Spatially Varying Exposure with 2-by-2 Multiplexing: Optimality and Universality

Xiangyu Qu, Student Member, IEEE, Yiheng Chi, Student Member, IEEE, and Stanley H. Chan, Senior Member, IEEE

Abstract—The advancement of new digital image sensors has enabled the design of exposure multiplexing schemes where a single image capture can have multiple exposures and conversion gains in an interlaced format, similar to that of a Bayer color filter array. In this paper, we ask the question of how to design such multiplexing schemes for *adaptive* high-dynamic range (HDR) imaging where the multiplexing scheme can be updated according to the scenes. We present two new findings.

(i) We address the problem of *design optimality*. We show that given a multiplex pattern, the conventional optimality criteria based on the input/output-referred signal-to-noise ratio (SNR) of the independently measured pixels can lead to flawed decisions because it cannot encapsulate the location of the saturated pixels. We overcome the issue by proposing a new concept known as the spatially varying exposure risk (SVE-Risk) which is a pseudoidealistic quantification of the amount of recoverable pixels. We present an efficient enumeration algorithm to select the optimal multiplex patterns.

(ii) We report a *design universality* observation that the design of the multiplex pattern can be decoupled from the image reconstruction algorithm. This is a significant departure from the recent literature that the multiplex pattern should be jointly optimized with the reconstruction algorithm. Our finding suggests that in the context of exposure multiplexing, an end-to-end training may not be necessary.

Index Terms—High Dynamic Range Imaging, Spatially Varying Exposure, Exposure Multiplexing, Computational Photography

I. INTRODUCTION

D IGITAL image sensors today, at least for the majority of them, pick and choose a *global* exposure and conversion gain across the entire pixel array to control the amount of photon flux reaching the sensor. For high dynamic range (HDR) scenes, this global configuration requires the camera to capture a bracket of exposures and use post-processing algorithms to fuse an HDR image. However, in the presence of motion and noise, HDR fusion is known to be difficult.

Approximately two decades ago, Nayar and Mitsunaga proposed the idea of spatially multiplexing the exposure and conversion gain [1]. The argument was that we could capture multiplexed exposures like color filter arrays in a single-shot to avoid the motion problem. The reduction of the spatial resolution can be, in principle, recovered by an appropriately designed interpolation algorithm. Nayar and Mitsunaga's idea led to a series of very interesting work in coded exposures,



1

Fig. 1. Given an HDR scene, the choice of spatially-varying exposure pattern significantly influences the quality of the reconstructed image. We propose a new risk estimator to determine the optimal exposure pattern for capturing high dynamic range scenes. [Left] HDR scene reconstruction using an arbitrary spatially-varying exposure pattern. [Right] HDR scene reconstruction using the optimal exposure pattern for the scene.

including adaptive control schemes [2], hardware multi-bucket sensor designs [3], and some of the most recent works in co-optimizing the multiplex pattern and the reconstruction algorithm via deep learning [4], [5].

Despite the large number of prior work, we seldom ask the question of how to *design* the multiplex pattern. This is a meaningful question, because a poorly chosen multiplex pattern can produce a substantially worse image than the optimal one, as illustrated in Figure 1. However, if we want to choose the optimal pattern, we must first answer the question:

What is the *optimality criteria* for exposure multiplex?

"Optimality criteria" may seem trivial — just pick a pattern that maximizes the signal-to-noise ratio (SNR)! But SNR of what? If it is the SNR of the measured pixel values, then we need a way to quantify the locations of the saturated pixels: a group of sparsely located saturated pixels are easy to recover (think about the color filter arrays) whereas a group of densely concentrated pixels are hard to recover (think about inpainting a large hole in an image). How about the SNR of the reconstructed image? This seems more plausible because if a final reconstruction has the highest PSNR, then the pattern is optimal. However, to compare the final reconstruction of every pattern, we need to first capture the scene using every possible pattern. This exhaustive capturing and reconstruction approach defeats the purpose of finding a good multiplex pattern; if one already has all possible captures and associated reconstructions, why bother finding the pattern to be used?

The authors are with the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907, USA. Email: {qu27, chi14, stanchan}@purdue.edu. This work is supported, in part, by the National Science Foundation under grant 2030570.

Therefore, as we can see, the above seemingly easy task can quickly evolve into a challenging research problem.

The goal of this paper is to clarify the difficulties and propose solutions. Our two contributions are:

- We introduce the concept of spatially varying exposure risk (SVE-Risk). SVE-Risk is customized to measure the usefulness of a multiplex pattern. It is universal in the sense that the risk does not require knowledge about a particular image reconstruction algorithm but can still be used to predict what pattern will likely yield the optimal final image for the given scene. We also propose efficient computation methods for SVE-Risk and demonstrate practicality of using our suite of algorithms for multiplex pattern selection.
- We discover, through large-scale experiments, that a multiplex pattern and the image reconstruction algorithm do not need to be co-optimized. This is a departure from prior work that argue the necessity of co-optimization. We hope that our finding can stimulate discussions about the future sensor-algorithm co-design problems.

II. PROBLEM STATEMENT AND RELATED WORK

In this section we define notations, state the problem, and comment on the related work.

A. Problem Statement and Notations

We let $\boldsymbol{\theta} = [\theta_1, \dots, \theta_N] \in \mathbb{R}^N$ be the ground truth image denoting the incoming photon flux, and let $\mathbf{y} \in \mathbb{R}^N$ be the observed photoelectric signal produced by the sensor. If the sensor uses a global exposure τ and a global gain α to capture the image \mathbf{y} , then \mathbf{y} can be simulated according to the equation

$$\mathbf{y}(\alpha, \tau) = \text{ADC} \left\{ \text{Clip} \left\{ \text{CRF} \left\{ \left(\mathcal{P}[\tau \boldsymbol{\theta} \cdot \text{QE}] + \tau \mu_{\text{dark}} \right) \cdot \alpha \right\} \right\} + \mathcal{N}(0, \sigma_{\text{read}}^2) \right\}.$$
(1)

We regard Eq. (1) as the first order approximation to the actual image sensor. The symbols we used here are defined in Table I.

TABLE I Image sensor model parameters

Symbol	Meaning	Typical Value
ADC	Analog-digital	10~12-bit
Clip	Full well limit	5000 e-
au	Exposure	10ms, 20ms, 40ms
μ_{dark}	dark current	0.002 e-/s
QE	Quantum efficiency	80%
α	Conversion gain	1
$\sigma_{\rm read}$	Read noise	0.3 e-
CRF	Camera Response Function	
\mathcal{P}	Poisson distribution	
\mathcal{N}	Gaussian distribution	

The concept of multiplexing is to assign, periodically, an exposure pattern and a gain pattern such that each pixel will be subject to a different exposure and gain. Mathematically, for a 2×2 exposure and gain pattern, we define

$$\boldsymbol{\tau} = \{\tau_{\ell}\} = \begin{bmatrix} \tau_1 & \tau_2 \\ \tau_3 & \tau_4 \end{bmatrix} \qquad \boldsymbol{\alpha} = \{\alpha_{\ell}\} = \begin{bmatrix} \alpha_1 & \alpha_2 \\ \alpha_3 & \alpha_4 \end{bmatrix}$$

where each τ_{ℓ} is sampled from a set of exposure levels, e.g., $\{1, 2, 4, 8, ...\}$. The same holds for the gain α_{ℓ} .

The core research question we ask in this paper is the choice of τ and α .

Given the scene radiance θ , how do we select $\boldsymbol{\tau} = \{\tau_{\ell}\}$
and $\alpha = \{\alpha_\ell\}$ to generate a y such that the reconstructed
image $\widehat{\theta}(\mathbf{y})$ has the highest PSNR?

Why limit to 2×2 patterns? We limit the scope of this paper to 2×2 multiplex patterns. Readers may say: This is too restrictive. Why not analyze 4×4 or 16×16 ? Our short answer is that 2×2 patterns are more hardware friendly than other options.¹ However, even so, the purpose of this paper is not to argue that 2×2 is the best option. Instead, the question we ask is that given the problem of using 2×2 , how to determine the optimal one? Our analysis on 2×2 can be generalized to other pattern sizes.

What other options do we have for exposure control? The theme of this paper is exposure and gains controls. In the literature, there are three mainstream approaches:

- A) One fixed pattern for all. Mount a static mask with spatially-varying light transmittance on top of the sensor array. This requires minimal/no additional circuit as compared with traditional CMOS sensors. However, the spatially-varying exposure pattern is then fixed and cannot be changed to adapt to the scene. Most traditional work on spatially-varying exposure image reconstruction explicitly/implicitly use this approach [1], [6]–[12].
- B) (This paper) **Periodic** 2×2 , **updated on-the-fly**. Use different signal lines to control the transfer timing (and hence control the exposure time) of different pixels. This approach allows the spatially-varying exposure pattern to be set on-the-fly and adapt to the scene. However, the number of signal lines increases linearly with the number of independent controls, and so the number of signal lines needs to be small. Overall, it is more functionally versatile than mounted static mask at the cost of some additional circuitry.
- C) Per-pixel or per-block, updated on-the-fly. Add inpixel latches/flip-flops/logic circuit (also called digital pixel sensor as compared with traditional active pixel sensor). This setup allows maximal level of flexibility. Usually, the exposure of each pixel can be independently controlled. The downside of this approach is that the required additional circuitry is gigantic and that it results in a low fill factor (as low as 6% [13]) as well as higher circuit noise. Most of the focal plane coded exposure work adopt this approach [13]–[20].

Among the three options, we do not prefer Option A because if we want it to be applicable to all images, then it must optimize for the average case. This will very likely lead to a periodic pattern. Option B is preferred over Option C because its hardware requirement is lower.

¹An analogy worth mentioning is the color filter arrays: While we all agree that the 2×2 Bayer pattern is sub-optimal, today we only see a handful of non-Bayer color filters in camera products. Even the latest 4-cell Quad-Bayer patterns by Sony, Samsung and OmniVision are just variants of Bayer.

Our work stands at the intersection of spatially-varyingexposure (SVE) imaging and HDR imaging exposure control. Some existing works are worth noting.

SVE Imaging: After Nayar and Mitsunaga [1], a number of methods have been proposed to: improve image reconstruction quality [6], improve reconstruction speed as well as robustness against non-uniform noise strength [7], [21], adapt the application of SVE HDR imaging from a single image to videos [8], leverage the higher single frame dynamic range advantage of SVE imaging to tackle motion registration problem of HDR video capturing [9], extend the original concept of SVE to the idea of generalized assorted pixel [10], in which image resolution, dynamic range, and spectral profile can be balanced post-capture by imaging with an optimized complex SVE mask.

Among these SVE imaging works, [10] is closest to our problem and we emphasize the difference between their problem setting and ours in Table II.

TABLE II Comparison of problem settings

	Yasuma et al. [10]	Ours
Pattern	fixed	adjustable on-the-fly
Optimization	optimal for the average	optimal for every single
	over all lighting scenarios	scene
Objective	obtain a universal pattern	select best for current scene

HDR Exposure Control: Exposure bracketing is a very popular HDR imaging technique, which fuses multiple LDR images exposed at different levels to form one HDR image, for example, imaging with dual sampling sensors [22], onchip fusion [23], and recently using deep networks [24]-[32]. Despite its popularity, few studies focus on selecting proper exposure levels for the LDR frames. These exposure control algorithms devised for exposure bracketing mostly fall into two groups: (1) algorithms focusing on design simplicity and efficiency so as to be deployed on imaging devices [33]–[36], and (2) image formation modeling based algorithms focusing on optimality, i.e., to find the optimal set of exposures by some metric for reconstructing the scene [37], [38]. A recent work also demonstrates the possibility of using reinforcement learning to train an exposure bracketing selection network [39]. In terms of design, our SVE pattern selection algorithm is an image formation modeling based algorithm; however, methods like [37], [38] cannot be migrated to our problem without significant changes, in that spatial multiplexing and pixel interpolation are not within their problem scopes. Imaging with other types of image sensors, e.g., Quanta Image Sensors (QIS) [40]-[42] and Single-Photon Avalanche Diode (SPAD) [43], are also candidate solutions to HDR imaging. It is further suggested by [44], [45] that low bit-depth sensors provide wider dynamic range. In this work, we use a general sensor model Eq. (1).

III. SVE PATTERN SELECTION FOR HDR

In this section we present the core idea of this paper, which is the concept of SVE-Risk and efficient methods to evaluate the SVE-risk.

A. Limitations of SNR

To motivate the definition of the SVE-Risk, we first discuss the limitations of the per-pixel output-referred² signal-tonoise ratio (SNR). In the context of our problem and image formation model, the SNR at pixel i is

$$SNR_{i} = \begin{cases} \frac{(\alpha_{i}\tau_{i})\theta_{i}}{\sqrt{\alpha_{i}^{2}(\tau_{i}\theta_{i}+\mu_{dark})+\sigma_{read}^{2}}}, & (\alpha_{i}\tau_{i})\theta_{i} \leq V_{max}, \\ 0, & (\alpha_{i}\tau_{i})\theta_{i} > V_{max}, \end{cases}$$
(2)

where V_{max} is the maximum voltage allowed by the ADC.

Intuitively, the SNR highlights two aspects of the exposure/gain: (i) If the pixel is not saturated, then the SNR will increase with α_i and τ_i . (ii) If the pixel is saturated (so the signal received by ADC exceeds V_{max}), then the SNR is zero. The two cases are consistent with the classical model in [1].

With the per-pixel SNR_i defined, it is straightforward to define the risk of adopting a particular (α, τ) for capturing a given scene using average SNR of all pixels:

$$\operatorname{SNR-Risk}(\alpha, \tau) = \left(\frac{1}{N}\sum_{i=1}^{N}\operatorname{SNR}_{i}\right)^{-1},$$
 (3)

where the reciprocal is used to convert the SNR to a risk. SNR-Risk has two major drawbacks:

(i) SNR-Risk is agnostic to how saturated pixels distribute in the image. Consider the example shown in Figure 2. While the two patterns will give exactly the same SNR, only pattern A is recoverable because neighboring pixels are available. Pattern B contains a large region of saturated pixels, which is very difficult to recover.



Fig. 2. Two arrangements of saturated pixels share the same SNR-Risk; however, the image captured with arrangement A can be recovered by interpolation while it is more challenging to recover that with arrangement B.

(ii) SNR is a pixelwise calculation without considering its neighbors. Therefore, if there is a bright region, all four control elements of the 2×2 pattern will try to match the scene without coordinating among themselves. The consequence is that SNR-Risk will choose an all bright or all dark pattern.

²In this paper we are interested in full-well capacities that are sufficiently large. For pixels with extremely small full-well capacity, e.g., Quanta Image Sensors (QIS), one needs to use the more general formula known as the exposure-referred SNR. We refer to the article by Chan [46] for details.

The SNR-Risk computes the risk for a pixel depending solely on one pixel. This is in contrast to the fact that practical image reconstruction algorithms almost always aggregate spatial information before predicting a pixel's output value. Hence, SNR-Risk tends to over-estimate the risks associated with easily recoverable pixels. To overcome this issue, we want to design a risk with the neighborhood relations of pixels taken into consideration. To summarize, SNR-Risk is not a good metric because of the following.

SNR-Risk cannot comprehend the local structure of the exposure, hence it cannot tell us whether the acquired image is **recoverable**.

B. SVE-Risk

After explaining why SNR is not a good metric to assess the multiplex patterns, in this subsection we introduce a new concept called spatially varying exposure risk (SVE-Risk).

We first consider the definition of an ideal risk. Let y be the sensor readout and estimator $\hat{\theta} = [\hat{\theta}_1, \dots, \hat{\theta}_N]$ be the reconstruction mapping that produces an estimate $\hat{\theta}(\mathbf{y}, \alpha, \tau)$ of $\boldsymbol{\theta} = [\theta_1, \dots, \theta_N]$. The ideal risk is

$$\operatorname{Risk}(\boldsymbol{\alpha},\boldsymbol{\tau}) = \inf_{\widehat{\boldsymbol{\theta}}} \mathbb{E}_{\mathbf{y}} \left[\frac{1}{N} \sum_{i=1}^{N} \left(\frac{\widehat{\theta}_{i}(\mathbf{y},\boldsymbol{\alpha},\boldsymbol{\tau}) - \theta_{i}}{\theta_{i}} \right)^{2} \right], \quad (4)$$

where θ_i and $\hat{\theta}_i(\cdot)$ denote the *i*-th element of the ground truth radiance θ and the estimate $\hat{\theta}(\cdot)$, respectively. Note that if the estimator $\hat{\theta}$ is a pixel-wise maximum likelihood (ML) estimator using the forward model defined in Eq. (1), then the squared ratio in Eq. (4) is exactly the inverse of pixelwise output-referred SNR for non-overflow pixels. From this perspective, the SNR-Risk can be deemed as a special case of the ideal risk, in which the estimator is predetermined to be a pixel-wise ML estimator.

The caveat of Eq. (4) is that an oracle estimator θ for a scene is never known and it is also impossible to obtain the infimum by enumerating all possible reconstruction algorithms. To mitigate this issue, we approximate the risk by using a hypothetical ideal *local* estimator. This hypothetical estimator cannot be constructed in practice (any reconstruction algorithm is likely to be worse than the hypothetical estimator), but it can give us a meaningful approximation to the infimum.

Definition 1 (Local estimator): A local estimator θ_i at the pixel *i* is a function that maps the neighborhood observations $\mathbf{y}_i = \{y_j \mid j \in \mathcal{B}_i\}$ to an estimate $\hat{\theta}_i(\mathbf{y}_i)$, where \mathcal{B}_i denotes the neighborhood around pixel *i*. If y_i is saturated, $\hat{\theta}_i$ uses the neighborhood without y_i , i.e., $\mathbf{y}_{-i} = \{y_j \mid j \in \mathcal{B}_i\} \setminus \{y_i\}$.

As we define this hypothetical local estimator, we assume that it has the perfect knowledge about inter-pixel correlations of its neighborhood. Therefore, it allows us to achieve two things:

- When a pixel is not saturated, the estimator will return us the same value as the SNR-Risk.
- When a pixel is saturated, the estimator will make an interpolation. The interpolated pixel will have a risk no higher than the largest risk within the neighborhood.

Based on these properties, we can define the SVE-Risk by considering three situations:

The **SVE-Risk of the** *i*-**th pixel** is defined as

$$SVE-Risk_{i}(\alpha, \tau)$$
(5)
$$= \begin{cases} \frac{1}{\theta_{i}^{2}|\mathcal{B}_{i}^{*}|} \left[\frac{(\theta_{i}+\mu_{dark})}{\tau_{i}} + \frac{\sigma_{read}^{2}}{\alpha_{i}^{2}\tau_{i}^{2}} \right], & \text{unsaturated,} \\ \max_{j \in \mathcal{B}_{i} \setminus \{i\}} \frac{1}{\theta_{j}^{2}|\mathcal{B}_{j}^{*}|} \left[\frac{(\theta_{j}+\mu_{dark})}{\tau_{j}} + \frac{\sigma_{read}^{2}}{\alpha_{j}^{2}\tau_{j}^{2}} \right], & \text{sat., } \checkmark \mathcal{B}_{i}, \\ \left(\frac{V_{max}}{\alpha_{i}\tau_{i}} - \theta_{i} \right)^{2}, & \text{sat., } \times \mathcal{B}_{i}, \end{cases}$$

where $\mathcal{B}_i^* = \mathcal{B}_i \setminus \mathcal{S}$ is the neighborhood around pixel *i* minus the set of saturated pixels \mathcal{S} .

Let's elaborate on the three cases in the definition:

- Unsaturated: If a pixel is unsaturated, the risk is defined as the variance of the measurement (which is the denominator of Eq. (2), squared and normalized). The normalization is necessary for preventing bright regions in the image from dominating darker region risks and is crucial for HDR imaging. For unsaturated pixels, the risk calculated in Eq. (5) is essentially SNR-Risk value scaled by the number of observations in the neighborhood.
- Saturated, ✓ B_i: The neighborhood contains pixels that can be used for interpolation. In this case, the risk is defined as the worst variance in the neighborhood. The intuition is that since SVE-Risk uses neighboring pixels to determine the risk, it can be deemed as an extension of the SNR-Risk where we combine independent exposures.
- Saturated, $\times \mathcal{B}_i$: The neighborhood does not contain any useful pixels. The risk is defined as the squared error between the cutoff $\frac{V_{\text{max}}}{\alpha_i \tau_i}$ and expected radiance θ_i . Note that this risk is worse than the second case because, in the second case, the substitution comes from one of the unsaturated pixels. For the third case, the difference between $\frac{V_{\text{max}}}{\alpha_i \tau_i}$ and θ_i can be very large if θ_i is far from the cutoff.

With pixel-wise SVE-Risk defined, we define overall SVE-Risk by taking sum of each pixel's risk.

The SVE-Risk of the whole image is defined as $SVE-Risk(\alpha, \tau) = \sum_{i=1}^{N} SVE-Risk_i(\alpha, \tau).$ (6)

C. Computing SVE-Risk

After defining the SVE-Risk, the next big question is how to compute this seemingly "uncivilized" Eq. (5). However, before we explain how we calculate the SVE-Risk, we first explain the overall imaging pipeline outlined in Figure 3.

Since our imaging goal is to dynamically control the exposure, our decision on the 2×2 pattern needs to be fast. As illustrated in Figure 3, the way we compute the SVE-Risk is based on the histogram of the scene radiance. This histogram does not need to be perfectly precise. Therefore, we can use a lower resolution of the same scene, and we can use any average 2×2 pattern such as high-medium-medium-low. The



Fig. 3. Our imaging pipeline is for *dynamical* control of the exposure. We capture a low-resolution pilot image using a fixed 2×2 exposure pattern (typically high-medium-medium-low) to construct a histogram. From the histogram we compute the risk, and guide by the risk we capture the full image. Since the SVE-Risk is calculated based on the histogram instead of the image, its computational cost is very low.

purpose of the pilot capture is to construct a histogram so that we can make decisions for choosing the 2×2 pattern. Once we have the 2×2 pattern determined, we perform a full capture and image reconstructions.

We now discuss the histogram. Assume we have collected a pilot image. We denote the distribution of the radiance as $p(\theta)$. An example is shown in Figure 4. We stress again that operating on the radiance distribution rather than individual pixels is more efficient, because, once the histogram is built, the complexity of pattern risk estimation becomes proportional to the number of bins in the histogram (order of magnitude is hundreds to thousands) instead of the number of pixels (order of magnitude is hundreds of thousands to millions).



Fig. 4. An example radiance histogram. The overall radiance histogram is partitioned by the 4 exposures/gains into 5 intervals. For the unsaturated pixels in intervals A to D, the risk associated with them corresponds to the unsaturated case in Eq. (5). Pixels in the interval above the threshold $V_{\text{max}}/\tau_0\alpha_0$ (interval E) are all saturated and unlikely to be recovered. Their risks are computed as the empirical mean squared error as the saturated, $\times \mathcal{B}_i$ case in Eq. (5). Risks of saturated pixels in intervals below $V_{\text{max}}/\tau_0\alpha_0$ (intervals A to D) can be substituted with the worst risks of their neighbors, which is the saturated, $\checkmark \mathcal{B}_i$ case in Eq. (5).

Given a pattern $\{(\tau_{\ell}, \alpha_{\ell}) | \ell = 0, ..., 3\}$, without loss of generality, we assume that the exposure/gains are sorted so that $\alpha_0 \tau_0 \leq \alpha_1 \tau_1 \leq \alpha_2 \tau_2 \leq \alpha_3 \tau_3$. The pattern gives us four cutoff radiance levels $\frac{V_{\text{max}}}{\alpha_{\ell} \tau_{\ell}}$ which partition the entire radiance range into five intervals, as shown in Figure 4.

As shown in Figure 4, a radiance level θ saturates an exposure/gain element (τ_l, α_l) if it is greater than the corresponding cutoff $\frac{V_{\text{max}}}{\alpha_\ell \tau_\ell}$. For $\theta > \frac{V_{\text{max}}}{\alpha_0 \tau_0}$ (interval E), they saturate all exposure/gain elements and it is very likely that the pixels corresponding to these radiance levels are not recoverable by any reconstruction algorithm. In this case, risk associated with these radiance levels is, according to Case 3 in Eq. (5):

$$\overline{\text{SVE-Risk}}_{\text{nonrecoverable}}(\boldsymbol{\alpha}, \boldsymbol{\tau}, \boldsymbol{\theta}) = \left(\frac{V_{\text{max}}}{\alpha_0 \tau_0} - \boldsymbol{\theta}\right)^2.$$
(7)

Now consider a radiance level between two cutoffs $\frac{V_{\text{max}}}{\alpha_{\ell}\tau_{\ell}} < \theta < \frac{V_{\text{max}}}{\alpha_{\ell-1}\tau_{\ell-1}}$ (i.e., a radiance level in intervals B, C, D). The radiance level θ saturates all exposure/gain elements below it; however, since there are neighboring elements not saturated by θ , it is likely that saturated pixels at this radiance level can be recovered by exploring their neighbor pixels. The risk associated with this radiance level is then, according to Case 1 and Case 2 in Eq. (5):

SVE-Risk_{recoverable}
$$(\boldsymbol{\alpha}, \boldsymbol{\tau}, \boldsymbol{\theta})$$

$$= \frac{1}{4} \left[\sum_{j=0}^{\ell-1} \frac{1}{B_j \boldsymbol{\theta}^2} \left(\frac{(\boldsymbol{\theta} + \boldsymbol{\mu}_{dark})}{\tau_j} + \frac{\sigma_{read}^2}{\alpha_j^2 \tau_j^2} \right) \right]_{\text{contribution from non-saturated elements}} + \sum_{j=\ell}^{3} \frac{1}{B_0 \boldsymbol{\theta}^2} \left(\frac{(\boldsymbol{\theta} + \boldsymbol{\mu}_{dark})}{\tau_0} + \frac{\sigma_{read}^2}{\alpha_0^2 \tau_0^2} \right) \right]_{\text{saturated, use risk of worst non-saturated neighbor}}, \quad (8)$$

where B_j is the number of non-saturated pixels within the neighborhood (with predetermined size) of exposure/gain element (α_l, τ_l) . Note that, since we know how pattern is tiled across the entire sensor array, B_l can be calculated for each saturation scenario (i.e., 0, ..., 3 exposure/gain elements saturated) beforehand and be stored in memory (see supplementary).

The risk is therefore the sum of Eq. (7) and Eq. (8).

To **numerically compute the SVE-Risk**, we construct
the radiance histogram (like Figure 4), and calculate
SVE-Risk(
$$\alpha, \tau$$
)
$$= \int_{\frac{V_{\text{max}}}{\alpha_0 \tau_0}}^{\infty} \overline{\text{SVE-Risk}}_{\text{nonrecoverable}}(\alpha, \tau, \theta) p(\theta) d\theta$$
$$+ \int_{0}^{\frac{V_{\text{max}}}{\alpha_0 \tau_0}} \overline{\text{SVE-Risk}}_{\text{recoverable}}(\alpha, \tau, \theta) p(\theta) d\theta,$$
where $p(\theta)$ is the radiance histogram of the pilot capture.

We show a comparison of the run time of calculating the SVE-Risk and the SNR-Risk in Table III. Since SVE-Risk is calculated using the histogram, it is insensitive to the image resolution. In contrast, since SNR needs to evaluate every single pixel, its computation grows with the number of pixels.

D. Efficient Pattern Enumeration

There is one final design question we need to answer before using the SVE-Risk. It is the problem of candidate patterns to evaluate. Suppose we have 9 exposure levels for a 2×2

TABLE III Comparison of risk estimator run time for sensors of different resolutions. Numbers shown are run time of evaluating the corresponding risk of 495 patterns averaged over 100 executions.

Resolution	Runtime for	Runtime for
	calculating SVE-Risk	calculating SNR-Risk
512x896	0.20±0.01 sec	0.69±0.01 sec
1024x1792	0.27±0.01 sec	6.13±0.22 sec
2048x3584	0.52±0.01 sec	27.35±0.12 sec

pattern, we will have a total of $9^4 = 6561$ candidates. It will be too much computation if we need to calculate the SVE-Risk for each candidate pattern. Therefore, in this subsection, we present a method to eliminate low priority patterns.

To remove the low-priority patterns, we made an observation that the majority of all candidate patterns are redundant. For example, from the reconstruction point-of-view, an exposure pattern $\tau = \{1, 10, 10, 10\}$ is almost identical to patterns $\{10, 1, 10, 10\}$, $\{10, 10, 1, 10\}$, and $\{10, 10, 10, 1\}$. To justify this claim, we experiment with three representative reconstruction algorithms on sensor readouts synthesized with these four patterns. We test on a dataset containing 46 images and show the results in Figure IV. Across the four exposure patterns, the reconstruction results are identical for any fixed algorithm.

TABLE IV μ PSNR (db) of the reconstructed images for the four equivalent multiplex patterns. The PSNR is averaged across 46 test images and rounded to first decimal place.

	1	10	10	1	10	10	10	10	
	10	10	10	10	1	10	10	1	
ADMM-TV [47]	31.1dB		31.1dB		31.1dB		31.1dB		
Restormer [48]	30.0dB		30.0)dB	30.0dB		30.0dB		
LPA [7]	29.7	7dB	29.7dB		29.'	7dB	29.7dB		

Based on the observations above, we define the concept of pattern equivalence: *two patterns are equivalent for capturing a scene if they are permutations of each other*. Our proposed strategy is to enumerate on **pattern equivalence classes** and compute canonical form of each class. The canonical pattern is the one with maximum variation in the 2×2 grid. Specifically, if the input pattern is $\boldsymbol{\tau} = [\tau_1, \tau_2, \tau_3, \tau_4]$, we sort the sequence to obtain $\texttt{sort}(\boldsymbol{\tau}) = \{\tau_{[1]}, \tau_{[2]}, \tau_{[3]}, \tau_{[4]}\}$ where $\tau_{[1]} \leq \tau_{[2]} \leq \tau_{[3]} \leq \tau_{[4]}$. The canonical pattern is then defined as

$$ext{canon}(oldsymbol{ au}) = egin{bmatrix} au_{[1]} & au_{[3]} \ au_{[4]} & au_{[2]} \end{bmatrix}.$$

The intuition here is that since $\text{sort}(\tau)$ is already sorted, the alternating allocation of the exposure values will maximize the variation within the 2×2 grid. For example, the canonical forms of the patterns $\tau = [8, 32, 1, 10]$ and $\tau = [1, 10, 1, 10]$ are

$$\operatorname{canon}(\boldsymbol{\tau}) = \begin{bmatrix} 1 & 10 \\ 32 & 8 \end{bmatrix}$$
 and $\operatorname{canon}(\boldsymbol{\tau}) = \begin{bmatrix} 1 & 10 \\ 10 & 1 \end{bmatrix}$,

By enumerating on equivalence class instead of all realizable patterns, we reduce the complexity to $\sum_{k=1}^{\min(L,m)} {L \choose k} {m-1 \choose k-1}$ while maintain same coverage of realizable pattern space³. When L = 9 and m = 4, this reduces enumeration size from 6561 to 495.

check whether the two canonical forms are identical.

Remark: Readers may ask: The proposed algorithm is largely an exhaustive search and it requires knowledge about the radiance distribution. Is it possible to improve the search? We note that the pilot estimate is designed to be *coarse*. As long as the shape of the radiance distribution is obtained, we can perform the histogram-based calculation. For faster algorithms, we do not think the typical gradient-based algorithm would work here because our problem is discrete with many stationary points. There might be some discrete optimization methods. We are open to explore them in our future work.

IV. OPTIMALITY AND UNIVERSALITY

In the beginning of the paper, we mentioned two key findings of this paper. Firstly, we claim that for exposure multiplexing, there exists a better *optimality* criteria than the SNR. We have elaborated on the SVE-Risk in the previous section. In this section, we evaluate SVE-Risk by justifying the following statement.

Optimality. The SVE-Risk optimal exposure multiplex pattern can generate a raw image, if passed through a reconstruction algorithm, with nearly the highest PSNR.

The second claim we made is *universality*. In this section, we justify the following statement.

Universality. The optimal exposure multiplex pattern is universally good for *all* image reconstruction algorithms.

Because of the empirical nature of both statements, we answer them through experiments. Our experiments involve largescale datasets, several new ways of visualizing the results, and a collection of real data.

A. Datasets

Before diving into experiment design and results, we describe datasets and synthesis parameters in more details. We use linearized ground truth 16-bit HDR images from NTIRE [49], HDR-Eye [50], and SIGGRAPH17 [51] datasets as scene radiance maps. These radiance maps are normalized such that, under minimal exposure and unit gain, the photon flux corresponding to 99 percentile is within the ADC range. This setting is practical and realizable on hardware using modern auto-exposure control. We do not attempt to accommodate for the brightest 1% of pixels, because these pixels usually are light sources that directly shine on the sensor. Each normalized radiance map is resized such that the short edge has a length of

³See supplementary for the derivation and pseudo-code.

512 pixels. The radiance maps are then used to synthesize raw sensor readout using image formation model in Section II-A. We use HDR-Eye data for reconstruction algorithms hyperparameter tuning and model training, SIGGRAPH17 data for hyper-parameter validation, and NTIRE data for testing and empirical study.

B. Reconstruction Algorithms

An important task of our verification is to evaluate the quality of the *reconstructed* image. Thus, it is necessary to consider image reconstruction algorithms. However, given the sheer volume of reconstruction methods, it would be impossible to evaluate everyone. A surrogate we take here is to consider three representative classes of algorithms. Within each class, we consider a representative method which we can either implement or we have access to the original source code.

- (i) Non-data-driven non-iterative: Algorithms in this class do not require any training and usually make relatively simple or even no assumptions about image structure. During reconstruction, the estimation for each pixel is only carried out once (hence non-iterative). Traditional bi-linear/quadratic/cubic interpolation, median filters, filter banks and more fall into this class. We adopt and implement a local polynomial approximation (LPA) [7] as a representative of this class.
- (ii) Non-data-driven iterative: These are classical tools for solving an inverse problem. Compared with non-iterative approaches, algorithms in this class usually model both the forward imaging process and underlying image structure/prior. This class of algorithms alternate between a forward step to handle the data fidelity, and a backward step that integrates the scene prior. Typical examples include Plug-and-Play [47] and Piecewise Linear Estimators [52], etc. We implement a Plug-and-Play ADMM with total variation prior (ADMM-TV) to represent this category.
- (iii) Data-driven: Dictionary learning [53] and neural networks [48], [54], [55] based image reconstruction/restoration methods fall into this category. It should be noted that although some non-data-driven iterative approaches may also adopt a dictionary or network as a sub-component (e.g., one may use a denoiser network in Plug-and-Play framework as prior step), and there has been numerous efforts [56], [57] to try to bring together the best of both tools, we do not consider them as purely data-driven approaches. We limit the scope of this category to methods that are one-pass (i.e., non-iterative) and trained directly for the inference task. We use Restormer⁴ [48] as an example of this category.

C. Metrics

Because of the unique problem setting we have, there is no prior standardized evaluation criterion. To this end, we consider a few known metrics and introduce a few new ones.

(i) μ **PSNR**, μ **SSIM**, μ **LPIPS**. In high dynamic range images, high exposure regions can easily dominate losses or metrics over low exposure regions; therefore, evaluating reconstruction quality in linear scale is usually less meaningful. Similar to other HDR related works [39], we evaluate reconstruction quality on μ -tone-mapped images, which is defined as

$$\mathbf{x}_{\mu} = \frac{\log(1 + \mathbf{x} \cdot \mu)}{\log(1 + \mu)}$$

where x is a linear scale image normalized to [0, 1] and μ is a hand-picked hyper-parameter controlling the strength of dynamic range compression. In our experiment, we set μ to the maximum reference level of ADC (see sec. II-A).

In tone-mapped space, we measure the PSNR (μ PSNR, higher is better), structural similarity (μ SSIM, higher is better) [58], and perceptual distance (μ LPIPS, lower is better) [59] between reconstructions and ground truth images.

(ii) **SNR-Risk, SVE-Risk, and their variants**. Since one main objective of this paper is to propose SVE-Risk, it is necessary to compare it with SNR. In addition to the standard SNR-Risk and SVE-Risk described in previous sections, we also evaluate following two variants of the risks, which arise naturally as one contemplate why one risk works while the other does not. Thus, we have four risk terms to consider:

- SNR-Risk, as defined in Eq. (3).
- SVE-Risk, as defined in Eq. (6).
- SNR_{MSE}: SVE-Risk assigns coarse estimates of mean squared error (MSE) as risk to unrecoverable pixels. Given the close relationship between MSE and PSNR/SSIM, one may wonder if this assignment gives SVE-Risk an unfair advantage over SNR-Risk, as SNR-Risk is purely forward model based. To answer this, we modify SNR-Risk by assigning the actual MSE between the normalized sensor readout (i.e., $y/(\alpha \tau)$) and the corresponding ground truth to overflowing pixels. We denote this variant as SNR_{MSE} in results.
- SVE_{w/o}: The idea of our SVE-Risk design is that it penalizes neighborhoods with too many saturated pixels through an auto-tuned parameter $|\mathcal{B}|$. We evaluate SVE-Risk without the penalty term $|\mathcal{B}|$, denoted SVE_{w/o} in results, to show that this idea is indeed imperative for achieving optimal performance instead of being a dubious add-on.

D. Verify the Optimality of SVE-Risk

In this subsection we discuss our experiments to assess the optimality of the SVE-Risk. We first discuss the protocol of the experiment, and then the results.

Protocol of experiment. We would like to compare SNR-Risk and SVE-Risk. The evaluation of SNR-Risk requires access to the radiance map. For convenience, we directly use ground truth radiance map as the input. The evaluation of SVE-Risk is easier because we only need the histogram.

⁴We discovered in our experiment that networks cannot be trained well when the input to a network has a very high dynamic range, and this training failure cannot be saved by input normalization. Therefore, instead of operating on raw sensor readout in linear scale and predict a linear/log scale output (as most of network-based HDR works do. Their inputs are usually LDR images and their task is to combine LDR images into HDR images, so the domain of their problems does not align exactly with ours), we take a log scale normalized sensor readout as input and predict a log scale radiance map.

TABLE V

LEFT EIGHT COLUMNS: AVERAGE RECONSTRUCTION QUALITY DROP (MEASURED BY μPSNR, μSSIM, μLPIPS) BETWEEN USING THE ORACLE PATTERN AND THE TOP-1/TOP-5 PATTERN(S) FOR VARIOUS RECONSTRUCTION ALGORITHMS. RIGHT EIGHT COLUMNS: PROBABILITY OF THE RECONSTRUCTION QUALITY DIFFERENCE BETWEEN THE ORACLE PATTERN AND THE TOP-1 PATTERN SELECTED BY A RISK BEING GREATER THAN PRE-DETERMINED THRESHOLDS. FOR ALL CATEGORIES, SMALLER NUMBERS CORRESPOND TO BETTER SELECTIONS OF EXPOSURE PATTERNS. THE BEST SCORE IN EACH COMPARISON CATEGORY IS COLORED IN BLUE. WE USE DIFFERENT Q-SCORE THRESHOLD FOR μLPIPS BECAUSE ITS SCALE AND SPREAD ARE MUCH WIDER THAN THOSE OF THE OTHER TWO METRICS.

								$\mu PSNR$	۱ ۱							
		Δ	Δ_1		Δ_5			Q(1%)			Q(5%)					
	SVE	$SVE_{w/o}$	SNR	SNR_{mse}	SVE	$SVE_{w/o}$	SNR	SNR_{mse}	SVE	$SVE_{w/o}$	SNR	SNR _{mse}	SVE	$SVE_{w/o}$	SNR	SNR_{mse}
LPA	0.66	1.03	10.32	9.66	1.04	0.98	8.5	8.81	53.8	71.4	99.9	98.4	10	18.6	99.9	96.5
ADMM-TV	1.3	1.7	8.43	7.99	1.43	1.8	7.44	7.38	78.9	81.9	99.9	98.4	26.5	34.5	99.5	92.5
Restormer	0.49	1.54	6.19	5.38	0.62	1.31	4.57	4.49	49.3	87.4	99.9	98.7	4.3	29.8	97.7	96
	μ SSIM															
	Δ_1				Δ	Δ_5		Q(1%)			Q(5%)					
	SVE	$SVE_{w/o}$	SNR	SNR_{mse}	SVE	$SVE_{w/o}$	SNR	SNR_{mse}	SVE	$SVE_{w/o}$	SNR	SNR _{mse}	SVE	$SVE_{w/o}$	SNR	SNR_{mse}
LPA	0.0048	0.0050	0.1056	0.1255	0.0093	0.0065	0.0887	0.1096	14.3	17.4	95.9	93.6	0.3	0.1	62.2	81.3
ADMM-TV	0.0080	0.0078	0.1099	0.1070	0.0103	0.0098	0.0829	0.0892	27.2	29	84	85.5	1.9	0.7	45.9	58.5
Restormer	0.0063	0.0106	0.0469	0.0573	0.0075	0.0103	0.0391	0.0443	21.6	37.4	69.5	92.2	0	0	22.5	54.9
								μ LPIPS	5							
	Δ_1				Δ	Δ_5			Q(2	0%)		Q(80%)				
	SVE	$SVE_{w/o}$	SNR	SNR_{mse}	SVE	$SVE_{w/o}$	SNR	SNR _{mse}	SVE	$SVE_{w/o}$	SNR	SNR _{mse}	SVE	$SVE_{w/o}$	SNR	SNR_{mse}
LPA	0.007	0.018	0.179	0.218	0.019	0.020	0.147	0.191	10.4	40.1	99.6	97.1	0.1	2.3	83.5	94.6
ADMM-TV	0.026	0.031	0.169	0.169	0.032	0.035	0.128	0.140	58.4	71.6	87.9	97.5	8.6	13.5	72.5	78
Restormer	0.03	0.046	0.08	0.112	0.035	0.044	0.070	0.093	75.8	92.6	92.2	94.5	6.4	26.2	42.6	60.5

We assume we have access to 4 exposures evenly distributed across the total exposure levels. The histogram is then built using all non-saturating pixels.

Given a set of candidate patterns (495 patterns in our experiment), we define the **oracle pattern** as the one that gives the highest PSNR. We do not have access to this oracle pattern. We want to use a risk to estimate the best pattern and to rank all patterns. Note that the ranked top-1 pattern by a risk is usually NOT the oracle pattern.

To gauge the ranking power of a risk, it is not informative to compare the rank of the oracle pattern rated by different estimators, because the ranks do no reflect reconstruction quality differences. It is also not enough to only look at the reconstruction quality difference between the top-1 pattern and the oracle pattern for two reasons. Firstly, a small quality difference can be a coincidence due to specific textures or scenes being insensitive to choices of pattern. Secondly, even a risk that rank patterns poorly may find an acceptable pattern once in a while (as illustrated in Figure 6 (d)).

Therefore, in our evaluation protocol, we propose to measure two descriptive statistics of risk estimators:

A) Average quality difference between using the oracle and the top-K patterns ranked by an estimator. This statistic evaluates the absolute reconstruction quality drop when one uses top patterns selected by a risk estimator compared to using the oracle pattern. Furthermore, if top-K average difference is an increasing function of K, then the risk estimator likely has good ranking power on patterns. Mathematically, we define

$$\Delta_K = \frac{1}{NK} \sum_{n=0}^{N-1} \sum_{i=0}^{K-1} (\underbrace{s_n^*}_{\text{oracle score}} - \underbrace{s_{n,i}}_{i\text{-th score}}), \quad (9)$$

where Δ_K is the top-K average difference, s_n^* is the score of the oracle pattern on the *n*-th radiance map (n = 0, 1, ..., N-1), with N = 1494 in this paper), $s_{n,i}$ is the

score of the *i*-th top pattern as ranked by the risk.

B) Probability that the reconstruction quality difference between the oracle pattern and the top-1 pattern is above certain pre-determined threshold. This statistic measures: given a threshold that one considers as critical, what is the probability that using a particular risk estimator will not yield satisfactory results. Formally, we define

$$Q(\eta) = \frac{1}{N} \sum_{n=0}^{N-1} \mathbb{I}\left\{\frac{s_n^* - s_{n,1}}{s_n^*} > \eta\right\},$$
 (10)

where Q is the probability of having a difference above a threshold η , $\mathbb{I}[\cdot]$ is an indicator function.

Results. We show in Table V the top-K average differences and Q scores at two thresholds across the pattern ranking dataset. The full SVE-Risk is capable of selecting a pattern that will yield a reconstruction with close to oracle performance, with an average μ PSNR drop for the top-1 pattern around 1 dB. SVE-Risk without the neighborhood penalty term can still pick a reasonably good pattern, but is almost always subpar compared to pattern selected by full SVE-Risk. SNR-Risk as well as its variant SNR-Risk_{MSE} are incapable of picking a good pattern in almost all scenarios, and equipping SNR-Risk with MSE for overflowing pixels does not help SNR-Risk. This experiment shows the significance of properly assigning surrogate risk to recoverable overflowing pixels and exposure/gain control element binding.

How do the Optimal Patterns Look Like? To give readers an idea of how the optimal exposure/gain patterns look like, we show in Figure 5 four randomly selected scenes and their corresponding optimal exposure/gain patterns. As the radiance of the scenes change from all dark to all bright, the optimal patterns change from all-high to all-low. This variety of scenes with the experimental results suggest that our proposed scheme is able to adaptively select the exposure and gain based on the radiance.

Visualize Ranking Power. To illustrate the ranking power

	Scene 1	Scene	2 Scene 3	Scene 4								
:	Optimal patterns (oracle scheme)											
	Image	How bright?	Optimal exposure τ	Optimal gain α								
	Scene 1	All dark	$\{10,1,10,10\}$	$\{4,2,4,4\}$								
	Scene 2	All bright	$\{1,1,1,1\}$	$\{2,1,2,2\}$								
	Scene 3 Half-half		$\{1,10,10,1\}$	{1,2,4,2}								
	Scene 4	More dark	{1,10,10,10}	{1,4,4,1}								

Fig. 5. Scenes and their corresponding oracle patterns.

of SNR-Risk and SVE-Risk, we show a scatter plot of μ PSNR of a scene sorted by risk values in Figure 6. We remark that this is a novel visualization of the performance, as we have not seen a similar plot in the literature.

To interpret the results of this plot, we note that the x-axis of the plots is the risk ranked from high to low. There are four risks: SNR, SVE, SVE_{w/o} and SNR_{MSE}. The ideal risk for our task is μ PSNR. If we use μ PSNR as the metric to rank the patterns, we will have a scatter plot shown in Figure 6(a). A better pattern ranked by μ PSNR will, of course, give a higher μ PSNR. When we evaluate other risks, we see that the proposed SVE-Risk has the closest behavior to the ideal risk. In contrast, SNR-based risks show an overlapped behavior. This means that if we use SNR to pick the pattern, we will not be able to tell which pattern is the best because for the same SNR (x-axis), we have multiple patterns on the y-axis.

Staircase μ PSNR Behavior. Readers may wonder about the step-wise behavior. This is due to the image histogram, as the scene may contain large flat regions of similar radiance values. As the minimum exposure and gain go above certain thresholds, the brightest large region of scene becomes completely saturated and irrecoverable, causing significant quality drop. Such drop is intrinsic to the scene itself and related to values that local exposure and gain can take, but no pattern is guaranteed to be in any particular cluster as the scene varies (i.e., no intrinsically bad pattern).

Visualize the Patterns. A visualization of adopting the top pattern selected by different risks for capturing is shown in Figure 7. In Figure 8, we show an example of reconstructing simulated readouts captured with SVE-Risk top pattern and SNR-Risk top pattern using different reconstruction algorithms.

Can SNR-Risk Work if We Discard "Bad" Patterns? A common question people ask is that would SNR-Risk perform better if we throw away the bad patterns. Our answer is no. Firstly, we simply cannot discard bad patterns when there is no intrinsically bad pattern. Secondly, even if we analyze current scene and discard all patterns that may yield large saturated region, pixel-wise SNR will still not use nonuniform exposure levels. This can be seen in the μ PSNR v.s. risk rank scatter plots as shown Figure 9.

E. Verify the Universality of Patterns

In this subsection, we describe our discovery that an exposure/gain pattern is universal for *many* image reconstruction algorithms. This is a significant departure from the recent trend of camera-algorithm co-optimization where people have been arguing that jointly optimizing the pattern and the algorithm is essential. Our experiments in this subsection show the opposite. We find that the design of the exposure/gain pattern can be completely decoupled from the design of the image reconstruction algorithm.

Spearman's ranking correlation. To evaluate the dependency of the pattern and the algorithm, we need some notion of correlation between the two factors. The metric we consider in this paper is the Spearman's ranking correlation [60], although other types of correlations can also be used.

For each ground truth radiance image θ , we synthesize 495 raw sensor readouts { $\mathbf{x}_0, ..., \mathbf{x}_{494}$ }, one for each pattern equivalence class. For each readout, we use three distinct algorithms { f_0, f_1, f_2 } to reconstruct the radiance image and evaluate the reconstruction quality using three different metrics $\mathcal{M}_{\mu \text{PSNR}}, \mathcal{M}_{\mu \text{SSIM}}, \mathcal{M}_{\mu \text{LPIPS}}$. We define a score as

$$s_{i,j,k} = \mathcal{M}_k(f_j(\mathbf{x}_i), \boldsymbol{\theta}), \tag{11}$$

where $s_{i,j,k}$ denotes the score using *i*th pattern, *j*th algorithm, and metric *k*. The exhaustive evaluation results are collated to create a pattern ranking data set. We assess whether a monotonic relationship exists between a pair of reconstruction algorithms as pattern varies by computing Spearman's ranking correlation coefficient [60] over scores of a metric

$$o_{j,j',k} =$$
Spearman $(\{(s_{i,j,k}, s_{i,j',k}) \mid i = 0, \cdots, 494\}),$

where a tuple of scores $(s_{i,j,k}, s_{i,j',k})$ is treated as an observation, associated with which a p-value describes the probability that no monotonic relationship exists between them. The correlation coefficients are computed for every pair of algorithms over all 1494 images from NTIRE dataset.

Experiment Protocol. The overall procedure of the experiments is as follows. Given a scene radiance map, we exhaustively evaluate reconstruction quality of distinct reconstruction algorithms on synthesized sensor readouts for all patterns returned by our enumeration algorithm. Taking reconstruction quality scores as data samples, we calculate Spearman's ranking correlation coefficient for every pair of reconstruction algorithms. We conduct hypothesis testing

- \mathcal{H}_0 : the reconstruction quality between two algorithms are uncorrelated as exposure pattern varies,
- \mathcal{H}_a : the reconstruction quality are positively correlated.

This procedure is repeated on every sample of NTIRE [49] dataset. We report the average and median of correlation coefficients across 1494 images for each metric in Table VI.

Results. To give a sense of the reported value, we also show scatter plots of μ PSNR of pairs of reconstruction algorithms in Figure 10, with detailed numbers shown in Table VI.

The three scattered plots in Figure 10 are worth discussing. These three subplots are the μ PSNR comparison between LPA, ADMM-TV, and Restormer. The scattered plot shows a surprisingly strong correlation between any pair of the methods: If a 2 × 2 pattern favors LPA, it also favors ADMM-TV, and similarly for other pairs. Therefore, at least based on this limited set of experiments, we find that if a pattern is good,



Fig. 6. Scatter plots of μ PSNR of exposure patterns with their x-coordinates sorted. Each dot represents a pattern. (a) Exposure patterns sorted by their corresponding μ PSNR scores from low to high, representing what an *ideal* risk estimator should achieve. An ideal risk estimator always assigns lower risks to patterns yielding higher μ PSNR. (b) – (e) Exposure patterns sorted by their estimated risk values using various risk estimators. The y-coordinate of a dot is the μ PSNR of reconstruction using that pattern.



(a)



(b)

Fig. 7. Qualitative comparisons of risk estimators. Images in the middle are the ground truth. Images on the sides are reconstructions using simulated readouts captured with spatially-varying exposure patterns selected by different risk estimators. Oracle denotes the best possible pattern through exhaustive search. (a) SNR-Risk is dominated by relatively darker regions, so it selects an exposure pattern such that the high-light textures are irrecoverable. SVE-Risk_{w/o} fails to retain low-light details, e.g., characters on the license plate. Oracle and SVE-Risk selections preserve most information across high dynamic ranges. (b) SNR-Risk is unable to pick a spatially-varying exposure/gain, and therefore compromises texture in high-light regions and introduces irrecoverably heavy noise in low-light region. SVE-Risk_{w/o} loses details in medium exposure regions such as the human face and the white shirt. Oracle and SVE-Risk selections trade off high/low flux regions differently: the oracle pattern better preserves the face but loses some details of the curtain, while SVE-Risk's top pattern preserves all curtain details but results in a slightly more blurry face. The reconstruction algorithm is Restormer for all images. All images are tone-mapped.

TABLE VI Average (standard deviation) and median correlation coefficients and average p-value of each correlation coefficient across NTIRE dataset

		$\mu PSNR$			μ SSIM		μ LPIPS		
	average ρ	median ρ	average p-value	average ρ	median ρ	average p-value	average ρ	median ρ	average p-value
ADMM-TV v.s. LPA	0.937 ± 0.035	0.952	$< 10^{-7}$	0.805 ± 0.116	0.830	$< 10^{-7}$	0.740 ± 0.195	0.775	0.008 ± 0.079
LPA v.s. Restormer	0.877 ± 0.070	0.878	$< 10^{-7}$	0.826 ± 0.053	0.826	$< 10^{-7}$	0.773 ± 0.112	0.770	$< 10^{-7}$
Restormer v.s. ADMM-TV	0.891 ± 0.042	0.885	$< 10^{-7}$	0.701 ± 0.130	0.719	$< 10^{-7}$	0.564 ± 0.306	0.625	0.071 ± 0.252

it is good for all reconstruction algorithms; if it is bad, it is bad for all reconstruction algorithms. By inspecting the numbers in Table VI, we further note that the Spearman's correlation coefficients are all in the range of 0.87 or above (for μ PSNR). For other evaluation metrics μ SSIM and μ LPIPS, we also see a high correlation coefficient.

We believe that this finding is new and perhaps less expected. The implication is that if we need to design the

multiplexing pattern, there is no need to consider the image reconstruction algorithm. This is a good news from the point of a designer's perspective. Co-optimization is not always preferred because we do not want the patterns to be dependent on a particular algorithm. If we can modularize the designs of the two, the debugging and analysis of the methods will be significantly easier.



Fig. 8. Images reconstructed by different algorithms. Using the sub-optimal spatially-varying exposure pattern selected by SNR-Risk, all three reconstruction algorithms fail to recover the texture of the hardwood flooring.



Fig. 9. μ PSNR versus risk rank scatter plots. Even if we throw away bad pixels, SNR-Risk still cannot generate a smooth curve. In contrast, SVE-Risk produces a largely smooth curve. Color code: (SVE, SNR)

F. Real Experiments

In this section, we test the SVE-Risk and SNR-Risk on real camera raw readouts and show the feasibility of the proposed risk on real hardware for exposure pattern selection. Since no SVE sensor is available to us, we interlace real camera raw readouts to synthesize images captured with SVE patterns.

1) Experiment Settings and Procedure: For each high dynamic range scene, we use a Sony Alpha 7 II camera to capture 9 differently exposed LDR frames. We capture 5 HDR scenes in total. For ease of camera parameter calibration, we keep the camera ISO at 100 and gradually increased the exposure time from 1/80 sec to 3.2 sec, doubling from one frame to the next. Our imaging model based risks require knowledge of dark current and read noise level. For these two parameters, we estimate them by capturing three dark frames (ISO 100, exposure 1/80 sec, 0.2 sec, 1.6 sec). We generate a pseudo ground truth radiance map by fusing all 9 frames. We synthesize all possible SVE captures using following procedure:

- A) Enumerate all possible SVE patterns with 9 different exposures.
- B) For each pattern, pick the corresponding frames from the 9 raw frames.
- C) Interlace picked frames to generate an SVE frame by taking every other pixel.

Then, we use our trained Restormer to reconstruct scene radiance for every SVE captures and compare the reconstruction against the pseudo ground truth. We build the empirical radiance histogram of a scene by treating the SVE capture with exposure (1/80 sec, 1/20 sec, 1/5 sec, 4/5 sec) as the pilot frame. We use this histogram to calculate SVE-Risk for every pattern. For SNR-Risk, we used the pseudo ground truth of a scene for calculation.

The camera response function (CRF) of Sony Alpha 7 II camera raw readout is close to linear at ISO 100 except when the photon charge accumulated is near full well capacity. Therefore, we use a linear CRF and aggregate the entire conversion from charge to final ADC readout (voltage follower gain, column amplifier gain, and output amplifier gain) into conversion gain term α in Eq. (1). The final parameters used in this experiment are listed in table VII.

2) Results: We show the quantitative reconstruction quality in terms of μ PSNR in table VIII and two sets of qualitative results in Figure 11. As evidenced in both quantitative and qualitative results, using top pattern selected by SVE-Risk can yield near-optimal final reconstruction.

TABLE VII Measured parameters in terms of equivalent charges for real camera experiment

Gain α	5.25 e- per ADC Unit
Exposure τ	1/80 sec, 1/40 sec,, 1.6 sec
Read noise σ_{read}	1.3 e-
Camera response function	Linear
dark current	1.4e-/sec
Clip threshold	84667e-

TABLE VIII μ PSNR of images reconstructed from raw images interlaced
according to best possible patterns (oracle), patterns
selected by SVE-Risk, and patterns selected by SNR-Risk,
compared to the pseudo ground truth

	scene 1	scene 2	scene 3	scene 4	scene 5
oracle	44.5	44.4	35.5	40.3	49.5
SVE-Risk top-1	44.5	42.3	33.9	38.8	49.3
SNR-Risk top-1	41.8	38.1	30.0	27.5	47.6

V. CONCLUSION

In this paper, we report two findings about the design of a spatially varying exposure multiplexing scheme. Firstly, we show that the pixel-wise SNR is a poor metric to quantify the performance of a multiplex pattern because it fails to



Fig. 10. Tone-mapped ground truth HDR scene and cross comparisons between reconstruction algorithms in μ PSNR. Each dot in the scatter plots represents an exposure pattern. The x and y coordinates are μ PSNR scores of the reconstructions by different algorithms on the image captured using that exposure pattern. The average correlation coefficient of the three scatter plots is approximately $\rho = 0.86$.



Fig. 11. Reconstruction results on real images and experimental setup. Input images with spatially-varying patterns are generated by interlacing raw images captured at various exposures. For (a) and (b), we show full frames of pseudo ground truth in both linear scale and tone-mapped scale (top-left and bottom-left, respectively). The linear scales are clipped at 95 percentile radiance level and are normalized to [0, 255]. We also show zoomed-in regions of tone-mapped reconstructions on the best possible spatially-varying pattern via exhaustive search (oracle), the best pattern selected by SVE-Risk (SVE), and the best by SNR-Risk (SNR). The reconstruction algorithm is Restormer for all images. (a) The texts and textures at low light shown in the orange box are preserved equally well in all three reconstructions, but SNR-Risk reconstruction loses all details of high-light regions. (b) The dark regions (QR code, wall, etc.) of all three reconstructions are indistinguishable, but SNR-Risk reconstruction loses outdoor high-light details (red arrows). (c) Experimental setup. We capture real images of HDR scenes using a Sony Alpha 7 II camera.

differentiate the recoverable cases and the non-recoverable cases. We circumvent the difficulty by proposing the SVE-Risk. Our experiments show that the pattern ranking provided by the SVE-Risk correlates extremely well with the ideal ranking. Secondly, through a large-scale experiment, we find that for spatially-varying-exposure imaging with tiled exposure patterns, it is not necessary to design a pattern selection algorithm tailored for specific reconstruction algorithm; the margin of improvement for using tailored/co-designed pattern selection algorithm is limited. Our finding is a significant departure from recent work in computational photography that advocates for sensor-algorithm co-optimization. To sensor designers, this could be good news because sensor-algorithm co-design is significantly more costly for production. However, our bigger hope is that this counterexample can stimulate more discussions about the necessity of sensor-algorithm co-optimization, and under what context would it become beneficial *not* to co-optimize.

REFERENCES

 S. Nayar and T. Mitsunaga, "High dynamic range imaging: spatially varying pixel exposures," in Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), vol. 1, pp. 472-479, 2000.

- [2] S. K. Nayar and V. Branzoi, "Adaptive dynamic range imaging: optical control of pixel exposures over space and time," in *Proc. IEEE International Conference on Computer Vision (ICCV)*, vol. 2, pp. 1168–1175, 2003.
- [3] G. Wan, X. Li, G. Agranov, M. Levoy, and M. Horowitz, "CMOS image sensors with multi-bucket pixels for computational photography," *IEEE Journal of Solid-State Circuits*, vol. 47, no. 4, pp. 1031–1042, 2012.
- [4] C. M. Nguyen, J. N. P. Martel, and G. Wetzstein, "Learning spatially varying pixel exposures for motion deblurring," in *IEEE International Conference on Computational Photography (ICCP)*, pp. 1–11, 2022. Available online at https://arxiv.org/abs/2204.07267.
- [5] T. Klinghoffer, S. Somasundaram, K. Tiwary, and R. Raskar, "Physics vs. learned priors: Rethinking camera and algorithm design for task-specific imaging," 2022. Available online at https://arxiv.org/abs/2204.09871.
- [6] C. Aguerrebere, A. Almansa, Y. Gousseau, J. Delon, and P. Musé, "Single shot high dynamic range imaging using piecewise linear estimators," in 2014 IEEE International Conference on Computational Photography (ICCP), pp. 1–10, 2014.
- [7] S. Hajisharif, J. Kronander, and J. Unger, "Adaptive Dual ISO HDR Reconstruction," *EURASIP Journal on Image and Video Processing*, Dec 2015.
- [8] M. Schoberl, A. Belz, A. Nowak, J. Seiler, A. Kaup, and S. Foessel, "Building a high dynamic range video sensor with spatially nonregular optical filtering," in *Applications of Digital Image Processing XXXV* (A. G. Tescher, ed.), vol. 8499, p. 84990C, International Society for Optics and Photonics, SPIE, 2012.
- [9] U. Cogalan, M. Bemana, K. Myszkowski, H. P. Seidel, and T. Ritschel, "Learning HDR video reconstruction for dual-exposure sensors with temporally-alternating exposures," *Computers & Graphics*, vol. 105, pp. "57–72", 2022.
- [10] F. Yasuma, T. Mitsunaga, D. Iso, and S. K. Nayar, "Generalized assorted pixel camera: Postcapture control of resolution, dynamic range, and spectrum," *IEEE Transactions on Image Processing*, vol. 19, no. 9, pp. 2241–2253, 2010.
- [11] M. Alghamdi, Q. Fu, A. Thabet, and W. Heidrich, "Reconfigurable snapshot HDR imaging using coded masks and inception network," in *Vision Modeling and Visualization*, 01 2019.
- [12] Y. Jiang, I. Choi, J. Jiang, and J. Gu, "HDR video reconstruction with tri-exposure quad-bayer sensors," 2021.
- [13] S. J. Carey, D. R. Barr, B. Wang, A. Lopich, and P. Dudek, "Mixed signal SIMD processor array vision chip for real-time image processing," *Analog Integr. Circuits Signal Process.*, vol. 77, p. 385–399, dec 2013.
- [14] Y. Luo, D. Ho, and S. Mirabbasi, "Exposure-programmable CMOS pixel with selective charge storage and code memory for computational imaging," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 65, no. 5, pp. 1555–1566, 2018.
- [15] H. Ke, N. Sarhangnejad, R. Gulve, Z. Xia, N. Gusev, N. Katic, K. N. Kutulakos, and R. Genov, "Extending image sensor dynamic range by scene-aware pixelwise-adaptive coded exposure," in *Proc. Int. Image Sensor Workshop*, pp. 111–114, 2019.
- [16] H. Reyserhove, A. S. Berkovich, and X. Liu, "Programmable pixel array." U.S. Patent 20200195828A1, Jun. 2020.
- [17] J. Zhang, R. Etienne-Cummings, T. Xiong, T. D. Tran, and S. H. Chin, "Flexible pixel-wise exposure control and readout." U.S. Patent 20180115725A1, Nov. 2019.
- [18] N. Sarhangnejad, N. Katic, Z. Xia, M. Wei, N. Gusev, G. Dutta, R. Gulve, H. Haim, M. M. Garcia, D. Stoppa, K. N. Kutulakos, and R. Genov, "5.5 dual-tap pipelined-code-memory coded-exposurepixel cmos image sensor for multi-exposure single-frame computational imaging," in 2019 IEEE International Solid- State Circuits Conference - (ISSCC), pp. 102–104, 2019.
- [19] J. N. P. Martel, L. K. Müller, S. J. Carey, P. Dudek, and G. Wetzstein, "Neural sensors: Learning pixel exposures for HDR imaging and video compressive sensing with programmable sensors," *IEEE Transactions* on Pattern Analysis and Machine Intelligence, vol. 42, no. 7, pp. 1642– 1653, 2020.
- [20] J. Zhang, J. P. Newman, X. Wang, C. S. Thakur, J. Rattray, R. Etienne-Cummings, and M. A. Wilson, "A closed-loop, all-electronic pixel-wise adaptive imaging system for high dynamic range videography," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 67, no. 6, pp. 1803–1814, 2020.
- [21] J. Kronander, S. Gustavson, and J. Bonnet, G.and Unger, "Unified HDR reconstruction from raw CFA data," in *IEEE International Conference* on Computational Photography (ICCP), pp. 1–9, 2013.

- [22] O. Yadid-Pecht and E. R. Fossum, "Wide intrascene dynamic range cmos aps using dual sampling," *IEEE Transactions on Electron Devices*, vol. 44, no. 10, pp. 1721–1723, 1997.
- [23] Y. Wang, S. Barna, S. Campbell, and E. R. Fossum, "A high dynamic range cmos aps image sensor," in *IEEE Workshop CCD and Advanced Image Sensors, Lake Tahoe, Nevada, USA*, 2001.
- [24] N. K. Kalantari and R. Ramamoorthi, "Deep high dynamic range imaging of dynamic scenes," ACM Transactions on Graphics (TOG), vol. 36, no. 4, pp. 1–12, 2017.
- [25] S. Wu, J. Xu, Y.-W. Tai, and C.-K. Tang, "Deep high dynamic range imaging with large foreground motions," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 117–132, 2018.
- [26] Q. Yan, D. Gong, Q. Shi, A. v. d. Hengel, C. Shen, I. Reid, and Y. Zhang, "Attention-guided network for ghost-free high dynamic range imaging," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1751–1760, 2019.
- [27] Q. Yan, D. Gong, P. Zhang, Q. Shi, J. Sun, I. Reid, and Y. Zhang, "Multiscale dense networks for deep high dynamic range imaging," in 2019 *IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 41–50, IEEE, 2019.
- [28] Q. Yan, L. Zhang, Y. Liu, Y. Zhu, J. Sun, Q. Shi, and Y. Zhang, "Deep HDR imaging via a non-local network," *IEEE Transactions on Image Processing*, vol. 29, pp. 4308–4322, 2020.
- [29] Y. Deng, Q. Liu, and T. Ikenaga, "Multi-scale contextual attention based hdr reconstruction of dynamic scenes," in *Twelfth International Conference on Digital Image Processing (ICDIP 2020)*, vol. 11519, pp. 413–419, SPIE, 2020.
- [30] Q. Yan, S. Zhang, W. Chen, Y. Liu, Z. Zhang, Y. Zhang, J. Q. Shi, and D. Gong, "A lightweight network for high dynamic range imaging," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 824–832, 2022.
- [31] L. Zhu, F. Zhou, B. Liu, and O. Göksel, "HDRfeat: A feature-rich network for high dynamic range image reconstruction," arXiv preprint arXiv:2211.04238, 2022.
- [32] Y. Chi, X. Zhang, and S. H. Chan, "HDR imaging with spatially varying signal-to-noise ratios," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5724–5734, 2023.
- [33] R. Pourreza-Shahri and N. Kehtarnavaz, "Exposure bracketing via automatic exposure selection," in 2015 IEEE International Conference on Image Processing (ICIP), pp. 320–323, 2015.
- [34] N. Barakat, A. N. Hone, and T. E. Darcie, "Minimal-bracketing sets for high-dynamic-range image capture," *IEEE Transactions on Image Processing*, vol. 17, no. 10, pp. 1864–1875, 2008.
- [35] B. Guthier, S. Kopf, and W. Effelsberg, "Optimal shutter speed sequences for real-time HDR video," in 2012 IEEE International Conference on Imaging Systems and Techniques Proceedings, pp. 303–308, 2012.
- [36] K. F. Huang and J. C. Chiang, "Intelligent exposure determination for high quality HDR image generation," in 2013 IEEE International Conference on Image Processing, pp. 3201–3205, 2013.
- [37] K. Hirakawa and P. J. Wolfe, "Optimal exposure control for high dynamic range imaging," in 2010 IEEE International Conference on Image Processing, pp. 3137–3140, 2010.
- [38] S. W. Hasinoff, F. Durand, and W. T. Freeman, "Noise-optimal capture for high dynamic range photography," in 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 553–560, 2010.
- [39] Z. Wang, J. Zhang, M. Lin, J. Wang, P. Luo, and J. Ren, "Learning a reinforced agent for flexible exposure bracketing selection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1820–1828, 2020.
- [40] Y. Chi, A. Gnanasambandam, V. Koltun, and S. H. Chan, "Dynamic low-light imaging with Quanta Image Sensors," in *Computer Vision– ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28,* 2020, Proceedings, Part XXI 16, pp. 122–138, Springer, 2020.
- [41] C. Li, X. Qu, A. Gnanasambandam, O. A. Elgendy, J. Ma, and S. H. Chan, "Photon-limited object detection using non-local feature matching and knowledge distillation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 3976–3987, 2021.
- [42] A. Gnanasambandam and S. H. Chan, "HDR imaging with Quanta Image Sensors: Theoretical limits and optimal reconstruction," *IEEE transactions on computational imaging*, vol. 6, pp. 1571–1585, 2020.
- [43] A. Ingle, T. Seets, M. Buttafava, S. Gupta, A. Tosi, M. Gupta, and A. Velten, "Passive inter-photon imaging," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8585– 8595, 2021.

- [44] S. H. Chan, "What does a one-bit Quanta Image Sensor offer?," *IEEE Transactions on Computational Imaging*, vol. 8, pp. 770–783, 2022.
- [45] S. H. Chan, "On the insensitivity of bit density to read noise in one-bit Quanta Image Sensors," *IEEE Sensors Journal*, 2023.
- [46] A. Gnanasambandam and S. H. Chan, "Exposure-referred signal-tonoise ratio for digital image sensors," *IEEE Transactions on Computational Imaging*, vol. 8, pp. 561–575, 2022.
- [47] S. H. Chan, X. Wang, and O. A. Elgendy, "Plug-and-play ADMM for image restoration: Fixed-point convergence and applications," *IEEE Transactions on Computational Imaging*, vol. 3, no. 1, pp. 84–98, 2017.
- [48] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, and M. H. Yang, "Restormer: Efficient transformer for high-resolution image restoration," in CVPR, 2022.
- [49] E. P'erez-Pellitero, S. Catley-Chandar, A. Leonardis, and R. Timofte, "Ntire 2021 challenge on high dynamic range imaging: Dataset, methods and results," 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 691–700, 2021.
- [50] P. Korshunov, H. Nemoto, A. Skodras, and T. Ebrahimi, "Crowdsourcing-based Evaluation of Privacy in HDR Images," in Optics, Photonics, and Digital Technologies for Multimedia Applications III (P. Schelkens, T. Ebrahimi, G. Cristóbal, F. Truchetet, and P. Saarikko, eds.), vol. 9138, p. 913802, International Society for Optics and Photonics, SPIE, 2014.
- [51] N. K. Kalantari and R. Ramamoorthi, "Deep high dynamic range imaging of dynamic scenes," ACM Transactions on Graphics (Proceedings of SIGGRAPH 2017), vol. 36, no. 4, 2017.
- [52] G. Yu, G. Sapiro, and S. Mallat, "Solving inverse problems with piecewise linear estimators: From gaussian mixture models to structured sparsity," *IEEE Transactions on Image Processing*, vol. 21, no. 5, pp. 2481–2499, 2012.
- [53] C. Bao, J. F. Cai, and H. Ji, "Fast sparsity-based orthogonal dictionary learning for image restoration," in 2013 IEEE International Conference on Computer Vision, pp. 3384–3391, 2013.
- [54] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142–3155, 2017.
- [55] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, M. H. Yang, and L. Shao, "Multi-stage progressive image restoration," in *CVPR*, 2021.
- [56] V. Monga, Y. Li, and Y. C. Eldar, "Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing," *IEEE Signal Processing Magazine*, vol. 38, no. 2, pp. 18–44, 2021.
- [57] D. Gilton, G. Ongie, and R. Willett, "Deep equilibrium architectures for inverse problems in imaging," *IEEE Transactions on Computational Imaging*, vol. 7, pp. 1123–1133, 2021.
- [58] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions* on *Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.
- [59] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *CVPR*, 2018.
- [60] D. Zwillinger and S. Kokoska, CRC Standard Probability and Statistics Tables and Formulae. Chapman & Hall, 2000.

Supplementary Material

This supplementary document contains following:

- More details on parameters used for synthesis and reconstruction algorithms (I)
- More details on counting of number of non-saturated pixels within a neighborhood (II)
- Derivation for pattern equivalence class size and pseudo-code for efficient enumeration algorithm (III)

I. MORE EXPERIMENT DETAILS

We describe in more detail the parameters used for the data synthesis and reconstruction algorithms in this section.

A. Synthesis Parameters

Exact parameters used for synthesis is shown in table I

Symbol	Meaning	Values Used for Experiment
ADC	Analog-to-Digital Converter	10-bit
	ADC least significant bit resolution	8 e- at lowest gain, 0.1 e- at highest gain
	ADC lower bound	1 e- at lowest gain, 0.0125 e- at highest gain
	ADC upper bound	8185 e- at lowest gain, 102 e- at highest gain
Clip	full well limit	8200 e-
au	exposure of individual pixel	$\{\times 0.25, \times 0.5, \times 1\}$ global exposure duration
	global exposure	30ms
μ_{dark}	dark current	0.2 e- at shortest exposure
QE	quantum efficiency	80%
α	conversion gain	{1, 10, 80} (roughly correspond to ISO 100, 1000, 8000)
$\sigma_{\rm read}$	read noise	20 e- at lowest gain, 0.25 e- at highest gain
CRF	Camera Response Function	Linear
		TABLE I

SYNTHESIS PARAMETERS

Note that

- wherever applicable, values shown are mapped from their original domain to number of electrons held in the potential well of a pixel for ease of understanding.
- the spatially-varying-exposure can be achieved by either adjusting pixel-wise exposure duration or modulating light right on top of the sensor array (sensor side modulation). In our experiment, we assumed using pixel-wise exposure duration control.
- CRF is in general non-linear on most modern cameras, but since our data is synthesized and a non-linear CRF only uniformly extends the dynamic range of all pixels, we adopt a linear CRF for simplicity

B. Reconstruction Algorithm Parameters

Parameters of LPA [1]/ADMM-TV [2] are tuned by performing grid search on HDR-Eye data set. Model of Restormer [3] is trained on HDR-Eye [4] data set and hyper-parameters are set by evaluating trained model on SIGGRAPH17 [5] data set.

- LPA For LPA algorithm, we estimate each pixel with a window of size 7-by-7 and a Gaussian radial basis function with scale parameter of 1.
- ADMM-TV For ADMM-TV, we use regularization coefficient of 1 and run at most 30 iterations for reconstruction. We use scikit-image Chambolle total variation denoising implementation with default parameters for prior step.
- **Restormer** We use original author's implementation¹ with following hyper-parameters "*inp_channels*" : 1, "*out_channels*" : 1, "*dim*" : 48, "*num_blocks*" : [4, 6, 6, 8], "*num_refinement_blocks*" : 4, "*heads*" : [1, 2, 4, 8], "*ffnexpansionfactor*" : 2.66, "*bias*" : *false*, "*LayerNormtype*" : "*WithBias*", "*dual_pixel_task*" : *false*. Since the network consumes a gigantic amount of memory even at inference time, we train the network on a fixed spatial resolution of 128-by-128. During test time, we partition image to 128-by-128 patches, reconstruct each single patch, and finally stitch patches back to original image.



Fig. 1. Example of counting number of non-saturated pixels with 3x3 neighborhood. The radiance levels and exposure/gain setting determines the saturation case, but for each case, the number of saturated pixels is fixed. Note that this is only true under our small neighborhood and pixel variation assumptions. We do not care about number of non-saturated pixels in a neighborhood for saturated exposure/gain element because saturated pixels' risks are replaced with their neighbors' and the number is not used.

II. COUNTING NON-SATURATED NEIGHBORS FOR COMPUTING SVE-RISK

Under the assumptions of using small neighborhood (e.g. 3x3, 5x5) for computing SVE-Risk as well as adjacent pixels having similar values, given our knowledge about how exposure/gain pattern is tiled across the entire sensor array, we can count number of non-saturated pixels within the neighborhood of any pixel at any radiance level before we calculate risk. In fact, we don't even need the knowledge of radiance distribution or the exact exposure/gain value for the counting.

As illustrated in figure 1, given maximum reference voltage of ADC (V_{max}) and exact exposure/gain values ({ τ_l }, { α_l }), a radiance level is mapped to a saturation case. This mapping is dependent on the exact values of θ as well as { τ_l }, { α_l } and cannot be known beforehand. However, the mapping from saturation case to number of non-saturated pixels within a pre-specified neighborhood is fixed. Therefore, once we determine the neighborhood size that we are gonna use for computing the SVE-Risk, we can do the counting for every saturation case and each exposure/gain control element, and then construct a hash table for usage when we calculate risk for a specific scene.

Numbers counted in this way are good approximations to the ground truth numbers for majority part of an image, and only become inaccurate around thresholds $\frac{V_{max}}{\tau_1 \alpha_1}$ and dark-bright region boundaries.

III. ENUMERATION SIZE DERIVATION AND IMPLEMENTATION

Given an SVE array with L unique exposure/gain levels (9 in our experiment) and M independently configurable exposure/gain control elements (4 in our experiment), we want to evaluate risk on canonical patterns of each pattern equivalence class. To do this, we need a method to enumerate these equivalence classes. The most straight forward, albeit naive, approach is to enumerate all patterns and for each pattern, check whether another pattern within the same class has appeared before it. This implementation requires at least L^M enumerations and comparisons. Although one could carry out the naive approach offline and cache equivalence classes for later usage, we show here that there is a much more elegant and efficient approach. The efficient approach comes naturally as we derive equation for calculating number of equivalence classes.

Two patterns are equivalent if and only if they are permutations of each other. This means that M exposure/gain control elements are indistinguishable, while L exposure/gain levels are distinguishable. The question of "how many equivalence classes are there" can therefore be rephrased as "how many ways are there to assign M indistinguishable elements to L distinguishable levels".

We can assign M elements to 1 level (i.e. all exposure/gain control element use same value), 2 levels, or up to min(M, L) different levels. Apparently, assigning M elements to 1 level is guaranteed to be non-redundant compared to assigning M elements to 2 levels, if we ensure that each level chosen receive at least 1 element. It is also easy to see that given a subset of chosen levels, if we replace any level within subset with another level outside of the subset, the possible assignments generated by old subset is guaranteed to be non-redundant as compared to assignments generated by new subset.

¹https://github.com/swz30/Restormer

Now, suppose we have decided to use k levels and each level must be assigned at least one element, then there are $\binom{L}{k}$ unique combinations of levels, and each combination guarantees generated assignments are non-redundant. How many non-redundant assignments are there for each combination? Well, imagine that M elements are in a list, then the question is same as how many ways there are to partition the list into k non-empty sub-lists, which is $\binom{M-1}{k-1}$. Therefore, we have $\binom{L}{k}\binom{M-1}{k-1}$ unique assignments if we use k levels. Summing k from 1 to min(M, L) yields total number of unique assignments.

Following the counting process, we obtain an algorithm that directly enumerate on equivalence classes. We show pseudo-code for an implementation in algorithm 1, 2

Al	gorithm	1	Pattern	equival	lence	class	enumeration
----	---------	---	---------	---------	-------	-------	-------------

1: procedure ENUMERATE(Levels, M) Enumerate pattern equivalence classes given allowed exposure/gain Levels and M		
	exposure/gain control elements	
2:	$locations \leftarrow [0,, M - 1]$	▷ split location for partitioning exposure/gain control elements list
3:	for $k \leftarrow 1,, min(L, M)$ do	\triangleright Use k of L levels
4:	$indices \leftarrow [0,,k-1]$	▷ index into exposure/gain level values
5:	$loc_indices \leftarrow [0,,k-1)$	\triangleright index into <i>locations</i>
6:	for $selected \leftarrow NCHOOSEK$	$(k, indices, Levels)$ do \triangleright retrieve k levels to use
7:	for $splits \leftarrow NCHOOSE$.	$K(k-1, loc_indices, locations)$ do \triangleright retrieve $k-1$ partition locations
8:	$Pattern \leftarrow [M \text{ copies of}]$	selected[0]] > initialize pattern to first selected level
9:	for $j \leftarrow [0,, k-1)$ de	▷ assign desired levels to rest of pattern
10:	$Pattern\left[splits[j]:$	$plits[j+1]] \leftarrow selected[j+1]$
11:	end for	
12:	Do something with the	enerated Pattern (e.g. evaluate risk)
13:	end for	
14:	end for	
15:	end for	
16: end procedure		

Algorithm 2 Retrieve non-redundant sub-list from a list

1: function NCHOOSEK(k, indices, optionList) \triangleright retrieve k elements from optionList using indices and update indices for next retrieval 2: $ret \leftarrow []$ 3: for $i \leftarrow 0, ..., k-1$ do ▷ Retrieve elements from *optionList* ret.append(optionList[indices[i]])4: end for 5: for $i \leftarrow k-1, ..., 0$ do > update indices for next retrieval 6: if indices[i] < optionList.length - k + i then 7: $indices[i] \leftarrow indices[i] + 1$ 8: break 9: else 10: mark indices[i] as undetermined 11: end if 12: 13: end for for $i \leftarrow 0, ..., k-1$ do 14: if indices[i] is undetermined then 15: $indices[i] \leftarrow indices[i-1] + 1$ if i > 0 else 0 16: end if 17: end for 18: 19: return ret 20: end function

REFERENCES

- S. Hajisharif, J. Kronander, and J. Unger, "Adaptive Dual ISO HDR Reconstruction," *EURASIP Journal on Image and Video Processing*, no. 41, Dec 2015.
 S. H. Chan, X. Wang, and O. A. Elgendy, "Plug-and-play ADMM for image restoration: Fixed-point convergence and applications," *IEEE Transactions*
- [2] S. H. Chan, X. Wang, and O. A. Elgendy, "Plug-and-play ADMM for image restoration: Fixed-point convergence and applications," *IEEE Transactions on Computational Imaging*, vol. 3, no. 1, pp. 84–98, 2017.
- [3] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, and M. H. Yang, "Restormer: Efficient transformer for high-resolution image restoration," in CVPR, 2022.
- [4] P. Korshunov, H. Nemoto, A. Skodras, and T. Ebrahimi, "Crowdsourcing-based Evaluation of Privacy in HDR Images," in *Optics, Photonics, and Digital Technologies for Multimedia Applications III*, P. Schelkens, T. Ebrahimi, G. Cristóbal, F. Truchetet, and P. Saarikko, Eds., vol. 9138, International Society for Optics and Photonics. SPIE, 2014, p. 913802. [Online]. Available: https://doi.org/10.1117/12.2054541
- [5] N. K. Kalantari and R. Ramamoorthi, "Deep high dynamic range imaging of dynamic scenes," ACM Transactions on Graphics (Proceedings of SIGGRAPH 2017), vol. 36, no. 4, 2017.