Learning Hierarchical Dynamics with Spatial Adjacency for Image Enhancement

Yudong Liang[†*]
School of Computer and Information Technology, Shanxi University & Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education
Taiyuan, China

Bin Wang[*]
School of Computer and Information Technology, Shanxi University & Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education
Taiyuan, China

Wenqi Ren
School of Cyber Science and Technology, Shenzhen Campus, Sun Yat-sen University
Shenzhen, China

Jiaying Liu
Wangxuan Institute of Computer Technology, Peking University
Beijing, China

Wenjian Wang
School of Computer and Information Technology, Shanxi University & Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education
Taiyuan, China

Wangmeng Zuo
School of Computer Science at Harbin Institute of Technology
Haerbin, China

ABSTRACT

In various real-world image enhancement applications, the degradations are always non-uniform or non-homogeneous and diverse, which challenges most deep networks with fixed parameters during the inference phase. Inspired by the dynamic deep networks that adapt the model structures or parameters conditioned on the inputs, we propose a DCP-guided hierarchical dynamic mechanism for image enhancement to adapt the model parameters and features from local to global as well as to keep spatial adjacency within the region. Specifically, channel-spatial-level, structure-level, and region-level dynamic components are sequentially applied. Channel-spatial-level dynamics obtain channel- and spatial-wise representation variations, and structure-level dynamics enable modeling geometric transformations and augment sampling locations for the varying local features to better describe the structures. In addition, a novel region-level dynamic is proposed to generate spatially continuous masks for dynamic features which capitalizes on the Dark Channel Priors (DCP). The proposed region-level dynamics benefit from exploiting the statistical differences between distorted and undistorted images. Moreover, the DCP-guided region generations are inherently spatial coherent which facilitates capturing local coherence of the images. The proposed method achieves state-of-the-art performance and generates visually pleasing images for multiple enhancement tasks, *i.e.*, image dehazing, image deraining and low-light image enhancement. The codes are available at https://github.com/DongLiangSXU/HDM.

CCS CONCEPTS

• **Computing methodologies** → **Computer vision**.

KEYWORDS

Hierarchical Dynamics, Image Enhancement, Dark Channel Priors, Region-level Dynamics, Spatial Adjacency, Depth

ACM Reference Format:
Yudong Liang, Bin Wang, Wenqi Ren, Jiaying Liu, Wenjian Wang, and Wangmeng Zuo. 2022. Learning Hierarchical Dynamics with Spatial Adjacency for Image Enhancement. In *Proceedings of the 30th ACM International Conference on Multimedia (MM '22), October 10–14, 2022, Lisboa, Portugal.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3503161.3548322

1 INTRODUCTION

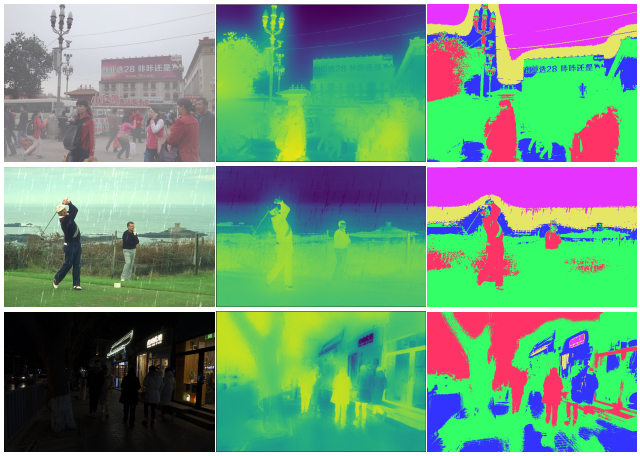
Image enhancement is a classic low-level computer vision problem with high practical values, which has attracted lots of attentions from the communities. Various kinds of degradations exist in the low-quality images, dues to adverse imaging conditions, such as fog, rain, or low light, which are quite diverse and hard to describe by a universal physical model. Multiple priors and skillfully handcrafted features are exploited to alleviate the ill-posedness of the image enhancement problems [3, 14, 36, 37]. He *et al.* [14] proposed the dark channel prior (DCP) for image dehazing. Unfortunately, although handcrafted methods have achieved satisfying results for specific pictures, the priors are easily violated in real applications.

With rapid developments of deep learning, convolutional neural networks (CNN) based dehazing methods have largely boosted the performances of the area [10, 23, 26]. Standard convolutional filters are shared across the spatial domain, producing responses to specific structures or features since some local structures repeatedly appear within a single image or across many pictures. A

[*]Both authors contributed equally to this research, Yudong Liang is the corresponding author (liangyudong006@163.com)

MM '22, October 10–14, 2022, Lisboa, Portugal

ACM ISBN 978-1-4503-9203-7/22/10…$15.00
https://doi.org/10.1145/3503161.3548322



(a) Input (b) Dark channel values (c) Dynamic region masks

Figure 1: Hazy, rainy and low-light images and the corresponding dark channel values as well as the DCP-guided region masks by our method.

tremendous number of filters are exploited in deep models to recover the potential structures in the target images. However, most of the current deep models perform inference in a static manner and fix the parameters for different inputs, which severely limits the further improvement of model performances. On the other hand, the degradations or adverse imaging conditions are always non-homogeneous or non-uniform, whether it is fog or rain or low light. In these static manners, diverse degraded images or nonhomogeneous areas in low-quality images are processed with the same filters from a trained deep model, which inevitably lacks representation power conditioned on various inputs and leads to some failure cases.

To further improve the representation ability of the deep models, dynamic networks [13] that adapt the model structures or parameters during inferences have recently aroused lots of interests from communities [5, 6, 17, 28]. The dynamic models have great potentials to restore clear images according to different areas of nonuniform low-quality inputs. Typical examples are attention mechanisms which are widely applied for image enhancement tasks [23, 26, 32], where attention weights are calculated to focus the important part of the input, for example, adapting parameters individually for each pixel. However, adapting parameters individually for each pixel may ignore the local coherence of the images and lose the translation invariance of the convolutional operations. To solve this problem, Dynamic Region-Aware Convolution [5] is proposed to assign multiple convolutional filters to different regions separately and share the same filters in each region, which obtains great performance gain. Nevertheless, the 'Region' defined in [5] has no direct connotations for spatial adjacency but shares similarities in high dimensional feature spaces captured by the guided mask separations.

It is quite challenging that the region generation for the enhancement process should consider the semantic information and the degradations. Intuitively, the regions are a group of connected pixels with similar properties which should have spatial adjacency within the regions. Modeling spatial adjacency of the region-level dynamics for local coherence is still absent. Thus, balancing representation ability and modeling local coherence remains an important issue for developing a dynamic model. Moreover, how to develop and combine dynamic techniques for better performance still needs to be explored.

To face these challenges, a hierarchical dynamic mechanism (HDM) with spatial adjacency is proposed for image enhancements, which hierarchically adapts the parameters and features of the model channel-spatial-wise, structure-wise, and region-wise from local to global. The channel-spatial-level dynamics are built by the attention model, which enable channel-wise and spatial-wise variations for representations and obtain the finest dynamics for the input. The structure-level dynamics apply deformable convolutions [6] to model geometric transformations and augment sampling locations for describing the structures. The region-level dynamics propose to generate spatial connected region masks which apply different filters to different regions adaptively and keep translation-invariance property in each region. Dark Channel Priors(DCP) [14] are revisited to measure the degradation degrees of regions to generate the DCP-guided region-level dynamic masks, which inherently generate local coherent regions where the pixels are spatially connected. Instead of simultaneously learning filters and guided masks, the DCP-guided region-level dynamic masks are generated without learning which is more efficient and enables better optimizations for filters. For hazy images, our DCP-guided region-level dynamics may exploit the depth information for better dynamic region generations as DCP priors relates to depth information. In Fig. 1, the dark channel values clearly reflect a grouping of the local pixels which leads to a good separation of regions for region-level dynamics. The dcp-guided region separations are consistent with separations according to the semantic and degradation information. The core of the proposed hierarchical dynamic mechanism is the local to global philosophy. The pixel ranges of the input involved in determining the output dynamic features or dynamic parameters are gradually enlarged, that is, from local to global. The components in each dynamic level can evolve as the techniques of the dynamic model develops. Performances are obviously dropped by different cascading orders of the same dynamic components, which demonstrate the importance of the local to global philosophy.

Our framework has achieved state-of-the-art performances with a relatively small parameter number and restored visually pleasing clear images for different image enhancement tasks. It could be further improved when specialized techniques or loss functions for certain degradation are combined. In addition, our models trained by different types of degraded training images could be combined to better solve the real degraded images, such as applying deraining and dehazing models to real rainy images where both rainy and hazy degradations exist. To summarize, our paper makes the following main contributions,

a) A hierarchical dynamic mechanism (HDM) is proposed to gradually adapt the model parameters and build local and global dynamics. The channel-spatial-level dynamics enable channel-wise and spatial-wise representation variations. The structure-level dynamics enable modeling geometric transformations and augmenting sampling locations for varied local features to better describe

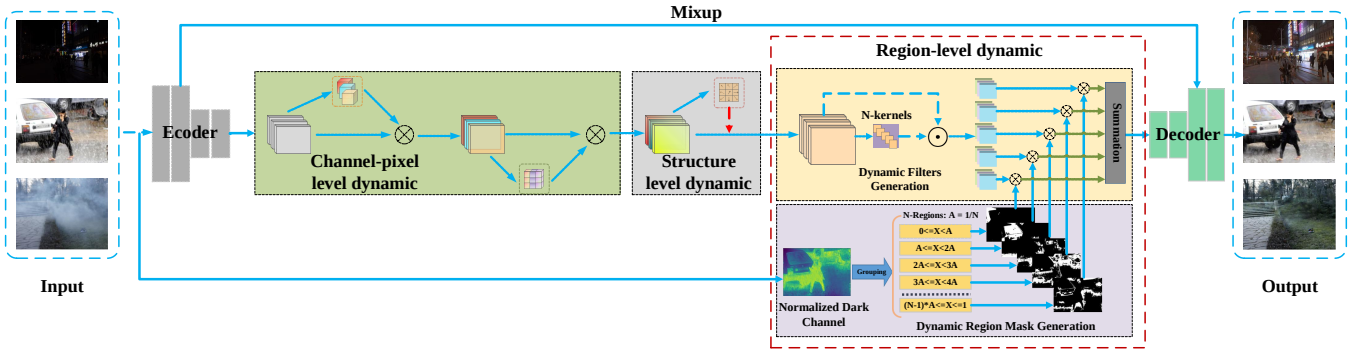


Figure 2: The architecture of the deep model with the proposed DCP-guided hierarchical dynamic mechanism (HDM). Only a rainy image is applied to illustrate the dynamic region mask generation due to the space limitation.

the structures. Finally, the region-level dynamics integrate the structure-level dynamic features and obtain better local coherence.

b) A DCP-guided region-level dynamic component is designed to measure the degradation degrees according to DCP priors and provides spatial adjacency within the region, facilitating local coherence for the images much more efficiently.

c) The proposed DCP-guided hierarchical dynamics demonstrate the state-of-the-art performances for different image enhancement tasks, *i.e.*, dehazing, deraining and low-light image enhancement.

2 RELATED WORK

Generally, the existing image enhancement methods can be classified into learning-based [9, 17, 25, 34, 52] and non-learning-based image enhancement methods [3, 14, 36, 37]. For deep learning based image enhancement methods, dynamic models have demonstrated obvious improvements over static deep models and have been a hot topic in the last few years. Dai *et al.*[6] introduced deformable convolutions and deformable ROI pooling for detections and semantic segmentations to enhance the geometric transformation modeling capability of deep models. Qin *et al.*[32] applied channel and spatial attentions as well as feature attention modules for image dehazing, which brought large improvements. Essentially, the dynamic models generate or re-weight the parameters or features of the inference models conditioned on inputs, which has dramatically expanded the parameter spaces and increased the model capacities as well as the representation ability.

Dynamic Region-Aware Convolution (DRConv) [5] achieves state-of-the-art performances on classification, face recognition, detection, and segmentation tasks. DRConv applies masks to separate regions, enables weight sharing within the separate regions, and captures similarities in high dimensional feature spaces. The mask M_a can be learned from the input deep feature F_R^{in}. In specific, if N regions are separated, N groups of filters would be generated and applied to calculate N groups of feature maps f_i $(i = 1, ..., N)$. Then these N groups of guided masks are calculated as

$$M_a^i(x) = \delta(i == argmax(f_1(x), ..., f_i(x), ...f_N(x))), \quad (1)$$

where δ is an indicator function. The resulted guided masks are the N groups of the one-hot feature maps. To enable the end-to-end learning for the mask generations, the softmax operation is exploited during the error propagation process resulting a very dispersive or sparse mask set. The region dynamic features can be calculated as multiplications of guided masks M_a and the input deep features F_a. However, spatial adjacency is not involved in this work and region-level dynamics remain open issues.

2.1 The Dark Channel Prior (DCP)

DCP [14] is a famous handcrafted dehazing method by which the transmission map can be easily estimated. He *et al.*[14] statistically found the minimum values of rgb channel for each pixel form a dark channel, which should approximate zero for the hazy-free images. The dark channel can be calculated as:

$$J_{dark} = \min_{c \in \{r,g,b\}} (J^c), \quad (2)$$

where J^c is the c channel of hazy-free images J. Thus, the dark channel prior is violated if the dark channel intensity is greater than a threshold. The DCP priors are mainly utilized in the dehazing problem. In fact, the DCP priors hold for the targeted enhancement results of ideal image quality. In this paper, DCP priors are exploited for universal image enhancement problems and the dark channel intensities are grouped to form region masks for dynamic models.

3 OUR APPROACH

Real-world low-quality images are always nonhomogeneous and diverse but with local coherence. A hierarchical dynamic mechanism (HDM) that enables feature dynamics from local to global is proposed, which gradually builds channel-spatial-wise, structure-wise, and region-wise dynamics. An encoder-decoder deep model is applied with the proposed hierarchical dynamic mechanism. As shown in Fig. 2, the proposed model first applies conventional convolutions to downsample the images, and then the proposed hierarchical dynamic components are utilized to enhance the feature mapping process. Finally, our model employs deconvolutions to upsample the extracted feature maps to the original size and uses additional conventional convolutions to alleviate the aliasing effect. The number of filter channels is fixed as 64 in the whole network, which is a relatively small feature number compared with recent enhancement networks [7, 32]. All the activation functions apply the ReLU function. In addition, the L1 loss and perceptual loss are applied in our implementations.

The downsampling process utilizes two stride convolution layers with a stride of 2 to downscale the feature map to quarter size. The downsampling process largely reduces the computational burdens for the following feature mapping, especially the hierarchical dynamic feature mapping process. Instead of feature concatenations like U-Net [35] or feature summations like residual networks [15], the adaptive Mixup operations [41] are integrated to enable information flow from downsampling layers to upsampling layers.

The hierarchical dynamic blocks consist of channel-spatial-level, structure-level and region-level dynamic components sequentially. The region-level dynamics are achieved by generating a DCP-guided mask from DCP priors and then applying each mask separately to the extracted features. The generated DCP-guided dynamic masks inherently impose spatial coherence constraints for dynamic region generations and benefit from the spatial adjacency within the regions to better capture the information. As DCP priors could be applied to calculate the transmission map for image dehazing problem, the generated mask may exploit the inner relationship between image enhancement and depth.

3.1 Channel-Spatial-Level Dynamics

The degradation is nonhomogeneous which should be adaptively recovered across spatial domains. Moreover, as demonstrated by [14], the distributions of pixel values in each channel of the degraded images are different which should be treated differently. Inspired by [32], channel attention and spatial attention are cascaded in our model to adapt the feature map channel-wise and spatial-wise which generate channel-spatial-level dynamics as Eq. (3) and Eq. (4). For channel-level dynamics,

$$\begin{aligned} W_c &= \sigma(Conv(ReLU(Conv(GAP(F_{in}^c))))), \\ F_{out}^c &= F_{in}^c \otimes W_c \end{aligned} \tag{3}$$

where $\otimes$ stands for the element-wise multiplication, F_{in}^c denotes the input for the channel attention component and W_c denotes the attention weights calculated for the different channels. Sigmoid function σ, convolution operation $Conv$ and ReLU function are applied after a global average pooling function (GAP) for each channel of input feature F_{in}^c.

For spatial-level dynamics, the feature can be adapted as:

$$F_{out}^{sp} = F_{in}^{sp} \otimes \sigma(Conv(ReLU(Conv(F_{in}^{sp})))), \tag{4}$$

where F_{in}^{sp} is the input for the spatial attention component which equals F_{out}^c. The channel-spatial-level dynamics enable features in different channels and different spatial locations varied adaptively, which could extract the pixel varied features conditioned on the input and achieve greater representation power.

3.2 Structure-Level Dynamics

The focus of the channel-spatial-level dynamics is mainly limited to the pixels individually which loses the translational invariance. Pixel varied features are not good at modeling large irregular variations for the input, especially for the distorted images. To benefit from the local coherence of structures as well as increasing the representation power for geometric transformations, deformable convolutions are applied right after channel-spatial-level dynamic components to better capture the features of local structures. The calculated channel-spatial-level dynamic features F_{out}^{sp} are fed into the structure-level dynamic component. The regular sampling grid of the conventional convolution is replaced by a calculated sampling offset. The calculation of the deformable convolution can be expressed as

$$F_s^{out}(x) = \sum_{x_n \in G} W_s(x_n) \cdot F_s^{in}(x + x_n + \triangle x), \tag{5}$$

where $W_s(x_n)$ are the wights, G is the traditional regular sampling grid, x_n is the regular sampling points and $\triangle x$ is the calculated sampling offset, F_s^{in} and F_s^{out} denote the input of the operations and the calculated structure-level dynamic features respectively. The spatial sampling locations are largely augmented from the learning of the restoration task which can better represent the features for the irregular structures. After the structure-level dynamic components, the focus of the features is obviously enlarged.

3.3 Region-Level Dynamics

The structure-level dynamics mainly focus on local structures as the irregular sampling offsets of deformable convolutions are still limited to local structure areas. To better integrate the local information for modeling the local coherence, region-level dynamic components are cascaded after structure-level dynamic components.

The region-level dynamics mean to apply different convolution filters according to the different input regions. The region separation is accomplished by an efficient DCP-guided mask generation. The corresponding filters for each region are generated by a filter generation network, which are applied for the corresponding input features according to the generated masks based on DCP priors.

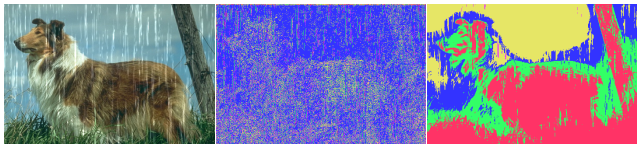
Dynamic region masks are groups of the one-hot feature maps that spatially indicate which pixel belongs to which region. The mask is generated according to a grouping of normalized dark channel values ($[0, 1]$) as Fig.2. N ($N = 5$) equally spaced intervals are applied to decide the separation of the regions. As the intensities of the dark channel for an ideal high-quality image should approximate zeros, the intensity deviations from the zeros in the dark channel could reflect the degradation-degrees for the low-quality images. DCP priors are priors to describe statistical properties of images with ideal image qualities although the DCP priors are mainly applied in the image dehazing problems. In the implementations, as the soft matting step applied by the DCP method is quite time-consuming, the soft matting step is abandoned.

For efficiency, different groups of filters for different regions are applied for the whole input F_s^{out} to get the dynamic features, then the dynamic features are multiplied by the generated dynamic region masks. Finally, a summation is applied to obtain the final results for the region-level dynamic components as

$$F_R^{out} = \sum_i F_R^i \otimes M_{dcp}^i = \sum_i (F_s^{out} \odot W_i) \otimes M_{dcp}^i \tag{6}$$

where F_R^i is the i_{th} feature map calculated by the filters W_i, M_{dcp}^i is the generated DCP-guided region masks, $\otimes$ denotes the pixel-wise multiplications and $\odot$ denotes the convolutional operations. Thus the network splits into two branches for the mask generation and corresponding filter generations as Fig.2.

The components of hierarchical dynamics are not limited to the proposed components and could evolve as the development



(a)Input (b) Drconv-mask (c) DCP-mask

Figure 3: A visual comparison of DCP-guided masks and the learned masks by DRconv [5]

of the dynamic models. The existing techniques such as [5] could be also inserted into our framework. Nevertheless, the proposed DCP-guided mask generation is independent of the corresponding dynamic feature learning process of certain regions, which leads to a faster and better feature learning process compared with our hierarchical dynamic mechanisms equipped with Dynamic Region-aware convolutions [5] denoted as HDM-F.

3.4 DCP Priors for Image Dehazing

DCP priors directly relate to the transmission map for image dehazing problem. Light attenuates in propagation and the transmission $t(x)$ decays exponentially in the medium with the scene distance or depth $d(x)$,

$$t(x) = e^{-\beta d(x)}, \tag{7}$$

where x indicates the location of the pixel in the image, β is a constant value. Clustering the dark channel values means grouping the negative exponential space of depth information.

Although Eq. 7 and estimated transmissions by DCP method may have some biases for the nonhomogeneous hazy images, it produces a rough prediction of the depth ranges that could be applied for generating a mask for our dynamic region separations. The DCP-guided dynamic calculations exploit the inner relationship between depth and dehazing process. Moreover, the depths are mostly local coherent and the spatial coherence constraints for dynamic region generations are inherently imposed. In Fig. 3, visualization comparisons of the DCP-guided masks and learned masks by [5] for an image are represented. The masks by [5] for the region generations include some rich features but are dispersive or sparse, while the proposed DCP-guided masks have a better spatial coherence and spatial adjacency within each region which could benefit the following enhancement process.

4 EXPERIMENTS

Our approaches denoted as HDM are evaluated on different image enhancement tasks of extreme weather or bad illuminations: **(a)** image dehazing, **(b)** image deraining, **(c)** low-light enhancement. The evaluated low-quality images mainly include synthetic images which are generated following the assumed physical models and real-world images.

Our proposed network is implemented by PyTorch 1.4.0 with one NVIDIA TITAN xp GPU. The models are trained using an Adam optimizer with exponential decay rates β_1 and β_2 of 0.9 and 0.999, respectively. The initial learning rate and batch size are set to 0.0002 and 16, respectively. The cosine annealing strategy is applied to adjust the learning rate and the total number of iterations is 100 epochs. For the three different types of data, the applied losses are all weighted combinations of L1 loss and perceptual loss, which are commonly used in various image enhancement tasks. All the report PSNR are tested on the RGB color space.

For all the three enhancement tasks, our method achieves the best performances with a small parameter number, which well balances performances and complexities. More visual comparisons and details could be found in the supplementary materials.

4.1 Image Dehazing

For synthetic datasets, the widely applied large-scale benchmark RESIDE dataset [22] is utilized. Following the common practice, the subsets of RESIDE, Indoor Training Set (ITS) and Synthetic Objective Testing Set (SOTS) are used for training and testing respectively. Two most commonly applied real datasets: NH-HAZE [2] and Dense-Haze [1] are investigated for real applications. Typical and the state-of-the-art dehazing methods are compared in Table 1. The compared methods include: traditional method **DCP** [14], physical model-driven deep learning method **DCPDN** [48] (For a fair comparison, we have finetuned it on the RESIDE dataset) and end-to-end dehazing method such as **AOD** [21], **GridDehaze** [30], **FFA** [32], **MSBDN** [7], **KDDN** [16] and **AECR** [41]. Since the code of **FDU** [8] is not released, only the metrics reported in their paper are referenced for comparisons.

Our method outperforms all the compared methods for all the test datasets. Visual comparisons are given in Fig.4 which proves our method obtains better visual results for real nonhomogeneous hazy images. Our method has restored fewer artifacts and correct color tunes of the image. For example, our result of the tree canopy in the first picture is more faithful and there are severe color shifts in the grounds of other restorations except ours.

4.2 Image Deraining

Careful comparisons with 9 state-of-the-art are performed on synthetic benchmarks Rain100L, Rain200H and DID-data in Table 3. Rain100L and Rain200H are synthesized with one type of light rain streaks and heavy rain streaks of five streak directions respectively [42], while DID-data [50] emphasizes different densities of rain and generates rain streaks with different orientations and scales. In addition, the real rainy images provided by [31, 39] are evaluated by our model with the parameters learned from Rain200H. The compared methods include a traditional method: **LP** [27], and 8 deep learning-based methods: **DID** [49], **SPANet** [39], **UMRL** [45], **PReNet** [33], **MSPFN** [18], **RCDNet** [38], **RLNet** [4], **MPRNet** [47].

Our method has achieved better PSNR/SSIM performances and visual results in all the synthetic data. Our method restores more texture for the bear in Fig. 5. More importantly, our method works effectively for the real rainy images, where both rainy and hazy distortions existed. Our model removes all the rain streaks with less artifacts in Fig.6. The real hazy images could be sequentially processed by our HDM models trained with Rain200H rainy dataset and ITS dehazing dataset respectively. With a combination of deraining and dehazing operations, our method could generate significant better enhancement results for real rainy images.

4.3 Low-Light Enhancement

The LOL [40] dataset is tested for comparisons which is commonly used in low-light enhancement tasks. The LOL dataset consists of

Table 1: Quantitative PSNR / SSIM comparisons on different hazy datasets as well as comparisons of parameter numbers.

	DCP	AOD-Net	DCPDN	GridDehaze	FFA	MSBDN	KDDN	FDU	AECR	Ours
SOTS	15.09/0.765	19.82/0.818	25.64/0.927	32.16/0.984	36.39/0.989	33.79/0.984	34.72/0.985	32.68/0.976	37.09/0.990	38.56/0.991
NH-HAZE	10.57/0.520	15.40/0.569	14.55/0.601	13.80/0.537	19.87/0.692	19.23/0.706	17.39/0.590	-/-	19.88/0.717	22.48/0.737
Dense-Haze	10.06/0.386	13.14/0.414	12.71/0.342	13.31/0.368	14.39/0.452	15.37/0.486	14.28/0.407	-/-	15.80/0.466	15.97/0.507
Params(Mb)	-	0.002	66.89	0.96	4.68	31.35	5.99	-	2.61	2.32

Table 2: Quantitative PSNR / SSIM comparisons on different rainy datasets as well as comparisons of parameter numbers.

	LP	DID	SPANet	UMRL	PReNet	MSPFN	RCDNet	RLNet	MPRNet	Ours
Rain100L	29.11/0.881	23.79/0.773	27.85/0.881	32.39/0.921	36.28/0.979	33.50/0.948	38.60/0.983	37.38/0.980	34.95/0.960	38.94/0.983
Rain200H	14.26/0.420	15.54/0.520	13.27/0.412	23.01/0.744	27.64/0.884	24.30/0.748	28.83/0.886	28.87/0.895	27.63/0.874	29.93/0.898
DID-data	22.46/0.801	27.93/0.861	22.96/0.720	30.05/0.891	30.40/0.891	30.34/0.881	29.81/0.859	32.62/0.917	31.29/0.894	32.91/0.919
Params(Mb)	-	0.37	0.28	0.98	0.17	3.17	13.22	4.73	20.15	2.32

Table 3: Quantitative PSNR / SSIM comparisons on LOL dataset and Quantitative NIQE comparisons on unlabeled low-light dataset as well as comparisons of parameter numbers.

	LIME	RetinexNet	KinD	Zero-Dce	EnlightenGAN	RUAS	KindD++	Zero-Dce++	MIRnet	Ours
LOL(PSNR	14.92	13.10	20.87	15.51	15.64	18.23	21.30	15.35	24.14	23.45
SSIM)	0.516	0.429	0.810	0.553	0.578	0.717	0.823	0.570	0.83	0.852
Params(Mb)	-	0.84	8.54	0.08	8.64	0.01	8.28	0.09	31.79	2.32
DICM(NIQE)	3.5347	4.4654	4.1383	3.5602	3.5458	5.2103	3.7860	3.5391	3.5642	3.5328
LIME(NIQE)	3.5593	3.6927	3.7111	3.3350	3.4895	3.9603	3.4936	3.5430	3.8161	3.4692
Darkface(NIQE)	3.6245	4.6648	4.1061	3.3723	3.1047	4.3489	3.1902	3.3578	3.2966	2.9255

500 pairs of images captured in real scenes and each pair contains a low-light image and the corresponding normal-light image. To investigate the generalization ability of the model, the model trained from the LOL dataset is also evaluated on three unlabeled datasets, *i.e.*, LIME[12], DICM[20], Darkface[44].

For low-light enhancement task, our method is compared with 8 state-of-the-art methods, which includes 1 traditional method: **LIME**[12], and 7 deep learning-based methods: **RetinexNet**[40], **KinD**[52], **Zero-DCE**[11], **EnlightenGAN**[19], **DRBN**[43], **RUAS**[29], **MIRnet** [46] and **Zero-DCE++**[51].

Our method obtain the best performances in LOL [40] dataset and obtains better or comparable no-reference image quality assessment index NIQE in three unlabeled datasets. Our method especially works well for the largest real-world low-light image dataset Darkface. In Fig. 7, our results appear lighter but keep the details of the original images. The pulled wires in the sky of the second picture in Fig.7 can be seen clearly while other methods may lose these details or introduce artifacts.

4.4 Ablation Study

4.4.1 The importance of local to global philosophy. The cascade order of the dynamic components for the hierarchical dynamic mechanisms largely affects the performances as shown in Table 4 which demonstrates the importance of local to global philosophy. Blind combinations of different techniques could largely degrade the performances.

Table 4: The comparisons of different cascaded orders for different dynamic components on NH-HAZE dataset, P,S,R indicate the channel-spatial-level, structure-level, region-level dynamic components respectively.

NH-HAZE	PSR(Ours)	PRS	SRP	SPR	RPS	RSP	AECR
PSNR (dB)	**22.48**	15.12	11.44	18.84	17.23	11.55	19.88

4.4.2 Architectures with different hierarchical dynamic deep components. The influences of abandoning some dynamic deep components on performances and parameter numbers are investigated in Table 5. The performances of five architectures are compared on RESIDE and NH-HAZE datasets. The five architectures are (a) **B**: baseline model that only encoder-decoder structures are applied and no dynamic components are utilized, (b) **B+FA**: channel and spatial attentions are applied in addition to the baseline model, (c) **B+FA+S**: deformable convolutions are applied in addition to the setting (b) to further obtain structure-level dynamics, (d) **HDM-F**: models following our hierarchical dynamic mechanism but with the learned mask by [5].(e) **HDM**: models following our DCP-guided hierarchical dynamic mechanism. The performances are constantly improved as more components of the hierarchical dynamics are constructed.

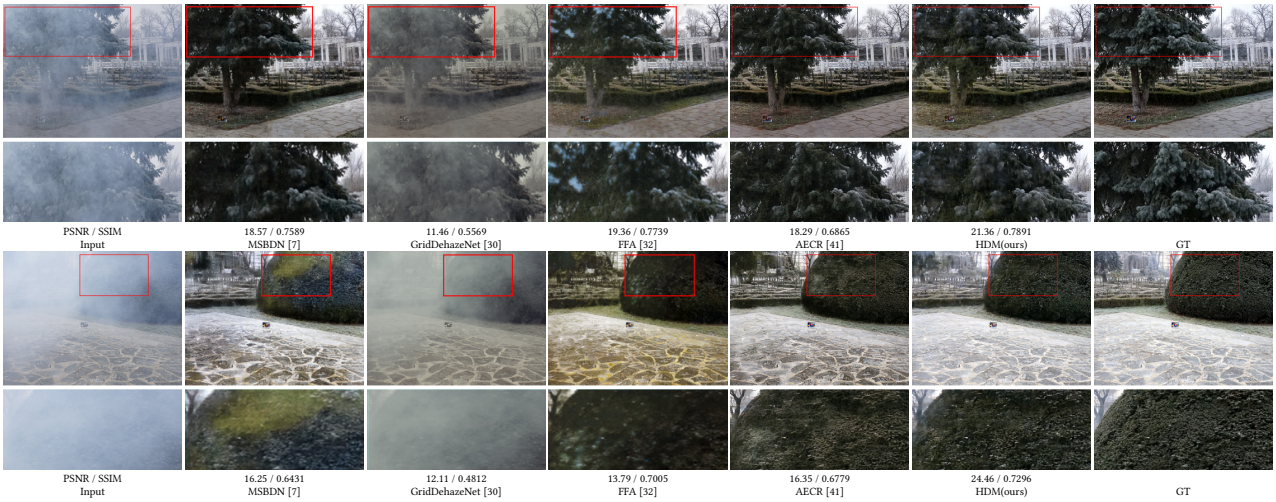


Figure 4: Visual comparisons of different methods for the real nonhomogeneous hazy images.

Figure 5: Visual comparisons of different methods for Rain200H

Figure 6: Visual comparisons of different methods for Real rainy images from Internet provided by [39]

Figure 7: Visual comparisons of different methods for Real dark images

Although adding dynamic deep components increases the parameter number of the model, the performances are largely boosted. The major growth of parameter numbers for the hierarchical dynamic deep components comes from the channel-spatial-level dynamic components but is still acceptable. The channel-spatial-level dynamics enable local varied representation for each pixel.Parameter numbers are only slightly increased when the structure-level and the region-level dynamic components are further added, but significant performance gains have been achieved. Table 5 proves the effectiveness of the proposed hierarchical dynamic mechanism.

In Fig. 8, our hierarchical dynamic mechanism gradually improves the enhancements as channel-spatial-level, structure-level,

Table 5: The comparisons of PSNR / SSIM and parameter numbers (Mb) of architectures with different hierarchical dynamic deep components.

	SOTS	NH-HAZE	Parms
a) **B**	28.74/0.9631	14.78/0.6064	0.49
b) **B+FA**	33.53/0.9812	16.46/0.6089	2.19
c) **B+FA+S**	36.19/0.9849	16.62/0.6268	2.24
d) **HDM-F**	37.57/0.9902	20.77/0.7267	2.37
e) **HDM**	**38.56/0.9909**	**22.48/0.7373**	2.32

Table 6: PSNR/SSIM comparisons of different region generation manners.

	GT-Depth	**Predicted Depth**	**DRconv**[5]	**DCP**(Ours)
SOTS	38.53/0.9908	36.11/0.9811	37.57/0.9902	**38.56/0.9909**
NH-HAZE	-/-	9.43/0.2579	20.77/0.7267	**22.48/0.7373**

region-level dynamic components are added. Applying Our DCP-guided dynamic region generation method further gets rid of the thick haze in the distant places marked by the red rectangles.Compared with images (b) in Fig. 8, structure-level dynamic components enhance some textures of the enhancements such as the top of the slide. Applying the feature-guided region dynamic component (HDM-F) further captures some general characteristics of the restoration, for example, the color shifts in (d) of Fig. 8 are largely reduced compared with (b) and (c), but some minor shifts still exist. Finally, the DCP-guided dynamic component improves the restorations in distant places such as the background marked by the red rectangle, benefiting from exploiting the inner-relationship between image dehazing and depth information. The color shifts are further reduced such as the image area of the yellow gym equipment.

4.4.3 The impact of different region generation manners. How to generate regions is the key factor of the region-level dynamics. In this section, for image dehazing task, dynamic region masks generated by groundtruth depth information, by depth predictions of a pretrained depth estimation model for clear images, by Dynamic Region-aware convolutions [5] and by DCP priors are compared in Table 6. It is clear that the depth information is really helpful to the region generations for image dehazing task. Our proposed DCP-guided region generations perform much better than the learned region generations by [5]. As shown by Fig. 3, although both masks capture some meaningful features, our DCP-guided generated regions are spatial continues and keep better spatial adjacency within the regions while learned masks by DRconv [5] are quite dispersive and sparse. Inaccurate depth predictions would be harmful to the region generations and final performances.

4.4.4 The influence of the region mask number . The numbers of the DCP-guided region masks applied in our paper are 5. The numbers of region masks affect the performances as the number relates to the granularity of the region separations and the coherence within each region. Too coarse separations can not describe the uneven distributions of the degraded images. Too fine separations would fail to preserve the global information and also lead to heavier computational burdens. The variations of PSNR for NH-Haze dataset vs. region mask number is represented in the supplementary materials.

Figure 8: Visual comparisons of enhancements by deep architectures with different dynamic components for the ablation study of the hierarchical dynamic mechanism.

5 CONCLUSION

In this paper, we propose a DCP-guided hierarchical dynamic mechanism from local to global that gradually builds channel-spatial-level, structure-level, and region-level dynamics. DCP-guided region-level dynamics, which measure the discrepancy between degradation images and images of ideal image qualities, could implicitly impose spatial coherency constraints inherently for better feature representations and keep spatial adjacency within regions. Our algorithm achieves state-of-the-art performances and restores visually pleasing results for image dehazing, image deraining and low-light image enhancement tasks with a small parameter number.

ACKNOWLEDGMENTS

This work is partially supported by the National Natural Science Foundation of China (Nos. 61802237, U21A20513, U19A2073, 621724090, 62076154), the Natural Science Foundation of Shanxi Province (201901D211176), Scientific and Technological Innovation Programs of Higher Education Institutions in Shanxi (STIP) (2019L0066), the Science and Technology Major Project of Shanxi Province (202101020101019), the Key R&D program of Shanxi Province (International Cooperation, 201903D421050, 201903D421041).

REFERENCES

[1] Codruta O Ancuti, Cosmin Ancuti, Mateu Sbert, and Radu Timofte. 2019. Dense-haze: A benchmark for image dehazing with dense-haze and haze-free images. In *2019 IEEE international conference on image processing (ICIP)*. IEEE, 1014–1018.

[2] Codruta O. Ancuti, Cosmin Ancuti, and Radu Timofte. 2020. NH-HAZE: An Image Dehazing Benchmark with Non-Homogeneous Hazy and Haze-Free Images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (Washington, US) *(IEEE CVPR 2020)*.

[3] Turgay Celik and Tardi Tjahjadi. 2011. Contextual and variational contrast enhancement. *IEEE Transactions on Image Processing* 20, 12 (2011), 3431–3441.

[4] Chenghao Chen and Hao Li. 2021. Robust representation learning with feedback for single image deraining. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 7742–7751.

[5] Jin Chen, Xijun Wang, Zichao Guo, Xiangyu Zhang, and Jian Sun. 2021. Dynamic region-aware convolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 8064–8073.

[6] Jifeng Dai, Haozhi Qi, Yuwen Xiong, Yi Li, Guodong Zhang, Han Hu, and Yichen Wei. 2017. Deformable convolutional networks. In *Proceedings of the IEEE international conference on computer vision*. 764–773.

[7] Hang Dong, Jinshan Pan, Lei Xiang, Zhe Hu, Xinyi Zhang, Fei Wang, and Ming-Hsuan Yang. 2020. Multi-scale boosted dehazing network with dense feature fusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2157–2167.

[8] Jiangxin Dong and Jinshan Pan. 2020. Physics-based feature dehazing networks. In *European Conference on Computer Vision*. Springer, 188–204.

[9] Alona Golts, Daniel Freedman, and Michael Elad. 2019. Unsupervised single image dehazing using dark channel prior loss. *IEEE Transactions on Image Processing* 29 (2019), 2692–2701.

[10] Jie Gui, Xiaofeng Cong, Yuan Cao, Wenqi Ren, Jun Zhang, Jing Zhang, and Dacheng Tao. 2021. A Comprehensive Survey on Image Dehazing Based on Deep Learning. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI 2021)*.

[11] Chunle Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, and Runmin Cong. 2020. Zero-reference deep curve estimation for low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 1780–1789.

[12] Xiaojie Guo, Yu Li, and Haibin Ling. 2016. LIME: Low-light image enhancement via illumination map estimation. *IEEE Transactions on image processing* 26, 2 (2016), 982–993.

[13] Yizeng Han, Gao Huang, Shiji Song, Le Yang, Honghui Wang, and Yulin Wang. 2021. Dynamic neural networks: A survey. *arXiv preprint arXiv:2102.04906* (2021).

[14] Kaiming He, Jian Sun, and Xiaoou Tang. 2010. Single image haze removal using dark channel prior. *IEEE transactions on pattern analysis and machine intelligence* 33, 12 (2010), 2341–2353.

[15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.

[16] Ming Hong, Yuan Xie, Cuihua Li, and Yanyun Qu. 2020. Distilling image dehazing with heterogeneous task imitation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 3462–3471.

[17] Xu Jia, Bert De Brabandere, Tinne Tuytelaars, and Luc V Gool. 2016. Dynamic filter networks. *Advances in neural information processing systems* 29 (2016), 667–675.

[18] Kui Jiang, Zhongyuan Wang, Peng Yi, Chen Chen, Baojin Huang, Yimin Luo, Jiayi Ma, and Junjun Jiang. 2020. Multi-scale progressive fusion network for single image deraining. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 8346–8355.

[19] Yifan Jiang, Xinyu Gong, Ding Liu, Yu Cheng, Chen Fang, Xiaohui Shen, Jianchao Yang, Pan Zhou, and Zhangyang Wang. 2021. Enlightengan: Deep light enhancement without paired supervision. *IEEE Transactions on Image Processing* 30 (2021), 2340–2349.

[20] Chulwoo Lee, Chul Lee, and Chang-Su Kim. 2012. Contrast enhancement based on layered difference representation. In *2012 19th IEEE international conference on image processing*. IEEE, 965–968.

[21] Boyi Li, Xiulian Peng, Zhangyang Wang, Jizheng Xu, and Dan Feng. 2017. Aod-net: All-in-one dehazing network. In *Proceedings of the IEEE International Conference on Computer Vision*. 4770–4778.

[22] Boyi Li, Wenqi Ren, Dengpan Fu, Dacheng Tao, Dan Feng, Wenjun Zeng, and Zhangyang Wang. 2018. Benchmarking single-image dehazing and beyond. *IEEE Transactions on Image Processing* 28, 1 (2018), 492–505.

[23] Chongyi Li, Chunle Guo, Ling-Hao Han, Jun Jiang, Ming-Ming Cheng, Jinwei Gu, and Chen Change Loy. 2021. Low-light image and video enhancement using deep learning: a survey. *IEEE Transactions on Pattern Analysis & Machine Intelligence* 01 (2021), 1–1.

[24] Chongyi Li, Chunle Guo, and Chen Change Loy. 2021. Learning to Enhance Low-Light Image via Zero-Reference Deep Curve Estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2021). https://doi.org/10.1109/TPAMI.2021.3063604

[25] Ruoteng Li, Loong-Fah Cheong, and Robby T Tan. 2019. Heavy rain image restoration: Integrating physics model and conditional adversarial learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 1633–1642.

[26] Siyuan Li, Iago Breno Araujo, Wenqi Ren, Zhangyang Wang, Eric K Tokuda, Roberto Hirata Junior, Roberto Cesar-Junior, Jiawan Zhang, Xiaojie Guo, and Xiaochun Cao. 2019. Single image deraining: A comprehensive benchmark analysis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 3838–3847.

[27] Yu Li, Robby T Tan, Xiaojie Guo, Jiangbo Lu, and Michael S Brown. 2016. Rain streak removal using layer priors. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2736–2744.

[28] Ming Liu, Zhilu Zhang, Liya Hou, Wangmeng Zuo, and Lei Zhang. 2020. Deep adaptive inference networks for single image super-resolution. In *European Conference on Computer Vision*. Springer, 131–148.

[29] Risheng Liu, Long Ma, Jiaao Zhang, Xin Fan, and Zhongxuan Luo. 2021. Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 10561–10570.

[30] Xiaohong Liu, Yongrui Ma, Zhihao Shi, and Jun Chen. 2019. Griddehazenet: Attention-based multi-scale network for image dehazing. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 7314–7323.

[31] Rui Qian, Robby T Tan, Wenhan Yang, Jiajun Su, and Jiaying Liu. 2018. Attentive generative adversarial network for raindrop removal from a single image. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2482–2491.

[32] Xu Qin, Zhilin Wang, Yuanchao Bai, Xiaodong Xie, and Huizhu Jia. 2020. FFA-Net: Feature fusion attention network for single image dehazing. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 11908–11915.

[33] Dongwei Ren, Wangmeng Zuo, Qinghua Hu, Pengfei Zhu, and Deyu Meng. 2019. Progressive image deraining networks: A better and simpler baseline. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 3937–3946.

[34] Wenqi Ren, Si Liu, Hua Zhang, Jinshan Pan, Xiaochun Cao, and Ming-Hsuan Yang. 2016. Single image dehazing via multi-scale convolutional neural networks. In *European conference on computer vision*. Springer, 154–169.

[35] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*. Springer, 234–241.

[36] Shao-Hua Sun, Shang-Pu Fan, and Yu-Chiang Frank Wang. 2014. Exploiting image structural similarity for single image rain removal. In *2014 IEEE International Conference on Image Processing (ICIP)*. IEEE, 4482–4486.

[37] Robby T Tan. 2008. Visibility in bad weather from a single image. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 1–8.

[38] Hong Wang, Qi Xie, Qian Zhao, and Deyu Meng. 2020. A model-driven deep neural network for single image rain removal. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 3103–3112.

[39] Tianyu Wang, Xin Yang, Ke Xu, Shaozhe Chen, Qiang Zhang, and Rynson WH Lau. 2019. Spatial attentive single-image deraining with a high quality real rain dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 12270–12279.

[40] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. 2018. Deep retinex decomposition for low-light enhancement. In *British Machine Vision Conference*.

[41] Haiyan Wu, Yanyun Qu, Shaohui Lin, Jian Zhou, Ruizhi Qiao, Zhizhong Zhang, Yuan Xie, and Lizhuang Ma. 2021. Contrastive Learning for Compact Single Image Dehazing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 10551–10560.

[42] Wenhan Yang, Robby T Tan, Jiashi Feng, Jiaying Liu, Zongming Guo, and Shuicheng Yan. 2017. Deep joint rain detection and removal from a single image. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1357–1366.

[43] Wenhan Yang, Shiqi Wang, Yuming Fang, Yue Wang, and Jiaying Liu. 2020. From fidelity to perceptual quality: A semi-supervised approach for low-light image enhancement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 3063–3072.

[44] Wenhan Yang, Ye Yuan, Wenqi Ren, Jiaying Liu, Walter J Scheirer, Zhangyang Wang, Taiheng Zhang, Qiaoyong Zhong, Di Xie, Shiliang Pu, et al. 2020. Advancing image understanding in poor visibility environments: A collective benchmark study. *IEEE Transactions on Image Processing* 29 (2020), 5737–5752.

[45] Rajeev Yasarla and Vishal M Patel. 2019. Uncertainty guided multi-scale residual learning-using a cycle spinning cnn for single image de-raining. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 8405–8414.

[46] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. 2020. Learning enriched features for real image restoration and enhancement. In *European Conference on Computer Vision*. Springer, 492–511.

[47] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. 2021. Multi-stage progressive image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 14821–14831.

[48] He Zhang and Vishal M Patel. 2018. Densely connected pyramid dehazing network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3194–3203.

[49] He Zhang and Vishal M Patel. 2018. Density-aware single image de-raining using a multi-stream dense network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 695–704.

[50] He Zhang, Vishwanath Sindagi, and Vishal M Patel. 2019. Image de-raining using a conditional generative adversarial network. *IEEE transactions on circuits and systems for video technology* 30, 11 (2019), 3943–3956.

[51] Yonghua Zhang, Xiaojie Guo, Jiayi Ma, Wei Liu, and Jiawan Zhang. 2021. Beyond brightening low-light images. *International Journal of Computer Vision* 129, 4 (2021), 1013–1037.

[52] Yonghua Zhang, Jiawan Zhang, and Xiaojie Guo. 2019. Kindling the darkness: A practical low-light image enhancer. In *Proceedings of the 27th ACM international conference on multimedia*. 1632–1640.