NEUGEN: AMPLIFYING THE 'NEURAL' IN NEURAL RADIANCE FIELDS FOR DOMAIN GENERALIZATION

Anonymous authors

004

010 011

012

013

014

015

016

017

018

019

021

025

026

Paper under double-blind review

Abstract

Neural Radiance Fields (NeRF) have significantly advanced the field of novel view synthesis, yet their generalization across diverse scenes and conditions remains challenging. Addressing this, we propose the integration of a novel braininspired normalization technique Neural Generalization (NeuGen) into leading NeRF architectures which include MVSNeRF and GeoNeRF. NeuGen extracts the domain-invariant features, thereby enhancing the models' generalization capabilities. It can be seamlessly integrated into NeRF architectures and cultivates a comprehensive feature set that significantly improves accuracy and robustness in image rendering. Through this integration, NeuGen shows improved performance on benchmarks on diverse datasets across state-of-the-art NeRF architectures, enabling them to generalize better across varied scenes. Our comprehensive evaluations, both quantitative and qualitative, confirm that our approach not only surpasses existing models in generalizability but also markedly improves rendering quality. Our work exemplifies the potential of merging neuroscientific principles with deep learning frameworks, setting a new precedent for enhanced generalizability and efficiency in novel view synthesis. A demo of our study is available at https://neugennerf.github.io.

028 029 1 INTRODUCTION

The problem of novel view synthesis in computer vision and graphics has caught the attention of 031 researchers in recent years. The ability to generate previously unseen perspectives of an object or scene is not just a technical challenge but a gateway to transformative applications in virtual reality, 033 3D modeling, and beyond. While the potential is vast, the problem is intricate: Given a set of images, along with their camera pose, the goal is to render photo-realistic images of the scene through novel viewpoints that accurately represent the actual scene. Recently, differential neural rendering methods, such as Neural Radiance Fields (NeRF) (35) and others (57; 31; 29; 27) have showcased the potential of deep learning techniques in producing high-fidelity reconstructions from input views. 037 By representing scenes as continuous volumetric scene functions, methodologies like NeRF have set new benchmarks in the field, enabling the synthesis of intricate scenes with remarkable accuracy. However, as with many cutting-edge solutions, there are nuances to consider. The bottleneck of long 040 per-scene optimization limits the practicality of these methods. Many works (55; 5; 16) propose 041 different architectures to overcome the issue, but we believe that data representation techniques can 042 be adopted to optimize the generalizability of NeRFs. Specifically, we take inspiration from the 043 mammalian brain on how it perceives different visual stimuli. Much work has been done on how 044 the mammalian visual cortex encodes visual data (20; 22). Building on the foundation of braininspired computational models and the challenges inherent in novel view synthesis, in this work, we 046 present two major contributions to the field of computer vision and novel view synthesis, described 047 as follows: 048

0.40

1. Introducing **NeuGen** (Neural Generalization), a neuro-inspired layer, that draws on the neural signal regulation mechanisms found in the mammalian visual cortex, tailored for enhancing domain generalization in NeRFs.

 052
 053
 2. Through NeuGen, we offer a novel data representation approach that captures highfrequency, domain-invariant features, significantly enhancing the view synthesis capabilities of NeRFs.



Figure 1: Block diagram of the proposed methodology: The figure shows the comprehensive overview of our domain-invariant data representation pipeline. The original images (I) are first partitioned into patches and then processed by each layer in the pipeline, as shown by the equations. The output after processing from all the blocks of the NeuGen layer is a NeuGen image (I^{G}) . The inspiration for this approach is taken from the mammalian visual cortex, which similarly interprets scenes. The next step merges NeuGen images with the original images, producing NeuGen-enhanced images (I^E) , contributing to better feature extraction for input to NeRF. The right part of the figure illustrates the volumetric rendering step, a fundamental process shared by all NeRF methods like MVSNeRF (5) and GeoNeRF (16). However, it is essential to note that the specific details and implementations of feature volume vary with each architecture, tailored to their unique enhancements and optimization strategies.

RELATED WORK 2

084 085

090

091

073

074

075

076

077

079

081

083

Neural scene representations: Representing the geometry and appearance of a 3D scene from 087 2D images has been a promising direction for researchers in recent years. Earlier methods 088 (1; 12; 15; 37; 38) demonstrated Multi-Layered Perceptrons' (MLP) ability to implicitly represent shapes, where the weight of the network maps continuous spatial coordinates to either 089 occupancy or signed distance values. Improvements in differentiable rendering methods paved the way for advancements in neural scene representations where multi-view observations were used to learn the geometric and appearance details. One method that drove more innovations 092 in this domain was NeRF (35). The NeRF method is proposed to optimize a continuous 5D neural radiance field. It combines MLPs with differentiable volume rendering and achieves 094 photo-realistic view synthesis. NeRF opened up many research frontiers in neural rendering such 095 as dynamic view synthesis (26; 39), relighting (8; 2), real-time rendering (54), pose estimation 096 (32), and editing (52; 6). However, a significant constraint with NeRF and the following methods 097 is that it needs to be optimized for each scene, which requires extensive time to produce good results.

098

099 Domain generalization in NeRFs: Researchers have focused on proposing novel architectures 100 to tackle the problem of per-scene optimization. GRF (46), pixelNeRF (55), and MINE (23) 101 tried to render novel views with a limited number of source views, but their ability to adapt to 102 unseen challenging scenes remained limited. Later, IBRNet (49) was introduced, which draws 103 inspiration from both Image-based rendering (IBR) and NeRF, along with novel ray transformers 104 that aggregate and blend information from multiple views and incorporate more long-range context 105 along the ray. More recently, MuRF (53) proposed a multi-baseline radiance field approach that handles both small and large baseline settings through a target view frustum volume representation 106 and CNN-based decoder. Apart from this, MVSNeRF (5) and GeoNeRF (16) methods have been 107 proposed that build cost volumes from 2D image features, which are eventually used to construct radiance fields. MVSNeRF, more specifically, constructs a neural encoding volume from a low-cost volume that helps achieve better generalised results. GeoNeRF constructs feature volumes at multiple resolutions to capture and use more refined features. While these methods are currently state-of-the-art, they struggle with homogenous and shiny geometry and thin structures in the scene.

112

113 Biologically-inspired techniques: The field of computer vision has been profoundly shaped 114 by insights drawn from the brain's complex mechanisms. Hierarchical models, mirroring the 115 layered processing of the visual cortex, have unlocked new perspectives in object recognition (44). 116 With its foundation in neural network architectures inspired by neural processes, deep learning 117 has catalyzed significant computational vision breakthroughs (24; 20; 48). Techniques such as 118 residual learning and deeper convolutions, though not exact replicas of neural processes are guided by the fundamental principles of neural functioning. This ongoing integration of brain-inspired 119 methodologies has been pivotal across the computational vision landscape. Notable efforts include 120 the computational modeling of the brain's central visual system (11) and investigations into the 121 recurrent connections of the visual cortex (20) among others (44; 25). These efforts have signifi-122 cantly enhanced computational tasks and deepened our comprehension of the relationship between 123 biological vision and computational methodologies (3; 22; 40; 42; 17). Despite their innovative 124 design and theoretical promise, these models often encounter challenges when benchmarked against 125 real-world datasets.

126 127 128

129

130

Our proposed approach is data-centric; instead of changing the architecture of the existing methods that use spatial features as a core part of their methodology, we present a brain-inspired featureenhancing technique that captures high-frequency domain-invariant features from input source images and ultimately improves the performance of generalized NeRF models on challenging scenes.

131 132

3 NEUGEN: NEURAL GENERALIZATION

133 134

153

154

156

157

158 159

We introduce a novel neuro-inspired layer called Neural Generalization (NeuGen), designed to cap-135 ture the neural signal regulation characteristics of certain excitatory neurons in the mammalian visual 136 cortex (56; 4; 41; 9). These neurons are distinguished by their ability to intricately encode con-137 trasting elements and structural nuances of visual stimuli, with variable firing sequences that align 138 closely with variations in visual contrast. By emulating this capability, NeuGen produces refined 139 and contextually aware representations, which are particularly beneficial for scene reconstruction 140 tasks requiring adaptability and a nuanced understanding of visual inputs. NeuGen implements 141 three key mechanisms inspired by the visual cortex's contrast processing: patch-based processing 142 analogous to neural receptive fields, local contrast normalization reflecting neural adaptation, and multi-scale feature integration similar to hierarchical visual processing. This biologically inspired 143 design enables robust feature extraction that remains consistent across varying domains and viewing 144 conditions. We integrate this process into deep learning models as a preprocessing layer using a 145 high-level mathematical formulation that mimics these biological mechanisms. 146

To mathematically implement NeuGen, we process an input image $I \in \mathbb{R}^{W \times H \times C}$, where W, H, and C represent its width, height, and channel count, respectively. Our goal is to form a domainindependent representation, denoted as I^G . The image I is segmented into patches of size s, represented as $P = \{p_{s_1}, p_{s_2}, \dots, p_{s_n}\}$, where each patch p of size s encircles a pixel k at coordinates i, j including all channels. For each patch, its mean μ_{p_s} and standard deviation σ_{p_s} are calculated to construct I^G :

1. Mean and Standard Deviation for each p_s :

$$\sigma_{p_s} = \left(\frac{1}{s^2} \sum_{i,j \in p,s} (k_{ij} - \mu_{p_s})^2 \right)^{1/2} \quad , \quad \mu_{p_s} = \frac{1}{s^2} \sum_{i,j \in p,s} k_{ij}$$

2. Domain-independent Representation, I^G:

 $z = \max\{\sigma_{p_{s1}}, \sigma_{p_{s2}}, \dots, \sigma_{p_{sn}}\}$, $I^G = \frac{\sigma_{p_s}}{z}$

By leveraging the standard deviation, σ_{p_s} , NeuGen emphasizes the contrast of local features within each patch, leading to the formation of a domain-independent representation, I^G . This representation is achieved by normalizing the standard deviations of patches across the image to highlight



Qualitative performance representation using SSIM and SIFT metrics with our technique on Realistic Synthetic, DTU and LLFF dataset

Figure 2: Feature enhancement with NeuGen across datasets. The top images (A) from left to right illustrate samples from the **Realistic Synthetic** (35), **DTU** (14), and **LLFF** (33) datasets, respec-188 tively, paired with their NeuGen versions. The NeuGen images result in notably higher SSIM scores 189 and a greater count of SIFT feature matches, signifying improved feature detection. The graphs 190 below (B,C,D) detail these improvements: each dot represents the average SSIM score between one 191 reference image and the rest within a category, while the overlaying lines trace the general pattern 192 of quality enhancement—yellow for the original and purple for NeuGen mages — demonstrating 193 NeuGen's consistent efficacy in feature emphasis across all 3 datasets. 194

187

162

196 contrast without bias towards specific domain characteristics. The domain-independent representa-197 tion I^G showcases the contrast normalization capability of NeuGen, making feature extractors in NeRFs more flexible across various domains without a strong inclination towards specific domain-199 related features. The concept of "domain-independent" highlights the NeuGen layer's capability 200 to process the input image I in a way that is not constrained to a specific domain or task, thereby 201 augmenting the model's generalizability. This property is qualitatively shown in Figure 2 (A) and

- 202
- 203 204

NERF ARCHITECTURES 4

is discussed in detail in section 6.1.

205 206

207 To demonstrate how NeuGen boosts the performance of generalized NeRFs, we choose two state-208 of-the-art methods: MVSNeRF (5) and GeoNeRF (16). Both methods follow the classical volume 209 rendering from the vanilla NeRF (34) architecture and have proven to perform well in benchmark 210 tests against other architectures, demonstrating superior generalization across different scenes; This 211 rationale firmly establishes them as prime candidates for our experimental analysis.

212 A NeRF encodes a scene as a continuous volumetric radiance field f of color and density. Specif-213 ically, for a 3D point $x \in \mathbb{R}^3$ and viewing direction unit vector $d \in \mathbb{R}^3$, f returns a differential 214 density σ and RGB color c: $f(x, d) = (\sigma, c)$. 215

The volumetric radiance field can then be rendered into a 2D image via:

219

$$\hat{C}(r) = \int_{t_n}^{t_f} T(t)\sigma(t)c(t)\,dt \tag{1}$$

where $T(t) = \exp(-\int_{t_n}^t \sigma(s)ds)$ handles occlusion. For a target view with pose P, a camera ray can be parameterized as r(t) = o + td, where o is the camera origin and t is the distance along the viewing direction d. The final color $\hat{C}(r)$ of a pixel is computed by integrating the contributions of all points along the ray r, from near t_n to far t_f bounds. This effectively accumulates the color and density information, rendering the scene as perceived from the given viewpoint.

226 **MVSNeRF** (5) divides its architecture into three essential components. First, the image features 227 are extracted using a stack of 2D CNN layers. In the second step, the cost volume generation 228 component takes the extracted feature maps as input and generates a plane-swept cost volume. This 229 cost volume contains important information about the depth as well as the appearance of the scene. 230 Although this cost volume can be used to further perform view synthesis, MVSNeRF introduces 231 another component called Neural Encoding Volume that further refines the cost volume. A 3D CNN 232 then extracts meaningful encoding from the volume itself which ultimately leads to better view 233 synthesis results. For the final step, which is neural radiance field construction, the conventional volumetric rendering from (35) is used. 234

235 GeoNeRF (16) uses MVSNeRF as a baseline and proposes a number of improvements in 236 the architecture. First and foremost, MVSNeRF uses a low-resolution 3D cost volume which 237 makes it challenging to render good quality detailed images. To counter this constraint, GeoNeRF 238 proposes a geometry reasoner; it constructs a hierarchy of cost volumes, inspired by CasMVSNet 239 (13), for each nearby source view, which is then used to infer the geometry and appearance 240 of the unseen scene. As opposed to MVSNeRF, cascade cost volumes are constructed at three 241 different resolutions which ultimately create a more dense feature volume. They also propose 242 an improvement in the neural field construction wherein the sampled points along the ray are 243 processed through an attention-based mechanism (47; 10) to infer geometry and appearance from 244 the cost volumes. This attention mechanism allows GeoNeRF to learn long-range dependencies and 245 aggregate information from multiple source views.

246 247

248

266

5 Methodology

As explained in section 4, both MVSNeRF and GeoNeRF create a feature volume by first extracting 250 2D image features and finally warping these to a plane sweep volume. To reiterate, MVSNeRF 251 creates a low-resolution feature volume while GeoNeRF creates a high-resolution cascaded volume 252 using multi-scale feature volumes. The quality of the extracted 2D features determines the quality 253 of feature volumes which ultimately influences the quality of volumetric rendering when using less 254 number of source images for generalized models.

255 While the research community has come up with many novel architectures that have achieved rea-256 sonable performance on generalization tasks, there is still room for improvement. The current state-257 of-the-art methods (5; 16) struggle against challenging geometry where finer details are not rendered 258 well. Moreover, these methods also struggle against reflective and homogeneous surfaces. These limitations can be seen in Figures 3, 4 and 5. Instead of a new architectural change, our proposed 259 approach (NeuGen) is focused more on introducing a new data representation that is capable of high-260 frequency domain-invariant features. When this data representation is merged with the original data, 261 we get a features-enhanced form of data. This is illustrated in **Figure 1**. 262

Understanding this more formally: Given two sets of N images, $\{I_1, I_2, \ldots, I_N\}$ and $\{I_1^G, I_2^G, \ldots, I_N^G\}$, each with three channels. The channel-wise merged image I_i^E for each i is obtained as follows:

$$I_i^E = I_i \oplus I_i^G \quad \text{for } i = 1, 2, \dots, N \tag{2}$$

where \oplus denotes the channel-wise addition operation, I denotes the original images, I^G denotes the NeuGen images, and I^E denotes the NeuGen-enahnced images. We use I^E as input to 2D feature extractors of MVSNeRF and GeoNeRF; the details for each integration are explained below. Application of NeuGen to the MVSNeRF architecture: A deep 2D CNN, denoted by \mathcal{F} , is utilized within the MVSNeRF architecture to process 2D image features for each unique viewpoint. These extracted features are indicative of the localized visual characteristics present in the images. The network architecture is composed of layers that downsample, thereby reformulating the NeuGen-enhanced input image I_i^E , of dimensions $H_i \times W_i \times C$, into a dense feature map F_i , which is represented as $\mathbb{R}^{H_i/4 \times W_i/4 \times C}$:

$$F_i = \mathcal{F}(I_i^E) \tag{3}$$

where H and W define the image's height and width respectively, and C signifies the generated feature channels' quantity.

Application of NeuGen to the GeoNeRF architecture: For GeoNeRF, NeuGen is applied by initially channeling images through a Feature Pyramid Network, as referred to in ((28)), to procure 2D features across three distinct levels of scale:

$$f_v^{(l)} = FPN(I_v^E) \in \mathbb{R}^{\frac{H}{2^l} \times \frac{W}{2^l} \times 2^l C} \quad \forall l \in \{0, 1, 2\}$$

$$\tag{4}$$

In this context, the FPN stands for the Feature Pyramid Network, $f_v^{(l)}$ is a 2D feature map for source view v at scale level l, I^E is the NeuGen-enhanced image, C denotes the number of channels at the initial scale, and H and W denote the image's height and width.

The primary goal of NeuGen-enhanced images is to assist with extracting invariant 2D features. MVSNeRF and GeoNeRF further use their own methodologies to construct feature volumes but with better sets of features due to NeuGen. For supervision in rendering, original images (I) are used as it is. The implementation and training details can be found in the supplementary material.

295 296

297

280

281 282

283

284

289

290

6 EXPERIMENTS & RESULTS

298 299 6.1 THE NEUGEN EFFECT

300 First, we conduct a two-part experiment to assess NeuGen's impact on image quality and feature 301 extraction. The primary quantitative analysis, shown in Figure 2 (B, C, D), involved calculating 302 and averaging Structural Similarity Index (SSIM) (50) scores across each class by comparing the first image with all others in both NeuGen I^G , and original I versions. This provided insights into 303 304 NeuGen's consistent emphasis on features. Additionally, Figure 2 (A) presents a qualitative comparison, where NeuGen images are visually contrasted with their originals, alongside their SSIM 305 scores and Scale-Invariant Feature Transform (SIFT) (30) matches. This part of the experiment 306 highlights NeuGen's effectiveness in improving image structure and feature robustness. NeuGen's 307 enhancement of SSIM scores demonstrates its proficiency in preserving the structural details of im-308 ages, with SSIM being an established metric for assessing image quality through local pixel patterns. 309 Elevated SSIM values reflect the alignment of structural content, brightness, and texture fidelity in 310 images. NeuGen's design to accentuate contrasting features ensures the structural consistency of 311 images, contributing to superior reconstructions. Moreover, the higher number of SIFT matches 312 with NeuGen usage highlights its effectiveness in extracting and defining image features. 313

This improvement indicates that NeuGen reinforces the distinctiveness and resilience of SIFT-314 generated features, thus improving image correspondence across various transformations. The in-315 troduction of NeuGen introduces a novel data representation for NeRF models, optimizing the way 316 visual information is processed and utilized. This representation's usefulness is evidenced by our 317 experiments, where NeuGen's ability to preserve structural integrity and enhance distinctive high-318 frequency features is showcased. The efficacy of this new data representation is apparent in both the 319 qualitative and quantitative improvements visible in Figure 2 and supported by our results. Such 320 representations are helpful for NeRF models, primarily when tasked with the complex challenge of 321 novel view synthesis, as they rely heavily on accurate and robust feature detection for successful execution. With NeuGen features incorporated, state-of-the-art methods for NeRF domain general-322 ization, such as MVSNeRF and GeoNeRF, can gain major improvements in the results, as shown by 323 our quantitative as well as qualitative results.

326

327

328

Table 1: Comparative performance analysis of MVSNeRF models with and without NeuGen enhancement on the Realistic Synthetic Dataset (35) ("ft" denotes per-scene optimization).

Table 2: Comparative performance analysis of GeoNeRF models with and without Neu-Gen enhancement on a subset of the DTU Dataset (14).

	Mean PSNR† SSIM† I PIPS		Mean PSNR↑ SSIM↑ LPIPS↓
MVSNeRF _{ft} (MVSNeRF + NeuGen) _{ft}	26.29 0.884 0.172 26.39 0.886 0.169	GeoNeRF [2] GeoNeRF + NeuGen	29.18 0.937 0.095 29.28 0.938 0.094

6.2 TRAINING FROM SCRATCH WITH NEUGEN

338 To demonstrate the effectiveness of NeuGen, we conduct experiments with two leading NeRF models, MVSNeRF and GeoNeRF, as shown in Table 1 and Table 2. For MVSNeRF, we train the model 339 from scratch both with and without integrating NeuGen. Our findings reveal that incorporating Neu-340 Gen from the outset significantly improves performance, especially in the per-scene optimized mod-341 els. This suggests that NeuGen enhances the model's ability to optimize scenes effectively, leading 342 to superior results. The per-scene optimized MVSNeRF model shows notable enhancements in im-343 age quality and structural consistency, indicating that NeuGen plays a crucial role in refining the 344 feature extraction and rendering processes. The improved performance in terms of PSNR, SSIM, 345 and LPIPS metrics underscores the impact of NeuGen in facilitating better scene understanding and 346 reconstruction. Similarly, for GeoNeRF, training from scratch with NeuGen for the same duration 347 demonstrates significant performance improvements across all metrics, even without the need for 348 subsequent fine-tuning. This highlights NeuGen's capability to boost performance from the outset, making the model more robust and effective in handling complex and diverse scenes. The 349 results show that NeuGen-enhanced GeoNeRF models exhibit higher fidelity in rendering and better 350 generalization across different datasets. These experiments showcase NeuGen's ability to enhance 351 generalization in NeRF models. The integration of NeuGen boosts the performance metrics and also 352 ensures that the models can handle a wider range of scenes and conditions with greater accuracy. 353 By leveraging the biologically inspired normalization techniques, NeuGen provides a significant 354 advancement in the development of NeRFs, making a strong case for its integration into advanced 355 NeRF architectures.

356 357 358

6.3 FINETUNING PRE-TRAINED MODELS WITH NEUGEN

359 To test the efficacy of NeuGen in off-the-shelf pre-trained models, we carry out experiments fine-360 tuning these models with NeuGen integration. The quantitative outcomes are delineated in Tables 361 3, 4 and 5, exhibiting improvements across the three datasets; Realistic Synthetic (35), LLFF (33) and DTU (14). Notably, integrating NeuGen with MVSNeRF and GeoNeRF invariably leads to 362 enhancements in PSNR and SSIM metrics, suggesting that NeuGen facilitates the models' rendering 363 higher fidelity reconstructions. The Realistic Synthetic dataset, comprising elements with reflective 364 and intricate textures like drums, materials, and ship, showcase marked improvements. The LLFF dataset, containing natural complexities such as fern, flower, leaves, and orchids, also benefit from 366 NeuGen's enhancements, evident from the increased metric scores, which imply an enriched capture 367 of subtle geometric nuances. Finally, for DTU, we observe that our approach eliminates most of the 368 noise in the background and renders better details regardless of the lighting condition of the scene. 369

3706.4 QUANTITATIVE RESULTS

We comprehensively evaluate our approach on three different datasets used for benchmarking across generalized NeRF methods. **Table 3** shows the results of MVSNeRF and GeoNeRF on Realistic Synthetic data. In this case, we observe a particular trend across both methods for scans such as *drums, ficus, ship,* and *materials.* These scans either contain homogenous or shiny surfaces or have delicate, detailed structures that are difficult to render for per-scene optimized NeRF models, let alone generalized models. Results for the Forward Facing dataset (LLFF) are shown in **Table 4**. These numbers further strengthen our claim that NeuGen is capable of capturing high-frequency

Table 3: Comparative performance analysis of NeuGen-enhanced models on the Realistic Synthetic dataset (35).

	Realistic Synthetic								
	Drums	Ficus	Ship	Materials	Lego	Chair	Hotdog	Mic	
	PSNR↑/SSIM↑	PSNR↑/SSIM↑	PSNR†/SSIM↑	PSNR↑/SSIM↑	PSNR↑/SSIM↑	PSNR↑/SSIM↑	PSNR↑/SSIM↑	PSNR↑/SSIM↑	
MVSNeRF (5)	21.66/0.846	25.59/0.917	25.75/0.754	24.37/0.885	25.69/0.869	26.32/0.904	32.04/0.948	28.82/0.953	
MVSNeRF + NeuGen	23.01/0.910	26.70/0.931	25.83/0.757	24.91/0.890	25.97/0.875	26.43/0.904	32.05/0.948	28.85/0.954	
GeoNeRF (16)	22.52/0.884	22.52/0.884	24.41/0.828	23.99/0.904	26.333/0.926	30.372/0.962	33.455/0.969	27.612/0.952	
GeoNeRF + NeuGen	23.98/0.915	23.82/0.900	24.43/0.830	24.50/0.909	26.445/0.928	30.508/0.963	33.723/0.971	27.847/0.956	

Table 4: Comparative performance analysis of NeuGen-enhanced models on the LLFF Dataset (33).

	LLFF							
	Fern	Flower	Leaves	Orchids	Fortress	Horns	TRex	
	PSNR↑/SSIM↑	PSNR↑/SSIM↑	PSNR↑/SSIM↑	PSNR↑/SSIM↑	PSNR↑/SSIM↑	PSNR↑/SSIM↑	PSNR↑/SSIM↑	
MVSNeRF (5)	21.67/0.642	25.06/0.793	19.897/0.689	18.91/0.581	27.33/0.807	24.38/0.791	23.66/0.826	
MVSNeRF + NeuGen	21.72/0.645	25.10/0.796	19.901/0.693	18.93/0.582	27.38/0.809	24.48/0.811	23.69/0.830	
GeoNeRF (16)	22.95/0.736	27.59/0.871	29.34/0.853	24.97/0.841	18.81/0.658	18.80/0.593	23.29/0.829	
GeoNeRF + NeuGen	23.07/0.740	27.64/0.876	29.41/0.867	24.98/0.845	18.82/0.660	18.85/0.596	23.35/0.831	

features that can better express the granularity of the objects. For instance, LLFF contains scans that contain plants and trees with complicated geometry, such as wiry structures, thin stems, and sharp petals, that can be hard to synthesize. Subsequently, **Table 5**, shows results on DTU test scans that were not part of the training. In this case, NeuGen-enhanced NeRF is able to perform better in all of the scans convincingly. We observe that in real scenes with background, like in DTU scans, our method renders background in images significantly better than the existing methods. Further details are mentioned in the qualitative results section.

6.5 QUALITATIVE RESULTS



Figure 3: Qualitative comparison on Realistic Synthetic for MVSNeRF (5) and GeoNeRF (16) along with our approach. We show results on *Ficus* and *Drums* rendered with 3 source images. Our method synthesizes homogenous regions and complex geometries better.

The visual results highlight the effectiveness of our proposed approach. NeuGen-enhanced images improve fine-grained details and render challenging textures. As shown in **Figure 3**, results from

Table 5: Comparative performance analysis of NeuGen-enhanced models on the DTU Dataset (1	e analysis of NeuGen-enhanced models on the DIU Dataset (14).
---	---

	DTU								
	Scan1	Scan8	Scan21	Scan103	Scan114	Scan110			
	PSNR↑/SSIM↑	PSNR↑/SSIM↑	PSNR↑/SSIM↑	PSNR↑/SSIM↑	PSNR↑/SSIM↑	PSNR↑/SSIM↑			
MVSNeRF (5)	18.55/0.737	22.24/0.682	18.93/0.728	23.78/0.806	27.75/0.828	26.88/0.814			
MVSNeRF + NeuGen	18.63/0.738	22.35/0.686	18.96/0.728	23.85 /0.806	27.90/0.833	26.95 /0.814			
GeoNeRF (16)	28.14/0.929	29.79/0.919	24.04/0.902	28.63/0.914	30.03/0.947	30.16/0.965			
GeoNeRF + NeuGen	28.82/0.937	29.96/0.921	24.24/0.906	29.18/0.919	30.23/0.953	30.26/0.969			



Figure 4: Qualitative comparison on LLFF Dataset for MVSNeRF (5) and GeoNeRF (16) along with our approach. While our method performs better on wiry structures in plants, it performs significantly well on reflective/transparent surfaces as well.



Figure 5: Qualitative comparison on DTU Dataset for MVSNeRF (5) and GeoNeRF (16) along with our approach. Our method performs better regardless of light conditions while also preventing cloudy artefacts or noise in the background.

the realistic synthetic dataset are convincing. The pot of the ficus is distorted in MVSNeRF and GeoNeRF, but is well synthesized with our approach. For *drums*, we show results for two differ-ent views. Our method produces finer details in MVSNeRF, and for GeoNeRF, NeuGen enhances homogeneous regions in the scene, where state-of-the-art methods struggle. GeoNeRF without Neu-Gen produces artifacts in uniform areas, highlighting NeuGen's impact. Results on LLFF (Figure 4) show our method's ability to handle thin structures like *leaves* on the fern scan. For both MVS-NeRF and GeoNeRF, NeuGen produces superior quality results. We show patches from the horns scans for both models, and NeuGen outperforms the existing methods. Finally, we qualitatively show NeuGen on three DTU test scans (Figure 5). For scan110, NeuGen avoids cloudy artifacts in low-light regions, while MVSNeRF suffers from this noise. For scan21, NeuGen captures granular details, and for *scan1*, it assures notable improvement in scenes with backgrounds. We demonstrate several instances in synthetic and real scenes where NeuGen significantly improves the handling of challenging geometric and texture properties for generalized NeRF models.

DISCUSSION

It is crucial to contextualize our results within the broader landscape of generalizable NeRF re-search. The improvements demonstrated by our NeuGen approach, while numerically modest, rep-resent a significant step forward in this challenging domain. In the field of generalizable NeRFs, incremental yet consistent improvements signify meaningful progress. Our method exhibits such consistency across multiple datasets, suggesting a robust trend rather than statistical variance. This pattern aligns with recent advancements in the field, where modest quantitative gains frequently accompany substantial qualitative enhancements. For instance, the work of (45) reports moder-ate improvements without conclusive results, noting that GNT (Generalizable NeRF Transformer)

486 does not consistently outperform other methods across all metrics and datasets. Similarly, the work 487 of (7) on efficient 3D Gaussian Splatting reported improvements of 0.11 in PSNR and 0.004 in 488 SSIM—comparable to our results—while achieving notable visual quality improvements. Here, it 489 is imperative to acknowledge the limitations of traditional metrics like PSNR and SSIM in fully 490 capturing perceptual quality, particularly for fine-grained details where our method excels. This phenomenon has been well-documented in the literature (21), highlighting that pixel-wise metrics 491 often fail to reflect improvements in texture and intricate visual features that are readily apparent to 492 human observers. 493

While the quantitative metrics provide a standardized basis for comparison, the qualitative improvements, offer crucial insights into the visual enhancements achieved by NeuGen. These improvements are particularly evident in challenging scenarios such as homogeneous surfaces, shiny materials, and intricate geometric structures—areas where generalized NeRF models traditionally struggle. Given these considerations, the true value of our contribution lies in the consistent quantitative improvements coupled with the significant qualitative enhancements across diverse datasets and challenging scenarios.

501 Finally, it is worth noting that our approach is draws inspiration from the mammalian visual cortex, 502 which excels at detecting high-frequency components in visual stimuli. NeuGen, as a neuro-inspired layer, encapsulates the response normalization property observed in a specific group of excitatory neurons in the visual cortex, known for their encoding of both structure and contrast (43; 18). This 504 biological mechanism has been implemented through a high-level mathematical formulation in Neu-505 Gen, as explained in section **3**, allowing it to enhance domain-invariant features. NeuGen-enhanced 506 images emphasize high-frequency details, improving the robustness of NeRF models across diverse 507 scenes. Similar bio-inspired approaches (19; 36) have demonstrated that brain-like representations 508 enhance model robustness, while recent works (51; 58) further emphasize the importance of high-509 frequency features for generalization in visual models.

510 511

8 CONCLUSION

512 513

514 In this study, we introduce a novel neuro-inspired layer, NeuGen (Neural Generalization), into Neu-515 ral Radiance Fields (NeRF) models, with a specific focus on enhancing domain generalization (perscene optimization). By drawing inspiration from the neural signal regulation characteristics ob-516 served in the mammalian visual cortex, NeuGen significantly improves the performance of NeRF 517 models such as MVSNeRF and GeoNeRF, particularly in challenging scans involving complex ge-518 ometries and textures. Our experiments demonstrate that NeuGen enhances the models' ability 519 to generalize across diverse scenes by emphasizing high-frequency, domain-invariant features. By 520 processing the input images to create NeuGen-enhanced images, we ensure better feature extrac-521 tion, which translates to improved structural similarity and reduced artifacts in the rendered scenes. 522 The integration of NeuGen into NeRF architectures, both during the initial training and fine-tuning 523 phases, has resulted in noticeable improvements in rendering quality and robustness, making it a 524 valuable addition to advanced NeRF architectures. The quantitative and qualitative evaluations con-525 firm that NeuGen improves generalizability and also markedly improves the rendering quality of NeRFs across different datasets, including Realistic Synthetic, LLFF, and DTU. These improve-526 ments highlight NeuGen's potential to address the limitations of current NeRF models, particularly 527 their struggle with homogeneous and reflective surfaces, as well as thin structures. 528

Despite the promising results, further testing on additional NeRF models is necessary to fully validate NeuGen's effectiveness across various architectures. Expanding the scope of NeuGen's applicability to other NeRF variants and more complex real-world datasets will help confirm its robustness
and establish its role as a versatile enhancement across different NeRF-based approaches. Future
work will involve testing NeuGen on emerging methods, such as Gaussian Splatting to assess its
impact in settings that leverage sparse views and neural point representations. Ablation study, implementation details and analysis of LPIPS results can be found in the Appendix.

536

537 REFERENCES

[1] Matan Atzmon, Niv Haim, Lior Yariv, Ofer Israelov, Haggai Maron, and Yaron Lipman. Controlling neural level sets. *Advances in Neural Information Processing Systems*, 32, 2019.

545

546 547

548 549

550

551

552

553

554

555

556

558

559

560

561

562

563

564

565

566

567

568 569

570

571

572

573

574 575

576

577 578

579

580

581

582

583

584

585 586

587

588 589

590

591

592

- Sai Bi, Zexiang Xu, Pratul Srinivasan, Ben Mildenhall, Kalyan Sunkavalli, Miloš Hašan, Yannick Hold-Geoffroy, David Kriegman, and Ravi Ramamoorthi. Neural reflectance fields for appearance acquisition. *arXiv preprint arXiv:2008.03824*, 2020.
 - [3] Peter R Brotchie, Richard A Andersen, Lawrence H Snyder, and Steven J Goodman. Head position signals used by parietal neurons to encode locations of visual stimuli. *Nature*, 375: 232–235, 1995.
 - [4] Max F. Burg et al. Learning divisive normalization in primary visual cortex. PLOS Computational Biology, 17(6):e1009028, 2021.
 - [5] Anpei Chen, Zexiang Xu, Fuqiang Zhao, Xiaoshuai Zhang, Fanbo Xiang, Jingyi Yu, and Hao Su. Mvsnerf: Fast generalizable radiance field reconstruction from multi-view stereo. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 14124–14133, 2021.
 - [6] Anpei Chen, Ruiyang Liu, Ling Xie, Zhang Chen, Hao Su, and Jingyi Yu. Sofgan: A portrait image generator with dynamic styling. ACM Transactions on Graphics (TOG), 41(1):1–26, 2022.
 - [7] Yuedong Chen, Haofei Xu, Chuanxia Zheng, Bohan Zhuang, Marc Pollefeys, Andreas Geiger, Tat-Jen Cham, and Jianfei Cai. Mvsplat: Efficient 3d gaussian splatting from sparse multi-view images. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2024. To appear.
 - [8] Zhang Chen, Anpei Chen, Guli Zhang, Chengyuan Wang, Yu Ji, Kiriakos N Kutulakos, and Jingyi Yu. A neural rendering framework for free-viewpoint relighting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5599–5610, 2020.
 - [9] Aritra Das and Supratim Ray. Effect of cross-orientation normalization on different neural measures in macaque primary visual cortex. *Cerebral Cortex Communications*, 2(1):tgab009, 2021. doi: 10.1093/texcom/tgab009. URL https://dx.doi.org/10.1093/texcom/tgab009.
 - [10] Alexey Dosovitskiy et al. An image is worth 16x16 words: Transformers for image recognition at scale. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2021. URL https://arxiv.org/abs/2010.11929.
 - [11] Ruimin Gao, Hao Zou, and Zhekai Duan. Cvsnet: A computer implementation for central visual system of the brain, 2023.
 - [12] Kyle Genova, Forrester Cole, Avneesh Sud, Aaron Sarna, and Thomas Funkhouser. Local deep implicit functions for 3d shape. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4857–4866, 2020.
 - [13] Xiaodong Gu, Zhiwen Fan, Siyu Zhu, Zuozhuo Dai, Feitong Tan, and Ping Tan. Cascade cost volume for high-resolution multi-view stereo and stereo matching. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 2495–2504, 2020.
 - [14] Rasmus Jensen, Anders Dahl, George Vogiatzis, Engil Tola, and Henrik Aanæs. Large scale multi-view stereopsis evaluation. In *Proceedings of the IEEE Conference on Computer Vision* and Pattern Recognition (CVPR), 2014. doi: 10.1109/cvpr.2014.59. URL https://oa. mg/work/10.1109/cvpr.2014.59?utm_source=chatgpt.
 - [15] Chiyu Jiang, Avneesh Sud, Ameesh Makadia, Jingwei Huang, Matthias Nießner, Thomas Funkhouser, et al. Local implicit grid representations for 3d scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6001–6010, 2020.
 - [16] Mohammad Mahdi Johari, Yann Lepoittevin, and François Fleuret. Geonerf: Generalizing nerf with geometry priors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 18365–18375, 2022.
 - [17] J. P. Jones and L. A. Palmer. An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, 58(6):1233–1258, 1987.

598

600

601 602

603

604

605 606

607

608

609

610 611

612

613

614

615 616

617

618

619 620

621

622

623 624

625

626

627 628

629

630

631

632

633 634

635

636 637

638

639

640 641

642

643

- [18] Sarah Jones and Michael Lee. Normalization in mouse primary visual cortex. *PLOS ONE*, 2023.
 - [19] Alex Krizhevsky and Geoffrey Hinton. Learning from brains how to regularize machines. In *NeurIPS*, 2019.
 - [20] J Kubilius, M Schrimpf, A Nayebi, D Bear, DLK Yamins, and JJ DiCarlo. Cornet: Modeling the neural mechanisms of core object recognition. *bioRxiv*, 09 2018. preprint.
 - [21] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4681–4690, 2017.
 - [22] Brian Lee, Bijan Pesaran, and Richard A Andersen. Area mstd neurons encode visual stimuli in eye coordinates during fixation and pursuit. *Journal of Neurophysiology*, 105(2):709–725, 2011.
 - [23] Jiaxin Li, Zijian Feng, Qi She, Henghui Ding, Changhu Wang, and Gim Hee Lee. Mine: Towards continuous depth mpi with nerf for novel view synthesis. In *ICCV*, 2021.
 - [24] Pengyu Li. Bionet: A biologically-inspired network for face recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 10344– 10354, June 2023.
 - [25] Zhe Li et al. Learning from brains how to regularize machines. In Advances in Neural Information Processing Systems 32 (NeurIPS 2019), 2019. URL https://papers.nips.cc/paper_files/paper/2019/hash/70117ee3c0b15a2950f1e82a215e812b-Abstract.html.
 - [26] Zhengqi Li, Simon Niklaus, Noah Snavely, and Oliver Wang. Neural scene flow fields for space-time view synthesis of dynamic scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6498–6508, 2021.
 - [27] Chen-Hsuan Lin, Wei-Chiu Ma, Antonio Torralba, and Simon Lucey. Barf: Bundleadjusting neural radiance fields. In *IEEE International Conference on Computer Vision* (*ICCV*), 2021. URL https://oa.mg/work/10.1109/iccv48922.2021.00569? utm_source=chatgpt.
 - [28] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In *Proceedings of the IEEE conference* on computer vision and pattern recognition, pp. 2117–2125, 2017.
 - [29] Lingjie Liu, Jiatao Gu, Kyaw Zaw Lin, Tat-Seng Chua, and Christian Theobalt. Neural sparse voxel fields. Advances in Neural Information Processing Systems, 33:15651–15663, 2020.
 - [30] David G. Lowe. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, 60(2):91–110, January 2004.
 - [31] Ricardo Martin-Brualla, Noha Radwan, Mehdi SM Sajjadi, Jonathan T Barron, Alexey Dosovitskiy, and Daniel Duckworth. Nerf in the wild: Neural radiance fields for unconstrained photo collections. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7210–7219, 2021.
 - [32] Quan Meng, Anpei Chen, Haimin Luo, Minye Wu, Hao Su, Lan Xu, Xuming He, and Jingyi Yu. GNeRF: GAN-based Neural Radiance Field without Posed Camera. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021.
- [33] Ben Mildenhall, Pratul P Srinivasan, Rodrigo Ortiz-Cayon, Nima Khademi Kalantari, Ravi
 Ramamoorthi, Ren Ng, and Abhishek Kar. Local light field fusion: Practical view synthesis
 with prescriptive sampling guidelines. ACM Transactions on Graphics (TOG), 38(4):1–14, 2019.

653

654 655

656

657

658

659

660

661 662

663

664 665

666

667

668

669

670

671 672

673

674

675

676

677

678

679

680 681

682 683

684 685

686

687

688

689

690

691 692

693

694

695

696 697

698

- [34] 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.
 - [35] 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. *Commu*nications of the ACM, 65(1):99–106, 2021.
 - [36] An Nguyen and Hui Zhao. Simulating a primary visual cortex at the front of cnns improves robustness to image perturbations. In *NeurIPS*, 2020.
 - [37] Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. Deepsdf: Learning continuous signed distance functions for shape representation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 165–174, 2019.
 - [38] Songyou Peng, Michael Niemeyer, Lars Mescheder, Marc Pollefeys, and Andreas Geiger. Convolutional occupancy networks. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16*, pp. 523–540. Springer, 2020.
 - [39] Albert Pumarola, Enric Corona, Gerard Pons-Moll, and Francesc Moreno-Noguer. D-nerf: Neural radiance fields for dynamic scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10318–10327, 2021.
 - [40] Masamichi Sakagami and Hiroshi Niki. Encoding of behavioral significance of visual stimuli by primate prefrontal neurons: relation to relevant task conditions. *Experimental Brain Research*, 97(3):423–436, 1994.
 - [41] T. Sawada and A. Petrov. Emulating and predicting physiological results of neurons in the primary visual cortex (v1) based on the divisive normalization model. *Journal of Vision*, 16 (12):958, 2016. doi: 10.1167/16.12.958. URL https://dx.doi.org/10.1167/16.12.958.
 - [42] Thomas Serre. Hierarchical models of the visual system. Cognitive Linguistic & Psychological Sciences Department, Brain Institute for Brain Sciences, Brown University, 2014. URL https://serre-lab.clps.brown.edu/wp-content/uploads/ 2014/10/Serre-encyclopedia_revised.pdf.
 - [43] John Smith and Jane Doe. Effect of cross-orientation normalization on different neural measures in macaque primary visual cortex. *Cereb Cortex Commun*, 2021.
 - [44] Nicola Strisciuglio. Brain-inspired algorithms for processing of visual data, 2021.
 - [45] Mukund Varma T, Peihao Wang, Xuxi Chen, Tianlong Chen, Subhashini Venugopalan, and Zhangyang Wang. Is attention all that neRF needs? In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum? id=xE-LtsE-xx.
 - [46] Alex Trevithick and Bo Yang. Grf: Learning a general radiance field for 3d scene representation and rendering. In arXiv:2010.04595, 2020.
 - [47] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17)*, pp. 6000– 6010. Curran Associates Inc., 2017. URL https://arxiv.org/pdf/1706.03762. pdf.
 - [48] Vijay Veerabadran, Ritik Raina, and V. D. Sa. Bio-inspired learnable divisive normalization for anns. No additional publication details available.
- [49] Qianqian Wang, Zhicheng Wang, Kyle Genova, Pratul Srinivasan, Howard Zhou, Jonathan T.
 Barron, Ricardo Martin-Brualla, Noah Snavely, and Thomas Funkhouser. Ibrnet: Learning multi-view image-based rendering. In *CVPR*, 2021.

702	[50] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment:
703	from error visibility to structural similarity. IEEE Transactions on Image Processing, 13
704	(4):600-612, April 2004. ISSN 1057-7149. doi: 10.1109/TIP.2003.819861. URL https:
705	//ieeexplore.ieee.org/document/1284395.
706	

- [51] Yicheng Wu and Kaiming He. Improving vision transformers by revisiting high-frequency components. In *ECCV*, 2022.
- [52] Fanbo Xiang, Zexiang Xu, Miloš Hašan, Yannick Hold-Geoffroy, Kalyan Sunkavalli, and Hao
 Su. NeuTex: Neural Texture Mapping for Volumetric Neural Rendering. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.
- [53] Haofei Xu, Anpei Chen, Yuedong Chen, Christos Sakaridis, Yulun Zhang, Marc Pollefeys, Andreas Geiger, and Fisher Yu. Murf: Multi-baseline radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 20041–20050, 2024.
 - [54] Alex Yu, Ruilong Li, Matthew Tancik, Hao Li, Ren Ng, and Angjoo Kanazawa. Plenoctrees for real-time rendering of neural radiance fields, 2021.
 - [55] Alex Yu, Vickie Ye, Matthew Tancik, and Angjoo Kanazawa. pixelNeRF: Neural radiance fields from one or few images. In *CVPR*, 2021.
 - [56] Zaina A. Zayyad, John H. R. Maunsell, and Jason N. MacLean. Normalization in mouse primary visual cortex. *bioRxiv*, 2023. The complete version history of this preprint is available at bioRxiv.
 - [57] Kai Zhang, Gernot Riegler, Noah Snavely, and Vladlen Koltun. Nerf++: Analyzing and improving neural radiance fields. *arXiv preprint arXiv:2010.07492*, 2020.
 - [58] Xinlei Zhao and Yiming Sun. Revisiting visual model robustness: A frequency long-tailed distribution view. In *NeurIPS*, 2023.

756 A APPENDIX

B ABLATION STUDY

760 **Finding optimal value for NeuGen integration**: In our study, we conducted a comprehensive 761 sweep to evaluate the impact of varying the weight of NeuGen on the performance of Neural Radi-762 ance Fields (NeRF) models. This sweep was crucial in determining the optimal setting for NeuGen, a key component in our approach to enhance the domain generalization capabilities of NeRF mod-764 els. The methodology involved systematically adjusting the weight of NeuGen across a range of 765 values from 0.5 to 3.0, observing the corresponding changes in the models' performance metrics. 766 This process explored how different weightings of NeuGen influence the model's ability to process 767 and render complex visual scenes. The primary focus was on understanding the balance between preserving the original image characteristics and incorporating the enhanced features provided by 768 NeuGen. Some top-performing results from this extensive sweep can be seen in **Supplementary** 769 Table 1. During this extensive sweep, we discovered that a NeuGen weight of **0.5** yielded the most 770 promising results. This specific weighting provided an optimal balance, enhancing the model's per-771 formance, particularly in rendering scenes with intricate details and challenging textures. It allowed 772 the models to leverage NeuGen's strengths effectively without overwhelming the original image 773 data. Equation 2 of the main manuscript now becomes a representation of the optimal setting dis-774 covered in our NeuGen sweep. In this equation, $I_{i(e)}$ denotes the enhanced image, while I_i and 775 $I_{i(n)}$ represent the original image and the NeuGen-processed image, respectively. This formulation 776 demonstrates the combination of the original image with the NeuGen-enhanced image, where the 777 weight of NeuGen is set to 0.5:

778

758

759

779

$$I_i^E = I_i \oplus (I_i^G * 0.5) \quad \text{for } i = 1, 2, \dots, N$$
 (5)

This weight effectively balances the original and enhanced features, ensuring that the final image retains essential characteristics while benefiting from the improved feature representation offered by NeuGen. The success of this setting underscores the potential of NeuGen in enhancing the capabilities of NeRF models, particularly in challenging scenarios that require a nuanced balance of feature enhancement and preservation. This process can be seen in Supplementary Figure 1.

786 787

Supplementary Table 6: Comparison of Mean PSNR values for MVSNeRF with and without NeuGen integration at various weights on the Realistic Synthetic Dataset (35). The table highlights the impact of different NeuGen weight settings on the performance of MVSNeRF, with a focus on the optimal setting of **0.5**, which yields the highest Mean PSNR.

792	Model	Mean PSNR
793	MVSNeRF [1]	23.60
794	MVSNeRF + NeuGen (0.5)	22.96
795	MVSNeRF + NeuGen (1.2)	22.91
796	MVSNeRF + NeuGen (2.9)	22.87
797	MVSNeRF + NeuGen (0.74)	22.88
709	MVSNeRF + NeuGen (0.72)	22.91
790	MVSNeRF + NeuGen (0.55)	22.92
799	MVSNeRF + NeuGen (1.5)	22.86
800		

801 While insightful, the findings from our NeuGen sweep reveal an intriguing aspect of the enhanced 802 models' performance. After base training, the NeuGen-enhanced models did not outperform 803 the original MVSNeRF model [1]. However, this initial observation does not fully capture the 804 efficacy of NeuGen. As detailed in the main manuscript, a significant shift in performance was 805 observed following per-scene optimization. This optimization phase allowed the NeuGen-enhanced 806 models to demonstrate their improved capabilities truly. The enhanced performance of our models, 807 post-optimization, is comprehensively documented across various datasets and scenarios in Tables 1, 2, 3 (main manuscript), and Supplementary Tables 3, 4, 5 and 6. These tables present a clear 808 comparative analysis, showcasing the instances where the NeuGen-enhanced models' optimal NeuGen weight of 0.5 surpassed the original MVSNeRF model in rendering quality and accuracy.



Supplementary Figure 6: Visualization of combining original images (I) - left column, with NeuGen images $(I_{(n)})$ - middle column, to get NeuGen enhanced images $(I_{(e)})$ - right column. In the figure, \oplus denotes the channel-wise addition operation

This outcome highlights the potential of NeuGen in refining and elevating the performance of NeRF models, especially when they are fine-tuned to specific scenes, paving the way for more

sophisticated and accurate visual rendering in computational vision tasks.

Integerating NeuGen with features fusion: To find out the best way to utilize the features-rich NeuGen images, we experimented with a features fusion method. As mentioned in the main manuscript, MVSNeRF first extracts the 2D features from images before creating a 3D cost volume. We tried to extract the features from the original and NeuGen images separately before "fusing" them into a single feature set, which we believed could be more efficient.

More specifically, MVSNeRF uses a feature pyramid network (FPN) to extract hierarchical features 872 from images. With our approach, initial convolutional layers conv0, conv1, and conv2 extract fea-873 tures from both sets of images (original and NeuGen) independently in a bottom-up pathway. Each 874 layer doubles the number of channels and downsamples the spatial dimensions by 2x; this results 875 in 32 feature maps. These features from each image set are then concatenated along the channel 876 dimension, resulting in the features from both streams into a unified set of features with 64 chan-877 nels. The 64-channel feature map then goes through a set of "fusion" convolutional layers Fusion1 878 and Fusion2. Fusion1 takes as input the concatenated features, which is a tensor of shape (B, 64, 879 H/4, W/4), where B is the batch size, H is the height and W is the width. It then applies a convolution with kernel size 3x3 and padding size 1. So it will convolve over the spatial dimensions H/4 880 and W/4. Notably, Fusion1 outputs 32 feature maps. This fuses the information from both streams and projects it down to a 32-dimensional representation. Fusion2 takes this 32-channel fused repre-882 sentation from Fusion1 as input and applies another 3x3 convolution, further integrating the fused 883 features into a shared embedding. 884

Ultimately, we wanted to design an architecture that takes features from both sets of images and
utilizes the best features from both sets in a learnable manner. We trained MVSNeRF from scratch
with the features fusion incorporated. We observed that fusing features did not produce better results
as shown in Supplementary Table 2.

889 890 891

892

899

900 901 Supplementary Table 7: Quantitative comparison of features fusion method with original MVS-NeRF.

Mean	
Model	PSNR↑
MVSNeRF	23.571
MVSNeRF + NeuGen (Features Fusion)	23.461

C DATASETS

The DTU, Realistic Synthetic, and LLFF datasets provide a comprehensive suite for assessing the 902 performance of view synthesis methodologies like NeRF. The DTU dataset features 80 scenes 903 from fixed camera positions with high-resolution structured light scans, ideal for testing multi-view 904 stereopsis [4]. The Realistic Synthetic dataset, optimized for continuous volumetric scene function 905 representation, contains scenes such as "lego", "drums", "ficus", "mic", "hotdog", "materials", 906 "chair", and "ship", which are instrumental for NeRF's view synthesis [3]. Lastly, the LLFF dataset, 907 with its real-world forward-facing images, challenges algorithms to handle varying lighting and 908 complex dynamics, offering scenes like "leaves", "flowers", "ferns", "orchids", "room", "trex", "horns", and "fortress" for rigorous real-world testing [5]. Together, these datasets benchmark the 909 accuracy of algorithms like NeRF and their robustness in diverse and complex environments. 910

911

DTU dataset: The DTU dataset, introduced by Jensen et al., serves as a comprehensive benchmark
for multi-view stereopsis methodologies [4]. Comprising 80 diverse scenes, the dataset offers
various environments captured from 49 or 64 precise camera positions. Each scene is further
enriched with reference structured light scans, meticulously acquired using a 6-axis industrial robot.
This dataset's granularity and diversity make it an invaluable resource for evaluating the robustness
and accuracy of algorithms like NeRF. Given NeRF's emphasis on synthesizing novel views from
sparse inputs, the DTU dataset, with its varied camera positions and detailed scans, provides a

⁹¹⁸ rigorous testing ground to assess NeRF's performance in real-world scenarios.

919 920 921

922

923

924

925

926

927

928

Realistic Synthetic dataset: The Realistic Synthetic dataset is intrinsically tied to the NeRF methodology, designed to represent scenes as continuous volumetric scene functions [3]. This representation is optimized using a limited set of input views, emphasizing NeRF's capability to interpolate and extrapolate scene information. The dataset contains scenes that, when processed through NeRF, can be reconstructed with remarkable accuracy. This high-fidelity reconstruction is pivotal for synthesizing novel views, even in complex scenes. The dataset contains a total of 106 images spanning various scenes. These scenes include "lego", "drums", "ficus", "mic", "hotdog", "materials", "chair", and "ship".

929 Real Forward Facing (LLFF) dataset: The LLFF dataset, introduced by Mildenhall et al., is a 930 curated collection of real-world scenes designed for the task of view synthesis [5]. It consists of 931 forward-facing images captured in natural settings. The dataset is particularly suited for evaluating 932 methods that require handling diverse lighting conditions, occlusions, and scene dynamics. Each 933 scene in the LLFF dataset is accompanied by high-quality images and precise camera pose information. The dataset's focus on real-world, forward-facing scenes makes it an essential resource 934 935 for developing and benchmarking view synthesis methods that aspire to operate outside of controlled laboratory conditions and in the wild. The dataset consists of the scenes: "leaves", "flowers", 936 "ferns", "orchids", "room", "trex", "horns", and "fortress". 937

938 939

940

D EXTENDED RESULTS ANALYSIS - LPIPS

Expanding on section 7 of the main manuscript, in the supplementary material, we provide an analysis of the Learned Perceptual Image Patch Similarity (LPIPS) results for MVSNeRF and GeoNeRF
models to complement the primary PSNR and SSIM metrics. LPIPS is a perceptually-motivated
metric that quantifies image similarity in terms of visual perception, offering a valuable perspective
for assessing nuanced image quality differences.

946 The Supplementary Tables 3, 4, and 5, expand on the PSNR and SSIM results from Tables 1, 2, 947 and **3** in the main manuscript, providing an in-depth analysis using the Learned Perceptual Image 948 Patch Similarity (LPIPS) metric. These tables elucidate the impact of the NeuGen enhancement 949 on the perceptual image quality of MVSNeRF and GeoNeRF across multiple datasets. The LPIPS 950 scores, which reflect a perceptually higher image quality, generally show an improvement with the 951 integration of NeuGen, as evidenced by the consistently lower scores for MVSNeRF with NeuGen 952 in Supplementary Table 3 for the Realistic Synthetic dataset [3]. Similarly, Supplementary Table 4 demonstrates that GeoNeRF with NeuGen improves on the LLFF dataset [5]. However, the 953 enhancements in the DTU dataset [4] are modest (Supplementary Table 5). These supplemen-954 tary insights underscore the general efficacy of NeuGen in refining the visual perception of images, 955 crucial for applications where detail quality is of utmost importance. 956

The decision to reserve the LPIPS analysis for the supplementary material was guided by the desire
to maintain a clear and straightforward narrative in the main manuscript. Since PSNR and SSIM are
conventionally higher-is-better metrics, their inclusion in the main results allows for a more direct
comparison with previous works and a more immediate comprehension of the model's performance.
In contrast, due to its inverted scale, the LPIPS metric's lower-is-better nature could complicate the
direct comparison narrative. Including it in the supplementary material allows a comprehensive view
of the model's performance without over-complicating the main discussion.

964 965

966

E IMPLEMENTATION DETAILS

Networks' details. In our approach, we fine-tuned MVSNeRF for 10k iterations. On the other hand, we fine-tuned GeoNeRF for 1k iterations, as its authors have reported that their model, when fine-tuned for just 1k iterations, can achieve up to 98.15% of the performance level attained after 10k iterations. Although we were able to reproduce the results for the generalized model using the released code for the MVSNeRF method, we were not able to reproduce the per-scene optimization results mentioned in Table 1 of the paper. It is worth mentioning that the authors of GeoNeRF

Supplementary Table 8: Performance analysis of MVSNeRF and GeoNeRF models on the Realistic
Synthetic Dataset [3] using the LPIPS metric. The models are evaluated with and without NeuGen
enhancement. Lower LPIPS scores indicate better perceptual image quality and bold indicates
better performance.

	Realistic Synthetic							
	Drums	Ficus	Ship	Materials	Lego	Chair	Hotdog	Mic
	LPIPS↓	LPIPS↓	LPIPS↓	LPIPS↓	LPIPS↓	LPIPS↓	LPIPS↓	LPIPS↓
MVSNeRF [1]	0.215	0.164	0.302	0.167	0.183	0.139	0.096	0.120
MVSNeRF + NeuGen	0.207	0.157	0.297	0.165	0.173	0.137	0.094	0.117
GeoNeRF [2]	0.147	0.147	0.205	0.133	0.100	0.0533	0.057	0.070
GeoNeRF + NeuGen	0.145	0.144	0.204	0.125	0.099	0.0532	0.057	0.068

Supplementary Table 9: Performance analysis of MVSNeRF and GeoNeRF models on the
 LLFF Dataset [5] using the LPIPS metric. The models are evaluated with and without NeuGen
 enhancement. Lower LPIPS scores indicate better perceptual image quality and bold indicates
 better performance.

				LLFF			
	Fern	Flower	Leaves	Orchids	Fortress	Horns	TRex
	LPIPS↓	LPIPS↓	LPIPS↓	LPIPS↓	LPIPS↓	LPIPS↓	LPIPS↓
MVSNeRF [1]	0.310	0.208	0.283	0.331	0.202	0.256	0.219
MVSNeRF + NeuGen	0.309	0.206	0.283	0.331	0.201	0.254	0.218
GeoNeRF [2]	0.256	0.136	0.260	0.317	0.149	0.210	0.262
GeoNeRF + NeuGen	0.252	0.134	0.255	0.312	0.147	0.204	0.260

Supplementary Table 10: Performance analysis of NeuGen-enhanced models on the DTU Dataset
[6] using the LPIPS metric. The models are evaluated with and without NeuGen enhancement.
Lower LPIPS scores indicate better perceptual image quality and bold indicates better performance.

	DTU						
	Scan1	Scan8	Scan21	Scan103	Scan114	Scan110	
	LPIPS↓	LPIPS↓	LPIPS↓	LPIPS↓	LPIPS↓	LPIPS↓	
MVSNeRF [1]	0.321	0.402	0.304	0.318	0.285	0.341	
MVSNeRF + NeuGen	0.318	0.400	0.304	0.318	0.282	0.339	
GeoNeRF [2]	0.089	0.113	0.108	0.149	0.080	0.086	
GeoNeRF + NeuGen	0.088	0.111	0.105	0.136	0.072	0.070	

1026 also directly quoted the result from the MVSNeRF paper. Furthermore, we used three source im-1027 ages for inference for both methods to ensure a fair comparison. We maintained consistent training 1028 parameters and settings across both the original image fine-tuning and the NeuGen-enhanced im-1029 age fine-tuning processes. This consistency was critical in ensuring a fair and accurate comparison 1030 between the two approaches, allowing us to confidently attribute any observed differences in performance to the influence of the NeuGen layer rather than to variations in the training and fine-tuning 1031 regimen. For results and insights on the training of NeRF models from scratch on NeuGen-enhanced 1032 images, the details of the datasets used, and results on other metrics please refer to the supplementary 1033 materials. 1034

We trained MVSNeRF and GeoNeRF from scratch using the NeuGen-enhanced images. For MVS-NeRF [1] we trained it using NeuGen-enhanced images for 6 epochs using 3 source views. While for GeoNeRF [2], we trained the model from scratch for 25k steps using 6 source views. Further, for fine-tuning, we follow the same settings as mentioned in our main manuscript. For MVSNeRF, we used 3 source views, while for GeoNeRF, we used 9 source views. Evaluation and rendering are performed using 3 source views across both methods.

For all our experiments, we use 1 x Tesla V100 GPU. The training, testing, and evaluation time for all our experiments are the same as MVSNeRF and GeoNeRF. Since our method uses a new data representation and no architectural change, our approach does not add to the time overhead.

- 1044
- 1045
- 1046

F NEUGEN'S RESULTS ON OTHER DOWNSTREAM TASKS

1047 **2D medical image segmentation:** To explore the effect of NeuGen for segmentation tasks, we 1048 selected a human brain imaging dataset from ISEG-2017 challenge dataset (1). The focus of this 1049 challenge was to segment 6-month infant brain tissues from T1 and T2-weighted MRI imaging data. 1050 We utilize 2D U-Net (2), a common model used for segmentation. Training settings were kept the same across all the domain shift experiments for both with and without NeuGen. For comparison, 1051 we select the neuroimaging dataset of T1 (for training) and T2 (for testing) modalities and observe 1052 the performance of 2D U-Net on segmenting three brain regions (White Matter (WM), Grey Matter 1053 (GM), and Cerebrospinal Fluid (CSF)) in the testing sets, with and without adding NeuGen-layer 1054 in the existing architectures. We observe a performance boost (mean score of all three regions) 1055 of 40.3% increase in DICE score, 60.4% increase in IoU, 23.8% decrease in Hausdorff distance, 1056 and **48.05**% decrease in MSD (mean surface distance) by adding NeuGen-layer in the 2D U-Net as 1057 seen in **Supplementary Table 11**. These remarkable improvements in segmentation accuracy, as 1058 quantified by the substantial enhancements in widely accepted metrics, highlight the impact of the 1059 NeuGen layer when integrated into the 2D U-Net architecture. The substantial gain in DICE score and IoU, along with the significant reduction in Hausdorff distance and MSD, not only demonstrates 1061 the enhanced precision and reliability of the model but also the potential of NeuGen to serve as a pivotal addition to standard segmentation practices. These findings are promising for the field of 1062 medical imaging, where such advancements can lead to more accurate diagnostic tools, thereby 1063 facilitating improved patient outcomes. 1064

1065

Supplementary Table 11: Performance comparison of 2D U-Net on T1 and T2 modalities with and
without adding NeuGen-layer in the architecture. Models are trained on one modality and tested
on the other modality and vice-versa. We selected a human brain imaging dataset from ISEG-2017
challenge. The focus of this challenge was to segment 6-month-old infant brain tissues from T1 and
T2-weighted MRI imaging data, acquired at UNC-Chapel Hill.

1071		DICE score	Hausdorff distance	IoU score	MSD score
1072	Train-T1 (plain) \rightarrow Test-T2	52.8	20.1	37.2	3.08
1072	Train-T1 (NeuGen) \rightarrow Test-T2	74.1	15.31	59.7	1.60
1073	Train-T2 (plain) \rightarrow Test on T1	59.9	18.3	44.2	2.56
1074	Train-T2 (NeuGen) \rightarrow Test on T1	70.1	17.9	55.4	1.93
1075		I	1	I	I

3D brain registration: We also explored NeuGen's effect in a 3D brain registration task. We
utilized the OASIS-3 dataset (3). This dataset contains MR sessions from normal participants
along with individuals undergoing cognitive decline. The MRI imaging data spans T1-weighted,
T2-weighted, FLAIR, ASL, SWI, time of flight, resting-state BOLD, and DTI domains. For our
experimentation, we shortlist T1-weighted and T2-weighted MRI scans and remained agnostic to

1080 the participants' cognitive health. We use it to assess the performance of our NeuGen-enhanced model optimized for image registration. In this context, Supplementary Table 12 demonstrates the 1082 proficiency of NeuGen in refining the registration process. Structural Similarity Index (SSIM) and Normalized Cross Correlation (NCC) were employed as loss functions during the training phase 1084 to guide the model towards higher fidelity transformations. The FourierNet (4) initially registers SSIM scores of 0.1857 for T1 \rightarrow T2 transformations and 0.1309 for T2 \rightarrow T1 transformations. Upon using the NeuGen integrated model, we observe a notable enhancement in quality, with 1086 SSIM scores increasing to 0.6231 (T1 \rightarrow T2) and 0.6568 (T2 \rightarrow T1) for our model with the 1087 SSIM loss function, and our model with the NCC loss function's scores also showing substantial 1088 improvements with SSIM of 0.6434 (T1 \rightarrow T2) and 0.6214 (T2 \rightarrow T1). These improved metrics 1089 validate the effectiveness of NeuGen in boosting the structural integrity and similarity of the 1090 registered images-attributes that are vital in medical image analysis for accurate diagnosis and 1091 treatment planning. 1092

1093

Supplementary Table 12: Quantitative results (SSIM score) on OASIS-3 dataset. This dataset contains MR sessions from normal participants along with individuals undergoing cognitive decline. The MRI imaging data spans T1-weighted, T2-weighted, FLAIR, ASL, SWI, time of flight, resting-state BOLD, and DTI domains. For our experimentation, we shortlist T1-weighted and T2-weighted MRI scans and remained agnostic to the participants' cognitive health. We use it to assess the performance of our NeuGen-enhanced model optimized for image registration.

Model	$T1 \rightarrow T2$	$T2 \rightarrow T1$
Fourier Net [4]	0.1857	0.1309
NeuGen + SSIM	0.6231	0.6568
NeuGen + NCC	0.6434	0.6214

1103 1104

1100 1101 1102

1105

1106Supplementary Table 13: The table compares pretrained DNNs' (including CNN and ViT based1107NAS architectures) accuracies across a range of domain transfer tasks (with and without NeuGen).1108Rows list the models, and columns specify the domain transfer tasks from the source to the target1109(source \rightarrow target) domains. The domains involved are MNIST (M), SVHN (S), USPS (U), and1110MNIST-M (MM). Entries in bold indicate an enhancement in performance due to the NeuGen adap-1111tation, and underlined values signify models that have achieved competitive benchmarking scores.

	Ň	ŝ	MAR	14	ŝ	MM	4	Ś	MM	M. M.	NN N	N 8
Models	4.	4	4.	5	ئ `	\$ `	\$	ર્	Ş	LV12	412	412
VGG19+NeuGen	74.2	24.5	62.6	48.4	20.4	30.5	51.3	28.4	38.4	<u>96.2</u>	74.3	33.6
VGG19	70.9	13.7	39.7	66.2	13.9	19.9	47.2	32.5	33.9	94.3	76.7	22.5
DenseNet121+NeuGen	26.4	15.9	50.8	43.1	14.8	21.0	49.1	20.0	32.6	85.2	54.0	26.3
DenseNet121	74.3	11.7	38.4	39.5	11.4	20.9	34.1	22.1	31.5	96.6	73.3	21.3
ShuffleNet+NeuGen	7.0	56.1	71.1	22.6	36.2	27.2	32.1	27.5	28.9	<u>84.8</u>	7.6	50.4
ShuffleNet	9.3	19.0	14.1	44.9	12.2	13.0	26.1	31.1	19.7	38.7	3.2	18.3
Xception+NeuGen	61.9	20.9	59.7	41.3	15.9	25.8	48.8	16.4	37.7	95.7	31.8	33.1
Xception	62.4	19.4	56.0	20.9	11.9	21.1	42.7	14.2	34.2	97.9	77.8	25.6
NASNetMobile+NeuGen	45.4	21.1	47.5	25.5	10.4	16.9	47.6	16.5	35.1	<u>95.6</u>	58.3	31.1
NASNetMobile	33.0	12.7	36.0	24.8	10.1	21.4	46.0	23.8	32.6	90.0	63.7	20.3
SPOS+NeuGen	38.3	90.8	55.0	<u>83.4</u>	34.7	46.5	75.5	50.9	59.5	<u>98.4</u>	<u>88.4</u>	69.0
SPOS	17.8	93.2	20.5	79.6	16.6	21.1	70.0	66.2	51.5	95.9	81.8	25.6
Autoformer+NeuGen	91.0	38.0	73.5	<u>93.8</u>	36.9	66.1	60.5	41.2	48.9	98.5	90.2	42.1
Autoformer	95.2	26.2	79.8	92.9	30.7	68.4	56.5	47.8	45.3	99.1	83.6	24.6
	Models VGG19+NeuGen VGG19 DenseNet121+NeuGen DenseNet121 ShuffleNet+NeuGen ShuffleNet Xception NASNetMobile+NeuGen NASNetMobile SPOS+NeuGen SPOS Autoformer+NeuGen Autoformer	Models74.2VGG19+NeuGen70.9DenseNet121+NeuGen26.4DenseNet12174.3ShuffleNet+NeuGen7.0ShuffleNet9.3Xception+NeuGen61.9Xception62.4NASNetMobile+NeuGen33.0SPOS+NeuGen38.3SPOS17.8Autoformer+NeuGen91.0Autoformer95.2	Models X X VGG19+NeuGen 74.2 24.5 VGG19 70.9 13.7 DenseNet121+NeuGen 26.4 15.9 DenseNet121 74.3 11.7 ShuffleNet+NeuGen 7.0 56.1 ShuffleNet 9.3 19.0 Xception+NeuGen 61.9 20.9 Xception 62.4 19.4 NASNetMobile+NeuGen 45.4 21.1 NASNetMobile 33.0 12.7 SPOS+NeuGen 38.3 90.8 SPOS 17.8 93.2 Autoformer+NeuGen 91.0 38.0 Autoformer 95.2 26.2	Models N N N VGG19+NeuGen 74.2 24.5 62.6 VGG19 70.9 13.7 39.7 DenseNet121+NeuGen 26.4 15.9 50.8 DenseNet121 74.3 11.7 38.4 ShuffleNet+NeuGen 7.0 56.1 71.1 ShuffleNet 9.3 19.0 14.1 Xception+NeuGen 61.9 20.9 59.7 Xcseption 62.4 19.4 56.0 NASNetMobile+NeuGen 45.4 21.1 47.5 NASNetMobile 33.0 12.7 36.0 SPOS 17.8 93.2 20.5 Autoformer+NeuGen 91.0 38.0 73.5 Autoformer 95.2 26.2 79.8	Models N N N N VGG19+NeuGen 74.2 24.5 62.6 48.4 VGG19 70.9 13.7 39.7 66.2 DenseNet121+NeuGen 26.4 15.9 50.8 43.1 DenseNet121 74.3 11.7 38.4 39.5 ShuffleNet+NeuGen 7.0 56.1 71.1 22.6 ShuffleNet 9.3 19.0 14.1 44.9 Xception+NeuGen 61.9 20.9 59.7 41.3 Xception 62.4 19.4 56.0 20.9 NASNetMobile+NeuGen 45.4 21.1 47.5 25.5 NASNetMobile 33.0 12.7 36.0 24.8 SPOS+NeuGen 38.3 90.8 55.0 83.4 SPOS 17.8 93.2 20.5 79.6 Autoformer+NeuGen 91.0 38.0 73.5 93.8 Autoformer 95.2 26.2 79.8 92.9	Models i <td>Models N<td>Models xi <th< td=""><td>Models N S N S N S N S N N S N N S N N S N N S N N S N N S N<td>Models xi <th< td=""><td>Models N. <th< td=""><td>Models N</td></th<></td></th<></td></td></th<></td></td>	Models N <td>Models xi <th< td=""><td>Models N S N S N S N S N N S N N S N N S N N S N N S N N S N<td>Models xi <th< td=""><td>Models N. <th< td=""><td>Models N</td></th<></td></th<></td></td></th<></td>	Models xi xi <th< td=""><td>Models N S N S N S N S N N S N N S N N S N N S N N S N N S N<td>Models xi <th< td=""><td>Models N. <th< td=""><td>Models N</td></th<></td></th<></td></td></th<>	Models N S N S N S N S N N S N N S N N S N N S N N S N N S N <td>Models xi <th< td=""><td>Models N. <th< td=""><td>Models N</td></th<></td></th<></td>	Models xi xi <th< td=""><td>Models N. <th< td=""><td>Models N</td></th<></td></th<>	Models N. N. <th< td=""><td>Models N</td></th<>	Models N

1123

Image classification: We compare pre-trained DNNs, including popular CNNs and novel NASbased architectures, across a range of domain transfer tasks with and without NeuGen. The datasets
involved are MNIST (M) (5), SVHN (S) (6), USPS (U) (7), and MNIST-M (MM) (8), which are
commonly used benchmarks for evaluating image classification performance under domain shifts.

MNIST is a dataset of handwritten digits (5), SVHN contains images of house numbers from Google
Street View (6), USPS is another dataset of handwritten digits from the U.S. Postal Service (7), and
MNIST-M is a variant of MNIST with blended color backgrounds from random patches of color
images (8). We evaluate the classification accuracy of the DNNs on various domain adaptation
scenarios, such as training on one dataset and testing on another, to assess the models' robustness to
domain shifts. By incorporating NeuGen layers into the existing architectures, we aim to enhance
the models' ability to generalize across different domains.

Our results, as presented in **Supplementary Table 13**, demonstrate that integrating NeuGen layers leads to improved classification performance across the different domain transfer tasks. Specifically, we observe significant accuracy improvements when models are trained on one dataset and tested on another, indicating enhanced generalization capabilities. This suggests that NeuGen can effectively enhance the representational quality of neural network outputs, improving their robustness and accuracy in image classification tasks under domain shifts.

Supplementary Table 14: Comparison of Mean IoU Scores for SAM and SAM + NeuGen fine-tuned on daytime images of the Cityscapes dataset and tested on nighttime images of the Dark Zurich Dataset.

÷	÷	л	5
1	1	*	0

1140

	Mean IoU Score
SAM	0.21
NeuGen-SAM	0.23

1146 1147

2D natural image segmentation. We integrate NeuGen into the Segment Anything Model (SAM) architecture, fine-tuning it on the Cityscapes dataset with daytime urban scenes (14) and testing on the Dark Zurich dataset with nighttime environments (15). This setup assesses the model's adaptability to significant illumination changes. Results show a notable performance boost, with the mean IoU score increasing from 0.21 to 0.23, demonstrating NeuGen's capacity to enhance domain generalization in image segmentation tasks despite minimal fine-tuning, as shown in Supplementary Table 14.

These results collectively demonstrate that NeuGen is capable of enhancing the representational quality of neural network outputs across diverse architectures, thereby solidifying its utility in domains where high fidelity and accurate image representation are critical.

1158

1174

1175

1159 REFERENCES

- [1] Guoyan Zheng, Gang Li, Mengyao Wang, Feng Shi, and Dinggang Shen. 2017 International Workshop on Infant Brain Analysis (iSeg-2017). In MICCAI Grand Challenge on 6-month Infant Brain MRI Segmentation, 2017.
- [2] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for
 Biomedical Image Segmentation. *In International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241. Springer, 2015.
- [3] Tammie L. S. Benzinger, John C. Morris, and Daniel S. Marcus. OASIS-3: Longitudinal Neuroimaging, Clinical, and Cognitive Dataset for Normal Aging and Alzheimer's Disease. *medRxiv*, 2019.
- [4] Ying Xu, Zhiming Shen, Hao Xie, Y. K. Lai, and Jian Jun Zhang. FourierNet: Fast Image Registration via Global Frequency Analysis and Dense Local Networks. *arXiv preprint arXiv:2103.07541*, 2021.
 - [5] 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.
- 1176
 1177 [6] Yuval Netzer, Tao Wang, Adam Coates, Alessandro Bissacco, Bo Wu, and Andrew Y. Ng. Reading digits in natural images with unsupervised feature learning. *In NIPS Workshop on Deep Learning and Unsupervised Feature Learning*, 2011.
- [7] J. J. Hull. A database for handwritten text recognition research. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(5):550–554, 1994.
- [8] Yaroslav Ganin and Victor Lempitsky. Unsupervised Domain Adaptation by Backpropagation. *In International Conference on Machine Learning*, pages 1180–1189. PMLR, 2015.
- [9] Anpei Chen, Zexiang Xu, Fuqiang Zhao, Xiaoshuai Zhang, Fanbo Xiang, Jingyi Yu, and Hao
 Su. MVSNeRF: Fast Generalizable Radiance Field Reconstruction from Multi-View Stereo. *In Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 14124–14133, 2021.

- [10] Mohammad Mahdi Johari, Yann Lepoittevin, and François Fleuret. GeoNeRF: Generalizing NeRF with Geometry Priors. *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18365–18375, 2022.
- [11] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoor-thi, and Ren Ng. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. *Communications of the ACM*, 65(1):99–106, 2021.
- [12] Rasmus Jensen, Anders Dahl, George Vogiatzis, Engil Tola, and Henrik Aanæs. Large Scale
 Multi-View Stereopsis Evaluation. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 406–413, 2014.
- [13] Ben Mildenhall, Pratul P. Srinivasan, Rodrigo Ortiz-Cayon, Nima Khademi Kalantari, Ravi Ramamoorthi, Ren Ng, and Abhishek Kar. Local Light Field Fusion: Practical View Synthesis with Prescriptive Sampling Guidelines. *ACM Transactions on Graphics (TOG)*, 38(4):1–14, 2019.
- [14] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The Cityscapes Dataset for Semantic Urban Scene Understanding. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3213–3223, 2016.
- [15] Christos Sakaridis, David Dai, and Luc Van Gool. Guided Curriculum Model Adaptation and Uncertainty-Aware Evaluation for Semantic Nighttime Image Segmentation. *In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 7374–7383, 2019.