Project page

IPoD: Implicit Field Learning with Point Diffusion for Generalizable 3D Object Reconstruction from Single RGB-D Images

CVPR2024 (**Highlight**)

Yushuang Wu, Luyue Shi, Junhao Cai, Weihao Yuan, Lingteng Qiu

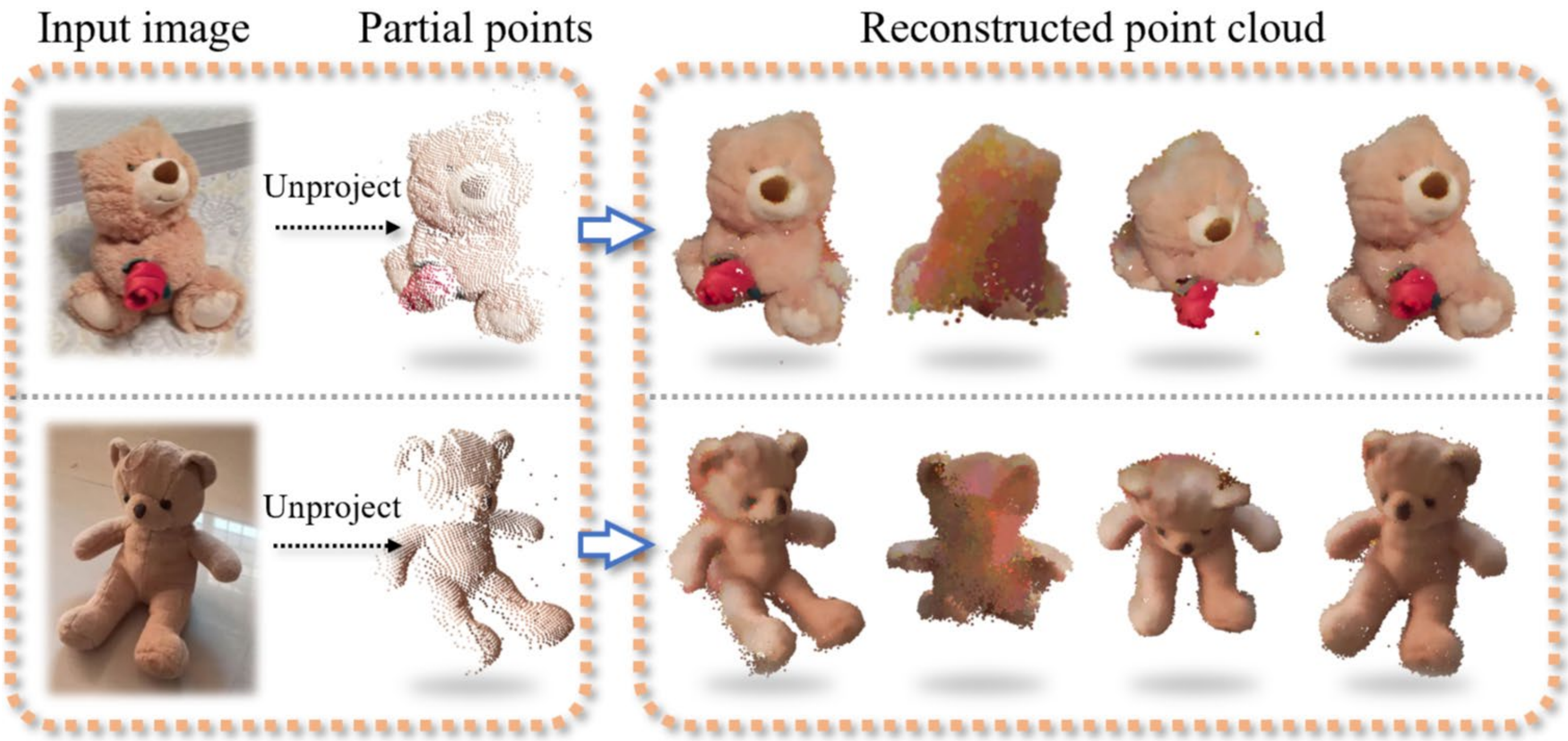
Zilong Dong, Liefeng Bo, Shuguang Cui, Xiaoguang Han

香港中文大學(深圳)
The Chinese University of Hong Kong, Shenzhen

Alibaba Cloud

Introduction

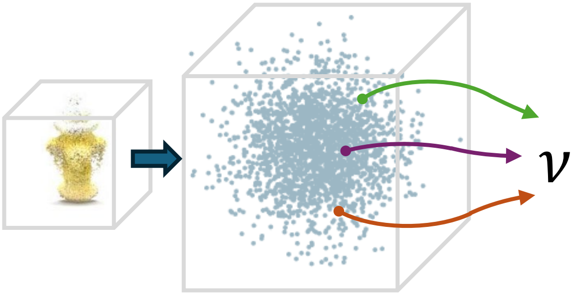
- Task: 3D object reconstruction from single-view RGB-D images



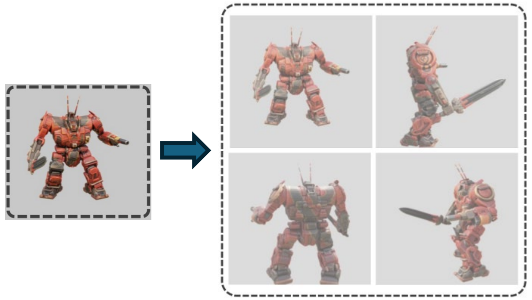
香港中文大學(深圳)
The Chinese University of Hong Kong, Shenzhen

Introduction

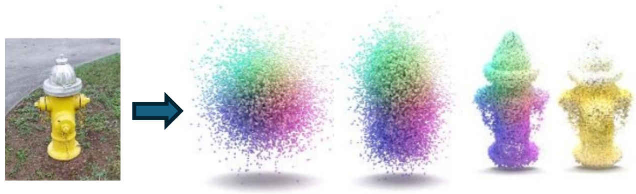
- Background



Implicit Field Learning:
MCC, NU-MCC



2D Multi-view Diffusion:
ImageDream, One2345



3D Point Diffusion:
PC^2, PVD

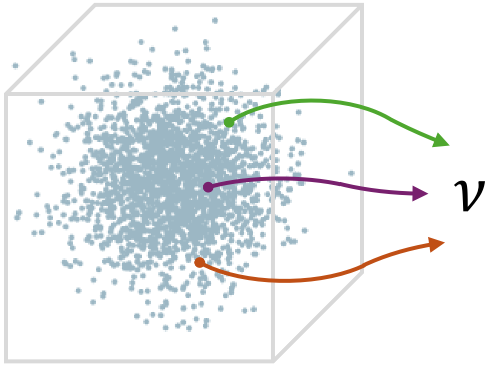
香港中文大學(深圳)
The Chinese University of Hong Kong, Shenzhen

Introduction

- Motivation:

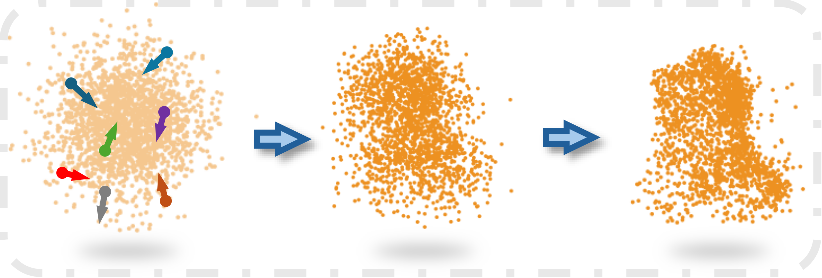
Input

Implicit Field Learning



Classic: random query sampling

VS



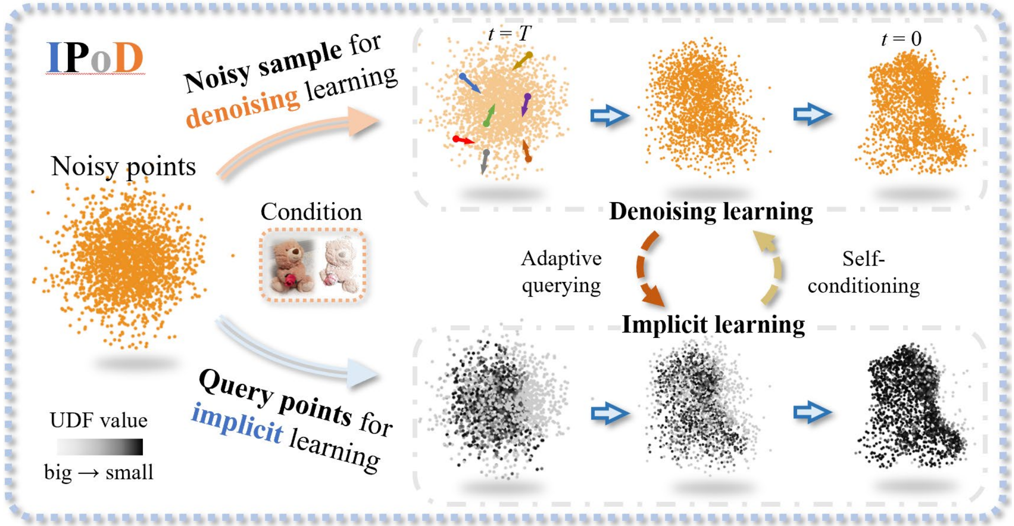
Ours: adaptive query sampling

香港中文大學(深圳)
The Chinese University of Hong Kong, Shenzhen

Introduction

- **IPoD**: **I**mplicit Field Learning with **Po**int **D**iffusion

D ⇨ I: Queries in implicit learning are view as a whole point cloud that can be adapted to the target shape via point denoising learning.



I ⇨ D: Implicit predictions at each point serve as self-condition to provide point-wise guidance for point diffusion-denoising learning.

The implicit field learning and diffusion-denoising learning in IPoD form a **cooperative** system!

香 港 中 文 大 學 (深 圳)
The Chinese University of Hong Kong, Shenzhen

CVPR SEATTLE, WA JUNE 17-21, 2024

Methodology

- Preliminary

Implicit field learning:

$$f_\theta(Q \mid P, I) \rightarrow \nu$$

$$\mathcal{L}_{\text{imp}} = \left\| f_\theta(Q \mid P, I) - \nu \right\|_1$$

Diffusion learning:

$$g_\theta(X_t, t \mid P, I) \rightarrow \epsilon$$

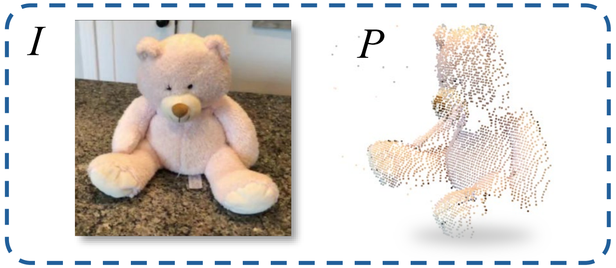
$$\mathcal{L}_{\text{diff}} = \left\| g_\theta(X_t, t \mid P, I) - \epsilon \right\|_2$$

Ours:

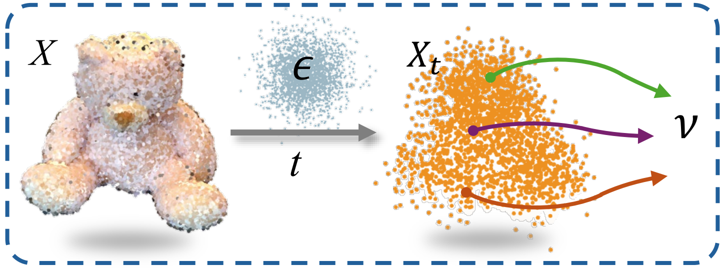
$$h_\theta(X_t, t \mid P, I) \rightarrow (\epsilon, \nu)$$

$$\mathcal{L}_{\text{uni}} = \left\| \nu' - \nu \right\|_1 + \lambda \left\| \epsilon' - \epsilon \right\|_2$$

Input: image and seen point cloud



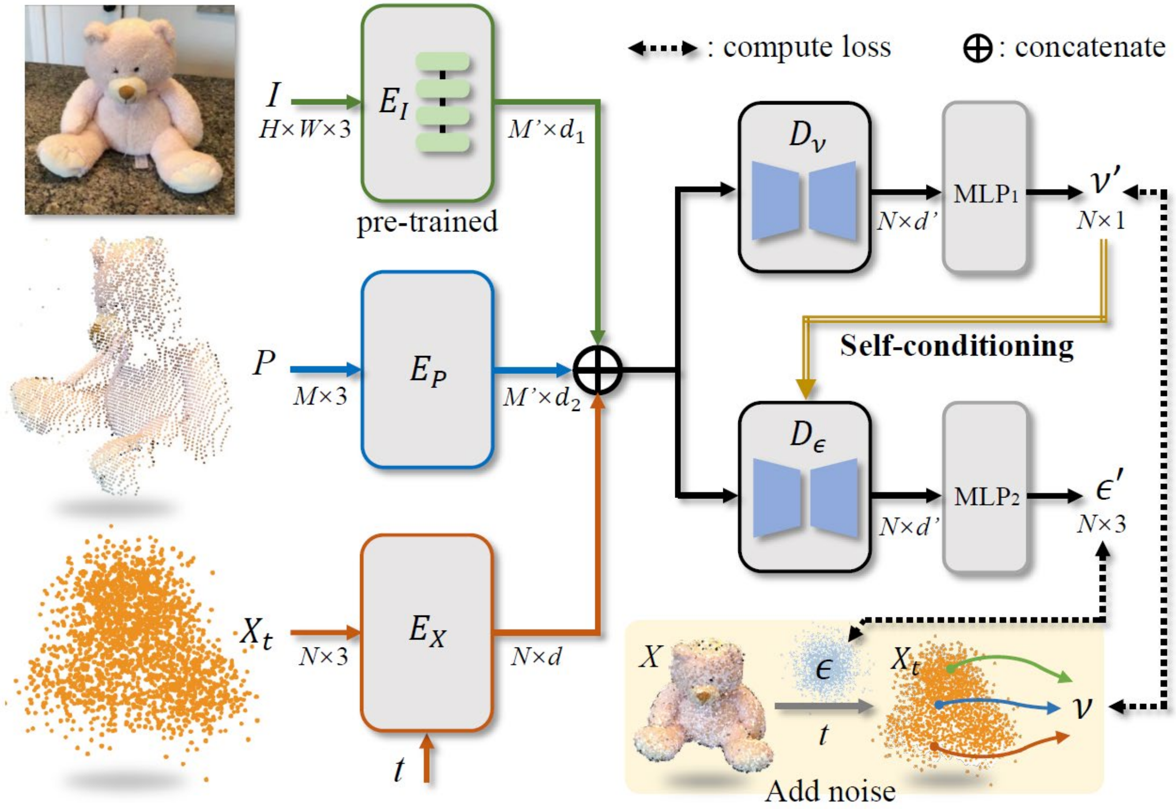
Supervision: GT pc, implicit value, and noise



香 港 中 文 大 學 (深 圳)
The Chinese University of Hong Kong, Shenzhen

Methodology

- Pipeline



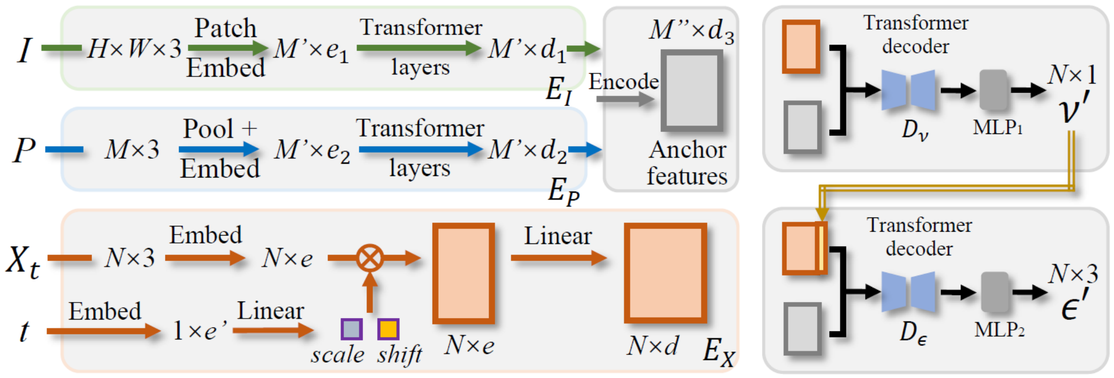
香港中文大學(深圳)
The Chinese University of Hong Kong, Shenzhen

CVPR
SEATTLE, WA JUNE 17-21, 2024

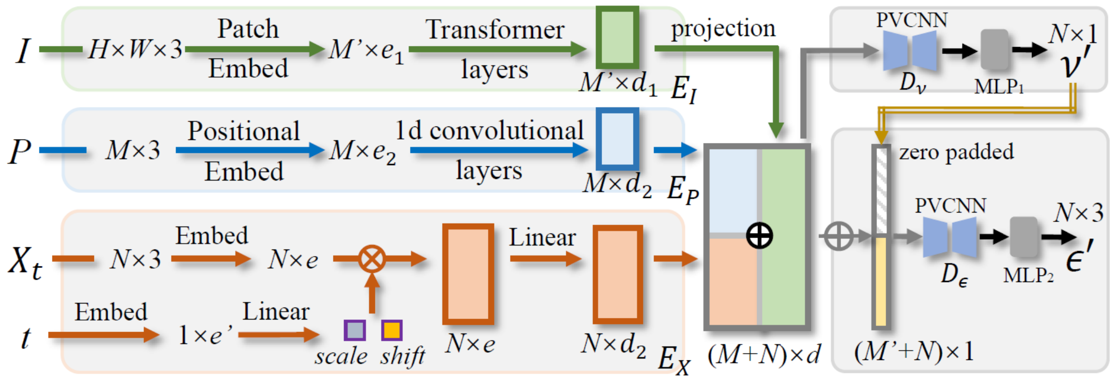
Methodology

- Implementation

Transformer-based implementation:



PVCNN-based implementation:



香港中文大學(深圳)
The Chinese University of Hong Kong, Shenzhen

Experiments

- Denoising process visualization

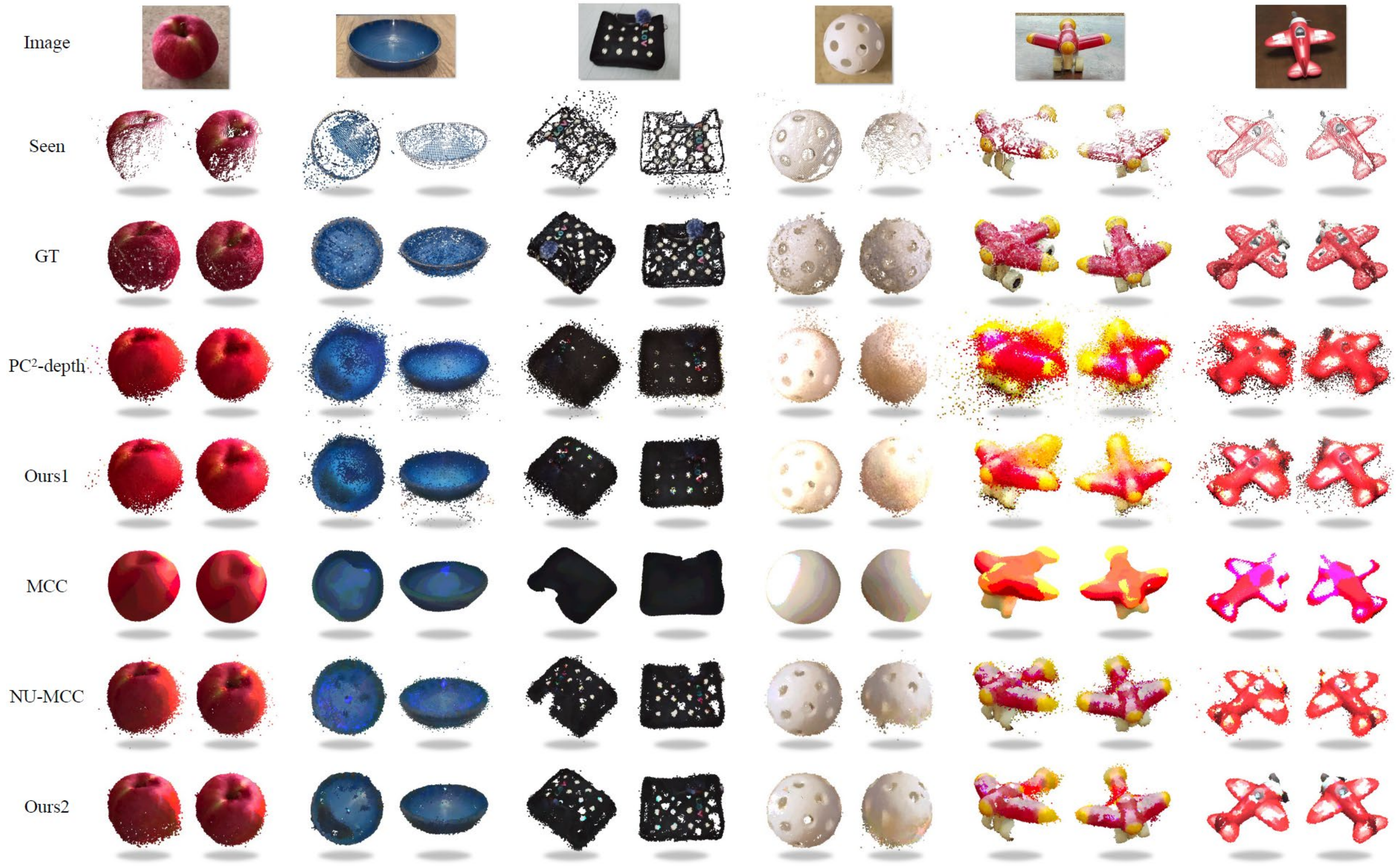
Experiments

- Quantitative results on CO3D-v2 (10 held-out categories)

Method	Backbone	Acc↓	Comp↓	CD↓	Prec↑	Recall↑	F1↑
PC^2	PVCNN	0.342	0.214	0.556	24.2	56.2	33.0
PC^2-depth	PVCNN	0.209	0.103	0.312	61.7	87.6	70.7
MCC	Transformer	0.172	0.144	0.316	68.9	72.7	69.8
NU-MCC	Transformer	0.121	0.146	0.266	79.2	84.0	80.9
Ours1	PVCNN	0.163	0.089	0.252	69.0	89.7	76.2
Ours2	Transformer	**0.104**	**0.087**	**0.190**	**85.1**	**90.1**	**87.2**

Experiments

- Qualitative results on CO3D-v2 (held-out categories)



香港中文大學(深圳)
The Chinese University of Hong Kong, Shenzhen

Experiments

- Generalization results on MVImgNet

Experiments

- Qualitative results on CO3D-v2 (held-in categories)

香港中文大學(深圳)
The Chinese University of Hong Kong, Shenzhen

End

- Thanks!

IPoD

Project Page

Our Lab

Yushuang Wu

My Homepage