CLIPXPlore: Coupled CLIP and Shape Spaces for 3D Shape Exploration

Jingyu Hu*
jyhu@cse.cuhk.edu.hk
The Chinese University of Hong Kong
HK SAR, China

Ka-Hei Hui*
khhui@cse.cuhk.edu.hk
The Chinese University of Hong Kong
HK SAR, China

Zhengzhe Liu
zzliu@cse.cuhk.edu.hk
The Chinese University of Hong Kong
HK SAR, China

Hao Zhang
haoz@sfu.ca
Simon Fraser University
Canada

Chi-Wing Fu
cwfu@cse.cuhk.edu.hk
The Chinese University of Hong Kong
HK SAR, China

ABSTRACT

This paper presents CLIPXPlore, a new framework that leverages a vision-language model to guide the exploration of the 3D shape space. Many recent methods have been developed to encode 3D shapes into a learned latent shape space to enable generative design and modeling. Yet, existing methods lack effective exploration mechanisms, despite the rich information. To this end, we propose to leverage CLIP, a powerful pre-trained vision-language model, to aid the shape-space exploration. Our idea is threefold. First, we couple the CLIP and shape spaces by generating paired CLIP and shape codes through sketch images and training a mapper network to connect the two spaces. Second, to explore the space around a given shape, we formulate a co-optimization strategy to search for the CLIP code that better matches the geometry of the shape. Third, we design three exploration modes, binary-attribute-guided, text-guided, and sketch-guided, to locate suitable exploration trajectories in shape space and induce meaningful changes to the shape. We perform a series of experiments to quantitatively and visually compare CLIPXPlore with different baselines in each of the three exploration modes, showing that CLIPXPlore can produce many meaningful exploration results that cannot be achieved by the existing solutions.

CCS CONCEPTS

• **Computing methodologies** → **Shape analysis**; **Neural networks**; **Mesh models**.

KEYWORDS

3D shape generation, shape space exploration

*indicates joint first authors

SA Conference Papers '23, December 12–15, 2023, Sydney, NSW, Australia

ACM ISBN 979-8-4007-0315-7/23/12…$15.00
https://doi.org/10.1145/3610548.3618144

ACM Reference Format:
Jingyu Hu, Ka-Hei Hui, Zhengzhe Liu, Hao Zhang, and Chi-Wing Fu. 2023. CLIPXPlore: Coupled CLIP and Shape Spaces for 3D Shape Exploration. In *SIGGRAPH Asia 2023 Conference Papers (SA Conference Papers '23), December 12–15, 2023, Sydney, NSW, Australia.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3610548.3618144

1 INTRODUCTION

Generative design and modeling of 3D shapes has been a long-standing problem in computer graphics. Since the design goals are often open-ended, even vague, the modeling process is naturally *iterative* and *exploratory*. Early works have already studied exploratory modeling as a means to empower novice users to create high-quality 3D shapes [Tal 2009; Chaudhuri and Koltun 2010]. Even earlier, both language [Coyne and Sproat 2001] and sketching [Igarashi et al. 1999] interfaces have been developed to best connect the user and the 3D modeling endeavor. With the proliferation of large-scale 3D data, the modern approach is to learn a high-dimensional 3D shape latent space, where each latent code can be decoded into a 3D shape, and perform shape generation to explore the latent space.

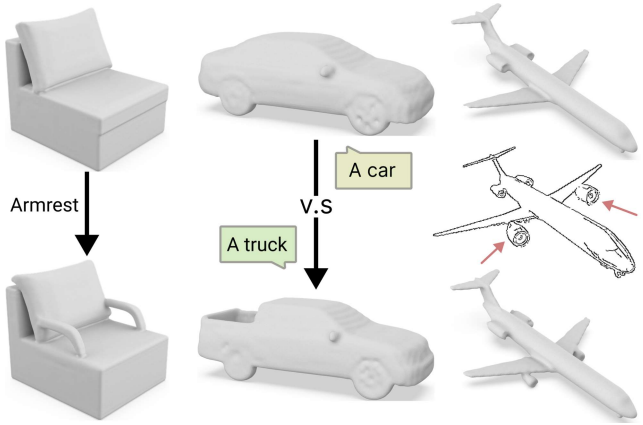

Figure 1: CLIPXPlore demonstrates new capabilities of leveraging a vision-language model to explore the 3D shape space: binary-attribute-guided (left), text-guided (middle), and sketch-guided (right). Top row shows input shapes, while bottom row shows associated shapes from exploration.