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

This supplementary material provides additional experiments, method details, and implementation specifications to complement the main paper.

Sec. A presents extended experimental results, including a detailed comparison with ADOP's image formation model [25] and additional experiments on components (camera calibration and exposure identifiability).

Sec. B provides further method details, *i.e.*, mathematical derivations of our color correction formulation and specifications of our per-frame controller architecture.

Sec. C details optimization settings, regularization weights, learning rate schedules, and dataset specifications used throughout our experiments.

Finally, Sec. D discusses interactive manual control capabilities of our method.

A. Additional Experiments

To complete the main paper experiments, we provide further qualitative results in Fig. 5 and present the detail of the novel-view PSNR for every scene in Tab. 6.

A.1. Detailed Comparison with ADOP [25]

In the related work (Sec. 2), we mention that ADOP [25] implements a similar image formation model as ours. We deviate in the color correction and CRF. Here, we provide a detailed comparison, expanding on the main results in Sec. 5.

White balance and exposure decoupling. In Sec. 4.3, we claim that our color correction method, which operates on 2D chromaticities instead of 3D color and normalizes the intensity post-transformation, decouples the white balance from the exposure correction. We evaluate this by computing the Pearson correlation coefficient (PCC) between the estimated exposure offset and the white point offset, $\Delta\mathbf{c}_W$, which controls the white balance and compare our method against ADOP's which uses per-channel white-point gains.

The PCC is defined as:

$$r_{X,Y} = \frac{\mathrm{cov}(X, Y)}{\sigma_X \sigma_Y} \tag{23}$$

where $r_{X,Y}$ is the Pearson correlation coefficient between variables X and Y, $\mathrm{cov}(X, Y)$ is the covariance between X and Y, and σ_X and σ_Y are the standard deviations of X and Y, respectively. A PCC near 1 indicates strong linear correlation, and a PCC near 0 indicates weak or no correlation.

A representative result is shown in Fig. 6. We find that the PCC numbers for our method are substantially lower as compared to ADOP's method on all sequences, indicating an improved decoupling of white balance and exposure correction.

Figure 7 further highlights the importance of decoupling color and exposure corrections: When exposure and color are coupled, the CRF will also be entangled in order to compensate for the value-dependent color shift. That, in turn, hinders the controllability of both aspects since neither can be changed without also affecting the other.

CRF stability in challenging sequences. In Sec. 4.4, we provide a formulation for the camera response function that is constrained to be monotonically increasing and smooth by design. This ensures that the optimization remains stable. In some sequences, particularly when large photometric variations were present, we found that this offers an improvement over ADOP's [25] CRF formulation, which uses 25 discrete nodes which are interpolated linearly and requires a smoothness loss. A degenerate case of ADOP's CRF is illustrated in Fig. 7 (third row), where the learned green and red channels of the CRF are split into lower and upper sections with a reversal. This violates the assumption that the CRF is monotonically increasing. While the post-processed image still remains close in brightness and color to the actual scene due to corrections being self-consistent, it falls apart with strong color artifacts when applying a controlled exposure offset.

A.2. Online Camera Calibration

Since certain parts of the PPISP pipeline, namely the vignetting (Sec. 4.2) and CRF (Sec. 4.4), are shared across all frames of a camera, the process of jointly optimizing them with the radiance field reconstruction can be understood as an online camera calibration. We compared the recovered per-camera parameters across multiple sequences qualitatively in Fig. 8, where multiple plots are overlaid. Same color implies same dataset. The close overlap of the curves from the same datasets and the distinct shapes between datasets indicate that our method can robustly extract these calibrations. It also suggests that the camera-specific curves are disentangled from scene radiance and other corrective effects, otherwise we would expect an ambiguous mixing of them.

A.3. Identifiability of Exposure Offsets

In Sec. 5.2, we tested the effectiveness of using image exposure metadata to guide the image formation process. Here, we consider the inverse problem of identifying calibrated exposure offsets. In this experiment, per-frame exposure

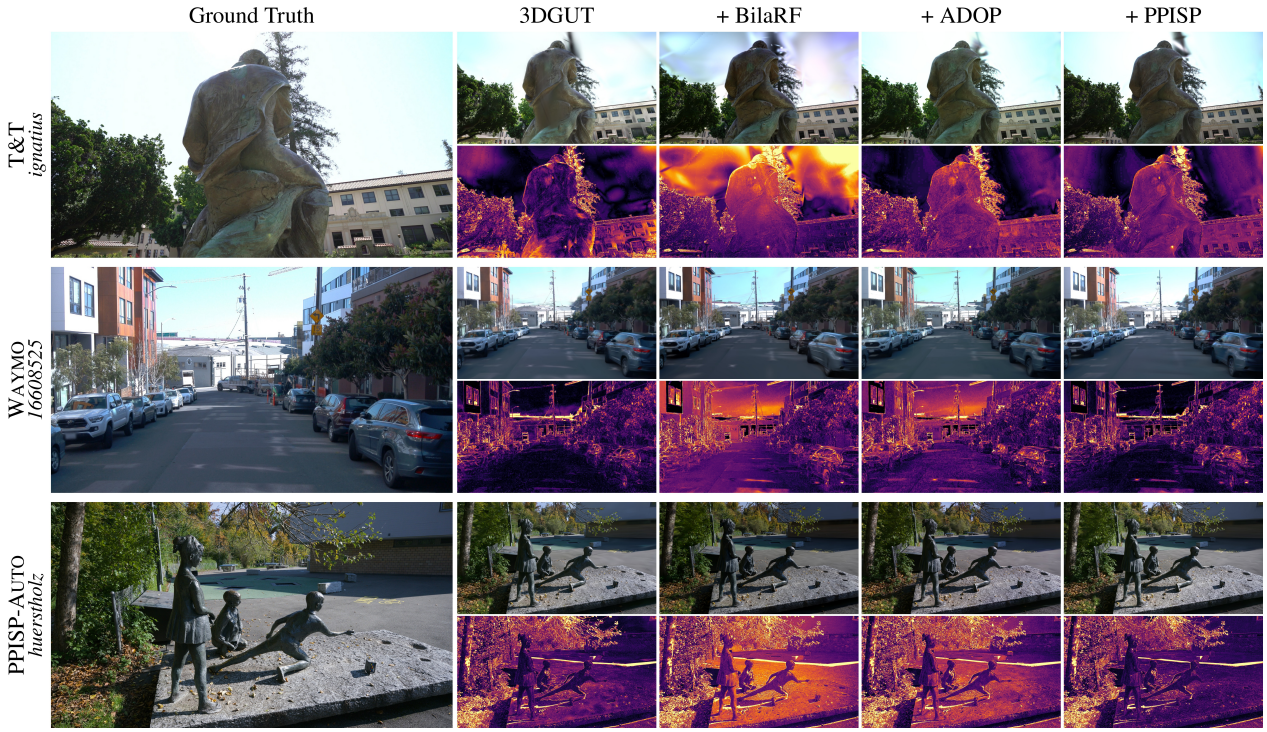

Figure 5. Qualitative comparison of novel view synthesis, additional examples. Row labels indicate datasets and sequences (in *italics*). Column labels indicate methods. Heat maps show perceptual CIEDE2000 [27] error (colormap range: 0–20 ΔE_{00}).

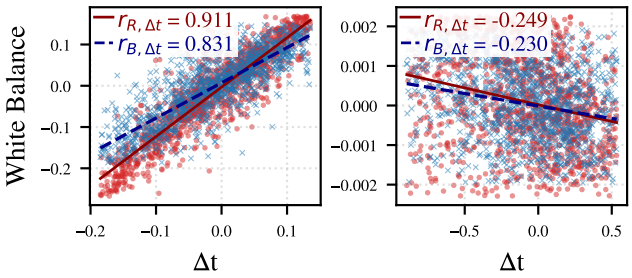

Figure 6. Correlation between optimized exposure offset and white balancing variables in SMERF's [8] *alameda* sequence. Left: ADOP's [25] red and blue channel scaling. Right: The offsets of the white point of our homography-based correction. The Pearson correlation coefficient for each component is inset.

offsets are freely optimized and compared against the relative exposure metadata present in the HDR-NeRF [13] and PPISP datasets.

According to Grossberg and Nayar [11], there is an "exponential ambiguity", which states that transforming both the inverse of the CRF and the radiance by some power produces exactly the same image intensities. Since our exposure offsets are parameterized in log-space, applying a power to the radiance corresponds to a scaling in parameter space. Thus, for this experiment, we apply an optimal affine transform on the recovered exposure offsets and compute the error on the transformed data.

As illustrated in Fig. 9 for a representative sequence, calibrated exposure metadata is matched closely.

B. Additional Method Details

B.1. Color Correction

In Sec. 4.3, we propose a color correction method based on a 3×3 homography matrix $\mathbf{H}$, applied on RG chromaticities and intensity, followed by an intensity normalization. For the parameterization of $\mathbf{H}$, we show a construction from chromaticity offsets $\Delta\mathbf{c}_k$ that control the mapping from source to target chromaticities. In this section, we provide a more detailed derivation.

Furthermore, we detail the preconditioning we apply to the chromaticity offsets $\Delta\mathbf{c}_k$.

Derivation and equivalence to direct linear transformation. We derive the construction of $\mathbf{H}$ in detail and show that the resulting matrix is equivalent to the standard method for constructing homography matrices from source-target pairs, the direct linear transformation (DLT).

In Sec. 4.3, we define source and target chromaticity vector pairs $\mathbf{c}_{\{s,t\},\{R,G,B,W\}}$. The homogeneous lifts of these vectors are denoted with a tilde, $\tilde{\mathbf{c}}_{\{s,t\},\{R,G,B,W\}}$. The $\mathbf{S}$ and $\mathbf{T}$ matrices are built by stacking the lifted source and

Table 6. **Per-scene novel view PSNR comparison.** We compare post-processing methods applied on top of 3DGUT reconstruction across all sequences. Higher is better (↑).

Dataset	Scene	3DGUT [34]	+ BilaRF [33]	+ ADOP [25]	+ PPISP (w/o ctrl.)	+ PPISP (w/ ctrl.)
			BILARF			
	building	24.85	22.81	25.30	26.36	**26.46**
	chinesearch	18.34	20.44	21.27	**22.13**	21.62
	lionpavilion	24.16	24.11	22.89	**25.06**	24.76
	nighttimepond	27.11	21.54	25.07	27.68	**28.16**
	pondbike	25.28	21.17	24.96	**26.33**	26.04
	statue	22.40	21.01	**22.84**	**22.84**	22.26
	strat	16.06	18.76	18.34	18.17	**19.55**
			MIP-NERF 360			
	bicycle	25.28	24.26	24.54	24.95	**25.72**
	bonsai	32.52	28.57	30.33	32.10	**33.02**
	counter	29.36	26.30	27.58	28.89	**29.50**
	flowers	21.80	20.10	21.54	21.76	**21.95**
	garden	26.85	24.06	26.10	27.14	**27.31**
	kitchen	31.86	27.50	28.08	30.51	**32.14**
	room	32.11	29.53	30.76	32.95	**32.84**
	stump	26.90	24.90	26.59	27.03	**27.28**
	treehill	22.97	19.46	22.25	22.59	**23.55**
			TANKS AND TEMPLES			
	caterpillar	22.61	19.19	18.15	19.74	**25.18**
	ignatius	22.03	20.01	20.47	20.77	**24.04**
	train	22.06	19.04	18.95	20.17	**23.74**
	truck	24.72	20.88	23.56	25.38	**25.51**
			WAYMO			
	10275144660749673822_5755_561_5775_561	24.73	20.59	23.68	24.30	**25.17**
	1265122081809781363_2879_530_2899_530	**28.39**	24.47	26.30	27.50	28.31
	15959580576639476066_5087_580_5107_580	27.52	24.06	26.54	27.04	**27.77**
	16470190748368943792_4369_490_4389_490	23.82	20.17	22.09	23.69	**24.21**
	16608525782988721413_100_000_120_000	**23.29**	19.86	22.62	22.91	23.27
	16646360389507147817_3320_000_3340_000	**26.65**	23.71	24.84	25.86	26.48
	17244566492658384963_2540_000_2560_000	27.25	22.19	26.00	26.31	**27.39**
	1999080374382764042_7094_100_7114_100	24.10	20.85	23.18	23.65	**24.34**
	744006317457557752_2080_000_2100_000	24.26	20.53	23.31	24.05	**24.30**
			PPISP-AUTO			
	huerstholz_auto	19.23	18.76	18.88	19.24	**19.81**
	struktur28_auto	24.21	22.80	21.97	22.25	**25.28**
	toro_auto	22.24	20.56	18.44	20.20	**23.01**
	valiant_auto	22.51	21.14	20.47	22.58	**23.39**

target red, green, and blue chromaticity vectors, respectively. We note that $\mathbf{S}$ is constant and has an inverse $\mathbf{S}^{-1}$.

Reduction using three correspondences. By definition, a homography is a collinear transformation (collineation), *i.e.*, transformed vectors are identical to the original ones up to scale: $\mathbf{H}\,\tilde{\mathbf{c}}_{s,i} \sim \tilde{\mathbf{c}}_{t,i}$ for $i \in \{R, G, B\}$. Using the stacked matrices $\mathbf{S}$ and $\mathbf{T}$, it follows that there exist nonzero $\mathbf{k} = (k_R, k_G, k_B)^\top$ such that

$$\mathbf{H}\,\mathbf{S} = \mathbf{T}\,\mathrm{diag}(\mathbf{k}) \Longrightarrow \mathbf{H}(\mathbf{k}) = \mathbf{T}\,\mathrm{diag}(\mathbf{k})\,\mathbf{S}^{-1}\,. \quad (24)$$

Thus, the homography is reduced to three column scales up to a common factor.

Fourth correspondence via collinearity. To find $\mathbf{k}$, we write the source white point as $\tilde{\mathbf{c}}_{s,W} = \mathbf{S}\,\mathbf{b}$ with barycentric $\mathbf{b} = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})^\top$.

We require $\mathbf{H}\,\tilde{\mathbf{c}}_{s,W} \sim \tilde{\mathbf{c}}_{t,W}$. Another way to express this collinearity constraint is $\tilde{\mathbf{c}}_{t,W} \times \big(\mathbf{T}\,\mathrm{diag}(\mathbf{b})\,\mathbf{k}\big) = \mathbf{0}$. Using the skew-symmetric matrix $[\cdot]_\times$ with $[\mathbf{x}]_\times \mathbf{y} = \mathbf{x} \times \mathbf{y}$, this yields the homogeneous linear system

$$[\tilde{\mathbf{c}}_{t,W}]_\times\,\mathbf{T}\,\mathrm{diag}(\mathbf{b})\,\mathbf{k} = \mathbf{0}\,.$$

For the white point, $\mathrm{diag}(\mathbf{b}) \propto \mathbf{I}$, so the constraint reduces to the 3×3 system $\mathbf{M}\,\mathbf{k} = \mathbf{0}$ with $\mathbf{M} = [\tilde{\mathbf{c}}_{t,W}]_\times\,\mathbf{T}$. Generically $\mathrm{rank}(\mathbf{M}) = 2$, so the right nullspace is 1D and determines $\mathbf{k}$ up to scale. A practical closed form is

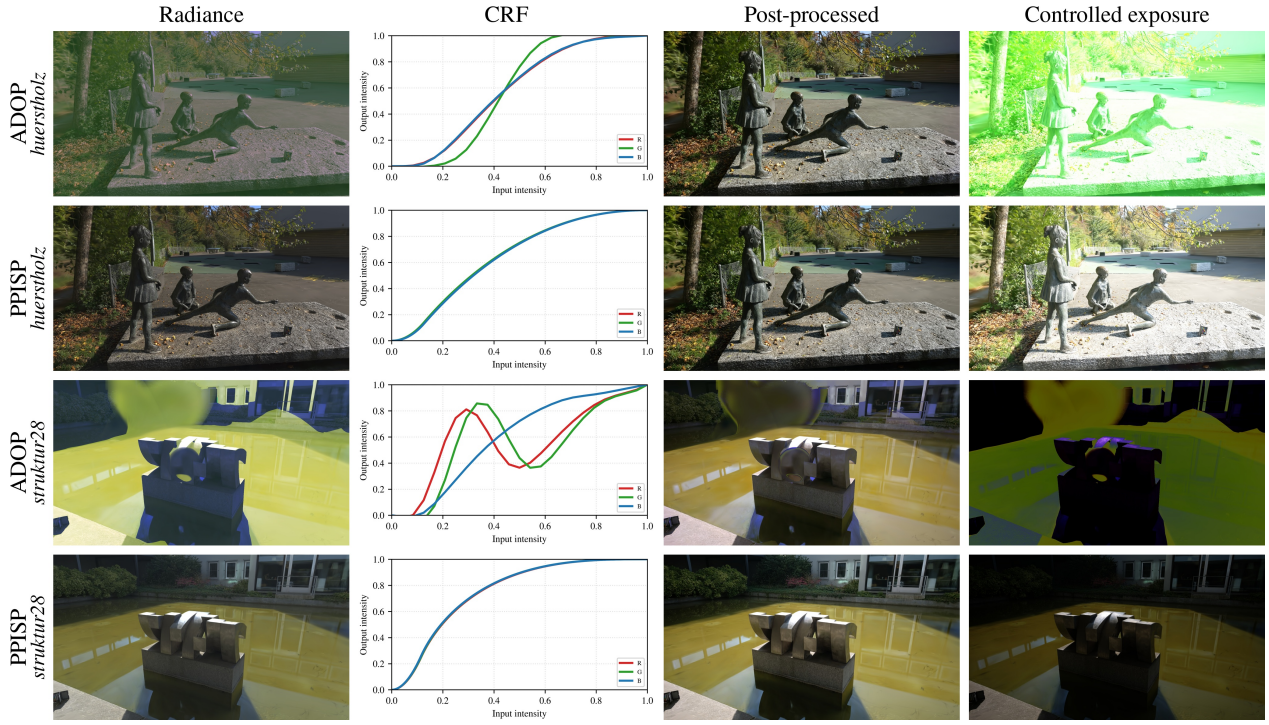

Figure 7. Comparison of ADOP [25]-style post-processing including exposure control against our method. Row labels indicate the post-processing method and the sequence name (in italics). The CRF for ADOP's formulation compensates for the color artifacts baked into the radiance field only at a specific exposure value. But when controlling exposure for novel views, color artifacts are exacerbated. In contrast, both our method's radiance field and output remain neutral since all corrections are decoupled.

to take any cross of two independent rows $\mathbf{r}_i, \mathbf{r}_j$ of $\mathbf{M}$, *i.e.*: $\mathbf{k} \propto \mathbf{r}_i \times \mathbf{r}_j$. Substituting $\mathbf{k}$ into $\mathbf{H}(\mathbf{k})$ and normalizing by an arbitrary scalar (*e.g.*, set $[\mathbf{H}]_{3,3} = 1$) gives the desired homography.

Equivalence to the 4-point DLT. The classical DLT stacks the four constraints into $\mathbf{A}\,\mathbf{h} = \mathbf{0}$ for the 9-vector $\mathbf{h}$ of $\mathbf{H}$ (up to scale), and solves for the 1D right-nullspace of $\mathbf{A}$. Our construction enforces the same constraints factorized through the invertible $\mathbf{S}$: three correspondences reduce to the column scales $\mathbf{k}$, and the fourth yields $\mathbf{M}\,\mathbf{k} = \mathbf{0}$. Under non-degenerate configurations (*i.e.*, the columns of $\mathbf{T}$ are not collinear and $\mathrm{rank}(\mathbf{M}) = 2$), both methods recover the same $\mathbf{H}$ up to an overall scalar.

Degeneracies and identity case. If $\mathrm{rank}(\mathbf{T}) < 2$ or $\mathrm{rank}(\mathbf{M}) < 2$, $\mathbf{k}$ is ill-defined, mirroring DLT degeneracies. When targets equal sources, $\mathbf{T} = \mathbf{S}$, $\tilde{\mathbf{c}}_{t,W} = \tilde{\mathbf{c}}_{s,W}$, and $\mathbf{k} \propto (1, 1, 1)$, yielding $\mathbf{H}$ proportional to the identity after normalization.

Preconditioning of the chromaticity offsets. Our color correction method involves a conversion from RGB color to RGI (red-green chromaticity and intensity) and back, with $I = R+G+B$ and $B = I-R-G$ in terms of components. In our optimization setting, this correlates the gradients of the individual chromaticity offsets $\{\Delta\mathbf{c}_i\}$ with the blue channel. In addition to that, the output image is generally more sensitive to changes in the white point than an offset in the RGB primaries.

In order to whiten the color correction and decorrelate the individual components, we apply ZCA preconditioning with proxy Jacobians following [16, 23]. We precondition the 8-dimensional vector of chromaticity offsets $\{\Delta\mathbf{c}_i\}_{i\in\{R,G,B,W\}}$. We use a block decomposition into four 2×2 blocks (one per control point) in place of the full 8×8 transform.

B.2. Controller Architecture

The overall architecture of the per-frame ISP controller is given in Sec. 4.5. Here, we provide the complete architectural specifications.

Input and output. The controller takes as input the rendered scene radiance $\mathbf{L} \in \mathbb{R}^{H\times W\times 3}$. Extra inputs, such as image metadata, are input at the beginning of the parameter regression stage.

The controller outputs 9 parameters: an exposure offset $\Delta t \in \mathbb{R}$ and eight color correction offsets $\{\Delta\mathbf{c}_i\}_{i\in\{R,G,B,W\}}$.

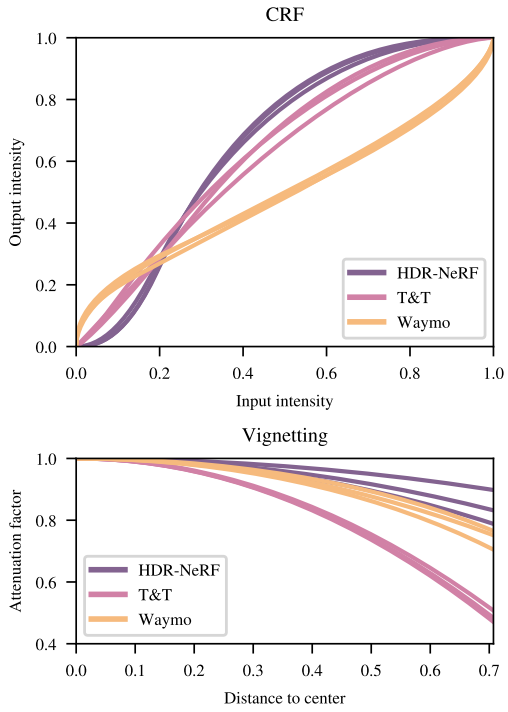

Figure 8. Recovered camera-specific parameters across datasets. Top: The calibrated CRF of three sequences of each of the HDR-NeRF [13], Tanks and Temples [17], and Waymo Open Drive [29] dataset are overlaid. Bottom: For the same sequences and datasets, the vignetting falloff curves are compared.

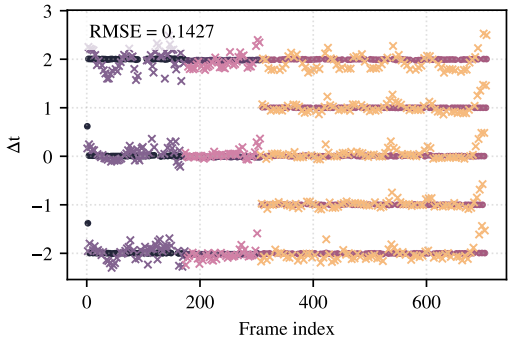

Figure 9. Optimized exposure parameters per frame and given exposure metadata for the *huerstholz* sequence in the PPISP dataset. Colors indicate individual cameras.

Feature extraction stage. The feature extractor processes the input radiance using a sequence of 1x1 convolutions and pooling operations.

First, a 1x1 convolution maps the 3-channel input to 16 feature channels. This is followed by max pooling with a factor of 3 in each spatial dimension, reducing the resolution to $H/3 \times W/3$. A ReLU activation is then applied. Next, a second 1x1 convolution expands the features to 32 channels, followed by ReLU. A third 1x1 convolution produces 64 feature channels, yielding a feature map $\mathbf{F} \in \mathbb{R}^{H/3 \times W/3 \times 64}$.

Then, spatial aggregation is performed. An adaptive average pooling operation reduces the spatial dimensions to a 5×5 grid, producing a coarse feature representation $\mathbf{F}_{\text{pool}} \in \mathbb{R}^{5 \times 5 \times 64}$. This grid captures multi-scale spatial statistics of the scene while maintaining spatial locality, analogous to metering zones in conventional cameras.

Parameter regression stage. The pooled features are flattened into a 1600-dimensional vector ($5 \times 5 \times 64$). If available, image metadata may be concatenated at this stage. This is input into an MLP with three hidden layers, each containing 128 neurons with ReLU activations. The output consists of two parallel linear heads: one producing the exposure offset and the other producing the 8 color correction parameters.

C. Additional Experiment Details

We provide optimization hyperparameters, regularization weights, and dataset specifications used throughout our experiments.

C.1. Optimization settings

Regularization weights. In Sec. 4.6, we specify the regularizer terms that break brightness and color ambiguities and ensure physically-plausible vignetting. In Tab. 7, we detail the numerical values used for each λ term.

Table 7. Regularization coefficients.

Term	λ
λ_b	1.0
λ_c	1.0
λ_{var}	0.1
λ_v	0.01

Optimizer, learning rates, and schedules. For all post-processing modules including BilaRF [33], ADOP's formulation [25], and our method, we use the Adam optimizer. We use the following learning rate scheduling with an initial delay (zero learning rate), linear warmup, and exponential decay.

$$lr(s) = \begin{cases} 0, & s < s_d, \\ lr_0 \left[f_s + (1 - f_s) \dfrac{s - s_d}{s_w} \right], & s_d \le s < s_d + s_w, \\ lr_0 \left(f_f^{1/s_{\max}} \right)^{s - s_d - s_w}, & s \ge s_d + s_w. \end{cases} \tag{25}$$

Where:

- lr_0 — base learning rate.
- s — current training step.
- s_d — delay steps (learning rate held at zero).

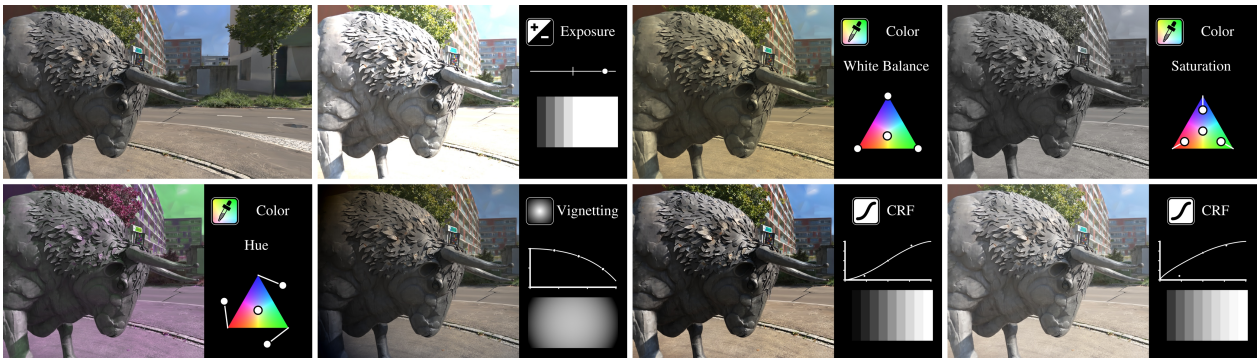

Figure 10. Our low-parametric formulation of the different image processing steps enables manual editing. Top left shows the input image. Other images have details overlaid, such as the primary effect being applied and an abstract visualization. In the color correction examples, the white dots correspond to the four target chromaticities $\mathbf{c}_{t,\{R,G,B,W\}}$, which can be intuitively manipulated.

- s_w — warmup steps (linear ramp from $f_s lr_0$ to lr_0).
- $s_{\max}$ — number of decay steps.
- f_s — start factor for warmup (e.g., 0.01).
- f_f — final factor reached after decay (e.g., 0.01).

Tab. 8 details the values used during experiments.

Table 8. Learning rate scheduler hyperparameters.

Term	Value
lr_0	0.002
s_d	0
s_w	500
f_s	0.01
$s_{\max}$	30000
f_f	0.01

In Sec. 5.4, we experiment with combined post-processing methods. In these cases, the BilaRF module as combined with PPISP and per-camera bilateral grids uses $s_d = 5000$ and $s_w = 1000$ with otherwise the same hyperparameters as in Tab. 8.

C.2. Datasets

In Sec. 5, we outline the datasets used for experiments. In this section, we define the datasets in more detail.

Specific choice of sequences. We chose the following sequences from each dataset:

- Mip-NeRF 360 [2]: All nine sequences,
- Tanks and Temples [17]: Four sequences, namely *train*, *truck*, *caterpillar*, and *ignatius*,
- BilaRF [33]: All seven sequences,
- HDR-NeRF [13]: All four real-camera sequences,
- Waymo Open Dataset [29]: Nine mostly static sequences, explicitly listed in Tab. 9; All five cameras used.

PPISP dataset details. As stated in Sec. 5, we captured our own dataset using three cameras, including two modern mirrorless and a smartphone camera. We provide further context here.

Table 9. Waymo Open Dataset [29] sequence names.

Sequence Name
7440631745755752_2080_000_2100_000
126512208180978136_2879_530_2899_530
199908037438276404_7094_100_7114_100
102751446607496738_5755_561_5775_561
159595805766394760_5087_580_5107_580
164701907483689437_4369_490_4389_490
166085257829887214_100_000_120_000
166463603895071478_3320_000_3340_000
172445664926583849_2540_000_2560_000

For all cameras and scenes, we used exposure bracketing of ± 2 EV to capture HDR data. The aperture and focus were set manually and remained fixed. Image stabilization was disabled. Each scene was captured in raw format. The raw photos were developed with NX Studio and OM Workspace for the Nikon and OM System photos, and Adobe Lightroom Classic for the iPhone photos, respectively. A color calibration target placed in the scene was used to white balance.

For each scene, we additionally picked certain exposures out of the brackets and re-developed them with normalized, automatic exposure compensation and white balancing, creating a more challenging setting for the controller module. We denote this derived dataset *PPISP-auto*.

Pre-processing. For all datasets including our own, where camera poses or sparse point clouds were not originally available, we processed them through COLMAP [26] and GLOMAP [22] to produce the necessary inputs for the radiance field reconstruction.

We used downsampled versions of the original camera images so that the maximum effective side length of each

input image did not exceed 2000 pixels. *E.g.*, for Mip-NeRF 360's [2] *garden* sequence, we used $4\times$ downsampling, and for *bonsai*, we used $2\times$.

We used a seven to one split of test views to validation views for evaluation throughout.

D. Manual Control

Our parametric ISP formulation enables intuitive manual editing and artistic control. Fig. 10 demonstrates various edits applied to a reconstructed scene, including adjustments to exposure, white balance, vignetting, and camera response. The low-dimensional and disentangled representation ensures meaningful and predictable edits, facilitating interactive workflows for applications such as artistic rendering, temporal consistency enforcement, or selective photometric matching.