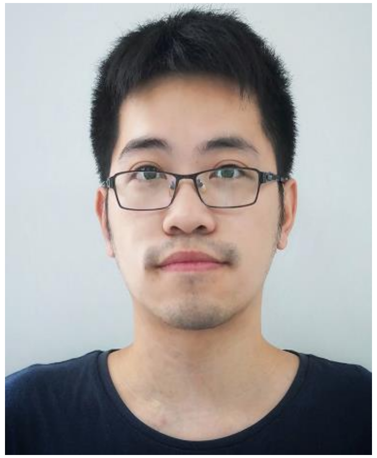
Point2Sequence: Learning the Shape Representation of 3D Point Clouds with an Attention-based Sequence to Sequence Network




Xinhai Liu[1] **Zhizhong Han**[1, 2] **Yu-Shen Liu**[1] **Matthias Zwicker**[2]

[1]School of Software, Tsinghua University, Beijing, China

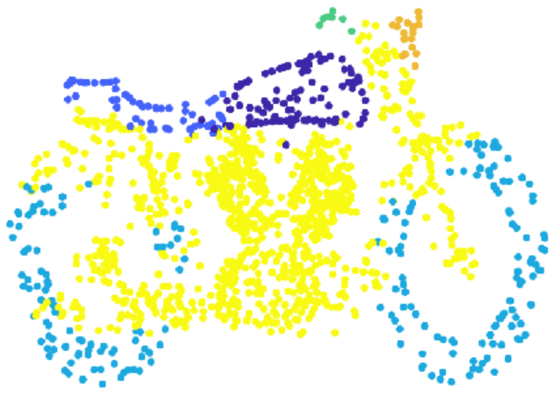
[2]Department of Computer Science, University of Maryland, College Park, USA

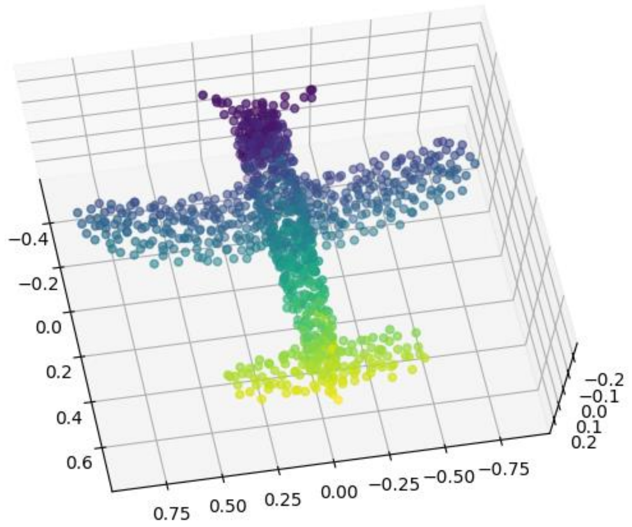
Content

- Introduction
- Motivation
- Technical details
- Experiment
- Contributions

Introduction

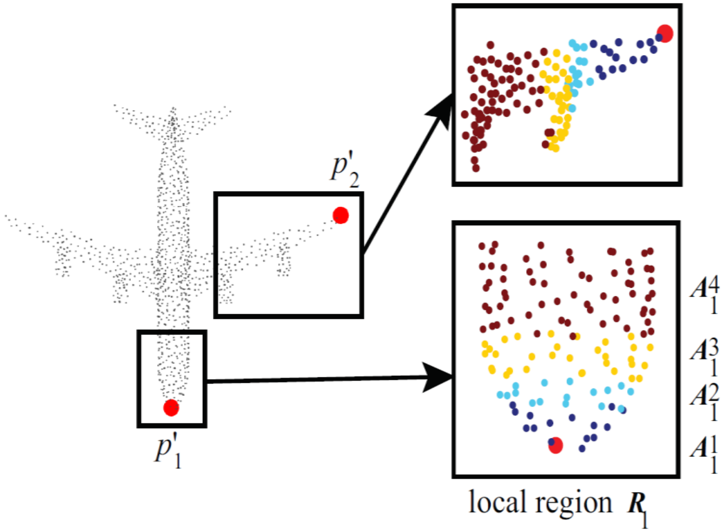
- Point cloud is considered as one of the simplest 3D shape representations
- 3D sensors: LiDAR, conventional cameras, or RGB-D cameras
- Applications: 3D modeling, autonomous driving, indoor navigation and robotics





Motivation

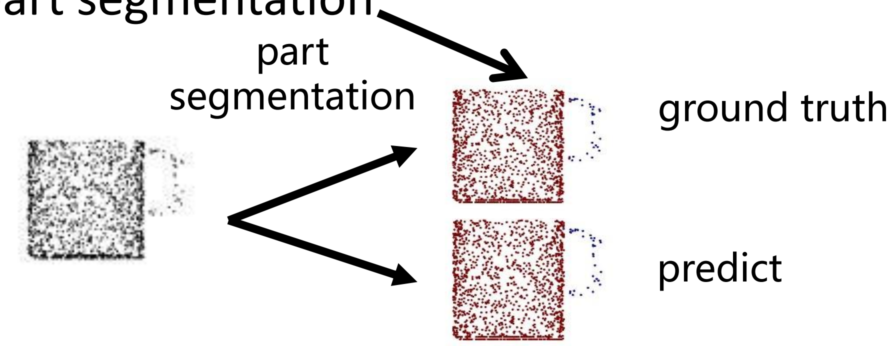
- Existing studies: encode contextual information of local regions in hand-crafted or explicit ways
- Capture fine-grained contextual information, such as the correlation between different areas in a local region.
- Point2Sequence: capture the correlations by aggregating multi-scale areas of each local region with attention



Technical details

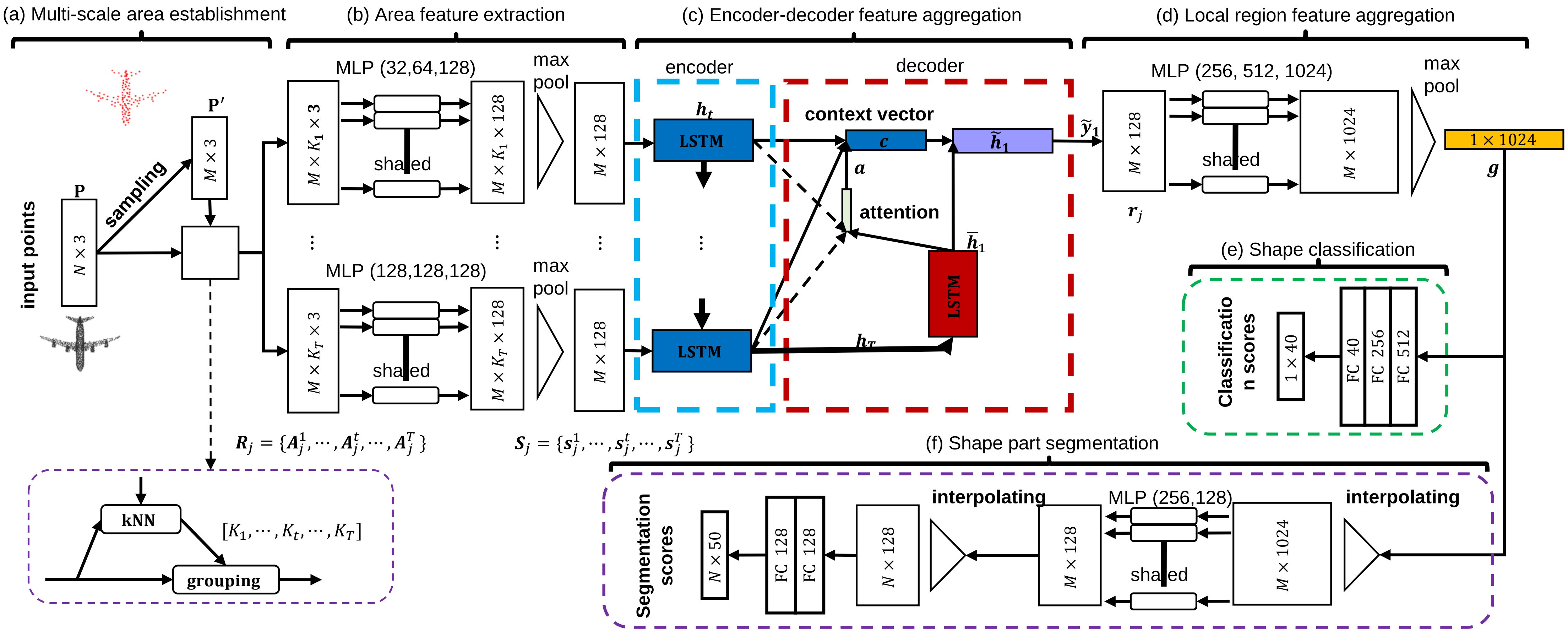
- Overview
  - focuses on capturing the contextual information of local regions
  - embeds the feature of multi-scale areas in each local region by RNN-based model
  - employs an attention mechanism to highlight the importance of different scale areas
  - has a wide range of applications, e.g.
    - 3D shape classification

    - 3D shape part segmentation



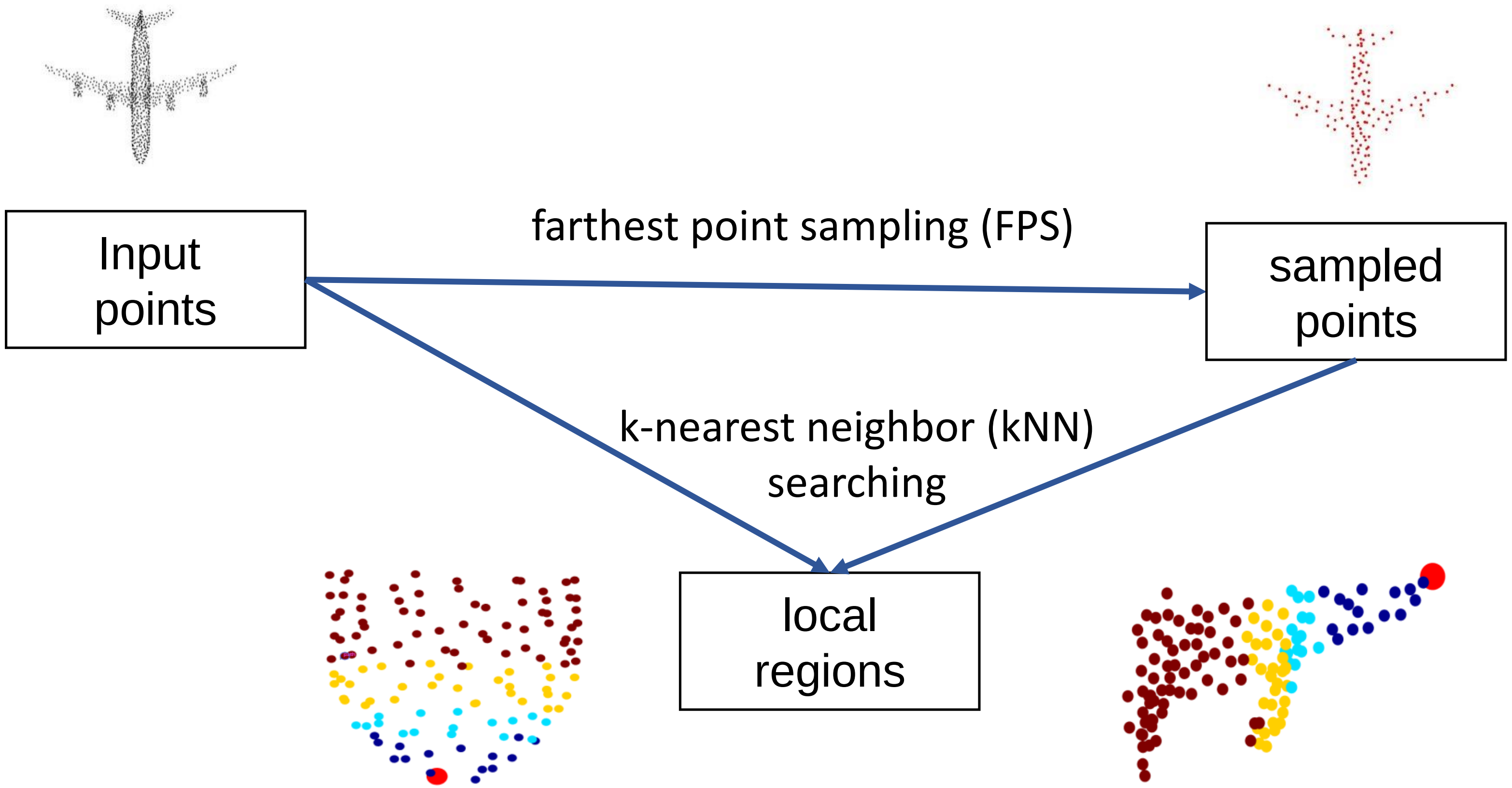
Technical details

- Framework



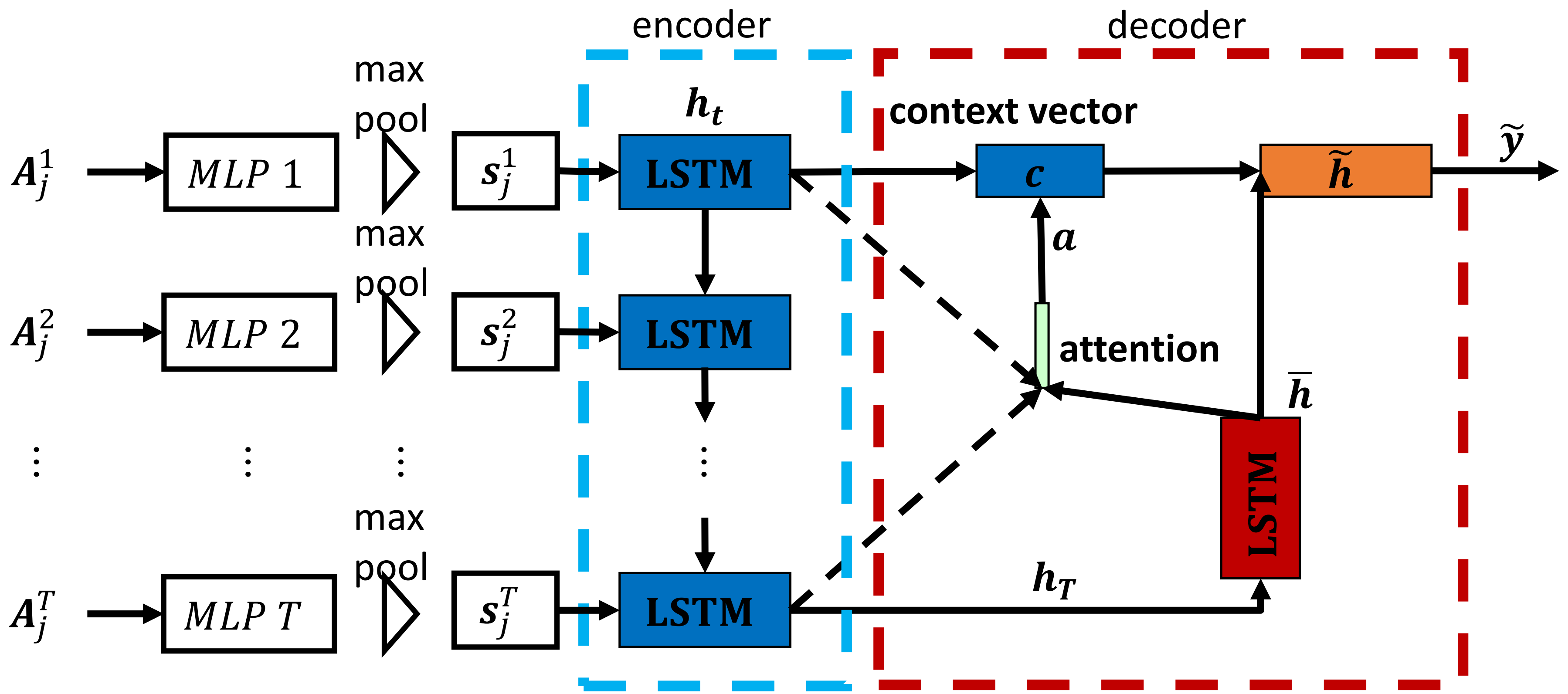
Technical details

- Multi-scale area establishment



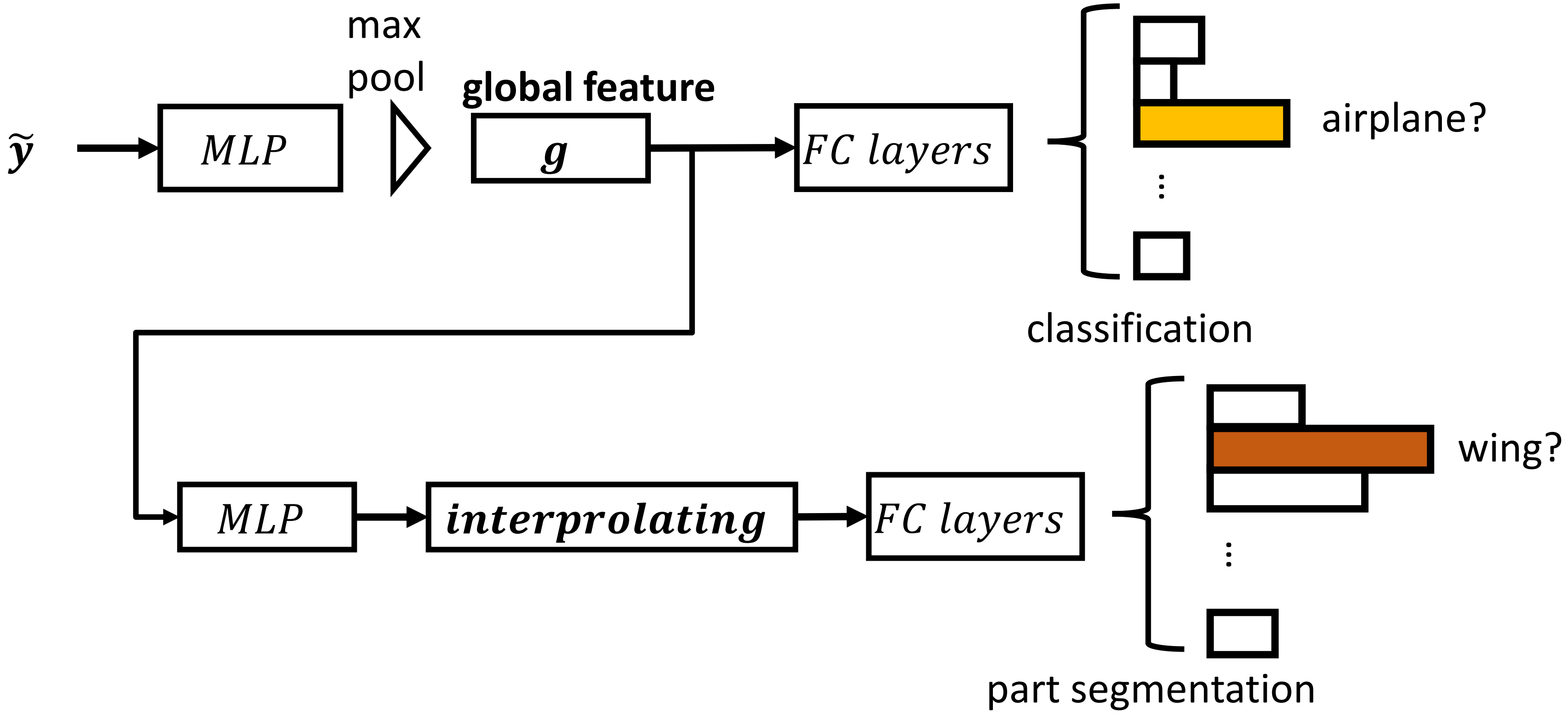
Technical details

- Area feature extraction & Encoder-decoder feature aggregation



Technical details

- Shape classification & shape part segmentation



Experiments

Table 3: The effect of the number of sampled points M on ModelNet40.

M	128	256	384	512
Acc (%)	91.86	92.34	**92.54**	91.86

Table 4: The effects of the type of RNN cell (CT) and hidden state dimension h on ModelNet40.

Metric	RT=LSTM	GRU	h=64	256
Acc (%)	**92.54**	92.18	92.46	92.18

Table 5: The effects of the attention mechanism (Att) and decoder (Dec) on ModelNet40.

Metric	Att+ED	No Att	No Dec	Con	MP
Acc (%)	**92.54**	92.26	92.42	92.06	91.73

Table 6: The effect of the number of scales T on ModelNet40.

T	4	3	2	1
Acc (%)	92.54	92.46	**92.63**	91.94

Table 1: The shape classification accuracy (%) comparison on ModelNet10 and ModelNet40.

Method	Input	ModelNet10		ModelNet40	
		Class	Instance	Class	Instance
PointNet (Qi et al. 2017b)	1024×3	-	-	86.2	89.2
PointNet++ (Qi et al. 2017c)	1024×3	-	-	-	90.7
ShapeContextNet (Xie et al. 2018)	1024×3	-	-	87.6	90.0
Kd-Net (Klokov and Lempitsky 2017)	$2^{15} \times 3$	93.5	94.0	88.5	91.8
KC-Net (Shen et al. 2018)	1024×3	-	94.4	-	91.0
PointCNN (Li et al. 2018)	1024×3	-	-	-	91.7
DGCNN (Wang et al. 2018)	1024×3	-	-	90.2	92.2
SO-Net (Li, Chen, and Lee 2018)	2048×3	93.9	94.1	87.3	90.9
Ours	1024×3	**95.1**	**95.3**	**90.4**	**92.6**

Experiments

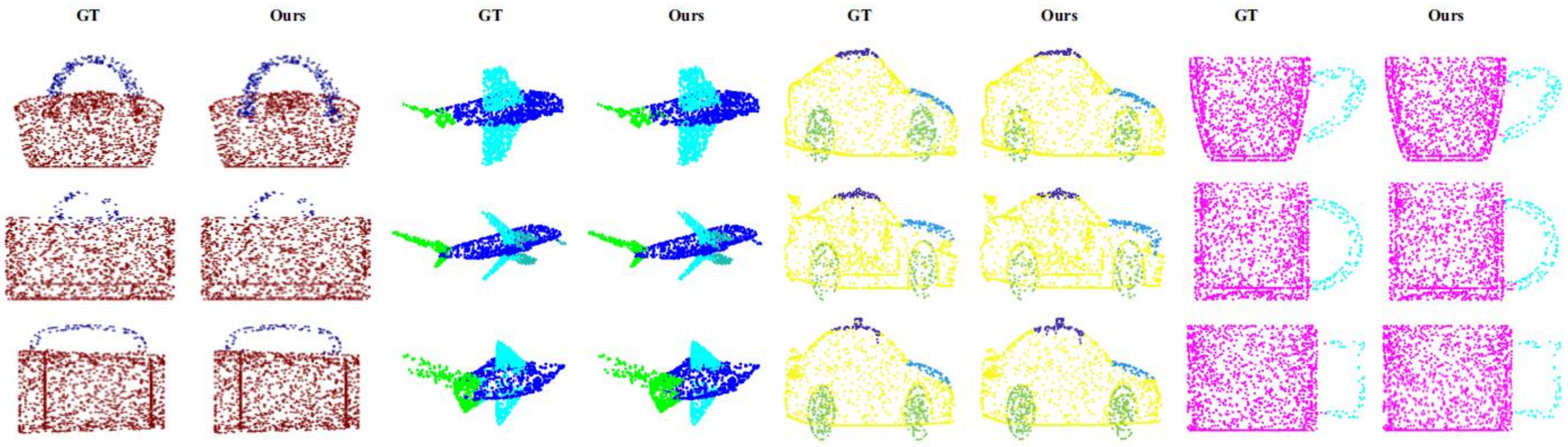


Table 2: The accuracies (%) of part segmentation on ShapeNet part segmentation dataset.

	mean	Intersection over Union (IoU) air.	bag	cap	car	cha.	ear.	gui.	kni.	lam.	lap.	mot.	mug	pis.	roc.	ska.	tab.
# SHAPES		2690	76	55	898	3758	69	787	392	1547	451	202	184	283	66	152	5271
PointNet	83.7	83.4	78.7	82.5	74.9	89.6	73.0	91.5	85.9	80.8	95.3	65.2	93.0	81.2	57.9	72.8	80.6
PointNet++	85.1	82.4	79.0	87.7	77.3	90.8	71.8	91.0	85.9	83.7	95.3	**71.6**	94.1	81.3	58.7	**76.4**	82.6
ShapeContextNet	84.6	83.8	80.8	83.5	**79.3**	90.5	69.8	**91.7**	86.5	82.9	**96.0**	69.2	93.8	82.5	**62.9**	74.4	80.8
Kd-Net	82.3	80.1	74.6	74.3	70.3	88.6	73.5	90.2	87.2	81.0	94.9	57.4	86.7	78.1	51.8	69.9	80.3
KCNet	84.7	82.8	81.5	86.4	77.6	90.3	76.8	91.0	87.2	**84.5**	95.5	69.2	94.4	81.6	60.1	75.2	81.3
DGCNN	85.1	**84.2**	**83.7**	84.4	77.1	**90.9**	**78.5**	91.5	**87.3**	82.9	**96.0**	67.8	93.3	**82.6**	59.7	75.5	82.0
SO-Net	84.9	82.8	77.8	**88.0**	77.3	90.6	73.5	90.7	83.9	82.8	94.8	69.1	94.2	80.9	53.1	72.9	**83.0**
Ours	**85.2**	82.6	81.8	87.5	77.3	90.8	77.1	91.1	86.9	83.9	95.7	70.8	**94.6**	79.3	58.1	75.2	82.8

Contributions

- Capture the correlation between different areas in a local region
- Introduce an attention mechanism to highlight the importance of different scale areas
- Our outperforming results verify the feasibility of RNNs to effectively understand point clouds.