西北工业大学

数字图像处理-论文翻译

Specularity Factorization for Low-Light Enhancement

陈佐维

计算机学院

计算机科学与技术

2024.11.12

学号：2022302649

Northwestern Polytechnical University

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

2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) | 979-8-3503-5300-6/24/$31.00 ©2024 IEEE | DOI: 10.1109/CVPR52733.2024.00009

[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-

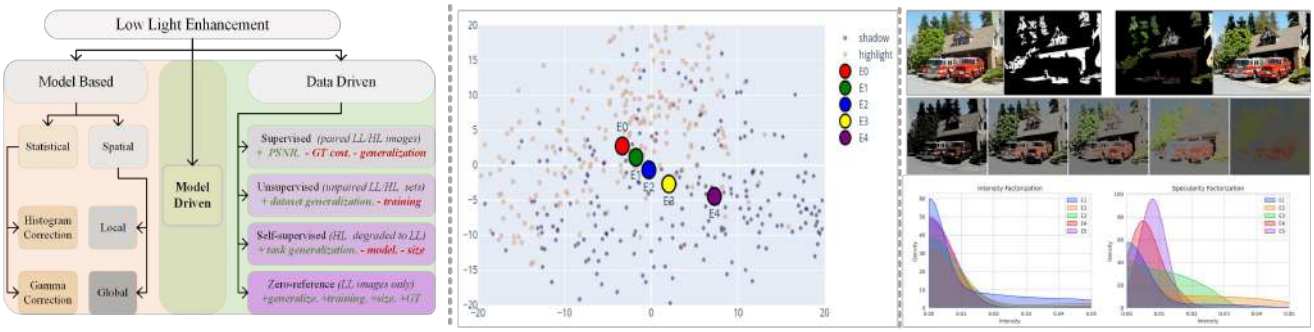

Figure 2. **Categorization and Motivation:** Left shows categorization of various LLE solution types (Sec. 2). Middle plot shows the relationship between five factor cluster centers w.r.t each other and the background comprising of shadow/non-shadow regions estimated using PCA dimensionality reduced 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).

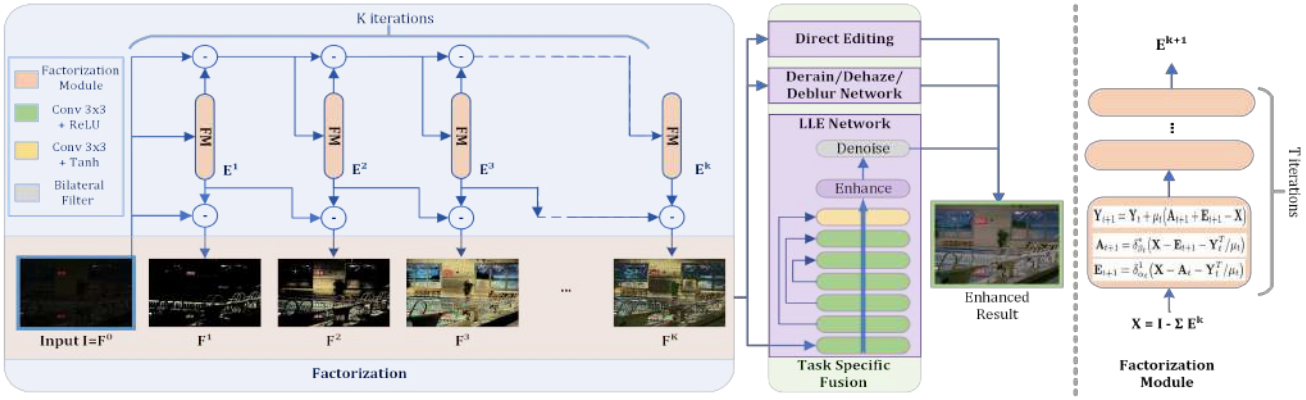

Figure 3. **Block Diagram:** Our factorization module (RSFNet) splits a given image into multiple specular components using model-driven unrolled optimization. Then fusion module combines all the factors to generate the enhanced output.

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:

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

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 $\boldsymbol{\alpha} : \{\alpha\}_0^T$, $\boldsymbol{\beta} : \{\beta\}_0^T$ and $\boldsymbol{\mu} : \{\mu\}_0^T$.

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.*:

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

with learnable parameters $(\alpha_t, \mathbf{w}_t^1, \mathbf{w}_t^2)$ 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 $e \leftarrow 0$ **to** *num of epochs* **do**
  // Train Factorization Module
  for $k \leftarrow 0$ **to** K **do**
    for $t \leftarrow 0$ **to** T **do**
      Initialize E_0^k, A_0^k, Y_0^k ; *// Eqn. 4*
      $E_t, A_t, Y_t \leftarrow$ ADMM updates ; *// Eqn. 2*
    end
    $F^k \leftarrow E^k - E^{k-1}$; *// Eqn. 7*
  end
  Compute L_f ; *// Eqn. 6*

  // Train Fusion Module
  if $e >$ *freeze epoch* **then**
    Freeze all $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\mu}$;
    $L_f \leftarrow 0$;
  end
  $I_{fuse} \leftarrow$ *Concatenate* $[I, F^1, ..., F^K]$;
  $O \leftarrow$ *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:

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

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:

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

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:

$$L = \lambda_f L_f + \lambda_c L_c + \lambda_e L_e + \lambda_s L_s. \tag{9}$$

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.

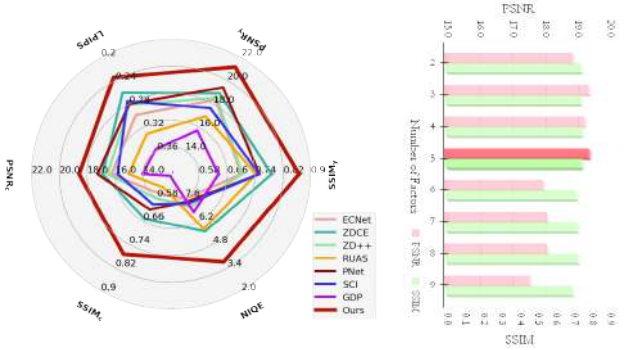

Figure 5. **Analysis:** On left, our average score on all datasets *vs*. other methods (more area implies better). On right, ablation analysis with varying number of factors.

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:

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

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)* Debluring on GoPro dataset [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.

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.

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

低光增强的高光分解

Saurabh Saini, P J Narayanan

2024 年 11 月 12 日

图 1: 高光分解：我们将单个输入图像（蓝框，第一行）分解为多个柔和的高光因子（为便于视觉效果已重新缩放），这些因子具有相似的照明特征（例如桌面阴影和灯光反射）。我们的因子直接支持无参考的低光增强和用户可控的图像重光（左下角）。此外，它们还可以作为即插即用的先验，用于去雾、去雨、去模糊等各种监督图像增强任务。右侧为我们的概念性框图。

摘要

我们提出了一种新的加性图像分解技术，将图像视为由多个潜在的高光成分组成，这些成分可以通过在分解过程中递归地调节稀疏性来简单地估计。我们的模型驱动型 RSFNet 通过将优化过程展开为网络层来估计这些因子，仅需要学习少量的标量。生成的因子具有可解释性，并且可以通过网络用于不同的图像增强任务，也可以由用户直接以可控的方式进行组合。基于 RSFNet，我们详细介绍了一个零参考的低光增强（LLE）应用，训练过程中无需成对或非成对的监督。我们的系统在标准基准测试上提高了当前最佳性能，并在多个其他数据集上实现了更好的泛化效果。此外，我们将这些因子与其他特定任务的融合网络相集成，用于去雨、去模糊和去雾等应用，几乎不增加额外开销，从而突显了我们提出的 RSFNet 在多领域和多任务中的广泛适应性。代码和数据已发布在项目主页上以供复现。

1 介绍

低光图像由于曝光设置过低或场景光线不足，使大多数区域过于黑暗，从而导致图像对计算机处理极具挑战性，且在视觉上不够美观。低光增强（LLE）旨在从低光输入中恢复出曝光良好的图像 [46]。LLE 可以作为下游应用的关键预处理步骤 [55, 57]。LLE 的核心挑战在于对退化函数的建模，该函数在空间上具有变化性，并且复杂地依赖于多种变量，如颜色、相机灵敏度、光源光谱和场景几何结构等。

低光图像由于曝光设置过低或场景光线不足，使大多数区域过于黑暗，从而导致图像对计算机处理极具挑战性，且在视觉上不够美观。低光增强（LLE）旨在从低光输入中恢复出曝光良好的图像[46]。LLE 可以作为下游应用的关键预处理步骤[55,57]。LLE的核心挑战在于对退化函数的建模，

该函数在空间上具有变化性，并且复杂地依赖于多种变量，如颜色、相机灵敏度、光源光谱和场景几何结构等。

大多数 LLE 解决方案基于相关的光学特性，将图像分解为有意义的潜在因子（表1），从而可以对每个因子进行独立操作，简化增强操作。一种常见的分解方法基于 Retinex 近似[39,58],假设图像 I 可被乘法解耦为两个内在因子：照明不变、分段恒定的反射率 R 和颜色不变、平滑的阴影 S，即 $I = R \cdot S$。其他分解标准包括频率[35,89]、空间尺度[3,50]、空间-频率表示 [18,67]、强度[33]、反射率等级[69,76] 等。固定数量的因子[35,69,87]和允许更好表示的可变数量因子[3,33,50]均已被使用。一些方法如 Retinex[35,87]将图像进行乘法分解，而其他方法则分解为加性因子，数值上更稳定[21,32,67]。像素分割可以基于因子间的成员关系进行软分割或硬分割，前者可引入更少的伪影[4]。LLE 解决方案可以是全局的或局部的。全局方法使用整个图像级别的统计信息，如伽马 [27]、直方图[96]等来增强图像，而局部方法则采用如照明图[87]、强度/分割掩码[33,63]等空间变化特征。全局方法更简单，而局部方法可以更好地捕捉场景语义。

传统的 LLE 方法使用基于手动设计的模型优化，通过从图像本身提取特定先验知识[23,29,100]，无需训练。基于数据驱动的机器学习解决方案最近表现更佳。它们使用训练数据集来调整模型，使其可以推广到其他图像[3,87,90]。监督方法需要成对的输入-输出图像[88,90,102]，而无监督方法则只需要标注的训练数据，不需要成对[36,91]。零参考方法不需要标注数据，通过显式编码来自训练图像的领域知识来解决问题 [27,57,72]。这些方法在设计上更具泛化性、简单、轻量且更具解释性。

在本文中，我们提出了一种零参考的 LLE 方法，在平均上优于先前的方法。其核心是一种基于图像高光的递归高光分解（RSF）新方法。我们通过递归地对每个像素的稀疏程度进行阈值处理，将图像分解为加性高光因子。相邻因子差异标记出新发现的图像区域，随后对这些区域分别进行增强。我们的 RSFNet 实现了因子分解，具有模型驱动、任务无关且轻量的特点，仅需约 200 个可训练参数。图像因子通过一个基于 UNet 的任务特定模块进行融合，以适当增强每个区域。RSF 在结合其他任务特定的融合模块时对于其他应用同样有效。我们的主要贡献包括：

- 基于递归高光估计的全新图像分解准则和优化公式。
- 使用算法展开的模型驱动 RSFNet 来学习分解阈值的基于数据的方式。
- 一种简单灵活的零参考 LLE 解决方案，在多个基准上超越了现有的技术，并在平均泛化性能上表现更佳。
- 演示了 RSF 在去雾、去雨和去模糊等任务中的可用性。RSF 作为多种图像理解与增强任务的结构先验具有极高的潜力。

2 背景

模型驱动的 LLE：早期的 LLE 解决方案使用传统的优化模型，采用全局统计方法 [15, 43, 66, 70] 或局部编辑的空间变化照明图 [22, 29, 85, 99]。这些方法具有较好的可解释性，但需要手工设计的算法和启发式方法。

数据驱动的 LLE：现代解决方案从传统技术中汲取灵感，通过损失项或网络架构设计引入领域知识，并从大型数据集中以数据驱动的方式学习。这些方法属于五种训练范式之一 [46]。监督的 LLE 方法需要低光和良好照明的配对图像，如 Sharma 和 Tan [78]，Wei 等人 [87]，Xu 等人 [90]，Yang 等人 [95]，Zhang 等人 [103]。另一方面，无监督方法（如 Jiang 等人 [36]，Ni 等人 [64]，Zhang 等人 [96]）仅需要未配对的低光和良好照明图像集。半监督方法结合了前两种技术，使用配对和非配对标注 [73, 94]。自监督解决方案 [48, 63] 使用伪标签或合成退化生成自己的标注。与所有这些方法相反，零参考方法不使用真实重建损失，而是基于编码的先验项评估输出质量 [27, 47, 57, 72, 98, 107]。这些方法（如我们的）由于显式引入领域知识并降

Criteria	No.	Type	Map	Seg	Example
Retinex	2	*	global	soft	[?, ?]
Frequency	2	+, low/high	global	hard	[?]
Spectral	2	*, fourier	global	soft	[?]
Low Rank	2	+	global	soft	[?, ?]
Wavelets	2^n	+, pyramid	global	soft	[?, ?]
Multiscale	2^n	+, pyramid	global	soft	[?, ?]
Glare/Shadow	3, 4	*, +	local	hard	[?, ?]
Intensity	var.	+, bands	local	hard	[?, ?]
Specularity	var.	+	local	soft	RSFNet

表 1: 不同的低光增强（LLE）分解标准，包括成分数量（var. 表示可变）、分解类型（+ 表示加性分解/* 表示乘性分解）、输出图类型（局部/全局）、跨图的像素分割方式（软分割/硬分割），以及对应的示例方法。我们的 RSFNet 提出了一个基于高光的新型分解方法，使得局部软分割因子的数量可变。

低过拟合的可能性，具有更好的泛化能力 [27]。零参考的见解还可以为其他范式的后续解决方案提供直接的有价值补充。

模型驱动网络：数据驱动的解决方案性能良好，但缺乏可解释性，而基于模型的方法具有可解释性设计，但通常在性能上有所妥协。模型驱动的网络 [61] 是一种混合方法，将两者的优点结合起来。这些网络将优化步骤展开为可微分的层，并具有可学习参数，以数据驱动的先验替代手工设计的启发式方法。尽管有很多数据驱动的解决方案，但只有少数模型驱动的解决方案用于低级视觉任务，如图像恢复 [40, 41]、去除阴影 [108]、去雾 [54]、去雨 [83]、去噪 [68] 和超分辨率 [7, 8]。由于其特定任务的公式化，这些解决方案简洁且高效。

模型驱动的 LLE 解决方案是最近才出现的。URetinexNet [88] 和 UTVNet [104] 都是监督方法，分别展开了 Retinex 和全变差 LLE 公式。RUAS [72] 和 SCI [57] 最接近我们的方法，因为它们都提出了模型驱动的零参考 LLE 解决方案。RUAS [72] 将照明估计和噪声去除步骤展开到它们的优化中，并通过可学习的架构搜索补充，以实现动态 LLE 框架。另一方面，SCI [57] 提出了一个残差框架，其中反射率估计由自校准模块完成，然后用于迭代细化照明图。相比之下，我们的方法直接受图像生成基本原理的启发，提出了一种新的分解准则，提供了更好的可解释性、性能和灵活性。

Retinex 分解用于 LLE：Retinex [39, 58] 是 LLE 中最广泛使用的分解策略 [46, 72, 75, 87, 88]，甚至在其他领域中也被应用 [9, 25, 74, 77]。Retinex 的一个主要限制是其基于朗伯反射假设 [38]，即将所有表面近似为漫反射，从而忽略了真实场景中普遍存在的非朗伯效应，如高光、半透明、焦散等。另一个问题是 Retinex 因子的逐像素乘法特性在数值处理上很复杂（尤其在 LLE 中像素值接近零时），且所得到的照明图需要进一步的语义分析以用于下游应用。Retinex 的扩展方法，如二色模型 [81] 和阴影分割 [5]，在漫反射 R和阴影S 之外分别多分离出一个额外成分，例如 Sharma 和 Tan [78] 以及 Baslamisli 等人 [5] 分别使用了眩光和阴影图像分解。从这个角度看，我们的递归高光分解可以理解为同一思路的扩展，以从亮斑到暗影的连续变化照明特性来实现（详情见图 2 和第 3 节）。

其他分解策略：除了 Retinex，其他分解技术如表 1 所示。Afifi 等人 [3]、Lim 和 Kim [50]、Mertens 等人 [59]、Xu 等人 [89] 使用了基于空间或频率的图像分解。最近，Yang 等人 [94] 使用了来自监督编码器的递归串接特征，Huang 等人 [35] 提出了基于傅里叶解耦的解决方案。除了这些监督分解，Zheng 和 Gupta [106] 提出了使用预训练的分割网络进行基于语义分类的 ROI 识别。[27, 63]

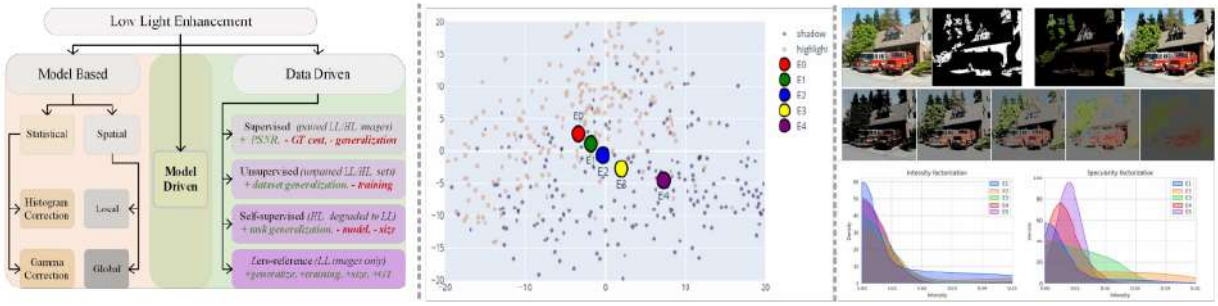

图 2: 分类和动机：左侧展示了各种 LLE 解决方案类型的分类（第 2 节）。中间图展示了五个因子聚类中心之间的关系，以及通过主成分分析（PCA）降维的 DINO 特征 [14] 估算出的背景（包括阴影/非阴影区域）。特征聚类中心从高光区域逐渐过渡到阴影区域，表明它们能够捕捉图像中的不同照明区域。右上角展示了 CHUK 数据集 [34] 中的一个数据点，包括遮罩、处理后的阴影/高光区域以及提取的因子。右下角的图则将我们的高光模糊因子与基于强度阈值的二值划分区分开来，我们的方法允许更丰富的分布和更丰富的表示。

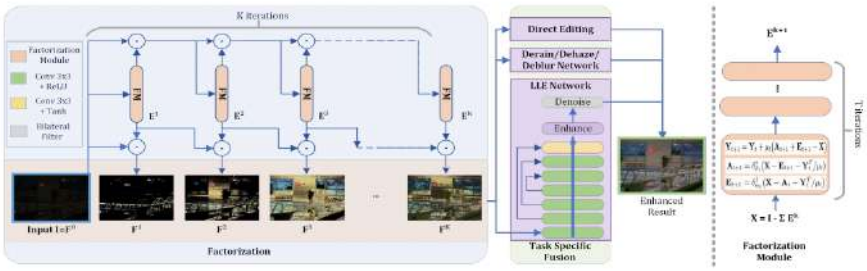

图 3: 模块图：我们的分解模块（RSFNet）通过模型驱动的展开优化将给定图像分解为多个高光成分。然后，融合模块将所有因子组合生成增强后的输出。

预测了多个用于增强的伽马校正图。[32, 33] 使用分段阈值的强度函数模拟单图像曝光爆发，而 [69] 则使用低秩分解来表示反射率。每种分解策略都利用了关键的光学观察，为低级视觉研究提供了宝贵的见解。据我们所知，我们提出的方法是首个将递归高光估计用作 LLE 和其他增强任务分解策略的方法。

3 方法

概述：我们的整个流程包含两个部分。首先，我们使用递归高光分解网络（RSFNet）将图像分解为 K个因子，其中 RSFNet 包含多个分解模块（FM），每个优化步骤被编码为一个可微分的网络层。然后，我们使用融合网络对这些因子进行融合、增强和去噪，融合网络基于任务依赖的现有架构构建。这种模块化设计使我们的技术能够轻松应用于其他任务和学习范式（第 4 节）。

3.1 分解网络：RSFNet

高光估计：高光去除是一个研究较多的问题。大多数高光去除方法[1,28,79]利用高光稀疏性的特性，使用预定义的固定稀疏性阈值来隔离高光成分。根据二色反射模型[81]，图像由漫反射项 A 和高光项 E 组成：对于输入 X，有 $X = A + E$，其中高光成分可以通过最小化 L_0 范数近似来估计：

$$\begin{aligned} &\arg\min_{E,A} \|A\|_* + \lambda\|E\|_1 \\ &\text{s.t.} \quad \mathbf{X} = \mathbf{A} + \mathbf{E}, \end{aligned} \tag{1}$$

其中 L_1 是 L_0 的松弛，$*$ 是 Frobenius 范数正则化器，λ 是稀疏参数，值越高则结果越稀疏。式 (1) 可以通过引入对偶形式和辅助参数 (Y,μ) 重新表述为增广拉格朗日[10]，然后可以使用迭代 ADMM 更新（在 $t\in[0,T]$ 中）来求解，如下所示：

$$\begin{aligned} E_{t+1} &= \delta^1_{\alpha_t}\left(X - A_t - \frac{Y_t^T}{\mu_t}\right),\\ A_{t+1} &= \delta^*_{\beta_t}\left(X - E_{t+1} - \frac{Y_t^T}{\mu_t}\right),\\ Y_{t+1} &= Y_t + \mu_t\left(A_{t+1} + E_{t+1} - X\right). \end{aligned} \tag{3}$$

其中，$\alpha:\mathcal{F}(\lambda,\mu)$，$\beta:\mathcal{F}(\mu)$，$\mu:\mathcal{F}(X)$。

这里，δ^p_α 是逐元素软阈值算子[65]：

$$\delta^p_\alpha(x) = \max\left(1 - \frac{\alpha}{|x|_p}, 0\right)\cdot x.$$

我们可以通过公式 (2) 中的更新进行反向传播[88,104]，因此可以将其展开为具有可学习参数的神经网络层，其中 $\alpha:\{\alpha_0\}^p_T$，$\beta:\{\beta_0\}^p_T$ 和 $\mu:\{\mu_0\}^p_T$。 **与 ISTA 的关系：**通过分析公式 (2) 的结构，我们可以将其与 ISTA 问题[17] 进行类比，ISTA 寻求稀疏解 E 满足条件 $X = GE+\epsilon$，其中 G 是一个可学习的字典，且 ϵ 可忽略不计。与之相反，我们有一个不可忽略的残差和单位字典。Gregor 和 LeCun[26]的 LISTA 展示了 E 的更新步骤可以表示为一个加权函数，该函数可以近似为有限的网络层，即：

$$E_{t+1} = \delta_{\alpha_t}\left(w^1_t E_t + w^2_t X\right), \tag{3}$$

其中每次迭代 $t\in[0,T]$ 的可学习参数为 (α_t, w^1_t, w^2_t)。基于 w_1 和 w_2 之间的权重耦合，Chen 等人[16] 通过从单一权重项导出 w_1 和 w_2，简化了公式 (3)，从而将计算成本减半。Liu 等人[51]在此基础上进一步提出了 ALISTA，证明了在已知字典的情况下可以解析地获得所有权重项，从而只需估计步长和阈值，即 μ 和 α_t。随后，该思想被扩展到其他类似的优化公式，并通过额外的简化和保证得到了改进，例如 Cai 等人[12]将他们的 ADMM 更新展开成一个网络，用于鲁棒主成分分析。

递归分解：借鉴 ALISTA[51]及其应用[12]的思想，我们提出通过展开的网络层来学习解析化简后的稀疏阈值和步长。在优化上述目标函数（公式 (1)）后，我们获得一个高光因子 E_k，其中索引 $k\in[1,K]$ 表示因子编号。对于多个因子，我们通过递归求解公式 (1) 来实现，每次在去除前一个高光输出后重置输入 X，并放宽初始稀疏权重。我们在 $t=0$ 处初始化每个因子的变量如下：

$$\begin{aligned} X_{k+1} &= X_k - E_k,\\ Y_k &= \frac{X_k}{\|X_k\|_2},\\ \alpha_k &= (1-\nu_k)\hat{X}_k,\\ \beta_k &= \nu_k\hat{X}_k,\\ \nu_k &= \frac{k}{K}, \end{aligned} \tag{4}$$

其中，$\hat{X}$ 表示输入的均值，且 $X_0 = I$。直观上，这可以理解为通过逐渐放松稀疏权重（$\alpha_{k+1} < \alpha_k$）从原始图像中逐步去除高光 E_k。这使我们可以将原始图像分解为多个加性因子：

$$I = E_1 + E_2 + \cdots + E_K = \sum_{k=1}^{K} E_k. \tag{5}$$

展开：基于上述讨论，我们提出了一个展开的网络，将所有参数收集在单个向量 θ 中。在每次迭代 t 中，我们估计三个标量：两个分量的阈值 (α_t,β_t) 和步长 (μ_t)。因此，对于一个因子 k，我们有 $3T$ 个参数 $\theta_k := (\alpha_k,\beta_k,\mu_k)$，总体上只有 $3KT$ 个参数 $\theta := \{\theta_k\}^K_{k=1}$。因此，与其他分解方法相比（表 2），我们的模型驱动分解模块极其轻量。

我们提出了以下新颖的分解损失函数：

$$L_f = \lambda_f\sum_{k=1}^{K} L^k_f \quad \text{其中 } L^k_f = \left\|\frac{\hat{E}_k}{\hat{X}_k} - \nu_k\right\|. \tag{6}$$

该损失约束了第 k 个因子的信号能量与输入的比值为 ν_k。随着 ν_k 对于更高阶因子的增大，我们的分解损失放松了稀疏性约束，从而逐步增加高光成分中的像素数量。经过训练后，我们得到 K 个高光因子，它们之和等于 I。如图 1 和图 2 所示，这些因子中的每一个都突出了具有相似照明特征的特定图像区域，可以单独进行增强。 **动机/验证：**我们分解的核心假设是图像可以分解为多个高光因

子，每个因子代表特定的照明特征。尽管这种分解质量的评估较难[9,25,30]，但我们进行了一个简单实验，以验证我们的假设，使用了包含复杂真实世界图像中的二值阴影掩码的阴影检测数据集[34]（如图 2 所示）。我们在掩盖阴影和非阴影图像区域后提取了具有丰富语义的 DINO 图像特征[14]，并使用 PCA 在二维中进行可视化。这展示了背景中阴影和高光区域在特征空间中的分离。具有逐渐衰减照明特征的区域（如眩光、直接光、间接光、柔和阴影、深阴影等）预计将逐步位于两个极端之间。

接下来，我们使用该方法将每张图像分解为五个因子，并在同一图上绘制每个因子的特征分布聚类均值。如图 2 所示，我们可以观察到，连续的因子逐渐从非阴影区域向阴影特征空间区域转移，反映出预期的照明顺序。这验证了我们的分解能够将像素值分解为基本的照明类型，如眩光、直接光、间接光、阴影等。

我们还绘制了强度分解[32,33]和我们的高光分解的相应因子分布密度（图 2，右下角）。强度分解在底层因子分布上允许的变化较小，并使用二值像素掩码施加硬分割约束。相比之下，我们的高光因子允许更高的变异性和软掩码，使每个像素值在多个因子间分布。这提供了更灵活的表示和更好的光学近似。

Algorithm 1 LLE Training

Input: Lowlight: I; Hyperparams: $\lambda_{c|e|s}, K, T$
Output: Enhanced: O; Params: $\theta = \{\alpha\}_0^K, \{\beta\}_0^K, \{\mu\}_0^K$
for each epoch $e = 0$ to num of epochs **do**
  // **Train Factorization Module**
  for each factor $k = 0$ to K **do**
    for each iteration $t = 0$ to T **do**
      Initialize E_k^0, A_k^0, Y_k^0
      $E_t, A_t, Y_t \leftarrow$ ADMM updates // Eqn. 2
    end for
    $F^k \leftarrow E^k - E^{k-1}$ // Eqn. 7
  end for
  Compute L_f // Eqn. 6
  // **Train Fusion Module**
  if $e >$ freeze epoch **then**
    Freeze all α, β, μ
    $L_f \leftarrow 0$
  end if
  $I_{\text{fuse}} \leftarrow \text{Concatenate}\left(I, F^1, \ldots, F^K\right)$
  $O \leftarrow \text{Forward}\left(I_{\text{fuse}}\right)$
  Compute L // Eqn. 9
  Backpropagate L
end for

3.2 融合网络

为了遵循零参考范式，我们选择将融合模块设计为一个小型的全卷积 UNet 类架构，带有对称跳跃连接，类似于其他零参考方法[27,63,106]。一个根本的不同在于，我们修改了架构，以利用多个因子并同时执行融合、增强和去噪。具体来说，它由七个 3×3 的卷积层组成，带有对称跳跃连接。我们首先通过相邻因子的相减来预处理所有因子，以发现当前因子相对于前一个因子所允许的附加像素值，形成软掩码：

$$F_k = E_k - E_{k-1} \quad \text{其中 } F_1 = E_1\Theta \tag{7}$$

如果需要，这些因子会使用固定的标量值进行加权处理，随后以连接的张量形式输入到融合网络中。输出的伽马映射 R_k 对不同的图像区域进行不同的重新缩放，并直接应用于原始图像，使用曲线调整公式[27]以获得融合结果：

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

融合输出最终通过一个可微分的双边滤波层 Φ[71]，生成最终的增强结果 O。需要注意的是，分解网络和融合网络的所有参数是一起端到端训练的。

损失项：我们使用了两种广泛应用的零参考损失来进行增强[27,63,96]，以及一种图像平滑损失来进行去噪。第一个颜色损失 L_c[27,96] 基于灰世界

假设，旨在最小化每一对颜色通道之间的均值差异：

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

第二个损失是曝光损失 L_e[27,33,59] ，它惩罚灰度强度偏离中间色调的情况：

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

其中 ϕ 表示在 16×16 窗口上的平均值。

我们的第三个损失是像素级平滑损失，它控制最终输出中的局部梯度 $\nabla_{x|y}$：

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

请注意，这与之前的研究不同，后者主要关注伽马映射的全变分损失。我们的最终训练损失带有各损失的权重 λ：

$$L = \lambda_f L_f + \lambda_c L_c + \lambda_e L_e + \lambda_s L_s. \tag{9}$$

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

表2：我们的方法 RSFNet 与其他传统和零参考解决方案在多个低光基准和六个评价指标上的定量比较。此处显示的是两个数据集 LOLv1 和 LOLv2-real 的评分，以及在最后一个子表中的所有数据集的平均值（注：↑ 表示值越高越好；↓ 表示值越低越好；加粗表示最佳值；下划线表示次优值）。

4 实验与结果

我们现在报告我们的实现细节、结果和扩展。更多的细节和结果请见补充文件。

设置：我们在一块 Nvidia 1080Ti GPU 上使用 PyTorch 端到端实现了我们的联合网络。我们直接使用低光 RGB 图像作为输入，无需任何额外的预处理。首先，我们训练分解模块 25 个 epoch，然后冻结该模块并优化融合模块接下来的 25 个 epoch。我们使用随机梯度下降（SGD）进行优化，批量大小为 10，学习率为 0.01。模型的超参数通过网格搜索固定，整个训练过程耗时不到 30 分钟。

数据集：我们使用多个 LLE 基准数据集评估我们的方法（Lolv1[87]、Lolv2-real[95] 、Lolv2-synthetic[95]和 VE-Lol[52]，并使用标准的训练/测试划分（表 2）。这些数据集包含了多个曝光不足的小光圈输入图像及其相应的高曝光真值对。这里我们报告了两个数据集（Lolv1 和 Lolv2-real）上的结果，并在表 2 和图 5 中的最后一个子表中展示了所有四个数据集的平均得分。此外，我们还在五个具有显著领域偏移的无参考数据集上报告了泛化结果（表 4）：DICM[42]、LIME[29]、MEF[56]、

NPE[85]和 VV[82]。

指标：我们报告了单通道（从 YCbCr 的 Y 通道）和多通道（RGB）性能得分。作为全参考指标（需要真值），我们使用了峰值信噪比（PSNR）、结构相似性指标（SSIM[86]）和学习的感知图像块相似性（LPIPS）[101]。对于无参考评估（无需真值），我们报告了自然图像质量评价（NIQE）[60]和亮度顺序误差[84]。注意，PSNR 和 SSIM 主要用于量化性能，而其他三个指标用于评估感知质量。

Variants	$PSNR_y \uparrow$	$SSIM_y \uparrow$
$w/o\ L_c$	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**

表 3: 我们的 RSFNet 的五个变体的消融分析（第 4 节）。

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***

表 4: 使用自然度保持指标（NIQE ↓ — LOE ↓）在五个无参考基准数据集上的定性比较：DICM、LIME、MEF、NPE 和 VV（最佳分数为加粗，越低越好）。

比较：我们与三种基于模型的传统优化方法进行了比较：LIME [29]、DUAL [100] 和 SDD [31]（其他方法因性能较低而被忽略）。对于数据驱动的方法，我们使用了七种最近的零参考方法（按时间顺序排列）：ECNet [98]、zeroDCE [27]、zeroDCE++ [47]、RUAS [72]、SCI [57]、PNet [63] 和 GDP [20]。我们使用官方发布的代码、预训练权重和默认参数生成结果。定量和定性性能比较分别在表 2 和图 4 中展示。定性来看，我们的方法更干净，伪影更少，光照自然（见图 4）。这一点通过感知指标（如 NIQE、LPIPS 和 LOE 分数）得到了验证（见表 2 和表 4）。我们的方法在多个指标上优于其他同类的现代方法，并在各个数据集上实现了最佳的泛化性能。为了获得综合性能，我们对所有基准数据集的分数取平均，并在图 5 中的极坐标图中以图形方式展示。每个多边形代表一种不同的 LLE 方法，内部面积越大表示性能越好。

图 4: ：结果：定性比较我们的方法（绿框）与其他解决方案（从左上角3每行： SDD [31]，ECNet [98]，ZDCE [27]；ZD++ [47]，RUAS [72]，SCI [57]；PNet [63]，GDP [20]和我们的RSFNet分别）。我们的方法通过同样好地处理暴露区域下的噪声来产生自然的图像，而不会过度饱和颜色或失去几何细节。

图 5: 使用我们的高光因子作为输入，在 AirNet [45] 基础模型上的图像增强应用。左到右分别显示了我们在去雾 [44]、去雨 [93] 和去模糊 [62] 任务中的结果，均使用 AirNet [45] 作为基础模型。

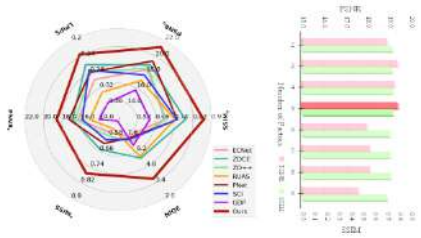

图 6: 分析：在左边，我们在所有数据集上的平均得分与其他方法上的平均得分（面积更大意味着更好）。右图，不同数量因素的消融分析。

消融实验：为了验证我们的设计选择，我们在 Lolv1 数据集上对我们方法的多个变体进行了消融研究。不同因子数 K 对最终 PSNR 和 SSIM 得分的影响在图 5 的右侧展示。我们选择最佳的超参数设置 $K = 5$ 用于所有实验。在一次移除不同损失项（即 w/o $L_e|L_c|L_s$）后的影响，以及最终去噪步骤的效果展示在表 3 中。最后一个变体（w/o Fusion）表示一个特别有趣的设置，其中完全移除了融合网络，推理仅使用 3KT (=3*5*3=45) 个参数。此时的融合简化为当前图像和下一个因子的加权平均，权重由归一化均值确定：

$$O_{k+1} = (1 - w_k)O_k + w_k F_k, \quad \text{其中 } w_k = \frac{\hat{F}_k}{\sum_k \hat{F}_k}. \tag{10}$$

即使不使用其他任何零参考损失并仅使用简单的线性融合，这种方法也表现良好，这表明我们的因子具有有效性。注意，此时我们拥有比 SCI 小一个数量级的网络大小（表 2 中为 0.045 vs. 0.26 千个参数）。

扩展：我们的高光因子具有易于解释的特性，可以直接用于图像编辑工具（如 GIMP[80]、Photoshop[2]）中的图像操作，作为图层使用。我们展示了一个图像重新照明的示例，通过调整因子的颜色和混合模式来实现（图 1 左下角，图 7）。这表明我们的分解方法具有应用于复杂下游任务的潜力。我们探索了三个不同的图像增强任务：去雾、去雨和去模糊。这里，我们的目标是评估高光分解作为现有基础模型的预处理步骤的效果。我们选择了最近的 AirNet[45]，因为它允许在多个图像增强任务中进行实验，只需对主干结构进行微小修改。为了将我们的因子作为先验信息引入，我们将它们与原始输入连接在一起，并修改第一个卷积层的输入通道。请注意，我们没有引入新的损失函数或层，并直接逐一训练模型以完成三项任务：(i) 在 RESIDE 数据集[44]上的去雾任务；(ii) 在 Rain100L 数据集[93] 上的去雨任务；(iii) 在 GoPro 数据集上的去模糊任务。如图6和表5所示,我们的结果在感知上更令人满意，并且在多个任务方法上始终提高了之前报告的分数[45,97]。我们认为这得益于以基于光照的区域分类形式引入的结构先验信息，因为像素的光照强度和顺序取决于场景结构。更多结果请见补充材料。

图 7: 使用我们的因素作为第[80]层的可控可靠性应用（顶部：输入；底部：结果；左侧：分别编辑的光镜、室内颜色和室外强度）

局限性：与其他基础算法类似，我们的方法对初始化条件较为敏感[12,51]。作为一种启发式方法，我们使用数据集的均值作为初始化。另一种有待探

索的思路是根据每个输入动态调整，这有望进一步提升性能。

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**

表5:我们的因素可以在现有基础模型中引入结构，提高多个增强任务的性能

致谢： 我们感谢 TCS Foundation 和 Kohli Centre on Intelligent Systems 对本研究的支持。

5 结论

在本文中，我们提出了一种递归高光分解（RSF）及其在零参考 LLE 中的应用。通过将优化阶段展开为一个小型神经网络，我们以数据驱动的方式学习优化超参数。各因子通过网络融合，生成最终结果。我们还展示了 RSF 在图像重新照明以及去雾、去雨和去模糊等图像增强任务中的实用性。目前我们正在探索将 RSF 扩展到图像和谐化、前景抠图、白平衡、深度估计等应用，并将该技术扩展到可见光谱之外的其他信号。

伦理问题： 该工作旨在增强捕获的图像，我们未发现任何需要特别注意的伦理问题。

参考文献

[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] Yagız 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, Herve Jegou, 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 image 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, Jurgen 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 un-

reasonable 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