Specularity Factorization for Low-Light Enhancement

Saurabh Saini
CVIT, KCIS
IIIT-Hyderabad, India
saurabh.saini@research.iiit.ac.in

P J Narayanan
CVIT, KCIS
IIIT-Hyderabad, India
pjn@iiit.ac.in

Figure 1. **Sepcularity Factorization**: We factorize a single input image (blue box, top row) into multiple *soft* specular factors (rescaled for visualization) based on their similar illumination characteristics (note table shadow and lamp reflection). Our factors directly enable zero-reference low-light enhancement and user controlled image relighting (bottom left). Additionally, they can also be used as a plug-and-play prior for various supervised image enhancement tasks like dehazing, deraining and deblurring. On right, our conceptual block diagram.

Abstract

We present a new additive image factorization technique that treats images to be composed of multiple latent specular components which can be simply estimated recursively by modulating the sparsity during decomposition. Our model-driven RSFNet *estimates these factors by unrolling the optimization into network layers requiring only a few scalars to be learned. The resultant factors are interpretable by design and can be fused for different image enhancement tasks via a network or combined directly by the user in a controllable fashion. Based on RSFNet, we detail a zero-reference Low Light Enhancement (LLE) application trained without paired or unpaired supervision. Our system improves the state-of-the-art performance on standard benchmarks and achieves better generalization on multiple other datasets. We also integrate our factors with other task specific fusion networks for applications like deraining, deblurring and dehazing with negligible overhead thereby highlighting the multi-domain and multi-task generalizability of our proposed RSFNet. The code and data is released for reproducibility on the project homepage* [1].

[1] https://sophont01.github.io/data/projects/RSFNet/

1. Introduction

A low-light image has most regions too dark for comprehension due to low exposure setting or insufficient scene lighting which makes images highly challenging for computer processing and aesthetically unpleasant. Low-Light Enhancement (LLE) aims to recovers a well-exposed image from a low-light input [46]. LLE can be a critical pre-processing step before the downstream applications [55, 57]. Core LLE challenge lies in modeling the degradation function which is spatially varying and has complex dependence on multiple variables like color, camera sensitivity, illuminant spectra, scene geometry, *etc*.

Most LLE solutions decompose the image into meaningful latent factors based on a relevant optical property (Tab. 1). This allows individual manipulation of each factor which simplifies the enhancement operation. A common factorization is based on the Retinex approximation [39, 58], which assumes a multiplicative disentanglement of image I into two intrinsic factors: illumination-invariant, piecewise constant *reflectance* R and color-invariant, smooth *shading* S as $I = R \cdot S$. Other factorization criteria include frequency [35, 89], spatial scale [3, 50], spatio-frequency representation [18, 67], intensity

[33], reflectance rank [69, 76], *etc*. Fixed number of factors [35, 69, 87] and variable number that allow better representation [3, 33, 50] have been used. Some decompose image multiplicatively like Retinex [35, 87], while others split into additive factors which are numerically more stable [21, 32, 67]. Pixel segmentation could be soft or hard based on the membership across factors, with the former introducing fewer artifacts [4]. LLE solutions can be *global* or *local*. Global methods use whole image level statistics like gamma [27], histograms [96], *etc*., to enhance the images. Local methods employ spatially varying features like illumination maps [87], intensity/segmentation masks [33, 63], *etc*., for the same. Global methods are simpler but local ones can capture scene semantics better.

Traditional LLE methods used manually-designed model-based optimisation by deriving specific priors from the image itself [23, 29, 100], needing no training. Data-driven, machine learning based solutions have done better recently. They use training datasets to tune the model which generalizes to other images [3, 87, 90]. *Supervised* methods require annotated input-output pairs of images [88, 90, 102]. *Unsupervised* methods require annotated training data but not necessarily paired [36, 91]. *Zero-reference* methods do not need annotated data and approach the problem by explicitly encoding the domain knowledge from training images [27, 57, 72]. They generalize better and are simpler, lighter, and more interpretable by design.

In this paper, we present a zero-reference LLE method that outperforms prior methods on the average. At core is a novel *Recursive Specularity Factorization (RSF)* of the image factorization based on image specularity. We decompose an image into additive specular factors by thresholding the amount of sparsity of each pixel recursively. Successive factor differences mark out newly discovered image regions which are then individually targeted for enhancement. Our *RSFNet* that computes the factorization is model-driven, task-agnostic, and light-weight, needing about 200 trainable parameters. The image factors are fused using a task-specific UNet-based module to enhance each region appropriately. RSF is useful to other applications when combined with other task-specific fusion modules. Our main contributions are:

- A novel image factorization criterion and optimization formulation based on recursive specularity estimation.
- A model-driven RSFNet to learn factorization thresholds in a data-driven fashion using algorithm unrolling.
- A simple and flexible zero-reference LLE solution that surpasses the state of the art on multiple benchmarks and in the average generalization performance.
- Demonstration of RSF's usability to tasks like dehazing, deraining, and deblurring. RSF has high potential as a structural prior for several image understanding and enhancement tasks.

Criteria	No.	Type	Map	Seg	Example
Retinex	2	$*$	global	soft	[72, 88]
Frequency	2	$+$, low/high	global	hard	[89]
Spectral	2	$*$, fourier	global	soft	[35]
Low Rank	2	$+$	global	soft	[69, 76]
Wavelets	2^n	$+$, pyramid	global	soft	[18, 67]
Multiscale	2^n	$+$, pyramid	global	soft	[3, 50]
Glare/Shadow	3, 4	$*, +$	local	hard	[5, 78]
Intensity	var.	$+$, bands	local	hard	[32, 33]
Specularity	var.	$+$	local	soft	RSFNet

Table 1. Various LLE factorization criteria, with number of components (var. implies variable), type of factorization ($+$ additive/$*$ multiplicative), types of output maps (local/global), pixel segmentation across maps (soft/hard) and corresponding examplar methods. Our RSFNet proposes a novel specularity based factorization which allows a variable number of local soft-segmented factors.

2. Background

Model-based LLE: Early LLE solutions used traditional optimization models using either global statistics [15, 43, 66, 70] or spatially varying illumination maps for local editing [22, 29, 85, 99]. They were more interpretable but required hand-crafted algorithms and heuristics.

Data-driven LLE: Modern solutions take inspiration from traditional techniques and induce domain knowledge via loss terms or designed within the network architecture which are learned from large datasets in a data-driven fashion. They belong to one of the five training paradigms [46]. *Supervised* LLE methods require both low-light and well-lit *paired images* like Sharma and Tan [78], Wei et al. [87], Xu et al. [90], Yang et al. [95], Zhang et al. [103]. On the other hand, *unsupervised* methods like Jiang et al. [36], Ni et al. [64], Zhang et al. [96], require only unpaired low-light and well-lit *image sets*. *Semi-supervised* methods combine the previous two techniques and use both paired and unpaired annotations [73, 94]. *Self-supervised* solutions [48, 63] generate their own annotations using pseudo-labels or synthetic degradations. Contrary to all of these, *zero-reference* methods do not use ground truth reconstruction losses and assess the quality of output based upon encoded prior terms [27, 47, 57, 72, 98, 107]. These methods, like ours, possess improved generalizability due to explicit induction of domain knowledge and reduced chances of overfitting [27]. Zero-reference insights also provide direct valuable additions to the subsequent solutions in other paradigms.

Model-Driven Networks: Data-driven solutions have good performance but lack interpretability, whereas model-based methods are explainable by design but often compromise with lower performance. Model-driven networks [61] are hybrids which bring the best of both together. Such networks *unroll* optimization steps as differentiable layers with learnable parameters, inducing data-driven priors in place of hand-crafted heuristics. Although data-driven solutions are plenty, only a few model-driven solutions exist for low-level vision tasks like image restoration [40, 41], shadow re-

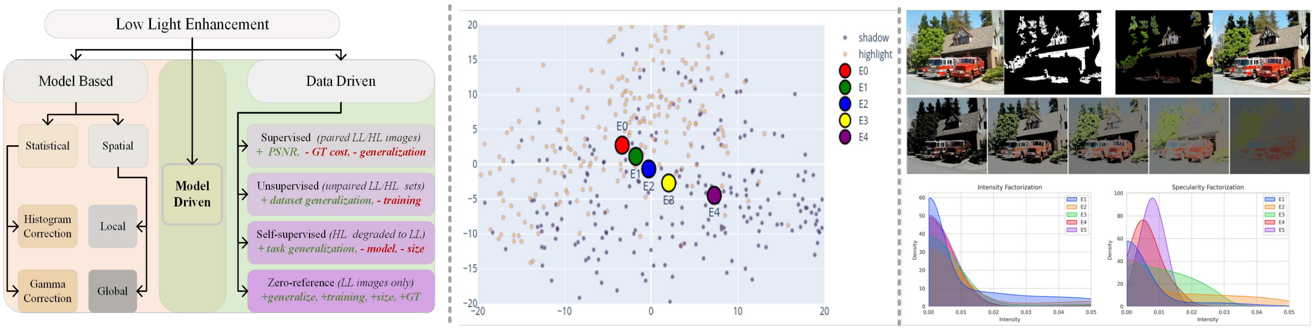

Figure 2. **Categorization and Motivation:** Left shows categorization of various LLE solution types (Sec. 2). Middle plot shows the relationship between five factor cluster centers w.r.t each other and the background comprising of shadow/non-shadow regions estimated using PCA dimensionality reduced DINO features [14]. Gradual progression of feature cluster centers from highlight region to shadow region indicates their capability to capture various illumination regions in an image. Top right shows one data point from CHUK dataset [34] with mask, processed shadow/highlight regions and extracted factors. Bottom right plots distinguish our specular fuzzy factors from intensity thresholding based binary division, with ours allowing more diverse distributions and richer representation.

moval [108], dehazing [54], deraining [83], denoising [68] and super-resolution [7, 8]. Such solutions are concise and efficient due to underlying task specific formulation.

Model-driven LLE solutions are very recent. UretinexNet [88] and UTVNet [104] are both supervised methods which respectively unroll the Retinex and total variational LLE formulations. RUAS [72] and SCI [57] are closest to our approach as they both propose model-driven zero-reference LLE solutions. RUAS [72] unrolls illumination estimation and noise removal steps in their optimization and compliment it with learnable architecture search, towards a dynamic LLE framework. SCI [57] on the other hand propose a residual framework wherein reflectance estimation is done by a self-calibration module which is then used to iteratively refine illumination maps. In contrast, our method is inspired directly by image formation fundamentals and presents a novel factorization criterion which provides better interpretability, performance and flexibility.

Retinex Factorization for LLE: Retinex [39, 58] is the most widely used factorization strategy for LLE [46, 72, 75, 87, 88] and beyond [9, 25, 74, 77]. One major Retinex limitation is due to the Lambertian reflection [38] assumption which approximates all surfaces as diffuse, thereby ignoring prevalent non-Lambertian effects in a real scene like specularity, translucency, caustics *etc*. Another issue is that pixel-wise multiplicative nature of Retinex factors is cumbersome to handle numerically (especially in LLE with near zero pixel values) and the obtained illumination maps require further semantic analysis for downstream applications. Extensions of Retinex like dichromatic model [81] and shadow segmentation [5], separate one extra component each in addition to diffuse R and S *e.g.* Sharma and Tan [78] and Baslamisli et al. [5] used glare and shadow image decomposition respectively. From this perspective our recursive specular factorization can be understood as an extension of the same idea with continuously varying illumination characteristics starting from bright glares and ending with dark shadows (see Fig. 2 and Sec. 3 for details).

Others Factorization Strategies: Apart from Retinex, other factorization techniques are listed in Tab. 1. Afifi et al. [3], Lim and Kim [50], Mertens et al. [59], Xu et al. [89] employ spatial or frequency based image decomposition. Recently, Yang et al. [94] used recursively concatenated features from a supervised encoder and Huang et al. [35] proposed a Fourier disentanglement based solution. Apart from these supervised factorizations, Zheng and Gupta [106] proposed semantic classification based ROI identification using a pretrained segmentation network. [27, 63] predict multiple gamma correction maps for enhancement. [32, 33] simulate single image exposure burst using piece-wise thresholded intensity functions whereas [69] uses low-rank decomposition for reflectance. Each factorization strategy harnesses crucial underlying optical observations and adds valuable insights to the low-level vision research. To the best of our knowledge, our proposed method here is the first to use recursive specularity estimation as a factorization strategy for LLE and other enhancement tasks.

3. Approach

Outline: Our entire pipeline consists of two parts. We first decompose the image into K factors using our *Recursive Specularity Factorization Network* (RSFNet), which consists of multiple factorization modules (FM) with each optimization step encoded as a differentiable network layer. Then we fuse, enhance, and denoise the factors using a *fusion network*, which is built using task dependent pre-existing architectures. This modular design allows easy adoption of our technique in several other tasks and learning paradigms (Sec. 4).

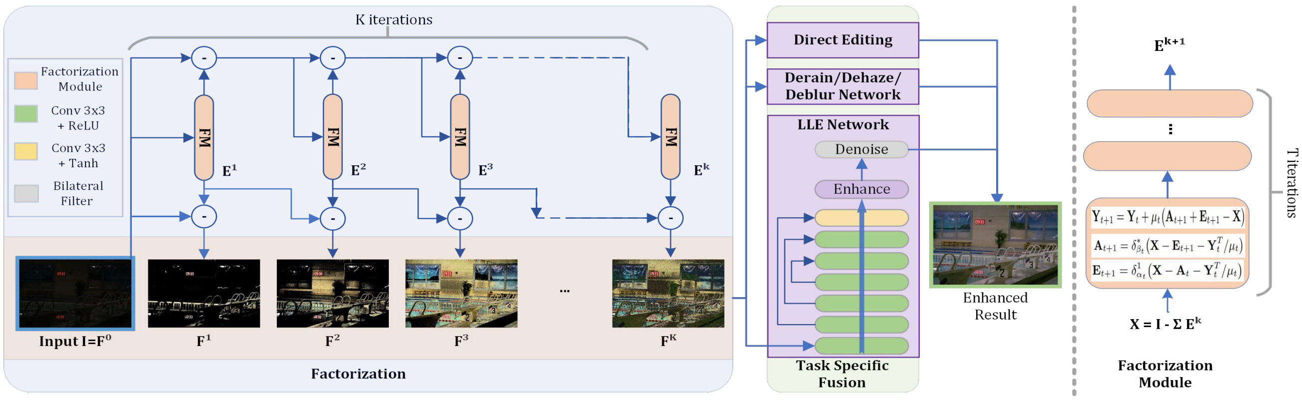

Figure 3. **Block Diagram:** Our factorization module (RSFNet) splits a given image into multiple specular components using model-driven unrolled optimization. Then fusion module combines all the factors to generate the enhanced output.

3.1. Factorization Network: RSFNet

Specularity Estimation: Specularity removal is a well studied problem. Most specularity removal methods [1, 28, 79] exploit the relative sparsity of specular highlights and use pre-defined fixed sparsity thresholds to isolate the specular component. According to dichromatic reflection model [81] image consists of a diffuse $\mathbf{A}$ and a specular $\mathbf{E}$ term: $\mathbf{X} = \mathbf{A} + \mathbf{E}$ for input X where specular component can be estimated by minimizing the L_0 norm approximated as:

$$\underset{\mathbf{E},\mathbf{A}}{\operatorname{argmin}} \, \|\mathbf{A}\|_* + \lambda \|\mathbf{E}\|_1 \quad \text{s.t.} \;\; \mathbf{X} = \mathbf{A} + \mathbf{E}, \tag{1}$$

where L_1 is relaxation of L_0, $*$ is Frobenius norm regularizer and λ is the sparsity parameter with higher values encouraging sparser results. Eq. (1) can be restated as augmented Lagrangian [10] using dual form and auxiliary parameters $(\mathbf{Y}, \mu)$, which are then solvable using iterative ADMM updates ($t \in [0, T]$) [11] as given below:

$$\begin{aligned} \mathbf{E}_{t+1} &= \delta^1_{\alpha_t}\left(\mathbf{X} - \mathbf{A}_t - \mathbf{Y}^T_t/\mu_t\right) && \text{where } \alpha : \mathcal{F}(\lambda, \mu), \\ \mathbf{A}_{t+1} &= \delta^*_{\beta_t}\left(\mathbf{X} - \mathbf{E}_{t+1} - \mathbf{Y}^T_t/\mu_t\right) && \text{where } \beta : \mathcal{F}(\mu), \\ \mathbf{Y}_{t+1} &= \mathbf{Y}_t + \mu_t\left(\mathbf{A}_{t+1} + \mathbf{E}_{t+1} - \mathbf{X}\right) && \text{where } \mu : \mathcal{F}(\mathbf{X}). \end{aligned} \tag{2}$$

Here δ^p_α is element-wise soft-thresholding operator [65]:

$$\delta^p_\alpha(x) = max(1 - \alpha/|x|_p,\ 0) \cdot x.$$

We can back-propagate through updates in Eq. (2) [88, 104] and hence can unroll them as neural network layers with learnable parameters $\boldsymbol{\alpha} : \{\alpha\}^T_0, \boldsymbol{\beta} : \{\beta\}^T_0$ and $\boldsymbol{\mu} : \{\mu\}^T_0$.

Relation with ISTA: Analyzing the structure of Eq. (2), we can draw parallels with the ISTA problem [17], which seeks a sparse solution to $\mathbf{E}$ for the condition $\mathbf{X} = \mathcal{G}\mathbf{E} + \epsilon$, with $\mathcal{G}$ as a learnable dictionary and negligible ϵ. In contrast, we have a non-negligible residue and identity dictionary. LISTA by Gregor and LeCun [26] showed how $\mathbf{E}$ update step can be represented as a weighted function which can then be approximated as finite network layers *i.e.*:

$$\mathbf{E}_{t+1} = \delta_{\alpha_t}(\mathbf{w}^1_t \mathbf{E}_t + \mathbf{w}^2_t \mathbf{X}), \tag{3}$$

with learnable parameters $(\alpha_t, \mathbf{w}^1_t, \mathbf{w}^2_t)$ for each iteration $t \in [0, T]$. Based on the weight coupling between $\mathbf{w}^1$ and $\mathbf{w}^2$, Chen et al. [16] simplified Eq. (3) by deriving both $\mathbf{w}^1$ and $\mathbf{w}^2$ from a single weight term, thereby halving the computation cost. A major simplification was further proposed by Liu et al. [51] as ALISTA, who proved how all weight terms could be analytically obtained for a known dictionary, thereby leaving only step sizes and thresholds *i.e.* μ and α_t to be estimated. Later on this idea was extended to other similar optimization formulations and improved upon by additional simplifications and guarantees *e.g.* Cai et al. [12] unrolled their ADMM updates into a network for robust principal component analysis.

Recursive Factorization: Drawing parallels from ALISTA [51] and its applications [12], we propose to learn the analytically reduced sparsity thresholds and step sizes via unrolled network layers. After optimizing the above mentioned objective Eq. (1) we obtain one specular factor $\mathbf{E}^k$ where index $k \in [1, K]$ indicates the factor number. For multiple factors, we recursively solve Eq. (1) by resetting the input X after removing the previous specular output and relaxing the initial sparsity weight. We initialize variables for each factor at $t = 0$ as:

$$\begin{aligned} \mathbf{X}^{k+1} = \mathbf{X}^k - \mathbf{E}^k, \qquad \mathbf{Y}^k = \mathbf{X}^k / \left\|\mathbf{X}^k\right\|_2 \\ \alpha^k = (1 - \nu^k)\hat{X}^k, \quad \beta^k = \nu^k \hat{X}^k, \quad \nu^k = k/K, \end{aligned} \tag{4}$$

where $\hat{X}$ indicates input mean and $\mathbf{X}^0$=$\mathbf{I}$. Intuitively, this can be understood as progressively removing specularity (E^k) from the original image by gradually relaxing the sparsity weight ($\alpha^{k+1} < \alpha^k$). This lets us split the original image into multiple additive factors as:

$$\mathbf{I} = \mathbf{E}^1 + \mathbf{E}^2 + \ldots + \mathbf{E}^K = \textstyle\sum_{k=1}^K \mathbf{E}^k \tag{5}$$

Unrolling: Based upon above discussion, we propose an unrolled network collecting all parameters in a single vector $\boldsymbol{\theta}$. In each iteration t, we estimate three scalars: thresholds for both components (α_t, β_t) and the step size (μ_t). Hence for a factor k, we have $3T$ parameters $\boldsymbol{\theta}^k := (\alpha^k, \beta^k, \mu^k)$ and overall we have only $3KT$ parameters $\boldsymbol{\theta} := \{\boldsymbol{\theta}^k\}_1^K$. Hence our model-driven factorization module is extremely light-weight compared to other decompositions (Tab. 2). We propose the following novel factorization loss:

$$L_f = \lambda_f \sum_{k=1}^{K} L_f^k \qquad \text{where} \qquad L_f^k = \left| \hat{E}^k / \hat{X}^k - \nu^k \right|. \tag{6}$$

This constraints the ratio of signal energy in the k^{th} factor compared to the input, to ν^k. As ν^k increases for higher factors, our factorization loss relaxes the sparsity constraint, thereby gradually increasing the number of pixels in the specular component. After training, we are left with K specular factors which sum to I. As shown in Fig. 1 and Fig. 2, each one of these factors highlights specific image regions with similar illumination characteristics which can be individually targeted for enhancement.

Motivation/Validation: The core assumption behind our factorization is that an image can be split into multiple specular factors with each representing specific illumination characteristic. Although such factorization quality assessment is difficult to estimate [9, 25, 30], we performed a toy experiment to validate our hypothesis using shadow detection dataset [34] which contains binary shadow masks in complex real world images (Fig. 2). We extract semantics-rich DINO image features [14] after masking shadow and non-shadow image regions and visualize them in 2D using PCA. This marks separation of feature space between shadowed and highlighted regions in the background. The regions with progressively degrading illumination characteristics (glare, direct light, indirect light, soft shadow, dark shadow, *etc*.) are expected to gradually lie between the two extremes. Next we factorize each image into five factors using our approach and plot the cluster mean for each factor feature distribution on the same graph. We can observe in Fig. 2 that successive factors gradually shift from the non-shadow towards the shadowed feature space region mirroring the expected illumination order. This confirms that our factorization decomposes the pixel values across fundamental illumination types like glare, direct light, indirect light, shadow, *etc*.

We also plot the respective factor distribution densities of intensity factorization [32, 33] and our specularity factorization (Fig. 2, bottom right). Intensity factorization allows little variation in the underlying factor distributions and imposes hard segmentation constraints with binary pixel masks. Our specular factors, on the other hand, permit higher variability and soft masks, with each pixel value spread across multiple factors. This provides more flexible representation and better optical approximation.

Algorithm 1: LLE Training

Input: Lowlight: I; Hyperparams: $\lambda_{c|e|s}$, K, T
Output: Enhanced: O; Params: $\boldsymbol{\theta}$=$\{\boldsymbol{\alpha}\}_0^K$, $\{\boldsymbol{\beta}\}_0^K$, $\{\boldsymbol{\mu}\}_0^K$

```
for e ← 0 to num of epochs do
    // Train Factorization Module
    for k ← 0 to K do
        for t ← 0 to T do
            Initialize E_0^k, A_0^k, Y_0^k ;            // Eqn. 4
            E_t, A_t, Y_t ← ADMM updates ;             // Eqn. 2
        end
        F^k ← E^k − E^{k−1} ;                          // Eqn. 7
    end
    Compute L_f ;                                       // Eqn. 6

    // Train Fusion Module
    if e > freeze epoch then
        Freeze all α, β, μ ;
        L_f ← 0 ;
    end
    I_fuse ← Concatenate [I, F^1, ..., F^K] ;
    O ← Forward (I_fuse) ;
    Compute L ;                                         // Eqn. 9
    Backpropagate L;
end
```

3.2. Fusion Network

In order to adhere to the zero-reference paradigm, we choose our fusion module to be a small fully-convolutional UNet like architecture with symmetric skip connections similar to other zero-reference methods [27, 63, 106]. One fundamental difference is that we modify the architecture to harness multiple factors and simultaneously perform fusion, enhancement and denoising. Specifically, it comprises of seven 3×3 convolutional layers with symmetric skip connections. We first pre-process all of our factors by subtracting the adjacent factors to discover the additional pixel values allowed in the current factor compared to the previous one as a soft mask:

$$\mathbf{F}^k = \mathbf{E}^k - \mathbf{E}^{k-1} \;\; \text{where} \;\; \mathbf{F}^1 = \mathbf{E}^1. \tag{7}$$

These factors are weighted if required using fixed scalar values and are then passed as a concatenated tensor into the fusion network. The output gamma maps $\mathbf{R}^k$ rescale different image regions differently and are applied directly on the original image inside the curve adjustment equation [27] for the fused result:

$$O = \Phi(\sum_{k=0}^{K} I + R^k.((I)^2 - I)). \tag{8}$$

The fused output is finally passed through a differentiable bilateral filtering layer Φ [71] for the final enhanced result O. Note that all the parameters from both factorization and fusion networks are trained together in end-to-end manner.
Loss Terms: We use two widely employed zero-reference losses for enhancement [27, 63, 96] and one image smoothing loss for denoising. First *color loss* L_c [27, 96] is based

Paradigm	Traditional Model Based			Zero-reference							
Method	LIME [29]	DUAL [100]	SDD [31]	ECNet [98]	ZDCE [27]	ZD++ [47]	RUAS [72]	SCI [57]	PNet [63]	GDP [20]	**RSFNet** (Ours)
Params x10^3	-	-	-	16.5x10^3	79.42	10.56	3.43	**0.26**	15.25	552x10^3	2.11
Lolv1 [87] (dataset split: 485/15, mean≈ 0.05, resolution: 400 × 600)											
PSNR$_y$ ↑	16.20	15.97	15.14	18.01	16.76	16.38	18.45	16.45	19.85	17.68	**22.17**
SSIM$_y$ ↑	0.695	0.692	0.754	0.644	0.734	0.645	0.766	0.709	0.718	0.678	**0.860**
PSNR$_c$ ↑	14.22	14.02	13.34	15.81	14.86	14.74	16.40	14.78	17.50	15.80	**19.39**
SSIM$_c$ ↑	0.521	0.519	0.634	0.469	0.562	0.496	0.503	0.525	0.550	0.539	**0.755**
NIQE ↓	8.583	8.611	3.706	8.844	8.223	8.195	5.927	8.374	8.629	6.437	**3.129**
LPIPS↓	0.344	0.346	0.278	0.358	0.331	0.346	0.303	0.327	0.340	0.375	**0.265**
Lolv2-real [95] (dataset split: 689/100, mean≈ 0.05, resolution: 400 × 600)											
PSNR$_y$ ↑	19.31	19.10	18.47	18.86	20.31	19.36	17.49	19.37	20.08	15.83	**21.46**
SSIM$_y$ ↑	0.705	0.704	0.792	0.613	0.745	0.585	0.742	0.722	0.691	0.627	**0.836**
PSNR$_c$ ↑	17.14	16.95	16.64	16.27	18.06	17.36	15.33	17.30	17.63	14.05	**19.27**
SSIM$_c$ ↑	0.537	0.535	0.678	0.459	0.580	0.442	0.493	0.540	0.539	0.502	**0.738**
NIQE ↓	9.076	9.083	4.191	9.475	4.191	8.709	6.172	8.739	9.152	6.867	**3.769**
LPIPS↓	0.322	0.324	**0.280**	0.360	0.310	0.340	0.325	0.294	0.340	0.390	**0.280**
GENERALIZED PERFORMANCE **Mean Scores** (**Lolv1** [87], **Lolv2-real** [95], **Lolv2-syn** [95] and **VE-Lol** [52])											
PSNR$_y$ ↑	18.50	17.83	17.50	18.45	19.26	18.73	17.09	18.07	19.65	15.88	**21.16**
SSIM$_y$ ↑	0.737	0.728	0.781	0.677	0.777	0.674	0.743	0.745	0.743	0.634	**0.854**
PSNR$_c$ ↑	16.53	15.88	15.77	16.25	17.19	16.76	15.12	16.20	17.35	14.15	**18.45**
SSIM$_c$ ↑	0.596	0.583	0.679	0.538	0.634	0.548	0.532	0.587	0.605	0.504	**0.758**
NIQE ↓	7.855	7.478	4.077	7.543	4.270	7.468	5.841	7.626	7.791	6.726	**3.763**
LPIPS↓	0.291	0.297	**0.266**	0.329	0.273	0.296	0.346	0.295	0.302	0.379	0.276

Table 2. Quantitative comparison of our method RSFNet with other traditional and zero-reference solutions on multiple lowlight benchmarks and six evaluation metrics. Shown here are scores for two datasets LOLv1 and LOLv2-real with mean value across all datasets in the last sub-table (key: ↑ higher better; ↓ lower better; **bold**: best; underline: second best).

on the gray-world assumption which tries to minimize the mean value difference between each color channel pair:

$$L_c = \sum_{(i,j)\in C} (\hat{O}^i - \hat{O}^j)^2, \quad C \in \{(r,g),(g,b),(b,r)\}.$$

Second is the *exposure loss* L_e [27, 33, 59], which penalizes grayscale intensity deviation from the mid-tone value:

$$L_e = \frac{1}{|\Omega|}\sum_{\Omega} (\phi(O) - 0.6)^2 \text{ where } \Omega \in \{c \times h \times w\},$$

where ϕ represents the average value over a 16×16 window. Our third loss is the pixel-wise *smoothing loss* which controls the local gradients $\nabla_{x|y}$ in the final output:

$$L_s = \frac{1}{|\Omega|}\sum_{\Omega} ((\nabla_x O)^2 + (\nabla_y O)^2),$$

Note that this differs from the previous works who focus on total variational loss of the gamma maps instead. Our final training loss with λ's as respective loss weights, is given as:

$$L = \lambda_f L_f + \lambda_c L_c + \lambda_e L_e + \lambda_s L_s. \quad (9)$$

4. Experiments and Results

We now report our implementation details, results and extensions. Please see the supplementary document for additional details and results.

Setup: We implement our combined network end-to-end on a single Nvidia 1080Ti GPU in PyTorch. We directly use low-light RGB images as inputs without any additional pre-processing. We first train factorization module for 25 epochs which we freeze and then optimize the fusion module for next 25 epochs. We use stochastic gradient descent for optimization with batch size of 10 and 0.01 as learning rate. Model hyper-parameters are fixed using grid search and the entire training take less than 30 minutes.

Datasets: We evaluate our method using multiple LLE benchmark datasets (Lolv1 [87], Lolv2-real [95], Lolv2-synthetic [95] and VE-Lol [52]) with standard train/test splits (Tab. 2). These datasets comprise of several underexposed small-aperture inputs and corresponding well-exposed ground-truth pairs. Here we report results on two datasets: Lolv1 and Lolv2-real and finally show the mean scores on all four datasets combined in the last sub-table Tab. 2 and in Fig. 5. Furthermore, we report generalization results (Tab. 4) on five additional no-reference datasets which have significant domain shift: DICM [42], LIME [29], MEF [56], NPE [85] and VV [82].

Metrics: We report both single channel (Y from YCbCr) and multichannel (RGB) performance scores. As full-reference metrics (which require ground truth), we use Peak Signal to Noise-Ratio (PSNR), Structural Similarity Index Metric (SSIM) [86] and Learned Perceptual Image Patch Similarity (LPIPS) [101]. For no-reference assessment (without ground truth), we report Naturalness Image

Variants	PSNR$_y$ ↑	SSIM$_y$ ↑
w/o L_e	8.12	0.238
w/o L_c	16.05	0.724
w/o L_s	20.13	0.846
w/o Denoise	19.51	0.756
w/o Fusion	19.32	0.830
Full	**22.17**	**0.860**

Table 3. Ablation analysis on five variants of our RSFNet (Sec. 4).

NIQE↓ & *LOE*↓	ECNet [98]	ZDCE [27]	ZD++ [47]	RUAS [72]	PNet [63]	SCI [57]	**RSFNet** (Ours)
DICM [42]	3.37—*676.7*	3.10—*340.8*	**2.94**—*511.9*	4.89—*1421*	3.00—*590.3*	3.61—*321.9*	3.23—***303.1***
LIME [29]	**3.75**—*685.1*	3.79—*135.0*	3.89—*332.2*	4.26—*719.9*	3.84—*223.2*	4.14—*75.5*	3.80—***68.3***
MEF [56]	3.30—*863.3*	3.31—*164.3*	3.18—*458.5*	4.08—*784.2*	3.25—*363.0*	3.43—***95.0***	**3.00**—*100.7*
NPE [85]	**3.24**—*936.1*	3.52—*312.9*	3.27—*532.2*	5.75—*1399*	3.29—*601.1*	3.89—*239.8*	3.31—***221.5***
VV [82]	2.15—292.4	2.75—*145.4*	2.53—*222.9*	3.82—*583.7*	2.56—*260.2*	2.30—***109.0***	**1.96**—***109.0***
Mean	3.16—*690.7*	3.29—*219.7*	3.16—*411.5*	4.56—*981.7*	3.19—*407.5*	3.47—*168.2*	**3.06**—***160.5***

Table 4. Qualitative comparison using naturalness preserving metrics (NIQE ↓ — LOE ↓) on five no-reference benchmarks: DICM, LIME, MEF, NPE and VV (**best** scores in bold, lower is better).

Figure 4. **Results:** Qualitative comparison of our method (green box) with other solutions (from top left 3 per row: SDD [31], ECNet [98], ZDCE [27]; ZD++ [47], RUAS [72], SCI [57]; PNet [63], GDP [20] and our RSFNet respectively). Our method generate natural looking images by handling noisy over and under exposed regions equally well, without over-saturating color or losing geometric details.

Quality Evaluator (NIQE) [60] and Lightness Order Error [84]. Note while PSNR and SSIM gauge performance quantitatively, other three metrics estimate perceptual quality.

Comparisons: We compare against three model-based traditional optimization methods: LIME [29], DUAL [100] and SDD [31] (others ignored due to low performance). For data-driven methods we use seven recent zero-reference methods (chronologically ordered): ECNet [98], zeroDCE [27], zeroDCE++ [47], RUAS [72], SCI [57], PNet [63] and GDP [20]. We use the official code releases with pre-trained weights and default parameters for results generation. Quantitative and qualitative performance comparison is shown in Tab. 2 and Fig. 4 respectively. Qualitatively, our method is cleaner with fewer artifacts and natural illumination (Fig. 4). This is validated by perceptual metrics like NIQE, LPIPS and LOE scores (Tabs. 2 and 4). Our method outperforms other similar category contemporary solutions on multiple metrics and achieves the best generalization performance across datasets. For a generalized performance, we take mean of all the scores across benchmarks and graphically show them in the polar plot in Fig. 5. Each polygon represents a separate LLE method with higher area inside indicating better performance.

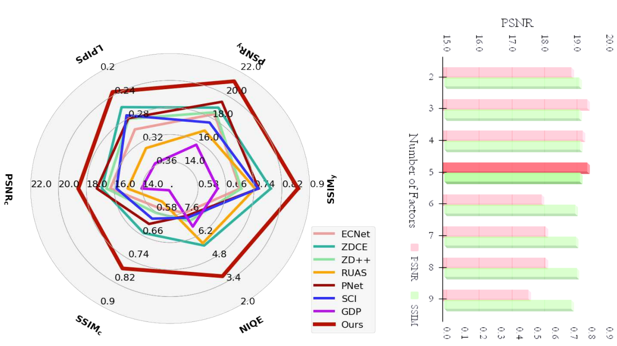

Figure 5. **Analysis:** On left, our average score on all datasets *vs*. other methods (more area implies better). On right, ablation analysis with varying number of factors.

Ablation: To validate our design choices, we conduct ablation study on several variants of our methods using Lolv1 dataset. The effect of different number of factors K on the final PSNR and SSIM scores are shown on right in Fig. 5. We choose the best observed hyper-parameter settings $K{=}5$ for all our experiments. The effect of various loss terms after removing them one at a time (*i.e.* w/o $L_{e|c|s}$) and the effect of the final denoising step are shown in Tab. 3. The last variant (*w/o* Fusion) represents an especially interesting setting where the fusion network is totally removed and inference uses only $3KT$ (=3*5*3=45) parameters. Fusion now reduces to a running average of the current image and

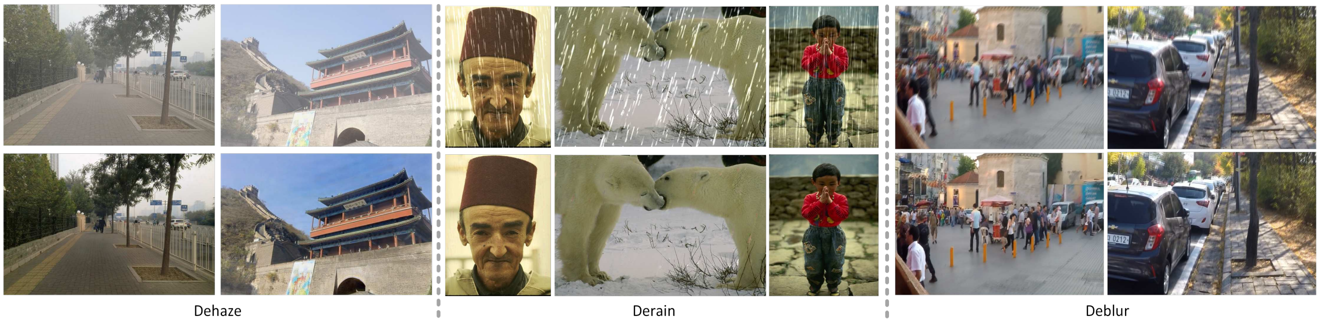

Figure 6. Image enhancement applications using our specular factors as inputs on the AirNet [45] base model. Shown here left-to-right are our results for Dehazing [44], Deraining [93] and Deblurring [62] tasks respectively using AirNet [45] as base model.

the next factor, weighted by the normalized mean:

$$O^{k+1} = (1-w^k)O^k + w^k F^k, \text{ where } w^k = \hat{F}^k / \textstyle\sum_k \hat{F}^k. \quad (10)$$

Even without any other zero-reference losses and using only a simple linear fusion, this method performs well, which demonstrates the effectiveness of our factors. Note here we have an order of magnitude smaller network size than SCI (0.045 *vs*. 0.26 thousand parameters in Tab. 2).

Extensions: Our specular factors are easily interpretable and can be used directly for image manipulation as image layers in standard image editing tools like GIMP [80], Photoshop [2], *etc*. We show an image relighting example by varying the color and blending modes of factors in (Fig. 1 bottom left, Fig. 7). This indicates the potential of our factorization to complex downstream applications. We explore three diverse image enhancement tasks: dehazing, deraining and deblurring. Here our goal is to evaluate the use of specularity factorization as a pre-processing step on an existing base model. We chose the recent AirNet [45] as it allows experimentation on multiple image enhancement tasks with minor backbone modification. To induce our factors as prior information, we concatenate them along with the original input and alter the first convolutional layer input channels. Note that we do not introduce any new loss or layers and directly train the model for three tasks one by one: *(i)* Dehazing on RESIDE dataset [44] *(ii)* Deraining on Rain100L dataset [93] and *(iii)* Deblurring on GoPro dataset


Figure 7. Controlable relighting applications using our factors as layers [80] (top:inputs; bottom:results; from left:edited light specularity, indoor color and outdoor intensity respectively).

TASK →	DEHAZE [44]		DERAIN [93]		DEBLUR [62]	
Method	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
AirNet (multi-task)	21.04	0.884	32.98	0.951	24.35	0.781
AirNet (uni-task)	23.18	0.900	34.90	0.9657	26.42	0.801
AirNet + Ours	**24.96**	**0.9292**	**36.19**	**0.9718**	**27.29**	**0.827**

Table 5. Our factors can induce structure prior in an existing base model and improve performance for multiple enhancement tasks.

[62]. As seen in Fig. 6 and Tab. 5, our results are perceptually more pleasing and improve the previously reported scores from multi-task methods consistently [45, 97]. We believe this is due to the induction of structural prior in the form of illumination based region categorization as the intensity and order of illumination at a pixel depends on the scene structure. See the supplementary for more results.

Limitations: Our method is sensitive to initialization conditions like the underlying algorithms [12, 51]. As a heuristic we use dataset mean for initialization. Another idea, to be explored in future, is to dynamically adapt to each input which is expected to further increase the performance.

Acknowledgements: We acknowledge the support of TCS Foundation and the Kohli Centre on Intelligent Systems for this research.

5. Conclusions

In this paper, we presented a recursive specularity factorization (RSF) and its application to zero-reference LLE. We learn optimization hyperparameters in a data-driven fashion by unrolling the stages into a small neural network. The factors are fused using a network to yield the final result. We also demonstrate the utility of RSFs for image relighting as well as for image enhancement tasks like dehazing, deraining and deblurring. We are exploring the extension of RSFs to applications like image harmonization, foreground matting, white-balancing, depth estimation, *etc*., and extend the technique to other signals beyond the visible spectrum.

Ethical Concerns: This work enhances captured images and poses no special ethical issues we are aware of.

References

[1] Amir Adler, Michael Elad, Yacov Hel-Or, and Ehud Rivlin. Sparse coding with anomaly detection. In *2013 IEEE MLSP*, 2013. 4

[2] Adobe Inc. Adobe photoshop, 2023. 8

[3] Mahmoud Afifi, Konstantinos Derpanis, Bjorn Ommer, and Michael Brown. Learning multi-scale photo exposure correction. In *CVPR*, 2021. 1, 2, 3

[4] Yağız Aksoy, Tae-Hyun Oh, Sylvain Paris, Marc Pollefeys, and Wojciech Matusik. Semantic soft segmentation. *ACM ToG (SIGGRAPH)*, 37(4), 2018. 2

[5] Anil S. Baslamisli, Partha Das, Hoang-An Le, Sezer Karaoglu, and Theo Gevers. Shadingnet: Image intrinsics by fine-grained shading decomposition. *IJCV*, 129(8), 2021. 2, 3

[6] Sean Bell, Kavita Bala, and Noah Snavely. Intrinsic images in the wild. *ACM Trans. on Graphics (SIGGRAPH)*, 33(4), 2014. 6, 8

[7] Goutam Bhat, Martin Danelljan, Luc Van Gool, and Radu Timofte. Deep burst super-resolution. *CVPR*, 2021. 3

[8] Goutam Bhat, Martin Danelljan, Fisher Yu, Luc Van Gool, and Radu Timofte. Deep reparametrization of multi-frame super-resolution and denoising. *ICCV*, 2021. 3

[9] Nicolas Bonneel, Balazs Kovacs, Sylvain Paris, and Kavita Bala. Intrinsic decompositions for image editing. *Computer Graphics Forum (Eurographics State of the Art Reports)*, 36(2), 2017. 3, 5

[10] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004. 4, 1

[11] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1), 2011. 4, 1

[12] HanQin Cai, Jialin Liu, and Wotao Yin. Learned robust pca: A scalable deep unfolding approach for high-dimensional outlier detection. *NeurIPS*, 34, 2021. 4, 8, 2, 3

[13] Yuanhao Cai, Hao Bian, Jing Lin, Haoqian Wang, Radu Timofte, and Yulun Zhang. Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In *ICCV*, 2023. 5, 17

[14] Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In *Proceedings of the International Conference on Computer Vision (ICCV)*, 2021. 3, 5, 1

[15] Turgay Celik and Tardi Tjahjadi. Contextual and variational contrast enhancement. *IEEE TIP*, 20(12), 2011. 2

[16] Xiaohan Chen, Jialin Liu, Zhangyang Wang, and Wotao Yin. Theoretical linear convergence of unfolded ista and its practical weights and thresholds. In *NeurIPS*, 2018. 4

[17] Ingrid Daubechies, Michel Defrise, and Christine De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on Pure and Applied Mathematics*, 57, 2003. 4

[18] Chi-Mao Fan, Tsung-Jung Liu, and Kuan-Hsien Liu. Half wavelet attention on m-net+ for low-light image enhancement. In *IEEE ICIP*, 2022. 1, 2

[19] Qingnan Fan, Dongdong Chen, Lu Yuan, Gang Hua, Nenghai Yu, and Baoquan Chen. A general decoupled learning framework for parameterized image operators. 2019. 5

[20] Ben Fei, Zhaoyang Lyu, Liang Pan, Junzhe Zhang, Weidong Yang, Tianyue Luo, Bo Zhang, and Bo Dai. Generative diffusion prior for unified image restoration and enhancement. In *CVPR*, 2023. 6, 7, 1, 11, 16

[21] Gang Fu, Qing Zhang, Chengfang Song, Qifeng Lin, and Chunxia Xiao. Specular highlight removal for real-world images. *Computer Graphics Forum*, 38 (7), 2019. 2

[22] Xueyang Fu, Delu Zeng, Yue Huang, Yinghao Liao, Xinghao Ding, and John Paisley. A fusion-based enhancing method for weakly illuminated images. *Signal Processing*, 129, 2016. 2

[23] Xueyang Fu, Delu Zeng, Yue Huang, Xiao-Ping Zhang, and Xinghao Ding. A weighted variational model for simultaneous reflectance and illumination estimation. In *CVPR*, 2016. 2

[24] Zhenqi Fu, Yan Yang, Xiaotong Tu, Yue Huang, Xinghao Ding, and Kai-Kuang Ma. Learning a simple low-light image enhancer from paired low-light instances. In *CVPR*, 2023. 1, 17

[25] Elena Garces, Carlos Rodriguez-Pardo, Dan Casas, and Jorge Lopez-Moreno. A survey on intrinsic images: Delving deep into lambert and beyond. *IJCV*, 2022. 3, 5

[26] Karol Gregor and Yann LeCun. Learning fast approximations of sparse coding. In *ICML*, 2010. 4

[27] Chunle Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, and Cong Runmin. Zero-reference deep curve estimation for low-light image enhancement. *CVPR*, 2020. 2, 3, 5, 6, 7, 1, 11, 16

[28] Jie Guo, Zuojian Zhou, and Limin Wang. Single im-

age highlight removal with a sparse and low-rank reflection model. In *ECCV*, 2018. 4

[29] Xiaojie Guo, Yu Li, and Haibin Ling. Lime: Low-light image enhancement via illumination map estimation. *IEEE TIP*, 26(2), 2016. 2, 6, 7, 3, 5, 16

[30] Avani Gupta, Saurabh Saini, and P. J. Narayanan. Interpreting intrinsic image decomposition using concept activations. In *ACM ICVGIP*, 2022. 5

[31] Shijie Hao, Xu Han, Yanrong Guo, Xin Xu, and Meng Wang. Low-light image enhancement with semi-decoupled decomposition. *IEEE TMM*, 22(12), 2020. 6, 7, 11, 16

[32] Charles Hessel. Simulated Exposure Fusion. *Image Processing On Line*, 9, 2019. 2, 3, 5

[33] Charles Hessel and Jean-Michel Morel. An extended exposure fusion and its application to single image contrast enhancement. In *WACV*, 2020. 2, 3, 5, 6

[34] Xiaowei Hu, Tianyu Wang, Chi-Wing Fu, Yitong Jiang, Qiong Wang, and Pheng-Ann Heng. Revisiting shadow detection: A new benchmark dataset for complex world. *IEEE TIP*, 30, 2021. 3, 5, 1, 2, 6, 7

[35] Jie Huang, Yajing Liu, Feng Zhao, Keyu Yan, Jinghao Zhang, Yukun Huang, Man Zhou, and Zhiwei Xiong. Deep fourier-based exposure correction network with spatial-frequency interaction. In *ECCV*, 2022. 1, 2, 3

[36] Yifan Jiang, Xinyu Gong, Ding Liu, Yu Cheng, Chen Fang, Xiaohui Shen, Jianchao Yang, Pan Zhou, and Zhangyang Wang. Enlightengan: Deep light enhancement without paired supervision. *IEEE TIP*, 30, 2021. 2, 1, 17

[37] Jeya Maria Jose Valanarasu, Rajeev Yasarla, and Vishal M. Patel. Transweather: Transformer-based restoration of images degraded by adverse weather conditions. In *CVPR*, 2022. 5

[38] Johann Heinrich Lambert. *Photometria sive de mensura et gradibus luminis, colorum et umbrae*. Klett, 1760. 3

[39] Edwin Herbert Land. The retinex theory of color vision. *Scientific American*, 237 6, 1977. 1, 3

[40] Bruno Lecouat, Jean Ponce, and Julien Mairal. Designing and learning trainable priors with non-cooperative games. *NeurIPS*, 2020. 2

[41] Bruno Lecouat, Jean Ponce, and Julien Mairal. Fully trainable and interpretable non-local sparse models for image restoration. *ECCV*, 2020. 2

[42] Chulwoo Lee, Chul Lee, and Chang-Su Kim. Contrast enhancement based on layered difference representation of 2d histograms. *IEEE TIP*, 22(12), 2013. 6, 7, 3, 5

[43] Chulwoo Lee, Chul Lee, and Chang-Su Kim. Contrast enhancement based on layered difference representation of 2d histograms. *IEEE TIP*, 22(12), 2013. 2

[44] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. Benchmarking single-image dehazing and beyond. *IEEE TIP*, 28(1), 2019. 8, 5, 6, 9, 13

[45] Boyun Li, Xiao Liu, Peng Hu, Zhongqin Wu, Jiancheng Lv, and Xi Peng. All-In-One Image Restoration for Unknown Corruption. In *CVPR*, 2022. 8, 5

[46] Chongyi Li, Chunle Guo, Linghao Han, Jun Jiang, Ming-Ming Cheng, Jinwei Gu, and Chen Change Loy. Low-light image and video enhancement using deep learning: A survey. *IEEE TPAMI*, 2021. 1, 2, 3

[47] Chongyi Li, Chunle Guo, and Chen Change Loy. Learning to enhance low-light image via zero-reference deep curve estimation. *IEEE TPAMI*, 2021. 2, 6, 7, 1, 11, 16

[48] Jinxiu Liang, Yong Xu, Yuhui Quan, Boxin Shi, and Hui Ji. Self-supervised low-light image enhancement using discrepant untrained network priors. *IEEE TCSVT*, 32(11), 2022. 2

[49] Zhexin Liang, Chongyi Li, Shangchen Zhou, Ruicheng Feng, and Chen Change Loy. Iterative prompt learning for unsupervised backlit image enhancement. In *ICCV*, 2023. 1, 17

[50] Seokjae Lim and Wonjun Kim. Dslr: Deep stacked laplacian restorer for low-light image enhancement. *IEEE TMM*, 23, 2021. 1, 2, 3

[51] Jialin Liu, Xiaohan Chen, Zhangyang Wang, and Wotao Yin. ALISTA: Analytic weights are as good as learned weights in LISTA. In *ICLR*, 2019. 4, 8, 3

[52] Jiaying Liu, Xu Dejia, Wenhan Yang, Minhao Fan, and Haofeng Huang. Benchmarking low-light image enhancement and beyond. *IJCV*, 129, 2021. 6, 3, 16

[53] Lin Liu, Lingxi Xie, Xiaopeng Zhang, Shanxin Yuan, Xiangyu Chen, Wengang Zhou, Houqiang Li, and Qi Tian. Tape: Task-agnostic prior embedding for image restoration. In *ECCV*, 2022. 5

[54] Yang Liu, Jinshan Pan, Jimmy Ren, and Zhixun Su. Learning deep priors for image dehazing. In *ICCV*, 2019. 3

[55] Yuen Peng Loh and Chee Seng Chan. Getting to know low-light images with the exclusively dark dataset. *CVIU*, 178, 2019. 1

[56] Kede Ma, Kai Zeng, and Zhou Wang. Perceptual quality assessment for multi-exposure image fusion. *IEEE TIP*, 24(11), 2015. 6, 7, 3, 5

[57] Long Ma, Tengyu Ma, Risheng Liu, Xin Fan, and Zhongxuan Luo. Toward fast, flexible, and robust low-light image enhancement. In *CVPR*, 2022. 1, 2, 3, 6, 7, 11, 16

[58] John J. McCann. Retinex at 50: color theory and spatial algorithms, a review. *Journal of Electronic Imaging*, 26, 2017. 1, 3

[59] Tom Mertens, Jan Kautz, and Frank Van Reeth. Exposure fusion: A simple and practical alternative to high dynamic range photography. *Computer Graphics Forum*, 28, 2009. 3, 6

[60] Anish Mittal, Rajiv Soundararajan, and Alan C. Bovik. Making a "completely blind" image quality analyzer. *IEEE Signal Processing Letters*, 20(3), 2013. 7, 5, 6

[61] Vishal Monga, Yuelong Li, and Yonina C. Eldar. Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing. *IEEE Signal Processing Magazine*, 38(2), 2021. 2

[62] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *CVPR*, 2017. 8, 5, 6, 9, 14

[63] Hue Nguyen, Diep Tran, Khoi Nguyen, and Rang Nguyen. Psenet: Progressive self-enhancement network for unsupervised extreme-light image enhancement. In *WACV*, 2023. 2, 3, 5, 6, 7, 1, 11, 16

[64] Zhangkai Ni, Wenhan Yang, Shiqi Wang, Lin Ma, and Sam Kwong. Towards unsupervised deep image enhancement with generative adversarial network. *IEEE TIP*, 29, 2020. 2

[65] Neal Parikh and Stephen Boyd. Proximal algorithms. *Foundations and Trends in Optimization*, 1(3), 2014. 4, 1

[66] Stephen M Pizer, E Philip Amburn, John D Austin, Robert Cromartie, Ari Geselowitz, Trey Greer, Bart ter Haar Romeny, John B Zimmerman, and Karel Zuiderveld. Adaptive histogram equalization and its variations. *CVGIP*, 39(3), 1987. 2

[67] Densen Puthussery, Hrishikesh Panikkasseril Sethumadhavan, Melvin Kuriakose, and Jiji Charangatt Victor. Wdrn: A wavelet decomposed relightnet for image relighting. In *ECCV workshop*, 2020. 1, 2

[68] Chao Ren, Yizhong Pan, and Jie Huang. Enhanced latent space blind model for real image denoising via alternative optimization. In *NeurIPS*, 2022. 3

[69] Xutong Ren, Wenhan Yang, Wen-Huang Cheng, and Jiaying Liu. Lr3m: Robust low-light enhancement via low-rank regularized retinex model. *IEEE TIP*, 29, 2020. 2, 3

[70] Ali M. Reza. Realization of the contrast limited adaptive histogram equalization (clahe) for real-time image enhancement. *J. VLSI Signal Process. Syst.*, 38(1), 2004. 2

[71] E. Riba, D. Mishkin, D. Ponsa, E. Rublee, and G. Bradski. Kornia: an open source differentiable computer vision library for pytorch. In *WACV*, 2020. 5

[72] Liu Risheng, Ma Long, Zhang Jiaao, Fan Xin, and Luo Zhongxuan. Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. In *CVPR*, 2021. 2, 3, 6, 7, 1, 11, 16

[73] Thomas Robert, Nicolas Thome, and Matthieu Cord. Hybridnet: Classification and reconstruction cooperation for semi-supervised learning. In *ECCV*, 2018. 2

[74] Saurabh Saini and P. J. Narayanan. Semantic priors for intrinsic image decomposition. In *BMVC*, 2018. 3

[75] Saurabh Saini and P. J. Narayanan. Semantic hierarchical priors for intrinsic image decomposition. *ArXiv*, abs/1902.03830, 2019. 3

[76] Saurabh Saini and P. J. Narayanan. Quaternion factorized simulated exposure fusion. In *ACM ICVGIP*, 2023. 2

[77] Saurabh Saini, Parikshit Sakurikar, and P. J. Narayanan. Intrinsic image decomposition using focal stacks. In *ACM ICVGIP*, 2016. 3

[78] Aashish Sharma and Robby T. Tan. Nighttime visibility enhancement by increasing the dynamic range and suppression of light effects. *CVPR*, 2021. 2, 3

[79] Sumit Shekhar, Max Reimann, Maximilian Mayer, Amir Semmo, Sebastian Pasewaldt, Jürgen Döllner, and Matthias Trapp. Interactive photo editing on smartphones via intrinsic decomposition. *Computer Graphics Forum*, 40(2), 2021. 4

[80] The GIMP Development Team. Gimp, 2023. 8, 6, 15

[81] Shoji Tominaga. Dichromatic reflection models for a variety of materials. *Color Research and Application*, 19, 1994. 3, 4

[82] Vassilios Vonikakis. Busting image enhancement and tone-mapping algorithms. `https://sites.google.com/site/vonikakis/datasets/`, 2007. [Online; accessed 26-Oct-2023]. 6, 7, 3, 5

[83] Hong Wang, Qi Xie, Qian Zhao, and Deyu Meng. A model-driven deep neural network for single image rain removal. 2020. 3

[84] Shuhang Wang, Jin Zheng, Hai-Miao Hu, and Bo Li. Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE TIP*, 22 (9), 2013. 7

[85] Shuhang Wang, Jin Zheng, Hai-Miao Hu, and Bo Li. Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE TIP*, 22 (9), 2013. 2, 6, 7, 3, 5

[86] Zhou Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: From error

visibility to structural similarity. *IEEE TIP*, 13(4), 2004. 6

[87] Chen Wei, Wenjing Wang, Yang Wenhan, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. In *BMVC*, 2018. 2, 3, 6, 10, 16, 17

[88] Wenhui Wu, Jian Weng, Pingping Zhang, Xu Wang, Wenhan Yang, and Jianmin Jiang. Uretinex-net: Retinex-based deep unfolding network for low-light image enhancement. In *CVPR*, 2022. 2, 3, 4, 17

[89] Ke Xu, Xin Yang, Baocai Yin, and Rynson W.H. Lau. Learning to restore low-light images via decomposition-and-enhancement. In *CVPR*, 2020. 1, 2, 3

[90] Xiaogang Xu, Ruixing Wang, Chi-Wing Fu, and Jiaya Jia. Snr-aware low-light image enhancement. In *CVPR*, 2022. 2, 5, 17

[91] Wending Yan, Robby T Tan, and Dengxin Dai. Nighttime defogging using high-low frequency decomposition and grayscale-color networks. In *ECCV*, 2020. 2, 1

[92] Shuzhou Yang, Moxuan Ding, Yanmin Wu, Zihan Li, and Jian Zhang. Implicit neural representation for cooperative low-light image enhancement. In *ICCV*, 2023. 1, 5, 17

[93] Wenhan Yang, Robby T. Tan, Jiashi Feng, Jiaying Liu, Zongming Guo, and Shuicheng Yan. Deep joint rain detection and removal from a single image. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1685–1694, 2017. 8, 5, 6, 9, 12

[94] Wenhan Yang, Shiqi Wang, Yapplicationsuming Fang, Yue Wang, and Jiaying Liu. Band representation-based semi-supervised low-light image enhancement: Bridging the gap between signal fidelity and perceptual quality. *IEEE TIP*, 30, 2021. 2, 3

[95] Wenhan Yang, Wenjing Wang, Haofeng Huang, Shiqi Wang, and Jiaying Liu. Sparse gradient regularized deep retinex network for robust low-light image enhancement. *IEEE TIP*, 30, 2021. 2, 6, 3, 4, 10, 16, 17

[96] Feng Zhang, Yuanjie Shao, Yishi Sun, Kai Zhu, Changxin Gao, and Nong Sang. Unsupervised low-light image enhancement via histogram equalization prior. *arXiv:2112.01766*, 2021. 2, 5, 1, 17

[97] Jinghao Zhang, Jie Huang, Mingde Yao, Zizheng Yang, Huikang Yu, Man Zhou, and Fengmei Zhao. Ingredient-oriented multi-degradation learning for image restoration. *CVPR*, 2023. 8, 6

[98] Lin Zhang, Lijun Zhang, Xinyu Liu, Ying Shen, Shaoming Zhang, and Shengjie Zhao. Zero-shot restoration of back-lit images using deep internal learning. *ACM MM*, 2019. 2, 6, 7, 1, 11, 16

[99] Qing Zhang, Ganzhao Yuan, Chunxia Xiao, Lei Zhu, and Wei-Shi Zheng. High-quality exposure correction of underexposed photos. In *ACM MM*, 2018. 2

[100] Qing Zhang, Yongwei Nie, and Weishi Zheng. Dual illumination estimation for robust exposure correction. *Computer Graphics Forum*, 38, 2019. 2, 6, 7, 16

[101] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, 2018. 6

[102] Yonghua Zhang, Jiawan Zhang, and Xiaojie Guo. Kindling the darkness: A practical low-light image enhancer. In *ACM MM*, 2019. 2

[103] Yonghua Zhang, Xiaojie Guo, Jiayi Ma, Wei Liu, and Jiawan Zhang. Beyond brightening low-light images. *IJCV*, 129, 2021. 2

[104] Chuanjun Zheng, Daming Shi, and Wentian Shi. Adaptive unfolding total variation network for low-light image enhancement. *ICCV*, 2021. 3, 4

[105] Naishan Zheng, Man Zhou, Yanmeng Dong, Xiangyu Rui, Jie Huang, Chongyi Li, and Fengmei Zhao. Empowering low-light image enhancer through customized learnable priors. 2023. 17

[106] Shen Zheng and Gaurav Gupta. Semantic-guided zero-shot learning for low-light image/video enhancement. In *WACV*, 2022. 3, 5

[107] Anqi Zhu, Lin Zhang, Ying Shen, Yong Ma, Shengjie Zhao, and Yicong Zhou. Zero-shot restoration of underexposed images via robust retinex decomposition. *ICME*, 2020. 2

[108] Yurui Zhu, Zeyu Xiao, Yanchi Fang, Xueyang Fu, Zhiwei Xiong, and Zheng-Jun Zha. Efficient model-driven network for shadow removal. *AAAI*, 2022. 3