End-to-end High Dynamic Range Camera Pipeline Optimization
Supplemental Document

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

In this supplemental document, we provide additional information and experimental results in support of the main document with the same title.

Contents

Table 1: ON Semiconductor AP0202AT ISP hyperparameters optimized at low gain, for human vision on an LCD, with their operational ranges. Hyperparameters are discrete; they are relaxed to continuous values in the optimization process. (Color Correction Matrix (CCM) coefficients are optimized only so that other ISP components be optimized with a correct CCM. Final values were discarded.)

Aperture correction

Hyperparameter	Operational range
ap gain (low gain)	$\{0, \ldots, 15\}$
ap thresh (low gain)	$\{0, \ldots, 255\}$

Global Tone Mapping

Hyperparameter	Operational range
contrast gradient (low gain)	$\{0, \ldots, 255\}$
contrast intercept point (low gain)	$\{0, \ldots, 255\}$

Demosaicking

Hyperparameter	Operational range
demosaic (low gain)	$\{0, \ldots, 255\}$

Color Correction Matrix*

Hyperparameter	Operational range
ccm_1	$\{0, \ldots, 65535\}$ but narrowed to $\{32743, \ldots, 32943\}$
ccm_2	$\{0, \ldots, 65535\}$ but narrowed to $\{32685, \ldots, 32885\}$
ccm_3	$\{0, \ldots, 65535\}$ but narrowed to $\{32644, \ldots, 32844\}$
ccm_5	$\{0, \ldots, 65535\}$ but narrowed to $\{32720, \ldots, 32920\}$
ccm_6	$\{0, \ldots, 65535\}$ but narrowed to $\{32678, \ldots, 32878\}$
ccm_7	$\{0, \ldots, 65535\}$ but narrowed to $\{32699, \ldots, 32899\}$

1. Overview

Sec. 2 details the sensor and ISP hyperparameters optimized in this work. In Sec. 2.1, we provide the non-HDR hyperparameters of the ON Semiconductor AP0202AT ISP optimized for perceptual IQ (Sec. 7.1 of the main document). In Sec. 2.2, we describe the HDR AP0202AT ISP hyperparameters optimized in the second stage (also Sec. 7.1 of the main document). In Sec. 2.3, we discuss the hyperparameters of the Sony IMX490 sensor and Renesas REN_AC_085 HDR ISP emulator used for the assessment of the proposed method.

In Sec. 3, we list the end-to-end loss functions used in this work to optimize for human viewing (Sec. 7.1 of the main document) and for object detection and classification evaluation metrics (Sec. 7.2 of the main document). This includes the novel "perceptual" image difference metric CWLP (Contrast Weighted Lp-norm). Evaluation metrics that are measured but not used for optimization are also listed.

In Sec. 4, details are provided regarding the non-HDR ("linear") first stage of ISP hyperparameter optimization for perceptual IQ (Sec. 7.1 of the main document); sample first stage results are shown.

In Sec. 5, additional details are provided regarding the second stage of ISP hyperparameter optimization for perceptual IQ (Sec. 7.1 of the main document).

In Sec. 6, the novel boundary-stabilizing CMA-ES centroid weights mentioned in Sec. 5 of the main document are derived.

Sec. 7 shows the effect of the SNR-drop emulation used to augment image detection and classification training sets. This HDR fusion simulation method is described in Sec. 4 and used in Sec. 7.2 of the main document.

Sec. 8 details the method used to acquire automotive field data generated with a representative sample of sensor hyperparameter values, and we explain how we work around acquiring sufficient training data for each hyperparameter settings or returning to the field every time the block coordinate descent method described in Algorithm 2 of Sec. 6 of the main document enters CNN weight training. This is used in Sec. 7.2 of the main document.

Sec. 9 describes the false color rendering of HDR RAW used in Fig. 7 of the main document.

Finally, Sec. 10 provides additional sample object detection and classification results (Sec. 7.2 of the main document).

2. Description of Optimized Hyperparameters

2.1. Non-HDR ISP Hyperparameters Optimized for Perceptual IQ

The ON Semiconductor AP0202AT ISP is a state-of-the-art automotive ISP targeting high dynamic range (HDR) sensors. In this work, it is paired with an ON Semiconductor AR0231AT imaging sensor. We keep sensor hyperparameters at their vendor-optimized values when optimizing for perceptual IQ; only ISP hyperparameters are modified.

The AP0202AT ISP is vendor-optimized to provide "full auto-functions support" [1]; its hyperparameter space is considerably smaller than other ISPs. Unlike other systems, the AP0202AT ISP modulation tables are obtained by interpolating between exactly two threshold values of modulation parameters, a low threshold and a high threshold, under which and above which dependent hyperparameter values are constant. This configuration, compatible with gain-based divide-and-conquer tuning (*of non-HDR hyperparameters*), simplifies hyperparameter selection. We optimize ISP hyperparameter values at low gain, and separately at high gain. Table 1 lists the ON Semiconductor AP0202AT ISP hyperparameters optimized at low

Table 2: ON Semiconductor AP0202AT ISP hyperparameters optimized at high gain, for human vision on an LCD, with their operational ranges. Hyperparameters are discrete; they are relaxed to continuous values in the optimization process.

Aperture correction

Hyperparameter	Operational range
ap gain (high gain)	$\{0, \ldots, 15\}$
ap thresh (high gain)	$\{0, \ldots, 255\}$

Global tone mapping

Hyperparameter	Operational range
contrast gradient (high gain)	$\{0, \ldots, 255\}$
contrast intercept point (high gain)	$\{0, \ldots, 255\}$

Demosaicking

Hyperparameter	Operational range
demosaic (high gain)	$\{0, \ldots, 255\}$

Color processing

Hyperparameter	Operational range
saturation (high gain)	$\{0, \ldots, 255\}$ but narrowed to $\{0, \ldots, 127\}$

gain; Table 2 lists those optimized at high gain.

The default value of the 'saturation' hyperparameter at low gain (which we use) is 127. Given that color saturation should decrease with gain, we narrow the search range of the high gain 'saturation' hyperparameter to $\{0, \ldots, 127\}$ (see Table 2).

CCM (Color Correction Matrix) coefficients (see Table 1) are only optimized so that, at convergence, other ISP components work with well-adapted values. Final ("optimal") CCM values are discarded after optimization. From the chosen six CCM degrees of freedom, namely the six off-diagonal values of the 3×3 CCM matrix, the remaining, diagonal, matrix values ('ccm_0', 'ccm_4' and 'ccm_8') are derived using the usual constraint that rows sum to 1. (This constraint assumes that the CCM is used by the ISP to process white balanced image data, as is the case; prior to optimization, white balance is performed separately at low and high gain using a simple method equalizing RGB color values within four middle grey chart patches.) CCM coefficients' narrowed search ranges are centered on the vendor-optimized settings for the laboratory rig's low gain capture conditions.

Except for CCM coefficients, hyperparameter values obtained in the first, non-HDR ("linear"), optimization stage (Sec. 4) are used in the second, HDR hyperparameter optimization stage, during which they are fixed to the values obtained in the first stage. See Sec. 7.1 of the main document.

2.2. HDR ISP Hyperparameters Optimized for Perceptual IQ

As indicated in Sec. 7.1 of the main document and in the last section, the same hardware, namely ON Semiconductor AP0202AT ISP paired with ON Semiconductor AR0231AT imaging sensor, is further optimized for human viewing on an LCD. In this second stage, only HDR ISP hyperparameters are optimized. Because, when optimizing in HDR mode, multiple illumination scenarios are used at the same time within the optimization loop, hyperparameter values associated with different gains are optimized simultaneously.

In HDR mode, the ON Semiconductor AP0202AT ISP has a brightness statistic which is the independent variable of some of the modulation tables; hyperparameter values associated with the high and low threshold values of this statistic are labeled with "bright" and "dark". Also, some hyperparameters have a value used with low bit values, and another with high bit values, labeled "(low)" and "(high)".

Several search ranges are narrowed by trimming off lowest or highest values when they have the same effect as a value kept within the range. This was determined from the ISP documentation as well as empirically. For example, there is little difference between image results obtained with 'acacd gr weights strength' values in the range $\{7, \ldots, 15\}$ and those obtained with the value 6. A less interesting example is 'altm sharpness strength': although in principle it has a 16-bit range, this hyperparameter acts as if the ISP simply cycles through the first 64 values; higher values are simply redundant and consequently are removed from the search range. In summary, HDR optimization is performed with "sane" hyperparameter search ranges, as described in Sec. 5 of the main document.

2.3. HDR Sensor and ISP Hyperparameters Optimized for Object Detection and Classification

The Renesas REN_AC_085 ISP is a state-of-the-art automotive ISP targeting HDR based on the ARM Mali ISP [3]. As discussed in Sec. 7.2 of the main document, it is paired with the state-of-the-art Sony IMX490 HDR automotive imaging sensor [2]. Jointly with downstream CNN detector weights, both the sensor and the ISP are optimized for object detection and classification.

41 hyperparameters, namely 6 sensor and 35 ISP hyperparameters, are optimized. Optimized sensor components include on-sensor noise reduction, motion compensation, and motion-driven false color correction. Optimized ISP components include denoising, local tone mapping, demosaicking (edge detection, edge enhancement, false color correction), color correction and global tone mapping. Noise profile calibration and white balance are performed before optimization. Table 4 lists

Table 3: ON Semiconductor AP0202AT ISP hyperparameters optimized in HDR mode, with their operational ranges. Hyperparameters are discrete; they are relaxed to continuous values in the optimization process.

Denoising (adaptive color difference)

Hyperparameter	Operational range
adacd gr weights strength (low)	$\{0, \ldots, 15\}$ but narrowed to $\{1, \ldots, 6\}$
adacd gr weights strength (high)	$\{0, \ldots, 15\}$ but narrowed to $\{1, \ldots, 6\}$

Adaptive local tone mapping

Hyperparameter	Operational range
altm key k_0	$\{0, \ldots, 65535\}$ but narrowed to $\{20, \ldots, 400\}$
altm gamma bright (low)	$\{0, \ldots, 65535\}$ but narrowed to $\{0, \ldots, 128\}$
altm gamma bright (high)	$\{0, \ldots, 65535\}$ but narrowed to $\{0, \ldots, 256\}$
altm gamma dark (low)	$\{0, \ldots, 65535\}$ but narrowed to $\{0, \ldots, 128\}$
altm gamma dark (high)	$\{0, \ldots, 65535\}$ but narrowed to $\{0, \ldots, 256\}$
altm sharpness strength (bright)	$\{0, \ldots, 65535\}$ but narrowed to $\{0, \ldots, 63\}$
altm sharpness strength (dark)	$\{0, \ldots, 65535\}$ but narrowed to $\{0, \ldots, 63\}$

the Sony IMX490 sensor and Renesas REN_AC_085 ISP hyperparameters modified in our experiments, with their operational search ranges.

The six CCM coefficients 'coef a 12' up to 'coef a 32' are, like in Sec. 2.1, the off-diagonal ones, and the "missing" diagonal entries of the 3×3 matrix, namely 'coef a 11', 'coef a 22' and 'coef a 33', are automatically generated so that the CCM has unit row sums. Search ranges for the off-diagonal coefficients are restricted to values that correspond to $[-1, 0.5]$ once converted to floating point values, range that produces diagonally dominant CCM matrices (unless possibly multiple entries reach 0.5) with the row sum 1 constraint, and also found empirically to be sufficient for most sensors and illuminants when mapping RGB to RGB (in extreme situations, values below -1 may be needed).

ISP threshold hyperparameters ('thresh 1' and 'thresh 4') which can have different values horizontally and vertically were set to be equal. (In the future, they may be decoupled given that patterned noise does not appear to be the same with respect to the two directions with every sensor and ISP combination.)

So as to obtain a reasonably sized field captures dataset with a fixed set of sampled values, given that sensors do not support RAW injection (only some ISPs) so that all sensor processing needs to be done at the capture time, sensor hyperparameters that parameterize false color correction are sampled at only 5 carefully chosen values. In addition, two of the hyperparameters ('cf1' and 'cf2') are set to be equal and consequently only sampled with equal values, given that they were found empirically to impact the output image in approximately the same way.

IMX490 Sensor hyperparameter values impact the output in strongly coupled ways. For example, if the 'noise reduction' toggle is 0, none of the other hyperparameters have any effect. Similarly, if the 'motion compensation' toggle is 0, the sensor hyperparameters that follow it in the list have no effect; likewise for 'false color correction'. In all, 254 combinations of sensor hyperparameters were sampled. Before the usual affine mapping normalizing values to the continuous interval $\mathbb{R}_{[0,1]}$, the values $\{512, 1448, 4096, 11585, 32767\}$ of the 'cf' sensor hyperparameters were mapped to $\{0, 1, 2, 3, 4\}$ so as to equalize the sizes of the continuous basins of attraction of the five sampled values as "seen" by the optimizer. A more detailed discussion of the sampling of 254 combinations of sensor hyperparameters is found in Sec. 8.

3. End-To-End Loss Functions

Unlike evaluation metrics used to compare *different* systems [7, 11, 16, 17, 21, 23], the evaluation metrics used in this work are directly used as losses. Specifically, we optimize image sensor and ISP hyperparameters for human viewing on an 8-bit display (for *perceptual Image Quality (IQ)*). We also optimize image sensor and ISP hyperparameters as well as the weights of a downstream deep CNN so that the entire pipeline, from sensor to ISP all the way to the CNN prediction, maximize object detection and classification performance. This sections details the end-to-end loss functions used for each purpose.

3.1. Perceptual Image Quality (IQ) Losses

Several loss functions are used to optimize ISP hyperparameters for human vision [18, 22]. These losses drive the optimizer toward image characteristics that human viewers find desirable like detail preservation and contrast, and away from undesirable ones like structured noise and jagged or smeared edges.

We use the following pixel loss ℓ_i is computed for every output image pixel i within the Region of Interest (far enough

Table 4: Sony IMX490 Sensor and Renesas REN_AC_085 ISP hyperparameters and their operational ranges. Hyperparameters are discrete; they are relaxed to continuous values in the optimization process.

Component	Hyperparameter	Operational Range
Sensor	noise reduction (RNR)	$\{0, 1\}$
	motion compensation (MDET)	$\{0, 1\}$
	false color correction (UP_ON)	$\{0, 1\}$
	cf0	$\{0, \ldots, 32767\}$ but restricted to $\{512, 1448, 4096, 11585, 32767\}$
	cf1 = cf2	$\{0, \ldots, 32767\}$ but restricted to $\{512, 1448, 4096, 11585, 32767\}$
	cf3	$\{0, \ldots, 32767\}$ but restricted to $\{512, 1448, 4096, 11585, 32767\}$
ISP: Denoiser	thresh 1h = thresh 1v	$\{0, \ldots, 255\}$
	thresh 4h = thresh 4v	$\{0, \ldots, 255\}$
	strength 1	$\{0, \ldots, 255\}$
	strength 4	$\{0, \ldots, 255\}$
	thresh long	$\{0, \ldots, 255\}$
	be power 0	$\{0, \ldots, 1000\}$
	be gamma	$\{0, \ldots, 250\}$
ISP: Local Tone Mapper	asymmetry power	$\{0, \ldots, 100\}$
	slope min	$\{0, \ldots, 255\}$
	slope max	$\{0, \ldots, 255\}$
	variance space	$\{0, \ldots, 15\}$
	variance intensity	$\{0, \ldots, 15\}$
ISP: Demosaicking	vh slope	$\{0, \ldots, 255\}$
	vh thresh	$\{0, \ldots, 10000\}$
	va slope	$\{0, \ldots, 255\}$
	va thresh	$\{0, \ldots, 10000\}$
	aa slope	$\{0, \ldots, 255\}$
	aa thresh	$\{0, \ldots, 10000\}$
	uu slope	$\{0, \ldots, 255\}$
	uu thresh	$\{0, \ldots, 10000\}$
	sharp alt ld	$\{0, \ldots, 255\}$
	sharp alt lu	$\{0, \ldots, 255\}$
	sharp alt ldu	$\{0, \ldots, 255\}$
	fc alias slope	$\{0, \ldots, 255\}$
	fc alias thresh	$\{0, \ldots, 255\}$
	fc slope	$\{0, \ldots, 255\}$
	np offset	$\{0, \ldots, 255\}$
ISP: Color Space Conversion	coef a 12	$\{0, \ldots, 5880\}$ but restricted to $\{3839, \ldots, 4223\}$
	coef a 13	$\{0, \ldots, 5880\}$ but restricted to $\{3839, \ldots, 4223\}$
	coef a 21	$\{0, \ldots, 5880\}$ but restricted to $\{3839, \ldots, 4223\}$
	coef a 13	$\{0, \ldots, 5880\}$ but restricted to $\{3839, \ldots, 4223\}$
	coef a 31	$\{0, \ldots, 5880\}$ but restricted to $\{3839, \ldots, 4223\}$
	coef a 32	$\{0, \ldots, 5880\}$ but restricted to $\{3839, \ldots, 4223\}$
ISP: Global Tone Mapper	mu	$\{0, \ldots, 20000\}$
	gamma	$\{0, \ldots, 1000\}$

from its boundary). The ℓ_is (assumed ≥ 0) are then aggregated with a normalized Lp-norm into an overall loss

$$\mathcal{L} = \sqrt[p]{\frac{1}{\sum_i 1} \sum_i \ell_i^p}, \tag{1}$$

the denominator being the number of pixels over which the loss is averaged. With $p = 1$, this construction is used for all of this work's IQ losses.

Reduced-reference metrics [6] quantify the occurrence of targeted artifacts within ROIs that have suitable visual content; one such evaluation metric, Zippering, is used in this work. The full-reference image difference metrics used in this work compare the output image to a higher bit-depth version (FSITM), or to an aligned and enhanced rendering of the chart captured by the camera, the *aligned reference* (CWLP).

3.1.1 Feature Similarity Index for Tone-Mapped images (FSITM)

Because LCDs generally have a lower bit-depth than HDR sensors, Tone Mapping Operators (TMOs) are often used to lower image bit-depth (e.g., from 20- or 14-bit to 8-bit). FSITM, proposed by Nafchi *et al.* [19] for TMO assessment, uses the HDR

RAW as reference; the lower bit-depth output of the ISP is compared to it using the locally weighted phase angle, a robust noise-independent measure that quantifies whether RAW details are preserved. These per pixel values are aggregated with (1) and converted to a dissimilarity index by subtracting from 1.

In this work, 20-bit decompanded demosaicked linear RGB was used as reference by FSITM.

3.1.2 Contrast-Weighted Lp-Norm (CWLP)

CWLP is a novel image difference metric. It is the only evaluation metric used to drive the optimizer in both stages of optimization for human viewing (Sec. 7.1 of the main document).

In this work, the only CWLP Lp-norm involved is the L1-norm (Lp with p=1) by way of the corresponding Minkowski distance, namely the average of the absolute differences of the pixel values of the ISP output and the reference image across all RGB channels.

CWLP is a convex combination (linear combination with non-negative weights) of an *unweighted* L1-norm of the differences between RGB value vectors, used to make the global tone map of the ISP output similar to the tone map of the reference [18, 22, 24], and of a *weighted* L1-norm targeting the detail layer. CWLP's per-pixel weights use a pair of five-pixel masks, a "+" and an "×", suited for the optimization of demosaicking, sharpening and denoising modules that operate near Nyquist. The contribution of one channel and one mask to the weight of one pixel's RGB differences is the square of Larkin's universal Noise Visibility Function DQ [14]:

$$\text{weight contribution} = \frac{\sigma_-^2 + \varepsilon}{\sigma_-^2 + \sigma_+^2 + 2\varepsilon}, \tag{2}$$

where σ_- (resp. σ_+) is the standard deviation of the channel value differences (resp. sums) between the two images over the mask, and ε is a small positive constant. Weights obtained with the "+" and "×" masks are aggregated by convex combination with weights inverse proportional to the diameter of the mask, and then averaged across channels. This per-pixel weight multiplies the L1-norm of the central pixel's RGB difference vector, which is then aggregated with (1) before combination with the unweighted L1-norm. Finally, the result is converted to log-scale after adding 1 so that image differences scale roughly like bits.

When optimizing HDR hyperparameters (the second stage of the procedure described in Sec. 7.1 of the main document), before the aligned reference and ISP output images are compared, output image ROIs are individually scaled (in linear RGB) so that a central estimate related to the trimean matches the reference's. Following normalization, pixel value differences and variances were then computed in gamma 3 RGB so as to penalize shadow differences more than highlight differences.

Normalization was not performed when optimizing non-HDR ("linear") ISP hyperparameters (the first stage of the procedure described in Sec. 7.1 of the main document; also see Sec. 4). When optimizing "linear-mode" ISP hyperparameters, pixel value differences and variances were computed in gamma 2.2 RGB.

3.1.3 Zippering

The Zippering loss [22] quantifies near-Nyquist structured noise that corrupts smooth lines, interfaces and regions. It can be used to detect zippering, an artifact that consists of contrasting single pixels jutting in and out of an interface like a zipper. Zippering loss ROIs should only contain smooth-edged features like hyperbolic wedges; in this work, the ROIs are smooth (they are featureless). After scaling so that the capture has the same mean as the chart within the ROI, per channel, per pixel maximins are obtained as follows: Compute, within a 3×3 mask, the distance between the *closest* pixel values within a 4-pixel "T" and the pixel values of its complement within the 3×3, a 5-pixel "U", *when the values are separated*, that is, when all the values within one of the masks are larger than those within the other (if not separated, assign 0); then take the maximum over the four orientations. That is: Over one image channel o and one orientation ("upright", as opposed to "rotated left", "rotated right", or "upside down"), compute

$$\max\left(\max(0, \min_{i\in\mathrm{T}} o_i - \max_{i\in\mathrm{U}} o_i), \max(0, \min_{i\in\mathrm{U}} o_i - \max_{i\in\mathrm{T}} o_i)\right), \tag{3}$$

and average over all orientations and channels before aggregating with (1). Following normalization, pixel value differences are computed in linear RGB (gamma 1).

3.1.4 Additional Reported Evaluation Metric (Not Used for Optimization): SNR

When optimizing for perceptual IQ, Signal-to-Noise Ratio (SNR) was monitored using using flat patches of the light box Imatest chart (see Table 1 and Fig. 3 of the main articles). It was computed as follows: Convert the sRGB output image to linear RGB with sRGB primaries. From the linear RGB values, compute a luminance Y using Rec. 709 coefficients

$$Y = 0.2126R + 0.7152G + 0.0722B. \tag{4}$$

The SNR of a (flat) patch is then

$$\text{SNR} = 20 \log_{10} \frac{\text{average of } Y \text{ within the patch}}{\text{standard deviation of } Y \text{ within the patch}}. \tag{5}$$

3.2. Object Detection and Classification Losses

The following losses are used in the assessment of the proposed method for image understanding presented in Sec. 7.2 of the main document.

3.2.1 Mean Average Precision with IoU > 0.5

In this work, the downstream image understanding module returns a collection of bounding boxes together with a classification of the content into one of the search classes. With the result of hand annotation treated as ground truth, Mean Average Precision (mAP) with IoU (Intersection over Union) > 0.5 is used as evaluation metric [15, 20]. (With mAP, higher is better. mAP is turned into a loss by subtracting it from 1.) Obtaining the mAP score corresponding to a sensor and ISP hyperparameter setting is computationally expensive because a meaningful score requires pushing a large number of captures through the sensor, ISP and the downstream CV module, and because meaningful mAP scores require large numbers of objects to available for detection.

3.2.2 Additional Reported Evaluation Metrics (Not Used for Optimization): mAP with IoU > 0.75 and mAR

The values of mAP with IoU > 0.75 and mAR (mean Average Recall [20]) are also reported in the main article; they are not used to optimize.

4. "Linear" (Non-HDR) Hyperparameter Optimization for Perceptual Image Quality

A two-stage optimization procedure is used to optimize the hyperparameters of an ISP for perceptual image quality efficiently and reproducibly (see Sec. 7.1 of the main article). In this section, the first stage ("linear", that is, non-HDR) is described in detail.

Following Mosleh *et al*. [18], we minimize the distance between the ISP output image and a reference image, namely the random ellipses ROI found in the central region of the aligned and resampled Rainbow Chart. The full resolution Rainbow Chart, without additional tone mapping and sharpening, is displayed on a ViewSonic VP2780-4K display colorimetrically calibrated to the white point of CIE Standard Illuminant D65. The camera and LCD are enclosed inside a dark optimization box. This setup, with Rainbow Chart displayed, is shown in Fig. 1. (A variant of the same Rainbow Chart was also used by Tseng *et al*. [22].) The proposed method differs from that of Mosleh *et al*. in the use of a novel image difference metric, Contrast Weighted Lp-Norm (CWLP) (see Sec. 3.1.2). The resampled, so as to be aligned, Rainbow Chart is tone-mapped with a sigmoid-like curve to boost global contrast and, when optimizing at low gain, it is sharpened by unsharp masking (in gamma 2.2, using a method that preserves color lines, and without a threshold since there is no noise) to raise artificial acutance. The modified aligned reference is used as reference in the image difference metric; the ISP is consequently driven by the optimizer to reproduce the perceptual image quality attributes of the enhanced aligned reference as they manifest themselves within the random ellipses ROI.

Optimization is performed separately for two scenarios: bright (low gain) and dark (high gain). With the ON Semiconductor AP0202AT ISP, this is enough to fill a modulation table (see Sec. 2.1). Bright conditions were achieved by setting the display peak luminance to 435 cd/m^2, whereas for dark conditions luminance went down to 1 cd/m^2. For each scenarios, an exposure suitable for optimization exposure was found by finding the minimizer of the CWLP loss (computed from the output of the ISP configured with vendor-optimized values) within an exposure sweep. In addition, white balance gains were found by equalizing the green-to-red and green-to-blue channel ratios of the four mid-gray patches of the Rainbow Chart

Figure 1: Non-HDR ("linear") ISP optimization laboratory rig with Rainbow Chart displayed on the LCD. During optimization, the camera (attached to a tripod head) and LCD are isolated from their surroundings by an opaque black curtain.

Table 5: Perceptual IQ loss values at the conclusion of non-HDR ("linear") optimization. Lower is better.

Loss	Low gain		High gain	
	Expert-tuned	Optimized	Expert-tuned	Optimized
CWLP	4.314	**4.230**	4.758	**4.721**

with a simple search algorithm based on the reciprocal of each ratio to update the corresponding white balance gain. The resulting exposure times and white balance coefficients were kept fixed during optimization.

Sec. 2.1 specifies the ISP hyperparameters that were optimized with each scenario. Unless there are ties (and, consequently, a kernel), an optimization run with respect to one single loss (CWLP, here) produces one single optimal solution. One single solution is indeed obtained for each scenario.

Table 5 shows the values of the CWLP loss measured on the Rainbow Chart random ellipses ROI of the output image generated using expert-tuned ISP hyperparameters on the one hand, and ISP hyperparameters optimized with the proposed method on the other, exactly as the loss is computed while optimizing. As shown in Fig. 2 and 3, the proposed method generally produces images with more contrast and detail, better colors and less noise, in particular, less structured noise.

5. Additional Details on HDR ISP Hyperparameter Optimization for Human Viewing

In the second stage of the procedure described in Sec. 7.1 of the main article, FSITM was evaluated on an ROI encompassing all the targets within the scene (including the halogen lamp and the small paper chart), normalized CWLP was evaluated on the three LCDs and the light box (when turned on), and the Zippering loss was only measured on the Dell display in one of the scenarios, scenario in which the Dell LCD shows a smooth, featureless greyscale chart. A small paper chart, printed with a small, dark version of the random ellipses chart displayed on the LCDs, was placed in the shadows (behind the light

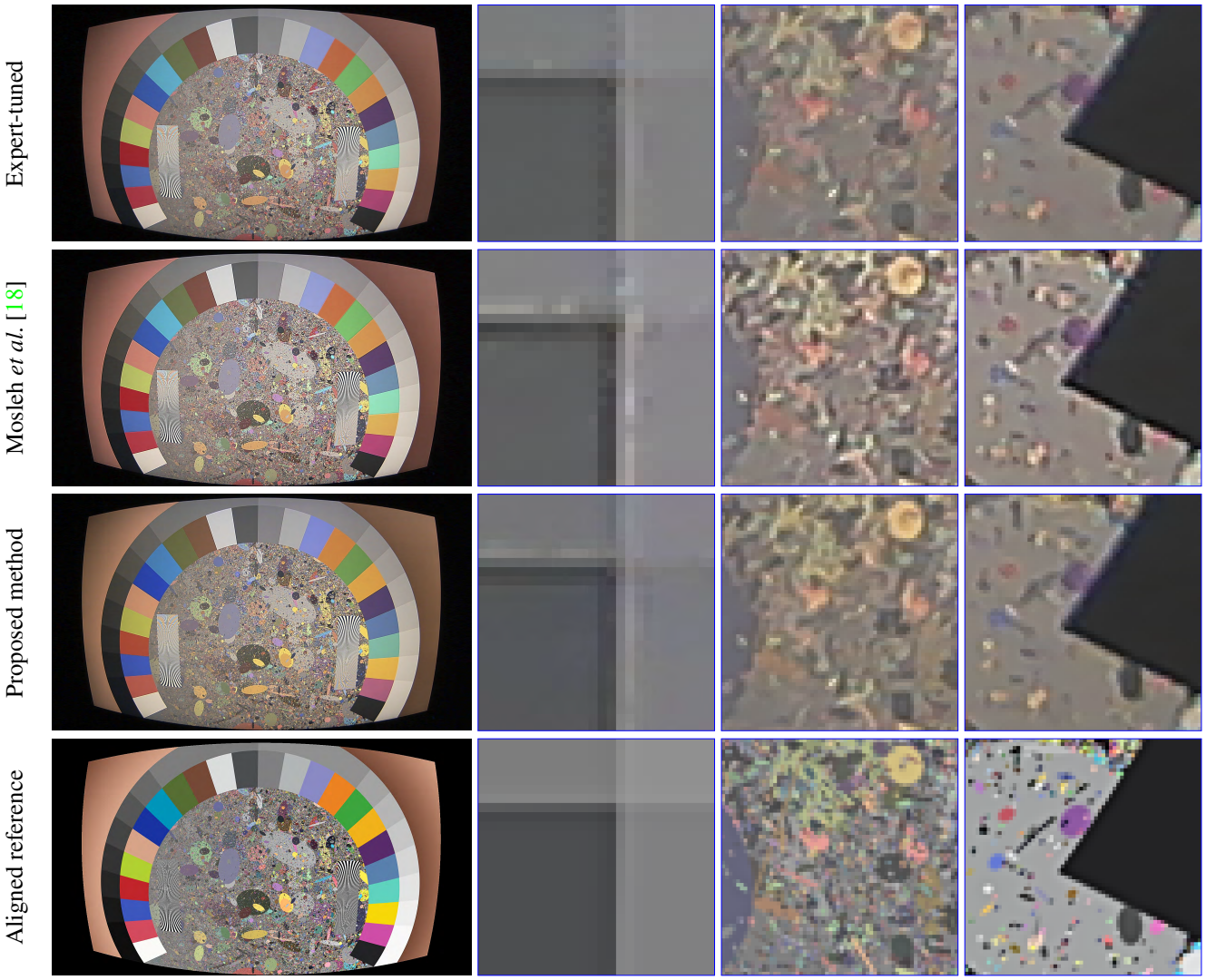

Figure 2: Non-HDR ("linear") low gain optimization results. Images processed with ISP settings obtained with the proposed method are nearly free of the zippering artifact (first crop) unlike those expert-tuned and, especially, those obtained by Mosleh *et al*. The proposed method also has the cleanest interfaces among the alternatives (last crop). Compared to the expert-tuned result, the proposed method has more fine detail and less artifacts (compare with the aligned reference, at the bottom). Although the method of Mosleh *et al*. has more contrast and detail, it also has considerably more artifacts and some of the details do not reflect the aligned reference. Zoom into the full size capture to view other areas. Note: Loss based solely on the random ellipses area; the rest of the aligned reference is rendered to match.

sources); this paper chart was only "seen" by FSITM.

Displays were colorimetrically calibrated to the CIE Standard Illuminant D65 white point. The light box illuminant was set to D65. As in the first stage, the rig was enclosed within a dark optimization box (see Fig. 3 of the main article). The white balance gains used in the first stage were also used in the second.

Charts were designed to include a combination of flat regions and variable-contrast texture patterns using random ellipses generated so that a non-clipping central estimator related to the trimean be equal in the random ellipse area and in the flats.

The nineteen CWLP losses (each from a different combination of a scenario and an LCD or light box), seven FSITM losses (one per scenario) and one Zippering loss were aggregated with the weighted max-rank loss [18], with larger weights given to the CWLP losses corresponding to darker ROIs within a scene and to the Zippering and FSITM losses.

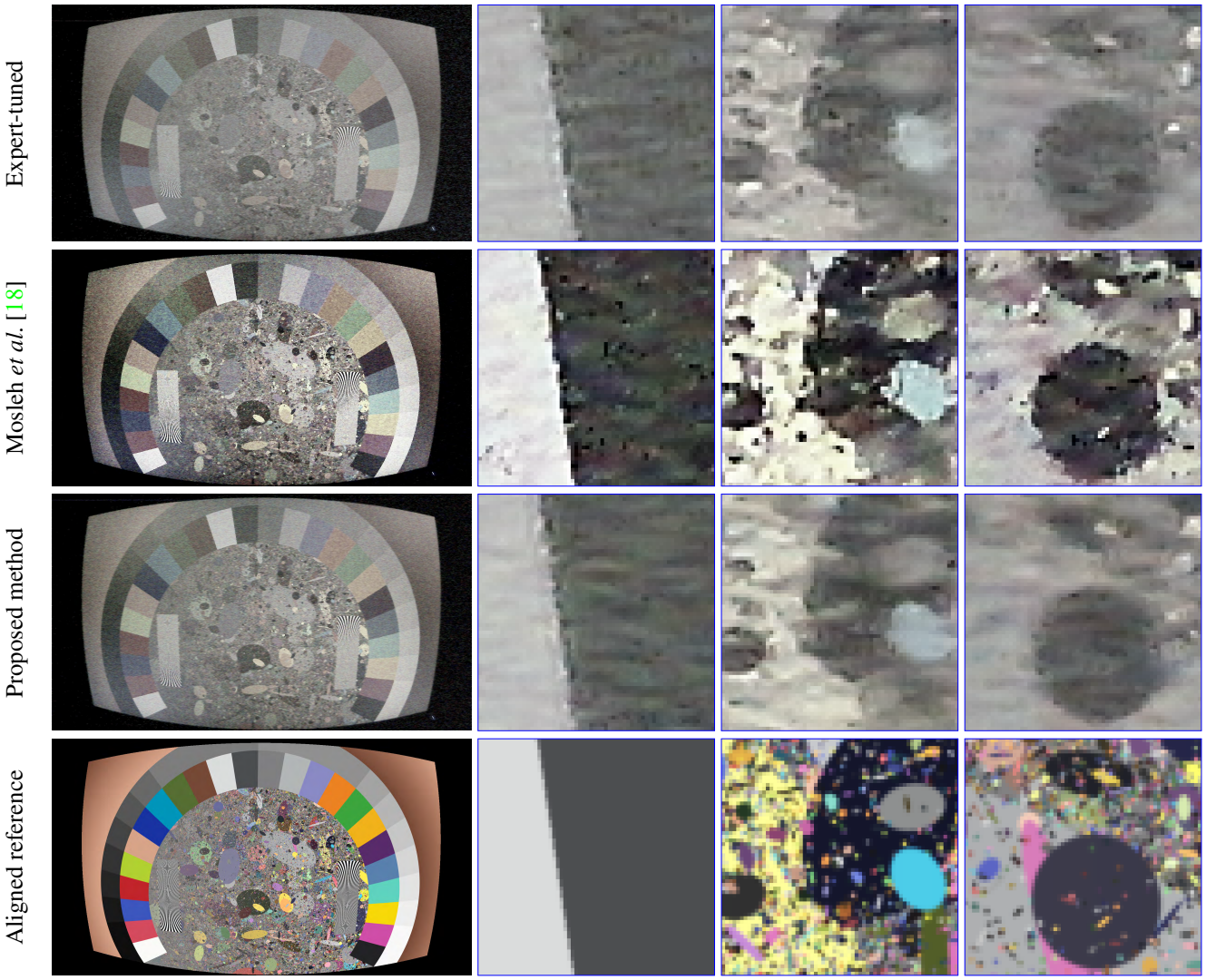

Figure 3: Non-HDR ("linear") high gain optimization results. Images processed with ISP settings obtained with the proposed method have the fewest artifacts and show faithful details. The proposed results show significantly fewer vertical ripples and have cleaner interfaces than the expert-tuned ones. Although the Mosleh *et al.* results have more detail and contrast, they contain a lot more spurious details and artifacts. (Compare with the aligned reference (bottom row).) Zoom into the full size capture to view other areas. Note: Loss based solely on the random ellipses area; the rest of the aligned reference is rendered to match.

6. Boundary-Stabilizing CMA-ES Centroid Weights

The most significant difference between the CMA-ES variant used as 0^{th}-order solver by the proposed sensor and ISP optimization method is, as mentioned in the Introduction and Sec. 5 of the main article, novel CMA-ES centroid weights that stabilize boundary minima when search domain boundaries are dealt with by mirroring.

The most commonly used variants of CMA-ES [9] set the next generation's centroid $\boldsymbol{\Theta}$ to be a weighted linear combination, with fixed weights (unless there are ties, in which case weights may be redistributed; we will ignore ties in the following), of the current generation's hyperparameter vectors

$$\boldsymbol{\Theta} = \sum_p w_p(\mathrm{rank}_p)\boldsymbol{\Theta}_p, \tag{6}$$

where rank_p is the rank of $\boldsymbol{\Theta}_p$ within the generation, rank based on the loss value: If, for example, $\boldsymbol{\Theta}_p$ has the very best loss

value within its generation, $\mathrm{rank}_p = 0$; if it has, for example, the very worst rank of its generation, and generations have λ samples (in Algorithm 1 of the main article, $\lambda = 2\mathrm{P}$, where P is the number of hyperparameters being optimized), then $\mathrm{rank}_p = \lambda - 1$. (Ranks could start at 1 instead of 0. Starting at 0 is convenient when using the max-rank loss [18].)

CMA-ES draws the next generation's trials using a Gaussian distribution centered at the current value of $\mathbf{\Theta}$, fixed within the generation. (This Gaussian distribution is modified with the so-called covariance matrix $\sigma\mathrm{C}$; again, we will ignore this complication in the following.) The mirroring boundary condition [5, 9, 10, 13], reflects randomly generated $\mathbf{\Theta}_p$ hyperparameter vectors that fall outside of the search domain back inside. When optimizing within an hypercube, the novel weights are such a minimum located at the boundary of the hypercube are stable in the following (weak) sense.

Definition *Rank-based centroid weights* $w_p(\mathrm{rank}_p)$ *are* weakly boundary stable *(with respect to the mirroring boundary condition) if whenever the loss function strictly monotonically depends only on one hyperparameter, and the current centroid is a minimum located at the boundary, the (statistical) expectation* $\mathbf{E}(\mathbf{\Theta})$ *of the updated centroid is also on that boundary provided it is the only boundary about which reflection occurs as the result of using mirroring boundary conditions.*

The rather strong separability assumption allows to reduce the issue of stability to 1D. Without loss of generality we can also assume that the boundary is located at 0, and that the domain is $\mathbb{R}_{[0,1]}$.

No method which only uses positive weights can be weakly boundary stable. This is seen in that mirroring a Gaussian distribution centered at 0 back into the positive real axis generically yields samples which are strictly positive, and no linear combination of positive numbers with positive weights *ever* equals 0. Leaving aside pathological methods in which all the weights are zero, this means that only *active* CMA-ES methods, that is, methods that use both positive and negative weights [12], are candidates for weak boundary stability.

First, let us assume an affine functional form for the active rank-based weighting function

$$w_p(\mathrm{rank}_p) = \frac{1}{s}\left(1 - \frac{m}{\lambda - 1}\,\mathrm{rank}_p\right), \tag{7}$$

where λ is the number of drawn samples per generation and s is chosen so that the sum of the weights is 1 as befits weighted averaging, which leaves one degree of freedom, the slope factor m. (The following construction applies to other functional forms. Another is discussed below.)

Remembering that mirroring a normal distribution about its center results in a half-normal (a.k.a. folded normal a.k.a. positive normal) distribution, weak boundary stability boils down to the statistical expectation of the centroid update vanishing, that is,

$$\mathbf{E}\left(\sum_{p=1}^{\lambda} w_p(\mathrm{rank}_p)\Theta_p\right) = 0, \tag{8}$$

where Θ_p are λ samples drawn from the half-normal distribution, and rank_p is the natural ordering with respect to the (positive) real axis. Monte-Carlo simulation rather quickly leads one to conjecture that $m = \sqrt{2}$ for every value of λ. This conjecture gets independent numerical validation by translating the problem into one involving the chi order statistics of the half-normal and using the numerically derived tables published in [8].

In summary, when all the drawn hyperparameter vectors get a weight, the weakly boundary stabilizing weights are obtained by giving a rank of 1 to the best ranked vector, a rank of $1 - \sqrt{2}$ to the worst rank, and interpolating linearly between these weights based on rank. These weights were used in this work.

Most CMA-ES methods *discard* the worst (with respect to the loss) drawn hyperparameter vectors. In other words, some proportion of the hyperparameter vectors of one generation, the worst ones, get a weight of 0, and only the better ranked hyperparameter vectors get a nonzero weight in the centroid update. The most common discard proportion is half. Given that μ is often used for the number of non-discarded hyperparameter vectors in CMA-ES methods, this means that $\mu = \lambda/2$ (for λ even). In this situation, Monte-Carlo simulation suggests that the weakly boundary stabilizing value of m for

$$w_p(\mathrm{rank}_p) = \begin{cases} \frac{1}{s}\left(1 - \frac{m}{\mu - 1}\,\mathrm{rank}_p\right) & \text{when } \mathrm{rank}_p < \mu, \text{ and} \\ 0 & \text{when } \mu \leq \mathrm{rank}_p\,. \end{cases} \tag{9}$$

is approximately equal to 1.489135. This, again, is numerically confirmed by [8].

The exact functional form of the dependence of the weight on the rank is not important. What matters is that it be monotone decreasing—as it should—and smooth as a function of the ranks of the non-discarded hyperparameter vectors, and then Monte Carlo simulation can be used to obtain weakly boundary stabilizing weights.

Figure 4: SNR-drop emulation example. Top: Fully processed from unmodified stitched RAW. Bottom: Fully processed from stitched RAW augmented with emulated SNR-drop artifacts. Texture changes are most obvious in the pavement about one third of the way from the bottom of the image and in the rear window of the car closest to the camera.

7. SNR-Drop Artifact Emulation

The emulation of fusion artifacts for the purpose of augmenting field data is justified and described in Sec. 4 of the main article; it is used in Sec. 7.2.

Fig 4 shows an example of what the augmentation does to a capture. Even fully processed by the ISP (instead of in the RAW), the artifacts look realistic.

8. One-Time Acquisition of Field Data Without RAW Injection

The methods described in this section were used in Sec. 7.2 of the main article.

The sensors used in this work do not allow for RAW injection of pre-captured RAW data, i.e. emulation of a live sensor feed after the actual capture process. Consequently, the optimization of sensor hyperparameters requires the acquisition of new captures every time sensor hyperparameter values are changed. In-the-field automotive field data, ideally, should only be acquired once. Thus, representative data, suitable for both optimization and training, should be acquired the first time around.

Algorithm 1 Sensor Hyperparameter Sweep Method.

value $\leftarrow [512, 1448, 4096, 11585, 32767]$
while driving **do**
  CF3_up $\leftarrow$ **false**
  CF0_up $\leftarrow$ **false**
  CF12_up $\leftarrow$ **false**
  RNR $\leftarrow 1$
  MDET $\leftarrow 1$
  UP_ON $\leftarrow 1$
  for $i = 1$ **to** 4 **do**
    if CF3_up **then**
      CF3 $\leftarrow$ value$[i]$
    else
      CF3 $\leftarrow$ value$[4 - i]$
    end if
    for $j = 0$ **to** 4 **do**
      if CF0_up **then**
        CF0 $\leftarrow$ value$[j]$
      else
        CF0 $\leftarrow$ value$[4 - j]$
      end if
      for $k = 0$ **to** 4 **do**
        if CF12_up **then**
          CF1 $\leftarrow$ value$[k]$
          CF2 $\leftarrow$ value$[k]$
        else
          CF1 $\leftarrow$ value$[4 - k]$
          CF2 $\leftarrow$ value$[4 - k]$
        end if
        Take capture with current hyperparameter combination
      end for
      CF12_up $\leftarrow$ **not** CF12_up (Flip sweep direction)
    end for
    CF0_up $\leftarrow$ **not** CF0_up (Flip sweep direction)
  end for
  CF3_up $\leftarrow$ **not** CF3_up (Flip sweep direction)
  UP_ON $\leftarrow 0$
  Take capture with current hyperparameter combination
  MDET $\leftarrow 0$
  Take capture with current hyperparameter combination
  RNR $\leftarrow 0$
  Take capture with current hyperparameter combination
  MDET $\leftarrow 1$
  Take capture with current hyperparameter combination
  UP_ON $\leftarrow 1$
  Redo above for $i = 1$ up to 4 loop
  CF3_up $\leftarrow$ **not** CF3_up (Flip sweep direction)
end while

If one samples too many sensor hyperparameter combinations, or if one samples them randomly, the time elapsed between captures made with similar hyperparameters, let alone very different ones, will be such that scene changes may dominate loss function variations. In order to minimize this risk, 254 representative sensor hyperparameter settings were sampled (see Table 4) while the data acquisition car was capturing field data samples.

8.1. In-Field Acquisition of Sensor-Processed RAWs With Variable Settings

Past coarse (initial) convergence, the 0^{th}-order optimizer compares nearby hyperparameter settings. It consequently makes sense to sample nearby hyperparameter values as close to each other as possible in time when in the field, so that scene

variations not overwhelm the loss differences that arise from changing hyperparameter values. Taking into account that some hyperparameters are toggles that deactivate other hyperparameters, the sampling method shown in Algorithm 1 modifies exactly one single hyperparameter value *that has an effect* from one capture to the next *and* covers the 254 chosen sensor hyperparameter combinations within every sequence of consecutive 254 captures.

8.2. Reuse of Sensor-Modulated Field Data for CNN Training

The following procedure for selecting RAWs from a collection that samples sensor settings takes into account that we optimize sensor hyperparameters which are toggles that deactivate the others, so that that large subsets of the unit hypercube $\mathbb{R}^{\mathrm{P}}_{[0,1]}$ have the exact same effect. So, aggregate hyperparameter combinations by partitioning $\mathbb{R}^{\mathrm{P}}_{[0,1]}$ with respect to the equivalence relation defined by having the same effect on-sensor. Because equivalence classes are subsets of Euclidean space, they inherit its distance function. To train the CNN weights for a fixed $\boldsymbol{\Theta}$, select the RAWs associated with the equivalence classes closest, with respect to the Euclidean distance, to $\boldsymbol{\Theta}$'s equivalence class. From this pool of RAWs acquired with settings close to those currently considered optimal, random samples can be extracted for each iteration of stochastic gradient descent. Note that because the ISP which further processes sensor output accepts injected data, output images resulting from modulated ISP hyperparameters values are computed on demand.

9. False Color HDR Rendering

In the main document, false color HDR renderings of two of the stitched RAW used to generate the results of Fig. 6 are shown in Fig. 7. Following [4], they are obtained by mapping L (specifically the average of the four linear decompanded RAW pixel values of 2×2 Bayer cells) to the HSV hue, by way of $\mathrm{H} = 240° \frac{\log(1+\mathrm{L_{max}})-\log(1+\mathrm{L})}{\log(1+\mathrm{L_{max}})-\log(1+\mathrm{L_{min}})}$, so that $\mathrm{L_{min}}$ is mapped to 240° (blue) and $\mathrm{L_{max}}$ to 0° (red).

10. Additional Automotive Object Detection Results

In this section, additional sample results of the joint sensor, ISP and deep CNN optimization are shown; a few are found in Sec. 7.2 of the main article. As shown in Figure 5 and 6, the proposed joint optimization outperforms expert-tuned pipelines and Mosleh *et al*. [18] in low light conditions which tend to be harsher and noisy. Even in high light conditions where the data tends to be less noisy, the proposed method improves slightly on already well performing baselines. This improvement in performance can be attributed to better signal quality, lower noise and improved data compression in 14-bit HDR output of the ISP, which when processed with the downstream YOLOv4 model resulted in improved detections in harsher lighting conditions and detections of smaller (distant) objects.

References

[1] AP0202AT high-dynamic range (HDR) image signal processor (ISP). https://www.mouser.com/datasheet/2/308/AP0202AT-D-932936.pdf. Accessed: 2020-03-30. 2

[2] IMX490, SONY's CMOS sensor for automotive. https://www.sony-semicon.co.jp/products/common/pdf/IMX490.pdf. Accessed: 2020-11-22. 3

[3] MALI CAMERA C71 image signal processing for automotive. https://www.arm.com/products/silicon-ip-multimedia/image-signal-processor/mali-c71. Accessed: 2020-11-11. 3

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

[5] 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. 11

[6] Shahrukh Athar and Zhou Wang. A comprehensive performance evaluation of image quality assessment algorithms. *IEEE Access*, 7:140030–140070, 2019. 5

[7] *IEEE Standard for Camera Phone Image Quality IEEE Std 1858-2016 (Incorporating IEEE Std 1858-2016/Cor 1-2017)*, May 2017. 4

[8] Zakkula Govindarajulu and Stan Eisenstat. Best estimates of location and scale parameters of a chi (1 df) distribution, using ordered observations. *Reports of statistical application research, Union of Japanese Scientists and Engineers*, 12:149–164, 1965. 11

[9] Nikolaus Hansen. The CMA evolution strategy: A tutorial. *CoRR*, abs/1604.00772, 2016. 10, 11

[10] 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. 11

[11] IEEE. *White Paper - IEEE P2020 Automotive Imaging*, 2018. 4

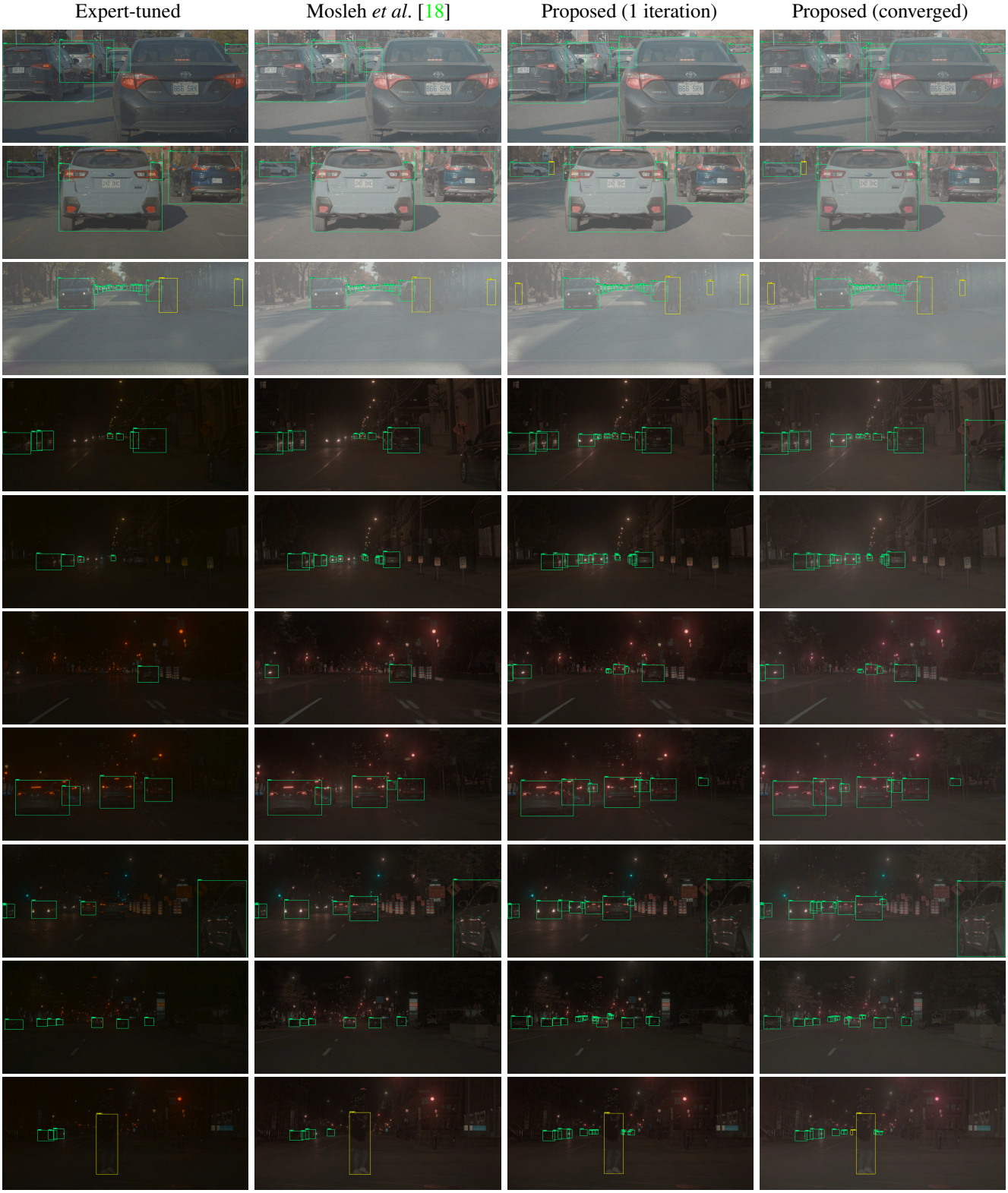

Figure 5: Joint sensor, ISP and CNN optimization for car and pedestrian detection with an automotive camera. By raising contrast within the 14-bit ISP output, the proposed method significantly outperforms both expert-tuned pipelines and Mosleh *et al*. [18]. (Please zoom into the electronic version of this document for confidence scores and class predictions.)

[12] Grahame A Jastrebski and Dirk V Arnold. Improving evolution strategies through active covariance matrix adaptation. In *2006 IEEE*

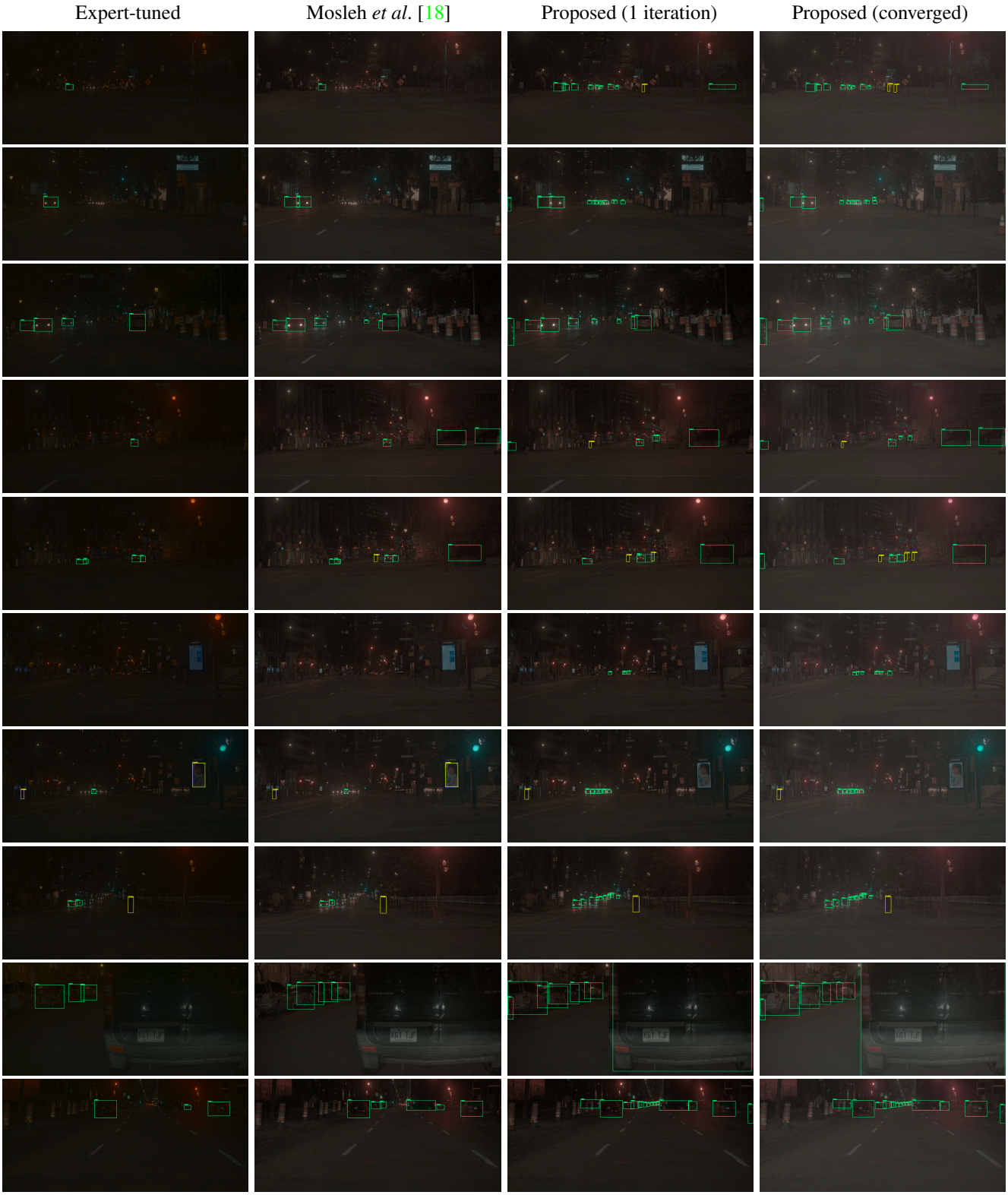

Figure 6: Joint sensor, ISP and CNN optimization for car and pedestrian detection with an automotive camera. By raising contrast within the 14-bit ISP output, the proposed method significantly outperforms both expert-tuned pipelines and Mosleh *et al*. [18]. (Please zoom into the electronic version of this document for confidence scores and class predictions.)

international conference on evolutionary computation, pages 2814–2821. IEEE, 2006. 11

[13] Oliver Kramer. A review of constraint-handling techniques for evolution strategies. *Appl. Comput. Intell. Soft Comput.*,

2010:185063:1–185063:11, 2010. 11

[14] 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. 6

[15] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft COCO: Common objects in context. In *European Conference on Computer Vision (ECCV)*, pages 740–755, 2014. 7

[16] Microsoft. *Skype & Lync Video Capture Specification*, 1.0 edition, Aug. 2013. Doc. No H100693. 4

[17] Microsoft. *Microsoft Teams Video Capture Specification*, 4.0 edition, Apr. 2019. 4

[18] 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. 4, 6, 7, 9, 10, 11, 14, 15, 16

[19] 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. 5

[20] 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. 7

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

[22] 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. 4, 6, 7

[23] 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. 4

[24] Hang Zhao, Orazio Gallo, Iuri Frosio, and Jan Kautz. Loss functions for image restoration with neural networks. *IEEE Transactions on Computational Imaging*, 3(1):47–57, 2017. 6