

Claims

What is claimed is:

1. A method for object recognition comprising:
obtaining a plurality of input points in three dimensional space wherein each point of the plurality of input points has at least one colour feature and at least one geometric feature;
selecting a set of query points from the plurality of input points;
determining, for each query point, a set of neighbouring points from the plurality of input points;
associating each query point with a respective input feature set and each neighbouring point of each query point with a respective input feature set;
determining, by a neural network, for each combination of query point and neighbouring point of the query point an attention value from the at least one colour feature and the at least one geometric feature of the query point and the neighbouring point;
determining, for each query point, an output feature set from the input feature sets of the neighbouring points of the query point in accordance with the attention values determined for the neighbouring points; and
performing object recognition based on the output feature set of the query points.
2. The method of claim 1, wherein the query points are selected from the plurality of input points using furthest point sampling.
3. The method of claim 2, wherein the neighbouring points of each query point are determined using k-nearest-neighbour with dilation.
4. The method of any one of claims 1 to 3, wherein the attention values are determined by at least one multi-layer perceptron.

5. The method of any one of claims 1 to 4, wherein determining the attention value for each combination of query point and neighbouring point of the query point comprises determining a colour-based attention value for the combination from the at least one colour feature of the query point and the neighbouring point and a geometric feature-based attention value for the combination from the at least one geometric feature of the query point and the neighbouring point and determining the attention value based on the colour-based attention value and the geometric feature-based attention value.
6. The method of claim 5, wherein determining the attention value for each combination of query point and neighbouring point of the query point further comprises determining a position-based attention value for the combination from the three-dimensional position of the query point and the three-dimensional position of the neighbouring point and determining the attention value based on the position-based attention value.
7. The method of any one of claims 5 or 6, wherein determining the attention value for each combination of query point and neighbouring point of the query point comprises determining an input feature-based attention value for the combination from the input feature set of the query point and the input feature set of the neighbouring point and determining the attention value based on the input feature-based attention value.
8. The method of claims 6 and 7, comprising determining the attention value by feeding the colour-based attention value, the geometric feature-based attention value, the position-based attention value and the input feature-based attention value to a partial attention value combining multi-layer perceptron.
9. The method of any one of claims 1 to 8, comprising determining, for each combination of query point and neighbouring point of the query point, the attention value from a difference of the at least one colour feature between the

query point and the neighbouring point and a difference of the at least one geometric feature between the query point and the neighbouring point.

10. The method of any one of claims 1 to 9, wherein determining the output feature set comprises max-pooling of the processed feature sets over the neighbouring points.
11. A method for object recognition comprising:
 - processing a point cloud in a sequence of neural network layers starting with a first neural network layer and ending with a last neural network layer, comprising, for each neural network layer;
 - obtaining a plurality of input points in three dimensional space for the neural network layer wherein each point of the plurality of input points has at least one colour feature and at least one geometric feature;
 - selecting a set of query points from the plurality of input points for the neural network layer;
 - determining, for each query point, a set of neighbouring points from the plurality of input points;
 - associating each query point and each neighbouring point of each query point with an input feature set for the neural network layer;
 - determining, for each combination of query point and neighbouring point of the query point an attention value from the at least one colour feature and the at least one geometric feature of the query point and the neighbouring point; and
 - determining, for each query point, an output feature set for the neural network layer from the input feature sets of the neighbouring points of the query point for the neural network layer for the neural network layer in accordance with the attention values determined for the neighbouring points.
12. The method of claim 11, wherein the method comprises performing object recognition based on the output feature set of the query points of the last neural network layer.

13. The method of claim 12, wherein the sequence of neural network layers implements an encoder and the method comprises performing the object recognition based on the output feature set of the query points of the last neural network layer by a decoder.
14. The method of any one of claims 11 to 13, wherein the input feature set of each point of the query points and the neighbouring points for the first neural network layer comprises the at least one colour feature and the at least one geometric feature of the point.
15. The method of any one of claims 11 to 14, wherein, for each neural network layer except for the first neural network layer, the input points comprise the query points of the neural network layer preceding the neural network layer in the sequence of neural network layers and the input feature set of each point of the query points and the neighbouring points comprises the output feature set for the query points of the neural network layer preceding the neural network layer in the sequence of neural network layers.
16. The method of any one of claims 11 to 15, wherein determining the output feature set comprises determining a processed feature set for each neighbouring point and determining the output feature set by combining the processed feature set of the neighbouring points.
17. The method of any one of claims 11 to 16, wherein determining the output feature set comprises max-pooling of the processed feature sets over the neighbouring points.
18. An object recognition device comprising a processor configured to perform the method of any one of claims 1 to 17.

19. A robotic control device comprising an object recognition device according to claim 18 and configured to control a robot device based on results of the object recognition.
20. A computer program element comprising program instructions, which, when executed by one or more processors, cause the one or more processors to perform the method of any one of claims 1 to 17.