

深度估计方法和装置、电子设备、程序和介质

本公开要求在 2018 年 02 月 01 日提交中国专利局、申请号为 CN 201810103195.0、发明名称为“深度估计方法和装置、电子设备、程序和介质”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，尤其是一种深度估计方法和装置、电子设备、计算机程序和计算机可读存储介质。

背景技术

深度估计是计算机视觉领域内的重要问题。准确的深度估计方法在许多领域，例如自动驾驶、三维场景重建以及增强现实等领域，有着重要的利用价值。

在卷积神经网络发展的驱动下，深度估计的相关技术得到了快速发展。

发明内容

本公开实施例提供一种深度估计技术方案。

根据本公开实施例的一个方面，提供一种深度估计方法，包括：

以单张图片作为双目图像中的第一图像，经第一神经网络，基于所述第一图像获取所述双目图像中的第二图像；

经第二神经网络，通过对所述第一图像与所述第二图像进行双目立体匹配，获取所述第一图像对应的深度信息。

根据本公开实施例的另一个方面，提供一种深度估计装置，包括：

图像获取模块，用于以单张图片作为双目图像中的第一图像，经第一神经网络，基于所述第一图像获取所述双目图像中的第二图像；

立体匹配模块，用于经第二神经网络，通过对所述第一图像与所述第二图像进行双目立体匹配，获取所述第一图像对应的深度信息。

根据本公开实施例的又一个方面，提供一种电子设备，包括：

存储器，用于存储可执行指令；以及

处理器，用于与所述存储器通信以执行所述可执行指令从而完成本公开上述任一实施例所述方法的操作。

根据本公开实施例的再一个方面，提供一种计算机程序，包括计算机可读代码，当所述计算机可读代码在设备上运行时，所述设备中的处理器执行用于实现本公开上述任一实施例所述方法中操作的指令。

根据本公开实施例的再一个方面，提供一种计算机可读存储介质，用于存储计算机可读的指令，所述指令被执行时实现本公开上述任一实施例所述方法中的操作。

基于本公开上述实施例提供的深度估计方法和装置、电子设备、计算机程序和计算机可读存储介质，以单张图片作为双目图像中的第一图像，经第一神经网络，基于第一图像获取双目图像中的第二图像，经第二神经网络，基于对第一图像与第二图像进行双目立体匹配，获取第一图像对应的深度信息，由此基于单张图片实现了该单张图片中场景的深度估计，而不需要双目相机，避免了双目相机产生的额外硬件开销，降低了成本；并且，可以避免双目相机设定不准确导致获得的深度信息错误，提高了深度估计的准确性。

下面通过附图和实施例，对本公开的技术方案做进一步的详细描述。

附图说明

构成说明书的一部分的附图描述了本公开的实施例，并且连同描述一起用于解释本公开的原理。

参照附图，根据下面的详细描述，可以更加清楚地理解本公开，其中：

图 1 为本公开深度估计方法一个实施例的流程图；

图 2 为本公开深度估计方法另一个实施例的流程图；

图 3 为本公开深度估计方法一个应用实施例的流程图；

图 4 为图 3 所示实施例对应的示例性框图；

图 5 为本公开深度估计装置一个实施例的结构示意图；

图 6 为本公开深度估计装置另一个实施例的结构示意图；

图 7 为本公开深度估计装置又一个实施例的结构示意图；

图 8 为本公开电子设备一个应用实施例的结构示意图。

具体实施方式

现在将参照附图来详细描述本公开的各种示例性实施例。应注意到：除非另外具体说明，否则在这些实施例中阐述的部件和操作的相对布置、数字表达式和数值不限制本公开的范围。

同时，应当明白，为了便于描述，附图中所示出的各个部分的尺寸并不是按照实际的比例关系绘制的。

以下对至少一个示例性实施例的描述实际上仅仅是说明性的，决不作为对本公开及其应用或使用的任何限制。

对于相关领域普通技术人员已知的技术、方法和设备可能不作详细讨论，但在适当情况下，所述技术、方法和设备应当被视为说明书的一部分。

应注意到：相似的标号和字母在下面的附图中表示类似项，因此，一旦某一项在一个附图中被定义，则在随后的附图中不需要对其进行进一步讨论。

本公开实施例可以应用于终端设备、计算机系统、服务器等电子设备，其可与众多其它通用或专用计算系统环境或配置一起操作。适于与终端设备、计算机系统、服务器等电子设备一起使用的众所周知的终端设备、计算系统、环境和/或配置的例子包括但不限于：个人计算机系统、服务器计算机系统、瘦客户机、厚客户机、手持或膝上设备、基于微处理器的系统、机顶盒、可编程消费电子产品、网络个人电脑、小型计算机系统、大型计算机系统和包括上述任何系统的分布式云计算技术环境，等等。

终端设备、计算机系统、服务器等电子设备可以在由计算机系统执行的计算机系统可执行指令（诸如程序模块）的一般语境下描述。通常，程序模块可以包括例程、程序、目标程序、组件、逻辑、数据结构等等，它们执行特定的任务或者实现特定的抽象数据类型。计算机系统/服务器可以在分布式云计算环境中实施，分布式云计算环境中，任务是由通过通信网络链接的远程处理设备执行的。在分布式云计算环境中，程序模块可以位于包括存储设备的本地或远程计算系统存储介质上。

图 1 为本公开深度估计方法一个实施例的流程图。如图 1 所示，该实施例的深度估计方法包括：

102，以单张图片作为双目图像中的第一图像，经第一神经网络，基于该第一图像获取双目图像中的第二图像。

双目图像为经双目相机拍摄的两张图像、或者经多目相机拍摄的多张图像中的两张图像，可以称为左图和右图。其中，第一图像为左图时，第二图像为右图；或者，第一图像

为右图时，第二图像为左图。另外，双目图像也可以称为主图和副图，双目图像中的任意一张图像作为主图时，另一张图像即作为副图。

在一个可选示例中，该操作 102 可以由处理器调用存储器存储的相应指令执行，也可以由被处理器运行的图像获取模块执行。

104，经第二神经网络，通过对上述第一图像与第二图像进行双目立体匹配，获取上述第一图像对应的深度信息。

在本公开至少一个实施例中，第一神经网络、第二神经网络，分别可以是一个多层神经网络（即：深度神经网络），例如多层的卷积神经网络，例如可以是 LeNet、AlexNet、GoogLeNet、VGG、ResNet 等任意神经网络。第一神经网络和第二神经网络可以采用相同类型和结构的神经网络，也可以采用不同类型和结构的神经网络。

在一个可选示例中，该操作 104 可以由处理器调用存储器存储的相应指令执行，也可以由被处理器运行的立体匹配模块执行。

在实现本公开的过程中，发明人通过研究发现，目前的深度估计方法主要可以分为两类。一类是使用大量的像素级别的深度标签对神经网络进行监督、通过训练好的神经网络获取进行深度估计，但是获取深度标签不仅花费极大，现有的技术也不能得到高质量且稠密的深度标签。第二类是基于双目立体匹配的深度估计方法，在这种深度估计方法中，两张从不同方位拍摄的图像将被作为输入，基于几何空间的规则，深度可以通过计算两张图相对应像素的视差而获得。但是，该类预测方法的准确率受限于双目相机的设定，而且该类方法由于需要双目相机产生额外的硬件开销。而基于本公开上述实施例提供的深度估计方法，以单张图片作为双目图像中的第一图像，经第一神经网络，基于第一图像获取双目图像中的第二图像，经第二神经网络，通过对第一图像与第二图像进行双目立体匹配，获取深度信息，由此基于单张图片实现了该单张图片中场景的深度估计，而不需要双目相机，避免了双目相机产生的额外硬件开销，降低了成本；并且，可以避免双目相机设定不准确导致获得的深度信息错误，提高了深度估计的准确性。

图 2 为本公开深度估计方法另一个实施例的流程图。如图 2 所示，该实施例的深度估计方法包括：

202，以单张图片作为双目图像中的第一图像，经第一神经网络对该第一图像进行处理，输出 N 个通道的视差概率图。

其中，N 个通道中不同的通道对应不同的视差，每个通道的视差概率图的表示第一图像上像素向第一水平方向偏移 i 个视差的概率，概率归一化后，同一个像素在所有通道上的概率值之和为 1； $i=0,1,\dots,N-1$ ，N 的取值为大于 1 的整数。第一图像为左图时，第一水平方向为水平向左的方向；第一图像为右图时，第一水平方向为水平向右的方向。即，第一图像为左图时，第 i 个通道的视差概率图表示该左图上像素水平向左偏移某 i 个视差的概率，例如，假设 N 的取值为 5、第一图像为左图，第 1、2、3、4、5 个通道的视差概率图分别表示该左图上像素水平向左偏移 0、1、2、3、4 个视差的概率，该左图上某一个像素水平向左偏移 0、1、2、3、4 个视差的概率例如可以分别为 0.3、0.4、0.2、0.1 和 0。

在一个可选示例中，该操作 202 可以由处理器调用存储器存储的相应指令执行，也可以由被处理器运行的图像获取模块的第一子神经网络单元执行。

204，根据上述 N 个通道的视差概率图，将第一图像分别向第一水平方向偏移 i 个像素，得到 N 张偏移图。

其中， $i=0,1,\dots,N-1$ ，N 的取值为大于 1 的整数。

在一个可选示例中，该操作 204 可以由处理器调用存储器存储的相应指令执行，也可以由被处理器运行的图像获取模块的偏移单元执行。

206，将上述 N 张偏移图中的各张偏移图分别点乘对应通道的视差概率图，得到 N 个

点乘结果。

在一个可选示例中，该操作 206 可以由处理器调用存储器存储的相应指令执行，也可以由被处理器运行的图像获取模块的点乘单元执行。

208，将上述 N 个点乘结果基于像素进行叠加，得到上述双目图像中的第二图像。

由于双目相机拍摄的左、右图像遵守基本的空间规则，当双目图像中的第一图像为双目相机拍摄的左图时，双目图像中的第二图像为双目相机拍摄的右图；或者，当双目图像中的第一图像为双目相机拍摄的右图时，双目图像中的第二图像为双目相机拍摄的左图，因此本公开实施例获取到的第二图像与第一图像相对应像素的视差遵守空间几何规律。在本公开实施例的其中一个实施方式中，第一图像中对应前景对象的像素在视差概率图中的对应像素位置，在对应较大视差的通道的视差概率图中具有较大的概率值；第一图像中对应背景对象的像素在视差概率图中的对应像素位置，在对应较小视差的通道的视差概率图中具有较大的概率值。例如，第一图像中包括背景和作为前景对象的人脸，对应人脸的像素在 N 个通道的视差概率图中对应较大视差的通道的视差概率图中的概率值为 0.8，在 N 个通道的视差概率图中对应较小视差的通道的视差概率图中的概率值为 0.1；对应背景的像素在 N 个通道的视差概率图中对应较小视差的通道的视差概率图中的概率值为 0.9，在 N 个通道的视差概率图中对应较大视差的通道的视差概率图中的概率值为 0。

在一个可选示例中，该操作 208 可以由处理器调用存储器存储的相应指令执行，也可以由被处理器运行的图像获取模块的加法单元执行。

210，通过第二神经网络，获取用于表示双目图像中一个图像相对于另一个图像中像素的位置关系的相关系数。

因为将第一图像中的像素在第一水平方向上移动即可得到第二图像中的像素，可以将第一图像中的每个像素位置分别视为一个变量，该变量的取值为视差概率图中对应的视差的取值，可以是 0, 1, ..., N-1 共 N 个。获取第一图像中每个像素位置的变量与第二图像中相邻 d 个像素位置的变量的相关系数并保存在对应的像素位置和通道，由第一图像中所有像素位置的相关系数即得到第一图像相对于第二图像中像素的位置关系的相关系数。其中，d 的取值为整数，例如可以取值 -40~+40。双目图像中一个图像相对于另一个图像中像素的位置关系的相关系数包括一个图像中所有像素位置的变量与第二图像中相邻 d 个像素位置的变量的相关系数，其可以表示为一个 $W \times H \times N$ 的相关系数图或者一个相关系数矩阵。其中，W、H、N 分别表示一个图像的宽度、高度和通道数，W、H、N 的取值均为大于 0 的整数。

假设上述一个图像为左图，另一个图像为右图，由于左图和右图在竖直方向上是对齐的，将左图的像素在水平方向上向左移动即可得到右图的像素。因此，在水平方向上计算至少一个像素位置的相关系数有助于在水平方向上更好的匹配左右图对应的像素。

其中，双目图像中一个图像包括第一图像或第二图像，另一个图像对应为该双目图像中的第二图像或第一图像。

在一个可选示例中，该操作 210 可以由处理器调用存储器存储的相应指令执行，也可以由被处理器运行的立体匹配模块的第一获取单元执行。

212，基于另一个图像与相关系数，生成第一图像与第二图像的视差图 (disparity)。

其中，视差图中至少一个像素的取值分别表示第一图像拍摄场景中的某一点的视差，即：该某一点在第一图像坐标系中的坐标与在第二图像坐标系中的坐标之间的差值。

在一个可选示例中，该操作 212 可以由处理器调用存储器存储的相应指令执行，也可以由被处理器运行的立体匹配模块的生成单元执行。

214，基于上述视差图获取第一图像对应的深度信息。

在本公开实施例的又一个实施方式中，可以基于上述视差图和相机参数获取上述第一

图像对应的深度信息，例如，可以基于上述视差图、拍摄第一图像的相机的焦距和双目图像对应的双目相机之间的距离，获取上述第一图像对应的深度信息。

例如，可以通过但不限于公式： $Z=f \times B / \text{Disp}$ ，获取上述第一图像对应的深度信息。

其中，Disp 代表预测的视差图，f 为拍摄第一图像的相机的焦距，B 为双目相机之间的距离，Z 即为所要预测的单目全局深度图。

在一个可选示例中，该操作 214 可以由处理器调用存储器存储的相应指令执行，也可以由被处理器运行的立体匹配模块的第三获取单元执行。

本公开实施例，将单目深度估计的问题转化为双目立体匹配的问题，通过这种方式，将较为困难的深度估计问题转化为了匹配两张图像相似像素点的问题，这种匹配不再需要推测单张图像中至少两个像素间的几何关系，降低了计算的复杂度。另外，本公开实施例利用深度学习方法，可以更好的实现合成第二图像和双目立体匹配两个操作，并通过将几何的变换显性地设置在第一神经网络和第二神经网络中，提高了操作结果的准确性。

本公开实施例通过使用单张图片合成相对应的右图再进行双目立体匹配，在用于对第一神经网络和第二神经网络进行训练时，不再需要大量精密的深度标签作为监督，只需要使用容易得到的矫正好的双目图像训练第一神经网络（也称为图像合成网络），使用大量计算机渲染的双目图像和深度图训练第二神经网络（也称为：双目立体匹配网络），相比于大量精密的深度标签，降低了训练数据开销。

在本公开实施例的其中一个实施方式中，操作 202 中，经第一神经网络对第一图像进行处理，输出视差概率图，可以包括：

分别通过第一神经网络中的两个或以上网络深度的网络层对第一图像进行特征提取，得到两个或以上尺度（即：大小）的特征图；在本公开中，至少两个为两个或者两个以上。

基于上述两个或以上尺度的特征图分别获取两个或以上分辨率的、N 个通道的初步视差概率图；

分别针对每个通道，将上述两个或以上分辨率的初步视差概率图放大至第一图像的分辨率下进行叠加，得到 N 个通道的视差概率图。

因为神经网络中的池化层效果，在神经网络的不同阶段会产生不同大小、分辨率的特征图，基于不同大小、分辨率的特征图能够产生不同大小、分辨率的初步视差概率图，从而有助于为预测深度信息提供不同的局部信息和全局信息。

例如，第一图像是 $W \times H \times N$ 为 $200 \times 200 \times 3$ 的红绿蓝（RGB）图，通过第一神经网络的某一网络深度的网络层得到 $100 \times 100 \times 64$ 的特征图，再继续经过另一网络深度的网络层得到 $50 \times 50 \times 128$ 的特征图，基于这两个不同大小的特征图，可以获得不同大小、分辨率的初步视差概率图，例如得到 $100 \times 100 \times N$ 以及 $50 \times 50 \times N$ 的初步视差概率图。其中，第一图像的分辨率为 200×200 ，两个初步视差概率图的分辨率分别为 100×100 和 50×50 ，两个初步视差概率图的分辨率分别为第一图像的分辨率大小的 $1/2 \times 1/2$ 、 $1/4 \times 1/4$ 。

基于本实施例，第一神经网络中的两个或以上网络深度的网络层提取的特征的尺寸不同，其中，网络深度较浅的网络层提取的特征感受野较小，体现第一图像中较小区域的信息，网络深度较深的网络层提取的特征感受野较大，可以体现第一图像中较大区域的信息、甚至全局信息，使用不同分辨率的特征图同时提供不同视野域的信息，可以产生更准确的概率视差图。

在本公开实施例的其中一个实施方式中，操作 210 可以包括：

分别对一个图像与另一个图像进行特征提取。例如，可以通过一段卷积神经网络，分别对一个图像与另一个图像进行特征提取；

通过第二神经网络，基于提取的一个图像的特征与另一个图像的特征，获取一个图像

与另一个图像中像素的位置关系，并输出相关系数。

相应地，在本公开实施例的另一个实施方式中，操作 212 可以包括：将另一个图像的特征与相关系数进行叠加，生成第一图像与第二图像的视差图。这样，基于单张图像即可获得该单张图像拍摄场景在双目图像中的视差图，将较为困难的深度估计问题转化为了匹配两张图像相似像素点的问题，这种匹配不再需要推测单张图像中至少两个像素间的几何关系，降低了计算的复杂度。另外，本实施例利用深度学习方法，将几何变换显性地设置在第二神经网络中，提高了操作结果的准确性。

可选地，在本公开实施例的又一个实施方式中，该操作 212 可以包括：将另一个图像的特征与相关系数进行叠加，得到叠加结果，该叠加结果例如可以是一个特征图；提取叠加结果的特征，并将提取的叠加结果的特征与叠加结果进行融合，获得第一图像与第二图像的视差图。

在其中一个实施方式中，可以通过一段卷积神经网络提取叠加结果的特征，该卷积神经网络例如可以示例性地包括但不限于一层卷积层和一层激活层（ReLU）。该卷积神经网络例如可以通过一个编码-解码模型实现，通过卷积层对叠加结果进行特征提取，得到一个与叠加结果相同大小的特征图，将该特征图与叠加结果进行融合（concat），获得第一图像与第二图像的视差图。

本公开实施例通过对叠加结果进行进一步特征提取，可以加大感受野的范围，再将提取的叠加结果的特征与叠加结果进行融合，获得第一图像与第二图像的视差图，使得视差图可以融合较多的信息，能够获取更多的全局信息，从而有助于提升后续预测的第一图像对应的深度信息。

图 3 为本公开深度估计方法一个应用实施例的流程图。图 4 为图 3 所示实施例对应的示例性框图。该应用实施例中，分别以左图、右图作为本公开上述至少一个实施例中的第一图像和第二图像进行说明。参见图 3 和图 4，该应用实施例包括：

302，以单张图片作为双目图像中的左图，经第一神经网络对该左图进行处理，输出 N 个通道的视差概率图。

其中，每个通道分别表示左图上像素向水平向左偏移 i 个视差的概率； $i=0,1,\dots,N-1$ ， N 的取值为大于 1 的整数。

在一个可选示例中，该操作 302 可以由处理器调用存储器存储的相应指令执行，也可以由被处理器运行的图像获取模块的第一子神经网络单元执行。

304，根据上述 N 个通道的视差概率图，将左图分别向第一水平方向偏移 i 个像素，得到 N 张偏移图。

在一个可选示例中，该操作 304 可以由处理器调用存储器存储的相应指令执行，也可以由被处理器运行的图像获取模块的偏移单元执行。

306，将上述 N 张偏移图中的各张偏移图分别点乘对应通道的视差概率图，得到 N 个点乘结果。

在一个可选示例中，该操作 306 可以由处理器调用存储器存储的相应指令执行，也可以由被处理器运行的图像获取模块的点乘单元执行。

308，将上述 N 个点乘结果基于像素进行叠加，得到双目图像中的右图。

在一个可选示例中，该操作 308 可以由处理器调用存储器存储的相应指令执行，也可以由被处理器运行的图像获取模块的加法单元执行。

310，通过第二神经网络，分别对左图与右图进行特征提取。

在一个可选示例中，该操作 310 可以由处理器调用存储器存储的相应指令执行，也可以由被处理器运行的立体匹配模块的第一获取单元的第二子神经网络单元执行。

312，通过第二神经网络，基于提取的左图的特征与右图的特征，获取左图与右图中

像素的位置关系，并输出相关系数。

在一个可选示例中，该操作 312 可以由处理器调用存储器存储的相应指令执行，也可以由被处理器运行的立体匹配模块的第一获取单元的获取子单元执行。

314，将左图的特征与相关系数进行叠加，得到叠加结果，该叠加结果例如可以是一个特征图。

可选地，为了得到与左图像素对齐的视差图，可以先将从左图得到的特征图通过神经网络进行进一步特征提取，再将提取到的特征与相关系数进行叠加。其中，该神经网络可以示例性地由一层卷积层和一层激活层组成，对从左图得到的特征图进行进一步特征提取，可以加大感受野的范围，得到进一步处理后的特征图（feature），再与相关系数进行叠加，从而使得叠加结果可以包括更多的全局信息，提高后续获得的视差图和深度信息的准确性。

在一个可选示例中，该操作 314 可以由处理器调用存储器存储的相应指令执行，也可以由被处理器运行的立体匹配模块的生成单元的叠加子单元执行。

316，通过第二神经网络，提取叠加结果的特征，并将提取的叠加结果的特征与叠加结果进行融合，获得第一图像与第二图像的视差图。

在一个可选示例中，该操作 316 可以由处理器调用存储器存储的相应指令执行，也可以由被处理器运行的立体匹配模块的生成单元的融合子单元执行。

318，基于上述视差图、拍摄第一图像的相机的焦距和双目图像对应的双目相机之间的距离，获取第一图像对应的深度信息。

在一个可选示例中，该操作 318 可以由处理器调用存储器存储的相应指令执行，也可以由被处理器运行的立体匹配模块的第三获取单元执行。

基于本公开上述至少一个实施例，获得深度信息后，还可以基于第一图像对应的深度信息与第二图像，获取该第一图像中场景的三维空间场景。

本公开实施例例如可以应用但不限于以下方面：

1，对单张图片进行全局深度估计；

2，本公开实施例可以被应用于三维场景重建，使用预测得到的第一图像对应的深度信息（也称为：全局深度图）可被应用到多种场景中，例如自动驾驶、三维场景恢复、3D 电影制作等等。使用本公开实施例，只需要单张图片即可得到较好的效果，降低了成本。

使用预测得到的第一图像对应的深度信息、结合合成的右图，可以将原图（即：第一图像）中的整个场景的三维空间场景恢复出来，恢复出来的三维空间场景有许多应用场景，例如 3D 电影、自动驾驶等等。

可选的，本公开上述至少一个实施例中，第一神经网络可以利用第一样本集中的样本双目图像预先训练而得，该第一样本集包括至少一组第一样本双目图像；和/或，第二神经网络可以利用第二样本集中的样本双目图像预先训练而得。

相应地，在本公开上述至少一个实施例之前，还可以包括：

利用第一样本集中的样本双目图像对第一神经网络进行训练，以及利用第二样本集中的样本双目图像和深度图对第二神经网络进行训练。其中，第一样本集包括至少一组第一样本双目图像，每组第一样本双目图像包括第一图像和第二图像；第二样本集包括至少一组第二样本双目图像和视差图标签。

在其中一个实施方式中，利用第一样本集中的样本双目图像对第一神经网络进行训练，可以包括：

经第一神经网络，由至少一组第一样本双目图像中的第一图像，获取至少一组第一样本双目图像中的第二图像并输出；

获取第一神经网络输出的第二图像与至少一组第一样本双目图像中的第二图像之间的第一差异，并基于第一差异对第一神经网络进行训练，直至满足第一训练完成条件。

在其中一个可选示例中，获取第一神经网络输出的第二图像与至少一组第一样本双目图像中的第二图像之间的第一差异，并基于第一差异对第一神经网络进行训练，直至满足第一训练完成条件，可以包括：

获取第一神经网络输出的第二图像与至少一组第一样本双目图像中的第二图像之间在像素上的第一差异，例如，在像素上差值的绝对值之和；

基于第一差异调整第一神经网络中网络参数的参数值，直至满足第一训练完成条件。

其中，第一训练完成条件例如可以包括但不限于：第一差异小于第一预设阈值，和/或，对第一神经网络的训练次数达到第一预设次数。

在另一个实施方式中，利用第二样本集中的样本双目图像和视差图标签对第二神经网络进行训练，可以包括：

经第二神经网络，获取至少一组第二样本双目图像的视差图并输出；

获取第二神经网络输出的视差图与至少一组第二样本双目图像的视差图标签之间的第二差异，并基于第二差异对第二神经网络进行训练，直至满足第二训练完成条件。

在其中一个可选示例中，获取第二神经网络输出的视差图与至少一组第二样本双目图像的视差图标签之间的第二差异，并基于第二差异对第二神经网络进行训练，直至满足第二训练完成条件，可以包括：

获取第二神经网络输出的视差图与至少一组第二样本双目图像的视差图标签之间在像素上的第二差异，例如在像素上差值的绝对值之和；

基于第二差异调整第二神经网络中网络参数的参数值，直至满足第二训练完成条件。

其中，第二训练完成条件例如可以包括但不限于：第二差异小于第二预设阈值，和/或，对第二神经网络的训练次数达到第二预设次数。

由于真实采集的深度图标签不容易获取，基于本公开实施例在实际应用中，可以使用计算机合成的左图、右图和深度图标签作为第二样本集中的第二样本双目图像和第二样本图像对应的深度图标签，训练第二神经网络。

另外，通过本公开上述实施例对第一神经网络和第二神经网络分阶段训练完成后，还可以包括：

利用第三样本集中的样本双目图像和第三样本图像对应的深度图标签对第一神经网络和第二神经网络进行训练。

其中，第三样本集包括至少一组第三样本双目图像和第三样本图像对应的深度图标签。

在其中一个实施方式中，利用第三样本集中的样本双目图像和第三样本图像对应的深度图标签对第一神经网络和第二神经网络进行训练，可以包括：

经第一神经网络，由至少一组第三样本双目图像中的第一图像，获取至少一组第三样本双目图像中的第二图像；

经第二神经网络，获取至少一组第三样本双目图像的视差图；

基于至少一组第三样本双目图像的视差图获取至少一组深度信息；

获取至少一组深度信息与至少一组第三样本双目图像的深度图标签之间的第三差异，该第三差异例如可以是在像素上差值的绝对值之和；

基于第三差异调整第一神经网络和第二神经网络中网络参数的参数值，直至满足第三训练完成条件。

其中，第三训练完成条件例如可以包括但不限于：第三差异小于第三预设阈值，和/或，对第一神经网络和第二神经网络的训练次数达到第三预设次数。

本公开实施例提供的任一种深度估计方法可以由任意适当的具有数据处理能力的设备执行，包括但不限于：终端设备和服务器等。或者，本公开实施例提供的任一种深度估计方法可以由处理器执行，如处理器通过调用存储器存储的相应指令来执行本公开实施例

提及的任一种深度估计方法。下文不再赘述。

本领域普通技术人员可以理解：实现上述方法实施例的全部或部分操作可以通过程序指令相关的硬件来完成，前述的程序可以存储于一计算机可读取存储介质中，该程序在执行时，执行包括上述方法实施例的操作；而前述的存储介质包括：ROM、RAM、磁碟或者光盘等至少一个种可以存储程序代码的介质。

图 5 为本公开深度估计装置一个实施例的结构示意图。该实施例的深度估计装置可用于实现本公开上述至少一个深度估计方法实施例。如图 5 所示，该实施例的深度估计装置包括：图像获取模块和立体匹配模块。其中：

图像获取模块，用于以单张图片作为双目图像中的第一图像，经第一神经网络，基于第一图像获取双目图像中的第二图像。

立体匹配模块，用于经第二神经网络，通过对第一图像与第二图像进行双目立体匹配，获取第一图像对应的深度信息。

基于本公开上述实施例提供的深度估计装置，以单张图片作为双目图像中的第一图像，经第一神经网络，基于第一图像获取双目图像中的第二图像，经第二神经网络，基于对第一图像与第二图像进行双目立体匹配，获取第一图像对应的深度信息，由此基于单张图片实现了该单张图片中场景的深度估计，而不需要双目相机，避免了双目相机产生的额外硬件开销，降低了成本；并且，可以避免双目相机设定不准确导致获得的深度信息错误，提高了深度估计的准确性。

图 6 为本公开深度估计装置另一个实施例的结构示意图。如图 6 所示，在本公开至少一个实施例的其中一个实施方式中，图像获取模块包括：第一子神经网络单元，偏移单元，点乘单元和加法单元。其中：

第一子神经网络单元，用于对第一图像进行处理，输出 N 个通道的视差概率图；其中，每个通道的视差概率图表示第一图像上像素向第一水平方向偏移 i 个视差的概率， $i=0,1,\dots,N-1$ ， N 的取值为大于 1 的整数。

偏移单元，用于根据 N 个通道的视差概率图，将第一图像分别向第一水平方向偏移 i 个像素，得到 N 张偏移图。

点乘单元，用于将 N 张偏移图中的各张偏移图分别点乘对应通道的视差概率图中，得到 N 个点乘结果。

加法单元，用于将 N 个点乘结果基于像素进行叠加，得到第二图像。

在其中一个可选示例中，第一子神经网络单元，包括两个或以上网络深度的网络层，用于：分别通过两个或以上网络深度的网络层对第一图像进行特征提取，获得两个或以上尺度的特征图；基于两个或以上尺度的特征图分别获取两个或以上分辨率的、 N 个通道的初步视差概率图；以及分别针对每个通道，将两个或以上分辨率的初步视差概率图放大至第一图像的分辨率下进行叠加，得到 N 个通道的视差概率图。

可选地，第一图像中对应前景对象的像素在 N 个通道的视差概率图中的对应像素位置，在对应较大视差的通道的视差概率图中具有较大的概率值；第一图像中对应背景对象的像素在 N 个通道的视差概率图中的对应像素位置，在对应较小视差的通道的视差概率图中具有较大的概率值。

另外，再参见图 6，在本公开至少一个实施例的另一个实施方式中，立体匹配模块可以包括：第一获取单元，生成单元和第三获取单元。其中：

第一获取单元，用于获取用于表示双目图像中一个图像相对于另一个图像中像素的位置关系的相关系数；双目图像中一个图像包括第一图像或第二图像，另一个图像对应包括第二图像或第一图像。

生成单元，用于基于另一个图像与相关系数，生成第一图像与第二图像的视差图。

第三获取单元，基于视差图获取第一图像对应的深度信息。

在其中一个可选示例中，第一获取单元可以包括：第二子神经网络单元，用于分别对一个图像与另一个图像进行特征提取；获取子单元，用于基于提取的一个图像的特征与另一个图像的特征，获取一个图像与另一个图像中像素的位置关系，并输出用于表示双目图像中一个图像相对于另一个图像中像素的位置关系的相关系数。

相应地，在另一个可选示例中，生成单元用于将另一个图像的特征与相关系数进行叠加，生成第一图像与第二图像的视差图。

在另一个可选示例中，生成单元可以包括：叠加子单元，用于将另一个图像的特征与相关系数进行叠加，得到叠加结果；融合子单元，用于提取叠加结果的特征，并将提取的叠加结果的特征与叠加结果进行融合，获得第一图像与第二图像的视差图。

在又一个可选示例中，第三获取单元用于基于视差图、拍摄第一图像的相机的焦距和双目图像对应的双目相机之间的距离，获取第一图像对应的深度信息。

另外，本公开上述至少一个实施例的深度估计装置中，还可以包括：获取模块，用于基于第一图像对应的深度信息与第二图像，获取第一图像中场景的三维空间场景。

如图7所示，为本公开深度估计装置又一个实施例的结构示意图。图7中，图像获取模块和立体匹配模块可以选择性地采用图6所示任一实施例的结构实现，也可以采用其他结构实现。

另外，本公开上述至少一个实施例的深度估计装置中，第一神经网络可以利用第一样本集中的样本双目图像预先训练而得，第一样本集包括至少一组第一样本双目图像。和/或，第二神经网络可以利用第二样本集中的样本双目图像预先训练而得；第二样本集包括至少一组第二样本双目图像和视差图标签。

再参见图7，在本公开深度估计装置的再一个实施例中，还包括第一训练模块。该实施例中，第一神经网络，用于由至少一组第一样本双目图像中的第一图像，获取至少一组第一样本双目图像中的第二图像并输出。第一训练模块，用于获取第一神经网络输出的第二图像与至少一组第一样本双目图像中的第二图像之间的第一差异，并基于第一差异对第一神经网络进行训练，直至满足第一训练完成条件。

在其中一个实施方式中，第一训练模块用于：获取第一神经网络输出的第二图像与至少一组第一样本双目图像中的第二图像之间在像素上的第一差异；

基于第一差异调整第一神经网络中网络参数的参数值，直至满足第一训练完成条件。其中，第一训练完成条件例如可以包括但不限于：第一差异小于第一预设阈值，和/或，对第一神经网络的训练次数达到第一预设次数。

再参见图7，在本公开深度估计装置的再一个实施例中，还可以包括第二训练模块。该实施例中，第二神经网络，用于获取至少一组第二样本双目图像的视差图并输出。第二训练模块，用于获取第二神经网络输出的视差图与至少一组第二样本双目图像的视差图标签之间的第二差异，并基于第二差异对第二神经网络进行训练，直至满足第二训练完成条件。

在其中一个实施方式中，第二训练模块具体用于：获取第二神经网络输出的视差图与至少一组第二样本双目图像的视差图标签之间在像素上的第二差异；

基于第二差异调整第二神经网络中网络参数的参数值，直至满足第二训练完成条件。其中，第二训练完成条件例如可以包括但不限于：第二差异小于第二预设阈值，和/或，对第二神经网络的训练次数达到第二预设次数。

进一步地，再参见图7，在本公开深度估计装置的再一个实施例中，还可以包括第三训练模块，用于利用第三样本集中的样本双目图像和第三样本图像对应的深度图标签对第一神经网络和第二神经网络进行训练。其中，第三样本集包括至少一组第三样本双目图像

和第三样本图像对应的深度图标签。

在其中一个实施方式中，第一神经网络用于由至少一组第三样本双目图像中的第一图像，获取至少一组第三样本双目图像中的第二图像；第二神经网络，用于获取至少一组第三样本双目图像的视差图。

第三训练模块用于：基于至少一组第三样本双目图像的视差图获取至少一组深度信息；获取至少一组深度信息与至少一组第三样本双目图像的深度图标签之间的第三差异；基于第三差异调整第一神经网络和第二神经网络中网络参数的参数值，直至满足第三训练完成条件。其中，第三训练完成条件例如可以包括但不限于：第三差异小于第三预设阈值，和/或，对第一神经网络和第二神经网络的训练次数达到第三预设次数。本公开深度估计装置实施例中至少一个方案的技术效果，可参加相应方法实施例中的相应描述，在此不再赘述。

另外，本公开实施例提供一种电子设备，包括：

存储器，用于存储可执行指令；以及

处理器，用于与所述存储器通信以执行所述可执行指令从而完成本公开上述任一实施例所述深度估计方法的操作。

图8为本公开电子设备一个应用实施例的结构示意图。下面参考图8，其示出了适于用来实现本公开实施例的终端设备或服务器的电子设备的结构示意图。如图8所示，该电子设备包括一个或多个处理器、通信部等，所述一个或多个处理器例如：一个或多个中央处理单元（CPU）801，和/或一个或多个加速单元（GPU）813等，加速单元813可包括但不限于GPU、FPGA、其他类型的专用处理器等，处理器可以根据存储在只读存储器（ROM）802中的可执行指令或者从存储部分808加载到随机访问存储器（RAM）803中的可执行指令而执行各种适当的动作和处理。通信部812可包括但不限于网卡，所述网卡可包括但不限于IB（Infiniband）网卡，处理器可与只读存储器802和/或随机访问存储器803中通信以执行可执行指令，通过总线804与通信部812相连、并经通信部812与其他目标设备通信，从而完成本公开实施例提供的任一方法对应的操作，例如，以单张图片作为双目图像中的第一图像，经第一神经网络，基于所述第一图像获取所述双目图像中的第二图像；经第二神经网络，通过对所述第一图像与所述第二图像进行双目立体匹配，获取所述第一图像对应的深度信息。

本公开电子设备中各方案的技术效果，可参加相应方法实施例中的相应描述，在此不再赘述。

此外，在RAM803中，还可存储有装置操作所需的各种程序和数据。CPU801、ROM802以及RAM803通过总线804彼此相连。在有RAM803的情况下，ROM802为可选模块。RAM803存储可执行指令，或在运行时向ROM802中写入可执行指令，可执行指令使处理器执行本公开上述任一方法对应的操作。输入/输出（I/O）接口805也连接至总线804。通信部812可以集成设置，也可以设置为具有多个子模块（例如多个IB网卡），并在总线链路上。

以下部件连接至I/O接口805：包括键盘、鼠标等的输入部分806；包括诸如阴极射线管（CRT）、液晶显示器（LCD）等以及扬声器等的输出部分807；包括硬盘等的存储部分808；以及包括诸如LAN卡、调制解调器等网络接口卡的通信部分809。通信部分809经由诸如因特网的网络执行通信处理。驱动器810也根据需要连接至I/O接口805。可拆卸介质811，诸如磁盘、光盘、磁光盘、半导体存储器等等，根据需要安装在驱动器810上，以便于从其上读出的计算机程序根据需要被安装入存储部分808。

需要说明的，如图8所示的架构仅为一种可选实现方式，在具体实践过程中，可根据实际需要对上述图8的部件数量和类型进行选择、删减、增加或替换；在不同功能部件设置上，也可采用分离设置或集成设置等实现方式，例如加速单元813和CPU801可分离设

置或者可将加速单元 813 集成在 CPU801 上, 通信部 812 可分离设置, 也可集成设置在 CPU801 或加速单元 813 上, 等等。这些可替换的实施方式均落入本公开公开的保护范围。

特别地, 根据本公开的实施例, 上文参考流程图描述的过程可以被实现为计算机软件程序。例如, 本公开的实施例包括一种计算机程序产品, 其包括有形地包含在机器可读介质上的计算机程序, 计算机程序包含用于执行流程图所示的方法的程序代码, 程序代码可包括对应执行本公开实施例提供的深度估计方法操作对应的指令。在这样的实施例中, 该计算机程序可以通过通信部分从网络上被下载和安装, 和/或从可拆卸介质被安装。在该计算机程序被 CPU 执行时, 执行本公开的方法中限定的上述功能。

另外, 本公开实施例还提供了一种计算机程序, 包括计算机可读代码, 当所述计算机可读代码在设备上运行时, 所述设备中的处理器执行用于实现本公开任一实施例所述深度估计方法中操作的指令。

另外, 本公开实施例还提供了一种计算机可读存储介质, 用于存储计算机可读取的指令, 其特征在于, 指令被执行时实现本公开任一实施例所述深度估计方法中的操作。

本领域普通技术人员可以理解: 实现上述方法实施例的全部或部分操作可以通过程序指令相关的硬件来完成, 前述的程序可以存储于一计算机可读取存储介质中, 该程序在执行时, 执行包括上述方法实施例的 c 操作; 而前述的存储介质包括: ROM、RAM、磁碟或者光盘等各种可以存储程序代码的介质。

本说明书中各个实施例均采用递进的方式描述, 每个实施例重点说明的都是与其它实施例的不同之处, 各个实施例之间相同或相似的部分相互参见即可。对于系统实施例而言, 由于其与方法实施例基本对应, 所以描述的比较简单, 相关之处参见方法实施例的部分说明即可。

可能以许多方式来实现本公开的方法和装置。例如, 可通过软件、硬件、固件或者软件、硬件、固件的任何组合来实现本公开的方法和装置。用于所述方法的操作的上述顺序仅是为了进行说明, 本公开的方法的操作不限于以上具体描述的顺序, 除非以其它方式特别说明。此外, 在一些实施例中, 还可将本公开实施为记录在记录介质中的程序, 这些程序包括用于实现根据本公开的方法的机器可读指令。因而, 本公开还覆盖存储用于执行根据本公开的方法的程序的记录介质。

本公开的描述是为了示例和描述起见而给出的, 而并不是无遗漏的或者将本公开限于所公开的形式。很多修改和变化对于本领域的普通技术人员而言是显然的。选择和描述实施例是为了更好说明本公开的原理和实际应用, 并且使本领域的普通技术人员能够理解本公开从而设计适于特定用途的带有各种修改的各种实施例。