ADVANCES IN IMAGE-BASED SENTIMENT ANALYSIS: A COMPREHENSIVE REVIEW

Pratishtha Gautam[1*], Prof. Puneet Kumar[2]

Article History: **Received:** 14/04/2023 **Revised:** 6/05/2023 **Accepted:** 12/06/2023

Abstract

Visual media makes image analysis more important and complicated. Images and photos are increasingly used to provide information to the web's enormous audience. Start using photos to understand image propagation patterns. Images, like words, should evoke complicated emotions. Unlike text, where semantics and context are easily accessible for sentiment analysis, extracting and understanding image sentiment is difficult. This paper presents the intermediate characteristics-based picture emotional tone prediction algorithm. Thus, feature selection methods yield simpler results than using the image's low-level attributes. To improve face-containing picture accuracy, it includes many facial expression detection algorithms. The exploratory investigation shows that the proposed framework outperforms earlier techniques in predicting accuracy. Analyzing the predictions could also reveal the relationships between intermediate characteristics and visual emotion.

Keywords: Attention-based Heterogeneous Relational Model (AHRM), Adjective Noun Pairs (ANPs), Long shortterm memory (LTSM), Region Convolutional Neural Network (R-CNN), Segmentation

[1*]Department of Computer Science and Engineering, Chandigarh University, Gharuan, Punjab.
gautampratishtha@gmail.com
[2]Department of Computer Science and Engineering, Chandigarh University, Gharuan, Punjab.
puneet.e11454@cumail.in

[*]**Corresponding Author:** Pratishtha Gautam
[*]Department of Computer Science and Engineering, Chandigarh University, Gharuan, Punjab.
gautampratishtha@gmail.com

DOI: 10.53555/ecb/2023.12.si10.00570

1. Introduction

These days, tweets serve as a common information carrier on the many social networks that have mushroomed into key venues for user-to-user information sharing and conversation. With the billions of bits of textual and visual data provided by social networks, it is now feasible to identify the emotion conveyed by both types of data. However, visual-based sentiment analysis remains in its infancy. Another method of sentiment analysis that does not rely on textual data is to use semantics and concept learning methodologies based on visual characteristics. Current methods for doing sentiment analysis on visual information rely on low-level elements like facial expression identification and user intent to a conclusion. High-level feature learning could also benefit from data included in picture metadata.[1] Methods that analyze photos for objects, human activities, and other visual cues connected to human emotions are becoming more prominent. However, this method is frequently inadequate since the same objects/actions might express various moods depending on the situation in which the shot was taken.[2]

Assessment of visual sentiment was founded on the evaluation of the aesthetic quality of images and the retrieval of emotional semantic images. Mid-level feature representations, such as the face emotion feature Sentribute or Adjective Noun Pairs (ANPs), are also commonly utilized in visual sentiment analysis, in addition to the more standard low-level features like color, texture, and contour .[3]This study goes deep into the architectures that use human attention to improve sentiment analysis in images.[4] The picture categorization considered the following groups of emotions: amusement, anger, excitement, awe, contentment, fear, disgust, and sadness.[5] Figure 1 shows the step-bystep architecture of sentiment analysis.

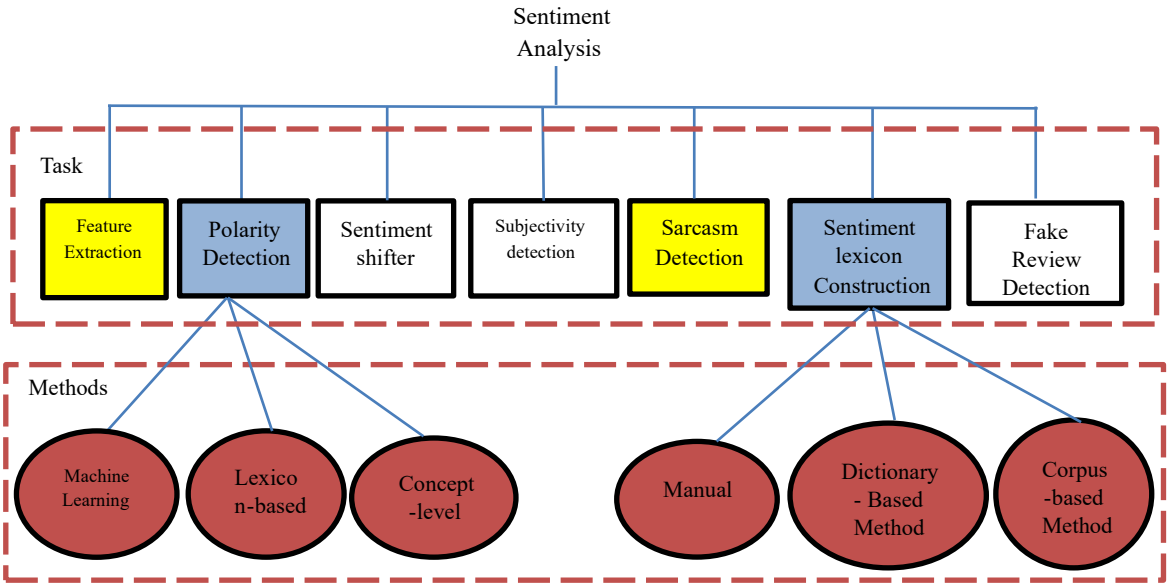

Figure 1: Sentiments Analysis Preview.[6]

The advances made in technology, safety procedures, computing power, and data storage capacity have been very beneficial to the medical imaging field. Segmentation, classification, and anomaly detection in pictures produced by a broad variety of clinical imaging modalities are now key application areas of medical image analysis.[7] Algorithms are based on dictionaries and machine learning. Support Vector Machine (SVM), Neural Networks, Naive Bayes, Bayesian Networks, and Maximum Entropy are all examples of algorithms based on machine learning, while algorithms based on dictionaries make use of statistical and semantic methods. Deep learning, a branch or method of machine learning, gives computers the ability to learn from their mistakes and gain insight into the abstract realm of ideas via exposure to data and practice. Training Deep Learning models requires a vast amount of labeled data and a certain kind of network design called a Neural Network Architecture, which allows characteristics and properties to be learned automatically from the data without human participation or feature detection Region Neural Networks (RNNs), Convolutional Neural Networks (CNNs), Deep Belief Networks (DBNs), Deep Neural Networks (DNNs), and many more deep learning approaches play a significant role in achieving accurate findings in picture sentiment analysis. In the context of picture classification, Deep Learning is thought of as a

framework that results in reliable parameter learning. Since its inception, dictionary learning has proven to be an effective method for learning sparse image features in unsupervised contexts, these features are then applied to image classification and object recognition. Exclusionary dictionaries have been suggested for spatial-spectral sparse representation and image analysis, sparse kernel networks have been introduced for classification, sparse representations over discovered dictionaries have been developed for image pan-sharpening, and saliency-based dictionary learning has been developed for image classification. While these techniques characterize the input pictures, done so in sparse representation spaces that fail to use the high nonlinear character of supervised learning. As a result, deep CNN architectures that seek sparse representations have not yet been used for unsupervised learning of features in the field of remote sensing.[8]

1.1 Deep Neural Network (DNN)

In addition to verbal sentiment analysis, DNN is also utilized for visual sentiment analysis. Multiple layers are used in neural networks, with the input layer receiving the input picture and the output layer providing the result. Deep Neural Networks are Neural Networks with several hidden layers (beyond the input and output layers) that are used to process the input picture in depth. Pixel values signify activations, and computers can only interpret pictures in a matrix form where each pixel has a value. Every neuron has connections with other neurons, and it is the activity of the first layer that controls the activation of the second. There is a total of n layers in the model, labeled $l_1, l_2 \ldots . . l_n$ the input layer is l_1 that cares about the sequence x=x_1, x_2..., x_f, w_1,w_2 ..., w_i are the connection weights, and b_1, b_2.....b_i is the intercepted bias vector. Layer l_1 corresponds to the visible output vector y_1, y_2..., y_i, whereas layers through l_n represent the hidden units, whose outputs are not readily apparent. Input layer values stand in for data delivered into the network, while hidden layer and output layer components represent neurons, the basic computational building blocks. These neurons, also known as activation functions, are tasked with performing a non-linear functional mapping among inputs and a response variable as shown in Figure 2.

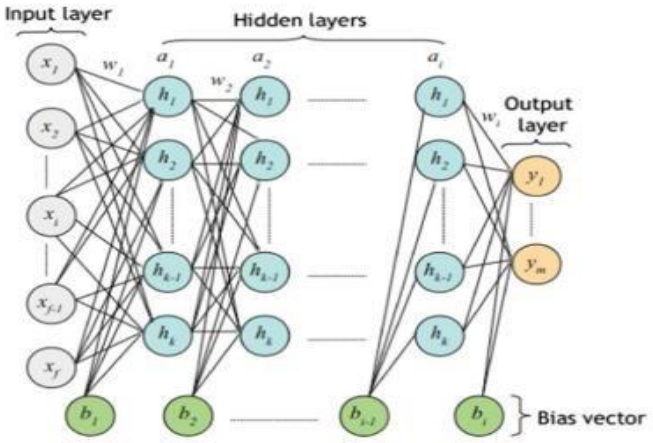

Figure 2: A generic architecture of DNN image extraction.[9]

Each hidden layer's activation g is thus the bias-weighted sum of the inputs from the layer below it. In mathematics, it is represented by the symbol Eq. 1.

$$h_i(x) = g(w_i h_{i-1} + b_i) \quad (1)$$

According to Eq. 2, the final output of the network is determined through the addition of an activation function to the output of the previously hidden layer.

$$y_m(x) = O(h_i(x)) \quad (2)$$

Depending on the issue at hand, a certain activation function O could be selected as the output. The computer calculates the probabilities that each input belongs to each category and then adjusts the output values for each category so that they all lie within the interval. Getting the right values for these parameters is essential for any model to provide accurate results.

1.2 Convolutional Neural Network (CNN)

As a flow neural network, CNN is most often used in the study of images, however, it also has other significant applications in areas such as image classification and picture prediction. To analyze an image's sentiment using CNN, a series of steps must be taken. When an image is fed into a CNN for classification, the first layer is the convolutional layer, where filters are used to pick a tiny segment or matrix of the picture to analyze. This process assumes that the analysis or reading of an image

begins at the top left corner. As the picture moves from one convolutional layer to the next, the output from each layer becomes the input for the preceding layer, resulting in numerous convolutional networks. Following the convolution step, the pipeline's second function is the Nonlinear Layer. The activation function is a part of the CNN that causes it to behave in a nonlinear fashion. Following a nonlinear layer, the pooling layer decreases the size of a large image, or the workload associated with processing that image by decreasing the number of its features. The output is indeed expected after the pooling layer, but the fully linked layer is presented in the meantime. It utilizes the information generated by CNN. The matrix is transformed into a vector and then sent to the fully linked layer.[10] Using the convolution process, CNN can extract a localized set of features from a larger dataset, and it can also consider the relationships between these features.[11] Figure 3 illustrates the salient regions extraction of CNN.

As a visual sentiment analyzer, CNN could be used for either small picture segments or whole images. To ensure that the suggested method works, the widely used model VGGNet was used as a test subject. Since CNN performed well on ImageNet, the author decided to employ a fine-tuning strategy based on a model already trained on ImageNet. VGGNet convolutional layers were left alone, and only the number of outputs from the final fully connected layer was modified.

Figure 3: Illustrate the salient regions of images for feature extraction.[12]

$$L = -\frac{1}{N}\sum_{i=1}^{N}(y_i * \log P_i + (1 - y_i) * \log(1 - P_i)) \quad (3)$$

where N is the total number of pictures used in training, y_i is the actual label, and P_i is the probability of correctly labeling a positive emotion type (1 for positive and 0 for negative). Here is how the author express P_i: $P_i = \sum\_\_\_\_\_ cj^{e=a}1^2 eaj$ (4)

where c is the total number of emotion categories (here, c = 2) and aj is the output of the last completely linked layer.

Through evaluation, it was initially checked whether the input photos included any conspicuous object locations. If so, the relevant sub-images were edited, and the results of those cropping were fed into a model meant to predict the likelihood of a certain emotional response. Every part of the picture was fed into a model that attempted to guess how the subject felt. Finally, the overall image's sentiment probability and its subpictures depending on the salient object's sentiment were combined to forecast the final sentiment probability.[12]

1.3 Region Convolutional Neural Network (R-CNN)

R-CNN is a technique for assessing the sentiment of an image with the aid of objects in the picture by scanning chosen areas to recognize items in the provided image. The name Region Convolutional Neural Network (R-CNN) comes from the fact that it takes an input picture and uses it to create a box border around 2000 areas, which are then fed into CNN. To categorize the item in an input area, SVMs are fed the features (or parameters) that were retrieved using CNNs. The main problem with RCNN is that it requires more time and much more memory since it instructs the network to identify and evaluate 2000 separate areas for a single picture. Figure 4 shows the feature extraction analysis of images by using the Region-CNN technique.

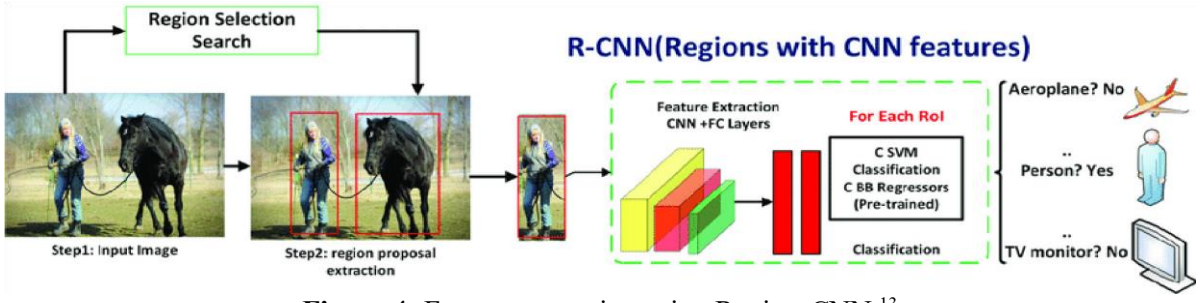

Figure 4: Feature extraction using Region-CNN.[13]

1.4Fast Region Convolutional Neural Network (Fast R-CNN)

It was recommended that a faster object identification method, which is often referred to as Fast R-CNN and is comparable to R-CNN, be developed to overcome the constraints of R-CNN. These drawbacks include a longer processing time and reduced memory use. Before feeding an image into a CNN, the picture is automatically segmented into smaller pieces using Fast R-CNN. This eliminates the need to manually segment the image. First, using a convolutional feature map, regions are detected, and a box boundary is drawn; then, using a n ROI (Region of Interest) pooling layer, the input to the completely connected layer is reshaped such that all images in the fully connected layer are of the same size; and finally, a SoftMax layer is used to predict the image. As a result, Fast R-CNN doesn't need to feed CNN 2000 areas every time. A single iteration of the convolution procedure on an image is all that is required.

2. Literature of Review

This method has been used by several writers, who have reported their findings following a survey of the relevant literature.

Zhou et al., (2021)[14] computed that attention to the new job, dubbed multimodal aspect-based sentiment analysis (MASA), which uses social media material with many types of media to do aspect-based sentiment analysis. It recommends a multimodal interaction model for this assignment, which efficiently studies the relationship between images, text, and sound. Then, it is introduced and makes available a publicly available version of a massive dataset for multimodal aspect-based sentiment analysis called MASAD. In all, there are 57 different categories and 7 different domains represented in MASAD's 38,000 example images and accompanying text. This dataset not only serves as a new standard for MASA but also offers a fresh perspective on aspect-based sentiment categorization.

Feiran et al., (2020)[15] offered a unique method, AMGN, after analyzing the challenge of using similarity and exclusionary data for photo sentiment analysis. Word-related visual characteristics are learned using a visual-verbal attention paradigm. Next, it describes a modality-gated LSTM for dynamically choosing between different affective modalities. Finally, an automated system is used to pick out characteristics with high discriminatory power for use in sentiment prediction, thanks to the semantic self-attention model. Next, it presents a semantic self-attention model that can intelligently zero in on the distinguishing characteristics needed for sentiment analysis.

Lifang et al., (2020)[16] explained CNN has the potential to improve the accuracy of both part and full-picture sentiment analysis. To guarantee the efficacy of the proposed strategy, it was applied to the popular VGGNet model. After seeing how well CNN performed on ImageNet, opted for a fine-tuning method that used a model that had been pre-trained using the data set. During work with VGGNet, additional care was taken to protect its convolutional layers. To make the sub-images more helpful in the long run, a model is also built. Researchers showed that including local information led to an improvement in the accuracy of sentiment estimation when using widely available datasets.

Jie Xu et al., (2020)[17] suggested the multi-modal sentiment categorization using content data and societal connections was suggested using an Attention-based Heterogeneous Relational Model (AHRM). It presents a unique progressive dual attention to draw attention to the emotional conceptual regions and develops a combined image-text representation, thereby capitalizing on the emotional connections between images and texts. Next, it extends the Graph Convolutional Network (GCN) to consolidate the content information from social settings as a supplement to learning high-quality representations, using the extracted social connections to build a heterogeneous relation network. Two standard datasets are used in experiments, and the findings show that perform better than state-of-the-art benchmarks.

Bawa and Kumar (2019)[18] provided a unique methodology for assessing the underlying emotional content of group portraits. The Group Effect Database is used to build and test the framework. The suggested system is split into two modules, the first of which is concerned with global characteristics at the scene level, while the second is concerned with local features based on the expressions of the individuals in the scene. Two separate convolutional neural networks, one for learning scene information and another for learning facial expressions, are created, and trained. The LSTM network is fed data from both the scene-extracting features networks and the face feature extraction networks to learn the joint probability distribution of features.

Fortin et al., (2019)[19] developed a multi-task strategy for analyzing and identifying emotions across several media.

The suggested method extends the common multimodal framework by including two additional image- and text-based classifiers, allowing for the

inclusion of a missing modality both during testing and during the training phase. Experiments have shown that multitask learning acts as a regularization process that could boost generalization and that it also gives a practical and easy solution to a missing modality. These findings support the hypothesis that generalization could be improved by multitasking learning. To the best of my knowledge, this is the first time that monomodal training data has been utilized to enhance multimodal categorization.

Kumar and Jaiswal (2018)[20] project to suggest a deep convolutional neural network-based visual sentiment idea categorization model. Caffe is used to train the deep CNNs model. The produced Adjective Noun Pairs (ANPs) are more useful for visual sentiment categorization after being run through SVM, according to a comparison of the ANPs' efficiency with that of simple text analysis. One of the outcomes of sentiment analysis was compared to a combination of textual and visual sentiment analysis to see which approach yields the most desired results for accurately categorizing emotions.

Paolanti et al., (2017)[21] employed a deep learning approach that considers both visual and verbal data to ascertain the mood conveyed in a business's images. Using information from two independently built deep convolutional neural networks, a machine learning classifier can determine an image's emotional intent. The technique can develop a highlevel description of both visual and textual material, as well as obtain superior accuracy and retention for sentiment classification, by merging DCNNs with machine learning algorithms such as RF, CNN, SVM, NB, DT, and ANN.

Jayalekshmi et al., (2017)[22] created a system that automatically detects facial expressions from the picture and categorizes emotions for a conclusion. Viola Jones 7 is used to recognize facial features, namely those associated with the most salient facial expressions. Local Binary Patterns, Zernike moments, and Discrete Cosine Transform are explained as three techniques for extracting feature points. As soon as the features have been extracted, to combine all the separate pieces of information into one coherent whole, the method of Normalized Mutual Information Selection could be used. SVM, RF, and KNN classifiers are used for development and classification across the whole system. Experiments are conducted using the JAFFE database, and the results show the efficiency of the system in terms of recognition.

Marie et al., (2016)[23] provide an innovative approach to analyzing the emotional tone of a picture by using implicit connections between different perspectives on the same training images. It begins by parsing out characteristics from many points of view, including those that can be seen, read, and felt. The features from various perspectives are then projected utilizing an explicit feature transformation framework inside a multi-view CCA. A sentiment polarity classification is then trained using the projected features in the subspace. Also, include deep learning-based features into the suggested framework since have shown to be very beneficial to many computer vision applications.

Cai and Xia (2015)[24] suggested a novel CNN architecture for carrying out multimodal sentiment analysis that makes extensive use of simultaneous text-level and image-level representation. The suggested technique makes use of the inherent connection between text and picture in image tweets, based on the premise that the two representations have a complementary impact as sentiment characteristics, leading to improved performance in sentiment prediction. The goal is to expand current work on multimedia sentiment analysis to include a wider variety of social media formats, including text, photos, and video.

3. Comparative Analysis

In this section, the authors analyze the models used by various authors and draw comparisons between them. The proposed method was evaluated using several different classification strategies, including CNN, AHRM, LTSM, CCA, Deep Learning, MASA, VGGNet, and AMGN. AMGN has a higher accuracy of 88.2% than CNN, AHRM, MASA, and VGGNet, which have an accuracy of 82.5%, 87.5%, 74.7%, 70.32%, and 84.3% respectively. The accuracy of a forecast is measured by how many instances it is correctly predicted. The following Equation 5 may be used to represent a two-class classification issue, often known as a binary classification problem. True positive, True negative, False positive, and False negative are abbreviated as TP, TN, FP, and FN respectively to calculate the accuracy of the following techniques used.

$$\text{Accuracy} = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad (5)$$

Comparative studies of various models are listed in Table 1.

Table 1: The accuracy of various techniques

Author	Technique	Accuracy
Feiran et al., (2022)	AMGN, LTSM, Self-Attention Semantic Model	88.2%
Marie et al., (2016)	CCA features learning, Deep Learning	74.7%
Lifang et al., (2020)	CNN and VGGNet	84.3%
Zhou et al., (2021)	MASA and MASAD	70.3%
Jie Xu et al., (2020)	AHRM, GCN	87.5%

Figure 5 comprises the comparison graph of various techniques used by the authors in this paper.

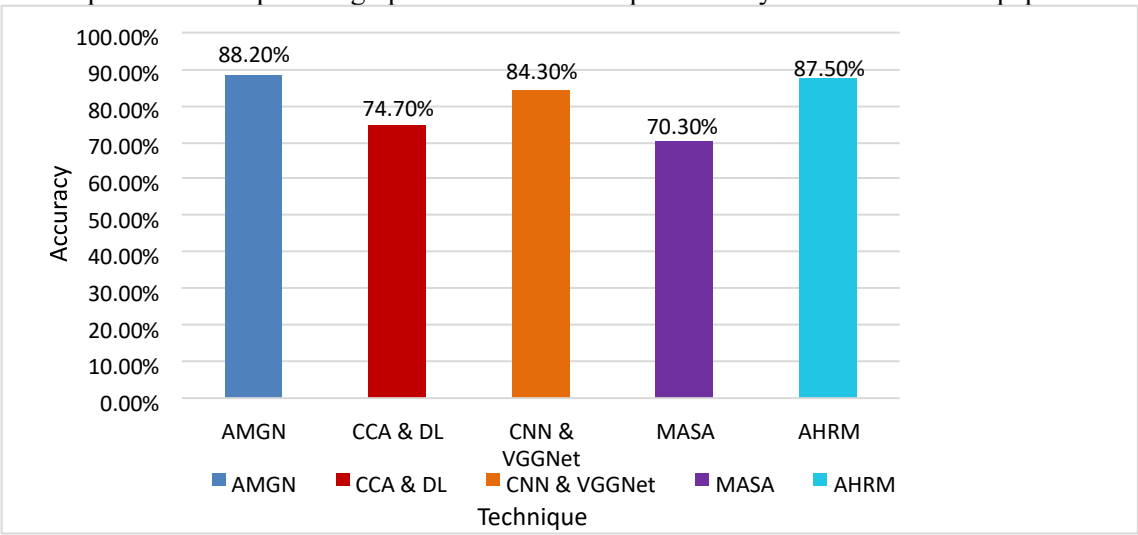

Figure 5: Comparison Analysis of Accuracy

4. Conclusion and Future Scope

In the realm of computer technology, analyzing the emotional content of images is a topic that is both interesting and inspiring. Existing state-of-the-art approaches to sentiment classification mostly focus on visual (midlevel picture) information from the complete image. This study proves the effectiveness of a method for focusing on a localized feeling. Based on these limited visual characteristics, a novel method of sentiment categorization was developed. The key concept is to provide textual labels to certain parts of images. This study presented a deep learning-based selfattention technique for Sentiment Analysis and prediction using the different DL models. Considering different techniques namely, LSTM, CNN, R-CNN, AMGN, AHRM, GCN, VGGNet, and LTSM has been reviewed.

Feiran et al. (2022)[15] performed the recognition, and their AMGN-based method appears to be the most accurate (88.2% accuracy). Finally, by combining data acquired using AMGN, authors have achieved better accuracy. To evaluate the efficacy of the suggested model, several methods are to be explored for creating visual qualities. Considering these findings, the reason is that better quality and more trustworthy visual sentiment analysis might be achieved by enhancing the accuracy of visual features. More remarkably, it can create a map of the specific regions of an image using the AMGN (Attention) model. Using local picture areas and varying strategic applications of deep classification in the last layer of CNN is something that have been working on and anticipate that the results of the work would lead to a rise in the quality of visual sentiment analysis. Planning must entail incorporating multimedia viewpoints and large-scale viewer movies into the process of building effective feature identifiers. To train a robust visual sentiment classifier, it should isolate parts of a picture that are most relevant to the intended sentiment.

Acknowledgment

The author would like to express her sincere gratitude to the members of "Chandigarh University, Gharuan,
Punjab" for their invaluable contributions to this research work. Without their support, this study would not have been possible. The author would like to extend my special thanks to her co-author "Prof. Puneet Kumar" for their guidance, encouragement, and constructive criticism throughout the research process.

In addition, the author emphasizes that no specific grant was awarded for this study by any public, commercial, or non-profit funding source. The author(s) of this article provided all funding for the study.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

1. Yuan, Jianbo, Sean Mcdonough, Quanzeng You, and Jiebo 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*, pp. 1-8. 2013.
2. Wang, Yilin, and Baoxin Li. "Sentiment analysis for social media images." In *2015 IEEE international conference on data mining workshop (ICDMW)*, pp. 1584-1591. IEEE, 2015.
3. Zhao, Ziyuan, Huiying Zhu, Zehao Xue, Zhao Liu, Jing Tian, Matthew Chin Heng Chua, and Maofu Liu. "An image-text consistency driven multimodal sentiment analysis approach for social media." *Information Processing & Management* 56, no. 6 (2019): 102097.
4. Ghosh, Shatadal, Manash Bagchi, Jayant Gangopadhyay, Nataraj Dasgupta, and Anurag Kumar. "Visualizing museums through the visitors' eye: An n-gram model-based text analysis approach." *Journal of Scientific Temper (JST)* 9, no. 1-2 (2022).
5. Ortis, Alessandro, Giovanni Maria Farinella, and Sebastiano Battiato. "An Overview on Image Sentiment Analysis: Methods, Datasets and Current Challenges." *ICETE (1)* (2019): 296-306.
6. Rajabi, Zeinab, and MohammadReza Valavi. "A survey on sentiment analysis in Persian: A comprehensive system perspective covering challenges and advances in resources and methods." *Cognitive Computation* 13, no. 4 (2021): 882-902.
7. Rajkumar, Swetha, and Subasree Palanisamy. "Online detection and diagnosis of sensor faults for a non-linear system." *The Scientific Temper* 14, no. 01 (2023): 216-221.
8. Singh, Rajesh Kumar, Abhishek Kumar Mishra, and Ramapati Mishra. "Hand Gesture Identification for Improving Accuracy Using Convolutional Neural Network (CNN)." *The Scientific Temper* 13, no. 02 (2022): 327-335.
9. Wadawadagi, Ramesh, and Veerappa Pagi. "Sentiment analysis with deep neural networks: comparative study and performance assessment." *Artificial Intelligence Review* 53, no. 8 (2020): 6155-6195.
10. Mittal, Namita, Divya Sharma, and Manju Lata Joshi. "Image sentiment analysis using deep learning." In *2018 IEEE/WIC/ACM international conference on web intelligence (WI)*, pp. 684-687. IEEE, 2018.
11. Liao, Shiyang, Junbo Wang, Ruiyun Yu, Koichi Sato, and Zixue Cheng. "CNN for situations understanding based on sentiment analysis of twitter data." *Procedia computer science* 111 (2017): 376-381.
12. Chandrasekaran, Ganesh, Tu N. Nguyen, and Jude Hemanth D. "Multimodal sentimental analysis for social media applications: A comprehensive review." *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 11, no. 5 (2021): e1415.
13. Aziz, Lubna, Md Sah Bin Haji Salam, Usman Ullah Sheikh, and Sara Ayub. "Exploring deep learning-based architecture, strategies, applications and current trends in generic object detection: A comprehensive review." *IEEE Access* 8 (2020): 170461-170495.
14. Zhou, Jie, Jiabao Zhao, Jimmy Xiangji Huang, Qinmin Vivian Hu, and Liang He. "MASAD: A large-scale dataset for multimodal aspect-based sentiment analysis." *Neurocomputing* 455 (2021): 47-58.
15. Huang, Feiran, Kaimin Wei, Jian Weng, and Zhoujun Li. "Attention-based modality-gated networks for imagetext sentiment analysis." *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 16, no. 3 (2020): 1-19.
16. Wu, Lifang, Mingchao Qi, Meng Jian, and Heng Zhang. "Visual sentiment analysis by combining global and local information." *Neural Processing Letters* 51 (2020): 2063-2075.
17. Xu, Jie, Zhoujun Li, Feiran Huang, Chaozhuo Li, and S. Yu Philip. "Social image sentiment analysis by exploiting multimodal content and heterogeneous relations." *IEEE Transactions on Industrial Informatics* 17, no. 4 (2020): 2974-2982.
18. Bawa, Vivek Singh, and Vinay Kumar. "Emotional sentiment analysis for a group of people based on transfer learning with a multi-modal system." *Neural Computing and Applications* 31 (2019): 9061-9072.
19. Fortin, Mathieu Pagé, and Brahim Chaib-Draa. "Multimodal Sentiment Analysis: A Multitask Learning Approach." In *ICPRAM*, pp. 368-376. 2019.
20. Kumar, Akshi, and Arunima Jaiswal. "Image sentiment analysis using convolutional neural network." In *Intelligent Systems Design and Applications: 17th International Conference on Intelligent Systems Design and Applications (ISDA 2017) held in Delhi, India, December 14-*

16, 2017, pp. 464-473. Springer International Publishing, 2018.
21. Paolanti, Marina, Carolin Kaiser, René Schallner, Emanuele Frontoni, and Primo Zingaretti. "Visual and textual sentiment analysis of brand-related social media pictures using deep convolutional neural networks." In *Image Analysis and Processing-ICIAP 2017: 19th International Conference, Catania, Italy, September 11-15, 2017, Proceedings, Part I 19*, pp. 402-413. Springer International Publishing, 2017.
22. Jayalekshmi, J., and Tessy Mathew. "Facial expression recognition and emotion classification system for sentiment analysis." In *2017 International Conference on Networks & Advances in Computational Technologies (NetACT)*, pp. 1-8. IEEE, 2017.
23. Katsurai, Marie, and Shin'ichi Satoh. "Image sentiment analysis using latent correlations among visual, textual, and sentiment views." In *2016 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 2837-2841. IEEE, 2016.
24. Cai, Guoyong, and Binbin Xia. "Convolutional neural networks for multimedia sentiment analysis." In *Natural Language Processing and Chinese Computing: 4th CCF Conference, NLPCC 2015, Nanchang, China, October 9-13, 2015, Proceedings 4*, pp. 159-167. Springer International Publishing, 2015.

Tables and Figure captions

1. **Figure 1:** Sentiments Analysis Preview (Rajabi et al., 2021).
2. **Figure 2:** A generic architecture of DNN image extraction (Wadawadagi et al., 2020).
3. **Figure 3**: Illustrate the salient regions of images for feature extraction (Chandrasekaran et al., 2021).
4. **Figure 4:** Feature extraction using Region-CNN (Aziz et al., 2020).
5. **Table 1:** The accuracy of various techniques
6. **Figure 5:** Comparison Analysis of Accuracy