Few-shot Implicit Function Generation via Equivariance

Suizhi Huang[1,2], Xingyi Yang[2], Hongtao Lu[1*], Xinchao Wang[2*]

[1] Shanghai Jiao Tong University   [2] National University of Singapore

huangsuizhi@sjtu.edu.cn, xyang@u.nus.edu, htlu@sjtu.edu.cn, xinchao@nus.edu.sg

Abstract

*Implicit Neural Representations (INRs) have emerged as a powerful framework for representing continuous signals. However, generating diverse INR weights remains challenging due to limited training data. We introduce **Few-shot Implicit Function Generation**, a new problem setup that aims to generate diverse yet functionally consistent INR weights from only a few examples. This is challenging because even for the same signal, the optimal INRs can vary significantly depending on their initializations. To tackle this, we propose EQUIGEN, a framework that can generate new INRs from limited data. The core idea is that functionally similar networks can be transformed into one another through weight permutations, forming an equivariance group. By projecting these weights into an equivariant latent space, we enable diverse generation within these groups, even with few examples. EQUIGEN implements this through an equivariant encoder trained via contrastive learning and smooth augmentation, an equivariance-guided diffusion process, and controlled perturbations in the equivariant subspace. Experiments on 2D image and 3D shape INR datasets demonstrate that our approach effectively generates diverse INR weights while preserving their functional properties in few-shot scenarios. Code is available at https://github.com/JeanDiable/EquiGen.*

1. Introduction

Open-sourced models have been the driving force behind the incredible progress of Artificial Intelligence (AI) [27, 52, 64]. Beyond just making powerful models accessible, this openness has created new opportunities for meta-learning: treating model weights themselves as data sources. Recent works have demonstrated the potential of learning from [13, 24, 36, 46, 71] and generating neural network weights [15, 28, 51, 80] to promote further research.

One type of neural network well-suited for the weight generation paradigm is the Implicit Neural Representation (INR). These networks use a simple multi-layer perceptron
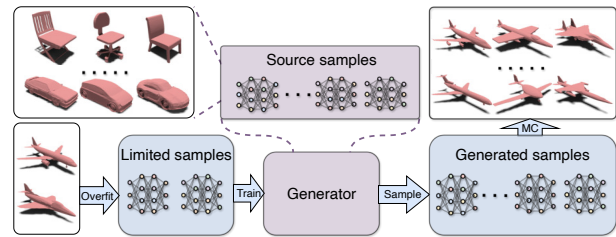
---

*Corresponding author.

Figure 1. Illustration of the Few-shot Implicit Function Generation setting with 3D INR data examples. The goal is to generate diverse INR weights from limited target samples. Source samples (top) show previously observed INRs of diverse shape categories. In practice, only limited target samples (bottom left) are available for training. The framework aims to learn a generator that can produce diverse generated samples (right) despite the limited training data. This setting addresses the practical scenario where only a few examples of new shapes are available for training.

(MLP) to fit continuous signals [8, 54, 61]. INRs serve as powerful tools for representing continuous data, offering several advantages: high-fidelity reconstruction [44, 49], smooth interpolation properties [8, 75], and infinite output resolution [3]. Their consistent and comparative simple architecture makes them ideal candidates for weight-based generative modeling.

However, training generative model on weight comes with its challenges, primarily due to a shortage of data. Currently, collecting large-scale INR data is costly, as it involves time-consuming gradient-based optimization. Existing datasets for INR data are limited from the diversity of sources [40, 46]. This data gap imposes restrictions on further investigation of weight space.

Given these constraints, an intuitive motivation is to target on a more practical task of **Few-shot Implicit Function Generation**. As illustrated in Fig. 1, with only a few new INR checkpoints, we aim to train a generative model that can produce diverse weights while preserving the original distribution's function of these few INR data. These generated checkpoints serve multiple purposes: they can be utilized as new data sources for weight space learning, or they can be directly applicable for practical downstream tasks like 3D shape generation.

This new problem comes with two fundamental challenges. First, the curse of dimensionality [4] exacerbates data demands. Since model weights reside in high-dimensional spaces, a limited number of samples cannot adequately capture the full weight distribution, resulting in poor generalization. Second, conventional few-shot learning methods assume that similar samples have element-wise similarity [26, 69]. However, this assumption doesn't hold for INR checkpoints. INR data can vary significantly in individual weights yet still yield similar input-output behavior, especially under weight permutations[2, 53]. This structural difference makes it difficult to apply standard knowledge transfer techniques effectively, results in potential mode collapse.

Previous approaches to parameter generation, while potentially applicable to Few-shot Implicit Function Generation, face significant limitations and fail to address these fundamental challenges. Some methods use generative models to directly reconstruct the weight distributions [16, 51], they struggle with the curse of dimensionality. While some other methods attempt to address this through autoencoder-based dimensionality reduction [13, 59, 67], they fail to respect the inherent symmetry of the weight space, leading to poor generalization performance.

To address these challenges, we resort to the principle of **equivariance**, *i.e.*, weight space symmetry [22, 46, 76]. It means that, for arbitrary neural networks, permuting output channels in one layer and reversing this permutation in the next layer preserves the network's functionality. Thus, all weights that result from such permutations form an equivariance group [46]. Leveraging this property, we aim to design a generative model capable of producing all samples within an equivariance group from a few checkpoints. In this way, we significantly reduce the data requirements in the few-shot setting.

Based on that, we design a framework named EQUIGEN. Our key insight of equivariance in few-shot setting is that, by projecting weights into an equivariant latent space, we could achieve diverse generation inside an equivariance group from limited examples.

The complete framework contains three stages.

- **Equivariant Encoder.** We design an encoder that projects weights into an equivariant latent space. Using contrastive learning with smooth augmentations, the encoder maps functionally equivalent weights to similar representations. This establishes a structured latent space that captures the concept of equivariance.
- **Equivariance-Guided Diffusion.** We develop a diffusion process guided by equivariance. The denoising steps are conditioned on previous equivariant features and are regularized using an equivariant loss, ensuring that the generated weights retain both distributional properties and equivariance.

- **Controlled Perturbations for Diversity.** To address the challenge of generating diverse yet consistent weights, we apply controlled perturbations in the equivariant subspace. These perturbations enable us to explore the full equivariance group of each example, generating weights that are both diverse and functionally consistent.

We evaluate EQUIGEN on various INR data including 2D images and 3D shapes. Our findings reveal that EQUIGEN not only boosts high quality generation, but also encourages diversity. Ablation studies support the effectiveness of each designed module, emperically prove the importance of equivariance.

Our contributions are summarized as follows:
- We introduce a practical setting named **Few-shot Implicit Function Generation** with an equivariance-based framework EQUIGEN that respect to weight equivariance to enable diverse yet functionally consistent weight generation from limited samples.
- We present three key innovations that systematically exploit equivariance: 1) An equivariant encoder learned through a contrastive learning paradigm with smooth augmentation for robust equivariant feature learning; 2) Equivariance-guided diffusion with explicit equivariance regularization; 3) Controlled equivariant subspace perturbation for diverse generation.
- We demonstrate the effectiveness of our EQUIGEN through comprehensive experiments on INR data from both 2D images and 3D shapes.

2. Related Work

2.1. Implicit Neural Representation

Implicit Neural Representations (INRs) have demonstrated remarkable efficacy in representing diverse forms of complex signals, including spatial occupancy [43, 58, 61], 3D geometric morphology [5, 9, 20], signed distance functions [34, 49], 3D scene appearence [10, 32, 38, 45, 70] and some other complex signals [8, 39, 42, 72] with the help of a small neural network, usually a MLP with few layers.

2.2. Equivariant Architectures

A wide range of studies have tried to build equivariant architecture to respect underlying symmetries in the data. This leads to advantages including smaller parameter space, efficient implementation, and better generalization abilities [12, 29, 46, 82]. The standard construction pattern involves identifying basic equivariant functions which are often linear [22, 41, 76], and composing them with pointwise nonlinearities to build deep networks. Of particular relevance to our work are architectures designed for set-structured data, where the input represents ordered elements requiring equivariance to permutations. A recent study incorporates advantages of several previous works to build
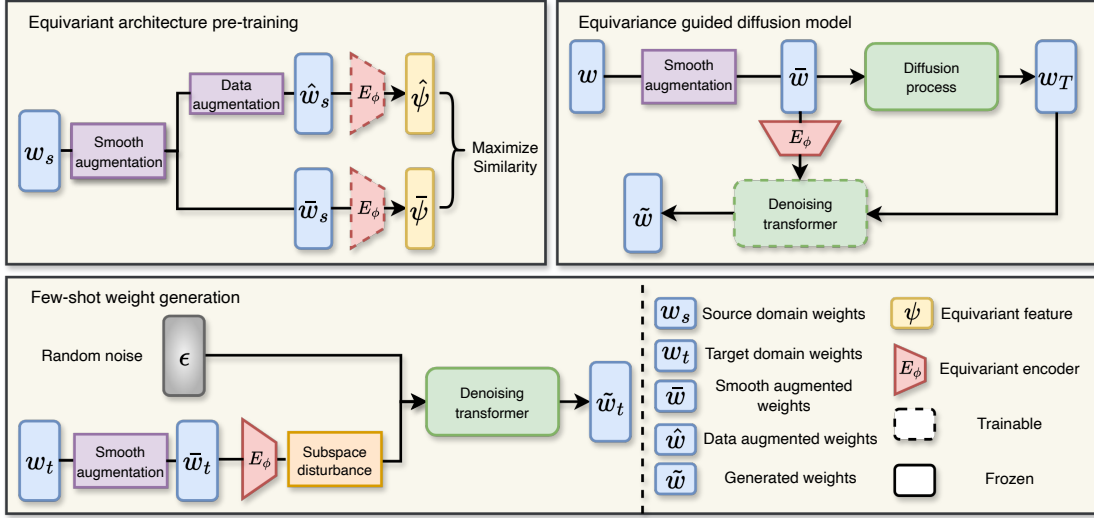
Figure 2. Overview of our EQUIGEN framework. The method consists of three stages: (1) Equivariant Encoder Pre-training through contrastive learning with smooth and INR-based augmentations, (2) Distribution Modeling via a diffusion process conditioned on learned equivariant features, and (3) Few-shot Adaptation using equivariant subspace disturbance for diverse weight generation. Our framework leverages the inherent equivariance of neural network weights to address both generalization and mode collapse challenges.

an efficient and highly expressive architecture [46]. These foundations prove crucial for our work, as neural network weight spaces naturally exhibit similar symmetry structures.

2.3. Neural Network Weight Generation

Recently, neural network weight generation has gained significant attention, from early meta-learning approaches [19, 21] to recent generative modeling techniques. Existing methods broadly split into two categories: direct weight space diffusion methods [16, 51], which struggle with high-dimensional spaces, and dimensionality reduction approaches using autoencoders [13, 59, 67], which often fail to preserve the inherent structural characteristics of weight spaces. The key distinction of our method lies in its systematic implementation of equivariance, which aims at preserving these inherent characteristics.

3. Preliminaries

In this part, we formalize the **Few-shot Implicit Function Generation** problem, interpret fundamental properties of equivariance and corresponding equivariant architectures.

3.1. Problem Definition

Consider a class space $\mathcal{F}$ where each $f \in \mathcal{F}$ represents a category of signals (*e.g.*, images of a specific digit or 3D shapes from a particular object class such as cars or planes) that can be encoded by implicit neural representations. Let $\mathcal{W}$ denotes the weight space of MLPs with a fixed architecture, and $\Phi : \mathcal{W} \to (\mathcal{X} \to \mathcal{Y})$ maps weights to their corresponding implicit functions. We define two key distributions: 1. S_{source}: The source distribution from which we

have abundant training data. 2. S_{target}: The target distribution for which we have only a few examples and want to generate more samples.

Definition 1 (Few-shot INR Generation). *Given k example weights $S_f = \{w_1, ..., w_k\}$ from the target distribution S_{target} that encode valid instances of a target class f, the objective is to generate new weights that encode diverse yet valid instances of the same class. Formally, we aim to learn a generator $G : S_f \mapsto \tilde{p}_f(w)$ that estimates the weight distribution of class f in S_{target}, where validity requires the reconstructed signal $\Phi(w)$ to maintain class-specific properties while allowing intra-class variation.*

3.2. Equivariance and Equivariant architecture

Previous studies revealed that neural networks exhibit inherent equivariance in their weight space. Different weight configurations can represent the same function due to the permutation invariance of neurons within each layer [41, 55, 68, 73]. Understanding and leveraging this equivariance is crucial for Few-shot Implicit Function Generation, as it enables efficient learning from limited examples by exploiting the underlying structure of the weight space [12, 17, 29, 81]. Since INR encoding networks typically employ MLPs that inherently possess this property, we present key concepts of it and corresponding equivariant architectures.

Equivariance. Consider a shallow network with weight matrices W_1 and W_2. For any permutation matrix P, the transformed weights PW_1 and W_2P^T yield a functionally equivalent network. This symmetry emerges from the equivariance of pointwise activation functions:

$$P\sigma(x) = \sigma(Px). \tag{1}$$

This property implies that the space of functionally equiva-

lent networks forms orbits under permutation groups, effectively reducing the sample complexity required for learning.

Equivariant Architectures. To exploit this equivariance, we construct architectures that respect the inherent structure of the weight space. A function $L : \mathcal{V} \to \mathcal{W}$ is *equivariant* if it commutes with group actions ρ_1 and ρ_2 for all $v \in \mathcal{V}, g \in \mathcal{G}$:

$$L(\rho_1(g)v) = \rho_2(g)L(v), \qquad (2)$$

where $\mathcal{V}$ and $\mathcal{W}$ are vector spaces, and $\mathcal{G}$ is a group. Following previous works [22, 41, 76], we define the canonical form of an equivariant encoder:

$$F_{\text{equi}}(x) = L_k \circ \sigma \circ \cdots \circ \sigma \circ L_1(x), \qquad (3)$$

where each L_i is an equivariant affine transformation and σ is a pointwise activation function. By incorporating these symmetries into our architecture, we significantly enhance sample efficiency in few-shot learning. The generator learns to explore the entire orbit of functionally equivalent weights from a few examples, amplifying the limited training data.

4. Methodology

4.1. Overview

Follow the definition of Few-shot Implicit Function Generation problem, our goal is to learn how to generate weights that follow a specific distribution, using only a small number of samples as reference. This problem presents two fundamental challenges as stated in Sec. 1: (1) poor generalization due to the high dimensionality of weight space, and (2) ineffective knowledge transfer and potential mode collapse due to weak element-wise similarity between functionally equivalent weights. We address both challenges through **equivariance** which is interpreted in Sec. 3.2.

Our key insight is that by projecting weights into an equivariant latent space, we can generate diverse samples within an equivariance group using only a few reference examples inside this latent space. This projection effectively reduces dimensionality while preserving functional relationships, enabling both better generalization and meaningful knowledge transfer. As illustrated in Fig. 2, we systematically leverage equivariance across three stages:

Equivariant Feature Learning: We establish a foundation through an equivariant encoder trained via contrastive learning with weight space smooth augmentations, ensuring learned representations preserve essential equivariance property while remaining expressive.

Equivariance-Guided Distribution Modeling: A diffusion-based generator, conditioned on these equivariant features, models the weight distribution while preserving symmetry properties through a designed explicit equivariance regularization.

Equivariant Few-shot Adaptation: For k-shot generation, we introduce controlled disturbances in the equivariant sub-

space, enabling diverse yet functionally consistent generation by exploiting the inherent symmetry structure from the equivariant features.

4.2. Equivariant Encoder Pre-training

A compact and expressive latent space of equivariant features is crucial for effective distribution modeling, improving diffusion convergence and stability. To learn such features, we employ a contrastive learning framework specifically adapted for our equivariant architecture. Additionally, we introduce a weight-specific smooth augmentation strategy to enhance the robustness and representation power of the equivariant encoder.

Equivariant Architecture. Similar to previous work, our equivariant encoder, inspired by the symmetry structure of deep weight spaces, operates directly on weight matrices through a carefully designed architecture that preserves functional equivariance [46, 76]. The encoder decomposes the input weight space into meaningful sub-representations corresponding to different components of a neural network and implements equivariant mappings between these sub-representations through a combination of linear equivariant layers, invariant layers, and efficient pooling and broadcasting operations. Formally, the encoder can be denoted as a mapping $E_\phi : \mathcal{V} \to \mathcal{V}$ parameterized by ϕ following the canonical form demonstrated in Sec. 3.2.

This architecture enables effective projection of the input weight space into an expressive low-dimensional equivariant subspace while preserving the underlying functional relationships. Detailed architecture design can be found in Appendix.

Smooth Augmentation.

The key idea of this encoder is to map weights from an equivariance group to a compact cluster in the equivariant subspace. Yet it can indeed learn from unconstrained weight spaces, recent studies reveal that modern neural networks fundamentally rely on smooth signal modeling [53, 73]. Consequently, to enhance our encoder's feature learning capability, we introduce a smooth augmentation technique for the input weight space. This augmentation could be seen as an optimization of the starting point, and could lead to better representation power of the pre-trained encoder. The procedures are as follows: 1. We represent neural networks as dependency graphs $G = (V, E)$ [18], where each node $v_i \in V$ represents an operation with weight W_i and each edge $e_{ij} \in E$ indicates inter-connectivity between v_i and v_j. 2. Each graph clique $C = (V_C, E_C) \subset G$ is fully connected, representing a group of interconnected operations. Each clique C corresponds to a unique permutation matrix P. 3. Our objective is to determine all permutation matrices P in a way that minimizes the total variation [57] across the whole network. This complex optimization can be broken down into multiple independent optimizations,
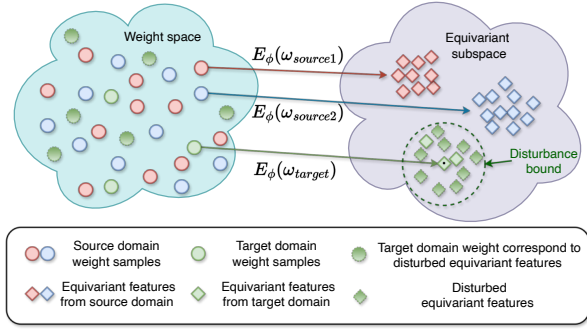
Figure 3. Equivariant architecture E_ϕ aims to map weights from the same equivariance group to similar representations, creating a structured latent space that captures the inherent symmetries of neural networks. By leveraging this equivariant subspace, we can implement a controlled disturbance strategy that sample diverse equivariant features while maintaining class consistency.

each focused on a single clique [73]. 4. We define this as a multi-objective Shortest Hamiltonian Path (mSHP) problem [23]. To address each mSHP problem, we transform it into a Traveling Salesman Problem (TSP) by adding a dummy node with zero-distance edges to all other nodes in the clique. 5. We solve this TSP using a 2.5-opt local search algorithm [63]. The resulting optimal permutation P^* is applied to all weight matrices within the clique.

Contrastive Learning. Following the SimCLR framework [7], we pre-train our equivariant encoder to maximize similarity between different augmented views of the same weight data in the equivariant latent space. The process involves three key steps:

1) We generate positive pairs by applying stochastic INR-based augmentations to smoothed weights $\bar{w}$, obtaining augmented weights $\hat{w}$. These augmentations include rotation, translation, scaling, and our proposed color jittering and bias perturbation (see Appendix for details).

2) Extract equivariant features through pre-trained equivariant encoder E_ϕ: $\bar{\psi} = E_\phi(\bar{w})$ and $\hat{\psi} = E_\phi(\hat{w})$.

3) Optimize using contrastive loss:

$$\ell_{i,j} = -\log \frac{\exp(sim(E_\phi(\check{w}_i), E_\phi(\check{w}_j))/\tau)}{\sum_{k \neq i} \exp(sim(E_\phi(\check{w}_i), E_\phi(\check{w}_k))/\tau)}, \quad (4)$$

where $sim(u, v)$ denotes cosine similarity, τ is a temperature parameter. $\{\check{w}_k\}$ contains both the smoothed weights $\{\bar{w}_k\}$ and the augmented weights $\{\hat{w}_k\}$. Negative pairs are implicitly formed within the contrastive loss.

4.3. Equivariance Guided Diffusion

Within this stage, we employ an equivariance-conditioned diffusion model to reconstruct the input weight distributions, utilizing their corresponding equivariant features. This approach enables a robust reconstruction of the weight space, while maintaining the equivariance inherent in the original data structure.
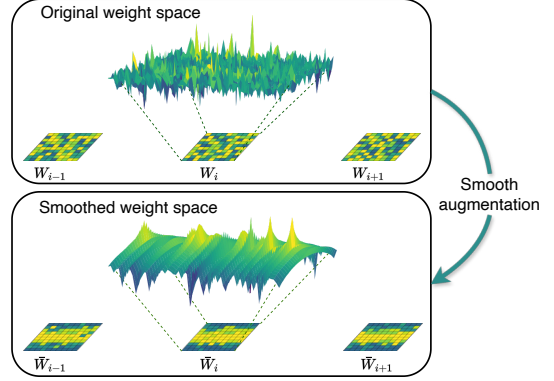

Figure 4. To optimize the equivariant feature extraction, we seek permutations that minimize the weight matrix's total variation. This smoothing operation reduces abrupt discontinuities in the weight space, facilitates more effective learning of inherent equivariant properties by starting from an optimized point. The smoother weight manifold (bottom) enables better equivariant feature capture compared to the original space (up).

Weight-Space Diffusion. The modeling begins by transforming each set of smoothed weights $\{\bar{w}_i\}$ into a flattened one-dimensional vector representation. Subsequently, these smoothed weights $\bar{w}_i$ undergo a diffusion process, yielding noised samples $\bar{w}_T$. We leverage the transformer architecture [50, 66, 74] G_θ as the denoising network. This choice leverages transformers' established capability to efficiently process long vector sequences, advantageous for weight space manipulation [16, 51]. Our denoising network follows previous works to predict the denoised weightsdirectly [51], to faster the convergence. The architecture also incorporates cross-attention layers following each transformer block [56], enabling the sampled weights $\tilde{w}_i$ to be conditioned on equivariant features ψ_i. These features are derived by processing the smoothed weights $\bar{w}_i$ through the former pre-trained equivariant encoder E_ϕ. The comprehensive denoising procedure can be expressed as:

$$\tilde{w}_i = G_\theta(\bar{w}_T, E_\phi(\bar{w}_i)). \quad (5)$$

During inference, we use the DDIM sampler with 1000 steps to generate the final weights, transforming $\bar{w}_T$ into $\tilde{w}_i$ through a series of intermediate denoising steps.

Training and Optimization. The primary goal of our approach is to optimize the denoising network to model the input distribution accurately. We apply a simple Mean Squared Error (MSE) loss between the denoised weights $\tilde{w}_i$ and the smoothed weight $\bar{w}_i$ denoted as $\mathcal{L}_{recon}(\bar{w}, \tilde{w})$.

However, in order to leverage the characteristic of equivariance in the following few-shot generation stage, we also designed a specific loss to regulate the approximation of the equivariant features of the generated weights and the original smoothed weights:

$$\mathcal{L}_{eq}(\bar{w}, \tilde{w}) = \frac{1}{N} \sum_{i=1}^{N} \|E_\phi(\bar{w}_i) - E_\phi(\tilde{w}_i)\|_2^2. \quad (6)$$

By composing these two different objectives, our final target is to minimize the composite objective by sampling all the smoothed weight and generated weight:

$$\min_{\theta} \mathbb{E}_{\bar{w}_i, \hat{w}_i} \left[\mathcal{L}_{total} \right] = \min_{\theta} \mathbb{E}_{\bar{w}_i, \hat{w}_i} \left[\mathcal{L}_{recon} + \lambda \mathcal{L}_{eq} \right]. \quad (7)$$

This loss function not only ensuring accurate distribution reconstruction, but also boosting the generalization.

4.4. Few-shot Fine-tuning

Through latent space guidance, the diffusion model demonstrates adaptability to previously unseen distributions, given only a limited set of weights $\{w'_1, \ldots, w'_k\}$ from the target distribution. This adaptation is facilitated through the utilization of equivariant features $\{\psi'_1, \ldots, \psi'_k\}$ extracted via the pre-trained equivariant encoder, serving as novel guidance vectors that enable knowledge transfer. The adaptation process need only a few iterations of fine-tuning, following the same procedures in Sec. 4.3.

Subspace Disturbance. To enhance the diversity of generated weights, we introduce an equivariant subspace disturbance strategy during the final generation. This approach is motivated by the observation that the diffusion model's training process inherently embeds knowledge within the equivariant subspace through feature-guided denoising. Consequently, the diversity of the guiding equivariant features directly influences the variability of the generated weights, establishing a foundation for controlled diversity in weight generation.

Specifically, we apply a random Gaussian noise $\epsilon \sim \mathcal{N}(0, I)$ on the equivariant feature ψ'_i with a controlling parameter γ: $\tilde{\psi}'_i = \psi'_i + \gamma \epsilon$. We impose this constraint on the magnitude of the disturbance to make it bounded and controllable. Finally, the generation of target weights is accomplished through a denoising procedure applied to k independent Gaussian noise vectors $\mathcal{E} = \{\epsilon_1, \ldots, \epsilon_k\}$. This process is guided by disturbed equivariant features $\{\tilde{\psi}'_1, \ldots, \tilde{\psi}'_k\}$ derived from the support set.

5. Experiments

5.1. Experimental Setup

We establish two Few-shot Implicit Function Generation scenarios: 1) Given a set of MLP encoded INRs of different **2D images**, we split them into two disjoint parts: the seen categories S_{source} and the unseen categories S_{test}, where category is defined by their rendered image. 2) Similarly, given a set of MLP encoded INRs of different **3D shapes**, which are split into two disjoint parts: the seen categories S_{source} and the unseen categories S_{test}, where category is defined by their rendered 3D shape.

Datasets. For the 2D image scenario, we evaluate our methodology on two benchmark datasets: MNIST (2D greyscale images) [31] and CIFAR-10 (2D RGB im-

ages) [30]. We utilize the pre-overfitted INR dataset for MNIST (MNIST-INRs) provided by [46] and the corresponding dataset for CIFAR-10 (CIFAR-10-INRs) from [40]. MNIST-INRs encompasses 50K INR instances, with 5K samples per category, while CIFAR-10-INRs contains 60K INR instances, comprising 6K samples per category. For the 3D shape domain, we evaluate our approach on the ShapeNet dataset [6], focusing on three representative categories: airplane (4045 shapes), car (6778 shapes) and chair (3533 shapes). These shapes are encoded into MLPs, denoted as ShapeNet-INRs. Regarding architectural specifications, CIFAR-10-INRs employs a 3-layer SIREN MLP with a width of 64 [61], while MNIST-INRs utilizes a 3-layer SIREN MLP with a width of 32 [61]. For ShapeNet-INRs, it implements a 3-layer standard MLP with a width of 128, incorporating ReLU activation functions and input positional encoding [16].

In the context of few-shot weight generation, we partition each dataset into two distinct subsets [14, 48, 78, 79]. The unseen category support set S_{test} comprises randomly sampled INR data from a single category, with the sample size varying from 1 to 10 to evaluate different levels of difficulty. The remaining data from other categories constitute the seen category support set S_{source}.

Implementation We provide important configurations here, please refer to Appendix for more detailed information. Our equivariant architecture is implemented with four hidden equivariant layers [46], with the output equivariant feature dimension set to 128. Across the three stages: In the equivariance-guided diffusion stage, we utilize a squared cosine beta scheduler [47] across 1000 timesteps and implement DDIM [62] for sampling. The equivariance loss proportion parameter λ is set to 0.1 unless otherwise specified. Finally, during the generation process, the subspace disturbance parameter γ, which controls the noise intensity, is set to 0.3 by default.

Metrics. Since there is no direct evaluation metrics for generated weights, we adopt an indirect evaluation strategy aligned with existing works [13, 15, 16, 59, 67]. For 2d images, a direct rendering is applied and for 3d shape scenario, we extract the underlying isosurface from the MLPs with Marching Cubes [35]. Our evaluation framework assesses both the quality and diversity of generated results for both two scenarios: For the 2D image scenario, we employ FID [25] and LPIPS [77] as the metrics. Specifically, we apply the intra-cluster version of LPIPS to quantify the diversity of generated unseen images in our few-shot generation context like many other few-shot image generation settings do [14, 56, 79]. Lower FID indicates better quality and higher LPIPS indicates enhanced diversity. For the 3D shape scenario, we follow prior works [37, 60, 65, 83] in evaluating MMD, COV, and 1-NNA. For MMD, lower is better; for COV, higher is better; for 1-NNA, 50% is the op-

Table 1. Quantitative comparison of 10-shot generation performance on ShapeNet-INRs, where each test set consists of INRs from a single unseen object class. **Bold** indicates the best result and underline indicates the second best result.

Category	Method	MMD↓	COV(%)↑	1-NNA(%)↓
Airplane	PVD [83]	5.6	26	84.7
	DPC [37]	4.1	29	84.6
	INR2Vec [13]	5.1	29	80.9
	Voxel Baseline [16]	8.5	20	98.3
	HyperDiffusion [16]	5.0	33	80.6
	EQUIGEN (Ours)	**3.4**	**35**	**73.0**
Car	PVD [83]	4.7	24	82.1
	DPC [37]	4.2	20	85.3
	INR2Vec [13]	4.8	19	87.0
	Voxel Baseline [16]	5.8	11	97.9
	HyperDiffusion [16]	4.4	22	83.4
	EQUIGEN (Ours)	**3.5**	**31**	**76.5**
Chair	PVD [83]	7.1	21	80.0
	DPC [37]	6.5	23	81.5
	INR2Vec [13]	6.8	25	71.7
	Voxel Baseline [16]	12.4	13	88.5
	HyperDiffusion [16]	7.2	28	73.1
	EQUIGEN (Ours)	**4.2**	**41**	**67.1**

Table 2. Quantitative comparison of 10-shot generation performance on MNIST-INRs and CIFAR-10-INRs.

Dataset	Method	FID↓	LPIPS↑
MNIST-INRs	FIGR [11]	160.26	0.1399
	DAGAN [1]	149.77	0.2001
	DAWSON [33]	164.90	0.1830
	INR2Vec [13]	157.20	0.2185
	HyperDiffsion [16]	165.91	0.2727
	EQUIGEN (Ours)	**121.24**	**0.4133**
CIFAR10-INRs	FIGR [11]	225.05	0.1728
	DAGAN [1]	200.11	0.2250
	DAWSON [33]	204.62	0.2191
	INR2Vec [13]	188.39	0.2712
	Hiperdiffusion [16]	186.58	0.3104
	EQUIGEN (Ours)	**164.14**	**0.4926**

timal. We compute these metrics using Chamfer Distance (CD) as the underlying distance measure, with reported CD values scaled by a factor of 10^2 for clarity.

Baselines. Our method is benchmarked against two different types of methods: modality-based methods and INR-based methods. In the first category, we benchmark against methods that directly manipulate domain-specific representations. For the 2D image domain, we compare our approach with representative few-shot image generation methods, including FIGR [11], DAGAN [1] and DAWSON [33]. Similarly, in the 3D shape domain, we evaluate against prominent 3D shape generation methods such as PVD [83] and DPC [37]. In the second category, we evaluate against methods that operate directly on INR representations rather than domain-specific data formats. Specifically, we compare our approach with two representative frameworks: INR2Vec [13] and HyperDiffusion [16], both of which are assessed across 2D image and 3D shape scenarios to ensure comprehensive evaluation. Specifically, HyperDiffu-
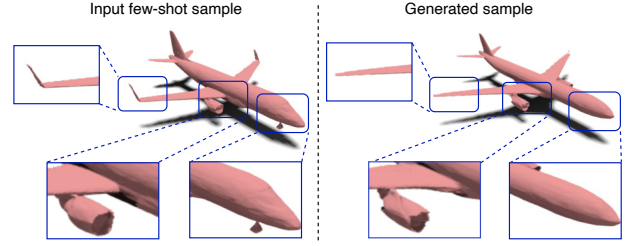

Figure 5. Visualizations of generated ShapeNet-INRs from the few-shot airplane example. The outputs exhibit diverse shape variations while preserving the airplane category characteristics.
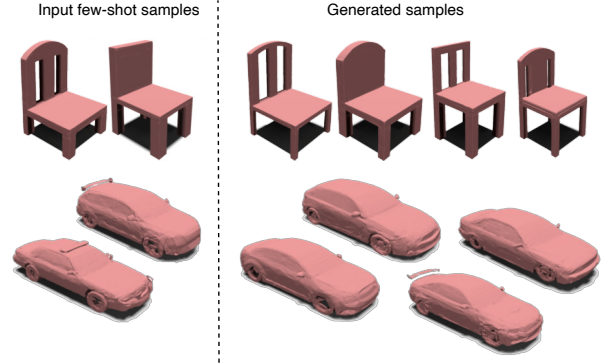

Figure 6. Visualizations of generated ShapeNet-INRs: few-shot examples and generated samples for both chair and car categories.

sion has a voxel version (Voxel Baseline) whose INR data are designed to encode voxel information instead of meshes.

5.2. Main Results

The empirical results illustrated in Tab. 1 and Tab. 2 demonstrate the consistent superior performance of our method across all evaluation metrics in the 10-shot setting. In Tab. 2, our comprehensive evaluation on both MNIST-INRs and CIFAR10-INRs reveals that our approach surpasses both modality-based and INR-based methods in terms of quality and diversity metrics. Notably, while most existing few-shot generation frameworks address style transfer, our method is designed to tackle the more challenging task of cross-class knowledge transfer. Similar evaluation results are observed in the 3D domain, as demonstrated in Tab. 1. Qualitative results are shown in Fig. 5 and Fig. 6. There are various differences between input INR and output INR's rendered results, indicating the effectiveness of our method.

5.3. Indepth Analysis

Exploration of equivariant subspace. The efficacy of our proposed method is fundamentally rooted in the expressive power of the equivariant subspace, which enables the incorporation of category-specific knowledge into distribution modeling. To illustrate the learning dynamics within the weight space, we conduct a 2D t-SNE visualization of the equivariant subspace, as shown in Fig. 7. We present comparative visualizations both with and without

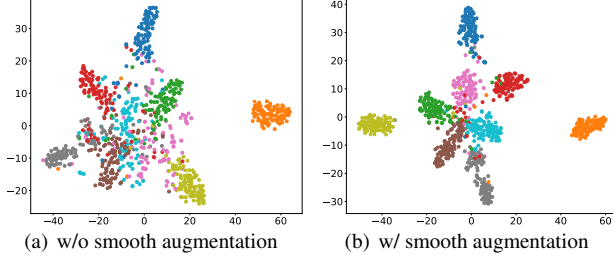(a) w/o smooth augmentation   (b) w/ smooth augmentation

Figure 7. 2D t-SNE visualization of equivariant subspace. Smooth augmentation (right) produces more compact and discriminative category clusters compared to baseline (left), demonstrating enhanced equivariant feature learning.

(a) COV with different γ (b) MMD with different γ
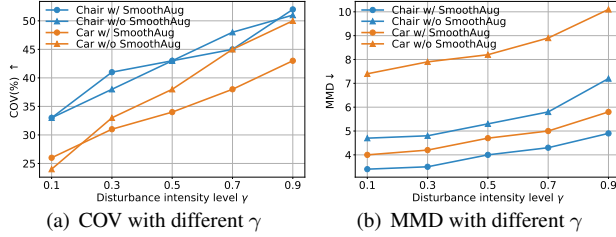
Figure 8. The evaluation on ShapeNet-INRs chair and car categories with respect to different subspace disturbance intensity. (a) Higher disturbance leads to increased COV, indicating greater sample diversity. (b) However, larger disturbances result in higher MMD, reflecting decreased generation quality.

smooth augmentation of the original weight space. The results demonstrate that the pre-trained equivariant encoder successfully projects the original weight space into a discriminative manifold where categories exhibit clear clustering behavior. Furthermore, the application of smooth augmentation to the original space alleviates the complexity of equivariant subspace learning, resulting in more compact and category-specific embeddings.

Impact of subspace disturbance. We conduct a systematic analysis of equivariant subspace disturbance intensity and its impact on generation performance, as illustrated in Fig. 8. Our findings demonstrate a clear trade-off between sample diversity and functional fidelity. The intensity of subspace disturbance exhibits a direct correlation with final generation variance. As shown in Fig. 8(a), higher disturbance intensities lead to increased COV, indicating greater diversity among generated samples. However, Fig. 8(b) reveals that this enhanced diversity comes at a cost to generation quality, as measured by increasing MMD.

Ablation study. To evaluate the importance of equivariance, we conduct ablation studies comparing three variants: (1) our full EQUIGEN framework, (2) a variant that omits the equivariant encoder and uses unconditional diffusion, and (3) a variant that replaces equivariant features with class-label conditioning. As shown in Tab. 3, the framework with equivariance modules demonstrates superior performance in both generation quality and diversity. The uncon-

Table 3. Ablation study of the equivariant encoder on MNIST-INRs. The incorporation of equivariance largely enhances the overall generation quality and diversity compared to different variants that omit equivariance.

Methods	FID↓	LPIPS↑
EQUIGEN w/o condition	164.78	0.2432
EQUIGEN w/ label condition	162.56	0.2437
EQUIGEN w/ equivariant condition	**121.24**	**0.4133**

Table 4. Ablation study of weight space smooth augmentation and equivariant subspace disturbance on ShapeNet-INRs, where each test set consists of INRs from a single unseen object class.

Category	Smooth	Disturbance	MMD↓	COV(%)↑	1-NNA(%)↓
Airplane	✗	✗	5.5	29	79.4
	✓	✗	**3.2**	30	77.5
	✗	✓	5.7	**35**	<u>75.2</u>
	✓	✓	<u>3.4</u>	**35**	**73.0**
Car	✗	✗	4.7	22	83.9
	✓	✗	**3.4**	25	<u>78.4</u>
	✗	✓	4.8	**33**	78.5
	✓	✓	<u>3.5</u>	<u>31</u>	**76.5**
Chair	✗	✗	7.4	30	77.5
	✓	✗	**4.2**	30	<u>76.3</u>
	✗	✓	7.9	<u>38</u>	77.0
	✓	✓	**4.2**	**41**	**67.1**

ditional variant exhibits mode collapse, and class-label conditioning variant offers only minimal improvement, demonstrating the limitations of direct class conditioning.

Another more fine-grained ablation study is conducted to discern the individual contributions of weight space smooth augmentation and equivariant subspace disturbance. The results are presented in Tab. 4. Our analysis reveals that smooth augmentation alone enhances both generation quality and diversity. This empirical evidence supports the idea that starting from an optimized start point, the equivariant encoder could gain better expressive power through contrastive learning. Conversely, applying subspace disturbance yields increased diversity, although a modest decrease in generation quality appears when smooth augmentation is absent. This trade-off can be attributed to the fact that equivariant features with same disturbance may locate at different category clusters with and without the smooth augmentation. The simultaneous employment of both smooth augmentation and subspace disturbance enables our method to achieve optimal performance.

6. Conclusion

In this work, we present a practical setting named Few-shot Implicit Function Generation and introduce solution to it by leveraging the principle of equivariance. Our three-stage framework addresses both the generalization and mode collapse challenges inherent in weight space generation. Extensive experiments across INR data for 2D image and 3D shape domains validate our approach, demonstrating superior performance in both generation quality and diversity.

7. Acknowledgement

References

[1] A Antoniou. Data augmentation generative adversarial networks. *arXiv preprint arXiv:1711.04340*, 2017. 7

[2] Maor Ashkenazi, Zohar Rimon, Ron Vainshtein, Shir Levi, Elad Richardson, Pinchas Mintz, and Eran Treister. Nern: Learning neural representations for neural networks. In *ICLR*, 2023. 2

[3] Jonathan T Barron, Ben Mildenhall, Matthew Tancik, Peter Hedman, Ricardo Martin-Brualla, and Pratul P Srinivasan. Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. In *CVPR*, pages 5855–5864, 2021. 1

[4] Richard Bellman. Dynamic programming. *Science*, 153 (3731):34–37, 1966. 2

[5] Rohan Chabra, Jan E Lenssen, Eddy Ilg, Tanner Schmidt, Julian Straub, Steven Lovegrove, and Richard Newcombe. Deep local shapes: Learning local sdf priors for detailed 3d reconstruction. In *ECCV*, pages 608–625. Springer, 2020. 2

[6] Angel X Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, et al. Shapenet: An information-rich 3d model repository. *arXiv preprint arXiv:1512.03012*, 2015. 6

[7] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *ICML*, pages 1597–1607. PMLR, 2020. 5

[8] Yinbo Chen, Sifei Liu, and Xiaolong Wang. Learning continuous image representation with local implicit image function. In *CVPR*, pages 8628–8638, 2021. 1, 2

[9] Zhiqin Chen and Hao Zhang. Learning implicit fields for generative shape modeling. In *CVPR*, pages 5939–5948, 2019. 2

[10] Zhang Chen, Zhong Li, Liangchen Song, Lele Chen, Jingyi Yu, Junsong Yuan, and Yi Xu. Neurbf: A neural fields representation with adaptive radial basis functions. In *CVPR*, pages 4182–4194, 2023. 2

[11] Louis Clouâtre and Marc Demers. Figr: Few-shot image generation with reptile. *arXiv preprint arXiv:1901.02199*, 2019. 7

[12] Taco S Cohen, Mario Geiger, Jonas Köhler, and Max Welling. Spherical cnns. In *ICLR*, 2018. 2, 3

[13] Luca De Luigi, Adriano Cardace, Riccardo Spezialetti, Pierluigi Zama Ramirez, Samuele Salti, and Luigi di Stefano. Deep learning on implicit neural representations of shapes. In *ICLR*, 2023. 1, 2, 3, 6, 7

[14] Yuxuan Duan, Li Niu, Yan Hong, and Liqing Zhang. Weditgan: Few-shot image generation via latent space relocation. In *AAAI*, pages 1653–1661, 2024. 6

[15] Emilien Dupont, Hyunjik Kim, SM Ali Eslami, Danilo Jimenez Rezende, and Dan Rosenbaum. From data to functa: Your data point is a function and you can treat it like one. In *ICML*, pages 5694–5725. PMLR, 2022. 1, 6

[16] Ziya Erkoç, Fangchang Ma, Qi Shan, Matthias Nießner, and Angela Dai. Hyperdiffusion: Generating implicit neural fields with weight-space diffusion. In *CVPR*, pages 14300–14310, 2023. 2, 3, 5, 6, 7

[17] Carlos Esteves, Christine Allen-Blanchette, Ameesh Makadia, and Kostas Daniilidis. Learning so (3) equivariant representations with spherical cnns. In *ECCV*, pages 52–68, 2018. 3

[18] Gongfan Fang, Xinyin Ma, Mingli Song, Michael Bi Mi, and Xinchao Wang. Depgraph: Towards any structural pruning. In *CVPR*, pages 16091–16101, 2023. 4

[19] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In *ICML*, pages 1126–1135. PMLR, 2017. 3

[20] Kyle Genova, Forrester Cole, Daniel Vlasic, Aaron Sarna, William T Freeman, and Thomas Funkhouser. Learning shape templates with structured implicit functions. In *CVPR*, pages 7154–7164, 2019. 2

[21] David Ha, Andrew Dai, and Quoc V Le. Hypernetworks. *arXiv preprint arXiv:1609.09106*, 2016. 3

[22] Jason Hartford, Devon Graham, Kevin Leyton-Brown, and Siamak Ravanbakhsh. Deep models of interactions across sets. In *ICLR*, pages 1909–1918. PMLR, 2018. 2, 4

[23] Juris Hartmanis. Computers and intractability: a guide to the theory of np-completeness (michael r. garey and david s. johnson). *Siam Review*, 24(1):90, 1982. 5

[24] Vincent Herrmann, Francesco Faccio, and Jürgen Schmidhuber. Learning useful representations of recurrent neural network weight matrices. In *ICML*, 2024. 1

[25] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In *NeurIPS*, 2017. 6

[26] Teng Hu, Jiangning Zhang, Ran Yi, Yuzhen Du, Xu Chen, Liang Liu, Yabiao Wang, and Chengjie Wang. Anomalydiffusion: Few-shot anomaly image generation with diffusion model. In *AAAI*, pages 8526–8534, 2024. 2

[27] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon Kohl, Andrew Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, and Demis Hassabis. Highly accurate protein structure prediction with alphafold. *Nature*, 596:1–11, 2021. 1

[28] Boris Knyazev, Michal Drozdzal, Graham W Taylor, and Adriana Romero Soriano. Parameter prediction for unseen deep architectures. In *NeurIPS*, pages 29433–29448, 2021. 1

[29] Risi Kondor and Shubhendu Trivedi. On the generalization of equivariance and convolution in neural networks to the action of compact groups. In *ICML*, pages 2747–2755. PMLR, 2018. 2, 3

[30] Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. 2009. 6

[31] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998. 6

[32] Zhengqi Li, Qianqian Wang, Forrester Cole, Richard Tucker, and Noah Snavely. Dynibar: Neural dynamic image-based rendering. In *CVPR*, pages 4273–4284, 2023. 2

[33] Weixin Liang, Zixuan Liu, and Can Liu. Dawson: A domain adaptive few shot generation framework. *arXiv preprint arXiv:2001.00576*, 2020. 7

[34] Shaohui Liu, Yinda Zhang, Songyou Peng, Boxin Shi, Marc Pollefeys, and Zhaopeng Cui. Dist: Rendering deep implicit signed distance function with differentiable sphere tracing. In *CVPR*, pages 2019–2028, 2020. 2

[35] William E Lorensen and Harvey E Cline. Marching cubes: A high resolution 3d surface construction algorithm. In *Seminal graphics: pioneering efforts that shaped the field*, pages 347–353. 1998. 6

[36] Yuxiang Lu, Suizhi Huang, Yuwen Yang, Shalayiding Sirejiding, Yue Ding, and Hongtao Lu. Fedhca2: Towards hetero-client federated multi-task learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5599–5609, 2024. 1

[37] Shitong Luo and Wei Hu. Diffusion probabilistic models for 3d point cloud generation. In *CVPR*, pages 2837–2845, 2021. 6, 7

[38] Qi Ma, Danda Pani Paudel, Ajad Chhatkuli, and Luc Van Gool. Deformable neural radiance fields using rgb and event cameras. In *CVPR*, pages 3590–3600, 2023. 2

[39] Qi Ma, Danda Pani Paudel, Ajad Chhatkuli, and Luc Van Gool. Continuous pose for monocular cameras in neural implicit representation. In *CVPR*, pages 5291–5301, 2024. 2

[40] Qi Ma, Danda Pani Paudel, Ender Konukoglu, and Luc Van Gool. Implicit-zoo: A large-scale dataset of neural implicit functions for 2d images and 3d scenes. *arXiv preprint arXiv:2406.17438*, 2024. 1, 6

[41] Haggai Maron, Heli Ben-Hamu, Nadav Shamir, and Yaron Lipman. Invariant and equivariant graph networks. In *ICLR*, 2018. 2, 3, 4

[42] Julian McGinnis, Suprosanna Shit, Hongwei Bran Li, Vasiliki Sideri-Lampretsa, Robert Graf, Maik Dannecker, Jiazhen Pan, Nil Stolt-Ansó, Mark Mühlau, Jan S Kirschke, et al. Single-subject multi-contrast mri super-resolution via implicit neural representations. In *MICCAI*, pages 173–183. Springer, 2023. 2

[43] Lars Mescheder, Michael Oechsle, Michael Niemeyer, Sebastian Nowozin, and Andreas Geiger. Occupancy networks: Learning 3d reconstruction in function space. In *CVPR*, pages 4460–4470, 2019. 2

[44] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021. 1

[45] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM transactions on graphics (TOG)*, 41(4):1–15, 2022. 2

[46] Aviv Navon, Aviv Shamsian, Idan Achituve, Ethan Fetaya, Gal Chechik, and Haggai Maron. Equivariant architectures for learning in deep weight spaces. In *ICML*, pages 25790–25816. PMLR, 2023. 1, 2, 3, 4, 6

[47] Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *ICLR*, pages 8162–8171. PMLR, 2021. 6

[48] Atsuhiro Noguchi and Tatsuya Harada. Image generation from small datasets via batch statistics adaptation. In *CVPR*, pages 2750–2758, 2019. 6

[49] Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. Deepsdf: Learning continuous signed distance functions for shape representation. In *CVPR*, pages 165–174, 2019. 1, 2

[50] William Peebles and Saining Xie. Scalable diffusion models with transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4195–4205, 2023. 5

[51] William Peebles, Ilija Radosavovic, Tim Brooks, Alexei Efros, and Jitendra Malik. Learning to learn with generative models of neural network checkpoints. *arXiv preprint arXiv:2209.12892*, 2022. 1, 2, 3, 5

[52] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *ICML*, pages 8748–8763. PMLR, 2021. 1

[53] Nasim Rahaman, Aristide Baratin, Devansh Arpit, Felix Draxler, Min Lin, Fred Hamprecht, Yoshua Bengio, and Aaron Courville. On the spectral bias of neural networks. In *ICML*, pages 5301–5310. PMLR, 2019. 2, 4

[54] Sameera Ramasinghe and Simon Lucey. Beyond periodicity: Towards a unifying framework for activations in coordinate-mlps. In *ECCV*, pages 142–158. Springer, 2022. 1

[55] Siamak Ravanbakhsh, Jeff Schneider, and Barnabas Poczos. Equivariance through parameter-sharing. In *ICML*, pages 2892–2901. PMLR, 2017. 3

[56] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *CVPR*, pages 10684–10695, 2022. 5, 6

[57] Leonid I Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena*, 60(1-4):259–268, 1992. 4

[58] Vishwanath Saragadam, Daniel LeJeune, Jasper Tan, Guha Balakrishnan, Ashok Veeraraghavan, and Richard G Baraniuk. Wire: Wavelet implicit neural representations. In *CVPR*, pages 18507–18516, 2023. 2

[59] Konstantin Schürholt, Boris Knyazev, Xavier Giró-i Nieto, and Damian Borth. Hyper-representations as generative models: Sampling unseen neural network weights. In *NeurIPS*, pages 27906–27920, 2022. 2, 3, 6

[60] 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. In *CVPR*, pages 19615–19625, 2024. 6

[61] Vincent Sitzmann, Julien Martel, Alexander Bergman, David Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. In *NeurIPS*, pages 7462–7473, 2020. 1, 2, 6

[62] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *ICLR*, 2021. 6

[63] Günther Stattenberger, Markus Dankesreiter, Florian Baumgartner, and Johannes J Schneider. On the neighborhood structure of the traveling salesman problem generated by local search moves. *Journal of Statistical Physics*, 129:623–648, 2007. 5

[64] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*, 2023. 1

[65] Arash Vahdat, Francis Williams, Zan Gojcic, Or Litany, Sanja Fidler, Karsten Kreis, et al. Lion: Latent point diffusion models for 3d shape generation. In *NeurIPS*, pages 10021–10039, 2022. 6

[66] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017. 5

[67] Kai Wang, Zhaopan Xu, Yukun Zhou, Zelin Zang, Trevor Darrell, Zhuang Liu, and Yang You. Neural network diffusion. *arXiv preprint arXiv:2402.13144*, 2024. 2, 3, 6

[68] Jeffrey Wood and John Shawe-Taylor. Representation theory and invariant neural networks. *Discrete applied mathematics*, 69(1-2):33–60, 1996. 3

[69] Jiayu Xiao, Liang Li, Chaofei Wang, Zheng-Jun Zha, and Qingming Huang. Few shot generative model adaption via relaxed spatial structural alignment. In *CVPR*, pages 11204–11213, 2022. 2

[70] Yiheng Xie, Towaki Takikawa, Shunsuke Saito, Or Litany, Shiqin Yan, Numair Khan, Federico Tombari, James Tompkin, Vincent Sitzmann, and Srinath Sridhar. Neural fields in visual computing and beyond. In *Computer Graphics Forum*, pages 641–676. Wiley Online Library, 2022. 2

[71] Dejia Xu, Peihao Wang, Yifan Jiang, Zhiwen Fan, and Zhangyang Wang. Signal processing for implicit neural representations. In *NeurIPS*, pages 13404–13418, 2022. 1

[72] Junshen Xu, Daniel Moyer, Borjan Gagoski, Juan Eugenio Iglesias, P Ellen Grant, Polina Golland, and Elfar Adalsteinsson. Nesvor: implicit neural representation for slice-to-volume reconstruction in mri. *IEEE TMI*, 42(6):1707–1719, 2023. 2

[73] Xingyi Yang and Xinchao Wang. Neural metamorphosis. In *ECCV*, pages 1–19. Springer, 2025. 3, 4, 5

[74] Xingyi Yang and Xinchao Wang. Kolmogorov-arnold transformer. In *The Thirteenth International Conference on Learning Representations*, 2025. 5

[75] Tackgeun You, Mijeong Kim, Jungtaek Kim, and Bohyung Han. Generative neural fields by mixtures of neural implicit functions. In *NeurIPS*, 2024. 1

[76] Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Russ R Salakhutdinov, and Alexander J Smola. Deep sets. In *NeurIPS*, 2017. 2, 4

[77] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, pages 586–595, 2018. 6

[78] Yunqing Zhao, Chao Du, Milad Abdollahzadeh, Tianyu Pang, Min Lin, Shuicheng Yan, and Ngai-Man Cheung. Exploring incompatible knowledge transfer in few-shot image generation. In *CVPR*, pages 7380–7391, 2023. 6

[79] Chenxi Zheng, Bangzhen Liu, Huaidong Zhang, Xuemiao Xu, and Shengfeng He. Where is my spot? few-shot image generation via latent subspace optimization. In *CVPR*, pages 3272–3281, 2023. 6

[80] Andrey Zhmoginov, Mark Sandler, and Maksym Vladymyrov. Hypertransformer: Model generation for supervised and semi-supervised few-shot learning. In *ICML*, pages 27075–27098. PMLR, 2022. 1

[81] Allan Zhou, Kaien Yang, Kaylee Burns, Adriano Cardace, Yiding Jiang, Samuel Sokota, J Zico Kolter, and Chelsea Finn. Permutation equivariant neural functionals. In *NeurIPS*, 2024. 3

[82] Allan Zhou, Kaien Yang, Yiding Jiang, Kaylee Burns, Winnie Xu, Samuel Sokota, J Zico Kolter, and Chelsea Finn. Neural functional transformers. In *NeurIPS*, 2024. 2

[83] Linqi Zhou, Yilun Du, and Jiajun Wu. 3d shape generation and completion through point-voxel diffusion. In *CVPR*, pages 5826–5835, 2021. 6, 7