DOI: 10.1111/cgf.70337

EUROGRAPHICS 2026 / B. Masia and J. Thies
(Guest Editors)

COMPUTER GRAPHICS *forum*
Volume 45 (2026), Number 2

Multi-Spectral Gaussian Splatting with Neural Color Representation

Lukas Meyer[1], Josef Grün[1], Maximilian Weiherer[1], Bernhard Egger[1], Marc Stamminger[1] and Linus Franke[1,2,*]

[1] Visual Computing Erlangen, Friedrich-Alexander-Universität Erlangen-Nürnberg-Fürth, Germany
[2] Inria, Université Côte d'Azur, France

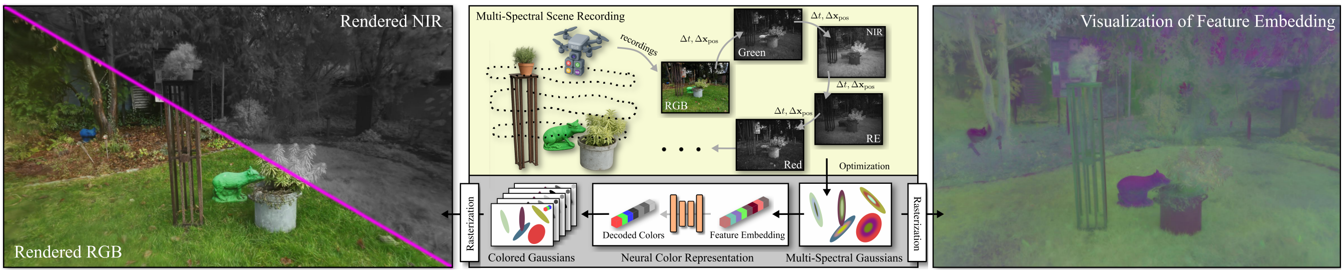

Figure 1: *Our approach integrates multi-spectral images taken from different capture positions and cameras (middle top) into a single 3D Gaussian-based [KKLD23] representation. Therefore, we introduce a neural color model to jointly optimize and store all spectral bands. To render individual spectral bands, we use a shallow MLP to decompose the feature embedding into its constituent bands (middle bottom). The left image shows a rendering of RGB and near-infra red, while the right image visualizes the feature embedding. Notice how regions with similar spectral characteristics coalesce in the learned feature space.*

Abstract

3D Gaussian Splatting (3DGS) [KKLD23] has transformed novel-view synthesis from RGB images, yet remains restricted to the visible spectrum. Many applications, including agricultural monitoring, rely on multi-spectral imaging, where spectral camera alignment and scalability pose major challenges.

We present MS-Splatting—a multi-spectral 3DGS framework enabling unified multi-view consistent reconstruction and rendering across both visible and invisible spectra. Our key component is a neural color representation that encodes per-primitive features shared across spectral bands, decoded through a shallow multi-layer perceptron into spectrum-specific radiance. By leveraging inter-band correlations, this formulation enhances detail while reducing memory consumption compared to independent band modeling via per-channel modeling with spherical harmonics.

Our method enables accurate parallax-free novel-view vegetation index rendering for plant monitoring and enhances RGB novel view synthesis quality by exploiting details revealed through multi-spectral bands. Our evaluation demonstrates that MS-Splatting exceeds the current leading methods in both categories. In addition, we introduce a multi-spectral dataset from aerial captures covering outdoor environments, specifically designed for evaluating these applications. We will release our code and dataset to facilitate further research. The project page is located at: `https://meyerls.github.io/ms_splatting`

CCS Concepts

• ***Computing methodologies*** → ***Rendering; Reconstruction; Hyperspectral imaging;*** • ***Applied computing*** → ***Agriculture;***

1. Introduction

The rise of recent (neural) rendering techniques such as Neural Radiance Fields (NeRFs) [MST*20] and 3D Gaussian Splatting (3DGS) [KKLD23], which are optimized from captured images in the *visible* (RGB) light spectrum, has greatly enhanced interaction and processing of 3D scenes. Specifically, 3DGS provides the state-of-the-art technique in this field by reconstructing the radiance field of the scene through a set of anisotropic 3D Gaussian primitives, allowing for efficient rendering and optimization [MGK*24, MSL*25] and scaling to large scenes [KMK*24, LLT*24].

* Corresponding author
Email: {firstname.lastname}@fau.de

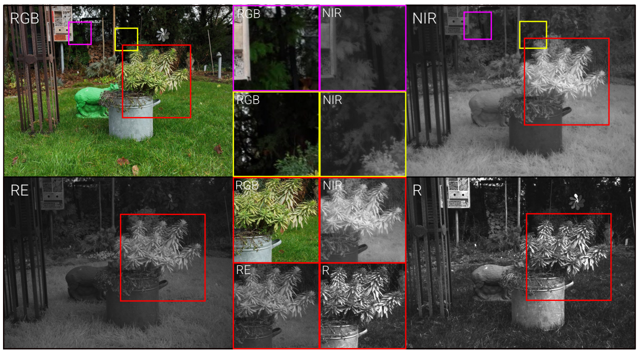

Figure 2: *Visualization of different spectral images, which allow for quality enhancement and compression for radiance fields. Near-infra red (NIR) reveals details not visible in RGB (magenta and yellow box) which allows for finer reconstruction. Light reflectance characteristics between bands (RE: red-edge; R: narrow 32nm red band) are shared (red box), which allows for compression.*

On the other hand, *multi-spectral* cameras, used to obtain images across different spectral bands (e.g., near-infra red or specific narrow bands of visible light) in expansive outdoor environments, have become prevalent and are available in both handheld and aerial formats; however, usually as independent camera sensors without calibration or synchronized shutters. Such cameras enable the acquisition of detailed surface characteristics [Anu70, AQT24] and material properties [NP12] by observing the isolated or combined emitted multi-spectral surface radiance of different spectral bands. The increasing affordability of these cameras has also enabled widespread adoption in agriculture and forestry, particularly for plant monitoring using *Vegetation Indices* [HRLL20]. Most processing pipelines involving these indices, however, demand precise co-registration of spectral channels, which is challenging due to parallax introduced by motion capture, wind effects, and variations in shutter timing between spectral cameras.

In this paper, we integrate multi-spectral capturings with radiance field rendering techniques into a unified 3DGS reconstruction model, allowing us to both increase novel view synthesis (NVS) quality in visible (RGB) light spectra, as well as to use NVS to precisely render aligned multi-spectral images, which form the base for downstream tasks such as agricultural monitoring. Furthermore, different from previous methodologies in the field, we take care not to introduce scaling constraints, e.g., positional dependencies common in NeRF or other neural field frameworks, which limit application in large-scale scenarios such as drone-scanned agriculture fields.

Since we have to merge multi-spectral information from multiple independent cameras, we face the challenges of reconstructing the trajectories in a joint space, merging the single bands into a joint representation, and efficiently querying and scaling this information to the requirements of agricultural applications.

Our key insight is that we can exploit the correlations in the interaction of different light spectra with scene surfaces for efficient reconstruction and rendering. The correlations are twofold: Firstly, different spectra reveal different details of materials and surfaces, resolving ambiguities in the optimization and also sharpening details. Secondly, nearby wavelengths of light are reflected and absorbed similarly by most surfaces, which allows to compress information. A visualization of these effects can be seen in Fig. 2

Consequently, instead of modeling all spectra in isolation, e.g., with separate sets of spherical harmonics per band [LCZ*24, GMW*25], we introduce a unified multi-spectral color representation for 3DGS. We refer to this representation as *neural color representation*, for which we split the radiance emittance model into two parts: (1) a neural feature vector per primitive, which acts as the surface characteristic convolved with the multi-spectral irradiance[†], and (2) a global function (approximated with an MLP) for generating emittance that calculates per-primitive radiance based on the outgoing direction and the demanded light spectrum.

This formulation efficiently exploits multi-spectral correlations and allows the per-primitive representation to be concise; however, it requires the underlying primitives to be grouped by spectral characteristics. To further facilitate this, we incorporate a multi-spectral aware densification formulation, allowing for each primitive in the reconstruction to achieve consistent surface properties. With our pipeline, we are able to increase RGB rendering quality by up to 1 dB in PSNR compared to a common 3DGS model [KKLD23] through cross-spectral information exchange, commonly called *spectral cross-talk*. Furthermore, we reduce memory consumption by 88% compared to 3DGS extended to the multi-spectral domain [GMW*25].

To evaluate our approach for agricultural applications, we introduce a new drone-captured multi-spectral dataset and showcase the effectiveness of our method in large-area capturing and computing vegetation indices. For the latter, we exploit the fact that rendering novel views from our multi-spectral model allows synthesizing a shared optical sensor not present in real-world capturing devices, something that is not leveraged in previous approaches.

In summary, our key contributions are as follows:

- We introduce MS-Splatting, a multi-spectral neural scene representation based on 3DGS which seamlessly incorporates multiple different visible and invisible spectra of light capturings.
- We propose a novel *neural color representation* that efficiently encodes multiple spectral bands into a joint representation shared across different spectra.
- We are the first to compute vegetation indices using novel view synthesis, allowing detailed parallax-free plant monitoring for agricultural applications.

We will release both the code and the multi-spectral dataset.

2. Related Work

2.1. Novel View Synthesis and Radiance Fields

Methods for novel view synthesis historically build on image-based rendering [SK00, SCK08], whereby images are warped

† This is because we do not model light transport but optimize the radiance field. 3DGS and similar particle-based radiance field methods represent the scene as a set of emitting primitives, each with a view-dependent radiance.

onto a geometry proxy obtained through active depth measuring or Structure-from-Motion [SSS06, SF16] and multi-view stereo [GSC*07, SZPF16]. Recent advances in neural rendering [TFT*20, TTM*22] allow direct entangling of 3D reconstruction and rendering through differentiable rendering. Neural Radiance Fields (NeRFs) [MST*20] optimize a 3D volume abstracted in an MLP, sparking a field with many subsequent quality and speed increases [ZRSK20, BMV*22, BMT*21, MESK22, BMV*23]. These methods abstract the complete appearance in an MLP or feature grid; however, scaling this methodology to large-scale areas is computationally intensive, both in optimization and rendering [TRS22].

Recently, point-based approaches proved to be a great alternative in the speed-to-quality tradeoff, first represented through point sampling or splat rendering with neural filtering [ASK*20]. Thereby, high-dimensional per-point features are optimized for each point of a geometry proxy to capture the local appearance and decoded with a CNN [RFS22, FRFS24, KPLD21] or an MLP [KLR*22]. We draw inspiration from these approaches, as we optimize high-dimensional appearance features per primitive to capture multi-spectral appearances and radiance; however, we specifically use this as a scheme to unify appearance channels.

2.2. 3D Gaussian Splatting

3D Gaussian Splatting (3DGS) [KKLD23] revolutionized point-based radiance field rendering by representing points as 3D Gaussian distributions $G(\mathbf{x}) = e^{-\frac{1}{2}(\mathbf{x}-\mu)^T\Sigma^{-1}(\mathbf{x}-\mu)}$, with position μ and covariance Σ, and omitting neural components entirely. Rendering is done with alpha blending $\mathbf{C} = \sum_{i\in N} \mathbf{c}_i\alpha_i T_i$, with $T_i = \prod_{j=1}^{i-1}(1-\alpha_j)$ for each pixel with all contributing Gaussians N with colors c. Notably, colors are hereby stored in spherical harmonics with three degrees.

Subsequent works improved performance [RSP*24, HFW*25, YCH*24, KRS*24, NMR*24], compression [BKL*24, NSW24], scalability [KMK*24, LLT*24], and dynamic scenes [LKLR24, WYF*24]; however, the original algorithm remains remarkably consistent in reconstruction quality. We refer to recent surveys for more details [CW24, DHRK24, FXZ*24, WYZ*24].

Neural components in radiance fields have been excessively used by *in-the-wild* scenarios [MBRS*21] adapted to 3DGS, which we take inspiration from. WildGaussians [KPK*24] use per-Gaussian optimizable embeddings during training, combined with per-image embeddings to harmonize different illuminations in the dataset. The embeddings serve as affine color transformations, allowing control of the illumination, with Wild-GS [XMP24] and GS-W [ZWW*24] following similar methodologies. Scaffold-GS [LYX*24] and Octree-GS [RJL*24] use neural components to spawn Gaussians on-the-fly, decoding the anchor Gaussians' neural attributes to all parameters, including color. Compression for Gaussian representations [CWL*24, BKL*24] similarly uses neural condensation; however, not due to physical properties or in relation to spectral imaging.

Also not related to multi-spectral rendering, *FeatSplat* [MC24] is methodologically closest to our approach. They learn a feature vector per Gaussian that encodes both color and semantic information. The important difference to our method is that they first alpha-blend the feature channels into one feature image, and then decode this image in *screen space* to RGB colors—an idea also explored in temporal Gaussian processing [LCLX24]. For the multi-spectral reconstruction domain, this blending of neural surface characteristics diminishes the effect of spectral correlations between surface materials and limits spectral cross-talk.

Compared to the aforementioned approaches, we introduce a per-Gaussian neural encoding and decoding, which, to the best of our knowledge, is unexplored in the 3DGS literature. We note, however, that this color model shines in a multi-spectral or multi-model scenario and shows little benefit when employed in the RGB domain only, as also shown in our evaluation.

2.3. Multi-Modal Radiance Fields

Multi-modal strategies integrate various sensing modalities, including passive reflectance (multi-spectral), passive emission (thermal), and active illumination (LiDAR or radar). Conversely, single-modal methods utilize only one sensing modality. Multi-spectral and hyper-spectral imaging are encompassed within this category, varying in spectral resolution: multi-spectral sensors collect approximately 4–10 distinct channels, whereas hyper-spectral cameras acquire hundreds of narrower, continuous bands, up to about 600 channels covering 400 nm to 2500 nm [AMR10]. This work primarily concentrates on passive multi-spectral reflectance.

Thermal Imaging. Although our method does not specifically target it, we assess thermal imaging, which detects emissions using infrared light, owing to its recent surge in popularity. Thermal-NeRF [LPFKW24] and Xu et al. [XLKP25] use thermal data to enhance low-light reconstruction, with the latter also improving RGB through thermal cues. Ye et al. [YWD*24] focus on thermal input training alone, while Özer et al. [OWHE24] assess various NeRF architectures to find optimal multi-modal configurations across thermal, near-infra red, and depth. Both *ThermoNeRF* [HFFM24] and Thermal-NeRF [YWD*24] utilize dual neural fields for RGB and thermal synthesis. In the 3DGS realm, Thermal3D-GS [CSB24] pioneered thermal-only splatting. *ThermalGaussian* [LCZ*24] builds on this by combining RGB and thermal Gaussians. They employ spherical harmonics for separate color modeling.

Multi- and Hyperspectral Imaging. Recent advances have explored the integration of multi- and hyperspectral imagery into neural scene representations. X-NeRF [PZRT*22] introduces a cross-modal neural field that jointly models RGB, multi-spectral, and near-infra red bands, learning camera extrinsics during optimization rather than relying on pre-calibration. Similarly, Spec-NeRF [LLS*24] captures separate visible-spectrum bands while simultaneously estimating each camera's spectral sensitivity functions to enable accurate reconstruction. SpectralNeRF [LLL*24] takes a different approach by training on eleven discrete spectral channels to predict per-band spectral maps, which are then fused into high-quality RGB images using a U-Net–style architecture. Extending to hyper-spectral data, Hyperspectral NeRF [CNO*24] embeds continuous wavelength information directly into its radiance field, allowing for detailed, per-wavelength radiance predictions across more than 128 channels.

HyperGS [TML*24] abstracted and reduced the spectral high-dimensional hyperspectral data domain to a latent space using a pre-trained autoencoder. During training, the Gaussians optimize

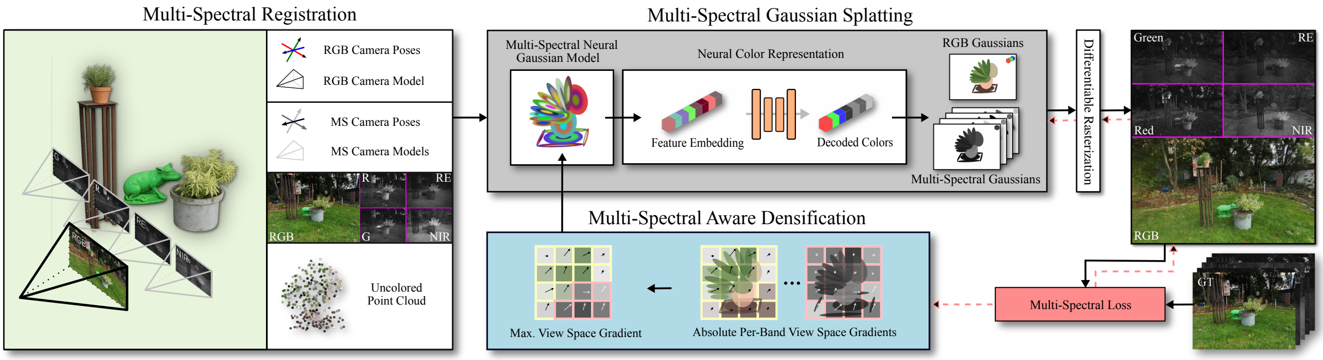

Figure 3: *Overview of MS-Splatting pipeline. After initial Structure-from-Motion Registration across all channels simultaneously, we initialize a multi-spectral neural Gaussian model shared for all spectral channels. Thereby, colors of all spectral bands are encoded in a per-Gaussian feature vector and decoded with a tiny MLP. During optimization, the shared Gaussian model and per Gaussian features are optimized by drawing a random view and spectral band under a multi-spectral loss formulation.*

latent spectral signatures, which a view-conditioned MLP maps to view-dependent spectral and opacity features that are then decoded back into full spectra at rendering time. Importantly, this method targets only object-centric scans in the hyperspectral domain, and it uses a positional neural field for color embeddings. For our domain, their methodology unfortunately does not suffice as we target *unbounded* scenes, and multi-spectral cameras lack the high spectral resolution of hyperspectral imaging. Closest to our approach is Grün et al. [GMW*25], who represent multi-spectral data in Gaussian Splatting as separate sets of spherical harmonics (similar to *ThermalGaussian* [LCZ*24] for thermal data), but optimized jointly after an initial RGB-only reconstruction. This, however, limits cross-talking effects and has high memory consumption due to the large number of SH parameters.

In contrast to all these methods, our approach is the first to use a physically inspired *neural color representation* to jointly optimize spectral bands. As we show in the results, this approach leads to superior performance in the reconstruction of multi-spectral scenes and increases quality in RGB renderings through allowing spectral cross-talk.

3. Overview

The input to our method comes from a set of n cameras, C_j. Each of these cameras has its own intrinsic parameters and covers one or multiple (spectral) bands. We aim our method to be versatile; as such, we assume no calibration or rig between the different cameras. With this, recordings can include independent trajectories, resulting in an input of n sets, $\mathcal{I} = \{\mathcal{I}_1, \mathcal{I}_2, \dots, \mathcal{I}_n\}$, where $\mathcal{I}_j$ contains images captured from camera C_j. For our recorded dataset (see Sec. 6.1) we use $n = 5$ cameras: an RGB camera, and four multi-spectral cameras capturing red (R), green (G), red edge (RE), and near-infra red (NIR). Hereby, the RGB camera captures the three wide bands, while the multi-spectral cameras capture narrower bands of the light (see Sec. 6.1 and the supplemental). In this capturing, the cameras mounted on a drone exhibit an irregular capturing pattern due to winds, drone motion and unsynchronized shutters, and are treated as independent shuffled sets.

Based on the input images, we optimize a 3D Gaussian-based scene representation with our *neural color representation* which allows for novel view synthesis across multiple spectral bands. Our framework operates in two stages: In the first step (Section 4), we compute camera poses and a sparse point cloud from $\mathcal{I}$ using classical Structure-from-Motion (SfM) [SZPF16] *without* assuming shared camera parameters across multiple sensors. Then, in the second stage (Section 5), we use previously computed poses and point cloud to initialize and train the unified *neural color representation* within our multi-spectral Gaussian Splatting model, which, eventually, yields a compact multi-spectral scene representation.

4. Multi-Spectral Camera Calibration

Existing multi-spectral and multi-modal scene representations [OWHE24, LCZ*24, LPFKW24] usually assume *shared* camera parameters (intrinsics *and* extrinsics) across multiple channels, and hence need some form of cross-modal camera calibration to superimpose images captured using different sensors. Once aligned, camera parameters are obtained by applying SfM only to the RGB images. Although simple, this workflow requires either a fixed setup and special hardware for camera calibration, or the use of error-prone methods to precisely align RGB and multi-spectral images after they have been captured.

In this work, we neither assume shared camera intrinsics nor extrinsics across multiple bands, allowing for images to be captured from different viewpoints and/or points in time, assuming the overall lighting conditions did not change drastically, e.g., no change from daytime to nighttime. Specifically, we found that processing all images in $\mathcal{I}$ with SfM *at once* leads to satisfying results if we use separate, per-camera intrinsics. This works because multi-spectral and RGB images show similar high-frequency geometric details when having a comparable spatial resolution [IA98], as is the case for most commercially available multi-spectral cameras. Thus, standard SfM's feature detection (e.g., using SIFT [Low99]) and matching is well applicable to match RGB and multi-spectral features *across spectral bands*. Please see Fig. S.F3 in the supplemental for an illustration.

The drawback of this initialization approach is that typical 3DGS initialization cannot be implemented. This occurs because the SfM point cloud, which consists of 2D features elevated and refined in 3D, is not separated across different bands. Consequently, initially coloring the point cloud becomes challenging since some feature points lack discernible RGB samples from the images, e.g., as feature points are similar in spectral bands, but wildly different in RGB. Standard procedures [SZPF16] involve reusing spectral band colors for those points, resulting in a mixed colored point cloud. We solve this issue by discarding colorization and replacing it with a brief color warm-up phase at the start of training (see Secs. 5.2 and 6.2).

5. Multi-Spectral Gaussian Splatting

The aim of our method is to optimize all available spectral channels into a single representation, which is both scalable and allows cross-spectral correlations to be exploited. Compared to grid- or neural field-informed approaches [PZRT*22, TML*24], we opt for a solely primitive-based representation, which is the base for our *shared geometry*. This particle-based representation automatically extends to unbounded scenarios and is extendable with spatial partitioning [KMK*24]. We extend this shared geometry formulation with a single *neural color representation* instead of per-band spherical harmonics, effectively enabling information to be shared *across* multiple channels. Fig. 3 provides an overview of our pipeline.

Our neural color-based model (Sec. 5.1) is first initialized (Sec. 5.2) and then optimized using a differentiably Gaussian renderer [KKLD23, YLK*25] with a randomly drawn spectral image $\mathcal{I}_j$ and our multi-spectral loss function including feature regularization (Sec. 5.3). To ensure sufficiently detailed reconstruction, we employ a multi-spectral aware densification strategy (Sec. 5.4).

5.1. Neural Color Representation

Our key insight is that RGB and multi-spectral channels profit from exchanging information between each other (spectral cross-talking) instead of being treated separately. To implement this finding, we replace traditional spherical harmonics with an optimizable, per primitive feature vector $f_i \in \mathbb{R}^d$, which stores the *view-dependent* multi-spectral radiance shared over the full vector.

In comparison to separate spherical harmonics [GMW*25, LCZ*24] using a high per-Gaussian number of parameters (usually 16 per modality/band), this necessitates compression across spectral bands and promotes contributions from various bands to refine details like appearance and edges, aiding in resolving ambiguities during reconstruction. This is further supported by our use of a low dimensionality ($d = 8$ for all bands jointly) as discussed in Sec. 6.7.

This feature vector is decoded before rendering (thus in world space), resulting in the full-spectrum color vector $\hat{\mathbf{c}}_i \in \mathbb{R}^B$, where B_j represents the channels count per camera. In our captured dataset (Sec. 6.1), this equates to $B = 7$, comprising three RGB and four multi-spectral channels. For decoding, we use a shallow MLP Φ:

$$\hat{\mathbf{c}}_i = \Phi(f_i \oplus s; \Theta), \tag{1}$$

where $s \in \mathbb{S}^2$ is the normalized outgoing viewing direction, in spherical coordinates. Θ are the learnable parameters of the MLP and $\oplus$ denotes vector concatenation.

Importantly, Φ is *not* a neural field, as it does not depend on the Gaussians' positions. Compared to NeRF [PZRT*22] or Gaussian Splatting sampling a neural field [TML*24], our formulation does *not* store multi-spectral radiance in the MLP weights Θ, but uses the MLP as an emittance function. This is similar in spirit to a BRDF, which emits light based on outgoing direction, surface parameters, and incoming light. Hereby, for us, surface parameters are combined with the incoming light in the feature vector f_i. This is an important distinction, as it separates the required size of Φ from the radiance model, thus allowing large unbound scenes only based on primitives.

For Φ, we use a shallow 3-layer MLP with one hidden layer of dimension 32. We apply exponential linear unit (ELU) activations [CUH15] to all layers except the output, which uses a sigmoid activation to constrain outputs between [0, 1]. No encodings [MESK22, TSM*20] are used.

To differentiable render the individual colors of each spectral band, we follow the default 3DGS rasterization pipeline [KKLD23]: The pixel color is obtained by alpha-blending the decoded color band $\hat{\mathbf{c}}_i$ of each contributing projected primitive.

5.2. Initialization

We initialize Gaussian primitives following Kerbl et al. [KKLD23], with positions from the SfM point cloud and scale through kNN-distance. Gaussian Splatting benefits also from good color initialization [KKLD23], which is not complete in the multi-spectral calibration scenario (see Sec. 4) due to unreliable point colors in the multi-spectral SfM point cloud.

To resolve this and stabilize reconstruction, we use a *warm-up initialization* phase. Point features f_i are initialized from a zero-mean normal distribution with a standard deviation of 0.2, and MLP weights Θ are initialized Kaiming uniformly. For the first 500 iterations, we freeze the covariance, mean, scaling, and opacity to solely let the feature vectors and MLP parameters optimize. This effectively recovers per-channel color values and successfully prevents our model from large, color-induced gradients which would cause the initial primitive positions to jump around.

We use this warm-up initialization phase in addition to the default warm-up used by Kerbl et al. [KKLD23], whereby earlier training iterations use lower-resolution images.

5.3. Loss Function and Training Regime

We employ the following loss function to train our model:

$$\mathcal{L}_{\text{MS}} = \mathcal{L}_{\text{3DGS}}(\hat{I}, I) + \lambda_{\text{norm}} \sum_{i=1}^{S} \left(\|f_i\|_2 - 1\right)^2, \tag{2}$$

where I denotes a randomly chosen image from the set $\mathcal{I}_j$ of all images captured using the j-th camera, $\hat{I}$ is the corresponding rendered image, and S is the total number of primitives in the scene. $\mathcal{L}_{\text{3DGS}}$ is the standard 3DGS loss,

$$\mathcal{L}_{\text{3DGS}}(\hat{I}, I) = (1-\lambda)\mathcal{L}_1(\hat{I}, I) + \lambda \mathcal{L}_{\text{D-SSIM}}(\hat{I}, I). \tag{3}$$

The second term in Eq. (2) regularizes the per-Gaussian feature vectors f_i. We back-propagate this loss through our model to update the parameters of the MLP Θ, feature vectors f_i, as well as the conventional 3DGS parameters $\{\mu, \Sigma, \sigma, s\}$.

When sampling the rendering camera C_j, we observe that RGB guides reconstruction quality over the whole image. Therefore, we determined empirically that the sampling weight for RGB cameras should be increased to four times that of multi-spectral cameras, so that, after four multi-spectral images, an RGB image is used.

We use $\lambda = 0.2$ and $\lambda_{\text{norm}} = 0.1$ in all our tests, and evaluate our loss function in Eq. (2) with a single randomly chosen image per iteration. For an ablation of the regularization, see the supplemental.

5.4. Multi-Spectral Aware Densification

We utilize the Gaussian densification strategy first introduced by [KKLD23] and later refined by [YLL*24], but adapt it to better respond to multi-spectral input data. In every iteration, for every rendered Gaussian G_i, we calculate the homodirectional view-space positional gradient

$$\hat{g}_i^{(j)} = \left(\sum_{k=1}^{m} \left| \frac{\partial L_k}{\partial \mu_{i,x}^{(2d)}} \right|, \sum_{k=1}^{m} \left| \frac{\partial L_k}{\partial \mu_{i,y}^{(2d)}} \right| \right), \tag{4}$$

where $j \in [n] := \{1, 2, \ldots, n\}$, $\mu_i^{(2d)}$ is the view-space position of G_i, m is the number of pixels rendered for the Gaussian, and L_k is the loss function in Eq. (2) computed for the k-th pixel. We choose a densification interval of 300 iterations, during which the calculated $\hat{g}_i^{(j)}$ are accumulated. In the next densification step, if the criterion

$$\max_{j \in [n]} \left(\frac{\sum \left\| \hat{g}_i^{(j)} \right\|}{n^{(j)}} \right) > \tau_{grad} \tag{5}$$

is met for a Gaussian, we split or clone it according to the approach laid out by [KKLD23], where $n^{(j)}$ is the number of view-space gradients accumulated in the densification interval for the current band. We choose $\tau_{grad} = 0.0008$ in our experiments.

Note that we separately accumulate view-space gradients for each spectral band, and then select the maximum average gradient for the densification criterion. Since high view-space gradients correlate with under- or over-reconstructed areas of the scene [KKLD23], this approach recognizes when a region is properly modeled for one spectral band, but is insufficiently reconstructed for another, and introduces new Gaussians to the area accordingly. Therefore, our modification to the densification strategy allows for proper reconstruction of high-frequency details that are only visible in one or a few of the captured spectral bands. See Sec. 6.7 for an a evaluation compared to the default densification strategy.

6. Evaluation

We compare our method against current state-of-the-art approaches to confirm the effectiveness of our *neural color representation* and to show that our framework can be readily extended to data from additional modalities, such as thermal. Moreover, in Sec. 7, we demonstrate how MS-Splatting can be used to compute vegetation indices for agriculture applications, and in the supplemental how to leverage our learned feature embedding to perform multi-spectral clustering to visualize areas with similar spectral information.

Our model builds upon *gsplat* [YLK*25] and is implemented within the *Nerfstudio* [TWN*23] framework. Unless otherwise noted, we trained our model for 120,000 iterations using the Adam optimizer with a learning rate of 0.005, as described in the previous section.

6.1. Multi-Spectral Datasets

6.1.1. MS-Splatting Dataset

We recorded our own dataset as as no suitable outdoor multi-view dataset with multi-spectral imagery was available. The dataset consists out of seven outdoor scenes recorded with a DJI Mavic 3M drone [DJI25]. This drone features a dual-camera system: one RGB sensor and four multi-spectral cameras that cover the red (R), green (G), red-edge (RE), and near-infra red (NIR) bands. While the green and red bands overlap with the RGB spectrum, their narrow capture bandwidth (32 nm) provides higher spectral precision. We note that spectral bands can only be captured sequentially—due to hardware restrictions—which can introduce spatial misalignment, especially in windy scenarios and due to the drone's stabilization system. Fig. S.F4 in the supplemental illustrates a typical misalignment caused by this sequential capture. Further details on the drone's imaging system, spectral band characteristics, and an overview of the dataset's evaluation images can also be found in the supplemental.

The dataset includes 360° panoramas of a garden, a dormant cherry tree, an apple orchard during bud swelling, and a lake with a construction trailer; 180° views of a house with solar panels; forward-facing sequences of an apple orchard at harvest; and a top-down survey of a golf course. An overview of the datasets can be seen in the evaluation images in the supplements. The dataset comprises out of 81-136 images per channel in total out of 405-680 images. RGB images are captured with a focal length of 24 mm, while each multi-spectral camera has a focal length of 25 mm. For training we split each dataset in 90% training and 10% evaluation data. We also release the indices of the evaluation data for all datasets. The dataset of default and warped is cropped to 1800px×1350px.

6.1.2. Dataset Overview

Voynov et al. [VBK*23] provide an indoor multi-sensor dataset containing RGB, near-infra red, and depth data; Poggi et al. [PZRT*22] offer an unposed, forward-facing indoor dataset with RGB, multi-spectral, and near-infra red channels (16 scenes, 10 bands), although they estimate poses by learning relative sensor transformations, which lie outside the scope of our work. Finally, Hyperspectral NeRF [CNO*24] captured an indoor hyperspectral dataset of eight scenes with 128 bands, which is not yet publicly released.

6.2. Comparison on Multi-Spectral Data

Methods. Multi-spectral radiance field methods rarely provide their code open-source; as such, we are limited in methods to fairly compare with. We compare against 3D Gaussian Splatting (*3DGS*) [KKLD23], trained on each spectral band individually, resulting in individual models. Furthermore, we compare against the joint-optimized setup of Grün et al. [GMW*25] who use 3DGS with

Table 1: *Results over all seven scenes of our captured dataset. See also Fig. S.F8 as well as Figs. S.F9, S.F10, S.F11, S.F12 and S.F13 in the supplemental for visual comparison. The colors indicate best and second best. Independent Cameras (IC) indicates if a method is able to fuse multiple independent cameras into the same model. For methods supporting IC, we use the original dataset, for the other we train with warped multi-spectral images. † indicates our re-implementation described in the supplementary Sec. S4.*

		PSNR (↑)						SSIM (↑)						LPIPS(↓)	SAM(↓)	SCM(↑)	SID(↓)
	IC	All	RGB	G	R	RE	NIR	All	RGB	G	R	RE	NIR	RGB	All-MS	All-MS	All-MS
ThermalGaus. [LCZ*24]	×	23.54	19.81	23.98	25.66	23.78	24.49	0.700	0.565	0.710	0.763	0.717	0.743	0.371	-	-	-
ThermoNeRF [HFFM24]	×	18.91	15.39	19.48	20.62	18.78	20.26	0.500	0.255	0.542	0.616	0.514	0.575	0.723	-	-	-
3DGS [KKLD23]	✓	23.62	20.86	22.92	25.23	23.33	25.75	0.721	0.623	0.667	0.769	0.730	0.815	0.283	0.146	0.812	0.044
ThermalGaussian †	✓	24.18	20.17	24.04	26.09	24.14	26.50	0.718	0.564	0.695	0.788	0.732	0.814	0.387	0.127	0.843	0.035
TIMS [GMW*25]	✓	24.99	21.29	24.62	26.61	25.07	27.42	0.755	0.647	0.719	0.810	0.766	0.834	0.283	0.116	0.860	0.031
MultiSpec-FeatSplat	✓	25.20	20.75	25.12	27.49	25.13	27.57	0.766	0.641	0.733	0.828	0.781	0.848	0.301	0.114	0.879	0.026
Ours	✓	25.65	21.17	25.42	27.79	25.79	28.13	0.763	0.633	0.729	0.827	0.780	0.849	0.306	0.109	0.888	0.024

shared Spherical Harmonics for spectral color channels (denoted as TIMS). A comparison to other multi-spectral methods was, due to non-availability of source code, not possible.

We further compare our method against two open-source thermal baselines: *ThermalGaussian* [LCZ*24] and *ThermoNeRF* [HFFM24]. As they normally use only RGB and thermal data, we trained *ThermalGaussian* and *ThermoNeRF* four times, pairing RGB with one additional multi-spectral band per run.

Furthermore, we re-implemented *ThermalGaussian*'s "One Multi Modal Gaussian" variant within our framework, denoted with †. However, our implementation extends each Gaussian with extra spherical harmonics to handle multiple (i.e., not just one as in the original implementation) spectral bands simultaneously. More implementation details can be found in the supplemental.

Finally, we also compare *MultiSpec-FeatSplat*, which is our multi-spectral adaptation of the approach of FeatSplat [MC24]. In this method, the Gaussian feature vectors are first rendered to a 2D image, before the MLP is applied in screen-space, producing a colored image. The MLP is also provided with the respective camera position in the scene. We utilize the same feature dimension, MLP size, training strategy, and training duration as in our method.

Datasets. We compare on our captured dataset (Sec. 6.1) as well as the X-NeRF dataset [PZRT*22]. Note that due to the unavailability of the source code and evaluation protocol, direct comparison to X-NeRF was not possible. For our dataset, we hold out every 10th image for test, while we use every 6th image for X-NeRF (see supplemental Tab. S.T5), aligning with the number of test images stated [PZRT*22]. Importantly, the test images may differ from those selected by X-NeRF, as those specifics are unknown.

To evaluate methods such as *ThermalGaussian* [LCZ*24] and *ThermoNeRF* [HFFM24] which assume shared camera parameters across all spectral bands and are therefore not directly applicable to our dataset, we first superimpose RGB and multi-spectral images by estimating pairwise homographies between each of them (i.e., between RGB + R, RGB + G, RGB + RE, and so on). We then warp multi-spectral images into the RGB frames. Due to focal-length differences, however, the warped multi-spectral images exhibit black borders; we, therefore, crop all images to 1800px × 1350px to effectively remove these margins. We trained our model and the other methods on the original, unaligned data, but cropped images to the same regions to ensure a fair comparison.

Table 2: *X-NeRF dataset [PZRT*22], results averaged on all scenes on RGB, multi-spectral (MS) and infra-red (IR). In this dataset we have 10 multi-spectral bands and 1 band for IR.*

	PSNR (↑)				SSIM (↑)				LPIPS(↓)	SAM(↓)	SCM(↑)	SID(↓)
	All	RGB	MS	IR	All	RGB	MS	IR	RGB	All-MS	All-MS	All-MS
3DGS	30.99	35.36	30.54	31.14	0.896	0.948	0.888	0.922	0.243	0.091	0.816	0.049
TG.†	31.47	32.87	31.51	29.63	0.916	0.929	0.914	0.918	0.311	0.065	0.869	0.021
TIMS	29.07	31.45	29.02	27.20	0.889	0.932	0.884	0.895	0.288	0.069	0.858	0.021
Ours	32.02	32.88	32.15	29.93	0.930	0.940	0.929	0.926	0.280	0.063	0.879	0.016

For the other Gaussian Splatting-based methods, we also employ the warmup strategy described in Sec. 5.2 optimizing the SH base color; otherwise, instabilities occur.

Metrics. To evaluate our results, we relied on both color, perceptual and spectral metrics on the test images. On the one hand, we utilized standard metrics such as PSNR, SSIM, and LPIPS. While LPIPS is a learned perceptual metric trained on RGB images, it is not very expressive in the settings of multi-spectral imagery, but it is included for the sake of completeness in the supplemental. On the other hand, we evaluated the multi-spectral images in terms of their spectral accuracy. Therefore, we employed the spectral-similarity metrics Spectral Angle Mapper (SAM) [KLB*93], Spectral Correlation Mapper (SCM) [CJM00], and Spectral Information Divergence (SID) [Cha99]. A detailed explanation of these metrics is provided in the supplementary material.

Results. Tab. 1 presents the results, averaged over all scenes of our dataset. RGB metrics for *ThermalGaussian* and *ThermoNeRF* were computed by averaging respective results over the four runs. As seen, our approach improves PSNR by over 1.4 dB, increasing SSIM by 6 %, and reducing LPIPS by 23 % compared to the previous state-of-the-art *ThermalGaussian*, and also outperforms TIMS [GMW*25] as well as our screen-space splat implementation. Notably, even on RGB data alone, our method outperforms vanilla 3DGS in PSNR and SSIM, demonstrating the cross-spectral benefit of incorporating additional bands. Furthermore, our method provides state-of-the-art results on all three spectral metrics, showcasing exceptional spectral reconstruction quality.

Beyond improvements in perceptual metrics, we also demonstrate enhancements in spectral-similarity metrics across all compared methods. Compared to the state-of-the-art *ThermalGaussian*, we reduce the spectral angle (SAM) by 17%, increase the spectral correlation (SCM) by 5%, and shrink the spectral information divergence (SID) by 50%. This highlights the stronger joint representational power of our approach on RGB and multi-spectral data. A per dataset listing of the spectral similarity metrics can be found in the supplemental in Tab. S.T17. In addition, we provide radar (spider) charts – analogous to spectral response curves of hyperspectral cameras – showing raw values across RGB and multi-spectral pixels in Fig. S.F6 and per-band errors in Fig. S.F7.

Visual comparisons for all spectral channels are shown in Fig. 4. Hereby, our method is able to reconstruct the fine leaves of the fruit treetop in all bands (top, blue crop), as well as provides great background reconstructions (bottom, red crop). Further visual comparisons are in the supplemental in Fig. S.F8, S.F9, S.F10 S.F11, S.F12 and S.F13. Per-scene results are provided in the supplements.

In Tab. 2, we evaluate 3DGS, *ThermalGaussian*†, TIMS, and our method on all scenes of the X-NeRF dataset [PZRT*22]. For 3DGS, we trained each spectral band independently, including the multi-spectral channels. More information on the training can be found in the supplemental Sec. S5. Overall, the results demonstrate that our method achieves the best average PSNR and SSIM across RGB, multi-spectral, and infrared (IR) data. Compared to the 3DGS baseline, we improve the PSNR on multi-spectral data by 1.7 dB. The slightly better results on RGB and IR can be attributed to the noisy registration of RGB with multi-spectral and IR images, since the multi-spectral images with a size of 510px$\times$240px are relatively small, IR has a low resolution with 1024px$\times$1024px with a large field of view and as we treat the 10 multi-spectral bands and the single IR band as separate camera models. For the multi-spectral images, we further show that our method achieves the best scores on spectral-similarity metrics, confirming once more the representational power of our proposed *neural color representation*. Additional tables with per-scene spectral-similarity results as well as per-band evaluations are provided in the supplementary material.

6.3. Comparison on Thermal Data

We also compared the original implementations of ThermalGaussian [LCZ*24] and *ThermoNeRF* [HFFM24] against our method on thermal data. To do so, we selected three scenes from the ThermalMix dataset [OWHE24] and three scenes from the dataset used in *ThermoNeRF* [HFFM24] (details can be found in the supplementary). Here, the datasets do provide calibrated pairs of thermal and RGB images, and we used the provided camera parameters for all methods. We evaluated our approach using the same configuration as in Sec. 6.2. Additionally, we tested the influence of the smoothness loss $\mathcal{L}_{\text{smooth}}$ introduced by [LCZ*24] on our method. More information can be found in the supplements.

The averaged results across all scenes are presented in Tab. 3, with per-scene details provided in the supplementary material. Visual comparisons can be found in Fig. 5 and show sharper RGB reconstructions (e.g., PAN, blue crop) and less noisy thermal results (FACE, red crop and background). As presented, our approach outperforms both *ThermalGaussian* and *ThermoNeRF*, with the exception of comparable quality in thermal PSNR. This demonstrates that MS-Splatting generalizes to other modalities without further adaptation. Please note that thermal LPIPS was evaluated on grayscale images; thus, its expressiveness is limited.

Table 3: *Average performance across six thermal scenes from the ThermalMix dataset [OWHE24] and the dataset from ThermoNeRF [HFFM24]. Details can be found in the supplemental.*

Method	PSNR (↑) RGB	PSNR (↑) T	SSIM (↑) RGB	SSIM (↑) T	LPIPS (↓) RGB	LPIPS (↓) T
ThermalGaussian	24.91	31.34	0.824	0.931	0.226	0.107
ThermoNeRF	16.54	22.45	0.546	0.765	0.376	0.295
Ours	26.31	31.18	0.883	0.934	0.166	0.104
Ours + $\mathcal{L}_{\text{smooth}}$	26.66	31.23	0.892	0.935	0.157	0.095

6.4. Spectral Cross-Talking Evaluation

To further investigate multi-spectral rendering in our method and the effects of spectral correlations and spectral crosstalk, we conducted the following study on our captured dataset. We systematically extend RGB with successively added different spectral bands in every permutation. The results of this study can be found in Tab. 4.

Only using RGB with our *neural color representation* decreases quality by 0.28 dB, suggesting that neurally compressing from 3DGS' full Spherical Harmonics (48 floats) to our neural feature vector ($d = 8$ floats) introduces slight quality loss. To achieve optimal RGB rendering, combining it with a narrow band red channel (R) and near-infra red (NIR) yields superior results, outperforming baseline 3DGS by 0.76 dB and the neural variant by over 1 dB. Interestingly, the NIR and R channels directly formulate the Normalized Vegetation Index (refer to Sec. 7), providing insights into plant health. This implies our approach benefits from knowledge of plant material characteristics for enhanced RGB NVS.

For all bands, the overall quality drops slightly across channels compared to the ideal configurations. While this suggests too strong compression, increasing feature vector sizes do benefit results (see Sec. 6.7). All in all, this configuration still shows a great tradeoff between multi-spectral reconstruction quality and ease of use.

Table 4: *Spectral cross-talking effects with RGB, green (G), red (R), red-edge (RE), and near-infra red (NIR) on MS-Splatting. The colors indicate the best and worst results.*

	PSNR (↑)					SSIM (↑)					LPIPS (↓)
	RGB	G	R	RE	NIR	RGB	G	R	RE	NIR	RGB
RGB	20.58	-	-	-	-	0.611	-	-	-	-	0.316
RGB + G	21.24	25.13	-	-	-	0.641	0.720	-	-	-	0.283
RGB + R	21.51	-	27.77	-	-	0.649	-	0.827	-	-	0.273
RGB + RE	21.45	-	-	25.65	-	0.650	-	-	0.777	-	0.276
RGB + NIR	21.47	-	-	-	28.14	0.644	-	-	-	0.845	0.287
RGB + G + R	21.38	25.11	27.66	-	-	0.640	0.726	0.826	-	-	0.289
RGB + G + RE	21.30	25.35	-	25.59	-	0.643	0.727	-	0.776	-	0.290
RGB + G + NIR	21.48	25.48	-	-	28.17	0.645	0.730	-	-	0.847	0.289
RGB + R + RE	21.52	-	27.75	25.73	-	0.650	-	0.829	0.781	-	0.283
RGB + R + NIR	21.62	-	27.84	-	28.16	0.651	-	0.832	-	0.850	0.284
RGB + RE + NIR	21.34	-	-	25.86	28.23	0.645	-	-	0.782	0.847	0.291
RGB + G + R + RE	21.27	25.18	27.54	25.46	-	0.642	0.729	0.828	0.779	-	0.293
RGB + G + R + NIR	21.30	25.27	27.67	-	27.88	0.642	0.731	0.830	-	0.849	0.292
RGB + G + RE + NIR	21.24	25.30	-	25.79	28.06	0.642	0.729	-	0.783	0.850	0.298
RGB + R + RE + NIR	21.33	-	27.59	25.48	27.76	0.644	-	0.829	0.783	0.850	0.293
All Bands	21.17	25.42	27.79	25.79	28.13	0.633	0.729	0.827	0.780	0.849	0.306

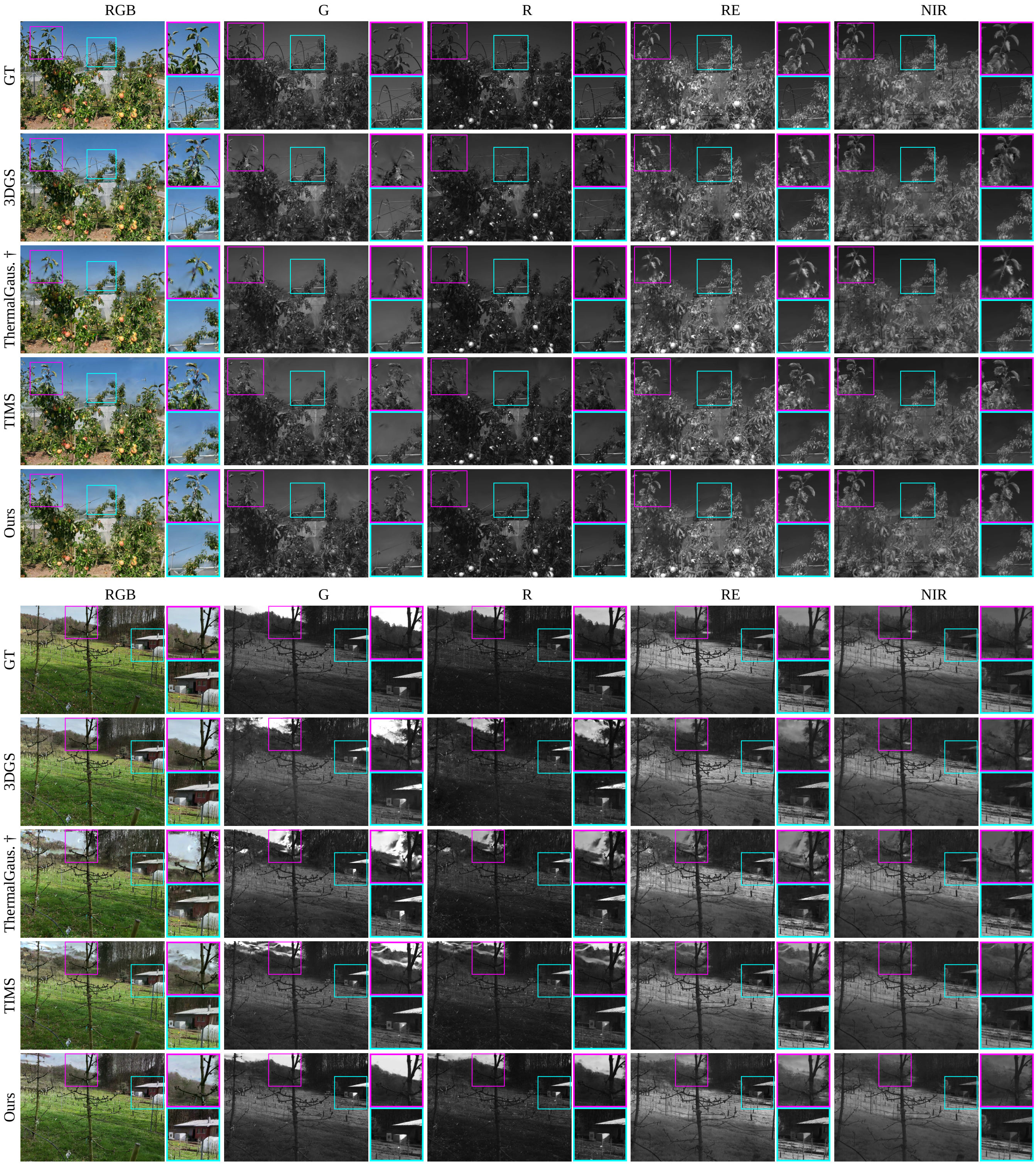

Figure 4: *Visual comparison on the FRUIT TREES (top) and SINGLE TREE (bottom) scenes with all available spectral-bands. The used ThermalGaussian method in this comparison is our multi-spectral re-implementation.*

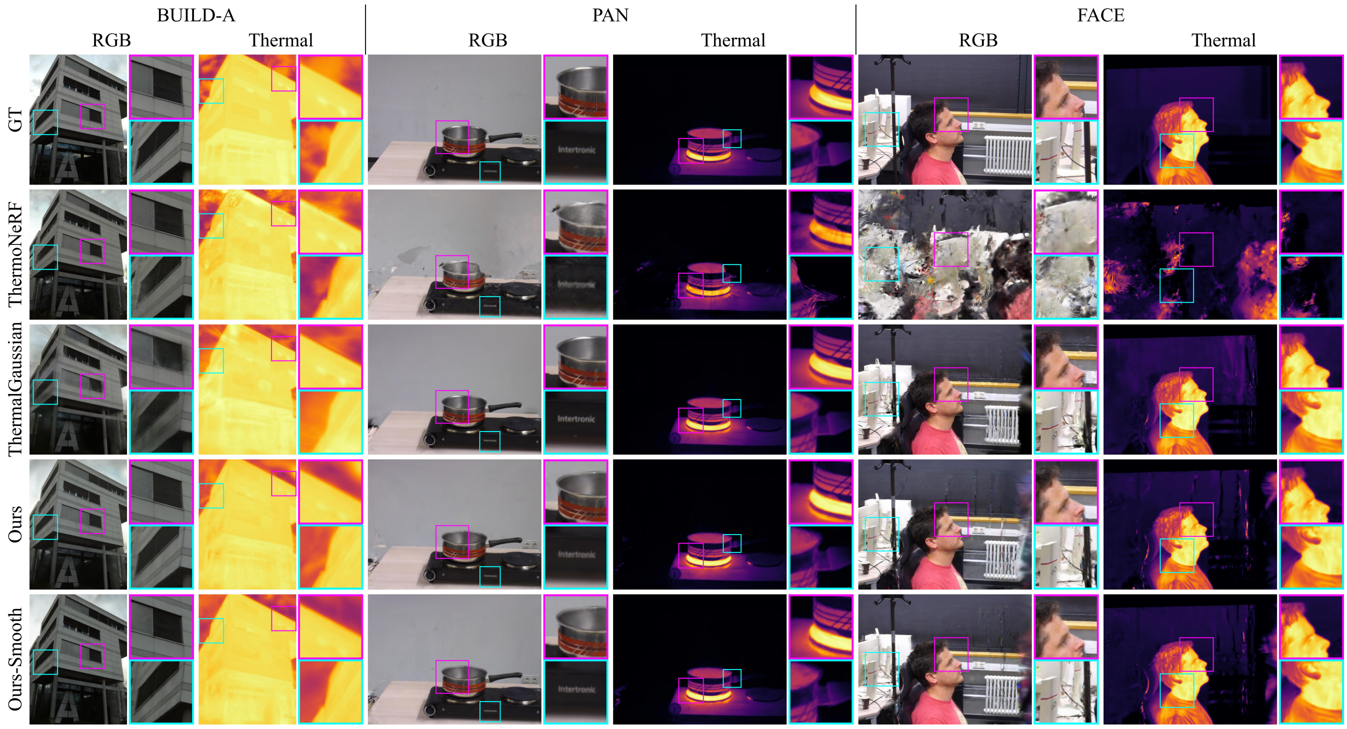

Figure 5: *Visual comparison of the thermal scenes BUILD-A, PAN and FACE on ThermalGaussian, ThermoNeRF and ours. The used ThermalGaussian method is the original implementation by [LCZ*24]. Ours-Smooth uses the smoothness loss proposed by [LCZ*24].*

6.5. Memory & Training Time

When considering real-time visualization on edge devices, such as remote controls in the agricultural domain (e.g., immersive NDVI exploration, see Sec. 7), the efficiency and compression capabilities of our method are of particular importance. To this end, we analyzed the memory consumption of the evaluated methods on two datasets (Lake and Garden) after training. Compared to 3DGS, our method MS-Splatting requires only 22% of the original storage. Relative to TIMS, we achieved an 88% reduction on the Lake scene, decreasing memory usage from 2.7 GB to just 326 MB. This remarkable efficiency results from our *neural color representation*, which jointly encodes cross-spectral information and proves more effective than spherical harmonics. By contrast, the immense storage demand of TIMS and *ThermalGaussian*† arises from the large number of spherical harmonic coefficients (16 floats per channel/band per Gaussian). An overview of the required storage and the number of primitives per method is provided in the supplements in Tab. S.T20.

In addition to reduced memory consumption, our method also achieves faster training. Averaged over all datasets, MS-Splatting reaches a training time of 02:30:23, which is 16% faster than *ThermalGaussian*† (02:59:16) and 38% faster than TIMS (04:01:35). This speedup originates from the optimization of our compact MLP and feature vectors, in contrast to the substantially larger number of spherical harmonic coefficients.

6.6. Sparse Captures

We conducted an experiment in which we lowered the amount of non-RGB spectral images for reconstruction to evaluate performance under the assumption that RGB captures are cheaper and more easily available. As seen in Tab. S.T21 in the supplemental, even with an 80% reduction in images, the overall reconstruction quality of spectral channels only drops by about 2-3 dB, which still enables passable renderings.

6.7. Ablation

In this section, we evaluate the impact of our key design choices. Specifically, we assess the new densification strategy and the effect of the dimension of the feature embedding and network size. Additional ablations can be found in the supplements.

Densification Strategy To evaluate our new multi-spectral densification strategy, we compared the standard 3DGS against our *Multi-Spectral Aware Densification* strategy in Tab. 5. From the results shown, we conclude that our new densification strategy slightly improves our overall performance and peaks at a densification interval of 300. The larger interval compared to 3DGS can be attributed to the average being accumulated over multiple steps to compute the gradient correctly.

Feature Dimension To isolate the effect of the feature embedding dimension, we swept this parameter while keeping the MLP architecture fixed (one hidden layer of width 32). Fig. 6 shows the

Table 5: *Densification strategy with alternating densification intervals. 3DGS (Default) uses an interval of 100 iterations.*

Split Interval	PSNR	SSIM	LPIPS
Default-100	25.02	0.748	0.310
Default-200	25.02	0.748	0.317
Default-300	25.08	0.748	0.319
Default-400	24.93	0.749	0.320
Default-500	24.86	0.747	0.323
Ours - 100	24.93	0.750	0.267
Ours - 200	25.08	0.754	0.271
Ours - 400	25.09	0.757	0.275
Ours - 300	25.16	0.758	0.271
Ours - 500	24.85	0.756	0.281

Table 6: *Impact of MLP size (L: Layer depth, W: Layer width) on rendering quality. Results are averaged over all datasets.*

Hidden Size	PSNR	SSIM	LPIPS
$L=0, W=16$	24.89	0.747	0.312
$L=0, W=32$	24.91	0.746	0.314
$L=0, W=64$	24.96	0.747	0.312
$L=0, W=128$	24.97	0.747	0.311
$L=1, W=16$	24.59	0.742	0.317
$L=1, W=32$	25.02	0.748	0.310
$L=1, W=64$	25.03	0.748	0.311
$L=1, W=128$	25.04	0.749	0.310
$L=2, W=16$	24.97	0.747	0.312
$L=2, W=32$	24.74	0.746	0.314
$L=2, W=64$	24.81	0.746	0.315
$L=2, W=128$	23.85	0.721	0.373

relative change in PSNR and LPIPS (SSIM follows a similar trend with smaller fluctuations) as functions of the embedding size d. The results indicate that a shallow MLP with $d = 8$ yields the best overall performance. Notably, even $d = 2$ provides sufficient capacity to encode substantial information across all input bands. Please note that even though we use seven distinct bands in our dataset, representing the view-dependent band radiance can require more than seven parameters, similar to 3DGS using 48 parameters for view-dependent RGB.

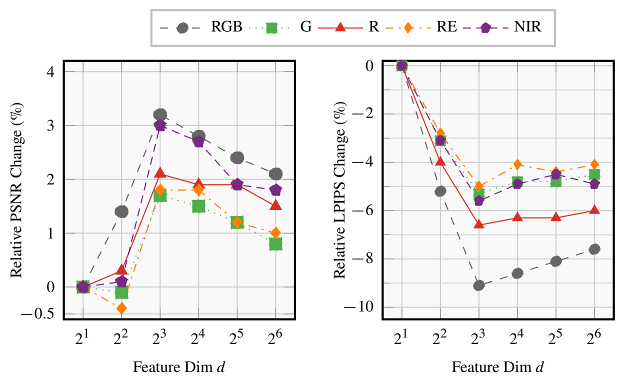

Figure 6: *Relative changes in PSNR and LPIPS for the lightweight MLP and with varying feature-embedding dimension d, reported in percentage (%). The first feature dimension, $d = 2^1$, is used as the reference for both plots. All reported changes are relative to this baseline.*

Network Size To assess the impact of network capacity, we evaluated our pipeline using MLPs of various sizes. Specifically, we varied the number of hidden layers (0, 1, or 2) and the width of each layer (from 16 to 128 neurons). Tab. 6 summarizes the results, averaged over all datasets. Overall, a single hidden layer yields the best trade-off across configurations. Extremely shallow networks (no hidden layers) or very deep, wide networks (two layers of width 128) struggle to process to accurate color representations, resulting in washed-out and dull renderings.

7. Agriculture Applications

For MS-Splatting, we regard the agricultural domain as the most promising field of application. Novel view synthesis has recently been adopted in the agricultural domain, with applications ranging from horticulture [MGSS24, MAWS25, COMK24], to greenhouse monitoring [SHZ*23, ZAQ*24, CPPL24], in-field crop observation [YLX*24, ERC*23, BKW24], and laboratory-based plant phenotyping [OLMS24, SSOA*23].

Vegetation Indices. We place particular emphasis on computing various Vegetation Indices (VI). VIs are a spectral imaging transformation designed to quantify vegetation type, plant mass, and health in a scene [JH91]. When light penetrates a surface, it is partly reflected, absorbed, or transmitted depending on the material and wavelength. Soil typically absorbs and reflects visible wavelengths while only little light is being transmitted, whereas vegetation absorbs most viable light but strongly reflects and transmits near-infra red wavelengths. This pronounced NIR reflectance arises from the internal leaf's structure, which efficiently scatters light [JH91].

Consequently, many vegetation indices, such as the normalized difference vegetation index (NDVI), the green NDVI (GNDVI), and the soil-adjusted vegetation index (SAVI), all exploit the strong NIR reflectance of vegetation. In the simplest form, NDVI is computed on a per-pixel basis as

$$\mathrm{NDVI} = \frac{\mathrm{NIR} - R}{\mathrm{NIR} + R}. \tag{6}$$

This formula implicitly assumes that the red and NIR sensors are perfectly co-registered or that the scene distance is large relative to the sensor baseline, as is typically the case for top-down drone with high altitude or satellite imagery. However, in near-field scenarios (orchards, greenhouses, phenotyping rigs), separate multi-spectral sensors often capture images from slightly different viewpoints. Without precise image registration, even small misalignment (for example, due to wind or vehicle motion) can introduce significant artifacts.

Evaluation. Mutual information (MI)–based registration is deemed the baseline, since MI is well known for aligning multi-modal imagery without assuming any specific intensity relationships between channels [KP07, ZF03, MCV*97]. Details are provided in the supplementary material, however, results as shown in the left panel of Fig. S.F4 of the supplement are unsatisfying.

To avoid these registration challenges entirely, we exploit our MS-Splatting model for NDVI (and other VI) rendering. An example NDVI output from MS-Splatting is shown in Fig.7. For qualitative evaluation, NDVI through NVS relies on the NVS quality of near-infra red and narrow-band red, for which the evaluation of Tabs. 1 and 2 showcase state of the art results, outperforming related works by more than 0.5 dB. We also note that for NDVI rendering, only reconstructing the model from RGB, R and NIR show even slightly better results (see Tab. 4).

The supplement included several additional experiments: As we test in Sec. S8 our representation is also able to achieve great results in a synthetic setup.

Figure 7: *NDVI rendered by synthesizing the red and near-infra red channels via MS-SPLATTING from a novel perspective. Values between* -1 *and* 0 *indicate inanimate object,* 0 *to* 0.33 *diseased,* 0.33 *to* 0.66 *moderately healthy and* 0.66 *to* 1 *very healthy plants.*

8. Limitations and Future Work

One limitation of our work is that, despite strong quantitative performance, the reconstructed RGB images can occasionally appear dull. We attribute this effect to the fact that both the per-primitive feature embeddings and the neural network continue to adapt until the end of training, allowing subtle color shifts caused by spectral cross-talk to persist. This behavior arises because extremely shallow networks tend to underfit color, whereas very deep and wide networks may require substantially more training iterations to achieve accurate color reproduction. Incorporating a multispectral-aware color loss or an additional RGB color-regularization term could mitigate these effects and help better preserve the original color appearance. In the same context, it would be interesting to investigate whether all rasterized Gaussian's are relevant for all spectral bands, or whether certain Gaussian's can be ignored for specific bands (e.g., NIR, which is typically less textured). This could be achieved by additionally optimizing a per-band opacity term [MSDT*25] that selectively amplifies or suppresses individual Gaussian's for particular spectral channels.

Additionally in the pipeline of SfM, feature matching across multi-spectral images should be investigated further. While the matching is effective and reliable when images are captured from approximately the same viewpoint and with similar (high) spatial resolution, it remains to be investigated how strongly varying image resolutions and significantly different camera poses affect reconstruction quality.

For future work, it would be valuable to evaluate this framework on truly high-dimensional hyperspectral data. Additionally, following the example of Hyperspectral NeRF [CNO*24], learning a continuous per-wavelength representation could enable interpolation between wavelengths and further improve spectral reconstruction. In agricultural settings, a real-time implementation (e.g., LiveNVS [FRF*23]) or even an immersive multi-spectral Gaussian splatting system [FFS25] could make this use case far more accessible. While our approach is able to render in real-time, this was not the focus of our work. As such, performance can be increased with e.g. selected MLP evaluations based on frustum and/or occlusion culling.

Moreover, incorporating hierarchical partitioning methods [KMK*24] may enable larger unrestricted captures. While our current setup manages memory conservatively, this extension would be beneficial with the advent of drones with enhanced battery capabilities for long capture sequences.

Finally, generating artificial dataset [MEY*24] using 3D Gaussian Splatting could benefit from the extension to multi-spectral data and the improved RGB rendering quality.

9. Conclusion

To conclude, we present MS-SPLATTING, a novel pipeline for view synthesis on multi-spectral images. At its core lies a unified *neural color representation* that jointly encodes all spectral channels into a shared feature embedding. A tiny MLP then decodes this embedding to accurately recover each band's appearance. Our extensive experiments show that this collaboration through a joint feature space leverages both spatial and spectral coherence, yielding consistent cross-spectral improvements across every evaluation metric. Furthermore, we introduced a multi-spectral outdoor dataset spanning RGB, green, red, red-edge, and near-infra red bands, which together with our code will be available openly.

Acknowledgments

We thank **Martin Hundhausen** for supporting the scene recordings. Lukas Meyer was funded by the 5G innovation program of the German Federal Ministry for Digital and Transport under the funding code 165GU103C. Maximilian Weiherer was funded by the German Federal Ministry of Education and Research (BMBF), FKZ: 01IS22082 (IRRW). Linus Franke was supported in part by the ERC Advanced Grant NERPHYS (101141721, https://project.inria.fr/nerphys). The authors are responsible for the content of this publication. The authors gratefully acknowledge the scientific support and HPC resources provided by the Erlangen National High Performance Computing Center (NHR@FAU) of the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) under the NHR project *b162dc*. NHR funding is provided by federal and Bavarian state authorities. NHR@FAU hardware is partially funded by the German Research Foundation (DFG) – 440719683. Open Access funding enabled and organized by Projekt DEAL.

References

[AACS*24] ATHERTON J., ALONSO CHORDA L., SUOMALAINEN J., MIETTINEN I., KUURASUO J., HAKALA T.: Dji mavic 3 multispectral edition spectral response. Zenodo, 2024. URL: https://doi.org/10.5281/zenodo.11102543, doi:10.5281/zenodo.11102543. 16

[AMR10] ADAM E., MUTANGA O., RUGEGE D.: Multispectral and hyperspectral remote sensing for identification and mapping of wetland vegetation: a review. *Wetlands Ecology and Management 18*, 3 (jun 2010), 281–296. 3

[Anu70] ANUTA P. E.: Spatial registration of multispectral and multitemporal digital imagery using fast fourier transform techniques. *IEEE Transactions on Geoscience Electronics 8*, 4 (1970), 353–368. doi:10.1109/TGE.1970.271435. 2

[AQT24] ALBANWAN H., QIN R., TANG Y.: Image fusion in remote sensing: An overview and meta analysis, 2024. URL: https://arxiv.org/abs/2401.08837, arXiv:2401.08837. 2

[ASK*20] Aliev K.-A., Sevastopolsky A., Kolos M., Ulyanov D., Lempitsky V.: Neural point-based graphics. In *European Conference on Computer Vision (ECCV)* (Cham, 2020), Springer, Springer International Publishing, pp. 696–712. 3

[BKL*24] Bagdasarian M. T., Knoll P., Li Y.-H., Barthel F., Hilsmann A., Eisert P., Morgenstern W.: 3dgs.zip: A survey on 3d gaussian splatting compression methods, 2024. arXiv:2407.09510. 3

[BKW24] Baar S., Kobayashi Y., Watanabe S.: Combining close range uavs and gaussian splatting for high-speed tomato leaf size estimation and counting. In *2024 IEEE 13th Global Conference on Consumer Electronics (GCCE)* (2024), pp. 215–218. doi:10.1109/GCCE62371.2024.10760832. 11

[BMT*21] Barron J. T., Mildenhall B., Tancik M., Hedman P., Martin-Brualla R., Srinivasan P. P.: Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)* (2021), pp. 5855–5864. 3

[BMV*22] Barron J. T., Mildenhall B., Verbin D., Srinivasan P. P., Hedman P.: Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2022), pp. 5470–5479. 3

[BMV*23] Barron J. T., Mildenhall B., Verbin D., Srinivasan P. P., Hedman P.: Zip-nerf: Anti-aliased grid-based neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)* (October 2023), pp. 19640–19648. 3

[Cha99] Chang C.-I.: Spectral information divergence for hyperspectral image analysis. In *IEEE 1999 International Geoscience and Remote Sensing Symposium. IGARSS'99 (Cat. No.99CH36293)* (1999), vol. 1, pp. 509–511 vol.1. doi:10.1109/IGARSS.1999.773549. 7, 17

[CJM00] Carvalho Jr. O. A. d., Meneses P. R.: Spectral correlation mapper (scm): An improvement on the spectral angle mapper (sam). In *Ninth JPL Airborne Earth Science Workshop* (Pasadena, CA, 2000), Jet Propulsion Laboratory, pp. 65–74. 7, 17

[CNO*24] Chen G., Narayanan S. K., Ottou T. G., Missaoui B., Muriki H., Pradalier C., Chen Y.: Hyperspectral neural radiance fields. *arXiv preprint arXiv:2403.14839* (2024). 3, 6, 12

[COMK24] Chopra S., Ojeda F. C., Murali V., Kumar V.: Agrinerf: Neural radiance fields for agriculture in challenging lighting conditions. *ArXiv abs/2409.15487* (2024). URL: https://api.semanticscholar.org/CorpusID:272831842. 11

[CPPL24] Choi H.-B., Park J.-K., Park S. H., Lee T.: Nerf-based 3d reconstruction pipeline for acquisition and analysis of tomato crop morphology. *Frontiers in Plant Science 15* (10 2024). doi:10.3389/fpls.2024.1439086. 11

[CSB24] Chen Q., Shu s., Bai X.: Thermal3d-gs: Physics-induced 3d gaussians for thermal infrared novel-view synthesis. In *European Conference on Computer Vision* (2024). 3

[CUH15] Clevert D.-A., Unterthiner T., Hochreiter S.: Fast and accurate deep network learning by exponential linear units (elus). *arXiv: Learning* (2015). URL: https://api.semanticscholar.org/CorpusID:5273326. 5

[CW24] Chen G., Wang W.: A survey on 3d gaussian splatting. *arXiv preprint arXiv:2401.03890* (2024). 3

[CWL*24] Chen Y., Wu Q., Lin W., Harandi M., Cai J.: Hac: Hash-grid assisted context for 3d gaussian splatting compression. In *European Conference on Computer Vision* (2024). 3

[DHRK24] Dalal A., Hagen D., Robbersmyr K. G., Knausgård K. M.: Gaussian splatting: 3d reconstruction and novel view synthesis, a review. *arXiv preprint arXiv:2405.03417* (2024). 3

[DJI23] DJI: *Mavic 3M Image Processing Guide*. Shenzhen, China, Aug. 2023. Version 1.0. 19

[DJI24] DJI: *DJI Mavic 3M User Manual, v1.8*, Aug. 2024. Version 1.8, English. URL: https://dl.djicdn.com/downloads/DJI_Mavic_3_Enterprise/20221216/DJI_Mavic_3M_User_Manual_EN.pdf. 16

[DJI25] DJI: Mavic 3 multispectral (3m). https://ag.dji.com/mavic-3-m, 2025. Accessed: 2025-01-24. 6, 16

[ERC*23] Esser F., Rosu R. A., Cornelißen A., Klingbeil L., Kuhlmann H., Behnke S.: Field robot for high-throughput and high-resolution 3d plant phenotyping: Towards efficient and sustainable crop production. *IEEE Robotics; Automation Magazine 30*, 4 (Dec. 2023), 20–29. URL: http://dx.doi.org/10.1109/MRA.2023.3321402, doi:10.1109/mra.2023.3321402. 11

[FFS25] Franke L., Fink L., Stamminger M.: Vr-splatting: Foveated radiance field rendering via 3d gaussian splatting and neural points. *Proc. ACM Comput. Graph. Interact. Tech. 8*, 1 (May 2025). URL: https://doi.org/10.1145/3728302, doi:10.1145/3728302. 12

[FRF*23] Fink L., Rückert D., Franke L., Keinert J., Stamminger M.: Livenvs: Neural view synthesis on live rgb-d streams. 12

[FRFS24] Franke L., Rückert D., Fink L., Stamminger M.: Trips: Trilinear point splatting for real-time radiance field rendering. *Computer Graphics Forum 43*, 2 (2024). doi:https://doi.org/10.1111/cgf.15012. 3

[FXZ*24] Fei B., Xu J., Zhang R., Zhou Q., Yang W., He Y.: 3d gaussian as a new vision era: A survey. *arXiv preprint arXiv:2402.07181* (2024). 3

[GMW*25] Grün J., Meyer L., Weiherer M., Egger B., Stamminger M.: Towards integrating multi-spectral imaging with gaussian splatting. In *Proceedings of the Vision, Modeling, and Visualization Conference (VMV)* (Sept. 2025), Eurographics Association. URL: https://meyerls.github.io/towards_multi_spec_splat. 2, 4, 5, 6, 7, 18, 24, 26

[GSC*07] Goesele M., Snavely N., Curless B., Hoppe H., Seitz S. M.: Multi-view stereo for community photo collections. In *2007 IEEE 11th International Conference on Computer Vision* (2007), IEEE, pp. 1–8. 3

[HFFM24] Hassan M., Forest F., Fink O., Mielle M.: Thermonerf: Multimodal neural radiance fields for thermal novel view synthesis, 2024. arXiv:2403.12154. 3, 7, 8, 18

[HFW*25] Hahlbohm F., Friederichs F., Weyrich T., Franke L., Kappel M., Castillo S., Stamminger M., Eisemann M., Magnor M.: Efficient perspective-correct 3d gaussian splatting using hybrid transparency. *Computer Graphics Forum 44*, 2 (2025). doi:10.1111/cgf.70014. 3

[HRLL20] Huang Y., Ren Z., Li D., Liu X.: Phenotypic techniques and applications in fruit trees: a review. *Plant Methods 16*, 1 (2020), 107. URL: https://doi.org/10.1186/s13007-020-00649-7, doi:10.1186/s13007-020-00649-7. 2

[IA98] Irani M., Anandan P.: Robust multi-sensor image alignment. In *Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)* (1998), pp. 959–966. doi:10.1109/ICCV.1998.710832. 4

[JH91] Jackson R. D., Huete A. R.: Interpreting vegetation indices. *Preventive Veterinary Medicine 11*, 3 (1991), 185–200. URL: https://www.sciencedirect.com/science/article/pii/S0167587705800042, doi:https://doi.org/10.1016/S0167-5877(05)80004-2. 11

[KKLD23] Kerbl B., Kopanas G., Leimkühler T., Drettakis G.: 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics 42*, 4 (July 2023). URL: https://repo-sam.inria.fr/fungraph/3d-gaussian-splatting/. 1, 2, 3, 5, 6, 7, 18, 24, 26

[KLB*93] Kruse F., Lefkoff A., Boardman J., Heidebrecht K., Shapiro A., Barloon P., Goetz A.: The spectral image processing system (sips)—interactive visualization and analysis of imaging spectrometer data. *Remote Sensing of Environment 44*, 2 (1993), 145–163. Airbone Imaging Spectrometry. 7, 17

[KLR*22] Kopanas G., Leimkühler T., Rainer G., Jambon C., Drettakis G.: Neural point catacaustics for novel-view synthesis of reflections. *ACM TOG 41*, 6 (2022). 3

[KMK*24] Kerbl B., Meuleman A., Kopanas G., Wimmer M., Lanvin A., Drettakis G.: A hierarchical 3d gaussian representation for real-time rendering of very large datasets. *ACM Transactions on Graphics (TOG) 43*, 4 (2024), 1–15. 1, 3, 5, 12

[KP07] Kern J. P., Pattichis M. S.: Robust multispectral image registration using mutual-information models. *IEEE Transactions on Geoscience and Remote Sensing 45*, 5 (2007), 1494–1505. doi:10.1109/TGRS.2007.892599. 11, 19

[KPK*24] Kulhanek J., Peng S., Kukelova Z., Pollefeys M., Sattler T.: WildGaussians: 3D gaussian splatting in the wild. In *Proceedings of the 38th International Conference on Neural Information Processing Systems* (2024). 3

[KPLD21] Kopanas G., Philip J., Leimkühler T., Drettakis G.: Point-based neural rendering with per-view optimization. *Computer Graphics Forum* (2021). doi:10.1111/cgf.14339. 3

[KRS*24] Kheradmand S., Rebain D., Sharma G., Sun W., Tseng Y.-C., Isack H., Kar A., Tagliasacchi A., Yi K. M.: 3d gaussian splatting as markov chain monte carlo. In *Advances in Neural Information Processing Systems (NeurIPS)* (2024). Spotlight Presentation. 3

[Kuu25] Kuurasuo J. P.: *Towards the Development of a Reflectance Estimation Pipeline for Multispectral and RGB Drone Imagery*. Master's thesis, University of Helsinki, Feb. 2025. URL: http://hdl.handle.net/10138/593074. 16

[LCLX24] Li Z., Chen Z., Li Z., Xu Y.: Spacetime gaussian feature splatting for real-time dynamic view synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2024), pp. 8508–8520. 3

[LCZ*24] Lu R., Chen H., Zhu Z., Qin Y., Lu M., Zhang L., Yan C., Xue A.: Thermalgaussian: Thermal 3d gaussian splatting, 2024. URL: https://arxiv.org/abs/2409.07200, arXiv:2409.07200. 2, 3, 4, 5, 7, 8, 10, 17, 18

[LKLR24] Luiten J., Kopanas G., Leibe B., Ramanan D.: Dynamic 3d gaussians: Tracking by persistent dynamic view synthesis. In *3DV* (2024). 3

[LLL*24] Li R., Liu J., Liu G., Zhang S., Zeng B., Liu S.: Spectralnerf: physically based spectral rendering with neural radiance field. In *Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence and Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence and Fourteenth Symposium on Educational Advances in Artificial Intelligence* (2024), AAAI'24/IAAI'24/EAAI'24, AAAI Press. URL: https://doi.org/10.1609/aaai.v38i4.28099, doi:10.1609/aaai.v38i4.28099. 3

[LLS*24] Li J., Li Y., Sun C., Wang C., Xiang J.: Spec-nerf: Multi-spectral neural radiance fields. In *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (2024), pp. 2485–2489. doi:10.1109/ICASSP48485.2024.10446015. 3

[LLT*24] Lin J., Li Z., Tang X., Liu J., Liu S., Liu J., Lu Y., Wu X., Xu S., Yan Y., Yang W.: Vastgaussian: Vast 3d gaussians for large scene reconstruction. In *CVPR* (2024). 1, 3

[Low99] Lowe D. G.: Object recognition from local scale-invariant features. In *Proceedings of the seventh IEEE international conference on computer vision* (1999), vol. 2, Ieee, pp. 1150–1157. 4, 17

[LPFKW24] Lin Y. Y., Pan X.-Y., Fridovich-Keil S., Wetzstein G.: ThermalNeRF: Thermal radiance fields. In *IEEE International Conference on Computational Photography (ICCP)* (2024), IEEE. 3, 4

[LYX*24] Lu T., Yu M., Xu L., Xiangli Y., Wang L., Lin D., Dai B.: Scaffold-gs: Structured 3d gaussians for view-adaptive rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2024), pp. 20654–20664. 3

[MAWS25] Meyer L., Ardelean A.-T., Weyrich T., Stamminger M.: Fruitnerf++: A generalized multi-fruit counting method utilizing contrastive learning and neural radiance fields. *ArXiv* (June 2025). URL: https://meyerls.github.io/fruit_nerfpp. 11

[MBRS*21] Martin-Brualla R., Radwan N., Sajjadi M. S. M., Barron J. T., Dosovitskiy A., Duckworth D.: NeRF in the Wild: Neural Radiance Fields for Unconstrained Photo Collections. In *CVPR* (2021). 3

[MC24] Martins T. B., Civera J.: Feature splatting for better novel view synthesis with low overlap, 2024. URL: https://arxiv.org/abs/2405.15518, arXiv:2405.15518. 3, 7, 16

[MCV*97] Maes F., Collignon A., Vandermeulen D., Marchal G., Suetens P.: Multimodality image registration by maximization of mutual information. *IEEE Transactions on Medical Imaging 16*, 2 (1997), 187–198. doi:10.1109/42.563664. 11, 19

[MESK22] Müller T., Evans A., Schied C., Keller A.: Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics (TOG) 41*, 4 (jul 2022). 3, 5

[MEY*24] Meyer L., Erich F., Yoshiyasu Y., Stamminger M., Ando N., Domae Y.: Pegasus: Physically enhanced gaussian splatting simulation system for 6dof object pose dataset generation. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (2024), pp. 10710–10715. doi:10.1109/IROS58592.2024.10802037. 12

[MGK*24] Mallick S. S., Goel R., Kerbl B., Steinberger M., Carrasco F. V., De La Torre F.: Taming 3dgs: High-quality radiance fields with limited resources. In *SIGGRAPH Asia 2024 Conference Papers* (New York, NY, USA, 2024), SA '24, Association for Computing Machinery. URL: https://doi.org/10.1145/3680528.3687694, doi:10.1145/3680528.3687694. 1

[MGSS24] Meyer L., Gilson A., Schmid U., Stamminger M.: Fruitnerf: A unified neural radiance field based fruit counting framework, April 2024. URL: https://meyerls.github.io/fruit_nerf. 11

[MHA17] McInnes L., Healy J., Astels S.: hdbscan: Hierarchical density based clustering. *The Journal of Open Source Software 2* (03 2017). doi:10.21105/joss.00205. 20

[MSDT*25] Malarz D., Smolak-Dyżewska W., Tabor J., Tadeja S., Spurek P.: Gaussian splatting with nerf-based color and opacity. *Comput. Vis. Image Underst. 251*, C (Feb. 2025). URL: https://doi.org/10.1016/j.cviu.2024.104273, doi:10.1016/j.cviu.2024.104273. 12

[MSL*25] Meuleman A., Shah I., Lanvin A., Kerbl B., Drettakis G.: On-the-fly reconstruction for large-scale novel view synthesis from unposed images. *ACM Transactions on Graphics 44*, 4 (2025). 1

[MST*20] Mildenhall B., Srinivasan P. P. ., Tancik M., Barron Jonathan T.and Ramamoorthi R., Ng R.: Nerf: Representing scenes as neural radiance fields for view synthesis. In *European Conference on Computer Vision* (2020), pp. 405–421. 1, 3

[NMR*24] Niemeyer M., Manhardt F., Rakotosaona M.-J., Oechsle M., Duckworth D., Gosula R., Tateno K., Bates J., Kaeser D., Tombari F.: Radsplat: Radiance field-informed gaussian splatting for robust real-time rendering with 900+ fps. *arXiv preprint arXiv:2403.13806* (2024). 3

[NP12] Namin S. T., Petersson L.: Classification of materials in natural scenes using multi-spectral images. In *International Conference on Intelligent Robots and Systems* (2012), IEEE, pp. 1393–1398. 2

[NSW24] Niedermayr S., Stumpfegger J., Westermann R.: Compressed 3d gaussian splatting for accelerated novel view synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2024), pp. 10349–10358. 3

[OLMS24] Ojo T., La T., Morton A., Stavness I.: Splanting: 3d plant capture with gaussian splatting. SA '24, Association for Computing Machinery. URL: https://doi.org/10.1145/3681758.3698009, doi:10.1145/3681758.3698009. 11

[OWHE24] Özer M., Weiherer M., Hundhausen M., Egger B.: Exploring multi-modal neural scene representations with applications on thermal imaging. *ArXiv abs/2403.11865* (2024). URL: https://api.semanticscholar.org/CorpusID:268531403. 3, 4, 8, 18

[Pow64] POWELL M. J. D.: An efficient method for finding the minimum of a function of several variables without calculating derivatives. *The Computer Journal 7*, 2 (01 1964), 155–162. 19

[PZRT*22] POGGI M., ZAMA RAMIREZ P., TOSI F., SALTI S., DI STEFANO L., MATTOCCIA S.: Cross-spectral neural radiance fields. In *Proceedings of the International Conference on 3D Vision* (2022). 3DV. 3, 5, 6, 7, 8, 18, 20

[RFS22] RÜCKERT D., FRANKE L., STAMMINGER M.: Adop: approximate differentiable one-pixel point rendering. *ACM Trans. Graph. 41*, 4 (July 2022). URL: https://doi.org/10.1145/3528223.3530122, doi:10.1145/3528223.3530122. 3

[RJL*24] REN K., JIANG L., LU T., YU M., XU L., NI Z., DAI B.: Octree-gs: Towards consistent real-time rendering with lod-structured 3d gaussians. *arXiv preprint arXiv:2403.17898* (2024). 3

[RSP*24] RADL L., STEINER M., PARGER M., WEINRAUCH A., KERBL B., STEINBERGER M.: StopThePop: Sorted Gaussian Splatting for View-Consistent Real-time Rendering. *ACM TOG 4*, 43 (2024). 3

[SCK08] SHUM H.-Y., CHAN S.-C., KANG S. B.: *Image-based rendering*. Springer Science & Business Media, 2008. 2

[SF16] SCHÖNBERGER J. L., FRAHM J.-M.: Structure-from-motion revisited. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2016), pp. 4104–4113. 3, 17

[SHZ*23] SMITT C., HALSTEAD M., ZIMMER P., LÄBE T., GUCLU E., STACHNISS C., MCCOOL C.: Pag-nerf: Towards fast and efficient end-to-end panoptic 3d representations for agricultural robotics, 2023. URL: https://arxiv.org/abs/2309.05339, arXiv:2309.05339. 11

[SK00] SHUM H., KANG S. B.: Review of image-based rendering techniques. In *Visual Communications and Image Processing 2000* (2000), vol. 4067, SPIE, pp. 2–13. 2

[SSOA*23] SAEED F., SUN J., OZIAS-AKINS P., CHU Y. J., LI C. C.: Peanutnerf: 3d radiance field for peanuts. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (2023), pp. 6254–6263. doi:10.1109/CVPRW59228.2023.00665. 11

[SSS06] SNAVELY N., SEITZ S. M., SZELISKI R.: Photo tourism: exploring photo collections in 3d. vol. 25. Association for Computing Machinery, New York, NY, USA, July 2006, p. 835–846. 3

[SZPF16] SCHÖNBERGER J. L., ZHENG E., POLLEFEYS M., FRAHM J.-M.: Pixelwise view selection for unstructured multi-view stereo. In *European Conference on Computer Vision (ECCV)* (Cham, 2016), Springer International Publishing. 3, 4, 5

[TFT*20] TEWARI A., FRIED O., THIES J., SITZMANN V., LOMBARDI S., SUNKAVALLI K., MARTIN-BRUALLA R., SIMON T., SARAGIH J., NIEßNER M., PANDEY R., FANELLO S., WETZSTEIN G., ZHU J.-Y., THEOBALT C., AGRAWALA M., SHECHTMAN E., GOLDMAN D. B., ZOLLHÖFER M.: State of the art on neural rendering. vol. 39, pp. 701–727. 3

[TML*24] THIRGOOD C., MENDEZ O., LING E., STOREY J., HADFIELD S.: Hypergs: Hyperspectral 3d gaussian splatting, 12 2024. doi:10.48550/arXiv.2412.12849. 3, 5

[TRS22] TURKI H., RAMANAN D., SATYANARAYANAN M.: Mega-nerf: Scalable construction of large-scale nerfs for virtual fly-throughs. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2022), pp. 12922–12931. 3

[TSM*20] TANCIK M., SRINIVASAN P. P., MILDENHALL B., FRIDOVICH-KEIL S., RAGHAVAN N., SINGHAL U., RAMAMOORTHI R., BARRON J. T., NG R.: Fourier features let networks learn high frequency functions in low dimensional domains. *NeurIPS* (2020). 5

[TTM*22] TEWARI A., THIES J., MILDENHALL B., SRINIVASAN P., TRETSCHK E., WANG Y., LASSNER C., SITZMANN V., MARTIN-BRUALLA R., LOMBARDI S., SIMON T., THEOBALT C., NIESSNER M., BARRON J. T., WETZSTEIN G., ZOLLHOEFER M., GOLYANIK V.: Advances in neural rendering. vol. 41, pp. 703–735. 3

[TWN*23] TANCIK M., WEBER E., NG E., LI R., YI B., WANG T., KRISTOFFERSEN A., AUSTIN J., SALAHI K., AHUJA A., MCALLISTER D., KERR J., KANAZAWA A.: Nerfstudio: A modular framework for neural radiance field development. In *ACM SIGGRAPH 2023 Conference Proceedings* (New York, NY, USA, 2023), SIGGRAPH '23, Association for Computing Machinery. URL: https://doi.org/10.1145/3588432.3591516, doi:10.1145/3588432.3591516. 6

[VBK*23] VOYNOV O., BOBROVSKIKH G., KARPYSHEV P., GALOCHKIN S., ARDELEAN A.-T., BOZHCNKO A., KARMANOVA E., KOPANEV P., LABUTIN-RYMSHO Y., RAKHIMOV R., SAFIN A., SERPIVA V., ARTEMOV A., BURNAEV E., TSETSERUKOU D., ZORIN D.: Multi-sensor large-scale dataset for multi-view 3d reconstruction. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2023), pp. 21392–21403. doi:10.1109/CVPR52729.2023.02049. 6

[WYF*24] WU G., YI T., FANG J., XIE L., ZHANG X., WEI W., LIU W., TIAN Q., WANG X.: 4d gaussian splatting for real-time dynamic scene rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2024), pp. 20310–20320. 3

[WYZ*24] WU T., YUAN Y.-J., ZHANG L.-X., YANG J., CAO Y.-P., YAN L.-Q., GAO L.: Recent advances in 3d gaussian splatting. *Computational Visual Media 10*, 4 (2024), 613–642. 3

[XLKP25] XU J., LIAO M., KATHIRVEL R. P., PATEL V. M.: Leveraging thermal modality to enhance reconstruction in low-light conditions. In *Computer Vision – ECCV 2024* (Cham, 2025), Leonardis A., Ricci E., Roth S., Russakovsky O., Sattler T., Varol G., (Eds.), Springer Nature Switzerland. 3

[XMP24] XU J., MEI Y., PATEL V. M.: Wild-GS: Real-time novel view synthesis from unconstrained photo collections. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems* (2024). 3

[YCH*24] YU Z., CHEN A., HUANG B., SATTLER T., GEIGER A.: Mip-splatting: Alias-free 3d gaussian splatting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2024), pp. 19447–19456. 3

[YLK*25] YE V., LI R., KERR J., TURKULAINEN M., YI B., PAN Z., SEISKARI O., YE J., HU J., TANCIK M., KANAZAWA A.: gsplat: An open-source library for gaussian splatting. *Journal of Machine Learning Research 26*, 34 (2025), 1–17. 5, 6

[YLL*24] YE Z., LI W., LIU S., QIAO P., DOU Y.: Absgs: Recovering fine details for 3d gaussian splatting, 2024. arXiv:2404.10484. 6

[YLX*24] YANG X., LU X., XIE P., ZIYUE G., FANG H., FU H., HU X., SUN Z., CEN H.: Paniclenerf: Low-cost, high-precision in-field phenotyping of rice panicles with smartphone. *Plant Phenomics 6* (12 2024). doi:10.34133/plantphenomics.0279. 11

[YWD*24] YE T., WU Q., DENG J., LIU G., LIU L., XIA S., PANG L., YU W., PEI L.: Thermal-nerf: Neural radiance fields from an infrared camera, 2024. URL: https://arxiv.org/abs/2403.10340, arXiv:2403.10340. 3

[ZAQ*24] ZHENG X., AI X., QIN H., RONG J., ZHANG Z., YANG Y., YUAN T., LI W.: Tomato-nerf: Advancing tomato model reconstruction with improved neural radiance fields. *IEEE Access 12* (2024), 184206–184215. doi:10.1109/ACCESS.2024.3424908. 11

[ZF03] ZITOVÁ B., FLUSSER J.: Image registration methods: a survey. *Image and Vision Computing 21*, 11 (2003), 977–1000. URL: https://www.sciencedirect.com/science/article/pii/S0262885603001379, doi:https://doi.org/10.1016/S0262-8856(03)00137-9. 11, 19

[ZRSK20] ZHANG K., RIEGLER G., SNAVELY N., KOLTUN V.: Nerf++: Analyzing and improving neural radiance fields. *arXiv preprint arXiv:2010.07492* (2020). 3

[ZWW*24] ZHANG D., WANG C., WANG W., LI P., QIN M., WANG H.: Gaussian in the wild: 3d gaussian splatting for unconstrained image collections. In *European Conference on Computer Vision* (2024), Springer, pp. 341–359. 3