SpectralGaussians

Semantic, spectral 3D Gaussian splatting for multi-spectral scene representation, visualization and analysis

Sinha, Saptarshi Neil; Graf, Holger; Weinmann, Michael

DOI
10.1016/j.isprsjprs.2025.06.008

Publication date
2025

Document Version
Final published version

Published in
ISPRS Journal of Photogrammetry and Remote Sensing

ISPRS Journal of Photogrammetry and Remote Sensing 227 (2025) 789–803

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

SpectralGaussians: Semantic, spectral 3D Gaussian splatting for multi-spectral scene representation, visualization and analysis

Saptarshi Neil Sinha [a,*], Holger Graf [a], Michael Weinmann [b]

[a] Fraunhofer IGD, Fraunhoferstr. 5, Darmstadt, 64283, Germany
[b] Delft University of Technology, Van Mourik Broekmanweg 6, Delft, 2628 XE, Netherlands

ARTICLE INFO

Keywords:
Computer graphics
Deep learning
Spectral imaging
3D reconstruction
3D gaussian splatting
Appearance modeling
Scene understanding and editing
Novel view synthesis

ABSTRACT

We propose a novel cross-spectral rendering framework based on 3D Gaussian Splatting (3DGS) that generates realistic and semantically meaningful splats from registered multi-view spectrum and segmentation maps. This extension enhances the representation of scenes with multiple spectra, providing insights into the underlying materials and segmentation. We introduce an improved physically-based rendering approach for Gaussian splats, estimating reflectance and lights per spectra, thereby enhancing accuracy and realism. In a comprehensive quantitative and qualitative evaluation, we demonstrate the superior performance of our approach with respect to other recent learning-based spectral scene representation approaches (i.e., XNeRF and SpectralNeRF) as well as other non-spectral state-of-the-art learning-based approaches. Our work also demonstrates the potential of spectral scene understanding for precise scene editing techniques like style transfer, inpainting, and removal. Thereby, our contributions address challenges in multi-spectral scene representation, rendering, and editing, offering new possibilities for diverse applications.

1. Introduction

Accurate scene representation is an essential prerequisite for numerous applications. The way we perceive our surroundings in terms of a mixture of light gives us a particular scene understanding, thereby determining how we interact with our environment. However, representing scenes in terms of red, green and blue color channels suffers from both a bad reproduction of the scene's appearance due to metamerism effects and lacking characteristics only observable in certain of the spectral bands. Therefore, multi-spectral scene capture and representation has become of high relevance, where light and reflectance spectra are given with a higher resolution thereby surpassing the limitations of the broad-band RGB color model.

In domains such as architecture, automotive industries, advertisement, and design, accurate modeling of light transport and considering the full spectrum of light is crucial for virtual prototyping. Predictive rendering, which involves simulating the spectral transport of light, is necessary to assess and evaluate the visual quality of products before physical production. This ensures reliable assessment and enables color-correct scene reproduction. Furthermore, spectral information as captured by multi-spectral (MS) cameras (Micasense, 2024; Silios, 2024), infrared (IR) cameras (JENOPTIK, 2024), and UV sensors (Lanteri and Pelosi, 2021) extends scene understanding in terms of insights on underlying material characteristics and behavior (including anomalies, defects, etc.) revealed only in certain sub-ranges of the light spectrum which empowers experts and autonomous systems to gain valuable insights and make informed decisions in the respective scenarios. For *precision* farming applications, multispectral scene monitoring enables early detection and monitoring of harmful algal bloom in bodies of water (Kwon et al., 2023), facilitates the detection and classification of plant diseases (Moghadam et al., 2017; Jung et al., 2022) to allow farmers to maintain crop health, optimize agricultural practices, and conduct quantitative and qualitative analysis of agro products (Zhang et al., 2021c), and allows getting insights on precise and objective plant parameters through 3D vision and multi-spectral imaging via phenotyping sensors like PlantEye (Phenospex, 2024). In the context of cultural heritage, multi-spectral information is essential for gaining insights on production processes of artifacts or artworks and used materials, as e.g. relevant for the analysis of historical paintings (Alfeld et al., 2018; Landi and Maino, 2011; Grillini et al., 2024) or for revealing hidden or altered features withing documents (Qureshi et al., 2019), thereby also providing crucial hints on restoration of eroded parts by utilizing information from individual spectral bands that may exceed the visible range. Among the many further application scenarios where multi-spectral scene monitoring and representation also allows for a more comprehensive understanding are facial recognition systems (Vetrekar et al., 2016), medical sciences, forensic sciences

* Corresponding author.
E-mail addresses: saptarshi.neil.sinha@igd.fraunhofer.de (S.N. Sinha), holger.graf@igd.fraunhofer.de (H. Graf), M.Weinmann@tudelft.nl (M. Weinmann).

https://doi.org/10.1016/j.isprsjprs.2025.06.008
Received 12 August 2024; Received in revised form 8 June 2025; Accepted 12 June 2025
Available online 12 July 2025

and remote sensing (Zahra et al., 2024), where land cover and usage can be monitored more accurately.

Depending on the respective scenario, multi-spectral information can be either stored in terms of multi-channel image-like representations (Weinmann and Weinmann, 2017; Zhan et al., 2017; che, 2018; Chen et al., 2018; Palos Sánchez et al., 2019; Weinmann and Weinmann, 2019; Sun et al., 2020; Shang et al., 2020; Zhang and Chi, 2020; Du et al., 2021; Senchuri et al., 2021; Florath et al., 2022) (as typically used for airborne or satellite-based surveillance) or – in case of a more detailed geometric scene representation – in terms of multi-spectral surface reflectance characteristics directly parameterized on 3D point clouds (wei, 2019; Mitschke et al., 2022; Afifi et al., 2023; Rizaldy et al., 2023, 2024a,b) or meshes (Merzbach et al., 2017; Koutsoudis et al., 2021). Recently, learning-based scene representations such as neural radiance fields (NERFs) (Mildenhall et al., 2020) and 3D Gaussian Splatting (3DGS) (Kerbl et al., 2023) have gained a lot of attention due to their potential of capturing fine-scale details and allowing their reproduction for novel view synthesis.

Implicit scene representation using NeRFs (Mildenhall et al., 2020) has been demonstrated to allow high-fidelity scene representation based on training a neural network to predict view-dependent color and view-independent density information for points in the scene volume and leveraging volume rendering to predict the scene's appearance for particular viewpoints, while optimizing the network to produce images that match the original input images. This means that scene geometry is not directly available but instead we rely on the inference of density information for densely-sampled points in the scene volume from which, in turn, scene geometry can be inferred. However, involving a neural network that way introduces a high computational burden. This has severe implications on scene editing applications like inpainting, style transfer, or object manipulation, preventing real-time performance. When densely storing multi-spectral characteristics instead of RGB-only data, this problem gets further amplified. This computational challenge has been addressed by the more recently introduced 3D Gaussian Splatting (3DGS) (Kerbl et al., 2023) which has been demonstrated to allow superior performance and quality compared to NERF-based scene representation and visualization at less computational burden. This explicit scene representation replaces the neural network used in NeRF approaches with a set of Gaussians and the number and arrangement of Gaussians is optimized to best match the input data. Thereby, the representation results in improved rendering efficiency, while also offering interpretability in contrast to black-box neural network representations.

Among the many extensions of NeRFs, only a few have explored extensions towards spectral scene representations. XNeRF (Poggi et al., 2022) and SpectralNeRF (Li et al., 2024b), despite their advancements in handling spectral scene representations, do not include reflectance and lighting estimation, segmentation of the spectral scene, and explicit geometry. These limitations can impact the accuracy, relightability, and comprehensive understanding of spectral scenes. Furthermore, the extension of 3DGS towards spectral scene representation and visualization has not been investigated so far, despite their potential for impactful extensions. Advanced 3DGS methods (Ye et al., 2023; Qin et al., 2023) go beyond appearance and geometry modeling by supporting complex and fine-grained scene understanding. They exceed the capabilities of NeRF-based approaches, like Semantic-NeRF (Zhi et al., 2021), which incorporate semantic information into radiance fields for 3D scene modeling. However, these methods struggle to generalize to complex scenarios. Distilled Feature Fields (Kobayashi et al., 2022) and LERF (Liu et al., 2023b) explore distilling 2D features to aid in real-world 3D semantics, but they have limitations in accurate segmentation and cannot match the segmentation quality and efficiency of Gaussian-based methods (Ye et al., 2023; Qin et al., 2023). We extend the semantic scene understanding per spectrum which has not been leveraged so far in learning-based multi-view scene representation methods allowing us to have enhanced per spectrum scene understanding.

In this paper, we present a novel concept of sparse spectral scene representation based on Gaussian Splatting that allows efficient multi-spectral scene representation and visualization. To provide a trade-off between computational demands and more accurate scene representation, we extend Gaussian Splatting in terms of a material-wise scene segmentation based on the captured multi-spectral information as well as the segmentwise spectral appearance representation based on a reflectance model, thereby allowing the separate representation of materialwise reflectance behavior in contrast to a view-dependent scene representation with baked-in lighting conditions as in the original spherical-harmonics-based Gaussian Splatting approach. The segmentwise storage and processing present the potential for constructing a scene graph that enables conditional loading of various primitives based on semantic information. This capability allows for selective loading of scene graph nodes and facilitates out-of-core rendering. By rendering only the enabled segments and nodes of the scene graph, this offers significant potential for improved performance which is not addressed in this paper. The cross-spectral scene understanding using Gaussian Splatting offers potential for scene analysis tasks such as depicting disease areas in plants, restoring statues or paintings, or the visualization of tissues for better diagnosis, etc.. Furthermore, due to the rather low number of spectra considered in the datasets used for benchmarking (Li et al., 2024b; Poggi et al., 2022) we consider the representation of all spectra in this work. However, concepts for reducing the data redundancy as especially contained in hyper-spectral data could be combined with our approach due to the orthogonality of such tasks, e.g. by combining feature selection strategies with the local clustering into segments of similar attributes. Due to the lacking availability of respective datasets, we leave this for future work.

In summary, the key contributions offered in this work are:

- We present a novel cross-spectral rendering framework that extends the scene representation based on 3D Gaussian Splatting (3DGS) to generate realistic and semantically meaningful splats from registered multi-view spectrum and segmentation maps.
- We present an improved physically-based rendering approach for Gaussian splats, estimating reflectance and lights per spectra, which enhances the accuracy and realism of the rendered output by considering the unique characteristics of different spectra, resulting in visually convincing and physically accurate scene representations.
- We generated two synthetic spectral datasets by extending the shiny Blender dataset (Verbin et al., 2022) and the synthetic NERF dataset (Mildenhall et al., 2020) in terms of their spectral properties. The datasets were created through simulations using Mitsuba (Jakob et al., 2022), where scenes were rendered at various wavelengths across the visible spectrum. These datasets are expected to serve as valuable resources for researchers and practitioners, offering a diverse range of spectral scenes for experimentation, evaluation, and advancements in the field of image-based/multi-view spectral rendering. We plan to release both the dataset and the code in our website to generate similar datasets using Mitsuba (Jakob et al., 2022), promoting reproducibility and further contributions to the field.
- In the scope of a detailed evaluation on our datasets, as well as the spectral NeRF dataset (Li et al., 2024b), we showcase the potential of our approach in spectral scene understanding. Through our evaluation, we demonstrate that spectral scene understanding enables efficient and accurate scene editing techniques, including style transfer, in-painting, and removal. These techniques leverage the specific spectral characteristics of objects in the scene, facilitating more precise and context-aware modifications.

2. Related work

2.1. Learning-based scene representation

In recent years, significant advancements have been made in generating photo-realistic novel views through the use of novel learning-based scene representations combined with volume rendering techniques. Neural Radiance Fields (NeRF) (Mildenhall et al., 2020; Tewari et al., 2022) represent the scene based on a neural network that predicts local density and view-dependent color for points in the scene volume. This information can then be used to synthesize images of the scene using volume rendering techniques. The network representing the scene is trained by minimizing the deviation of the predicted images to their respective given input images under the respective view conditions, thereby exploiting the observation that an accurate scene representation by the network leads to an accurate image synthesis. The remarkable potential of the NeRF approach for novel view synthesis has given rise to several notable extensions. Researchers have focused on improving rendering quality by addressing issues such as aliasing (Barron et al., 2021; Wang et al., 2022; Barron et al., 2022, 2023), as well as accelerating network training (Reiser et al., 2021; Fridovich-Keil et al., 2022; Müller et al., 2022; Chen et al., 2022b; Yariv et al., 2023). Furthermore, there have been efforts to handle more complex inputs, including unconstrained image collections (Martin-Brualla et al., 2021; Chen et al., 2022c; Jun-Seong et al., 2022), image collections requiring the refinement or complete estimation of camera pose parameters (Wang et al., 2021b; Yen-Chen et al., 2021; Lin et al., 2021; Jeong et al., 2021), deformable scenes (Park et al., 2021; Pumarola et al., 2021) and large-scale scenarios (Tancik et al., 2022; Turki et al., 2022; Mi and Xu, 2023). Further works aimed at guiding the training and handling textureless regions by incorporating depth cues (Wei et al., 2021; Deng et al., 2022; Roessle et al., 2022; Rematas et al., 2022; Attal et al., 2021).

Despite the great success of NeRFs for novel view synthesis applications, the neural network lacks interpretability and the extraction of surface information requires network evaluations on a dense grid and a subsequent derivation of surface information from the volumetric density information based on techniques like Marching Cubes (Lorensen and Cline, 1998), which limits real-time applications. Therefore, further works focused on representing scenes in terms of implicit surfaces (Wang et al., 2021a, 2023a; Ge et al., 2023), explicit representations using points (Xu et al., 2022), meshes (Munkberg et al., 2022), and 3D Gaussians (Kerbl et al., 2023). Point-based neural rendering techniques, such as Point-NeRF (Xu et al., 2022), merge precise view synthesis from NeRF with the fast scene reconstruction abilities of deep multi-view stereo methods. These techniques employ neural 3D point clouds to enable efficient rendering, thereby facilitating accelerated training processes. Furthermore, a recent approach (Zhang et al., 2023) has shown that point-based methods are well-suited for scene editing purposes. Recently, 3D Gaussian Splatting (Kerbl et al., 2023) has been introduced as the state-of-the-art, learning-based scene representation based on optimized Gaussians for novel view synthesis, surpassing existing implicit neural representation methods such as NeRFs in terms of both quality and efficiency. This approach utilizes anisotropic 3D Gaussians as an explicit scene representation and employs a fast tile-based differentiable rasterizer for image rendering. Similar to the developments for NeRF approaches (Li et al., 2024d), Gaussian Splatting has been extended in terms of speeding up rendering and training time (Zhang et al., 2025a; LI et al., 2025; Park et al., 2025; Peng et al., 2025; Bouzidi et al., 2025; Zheng et al., 2025; Zhang et al., 2025b; Li et al., 2024a; Yang et al., 2024b), geometry-aware scene representation (Deng et al., 2025a; Zhang et al., 2025c; Wu et al., 2025a; Zheng et al., 2025; Zhang et al., 2025b; Huang et al., 2024), allowing the handling of dynamic scenes (Song et al., 2025; Yuan et al., 2025; Gao et al., 2025; Zheng et al., 2025; Zhang et al., 2025b; Li et al., 2024c) and large-scale scenes (Deng et al., 2025b; Ong et al., 2025; Yang et al., 2024a; Wu et al., 2025b), model compression (Chen et al., 2025; Yuan et al., 2025; Deng et al., 2025b; Zheng et al., 2025; Gao et al., 2025), decoupling reflectance and illumination effects (Skorokhodov et al., 2025; Han and Wu, 2025; Yong et al., 2025; Park et al., 2025), adding semantic information (Li et al., 2025; Maggio and Carlone, 2025; Zhang et al., 2025b; Park et al., 2025) and estimating or refining camera parameters (Li et al., 2024a; Bortolon et al., 2024; Sun et al., 2024).

However, extending these novel scene representations to the spectral domain beyond RGB channels remains an open challenge, with only a few seminal works addressing this so far. Spectral variants of NERF, such as xNERF (Poggi et al., 2022) for cross-spectral spectrum-maps and SpectralNeRF (Li et al., 2024b) for multi-spectral spectrum-maps, have shown effectiveness in generating novel views across different spectral domains. The cross-spectral splats generated by our approach can be visualized via an interactive spectral viewer (Sinha et al., 2024) based on Viser (Tancik et al., 2023). Besides view synthesis, the viewer allows to visualize splats, even with spectral characteristics, as well as visualizing residuals between different versions of splats such as splats from different iterations during training or comparing differences between splats in different spectral ranges. Furthermore, the user study conducted in their work (Sinha et al., 2024) validates the effectiveness and practicality of the reconstructed 3D splats derived from the spectrum maps, confirming their utility in spectral visualization and analysis. However, the framework of reconstructing a spectral Gaussian Splatting scene representation is a novel contribution in this paper and has not been considered in their work (Sinha et al., 2024).

2.2. Radiance based appearance capture

Instead of focusing on the pure reproduction of a scene according to the original NeRF formulation without explicitly modeling reflectance and illumination characteristics, several NeRF extensions focused on modeling reflectance by separating visual appearance into lighting and material properties. Respective approaches have the capability to jointly predict environmental illumination and surface reflectance properties even in the presence of unknown or varying lighting conditions (Bi et al., 2020; Zhang et al., 2021a; Boss et al., 2021a; Srinivasan et al., 2021; Boss et al., 2021b; Yariv et al., 2021; Zhang et al., 2021b).

One notable contribution is Ref-NeRF (Verbin et al., 2022), which introduces a novel parameterization and structuring of view-dependent outgoing radiance, along with a regularizer on normal vectors. This enhances the accuracy in predicting reflectance properties. To address the challenge of learning geometry from highly specular surfaces, recent works (Liu et al., 2023a; Liang et al., 2023, 2022) have utilized SDF-based representations. This enables more precise estimation of surface normals for physically based rendering. Further work (Van Holland et al., 2025) achieved the detection of mirroring surfaces based on photometric inconsistencies observed in the result of a NeRF approach with additional reprojection loss, which, in turn, were then used to jointly optimize the radiance field and mirror geometry in a second refinement step. However, these methods suffer from time-consuming optimization and slow rendering speed, limiting their practical application in real-world scenarios. Furthermore, NVDiffRec (Hasselgren et al., 2022) is an explicit representation method that directly optimizes triangle meshes with materials and environment map lighting, enabling real-time interactive applications, unlike MLP-based methods that tend to be slower.

Relightable Gaussians (Gao et al., 2023) presents a differentiable point-based rendering framework for material and lighting decomposition from multi-view images, enabling real-time relighting and editing of 3D point clouds. It surpasses existing material estimation approaches and offers improved results. GaussianShader (Jiang et al., 2023) is another method that enhances neural rendering in scenes with reflective surfaces by applying a simplified shading function on 3D Gaussians. It addresses the challenge of accurate normal estimation on discrete 3D Gaussians, achieving a balance between efficiency and rendering quality. Our shading model is inspired by this method where we use the model without the residual color in the reflectance estimation.

2.3. Sparse spectral scene understanding

Gaussian splatting based semantic segmentation frameworks, such as Gaussian Grouping (Ye et al., 2023) and LangSplat (Qin et al., 2023), have successfully utilized foundation models like Segment Anything (Kirillov et al., 2023) to segment scenes. LangSplat is a 3D language field that enables precise and efficient open-vocabulary querying within 3D spaces by representing language features using a collection of 3D Gaussians distilled from CLIP (Radford et al., 2021). Gaussian Grouping extends Gaussian Splatting by incorporating object-level scene understanding and introducing Identity Encodings to reconstruct and segment objects in real-world 3D scenes. We utilized this method for accurate semantic segmentation of spectral scenes. Segmenting the scene per spectra provides valuable information about regions that are visible in specific spectral ranges, enabling us to obtain finer details that can be leveraged in various domains such as cultural heritage (Alfeld et al., 2018; Landi and Maino, 2011; Grillini et al., 2024), smart farming (Kwon et al., 2023; Zhang et al., 2021c; Jung et al., 2022), document analysis (Qureshi et al., 2019), face recognition (Vetrekar et al., 2016), and other fields. This spectral segmentation approach offers insights and solutions for diverse applications in these domains. In the scope of the evaluation, we demonstrate that spectral scene understanding enables efficient and accurate scene editing techniques, including style transfer, in-painting, and removal.

2.4. Spectral renderers

Spectral rendering engines such as ART (The ART development team, 2018), PBRT v3 (Pharr et al., 2016), and Mitsuba (Jakob, 2010) are commonly utilized by the scientific community. While CPU-based renderers are more prevalent, there is a growing trend of GPU-based spectral renderers that leverage GPU acceleration. Some examples of GPU-based spectral renderers include Mitsuba 2 (Nimier-David et al., 2019), PBRT v4 (Pharr, 2020), and Malia (Dufay et al., 2019). These renderers play a crucial role in simulating real-world spectral data and are gaining recognition in the field. To achieve computational efficiency in deep learning and focus on relevant spectral information, we adopt a sparse spectral rendering approach using multi-view spectrum maps. This technique enables faster computations by reducing unimportant spectral data while preserving the necessary information for realistic rendering of spectral scenes. By leveraging spectrum maps from multiple viewpoints, high-quality spectral renderings are generated with a reduced computational cost compared to full-resolution spectral rendering methods.

3. Background

The human eye is sensitive to only a certain range in the electromagnetic spectrum (for wavelengths between about 380 nm and 780 nm) which varies between subjects. The response curve of the human eye is to the red, green and blue wavelengths were determined using color matching functions which has been standardized by CIE in 1932 (Wyszecki and Stiles, 2000). Given a spectral power distribution $L(\lambda)$, its corresponding CIE tristimulus values X, Y and Z can be computed by convolution of the $L(\lambda)$ with the appropriate color matching functions $f_X(\lambda)$, $f_Y(\lambda)$, $f_Z(\lambda)$ as represented in the following equations (Devlin et al., 2002):

$$\begin{cases} X = \int_{380}^{780} f_X(\lambda) L(\lambda) d\lambda \\ Y = \int_{380}^{780} f_Y(\lambda) L(\lambda) d\lambda \\ Z = \int_{380}^{780} f_Z(\lambda) L(\lambda) d\lambda \end{cases} \tag{1}$$

The spectral power distribution $L(\lambda)$ at a point x for incoming wavelength λ_i and outgoing wavelength λ_o can be computed as follows:

$$L(x, \omega_i, \omega_o, \lambda_i, \lambda_o) = \int_{\Omega} f_r(x, \omega_i, \omega_o, \lambda_i, \lambda_o) L_i(x, \omega_i, \omega_o, \lambda_i) \cos\theta \, d\omega_i \tag{2}$$

where Ω represents the hemisphere above a surface point x, f_r is the bidirectional reflectance function, L_i is the incoming radiance coming from incident direction ω_i and ω_o is the direction of the outgoing radiance.

The final RGB image is obtained based on the conversion from the XYZ color space to the sRGB space which involves the following steps.

- **Conversion to linear RGB**: This step involves using a matrix multiplication to convert XYZ values to linear RGB values.

$$\begin{pmatrix} R \\ G \\ B \end{pmatrix} = \left(M^l\right) \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \tag{3}$$

  There are many methods (Smits, 2000) to convert XYZ to linear RGB and the value of the matrix M^l depends on it.
- **Gamma correction**: Linear RGB values are gamma-corrected to get sRGB values. This involves applying a power function with a specific gamma value (≈ 2.2).
- **Clipping**: All RGB values are clipped within the range [0, 1].

The above steps can be combined to get the final transformation matrix (M^c) directly get the sRGB values:

$$\begin{pmatrix} R \\ G \\ B \end{pmatrix} = \left(M^c\right) \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \tag{4}$$

Based on Eqs. (1), (2), and (4), the RGB values per spectra maps (Li et al., 2024b) can be computed according to

$$\begin{pmatrix} R_\lambda \\ G_\lambda \\ B_\lambda \end{pmatrix} = \begin{pmatrix} M^c_{11} f_X(\lambda) + M^c_{12} f_Y(\lambda) + M^c_{13} f_Z(\lambda) \\ M^c_{21} f_X(\lambda) + M^c_{22} f_Y(\lambda) + M^c_{23} f_Z(\lambda) \\ M^c_{31} f_X(\lambda) + M^c_{32} f_Y(\lambda) + M^c_{33} f_Z(\lambda) \end{pmatrix} L(\lambda) \Delta\lambda \tag{5}$$

4. Methodology

In this section, we first provide an overview on our Spectral Gaussian splatting approach for addressing efficient and accurate multi-spectral scene representation from input images depicting the scene characteristics in multiple wavelengths based on dedicated extensions. This is followed by a brief description of the involved appearance modeling from the multi-spectral input data as well as the involved scene segmentation based on the spectral characteristics that we use to allow a more compact per-spectrum grouping of similar Gaussians, thereby enabling a reasonable representation of multi-spectral scene characteristics. Finally, we discuss the combination of these parts into the proposed Spectral Gaussian splatting framework.

4.1. Spectral Gaussian splatting

We propose an end-to-end spectral Gaussian splatting approach that enables physically-based rendering, relighting, and semantic segmentation of a scene. Our method is built upon the Gaussian splatting architecture (Kerbl et al., 2023) and leverages the Gaussian shader (Jiang et al., 2023) for the accurate estimation of BRDF parameters and illumination. By employing Gaussian grouping (Ye et al., 2023), we effectively group 3D Gaussian splats with similar semantic information. Our framework excels in generating full spectra rendering and conveniently initializes common features from other spectra trained to a specific iteration, ensuring improved reconstruction of splats. In Fig. 1, we showcase our proposed spectral Gaussian splatting framework, which uses a Spectral Gaussian model to predict BRDF parameters, distilled feature fields, and light per spectrum from multi-view spectrum-maps. Our method combines segmentation, appearance modeling, and sparse spectral scene representation in an end-to-end manner. Thereby it enhances BRDF estimation by incorporating spectral information. The framework has applications in material recognition, spectral analysis, reflectance estimation, segmentation, illumination correction, and inpainting.

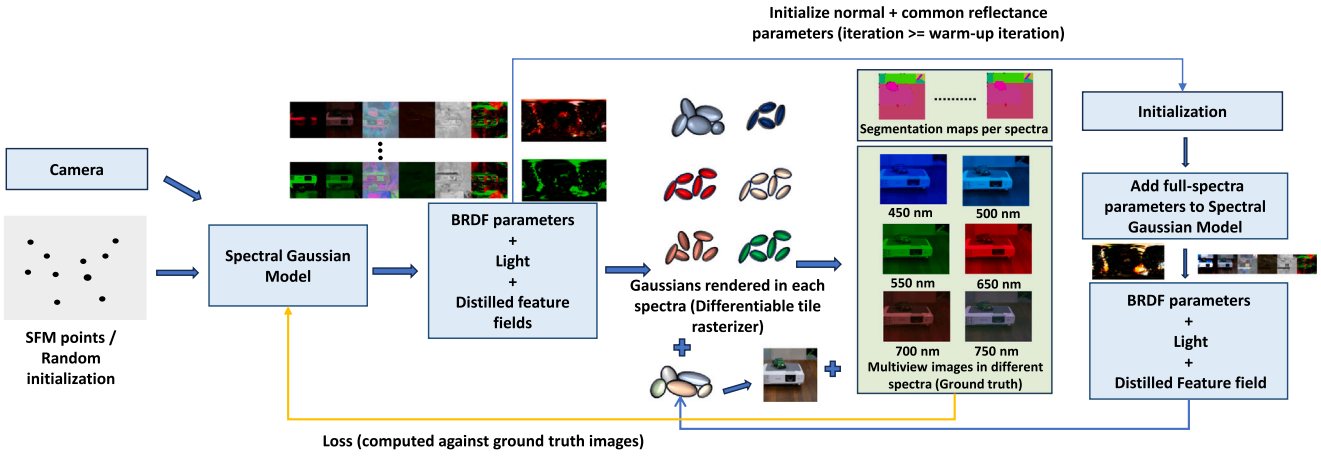

Fig. 1. The proposed spectral Gaussian splatting framework: Spectral Gaussian model predicting BRDF parameters, distilled feature fields, and light per spectrum from multi-view spectrum-maps. The full-spectra maps and learnable parameters are introduced later in the training process by initializing them with priors from all other spectra.

In the following subsections, we provide further details regarding the spectral model, covering topics such as appearance modeling, spectral semantic scene representation, spectral scene editing, and the seamless integration of these aspects into the 3DGS framework.

4.2. Spectral appearance modeling

In order to support material editing and re-lighting, we use an enhanced representation of appearance by replacing the spherical harmonic co-efficients by a shading function, which incorporates diffuse color, roughness, specular tint and normal information and a differentiable environment light map to model direct lighting similar to the Gaussian shader (Jiang et al., 2023).

Thereby, the rendered color per spectrum of a 3D Gaussian can be computed by considering its diffuse color, specular tint, direct specular light, normal vector and roughness according to

$$c(\omega_o)_\lambda = \gamma\left(c_{d_\lambda} + s_\lambda \odot L_{s_\lambda}(\omega_o, n, \rho_\lambda)\right) \tag{6}$$

where $c(\omega_o)_\lambda$ represents the rendered color per spectrum for the viewing direction ω_o. The function γ is a gamma tone mapping function that adjusts the color values for display purposes in terms of converting the wide range of brightness levels of the perceived light to a lower dynamic range representation that fits within the display's capabilities.. $c_{d_\lambda} \in [0,1]^3$ denotes the diffuse color of the 3D Gaussian, specifying the color appearance under diffuse lighting per spectrum. $s_\lambda \in [0,1]^3$ is the specular tint on the sphere, indicating the color of the specular highlights per spectrum. $L_{s_\lambda}(\omega_o, n, \rho_\lambda)$ describes the direct specular light for the 3D Gaussian in the viewing direction ω_o per spectrum, considering the surface normal n and roughness ρ_λ. n is the normal vector indicating the surface orientation, and $\rho_\lambda \in [0,1]$ represents the surface smoothness or roughness per spectrum.

To allow representing a wide range of appearance characteristics, we consider a grouping of similar scene characteristics and the fitting of per-segment di-chromatic shading models consisting of the following components:

- We use a diffuse reflectance component (c_{d_λ}) to represent the view-independent appearance of Gaussians.
- To account for most of the view-dependent reflections in rendering, we use a specular reflectance term $s_\lambda \odot L_{s_\lambda}(\omega_o, \mathbf{n}, \rho_\lambda)$to describe the interaction between the intrinsic surface color s_λ (specular tint) and the direct specular light L_{s_λ}.

To compute the contribution of the specular light perceived at a surface point with surface normal $\mathbf{n}$ from the direction ω_o and roughness ρ_λ, we have to combine the incoming light from different directions weighted by the respective shading model $D(\mathbf{r})$ for the specular component (where we use the specular GGX Normal Distribution Function D (Walter et al., 2007) in this work) according to the integration

$$L_{s_\lambda}(\omega_o, \mathbf{n}, \rho_\lambda) = \int_\Omega L(\omega_i) D(\mathbf{r}, \rho_\lambda)(\omega_i \cdot \mathbf{n}) d\omega_i \tag{7}$$

Here, Ω represents the whole upper hemi-sphere, ω_i is the direction for the input radiance, and D characterizes the specular lobe (effective integral range). The reflective direction $\mathbf{r}$ is calculated using the view direction ω_o and the surface normal $\mathbf{n}$ as $\mathbf{r} = 2(\omega_o \cdot \mathbf{n})\ \mathbf{n} - \omega_o$. L_{s_λ} represents the direct specular light per spectral band λ. The appearance model employed for training our network can be seamlessly substituted with alternative parametric shading models that incorporate different scene reflectance parameters. Consequently, the model parameters within the training pipeline must be appropriately updated to reflect these changes. Although, applying various parametric appearance models for different scenes is possible, rendering different objects/segments within a scene with different shading models is an avenue for future research. Instead, we only consider increasing the compactness of the Gaussian splatting representation by fitting a shading model as described before to groups of Gaussians with similar characteristics. The respective segmentation will be discussed in the following section.

4.3. Spectral semantic scene representation

Per-spectrum segmentation maps serve multiple purposes in various applications. They enable sparse scene representation, allowing for detailed identification of specific regions of interest and the detection of attributes like material composition or texture under a more compact scene representation. These maps are beneficial for tasks like inpainting and statue restoration, where spectral information is crucial for accurate and realistic results. Additionally, per-spectrum segmentation maps aid in anomaly detection by analyzing the spectral properties of different regions and identifying deviations from expected patterns. This approach of segmenting different spectra enables the identification of specific regions of interest, such as the detection of gray mould disease in strawberry plants (Jung et al., 2022). Overall, these maps provide valuable insights into the scene, allowing for more robust and precise image processing and analysis. Our framework utilizes the Gaussian grouping method (Ye et al., 2023) to generate per-spectrum segmentation of the splats. This ensures consistent mask identities across different views of the scene and groups 3D Gaussian splats with the same semantic information. To create ground truth multi-view segmentation maps for each spectrum, we employ the Segment Anything Model (SAM) (Kirillov et al., 2023) along with a zero-shot tracker (Cheng et al., 2023). This combination automatically generates

masks for each image in the multi-view collection per spectrum, ensuring that each 2D mask corresponds to a unique identity in the 3D scene. By associating masks of the same identity across different views, we can determine the total number of objects present in the 3D scene.

In addition to the existing appearance and lighting properties, we assign a novel attribute that we denote as *Identity Encoding* to each spectral Gaussian, similar to Gaussian grouping (Ye et al., 2023). The Identity Encoding is a compact and learnable vector (of length 16) that effectively distinguishes different objects or parts within the scene. During training, similar to using Spherical Harmonic coefficients to represent color, the method optimizes the Identity Encoding vector to represent the instance ID of the scene. Unlike view-dependent appearance modeling, the instance ID remains consistent across different rendering views, as only the direct-current component of the Identity Encoding is generated by setting the Spherical Harmonic degree to 0.

The final rendered 2D mask identity feature, denoted as E_{id_λ}, for each pixel per spectrum λ is calculated by taking a weighted sum over the Identity Encoding (e_{i_λ}) of each Gaussian per spectrum. The weights are determined by the influence factor α'_{i_λ} of the respective Gaussian on that pixel per spectrum. Mathematically, this can be expressed as

$$E_{id_\lambda} = \sum_{i \in N} e_{i_\lambda} \alpha'_{i_\lambda} \prod_{j=1}^{i-1} (1 - \alpha'_{j_\lambda}) \tag{8}$$

where N represents the total number of Gaussians.

To group the 3D Gaussians based on their object mask identities, a grouping loss L_{id_λ} is computed per spectra. This loss has two components , i.e. it can be formulated as

$$L_{id_\lambda} = L_{2d_\lambda} + L_{3d_\lambda} \tag{9}$$

where the first component L_{2d_λ} denotes the 2D Identity Loss, which involves a softmax function to classify the rendered 2D features E_{id} (see Eq. (8)) into $K_s + 1$ categories, representing the total number of masks per spectrum in the 3D scene. The standard cross-entropy loss $\mathcal{L}_{2d_\lambda}$ for the classification of $K_s + 1$ categories is applied to allow indirect 2D classification. So given the rendered 2D features E_{id_λ} as input, a linear layer is first applied f to restore its feature dimension back to K:

$$f(E_{id_\lambda}) = W \cdot E_{id_\lambda} + b, \tag{10}$$

where W represents the learnable weight matrix and b is the bias term.

To obtain the probabilities for each category, we apply the softmax function:

$$\text{softmax}(f(E_{id_\lambda})) = \frac{\exp(f(E_{id_\lambda}))}{\sum_{i=1}^{K} \exp(f(E_{id_\lambda}))}, \tag{11}$$

For the identity classification task with K categories per spectrum λ, we utilize the standard cross-entropy loss:

$$L_{2d_\lambda} = -\sum_{i=1}^{K} y_{i_\lambda} \log(\text{softmax}(f(E_{id_\lambda}))), \tag{12}$$

where y is the ground truth label for each category.

The second component is the 3D Regularization Loss $\mathcal{L}_{3d_s}$, which capitalizes on the 3D spatial consistency to regulate the learning process of the Identity Encoding e_{i_λ} per spectrum λ. This loss ensures that the Identity Encodings of the top k-nearest 3D Gaussians are similar in terms of their feature distance, thereby promoting spatially consistent grouping. The 3D grouping loss per spectrum λ and sampled m points is computed as:

$$\mathcal{L}_{3d_\lambda} = \frac{1}{m} \sum_{j=1}^{m} D_{kl}(P \parallel Q) = \frac{1}{mk} \sum_{j=1}^{m} \sum_{i=1}^{k} F(e_{j_\lambda}) \log\left(\frac{F(e_{j_\lambda})}{F(e'_{i_\lambda})}\right) \tag{13}$$

Here, P contains the sampled Identity Encoding e_λ of a 3D Gaussian, and $Q = e'_{1_\lambda}, e'_{2_\lambda}, \ldots, e'_{k_\lambda}$ represents its k nearest neighbors in 3D Euclidean space.

Fig. 2. Spectral scene editing: The segmented scene at 450 nm (middle) is used to perform a semantic style-transfer on the full spectra (left). The semantic stylized scene (right) has been generated using by applying a style transfer on the multi-view maps (full-spectra) and then in-painting the splats using the semantic object-ID in spectrum 450 nm.

4.4. Combined spectral model (involving semantics and appearance)

Combined with the original 3D Gaussian loss (Kerbl et al., 2023) (we use γ instead of λ as we use λ to denote the spectral bands) on image rendering (we use the appearance model as explained in the Section 4.2 instead of spherical harmonics), the total loss per spectra $\mathcal{L}_{render_s}$ for fully end-to-end training is given by

$$\mathcal{L}_{\text{render}_\lambda} = (1 - \gamma) L_{1_\lambda} + \gamma \cdot \mathcal{L}_{\text{D-SSIM}_\lambda} + \gamma_{2d_\lambda} \mathcal{L}_{2d_\lambda} + \gamma_{3d} \mathcal{L}_{3d_\lambda} \tag{14}$$

The total loss is given by

$$\mathcal{L}_{\text{render}_{total}} = \sum_{\lambda=1}^{n_\lambda} \mathcal{L}_{\text{render}_\lambda} \tag{15}$$

where n_λ is the total number of spectral bands.

To enhance the optimization process and improve robustness, the model is initially trained for a specific warm-up iteration (1000 iterations) without incorporating the full-spectra spectrum maps. Following this, the estimated BRDF parameters and normals per 3D Gaussian for the full-spectra are initialized (see Fig. 1) using the average values from all other spectra, and this initialization step is integrated into the training process. By including these adequate priors, the optimization of parameters is guided more effectively, leading to better outcomes as demonstrated in the quantitative and qualitative analysis.

4.5. Spectral scene editing

Our framework extends scene editing techniques, such as Gaussian Grouping (Ye et al., 2023), into the spectral domain, unlocking a wide range of possibilities. By leveraging the semantic information present in any of the spectrum maps, we can achieve object deletion, in-painting, and style-transfer. Fig. 2 illustrates the utilization of segmentation maps obtained from the 450 nm spectrum for the stylization of the splats across the full spectra. To accomplish this, we transfer the style to the multi-view full spectra maps and perform object in-painting through a fine-tuning of the splats, similar to Gaussian grouping (Ye et al., 2023), using the new ground truth (multi-view semantic stylized maps). The significance of this capability is particularly evident in fields like cultural heritage, where the retrieval of color information from a specific spectral band enables the accurate restoration of missing color details throughout the full-spectrum. By leveraging these advancements, we can enhance various applications and open up new avenues for exploration. To demonstrate scene editing capabilities in real-world settings, we apply style transfer to the scene considered in the Freiburg forest dataset (Valada et al., 2016). Specifically, we use a marble stone material as material to replace the style of a gravel road, thereby achieving a semantic stylization effect (see Fig. 3). The segments generated in a multi-spectral setup more accurately will be able to delineate the road in comparison to the purely RGB-based segmentation in challenging lighting conditions, thereby enabling more precise scene editing.

Table 1
Quantitative comparisons (PSNR/SSIM/LPIPS) on spectral NeRF **synthetic** dataset (Li et al., 2024b).

Method	Spectral NeRF Synthetic Dataset (Li et al., 2024b)						Average
	kitchen	Living room	Digger	Spaceship	Vintage car	Cartoon knight	
	PSNR ↑						
NeRF (Mildenhall et al., 2020)	34.583	33.172	30.658	30.126	33.478	34.485	32.400
Mip-NeRF (Barron et al., 2021)	-	-	33.301	31.495	33.883	35.102	33.945
Aug-NeRF (Chen et al., 2022a)	34.480	32.540	31.538	30.929	33.639	33.908	32.677
SpectralNeRF (Li et al., 2024b)	35.115	33.665	33.378	31.951	34.480	34.915	33.610
Ours	37.035	37.989	40.218	41.233	42.636	36.723	38.456
	SSIM ↑						
NeRF (Mildenhall et al., 2020)	0.8943	0.9929	0.9187	0.9358	0.7958	0.9273	0.9123
Mip-NeRF (Barron et al., 2021)	-	-	0.9290	0.9475	0.8166	0.9572	0.9126
Aug-NeRF (Chen et al., 2022a)	0.9026	0.9649	0.9248	0.9402	0.8002	0.9287	0.9163
SpectralNeRF (Li et al., 2024b)	0.9117	0.9931	0.9357	0.9482	0.8169	0.9573	0.9349
Ours	0.9747	0.9733	0.9923	0.9951	0.9893	0.9572	0.9801
	LPIPS ↓						
NeRF (Mildenhall et al., 2020)	0.1650	0.0578	0.0413	0.0275	0.1319	0.1545	0.0722
Mip-NeRF (Barron et al., 2021)	-	-	0.0435	0.0535	0.1747	0.1526	0.1061
Aug-NeRF (Chen et al., 2022a)	0.1603	0.0706	0.0341	0.0389	0.1536	0.1705	0.0973
SpectralNeRF (Li et al., 2024b)	0.1637	0.0479	0.0259	0.0250	0.1499	0.1510	0.0733
Ours	0.0739	0.0525	0.0109	0.0084	0.0527	0.0741	0.0438

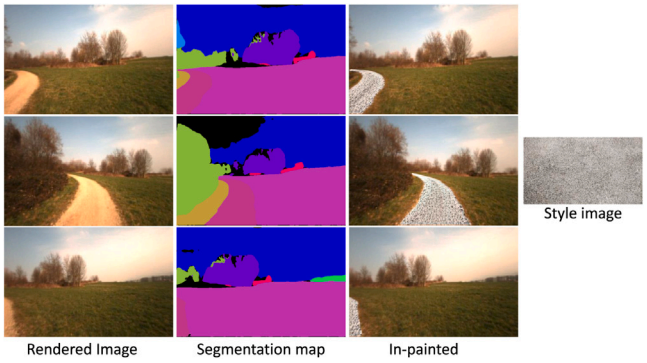

Fig. 3. Scene editing on the Freiburg forest dataset (Valada et al., 2016) by adding marble style to the gravel road using segments generated in a multi-spectral manner.

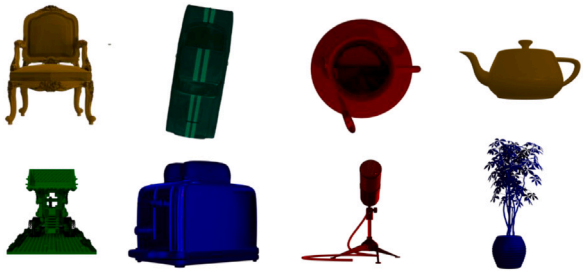

Fig. 4. Snapshot of the different scenes in the Spectral NeRF synthetic and Spectral shiny Blender datasets.

Table 2
Dataset overview.

Dataset	Scenes	Number of multi-view images	Number of iterations	Number of spectral bands
SpectralNeRF	6 synthetic and 2 real-world (MS)[a]	20 (Digger, Spaceship, Vintage car), 40 (cartoon knight) and 120 (all other scenes)	40,000 (Digger, Spaceship, Vintage car), 30,000 (all other scenes)	5 (Synthetic) and 8 (Real)
CrossSpectralNeRF	16 real-world (MS + IR)[a]	30 – 32	30,000	10 (MS) and 1(IR)
Spectral ShinyBlender	5 synthetic (MS)[a]	120	30,000	5
Spectral SyntheticNeRF	4 Synthetic (MS)[a]	120	30,000	5

[a] MS = Multispectral, IR = Infrared.

5. Experiments

To demonstrate the potential of our approach, we provide both quantitative and qualitative evaluations with comparisons to baseline techniques.

5.1. Baseline techniques used for comparison

The techniques used as a reference in the scope of the evaluation include several state-of-the-art variants of Neural Radiance Fields (NeRF) (i.e., NeRF (Mildenhall et al., 2020), MIP-NeRF (Barron et al., 2021), Aug-NeRF (Chen et al., 2022a), Ref-NeRF (Verbin et al., 2022)) (which considers appearance parameters) and Gaussian splatting (i.e., Gaussian splatting without special reflectance modeling (Kerbl et al., 2023) and Gaussian Shader that specifically models reflectance (Jiang et al., 2023)) as well as the respective extensions of such modern scene representation approaches to the spectral domain (i.e., SpectralNeRF (Li et al., 2024b) and Cross-spectral NeRF (Poggi et al., 2022)).

5.2. Datasets

For the comparison with SpectralNeRF, we use both synthetic and real-world multi-spectral videos (Li et al., 2024b). The poses for the digger, spaceship, and vintage car models were estimated using DUSt3R (Wang et al., 2023b) since reconstruction failed with COLMAP (Schönberger and Frahm, 2016). For the remaining scene videos (kitchen, living room, projector, and dragon doll), COLMAP was used to generate the poses.

To demonstrate the adaptability of our method in handling cross-spectral data (infrared and multi-spectral), we conducted a comparative analysis using the cross-spectral NeRF dataset (Poggi et al., 2022). We created the ground truth full spectrum image from the cross-spectral spectrum-maps. For this, we averaged the images from all spectra and applied the colormaps *viridis* and *magma* for the multi-spectral and infrared dataset respectively, similar to the approach used in cross-spectral NeRF (Poggi et al., 2022). To further validate that the spectral appearance estimation produces plausible results for different types of scenes (having also highly-reflective objects in the scene), we created a synthetic multi-spectral dataset from the shiny blender dataset (Verbin et al., 2022) and synthetic NeRF dataset (Mildenhall et al., 2020) (see Fig. 4). We generated this multi-spectral dataset using Mitsuba (Jakob et al., 2022) for 5 bands from 460 nm to 620 nm similar to SpectralNeRF (Li et al., 2024b). We generated the data for the scenes where the shading model supported in Mitsuba corresponded to the shading model in Blender in order to get representative data. We utilized this dataset to conduct a comparative analysis of our method against state-of-the-art NeRF and Gaussian splatting techniques. The results are presented in Table 7 and Table 6.

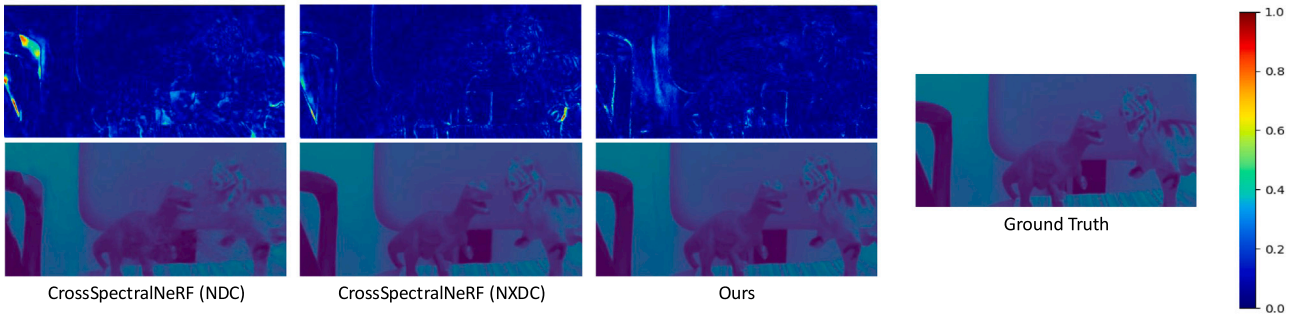

Fig. 5. Qualitative comparison of CrossSpectralNerf (Poggi et al., 2022) with our method with the dino dataset.

Table 3
Quantitative comparisons (PSNR/SSIM/LPIPS) on spectral NeRF **real** dataset (Li et al., 2024b).

Method	Spectral NeRF **real** Dataset	
	Projector	Dog doll
	PSNR ↑	
NeRF	28.9670	22.5040
Aug-NeRF	30.0795	21.2073
SpectralNeRF	31.2535	25.1257
Ours	34.2659	32.5517
	SSIM ↑	
NeRF	0.9429	0.8740
Aug-NeRF	0.9573	0.8915
SpectralNeRF	0.9449	0.8932
Ours	0.9746	0.97245
	LPIPS ↓	
NeRF	0.0472	0.1319
Aug-NeRF	0.0354	0.1168
SpectralNeRF	0.0605	0.14501
Ours	0.0709	0.07650

Table 4
Quantitative comparison (PSNR/SSIM) with the cross-spectral NeRF dataset (Poggi et al., 2022).

Configuration				Avg.	
Model	Train	NXDC	Test	PSNR	SSIM
NeRF	MS	-	MS	33.53	0.917
X-NeRF	RGB+MS	×	MS	31.96	0.897
X-NeRF	RGB+MS	✓	MS	33.87	0.918
NeRF	MS	-	MS	33.53	0.917
X-NeRF	RGB+MS+IR	×	MS	30.87	0.870
X-NeRF	RGB+MS+IR	✓	MS	33.53	0.914
Ours	RGB+MS	-	MS	35.17	0.962
NeRF	IR	-	IR	33.26	0.897
X-NeRF	RGB+MS+IR	×	IR	31.60	0.869
X-NeRF	RGB+MS+IR	✓	IR	32.44	0.879
Ours	RGB+IR	-	IR	33.19	0.952

Table 5
Quantitative comparisons (PSNR/SSIM/LPIPS) on the real-world Freiburg forest dataset (Valada et al., 2016). (a) without initialization from other spectra (b) with initialization from other spectra (with warm-up iteration of 1000).

Method	PSNR ↑	SSIM ↑	LPIPS ↓
Gaussian Splatting (Kerbl et al., 2023)	38.07	0.977	0.090
Ours (a)	40.17	0.974	0.086
Ours (b)	39.33	0.981	0.075

5.3. Implementation details

The evaluations were conducted on an Nvidia RTX 3090 graphics card. In most scenes, we used a total of 30,000 iterations, except for the digger, spaceship, and vintage car scenes where we used 40,000 iterations. For the comparison to other methods, we used the results reported in their original publications. In addition, we set the weights for the different terms in the loss formulation in Eq. (14) to $\gamma = 0.2, \gamma_{2d_\lambda} = 1.0$ and $\gamma_{3d_\lambda} = 2.0$ respectively.. These values were chosen heuristically, also taking into account their suitability for different scenarios. The computational costs depend on the resolution of the images. In case of hardware limitations or to improve the processing speed, the input images could be downsampled. With high-end consumer GPUs, it should however already be possible to infer the respective scene representation from high-resolution images. Furthermore, we already improved the efficiency by training a specific spectra in an iteration rather than setting the gradient of all the spectra to true.

5.4. Quantitative analysis

Quantitative analysis was performed on all datasets mentioned in Section 5.2 and overview of the number of scenes, multi-view images and number of iterations for which each scene was trained is presented in Table 2. We compute the PSNR (Fardo et al., 2016), SSIM (Nilsson and Akenine-Möller, 2020) and LPIPS (Zhang et al., 2018) for all camera-views and report average the average result. The orange in the tables represents the best result and yellow represents the second best results.

5.4.1. Evaluation on challenging real-world dataset

To demonstrate that our method effectively supports handling challenging real-world datasets, we also trained our model and conducted quantitative analyses on the Freiburg Forest dataset (Valada et al., 2016). This dataset, collected using the Viona autonomous mobile robot, features multi-spectral and multi-modal images, along with manually annotated pixel-wise ground truth segmentation masks for six classes, recorded under varying lighting conditions to enhance data variability. Table 5 demonstrates that our model achieves superior results under challenging lighting conditions when incorporating priors from other spectra during training. The LPIPS metric shows significant improvement, indicating enhanced perceptual reconstruction in difficult real-world static scenes.

5.4.2. Comparison with radiance-field-based spectral methods

The quantitative analysis shows that our method overall outperforms the existing spectral methods (Poggi et al., 2022; Li et al., 2024b) for both multi-spectral and cross-spectral data. The results presented in Table 1 indicate that our method outperforms SpectralNeRF in most scenes and on average for the synthetic dataset. Additionally, our analysis, as shown in Table 3, reveals that our method also surpasses SpectralNeRF when applied to the real-world dataset. It is important to note that due to the unavailability of all datasets and test views from the original paper, our evaluation was limited to only one real-world dataset (see Table 3) for the SpectralNeRF method. However, we also compare our method based on the Cross-spectral NeRF dataset which contains only real-world scenes. Here, our method clearly performs better for all scenes (multi-spectral and infrared datasets) as presented in Table 4. This shows that our method produces plausible results with real-world scenes and outperforms state-of-the-art spectral methods.

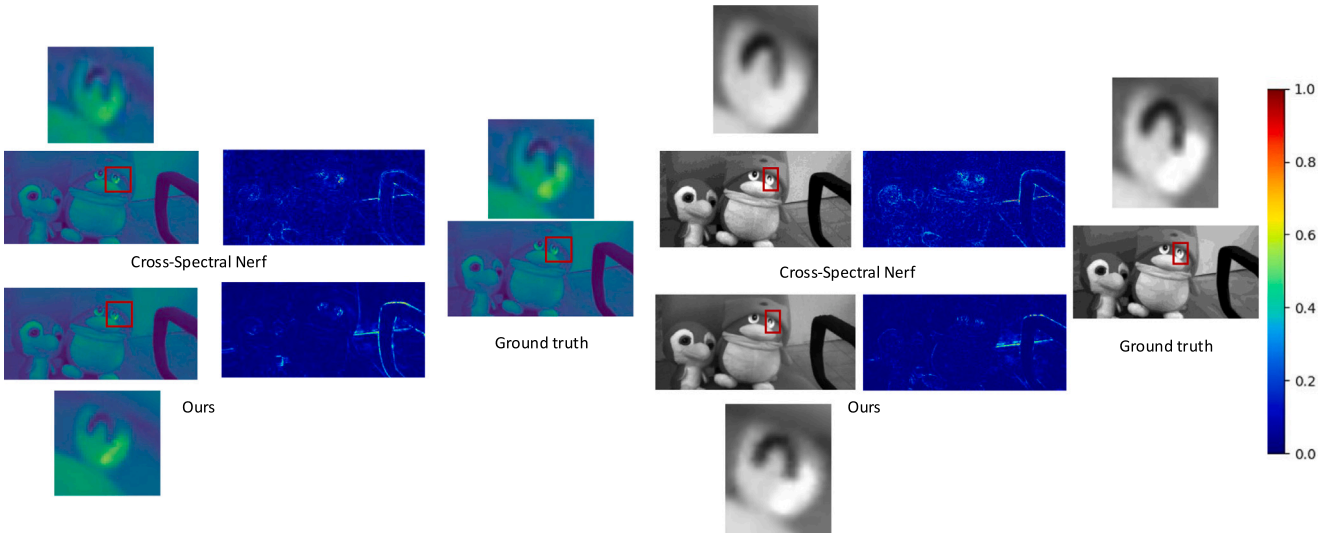

Fig. 6. Qualitative comparison of CrossSpectralNerf (Poggi et al., 2022) with our method with the Penguin dataset. The comparison shows the average of the 10 spectra colored with colormap viridis (left) and one such spectra (right). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 6
Quantitative comparisons (PSNR/SSIM/LPIPS) on spectral shiny blender dataset.

Method	Spectral Shiny Blender Dataset					
	Car	Helmet	Teapot	Toaster	Coffee	Avg.
			PSNR ↑			
NVDiffRec (Munkberg et al., 2022)	27.98	26.97	40.44	24.31	30.74	28.70
NVDiffMC (Hasselgren et al., 2022)	25.93	26.27	38.44	22.18	29.60	28.88
Ref-NeRF (Verbin et al., 2022)	30.41	29.92	45.19	25.29	33.99	32.32
NeRO (Liu et al., 2023a)	25.53	29.20	38.70	26.46	28.89	29.84
ENVIDR (Liang et al., 2023)	28.46	32.73	41.59	26.11	29.48	32.88
Guassian Splatting (Kerbl et al., 2023)	27.24	28.32	45.68	20.99	32.32	30.37
Gaussian shader (Jiang et al., 2023)	27.90	28.32	45.86	26.21	32.39	31.94
Ours	30.37	36.39	44.42	24.82	36.62	34.524
			SSIM ↑			
NVDiffRec (Munkberg et al., 2022)	0.963	0.951	0.996	0.928	0.973	0.945
NVDiffMC (Hasselgren et al., 2022)	0.940	0.940	0.995	0.886	0.965	0.944
Ref-NeRF (Verbin et al., 2022)	0.949	0.955	0.995	0.910	0.972	0.956
NeRO (Liu et al., 2023a)	0.949	0.971	0.995	0.929	0.956	0.962
ENVIDR (Liang et al., 2023)	0.961	0.980	0.996	0.939	0.949	0.969
Guassian Splatting (Kerbl et al., 2023)	0.930	0.951	0.996	0.895	0.971	0.947
Gaussian shader (Jiang et al., 2023)	0.931	0.950	0.996	0.929	0.971	0.957
Ours	0.970	0.970	0.992	0.942	0.973	0.969
			LPIPS ↓			
NVDiffRec (Munkberg et al., 2022)	0.045	0.118	0.011	0.169	0.076	0.119
NVDiffMC (Hasselgren et al., 2022)	0.077	0.157	0.014	0.225	0.097	0.147
Ref-NeRF (Verbin et al., 2022)	0.051	0.087	0.013	0.118	0.082	0.109
NeRO (Liu et al., 2023a)	0.074	0.050	0.012	0.089	0.110	0.072
ENVIDR (Liang et al., 2023)	0.049	0.051	0.011	0.116	0.139	0.072
Guassian Splatting (Kerbl et al., 2023)	0.047	0.079	0.007	0.126	0.078	0.083
Gaussian Shader (Jiang et al., 2023)	0.045	0.076	0.007	0.079	0.078	0.068
Ours	0.049	0.043	0.026	0.079	0.068	0.053

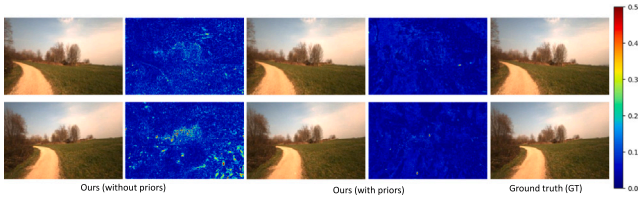

Fig. 7. Qualitative comparison of results (difference map of the renderings with the ground-truth images) on the Freiburg forest dataset (Valada et al., 2016), using priors from other spectra, with a warm-up iteration of 1000 and without adding priors.

5.4.3. Comparison with non-spectral radiance-based methods

To demonstrate that our method produces plausible results compared to existing state-of-the-art Gaussian splatting methods (Jiang et al., 2023; Kerbl et al., 2023), we conducted a comparison using spectral datasets created from both the NeRF synthetic dataset and the shiny Blender dataset, as described in Section 5.2. The analysis reveals that our method consistently outperforms existing methods on average for the Shiny Blender dataset, as shown in Table 6. This indicates that extending Gaussian splatting to the spectral domain improves the accuracy of reflectance estimation, particularly for shiny objects. Additionally, our method performs quite well on the synthetic NeRF dataset, as evidenced by the average PSNR and SSIM values in Table 7.

5.5. Qualitative analysis

We conducted a qualitative comparison between our method and the Cross-spectral NeRF (Poggi et al., 2022) using the dino and penguin datasets. The results, shown in Fig. 5 for the dino dataset and Fig. 6 for the penguin dataset, highlight the superior performance of our method in reconstructing scene appearance. In particular, Fig. 6 demonstrates the accurate rendering of specular effects in the eyes of the penguin, showcasing the effectiveness of our approach. Additionally, Fig. 5 reveals that our method produces better reflectance reconstruction, as evidenced by the shading effects on the surface of the dino. We also

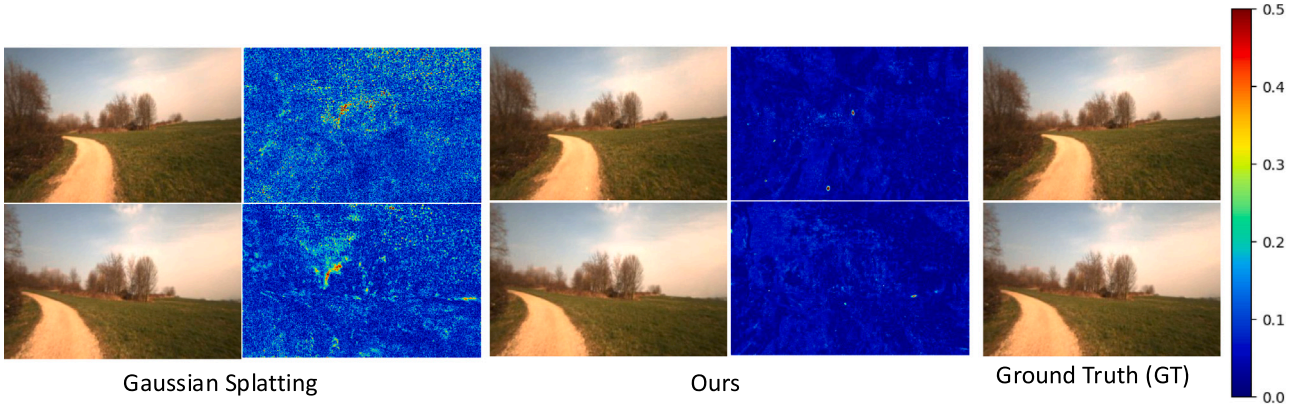

Fig. 8. Qualitative comparison of results (difference map of the renderings with the ground-truth images) on the Freiburg forest dataset (Valada et al., 2016) : In comparison to the original Gaussian splatting (Kerbl et al., 2023) our approach allows a significant reduction of the deviations from the ground truth.

Table 7
Quantitative comparisons (PSNR/SSIM/LPIPS) on spectral synthetic NeRF dataset.

Method	Spectral Synthetic NeRF Dataset				
	Chair	Lego	Mic	Ficus	Avg.
			PSNR ↑		
NeRF (Mildenhall et al., 2020)	33.00	32.54	32.91	30.13	32.64
VolSDF (Yariv et al., 2021)	30.57	29.46	30.53	22.91	28.87
Ref-NeRF (Verbin et al., 2022)	33.98	35.10	33.65	28.74	32.11
ENVIDR (Liang et al., 2023)	31.22	29.55	32.17	26.60	29.88
Gaussian Splatting (Kerbl et al., 2023)	35.82	35.69	35.34	34.83	35.17
Gaussian Shader (Jiang et al., 2023)	35.83	35.87	35.23	34.97	35.22
Ours	38.93	34.26	36.80	36.57	36.39
			SSIM ↑		
NeRF (Mildenhall et al., 2020)	0.967	0.961	0.980	0.964	0.968
VolSDF (Yariv et al., 2021)	0.949	0.951	0.969	0.929	0.949
Ref-NeRF (Verbin et al., 2022)	0.974	0.975	0.983	0.954	0.971
ENVIDR (Liang et al., 2023)	0.976	0.961	0.984	0.987	0.977
Gaussian Splatting (Kerbl et al., 2023)	0.987	0.983	0.991	0.987	0.987
Gaussian Shader (Jiang et al., 2023)	0.987	0.983	0.991	0.985	0.986
Ours	0.990	0.977	0.990	0.994	0.987
			LPIPS ↓		
NeRF (Mildenhall et al., 2020)	0.046	0.050	0.028	0.044	0.042
VolSDF (Yariv et al., 2021)	0.056	0.054	0.191	0.068	0.092
Ref-NeRF (Verbin et al., 2022)	0.029	0.025	0.018	0.056	0.032
ENVIDR (Liang et al., 2023)	0.031	0.054	0.021	0.010	0.029
Gaussian Splatting (Kerbl et al., 2023)	0.012	0.016	0.006	0.012	0.012
Gaussian Shader (Jiang et al., 2023)	0.012	0.014	0.006	0.013	0.011
Ours	0.017	0.031	0.014	0.006	0.017

Fig. 9. Spectral renderings and rendered segments of the Freiburg forest dataset (Valada et al., 2016) with enhanced vegetation index (EVI) and near-infrared (NIR) components for a specific camera viewpoint.

analyze our method qualitatively using a real-world spectral dataset in the wild (Freiburg forest dataset (Valada et al., 2016)). Our results, as shown in Fig. 8, outperform the original Gaussian splatting (Kerbl et al., 2023), as indicated by the difference maps compared to the ground truth. Additionally, Fig. 7 demonstrates that integrating priors from other spectra significantly improves the final renderings, as reflected in the difference maps across various camera views. We also present spectral renderings of this dataset, along with rendered segments from a specific camera viewpoint, using segment-anything (Kirillov et al., 2023) predictions on the multi-view images as ground truth, as shown in Fig. 9.

As depicted in Fig. 10, our framework successfully estimates the lighting and BRDF parameters within the individual spectra, while also providing segmented object IDs. This showcases the effectiveness and accuracy of our framework in capturing and analyzing the desired parameters for the given scene.

5.6. Ablation study

In this section, we conduct ablations by eliminating the warm-up iterations that we introduced to enhance reflectance and light estimations in the scene through the inclusion of appropriate priors from other spectra. For this, we use three real-world scenes: dragon doll (from the SpectralNeRF dataset (Li et al., 2024b)), orange, and tech scenes (from the Cross-SpectralNeRF dataset (Poggi et al., 2022)). The dragon doll scene has 8 bands, while the orange and tech scenes have 10 bands (see Table 8)

The ablation studies in Tables 9 and 10 evaluate the influence of integrating spectral information with material estimation. The results suggest that incorporating spectral priors positively affects the final appearance, as evidenced by improvements in PSNR and SSIM metrics, while LPIPS values indicate varying performance across the datasets.

To evaluate the impact of including priors from different spectra, we also conducted a qualitative analysis (see Fig. 11), after initializing the common model parameters with the average of all other spectra following a warm-up phase of 1000 iterations.

Qualitative analysis. We conducted a qualitative assessment by comparing the rendered outputs with the ground truth for the aforementioned scenes. The results reveal noticeable improvements in capturing finer

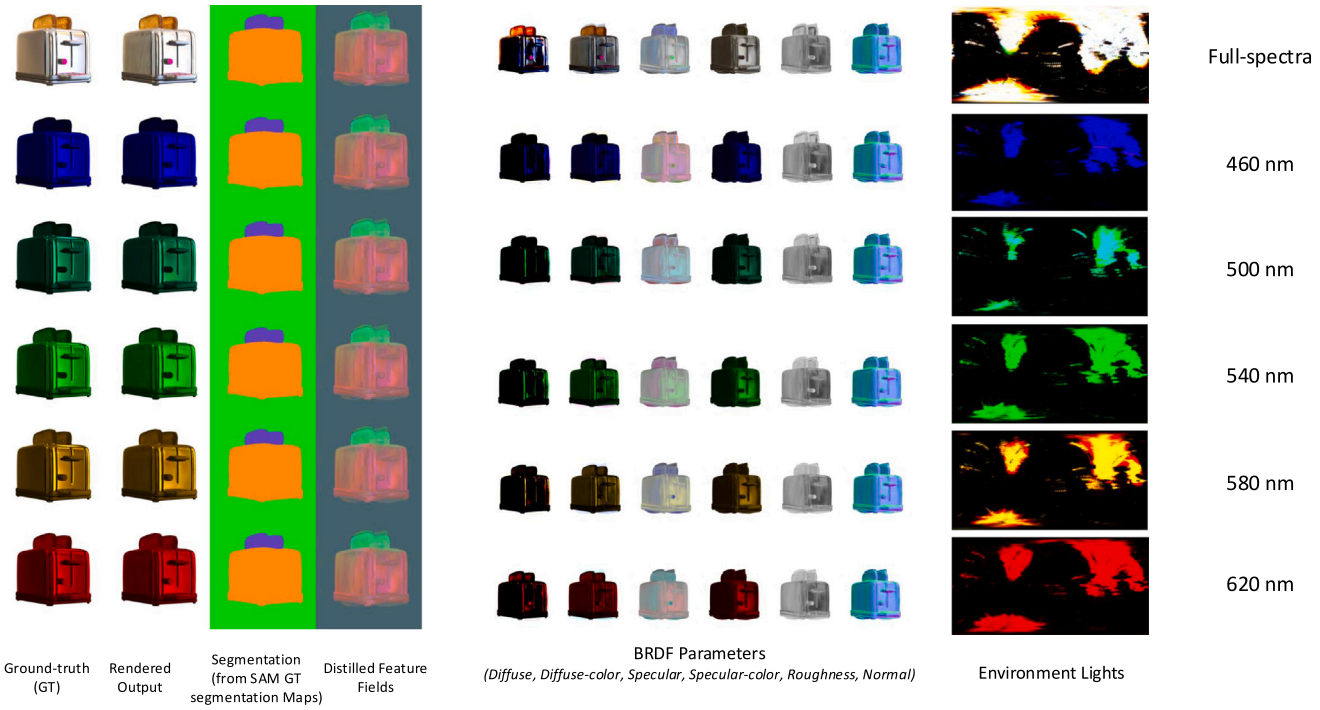

Fig. 10. Qualitative analysis (Rendering of the toaster scene using our framework from multi-spectral data).

Table 8
Ablation studies comparing full-spectra reconstructions with (a) no prior initialization from other spectra and (b) learnable parameters in full spectra initialized with priors from other spectra after a warm-up iteration show that the latter approach yields better results.

	Dragon doll	Orange	Tech	hall4	Avg.
		PSNR ↑			
(a)	36.55	42.98	40.17	40.99	40.17
(b)	38.52	44.13	40.73	42.14	41.63
		SSIM ↑			
(a)	0.972	0.992	0.986	0.989	0.985
(b)	0.980	0.994	0.987	0.991	0.988
		LPIPS ↓			
(a)	0.047	0.017	0.045	0.018	0.031
(b)	0.029	0.013	0.051	0.017	0.027

details, such as the edges of the shuttlecock in the Dragon doll scene, as well as enhanced reconstruction of objects like the speaker in the tech scene (see Fig. 11) . These findings further reinforce the effectiveness of incorporating information from other spectra in achieving more accurate and detailed rendering results.

5.7. Limitations

While the presented framework offers promising capabilities, it is important to acknowledge its limitations. One such limitation is the requirement for spectrum-maps to be co-registered, which can be a complex and time-intensive process. Moreover, as the resolution of images increases and more spectra are incorporated, the training time escalates significantly. To overcome these challenges, future research can explore the integration of alternative deep learning algorithms that support end-to-end training specifically for co-registering maps. Additionally, improving the encoding methods to efficiently accommodate a larger number of spectra would enhance the framework's capabilities.

Another limitation to consider is that the shading model currently used in the framework is fixed. However, the framework can be modified to have a flexible number of learnable parameters based on the shading model. This would allow users to configure the framework to their specific needs and enable more customized and adaptable shading models. By addressing these limitations, the framework can be made more practical and effective, enabling seamless co-registration, support for an expanded range of spectra, reduced training time for high-resolution images, and user-configurable shading models.

6. Conclusion

We presented 3D Spectral Gaussian Splatting, a cross-spectral rendering framework that utilizes 3D Gaussian Splatting to generate realistic and semantically meaningful splats from registered multi-view spectrum and segmentation maps. This framework enhances scene representation by incorporating multiple spectra, providing valuable insights into material properties and segmentation. Additionally, the paper introduces an improved physically-based rendering approach for Gaussian splats, enabling accurate estimation of reflectance and lights per spectra, resulting in enhanced realism. Furthermore, the paper showcases the potential of spectral scene understanding for precise scene editing techniques such as style transfer, in-painting, and removal. The contributions of this work address challenges in multi-spectral scene representation, rendering, and editing, opening up new possibilities for diverse applications.

Future work can focus on improving the accuracy of lighting and reflectance estimation in the proposed framework. While we demonstrated our approach to outperform other recent, spectral learning-based scene representations (Poggi et al., 2022; Li et al., 2024b) for different scenes, the evaluation of its potential for high-precision scanning with costly devices like the TAC7 (Merzbach et al., 2017), that allow capturing lots of photographs under controlled light-view conditions, might be interesting as well. There might be a chance that our learning-based spectral scene representation offers advantages over the parametric models used as a default option for the TAC7 due to the flexibility of the learnable models. Additionally, the utilization of spectral data, which has not been used in learning-based scene representation techniques like NeRFs or 3D Gaussian Splatting with a careful reflectance modeling so far, can open up new possibilities for achieving better results in this field. Additionally, integrating a registration process into the pipeline would allow for end-to-end training of non-co-registered spectrum maps, which is common with many spectral

Table 9
Ablation studies on the Shiny Blender Dataset to evaluate the impact of integrating spectral information with material information. Metrics are averages across all scenes in the dataset.

Method	Material Estimation	Spectral Priors	PSNR ↑	SSIM ↑	LPIPS ↓
Guassian Splatting (Kerbl et al., 2023)	×	×	30.37	0.947	0.083
Gaussian shader (Jiang et al., 2023)	✓	×	31.94	0.957	0.068
Ours	✓	✓	34.524	0.969	0.053

Table 10
Ablation studies on the Synthetic NeRF Dataset to evaluate the impact of integrating spectral information with material information. Metrics are averages across all scenes in the dataset.

Method	Material Estimation	Spectral Priors	PSNR ↑	SSIM ↑	LPIPS ↓
Guassian Splatting (Kerbl et al., 2023)	×	×	35.17	0.987	0.012
Gaussian shader (Jiang et al., 2023)	✓	×	35.22	0.986	0.011
Ours	✓	✓	36.39	0.987	0.017

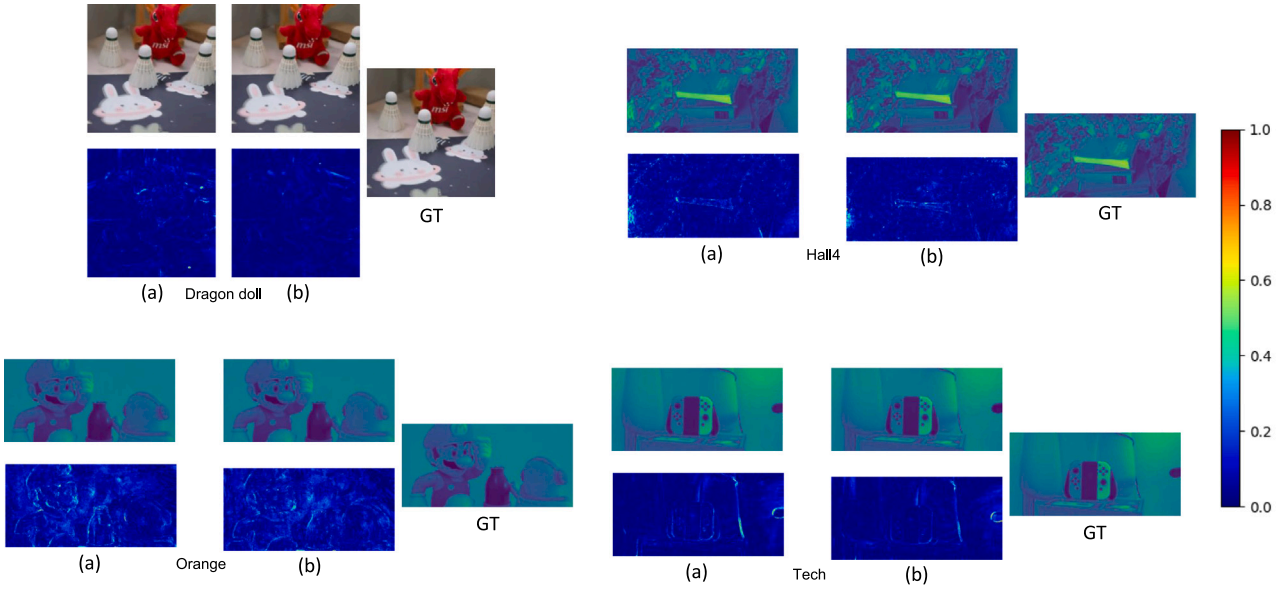

Fig. 11. Ablation studies were conducted to assess the differences with ground truth for scenes dragon-doll (Li et al., 2024b), hall4, orange, and tech (Poggi et al., 2022). The models were trained under two conditions: (a) without initialization of full-spectra model parameters from other spectra, and (b) with initialization of full-spectra model parameters using the average of common model parameters from other spectra. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

cameras. Exploring these areas can lead to better results and expand the possibilities of research in this field and open new opportunities for several applications where spectral characteristics are of great importance.

CRediT authorship contribution statement

Saptarshi Neil Sinha: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Holger Graf:** Validation, Resources, Project administration. **Michael Weinmann:** Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration, Methodology, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Saptarshi Neil Sinha reports financial support was provided by Fraunhofer Gesellschaft IuK Verbund (FhG). Saptarshi Neil Sinha reports a relationship with Fraunhofer Gesellschaft IuK Verbund (FhG) that includes: employment. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The work presented in this paper has been partially funded by the European Commission during the project PERCEIVE under grant agreement 101061157.

References

2018. Residual shuffling convolutional neural networks for deep semantic image segmentation using multi-modal data. ISPRS Ann. Photogramm. Remote. Sens. Spat. Inf. Sci. 4, 65–72.

2019. Fusion of hyperspectral, multispectral, color and 3D point cloud information for the semantic interpretation of urban environments. Int. Arch. Photogramm. Remote. Sens. Spat. Inf. Sci. 42, 1899–1906.

Afifi, A.J., Thiele, S.T., Rizaldy, A., Lorenz, S., Ghamisi, P., Tolosana-Delgado, R., Kirsch, M., Gloaguen, R., Heizmann, M., 2023. Tinto: Multisensor benchmark for 3d hyperspectral point cloud segmentation in the geosciences. IEEE Trans. Geosci. Remote Sens..

Alfeld, M., Mulliez, M., Devogelaere, J., de Viguerie, L., Jockey, P., Walter, P., 2018. MA-XRF and hyperspectral reflectance imaging for visualizing traces of antique polychromy on the frieze of the Siphnian treasury. Microchem. J. 141, 395–403.

Attal, B., Laidlaw, E., Gokaslan, A., Kim, C., Richardt, C., Tompkin, J., O'Toole, M., 2021. TöRF: Time-of-flight radiance fields for dynamic scene view synthesis. NeurIPS 34, 26289–26301.

Barron, J.T., Mildenhall, B., Tancik, M., Hedman, P., Martin-Brualla, R., Srinivasan, P.P., 2021. Mip-NeRF: A multiscale representation for anti-aliasing neural radiance fields. arXiv:2103.13415.

Barron, J.T., Mildenhall, B., Verbin, D., Srinivasan, P.P., Hedman, P., 2022. Mip-NeRF 360: Unbounded anti-aliased neural radiance fields. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5460–5469.

Barron, J.T., Mildenhall, B., Verbin, D., Srinivasan, P.P., Hedman, P., 2023. Zip-NeRF: Anti-aliased grid-based neural radiance fields. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 19697–19705.

Bi, S., Xu, Z., Srinivasan, P., Mildenhall, B., Sunkavalli, K., Hašan, M., Hold-Geoffroy, Y., Kriegman, D., Ramamoorthi, R., 2020. Neural reflectance fields for appearance acquisition. arXiv preprint arXiv:2008.03824.

Bortolon, M., Tsesmelis, T., James, S., Poiesi, F., Bue, A.D., 2024. 6DGS: 6D pose estimation from a single image and a 3D Gaussian splatting model. URL: https://arxiv.org/abs/2407.15484.

Boss, M., Braun, R., Jampani, V., Barron, J.T., Liu, C., Lensch, H., 2021a. Nerd: Neural reflectance decomposition from image collections. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 12684–12694.

Boss, M., Jampani, V., Braun, R., Liu, C., Barron, J., Lensch, H., 2021b. Neural-pil: Neural pre-integrated lighting for reflectance decomposition. Adv. Neural Inf. Process. Syst. 34, 10691–10704.

Bouzidi, A., Laga, H., Wannous, H., 2025. Gaussian RBFNet: Gaussian radial basis functions for fast and accurate representation and reconstruction of neural fields. URL: https://arxiv.org/abs/2503.06762.

Chen, K., Fu, K., Sun, X., Weinmann, M., Hinz, S., Jutzi, B., Weinmann, M., 2018. Deep semantic segmentation of aerial imagery based on multi-modal data. In: IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium. pp. 6219–6222. http://dx.doi.org/10.1109/IGARSS.2018.8519225.

Chen, Y., Li, M., Wu, Q., Lin, W., Harandi, M., Cai, J., 2025. PCGS: Progressive compression of 3D Gaussian splatting. URL: https://arxiv.org/abs/2503.08511.

Chen, T., Wang, P., Fan, Z., Wang, Z., 2022a. Aug-NeRF: Training stronger neural radiance fields with triple-level physically-grounded augmentations. In: IEEE Conference on Computer Vision and Pattern Recognition. CVPR.

Chen, A., Xu, Z., Geiger, A., Yu, J., Su, H., 2022b. TensoRF: Tensorial radiance fields. In: Proceedings of the European Conference on Computer Vision. pp. 333–350.

Chen, X., Zhang, Q., Li, X., Chen, Y., Feng, Y., Wang, X., Wang, J., 2022c. Hallucinated neural radiance fields in the wild. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12943–12952.

Cheng, H.K., Oh, S.W., Price, B., Schwing, A., Lee, J.-Y., 2023. Tracking anything with decoupled video segmentation. URL: https://arxiv.org/abs/2309.03903.

Deng, K., Liu, A., Zhu, J.-Y., Ramanan, D., 2022. Depth-supervised NeRF: Fewer views and faster training for free. In: CVPR. IEEE, pp. 12882–12891. http://dx.doi.org/10.1109/cvpr52688.2022.01254.

Deng, K., Yang, J., Wang, S., Xie, J., 2025a. GigaSLAM: Large-scale monocular SLAM with hierachical Gaussian splats. URL: https://arxiv.org/abs/2503.08071.

Deng, K., Yang, J., Wang, S., Xie, J., 2025b. GigaSLAM: Large-scale monocular SLAM with hierachical Gaussian splats. URL: https://arxiv.org/abs/2503.08071.

Devlin, K., Chalmers, A., Wilkie, A., Purgathofer, W., 2002. Tone Reproduction and Physically Based Spectral Rendering. The Eurographics Association, pp. 101–123, Conference Proceedings/Title of Journal: State of the Art Reports, Eurographics 2002.

Du, S., Du, S., Liu, B., Zhang, X., 2021. Incorporating DeepLabv3+ and object-based image analysis for semantic segmentation of very high resolution remote sensing images. Int. J. Digit. Earth 14 (3), 357–378.

Dufay, A., Murray, D., Pacanowski, R., et al., 2019. The malia rendering framework. https://pacanows.gitlabpages.inria.fr/MRF.

Fardo, F.A., Conforto, V.H., de Oliveira, F.C., Rodrigues, P.S., 2016. A formal evaluation of PSNR as quality measurement parameter for image segmentation algorithms. URL: https://arxiv.org/abs/1605.07116, arXiv:1605.07116.

Florath, J., Keller, S., Abarca-del Rio, R., Hinz, S., Staub, G., Weinmann, M., 2022. Glacier monitoring based on multi-spectral and multi-temporal satellite data: A case study for classification with respect to different snow and ice types. Remote. Sens. 14 (4), 845.

Fridovich-Keil, S., Yu, A., Tancik, M., Chen, Q., Recht, B., Kanazawa, A., 2022. Plenoxels: Radiance fields without neural networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5491–5500.

Gao, J., Gu, C., Lin, Y., Zhu, H., Cao, X., Zhang, L., Yao, Y., 2023. Relightable 3D Gaussian: Real-time point cloud relighting with BRDF decomposition and ray tracing. ArXiv:2311.16043.

Gao, Z., Planche, B., Zheng, M., Choudhuri, A., Chen, T., Wu, Z., 2025. 7DGS: Unified spatial-temporal-angular Gaussian splatting. URL: https://arxiv.org/abs/2503.07946.

Ge, W., Hu, T., Zhao, H., Liu, S., Chen, Y.-C., 2023. Ref-NeuS: Ambiguity-reduced neural implicit surface learning for multi-view reconstruction with reflection. arXiv preprint arXiv:2303.10840.

Grillini, F., de Ferri, L., Pantos, G.A., George, S., Veseth, M., 2024. Reflectance imaging spectroscopy for the study of archaeological pre-columbian textiles. Microchem. J. 200, 110168.

Han, J.J., Wu, J.Y., 2025. EndoPBR: Material and lighting estimation for photorealistic surgical simulations via physically-based rendering. URL: https://arxiv.org/abs/2502.20669.

Hasselgren, J., Hofmann, N., Munkberg, J., 2022. Shape, Light, and Material Decomposition from Images using Monte Carlo Rendering and Denoising. ArXiv:2206.03380.

Huang, L., Guo, J., Dan, J., Fu, R., Wang, S., Li, Y., Guo, Y., 2024. Spectral-GS: Taming 3D Gaussian splatting with spectral entropy. URL: https://arxiv.org/abs/2409.12771, arXiv:2409.12771.

Jakob, W., 2010. Mitsuba 2: Physically based renderer. http://www.mitsuba-renderer.org.

Jakob, W., Speierer, S., Roussel, N., Nimier-David, M., Vicini, D., Zeltner, T., Nicolet, B., Crespo, M., Leroy, V., Zhang, Z., 2022. Mitsuba 3 renderer. https://mitsuba-renderer.org.

JENOPTIK, 2024. EVIDIR alpha thermal imaging camera and infrared modules – one size for all variants. https://www.jenoptik.com/products/cameras-and-imaging-modules/thermographic-camera/thermal-imaging-camera. (Accessed 31 May 2024).

Jeong, Y., Ahn, S., Choy, C., Anandkumar, A., Cho, M., Park, J., 2021. Self-calibrating neural radiance fields. In: ICCV. IEEE, pp. 5846–5854. http://dx.doi.org/10.1109/iccv48922.2021.00579.

Jiang, Y., Tu, J., Liu, Y., Gao, X., Long, X., Wang, W., Ma, Y., 2023. GaussianShader: 3D Gaussian splatting with shading functions for reflective surfaces. arXiv preprint arXiv:2311.17977.

Jun-Seong, K., Yu-Ji, K., Ye-Bin, M., Oh, T.-H., 2022. HDR-Plenoxels: Self-calibrating high dynamic range radiance fields. In: Proceedings of the European Conference on Computer Vision. pp. 384–401.

Jung, D.-H., Kim, J.D., Kim, H.-Y., Lee, T.S., Kim, H.S., Park, S.H., 2022. A hyperspectral data 3D convolutional neural network classification model for diagnosis of gray mold disease in strawberry leaves. Front. Plant Sci. 13, 837020.

Kerbl, B., Kopanas, G., Leimkühler, T., Drettakis, G., 2023. 3D Gaussian splatting for real-time radiance field rendering. ACM Trans. Graph. 42 (4).

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.-Y., Dollár, P., Girshick, R., 2023. Segment anything. ArXiv:2304.02643.

Kobayashi, S., Matsumoto, E., Sitzmann, V., 2022. Decomposing NeRF for editing via feature field distillation. arXiv:2205.15585.

Koutsoudis, A., Ioannakis, G., Pistofidis, P., Arnaoutoglou, F., Kazakis, N., Pavlidis, G., Chamzas, C., Tsirliganis, N., 2021. Multispectral aerial imagery-based 3D digitisation, segmentation and annotation of large scale urban areas of significant cultural value. J. Cult. Herit. 49, 1–9.

Kwon, D.H., Hong, S.M., Abbas, A., Park, S., Nam, G., Yoo, J.-H., Kim, K., Kim, H.T., Pyo, J., Cho, K.H., 2023. Deep learning-based super-resolution for harmful algal bloom monitoring of inland water. GIScience Remote. Sens. 60 (1), 2249753.

Landi, M., Maino, G., 2011. Multispectral imaging and digital restoration for paintings documentation. In: Maino, G., Foresti, G.L. (Eds.), Image Analysis and Processing. ICIAP 2011, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 464–474.

Lanteri, L., Pelosi, C., 2021. 2D and 3D ultraviolet fluorescence applications on cultural heritage paintings and objects through a low-cost approach for diagnostics and documentation. In: Liang, H., Groves, R. (Eds.), In: Optics for Arts, Architecture, and Archaeology VIII, vol. 11784, International Society for Optics and Photonics, SPIE, 1178417. http://dx.doi.org/10.1117/12.2593691.

Li, H., Gao, Y., Wu, C., Zhang, D., Dai, Y., Zhao, C., Feng, H., Ding, E., Wang, J., Han, J., 2024a. GGRt: Towards generalizable 3D Gaussians without pose priors in real-time.

Li, H., Li, J., Zhang, D., Wu, C., Shi, J., Zhao, C., Feng, H., Ding, E., Wang, J., Han, J., 2024c. VDG: Vision-only dynamic Gaussian for driving simulation.

Li, R., Liu, J., Liu, G., Zhang, S., Zeng, B., Liu, S., 2024b. SpectralNeRF: Physically based spectral rendering with neural radiance field. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 38, (4), pp. 3154–3162.

LI, Y., Wang, J., Chu, L., Li, X., hong Kao, S., Chen, Y.-C., Lu, Y., 2025. StreamGS: Online generalizable Gaussian splatting reconstruction for unposed image streams. URL: https://arxiv.org/abs/2503.06235.

Li, K., Wang, J., Han, W., Zhao, D., 2025. Feature-EndoGaussian: Feature distilled Gaussian splatting in surgical deformable scene reconstruction. URL: https://arxiv.org/abs/2503.06161.

Li, H., Zhang, D., Dai, Y., Liu, N., Cheng, L., Li, J., Wang, J., Han, J., 2024d. GP-NeRF: Generalized perception NeRF for context-aware 3D scene understanding. CVPR.

Liang, R., Chen, H., Li, C., Chen, F., Panneer, S., Vijaykumar, N., 2023. ENVIDR: Implicit differentiable renderer with neural environment lighting. arXiv preprint arXiv:2303.13022.

Liang, R., Zhang, J., Li, H., Yang, C., Guan, Y., Vijaykumar, N., 2022. Spidr: Sdf-based neural point fields for illumination and deformation. arXiv preprint arXiv:2210.08398.

Lin, C.-H., Ma, W.-C., Torralba, A., Lucey, S., 2021. BaRF: Bundle-adjusting neural radiance fields. In: ICCV. IEEE, pp. 5741–5751. http://dx.doi.org/10.1109/iccv48922.2021.00569.

Liu, Y., Wang, P., Lin, C., Long, X., Wang, J., Liu, L., Komura, T., Wang, W., 2023a. NeRO: Neural geometry and BRDF reconstruction of reflective objects from multiview images. In: SIGGRAPH.

Liu, K., Zhan, F., Chen, Y., Zhang, J., Yu, Y., El Saddik, A., Lu, S., Xing, E., 2023b. Stylerf: Zero-shot 3D style transfer of neural radiance fields. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8338–8348.

Lorensen, W.E., Cline, H.E., 1998. Marching cubes: A high resolution 3D surface construction algorithm. In: Seminal Graphics: Pioneering Efforts that Shaped the Field. pp. 347–353.

Maggio, D., Carlone, L., 2025. Bayesian fields: Task-driven open-set semantic Gaussian splatting. URL: https://arxiv.org/abs/2503.05949.

Martin-Brualla, R., Radwan, N., Sajjadi, M.S., Barron, J.T., Dosovitskiy, A., Duckworth, D., 2021. 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. 7206–7215.

Merzbach, S., Weinmann, M., Klein, R., 2017. High-quality multi-spectral reflectance acquisition with X-rite TAC7. In: Proceedings of the Workshop on Material Appearance Modeling. pp. 11–16.

Mi, Z., Xu, D., 2023. Switch-NeRF: Learning scene decomposition with mixture of experts for large-scale neural radiance fields. In: ICLR.

Micasense, 2024. Micasense RedEdge-MX DUAL. https://drones.measurusa.com/products/micasense-rededge-mx-dual. (Accessed 24 April 2024).

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: ECCV.

Mitschke, I., Wiemann, T., Igelbrink, F., Hertzberg, J., 2022. Hyperspectral 3D point cloud segmentation using RandLA-Net. In: International Conference on Intelligent Autonomous Systems. Springer, pp. 301–312.

Moghadam, P., Ward, D., Goan, E., Jayawardena, S., Sikka, P., Hernandez, E., 2017. Plant disease detection using hyperspectral imaging. In: 2017 International Conference on Digital Image Computing: Techniques and Applications. DICTA, pp. 1–8. http://dx.doi.org/10.1109/DICTA.2017.8227476.

Müller, T., Evans, A., Schied, C., Keller, A., 2022. Instant neural graphics primitives with a multiresolution hash encoding. ACM Trans. Graph. 41 (4), 102:1–102:15.

Munkberg, J., Hasselgren, J., Shen, T., Gao, J., Chen, W., Evans, A., Müller, T., Fidler, S., 2022. Extracting Triangular 3D Models, Materials, and Lighting From Images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. CVPR, pp. 8280–8290.

Nilsson, J., Akenine-Möller, T., 2020. Understanding SSIM. URL: https://arxiv.org/abs/2006.13846, arXiv:2006.13846.

Nimier-David, M., Vicini, D., Zeltner, T., Jakob, W., 2019. Mitsuba 2: A retargetable forward and inverse renderer. ACM Trans. Graph. (Proc. SIGGRAPH Asia) 38 (6), 203:1–203:17.

Ong, D., Tao, Y., Murali, V., Spasojevic, I., Kumar, V., Chaudhari, P., 2025. ATLAS navigator: Active task-driven LAnguage-embedded Gaussian splatting. URL: https://arxiv.org/abs/2502.20386.

Palos Sánchez, P.R., Saura, J.R., Reyes Menéndez, A., 2019. Mapping multispectral digital images using a cloud computing software: applications from UAV images. Heliyon 5.

Park, W., Nam, M., Kim, S., Jo, S., Lee, S., 2025. ForestSplats: Deformable transient field for Gaussian splatting in the wild. URL: https://arxiv.org/abs/2503.06179.

Park, K., Sinha, U., Barron, J.T., Bouaziz, S., Goldman, D.B., Seitz, S.M., Martin-Brualla, R., 2021. Nerfies: Deformable neural radiance fields. In: ICCV. IEEE, pp. 5865–5874. http://dx.doi.org/10.1109/iccv48922.2021.00581.

Peng, L., Wu, A., Li, W., Xia, P., Dai, X., Zhang, X., Di, X., Sun, H., Pei, R., Wang, Y., Cao, Y., Zha, Z.-J., 2025. Pixel to Gaussian: Ultra-fast continuous super-resolution with 2D Gaussian modeling. URL: https://arxiv.org/abs/2503.06617.

Pharr, M., 2020. PBRT version 4. https://github.com/mmp/pbrt-v4.

Pharr, M., Jakob, W., Humphreys, G., 2016. Physically Based Rendering: From Theory to Implementation, third ed. Morgan Kaufmann.

Phenospex, B.V., 2024. PlantEye F600: Multispectral 3D Scanner for Plants. https://phenospex.com/products/plant-phenotyping/planteye-f600-multispectral-3d-scanner-for-plants/. (Accessed 25 June 2024).

Poggi, M., Zama Ramirez, P., Tosi, F., Salti, S., Di Stefano, L., Mattoccia, S., 2022. Cross-spectral neural radiance fields. In: Proceedings of the International Conference on 3D Vision. 3DV.

Pumarola, A., Corona, E., Pons-Moll, G., Moreno-Noguer, F., 2021. D-NeRF: Neural radiance fields for dynamic scenes. In: CVPR. IEEE, pp. 10318–10327. http://dx.doi.org/10.1109/cvpr46437.2021.01018.

Qin, M., Li, W., Zhou, J., Wang, H., Pfister, H., 2023. LangSplat: 3D language Gaussian splatting. arXiv preprint arXiv:2312.16084.

Qureshi, R., Uzair, M., Khurshid, K., Yan, H., 2019. Hyperspectral document image processing: Applications, challenges and future prospects. Pattern Recognit. 90, 12–22.

Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al., 2021. Learning transferable visual models from natural language supervision. In: International Conference on Machine Learning. PMLR, pp. 8748–8763.

Reiser, C., Peng, S., Liao, Y., Geiger, A., 2021. Kilonerf: Speeding up neural radiance fields with thousands of tiny mlps. In: ICCV. pp. 14335–14345.

Rematas, K., Liu, A., Srinivasan, P.P., Barron, J.T., Tagliasacchi, A., Funkhouser, T., Ferrari, V., 2022. Urban radiance fields. In: CVPR. IEEE, pp. 12932–12942. http://dx.doi.org/10.1109/cvpr52688.2022.01259.

Rizaldy, A., Afifi, A.J., Ghamisi, P., Gloaguen, R., 2023. Transformer-based models for hyperspectral point clouds segmentation. In: 2023 13th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing. WHISPERS, pp. 1–5. http://dx.doi.org/10.1109/WHISPERS61460.2023.10431346.

Rizaldy, A., Afifi, A.J., Ghamisi, P., Gloaguen, R., 2024a. Improving mineral classification using multimodal hyperspectral point cloud data and multi-stream neural network. Remote. Sens. 16 (13), 2336.

Rizaldy, A., Ghamisi, P., Gloaguen, R., 2024b. Channel attention module for segmentation of 3D hyperspectral point clouds in geological applications. Int. Arch. Photogramm. Remote. Sens. Spat. Inf. Sci. 48, 103–109.

Roessle, B., Barron, J.T., Mildenhall, B., Srinivasan, P.P., Nießner, M., 2022. Dense depth priors for neural radiance fields from sparse input views. In: CVPR. IEEE, pp. 12892–12901. http://dx.doi.org/10.1109/cvpr52688.2022.01255.

Schönberger, J.L., Frahm, J.-M., 2016. Structure-from-motion revisited. In: Conference on Computer Vision and Pattern Recognition. CVPR.

Senchuri, R., Kuras, A., Burud, I., 2021. Machine learning methods for road edge detection on fused airborne hyperspectral and lidar data. In: 2021 11th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing. WHISPERS, IEEE, pp. 1–5.

Shang, R., Zhang, J., Jiao, L., Li, Y., Marturi, N., Stolkin, R., 2020. Multi-scale adaptive feature fusion network for semantic segmentation in remote sensing images. Remote. Sens. 12 (5), 872.

Silios, 2024. Off-the-shelf snapshot multispectral cameras. https://www.silios.com/multispectral-imaging. (Accessed: 2024-04-24).

Sinha, S.N., Kühn, J., Graf, H., Weinmann, M., 2024. SpectralSplatsViewer: An interactive web-based tool for visualizing cross-spectral Gaussian splats. In: WEB3D '24: The 29th International ACM Conference on 3D Web Technology. ACM, Guimarães, Portugal, http://dx.doi.org/10.1145/3665318.3677151.

Skorokhodov, V., Durasov, N., Fua, P., 2025. D3DR: Lighting-aware object insertion in Gaussian splatting. URL: https://arxiv.org/abs/2503.06740.

Smits, B., 2000. An RGB to spectrum conversion for reflectances. J. Color. Sci. 10 (4), 200–215.

Song, R., Liang, C., Xia, Y., Zimmer, W., Cao, H., Caesar, H., Festag, A., Knoll, A., 2025. CoDa-4DGS: Dynamic Gaussian splatting with context and deformation awareness for autonomous driving. URL: https://arxiv.org/abs/2503.06744.

Srinivasan, P.P., Deng, B., Zhang, X., Tancik, M., Mildenhall, B., Barron, J.T., 2021. Nerv: Neural reflectance and visibility fields for relighting and view synthesis. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7495–7504.

Sun, Y., Wang, X., Zhang, Y., Zhang, J., Jiang, C., Guo, Y., Wang, F., 2024. iComMa: Inverting 3D Gaussian splatting for camera pose estimation via comparing and matching. URL: https://arxiv.org/abs/2312.09031.

Sun, L., Wu, F., Zhan, T., Liu, W., Wang, J., Jeon, B., 2020. Weighted nonlocal low-rank tensor decomposition method for sparse unmixing of hyperspectral images. IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens. 13, 1174–1188.

Tancik, M., Casser, V., Yan, X., Pradhan, S., Mildenhall, B., Srinivasan, P.P., Barron, J.T., Kretzschmar, H., 2022. Block-NeRF: Scalable large scene neural view synthesis. In: CVPR. IEEE, pp. 8248–8258. http://dx.doi.org/10.1109/cvpr52688.2022.00807.

Tancik, M., Weber, E., Ng, E., Li, R., Yi, B., Kerr, J., Wang, T., Kristoffersen, A., Austin, J., Salahi, K., Ahuja, A., McAllister, D., Kanazawa, A., 2023. Nerfstudio: A modular framework for neural radiance field development. In: ACM SIGGRAPH 2023 Conference Proceedings. SIGGRAPH '23.

Tewari, A., Thies, J., Mildenhall, B., et al., 2022. Advances in neural rendering. In: CGF, vol. 41, (2), Wiley Online Library, Wiley, pp. 703–735.

The ART development team, 2018. The Advanced Rendering Toolkit. https://cgg.mff.cuni.cz/ART.

Turki, H., Ramanan, D., Satyanarayanan, M., 2022. Mega-NeRF: Scalable construction of large-scale NeRFs for virtual fly-throughs. In: CVPR. IEEE, pp. 12922–12931. http://dx.doi.org/10.1109/cvpr52688.2022.01258.

Valada, A., Oliveira, G., Brox, T., Burgard, W., 2016. Deep multispectral semantic scene understanding of forested environments using multimodal fusion. In: International Symposium on Experimental Robotics. ISER.

Van Holland, L., Weinmann, M., Müller, J.U., Stotko, P., Klein, R., 2025. NeRFs are mirror detectors: Using structural similarity for multi-view mirror scene reconstruction with 3D surface primitives. arXiv preprint arXiv:2501.04074.

Verbin, D., Hedman, P., Mildenhall, B., Zickler, T., Barron, J.T., Srinivasan, P.P., 2022. Ref-NeRF: Structured view-dependent appearance for neural radiance fields. CVPR.

Vetrekar, N., Raghavendra, R., Gad, R., 2016. Low-cost multi-spectral face imaging for robust face recognition. In: 2016 IEEE International Conference on Imaging Systems and Techniques. IST, pp. 324–329. http://dx.doi.org/10.1109/IST.2016.7738245.

Walter, B., Marschner, S., Li, H., Torrance, K.E., 2007. Microfacet models for refraction through rough surfaces. In: Rendering Techniques. URL: https://api.semanticscholar.org/CorpusID:8061072.

Wang, Y., Han, Q., Habermann, M., Daniilidis, K., Theobalt, C., Liu, L., 2023a. Neus2: Fast learning of neural implicit surfaces for multi-view reconstruction. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 3272–3283.

Wang, S., Leroy, V., Cabon, Y., Chidlovskii, B., Jerome, R., 2023b. DUSt3R: Geometric 3D Vision Made Easy. ArXiv Preprint 2312.14132.

Wang, P., Liu, L., Liu, Y., Theobalt, C., Komura, T., Wang, W., 2021a. NeuS: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. Adv. Neural Inf. Proc. Syst. 354, 27171–27183.

Wang, C., Wu, X., Guo, Y.-C., Zhang, S.-H., Tai, Y.-W., Hu, S.-M., 2022. NeRF-SR: High quality neural radiance fields using supersampling. In: Proceedings of the 30th ACM International Conference on Multimedia. pp. 6445–6454.

Wang, Z., Wu, S., Xie, W., Chen, M., Prisacariu, V.A., 2021b. NeRF–: Neural radiance fields without known camera parameters. arXiv preprint arXiv:2102.07064.

Wei, Y., Liu, S., Rao, Y., Zhao, W., Lu, J., Zhou, J., 2021. NerfingMVS: Guided optimization of neural radiance fields for indoor multi-view stereo. In: ICCV. IEEE, pp. 5610–5619. http://dx.doi.org/10.1109/iccv48922.2021.00556.

Weinmann, M., Weinmann, M., 2017. Geospatial computer vision based on multi-modal data—How valuable is shape information for the extraction of semantic information? Remote. Sens. 10 (1), 2.

Weinmann, M., Weinmann, M., 2019. Urban scene labeling based on multi-modal data acquired from aerial sensor platforms. In: 2019 Joint Urban Remote Sensing Event. JURSE, pp. 1–4. http://dx.doi.org/10.1109/JURSE.2019.8809035.

Wu, D., Liu, L., Linli, Z., Huang, A., Song, L., Yu, Q., Wu, Q., Lu, C., 2025a. REArtGS: Reconstructing and generating articulated objects via 3D Gaussian splatting with geometric and motion constraints. URL: https://arxiv.org/abs/2503.06677.

Wu, K., Zhang, Z., Tie, M., Ai, Z., Gan, Z., Ding, W., 2025b. VINGS-mono: Visual-inertial Gaussian splatting monocular SLAM in large scenes. URL: https://arxiv.org/abs/2501.08286.

Wyszecki, G., Stiles, W., 2000. In: Wyszecki, G., Stiles, W.S. (Eds.), Color Science: Concepts and Methods, Quantitative Data and Formulae, second Edition Wiley-VCH, ISBN: 0-471-39918-3, p. 968.

Xu, Q., Xu, Z., Philip, J., Bi, S., Shu, Z., Sunkavalli, K., Neumann, U., 2022. Point-nerf: Point-based neural radiance fields. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5438–5448.

Yang, J., Huang, J., Chen, Y., Wang, Y., Li, B., You, Y., Sharma, A., Igl, M., Karkus, P., Xu, D., Ivanovic, B., Wang, Y., Pavone, M., 2024a. STORM: Spatio-temporal reconstruction model for large-scale outdoor scenes. URL: https://arxiv.org/abs/2501.00602.

Yang, R., Zhu, Z., Jiang, Z., Ye, B., Chen, X., Zhang, Y., Chen, Y., Zhao, J., Zhao, H., 2024b. Spectrally pruned Gaussian fields with neural compensation. arXiv:2405.00676.

Yariv, L., Gu, J., Kasten, Y., Lipman, Y., 2021. Volume rendering of neural implicit surfaces. In: Thirty-Fifth Conference on Neural Information Processing Systems.

Yariv, L., Hedman, P., Reiser, C., Verbin, D., Srinivasan, P.P., Szeliski, R., Barron, J.T., Mildenhall, B., 2023. BakedSDF: Meshing neural SDFs for real-time view synthesis. In: Brunvand, E., Sheffer, A., Wimmer, M. (Eds.), Proceedings of the ACM SIGGRAPH 2023 Conference. 46:1–46:9.

Ye, M., Danelljan, M., Yu, F., Ke, L., 2023. Gaussian grouping: Segment and edit anything in 3D scenes. arXiv preprint arXiv:2312.00732.

Yen-Chen, L., Florence, P., Barron, J.T., Rodriguez, A., Isola, P., Lin, T.-Y., 2021. iNeRF: Inverting neural radiance fields for pose estimation. In: IROS. IEEE, pp. 1323–1330. http://dx.doi.org/10.1109/iros51168.2021.9636708.

Yong, S., Manivannan, V.N.P., Kerbl, B., Wan, Z., Stepputtis, S., Sycara, K., Xie, Y., 2025. OMG: Opacity matters in material modeling with Gaussian splatting. URL: https://arxiv.org/abs/2502.10988.

Yuan, Z., Pu, Y., Luo, H., Lang, F., Chi, C., Li, T., Shen, Y., Sun, H., Wang, B., Yang, X., 2025. Uni-Gaussians: Unifying camera and lidar simulation with Gaussians for dynamic driving scenarios. URL: https://arxiv.org/abs/2503.08317.

Zahra, A., Qureshi, R., Sajjad, M., Sadak, F., Nawaz, M., Khan, H.A., Uzair, M., 2024. Current advances in imaging spectroscopy and its state-of-the-art applications. Expert Syst. Appl. 238, 122172.

Zhan, Y., Hu, D., Wang, Y., Yu, X., 2017. Semisupervised hyperspectral image classification based on generative adversarial networks. IEEE Geosci. Remote. Sens. Lett. 15 (2), 212–216.

Zhang, Y., Chi, M., 2020. Mask-R-FCN: A deep fusion network for semantic segmentation. IEEE Access 8, 155753–155765.

Zhang, Y., Huang, X., Ni, B., Li, T., Zhang, W., 2023. Frequency-modulated point cloud rendering with easy editing. arXiv preprint arXiv:2303.07596.

Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O., 2018. The unreasonable effectiveness of deep features as a perceptual metric. In: CVPR.

Zhang, K., Luan, F., Wang, Q., Bala, K., Snavely, N., 2021a. Physg: Inverse rendering with spherical gaussians for physics-based material editing and relighting. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5453–5462.

Zhang, X., Meng, J., Xu, Z., Yang, S., Wu, Y., Wang, R., Zhang, J., 2025a. SecureGS: Boosting the security and fidelity of 3D Gaussian splatting steganography. URL: https://arxiv.org/abs/2503.06118.

Zhang, X., Srinivasan, P.P., Deng, B., Debevec, P., Freeman, W.T., Barron, J.T., 2021b. Nerfactor: Neural factorization of shape and reflectance under an unknown illumination. ACM Trans. Graph. (ToG) 40 (6), 1–18.

Zhang, X., Yang, J., Lin, T., Ying, Y., 2021c. Food and agro-product quality evaluation based on spectroscopy and deep learning: A review. Trends Food Sci. Technol. 112, 431–441.

Zhang, F., Yang, H., Zhang, Z., Huang, Z., Luo, Y., 2025b. TT-GaussOcc: Test-time compute for self-supervised occupancy prediction via spatio-temporal Gaussian splatting. URL: https://arxiv.org/abs/2503.08485.

Zhang, T., Zhi, W., Mangelson, J., Johnson-Roberson, M., 2025c. Infinite leagues under the sea: Photorealistic 3D underwater terrain generation by latent fractal diffusion models. URL: https://arxiv.org/abs/2503.06784.

Zheng, G., Deng, J., Chu, X., Yuan, Y., Li, H., Zhang, Y., 2025. S3R-GS: Streamlining the pipeline for large-scale street scene reconstruction. URL: https://arxiv.org/abs/2503.08217.

Zhi, S., Laidlow, T., Leutenegger, S., Davison, A.J., 2021. In-place scene labelling and understanding with implicit scene representation. URL: https://arxiv.org/abs/2103.15875.