Luo-yang XUE, Qi-rong MAO, Xiao-hua HUANG, Jie CHEN, 2020. NLWSNet: a weakly supervised network for visual sentiment analysis in mislabeled web images. *Frontiers of Information Technology & Electronic Engineering*, 21(9):1321-1333. https://doi.org/10.1631/FITEE.1900618

NLWSNet: a weakly supervised network for visual sentiment analysis in mislabeled web images

Key words: Visual sentiment analysis; Weakly supervised learning; Mislabeled samples; Significant sentiment regions

Corresponding author: Qi-rong MAO
E-mail: mao_qr@ujs.edu.cn
ORCID: https://orcid.org/0000-0002-0616-4431

Motivation

1. The annotation of large-scale datasets is expensive and time consuming. Instead, it is easy to obtain weakly labeled web images from the Internet.

2. Noisy labels still lead to seriously degraded performance when we use images directly from the web to train networks.

3. To address this drawback, we propose an end-to-end weakly supervised learning network, which is robust to mislabeled web images.

Main idea

1. Annotation cost for training samples is reduced in our system because mislabeled web images can be used directly for the training process by the proposed non-extreme channel attention (NECA) module, which will reduce the negative influence of mislabeled samples in sentiment recognition.

2. We propose a spatial-class activation map (SCAM) module to stimulate the network to discover the special-class and significant regions, which may assist complicated sentiment analysis.

3. We introduce the regularization for visual sentiment classification to effectively learn the discriminative representation, which could be favorable in sentiment classification.

Method

1. The NECA module is proposed to use mislabeled web images directly for the training process.

2. The SCAM module is proposed to stimulate the network to discover the special-class and significant regions.

3. The regularization for visual sentiment classification is proposed to effectively learn the discriminative representation.

Major results

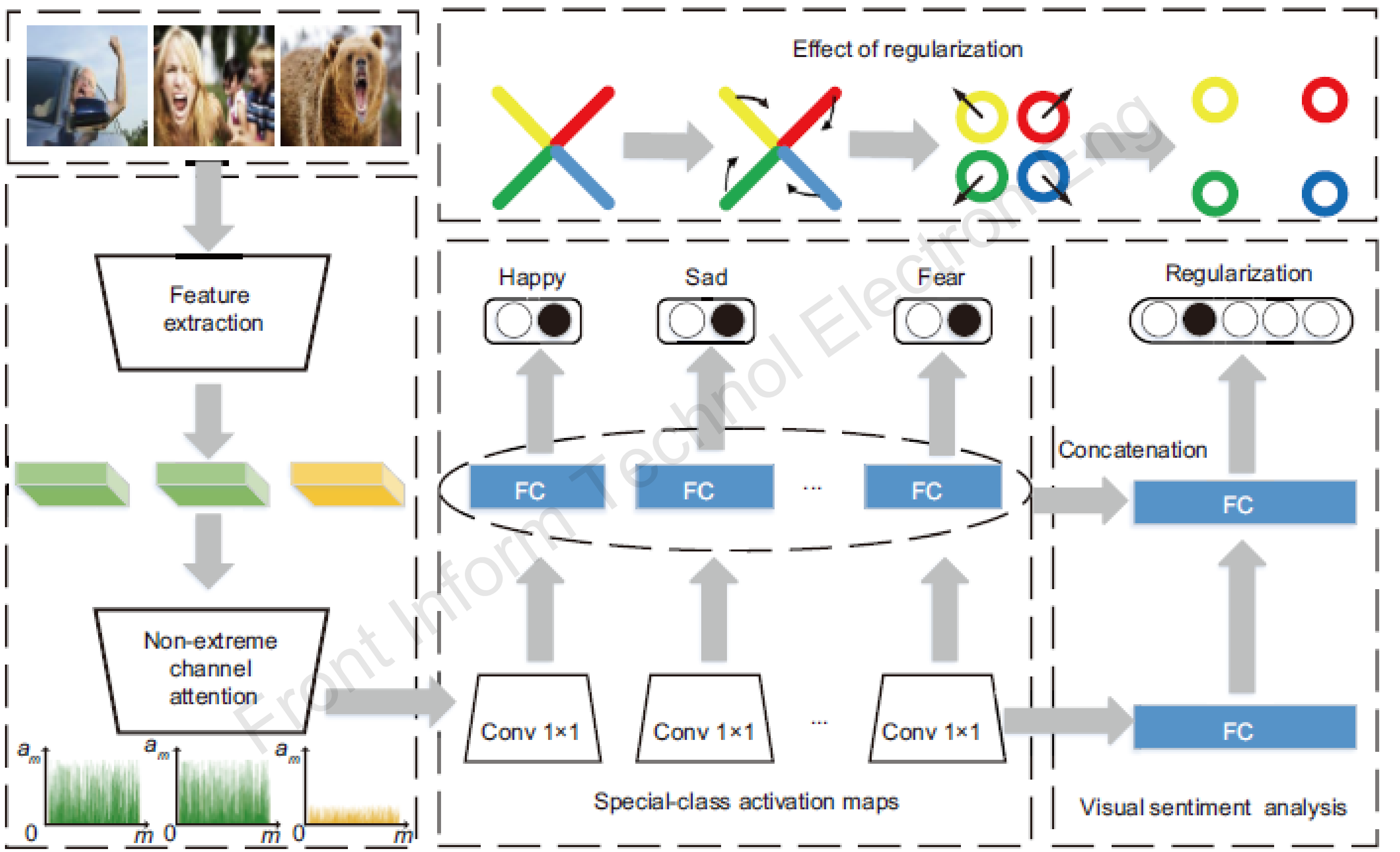

Fig. 2 Architecture of the weakly supervised learning network (NLWSNet)

NLWSNet incorporates a non-extreme channel-attention module, a spatial-class activation map module, and classification regularization. It can reduce the annotation cost tremendously and enhance the robustness of sentiment analysis on mislabeled web images. FC: fully connected layer; Conv: convolutional layer. Feature maps of correctly and incorrectly labeled samples are in green and orange, respectively. References to color refer to the online version of this figure

Major results (Cont’d)

1. Test results of our model and related methods

Table 3 Classification accuracy comparison for model ablation on Twitter II, EmotionROI, Flickr, and Instagram datasets

Method	Classification accuracy (%)			
	Twitter II	EmotionROI	Flickr	Instagram
BASED	73.34	41.02	73.46	68.35
BASED+NECA	80.22	54.63	82.45	81.10
BASED+SCAM	80.34	55.75	81.54	79.97
BASED+NECA+SCAM	82.46	57.84	81.94	82.48
BASED+NECA+C	84.35	60.06	82.98	82.97
BASED+NECA+T	82.96	57.94	83.17	83.12
BASED+NECA+SCAM+R	**85.43**	**61.25**	**84.62**	**83.55**

The pre-trained DenseNet121 using ImageNet is used as the BASED method. Best results are in bold

Major results (Cont'd)

2. Test results of our model and related methods

Table 4 Classification accuracy comparison on the mislabeled datasets

Dataset	Method	Classification accuracy (%)										
		δ=0	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
Flickr	BASED	73.46	69.80	66.14	65.34	62.63	57.87	53.21	49.76	48.45	42.14	40.21
	CWA (Chen L et al., 2017)	75.54	72.23	70.10	66.43	62.12	58.46	55.16	50.13	48.61	43.17	41.84
	SA (Chen L et al., 2017)	79.34	75.01	72.14	70.12	63.26	58.93	56.33	51.11	48.95	43.44	41.92
	SA (Zhuang et al., 2017)	79.89	77.75	73.41	69.96	63.67	59.05	56.60	50.98	48.79	45.53	43.03
	Ours	84.62	84.33	80.60	82.14	79.93	80.73	75.73	73.87	72.12	70.33	67.42
Instagram	BASED	68.35	64.32	60.73	54.17	52.86	48.57	45.53	43.32	39.45	35.57	33.63
	CWA (Chen L et al., 2017)	75.22	70.01	65.88	62.53	59.84	55.41	51.75	46.92	42.46	37.18	32.98
	SA (Chen L et al., 2017)	77.13	73.26	70.85	67.69	63.24	59.60	56.06	52.70	48.28	44.07	39.24
	SA (Zhuang et al., 2017)	79.57	76.14	72.33	67.44	64.26	60.50	56.99	53.67	48.53	44.67	40.05
	Ours	83.55	79.72	78.10	78.85	81.19	77.50	74.36	71.61	67.48	69.34	66.62

Results become smaller as δ becomes larger, but our method still works when the value of δ is close to 0.50. δ is the proportion of the mixed samples from Cifar10

Major results (Cont'd)

3. Test results of our model and related methods

Table 5 Classification accuracy comparison with several baseline methods on Twitter, EmotionROI, Flickr, and Instagram

Category	Method	Classification accuracy (%)			
		Twitter II	EmotionROI	Flickr	Instagram
Hand-crafted	Zhao et al. (2014)'s	67.92	34.84	66.61	64.17
	SentiBank (Borth et al., 2013)	66.63	35.24	69.26	66.53
	DeepSentiBank (Chen T et al., 2014a)	71.25	42.35	70.16	67.13
CNN-based	ImageNet VGG16 (Simonyan and Zisserman, 2014)	67.49	37.26	69.88	63.44
	ImageNet DenseNet121 (Huang et al., 2017)	73.34	41.02	73.46	68.35
	Fine-tuned VGG16 (Simonyan and Zisserman, 2014)	76.99	45.46	78.14	77.41
	Fine-tuned DenseNet121 (Huang et al., 2017)	78.81	52.63	81.23	80.12
	Sun et al. (2016)'s	81.06	–	79.85	78.67
	Yang et al. (2017a)'s	–	52.40	–	–
CAM-related	WILDCAT (Durand et al., 2017)	79.53	55.05	80.67	80.31
	SPN (Zhu Y et al., 2017)	81.67	52.70	79.71	79.53
	WSCNet (Yang et al., 2018b)	84.25	58.25	81.36	81.81
	Ours	**85.43**	**61.25**	**84.62**	**83.55**

Datasets with incompatible class numbers cannot be evaluated and the classification accuracies are denoted as "–." Best results are in bold

Conclusions

1. We have proposed a weakly supervised network (NLWSNet) to learn visual sentiment representation from a mislabeled dataset.

2. NLWSNet can effectively handle label noises through a non-extreme attention mechanism and a special-class activation map module.

3. The efficacy of our method has been demonstrated by extensive experiments.

Luo-yang XUE received his BS degree in computer science and technology from Jiangsu University, China, in 2017. He is currently a MS candidate in computer application technology of the School of Computer Science and Communication Engineering at Jiangsu University. His research interests include affective computing and pattern recognition.

Qi-rong MAO received her MS and PhD degrees both in computer application technology from Jiangsu University, China, in 2002 and 2009, respectively. She is currently a professor at the School of Computer Science and Communication Engineering, Jiangsu University. She has published over 40 technical articles, some in premium journals and conferences. Her research interests include affective computing, pattern recognition, and multimedia analysis.