Variational Object Point Cloud Encoder (VOPCE)

Shuo Jiang
Northeastern University
jiang.shuo@husky.neu.edu

Abstract

In this paper, we proposed an object encoding algorithm on point cloud data. Such algorithm can embed an object as point cloud with any number of points into a latent vector with fixed length.

1 Method

We model the point cloud data as a set contains multiple points $X = \{x_1, x_2, \cdots, x_N\}$, N can be arbitrary. Each data point $x_i \in \mathbb{R}^3$ for x, y, z positions. The latent vector is represented by a random vector with fixed length $z \sim Z \in \mathbb{R}^d$.

1.1 Likelihood Model

The likelihood probability $p_\theta (X|Z)$ can be model as a Gaussian Mixture Model (GMM) whose parameter is conditioned on given latent variable z. We use a Mixture Density Network (MDN) [1] to model such GMM. The neural network incorporate latent representation z and generate mean μ_k, variance σ_k and proportion π_k for each cluster k as vector of dimension l. We can then sample N times by the parameters given by MDN as a GMM. GMM is shown to be effective representing 3D point cloud data [2], in this work, we use neural network to represent it. Such model can be trained by maximizing the log likelihood function

$$\underset{\theta}{argmax} L(\theta) = \log p_\theta (X|z) = \frac{1}{N} \sum_{i=1}^{N} \log p_\theta (x_i|z) \tag{1}$$

Such maximization can be done by gradient descent method. The likelihood for each x_i is given by

$$p_\theta (x_i|z) = \log \sum_{k=1}^{K} \pi_k \cdot \frac{1}{\sqrt{(2\pi)^l |\sigma_k \mathbf{I}|}} \exp \left[-\frac{1}{2} (x_i - \mu_k)^T \sigma_k^{-1} \mathbf{I} (x_i - \mu_k) \right] \tag{2}$$

1.2 Posterior Model

The function of Posterior Model $q_\phi (Z|X)$ is to encode a point cloud set to latent vector. We employed the DeepSets Network for such purpose [4]. The network includes permutation invariant layers to process set data. The output of the network is the mean μ_z and variance σ_z of z as z is modelled by a multivariate normal distribution.

1.3 Variational Lower Bound

We train the model with one Likelihood Model and one Posterior Model end-to-end by maximizing Variational Lower Bound.

$$ELBO (\theta, \phi, X) = \mathbb{E}_{q_\phi(z|X)} [\log p_\theta (X|z)] - D_{KL} (q_\phi (z|X) || p(z)) \tag{3}$$

Preprint. Under review.

The first term can be calculated by 1 and 2. The prior of z in KL divergence is a multivariate normal distribution with zero mean value and identity matrix as covariance. The KL divergence term has a simple form as

$$D_{KL}\left(q_{\phi}\left(z|X\right)||p\left(z\right)\right)=\frac{1}{2}\sum_{m}\left[1+\log\left(\left(\sigma_{z}^{(m)}\right)^{2}\right)-\left(\mu_{z}^{(m)}\right)^{2}-\left(\sigma_{z}^{(m)}\right)^{2}\right] \tag{4}$$

m is the dimension of μ_z and σ_z.

2 Results

We used some ModelNet [3] point cloud data as our baseline of reconstruction. As shown in Figure

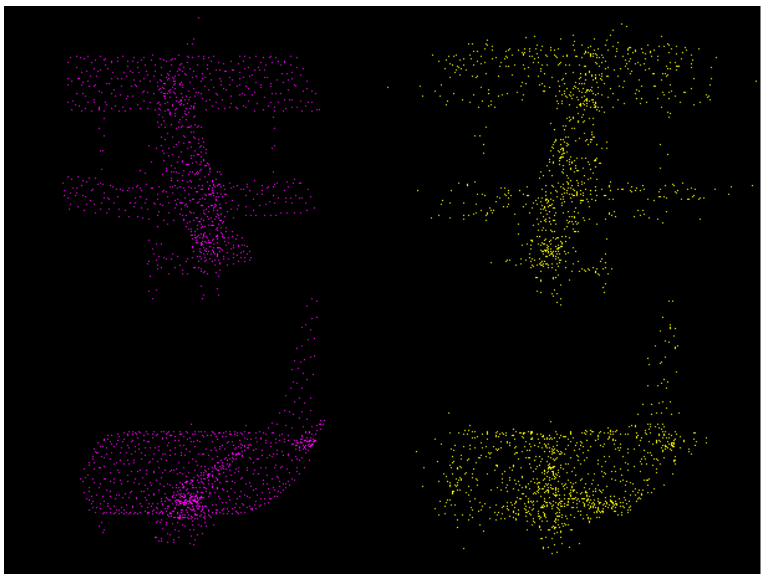

Figure 1: Reconstruction of ModelNet data

1, our model catches the point cloud well. We used a 100 dimension of latent variable z.
Use the code here
https://github.com/Bigpig4396/Variational-Object-Point-Cloud-Encoder-VOPCE

References

[1] Christopher M Bishop. Mixture density networks. 1994.

[2] Benjamin Eckart, Kihwan Kim, Alejandro Troccoli, Alonzo Kelly, and Jan Kautz. Accelerated generative models for 3d point cloud data. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5497–5505, 2016.

[3] Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 3d shapenets: A deep representation for volumetric shapes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1912–1920, 2015.

[4] Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Russ R Salakhutdinov, and Alexander J Smola. Deep sets. In *Advances in neural information processing systems*, pages 3391–3401, 2017.