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 [1] and unconditional triangle-mesh generation using transformers [3] 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

Neural Implicit Fields and Diffusion Models Recent advancements have demonstrated the effectiveness of neural implicit fields in representing high-fidelity 3D geometries and radiance fields. For instance, DeepSDF encodes the shapes of objects as signed distance functions using a multilayer neural network, and NeRF uses MLPs to encode radiance fields for photorealistic rendering from novel views (2303.17015v1). Additionally, methods like Fourier features and periodic activation functions have been proposed to improve the representation of complex signals by addressing the bias towards learning low-frequency details in standard MLP (2303.17015v1).

Transformer-Based 3D Structure Generation Transformer architectures have shown promise in generating 3D structures. MeshGPT, for example, uses a decoder-only transformer to autoregressively generate triangle meshes, representing them as sequences of geometric embeddings. This approach has demonstrated improvements in mesh generation quality, emphasizing the capability of transformers to handle complex geometric data efficiently (2311.15475v1).

Diffusion Models in Generative Modeling Diffusion probabilistic models have emerged as powerful alternatives to GANs and energy-based models for generative tasks. Specifically, HyperDiffusion operates directly on the MLP weights of neural fields, enabling high-fidelity synthesis of 3D and 4D shapes. This method leverages a transformer-based architecture to model the diffusion process, achieving state-of-the-art performance in generating compact and coherent neural implicit fields (2303.17015v1)

Equivariant and Graph-Based Methods Research into leveraging permutation symmetries in weight matrices has also been significant. Equivariant weight space representations exploit these symmetries for more effective learning, while graph-based methods utilize the structure of neural networks to uncover underlying patterns and find embeddings (2303.17015v1). Extensions to transformers for graph-based and continuous data have further expanded their applicability in generative modeling (2303.17015v1).

Generative Adversarial Networks (GANs) GANs have been widely used for generating high-resolution images and 3D structures. However, they often suffer from training instability. This has led researchers to explore diffusion models as alternatives, which offer more stable training dynamics and improved generative performance (2303.17015v1).

Unifying Cross-Modality Generative Models Efforts to unify generative models across different data dimensions and modalities include methods like Functa and Diffusion Probabilistic Fields, which propose dimension-agnostic frameworks. These approaches aim to integrate the strengths of various generative models to handle diverse data types effective (2303.17015v1).

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 [2], 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 [1], 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.

References

[1] Ziya Erkoç, Fangchang Ma, Qi Shan, Matthias Nießner, and Angela Dai. Hyperdiffusion: Generating implicit neural fields with weight-space diffusion, 2023. 1, 2

[2] 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. 2

[3] 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