Benchmarking Burst Super-Resolution for Polarization Images: Noise Dataset and Analysis

Supplementary Material

A.1. Polarization Super-resolution Image Processing Details

The camera used is the same model as the one used to obtain the noise statistics. To achieve a 4× super-resolution configuration like that of the BurstSR dataset, we capture the burst images with an 8mm lens, which has a focal length approximately four times shorter. The reference frame is set to the first frame of the burst image. The ground truth image is debayered into 12-channel images, consisting of three colors multiplied by four polarization angles, using bilinear interpolation. The polarization angle images are treated as occupying the same pixel locations, which results in the spatial resolution of the ground truth being halved at this stage. The homography between the ground truth and the reference frame is calculated using the SIFT feature [6] and RANSAC matching [4]. The ground truth images are warped to align with the reference frame of the burst images using the calculated homography. While warping the image pixels, the s_1 and s_2 values are adjusted based on the rotation from the homography. Due to the multiple device acquisitions, the burst frames and ground truth images exhibit misalignments caused by disparity and color differences. We exclude image pairs with a normalized cross-correlation lower than 0.9 and adjust the image colors using a per-image color correction matrix. The images are cropped to sizes of 192×192 and 384×384 for frames, and 384×384 for the ground truth, resulting in an overall super-resolution of 2×. The actual dimensions for input and output in the networks are 48×48×16 and 384×384×9, respectively, yielding a spatial resolution ratio of 8×, which aligns with the original BurstSR and many burst super-resolution models.

A.2. Training Details

Training the network requires a massive labeled dataset. However, acquiring a real burst image dataset involves high costs and significant effort, making it challenging to collect a sufficient amount of data through real captures. Instead, we generate synthetic data from other unlabeled polarization image datasets. We first train the network using synthetic data and then fine-tune it with real image data.

Synthetic data training. In the synthetic data training, we follow the synthetic burst image generation method outlined in DBSR[1]. We generate synthetic burst polarization images from the RSP dataset[5]. The RSP dataset includes 1586 and 176 synthetic polarization images in the training and validation sets, respectively, along with 238 carefully captured real polarization images in the test set. We maintain the original order and use their separation directly. Unlike DBSR, the RSP dataset is already in linear RGB color space, so we do not apply the unprocessing pipeline[2] to extract raw pixel values. For each generation, we first randomly crop the image to obtain the labeled ground truth and the 0th frame. For the other frames, we include additional random translations and rotations within the ranges of [-24, 24] pixels and [-1, 1] degrees. Since the input to FBANet[8] consists of aligned images, we do not introduce random movement for FBANet in the synthetic data training. Subsequently, the frames are downsampled by a factor of 2 and masked using a polarization Bayer pattern that matches the off-the-shelf polarization image sensor. Finally, random synthetic noise is added to each pixel based on the sensor noise distribution defined by Brooks et al.[2]. The final ground truth has 384×384×12, which means rows×columns×RGB and polarization channels, while each frame has 48×48×16, which means rows of Bayer pattern lattices ×columns of that×variations of the polarization Bayer pattern.

For actual training, we use 14 generated burst frames for each labeled sample. We implement the network in PyTorch. The network is trained using the ADAM optimizer with a fixed learning rate of 1e-4 for 500,000 iterations, with a batch size of 12. The loss function used is the L1 loss, which is commonly employed in previous burst SR works.

Real data training. The frames and GTs in the real burst SR dataset have positional misalignment and color differences resulting from being captured by two different devices in varying positions [8]. Our dataset also has these intrinsic problems, so we use the aligned losses and metrics proposed in DBSR [1]. The original metric is developed for RGB images; therefore, the calculation of the optical flow, color mapping, and validation mask utilizes s_0 images. Training begins with the trained network from the synthetic data, and the network is fine-tuned over 50k iterations. The other settings remain the same as those used in synthetic data training.

A.3. Validation of Stokes Vector Noise Model

To demonstrate the dependency of the noise in the Stokes vector components and the true s_0 value, we obtain a histogram of the signal-to-noise ratio for each Stokes vector component. We calculate the variance of the observed values to illustrate the linear relationship with the true value. Histograms are generated for the three Stokes vector components of signal and noise, resulting in a total of nine histograms. The results are presented in Figure 1. The rows represent the variance of each component, indicating the noise, while the columns denote the mean of each component, reflecting the signal. The first column clearly demonstrates the linear relationship between the true s_0 value and each component of the Stokes vector. In contrast, the histograms comparing the noise with the s_1 and s_2 signals do not show a clear dependency in the second and third columns. In contrast, the histograms comparing the noise with the s_1 or s_2 signals do not exhibit a clear dependency in the second and third columns.

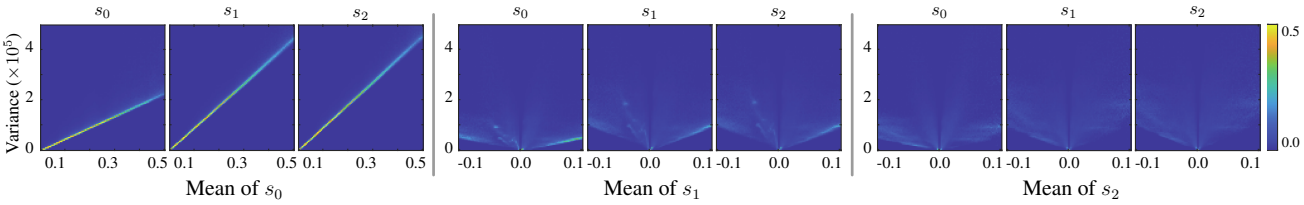

Figure 1. Histogram of noise for each component based on the values of the Stokes vector components. The noise of s_0, s_1, and s_2 is strongly correlated with the s_0 value.

A.4. Detail Derivation Process of the AoLP Distribution

The noise model of AoLP is the marginal distribution of the phase of $\hat{s}_1$ and $\hat{s}_2$. It can be represented by the integration of polar coordinates, which is expressed as:

$$\begin{aligned} f_\phi\left(2\hat{\phi}|s_1, s_2, \sigma_v^2\right) &= \int_0^\infty f\left(\hat{s}_1, \hat{s}_2\right) r dr \\ &= \int_0^\infty f\left(r\cos 2\hat{\phi}, r\sin 2\hat{\phi}\right) r dr \end{aligned}$$

We expand the Gaussian distribution and organize it using the angle addition and subtraction of trigonometric functions. It is expressed as follows:

$$\begin{aligned} f_\phi\left(2\hat{\phi}|s_1, s_2, \sigma_v^2\right) &= \int_0^\infty \frac{1}{2\pi\sigma_v^2} \exp\left(-\frac{\left(r\cos 2\hat{\theta} - s_1\right)^2 + \left(r\sin 2\hat{\theta} - s_2\right)^2}{2\sigma_v^2}\right) r dr \\ &= \int_0^\infty \frac{1}{2\pi\sigma_v^2} \exp\left(-\frac{\left(r\cos 2\hat{\phi} - s_{pol}\cos 2\phi\right)^2 + \left(r\sin 2\hat{\phi} - s_{pol}\sin 2\phi\right)^2}{2\sigma_v^2}\right) r dr \\ &= \int_0^\infty \frac{1}{2\pi\sigma_v^2} \exp\left(-\frac{r^2 + {s_{pol}}^2 - 2rs_{pol}\left(\cos 2\hat{\phi}\cos 2\phi + \sin 2\hat{\phi}\sin 2\phi\right)}{2\sigma_v^2}\right) r dr \\ &= \int_0^\infty \frac{1}{2\pi\sigma_v^2} \exp\left(-\frac{r^2 + {s_{pol}}^2 - 2rs_{pol}\left(\cos\left(2\hat{\phi} - 2\phi\right)\right)}{2\sigma_v^2}\right) r dr \end{aligned}$$

From the integral above, we can see that the ratio of the intensity of the polarization component and Stokes vector noise,s_{pol}/σ_v, is a parameter of the function. Using the following substitution $t = \frac{r}{\sigma_v}$, $dt = \frac{1}{\sigma_v}dr$, and $A =$

$\frac{1}{2\pi}\exp\left(-\frac{{s_{pol}}^2\sin^2\left(2\hat{\phi}-2\phi\right)}{2\sigma_v^2}\right)$, the equation is summarized as follows:

$$
\begin{aligned}
f_\phi\left(2\hat{\phi}|s_1,s_2,\sigma_v^2\right) &= \int_0^\infty \frac{1}{2\pi}\exp\left(-\frac{t^2}{2}-\frac{{s_{pol}}^2}{2\sigma_v^2}-\frac{s_{pol}t\left(\cos\left(2\hat{\phi}-2\phi\right)\right)}{\sigma_v}\right)tdt \\
&= \int_0^\infty \frac{1}{2\pi}\exp\left(-\frac{1}{2}\left(t-\frac{s_{pol}\cos\left(2\hat{\phi}-2\phi\right)}{\sigma_v}\right)^2-\frac{{s_{pol}}^2\sin^2\left(2\hat{\phi}-2\phi\right)}{2\sigma_v^2}\right)tdt \\
&= \int_0^\infty At\exp\left(-\frac{1}{2}\left(t-\frac{s_{pol}\cos\left(2\hat{\phi}-2\phi\right)}{\sigma_v}\right)^2\right)dt
\end{aligned}
$$

Finally, by performing integration using the differentiation $\frac{d}{dx}\exp\left(-\frac{1}{2}x^2\right)=-x\exp\left(-\frac{1}{2}x^2\right)$ and solving the substitution, it is expressed as follows:

$$
\begin{aligned}
f_\phi\left(2\hat{\phi}|s_1,s_2,\sigma_v^2\right) &= \int_0^\infty A\left(t-\frac{s_{pol}\cos\left(2\hat{\phi}-2\phi\right)}{\sigma_v}\right)\exp\left(-\frac{1}{2}\left(t-\frac{s_{pol}\cos\left(2\hat{\phi}-2\phi\right)}{\sigma_v}\right)^2\right)dt \\
&+ \int_0^\infty \frac{As_{pol}\cos\left(2\hat{\phi}-2\phi\right)}{\sigma_v}\exp\left(-\frac{1}{2}\left(t-\frac{s_{pol}\cos\left(2\hat{\phi}-2\phi\right)}{\sigma_v}\right)^2\right)dt \\
&= A\exp\left(-\frac{{s_{pol}}^2\cos^2\left(2\hat{\phi}-2\phi\right)}{2\sigma_v^2}\right)+\frac{\sqrt{2\pi}As_{pol}\cos\left(2\hat{\phi}-2\phi\right)}{\sigma_v}\Phi\left(\frac{s_{pol}\cos\left(2\hat{\phi}-2\phi\right)}{\sigma_v}\right) \\
&= \frac{1}{2\pi}\exp\left(-\frac{{s_{pol}}^2}{2\sigma_v^2}\right)+\frac{\psi\cos\left(2\hat{\phi}-2\phi\right)}{\sqrt{2\pi}\sigma_v}\exp\left(-\frac{{s_{pol}}^2\sin^2\left(2\hat{\phi}-2\phi\right)}{2\sigma_v^2}\right)\Phi\left(\frac{s_{pol}\cos\left(2\hat{\phi}-2\phi\right)}{\sigma_v}\right)
\end{aligned}
$$

where Φ is the cumulative distribution function of the standard normal distribution.

A.5. Discussion

Although we built the dataset and validated the noise model as elaborately as possible, there are still several considerable tasks left for future work.

Assumptions in the noise model The polarization noise analysis model is induced from basic assumptions. In this paragraph, we outline the basic assumptions and explain why they are justified. We adopted the shot-and-read noise model. It modeled quantum properties of photons, which were described as a Poisson distribution, and the other noise sources as Gaussian noise. It is the basic model of image noise, and is adopted by noise image synthesis in denoising [2], low-light image enhancement [7], and burst super-resolution [1]. We assume that each observation is independent; this assumption is typically taken for granted in statistical analysis, and in an ideal sensor, each observation should be both spatially and temporally independent. Similarly, we assume an ideal polarizing filter, which means that circular polarization has no effect on linear polarization filters, the perpendicular component to the direction of the polarizing filter is completely blocked, and the transmittance of the parallel components is the same at any angle of the polarizing filter. With the above assumptions, most of our model is derived analytically, with the only additional approximation being that s_0 and $\hat{s}_0$ are equal for the DoLP model.

Noise comparison among datasets A single capture of the polarization image may not provide sufficient effective polarization information unless its noise level is sufficiently low. Therefore, we propose presenting the noise statistics data for

the first time. As a result, we were unable to compare our noise model and physical values in polarization with the previous dataset; instead, we only compared them using noise level estimation from a single image in the s_0 domain. However, this comparison is not an ideal way to assess the robustness of the polarization data against noise.

The model for single-image noise detection is not the shot and read noise model, but rather the additive white Gaussian noise model. This model creates a gap between the effectiveness of the data's polarization properties and the quality of the estimated noise level. The estimated noise level increases as the overall intensity of the images becomes larger because shot noise depends on the number of photons arriving at the sensor. However, the SNR of s_0 or s_{pol}, which are key parameters for the effectiveness of polarimetric properties, increases when the overall captured intensity is larger. Therefore, this metric might not distinguish between sophisticated captures of bright scenes and noisy scenes with underexposure. For example, the KAUST polarization image dataset has the highest estimated noise level, suggesting that the capture setup of the KAUST dataset could be the best among the comparison datasets. They captured images using a DSLR with a rotating polarizer, 2×2 pixel binning, and an integration of 100 burst images.

We provide noise statistics, equipment specifications, and parameter settings. Additionally, new comparison methods for evaluating the effectiveness of polarimetric data without provided noise statistics could also be considered for future work.

Limitation of the noise model We adopted a shot-and-read noise model with a Gaussian distribution as the base noise model in the sensor domain for deriving the noise distribution of the polarization properties. The assumption of Gaussian noise is very useful for derivation, and it yields quite good results in many cases. However, it can deviate from the model when the image has low intensity. First, the effect of quantization noise increases when the digital number becomes smaller. Second, the distribution of read noise does not strictly follow a Gaussian distribution. For instance, in Figure 4 of the main paper, the high SNR aligns well with the model, but the low SNR displays a spiky shape, despite the entire distribution adhering to our model. This might be the effect of quantization noise, which can limit the number of cases of the DoLP and AoLP in low-intensity data. A more sophisticated model for low intensity could be future work.

Scene variety We reduce noise using massive images in the temporal domain. This approach effectively and predictably minimizes noise. However, it limits scene variety due to the long acquisition time. For outdoor and indoor scene capture, we capture approximately 1,000 images over a period of around 2 minutes. During this capture time, the scene must remain motionless and free of illumination changes. Therefore, our dataset excludes objects affected by their own movement or wind, such as plants, animals, clouds, rivers, or lakes. Moreover, the illumination conditions must also be static, so we cannot capture images in rainy, snowy, cloudy, or foggy weather, nor during sunrise or sunset, even though these varying weather conditions may produce meaningful polarization effects from scattering or large incident angles. The acquisition method for a polarization noise-reduced image under various objects and illumination conditions remains future work.

Incoherency between burst images and GT As criticized by Wei et al. [8], burst SR datasets acquired with multiple devices simultaneously have incoherencies between different devices, as does the PolarBurstSR dataset. Burst SR datasets include two types of images in a sample: hand-held burst images and a GT image. These should have different focal lengths, so datasets select temporal multiplexing [3, 8] or multiple device capturing. Acquiring on multiple devices inherently causes misalignment from different views and color differences among devices [1]. Despite these problems, we captured images using two cameras with different lenses. Due to the susceptibility of the polarization properties to noise, we captured around 1000 images for GT using a tripod. Additionally, since the polarization cameras did not perform like commercial DSLR cameras, we cannot use zoom lenses for different focal lengths in a single optical system. Therefore, temporal multiplexing requires the mounting and unmounting of elements in the setup, which could undermine the benefits of single-device acquisition.

A.6. Synthetic Dataset Results

Both code and dataset are publicly available on https://github.com/KAIST-VCLAB/polarns.

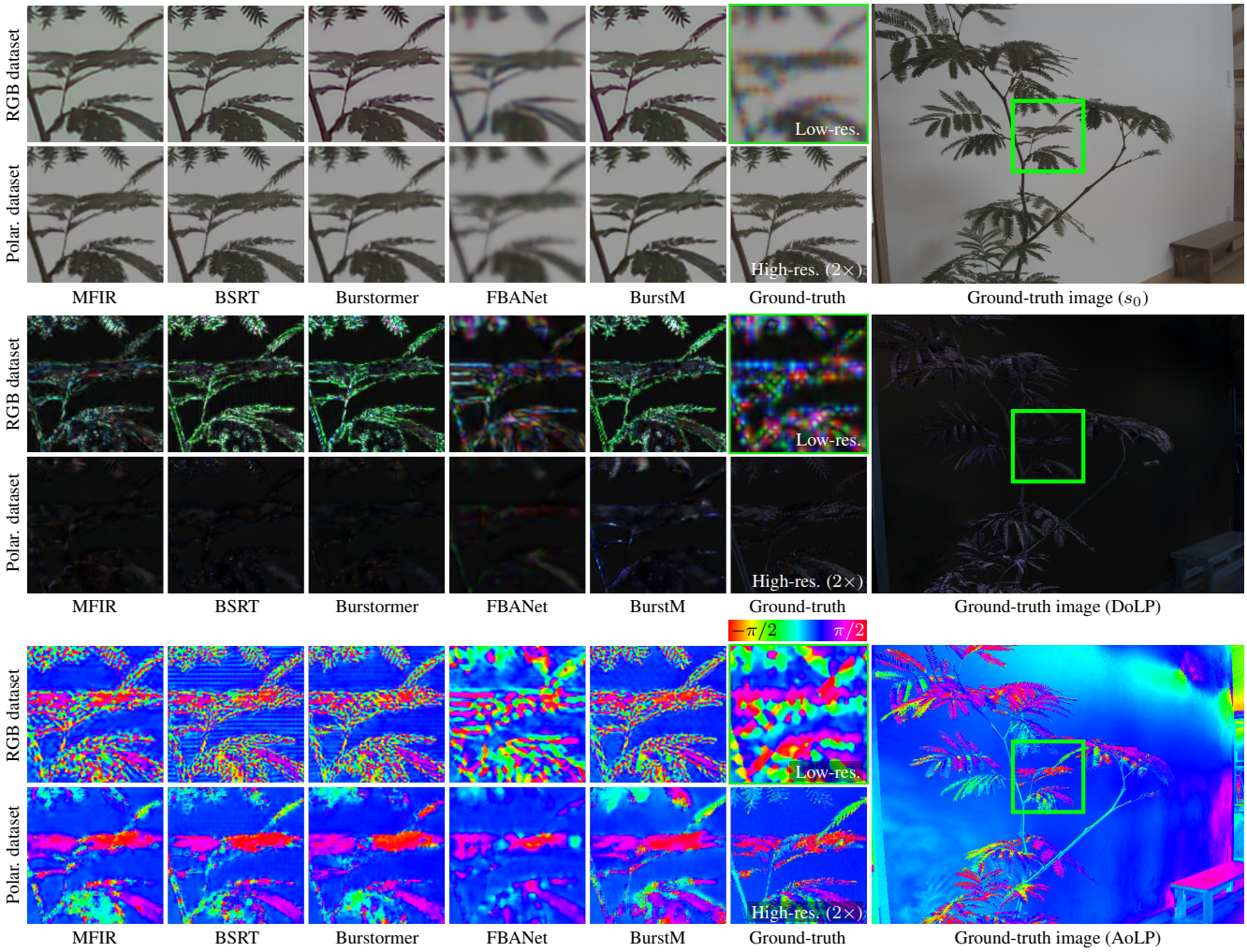

Figure 2. Qualitative comparisons on the polarization synthetic dataset.

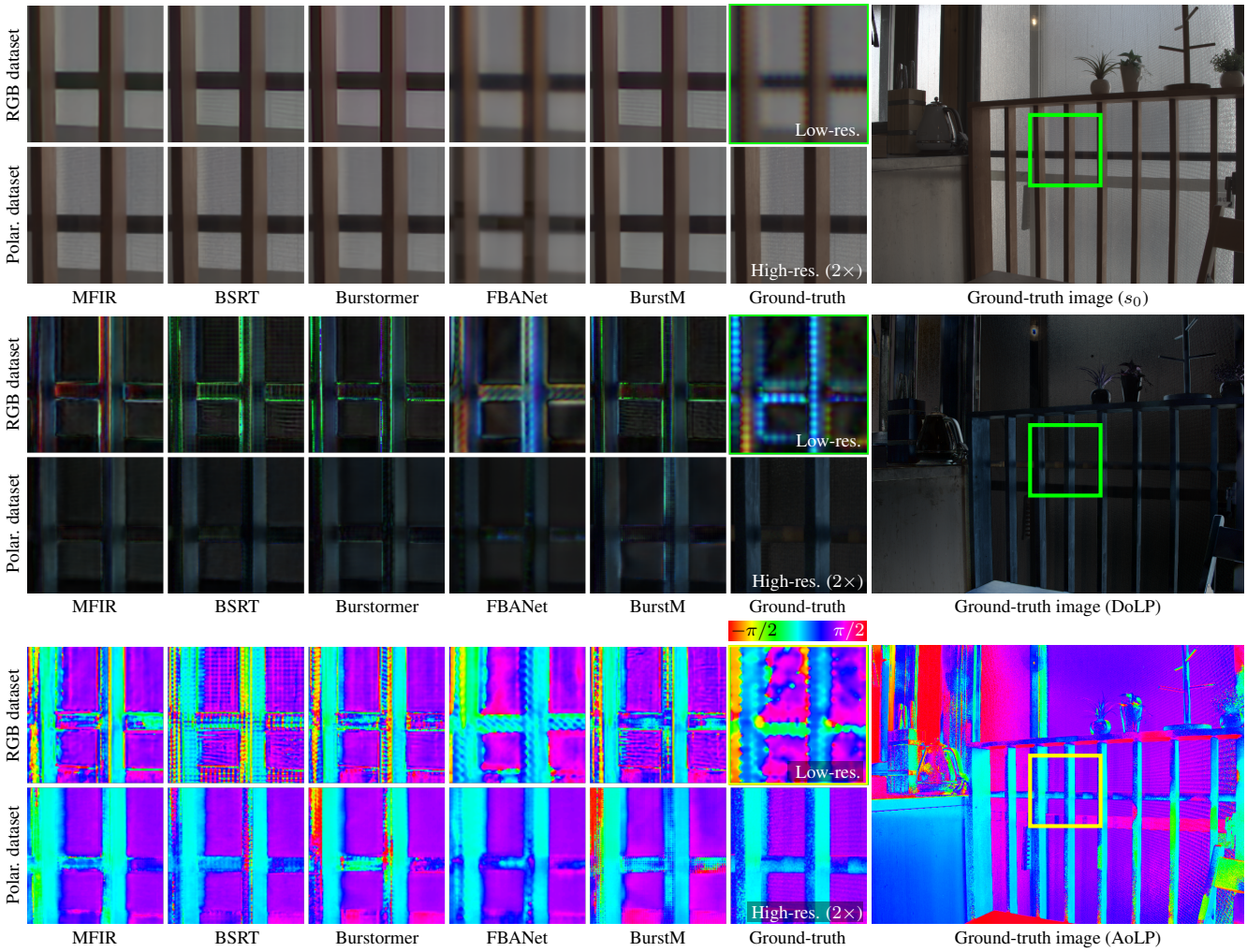

Figure 3. Qualitative comparisons on the polarization synthetic dataset.

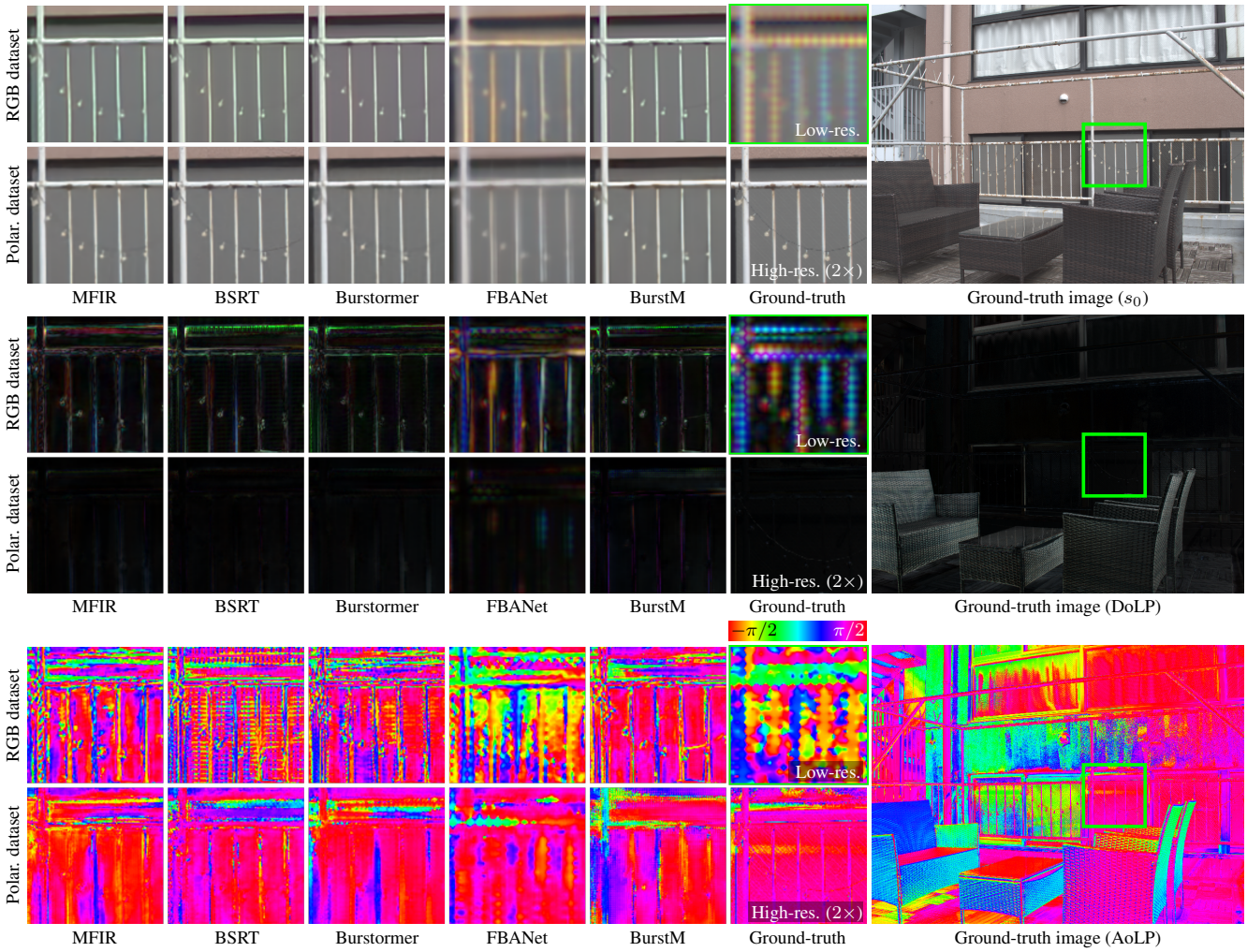

Figure 4. Qualitative comparisons on the polarization synthetic dataset.

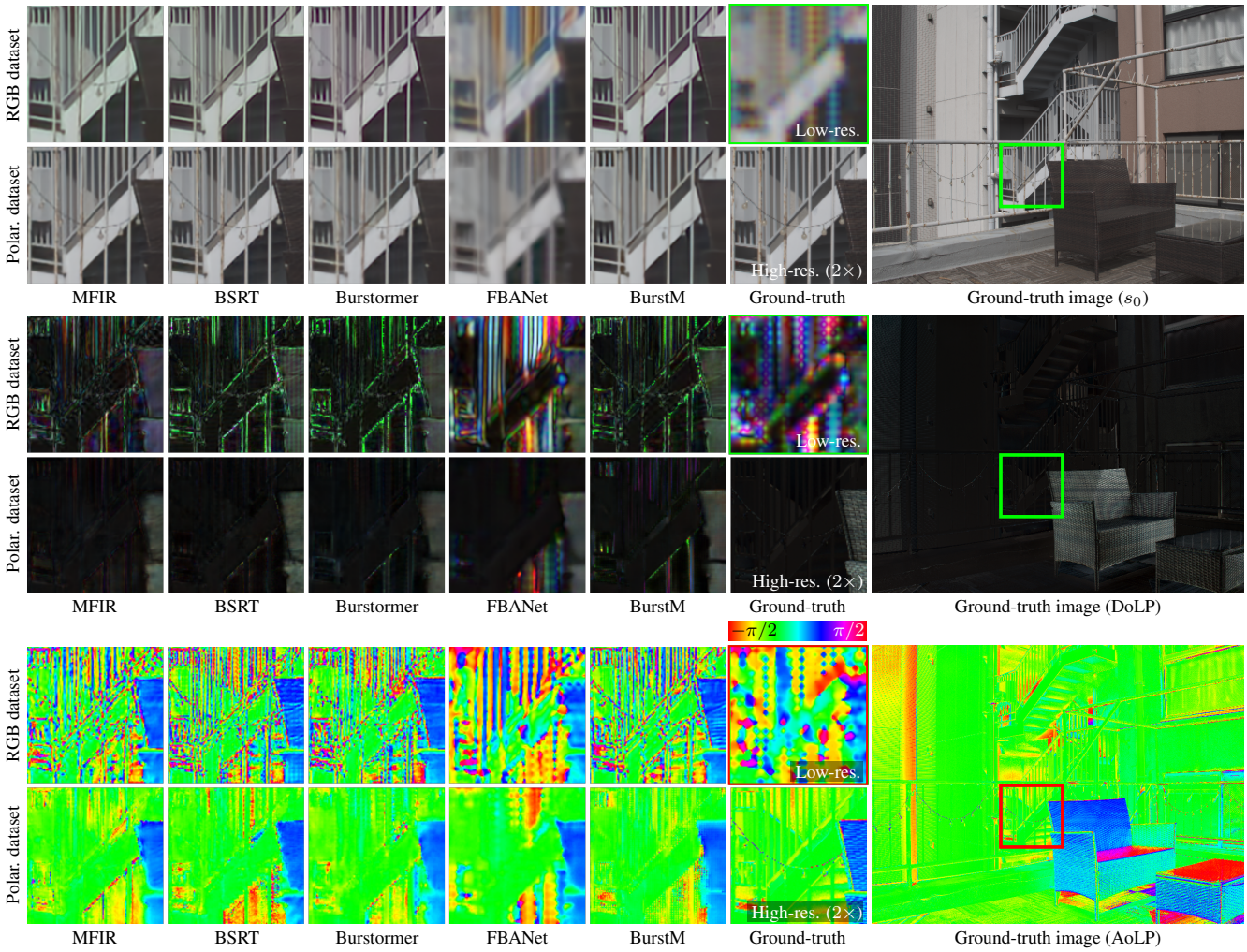

Figure 5. Qualitative comparisons on the polarization synthetic dataset.

References

[1] Goutam Bhat, Martin Danelljan, Luc Van Gool, and Radu Timofte. Deep burst super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9209–9218, 2021. 1, 3, 4

[2] Tim Brooks, Ben Mildenhall, Tianfan Xue, Jiawen Chen, Dillon Sharlet, and Jonathan T Barron. Unprocessing images for learned raw denoising. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 1, 3

[3] Jianrui Cai, Hui Zeng, Hongwei Yong, Zisheng Cao, and Lei Zhang. Toward real-world single image super-resolution: A new benchmark and a new model. In *Proceedings of the IEEE International Conference on Computer Vision*, 2019. 4

[4] Martin A. Fischler and Robert C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, 1981. 1

[5] Teppei Kurita, Yuhi Kondo, Legong Sun, and Yusuke Moriuchi. Simultaneous acquisition of high quality rgb image and polarization. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 178–188, 2023. 1

[6] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004. 1

[7] Feifan Lv, Yu Li, and Feng Lu. Attention guided low-light image enhancement with a large scale low-light simulation dataset. *International Journal of Computer Vision*, 129(7):2175–2193, 2021. 3

[8] Pengxu Wei, Yujing Sun, Xingbei Guo, Chang Liu, Guanbin Li, Jie Chen, Xiangyang Ji, and Liang Lin. Towards real-world burst image super-resolution: Benchmark and method. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 13233–13242, 2023. 1, 4