IMAGE SENTIMENT ANALYSIS ON PAST WORKS - AN EXHAUSTIVE STUDY

Noora.C.T, P. Tamil selvan*

Abstract

Many affective reactions can be triggered in individuals from visuals (example images and videos). Sentiment analysis from visual content isnotasstraight-forward assentiment analysis from text. The early research studies were mainlyfocusedontext-based sentiment analysis[1–4].Text sentiment can be generally classified intopositive,neutral, and negative polarity. While considering visual content, apart from considering these basic polarities, we have to godeeper into emotion prediction. The explosive growth of social networks supplied a significant amount of internet data. The entire world startedtoevaluatethesepublicperceptionsfortheircommercial plans.Analyzingthatinformationhas a great impact on market prediction, political voting fore casting, and brand monitoring. The most widely used application is in the monitoring of people's opinions towards their products or feedback to evaluate the worth of a product.

This paper reviews some publications based on sentiment analysis. Thus, it is categorised based on the different modalities considered by there searcher. As the paper mainly focuses on visual sentiment analysis, most of the journals revie we dare based on visual modalities such as images and videos rather than text.

Keywords: Visual Sentiment Analysis, Visual Content, Emotion, Sentiments, Neural Networks

I. INTRODUCTION

It is challenging to uncover hidden emotions in ani mage because a single image is more powerful than thousand words.People tend to usev isual cues to express their emotions. There is an increasing tendency to express such emotions on social media.

Thus, researchers get a great chance to explore the emotions behind this visual content. Eight specific attitudes were estimated by the majority of earlier studies (Anger, Awe, Amusement, Excitement, Contentment, Disgust, Fear, and Sadness), which are not enough to capture human emotions.

Precisee valuation of certaine motions invisuals has become an explorative area in affective computing. Such visual cues help to produce more personalized predictions with real emotions.

Thus, many practical applications can be seen in education, business, entertainment, advertisements etc.The early multimodal sentiment predictions were based on hand crafted features. This paper attempts to highlight the main contributions on visual sentiment analysis found in papers published in various journals around the world.

II. SENTIMENT ANALYSIS

Sentiment Analysis is a term that is widely used by researchers, business industries, entertainment, and many other fields to extract opinions from the user and thus use them for real-time usage and application. The early studies were mainly focused on a single modality, text. We can track down applications for this mining of texts. For example, analyzing users' purpose behind a message and recognising whether it relates to judgement, news, complaint, objection, idea, appreciation, orinquiry. Recent advances in deeplearning have considerably improved algorithms' capacity for text analysis

[1]Department of Computer Science
[2]Department of Computer Applications
Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu, India
*Corresponding Author

Categorizing visuals as "positive" or "negative" are the goal of the initial studies on visual sentiment analysis. Additionally, by giving a numerical value for each sentiment from the text associated with an image, researchers attempted to determine the relationship of image sentimenttometa-data associated with it. Visual content has mainly 3 common modalities-image, audio, or video. In some studies, researchers attempt to analyse sentiment inany modality by considering features of that modality such as image quality, background, color, and texture. However, the majority of recent work shave primarily focusedonthesemulti-modalities. They are considering two or more modalities together for sentiment detection. Such studies produce better results than previous studies on a single modality. Thus, the research on multi-modal visual sentiment analysis explores great possibilities.

III. EMOTIONAL MODELS

The primary goal of this analysis is to fore see the image's extreme sentiment by positivity, negativity, or neutrality. There are many approaches that took more than three levels of emotions. Shaveret al in, 1987, described a profound list of emotions, in which emotions are arranged in ahierarchy. Numerous studies attempted to characterize fundamental emotions, asitis difficult to decide which emotions belong in which categories. Plutchik's Wheel of emotion from 1980 is the one that is most commonly adopted. It describes 8 fundamental emotions and their three valences. Asper Ekman's hypothesis in 1987, only five fundam-ental emotions-anger,fear,disgust, surprise, and sad-ness-shouldbetakenintoaccount.Someotheremotion classification is described by psychologist Mikels et al., 2005,where the researchers implement an intensive study to bring out a categorical structure of the data set IAPS [Lang et al., 1999]. As a result, a fraction of the IAPS (International Affective Picture System) has been categorized into eight different emotional categories includingamusement,awe,anger,contentment,disgust,excite ment,fearandsad.

IV. LITERATURE SURVEY

Recentre search has focused on visual sentiment analysis. The studies in this growing area focuses on earlier research on retrieving emotional characteristics, which establishes correlations between emotions and low-level image characterist-ics. Additionally, these works had impacted by empirical psychology and art theory[45,46,47]. "Alargeportion of the Sentiment Analysis has been given in the analysis of text"[PangandLee,2008].Despite few studies being utilized to interprete motions from visuals. The greater part of the works in Image Sentiment Analysis depends on past studies close to emotional-aware image retrieval by Colombo et al(1999)and Schmidt, stock(2000), that attempt to track down relationship among emotion and visual cues. In(Siersdorferetal.,2010)there searchers took advantageofmeta-information. They assigned sentiments cores to images and concentrated on the relationships between the sentiment of visuals, regarding different polarity, and visual features like local/globalcolorhistogram.[Borth et al., 2013] proposed “the Visual Sentiment Ontology (VSO) of semantic concepts in view of psychological theories and webmining.”

As per cognitive psychology studies, visuals can strength enrelationships in many diversemanners, such as capturing attention, evokinge motions, and effectively delivering enormous information in a brief amount of time. Such information helps in understanding visual content beyond the semantic concept. Images are the most straight forward medium through which individuals can communicate their feeling soninter personal interaction locales. As a way to express, and share view points and experiences, online entertainment users are increasingly turning to images and videos.

[Chen et al., 2014] calibrated a CNN model in one of the 2.096 ANP categories by extending Senti Bank (Borthetal.,2013),which had been previously trained for the purpose of objectclassification. Deep Senti Bank is the name of the succeeding CNN model. [Xuetal,2014]took advantage of a pre-trained convolutional model forremoving the

activation of fc7 and fc8 significant level features ,furthertrained two logistic regressors for sentiment detection. This took into account the following five sorts ofe motions:neutral,weakpositive, strong positive, and strong negative.[You etal,2015] used a “Progressive CNN (PCNN) method”toperformbinarysentimentimageclassification.[Camposetal.,2015]Identifyingimagesentimentwith pre-trained CNN is a problem that is enhanced by object classification. In order to understand how each CNN layer contributed to the task, there searchers also performed a thorough layer-by-layer study of there fined model. By proposing anew data set called "Product Reviews-150K (PR-150K)",Jin Ye et al[14] address sentiment analysis on multimodalities with the deep tucker fusion approach. Additional testing withVSO, and MVSO results performs better than most of the popular fusion methods. Jie chen, et al[15] in their paper provide a deep CNN method for visual sentiment analysis. By evaluating the classification probability values of the output softwoclassifiers, thesuggesteduniquemethod uses a new sampling strategy to gradually choose a subset of the most useful affective J data.

According to Syed Zohaib Hassan,etal.[26],crowd source study is a good method for collecting, l earning, and extracting public perceptions about the images, but selecting labels

The relevant special challenges with emotions representation models and the available data sets are discussed in [5]. A generalizable analysis is made on the visual sentiment, by identifying and analyzing the components. This work also provides some other problems and methodologies that might be researched, including suggestions for future techniques, features, and datasets. Using recently created ASR and CCM methods, [22] proposed anew model for image sentiment analysis. Two limitations of this method are mentioned in the paper itself. Itisatraditional machine learning concept, despite the model's great performance. It lacks end-to-end, which would make it less useful. Sonali BGaikwad and Prof. S. R. Durugkar [7] proposed a technique forsentiment analysis of images that utilizes the latent correlations with different angles of training images.

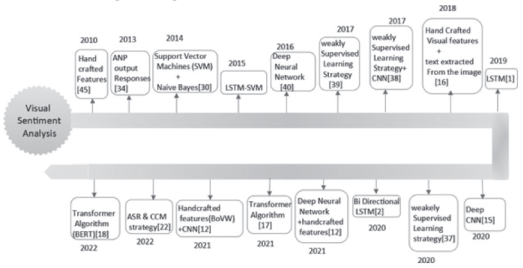

Figure1: Temporal Evolution of Algorithms / Methods in Visual Sentiment Analysis

is not an easy task for a successful study. The study was conducted on disaster images. It reveals that most of the images on social media convey negative sentiments. However, there exist some samples that make people feel joy and relief. To effectively convey people's emotions and sentiments, it can be useful to consider the objects in images such as gadgets, clothes, and destroyed houses, and visual elements like back ground and land marks. This proposed multi-modal classification model evaluates the performance by using predetermined models like VGG Net, Efficient Net, and Dense Net.

They begin by extracting sentiment features from visuals. Based on the projected features, a sentiment polarity classifier is trained in the embedding space. They constructed image data sets via crowd-sourcing for verifying the mentioned approach's efficiency.

Based on Plutchik's wheel of emotions, Tetsuya As akawa and Masaki Aono[12] devised a methodology for visual sentiment analys is that precisely forecasts multi-label multi-class challenges of the images. Using ahy brid deep neural network model, they created a multi-label visual sentiment analysis data set that allowed inputs to come from both manually created features such as Bag of Visual Words-BoVW and CNN features. A threshold-basedmulti-label prediction technique was also introduced."Animagetextinteractionnetwork"onmulti modalities was proposed by Tong Zhu et al.[27].A cross-modal gating module and a cross-modal alignment module

make up this concept. For each region, the word-level textual information is chosen in the first module.The gating module uses a soft gate to combine multi-modal features, which can help to reduce the impact of out-of-alignment area word pairs. Lukas Stappen et al[20], explored sub-symbolic representations from sentic concepts to get in sights into the emotional and contextual information offered by video transcriptions. Additionally, they were able to use the deduced attributes to accurately categorize video clips based on domain-specific features, arousal, and valence. Yash Gherkar, etal[19]implemented sentiment analysis on images-mainly clear faces from the restaurant's review revealings. The focus of this study is to provide an opportunity for the users to analyze business using reviews of images left by several previous customers.

The issue of image sentiment analysis is addressed byLifangWuetal.[8] by exploiting sentiment interaction information of objects. They presented the Sentiment Interaction Distillation which has two branches- the object branch and the global context branch. "Sentiment graph" was created to show the interaction among objects without the use of human remarks, which depicts objects with visual qualities and that characterizes edges with sentimental resemblance. Additionally, they use a technique for knowledge distillation to prevent the noise caused by segmentation errors and the varying amount of data. The method has been tested on five well-known data sets, and the findings show that it is effective. Astimuli-aware VEA network, which includes stimuliselection, feature extraction, andemotion prediction, was proposed by Jingyuan Yanget al.[13] under the influence of the S-O-R model in psychology. By using commercially available tools to select particular emotional stimuli, they first introduced the conceptofstimulusselectionintoVEA. Then,to simultaneously extract various emotio-nal elements from various stimuli, they design three independent sub-networks. To assist the network to learn an emotion, a hierarchical cross-entropy loss is presented in order to extract bogus cases, which aids organizations to learn in an explicit way. Studies showed that on four popular emotional data sets, the suggested strategy consistently outperformed the leading methods.

The primary concerns and methods of analyzing sentiment on visuals were addressed by Alessandro Ortisetal[9].The present status of the art has been examined thoroughly. This survey analyses every viewpoint connected with Visual Sentiment Analysis. Further, for each problem, it suggests a critical viewpoint. Researchers who deal with Visual Sentiment Analysis and its associated challenges and difficulties can use this paper as are source. It is introduced as an organized study of prior researches, depiction of accessible data sets, attributes, and methods. It suggests novel exciting techniques and information resources to explore the problems. Incorporating the BERT and dilated convolutionalBi-LSTM, Sayyida Tabinda Kokabetal.[18]created an improved feature extraction and classification model. It was suggested to use the BERT based model for sentence level categorization. Zero-shot BERT annotated data was applied on pre-trained BERT to extract sentence level semantics and contextual characteristics. Relevant embedding obtained in this way had been given to the Bi-LSTM and convolution neural network. For detecting the long- term sentence sequencing Bi-LSTM is employed. The suggest edhy brid model (CBRNN)was test edusing four different datasets:moviereviews,self-driving carreviews, airlinereviews,and US presidential election results. According to the experimental findings, this CBRNN model is more effective than the other models. Transformer-basedfeaturereconstructionnamedasTFR-NetwascreatedbyZiqiYuanetal[17].ThecoreofTFR-Netisthefeaturereconstructionmodule,whichdirects the extractor in gaining the semantics of the missing modalities features.On two benchmark MSA data sets, all experimental results show that this model still functions effectively even when non aligned features are incomplete in varied modalities and degrees. By suggesting anew source of text, Alessandro Ortis et al[16] address the difficulty of images entiment polarity evaluation. The point is to manage the difficulties with user-provided text, which is usually utilized

in the majority of past studies. This concentrates on a few disadvantages induced by subjective text by its inherent nature, even though it conveys improved performance on subjective text that the user-supplied than that of the use of objective text associated with images. This method uses objective text, which is automatically retrieved from the image and does not exhibit the concerns that have been identified.

Year	Author	method	Output/features
2010	Machajdik andHanbury [24]	Color,texture,ands hape	Sentimentextremityofvisuals
2010	StefanSiersdorferetal.[25]	Handcraftedattributes	each image is analyzed to find the relationshipbetweensentimentandvisualcontent,byassigninganumerical score based on the text that accompaniedit.
2015	Stutijindaletal[28]	CNN	usinglarge-scalevisualsentimentimagesdeterminehowusersfeel aboutacertainevent
2016	Jyoti Islam, YanqingZhang[29]	Transfer LearningApproach onvisualcontent	DeepCNN-basedframeworkdevelopedbyusinghyper-parameters
2014	Malhar Anjaria etal[30]	Supervised LearningMethods	Proposed a hybrid method for extracting opinionfrom Twitter data with direct and indirectcharacteristics built with SVM, Naive Bayes,maximumentropy,andartificialneuralnetworks.
2016	YuhaiYuletal.[31]	Micro-blog analysiswithDeep CNN(Visualand Textual)	DeepCNNwithgeneralizeddropoutforvisualandCNN-based pre-trained word vectors for textualsentimentanalysis.
2010	S.Siersdorferetal[32]	Socialmediaimagess entimentextraction	GlobalaswellaslocalRGBhistogramandSIFT-basedbagofvisuals
2014	S.Zhaoetal[33]	Image emotionrecognition	characteristics such as Balance, emphasis, harmony,diversity, gradation, and movement were consideredtoexaminetherelationshipofartisticprinciples withemotions
2013	D.Borthetal[34]	Adjective-Noun Pairmethod	visualsentimentontologyanddetectorswereemployedinthemodel. Preservesboththelocationandemotionalcontentofobjects inanimage.

Table1: Some Pertinent Publications on Sentiment Analysis of Visuals

Such observations and test results justify the automatic generation of objective text from images

Quoc-Tuan Truong, et al. [3] introduced a model to predict the sentiment from review images posted by customers. They find out that the emotion expressed in the review images can be represented upon three factors :the image viewpoint , user feature, and the factor of the reviewer, which uses two different models, the user-oriented Model:uVSCNNanditem-oriented CNN. Opinion assessment from visuals is a difficult exploration issue in visual sentiment analysis. Past investigations zeroed on a couple of explicit sentiments and have not caught plentiful mental human sentiments. These days, videos are an important method for expressing emotions on the internet. But sentiment detection of these contents is still in its early stages. The initial works on multi-modal analysis were mainly based on handcrafted features, even though they lead to sub-optimal results. Recent research [13] has shown that the affective regions in an image are mostly responsible for triggering human emotion. We can find out a sub conscious connection between affective regions and the comments of people who have seen the visuals. Most of the studies on visual sentiment analysis especially carried out facial expression analysis with no genuine application.

Due to the wide range of objects and depend encies among them, sentiment analysis is a difficult task. The ability of CNN to learn complex features has been proved. Das et al.[35] created an attention mechanism model of deep learning which concentrates on particular areas and figuring out the necessary emotion. Weakly supervised coupled network (WSCNet) is a model that She et al. [37]devised to find significant regions by lowering the load of emotions. Aweakly supervised learningmethod and CNN model were integrated to achieveend-to-end image sentiment extraction Zhuetal [38]. Durandetal, [39] designed WILDCAT by using deep ConvNets with a weakly supervised learning strategy.[36–39]methods reduc-ed the load of annotation. Similar to this, Sunetal. [40] used a deep network to identify the affective areas. To study the sentiment reaction of local regions, a multilevel region-based approach was developed by Raoetal[41]. Mldr Net model was proposed to learn multi-level deep representations [42].Wu et al. [42] also promoted a multi-attention model for collaboratively identifying and localizing numerous pertinent local regions that provided expected features.

Although the issue with sentiment annotation has perhaps been resolved, most weakly supervised models' [37–42] classification performance is still poor. In contrast to the a fore mentioned analyses, Simonyan and Zisserman[43] developed Smiley Net, which would reinforce the deep association of emojis and photos invast amounts of easily

accessible social media data. They achieved this by employing a novel sentiment-aligned image embedding technique. Emoji data may be sparse in the largest publicly available dataset, but the SmileyNet model makes use of additional emoji information.

V CONCLUSION

This paper intended to give an exhaustive outline of the issues in visual sentiment extraction, the algorithms prefered in various studies. Visual content is having plentiful cues to express one's exact emotion. On the other hand,wecanfinddifferentdimensionstoaparticularvisualinthe lightofthevarious objects, colors, interactions, and soon. Hence it is a complex task to go beyond these mantic concept. This work examined some concerns and methods connected with VSA.

REFERENCES

[1] S. Y. Tseng, S. Narayanan, and P. Georgiou, "Multimodal embeddings from language models for emotion recognition in the wild," IEEE Signal Processing Letters, vol. 28, pp. 608–612, 2019.

[2] H. Tang, D. H. Ji, C. L. Li, and Q. J. Zhou, "Dependency graph enhanced dual-transformer structure for aspct-based sentiment classification," in Proceedings f the Fifty Eight Annual Meeting of the Association for Computational Linguistics (ACL), pp. 6578–6588, July 2020.

[3] M. Phan and P. Ogunbona, "Modelling context and syntactical features for aspect-based sentiment analysis," in Proceedings of the Fifty Eight Annual Meeting of the Association for Computational Linguistics (ACL), pp. 3211–3220, July 2020.

[4] N. Charlton, C. Singleton, and D. V. Greetham, "In the mood: the dynamics of collective sentiments on Twitter," Royal Society Open Science, vol. 3, no. 6, Article ID 160162, 2016.

[5] Alessandro Ortis, Giovanni Maria Farinella and Sebastiano Battiato "An Overview on Image Sentiment Analysis: Methods, Data sets and Current Challenges" published on: January 2019 as the proceedings of 16th International Conference on Signal Processing and Multimedia Applications

[6] Shaojing Fan, Zhiqi Shen, Ming Jiang, Bryan L. Koenig, Juan Xu, Mohan S. Kankanhalli, and Qi Zhao "Emotional Attention: A Study of Image Sentiment and Visual Attention" Publisher: IEEE, 16 December 2018

[7] Ms. Sonali B Gaikwad*, Prof. S. R. Durugkar "image sentiment analysis using different methods: a recent survey" journal: international journal of engineering sciences & research technology, published on: january 5, 2017

[8] Lifang Wu, Sinuo Deng, Heng Zhang and Ge Sh "Sentiment Interaction Distillation Network for Image Sentiment Analysis" published on: 29 March 2022

[9] Alessandro, Giovanni Maria Farinella ,Sebastiano Battiato "Survey on Visual Sentiment Analysis "published on:14 May 2020

[10] Ganesh Chandra sekaran, Naaji Antoanela, Gabor Andrei, Ciobanu Monica and Jude Hemanth "Visual Sentiment Analysis Using Deep Learning Models with Social Media Data" Published: 19 January 2022

[11] Zeyu wang, Yutong Bai,Yuyin Zhou and Cihang Xie "Can CNN Be More Robust Than transformers" published on:7 Jun 2022

[12] Tetsuya Asakawa, Masaki Aono, "Multi label prediction for visual sentiment analysis using eight different emotion based on psychology" published on: August 2021as the conference proceeding of 4th International Conference on Control and Computer Vision

[13] Jingyuan Yang, Jie Li, Xiumei Wang, Yuxuan Ding and Xinbo Gao "Stimuli-Aware Visual Emotion Analysis"

published on: 27 August 2021,Publisher: IEEE

[14] Jin Ye, Xiaojiang Peng, Yu Qiao, Hao Xing, Junli Li, and Rongrong Ji "visual-textual sentiment analysis in product review"

[15] jie chen, qirong mao, and luoyang xue1 "visual sentiment analysis with active learning", October 21, 2020, publisher: IEEE access

[16] Alessandro, Giovanni M, Giovanni Torrisi, Sebastiano Battiato "visual Sentiment Analysis Based on Objective Text Description of Images"

[17] Ziqi Yuan, Wei Li, Hua Xu, Wenmeng Yu "Transformer-based Feature Reconstruction Network for Robust Multimodal Sentiment Analysis"

[18] Sayyida Tabinda Kokab, Sohail Asghar, Shehneela Naz "Transformer-based deep learning models for the sentiment analysis of social media data" publisher: Elsevier

[19] Yash Gherkar, Parth Gujar, Amaan Gaziyani, and Siddhi Kadu "Sentiment Analysis of Images using Machine Learning Techniques"

[20] Alice Baird ,Erik Cambria, Bjorn W. Schuller" Sentiment analysis and Topic Recognition in Video Transcriptions"

[21] Tetsuya Asakawa, T, Asakawa,Masaki Aono, M, Aono "Multi-label Prediction for Visual Sentiment Analysis using Eight Different Emotions based on Psychology"

[22] Hongbin Zhang, Haowei Shi, Jingyi Hou, Qipeng Xiong and Donghong Ji "Image Sentiment Analysis via Active Sample Refinement and Cluster Correlation Mining"

[23] Alessandro Ortis a, Giovanni Maria Farinella b and Sebastiano Battiato "An Overview on Image Sentiment Analysis: Methods, Datasets and Current Challenges"

[24] J. Machajdik and A. Hanbury, "Affective image classification using features inspired by psychology and art theory," in Proceedings of the ACM International Conference on Multimedia, pp. 83–92, ACM MM), Firenze, Italy, October 2010

[25] Stefan Siersdorfer, Enrico Minack, Fan Deng, and Jonathon Hare. Analyzing and predicting sentiment of images on the social web. In Proceedings of the 18th ACM international conference on Multimedia, pages 715–718. ACM, 2010.

[26] Syed Zohaib Hassan, Kashif Ahmad, Steven Hicks, Pål Halvorsen 1, Ala Al-Fuqaha 2 and Nicola Conci 3 and Michael Riegler 1" Visual Sentiment Analysis from Disaster Images in Social Media"

[27] Tong Zhu; Leida Li; Jufeng Yang; Sicheng Zhao; Hantao Liu; Jiansheng Qian "Multimodal Sentiment Analysis With Image-Text Interaction Network", Journal: IEEE, Date of Publication: 16 March 2022

[28] Stuti Jindal and Sanjay Singh "Image Sentiment Analysis using Deep Convolutional Neural Networks with Domain Specific Fine Tuning", International Conference on Information Processing (ICIP) 2015.

[29] Jyoti Islam, Yanqing Zhang "Visual Sentiment Analysis for Social Images Using Transfer Learning Approach" IEEE International Conferences on Big Data and Cloud Computing 2016.

[30] Malhar Anjaria, Ram Mahana Reddy Guddeti "Influence Factor or Based Opinion Mining of Twitter Data Using Supervised Learning" 978-1-4799-3635-9/14 2014 IEEE.

[31] Yuhai Yu1, Hongfei Lin, Jiana Meng, and Zhehuan Zhao "Visual and Textual Sentiment Analysis of a Micro-blog Using Deep Convolutional Neural Networks."

[32] S.Siersdorfer, E. Minack, F. Deng, and J. Hare, "Analyzing and predicting sentiment of images on the social web," in ACMMM, 2010, pp. 715–718.

[33] S. Zhao, Y. Gao, X. Jiang, H. Yao, T.-S. Chua, and X. Sun, "Exploring principles-of-art features for image emotion recognition," in ACMMM, 2014, pp. 47–56

[34] D. Borth, R. Ji, T. Chen, T. Breuel, and S.-F. Chang, "Large-scale visual sentiment ontology and detectors using adjective noun pairs," in ACMMM, 2013, pp. 223–232.

[35] P. Das, A. Ghosh, and R. Majumdar, "Determining attention mechanism for visual sentiment analysis of an image using svm classifier in deep learning-based architecture," in Proceedings of the Eighth International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions), pp. 339–343, ICRITO), Noida, India, June 2020.

[37] D. She, J. Yang, M.-M. Cheng, Y.-K. Lai, P. L. Rosin, and L. Wang, "WSCNet: weakly supervised coupled networks for visual sentiment classification and detection," IEEE Transactions on Multimedia, vol. 22, no. 5, pp. 1358–1371, 2020.

[38] Y. Zhu, Y. Z. Zhou, Q. X. Ye, Q. Qiu, and J. B. Jiao, "Soft proposal networks for weakly supervised object localization," in Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 1841–1850, Venice, Italy, October 2017.

[39] T. Durand, T. Mordan, N. *ome, and M. Cord, "WILDCAT: weakly supervised learning of deep ConvNets for image classification, pointwise localization and segmentation," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5957–5966, CVPR), Honolulu, HI, USA, July 2017.

[40] M. Sun, J. F. Yang, K. Wang, and H. Shen, "Discovering affective regions in deep convolutional neural networks for visual sentiment prediction," in Proceedings of the IEEE International Conference on Multimedia and Expo (ICME), pp. 1–6, Seattle, WA, USA, July 2016.

[41] T. Rao, X. Li, H. Zhang, and M. Xu, "Multi-level region-based convolutional neural network for image emotion classification," Neuro computing, vol. 333, pp. 429–439, 201

[42] Z.Wu, M. Meng, and J. Wu, "Visual sentiment prediction with attribute augmentation and multi-attention mechanism," Neural Processing Letters, vol. 51, no. 3, pp. 2403–2416, 2020.

[43] K. Simonyan and A. Zisserman, "Smile, be Happy:) emoji embedding for visual sentiment analysis," in Proceedings of the IEEE International Conference on Computer Vision Workshop.
, October 2019.

[45] Margaret M Bradley. Emotional memory: A dimensional analysis. Emotions: Essays on emotion theory, pages 97–134, 1994.

[46] Johannes Itten. The Art of Color: The Subjective Experience and Objective Rationale of Color. John Wiley & Sons Inc, 1973

.

[47] Peter J Lang. The network model of emotion: Motivational connections. Perspectives on anger and emotion: Advances in social cognition, 6:109–133, 1993.