Design and Development of Novel Ensemble Classifier for Visual Sentiment Analysis

S.Sai Satyanarayana Reddy
Professor & Principal, Sreyas Institute of Engineering and Technology, Hyderabad, India
saisn90@gmail.com

Abstract. Visual media is now become one of the pre-eminent methods utilized for conveying feelings or suppositions on the web. A huge number of photographs are being transferred by individuals on popular social network platform for interpersonal communication. Visual sentiment analysis (VSA) has pulled in wide consideration since an ever increasing number of individuals will communicate their feeling and conclusions through visual substance via online media. The domain of visual emotion evaluation is impractical because of the great degree of discrimination in the human realization measure. In this paper, a novel ensemble classifier for visual sentiment analysis, which will use few labelled training specimens to attain an effective SA model, is proposed. The classified output is provided by the Ensemble classifier. The steps adopted in the proposed Visual sentimental analysis for the effective classification are pre-processing and the classification. The pre-processing is accomplished by the Resnet-101, which will boost the potential of the classifier. The effectiveness of the proposed sentimental analysis method is revealed by the comparative analysis and it is expected that the proposed method outperforms all the comparative methods. Then, broad datasets enable the quick advancement of profound neural organizations for recognition process. However, the explanation of enormous scope datasets is over the top expensive and tedious. The efficacy of the proposed sentimental analysis method will be revealed by the comparative analysis and it is expected that the proposed method will exceed all the comparative methods.

Keywords: Visual sentiment analysis, Human posture assessment, Classifier, Deep neural network

1. Introduction
The social media greatly influence the lifestyles of the human beings as it enables the sharing of the multimedia data in the social networks. The users can express their opinion in the social media platform through in the configuration of audios, images, videos, emotions and text. This has created a solid perception for assessing the emotions of clients from various strategies. Additionally, the opinions communicated by visual modalities like emojis, pictures, and recordings enhance the extraction of sentiments from the content methodology. These days, visual media are improving as the clients feel more advantageous to communicate their assessment as emojis or pictures [1]. VSA intends to construe the assessment invoked by pictures regarding negative or positive extremity. Traditional methods used in this domain concentrated only on perception highlights (disregarding the content related to the pictures) or have utilized content to characterize an estimation ground truth [9, 10]. Later methodologies utilize the blend of visual and text highlights in a unique nature. A large portion of them, consider SentiWordNet and the WordNet vocabularies as outside information to separate valuable semantic data from printed information. Specifically, SentiWordNet gives three sorts of conclusion extremity outcomes for each word characterized in WordNet, and depicts possibility of sure or negative are the provided terms [2]. VSA is expected to abstractly depict pictures and empower PCs to distinguish and recognize feeling [3]. It can anticipate the extremity of assumption in text by evaluating the hidden data like mentalities, conclusions, and sentiments in the content. This secret data is exceptionally valuable for acquiring understanding the writer’s inclinations and repugnance [4]. Since, the quantity of dynamic Internet clients, who expose their feeling and suppositions on the web is quickly expanding around the world, the programmed evaluation of picture has become a new research topic and generally been utilized in numerous fields (e.g., instruction, diversion and notice) [11], [12]. In the early years, numerous
techniques for hand-created highlights have been presented, which concentrates on the low-level perception highlights (e.g., shading, surface and shape) to recognize the emotion in the image. Now-a-days, the efficacy of the AI dependent methods dominates the traditional methods by automatic recognition of emotion through the huge training samples. An ever increasing number of researchers apply various existing ways to deal with VSA [13][3]. Recent and effective procedures can significantly improve and simplify the process of VSA, given that this evaluation ought to be comprehensive data sharing, involving full of sentiment image recovery, proper element arrangement, conclusion mining, remark associate, and so on. Researcher have thought of and used a few qualities like tone, surface, and state of the image with the goal that PCs, similar to people, can comprehend individuals’ assumptions [7].

For image categorization process, deep neural network (DNN) has exhibited the capacity of learning agent presented in numerous literary works [14]. It can improve characterization exactness when utilizing more analysed samples. Nonetheless, the labelled specimens are not accessible and preferable for recognition task and with respect to time, which is a major issue in the VSA. Furthermore, picture labelling is more troublesome than customary vision undertakings for that the feelings are not totally free in the part of semantic. Despite the most concentrated objects such as feline, canine and bird are more independent [3]. With the developing interest for profound learning models, CNNs are widely utilized in the space of PC vision and have shown leap forward in different errands like activity acknowledgment [11], individual re-identification [12], human posture assessment [13], object identification and limitation [14, 15], and object division [16]. Article acknowledgment includes distinguishing the items in a given picture, while visual assessment investigation (VSA) alludes to the undertaking of perceiving the articles, scenes alongside their enthusiastic setting [17]. The issue of widespread article acknowledgment is all around characterized, while VSA is dynamic in nature because of the great degree of subjectivity in the human acknowledgment measure [18][1]. Existing innovation faces genuine difficulties in building an estimation model because of the “full of feeling hole” between the low-level apparent qualities and significant level feelings [7]. The flow of sentiment analysis method is shown in the figure 1.

Figure 1. Flow of Visual sentimental analysis.
2. Literature review:

| S.no | Author | Method | Pros | Cons |
|------|--------|--------|------|------|
| 1    | Ashima Yadav and Dinesh Kumar Vishwakarma [1] | Residual Attention-based Deep Learning Network (RA-DLNet) | RA-DLNet architecture provides better performance on the six publicly available, real-world challenging datasets for binary sentiment classification | The affect of invoked sentiments on individuals from the Cine trailers are not determined |
| 2    | Alessandro Ortis et al. [2] | Support Vector Machine (SVM) | SVM-based sentiment analysis provides better sentiment polarity classification | SVM-based SA approach is not preferable for large dataset |
| 3    | Jie Chen et al. [3] | Convolution Neural Network (CNN) and texture model | Obtain utmost instructive samples to further evaluation of sentimental attributes for VSA task | CNN requires large training data |
| 4    | Sanghyun Seo et al. [4] | Heterogeneous modality transfer learning (HMTL) | The effectiveness of the target performance of the target representation is improved to great extent | HMTL is limited by its affectability to the qualities of the preparation dataset. |
| 5    | Luo-yang XUE et al. [5] | weakly supervised network (NLWSNet) | NLWSNet can constructively deal the label noise. | Insufficient quantity of labeled data |
| 6    | Papiya Das et al. [6] | SVM Classifier in Deep learning based Architecture | Achieves precise execution on visual sentiment analysis | The inaccurate visual feature degrades the performance of |
VSA system

|    | N. Desai et al.[7] | Convolution Neural Network | It robotically classifies text, image, and video documents | The high dimensional data are required for the training process |
|----|-------------------|----------------------------|----------------------------------------------------------|---------------------------------------------------------------|
| 8  | Zhuanghui Wu et al.[8] | Multi-task learning based approach | Significant improvement in visual sentiment prediction. | Computational complexity |

2.1 Probability of Default (PD)
In this a customer generally borrower will default on debit like credit card and mortgage loan over a time. It basically returns the expected probability of users who fail to repay the loan back to the bank. In this 0% and 100% percentage is used to represent the Probability. Greater probability shows the greater chance of default.

3. Challenges:

- The influence of invoked emotions on humans from the movie trailers is not determined in the RA DLNet approach for the sentiment analysis. Hence, evaluating the impact of evoked emotion is the hectic challenge experienced in RA DLNet [1] approach for the sentiment analysis.

- SVM-based SA approach is not preferable for large dataset, therefore executing the high dimensional data is the prime issue encountered in the SVM-based sentiment [2] analysis.

- CNN [3] requires large training data for the training the entire network. Hence the high execution time is the prime issue that deteriorates the performance of the system.

- HMTL is limited by its affectability to the qualities of the preparation dataset, which is the hectic issue encountered in HMTL [4].

- The prime issue that degrades the performance of NLWSNet is that it suffers from insufficient quantity of labeled data [5].

4. Objectives:
To design and develop an effective visual sentimental analysis model based on deep learning through ensemble classifier to render high classification accuracy and minimize the computational complexity.

5. Proposed method of visual sentiment analysis:
The prime intention of the research is to design and develop the pre-eminent visual sentiment analysis technique. The input sample images will be extracted from the Art_photo [19] dataset. Then, the gathered image is subjected to the pre-processing technique, which is accomplished by the Resnet-50 [20] model. The pre-trained image is fed to the feature extraction process in order to extricate the most significant bits from the image. The final step employed in the proposed technique is the classification of the image as the positive and negative image by the ensemble classifier. The system
framework of the proposed technique is shown in the Figure 1. The Ensemble classifier is developed using the KNN classifier, SVM classifier, and neural network, combined using the fusion constant.

![System framework of proposed technology.](image)

**Figure 2.** System framework of proposed technology.

6. Results and Discussion

The experimentation is done in MATLAB and the efficacy of the proposed method is analyzed based on the performance metrics, such as specificity, accuracy and sensitivity. The methods used for the comparison are RA DLNet [1], SVM [2], CNN + texture model [3], HMTL [4] and NLWSNet [5].

6.1 Experimental result:

The image from the Art photo dataset is taken as the input and this image is pre-trained through the Resnet-101 model. The pre-trained image is then subjected to the Ensemble classifier, which is the well organized classifier widely utilized classifier to attain the best classification output. The input sample are shown in the Figure 1 a) and b) respectively and the classification output is given in the Figure 1 c) and d)
6.2 Comparative analysis of the proposed method with respect to accuracy:

The comparative evaluation in accordance to the accuracy is depicted in the Table 1. At 80% of training the maximum value of the accuracy gained by the proposed Visual sentimental analysis technique is 76.8%, whereas the competent techniques such as RA DLNet, SVM, CNN + texture model, HTML attain the accuracy of 65.3%, 66.2%, 69.3% and 76.0% respectively. The obtained accuracies of the competent techniques show that the proposed Visual sentimental analysis method attains high value of accuracy.

Table 1. Comparative analysis in terms of accuracy

| Methods            | Accuracy in accordance with the training percentage |
|--------------------|----------------------------------------------------|
|                    | 40% of training | 50% of training | 60% of training | 70% of training | 80% of training |
| RA-DLNet           | 64%            | 64.3%            | 64.6%            | 64.9%            | 65.3%            |
| SVM                | 64.6%          | 65.2%            | 65.4%            | 65.5%            | 66.2%            |
| CNN+texture Model  | 64.7%          | 67.3%            | 68.9%            | 69.2%            | 69.3%            |
| HTML               | 71.9%          | 73.7%            | 75.0%            | 75.6%            | 76.0%            |
| Proposed Method    | 73.1%          | 76.1%            | 76.4%            | 76.7%            | 76.8%            |
7. Conclusion

In this paper, a novel ensemble classifier for visual sentiment analysis is presented, which uses few labelled training samples to obtain an pre-eminent SA model is proposed. The classified output is provided by the ensemble classifier. The steps adopted in the proposed visual sentimental analysis for the effective classification are pre-processing and the classification. The pre-processing is accomplished by the Resnet-101, which will boost the effectiveness of the classifier. The effectiveness of the proposed sentimental analysis method is revealed by the comparative analysis and it is expected that the proposed method suppress all the comparative methods.

References

[1] Ashima Yadav and Vishwakarma, D.K., "A deep learning architecture of RA-DLNet for visual sentiment analysis", Multimedia Systems, vol.26, pp.431-451, 2020.

[2] Alessandro Ortis, Farinella, G.M., Torrisi, G. and Battiato, S., "Exploiting objective text description of images for visual sentiment analysis", Multimedia Tools and Applications, pp.1-24, 2020.

[3] Jie Chen, Mao, Q. and Xue, L., "Visual Sentiment Analysis With Active Learning", IEEE Access, vol.8, pp.185899-185908, 2020.

[4] Sanghyun Seo, Na, S. and Kim, J., "HMTL: Heterogeneous Modality Transfer Learning for Audio-Visual Sentiment Analysis", IEEE Access, vol.8, pp.140426-140437, 2020.

[5] Luo-yang Xue, Mao, Q.R., Huang, X.H. and Chen, J., "NLWSNet: a weakly supervised network for visual sentiment analysis in mislabeled web images", Frontiers of Information Technology & Electronic Engineering, vol.21, no.9, pp.1321-1333, 2020.

[6] Papiya Das, Ghosh, A. and Majumdar, R., "Determining Attention Mechanism for Visual Sentiment Analysis of an Image using SVM Classifier in Deep learning based Architecture", In proceedings of 8th IEEE International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO), pp. 339-343, 2020.

[7] Desai, N., Venkatramana, S. and Sekhar, B.V.D.S., "Automatic Visual Sentiment Analysis with Convolution Neural network", International Journal of Industrial Engineering & Production Research, vol.31, no.3, pp.351-360, 2020.

[8] Zhuanghui Wu, Min Meng and Wu, Jiang., "Visual Sentiment Prediction with Attribute Augmentation and Multi-attention Mechanism", Neural Processing Letters, pp.1-14, 2020.

[9] Siersdorfer, S., Minack, E., Deng, F. and Hare, J., "Analyzing and predicting sentiment of images on the social web", In the Proceedings of the 18th ACM international conference on Multimedia, pp. 715-718, 2010.

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

[11] Li, B., Xiong, W., Hu, W. and Ding, X., "Context-aware affective images classification based on bilayer sparse representation", In Proceedings of the 20th ACM international conference on Multimedia, pp. 721-724, 2012.
[12] You, Q., Jin, H. and Luo, J., "Visual sentiment analysis by attending on local image regions", In Proceedings of the AAAI conference on artificial intelligence, Vol. 31, No. 1, 2017.

[13] Peng, K.C., Chen, T., Sadovnik, A. and Gallagher, A.C., "A mixed bag of emotions: Model, predict, and transfer emotion distributions", In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 860-868, 2015.

[14] Ge, W., Lin, X. and Yu, Y., "Weakly supervised complementary parts models for fine-grained image classification from the bottom up", In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3034-3043, 2019.

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

[16] Chen, T., Yu, F.X., Chen, J., Cui, Y., Chen, Y.Y. and Chang, S.F., "Object-based visual sentiment concept analysis and application", In Proceedings of the 22nd ACM international conference on Multimedia, pp. 367-376, 2014.

[17] Solli, M. and Lenz, R.,"Color based bags-of-emotions", In proceedings of International Conference on Computer Analysis of Images and Patterns, Springer, pp. 573-580, 2009.

[18] Hanjalic, A., "Extracting moods from pictures and sounds: Towards truly personalized TV", IEEE Signal Processing Magazine, vol.23, no.2, pp.90-100, 2006.

[19] Peng, Kuan-Chuan, Amir Sadovnik, Andrew Gallagher, and Tsuhan Chen. "Where do emotions come from? predicting the emotion stimuli map." In 2016 IEEE International Conference on Image Processing (ICIP), pp. 614-618. IEEE, 2016.

[20] Yadav, A. and Vishwakarma, D.K., "A deep learning architecture of RA-DLNet for visual sentiment analysis", Multimedia Systems, vol.26, pp.431-451, 2020.

[21] Ashwani Kumar, S. S. S. Reddy and V. Kulkarni, "An Object Detection Technique For Blind People in Real-Time Using Deep Neural Network," 2019 Fifth International Conference on Image Information Processing (ICIIP), Shimla, India, 2019, pp. 292-297, doi: 10.1109/ICIIP47207.2019.8985965.

[22] Ashwani Kumar, “A Review on Implementation of Digital Image Watermarking Techniques Using LSB and DWT” in the Third International Conference on Information and Communication Technology for Sustainable Development (ICT4SD 2018), held during August 30-31,2018 at Hotel Vivanta by Taj, GOA, INDIA.

[23] Sai Satyanarayana Reddy S., Kumar A. (2020) Edge Detection and Enhancement of Color Images Based on Bilateral Filtering Method Using K-Means Clustering Algorithm. In: Tuba M., Akashe S., Joshi A. (eds) ICT Systems and Sustainability. Advances in Intelligent Systems and Computing, vol 1077. Springer, Singapore.

[24] Kumar, Ashwani and Seelam Sai Satyanarayana Reddy, editors. Advancements in Security and Privacy Initiatives for Multimedia Images. IGI Global, 2021. http://doi:10.4018/978-1-7998-2795-5.
[25] Kumar, Ashwani. "Design of Secure Image Fusion Technique Using Cloud for Privacy-Preserving and Copyright Protection." International Journal of Cloud Applications and Computing (IJCAC) 9.3 (2019): 22-36.