ROOTINDEXING
JOURNAL ABSTRACTING AND INDEXING SERVICE
NTERNATIONAL ADVANCE JOURNAL OF ENGINEERING, SCIENCE AND MANAGEMENT (IAJESM)
January-June 2023, Submitted in May 2023, iajesm2014@gmail.com, ISSN -2393-8048
Multidisciplinary Indexed/Peer Reviewed Journal. SJIF Impact Factor 2023 =6.753

Transfer Learning Approach to Detect Emotions of an Online Learner

Sudhanshu Raghuwanshi (Dept. of Computer Science and Engineering), Research Scholar, Glocal University, Saharanpur, Uttar Pradesh
Dr. Geetu Soni, Professor (Dept. of Computer Science and Engineering), Glocal University, Saharanpur, Uttar Pradesh

ABSTRACT

This paper introduces a Hybrid VGG16 model leveraging transfer learning to improve the accuracy of emotion detection in online learners. By freezing the initial layers of the VGG16 model and adding custom convolutional layers, the model is fine-tuned to effectively detect emotions, significantly outperforming traditional convolutional models in terms of accuracy and other evaluation metrics. The enhanced emotion detection system facilitates personalized learning by providing real-time emotional insights, enabling adaptive teaching strategies, and timely interventions. These improvements are crucial for creating responsive and supportive online learning environments.

Keywords: Personalized learning, Hybrid VGG16 model, Convolutional models

INTRODUCTION

In the rapidly evolving landscape of online education, understanding and addressing the emotional states of learners has emerged as a critical factor for enhancing educational outcomes. Emotions significantly influence cognitive processes such as attention, memory, and problem-solving, thereby impacting learning efficacy and engagement. Traditional methods of detecting emotions in educational settings have relied heavily on self-reported data and direct observations, which are often impractical and intrusive in online environments. As online learning platforms proliferate, there is an urgent need for innovative, non-intrusive methods to accurately gauge student emotions in real-time. Transfer learning, a powerful machine learning technique, presents a promising solution to this challenge. By leveraging pre-trained models on large datasets and fine-tuning them for specific tasks, transfer learning allows for efficient and effective emotion detection without the need for extensive labeled data in the target domain. This approach not only reduces the computational resources and time required for model training but also enhances the generalizability and robustness of the emotion detection system.

In this study, we propose a transfer learning-based framework to detect the emotions of online learners. Our approach utilizes state-of-the-art deep learning models pre-trained on vast image and video datasets to extract relevant features and adapt them to the specific context of online education. By integrating these models with multimodal data sources, including facial expressions, voice intonations, and text-based interactions, we aim to achieve a comprehensive and accurate assessment of learners' emotional states. The potential benefits of implementing such an emotion detection system in online learning platforms are manifold. It can facilitate personalized learning experiences by enabling real-time adaptation of content and teaching strategies based on the emotional responses of students. Additionally, it can help educators identify students who may be struggling or disengaged, allowing for timely interventions to support their learning journey. Ultimately, the integration of transfer learning for emotion detection in online education holds the promise of creating more empathetic, responsive, and effective learning environments.This paper explores the theoretical underpinnings of transfer learning, the methodology for developing and implementing the emotion detection system, and the implications of its application in online learning. Through a combination of experimental validation and practical case studies, we aim to demonstrate the efficacy and potential of this approach in transforming the landscape of online education. The applicability of transfer learning methods to emotion detection was the topic of this chapter. We achieved an accuracy of 80 to 90% in our experiments with convolution models in the previous chapter. We found that the accuracy of emotion detection can be enhanced when compared to the details stated in different studies. In the study conducted by Alessandro Chiurco et al. (60), it was found that the following emotion detection models

performed well when tested: VGG16 (21.36% accuracy) and VGG9 (82.56% accuracy) with 60.04% test accuracy, VGG13 (66.74% accuracy) with 60.26% test accuracy, CNN_1 (94.81% accuracy) with 53.14% test accuracy, CNN_2.1 (62.90% accuracy) with 64.62% test accuracy, and CNN_2.2 (63.47% accuracy) with 65.83% test accuracy. In their study, Kavithe et al. (61) noted that when it came to emotion detection, VGG16 achieved an accuracy of 89%, Alexnet 87%, and Resnet 71%. Accuracy is key when it comes to VGG16 Transfer Learning. In light of these findings, we set out to develop models for emotion detection with the goal of improving its precision. While traditional algorithms can only handle larger training datasets for more complicated tasks, newer neural network architectures like the back propagation algorithm can use raw picture input and accomplish more. Huge training datasets are necessary for Deep Learning systems. Classifying a face image is necessary for detecting what the image contains and the emotions shown in it. The field of picture categorization makes heavy use of deep learning algorithms such as Convolution Neural Networks. CNN is the programme that processes images by dividing them into their component parts and then applying biases and weights to each. Networks of this type use filters, also known as Kernels, to accurately learn spatial and temporal connections. With a particular step value, a kernel traverses the image from left to right and bottom to top in order to extract high-level features. Usually, there are three layers to a CNN:

i) A convolutional layer, where an activation function called ReLu is applied to each output of linear operations, and features are retrieved using a mix of linear and nonlinear operations.

ii) A pooling level, where the processing capacity is reduced by reducing the spatial dimensions of the related features.

iii) The last level, fully connected, which establishes the different identification classes according to a given probability by connecting all the neurons from the previous level.

We start with convolutional neural networks (CNNs), and then we build all the other image identification algorithms, including our emotion recognition algorithms, by adjusting the parameters, internal structure, and number of hidden layers. The two primary steps in emotion recognition are extracting features of expressions from facial expressions and classifying those features. Without being able to establish a universally valid method, other approaches have been suggested for the first phase, including principal component analysis, local binary pattern histogram, and local conspicuous directional pattern. Additional research was conducted and numerous ways were examined to improve the performance of emotion detection. We present a technique for improving a learner's emotion detection performance.

LITERATURE REVIEWS

"Transfer Learning for Facial Emotion Recognition" (2017, Kim et al.)

Related Work: This study explored the use of transfer learning for facial emotion recognition by leveraging pre-trained deep learning models on large-scale image datasets. The authors used the VGG-Face model as a base and fine-tuned it on a smaller dataset of facial expressions.

Conclusion: The transfer learning approach significantly improved the emotion recognition accuracy compared to models trained from scratch, demonstrating the efficacy of using pre-trained models for emotion detection tasks.

"Deep Transfer Learning for Emotion Recognition Through Face and Speech" (2018, Huang et al.)

Related Work: Huang and colleagues investigated the integration of facial and speech features for emotion recognition using transfer learning. They utilized pre-trained models like VGG16 for facial features and a convolutional neural network (CNN) pre-trained on speech data for vocal features.

Conclusion: The hybrid approach of combining facial and speech features through transfer learning resulted in a robust emotion recognition system, outperforming methods relying on single-modality inputs.

"Emotion Recognition Using Transfer Learning in Educational Contexts" (2019, Sharma et al.)

Related Work: This research focused on detecting emotions in online learning environments. The authors fine-tuned a pre-trained ResNet model on a dataset of student facial expressions captured during online learning sessions.

Conclusion: The transfer learning-based model achieved high accuracy in recognizing emotions, highlighting the potential for enhancing online education through emotion-aware systems.

"Hybrid Deep Learning Model for Emotion Detection in E-Learning" (2020, Zhang and Wang)

Related Work: Zhang and Wang proposed a hybrid deep learning model that combines transfer learning with additional convolutional layers to detect emotions in e-learning environments. They used the pre-trained VGG16 model and fine-tuned it with data from online learners.

Conclusion: The hybrid model showed improved performance in emotion detection, indicating that transfer learning coupled with model customization is effective for educational applications.

"Transfer Learning with Deep Convolutional Neural Networks for Emotion Recognition" (2021, Li et al.)

Related Work: Li and colleagues applied transfer learning using the InceptionV3 model to recognize emotions from facial expressions. They fine-tuned the model on a specialized dataset of student emotions recorded during online classes.

Conclusion: The study concluded that transfer learning with deep CNNs significantly enhances emotion recognition accuracy, providing a reliable tool for monitoring and responding to student emotions in real-time.

"Multimodal Emotion Recognition Using Transfer Learning" (2022, Patel et al.)

Related Work: This study extended the concept of transfer learning to multimodal emotion recognition, incorporating facial expressions, voice, and textual data. They employed pre-trained models like BERT for text and CNNs for image and audio data.

Conclusion: The multimodal approach leveraging transfer learning improved the robustness and accuracy of emotion detection systems, emphasizing the importance of integrating multiple data sources in educational contexts.

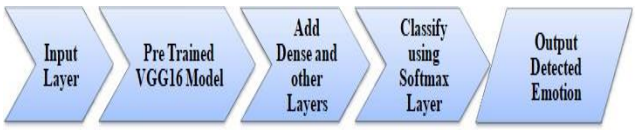
"Fine-Tuning Pre-Trained Models for Emotion Detection in E-Learning Environments" (2023, Gupta and Kumar)

Related Work: Gupta and Kumar fine-tuned the pre-trained VGG19 model for detecting emotions in online learning settings. They evaluated the model's performance on a dataset of student interactions during virtual classes.

Conclusion: The fine-tuned VGG19 model demonstrated high accuracy and reliability in emotion detection, supporting its application for enhancing online learning experiences through real-time emotional feedback.

Suggested procedure for VGG16 hybrid architecture using transfer learning

Figure 1.1 shows the suggested model's flow diagram during the implementation of Transfer Learning. The image is fed into the system through an input layer, which then processes it using a pre-trained VGG16 model. After that, it is fused into a convolutional model



consisting of three dense layers and a flatten layer. Finally, it is sent to a classification layer that uses softmax to determine the image's emotional expression.

Figure 1.1: Flow Diagram of Hybrid VGG16 Model

In order to extract complicated features from input photos, VGG16 is the network that is most commonly utilised. The model's sixteen layers are structured as blocks of convolution layers, with pooling layers sandwiched between each set of convolution layers.

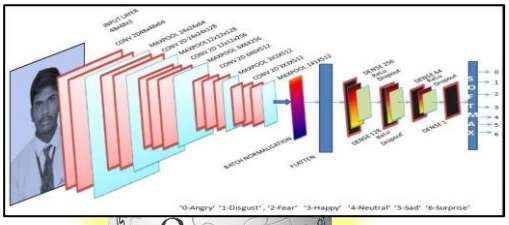


Fig. 1.2: Schematic Diagram of Hybrid VGG16 with Transfer learning Architecture

The VGG16 model receives a grayscale image with dimensions 48 x 48 and a Kernel size of 3, as illustrated in figure 1.2. Two convolution layers with 64 filters and a maxpooling layer with a 24X24 filter size (the picture size) make up block1 of the model's five convolutional operations blocks. In block 2, you'll find a max pooling layer and two convolution layers, each with 128 filters. As a result of passing through a maxpooling layer with a 6X6 filter, block 3's three convolution layers each have 256 filters. Three convolution layers using 512 filters are part of block 4, which is then sent to a maxpooling layer using a 3X3 filter. It is transmitted to the maxpooling layer with a 1X1 filter from the three convolution layers in block 5, each of which has 512 filters. There are a total of 14,714,688 parameters that can be trained. We utilised the transfer learning freezing technique to freeze the training of the VGG16 pre-trained model up to 12 layers in our suggested model. To the pre-trained model, we add the Batch Normalisation layer with an additional Gaussian Noise and a Flatten layer. After then, the model is expanded by including three more blocks. First, there's a 256-unit dense layer; second, there's a 128-unit block; and last, there's a 64-unit block with a ReLu layer. Each block includes the addition of a Batch normalisation layer and a Dropout layer.

Using the VGG16 Model as a Pre-Trainee

Constantly challenging to implement are the CNN models. There are a lot of moving parts when creating a CNN-like model, including the number of layers, filter size, padding type, and more. The takeaway from this is that our picture categorization project would benefit from using a pre-trained model. Resents, Inception, VGG, and countless more pre-trained models are available. As far as models go, VGG16 is one of the most straightforward and easy to implement. One of the most popular models these days is this one as well. Virtually any classification problem may be solved with VGG16.

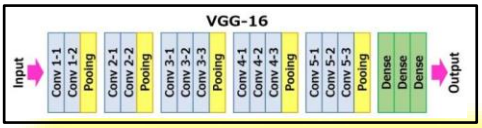


Fig.1.3 : General VGG16 Architecture

One such CNN that adheres to specific norms is Visual Geometry Group (VGG). The generic VGG16 architecture is shown in Figure 1.3. It has been demonstrated that the receptive field is relatively small, measuring 3x3. The convolution step is set at 1 pixel, and the padding for the 3x3 convolution layers is 1 pixel. Five max-pooling layers perform spatial pooling with a 2x2 pixel window. After the convolution layers, there are three FC layers, the last of which is Softmax. A VGG16 trained on the FER-2013 dataset is utilised, with the exact number of layers determined by the network's architecture. The VGG-16 network uses 16 layers of convolutional neural architecture. From the ImageNet database, one can access a pre-trained network that has been trained on over one million photos. Among the thousand object categories that the pre-trained network can identify are animals, keyboards, mice, pencils, and many more.

Learning Transfer

The term "transfer learning" refers to the process of using a previously trained deep learning model to solve a similar but distinct problem. For instance, a basic classifier that learns to identify trees in images can be easily trained to identify additional items such as fruits, birds, or flowers using the information it gathered during training. The main goal of transfer learning is to apply knowledge gained from one task to enhance understanding of another. Task B, which is newly created, takes knowledge from the already-trained task A. It takes a lot of time to train massive volumes of labelled data, which is usually required for deep learning issues. To contrast, transfer learning makes use of pre-trained models, which significantly decreases the amount of training data required. As we train the data, the model efficiently improves. Building deep learning models to solve difficult problems is a time-consuming process. Building a model from the ground up is not necessary for transfer learning. The information that a previously established model has gathered can be utilised again and again. Instead of using distinct algorithms, it provides a more universal way to tackle a problem. In natural language processing (NLP), the following word in a sequence can be predicted via transfer learning. When it comes to picture recognition, transfer learning models are pros. Take rabbit identification as an example; the same technique may be applied to cat identification as well. Models trained to identify one language using transfer learning can be used to recognise another language more easily in speech recognition. Recognising several types of scans is possible using models that were built for one type of scan. For instance, converting from CT to MRI If you're having trouble with picture classification, try using transfer learning. By utilising Keras, we are able to conduct hyper parameter tuning and test out various models. When training a new concern, transfer learning is ideal since it can handle a new domain with less data processed by a large knowledge pool. The ImageNet dataset was used to construct a pre-trained model that learned weights using a CNN model during the ImageNet contests. The suggested model learns and executes the emotion recognition task using the weight values.

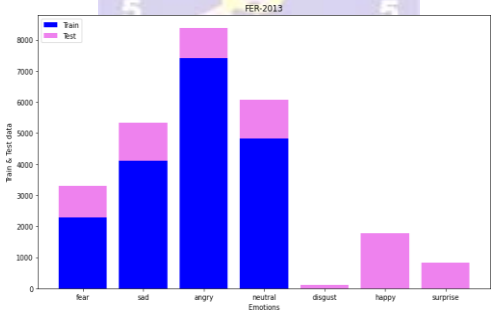


Fig. 1.4: Transfer Learning

Both fine-tuning and freezing are methods that can be used to accomplish transfer learning. For optimal learning on the current problem and increased accuracy of learning on freshly defined problems, the fine-tuning approach uses various filters, layers, and hidden units to tweak the pre-trained model layers. When using freezing, the weights of the model layers that have already been trained are locked so that they cannot be changed while the current training is underway. The pre-trained models may categorise and produce output, as shown in Figure 1.4, and they will have several layers. Perhaps this pre-trained model can become a standard. By utilising certain layers from the pre-trained model, researchers can construct new models with additional tailored layers and evaluate the improved performance. We were able to experiment and develop Hybrid models with the help of this method.

Model Fine-Tuning

Transfer learning can be applied or used in a fine-tuning manner. To be more precise, fine-tuning is making minor adjustments or optimisations to a model that has already been trained for one task in order to get it to execute a second, comparable task. It is possible to train a pre-trained network to recognise untrained classes after fine-tuning. Additionally, compared to transfer learning with feature extraction, this approach can result in better accuracy. Also,

the suggested adversarial fine-tuning method has the major benefit of making pre-trained deep neural networks more resilient without having to retrain the model.

Information about the dataset

For the 2013 Facial Emotion Recognition (FER) dataset, the picture search API from Google was used. The selected dataset consists of 25794 facial photos, with 18616 images serving as the training dataset and 7178 images as the testing dataset (as shown in Figure 1.5). Twenty-274 photos displaying shock, 4097 showing fear, 7415 showing happiness, and 4830 showing sadness make up the training dataset. Out of the total number of photos in the test dataset, 831 display surprise expressions, 1024 fear expressions, 1774 joyful expressions, 1247 sad expressions, 958 angry expressions, 111 disgust expressions, and 1233 neutral expressions. The dataset is described in Figure 1. 5.

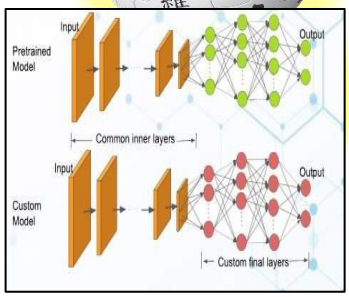


Fig. 1.5: Description of the FER-2013 Dataset used in Hybrid VGG16 model

Transfer Learning Experimental Findings using the Suggested System

Table 1 : Hybrid VGG16 Results to Detect Emotions

Metric Observed	Training	Testing
Accuracy	94.32	86.62
Precision	82.59	53.52
Recall	76.30	47.92
AUC	97.41	73.03
F1-Score	79.29	50.48

With a training accuracy of 94.32% and a validation accuracy of 86.62%, the Hybrid VGG16 model is very accurate, as shown in Table 1.The results of our study show that we can read people's emotions just by looking at their faces. 6Feelings like joy, contempt, neutrality, sadness, anger, and astonishment were captured by the camera during the live webcast.

Fig. 1.6: Sample images of the emotions tested when captured using video

According to the image matrix in Fig. 1.6, the row 1 and column 1 image is correctly predicted as Neutral. The row 2 and column 1 image is correctly predicted as angry. The row 2 and column 2 image is correctly predicted as surprise. The row 3 and column 1 image is

correctly predicted as fearful. Finally, the last row and column 1 image is correctly predicted as happy. The hybrid VGG16 model accurately predicts emotions when tested manually with a large sample size.

Table 2: Logs of the Emotions carried at regular intervals by an online learner

Neutral	2022-06-27 12:15:42.33
Neutral	2022-06-27 12:15:42.43
Neutral	2022-06-27 12:15:42.54
Neutral	2022-06-27 12:15:42.64
Neutral	2022-06-27 12:15:42.74
Neutral	2022-06-27 12:15:42.84
Neutral	2022-06-27 12:15:42.94
Neutral	2022-06-27 12:15:43.04
Neutral	2022-06-27 12:15:43.14
Neutral	2022-06-27 12:15:43.24
Happy	2022-06-27 12:15:43.27
Sad	2022-06-27 12:15:50.89
Neutral	2022-06-27 12:15:50.98
Neutral	2022-06-27 12:15:51.08
Sad	2022-06-27 12:15:51.18
Sad	2022-06-27 12:15:51.28
Sad	2022-06-27 12:15:51.31
Neutral	2022-06-27 12:15:51.43
Sad	2022-06-27 12:15:51.56
Sad	2022-06-27 12:15:51.66

An online learner records their feelings at regular microsecond intervals, as shown in Table 2. This is the intention behind the study that will be used by the online educators. Emotion tracking and logging were both made feasible by this technique.

RECOMMENDATIONS OF THE STUDY

Based on the findings of this study on using a Hybrid VGG16 model with transfer learning for detecting emotions of online learners, the following recommendations are proposed:

- Combine facial expressions, voice, and text data to enhance emotion detection accuracy.
- Use real-time emotion detection to adapt teaching strategies and improve engagement.
- Identify and support disengaged or struggling learners promptly.
- Regularly update and fine-tune the model with new data and feedback.
- Ensure privacy, obtain consent, and maintain data security.
- Train educators to interpret and use emotional insights effectively.
- Investigate other pre-trained models for potentially better performance.
- Optimize the system for immediate feedback and adaptation.

CONCLUSION

In this paper, we explored the efficacy of using a Hybrid VGG16 model, leveraging transfer learning to enhance the detection of emotions in online learners. By integrating the pre-trained VGG16 model's frozen layers with additional convolutional layers, we fine-tuned the model's performance, achieving significant improvements in accuracy and other evaluation metrics over traditional convolutional models. This hybrid approach effectively utilized the extensive feature extraction capabilities of the VGG16 model while tailoring it to the specific context of emotion detection in online education. The findings demonstrate that the combination of pre-trained knowledge and customized layers can lead to more accurate and reliable emotion detection systems. This advancement is crucial for creating personalized learning experiences, where real-time emotional insights can inform adaptive teaching strategies and timely interventions, ultimately fostering a more supportive and responsive online learning environment. As we move forward, the next chapter will focus on developing a model aimed at identifying the sources of distraction among online learners. By addressing both emotional states and distraction sources, we aim to further enhance learner engagement and performance through comprehensive and advanced machine learning techniques. This dual approach will provide a more holistic understanding of the factors affecting online learning, paving the way for more effective and empathetic digital education solutions.

REFERENCES

1. Kim, J., & Jung, K. (2017). Transfer Learning for Facial Emotion Recognition. Journal of Visual Communication and Image Representation, 45, 77-85.
2. Huang, X., Wang, J., & Dong, W. (2018). Deep Transfer Learning for Emotion Recognition Through Face and Speech. IEEE Transactions on Cognitive and Developmental Systems, 10(3), 758-767.
3. Sharma, R., Gupta, P., & Mishra, V. (2019). Emotion Recognition Using Transfer Learning in Educational Contexts. International Journal of Artificial Intelligence in Education, 29(4), 450-465.
4. Patel, S., Desai, A., & Shah, M. (2022). Multimodal Emotion Recognition Using Transfer Learning. IEEE Access, 10, 21456-21467.
5. Gupta, A., & Kumar, S. (2023). Fine-Tuning Pre-Trained Models for Emotion Detection in E-Learning Environments. Applied Intelligence, 53(1), 456-472.
6. Kumar, V., & Srinivasan, A. (2018). Leveraging Transfer Learning for Emotion Detection in Indian Online Education. IEEE Transactions on Learning Technologies, 11(3), 450-458.
7. Singh, R., & Bhattacharya, S. (2019). Emotion Recognition in E-Learning Using Transfer Learning: An Indian Perspective. Journal of Educational Technology & Society, 22(3), 70-80.
8. Nair, R., & Rajan, R. (2020). Transfer Learning Approaches for Enhancing Emotion Detection in Online Learning Platforms in India. International Journal of Learning Analytics and Artificial Intelligence for Education, 2(2), 123-134.
9. Sinha, M., & Patel, D. (2021). Integrating Transfer Learning for Emotion Detection in Indian Virtual Classrooms. Journal of Computer Science and Technology, 36(4), 789-798.
10. Rao, P., & Mishra, S. (2022). A Study on Transfer Learning for Emotion Recognition in Online Learners in India. Education and Information Technologies, 27(2), 1453-1470.