Geo-NVS-w: Geometry-Aware Novel View Synthesis In-the-Wild with an SDF Renderer

Angelos Kanlis[1*] Anastasios Tsalakopoulos[1*] Evangelos Chatzis[1] Antonis Karakottas[1]
Dimitrios Zarpalas[1]
[1]Centre for Research and Technology Hellas (CERTH), Thessaloniki, Greece
{tsalakop, a.kanlis, chatzise, ankarako, zarpalas}@iti.gr

Abstract

We introduce Geo-NVS-w, a geometry-aware framework for high-fidelity novel view synthesis from unstructured, in-the-wild image collections. While existing in-the-wild methods already excel at novel view synthesis, they often lack geometric grounding on complex surfaces, sometimes producing results that contain inconsistencies.

Geo-NVS-w addresses this limitation by leveraging an underlying geometric representation based on a Signed Distance Function (SDF) to guide the rendering process. This is complemented by a novel Geometry-Preservation Loss which ensures that fine structural details are preserved. Our framework achieves competitive rendering performance, while demonstrating a 4–5× reduction reduction in energy consumption compared to similar methods. We demonstrate that Geo-NVS-w is a robust method for in-the-wild NVS, yielding photorealistic results with sharp, geometrically coherent details.

1. Introduction

Novel view synthesis from unconstrained, in-the-wild image collections has emerged as a cornerstone problem in computer graphics and vision. The ability to explore a 3D scene by rendering photorealistic views from any arbitrary viewpoint has profound applications in virtual reality, digital heritage, and visual effects. However, in-the-wild datasets, such as the widely used IMC-Phototourism dataset [11] consisting of tourist photos of landmarks, present a formidable challenge. They are characterized by inconsistent illumination, varying camera settings, and transient occluders such as pedestrians, vehicles, or temporary structures.

Pioneering methods like Neural Radiance Fields [7] and its variants for in-the-wild data [1, 6] have made remarkable progress. However, they inherit a fundamental limitation of volumetric density representations, which is geometric ambiguity that tends to produce semi-transparent artifacts and blur sharp architectural details.

We argue that to achieve greater photorealism and consistency in NVS, the rendering process can benefit from being explicitly guided by the scene's underlying geometry. To this end, we introduce Geo-NVS-w, a framework that uses a high-fidelity Signed Distance Function (SDF) representation [13] as its geometric backbone. An accurate SDF provides a strong prior for surface locations, enabling our renderer to create sharp depth discontinuities and, consequently, sharper images.

Our core contributions are:

- **Octree-Accelerated feature-based volume**: We pair an SDF-guided renderer with an octree-based feature volume, interpolating features within geometry-bearing regions to preserve detail while accelerating training.
- **Geometry-Preservation Loss (GPL).** We introduce a novel loss that explicitly penalizes the model for incorrectly masking out geometrically significant regions as transient, ensuring textures and features remain coherent across viewpoints.
- **Quantified Energy Efficiency Analysis:** We instrument training with GPU power logging to quantify the energy–quality trade-off, documenting markedly lower training time and energy use.

Our approach is founded on the principle that an accurate and robust geometric representation is essential for high-quality view synthesis. By grounding our rendering in an SDF, we provide a strong inductive bias for surfaces, which directly contributes to the quality of our synthesized views, especially on complex scenes. Geo-NVS-w thus demonstrates that high-quality NVS can be achieved through the synergy of advanced rendering techniques and strong geometric grounding.

*Equal contribution

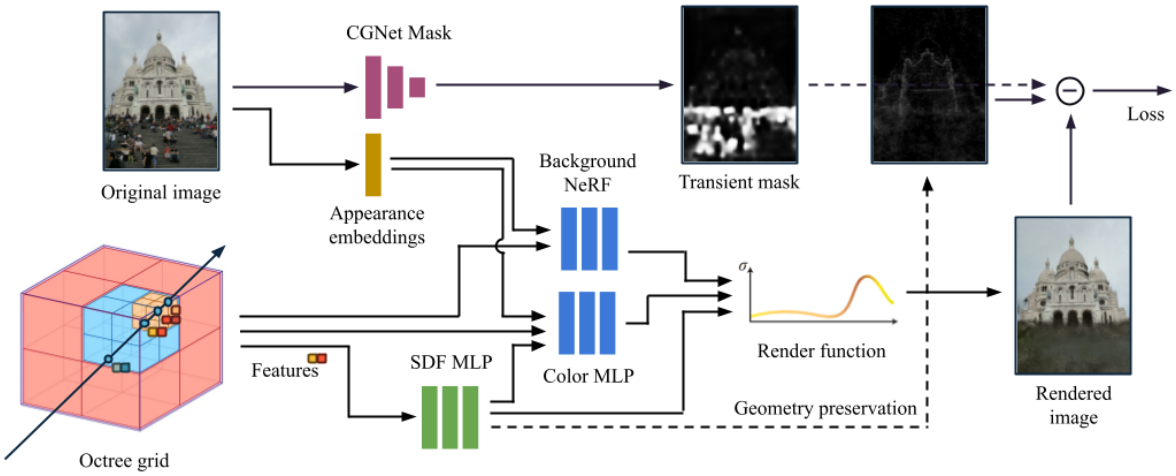

Figure 1. **Overview of the Geo-NVS-w Framework.** For a given camera ray, we march through an octree containing feature grids for the foreground (SDF-based) and background (NeRF-based). Within the foreground unit sphere, interpolated features are passed to MLPs to predict an SDF value $s(\mathbf{x})$ and a color $c(\mathbf{x}, \mathbf{d})$. We use the NeuS rendering formula to convert SDF values into alpha-compositing weights w_i, ensuring rendered colors are tightly coupled to the underlying surface. These weights are used to accumulate color, which is then composed with the background. Our Geometry-Preservation Loss (GPL) ensures the transient mask does not erode high-curvature details.

2. Related Work

Our work builds upon advances in novel view synthesis and neural surface representation.

2.1. Novel View Synthesis

Neural Radiance Fields (NeRF) [7] model a scene as a continuous mapping from 5D coordinates (position and viewing direction) to volumetric density and color, enabling photo-realistic rendering but requiring dense sampling and long training, and struggling with large-scale scenes. For faster rendering, PlenOctrees [17] bake the network into an explicit octree for real-time rendering. To handle unbounded scenes, NeRF++ [19] uses dual MLPs to separately model foreground and background. For in-the-wild data, NeRF-W [6] adds per-image appearance embeddings and a transient component for variable lighting and dynamics, while Ha-NeRF [1] couples an appearance-hallucinating CNN with a transient 2D mask MLP.

2.2. Neural Surface Representation

A parallel thread targets high-quality geometry via implicit representations. NeuS [13] introduces SDF volume rendering, ensuring the opacity-weighted color of a ray corresponds to the underlying surface. To scale, acceleration structures are crucial: Instant-NGP [8] employs multi-resolution hash grids for NeRFs, and Neuralangelo [12] adapts them to SDFs for high-fidelity, large-scale reconstruction, highlighting the importance of numerical gradients and coarse-to-fine optimization for stabilizing SDF training on hash grids.

Geo-NVS-w unifies these lines: we adopt NeuS-style SDF rendering for geometric fidelity and grid efficiency, analogous to Neuralangelo but using an octree feature grid. Unlike reconstruction-focused work, we use this pipeline as the backbone for high-quality novel view synthesis.

Geo-NVS-w is engineered primarily for high-fidelity novel view synthesis. Its architecture is built upon a robust geometric foundation—a Signed Distance Function (SDF)—which is key to rendering consistent views. The framework combines an efficient octree feature volume, an SDF-guided rendering process, and our novel Geometry-Preservation Loss (GPL). This geometric foundation avoids artifacts typical in NeRF renderings—such as density clouding from inconsistent ray sampling—by enforcing coherent surface geometry.

2.3. Octree Feature Volume

To balance performance and efficiency, we represent the scene using two separate, sparse octree feature grids: one for the foreground, modeled with an SDF, and a second for the background environment, modeled with a conventional NeRF. This dual-grid octree structure concentrates computational resources on regions containing geometry, pruning vast empty spaces. For any 3D point $\mathbf{x}$ inside the foreground's unit sphere, we query features via trilinear interpolation from the SDF grid. These features are then decoded by small MLPs to produce SDF and color values. For points outside this sphere, features are queried from the background grid and passed to a small background NeRF network.

2.4. SDF-guided Volumetric Rendering

Accurate surface localization is key for sharp rendering. An SDF defines the surface by the zero level set $s(\mathbf{x}) = 0$.

We use NeuS [13] to link SDFs to alpha compositing. For

a ray sample $\mathbf{x}_i$ with SDF $s_i = s(\mathbf{x}_i)$, the occupancy is

$$\alpha_i = \text{sigmoid}(\zeta s_i + \beta_i) - \text{sigmoid}(\zeta s_{i+1} + \beta_i), \quad (1)$$

where ζ is a learned global deviation and $\beta_i = \langle \nabla s_i, \mathbf{d} \rangle \Delta t_i / 2$ corrects for the angle between ray $\mathbf{d}$ and normal ∇s_i. A compact *Deviation Network* learns ζ, and normals/derivatives use finite differences [12]. Rendering weights follow discrete compositing:

$$w_i = T_i \alpha_i, \quad T_i = \prod_{j<i} (1 - \alpha_j), \quad (2)$$

and the final color $C(\mathbf{r})$ accumulates the foreground over the background NeRF. This concentrates color integration to a narrow band around the surface, producing geometrically consistent, sharp views.

2.5. Appearance and Transient Modules

A per-image latent appearance code is used to model variations in lighting and camera parameters, inspired by NeRF-W [6]. Regarding transient occluders, we follow CR-NeRF [15] in using an unsupervised segmentation-based approach, where a lightweight CGNet [14] is trained to generate transient object masks.

2.6. Geometry-Preservation Loss (GPL)

A standard challenge for in-the-wild NVS is disentangling static background from transient objects (e.g., people, cars). This is often handled with a transient mask network, which learns to down-weight pixels corresponding to dynamic elements. However, training multiple networks using a photometric loss often leads to the transient mask network removing static structure edges, leading to degradation of parts of the SDF representation, particularly those with complex geometry.

To counteract this, we introduce the **Geometry-Preservation Loss (GPL)**. The intuition is to penalize the transient mask for being active on rays that pass through geometrically significant regions. We define the loss as:

$$\mathcal{L}_{\text{GPL}} = \mathbb{E}_{\mathbf{r}} \left[M(\mathbf{r}) \cdot \Phi(\mathbf{r}) \right], \quad (3)$$

where $M(\mathbf{r}) \in [0, 1]$ is the predicted mask value for ray $\mathbf{r}$, and $\Phi(\mathbf{r})$ is an edge indicator function derived from the foreground SDF. This indicator is designed to be high for rays intersecting geometrically complex regions. For each ray, we compute the ray-wise average of the per-sample eikonal error, $E_{\text{grad}}(r) = (\|\nabla s\| - 1)^2$, and the absolute curvature estimate, $E_{\text{curv}}(r) = |\Delta s|$. The indicator function is then:

$$\Phi(\mathbf{r}) = \sigma(\lambda_{\text{grad}} E_{\text{grad}}(\mathbf{r}) + \lambda_{\text{curv}} E_{\text{curv}}(\mathbf{r})), \quad (4)$$

where σ is a calibrated sigmoid function. Intuitively, $\Phi(\mathbf{r})$ peaks on rays that traverse sharp edges or areas of high curvature. By multiplying the mask value with this indicator, our geometry preservation loss encourages the mask to remain zero (i.e., fully foreground) for rays that are crucial for defining the scene's structure. This is designed to preserve sharp features, leading to visibly sharper and more coherent novel views. We set the loss weights $\lambda_{\text{GPL}} = 0.05$, $\lambda_{\text{grad}} = 10.0$, and $\lambda_{\text{curv}} = 2.0$ in all our experiments.

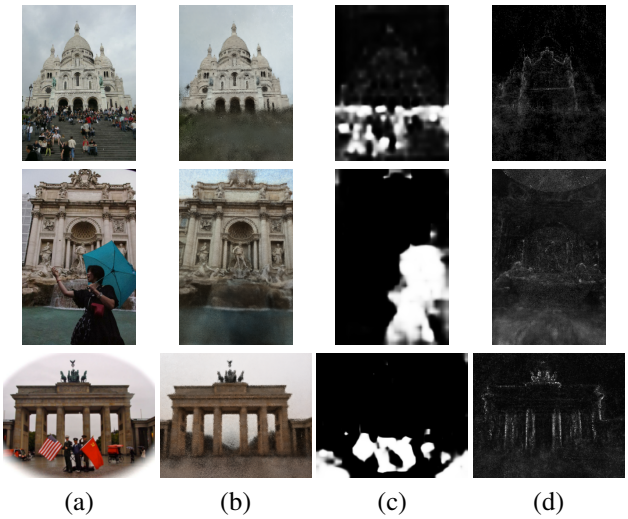

Figure 2. **Qualitative results on Phototourism scenes.** Top to bottom: Sacré-Cœur, Trevi Fountain, Brandenburg Gate. (a) Input image. (b) Rendered result with Geo-NVS-w from the same viewpoint and appearance embedding. (c) Estimated transiency mask. (d) Visualization of the geometry-preservation map (accumulated SDF gradients along each ray).

2.7. Learning Objective

The complete framework is trained end-to-end by minimizing a composite loss:

$$\begin{aligned} \mathcal{L} = \lambda_{\text{rgb}} \mathcal{L}_{\text{rgb}} + \lambda_{\text{eik}} \mathcal{L}_{\text{eik}} + \lambda_{\text{curv}} \mathcal{L}_{\text{curv}} \\ + \lambda_{\text{mask}} L_{\text{reg}} + \lambda_{\text{GPL}} \mathcal{L}_{\text{GPL}} + \lambda_{\text{lps}} \mathcal{L}_{\text{lps}}. \end{aligned} \quad (5)$$

The key components include: the primary photometric loss $\mathcal{L}_{\text{rgb}}$ (L1 difference); an eikonal loss $\mathcal{L}_{\text{eik}} = (\|\nabla s\| - 1)^2$; our novel $\mathcal{L}_{\text{GPL}}$; the Lipschitz loss $\mathcal{L}_{\text{lps}}$ [4], on the color network, following prior work [10]; a curvature loss $\mathcal{L}_{\text{curv}} = |\Delta s|$; and an off-surface penalty. The regularized transient mask, L_{reg}, combines the transient CNN mask with penalties for intensity and bimodality to encourage a clean separation of static and dynamic elements. During a warm-up phase, we also use a sphere-initialization loss, $\mathcal{L}_{\text{sphere}}$, to provide a stable initial geometry.

3. The Geo-NVS-w Framework

4. Experiments

We evaluate Geo-NVS-w on all four currently available scenes from the IMC-PT dataset[11] also used in NeRF-W and Ha-NeRF, as the rest have been removed due to data

Table 1. **Novel View Synthesis (NVS) quality.** We report PSNR (↑), SSIM (↑), and LPIPS (↓). Geo-NVS-w achieves strong results on all IMC-PT dataset scenes.

Method	PSNR ↑	SSIM ↑	LPIPS ↓
Brandenburg Gate			
Ha-NeRF	23.45	0.811	0.247
NeRF-W	23.98	0.915	0.198
Geo-NVS-w (Ours)	**25.4**	**0.944**	**0.158**
Sacré-Coeur			
Ha-NeRF	**25.40**	**0.877**	**0.124**
NeRF-W	25.11	0.859	0.141
Geo-NVS-w (Ours)	23.23	0.85	0.16
Trevi Fountain			
Ha-NeRF	22.15	0.695	**0.117**
NeRF-W	23.01	0.751	**0.109**
Geo-NVS-w (Ours)	**24.5**	**0.831**	0.203
Taj Mahal			
Ha-NeRF	22.72	0.767	0.301
NeRF-W	**25.15**	0.833	0.195
Geo-NVS-w (Ours)	24.19	**0.86**	**0.194**

inconsistencies [16]. We also view CR-NeRF [15] as complementary and plan to include a controlled comparison in future work.

4.1. Implementation Details

Our model is implemented in PyTorch [9] and trained on a single NVIDIA A10G GPU using mixed-precision computation to accelerate training. We measure energy consumption by incorporating GPU power measurement directly into our training pipeline, logging cumulative usage over time.

4.2. Novel View Synthesis Benchmarks

Our method matches or exceeds baselines [1, 6] on several metrics, as shown in Tab. 1, while achieving superior image quality metrics in most scenes. Fig. 3 corroborates these findings, illustrating that our method delivers enhanced visual fidelity at higher processing speed and reduced energy expenditure.

4.3. Energy Analysis

Beyond raw performance, practical usability and scalability depend on computational efficiency. As shown in Fig. 3, Geo-NVS-w is not only faster but also more energy-efficient. Our method completes a 300,000-iteration run using approximately 2.05 kWh, whereas NeRF-W consumes 9.7 kWh and Ha-NeRF consumes 7.71 kWh. Beyond achieving superior geometric fidelity, our SDF-based approach also demonstrates efficiency gains, reaching peak PSNR on our runs using approximately 2.05 kWh, a beneficial effect of an architecture that allows for significant downsizing of the MLPs.

5. Conclusion

We presented Geo-NVS-w, a framework advancing in-the-wild novel view synthesis by placing geometric preservation

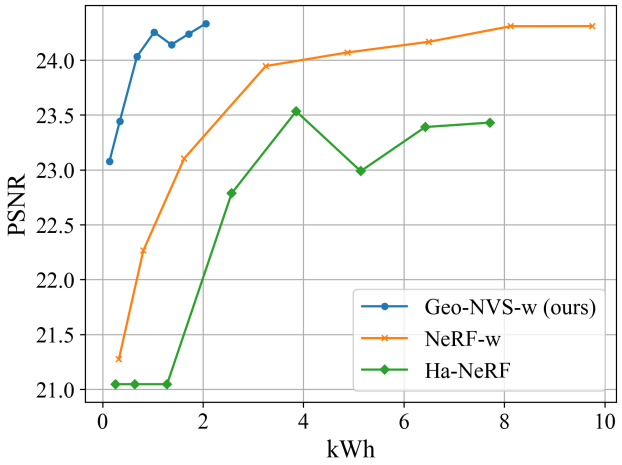

Figure 3. **Energy vs. quality trade-off.** Geo-NVS-w achieves high PSNR with significantly less training time and cumulative energy consumption (kWh) compared to baseline NeRF-W, making it a more efficient and scalable solution.

at the core of the rendering process. Incorporating a Signed Distance Function into our architecture and introducing a Geometry-Preservation Loss, our method produces renderings with consistent sharpness and geometrical coherence. The framework offers a strong, geometry-consistent alternative with favorable efficiency–quality trade-offs compared to prior approaches. While newer paradigms such as Gaussian Splatting have emerged, our findings confirm that geometry-aware methods remain effective for high-fidelity, in-the-wild view synthesis.

Future work could explore integrating 3D Gaussian Splatting techniques [2] within in-the-wild settings [3, 18], leveraging their speed advantages to further reduce rendering latency and computational overhead while maintaining the surface coherence enabled by our SDF methodology, as presented in [5].

6. Acknowledgments

This research has been supported by the European Commission funded program XReco, under Horizon Europe Grant Agreement 101070250. The computational resources were granted with the support of GRNET.

References

[1] Xiaoshuai Chen, Zizhang Cai, Jia Li, and Guofeng Zhang. HA-NeRF: Hallucinated neural radiance fields in the wild. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 12652–12661, 2022. 1, 2, 4

[2] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42(4):139:1–139:1, 2023. 4

[3] Jonas Kulhanek, Songyou Peng, Zuzana Kukelova, Marc Pollefeys, and Torsten Sattler. Wildgaussians: 3d gaussian splatting in the wild. *NeurIPS*, 2024. 4

[4] Hsueh-Ti Derek Liu, Francis Williams, Alec Jacobson, Sanja Fidler, and Or Litany. Learning smooth neural functions via lipschitz regularization. In *ACM SIGGRAPH 2022 Conference Proceedings*, pages 31:1–31:13. ACM, 2022. 3

[5] Jianheng Liu et al. GS-SDF: LiDAR-Augmented Gaussian Splatting and Neural SDF for Geometrically Consistent Rendering and Reconstruction. *arXiv preprint arXiv:2503.10170*, 2025. 4

[6] Ricardo Martin-Brualla, Noha Radwan, Mehdi S. M. Sajjadi, Jonathan T. Barron, Alexey Dosovitskiy, and Daniel Duckworth. NeRF-W: Neural radiance fields for unconstrained photo collections. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7210–7219, 2021. 1, 2, 3, 4

[7] 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. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 405–421, 2020. 1, 2

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

[9] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates Inc., 2019. 4

[10] Radu Alexandru Rosu and Sven Behnke. Permutosdf: Fast multi-view reconstruction with implicit surfaces using permutohedral lattices. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8466–8475, 2022. 3

[11] Noah Snavely, Steven M. Seitz, and Richard Szeliski. Photo tourism: Exploring photo collections in 3d. *ACM Transactions on Graphics*, 25(3):835–846, 2006. 1, 3

[12] Matthew Tancik, Jonathan T. Barron, Ben Mildenhall, Thomas Müller, Alex Evans, Ricardo Martin-Brualla, Pratul P. Srinivasan, Mehdi S. M. Sajjadi, Daniel Duckworth, and Tero Karras. Neuralangelo: High-fidelity neural surface reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8548–8558, 2023. 2, 3

[13] Peng Wang, Lingjie Liu, Yuan Liu, Christian Theobalt, Taku Komura, and Wenping Wang. NeuS: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 26998–27010, 2021. 1, 2

[14] Tianyi Wu, Sheng Tang, Rui Zhang, and Yongdong Zhang. Cgnet: A light-weight context guided network for semantic segmentation. *IEEE Transactions on Image Processing*, 30: 1169–1179, 2021. 3

[15] Yifan Yang, Shuhai Zhang, Zixiong Huang, Yubing Zhang, and Mingkui Tan. Cross-ray neural radiance fields for novel-view synthesis from unconstrained image collections, 2023. 3, 4

[16] Kwang Moo Yi. IMC-PT 2020 dataset, 2020. University of British Columbia. Accessed: 2025-09-26. 4

[17] Alex Yu, Ruilong Li, Matthew Tancik, Hao Li, Ren Ng, and Angjoo Kanazawa. PlenOctrees: For real-time rendering of neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 5752–5761, 2021. 2

[18] Dongbin Zhang, Chuming Wang, Weitao Wang, Peihao Li, Minghan Qin, and Haoqian Wang. Gaussian in the wild: 3d gaussian splatting for unconstrained image collections. *European Conference on Computer Vision (ECCV)*, 2024. 4

[19] Kai Zhang, Gernot Riegler, Noah Snavely, and Vladlen Koltun. Nerf++: Analyzing and improving neural radiance fields. *arXiv preprint arXiv:2010.07492*, 2020. 2