Energy-Based Residual Latent Transport for Unsupervised Point Cloud Completion

Ruikai Cui[1] Shi Qiu [1,2] Saeed Anwar[1,2], Jing Zhang[1], Nick Barnes[1]

[1]Australian National University [2]Data61-CSIRO

Introduction

Point cloud data is a fundamental representation of 3D geometry, contributing to numerous applications in robotics, auto-navigation, augmented reality, *etc*. Limited by viewing angle, occlusion, and acquisition resolution, raw point clouds are generally sparse and incomplete.

Figure 1. Incomplete point cloud samples

Unsupervised point cloud completion aims to infer the whole geometry of a partial object observation without requiring partial-complete correspondence.

Motivation

- **Generative Modelling** Existing method leverages a one-to-one deterministic mapping to complete a partial shape. However, recovering missing geometries corresponds to a one-to-many stochastic mapping, so we proposed the first unsupervised completion method based on generative modelling.
- **Latent Transport** We assume that representations of a partial point cloud and its valid complete forms are close in the latent space, so we designed a latent-space energy-based model (EBM) in an encoder-decoder architecture, aiming to learn a probability distribution conditioned on the partial shape encoding.
- **Residual Sampling** To train the latent code transport module and the encoder-decoder network jointly, we introduce a residual sampling strategy, where the residual captures the domain gap between partial and complete shape latent spaces.

Contribution

Our main contributions are summarized below:

- We propose a novel energy-based latent transport mechanism, enabling generative modeling of the unsupervised point cloud completion task for the first time.
- We present a residual sampling strategy that allows joint training of a latent-space EBM and an encoder-decoder for the first time.
- Experimental results indicate that our model not only achieves state-of-the-art performance on synthetic (ShapeNet) and real-world (KITTI, ScanNet, MatterPort3D) datasets, but is also capable of generating explainable uncertainty maps.

Generative Modelling Framework

Let $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ be samples from the partial and complete point cloud domain, respectively. Our model consists of three main parts, namely

1. an encoder for point cloud code extraction and a decoder for point cloud reconstruction.
2. a latent-space energy-based model with energy function E_θ to fill the code gap between the partial point cloud and its corresponding completion.
3. a point domain discriminator to achieve adversarial learning.

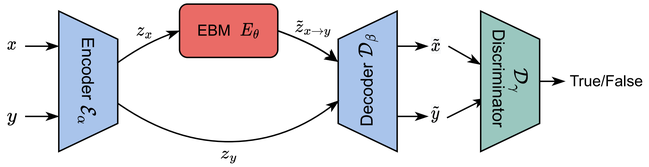

Figure 2. Model overview

Latent-Space Energy Based Model and Residual Sampling

Challenges and solution of applying EBM for unsupervised point cloud completion

To define the conditional distribution of the complete shape latent code, we use an energy function $E_\theta(z_y, z_x)$ parameterized by a deep neural network to map the code pair (z_y, z_x) to a scalar that measures their compatibility. However, in the unsupervised setting, we are incapable of directly modelling the compatibility of z_x and z_y due to lack of paired samples. Alternatively, we model the conditional distribution of the residual r_{xy} instead, representing the distribution gap between the latent space of the two domains as:

$$p_\theta(r_{xy}|z_x) = \frac{p_\theta(r_{xy}, z_x)}{\int_\theta p_\theta(r_{xy}, z_x) dz_x} = \frac{\exp[-E_\theta(r_{xy}, z_x)]}{Z(r_{xy}; \theta)}. \quad (1)$$

Joint Training of EBM and the Encoder-Decoder Framework

Given the above formulation, a residual r_{xy} can be generated from the conditional distrubution by sampling with Langevin dynamics, then a complete shape code $\tilde{z}_{x\to y}$ can be obtained as:

$$\tilde{z}_{x\to y} = z_x + \Omega(r_{xy}) \quad (2)$$

Here, we apply the stop gradient operation ($\Omega(\cdot)$) to avoid unfolding the Langevin dynamics iteration and involving the second-order gradient of r_{xy} in future gradient computation. Moreover, the parameter θ of the energy function can be updated with gradient defined as:

$$\Delta\theta = \frac{1}{N}\sum_{i=1}^{N} \mathbb{E}_{\tilde{z}_{x\to y}\sim p_\theta}[\frac{\partial}{\partial\theta} E_\theta(\tilde{z}^i_{x\to y})] - \mathbb{E}_{z_y\sim p_{z_y}}[\frac{\partial}{\partial\theta} E_\theta(z_y)] + C, \quad (3)$$

where C is a constant which can be ignored during training.

Experiments

Table 1. Shape Completion results of supervised (upper three rows) and unsupervised (lower five rows) methods on ShapeNet. The numbers shown are ℓ^2 CD↓ scaled by 10^4.

Method	Avg.	Plane	Cabinet	Car	Chair	Lamp	Sofa	Table	Boat
3D-EPN	29.1	60.0	27.0	24.0	16.0	38.0	45.0	14.0	9.0
FoldingNet	9.2	2.4	8.5	7.2	10.3	14.1	9.1	13.6	8.8
PCN	7.6	2.0	8.0	5.0	9.0	13.0	8.0	10.0	6.0
PcI2PcI	17.4	4.0	19.0	10.0	20.0	23.0	26.0	26.0	11.0
C4C.	14.3	3.7	12.6	8.1	14.6	18.2	26.2	22.5	8.7
ShapeInv.	23.6	4.3	20.7	11.9	20.6	25.9	54.8	38.0	12.8
Cai *et al.*	13.6	3.5	12.2	9.0	12.1	17.6	26.0	19.8	8.6
Ours	**9.4**	**2.3**	**12.2**	**5.8**	**12.0**	**12.8**	**10.3**	**13.8**	**5.7**

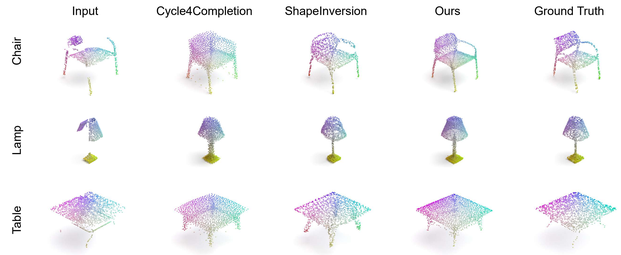

Figure 3. Qualitative Result on the ShapeNet

Our framework can provide uncertainty maps that summarize the stochasticity.

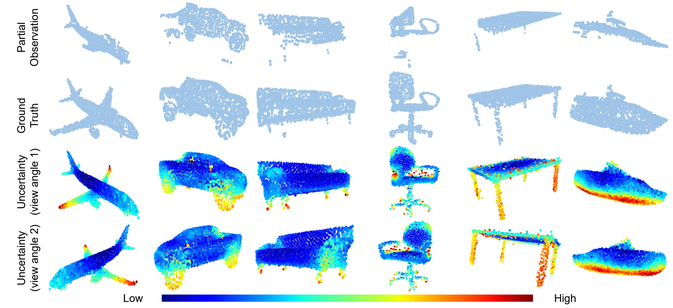

Figure 4. Uncertainty maps. The first and second rows show partial observations and their complete ground truth, and the third and fourth rows show two views of uncertainty maps.