

单目深度估计方法及其装置、设备和存储介质

相关申请的交叉引用

本申请基于申请号为 201810496541.6、申请日为 2018 年 05 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申请。

技术领域

本申请实施例涉及人工智能领域，尤其涉及一种单目深度估计方法及其装置、设备和存储介质。

背景技术

单目深度估计是计算机视觉中的重要问题，单目深度估计的具体任务指的是预测一张图片中每个像素点的深度。其中，由每个像素点的深度值组成的图片又称为深度图。单目深度估计对于自动驾驶中的障碍物检测、三维场景重建，场景立体分析有着重要的意义。另外单目深度估计可以间接地提高其他计算机视觉任务的性能，比如物体检测、目标跟踪与目标识别。

目前存在的问题是训练用于单目深度估计的神经网络需要大量标记的数据，但是获取标记数据成本很大。在室外环境下标记数据可以通过激光雷达获取，但是获取的标记数据是非常稀疏的，用这样的标记数据训练得到的单目深度估计网络没有清晰的边缘以及不能捕捉细小物体的正确深度信息。

发明内容

本申请实施例提供一种单目深度估计方法及其装置、设备和存储介质。

本申请实施例的技术方案是这样实现的：

本申请实施例提供一种单目深度估计方法，所述方法包括：获取待处理图像；将所述待处理图像输入至经过训练得到的单目深度估计网络模型，得到所述待处理图像的分析结果，其中，所述单目深度估计网络模型是通过第一双目匹配神经网络模型输出的视差图进行监督训练的；输出所述待处理图像的分析结果。

本申请实施例提供一种单目深度估计装置，所述装置包括：获取模块、执行模块和输出模块，其中：所述获取模块，配置为获取待处理图像；所述执行模块，配置为将所述待处理图像输入至经过训练得到的单目深度估计网络模型，得到所述待处理图像的分析结果，其中，所述单目深度估计网络模型是通过第一双目匹配神经网络模型输出的视差图进行监督训练的；所述输出模块，配置为输出所述待处理图像的分析结果。

本申请实施例提供一种单目深度估计设备，包括存储器和处理器，所述存储器存储有可在处理器上运行的计算机程序，所述处理器执行所述程序时实现本申请实施例提供的单目深度估计方法中的步骤。

本申请实施例提供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本申请实施例提供的单目深度估计方法中的步骤。

本申请实施例中，通过获取待处理图像；将所述待处理图像输入至经过训练得到的单目深度估计网络模型，得到所述待处理图像的分析结果，其中，所述单目深度估计网络模型是通过第一双目匹配神经网络模型输出的视差图进行监督训练的；输出所述待处理图像的分析结果；从而能够使用更少或者不使用有深度图标记的数据训练单目深度估计网络，并且提出了一种更有效的无监督微调双目视差网络的方法，从而间接提高了单目深度估计的效果。

附图说明

图 1A 为本申请实施例单目深度估计方法的实现流程示意图一；

图 1B 为本申请实施例单个图片深度估计示意图；

图 1C 为本申请实施例训练第二双目匹配神经网络模型示意图；

图 1D 为本申请实施例训练单目深度估计网络模型示意图；

图 1E 为本申请实施例损失函数相关图片示意图；

图 2A 为本申请实施例单目深度估计方法的实现流程示意图二；

图 2B 为本申请实施例损失函数效果示意图；

图 2C 为本申请实施例可视化深度估计结果示意图；

图 3 为本申请实施例单目深度估计装置的组成结构示意图；

图 4 为本申请实施例单目深度估计设备的一种硬件实体示意图。

具体实施方式

为使本申请实施例的目的、技术方案和优点更加清楚，下面将结合本申请实施例中的附图，对申请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申请，但不用来限制本申请的范围。

在后续的描述中，使用用于表示元件的诸如“模块”、“部件”或“单元”的后缀仅为了有利于本申请的说明，其本身没有特定的意义。因此，“模块”、“部件”或“单元”可以混合地使用。

一般地，利用深度神经网络来预测单张图片的深度图，仅需要一张图片即可以对图片对应的场景进行三维建模，得到每个像素点的深度。本申请实施例提出的单目深度估计方法使用神经网络训练得到，训练数据来自双目匹配输出的视差图数据，而不需要昂贵的深度采集设备如激光雷达。提供训练数据的双目匹配算法也是通过神经网络实现，该网络通过渲染引擎渲染的大量虚拟双目图片对进行预训练即可达到很好的效果，另外可以在真实数据上再进行微调训练以达到更好的效果。

下面结合附图和实施例对本申请的技术方案进一步详细阐述。

本申请实施例提供一种单目深度估计方法，该方法应用于计算设备，该方法所实现的功能可以通过服务器中的处理器调用程序代码来实现，当然程序代码可以保存在计算

机存储介质中，可见，该服务器至少包括处理器和存储介质。图 1A 为本申请实施例单目深度估计方法的实现流程示意图一，如图 1A 所示，该方法包括：

步骤 S101、获取待处理图像；

这里，可以由移动终端来获取待处理图像，所述待处理图像，可以包含任意场景的图片。一般来说，移动终端在实施的过程中可以为各种类型的具有信息处理能力的设备，例如所述移动终端可以包括手机、个人数字助理（Personal Digital Assistant, PDA）、导航仪、数字电话、视频电话、智能手表、智能手环、可穿戴设备、平板电脑等。服务器在实现的过程中可以是移动终端如手机、平板电脑、笔记本电脑，固定终端如个人计算机和服务器集群等具有信息处理能力的计算设备。

步骤 S102、将所述待处理图像输入至经过训练得到的单目深度估计网络模型，得到所述待处理图像的分析结果，其中，所述单目深度估计网络模型是通过第一双目匹配神经网络模型输出的视差图进行监督训练的；

本申请实施例中，所述单目深度估计网络模型主要是通过以下三个步骤获取的：第一步是使用渲染引擎渲染的合成双目数据预训练一个双目匹配神经网络；第二步是使用真实场景的数据对第一步得到的双目匹配神经网络进行微调训练；第三步是使用第二步得到的双目匹配神经网络对单目深度估计网络提供监督，从而训练得到单目深度估计网络。现有技术中，单目深度估计一般使用大量的真实标记数据进行训练，或者使用无监督的方法训练单目深度估计网络。但是，大量的真实标记数据获取成本很高，直接用无监督的方法训练单目深度估计网络又无法处理遮挡区域的深度估计，得到的效果较差。而本申请中所述单目深度估计网络模型的样本数据来自第一双目匹配神经网络模型输出的视差图，也就是说，本申请利用了双目视差来指导单目深度的预测。因此，本申请中的方法无需大量的标记数据，并且可以得到较好的训练效果。

步骤 S103、输出所述待处理图像的分析结果。这里，所述待处理图像的分析结果，指的是所述待处理图像对应的深度图。获取待处理图像后，将所述待处理图像输入至经过训练得到的单目深度估计网络模型，所述单目深度估计网络模型一般输出的是所述待处理图像对应的视差图，而不是深度图；因此，还需要根据所述单目深度估计网络模型输出的视差图、拍摄待处理图像的摄像机的镜头基线距离和拍摄待处理图像的摄像机的镜头焦距，确定所述待处理图像对应的深度图。

图 1B 为本申请实施例单个图片深度估计示意图，如图 1B 所示，标号为 11 的图片 11 为待处理图像，标号为 12 的图片 12 为标号为 11 的图片 11 对应的深度图。

在实际应用中，可以将所述镜头基线距离和所述镜头焦距的乘积，与所述输出的待处理图像对应的视差图的比值，确定为所述待处理图像对应的深度图。

基于上述方法实施例，本申请实施例再提供一种单目深度估计方法，该方法包括：

步骤 S111、获取有深度标记的合成的双目图片作为合成样本数据，其中，所述合成的双目图片包括合成的左图和合成的右图；

在一些实施例中，所述方法还包括：步骤 S11、通过渲染引擎构造虚拟 3D 场景；

步骤 S12、通过两个虚拟的摄像机将所述 3D 场景映射成双目图片；步骤 S13、根据构造所述虚拟 3D 场景时的位置、构造所述虚拟 3D 场景时的方向和所述虚拟的摄像机的镜头焦距获取所述合成双目图片的深度数据；步骤 S14、根据所述深度数据标记所述双目图片，得到所述合成的双目图片。

步骤 S112、根据获取的合成样本数据训练第二双目匹配神经网络模型；

这里，在实际应用中，所述步骤 S112 可以通过以下步骤实现：步骤 S1121、根据所述合成的双目图片对第二双目匹配神经网络模型进行训练，得到训练后的第二双目匹配神经网络模型，其中，所述训练后的第二双目匹配神经网络模型的输出为视差图和遮挡图，所述视差图描述了所述左图中每个像素点与所述右图中对应的像素点的视差距离，所述视差距离以像素为单位；所述遮挡图描述了所述左图中每个像素点在所述右图中对应的像素点是否被物体遮挡。

图 1C 为本申请实施例训练第二双目匹配神经网络模型示意图，如图 1C 所示，标号为 11 的图片 11 为合成的双目图片的左图，标号为 12 的图片 12 为合成的双目图片的右图， I^L 为标号为 11 的左图图片 11 中包含的所有像素点的像素值， I^R 为标号为 12 的右图图片 12 中包含的所有像素点的像素值；标号为 13 的图片 13 为第二双目匹配神经网络模型经过训练后输出的遮挡图，标号为 14 的图片 14 为第二双目匹配神经网络模型经过训练后输出的视差图，标号为 15 的图片 15 为第二双目匹配神经网络模型。

步骤 S113、根据获取的真实样本数据对训练后的第二双目匹配神经网络模型的参数进行调整，得到第一双目匹配神经网络模型；

这里，所述步骤 S113 可以通过两种方式实现，其中，第一种实现方式按照以下步骤实现：步骤 S1131a、根据获取的带深度标记的真实双目数据对训练后的第二双目匹配神经网络模型进行监督训练，以调整所述训练后的第二双目匹配神经网络模型的权值，得到第一双目匹配神经网络模型。这里，获取的是带有深度标记的真实双目数据，这样，就可以直接用带有深度标记的真实双目数据，对步骤 S112 中训练后的第二双目匹配神经网络进行监督训练，以调整所述训练后的第二双目匹配神经网络模型的权值，进一步提高训练后的第二双目匹配神经网络模型的效果，得到第一双目匹配神经网络模型。在这一部分中，双目视差网络需要对真实数据进行适配。可以使用真实的带有深度标记的双目数据，通过有监督的训练对双目视差网络直接进行微调训练调整网络权值。第二种实现方式按照以下步骤实现：步骤 S1131b、根据获取的不带深度标记的真实双目数据对训练后的第二双目匹配神经网络模型进行无监督训练，以调整所述训练后的第二双目匹配神经网络模型的权值，得到第一双目匹配神经网络模型。本申请实施例中，还可以使用不带深度标记的真实双目数据对训练后的第二双目匹配神经网络模型进行无监督训练，以调整所述训练后的第二双目匹配神经网络模型的权值，得到第一双目匹配神经网络模型。这里无监督训练指的是在没有深度数据标记的情况下，仅仅使用双目数据进行训练，可以使用无监督微调方法对此过程进行实现。

步骤 S114、通过所述第一双目匹配神经网络模型输出的视差图对单目深度估计网络

模型进行监督,从而训练所述单目深度估计网络模型;

这里,所述步骤 S114 以通过两种方式实现,其中,第一种实现方式按照以下步骤实现:步骤 S1141a、获取所述带深度标记的真实双目数据中的左图或右图作为训练样本,其中,所述带深度标记的真实双目数据包括左图和右图;步骤 S1142a、根据所述带深度标记的真实双目数据中的左图或右图对单目深度估计网络模型进行训练。这里,利用深度神经网络来预测单张图片的深度图,仅需要一张图片即可以对图片对应的场景进行三维建模,得到每个像素点的深度。因此,可以根据所述带深度标记的真实双目数据中的左图或右图对单目深度估计网络模型进行训练,其中,所述带深度标记的真实双目数据为步骤 S1131a 中使用的带深度标记的真实双目数据。第二种实现方式按照以下步骤实现:步骤 S1141b、所述不带深度标记的真实双目数据输入到所述第一双目匹配神经网络模型,得到对应的视差图,其中,所述不带深度标记的真实双目数据包括左图和右图;步骤 S1142b、根据所述对应的视差图、拍摄所述不带深度标记的真实双目数据的摄像机的镜头基线距离和拍摄所述不带深度标记的真实双目数据的摄像机的镜头焦距,确定所述视差图对应的深度图;步骤 S1143b、所述不带深度标记的真实双目数据中的左图或右图作为样本数据,根据所述视差图对应的深度图对单目深度估计网络模型进行监督,从而训练所述单目深度估计网络模型。这里,利用深度神经网络来预测单张图片的深度图,仅需要一张图片即可以对图片对应的场景进行三维建模,得到每个像素点的深度。因此,可以根据步骤 S1131b 中使用的不带深度标记的真实双目数据中的左图或右图作为样本数据,也是步骤 S1141b 中使用的不带深度标记的真实双目数据中的左图或右图作为样本数据,根据步骤 S1141b 中输出的视差图对应的深度图对单目深度估计网络模型进行监督,从而训练所述单目深度估计网络模型,得到训练后的单目深度估计网络模型。

图 1D 为本申请实施例训练单目深度估计网络模型示意图,如图 1D 所示,图(a)表示了将不带深度标记的真实双目数据输入到所述第一双目匹配神经网络模型,得到对应的标号为 13 的视差图图片 13,其中,所述不带深度标记的真实双目数据包括标号为 11 的左图图片 11 和标号为 12 的右图图片 12,标号为 15 的图片 15 为第一双目匹配神经网络模型。图 1D 中的图(b)表示了将所述不带深度标记的真实双目数据中的左图或右图作为样本数据,根据所述标号为 13 的视差图图片 13 对应的深度图对单目深度估计网络模型进行监督,从而训练所述单目深度估计网络模型,其中所述样本数据经过所述单目深度估计网络模型的输出为标号为 14 的视差图图片 14,标号为 16 的图片 16 为单目深度估计网络模型。

步骤 S115、获取待处理图像;

这里,在得到训练后的单目深度估计网络模型后,就可以使用此单目深度估计网络模型。即利用此单目深度估计网络模型,获取待处理图像对应的深度图。

步骤 S116、将所述待处理图像输入至经过训练得到的单目深度估计网络模型,得到所述待处理图像的分析结果,其中,所述单目深度估计网络模型是通过第一双目匹配神经网络模型输出的视差图进行监督训练的;

步骤 S117、输出所述待处理图像的分析结果，其中，所述待处理图像的分析结果包括所述单目深度估计网络模型输出的视差图；

步骤 S118、根据所述单目深度估计网络模型输出的视差图、拍摄输入所述单目深度估计网络模型的图片的摄像机的镜头基线距离和拍摄输入所述单目深度估计网络模型的图片的摄像机的镜头焦距，确定所述视差图对应的深度图；

步骤 S119、输出所述视差图对应的深度图。

基于上述方法实施例，本申请实施例再提供一种单目深度估计方法，该方法包括：

步骤 S121、获取有深度标记的合成的双目图片作为合成样本数据，其中，所述合成的双目图片包括合成的左图和合成的右图。

步骤 S122、根据获取的合成样本数据训练第二双目匹配神经网络模型；

这里，使用合成数据用于训练第二双目匹配神经网络模型具有更好的泛化能力。

步骤 S123、利用公式（1）确定所述损失函数： $L_{\text{stereo-unsup fit}} = L_{\text{photo}} + \gamma_1 L_{\text{abs}} + \gamma_2 L_{\text{rel}}$ （1）；其中，所述 $L_{\text{stereo-unsup fit}}$ 表示本申请实施例提出的损失函数；所述 L_{photo} 表示重建误差，所述 L_{abs} 表示所述第一双目匹配网络模型输出的视差图与所述训练后的第二双目匹配网络模型输出的视差图相比偏离较小；所述 L_{rel} 表示约束所述第一双目匹配网络模型的输出梯度与所述训练后的第二双目匹配网络模型的输出梯度一致；所述 γ_1 和 γ_2 表示强度系数。这里， L_{abs} 和 L_{rel} 为正则项。

在一些实施例中，步骤 S123 中的公式（1）还可以通过以下步骤中的公式进行细化，即所述方法还包括：步骤 S1231、利用公式（2）或公式（3）确定所述重建误差：

$$L_{\text{photo}} = \frac{1}{N} \sum_{i,j} \bar{M}_{\text{old}(ij)} |I_{ij}^L - I_{ij}^{wR}| = \frac{1}{N} \sum_{i,j} \bar{M}_{\text{old}(ij)} |I_{ij}^L - I_{i(j-\bar{D}_{ij}^R)}^R| \quad (2); \quad L_{\text{photo}} = \frac{1}{N} \sum_{i,j} \bar{M}_{\text{old}(ij)} |I_{ij}^R - I_{ij}^{wL}| = \frac{1}{N} \sum_{i,j} \bar{M}_{\text{old}(ij)} |I_{ij}^R - I_{i(j+\bar{D}_{ij}^L)}^L| \quad (3);$$

其中，所述 N 表示图片中像素的个数；所述 $\bar{M}_{\text{old}(ij)}$ 表示所述训练后的第二双目匹配网络模型输出的遮挡图的像素值；所述 I_{ij}^L 表示不带深度标记的真实双目数据中的左图的像素值；所述 I_{ij}^R 表示不带深度标记的真实双目数据中的右图的像素值；所述 I_{ij}^{wR} 表示将右图采样后合成的图片的像素值，即重建的左图；所述 I_{ij}^{wL} 表示将左图采样后合成的图片的像素值，即重建的右图；所述 $\bar{D}_{ij}^L$ 表示不带深度标记的真实双目数据中的左图经第一双目匹配网络模型输出的视差图的像素值；所述 $\bar{D}_{ij}^R$ 表示不带深度标记的真实双目数据中的右图经第一双目匹配网络模型输出的视差图的像素值；所述 ij 表示像素点的像素坐标；所述 old 表示训练后的第二双目匹配网络模型的输出；所述 R 表示右图或右图的相关数据，所述 L 表示左图或左图的相关数据；所述 I 表示图片像素点的 RGB（Red Green Blue，红色、绿色和蓝色）值。步骤 S1232、利用公式（4）或公式（5）确定所述第一双目匹配网络模型输出的视差图与所述训练后的第二双目匹配网络模型输出的视差图相比偏离较小：

$$L_{\text{abs}} = \frac{1}{N} \sum_{i,j} (1 - \bar{M}_{\text{old}(ij)} + \gamma_3) |\bar{D}_{ij}^L - \bar{D}_{\text{old}(ij)}^L| \quad (4); \quad L_{\text{abs}} = \frac{1}{N} \sum_{i,j} (1 - \bar{M}_{\text{old}(ij)} + \gamma_3) |\bar{D}_{ij}^R - \bar{D}_{\text{old}(ij)}^R| \quad (5);$$

其中，所述 N 表示图片中像素的个数，所述 $\bar{M}_{\text{old}(ij)}$ 表示所述训练后的第二双目匹配网络模型输出的遮挡图的像素值，所述 $\bar{D}_{\text{old}(ij)}^L$ 表示样本数据中的左图经训练后的第二双目匹配网络输出的视差图的像素值，所述 $\bar{D}_{\text{old}(ij)}^R$ 表示样本数据中的右图经训练后的第二双目匹配网络输出的视差图

的像素值, 所述 $\bar{D}_{ij}^L$ 表示不带深度标记的真实双目数据中的左图经第一双目匹配网络输出的视差图的像素值, 所述 $\bar{D}_{ij}^R$ 表示不带深度标记的真实双目数据中的右图经第一双目匹配网络输出的视差图的像素值, 所述 ij 表示像素点的像素坐标, 所述 old 表示训练后的第二双目匹配网络模型的输出, 所述 R 表示右图或右图的相关数据, 所述 L 表示左图或左图的相关数据, 所述 γ_3 表示强度系数。步骤 S1233、利用公式 (6) 或公式 (7) 确定所述第一双目匹配网络模型的输出梯度与所述第二双目匹配网络模型的输出梯度一致:

$$L_{rel} = \frac{1}{N} \sum_{i,j} |\nabla \bar{D}_{ij}^L - \nabla \bar{D}_{old(ij)}^L| \quad (6); \quad L_{rel} = \frac{1}{N} \sum_{i,j} |\nabla \bar{D}_{ij}^R - \nabla \bar{D}_{old(ij)}^R| \quad (7);$$

其中, 所述 N 表示图片中像素的个数, 所述 $\nabla \bar{D}_{ij}^L$ 表示不带深度标记的真实双目数据中的左图经第一双目匹配网络输出的视差图的梯度, 所述 $\nabla \bar{D}_{ij}^R$ 表示不带深度标记的真实双目数据中的右图经第一双目匹配网络输出的视差图的梯度, 所述 $\nabla \bar{D}_{old(ij)}^L$ 表示样本数据中的左图经训练后的第二双目匹配网络输出的视差图的梯度, 所述 $\nabla \bar{D}_{old(ij)}^R$ 表示样本数据中的右图经训练后的第二双目匹配网络输出的视差图的梯度, 所述 old 表示训练后的第二双目匹配网络模型的输出, 所述 R 表示右图或右图的相关数据, 所述 L 表示左图或左图的相关数据。

步骤 S124、使用损失函数 (Loss), 根据所述不带深度标记的真实双目数据对训练后的第二双目匹配神经网络模型进行无监督训练, 以调整所述训练后的第二双目匹配神经网络模型的权值, 得到第一双目匹配神经网络模型。

这里, 所述损失函数 (Loss) 利用了步骤 S122 中训练后的第二双目匹配神经网络的输出对微调训练进行正则化, 避免了现有技术中的无监督微调普遍存在的预测变模糊的问题, 提高了微调得到的第一双目匹配网络的效果, 从而间接提高了第一双目匹配网络监督得到的单目深度网络的效果。图 1E 为本申请实施例损失函数相关图片示意图, 如图 1E 所示, 图 (a) 为不带深度标记的真实双目数据的左图; 图 1E 中的图 (b) 为不带深度标记的真实双目数据的右图; 图 1E 中的图 (c) 为将图 (a) 和图 (b) 组成的不带深度标记的真实双目图片输入至经过训练后的第二双目匹配神经网络模型输出的视差图; 图 1E 中的图 (d) 为将图 (b) 表示的右图进行采样后, 结合图 (c) 表示的视差图, 对左图进行重建后的图片; 图 1E 中的图 (e) 为将图 (a) 表示的左图中的像素与图 (d) 表示的重建后的左图中的对应像素做差得到的图片, 即左图的重建误差图; 图 1E 中的图 (f) 为将图 (a) 和图 (b) 组成的不带深度标记的真实双目图片输入至经过训练后的第二双目匹配神经网络模型输出的遮挡图。其中, 图 (d) 中所有的红框 11 表示所述重建后的左图与图 (a) 标识的真实左图有差异的部分, 图 (e) 中所有的红框 12 表示所述重建误差图中有误差的部分, 即被遮挡的部分。这里, 实现步骤 S124 中描述的用无监督微调训练双目视差网络时, 需要使用右图对左图进行重建, 但是有遮挡区域是无法重建正确的, 因此, 用遮挡图来清理这一部分的错误训练信号来提高无监督微调训练的效果。

步骤 S125、通过所述第一双目匹配神经网络模型输出的视差图对所述单目深度估计网络模型进行监督, 从而训练所述单目深度估计网络模型。

这里，所述单目深度估计网络模型的样本图片，可以是不带深度标记的真实双目数据中的左图，也可以是不带深度标记的真实双目数据中的右图。其中，如果使用左图作为样本图片，则通过公式（1）、公式（2）、公式（4）和公式（6）来确定损失函数；如果使用右图作为样本图片，则通过公式（1）、公式（3）、公式（5）和公式（7）来确定损失函数。

本申请实施例中，所述通过所述第一双目匹配神经网络模型输出的视差图对所述单目深度估计网络模型进行监督，从而训练所述单目深度估计网络模型，指的是通过所述第一双目匹配神经网络模型输出的视差图对应的深度图对所述单目深度估计网络模型进行监督，也即使提供监督信息，从而训练所述单目深度估计网络模型。

步骤 S126、获取待处理图像；

步骤 S127、将所述待处理图像输入至经过训练得到的单目深度估计网络模型，得到所述待处理图像的分析结果，其中，所述单目深度估计网络模型是通过第一双目匹配神经网络模型输出的视差图进行监督训练的；

步骤 S128、输出所述待处理图像的分析结果，其中，所述待处理图像的分析结果包括所述单目深度估计网络模型输出的视差图。

步骤 S129、根据所述单目深度估计网络模型输出的视差图、拍摄输入所述单目深度估计网络模型的图片的摄像机的镜头基线距离和拍摄输入所述单目深度估计网络模型的图片的摄像机的镜头焦距，确定所述视差图对应的深度图；

步骤 S130、输出所述视差图对应的深度图。

本申请实施例中，当所述待处理图像为街景图片时，就可以使用所述训练后的单目深度估计网络模型预测所述街景图片的深度。

基于上述的方法实施例，本申请实施例再提供一种单目深度估计方法，图 2A 为本申请实施例单目深度估计方法的实现流程示意图二，如图 2A 所示，该方法包括：

步骤 S201、使用渲染引擎渲染的合成数据训练双目匹配网络，得到双目图片的视差图；

这里，所述双目匹配网络的输入为：一对双目图片（包含左图和右图），所述双目匹配网络的输出为：视差图、遮挡图，即双目匹配网络使用双目图片作为输入，输出视差图和遮挡图。其中，视差图用于描述左图中每个像素点与右图中对应的像素点的视差距离，以像素为单位；遮挡图用于描述左图每个像素在右图中对应的像素点是否被其他物体遮挡。由于视角的变化，左图中的一些区域在右图中会被其他物体遮挡，遮挡图则是用于标记左图中的像素是否在右图中被遮挡。这一部分，双目匹配网络使用计算机渲染引擎产生的合成数据进行训练，首先通过渲染引擎构造一些虚拟 3D 场景，然后通过两个虚拟的摄像机将 3D 场景映射成双目图片，从而获得合成数据，同时正确的深度数据和相机焦距等数据也可以从渲染引擎中得到，所以双目匹配网络可以直接通过这些标记数据进行监督训练。

步骤 S202、利用损失函数，通过无监督微调方法在真实双目图片数据上对步骤 S201

得到的双目匹配网络进行微调;

在这一部分中,双目视差网络需要对真实数据进行适配。即使用不带深度标记的真实双目数据对双目视差网络进行无监督训练。这里无监督训练指的是在没有深度数据标记的情况下,仅仅使用双目数据进行训练。本申请实施例提出了一种新的无监督微调方法,即使用上述实施例中的损失函数进行无监督微调。本申请实施例提出的损失函数的主要目的是希望在不降低预训练效果的情况下在真实双目数据上对双目视差网络进行微调,微调过程中借助了步骤 S201 得到的预训练双目视差网络的初步输出进行指导和正则化。图 2B 为本申请实施例损失函数效果示意图,如图 2B 所示,标号为 21 的图片 21 为使用现有技术中的损失函数时得到的视差图,标号为 22 的图片 22 为使用本申请实施例提出的损失函数时得到的视差图。现有技术的损失函数没有单独考虑遮挡区域,会将遮挡区域的图像重建误差也优化为零,这样会导致遮挡区域的预测视差错误,视差图的边缘也会模糊,而本申请中的损失函数用遮挡图来清理这一部分的错误训练信号来提高无监督微调训练的效果。

步骤 S203、使用步骤 S202 得到的双目匹配网络在真实数据上对单目深度估计进行监督,最终得到单目深度估计网络。这里,所述单目深度估计网络的输入为:单张单目图片,所述单目深度估计网络的输出为:深度图。在步骤 S202 中得到了在真实数据上微调过的双目视差网络,对于每一对双目图片,双目视差网络预测得到视差图,通过视差图 D 、双目镜头基线距离 b 以及镜头焦距 f ,可以计算得到视差图对应的深度图,即通过公式 (8),可以计算得到视差图对应的深度图 d : $d = bf / D$ (8);为了训练单目深度网络预测得到深度图,可以使用双目图片对中的左图作为单目深度网络的输入,然后使用双目视差网络输出计算得到的深度图进行监督,从而训练单目深度网络,得到最终结果。在实际应用中,可以本申请实施例中的单目深度估计方法训练得到用于无人驾驶的深度估计模块,从而对场景进行三维重建或者障碍物检测。且本申请实施例提出的无监督微调方法提高了双目视差网络的性能。

现有技术中,有监督的单目深度估计方法,获取准确的标记数据是非常有限也是非常难的。基于重建误差的无监督方法的性能通常受到像素匹配歧义的限制。为了解决这些问题,本申请实施例提出了一种新的单目深度估计方法,解决了现有技术中监督和无监督深度估计方法存在的局限性。本申请实施例中的方法是使用一个双目匹配网络在跨模态合成数据上训练,并用来监督单目深度估计网络。所述双目匹配网络是基于左右图的像素匹配关系来获得视差,而不是从语义特征中提取,因此,双目匹配网络可以很好地从合成数据泛化到真实数据。本申请实施例的方法主要包括三个步骤。第一,用合成数据对双目匹配网络进行训练,从双目图片中预测遮挡图和视差图。第二,根据可用的真实数据,在有监督或者无监督的情况下,对训练后的双目匹配网络有选择性地进行调整。第三,在第二步得到的用真实数据微调训练后的双目匹配网络的监督下,训练单目深度估计网络。这样可以间接利用双目匹配网络来使单目深度估计更好地利用合成数据来提高性能。

第一步、利用合成数据对双目匹配网络进行训练,包括:目前由图形渲染引擎可以生成很多的包含深度信息的合成图像。但是,直接将这些合成图像数据与真实数据合并来训练单目深度估计网络得到的性能通常较差,因为单目深度估计对输入场景的语义信息非常敏感。合成数据和真实数据之间的巨大模态差距使得使用合成数据辅助训练变得毫无用处。然而,双目匹配网络有更好的泛化能力,使用合成数据训练的双目匹配网络在真实数据上也能得到较好的视差图输出。因此,本申请实施例将双目匹配网络训练作为在合成数据和真实数据之间的桥梁来提高单目深度训练的性能。首先利用大量的合成双目数据对双目匹配网络进行预训练。与传统的结构不同,实施例中的双目匹配网络在视差图的基础上,还估计了多尺度遮挡图。其中,遮挡图表示在正确的图像中,左侧图像像素的在右图中的对应像素点是否被其他物体遮挡。在接下来的步骤中,无监督的微调方法会使用到所述遮挡图,以避免错误的估计。其中,可以使用左右视差一致性检验方法,利用公式(9)从正确标记的视差图中得到有正确标记的遮挡图 M_{ij}^* :

$M_{ij}^* = 1(|D_{ij}^{*L} - D_{ij}^{*wR}| \leq 1) = 1(|D_{ij}^{*L} - D_{i(j-D_{ij}^{*R})}^{*R}| \leq 1)$ (9); 其中,下标 i 表示图像中第 i 行的值,下标 j 表示图像中第 j 列的值。 $D^{*L/R}$ 表示左右幅图像的视差图, D^{*wR} 是用右图重建后的左图的视差图,对于非遮挡区域,左视差图和利用右图重建后的左图的视差图是一致的。一致性检查的阈值设置为 1。遮挡图在遮挡区域为 0,非遮挡区域为 1。因此,本实施例使用以下公式(10)计算使用合成数据训练双目匹配网络的损失(Loss),在此阶段,损失函数 L_{stereo} 由两部分组成,即视差图估计误差 L_{disp} 和遮挡图估计误差 L_{occ} 。双目视差网络的多尺度中间层也产生了视差和遮挡预测,并直接应用于多尺度预测的损失权重 w_m , L_{disp}^m 表示每一层对应的视差图估计误差, L_{occ}^m 表示每一层对应的遮挡图估计误差, m 表示第 m 层:

$L_{stereo} = L_{disp} + L_{occ} = \sum w_m (L_{disp}^m + L_{occ}^m)$ (10); 为了训练视差图,采用 L1 损失函数来避免异常值的影响,使训练过程更加鲁棒。为了训练遮挡图,利用公式(11)来表示遮挡图估计误差 L_{occ} ,采用二元交叉熵损失作为一种分类任务来训练遮挡图:

$L_{occ} = -\frac{1}{N} \sum_{ij} M_{ij}^* \log(\bar{M}_{ij}) + (1 - M_{ij}^*) \log(1 - \bar{M}_{ij})$ (11); 其中, N 是图像中像素的总数, M_{ij}^* 表示有正确标记的遮挡图, $\bar{M}_{ij}$ 表示经训练后的双目匹配网络输出的遮挡图。

第二步、使用有监督或无监督的微调方法在真实数据上训练第一步得到的训练后的双目匹配网络,包括:本申请实施例以两种方式对训练后的双目匹配网络进行微调。其中,有监督的微调方法,仅采用多尺度的 L1 回归损失函数 $L_{stereo-sup ft}$,即视差图估计误差 L_{disp} 来改进之前像素匹配预测的误差,见公式(12): $L_{stereo-sup ft} = L_{disp} = \sum_m L_{disp}^m$ (12); 结果表明,使用少量的监督数据,例如 100 幅图片,双目匹配网络也能够从合成模态数据适应到真实模态数据。无监督的微调方法。对于无监督的网络调优,现有技术中的无监督微调方法得到的视差图模糊,性能较差,如图 2B 中的图片 21 所示。这是由于无监督损失的局限性以及只有 RGB 值的像素匹配的歧义性导致的。因此,本申请实施例引入额外的正则项约束来提高性能。利用真实数据,从没有进行微调的训练后的双目匹配网络中得到

了相应的遮挡图和视差图，并且，将其分别标记为 $\overline{M}_{old}$ 和 $\overline{D}_{old}$ 。这两个数据用于帮助规范训练过程。进一步的，本申请实施例提出的无监督的微调损失函数，即损失函数 $L_{stereo-unsup fi}$ 的获取可以参见前面实施例中的描述。

第三步、训练单目深度估计网络，包括：到目前为止，我们已经通过大量的合成数据对双目匹配网络进行了跨模态训练，并使用真实数据进行了微调。为了训练最终的单目深度估计网络，本申请实施例采用训练后的双目匹配网络预测的视差图提供训练数据。

单目深度估计的损失 L_{mono} 由以下几个部分给出，参见公式 (13): $L_{mono} = \frac{1}{N} \sum_{i,j} |\bar{D}_i^L - \bar{D}_j^L|$ (13);

这里， N 为像素点的总和， $\bar{D}_i^L$ 指的是单目深度估计网络输出的视差图， $\bar{D}_j^L$ 指的是训练后的双目匹配网络输出的视差图，或者，将训练后的双目匹配网络进行微调，微调后的网络输出的视差图。需要指出的是，公式 (9) 至公式 (13) 都是以单目深度估计网络使用真实数据中的左图作为训练样本为例，进行说明的。实验：由于单目深度估计网络对视角变化敏感，所以不对训练数据使用裁剪和缩放。所述单目深度估计网络的输入和用于监督单目深度估计网络的视差图都是来自训练后的双目匹配网络。图 2C 为本申请实施例可视化深度估计结果示意图，如图 2C 所示，图 2C 中展示了使用现有技术和本申请实施例中的单目深度估计方法，获取的三幅不同的街景图片对应的深度图，其中，第一行为单目深度估计网络的输入，即三幅不同的街景图片；第二行为使用最近邻算法对稀疏激光雷达深度图插值得到的深度数据，第三行至第五行为现有技术中的三种不同的单目深度估计方法分别得到的三幅输入图片对应的深度图；本申请的结果见最后三行，直接利用本申请实施例中第一步得到的使用合成数据训练得到的双目匹配网络，对单目深度估计网络进行监督，得到的单目深度网络的三幅输入图片对应的深度图，即标号为 21 的图片 21、标号为 22 的图片 22、标号为 23 的图片 23；利用本申请实施例提出的无监督损失函数，对训练后的双目匹配网络进行微调，将微调后的网络输出的视差图，作为单目深度估计网络的训练数据，得到的单目深度网络的三幅输入图片对应的深度图，即标号为 24 的图片 24、标号为 25 的图片 25、标号为 26 的图片 26；对训练后的双目匹配网络进行有监督的微调，将微调后的网络输出的视差图，作为单目深度估计网络的训练数据，得到的单目深度网络的三幅输入图片对应的深度图，即标号为 27 的图片 27、标号为 28 的图片 28、标号为 29 的图片 29；从标号为 21 的图片 21 至标号为 29 的图片 29 可以看出，本申请实施例中的单目深度估计方法获得的模型可以捕捉到更细节的场景结构。

本申请实施例提供一种单目深度估计装置，图 3 为本申请实施例单目深度估计装置的组成结构示意图，如图 3 所示，所述装置 300 包括：获取模块 301、执行模块 302 和输出模块 303，其中：

所述获取模块 301，配置为获取待处理图像；

所述执行模块 302，配置为将所述待处理图像输入至经过训练得到的单目深度估计网络模型，得到所述待处理图像的分析结果，其中，所述单目深度估计网络模型是通过

第一双目匹配神经网络模型输出的视差图进行监督训练的；

所述输出模块 303，配置为输出所述待处理图像的分析结果。

在一些实施例中，所述装置还包括：第三训练模块，配置为通过所述第一双目匹配神经网络模型输出的视差图对所述单目深度估计网络模型进行监督，从而训练所述单目深度估计网络模型。

在一些实施例中，所述装置还包括：第一训练模块，配置为根据获取的合成样本数据训练第二双目匹配神经网络模型；第二训练模块，配置为根据获取的真实样本数据对训练后的第二双目匹配神经网络模型的参数进行调整，得到第一双目匹配神经网络模型。

在一些实施例中，所述装置还包括：第一获取模块，配置为获取有深度标记的合成双目图片作为所述合成样本数据，其中，所述合成的双目图片包括合成的左图和合成的右图。

在一些实施例中，所述第一训练模块，包括：第一训练单元，配置为根据所述合成的双目图片对第二双目匹配神经网络模型进行训练，得到训练后的第二双目匹配神经网络模型，其中，所述训练后的第二双目匹配神经网络模型的输出为视差图和遮挡图，所述视差图描述了所述左图中每个像素点与所述右图中对应的像素点的视差距离，所述视差距离以像素为单位；所述遮挡图描述了所述左图中每个像素点在所述右图中对应的像素点是否被物体遮挡。

在一些实施例中，所述装置还包括：构造模块，配置为通过渲染引擎构造虚拟 3D 场景；映射模块，配置为通过两个虚拟的摄像机将所述 3D 场景映射成双目图片；第二获取模块，配置为根据构造所述虚拟 3D 场景时的位置、构造所述虚拟 3D 场景时的方向和所述虚拟的摄像机的镜头焦距获取所述合成双目图片的深度数据；第三获取模块，配置为根据所述深度数据标记所述双目图片，得到所述合成的双目图片。

在一些实施例中，所述第二训练模块，包括：第二训练单元，配置为根据获取的带深度标记的真实双目数据对训练后的第二双目匹配神经网络模型进行监督训练，以调整所述训练后的第二双目匹配神经网络模型的权值，得到第一双目匹配神经网络模型。

在一些实施例中，所述第二训练模块中的第二训练单元，还配置为：根据获取的不带深度标记的真实双目数据对训练后的第二双目匹配神经网络模型进行无监督训练，以调整所述训练后的第二双目匹配神经网络模型的权值，得到第一双目匹配神经网络模型。

在一些实施例中，所述第二训练模块中的第二训练单元，包括：第二训练部件，配置为使用损失函数，根据所述不带深度标记的真实双目数据对训练后的第二双目匹配神经网络模型进行无监督训练，以调整所述训练后的第二双目匹配神经网络模型的权值，得到第一双目匹配神经网络模型。

在一些实施例中，所述装置还包括：第一确定模块，配置为利用公式 (14) 确定所述损失函数； $L_{\text{stereo-unsup fit}} = L_{\text{photo}} + \gamma_1 L_{\text{abs}} + \gamma_2 L_{\text{rel}}$ (14)；其中，所述 $L_{\text{stereo-unsup fit}}$ 表示损失函数，所述 L_{photo} 表示重建误差，所述 L_{abs} 表示所述第一双目匹配网络模型输出的视差图与所述训练后的第二双目匹配网络模型输出的视差图相比偏离较小，所述 L_{rel} 表示约束所述第一双目

匹配网络模型的输出梯度与所述训练后的第二双目匹配网络模型的输出梯度一致, 所述 γ_1 和 γ_2 表示强度系数。

在一些实施例中, 所述装置还包括: 第二确定模块, 配置为利用公式 (15) 或公式 (16) 确定所述重建误差; $L_{photo} = \frac{1}{N} \sum_{i,j} \bar{M}_{old(ij)} |I_{ij}^L - I_{ij}^{wR}| = \frac{1}{N} \sum_{i,j} \bar{M}_{old(ij)} |I_{ij}^L - I_{i(j-\bar{D}_i^L)}^R|$ (15); $L_{photo} = \frac{1}{N} \sum_{i,j} \bar{M}_{old(ij)} |I_{ij}^R - I_{ij}^{wL}| = \frac{1}{N} \sum_{i,j} \bar{M}_{old(ij)} |I_{ij}^R - I_{i(j+\bar{D}_i^R)}^L|$ (16); 其中, 所述 N 表示图片中像素的个数, 所述 $\bar{M}_{old(ij)}$ 表示所述训练后的第二双目匹配网络模型输出的遮挡图的像素值, 所述 I_{ij}^L 表示不带深度标记的真实双目数据中的左图的像素值, 所述 I_{ij}^R 表示不带深度标记的真实双目数据中的右图的像素值, 所述 I_{ij}^{wR} 表示将右图采样后合成的图片的像素值, 所述 I_{ij}^{wL} 表示将左图采样后合成的图片的像素值, 所述 $\bar{D}_i^L$ 表示不带深度标记的真实双目数据中的左图经第一双目匹配网络模型输出的视差图的像素值, 所述 $\bar{D}_i^R$ 表示不带深度标记的真实双目数据中的右图经第一双目匹配网络模型输出的视差图的像素值, ij 表示像素点的像素坐标。

在一些实施例中, 所述装置还包括: 第三确定模块, 配置为利用公式 (17) 或公式 (18) 确定所述第一双目匹配网络模型输出的视差图与所述训练后的第二双目匹配网络模型输出的视差图相比偏离较小; $L_{abs} = \frac{1}{N} \sum_{i,j} (1 - \bar{M}_{old(ij)} + \gamma_3) |\bar{D}_{ij}^L - \bar{D}_{old(ij)}^L|$ (17); $L_{abs} = \frac{1}{N} \sum_{i,j} (1 - \bar{M}_{old(ij)} + \gamma_3) |\bar{D}_{ij}^R - \bar{D}_{old(ij)}^R|$ (18); 其中, 所述 $\bar{D}_{old(ij)}^L$ 表示样本数据中的左图经训练后的第二双目匹配网络模型输出的视差图的像素值, 所述 $\bar{D}_{old(ij)}^R$ 表示样本数据中的右图经训练后的第二双目匹配网络模型输出的视差图的像素值, 所述 γ_3 表示强度系数。

在一些实施例中, 所述装置还包括: 第四确定模块, 配置为利用公式 (19), 或公式 (20), 确定所述第一双目匹配网络模型的输出梯度与所述第二双目匹配网络模型的输出梯度一致; $L_{rel} = \frac{1}{N} \sum_{i,j} |\nabla \bar{D}_{ij}^L - \nabla \bar{D}_{old(ij)}^L|$ (19); $L_{rel} = \frac{1}{N} \sum_{i,j} |\nabla \bar{D}_{ij}^R - \nabla \bar{D}_{old(ij)}^R|$ (20); 其中, 所述 $\nabla \bar{D}_{ij}^L$ 表示不带深度标记的真实双目数据中的左图经第一双目匹配网络模型输出的视差图的梯度, 所述 $\nabla \bar{D}_{ij}^R$ 表示不带深度标记的真实双目数据中的右图经第一双目匹配网络模型输出的视差图的梯度, 所述 $\nabla \bar{D}_{old(ij)}^L$ 表示样本数据中的左图经训练后的第二双目匹配网络模型输出的视差图的梯度, 所述 $\nabla \bar{D}_{old(ij)}^R$ 表示样本数据中的右图经训练后的第二双目匹配网络模型输出的视差图的梯度。

在一些实施例中, 所述带深度标记的真实双目数据包括左图和右图, 对应地, 所述第三训练模块, 包括: 第一获取单元, 配置为获取所述带深度标记的真实双目数据中的左图或右图作为训练样本; 第一训练单元, 配置为根据所述带深度标记的真实双目数据中的左图或右图对单目深度估计网络模型进行训练。

在一些实施例中, 所述不带深度标记的真实双目数据包括左图和右图, 对应地, 所述第三训练模块, 还包括: 第二获取单元, 配置为将所述不带深度标记的真实双目数据输入到所述第一双目匹配神经网络模型, 得到对应的视差图; 第一确定单元, 配置为根据所述对应的视差图、拍摄所述不带深度标记的真实双目数据的摄像机的镜头基线距离

和拍摄所述不带深度标记的真实双目数据的摄像机的镜头焦距，确定所述视差图对应的深度图；第二训练单元，配置为将所述不带深度标记的真实双目数据中的左图或右图作为样本数据，根据所述视差图对应的深度图对单目深度估计网络模型进行监督，从而训练所述单目深度估计网络模型。

在一些实施例中，所述待处理图像的分析结果包括所述单目深度估计网络模型输出的视差图，对应地，所述装置还包括：第五确定模块，配置为根据所述单目深度估计网络模型输出的视差图、拍摄输入所述单目深度估计网络模型的图片的摄像机的镜头基线距离和拍摄输入所述单目深度估计网络模型的图片的摄像机的镜头焦距，确定所述视差图对应的深度图；第一输出模块，配置为输出所述视差图对应的深度图。

这里需要指出的是：以上装置实施例的描述，与上述方法实施例的描述是类似的，具有同方法实施例相似的有益效果。对于本申请装置实施例中未披露的技术细节，请参照本申请方法实施例的描述而理解。本申请实施例中，如果以软件功能模块的形式实现上述的单目深度估计方法，并作为独立的产品销售或使用，也可以存储在一个计算机可读存储介质中。基于这样的理解，本申请实施例的技术方案本质上或者说对现有技术做出贡献的部分可以以软件产品的形式体现出来，该计算机软件产品存储在一个存储介质中，包括若干指令用以使得一台计算设备执行本申请各个实施例所述方法的全部或部分。而前述的存储介质包括：U盘、移动硬盘、ROM（Read Only Memory，只读存储器）、磁碟或者光盘等各种可以存储程序代码的介质。这样，本申请实施例不限制于任何特定的硬件和软件结合。对应地，本申请实施例提供一种单目深度估计设备，该设备包括存储器和处理器，所述存储器存储有可在处理器上运行的计算机程序，所述处理器执行所述程序时实现单目深度估计方法中的步骤。对应地，本申请实施例提供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单目深度估计方法中的步骤。这里需要指出的是：以上存储介质和设备实施例的描述，与上述方法实施例的描述是类似的，具有同方法实施例相似的有益效果。对于本申请存储介质和设备实施例中未披露的技术细节，请参照本申请方法实施例的描述而理解。

需要说明的是，图4为本申请实施例单目深度估计设备的一种硬件实体示意图，如图4所示，该单目深度估计设备400的硬件实体包括：存储器401、通信总线402和处理器403，其中，存储器401配置为存储由处理器403可执行的指令和应用，还可以缓存待处理器403以及单目深度估计设备400中各模块待处理或已经处理的数据，可以通过FLASH（闪存）或RAM（Random Access Memory，随机访问存储器）实现。通信总线402可以使单目深度估计设备400通过网络与其他终端或服务器通信，还可以实现处理器403和存储器401之间的连接通信。处理器403通常控制单目深度估计设备400的总体操作。

需要说明的是，在本文中，术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含，从而使得包括一系列要素的过程、方法、物品或者装置不仅包括那些要素，而且还包括没有明确列出的其他要素，或者是还包括为这种过程、方法、物品或者

装置所固有的要素。在没有更多限制的情况下，由语句“包括一个……”限定的要素，并不排除在包括该要素的过程、方法、物品或者装置中还存在另外的相同要素。

通过以上的实施方式的描述，本领域的技术人员可以清楚地了解到上述实施例方法可借助软件加必需的通用硬件平台的方式来实现，当然也可以通过硬件，但很多情况下前者是更佳的实施方式。基于这样的理解，本申请的技术方案本质上或者说对现有技术做出贡献的部分可以以软件产品的形式体现出来，该计算机软件产品存储在一个存储介质（如 ROM/RAM、磁碟、光盘）中，包括若干指令用以使得一台终端设备（可以是手机，计算机，服务器，空调器，或者网络设备等）执行本申请各个实施例所描述的方法。

本申请是参照根据本申请实施例的方法、设备（装置）、和计算机程序产品的流程图和/或方框图来描述的。应理解可由计算机程序指令实现流程图和/或方框图中的每一流程和/或方框、以及流程图和/或方框图中的流程和/或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理设备的处理器以产生一个机器，使得通过计算机或其他可编程数据处理设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的装置。这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理设备以特定方式工作的计算机可读存储器中，使得存储在该计算机可读存储器中的指令产生包括指令装置的制造品，该指令装置实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能。

这些计算机程序指令也可装载到计算机或其他可编程数据处理设备上，使得在计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理，从而在计算机或其他可编程设备上执行的指令提供用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的步骤。

以上仅为本申请的优选实施例，并非因此限制本申请的专利范围，凡是利用本申请说明书及附图内容所作的等效结构或等效流程变换，或直接或间接运用在其他相关的技术领域，均同理包括在本申请的专利保护范围内。

权利要求书

1、一种单目深度估计方法，其中，所述方法包括：获取待处理图像；将所述待处理图像输入至经过训练得到的单目深度估计网络模型，得到所述待处理图像的分析结果，其中，所述单目深度估计网络模型是通过第一双目匹配神经网络模型输出的视差图进行监督训练的；输出所述待处理图像的分析结果。

2、根据权利要求 1 所述的方法，其中，所述第一双目匹配神经网络模型的训练过程，包括：根据获取的合成样本数据训练第二双目匹配神经网络模型；根据获取的真实样本数据对训练后的第二双目匹配神经网络模型的参数进行调整，得到第一双目匹配神经网络模型。

3、根据权利要求 2 所述的方法，其中，所述方法还包括：获取有深度标记的合成的双目图片作为所述合成样本数据，其中，所述合成的双目图片包括合成的左图和合成的右图。

4、根据权利要求 3 所述的方法，其中，所述根据获取的合成样本数据训练第二双目匹配神经网络模型，包括：根据所述合成的双目图片对第二双目匹配神经网络模型进行训练，得到训练后的第二双目匹配神经网络模型，其中，所述训练后的第二双目匹配神经网络模型的输出为视差图和遮挡图，所述视差图描述了所述左图中每个像素点与所述右图中对应的像素点的视差距离，所述视差距离以像素为单位；所述遮挡图描述了所述左图中每个像素点在所述右图中对应的像素点是否被物体遮挡。

5、根据权利要求 2 所述的方法，其中，所述根据获取的真实样本数据对训练后的第二双目匹配神经网络模型的参数进行调整，得到第一双目匹配神经网络模型，包括：根据获取的带深度标记的真实双目数据对训练后的第二双目匹配神经网络模型进行监督训练，以调整所述训练后的第二双目匹配神经网络模型的权值，得到第一双目匹配神经网络模型。

6、根据权利要求 2 所述的方法，其中，所述根据获取的真实样本数据对训练后的第二双目匹配神经网络模型的参数进行调整，得到第一双目匹配神经网络模型，还包括：根据获取的不带深度标记的真实双目数据对训练后的第二双目匹配神经网络模型进行无监督训练，以调整所述训练后的第二双目匹配神经网络模型的权值，得到第一双目匹配神经网络模型。

7、根据权利要求 6 所述的方法，其中，所述根据获取的不带深度标记的真实双目数据对训练后的第二双目匹配神经网络模型进行无监督训练，以调整所述训练后的第二双目匹配神经网络模型的权值，得到第一双目匹配神经网络模型，包括：使用损失函数，根据所述不带深度标记的真实双目数据对训练后的第二双目匹配神经网络模型进行无监督训练，以调整所述训练后的第二双目匹配神经网络模型的权值，得到第一双目匹配神经网络模型。

8、根据权利要求 7 所述的方法，其中，所述方法还包括：利用公式 $L_{\text{stereo-unsup fit}} = L_{\text{photo}} + \gamma_1 L_{\text{abs}} + \gamma_2 L_{\text{rel}}$ 确定所述损失函数，其中，所述 $L_{\text{stereo-unsup fit}}$ 表示损失函数，所述 L_{photo} 表示重建误差，所述 L_{abs} 表示所述第一双目匹配网络模型输出的视差图与所述训练后的第二双目匹配网络模型输出的视差图相比偏离较小，所述 L_{rel} 表示约束所述第一双目匹配网络模型的输出梯度与所述训练后的第二双目匹配网络模型的输出梯度一致，所述 γ_1 和 γ_2 表示强度系数。

9、根据权利要求 8 所述的方法，其中，所述方法还包括：利用公式 $L_{\text{photo}} = \frac{1}{N} \sum_{i,j} \bar{M}_{\text{old}(ij)} |I_{ij}^L - I_{ij}^{wR}| = \frac{1}{N} \sum_{i,j} \bar{M}_{\text{old}(ij)} |I_{ij}^L - I_{i(j-\bar{D}_i^L)}^R|$ ，或， $L_{\text{photo}} = \frac{1}{N} \sum_{i,j} \bar{M}_{\text{old}(ij)} |I_{ij}^R - I_{ij}^{wL}| = \frac{1}{N} \sum_{i,j} \bar{M}_{\text{old}(ij)} |I_{ij}^R - I_{i(j+\bar{D}_i^R)}^L|$ 确定所述重建误差，其中，所述 N 表示图片中像素的个数，所述 $\bar{M}_{\text{old}(ij)}$ 表示所述训练后的第二双目匹配网络模型输出的遮挡图的像素值，所述 I_{ij}^L 表示不带深度标记的真实双目数据中的左图的像素值，所述 I_{ij}^R 表示不带深度标记的真实双目数据中的右图的像素值，所述 I_{ij}^{wR} 表示将右图采样后合成的图片的像素值，所述 I_{ij}^{wL} 表示将左图采样后合成的图片的像素值，所述 $\bar{D}_i^L$ 表示不带深度标记的真实双目数据中的左图经第一双目匹配网络模型输出的视差图的像素值，所述 $\bar{D}_i^R$ 表示不带深度标记的真实双目数据中的右图经第一双目匹配网络模型输出的视差图的像素值，所述 ij 表示像素点的像素坐标。

10、根据权利要求 8 所述的方法，其中，所述方法还包括：利用公式 $L_{\text{abs}} = \frac{1}{N} \sum_{i,j} (1 - \bar{M}_{\text{old}(ij)} + \gamma_3) |\bar{D}_{ij}^L - \bar{D}_{\text{old}(ij)}^L|$ ，或， $L_{\text{abs}} = \frac{1}{N} \sum_{i,j} (1 - \bar{M}_{\text{old}(ij)} + \gamma_3) |\bar{D}_{ij}^R - \bar{D}_{\text{old}(ij)}^R|$ 确定所述第一双目匹配网络模型输出的视差图与所述训练后的第二双目匹配网络模型输出的视差图相比偏离较小，其中，所述 N 表示图片中像素的个数，所述 $\bar{M}_{\text{old}(ij)}$ 表示所述训练后的第二双目匹配网络模型输出的遮挡图的像素值，所述 $\bar{D}_{ij}^L$ 表示不带深度标记的真实双目数据中的左图经第一双目匹配网络模型输出的视差图的像素值，所述 $\bar{D}_{ij}^R$ 表示不带深度标记的真实双目数据中的右图经第一双目匹配网络模型输出的视差图的像素值，所述 $\bar{D}_{\text{old}(ij)}^L$ 表示左图经训练后的第二双目匹配网络模型输出的视差图的像素值，所述 $\bar{D}_{\text{old}(ij)}^R$ 表示右图经训练后的第二双目匹配网络模型输出的视差图的像素值，所述 ij 表示像素点的像素坐标，所述 γ_3 表示强度系数。

11、根据权利要求 8 所述的方法，其中，所述方法还包括：利用公式 $L_{\text{rel}} = \frac{1}{N} \sum_{i,j} |\nabla \bar{D}_{ij}^L - \nabla \bar{D}_{\text{old}(ij)}^L|$ ，或， $L_{\text{rel}} = \frac{1}{N} \sum_{i,j} |\nabla \bar{D}_{ij}^R - \nabla \bar{D}_{\text{old}(ij)}^R|$ 确定所述第一双目匹配网络模型的输出梯度与所述第二双目匹配网络模型的输出梯度一致，其中，所述 N 表示图片中像素的个数，所述 $\nabla \bar{D}_{ij}^L$ 表示不带深度标记的真实双目数据中的左图经第一双目匹配网络模型输出的视差图的梯度，所述 $\nabla \bar{D}_{ij}^R$ 表示不带深度标记的真实双目数据中的右图经第一双目匹配网络模型输出的视差图的梯度，所述 $\nabla \bar{D}_{\text{old}(ij)}^L$ 表示左图经训练后的第二双目匹配网络模型输出的视差图的梯度，所述 $\nabla \bar{D}_{\text{old}(ij)}^R$ 表示右图经训练后的第二双目匹配网络模型输出的视差图的梯度，所述 ij 表示像素点的像素坐标。

12、根据权利要求 5 所述的方法，其中，所述带深度标记的真实双目数据包括左图

和右图，对应地，所述单目深度估计网络模型的训练过程，包括：获取所述带深度标记的真实双目数据中的左图或右图作为训练样本；根据所述带深度标记的真实双目数据中的左图或右图对单目深度估计网络模型进行训练。

13、根据权利要求 6 至 11 任一项所述的方法，其中，所述不带深度标记的真实双目数据包括左图和右图，对应地，所述单目深度估计网络模型的训练过程，包括：将所述不带深度标记的真实双目数据输入到所述第一双目匹配神经网络模型，得到对应的视差图；根据所述对应的视差图、拍摄所述不带深度标记的真实双目数据的摄像机的镜头基线距离和拍摄所述不带深度标记的真实双目数据的摄像机的镜头焦距，确定所述视差图对应的深度图；将所述不带深度标记的真实双目数据中的左图或右图作为样本数据，根据所述视差图对应的深度图对单目深度估计网络模型进行监督，从而训练所述单目深度估计网络模型。

14、根据权利要求 12 或 13 所述的方法，其中，所述待处理图像的分析结果包括所述单目深度估计网络模型输出的视差图，对应地，所述方法还包括：根据所述单目深度估计网络模型输出的视差图、拍摄输入所述单目深度估计网络模型的图像的摄像机的镜头基线距离和拍摄输入所述单目深度估计网络模型的图像的摄像机的镜头焦距，确定所述视差图对应的深度图；输出所述视差图对应的深度图。

15、一种单目深度估计装置，其中，所述装置包括：获取模块、执行模块和输出模块，其中：所述获取模块，配置为获取待处理图像；所述执行模块，配置为将所述待处理图像输入至经过训练得到的单目深度估计网络模型，得到所述待处理图像的分析结果，其中，所述单目深度估计网络模型是通过第一双目匹配神经网络模型输出的视差图进行监督训练的；所述输出模块，配置为输出所述待处理图像的分析结果。

16、根据权利要求 15 所述的装置，其中，所述装置还包括：第一训练模块，配置为根据获取的合成样本数据训练第二双目匹配神经网络模型；第二训练模块，配置为根据获取的真实样本数据对训练后的第二双目匹配神经网络模型的参数进行调整，得到第一双目匹配神经网络模型。

17、根据权利要求 16 所述的装置，其中，所述装置还包括：第一获取模块，配置为获取有深度标记的合成的双目图片作为所述合成样本数据，其中，所述合成的双目图片包括合成的左图和合成的右图。

18、根据权利要求 17 所述的装置，其中，所述第一训练模块，包括：第一训练单元，配置为根据所述合成的双目图片对第二双目匹配神经网络模型进行训练，得到训练后的第二双目匹配神经网络模型，其中，所述训练后的第二双目匹配神经网络模型的输出为视差图和遮挡图，所述视差图描述了所述左图中每个像素点与所述右图中对应的像素点的视差距离，所述视差距离以像素为单位；所述遮挡图描述了所述左图中每个像素点在所述右图中对应的像素点是否被物体遮挡。

19、根据权利要求 16 所述的装置，其中，所述第二训练模块，包括：第二训练单元，配置为根据获取的带深度标记的真实双目数据对训练后的第二双目匹配神经网络模

型进行监督训练,以调整所述训练后的第二双目匹配神经网络模型的权值,得到第一双目匹配神经网络模型。

20、根据权利要求 16 所述的装置,其中,所述第二训练单元,还配置为:根据获取的不带深度标记的真实双目数据对训练后的第二双目匹配神经网络模型进行无监督训练,以调整所述训练后的第二双目匹配神经网络模型的权值,得到第一双目匹配神经网络模型。

21、根据权利要求 20 所述的装置,其中,所述第二训练单元,包括:第二训练部件,配置为使用损失函数,根据所述不带深度标记的真实双目数据对训练后的第二双目匹配神经网络模型进行无监督训练,以调整所述训练后的第二双目匹配神经网络模型的权值,得到第一双目匹配神经网络模型。

22、根据权利要求 21 所述的装置,其中,所述装置还包括:第一确定模块,配置为利用公式 $L_{\text{stereo-unsup ft}} = L_{\text{photo}} + \gamma_1 L_{\text{abs}} + \gamma_2 L_{\text{rel}}$ 确定所述损失函数,其中,所述 $L_{\text{stereo-unsup ft}}$ 表示损失函数,所述 L_{photo} 表示重建误差,所述 L_{abs} 表示所述第一双目匹配网络模型输出的视差图与所述训练后的第二双目匹配网络模型输出的视差图相比偏离较小,所述 L_{rel} 表示约束所述第一双目匹配网络模型的输出梯度与所述训练后的第二双目匹配网络模型的输出梯度一致,所述 γ_1 和 γ_2 表示强度系数。

23、根据权利要求 22 所述的装置,其中,所述装置还包括:第二确定模块,配置为利用公式 $L_{\text{photo}} = \frac{1}{N} \sum_{i,j} \bar{M}_{\text{old}(ij)} |I_{ij}^L - I_{ij}^{wR}| = \frac{1}{N} \sum_{i,j} \bar{M}_{\text{old}(ij)} |I_{ij}^L - I_{i(j-D_0^R)}^R|$, 或,

$L_{\text{photo}} = \frac{1}{N} \sum_{i,j} \bar{M}_{\text{old}(ij)} |I_{ij}^R - I_{ij}^{wL}| = \frac{1}{N} \sum_{i,j} \bar{M}_{\text{old}(ij)} |I_{ij}^R - I_{i(j+D_0^L)}^L|$ 确定所述重建误差,其中,所述 N 表示图片中像素的个数,所述 $\bar{M}_{\text{old}(ij)}$ 表示所述训练后的第二双目匹配网络模型输出的遮挡图的像素值,所述 I_{ij}^L 表示不带深度标记的真实双目数据中的左图的像素值,所述 I_{ij}^R 表示不带深度标记的真实双目数据中的右图的像素值,所述 I_{ij}^{wR} 表示将右图采样后合成的图片的像素值,所述 I_{ij}^{wL} 表示将左图采样后合成的图片的像素值,所述 D_0^L 表示不带深度标记的真实双目数据中的左图经第一双目匹配网络模型输出的视差图的像素值,所述 D_0^R 表示不带深度标记的真实双目数据中的右图经第一双目匹配网络模型输出的视差图的像素值,所述 ij 表示像素点的像素坐标。

24、根据权利要求 22 所述的装置,其中,所述装置还包括:第三确定模块,配置为利用公式 $L_{\text{abs}} = \frac{1}{N} \sum_{i,j} (1 - \bar{M}_{\text{old}(ij)} + \gamma_3) |\bar{D}_{ij}^L - \bar{D}_{\text{old}(ij)}^L|$, 或, $L_{\text{abs}} = \frac{1}{N} \sum_{i,j} (1 - \bar{M}_{\text{old}(ij)} + \gamma_3) |\bar{D}_{ij}^R - \bar{D}_{\text{old}(ij)}^R|$ 确定所述第一双目匹配网络模型输出的视差图与所述训练后的第二双目匹配网络模型输出的视差图相比偏离较小,其中,所述 N 表示图片中像素的个数,所述 $\bar{M}_{\text{old}(ij)}$ 表示所述训练后的第二双目匹配网络模型输出的遮挡图的像素值,所述 $\bar{D}_{ij}^L$ 表示不带深度标记的真实双目数据中的左图经第一双目匹配网络模型输出的视差图的像素值,所述 $\bar{D}_{ij}^R$ 表示不带深度标记的真实双目数据中的右图经第一双目匹配网络模型输出的视差图的像素值,所述 $\bar{D}_{\text{old}(ij)}^L$ 表示左图经训练后的第二双目匹配网络模型输出的视差图的像素值,所述 $\bar{D}_{\text{old}(ij)}^R$ 表示右图经

训练后的第二双目匹配网络模型输出的视差图的像素值, 所述 ij 表示像素点的像素坐标, 所述 γ_3 表示强度系数。

25、根据权利要求 22 所述的装置, 其中, 所述装置还包括: 第四确定模块, 配置为利用公式 $L_{rel} = \frac{1}{N} \sum_{i,j} |\nabla \bar{D}_{ij}^L - \nabla \bar{D}_{old(ij)}^L|$, 或, $L_{rel} = \frac{1}{N} \sum_{i,j} |\nabla \bar{D}_{ij}^R - \nabla \bar{D}_{old(ij)}^R|$ 确定所述第一双目匹配网络模型的输出梯度与所述第二双目匹配网络模型的输出梯度一致, 其中, 所述 N 表示图片中像素的个数, 所述 $\nabla \bar{D}_{ij}^L$ 表示不带深度标记的真实双目数据中的左图经第一双目匹配网络模型输出的视差图的梯度, 所述 $\nabla \bar{D}_{ij}^R$ 表示不带深度标记的真实双目数据中的右图经第一双目匹配网络模型输出的视差图的梯度, 所述 $\nabla \bar{D}_{old(ij)}^L$ 表示左图经训练后的第二双目匹配网络模型输出的视差图的梯度, 所述 $\nabla \bar{D}_{old(ij)}^R$ 表示右图经训练后的第二双目匹配网络模型输出的视差图的梯度, 所述 ij 表示像素点的像素坐标。

26、根据权利要求 19 所述的装置, 其中, 所述带深度标记的真实双目数据包括左图和右图, 对应地, 所述装置还包括: 第三训练模块, 配置为获取所述带深度标记的真实双目数据中的左图或右图作为训练样本; 根据所述带深度标记的真实双目数据中的左图或右图对单目深度估计网络模型进行训练。

27、根据权利要求 20 至 25 任一项所述的装置, 其中, 所述不带深度标记的真实双目数据包括左图和右图, 对应地, 所述装置还包括: 第三训练模块, 配置为将所述不带深度标记的真实双目数据输入到所述第一双目匹配神经网络模型, 得到对应的视差图; 根据所述对应的视差图、拍摄所述不带深度标记的真实双目数据的摄像机的镜头基线距离和拍摄所述不带深度标记的真实双目数据的摄像机的镜头焦距, 确定所述视差图对应的深度图; 将所述不带深度标记的真实双目数据中的左图或右图作为样本数据, 根据所述视差图对应的深度图对单目深度估计网络模型进行监督, 从而训练所述单目深度估计网络模型。

28、根据权利要求 26 或 27 所述的装置, 其中, 所述待处理图像的分析结果包括所述单目深度估计网络模型输出的视差图, 对应地, 所述装置还包括: 第五确定模块, 配置为根据所述单目深度估计网络模型输出的视差图、拍摄输入所述单目深度估计网络模型的图片的摄像机的镜头基线距离和拍摄输入所述单目深度估计网络模型的图片的摄像机的镜头焦距, 确定所述视差图对应的深度图; 第一输出模块, 配置为输出所述视差图对应的深度图。

29、一种单目深度估计设备, 包括存储器和处理器, 所述存储器存储有可在处理器上运行的计算机程序, 其中, 所述处理器执行所述程序时实现权利要求 1 至 14 任一项所述单目深度估计方法中的步骤。

30、一种计算机可读存储介质, 其上存储有计算机程序, 其中, 该计算机程序被处理器执行时实现权利要求 1 至 14 任一项所述单目深度估计方法中的步骤。