Enhanced Vote Network for 3D Object Detection in Point Clouds

Min Zhong, Gang Zeng
Key Laboratory on Machine Perception, Peking University

Introduction:

- **3D point object detection** takes the point cloud as input and outputs the 3D bounding boxes and semantic classes of objects.
- Due to the sparse and unstructured nature of point clouds, encoding fine semantics and global context information are important for predicting the bounding boxes and its class

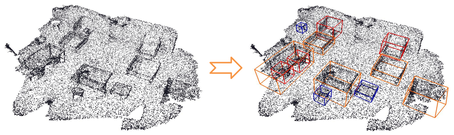

Fig. 1: 3D point object detection. The point cloud is sparse and unstructured.

The overall framework:

- In Fig.2, the input point cloud is feed into feature encoder to extract high-level features. Then, **Vote Feature Generation** module generates the vote features from the encoded features.
- To enhance vote features, **Semantic Feature Generation** module generates features with rich semantic information and **Semantic Aware** module combines it into the vote feature;
- **Object-sencetive Feature Generation** module outputs the object sensitive features that are used to aggregate the context for vote features by **Context Aggregation** module.
- Finally, the **Detection From Vote** module leverages the enhanced vote features to localize and classify the objects.

Feature Enhancement

- **Semantic Aware.** Semantic Feature Generation module generates semantic aware feature with a semantic segmentation loss, and Semantic Aware module add it to the vote feature to obtain the semantic-aware vote feature.
- **Context Aggregation.** we learn object-sensitive embedding with a discriminative loss which encourages points belong to the same object to lie close with each other, otherwise, lie far away from each other. Then the Context Aggregation Module produces an attention map use the embeddings and applies to the vote feature to aggregate context information.

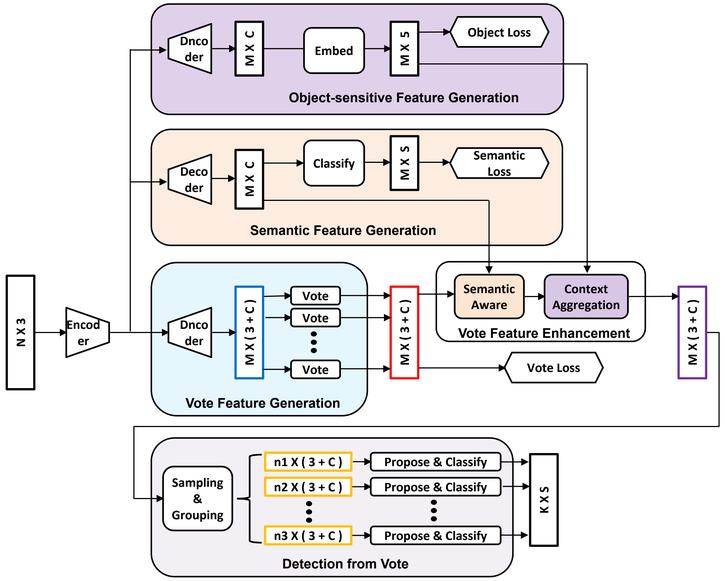

Fig. 2: The overall network.

Experiments:

- **Compare with other methods**.

	Input	mAP@0.25	mAP@0.5
DSS [4]	Geo + RGB	15.2	6.8
MRCNN 2D-3D [2]	Geo + RGB	17.3	10.5
F-PointNet [8]	Geo + RGB	19.8	10.8
GSPN [30]	Geo + RGB	30.6	17.7
3D-SIS [5]	Geo + 1 view	35.1	18.7
3D-SIS [5]	Geo + 3 views	36.6	19.0
3D-SIS [5]	Geo + 5 views	40.2	22.5
3D-SIS [5]	Geo only	25.4	14.6
VoteNet	Geo only	58.6	33.5
Ours	Geo only	**60.3**	**36.8**

Tab. 1: Compare with other methods on SUN RGB-D dataset.

- **Ablation studies on two modules**.

Method	+Context	+Semantic	mAP
Baseline			58.6
C-Module	✓		59.5
S-Module		✓	60.0
Full	✓	✓	**60.3**

Tab. 2: Ablation studies. C is Context Aggravation Module. S is Semantic-aware Feature Module.

- **Qualitative results**.

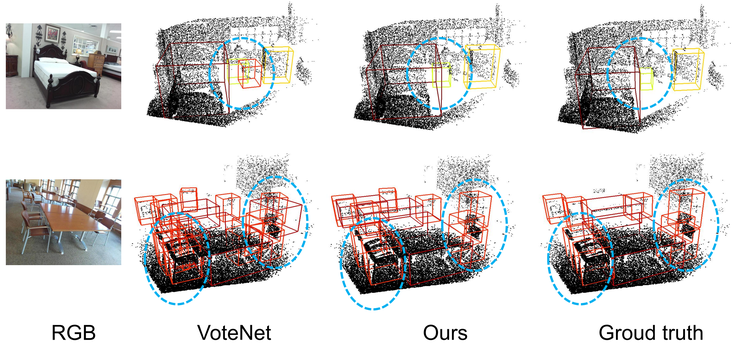

Fig. 4: Qualitative results on SUN RGB-D