VALSE 2019

April 11 – 14, 2019

Visual Sentiment Prediction based on Automatic Discovery of Affective Regions

Jufeng Yang[1], Dongyu She[1], Ming-Ming Cheng[1], P.L. Rosin[2], Liang Wang[3]

[1]Nankai University, [2]Cardiff University, [3]Chinese Academy of Sciences

Introduction

Two distinguishing characteristics of **Affective Regions** (ARs):

a) a salient region and probably contains one or more objects, which can attract people's attention, and

b) conveys significant sentiments

Algorithms

Algorithm 1 Visual Sentiment Analysis using Affective Regions

Input:

Input Image: I

The number of desired affective regions: K

Output:

Predicted sentiment label : $\vec{Y}$

1: Generate n bounding boxes with their objectness scores $B = \{b_i; Obj\_score_i^I\}_{i=1}^n$.

2: Apply candidate selection method to generate m candidate regions $H = \{h_i\}_{i=1}^m$.

3: Initialize the framework with pre-trained CNN.

4: Let $\vec{Y}_{Global}$ be the predictions of the whole image.

5: Pass H through the CNN model from the second layer to the last layer.

6: Let $y \in \mathbb{R}^{m \times c}$ be the sentiment probability of m proposal using the CNN model, compute the sentiment score in Eqn. (4)

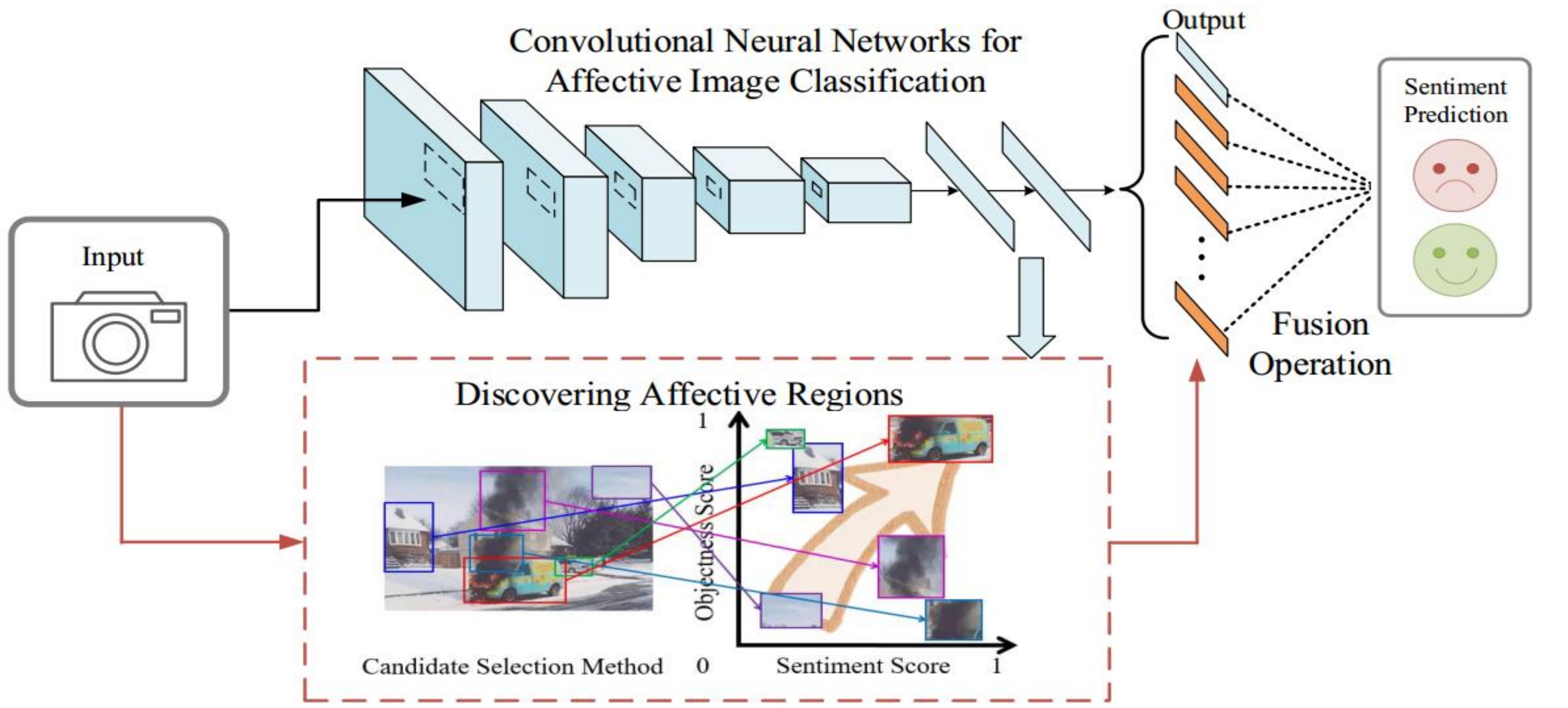
7: Compute the AR score for the each region in Eqn. (5).

8: Rank proposals with AR scores and select top K as affective regions.

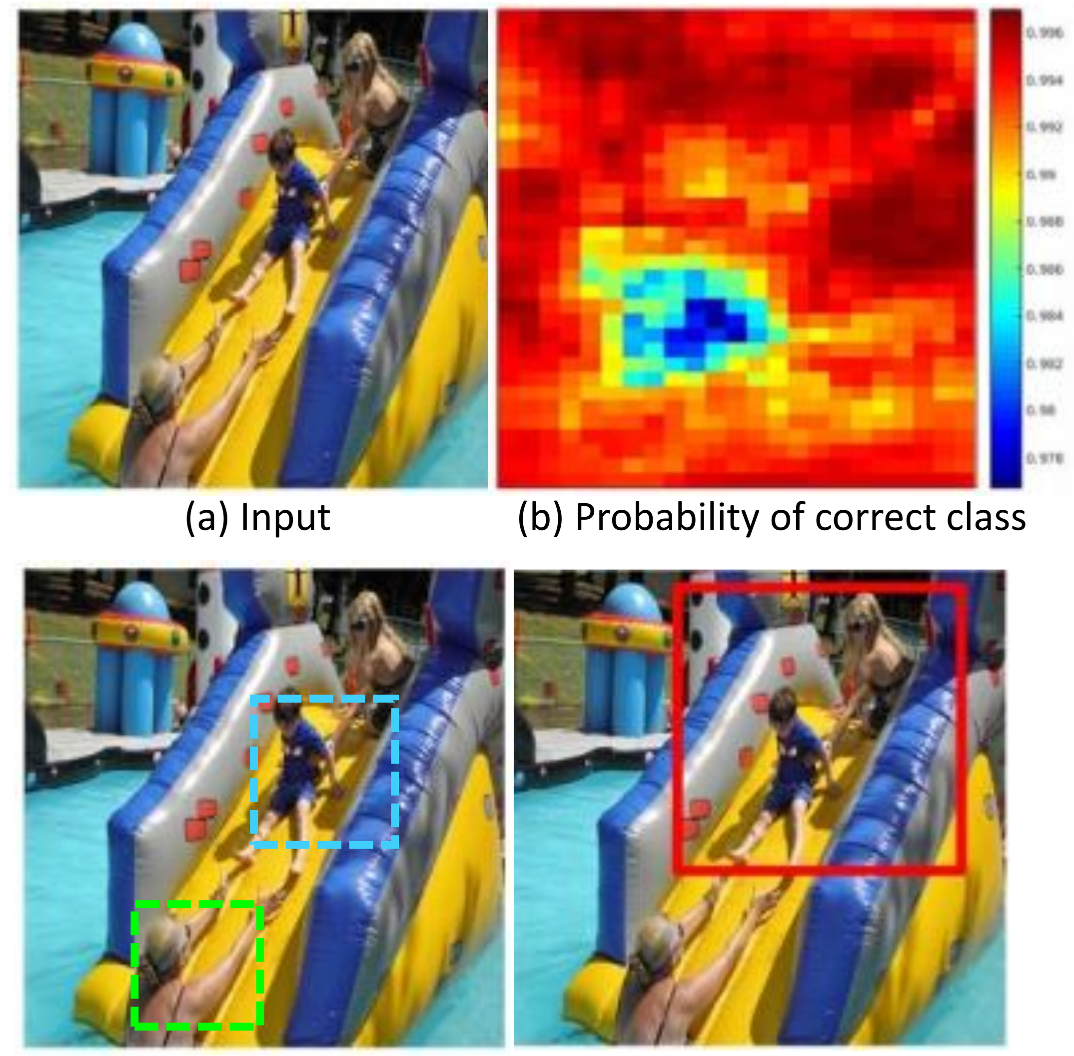
9: Predict the label $\vec{Y}$ using the cross-candidates pooling operation.

10: **return** $\vec{Y}$

Discovering Affective Regions



we aim to develop an algorithm to automatically discover ARs carrying significant sentiments and combine the standard holistic representation with a local representation.



(a) Input (b) Probability of correct class

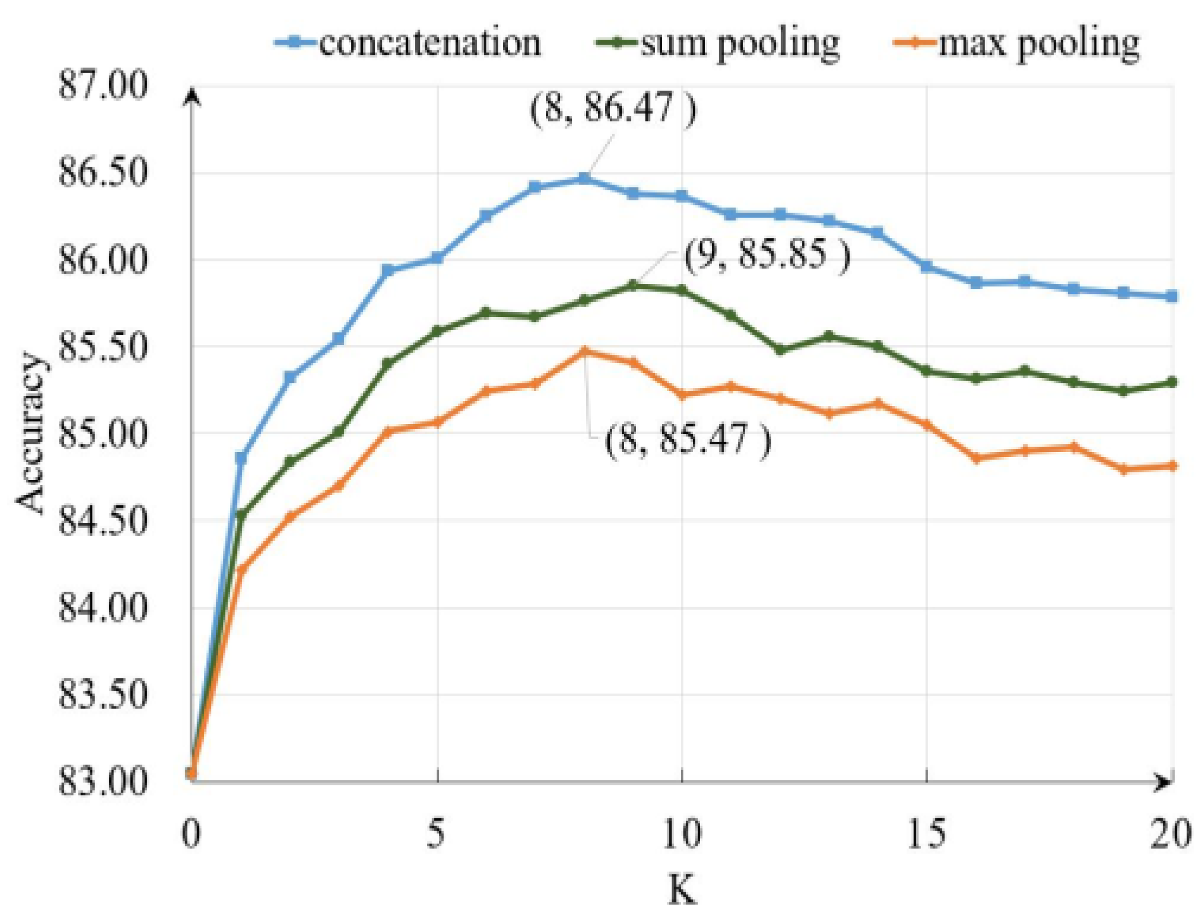
(c) object/sentiment region (d) affection region

Visualization of images from the FI dataset

Experimental Results

	Methods	FI	Flickr
Baseline	AlexNet [45]	60.54	55.13
	VGGNet [69]	70.64	61.28
	Fine-tuned AlexNet	72.43	61.85
	Fine-tuned VGGNet	83.05	70.12
	PCNN (VGGNet) [14]	75.34	70.48
	DeepSentiBank [49]	61.54	57.83
Ours	obj + concatenation	83.85	70.05
	senti + concatenation	84.07	70.10
	AR + concatenation	84.83	70.51
	AR + sum-pooling	84.50	70.46
	AR + max-pooling	84.21	70.49
	AR + concatenation ($K = 8$)	**86.35**	**71.13**

Classification accuracy (%) on the FI and Flickr datasets. We compare our proposed method with different deep methods. The proposed method with different configurations are also given, i.e., combining with the top-1 region (row 7-11), and leveraging more Affective Regions(row 12).



Impact of combining different number of affective regions (K) with three alternative fusion operations (i.e., concatenation, sum pooling and max pooling) on the validation set.

Conclusion

we address the problem of automatically recognizing sentiments in images. We propose to discover affective regions and combine both information using a CNN. The experimental results show that our method outperforms the state-of-the-art methods on the popular affective datasets.

Any comments are welcome.

Email: sherry6656@163.com

Homepage: http://dongyushe.cn