Autoregressive Generation of Neural Field Weights

Luca Fanselau
TUM Munich
go68vog@mytum.de

Luis Muschal
TUM Munich
go98zoh@mytum.de

Abstract

The absence of a clear structure of implicit neural fields (NeF) makes it difficult to apply generative modeling directly to synthesize new data. To this end, we propose a novel approach for generating Multi Layer Perceptron (MLP)-weights of neural fields in an autoregressive transformer-based fashion. Our approach is expected to unconditionally generate the MLP-weights of novel neural fields.

1. Introduction

Recent successes in neural fields for compressed scene representation and autoregressive transformer models have been remarkable. Additionally, first approaches towards generating novel neural fields using diffusion models [4] and unconditional triangle-mesh generation using transformers [10] have been proposed. Challenges persist, however, due to the unstructured nature of implicit neural fields and the mismatch between the continuous MLP-weights and the discrete vocabulary typically used by transformers. To address this challenge, we investigate embedding techniques for neural fields that are well-suited for integration with transformer architectures. We propose a naive approach with positional encoding, learned continuous embedding, as well as a learned vocabulary and use them in conjunction with popular transformer architectures.

Our contributions are: (1) Develop embedding strategies for neural fields. (2) Train an autoregressive transformer-based model to generate novel neural fields

2. Related Works

HyperDiffusion [4] demonstrated the generation of novel MLP-weights for NeF, while MeshGPT [10] employed an autoregressive transformer-based approach for novel 3D structure generation. We aim to adapt the general transformer architecture from MeshGPT [10] to a different domain, namely from polygon meshes to MLP-weights. Additionally, relevant literature includes research on equivariant weight space representations [7, 8] which leveraged permutation symmetries in weight matrices for effective learning, utilizing neural network structure to uncover underlying patterns and embeddings [1, 5]. To handle the different domains extensions of transformers have been proposed for graph-based [3] and continuous data [2, 6].

3. Proposed Methods

Firstly, the training Neural Fields will be fitted on a dataset of images and 3D structures. The resulting weights will then have to be transformed to be usable as a transformer input such that they can be used to unconditionally and autoregressively generate novel weights. To deal with the unstructured nature of NeFs we investigate the following approaches:

Naive Approach: Directly perform a regression task on the MLP-weights of the NeF and use a continuous loss function. Additionally, we also want to investigate the possibility to use positional encoding to inform the model about the structure within the MLP.

Learned Embedding: Since the MLP is a fully connected graph, we want to propose an encoder-decoder architecture that captures the underlying structure in the latent space, by for example using Graph-CNNs. This representation is then used to train the transformer on a regression task.

Learned Vocabulary: Since transformers usually excel with predicting tokenized sequences we also want to investigate methods to quantize the input, either directly from the weights or the learned embedding. Instead of a regression task the transformer would predict the probability distribution of the most likely next token.

4. Experiments

Data: The experiments leverage the dataset introduced by [9], which contains neural radiance fields overfitted on the SIREN architecture using the MNIST, CIFAR10, MicroImageNet and ShapeNet datasets. This allows for evaluation on both 2D image data as well as 3D shape representations.

Metrics and Baselines: Evaluating the quality of synthesized neural fields poses challenges due to the lack of ground truth data. We adopt the metrics proposed by [4], specifically the Minimum Matching Distance (MMD), Coverage (COV), and 1-Nearest-Neighbor Accuracy (1-NNA), to facilitate comparisons with the seminal works that inspired this research and serve as performance baselines.

Top K	Temperature	MMD	COV %	1-NNA %	FPD
3	0.8	0.10	38.77	90.77	31.34
	1	0.11	34.57	91.31	36.78
5	0.8	0.10	37.50	90.43	35.21
	1	0.11	35.64	91.36	31.87

Top K	Temperature	MMD	COV %	1-NNA %	FPD
3	0.8	0.10	20.12	98.58	28.77
	1	0.10	19.04	98.78	26.80
5	0.8	0.10	21.88	98.29	28.03
	1	0.11	21.39	99.07	27.80

Method	COV % $\uparrow$	1-NNA % $\downarrow$	FPD $\downarrow$
Diffusion Voxel Baseline	28	94.1	38.9
PVD	39	76.3	5.8
DPC	46	74.7	18.7
HyperDiffusion	49	69.3	3.5
Ours [3, 0.8]	**38.77**	90.77	**31.34**
Ours [3, 1]	34.57	91.31	36.78
Ours [5, 0.8]	37.50	**90.43**	35.21
Ours [5, 1]	35.64	91.36	31.87

References

[1] Bruno Andreis, Soro Bedionita, and Sung Ju Hwang. Set-based neural network encoding, 2023. 1

[2] Jannis Born and Matteo Manica. Regression transformer enables concurrent sequence regression and generation for molecular language modelling. *Nature Machine Intelligence*, 5(4):432–444, 2023. 1

[3] Cameron Diao and Ricky Loynd. Relational attention: Generalizing transformers for graph-structured tasks, 2023. 1

[4] Ziya Erkoç, Fangchang Ma, Qi Shan, Matthias Nießner, and Angela Dai. Hyperdiffusion: Generating implicit neural fields with weight-space diffusion, 2023. 1

[5] Derek Lim, Haggai Maron, Marc T. Law, Jonathan Lorraine, and James Lucas. Graph metanetworks for processing diverse neural architectures, 2023. 1

[6] Weian Mao, Yongtao Ge, Chunhua Shen, Zhi Tian, Xinlong Wang, Zhibin Wang, and Anton van den Hengel. Poseur: Direct human pose regression with transformers. In *European conference on computer vision*, pages 72–88. Springer, 2022. 1

[7] Aviv Navon, Aviv Shamsian, Idan Achituve, Ethan Fetaya, Gal Chechik, and Haggai Maron. Equivariant architectures for learning in deep weight spaces, 2023. 1

[8] Aviv Navon, Aviv Shamsian, Ethan Fetaya, Gal Chechik, Nadav Dym, and Haggai Maron. Equivariant deep weight space alignment, 2024. 1

[9] Samuele Papa, Riccardo Valperga, David Knigge, Miltiadis Kofinas, Phillip Lippe, Jan-Jakob Sonke, and Efstratios Gavves. How to train neural field representations: A comprehensive study and benchmark, 2023. 1

[10] Yawar Siddiqui, Antonio Alliegro, Alexey Artemov, Tatiana Tommasi, Daniele Sirigatti, Vladislav Rosov, Angela Dai, and Matthias Nießner. Meshgpt: Generating triangle meshes with decoder-only transformers, 2023. 1