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 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. 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 recover 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

Figure 1. Specularity 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.
[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 U-Net-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.

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

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 exemplar 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-
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).
### 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 component and a specular component. According to dichromatic reflection model, image can be expressed as $I = \hat{I} + \hat{E}$. Here $\hat{I}$ indicates input mean and $\hat{E}$ indicates input sparsity:

$$\hat{I} = \frac{I}{\hat{E}}$$

where $\hat{E}$ is Frobenius norm regularizer and $\lambda$ is the sparsity parameter with higher values encouraging sparser results. Eq. (1) can be rewritten as Eq. (2) as ALIST A, who proved how all weight could be analytically obtained for a known dictionary, thereby leaving only step sizes and thresholds $\mu$ 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] unraveled 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 $E^k$ 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:

$$X_k^{t+1} = X_k^t - E_k^t$$

and hence can unroll them as neural network layers with learnable parameters $\alpha_t, w^1_t, w^2_t$ for each iteration $t \in [0, T]$. Based on the weight coupling between $w^1$ and $w^2$, Chen et al. [16] simplified Eq. (3) by deriving both $w^1$ and $w^2$ 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 $\mu$ 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] unraveled their ADMM updates into a network for robust principal component analysis.

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

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

with learnable parameters $(\alpha_t, w^1_t, w^2_t)$ for each iteration $t \in [0, T]$. Based on the weight coupling between $w^1$ and $w^2$, Chen et al. [16] simplified Eq. (3) by deriving both $w^1$ and $w^2$ 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 $\mu$ 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] unraveled their ADMM updates into a network for robust principal component analysis.
Unrolling: Based upon above discussion, we propose an unrolled network collecting all parameters in a single vector $\theta$. In each iteration $t$, we estimate three scalars: thresholds for both components ($\alpha, \beta$) and the step size ($\mu$). Hence for a factor $k$, we have $3T$ parameters $\theta^k := (\alpha^k, \beta^k, \mu^k)$ and overall we have only $3KT$ parameters $\theta := \{\theta^k\}_1^K$. Hence our model-driven factorization module is extremely light-weight compared to other decompositions (Tab. 2).

We propose the following novel factorization loss:

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

This constraints the ratio of signal energy in the $k^{th}$ factor compared to the input, to $\nu^k$. As $\nu^k$ increases for higher factors, our factorization loss relays 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 semantically-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.

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 $3 \times 3$ 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:

$$F^k = E^k - E^{k-1} \quad \text{where} \; F^1 = E^1. \quad \text{(7)}$$

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 $R^k$ rescale different image regions differently and are applied directly on the original image inside the curve adjustment equation [27] for the fused result:

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

The fused output is finally passed through a differentiable bilateral filtering layer $\Phi$ [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
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:

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

where \( \hat{\phi} \) represents the average value over a 16 × 16 window. Our third loss is the pixel-wise smoothing loss which controls the local gradients \( \nabla_x O \) in the final output:

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

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

\[ L = \lambda_f L_f + \lambda_c L_c + \lambda_e L_e + \lambda_s L_s. \]  

\[ 4. \text{ Experiments and Results} \]

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

<table>
<thead>
<tr>
<th>Paradigm</th>
<th>Traditional Model Based</th>
<th>Zero-reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method</td>
<td>LIME</td>
<td>DUAL</td>
</tr>
<tr>
<td>Params x10^3</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Lolv1 [87] (dataset split: 485/15, mean±: 0.05, resolution: 400 × 600)

Lolv2-real [95] (dataset split: 689/100, mean±: 0.05, resolution: 400 × 600)

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).

**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 takes less than 30 minutes.

**Datasets:** We evaluate our method using multiple LLE benchmark datasets (Lolv1 [87], Lolv2-real [95], Lolv2-syn [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 Natural Image
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], 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.

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 $L_{enc}$) 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 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 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.

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.
the next factor, weighted by the normalized mean:

\[ O^{k+1} = (1 - w^k)O^k + w^k F^k, \]

where \( w^k = \hat{F}^k / \sum_{k} \hat{F}^k \).

(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 [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


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


[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


[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


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


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


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


[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


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


[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


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


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


[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


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


