

#### Abstract:

Affective state recognition from facial images is a challenging task in computer vision, with potential applications in various fields such as psychology, education, and human-computer interaction. In this study, we aimed to reproduce the results of the paper "Detecting Affect States Using VGG16, ResNet50 and SE ResNet50 Networks" and conduct a sensitivity analysis on the hyperparameters of the models. We will use the same dataset, AffectNet, and the three deep learning models, VGG16, ResNet50, and SE-ResNet50, to train and test our models

#### Introduction:

Affective state recognition is a crucial aspect of human communication and interaction, as emotions play a significant role in our daily lives. In recent years, computer vision and deep learning techniques have shown great potential in recognizing emotions from facial expressions. The study of "Detecting Affect States Using VGG16, ResNet50 and SE ResNet50 Networks" proposed a method for emotion recognition using three deep learning models, VGG16, ResNet50, and SE-ResNet50, and achieved promising results on the AffectNet dataset. However, the study did not investigate the sensitivity of the models to various hyperparameters, which can greatly affect the performance of the models. In this study, we aimed to reproduce the results of the original study and conduct a sensitivity analysis on the hyperparameters of the models. Our study can provide insights into the impact of hyperparameters on affective state recognition and guide the development of more effective models for this task.

#### Related Works:

Several studies have been published that focus on automatically detecting affective states. Most of these studies use systematic measurements created by Ekman and Friesen (1978), which have become the standard for subsequent affect analysis research.

Mahbub et al. (2018) introduce a new configuration called SPLITFACE, a deep convolutional neural network-based method that performs attribute detection in partially occluded faces. Several segments of the face are taken, and the authors identify which attributes are localized in which part of the face. The authors successfully demonstrate that their method outperforms other recent methods.

Although large convolutional network models have demonstrated good classification performance on the ImageNet benchmark, there is still a lack of understanding of why they perform so well. A paper, Zeiler and Fergus (2014) discusses a novel visualization technique that gives insight into the function of intermediate feature layers and the operation of the classifier.

VGG-16 is a deep CNN network published in 2015. The Visual Geometry Group (VGG) explores the effect of increasing the depth of the convolutional network on its accuracy, using architecture with small convolution filters of size  $3 \times 3$ , which show significant improvement compared to state-of-the-art configurations. These findings were submitted in the ImageNet Challenge 2014.

Squeeze-and-excitation networks are a recent advancement in deep learning. A new architectural unit called the squeeze-and-excitation (SE) block is introduced by Hu et al. (2017) to improve the quality of representations produced by a network. The SE block models interdependencies between the channels of its convolutional features and can be integrated into standard architectures such as VGG, ResNet, and Inception by stacking a collection of SE blocks.

#### Method/Algorithm:

The algorithms to be explored are modified versions of VGG-16, ResNet50 and SE-ResNet50.

VGG-16 is a deep CNN network that was published in 2015. The creators, Visual Geometry Group (VGG), explores the effect of increasing the depth of the convolutional network on its accuracy. The network consists of 16 convolutional layers and has a small receptive field of 3x3. It has a max pooling size 2x2 and has a total of 5 such layers. In addition, there are 3 fully connected layers after the last Max pooling layer.

ResNet50 stands for residual networks that have 50 layers. It's a pre-trained network with a 224x224 input size that can classify images into 1000 object categories. Compared to VGG-16, ResNet50 has an additional identity mapping capability, which allows the model to bypass a CNN weight layer if the current layer is unnecessary. This mechanism helps with reducing overfitting in the training set since there are a lot of layers in the network.

Squeeze-and-Excitation (SE) is a new architectural unit for neural networks that aims to improve the quality of representations that the network produces. The block is integrated into ResNet50 by inserting it after the non-linearity following each convolution. Therefore, SE-ResNet50 has improved channel interdependencies at a near-zero computational cost.

#### References:

<https://www.mathworks.com/help/deeplearning/ref/resnet50.html>

<https://link.springer.com/article/10.1007/s42979-020-0114-9>

Theckedath, D., & Sedamkar, R. R. (2020). Detecting Affect States Using VGG16, ResNet50 and SE-ResNet50 Networks. *SN Computer Science*, 1(2). <https://doi.org/10.1007/s42979-020-0114-9>

Zeiler, M. D., & Fergus, R. (2014). Visualizing and Understanding Convolutional Networks. *Computer Vision – ECCV 2014*, 818–833. [https://doi.org/10.1007/978-3-319-10590-1\\_53](https://doi.org/10.1007/978-3-319-10590-1_53)

Mahbub, U., Sarkar, S., & Chellappa, R. (2018). Segment-based Methods for Facial Attribute Detection from Partial Faces. <https://doi.org/10.1109/TAFFC.2018.2820048>

Ekman, P., and Friesen, W. V. (1978). *Facial Action Coding System: A Technique for The Measurement of Facial Movement*. California, CA : Consulting Psychologists Press, Palo Alto.

Hu, J., Shen, L., Albanie, S., Sun, G., & Wu, E. (2017). Squeeze-and-Excitation Networks. <https://doi.org/10.48550/arxiv.1709.01507>

Ekman, P., and Friesen, W. V. (1978). *Facial Action Coding System: A Technique for The Measurement of Facial Movement*. California, CA : Consulting Psychologists Press, Palo Alto.