# Visual Recognition Project

## Image Classification Between Selfie and Non-Selfie Images

**Group-14:**

Sameer Ahmed (IIT2020053)
Sanjeet Beniwal (IIT2020052)
Harsh Garg (IIT2020082)
Jitu Rajak (IIT2020218)
Rahul (IIT2020244)

# Introduction

- Focuses on developing an image classifier using computer vision to differentiate between selfies and non-selfie images.
- With the prevalence of smartphones and social media, the distinction between personal and impersonal images has become blurred. The motivation comes from the need of sharing images on social platforms and to address various applications, including content recommendations, advertising, and security.
- The project aims to contribute to the advancement of computer vision, offering accurate and robust models with broad applications in social media, user experience, and data analysis in today's digital world.

# Methodology

Overview of Model Architecture

- Base Model: ResNet50 (with initial weights set on “imagenet”)
- ResNet50 serves as Feature Extractor for the model
- Residual Blocks in ResNet - Addresses the vanishing gradient problem

Transfer Learning – Custom Head Layers:

- The custom head layers adapt to the nuances of the specific dataset, focusing on learning class-specific features.
- Average Pooling, Flatten Layer, Dense Layer (256 units, ReLU activation), Dropout Layer (50%), Output Layer (Softmax activation for classes)

# Model Training Setup

- Freezing Layers - All layers in the base model (ResNet50) are set as non-trainable to retain Pre-trained Weights
- Focus on Custom Head Training - Freezing Layers Directs the training process to primarily update and fine-tune the custom head for the specific task.
- Loss Function - Binary Cross-Entropy: Suitable for binary classification tasks, where the output is a probability distribution over two classes.
- Optimizer - Adam : Efficient for training deep neural networks, helps converge faster and handle sparse gradients.

# Experimental Settings

We are initializing the model with the weights pre-trained on ImageNet dataset.

The input images of the model are of dimensions 224X224 pixels with three colour channels (RGB).

In case of model the average pooling layer in the head of model has dimensions 7X7, the first dense layer uses 256 units and uses ReLu activation function. The dropout rate is 0.5. The final dense layer has 2 units and activation function is Softmax. The optimizer used here is Adam optimizer with learning rate as 1e-4. The loss function here is binary cross entropy loss, and the Batch Size is 32.

The models have been trained on kaggle VM instances with GPU P100 turned on.

# Dataset Description

- The dataset for the problem of detecting whether an image is a selfie or not has been taken from Kaggle.
- The entire dataset contains roughly 70,000 images.
- The dataset is categorized into two classes: ”Selfie” images and ”non-selfie” images.
- The dataset is further split into training data, validation data, and testing data.
  - The training set comprises a total of 10,000 images
  - The testing and validation set each consist of 3,930 images.

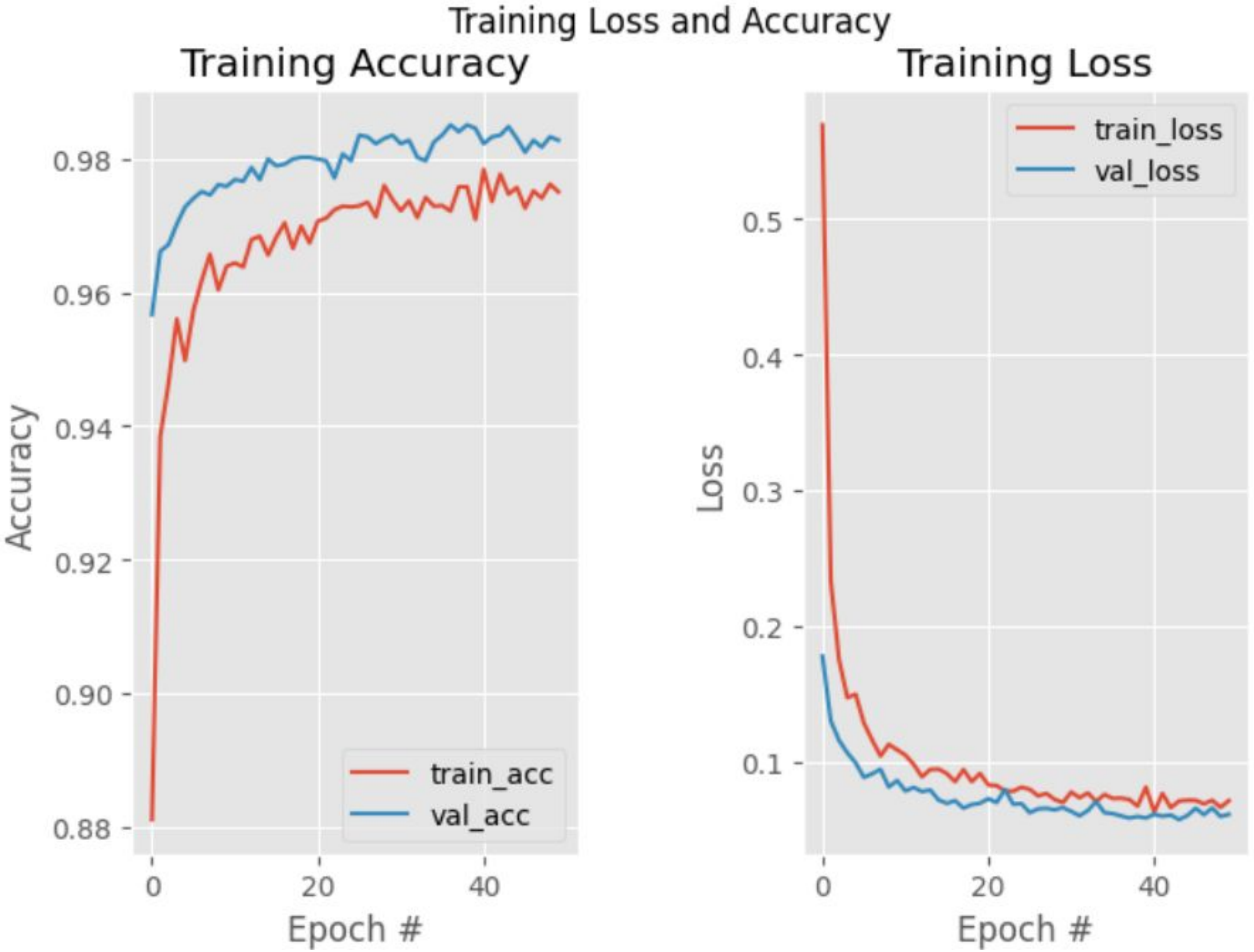
Non-selfies:

Selfies:

# Results and Analysis

# VGG16

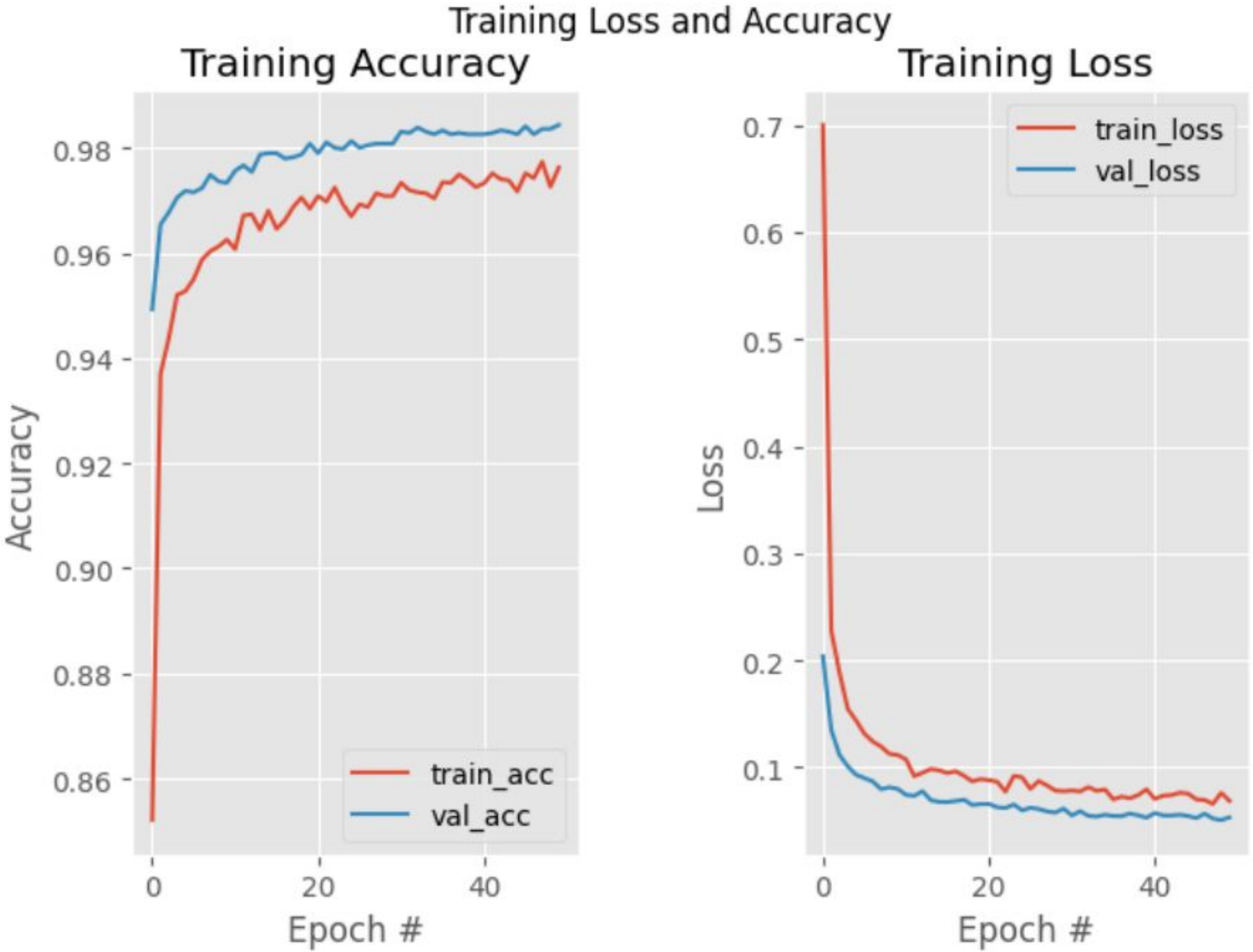
| | precision | recall | f1-score | support |
|---|---|---|---|---|
| NonSelfie | 0.99 | 0.98 | 0.98 | 1966 |
| Selfie | 0.98 | 0.99 | 0.98 | 1966 |
| accuracy | | | 0.98 | 3932 |
| macro avg | 0.98 | 0.98 | 0.98 | 3932 |
| weighted avg | 0.98 | 0.98 | 0.98 | 3932 |


Training Loss and Accuracy
Training Accuracy
Accuracy
0.98
0.96
0.94
0.92
0.90
0.88
0
20
40
Epoch #
train_acc
val_acc
Training Loss
Loss
0.5
0.4
0.3
0.2
0.1
0
20
40
Epoch #
train_loss
val_loss

Selfie: 56.83%
PLAYLIST
Selfie: 91.18%
NonSelfie: 100.00%

# VGG19

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| NonSelfie | 0.99 | 0.98 | 0.99 | 1966 |
| Selfie | 0.98 | 0.99 | 0.99 | 1966 |
| accuracy | | | 0.99 | 3932 |
| macro avg | 0.99 | 0.99 | 0.99 | 3932 |
| weighted avg | 0.99 | 0.99 | 0.99 | 3932 |

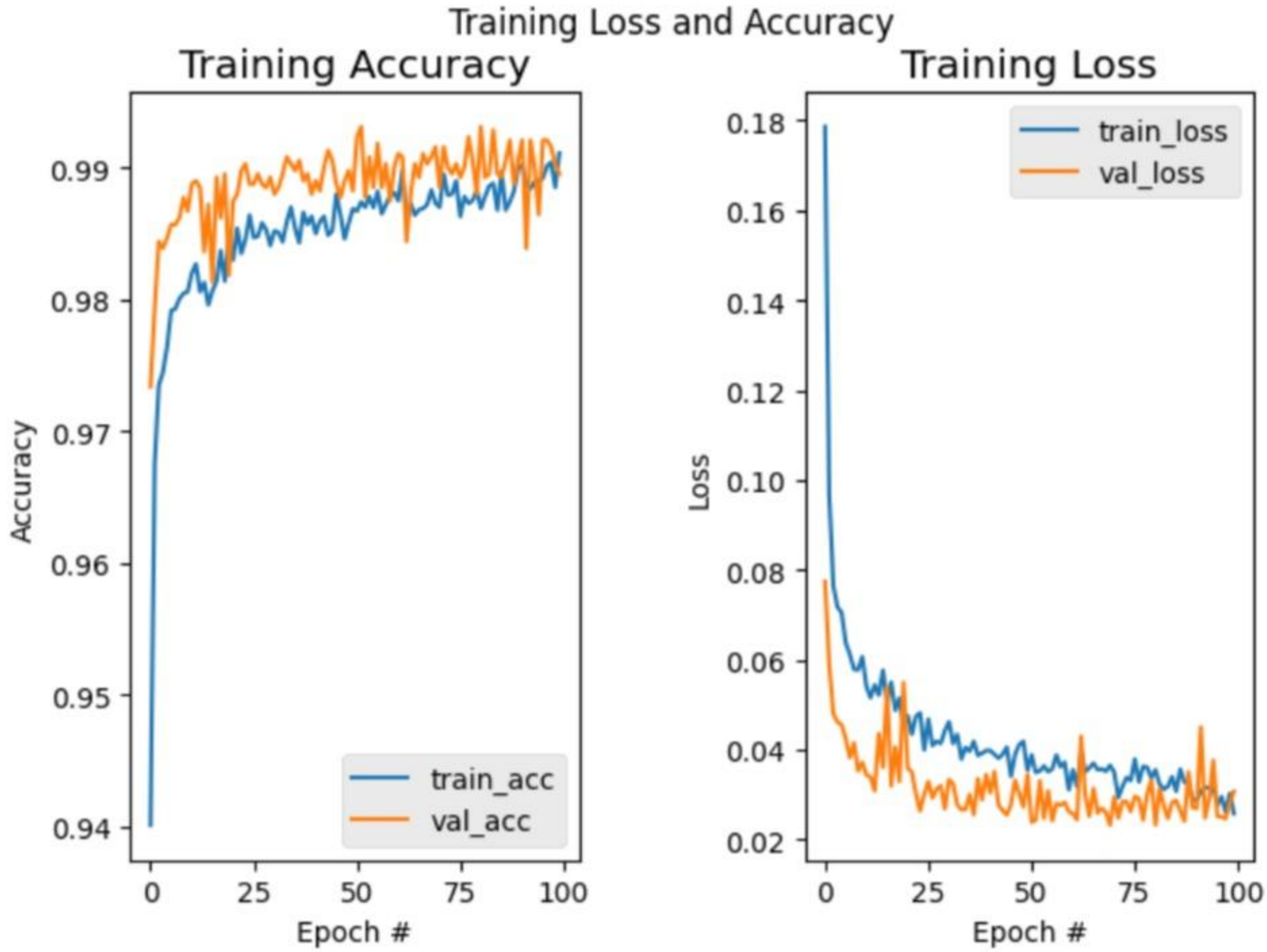


Training Loss and Accuracy
Training Accuracy
Training Loss
Accuracy
Loss
Epoch #
Epoch #
train_acc
val_acc
train_loss
val_loss
0.98
0.96
0.94
0.92
0.90
0.88
0.86
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
20
40
0
20
40


NonSelfie: 57.93%
PLAYLIST
NonSelfie: 99.96%
Selfie: 54.49%
NonSelfie: 99.94%

# ResNet50

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| NonSelfie | 1.00 | 0.98 | 0.99 | 1966 |
| Selfie | 0.98 | 1.00 | 0.99 | 1966 |
| accuracy | | | 0.99 | 3932 |
| macro avg | 0.99 | 0.99 | 0.99 | 3932 |
| weighted avg | 0.99 | 0.99 | 0.99 | 3932 |



Training Loss and Accuracy
Training Accuracy
Accuracy
0.99
0.98
0.97
0.96
0.95
0.94
0
25
50
75
100
Epoch #
train_acc
val_acc
Training Loss
Loss
0.18
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0
25
50
75
100
Epoch #
train_loss
val_loss


NonSelfie: 57.93%
NonSelfie: 99.96%
Selfie: 54.49%
NonSelfie: 99.94%

# Conclusion

- By looking at the loss curves of the three models, we can see that all three models perform very well on the test dataset.
- The performance of all three models is remarkably high.
- The models not only perform good on the data similar to images in the training dataset but also on the real world examples.
- These models can be used in the real world.

# References

[1] S. Kong, ”Self-supervised Image Classification Using Convolutional Neural Network,” 2023 IEEE 3rd International Conference on Power, Electronics and Computer Applications (ICPECA), Shenyang, China, 2023, pp. 1283-1290, doi: 10.1109/ ICPECA56706.2023.10075949.
[2] He, K., Fan, H., Wu, Y., Xie, S., Girshick, R. (2020). Momentum contrast for unsupervised visual representation learning. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 9729-9738).
[3] Chen, T., Kornblith, S., Norouzi, M., Hinton, G. (2020, November). A simple framework for contrastive learning of visual representations. In International conference on machine learning (pp. 1597-1607). PMLR.
[4] Kolesnikov, A., Zhai, X., Beyer, L. (2019). Revisiting self-supervised visual representation learning. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 1920-1929).
[5] Ericsson, L., Gouk, H., Loy, C. C., Hospedales, T. M. (2022). Selfsupervised representation learning: Introduction, advances, and challenges. IEEE Signal Processing Magazine, 39(3), 42-62.
[6] Goyal, P., Mahajan, D., Gupta, A., Misra, I. (2019). Scaling and benchmarking self-supervised visual representation learning. In Proceedings of the ieee/cvf International Conference on computer vision (pp. 6391- 6400).
[7] He, Kaiming, et al. ”Deep residual learning for image recognition.” Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
[8] Simonyan, Karen, and Andrew Zisserman. ”Very Deep Convolutional Networks for Large-Scale Image Recognition.” ArXiv, (2014). Accessed November 2, 2023. /abs/1409.1556.

Thank you