Aesthetic Critiques Generation for Photos
Kuang-Yu Chang, Kung-Hung Lu, and Chu-Song Chen
Institute of Information Science, Academia Sinica, Taipei, Taiwan

Problem & Contribution

- A picture is worth a thousand words. We extend the task to produce captions related to photo aesthetics and/or photography skills.
- We study a new problem, generating aesthetic critiques of photos. A worth noting issue of the photo aesthetics captioning is its multi-aspect nature.

Contribution:

1. From aesthetic quality (AQ) assessment to critiques
2. Photo aesthetics captioning dataset
3. Multi-aspect aesthetics captioning

Evaluation Criteria

- Traditional criteria (BLEU, METEOR & CIDEr) borrowed from machine translation community are unsuitable for image captioning task.
- Simple example on different criteria:
  (a) the ocean on the bottom of the image is too dark.
  (b) the tree on the foreground is too close to the right.
- SPICE: parses a sentence into a graph, and evaluates the similarity based on the parsed results between the generated and reference sentences.
- Diversity: takes the near-duplication sentences into consideration to establish an evaluation measure.

Table: Automatic evaluation of sentences (a) and (b).

<table>
<thead>
<tr>
<th>BLEU1</th>
<th>BLEU2</th>
<th>BLEU3</th>
<th>BLEU4</th>
<th>METEOR</th>
<th>CIDEr</th>
<th>SPICE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.452</td>
<td>0.196</td>
<td>0</td>
<td>0</td>
<td>0.156</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Our Framework

- **Aspect-oriented (AO)** approach: select an appealing aspect of the image and generate caption
- **Aspect-fusion (AF)** approach: aggregate 4 aspects using soft-attention mechanism

Experimental Results

- **Automatic evaluation**: SPICE and diversity are used for evaluation
- **Human evaluation**: rate the generations on a 3-point scale, Good, Common and Bad. 3 experts and 5 people are chosen for expert evaluation and Amazon Mechanical Turk (AMT), respectively.

Photo Critique Captioning Dataset (PCCD)

- PCCD contains 7 aspects rating from 1 to 10 per aspect for a photo, more than 60,000 captions and 4235 images.
- Codes and dataset are publicly available.

Generated Results

- Examples of the critiques generated by the three models, AF, AO, and CNN-LSTM-WD on PCCD. We apply the AF to a large-scale aesthetic quality assessment dataset, Aesthetic Visual Analysis (AVA) dataset.

Some failed-case captions generated by our approach.

- The clouds have very good detail and help maintain and or help the comp
- I like the way you have used the rule of thirds and this is a very good example of the colours
- The composition is good and is very close to the rule of thirds