Overview

The supplementary material contains:

- Conceptual Differences between Proposed Method and Existing Methods
- Loss Functions
- More Discussion about Ablation Study
- More Qualitative Results

1 Conceptual Differences between Proposed Method and Existing Methods

The conceptual differences between the networks of the existing methods and the proposed method are provided in Table S 1. The proposed method incorporates all the essential blocks required for effective aerial image dehazing.

2 Loss Functions

We train the proposed network using $L_1$ loss ($L_1$), edge loss ($L_{\text{Edge}}$) and perceptual loss ($L_p$). The detailed equations of these loss functions are:

2.1 $L_1$ Loss

$L_1$ loss aims at reduction of the per pixel difference between the output image ($I_O$) and ground truth ($I_{GT}$), which can be formulated as:

$$L_1 = \|I_O - I_{GT}\|_1 \quad (1)$$

2.2 Edge Loss

Edge loss ($L_{\text{Edge}}$) aims at improving the edge details by calculating the discrepancies between edges of the output image and ground truth image. It can be represented as:

$$L_{\text{Edge}} = \|\psi(I_O) - \psi(I_{GT})\|_1 \quad (2)$$

where, $\psi(\cdot)$ is Sobel operator.

2.3 Perceptual Loss

The perceptual loss ($L_p$) aims at improving the perceptual quality of the output by calculating the differences between output and ground truth at various feature levels. The pretrained layers of VGG-16 [1] model are used as feature space for the loss calculation. The perceptual loss can be represented mathematically as:

$$L_P = \|\Phi_{i\in(3,8,15)}(I_O) - \Phi_{i\in(3,8,15)}(I_{GT})\|_1 \quad (3)$$

where, $\Phi_{i}(\cdot)$ is $i^{th}$ layer of VGG-16 model.

3 More Discussion about Ablation Study

The description regarding which blocks and loss functions are used in each network setting in ablation study is provided in Table S 2.
Table S 1: Conceptual comparison of the proposed method with existing state-of-the-art methods.

<table>
<thead>
<tr>
<th>Method</th>
<th>Publication</th>
<th>Local Dependency</th>
<th>Global Dependency</th>
<th>Geometric Adaptability</th>
<th>Edge Boosting Skip Connections</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPA-GAN [7]</td>
<td>-</td>
<td>✓</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>Huang et al. [9]</td>
<td>WACV-20</td>
<td>✓</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>SkyGAN [10]</td>
<td>WACV-21</td>
<td>✓</td>
<td>×</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>UFormer [8]</td>
<td>CVPR-22</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>×</td>
</tr>
</tbody>
</table>

Proposed Method - ✓ ✓ ✓ ✓ ✓

Table S 2: Details regarding modules and losses used in each network setting (DW-MSA: DepthWise convolution based Multi-head Self Attention).

<table>
<thead>
<tr>
<th>Setting</th>
<th>Attention Type</th>
<th>Offset Extraction Type</th>
<th>Skip Connection Type</th>
<th>Loss Functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Network 1</td>
<td>Vanilla Attention [11]</td>
<td>Spatially Attentive</td>
<td>Edge Boosting Skip Connection</td>
<td>$L_1 + L_{Edge} + L_P$</td>
</tr>
<tr>
<td>Network 2</td>
<td>DW-MSA [8]</td>
<td>Spatially Attentive</td>
<td>Edge Boosting Skip Connection</td>
<td>$L_1 + L_{Edge} + L_P$</td>
</tr>
<tr>
<td>Network 3</td>
<td>Deformable MSA</td>
<td>Deformable [12]</td>
<td>Edge Boosting Skip Connection</td>
<td>$L_1 + L_{Edge} + L_P$</td>
</tr>
<tr>
<td>Network 4</td>
<td>Deformable MSA</td>
<td>Modulated Deformable</td>
<td>Edge Boosting Skip Connection</td>
<td>$L_1 + L_{Edge} + L_P$</td>
</tr>
<tr>
<td>Network 5</td>
<td>Deformable MSA</td>
<td>Spatially Attentive</td>
<td>No Skip Connection</td>
<td>$L_1 + L_{Edge} + L_P$</td>
</tr>
<tr>
<td>Network 6</td>
<td>Deformable MSA</td>
<td>Spatially Attentive</td>
<td>Regular Skip Connection</td>
<td>$L_1 + L_{Edge} + L_P$</td>
</tr>
<tr>
<td>Network 7</td>
<td>Deformable MSA</td>
<td>Spatially Attentive</td>
<td>Edge Boosting Skip Connection</td>
<td>$L_1$</td>
</tr>
<tr>
<td>Network 8</td>
<td>Deformable MSA</td>
<td>Spatially Attentive</td>
<td>Edge Boosting Skip Connection</td>
<td>$L_1 + L_{Edge}$</td>
</tr>
<tr>
<td>Proposed</td>
<td>Deformable MSA</td>
<td>Spatially Attentive</td>
<td>Edge Boosting Skip Connection</td>
<td>$L_1 + L_{Edge} + L_P$</td>
</tr>
</tbody>
</table>

4 More Qualitative Results

We provide more qualitative results on RICE dataset for aerial haze removal in Figure S 1. We compare the qualitative results with state-of-the-art methods GCANet [2], USID [3], MSBDN [4], TSDNet [5], RDNet [6], SPA-GAN [7] and UFormer [8]. As seen from the results, the proposed method is able to maintain more spatial content and color balance.
Figure S 1: Qualitative results comparison of the proposed method with existing state-of-the-art methods on RICE dataset for aerial haze removal.
References


