Image Sentiment Analysis
with Convolutional Networks

Zheng Lu, Yunhe Feng

Department of Electrical Engineering and Computer Science
University of Tennessee, Knoxville
Email: {zlu12, yfeng14}@vols.utk.edu

Abstract—**Understanding the underlying attitude of people towards a given image automatically is important for many applications. The rapid growth of social media provides new opportunities and challenges to design image sentiment inference systems. Due to the weak relation between low-level visual features and sentiment, recent work focus on constructing mid-level semantic representation that can both be easily detected from the input images and be mapped to the sentiment class. However, ... We propose a novel convolution network based end to end system that can automatically generate the most suitable mid-level features. We show the effectiveness of our approaches by detecting sentiment of image tweets and show significant improvement against baseline approaches.**

I. INTRODUCTION

Recent years, social media platforms have seen a rapid growth of user-generated multimedia contents. Billion of images are shared on multiple social media platforms such as Instagram or Twitter which contribute to a large portion of the shared links. Through image sharing, users are usually also express their emotions and sentiments to strengthen the opinion carried in the content. Understanding users' sentiments provides us reliable signals of people's real-world activities which is very helpful in many applications such as predicting movie box-office revenues [2], political voting forecasts [9]. It can also be used as the building block for other tasks such as the image captioning [13].

Automatic sentiment analysis recognize a person's position, attitude or opinion on an entity with computer technologies [11]. Text-based sentiment analysis has been the main concentration in the past. Only recently, sentiment analysis from online social media images has begun to draw more attentions. To simplify the task, previously the sentiment analysis mainly focus on the opinion's polarity, i.e., one's sentiment is classified into categories of positive, neutral and negative. However, as pointed out in many recent studies [1], [3], [5], [16], it faces the unique challenge of large "affective gap" between the low-level features and the high-level sentiment.

To overcome this challenge, recent work resort to extract manually designed mid-level representations from low-level features, e.g., visual sentiment ontology (VSO) in [3], mid-level attributes in [16]. Such approaches usually outperforms methods that inferring sentiment directly from low-level features. Rapid developments in Convolutional Neural Networks (CNNs) [6], [8], [10], [12] push the transformations of computer vision tasks. There are also several efforts to apply CNNs

to image sentiment analysis [4], [5], [15]. However, recent works on applying convolutional networks still borrow the network architectures from the image classification tasks [1], [4], [5], [15]. Such a methodology limits the proposed systems works only on images containing object, person or scene. We propose... **another weakness lies in the generalizability to cover different domains.**

A. Contributions

Our contributions can be summarized as follows:
- We propose ...;
- We design ...;
- Experiments on ....

The rest of this paper is organized as follows. In Section II we show previous works on image sentiment analysis. We explain ... in Section III. The evaluations of our proposed method are in Section IV. We conclude our work in Section V.

II. RELATED WORKS

simply write down what did each paper do for now

The majority of work in image sentiment analysis focus on ... .

[3] trys to fill in the "affective gap" between the low-level features and the high-level sentiment by a set of mid-level representation called visual sentiment ontology that consist of more than 3,000 Adjective Noun Pairs (ANP) such as "beautiful flower" or "disgusting food". The authors also publish a large-scale dataset called SentiBank that is widely used in later works. They also extends their work into a multilingual settings in one of their later work [7].

[15] applies CNN to image sentiment analysis. They propose a method to progressively training CNNs by keep training instances with distinct sentiment scores towards sentiment polars and discard training instances otherwise.

[16] adopts a similar methodology as [3]. The authors also construct a mid-level attributes for better classification except that they choose different scene-based mid-level attributes than [3]. In addition to that, they also include face detection to enhance the performance on images containing human face.

the challenge of collecting data: not only the image need to be manually labeled, but also each image normally require multiple people to label it as it is very common for people to have different opinion on a same image. In the end, only a portion of the labeled data can be used,

TABLE I
DATASETS

Dataset	size	class	p/n	workers
SentiBank-Flickr [3]	316,000	1553 ANPs	N/A	auto
SentiBank-twitter [3]	603	2	3.53	3/3
twitter [15]	1269	2	1.54	3/5
Bing [1]	8,812	3	2.76	2/3

Datasets with the no. of images and no. of mid-level representations if applicable. We also list how the label is generated, the number following manual method shows how many workers label each image. There are much more positive images than negative images as people tend to engage more with positive posts [**?**]. We discard all the neutral images in all datasets.

for example, images with all 5 annotator agree on the sentiment label.

Due to the challenge in manually collect labeled data for image sentiment analysis, [14] proposes an unsupervised method to facilitate social media images sentiment analysis with textual information associated with each image.

[5] try to apply CNN based on AlexNet [8] to automaticall extract features based on ANPs proposed in [3].

[1] studies the sentiment analysis of images of social events. It designs specific mid-level representations of each event class and classify sentiment of each image without the help of texts associated with the image.

[4] studies how to fine-tune AlexNet-styled CNNs to achieve better performance on image sentiment analysis tasks.

Our work falls into the category of ... .

III. THE DESIGN

We propose to solve the ... .

A. Preliminary

In our problem ...

B. Main design

C. More details

IV. EVALUATIONS

In this section, we show the results of experimental evaluation of ...

A. Datasets

We collect several public available data and list them in table I.

We evaluate our algorithm on ...

B. Experiment settings

We evaluate our algorithms in ... setting

C. Classification Accuracy

We show the results...

D. More

We show in this section that ...

V. CONCLUSION

In this paper, we describe a novel method to ...

REFERENCES

[1] U. Ahsan, M. De Choudhury, and I. Essa. Towards using visual attributes to infer image sentiment of social events. In *Neural Networks (IJCNN), 2017 International Joint Conference on*, pages 1372–1379. IEEE, 2017.

[2] S. Asur and B. A. Huberman. Predicting the future with social media. In *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on*, volume 1, pages 492–499. IEEE, 2010.

[3] D. Borth, R. Ji, T. Chen, T. Breuel, and S.-F. Chang. Large-scale visual sentiment ontology and detectors using adjective noun pairs. In *Proceedings of the 21st ACM international conference on Multimedia*, pages 223–232. ACM, 2013.

[4] V. Campos, B. Jou, and X. Giro-i Nieto. From pixels to sentiment: Fine-tuning cnns for visual sentiment prediction. *Image and Vision Computing*, 2017.

[5] T. Chen, D. Borth, T. Darrell, and S.-F. Chang. Deepsentibank: Visual sentiment concept classification with deep convolutional neural networks. *arXiv preprint arXiv:1410.8586*, 2014.

[6] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[7] B. Jou, T. Chen, N. Pappas, M. Redi, M. Topkara, and S.-F. Chang. Visual affect around the world: A large-scale multilingual visual sentiment ontology. In *Proceedings of the 23rd ACM international conference on Multimedia*, pages 159–168. ACM, 2015.

[8] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[9] B. O'Connor, R. Balasubramanyan, B. R. Routledge, and N. A. Smith. From tweets to polls: Linking text sentiment to public opinion time series. *ICWSM*, 11(122-129):1–2, 2010.

[10] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[11] M. Soleymani, D. Garcia, B. Jou, B. Schuller, S.-F. Chang, and M. Pantic. A survey of multimodal sentiment analysis. *Image and Vision Computing*, 65:3–14, 2017.

[12] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

[13] O. Vinyals, A. Toshev, S. Bengio, and D. Erhan. Show and tell: A neural image caption generator. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3156–3164, 2015.

[14] Y. Wang, S. Wang, J. Tang, H. Liu, and B. Li. Unsupervised sentiment analysis for social media images. In *IJCAI*, pages 2378–2379, 2015.

[15] Q. You, J. Luo, H. Jin, and J. Yang. Robust image sentiment analysis using progressively trained and domain transferred deep networks. In *AAAI*, pages 381–388, 2015.

[16] J. Yuan, S. Mcdonough, Q. You, and J. Luo. Sentribute: image sentiment analysis from a mid-level perspective. In *Proceedings of the Second International Workshop on Issues of Sentiment Discovery and Opinion Mining*, page 10. ACM, 2013.