INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

A Novel Fusion Attention Algorithm for Sentimental Image Analysis

V Roseline[1]*, G Heren Chellam[2]

1 Research Scholar, Register No. 18221172162009, Rani Anna Govt. College for Women, Affiliated to Manonmaniam Sundaranar University, Abishekapatti, Tirunelveli, 627 012, Tamil Nadu, India
2 Assistant Professor, Department of Computer Science, Rani Anna Govt. College for Women, Affiliated to Manonmaniam Sundaranar University, Abishekapatti, Tirunelveli, 627 012, Tamil Nadu, India

OPEN ACCESS

Received: 20.11.2021
Accepted: 04.02.2022
Published: 05.03.2022

Citation: Roseline V, Chellam GH (2022) A Novel Fusion Attention Algorithm for Sentimental Image Analysis. Indian Journal of Science and Technology 15(9): 386-394. https://doi.org/10.17485/IJST/v15i9.2159

* **Corresponding author**.

rose.vasee2014@gmail.com

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Abstract

Objectives: To implement a novel and hybrid methodology for finding out the positive features when using convolutional neural networks (CNNs) for visual sentiment analysis. To achieve increased accuracy, precision and recall by using this proposed fusion attention methodology. **Methods:** This study proposes a modified methodology encompassing spatial attention, channel attention as well as squeeze excitation modules. An enhanced approach on the basis of convolutional neural networks was used here which utilizes convolution operators by combining both spatial and channel-based data. Moreover, we have incorporated three considerations like spatial, channel as well as squeeze and excitation at various levels for attaining optimal results. **Findings:** The accuracy of the existing approaches was 59.88%, 60.06%, 59.28% and 62.89%, but the proposed fusion attention method showed increased accuracy of 64.15%. Similarly, the F1 score of existing approaches are 0.464804, 0.250164, 0.474129 and 0.2574, but the proposed method revealed increased F1 score of 0.512933. Furthermore, the proposed algorithm showed precision and recall of 0.560896 and 0.472526 which were better when compared with the existing approaches like Res-Target, Resnet50, Alexnet and VGG16. **Novelty:** The novel feature of this proposed fusion attention algorithm was that it incorporates a hybrid approach in which the image together with convolution passes through channel attention, spatial attention as well as squeeze and excitation so as to attain increased accuracy, but most of the existing approaches have used only channel attention and spatial attention modules. In this proposed method, the algorithm performs convolution in 64-bit, 128-bit and 256-bit respectively together in which the three attentions were interchanged in each convolution, which were not prevalent in the existing approaches.

Keywords: Fusion attention algorithm; Sentimental image analysis; Convolutional neural networks; Convolution and pooling; Deep neural network

1 Introduction

The existing research works revealed that hybrid methods were largely employed concerning deep learning networks. The pre-processing methods as well as post- processing method were evolving largely, in which the former generates optimal inputs for networks while the latter targets in improving the network output result [1–3]. Such integrated frameworks will permit higher-level feature extraction using CNN offering better results in terms of accuracy when compared with traditional approaches [4–6].

In this paper, a novel and hybrid methodology was proposed for finding out the positive features when using CNNs for visual sentiment analysis. This proposed work incorporates a hybrid approach in which the image together with convolution passes through channel attention, spatial attention as well as squeeze and excitation so as to attain more accurate results. Here, the channel attention provides a clear identification of different image views. Alternatively, spatial attention gives a representation on the area of the focus. Moreover, squeeze and excitation investigates the output to attain more precise classification.

Previous works showed that, visual sentiment analysis possesses significant portions to substitute the mid-level features for selecting highlights from low-level image features. Driven by the consideration that, a sentiment typically includes undeniable levels of reflection, and this might be simpler as classification of images. Only few papers have concentrated in utilizing the visual elements or characteristics for examination of visual sentiments [7–10]. The significant disadvantage of these methodologies is that, training requires lot of spatial information on psychological research or phonetics to characterize the mid-level characteristics as well as human mediation for calibrating the expectation results.

Deep convolution architectures [11] over-performed every available image classification methods for large-scale ImageNet visual recognition techniques. Deep convolutional neural network (DCNN) represents a layer-based classifier possessing larger quantity of input parameters. Considering larger datasets, fully supervised learning of CNNs was possible without over fitting additional measure of parameters [12]. Recent studies revealed that, the restrictions of CNN that was trained for large datasets, for instance, ILSVRC shall be used in object recognition or image classification tasks when the information was restricted, resulting in better execution of usual representations [13–15]. Very few research works were done on sentimental analysis methods dependent on CNN for visual sentimental predictions. Moreover, it could be seen that, the image representations from CNNs that were trained for larger datasets shall be effectively converted to sentimental analysis. Also, CNNs were found to be acquiring greater considerations among research expert [16] [17]. For creating larger datasets, for instance, ImageNet has driven the modest GPUs, permitted deep CNNs to show increased execution in Artificial intelligence (AI) applications [18]. An effective and gradually trained CNN was formulated towards visual sentimental analysis. Here, it was observed that, CNNs could attain improved accuracy over image sentiment analysis when compared with other existing classification approaches. The main idea was in using AlexNet, a deep neural network (DNN)-based model which was formulated for detecting the objects in ImageNet.

The present research carried out in this paper was an improvement over Convolutional block attention module (CBAM), a direct but effective responsive element over feed-forward CNNs, in which the attention-based feature enhancement was carried out using two individual segments, channel and spatial, thereby it is capable of achieving substantial improvements in performance but also keeps the overhead to be minimal [19,20].

This proposed work combines the Channel, Spatial and SE attention resulting in an enhanced model for convolution neural networks.

2 Materials and Methods

In this proposed research, Twitter dataset was used that consists of 1,269 images collected from tweets [21], with additional 6000 tweets from 100 users. The images were gathered initially and subsequently preprocessing was done. Then, the image was passed to the fusion attention algorithm by which the classification was performed. The image classification was based on the representation as given in Figure 1.

2.1 Input Image

The input information for this proposed model was gathered from online media platform like Twitter, where the images were employed with the objective to share and pass the information as a replacement for text information. Few images in these datasets were additionally composed from Twitter15 images that possess higher image quality so as to be applicable in this present work. In this paper, custom datasets were formed that combines the images that were taken from the internet platforms. For generating user profiles, tweets from 100 users were gathered and a total of 6000 tweets that were relative to this present work were selected based on the applicability of this present research. Some collected tweets for input image generation did not deliver any data about regarding this present work, subsequently this proposed system disregarded those tweets. In this experimentation, 20% of the images were selected in random manner for training, the remaining were utilized for testing.

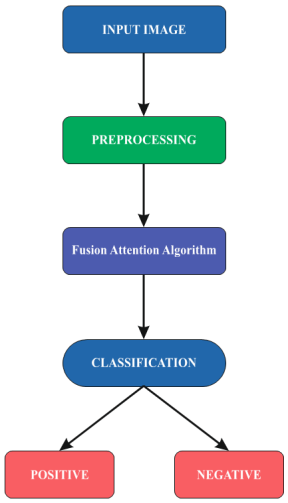

Fig 1. Stages Involved in the Proposed Strategy

2.2 Preprocessing

Preprocessing of datasets was considered to be the foremost and vital stage before continuing to subsequent stages. Therefore, greater number of datasets that were constructed should be totally processed initially by removing the prevalent noises. In this present work, processing was done in fragmented images or whichever recurrent images that do not correlate with the specified desirable standard. Moreover, the images were resized as per the network requirements.

2.3 Fusion Attention Algorithm

Here, fusion algorithm was adopted in which three attention layers were combined together. The algorithm performs convolution in 64-bit, 128-bit and 256-bit respectively, in which the three attentions were interchanged in each convolution along with max and average pooling.

The image was first passed to three dissimilar levels of 64-bit convolution followed by pooling, and then each level possesses three attentions respectively. Later, the process continues identically for 128-bit and 256-bit convolutions by interchanging the levels of attention. This interchange and repetition processes were performed to provide a clear view of the image as shown in Figure 2. The brief explanations of the attention layers were explained below in subsequent sections.

2.3.1 Spatial attention

Human beings normally show interests on emphasizing the object location together with higher prominence which was commonly referred as spatial attentions. Spatial attentions permit individuals in precisely handling visual information via prioritizing the spaces inside the visual areas. In this paper, a region of space within the visual fields was selected for consideration and the information within these areas gets subsequently processed. Investigation showed that, by using spatial attention, the observations will be faster and more precise in identifying the objectives which increases in normal areas when compared with unexpected areas.

Attentions are focused significantly more quickly at unexpected areas as these areas were being made significant with external visual sources inputs. This element self-learns the association of spatial attentions, thereby increases significant region, at the same time limits redundant regions.

The organization of spatial attention element was displayed in Figure 3. Initially, the feature maps U $\in \mathbb{R}^{C\times H\times W}$ were passed for aggregation operations that will produce spatial descriptors p $\in \mathbb{R}^{H\times W}$ with aggregation of feature maps in their channel dimensions (C). This produces global distributions of spatial feature as expressed in Eq.1.,

$$p_{hw} = F_{ac}(u_{hw}) = \frac{1}{c}\sum_{i=1}^{C} u_{hw}(i) \tag{1}$$

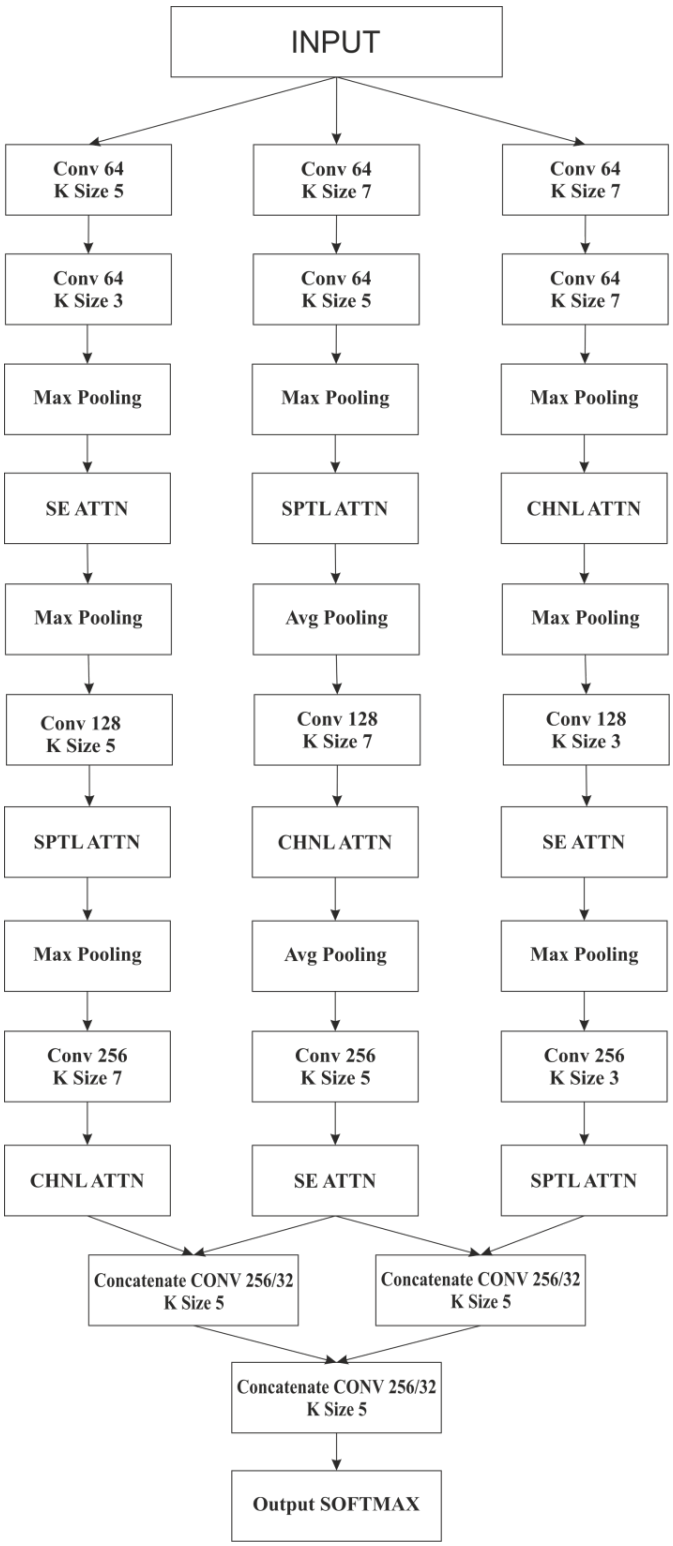

Fig 2. Flow Diagram Representation of the Proposed Algorithm

Here, $u_{hw} \in R^c$ corresponds to the local features corresponding to the spatial positions (h, w).

The aggregate function F_{ac} produces spatial descriptors $\mathbf{p} \in \mathbb{R}^{H \times W}$. Correspondingly, self-learning functions F_l were realized using two convolutional layers and this produces spatial weight maps $\mathbf{t} \in \mathbb{R}^{H \times W}$. Lastly, the function F_{re} employs $\mathbf{t}$ for generating the outputs of the SA unit.

Subsequently, this was followed by weighted self-learning operations and subsequently realized using the convolutional layer. The function $F_l(\mathbf{p}, \mathbf{f})$ targets in entirely capturing the spatial correlations and dynamically produces the spatial weight maps $\mathbf{t} \in \mathbb{R}^{H \times W}$. The assessment was carried out using the mathematical expression below,

$$t = F_1(p, f) = \sigma(g(p, f)) = \sigma(f_2 \delta(f_1 p)) \tag{2}$$

Here, $\mathbf{f}_1$ corresponds to the convolution and was represented as Conv(m), moreover $\mathbf{f}_2$ denotes the convolution and was represented as Conv(1). Also, m represents the number of channels of the hidden feature maps.

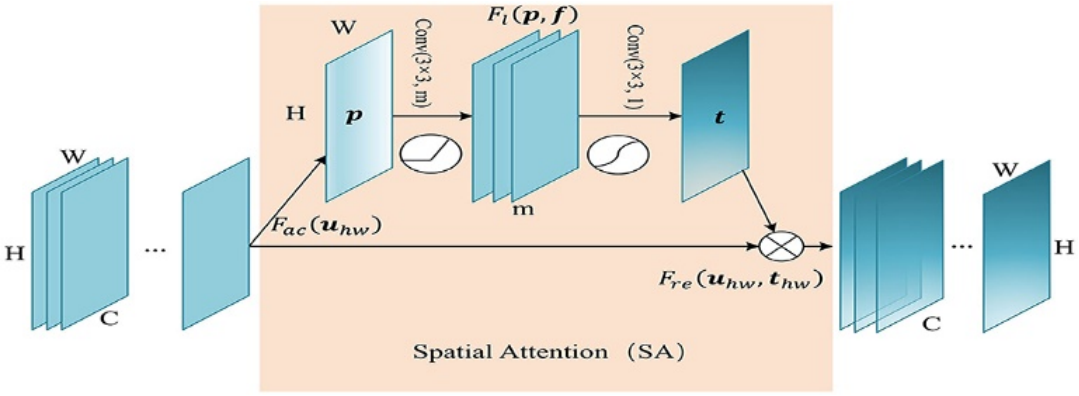

Fig 3. Articulation ofspatial attention (SA) unit

The parameter δ signifies the activation functions ReLU, also the parameter σ corresponds to the sigmoidal activation function which was employed for generating the spatial weights $\mathbf{t}_{hw} \in (0, 1)$ with respect to positions (h, w). However, convolution operations which acquires original spatial descriptors as input was represented as spatial-wise self-attention functions, and these captures non-linear inter-spatial correlation.

The weight calculated at preceding steps was employed at the feature maps $\mathbf{U}$. Using spatial-wise recalibration $F_{re}(\mathbf{u}_{hw}, \mathbf{t}_{hw})$, the feature value of dissimilar positions of $\mathbf{U}$ were multiplied by dissimilar weights for generating the outputs $\mathbf{U}'$ of the SA unit as expressed by,

$$U'_{hw} = F_{re}(U_{hw}, t_{hw}) = U_{hw} \cdot t_{hw} \tag{3}$$

2.3.2 Channel Attention

Similarly, channel attention modules were included in the final layer in encoders, as high-level feature maps mostly express composite features possessing greater receptive fields and additional channels. This procedure allows the networks in performing feature recalibrations, by assuming to take benefit in global information for precisely improving the beneficial features and limits unwanted features. Designing the channel attention unit was represented in Figure 4. Initially, feature maps $U \in \mathbb{R}^{C \times H \times W}$ were passed for performing aggregation operations that produces channel descriptors $q \in \mathbb{R}^C$ by combining feature maps of spatial dimensions (H×W), thereby generating global distributions of channel features as expressed by,

$$q_c = F_{as}(U_c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} W\, U_c(i, j) \tag{4}$$

Here, $U_c \in RH \times W$ corresponds to the local features of channel c.

These aggregate functions F_{as} generate channel descriptors $\mathbf{q} \in \mathbb{R}^C$. Moreover, self-learning functions F_l were realized using two connected layers that produce channel weight maps $\mathbf{v} \in \mathbb{R}^C$. Lastly, the function F_{re} employs $\mathbf{v}$ for generating outputs of CA unit. Moreover, followed by weight self-learning operation, they get realized by fully connected layer. The function $F_l(\mathbf{q}, \mathbf{w})$

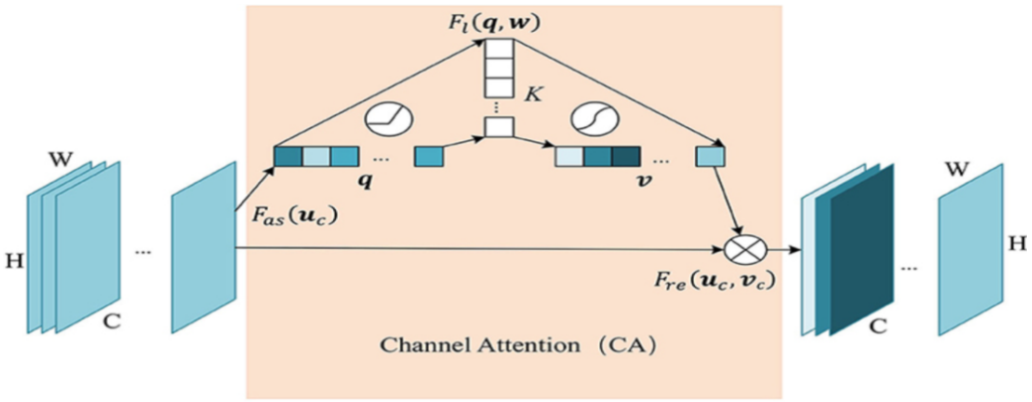

Fig 4. Representation of Channel Attention module

targets in completely capturing the dependencies among channel and dynamically generate channel weight maps $\mathbf{v} \in \mathbb{R}^C$ and the assessment was carried out using the expression below,

$$v = F_l(q,w) = \sigma(g(q,w)) = \sigma\left(w_2 \delta\left(w_1 q\right)\right) \tag{5}$$

Here, $w_1 \in R^{K\times C}$,$w_2 \in R^{C\times K}$. K corresponds to the hidden neurons, σ represents the sigmoid activation functions that were employed for generating the channel weight $\mathbf{v}_c \in (0, 1)$ over channel c. Using fully-connected hidden layer, this will be able in capturing non-linear interactions among the channels. The calculated weight from the preceding stage was passed to the feature maps U. Using channel-wise recalibration $F_{re}(u_c, v_c)$, feature values corresponding to the dissimilar channels at U were multiplied using dissimilar weights for generating the outputs U′ of the CA unit and shall be mathematically expressed as,

$$u'c_c = E_{ve}\left(uc_a v_c\right) = u_c \cdot v_a \tag{6}$$

2.3.3 Squeeze Excitation Module

The channel relations displayed using convolutions were fundamentally understood, excluding those top-most layers. Here, it was expected that the learning of convolutional features must be enhanced using explicit modelled channel interdependencies, thereby the network can build its accessibility to information features which can be taken advantage of ensuing changes. Hence, it might be required for accessing to the global information and recalibrating the filter's responses using two phases, squeeze and excitation [22–24] , before passing them to consequent transformations. Here, we have proposed the model for getting global spatial information within the channel descriptors, and was attained using the global average pooling for producing channel-wise indicators. Strictly, the statistics $z \in R\,C$ was produced using the reduction of U over its spatial dimension $H \times W$, thereby the c^{th} elements in z was assessed using the expression below,

$$z_c = F_{sq}\left(u_c\right) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} u_c(i,j) \tag{7}$$

For making usage of aggregated data at squeeze operations, additional operation was executed that aims in entirely capturing channel-wise dependency. For satisfying these objectives, the functions should meet two conditions: (1) it should possess better flexibility (specially, they have to be prepared in learning nonlinear interactions among channels), (2) it must possess expertise towards non-mutually-exclusive relationships as it might be required in making sure that several channels were allowed to be emphasized (other than realizing single activation functions). For meeting these constraints, elementary gating mechanism was employed together with sigmoid activations and shall be expressed mathematically as,

$$s = F_{ex}(z,W) = \sigma(g(z,W)) = \sigma\left(W_{2\delta}\left(W_1 z\right)\right) \tag{8}$$

Here, δ corresponds to ReLU functions $W_1 \in R^{\frac{C}{r}\times c}$ and W2 $\in R^{c\times\frac{C}{r}}$. For restricting the model complexities, gating mechanism was parameterized using association amid two fully-connected (FC) layers over the non-linearity. The finalized output from these blocks were attained by realization of U using the activations s and shall be expressed as,

$$\widetilde{x_c} = F_{re}\left(u_c, s_c\right) = u_c s_c \tag{9}$$

Here, $\widetilde{X} = [\widetilde{x}_1, \widetilde{x}_2, \ldots\ldots, \widetilde{x}_c]$ and $F_{re}(u_c, s_c)$ signifies the channel-wise multiplication among scalar sc together with feature maps $u_c \in R^{H \times W}$.

3 Results and Discussions

Tables 1 and 2 show the accuracy of the proposed Fusion Attention algorithm and its assessment with the existing methods like Res-Target[10], VGG16[15], Resnet50[17] and Alexnet[20]. Considering the existing VGG-16 networks, it possesses 13 convolutional layers as well as 3 fully connected layers. At convolutional layers, it includes 13 ReLU layers as well as 4 pooling layers. But, this proposed work specifies a hybrid approach in which the image together with convolution passes through channel attention, spatial attention as well as squeeze and excitation for attaining higher accuracy.

AlexNet comprises of 8 layers: that includes 5 convolutional layers, 2 fully-connected hidden layers, as well as a single fully-connected output layer. Moreover, AlexNet makes use of ReLU as an alternative to sigmoid as the activation functions. In case of this proposed method, the algorithm performs convolution in 64-bit, 128-bit and 256-bit respectively, in which the three attentions were interchanged in each convolution along with max and average pooling. The image was first passed to three different levels of 64-bit convolution followed by pooling, and then each level possesses three attentions respectively. Later, the process continues identically for 128-bit and 256-bit convolutions by interchanging the levels of attention.

Table 1. Performanceof the Proposed Algorithm with Existing Methods

Algorithm	Accuracy	Precision	Recall	F1 Score
Res-Target	0.598835	0.535344	0.419258	0.464804
Resnet50	0.600629	0.20021	0.333333	0.250164
Alexnet	0.592767	0.545455	0.459283	0.474129
VGG16	0.628931	0.209644	0.333333	0.2574
Fusion Attention	0.641509	0.560896	0.472526	0.512933

Table 2. Performance in % of Proposed and Existing Algorithms

Algorithm	Accuracy
Res-Target	59.88%
Resnet50	60.06%
Alexnet	59.28%
VGG16	62.89%
Fusion Attention	64.15%

Table 3. Classification Results of the Twitter15 images

	Positive	Negative	Neutral	Total
Train	928	368	1883	3179
Dev	303	149	670	1122

The novel feature of this proposed fusion attention algorithm was not only limited to the features of prevailing channel attention models, but also an enhanced methodology towards spatial attention i.e., in the image where the area of focus, and also provides a clear understanding of images along with the squeeze and excitation. From this comparison, it was clear that this proposed model was further equipped based on the custom datasets and this provided improved accuracy over the other two pre-trained models by an accuracy level of 64.15% and F1 Score of 51.29%. The classification of the input image of Twitter15 images was provided in Table 3 and in Figure 5. The proposed system was compared in terms of accuracy, precision, recall, F1 Score.

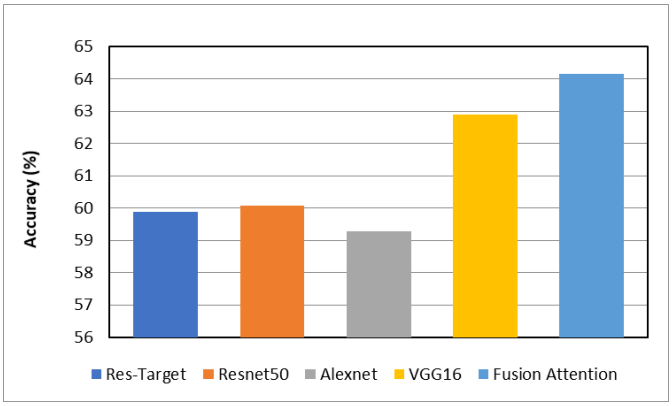

Fig 5. Performance of the Proposed Fusion Attention Algorithm with ExistingAlgorithms

4 Conclusion and Future works

This study presents a novel fusion attention algorithm for image sentiment analysis. It possesses comprehensive advantages of the characteristic features of CNN for yielding enhanced image features, the spatial, channel-wise, and SE. The main feature of this proposed fusion attention algorithm is not only limited with more dominant attention model, but also this was seen to be an improved understanding methodology towards Spatial attention i.e., in the image where the area of focus, and Channel attention i.e., what provide a clear view in the image along with the Squeeze and Excitation looking like in a CNN that evolves during image generation. The simulation results showed that, this proposed model provided enhanced accuracy when compared with other two pre-trained models by an accuracy level of 64.15% and F1 Score of 51.29%. This proposed approach similarly provides more clarity on the essential features by eliminating unwanted features of an image. As a future direction, temporal attention over SCA-CNN shall be incorporated for combining the features of dissimilar video frames for effective video captioning. Moreover, analysis shall be carried out on investigating by increasing the attentive layer count deprived of overfitting.

References

1) Ou H, Qing C, Xu X, Jin J. Multi-Level Context Pyramid Network for Visual Sentiment Analysis. *Sensors*. 2021;21(6):2136–2136. Available from: https://dx.doi.org/10.3390/s21062136.
2) Yan W, Zhou L, Qian Z, Xiao L, Zhu H. Sentiment Analysis of Student Texts Using the CNN-BiGRU-AT Model. *Scientific Programming*. 2021;2021:1–9. Available from: https://dx.doi.org/10.1155/2021/8405623.
3) Asakawa T, Aono M. Multi-label Prediction for Visual Sentiment Analysis using Eight Different Emotions based on Psychology. In: 2021 4th International Conference on Control and Computer Vision. ACM. 2021;p. 142–146. Available from: https://doi.org/10.1145/3484274.3484296.
4) Jamatia A, Swamy SD, Gambäck B, Das A, Debbarma S. Deep Learning Based Sentiment Analysis in a Code-Mixed English-Hindi and English-Bengali Social Media Corpus. *International Journal on Artificial Intelligence Tools*. 2020;29(05):2050014–2050014. Available from: https://dx.doi.org/10.1142/s0218213020500141.
5) Chen J, Mao Q, Xue L. Visual Sentiment Analysis With Active Learning. *IEEE Access*. 2020;8(1):185899–185908. Available from: https://dx.doi.org/10.1109/access.2020.3024948.
6) Ortis A, Farinella GM, Battiato S. A Survey on Visual Sentiment Analysis. . *IET Image Processing*. 2020. Available from: https://doi.org/10.1049/iet-ipr.2019.1270.
7) Balakrishnan V, Shi Z, Law CL, Lim R, Teh LL, Fan Y. A deep learning approach in predicting products' sentiment ratings: a comparative analysis. *The Journal of Supercomputing*. 2021. Available from: https://dx.doi.org/10.1007/s11227-021-04169-6.
8) Li B, Ren H, Jiang X, Miao F, Feng F, Jin L. SCEP—A New Image Dimensional Emotion Recognition Model Based on Spatial and Channel-Wise Attention Mechanisms. *IEEE Access*. 2021;9:25278–25290. Available from: https://dx.doi.org/10.1109/access.2021.3057373.
9) Chellam GH, Roseline V. Analysis of sentimental images using deep learning approach. *Journal of Mathematical and Computational Science*. 2021;11(5):5474–5486. Available from: https://doi.org/10.28919/jmcs/5835.
10) Kabiito D, Nakatumba-Nabende J. Targeted Aspect-Based Sentiment Analysis for Ugandan Telecom Reviews from Twitter. *Transactions on Computational Science and Computational Intelligence*. 2021;p. 311–322. Available from: https://doi.org/10.1007/978-3-030-70296-0_24.
11) Jiang T, juan Hu X, hua Yao X, ping Tu L, bin Huang J, xiang Ma X, et al. Tongue image quality assessment based on a deep convolutional neural network. *BMC Medical Informatics and Decision Making*. 2021;21(1):147–147. Available from: https://dx.doi.org/10.1186/s12911-021-01508-8.
12) Mhamed M, Sutcliffe R, Sun X, Feng J, Almekhlafi E, Retta EA. Improving Arabic Sentiment Analysis Using CNN-Based Architectures and Text Preprocessing. *Computational Intelligence and Neuroscience*. 2021;2021:1–12. Available from: https://dx.doi.org/10.1155/2021/5538791.
13) Li Z, Zhou A. Self-Selection Salient Region-Based Scene Recognition Using Slight-Weight Convolutional Neural Network. *Journal of Intelligent & Robotic Systems*. 2021;102(3). Available from: https://dx.doi.org/10.1007/s10846-021-01421-2.
14) Wu BX, Yang CG, Zhong JP. Research on Transfer Learning of Vision-based Gesture Recognition. *International Journal of Automation and Computing*. 2021;18(3):422–431. Available from: https://dx.doi.org/10.1007/s11633-020-1273-9.

15) Abdelgwad M, Soliman A, Taloba A, Farghaly MF. Arabic aspect based sentiment analysis using bidirectional GRU based models. *Journal of King Saud University - Computer and Information Sciences*. 2021. Available from: https://dx.doi.org/10.1016/j.jksuci.2021.08.030.
16) Tuan TQ, Hady WL, Martin A, Naoko N. Visual Sentiment Analysis for Review Images with Item-Oriented and User-Oriented CNN: Reproducibility Companion Paper. *Proceedings of the 28th ACM International Conference on Multimedia*. 2020;p. 4444–4447. Available from: https://ink.library.smu.edu.sg/sis_research/5955.
17) Xia K, Zhou Q, Jiang Y, Chen B, Gu X. Deep residual neural network based image enhancement algorithm for low dose CT images. *Multimedia Tools and Applications*. 2021. Available from: https://dx.doi.org/10.1007/s11042-021-11024-6.
18) Nguyen G, Dlugolinsky S, Bobák M, Tran V, Álvaro López García, Heredia I, et al. Machine Learning and Deep Learning frameworks and libraries for large-scale data mining: a survey. *Artificial Intelligence Review*. 2019;52(1):77–124. Available from: https://dx.doi.org/10.1007/s10462-018-09679-z.
19) Scheidegger F, Istrate R, Mariani G, Benini L, Bekas C, Malossi C. Efficient image dataset classification difficulty estimation for predicting deep-learning accuracy. *The Visual Computer*. 2021;37(6):1593–1610. Available from: https://dx.doi.org/10.1007/s00371-020-01922-5.
20) Sathish R, Ezhumalai P. Enhanced sentimental analysis using visual geometry group network-based deep learning approach. *Soft Computing*. 2021;25(16):11235–11243. Available from: https://dx.doi.org/10.1007/s00500-021-05890-3.
21) Das S, Kolya AK. Predicting the pandemic: sentiment evaluation and predictive analysis from large-scale tweets on Covid-19 by deep convolutional neural network. *Evolutionary Intelligence*. 2021. Available from: https://dx.doi.org/10.1007/s12065-021-00598-7.
22) Yu J, Jiang J, Xia R. Entity-Sensitive Attention and Fusion Network for Entity-Level Multimodal Sentiment Classification. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. 2020;28:429–439. Available from: https://dx.doi.org/10.1109/taslp.2019.2957872.
23) Zhao P, Zhang J, Fang W, Deng S. SCAU-Net: Spatial-Channel Attention U-Net for Gland Segmentation. *Frontiers in Bioengineering and Biotechnology*. 2020;8. Available from: https://dx.doi.org/10.3389/fbioe.2020.00670.
24) Alzubaidi L, Zhang J, Humaidi AJ, Al-Dujaili A, Duan Y, Al-Shamma O, et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*. 2021;8(1). Available from: https://dx.doi.org/10.1186/s40537-021-00444-8.