International Journal of Engineering and Management Research
e-ISSN: 2250-0758 | p-ISSN: 2394-6962
https://ijemr.vandanapublications.com

Peer Reviewed & Refereed Journal
Volume-14, Issue-3 (June 2024)
https://doi.org/10.5281/zenodo.12662936

Editable Neural Radiance Fields Convert 2D to 3D Furniture Texture

Chaoyi Tan[1], Chenghao Wang[2], Zheng Lin[3], Shuyao He[4] and Chao Li[5]
[1]Department of Electrical and Computer Engineering, Northeastern University, Boston, MA, USA
[2]Georgia Institute of Technology, USA
[3]University of California Santa Cruz, USA
[4]Information Systems, Northeastern University, Boston, USA
[5]Georgetown University, USA

[4]Corresponding Author: he.shuyao@northeastern.edu

Received: 18-05-2024 Revised: 07-06-2024 Accepted: 29-06-2024

ABSTRACT

Our work presents a neural network designed to convert textual descriptions into 3D models. By leveraging the encoder-decoder architecture, we effectively combine text information with attributes such as shape, color, and position. This combined information is then input into a generator to predict new furniture objects, which are enriched with detailed information like color and shape.[1] The predicted furniture objects are subsequently processed by an encoder to extract feature information, which is then utilized in the loss function to propagate errors and update model weights. After training the network, we can generate new 3D objects solely based on textual input, showcasing the potential of our approach in generating customizable 3D models from descriptive text.[2]

Keywords-- Neural Radiance, 2D, 3D, Texture

I. INTRODUCTION

This semester we need to realize the construction from textual information to 3D scene. At present, the number of text to 3d work is very limited, and the application of text to 2d is quite mature, especially the work of DALE of open-AI[3], which basically reaches the benchmark level of text to 2D images. Our work on text-to-image generation starts with the idea of text-to-image generation 2d. The purpose is to understand the multimodal structure of text-to-image models, including decoders and encoders. We made a model of a text-generating 3d object based on the work of others, this model can be customized in shape and material[4]

II. RELATED WORK

Related work focuses on text to 2d images, and open ai proposed DALL-E [5] and CLIP [6] derived a lot of work. The data for CLIP training comes from 400 million data pairs on the Internet. With this data, CLIP needs to complete the task of: [7]given an image, among 32,768 randomly sampled text snippets, find the one that matches. To accomplish this task, the CLIP model needs to learn to recognize various visual concepts in images and associate concepts with pictures. Therefore, the CLIP model can be applied to almost any visual classification task. For example, if a dataset is tasked with classifying photos of dogs and cats, and the CLIP model predicts which text description "a photo of a dog" and "a photo of a cat" is a better match. DALL·E is based on the CLIP/unCLIP mechanism, first in order to obtain a complete image generation model, the CLIP image embedding decoder is combined with a prior model, which generates possible CLIP image embeddings from a given text title. The full text conditional image generation stack is called unCLIP because it generates images by inverting the CLIP image encoder.[8]

In 3d area, there is very limited work that can be referenced. Faria Huq [9] investigates a new pipeline aimed at generating static and animated 3D scenes from different types of free-text scene descriptions without any major limitations. In particular, to make our study practical and tractable, we focus on a small subspace of all possible 3D scenes containing various combinations of cubes, cylinders, and spheres. We designed a two-stage pipeline. In the first stage, we encode free-form text using an encoder-decoder neural architecture. In the second stage, we generate a 3D scene from the generated encoding. [10]Our neural architecture utilizes a state-of-the-art language model as an encoder, exploiting rich contextual encoding and a new multi-head decoder to simultaneously predict multiple features of objects in a scene.

Peer Reviewed & Refereed Journal
Volume-14, Issue-3 (June 2024)
https://doi.org/10.5281/zenodo.12662936

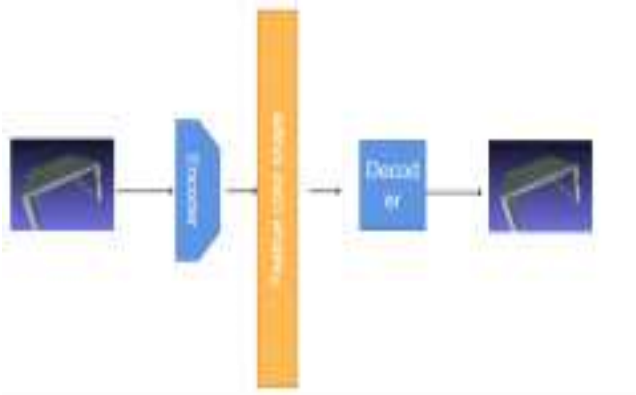

Figure 1: Object encoder

III. SOLUTION

The solution section covers all of your contributions (architecture, algorithms, formulas, findings). It explains in detail each contribution, if possible with figures/schematics.[11]

Based on the work of Faria Huq, the multimodality of text-generating 3D models is now able to generate very realistic geometry. However, most of the work still lacks the authenticity of 3D materials, mainly due to the lack of realism of materials. So we combine Implicit and 3d furniture generation, we can define the generated furniture color and material. This method can make the text generation 3D model more realistic.[12]

3.1 Text Encoder

We designed a set of architecture, the text information decoder is completed by bert [4] , Bert can decompose text information into word information, such as a /green /and /gray/ leg /vertical center/ table. We map the word level information to the feature combination of the next layer, and form a mapping relationship with the object information.[13]

3.2 Object Encoder

We extend the auto-encoder in [14] to jointly reconstruct the shape and color. As shown in Figure 2, our shape auto-encoder aims to map the input voxel-based shape $I \in R^{64\times64\times64}$ into a compact feature space.

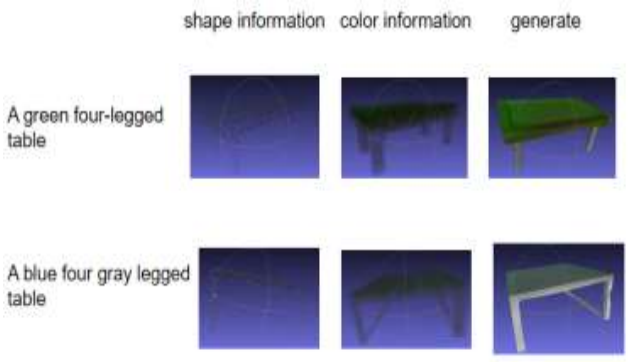

Figure 2: Result

Implicit [6] fields for learning generative models of shapes and introduce an implicit field decoder, called IM-NET, for shape generation, aimed at improving the visual quality of the generated shapes. An implicit field assigns a value to each point in 3D space, so that a shape can be extracted as an iso-surface.[15] IM-NET is trained to perform this assignment by means of a binary classifier. Specifically, it takes a point coordinate, along with a feature vector encoding a shape, and outputs a value which indicates whether the point is outside the shape or not. By replacing conventional decoders by our implicit decoder for representation learning (via IM-AE) and shape generation (via IM-GAN), we demonstrate superior results for tasks such as generative shape[16] modeling, interpolation, and single-view 3D reconstruction[17], particularly in terms of visual quality.

3.3 Transformer Decoder

We construct the spatial aware decoder D′ as the word-level spatial transformer(WLST) In short, we take the local features from WLST to improve the spatial correlation implied from feature vector.

IV. RESULTS AND DATA

Figure 3 shows the running results of our test set model. The model contains the color information and shape information of the object. 3d objects are generated when text is entered. And generate a point cloud object containing color and shape.[20]

4.1 Data Tables

We use Stanford's text2shape dataset [19], which has text information and two types of furniture information.

As an example, Table 4.1Contains text information, which describes the shape, color, and location of the item. There are two types of tags, table and chair.

Peer Reviewed & Refereed Journal
Volume-14, Issue-3 (June 2024)
https://doi.org/10.5281/zenodo.12662936

Dataset type	
table	chair
33134	42226

Table 1: Type and Destruction in dataset

Summary	Description
table	the table is round and has 3 legs. the table is rotating
chair	the chair is made of plastic and has 4 legs, it is black in colour
table	It is narrow console table. It is made from plywood. It is grey in colour.
table	A square shaped designery stylish table, with perfect supporting.
...	...

Table 2: Some sample in Dataset

V. CONCLUSION

This semester I tried everything from text to 3d images[21]. 3D cross-modal tasks can be said to be very few references. The conventional cross-modal approach is to break the gaps of different modalities through a decoder-generator structure. Our method is no exception using an implicit decoder-generator structure. This method completes some simple text-to-3d graphics demos. The next step is to continue to explore the generation of text-3d.

REFERENCES

[1] Li, Zhenglin, et al. (2023). Stock market analysis and prediction using LSTM: A case study on technology stocks. *Innovations in Applied Engineering and Technology*, 1-6.

[2] Hong, Bo, et al. (2024). The application of artificial intelligence technology in assembly techniques within the industrial sector. *Journal of Artificial Intelligence General Science (JAIGS)*, 1-12.

[3] Zhou, Chang, et al. (2024). *Optimizing search advertising strategies: Integrating reinforcement learning with generalized second-price auctions for enhanced ad ranking and bidding*. arXiv preprint arXiv:2405.13381.

[4] Li, Shaojie, Yuhong Mo & Zhenglin Li. (2022). Automated pneumonia detection in chest x-ray images using deep learning model. *Innovations in Applied Engineering and Technology*, 1-6.

[5] Zhou, Chang, et al. (2024). *Optimizing search advertising strategies: integrating reinforcement learning with generalized second-price auctions for enhanced ad ranking and bidding*. arXiv preprint arXiv:2405.13381.

[6] Mo, Yuhong, et al. (2024). Password complexity prediction based on roberta algorithm. *Applied Science and Engineering Journal for Advanced Research, 3*(3), 1-5.

[7] Jin, Jiajun, et al. (2024). Enhancing federated semi-supervised learning with out-of-distribution filtering amidst class mismatches. *Journal of Computer Technology and Applied Mathematics, 1*(1), 100-108.

[8] Dai, Shuying, et al. (2024). AI-based NLP section discusses the application and effect of bag-of-words models and TF-IDF in NLP tasks. *Journal of Artificial Intelligence General Science (JAIGS), 5*(1), 13-21.

[9] Mo, Yuhong, et al. (2024). Large language model (LLM) AI text generation detection based on transformer deep learning algorithm. *International Journal of Engineering and Management Research, 14*(2), 154-159.

[10] Song, Jintong, et al. (2024). A comprehensive evaluation and comparison of enhanced learning methods. *Academic Journal of Science and Technology 10*(3), 167-171.

[11] Dai, Shuying, et al. (2024). The cloud-based design of unmanned constant temperature food delivery trolley in the context of artificial intelligence. *Journal of Computer Technology and Applied Mathematics 1*(1), 6-12.

[12] Liu, Tianrui, et al. (2024). Spam detection and classification based on distilbert deep learning algorithm. *Applied Science and Engineering Journal for Advanced Research 3*(3), 6-10.

[13] Mo, Yuhong, et al. (2024). Make scale invariant feature transform “fly” with CUDA. *International Journal of Engineering and Management Research, 14*(3), 38-45.

[14] He, Shuyao, et al. (2024). Lidar and monocular sensor fusion depth estimation. *Applied Science and Engineering Journal for Advanced Research, 3*(3), 20-26.

Peer Reviewed & Refereed Journal
Volume-14, Issue-3 (June 2024)
https://doi.org/10.5281/zenodo.12662936

[15] Samir Elhedhli, Zichao Li, & James H. Bookbinder. (2017). Airfreight forwarding under system-wide and double discounts. EURO Journal on Transportation and Logistics, 6(2), 165–83. https://doi.org/10.1007/s13676-015-0093-5.

[16] Liu, Jihang, et al. (2024). Unraveling large language models: from evolution to ethical implications-introduction to large language models. *World Scientific Research Journal, 10*(5), 97-102.

[17] Lin, Zheng, et al. (2024). Text sentiment detection and classification based on integrated learning algorithm. *Applied Science and Engineering Journal for Advanced Research, 3*(3), 27-33.

[18] Zhao, Peng, et al. (2024). Task allocation planning based on hierarchical task network for national economic mobilization. *Journal of Artificial Intelligence General Science (JAIGS), 5*(1), 22-31.

[19] Zhu, Armando, et al. (2024). *Cross-task multi-branch vision transformer for facial expression and mask wearing classification*. arXiv preprint arXiv:2404.14606.

[20] Wang, Jin, et al. (2024). *Research on emotionally intelligent dialogue generation based on automatic dialogue system*. arXiv preprint arXiv:2404.11447.

[21] Zhang, Jingyu, et al. (2024). Research on detection of floating objects in river and lake based on AI image recognition. *Journal of Artificial Intelligence Practice, 7*(2), 97-106.

[22] Xiang, Ao, et al. (2024). Research on splicing image detection algorithms based on natural image statistical characteristics. *Journal of Image Processing Theory and Applications, 7*(1), 43-52.

[23] Cheng, Yu, et al. (2024). *Research on credit risk early warning model of commercial banks based on neural network algorithm*. arXiv preprint arXiv:2405.10762.

[24] Wang, Liyang, et al. (2024). *Application of natural language processing in financial risk detection*. arXiv preprint arXiv:2406.09765.

[25] Yang, Haowei et al. (2024). *Research on edge detection of LiDAR images based on artificial intelligence technology*.

[26] Zeyu Wang, Yue Zhu, Zichao Li, Zhuoyue Wang, Hao Qin & Xinqi Liu. (2024). Graph neural network recommendation system for football formation. *Applied Science and Biotechnology Journal for Advanced Research, 3*(3), 33-39. DOI: 10.5281/zenodo.12198843.

[27] Yang Wang, Chenghao Wang, Zichao Li, Zhuoyue Wang, Xinqi Liu & Yue Zhu. (2024). Neural radiance fields convert 2d to 3d texture. *Applied Science and Biotechnology Journal for Advanced Research, 3*(3), 40-44. DOI: 10.5281/zenodo.12200107.