

人脸和人手关联检测方法、装置、电子设备和存储介质

5 本公开要求在2019年09月18日提交中国专利局、申请号为201910882139.6、申请名称为“人脸和人手关联检测方法、装置、电子设备和存储介质”的中国专利申请的优先权，其全部内容通过引用结合在本公开中。

技术领域

本公开涉及计算机视觉技术领域，尤其涉及一种人脸和人手位置方法及装置、电子设备和存储介质。

10 背景技术

人体人脸人手关联是指将检测出的人脸和人手进行关联，从而通过这个关联的信息来将人手进行的某项操作与某个具体的人对应起来。

因为人体中人脸与人手之间有较远的距离，无法通过位置信息直接进行关联。所以，现有的技术中，通常使用关键点技术以及物体检测技术，来将对应的人脸框、人手框进行关联。

15 发明内容

本公开提出了一种图像处理中检测人脸和人手的技术方案。

根据本公开的一方面，提供了一种人脸和人手关联检测方法，其包括：获取第一图像，所述第一图像为人物对象的图像；对所述第一图像执行特征提取，得到多个尺度的第一特征图；对所述多个尺度的第一特征图执行特征融合处理，得到多个尺度的第二特征图，所述第二特征图的尺度与所述第一特征图的尺度一一对应；基于得到的所述多个尺度的第二特征图检测所述第一图像中针对同一人物对象的关联的人脸位置和人手位置。基于上述配置，本公开实施例可以简单方便的得到图像中相互关联的人脸和人手，同时还可以提高检测精度。

20 在一些可能的实施方式中，所述获取第一图像，包括：获取所述第二图像，所述第二图像为包括至少一个物体对象的图像；对所述第二图像执行人体目标检测，得到所述第一图像中所述至少一个物体对象中任一物体对象的检测框；将所述任一物体对象的所述检测框在所述第二图像中对应的图像区域，确定为所述任一物体对象的第一图像。基于上述配置，本公开实施例得到的第一图像中删除了其

25 他环境因素的影响，能够进一步提高检测精度。

在一些可能的实施方式中，所述对所述第一图像执行特征提取，得到多个尺度的第一特征图，包括：将所述第一图像调整为预设尺度的第三图像；将所述第三图像输入至残差网络，得到所述多个尺度的第一特征图。基于上述配置，可以实现图像的尺度统一，提高适用性。

30 在一些可能的实施方式中，所述对所述多个尺度的第一特征图执行特征融合处理，得到多个尺度的第二特征图，包括：将所述多个尺度的第一特征图输入至特征金字塔网络，通过所述特征金字塔网络执行所述特征融合处理，得到所述多个尺度的第二特征图。基于上述配置，可以提高得到的多个尺度的第二特征图的特征精度。

35 在一些可能的实施方式中，按照尺度从小到大的顺序，所述多个第一特征图表示为 $\{C_1, \dots, C_n\}$ ，其中， n 表示第一特征图的数量， n 为大于1的整数；所述对所述多个尺度的第一特征图执行特征融合处理，得到多个尺度的第二特征图，包括：利用第一卷积核对第一特征图 C_n 执行卷积处理，获得与所述第一特征图 C_n 对应的第二特征图 F_n ，其中，所述第一特征图 C_n 的尺度与所述第二特征图 F_n 的尺度相同；对所述第二特征图 F_n 执行线性插值处理获得与所述第二特征图 F_n 对应的第一中间特征图 F'_n ，其中，所述第一中间特征图 F'_n 的尺度与第一特征图 C_{n-1} 的尺度相同；利用第二卷积核对所述第

40 F'_n

一特征图 C_n 以外的第一特征图 C_i 执行卷积处理, 得到所述第一特征图 C_i 对应的第二中间特征图 C'_i , 所述第二中间特征图 C'_i 的尺度与第一中间特征图 F'_{i+1} 的尺度相同, 其中, i 为大于或者等于 1 且小于 n 的整数变量; 利用所述第二中间特征图 C'_i 和对应的所述第一中间特征图 F'_{i+1} 得到所述第二特征图 F_n 以外的第二特征图 F_i , 其中, 所述第一中间特征图 F'_{i+1} 由对应的所述第二特征图 F_{i+1} 经线性插值得到。基于上述配置, 可以融合不同尺度的特征信息, 进一步提高特征精度。

在一些可能的实施方式中, 所述利用所述第二中间特征图 C'_i 和对应的所述第一中间特征图 F'_{i+1} 得到所述第二特征图 F_n 以外的第二特征图 F_i , 包括: 将所述第二中间特征图 C'_i 与对应的所述第一中间特征图 F'_{i+1} 进行加和处理, 得到所述第二特征图 F_i 。基于上述配置, 可以有效的融合两个中间特征的特征信息。

10 在一些可能的实施方式中, 所述基于得到的所述多个尺度的第二特征图检测所述第一图像中针对同一人物对象的关联的人脸位置和人手位置, 包括: 对所述多个尺度的第二特征图中尺度最大的第二特征图执行卷积处理, 分别得到表示所述人脸位置的掩码图, 以及表示所述人手位置的掩码图; 基于所述人脸位置的掩码图以及所述人手位置的掩码图确定所述第一图像中关联的人手和人脸所在的位置区域。基于上述配置, 可以方便的预测和表示关联的人脸和人手的位置。

15 在一些可能的实施方式中, 所述多个尺度的第一特征图之间的尺度关系为: $L(C_{i-1})=2^{k_1} \cdot L(C_i)$ 且 $W(C_{i-1})=2^{k_1} \cdot W(C_i)$, 其中, C_i 表示各第一特征图, $L(C_i)$ 表示所述第一特征图 C_i 的长度, $W(C_i)$ 表示所述第一特征图 C_i 的宽度, k_1 为大于或者等于 1 的整数, i 为变量, 且 i 的范围为 $[2, n]$, n 表示第一特征图的数量。

20 在一些可能的实施方式中, 所述方法还包括以下方式中的至少一种: 在所述第一图像中突出显示所述关联的人手和人脸; 为所述第一图像中检测到的关联的人脸位置和人手位置分配相同的标签。基于上述配置, 可以直观的体现关联的人脸和人手所在的图像区域, 同时有效的区分不同人物对象的关联检测结果。

25 在一些可能的实施方式中, 所述方法通过神经网络实现, 其中, 训练所述神经网络的步骤包括: 获取训练图像, 所述训练图像为包括人物对象的图像, 所述训练图像具有真实关联的人脸位置和人手位置的标注信息; 将所述训练图像输入至所述神经网络, 通过所述神经网络预测所述训练图像中针对同一人物对象的关联的人脸位置和人手位置; 基于预测出的关联的所述人脸位置以及人手位置以及所述标注信息确定网络损失, 并根据所述网络损失调整所述神经网络的网络参数, 直至满足训练要求。基于上述配置, 可以实现神经网络的优化训练, 保证网络检测精度。

30 根据本公开的第二方面, 提供了一种人脸和人手关联检测装置, 其包括: 获取模块, 用于获取第一图像, 所述第一图像为物体对象的图像; 特征提取模块, 用于对所述第一图像执行特征提取, 得到多个尺度的第一特征图; 融合模块, 用于对所述多个尺度的第一特征图执行特征融合处理, 得到多个尺度的第二特征图, 所述第二特征图的尺度与所述第一特征图的尺度一一对应; 检测模块, 用于基于得到的所述多个尺度的第二特征图检测所述第一图像中针对同一物体对象的关联的人脸位置和人手位置。

在一些可能的实施方式中,所述获取模块包括:获取单元,用于获取所述第二图像,所述第二图像为包括至少一个物体对象的图像;目标检测单元,用于对所述第二图像执行人体目标检测,得到所述第一图像中所述至少一个物体对象中任一物体对象的检测框;确定单元,用于将所述任一物体对象的所述检测框在所述第二图像中对应的图像区域,确定为所述任一物体对象的第一图像。

- 5 在一些可能的实施方式中,所述特征提取模块还用于将所述第一图像调整为预设尺度的第三图像;将所述第三图像输入至残差网络,得到所述多个尺度的第一特征图。

在一些可能的实施方式中,所述融合单元还用于将所述多个尺度的第一特征图输入至特征金字塔网络,通过所述特征金字塔网络执行所述特征融合处理,得到所述多个尺度的第二特征图。

在一些可能的实施方式中,按照尺度从小到大的顺序,所述多个第一特征图表示为 $\{C_1, \dots, C_n\}$,

- 10 其中, n 表示第一特征图的数量, n 为大于1的整数;所述融合模块还用于利用第一卷积核对第一特征图 C_n 执行卷积处理,获得与所述第一特征图 C_n 对应的第二特征图 F_n ,其中,所述第一特征图 C_n 的尺度与所述第二特征图 F_n 的尺度相同;对所述第二特征图 F_n 执行线性插值处理获得与所述第二特征图 F_n 对应的第一中间特征图 F'_n ,其中,所述第一中间特征图 F'_n 的尺度与第一特征图 C_{n-1} 的尺度相同;利用第二卷积核将所述第一特征图 C_n 以外的第一特征图 C_i 执行卷积处理,得到所述第一特征图 C_i
- 15 对应的第二中间特征图 C'_i ,所述第二中间特征图 C'_i 的尺度与第一中间特征图 F'_{i+1} 的尺度相同,其中, i 为大于或者等于1且小于 n 的整数变量;利用所述第二中间特征图 C'_i 和对应的所述第一中间特征图 F'_{i+1} 得到所述第二特征图 F_n 以外的第二特征图 F_i ,其中,所述第一中间特征图 F'_{i+1} 由对应的所述第二特征图 F_{i+1} 经线性插值得到。

- 20 在一些可能的实施方式中,所述融合模块还用于将所述第二中间特征图 C'_i 与对应的所述第一中间特征图 F'_{i+1} 进行加和处理,得到所述第二特征图 F_i 。

在一些可能的实施方式中,所述检测模块还用于对所述多个尺度的第二特征图中尺度最大的第二特征图执行卷积处理,分别得到表示所述人脸位置的掩码图,以及表示所述人手位置的掩码图;基于所述人脸位置的掩码图以及所述人手位置的掩码图确定所述第一图像中关联的人手和人脸所在的位置区域。

- 25 在一些可能的实施方式中,所述多个尺度的第一特征图之间的尺度关系为: $L(C_{i-1})=2^{k_1} \cdot L(C_i)$ 且 $W(C_{i-1})=2^{k_1} \cdot W(C_i)$,其中, C_i 表示各第一特征图, $L(C_i)$ 表示所述第一特征图 C_i 的长度, $W(C_i)$ 表示所述第一特征图 C_i 的宽度, k_1 为大于或者等于1的整数, i 为变量,且 i 的范围为 $[2, n]$, n 表示第一特征图的数量。

- 30 在一些可能的实施方式中,所述装置还包括显示模块和分配模块中的至少一种,其中所述显示模块,用于在所述第一图像中突出显示所述关联的人手和人脸;所述分配模块,用于为所述第一图像中检测到的关联的人脸位置和人手位置分配相同的标签。

5 在一些可能的实施方式中，所述装置包括神经网络，所述特征提取模块、所述融合模块和所述检测模块应用所述神经网络，所述装置还包括训练模块，用于训练所述神经网络，其中，训练所述神经网络的步骤包括：获取训练图像，所述训练图像为包括人物对象的图像，所述训练图像具有真实关联的人脸位置和人手位置的标注信息；将所述训练图像输入至所述神经网络，通过所述神经网络预测所述训练图像中针对同一人物对象的关联的人脸位置和人手位置；基于预测出的关联的所述人脸位置以及人手位置以及所述标注信息确定网络损失，并根据所述网络损失调整所述神经网络的网络参数，直至满足训练要求。

10 根据本公开的第三方面，提供了一种电子设备，其包括：处理器；用于存储处理器可执行指令的存储器；其中，所述处理器被配置为调用所述存储器存储的指令，以执行第一方面中任意一项所述的方法。

根据本公开的第四方面，提供了一种计算机可读存储介质，其上存储有计算机程序指令，所述计算机程序指令被处理器执行时实现第一方面中任意一项所述的方法。

15 根据本公开的第五方面，提供了一种计算机程序，所述计算机程序包括计算机可读代码，当所述计算机可读代码在电子设备中运行时，所述电子设备中的处理器执行用于实现第一方面中任意一项所述的方法。

20 本公开实施例，可以从第一图像中确定一个人物对象所在的区域对应的第一图像，并对第一图像进行特征提取处理得到相应的特征图，而后对特征图进行多尺度的特征融合处理，得到多个尺度的第二特征图，其中第二特征图相对于第一特征图具有更精确的特征信息，通过对第二特征图进行处理可以得到第一图像中关联的人手和人脸的位置，提高人脸和人手检测精度。另外，本公开实施例的技术方案不需要获取人耳或者手腕的关键点，可以直接得到图像中关联的人手和人脸的位置，具有简单方便且精度高的特点。

应当理解的是，以上的一般描述和后文的细节描述仅是示例性和解释性的，而非限制本公开。

根据下面参考附图对示例性实施例的详细说明，本公开的其它特征及方面将变得清楚。

附图说明

25 此处的附图被并入说明书中并构成本说明书的一部分，这些附图示出了符合本公开的实施例，并与说明书一起用于说明本公开的技术方案。

图1示出根据本公开实施例的一种人脸和人手关联检测方法的流程图；

图2示出根据本公开实施例的一种人脸和人手关联检测方法中步骤S10的流程图；

图3示出根据本公开实施例的第二图像的示意图；

30 图4示出根据本公开实施例的一种人脸和人手关联检测方法的步骤S20的流程图；

图5示出根据本公开实施例的一种人脸和人手关联检测方法中步骤S30的流程图；

图6示出根据本公开实施例的特征提取和特征融合过程的示意图；

图7示出根据本公开实施例的一种人脸和人手关联检测方法中步骤S40的流程图；

图8示出根据本公开实施例的训练神经网络的流程图；

35 图9示出根据本公开实施例的一种人脸和人手关联检测装置的框图；

图10示出根据本公开实施例的一种电子设备的框图；

图11示出根据本公开实施例的另一种电子设备的框图。

具体实施方式

40 以下将参考附图详细说明本公开的各种示例性实施例、特征和方面。附图中相同的附图标记表示功能相同或相似的元件。尽管在附图中示出了实施例的各种方面，但是除非特别指出，不必按比例绘制附图。

在这里专用的词“示例性”意为“用作例子、实施例或说明性”。这里作为“示例性”所说明的任何实施例不必解释为优于或好于其它实施例。

45 本文中术语“和/或”，仅仅是一种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。另外，本文中术语“至少一

种”表示多种中的任意一种或多种中的至少两种的任意组合，例如，包括A、B、C中的至少一种，可以表示包括从A、B和C构成的集合中选择的任意一个或多个元素。

另外，为了更好地说明本公开，在下文的具体实施方式中给出了众多的具体细节。本领域技术人员应当理解，没有某些具体细节，本公开同样可以实施。在一些实例中，对于本领域技术人员熟知的方法、手段、元件和电路未作详细描述，以便于凸显本公开的主旨。

本公开实施例提供了一种人脸和人手关联检测方法，其可以应用在任意的图像处理装置中，例如，该方法可以应用在终端设备或服务器中，或者也可以应用在其它处理设备中，其中，终端设备可以包括用户设备（User Equipment, UE）、移动设备、用户终端、终端、蜂窝电话、无绳电话、个人数字处理（Personal Digital Assistant, PDA）、手持设备、计算设备、车载设备、可穿戴设备等。在一些可能的实现方式中，该人脸和人手关联检测方法可以通过处理器调用存储器中存储的计算机可读指令的方式来实现。

图1示出根据本公开实施例的一种人脸和人手关联检测方法的流程图，如图1所示，所述人脸和人手关联检测方法包括：

S10：获取第一图像；

在一些可能的实施方式中，第一图像可以为人物对象的图像，其中可以包括至少一个人脸和至少一个手，本公开实施例可以实现该第一图像中的人物对象的人手和人脸的关联检测，该关联是指得到的人脸和人手为同一人物对象的人脸和人手。

在一些可能的实施方式中，获取第一图像的方式可以包括：直接通过图像采集设备采集第一图像，其中图像采集设备可以为手机、摄像头、照相机等具有图像采集功能的设备。获取第一图像的方式也可以包括从其他设备接收传输的第一图像，或者从存储器中读取第一图像，或者第一图像也可以为从视频流中执行选帧操作后得到的图像帧，本公开对此不作具体限定。

在另一些可能的实施方式中，第一图像也可以为其他图像的部分图像区域，例如第一图像可以为通过接收的选择信息从其他图像中选择出的图像区域，或者也可以为通过目标检测的方式，如检测人体，检测得到的图像区域，本公开对此不作具体限定。

S20：对所述第一图像执行特征提取，得到多个尺度的第一特征图；

在一些可能的实施方式中，本公开实施例可以对第一图像执行特征提取处理，得到多个尺度的第一特征图，例如，本公开实施例可以将第一图像输入至特征提取网络，得到多个尺度的第一特征图，其中特征提取网络可以为卷积神经网络，如残差网络（Res-Net），通过该残差网络执行第一图像的特征提取，得到至少两个尺度的第一特征图。或者，在其他实施例中也可以采用其他类型的特征提取网络，得到该多个尺度的第一特征图，本公开对此不作具体限定。或者，在另一些可能的实施方式中，也可以通过对第一图像进行升采样或者降采样的方式得到多个尺度的第一特征图，例如通过不同的采样率得到相应的多个尺度的第一特征图。

S30：对所述多个尺度的第一特征图执行特征融合处理，得到多个尺度的第二特征图，所述第二特征图的尺度与所述第一特征图的尺度一一对应；

在一些可能的实施方式中，在得到多个尺度的第一特征图的情况下，可以对该多个尺度的第一特征图执行特征融合处理，得到相应尺度的第二特征图。其中，通过特征融合可以提高每个第二特征图内特征信息的精确度，从而可以进一步提高人脸和人手的关联检测精度。

在一些可能的实施方式中，可以通过特征金字塔网络执行该多个尺度的第一特征图的特征融合处理，其中，可以对相邻尺度的第一特征图的特征信息进行特征融合，通过依次从小尺度的第一特征图的特征信息融合到大尺度的第一特征图的特征信息，最终可以得到融合了所有尺度的第一特征图中的特征信息的第二特征图。

S40：基于得到的所述多个尺度的第二特征图检测所述第一图像中针对同一人物对象的关联的人脸位置和人手位置。

在一些可能的实施方式中，在得到多个尺度的第二特征图之后，可以基于该多个尺度的第二特征图执行人脸和人手的关联检测。其中，可以对各个尺度的第二特征图中的至少一个第二特征图执行卷

积处理,从而得到第一图像中关联的人脸位置和人手位置。例如,可以将尺度最大的第二特征图输入至卷积层执行卷积处理,分别得到关于人脸位置和人手位置的掩码图,其中可以包括一个人脸位置的第一掩码图,以及左手位置的第二掩码图,和右手位置的第三掩码图,通过得到的各掩码图可以对应的在第一图像中确定出关联的人手和人脸位置。

- 5 基于上述配置,本公开实施例可以不需要获取人耳或者手腕的关键点,以及也不需要分析是否满足高斯分布,可以直接通过第一图像的特征的多尺度提取以及特征融合得到关联的人手和人脸,具有简单方便且精度高的特点。

下面结合附图,对本公开实施例的过程进行详细说明。如上述实施例所述,本公开实施例得到的第一图像可以为人物对象的图像,其中,在实际应用过程中,得到的图像中可能包括多个人物对象,10 为了提高同一人物对象的人脸和人手的关联检测精度,本公开可以从得到的图像中得到每个人物对象的图像区域,再分别对每个图像区域执行特征提取和特征融合,最终得到每个人物对象的人脸和人手位置。图2示出根据本公开实施例的一种人脸和人手关联检测方法中步骤S10的流程图。所述获取第一图像,包括:

S101:获取第二图像,所述第二图像为包括至少一个人物对象的图像;

- 15 在一些可能的实施方式中,第一图像可以是基于第二图像获得的图像,其中,第二图像可以为至少一个人物对象的图像。获取第二图像的方式可以包括:直接通过图像采集设备采集第一图像,其中图像采集设备可以为手机、摄像头、照相机等具有图像采集功能的设备。获取第二图像的方式也可以包括从其他设备接收传输的第二图像,或者从存储器中读取第二像,或者第二图像也可以为从视频流中执行选帧操作后得到的图像帧,本公开对此不作具体限定。

- 20 图3示出根据本公开实施例的第二图像的示意图。其中,可以包括5个人物对象A、B、C、D和E。在其他实施例中,第二图像也可以仅包括一个人物对象,或者也可以包括其他数量的人物对象,本公开对此不作具体限定。

S102:对所述第二图像执行人体目标检测,得到所述第二图像中所述至少一个人物对象中任一15 人物对象的检测框;

- 25 在一些可能的实施方式中,在通过第二图像得到第一图像时,可以检测第一图像中针对每个人物对象的人体区域的位置,得到与该人物对象对应的第一图像,其中在第二图像中包括多个人物对象时,在得到的第一图像中可以包括一个人物对象的人体区域,同时也可以其他人物对象的至少一部分图像,如包括其他对象的人脸或者人手的至少一部分。本公开实施例通过对第一图像进行后续处理得到第一图像中为一个人物对象的人手和人脸。

- 30 如上所述,第二图像中可以包括至少一个人物对象,本公开可以对该第二图像执行目标检测,实现第二图像中人物对象的人体区域检测,得到每个人物对象的检测框。

在一些可能的实施方式中,可以通过能够执行人体目标检测神经网络检测第二图像中人物对象对应的检测框,该神经网络可以为卷积神经网络,其可以是经过训练能够精确的识别出图像中的每个人物对象,以及相应人物对象的位置区域(即检测框)的卷积神经网络,例如可以为R-CNN网络,或者35 也可以为其他能够实现目标检测的神经网络,本公开对此不作具体限定。

如图3所示,通过目标检测处理,得到了图像中的人物对象的人体区域对应的检测框,例如人物对象A的检测框A1,以及人物对象D的检测框D1,上述仅为示例性说明,也可以检测出其他人物对象的检测框。

- 其中,在得到检测框的过程中,可以识别图像中每个人物对象的检测框,也可以识别满足质量要求的检测框,例如图3中对于人物对象B、C和D,得到的检测框的质量值小于质量阈值,此时可以将人物对象B、C和D对应的检测框确定为不满足质量要求的检测框,做删除处理。其中检测框的质量值可以是在执行目标检测处理时,在得到检测框同时得到的关于该检测框的得分或者置信度,在该得分或者置信度大于质量阈值的情况下,则确定检测框满足质量要求。其中质量阈值可以为设定的数值,如80%,或者也可以为其他小于1的数值,本公开对此不作具体限定。

- 45 S103:将所述任一人物对象的所述检测框在所述第二图像中的图像区域,确定为所述任一人物对

象对应的第一图像。

在得到第二图像中每个人物对象的检测框的情况下，可以将第二图像中与检测框对应的图像区域，确定为该检测框对应的人物对象的第一图像。例如如图3示出的实施例，可以得到第二图像中人物对象A的检测框A1，以及人物对象D的检测框D1。对应的可以将A1对应的图像区域确定为人物对象A的第一图像，以及检测框D1对应的图像区域确定为人物对象D的第一图像。

基于上述配置，本公开实施例得到的第一图像中删除了其他环境因素的影响，能够进一步提高检测精度。另外基于上述可以从第二图像中得到针对一个人物对象的图像区域（第一图像），虽然得到的第一图像为针对一个人物对象的图像，但是在实际应用过程中，由于第二图像中包括的各人物可能相近，此时得到的第一图像中也可能包括其他人物对象的至少一部分，例如，图3中的检测框D1除了包括人物对象D还可以包括人物C的人脸的一部分，本公开可以通过后续的处理过程得到第一图像中为同一人物对象的人脸和人手的位置。

图4示出根据本公开实施例的一种人脸和人手关联检测方法的步骤S20的流程图，其中所述对所述第一图像执行特征提取，得到多个尺度的第一特征图，包括：

S201：将所述第一图像调整为预设规格的第三图像；

在一些可能的实施方式中，得到的第一图像的尺度可能不同，本公开实施例可以将得到的第一图像调整为同一尺度，即调整为预设尺度，从而可以对相同尺度的图像执行后续的特征提取处理。其中，本公开实施例的预设尺度可以根据网络的设计和配置确定，例如本公开实施例的预设尺度可以为256*192（高度*宽度），但不作为本公开的具体限定。

其中，调整图像尺度的方式可以包括、上采样、降采样、图像插值中的至少一种，本公开对此也不作具体限定，也可以通过其他方式得到预设尺度的第三图像。

S202：将所述第三图像输入至残差网络，得到所述多个尺度的第一特征图。

在得到预设尺度的第三图像的情况下，可以对第三图像执行特征提取处理，例如可以将第三图像输入至残差网络（如Resnet50）执行图像的特征提取处理，得到不同尺度的第一特征图。其中，可以通过残差网络的不同卷积层输出不同尺度的第一特征图。

或者，在其他实施方式中，也可以通过其他特征提取网络得到该多尺度的第一特征图，如金字塔特征提取网络，或者通过升采样或者降采样的方式得到多尺度的第一特征图，例如本公开实施例的采样频率可以为1/8、1/16、1/32等，但本公开实施例对此不进行限定。

在一些可能的实施方式中，得到的各第一特征图之间的关系为 $L(C_{i-1}) = 2^{k_1} \cdot L(C_i)$ 且

$W(C_{i-1}) = 2^{k_1} \cdot W(C_i)$ ，其中， C_i 表示各第一特征图， $L(C_i)$ 表示第一特征图 C_i 的长度， $W(C_i)$ 表示

第一特征图 C_i 的宽度， k_1 为大于或者等于1的整数， i 为变量，且 i 的范围为 $[2, n]$ ， n 为第一特征图的数量。即本公开实施例中的各第一特征图的长度和宽度之间的关系均为2的 k_1 次方倍。

在一个示例中，本公开得到的第一特征图的数量可以为4个，可以分别表示为第一特征图

C_1 、 C_2 、 C_3 和 C_4 ，其中，第一特征图 C_1 的长度和宽度可以分别对应的为第一特征图 C_2 的长度和宽

度的二倍，第二特征图 C_2 的长度和宽度可以分别对应的为第三特征图 C_3 的长度和宽度的二倍，以及

第三特征图 C_3 的长度和宽度可以分别对应的为第四特征图 C_4 的长度和宽度的二倍。本公开实施例上

述 C_1 和 C_2 之间、 C_2 和 C_3 之间，以及 C_3 和 C_4 之间的长度倍数以及宽度倍数均相同，即 k_1 取值为1。

在其他的实施例中， k_1 可以为不同的值，例如可以为：第一特征图 C_1 的长度和宽度可以分别对应的为

第一特征图 C_2 的长度和宽度的二倍，第二特征图 C_2 的长度和宽度可以分别对应的为第三特征图 C_3 的长度和宽度的四倍，以及第三特征图 C_3 的长度和宽度可以分别对应的为第四特征图 C_4 的长度和宽度的八倍。本公开实施例对此不进行限定。

在得到第一图像对应的多个尺度的第一特征图的情况下，可以进一步执行各第一特征图的特征融合处理，提高得到的第二特征图的特征信息的精确度。

在一些可能的实施方式中，对第一特征图执行特征融合处理，可以利用金字塔特征提取网络（FPN）执行。即可以将多个尺度的第一特征图输入至特征金字塔网络，通过所述特征金字塔网络执行所述特征融合处理，得到第一特征图对应的第二特征图。或者也可以通过其他方式执行特征融合处理，例如可以通过卷积处理和上采样处理得到多个尺度第二特征图。基于上述配置，可以提高得到的多个尺度的第二特征图的特征精度。

图5示出根据本公开实施例的一种人脸和人手关联检测方法中步骤S30的流程图，其中所述对所述多个尺度的第一特征图执行特征融合处理，得到多个尺度的第二特征图，包括：

S301：利用第一卷积核对第一特征图 C_n 执行卷积处理，获得与第一特征图 C_n 对应的第二特征图 F_n ，其中，第一特征图 C_n 的尺度与第二特征图 F_n 的尺度相同；

在一些可能的实施方式中，本公开实施例得到的第一特征图可以表示成 $\{C_1, \dots, C_n\}$ ，即 n 个第一特征图，且 C_n 可以为长度和宽度最小的特征图，即尺度最小的第一特征图。其中随着 n 值越大对应的第一特征图的尺度就越小，例如上述第一特征图 C_1 、 C_2 、 C_3 和 C_4 ，尺度依次降低。

在执行特征融合处理时，可以首先得到尺度最小的第一特征图 C_n 对应的第二特征图 F_n 。例如，可以通过第一卷积核对第一特征图 C_n 执行卷积处理，得到第一特征图 C_n 对应的第二特征图 F_n ，其中，第一特征图 C_n 的尺度与第二特征图 F_n 的尺度相同。同样的，第二特征图 F_n 也是第二特征图中尺度最小的特征图。通过第一卷积核执行的卷积处理可以得到相对于第一特征图 C_n 的特征信息更精确的第二特征图 F_n 。其中，第一卷积核可以为 3×3 的卷积核，或者也可以是其他类型的卷积核。

S302：对所述第二特征图 F_n 执行线性插值处理获得与第二特征图 F_n 对应的第一中间特征图 F_n' ，其中第一中间特征图 F_n' 的尺度与第一特征图 C_{n-1} 的尺度相同；

在得到第二特征图 F_n 之后，可以利用该第二特征图 F_n 获得与其对应的第一中间特征图 F_n' ，本公开实施例可以通过对第二特征图 F_n 执行线性插值处理获得与第二特征图 F_n 对应的第一中间特征图 F_n' ，其中，第一中间特征图 F_n' 的尺度与第一特征图 C_{n-1} 的尺度相同，例如，在 C_{n-1} 的尺度为 C_n 的尺度

的二倍时，第一中间特征图 F'_n 的长度为第二特征图 F_n 的长度的二倍，以及第一中间特征图 F'_n 的宽度为第二特征图 F_n 的宽度的二倍。

S303：利用第二卷积核对第一特征图 C_n 以外的第一特征图 C_i 执行卷积处理，得到所述第一特征图 C_i 对应的第二中间特征图 C'_i ，所述第二中间特征图 C'_i 的尺度与第一中间特征图 F'_{i+1} 的尺度相同，其中， i 为大于或者等于 1 且小于 n 的整数变量；

在一些可能的实施方式中，可以获得第一特征图 C_n 以外的各第一特征图 $C_1 \dots C_{n-1}$ 对应的第二中间特征图 $C'_1 \dots C'_{n-1}$ ，其中，可以利用第二卷积核分别对第一特征图 $C_1 \dots C_{n-1}$ 进行卷积处理，分别得到与各第一特征图 $C_1 \dots C_{n-1}$ 一一对应的第二中间特征图 $C'_1 \dots C'_{n-1}$ ，其中第二卷积核可以为 1×1 的卷积核，但本公开对此不作具体限定。通过第二卷积核的卷积处理得到的各第二中间特征图的尺度与对应的第一特征图的尺度分别相同。其中，本公开实施例可以按照第一特征图 $C_1 \dots C_{n-1}$ 的倒序，获得各第一特征图 $C_1 \dots C_{n-1}$ 的第二中间特征图 $C'_1 \dots C'_{n-1}$ 。即，可以先获得第一特征图 C_{n-1} 对应的第二中间特征图 C'_{n-1} ，而后获得第一特征图 C_{n-2} 的对应的第二中间图 C'_{n-2} ，以此类推，直至获得第一特征图 C_1 对应的第二中间特征图 C'_1 。

S304：利用所述第二中间特征图 C'_i 和对应的第一中间特征图 F'_{i+1} 得到所述第二特征图 F_n 以外的第二特征图 F_i ，其中，第一中间特征图 F'_{i+1} 由对应的第二特征图 F_{i+1} 经线性插值得到。

在获得各第二中间特征图的，或者获得各第二中间特征图之后还可以对应的获得第一中间特征图 F'_n 以外的其他第一中间特征图 $F'_1 \dots F'_{n-1}$ ，本公开实施例中，与第一特征图 $C_1 \dots C_{n-1}$ 中的第一特征图 C_i 对应的第二特征图 $F_i = C'_i + F'_{i+1}$ ，其中，第二中间特征图 C'_i 的尺度（长度和宽度）分别与第一中间特征图 F'_{i+1} 的尺度（长度和宽度）相等，并且第二中间特征图 C'_i 的长度和宽度与第一特征图 C_i 的长度和宽度相同，因此得到的第二特征图 F_i 的长度和宽度分别为第一特征图 C_i 的长度和宽度。其中， i 为大于或者等于 1 且小于 n 的整数。

具体的，本公开实施例依然可以采用倒序的处理方式获得第二特征图 F_n 以外的各第二特征图 F_i 。

即，本公开实施例可以首先获得第一中间特征图 F_{n-1} ，其中，可以利用第一特征图 C_{n-1} 对应的第二中间图 C'_{n-1} 与第一中间特征图 F'_n 进行加和处理得到第二特征图 F_{n-1} ，其中，第二中间特征图 C'_{n-1} 的长度

- 和宽度分别与第一中间特征图 F_n' 的长度和宽度相同，以及第二特征图 F_{n-1} 的长度和宽度为第二中间特征图 C_{n-1}' 和 F_n' 的长度和宽度。此时第二特征图 F_{n-1} 的长度和宽度分别为第二特征图 F_n 的长度和宽度的二倍（ C_{n-1} 的尺度为 C_n 的尺度的二倍）。进一步地，可以对第二特征图 F_{n-1} 进行线性插值处理，得到第一中间特征图 F_{n-1}' ，使得 F_{n-1}' 的尺度与 C_{n-1} 的尺度相同，继而可以利用第一特征图 C_{n-2} 对应的第二中间特征图 C_{n-2}' 与第一中间特征图 F_{n-1}' 进行加和处理得到第二特征图 F_{n-2} ，其中，第二中间特征图 C_{n-2}' 的长度和宽度分别与第一中间特征图 F_{n-1}' 的长度和宽度相同，以及第二特征图 F_{n-2} 的长度和宽度为第二中间特征图 C_{n-2}' 和 F_{n-1}' 的长度和宽度。例如第二特征图 F_{n-2} 的长度和宽度分别为第二特征图 F_{n-1} 的长度和宽度的二倍。以此类推，可以最终获得第一中间特征图 F_2' ，以及根据该第一中间特征图 F_2' 与第一特征图 C_1' 的加和处理得到第二特征图 F_1 ， F_1 的长度和宽度分别为与 C_1 的长度和宽度的相同。从而得到各第二特征图，并满足 $L(F_{i-1})=2^{k_1} \cdot L(F_i)$ 及 $W(F_{i-1})=2^{k_1} \cdot W(F_i)$ ，并且 $L(F_n)=L(C_n)$ ， $W(F_n)=W(C_n)$ 。

- 例如，以上述四个第一特征图 C_1 、 C_2 、 C_3 和 C_4 为例进行说明。图6示出根据本公开实施例的特征提取和特征融合过程的示意图。其中，可以通过残差网络a执行特征提取处理，以及利用残差网络中的四个卷积层分别输出四个不同尺度的第一特征图 C_1 、 C_2 、 C_3 和 C_4 ，而后利用特征提取网络b执行特征融合处理，获得多尺度的第二特征图。其中，首先可以将 C_4 经过一个 3×3 的第一卷积核计算得到一个新的特征图 F_4 （第二特征图）， F_4 的长度和宽度的大小与 C_4 相同。对 F_4 进行双线性插值的上采样（upsample）操作，得到一个长和宽都放大两倍的特征图，即第一中间特征图 F_4' 。 C_3 经过一个 1×1 的第二卷积核计算得到一个第二中间特征图 C_3' ， C_3' 与 F_4' 大小相同，两个中间特征图相加，得到新的特征图 F_3 （第二特征图），使得第二特征图 F_3 的长度和宽度分别为第二特征图 F_4 二倍，同时与第一特征图 C_3 的尺度相同。对 F_3 进行双线性插值的上采样（upsample）操作，得到一个长和宽都放大两倍的特征图，即第一中间特征图 F_3' 。 C_2 经过一个 1×1 的第二卷积核计算得到一个第二中间特征图 C_2' ， C_2' 与 F_3' 大小相同，两个中间特征图相加，得到新的特征图 F_2 （第二特征图），使得第二特征图 F_2 的长度和宽度分别为第二特征图 F_3 二倍。对 F_2 进行双线性插值的上采样（upsample）操作，得到一个长和宽都放大两倍的特征图，即第一中间特征图 F_2' 。 C_1 经过一个 1×1 的第二卷积核计算得到一个第二中间特征图 C_1' ， C_1' 与 F_2' 大小相同，两个特征图相加，得到新的特征图 F_1 （第二特征图），使得第二特征图 F_1 的长度和宽度分别为第二特征图 F_2 二倍。经过FPN之后，同样得到了四个不同尺度的第二特征图，分别记为 F_1 、 F_2 、 F_3 和 F_4 。并且 F_1 和 F_2 之间的长度和宽度的倍数与 C_1 和 C_2 之间的长

度和宽度的倍数相同,以及 F_2 和 F_3 之间的长度和宽度的倍数与 C_2 和 C_3 之间的长度和宽度的倍数相同, F_3 和 F_4 之间的长度和宽度的倍数与 C_3 和 C_4 之间的长度和宽度的倍数相同。

基于上述配置,可以融合不同尺度的特征信息,进一步提高特征精度。通过上述方式可以得到与多个尺度的第一特征图分别对应的第二特征图,第二特征图的特征信息相对于第一特征图的特征信息提高了精确度。

在得到第二特征图的情况下,可以根据第二特征图得到第一图像中针对同一人物对象的人脸和人手的位置。图7示出根据本公开实施例的一种人脸和人手关联检测方法中步骤S40的流程图。如上述实施例所述,本公开实施例中基于得到的所述多个尺度的第二特征图检测所述第一图像中针对同一人物对象的关联的人脸位置和人手位置,包括:

10 S401:对所述多个尺度的第二特征图中尺度最大的第二特征图执行卷积处理,分别得到表示所述人脸位置的掩码图,以及表示所述人手位置的掩码图;

在一些可能的实施方式中,可以将得到的多个尺度的第二特征图中的至少一个第二特征图输入至卷积层中,对该至少一个第二特征图执行进一步的特征融合,并对应的生成第一图像对应的同一人物对象的人脸位置的掩码图以及人手位置的掩码图。其中,由于尺度最高的第二特征图融合了各个尺度的特征图的特征信息,本公开可以将第二特征图输入至该卷积层中执行人手和人脸位置的关联检测。其中,得到的掩码图中的元素可以表示为由1和0构成,其中1表示人手或者人脸的位置区域。例如,本公开实施例可以得到同一人物对象的人脸位置的第一掩码图,左手位置的第二掩码图以及右手位置的第三掩码图,通过各掩码图中元素1的位置,即可以得到相应的关联的人脸和人手在第一图像中的位置。

20 在一些可能的实施方式中,如果只能检测到左手和右手中的一个手,则检测不到的人手对应的掩码图可以为全0掩码图。或者,如果检测不到关联的人脸和人手,则输出的掩码图也可以为全0掩码图。

在一些可能的实施方式中,得到的掩码图可以对应关联有人物对象标识以及类型标识,其中人物对象标识用于区分不同的人物对象,不同的人物对象可以具有不同的人物对象标识,类型标识可以用于表示掩码图对应的人脸位置、左手位置或者右手位置。通过上述人物对象标识以及类型标识可以清楚

25 S402:基于所述人脸位置的掩码图以及所述人手位置的掩码图确定所述第一图像中关联的人手和人脸所在的位置区域。

在得到关联的人手和人脸对应的掩码图的情况下,进一步得到第一图像中的关联的人手和人脸对应的位置区域。

30 本公开实施例得到的第一掩码图和第二掩码图的尺度可以和第一图像相同,从而可以根据掩码图确定的人脸位置映射到第一图像中相应的人脸图像区域,以及根据掩码图确定的人手位置映射到第一图像中的人手图像区域,进而得到关联的人手和人脸所在的位置区域。

在一些可能的实施方式中,在检测到所述第一图像中的关联的人脸和人手的位置后,可以基于得到的掩码图在第一图像中突出显示匹配的人脸和人手,例如将掩码图在第一图像中图像区域以检测框的方式表示,以提示关联的人脸和人手。如图3所示,在图像中可以显示人物对象D关联的人脸检测框D11和人手检测框D12和D13。同时,本公开实施例还可以为相关联的人脸和人手分配相同的标签,用以标识该人脸和人手为同一个人物对象的人脸和人手。

基于上述配置可以方便的预测和表示关联的人脸和人手的位置。

40 在一些可能的实施方式中,本公开实施例得到的关联人脸和人手的位置还可以用于确定人物对象姿态变化。例如,第一图像可以为基于视频流中的图像帧得到的,通过本公开实施例的方法可以检测出图像帧中针对同一任务对象的人脸位置的变化以及人手位置的变化,更进一步的,还可以通过对相应图像帧中的人脸位置进行表情识别,或者基于人手位置执行手势识别,从而可以得到表情的变化情况,或者手势的变化情况。

如上述实施例所述,本公开实施例中的人脸和人手关联检测方法可以应用在神经网络中,如卷积神经网络中,例如可以由残差网络和金字塔网络构建形成上述卷积神经网络。本公开还可以对神经网络

络进行训练得到满足精度要求的神经网络。图8示出根据本公开实施例的训练神经网络的流程图。其中，所述训练神经网络可以包括：

S501：获取训练图像，所述训练图像为包括人物对象的图像，所述训练图像具有真实关联的人脸位置和人手位置的标注信息；

- 5 在一些可能的实施方式中，训练图像可以为一个人物对象的图像，同时训练图像中也可以包括其余人物对象的人脸或者人手中的一部分，从而可以提高训练精度。其中训练图像的数量为多个，本公开对训练图像的数量不作限制。

- 10 在一些可能的实施方式中，训练图像可以关联有真实的标注信息，用以监督神经网络的训练。其中，每个训练图像均具有真实关联的人脸位置和人手位置的标注信息，用以表示训练图像中针对同一人物对象的人脸位置和人手位置（左手和右手），其中，标注信息可以表示为标注框，或者也可以表示为位置坐标的形式，或者也可以表示成真实的关联的人手和人脸位置的掩码图，只要是能够确定训练图像中的关联的人脸位置和人手位置就可以作为本公开实施例。S502：将所述训练图像输入至所述神经网络，通过所述神经网络预测所述训练图像中针对同一人物对象的关联的人脸位置和人手位置；

- 15 在一些可能的实施方式，可以将训练图像输入至神经网络，进行特征提取、特征融合、以及关联的人手和人脸位置的检测。

例如，可以通过残差网络等特征提取网络执行训练图像的多尺度特征提取，得到多个尺度的第一预测特征图，具体特征提取的过程可以参照上述实施例的说明，在此不再重复说明。

- 20 在得到多个尺度的第一特征图之后，可以对该多个尺度的第一预测特征图执行特征融合处理，例如利用金字塔网络FPN执行该多个第一预测特征图的特征融合，得到多个尺度的第二预测特征图，其中，特征融合的具体过程在此也不作重复说明，具体可以参照上述实施例的过程。

在得到多个第二预测特征图的情况下，可以基于每个第二预测特征图执行卷积处理，得到基于每个第二预测特征图预测的关联的人脸和人手的位置的预测掩码图。

S503：基于针对所述训练图像预测出的关联的所述人脸位置以及人手位置以及所述标注信息确定网络损失，并根据所述网络损失调整所述神经网络的网络参数，直至满足训练要求。

- 25 本公开实施例可以根据各个尺度的第二预测特征图预测得到的人脸预测掩码图和人手预测掩码图与真实的人脸和人手的对应的掩码图之间的差异，得到网络损失，其中可以通过对数损失函数确定网络损失。例如，本公开实施例可以直接利用对数损失函数处理，得到每个尺度的第二预测特征图得到的预测掩码图和标注的真实掩码图之间的损失，并将该损失作为网络损失，调整神经网络的参数。也就是说，可以将每个尺度对应的损失都作为网络损失，单独的优化神经网络参数。

- 30 或者，在其他实施方式中，本公开实施例可以通过对数损失函数确定每个尺度的第二预测特征图得到的人脸预测掩码图、人手预测掩码图与真实标注信息对应的掩码图之间的子网络损失，并利用各尺度对应得到的子网络损失的加权和确定网络损失。也就是说，可以根据每个尺度对应的损失的加权和确定网络损失，用以一起优化神经网络参数。

- 35 另外，由于本公开实施例可以基于每个第二预测特征图的预测结果得到网络损失，因此得到的神经网络无论哪个尺度的第二预测特征图的预测结果的精度都会较高，进而可以提高整个神经网络的检测精度。

- 40 在得到网络损失的情况下，基于网络损失和损失阈值的比较结果调整神经网络的网络参数，例如，在网络损失大于损失阈值的情况下，反馈调整神经网络的参数，如调整特征提取网络、金字塔特征网络以及得到掩码图的卷积层的参数，重新对训练图像进行处理，直至得到的网络参数小于损失阈值。以及在网络损失小于损失阈值的情况下，可以确定为神经网络满足训练要求，此时可以终止训练。基于上述配置，可以实现神经网络的优化训练，保证网络检测精度。

- 45 综上所述，本公开实施例，可以从第一图像中确定一个人体对象所在的区域对应的第一图像，并对第一图像进行特征提取处理得到相应的特征图，而后对特征图进行多尺度的特征融合处理，得到多个尺度的第二特征图，其中第二特征图相对于第一特征图具有更精确的特征信息，通过对第二特征图进行处理可以得到第一图像中关联的人手和人脸的位置，提高人脸和人手检测精度。另外，本公开实

施例的技术方案不需要获取人耳或者手腕的关键点，可以直接得到图像中关联的人手和人脸的位置，具有简单方便且精度高的特点。

本领域技术人员可以理解，在具体实施方式的上述方法中，各步骤的撰写顺序并不意味着严格的执行顺序而对实施过程构成任何限定，各步骤的具体执行顺序应当以其功能和可能的内在逻辑确定。

可以理解，本公开提及的上述各个方法实施例，在不违背原理逻辑的情况下，均可以彼此相互结合形成结合后的实施例，限于篇幅，本公开不再赘述。

此外，本公开还提供了人脸和人手关联检测装置、电子设备、计算机可读存储介质、程序，上述均可用来实现本公开提供的任一种人脸和人手关联检测方法，相应技术方案和描述和参见方法部分的相应记载，不再赘述。

图9示出根据本公开实施例的一种人脸和人手关联检测装置的框图，如图9所示，所述人脸和人手关联检测装置包括：

获取模块10，用于获取第一图像，所述第一图像为对象对象的图像；

特征提取模块20，用于对所述第一图像执行特征提取，得到多个尺度的第一特征图；

融合模块30，用于对所述多个尺度的第一特征图执行特征融合处理，得到多个尺度的第二特征图，所述第二特征图的尺度与所述第一特征图的尺度一一对应；

检测模块40，用于基于得到的所述多个尺度的第二特征图检测所述第一图像中针对同一对象对象的关联的人脸位置和人手位置。

在一些可能的实施方式中，所述获取模块包括：

获取单元，用于获取所述第二图像，所述第二图像为包括至少一个对象对象的图像；

目标检测单元，用于对所述第二图像执行人体目标检测，得到所述第一图像中所述至少一个对象对象中任一对象对象的检测框；

确定单元，用于将所述任一对象对象的所述检测框在所述第二图像中对应的图像区域，确定为所述任一对象对象的第一图像。

在一些可能的实施方式中，所述特征提取模块还用于获取所述第二图像，所述第二图像为包括至少一个对象对象的图像；

对所述第二图像执行人体目标检测，得到所述第一图像中所述至少一个对象对象中任一对象对象的检测框；

将所述任一对象对象的所述检测框在所述第二图像中对应的图像区域，确定为所述任一对象对象的第一图像。

在一些可能的实施方式中，所述融合单元还用于将所述多个尺度的第一特征图输入至特征金字塔网络，通过所述特征金字塔网络执行所述特征融合处理，得到所述多个尺度的第二特征图。

在一些可能的实施方式中，按照尺度从小到大的顺序，所述多个第一特征图表示为 $\{C_1, \dots, C_n\}$ ，

其中， n 表示第一特征图的数量， n 为大于1的整数；

所述融合模块还用于利用第一卷积核对第一特征图 C_n 执行卷积处理，获得与所述第一特征图 C_n

对应的第二特征图 F_n ，其中，所述第一特征图 C_n 的尺度与所述第二特征图 F_n 的尺度相同；

对所述第二特征图 F_n 执行线性插值处理获得与所述第二特征图 F_n 对应的第一中间特征图 F'_n ，其

中，所述第一中间特征图 F'_n 的尺度与第一特征图 C_{n-1} 的尺度相同；

利用第二卷积核对所述第一特征图 C_n 以外的第一特征图 C_1 执行卷积处理，得到所述第一特征图

C_i 对应的第二中间特征图 C'_i ，所述第二中间特征图 C'_i 的尺度与第一中间特征图 F'_{i+1} 的尺度相同，其中， i 为大于或者等于 1 且小于 n 的整数变量；

利用所述第二中间特征图 C'_i 和对应的所述第一中间特征图 F'_{i+1} 得到所述第二特征图 F_n 以外的第二特征图 F_i ，其中，所述第一中间特征图 F'_{i+1} 由对应的所述第二特征图 F_{i+1} 经线性插值得到。

在一些可能的实施方式中，所述融合模块还用于将所述第二中间特征图 C'_i 与对应的所述第一中间特征图 F'_{i+1} 进行加和处理，得到所述第二特征图 F_i 。

在一些可能的实施方式中，所述检测模块还用于对所述多个尺度的第二特征图中尺度最大的第二特征图执行卷积处理，分别得到表示所述人脸位置的掩码图，以及表示所述人手位置的掩码图；

基于所述人脸位置的掩码图以及所述人手位置的掩码图确定所述第一图像中关联的人手和人脸所在的位置区域。

在一些可能的实施方式中，所述多个尺度的第一特征图之间的尺度关系为： $L(C_{i-1}) = 2^{k_1} \cdot L(C_i)$ 且

$W(C_{i-1}) = 2^{k_1} \cdot W(C_i)$ ，其中， C_i 表示各第一特征图， $L(C_i)$ 表示所述第一特征图 C_i 的长度， $W(C_i)$

表示所述第一特征图 C_i 的宽度， k_1 为大于或者等于 1 的整数， i 为变量，且 i 的范围为 $[2, n]$ ， n 表示第一

特征图的数量。

在一些可能的实施方式中，所述装置还包括显示模块和分配模块中的至少一种，其中

所述显示模块，用于在所述第一图像中突出显示所述关联的人手和人脸；

所述分配模块，用于为所述第一图像中检测到的关联的人脸位置和人手位置分配相同的标签。

在一些可能的实施方式中，所述装置包括神经网络，所述特征提取模块、所述融合模块和所述检测模块应用所述神经网络，

所述装置还包括训练模块，用于训练所述神经网络，其中，训练所述神经网络的步骤包括：

获取训练图像，所述训练图像为包括人物对象的图像，所述训练图像具有真实关联的人脸位置和人手位置的标注信息；

将所述训练图像输入至所述神经网络，通过所述神经网络预测所述训练图像中针对同一人物对象的关联的人脸位置和人手位置；

基于预测出的关联的所述人脸位置以及人手位置以及所述标注信息确定网络损失，并根据所述网络损失调整所述神经网络的网络参数，直至满足训练要求。

在一些实施例，本公开实施例提供的装置具有的功能或包含的模块可以用于执行上文方法实施例描述的方法，其具体实现可以参照上文方法实施例的描述，为了简洁，这里不再赘述。

本公开实施例还提出一种计算机可读存储介质，其上存储有计算机程序指令，所述计算机程序指令被处理器执行时实现上述方法。计算机可读存储介质可以是非易失性计算机可读存储介质。

本公开实施例还提出一种电子设备，包括：处理器；用于存储处理器可执行指令的存储器；其中，所述处理器被配置为上述方法。

电子设备可以被提供为终端、服务器或其它形态的设备。

图10示出根据本公开实施例的一种电子设备的框图。例如，电子设备800可以是移动电话，计算机，数字广播终端，消息收发设备，游戏控制台，平板设备，医疗设备，健身设备，个人数字助理等终端。

参照图10, 电子设备800可以包括以下一个或多个组件: 处理组件802, 存储器804, 电源组件806, 多媒体组件808, 音频组件810, 输入/输出(I/O)的接口812, 传感器组件814, 以及通信组件816。

5 处理组件802通常控制电子设备800的整体操作, 诸如与显示, 电话呼叫, 数据通信, 相机操作和记录操作相关联的操作。处理组件802可以包括一个或多个处理器820来执行指令, 以完成上述的方法的全部或部分步骤。此外, 处理组件802可以包括一个或多个模块, 便于处理组件802和其他组件之间的交互。例如, 处理组件802可以包括多媒体模块, 以方便多媒体组件808和处理组件802之间的交互。

存储器804被配置为存储各种类型的数据以支持在电子设备800的操作。这些数据的示例包括用于在电子设备800上操作的任何应用程序或方法的指令, 联系人数据, 电话簿数据, 消息, 图片, 视频等。存储器804可以由任何类型的易失性或非易失性存储设备或者它们的组合实现, 如静态随机存取存储器(SRAM), 电可擦除可编程只读存储器(EEPROM), 可擦除可编程只读存储器(EPROM), 可编程只读存储器(PROM), 只读存储器(ROM), 磁存储器, 快闪存储器, 磁盘或光盘。

10 电源组件806为电子设备800的各种组件提供电力。电源组件806可以包括电源管理系统, 一个或多个电源, 及其他与为电子设备800生成、管理和分配电力相关联的组件。

多媒体组件808包括在所述电子设备800和用户之间提供一个输出接口的屏幕。在一些实施例中, 15 屏幕可以包括液晶显示器(LCD)和触摸面板(TP)。如果屏幕包括触摸面板, 屏幕可以被实现为触摸屏, 以接收来自用户的输入信号。触摸面板包括一个或多个触摸传感器以感测触摸、滑动和触摸面板上的手势。所述触摸传感器可以不仅感测触摸或滑动动作的边界, 而且还检测与所述触摸或滑动操作相关的持续时间和压力。在一些实施例中, 多媒体组件808包括一个前置摄像头和/或后置摄像头。当电子设备800处于操作模式, 如拍摄模式或视频模式时, 前置摄像头和/或后置摄像头可以接收外部的多媒体数据。每个前置摄像头和后置摄像头可以是一个固定的光学透镜系统或具有焦距和光学变焦能力。

音频组件810被配置为输出和/或输入音频信号。例如, 音频组件810包括一个麦克风(MIC), 当电子设备800处于操作模式, 如呼叫模式、记录模式和语音识别模式时, 麦克风被配置为接收外部音频信号。所接收的音频信号可以被进一步存储在存储器804或经由通信组件816发送。在一些实施例中, 25 音频组件810还包括一个扬声器, 用于输出音频信号。

I/O接口812为处理组件802和外围接口模块之间提供接口, 上述外围接口模块可以是键盘, 点击轮, 按钮等。这些按钮可包括但不限于: 主页按钮、音量按钮、启动按钮和锁定按钮。

传感器组件814包括一个或多个传感器, 用于为电子设备800提供各个方面的状态评估。例如, 传感器组件814可以检测到电子设备800的打开/关闭状态, 组件的相对定位, 例如所述组件为电子设备800的显示器和小键盘, 传感器组件814还可以检测电子设备800或电子设备800一个组件的位置改变, 用户与电子设备800接触的存在或不存在, 电子设备800方位或加速/减速和电子设备800的温度变化。传感器组件814可以包括接近传感器, 被配置用来在没有任何的物理接触时检测附近物体的存在。传感器组件814还可以包括光传感器, 如CMOS或CCD图像传感器, 用于在成像应用中使用。在一些实施例中, 该传感器组件814还可以包括加速度传感器, 陀螺仪传感器, 磁传感器, 压力传感器或温度传感器。

35 通信组件816被配置为便于电子设备800和其他设备之间有线或无线方式的通信。电子设备800可以接入基于通信标准的无线网络, 如WiFi, 2G或3G, 或它们的组合。在一个示例性实施例中, 通信组件816经由广播信道接收来自外部广播管理系统的广播信号或广播相关信息。在一个示例性实施例中, 所述通信组件816还包括近场通信(NFC)模块, 以促进短程通信。例如, 在NFC模块可基于射频识别(RFID)技术, 红外数据协会(IrDA)技术, 超宽带(UWB)技术, 蓝牙(BT)技术和其他技术来实现。

在示例性实施例中, 电子设备800可以被一个或多个应用专用集成电路(ASIC)、数字信号处理器(DSP)、数字信号处理设备(DSPD)、可编程逻辑器件(PLD)、现场可编程门阵列(FPGA)、控制器、微控制器、微处理器或其他电子元件实现, 用于执行上述方法。

45 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, 例如包括计算机程序指令的存

存储器804, 上述计算机程序指令可由电子设备800的处理器820执行以完成上述方法。

图11示出根据本公开实施例的另一种电子设备的框图。例如, 电子设备1900可以被提供为一服务器。参照图11, 电子设备1900包括处理组件1922, 其进一步包括一个或多个处理器, 以及由存储器1932所代表的存储器资源, 用于存储可由处理组件1922的执行的指令, 例如应用程序。存储器1932中存储的应用程序可以包括一个或一个以上的每一个对应于一组指令的模块。此外, 处理组件1922被配置为执行指令, 以执行上述方法。

电子设备1900还可以包括一个电源组件1926被配置为执行电子设备1900的电源管理, 一个有线或无线网络接口1950被配置为将电子设备1900连接到网络, 和一个输入输出(I/O)接口1958。电子设备1900可以操作基于存储在存储器1932的操作系统, 例如Windows Server™, Mac OS X™, Unix™, Linux™, FreeBSD™或类似。

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, 例如包括计算机程序指令的存储器1932, 上述计算机程序指令可由电子设备1900的处理组件1922执行以完成上述方法。

本公开可以是系统、方法和/或计算机程序产品。计算机程序产品可以包括计算机可读存储介质, 其上载有用于使处理器实现本公开的各个方面的计算机可读程序指令。

计算机可读存储介质可以是可以保持和存储由指令执行设备使用的指令的有形设备。计算机可读存储介质例如可以是——但不限于——电存储设备、磁存储设备、光存储设备、电磁存储设备、半导体存储设备或者上述的任意合适的组合。计算机可读存储介质的更具体的例子(非穷举的列表)包括: 便携式计算机盘、硬盘、随机存取存储器(RAM)、只读存储器(ROM)、可擦式可编程只读存储器(EPROM或闪存)、静态随机存取存储器(SRAM)、便携式压缩盘只读存储器(CD-ROM)、数字多功能盘(DVD)、记忆棒、软盘、机械编码设备、例如其上存储有指令的打孔卡或凹槽内凸起结构、以及上述的任意合适的组合。这里所使用的计算机可读存储介质不被解释为瞬时信号本身, 诸如无线电波或者其他自由传播的电磁波、通过波导或者其他传输媒介传播的电磁波(例如, 通过光纤电缆的光脉冲)、或者通过电线传输的电信号。

这里所描述的计算机可读程序指令可以从计算机可读存储介质下载到各个计算/处理设备, 或者通过网络、例如因特网、局域网、广域网和/或无线网下载到外部计算机或外部存储设备。网络可以包括铜传输电缆、光纤传输、无线传输、路由器、防火墙、交换机、网关计算机和/或边缘服务器。每个计算/处理设备中的网络适配卡或者网络接口从网络接收计算机可读程序指令, 并转发该计算机可读程序指令, 以供存储在各个计算/处理设备中的计算机可读存储介质中。

用于执行本公开操作的计算机程序指令可以是汇编指令、指令集架构(ISA)指令、机器指令、机器相关指令、微代码、固件指令、状态设置数据、或者以一种或多种编程语言的任意组合编写的源代码或目标代码, 所述编程语言包括面向对象的编程语言—诸如Smalltalk、C++等, 以及常规的过程式编程语言—诸如“C”语言或类似的编程语言。计算机可读程序指令可以完全地在用户计算机上执行、部分地在用户计算机上执行、作为一个独立的软件包执行、部分在用户计算机上部分在远程计算机上执行、或者完全在远程计算机或服务器上执行。在涉及远程计算机的情形中, 远程计算机可以通过任意种类的网络—包括局域网(LAN)或广域网(WAN)—连接到用户计算机, 或者, 可以连接到外部计算机(例如利用因特网服务提供商来通过因特网连接)。在一些实施例中, 通过利用计算机可读程序指令的状态信息来个性化定制电子电路, 例如可编程逻辑电路、现场可编程门阵列(FPGA)或可编程逻辑阵列(PLA), 该电子电路可以执行计算机可读程序指令, 从而实现本公开的各个方面。

这里参照根据本公开实施例的方法、装置(系统)和计算机程序产品的流程图和/或框图描述了本公开的各个方面。应当理解, 流程图和/或框图的每个方框以及流程图和/或框图中各方框的组合, 都可以由计算机可读程序指令实现。

这些计算机可读程序指令可以提供给通用计算机、专用计算机或其它可编程数据处理装置的处理单元, 从而生产出一种机器, 使得这些指令在通过计算机或其它可编程数据处理装置的处理单元执行时, 产生了实现流程图和/或框图中的一个或多个方框中规定的功能/动作的装置。也可以把这些计算机可读程序指令存储在计算机可读存储介质中, 这些指令使得计算机、可编程数据处理装置和/或其他设

- 备以特定方式工作，从而，存储有指令的计算机可读介质则包括一个制造品，其包括实现流程图和/或框图中的一个或多个方框中规定的功能/动作的各个方面的指令。
- 也可以把计算机可读程序指令加载到计算机、其它可编程数据处理装置、或其它设备上，使得在计算机、其它可编程数据处理装置或其它设备上执行一系列操作步骤，以产生计算机实现的过程，从而使得在计算机、其它可编程数据处理装置、或其它设备上执行的指令实现流程图和/或框图中的一个或多个方框中规定的功能/动作。
- 附图中的流程图和框图显示了根据本公开的多个实施例的系统、方法和计算机程序产品的可能实现的体系架构、功能和操作。在这点上，流程图或框图中的每个方框可以代表一个模块、程序段或指令的一部分，所述模块、程序段或指令的一部分包含一个或多个用于实现规定的逻辑功能的可执行指令。在有些作为替换的实现中，方框中所标注的功能也可以以不同于附图中所标注的顺序发生。例如，两个连续的方框实际上可以基本并行地执行，它们有时也可以按相反的顺序执行，这依所涉及的功能而定。也要注意的，框图和/或流程图中的每个方框、以及框图和/或流程图中的方框的组合，可以用执行规定的功能或动作的专用的基于硬件的系统来实现，或者可以用专用硬件与计算机指令的组合来实现。
- 以上已经描述了本公开的各实施例，上述说明是示例性的，并非穷尽性的，并且也不限于所披露的各实施例。在不偏离所说明的各实施例的范围和精神的情况下，对于本技术领域的普通技术人员来说许多修改和变更都是显而易见的。本文中所用术语的选择，旨在最好地解释各实施例的原理、实际应用或对市场中的技术的技术改进，或者使本技术领域的其它普通技术人员能理解本文披露的各实施例。