APSNet: Attention Based Point Cloud Sampling

Yang Ye, Xiulong Yang and Shihao Ji
Department of Computer Science, Georgia State University, USA

BMVC 2022
London, UK
Nov 21st - 24th

Introduction

This paper introduces APSNet, an attention-based autoregressive network for task-oriented 3D point clouds sampling, which aims to sample a subset of points that are tailored specifically to a downstream task of interest. APSNet employs a sequential autoregressive generation with a novel LSTM-based sequential model for sampling. Depending on the availability of labeled training data, APSNet can be trained in supervised learning or self-supervised learning via knowledge distillation. We also present a joint training of APSNet, yielding a single compact model that can generate arbitrary length of samples with prominent performances. Extensive experiments demonstrate the superior performance of APSNet against state-of-the-arts in various downstream tasks, including 3D point cloud classification, reconstruction, and registration.

Method

Given original point cloud $\boldsymbol{P}$, the goal of APSNet is to generate a point cloud $\boldsymbol{Q} = f_{\boldsymbol{\theta}}(\boldsymbol{P})$ to maximize the predictive performance of task network $\boldsymbol{T}$. The parameters of APSNet, $\boldsymbol{\theta}$, are optimized by minimizing a task loss and a sampling loss jointly as

$$\min_{\boldsymbol{\theta}} \ell_{task}(T(\boldsymbol{Q}), y) + \lambda L_{sample}(\boldsymbol{Q}, \boldsymbol{P})$$

The sampling loss L_{sample} encourages the sampled points in $\boldsymbol{Q}$ to be close to those of $\boldsymbol{P}$ and also have a maximal coverage w.r.t. $\boldsymbol{P}$.

- Sampling loss

$$L_{sample}(\boldsymbol{Q}, \boldsymbol{P}) = L_a(\boldsymbol{Q}, \boldsymbol{P}) + \beta L_m(\boldsymbol{Q}, \boldsymbol{P}) + (\gamma + \delta|\boldsymbol{Q}|) L_a(\boldsymbol{P}, \boldsymbol{Q})$$

- Average nearest neighbor loss

$$L_a(\boldsymbol{S}_1, \boldsymbol{S}_2) = \frac{1}{|\boldsymbol{S}_1|} \sum_{s_1 \in \boldsymbol{S}_1} \min_{s_2 \in \boldsymbol{S}_2} \|s_1 - s_2\|_2^2.$$

Maximal nearest neighbor loss

$$L_m(\boldsymbol{S}_1, \boldsymbol{S}_2) = \max_{s_1 \in \boldsymbol{S}_1} \min_{s_2 \in \boldsymbol{S}_2} \|s_1 - s_2\|_2^2$$

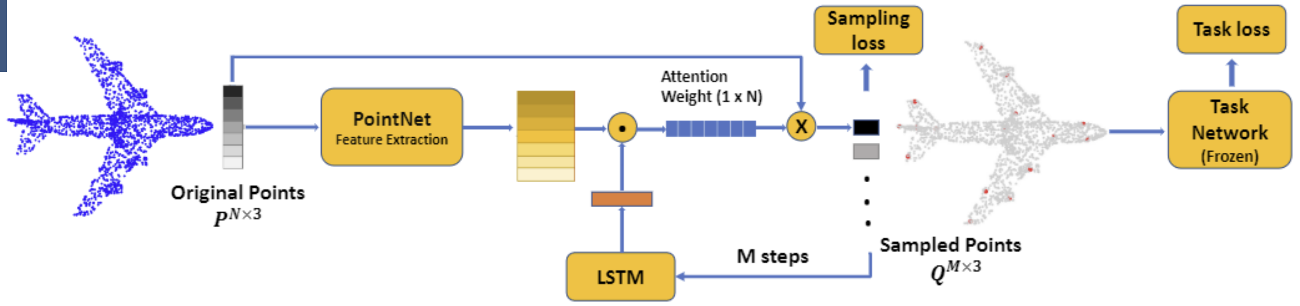

Fig. 1 Overview of APSNet. APSNet first extracts features with a simplified PointNet that preserves the geometric information of a point cloud. Then, an LSTM with attention mechanism is used to capture the relationship among points and select the most informative point sequentially. Finally, the sampled point cloud is fed to a task network for prediction. The whole pipeline is optimized by minimizing a task loss and a sampling loss jointly.

- Self-supervised Training with Knowledge Distillation

The task network $\boldsymbol{T}$ is the teacher model, and APSNet is the student model and use the soft predictions of $\boldsymbol{T}$ as the targets to train APSNet. In this scenario,

$$\ell_{task}(T(\boldsymbol{Q}), \tilde{y}), \text{ with } \tilde{y} = T(\boldsymbol{P})$$

- Joint Training

Given the autoregressive model of our method, APSNet can generate arbitrary length of samples from a single mode. We can train one APSNet with different sample sizes by

$$L_{joint} = \sum_{c \in C_s} \left(\ell_{task}(T(\boldsymbol{Q}_c), y) + \lambda L_{sample}(\boldsymbol{Q}_c, \boldsymbol{P}) \right)$$

where C_s is a set of sample sizes of interest.

3. Inference Time

m	32	128	256	512
SampleNet-G	7.63	7.54	7.79	7.94
SampleNet-M*	44.33	135.23	261.47	515.30
APSNet-G*	9.21	12.84	17.68	27.48
APSNet-M	45.91	139.83	269.40	525.38

Experimental Results

1. Classification

m	RS	FPS	DaNet	MOPS-Net		SampleNet			APSNet		APSNet-KD	
				G	M	G	M	M*	G	M	G	M
8	8.26	23.29	-	-	-	78.36	73.31	28.7	**81.42**	74.12	80.22	73.81
16	25.11	54.19	-	**84.7**	51.2	80.60	79.68	55.5	83.89	82.25	83.82	82.02
32	55.19	77.32	85.1	86.1	77.6	80.32	82.97	74.4	88.15	86.97	**88.76**	84.95
64	78.26	87.22	86.8	87.1	81.0	79.36	84.01	79.0	88.38	87.58	**88.66**	87.54
128	85.95	88.76	86.8	87.2	85.0	85.52	87.17	79.7	89.22	**89.38**	87.83	88.01
256	88.80	89.30	87.2	87.4	86.7	87.43	89.58	83.4	89.54	**89.86**	88.02	88.21
512	89.66	89.87	-	88.3	88.3	88.01	90.18	88.2	89.78	**90.18**	88.69	88.56

Classification accuracies with different sample sizes m on ModelNet40.

2. Reconstruction

m	RS	FPS	SampleNet			APSNet		APSNet-KD	
			G	M	M*	G	M	G	M
8	21.85	12.79	5.29	5.48	-	**4.27**	4.59	4.69	4.98
16	13.47	7.25	2.78	2.89	-	**2.51**	2.62	2.57	2.67
32	8.16	3.84	1.68	1.71	2.32	1.54	1.59	**1.47**	1.52
64	4.54	2.23	1.32	1.27	1.33	**1.07**	1.11	1.12	1.14

The normalized reconstruction errors with different sample sizes m on the ShapeNet Core55 dataset.

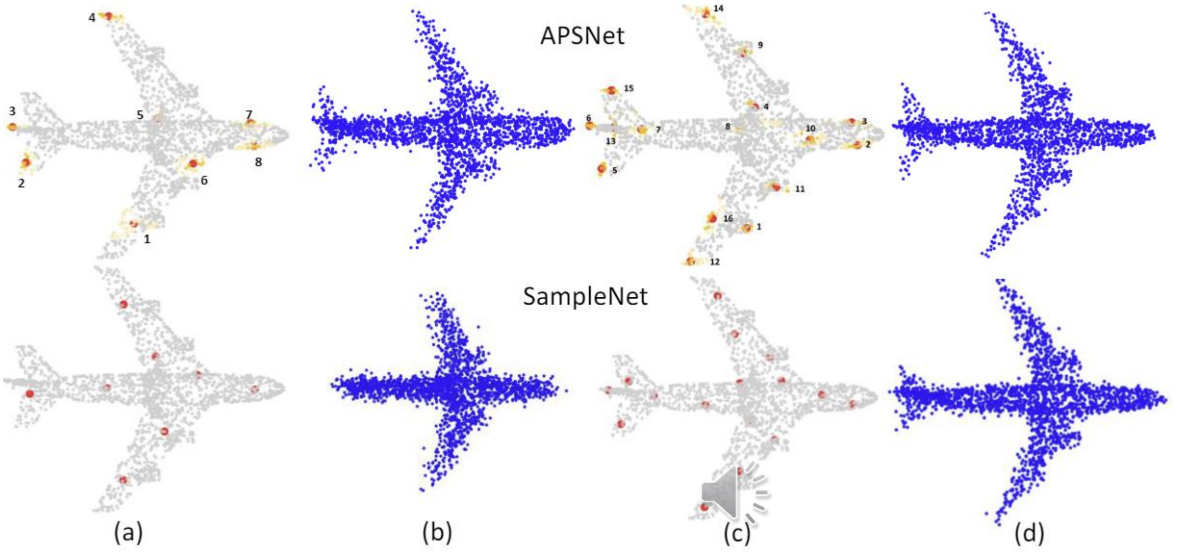

Fig. 2 Visualization of sampled points and reconstructed point clouds. APSNet focuses more on the outline of the airplane without losing details, which are critical for the reconstruction

Code: https://github.com/Yangyeeee/APSNet