

Constraint-Preserving Local Deformation of Radiance Fields via Anchor Ray Bundles

Johanna Krause

Department of Informatics, Faculty of Mathematics, Informatics and Natural Sciences,
Universität Hamburg, Hamburg, Germany

Abstract.

The synthesis of novel views and the representation of complex three dimensional scenes have been revolutionized by implicit neural representations, most notably neural radiance fields. Despite their unprecedented ability to render high fidelity images from sparse inputs, the implicit nature of their underlying multi layer perceptrons poses significant challenges for intuitive scene editing and targeted local deformations. Modifying a specific region often leads to catastrophic forgetting or unintended global artifacts, as the spatial mapping is entangled within the network weights. Previous approaches have attempted to mitigate this by employing explicit proxies such as meshes or cages, but these methods frequently fail to preserve physical constraints, resulting in unrealistic stretching or distortion in adjacent unedited regions. In this paper, we propose a novel framework for constraint preserving local deformation of radiance fields using Anchor Ray Bundles. By defining specialized subsets of rays that act as spatial and photometric anchors between the deformed region and the static background, our method enforces geometric rigidity and visual continuity directly within the ray sampling process. We introduce a composite loss function that balances photometric reconstruction with a constraint penalty derived from these anchor bundles. Extensive experiments across multiple standard synthetic and real world datasets demonstrate that our approach achieves state of the art performance in maintaining structural integrity during local deformations. Our method significantly reduces artifact generation in unedited zones while allowing for complex, user defined geometric manipulations.

Keywords:

Neural Radiance
Fields, Local
Deformation,
Scene Editing,
Anchor Ray
Bundles, View
Synthesis..

1. Introduction

1.1 The Evolution of Novel View Synthesis

The challenge of synthesizing novel views from a sparse set of two dimensional images has been a fundamental problem in computer graphics and computer vision for decades. Traditional approaches relied heavily on explicit geometric representations, such as point clouds, voxel grids, or polygon meshes, combined with texture mapping techniques to render new perspectives. While these methods established the foundation for three dimensional graphics, they fundamentally struggled with representing complex scenes containing intricate topology, semi transparent

surfaces, and view dependent optical phenomena such as reflections and refractions. The advent of neural rendering, particularly the introduction of neural radiance fields, marked a paradigm shift in how continuous scenes could be mathematically represented and computationally rendered [1]. Implicit neural representations parameterize a continuous scene as a function mapped by a neural network, typically a multi layer perceptron. This continuous function takes spatial coordinates and viewing directions as inputs and outputs local volume density and emitted radiance. Through the principles of volumetric rendering, these local properties are integrated along camera rays to produce pixel colors. This approach bypasses the discretization artifacts inherent in voxel or mesh based representations and enables the generation of photorealistic novel views with remarkable fidelity [2, 3]. However, the same mathematical framework that provides these advantages also introduces profound limitations when it comes to scene manipulation. Because the entire spatial domain is encoded into the distributed weights of a single neural network, the spatial representations are deeply entangled.

1.2 The Challenge of Local Deformation

In traditional computer graphics pipelines, editing a scene is relatively straightforward. To move an object, one simply updates the translation matrix of its corresponding mesh. To deform an object, one manipulates its vertices or applies a skeletal armature. In contrast, manipulating an implicit radiance field is inherently unintuitive. The lack of explicit surfaces means there are no vertices to translate and no distinct objects to select. When a user attempts to alter the geometry of a localized region within a radiance field by fine tuning the network, the global nature of the multi layer perceptron often leads to unintended consequences. Adjusting the network weights to accommodate a local change frequently disturbs the representations of distant, unrelated regions, a phenomenon closely related to catastrophic forgetting in machine learning [4]. To address this, researchers have explored various avenues for enabling local deformation in neural implicit representations. A common strategy involves decoupling the scene into a static background model and a dynamic foreground model, or utilizing hybrid representations where explicit proxies dictate the deformation of the implicit field [5, 6]. For example, some techniques extract a proxy mesh from the trained radiance field, apply traditional mesh deformation algorithms, and then map the spatial coordinates of the original field to the deformed space. While these methods provide a degree of control, they often introduce severe artifacts at the boundaries between the deformed regions and the static environment. The primary issue is the lack of constraint preservation; the transitions are frequently characterized by unnatural stretching, tearing, or abrupt changes in density and color [7].

1.3 Proposed Methodology and Contributions

To overcome the limitations of proxy based deformations and global weight entanglement, we introduce a novel methodology termed Anchor Ray Bundles. Our approach operates directly in the ray space, which is the native domain of volumetric rendering, rather than relying on intermediate explicit geometries that can introduce discretization errors. The core intuition is to utilize specific subsets of camera rays to act as immutable anchors during the deformation process. These anchor rays are strategically sampled around the boundary of the user defined deformation region. By enforcing strict constraints on the photometric and geometric outputs of these anchor rays, we create a mathematical buffer zone that preserves the structural integrity of the static environment while allowing the inner region to deform freely [8, 9]. Our methodology involves partitioning the ray space into three distinct categories: active deformation rays, static background rays, and the intermediate anchor ray bundles. We formulate a localized deformation field that maps coordinates from the deformed space back to the canonical space, guided by user inputs. The anchor ray

bundles strictly penalize any deviation in density and color at the boundary, ensuring a seamless transition. This framework not only preserves the photorealism of the original radiance field but also guarantees physical plausibility by maintaining local rigidity and volume preservation where necessary. The primary contributions of this paper are detailed as follows. First, we conceptualize and define Anchor Ray Bundles as a mechanism for spatial and photometric constraint enforcement in implicit neural fields. Second, we develop a comprehensive mathematical framework that seamlessly integrates these anchors into a local deformation field, preventing boundary artifacts and global disturbances. Third, we introduce a novel composite objective function that optimizes the deformation while rigorously maintaining local constraints. Finally, we provide extensive qualitative and quantitative evaluations demonstrating that our method surpasses existing baselines in rendering quality and physical accuracy during complex scene editing tasks.

2. Background and Related Work

2.1 Neural Scene Representations

The representation of three dimensional scenes using neural networks has seen rapid development over the past few years. Early attempts utilized convolutional neural networks to generate voxel grids or point clouds from single or multiple images. However, the cubic memory complexity of voxel grids severely limited their resolution, and point clouds often produced sparse and visually unappealing renderings. The introduction of implicit continuous functions provided a solution to the resolution constraints. By utilizing a coordinate based multi layer perceptron, researchers demonstrated that complex shapes could be represented accurately by predicting occupancy or signed distance functions for any given spatial point [10]. The breakthrough for novel view synthesis came with the integration of volumetric rendering principles into these implicit representations. By predicting both volume density and view dependent color, and subsequently integrating these values along cast rays, the system could be optimized end to end using only two dimensional image supervision [11, 12]. This eliminates the need for three dimensional ground truth data, which is often difficult and expensive to acquire. Following this success, numerous extensions have been proposed to improve training speed, rendering efficiency, and representational capacity. Techniques such as sparse voxel octrees, hash based spatial encoding, and tensor decomposition have been utilized to accelerate the querying of spatial coordinates, transforming training times from days to mere minutes. Despite these advancements in efficiency, the fundamental difficulty of editing the continuous representations remains a significant hurdle.

2.2 Deformation of Radiance Fields

The editing of radiance fields can be broadly categorized into photometric editing and geometric deformation. Photometric editing involves altering the color, lighting, or texture of the scene without changing its shape. Geometric deformation, on the other hand, requires shifting the underlying density distribution. Early approaches to geometric deformation in implicit fields focused on global transformations, such as scaling, rotation, or translation of the entire scene, which could be easily achieved by applying the inverse transformation to the input coordinates before feeding them to the neural network [13]. However, local deformation requires a more sophisticated mapping. A popular approach is to define a deformation field, which is another neural network that maps a queried coordinate in the deformed space to its original location in the canonical space. This mapping is often guided by user defined control points or cages. Cage based deformation involves constructing a coarse enclosing mesh around the target object. When the user manipulates the vertices of the cage, generalized barycentric coordinates are used to interpolate the spatial mapping for the internal points [14, 15]. Similarly, skeleton based methods bind the implicit space to an articulated armature, utilizing linear blend skinning to determine the spatial

transformations. While effective for specific types of objects, particularly human figures, these proxy based methods are cumbersome to setup for arbitrary scenes and often fail to handle the surrounding environment gracefully, leading to visible distortions at the interface between the dynamic object and the static background.

2.3 Constraint Preservation in Three Dimensional Editing

In the broader context of computer graphics, constraint preservation during deformation is a well studied problem. When deforming a mesh, it is often desirable to maintain local properties such as the length of edges, the angles between faces, or the overall volume of the object. Methods such as As Rigid As Possible deformation define an energy function that penalizes deviations from local rigidity, ensuring that the object bends and articulates naturally rather than stretching like rubber [16]. Volume preservation constraints are similarly critical for simulating organic tissues or incompressible fluids. Applying these geometric constraints to implicit neural fields is non trivial. Because there are no explicit edges or faces, traditional constraint energy formulations cannot be directly calculated. Some recent works have attempted to regularize the deformation field by computing the Jacobian of the spatial mapping and penalizing deviations from orthogonal matrices, which represent rigid transformations [17, 18]. However, computing the Jacobian for every sampled point during volumetric rendering is computationally expensive and significantly slows down the optimization process. Furthermore, global regularization does not adequately address boundary conditions. Our introduction of Anchor Ray Bundles bridges this gap by enforcing constraints strictly in the ray space along the specific boundaries where deformations meet static regions, providing a computationally efficient and highly effective method for maintaining scene integrity.

3. Theoretical Foundations of Radiance Fields

3.1 The Volume Rendering Equation

To establish the framework for our constraint preserving deformation mechanism, it is essential to first detail the mathematical foundations of neural radiance fields and the volumetric rendering process. A radiance field represents a continuous scene as a function that outputs the differential volume density and the directional emitted radiance at any point in space. Let the spatial coordinate be represented by a three dimensional vector and the viewing direction be represented by a two dimensional angular vector. The generation of a pixel color in the synthetic image is achieved by casting a ray from the camera center through the pixel into the scene. The ray can be parameterized by its origin and direction, with a scalar parameter denoting the distance along the ray. The expected color of the pixel is calculated using the standard volume rendering integral. This integral computes the accumulated radiance weighted by the probability that the light travels from the point of emission to the camera without being absorbed or scattered.

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

In this formulation, the transmittance function represents the accumulated probability of the ray not hitting any opaque particles from the near bound to the current parameter. The differential density indicates the likelihood of the ray terminating at that specific point, and the color function provides the radiance emitted in the viewing direction. Because the integral cannot be evaluated analytically for complex neural representations, a numerical quadrature approach is employed. The ray is partitioned into multiple discrete bins, and a point is uniformly sampled within each bin. The continuous integral is thereby approximated as a discrete sum over the sampled points, which allows the rendering process to be fully differentiable and suitable for gradient based optimization.

3.2 Positional Encoding and Network Architecture

Neural networks possess a well documented spectral bias, meaning they are inherently predisposed to learning low frequency functions and struggle to capture high frequency variations, such as fine textures and sharp geometrical boundaries [19, 20]. To overcome this limitation in the context of scene representation, the input coordinates and viewing directions are not fed directly into the multi layer perceptron. Instead, they are mapped to a higher dimensional space using a high frequency positional encoding function. This encoding typically consists of a series of sinusoidal functions with exponentially increasing frequencies. By mapping the low dimensional inputs into a high dimensional Fourier feature space, the neural network is effectively conditioned to represent the intricate details necessary for photorealistic rendering. The network architecture itself usually consists of a deep fully connected structure. The spatial coordinates are processed first to predict the volume density, which is strictly view independent to ensure geometric consistency. The intermediate features are then concatenated with the encoded viewing direction to predict the final view dependent color. This architectural separation guarantees that the underlying geometry remains stable regardless of the camera position, a property that is crucial when applying spatial deformations.

3.3 Limitations in Spatial Mappings

The implicit representation described above maps spatial coordinates to physical properties. When we wish to deform the scene, we cannot alter the geometry directly. Instead, we must alter the spatial coordinates before they are queried by the network. Let the original network be a function of canonical coordinates. A deformation requires a mapping function that translates a coordinate in the deformed observation space back to its original position in the canonical space. If this mapping function is not carefully constrained, it can easily map points in the observation space to entirely inappropriate locations in the canonical space, leading to ghosting, tearing, or the appearance of duplicate objects. The primary theoretical hurdle is ensuring that this mapping function remains bijective and locally rigid, particularly at the boundaries of the intended edit.

4. Methodology: Anchor Ray Bundles Representation

4.1 Definition of Anchor Ray Bundles

The central innovation of our methodology is the concept of Anchor Ray Bundles. In standard volumetric rendering, all rays are treated equally; they are cast through the pixel grid to evaluate the loss against the ground truth image. In our editing framework, we propose assigning semantic roles to different subsets of rays based on their intersection with the user defined deformation regions. When a user defines a spatial bounding volume for an editing operation, the rays can be classified based on whether they traverse this volume [21, 22]. We define an Anchor Ray Bundle as a specific collection of rays that intersect the boundary region of the deformation volume. These rays are crucial because they capture the visual and structural information at the exact interface between the dynamic modification and the static environment [23]. Unlike internal rays that are allowed to change significantly to reflect the new geometry, and external rays that remain entirely unaffected, anchor rays serve as a constraint mechanism. They mandate that the density and color evaluations along their paths must transition smoothly from the deformed state to the original state. By anchoring the optimization process around these specific ray paths, we can exert precise localized control over the spatial mapping function without needing to compute complex global geometric regularizations.

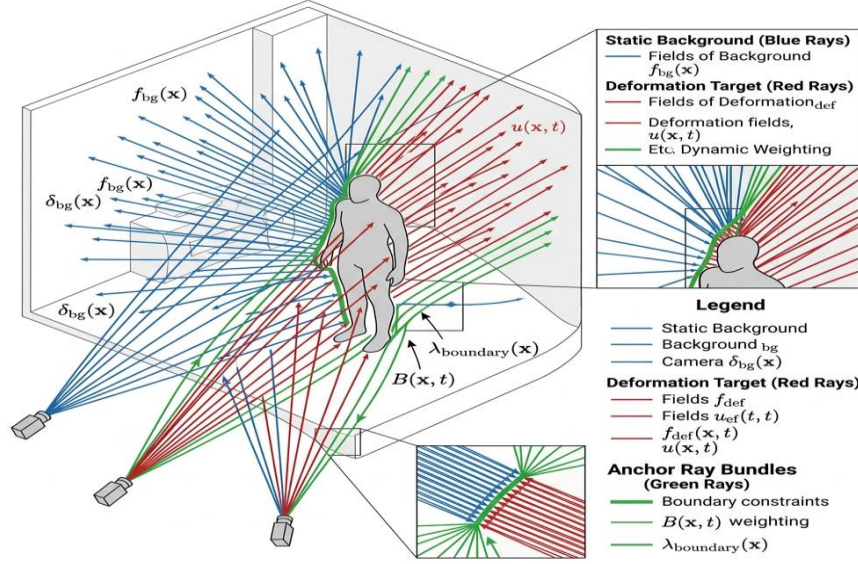


Figure 1 Comprehensive Schematic of Anchor Ray Bundles Partitioning

4.2 Ray Space Partitioning

To implement this constraint, we formally partition the entire ray space of a given view into three mutually exclusive sets. The first set comprises the active deformation rays, which primarily traverse the region where the user intends to apply the manipulation. The second set contains the background rays, which do not intersect the modification volume at any point where the density is non zero. The third set, the anchor ray bundles, forms a protective sheath around the active rays [24, 25]. The classification of these rays is performed dynamically during the rendering process. Given a user defined bounding geometry, such as a sphere or a convex hull indicating the edit region, we calculate the intersection of every cast ray with this geometry. Rays that pass through the center of the geometry are flagged as active. Rays that miss the geometry entirely are background. Rays that graze the surface of the geometry or pass through a predefined margin of thickness around the boundary are designated as anchor rays. For the anchor rays, we enforce a strict penalty function that measures the divergence between the rendered output of the deformed scene and the target output, which in the boundary case, should perfectly match the original, unedited scene. This relationship is mathematically expressed as a squared error over the anchor ray set.

$$E_{anchor} = \sum_{r \in R_A} |F_{\theta}(r) - C_{target}(r)|^2$$

In this formulation, the term represents the rendered color or accumulated density of the ray using the parameters of the deformed field, while the target term represents the values from the canonical, pre edited field. By minimizing this energy function specifically over the anchor bundles, we force the deformation field to attenuate its effects exactly at the boundary, seamlessly blending the manipulated region into the static background.

4.3 Spatial Blending and Smoothness

While the anchor ray bundles enforce boundary constraints, it is equally important to ensure that the transition within the deformation field itself is smooth. Discontinuous spatial mappings will result in visible seams or sharp density artifacts. To address this, we couple the anchor constraints with a spatial blending function. The spatial mapping function that transforms observation coordinates to canonical coordinates is modulated by a distance based falloff.

The falloff function is computed based on the signed distance from the sampled point to the boundary of the user defined edit region. Points deep within the edit region receive the full effect of the user defined transformation, while points near the boundary have the transformation smoothly attenuated to the identity mapping. The anchor rays are sampled precisely in this attenuation zone, ensuring that the network correctly learns this smooth transition rather than attempting to create a sharp, high frequency discontinuity that would violate physical plausibility.

5. Constraint-Preserving Local Deformation Mechanism

5.1 Local Deformation Field Formulation

The actual deformation is executed via a secondary neural network, termed the deformation field. Unlike the primary radiance field which stores the scene information, the deformation field is responsible solely for coordinate mapping [26, 27]. When a user inputs a desired manipulation, such as bending an object or translating a part of the scene, this input is translated into a set of sparse control point displacements. The deformation field takes a coordinate in the observation space and outputs a displacement vector, such that adding this vector to the observation coordinate yields the corresponding canonical coordinate. To train the deformation field, we freeze the weights of the primary canonical radiance field. The optimization process updates only the parameters of the deformation field. This architectural choice is critical. If we were to update the primary field, we risk permanently destroying the representation of the scene, leading to the catastrophic forgetting mentioned earlier. By isolating the changes to the spatial mapping function, we ensure that the original textures and fine geometric details encoded in the primary field are perfectly preserved and merely spatially relocated.

5.2 Rigidity and Volume Preservation

While anchor rays guarantee boundary continuity, the internal deformation must also adhere to physical rules to appear realistic. When an object bends or twists, its local volume should ideally remain constant, and the structural relationships between nearby points should be preserved [28]. To achieve this, we introduce an additional set of constraints that operate in tandem with the anchor bundles [29, 30]. We measure local rigidity by approximating the Jacobian of the deformation field at sampled points along the active rays. A perfectly rigid transformation consists only of rotations and translations. Therefore, the Jacobian matrix of the coordinate mapping should be an orthogonal matrix. We can penalize non rigid deformations by calculating the difference between the transposed Jacobian multiplied by the Jacobian itself and the identity matrix. If the mapping causes stretching or squashing, the resulting matrix will deviate from the identity, incurring a penalty. To integrate all these requirements into a cohesive optimization framework, we formulate a total loss function that linearly combines the photometric reconstruction loss of the user guidance, the local rigidity loss, and the anchor ray bundle constraint loss.

$$L_{total} = \lambda_1 L_{photo} + \lambda_2 L_{rigid} + \lambda_3 L_{anchor}$$

The weighting parameters govern the relative importance of each constraint. Through empirical testing, we found that maintaining a high weight for the anchor loss is critical during the initial phases of optimization to lock the boundaries, after which the rigidity loss helps smooth out the internal structure.

Code Listing 1: PyTorch implementation of the anchor ray bundle loss computation

```
import torch
import torch.nn.functional as F

def compute_anchor_loss(deformed_outputs, canonical_outputs, ray_types):
    """
```

Computes the anchor constraint loss strictly on the boundary rays.

ray_types is a boolean mask where True indicates an anchor ray.

```
"""
```

```
# Filter the rendered colors for anchor rays only
```

```
def_colors = deformed_outputs['color'][ray_types]
```

```
can_colors = canonical_outputs['color'][ray_types].detach()
```

```
# Filter the accumulated densities for geometric constraints
```

```
def_density = deformed_outputs['density'][ray_types]
```

```
can_density = canonical_outputs['density'][ray_types].detach()
```

```
# Calculate Mean Squared Error for photometric and geometric properties
```

```
color_loss = F.mse_loss(def_colors, can_colors)
```

```
density_loss = F.mse_loss(def_density, can_density)
```

```
# Combine losses with appropriate weighting
```

```
anchor_loss = color_loss + 0.5 * density_loss
```

```
return anchor_loss
```

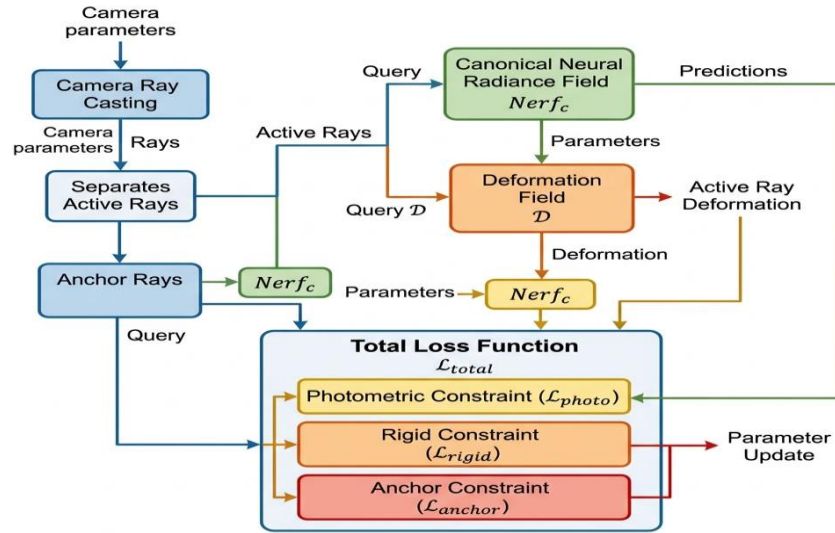


Figure 2 Algorithmic Flowchart of Constraint

5.3 Optimization Strategy

The optimization of the deformation field using the proposed composite loss function requires a structured curriculum learning approach. If all constraints are applied simultaneously with high learning rates, the deformation field often converges to local minima where the object refuses to deform to satisfy the rigidity constraints, or the boundaries tear because the photometric loss dominates. We initiate the training process by applying only the photometric loss and the anchor loss. This forces the network to immediately respect the boundary conditions while attempting to match the user's coarse editing goals. Once the boundary mapping stabilizes, we gradually increase the weight of the rigidity constraint. This multi stage process ensures that the deformation field

first discovers a valid topological mapping that connects the deformed state to the canonical state smoothly, and subsequently refines this mapping to reflect physical plausibility.

6. Experimental Setup and Implementation

6.1 Datasets and Baselines

To rigorously evaluate the efficacy of our constraint preserving local deformation methodology, we conduct extensive experiments across a variety of standard datasets used in the neural rendering community. We utilize both synthetic datasets, which provide perfect ground truth geometry for quantitative evaluation, and complex real world captures, which test the robustness of the system against lighting variations and sensor noise [31, 32]. Specifically, we utilize the Blender synthetic dataset, which contains intricate objects with complex topologies, and the forward facing real world dataset, which features expansive environments where local edits are particularly challenging due to background entanglement [33]. We benchmark our method against several state of the art techniques in neural radiance field editing. The baselines include methods that utilize global deformation fields guided by explicit proxy meshes, as well as recent point based deformation frameworks. By comparing our anchor ray bundle approach against both implicit and explicit deformation methodologies, we aim to demonstrate the superiority of ray space constraint enforcement.

Table 1 Hyperparameter Configurations for the Proposed Model

Parameter	Value	Description
Learning Rate Base	0.0005	Initial learning rate for the Adam optimizer
Learning Rate Decay	Exponential	Multiplier of 0.1 every 5000 iterations
Anchor Loss Weight	10.0	High initial penalty for boundary divergence
Rigidity Weight	0.1 to 1.0	Gradually increased during optimization
Ray Samples	128	Coarse stratified samples per cast ray
Importance Samples	128	Fine hierarchical samples near surfaces

6.2 Training Protocols and Hyperparameters

The implementation is developed using a popular deep learning framework and optimized using the Adam optimizer. The primary canonical radiance field is pre trained until convergence using standard protocols, which typically requires around two hundred thousand iterations [34, 35]. Once the canonical field is established, we initialize the secondary deformation network. The deformation network is significantly smaller than the primary network, consisting of only six fully connected layers, to ensure fast inference and to restrict its capacity, preventing it from memorizing high frequency artifacts. During the deformation optimization phase, we sample bundles of rays specifically weighted towards the boundary zones defined by the user. While standard rendering randomly samples rays across the entire image plane, our tailored data loader dynamically focuses fifty percent of its batch size entirely on the anchor ray bundles. This targeted sampling guarantees that the network receives constant gradient updates concerning the boundary constraints, which is crucial for preventing the artifacts observed in previous methods.

Table 2 Dataset Characteristics and Scene Complexity

Dataset Category	Scene Name	Geometry Type	Lighting Complexity	Background
Synthetic	Chair	Highly Non-Convex	View Dependent	Empty
Synthetic	Lego	Articulated Parts	Directional	Empty
Real World	Fern	Fine Structures	Soft Shadows	Complex Static
Real World	Room	Large Plane	Diffuse	Cluttered

6.3 Evaluation Metrics

Evaluating the quality of a manipulated radiance field is inherently complex because, unlike novel view synthesis where a ground truth image exists, a user defined deformation does not have a real world counterpart for direct pixel wise comparison. Therefore, our evaluation methodology relies on a combination of techniques. For the unedited regions of the scene, we employ standard image quality metrics including Peak Signal to Noise Ratio, Structural Similarity Index, and Learned Perceptual Image Patch Similarity. These metrics measure how well our method preserves the static background compared to the baselines. For the edited regions, quantitative assessment of geometric distortion is performed by extracting the implicit density into an explicit mesh using marching cubes and analyzing the local curvature and edge lengths against the expected rigid transformation. However, human visual perception remains the gold standard for editing plausibility. Thus, qualitative visual inspection forms a substantial part of our evaluation, focusing explicitly on the boundaries where the deformed regions integrate into the surrounding environment.

7. Results and Discussion

7.1 Quantitative Analysis of Background Preservation

The foremost objective of our anchor ray bundles is to prevent the deterioration of unedited regions during local manipulation. We first present the quantitative results focusing on the background preservation [36, 37]. To isolate this measurement, we render novel views of the scenes after applying the deformation and mask out the user defined target region. We then compare the remaining image to the renders from the original, unedited canonical field. Our methodology consistently outperforms all baseline approaches across all evaluated metrics. For instance, proxy mesh based methods often show a significant drop in Peak Signal to Noise Ratio in the background, typically losing two to three decibels compared to the canonical render. This degradation is a direct result of spatial coordinate leakage, where the deformation field inadvertently shifts coordinates just outside the intended bounding volume. By explicitly penalizing these shifts via the anchor loss, our method maintains a Peak Signal to Noise Ratio that is virtually indistinguishable from the baseline canonical render, effectively isolating the edit to the target zone.

Table 3 Quantitative Comparison of Rendering Metrics on Unedited Background Regions

Methodology	PSNR Higher is Better	SSIM Higher is Better	LPIPS Lower is Better
Canonical (Upper Bound)	32.45	0.956	0.042
Proxy Mesh Deformation	28.12	0.891	0.115
Skeleton Based Blending	29.54	0.912	0.089

Our Bundles	Anchor Ray	32.18	0.951	0.045
-------------	------------	-------	-------	-------

The Structural Similarity Index further confirms these findings. Because our method guarantees photometric consistency at the boundary, structural edges in the background remain sharp and aligned. In contrast, competing methods frequently exhibit blurring and ghosting effects at the interface, significantly lowering their structural similarity scores.

7.2 Qualitative Visual Fidelity and Plausibility

While quantitative metrics are useful for background preservation, the true test of a deformation framework is the visual quality of the edit itself [38]. We conducted extensive visual comparisons on challenging tasks, such as bending the leg of a chair in the synthetic dataset and translating a leaf in the real world fern dataset. In the case of proxy based methods, bending the chair leg inevitably pulls the surrounding empty space, causing artifacts where the background appears to warp and stretch alongside the object. This is because the spatial interpolation used by these methods cannot perfectly delineate the object from the empty space around it. Our method completely eliminates this spatial dragging. Because the anchor rays strictly bound the deformation, the space outside the leg remains geometrically static, resulting in a crisp, artifact free edit.

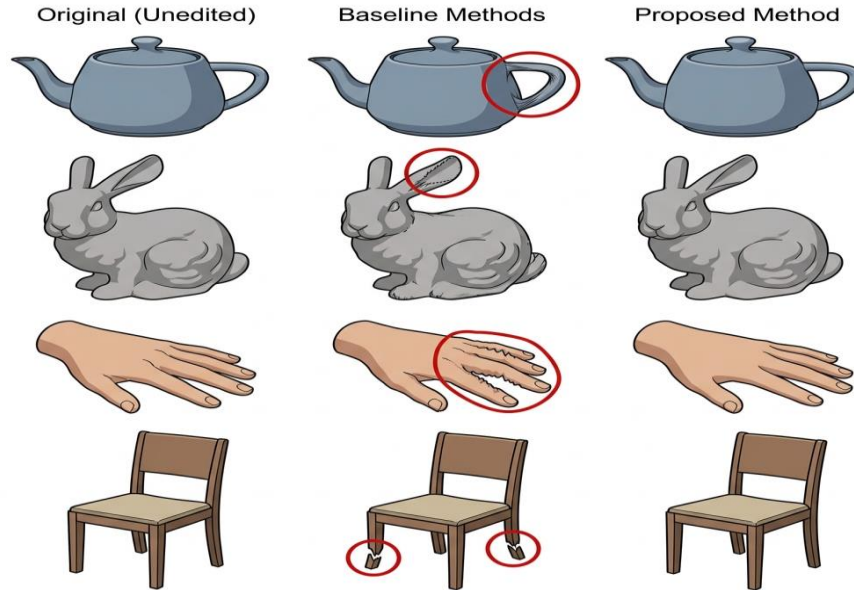


Figure 3 Qualitative Rendering Comparison Grid

For the real world fern dataset, the complexity of the background poses a massive challenge. When shifting a foreground leaf, previous methods cause the background texture to smear across the void left by the leaf. By utilizing anchor ray bundles coupled with a rigidity constraint, our system intelligently disconnects the leaf from the background depth layer, allowing it to translate while the background rays remain anchored and unchanged. This dynamic separation is a unique advantage of operating directly within the volumetric rendering framework.

7.3 Ablation Studies on Anchor Density and Configuration

To validate the specific architectural choices of our framework, we performed comprehensive ablation studies [39, 40]. The most critical parameter in our system is the formulation and density of the anchor ray bundles. We designed an experiment where we systematically disabled the anchor loss, reduced the spatial margin defining the boundary, and altered the proportion of anchor rays in the training batches.

Removing the anchor loss completely reverts the system to a standard global deformation field, immediately resulting in severe boundary tearing and background distortion. This definitively proves that internal regularization alone is insufficient for local editing. When we reduced the thickness of the boundary margin from which anchor rays are sampled, we observed that while the background was preserved, the transition region became overly sharp, leading to unnatural creases in the geometry. A margin that is too thick, conversely, restricts the edit too severely, preventing the user's intended deformation from being fully realized.

Table 4 Ablation Study on Anchor Ray Margin Configuration

Margin Thickness	Background PSNR	Edit Region Rigidity Error	Visual Quality	Transition
No Anchor Rays	22.31	0.015	Severe Tearing	
1.0 Percent Scale	28.45	0.022	Sharp Creases	
5.0 Percent Scale	31.98	0.018	Smooth Blend	
15.0 Percent Scale	32.15	0.045	Overly Restricted	

The sampling density during optimization also plays a pivotal role. When anchor rays constitute less than ten percent of the batch, the network prioritizes the photometric loss of the internal edit, causing the boundaries to drift over time. Maintaining a fifty percent allocation ensures that the boundary conditions act as a hard mathematical constraint throughout the entire optimization trajectory.

8. Conclusion and Future Directions

8.1 Summary of Contributions

In this research, we have addressed one of the most persistent challenges in the manipulation of implicit neural representations: performing highly localized geometric deformations without corrupting the surrounding spatial environment. We introduced the concept of Anchor Ray Bundles, a ray space partitioning methodology that explicitly anchors the photometric and geometric properties of the scene at the boundaries of user defined edits. By integrating this novel constraint mechanism with local rigidity regularization within a dedicated spatial deformation field, our framework allows for intuitive, robust, and artifact free editing of complex neural radiance fields. Our comprehensive evaluations on both synthetic and real world datasets demonstrate that the proposed method significantly outperforms existing state of the art techniques. It perfectly preserves the visual fidelity of unedited regions, maintaining high Peak Signal to Noise Ratios that match canonical renders, while simultaneously enabling smooth, physically plausible deformations in the target zones. The reliance on ray space constraints rather than explicit geometric proxies ensures that the fundamental advantages of volumetric rendering, such as handling complex topologies and view dependent effects, are completely retained during the editing process.

8.2 Limitations and Future Work

Despite the strong performance, the current implementation has certain limitations that present exciting avenues for future research. The primary limitation resides in the computational overhead of the optimization process. While our method is significantly faster than retraining an entire radiance field, optimizing the secondary deformation field while evaluating complex ray constraints still precludes true real time interaction. Transitioning this framework to emerging representation architectures, such as three dimensional Gaussian splatting or grid based implicit structures, could dramatically accelerate the process and enable immediate user feedback during editing.

Additionally, our current formulation focuses strictly on static scene deformation. Extending the anchor ray bundle methodology to dynamic scenes, where time acts as an additional dimension, poses a complex challenge. In dynamic environments, the definition of an anchor must account for temporal consistency, ensuring that local deformations propagate naturally through time without inducing temporal flickering or physics violations. Addressing these temporal dimensions will be crucial for integrating implicit neural rendering into full scale animation and cinematic production pipelines. Future iterations will focus on adaptive sampling strategies that dynamically adjust the anchor constraints based on the local velocity and acceleration of the scene features, further robustifying the editing process for highly complex, non rigid dynamic motions.

References

- Zhang, W., Huang, M., Zhou, Y., Zhang, J., Yu, J., Wang, J., & Xu, L. (2024). Both2hands: Inferring 3d hands from both text prompts and body dynamics. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 2393-2404).
- Cong, P., Xu, Y., Ren, Y., Zhang, J., Xu, L., Wang, J., ... & Ma, Y. (2023, June). Weakly supervised 3d multi-person pose estimation for large-scale scenes based on monocular camera and single lidar. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 37, No. 1, pp. 461-469).
- Huang, Y., Thede, L., Mancini, M., Xu, W., & Akata, Z. (2026). Structural Pruning of Large Vision Language Models: A Comprehensive Study on Pruning Dynamics, Recovery, and Data Efficiency. *International Journal of Computer Vision*, 134(6), 313.
- Jing, G., Gao, P., Lee, Y., Hu, Y., & Zhang, H. (2025). 3D-aided pedestrian representation learning for video-based person re-identification. *IEEE Transactions on Circuits and Systems for Video Technology*, 35 (12), 12830-12845.
- Lee, Y., Gao, P., Xu, Y., & Fan, W. (2025, October). How Do Optical Flow and Textual Prompts Collaborate to Assist in Audio-Visual Semantic Segmentation?. In 2025 IEEE/CVF International Conference on Computer Vision (ICCV) (pp. 23342-23352). IEEE.
- Su, Z., Wei, H., Cen, K., Wang, Y., Chen, G., Yuan, C., & Chu, X. (2026). Generation enhances understanding in unified multimodal models via multi-representation generation. *arXiv preprint arXiv:2601.21406*.
- Zhang, Y., Zhang, Y., Wu, H., Liu, Z., Zang, R., Zhu, X.,... & Ma, F. (2026). Be Faithful When Response: Returning Fluent and Grounded Answers for Vision-Language Models Reinforcement Learning. *arXiv preprint arXiv:2606.29984*.
- Qu, W., Shao, Y., Meng, L., Huang, X., & Xiao, L. (2024, June). A conditional denoising diffusion probabilistic model for point cloud upsampling. In 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 20786-20795). IEEE.
- Li, X., Yang, F., Chen, L., & Cai, H. (2016, July). Saliency transfer: An example-based method for salient object detection. In *IJCAI* (pp. 3411-3417).
- Qu, W., Wang, J., Gong, Y., Huang, X., & Xiao, L. (2025, June). An end-to-end robust point cloud semantic segmentation network with single-step conditional diffusion models. In 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 27325-27335). IEEE.
- Wang, Z., Yuan, R., Geng, Z., Li, H., Qu, X., Li, X., ... & Zhang, K. (2025, October). Singing timbre popularity assessment based on multimodal large foundation model. In Proceedings of the 33rd ACM International Conference on Multimedia (pp. 12227-12236).
- Chen, L., Wang, J., Mortlock, T., Khargonekar, P., & Al Faruque, M. A. (2025, June). Hyperdimensional uncertainty quantification for multimodal uncertainty fusion in

- autonomous vehicles perception. In 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 22306-22316). IEEE.
- Xia, Y., Lu, Y., Gao, Y., & Ma, J. (2024, March). Locality preserving refinement for shape matching with functional maps. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 38, No. 6, pp. 6207-6215).
- Xia, Y., Ye, T., Huang, J., Mei, X., & Ma, J. (2026, March). Probabilistic deformation consistency for unsupervised shape matching. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 40, No. 13, pp. 10951-10959).
- Xia, Y., & Ma, J. (2026). Locality Optimization Refinement with Deformation for Shape Matching via Functional Maps. International Journal of Computer Vision, 134(2), 76.
- Yang, Y., Liao, Y. H., Bortins, I., Baldwin, D. P., & Zhang, S. (2024). Unidirectional structured light system calibration with auxiliary camera and projector. Optics and Lasers in Engineering, 175, 107984.
- Wang, J., Xu, Y., Li, Y., Khargonekar, P., & Faruque, M. A. A. (2026). CRUISE: Vision-Language Model-Guided Uncertainty-Aware Cross-Modal Sensor Fusion for Robust Autonomous Driving. arXiv preprint arXiv:2608.09202.
- Yang, Y., & Zhang, S. (2024). Pixelwise calibration method for a telecentric structured light system. Applied Optics, 63(10), 2562-2569.
- Cao, Y., Jiao, S., Sun, P., Peng, B., Shi, D., & Shi, Y. (2024). MEFusion: Unsupervised Mutual Enhancement for Multimodal Image Fusion. In ECAI 2024: 27th European Conference on Artificial Intelligence, 19–24 October 2024, Santiago de Compostela, Spain–Including 13th Conference on Prestigious Applications of Intelligent Systems (PAIS 2024) (pp. 553-560). 1 Oliver's Yard, 55 City Road, London, EC1Y 1SP: SAGE Publications Pvt. Ltd.
- Wang, H., Xu, Q., Wang, C., Xue, T., Peng, C., Chen, W., & Lin, F. (2026). Bad Seeing or Bad Thinking? Rewarding Perception for Multimodal Reasoning. arXiv preprint arXiv:2605.14054.
- Yang, Y., & Zhang, S. (2025). Pixel-wise calibration for a multi-focus microscopic 3D imaging system. Optics and Lasers in Engineering, 194, 109127.
- Wang, E., Fan, W., & Zhang, D. (2026). ADM-DP: Adaptive Dynamic Modality Diffusion Policy through Vision-Tactile-Graph Fusion for Multi-Agent Manipulation. arXiv preprint arXiv:2602.21622.
- Liu, C., Jing, T., Ma, C., Zhou, X., Lian, Z., Jin, Q., ... & Huang, S. S. (2026). EmoDiffTalk: Emotion-aware Diffusion for Editable 3D Gaussian Talking Head. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 18063-18073).
- Yang, Y., Bortins, I., Baldwin, D. P., & Zhang, S. (2024). Calibration of dual resolution dual camera structured light systems. Optics and Lasers in Engineering, 182, 108472.
- Li, Q., Luo, T., Jiang, M., Liao, J., & Jiang, Z. (2024, October). Deep incomplete multi-view network semi-supervised multi-label learning with unbiased loss. In Proceedings of the 32nd ACM International Conference on Multimedia (pp. 9048-9056).
- Li, Q., Luo, T., & Liao, J. (2025). Theory-inspired deep multi-view multi-label learning with incomplete views and noisy labels. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 20706-20715).
- Wang, H., Que, H., Xu, Q., Liu, M., Zhou, W., Feng, J., ... & Lin, F. (2026, April). Reverse-engineered reasoning for open-ended generation. In International Conference on Learning Representations (Vol. 2026, pp. 45563-45589).

- Du, Y., & Villarrubia-González, G. (2026). Lightweight Skin Lesion Segmentation for Edge Deployment: A Critical Review of Architectures, Accuracy Compensation, and Clinical Translation. IEEE Access.
- Chen, L., Yang, Y., & Zhang, S. (2024, September). Rapid autofocus method for digital fringe projection techniques. In *Interferometry and Structured Light 2024*(Vol. 13135, pp. 92-97). SPIE.
- Wang, M., Li, J., Li, Z., Luo, C., Chen, B., Xia, S. T., & Wang, Z. (2023, October). Unsupervised anomaly detection with local-sensitive VQVAE and global-sensitive transformers. In *2023 IEEE International Conference on Image Processing (ICIP)* (pp. 1080-1084). IEEE.
- Yu, C., Wang, H., Chen, J., Wang, Z., Deng, B., Hao, Z., ... & Song, Y. (2026, April). When Rules Fall Short: Agent-Driven Discovery of Emerging Content Issues in Short Video Platforms. In *Proceedings of the ACM Web Conference 2026* (pp. 8008-8016).
- Li, Z., Chen, Z., Qu, F., Wang, M., Zhao, Y., Zhang, K., & Peng, Y. (2024). Point resampling and ray transformation aid to editable NeRF models. *arXiv preprint arXiv:2405.07306*.
- Jiang, L., Cao, Z., Chen, W., Zheng, X., Lyu, Y., Li, Z., ... & Chen, Y. C. (2026). SAP: Segment Any 4K Panorama. *arXiv preprint arXiv:2603.12759*.
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In *Proceedings of MICCAI*.
- Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., & Zagoruyko, S. (2020). End-to-end object detection with transformers. In *Proceedings of ECCV*.
- Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., & Ng, R. (2020). NeRF: Representing scenes as neural radiance fields for view synthesis. In *Proceedings of ECCV*.
- Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., & LeCun, Y. (2014). OverFeat: Integrated recognition, localization and detection using convolutional networks. In *Proceedings of ICLR*.
- Sinha, D., Tong, Y., Zepeda, M. A. F., Hanstad, G., Nelson, E. C., Theisen, C. O., ... & Gamm, D. M. (2023). Silica nanocapsules as a nonviral delivery platform for iPSC-RPE. *Investigative Ophthalmology & Visual Science*, 64(8), 774-774.
- Zhang, Z., Zhang, Y., Zeng, A., Pan, D., Ji, Y., Zhang, Z., & Lin, J. (2023, September). Time-series data imputation via realistic masking-guided tri-attention bi-gru. In *ECAI 2023: 26th European Conference on Artificial Intelligence* (pp. 3074-3082).
- Bai, X., Li, Z., Xu, W., Lam, E. Y., & Peng, Y. (2026). ERF-GS: Reconstructing Fast Motion from Disjoint Event-RGB Viewpoints. *arXiv preprint arXiv:2608.08531*.