International Journal of Industrial Engineering & Production Research September 2020 Vol. 31, No. 3: 351-360
DOI: 10.22068/ijiepr.31.3.351

RESEARCH PAPER

Automatic Visual Sentiment Analysis with Convolution Neural Network

N. Desai[*1], S. Venkatramana[2] & B. V. D. S. Sekhar[3]

Received 3 May 2020; Revised 24 August 2020; Accepted 9 September 2020; Published online 30 September 2020

ABSTRACT

There is strong demand for the application of automated sentiment analysis to visual and text contents in today's digital world so as to significantly reveal people's feelings, opinions, and emotions through texts, images, and videos in popular social networks. However, conventional visual sentimental analysis has been subject to some drawbacks including low accuracy and difficulty to detect people's opinions. In addition, a considerable number of images generated and uploaded every day across the world complicate the already given problem. This paper aims to resolve the problems of visual sentiment analysis using deep-learning Convolution Neural Network (CNN) and Affective Regions (ARs) approach to achieve comprehensible sentiment reports with high accuracy.

KEYWORDS: *Affective region; Convolution neural networks; Sentiment classification; Visual sentiment analysis.*

1. Introduction

In recent decades, growing popularity of social networks has provided a communal space where the public can share their feelings, emotions, and experience and express their opinions about almost every trending event or issue throughout social network platforms. The considerable challenge the digital world is facing today is to significantly understand the consumers' opinions about visual media records (such as pictures and videos) and to raise research awareness and developments. Modern and effective strategies can considerably improve and facilitate the application of visual sentiment analysis, given that this analysis should be inclusive of mega data comprising affective image retrieval, beautiful feature categorization, sentiment mining, comment assistant, etc. Researchers have considered and utilized several characteristics such as color, texture, and shape of the picture so that computers, like humans, can understand people's sentiments. Instead of manually creating visible characteristics, a popular neural network called Convolution Neural Network (CNN) which is able to automatically learn pictures and representations should be created. Visual sentiment analysis remains more controversial than conventional recognition since the former includes a significant degree of abstraction and subjectivity in the individual identification method.

Identifying and understanding people's different feelings through different pictures in social media is considerably more complicated than other visual identification methods including target analysis, scene identification, etc. since an essential technique is required to track a set of hints regarding visible feeling prediction.

Existing technology faces serious challenges in building a sentiment model due to the "affective gap" between the low-level visible characteristics and high-level emotions. Li et al. [1] suggested a context-aware classifier model regarding bilayer scattered description and subsequently, considered the local and global texts as input. However, this strategy is constrained by its considerable dependency on primary segmentation outcomes which are classified to represent various objects. Moreover, the authors assumed that all texts either in local or global contexts (all regions) would become equally important variables for sentiment prediction. Human recognition methods are performed similarly as stated above; in particular, the human vision method takes into account many different elements of an image in detail [2]. You et al. [3]

* **Corresponding author:** ***N. Desai***

desai@srkrec.ac.in.

1. *N.Desai, Department of IT, SRKREC, Bhimavaram, A.P, India.*
2. *B.V.D.S.Sekhar, Department of IT, SRKREC, Bhimavaram, A.P, India.*
3. *S.Venkatramana, Department of IT, SRKREC, Bhimavaram, A.P, India.*

attempted to create a balance between local image fields based on graphically visual characteristics in order to identify specific areas, although one may be constrained by the inadequate and generalizable nature of sentiment analysis.

To address the mentioned difficulties, it is recommended that local and global data be employed for visual sentiment investigation. An innovative approach called Affective Regions (ARs) includes two distinctive features: 1) an affective region stands a remarkable area which is characterized by various objects because it can catch people's attention; 2) an affective region carries meaningful sentiments. Fig. 1 presents some affective regions in popular datasets [4] [5]. As observed earlier, a visible opinion can be deducted from these regions inside pictures. For example, based on the feature (I), opinions are essentially extracted from the area of the bleeding fingers, while in (II), the lovely leaf slightly than the gray marble carries accurate sentiment. However, the task of manually specifying the affective regions of pictures so as to determine which indicator to use is extremely subjective and laborsome.

This paper recommends a new technique that could simply use picture-level labels to consider the affective regions more automatically; however, the annotation difficulties can be overcome significantly. First, an off-the-shelf tool is applied to make bounding box applicants simultaneously with their object-ness rate about their input picture, which is more excited by the powerful co existence relations inside objects and feelings [6]. Then, the candidate preference technique is used to omit irrelevant recommendations and instead, keeping many valuable ones. A strong CNN is attached to every candidate and is employed to calculate the sentiment rate. The abjectness and sentiment rates are combined to estimate the scope of Affective Regions based on top-K regions so that candidate areas can be re-ranked in terms of abjectness and sentiment rates.

Fig. 1. Sentiments represented by affective regions

Ultimately, the CNN outcomes in local and global terms merge into an alternating fusion process using the linking as well as max and sum pooling, thus facilitating the final prediction.

This research paper proposes a deep-learning framework to automatically find the affective regions of each image which make it possible to extract meaningful data. This process is more common in applicability than previous methods, is independent of object classification, and uses the bounding box solution. A visual opinion prediction pattern is detected by applying deep CNN, which is employed to holistically cover both the global and the local image regions. The final discovery is more efficient for visual feeling analysis and outperforms the state-of-the-art strategies on the affective data.

Test outcomes reveal that the recommended model with a limited range benchmark utilizing the learning variation can be properly used.

This research is structured as follows. Section II reviews the associated works on visual and text sentiment analysis. Section III presents the recommended technique to pinpoint the sentiment region of each image and make accurate predictions using latest deep framework. In Section IV gives image analytics approaches and binary pattern Bag of Visual Words. Section V focuses on classification by SVM and CNN. Section VI discusses experimental results on Twitter datasets. Section VII achieves the aim of visual sentimental analysis.

2. Related Work

Today's modern world improved and developed multiple methods for making an efficient visual sentiment analysis over a large number of images and videos. Image forecast and region-based CNNs fall specifically in the domain of this

research. Earlier techniques utilized to make an effective image prediction could be approximately classified into dimensional strategies and categorical strategies. The dimensional strategies express sentiment in 2D valence arousal parallel place or a 3D space [7]. Hanjalic [8] showed social affective replies utilizing three primary dimensions: valence, arousal, and control (dominance), where there is an equal value for each affective state. Zhao et al. [9], [10] recommended predicting the personalized sentiment perceptions of pictures in the valence-arousal space employing distributed sparse regression as a training standard. The categorical strategies categorize sentiments into one of the symbolic classes. There is also a remarkable performance that predicts the discrete likelihood of various sentiment categories [11]. Given that categorical strategies are more convenient, the categorical sentiment forecast is applied in this research work.

In the case of simple modeling approaches, our early attempts at forecasting affective image engage with common low-level characteristics. Machajdik et al. [12] specified a mixture of strong hand-crafted characteristics using scientific and psychological approaches including composition, color variation, image surface, etc. Lu et al. [13] discussed how the shape characteristics in general images that provoke sentiments could be employed in human beings and produce a proof for the importance of roundness angularity and simplicity-complexity for foretelling sentiment texts. Zhao et al. [14] proposed some strong and invariant visual characteristics based on art policies. These handcrafted optical characteristics are regarded as valid in different small datasets, whose pictures are selected from a few precise areas, e.g., abstract paintings and art photos [15]. To fill the

"affective gap" between the low-level characters and high-level feelings, Borth et al. [16] modeled a mid-level theory, i.e., Adjective Noun Pairs (ANPs), to identify image ideas instead of showing opinions immediately. Li et al. [17] estimated the total value of the textual opinions that ANPs represents in pictures. Yuan et al. [18] suggested using the Sentribute, an image-sentiment investigation algorithm based on 102 mid-level properties, making it easier to understand and quicker to apply to high-level cognition. Moreover, Zhao et al. [19] described the characteristics of various levels including low-level characteristics of elements of art, mid-level characteristics of principles of art, and high-level characteristics of grammatical notion indicators in a multi-graph training structure. Chen et al. [20] used object exposure standards to identify six common things like car, dog, dress, face, flower, and food and introduced a novel analysis model to manage the attributive and comparative relationships among visible emotion ideas. The proposed algorithm in this study examines whether an elected section includes objects or not, which is independent of object classes and more appropriate for real applications. The rest of this paper is structures as follows. Section II summarizes the related works on visual sentiment analysis and deep-learning technique. Section III introduces the proposed method of detecting the affective regions and the deep framework used for sentiment prediction. Sections IV and V present and visualize the experimental results on the popular datasets. Finally, Section VI concludes this study.

3. Deep-Learning Techniques

In an era dominated by technology and the Internet, CNNs are combined to several visual identification schemes. The efficiency of these techniques lies in their strength to learn discriminative characteristics from natural data inputs utilizing the back-propagation algorithm, in contradiction to the conventional identification pipelines which measure hand-engineered characteristics on pictures at a primary preprocessing level. Various modern techniques utilize deep CNNs for image feeling prediction. Based on their earlier research, Chen et al. [21] accommodated deep systems for building DeepSentiBank, a classifying pattern with respect to visual opinion aspects which presents meaningful developments in both annotation precision and retrieval execution. It has been proved that when labeled training data are scarce, supervised pre-training for an auxiliary task, followed by domain-specific fine-tuning, boosts performance significantly [22]. Girshick [23] introduced Fast R-CNN for additional decrease training and examination time, while enhancing detection precision and clarifying the training mechanisms. Fast R-CNN overcomes detection time and excludes the area recommendation calculation to 50– 300 ms per picture based on network design. This is quite in contradiction to the conventional approaches on field-based CNNs for detecting prominent things on images. In fact, its majin objective is to automatically recognize the ARs that extract feeling and apply the local data as the additional sentiment description. This is quite challenging in examining not only the areas including objects

but also the encompassing environment [24], which may have considerable impact on the elected areas. Furthermore, RCNN-based techniques demand ground truth bounding box explanations for training; in addition, they are time-consuming and make workers label the affected areas manually. In this paper, an off-the-shelf engine is implanted to produce object outlines regarded as the competitor affective areas and recommend choosing the AR recognition of the low-level, more-efficient level text.

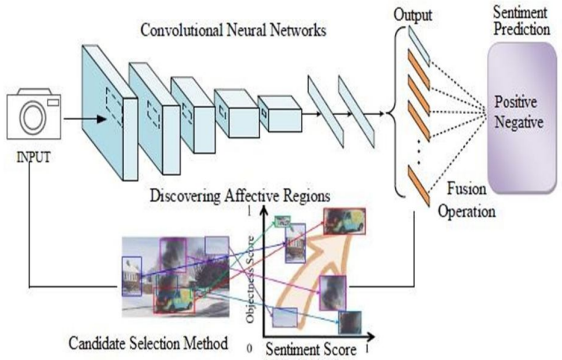


Fig. 2. Channel for CNN Recommended Framework

This research covers the shortcoming of our earlier performance [27] in four major steps: (1) The framework is enhanced in terms of combining the candidate choosing module to overcome the probably noisy aspects and degrade computational contents. (2) There are three fusion actions regarded as alternatives to combining the holistic description with the affective areas, which significantly supports obtaining the local data. As shown in Fig. 2. by considering the input image, thousands of candidates as well as objectness scores are generated, and the candidate selection method is applied to remove the minor candidates. The sentiment score of each proposal is roughly computed via CNN which is combined with the objectness score to discover the affective regions. Finally, the sentiment label is predicted by fusing the local information with the holistic representation using several alternative operations.

Algorithm: Sentiment Analysis on Image
Input: different pictures I
Image Regions: K
Output: people feelings
Step1: Create bound boxes with each object s rate B = fbi;Obj rate Ii gni =1.
Step 2: Use consumer choice technique to create m regions H = fhigmi =1.
Step 3: prepare the classifier using CNN.
Step 4: Predict the complete picture Y Global.
Step 5: Pass H parameter across the second layer to the last layer in CNN.
Step 6: Estimate the probability of m proposal.
Step 7: calculate the sentiment score

$$\text{Score} = \sum\_(j=1)\hat{}c(X\_ij+1).$$

Step 8: Calculates the AR score across every region in
Eq. .
Step 9: Level the corresponding AR scores and choose the highest level of K as sentimental area.
Step 10: Predict the tag X with the pooling process.
Step 10: generate X.

4. Image Analytics Module

The Image Analytics unit incorporates a Bag of Visual Words (BoVW) descent feature in which each learning picture is described by a graphical word event vector. The important facts should first be identified accompanied by the presentation of characteristics to view the picture as a text. This work uses the regional numeric architecture Local Binary Pattern (LBP), calculates a local material description to extirpate characteristics and then, maps these characteristics in language or reference manual to the already existing graphical phrase. Therefore, by contrasting each pixel with the surroundings, the LBP builds regional representation. The LBP functions collected from the train data construct a feature codebook (a visual word dictionary) by grouping all the projections produced. Typically, the picture analysis unit is motivated by text mining and introduces a Bag of Visual Words (BoVW) extracting feature in which each training picture is defined by a video words activity variable. To treat the picture as a text, it is necessary to identify the significant points and then, indicate the characteristics. In this research, the LBP)

functionality descriptive term is applied to obtain a characteristic that calculates a local representation of character and then, these characteristics are drawn round to the existing language or reference manual graphical term. Therefore, by contrasting through pixels due to its corresponding pixels community, the LBP creates regional representation. The clustering technique is applied to significantly cluster the vectors and the most similar characteristics make up the cluster center and reflect a single visual word in the dictionary. Therefore, the visual term frequency range of vector is generated, and the incidence of several visual words gives specific clues about the existence and sort of feeling in the picture. Ultimately, the SVM model could be employed for analysis of sentiments. The components of image analysis are shown in Fig. 3.

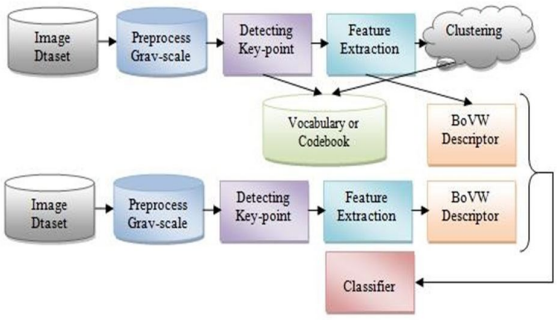


Fig. 3. Visual Analytics Section

Pre-processing is generally used for cleaning each text and then, converting the picture into an appropriate format; this is assumed to be an appropriate approach to doing efficient sentimental analysis. Here, the noise is omitted from the pictures which should be resized and transformed to a gray scale.

Feature extraction Images actually show several nearby landmarks across the edges and corners. Local identifiers have been used to identify the residents. Moreover, the points through Local descriptors features Extraction should be identified using only a local binary method which utilizes the texture detection itself. Specifically, the local binary method is employed to limit the middle pixel size of the frame. In addition, it encodes regional comparison and trends that make it discriminatory. It is also very simple to calculate. The picture is classified into frames of 16x 16 pixels or 32x 32 pixels to each cell for the creation of the feature vector. Each pixel in a cell is contrasted to its 8 surroundings (i.e., cells in top-left, top-right, left, etc.). The sum is allocated according to the principle that if the value of the central pixel is higher than the cost of the neighboring pixel, the value of 0 is specified in its 8x 8 region. In this approach, an 8-digit decimal integer is extracted and translated to its binary part to make it easy to interpret. This total must then be assigned to the pixel in the center. Next, the histogram is determined for every frame and the function vector for the picture is considered. Histogram fragmentation could be done until the truncation. The BoVW framework is described as an unsorted set picture (Tirilly et al.2008). It is equivalent to the description of the Bag-of-Words (BOW) used during text data retrieval of content. This is a histogram description generated ftom from separate elements. This system seems to work in encoding and pooling in 2 phases. Coding is a quite difficult process of allocating every local identifier to its nearest graphical term. Pooling seems are considered the mechanism by which regional predictor projections are performed on average. Therefore, a histogram is generated, following such phases, which includes the presence of each visual term in the picture as shown in Fig. 4.

Typically, every picture is expressed via different local spots and segments, utilizing LBP extracting features to classify these spots (S. and A., 2018). Such vectors are commonly regarded as descriptive words of the apps. After receiving the feature descriptive words, parameters of the very same aspect for each picture will be available. Such variables or descriptors of features are considered plots of the graphical vocabulary in the graphical sentences.

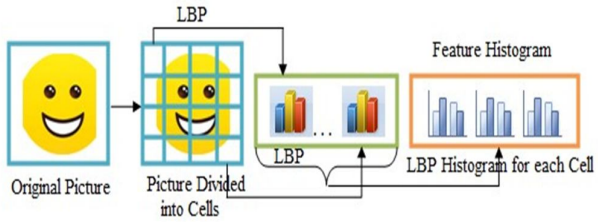
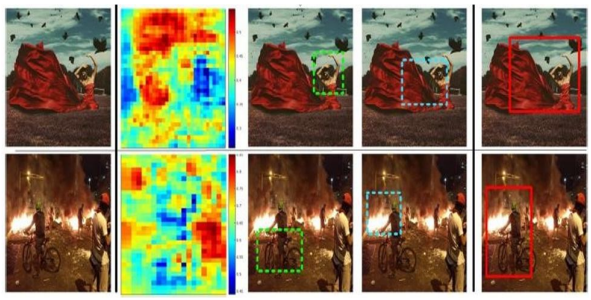


Fig. 4. Visual and Text Histogram

The width of the language could be as precise as the number of graphical words in the language. Different clusters are made from both the descriptive words of an specific element. The hub of every cluster could be used as the vocabulary of the graphical term relation. Whereas Bag of Visual Words (BoVW) is assumed to be an unorganized set of factors, the data of the spatial arrangement of characteristics are rejected; therefore, this technique offers a minimal definition. Precisely, BoVW cannot separate that item from the backdrop of the object.

Fig. 5. Predicted sentiment at the global scale green for positive and red for negative:

In addition, the geographical triangle match that frequently subdivides a picture and calculates histograms of picture characteristics over the corresponding sub-regions could be used to omit this downside of the simple BoVW design. It also leads to a precised production of the feature.



(1) Input Image (2) Probability of correct class (3) Object Region (4) Sentiment Region (5) Affective Region

Fig. 6. Visualization of images from the FI dataset

Given the input image (a), different parts of the image are systematically covered with a gray square and the classifier output (b) is changed. Column (b) denotes a map of the probabilities estimated by the CNN for the ground-truth class, indicating the relative importance of locations in the affective image for the CNN. Further, the most significant regions ranked by different scores are shown (i.e., Obj score, Senti score, AR) as object region (c), sentiment region (d), and affective region (e).

5. Classification

Using the BoVW function qualified via Support Vector Machine (SVM), the image analysis framework significantly predicts the emotion in the picture. SVM functions through discovering a hyper plane that could successfully categorize the number of objects into various classes. SVM taking a named data sets and produces an ideal hyper plane, which could be used to classify new instances. A decision plane differentiates entities from each other through subscriptions of various categories. The whole hyper-plan or decision border is a direct line for a 2D space. SVM evaluates images and identifies image trends throughout this element of picture analytics. The method is presented with a collection of training samples, and it establishes a border to distinguish between the learning groups and training instances.

Therefore, the classification mechanism consists of the following steps:

- Presentng each training picture with a variable that used a BoVW specification.
- Training the SVM algorithm to distinguish variables referring to positively and negatively training objects.
- Implementing the qualified classifier to the testing picture. In the following, the steps involved in this experiment are presented:

Step 1: Send the information (both train and test) as input; Step 2: Keep the activations from CNN as feature vectors; Step 3: Train each object across SVM classifier for each sentiment;

Step 4: Estimate the highest score to every test image;

Step 5: predict the classified label using SVM.

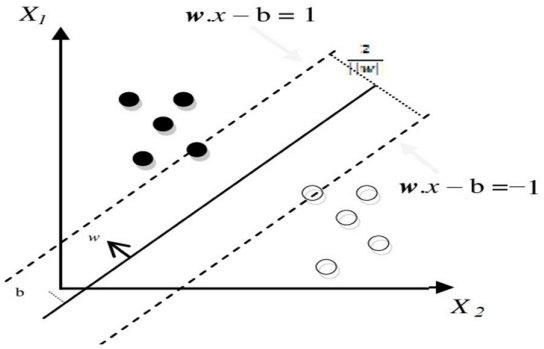


Fig. 7. Classification of hyper plane

$$L= (1-\lambda)Lcls(x, y) + \lambda Lsdl(x, l) \quad \ldots\ldots(1)$$

$$L_{cls}(x, y) = -\frac{1}{N}\left[\sum_{i=1}^{N}\sum_{j=1}^{c} l\left(y^{(i)}=j\right) ln p_j^{(i)}\right]$$

$$L_{sdl}(x, l) = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{c} l_j^{(i)} ln p_j^{(i)}$$

(2) and (3) SVM expresses the data feature vectors in the feature place and introduces to those examples from the train the closest information to the classified hyper plane. For linear separability data, SVM can transform the data into a special dimensional place with kernel objectives and then, change the linear separability difficulty to a linearly divisible difficulty. In other words, CNNs is useful in learning the features of the invariance, and SVM can obtain the optimal classifier covering for features using Formula 1, 2, and 3. Coupled with the earlier parts, this paper combines CNNs and SVM to achieve the text and visual emotional analysis. Since the output vectors of the pooling layer of convolution neural network are described by feature spread of input units, the feature spread description can be utilized as an input in support vector machine. Therefore, while CNNs could be utilized as an automated feature learner, SVM is employed as an emotional classifier; and the two can be combined to deal with the problem of text emotional analysis.

The feeling prediction of unimodality (text and picture individually) is actually carried out by the corresponding components of categorization. To evaluate the feeling of the multimodal information, an extra decision scheme is used.

Such decision method is a Boolean process with only an OR function which puts the results in to the categories of fantastic-grained feelings. The theory of using such a Binary decision-making method is driven by the fact that distinction of document and picture feeling could either increase or decrease the intensity of the overall feeling. The logical operator demands a minimum amount of two inputs to be considered as available; however, there is only one document or picture available as source and the system's corresponding secondary input is 0. It's a bed. 6 The Boolean decision framework is portrayed. Obviously, according to the above table, few cases present a difference in contradictions among the method of treatment of the picture and document. Such instances are referred to as neutral contradictions, as a specific case of sarcastic words is supported by contradictions inside the polarities of feelings. Sarcasm becomes contextual and is correlated to change in polarization intensity in using any reference from various approaches (message-supporting images or text-supporting images). Of note, detecting sarcasm or irony is essential for an enhanced role of classifying feelings that is beyond the considerations of this study.

The following section examines the designer's results and findings.

6. Results and Discussions

The analysis conducted on the most popular social media networking, Twitter, dataset contains 7000 posts including both text and image. The modalities on the dataset contain 50% text and 20% images. Table. 1 shows the real allocation of data in records. Different factors are exercised for both image elements and text analysis during the entire experiment.

Tab. 1. Training Data

Modality	Instances	Positive	Negative	Neutral
Images1000	150	260	280	270
Text 4000	900	1200	500	700

Tab. 2. Confusion Matrix for Binary Classification

	Actual Positive Class s	Actual Negative Class
Predicted Positive Class	True Positive (TP)	False Negative (FN)
Predicted Negative Class	False Positive (FP)	True Negative (TN)

Tab. 3. Metrics of Accuracy, Precision and, Recall

Metrics	Formula	Evaluation Focus
Accuracy	$\frac{TP + TN}{TP + FP + TN + FN}$	Which measures the ratio of correct predictions over the total No. of instances
Precision	$\frac{TP}{TP + FP}$	Which measures the positive elements that are rightly predicting over whole predicted patterns
Recall	$\frac{TP}{TP + TN}$	It calculates the fraction of positive sample that are rightly classify

Tab.4. Performance Results

Modality	Precision	Recall	Accuracy
Images	76.25	80.12	76.01
Text	84	86	86.3

The performance outcomes are estimated based on the source of classification in terms of precision, recall, and accuracy as represented in Table 2. This performance achieved an accuracy more than 92%, especially for the multimodal which is a great enhancement during independent authorization of text and image elements.

7. Conclusion

The findings in this study show that deep-learning performance brings about promising outcomes with a functionality comparable to some techniques that apply handcrafted features to sentiment categorization task. In addition, a few techniques implement deep learning for efficient sentiment analysis. Feeling categorization in images has applications in tagging the images within the emotional category in robotically classifies text, image, and video documents. The present research aims to introduce a hybrid approach to real-time sentiment analysis. Individual analytical strategies depending on modalities are also illustrated. Using Semicircle improves a deep convolution network technology so as to significantly treat the text modality. A bag of features (LBP features) accompanied by the SVM is used to evaluate the feelings hidden in the picture modality.

References

[1] Li, B., et al.,"Context-aware affective images classification based on bilayer sparse representation," in ACM Int. Confence Multimedia, (2012).

[2] Desimone , R., Duncan, J., "Neural mechanisms of selective visual attention," Annu. Rev. Neurosci, Vol. 18, (1995), pp. 193-222.

[3] You, Q.,"Visual sentiment analysis by attending on local image regions," in AAAI Conference. Artif. Intell, (2017).

[4] Luo, J., Jin, H., Yang, J., "Robust image sentiment analysis using progressively trained and domain transferred deep networks," in AAAI Conf. Artif. Intell, (2015).

[5] Borth, D.,"Large-scale visual sentiment ontology and detectors using adjective noun pairs," in ACM Int. Conf. Multimedia, (2013).

[6] Chen, T., "Objectbased visual sentiment concept analysis and application," in ACM Int. Conf. Multimedia, (2014).

[7] Solli, M., "Color based bags-of-emotions," in Int. Conf. Comput. Anal. Images Patterns, (2009).

[8] Hanjalic,A., "Extracting moods from pictures and sounds: Towards truly personalized tv," IEEE Signal Proc. Mag, Vol. 23, No. 2, (2006), pp. 90-100.

[9] Zhao, S., Yao, H., Gao, Y., Ji, R., Xie, W., Jiang, J., Chua, T., "Predicting personalized emotion perceptions of social images," in ACM Int. Conf. Multimedia, (2016).

[10] Zhao, S., "Predicting personalized image emotion perceptions in social networks," IEEE Trans. Affect. Comput, (2018).

[11] Zhao, S., Jiang, X., "Predicting continuous probability distribution of image emotions in valence-arousal space," in ACM Int. Conf. Multimedia, (2015).

[12] Machajdik, J., Hanbury, A., "Affective image classification using features inspired by psychology and art theory," in ACM Int. Conf. Multimedia, (2010).

[13] Lu, X., "On shape and the computability of emotions," in ACM Int. Conf. Multimedia, (2012).

[14] Zhao, S., "Exploring principles-of-art features for image emotion recognition," in ACM Int. Conf. Multimedia, (2014).

[15] Machajdik, J., Hanbury, A., "Affective image classification using features inspired by psychology and art theory," in ACM Int. Conf. Multimedia, (2010).

[16] Borth, D., "Large-scale visual sentiment ontology and detectors using adjective noun pairs," in ACM Int. Conf. Multimedia, (2013).

[17] Li, Z., "Image sentiment prediction based on textual descriptions with adjective noun pairs," Multimed. Tools Appl, (2017), pp. 1-18.

[18] Yuan, J., "Sentribute: image sentiment analysis from a mid-level perspective," in ACM International Workshop on Issues of Sentiment Discovery and Opinion Mining, (2013).

[19] Zhao, S., "Affective image retrieval via multi-graph learning," in ACM Int. Conf. Multimedia, (2014).

[20] Chen, T., "Objectbased visual sentiment concept analysis and application," in ACM Int.Conf. Multimedia, (2014).

[21] Chen, T., "DeepSentiBank: Visual sentiment concept classification with deep convolutional neural networks," ArXiv e-prints, (2014).

[22] Girshick, R., "Rich feature hierarchies for accurate object detection and semantic segmentation," in IEEE Conf. Comput. Vis. Pattern Recog., (2014).

[23] Girshick, R., "Region-based convolutional networks for accurate object detection and segmentation," IEEE Trans. Pattern Anal. Mach. Intell, Vol. 38, No. 1, (2016), pp. 142-158.

[24] Girshick, R., "Fast R-CNN," in Int. Conf. Comput. Vis, (2015).

[25] Peng, K,C., Where do emotions come from? predicting the emotion stimuli map," in IEEE Int. Conf. Image Process, (2016).

Follow This Article at The Following Site:

Desai N, Venkatramana S, Sekhar B. Automatic Visual Sentiment Analysis with Convolution Neural network. IJIEPR. 2020; 31 (3) :351-360
URL: http://ijiepr.iust.ac.ir/article-1-1070-en.html