Decoupling-and-Aggregating for Image Exposure Correction

Yang Wang[1], LongPeng[1*], Liang Li[2], Yang Cao[1†], Zheng-Jun Zha[1]
[1]University of Science and Technology of China, Hefei, China
[2] Key Lab of Intell. Info. Process., Inst. of Comput. Tech., CAS, Beijing, China
{ywang120, forrest, zhazj}@ustc.edu.cn, longp2001@mail.ustc.edu.cn, liang.li@ict.ac.cn

Abstract

The images captured under improper exposure conditions often suffer from contrast degradation and detail distortion. Contrast degradation will destroy the statistical properties of low-frequency components, while detail distortion will disturb the structural properties of high-frequency components, leading to the low-frequency and high-frequency components being mixed and inseparable. This will limit the statistical and structural modeling capacity for exposure correction. To address this issue, this paper proposes to decouple the contrast enhancement and detail restoration within each convolution process. It is based on the observation that, in the local regions covered by convolution kernels, the feature response of low-/high-frequency can be decoupled by addition/difference operation. To this end, we inject the addition/difference operation into the convolution process and devise a Contrast Aware (CA) unit and a Detail Aware (DA) unit to facilitate the statistical and structural regularities modeling. The proposed CA and DA can be plugged into existing CNN-based exposure correction networks to substitute the Traditional Convolution (TConv) to improve the performance. Furthermore, to maintain the computational costs of the network without changing, we aggregate two units into a single TConv kernel using structural re-parameterization. Evaluations of nine methods and five benchmark datasets demonstrate that our proposed method can comprehensively improve the performance of existing methods without introducing extra computational costs compared with the original networks. The codes will be publicly available.

1. Introduction

Images captured under improper exposure conditions often suffer from under-exposure or over-exposure problems [2, 14, 16]. Improper exposure will change the statistical distribution of image brightness, resulting in contrast degradation. Besides, improper exposure will also destroy the image's structural property and result in detail distortion. The contrast degradation and detail distortion will cause the low-frequency and high-frequency components to mix and inseparable across the image, making the image exposure correction extremely challenging [2, 4, 9, 31, 33, 39].

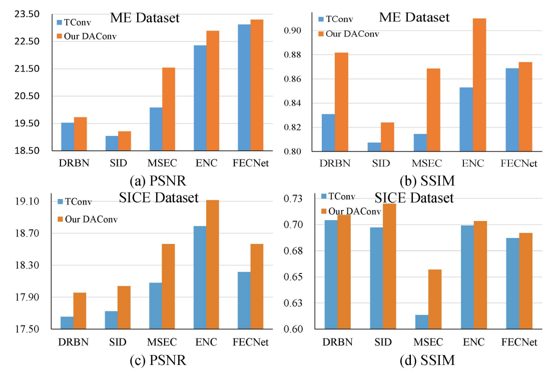


Figure 1. (a, b) The PSNR and SSIM comparison of TConv and our DAConv on the ME dataset. (c, d) The PSNR and SSIM comparison of TConv and our DAConv on the SICE dataset. Under the boosting of DAConv, the performance of existing methods has been comprehensively improved, reaching a new SOTA performance without introducing extra computational costs. Complete results can be found in Table 2.

In practice, one solution for this problem is to design an end-to-end architecture for learning contrast enhancement and detail restoration in shared feature space [14, 16]. However, the contrast-relevant features are primarily distributed in low-frequency components, while the detail-relevant features are primarily distributed in high-frequency components. Since low-frequency components are statistically dominant over high-frequency components, these methods mainly focus on contrast enhancement and cannot guarantee that the high-frequency details can be efficiently restored.

To achieve better contrast enhancement and detail restoration, some researchers propose to decompose and restore the input image's lightness and structure components, respectively [2, 20, 41]. For example, some researchers de-

*Co-first author. †Corresponding author.

compose images into illumination and reflectance components by utilizing Retinex theory and then design a specific network for each component [19, 20, 24]. Other researchers propose to decompose the input image into multi-scale components and adopt the coarse-to-fine strategy to progressively recover the lightness and fine-scale structures [2]. However, the decomposition operation inevitably destroys the relationship between brightness and detail and cannot balance the contrast and detail enhancement, leading to over-smooth problems or artifacts for enhanced results.

To address the above issues, this paper proposes to decouple the contrast enhancement and detail restoration during the convolution process. This method is based on statistical observations that the feature response in local regions can be decomposed into low-frequency components and high-frequency components by a difference operation. Based on this, we introduce a novel Contrast Aware (CA) unit in parallel with a Detail Aware (DA) unit to guide the contrast and detail modeling, termed Decoupling-and-Aggregating Convolution (DAConv). Different from TConv, we inject the addition/difference operation into the convolution process, which can guide the contrast and detail modeling in an explicit manner. Furthermore, to balance the contrast enhancement and detail restoration, we introduce a dynamic coefficient for each branch to adjust the amplitude of the feature response. Our proposed DAConv can be used as a general unit to substitute the TConv kernel in existing CNN-based exposure correction networks to facilitate contrast enhancement and detail restoration.

To reduce the computational costs, the CA, DA, and dynamic coefficients are aggregated into a single TConv kernel by structural re-parameterization in the inference phase. The aggregation is conducted before the activation function, and the linear superposition can reduce computational costs without changing the function of DAConv. After that, the performance of networks can be significantly improved without introducing extra computational costs compared with the original network. Evaluations of nine methods and five benchmark datasets demonstrate the effectiveness of our proposed method, as shown in Fig. 1.

The contribution can be summarized as follows:

(1) We propose a novel decoupling-and-aggregating scheme for image exposure correction, in which two parallel convolution processes are decoupled for contrast enhancement and detail restoration, respectively, and then aggregated into a single branch without additional computation compared with the original convolution scheme.

(2) To facilitate the contrast and detail relevant features extraction, a novel CA and DA unit are devised by injecting the addition and difference operation into the convolution process. Compared with traditional convolution kernels, our proposed CA and DA can explicitly model the contrast and detail relevant properties.

(3) Evaluations on the five prevailing benchmark datasets and nine SOTA image exposure correction methods demonstrate our proposed DC can comprehensively improve the contrast enhancement and detail restoration performances without introducing extra computational costs.

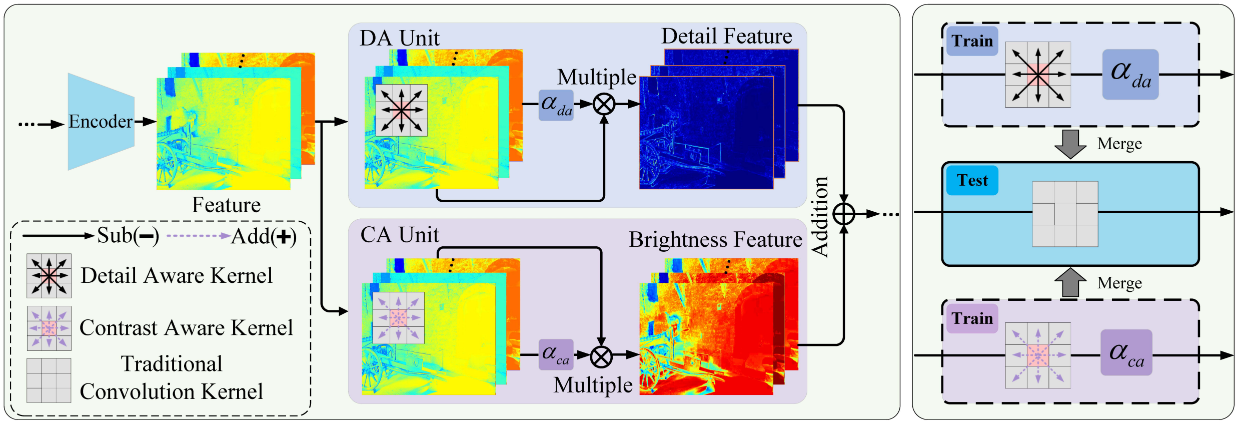
2. Related Work

Image Exposure Correction. Exposure correction is a hot research topic and has been studied for a long time in computational imaging [14, 16, 18, 21, 33, 37, 38, 40], which can be divided into traditional methods and learning-based methods. Traditional methods usually use Retinex theory and image histogram to enhance the contrast and detail [1,10,12,13,17,19,25,32,35]. However, suffering from the limitation of model capacity, it is difficult for traditional methods to deal with complex real-world conditions [6].

Learning-based methods can automatically learn the complex mapping function from datasets and have better performance in contrast enhancement and detail restoration [2,5,6,14,15,23,29]. Existing methods tend to decompose the image (*e.g.,* laplacian pyramid, frequency transformation) into different frequency components through preprocessing and enhance the different components one by one. Afifi *et al.* [2] propose a pyramid structure network to enhance image brightness and details in a coarse-to-fine manner. Huang *et al.* [15] propose a deep Fourier-based exposure correction network for image lightness and structure component reconstruction. CMEC [23] and ENC [14] learn to improve the contrast by learning exposure-invariant space. However, these preprocessing steps will disrupt the interrelationship between low-frequency and high-frequency, leading to the imbalance of the enhancement amplitudes of different components, leading to over-smooth or artifacts in enhanced results.

Structural Re-parameterization. Structural re-parameterization [7, 8] is a methodology of equivalently converting model structures via transforming the parameters. A widely used method is to design multiple convolutions parallel modules during training and merge them during inference. Different from the above structural re-parameterization, our DAConv can explicitly extract statistical and structural feature properties, respectively.

Pixel Difference Operation: Inspired by LBP, Yu *et al.* [36] propose central difference operation to improve the robustness of the face anti-spoofing network in the variable lighting environment. Su *et al.* [28] take the pixel relationship on different positions into consideration and proposes several different differential modes for extracting object edge information. Different from the above works, we propose a novel decoupling-and-aggregating scheme to facilitate the statistical and structural properties modeling for image exposure correction without introducing extra computational costs.



(a) Decoupling (b) Aggregating

Figure 2. (a) In the training phase, each TConv is substituted by a DAConv. (b) After training, the DA, CA, α_c and α_d are aggregated into a single TConv again.

3. Decoupling-and-Aggregating Convolution

The under-/over-exposure images suffer from both contrast degradation and detail distortion. The contrast degradation will change the statistical distribution of low-frequency components, while the detail distortion will disturb the structural properties of high-frequency components. Based on this frequency characteristic, some researchers propose decomposing under-/over-exposure images into a series of components and then performing contrast enhancement and detail restoration, respectively. However, the decomposition operation in existing methods will inevitably destroy the coupling relationship between contrast enhancement and detail restoration, resulting in over-smoothing or artifact problems in enhanced results. To better balance the relative relationship between contrast enhancement and detail restoration during the exposure correction, we propose a novel exposure correction method based on the decoupling-and-aggregating convolution, which contains two stages: the decoupling in the training phase and the aggregation in the testing phase.

3.1. Decoupling

Unlike existing methods of designing multiple subnetworks [2, 14], we dive into the convolution process within the network and decouple the convolution process into two parallel branches for statistical modeling and structural modeling, as shown in Fig. 2 (a). Our decoupling operation mainly uses the local smoothness assumption, which is widely used in image processing [22, 27] and is mathematically formulated as:

$$x(p_i) \approx x(p_j), \quad p_i,\ p_j \in R_n. \tag{1}$$

where, $x(p_i)$ and $x(p_j)$ represent the pixel intensity on location of p_i and p_j of local patch R_n and $x_{p_i}, x_{p_j} \in [0, 1]$. We conduct the following statistical experiment on improper exposure images to verify the local smoothness assumption. We randomly sample 10,000 image patches with size 3×3 from the ME dataset [2]. For each patch, we randomly select 5 pairs of pixels from different positions and calculate the average of intensity absolute difference for each pair of pixels as follows:

$$P_m = \frac{1}{10000} \left| x\left(p_i\right) - x\left(p_j\right) \right| \quad m = 1, 2, \cdots, 5. \tag{2}$$

The values of P_1 to P_5 are 0.00777, 0.00778, 0.00778, 0.00779, and 0.00780, respectively. We can infer that the pixel intensity at different positions is very close, which verifies the local smoothness assumption. Based on the above statistical experiment, we choose the central pixel within the local patch as the reference pixel, denoted as p_c, and the intensity of pixels in other positions can be expressed as the sum of the central pixel intensity and a bias n_i:

$$x(p_i) = x(p_c) + n_i. \tag{3}$$

where n_i changes from pixel to pixel, which is also known as the high-frequency components. Based on Eq. 3, the convolution process of TConv can be expressed as:

$$\begin{aligned} y(p_c) &= \sum_{p_i \in R} w(p_i) \cdot x(p_i) \\ &= \sum_{p_i \in R} w(p_i) \cdot \left(x(p_c) + n_i\right) \\ &= \underbrace{\sum_{p_i \in R} w(p_i) \cdot x(p_c)}_{\text{low-frequency response}} + \underbrace{\sum_{p_i \in R} w(p_i) \cdot n_i}_{\text{high-frequency response}} \end{aligned} \tag{4}$$

From Eq. 4, we can observe that the low-frequency response and high-frequency response are mixed in the traditional convolution feature response. To separate the high-frequency response from the above convolution response, we introduce a central-surrounding difference operation as:

$$\begin{aligned} y_h(p_c) &= \sum_{p_i \in R} w(p_i) \cdot (x(p_i) - x(p_c)) \\ &= \sum_{p_i \in R} w(p_i) \cdot n_i, \end{aligned} \tag{5}$$

With the central-surrounding difference operation, the low-frequency response can be significantly suppressed. We denote the convolution kernel injected with difference operation as Detail Aware (DA) kernel. After obtaining the high-frequency response, an intuitive option to get the low-frequency response is to subtract the high-frequency response from the $y(p_c)$ in Eq. 4. However, we empirically found that the direct subtraction operation would significantly drop the performance. The reason for this is that the obtained feature response will also contain enormous noise, especially in under-exposure conditions.

To this end, we propose to suppress the high-frequency response by increasing the proportion of low-frequency response, which is achieved by injecting the central-surrounding addition operation into the convolution process:

$$y_l(p_c) = \sum_{p_i \in R} w(p_i) \cdot (x(p_i) + x(p_c)), \tag{6}$$

In Eq. 6, the pixel at each position within the receptive field is superimposed with the same intensity value as the central pixel. Mathematically, the above operation is equivalent to adding a low-frequency response to the original response. We denote the above kernel as Contrast Aware (CA) kernel. Next, we connect DA and CA in parallel to substitute the TConv in existing networks. However, the difference and addition operation in DA and CA may result in the amplitude imbalance of high-frequency and low-frequency responses. To compensate for this, we introduce a dynamic adjustment coefficient on each branch to adjust the amplitude of feature response, as shown in Fig. 2 (a). Mathematically, it can be represented as:

$$\begin{aligned} y(p_c) = s(\alpha_{ca}) \cdot &\left(\sum_{p_i \in R} w_{ca}(p_i) \cdot (x(p_i) + x(p_c)) \right) \\ + s(\alpha_{da}) \cdot &\left(\sum_{p_i \in R} w_{da}(p_i) \cdot (x(p_i) - x(p_c)) \right). \end{aligned} \tag{7}$$

where s is the sigmoid activation function that is used to constrain the distribution of adjustment coefficients from 0 to 1. With continuous training, adjustment coefficients are constantly updated to balance the response magnitudes of CA and DA.

3.2. Aggregating

Under the boosting of DA unit and CA unit, the modeling capability and performance of networks can be significantly improved. However, this parallel structure will increase the model's complexity and parameters, resulting in low efficiency. In order to reduce computational costs, we introduce structural re-parameterization to merge CA and DA in parallel into a TConv kernel during inference, as shown in Fig. 2 (b). During training, we replace $\mathrm{k} \times \mathrm{k}$ TConv with $\mathrm{k} \times \mathrm{k}$ DAConv. After training, we perform an equivalent replacement to fuse $\mathrm{k} \times \mathrm{k}$ DAConv into a TConv kernel to maintain the computational costs of network without changing, as shown in the following formula [7, 8]:

Firstly, we can expand Eq. 7 as follows:

$$\begin{aligned} y(p_c) = &\underbrace{\left(\sum_{p_i \in R} (s(\alpha_{ca}) \cdot w_{ca}(p_i)) \cdot x(p_i) \right)}_{\mathrm{k \times k\ conv}} \\ &+ \underbrace{\left(\sum_{p_i \in R} (s(\alpha_{da}) \cdot w_{da}(p_i)) \cdot x(p_i) \right)}_{\mathrm{k \times k\ conv}} \\ &\underbrace{+ x(p_c) \cdot \left(\sum_{p_i \in R} s(\alpha_{ca}) \cdot w_{ca}(p_i) - \sum_{p_i \in R} s(\alpha_{da}) \cdot w_{da}(p_i) \right)}_{\mathrm{item1}}, \end{aligned} \tag{8}$$

Secondly, the weight accumulation operation in $\mathrm{item1}$ can be mathematically equivalent to a 1×1 convolution. Then, we expand 1×1 convolution to $\mathrm{k} \times \mathrm{k}$ convolution:

$$\begin{aligned} y(p_c) = &\underbrace{\left(\sum_{p_i \in R} (s(\alpha_{ca}) \cdot w_{ca}(p_i)) \cdot x(p_i) \right)}_{\mathrm{k \times k\ conv}} \\ &+ \underbrace{\left(\sum_{p_i \in R} (s(\alpha_{da}) \cdot w_{da}(p_i)) \cdot x(p_i) \right)}_{\mathrm{k \times k\ conv}} \\ &+ \underbrace{\left(\sum_{p_i \in R} w_c(p_i) \cdot x(p_i) \right)}_{\mathrm{k \times k\ conv}}, \end{aligned} \tag{9}$$

where $w_c\,(p_i)$ is defined by the following formula:

$$w_c\,(p_i) = \begin{cases} s(\alpha_{ca}) \cdot \mathrm{sum}\,(w_{ca}) - s(\alpha_{da}) \cdot \mathrm{sum}\,(w_{da}) & if\ p_i = p_c \\ 0 & if\ p_i \neq p_c \end{cases}, \tag{10}$$

Finally, we fuse all parallel $\mathrm{k} \times \mathrm{k}$ kernels into a single $\mathrm{k} \times \mathrm{k}$ kernel w_{all} by the linearity of convolution [8]:

$$\begin{aligned} y(p_c) &= \sum_{p_i \in R} (s(\alpha_{ca}) \cdot w_{ca}(p_i) + s(\alpha_{da}) \cdot w_{da}(p_i) + w_c(p_i)) \cdot x(p_i) \\ &= \sum_{p_i \in R} w_{all}(p_i) \cdot x(p_i). \end{aligned} \tag{11}$$

Table 1. Summary of multi-exposure correction and low-light image enhancement datasets.

	Multi-exposure correction		Low-light image enhancement		
	ME dataset [2]	SICE dataset [4]	LOLV1 [30]	LOL-v2-Real [34]	LOL-v2-Synthetic [34]
Train samples	17,675	3,988	485	689	900
Test samples	5,905	812	15	100	100

Table 2. Quantitative results for nine image exposure correction methods on ME dataset and SICE dataset. The **bold** represents performance of DAConv-based methods. The bold represents our cost-free improvement compared to the baselines. The bold represents a slight degradation after using DAConv. Note that DAConv-based methods are marked with as **bold***.

	ME dataset [2]				SICE dataset [4]			
	Under-exposure		Over-exposure		Under-exposure		Over-exposure	
	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑
RUAS [20]	13.430	0.681	6.390	0.466	7.507	0.246	5.806	0.089
RUAS*	**14.867**+1.437	**0.708**+0.027	**6.940**+0.550	**0.486**+0.020	**8.528**+1.021	**0.356**+0.110	**5.938**+0.132	**0.137**+0.048
Zero-DCE [11]	14.550	0.589	10.400	0.514	15.972	0.653	9.078	0.590
Zero-DCE*	**15.067**+0.517	**0.771**+0.182	**10.847**+0.447	**0.710**+0.196	**16.229**+0.257	**0.656**+0.003	**9.315**+0.237	**0.595**+0.005
RetinexNet [30]	12.130	0.621	10.470	0.595	15.239	0.613	16.863	0.638
RetinexNet*	**12.208**+0.078	**0.607**-0.014	**18.576**+8.106	**0.794**+0.199	**15.637**+0.398	**0.642**+0.029	**17.009**+0.146	**0.645**+0.007
UNet [26]	18.437	0.821	17.440	0.809	16.036	0.650	17.209	0.664
UNet*	**18.524**+0.087	**0.831**+0.010	**17.953**+0.513	**0.822**+0.013	**16.521**+0.485	**0.678**+0.028	**17.239**+0.030	**0.684**+0.020
DRBN [34]	19.740	0.829	19.370	0.832	17.249	0.707	18.275	0.700
DRBN*	**20.630**+0.890	**0.888**+0.059	**19.100**-0.270	**0.878**+0.046	**17.337**+0.088	**0.709**+0.002	**18.896**+0.621	**0.780**+0.080
SID [5]	19.370	0.810	18.830	0.806	17.065	0.692	18.728	0.706
SID*	**19.484**+0.114	**0.829**+0.019	**19.015**+0.185	**0.820**+0.014	**17.539**+0.474	**0.724**+0.032	**18.796**+0.068	**0.714**+0.008
MSEC [2]	20.520	0.813	19.790	0.816	18.291	0.606	17.755	0.626
MSEC*	**21.530**+1.010	**0.859**+0.046	**21.550**+1.760	**0.875**+0.059	**18.949**+0.658	**0.655**+0.049	**17.979**+0.224	**0.660**+0.034
ENC [14]	22.720	0.854	22.110	0.852	18.665	0.696	18.974	0.703
ENC*	**23.320**+0.600	**0.909**+0.055	**22.600**+0.490	**0.909**+0.057	**19.072**+0.407	**0.701**+0.005	**19.176**+0.202	**0.707**+0.004
FECNet [15]	22.960	0.860	23.220	0.875	18.012	0.685	18.496	0.691
FECNet*	**23.150**+0.190	**0.865**+0.005	**23.410**+0.190	**0.880**+0.005	**18.347**+0.335	**0.691**+0.006	**18.893**+0.397	**0.698**+0.007

Table 3. Ablation study for DAConv. The best performance is marked in **bold**.

	Under-exposure		Over-exposure	
	PSNR↑	SSIM↑	PSNR↑	SSIM↑
baseline	20.520	0.812	19.790	0.815
DA+CA	18.592	0.782	18.055	0.751
CA+DA	18.950	0.787	18.433	0.759
TConv//TConv	20.720	0.831	20.453	0.822
DA//TConv-DA	20.640	0.821	20.023	0.817
TConv//DA	21.010	0.842	21.243	0.838
TConv//CA	21.034	0.849	21.260	0.849
w/o α DA//CA	21.256	0.851	21.379	0.867
w/ α & DA//CA	**21.530**	**0.859**	**21.550**	**0.875**

4. Experiments and Analysis

4.1. Settings

Datasets. We evaluate our DAConv on five prevailing benchmarks for multi-exposure correction and low-light image enhancement: ME dataset [2], SICE dataset [4], LOLV1 [30], LOL-v2-Real [34] and LOL-v2-Synthetic [34]. The details of each dataset are summarized in Table 1. Different from ENC [14] and SICE [4], they only select a part of the exposure levels for evaluation. We take a further step and use all of the exposure levels for evaluation to verify the algorithm's performance under more practical multi-exposure conditions. We randomly select 489 scenes as the training set, and the rest of the 100 scenes are used as the test set, containing 3,988 and 812 paired images, respectively. The ME dataset contains five exposure levels for each scene, and we also use all exposure levels for training and evaluation. Following [14], we use Expert Cin [3] as ground truth. Referring to [14], we define the images at exposure level of 1-2 as under-exposure images, the rest as over-exposure images. For the SICE dataset, we define the average brightness on the Y channel of YCbCr space lower than that of ground truth as under-exposure images, and the rest of the images are used as over-exposure images.

Baselines. In order to evaluate the superiority and generality of DAConv on image exposure correction, nine public methods are selected for evaluation, including: RUAS [20], Zero-DCE [11], RetinexNet [30], U-Net [26], DRBN [34], SID [5], MSEC [2], ENC [14] and FECNet [15]. For low-light image enhancement, six baselines are selected for evaluation, including: Zero-DCE [11], U-Net [26], DRBN [34], SID [5], MSEC [2] and ENC [14].

Implementation Details. In all experiments, we keep the training setting (*e.g.,* loss function, batch size, training epoch, and active function) the same as the original setting, except that the TConv is replaced by DAConv.

Table 4. Quantitative comparison on LOLV1, LOL-V2-R, and LOL-V2-S datasets. The bold represents our cost-free improvement compared to the existing method. The performance of baseline can be comprehensively improved after using DConv.

	LOLV1 [30]		LOL-V2-R [34]		LOL-V2-S [34]	
	PSNR↑	SSIM↑	PSNR↑	SSIM↑	PSNR↑	SSIM↑
ZeroDCE [11]	15.296	0.518	12.382	0.448	16.954	0.810
ZeroDCE*	**16.206**+0.910	**0.522**+0.004	**13.445**+1.063	**0.460**+0.012	**17.372**+0.418	**0.820**+0.01
UNet [26]	17.480	0.753	18.449	0.668	18.131	0.843
UNet*	**17.671**+0.191	**0.764**+0.011	**18.533**+0.084	**0.718**+0.050	**20.079**+1.948	**0.878**+0.035
DRBN [34]	19.068	0.790	19.421	0.729	21.012	0.895
DRBN*	**19.190**+0.122	**0.812**+0.022	**19.855**+0.434	**0.747**+0.018	**21.100**+0.088	**0.899**+0.004
SID [5]	18.577	0.789	18.640	0.703	20.801	0.884
SID*	**19.260**+0.683	**0.812**+0.023	**18.892**+0.252	**0.713**+0.01	**22.267**+1.456	**0.910**+0.026
MSEC [2]	18.845	0.679	19.031	0.662	19.582	0.705
MSEC*	**20.895**+2.050	**0.748**+0.069	**20.192**+1.161	**0.670**+0.008	**20.745**+1.163	**0.813**+0.108
ENC [14]	22.310	0.837	21.004	0.802	21.608	0.887
ENC*	**22.856**+0.546	**0.843**+0.006	**21.764**+0.760	**0.839**+0.037	**22.337**+0.729	**0.902**+0.015

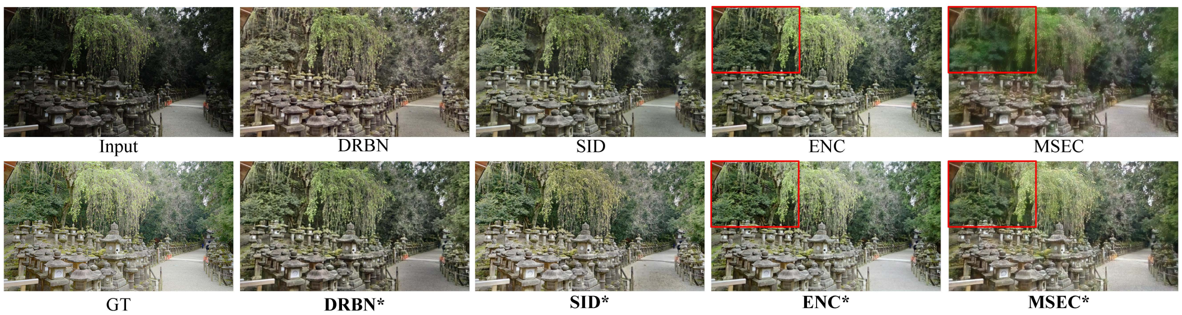


Figure 3. Visualization comparison on SICE dataset. DAConv-based methods have better results in image contrast and details.

Figure 4. Qualitative comparison on LOLV1. Compared with the TConv-based method, the DAConv-based method is closer to the ground truth in image contrast and image details.

4.2. Ablation study

To demonstrate the effectiveness of DAConv, we compare DAConv with the following settings: (a) DA and CA in serial; (b) CA and DA in serial; (c) TConv and TConv in parallel; (d) DA and TConv-DA in parallel. The "TConv-DA" represents the response of TConv subtract the response of DA; (e) TConv and DA in parallel; (f) TConv and CA in parallel; (g) DAConv without α. The setting from (a) to (g) is denoted as DA+CA, CA+DA, TConv//TConv, DA//TConv-DA, TConv//DA, TConv//CA, and w/o α DA//CA, respectively. We choose MSEC [2] as the baseline and replace each TConv within the network with the above settings. The experiments are conducted on the ME dataset, and the results are reported in Table 3.

We can observe that the performance of the serial connection is much lower than the parallel connection. The reason for this is that difference operation will lose the low-frequency components, resulting in the next layer cannot obtain sufficient information for correction. The performance of DA//(TConv-DA) is lower than that of w/o α DA//CA, which demonstrates that subtracting the response of DA from the response of TConv cannot obtain accurate statistical features. The performance of TConv//TConv is higher than the baseline. It is because the number of parameters has been doubled and representation capabilities have been improved. However, the performance of TConv//TConv is lower than TConv//DA and TConv//CA. The reason for this is that the DA and CA can explicitly guide the detail and contrast modeling. Thus, with the combination of DA and CA in parallel (*i.e.* w/o α & DA//CA), the exposure correction performance can be further improved. To balance the contrast enhancement and detail restoration, we further introduce a dynamic adjustment coefficient for each branch (*i.e.* w/ α & DA//CA), which achieves the best performance among all settings.

Figure 5. Qualitative comparison on ME dataset.

4.3. Quantitative results

In Table 2, we report the PSNR/SSIM performance of nine exposure correction methods on the ME dataset and SICE dataset. We can observe that the performance of most of these methods is comprehensively improved after utilizing the DAConv, demonstrating that our proposed DC is robust and can be embedded in various network architectures. It is worth mentioning that unsupervised algorithms, such as Zero-DCE [11] and RUAS [20], are also improved after using DAConv, indicating that DAConv is not sensitive to network learning methods. For MSEC [2], which focuses on detail restoration, DAConv still improves its ability to perceive detail and contrast features. Even for SOTA algorithms such as MSEC [2], FECNet [15], and ENC [14], using DAConv can still improve network performance and achieve new SOTA performance. To better demonstrate the performance in practical multi-exposure conditions, like [14], we calculate the average performance on all under-exposure and over-exposure images, as shown in Fig. 1. We can observe after using DAConv, the performance of each method gains comprehensive improvements.

In Table 4, we report the PSNR/SSIM performance of six methods on public LOLV1, LOL-V2-Real, and LOL-V2-Synthetic datasets. Particularly, LOLV1 and LOL-V2-Real datasets are captured in real dark environments, losing many image details, as shown in Fig. 4. Compared with TConv, DAConv can better perceive image details and texture, significantly improving network performance. For example, after using DAConv, the PSNR/SSIM score of MSEC on LOLV1 dataset is improved from 18.845/0.679 to 20.895/0.748. Furthermore, all these performance gains are cost-free without introducing extra computational costs. Thus, the DAConv can be used as a general computing unit to incorporate with various networks to improve low-light image enhancement performance.

4.4. Qualitative results

Fig. 3 shows the visual comparison before and after using DAConv on under-exposure images of the SICE dataset. Due to space limitations, we only show the exposure correction results of several methods. More results are provided in the supplementary material. We can observe that the TConv-based method suffers from blurred details and color distortion, especially in the red box in Fig. 3, while the DAConv-based method is better at detail restoration and contrast enhancement. For real dark scenes where a lot of image details have been lost, our DAConv can still improve the image details while improving the contrast of the image, as shown in Fig. 4.

Fig. 5 represents the visualization results on the over-exposure images of the ME dataset. It can be seen that the details of the building area in the over-exposure image background have been seriously damaged. The network based on TConv separates the processes of brightness

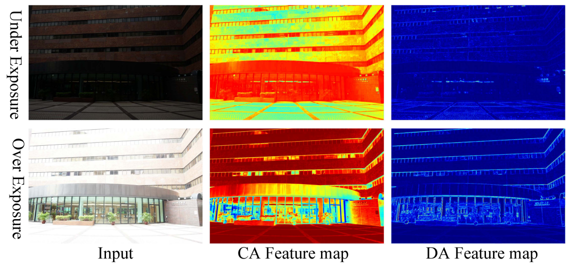
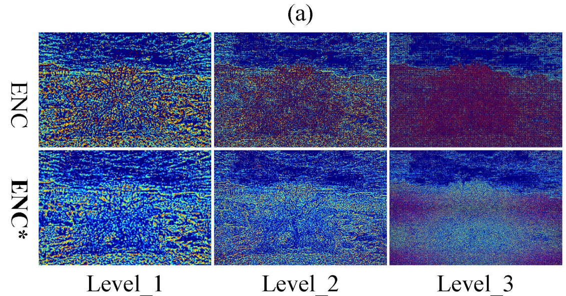


Figure 6. Feature visualization of CA and DA, which can capture image brightness distribution and image details, respectively

	Level_1	Level_2	Level_3	Image Level
Input	40.24	40.83	39.42	14.07
ENC	43.71	42.97	39.72	21.84
ENC*	**44.47+0.76**	**44.53+1.56**	**42.55+2.83**	**22.78+0.94**

(a)



(b)

Figure 7. The PSNR comparison of different high-frequency layers and the error map between enhanced high-frequency layers and corresponding GT.

enhancement and detail restoration, destroys the inner relationship between them, and leads to over-smoothness in these areas. However, the algorithm based on DAConv uses the decoupling-and-aggregating mechanism at each convolution, which can make full use of the mutual relationship between them to achieve a balance while performing contrast enhancement and detail restoration.

4.5. Performance analysis

Feature Visualization. To verify that DAConv can capture the image details and contrast-relevant information explicitly, we select over-exposure and under-exposure images in the same scene and visualize the feature maps of CA and DA in DAConv, respectively, as shown in the Fig. 6. For under-exposure images, CA can perceive the brightness distribution and pays more attention to dark areas, while DA focuses on extracting structural features. For over-exposure images, CA pays more attention to the overexposed area.

Detail Error Map. In order to verify the superiority of DAConv in detail restoration, we take ENC as the baseline and conduct the following experiment. Firstly, we randomly select 100 pairs of under-exposure and over-exposure image enhancement results from ENC and **ENC***. Secondly, following [2], we decompose the image into detail layers via the Laplace Image Pyramid, denoted as Level_1, Level_2, and Level_3. Finally, we calculate the average PSNR score of different layers, as shown in Fig 7 (a). We can observe that DAConv can significantly outperform TConv in detail restoration, especially for tiny textures (*i.e.* Level_3). We further visualize the error map between each layer with the corresponding GT, as shown in Fig.7 (b). We can observe that the error of DAConv is much lower than that of TConv, which demonstrates that DAConv can improve the detail restoration capability of existing networks.

Table 5. Running time comparison before and after aggregation as well as the original network. We calculate the average running time of 1000 images with 1024×1024 resolution on RTX3080.

	w/o aggregation	w/ aggregation	original
Running Time (s)↓	0.320	0.144	0.144

Running Times. We choose ENC as the baseline and compare the average inference time of 1000 images of 1024×1024 resolution before and after aggregating, as shown in Table 5. After aggregating, DA and CA can be merged into a convolution kernel, which keeps the inference time the same as the original network.

5. Conclusion and Limitation

Conclusion. This paper proposes a novel Decoupling-and-Aggregating Convolution (DAConv) for image exposure correction that can explicitly guide contrast and detail modeling. The DAConv can be used to substitute TConv in existing CNN-based exposure correction networks. Extensive experiments on under-exposure and over-exposure datasets verify the effectiveness of DAConv in contrast enhancement and detail restoration. It can significantly improve the performance of existing methods while maintaining the same computational costs as the original networks.

Limitation. In this paper, we propose to use DAConv to replace the TConv in the existing network to improve image exposure correction performance. However, the combination of DAConv with existing TConv-based networks is may not the best choice. In the future, we will design a new framework to fully exploit the capability of DAConv for image exposure correction.

6. Acknowledgments

This work was supported by the National Key R&D Program of China under Grant 2020AAA0105702, the National Natural Science Foundation of China (NSFC) under Grants 62206262, 62225207 and U19B2038, the University Synergy Innovation Program of Anhui Province under Grant GXXT-2019-025.

References

[1] Mohammad Abdullah-Al-Wadud, Md Hasanul Kabir, M Ali Akber Dewan, and Oksam Chae. A dynamic histogram equalization for image contrast enhancement. *IEEE Transactions on Consumer Electronics*, 53(2):593–600, 2007. 2

[2] Mahmoud Afifi, Konstantinos G Derpanis, Bjorn Ommer, and Michael S Brown. Learning multi-scale photo exposure correction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9157–9167, 2021. 1, 2, 3, 5, 6, 7, 8

[3] Vladimir Bychkovsky, Sylvain Paris, Eric Chan, and Frédo Durand. Learning photographic global tonal adjustment with a database of input/output image pairs. In *CVPR 2011*, pages 97–104. IEEE, 2011. 5

[4] Jianrui Cai, Shuhang Gu, and Lei Zhang. Learning a deep single image contrast enhancer from multi-exposure images. *IEEE Transactions on Image Processing*, 27(4):2049–2062, 2018. 1, 5

[5] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3291–3300, 2018. 2, 5, 6

[6] Yu-Sheng Chen, Yu-Ching Wang, Man-Hsin Kao, and Yung-Yu Chuang. Deep photo enhancer: Unpaired learning for image enhancement from photographs with gans. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6306–6314, 2018. 2

[7] Xiaohan Ding, Yuchen Guo, Guiguang Ding, and Jungong Han. Acnet: Strengthening the kernel skeletons for powerful cnn via asymmetric convolution blocks. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1911–1920, 2019. 2, 4

[8] Xiaohan Ding, Xiangyu Zhang, Ningning Ma, Jungong Han, Guiguang Ding, and Jian Sun. Repvgg: Making vgg-style convnets great again. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13733–13742, 2021. 2, 4

[9] Xingbo Dong, Wanyan Xu, Zhihui Miao, Lan Ma, Chao Zhang, Jiewen Yang, Zhe Jin, Andrew Beng Jin Teoh, and Jiajun Shen. Abandoning the bayer-filter to see in the dark. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17431–17440, 2022. 1

[10] Xueyang Fu, Yinghao Liao, Delu Zeng, Yue Huang, Xiao-Ping Zhang, and Xinghao Ding. A probabilistic method for image enhancement with simultaneous illumination and reflectance estimation. *IEEE Transactions on Image Processing*, 24(12):4965–4977, 2015. 2

[11] Chunle Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, and Runmin Cong. Zero-reference deep curve estimation for low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1780–1789, 2020. 5, 6, 7

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

[13] Shijie Hao, Xu Han, Yanrong Guo, Xin Xu, and Meng Wang. Low-light image enhancement with semi-decoupled decomposition. *IEEE transactions on multimedia*, 22(12):3025–3038, 2020. 2

[14] Jie Huang, Yajing Liu, Xueyang Fu, Man Zhou, Yang Wang, Feng Zhao, and Zhiwei Xiong. Exposure normalization and compensation for multiple-exposure correction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6043–6052, 2022. 1, 2, 3, 5, 6, 7

[15] 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 *Proceedings of the European conference on computer vision (ECCV)*, 2022. 2, 5, 7

[16] Jie Huang, Man Zhou, Yajing Liu, Mingde Yao, Feng Zhao, and Zhiwei Xiong. Exposure-consistency representation learning for exposure correction. In *Proceedings of the 30th ACM International Conference on Multimedia*, pages 6309–6317, 2022. 1, 2

[17] Edwin H Land. The retinex theory of color vision. *Scientific american*, 237(6):108–129, 1977. 2

[18] Chongyi Li, Chunle Guo, Ling-Hao Han, Jun Jiang, Ming-Ming Cheng, Jinwei Gu, and Chen Change Loy. Low-light image and video enhancement using deep learning: A survey. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (01):1–1, 2021. 2

[19] Mading Li, Jiaying Liu, Wenhan Yang, Xiaoyan Sun, and Zongming Guo. Structure-revealing low-light image enhancement via robust retinex model. *IEEE Transactions on Image Processing*, 27(6):2828–2841, 2018. 2

[20] Risheng Liu, Long Ma, Jiaao Zhang, Xin Fan, and Zhongxuan Luo. 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*, pages 10561–10570, 2021. 1, 2, 5, 7

[21] Long Ma, Tengyu Ma, Risheng Liu, Xin Fan, and Zhongxuan Luo. Toward fast, flexible, and robust low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5637–5646, 2022. 2

[22] Jean-Michel Morel, Antoni Buades, and Tomeu Coll. Local smoothing neighborhood filters. In *Handbook of Mathematical Methods in Imaging*. 2015. 3

[23] Ntumba Elie Nsampi, Zhongyun Hu, and Qing Wang. Learning exposure correction via consistency modeling. BMVC, 2021. 2

[24] Xutong Ren, Wenhan Yang, Wen-Huang Cheng, and Jiaying Liu. Lr3m: Robust low-light enhancement via low-rank regularized retinex model. *IEEE Transactions on Image Processing*, 29:5862–5876, 2020. 2

[25] Ali M Reza. Realization of the contrast limited adaptive histogram equalization (clahe) for real-time image enhancement. *Journal of VLSI signal processing systems for signal, image and video technology*, 38(1):35–44, 2004. 2

[26] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 5, 6

[27] Leonid I Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena*, 60(1-4):259–268, 1992. 3

[28] Zhuo Su, Wenzhe Liu, Zitong Yu, Dewen Hu, Qing Liao, Qi Tian, Matti Pietikäinen, and Li Liu. Pixel difference networks for efficient edge detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5117–5127, 2021. 2

[29] Yang Wang, Yang Cao, Zheng-Jun Zha, Jing Zhang, Zhiwei Xiong, Wei Zhang, and Feng Wu. Progressive retinex: Mutually reinforced illumination-noise perception network for low-light image enhancement. In *Proceedings of the 27th ACM international conference on multimedia*, pages 2015–2023, 2019. 2

[30] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560*, 2018. 5, 6

[31] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5901–5910, 2022. 1

[32] Ke Xu, Xin Yang, Baocai Yin, and Rynson WH Lau. Learning to restore low-light images via decomposition-and-enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2281–2290, 2020. 2

[33] Xiaogang Xu, Ruixing Wang, Chi-Wing Fu, and Jiaya Jia. Snr-aware low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17714–17724, 2022. 1, 2

[34] Wenhan Yang, Shiqi Wang, Yuming Fang, Yue Wang, and Jiaying Liu. 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*, pages 3063–3072, 2020. 5, 6

[35] Zhenqiang Ying, Ge Li, Yurui Ren, Ronggang Wang, and Wenmin Wang. A new image contrast enhancement algorithm using exposure fusion framework. In *International conference on computer analysis of images and patterns*, pages 36–46. Springer, 2017. 2

[36] Zitong Yu, Chenxu Zhao, Zezheng Wang, Yunxiao Qin, Zhuo Su, Xiaobai Li, Feng Zhou, and Guoying Zhao. Searching central difference convolutional networks for face anti-spoofing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5295–5305, 2020. 2

[37] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Learning enriched features for real image restoration and enhancement. In *European Conference on Computer Vision*, pages 492–511. Springer, 2020. 2

[38] Yonghua Zhang, Jiawan Zhang, and Xiaojie Guo. Kindling the darkness: A practical low-light image enhancer. In *Proceedings of the 27th ACM international conference on multimedia*, pages 1632–1640, 2019. 2

[39] Zhao Zhang, Huan Zheng, Richang Hong, Mingliang Xu, Shuicheng Yan, and Meng Wang. Deep color consistent network for low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1899–1908, 2022. 1

[40] Lin Zhao, Shao-Ping Lu, Tao Chen, Zhenglu Yang, and Ariel Shamir. Deep symmetric network for underexposed image enhancement with recurrent attentional learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12075–12084, 2021. 2

[41] Minfeng Zhu, Pingbo Pan, Wei Chen, and Yi Yang. Eemefn: Low-light image enhancement via edge-enhanced multi-exposure fusion network. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 13106–13113, 2020. 1