板图像的高频相似度和低频相似度。此外, 上述待处理图像可以包括样本图像自身和样本图像的特征图像二者。在步骤 S140 中, 可以根据步骤 S130 所确定的所有频率相似度, 计算每个样本图像和模板图像之间的总相似度。总相似度能够综合反映两个图像的整体相似程度, 可以更准确地评估两个图像之间的相似性。

步骤 S150: 根据多个样本图像中每个样本图像分别与模板图像的总相似度, 在多个样本图像中确定预设数量的目标图像。

预设数量可以表示样本图像经图像去重操作后所要保留的图像数量。其可以根据预期的去重效果进行设置。期望的去重比例越高, 该预设数量可以设置得越小。

根据每个样本图像的总相似度和预设数量, 可以筛选出符合要求的总相似度。根

据筛选出的总相似度所对应的样本图像，进而确定出所需要的目标图像。将除目标图像外的样本图像去除，以完成图像去重操作。

在上述实施例中，首先，从多个样本图像中确定模板图像；然后利用模板图像基于样本图像所对应的频率相似度对样本图像集进行去重。这不仅保证了图像去重处理的准确性，而且保证了处理速度，提高了实时性。

在一些实施例中，图 2 示出了根据本申请的一个实施例的确定模板图像的示意性流程图。如图 2 所示，步骤 S110 从多个样本图像中确定模板图像包括步骤 S111 至步骤 S113。

步骤 S111：计算多个样本图像中不同样本图像之间的余弦相似度。

可以计算多个样本图像中每两个样本图像之间的余弦相似度，以根据余弦相似度确定模板图像。将包含模板图像在内的每个样本图像转换为向量表示。计算每两个向量之间的余弦相似度，其可以用于衡量两个图像之间的相似性。

余弦相似度的取值范围可以在-1 到 1 之间，取值数值越接近于 1，可以表示为两个向量越相似，即两个图像之间的相似度越高。取值数值越接近于-1，可以表示两个向量越不相似，即两个图像之间的相似度越低。

步骤 S112：针对多个样本图像中的每个样本图像，统计第一样本图像的个数，其中第一样本图像与该样本图像的余弦相似度大于相似度阈值。

相似度阈值可以根据应用场景的需求进行设置，例如：0.9。

依次计算每个样本图像和其余样本图像之间的余弦相似度。当两个样本图像之间的余弦相似度大于相似度阈值时，则可以认为针对二者分别统计的第一样本图像的个数均增加 1。换言之，二者互为针对对方的第一样本图像。当两个样本图像之间的余弦相似度小于或等于相似度阈值时，则各自的第一样本图像的个数保持不变。当针对所有样本图像全部计算完成后，可以获得针对每个样本图像的第一样本图像的个数。

步骤 S113：比较针对每个样本图像所统计的第一样本图像的个数，并将针对其所统计的第一样本图像的个数最多的样本图像确定为模板图像。

示例性地，以多个样本图像的数量为 200 个，由于每个样本图像会与其他样本图像计算 199 次余弦相似度，则针对该样本图像所统计的第一样本图像的个数最小可以为 0，最高可以为 199。如果对于多个样本图像中的一个样本图像，针对其所统计的第一样本图像的个数最多，换言之，与该样本图像之间的余弦相似度大于相似度阈值的图像最多，则将该样本图像确定为模板图像。当针对两个或更多个样本图像所统计的第一样本图像的个数一样时，则可以在两个或更多个样本图像中随机选择其中一个作为模板图像。

在上述实施例中，通过计算多个样本图像中的余弦相似度，并将与其他样本图像的余弦相似度大于相似度阈值的情况出现最多的样本图像确定为模板图像。该模板图像更能够体现多个样本图像的特征。利用该模板图像进行后续图像去重操作，有效地提高了图像去重的效果。

如前所述，步骤 S130 中的频率相似度可以包括高频相似度和/或低频相似度。

在一些实施例中，步骤 S120 获取该待处理图像的频谱图包括：对该待处理图像执行傅里叶变换，以获得该待处理图像的频谱图。

傅里叶变换能够将图像从空间域转换到频率域。傅里叶变换可以将待处理图像的灰度分布函数变换为图像的频率分布函数。对待处理图像进行傅里叶变换得到的频谱图是图像梯度的分布图。频谱图上看到的明暗不一的亮点，实际上是待处理图像上的一个像素点与邻域像素点差异的强弱，即梯度的大小，也即该像素点的频率的大小。如果频谱图中暗的像素点更多，则待处理图像比较柔和；反之，如果频谱图中亮的像素点更多，则待处理图像比较尖锐，边界分明且边界两边像素差异较大。总之，待处理图像的频谱图从频率角度表示了待处理图像的特征。

步骤 S130 基于该待处理图像的频谱图和模板图像的频谱图，确定与该待处理图像与模板图像的频率相似度，可以包括以下步骤：首先，确定该待处理图像的频谱图中的高通滤波区域和/或低通滤波区域，并且确定模板图像的频谱图中的高通滤波区域和/或低通滤波区域。然后，对所确定的高通滤波区域和/或低通滤波区域进行傅里叶逆变换，以获得对应的高通滤波图像和/或低通滤波图像。最后，基于该待处理图像对应的高通滤波图像和模板图像对应的高通滤波图像，计算该待处理图像与模板图像的高频相似度；和/或，基于该待处理图像对应的低通滤波图像和模板图像对应的低通滤波图像，计算该待处理图像与模板图像的低频相似度。

待处理图像的频谱图可以包括高通滤波区域和低通滤波区域。越靠近于频谱图的中心点的区域对应的频率越低，反之，频谱图外围的边缘区域对应的频率更高。频谱图的中心点可以表示待处理图像的平均灰度，中心点的频率为 0。从频谱图的中心点向外，频率依次增高。即频谱图的中心区域对应低通滤波区域，外围区域则对应高通滤波区域。根据上述原理，可以分别确定待处理图像和模板图像各自的频谱图中的高通滤波区域和/或低通滤波区域。将待处理图像的频谱图的低通滤波区域的频谱数值设置为 0，高通滤波区域的频谱数值保持不变，则可以得到高通滤波区域。将上述频谱图的高通滤波区域的频谱数值设置为 0，低通滤波区域的频谱数值保持不变，则可以得到低通滤波区域。

将高通滤波区域和/或低通滤波区域进行傅里叶逆变换，可以得到对应的高通滤波图像和/或低通滤波图像。通过傅里叶逆变换，能够将图像的频率分布变换为其灰度分布。

在一些实施例中，可以基于待处理图像和模板图像各自对应的高通滤波图像计算待处理图像与模板图像的高频相似度。替代地，可以基于该待处理图像和模板图像各自对应的低通滤波图像，计算该待处理图像与模板图像的低频相似度。在优选实施例中，可以既计算上述高频相似度，也计算低频相似度。某些图像的高频细节较为显著，而低频部分被减弱或过滤掉，因此该图像的频谱图中高通滤波区域的亮度更大，更有意义。同理，某些图像的低频成分较为明显，而高频细节被模糊化或抑制，则该图像的频谱图中低通滤波区域的亮度更大，更有意义。在实际应用场景中，可以根据

需要计算高频相似度和低频相似度中的任一者或二者均计算。

在上述实施例中，利用待处理图像的高通滤波图像和/或低通滤波图像，分别与模板图像的高通滤波图像和/或低通滤波图像进行计算，得到相对应的高频相似度和/或低频相似度。该高频相似度和低频相似度表征了图像之间的相似关系，基于其可以有效提高图像去重的准确性，即保证所去掉的图像是与留下的图像重复度较高的图像。特别地，对于既计算高频相似度又计算低频相似度的实施例，可以更全面地比较图像的特征，进一步有效地提高图像去重的准确性。

在一些实施例中，上述确定该待处理图像的频谱图中的高通滤波区域和/或低通滤波区域并且确定模板图像的频谱图中的高通滤波区域和/或低通滤波区域的步骤，可以包括：对于该待处理图像的频谱图或模板图像的频谱图，执行以下操作。首先，在该频谱图中，确定基准点，其中，基准点是该频谱图的中心点；然后，在该频谱图中，确定以基准点为中心点的矩形区域为低通滤波区域，所确定的低通滤波区域以外区域为高通滤波区域。

示例性地，可以首先根据频谱图的长和宽，确定频谱图的中心点，以作为基准点。以该基准点为矩形区域的中心点，确定矩形区域。可以理解，所确定的矩形区域将位于频谱图的中心位置。示例性地，通过上述操作，在频谱图的中心位置设置一个 60×60 的矩形区域。矩形区域即为低通滤波区域，矩形区域外的区域为高通滤波区域。频谱图中的矩形区域的尺寸可以根据具体的实际需求进行调整。通过调整矩形区域的尺寸，可以控制频谱图中的低通滤波区域和高通滤波区域各自的范围。

在上述实施例中，通过在频谱图中设置矩形区域来确定低通滤波区域和高通滤波区域。该方案可以灵活地调整低通滤波区域和高通滤波区域的范围，并且计算简单，实现容易。

在一些实施例中，上述基于该待处理图像对应的高通滤波图像和模板图像对应的高通滤波图像计算该待处理图像与模板图像的高频相似度的步骤，可以包括：计算该待处理图像对应的高通滤波图像与模板图像对应的高通滤波图像的结构相似性（Structural Similarity Index，简称为 SSIM），以作为该待处理图像与模板图像的高频相似度。

结构相似性可以用于衡量两个图像之间相似性。结构相似性是基于感知的，其更符合人眼的直观感受。结构相似性主要考量图像的三个关键特征：亮度、对比度和结构。其中，亮度可以以平均灰度衡量，通过平均所有像素的像素值得到。对比度可以通过灰度标准差来衡量。结构可以基于相关性系数来衡量。

类似地，上述基于该待处理图像对应的低通滤波图像和模板图像对应的低通滤波图像，计算该待处理图像与模板图像的低频相似度，可以包括：计算该待处理图像对应的低通滤波图像与模板图像对应的低通滤波图像的结构相似性，以作为该待处理图像与模板图像的低频相似度。

对于待处理图像是样本图像的情况，将样本图像对应的高通滤波图像与模板图像对应的高通滤波图像进行 SSIM 相似性计算，可以得到图像高频相似度（ISH）。

对于待处理图像是样本图像的情况，将样本图像对应的低通滤波图像与模板图像对应的低通滤波图像进行 SSIM 相似性计算，可以得到图像低频相似度（ISL）。

对于待处理图像是样本图像的特征图像的情况，将样本图像的特征图像对应的高通滤波图像与模板图像的特征图像对应的高通滤波图像进行 SSIM 相似性计算，可以得到特征高频相似度（FSH）。

对于待处理图像是样本图像的特征图像的情况，将样本图像的特征图像对应的低通滤波图像与模板图像的特征图像对应的低通滤波图像进行 SSIM 相似性计算，可以得到特征低频相似度（FSL）。

可以理解，由于待处理图像包括多个样本图像中的每个样本图像和/或其特征图像，所以，待处理图像中包含有模板图像和/或其特征图像，因此在进行结构相似性计算时，当计算模板图像与其自身的相似度的情况或者计算模板图像的特征图像与其自身的相似度的情况下，其高频相似度和低频相似度可以为 1。

在上述实施例中，通过结构相似性计算进行图像去重，能够更符合人眼的观感，从而提高去重的准确性。

在一些实施例中，频率相似度包括高频相似度和低频相似度。

上述步骤 S140 基于频率相似度，确定该样本图像和模板图像的总相似度，包括以下步骤 S141 至步骤 S143。

在步骤 S141 中，计算该样本图像的特征图像与模板图像的特征图像的高频相似度和该样本图像与模板图像的高频相似度的乘积，以作为第一相似度。

在步骤 S142 中，计算该样本图像的特征图像与模板图像的特征图像的低频相似度和该样本图像与模板图像的低频相似度的乘积，以作为第二相似度。

在步骤 S143 中，计算第一相似度与第二相似度的和，以作为该样本图像和模板图像的总相似度。

具体地，总相似度 Y 可以通过如下公式进行计算：

$$Y=ISH\times FSH+ISL\times FSL$$

其中，ISH 表示为图像高频相似度；FSH 表示为特征高频相似度；ISL 表示为图像低频相似度；FSL 表示为特征低频相似度。图像高频相似度是样本图像与模板图像的高频相似度。特征高频相似度是样本图像的特征图像与模板图像的特征图像的高频相似度。图像低频相似度是样本图像与模板图像的低频相似度。特征低频相似度是样本图像的特征图像与模板图像的特征图像的低频相似度。

在上述实施例中，总相似度的计算不仅考虑了样本图像自身，还考虑了样本图像的特征图像，并且，不仅考虑了图像间的高频相似度，还考虑了图像间的低频相似度。由此，保证了图像去重的效果，此外，上述计算方式简单、计算速度较快、实现成本较低，可以适用于需要快速处理大量图像数据的场景。并且，总相似度计算方法可以适用于各种不同类型的图像，具有较强的普适性。

在一些实施例中，上述步骤 S150 根据多个样本图像中每个样本图像分别与模板图

像的总相似度，在多个样本图像中确定预设数量的目标图像，包括：首先，对多个样本图像分别与模板图像的总相似度进行聚类，其中，所聚类别为预设数量；然后，根据聚类结果，在每个所聚类别的样本图像中分别选择一个目标图像。

可以使用 K 均值聚类（K-Means Clustering）算法对多个样本图像分别与模板图像的总相似度进行聚类。K 均值聚类将数据点分成 K 个簇，其中 K 是预先指定的参数。在本申请实施例中，K 为预设的目标图像的数量，即图像去重操作后，所剩余的图像的数量。K 均值聚类算法通过迭代计算来移动簇的中心，直到找到最优的聚类。在本申请的实施例中，对聚类的方法不做任何限定。凡是可以实现数据的分组的方法均在本申请保护范围之内。

示例性地，以样本图像的数量为 200 个，去重比例设置为 90% 为例，则预设数量为 $200 \times (1 - 0.9) = 20$ 个。由此对总相似度进行 K 均值聚类，将样本图像的总相似度分为 20 个簇，且每个簇的中心为簇中所含值的均值。可以根据该簇的中心，在每个簇中选取与该中心最近的总相似度。根据所选取的总相似度，可以确定与之对应的样本图像。由此，在每个簇中确定了一个样本图像。由所有簇中所确定样本图像构成最终确定的经图像去重后所保留的 20 个图像。

在上述实施例中，使用聚类的方式对样本图像进行去重操作，可以简单而高效的处理大量的样本图像，有效地保证了处理速度。并且，可以进一步地提高了图像去重方法的准确性。

根据本申请的另一方面，还提供了一种图像去重处理装置。图 3 示出了根据本申请的一个实施例的图像去重处理装置的示意性框图。如图 3 所示，图像去重处理装置 300 包括选取模块 310、获取模块 320、确定模块 330、对比模块 340 和去重模块 350。

选取模块 310 用于从多个样本图像中确定模板图像。获取模块 320 用于针对多个样本图像中的每个样本图像，获取待处理图像的频谱图，其中，该待处理图像包括该样本图像和/或该样本图像的特征图像。确定模块 330 用于基于该待处理图像的频谱图和模板图像的频谱图，确定该待处理图像与模板图像的频率相似度。对比模块 340 用于基于频率相似度，确定该样本图像和模板图像的总相似度。去重模块 350 用于根据多个样本图像中每个样本图像分别与模板图像的总相似度，在多个样本图像中确定预设数量的目标图像。

根据本申请又一方面，还提供了一种电子设备。图 4 示出了根据本申请的一个实施例的电子设备 400 的示意性框图。如图 4 所示，电子设备 400 包括处理器 410 和存储器 420。其中，存储器 420 中存储有计算机程序指令，计算机程序指令被处理器运行时用于执行上述图像去重处理方法。

根据本申请另一方面，还提供了一种存储介质。在存储介质上存储了程序指令。在程序指令被计算机或处理器运行时使得计算机或处理器执行本申请实施例的上述图像去重处理方法的相应步骤，并且用于实现根据本申请实施例的上述图像去重处理装置的相应模块或上述电子设备中的相应模块。存储介质例如可以包括平板电脑的存储部件、个人计算机的硬盘、只读存储器(ROM)、可擦除可编程只读存储器(EPROM)、

便携式紧致盘只读存储器(CD-ROM)、USB 存储器、或者上述存储介质的任意组合。计算机可读存储介质可以是一个或多个计算机可读存储介质的任意组合。

本领域普通技术人员通过阅读上述关于图像去重处理方法的具体描述，能够理解上述图像去重处理装置、存储介质和电子设备的具体实现和有益效果，为了简洁，在此不再赘述。

实施例：

实施例 1：一种图像去重处理方法，其中，所述方法包括：

从多个样本图像中确定模板图像；

针对所述多个样本图像中的每个样本图像，

获取待处理图像的频谱图，其中，该待处理图像包括该样本图像和/或该样本图像的特征图像；

基于该待处理图像的频谱图和所述模板图像的频谱图，确定该待处理图像与所述模板图像的频率相似度；

基于所述频率相似度，确定该样本图像和所述模板图像的总相似度；

根据所述多个样本图像中每个样本图像分别与所述模板图像的总相似度，在所述多个样本图像中确定预设数量的目标图像。

实施例 2：根据实施例 1 所述的方法，其中，所述从多个样本图像中确定模板图像，包括：

计算所述多个样本图像中不同样本图像之间的余弦相似度；

针对所述多个样本图像中的每个样本图像，统计第一样本图像的个数，其中所述第一样本图像与该样本图像的余弦相似度大于相似度阈值；

比较针对每个样本图像所统计的第一样本图像的个数，并将针对其所统计的第一样本图像的个数最多的样本图像确定为所述模板图像。

实施例 3：根据实施例 1 或 2 所述的方法，其中，所述频率相似度包括高频相似度和/或低频相似度；

所述获取该待处理图像的频谱图包括：

对该待处理图像执行傅里叶变换，以获得该待处理图像的频谱图；

所述基于该待处理图像的频谱图和所述模板图像的频谱图，确定与该待处理图像与所述模板图像的频率相似度，包括：

确定该待处理图像的频谱图中的高通滤波区域和/或低通滤波区域，并且确定所述模板图像的频谱图中的高通滤波区域和/或低通滤波区域；

对所确定的高通滤波区域和/或低通滤波区域进行傅里叶逆变换，以获得对应的高通滤波图像和/或低通滤波图像；

基于该待处理图像对应的高通滤波图像和所述模板图像对应的高通滤波图像，计

算该待处理图像与所述模板图像的高频相似度，和/或，基于该待处理图像对应的低通滤波图像和所述模板图像对应的低通滤波图像，计算该待处理图像与所述模板图像的低频相似度。

5 实施例 4：根据实施例 1-3 中任一实施例介绍的方法，其中，所述确定该待处理图像的频谱图中的高通滤波区域和/或低通滤波区域，并且确定所述模板图像的频谱图中的高通滤波区域和/或低通滤波区域，包括：

对于该待处理图像的频谱图或所述模板图像的频谱图，
在该频谱图中，确定基准点，其中，所述基准点是该频谱图的中心点；
10 在该频谱图中，确定以所述基准点为中心点的矩形区域为低通滤波区域，所确定的低通滤波区域以外区域为高通滤波区域。

实施例 5：根据实施例 1-4 中任一实施例介绍的方法，其中，所述基于该待处理图像对应的高通滤波图像和所述模板图像对应的高通滤波图像，计算该待处理图像与所述模板图像的高频相似度，包括：

15 计算该待处理图像对应的高通滤波图像与所述模板图像对应的高通滤波图像的结构相似性，以作为该待处理图像与所述模板图像的高频相似度；
和/或

所述基于该待处理图像对应的低通滤波图像和所述模板图像对应的低通滤波图像，计算该待处理图像与所述模板图像的低频相似度，包括：

20 计算该待处理图像对应的低通滤波图像与所述模板图像对应的低通滤波图像的结构相似性，以作为该待处理图像与所述模板图像的低频相似度。

实施例 6：根据实施例 1-5 中任一实施例介绍的方法，其中，所述频率相似度包括高频相似度和低频相似度；

所述基于所述频率相似度，确定该样本图像和所述模板图像的总相似度，包括：
25 计算该样本图像的特征图像与所述模板图像的特征图像的高频相似度和该样本图像与所述模板图像的高频相似度的乘积，以作为第一相似度；

计算该样本图像的特征图像与所述模板图像的特征图像的低频相似度和该样本图像与所述模板图像的低频相似度的乘积，以作为第二相似度；以及

计算所述第一相似度与所述第二相似度的和，以作为该样本图像和所述模板图像的总相似度。

30 实施例 7：根据实施例 1-6 中任一实施例介绍的方法，其中，所述根据所述多个样本图像中每个样本图像分别与所述模板图像的总相似度，在所述多个样本图像中确定预设数量的目标图像，包括：

对所述多个样本图像分别与所述模板图像的总相似度进行聚类，其中，所聚类别为所述预设数量；

35 根据聚类结果，在每个所聚类别的样本图像中分别选择一个目标图像。

实施例 8：一种图像去重处理装置，包括：

选取模块，用于从多个样本图像中确定模板图像；

获取模块，用于针对所述多个样本图像中的每个样本图像，获取待处理图像的频谱图，其中，该待处理图像包括该样本图像和/或该样本图像的特征图像；

5 确定模块，用于基于该待处理图像的频谱图和所述模板图像的频谱图，确定该待处理图像与所述模板图像的频率相似度；

对比模块，用于基于所述频率相似度，确定该样本图像和所述模板图像的总相似度；

10 去重模块，用于根据所述多个样本图像中每个样本图像分别与所述模板图像的总相似度，在所述多个样本图像中确定预设数量的目标图像。

实施例 9：一种存储介质，在所述存储介质上存储了程序指令，其中，所述程序指令在运行时用于执行如实施例 1 至 7 中任一实施例介绍的图像去重处理方法。

15 实施例 10：一种电子设备，包括处理器和存储器，其中，所述存储器中存储有计算机程序指令，所述计算机程序指令被所述处理器运行时用于执行如实施例 1 至 7 中任一实施例介绍的图像去重处理方法。

尽管这里已经参考附图描述了示例实施例，应理解上述示例实施例仅仅是示例性的，并且不意图将本申请的范围限制于此。本领域普通技术人员可以在其中进行各种改变和修改，而不偏离本申请的范围和精神。所有这些改变和修改意在被包括在所附权利要求所要求的本申请的范围之内。

20 本领域普通技术人员可以意识到，结合本文中所公开的实施例描述的各示例的单元及算法步骤，能够以电子硬件、或者计算机软件和电子硬件的结合来实现。这些功能究竟以硬件还是软件方式来执行，取决于技术方案的特定应用和设计约束条件。专业技术人员可以对每个特定的应用来使用不同方法来实现所描述的功能，但是这种实现不应认为超出本申请的范围。

25 在本申请所提供的几个实施例中，应该理解到，所揭露的设备和方法，可以通过其它的方式实现。例如，以上所描述的设备实施例仅仅是示意性的，例如，所述单元的划分，仅仅为一种逻辑功能划分，实际实现时可以有另外的划分方式，例如多个单元或组件可以结合或者可以集成到另一个设备，或一些特征可以忽略，或不执行。

30 在此处所提供的说明书中，说明了大量具体细节。然而，能够理解，本申请的实施例可以在没有这些具体细节的情况下实践。在一些实例中，并未详细示出公知的方法、结构和技术，以便不模糊对本说明书的理解。

35 类似地，应当理解，为了精简本申请并帮助理解各个发明方面中的一个或多个，在对本申请的示例性实施例的描述中，本申请的各个特征有时被一起分组到单个实施例、图、或者对其的描述中。然而，并不应将该本申请的方法解释成反映如下意图：即所要求保护的本申请要求比在每个权利要求中所明确记载的特征更多的特征。更确切地说，如相应的权利要求书所反映的那样，其发明点在于可以用少于某个公开的单

个实施例的所有特征的特征来解决相应的技术问题。因此，遵循具体实施方式的权利要求书由此明确地并入该具体实施方式，其中每个权利要求本身都作为本申请的单独实施例。

5 本领域的技术人员可以理解，除了特征之间相互排斥之外，可以采用任何组合对本说明书（包括伴随的权利要求、摘要和附图）中公开的所有特征以及如此公开的任何方法或者设备的所有过程或单元进行组合。除非另外明确陈述，本说明书（包括伴随的权利要求、摘要和附图）中公开的每个特征可以由提供相同、等同或相似目的的替代特征来代替。

10 此外，本领域的技术人员能够理解，尽管在此所述的一些实施例包括其它实施例中有所包括的某些特征而不是其它特征，但是不同实施例的特征的组合意味着处于本申请的范围之内并且形成不同的实施例。例如，在权利要求书中，所要求保护的实施例的任意之一都可以以任意的组合方式来使用。

15 本申请的各个部件实施例可以以硬件实现，或者以在一个或者多个处理器上运行的软件模块实现，或者以它们的组合实现。本领域的技术人员应当理解，可以在实践中使用微处理器或者数字信号处理器（DSP）来实现根据本申请实施例的图像去重处理装置中的一些模块的一些或者全部功能。本申请还可以实现为用于执行这里所描述的方法的一部分或者全部的装置程序（例如，计算机程序和计算机程序产品）。这样的实现本申请的程序可以存储在计算机可读介质上，或者可以具有一个或者多个信号的形式。这样的信号可以从因特网网站上下载得到，或者在载体信号上提供，或者以任何
20 其他形式提供。

25 应该注意的是上述实施例对本申请进行说明而不是对本申请进行限制，并且本领域技术人员在不脱离所附权利要求的范围的情况下可设计出替换实施例。在权利要求中，不应将位于括号之间的任何参考符号构造成对权利要求的限制。单词“包含”不排除存在未列在权利要求中的元件或步骤。位于元件之前的单词“一”或“一个”不排除存在多个这样的元件。本申请可以借助于包括有若干不同元件的硬件以及借助于适当编程的计算机来实现。在列举了若干装置的单元权利要求中，这些装置中的若干个可以通过同一个硬件项来具体体现。单词第一、第二、以及第三等的使用不表示任何顺序。可将这些单词解释为名称。

30 以上所述，仅为本申请的具体实施方式或对具体实施方式的说明，本申请的保护范围并不局限于此，任何熟悉本技术领域的技术人员在本申请揭露的技术范围内，可轻易想到变化或替换，都应涵盖在本申请的保护范围之内。本申请的保护范围应以权利要求的保护范围为准。

权 利 要 求

- 1.一种图像去重处理方法，其特征在于，包括：
- 从多个样本图像中确定模板图像；
- 针对所述多个样本图像中的每个样本图像，
- 5 获取待处理图像的频谱图，其中，该待处理图像包括该样本图像和/或该样本图像的特征图像；
- 基于该待处理图像的频谱图和所述模板图像的频谱图，确定该待处理图像与所述模板图像的频率相似度；
- 基于所述频率相似度，确定该样本图像和所述模板图像的总相似度；
- 10 根据所述多个样本图像中每个样本图像分别与所述模板图像的总相似度，在所述多个样本图像中确定预设数量的目标图像。
- 2.根据权利要求1所述的方法，其特征在于，所述从多个样本图像中确定模板图像，包括：
- 计算所述多个样本图像中不同样本图像之间的余弦相似度；
- 15 针对所述多个样本图像中的每个样本图像，统计第一样本图像的个数，其中所述第一样本图像与该样本图像的余弦相似度大于相似度阈值；
- 比较针对每个样本图像所统计的第一样本图像的个数，并将针对其所统计的第一样本图像的个数最多的样本图像确定为所述模板图像。
- 3.根据权利要求1所述的方法，其特征在于，所述频率相似度包括高频相似度和/
- 20 或低频相似度；
- 所述获取该待处理图像的频谱图包括：
- 对该待处理图像执行傅里叶变换，以获得该待处理图像的频谱图；
- 所述基于该待处理图像的频谱图和所述模板图像的频谱图，确定与该待处理图像与所述模板图像的频率相似度，包括：
- 25 确定该待处理图像的频谱图中的高通滤波区域和/或低通滤波区域，并且确定所述模板图像的频谱图中的高通滤波区域和/或低通滤波区域；
- 对所确定的高通滤波区域和/或低通滤波区域进行傅里叶逆变换，以获得对应的高通滤波图像和/或低通滤波图像；
- 基于该待处理图像对应的高通滤波图像和所述模板图像对应的高通滤波图像，计
- 30 算该待处理图像与所述模板图像的高频相似度，和/或，基于该待处理图像对应的低通滤波图像和所述模板图像对应的低通滤波图像，计算该待处理图像与所述模板图像的低频相似度。
- 4.根据权利要求3所述的方法，其特征在于，所述确定该待处理图像的频谱图中的高通滤波区域和/或低通滤波区域，并且确定所述模板图像的频谱图中的高通滤波区域
- 35 和/或低通滤波区域，包括：

- 对于该待处理图像的频谱图或所述模板图像的频谱图，
在该频谱图中，确定基准点，其中，所述基准点是该频谱图的中心点；
在该频谱图中，确定以所述基准点为中心点的矩形区域为低通滤波区域，所确定的低通滤波区域以外区域为高通滤波区域。
- 5 5.根据权利要求3所述的方法，其特征在于，
所述基于该待处理图像对应的高通滤波图像和所述模板图像对应的高通滤波图像，计算该待处理图像与所述模板图像的高频相似度，包括：
计算该待处理图像对应的高通滤波图像与所述模板图像对应的高通滤波图像的结构相似性，以作为该待处理图像与所述模板图像的高频相似度；
- 10 和/或
所述基于该待处理图像对应的低通滤波图像和所述模板图像对应的低通滤波图像，计算该待处理图像与所述模板图像的低频相似度，包括：
计算该待处理图像对应的低通滤波图像与所述模板图像对应的低通滤波图像的结构相似性，以作为该待处理图像与所述模板图像的低频相似度。
- 15 6.根据权利要求1至5任一项所述的方法，其特征在于，所述频率相似度包括高频相似度和低频相似度；
所述基于所述频率相似度，确定该样本图像和所述模板图像的总相似度，包括：
计算该样本图像的特征图像与所述模板图像的特征图像的高频相似度和该样本图像与所述模板图像的高频相似度的乘积，以作为第一相似度；
- 20 计算该样本图像的特征图像与所述模板图像的特征图像的低频相似度和该样本图像与所述模板图像的低频相似度的乘积，以作为第二相似度；以及
计算所述第一相似度与所述第二相似度的和，以作为该样本图像和所述模板图像的总相似度。
- 25 7.根据权利要求1所述的方法，其特征在于，所述根据所述多个样本图像中每个样本图像分别与所述模板图像的总相似度，在所述多个样本图像中确定预设数量的目标图像，包括：
对所述多个样本图像分别与所述模板图像的总相似度进行聚类，其中，所聚类别为所述预设数量；
根据聚类结果，在每个所聚类别的样本图像中分别选择一个目标图像。
- 30 8.一种图像去重处理装置，包括：
选取模块，用于从多个样本图像中确定模板图像；
获取模块，用于针对所述多个样本图像中的每个样本图像，获取待处理图像的频谱图，其中，该待处理图像包括该样本图像和/或该样本图像的特征图像；
确定模块，用于基于该待处理图像的频谱图和所述模板图像的频谱图，确定该待
- 35 处理图像与所述模板图像的频率相似度；

对比模块，用于基于所述频率相似度，确定该样本图像和所述模板图像的总相似度；

去重模块，用于根据所述多个样本图像中每个样本图像分别与所述模板图像的总相似度，在所述多个样本图像中确定预设数量的目标图像。

5 9.一种存储介质，在所述存储介质上存储了程序指令，其特征在于，所述程序指令在运行时用于执行如权利要求 1 至 7 任一项所述的图像去重处理方法。

10.一种电子设备，包括处理器和存储器，其特征在于，所述存储器中存储有计算机程序指令，所述计算机程序指令被所述处理器运行时用于执行如权利要求 1 至 7 任一项所述的图像去重处理方法。