This CVPR paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

PPISP: Physically-Plausible Compensation and Control of Photometric Variations in Radiance Field Reconstruction

Isaac Deutsch*, Nicolas Moënne-Loccoz*, Gavriel State, Zan Gojcic
NVIDIA
{ideutsch, nicolasm, gstate, zgojcic}@nvidia.com
https://research.nvidia.com/labs/sil/projects/ppisp/



Photometrically Varying Input | Consistent Radiance Field Reconstruction | Automatic Control

Figure 1. We introduce a differentiable image processing pipeline applied to radiance field reconstruction. By modeling the behavior of conventional cameras, our approach disentangles image formation effects from the rest of the pipeline. Our physically-plausible model admits a controller module that predicts exposure and color changes for novel views.

Abstract

Multi-view 3D reconstruction methods remain highly sensitive to photometric inconsistencies arising from camera optical characteristics and variations in image signal processing (ISP). Existing mitigation strategies such as per-frame latent variables or affine color corrections lack physical grounding and generalize poorly to novel views. We propose the Physically-Plausible ISP (PPISP) correction module, which disentangles camera-intrinsic and capture-dependent effects through physically based and interpretable transformations. A dedicated PPISP controller, trained on the input views, predicts ISP parameters for novel viewpoints, analogous to auto exposure and auto white balance in real cameras. This design enables realistic and fair evaluation on novel views without access to ground-truth images. PPISP achieves SotA performance on standard benchmarks, while providing intuitive control and supporting the integration of metadata when available. The source code is available at: https://github.com/nv-tlabs/ppisp

* Equal contribution.

1. Introduction

State-of-the-art multi-view 3D reconstruction methods have significantly advanced the fidelity of novel view synthesis (NVS), transforming it into a technology with real-world applications in physical AI simulation, virtual production, and content creation. Despite these advances, the quality of reconstruction and view synthesis remains highly sensitive to the quality of the input data—both to the distribution of camera poses and to multi-view appearance inconsistencies. The latter often arise from variations in camera optical characteristics and image signal processing (ISP) settings over time. These variations result in differences in color tone, intensity, and contrast that violate the photometric consistency assumptions underlying 3D reconstruction.

A common strategy to mitigate these appearance variations is to introduce additional, optimizable per-frame or per-camera parameters designed to capture photometric residuals while preserving a consistent multi-view scene representation. Recent state-of-the-art approaches include low-dimensional generative latent optimization (GLO) vectors [18], learnable affine transformations [24], and bilateral grids (BilaRF) [33]. However, these mitigation strategies face several trade-offs and challenges:

- **Representation capacity:** higher-capacity and less-constrained modules tend to improve PSNR on the training views but risk modeling more than just photometric variations, often degrading novel view synthesis quality.
- **Interpretability and controllability:** the learned parameters are typically non-interpretable (*e.g.*, in GLO or BilaRF), making it difficult to intuitively adjust properties such as brightness or white balance.
- **Parameters for novel views:** since the parameters are optimized independently per frame, it is unclear how to assign appropriate values when synthesizing novel views.

The latter is especially challenging due to the tendency of these modules to conflate camera sensor intrinsic properties (*e.g.*, vignetting and camera response function) with capture-dependent settings that vary per frame or are adjusted by the ISP (*e.g.*, exposure time and white balance). As a consequence, evaluation protocols commonly assume access to the ground-truth novel view image and estimate a corrective mapping, such as an affine transform, quadratic polynomial, or direct parameter optimization, to minimize the difference between the synthesized and the ground-truth (GT) image before computing the evaluation metrics. But such protocols are inherently flawed as they: **(i)** deviate from real-world scenarios where GT novel views are unavailable, and **(ii)** conceal differences between methods by compensating for them through the corrective mapping.

To address these challenges, we propose a Physically-Plausible ISP (PPISP) correction module, grounded in the physical principles of camera image formation. Specifically, we disentangle sensor-intrinsic properties and capture-dependent settings through dedicated per-sensor and per-frame modules, respectively, and constrain their effects according to the image formation process (*e.g.*, the exposure module can only modify the overall image brightness). Our model acts as a post-processing step applied to the raw images rendered from the 3D representation, and enables direct controllability through manual change of the parameters. Moreover, we introduce a PPISP controller that predicts the parameters of the per-frame modules for novel views, analogous to the auto exposure and auto white balance mechanisms in conventional cameras.

2. Related Work

Appearance inconsistencies across multi-view input images significantly degrade the quality of radiance field reconstructions and subsequent novel-view synthesis. Such variations are common in unconstrained image collections, for instance when using internet photo collections or captures under uncontrolled lighting conditions.

Compensation during reconstruction. To mitigate these inconsistencies, NeRF-W [18] and GS-W [37] introduce GLO that are optimized jointly with the scene representation. These per-image latent embeddings enable smooth interpolation across observed appearances, but risk entangling scene geometry with reflectance when optimized end-to-end. Block-NeRF [30] extends this idea to city-scale scenes and additionally conditions on camera exposure metadata. To impose stronger constraints and better align with the image formation process, subsequent works model photometric transformations explicitly. URF [24] represents per-image variations using affine color transformations, while BilaRF [33] extends this idea to per-pixel affine mappings parameterized via bilateral grids. Several works instead model specific physical components of the image formation pipeline: Xian *et al*. [35] learn lens distortion and vignetting jointly with the radiance field, and HDR-NeRF [13] and HDR-GS [4] recover HDR radiance by learning a camera response function (CRF) from multi-exposure captures. Closest to our approach, ADOP's [25] post-processing models exposure, white balance, CRF, and vignetting effects as explicit calibration parameters. However, our formulation better disentangles exposure offset and white balance, while using a more compact CRF model. Recently, Huang *et al*. [14] and Niemeyer *et al*. [20] deviate from a frame-based correction and instead learn a 3D exposure neural field, predicting the optimal exposure values for each 3D point.

Harmonizing appearance during preprocessing. An alternative strategy is to decouple the compensation from reconstruction and harmonize the input images as a preprocessing step. Shin *et al*. [28] employ a transformer network to predict bilateral grids that harmonize each image to a chosen reference view. Alzayer *et al*. [1] instead use a diffusion model to relight images directly, but due to the lack of paired real data, they train their generative model only on synthetic data. To overcome this limitation, Trevithick *et al*. [32] use a generative video model to simulate capture-time inconsistencies on consistent multi-view images, creating pseudo-paired data to train a harmonization network.

Novel view synthesis with target appearance. The above methods reconstruct the scene in a canonical or reference appearance, but it remains unclear how to set the parameters of their appearance modules to render an image in a desired target appearance. This target appearance could be user-defined or selected to match the appearance that a camera with auto exposure and white balance would produce. This ambiguity poses practical challenges for novel view synthesis and complicates fair evaluation under photometric variation. Prior work typically applies post-render normalization that assumes access to the target image during evaluation: NeRF-W [18] fine-tunes latent embeddings on one half of each image and evaluates on the other, RawNeRF [19] performs channel-wise affine alignment, Mip-NeRF 360 [2] uses a quadratic color basis alignment, and ADOP [25] re-optimizes per-frame parameters.

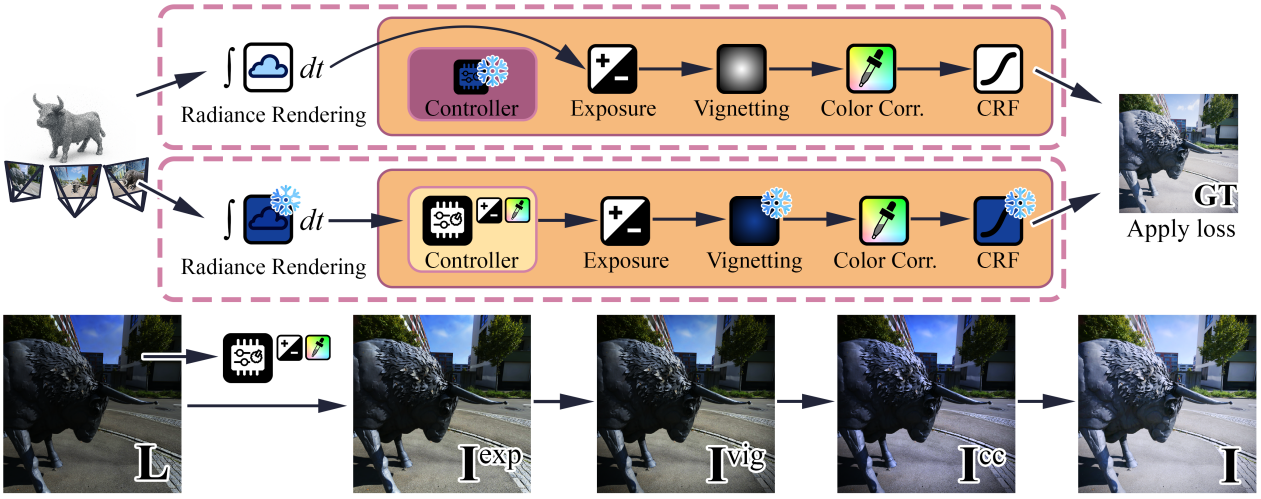


Figure 2. Our proposed pipeline applies a sequence of physically-grounded modules to the input reconstructed radiance (exposure offset, chromatic vignetting, linear color correction, and non-linear camera response function). Top: all modules except the controller are jointly optimized during the first training phase. Bottom: the controller is then trained to predict per-frame exposure and color correction for novel views while other modules are frozen. The image sequence illustrates the progressive effect of each module.

Such evaluation protocols, however, **(i)** mask differences between methods and **(ii)** are infeasible in real-world applications where access to the target image cannot be assumed. In line with the principle that novel views should be rendered solely from reconstructed data without access to target pixels, we introduce a PPISP controller that takes the *raw* radiance image rendered from the 3D representation as input and outputs the PPISP parameters. We optimize this network on the training views and then directly apply it to the novel views during inference. Somewhat related to our PPISP controller, [21, 31] train a network to predict exposure control for improved feature matching and object detection, respectively.

3. Preliminaries

Radiance Field Reconstruction aims to optimize a parametric representation of a scene's volumetric density $\sigma \in \mathbb{R}$ and emitted radiance $\mathbf{c} \in \mathbb{R}^3$. The radiance $\mathbf{L}(\mathbf{r})$ of a camera ray $\mathbf{r}(x) = \mathbf{o} + x\,\mathbf{d}$ with origin $\mathbf{o} \in \mathbb{R}^3$ and direction $\mathbf{d} \in \mathbb{R}^3$ is rendered from this representation as

$$\mathbf{L}(\mathbf{r}) = \int_{near}^{far} T(x)\,\sigma(\mathbf{r}(x))\,\mathbf{c}(\mathbf{r}(x))\,dt\,, \tag{1}$$

where $T(x) = \exp(-\int_{near}^{x} \sigma(\mathbf{r}(y))\,dy)$ denotes the transmittance along the ray. The optimization is supervised using ground truth images $\mathbf{I}$ captured by one or more cameras with known intrinsics and extrinsics. This standard formulation alone does not account for camera-specific imaging effects.

Camera Image Formation is the process through which the radiance $\mathbf{L}$ is converted to the final image:

$$\mathbf{I} = \mathcal{F}(\mathbf{L}; \mathbf{\Theta})\,. \tag{2}$$

Here, the function $\mathcal{F}(\cdot)$ models the complete image acquisition process, including lens distortions (e.g., vignetting, chromatic aberrations), exposure settings (aperture, shutter time), sensor characteristics (spectral response, noise, gain), and ISP operations according to some parameters $\mathbf{\Theta}$. While some components of this process remain constant across acquisition time, others may vary due to manual adjustments or automatic adaptation by the sensor controller.

Notation. Let $\mathbf{I} \in \mathbb{R}^{H \times W \times 3}$ be an RGB image. The color at spatial location $\mathbf{u} = (i, j)$ is $\mathbf{x} = \mathbf{I}_{i,j} \in \mathbb{R}^3$ and its k-th channel value is $x = \mathbf{x}_k = \mathbf{I}_{i,j,k} \in \mathbb{R}$, $k \in \{R, G, B\}$. Operations defined on channel values or colors are understood element-wise when applied to an image.

4. Method

We compensate for photometric inconsistencies across input images by jointly optimizing the scene representation together with a differentiable ISP pipeline that approximates the camera image formation function $\mathcal{F}(\cdot)$ defined in Eq. (2). During optimization, this pipeline models both camera-specific and time-varying effects. During inference (*i.e.*, when rendering novel views), the learned controller (Sec. 4.5) predicts the time-varying parameters directly from the radiance $\mathbf{L}$ rendered from the scene representation.
Our ISP pipeline consists of four sequential modules (see Fig. 2):

- *Exposure offset* accounts for aperture, shutter time and gain variations,
- *Vignetting* models optical attenuation across the sensor,
- *Color correction* models sensor spectral response and white balance adjustments,
- *Camera response function* (CRF) applies a non-linear transformation from sensor irradiance to image colors.

Following Debevec and Malik [6], the first three modules operate linearly on the scene radiance, while the CRF provides the final non-linear mapping. Fig. 2 puts the pipeline in the context of the radiance reconstruction and illustrates the individual parts and their effects.

4.1. Exposure Offset

We model exposure as a global, per-frame scale on the radiance using a base-2 exponent, mimicking photographic exposure values:

$$\mathbf{I}^{\text{exp}} \;=\; \mathcal{E}(\mathbf{L};\Delta t) \;=\; \mathbf{L}\,2^{\Delta t}\,, \tag{3}$$

where $\Delta t \in \mathbb{R}$ is an optimizable exposure offset. This offset represents the variation of the radiance intensity reaching the sensor and is specific to the capture. Thus, we estimate one such offset for each frame.

4.2. Vignetting

Following Goldman [10], we model per-channel radial intensity falloff using a polynomial in the squared radius around an optimizable optical center:

$$\mathbf{I}^{\text{vig}} \;=\; \mathcal{V}(\mathbf{I}^{\text{exp}};\boldsymbol{\mu},\boldsymbol{\alpha}) \;=\; \mathbf{I}^{\text{exp}}\cdot v(r;\boldsymbol{\alpha})\,, \tag{4}$$

where $\boldsymbol{\mu}\in\mathbb{R}^2$ is the optical center, $\boldsymbol{\alpha}\in\mathbb{R}^3$ are polynomial coefficients, and $r=\|\mathbf{u}-\boldsymbol{\mu}\|_2$ is the distance of the pixel location $\mathbf{u}$ to the optical center. The attenuation factor $v(r)$ is defined as:

$$v(r) \;=\; \text{clip}_{(0,1)}\big(1+\alpha_1\,r^2+\alpha_2\,r^4+\alpha_3\,r^6\big)\;. \tag{5}$$

At the start of optimization, we initialize $\boldsymbol{\alpha}=0$ and let $\boldsymbol{\mu}$ be the image center.

Since our vignetting model is chromatic, a falloff polynomial is defined for each color channel by distinct parameter values.

4.3. Color Correction

To model effects such as white balance, which may vary per-frame, and gamut differences between multiple cameras, we apply color correction. To disentangle it from exposure correction, we apply a 3×3 homography $\mathbf{H}$ on RG chromaticities and intensity — following Finlayson *et al.* [9] — and ensure normalization of the intensity after the transform. Inspired by DeTone *et al.* [7], we parameterize the color correction as four chromaticity offsets $\Delta\mathbf{c}_k$, construct $\mathbf{H}$ from them, and apply the color correction:

$$\mathbf{I}^{\text{cc}} \;=\; \mathcal{C}\big(\mathbf{I}^{\text{vig}};\,\{\Delta\mathbf{c}_k\}_{k\in\{R,G,B,W\}}\big) \;=\; h(\mathbf{I}^{\text{vig}};\mathbf{H})\;. \tag{6}$$

Let $\mathbf{C}\in\mathbb{R}^{3\times 3}$ denote the RGB→RGI conversion matrix and $\mathbf{C}^{-1}$ its inverse. The intensity normalization can then be defined as:

$$n(\mathbf{x};\mathbf{H}) \doteq \frac{\mathbf{x}_R+\mathbf{x}_G+\mathbf{x}_B}{\big[\mathbf{H}\cdot\mathbf{C}\,\mathbf{x}\big]_3+\varepsilon}\;. \tag{7}$$

Here, ε is a small constant for numerical stability. This normalization decouples exposure from chromatic correction. The color transform follows compactly as

$$h(\mathbf{x};\mathbf{H}) \;\doteq\; \mathbf{C}^{-1}\big(n(\mathbf{x};\mathbf{H})\cdot\big(\mathbf{H}\cdot\mathbf{C}\,\mathbf{x}\big)\big)\,. \tag{8}$$

To construct $\mathbf{H}$, we define four 2D source–target chromaticity pairs. Specifically, we fix the source RG chromaticities $\mathbf{c}_{s,\cdot}$ to the three primaries and a neutral white:

$$\begin{aligned}\mathbf{c}_{s,R}&=(1,0)^T\;; \quad \mathbf{c}_{s,G}=(0,1)^T\;;\\ \mathbf{c}_{s,B}&=(0,0)^T\;; \quad \mathbf{c}_{s,W}=\big(\tfrac{1}{3},\tfrac{1}{3}\big)^T\;,\end{aligned} \tag{9}$$

and define the targets $\mathbf{c}_{t,\cdot}$ as offsets from these sources $\mathbf{c}_{t,k} = \mathbf{c}_{s,k}+\Delta\mathbf{c}_k$ for $k\in\{R,G,B,W\}$. By lifting the 2D chromaticities to homogeneous coordinates and stacking them as $\mathbf{S} \doteq [\,\tilde{\mathbf{c}}_{s,R}\;\tilde{\mathbf{c}}_{s,G}\;\tilde{\mathbf{c}}_{s,B}\,]$ and $\mathbf{T} \doteq [\,\tilde{\mathbf{c}}_{t,R}\;\tilde{\mathbf{c}}_{t,G}\;\tilde{\mathbf{c}}_{t,B}\,]$, we can define

$$\mathbf{M} \;\doteq\; [\tilde{\mathbf{c}}_{t,W}]_\times\,\mathbf{T}\,, \tag{10}$$

where $[\cdot]_\times$ is the skew-symmetric cross-product matrix. Then, $\mathbf{k}\in\mathbb{R}^3$ can be obtained via a cross-product of any pair of linearly independent rows i and j,

$$\mathbf{k} \;\propto\; \mathbf{m}_i\times\mathbf{m}_j\,. \tag{11}$$

where $\mathbf{m}_1,\mathbf{m}_2,\mathbf{m}_3$ are the rows of $\mathbf{M}$. Finally, we form and normalize

$$\mathbf{H} \;=\; \mathbf{T}\,\text{diag}(\mathbf{k})\,\mathbf{S}^{-1}\,, \qquad \mathbf{H}\;\leftarrow\;\frac{\mathbf{H}}{[\mathbf{H}]_{3,3}}\,. \tag{12}$$

A precise derivation and further details are provided in the Supplementary.

4.4. Camera Response Function

Inspired by Grossberg and Nayar [12], we use a piecewise power curve to model non-linear chromatic transformations. The CRF operator $\mathcal{G}$ has four learned parameters:

$$\mathbf{I} \;=\; \mathcal{G}(\mathbf{I}^{\text{cc}};\tau,\eta,\xi,\gamma)\,. \tag{13}$$

For each channel, the basic S-shaped curve is given by:

$$f_0(x;\,\tau,\eta,\xi)=\begin{cases} a\left(\dfrac{x}{\xi}\right)^{\tau}, & 0\le x\le\xi\,,\\[2ex] 1-b\left(\dfrac{1-x}{1-\xi}\right)^{\eta}, & \xi<x\le 1\,,\end{cases} \tag{14}$$

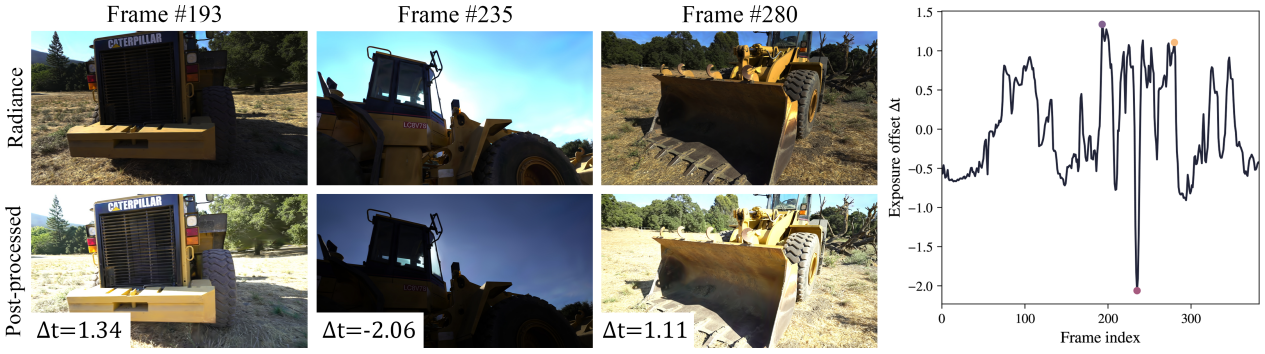


Figure 3. Dynamics of the controller module. The predicted exposure offset (inset) depends on the image content of the rendered radiance. Right side: Plot of exposure offsets as predicted for each frame of the *caterpillar* sequence, with the three displayed frames highlighted.

setting a and b to match the slope at the inflection point to ensure C^1 continuity:

$$a = \frac{\eta\,\xi}{\tau(1-\xi)+\eta\,\xi}\,, \qquad b = 1-a\,. \tag{15}$$

Finally, the CRF image operator $\mathcal{G}$ is a composition of this S-curve with a gamma correction:

$$\mathcal{G}(x;\,\tau,\eta,\xi,\gamma) = \big[f_0(x;\,\tau,\eta,\xi)\big]^{\gamma}\,. \tag{16}$$

4.5. Per-Frame ISP Parameter Controller

The exposure offsets and color correction transforms introduced above are valid only for a specific capture, *i.e.*, a single camera pose, and therefore cannot be directly reused for novel view rendering. To address this limitation, we introduce a controller that predicts these parameters from the rendered scene radiance, analogous to how auto exposure and auto white balance work in conventional cameras:

$$(\Delta t, \{\Delta\mathbf{c}_k\}_{k\in\{R,G,B,W\}}) = \mathcal{T}(\mathbf{L})\,. \tag{17}$$

Here, $\mathcal{T}(\cdot)$ is the camera-specific controller parametric function, which we design as a coarse feature extractor (1×1 convolutions with pooling to a 5×5 grid), followed by a parameter regressor (an MLP with separate output heads). The detailed architecture of the controller is provided in the Supplementary.

We optimize the controller in a separate stage once the optimization of the scene representation is complete. At that stage, the underlying reconstruction and all per-camera ISP parameters are frozen, the controller-predicted parameters are applied through the ISP, and the controller itself is trained using the same photometric loss as in the initial phase. A qualitative example of the controller's effects is given in Fig. 3. Optional scalar controls (*e.g.*, exposure compensation or EXIF-derived biases) can be concatenated to the regressor input.

4.6. Regularization

Joint optimization of the modules can introduce brightness and color ambiguities between scene radiance and the ISP parameters. To mitigate this, we apply regularization on the previously defined parameters, using the Huber loss $\mathcal{L}_\delta$, where δ denotes the threshold. We use superscripts to indicate parameters belonging to specific camera sensors$^{(s)}$ and frames$^{(f)}$.

Brightness. We penalize the mean exposure offset over frames:

$$\mathcal{L}_b \;=\; \lambda_b\,\mathcal{L}_{\delta=0.1}\left(\frac{1}{F}\sum_{f=1}^{F}\Delta t^{(f)}\right)\,. \tag{18}$$

Color. We penalize the frame-mean of the target chromaticity offsets (element-wise in $\mathbb{R}^2$):

$$\mathcal{L}_c \;=\; \lambda_c \sum_{k\in\{R,G,B,W\}} \mathcal{L}_{\delta=0.005}\left(\frac{1}{F}\sum_{f=1}^{F}\Delta\mathbf{c}_k^{(f)}\right)\,. \tag{19}$$

Because chromatic corrections, as done in vignetting and CRF modules, may also introduce localized color shifts, we shrink parameter variance across channels. Let $\boldsymbol{\theta}_{m,k}$ be the parameters of channel k for module $m\in\{\text{vig},\text{crf}\}$. We penalize their across-channel variance, averaged over parameters:

$$\mathcal{L}_{\text{var}} \;=\; \lambda_{\text{var}} \sum_{m\in\{\text{vig},\text{crf}\}} \mathrm{Var}_k(\boldsymbol{\theta}_{m,k})\,. \tag{20}$$

Physically-plausible vignetting. For each polynomial, we penalize the center and softly enforce $\alpha_j \le 0$:

$$\mathcal{L}_{\text{vig}} \;=\; \lambda_v\left(\|\boldsymbol{\mu}_k\|_2^2 \;+\; \sum_j \big[\alpha_j\big]_+^2\right)\,. \tag{21}$$

Here $[x]_+ = \max(x,0)$ is the elementwise rectifier. The overall regularizer is

$$\mathcal{L}_{\text{reg}} = \mathcal{L}_b + \mathcal{L}_c + \mathcal{L}_{\text{var}} + \mathcal{L}_{\text{vig}}\,. \tag{22}$$

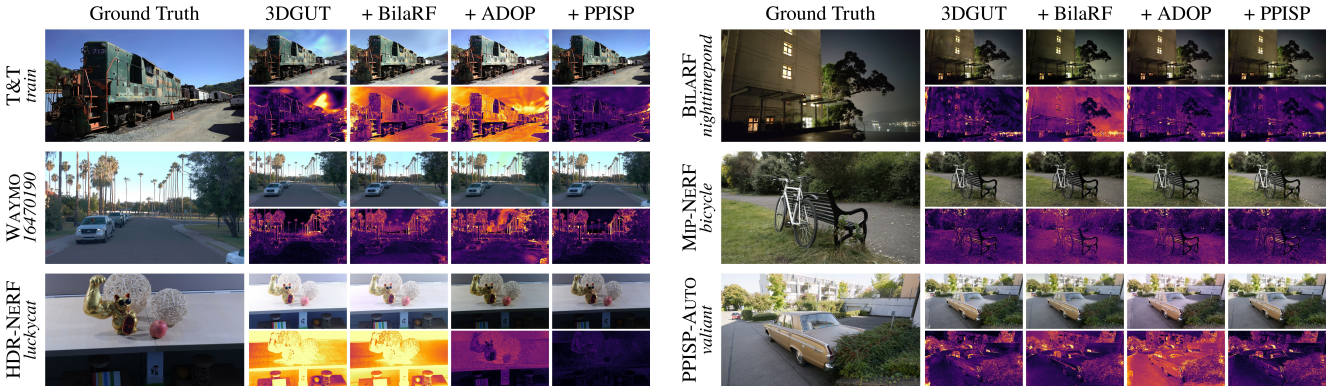


Figure 4. Qualitative comparison of novel view synthesis. Row labels indicate datasets and sequences (in *italics*). Column labels indicate methods. Heat maps show perceptual CIEDE2000 [27] error (colormap range: 0–20 ΔE_{00}). Our method achieves more consistent photometry and better color reproduction across various datasets and sequences. HDR-NeRF: When image metadata is available, our method can incorporate it to produce a more accurate novel view.

5. Experiments

We begin by evaluating the proposed PPISP correction module and controller on standard novel-view synthesis benchmarks, assessing both reconstruction fidelity and novel-view quality (Sec. 5.1). We then demonstrate how our formulation allows us to incorporate image metadata, such as relative exposure, when available (Sec. 5.2). We measure the runtime performance impact (Sec. 5.3). Finally, we analyze the relationship between model capacity, overfitting behavior, and novel-view synthesis performance (Sec. 5.4).

Setting. As a reconstruction-agnostic post-processing step, the PPISP module readily applies to different radiance field methods. We integrate it in 3DGUT [34], GSplat [36] (an accelerated implementation of 3DGS [15]), and Zip-NeRF [3].

Comparison baselines are the appearance correction approaches described in GLO [18], BilaRF [33], and ADOP [25]. For experiments, we rely on their reference hyperparameters and reference implementations available in the respective framework. To increase the stability of ADOP's method, we increase the strength of their CRF regularization about $100\times$ compared to the reference value.

We jointly train the reconstruction method and the post-processing operator for 30k iterations. For the PPISP controller, we freeze both and train the controller for an additional 5k iterations. For 3DGS [15, 36] and 3DGUT [34], we enable MCMC sampling.

Metrics. We evaluate the perceptual quality of the rendered views using peak signal-to-noise ratio (PSNR), structural similarity (SSIM), and learned perceptual image patch similarity (LPIPS) metrics.

As the PSNR metric is highly sensitive to global brightness shifts, and our baselines do not support appearance compensation for novel views, we additionally report the PSNR computed after affine color alignment, following RawNeRF [19]. We denote this as "PSNR-CC", but emphasize that such comparison masks the differences between the methods and assumes access to the GT target views, which are not available in practice.

Datasets. To show the robustness and generality of our method, we conducted experiments on a variety of publicly available datasets: Mip-NeRF 360 [2], Tanks and Temples [17], BilaRF [33], HDR-NeRF [13], and nine static sequences of the Waymo Open Dataset [29].

To further highlight the differences of the methods in challenging real-world scenarios, we captured a new *PPISP dataset* consisting of four scenes. Each of them was captured with three different cameras (Apple iPhone 13 Pro, Nikon Z7, and OM System OM-1 Mark II) to ensure variations. More details about the scenes, resolution, and training-test splits are available in the Supplementary.

5.1. Novel View Synthesis Benchmark

Quantitative results on the standard benchmark scenes are presented in Tab. 1, and qualitative comparisons are shown in Fig. 4. Our method achieves the best PSNR, SSIM, and LPIPS in the large majority of settings across datasets and base methods, and on most datasets even surpasses the BilaRF baseline [33] when that baseline is given privileged access to the target image, i.e., when comparing our PSNR against the baseline's PSNR-CC. These gains, established primarily on 3DGUT [34], extend to the 3DGS [15, 36] and Zip-NeRF [3] integrations.

The comparison between PSNR and PSNR-CC further highlights the effectiveness of our controller in reproducing the camera's auto-exposure and white-balance behavior. On most datasets, the controller achieves metrics close to those obtained after affine color alignment, indicating that it faithfully predicts the necessary per-frame appearance corrections. The only notable discrepancy appears on the BilaRF

Table 1. **Novel view synthesis results across methods and datasets.** We compare appearance compensation methods applied on radiance field reconstruction methods. When the PPISP controller is omitted (*w/o ctrl.*), novel views use zero per-frame corrections. PSNR-CC factors out global exposure and color differences.

	PSNR ↑	PSNR-CC ↑	SSIM ↑	LPIPS ↓
BILARF				
3DGUT [34]	22.60	23.57	0.804	0.371
+ BilaRF [33]	21.41	25.63	0.764	0.371
+ ADOP [25]	22.95	25.73	0.802	0.376
+ PPISP (w/o ctrl.)	24.08	26.16	**0.820**	**0.346**
+ PPISP (w/ ctrl.)	**24.12**	25.92	**0.820**	0.349
3DGS [15, 36]	22.55	24.20	0.799	0.367
+ BilaRF [33]	23.06	25.72	0.766	0.360
+ GLO [18]	21.76	23.28	0.744	0.398
+ ADOP [25]	23.42	26.14	0.804	0.360
+ PPISP (w/ ctrl.)	**25.39**	26.62	**0.824**	**0.340**
Zip-NeRF [3]	21.23	21.44	0.756	0.385
+ BilaRF [33]	**22.94**	24.91	**0.787**	**0.336**
+ PPISP (w/ ctrl.)	22.85	23.76	0.778	0.368
MIP-NERF 360				
3DGUT [34]	27.74	27.65	**0.821**	0.262
+ BilaRF [33]	24.97	26.64	0.801	**0.260**
+ ADOP [25]	26.42	27.75	0.815	0.271
+ PPISP (w/o ctrl.)	27.55	28.02	0.819	0.264
+ PPISP (w/ ctrl.)	**28.15**	28.06	**0.821**	0.264
3DGS [15, 36]	27.67	27.54	0.818	0.261
+ BilaRF [33]	25.46	26.48	0.803	0.262
+ GLO [18]	21.26	23.23	0.693	0.334
+ ADOP [25]	26.51	27.49	0.811	0.270
+ PPISP (w/ ctrl.)	**27.97**	27.87	**0.819**	**0.260**
Zip-NeRF [3]	27.80	27.64	0.810	0.257
+ BilaRF [33]	26.65	27.05	0.808	0.255
+ PPISP (w/ ctrl.)	**28.13**	28.08	**0.815**	**0.253**
TANKS & TEMPLES				
3DGUT [34]	22.86	23.46	0.790	0.312
+ BilaRF [33]	19.78	23.46	0.770	0.298
+ ADOP [25]	20.28	24.20	0.769	0.323
+ PPISP (w/o ctrl.)	21.52	24.87	0.783	0.296
+ PPISP (w/ ctrl.)	**24.62**	25.25	**0.809**	**0.285**
3DGS [15, 36]	23.08	23.73	0.789	0.303
+ BilaRF [33]	20.58	23.34	0.774	0.296
+ GLO [18]	18.74	20.95	0.707	0.358
+ ADOP [25]	20.63	24.28	0.771	0.314
+ PPISP (w/ ctrl.)	**24.36**	25.16	**0.807**	**0.281**
Zip-NeRF [3]	22.20	22.24	0.725	0.378
+ BilaRF [33]	20.82	22.48	0.747	0.327
+ PPISP (w/ ctrl.)	**23.54**	23.85	**0.767**	**0.324**
WAYMO				
3DGUT [34]	25.56	25.21	0.785	0.397
+ BilaRF [33]	21.83	23.66	0.768	0.397
+ ADOP [25]	24.28	25.18	0.781	0.405
+ PPISP (w/o ctrl.)	25.03	25.46	0.786	**0.391**
+ PPISP (w/ ctrl.)	**25.69**	25.48	**0.787**	**0.391**
PPISP-AUTO				
3DGUT [34]	22.05	22.20	0.677	0.453
+ BilaRF [33]	20.81	22.30	0.668	0.440
+ ADOP [25]	19.94	22.52	0.670	0.462
+ PPISP (w/o ctrl.)	21.07	23.14	0.677	0.438
+ PPISP (w/ ctrl.)	**22.87**	23.21	**0.687**	**0.434**
3DGS [15, 36]	22.29	22.42	0.679	0.442
+ BilaRF [33]	20.92	22.37	0.668	0.438
+ GLO [18]	19.62	20.53	0.613	0.488
+ ADOP [25]	20.22	22.49	0.670	0.449
+ PPISP (w/ ctrl.)	**22.86**	23.17	**0.687**	**0.426**

Table 2. Component ablation of PPISP on the Tanks and Temples dataset for novel views (NV). Each row shows performance when removing the specified component.

	NV PSNR ↑
PPISP (full)	24.62
PPISP - no exposure	23.33
PPISP - no vignetting	24.08
PPISP - no color correction	24.27
PPISP - no CRF	24.36

Table 3. Novel View PSNR across datasets with metadata. Our pipeline is able to leverage metadata (e.g. EXIF) from the sensor as a side data provided to the controller regressor.

		HDR-NERF [13]		PPISP	
	metadata	PSNR ↑	PSNR-CC ↑	PSNR ↑	PSNR-CC ↑
3DGUT [34]		17.81	27.37	12.44	18.59
3DGUT + BilaRF [33]		15.40	26.95	13.39	20.89
3DGUT + ADOP [25]		15.49	24.14	13.36	17.34
	✓	31.27	36.10	20.44	21.60
3DGUT + PPISP		17.86	27.78	14.69	21.19
	✓	**34.30**	37.10	**21.69**	21.94

dataset, likely due to the fact that this dataset contains some manual settings overrides (indicated by the metadata), which are not captured by our controller.

Both PPISP and ADOP [25] employ camera-specific components (vignetting and CRF), which generalize to novel views, leading to improved metrics over BilaRF [33]. Our base image formation model (*w/o ctrl.*) outperforms both of these baselines thanks to better separation of concerns of the individual modules and stronger constraints (see also Sec. 5.4; a detailed comparison to ADOP is provided in the Supplementary). Our full pipeline consistently improves upon the base model by providing plausible per-frame parameter estimates via the controller.

Ablation. We ablate the relative contribution of each module in our pipeline through an ablation study on the TANKS AND TEMPLES dataset. Tab. 2 presents the novel view PSNR when individual components are removed from the full pipeline. The results demonstrate that all modules contribute to the full pipeline's performance, with exposure and vignetting corrections being most critical.

5.2. Using Image Metadata

Because our formulation closely mirrors the camera image formation process, it can naturally incorporate image metadata, such as the relative exposure of each frame, whenever available. We demonstrate this capability on the HDR-NeRF [13] and PPISP datasets, both of which use exposure bracketing (*i.e.*, captures with positive and negative exposure compensation) and provide the corresponding

Table 4. Rendering times (ms) on NVIDIA RTX 5090 for the MipNeRF 360 [2] dataset.

	Time (ms) ↓	% overhead ↓
3DGUT [34]	3.24	—
BilaRF [33]	1.17	36%
ADOP	0.10	3%
PPISP (w/o ctrl.)	0.10	3%
PPISP (w/ ctrl.)	0.84	26%

Table 5. Average PSNR on the Tanks and Temples dataset comparing training views (TV) and novel views (NV) for ISP modules with varying capacity. The limited capacity of our proposed pipeline reduces overfitting and leads to better generalization.

	TV PSNR ↑	NV PSNR ↑
BilaRF + PC	26.83	21.80
PPISP + BilaRF [33]	26.66	23.52
BilaRF [33]	**26.87**	19.78
ADOP [25]	26.08	20.28
PPISP	25.85	**24.62**

metadata. We concatenate this metadata to the input of the controller MLP regressor, allowing it to map rendered radiance plus metadata to effective ISP parameters.

Since the ADOP-style post-processing also models per-frame exposure offsets explicitly, we initialize them from known exposure values as proposed in ADOP [25].

Quantitative results in Tab. 3 show that supplying calibrated exposure offsets substantially improves novel-view accuracy. Moreover, providing this metadata to the controller yields further gains compared to ADOP, demonstrating our method's ability to leverage metadata for more accurate novel view appearance prediction.

5.3. Runtime Performance

Tab. 4 presents the computational performance of the post-processing methods we evaluated compared to the scene rendering. PPISP (*w/o ctrl.*) and ADOP [25] have a similar and very small computational footprint (3% of the rendering). The controller is adding a substantial overhead due to the required processing of the input image, but our pipeline remains significantly faster (26% vs 36%) compared to BilaRF on an NVIDIA RTX 5090 GPU.

5.4. ISP Capacity vs. Training and Novel Views

Next, we investigate how the capacity of the correction module affects the overfitting (difference between the PSNR on training and novel views) and generalization to novel views. The bilateral grids used in BilaRF [33] provide a highly expressive mechanism for modeling image operations [5] extending beyond simple compensation of photometric inconsistencies. In BilaRF [33], this operation is learned independently for each frame, providing a high modeling capacity. In contrast, our PPISP module intentionally has limited capacity to prevent overfitting, but in turn cannot model complex image operations that mix spatial and intensity effects such as localized tone-mapping.

In Tab. 5, we therefore study hybrids of the two approaches. Adding more capacity to per-frame BilaRF [33] with additional per-camera bilateral grids (+PC) does not meaningfully change PSNR on the training views as the model already has sufficient capacity. However, it does slightly improve the generalization as per-camera corrections carry over to novel viewpoints. Increasing our method's capacity by adding per-frame bilateral grids boosts PSNR on the training views, but noticeably degrades performance on novel views due to overfitting. Overall, our formulation achieves a favorable balance between capacity and generalization to unseen views.

6. Conclusion and Limitations

Accurately reconstructing the radiance field of a scene requires accounting for variations in the camera imaging pipeline across the input frames. Ignoring these variations introduces strong biases, leading to spurious color shifts and geometric artifacts. In this work, we introduced a differentiable post-processing pipeline whose design permits simulating the imaging process while remaining highly constrained to prevent reconstruction bias. We further proposed a controller that improves generalization to novel views by predicting per-frame imaging parameters directly from the rendered radiance.

Limitations. Our method shows superior generalization to novel views (Tab. 1), but it sometimes struggles to match the baselines on the training views (Tab. 5). This can be partially attributed to overfitting, but our formulation also ignores some important optical effects such as localized tone-mapping commonly found in modern phone cameras; lens flares, which are prominent in night scenes; and similar spatially-varying effects. While the proposed controller enables generalization to novel views, its ability to infer exposure and color-correction parameters from rendered radiance depends on the existence of meaningful correlations in the data. When such correlations are absent, for example when the physical camera controls (*e.g.*, shutter, aperture, ISO) are manually overridden, the controller must rely on extra metadata to predict correct values.

7. Acknowledgments

We thank our colleagues Qi Wu, Janick Martinez Esturo, András Bódis-Szomorú, and Nick Schneider for their suggestions, feedback, and valuable discussions.

References

[1] Hadi Alzayer, Philipp Henzler, Jonathan T. Barron, Jia-Bin Huang, Pratul P. Srinivasan, and Dor Verbin. Generative multiview relighting for 3d reconstruction under extreme illumination variation. In *Proceedings of the Computer Vision and Pattern Recognition Conference (CVPR)*, pages 10933–10942, 2025. 2

[2] Jonathan T. Barron, Ben Mildenhall, Dor Verbin, Pratul P. Srinivasan, and Peter Hedman. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In *CVPR*, 2022. 2, 6, 8, 7

[3] Jonathan T. Barron, Ben Mildenhall, Dor Verbin, Pratul P. Srinivasan, and Peter Hedman. Zip-nerf: Anti-aliased grid-based neural radiance fields. *ICCV*, 2023. 6, 7

[4] Yuanhao Cai, Zihao Xiao, Yixun Liang, Minghan Qin, Yulun Zhang, Xiaokang Yang, Yaoyao Liu, and Alan Yuille. Hdr-gs: Efficient high dynamic range novel view synthesis at 1000× speed via gaussian splatting. In *NeurIPS*, 2024. 2

[5] Jiawen Chen, Andrew Adams, Neal Wadhwa, and Samuel W Hasinoff. Bilateral guided upsampling. *ACM Transactions on Graphics (TOG)*, 35(6):1–8, 2016. 8

[6] Paul E. Debevec and Jitendra Malik. Recovering high dynamic range radiance maps from photographs. In *SIGGRAPH*, pages 369–378, 1997. 4

[7] Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. Deep image homography estimation, 2016. Preprint. 4

[8] Daniel Duckworth, Peter Hedman, Christian Reiser, Peter Zhizhin, Jean-François Thibert, Mario Lučić, Richard Szeliski, and Jonathan T. Barron. Smerf: Streamable memory efficient radiance fields for real-time large-scene exploration, 2023. 2

[9] Graham Finlayson, Han Gong, and Robert B. Fisher. Color homography: theory and applications. *IEEE TPAMI*, 41(1): 20–33, 2017. 4

[10] Daniel B Goldman. Vignette and exposure calibration and compensation. *IEEE transactions on pattern analysis and machine intelligence*, 32(12):2276–2288, 2010. 4

[11] Michael D. Grossberg and Shree K. Nayar. Determining the camera response from images: What is knowable? *IEEE TPAMI*, 25(11):1455–1467, 2003. 2

[12] Michael D. Grossberg and Shree K. Nayar. Modeling the space of camera response functions. *IEEE TPAMI*, 26(10): 1272–1282, 2004. 4

[13] Xin Huang, Qi Zhang, Ying Feng, Hongdong Li, Xuan Wang, and Qing Wang. Hdr-nerf: High dynamic range neural radiance fields. In *CVPR*, pages 18398–18408, 2022. 2, 6, 7, 5

[14] Xin Huang, Qi Zhang, Ying Feng, Hongdong Li, and Qing Wang. Ltm-nerf: Embedding 3d local tone mapping in hdr neural radiance field. *IEEE TPAMI*, 2024. 2

[15] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42(4), 2023. 6, 7

[16] Agnan Kessy, Alex Lewin, and Korbinian Strimmer. Optimal whitening and decorrelation. *The American Statistician*, 72 (4):309–314, 2018. 4

[17] Arno Knapitsch, Jaesik Park, Qian-Yi Zhou, and Vladlen Koltun. Tanks and temples: Benchmarking large-scale scene reconstruction. *ACM Transactions on Graphics*, 36(4), 2017. 6, 5

[18] Ricardo Martin-Brualla, Noha Radwan, Mehdi S. M. Sajjadi, Jonathan T. Barron, Alexey Dosovitskiy, and Daniel Duckworth. Nerf in the wild: Neural radiance fields for unconstrained photo collections. In *CVPR*, pages 7210–7219, 2021. 1, 2, 6, 7

[19] Ben Mildenhall, Peter Hedman, Ricardo Martin-Brualla, Pratul P. Srinivasan, and Jonathan T. Barron. NeRF in the dark: High dynamic range view synthesis from noisy raw images. In *CVPR*, 2022. 2, 6

[20] Michael Niemeyer, Fabian Manhardt, Marie-Julie Rakotosaona, Michael Oechsle, Christina Tsalicoglou, Keisuke Tateno, Jonathan T. Barron, and Federico Tombari. Learning neural exposure fields for view synthesis. In *NeurIPS*, 2025. to appear. 2

[21] Emmanuel Onzon, Fahim Mannan, and Felix Heide. Neural auto-exposure for high-dynamic range object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021. 3

[22] Linfei Pan, Daniel Barath, Marc Pollefeys, and Johannes Lutz Schönberger. Global Structure-from-Motion Revisited. In *ECCV*, 2024. 6

[23] Keunhong Park, Philipp Henzler, Ben Mildenhall, Jonathan T. Barron, and Ricardo Martin-Brualla. Camp: Camera preconditioning for neural radiance fields. *ACM TOG*, 42(6):1–11, 2023. 4

[24] Konstantinos Rematas, Andrew Liu, Pratul P. Srinivasan, Jonathan T. Barron, Andrea Tagliasacchi, Tom Funkhouser, and Vittorio Ferrari. Urban radiance fields. *CVPR*, 2022. 1, 2

[25] Darius Rückert, Linus Franke, and Marc Stamminger. Adop: Approximate differentiable one-pixel point rendering. *ACM TOG*, 41(4):99:1–99:14, 2022. 2, 6, 7, 8, 1, 3, 4, 5

[26] Johannes Lutz Schönberger and Jan-Michael Frahm. Structure-from-motion revisited. In *CVPR*, 2016. 6

[27] Gaurav Sharma, Wencheng Wu, and Edul N Dalal. The ciede2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations. *Color Research & Application: Endorsed by Inter-Society Color Council, The Colour Group (Great Britain), Canadian Society for Color, Color Science Association of Japan, Dutch Society for the Study of Color, The Swedish Colour Centre Foundation, Colour Society of Australia, Centre Français de la Couleur*, 30(1):21–30, 2005. 6, 2

[28] Jisu Shin, Richard Shaw, Seunghyun Shin, Zhensong Zhang, Hae-Gon Jeon, and Eduardo Perez-Pellitero. Chroma: Consistent harmonization of multi-view appearance via bilateral grid prediction, 2025. Preprint. 2

[29] Pei Sun, Henrik Kretzschmar, Xerxes Dotiwalla, Aurelien Chouard, Vijaysai Patnaik, Paul Tsui, James Guo, Yin Zhou, Yuning Chai, Benjamin Caine, Vijay Vasudevan, Wei Han, Jiquan Ngiam, Hang Zhao, Aleksei Timofeev, Scott Ettinger, Maxim Krivokon, Amy Gao, Aditya Joshi, Yu Zhang, Jonathon Shlens, Zhifeng Chen, and Dragomir Anguelov. Scalability in perception for autonomous driving: Waymo open dataset. In *CVPR*, 2020. 6, 5

[30] Matthew Tancik, Vincent Casser, Xinchen Yan, Sabeek Pradhan, Ben Mildenhall, Pratul P. Srinivasan, Jonathan T. Barron, and Henrik Kretzschmar. Block-nerf: Scalable large scene neural view synthesis. In *CVPR*, pages 8248–8258, 2022. 2

[31] Justin Tomasi, Brandon Wagstaff, Steven L Waslander, and Jonathan Kelly. Learned camera gain and exposure control for improved visual feature detection and matching. *IEEE Robotics and Automation Letters*, 6(2):2028–2035, 2021. 3

[32] Alex Trevithick, Roni Paiss, Philipp Henzler, Dor Verbin, Rundi Wu, Hadi Alzayer, Ruiqi Gao, Ben Poole, Jonathan T. Barron, Aleksander Holynski, Ravi Ramamoorthi, and Pratul P. Srinivasan. Simvs: Simulating world inconsistencies for robust view synthesis. *arXiv*, 2024. 2

[33] Yuehao Wang, Chaoyi Wang, Bingchen Gong, and Tianfan Xue. Bilateral guided radiance field processing. *ACM TOG*, 43(4):148:1–148:13, 2024. 1, 2, 6, 7, 8, 3, 5

[34] Qi Wu, Janick Martinez Esturo, Ashkan Mirzaei, Nicolas Moenne-Loccoz, and Zan Gojcic. 3dgut: Enabling distorted cameras and secondary rays in gaussian splatting. In *CVPR*, 2025. 6, 7, 8, 3

[35] Wenqi Xian, Aljaž Božič, Noah Snavely, and Christoph Lassner. Neural lens modeling. In *CVPR*, pages 8435–8445, 2023. 2

[36] Vickie Ye, Ruilong Li, Justin Kerr, Matias Turkulainen, Brent Yi, Zhuoyang Pan, Otto Seiskari, Jianbo Ye, Jeffrey Hu, Matthew Tancik, and Angjoo Kanazawa. gsplat: An open-source library for gaussian splatting. *Journal of Machine Learning Research*, 26(34):1–17, 2025. 6, 7

[37] Dongbin Zhang, Chuming Wang, Weitao Wang, Peihao Li, Minghan Qin, and Haoqian Wang. Gaussian in the wild: 3d gaussian splatting for unconstrained image collections. In *European Conference on Computer Vision*, pages 341–359. Springer, 2024. 2