

INSTANCE SEGMENTATION METHOD, APPARATUS AND DEVICE, STORAGE MEDIUM, AND PROGRAM PRODUCT

TECHNICAL FIELD

[0001] The disclosure relates, but is not limited, to a computer vision technology, and in particular, to an instance segmentation method, apparatus and device, a storage medium, and a program product.

BACKGROUND

[0002] At present, a deep learning based point cloud instance segmentation method is usually a strong supervision method. That is to say, a large number of used training samples require to be labeled with semantic labels and instance labels, and the large amount of labeling work about labels increases the labor cost of instance segmentation, affecting the efficiency of instance segmentation. Meanwhile, the adoption of a weak supervision method can reduce a labeling quantity and reduce the labor cost, which enhances the efficiency of instance segmentation to a certain extent, while affecting the accuracy of an instance segmentation result.

SUMMARY

[0003] Embodiments of the disclosure at least provide an instance segmentation method, apparatus and device, a storage medium, and a program product, which enhance the efficiency and accuracy of instance segmentation.

[0004] The technical solutions of the disclosure are implemented by means of the following operations.

[0005] In an aspect, an embodiment of the disclosure provides an instance segmentation method. The method includes the following operations.

[0006] Semantic prediction and centroid shift prediction are performed on acquired to-be-segmented point cloud by using a first backbone network, to obtain a semantic prediction result and centroid shift vector of each point in the to-be-segmented point cloud. The first backbone network and a second backbone network are trained based on first training point cloud, and the first training point cloud is determined by means of original semantic point cloud carrying a semantic label. Objectness prediction is performed on the to-be-segmented point cloud by using the second backbone network, to obtain multi-level point cloud including multi-level objectness labels. The multi-level objectness labels are configured to characterize a shape of each instance. Each point in the multi-level point cloud is moved to a respective predicted center according to the centroid shift vector, to obtain influence point cloud. Core points in the influence point cloud are clustered according to the semantic prediction result, to obtain at least one core point cloud of each semantic category. The core points are points having objectness labels at preset core levels. Edge points in the influence point cloud are clustered into target core point cloud in the at least one core point cloud of the corresponding semantic

category, to obtain an instance segmentation result. The edge points are points having objectness labels at preset edge levels.

- [0007] In this way, the training of the first backbone network and the second backbone network in an instance segmentation apparatus does not adopt labeled instance label. Therefore, the workload of instance labeling is reduced, and network training efficiency is enhanced, so that the efficiency of instance segmentation is enhanced. In addition, since the instance segmentation apparatus separates core points located in an instance core area and edge points located in an edge area, the accuracy of edge point clustering is enhanced. Therefore, the accuracy of instance segmentation can be enhanced.
- [0008] In another aspect, an embodiment of the disclosure provides an instance segmentation apparatus. The apparatus includes a first prediction module, a second prediction module, a moving module, a core clustering module, and an edge clustering module.
- [0009] The first prediction module is configured to perform semantic prediction and centroid shift prediction on acquired to-be-segmented point cloud by using a first backbone network, to obtain a semantic prediction result and centroid shift vector of each point in the to-be-segmented point cloud. The first backbone network and a second backbone network are trained based on first training point cloud, and the first training point cloud is determined by means of original semantic point cloud carrying a semantic label.
- [0010] The second prediction module is configured to perform objectness prediction on the to-be-segmented point cloud by using the second backbone network, to obtain multi-level point cloud including multi-level objectness labels. The multi-level objectness labels are configured to characterize a shape of each instance, internal information, and external information.
- [0011] The moving module is configured to move each point in the multi-level point cloud to a respective predicted center according to the centroid shift vector, to obtain influence point cloud.
- [0012] The core clustering module is configured to cluster core points of the influence point cloud according to the semantic prediction result, to obtain at least one core point cloud of each semantic category. The core points are points having objectness labels at preset core levels.
- [0013] The edge clustering module is configured to cluster edge points in the influence point cloud into target core point cloud in the at least one core point cloud of the corresponding semantic category, to obtain an instance segmentation result. The edge points are points having objectness labels at preset edge levels.
- [0014] In still another aspect, an embodiment of the disclosure provides an instance segmentation device. The device includes a memory and a processor.
- [0015] The memory is configured to store a computer program.
- [0016] The processor is configured to implement the instance segmentation method when executing the computer program stored in the memory.
- [0017] In yet another aspect, an embodiment of the disclosure provides a computer-readable storage medium. The computer-readable storage medium stores a computer

program. The instance segmentation method is implemented when the computer program is executed by a processor.

- [0018] In yet another aspect, an embodiment of the disclosure provides a computer program product. The computer program product includes a non-transitory computer-readable storage medium storing a computer program. When the computer program is read and executed by a computer, the instance segmentation method is implemented.
- [0019] It should be understood that, the above general description and the following detailed description are merely exemplary and explanatory, and do not limit the disclosure.

BRIEF DESCRIPTION OF THE DRAWINGS

- [0020] Drawings are incorporated into the specification and constitute a part of the specification. The drawings illustrate embodiments in accordance with the disclosure and serve to describe the technical solutions of the disclosure together with the specification.
- [0021] FIG. 1 is a flowchart of an optional instance segmentation method according to an embodiment of the disclosure.
- [0022] FIG. 2 is a schematic effect diagram of a center point regression according to an embodiment of the disclosure.
- [0023] FIG. 3 is a schematic diagram of optional multi-level prediction point cloud according to an embodiment of the disclosure.
- [0024] FIG. 4 is a flowchart of an optional instance segmentation method according to an embodiment of the disclosure.
- [0025] FIG. 5 is an optional core point influence diagram according to an embodiment of the disclosure.
- [0026] FIG. 6A is a schematic diagram of an optional clustering method in the related art according to an embodiment of the disclosure.
- [0027] FIG. 6B is a schematic diagram of an optional clustering effect according to an embodiment of the disclosure.
- [0028] FIG. 7 is a schematic diagram of an optional instance segmentation method according to an embodiment of the disclosure.
- [0029] FIG. 8 is a flowchart of an optional instance segmentation method according to an embodiment of the disclosure.
- [0030] FIG. 9 is a flowchart of an optional instance segmentation method according to an embodiment of the disclosure.
- [0031] FIG. 10 is a flowchart of an optional instance segmentation method according to an embodiment of the disclosure.
- [0032] FIG. 11A is a schematic diagram of optional multi-level training point cloud according to an embodiment of the disclosure.
- [0033] FIG. 11B is a schematic diagram of optional recombined training point cloud according to an embodiment of the disclosure.

- [0034] FIG. 11C is a schematic diagram of optional recombined virtual scene according to an embodiment of the disclosure.
- [0035] FIG. 12 is a schematic diagram of an optional recombination template according to an embodiment of the disclosure.
- [0036] FIG. 13 is a flowchart of an optional instance segmentation method according to an embodiment of the disclosure.
- [0037] FIG. 14 is a flowchart of an optional instance segmentation method according to an embodiment of the disclosure.
- [0038] FIG. 15 is a flowchart of an optional instance segmentation method according to an embodiment of the disclosure.
- [0039] FIG. 16 is a flowchart of an optional instance segmentation method according to an embodiment of the disclosure.
- [0040] FIG. 17 is a schematic diagram of a training process of an optional instance segmentation network according to an embodiment of the disclosure.
- [0041] FIG. 18A is a schematic diagram of optional initial pseudo-instance point cloud according to an embodiment of the disclosure.
- [0042] FIG. 18B is a schematic diagram of optional filtered first training point cloud according to an embodiment of the disclosure.
- [0043] FIG. 18C is an optional instance segmentation effect diagram according to an embodiment of the disclosure.
- [0044] FIG. 18D is a schematic diagram of optional multi-level prediction point cloud according to an embodiment of the disclosure.
- [0045] FIG. 18E is a schematic diagram of optional core point clustering according to an embodiment of the disclosure.
- [0046] FIG. 18F is a schematic diagram of an optional instance segmentation effect according to an embodiment of the disclosure.
- [0047] FIG. 19A is a schematic diagram of optional multi-level prediction point cloud according to an embodiment of the disclosure.
- [0048] FIG. 19B is a schematic diagram of an optional core point clustering effect according to an embodiment of the disclosure.
- [0049] FIG. 19C is a schematic diagram of an optional core point clustering effect according to an embodiment of the disclosure.
- [0050] FIG. 20 is a schematic structural diagram of an optional instance segmentation apparatus according to an embodiment of the disclosure.
- [0051] FIG. 21 is a schematic diagram of a hardware entity of an optional instance segmentation device according to an embodiment of the disclosure.

DETAILED DESCRIPTION

- [0052] For making the purposes, technical solutions and advantages of the disclosure

clearer, the technical solutions of the disclosure will further be described below in combination with the drawings and embodiments in detail. The described embodiments should not be considered as limits to the disclosure. All other embodiments obtained by those of ordinary skill in the art without creative work shall fall within the scope of protection of the disclosure.

- [0053] "Some embodiments" involved in the following descriptions describes a subset of all possible embodiments. However, it can be understood that "some embodiments" may be the same subset or different subsets of all the possible embodiments, and may be combined without conflicts.
- [0054] Unless otherwise defined, all technical and scientific terms in the specification have the same meaning as those skilled in the art, belonging to the disclosure, usually understand. Terms used in the specification are only used for describing the purpose of the disclosure, but not intended to limit the disclosure.
- [0055] In order to better understand an instance segmentation method provided in an embodiment of the disclosure, an instance segmentation solution in the related art is first described below.
- [0056] With the rapid development of a 3D acquisition technology, high-performance 3D scanners such as LiDAR lidar and RGB-D depth cameras greatly reduce the difficulty of acquiring point cloud data, and also bring more and more emerging applications that need to perceive interaction between a 3D environment and an object. As a key and basic perception task in 3D scene understanding, three-dimensional instance segmentation needs to simultaneously assign a specific semantic label and instance label to each point in point cloud, so as to distinguish different units, which is widely used in real scenes such as indoor robot navigation, augmented reality and automatic driving.
- [0057] Currently, most three-dimensional instance segmentation methods are strong supervision methods. For example, target detection is first performed to obtain a 3D detection box. Then, points with low instance confidence are eliminated from the 3D detection box. Alternatively, it is also possible to not obtain the 3D detection box, but to group feature points. The strong supervision methods are all required to perform joint prediction on semantics and instances, and the annotation of instance labeling takes a lot of labor time.
- [0058] Exemplarily, in a scene, there are 20 chairs, which belong to same semantics-chair and may be labeled at one time. However, instance labeling needs to mark different chairs one by one, that is, more labeling time is required to be taken.
- [0059] Later, a weak supervision method of three-dimensional segmentation is proposed, which reduces point cloud required to be labeled. For example, only a 3D boundary box is labeled or only one point is labeled for each instance. However, the performance of an instance segmentation network trained with the labeled point cloud is far behind that of the strong supervision method.
- [0060] An embodiment of the disclosure provides an instance segmentation method. The method may be executed by a processor of an instance segmentation device. The instance segmentation device may be a device having an instance segmentation capability, such as a server, a notebook computer, a tablet computer, a desktop computer, a smart television, a set-top box, or a mobile device (for example, a mobile phone, a

portable video player, a personal digital assistant, a special message device, and a portable game device). FIG. 1 is a schematic diagram of an implementation process of an instance segmentation method according to an embodiment of the disclosure. As shown in FIG. 1, the method includes S101 to S105.

[0061] At S101, semantic prediction and centroid shift prediction are performed on acquired to-be-segmented point cloud by using a first backbone network, to obtain a semantic prediction result and centroid shift vector of each point in the to-be-segmented point cloud. The first backbone network and a second backbone network are trained based on first training point cloud, and the first training point cloud is determined by means of original semantic point cloud carrying a semantic label.

[0062] In this embodiment of the disclosure, the to-be-segmented point cloud is 3D point cloud acquired from a scene by using an image acquisition apparatus or a 3D scanner, and each point in the 3D point cloud is in a three-dimensional coordinate system. Herein, a manner of acquiring the 3D point cloud from the scene may be designed according to requirements, which is not limited in this embodiment of the disclosure.

[0063] In this embodiment of the disclosure, the first backbone network may perform semantic prediction and centroid shift prediction on the to-be-segmented point cloud, to obtain the semantic prediction result and the centroid shift prediction vector. The semantic prediction result is the to-be-segmented point cloud carrying the predicted semantic label of each point. The centroid shift vector characterizes the shift vector between predicted instance centroid and a point in a corresponding instance.

[0064] In this embodiment of the disclosure, the instance segmentation apparatus may train an initial first backbone network based on the first training point cloud, to obtain the first backbone network. The first training point cloud includes a pseudo-instance label, and instance centroid of a pseudo-instance in the first training point cloud may act as true instance centroid. In this way, the instance segmentation apparatus may predict the centroid shift vector through learning.

[0065] In this embodiment of the disclosure, the second backbone network is also trained based on the first training point cloud. The instance segmentation apparatus may perform labeling with multi-level objectness labels based on the first training point cloud, to obtain multi-level training point cloud. The multi-level training point cloud includes the multi-level objectness labels. The multi-level objectness labels are configured to characterize a shape of each instance in the to-be-segmented point cloud. An objectness-level label of each point in the multi-level training point cloud is configured to characterize whether the point belongs a core area of the corresponding instance or an edge area. An initial second backbone network is trained based on the multi-level training point cloud, to obtain the second backbone network.

[0066] It is to be noted that, the quantity of a plurality of levels may be set according to requirements, which is not limited in this embodiment of the disclosure. The plurality of levels includes a core level and an edge level.

[0067] In this embodiment of the disclosure, the first backbone network and the second backbone network may be a same network, or may be different networks. Herein, the first backbone network or the second backbone network may be a Convolutional Neural Network (CNN), such as a submanifold sparse convolution based 3D U-net, a

Point-Voxel CNN (PVCNN), or a persistent homology based graph convolution network, and this embodiment of the disclosure is not limited thereto.

[0068] At S102, objectness prediction is performed on the to-be-segmented point cloud by using the second backbone network, to obtain multi-level point cloud including multi-level objectness labels. The multi-level objectness labels are configured to characterize a shape of each instance, internal information, and external information.

[0069] In this embodiment of the disclosure, the instance segmentation apparatus may input the to-be-segmented point cloud into the second backbone network. The second backbone network may perform multi-level objectness prediction on the to-be-segmented point cloud, to obtain the multi-level point cloud. Each point in the multi-level point cloud includes objectness labels of the corresponding levels. The multi-level objectness labels may characterize the shape of each instance predicted by the second backbone network for the to-be-segmented point cloud, and distinguish other internal and external information of each instance, that is, the internal information of each instance and the external information of each instance.

[0070] It is to be noted that, the second backbone network is trained by the multi-level training point cloud carrying the multi-level objectness labels. The quantity of the plurality of levels of the objectness labels in the multi-level point cloud is the same as the quantity of the plurality of levels of the objectness labels in the multi-level training point cloud, and the label of each level is also the same. In this way, after the multi-level point cloud is obtained, the instance segmentation apparatus may distinguish a core point in the core area of each instance and an edge point of the edge area by means of the multi-level point cloud. Therefore, each instance is distinguished with other instances, and the boundary of each instance is perceived simultaneously.

[0071] At S103, each point in the multi-level point cloud is moved to a respective predicted center according to the centroid shift vector, to obtain influence point cloud.

[0072] In this embodiment of the disclosure, after the centroid shift vector is obtained, the instance segmentation apparatus may move each point in the multi-level point cloud to the respective predicted center by adopting a manner of center point regression, to obtain the influence point cloud. That is to say, the influence point cloud is the multi-level point cloud after each point is moved, and includes the multi-level objectness labels.

[0073] In this embodiment of the disclosure, the instance segmentation apparatus may use each point in the multi-level point cloud to add the corresponding centroid shift vector, so as to obtain each influence point in the influence point cloud. In this way, the instance segmentation apparatus may move the points in the multi-level point cloud to the predicted center, to obtain the influence point cloud corresponding to the multi-level point cloud, referring to a formula (1):

$$x_m' = x_m + d_m \quad \text{Formula (1).}$$

The multi-level point cloud includes M points; m is a positive integer, and $m \in M$; x_m is any point in the multi-level point cloud; d_m is the centroid shift vector corresponding to x_m ; and x_m' is the influence point corresponding to x_m .

- [0074] Exemplarily, as shown in FIG. 2, after point cloud A, point cloud B, point cloud C, point cloud D, and point cloud E are moved to the respective predicted centers, point cloud A', point cloud B', point cloud C', point cloud D', and point cloud E' may be obtained. It can be seen from the figure that, through moving, the points in the point cloud A' to E' located in the edge area are more marginalized than that in A to E. The points in the point cloud A' to E' located in the core area are more concentrated towards the core area than that in A to C.
- [0075] At S104, core points in the influence point cloud are clustered according to the semantic prediction result, to obtain at least one core point cloud of each semantic category. The core points are points having objectness labels at preset core levels.
- [0076] In this embodiment of the disclosure, the instance segmentation apparatus may cluster the core points in the influence point cloud according to the semantic category of each point in the semantic prediction result, to obtain the at least one core point cloud of each semantic category. Herein, the at least one core point cloud of each semantic category corresponds to at least one instance of each semantic category. The core points in the influence point cloud are labels that are at a preset core level in the multi-level objectness labels. The plurality of levels includes the preset core level and a preset edge level.
- [0077] Exemplarily, there are 5 multi-level objectness labels, which are 0, 1, 2, 3 and 4, respectively. Here, 3 and 4 are objectness labels of the preset core level, and 0 to 2 are objectness labels of the preset edge level.
- [0078] It is to be noted that, the instance segmentation apparatus may cluster the core points by using a clustering algorithm. Herein, the clustering algorithm may be a Breadth-First Search (BFS) algorithm, may be a PageRank algorithm, or may be a graphics algorithm such as a closeness centrality algorithm, and alternatively, Density-Based Spatial Clustering of Applications with Noise (DBSCAN). The clustering algorithm may be designed according to requirements, and this embodiment of the disclosure is not limited thereto.
- [0079] At S105, edge points in the influence point cloud are clustered into target core point cloud of the corresponding semantic category, to obtain an instance segmentation result. The edge points are points having objectness labels at preset edge levels.
- [0080] In this embodiment of the disclosure, after the at least one core point cloud of each semantic category is obtained, the instance segmentation apparatus may cluster each edge point into the target core point cloud in the at least one core point cloud of the corresponding semantic category, so that instance segmentation is completed, and the instance segmentation result is obtained. Each edge point and the corresponding target core point cloud belong to the same semantic category in the semantic prediction result.
- [0081] In this embodiment of the disclosure, the instance segmentation apparatus may calculate an influence value between the edge point and each core point cloud of the corresponding semantic category according to the distribution of the influence point cloud. In this way, the instance segmentation apparatus may calculate an influence value between each edge point in the influence point cloud and the at least one core point cloud of the same semantic category, and cluster the edge points in the multi-level point cloud into the target core point cloud according to the influence value.

- [0082] In this embodiment of the disclosure, the instance segmentation apparatus may use the core point cloud with the maximum influence value between the at least one core point cloud of the same semantic category and the corresponding edge point as the target core point cloud. By clustering each edge point into the target core point cloud, the classification of the point cloud is completed, that is, instance segmentation.
- [0083] In this embodiment of the disclosure, the influence value between the edge point and the at least one core point cloud of the same semantic category is inversely correlated to a distance between the edge point and the at least one core point cloud.
- [0084] In some embodiments of the disclosure, the influence value between the edge point and the core point cloud of the same semantic category may be an influence value between the edge point and a center point of the core point cloud.
- [0085] In some embodiments of the disclosure, the influence value between the edge point and the core point cloud of the same semantic category may be a mean value of the influence values between the edge points and all of the point of the core point cloud.
- [0086] Herein, the influence value between the edge point and the core point cloud may be set according to requirements, and this embodiment of the disclosure is not limited thereto.
- [0087] It is to be noted that, such a method of distinguishing the core point and the edge point and clustering the core point and the edge point by using different strategies to complete instance segmentation may be called an asymmetric object reasoning algorithm.
- [0088] Exemplarily, FIG. 3 is an influence diagram. As shown in FIG. 3, the influence diagram includes points of three levels of the objectness labels, which are a high objectness point, a medium objectness point, and a low objectness point. The high objectness point is located at the core of the instance, the medium objectness point is located at the secondary core of the instance, and the low objectness point is located on the edge of the instance. The high objectness point and the medium objectness point may be used as core points for clustering first, and then the low objectness point is clustered.
- [0089] It may be understood that, the instance segmentation apparatus includes the first backbone network and the second backbone network. Since the first backbone network and the second backbone network are trained based on the first training point cloud, and the first training point cloud is determined based on semantic standard point cloud, that is to say, the training of the first backbone network and the second backbone network in the instance segmentation apparatus does not adopt labeled instance label, the workload of instance labeling is reduced, and network training efficiency is enhanced, so that the efficiency of instance segmentation is enhanced. In addition, since the instance segmentation apparatus performs clustering by separating the core point located in the instance core area and the edge point located in the edge area, the accuracy of edge point clustering is enhanced. Therefore, the accuracy of instance segmentation can be enhanced.
- [0090] In some embodiments of the disclosure, as shown in FIG. 4, the implementation of S105 of clustering the edge points in the influence point cloud into the target core point cloud of the corresponding semantic category, to obtain the instance segmentation result may include S201 to S203.
- [0091] At S201, an influence value between each edge point and the at least one core

point cloud of the corresponding semantic category is calculated, to obtain at least one corresponding influence value of each edge point. The influence value is configured to characterize a mean value of the influence values between the edge point and all of the points in the core point cloud of the corresponding semantic category.

[0092] In this embodiment of the disclosure, after cores in the influence point cloud are clustered to obtain the at least one core point cloud of each semantic category, the instance segmentation apparatus may calculate at least one influence value between the edge point in the influence point cloud and the at least one core point cloud of the same semantic category, to obtain the at least one influence value of each edge point.

[0093] In this embodiment of the disclosure, the instance segmentation apparatus may calculate the influence value between each edge point and each core point in the corresponding core point cloud, to obtain each influence value, and then average each influence value to obtain the influence value between each edge point and the corresponding core point cloud, so that the at least one influence value between each edge point and the at least one core point cloud of the corresponding semantic category is obtained.

[0094] In this embodiment of the disclosure, the influence value between any edge point j and the core point cloud A may be calculated by using a formula (2):

$$I_j^A = \sum_{i \in P_A} \frac{w_{ij}}{|P_A|} \quad \text{Formula (2);}$$

The core point cloud A includes P_A points, and any point in the core point cloud A is i, and $i \in P_A$; and w_{ij} represents the influence value between the edge point j and the i-th point in the core point cloud A, which may be calculated by referring to a formula (3). That is to say, in a case that the core point cloud belonging to the same semantic category with the edge point j includes A, B,, K in total, the instance segmentation apparatus may obtain K influence values of the edge point j, representing as $I_j = \{I_j^A, I_j^B, \dots, I_j^K\}$.

$$w_{ij} = \begin{cases} 0, (i, j) \in P_{core} \\ e^{-\left(\frac{\|x_i - x_j\|^2}{2\varepsilon^2}\right)}, otherwise \end{cases} \quad \text{Formula (3)}$$

P_{core} represents the core point cloud, and ε is a hyper-parameter; ε may be set according to requirements; and the larger the ε , the larger the influence range of a point, and the more points influenced. It can be seen that, for any two points in the core point cloud, the influence value between the two points is not considered, that is, the influence value between the two points is 0. The larger the distance between the edge point j and the i-th point in the core point cloud A, the smaller the influence value between the edge point and the point in the core point cloud.

[0095] In this embodiment of the disclosure, in order to promote the core point cloud to absorb unclustered edge points, the instance segmentation apparatus needs to calculate the influence value between the core point cloud and other points in the same semantic

category for each core point cloud.

[0096] Exemplarily, FIG. 5 is a core point influence diagram. There are 4 influence values between the core point cloud A and the points e1 to e4, which are expressed by using 4 arrows. The closer the point is to the core point cloud A, the thicker the arrow, indicating the larger the influence. That is to say, in the points e1 to e4, the influence between the core point cloud A and the point e1 is the largest. There are 5 influence values between the core point cloud B and the points e5 to e9, which are expressed by using 5 arrows. The closer the point is to the core point cloud B, the thicker the arrow, indicating the larger the influence. That is to say, in the points e5 to e9, the influence between the core point cloud B and the point e9 is the largest.

[0097] At S202, core point cloud corresponding to a maximum value in the at least one influence value of each edge point is used as the target core point cloud corresponding to each edge point.

[0098] In this embodiment of the disclosure, after obtaining the at least one influence value of each edge point, the instance segmentation apparatus may use the influence core point cloud corresponding to the maximum value in the at least one influence value of each edge point as the target core point cloud. For example, in K influence values of the edge point j, the influence value of the influence core point cloud C is the largest, indicating that the target core point cloud of the edge point j is C.

[0099] At S203, each edge point is clustered into the corresponding target core point cloud, to obtain at least one instance segmentation point cloud of each semantic category, so as to obtain the instance segmentation result.

[00100] In this embodiment of the disclosure, after the target core point cloud of each edge point is determined, the instance segmentation apparatus may cluster each edge point into the corresponding target core point cloud, to obtain at least one instance segmentation point cloud of each semantic category, so as to obtain the instance segmentation result.

[00101] Exemplarily, FIG. 6A is a schematic diagram of a clustering effect in the related art. As shown in FIG. 6A, it can be seen that, under-segmentation and over-segmentation may be caused by only using a manner of radius search for clustering in the related art. That is to say, different instances A and B are not segmented, and points belonging to the instance B are segmented, but not clustered into the instance B. FIG. 6B is a schematic diagram of an optional clustering effect according to an embodiment of the disclosure. As shown in FIG. 6B, it can be seen that, during clustering, since the core point and the edge point are distinguished, the core point cloud a and b are first clustered, then the edge point a1 beside a and the edge point b1 beside b are clustered, and the instance A and the instance B are finally obtained.

[00102] In some embodiments of the disclosure, the asymmetric object reasoning algorithm may be embodied by using the following algorithm.

Input: scene point cloud $X \in R^{N \times 3}$; semantic label $S \in R^{N \times 1}$; centroid shift vector $D \in R^{N \times 3}$; multi-level objectness label $O \in R^{N \times 1}$.

Output: pseudo-instance label C

1: for $id \in \text{unique}(S)$ do

2: if $id \in \text{foreground semantic IDs}$ then

3: $N_{fg} \leftarrow |P_{fg}|$

4: for $i=1$ to N_{fg} do

5: if $O_{fg} \geq \hat{O}$ then

6: $P_{core} \leftarrow P_i$

7: else

8: $P_{boundary} \leftarrow P_i$

9: $X'_{fg} = X_{fg} + D_{fg}$

10: $C_{core}^K \leftarrow P_{core}$

11: for $k=1$ to K do

12: compute influence map I^k

13: $I \leftarrow I^k$

14: for $P_j \in P_{boundary}$ do

15: compare influence levels among I_j

16: $k \leftarrow \arg \max(I_j)$

17: $C_{core}^K \leftarrow P_j$

18: $C_{id} \leftarrow C_{core}$

19: return C .

[00103] It can be seen from steps 1 to 4, the instance segmentation apparatus only processes the point N_{fg} belonging to a foreground category in the scene point cloud, ignoring a background category. By using an indoor scene as an example, the background category is, for example, a wall or the floor. The foreground category is, for example, an indoor table, a chair, or a cabinet. In step 5, $\hat{O}$ is the highest level in the edge level. For example, $\hat{O}$ is 1, and the level of the core point is 2 to 4. That is to say, steps 5 to 8 are to distinguish the edge point and the core point. Step 9 is to move the point cloud to the predicted center of the corresponding instance. Step 10 is to cluster the core points as K core point cloud by using the instance as a unit. Steps 11 to 16 are to calculate influence diagrams, and determine the core point cloud having the largest influence on the edge point according to the influence diagrams. Step 17 is to cluster the edge point P_j into the

core point cloud C_{core}^K with the largest influence. Step 18 obtains an identifier ID of each instance point cloud. Step 19 returns the instance label to complete instance segmentation.

- [00104] It may be understood that, the instance segmentation apparatus may cluster the edge points into the core point cloud with the largest influence value according to each edge point in the influence point cloud and the entire core point cloud. In this way the accuracy of edge point clustering is enhanced, thereby enhancing the accuracy of instance segmentation.
- [00105] In some embodiments of the disclosure, as shown in FIG. 7, the implementation before S101 of performing semantic prediction and centroid shift prediction on the acquired to-be-segmented point cloud by using the first backbone network, to obtain the semantic prediction result and centroid shift vector of each point in the to-be-segmented point cloud, the method may include S301 to S304.
- [00106] At S301, original semantic point cloud is clustered to obtain initial pseudo-instance point cloud.
- [00107] In this embodiment of the disclosure, after the original semantic point cloud is acquired, the instance segmentation apparatus may cluster the original semantic point cloud to obtain the initial pseudo-instance point cloud. The original semantic point cloud is the point cloud carrying semantic labels.
- [00108] In this embodiment of the disclosure, the instance segmentation apparatus may cluster the original semantic point cloud by using a clustering algorithm according to the semantic labels, to obtain the initial pseudo-instance point cloud. The initial pseudo-instance point cloud includes clustered pseudo-instance labels.
- [00109] In some embodiments of the disclosure, the clustering algorithm may be a BFS algorithm. That is to say, the instance segmentation apparatus may use any point in each semantic category in the original semantic point cloud as a seed point. The seed point is used as a center, and r is a search radius. The points of the same semantic category in the original semantic point cloud are searched one by one, and the points within the radius r are clustered. Then the points that are clustered together with the seed points are used as new seed points. Searching with r being the radius is continuously performed, until all of the points in the original semantic point cloud are searched, so that initial instance segmentation of the original semantic point cloud is completed, so as to obtain the initial pseudo-instance point cloud.
- [00110] At S302, the initial pseudo-instance point cloud is filtered according to the quantity of points included in each pseudo-instance in the initial pseudo-instance point cloud, to obtain the first training point cloud.
- [00111] In this embodiment of the disclosure, after the initial pseudo-instance point cloud is obtained, the instance segmentation apparatus may filter the initial pseudo-instance point cloud is filtered according to the quantity of points included in each pseudo-instance in the initial pseudo-instance point cloud, to obtain the first training point cloud.
- [00112] In this embodiment of the disclosure, the pseudo-instance with excessive points may be an under-segmented instance, that is, a plurality of instances are clustered

together to become one pseudo-instance. The pseudo-instance with less points may be an over-segmented instance, that is, parts of the points belonging to one instance. That is to say, excessive and less points included in the initial pseudo-instance point cloud may be pseudo-instances with low confidence. Therefore, the instance segmentation apparatus needs to filter the pseudo-instance point cloud with excessive or less points, to obtain the first training point cloud with high confidence. Herein, the first training point cloud may include a plurality of instance point cloud with high confidence. For example, the first training point cloud may include the point cloud of a chair 1, a chair 2, a table, a bed, a lamp 1, and a lamp 2.

- [00113] It may be understood that, since the pseudo-instance point cloud with low confidence is filtered from the initial pseudo-instance point cloud, the first training point cloud with high confidence is obtained, so that the accuracy of instance labels in the first training point cloud is enhanced. Therefore, the accuracy of the first backbone network and the second backbone network is enhanced.
- [00114] In some embodiments of the disclosure, as shown in FIG. 8, the implementation of S302 of filtering the initial pseudo-instance point cloud according to the quantity of points included in each pseudo-instance in the initial pseudo-instance point cloud, to obtain the first training point cloud may include S401 to S402.
- [00115] At S401, the quantity of the points in each pseudo-instance point cloud of each semantic category is sorted, to obtain a quantity sequence of each semantic category.
- [00116] In this embodiment of the disclosure, after obtaining the initial pseudo-instance point cloud, the instance segmentation apparatus may sort the quantity of the points in each pseudo-instance point cloud of each semantic category in the initial pseudo-instance point cloud according to the semantic labels in the original semantic point cloud, to obtain the quantity sequence of each semantic category.
- [00117] In some embodiments of the disclosure, the instance segmentation apparatus may sort the quantity of the points in each pseudo-instance point cloud of each semantic category in the initial pseudo-instance point cloud in an ascending order, to obtain the ascending quantity sequence of each semantic category, or may sort the quantity of the points in each pseudo-instance point cloud of each semantic category in the initial pseudo-instance point cloud in a descending order, to obtain the descending quantity sequence of each semantic category. A sorting manner may be designed according to requirement, and this embodiment of the disclosure is not limited thereto.
- [00118] At S402, pseudo-instance point cloud within a preset quantity sequence range is selected from the quantity sequence of each semantic category, to obtain training point cloud of each semantic category. The first training point cloud includes the training point cloud of each semantic category.
- [00119] In this embodiment of the disclosure, after the quantity sequence of each semantic category is obtained, the instance segmentation apparatus may select the pseudo-instance point cloud within the preset quantity sequence range from the quantity sequence of each semantic category, to obtain the training point cloud of each semantic category, so that the first training point cloud is obtained.
- [00120] In some embodiments of the disclosure, the preset quantity sequence range may be an ordinal range in the quantity sequence.

- [00121] Exemplarily, the preset quantity sequence range may be an ordinal percentage range occupied in the quantity sequence by using a quantity center value as a center. For example, the pseudo-instance point cloud belonging to the semantic category 1 in the initial pseudo-instance point cloud includes 5. The quantity of the points in the pseudo-instance point cloud 1 is 17289, the quantity of the points in the pseudo-instance point cloud 2 is 3526, the quantity of the points in the pseudo-instance point cloud 3 is 5564, the quantity of the points in the pseudo-instance point cloud 4 is 4891, the quantity of the points in the pseudo-instance point cloud 5 is 15, and the quantity of the points in the pseudo-instance point cloud 6 is 32. The instance segmentation apparatus sorts the quantity of the points in the pseudo-instance point cloud 1 in the ascending order, to obtain the quantity sequence being 15, 32, 3526, 4891, 5564, and 17289. The preset quantity sequence range is 33%. That is to say, the first training point cloud may include the pseudo-instance point cloud 2 and the pseudo-instance point cloud 3.
- [00122] It may be understood that, by means of the quantity sequence and the preset quantity sequence range, the pseudo-instance point cloud of which the quantity of the points is close to a center value may be selected from the initial pseudo-instance point cloud, so that the point cloud with too high or too low points is filtered. Therefore, the initial pseudo-instance point cloud may be rapidly filtered, to obtain the first training point cloud.
- [00123] At S303, an initial second backbone network is trained based on the first training point cloud, to obtain the second backbone network.
- [00124] In this embodiment of the disclosure, after the first training point cloud is obtained, the instance segmentation apparatus may train the initial second backbone network based on the first training point cloud. After training is completed, the second backbone network may be obtained.
- [00125] In some embodiments of the disclosure, the instance segmentation apparatus may recombine the pseudo-instance point cloud in the first training point cloud, to obtain recombined training point cloud. The recombined training point cloud may characterize scenes that are randomly generated by a plurality of sets of pseudo-instance point cloud with high confidence. The initial second backbone network is trained based on the recombined training point cloud to obtain the second backbone network.
- [00126] In some embodiments of the disclosure, as shown in FIG. 9, the implementation of S203 of training the initial second backbone network based on the first training point cloud, to obtain the second backbone network may include S501 to S504.
- [00127] At S501, point cloud compression is performed on the first training point cloud, to obtain first compressed point cloud.
- [00128] In this embodiment of the disclosure, the instance segmentation apparatus may reduce a coordinate of each point in the first training point cloud according to a same proportion, to implement point cloud compression, so as to obtain the first compressed point cloud. Herein, the same proportion is preset, for example, 3:1, 4:1, and 2:1. A proportional value of the same proportion may be set according to requirements, and this embodiment of the disclosure is not limited thereto.
- [00129] In some embodiments of the disclosure, the same proportion is 2:1. The instance segmentation apparatus may divide the coordinate of each point in the first

training point cloud by 2, to obtain the coordinate of each point in the first compressed point cloud, so that the first compressed point cloud is obtained. The first compressed point cloud includes a plurality of compressed pseudo-instance point cloud. The plurality of compressed pseudo-instance point cloud in the first compressed point cloud are in one-to-one correspondence with a plurality of pseudo-instance point cloud in the first training point cloud. Compared with the pseudo-instance point cloud corresponding to the first training point cloud, a shape of each pseudo-instance point cloud in the first compressed point cloud remains unchanged, but the volume is reduced.

[00130] At S502, the multi-level training point cloud corresponding to each pseudo-instance point cloud is determined according to a compressed distance between each point in each pseudo-instance point cloud in the first training point cloud and corresponding compressed pseudo-instance point cloud in the first compressed point cloud. The multi-level training point cloud includes the multi-level objectness labels.

[00131] In this embodiment of the disclosure, after the first compressed point cloud is obtained, the instance segmentation apparatus may calculate a distance between each point in each pseudo-instance point cloud and the corresponding compressed pseudo-instance point cloud, to obtain the compressed distance between each point in each pseudo-instance point cloud and the corresponding compressed pseudo-instance point cloud. The quantity of a plurality of compressed distances of each pseudo-instance point cloud is the same as the quantity of the points in each pseudo-instance point cloud.

[00132] It may be understood that, since the compressed distance between each point in each pseudo-instance point cloud and the compressed pseudo-instance point cloud includes shape information of each pseudo-instance point cloud, in a case that the multi-level training point cloud is configured to train the second backbone network, the second backbone network may better perceive the shape of the pseudo-instance, thereby enhancing the accuracy of the second backbone network for predicting the objectness labels.

[00133] In some embodiments of the disclosure, the instance segmentation apparatus may calculate the distance between each point in each pseudo-instance point cloud and all of the points in the corresponding compressed pseudo-instance point cloud, to obtain a plurality of distances of each point, and then use the mean value of the plurality of distances as the compressed distance between each point in each pseudo-instance point cloud and the corresponding compressed pseudo-instance point cloud.

[00134] In some embodiments of the disclosure, as shown in FIG. 10, the implementation of S502 of determining the multi-level training point cloud corresponding to each pseudo-instance point cloud according to the compressed distance between each point in each pseudo-instance point cloud in the first training point cloud and the corresponding compressed pseudo-instance point cloud in the first compressed point cloud may include S601 to S603.

[00135] At S601, a distance between each point in each pseudo-instance point cloud and each point in the corresponding compressed pseudo-instance point cloud is calculated, to obtain a plurality of distances corresponding to each point in each pseudo-instance point cloud.

[00136] In this embodiment of the disclosure, the instance segmentation apparatus may calculate the distance between each point in each pseudo-instance point cloud and all of

the points in the corresponding compressed pseudo-instance point cloud, to obtain the plurality of distances corresponding to each point in each pseudo-instance point cloud. The quantity of the plurality of distances corresponding to each point in each pseudo-instance point cloud is the same as the quantity of the points included in each pseudo-instance point cloud.

[00137] At S602, a minimum distance in the plurality of distances corresponding to each point in each pseudo-instance point cloud is used as the compressed distance between each point in each pseudo-instance point cloud and corresponding compressed point cloud, to obtain a plurality of compressed distances of each pseudo-instance point cloud.

[00138] In this embodiment of the disclosure, after the plurality of distances corresponding to each point in each pseudo-instance point cloud is obtained, the instance segmentation apparatus may determine the minimum distance from the plurality of distances corresponding to each point in each pseudo-instance point cloud as the compressed distance between each point in each pseudo-instance point cloud and the corresponding compressed point cloud, so that the plurality of compressed distances corresponding to the plurality of points in each pseudo-instance point cloud.

[00139] At S603, the multi-level objectness labels of each pseudo-instance point cloud are determined according to the plurality of compressed distances of each pseudo-instance point cloud, to obtain the multi-level training point cloud corresponding to each pseudo-instance point cloud.

[00140] In this embodiment of the disclosure, the instance segmentation apparatus may determine the multi-level objectness labels of each pseudo-instance point cloud according to the plurality of compressed distances of each pseudo-instance point cloud, so that the multi-level training point cloud corresponding to each pseudo-instance point cloud is obtained.

[00141] In this embodiment of the disclosure, the instance segmentation apparatus may sort the plurality of compressed distances of each pseudo-instance point cloud to obtain a distance sequence. The distance sequence is classified according to a plurality of preset distance sequence ranges corresponding to the plurality of levels, to obtain the multi-level objectness labels. A sum of the plurality of preset distance sequence ranges is 1.

[00142] Exemplarily, the pseudo-instance point cloud P_x in the first training point cloud includes S points. A coordinate X of the point cloud P_x includes coordinates of S points, which is recorded as $X = \{x_1, x_2, \dots, x_S\} \in R^{S \times 3}$. The instance segmentation apparatus may compress P_x to obtain the compressed point cloud P_x of the point cloud P_x . A coordinate x of the compressed point cloud P_x also includes coordinates of S points, which is recorded as $x = \left\{ \frac{x_1}{2}, \frac{x_2}{2}, \dots, \frac{x_S}{2} \right\} \in R^{S \times 3}$. Then, a closest distance corresponding to each point in the point cloud P_x is found from the distance between each point in P_x and each point in the compressed point cloud P_x as the compressed

distance between the S points in the point cloud P_x and the compressed point cloud P_x , to obtain a compressed distance list T. the compressed distance list T includes the distances between the S points in the point cloud and the compressed point cloud P_x , which is recorded as $T = \{T_1, T_2, \dots, T_S\} \in R^{S \times 1}$. The objectness labels values $label_{ob-ID}$ of the plurality of levels and the plurality of preset distance sequence ranges are shown in a formula (4). The objectness labels values $label_{ob-ID}$ being 2 to 4 may be at the preset core level, and the objectness labels values $label_{ob-ID}$ being 0 to 1 may be at the preset edge level.

$$label_{ob-ID} = \begin{cases} 4, 0 \sim 20\% \text{ of } T \\ 3, 20 \sim 40\% \text{ of } T \\ 2, 40 \sim 60\% \text{ of } T \\ 1, 60 \sim 80\% \text{ of } T \\ 0, 80 \sim 100\% \text{ of } T \end{cases} \quad \text{Formula (4).}$$

- [00143] In this embodiment of the disclosure, the quantity of the multi-level objectness labels may be set according to requirements. The plurality of preset distance sequence ranges corresponding to the multi-level objectness labels may also be set according to requirements. The plurality of preset distance sequence ranges may be the same or different. The quantity, identifier and the corresponding preset distance sequence range of the multi-level objectness labels are not limited in this embodiment of the disclosure.
- [00144] It may be understood that, by means of the sorting of the plurality of compressed distances and the plurality of preset distance sequence ranges, the multi-level objectness labels may be rapidly determined, so that the multi-level training point cloud corresponding to each pseudo-instance point cloud is rapidly obtained and determined.
- [00145] At S503, point cloud recombination is performed based on the multi-level training point cloud corresponding to each pseudo-instance point cloud, to obtain recombined training point cloud.
- [00146] In this embodiment of the disclosure, after the multi-level training point cloud corresponding to each pseudo-instance point cloud is obtained, the instance segmentation apparatus may perform point cloud recombination based on the multi-level training point cloud corresponding to each pseudo-instance point cloud, to obtain the recombined training point cloud.
- [00147] In some embodiments of the disclosure, the instance segmentation apparatus may select at least one multi-level training point cloud from the multi-level training point cloud corresponding to all of the pseudo-instances, and perform spatial rotation and spatial position transformation on the at least one multi-level training point cloud, to obtain the recombined training point cloud.
- [00148] Exemplarily, FIG. 11A shows 4 multi-level training point cloud corresponding to 4 pseudo-instance point cloud. FIG. 11B shows that 5 multi-level training point cloud corresponding to 5 pseudo-instance point cloud are recombined to obtain the recombined training point cloud. A virtual scene corresponding to the recombined training point cloud

in FIG. 11B is shown in FIG. 11C.

- [00149] In some embodiments of the disclosure, the implementation of S503 of performing point cloud recombination based on the multi-level training point cloud corresponding to each pseudo-instance point cloud, to obtain the recombined training point cloud may include the following steps.
- [00150] At S5031, point cloud recombination is performed on the multi-level training point cloud according to a preset point cloud attribute in a recombination template, to obtain the recombined training point cloud. The preset point cloud attribute includes a preset point cloud quantity, a preset point cloud volume, and a preset point cloud position. There is no intersection area between the preset point cloud positions.
- [00151] In this embodiment of the disclosure, the preset point cloud attribute in the recombination template may include the preset point cloud quantity, the preset point cloud position, and the preset point cloud volume corresponding to the preset point cloud position. The preset point cloud quantity is configured to represent the quantity of the multi-level training point cloud included in the recombined training point cloud. The preset point cloud position is configured to represent a position corresponding to each point cloud in the preset point cloud quantity. The preset point cloud volume is configured to represent the volume of the multi-level training point cloud on each preset point cloud position. In some embodiments, in a case that a point cloud density is uniform at a certain value, the point cloud volume may be characterized by using the quantity of the points in the point cloud.
- [00152] It is to be noted that, each preset point cloud position in the recombination template does not intersect with each other spatially.
- [00153] Exemplarily, FIG. 12 shows an optional recombination template. The quantity of preset point cloud included in the recombination template is 9, corresponding to 9 preset point cloud positions: 0 to 8. The 0-th position is a center position. The preset point cloud volume on each point cloud position is different. The instance segmentation apparatus may acquire 9 pieces of the multi-level training point cloud from the first training point cloud according to 9 different point cloud volumes. The 9 multi-level training point cloud are respectively placed on the corresponding 9 preset point cloud positions, to obtain the recombined training point cloud.
- [00154] It may be understood that, the instance segmentation apparatus may perform point cloud recombination based on the recombination template, thereby enhancing the efficiency of point cloud recombination.
- [00155] At S504, the initial second backbone network is trained by using the recombined training point cloud, to obtain the second backbone network and the multi-level prediction point cloud outputted by the second backbone network.
- [00156] In this embodiment of the disclosure, after the recombined training point cloud is obtained, the instance segmentation apparatus may use the multi-level objectness labels in the recombined training point cloud as supervision data, use the recombined training point cloud with the multi-level objectness labels removed as input data, and train the initial second backbone network to obtain the second backbone network.
- [00157] It may be understood that, since a large number of virtual point cloud scenes may be randomly generated by means of point cloud recombination, that is, the

recombined training point cloud may include a large number of training data, the precision of the second backbone network may be improved.

- [00158] In some embodiments of the disclosure, the implementation of S504 of training the initial second backbone network by using the recombined training point cloud, to obtain the second backbone network may include S701 to S705.
- [00159] At S701, the second backbone network of a (k-1)th update is acquired, where k is a positive integer, and in a case that k equals to 1, the second backbone network of the (k-1)th update is the initial second backbone network.
- [00160] In this embodiment of the disclosure, the instance segmentation apparatus updates the initial second backbone network for a plurality of times. The second backbone network of the (k-1)th update represents that the initial second backbone network is updated for k-1 times. In the case that k is 1, the second backbone network of the (k-1)th update represents that the initial second backbone network is updated for 0 times, that is, the initial second backbone network is not updated. That is to say, in the case that k is 1, the second backbone network of the (k-1)th update is the initial second backbone network.
- [00161] At S702, feature extraction is performed on each point of the recombined training point cloud by means of the second backbone network of the (k-1)th update, to obtain a k-th second point cloud feature of each point.
- [00162] In this embodiment of the disclosure, after acquiring the second backbone network of the (k-1)th update, the instance segmentation apparatus may perform feature extraction on each point of the recombined training point cloud by means of the second backbone network of the (k-1)th update, to obtain the k-th second point cloud feature of each point.
- [00163] At S703, second linear transformation is performed on the k-th second point cloud feature of each point, to obtain a k-th second linear transformation feature. A dimension of the k-th second linear transformation feature is in one-to-one correspondence with a level quantity of the multi-level objectness labels.
- [00164] In this embodiment of the disclosure, after the k-th second point cloud feature of each point is obtained, the instance segmentation apparatus may perform second linear transformation on the k-th second point cloud feature of each point, to obtain the k-th second linear transformation feature. The dimension of the k-th second linear transformation feature is in one-to-one correspondence with the quantity of the multi-level objectness labels. A value of each dimension of the k-th second linear transformation feature represents a score of the corresponding level label.
- [00165] At S704, k-th second cross entropy between the k-th second linear transformation feature of each point and a true objectness label of a corresponding point is calculated, to obtain a mean value of the k-th second cross entropy of all of the points as a k-th objectness loss.
- [00166] In this embodiment of the disclosure, the instance segmentation apparatus may select the maximum value from the values of all of dimensions of the k-th second linear transformation feature of each point as a k-th second target dimension value of the corresponding point, calculate the k-th cross entropy between the k-th second target dimension values of all of the points and a k-th instance label value of the corresponding

level, and use the mean value of the k-th cross entropy corresponding to all of the points as the k-th objectness loss $L_{objectness}$, referring to a formula (5):

$$L_{objectness} = -\frac{1}{N_1} \sum_{n=1}^{N_1} HCE(q_n, o_n) \quad \text{Formula (5).}$$

The recombined training point cloud includes N_1 points; and for any point n, the k-th second linear transformation feature is q_n , and the true objectness label is o_n .

[00167] Exemplarily, for the point n, the k-th second linear transformation feature q_n is [0.3, 0.1, 0.1, 0.1, 0.9]. A prediction score that $label_{ob-ID}$ of the point n is 0 is 0.3, the prediction score that $label_{ob-ID}$ of the point n is 1 to 3 is 0.1, and the prediction score that $label_{ob-ID}$ of the point n is 4 is 0.9. The true objectness label o_n of the point n is [0, 0, 0, 0, 1], which indicates that the actual $label_{ob-ID}$ of the point n is 4.

[00168] At S705, in a case that a difference value of the k-th objectness loss between the k-th objectness loss and a (k-1)th objectness loss is consecutively less than or equal to a second difference value threshold and the number of consecutive times is greater than or equal to a second preset number, training is stopped, and the second backbone network of the (k-1)th update is used as the second backbone network; and the multi-level point cloud predicted by the second backbone network of the (k-1)th update is used as the multi-level prediction point cloud.

[00169] In this embodiment of the disclosure, after the k-th objectness loss is obtained, the instance segmentation apparatus may obtain the difference value of the k-th objectness loss by using the k-th objectness loss to subtract the (k-1)th objectness loss, stop training in a case that the difference value of the k-th objectness loss is consecutively less than or equal to the second difference value threshold for the Hth time and H is greater than or equal to the second preset number, and use the second backbone network of the (k-1)th update as the trained second backbone network. The second difference value threshold and the second preset number may be set according to actual requirements, and this embodiment of the disclosure is not limited thereto.

[00170] It may be understood that, by determining the difference value of the k-th objectness loss by using the second backbone network of the (k-1)th update, the instance segmentation apparatus may obtain an objectness loss difference value between the second backbone network of each update and the second backbone network of the last update, and stop training based on the condition that the objectness loss difference value is stable consecutively for many times. In this way, the second backbone network may achieve the highest accuracy, so that the accuracy of the multi-level prediction point cloud predicted by the second backbone network is enhanced.

[00171] In some embodiments of the disclosure, the implementation after S704 of calculating the k-th second cross entropy between the k-th second linear transformation feature of each point and an true objectness label of a corresponding point, to obtain the mean value of the k-th second cross entropy of all of the points as the k-th objectness loss may include the following steps.

- [00172] At S706, k-th update is performed on the (k-1)th second backbone network according to the k-th objectness loss in a case that the k-th objectness loss difference value is greater than the preset difference value threshold, or the difference value of the k-th objectness loss is consecutively less than or equal to the second preset difference value threshold and the number of consecutive times is less than the second preset number, so as to obtain the second backbone network of a k-th update.
- [00173] In this embodiment of the disclosure, the instance segmentation apparatus may perform k-th update on the (k-1)th second backbone network according to the k-th objectness loss in the case that the k-th objectness loss difference value is greater than the preset difference value threshold, so as to obtain the second backbone network of the k-th update.
- [00174] In this embodiment of the disclosure, the instance segmentation apparatus may perform k-th update on the (k-1)th second backbone network according to the k-th objectness loss in the case that the k-th objectness loss difference value is consecutively less than or equal to the second preset difference value threshold and the number of consecutive times is less than the second preset number, so as to obtain the second backbone network of the k-th update.
- [00175] It may be understood that, the instance segmentation apparatus may perform next update on the (k-1)th second backbone network in a case that the difference value of the k-th objectness loss does not meet the condition in S705, stop training until a k+h+1th loss difference value threshold obtained by the second backbone network of a (k+h)th update is consecutively less than or equal to the second preset difference value threshold and the number of consecutive times is greater than or equal to the second preset number, and then use the second backbone network of the (k+h)th update as the second backbone network, where h is a positive integer.
- [00176] In some embodiments of the disclosure, the implementation after S706 of performing k-th update on the (k-1)th second backbone network according to the k-th objectness loss may include the following steps.
- [00177] At S707, in a case that k is greater than or equal to a second update times threshold, training is stopped, and the second backbone network of the k-th update is used as the second backbone network.
- [00178] In this embodiment of the disclosure, the second update times threshold is the maximum update times of the second backbone network. After obtaining the second backbone network of the k-th update, the instance segmentation apparatus may determine whether k is greater than or equal to the second update times threshold, if so, stops training; otherwise, continues to acquire the (k+1)th loss. Referring to the step of S705, whether the second backbone network is trained for the next time is determined according to the (k+1)th loss.
- [00179] The second update times threshold may be set according to requirements, which is not limited in this embodiment of the disclosure.
- [00180] It may be understood that, the number of training times for the second backbone network may be limited by means of the second update times threshold, so that the training efficiency of the second backbone network is enhanced.
- [00181] At S304, an initial first backbone network is trained based on the first training

point cloud and the original semantic point cloud, to obtain the first backbone network.

- [00182] In this embodiment of the disclosure, after the first training point cloud is obtained, the instance segmentation apparatus may train the initial first backbone network based on the first training point cloud and the original semantic point cloud. After training is completed, the first backbone network is obtained.
- [00183] In some embodiments of the disclosure, the instance segmentation apparatus may use the pseudo-instance labels in the first training point cloud and the semantic labels in the original semantic point cloud as the supervision data, to train the initial first backbone network, so as to obtain the first backbone network.
- [00184] In some embodiments of the disclosure, the instance segmentation apparatus may use the pseudo-instance labels in the first training point cloud and the semantic labels in the original semantic point cloud as the supervision data, to perform a first round of training on the initial first backbone network, then perform instance segmentation according to the semantic segmentation result and the centroid shift vector outputted by the first backbone network after first round of training is completed and the multi-level prediction point cloud outputted by the trained second backbone network, so as to obtain updated pseudo-instance labels, and finally use the updated pseudo-instance labels and the semantic labels in the original semantic point cloud as the supervision data, to perform second round of training on the first backbone network after first round of training is completed.
- [00185] It may be understood that, the instance segmentation apparatus may obtain the first training point cloud by means of the original semantic point cloud, and may train the first backbone network and the second backbone network based on the first training point cloud, so that the point cloud carrying instance labels for supervision is not required. Therefore, the labor cost of training data may be reduced, and the efficiency of instance segmentation may be enhanced.
- [00186] In some embodiments of the disclosure, as shown in FIG. 13, the implementation of S304 of training the initial first backbone network based on the first training point cloud and the original semantic point cloud, to obtain the first backbone network may include S801 to S803.
- [00187] At S801, the initial first backbone network is trained for the first round based on a pseudo-instance label in the first training point cloud and a semantic label in the corresponding original semantic point cloud, to obtain an initial third backbone network, a third centroid shift vector, and a third semantic prediction result.
- [00188] In this embodiment of the disclosure, after obtaining the first training point cloud, the instance segmentation apparatus may use the pseudo-instance label in the first training point cloud and the semantic label in the original semantic point cloud as the supervision data, and train the initial first backbone network by using the original training point cloud as training data, so as to obtain the trained network as the initial third backbone network. The centroid shift vector predicted by the initial third backbone network is the third centroid shift vector. The third semantic prediction result is determined based on the semantic prediction result predicted based on the initial third backbone network. The third semantic prediction result includes the predicted semantic labels.

- [00189] In some embodiments of the disclosure, the implementation of S801 of performing first round of training on the initial first backbone network based on the pseudo-instance label in the first training point cloud and the semantic label in the corresponding original semantic point cloud, to obtain the initial third backbone network, the third centroid shift vector, and the third semantic prediction result may include the following steps.
- [00190] At S901, the first backbone network of an $(\alpha-1)$ th update is acquired, where α is a positive integer, and in a case that α equals to 1, the first backbone network of the $(\alpha-1)$ th update is the initial first backbone network.
- [00191] In this embodiment of the disclosure, the instance segmentation apparatus updates the initial first backbone network for a plurality of times. The first backbone network of the $(\alpha-1)$ th update represents that the initial first backbone network is updated for $\alpha-1$ times. In the case that α is 1, the first backbone network of the $(\alpha-1)$ th update represents that the initial first backbone network is updated for 0 times, that is, the initial first backbone network is not updated. That is to say, in the case that α is 1, the second backbone network of the $(\alpha-1)$ th update is the initial first backbone network.
- [00192] At S902, an α -th first semantic prediction result, an α -th first semantic loss, and an α -th first prediction shift vector are obtained based on the first backbone network of the $(\alpha-1)$ th update, the semantic label in the original semantic point cloud, and the pseudo-instance label in the first training point cloud. The α -th first shift loss is configured to characterize a difference between instance centroid predicted by the first backbone network of the $(\alpha-1)$ th update and true pseudo-instance centroid. The first training point cloud includes a true pseudo-instance.
- [00193] In this embodiment of the disclosure, the instance segmentation apparatus may perform semantic prediction on the original training point cloud by means of the first backbone network of the $(\alpha-1)$ th update and the semantic labels in the original semantic point cloud, to obtain the α -th first semantic prediction result and the α -th first semantic loss.
- [00194] In this embodiment of the disclosure, the instance segmentation apparatus may perform centroid shift prediction on the original training point cloud by means of the first backbone network of the $(\alpha-1)$ th update and the pseudo-instance labels in the first training point cloud, to obtain the first prediction shift vector.
- [00195] In some embodiments of the disclosure, the implementation of S902 of obtaining the α -th first semantic prediction result, the α -th first semantic loss, and the α -th first prediction shift vector based on the first backbone network of the $(\alpha-1)$ th update, the semantic label in the original semantic point cloud, and the pseudo-instance label in the first training point cloud may include S1001 to S1002.
- [00196] At S1001, α -th feature extraction is performed on each original point in the original training point cloud by using the first backbone network of the $(\alpha-1)$ th update, to obtain an α -th first point cloud feature of each original point.
- [00197] In this embodiment of the disclosure, the instance segmentation apparatus may perform feature extraction on each point in the original training point cloud by using the first backbone network of the $(\alpha-1)$ th update, to obtain the α -th first point cloud feature $F \in R^{N \times \text{dim}}$ of each original point.

[00198] Herein, the dimension dim of the first point cloud feature may be set according to requirements. For example, the dimension of the first point cloud feature may be 128 dimensions, 64 dimensions, or the same, which is not limited in this embodiment of the disclosure.

[00199] At S1002, semantic prediction is performed on the α -th first point cloud feature of each original point by using the first backbone network of the $(\alpha-1)$ th update, to obtain the α -th first semantic prediction result; the α -th first semantic loss is determined based on the α -th first point cloud feature of each original point and a semantic label of the corresponding original point; and feature transformation is performed on the α -th first point cloud feature of each original point, to obtain an α -th first prediction centroid shift vector of each original point.

[00200] In this embodiment of the disclosure, after obtaining the α -th first point cloud feature, the instance segmentation apparatus may perform semantic prediction on the α -th first point cloud feature of each original point by using the first backbone network of the $(\alpha-1)$ th update, to obtain the α -th first semantic prediction result. The α -th first semantic prediction result includes α -th semantic labels predicted by the first backbone network of the $(\alpha-1)$ th update.

[00201] In this embodiment of the disclosure, after obtaining the α -th first point cloud feature, the instance segmentation apparatus may perform first linear transformation on the α -th first point cloud feature of each original point, to obtain an α -th first linear transformation feature. The dimension of the first linear transformation feature is the same as the quantity of categories of the semantic labels in the original semantic point cloud. Each dimension value of the first linear transformation feature represents a score of the corresponding semantic label category.

[00202] In this embodiment of the disclosure, the instance segmentation apparatus may calculate a mean value of the cross entropy between the α -th first linear transformation feature of each point and the corresponding semantic label as the α -th first semantic loss. The first semantic loss may be calculated by referring to a formula (6):

$$L_{semantic} = -\frac{1}{N_2} \sum_{n=1}^{N_2} HCE(y_n, c_n) \quad \text{Formula (6).}$$

The original semantic point cloud includes N_2 points. For any point n , the α -th first linear transformation feature is y_n , and the corresponding semantic label is c_n .

[00203] In this embodiment of the disclosure, after the α -th first point cloud feature of each original point is obtained, the instance segmentation apparatus may perform feature transformation on the α -th first point cloud feature of each original point, to obtain an α -th prediction centroid shift vector of each original point.

[00204] Herein, the instance segmentation apparatus may perform feature transformation on the α -th first point cloud feature by using a Multi-Layer Perception (MLP) machine, to obtain the α -th prediction centroid shift vector $D \in R^{N \times 3}$.

[00205] In some embodiments, there may be two layers of MLP machines. The first point cloud feature of each original point is transformed into the prediction centroid shift vector of the corresponding original point by means of the two layers of MLP machines.

- [00206] It may be understood that, after the α -th first point cloud feature of each original point is extracted, the instance segmentation apparatus may obtain the α -th first semantic prediction result, the α -th first semantic loss, and the α -th first prediction shift vector.
- [00207] In some embodiments of the disclosure, the instance segmentation apparatus may perform semantic prediction on the α -th first point cloud feature of each original point by using the first backbone network of the $(\alpha-1)$ th update, and directly use the obtained semantic prediction result as the α -th first semantic prediction result.
- [00208] In some embodiments of the disclosure, as shown in FIG. 14, the implementation of S1002 of performing semantic prediction on the α -th first point cloud feature of each original point by using the first backbone network of the $(\alpha-1)$ th update, to obtain the α -th first semantic prediction result may include S10021 to S10022.
- [00209] At S10021, semantic prediction is performed on the α -th first point cloud feature of each original point by using the first backbone network of the $(\alpha-1)$ th update, to obtain an α -th initial semantic prediction point cloud.
- [00210] In this embodiment of the disclosure, the instance segmentation apparatus may perform semantic prediction on the α -th first point cloud feature of each original point by using the first backbone network of the $(\alpha-1)$ th update, to obtain α -th initial semantic prediction point cloud. Each point in the α -th initial semantic prediction point cloud has a corresponding semantic feature vector.
- [00211] At S10022, a semantic feature vector in each supervoxel in the α -th initial semantic prediction point cloud is smoothed by means of average pooling, to obtain a fine-tuned α -th initial semantic prediction point cloud as the α -th first semantic prediction result.
- [00212] In this embodiment of the disclosure, the supervoxel in any initial semantic prediction point cloud corresponds to the precomputed supervoxel in the original semantic point cloud. The instance segmentation apparatus may perform supervoxel clustering on the original semantic point cloud by using a supervoxel clustering algorithm, to obtain a plurality of supervoxel. The plurality of supervoxel in the original semantic point cloud is in one-to-one correspondence with the plurality of supervoxel in the initial semantic prediction point cloud.
- [00213] It is to be noted that, the supervoxel clustering algorithm may be a fuzzy C-means clustering method, may be a flow constrained clustering algorithm, or may be an unsupervised over-segmentation method. The supervoxel clustering algorithm may be designed according to requirements, which is not limited in this embodiment of the disclosure.
- [00214] In this embodiment of the disclosure, the instance segmentation apparatus may average each semantic feature vector in each of the plurality of supervoxel to obtain an average semantic vector of each supervoxel. The average semantic vector of each supervoxel acts as the semantic feature vector of each point in the corresponding supervoxel, so that a fine-tuned α -th initial semantic prediction point cloud is obtained as the α -th first semantic prediction result.
- [00215] It may be understood that, noise interference in the point cloud may be reduced by fine-tuning the semantic prediction result based on the supervoxel, so that the effect of

follow-up instance segmentation is enhanced.

- [00216] At S903, point cloud clustering is performed based on the α -th first semantic prediction result and the original semantic point cloud, to obtain an α -th first instance prediction result.
- [00217] In this embodiment of the disclosure, after obtaining the α -th first semantic prediction result, the instance segmentation apparatus may cluster the original semantic point cloud according to the semantic label of each point in the α -th first semantic prediction result, to obtain the α -th first instance prediction result. Herein, the point cloud clustering method may refer to the description in S103, which is not described here again.
- [00218] At S904, a first shift loss is determined according to the α -th first instance prediction result, the first training point cloud, and the α -th first prediction shift vector.
- [00219] In this embodiment of the disclosure, after obtaining the α -th first instance prediction result and the α -th first prediction shift vector, the instance segmentation apparatus may determine the first shift loss according to the α -th first instance prediction result, the first training point cloud, and the α -th first prediction shift vector.
- [00220] In some embodiments of the disclosure, as shown in FIG. 15, the manner of S904 of determining the first shift loss according to the α -th first instance prediction result, the first training point cloud, and the α -th first prediction shift vector may include S1101 to S1103.
- [00221] At S1101, an average coordinate of all of the points of each instance in the α -th first instance prediction result is determined as an α -th first instance centroid vector of each point in the corresponding instance.
- [00222] In this embodiment of the disclosure, after obtaining the α -th first instance prediction result, the instance segmentation apparatus may use the average coordinate of all of the points of each instance in the α -th first instance prediction result as the α -th first instance centroid vector $\hat{q}_n$ of each point in the corresponding instance.
- [00223] At S1102, a coordinate of each point corresponding to the first training point cloud is subtracted from the α -th first instance centroid vector of the corresponding point in the first instance prediction result, to obtain an α -th first vector difference value of each point as an α -th first true centroid shift vector of each point corresponding to the first training point cloud.
- [00224] In this embodiment of the disclosure, after obtaining the α -th first instance prediction result, the instance segmentation apparatus may determine the α -th first instance centroid vector of each point corresponding to the first training point cloud. Therefore, the instance segmentation apparatus may use the α -th first instance centroid vector of each point corresponding to the first training point cloud to subtract the coordinate p_n of the corresponding point, so as to obtain the α -th first vector difference value $(\hat{q}_n - p_n)$ of each point corresponding to the first training point cloud as the α -th first true centroid shift vector of each point corresponding to the first training point cloud.
- [00225] At S1103, the α -th first shift loss is determined according to the α -th first prediction centroid shift vector of each point corresponding to the first training point

cloud and the α -th first true centroid shift vector of the corresponding point.

- [00226] In this embodiment of the disclosure, after obtaining the α -th first true centroid shift vector of each point corresponding to the first training point cloud, the instance segmentation apparatus may determine the first shift loss of the first time according to the α -th first prediction centroid shift vector of each point corresponding to the first training point cloud and the α -th prediction centroid shift vector.
- [00227] It may be understood that, the instance segmentation apparatus may determine the true centroid shift vector according to the α -th first instance prediction result, and jointly determine the first shift loss according to the true centroid shift vector of each point corresponding to the first training point cloud and the prediction centroid shift vector, so that the first backbone network may learn the prediction of the centroid shift vector for the first training point cloud with high confidence, thereby enhancing the accuracy of the centroid shift vector.
- [00228] In some embodiments of the disclosure, as shown in FIG. 16, the implementation of S1103 of determining the α -th first shift loss according to the α -th first prediction centroid shift vector of each point corresponding to the first training point cloud and the α -th first true centroid shift vector of the corresponding point may include S1201 to S1206.
- [00229] At S1201, a norm of a difference value between the α -th first prediction centroid shift vector of each point corresponding to the first training point cloud and the α -th first true centroid shift vector of the corresponding point is calculated, to obtain an α -th first difference value norm of each point corresponding to the first training point cloud.
- [00230] In this embodiment of the disclosure, the instance segmentation apparatus may calculate the norm of the difference value between the α -th first prediction centroid shift vector d_n and the α -th first true centroid shift vector $(\hat{q}_n - p_n)$ for each point corresponding to the first training point cloud, as the α -th first difference value norm of each point.
- [00231] At S1202, the α -th first difference value norm of each point corresponding to the first training point cloud is averaged to obtain an α -th first shift sub-loss.
- [00232] In this embodiment of the disclosure, the instance segmentation apparatus may calculate the mean value of the α -th first difference value norm for all of the points corresponding to the first training point cloud, to obtain the α -th first shift sub-loss $L_{offset1}$, referring a formula (7):

$$L_{offset1} = \frac{1}{\sum_n M_n} \sum_n \|d_n - (\hat{q}_n - p_n) \cdot M_n\|$$

Formula (7);

n represents any point in the original semantic point cloud; and $M_n \in \{0,1\}$ is a binary mask. If n belongs to a point in the first training point cloud, the value of M_n is 0, otherwise, the value is 1. That is to say, the α -th first shift sub-loss $L_{offset1}$ only calculates the points corresponding to the first training point cloud.

[00233] At S1203, a ratio of the α -th first prediction centroid shift vector of each point corresponding to the first training point cloud to the norm of the α -th first prediction centroid shift vector of the corresponding point is calculated as an α -th first prediction ratio; a ratio of the α -th first true centroid shift vector of each point corresponding to the first training point cloud to the α -th first true centroid shift vector of the corresponding point is calculated as an α -th first actual ratio.

[00234] At S1204, a product of the α -th first prediction ratio of each point corresponding to the first training point cloud and the α -th first actual ratio of the corresponding point is determined, to obtain an α -th first direction product of each point corresponding to the first training point cloud.

[00235] At S1205, the α -th direction product of each point corresponding to the first training point cloud is averaged to obtain an α -th second shift sub-loss.

[00236] In this embodiment of the disclosure, the instance segmentation apparatus may calculate the ratio of the α -th first prediction centroid shift vector to the norm of the α -th first prediction centroid shift vector as the α -th first prediction ratio for each point corresponding to the first training point cloud, may calculate the ratio of the α -th first true centroid shift vector to the norm of the α -th first true centroid shift vector as the α -th first actual ratio, then may multiply the α -th first prediction ratio and the α -th first actual ratio, to obtain the α -th first direction product, and then may obtain the α -th second shift sub-loss $L_{offset2}$ by averaging the α -th first direction product of all of the points, referring to a formula (8):

$$L_{offset2} = \frac{1}{\sum_n M_n} \sum_n \frac{d_n}{\|d_n\|_2} \frac{\hat{q}_n - p_n}{\|\hat{q}_n - p_n\|_2} \cdot M_n \quad \text{Formula (8).}$$

Both the norm of the α -th first prediction centroid shift vector and the norm of the α -th first true centroid shift vector are A2 norms. It can be seen that, through the setting of M_n , the α -th second shift sub-loss $L_{offset2}$ only calculates the points corresponding to the first training point cloud.

[00237] It is to be noted that, the second shift sub-loss $L_{offset2}$ characterizes an additional direction loss based on cosine similarity.

[00238] At S1206, the α -th second shift sub-loss is subtracted from the α -th first shift sub-loss, to obtain the α -th first shift loss.

[00239] In this embodiment of the disclosure, after obtaining the α -th first shift sub-loss $L_{offset1}$ and the α -th second shift sub-loss $L_{offset2}$, the instance segmentation apparatus may use the α -th first shift sub-loss $L_{offset1}$ to subtract the α -th second shift sub-loss $L_{offset2}$, to obtain the α -th first shift loss, referring to a formula (9):

$$L_{offset} = L_{offset1} - L_{offset2} \quad \text{Formula (9).}$$

[00240] Herein, the α -th first shift sub-loss $L_{offset1}$ and the α -th second shift sub-loss $L_{offset2}$ in the formula (9) may be shown by using a formula (10) after expanding:

$$L_{offset} = \frac{1}{\sum_n M_n} \sum_n \|d_n - (\hat{q}_n - p_n) \cdot M_n\| - \frac{1}{\sum_n M_n} \sum_n \frac{d_n}{\|d_n\|_2} \frac{\hat{q}_n - p_n}{\|\hat{q}_n - p_n\|_2} \cdot M_n \quad \text{Formula (10).}$$

[00241] At S905, a sum of the α -th first semantic loss and the α -th first shift loss is used as an α -th first loss.

[00242] In this embodiment of the disclosure, after calculating the α -th first semantic loss $L_{semantic}$ and the α -th first shift loss L_{offset} , the instance segmentation apparatus may use the sum of the α -th first semantic loss $L_{semantic}$ and the α -th first shift loss L_{offset} as the α -th first loss. The α -th first loss is a joint loss of the α -th first semantic loss $L_{semantic}$ and the α -th first shift loss L_{offset} , which is recorded as L_{joint} , referring to a formula (11):

$$L_{joint} = L_{offset} + L_{semantic} \quad \text{Formula (11).}$$

[00243] It may be understood that, the instance segmentation apparatus updates the first backbone network by means of semantic loss and the joint loss of the shift loss for the centroid shift vector, so that the finally obtained first backbone network may simultaneously guarantee the prediction accuracy of the semantic prediction result and the centroid shift vector, thereby enhancing the accuracy of follow-up instance segmentation.

[00244] At S906, in a case that a α -th first loss difference value between the α -th first loss and the $(\alpha-1)$ th first loss is consecutively less than or equal to a first difference value threshold and the number of consecutive times is greater than or equal to a first preset number, training is stopped, and the first backbone network of the $(\alpha-1)$ th update is obtained as the initial third backbone network; and the α -th first prediction shift vector is used as the third centroid shift vector, and the α -th first semantic prediction result is used as the third semantic prediction result.

[00245] In this embodiment of the disclosure, after the α -th first loss is obtained, the instance segmentation apparatus acquires the $(\alpha-1)$ th first loss, and calculates a difference value between the α -th first loss and the $(\alpha-1)$ th first loss as the α -th first loss difference value.

[00246] In this embodiment of the disclosure, the instance segmentation apparatus may stop training by determining that an output area of the first backbone network is stable in a case that a α -th first loss difference value between the α -th first loss and the $(\alpha-1)$ th first loss is consecutively less than or equal to a first difference value threshold and the number of consecutive times is greater than or equal to a first preset number, and the first backbone network of the $(\alpha-1)$ th update is obtained as the initial third backbone network; and the α -th first prediction shift vector is used as the third centroid shift vector, and the α -th first semantic prediction result is used as the third semantic prediction result.

[00247] It may be understood that, in a case that the value of the first loss tends to be stable, the difference value between the current first loss and the last first loss is relatively small, which may be less than or equal to a first difference value threshold and the number of consecutive times is greater than or equal to the first preset number. In this way, the initial third backbone network may achieve the highest accuracy, so that the accuracy of the initial third backbone network for predicting the centroid shift vector and

the semantic segmentation result is enhanced.

- [00248] In some embodiments of the disclosure, in a case that the first backbone network is trained for the first round, the point cloud in a three-dimensional coordinate system may be used for training.
- [00249] In some embodiments of the disclosure, in a case that the first backbone network is trained for the first round, a depth coordinate may be ignored. For example, the three-dimensional coordinates of the point are (x, y, z), the z-coordinate may be ignored. Then two-dimensional point cloud is used for training, or the z coordinate is uniformly set to a value.
- [00250] It may be understood that, the training manner of ignoring the depth coordinate may reduce interference during sampling. For example, considering that the depth coordinate may cause a chair to be segmented into upper and lower instances, that is, over-segmentation, the situation may be improved by using the manner of ignoring the depth coordinate. Therefore, the effect of the first round of training of the first backbone network can be enhanced, thereby enhancing an instance segmentation effect.
- [00251] At S802, instance segmentation is performed on the multi-level prediction point cloud outputted by the second backbone network based on the third semantic prediction result and the third centroid shift vector, to obtain third training point cloud.
- [00252] In this embodiment of the disclosure, after obtaining the initial third backbone network by performing first round of training on the first backbone network, the instance segmentation apparatus may perform instance segmentation on the multi-level prediction point cloud outputted by the second backbone network based on the third semantic prediction result outputted based on the initial third backbone network and the third centroid shift vector, to obtain the third training point cloud.
- [00253] In some embodiments of the disclosure, the implementation of S802 of performing instance segmentation on the multi-level prediction point cloud outputted by the second backbone network based on the third semantic prediction result and the third centroid shift vector, to obtain the third training point cloud may include S1301 to S1303.
- [00254] At S1301, each point in the multi-level prediction point cloud is moved to a respective predicted center according to the third centroid shift vector, to obtain influence prediction point cloud.
- [00255] At S1302, core points of the influence prediction point cloud are clustered according to the third semantic prediction result, to obtain at least one third core point cloud of each semantic category.
- [00256] At S1303, edge points in the influence prediction point cloud are clustered into third target core point cloud in the at least one third core point cloud of the corresponding semantic category, to obtain the third training point cloud.
- [00257] In this embodiment of the disclosure, the implementation of S1301 to S1303 may refer to the description of S103 to S105, which is not described here again.
- [00258] It may be understood that, the instance segmentation apparatus may predict the core points in the core area and the edge points in the edge area in the multi-level prediction point cloud for the second backbone network. By using different manners for clustering, under-segmentation of the edge points is reduced, so that the accuracy of

instance segmentation in the third training point cloud.

- [00259] At S803, the initial third backbone network is trained for a second round based on a pseudo-instance label in the third training point cloud and the semantic label in the original semantic point cloud, to obtain the first backbone network.
- [00260] In this embodiment of the disclosure, the manner that the instance segmentation apparatus performs second round of training on the initial third backbone network based on the third training point cloud and the original semantic point cloud may refer to the implementation of S801 of performing the first round of training on the initial first backbone network based on the first training point cloud and the original semantic point cloud, to obtain the initial third backbone network, which is not described here again.
- [00261] It may be understood that, since the third training point cloud is the instance segmentation result that is obtained by respectively clustering the core points and the edge points based on the multi-level prediction point cloud, the third training point cloud includes more training data compared with the first training point cloud, and has higher accuracy of the pseudo-instance labels. By using the third training point cloud as new supervision data to replace the first training point cloud, the prediction accuracy of the first backbone network may be further enhanced.
- [00262] An embodiment of the disclosure provides a schematic diagram of a training process of an instance segmentation network. As shown in FIG. 17, the instance segmentation network includes a 3D U-net backbone network 1 and a 3D U-net backbone network 2. Solid arrows are configured to represent input and output of data. Dotted arrows are configured to represent input of the supervision data. Dashed arrows are configured to represent input of the updated supervision data. It can be seen from the figure that, the initial pseudo-instance point cloud may be obtained by using the BFS algorithm to perform instance segmentation on the original semantic point cloud. The first training point cloud may be obtained by performing point cloud filtering based on the quantity of the points in each pseudo-instance point cloud in the initial pseudo-instance point cloud. The first training point cloud and the original semantic point cloud are used as the supervision data of the first round of training of the 3D U-net backbone network 1. The original training point cloud is used as the training data to be inputted into the 3D U-net backbone network 1. After the first round of training is completed, the 3D U-net backbone network 1 may output the semantic prediction result and the predicted centroid shift vector. The labeling of the multi-level objectness labels is performed on the first training point cloud, and then point cloud recombination is performed, so that the recombined training point cloud may be obtained. The recombined training point cloud is used as the supervision data, and in addition, the recombined training point cloud without multi-level pseudo-instance labels is used as the training data, so that the 3D U-net backbone network 2 is trained. The trained 3D U-net backbone network 2 may output the multi-level prediction point cloud. The core points and the edge points may be distinguished by means of the multi-level objectness labels in the multi-level prediction point cloud. The multi-level prediction point cloud is moved by means of the centroid shift vector, to obtain the influence point cloud. The core points in the influence point cloud are clustered to obtain the core point cloud in the corresponding multi-level prediction point cloud. An influence value between the core points and other points is calculated based on the influence point cloud, to generate an influence diagram.

Then, the edge points are absorbed into the corresponding core point cloud, to finally complete instance segmentation, so as to obtain the third training point cloud. Finally, the first training point cloud is replaced with the third training point cloud as the supervision data for the 3D U-net backbone network 1, and the second round of training is performed on the 3D U-net backbone network 1, to obtain a trained 3D U-net backbone network 1.

[00263] Based on the instance segmentation method in FIG. 17, FIG. 18A shows the initial pseudo-instance point cloud. It can be seen that, the point cloud of three chairs around a table is connected and predicted as an instance. The three chairs may be filtered after the initial pseudo-instance is filtered, so as to obtain the first training point cloud in FIG. 18B. Learning is performed on the first training point cloud. The instance segmentation result predicted by the first backbone network is shown in FIG. 18C. Two of the three chairs are predicted as one instance, and the instance segmentation result is improved. FIG. 18D is the multi-level prediction point cloud predicted by the second backbone network. FIG. 18E is a result of core point clustering based on FIG. 18D. FIG. 18F shows the instance segmentation result after the edge points are absorbed based on the core point cloud in FIG. 18D. It can be seen that, 3 chairs are well distinguished in FIG. 18D, so that the instance segmentation effect of FIG. 18D is better than that of FIG. 18C.

[00264] Based on the instance segmentation network in FIG. 17, 3D U-net uses 7 layers of U-shaped meshes, the size of each supervoxel is 2 cm, the batch size of the training data of 3D U-net1 is set to 8, the batch size of the training data of 3D U-net2 is set to 16, the recombined training point cloud is 5000 virtual scenes, and an initial learning rate is 0.001. For the asymmetric reasoning algorithm, a hyper-parameter is set to 3. The search radius of the BFS algorithm of the initial pseudo-instance point cloud obtained from the original semantic point cloud is 5 cm. the preset edge level is 0 and 1, and the preset core level is 2 to 4. A training set adopts ScanNet-v2, to obtain three different average precisions of each pseudo-instance label in the third training point cloud, referring to Table 1.

[00265] Table 1

Evaluation index	Cabinet	Bed	Chair	Sofa	Table	Door	Window	Book-shelf	Picture	Evaluation index	Mean value
mAP	54.6	88.4	72.5	88.2	80.8	52.6	72.6	70.4	76	mAP	76.9
mAP@0.5	65.9	90.3	81.9	92.3	86.2	69.6	82.9	80.8	82.4		
mAP@0.25	73.1	90.3	84.8	93.8	87	73.4	86	82.7	84.9		
	Counter	Writing desk	Curtain	Refrigerator	Showers	Toilet	Sink	Bath-tub	Other	mAP@0.5	84.5
mAP	69.6	61.6	62.8	94.1	89	99	92.3	96.5	63.3		
mAP@0.5	81.7	72.8	79	96	93.5	99.3	95.9	97.5	73		
mAP@0.25	84.7	77.2	79.9	96	93.5	99.3	96.2	97.5	78.2	mAP@0.25	86.6

[00266] It can be seen from Table 1 that, for objects in different indoor scenes, the average score of each index exceeds 75. mAP represents a score range between 50% and 95%, and a step length is 5%. The average score of mAP@0.5 and mAP@0.25 is close to 85. That is to say, the instance segmentation methods provided in the embodiments of the disclosure all have desirable effects of instance segmentation.

[00267] As shown in Table 2, a verification set adopts ScanNet-v2. Instance segmentation is performed on the verification set by different instance segmentation methods, and the average accuracy of the different methods is obtained.

[00268] Table 2

Method	Sem.Supervision	Ins.Supervision	AP	AP ₅₀	AP ₂₅
PointGroup	100%	100%	34.8	56.9	71.3
CSC-20(PointGroup)	20pts/scene	20pts/scene	-	27.2	-
CSC-50(PointGroup)	50pts/scene	50pts/scene	-	35.7	-
CSC-100(PointGroup)	100pts/scene	100pts/scene	-	43.6	-
CSC-200(PointGroup)	200pts/scene	200pts/scene	-	50.4	-
SegGroup(PointGroup)	0.02%	0.02%	23.4	43.4	62.9
SPIB	100% Box sup.	100% Box sup.	-	38.6	61.4
AOIA(Ours)	100%	0%	33.7	54.9	70.4
AOIA [†] (Ours)	100%	0%	38.1	57	70.6

[00269] It can be seen from Table 2 that, a fully supervised method (PointGroup) and the instance segmentation method (AOIA) provided in the embodiments of the disclosure have basically the same average accuracy in different evaluation indexes. On the basis of the instance segmentation method provided in the embodiments of the disclosure, after the semantic prediction result is smoothed through the average pooling (AOIA[†]), in different evaluation indexes, 3 different AP values are higher than AOIA. For a weak supervision method, such as the 4 methods of CSC, the sparser the labeled points, the lower the AP₅₀ value. It may be seen that, the more points labeled in the training data, the higher the AP. However, SegGroup only labels one point for each instance, and the AP₅₀ value is basically the same as that of the method of CSC-100. SPIB labels 3D frames of each instance, and the obtained AP value is lower than that of SegGroup.

[00270] The verification set adopts ScanNet-v2. Table 3 shows the instance segmentation network after training with the filtered first training point cloud and with the unfiltered initial pseudo-instance point cloud. From the comparison of different AP values in Table 3, it can be seen that, the 3 AP values of the first training point cloud are higher than those of the initial pseudo-instance point cloud. That is to say, filtering the initial pseudo-instance point cloud may improve the effect of instance segmentation.

[00271] Table 3

Training point cloud	AP	AP ₅₀	AP ₂₅
Initial pseudo-instance point cloud	28.7	46.4	62.3
First training point cloud	30.3	48.5	65.5

[00272] The verification set adopts ScanNet-v2. Table 4 shows the impact of different preset core level settings on instance segmentation, the impact of whether to use center point regression, and the impact of whether to use projection. It can be seen from Table 4 that, the preset core level includes 1 to 4, that is, a manner of setting 80% of the points as the core points. Compared with the setting of 40% and 60%, three AP values are all increased, and higher than those of a baseline method of instance segmentation through BFS search. In addition, the AP value with center point regression is certainly higher than

the AP value without center point regression. According to the embodiments of the disclosure, the manner of center point regression is adopted, and the point cloud is moved to the predicted center to generate the influence point cloud. In the first round of training of 3D U-net1, in a case that projection is adopted to ignore the depth coordinate, that is, the z coordinate, three AP values are obviously increased compared with no projection.

[00273] Table 4

Instance segmentation method	Center point regression	Core point proportion	Projection	AP	AP ₅₀	AP ₂₅
BaselineBFS		100%		62	73.5	76.9
BaselineBFS	✓	100%		67.9	78.2	82.1
AOIA(Ours)	✓	Top40%		68.5	78.8	82.9
AOIA(Ours)	✓	Top60%		69.6	80	83.7
AOIA(Ours)	✓	Top80%		70.9	80.8	83.9
AOIA(Ours)	✓	Top80%	✓	75.6	83.9	86.3

[00274] FIG. 19A shows the multi-level point cloud. Based on the multi-level point cloud in FIG. 19A, FIG. 19B is a schematic diagram of a core point clustering effect using 40% of points. FIG. 19C is a schematic diagram of a core point clustering effect using 80% of points.

[00275] It is to be noted that, the smaller the percentage of core points, the better the instance segmentation is, but it is also easy to cause over-segmentation. Therefore, it is necessary to use an appropriate percentage to achieve the optimal effect.

[00276] Based on the foregoing embodiments, an embodiment of the disclosure provides an instance segmentation apparatus. The apparatus includes the included units and modules included in the units, and may be implemented by using a processor in a computer device. Definitely, the apparatus may alternatively be implemented by using a specific logic circuit. During implementation, the processor may be a Central Processing Unit (CPU), a Microprocessor Unit (MPU), a Digital Signal Processor (DSP), or a Field Programmable Gate Array (FPGA).

[00277] FIG. 20 is a schematic diagram of a composition structure of an instance segmentation apparatus according to an embodiment of the disclosure. As shown in FIG. 20, the instance segmentation apparatus 600 includes a first prediction module 610, a second prediction module 620, a moving module 630, a core clustering module 640, and an edge clustering module 650.

[00278] The first prediction module 610 is configured to perform semantic prediction and centroid shift prediction on acquired to-be-segmented point cloud by using a first backbone network, to obtain a semantic prediction result and centroid shift vector of each point in the to-be-segmented point cloud. The first backbone network and a second

backbone network are trained based on first training point cloud, and the first training point cloud is determined by means of original semantic point cloud carrying a semantic label.

- [00279] The second prediction module 620 is configured to perform objectness prediction on the to-be-segmented point cloud by using the second backbone network, to obtain multi-level point cloud including a multi-level objectness label. The multi-level objectness label is configured to characterize a shape of each instance, internal information, and external information.
- [00280] The moving module 630 is configured to move each point in the multi-level point cloud to a respective predicted center according to the centroid shift vector, to obtain influence point cloud.
- [00281] The core clustering module 640 is configured to cluster core points of the influence point cloud according to the semantic prediction result, to obtain at least one core point cloud of each semantic category. The core points are points having objectness label at preset core levels.
- [00282] The edge clustering module 650 is configured to cluster edge points in the influence point cloud into target core point cloud in the at least one core point cloud of the corresponding semantic category, to obtain an instance segmentation result. The edge points are points having objectness labels at preset edge levels.
- [00283] In some embodiments, the edge clustering module 650 is further configured to calculate an influence value between each edge point and the at least one core point cloud of the corresponding semantic category, to obtain at least one corresponding influence value of each edge point, where the influence value is configured to characterize a mean value of the influence values between the edge point and all of the points in core influence point cloud of the corresponding semantic category; use core point cloud corresponding to a maximum value in the at least one influence value of each edge point as the target core point cloud corresponding to each edge point; and cluster each edge point into the corresponding target core point cloud, to obtain at least one instance segmentation point cloud of each semantic category as the instance segmentation result.
- [00284] In some embodiments, the instance segmentation apparatus 600 further includes a training module. The training module is configured to: before the performing semantic prediction and centroid shift prediction on to-be-segmented point cloud by using a first backbone network, to obtain a semantic prediction result and centroid shift vector of each point in the to-be-segmented point cloud, cluster original semantic point cloud to obtain initial pseudo-instance point cloud; filter the initial pseudo-instance point cloud according to the quantity of points included in each pseudo-instance in the initial pseudo-instance point cloud, to obtain the first training point cloud; train an initial second backbone network based on the first training point cloud, to obtain the second backbone network; and train an initial first backbone network based on the first training point cloud and the original semantic point cloud, to obtain the first backbone network.
- [00285] In some embodiments, the training module is further configured to: sort the quantity of the points in each pseudo-instance point cloud of each semantic category, to obtain a quantity sequence of each semantic category; and select pseudo-instance point cloud within a preset quantity sequence range from the quantity sequence of each semantic category, to obtain training point cloud of each semantic category. The first

training point cloud includes the training point cloud of each semantic category.

- [00286] In some embodiments, the training module is further configured to: perform point cloud compression on the first training point cloud, to obtain first compressed point cloud; determine the multi-level training point cloud corresponding to each pseudo-instance point cloud according to a compressed distance between each point in each pseudo-instance point cloud and corresponding compressed pseudo-instance point cloud in the first compressed point cloud, where the multi-level training point cloud includes the multi-level objectness label; perform point cloud recombination based on the multi-level training point cloud corresponding to each pseudo-instance point cloud, to obtain recombined training point cloud; and train the initial second backbone network by using the recombined training point cloud, to obtain the second backbone network and the multi-level prediction point cloud outputted by the second backbone network.
- [00287] In some embodiments, the training module is further configured to: calculate a distance between each point in each pseudo-instance point cloud and each point in the corresponding compressed pseudo-instance point cloud, to obtain a plurality of distances corresponding to each point in each pseudo-instance point cloud; use a minimum distance in the plurality of compressed distances corresponding to each point in each pseudo-instance point cloud as the compressed distance between each point in each pseudo-instance point cloud and corresponding compressed point cloud, to obtain a plurality of compressed distances of each pseudo-instance point cloud; and determine the multi-level objectness labels of each pseudo-instance point cloud according to the plurality of compressed distances of each pseudo-instance point cloud, to obtain the multi-level training point cloud corresponding to each pseudo-instance point cloud.
- [00288] In some embodiments, the training module is further configured to perform point cloud recombination on the multi-level training point cloud according to a preset point cloud attribute in a recombination template, to obtain the recombined training point cloud. The preset point cloud attribute includes a preset point cloud quantity, a preset point cloud volume, and a preset point cloud position, and there is no intersection area between the preset point cloud positions.
- [00289] In some embodiments, the training module is further configured to: acquire the second backbone network of a $(k-1)$ th update, where k is a positive integer, and in a case that k equals to 1, the second backbone network of the $(k-1)$ th update is the initial second backbone network; perform feature extraction on each point of the recombined training point cloud by means of the second backbone network of the $(k-1)$ th update, to obtain a k -th second point cloud feature of each point; perform second linear transformation on the k -th second point cloud feature of each point, to obtain a k -th second linear transformation feature, where a dimension of the k -th second linear transformation feature is in one-to-one correspondence with a level quantity of the multi-level objectness labels; calculate k -th second cross entropy between the k -th second linear transformation feature of each point and an true objectness label of a corresponding point, to obtain a mean value of the k -th second cross entropy of all of the points as a k -th objectness loss; and stop training in a case that a difference value of the k -th objectness loss between the k -th objectness loss and a $(k-1)$ th objectness loss is consecutively less than or equal to a second preset difference value threshold and the number of consecutive times is greater than or equal to a second preset number, and use the second backbone network of the $(k-1)$ th update as the second backbone network.

- [00290] In some embodiments, the training module is further configured to: after the calculating k-th second cross entropy between a second target dimension value in the k-th second linear transformation feature of each point and a corresponding target level training label value, to obtain a mean value of the k-th second cross entropy of all of the points as a k-th objectness loss, perform k-th update on the (k-1)th second backbone network according to the k-th objectness loss in a case that the difference value of the k-th objectness loss is greater than the preset difference value threshold, or the difference value of the k-th objectness loss is consecutively less than or equal to a second difference value and the number of consecutive times is less than the second preset number, so as to obtain the second backbone network of a k-th update.
- [00291] In some embodiments, the training module is further configured to: train the initial first backbone network based on a pseudo-instance label in the first training point cloud and a semantic label in the corresponding original semantic point cloud, to obtain an initial third backbone network, a third centroid shift vector, and a third semantic prediction result; perform instance segmentation on the multi-level prediction point cloud outputted by the second backbone network based on the third semantic prediction result and the third centroid shift vector, to obtain third training point cloud; and train the initial third backbone network based on a pseudo-instance label in the third training point cloud and the semantic label in the original semantic point cloud, to obtain the first backbone network.
- [00292] In some embodiments, the training module is further configured to: acquire the first backbone network of an $(\alpha-1)$ th update, where α is a positive integer, and in a case that α equals to 1, the first backbone network of the $(\alpha-1)$ th update is the initial first backbone network; obtain an α -th first semantic prediction result, an α -th first semantic loss, and an α -th first prediction shift vector based on the first backbone network of the $(\alpha-1)$ th update, the semantic label in the original semantic point cloud, and the pseudo-instance label in the first training point cloud, where the α -th first shift loss is configured to characterize a difference between instance centroid predicted by the first backbone network of the $(\alpha-1)$ th update and true pseudo-instance centroid, and the first training point cloud includes a true pseudo-instance; perform point cloud clustering based on the α -th first semantic prediction result and the original semantic point cloud, to obtain an α -th first instance prediction result; determine a first shift loss according to the α -th first instance prediction result, the first training point cloud, and the α -th first prediction shift vector; use a sum of the α -th first semantic loss and the α -th first shift loss as an α -th first loss; stop training in a case that a α -th first loss difference value between the α -th first loss and the $(\alpha-1)$ th first loss is consecutively less than or equal to a first difference value threshold and the number of consecutive times is greater than or equal to a first preset number, and obtain the first backbone network of the $(\alpha-1)$ th update as the initial third backbone network; and use the α -th first prediction shift vector as the third centroid shift vector, and use the α -th first semantic prediction result as the third semantic prediction result.
- [00293] In some embodiments, the training module is further configured to: perform α -th feature extraction on each original point in the original training point cloud by using the first backbone network of the $(\alpha-1)$ th update, to obtain an α -th first point cloud feature of each original point; perform semantic prediction on the α -th first point cloud feature of each original point by using the first backbone network of the $(\alpha-1)$ th update,

to obtain the α -th first semantic prediction result; determine the α -th first semantic loss based on the α -th first point cloud feature of each original point and a semantic label of the corresponding original point; and perform feature transformation on the α -th first point cloud feature of each original point, to obtain an α -th first prediction centroid shift vector of each original point.

[00294] In some embodiments, the training module is further configured to: determine an average coordinate of all of the points of each instance in the α -th first instance prediction result as an α -th first instance centroid vector of each point in the corresponding instance; subtract a coordinate of each point corresponding to the first training point cloud from the α -th first instance centroid vector of the corresponding point, so as to obtain an α -th first vector difference value of each point corresponding to the first training point cloud as an α -th first true centroid shift vector of each point corresponding to the first training point cloud; and determine the α -th first shift loss according to the α -th first prediction centroid shift vector of each point corresponding to the first training point cloud and the α -th first true centroid shift vector of the corresponding point.

[00295] In some embodiments, the training module is further configured to: perform semantic prediction on the α -th first point cloud feature of each original point by using the first backbone network of the $(\alpha-1)$ th update, to obtain an α -th initial semantic prediction point cloud; and smooth a semantic feature vector in each supervoxel in the α -th initial semantic prediction point cloud by means of average pooling, to obtain a fine-tuned α -th initial semantic prediction point cloud as the α -th first semantic prediction result.

[00296] In some embodiments, the training module is further configured to: calculate a norm of a difference value between the α -th first prediction centroid shift vector of each point corresponding to the first training point cloud and the α -th first true centroid shift vector of the corresponding point, to obtain an α -th first difference value norm of each point corresponding to the first training point cloud; average the α -th first difference value norm of each point corresponding to the first training point cloud to obtain an α -th first shift sub-loss; calculate a ratio of the α -th first prediction centroid shift vector of each point corresponding to the first training point cloud to the norm of the α -th first prediction centroid shift vector of the corresponding point as an α -th first prediction ratio; calculate a ratio of the α -th first true centroid shift vector of each point corresponding to the first training point cloud to the norm of the α -th first true centroid shift vector as an α -th first actual ratio; determine a product of the α -th first prediction ratio of each point corresponding to the first training point cloud and the α -th first actual ratio of the corresponding point, to obtain an α -th first direction product of each point corresponding to the first training point cloud; average the α -th first direction product of each point corresponding to the first training point cloud to obtain an α -th second shift sub-loss; and subtract the α -th second shift sub-loss from the α -th first shift sub-loss, to obtain the α -th first shift loss.

[00297] In some embodiments, the training module is further configured to: after the using a sum of the α -th first semantic loss and the α -th first shift loss as an α -th first loss, perform α -th update on the first backbone network of the $(\alpha-1)$ th update according to the α -th first loss in a case that the α -th first loss difference value is greater than a first loss difference value threshold, or the α -th first loss difference value is consecutively less than

or equal to the first loss difference value threshold and the number of consecutive times is less than the first preset number, so as to obtain the first backbone network of an α -th update.

[00298] In some embodiments, the training module is further configured to: move each point in the multi-level prediction point cloud to a respective predicted center according to the third centroid shift vector, to obtain influence prediction point cloud; cluster core points of the influence prediction point cloud according to the third semantic prediction result, to obtain at least one third core point cloud of each semantic category; and cluster edge points in the influence prediction point cloud into third target core point cloud in the at least one third core point cloud of the corresponding semantic category, to obtain the third training point cloud.

[00299] The descriptions of the above apparatus embodiments are similar to the descriptions of the above method embodiments, and have similar beneficial effects to the method embodiments. In some embodiments, functions or modules of the apparatus provided in the embodiments of the disclosure may be configured to execute the method described in the above method embodiment. For technical details not disclosed in the apparatus embodiments of the disclosure, refer to the descriptions of the method embodiments of the disclosure for understanding.

[00300] It is to be noted that, in the embodiments of the disclosure, if the above instance segmentation method is implemented in the form of a software functional module and sold or used as an independent product, it can be stored in the computer-readable storage medium. Based on such an understanding, the technical solutions of the disclosure substantially or parts making contributions to the related art may be embodied in form of a software product, and the software product is stored in a storage medium, including a plurality of instructions for causing a computer device (which may be a personal computer, a server, a network device or the like) to execute all or part of the method in each embodiment of the disclosure. The foregoing storage medium includes: a USB flash disk, a mobile Hard Disk Drive (HDD), a Read-Only Memory (ROM), and various media that can store program codes, such as a magnetic disk, or an optical disk. In this way, the embodiments of the disclosure are not limited to any specific hardware, software, or firmware, or any combination of hardware, software, and firmware.

[00301] An embodiment of the disclosure provides a computer device. The computer device includes a memory and a processor. The memory stores a computer program executable on the processor. The processor, when executing the computer program, implements the above method.

[00302] An embodiment of the disclosure provides a computer-readable storage medium. The computer-readable storage medium stores a computer program. The above method is implemented when the computer program is performed by a processor. The computer-readable storage medium may be transitory or non-transitory.

[00303] An embodiment of the disclosure provides a computer program. The computer program includes a computer-readable code. When the computer-readable code is operated in a computer device, a processor in the computer device executes and implements part or all of steps is the above method.

[00304] An embodiment of the disclosure provides a computer program product. The computer program product includes a non-transitory computer-readable storage medium

storing a computer program. When the computer program is read and executed by a computer, part or all of steps in the above method are implemented. The computer program product can be specifically implemented in a manner of hardware, software or a combination thereof. In some embodiments, the computer program product is embodied as a computer storage medium. In some other embodiments, the computer program product is embodied as a software product, such as a Software Development Kit (SDK) and the like.

[00305] Herein, it is to be noted that, the above description of the various embodiments tends to emphasize the differences between the various embodiments, and the identicalness or similarities can be referred to each other. The descriptions of the above device, storage medium, computer program, and computer program product embodiments are similar to the descriptions of the above method embodiments, and have similar beneficial effects to the method embodiments. For technical details not disclosed in the embodiments of the device, storage medium, computer program and computer program product of the disclosure, refer to the description of the method embodiments of the disclosure for understanding.

[00306] It is to be noted that, FIG. 21 is a schematic diagram of a hardware entity of an instance segmentation device according to an embodiment of the disclosure. As shown in FIG. 21, a hardware entity of the instance segmentation device 700 includes a processor 701, a communication interface 702, and a memory 703.

[00307] The processor 701 usually controls an overall operation of the computer device 700.

[00308] The communication interface 702 may cause the computer device to achieve communication with other terminals or servers through a network.

[00309] The memory 703 is configured to store an instruction and an application executable by the processor 701, or may cache data to be processed or processed by the processor 701 and each module in the computer device 700 (such as, image data, audio data, voice communication data, and video communication data), which may be implemented by using FLASH or a Random Access Memory (RAM). Data transmission among the processor 701, the communication interface 702, and the memory 703 may be achieved by using a bus 704.

[00310] It is to be understood that "one embodiment" and "an embodiment" mentioned in the whole specification mean that specific features, structures or characteristics related to the embodiment is included in at least one embodiment of the disclosure. Therefore, "in one embodiment" or "in an embodiment" appearing at any place of the whole specification does not always refer to the same embodiment. In addition, these specific features, structures or characteristics may be combined in one or more embodiments in any proper manner. It is to be understood that, in various embodiments of the disclosure, the sequence number of each step/process does not mean the sequence of execution. The execution sequence of each step/process should be determined by its functions and internal logic, which should not constitute any limitation on the implementation process of the embodiments of the disclosure. The serial numbers of the foregoing embodiments of the disclosure are merely for description, and do not represent the superiority or inferiority of the embodiments.

[00311] It is to be noted that terms "include" and "comprise" or any other variant

thereof is intended to cover nonexclusive inclusions herein, so that a process, method, object or apparatus including a series of components not only includes those components but also includes other components which are not clearly listed or further includes components intrinsic to the process, the method, the object or the apparatus. Under the condition of no more limitations, a component defined by the statement "including a/an..." does not exclude existence of the same other components in a process, method, object or apparatus including the component.

[00312] In several embodiments provided by the disclosure, it is to be understood that the disclosed device and method may be implemented in other ways. The device embodiment described above is only schematic, and for example, division of the units is only logic function division, and other division manners may be adopted during practical implementation. For example, a plurality of units or components may be combined or integrated into another system, or some characteristics may be neglected or not executed. In addition, coupling or direct coupling or communication connection between each displayed or discussed component may be indirect coupling or communication connection, implemented through some interfaces, of the device or the units, and may be electrical and mechanical or adopt other forms.

[00313] The units described as separate parts may or may not be physically separated, and parts displayed as units may or may not be physical units, and namely may be located in the same place, or may also be distributed to multiple network units. Part of all of the units may be selected according to a practical requirement to achieve the purposes of the solutions of the embodiments.

[00314] In addition, each functional unit in each embodiment of the disclosure may be integrated into a processing unit, each unit may also serve as an independent unit and two or more than two units may also be integrated into a unit. The integrated unit may be implemented in a hardware form and may also be implemented in form of hardware and software functional unit.

[00315] Those of ordinary skill in the art should know that all or part of the steps of the method embodiment may be implemented by related hardware instructed through a program, the program may be stored in a computer-readable storage medium, and the program is executed to execute the steps of the method embodiment. The storage medium which may be a volatile storage medium includes: a mobile storage device, an ROM, and various media that can store program codes, such as a magnetic disk, or an optical disk.

[00316] If the integrated unit is implemented in the form of the software functional unit and sold or used as an independent product, it can be stored in the computer readable storage medium. Based on such an understanding, the technical solutions of the disclosure substantially or parts making contributions to the related art may be embodied in form of a software product, and the computer software product is stored in a storage medium, including a plurality of instructions for causing a computer device (which may be a personal computer, a server, a network device or the like) to execute all or part of the method in each embodiment of the disclosure. The foregoing storage medium includes a portable storage device, an ROM, and various media that can store program codes, such as a magnetic disk, or an optical disk.

[00317] The above is only the implementations of the disclosure and not intended to limit the scope of protection of the disclosure. Any variations or replacements apparent to

those skilled in the art within the technical scope disclosed by the disclosure shall fall within the scope of protection of the disclosure.

CLAIMS

1. An instance segmentation method, comprising:

performing semantic prediction and centroid shift prediction on acquired to-be-segmented point cloud by using a first backbone network, to obtain a semantic prediction result and centroid shift vector of each point in the to-be-segmented point cloud, wherein the first backbone network and a second backbone network are trained based on first training point cloud, and the first training point cloud is determined by means of original semantic point cloud carrying a semantic label;

performing objectness prediction on the to-be-segmented point cloud by using the second backbone network, to obtain multi-level point cloud comprising multi-level objectness labels, wherein the multi-level objectness labels are configured to characterize a shape of each instance, internal information, and external information;

moving each point in the multi-level point cloud to a respective predicted center according to the centroid shift vector, to obtain influence point cloud;

clustering core points of the influence point cloud according to the semantic prediction result, to obtain at least one core point cloud of each semantic category, wherein the core points are points having objectness labels at preset core levels; and

clustering edge points in the influence point cloud into target core point cloud in the at least one core point cloud of the corresponding semantic category, to obtain an instance segmentation result, wherein the edge points are points having objectness labels at preset edge levels.

2. The method of claim 1, wherein the clustering edge points in the influence point cloud into target core point cloud in the at least one core point cloud of the corresponding semantic category, to obtain an instance segmentation result comprises:

calculating an influence value between each edge point and the at least one core point cloud of the corresponding semantic category, to obtain at least one corresponding influence value of each edge point, wherein the influence value is configured to characterize a mean value of influence values between the edge point and all of the points in core influence point cloud of the corresponding semantic category;

using core point cloud corresponding to a maximum value in the at least one influence value of the each edge point as the target core point cloud corresponding to the each edge point; and

clustering the each edge point into the corresponding target core point cloud, to obtain at least one instance segmentation point cloud of each semantic category as the instance segmentation result.

3. The method of claim 1 or 2, wherein, before the performing semantic prediction and centroid shift prediction on to-be-segmented point cloud by using a first backbone network, to obtain a semantic prediction result and centroid shift vector of each point in the to-be-

segmented point cloud, the method further comprises:

clustering original semantic point cloud to obtain initial pseudo-instance point cloud;

filtering the initial pseudo-instance point cloud according to the quantity of points comprised in each pseudo-instance in the initial pseudo-instance point cloud, to obtain the first training point cloud;

training an initial second backbone network based on the first training point cloud, to obtain the second backbone network; and

training an initial first backbone network based on the first training point cloud and the original semantic point cloud, to obtain the first backbone network.

4. The method of claim 3, wherein the filtering the initial pseudo-instance point cloud according to the quantity of points comprised in each pseudo-instance in the initial pseudo-instance point cloud, to obtain the first training point cloud comprises:

sorting the quantity of the points in each pseudo-instance point cloud of each semantic category, to obtain a quantity sequence of each semantic category; and

selecting pseudo-instance point cloud within a preset quantity sequence range from the quantity sequence of each semantic category, to obtain training point cloud of each semantic category, wherein the first training point cloud comprises the training point cloud of each semantic category.

5. The method of claim 3 or 4, wherein the training an initial second backbone network based on the first training point cloud, to obtain the second backbone network, and outputting multi-level training point cloud comprises:

performing point cloud compression on the first training point cloud, to obtain first compressed point cloud;

determining the multi-level training point cloud corresponding to each pseudo-instance point cloud according to a compressed distance between each point in the each pseudo-instance point cloud in the first training point cloud and corresponding compressed pseudo-instance point cloud in the first compressed point cloud, wherein the multi-level training point cloud comprises the multi-level objectness labels;

performing point cloud recombination based on the multi-level training point cloud corresponding to the each pseudo-instance point cloud, to obtain recombined training point cloud; and

training the initial second backbone network by using the recombined training point cloud, to obtain the second backbone network and the multi-level prediction point cloud outputted by the second backbone network.

6. The method of claim 5, wherein the determining the multi-level training point cloud corresponding to each pseudo-instance point cloud according to a compressed distance between each point in the each pseudo-instance point cloud in the first training point cloud and corresponding compressed pseudo-instance point cloud in the first compressed point

cloud comprises:

calculating a distance between each point in the each pseudo-instance point cloud and each point in the corresponding compressed pseudo-instance point cloud, to obtain a plurality of distances corresponding to each point in the each pseudo-instance point cloud;

using a minimum distance in the plurality of distances corresponding to each point in the each pseudo-instance point cloud as the compressed distance between each point in the each pseudo-instance point cloud and corresponding compressed point cloud, to obtain a plurality of compressed distances of the each pseudo-instance point cloud; and

determining the multi-level objectness labels of the each pseudo-instance point cloud according to the plurality of compressed distances of the each pseudo-instance point cloud, to obtain the multi-level training point cloud corresponding to the each pseudo-instance point cloud.

7. The method of claim 5 or 6, wherein the performing point cloud recombination based on the multi-level training point cloud corresponding to the each pseudo-instance point cloud, to obtain recombined training point cloud comprises:

performing point cloud recombination on the multi-level training point cloud according to a preset point cloud attribute in a recombination template, to obtain the recombined training point cloud, wherein the preset point cloud attribute comprises a preset point cloud quantity, a preset point cloud volume, and a preset point cloud position, and there is no intersection area between preset point cloud positions.

8. The method of any of claims 5 to 7, wherein the training the initial second backbone network by using the recombined training point cloud, to obtain the second backbone network comprises:

acquiring the second backbone network of a (k-1)th update, wherein k is a positive integer, and in a case that k equals to 1, the second backbone network of the (k-1)th update is the initial second backbone network;

performing feature extraction on each point of the recombined training point cloud by means of the second backbone network of the (k-1)th update, to obtain a k-th second point cloud feature of each point;

performing second linear transformation on the k-th second point cloud feature of the each point, to obtain a k-th second linear transformation feature, wherein a dimension of the k-th second linear transformation feature is in one-to-one correspondence with a level quantity of the multi-level objectness labels;

calculating k-th second cross entropy between the k-th second linear transformation feature of the each point and an true objectness label of a corresponding point, to obtain a mean value of the k-th second cross entropy of all of the points as a k-th objectness loss; and

in a case that a difference value of the k-th objectness loss between the k-th objectness loss and a (k-1)th objectness loss is consecutively less than or equal to a second preset difference value threshold and the number of consecutive times is greater than or equal to a second preset number, stopping training, and using the second backbone network of the (k-

1)th update as the second backbone network.

9. The method of claim 8, wherein, after the calculating k-th second cross entropy between the k-th second linear transformation feature of the each point and a true objectness label of a corresponding point, to obtain a mean value of the k-th second cross entropy of all of the points as a k-th objectness loss, the method further comprises:

in a case that the difference value of the k-th objectness loss is greater than the preset difference value threshold, or the difference value of the k-th objectness loss is consecutively less than or equal to a second difference value and the number of consecutive times is less than the second preset number, performing k-th update on the (k-1)th second backbone network according to the k-th objectness loss, to obtain the second backbone network of a k-th update.

10. The method of any of claims 3 to 9, wherein the training an initial first backbone network based on the first training point cloud and the original semantic point cloud, to obtain the first backbone network comprises:

training the initial first backbone network based on a pseudo-instance label in the first training point cloud and a semantic label in the corresponding original semantic point cloud, to obtain an initial third backbone network, a third centroid shift vector, and a third semantic prediction result;

performing instance segmentation on the multi-level prediction point cloud outputted by the second backbone network based on the third semantic prediction result and the third centroid shift vector, to obtain third training point cloud; and

training the initial third backbone network based on a pseudo-instance label in the third training point cloud and the semantic label in the original semantic point cloud, to obtain the first backbone network.

11. The method of claim 10, wherein the training the initial first backbone network based on a pseudo-instance label in the first training point cloud and a semantic label in the corresponding original semantic point cloud, to obtain an initial third backbone network, a third centroid shift vector, and a third semantic prediction result comprises:

acquiring the first backbone network of an $(\alpha-1)$ th update, wherein α is a positive integer, and in a case that α equals to 1, the first backbone network of the $(\alpha-1)$ th update is the initial first backbone network;

obtaining an α -th first semantic prediction result, an α -th first semantic loss, and an α -th first prediction shift vector based on the first backbone network of the $(\alpha-1)$ th update, the semantic label in the original semantic point cloud, and the pseudo-instance label in the first training point cloud, wherein the α -th first shift loss is configured to characterize a difference between instance centroid predicted by the first backbone network of the $(\alpha-1)$ th update and true pseudo-instance centroid, and the first training point cloud comprises a true pseudo-instance;

performing point cloud clustering based on the α -th first semantic prediction result and

the original semantic point cloud, to obtain an α -th first instance prediction result;

determining a first shift loss according to the α -th first instance prediction result, the first training point cloud, and the α -th first prediction shift vector;

using a sum of the α -th first semantic loss and the α -th first shift loss as an α -th first loss;

in a case that a α -th first loss difference value between the α -th first loss and the $(\alpha-1)$ th first loss is consecutively less than or equal to a first difference value threshold and the number of consecutive times is greater than or equal to a first preset number, stopping training, and obtaining the first backbone network of the $(\alpha-1)$ th update as the initial third backbone network; and using the α -th first prediction shift vector as the third centroid shift vector, and using the α -th first semantic prediction result as the third semantic prediction result.

12. The method of claim 11, wherein the obtaining an α -th first semantic prediction result, an α -th first semantic loss, and an α -th first prediction shift vector based on the first backbone network of the $(\alpha-1)$ th update, the semantic label in the original semantic point cloud, and the pseudo-instance label in the first training point cloud comprises:

performing α -th feature extraction on each original point in the original training point cloud by using the first backbone network of the $(\alpha-1)$ th update, to obtain an α -th first point cloud feature of the each original point;

performing semantic prediction on the α -th first point cloud feature of the each original point by using the first backbone network of the $(\alpha-1)$ th update, to obtain the α -th first semantic prediction result;

determining the α -th first semantic loss based on the α -th first point cloud feature of the each original point and a semantic label of the corresponding original point; and

performing feature transformation on the α -th first point cloud feature of the each original point, to obtain an α -th first prediction centroid shift vector of the each original point.

13. The method of claim 11, wherein the determining a first shift loss according to the α -th first instance prediction result, the first training point cloud, and the α -th first prediction shift vector comprises:

determining an average coordinate of all of the points of each instance in the α -th first instance prediction result as an α -th first instance centroid vector of each point in the corresponding instance;

subtracting a coordinate of each point corresponding to the first training point cloud from the α -th first instance centroid vector of the corresponding point, to obtain an α -th first vector difference value of each point corresponding to the first training point cloud as an α -th first true centroid shift vector of each point corresponding to the first training point cloud; and

determining the α -th first shift loss according to the α -th first prediction centroid shift vector of each point corresponding to the first training point cloud and the α -th first true centroid shift vector of the corresponding point.

14. The method of claim 12 or 13, wherein the performing semantic prediction on the α -th first point cloud feature of each original point by using the first backbone network of the $(\alpha-1)$ -th update, to obtain the α -th first semantic prediction result comprises:

performing semantic prediction on the α -th first point cloud feature of each original point by using the first backbone network of the $(\alpha-1)$ -th update, to obtain an α -th initial semantic prediction point cloud; and

smoothing a semantic feature vector in each supervoxel in the α -th initial semantic prediction point cloud by means of average pooling, to obtain a fine-tuned α -th initial semantic prediction point cloud as the α -th first semantic prediction result.

15. The method of claim 13, wherein the determining the α -th first shift loss according to the α -th first prediction centroid shift vector of each point corresponding to the first training point cloud and the α -th first true centroid shift vector of the corresponding point comprises:

calculating a norm of a difference value between the α -th first prediction centroid shift vector of each point corresponding to the first training point cloud and the α -th first true centroid shift vector of the corresponding point, to obtain an α -th first difference value norm of each point corresponding to the first training point cloud;

averaging the α -th first difference value norm of each point corresponding to the first training point cloud to obtain an α -th first shift sub-loss;

calculating a ratio of the α -th first prediction centroid shift vector of each point corresponding to the first training point cloud to the norm of the α -th first prediction centroid shift vector of the corresponding point as an α -th first prediction ratio; calculating a ratio of the α -th first true centroid shift vector of each point corresponding to the first training point cloud to the norm of the α -th first true centroid shift vector as an α -th first actual ratio;

determining a product of the α -th first prediction ratio of each point corresponding to the first training point cloud and the α -th first actual ratio of the corresponding point, to obtain an α -th first direction product of each point corresponding to the first training point cloud;

averaging the α -th first direction product of each point corresponding to the first training point cloud to obtain an α -th second shift sub-loss; and

subtracting the α -th second shift sub-loss from the α -th first shift sub-loss, to obtain the α -th first shift loss.

16. The method of any of claims 11 to 15, wherein, after the using a sum of the α -th first semantic loss and the α -th first shift loss as an α -th first loss, the method further comprises:

in a case that the α -th first loss difference value is greater than a first loss difference value threshold, or the α -th first loss difference value is consecutively less than or equal to the first loss difference value threshold and the number of consecutive times is less than the first preset number, performing α -th update on the first backbone network of the $(\alpha-1)$ -th update according to the α -th first loss, to obtain the first backbone network of an α -th update.

17. The method of claim 10, wherein the performing instance segmentation on the multi-level prediction point cloud outputted by the second backbone network based on the third semantic prediction result and the third centroid shift vector, to obtain third training point cloud comprises:

moving each point in the multi-level prediction point cloud to a respective predicted center according to the third centroid shift vector, to obtain influence prediction point cloud;

clustering core points of the influence prediction point cloud according to the third semantic prediction result, to obtain at least one third core point cloud of each semantic category; and

clustering edge points in the influence prediction point cloud into third target core point cloud in the at least one third core point cloud of the corresponding semantic category, to obtain the third training point cloud.

18. An instance segmentation device, comprising:

a memory, configured to store a computer program; and

a processor;

wherein the processor is configured to execute the computer program to:

perform semantic prediction and centroid shift prediction on acquired to-be-segmented point cloud by using a first backbone network, to obtain a semantic prediction result and centroid shift vector of each point in the to-be-segmented point cloud, wherein the first backbone network and a second backbone network are trained based on first training point cloud, and the first training point cloud is determined by means of original semantic point cloud carrying a semantic label;

perform objectness prediction on the to-be-segmented point cloud by using the second backbone network, to obtain multi-level point cloud comprising multi-level objectness labels, wherein the multi-level objectness labels are configured to characterize a shape of each instance, internal information, and external information;

move each point in the multi-level point cloud to a respective predicted center according to the centroid shift vector, to obtain influence point cloud;

cluster core points of the influence point cloud according to the semantic prediction result, to obtain at least one core point cloud of each semantic category, wherein the core points are points having objectness labels at preset core levels; and

cluster edge points in the influence point cloud into target core point cloud in the at least one core point cloud of the corresponding semantic category, to obtain an instance segmentation result, wherein the edge points are points having objectness labels at preset edge levels.

19. A computer-readable storage medium, storing a computer program, wherein when executed by a processor, the computer program causes the processor to:

perform semantic prediction and centroid shift prediction on acquired to-be-segmented point cloud by using a first backbone network, to obtain a semantic prediction result and

centroid shift vector of each point in the to-be-segmented point cloud, wherein the first backbone network and a second backbone network are trained based on first training point cloud, and the first training point cloud is determined by means of original semantic point cloud carrying a semantic label;

perform objectness prediction on the to-be-segmented point cloud by using the second backbone network, to obtain multi-level point cloud comprising multi-level objectness labels, wherein the multi-level objectness labels are configured to characterize a shape of each instance, internal information, and external information;

move each point in the multi-level point cloud to a respective predicted center according to the centroid shift vector, to obtain influence point cloud;

cluster core points of the influence point cloud according to the semantic prediction result, to obtain at least one core point cloud of each semantic category, wherein the core points are points having objectness labels at preset core levels; and

cluster edge points in the influence point cloud into target core point cloud in the at least one core point cloud of the corresponding semantic category, to obtain an instance segmentation result, wherein the edge points are points having objectness labels at preset edge levels.

20. A computer program product, comprising a non-transitory computer-readable storage medium storing a computer program, wherein, when executed by a computer, the computer program causes the computer to:

perform semantic prediction and centroid shift prediction on acquired to-be-segmented point cloud by using a first backbone network, to obtain a semantic prediction result and centroid shift vector of each point in the to-be-segmented point cloud, wherein the first backbone network and a second backbone network are trained based on first training point cloud, and the first training point cloud is determined by means of original semantic point cloud carrying a semantic label;

perform objectness prediction on the to-be-segmented point cloud by using the second backbone network, to obtain multi-level point cloud comprising multi-level objectness labels, wherein the multi-level objectness labels are configured to characterize a shape of each instance, internal information, and external information;

move each point in the multi-level point cloud to a respective predicted center according to the centroid shift vector, to obtain influence point cloud;

cluster core points of the influence point cloud according to the semantic prediction result, to obtain at least one core point cloud of each semantic category, wherein the core points are points having objectness labels at preset core levels; and

cluster edge points in the influence point cloud into target core point cloud in the at least one core point cloud of the corresponding semantic category, to obtain an instance segmentation result, wherein the edge points are points having objectness labels at preset edge levels.