This CVPR 2021 paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

End-to-end High Dynamic Range Camera Pipeline Optimization

Nicolas Robidoux[1] Dong-eun Seo[1] Federico Ariza[1]
Luis E. García Capel[1] Avinash Sharma[1] Felix Heide[1,2]

[1]Algolux [2]Princeton University

Abstract

The real world is a 280 dB High Dynamic Range (HDR) world which imaging sensors cannot record in a single shot. HDR cameras acquire multiple measurements with different exposures, gains and photodiodes, from which an Image Signal Processor (ISP) reconstructs an HDR image. Dynamic scene HDR image recovery is an open challenge because of motion and because stitched captures have different noise characteristics, resulting in artifacts that ISPs must resolve in real time at double-digit megapixel resolutions. Traditionally, ISP settings used by downstream vision modules are chosen by domain experts; such frozen camera designs are then used for training data acquisition and supervised learning of downstream vision modules. We depart from this paradigm and formulate HDR ISP hyperparameter search as an end-to-end optimization problem, proposing a mixed 0^{th} and 1^{st}-order block coordinate descent optimizer that jointly learns sensor, ISP and detector network weights using RAW image data augmented with emulated SNR transition region artifacts. We assess the proposed method for human vision and image understanding. For automotive object detection, the method improves mAP and mAR by 33% over expert-tuning and 22% over state-of-the-art optimization methods, outperforming expert-tuned HDR imaging and vision pipelines in all HDR laboratory rig and field experiments.

1. Introduction

Real-world scenes have dynamic ranges that often exceed 1,000,000:1 (120 dB) [50] and, in extreme cases like tunnel exit in direct sunlight, reach over 200 dB. This dynamic range must be captured by vision algorithms for safety-critical decision making in robotics and navigation. Existing sensors cannot capture High Dynamic Range (HDR) in a single shot [9, 12, 38]. As a result, modern cameras rely on sequentially and spatially-multiplexed acquisition techniques, combining data acquired with different exposure times, gains and photodiodes.

Image Signal Processors (ISPs) are low-level pipelines implemented in hardware that convert RAW sensor pixel data into images suitable for human viewing or scene understanding tasks such as object detection and classification. ISPs thus form an essential interface and abstraction layer between the sensor and the display or computer vision module. ISP processing blocks are configured with tens to hundreds of adjustable hyperparameters which define its static and dynamic behavior [44, 46, 49, 58], for example adaptation to noise level. Choosing optimal ISP hyperparameter values is challenging as they depend on the context in which the camera is used (portraits and landscapes vs. all-weather autonomous driving), on the specifics of the lens and sensor (before the ISP), and on the downstream task (display to human viewers vs. object detection).

Traditionally, imaging experts have manually selected ISP hyperparameter values using charts and visual inspection [44, 58]. The potential of automated loss-based hardware ISP hyperparameter optimization in the low-dynamic range (LDR) context, using differentiable approximations [58] or 0^{th}-order (derivative-free) methods [44, 46], was recently established. These methods rely on gain separability and consequently are limited to LDR image processing; HDR optimization requires novel approaches. End-to-end loss-based optimization has not included sensor hyperparameters and work on the optimization of non-differentiable ISPs for CV [44, 58] has kept the downstream Convolutional Neural Network (CNN) detector fixed. In this work, we tackle HDR and jointly optimize the sensor, ISP and CNN.

The search for optimal HDR imaging pipelines is an open problem central to imaging and vision tasks in uncontrolled in-the-wild scenarios. Real-time applications, e.g., in robotics and autonomous driving, and high sensor resolutions, up to triple-digit megapixel counts, mandate efficient hardware implementations [6]. Multiplexing makes HDR processing an open challenge. Motion causes ghosting artifacts when captures acquired sequentially or with different exposure times are stitched together [14]. Split-pixel sensors, with two or more diodes per pixel [60], reduce motion blur discrepancies but are often used with multiple ex-

posure times. With few captures (four or less in automotive imaging [52]), signal-to-noise ratio (SNR) transition regions show sudden texture changes, resulting in spurious edge detections by the Human Visual System and CNN detectors. Complicating matters, some ISP nodes behave differently in HDR; for example, color artifacts occur near knee points of the companding curve.

Departing from handcrafted ISP hyperparameter tuning, we propose a task-specific, loss-driven, end-to-end approach to the joint optimization of the sensor, ISP and detector for downstream applications such as human viewing and object detection. Optimization for human viewing is performed with multiple losses, including Contrast Weighted Lp-Norm, a novel full reference image difference metric based on Larkin's universal Noise Visibility Function [32], and a dynamic HDR lab setup covering 123 dB. When optimizing for image understanding, instead of acquiring large datasets containing SNR transition region edge cases in semantic scene content, we augment data with a proposed SNR transition region artifact emulation method. The proposed block coordinate descent approach combines a 0^{th}-order evolutionary optimizer (with novel centroid weights that stabilize boundary minima) with 1^{st}-order Stochastic Gradient Descent optimization, demonstrating the first method that jointly optimizes hardware hyperparameters and downstream CNN detector weights. The method is validated with state-of-the-art hardware sensors and ISPs in an HDR lab and in outdoor, in-the-wild human viewing and automotive object detection HDR scenarios.

In summary, we make the following contributions:

- We propose the first end-to-end hardware-in-the-loop optimization method for the hyperparameters of multi-exposure HDR camera systems, and the first method for the joint optimization of sensor and ISP hardware hyperparameters and CNN weights of a vision module.
- We propose a dynamic HDR lab setup, a full reference perceptual image difference metric, and a data augmentation methodology targeting HDR stitching artifacts.
- With state-of-the-art automotive ISPs and sensors, we validate the proposed method experimentally and in simulation for human viewing and 2D object detection. Across all tasks considered in this paper, the proposed method outperforms existing methods.

The proposed method has the following limitations. Unlike Mosleh *et al*. [44], we only consider one image understanding task, namely object detection and classification. We sparsely sample *sensor* hyperparameters; a methodology with a finer grain, involving for example multiple cameras or coarse optimization followed by additional field data acquisition, is needed. We only optimize single frame image processing; RAW video sequences could be fed to an ISP to process temporal cues.

2. Related Work

High Dynamic Range Image Acquisition Hallucinating HDR from LDR content [12, 13, 35, 36, 40] is not an alternative for safety-critical applications. Actual HDR imaging increases dynamic range by stitching measurements made with different photodiodes, exposures and/or gains [4, 9, 38, 39, 51, 54, 59]. Temporal multiplexing introduces severe motion artifacts in dynamic scenes [9, 16, 19, 38, 41, 51]. They are addressed by a large body of work, from post-capture stitching [14, 15, 21, 26–28, 53] to optical flow [37] and deep learning [24, 25]. Split-pixel HDR sensors reduce motion artifacts by multiplexing with colocated photodiodes with different response sensitivities [10, 55, 57].

Optimization of Image Processing Pipelines Sensor and ISP Hyperparameter optimization should not be confused with adaptive capture controls like Auto-Exposure (AE) [48, 64]. Hyperparameters *configure* camera systems; they are persistent and fixed during normal operation.

ISPs have traditionally been manually optimized [6]. Recent work demonstrated the potential of automated loss-based ISP hyperparameter optimization for LDR. Human viewing loss functions parallel image quality metrics and standards [8, 22, 42, 43, 48, 62]. Computer vision loss functions are evaluated on the output of a downstream image understanding module [44]. Nishimura *et al*. [46, 61] optimized a model software ISP by combining a global swarm-intelligence optimization method with local Nelder-Mead. Portelli *et al*. [49] optimized a simple model software ISP with a Particle Swarm Optimization method. Tseng *et al*. [58] optimized hardware ISPs by training a CNN to mimic it and optimizing this differentiable proxy with Stochastic Gradient Descent. Very recently, Mosleh *et al*. [44] directly optimized hardware ISPs, without approximation, with a two-step method: search space remapping based on random sampling and statistical analysis, followed by CMA-ES [17, 23]. Like Tseng and Mosleh, we formulate the selection of ISP hyperparameter values as a black-box optimization problem driven by end-to-end losses; reliant on gain separability, their work does not extend to HDR. Furthermore, Tseng *et al*. rely on approximating the hardware, while Mosleh *et al*.'s 0^{th}-order solver is not suited for the optimization of CNNs. None considered sensors.

ISP Hyperparameter Optimization for Computer Vision The impact of ISP hyperparameter values on the performance of a downstream vision module was explored in [7, 11, 44, 58, 61, 63–65]. Image understanding optimization has been driven by various end-to-end losses. Tseng *et al*. [58] optimized hardware ISPs for object detection and classification using Intersection over Union loss (IoU [47]). Wu *et al*. [61] optimized a simple model ISP for object detection and classification using mean Average Precision (mAP [47]). Mosleh *et al*. [44] optimized hardware ISPs for object detection and classification using mAP and mAR

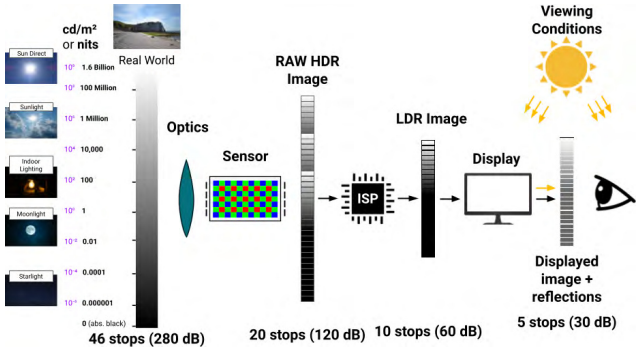

Figure 1: Camera image formation process. Scene radiance goes through the optical system and reaches the sensor. An HDR RAW image is built from multiple exposures. The ISP reverses the capture process and renders an image suitable for viewing or CV.

(mean Average Recall [47]), object segmentation with mAP, and panoptic segmentation with PQ [29]. None jointly optimized an hardware ISP and an image understanding module. We close this gap and jointly train a deep CNN. The work of Diamond *et al*. [11] comes closest. It is restricted to differentiable software ISPs: they jointly optimized a trainable software ISP and a downstream object classification model using Top-1 and Top-5 classification accuracy.

3. Background and Image Formation

The total dynamic range of the human eye is about 46 stops (280 dB), from 10^{-6} cd/m^2 at dimmest to 10^{8} cd/m^2 where retinal damage may occur [20]; the instantaneous dynamic range is much lower. Unlike still photographers who control lighting conditions, surveillance or automotive imaging applications must accurately capture up to 144 dB in rapidly changing light conditions. The range of light measurable by a sensor in a single capture is limited. The ratio between the brightest and darkest measurable signals is essentially fixed: Varying exposure time adapts to bright and dark conditions by shifting the light intensity capture window, but sensor dynamic range is limited at the top end by the electron well capacity and, at the bottom end, by the photodiode sensitivity, electronic noise and sensor bit depth.

HDR, Optics, Sensors and Multiplexing HDR image capture addresses situations in which 10 stops of light reaches the sensor, a common real-world scenario [51]. A typical camera image formation chain is shown in Fig. 1. Light sources and reflective surfaces send radiation towards the imaging system. It reaches the optical system and is focused onto the image sensor, which also receives internal reflections and scattering. Optical noise thus includes veiling glare, stray light and images of the aperture visible as lens flare. The veiling glare floor is a hard limit on the dynamic range. A raw measurement $\mathrm{I} \in \mathbb{R}^{\mathrm{W}\times\mathrm{H}}_{[0,\max]}$ captured by a sensor with resolution $\mathrm{W}\times\mathrm{H}$ is given by the response of all the elements between the lens and the image sensor:

$$\mathrm{I} = f(\mathbf{L}((\mathbf{E}\Delta t) * \mathbf{P}) + \boldsymbol{\eta}). \tag{1}$$

$\mathbf{E} \in \mathbb{R}^{\mathrm{W}\times\mathrm{H}}_{[0,\infty)}$ is the irradiance of the scene (the HDR ground truth), $\Delta t \in \mathbb{R}_{[0,\infty)}$ is the exposure time, $\mathbf{P} \in \mathbb{R}^{\mathrm{W}\times\mathrm{H}}_{[0,\infty)}$ is the optical Point Spread Function (PSF), $\mathbf{L}$ is glare formation [56], $\boldsymbol{\eta} \in \mathbb{R}^{\mathrm{W}\times\mathrm{H}}$ is sensor noise (fixed pattern noise included), and f is the sensor response, a non-linear mapping considered smooth and monotonic clipped to $[0, \max]$.

Multiple such measurements are combined to produce one HDR image. Existing methods to acquire the individual measurements are described in the following:

Temporal Multiplexing: Multiple LDR images are captured in rapid succession with different exposure times and merged into a single HDR image. This works well for still photography, but introduces artifacts in the transitions between captures around and within moving objects.

Spatial Multiplexing: Individual pixels in the sensor array, in an alternating pattern, use different exposure times or gains allowing them to be captured simultaneously. This reduces temporal discrepancies *and* spatial resolution.

Split-Pixel: Each sensor pixel has two photodiodes: one small and one large. The small photodiode captures fewer photons and acts like a *short* exposure; the large one, a *long* exposure. Different gains may also be used.

HDR Image Formation Model The pixel values of J different captures are combined into an HDR irradiance map. Assuming, without loss of generality, a split-pixel sensor, RAW data is modeled as a tuple of exposures by multiple diodes, with gains folded into effective exposure times Δt_j ($j \in \{1, \ldots, \mathrm{J}\}$). The Sony IMX490 sensor used in this work, for example, acquires 4-tuples: two diodes with two conversion gains. Assuming constant irradiance and disregarding the PSF, glare and noise, the *estimated relative log-irradiance* of the j^{th} capture at pixel location i is

$$\ln \tilde{\mathrm{E}}_{ji} = \ln\left(f^{-1}(\mathrm{I}_{ji})/\Delta t_j\right) = \ln f^{-1}(\mathrm{I}_{ji}) - \ln \Delta t_j, \tag{2}$$

where I_{ji} is the j^{th} sensor measurement value at pixel i and f^{-1} is the inverse camera curve [9] that returns I_{ji} to the linear domain. In principle, Eq. 2 holds everywhere but at under- and over-exposed pixels. In practice, however, temporal misalignment between captures can induce large deviations. Captures are aligned using an image warping function $\mathrm{I}'_j = h(\mathrm{I}_j)$ [51], from which the HDR log-irradiance map is reconstructed as the weighted average

$$\ln \tilde{\mathrm{E}}_i = \frac{1}{\sum_j w(\mathrm{I}'_{ji})} \sum_j w(\mathrm{I}'_{ji}) \left(\ln(f^{-1}(\mathrm{I}'_{ji}) - \ln \Delta t_j\right). \tag{3}$$

This extends dynamic range by lowering the effective noise floor [51], leaving optical noise as the dominant contributor in static scenes with large J. With small J or dynamic content however, existing methods introduce hard to correct artifacts like ghosting and SNR discontinuities (see Fig. 2).

Figure 2: Common HDR multiplexing artifacts. Crops (a) and (b): ghosting. Crop (c): SNR discontinuity.

4. HDR Fusion Simulation

Field data usually contains few samples where stitching artifacts impact detection. We augment field data with simulated SNR drop artifacts *post-stitching*, that is, we modify captures already passed through the (on-sensor) stitcher.

To optimize sensor hyperparameters, we acquire training data using consecutive captures that sample 254 combinations of sensor hyperparameter values (see Supplemental Document). To augment this data with simulated HDR fusion artifacts, we first determine, for each sequence, the RAW pixel level L above which noise is added: Each stitched frame is decompanded into "linear RAW" (approximately proportional to photon counts), the median pixel value within the largest bounding box of each frame is computed, and then L is the median of the frames' medians, ignoring frames with no bounding box. The Sony IMX490 sensor stitches four exposures; three gain ratios, fixed in hardware, consequently drive noise discontinuities. 70% of the field RAWs are not modified; each gain ratio is used with 10% of them. Sequences with the lowest L value use the gain ratio between the two highest gain exposures (17.5), those with the highest L value use the gain ratio between the two lowest gain exposures (65.0), and the rest use the other gain ratio (6.49). We call this ratio "R".

Sensor noise is modeled as Gaussian noise with intensity dependent standard deviation

$$\sigma_{\text{noise}}(I) = \sqrt{\mathrm{K}I + \mathrm{K}^2\,\sigma_{\text{dark}}^2}. \tag{4}$$

K is a gain and σ_{dark} is the standard deviation of the dark noise [3]. (The values of K (9.83 DN24/e-) and σ_{dark} (0.54) for the Sony IMX490 sensor were taken from calibration data published by LUCID Vision Labs [31].)

Within each frame of the sequence, the *absolute value* of random Gaussian noise with standard deviation

$$\sigma_{\text{added}} = \sqrt{\left(R\,\sigma_{\text{noise}}(L/R)\right)^2 - \left(\sigma_{\text{noise}}(L)\right)^2} \tag{5}$$

is added to pixel values $I > L$. The absolute value is used because the stitching process generally produces RAWs slightly lighter on the high gain side and because large negative noise values introduces unrealistic fusion artifacts.

5. HDR Sensor and ISP Optimization

HDR sensor and ISP optimization is an ill-conditioned problem which involves discrete hardware registers and computationally expensive losses. The proposed method allows us to obtain perceptually pleasing images as well as images with optimal IoU scores when input into an object detector. See the Supplemental Document for a review of the loss functions used in this work. Like Mosleh *et al*. [44], we pose parameter selection as an optimization problem, but we also include sensor functionality and the downstream detector in the optimization problem. Relaxing integers as real numbers, we model an imaging pipeline ϕ that reconstructs trichromatic color images $\mathbf{O}$ from J multiplexed RAW exposures as

$$\phi : \mathbb{R}^{\mathrm{J}\times\mathrm{W}\times\mathrm{H}} \times \mathbb{R}^{\mathrm{P}}_{[0,1]} \to \mathbb{R}^{\mathrm{W}\times\mathrm{H}\times 3}, \quad (\mathbf{I}, \boldsymbol{\Theta}) \mapsto \mathbf{O}. \tag{6}$$

The transformation ϕ is modulated using P continuous hyperparameters $\boldsymbol{\Theta}$ on the sensor *and* ISP with the range of values normalized to the unit interval $\mathbb{R}_{[0,1]}$. Hardware registers are actually discrete, each with its own operational range, for example $\{0, 1\}$ for an algorithmic branch toggle and $\{0, \ldots, 2^{10}-1\}$ for a noise threshold [58] but they are relaxed to the continuum [44].

We frame HDR hyperparameter selection as a Multi-Objective Optimization (MOO) problem [33] with solutions

$$\boldsymbol{\Theta}^* = \underset{\boldsymbol{\Theta}\in\mathbb{R}^{\mathrm{P}}_{[0,1]}}{\operatorname{argmin}} \boldsymbol{\mathcal{L}}(\mathbf{s}(\boldsymbol{\Theta})) := \underset{\boldsymbol{\Theta}\in\mathbb{R}^{\mathrm{P}}_{[0,1]}}{\operatorname{argmin}} (\mathcal{L}_1(\mathbf{s}(\boldsymbol{\Theta})), \ldots, \mathcal{L}_{\mathrm{L}}(\mathbf{s}(\boldsymbol{\Theta}))), \tag{7}$$

where

$$\mathbf{s}(\boldsymbol{\Theta}) = (\phi(\mathbf{I}_1, \boldsymbol{\Theta}), \ldots, \phi(\mathbf{I}_{\mathrm{S}}, \boldsymbol{\Theta})) \tag{8}$$

is the *output image stack*, a collection of HDR captures processed by the sensor and ISP with the same hyperparameter setting $\boldsymbol{\Theta}$ but S different RAW image inputs from the *HDR input image stack* $\mathbf{I}_1, \ldots, \mathbf{I}_{\mathrm{S}}$. The *objective* is the loss vector $\boldsymbol{\mathcal{L}}(\mathbf{s}(\boldsymbol{\Theta}))$. Each of its L components is a loss measured on the output image stack [44]. Specifically, each end-to-end loss component $\mathcal{L}_l(\mathbf{s}(\boldsymbol{\Theta}))$ is derived from an evaluation metric calculated on the output images produced by the $\boldsymbol{\Theta}$-modulated sensor and ISP. These metrics may include downstream vision tasks or even human observers [48].

When a deep vision CNN is involved, the loss does not depend directly on the output image stack. It is then computed with an evaluation metric that quantifies the output of the downstream CNN

$$\boldsymbol{\mathcal{L}}(\mathbf{s}(\boldsymbol{\Theta})) = \boldsymbol{\mathcal{L}}(\mathrm{CNN}(\boldsymbol{\Omega}, \mathbf{s}(\boldsymbol{\Theta}))), \tag{9}$$

where the downstream image understanding CNN and its weights $\boldsymbol{\Omega}$ are shown instead of being folded into the loss.

The set of MOO solutions is the *Pareto front* [33]. MOO problems generically have multiple solutions. For example, a first optimal solution may make $\mathcal{L}_1$ better but $\mathcal{L}_2$ worse than another optimal solution, each solution manifesting a different tradeoff between conflicting objectives. Multimodality aside, there may be multiple solutions even with a single objective (L=1). For example, the mapping between $\boldsymbol{\Theta}$ and the output image stack $\mathbf{s}(\boldsymbol{\Theta})$ may have a nontrivial

Algorithm 1 ISP Hyperparameter Optimization Method.

Require: $\boldsymbol{\Theta} \in \mathbb{R}^{\mathrm{P}}_{[0,1]}$ (sensor + ISP hyperparameter vector),
$\sigma \in \mathbb{R}_{(0,\infty)}$ (CMA-ES covariance matrix scaling factor),
$\mathbf{C} \in \mathbb{R}^{\mathrm{P}\times\mathrm{P}}$ (CMA-ES "directional" cov. matrix factor),
$\varepsilon \in \mathbb{R}_{(0,\infty)}$ (small bound), $\mathrm{N} \in \mathbb{N}^*$ (number of iterations)

1: $\mathbf{p} \leftarrow \mathbf{0}$, $\mathbf{c} \leftarrow \mathbf{0}$ (CMA-ES path vectors)
2: **for** $n = 1$ **to** N **do**
3: symmetrize $\mathbf{C}$
4: **if** smallest $\mathbf{C}$ eigenvalue $< \varepsilon$ **then**
5: clamp eigenvalues up to ε
6: bring eigenvalues λ closer to 1 by replacing by $\lambda^{0.99}$
7: $\mathbf{p} \leftarrow \mathbf{0}$, $\mathbf{c} \leftarrow \mathbf{0}$
8: **end if**
9: **if** $\sigma < \varepsilon$ **or** $\sigma > 1/2$ **then**
10: $\sigma \leftarrow \mathbf{median}(\varepsilon, \sigma, 1/2)$, $\mathbf{p} \leftarrow \mathbf{0}$, $\mathbf{c} \leftarrow \mathbf{0}$
11: **end if**
12: **if** largest $\mathbf{C}$ eigenvalue $> 1/(2\sigma)$ **then**
13: clamp eigenvalues down to $1/(2\sigma)$, $\mathbf{p} \leftarrow \mathbf{0}$, $\mathbf{c} \leftarrow \mathbf{0}$
14: **else if** $\|\mathbf{p}\| >$ CMA-ES bound **then**
15: $\mathbf{p} \leftarrow \mathbf{0}$, $\mathbf{c} \leftarrow \mathbf{0}$
16: **end if**
17: **for** $p = 1$ **to** 2P **do**
18: $\boldsymbol{\Theta}_p^{(n)} \leftarrow$ draw from Gaussian at $\boldsymbol{\Theta}$ with cov. matrix $\sigma\mathbf{C}$
19: $\boldsymbol{\Theta}_p^{(n)} \leftarrow$ draw from Gaussian at $\boldsymbol{\Theta}_p^{(n)}$ with diagonal cov. matrix with entries proportional to quantization grain
20: reflect $\boldsymbol{\Theta}_p^{(n)}$ back into $\mathbb{R}^{\mathrm{P}}_{[0,1]}$ (mirroring boundaries)
21: $\mathbf{s}(\boldsymbol{\Theta}_p^{(n)}) \leftarrow$ run ISP on $\mathbf{I}_1, \ldots, \mathbf{I}_{\mathrm{S}}$ with settings $\boldsymbol{\Theta}_p^{(n)}$
22: $\boldsymbol{\mathcal{L}}(\mathbf{s}(\boldsymbol{\Theta}_p^{(n)})) \leftarrow$ loss evaluated on ISP output $\mathbf{s}(\boldsymbol{\Theta}_p^{(n)})$
23: **end for**
24: update $\boldsymbol{\Theta}$, σ, $\mathbf{C}$, $\mathbf{p}$, $\mathbf{c}$ based on the loss
25: **end for**
26: **return** $\boldsymbol{\Theta}_p^{(n)}$ with smallest $\boldsymbol{\mathcal{L}}(\mathbf{s}(\boldsymbol{\Theta}_p^{(n)}))$

kernel, meaning that different hyperparameter settings drive the ISP to produce identical output images and, therefore, losses. Such kernels should be disambiguated by reducing the number of search space degrees of freedom so that $\mathbf{s}$ is one-to-one, at least near candidates for optimality. Well-balanced training data decreases the likelihood that widely different parameter settings produce similar output image stacks. This being said, the proposed solver robustly handles some kernels. For example, one of the optimized Sony IMX490 sensor hyperparameters is a toggle that deactivates all the others, and optimization proceeded without a hitch.

The pipelines optimized in the present work do not allow "re-injection" of RAW captures. So, optimization uses ever changing input image stack instances. When optimizing for human viewing for example, new lab captures are acquired whenever a new $\boldsymbol{\Theta}$ is evaluated through its loss.

The 0^{th}-order solver Algorithm 1 used to optimize sensor and ISP hyperparameters is a variant of CMA-ES (Covariance Matrix Adaptation Evolution Strategy) [17, 23] in which Line 26 is disambiguated using Mosleh *et al.*'s max-rank loss scalarization [44] when performing MOO. Key differences with Mosleh *et al.* [44] are discussed below.

Algorithm 2 Joint Sensor, ISP and CNN Optimization Method.

Require: $\boldsymbol{\Omega} \in \mathbb{R}^{\mathrm{Q}}_{[0,\infty)}$ (CNN weight vector),
$\boldsymbol{\Theta} \in \mathbb{R}^{\mathrm{P}}_{[0,1]}$ (sensor + ISP hyperparameter vector),
$\mathrm{L} \in \mathbb{N}^*$ (number of joint optimization cycles),
$\mathrm{M} \in \mathbb{N}^*$ (Stochastic Gradient Descent iterations per cycle),
$\mathrm{N} \in \mathbb{N}^*$, $\sigma_0 \in \mathbb{R}_{(0,\infty)}$, $\mathbf{C}_0 \in \mathbb{R}^{\mathrm{P}\times\mathrm{P}}$, $\varepsilon \in \mathbb{R}_{(0,\infty)}$ (Algorithm 1),
$\eta \in \mathbb{R}_{(0,\infty)}$ (CNN training learning rate)

1: **for** $l = 1$ **to** L **do**
2: $\sigma \leftarrow \sigma_0$, $\mathbf{C} \leftarrow \mathbf{C}_0$
3: $\boldsymbol{\Theta} \leftarrow$ Algorithm 1 with loss $\boldsymbol{\mathcal{L}}$ evaluated with fixed $\boldsymbol{\Omega}$
4: **for** $m = 1$ **to** M **do**
5: $\boldsymbol{\Omega} \leftarrow \boldsymbol{\Omega} - \eta\nabla_{\boldsymbol{\Omega}}\mathcal{L}_m$ (Stochastic Gradient Descent iteration for loss $\mathcal{L}_m$ evaluated with fixed $\boldsymbol{\Theta}$)
6: **end for**
7: $\sigma_0 \leftarrow \sigma_0/2$, $\eta \leftarrow \eta/10$
8: **end for**
9: **return** $(\boldsymbol{\Omega}, \boldsymbol{\Theta})$

Hyperparameter values at the boundary of the usable range are valid candidates for optimality, even more so when ISP output is fed to downstream image understanding modules. Existing CMA-ES methods, when used with mirroring boundary conditions, are biased away from the boundary (other boundary conditions also have issues) [2, 17, 18, 30]. We constructed CMA-ES centroid weights such that boundary minima in regions where one parameter dominates loss variation are stable in expectation (when the covariance matrix is consistent), that is, if the so-called centroid $\boldsymbol{\Theta}$ is on that boundary, its update is statistically expected to stay there. These so-called active [23, 44]) boundary stabilizing weights have been empirically found to work best with a different generation size (2P vs. $4\mathrm{P}/3$ in [44]) and discarded trials proportion (none vs. worst ranked quarter). With no discard, the novel weights are obtained by assigning a weight of 1 to the best trial of a generation, $1-\sqrt{2}$ to the worst, interpolating linearly based on rank to get the other weights, and normalizing to a unit sum, see the Supplemental Document. Other improvements over Mosleh *et al.* include that path variables are reset whenever CMA-ES internals are seatbelted (Lines 3–16 of Algorithm 1) leading to more reliable improvements past coarse convergence, and that warm-starting was found to be unnecessary.

6. Joint Sensor, ISP and CNN Optimization

We *jointly* optimize sensor/ISP hyperparameters *and* image understanding CNN *weights*, framing joint HDR hyperparameter optimization as a MOO minimization problem with optimal solutions

$$(\boldsymbol{\Omega}^*, \boldsymbol{\Theta}^*) = \underset{\boldsymbol{\Omega}\in\mathbb{R}^{\mathrm{Q}}_{[0,\infty)}, \boldsymbol{\Theta}\in\mathbb{R}^{\mathrm{P}}_{[0,1]}}{\operatorname{argmin}} \boldsymbol{\mathcal{L}}(\mathrm{CNN}(\boldsymbol{\Omega}, \mathbf{s}(\boldsymbol{\Theta}))). \quad (10)$$

Joint optimization is performed with Algorithm 2. Block coordinate descent alternates between a 0^{th}-order optimizer that improves ISP hyperparameters $\boldsymbol{\Theta}$ keeping

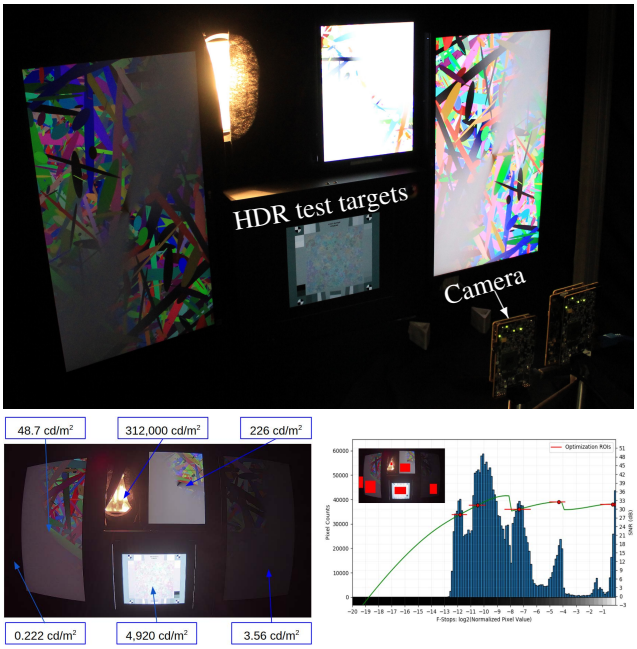

Figure 3: HDR lab setup covering 20.4 stops (123 dB). Top: overall view. Bottom: ISP output (left) and ROI (Region Of Interest) content distribution (right). Sensor covers 18.7 stops (112 dB) with SNR discontinuities at -4 and -8 stops.

Table 1: Test targets used in the HDR laboratory setup.

Test target / Source	Model / Type	Peak lum. (cd/m^2)
Printed chart	N/A	0.222
Display	LG 27UK600-W	3.56
Display	LiteMax SLO1568-ENB-I24	48.7
Display	Dell UP2718Q	226
Lightbox + Transmissive chart	IQL LED Lightbox 6500K + Imatest ISC-LED chart	4,920
Light source	Husky C 639975 with halogen Lamp	312,000

CNN weights fixed (Line 3), and Stochastic Gradient Descent [34] (Lines 4–6), a 1st-order optimizer that solves for optimal CNN weights $\boldsymbol{\Omega}$ keeping ISP hyperparameters fixed (the gradient of each loss component is taken with respect to CNN weights only). A partial input image stack is used by each of the two block steps. Sensors and ISPs process images locally; fewer training samples are needed to optimize them than to train CNN detectors performing non-local scene understanding. Also, acquiring a CNN training dataset for each of hundreds to thousands of sensor settings is not practical; a workaround is detailed in the Supplemental Document.

7. Assessment

7.1. HDR Optimization for Human Vision

We validate the method proposed in Sec. 5 with an ON Semiconductor AR0231AT sensor and AP0202AT HDR ISP. A two-stage approach is used to optimize hyperparameters efficiently and reproducibly. In the first stage, non HDR-specific ISP hyperparameters are optimized by minimizing the distance between the ISP output and a reference image, basically Mosleh *et al.* [44] except in the use of a novel image difference metric, Contrast Weighted Lp-Norm (CWLP), that uses Larkin's universal Noise Visibility Function [32] as a weight. Compared to expert-tuning and Mosleh *et al.*, the proposed method strikes a better balance between detail, noise and artifacts, especially at high gain.

Table 2: Human vision (perceptual) HDR optimization losses. Lower is better except for SNR (not used for optimization). Mean and worst values over all applicable ROIs. With respect to most metrics, the proposed method outperforms a combination of "linear-mode" optimization with [44] and expert-tuned HDR.

Loss	Mosleh *et al.* [44] + Expert-tuned	Proposed
FSITM mean	0.335	**0.333**
CWLP mean	4.083	**4.045**
Zippering mean = worst	**0.068**	0.071
SNR worst	23.60	**24.89**

In the second stage, we freeze all previously optimized hyperparameters except those associated with noise reduction, and also optimize adaptive local tone mapping. The lab setup with the reflective charts, displays and light sources listed in Table 1 is used. The dynamic range of this setup exceeds the camera's; see Fig. 3. Regions Of Interest (ROIs) are chosen to optimize detail preservation in the shadows and highlights, and LCD brightness was adjusted so that content straddles sensor SNR discontinuities (Fig. 3, bottom right). Three evaluation metrics were used: CWLP; Feature Similarity Index for Tone-Mapped images (FSITM) [45], an image difference metric that compares the 8-bit output with the full bit depth RAW; and Zippering, a semi-reference metric that quantifies structured noise [58]. See the Supplemental Document for additional details.

Seven illumination scenarios are cycled through by switching light sources and displays on and off, see Fig. 4. At the conclusion of the second stage, a small set of Pareto points taken from the latest iterations is analyzed visually and the setting with the best combination of contrast, details and low noise level throughout the luminance range is selected. Fig. 4 compares the output obtained with expert-tuned HDR hyperparameters together with"linear" hyperparameters optimized with the method of Mosleh *et al.* [44], with those obtained with the proposed method. As expected from comparing loss values (Table 2), the proposed method better preserves detail throughout the dynamic range.

Perceptual image quality is further evaluated using a second controlled lab setup and challenging field captures. The assessment lab setup consists of a light booth, light sources and several traffic signs, with several illumination scenarios, from very dark to very bright. Sample results are shown in Fig. 5 (left), please zoom into electronic version. The proposed method generally produces images with less noise and more contrast and detail; the Siemens star is more

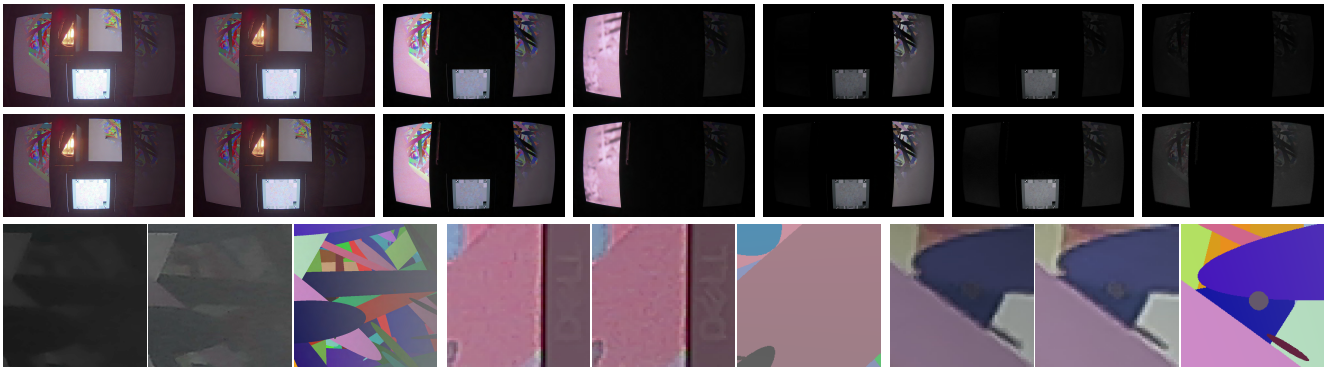

Figure 4: All seven scenarios used for HDR perceptual IQ optimization and corresponding ISP output. Top: Expert-tuned results. Middle: Outputs using the proposed method. Bottom: Zoom-ins where for each triple, the first enlargement shows a crop of one of the captures in the top row (expert-tuned), the second, optimized with the proposed method (from the second row), and the third, the corresponding area of the displayed chart (second triple also shows monitor logo outside of the chart). The proposed method preserves detail at all luminance levels. Gamma correction applied to facilitate crop visualization.

Figure 5: Results of ISP optimization for perceptual IQ. Left: HDR lab scene. Right: Real-life HDR scene. Top: Expert-tuned outputs. Bottom: Outputs of the proposed method. The proposed approach provides more detail and better dynamic range compression and local contrast. Please zoom into the electronic version of this document. Gamma correction applied to facilitate crop visualization.

clearly visible for example. With very bright light sources, the proposed method achieves better dynamic range compression by reducing artifacts in highlights; see the spotlight shining on the stop sign in the leftmost crop. Further assessment under challenging in-the-wild conditions confirmed that the proposed method preserves more contrast and detail. Sample results are shown in Fig. 5 (right). Expert-tuned HDR settings fail to preserve the crane's silhouette in the leftmost crop for example. Loss of detail is apparent elsewhere.

7.2. Joint HDR and CNN Optimization for Object Detection

We validate the method proposed in Sec. 6 with a Sony IMX490 sensor, a Renesas REN_AC_085 HDR ISP emulator, and the YOLOv4 [5] CNN for automotive object detection on the classes "pedestrian" and "car". For sensor and ISP optimization (Lines 2–3 of Algorithm 2), 40 groups of stitched and companded captures, each consisting of 254 consecutive frames sampling different combinations of sensor hyperparameter values, are randomly selected for each iteration, and the loss used by the 0^{th}-order optimizer is mAP with IoU$>$.5 measured on the output of YOLOv4. The same loss on a larger training dataset is used to train the CNN (Lines 4–6 of Algorithm 2). In all cases, 30% of the frames are augmented with emulated SNR drop artifacts (Sec. 6). The initial optimizer parameters σ is 0.25, the initial learning rate η is 10^{-4}, and there are 8128 training frames, 2032 validation frames and 565 test frames. See Supplemental Document for additional details.

Table 3: Joint ISP and CNN optimization object detection mAP and mAR scores. The proposed joint optimization method outperforms expert-tuned by 33% and [44] by 22% in mAP and mAR.

	mAP (IoU$>$0.5)	**mAP (IoU$>$0.75)**	**mAR**
Expert-tuned	0.250	0.244	0.235
Mosleh *et al*. [44]	0.367	0.356	0.352
Proposed (one iteration)	0.563	0.540	0.536
Proposed (converged)	**0.584**	**0.561**	**0.560**

As shown in Table 3, jointly optimized hyperparameters and CNN weights significantly outperform existing methods in both mAP and mAR, including expert-tuned (CNN fine-tuned for fairness) and Mosleh *et al*. [44] (extended to sensor and HDR hyperparameters). This results from joint optimization achieving better denoising and image compression throughout the 14-bit HDR range. As shown in Fig. 6–7, by preserving the local contrast in shadows (ex-

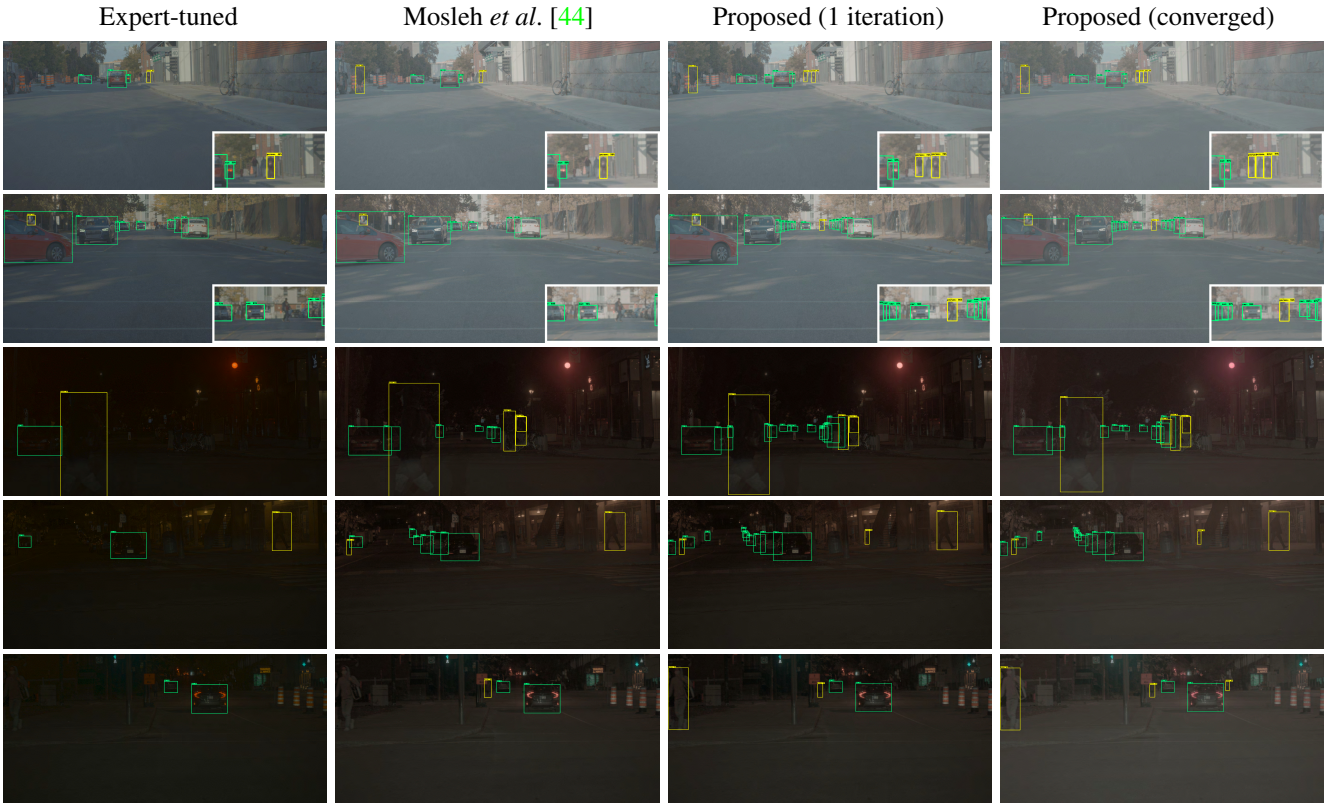

Figure 6: Joint sensor, ISP and CNN optimization for car and pedestrian detection. With higher contrast within lower bits, the proposed method outperforms expert-tuned and Mosleh *et al*. [44]. (Please zoom in for confidence scores and class predictions.)

panding the data in lower bits), the proposed method was able to significantly improve performance in low light conditions without loss of performance in high light conditions where, as a result of better denoising and detail preservation, the proposed method was able to detect significantly smaller objects. We note that this result is achieved purely by supervision using the downstream IoU loss without any additional image quality loss measured on intermediate ISP output images.

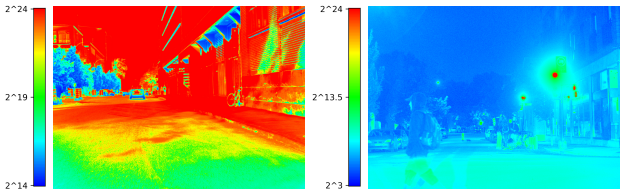

Figure 7: RAW luminance for the first and third rows of Fig. 6. Blue = minimum luminance ($21868 \approx 2^{14}$ for the first, $9 \approx 2^3$ for the third). Red = maximum luminance ($16513038 \approx 2^{24}$ for both). See the Supplemental Document for details [1].

8. Conclusions

We present an end-to-end optimization method that jointly learns optimal parameter values for an high dynamic range camera pipeline, both HDR sensor and hardware ISP parameters and downstream CNN weights of a vision module. Individual parameters are supervised only by downstream losses at the end of the pipeline—perceptual image quality losses for display, and an IoU loss for object detection—evaluated on captured training data. We jointly optimize network weights and ISP parameters with a block-coordinate descent method alternating between sensor and ISP optimization and CNN training. Because HDR imaging pipelines do not allow for gain separability like low dynamic range ones, optimization for human viewing is performed with a laboratory setup that cycles through challenging illumination conditions resulting in HDR multiplexing artifacts. As such artifacts are challenging to reproduce consistently outside the lab, we propose a method for simulating them in captured training data when optimizing for object detection. We validate the proposed method experimentally for human viewing and for 2D object detection with state-of-the-art automotive ISPs and sensors. Across all tasks considered in this paper, the proposed method outperforms existing methods, including manual expert tuning and existing optimization methods for low-dynamic range cameras.

Acknowledgments We thank Emmanuel Onzon, Doug Taylor and Jean-François Taillon for fruitful discussions.

References

[1] Ahmet Oğuz Akyüz and Osman Kaya. A proposed methodology for evaluating hdr false color maps. *ACM Trans. Appl. Percept.*, 14(1), July 2016. 8

[2] Jarosław Arabas, Adam Szczepankiewicz, and Tomasz Wroniak. Experimental comparison of methods to handle boundary constraints in differential evolution. In Robert Schaefer, Carlos Cotta, Joanna Kołodziej, and Günter Rudolph, editors, *Parallel Problem Solving from Nature, PPSN XI*, pages 411–420, Berlin, Heidelberg, 2010. Springer Berlin Heidelberg. 5

[3] European Machine Vision Association. EMVA standard 1288, standard for characterization of image sensors and cameras, Dec. 2016. www.emva.org. 4

[4] Francesco Banterle, Alessandro Artusi, Kurt Debattista, and Alan Chalmers. *Advanced High Dynamic Range Imaging, Second Edition*. A. K. Peters, Ltd., USA, 2nd edition, 2017. 2

[5] Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao. Yolov4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*, 2020. 7

[6] Michael S. Brown. Understanding the in-camera image processing pipeline for computer vision. *IEEE International Conference on Computer Vision (ICCV) - Tutorial*, 2019. 1, 2

[7] Mark Buckler, Suren Jayasuriya, and Adrian Sampson. Reconfiguring the imaging pipeline for computer vision. In *IEEE International Conference on Computer Vision (ICCV)*, pages 975–984, 2017. 2

[8] *IEEE Standard for Camera Phone Image Quality IEEE Std 1858-2016 (Incorporating IEEE Std 1858-2016/Cor 1-2017)*, May 2017. 2

[9] Paul E. Debevec and Jitendra Malik. Recovering high dynamic range radiance maps from photographs. In *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '97, page 369378, USA, 1997. ACM Press/Addison-Wesley Publishing Co. 1, 2, 3

[10] Brian M. Deegan. The effect of split pixel HDR image sensor technology on MTF measurements. In *Proc. SPIE 9023, Digital Photography X, 90230Z (7 March 2014)*, Mar. 2014. 2

[11] Steven Diamond, Vincent Sitzmann, Stephen P. Boyd, Gordon Wetzstein, and Felix Heide. Dirty pixels: Optimizing image classification architectures for raw sensor data. *CoRR*, abs/1701.06487, 2017. 2, 3

[12] Gabriel Eilertsen, Joel Kronander, Gyorgy Denes, Rafał K. Mantiuk, and Jonas Unger. Hdr image reconstruction from a single exposure using deep cnns. *ACM Trans. Graph.*, 36(6), Nov. 2017. 1, 2

[13] Konstantina Fotiadou, Grigorios Tsagkatakis, and Panagiotis Tsakalides. Snapshot high dynamic range imaging via sparse representations and feature learning. *IEEE Transactions on Multimedia*, 2019. 2

[14] Orazio Gallo, Natasha Gelfandz, Wei-Chao Chen, Marius Tico, and Kari Pulli. Artifact-free high dynamic range imaging. In *2009 IEEE International Conference on Computational Photography (ICCP)*, pages 1–7. IEEE, 2009. 1, 2

[15] Miguel Granados, Kwang In Kim, James Tompkin, and Christian Theobalt. Automatic noise modeling for ghost-free hdr reconstruction. *ACM Trans. Graph.*, 32:201:1–201:10, 2013. 2

[16] Michael D. Grossberg and Shree K. Nayar. High dynamic range from multiple images: Which exposures to combine. In *in Proc. ICCV Workshop on Color and Photometric Methods in Computer Vision (CPMCV)*, 2003. 2

[17] Nikolaus Hansen. The CMA evolution strategy: A tutorial. *CoRR*, abs/1604.00772, 2016. 2, 5

[18] Nikolaus Hansen, André S. P. Niederberger, Lino Guzzella, and Petros Koumoutsakos. A method for handling uncertainty in evolutionary optimization with an application to feedback control of combustion. *IEEE Trans. Evol. Comput.*, 13(1):180–197, 2009. 5

[19] Samuel W. Hasinoff, Frédo Durand, and William T. Freeman. Noise-optimal capture for high dynamic range photography. *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 553–560, 2010. 2

[20] Donald C. Hood and Marcia A. Finkelstein. *Sensitivity to light*, chapter 5, pages 5–1–5–66. New York: Wiley, 1986. 3

[21] Jun Hu, Orazio Gallo, Kari Pulli, and Xiaobai Sun. Hdr deghosting: How to deal with saturation? *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1163–1170, 2013. 2

[22] IEEE. *White Paper - IEEE P2020 Automotive Imaging*, 2018. 2

[23] Grahame A Jastrebski and Dirk V Arnold. Improving evolution strategies through active covariance matrix adaptation. In *2006 IEEE international conference on evolutionary computation*, pages 2814–2821. IEEE, 2006. 2, 5

[24] Nima Khademi Kalantari and Ravi Ramamoorthi. Deep high dynamic range imaging of dynamic scenes. *ACM Trans. Graph.*, 36:144:1–144:12, 2017. 2

[25] Nima Khademi Kalantari and Ravi Ramamoorthi. Deep hdr video from sequences with alternating exposures. *Comput. Graph. Forum*, 38:193–205, 2019. 2

[26] Nima Khademi Kalantari, Eli Shechtman, Connelly Barnes, Soheil Darabi, Dan B. Goldman, and Pradeep Sen. Patch-based high dynamic range video. *ACM Trans. Graph.*, 32:202:1–202:8, 2013. 2

[27] Sing Bing Kang, Matthew Uyttendaele, Simon A. J. Winder, and Richard Szeliski. High dynamic range video. *ACM Trans. Graph.*, 22:319–325, 2003.

[28] Erum Arif Khan, Ahmet Oguz Akyüz, and Erik Reinhard. Ghost removal in high dynamic range images. *2006 International Conference on Image Processing*, pages 2005–2008, 2006. 2

[29] Alexander Kirillov, Kaiming He, Ross Girshick, Carsten Rother, and Piotr Dollár. Panoptic segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9404–9413, 2019. 3

[30] Oliver Kramer. A review of constraint-handling techniques for evolution strategies. *Appl. Comput. Intell. Soft Comput.*, 2010:185063:1–185063:11, 2010. 5

[31] LUCID Vision Labs. Triton 5.4 mp model. https://thinklucid.com/product/triton-5-mp-imx490. Data sheet. Accessed: 2020-11-05. 4

[32] Kieran G. Larkin. Structural Similarity Index SSIMplified:

Is there really a simpler concept at the heart of image quality measurement? *CoRR*, abs/1503.06680, 2015. 2, 6

[33] Marco Laumanns, Lothar Thiele, Kalyanmoy Deb, and Eckart Zitzler. Combining convergence and diversity in evolutionary multiobjective optimization. *Evolutionary Computation*, 10(3):263–282, 2002. 4

[34] Yann LeCun, Léon Bottou, Genevieve B. Orr, and Klaus-Robert Müller. Efficient backprop. In Grégoire Montavon, Genevieve B. Orr, and Klaus-Robert Müller, editors, *Neural Networks: Tricks of the Trade - Second Edition*, volume 7700 of *Lecture Notes in Computer Science*, pages 9–48. Springer, 2012. 6

[35] Siyeong Lee, Gwon Hwan An, and Suk-Ju Kang. Deep chain hdri: Reconstructing a high dynamic range image from a single low dynamic range image. *IEEE Access*, 6:49913–49924, 2018. 2

[36] Siyeong Lee, Gwon Hwan An, and Suk-Ju Kang. Deep recursive hdri: Inverse tone mapping using generative adversarial networks. In *The European Conference on Computer Vision (ECCV)*, Sept. 2018. 2

[37] Ce Liu. *Beyond Pixels: Exploring New Representations and Applications for Motion Analysis*. PhD thesis, MIT, USA, 2009. 2

[38] Steve Mann and Rosalind W. Picard. Being 'undigital' with digital cameras: extending dynamic range by combining differently exposed pictures. In *Proceedings of IS&T*, 1994. 1, 2

[39] Rafa K. Mantiuk, Karol Myszkowski, and Hans-Peter Seidel. *High Dynamic Range Imaging*, pages 1–42. American Cancer Society, 2015. 2

[40] Demetris Marnerides, Thomas Bashford-Rogers, Jonathan Hatchett, and Kurt Debattista. Expandnet: A deep convolutional neural network for high dynamic range expansion from low dynamic range content. *CoRR*, abs/1803.02266, 2018. 2

[41] Tom Mertens, Jan Kautz, and Frank Van Reeth. Exposure fusion: A simple and practical alternative to high dynamic range photography. In *Computer graphics forum*, volume 28, pages 161–171. Wiley Online Library, 2009. 2

[42] Microsoft. *Skype & Lync Video Capture Specification*, 1.0 edition, Aug. 2013. Doc. No H100693. 2

[43] Microsoft. *Microsoft Teams Video Capture Specification*, 4.0 edition, Apr. 2019. 2

[44] Ali Mosleh, Avinash Sharma, Emmanuel Onzon, Fahim Mannan, Nicolas Robidoux, and Felix Heide. Hardware-in-the-loop end-to-end optimization of camera image processing pipelines. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. 1, 2, 4, 5, 6, 7, 8

[45] Hossein Ziaei Nafchi, Atena Shahkolaei, Reza Farrahi Moghaddam, and Mohamed Cheriet. FSITM: A feature similarity index for tone-mapped images. *CoRR*, abs/1704.05624, 2017. 6

[46] Jun Nishimura, Timo Gerasimow, Rao Sushma, Aleksandar Sutic, Chyuan-Tyng Wu, and Gilad Michael. Automatic ISP image quality tuning using nonlinear optimization. In *IEEE International Conference on Image Processing (ICIP)*, pages 2471–2475, 2018. 1, 2

[47] Rafael Padilla, Sergio L. Netto, and Eduardo A. B. da Silva. A survey on performance metrics for object-detection algorithms. In *2020 International Conference on Systems, Signals and Image Processing (IWSSIP)*, pages 237–242, 2020. 2, 3

[48] Jonathan B. Phillips and Henrik Eliasson. *Camera Image Quality Benchmarking*. Wiley Publishing, 1st edition, 2018. 2, 4

[49] Geoffrey Portelli and Denis Pallez. Image signal processor parameter tuning with surrogate-assisted particle swarm optimization. In Lhassane Idoumghar, Pierrick Legrand, Arnaud Liefooghe, Evelyne Lutton, Nicolas Monmarché, and Marc Schoenauer, editors, *Artificial Evolution - 14th International Conference, Évolution Artificielle, EA 2019, Mulhouse, France, October 29-30, 2019, Revised Selected Papers*, volume 12052 of *Lecture Notes in Computer Science*, pages 28–41. Springer, 2019. 1, 2

[50] Ana Radonjić, Sarah R. Allred, Alan L. Gilchrist, and David H. Brainard. The dynamic range of human lightness perception. *Curr Biol.*, 21(22):1931–1936, Nov. 2011. 1

[51] Erik Reinhard, Greg Ward, Sumanta N. Pattanaik, Paul E. Debevec, and Wolfgang Heidrich. *High Dynamic Range Imaging - Acquisition, Display, and Image-Based Lighting (2. ed.)*. Academic Press, 2010. 2, 3

[52] Ulrich Seger. Hdr imaging in automotive applications. In *High Dynamic Range Video*, pages 477–498. Elsevier, 2016. 2

[53] Pradeep Sen, Nima Khademi Kalantari, Maziar Yaesoubi, Soheil Darabi, Dan B. Goldman, and Eli Shechtman. Robust patch-based hdr reconstruction of dynamic scenes. *ACM Trans. Graph.*, 31:203:1–203:11, 2012. 2

[54] Ana Serrano, Felix Heide, Diego Gutierrez, Gordon Wetzstein, and Belen Masia. Convolutional sparse coding for high dynamic range imaging. *Computer Graphics Forum*, 35(2):153–163, 2016. 2

[55] Isao Takayanagi, Norio Yoshimura, Kazuya Mori, Shinichiro Matsuo, Shunsuke Tanaka, Hirofumi Abe, Naoto Yasuda, Kenichiro Ishikawa, Shunsuke Okura, Shinji Ohsawa, and Toshinori Otaka. An over 90 dB intra-scene single-exposure dynamic range CMOS image sensor using a 3.0 micron triple-gain pixel fabricated in a standard BSI process. *Sensors (Basel, Switzerland)*, 18, 01 2018. 2

[56] Eino-Ville Talvala, Andrew Adams, Mark Horowitz, and Marc Levoy. Veiling glare in high dynamic range imaging. *ACM Transactions on Graphics (TOG)*, 26(3):37–es, 2007. 3

[57] Shunsuke Tanaka, Toshinori Otaka, Kazuya Mori, Norio Yoshimura, Shinichiro Matsuo, Hirofumi Abe, Naoto Yasuda, Kenichiro Ishikawa, Shunsuke Okura, Shinji Ohsawa, Takahiro Akutsu, Ken Fu, Ho-Ching Chien, Kenny Liu, Alex Tsai, Stephen Chen, Leo Teng, and Isao Takayanagi. Single exposure type wide dynamic range CMOS image sensor with enhanced nir sensitivity. *ITE Transactions on Media Technology and Applications*, 6:195–201, 07 2018. 2

[58] Ethan Tseng, Felix Yu, Yuting Yang, Fahim Mannan, Karl St. Arnaud, Derek Nowrouzezahrai, Jean-François Lalonde, and Felix Heide. Hyperparameter optimization in black-box image processing using differentiable proxies. *ACM Transactions on Graphics (SIGGRAPH)*, 38(4):27, 2019. 1, 2, 4, 6

[59] Jonas Unger, Francesco Banterle, Gabriel Eilertsen, and

Rafal Mantiuk. The hdr-video pipeline. In *Eurographics 2016 Tutorials*. Eurographics Association, May 2016. 2

[60] Trygve Willassen, Johannes Solhusvik, Robert Johansson, Sohrab Yaghmai, Howard Rhodes, Sohei Manabe, Duli Mao, Zhiqiang Lin, Dajiang Yang, Orkun Cellek, et al. A 1280× 1080 4.2 μm split-diode pixel HDR sensor in 110 nm BSI CMOS process. In *Proceedings of the International Image Sensor Workshop, Vaals, The Netherlands*, pages 8–11, 2015. 1

[61] Chyuan-Tyng Wu, Leo F. Isikdogan, Sushma Rao, Bhavin Nayak, Timo Gerasimow, Aleksandar Sutic, Liron Ainkedem, and Gilad Michael. VisionISP: Repurposing the image signal processor for computer vision applications. In *IEEE International Conference on Image Processing (ICIP)*, pages 4624–4628, 2019. 2

[62] Dietmar Wüller and Ulla Bøgvad Kejser. Standardization of image quality analysis ISO 19264. In *Archiving Conference*, 2016, pages 111–116. Society for Imaging Science and Technology, Apr. 2016. 2

[63] Lucie Yahiaoui, Jonathan Horgan, Brian Deegan, Senthil Yogamani, Ciarán Hughes, and Patrick Denny. Overview and empirical analysis of isp parameter tuning for visual perception in autonomous driving. *J. Imaging*, 5(10):78, 2019. 2

[64] Lucie Yahiaoui, Jonathan Horgan, Senthil Yogamani, Ciaran Hughes, and Brian Deegan. Impact analysis and tuning strategies for camera image signal processing parameters in computer vision. In *Irish Machine Vision and Image Processing conference (IMVIP)*, 2011. 2

[65] Lucie Yahiaoui, Ciarán Hughes, Jonathan Horgan, Brian Deegan, Patrick Denny, and Senthil Yogamani. Optimization of ISP parameters for object detection algorithms. *Electronic Imaging*, 2019(15):44–1, 2019. 2