BRNO UNIVERSITY OF TECHNOLOGY FACULTY OF INFORMATION TECHNOLOGY

Excel@FIT 2024

http://excel.fit.vutbr.cz

Generative Models for 3D Shape Completion

Peter Zdravecký*

Abstract

In many real-world scenarios, scanned 3D models contain missing parts due to occlusion, scanning errors, or the incomplete nature of the data itself. To mitigate those errors, methods are employed for the shape completion of incomplete 3D models. The majority of existing methods that perform 3D model shape completion are not robust enough and cannot handle large missing areas. The goal of this paper is to create an automated process for the shape completion task using a supervised method based on deep learning. The proposed solution is to use a diffusion-based model and handle the task as a generative problem to create a complete shape from an incomplete one. The results showed a high capability of this model in the shape completion task with an 81.6 IoU metric score on the test dataset. The model also demonstrates strong generalization capabilities on shapes that are not part of the training distribution (average 70.9 IoU metric score). The strength of the proposed approach is in its processing in a low-resolution domain, which enhances the inference speed and reduces the computational demands, given that diffusion models are challenging in this respect.

*xzdrav00@vut.cz, *Faculty of Information Technology, Brno University of Technology*

1. Introduction

With the increasing accessibility of tools that generate 3D data from physical shapes, there is a need for solutions to address the potential drawbacks these tools may present. The issue lies in the fact that the scanned 3D model may have missing sections that need repair [1]. The problem of *filling the holes* is called shape completion. The existing method performs well in filling small areas or flat surfaces without extra details [2]. The ideal method for the shape completion method should be able to handle larger missing areas and complex geometry, which is usually the case in real-world shapes.

To address the shape completion problem, deep learning methods are employed. Chu *et al.* [3] introduced DiffComplete, a diffusion-based state-of-the-art approach to 3D shape completion on range scans represented by implicit shape representation. This approach yields encouraging results, even in instances that fall outside of the training distribution. However, the downside is that it requires substantial computational resources to infer the complete shape, and the quality of the final shape it provides, due to small grid resolution (TSDF representation), may not be adequate for potential production applications. ShapeFormer, introduced by Yan et al. [4], used for shape completion, takes advantage of a compact 3D representation known as a Vector Quantized Deep Implicit Function (VQDIF). This compact representation drastically reduces the sequence length from cubic to quadratic in terms of feature resolution.

The proposed solution to handle shape completion is based on the DiffComplete approach. The aim is to improve the shortcomings and explore the potential capabilities of the Diffcomplete approach. The primary improvement involves a new method of reducing the input resolution, processing it, and then subsequently rescaling it back to its original size. The additional enhancement incorporates user input in the form of a Region of Interest to more effectively guide the completion process and rectify failure cases.

2. Proposed method

Diffusion process

The approach uses the probabilistic diffusion model, which includes a *forward* and *backward* process (see Equation 1 and Equation 2). In the *forward pro-*

cess $q(x_{0:T})$, Gaussian noise is gradually added to obscure the ground-truth shape x_0, into a random noise volume x_T where T is the total number of time stamps. The *backward process* $p_\theta(x_{0:T}, c)$ utilizes a shape completion network, with learned parameters θ, to iteratively remove noise from the noise volume x_T. The iterative noise removal process is visualized in Figure 1. The network architecture used (see Figure 2) employs a dual branch strategy. One branch handles complete shapes, and another branch handles incomplete shapes. The primary branch takes as input a corrupted complete shape $TSDF_C^T$. The secondary branch processes incomplete shapes $TSDF_P$ and mirrors the structure of the primary branch. This branch focuses on efficient feature extraction and utilizes a projection layer after each encoder/middle block to forward multiscale features to the primary branch's decoder blocks.

Training and Inference

The training objective for this approach is as follows:

$$\arg\min_\theta E_{t,x_0,\epsilon,c}\left[\|\epsilon - \epsilon_\theta(x_t, t, c)\|^2\right], \quad \epsilon \in \mathcal{N}(0, \mathbf{I}),$$

To maximize the probability of generation $p_\theta(x_0)$ (to obtain the original shape), the mean square error loss function is used.

During the inference phase, a randomized 3D noise volume of the standard Gaussian distribution is used as input x_t. The trained completion network is then used for T iterations to produce x_0 from x_t, conditioned on the partial shape. To accelerate the inference process, the technique of subsampling a set of timestamps $[1, \ldots, T/10]$ is used [5]. The visualization of the inference process is shown in Figure 4.

Evaluation Metrics

To ensure a comprehensive and unbiased quantitative evaluation, three metrics are used: *Intersection over Union (IoU)*, *Chamfer Distance (CD)*, and *Mean Absolute Error (L1 Loss)*. IoU measures the overlap between two binary volumes, presented in a voxel grid: the predicted volume $\mathcal{V}_p$ and the ground truth volume $\mathcal{V}_{gt}$. The chamfer distance is a measure of similarity between two point clouds, defined as the average distance between each point in one cloud and its nearest neighbor in the other cloud. mean absolute error, also known as L1 loss, quantifies the average magnitude of errors in a set of predictions without considering their direction.

3. Experimental Results

The experiments first focused on the reproducibility of the proposed method to match the results of the original paper. The absence of code for the Diffcomplete approach and some ambiguities in the paper necessitated adaptations, resulting in a structure that might diverge from the original Diffcomplete. Quantitative results for various datasets, including those with out-of-distribution shapes, are shown in Table 1. The subsequent investigation delved into the network's perception of the condition by attempting to complete the same model using diverse inputs with differing missing parts, revealing a heavy reliance on repetitive parts. The following experiment tested the ability to complete shapes outside the training distribution, yielding promising results. To address the failure cases identified in the base approach, an additional input was introduced in the form of a Region of Interest (RoI). The model incorporating the RoI surpassed the performance of the base solution. The primary challenge was the computational requirements, which allowed for experiments with low-resolution grid sizes for the input representation. The experiment involving the proposed low-resolution processing yielded positive results, but at the cost of poorer generalization. The visual results of the base solution are presented on the left side of Figure 3, while the right side displays results at higher resolution, noting that the inputs retain the same low-resolution size.

4. Conclusions and Future Work

The improvements made to the existing DiffComplete method showed a great improvement over the original method. The proposed method is capable of producing high-quality shapes for various scenarios.

Future Work

Although the proposed solution shows strong results in shape completion, there is substantial room for improvement. Exploring efficient 3D network modules, such as SparseConv [6] or Octree-based [7] layers, could offer a viable way to handle high-resolution 3D shapes without incurring prohibitive computational costs. Further research could focus on the diffusion process, by experiment with number of iterations steps, using different sampling schedulers or noise schedulers could also be introduced to assess their impact on training efficiency and the quality of results. There is also an intention to modify the method to suit the medical data related to cranial implants. It should be noted that no literature has yet tried a similar approach.

Acknowledgements

I would like to thank my supervisor, Ing. Tibor Kubík for his help. Computational resources were provided by the e-INFRA CZ project (ID:90254), supported by the Ministry of Education, Youth and Sports of the Czech Republic.

References

[1] Xiaoyuan Guo, Jun Xiao, and Ying Wang. A survey on algorithms of hole filling in 3d surface reconstruction. *The Visual Computer*, 34:93–103, 2018.

[2] Tao Ju. Robust repair of polygonal models. *ACM Transactions on Graphics (TOG)*, 23(3):888–895, 2004.

[3] Ruihang Chu, Enze Xie, Shentong Mo, Zhenguo Li, Matthias Nießner, Chi-Wing Fu, and Jiaya Jia. Diffcomplete: Diffusion-based generative 3d shape completion. *arXiv preprint arXiv:2306.16329*, 2023.

[4] Xingguang Yan, Liqiang Lin, Niloy J Mitra, Dani Lischinski, Daniel Cohen-Or, and Hui Huang. Shapeformer: Transformer-based shape completion via sparse representation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6239–6249, 2022.

[5] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models, 2022.

[6] Benjamin Graham, Martin Engelcke, and Laurens Van Der Maaten. 3d semantic segmentation with submanifold sparse convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 9224–9232, 2018.

[7] Peng-Shuai Wang, Yang Liu, Yu-Xiao Guo, Chun-Yu Sun, and Xin Tong. O-cnn: Octree-based convolutional neural networks for 3d shape analysis. *ACM Transactions On Graphics (TOG)*, 36(4):1–11, 2017.