Autoregressive High-Order Finite Difference Modulo Imaging: High-Dynamic Range for Computer Vision Applications

Brayan Monroy[1], Kebin Contreras[2], and Jorge Bacca[1][*]

[1] Universidad Industrial de Santander, Bucaramanga, Colombia
[2] Universidad del Cauca, Popayán, Colombia
https://github.com/bemc22/AHFD

Abstract. High dynamic range (HDR) imaging is vital for capturing the full range of light tones in scenes, essential for computer vision tasks such as autonomous driving. Standard commercial imaging systems face limitations in capacity for well depth, and quantization precision, hindering their HDR capabilities. Modulo imaging, based on unlimited sampling (US) theory, addresses these limitations by using a modulo analog-to-digital approach that resets signals upon saturation, enabling estimation of pixel resets through neighboring pixel intensities. Despite the effectiveness of (US) algorithms in one-dimensional signals, their optimization problem for two-dimensional signals remains unclear. This work formulates the US framework as an autoregressive ℓ_2 phase unwrapping problem, providing computationally efficient solutions in the discrete cosine domain jointly with a stride removal algorithm also based on spatial differences. By leveraging higher-order finite differences for two-dimensional images, our approach enhances HDR image reconstruction from modulo images, demonstrating its efficacy in improving object detection in autonomous driving scenes without retraining.

Keywords: Modulo Imaging · Unlimited Sampling · Phase Unwrapping · Autonomous Driving · Object Detection.

1 Introduction

High dynamic range (HDR) imaging involves capturing the entire range of light intensity of a scene, preserving the details of the image ranging from dark to bright areas in contrast to under/overexposed images captured by standard digital cameras [10]. HDR images play an essential role in the production of high-quality images and videos, as they offer a more authentic and visually appealing representation of the real world. This is particularly important in industries such as photography [12], cinematography [30], and computer vision [17], where visual fidelity and detail are crucial [20]. Furthermore, HDR imaging of dynamic scenes reduces the dazzle effect caused by sudden changes in lighting conditions, which is essential for vehicle or traffic light recognition in autonomous driving [33,34].

However, standard digital cameras face significant challenges in capturing HDR images due to inherent deficiencies such as limited well capacity and quantization precision [28]. Several methods have been explored to capture HDR images using standard digital cameras. For instance, HDR imaging from multiple exposure acquisition involves capturing several images of the same scene at different exposure levels and then combining the pixel information from these images [9, 11, 21], which, knowing the sensor response, allows the estimation of an HDR image from these multiple low dynamic range (LDR) images [11]. While multiple exposure capture can produce high-quality HDR images, it is less effective in dynamic scenes with variable lighting, such as high-speed imaging or autonomous driving, where rapidly changing scenes make multiple exposures challenging to use. In response to these limitations, new approaches are emerging that manipulate pixel intensity as a coding strategy enabling HDR image reconstruction from a single-shot LDR image [15]. Traditional pixel coding techniques include pixel-wise variable exposure [24] or the incorporation of diffractive optical elements [3, 22, 29] or recently, modulo sensors [28, 38].

Encoding pixel intensity with smart vision sensors offers a promising strategy to reduce information loss due to saturation, particularly by incorporating the modulo operator in the processing of collected intensity values at the pixel level, also known as modulo imaging [37]. The foundation of modulo imaging involves altering the scene's intensity before sensing it and subsequently reconstructing it using computational methods. Modulo imaging takes advantage of the unlimited sampling framework (USF) [4], which states that using modulo analog-to-digital converters (modulo-ADC) that automatically reset or wrap the signal upon reaching a saturation point in continuous time, allows for the estimation of the number of times each pixel is reset using information from adjacent pixel intensities [5]. USF exploits the idea that spatial differences in a scene are the same as those in modulo measurements, where for the one-dimensional signal, the integration operators are effectively applied [6]. However, the exact optimization problem addressed within this framework, particularly for high-order finite differences in USF, remains unclear, limiting its generalization or extension to two-dimensional signals.

Due to the correlation between the discontinuities observed in modulo images and the challenges inherent in 2D phase unwrapping, various researchers have adapted phase unwrapping algorithms to address HDR modulo imaging [2, 28]. However, most of these algorithms predominantly rely on the first spatial finite difference, wasting the advantages offered by higher-order finite differences. In contrast to phase images, the inherent discontinuities in HDR images frequently violated Itoh's condition [16], which limits the image unwrapping using only the first finite-difference.

Consequently, this work proposes to formulate the USF framework as an autoregressive ℓ_2 phase unwrapping problem and provides computationally efficient analytical solutions in the discrete cosine domain. By adopting this approach, we offer guidelines on how to utilize high-order finite differences for two-dimensional images, thereby improving the reconstruction of HDR images from modulo mea-

surements. Furthermore, from a visual inspection, we notice that the proposed autoregressive method, using multiple differences, provides some stripe artifacts according to the vectorization ordering of the image. As a result, we propose to apply a stride artifact removal algorithm based on the sparsity in the spatial finite difference and employ closed-form solutions based on the cosine transform to remove the stripe artifacts in a non-iterative way. We demonstrate the potential of modulo sensors and the proposed reconstruction algorithm in object detection tasks for autonomous driving scenes, improving detection accuracy under overexposed conditions without retraining.

2   Background

The modulo-sensing process consists of analytically resetting the pixel intensity when a saturation threshold λ is reached, i.e., the photon count is reset every time the maximum λ value is reached (usually $\lambda = 2^8 - 1 = 255$) [6]. Assuming a vector form of the signal $\boldsymbol{x} \in \mathbb{R}^n$, the modulo-sensing model for modulo measurements can be mathematically defined as

$$\boldsymbol{y} = \mathcal{M}_\lambda(\boldsymbol{x}), \tag{1}$$

where $\mathcal{M}_\lambda(t) = \mathrm{mod}(t, \lambda)$ is known as the modulo operator and $\boldsymbol{y}$ is modulo measurements. Obtaining $\boldsymbol{x}$ from (1) results in an ill-posed problem due to the nonlinearity of the modulo operator.

However, following an interesting observation about the finite spatial difference between the signal $\boldsymbol{x}$ and the modulo measurements $\boldsymbol{y}$, it is possible to establish a linear equivalence, guaranteed for some bounded signals, known as Itoh's condition [16]. Specifically, under the condition that $||\Delta\boldsymbol{x}||_\infty < \lambda/2$ the following equation is maintained

$$\mathcal{M}_\lambda(\Delta\boldsymbol{y}) = \Delta\boldsymbol{x}. \tag{2}$$

This formulation of the problem on the finite difference is linear with respect to $\boldsymbol{x}$, and there are several methods available in the field of phase unwrapping to solve it [25, 36]. However, in contrast to phase images that are usually smooth, this assumption is difficult to satisfy for a range of images where some spatial differences may be larger than $\lambda/2$, as is the case for discontinuities caused by edges. Nonetheless, a more general assumption based on high-order finite differences can be made [6]

$$\mathcal{M}_\lambda(\Delta^N\boldsymbol{y}) = \Delta^N\boldsymbol{x}. \tag{3}$$

As is presented by authors in [6], and illustrated in Fig. 1 the second spatial difference or even higher order finite differences could reduce the bounded threshold, i.e., more images can satisfy $||\Delta^N\boldsymbol{x}||_\infty < \lambda/2$ for $N > 1$ instead of $N = 1$; consequently, recover the unwrapped signal can be made by iterative inversion of the high-order finite differences as is explained in the following section.
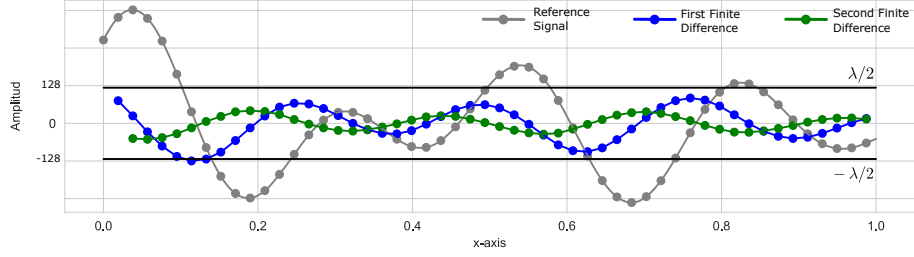
Fig. 1: Visual representation of Itoh's condition. For a given band-limited signal, the bounded threshold for the second finite difference its reduced, enabling the unwrapping of modulo samples in contrast to use the first finite difference. Figure inspired from [6].

2.1   Unlimited Sampling Framework (USF)

The solution of equation (3) can be interpreted by analyzing the information loss encountered due to the modulo operator. Specifically, we can decompose the HDR image by the addition of the modulo imaging and the unknown wrapped levels as follows

$$\boldsymbol{x} = \boldsymbol{y} + \boldsymbol{k}\lambda, \tag{4}$$

where $\boldsymbol{k} \in \mathbb{N}^n$ represents an integer-value vector that accounts for the number of times the modulo operator has been applied for each pixel. $\mathcal{M}_\lambda(\boldsymbol{k}\lambda) = 0$. This interpretation is used in the (USF) [6] to re-formulate the problem (7) to now recover $\boldsymbol{k}$ instead of $\boldsymbol{x}$ since an additional prior about the signal can be incorporated, i.e., $\boldsymbol{k}$ is an integer vector. Therefore, taking into account the assumption in (3) we have that

$$\Delta^N \boldsymbol{k} = \mathcal{M}_\lambda(\Delta^N \boldsymbol{y}) - \Delta^N \boldsymbol{y}, \tag{5}$$

where it can be iteratively solved for each difference, taking into account that $\Delta^N \boldsymbol{k} = \Delta \Delta^{N-1} \boldsymbol{k}$ using the summation operator followed by a projection of $\boldsymbol{k}$ into the integer set, as presented in [6]. The USF recovery is summarized in Algorithm 1, where Lines 1-3 compute the high-order finite differences into the modulo samples $\bar{\boldsymbol{y}}$, estimate the residual differences $\bar{\epsilon}_\lambda$ and initialize the first iteration $\boldsymbol{s}_{(1)}$, respectively. Lines 4-9 compute the iterative inversion of finite differences and estimated the associated constant bias from each iteration $\eta_{(n)}$. Finally, Line 10 computes the unwrapped signal following Equation (4). Although this methodology has worked for 1D signals, it has limitations when adapting it to the modulo image measurements to obtain the HDR images.


Algorithm 1 Unlimited Sampling Recovery

Input: $\boldsymbol{y} = \mathcal{M}_\lambda(\boldsymbol{x})$, $N \in \mathbb{N}$
Output: $\hat{\boldsymbol{x}} \approx \boldsymbol{x}$
1: Compute $\bar{\boldsymbol{y}} = \Delta^N \boldsymbol{y}$
2: Compute $\bar{\epsilon}_\lambda = \mathcal{M}_\lambda(\bar{\boldsymbol{y}}) - \bar{\boldsymbol{y}}$
3: Set $\boldsymbol{s}_{(1)} = \bar{\epsilon}_\lambda$
4: **for** $n = 1$ to $N - 1$ **do**
5:     $s_{(n+1)} = \mathbf{S}s_{(n)}$
6:     Round $s_{(n+1)}$ to $[\lambda\mathbb{Z}]$
7:     Compute $\eta_{(n)}$
8:     $s_{(n+1)} = s_{(n+1)} + \lambda\eta_{(n)}$
9: **end for**
10: Compute $\hat{\boldsymbol{x}} = \boldsymbol{y} + \mathbf{S}s_{(N)}$
11: **return** $\hat{\boldsymbol{x}}$

---

3   Proposed Method

In this section, we present the proposed method called Autoregressive High-Order Finite Difference (AHFD) to obtain HDR images from modulo measurements. The AHFD method comprises three key components: an adaptation of the spatial difference assumption in modulo measurements for matrices (Section 3.1); an iterative closed-form solution to handle high-order finite differences based on the discrete cosine transform (Section 3.2); and stripe artifact removal based on sparsity in the spatial differences (Section 3.3). A visual representation of the AHFD method is illustrated in Fig 2.
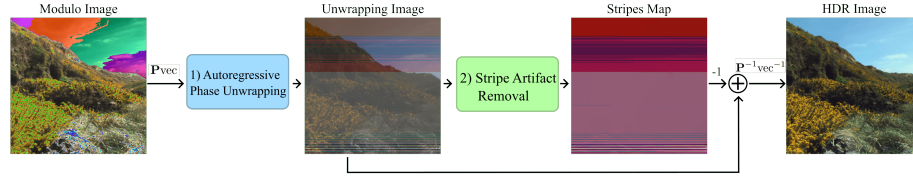
Fig. 2: Proposed AHFD method for HDR image restoration from modulo measurements composed of three components: 1) Autoregressive phaseunwrapping algorithm, and 2) Stripe Artifact Removal, 3) The operator **Pvec** to adapt AHFD for matrices.

3.1   Adaptation of USF for Modulo Images

To use USF recovery on 2D modulo measurements, we need to adapt the unwrapping Algorithm 1 of 1d signals for 2d signals. A straightforward alternative could consist of applying the Algorithm 1 row-wise or column-wise for each modulo image, however, this approach generates multiple offset artifacts that destroy the spatial image structure [2]. We propose vectorizing the image and addressing the unwrapping problem for the entire scene to mitigate multiple offsets. However, the traditional vectorization of the matrix $\boldsymbol{x} = \text{vec}(\mathbf{X})$, which stacks the columns of the matrix $\mathbf{X}$ on top of one another, increases the likelihood that condition $||\Delta\boldsymbol{x}||_\infty < \lambda/2$ is not met due to border discontinuities, as illustrated

in Fig. 3. Consequently, we proposed an neighborhood vectorization of the form

$$\boldsymbol{x} = \mathbf{P}\mathrm{vec}(\mathbf{X}), \tag{6}$$

where $\mathbf{P}$ is a permutation matrix that orders the pixels of the image. Equation (6) guarantees that the contiguous pixels in the 1d signal are associated with the contiguous pixels in the 2d images. With this vectorization trick, we evaluated the 1d unwrapping recovery as

$$\tilde{\boldsymbol{x}} \in \arg\min_{\boldsymbol{x}} ||\Delta^N \boldsymbol{x} - \mathcal{M}_\lambda(\Delta^N \boldsymbol{y})||_2^2, \tag{7}$$

where the HDR image is obtained as $\tilde{\mathbf{X}} = \mathbf{P}^{-1}\mathrm{vec}^{-1}(\tilde{\boldsymbol{x}})$. Specifially, we proposed two vectorization orderings, following horizontal and vertical vectorization paths as is illustrated in Fig 3.
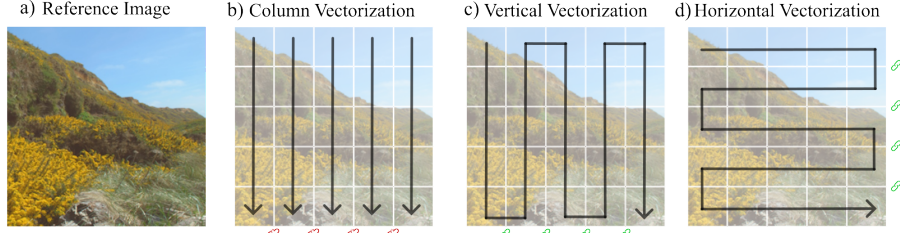
Fig. 3: Pixel neighborhood vectorization. a) Reference image, b) column vectorization order using the vec($\cdot$) operator, c-d) proposed pixel neighborhood vectorization $\mathbf{P}\mathrm{vec}(\cdot)$ which follows either a vertical or horizontal trajectory. Column vectorization creates artificial discontinuities at the end of each column, whereas our strategy maintains a next-pixel vectorization path.

3.2   Autoregressive phase unwrapping

Similarly to [4], we proposed to estimate $\boldsymbol{k}$ instead of $\boldsymbol{x}$ based on the assumption presented in Eq (4) . Therefore, we proposed the following optimization formulation

$$\arg\min_{\boldsymbol{k} \in \mathbb{Z}} ||\Delta^N \boldsymbol{k} - (\mathcal{M}_\lambda(\Delta^N \boldsymbol{y}) - \Delta^N \boldsymbol{y})||_2^2, \tag{8}$$

using Eq (3), i.e., $\mathcal{M}_\lambda(\Delta^N \boldsymbol{y}) = \Delta^N \boldsymbol{x}$. In this sense, we explore the formulation of high-order finite differentiation from USF recovery as an autoregressive phase unwrapping problem. Specifically, based on the fact that $\Delta^N \boldsymbol{k} = \Delta\Delta^{N-1}\boldsymbol{k} = \Delta s^{(1)}$, we can reformulate the problem as a subset of iterative optimization problems where

$$\boldsymbol{s}^{(n+1)} = \arg\min_{\Delta^{N-1-n}\boldsymbol{k} \in \mathbb{Z}} ||\Delta(\Delta^{N-1-n}\boldsymbol{k}) - \boldsymbol{s}^{(n)}||_2^2 \quad \text{for } n = 0, \cdots, N-1, \tag{9}$$

with $\boldsymbol{s}^{(0)} = \mathcal{M}_\lambda(\Delta^N \boldsymbol{y}) - \Delta^N \boldsymbol{y}$ and $\boldsymbol{s}^{(N)} \approx \boldsymbol{k}$. Interestingly, each sub-problem can also be solved for each component using the DCT transform as the conventional

phase unwrapping recovery [26], taking the form

$$s_i^{(n+1)} = \mathcal{D}^{-1}\Big(\frac{\mathcal{D}(\Delta^\top \boldsymbol{s}^{(n)})_i}{2\cos(\pi i/N) - 1}\Big). \tag{10}$$

Similarly to Algorithm 1, the projection operator into integer set $\mathbb{Z}$ can be employed in each iteration. Our proposed autoregressive phase unwrapping is summarized in Algorithm 1.

Algorithm 1 Autoregressive Phase Unwrapping

Input: $\boldsymbol{y} = \mathcal{M}_\lambda(\boldsymbol{x})$, $N \in \mathbb{N}$
Output: $\hat{\boldsymbol{x}} \approx \boldsymbol{x}$
 1: Compute $\bar{\boldsymbol{y}} = \Delta^N \boldsymbol{y}$
 2: Compute $\bar{\epsilon}_\lambda = \mathcal{M}_\lambda(\bar{\boldsymbol{y}}) - \bar{\boldsymbol{y}}$
 3: Set $\boldsymbol{s}^{(0)} = \bar{\epsilon}_\lambda$
 4: **for** $n = 0$ to $N - 1$ **do**
 5:    $s_i^{(n+1)} = \mathcal{D}^{-1}\Big(\frac{\mathcal{D}(\Delta^\top \boldsymbol{s}^{(n)})_i}{2\cos(\pi i/N) - 1}\Big)$
 6:    Project $\boldsymbol{s}^{(n+1)}$ into the integer $\mathbb{Z}$
 7: **end for**
 8: Compute $\hat{\boldsymbol{x}} = \boldsymbol{y} + \boldsymbol{s}^{(N)}\lambda$
 9: **return** $\hat{\boldsymbol{x}}$

3.3 Stripe Artifact Removal

Our HDR method from modulo measurements in Section 3.2, similar to state-of-the-art methods [2,4,26], relies heavily on the assumption of band-limited signals, i.e., $\|\Delta\boldsymbol{x}\|_\infty < \lambda/2$ is satisfied. When this assumption is violated, ambiguities arise in the high-order finite differences associated with modulo or natural image edges, leading to visual artifacts as shown in Fig.4. However, we observe that the artifacts generated using pixel neighborhood vectorization are similar to those found in striping noise [8, 31], which are more manageable than artifacts from other phase unwrapping methods, as illustrated in Fig.4. Consequently, we propose to address this problem after Algorithm 1. Specifically, the presence of stripe artifacts in the estimated HDR image can be mathematically modeled as

$$\tilde{\boldsymbol{x}} = \boldsymbol{x} + \boldsymbol{s}. \tag{11}$$

Following the idea of spatial difference, we assume that the stripes are sparse and exhibit high frequency in the spatial difference dimensions. Consequently, we propose estimating the stripes by solving the following optimization problem:

$$\tilde{\boldsymbol{s}} \in \underset{\boldsymbol{s}\in\mathbb{R}}{\arg\min}\quad \|\Delta_{xy}\tilde{\boldsymbol{x}} - \Delta_{xy}(\boldsymbol{x} + \boldsymbol{s})\|_2^2 + \gamma\|\Delta_{x,y}\boldsymbol{s}\|_0, \tag{12}$$

where $\Delta_{xy} = [\Delta_x^\top, \Delta_y^\top]^\top$. This problem represents a conventional $\ell_2 - \ell_0$ optimization, wherein compressive sensing techniques advocate substituting for a $\ell_2 - \ell_1$ formulation and resolving the equations through iterative methods [7,23]. However, following the idea of a non-iterative process [1], we propose to assume
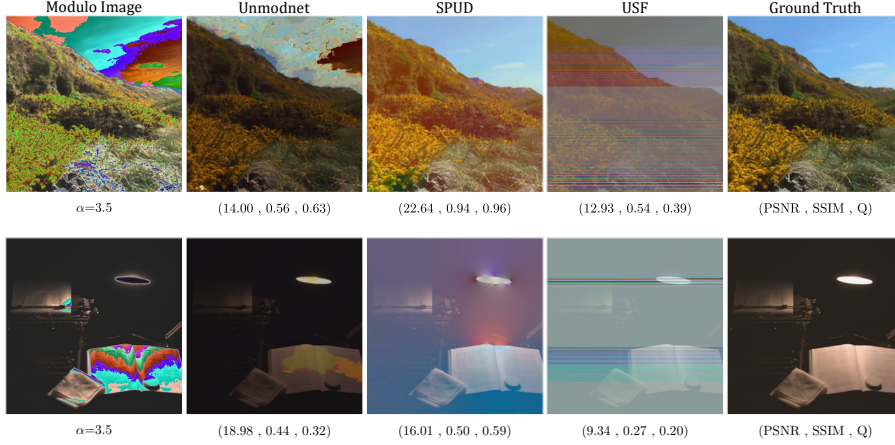
Fig. 4: Artifacts from various image unwrapping algorithms. In Undmodnet [38], overestimating unwrapping levels results in "stains" in the images. SPUD [25] when failure to maintain Itoh's condition causes "light leaks" around synthetic wraps at the image borders. Alternatively, USF [6] artifacts produce structural lines that overlap with correctly unwrapped images, often referred to as "stripes" in remote sensing images.

that $\Delta_{xy}\boldsymbol{s} = \Delta_{xy}(\tilde{\boldsymbol{x}} - \boldsymbol{x}) = \mathcal{H}_\gamma(\tilde{\boldsymbol{x}})$, where $\mathcal{H}_\gamma$ is the hard thresholding operator defined element-wise as $\mathcal{H}_\gamma(\boldsymbol{x})_i = \boldsymbol{x}_i \cdot \mathbf{1}_{|\boldsymbol{x}| \leq \gamma}$ for an input vector $\boldsymbol{x}$. Consequently, Eq. (12) can now be expressed as.

$$\tilde{\boldsymbol{s}} \in \underset{\boldsymbol{s} \in \mathbb{R}}{\arg\min} \quad \|\Delta_{xy}\mathcal{H}_\gamma(\tilde{\boldsymbol{x}}) - \Delta_{xy}\boldsymbol{s}\|_2^2. \tag{13}$$

Interestingly, this problem results in a 2D generalization of a single problem such as in Eq (9) and can be solved using 2D DCT [26] instead of 1D DCT as summarized in Algorithm 2. With the inclusion of Algorithm 2, our proposed method significantly overcomes the performance in cases where modulo measurements present high ambiguity between modulo edges and image edges.

Finally, the proposed modulo recovery algorithm that combines a vectorization operator $\mathbf{P}$, the auto-regressive phase unwrapping algorithm, and the strip artifact removal algorithm are summarized in Algorithm 3. The proposed method computes ordered vectorization $\boldsymbol{y}$ for a given modulo image $\mathbf{Y}$ (Line 1) and subsequently applies Algorithms 1-2 (Lines 2-3); finally, the estimated vector is reshaped to match the image dimensions. For color images, Algorithm 3 is broadcast along the channels dimensions. In terms of time complexity, the main computational bottleneck manifests itself during iterative integration within the Discrete Cosine Transform (DCT) domain for the vectorized modulo image, as detailed in Algorithm 1. This process exhibits a computational complexity of $\mathcal{O}(Kn\log(n))$, where n denotes the total pixel count of the modulo image and N represents the number of finite high-order differences.

Algorithm 2 Stripe Artifact Removal Algorithm

Input: $\tilde{\boldsymbol{x}}, \gamma$
Output: $\hat{\boldsymbol{x}} \approx \boldsymbol{x}$
1: $\boldsymbol{\rho} = \texttt{DTC-2D}(\Delta_{xy}^\top \mathcal{H}_\gamma(\Delta_{xy}\tilde{\boldsymbol{x}}))$
2: $\tilde{\boldsymbol{n}}_{mn+n} = \texttt{DTC-2D}^{-1}(\boldsymbol{\rho} \cdot [2\cos(\frac{\pi m}{M}) + 2\cos(\frac{\pi n}{N}) - 4]^{-1})$
3: $\hat{\boldsymbol{x}} = \tilde{\boldsymbol{x}} - \tilde{\boldsymbol{n}}$
4: **return** $\hat{\boldsymbol{x}}$

Algorithm 3 Proposed Modulo Recovery Algorithm

Input: $\mathbf{Y} = \mathcal{M}_\lambda(\mathbf{X})$, $N \in \mathbb{N}, \gamma$
Output: $\tilde{\boldsymbol{X}} \approx \mathbf{X}$
1: Computes $\boldsymbol{y} = \mathbf{P}\texttt{vec}(\mathbf{Y})$
2: Computes $\hat{\boldsymbol{x}}$ from Algorithm 1.   #Autoregressive Phase Unwrapping
3: Computes $\tilde{\boldsymbol{x}}$ from Algorithm 2.   #Stripes Artifact Removal
4: Computes $\tilde{\mathbf{X}} = \texttt{vec}^{-1}(\mathbf{P}^{-1}\tilde{\boldsymbol{x}})$
5: **return** $\tilde{\mathbf{X}}$

---

4 Simulations and Results

This section presents a detailed analysis of the performance and efficacy of the proposed methodology through various experiments on HDR image restoration and object detection in autonomous driving scenes. The following subsections will detail the experimental settings, the evaluation metrics used for the assessment, and a comprehensive overview of the results obtained for both tasks.

For the validation of the proposed recovery method in HDR image reconstruction from modulation measurements, we conducted a comparison with state-of-the-art optimization and deep learning-based unwrapping algorithms. Specifically, for optimization algorithms, we selected the simultaneous phase unwrapping and denoising algorithm (SPUD) [25] and the Plug-and-Play Unwrapping Algorithm (PnP-UA) [2]; they proposed non-iterative and deep prior iterative solutions for the phase unwrapping problem based on the first spatial finite difference, respectively. In the case of deep learning-based recovery network, we select Unmodnet [38], which employs an iterative estimation of the binary mask to estimate the wrapping levels for modulo measurements and includes Laplacian in the modulo image as guided information.

On the other hand, we employed YOLOv10x [32] directly to the measurements to evaluate the object detection task. Specifically, we simulate acquisition of saturated images with a CCD sensor, modulo measurements, and finally, HDR image reconstruction using the proposed method. For this experiment, the network weights are the same for all scenarios and were obtained from [32].

4.1 Validation on HDR Image Restoration.

Dataset. We use the Korshunov dataset as a benchmark for HDR image restoration. This dataset includes 20 HDR images with resolutions varying from full HD (1920×1080) to greater than 4K (6032×4018) and 12 bits per color channel [18]. The images encompass various scenes, dynamic ranges, and acquisition

techniques, featuring architecture, landscapes, portraits, and CGI-generated visuals. All images were resized to a resolution of 1024×1024. It is important to highlight that the proposed method is non-data dependent. Therefore, this dataset is only used for testing purposes.

Metrics. To evaluate the quality of the unwrapped restoration, we utilized three specific metrics: Q-index, peak signal-to-noise ratio (PSNR), and structural similarity index (SSIM). The Q-index is frequently used in phase unwrapping methods, whereas PSNR and SSIM are common image quality metrics. All metrics were applied within the HDR range as per [25]. The Reinhard tone mapping function is used for image visualization [27].

Simulations. The ability of the proposed reconstruction method was evaluated by comparing it against SPUD [25], Unmodnet [38], and PnP-UA [2]. For this experiment, different illumination conditions were evaluated by changing the saturation faction under the synthetic testing dataset. The generation of this dataset is represented as

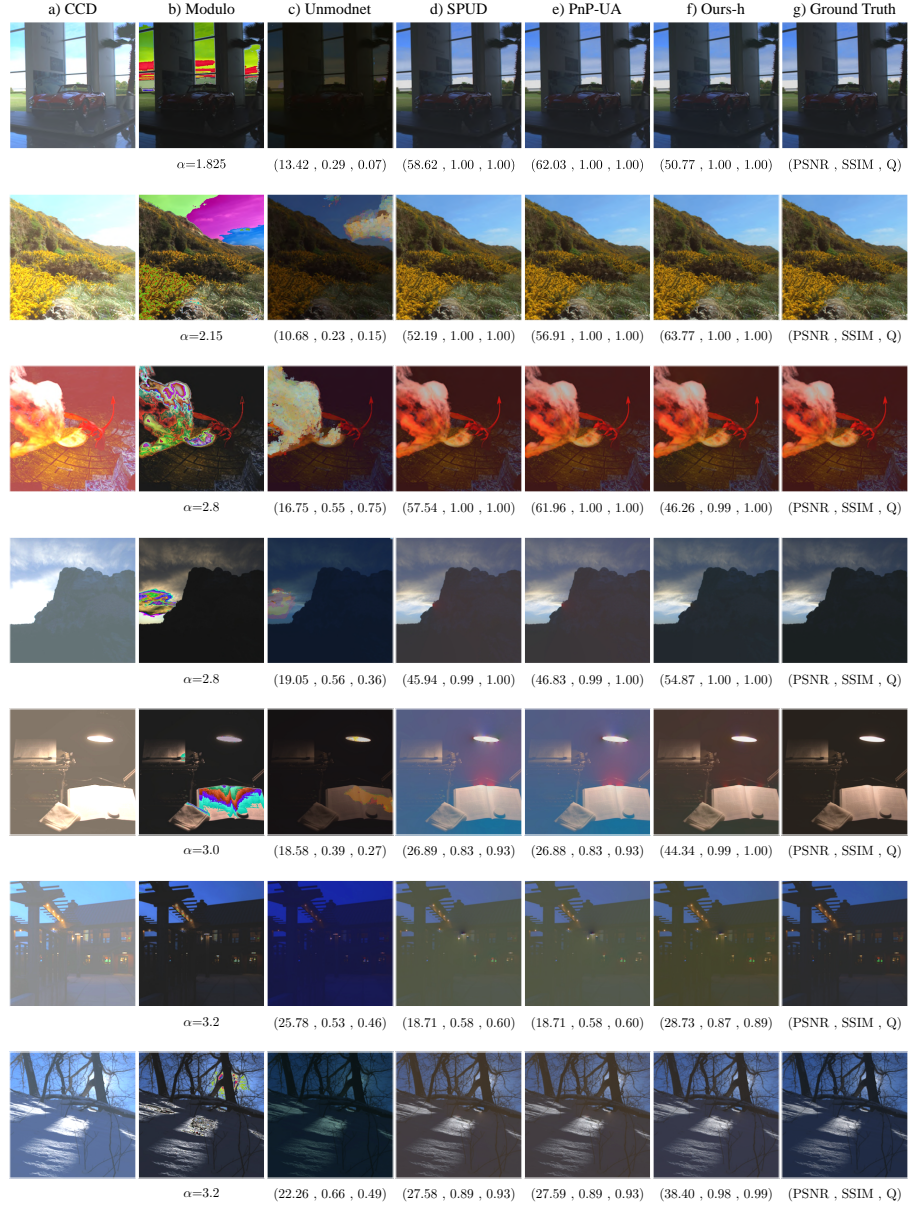$$\mathbf{Y} = \mathcal{M}_\lambda(\alpha\mathbf{X}), \tag{14}$$

where $\mathbf{X}$ is a normalized HDR image within the range [0, 1] with three channels (RGB) and 1024×1024 spatial resolution, $\alpha > 1$ is the saturation factor, and $\lambda = 1$. Our evaluations cover a variety of saturation factors, specifically $\alpha \in \{1.825, 2.15, 2.475, 2.8, 3.0, 3.2\}$.

Quantitative results are summarized in Table 1. Specifically, our proposed method delivers performance comparable to SPUD and PnP-UA for α values between [2.15, 2.475] and achieves better results for $\alpha \geq 2.8$. For deep learning methods such as Unmodnet, we observed a poor generalization across all illumination conditions. These quantitative results are supported by Figure 5, which

Table 1: Quantitative comparison results using different saturation factors for state-of-the-art recovery methods from modulo measurements. Our-h and Ours-v stands for our proposed recovery algorithm with horizontal and vertical vectorization, respectively.

| Metric | Method | Saturation | Factor | | | | |
		1.825	2.15	2.475	2.8	3.0	3.2
PSNR (↑)	Unmodnet [38]	22.77	21.93	19.63	19.26	18.78	19.29
	SPUD [25]	57.63	57.48	47.24	36.32	32.55	29.07
	PnP-UA [2]	61.26	**60.85**	**49.78**	37.01	33.11	29.53
	Ours-h	60.37	59.80	44.58	**43.19**	36.26	32.81
	Ours-v	**61.86**	59.24	40.03	43.61	**37.96**	**34.05**
SSIM (↑)	Unmodnet [38]	0.6844	0.6907	0.5602	0.5657	0.5073	0.5897
	SPUD [25]	0.9993	0.9993	0.9861	0.9299	0.8826	0.8410
	PnP-UA	0.9995	**0.9996**	**0.9869**	0.9310	0.8837	0.8422
	Ours-h	0.9994	0.9994	0.9735	**0.9913**	0.9607	**0.9364**
	Ours-v	**0.9996**	0.9992	0.8697	0.9742	**0.9333**	0.8791
Q-score (↑)	Unmodnet [38]	0.6194	0.6420	0.4853	0.4852	0.4189	0.5010
	SPUD [25]	0.9999	0.9999	0.9904	0.9400	0.9027	0.8527
	PnP-UA [2]	0.9999	0.9999	**0.9906**	0.9404	0.9031	0.8530
	Ours-h	0.9999	**0.9999**	0.9777	**0.9957**	0.9412	**0.9418**
	Ours-v	**1.0000**	**0.9999**	0.7547	0.9783	0.8933	0.8265

a) CCD	b) Modulo	c) Unmodnet	d) SPUD	e) PnP-UA	f) Ours-h	g) Ground Truth
α=1.825		(13.42 , 0.29 , 0.07)	(58.62 , 1.00 , 1.00)	(62.03 , 1.00 , 1.00)	(50.77 , 1.00 , 1.00)	(PSNR , SSIM , Q)
α=2.15		(10.68 , 0.23 , 0.15)	(52.19 , 1.00 , 1.00)	(56.91 , 1.00 , 1.00)	(63.77 , 1.00 , 1.00)	(PSNR , SSIM , Q)
α=2.8		(16.75 , 0.55 , 0.75)	(57.54 , 1.00 , 1.00)	(61.96 , 1.00 , 1.00)	(46.26 , 0.99 , 1.00)	(PSNR , SSIM , Q)
α=2.8		(19.05 , 0.56 , 0.36)	(45.94 , 0.99 , 1.00)	(46.83 , 0.99 , 1.00)	(54.87 , 1.00 , 1.00)	(PSNR , SSIM , Q)
α=3.0		(18.58 , 0.39 , 0.27)	(26.89 , 0.83 , 0.93)	(26.88 , 0.83 , 0.93)	(44.34 , 0.99 , 1.00)	(PSNR , SSIM , Q)
α=3.2		(25.78 , 0.53 , 0.46)	(18.71 , 0.58 , 0.60)	(18.71 , 0.58 , 0.60)	(28.73 , 0.87 , 0.89)	(PSNR , SSIM , Q)
α=3.2		(22.26 , 0.66 , 0.49)	(27.58 , 0.89 , 0.93)	(27.59 , 0.89 , 0.93)	(38.40 , 0.98 , 0.99)	(PSNR , SSIM , Q)

Fig. 5: Comparison with state-of-the-art recovery methods from the modulo images. (a-b) correspond to saturated and modulo measurements under different intensity levels, respectively. (c-f) Corresponding to different recovery methods and the proposed Ours-h method, finally, (g) corresponds to the ground truth image.

shows that our method surpasses state-of-the-art methods under almost all illumination conditions. At high saturation, it effectively mitigates unwrapping artifacts in areas with high wrapping concentration, such as near light sources like lamps or the Sun. Despite this, there is still room for improvement, as seen in the case of $\alpha = 3.0$ in Figure 5, where unwrapped artifacts are still present in our proposed method.

Finally, we present a computational resource consumption analysis in Table 2. CNN-based methods like Undmodnet and PnP-UA have high memory usage and long running times due to deep neural model parameters, with Undmodnet being the slowest. In contrast, SPUD is the fastest and most memory-efficient due to its non-iterative and low hyperparameter design. Our method is the second fastest and most memory-efficient, with significant performance improvement. This analysis underscores the limitations of CNN-based HDR image restoration on edge devices, whereas our algorithm offers a computationally efficient alternative with competitive HDR restoration results.

Table 2: Time and Memory consumption between proposed and state-of-the-art HDR modulo imaging methods. All method were tested on a RTX 3090 and R5 5600X CPU.

Method	Undmodnet	PnP-UA	SPUD	Ours-h/v
Running Time [ms]	5327	456	5.51	8.34
VRAM [MiB]	2697	6717	70	151

4.2   Validation on Object Detection for Autonomous Driving.

Dataset. We select the KITTI database, a reference in object detection task due to its focus on autonomous driving. This database contains 7,481 images with dimensions of 1242×375 and covers urban scenes with a variety of objects such as vehicles, pedestrians, cyclists, cars, cyclists, miscellaneous objects, pedestrians, people sitting, trams, trucks, and vans, providing a real challenge for vision-based object detection systems [13].

Model. The efficacy of object detection was evaluated using the YOLOv10x model, a neural network architecture designed for real-time object detection with 29.5 million parameters [32]. This model consists in a convolutional neural network backbone with multiple detection layers that utilize anchor boxes to predict bounding boxes for objects within an image. The architecture of YOLOv10x incorporates a consistent dual assignment design that eliminates the need for non-maximum suppression during inference, improving efficiency. Additionally, it employs a lightweight classification head, spatial-channel decoupled downsampling, and a rank-guided block design to decrease computational redundancy and enhance efficiency. It also uses large kernel convolution and a partial self-attention module to improve feature extraction capability and localization accuracy.

Metrics. The performance of the object detection task is evaluated using three specific metrics: Accuracy, Intersection over Union (IoU), and the F1 score. These metrics serve as standards in the evaluation of object detection and enable an objective comparison of the performance of the YOLOv10x model under different image manipulation conditions.

Simulations. The results were obtained by varying the saturation factor α with three specific values: $\alpha = \{1.5, 2, 3\}$, in accordance with Equation 14, which adjusted the saturation intensity of the images. The images from the KITTI dataset were processed by simulating four scenarios and then used as input in YOLOv10x to obtain the boundary detection boxes. Specifically, we simulated the acquisition of i) **CCD**, which refers to the saturated image based on the specific saturation factor, ii) **Modulo**, which refers to the modulo measurements obtained using Eq. 14, iii) **Recovery**, which refers to the reconstructed HDR from the modulo measurements using Algorithm 3, and iv) **Ideal HDR**, which refers to the original image without modification, assuming a perfect acquisition of the HDR image.

The numerical findings are presented in Table 3. In particular, we notice that when α increases, the CCD results decrease drastically since more spatial zones are lost due to saturation. However, *Modulo* and *Recovery* provide better results and a lower drop in the metrics. Specifically, for small saturation levels $\alpha = 1.5$, both *Modulo* and *Recovery* outperform the *CCD* results where, for the *Recovery*, provide similar results to ideal HDR. The visual results in Fig 6 further corrob-

Table 3: Object Detection Evaluation with Different Saturation Levels

Method	Metrics								
	$\alpha = 1.5$			$\alpha = 2$			$\alpha = 3$		
	IOU ($\uparrow$)	F1 ($\uparrow$)	ACC ($\uparrow$)	IOU ($\uparrow$)	F1 ($\uparrow$)	ACC ($\uparrow$)	IOU ($\uparrow$)	F1 ($\uparrow$)	ACC ($\uparrow$)
CCD	64.69	47.05	36.70	61.56	45.67	34.45	59.60	37.98	28.29
Modulo	74.59	61.15	50.41	73.04	58.86	47.93	70.40	55.42	44.70
Recovery	**74.61**	**61.25**	**50.64**	**74.44**	**60.95**	**50.28**	**71.06**	**55.86**	**44.92**
Ideal HDR	75.33	63.52	52.57	75.33	63.52	52.57	75.33	63.52	52.57

Fig. 6: Comparative of input modalities applied to a real urban scene with YOLO object detection. The first row shows images with saturation $\alpha = 1.5$, the second with $\alpha = 2$, and the third with $\alpha = 3$. Each column represents different methods: GT, CDD, Modulo, and Recovery, visually demonstrating how saturation influences the detections.

orate these quantitative results. Interestingly, for the most challenging scenario $\alpha = 3$, the proposed *Recovery* method does not provide perfect recovery; however, it allows us to estimate better the object in the scene compared with CCD or raw modulo measurements. This general performance of the modulo sensor is based on addressing the saturation areas typically generated by excessive lighting from sources like car headlights or the sun. Information about these areas is often lost in the saturated image. In contrast, the modulo image captures these details, and our proposed method effectively recovers the wrapping levels, thereby preserving crucial information and improving object detection.

4.3   Discussion and Future Work

We suggest that the proposed autoregressive phase unwrapping formulation within the USF framework, combined with incorporating pixel neighborhood vectorization for HDR modulo imaging, opens the way for innovative optimization-based unwrapping algorithms. These algorithms can use well-established signal priors in natural images, enhancing optimization strategies derived from the phase unwrapping problem. For instance, a promising research direction is the incorporation of denoising regularization such as that indicated for the first spatial finite difference [2,25]. This approach leverages powerful denoising algorithms and image-denoising neural networks to enhance HDR modulo imaging performance under noisy conditions without model retraining.

Furthermore, since our proposed method handles multi-channel and multi-frame HDR image restoration, the buffering of multi-frame images or burst images could further enhance HDR image quality for HDR video restoration, as presented in the literature [14,19,35]. Finally, an intriguing research path could involve the joint and iterative estimation of the HDR image and the stripe artifact map. This could be achieved by incorporating iterative image de-striping algorithms like those used in remote sensing images. Even more, one could explore different ordering paths in the image vectorization by neighborhood pixels. This alternative exploits the specific spatial structures presented in each image. Our experimental results indicate that no ordering path yields the best performance; it varies based on the image structure, wrapping location, and specific saturation factor.

5   Conclusion

This work introduces an autoregressive phase unwrapping method for HDR image reconstruction from modulo measurements. It employs high-level finite differences for 2D images, providing efficient solutions in the discrete cosine domain with a stride removal algorithm based on spatial sparsity. We show improved HDR restoration performance and evaluate our method against state-of-the-art optimization and deep learning HDR restoration algorithms using first-finite differences. We highlight the potential of modulo-ADC sensors and our HDR recovery algorithm in enhancing YOLO detector accuracy under overexposed conditions without retraining.

References

1. Bacca, J., Correa, C.V., Arguello, H.: Noniterative hyperspectral image reconstruction from compressive fused measurements. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing **12**(4), 1231–1239 (2019) 7

2. Bacca, J., Monroy, B., Arguello, H.: Deep plug-and-play algorithm for unsaturated imaging. In: ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 2460–2464. IEEE (2024) 2, 5, 7, 9, 10, 14

3. Baek, S.H., Ikoma, H., Jeon, D.S., Li, Y., Heidrich, W., Wetzstein, G., Kim, M.H.: Single-shot hyperspectral-depth imaging with learned diffractive optics. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 2651–2660 (2021) 2

4. Bhandari, A., Krahmer, F., Poskitt, T.: Unlimited sampling from theory to practice: Fourier-prony recovery and prototype adc. IEEE Transactions on Signal Processing **70**, 1131–1141 (2021) 2, 6, 7

5. Bhandari, A., Krahmer, F., Raskar, R.: On unlimited sampling. In: 2017 International Conference on Sampling Theory and Applications (SampTA). pp. 31–35. IEEE (2017) 2

6. Bhandari, A., Krahmer, F., Raskar, R.: On unlimited sampling and reconstruction. IEEE Transactions on Signal Processing **69**, 3827–3839 (2020) 2, 3, 4, 8

7. Bioucas-Dias, J.M., Figueiredo, M.A.: A new twist: Two-step iterative shrinkage/thresholding algorithms for image restoration. IEEE Transactions on Image processing **16**(12), 2992–3004 (2007) 7

8. Cao, Y., Yang, M.Y., Tisse, C.L.: Effective strip noise removal for low-textured infrared images based on 1-d guided filtering. IEEE transactions on circuits and systems for video technology **26**(12), 2176–2188 (2015) 7

9. Cho, W.h., Hong, K.S.: Extending dynamic range of two color images under different exposures. In: Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004. vol. 4, pp. 853–856. IEEE (2004) 2

10. Debevec, P.E., Malik, J.: Recovering high dynamic range radiance maps from photographs. In: Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques. p. 369–378. SIGGRAPH '97, ACM Press/Addison-Wesley Publishing Co., USA (1997). https://doi.org/10.1145/258734.258884, https://doi.org/10.1145/258734.258884 1

11. Debevec, P.E., Malik, J.: Recovering High Dynamic Range Radiance Maps from Photographs. Association for Computing Machinery, New York, NY, USA, 1 edn. (2023), https://doi.org/10.1145/3596711.3596779 2

12. Fairchild, M.D.: The hdr photographic survey. In: Color and imaging conference. vol. 15, pp. 233–238. Society of Imaging Science and Technology (2007) 1

13. Geiger, A., Lenz, P., Urtasun, R.: Are we ready for autonomous driving? the kitti vision benchmark suite. In: 2012 IEEE conference on computer vision and pattern recognition. pp. 3354–3361. IEEE (2012) 12

14. Hasinoff, S.W., Sharlet, D., Geiss, R., Adams, A., Barron, J.T., Kainz, F., Chen, J., Levoy, M.: Burst photography for high dynamic range and low-light imaging on mobile cameras. ACM Transactions on Graphics (ToG) **35**(6), 1–12 (2016) 14

15. Hirakawa, K., Simon, P.M.: Single-shot high dynamic range imaging with conventional camera hardware. In: 2011 International Conference on Computer Vision. pp. 1339–1346. IEEE (2011) 2

16. Itoh, K.: Analysis of the phase unwrapping algorithm. Applied optics **21**(14), 2470–2470 (1982) 2, 3

17. Johnson, G.M., Fairchild, M.D.: Rendering hdr images. In: Color and Imaging Conference. vol. 11, pp. 36–41. Society of Imaging Science and Technology (2003) 1

18. Korshunov, P., Hanhart, P., Richter, T., Artusi, A., Mantiuk, R., Ebrahimi, T.: Subjective quality assessment database of hdr images compressed with jpeg xt. In: 2015 seventh international workshop on quality of multimedia experience (QoMEX). pp. 1–6. IEEE (2015) 9

19. Li, Y., Qiao, Y., Ruichek, Y.: Multiframe-based high dynamic range monocular vision system for advanced driver assistance systems. IEEE Sensors Journal **15**(10), 5433–5441 (2015) 14

20. McCann, J.J., Rizzi, A.: The art and science of HDR imaging. John Wiley & Sons (2011) 1

21. Mertens, T., Kautz, J., Van Reeth, F.: Exposure fusion: A simple and practical alternative to high dynamic range photography. In: Computer graphics forum. vol. 28, pp. 161–171. Wiley Online Library (2009) 2

22. Metzler, C.A., Ikoma, H., Peng, Y., Wetzstein, G.: Deep optics for single-shot high-dynamic-range imaging. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1375–1385 (2020) 2

23. Monroy, B., Bacca, J., Arguello, H.: Jr2net: a joint non-linear representation and recovery network for compressive spectral imaging. Applied Optics **61**(26), 7757–7766 (2022) 7

24. Nayar, S.K., Mitsunaga, T.: High dynamic range imaging: Spatially varying pixel exposures. In: Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No. PR00662). vol. 1, pp. 472–479. IEEE (2000) 2

25. Pineda, J., Bacca, J., Meza, J., Romero, L.A., Arguello, H., Marrugo, A.G.: Spud: simultaneous phase unwrapping and denoising algorithm for phase imaging. Applied Optics **59**(13), D81–D88 (2020) 3, 8, 9, 10, 14

26. Ramirez, J., Arguello, H., Bacca, J.: Phase unwrapping for phase imaging using the plug-and-play proximal algorithm. Applied Optics **63**(2), 535–542 (2024) 7, 8

27. Reinhard, E., Stark, M., Shirley, P., Ferwerda, J.: Photographic tone reproduction for digital images. In: Seminal Graphics Papers: Pushing the Boundaries, Volume 2, pp. 661–670 (2023) 10

28. So, H.M., Martel, J.N., Wetzstein, G., Dudek, P.: Mantissacam: Learning snapshot high-dynamic-range imaging with perceptually-based in-pixel irradiance encoding. In: 2022 IEEE International Conference on Computational Photography (ICCP). pp. 1–12. IEEE (2022) 2

29. Sun, Q., Tseng, E., Fu, Q., Heidrich, W., Heide, F.: Learning rank-1 diffractive optics for single-shot high dynamic range imaging. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 1386–1396 (2020) 2

30. Tocci, M.D., Kiser, C., Tocci, N., Sen, P.: A versatile hdr video production system. ACM Transactions on Graphics (TOG) **30**(4), 1–10 (2011) 1

31. Tsai, F., Chen, W.W.: Striping noise detection and correction of remote sensing images. IEEE Transactions on Geoscience and remote sensing **46**(12), 4122–4131 (2008) 7

32. Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J., Ding, G.: Yolov10: Real-time end-to-end object detection. arXiv preprint arXiv:2405.14458 (2024) 9, 12

33. Wang, J.G., Zhou, L.B.: Traffic light recognition with high dynamic range imaging and deep learning. IEEE Transactions on Intelligent Transportation Systems **20**(4), 1341–1352 (2018) 1

34. Wang, J.G., Zhou, L., Song, Z., Yuan, M.: Real-time vehicle signal lights recognition with hdr camera. In: 2016 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData). pp. 355–358. IEEE (2016) 1

35. Xiao, J., Ye, Q., Liu, T., Zhang, C., Lam, K.M.: Deep progressive feature aggregation network for multi-frame high dynamic range imaging. Neurocomputing **594**, 127804 (2024) 14

36. Yu, H., Lan, Y., Yuan, Z., Xu, J., Lee, H.: Phase unwrapping in insar: A review. IEEE Geoscience and Remote Sensing Magazine **7**(1), 40–58 (2019) 3

37. Zhao, H., Shi, B., Fernandez-Cull, C., Yeung, S.K., Raskar, R.: Unbounded high dynamic range photography using a modulo camera. In: 2015 IEEE International Conference on Computational Photography (ICCP). pp. 1–10. IEEE (2015) 2

38. Zhou, C., Zhao, H., Han, J., Xu, C., Xu, C., Huang, T., Shi, B.: Unmodnet: Learning to unwrap a modulo image for high dynamic range imaging. Advances in Neural Information Processing Systems **33**, 1559–1570 (2020) 2, 8, 9, 10