Inharmonious Region Localization

Jing Liang, Li Niu, Liqing Zhang
Shanghai Jiao Tong University
1. Background—What is inharmonious region?
1. Background—What is inharmonious region?

- Image
- Ground Truth

(a) Color inconsistency
(b) Lighting inconsistency

- Image
- Ground Truth

(c) Color inconsistency
(d) Lighting inconsistency
1. Background—Definition & Challenges

1) Definition
The task of inharmonious region localization is aimed to localize the inharmonious region in a synthetic image.

2) Challenges
- An area should be compared with multi-scale nearby regions to determine whether it is an inharmonious region.
- Inharmonious factors like color, lighting inconsistency are hard to capture.
01 Background
02 Methods
03 Experiments
04 Discussion
2. Methods—DIRL

- ResNet Block
- Bi-directional Feature Integration
- MDA
- Mask-guided Dual Attention
- GGD
- Global-context Guided Decoder

Feature Flow → Adjacent Feature Flow → Global-context Feature Flow → Supervision

I

Res
256x256x64

Res
128x128x128

Res
64x64x256

Res
32x32x512

Res
16x16x512

r_1

r_2

r_3

r_4

r_5

BFI
256x256x32

BFI
128x128x32

BFI
64x64x32

BFI
32x32x32

BFI
16x16x32

b_1

b_2

b_3

b_4

b_5

MDA

MDA

MDA

MDA

MDA

a_1

a_2

a_3

a_4

a_5

GGD
256x256x32

GGD
128x128x32

GGD
64x64x32

GGD
32x32x32

L_{bce} + L_{ssim}

L_{aux}

L_{aux}

L_{aux}

L_{aux}

L_{aux}

I → M → \hat{M}

Encoding → Transition → Refinement → Decoding
2. Methods—DIRL

- **DIRL**
  - Concatenation
  - Element-wise Multiplication
  - Element-wise Addition

---

---

**Bi-directional Feature Integration**

**Dual Attention**

---

**Global-context Guided Decoder**

- Concatenation
- Element-wise Multiplication
- Element-wise Addition
2. Methods—DIRL

Loss functions:

\[ \mathcal{L}_{bce} = - \sum_{i,j} M_{i,j} \log(\hat{M}_{i,j}) - \sum_{i,j} (1 - M_{i,j}) \log(1 - \hat{M}_{i,j}) \]

\[ \mathcal{L}_{ssim} = 1 - \frac{(2\mu_x \mu_y + C_1)(2\sigma_{xy} + C_2)}{\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2} \]

\[ \mathcal{L}_{aux} = - \sum_k \sum_{i,j} M_{i,j} \log(A_{k,i,j}) \]

\[ - \sum_k \sum_{i,j} (1 - M_{i,j}) \log(1 - A_{k,i,j}) \]

\[ \mathcal{L}_{total} = \mathcal{L}_{bce} + \mathcal{L}_{ssim} + \lambda \mathcal{L}_{aux} \]
3. Experiments

We conduct all experiments on the iHarmony4 dataset.

(a) Examples of HCOCO.

(b) Examples of HAdobe5k.

(c) Examples of HFlickr.

(d) Examples of Hday2night.
### 3. Experiments—Quantitative Comparisons

<table>
<thead>
<tr>
<th>Method</th>
<th>AP(%)</th>
<th>$F_1$</th>
<th>IoU(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>UNet [25]</td>
<td>74.90</td>
<td>0.6717</td>
<td>64.74</td>
</tr>
<tr>
<td>DeepLabv3 [29]</td>
<td>75.69</td>
<td>0.6902</td>
<td>66.01</td>
</tr>
<tr>
<td>HRNet [30]</td>
<td>75.33</td>
<td>0.6765</td>
<td>65.49</td>
</tr>
<tr>
<td>MFCN [19]</td>
<td>45.63</td>
<td>0.3794</td>
<td>28.54</td>
</tr>
<tr>
<td>MantraNet [15]</td>
<td>64.22</td>
<td>0.5691</td>
<td>50.31</td>
</tr>
<tr>
<td>MAGritte [20]</td>
<td>71.16</td>
<td>0.6907</td>
<td>60.14</td>
</tr>
<tr>
<td>H-LSTM [16]</td>
<td>60.21</td>
<td>0.5239</td>
<td>47.07</td>
</tr>
<tr>
<td>S2AM [2]</td>
<td>43.77</td>
<td>0.3029</td>
<td>22.36</td>
</tr>
<tr>
<td>DIRL</td>
<td>80.02</td>
<td>0.7317</td>
<td>67.85</td>
</tr>
</tbody>
</table>
3. Experiments—Qualitative Comparisons

<table>
<thead>
<tr>
<th>Image</th>
<th>GT</th>
<th>Ours</th>
<th>DeepLabv3</th>
<th>HRNet</th>
<th>UNet</th>
<th>MAGritte</th>
</tr>
</thead>
</table>

[Image showing qualitative comparisons for various datasets and models]
3. Experiments—MDA Visualization
Thank You!