

三维点云处理模型和语义分割模型的训练方法和相关设备

相关申请的交叉引用

本申请是以中国申请号为 202310518670.1, 申请日为 2023 年 5 月 9 日的申请为基础, 并主张其优先权, 该中国申请的公开内容在此作为整体引入本申请中。

技术领域

本公开涉及三维点云技术领域, 特别涉及一种三维点云处理模型和语义分割模型的训练方法和相关设备。

背景技术

三维点云是指由激光雷达、深度相机等扫描设备获取的空间三维点信息, 一般包含空间三维位置信息、颜色信息和强度信息等。三维点云语义分割是指对点云中的每个点都赋予特定的语义标签, 是场景理解的关键步骤, 在机器人、增强现实和自动驾驶等许多领域都有广泛的应用。因此, 三维点云语义分割一直是计算机视觉领域的研究热点。

在相关技术中, 基于场景分类监督信号的弱监督点云语义分割方法通常采用点云类别激活图的思想来生成每个点的伪标签, 这种方式根据点云数据生成伪标签。

发明内容

根据本公开一些实施例的第一个方面, 提供一种三维点云处理模型的训练方法, 包括: 利用三维点云处理模型对三维点云数据进行处理, 以获取三维点云数据中点的信息; 获取二维图像数据中像素的信息, 其中, 二维图像数据和三维点云数据覆盖相同的区域; 利用点的信息以及像素的信息, 确定三维点云数据和二维图像数据的对齐约束损失; 利用三维点云处理模型预测的、三维点云数据中点的分类信息以及预先标记的场景分类信息, 确定三维点云处理模型的分

类损失; 利用对齐约束损失和分类损失, 训练三维点云处理模型。

在一些实施例中, 点的信息包括点的分类信息、特征信息中的至少一种; 像素的信息包括像素的分类信息、特征信息中的至少一种。

在一些实施例中, 利用点的信息以及像素的信息, 确定三维点云数据和二维图像

数据的对齐约束损失包括：利用点的信息以及像素的信息，确定三维点云数据和二维图像数据的匹配程度；根据匹配程度，确定对齐约束损失。

在一些实施例中，匹配程度包括：点与像素的特征匹配程度、三维点云数据与二维图像数据的全局特征匹配程度、点与像素的语义匹配程度中的至少一种。

5 在一些实施例中，点与像素的特征匹配程度根据点和相应像素之间的相似度确定。

在一些实施例中，在匹配程度包括点与像素的特征匹配程度的情况下，根据匹配程度，确定对齐约束损失包括：将点的特征、和像素的特征映射到同一维度，分别获得每个点的第一点云特征和每个像素的第一像素特征；随机获取第一数量的点以及相应的像素；对于第一数量的点中的每一个，利用点的第一点云特征与相应像素的第一像素特征的相似度、以及点的第一点云特征与每个像素的第一像素特征的相似度的总和，
10 确定点与相应像素的对比学习损失；根据第一数量的点中的每一个的对比学习损失的总和，确定对齐约束损失。

在一些实施例中，对于第一数量的点中的每一个：点与相应像素的对比学习损失，和点的第一点云特征与相应像素的第一像素特征的相似度成负相关关系、和点的第一点云特征与每个像素的第一像素特征的相似度的总和成正相关关系。
15

在一些实施例中，全局特征匹配程度根据三维点云数据中任意两个点之间的相似度、以及二维图像数据中任意两个像素之间的相似度确定。

在一些实施例中，在匹配程度包括全局特征匹配程度的情况下，根据匹配程度，确定对齐约束损失包括：将点的特征、和像素的特征映射到同一维度，分别获得每个点的第二点云特征和每个像素的第二像素特征；随机获取第二数量的点以及相应的像素；对于第二数量的点中的任意两个点，确定两个点的第二点云特征之间的第一相似性信息、以及第二数量的像素中两个相应像素的第二像素特征之间的第二相似性信息，并根据第一相似性信息和第二相似性信息之间的差距确定两个点对应的相似性损失；根据三维点云数据中的点的相似性损失的总和，确定对齐约束损失。
20

25 在一些实施例中，第一相似性信息和第二相似性信息为相似度矩阵。

在一些实施例中，语义匹配程度根据每个点的第一分类信息、以及每个像素的第二分类信息确定。

在一些实施例中，其中，第二分类信息通过利用二维图像处理模型对每个像素进行处理获得，并且在匹配程度包括语义匹配程度的情况下，根据匹配程度，确定对齐约束损失包括：将三维点云数据映射到与二维图像数据相同的维度；利用二维图像处
30

理模型，对映射后的三维点云数据进行处理，获得每个点的第三分类信息；利用第二分类信息和第三分类信息确定语义匹配损失，并根据语义匹配损失确定对齐约束损失。

在一些实施例中，对齐约束损失与第二分类信息成正相关关系、与第三分类信息成负相关关系。

5 在一些实施例中，获取二维图像数据的信息包括：利用二维图像处理模型对二维图像数据进行处理，获得二维图像数据的信息。

在一些实施例中，利用三维点云处理模型预测的、三维点云数据中点的分类信息以及预先标记的场景分类信息，确定三维点云处理模型的分类损失包括：根据三维点云处理模型预测的、三维点云数据中点的分类信息以及预先标记的场景分类信息，确定第一分类损失；根据二维图像处理模型预测的、二维图像数据中像素的分类信息以及场景分类信息，确定第二分类损失；根据第一分类损失和第二分类损失，确定三维点云处理模型的分类损失。

15 在一些实施例中，根据三维点云处理模型预测的、三维点云数据的分类信息的池化处理结果以及场景分类信息，确定第一分类损失；根据二维图像处理模型预测的、二维图像数据中像素的分类信息以及场景分类信息，确定第二分类损失。

在一些实施例中，利用对齐约束损失和分类损失的加权结果，训练三维点云处理模型。

在一些实施例中，三维点云处理模型利用三维残差网络实现。

在一些实施例中，二维图像处理模型利用二维残差网络实现。

20 根据本公开一些实施例的第二个方面，提供一种语义分割模型的训练方法，包括：利用三维点云处理模型预测待处理的三维点云数据中的分类信息，其中，三维点云处理模型是利用前述任意一种三维点云处理模型的训练方法训练的；根据分类信息生成待处理的三维点云数据中的点的伪标签；利用伪标签，训练三维点云的语义分割模型。

根据本公开一些实施例的第三个方面，提供一种三维点云处理模型的训练装置，
25 包括：第一信息获取模块，被配置为利用三维点云处理模型对三维点云数据进行处理，以获取三维点云数据中点的信息；第二信息获取模块，被配置为获取二维图像数据中像素的信息，其中，二维图像数据和三维点云数据覆盖相同的区域；对齐约束损失确定模块，被配置为利用点的信息以及像素的信息，确定三维点云数据和二维图像数据的对齐约束损失；分类损失确定模块，被配置为利用三维点云处理模型预测的、三维点云数据中点的分类信息以及预先标记的场景分类信息，确定三维点云处理模型的分
30

类损失；三维点云处理模型训练模块，被配置为利用对齐约束损失和分类损失，训练三维点云处理模型。

根据本公开一些实施例的第四个方面，提供一种三维点云处理模型的训练装置，包括：存储器；以及耦接至存储器的处理器，处理器被配置为基于存储在存储器中的指令，执行前述任意一种三维点云处理模型的训练方法。

根据本公开一些实施例的第五个方面，提供一种语义分割模型的训练系统，包括：前述任意一种三维点云处理模型的训练装置；预测模块，被配置为利用三维点云处理模型预测待处理的三维点云数据中的分类信息；生成模块，被配置为根据分类信息生成待处理的三维点云数据中的点的伪标签；语义分割模型训练模块，被配置为利用伪标签，训练三维点云的语义分割模型。

根据本公开一些实施例的第六个方面，提供一种计算机可读存储介质，其上存储有计算机程序，该程序被处理器执行时实现前述任意一种三维点云处理模型的训练方法。

根据本公开一些实施例的第六个方面，提供一种计算机程序，包括：指令，所述指令当由处理器执行时使所述处理器执行前述任意一种三维点云处理模型的训练方法。

通过以下参照附图对本公开的示例性实施例的详细描述，本公开的其它特征及其优点将会变得清楚。

附图说明

为了更清楚地说明本公开实施例或现有技术中的技术方案，下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本公开的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动性的前提下，还可以根据这些附图获得其他的附图。

- 图 1 示出了根据本公开一些实施例的处理模型的训练方法的流程示意图。
- 图 2 示出了根据本公开一些实施例的对齐约束损失确定方法的流程示意图。
- 图 3 示出了根据本公开另一些实施例的对齐约束损失确定方法的流程示意图。
- 图 4 示出了根据本公开又一些实施例的对齐约束损失确定方法的流程示意图。
- 图 5 示出了根据本公开一些实施例的分类损失确定方法的流程示意图。
- 图 6 示出了根据本公开一些实施例的三维点云的语义分割模型的训练方法的流程

示意图。

图 7 示出了根据本公开一些实施例的三维点云处理模型的训练装置的结构示意图。

图 8 示出了根据本公开一些实施例的语义分割模型的训练系统的结构示意图。

图 9 示出了根据本公开另一些实施例的三维点云处理模型的训练装置的结构示意图。

图 10 示出了根据本公开又一些实施例的三维点云处理模型的训练装置的结构示意图。

具体实施方式

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。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。基于本公开中的实施例，本领域普通技术人员在没有作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。

除非另外具体说明，否则在这些实施例中阐述的部件和步骤的相对布置、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。

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。

在这里示出和讨论的所有示例中，任何具体值应被解释为仅仅是示例性的，而不是作为限制。因此，示例性实施例的其它示例可以具有不同的值。

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。

经过分析后发现，基于点云类别激活图生成伪标签的方式仅考虑了点云数据这一个模态，而生成的伪标签不准确又会使得三维点云的语义分割结果不准确。

本公开实施例所要解决的一个技术问题是：如何提高三维点云数据的标注准确性。

图 1 示出了根据本公开一些实施例的处理模型的训练方法的流程示意图。如图 1 所示，该实施例的处理模型的训练方法包括步骤 S102~S110。

在步骤 S102 中，利用三维点云处理模型对三维点云数据进行处理，以获取三维

点云数据中点的信息。

三维点云数据包括若干点的数据，每个点的数据可以包括位置信息和颜色信息。在一些实施例中，三维点云数据以 $\mathbf{P} \in \mathbb{R}^{N \times 6}$ 表示，其中 N 表示点的数量， 6 表示 6 个维度（ 3 维颜色信息以及 3 维位置信息）。

- 5 三维点云处理模型用于对三维点云数据中的每个点进行分类。例如，可以输出每个点属于每种类别的概率，然后根据概率阈值来确定该点所属的一个或多个类。

三维点云处理模型还可以输出每个点的特征。例如，可以将三维点云处理模型的中间计算结果输出，该输出表示点的特征。

- 10 在一些实施例中，三维点云数据中点的信息包括三维点云数据中的点的分类信息、特征信息中的至少一种。

在一些实施例中，三维点云处理模型利用三维残差网络实现。三维点云处理模型可以同时输出三维点云数据的逐点特征图 $\mathbf{F}^{3d} \in \mathbb{R}^{N \times L}$ 和分类预测概率 $\mathbf{M}^{3d} \in \mathbb{R}^{N \times C}$ ，其中 N 表示点的数量， L 表示三维点云点特征的维度， C 表示类别数。

- 15 在步骤 S104 中，获取二维图像数据中像素的信息，其中，二维图像数据和三维点云数据覆盖相同的区域。

二维图像数据包括若干像素的数据，每个像素的数据可以包括颜色信息，当然也可以采用其他方式表示像素。在一些实施例中，二维图像数据以 $\mathbf{I} \in \mathbb{R}^{H \times W \times 3}$ 表示，其中 $H \times W$ 表示图像的长乘以宽， 3 表示三个颜色通道（例如红、绿、蓝通道）。

- 20 由于采集设备的差异，在部分情况下，三维点云数据和二维图像数据所覆盖的区域可能不完全一致。但是，二者覆盖的区域存在交集。

在一些实施例中，二维图像数据中像素的信息包括像素的分类信息、特征信息中的至少一种。二维图像数据的信息类型可以与三维点云数据的信息类型对应，例如，二者都为特征信息、都为分类信息或者都同时包括特征信息和分类信息。从而，可以从特征角度、语义角度描述三维点云数据和二维图像数据。

- 25 在一些实施例中，利用二维图像处理模型对二维图像数据进行处理，获得二维图像数据的信息。从而，可以提高信息获取的效率。根据需要，二维图像处理模型也可以与三维点云处理模型共同训练，从而从整体上提高预测的准确率。

- 30 在一些实施例中，二维图像处理模型利用二维残差网络实现。同时输出二维图像的逐像素特征图 $\mathbf{F}^{2d} \in \mathbb{R}^{H \times W \times D}$ 和分类预测概率 $\mathbf{M}^{2d} \in \mathbb{R}^{H \times W \times C}$ ，其中 $H \times W$ 表示图像的长乘以宽、即像素数量， D 表示二维图像像素特征的维度， C 表示语义分割的类

别个数。

在步骤 S106 中，利用点的信息以及像素的信息，确定三维点云数据和二维图像数据的对齐约束损失。

5 对齐约束是指将三维点云数据和二维图像数据对齐的约束。在一些实施例中，可以基于点和像素的特征进行约束、也可以基于点和像素的语义进行约束。此外，还可以基于三维点云和二维图像的全局属性（例如全局特征）进行约束。

10 在一些实施例中，利用点的信息以及像素的信息，确定三维点云数据和二维图像数据的匹配程度；根据匹配程度，确定对齐约束损失。基于三维点云数据和二维图像数据的匹配程度确定对齐约束损失、并进行训练，可以使得三维点云处理模型将三维点云数据和二维图像数据进行对齐，例如，将相似的点和像素拉近、将不相似的点和像素拉远，或者尽量使得二者在对应位置上具有相同的语义。

15 在一些实施例中，三维点云数据和二维图像数据之间的匹配程度包括：三维点云数据中的点与二维图像数据中的像素的特征匹配程度、三维点云数据与二维图像数据的全局特征匹配程度、三维点云数据中的点与二维图像数据中的像素的语义匹配程度中的至少一种。后文将示例性地描述这三种约束的确定方式的实施例。根据需要，本领域技术人员还可以使用其他方式确定匹配程度。

在步骤 S108 中，利用三维点云处理模型预测的、三维点云数据中点分类信息以及预先标记的场景分类信息，确定三维点云处理模型的分类损失。即，预测值和标记值的差距越大，则分类损失越大。

20 场景分类信息是描述整个场景的分类结果的信息，例如，某个场景中包括桌子和椅子，则其分类信息可以包括“桌子”“椅子”对应的信息。也就是说，在本公开的实施例中，在对三维点云处理模型进行训练之前不需要逐点标注三维点云数据，而是对其进行整体标注。

25 由于预测的分类信息是逐点信息，为了将其与场景分类信息对应，可以通过池化处理获得对整个场景的预测结果，并进一步计算分类损失。从而，能够节约预先标注的工作量，从整体上提高了训练效率。

在一些实施例中，分类损失基于交叉熵损失函数确定。根据需要，也可以采用其他方式计算分类损失，这里不再赘述。

在步骤 S110 中，利用对齐约束损失和分类损失，训练三维点云处理模型。

30 在一些实施例中，利用对齐约束损失和分类损失的加权结果，训练三维点云处理

模型。在对齐约束损失包括多个种类的情况，或者分类损失包括多个种类的情况，每个种类也可以对应有预设的权重。

5 在一些实施例中，在使用了二维图像处理模型得到二维图像数据的信息的情况下，还可以利用二维图像数据的分类损失训练二维图像处理模型。从而，使得二维图像处理模型能够给出更准确的预测结果。

本公开的实施例在训练三维点云处理模型的过程中，参考了三维点云数据和二维图像数据的对齐约束损失，从而将二维图像引入到三维点云处理模型的训练过程中，实现了二维图像对三维点云的跨模态引导，提升了模型训练的准确性。

下面分别示例性地描述基于三种匹配程度确定对齐约束的方式。

10 在一些实施例中，点与像素的特征匹配程度根据点和相应像素之间的相似度确定。下面参考图 2 描述对齐约束损失确定方法的实施例。

图 2 示出了根据本公开一些实施例的对齐约束损失确定方法的流程示意图。如图 2 所示，该实施例的对齐约束损失确定方法包括步骤 S202~S208。

15 在步骤 S202 中，将点的特征、和像素的特征映射到同一维度，分别获得每个点的第一点云特征和每个像素的第一像素特征。

例如，将点的特征和像素的特征都处理成一个新的 E 维数据。点和像素映射可以分别通过两个全连接层网络实现。

在步骤 S204 中，随机获取第一数量的点以及相应的像素。

20 相应的点和像素可以预先通过特征匹配等手段确定。例如，当场景中包括一个桌角时，三维点云数据中桌角对应的点、以及二维图像数据中桌角对应的像素就是相应的点和像素。

在步骤 S206 中，对于第一数量的点中的每一个，利用点的第一点云特征与相应像素的第一像素特征的相似度、以及点的第一点云特征与每个像素的第一像素特征的相似度的总和，确定点与相应像素的对比学习损失。

25 为了便于描述，设相应的点和像素具有相同的标识，即第 i 个点和第 i 个像素对应。则对于第 i 个点，利用第 i 个点的第一点云特征与第 i 个像素的第一像素特征的相似度、以及第 i 个点的第一点云特征与第 1、2…… N_c （表示第一数量）个像素中每个像素的第一像素特征的相似度的总和，确定第 i 个点和第 i 个像素的对比学习损失。

30 在一些实施例中，点与相应像素的对比学习损失，和该点的第一点云特征与相应像素的第一像素特征的相似度成负相关关系、和该点的第一点云特征与每个像素的第

一像素特征的相似度的总和成正相关关系。

在步骤 S208 中，根据第一数量的点中的每一个的对比学习损失的总和，确定对齐约束损失。对齐约束损失与对比学习损失的总和成正相关关系。

设对比学习损失的总和为“点到像素对比损失”。在一些实施例中，点到像素对比损失 $\mathcal{L}_c$ 根据公式 (1) 确定。

$$\mathcal{L}_c = - \sum_{i=1}^{N_c} \log \frac{\exp(F_c^{3d}(i) \cdot F_c^{2d}(i)/\tau)}{\sum_{k=1}^{N_c} \exp(F_c^{3d}(i) \cdot F_c^{2d}(k)/\tau)} \quad (1)$$

其中， F_c^{3d} 为三维点云数据中点的映射后的特征， F_c^{2d} 为二维图像数据中像素的映射后的特征， N_c 为点和像素的数量， τ 为调节参数， i 和 k 为点和像素的标识。该损失函数的目的是增大点与其对应的像素（正样例）之间的特征相似度，同时缩小与其非对应的像素（负样例）之间的相似度。而调节参数控制的是该损失函数对困难负样例（相似度较高的负样例）的关注程度。调节参数越大，关注度越低；调节参数越小，关注度越高。在具体应用中，可以通过实验来选择合适的调节参数。

通过上述实施例，对应的像素的特征和点的特征被拉近、不对应的特征被拉远，实现了将点的特征对齐到图像的特征的目的。

在一些实施例中，全局特征匹配程度根据三维点云数据中任意两个点之间的相似度、以及二维图像数据中任意两个像素之间的相似度确定。下面参考图 3 描述对齐约束损失确定方法的实施例。

图 3 示出了根据本公开另一些实施例的对齐约束损失确定方法的流程示意图。如图 3 所示，该实施例的对齐约束损失确定方法包括步骤 S302~S308。

在步骤 S302 中，将点的特征、和像素的特征映射到同一维度，分别获得每个点的第二点云特征和每个像素的第二像素特征。

例如，将点的特征和像素的特征都处理成一个新的 F 维数据。点和像素映射可以分别通过两个全连接层网络实现。步骤 S302 可以是步骤 S202，也可以是在步骤 S202 以外进行的额外的处理。即，E 和 F 可以相等、也可以不相等，步骤 S302 使用的全连接层网络可以与步骤 S202 使用的全连接层网络相同、也可以不同。

在步骤 S304 中，随机获取第二数量的点以及相应的像素。第二数量与第一数量可以相同、也可以不同。

在步骤 S306 中，对于第二数量的点中的任意两个点，确定这两个点的第二点云特征之间的第一相似性信息、以及第二数量的像素中两个相应像素的第二像素特征之间的第二相似性信息，并根据第一相似性信息和第二相似性信息之间的差距确定这两

个点对应的相似性损失。

在一些实施例中，第一相似性信息和第二相似性信息为相似度矩阵。相似性矩阵描述了任意两个点或任意两个像素之间的相似性，能够被看作是一种全局结构信息。

在步骤 S308 中，根据三维点云数据中的点的相似性损失的总和，确定对齐约束损失。

设相似性损失的总和为“全局结构相似性损失”。在一些实施例中，全局结构相似性损失 $\mathcal{L}_a$ 根据公式（2）-（4）确定。

$$A^{3d}(i, j) = \frac{F_a^{3d}(i) \cdot F_a^{3d}(j)}{\|F_a^{3d}(i)\|_2 \|F_a^{3d}(j)\|_2} \quad (2)$$

$$A^{2d}(i, j) = \frac{F_a^{2d}(i) \cdot F_a^{2d}(j)}{\|F_a^{2d}(i)\|_2 \|F_a^{2d}(j)\|_2} \quad (3)$$

$$\mathcal{L}_a = \frac{1}{N_a} \sum_{i=1}^{N_a} \sum_{j=1}^{N_a} \|A^{3d}(i, j) - A^{2d}(i, j)\|_2^2 \quad (4)$$

其中， F_a^{3d} 为三维点云数据中点的特征， F_a^{2d} 为二维图像数据中像素的特征， $A^{3d}(i, j)$ 为第一相似性信息， $A^{2d}(i, j)$ 为第二相似性信息， N_a 为点和像素的数量， i 、 j 和 k 为点和像素的标识。

通过上述实施例，三维点云的处理模型能够学习到图像特征的全局结构信息，从而实现三维点云和二维图像的全局结构层面的对齐。

在一些实施例中，语义匹配程度根据每个点的第一分类信息、以及每个像素的第二分类信息确定。每个点的第一分类信息可以通过三维点云的处理模型的输出确定，每个像素的第二分类信息可以通过二维图像的处理模型的输出确定。下面参考图 4 描述对齐约束损失确定方法的实施例。

图 4 示出了根据本公开又一些实施例的对齐约束损失确定方法的流程示意图。如图 4 所示，该实施例的对齐约束损失确定方法包括步骤 S402~S406。

在步骤 S402 中，将三维点云数据映射到与二维图像数据相同的维度。

在步骤 S404 中，利用二维图像处理模型，对映射后的三维点云数据进行处理，获得每个点的第三分类信息。

在步骤 S406 中，利用第二分类信息和第三分类信息确定语义匹配损失，并根据语义匹配损失对齐约束损失。

在一些实施例中，对齐约束损失与第二分类信息成正相关关系、与第三分类信息成负相关关系。

在一些实施例中，语义匹配损失 $\mathcal{L}_m$ 根据公式（5）确定。

$$\mathcal{L}_m = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C M^{2d}(i, c) \log \frac{M^{2d}(i, c)}{\hat{M}^{3d}(i, c)} \quad (5)$$

其中, N 表示点的数量, C 表示类别数量, $M^{2d}(i, c)$ 表示二维图像处理模型对二维图像数据中像素的预测结果, $\hat{M}^{3d}(i, c)$ 表示二维图像处理模型对映射后的三维点云数据中点的预测结果。

- 5 通过上述实施例, 三维点云处理模型能够学习到图像特征的语义信息, 从而实现语义层面的对齐。

在利用二维图像处理模型获得二维图像数据的信息时, 在确定分类损失时, 还可以考虑二维图像处理模型的分分类损失。下面参考图 5 描述本公开分类损失确定方法的实施例。

- 10 图 5 示出了根据本公开一些实施例的分类损失确定方法的流程示意图。如图 5 所示, 该实施例的分类损失确定方法包括步骤 S502~S506。

在步骤 S502 中, 根据三维点云处理模型预测的、三维点云数据中点的分类信息以及预先标记的场景分类信息, 确定第一分类损失。

- 15 在步骤 S504 中, 根据二维图像处理模型预测的、二维图像数据中像素的分类信息以及场景分类信息, 确定第二分类损失。

由于三维点云数据和二维图像数据覆盖同一区域、即针对同一场景, 因此二者可以共享预先标记的场景分类信息。

在步骤 S506 中, 根据第一分类损失和第二分类损失, 确定三维点云处理模型的分分类损失。

- 20 在一些实施例中, 分类信息为分类概率。分类损失包括根据全局平均池化后的、三维点云数据分类信息以及场景分类信息确定的第一分类损失、和根据全局平均池化后的分类信息以及场景分类信息确定的第二分类损失。

- 25 例如, 可以通过公式 (6) 和 (7) 计算池化后的分类信息, 其中, Z^{3d} 表示池化后的、三维点云处理模型预测的分类信息, $M^{3d}(n)$ 表示池化前的、三维点云处理模型预测的分类信息, Z^{2d} 表示池化后的、二维图像处理模型预测的分类信息, $M^{2d}(h, w)$ 表示池化前的、二维图像处理模型预测的分类信息, N 表示三维点云中点的数量 H 和 W 分别表示二维图像的长和宽。

$$Z^{3d} = \frac{1}{N} \sum_{n=1}^N M^{3d}(n) \quad (6)$$

$$Z^{2d} = \frac{1}{HW} \sum_{h=1}^H \sum_{w=1}^W M^{2d}(h, w) \quad (7)$$

- 30 然后, 可以通过公式 (8) 和 (9) 计算第一分类损失 $\mathcal{L}_{cls}^{3d}$ 和第二分类损失 $\mathcal{L}_{cls}^{2d}$, 其

中, C 表示类别数量, c 表示类别标识, Y_c 表示场景的多类别分类标签, σ 表示 sigmoid 激活函数。

$$\mathcal{L}_{cls}^{3d} = -\frac{1}{C} \sum_{c=1}^C [Y_c \cdot \log \sigma(Z_c^{3d}) + (1 - Y_c) \cdot \log(1 - \sigma(Z_c^{3d}))] \quad (8)$$

$$\mathcal{L}_{cls}^{2d} = -\frac{1}{C} \sum_{c=1}^C [Y_c \cdot \log \sigma(Z_c^{2d}) + (1 - Y_c) \cdot \log(1 - \sigma(Z_c^{2d}))] \quad (9)$$

- 5 通过上述实施例, 能够在计算分类损失时, 在考虑三维点云处理模型的分类准确率以外, 还考虑二维图像处理模型的分类准确率, 从而考虑了预测结果中受到二维图像处理模型的影响, 进一步提高了三维点云处理模型的训练效果。

在完成了三维点云处理模型的训练后, 可以将其应用于三维点云的语义分割模型的训练。下面参考图 6 描述三维点云的语义分割模型的训练方法的实施例。

- 10 图 6 示出了根据本公开一些实施例的三维点云的语义分割模型的训练方法的流程示意图。如图 6 所示, 该实施例的训练方法包括步骤 S602~S606。

在步骤 S602 中, 利用三维点云处理模型预测待处理的三维点云数据中的分类信息。

该三维点云处理模型是利用前述任意一个实施例中的方法训练的。

- 15 在步骤 S604 中, 根据分类信息生成待处理的三维点云数据中的点的伪标签。每个伪标签表示点的类别。

在步骤 S606 中, 利用伪标签, 训练三维点云的语义分割模型。即, 将三维点云处理模型的预测结果作为训练数据的标记信息, 进一步利用训练数据训练语义分割模型。

- 20 在一些实施例中, 在生成伪标签类别时, 对于输入的点云 $P \in \mathbb{R}^{N \times 6}$, 首先通过三维点云处理模型得到每个点的分类概率 $M^{3d} \in \mathbb{R}^{N \times C}$, 然后对于每个点在 C 个类别中选择的预测概率最大的类别作为伪标签, 得到逐点的伪标签 $\tilde{Y} \in \mathbb{R}^N$, 再将其作为监督信号训练语义分割模型。

在一些实施例中, 采用交叉熵损失训练基于三维残差网络的三维 U-Net 分割模型。

- 25 根据需要, 也可以训练其他类型的语义分割网络, 或者, 将训练后的三维点云处理模型应用到其他三维点云弱监督学习任务中。

下面参考图 7 描述三维点云处理模型的训练装置的实施例。

图 7 示出了根据本公开一些实施例的三维点云处理模型的训练装置的结构示意图。

如图 7 所示, 该实施例的三维点云处理模型的训练装置 700 包括: 第一信息获取模块

- 30 7100, 被配置为利用三维点云处理模型对三维点云数据进行处理, 以获取三维点云数

据中点的信息；第二信息获取模块 7200，被配置为获取二维图像数据中像素的信息，其中，二维图像数据和三维点云数据覆盖相同的区域；对齐约束损失确定模块 7300，被配置为利用点的信息以及像素的信息，确定三维点云数据和二维图像数据的对齐约束损失；分类损失确定模块 7400，被配置为利用三维点云处理模型预测的、三维点云数据中点的分类信息以及预先标记的场景分类信息，确定三维点云处理模型的分

5 类损失；三维点云处理模型训练模块 7500，被配置为利用对齐约束损失和分类损失，训练三维点云处理模型。

在一些实施例中，点的信息包括点的分类信息、特征信息中的至少一种；像素的信息包括像素的分类信息、特征信息中的至少一种。

10 在一些实施例中，对齐约束损失确定模块 7300 进一步被配置为利用点的信息以及像素的信息，确定三维点云数据和二维图像数据的匹配程度；根据匹配程度，确定对齐约束损失。

在一些实施例中，匹配程度包括：点与像素的特征匹配程度、三维点云数据与二维图像数据的全局特征匹配程度、点与像素的语义匹配程度中的至少一种。

15 在一些实施例中，点与像素的特征匹配程度根据点和相应像素之间的相似度确定。

在一些实施例中，对齐约束损失确定模块 7300 进一步被配置为在匹配程度包括点与像素的特征匹配程度的情况下：将点的特征、和像素的特征映射到同一维度，分别获得每个点的第一点云特征和每个像素的第一像素特征；随机获取第一数量的点以及相应的像素；对于第一数量的点中的每一个，利用点的第一点云特征与相应像素的第一像素特征的相似度、以及点的第一点云特征与每个像素的第一像素特征的相似度的总和，确定点与相应像素的对比学习损失；根据第一数量的点中的每一个的对比学习损失的总和，确定对齐约束损失。

20

在一些实施例中，对于第一数量的点中的每一个：点与相应像素的对比学习损失，和点的第一点云特征与相应像素的第一像素特征的相似度成负相关关系、和点的第一点云特征与每个像素的第一像素特征的相似度的总和成正相关关系。

25

在一些实施例中，全局特征匹配程度根据三维点云数据中任意两个点之间的相似度、以及二维图像数据中任意两个像素之间的相似度确定。

在一些实施例中，对齐约束损失确定模块 7300 进一步被配置为在匹配程度包括全局特征匹配程度的情况下：将点的特征、和像素的特征映射到同一维度，分别获得每个点的第二点云特征和每个像素的第二像素特征；随机获取第二数量的点以及相应

30

的像素；对于第二数量的点中的任意两个点，确定两个点的第二点云特征之间的第一相似性信息、以及第二数量的像素中两个相应像素的第二像素特征之间的第二相似性信息，并根据第一相似性信息和第二相似性信息之间的差距确定两个点对应的相似性损失；根据三维点云数据中的点的相似性损失的总和，确定对齐约束损失。

5 在一些实施例中，第一相似性信息和第二相似性信息为相似度矩阵。

在一些实施例中，语义匹配程度根据每个点的第一分类信息、以及每个像素的第二分类信息确定。

10 在一些实施例中，其中，第二分类信息通过利用二维图像处理模型对每个像素进行处理获得，并且对齐约束损失确定模块 7300 进一步被配置为在匹配程度包括语义匹配程度的情况下：将三维点云数据映射到与二维图像数据相同的维度；利用二维图像处理模型，对映射后的三维点云数据进行处理，获得每个点的第三分类信息；利用第二分类信息和第三分类信息确定语义匹配损失，并根据语义匹配损失确定对齐约束损失。

15 在一些实施例中，对齐约束损失与第二分类信息成正相关关系、与第三分类信息成负相关关系。

在一些实施例中，第二信息获取模块 7200 进一步被配置为利用二维图像处理模型对二维图像数据进行处理，获得二维图像数据的信息。

20 在一些实施例中，分类损失确定模块 7400 进一步被配置为根据三维点云处理模型预测的、三维点云数据中点的分类信息以及预先标记的场景分类信息，确定第一分类损失；根据二维图像处理模型预测的、二维图像数据中像素的分类信息以及场景分类信息，确定第二分类损失；根据第一分类损失和第二分类损失，确定三维点云处理模型的分类损失。

25 在一些实施例中，分类损失确定模块 7400 进一步被配置为根据三维点云处理模型预测的、三维点云数据的分类信息的池化处理结果以及场景分类信息，确定第一分类损失；根据二维图像处理模型预测的、二维图像数据中像素的分类信息以及场景分类信息，确定第二分类损失。

在一些实施例中，训练装置 700 还包括：二维图像处理模型训练模块 7600，被配置为利用对齐约束损失和分类损失，训练二维图像处理模型。

30 在一些实施例中，三维点云处理模型训练模块 7500 进一步被配置为利用对齐约束损失和分类损失的加权结果，训练三维点云处理模型。

在一些实施例中，三维点云处理模型利用三维残差网络实现。

在一些实施例中，二维图像处理模型利用二维残差网络实现。

下面参考图 8 描述语义分割模型的训练系统的实施例。

图 8 示出了根据本公开一些实施例的语义分割模型的训练系统的结构示意图。如图 8 所示，该实施例的语义分割模型的训练系统 80 包括：三维点云处理模型的训练装置 810，其具体实施方式可以参考本公开任一实施例；预测模块 820，被配置为利用三维点云处理模型预测待处理的三维点云数据中的分类信息；生成模块 830，被配置为根据分类信息生成待处理的三维点云数据中的点的伪标签；语义分割模型训练模块 840，被配置为利用伪标签，训练三维点云的语义分割模型。

图 9 示出了根据本公开另一些实施例的三维点云处理模型的训练装置的结构示意图。如图 9 所示，该实施例的三维点云处理模型的训练装置 90 包括：存储器 910 以及耦接至该存储器 910 的处理器 920，处理器 920 被配置为基于存储在存储器 910 中的指令，执行前述任意一个实施例中的三维点云处理模型的训练方法。

其中，存储器 910 例如可以包括系统存储器、固定非易失性存储介质等。系统存储器例如存储有操作系统、应用程序、引导装载程序 (Boot Loader) 以及其他程序等。

图 10 示出了根据本公开又一些实施例的三维点云处理模型的训练装置的结构示意图。如图 10 所示，该实施例的三维点云处理模型的训练装置 100 包括：存储器 1010 以及处理器 1020，还可以包括输入输出接口 1030、网络接口 1040、存储接口 1050 等。这些接口 1030，1040，1050 以及存储器 1010 和处理器 1020 之间例如可以通过总线 1060 连接。其中，输入输出接口 1030 为显示器、鼠标、键盘、触摸屏等输入输出设备提供连接接口。网络接口 1040 为各种联网设备提供连接接口。存储接口 1050 为 SD 卡、U 盘等外置存储设备提供连接接口。

本公开的实施例还提供一种计算机可读存储介质，其上存储有计算机程序，其特征在于，该程序被处理器执行时实现前述任意一种三维点云处理模型的训练方法。

本领域内的技术人员应当明白，本公开的实施例可提供为方法、系统、或计算机程序产品。因此，本公开可采用完全硬件实施例、完全软件实施例、或结合软件和硬件方面的实施例的形式。而且，本公开可采用在一个或多个其中包含有计算机可用程序代码的计算机可用非瞬时性存储介质 (包括但不限于磁盘存储器、CD-ROM、光学存储器等) 上实施的计算机程序产品的形式。

本公开是参照根据本公开实施例的方法、设备 (系统)、和计算机程序产品的流

程图和 / 或方框图来描述的。应理解为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。

以上所述仅为本公开的较佳实施例，并不用以限制本公开，凡在本公开的精神和原则之内，所作的任何修改、等同替换、改进等，均应包含在本公开的保护范围之内。

权 利 要 求

1. 一种三维点云处理模型的训练方法，包括：

利用三维点云处理模型对所述三维点云数据进行处理，以获取所述三维点云数据中点的信息；

获取二维图像数据中像素的信息，其中，所述二维图像数据和所述三维点云数据覆盖相同的区域；

利用所述点的信息以及所述像素的信息，确定所述三维点云数据和所述二维图像数据的对齐约束损失；

利用所述三维点云处理模型预测的、所述三维点云数据中点的分类信息以及预先标记的场景分类信息，确定所述三维点云处理模型的分​​类损失；

利用所述对齐约束损失和所述分类损失，训练所述三维点云处理模型。

2. 根据权利要求 1 所述的训练方法，其中：

所述点的信息包括所述点的分类信息、特征信息中的至少一种；

所述像素的信息包括所述像素的分类信息、特征信息中的至少一种。

3. 根据权利要求 1 或 2 所述的训练方法，其中，所述利用所述点的信息以及所述像素的信息，确定所述三维点云数据和所述二维图像数据的对齐约束损失包括：

利用所述点的信息以及所述像素的信息，确定所述三维点云数据和所述二维图像数据的匹配程度；

根据所述匹配程度，确定所述对齐约束损失。

4. 根据权利要求 3 所述的训练方法，其中，所述匹配程度包括：所述点与所述像素的特征匹配程度、所述三维点云数据与所述二维图像数据的全局特征匹配程度、所述点与所述像素的语义匹配程度中的至少一种。

5. 根据权利要求 4 所述的训练方法，其中，所述点与所述像素的特征匹配程度根据点和相应像素之间的相似度确定。

6. 根据权利要求 5 所述的训练方法, 其中, 在所述匹配程度包括所述点与所述像素的特征匹配程度的情况下, 所述根据所述匹配程度, 确定所述对齐约束损失包括:

将所述点的特征、和所述像素的特征映射到同一维度, 分别获得每个点的第一点云特征和每个像素的第一像素特征;

随机获取第一数量的点以及相应的像素;

对于第一数量的点中的每一个, 利用所述点的第一点云特征与相应像素的第一像素特征的相似度、以及所述点的第一点云特征与每个像素的第一像素特征的相似度的总和, 确定所述点与相应像素的对比学习损失;

根据第一数量的点中的每一个的对比学习损失的总和, 确定所述对齐约束损失。

7. 根据权利要求 6 所述的训练方法, 其中, 对于第一数量的点中的每一个:

所述点与相应像素的对比学习损失, 和所述点的第一点云特征与相应像素的第一像素特征的相似度成负相关关系、和所述点的第一点云特征与每个像素的第一像素特征的相似度的总和成正相关关系。

8. 根据权利要求 4~7 中任一项所述的训练方法, 其中, 所述全局特征匹配程度根据所述三维点云数据中任意两个点之间的相似度、以及所述二维图像数据中任意两个像素之间的相似度确定。

9. 根据权利要求 8 所述的训练方法, 其中, 在所述匹配程度包括所述全局特征匹配程度的情况下, 所述根据所述匹配程度, 确定所述对齐约束损失包括:

将所述点的特征、和所述像素的特征映射到同一维度, 分别获得每个点的第二点云特征和每个像素的第二像素特征;

随机获取第二数量的点以及相应的像素;

对于第二数量的点中的任意两个点, 确定所述两个点的第二点云特征之间的第一相似性信息、以及第二数量的像素中两个相应像素的第二像素特征之间的第二相似性信息, 并根据所述第一相似性信息和所述第二相似性信息之间的差距确定所述两个点对应的相似性损失;

根据所述三维点云数据中的点的相似性损失的总和, 确定所述对齐约束损失。

10. 根据权利要求 9 所述的训练方法，其中，所述第一相似性信息和第二相似性信息为相似度矩阵。

11. 根据权利要求 4~10 中任一项所述的训练方法，其中，所述语义匹配程度根据每个点的第一分类信息、以及每个像素的第二分类信息确定。

12. 根据权利要求 11 所述的训练方法，其中，所述第二分类信息通过利用二维图像处理模型对每个像素进行处理获得，并且在所述匹配程度包括所述语义匹配程度的情况下，所述根据所述匹配程度，确定所述对齐约束损失包括：

将所述三维点云数据映射到与所述二维图像数据相同的维度；

利用二维图像处理模型，对映射后的三维点云数据进行处理，获得每个点的第三分类信息；

利用所述第二分类信息和所述第三分类信息确定语义匹配损失，并根据所述语义匹配损失确定所述对齐约束损失。

13. 根据权利要求 12 所述的训练方法，其中，所述对齐约束损失与所述第二分类信息成正相关关系、与所述第三分类信息成负相关关系。

14. 根据权利要求 1~13 中任一项所述的训练方法，其中，所述获取二维图像数据的信息包括：

利用二维图像处理模型对所述二维图像数据进行处理，获得所述二维图像数据的信息。

15. 根据权利要求 14 所述的训练方法，其中，所述利用所述三维点云处理模型预测的、所述三维点云数据中点的分类信息以及预先标记的场景分类信息，确定所述三维点云处理模型的分类损失包括：

根据所述三维点云处理模型预测的、所述三维点云数据中点的分类信息以及预先标记的场景分类信息，确定第一分类损失；

根据所述二维图像处理模型预测的、所述二维图像数据中像素的分类信息以及所述场景分类信息，确定第二分类损失；

根据所述第一分类损失和所述第二分类损失，确定所述三维点云处理模型的分类损失。

16. 根据权利要求 15 所述的训练方法，其中：

根据所述三维点云处理模型预测的、所述三维点云数据的分类信息的池化处理结果以及所述场景分类信息，确定第一分类损失；根据所述二维图像处理模型预测的、所述二维图像数据中像素的分类信息以及所述场景分类信息，确定第二分类损失。

17. 根据权利要求 1~16 中任一项所述的训练方法，其中，利用所述对齐约束损失和所述分类损失的加权结果，训练所述三维点云处理模型。

18. 根据权利要求 1~17 中任一项所述的训练方法，其中，所述三维点云处理模型利用三维残差网络实现。

19. 根据权利要求 14~18 中任一项所述的训练方法，其中，所述二维图像处理模型利用二维残差网络实现。

20. 一种语义分割模型的训练方法，包括：

利用三维点云处理模型预测待处理的三维点云数据中的分类信息，其中，所述三维点云处理模型是利用权利要求 1~19 中任一项的三维点云处理模型的训练方法训练的；

根据所述分类信息生成所述待处理的三维点云数据中的点的伪标签；

利用所述伪标签，训练三维点云的语义分割模型。

21. 一种三维点云处理模型的训练装置，包括：

第一信息获取模块，被配置为利用三维点云处理模型对所述三维点云数据进行处理，以获取所述三维点云数据中点的信息；

第二信息获取模块，被配置为获取二维图像数据中像素的信息，其中，所述二维图像数据和所述三维点云数据覆盖相同的区域；

对齐约束损失确定模块，被配置为利用所述点的信息以及所述像素的信息，确定

所述三维点云数据和所述二维图像数据的对齐约束损失；

分类损失确定模块，被配置为利用所述三维点云处理模型预测的、所述三维点云数据中点的分类信息以及预先标记的场景分类信息，确定所述三维点云处理模型的分

类损失；

三维点云处理模型训练模块，被配置为利用所述对齐约束损失和所述分类损失，训练所述三维点云处理模型。

22. 一种三维点云处理模型的训练装置，包括：

存储器；以及

耦接至所述存储器的处理器，所述处理器被配置为基于存储在所述存储器中的指令，执行如权利要求 1~19 中任一项所述的三维点云处理模型的训练方法。

23. 一种语义分割模型的训练系统，包括：

权利要求 21 或 22 所述的三维点云处理模型的训练装置；

预测模块，被配置为利用三维点云处理模型预测待处理的三维点云数据中的分类信息；

生成模块，被配置为根据所述分类信息生成所述待处理的三维点云数据中的点的伪标签；

语义分割模型训练模块，被配置为利用所述伪标签，训练三维点云的语义分割模型。

24. 一种计算机可读存储介质，其上存储有计算机程序，该程序被处理器执行时实现权利要求 1~19 中任一项所述的三维点云处理模型的训练方法。

25. 一种计算机程序，包括：

指令，所述指令当由处理器执行时使所述处理器执行根据权利要求 1~19 中任一项所述的三维点云处理模型的训练方法。