3D SHAPE COMPLETION VIA SPARSE IRREGULAR REPRESENTATION

Jiahui Li & Pourya Shamsolmoali
Shanghai Key Laboratory of Multidimensional Information Processing
East China Normal University
Shanghai, China
{jiahuili0331,pshams55}@gmail.com

ABSTRACT

The task of 3D shape completion involves completing missing regions of an object from partial observation. The current methods accomplish this task by modeling latent completion distributions based on an autoregressive model. However, this approach often struggles with geometric details, as it represents 3D shapes with variable latent sequences, leading to gaps (local missing) in the completed shape. In this paper, we introduce a multiple 3D shape completion method using a transformer-based autoregressive model and a fixed-length sparse irregular latent sequence. Experiments demonstrate that our method outperforms state-of-the-art methods in terms of both quality and fidelity.

1 INTRODUCTION

3D shape completion is becoming more important in computer vision due to challenges in acquiring complete object scans stemming from factors like the varying angles of 3D scanning devices and issues with object occlusion. cGAN Wu et al. (2020) addresses this challenge by using the GAN model Goodfellow et al. (2014). However, this model encodes the 3D shape using a global latent vector, failing to capture the fine-grained details of a 3D object, which leads to blurry completion. ConvONet Peng et al. (2020) and IF-Net Chibane et al. (2020) present alternative methods for representing 3D shapes using voxelized latent grids. They interpolate a 3D shape into grid-based local latent vectors, effectively preserving the shape's geometric details. However, they fail to make reasonable generalizations for the unseen parts due to the deterministic nature of gird. On the other hand, Pointr Yu et al. (2021) processes the 3D shape as a sequence of tokens and adopted transformers Vaswani et al. (2017) to predict missing parts. However, the missing tokens predicted by the linear projection layer overlook the contextual relationship with other tokens. To address this issue, ShapeFormer Yan et al. (2022) utilizes an autoregressive model Van Den Oord et al. (2016) to construct shape completion in a recurrent manner. However, since ShapeFormer represents 3D shapes as variable-length latent sequences, it requires exploring the entire dataset to determine the maximum sequence length. This requirement can pose challenges in generative modeling and may result in local geometric gaps in the completed shapes.

To address the above limitation, we replace the variable representation in ShapeFormer with a sparse irregular representation Zhang et al. (2022), which encodes the 3D shape into a fixed-length discrete latent sequence. With this design, our model is more sensitive to local geometry variations, which helps to accurately capture incomplete shape structures when confronted with different ambiguities. More specifically, our transformer-based autoregressive method Radford et al. (2019) can produce a more accurate and plausible completion conditioned on partial input. Experiments demonstrate that our method achieves state-of-the-art results in completion quality, diversity, and fidelity.

2 METHOD

Shape Encoding: Similar to Zhang et al. (2022), we learn a fixed-length sparse irregular discrete latent sequence $\mathcal{S} = \{x_i, y_i, z_i, v_i\}_{i=0}^{M-1}$, which represent M local latent vectors with four different

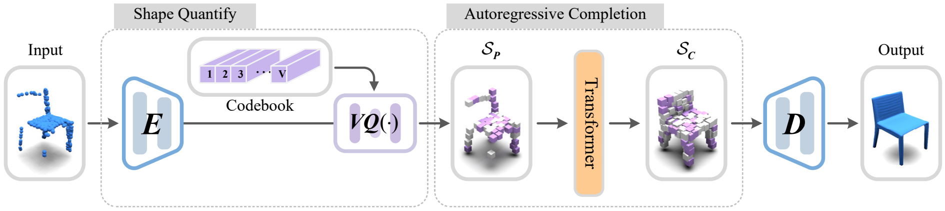

Figure 1: Overview of our shape completion method. Given a partial input, our method first encode it as incomplete latent codes, then the transformer autoregressively constructs latent completion.

attributes (v, x, y, z), where x_i, y_i, and z_i are the 3D coordinates and v_i is the discrete form of each local latent vector (refer to VQVAE Van Den Oord et al. (2017) for more details). Due to the spatial nature of the latent sequence S, arbitrary inputs are able to keep the topology structure, which can sensitively capture incomplete shape structure, leading to more faithful completion. With the shape encoding, we represent the complete shape and the incomplete shape as $\mathcal{S}_C$ and $\mathcal{S}_P$, respectively.

Autoregressive Sequence Completion: As we discussed above, the problem of 3D shape completion can be defined as the conditional probability distribution between partial and complete sequences, $p(\mathcal{S}_C|\mathcal{S}_P)$. Our objective is to predict the probability distribution of the next element based on the previous elements. Therefore, the likelihood can be written as:

$$p(\mathcal{S}_C|\mathcal{S}_P) = \prod_{i=0}^{M-1} p(\mathcal{S}_{C_i}|\mathcal{S}_{C_{<i}}, \mathcal{S}_P) = \prod_{i=0}^{M-1} p_{x_i} \cdot p_{y_i} \cdot p_{z_i} \cdot p_{v_i} \tag{1}$$

in which $p_{x_i} = p(x_i|\mathcal{S}_{C_{<i}}, \mathcal{S}_P)$, $p_{y_i} = p(y_i|x_i, \mathcal{S}_{C_{<i}}, \mathcal{S}_P)$, $p_{z_i} = p(z_i|y_i, x_i, \mathcal{S}_{C_{<i}}, \mathcal{S}_P)$, $p_{v_i} = p(v_i|z_i, y_i, x_i, \mathcal{S}_{C_{<i}}, \mathcal{S}_P)$. The completion process is illustrated in Figure 1.

3 Experiments

In this section, we demonstrate that our method outperforms state-of-the-art models for shape completion in different scan ambiguities. Following Yan et al. (2022) 13 classes of the ShapeNet Chang et al. (2015) dataset are used, and we process the data like OccNet Mescheder et al. (2019). The qualitative and quantitative results show that our method achieves more accurate completion in the presence of various scan ambiguities. For example, our method shows a 4.3% improvement in FPD over ShapeFormer (SFr.) at the high ambiguity level. This success is largely due to our irregular encoding schedule, which is more sensitive to different levels of incompleteness and provides sufficient condition perception for better completion.

Table 1: Quantitative results on ShapeNet with [Low/High] scan ambiguity. CD is scaled by 10^3.

	OccNet	CONet	IF-Net	Pointr	cGAN	SFr.	Ours
CD↓	1.48 / **2.79**	0.81 / 3.14	0.79 / 18.4	0.80 / 3.11	1.33 / 3.49	0.74 / 4.72	**0.72** / 3.30
F-score↑	63.2 / 50.4	72.9 / 60.4	**73.8** / 51.5	70.1 / 59.3	62.1 / 59.3	70.3 / 60.5	**73.8 / 61.6**
FPD↓	0.34 / 3.12	0.23 / 2.85	0.25 / 3.66	0.23 / 3.29	1.36 / 2.55	0.24 / 1.45	**0.22 / 1.39**

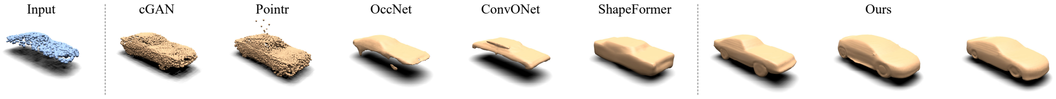

Figure 2: Visual comparison for shape completion in the car category of ShapeNet.

4 Conclusion

We presented a new 3D shape completion method based on a sparse irregular representation. By encoding 3D shapes into fixed-length discrete sequences, our method uses a transformer to autoregressively generate multiple plausible shape completions from incomplete input. The quantitative and qualitative comparisons demonstrate that our method outperforms other approaches in terms of completion quality and fidelity.

URM Statement

The authors acknowledge that at least the key author of this work meets the URM criteria of the ICLR 2024 Tiny Papers Track.

References

Angel X Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, et al. Shapenet: An information-rich 3d model repository. *arXiv preprint arXiv:1512.03012*, 2015.

Julian Chibane, Thiemo Alldieck, and Gerard Pons-Moll. Implicit functions in feature space for 3d shape reconstruction and completion. In *Proc. CVPR*, 2020.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Proc. NIPS*, 27, 2014.

Lars Mescheder, Michael Oechsle, Michael Niemeyer, Sebastian Nowozin, and Andreas Geiger. Occupancy networks: Learning 3d reconstruction in function space. In *Proc. CVPR*, 2019.

Songyou Peng, Michael Niemeyer, Lars Mescheder, Marc Pollefeys, and Andreas Geiger. Convolutional occupancy networks. In *Proc. ECCV*, 2020.

Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proc. CVPR*, 2017.

Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9, 2019.

Aäron Van Den Oord, Nal Kalchbrenner, and Koray Kavukcuoglu. Pixel recurrent neural networks. In *Proc. ICML*, pp. 1747–1756. PMLR, 2016.

Aaron Van Den Oord, Oriol Vinyals, et al. Neural discrete representation learning. *Proc. NIPS*, 2017.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Proc. NIPS*, 2017.

Rundi Wu, Xuelin Chen, Yixin Zhuang, and Baoquan Chen. Multimodal shape completion via conditional generative adversarial networks. In *Proc. ECCV*, 2020.

Xingguang Yan, Liqiang Lin, Niloy J Mitra, Dani Lischinski, Daniel Cohen-Or, and Hui Huang. Shapeformer: Transformer-based shape completion via sparse representation. In *Proc. CVPR*, 2022.

Xumin Yu, Yongming Rao, Ziyi Wang, Zuyan Liu, Jiwen Lu, and Jie Zhou. Pointr: Diverse point cloud completion with geometry-aware transformers. In *Proc. ICCV*, 2021.

Biao Zhang, Matthias Nießner, and Peter Wonka. 3dilg: Irregular latent grids for 3d generative modeling. *Proc. NIPS*, 35:21871–21885, 2022.

A Appendix

A.1 Experiment Setting

Metric: Same to ShapeFormer Yan et al. (2022), we utilize Chamfer L2 Distance (CD), F-score, and Fréchet Point Cloud Distance (FPD) as evaluation metrics. **Quantitative Evaluation:** Following previous method Yan et al. (2022), we only generate one completion result for quantitative comparison. **Qualitative Evaluation:** For qualitative comparison, we generate multiple completion results to showcase the diversity of our model.

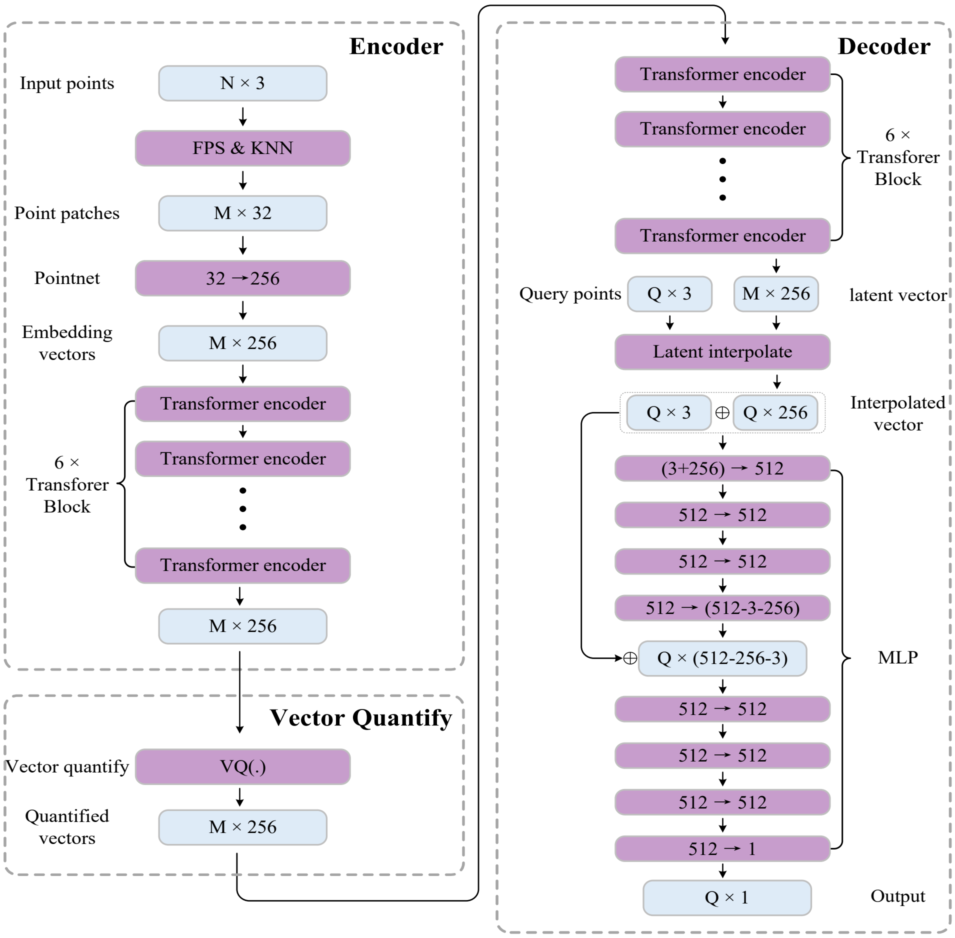

Figure 3: Network architecture of our shape encoding schedule. We show the detailed module inside our autoencoder.

A.2 Training Details

The training process of our encoding schedule is set to $T = 800$ epochs with 32 batch size. The learning rate is linearly increased to $lr_{max} = 1e - 3$ in the first $t_0 = 40$ epochs. Then gradually decreased using the cosine decay schedule $lr_{max} * 0.5^{1+\cos\frac{t-t_0}{T-t_0}}$, until the minimum value of $lr_{min} = 1e - 6$. Also, Adamw optimizer is used with its default parameters. The training process of the autoregressive model is the same as for the encoding part, but $T = 400$ with batch size 16. We set top-$p = 0.8$ for sampling.

A.3 Dataset Details

We train our model using the objects from 13 categories of the ShapeNet dataset, including [airplane, bench, cabinet, car, chair, display, lamp, speaker, rifle, sofa, table, telephone, vessel]. We extract the occupancy value as Mescheder et al. (2019), including 50k volume query points from the bounding volume ($[-1, 1]^3$) and 50k near query points from the near surface region.

A.4 Network Architecture of Shape Autoencoder

We represent each 3D shape as a fixed-length sparse irregular discrete latent sequence as 3DILG Zhang et al. (2022). The network architecture is shown in Figure 3. Given an input point cloud with size $N \times 3$, we process it into M point patches by FPS and KNN. Next, we use pointnet Qi et al. (2017) to project each patch to an embedding vector and use a six layers transformer encoder

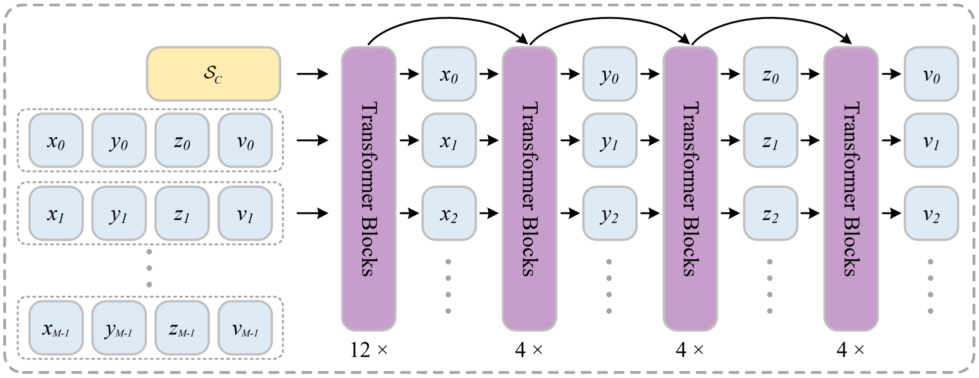

Figure 4: Network architecture of our autoregressive model. Here, we show the generation process of our autoregressive sampling.

Table 2: Quantitative comparison of computational cost with autoregressive style model ShapeFormer.

Method	Training Time	Inference Time / shape
ShapeFormer	144 hours	0.06
Ours	131 hours	0.075

Dosovitskiy et al. (2020) to build embedding vectors into latent vectors. Following the vector quantization Van Den Oord et al. (2017), we replace the latent vectors with the closest quantified vectors in the codebook. By preserving the index of each quantified vector and the discrete coordinates of each latent vector, we can represent the 3D shape as a discrete latent sequence $\mathcal{S} = \{x_i, y_i, z_i, v_i\}_{i=0}^{M-1}$. After being processed by another six layers transformer encoder, we interpolate quantified vectors to each query point by:

$$\mathrm{z_q} = \sum_{i=0}^{M-1} \frac{\exp(-\beta \|\mathrm{q} - \mathrm{c}_i\|^2)}{\sum_{j=0}^{M-1} \exp(-\beta \|\mathrm{q} - \mathrm{c}_j\|^2)} \mathrm{z}_i. \tag{2}$$

In which, q is the query point used for surface reconstruction, while c_i is the position of each quantified vector z_i (which has the same position as the latent vectors). And β is a learnable parameter controlling the smoothness of interpolation. Consequently, we get interpolated vectors with the same size as query points q. With a MLP we can predict the occupancy value Mescheder et al. (2019) of each query point.

A.5 Autoregressive Sampling

We show the sampling process of the autoregressive model in Figure 4. The model Radford et al. (2019) consists of twenty four layers of transformer modules. Given a partial sequence, the model autoregressively generate new element based on previous elements. With M times iteration, the model autoregressively construct sequence completion as shown in Figure 4. Note that, the training process only runs once, since the input sequences and the ground truth sequences are trained in a displacement manner.

A.6 Efficiency Analysis and Qualitative Results

We show the computational experiment with autoregressive style model ShapeFormer in Table 2. Similarly as long-range iterative generation model, our model has similar training time as ShapeFormer. However, our model has more faster inference speed than ShapeFormer which needs to process variable-length latent decoding to 3D shapes. Furthermore, our model can achieve better completion quality as shown in Table 1.

Moreover, in Figure 5 we have shown more completion results of our method. For each partial input, we generate five completion results to showcase the diversity of our model. From the completion

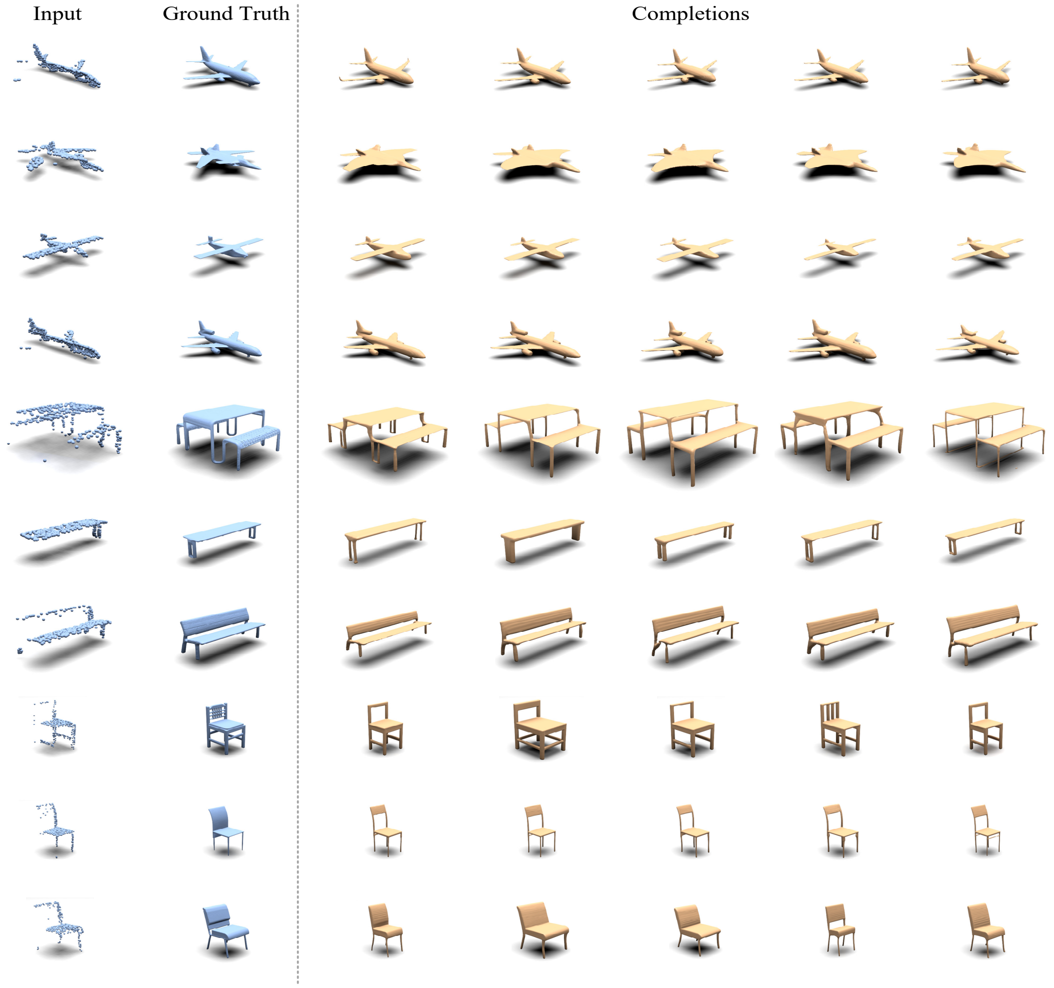

Figure 5: Completion results on ShapeNet dataset.

results, it can be proved that our model is capable of generating multiple high-quality completions while remaining aligned with the input.