Decoupling-and-Aggregating for Image Exposure Correction

Yang Wang¹, LongPeng¹*, Liang Li², Yang Cao¹†, Zheng-Jun Zha¹
¹University of Science and Technology of China, Hefei, China
²Key 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. Complete results can be found in Table 2.

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].

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-
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.
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 sub-networks [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:

$$x(p_i) \approx x(p_j), \quad p_i, p_j \in R_n.$$  (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 $3 \times 3$ 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:

$$P_m = \frac{1}{10000} |x(p_i) - x(p_j)|, \quad m = 1, 2, \ldots, 5.$$  (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 $n_i$:

$$x(p_i) = x(p_c) + n_i.$$  (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:

$$y(p_c) = \sum_{p_i \in R} w(p_i) \cdot x(p_i)$$
$$= \sum_{p_i \in R} w(p_i) \cdot (x(p_c) + n_i)$$
$$= \sum_{p_i \in R} w(p_i) \cdot x(p_c) + \sum_{p_i \in R} w(p_i) \cdot n_i$$

$$= \sum_{p_i \in R} w(p_i) \cdot x(p_c) + \sum_{p_i \in R} w(p_i) \cdot n_i$$  (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:

\[
y_{h}(p_c) = \sum_{p_i \in R} w(p_i) \cdot (x(p_i) - x(p_c)) = \sum_{p_i \in R} w(p_i) \cdot n_i, \tag{5}
\]

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:

\[
y_r(p_c) = \sum_{p_i \in R} w(p_i) \cdot (x(p_i) + x(p_c)), \tag{6}
\]

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:

\[
y(p_c) = s(\alpha_{ca}) \cdot \left( \sum_{p_i \in R} w_{ca}(p_i) \cdot (x(p_i) + x(p_c)) \right) + s(\alpha_{da}) \cdot \left( \sum_{p_i \in R} w_{da}(p_i) \cdot (x(p_i) - x(p_c)) \right). \tag{7}
\]

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 \(k \times k\) TConv with \(k \times k\) DAConv. After training, we perform an equivalent replacement to fuse \(k \times k\) 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:

\[
y(p_c) = \left( \sum_{p_i \in R} (s(\alpha_{ca}) \cdot w_{ca}(p_i)) \cdot x(p_i) \right)_{\text{k \times k conv}} + \left( \sum_{p_i \in R} (s(\alpha_{da}) \cdot w_{da}(p_i)) \cdot x(p_i) \right)_{\text{k \times k conv}} + x(p_c) \cdot \left( \sum_{p_i \in R} s(\alpha_{ca}) \cdot w_{ca}(p_i) - \sum_{p_i \in R} s(\alpha_{da}) \cdot w_{da}(p_i) \right)_{\text{item1}}, \tag{8}
\]

Secondly, the weight accumulation operation in item1 can be mathematically equivalent to a \(1 \times 1\) convolution. Then, we expand \(1 \times 1\) convolution to \(k \times k\) convolution:

\[
y(p_c) = \left( \sum_{p_i \in R} (s(\alpha_{ca}) \cdot w_{ca}(p_i)) \cdot x(p_i) \right)_{\text{k \times k conv}} + \left( \sum_{p_i \in R} (s(\alpha_{da}) \cdot w_{da}(p_i)) \cdot x(p_i) \right)_{\text{k \times k conv}} + \left( \sum_{p_i \in R} w_{c}(p_i) \cdot x(p_i) \right)_{\text{k \times k conv}}, \tag{9}
\]

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

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

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

\[
y(p_c) = \sum_{p_i \in R} (s(\alpha_{ca}) \cdot w_{ca}(p_i) + s(\alpha_{da}) \cdot w_{da}(p_i) + w_{c}(p_i)) \cdot x(p_i) = \sum_{p_i \in R} w_{all}(p_i) \cdot x(p_i). \tag{11}
\]
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 as over-exposure images. The SICE dataset contains 100 scenes, and the rest of the 100 scenes are used as the test set, containing 3,988 and 812 paired images, respectively. The SICE dataset contains five exposure levels for each scene, and we also use all exposure levels for training and evaluation.

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], SICE dataset [4], 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], SICE dataset [4], ENC [14] and MSEC [2].

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.

<table>
<thead>
<tr>
<th></th>
<th>LOLV1 [30]</th>
<th>LOL-V2-R [34]</th>
<th>LOL-V2-S [34]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PSNR↑</td>
<td>SSIM↑</td>
<td>PSNR↑</td>
</tr>
<tr>
<td>ZeroDCE [11]</td>
<td>15.296</td>
<td>0.518</td>
<td>12.382</td>
</tr>
<tr>
<td>ZeroDCE*</td>
<td>16.206+0.910</td>
<td>0.522</td>
<td>0.004</td>
</tr>
<tr>
<td>UNet [26]</td>
<td>17.480</td>
<td>0.753</td>
<td>18.449</td>
</tr>
<tr>
<td>UNet*</td>
<td>17.671+0.191</td>
<td>0.764</td>
<td>0.011</td>
</tr>
<tr>
<td>DRBN [34]</td>
<td>19.068</td>
<td>0.790</td>
<td>19.421</td>
</tr>
<tr>
<td>DRBN*</td>
<td>19.190+0.122</td>
<td>0.818</td>
<td>0.022</td>
</tr>
<tr>
<td>SID [5]</td>
<td>18.577</td>
<td>0.789</td>
<td>18.640</td>
</tr>
<tr>
<td>SID*</td>
<td>19.260+0.683</td>
<td>0.812</td>
<td>0.023</td>
</tr>
<tr>
<td>MSE [2]</td>
<td>18.845</td>
<td>0.679</td>
<td>19.031</td>
</tr>
<tr>
<td>MSE*</td>
<td>20.095+0.050</td>
<td>0.748</td>
<td>0.069</td>
</tr>
<tr>
<td>ENC [14]</td>
<td>22.310</td>
<td>0.837</td>
<td>21.004</td>
</tr>
<tr>
<td>ENC*</td>
<td>22.856+0.546</td>
<td>0.843</td>
<td>0.006</td>
</tr>
</tbody>
</table>

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 MSE [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.
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.

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.

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


