

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.



Photo Critique Captioning: racing makes for interesting pictures because of the speed the movement the bright colors

Image Captioning: a man riding a motorcycle down a street

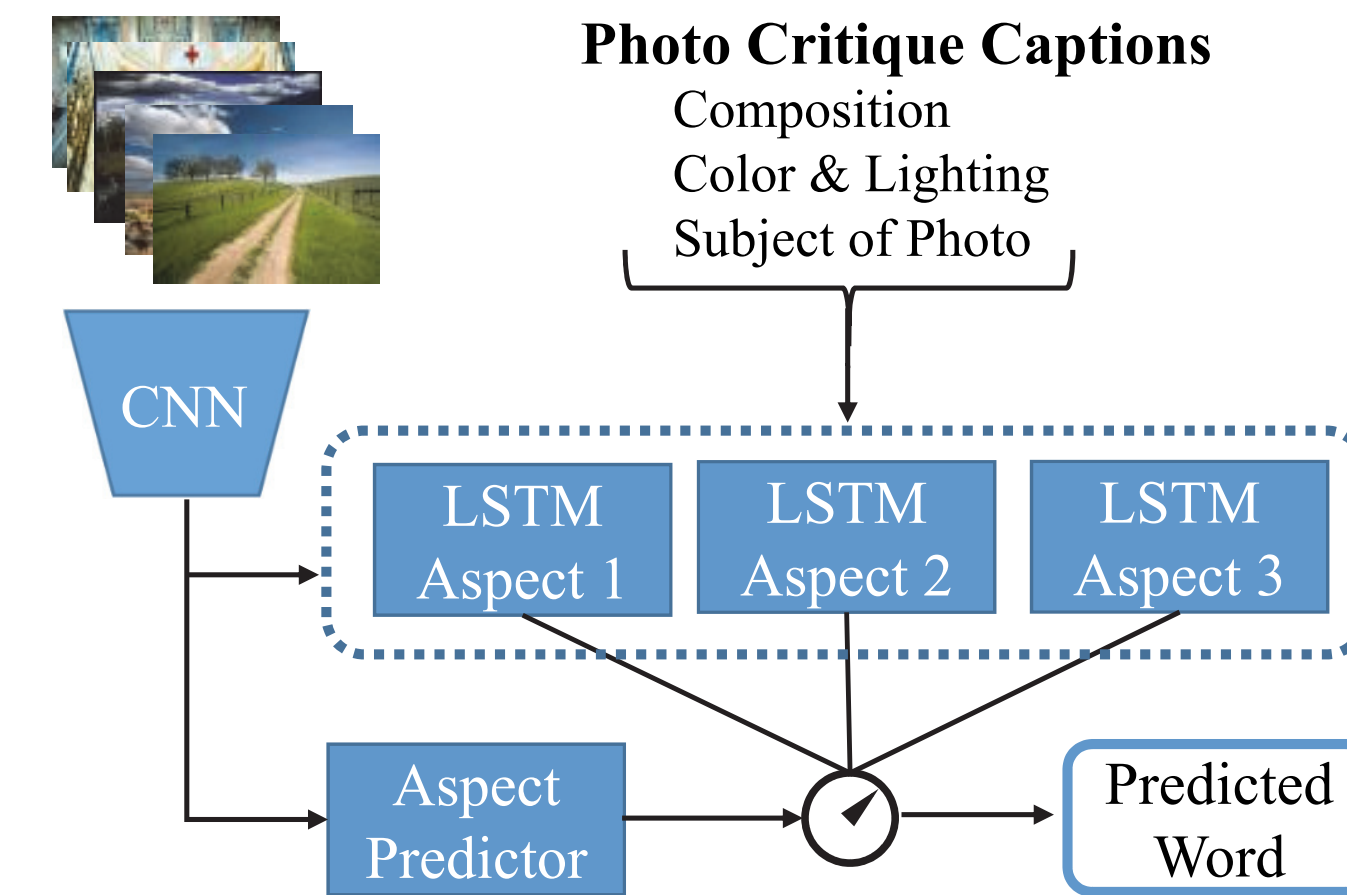


Color: the color is great i love the bright greens

Composition: i like the way you have used the rule of thirds to make the image more interesting composition and i would like to see more of the sky and the tree branches

Our Framework

- Aspect-oriented (AO)** approach: select an appealing aspect of the image and generate caption



- Aspect-fusion (AF)** approach: aggregate L aspects using soft-attention mechanism

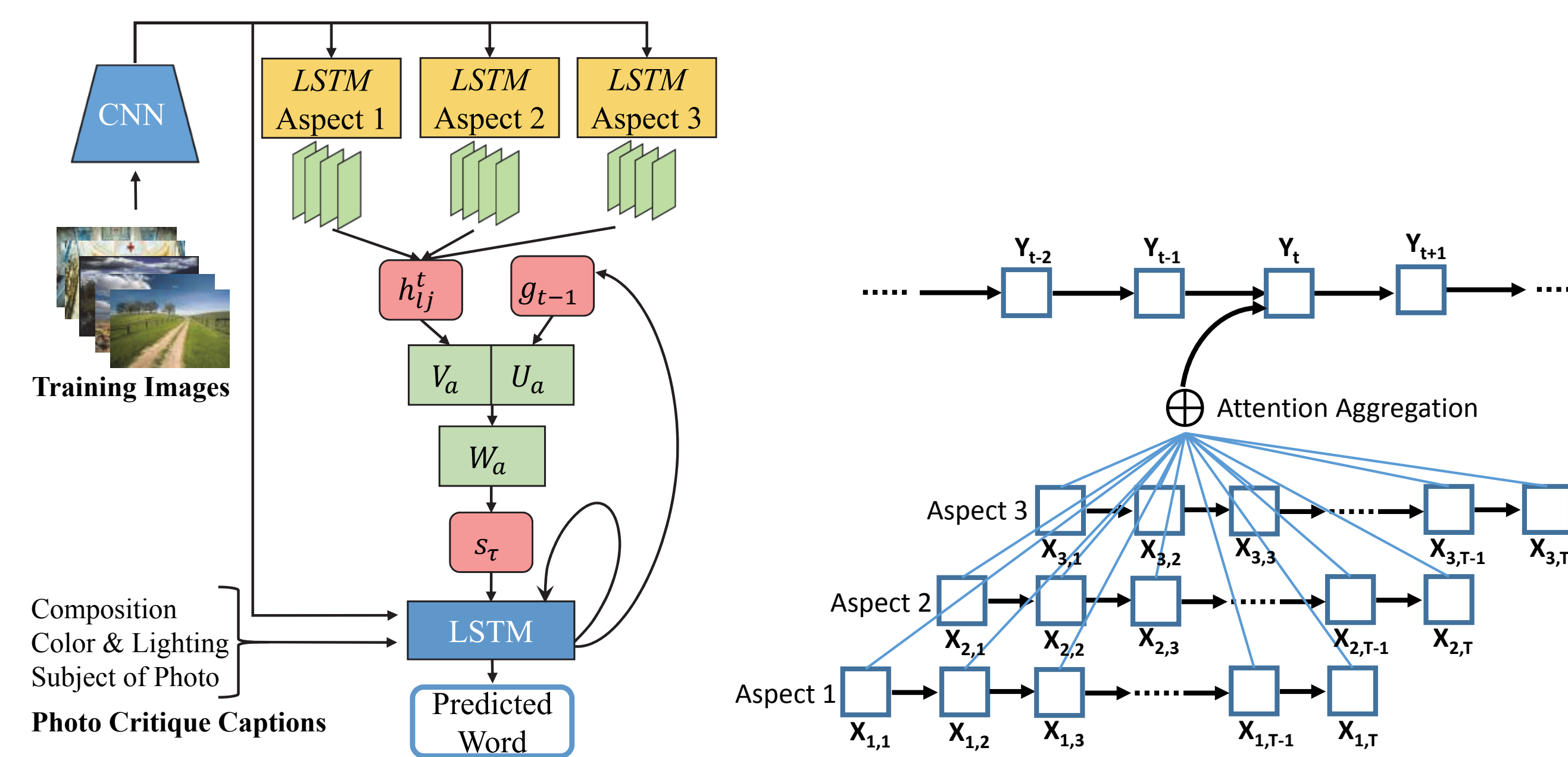



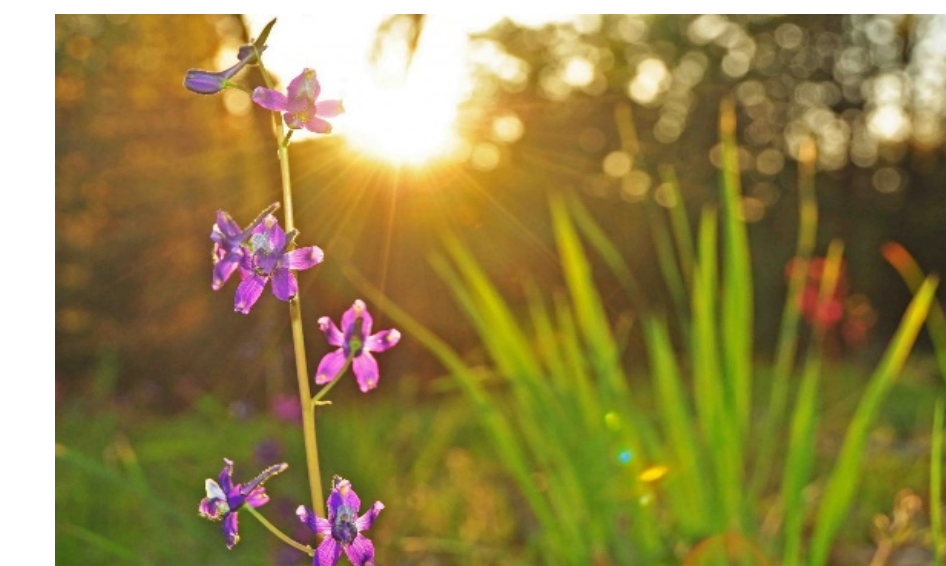
Photo Critique Captioning Dataset (PCCD)

- 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.

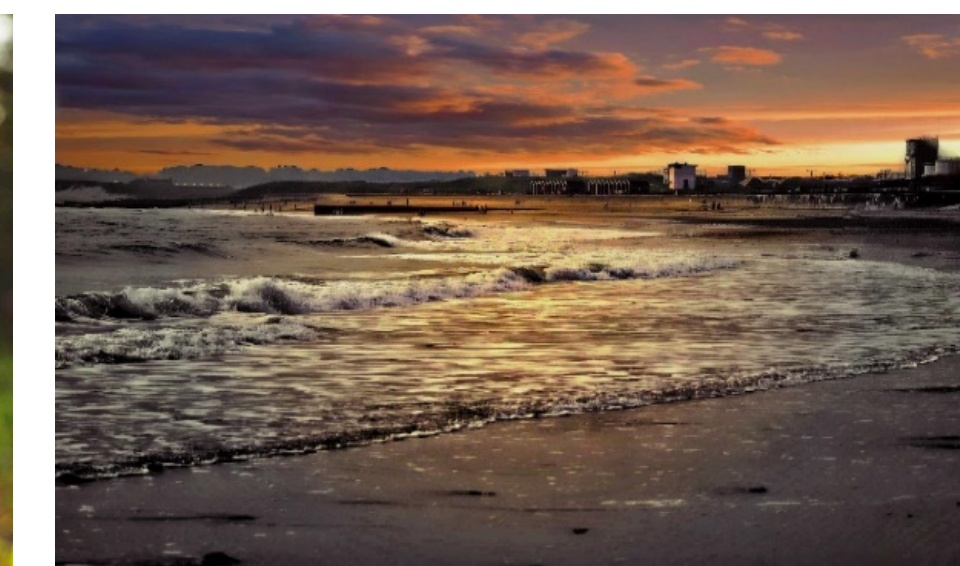
	General Impression	Great shot Nick the composition is outstanding! and the wave action adds , I can just hear the waves launching on the shore, leaves one with a calming effect. The clouds also add to complete the mood.	10/10
	Subject of Photo	Hello Nick, Great shot the composition is just a wonderful mix of center composition to diagonal lines and triangles, and density portions. It's so serene calming with the explosion of the sun during sun down. Like the ocean flow between the walls add the action to the image.	9/10
	Composition & Perspective	As mentioned above Nick, excellent composition, five factors happening all at once. Just lacks a little contrast more so density in the center.	10/10
	Use of Camera, Exposure & Speed	all settings observed are good and well used to achieve the effects including both nd filters. To gain some of the lost density in the center, a slightly longer exposure would have achieved this.	8/10
	Depth of Field	depth is great with f 16 and the foreground and ocean horizon along with the wooden walls on both sides, this really adds Nick to depth.	10/10
	Color & Lighting	color is great in terms of the time when the shot was taken, nice and subtle pastel sort of. again for me just the center with a longer exposure would have helped to draw out more color and density, and nick I am speaking of a little.	9/10
	Focus	Great focus by stopping down to f 16.	10/10

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.



AF: the color of the flower is fantastic and the tones of the purple are nice



AF: i like the way the reflection of the water in the foreground



AF: i like the way the river flows down the centre of the image and draws the viewers eye into the centre of the frame

AO: the color is great but i would have liked to have seen more of the image as a result of some of the foliage

AO: i like the subject of this photo

AO: i like the colors in the trees

CNN-LSTM: the composition is good but the composition is a bit too much

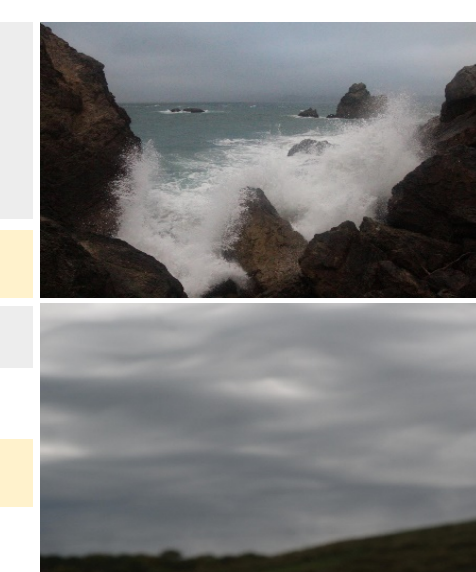
CNN-LSTM: i think the composition here is pretty good

CNN-LSTM: i think the composition is very good

- Some failed-case captions generated by our approach.



the UNK UNK in the foreground is also interesting to UNK the boat and elements are in balance with the way the trees on the left of the frame
Error: too many UNK
the tree is a bit dark
Error: no tree



the horizon is in the bottom of the frame
Error: in the bottom → in the top
i like the composition of this image
Error: aesthetic description error

- Cross-dataset results of the critiques generated by AF model on AVA.



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

Evaluation Criteria

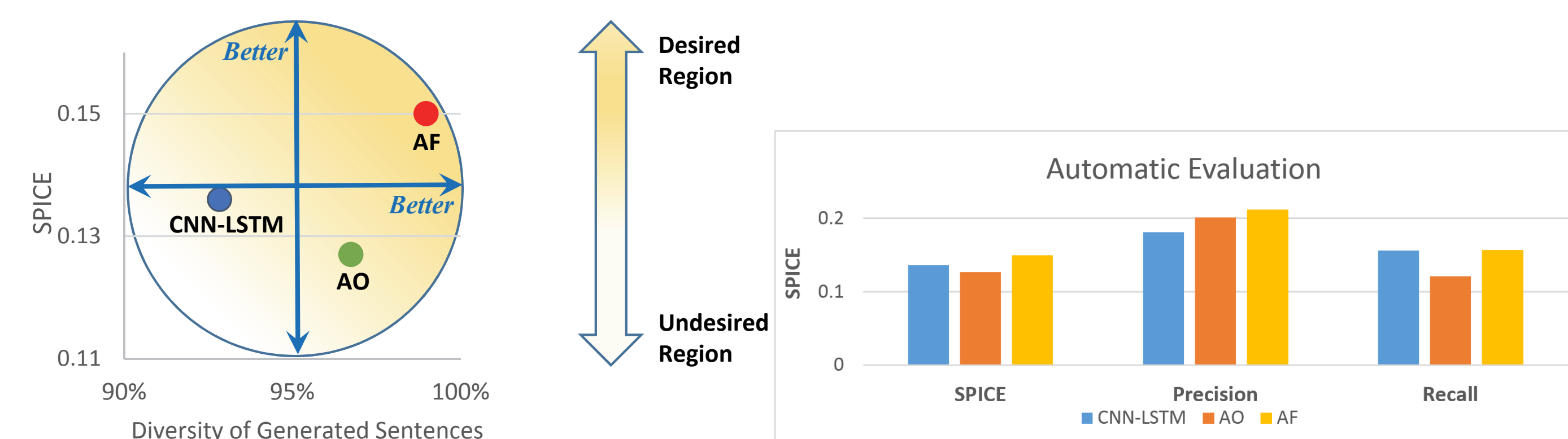
- Traditional criteria (BLEU, METEOR & CIDEr) borrowed from machine translation community are unsuitable for image captioning task.
- Simple example on different criteria
 - (a) the ocean on the bottom of the image is too dark.
 - (b) the tree on the foreground is too close to the right.
- SPICE:** parses 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).

BLEU1	BLEU2	BLEU3	BLEU4	METEOR	CIDEr	SPICE
0.462	0.196	0	0	0.156	0	0

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.

