Pattern Recognition 161 (2025) 111314

Contents lists available at ScienceDirect

Pattern Recognition

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

Bayesian nights: Optimizing night photography rendering with Bayesian derivative-free methods

Simone Zini *, Marco Buzzelli

Department of Informatics Systems and Communication, University of Milano–Bicocca, Viale Sarca 336, Milan, 20126, Italy

ARTICLE INFO

Keywords:
Bayesian optimization
Low-light image enhancement
Image signal processing
Camera pipeline
Machine learning

ABSTRACT

We introduce a novel approach for optimizing Image Signal Processing (ISP) rendering pipelines for night photography through a Bayesian derivative-free procedure. Traditional neural-network-based ISPs depend on differentiable operations to enable backpropagation-based optimization, a requirement that can impose significant constraints. Our method circumvents this by employing Bayesian optimization to fine-tune the pipeline's parameters, independently of their differentiability. Additionally, we address the need for paired data to enable supervised optimization: while such paired data is available on public datasets, it is expensive to collect for new imaging devices. To this extent, we design a raw-to-raw mapping procedure, that aligns images from an available paired dataset to the target unpaired dataset. This allows us to supervise the optimization of our solution directly within the target space, without the need for device-specific paired data. We validate our approach with extensive experimentation on paired and unpaired datasets, demonstrating its efficacy using both subjective and objective evaluation metrics. Our code is made available for public download at https://github.com/TheZino/Bayesian-pipeline-optimization.

1. Introduction

Night photography rendering is a critical yet challenging domain in image processing. Its significance lies in the ability to capture vivid and clear images under low-light conditions, a scenario frequently encountered in various applications ranging from personal photography [1] to surveillance [2]. However, the complexity of this task arises from the inherent limitations of capturing high-quality images in the absence of adequate lighting, often resulting in color distortion, noise, and loss of detail.

Image Signal Processing (ISP) pipelines play a pivotal role in addressing these challenges. The concept of a pipeline refers to the sequence of processing steps that convert raw sensor data from a digital camera into a final image. This process typically includes several low-level functions, such as demosaicing, as well as higher-level operations designed to produce a visually pleasing image while correcting for various distortions inherent in the capture process. Among these, figures the adjustment (fixed or adaptive) of image brightness and contrast: this aspect is crucial, as it directly influences the perceptual quality of the image, balancing the illumination and enhancing the details without introducing artifacts.

Different vendors develop unique ISP pipelines [3], each with a set of parameters fine-tuned to specific imaging conditions. This diversity highlights the need of optimizing these pipelines for night or low-light photography, ensuring that images captured under these conditions are rendered appropriately. In this paper we design a procedure for the optimization of ISP pipelines, i.e. for the search of optimal values in the free parameters of existing pipelines. We design our solution to be based on derivative-free Bayesian optimization. This is a significant departure from deep learning methods: although proven effective for many tasks related to image processing and computer vision, they pose a limiting requirement of differentiability of the pipeline steps, so that these can be optimized via gradient backpropagation. By relaxing this constraint, our approach offers greater flexibility and adaptability in optimizing ISPs.

Many studies in scientific literature depend on paired data, which involves using a reference "high light" rendered image to train supervised models. However, this dependency can limit the transferability of the developed rendering pipelines to different devices, as each device may require its own unique set of paired data for effective training, which is expensive to collect. To this extent, we introduce an innovative approach that shifts the need for paired data from the target camera space to an external source, using a dedicated raw-to-raw mapping procedure.

The primary problem addressed in this paper is the challenge of optimizing ISP pipelines for night photography, where low-light conditions result in raw images with poor value distribution, color distortions, and noise. Current deep learning-based methods impose constraints such as the need for differentiable operations (which limit

* Corresponding author.
E-mail addresses: simone.zini@unimib.it (S. Zini), marco.buzzelli@unimib.it (M. Buzzelli).

https://doi.org/10.1016/j.patcog.2024.111314
Received 4 July 2024; Received in revised form 18 November 2024; Accepted 23 December 2024
Available online 30 December 2024

the set of possible algorithms that may be integrated in the pipeline) and device-specific paired data for supervised optimization (which are costly and challenging to obtain for new imaging devices). Our leading objective is therefore to design a flexible, derivative-free Bayesian optimization procedure to fine-tune ISP parameters for night photography without relying on paired data specific to each device. This method enables the optimization of camera pipelines in low-light conditions by employing a raw-to-raw mapping procedure, making it adaptable to different camera sensors and reducing the need for manual intervention or device-specific data collection.

The main contributions of this paper are:

1. Introducing a Bayesian derivative-free method to optimize ISP parameters for night photography, suitable for various camera setups.
2. Employing a raw-to-raw conversion using external paired datasets, enabling effective ISP optimization.
3. Demonstrating the potential to enhance existing camera ISPs using external data.

2. Related works

Several studies have adopted the approach of using paired data for night image processing [4,5]. However, acquiring such data presents a significant cost, and consequently it reduces the applicability of night rendering methods to new imaging devices. An alternative strategy to the acquisition of paired data involves synthesizing input data from a clean source combined with a detailed analysis of the distortions that are characteristic of night imaging setups. For instance, Wei et al. [6] create realistic noisy raw data that accurately reflects the physical processes behind noise generation. While innovative, this method requires comprehensive modeling of various aspects of the imaging pipeline, focusing primarily on noise models, which may limit its applicability to broader ISP optimizations.

Our work reimagines this modelization approach by exploiting existing paired data from a third party, adapting its raw images to match the device-specific raw space of our unpaired domain of interest. This adaptation is inspired by raw-to-raw mapping techniques, which are, however traditionally designed to also work with paired data from the two raw spaces [7,8]. Our methodology, as further detailed in Section 3, circumvents the limitations of relying on paired datasets for the mapping step itself. In recent advancements, Shi et al. [9] proposed Zero-IG, a novel zero-shot method for low-light image enhancement and denoising that does not require paired training data, outperforming previous methods by effectively integrating illumination guidance in the enhancement process. Wang et al. [10] introduced QuadPrior: a zero-reference low-light enhancement framework, utilizing physical quadruple priors for illumination-invariant processing, which bypasses the need for paired low-light training data, marking a significant step in unsupervised low-light enhancement.

Zhu et al. [11] and Park et al. [12] exploited generative models to perform general-purpose domain adaptation. These models are found to be capable of simulating complex scenarios and transformations, at the cost of introducing local artifacts [13]. In the specific domain of low light imaging, Zamir et al. [14] addressed this issue by developing loss functions that assess the perceptual quality of images, using them to train a CNN encoder–decoder (that also necessitates paired data). However, these handcrafted loss functions may overlook specific aspects related to image quality, potentially leading to new unforeseen artifacts at inference time. Given these challenges, exploring simpler pipelines that minimize the risk of introducing artifacts while still enhancing image quality in low-light conditions is a promising direction.

The NTIRE night imaging challenge [15,16], initiated in 2022 and held yearly, has been a significant catalyst for research in low light and night image rendering. This competition has spurred advancements by encouraging the development of innovative solutions for low light image enhancement.

Most notably, Li et al. [17] introduced CBUnet: a neural image processing pipeline that involves two U-net-type sub-architectures to correct for color and brightness, respectively. The authors resorted to manual correction of images from the challenge training set in order to construct a paired dataset for the training of their own contribution to night imaging. While effective, this manual approach limits scalability to other datasets or imaging devices, suggesting a need for more adaptable solutions.

Similarly, Liu et al. [18] presented a method for night imaging based on the cascaded application of three independent modules, each trained on its own set of paired data: a U-Net denoising module trained on a custom dataset, the popular FC4 white balancing module [19] trained on the ColorChecker dataset [20], and MWISP-inspired module [21] for tone mapping, trained on manually-corrected images from the challenge dataset. The separation of tasks into independent subtasks reduces the reliance on sensor-specific data, though paired data remain essential.

Desai et al. [22] introduced LightNet, a multiscale encoder–decoder network trained on a dedicated external dataset of paired images, acquired across various ISO settings. This approach aimed to account for different exposure levels during inference, but did not address cross-sensor variability, which limited its effectiveness. This highlights the importance of explicitly addressing the challenge of raw-to-raw sensor mapping to enhance performance across different camera sensors.

In the 2023 edition of the night rendering challenge, Zini et al. [23] presented “Back to the future”, a night photography rendering ISP that makes no use of deep learning, yet won first place in people’s choice track and third place in photographer’s choice track. Building on top of the previous year’s iteration [24], the authors’ solution implements a “traditional” ISP pipeline to render visually pleasing photographs of night scenes, characterized by a shallow structure, explainable steps, and a low parameter count, resulting in computationally efficient processing. This achievement underlines the potential of lightweight pipelines to deliver high-quality night scene renderings, balancing proper tone mapping with minimization of visual artifacts. Although no explicit training was required by this approach, the authors manually tuned the free parameters of their solution to match a desired visual aesthetic. This effectively means fitting for the target test environment with manual labor, thus reducing the scalability of the approach. This limitation prompts the exploration of a learning-based method to determine optimal parameters using an independent dataset, aiming for a process that eliminates the need for manual intervention when adapting to the final application domain, thus enhancing the method’s applicability and efficiency.

In Table 1 we reported each of the methods that are directly comparable with the proposed approach, along with a brief description. We also report the corresponding pros and cons, which motivate our current proposal as described in the text.

3. Proposed method

In this section we describe our approach to night image rendering, with a focus on the handling of parameters optimization on an unpaired dataset. Based on this requirement, our solution is composed of:

1. A raw-to-raw mapping step, designed to delegate the need for paired data to an external source independent from the device whose pipeline is being optimized.
2. A Bayesian-based optimization of the specific parameters that characterize the pipeline at hand.

Let D_t be a dataset of low light images L_t:

$$D_t = \{L_t\}. \tag{1}$$

Table 1
Description of each state-of-the-art methods comparable with the proposed approach.

Method (year)	Approach	Pros	Cons
CBUnet [17] (2022)	Use of two U-net sub-architectures for separate color and brightness correction of images.	Cascade architecture for separate processing of components. Presents a custom dataset.	Necessity of paired data obtained by manual editing.
Deep-flexisp [18] (2022)	Cascade application of three modules based on Deep Learning: U-Net (Denoise), FC4 (Color balancing), MWISP-inspired network (tone mapping).	Flexible with independent modules. Sensor independent.	Necessity of manually-corrected images.
LightNet [22] (2023)	LightNet, a multiscale encoder–decoder network trained in a GAN framework.	Handles multiple exposure levels simultaneously. Hierarchical encoder–decoder effective for processing fine details. Presents a customized data set.	Custom collected dataset required for training. Computational complexity high due to hierarchical generator.
BttF [23] (2023)	Night image processing pipeline with low number of parameters. NTIRE 2023 winner.	Shallow architecture and low number of parameters. Baseline for night image rendering. Deep-learning free approach	Optimized manually with human experience

Low-light images in this dataset are encoded in a sensor-specific "target" raw space. Our goal is to produce high-light versions H_t in a standard space (e.g. sRGB), through the optimization of an image signal processing pipeline $p_{\theta_{opt}}(\cdot)$ that depends from parameters θ:

$$H_t = p_{\theta_{opt}}\left(L_t\right). \tag{2}$$

No reference high-light version is available for direct optimization on this dataset, therefore we aim to optimize θ on a second "source" dataset D_s, where low-light/high-light paired data is available:

$$D_s = \{L_s, H_s\}. \tag{3}$$

Here, L_s is encoded in a sensor-specific raw space different from that of L_t, while H_s determines the standard space that will be used to also produce H_t.

Due to the different low-light raw spaces, an optimization of parameters θ on dataset D_s would yield suboptimal results on the target dataset D_t. For this reason, it is necessary to first map source low-light images L_s into the target raw space via mapping function $m_{\gamma_{opt}}(\cdot)$, depending on parameters γ:

$$L_m = m_{\gamma_{opt}}(L_s). \tag{4}$$

The following subsections deal, respectively, with: defining the image signal processing pipeline p, optimizing the mapping function m_γ, and optimizing the pipeline parameters θ.

3.1. Night image rendering pipeline

In order to optimize an already established pipeline, and to have a baseline approach to start with, we select "Back to the future": the night photography rendering pipeline proposed by Zini et al. [23] that won the NTIRE 2023 Night Photography Rendering challenge.

Despite our specific choice, the optimization techniques described in the following are general-purpose, and may be applied to other ISP pipelines.

The current pipeline, represented in Fig. 1, processes the images starting from the raw data from the camera sensor, with metadata associated, giving final 8-bit processed images.

Since the total number of parameters of this pipeline is very low (specifically 14), and since certain steps of the pipeline are not differentiable functions (e.g. Non-local Means denoising step [25]), Bayesian optimization represents the best solution for the optimization of these parameters, as developed later on in Section 3.3.

Table 2 reports the total amount of parameters and the details regarding each step of the pipeline, while Fig. 1 depicts the entire pipeline.

We can define three main parts in the pipeline: a set of preliminary camera-specific operations, a group of common fixed operations, and a final group of optimizable functions.

As shown in Fig. 2, in the first group (*a.1* in Fig. 1) two processing steps are performed on the input raw data L. These *normalization* and *demosaicing* operations are dependent on the metadata coming from the camera and so are defined as camera-specific.

After this first preliminary group of operations the mapping function $m_{\gamma opt}(\cdot)$ on the input is applied in order to obtain their counterpart L_m in the target camera space. Once the mapping has been applied, the images pass through two fixed preliminary steps (*a.2* in Fig. 1): the Gray World AWB and the conversion from the camera to sRGB space. As shown in Fig. 2, the mapping function $m_{\gamma opt}(\cdot)$ is used only during the optimization phase, since the dataset used is the D_s one which needs to be transformed into the target dataset space. During the actual rendering phase the camera-specific and the preliminary steps are performed directly one after the other. The color space conversion is performed in both cases using the camera color matrix of the target dataset D_t.

Subsequently, the images are processed by the rest of the pipeline modules, of which parameters are optimized by the Bayesian optimization procedure.

3.2. Optimization of raw-to-raw space mapping

Our objective here is to develop a mapping function m that transforms low-light source images L_s into a format that matches properties of a target raw space, represented by low-light images L_t. This transformation is aimed not at visual aesthetics, i.e. it is not used to directly perform the final rendering. Rather, it is aimed at aligning the data distribution for the subsequent optimization process, which assumes data with similar distributions. We formulate this mapping as a per-channel non linear adjustment curve, and we motivate this choice in the following.

The core challenge in raw-to-raw mapping lies in the different sensitivities of different camera sensors: these influence the linearity of data distribution within channels, as well as the interplay of colors across them. The scientific literature indicates that simple linear or polynomial transformations are not sufficient to model the complex, non-linear nature of these differences [7], requiring instead more complex transformations.

In terms of data, existing methods typically leverage a set of paired images to optimize the mapping between the two devices [8]. In our configuration, however, no paired data can be assumed to be available. This impacts the optimization strategy at multiple levels. First of all, we are forced to compare the mapped data L_m and the target data L_t

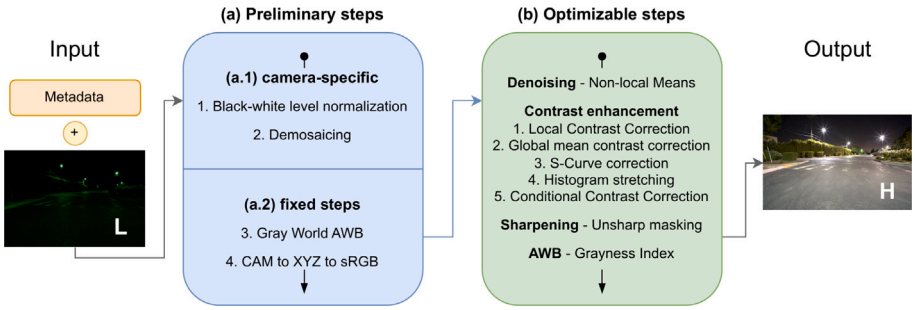

Fig. 1. Overview of the baseline pipeline for night photography rendering by Zini et al. [23]. The pipeline can be subdivided into three main components: Preliminary steps (camera-specific), Preliminary steps (fixed), and Optimizable steps.

Table 2
Default values of each parameter of the baseline pipeline. For each step of are here reported the number of parameters, their role, and the default values taken from the original paper by Zini et al. [23].

Processing step	# par.	Default parameter values
Non-local means denoising	2	Luminance channel weight: 4.5 Chrominance channel weight: 20
Local Contrast Correction	1	Gaussian blur deviation σ: 7.24
Global Mean Contrast	1	Value stretch factor β: 1.5
S-Curve	2	Curve node α: 0 Curve divergence λ: 0.556
Histogram stretch	2	Lower percentile: 0.0001 Higher percentile: 0.9999
Conditional Contrast Correction	3	First S-curve λ_1: 0.556 Second S-curve λ_2: 0.714 Gamma correction γ: 2.2
Sharpening	2	Gaussian blur deviation σ: 2 Scale factor: 1
Grayness Index White Balancing	2	Best grey pixel proportion N: 0.1 Threshold: 0.0001
Total number of parameters	15	

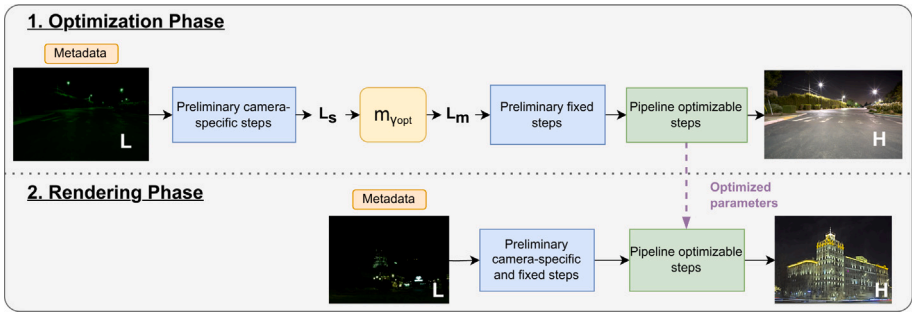

Fig. 2. Our optimization and rendering phases. In the optimization phase the raw images from the source dataset are processed with the camera-specific steps, mapped to the target camera color space and then used for parameter optimization. In the rendering phase, the optimized parameters are directly used on the images of the target dataset.

via global statistics such as histogram comparison, as opposed to pixel-wise operations. This, in turn, poses a serious constraint on the type of raw-to-raw transformation to be implemented: a generative and/or an encoder–decoder model [11,12] would have the freedom to introduce local artifacts that would not be detected by a global metric. It follows that a stronger control on the output of the mapping is required.

Aiming to strike a balance in terms of mapping complexity, and guided by the requirement of global comparison, we identify color curves as the basis for our solution for raw-to-raw mapping. The procedure depends on specific architectural and implementative choices, as delineated in the following.

1. A training set of target images is first isolated to ensure fairness in the subsequent procedure and its evaluation:

$$L_{t_{\mathrm{TR}}} \subset L_t \tag{5}$$

2. Images from the source set, and from the target training set, are downscaled with anti-aliased sampling:

$$\{\downarrow L_s, \downarrow L_{t_{\mathrm{TR}}}\} \tag{6}$$

This operation has two main effects: it leads to a drastic reduction in the computational requirements, and it filters out noise during optimization.

3. Color histograms are computed for all images from both sets, collecting Red Green and Blue information independently:

$$hist\,(I[\mathrm{ch}])\ \forall I \in \{\downarrow L_s, \downarrow L_{t_{\mathrm{TR}}}\}\ \forall \mathrm{ch} \in \{R, G, B\} \tag{7}$$

Working in RGB allows to adjust the relative color distribution, which is essential for accurately simulating the target raw space's characteristics. Conversely, histogram matching on a luminance-like channel would fail the objective of redistributing the colors.

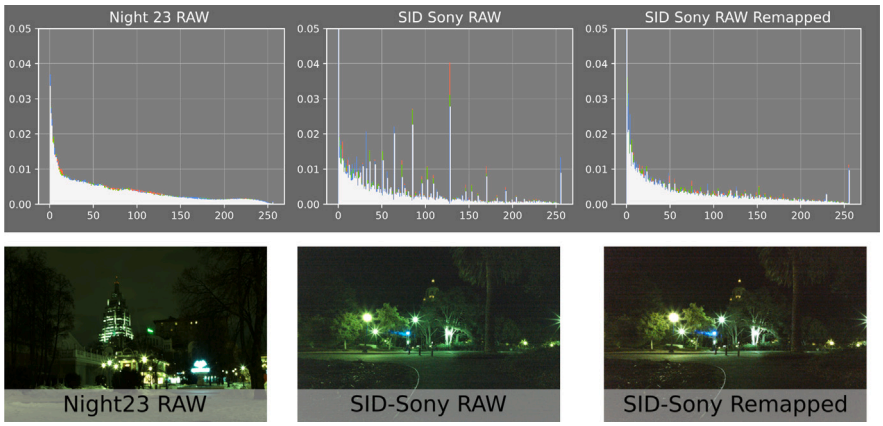

Fig. 3. Comparison between color histograms of the Night23 raw data, SID-Sony raw data and SID-Sony raw data after applying the mapping function $m_{\gamma_{opt}}(\cdot)$. The raw images are processed with a fixed gamma correction for visualization purposes.

4. For each channel, one global histogram is aggregated from the histograms of all source images, and one from the histograms of all training target images:

$$g_{s,\text{ch}} = \sum_{I \in \downarrow I} hist\,(I[\text{ch}]) \tag{8}$$

$$g_{t_{\text{TR}},\text{ch}} = \sum_{I \in \downarrow L_{t_{\text{TR}}}} hist\,(I[\text{ch}]) \tag{9}$$

5. Piecewise Cubic Hermite Interpolating Polynomial (PCHIP) [26] is used to create a mapping function that aligns $g_{L_s,\text{ch}}$ and $g_{L_{t_{\text{TR}}},\text{ch}}$ for each color channel independently.
   In other words, we do not match each source image to a given target image, instead we define three global matching functions (different for each channel). This ensures that the transformation is consistent across all images, preserving the inherent differences between images, such as variations between blue-light and yellow-light scenes.

A visual representation of the effect of raw-to-raw space mapping is presented in Fig. 3: global RGB histograms are shown for a target dataset Night23, for a source dataset SID Sony, and for its remapped version. Sample images from the datasets are also visualized.

3.3. Optimization of image signal processing pipeline

Algorithm 1 Bayesian optimization of the night rendering pipeline.
Θ: hyper parameter space.
$n_{iterations}$: number of total optimization iterations.

```
Select initial configuration θ_0 ∈ Θ
Evaluate initial score y_0 = ℓ(θ_0)
Set θ* ← θ_0 and y* ← y_0
for n = 1, ..., n_iterations do
    sample new θ ∈ Θ using the acquisition function
    evaluate y = ℓ(θ)
    update surrogate model π(y|θ)
    if y > y* then
        Set θ* ← θ and y* ← y
    end if
end for
output θ*, y*
```

Let ℓ be a function that computes the distance of two *paired* image sets, where each image in the first set has a direct correspondence in the second set. Our goal is to minimize distance ℓ between the source low-light images processed with pipeline p, and their high-light counterpart:

$$\theta_{opt} = \underset{\theta}{\text{argmin}} \left(\ell \left(p_\theta(L_m), H_s \right) \right). \tag{10}$$

Here, p is a previously defined image processing pipeline and θ is the set of free parameters to be optimized. We approach the search for the minimum in ℓ by means of Bayesian optimization [27]. This approach considers ℓ as a black-box function: the evaluation of ℓ is performed by only observing its application to parameters θ, without the necessity of computing first- or second-order derivatives. This particular condition lets us optimize a pipeline containing any kind of algorithm, without the restriction of using only differentiable functions, a characteristic that suits well the use case we are considering, where several operations, like histogram-based operators or denoising algorithms, are in general non-differentiable steps.

The typical form of Bayesian optimization algorithms involves two primary components:

1. A Bayesian statistical model for modeling the probability of improvement, defined as $\pi(y|\theta)$, where y represents an observation of the objective function ℓ. It is usually modeled with Gaussian Processes.
2. An acquisition function for deciding where to sample next. This function measures the value that would be generated by evaluation of the objective function at a new point θ, based on the current posterior distribution modeled over ℓ.

We reported the pseudo-code of the Bayesian optimization procedure in Algorithm 1.

In our configuration we adopt the Tree-structured Parzen Estimator (TPE) algorithm [28], a variant of BO that adopt an acquisition function similar in behavior to the probability of improvement, and exploits Kernel Density Estimators (KDEs) and tree-structured search spaces. Whereas the Gaussian-process based approach models the $\pi(y|\theta)$ directly, this strategy models $\pi(\theta|y)$ and $\pi(y)$. At each iteration, for each parameter, TPE fits one Gaussian Mixture Model (GMM) $l(\theta)$ to the set of parameter values associated with the best objective values, and another GMM $g(\theta)$ to the remaining parameter values. Two such densities define $\pi(\theta|y)$ as:

$$\pi(\theta|y) = \begin{cases} l(\theta) & if\ y < y* \\ g(\theta) & if\ y \geq y*, \end{cases} \tag{11}$$

where $l(\theta)$ is the density formed by using the observations θ_i such that corresponding loss ℓ is less than $y*$, and $g(\theta)$ is the density formed by using the remaining observations. The TPE algorithm chooses $y*$ to be some quantile q of the observed y values, so that $\pi(y < y*) = q$, without the necessity of specifying a specific model for $\pi(y)$. TPE chooses the parameter value θ that maximizes the ratio $\frac{l(\theta)}{g(\theta)}$.

We adopt this optimization strategy to optimize a image rendering pipeline with a limited number of parameters, without any restriction on algorithms used in such pipeline.

Table 3
Ranges of the search spaces of each parameter, for each step of the rendering pipeline. Each parameter has been sampled as a floating point value 32-bit precision. The values at the borders are included in the search space.

Processing step	Default parameter values	Search space range
Non-local means denoising	Luminance channel weight: 4.5	[0, 30]
	Chrominance channel weight: 20	[0, 30]
Local Contrast Correction	Gaussian blur deviation σ: 7.24	[0, 12]
Global Mean Contrast	Value stretch factor β: 1.5	[0.8, 2.0]
S-Curve	Curve node α: 0	[0, 1]
	Curve divergence λ: 0.556	[0.1, 10]
Histogram stretch	Lower percentile: 0.0001	[0, 0.5]
	Higher percentile: 0.9999	[0.5 , 1]
Conditional Contrast Correction	First S-curve λ_1: 0.556	[0, 1]
	Second S-curve λ_2: 0.714	[0, 1]
	Gamma correction γ: 2.2	[1, 10]
Sharpening	Gaussian blur deviation σ: 2	[1, 3]
	Scale factor: 1	[0, 2]
Grayness Index White Balancing	Best gray pixel proportion N: 0.1	[0, 1]
	Threshold: 0.0001	$[1e^{-6}, 1e^{-2}]$

The distance function ℓ, adopted for the optimization process is the negative of Peak Signal-to-Noise Ratio (PSNR) [29]. PSNR is computed on the three channel H_t output image and the target H_s image and is defined as:

$$\text{PSNR} = 20 \cdot log_{10}\left(\frac{1}{\text{MSE}(H_t, H_s)}\right) \tag{12}$$

where MSE is the Mean Squared Error between the two images H_t and H_s.

Due to the fact that the PSNR value is close to zero for images which are completely different and goes to infinity for images equal to each other, we formalize the minimization problem by targeting the negative of the PSNR.

4. Experiments

The night rendering pipeline has been developed in Python 3.8, and the Bayesian optimization has been implemented with the Optuna library, version 2.10.1. The raw-to-raw space mapping has been implemented in MATLAB R2022b.

In the following, we will describe the datasets used for the optimization and the testing of the pipeline, as well as the configuration of the optimization process.

4.1. Dataset

In order to optimize the pipeline defined in Section 3.1 with the proposed optimization approach, we require a third-party dataset that includes raw data and reference ground truth (images processed for high-lighting). We consequently selected the "Learning to See In the Dark" dataset (SID) from Chen et al. [30], which is a collection of pictures taken in low light conditions (mainly night scenes and indoor scenes) with two different cameras: a Sony α7S and a Fujifilm X-T2. For each image of the two cameras, the dataset offers a single long-exposure processed image, which is considered the ground truth, and multiple short-exposure raw data images in the camera's proprietary format. Each input raw image keeps the original camera metadata in EXIF format, while explicit information on exposure time is encoded in the file names. We specifically used the data from the Sony camera as our source dataset D_s, to perform the optimization procedure. This part of the SID dataset contains 8091 image pairs.

For the validation and testing of the optimization procedure, we adopted the dataset from the NTIRE Night Photography Rendering Challenges 2022 [15] (Night22) and 2023 editions [16] (Night23); both the datasets are made of around 400 images (482 the 2022 edition and 400 the 2023 one), in raw 16-bit format with metadata associated. Since the images have been acquired with the same camera model, we can effectively exploit the two datasets as training and test sets respectively. More specifically: the Night22 images have been used to compute the mapping function $m_{\gamma opt}(\cdot)$, while Night23 images have been used for the actual evaluation of the optimized pipeline. Due to the nature of the challenges, reference ground truth images are not provided, so evaluation has been performed on the basis of psycho-visual tests. Datasets that do not provide raw data alongside the reference processed images (e.g., the LOL Low-Light dataset [31]) were excluded from the training and evaluation of our optimization approach. This is because our method specifically targets raw camera data to optimize the image signal processing pipeline. Additionally, the LOL dataset consists of images captured in daylight conditions but under incorrect exposure settings to simulate low-light scenarios, rather than actual night landscapes. As a result, this dataset does not align with the real-world conditions we aim to address, rendering it unsuitable for our task.

4.2. Optimization setup

The Bayesian optimization process requires defining a search range for each parameter to be investigated. In Table 3 are reported, for each parameter, the corresponding ranges of each search space. Each range has been manually defined, based on previous knowledge of the semantics of each parameter, in order to have spaces large enough for and exhaustive exploration during the optimization procedure, while preventing combinations of parameters that may bring undesired wrong behaviors (e.g. gamma values lower or equal to 0, etc...). All of the parameters have been sampled as float values in the reported ranges $[i,\ j]$, where i and j are included in the set of possible values. To optimize those parameters a *thumbnail* version of the SID dataset has been used; each image of the Sony-SID dataset has been reduced at $\frac{1}{16}^{th}$ of the original dimension in order to reduce the computational costs. The parameters dependent on the image resolution have been optimized during the optimization phase but set back to the default values during the inference/testing phase.

The optimization has been done for a total of 400 trials, adopting the multivariate TPE sampler [32] with a number of candidates for the expected improvement computation set to 64.

4.3. Evaluation

In the following we describe the metrics and strategies used to evaluate the effectiveness of the proposed ISP optimization pipeline. We cover three main aspects: distributional similarity, image quality assessment in paired settings, and subjective quality evaluation in unpaired settings.

4.3.1. Kullback–Leibler divergence for distribution analysis

To assess the effectiveness of our raw-to-raw mapping in reducing the discrepancy between the source (SID) and target (Night23) image set distributions, we utilize the Kullback–Leibler (KL) divergence [33]. The KL divergence quantifies how one probability distribution Q diverges from a second, expected probability distribution P:

$$\mathrm{KL}(P \parallel Q) = \sum_i P(i) \log \left(\frac{P(i)}{Q(i)} \right), \tag{13}$$

where $P(i)$ and $Q(i)$ are the probabilities of the ith element in distributions P and Q, respectively.

4.3.2. SSIM and PSNR for image quality

To evaluate the image quality after applying the optimized ISP pipeline in a controlled, paired data setup (i.e. within the SID dataset), we use the Structural Similarity Index (SSIM) [34] and the Peak Signal-to-Noise Ratio (PSNR) [29]. SSIM provides a measure of similarity between two images, capturing texture and structural information:

$$\mathrm{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{14}$$

where μ_x, μ_y are the averages, σ_x, σ_y are the variances, and σ_{xy} is the covariance of images x and y. PSNR is defined as:

$$\mathrm{PSNR} = 20 \cdot \log_{10} \left(\frac{MAX_I}{\sqrt{MSE}} \right) \tag{15}$$

where MAX_I is the maximum possible pixel value of the image, and MSE is the mean squared error between the optimized and reference images.

4.3.3. Psychovisual study for unpaired setup

To evaluate the perceptual quality of the optimized pipeline in an unpaired setup (i.e. on the Night23 dataset), where no reference images exist, we conduct a psychovisual study. This involves subjective assessments from a group of ten observers who rate the visual quality of images processed by our pipeline, and by other existing solutions from the state of the art.

4.3.4. Quantitative evaluation for unpaired setup

In addition to the psychovisual study for evaluating the perceptual quality in an unpaired setup, we implement a quantifiable assessment in the Unpaired setup of the Night 2023 dataset.

The Light Order Error (LOE) [35] measures the accuracy of lightness order between the processed image and a reference (input raw), assessing how well the processing preserves the natural luminance ranking among different areas within an image:

$$LOE = \frac{1}{N} \sum_{i,j} |\mathrm{sign}(H(i) - H(j)) - \mathrm{sign}(L(i) - L(j))|, \tag{16}$$

where L and H are, respectively, the original and processed images, indexed with i and j over the total number of pixels N.

We also consider no-reference image quality assessments using BRISQUE [36], NIQE [37], and PIQE [38]. These metrics evaluate the naturalness and perceptual quality of images without the need for a reference image:

- BRISQUE (Blind/Referenceless Image Spatial Quality Evaluator) quantifies possible losses in naturalness due to processing.
- NIQE (Natural Image Quality Evaluator) uses a statistical model of natural image features to evaluate quality.
- PIQE (Perception-based Image Quality Evaluator) provides scores based on perceptual aspects of image quality, such as texture and noise.

Table 4
Kullback–Leibler divergence between SID-Sony images and Night23 images (top), and between SID-Fuji images and Night23 images (bottom). We report the effect without and with raw-to-raw space mapping on all three color channels (RGB), as well as the average of the three color channels (μ). The lower, the better.

SID-Sony	KL divergence			
	R	G	B	μ
No raw-to-raw mapping	0.6747	0.5209	0.8079	0.6678
Raw-to-raw mapping	0.2304	0.1531	0.2722	0.2186
SID-Fuji	KL divergence			
	R	G	B	μ
No raw-to-raw mapping	0.6362	0.4566	0.9393	0.6774
Raw-to-raw mapping	0.2910	0.2006	0.5919	0.3612

Table 5
PSNR and SSIM values computed on the processed images of the SID-Sony and SID-Fuji datasets, using different versions of the pipeline (default and optimized parameters). The higher, the better.

SID-Sony		
	PSNR	SSIM
Default parameters	16.9656	0.7041
Bayes optimization (from random values)	18.2684	0.7286
Bayes optimization (from default values)	**18.3734**	**0.7680**
SID-Fuji		
	PSNR	SSIM
Default parameters	15.8602	0.7405
Bayes optimization (from random values)	**16.8554**	**0.7792**
Bayes optimization (from default values)	16.4109	0.7663

4.4. Experimental results

In this section we report the effect of the mapping function on the color distribution of the images of the different datasets, and the impact of the optimization on the SID dataset and the NTIRE photography rendering challenge datasets. Here a visual comparison of the results of the state-of-the-art approaches and the proposed one is reported, alongside the results of the psycho-visual test using the Night23 dataset.

In order to quantify the effectiveness of our raw-to-raw space mapping, we compute the Kullback–Leibler divergence for several versions of the SID image dataset against the Night23 test set. The results are reported in Table 4, concerning both the Sony and Fuji cameras, without and with raw-to-raw mapping. Since our mapping takes place on the three RGB channels independently, as motivated in Section X, we report the resulting divergences on all channels, as well as an aggregate average measure (μ). It can be observed that the Kullback–Leibler divergence drops significantly thanks to our mapping, specifically by 67% on Sony and by 47% on Fuji.

Fig. 3 offers a qualitative reference for the effect of raw-to-raw mapping, highlighting visually the difference in values distribution of a sample Sony raw image against a Night23 raw image, and how this difference is reduced after mapping.

In Table 5 are reported the values of PSNR and SSIM obtained on source paired data, focusing on two different cameras (Sony and Fuji) of the SID dataset. The metrics are computed on thumbnail images, as done during the optimization setup: results on full resolution images are presented later on, on the target Night23 dataset. The reported experiments refer to different configurations of parameters in the night rendering pipeline: “Default” indicates the original setup from [24], while “Bayes optimization” refers to the configuration reached with the approach presented in this paper. Specifically, two variants are explored: “from default values” uses the default parameters as a starting point for optimization, whereas “from random values” does not exploit this information. The rationale is to investigate to what extent a priori domain knowledge may facilitate the Bayesian optimization process.

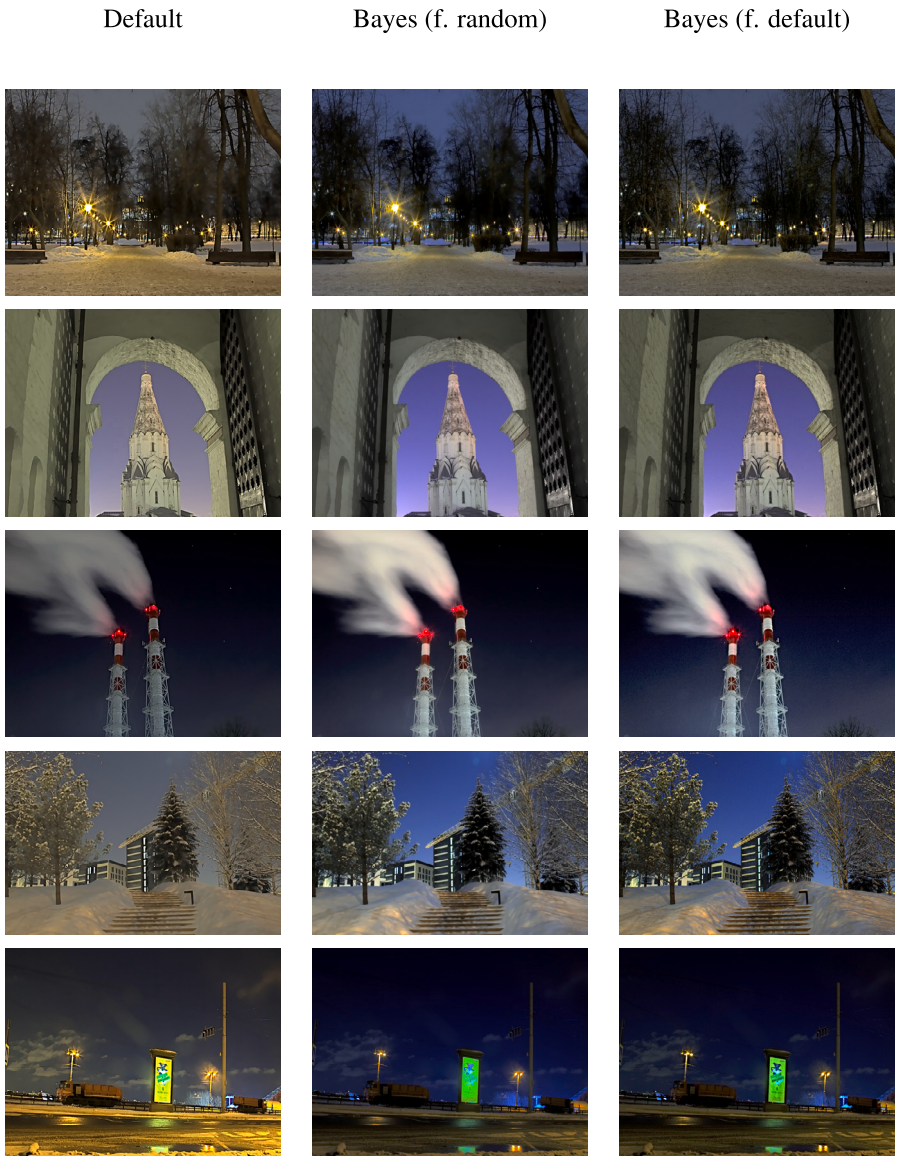

Fig. 4. Visual comparison of the results obtained on the Night 23 dataset. The first column shows the results obtained using the default set of parameters, while the second and the third show the results obtained with the two sets of optimized parameters.

As can be seen in Table 5, the Bayesian optimization brings an improvement on both metrics, compared to the use of default parameters. This is related to the fact that the original default parameters of the pipelines had been specifically selected for the Night23 dataset processing, and hence they do not fit well with the SID dataset. Comparing the two starting points for Bayesian optimization, there are small, and inconsistent, performance variations between the two cameras: while Sony favors starting from default values, Fuji favors starting from random values. This suggests that there is no significant impact given by providing a priori knowledge on the starting point. The observation is further corroborated by Fig. 4, where only slight differences between the two optimizations are visible.

As previously described, the proposed optimization procedure exploits the ground truth of a secondary dataset (SID, in our setup) to optimize the parameters of a rendering pipeline. For this reason, the results necessarily fit the color distribution of the dataset used for the optimization. In Fig. 5 are reported the mean RGB histograms of the Night23 dataset images processed with different parameters and the SID Sony high light images.

4.4.1. *Psychovisual study*

We conduct a psychovisual study, structured in two parts, to qualitatively assess the results of our night rendering pipeline optimization.

In the first part, participants to the study assessed three distinct sets of images processed through different pipelines:

1. The original, manually-tuned pipeline by Zini et al. serving as a baseline.
2. Our Bayesian-based optimizations of the pipeline, using SID-Sony images as a reference.
3. Our Bayesian-based optimizations of the pipeline, using SID-Fuji images as a reference.

The results, depicted in Fig. 6 (top), show that the original pipeline is outperformed by either Bayesian-based optimization. Specifically, the gathered feedback revealed a clear preference for the SID-Sony optimized pipeline, which participants found to produce images with superior clarity and color accuracy. The SID-Fuji pipeline, while also improving upon the original, was slightly less favored, indicating that

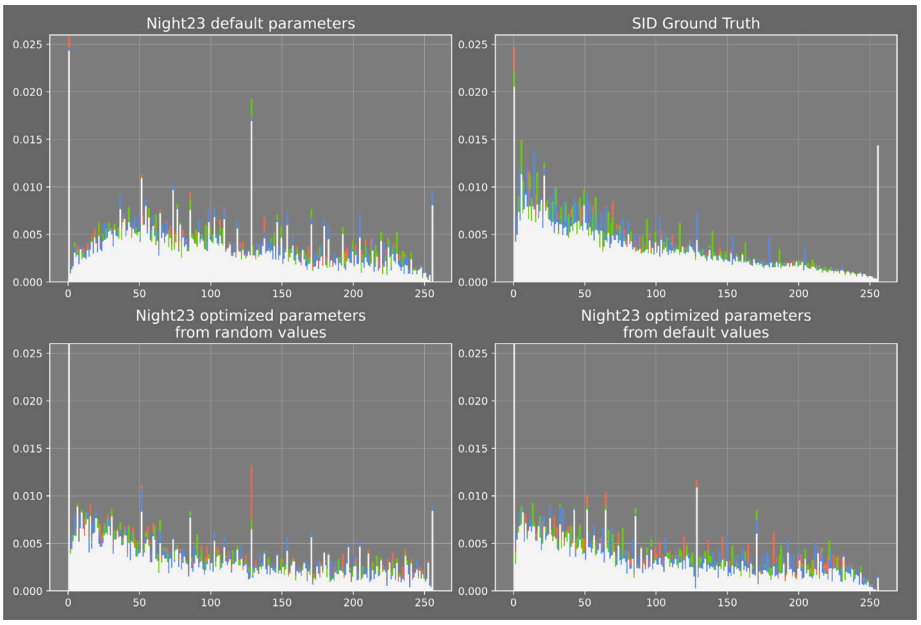

Fig. 5. Color distributions of the Night23 dataset and the SID-Sony dataset. In the first row are reported the distributions of the Night23 processed by the original pipeline and the SID-Sony Ground Truth; in the second row are reported the color distributions of the results obtained on the Night23 dataset with the two different sets of optimized parameters, using the SID-Sony dataset for optimization.

the choice of reference dataset plays a critical role in the optimization's effectiveness.

For the second part, we selected eight recent state-of-the-art approaches for low-light image enhancement and performed a psychovisual evaluation test using the same 50 images from the validation set, processed by the selected methods. More precisely, we selected EnlightenGAN [39], ExCNet [40], Kind++ [41], Zero-DCE [42], Zero-DCE++ [43], QuadPrior [10], and two variants of ZeroIG [9] trained respectively on LOL images and Huawei smartphone images. Since not all these methods are originally designed to handle the sensor-specific green tint typical of raw images, we preprocess the images through a Gray world white balancing algorithm in order to provide a more fair setup for such methods. The results, depicted in Fig. 6 (bottom), indicate that our Bayesian-Optimized SID-Sony pipeline was rated the best in perceptual quality, followed directly by Zero-DCE++ and ZeroIG in its Huawei variant. In addition, in Fig. 7 we show a visual comparison of the results of selected state-of-the-art methods and our B.O. SID Sony optimized pipeline, providing a general view of the rendering capabilities of different solutions. Detail analysis is shown later on.

We also evaluate the results of our Bayesian pipeline optimization against the same methods in the state of the art, by resorting to lightness order error, as well as standard no-reference image quality metrics. The results, reported in Table 6, show that our solution is competitive throughout most considered metrics, resulting in the top 3 solutions for LOE, NIQE, and BRISQUE. Intermediate results are obtained through PIQE, which seems to favor QuadPrior. This result contrasts with the psychovisual study, suggesting a possible discrepancy between said metric and perceived quality in night images. This hypothesis opens an interesting direction for future investigation. We have also reported the number of parameters and floating-point operations of the models [44]. It is worth noting that the proposed solution not only has a much smaller number of parameters, as already mentioned, but is also the second best in terms of GFLOPs, second only to Zero-DCE++, which represents the most efficient solution.

Finally, Fig. 8 offers a close-up view of the details generated by our B.O. SID Sony solution, compared to recent methods QuadPrior and ZeroIG (LOL), and the second best method according to quantitative analysis. QuadPrior is shown to introduce a loss in image resolution that is tied to the underlying diffusion-based generative process. Conversely, both Zero-DCE++ and ZeroIG (LOL) produce a competitive level of detail rendering, with ZeroIG (LOL) losing partial detail in highly illuminated local areas.

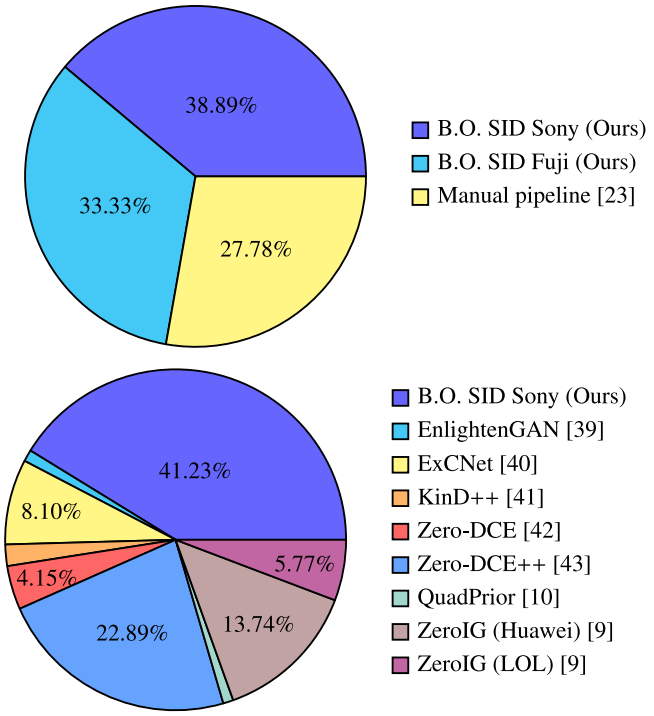

Fig. 6. Pie charts reporting the distribution of the results of the psycho-visual test, in terms of number of preferences. On the top we compare different versions of our solution. On the bottom, we compare our best configuration against other state-of-the-art low-light image enhancement approaches.

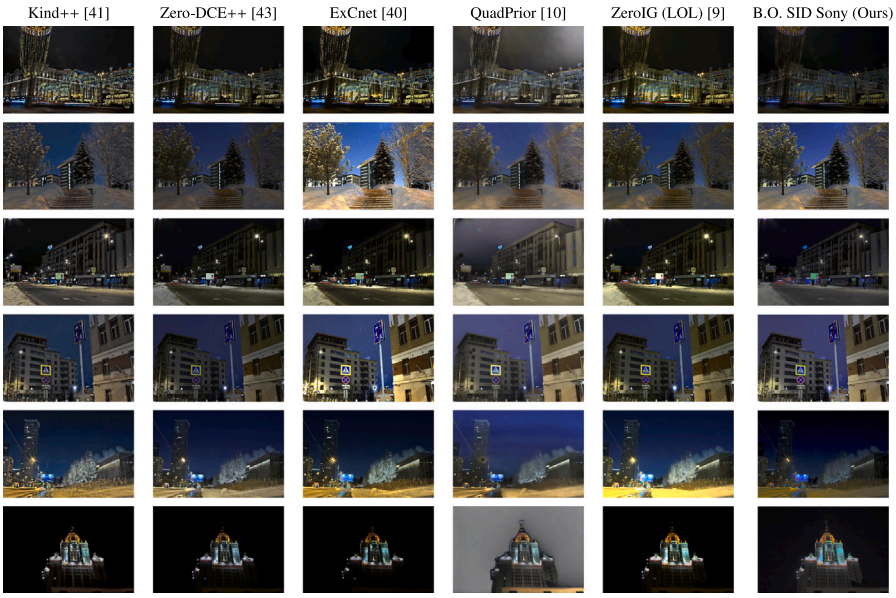

Fig. 7. Visual comparison of images from the Night23 dataset rendered with our optimized pipeline (B.O. SID-Sony) and with state-of-the-art approaches.

Table 6
Comparison with the state-of-the-art approaches using LOE, no reference metrics, alongside the number of parameters and floating point operations (FLOPs). For all values except LOE, the lower, the better. In parentheses are reported the rankings for each metric, with the best per column highlighted in boldface, and the second best underlined. The last column reports the average ranking.

Method	Parameters (M) ↓	FLOPs (G) ↓	LOE ↑	NIQE ↓	BRISQUE ↓	PIQE ↓	Average rank
Manual [23]	$15*e^{-6}$	3.998	0.9132 (6)	3.8966 (9)	30.4424 (10)	44.7492 (10)	8.75
B.O. SID Fuji (ours)	$15*e^{-6}$	3.998	0.9133 (5)	2.9371 (4)	20.3956 (4)	33.5735 (7)	5
B.O. SID Sony (ours)	$15*e^{-6}$	3.998	<u>0.9194</u> (2)	2.9133 (3)	19.2530 (3)	34.0091 (8)	<u>4</u>
EnlightenGAN [39]	8.637	273.240	0.7537 (11)	<u>2.8252</u> (2)	27.7384 (9)	33.4390 (6)	7
ExCnet [40]	8.274	–	0.8970 (8)	4.1445 (11)	26.2683 (6)	27.4960 (4)	7.25
KinD++ [41]	8.275	12 238.026	0.8859 (9)	3.8346 (8)	36.0030 (11)	49.9703 (11)	9.75
Zero-DCE [42]	0.079	84.990	0.9119 (7)	**2.6934** (1)	27.1583 (7)	31.0900 (5)	5
Zero-DCE++ [43]	0.010	0.115	**0.9371** (1)	3.3054 (5)	27.2950 (8)	41.7243 (9)	5.75
QuadPrior [10]	1313.603	–	0.7600 (10)	4.1380 (10)	23.8110 (5)	**21.3203** (1)	6.5
ZeroIG (Huawei) [9]	0.086	5013.563	0.9146 (4)	3.6834 (7)	**16.2269** (1)	<u>24.1564</u> (2)	**3.5**
ZeroIG (LOL) [9]	0.086	5013.563	0.9159 (3)	3.6456 (6)	<u>16.3019</u> (2)	24.3302 (3)	**3.5**

5. Conclusions

In this paper, we have introduced a novel approach for optimizing Image Signal Processing pipelines for night photography using Bayesian derivative-free methods. Our method addresses several key challenges in the domain, including the need for paired data and the limitations imposed by differentiability requirements in conventional deep learning methods. By leveraging a raw-to-raw conversion process and external paired datasets, we have demonstrated the ability to adapt and optimize existing ISP pipelines for specific camera setups without manual tuning.

Our experimental results show improvements in image quality as evaluated by both objective metrics and subjective assessments through psychovisual studies. The SID-Sony optimized pipeline, in particular, was found to be highly effective, outperforming the original pipeline and competing closely with state-of-the-art methods while maintaining a low computational complexity.

Our approach delegates the requirement for paired data to a third-party dataset, significantly improving its applicability to new sensors. However, this introduces a dependency on the specific characteristics of the training data, since different rendering styles in the third-party dataset would lead to different optimal pipeline parameters, and to a different final rendering. This suggests the opportunity to fine tune different variants of a rendering style by resorting to a limited number of user-specific rendering examples [45]. Additionally, while the Bayesian derivative-free optimization removes the dependency on differentiability that is characteristic of backpropagation, it shares its limited applicability to parameter optimization, thus ignoring structure optimization. To this extent, we consider expanding our work by integrating it into a reinforcement learning paradigm, or by exploiting techniques from the field of cartesian genetic programming [46].

Overall, our approach not only enhances the performance of night photography rendering but also provides a scalable and flexible framework that can be adapted to various imaging conditions and camera devices. This work paves the way for future research in ISP optimization, emphasizing the importance of flexible, data-efficient methods over traditional, labor-intensive tuning processes.

CRediT authorship contribution statement

Simone Zini: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Marco Buzzelli:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

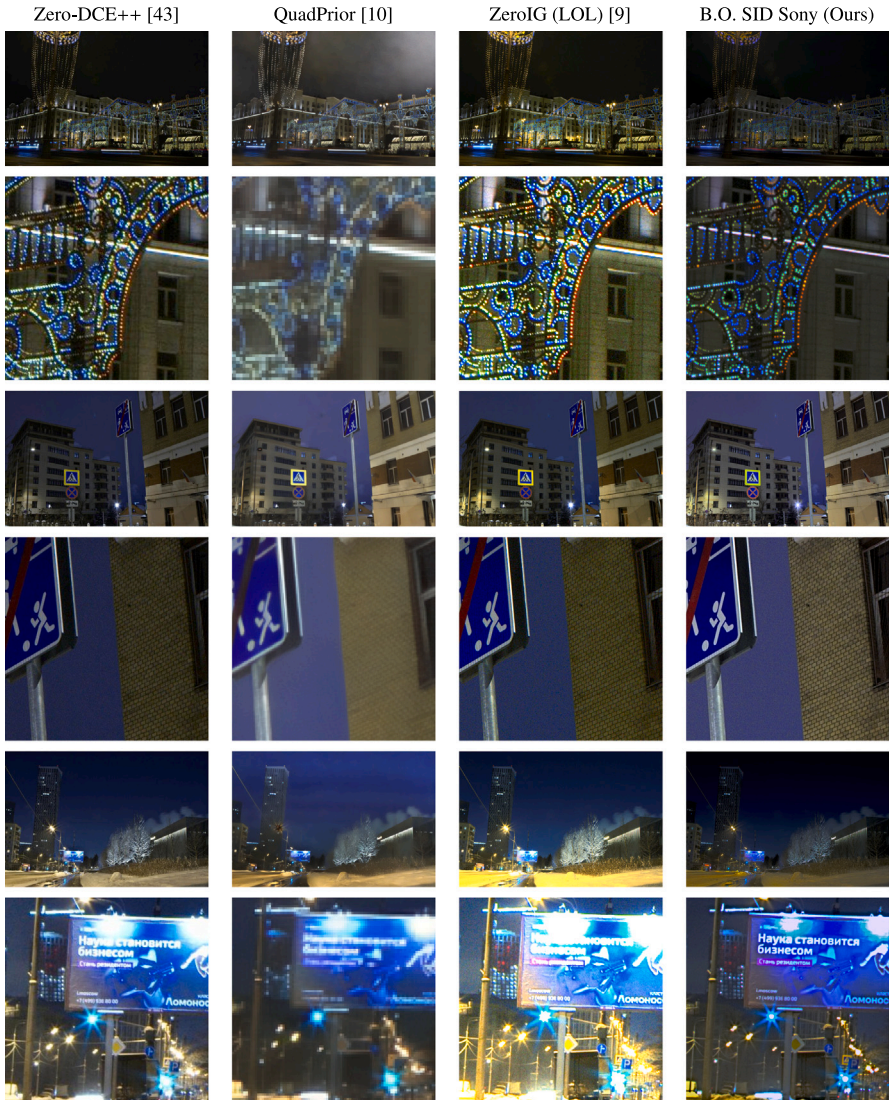

Fig. 8. Visual comparison of images from the Night23 dataset rendered with our optimized pipeline (B.O. SID-Sony) and with state-of-the-art approaches. Zoomed sections of different test images are reported here to compare how the approaches perform on small details in the images.

Data availability

Data is already publicly available.

References

[1] S. Oishi, N. Fukushima, Retinex-based relighting for night photography, Appl. Sci. 13 (3) (2023) 1719.
[2] A. Yamasaki, H. Takauji, S. Kaneko, T. Kanade, H. Ohki, Denighting: Enhancement of nighttime images for a surveillance camera, in: 2008 19th International Conference on Pattern Recognition, IEEE, 2008, pp. 1–4.
[3] S.J. Kim, H.T. Lin, Z. Lu, S. Süsstrunk, S. Lin, M.S. Brown, A new in-camera imaging model for color computer vision and its application, IEEE Trans. Pattern Anal. Mach. Intell. 34 (12) (2012) 2289–2302.
[4] X. Liu, W. Ma, X. Ma, J. Wang, LAE-net: A locally-adaptive embedding network for low-light image enhancement, Pattern Recognit. 133 (2023) 109039.
[5] F. Zhou, X. Sun, J. Dong, X.X. Zhu, SurroundNet: Towards effective low-light image enhancement, Pattern Recognit. 141 (2023) 109602.
[6] K. Wei, Y. Fu, Y. Zheng, J. Yang, Physics-based noise modeling for extreme low-light photography, IEEE Trans. Pattern Anal. Mach. Intell. 44 (11) (2021) 8520–8537.
[7] R. Nguyen, D.K. Prasad, M.S. Brown, Raw-to-raw: Mapping between image sensor color responses, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 3398–3405.
[8] M. Afifi, A. Abuolaim, Semi-supervised raw-to-raw mapping, 2021, arXiv preprint arXiv:2106.13883.
[9] Y. Shi, D. Liu, L. Zhang, Y. Tian, X. Xia, X. Fu, ZERO-IG: Zero-shot illumination-guided joint denoising and adaptive enhancement for low-light images, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, pp. 3015–3024.
[10] W. Wang, H. Yang, J. Fu, J. Liu, Zero-reference low-light enhancement via physical quadruple priors, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, pp. 26057–26066.
[11] J.-Y. Zhu, T. Park, P. Isola, A.A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2223–2232.
[12] T. Park, A.A. Efros, R. Zhang, J.-Y. Zhu, Contrastive learning for unpaired image-to-image translation, in: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IX 16, Springer, 2020, pp. 319–345.
[13] M. Buzzelli, R. Riva, S. Bianco, R. Schettini, Consensus-driven illuminant estimation with GANs, in: Thirteenth International Conference on Machine Vision, vol. 11605, SPIE, 2021, pp. 578–584.
[14] S.W. Zamir, A. Arora, S. Khan, F.S. Khan, L. Shao, Learning digital camera pipeline for extreme low-light imaging, Neurocomputing 452 (2021) 37–47.
[15] E. Ershov, A. Savchik, D. Shepelev, N. Banić, M.S. Brown, R. Timofte, K. Koščević, M. Freeman, V. Tesalin, D. Bocharov, et al., NTIRE 2022 challenge on night photography rendering, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 1287–1300.
[16] A. Shutova, E. Ershov, G. Perevozchikov, I. Ermakov, N. Banić, R. Timofte, R. Collins, M. Efimova, A. Terekhin, S. Zini, et al., NTIRE 2023 challenge on night photography rendering, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 1981–1992.

[17] Z. Li, S. Yi, Z. Ma, Rendering nighttime image via cascaded color and brightness compensation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 897–905.
[18] S. Liu, C. Feng, X. Wang, H. Wang, R. Zhu, Y. Li, L. Lei, Deep-flexisp: A three-stage framework for night photography rendering, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 1211–1220.
[19] Y. Hu, B. Wang, S. Lin, Fc4: Fully convolutional color constancy with confidence-weighted pooling, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4085–4094.
[20] P.V. Gehler, C. Rother, A. Blake, T. Minka, T. Sharp, Bayesian color constancy revisited, in: 2008 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2008, pp. 1–8.
[21] P. Liu, H. Zhang, K. Zhang, L. Lin, W. Zuo, Multi-level wavelet-CNN for image restoration, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2018, pp. 773–782.
[22] C. Desai, N. Akalwadi, A. Joshi, S. Malagi, C. Mandi, R.A. Tabib, U. Patil, U. Mudenagudi, LightNet: Generative model for enhancement of low-light images, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 2231–2240.
[23] S. Zini, C. Rota, M. Buzzelli, S. Bianco, R. Schettini, Back to the future: a night photography rendering ISP without deep learning, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 1465–1473.
[24] S. Zini, C. Rota, M. Buzzelli, S. Bianco, R. Schettini, Shallow camera pipeline for night photography enhancement, in: International Conference on Image Analysis and Processing, Springer, 2023, pp. 51–61.
[25] A. Buades, B. Coll, J.-M. Morel, Non-local means denoising, Image Process. Line 1 (2011) 208–212.
[26] F.N. Fritsch, R.E. Carlson, Monotone piecewise cubic interpolation, SIAM J. Numer. Anal. 17 (2) (1980) 238–246.
[27] P.I. Frazier, A tutorial on Bayesian optimization, 2018, arXiv preprint arXiv:1807.02811.
[28] J. Bergstra, R. Bardenet, Y. Bengio, B. Kégl, Algorithms for hyper-parameter optimization, in: Advances in Neural Information Processing Systems, vol. 24, 2011.
[29] Z. Wang, A.C. Bovik, Modern image quality assessment, Synthesis Lectures on Image, Video, and Multimedia Processing, vol. 2 (1) (2006) 1–156.
[30] C. Chen, Q. Chen, J. Xu, V. Koltun, Learning to see in the dark, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 3291–3300.
[31] C. Wei, W. Wang, W. Yang, J. Liu, Deep retinex decomposition for low-light enhancement, in: British Machine Vision Conference 2018, BMVC 2018, Newcastle, UK, September 3-6, 2018, BMVA Press, 2018, p. 155, URL http://bmvc2018.org/contents/papers/0451.pdf.
[32] S. Falkner, A. Klein, F. Hutter, BOHB: Robust and efficient hyperparameter optimization at scale, in: International Conference on Machine Learning, PMLR, 2018, pp. 1437–1446.
[33] S. Kullback, R.A. Leibler, On information and sufficiency, Ann. Math. Stat. 22 (1) (1951) 79–86.
[34] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612.
[35] S. Wang, J. Zheng, H.-M. Hu, B. Li, Naturalness preserved enhancement algorithm for non-uniform illumination images, IEEE Trans. Image Process. 22 (9) (2013) 3538–3548.
[36] A. Mittal, A.K. Moorthy, A.C. Bovik, No-reference image quality assessment in the spatial domain, IEEE Trans. Image Process. 21 (12) (2012) 4695–4708.
[37] A. Mittal, R. Soundararajan, A.C. Bovik, Making a "completely blind" image quality analyzer, IEEE Signal Process. Lett. 20 (3) (2012) 209–212.
[38] N. Venkatanath, D. Praneeth, M.C. Bh, S.S. Channappayya, S.S. Medasani, Blind image quality evaluation using perception based features, in: 2015 Twenty First National Conference on Communications, NCC, IEEE, 2015, pp. 1–6.
[39] Y. Jiang, X. Gong, D. Liu, Y. Cheng, C. Fang, X. Shen, J. Yang, P. Zhou, Z. Wang, Enlightengan: Deep light enhancement without paired supervision, IEEE Trans. Image Process. 30 (2021) 2340–2349.
[40] L. Zhang, L. Zhang, X. Liu, Y. Shen, S. Zhang, S. Zhao, Zero-shot restoration of back-lit images using deep internal learning, in: Proceedings of the 27th ACM International Conference on Multimedia, 2019, pp. 1623–1631.
[41] Y. Zhang, X. Guo, J. Ma, W. Liu, J. Zhang, Beyond brightening low-light images, Int. J. Comput. Vis. 129 (2021) 1013–1037.
[42] C. Guo, C. Li, J. Guo, C.C. Loy, J. Hou, S. Kwong, R. Cong, Zero-reference deep curve estimation for low-light image enhancement, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 1780–1789.
[43] C. Li, C. Guo, C.C. Loy, Learning to enhance low-light image via zero-reference deep curve estimation, IEEE Trans. Pattern Anal. Mach. Intell. 44 (8) (2021) 4225–4238.
[44] C. Li, C. Guo, L. Han, J. Jiang, M.-M. Cheng, J. Gu, C.C. Loy, Low-light image and video enhancement using deep learning: A survey, IEEE Trans. Pattern Anal. Mach. Intell. 44 (12) (2021) 9396–9416.
[45] S. Bianco, C. Cusano, F. Piccoli, R. Schettini, Personalized image enhancement using neural spline color transforms, IEEE Trans. Image Process. 29 (2020) 6223–6236.
[46] J.F. Miller, Cartesian genetic programming: its status and future, Genet. Program. Evolvable Mach. 21 (1) (2020) 129–168.

Simone Zini is an Assistant Professor at the Department of Informatics, Systems and Communication (DISCo) of the University of Milano-Bicocca. He obtained his Bachelor Degree and Master Degree in Computer Science at University of Milano-Bicocca (Italy), respectively in 2015 and 2018, focusing on image processing and enhancement. In 2018 he worked in Tokyo, Japan, at NEC Corporation as a research intern, on the topic of Image denoising and enhancement, and obtained a research grant at the Imaging and Vision Laboratory of the University of Milano-Bicocca. He obtained his Ph.D. in 2022 at the University of Milano-Bicocca (Italy), where he is currently employed as assistant professor, after two years as post-doctoral researcher at National Institute of Nuclear Physics (INFN). His main topics of research are machine learning, image and signal processing and enhancement, computational photography and remote sensing . He is currently working on the AdvaNced Technologies for HumancentrEd Medicine (ANTEHM) project.

Marco Buzzelli obtained his Ph.D. in Computer Science in 2019 at the University of Milano-Bicocca (Italy), where he currently serves as assistant professor at the Department of Informatics, Systems and Communication. He is actively engaged in conducting cutting-edge research in the field of signal/image/video processing and understanding, using machine learning techniques. He is particularly passionate about color imaging. Marco Buzzelli has actively collaborated with European institutions, including but not limited to Universitat Autònoma de Barcelona, Universidade Nova de Lisboa, Université Jean Monnet, Universidad de Granada. These collaborations have allowed him to contribute to the European AI landscape as an active ELLIS member, while also gaining valuable insights and exposure to diverse perspectives.