

人脸识别方法及装置

相关申请的交叉引用

本公开基于申请号为 201910220321.5、申请日为 2019 年 3 月 22 日的中国专利申请提出，并要求该中国专利申请的优先权，该中国专利申请的全部内容在此以全文引入的方式引入本公开。

技术领域

本公开实施例涉及图像处理技术领域，尤其涉及一种人脸识别方法及装置。

背景技术

安防、社保、通信等领域需要识别不同图像中包括的人物对象是否是同一个人，以实现面部跟踪、实名认证、手机解锁等操作。目前，通过人脸识别算法对不同图像中的人物对象分别进行人脸识别，可识别不同图像中包括的人物对象是否是同一个人，但识别准确率较低。

发明内容

本公开提供一种人脸识别方法，以识别不同图像中的人物对象是否是同一个人。

第一方面，提供了一种人脸识别方法，包括：获得待识别图像；基于跨模态人脸识别网络对所述待识别图像进行识别，得到所述待识别图像的识别结果，其中，所述跨模态人脸识别网络基于不同模态的人脸图像数据训练得到。

在一种可能实现的方式中，所述基于不同模态的人脸图像数据训练得到所述跨模态人脸识别网络的过程，包括：基于第一模态网络和第二模态网络进行训练得到所述跨模态人脸识别网络。

在另一种可能实现的方式中，在所述基于第一模态网络和第二模态网络进行训练得到所述跨模态人脸识别网络之前，还包括：基于第一图像集和第二图像集对所述第一模态网络训练，其中，所述第一图像集中的对象属于第一类别，所述第二图像集中的对象属于第二类别。

在又一种可能实现的方式中，所述基于第一图像集和第二图像集对所述第一模态网络训练，包括：基于所述第一图像集和所述第二图像集对所述第一模态网络进行训练，得到所述第二模态网络；按预设条件从所述第一图像集中选取第一数目的图像，并从所述第二图像集中选取第二数目的图像，并根据所述第一数目的图像和所述第二数目的图像得到第三图像集；基于所述第三图像集对所述第二模态网络进行训练，得到所述跨模态人脸识别网络。

在又一种可能实现的方式中，所述预设条件包括：所述第一数目与所述第二数目相同，所述第一数目与所述第二数目的比值等于所述第一图像集包含的图像数目与所述第二图像集包含的图像数目的比值，所述第一数目与所述第二数目的比值等于所述第一图像集包含的人数与所述第二图像集包含的人数的比值中的任意一种。

在又一种可能实现的方式中，所述第一模态网络包括第一特征提取分支、第二特征提取分支以及第三特征提取分支；所述基于所述第一图像集和所述第二图像集对所述第一模态网络进行训练，得到所述第二模态网络，包括：将所述第一图像集输入至所述第一特征提取分支，并将所述第二图像集输入至所述第二特征提取分支，并将第四图像集输入至所述第三特征提取分支，对所述第一模态网络进行训练，其中，所述第四图像集包括的图像为同一场景下采集的图像或同一采集方式采集的图像；将训练后的第一特征提取分支或训练后的第二特征提取分支或训练后的第三特征提取分支作为所述第二模

态网络。

在又一种可能实现的方式中，所述将所述第一图像集输入至所述第一特征提取分支，并将所述第二图像集输入至所述第二特征提取分支，并将第四图像集输入至所述第三特征提取分支，对所述第一模态网络进行训练，包括：将所述第一图像集、所述第二图像集以及所述第四图像集分别输入至所述第一特征提取分支、所述第二特征提取分支以及所述第三特征提取分支，分别得到第一识别结果、第二识别结果以及第三识别结果；获取所述

5 所述第一特征提取分支的第一损失函数、所述第二特征提取分支的第二损失函数以及所述第三特征提取分支的第三损失函数；根据所述第一图像集、所述第一识别结果以及所述第一损失函数，所述第二图像集、所述第二识别结果以及所述第二损失函数，所述

10 第四图像集、所述第三识别结果以及所述第三损失函数，调整所述第一模态网络的参数，得到调整后的第一模态网络，其中，所述第一模态网络的参数包括第一特征提取分支参数、第二特征提取分支参数以及第三特征提取分支参数，所述调整后的第一模态网络的各分支参数相同。

在又一种可能实现的方式中，所述第一图像集中的图像包括第一标注信息，所述第二图像集中的图像包括第二标注信息，所述第四图像集中的图像包括第三标注信息；所述根据所述第一图像集、所述第一识别结果以及所述第一损失函数，所述第二图像集、所述第二识别结果以及所述第二损失函数，所述第四图像集、所述第三识别结果以及所述第三损失函数，调整所述第一模态网络的参数，得到调整后的第一模态网络，包括：根据所述

15 所述第一标注信息、所述第一识别结果、所述第一损失函数以及所述第一特征提取分支的初始参数，得到第一梯度，以及根据所述第二标注信息、所述第二识别结果、所述第二损失函数以及所述第二特征提取分支的初始参数，得到第二梯度，以及根据所述

20 第三标注信息、所述第三识别结果、所述第三损失函数以及所述第三特征提取分支的初始参数，得到第三梯度；将所述第一梯度、所述第二梯度以及所述第三梯度的平均值作为所述第一模态网络的反向传播梯度，并通过所述反向传播梯度调整所述第一模态网络的参数，使所述

25 所述第一特征提取分支的参数、所述第二特征提取分支的参数以及所述第三特征提取分支的参数相同。

在又一种可能实现的方式中，所述按预设条件从所述第一图像集中选取第一数量张图像，并从所述第二图像集中选取第二数量张图像，得到第三图像集，包括：从所述第一图像集以及所述第二图像集中分别选取 f 张图像，使所述 f 张图像中包含的人数为阈值，得到所述第三图像集；或，从所述第一图像集以及所述第二图像集中分别选取 m 张

30 图像以及 n 张图像，使所述 m 与所述 n 的比值等于所述第一图像集包含的图像数量与所述第二图像集包含的图像数量的比值，且所述 m 张图像以及所述 n 张图像中包含的人数均为所述阈值，得到所述第三图像集；或，从所述第一图像集以及所述第二图像集中分别选取 s 张图像以及 t 张图像，使所述 s 与所述 t 的比值等于所述第一图像集包含的人

35 数与所述第二图像集包含的人数的比值，且所述 s 张图像以及所述 t 张图像中包含的人数均为所述阈值，得到所述第三图像集。

在又一种可能实现的方式中，所述基于所述第三图像集对所述第二模态网络进行训练，得到所述跨模态人脸识别网络，包括：对所述第三图像集中的图像依次进行特征提取处理、线性变换、非线性变换，得到第四识别结果；根据所述第三图像集中的图像、所述第四识别结果以及所述第二模态网络的第四损失函数，调整所述第二模态网络的参

40 数，得到所述跨模态人脸识别网络。

在又一种可能实现的方式中，所述第一类别以及所述第二类别分别对应不同人种。

第二方面，提供了一种人脸识别装置，包括：获取单元，配置为获得待识别图像；识别单元，配置为基于跨模态人脸识别网络对所述待识别图像进行识别，得到所述待识

别图像的识别结果,其中,所述跨模态人脸识别网络基于不同模态的人脸图像数据训练得到。

在一种可能实现的方式中,所述识别单元包括:训练子单元,配置为基于第一模态网络和第二模态网络进行训练得到所述跨模态人脸识别网络。

5 在另一种可能实现的方式中,所述训练子单元还配置为:基于第一图像集和第二图像集对所述第一模态网络训练,其中,所述第一图像集中的对象属于第一类别,所述第二图像集中的对象属于第二类别。

10 在又一种可能实现的方式中,所述训练子单元还配置为:基于所述第一图像集和所述第二图像集对所述第一模态网络进行训练,得到所述第二模态网络;以及按预设条件从所述第一图像集中选取第一数目的图像,并从所述第二图像集中选取第二数目的图像,并根据所述第一数目的图像和所述第二数目的图像得到第三图像集;以及基于所述第三图像集对所述第二模态网络进行训练,得到所述跨模态人脸识别网络。

15 在又一种可能实现的方式中,所述预设条件包括:所述第一数目与所述第二数目相同,所述第一数目与所述第二数目的比值等于所述第一图像集包含的图像数目与所述第二图像集包含的图像数目的比值,所述第一数目与所述第二数目的比值等于所述第一图像集包含的人数与所述第二图像集包含的人数的比值中的任意一种。

20 在又一种可能实现的方式中,所述第一模态网络包括第一特征提取分支、第二特征提取分支以及第三特征提取分支;所述训练子单元还配置为:将所述第一图像集输入至所述第一特征提取分支,并将所述第二图像集输入至所述第二特征提取分支,并将第四图像集输入至所述第三特征提取分支,对所述第一模态网络进行训练,其中,所述第四图像集包括的图像为同一场景下采集的图像或同一采集方式采集的图像;以及将训练后的第一特征提取分支或训练后的第二特征提取分支或训练后的第三特征提取分支作为所述第二模态网络。

25 在又一种可能实现的方式中,所述训练子单元还配置为:将所述第一图像集、所述第二图像集以及所述第四图像集分别输入至所述第一特征提取分支、所述第二特征提取分支以及所述第三特征提取分支,分别得到第一识别结果、第二识别结果以及第三识别结果;以及获取所述第一特征提取分支的第一损失函数、所述第二特征提取分支的第二损失函数以及所述第三特征提取分支的第三损失函数;以及根据所述第一图像集、所述第一识别结果以及所述第一损失函数,所述第二图像集、所述第二识别结果以及所述第二损失函数,所述第四图像集、所述第三识别结果以及所述第三损失函数,调整所述第一模态网络的参数,得到调整后的第一模态网络,其中,所述第一模态网络的参数包括第一特征提取分支参数、第二特征提取分支参数以及第三特征提取分支参数,所述调整后的第一模态网络的分参数相同。

35 在又一种可能实现的方式中,所述第一图像集中的图像包括第一标注信息,所述第二图像集中的图像包括第二标注信息,所述第四图像集中的图像包括第三标注信息;所述训练子单元还配置为:根据所述第一标注信息、所述第一识别结果、所述第一损失函数以及所述第一特征提取分支的初始参数,得到第一梯度,以及根据所述第二标注信息、所述第二识别结果、所述第二损失函数以及所述第二特征提取分支的初始参数,得到第二梯度,以及根据所述第三标注信息、所述第三识别结果、所述第三损失函数以及所述第三特征提取分支的初始参数,得到第三梯度;以及将所述第一梯度、所述第二梯度以及所述第三梯度的平均值作为所述第一模态网络的反向传播梯度,并通过所述反向传播梯度调整所述第一模态网络的参数,使所述第一特征提取分支的参数、所述第二特征提取分支的参数以及所述第三特征提取分支的参数相同。

在又一种可能实现的方式中,所述训练子单元还配置为:从所述第一图像集以及所

述第二图像集中分别选取 f 张图像, 使所述 f 张图像中包含的人数为阈值, 得到所述第三图像集; 或, 以及从所述第一图像集以及所述第二图像集中分别选取 m 张图像以及 n 张图像, 使所述 m 与所述 n 的比值等于所述第一图像集包含的图像数量与所述第二图像集包含的图像数量的比值, 且所述 m 张图像以及所述 n 张图像中包含的人数均为所述阈值, 得到所述第三图像集; 或, 以及从所述第一图像集以及所述第二图像集中分别选取 s 张图像以及 t 张图像, 使所述 s 与所述 t 的比值等于所述第一图像集包含的人数与所述第二图像集包含的人数的比值, 且所述 s 张图像以及所述 t 张图像中包含的人数均为所述阈值, 得到所述第三图像集。

在又一种可能实现的方式中, 所述训练子单元还配置为: 对所述第三图像集中的图像依次进行特征提取处理、线性变换、非线性变换, 得到第四识别结果; 以及根据所述第三图像集中的图像、所述第四识别结果以及所述第二模态网络的第四损失函数, 调整所述第二模态网络的参数, 得到所述跨模态人脸识别网络。

在又一种可能实现的方式中, 所述第一类别以及所述第二类别分别对应不同人种。

第三方面, 提供了一种电子设备, 包括: 包括处理器、存储器; 所述处理器被配置为支持所述装置执行上述第一方面及其任一种可能的实现方式的方法中相应的功能。存储器用于与处理器耦合, 其保存所述装置必要的程序(指令)和数据。可选的, 所述装置还可以包括输入/输出接口, 用于支持所述装置与其他装置之间的通信。

第四方面, 提供了一种计算机可读存储介质, 所述计算机可读存储介质中存储有指令, 当其在计算机上运行时, 使得计算机执行上述第一方面及其任一种可能的实现方式的方法。

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。

附图说明

为了更清楚地说明本公开实施例或背景技术中的技术方案, 下面将对本公开实施例或背景技术中所需要使用的附图进行说明。

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 为本公开实施例提供的一种人脸识别方法的流程示意图;

图 2 为本公开实施例提供的一种基于第一图像集和第二图像集对第一模态网络训练的流程图;

图 3 为本公开实施例提供的另一种人脸识别神经网络的训练方法的流程图;

图 4 为本公开实施例提供的另一种人脸识别神经网络的训练方法的流程图;

图 5 为本公开实施例提供的一种基于按人种分类得到的图像集对神经网络进行训练的流程图;

图 6 为本公开实施例提供的一种人脸识别装置的结构示意图;

图 7 为本公开实施例提供的一种人脸识别装置的硬件结构示意图。

具体实施方式

为了使本技术领域的人员更好地理解本公开方案, 下面将结合本公开实施例中的附图, 对本公开实施例中的技术方案进行清楚、完整地描述, 显然, 所描述的实施例仅仅是本公开一部分实施例, 而不是全部的实施例。基于本公开中的实施例, 本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例, 都属于本公开保护的范围。

本公开的说明书和权利要求书及上述附图中的术语“第一”、“第二”等是用于区别不同对象, 而不是用于描述特定顺序。此外, 术语“包括”和“具有”以及它们任何变

形，意图在于覆盖不排他的包含。例如包含了一系列步骤或单元的过程、方法、系统、产品或设备没有限定于已列出的步骤或单元，而是可选地还包括没有列出的步骤或单元，或可选地还包括对于这些过程、方法、产品或设备固有的其他步骤或单元。

5 在本文中提及“实施例”意味着，结合实施例描述的特定特征、结构或特性可以包含在本公开的至少一个实施例中。在说明书中的各个位置出现该短语并不一定均是指相同的实施例，也不是与其它实施例互斥的独立的或备选的实施例。本领域技术人员显式地和隐式地理解的是，本文所描述的实施例可以与其它实施例相结合。

10 在本公开实施例中，人数并不等同于人物对象的数量，如：图像 A 包含 2 个对象，分别为张三和李四；图像 B 包含 1 个对象，为张三；图像 C 包含 2 个对象，分别为张三和李四，则图像 A、图像 B 以及图像 C 包含的人数为 2（张三和李四），图像 A、图像 B 以及图像 C 包含的对象的数量为 $2+1+2=5$ ，即人数为 5。

为了更清楚地说明本公开实施例或背景技术中的技术方案，下面将对本公开实施例或背景技术中所需要使用的附图进行说明。

下面结合本公开实施例中的附图对本公开实施例进行描述。

15 请参阅图 1，图 1 是本公开实施例提供的一种人脸识别方法的流程示意图。

101、获得待识别图像。在本公开实施例中，待识别图像可以是存储于本地终端（如：手机、平板电脑、笔记本电脑等等）的图像集；也可以将视频中的任意帧图像作为待识别图像，还可以从视频中任意帧图像中检测出脸部区域图像，并将该脸部区域图像作为待识别图像。

20 102、基于跨模态人脸识别网络对待识别图像进行识别，得到待识别图像的识别结果，其中，跨模态人脸识别网络基于不同模态的人脸图像数据训练得到。在本公开实施例中，跨模态人脸识别网络可对包含不同类别的对象的图像进行识别，例如，可识别两张图像中的对象是否是同一个人。其中，类别可以按人的年龄划分，也可以按人种划分，还可以按地区划分，如：可以将 0~3 岁的人划分为第一类别，将 4~10 岁的人划分为第二类别，将 11~20 岁的人划分为第三类别...；也可以将黄种人划分为第一类别，将白种人划分为第二类别，将黑种人划分为第三类别，将棕种人划分为第四类别；还可以将中国地区的人划分为第一类别，将泰国地区的人划分为第二类别，将印度地区的人划分为第三类别，将开罗地区的人划分为第四类别，将非洲地区的人划分为第五类别，将欧洲地区的人划分为第六类别。本公开实施例对类别的划分不做限定。

30 在一些可能实现的方式中，将手机摄像头采集的包括对象脸部区域图像以及事先存储的脸部区域图像作为待识别图像集输入至人脸识别神经网络，识别出待识别图像集包含的对象是否是同一个人。在另一些可能实现的方式中，摄像头 A 在第一时刻采集到第一待识别图像，摄像头 B 在第二时刻采集到第二待识别图像，将第一待识别图像以及第二待识别图像作为待识别图像集输入至人脸识别神经网络，识别这两张待识别图像中包含的对象是否是同一个人。在本公开实施例中，不同模态的人脸图像数据指包含的不同类别的对象的图像集。跨模态人脸识别网络是以不同模态的人脸图像集为训练集预先进行训练得到的，其中，跨模态人脸识别网络可以是任意具备从图像中提取特征中功能的神经网络，如：可以基于卷积层、非线性层、全连接层等网络单元按照一定方式堆叠或组成，也可以采用现有的神经网络结构，本公开对跨模态人脸识别网络的结构不做具体限定。

40 在一种可能实现的方式中，将两张待识别图像输入至跨模态人脸识别网络，跨模态人脸识别网络分别对待识别图像进行特征提取处理，得到不同的特征，再将提取出的特征进行对比，得到特征匹配度，在特征匹配度达到特征匹配度阈值的情况下，识别两张待识别图像中的对象是同一个人，反之，在特征匹配度未达到特征匹配度阈值的情况下，

识别两张待识别图像中的对象不是同一个人。本实施例通过由按类别划分的图像集训练神经网络得到跨模态人脸识别网络,通过跨模态人脸识别网络对各个类别的对象是否是同一个人进行识别,可提高识别准确率。

以下实施例为本公开提供的人脸识别方法中步骤 102 的一些可能的实现方式。

- 5 基于第一模态网络和第二模态网络进行训练得到跨模态人脸识别网络,其中,第一模态网络和第二模态网络可以是任意具备从图像中提取特征中功能的神经网络,如:可以基于卷积层、非线性层、全连接层等网络单元按照一定方式堆叠或组成,也可以采用现有的神经网络结构,本公开对跨模态人脸识别网络的结构不做具体限定。在一些可能实现的方式中,以不同的图像集为训练集分别对第一模态网络和第二模态网络进行训练,使第一模态网络分别学习到不同类别的对象的特征,再总和第一模态网络和第二模态网络学习到的特征得到跨模态网络,使跨模态网络能对不同类别的对象进行识别。可选地,在基于第一模态网络和第二模态网络进行训练得到跨模态人脸识别网络之前,基于第一图像集和第二图像集对第一模态网络训练,其中,第一图像集和第二图像集中的对象可以只包括人脸,也可以包括人脸以及躯干等其他部分,本公开对此不做具体限定。在一些可能实现的方式中,以第一图像集为训练集对第一模态网络进行训练,得到第二模态神经网络,使第二模态网络可以识别多张包含第一类别的对象的图像中的对象是否是同一个人,以第二图像集为训练集对第二模态网络进行训练,得到跨模态人脸识别网络,使跨模态人脸识别网络可以识别多张包含第一类别的对象的图像中的对象是否是同一个人,以及多张包含第二类别的对象的图像中的对象是否是同一个人,这样,跨模态人脸识别网络既在对第一类别的对象进行识别时的识别率高,且在对第二类别的对象进行识别时的识别率高。

- 在另一些可能实现的方式中,将第一图像集和第二图像集中的所有图像作为训练集对第一模态网络进行训练,得到跨模态人脸识别网络,使跨模态人脸识别网络可以识别多张包含第一类别或第二类别的对象的图像中的对象是否是同一个人。在又一些可能实现的方式中,从第一图像集中选取 a 张图像、从第二图像集中选取 b 张图像,得到训练集,其中, a:b 满足预设比例,再以训练集对第一模态网络进行训练,得到跨模态人脸识别网络,使跨模态人脸识别网络识别多张包含第一类别或第二类别的对象的图像中的人物对象是否是同一个人的识别准确率高。

- 跨模态人脸识别网络通过特征匹配度确定不同图像中的对象是否是同一个人,而不同类别的人的脸部特征会存在较大差异,因此,不同类别的人的特征匹配度阈值(即达到这个阈值,将被识别为同一个人)均不相同,本实施例提供的训练方法通过将包含不同类别的对象的图像集放到一起进行训练,可使减小跨模态人脸识别网络识别不同类别的人物对象的特征匹配度之间的差异。

- 本实施例通过由按类别划分的图像集训练神经网络(第一模态网络和第二模态网络),使神经网络同时学习不同类别的对象的人脸特征,这样,通过训练得到的跨模态人脸识别网络对各个类别的对象是否是同一个人进行识别,可提高识别准确率;通过不同类别的图像集同时训练神经网络,可减小神经网络识别不同类别的人物对象的识别标准之间的差异。

- 请参阅图 2,图 2 是本公开实施例提供的基于第一图像集和第二图像集对第一模态网络训练的一些可能的实现方式的流程示意图。

201、基于第一图像集和第二图像集对第一模态网络进行训练,得到第二模态网络,其中,第一图像集中的对象属于第一类别,第二图像集中的对象属于第二类别。在本公开实施例中,可以通过多种方式获取第一模态网络。在一些可能的实现方式中,可以从其他设备处获取第一模态网络,例如接收终端设备发送的第一模态网络。在另一些可能

的实现方式中,第一模态网络存储于本地终端,可从本地终端中调用第一模态网络。如上所述,第一图像集包括的第一类别与第二图像集包括的第二类别不同,分别以第一图像集以及第二图像集为训练集对第一模态网络进行训练,可使第一模态网络学习到第一类别以及第二类别的特征,提高并识别第一类别以及第二类别的对象是否是同一个人的准确率。在一些可能实现的方式中,第一图像集包括的对象为 11~20 岁的人,第二图像集包括的对象为 20~30 岁的人。以第一图像集、第二图像集为训练集对第一模态网络进行训练,得到的第二模态网络对对象为 11~20 岁以及 20~30 岁的对象的识别准确率高

202、按预设条件从第一图像集中选取第一数目的图像,并从第二图像集中选取第二数目的图像,并根据第一数目的图像和第二数目的图像得到第三图像集。由于第一类别的特征与第二类别的特征的差异较大,神经网络在识别第一类别的对象是否是同一个人的识别标准与识别第二类别的对象是否是同一个人的识别标准也会不同,其中,识别标准可以为提取出的不同对象的特征匹配度,如:由于 0~3 岁的人的五官以及脸部轮廓特征没有 20~30 岁的人的五官以及脸部轮廓特征明显,在训练过程中,神经网络学习到的 20~30 岁的对象的特征比 0~30 岁的对象的特征多,这样,训练后的神经网络需要以更大的特征匹配度来识别 0~3 岁的对象是否是同一个人。举例来说,在识别 0~3 岁的对象是否是同一个人时,确定特征匹配度大于或等于 0.8 的两个对象为同一个人,确定特征匹配度小于 0.8 的两个对象不是同一个人;神经网络在识别 20~30 岁的对象是否是同一个人时,确定特征匹配度大于或等于 0.65 的两个对象为同一个人,确定特征匹配度小于 0.65 的两个对象不是同一个人。此时,若用 0~3 岁的对象的识别标准去识别 20~30 岁的对象易导致本来是同一个人的两个对象被识别为不是同一个人,反之,若用 20~30 岁的对象的识别标准去识别 0~3 岁的对象易导致本来不是同一个人的两个对象被识别为同一个人。

本实施例按预设条件从第一图像集中选取第一数目的图像,并从第二图像集中选取第二数目的图像,并将第一数目的图像和第二数目的图像作为训练集,可使第二模态网络在训练过程中学习不同类别的特征的比例更均衡,减小不同类别的对象的识别标准的差异。在一些可能实现的方式中,设第一图像集中选取的第一数目的图像包括的人数以及第二图像集中选取的第二数目的图像包括的人数均为 X,则只需使分别从第一图像集以及第二图像集中选取的图像包括的人数达到 X 即可,不限定从第一图像集以及第二图像集中选取的图像的数量。

203、基于第三图像集对第二模态网络进行训练,得到跨模态人脸识别网络。第三图像集包括第一类别以及第二类别,且第一类别的人数与第二类别的人数是按预设条件选取的,这也是第三图像集不同于随机选取的图像集的地方,以第三图像集为训练集对第二模态网络进行训练,可使第二模态网络对第一类别的特征的学习和对第二类别的特征的学习更均衡。此外,若对第二模态网络的进行监督训练,在训练过程中,可通过 softmax 函数对每一张图像中的对象所属类别进行分类,并通过监督标签、分类结果以及损失函数对第二模态网络的参数进行调整。在一些可能实现的方式中,第三图像集中的每个对应一个标签,如:图像 A 与图像 B 中的同一个对象的标签均为 1,图像 C 中另一个对象的标签为 2。softmax 函数的表达式如下:

$$S_j = \frac{e^{P_j}}{\sum_{k=1}^t e^{P_k}} \dots \text{公式 (1)};$$

其中,t为第三图像集包括的人数, S_j 为对象为j类的概率, P_j 为输入 softmax 层的特征向量中的第j个数值,k为输入 softmax 层的特征向量中的第k个数值。在 softmax 层后连接包含有损失函数的损失函数层,通过 softmax 层输出的概率值、第三图像集的标签,以及损失函数,可得到第二待训练神经网络的反向传播梯度,再根据反向传播梯度对第二待训练神经网络进行梯度反向传播,可得到跨模态人脸识别网络。由于第三图像集中

包含第一类别的对象以及第二类别的对象,且第一类别的人数与第二类别的人数是满足预设条件,因此,以第三图像集为训练集对第二模态网络进行训练,可使第二模态网络平衡第一类别的人脸特征以及第二类别的人脸特征的学习比例,这样,可使最终得到的跨模态人脸识别网络在识别第一类别的对象是否是同一个人的识别率高,同时在识别第二类别的对象是否是同一个人的识别率也高。在一些可能实现的方式中,损失函数的表达式可参见下式:

$$L = -\sum_0^t y_j \log S_j \dots \text{公式 (2)};$$

其中, t 为第三图像集包括的人数, S_j 为图像对象为 j 类的概率, y_j 为第三图像集中图像对象为 j 类的标签,如:第三图像集包括张三的图像,标签为 1,则对象为 1 类的标签 1,且该对象为其他任意类别的标签都为 0。本公开实施例通过以按类别划分的第一图像集以及第二图像集为训练集对第一模态网络进行训练,提高第一模态网络对第一类别以及第二类别的识别准确率;通过以第三图像集为训练集对第二模态网络进行训练,可使第二模态网络平衡第一类别的人脸特征以及第二类别的人脸特征的学习比例,这样,训练得到的跨模态人脸识别网络不仅对第一类别的对象是否是同一个人的识别准确率高,而且对第二类别的对象是否是同一个人的识别准确率高。

请参阅图 3,图 3 是本公开实施例提供的步骤 201 的一种可能实现方式的流程示意图。

301、将第一图像集输入至第一特征提取分支,并将第二图像集输入至第二特征提取分支,并将第四图像集输入至第三特征提取分支,对第一模态网络进行训练,其中,第四图像集包括的图像为同一场景下采集的图像或同一采集方式采集的图像。在本公开实施例中,第四图像集包括的图像为同一场景下采集的图像或同一采集方式采集的图像,例如:第四图像集包括的图像均是用手机拍摄的图像;再例如:第四图像集包括的图像均是室内拍摄的图像;又例如:第四图像集包括的图像均是在港口拍摄的图像,本公开实施例对第四图像集中的图像的采集方式和采集方式不做限定。在本公开实施例中,第一模态网络包括第一特征提取分支、第二特征提取分支以及第三特征提取分支,其中,第一特征提取分支、第二特征提取分支以及第三特征提取分支均可以是任意具备从图像中提取特征功能的神经网络结构,如:可以基于卷积层、非线性层、全连接层等网络单元按照一定方式堆叠或组成,也可以采用现有的神经网络的结构,本公开对第一特征提取分支、第二特征提取分支以及第三特征提取分支的结构不做具体限定。在本实施例中,第一图像集、第二图像集以及第四图像集中的图像分别包括第一标注信息、第二标注信息以及第三标注信息,其中,标注信息包括图像中包含的对象的编号,例如:第一图像集、第二图像集以及第四图像集中包含的人数均为 Y (Y 为大于 1 的整数),对第一图像集、第二图像集以及第四图像集中的任意一张图像均包含对象对应的编号均为 $1 \sim Y$ 之间任意一个数字。需要理解的是,同一个人的对象在不同图像中的编号相同,例如:图像 A 中的对象为张三,图像 B 中的对象也为张三,则图像 A 中的对象与图像 B 中的对象的编号相同,反之,图像 C 中的对象为李四,则图像 C 中的对象的编号与图像 A 中的对象的编号不同。为使各图像集包含的对象的人脸特征可起到对应该类别人脸特征的代表性的作用,可选地,每个图像集包含的人数均在 5000 人以上,需要理解的是,本公开实施例对图像集中图像的数量不做限定。在本公开实施例中,第一特征提取分支的初始参数、第二特征提取分支的初始参数以及第三特征提取分支的初始参数分别指未调整参数前的第一特征提取分支的参数、未调整参数前的第二特征提取分支的参数以及未调整参数前的第三特征提取分支的参数。第一模态网络的各分支包括第一特征提取分支、第二特征提取分支以及第三特征提取分支。将第一图像集输入至第一特征提取分支,并将第二图像集输入至第二特征提取分支,并将第四图像集输入至第三特征提取分支,即

用第一特征提取分支去学习第一图像集包含的对象的人脸特征,用第二特征提取分支去学习第二图像集包含的对象的人脸特征,用第三特征提取分支去学习第四图像集包含的对象的人脸特征,并根据各个特征提取分支的 softmax 函数以及损失函数确定各个特征提取分支的反向传播梯度,最后根据各个特征提取分支的反向传播梯度确定第一模态网络的反向传播梯度,并对第一模态网络的参数进行调整。需要理解的是,对第一模态网络的参数进行调整即对所有特征提取分支的初始参数进行调整,由于每个特征提取分支的反向传播梯度均相同,最终调整后的参数也都相同,每个分支的反向传播梯度代表每个特征提取分支参数的调整方向,即通过特征提取分支的反向传播梯度调整分支的参数,可提高特征提取分支识别对应类别(与输入的图像集包含的类别相同)的对象准确率。

通过第一特征提取分支和第二特征提取分支的反向传播梯度调整神经网络的参数,可综合各个分支参数的调整方向,得到一个平衡的调整方向,由于第四图像集包含特定场景下或特定拍摄方式采集得到的图像,通过第三特征提取分支的反向传播梯度调整第一模态网络的参数可提高第一模态网络的鲁棒性(即对图像采集场景和图像采集方式的鲁棒性高)。通过三个特征提取分支的反向传播梯度得到的反向传播梯度来调整第一模态网络的参数可使任意一个特征提取分支识别对应类别(第一图像集以及第二图像集包含的类别中的任意一个)的对象都有较高的准确率,且可提高任意一个特征提取分支在图像采集场景和图像采集方式方面的鲁棒性。

在一些可能实现的方式中,将第一图像集输入至第一特征提取分支,并将第二图像集输入至第二特征提取分支,并将第四图像集输入至第三特征提取分支,依次经过特征提取处理、全连接层的处理、softmax 层的处理,分别得到第一识别结果、第二识别结果以及第三识别结果,其中,softmax 层包含 softmax 函数,可参见公式(1),此处将不再赘述,第一识别结果、第二识别结果以及第三识别结果包括每个对象的编号为不同编号的概率,例如:第一图像集、第二图像集以及第四图像集中包含的人数为 Y (Y 为大于 1 的整数),对第一图像集、第二图像集以及第四图像集中的任意一张图像均包含人物对象对应的编号均为 $1\sim Y$ 之间任意一个数字,则第一识别结果包括第一图像集包含的人物对象的编号分别是 $1\sim Y$ 的概率,即每个对象的第一识别结果有 Y 个概率。同理,第二识别结果包括第二图像集包含的对象的编号分别是 $1\sim Y$ 的概率,第三识别结果包括第四图像集包含的对象的编号分别是 $1\sim Y$ 的概率。在每个分支中,softmax 层后连接包含有损失函数的损失函数层,获取第一分支的第一损失函数、第二分支的第二损失函数以及第三分支的第三损失函数,根据第一图像集的第一标注信息、第一识别结果以及第一损失函数,得到第一损失,根据第二图像集的第二标注信息、第二识别结果以及第二损失函数,得到第二损失,根据第四图像集的第三标注信息、第三识别结果以及第三损失函数,得到第三损失。第一损失函数、第二损失函数以及第三损失函数可参见公式(2),此处将不再赘述。获得第一特征提取分支的参数、第二特征提取分支的参数以及第三特征提取分支的参数,根据第一特征提取分支的参数以及第一损失,得到第一梯度,以及根据第二特征提取分支的参数以及第二损失,得到第二梯度,以及根据第三特征提取分支的参数以及第三损失,得到第三梯度,其中,第一梯度、第二梯度以及第三梯度分别为第一特征提取分支、第二特征提取分支以及第三特征提取分支的反向传播梯度。根据第一梯度、第二梯度以及第三梯度,得到第一模态网络的反向传播梯度,并通过梯度反向传播的方式调整第一模态网络的参数,使第一特征提取分支的参数、第二特征提取分支以及第三特征提取分支的参数相同。在一些可能实现的方式中,将第一梯度、第二梯度以及第三梯度的平均值作为第一待训练神经网络的反向传播梯度,并根据反向传播梯度对第一模态网络进行梯度方向传播,调整第一特征提取分支的参数、第二特征提取分支以及第三特征提取分支的参数,使调整参数后的第一特征提取分支、第二特征提取分

支以及第三特征提取分支的参数相同。

302、将训练后的第一特征提取分支或训练后的第二特征提取分支或训练后的第三特征提取分支作为第二模态网络。通过 301 的处理，训练后的第一特征提取分支、训练后的第二特征提取分支以及训练后的第三特征提取分支的参数相同，即对第一类别（第一图像集包含的类别）、第二类别（第二图像集包含的类别）的对象识别准确率高，且识别不同场景采集的图像和不同采集方式采集的图像的鲁棒性好。因此，将训练后的第一特征提取分支或训练后的第二特征提取分支或训练后的第三特征提取分支作为下一步训练的网络，即第二模态网络。本公开实施例中，第一图像集以及第二图像集均是按类别选取得到的图像集，第四图像集为按照场景和拍摄方式选取的图像集，以第一图像集对第一特征提取分支进行训练，可使第一特征提取分支着重学习第一类别的人脸特征，以第二图像集对第二特征提取分支进行训练，可使第二特征提取分支着重学习第二类别的人脸特征，而以第四图像集对第三特征提取分支进行训练，可使第三特征提取分支着重学习第四图像集包括的对象的人脸特征，提高第三特征提取分支的鲁棒性；根据第一特征提取分支的反向传播梯度、第二特征提取分支的反向传播梯度以及第三特征提取分支的反向传播梯度得到第一模态网络的反向传播梯度，并以该梯度对第一模态网络进行梯度反向传播，可同时兼顾三个特征提取分支的参数调整方向，并使调整参数后的第一模态网络的鲁棒性好，且对第一类别以及第二类别的人物对象的识别准确率高。以下实施例为步骤 202 的一些可能的实现方式。为使第二模态网络在基于第三图像集进行训练时，更均衡的学习第一类别和第二类别的特征，预设条件可以为第一数目与第二数目相同，在一种可能实现的方式中，从第一图像集以及第二图像集中分别选取 f 张图像，使 f 张图像中包含的人数为阈值，得到第三图像集。在一些可能实现的方式中，阈值为 1000，从第一图像集以及第二图像集中分别选取 f 张图像，使 f 张图像中包含的人数为 1000 即可，其中， f 可为任意正整数，最后将从第一图像集中选出的 f 张图像以及从第二图像集中选出的 f 张图像作为第三图像集。为使第二模态网络在基于第三图像集进行训练时，更有针对性的学习第一类别和第二类别的特征，预设条件可以为第一数目与第二数目的比值等于第一图像集包含的图像数目与第二图像集包含的图像数目的比值，或第一数目与第二数目的比值等于第一图像集包含的人数与第二图像集包含的人数的比值，这样，第二模态网络学习第一类别的特征与第二类别的特征的比值均为定值，可弥补第一类别的识别标准与第二类别的识别标准的差异。在一种可能实现的方式中，从第一图像集以及第二图像集中分别选取 m 张图像以及 n 张图像，使 m 与 n 的比值等于第一图像集包含的图像数量与第二图像集包含的图像数量的比值，且 m 张图像以及 n 张图像中包含的人数均为阈值，得到第三图像集。在一些可能实现的方式中，第一图像集包含 7000 张图像，第二图像集包含 8000 张图像，阈值为 1000，从第一图像集选取的 m 张图像以及从第二图像集中选取的 n 张图像中包含的人数均为 1000，且 $m:n=7:8$ ， m 、 n 可为任意正整数，最后将从第一图像集中选出的 m 张图像以及从第二图像集中选出的 n 张图像作为第三图像集。在另一种可能实现的方式中，从第一图像集以及第二图像集中分别选取 s 张图像以及 t 张图像，使 s 与 t 的比值等于第一图像集包含的人数与第二图像集包含的人数的比值，且 s 张图像以及 t 张图像中包含的人数均为阈值，得到第三图像集。在一些可能实现的方式中，第一图像集包含的人数为 6000，第二图像集包含的人数为 7000，阈值为 1000，从第一图像集选取的 s 张图像以及从第二图像集中选取的 t 张图像中包含的人数均为 1000，且 $s:t=6:7$ ， s 、 t 可为任意正整数，最后将从第一图像集中选出的 s 张图像以及从第二图像集中选出的 t 张图像作为第三图像集。

本实施例提供了几种从第一图像集以及第二图像集中选取图像的方式，通过不同的选取方式可得到不同的第三图像集，可根据具体训练效果以及需求选择不同的选取方式。

请参阅图 4，图 4 是本公开实施例提供的步骤 203 的一种可能的实现方式的流程图。

401、对第三图像集中的图像依次进行特征提取处理、线性变换、非线性变换，得到第四识别结果。首先，第二模态网络对第三图像集中的图像进行特征提取处理，特征提取处理可以通过多种方式实现，例如卷积、池化等，本公开实施例对此不做具体限定。在一些可能的实现方式中，第二模态网络包括多层卷积层，通过多层卷积层对第三图像集中的图像逐层进行卷积处理完成对第三图像集中的图像的特征提取处理，其中，每个卷积层提取出的特征内容及语义信息均不一样，具体表现为，特征提取处理一步步地将图像的特征抽象出来，同时也将逐步去除相对次要的特征，因此，越到后面提取出的特征尺寸越小，内容及语义信息就越浓缩。通过多层卷积层逐级对第三图像集中的图像进行卷积处理，并提取相应的特征，最终得到固定大小的特征图像，这样，可在获得待处理图像主要内容信息（即第三图像集中的图像的特征图像）的同时，将图像尺寸缩小，减小系统的计算量，提高运算速度。在一种可能实现的方式中，卷积处理的实现过程如下：卷积层对待处理图像做卷积处理，即利用卷积核在第三图像集中的图像上滑动，并将第三图像集中的图像上的像素与对应的卷积核上的数值相乘，然后将所有相乘后的值相加作为卷积核中间像素对应的图像上像素值，最终滑动处理完第三图像集中的图像中所有的像素，并提取出相应的特征图像。在卷积层后连接的是全连接层，通过全连接层对卷积层提取出的特征图像进行线性变换，可将特征图像中的特征映射到样本（即对象的编号）标记空间。在全连接层后连接有 softmax 层，通过 softmax 层对提取出的特征图像进行处理，得到第四识别结果，softmax 层具体组成以及对特征图像的处理过程可参见 301，此处将不再赘述，其中，第四识别结果包括第三图像集包含的对象的编号分别是 1~Z（第三图像集包括的人数为 Z）的概率，即每个对象的第四识别结果有 Z 个概率。

402、根据第三图像集中的图像、第四识别结果以及第二模态网络的第四损失函数，调整第二模态网络的参数，得到跨模态人脸识别网络。在 softmax 层后连接有包含第四损失函数的损失函数层，第四损失函数的表达式可参见公式（2）。由于输入至第二待训练神经网络的第三图像集包含不同类别的对象，因此，在通过 softmax 函数得到第四识别结果的过程中，将不同类别的对象的人脸特征放在一起进行比较，对不同类别的识别标准归一化，即以相同的识别标准识别不同类别的对象，最后通过第四识别结果和第四损失函数调整第二模态网络的参数，使调整参数后的第二模态网络以相同的识别标准识别不同类别的对象，提高了不同类别的对象的识别准确率，在一些可能实现的方式中，第一类别的识别标准是 0.8，第二类别的识别标准是 0.65，通过 402 的训练，调整第二模态网络的参数以及识别标准，最终确定识别标准为 0.72。由于第二模态网络的参数随着识别标准的调整也会相应地调整，因此，使调整参数后得到的跨模态人脸识别网络通过减少第一类别的识别标准与第二类别的识别标准之间的差异。

本公开实施例中，以第三图像集为训练集对第二模态网络进行训练，可将不同类别的对象的人脸特征放在一起进行比较，对不同类别的识别标准归一化；通过调整第二模态网络的参数，使调整参数后得到的跨模态人脸识别网络不仅对识别第一类别的对象的是否是同一个人的识别准确率高，而且对识别第二类别的对象的是否是同一个人的识别准确率高，减小了识别不同类别的对象是否是同一个人时的识别标准的差异。如上所述，训练用的图像集包含的人物对象的类别可以是按人的年龄划分的，也可以是按人种划分的，还可以是按地区划分的，本公开提供一种基于按人种分类得到的图像集对神经网络进行训练的方法，即第一类别以及第二类别分别对应不同人种，可提高神经网络对不同人种的对象的识别准确率。

请参见图 5，图 5 为本公开提供的一种基于按人种分类得到的图像集对神经网络进行训练的方法流程。

501、获得基础图像集、人种图像集，以及第三模态网络。在本公开实施例中，基础图像集可以包括一个或多个图像集，具体地，第十一图像集中的图像均是在室内采集的图像，第十二图像集中的图像均是在港口采集的图像，第十三图像集中的图像均是在野外采集的图像，第十四图像集中的图像均是在人群中采集的图像，第十五图像集中的图像均是证件图像，第十六图像集中的图像均是通过手机拍摄的图像，第十七图像集中的图像均是通过摄像机采集的图像，第十八图像集中的图像均是从视频中截取的图像，第十九图像集中的图像均是从互联网下载的图像，第二十图像集中的图像均是对名人图像进行处理后得到的图像。需要理解的是，基础图像集中的任意一个图像集包括的图像均为同一场景下采集的图像或同一采集方式采集的图像，即基础图像集中的图像集对应与 301 中的第四图像集。将中国地区的人划分为第一人种，将泰国地区的人划分为第二人种，将印度地区的人划分为第三人种，将开罗地区的人划分为第四人种，将非洲地区的人划分为第五人种，将欧洲地区的人划分为第六人种，对应地，就有 6 个人种图像集，分别为包含以上 6 个人种，具体地，第五图像集包含第一人种，第六图像集包含第二人种...第十图像集包含第六人种。需要理解的是，人种图像集中的任意一个图像集包括的对象均为同一人种（即同一类别），即人种图像集中的图像集对应与 101 中的第一图像集或第二图像集。

为使各图像集包含的对象的人脸特征可起到对应该类别人脸特征的代表性的作用，可选地，每个图像集包含的人数均在 5000 人以上，需要理解的是，本公开实施例对图像集中图像的数量不做限定。需要理解的是，人种划分还可以是其他方式，例如：按肤色划分人种，可分为黄色人种、白色人种、黑色人种和棕色人种四个人种，本实施例对人种划分的方式不做限定。基础图像集以及人种图像集中的对象可以只包括人脸，也可以包括人脸以及躯干等其他部分，本公开对此不做具体限定。在本实施例中，第三模态网络可以是任意具备从图像中提取特征中功能的神经网络，如：可以基于卷积层、非线性层、全连接层等网络单元按照一定方式堆叠或组成，也可以采用现有的神经网络结构，本公开对第三模态网络的结构不做具体限定。

502、基于基础图像集和人种图像集第三模态网络进行训练，得到第四模态网络。此步骤具体可参见 201 以及 301~302，此处将不再赘述。需要理解的是，由于基础图像集中包括 10 个图像集，人种图像集中包括 6 个图像集，相应地，第三模态网络包括 16 个特征提取分支，即每个图像集对应一个特征提取分支。通过 502 的处理，可提高第四模态网络对不同人种的对象是否是同一个人的识别准确率，即提高各个人种内的识别准确率，具体地，用第四模态网络分别识别第一人种、第二人种、第三人种、第四人种、第五人种、第六人种的对象是否是同一个人，均有较高的准确率，且第四待模态网络对识别不同场景下或以不同采集方式采集到的图像的鲁棒性好。

503、基于人种图像集对第四模态网络进行训练，得到跨人种人脸识别网络。此步骤具体可参见 202~203 以及 401~402，此处将不再赘述。通过 503 的处理，可减小得到的跨人种人脸识别网络识别不同人种的对象是否是同一个人时的识别标准的差异，跨人种人脸识别网络可提高不同人种的对象识别准确率。具体地，跨人种人脸识别网络对不同图像中属于第一人种的对象是否是同一个人的识别准确率，以及对不同图像中属于第二人种的对象是否是同一个人的识别准确率，以及...，以及对不同图像中属于第六人种的对象是否是同一个人的识别准确率都在预设值之上，需理解，预设值表示跨人种人脸识别网络对各个人种的识别准确率都很高，本公开对预设值的具体大小不做限定，可选地，预设值为 98%。可选地，为同时提高人种内的识别准确率以及减小不同人种的识

别标准的差异,可多次重复 502 以及 503,在一些可能实现的方式中,以 502 的训练方式对第三模态网络训练 10 万轮,然后在接下来的 10~15 万轮训练中,502 的训练方式的比重逐渐降低为 0,而 503 的训练方式的比重逐提升至 1,15~25 万轮的训练均通过 503 的训练方式完成,在接下来的 25~30 万轮训练中,503 的训练方式的比重逐渐降低为 0,而 502 的训练方式的比重逐提升至 1;最后,在第 30~40 万轮训练中,502 的训练方式以及 503 的训练方式各占一半比重。需要理解的是,本公开实施例对各个阶段的轮数具体数值、502 的训练方式以及 503 的训练方式的比重均不做限定。应用本实施例得到的跨人种人脸识别网络可对识别多个人种的对象是否是同一个人,且识别准确率高,如:应用跨人种人脸识别网络即可对中国地区的人种进行识别,也可对开罗地区的人种进行识别,还可对欧洲地区的人种进行识别,且每个人种的识别准确率高,这样,可解决人脸识别算法在对某一类人种识别准确率高,但对其他人种识别准确率低的问题。此外,应用本实施例可提高跨人种人脸识别网络识别不同场景下或以不同采集方式采集到的图像的鲁棒性。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。

上述详细阐述了本公开实施例的方法,下面提供了本公开实施例的装置。

请参阅图 6,图 6 为本公开实施例提供的一种人脸识别装置的结构示意图,该识别装置 1 包括:获取单元 11 以及识别单元 12。其中:获取单元 11,配置为获得待识别图像;识别单元 12,配置为基于跨模态人脸识别网络对所述待识别图像进行识别,得到所述待识别图像的识别结果,其中,所述跨模态人脸识别网络基于不同模态的人脸图像数据训练得到。

进一步地,所述识别单元 12 包括:训练子单元 121,配置为基于第一模态网络和第二模态网络进行训练得到所述跨模态人脸识别网络。

进一步地,所述训练子单元 121 还配置为:基于第一图像集和第二图像集对所述第一模态网络训练,其中,所述第一图像集中的对象属于第一类别,所述第二图像集中的对象属于第二类别。进一步地,所述训练子单元 121 还配置为:基于所述第一图像集和所述第二图像集对所述第一模态网络进行训练,得到所述第二模态网络;以及按预设条件从所述第一图像集中选取第一数目的图像,并从所述第二图像集中选取第二数目的图像,并根据所述第一数目的图像和所述第二数目的图像得到第三图像集;以及基于所述第三图像集对所述第二模态网络进行训练,得到所述跨模态人脸识别网络。进一步地,所述预设条件包括:所述第一数目与所述第二数目相同,所述第一数目与所述第二数目的比值等于所述第一图像集包含的图像数目与所述第二图像集包含的图像数目的比值,所述第一数目与所述第二数目的比值等于所述第一图像集包含的人数与所述第二图像集包含的人数的比值中的任意一种。进一步地,所述第一模态网络包括第一特征提取分支、第二特征提取分支以及第三特征提取分支;所述训练子单元 121 还配置为:将所述第一图像集输入至所述第一特征提取分支,并将所述第二图像集输入至所述第二特征提取分支,并将第四图像集输入至所述第三特征提取分支,对所述第一模态网络进行训练,其中,所述第四图像集包括的图像为同一场景下采集的图像或同一采集方式采集的图像;以及将训练后的第一特征提取分支或训练后的第二特征提取分支或训练后的第三特征提取分支作为所述第二模态网络。进一步地,所述训练子单元 121 还配置为:将所述第一图像集、所述第二图像集以及所述第四图像集分别输入至所述第一特征提取分支、所述第二特征提取分支以及所述第三特征提取分支,分别得到第一识别结果、第二识别结果以及第三识别结果;以及获取所述第一特征提取分支的第一损失函数、所述第二特征提取分支的第二损失函数以及所述第三特征提取分支的第三损失函数;以及根据所述第

一图像集、所述第一识别结果以及所述第一损失函数，所述第二图像集、所述第二识别结果以及所述第二损失函数，所述第四图像集、所述第三识别结果以及所述第三损失函数，调整所述第一模态网络的参数，得到调整后的第一模态网络，其中，所述第一模态网络的参数包括第一特征提取分支参数、第二特征提取分支参数以及第三特征提取分支参数，所述调整后的第一模态网络的各分支参数相同。进一步地，所述第一图像集中的图像包括第一标注信息，所述第二图像集中的图像包括第二标注信息，所述第四图像集中的图像包括第三标注信息；所述训练子单元 121 还配置为：根据所述第一标注信息、所述第一识别结果、所述第一损失函数以及所述第一特征提取分支的初始参数，得到第一梯度，以及根据所述第二标注信息、所述第二识别结果、所述第二损失函数以及所述第二特征提取分支的初始参数，得到第二梯度，以及根据所述第三标注信息、所述第三识别结果、所述第三损失函数以及所述第三特征提取分支的初始参数，得到第三梯度；以及将所述第一梯度、所述第二梯度以及所述第三梯度的平均值作为所述第一模态网络的反向传播梯度，并通过所述反向传播梯度调整所述第一模态网络的参数，使所述第一特征提取分支的参数、所述第二特征提取分支的参数以及所述第三特征提取分支的参数相同。进一步地，所述训练子单元 121 还配置为：从所述第一图像集以及所述第二图像集中分别选取 f 张图像，使所述 f 张图像中包含的人数为阈值，得到所述第三图像集；或，以及从所述第一图像集以及所述第二图像集中分别选取 m 张图像以及 n 张图像，使所述 m 与所述 n 的比值等于所述第一图像集包含的图像数量与所述第二图像集包含的图像数量的比值，且所述 m 张图像以及所述 n 张图像中包含的人数均为所述阈值，得到所述第三图像集；或，以及从所述第一图像集以及所述第二图像集中分别选取 s 张图像以及 t 张图像，使所述 s 与所述 t 的比值等于所述第一图像集包含的人数与所述第二图像集包含的人数的比值，且所述 s 张图像以及所述 t 张图像中包含的人数均为所述阈值，得到所述第三图像集。进一步地，所述训练子单元 121 还配置为：对所述第三图像集中的图像依次进行特征提取处理、线性变换、非线性变换，得到第四识别结果；以及根据所述第三图像集中的图像、所述第四识别结果以及所述第二模态网络的第四损失函数，调整所述第二模态网络的参数，得到所述跨模态人脸识别网络。进一步地，所述第一类别以及所述第二类别分别对应不同人种。在一些实施例中，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。

图 7 为本公开实施例提供的一种人脸识别装置的硬件结构示意图。该识别装置 2 包括处理器 21，还可以包括输入装置 22、输出装置 23 和存储器 24。该输入装置 22、输出装置 23、存储器 24 和处理器 21 之间通过总线相互连接。存储器包括但不限于是随机存储记忆体（random access memory，RAM）、只读存储器（read-only memory，ROM）、可擦除可编程只读存储器（erasable programmable read only memory，EPROM）、或便携式只读存储器（compact disc read-only memory，CD-ROM），该存储器用于相关指令及数据。输入装置用于输入数据和/或信号，以及输出装置用于输出数据和/或信号。输出装置和输入装置可以是独立的器件，也可以是一个整体的器件。处理器可以包括是一个或多个处理器，例如包括一个或多个中央处理器（central processing unit，CPU），在处理器是一个 CPU 的情况下，该 CPU 可以是单核 CPU，也可以是多核 CPU。存储器用于存储网络设备的程序代码和数据。处理器用于调用该存储器中的程序代码和数据，执行上述方法实施例中的步骤。具体可参见方法实施例中的描述，在此不再赘述。可以理解的是，图 7 仅仅示出了一种人脸识别装置的简化设计。在实际应用中，人脸识别装置还可以分别包含必要的其他元件，包括但不限于任意数量的输入/输出装置、处理器、控制器、存储器等，而所有可以实现本公开实施例的人脸识别装置都在本公开的保护范围之

内。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公开的范围。所属领域的技术人员可以清楚地了解到，为描述的方便和简洁，上述描述的系统、装置和单元的具体工作过程，可以参考前述方法实施例中的对应过程，在此不再赘述。所属领域的技术人员还可以清楚地了解到，本公开各个实施例描述各有侧重，为描述的方便和简洁，相同或类似的部分在不同实施例中可能没有赘述，因此，在某一实施例未描述或未详细描述的部分可以参见其他实施例的记载。在本公开所提供的几个实施例中，应该理解到，所揭露的系统、装置和方法，可以通过其它的方式实现。例如，以上所描述的装置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系统，或一些特征可以忽略，或不执行。另一点，所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口，装置或单元的间接耦合或通信连接，可以是电性，机械或其它的形式。所述作为分离部件说明的单元可以是或者也可以不是物理上分开的，作为单元显示的部件可以是或者也可以不是物理单元，即可以位于一个地方，或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本实施例方案的目的。

另外，在本公开各个实施例中的各功能单元可以集成在一个处理单元中，也可以是各个单元单独物理存在，也可以两个或两个以上单元集成在一个单元中。在上述实施例中，可以全部或部分地通过软件、硬件、固件或者其任意组合来实现。当使用软件实现时，可以全部或部分地以计算机程序产品的形式实现。所述计算机程序产品包括一个或多个计算机指令。在计算机上加载和执行所述计算机程序指令时，全部或部分地产生按照本公开实施例所述的流程或功能。所述计算机可以是通用计算机、专用计算机、计算机网络、或者其他可编程装置。所述计算机指令可以存储在计算机可读存储介质中，或者通过所述计算机可读存储介质进行传输。所述计算机指令可以从一个网站站点、计算机、服务器或数据中心通过有线（例如同轴电缆、光纤、数字用户线（digital subscriber line, DSL））或无线（例如红外、无线、微波等）方式向另一个网站站点、计算机、服务器或数据中心进行传输。所述计算机可读存储介质可以是计算机能够存取的任何可用介质或者是包含一个或多个可用介质集成的服务器、数据中心等数据存储设备。所述可用介质可以是磁性介质，（例如，软盘、硬盘、磁带）、光介质（例如，数字通用光盘（digital versatile disc, DVD））、或者半导体介质（例如固态硬盘（solid state disk, SSD））等。

本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程，该流程可以由计算机程序来指令相关的硬件完成，该程序可存储于计算机可读取存储介质中，该程序在执行时，可包括如上述各方法实施例的流程。而前述的存储介质包括：只读存储器（read-only memory, ROM）或随机存储存储器（random access memory, RAM）、磁碟或者光盘等各种可存储程序代码的介质为使本公开实施例的目的、技术方案和优点更加清楚，下面将结合本公开实施例中的附图，对发明的具体技术方案做进一步详细描述。以下实施例用于说明本公开，但不用来限制本公开的范围。

权利要求书

- 1、一种人脸识别方法，其中，包括：
获得对待识别图像；
基于跨模态人脸识别网络对所述待识别图像进行识别，得到所述待识别图像的识别
5 结果，其中，所述跨模态人脸识别网络基于不同模态的人脸图像数据训练得到。
- 2、根据权利要求 1 所述的方法，其中，所述基于不同模态的人脸图像数据训练得
到所述跨模态人脸识别网络的过程，包括：
基于第一模态网络和第二模态网络进行训练得到所述跨模态人脸识别网络。
- 3、根据权利要求 2 所述的方法，其中，在所述基于第一模态网络和第二模态网络
10 进行训练得到所述跨模态人脸识别网络之前，还包括：
基于第一图像集和第二图像集对所述第一模态网络训练，其中，所述第一图像集中
的对象属于第一类别，所述第二图像集中的对象属于第二类别。
- 4、根据权利要求 3 所述的方法，其中，所述基于第一图像集和第二图像集对所述
第一模态网络训练，包括：
15 基于所述第一图像集和所述第二图像集对所述第一模态网络进行训练，得到所述第
二模态网络；
按预设条件从所述第一图像集中选取第一数目的图像，并从所述第二图像集中选取
第二数目的图像，并根据所述第一数目的图像和所述第二数目的图像得到第三图像集；
基于所述第三图像集对所述第二模态网络进行训练，得到所述跨模态人脸识别网络。
- 5、根据权利要求 4 所述的方法，其中，所述预设条件包括：所述第一数目与所述
20 第二数目相同，所述第一数目与所述第二数目的比值等于所述第一图像集包含的图像数
目与所述第二图像集包含的图像数目的比值，所述第一数目与所述第二数目的比值等于
所述第一图像集包含的人数与所述第二图像集包含的人数的比值中的任意一种。
- 6、根据权利要求 2 或 4 所述的方法，其中，所述第一模态网络包括第一特征提取
25 分支、第二特征提取分支以及第三特征提取分支；
所述基于所述第一图像集和所述第二图像集对所述第一模态网络进行训练，得到所
述第二模态网络，包括：
将所述第一图像集输入至所述第一特征提取分支，并将所述第二图像集输入至所述
第二特征提取分支，并将第四图像集输入至所述第三特征提取分支，对所述第一模态网
30 络进行训练，其中，所述第四图像集包括的图像为同一场景下采集的图像或同一采集方
式采集的图像；
将训练后的第一特征提取分支或训练后的第二特征提取分支或训练后的第三特征
提取分支作为所述第二模态网络。
- 7、根据权利要求 6 所述的方法，其中，所述将所述第一图像集输入至所述第一特
35 征提取分支，并将所述第二图像集输入至所述第二特征提取分支，并将第四图像集输入
至所述第三特征提取分支，对所述第一模态网络进行训练，包括：
将所述第一图像集、所述第二图像集以及所述第四图像集分别输入至所述第一特征
提取分支、所述第二特征提取分支以及所述第三特征提取分支，分别得到第一识别结果、
第二识别结果以及第三识别结果；
40 获取所述第一特征提取分支的第一损失函数、所述第二特征提取分支的第二损失函
数以及所述第三特征提取分支的第三损失函数；
根据所述第一图像集、所述第一识别结果以及所述第一损失函数，所述第二图像集、

所述第二识别结果以及所述第二损失函数,所述第四图像集、所述第三识别结果以及所述第三损失函数,调整所述第一模态网络的参数,得到调整后的第一模态网络,其中,所述第一模态网络的参数包括第一特征提取分支参数、第二特征提取分支参数以及第三特征提取分支参数,所述调整后的第一模态网络的各分支参数相同。

- 5 8、根据权利要求7所述的方法,其中,所述第一图像集中的图像包括第一标注信息,所述第二图像集中的图像包括第二标注信息,所述第四图像集中的图像包括第三标注信息;

所述根据所述第一图像集、所述第一识别结果以及所述第一损失函数,所述第二图像集、所述第二识别结果以及所述第二损失函数,所述第四图像集、所述第三识别结果以及所述第三损失函数,调整所述第一模态网络的参数,得到调整后的第一模态网络,包括:

根据所述第一标注信息、所述第一识别结果、所述第一损失函数以及所述第一特征提取分支的初始参数,得到第一梯度,以及根据所述第二标注信息、所述第二识别结果、所述第二损失函数以及所述第二特征提取分支的初始参数,得到第二梯度,以及根据所述第三标注信息、所述第三识别结果、所述第三损失函数以及所述第三特征提取分支的初始参数,得到第三梯度;

将所述第一梯度、所述第二梯度以及所述第三梯度的平均值作为所述第一模态网络的反向传播梯度,并通过所述反向传播梯度调整所述第一模态网络的参数,使所述第一特征提取分支的参数、所述第二特征提取分支的参数以及所述第三特征提取分支的参数相同。

- 20 9、根据权利要求4或5所述的方法,其中,所述按预设条件从所述第一图像集中选取第一数量张图像,并从所述第二图像集中选取第二数量张图像,得到第三图像集,包括:

从所述第一图像集以及所述第二图像集中分别选取f张图像,使所述f张图像中包含的人数为阈值,得到所述第三图像集;或,

从所述第一图像集以及所述第二图像集中分别选取m张图像以及n张图像,使所述m与所述n的比值等于所述第一图像集包含的图像数量与所述第二图像集包含的图像数量的比值,且所述m张图像以及所述n张图像中包含的人数均为所述阈值,得到所述第三图像集;或,

30 从所述第一图像集以及所述第二图像集中分别选取s张图像以及t张图像,使所述s与所述t的比值等于所述第一图像集包含的人数与所述第二图像集包含的人数的比值,且所述s张图像以及所述t张图像中包含的人数均为所述阈值,得到所述第三图像集。

- 10、根据权利要求3所述的方法,其中,所述基于所述第三图像集对所述第二模态网络进行训练,得到所述跨模态人脸识别网络,包括:

35 对所述第三图像集中的图像依次进行特征提取处理、线性变换、非线性变换,得到第四识别结果;

根据所述第三图像集中的图像、所述第四识别结果以及所述第二模态网络的第四损失函数,调整所述第二模态网络的参数,得到所述跨模态人脸识别网络。

- 40 11、根据权利要求1至5、7、8、10中任意一项所述的方法,其中,所述第一类别以及所述第二类别分别对应不同人种。

12、一种人脸识别装置,其中,包括:

获取单元,配置为获得待识别图像;

识别单元,配置为基于跨模态人脸识别网络对所述待识别图像进行识别,得到所述待识别图像的识别结果,其中,所述跨模态人脸识别网络基于不同模态的人脸图像数据

训练得到。

13、根据权利要求 12 所述的装置，其中，所述识别单元包括：

训练子单元，配置为基于第一模态网络和第二模态网络进行训练得到所述跨模态人脸识别网络。

5 14、根据权利要求 13 所述的装置，其中，所述训练子单元还配置为：

基于第一图像集和第二图像集对所述第一模态网络训练，其中，所述第一图像集中的对象属于第一类别，所述第二图像集中的对象属于第二类别。

15、根据权利要求 14 所述的装置，其中，所述训练子单元还配置为：

10 基于所述第一图像集和所述第二图像集对所述第一模态网络进行训练，得到所述第二模态网络；

以及按预设条件从所述第一图像集中选取第一数目的图像，并从所述第二图像集中选取第二数目的图像，并根据所述第一数目的图像和所述第二数目的图像得到第三图像集；

15 以及基于所述第三图像集对所述第二模态网络进行训练，得到所述跨模态人脸识别网络。

16、根据权利要求 15 所述的装置，其中，所述预设条件包括：所述第一数目与所述第二数目相同，所述第一数目与所述第二数目的比值等于所述第一图像集包含的图像数目与所述第二图像集包含的图像数目的比值，所述第一数目与所述第二数目的比值等于所述第一图像集包含的人数与所述第二图像集包含的人数的比值中的任意一种。

20 17、根据权利要求 13 或 15 所述的装置，其中，所述第一模态网络包括第一特征提取分支、第二特征提取分支以及第三特征提取分支；所述训练子单元还配置为：

25 将所述第一图像集输入至所述第一特征提取分支，并将所述第二图像集输入至所述第二特征提取分支，并将第四图像集输入至所述第三特征提取分支，对所述第一模态网络进行训练，其中，所述第四图像集包括的图像为同一场景下采集的图像或同一采集方式采集的图像；

以及将训练后的第一特征提取分支或训练后的第二特征提取分支或训练后的第三特征提取分支作为所述第二模态网络。

18、根据权利要求 17 所述的装置，其中，所述训练子单元还配置为：

30 将所述第一图像集、所述第二图像集以及所述第四图像集分别输入至所述第一特征提取分支、所述第二特征提取分支以及所述第三特征提取分支，分别得到第一识别结果、第二识别结果以及第三识别结果；

以及获取所述第一特征提取分支的第一损失函数、所述第二特征提取分支的第二损失函数以及所述第三特征提取分支的第三损失函数；

35 以及根据所述第一图像集、所述第一识别结果以及所述第一损失函数，所述第二图像集、所述第二识别结果以及所述第二损失函数，所述第四图像集、所述第三识别结果以及所述第三损失函数，调整所述第一模态网络的参数，得到调整后的第一模态网络，其中，所述第一模态网络的参数包括第一特征提取分支参数、第二特征提取分支参数以及第三特征提取分支参数，所述调整后的第一模态网络的各分支参数相同。

40 19、根据权利要求 18 所述的装置，其中，所述第一图像集中的图像包括第一标注信息，所述第二图像集中的图像包括第二标注信息，所述第四图像集中的图像包括第三标注信息；所述训练子单元还配置为：

根据所述第一标注信息、所述第一识别结果、所述第一损失函数以及所述第一特征提取分支的初始参数，得到第一梯度，以及根据所述第二标注信息、所述第二识别结果、所述第二损失函数以及所述第二特征提取分支的初始参数，得到第二梯度，以及根据所

述第三标注信息、所述第三识别结果、所述第三损失函数以及所述第三特征提取分支的初始参数，得到第三梯度；

5 以及将所述第一梯度、所述第二梯度以及所述第三梯度的平均值作为所述第一模态网络的反向传播梯度，并通过所述反向传播梯度调整所述第一模态网络的参数，使所述第一特征提取分支的参数、所述第二特征提取分支的参数以及所述第三特征提取分支的参数相同。

20、根据权利要求 15 或 16 所述的装置，其中，所述训练子单元还配置为：

10 从所述第一图像集以及所述第二图像集中分别选取 f 张图像，使所述 f 张图像中包含的人数为阈值，得到所述第三图像集；或，

以及从所述第一图像集以及所述第二图像集中分别选取 m 张图像以及 n 张图像，使所述 m 与所述 n 的比值等于所述第一图像集包含的图像数量与所述第二图像集包含的图像数量的比值，且所述 m 张图像以及所述 n 张图像中包含的人数均为所述阈值，得到所述第三图像集；或，

15 以及从所述第一图像集以及所述第二图像集中分别选取 s 张图像以及 t 张图像，使所述 s 与所述 t 的比值等于所述第一图像集包含的人数与所述第二图像集包含的人数的比值，且所述 s 张图像以及所述 t 张图像中包含的人数均为所述阈值，得到所述第三图像集。

21、根据权利要求 14 所述的装置，其中，所述训练子单元还配置为：

20 对所述第三图像集中的图像依次进行特征提取处理、线性变换、非线性变换，得到第四识别结果；

以及根据所述第三图像集中的图像、所述第四识别结果以及所述第二模态网络的第二损失函数，调整所述第二模态网络的参数，得到所述跨模态人脸识别网络。

22、根据权利要求 12 至 16、18、19、21 中任意一项所述的装置，其中，所述第一类别以及所述第二类别分别对应不同人种。

25 23、一种电子设备，其中，包括存储器和处理器，所述存储器上存储有计算机可执行指令，所述处理器运行所述存储器上的计算机可执行指令时实现权利要求 1 至 11 任一项所述的方法。

24、一种计算机可读存储介质，其上存储有计算机程序，该计算机程序被处理器执行时，实现权利要求 1 至 11 任一项所述的方法。

30