UNIVERSITAS
GADJAH MADA

Video and Image Manipulation Detection Based on ResNet and EfficientNet

Aisya Ulul Asmi (25/564969/PPA/07123)
Ariq Sudibyo (24/551515/PPA/06976)
Fiqki Haidar A (25/563337/PPA/07109)
Rafli Irfansyah K (24/551889/PPA/06988)
Stephanus H S (24/552378/PPA/07009)

UNIVERSITAS GADJAH MADA

Introduction

- Deepfake technology utilizes machine learning (ML) algorithms to produce fake, highly realistic content.
- The technology is increasingly accessible and affordable, leading to a rise in misuse and cybercrimes.
- Examples of misuse include the spread of manipulated photos or videos intended to damage reputations.

Polda Jatim Ungkap Kasus Penipuan Deepfake AI Kepala Daerah, Pelaku Kantongi Keuntungan Hingga Rp87 Juta

Diunggah pada: 28 April 2025 20:33:41 7418

UNIVERSITAS GADJAH MADA

Reference to Previous Research

Contents lists available at ScienceDirect

SoftwareX

journal homepage: www.elsevier.com/locate/softx

ELSEVIER

Original software publication

FDT: A python toolkit for fake image and video detection

Surbhi Raj *, Jimson Mathew, Arijit Mondal

Indian Institute of Technology Patna, India

ARTICLE INFO

Article history:
Received 3 November 2022
Received in revised form 21 April 2023
Accepted 23 April 2023

Keywords:
Deepfake video
Generative Adversarial Networks (GANs)
Fake Detection Tool (FDT)
Copy-move
Splicing
Twitter

ABSTRACT

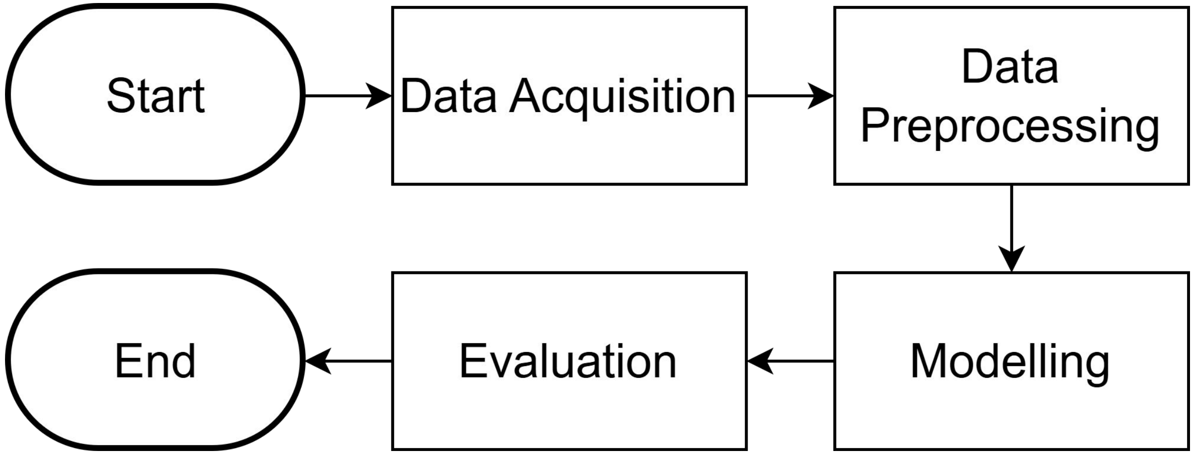
With the advent of readily and widely available applications based on deepfake technology, several cybersecurity threats are on the rise. It is challenging to curtail these threats as deepfakes are realistic and very difficult to detect. The present work proposes a Fake Detection Tool (FDT) that streamlines the procedure of fake detection by incorporating various manipulation techniques and aids users in detecting and visualizing the same. The tool is also integrated with Twitter for streaming facial image posts based on hashtags. It provides an output dataframe and presents statistics of virality, sentiments, etc, using pie charts for better visualization. The proposed tool uses a wide variety of large-scale datasets for training to deal with the fakes in the wild and deploys models that are at par with the cutting-edge models. It is an efficient, user-friendly, and freely available software for fake detection. The source code of the FDT package toolkit is available at https://github.com/surbhiraj786/GUI_Fake-Detection.

Research Proposal: This study proposes a focused experimental framework to benchmark two complementary deep architectures:

- ResNet-based models
- EfficientNet-based models

Research Method

UNIVERSITAS GADJAH MADA



Research Method (Data Acquisition)

UNIVERSITAS GADJAH MADA

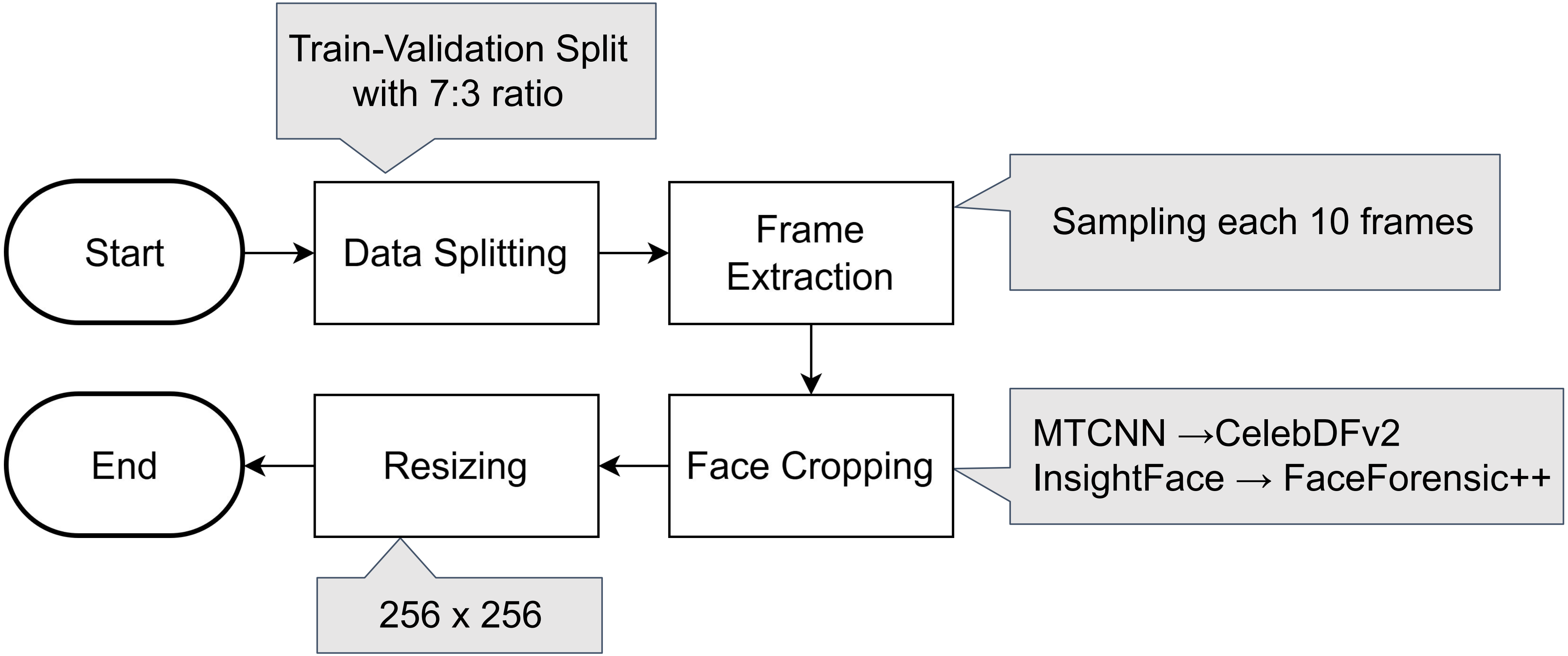
(FaceForensic++ Sample Data)

(CelebDFv2 Sample Data)

FaceForensics++: contains 1,000 real videos, and their respective Deepfakes, Face2Face, FaceSwap, and NeuralTextures version.

CelebDFv2: contains 590 real videos and 5,639 fake videos generated from those real videos.

Research Method (Data Preprocessing)



UNIVERSITAS GADJAH MADA

Research Method (Modelling)

Dataset	Model	Training Parameter
CelebDFv2	EfficientNetB7	Epoch: 10(20); Batch Size: 1; Optimizer: Adam; Learning Rate: 1e-4
	ResNet50	Epoch: 10(20); Batch Size: 1; Optimizer: Adam; Learning Rate: 1e-4
FaceForensic++	EfficientNetB7	Epoch: 20; Batch Size: 8; Optimizer: Adam; Learning Rate: 4e-4
	ResNet50	Epoch: 20; Batch Size: 8; Optimizer: Adam; Learning Rate: 1e-5

Research Method (Evaluation Metrics)

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

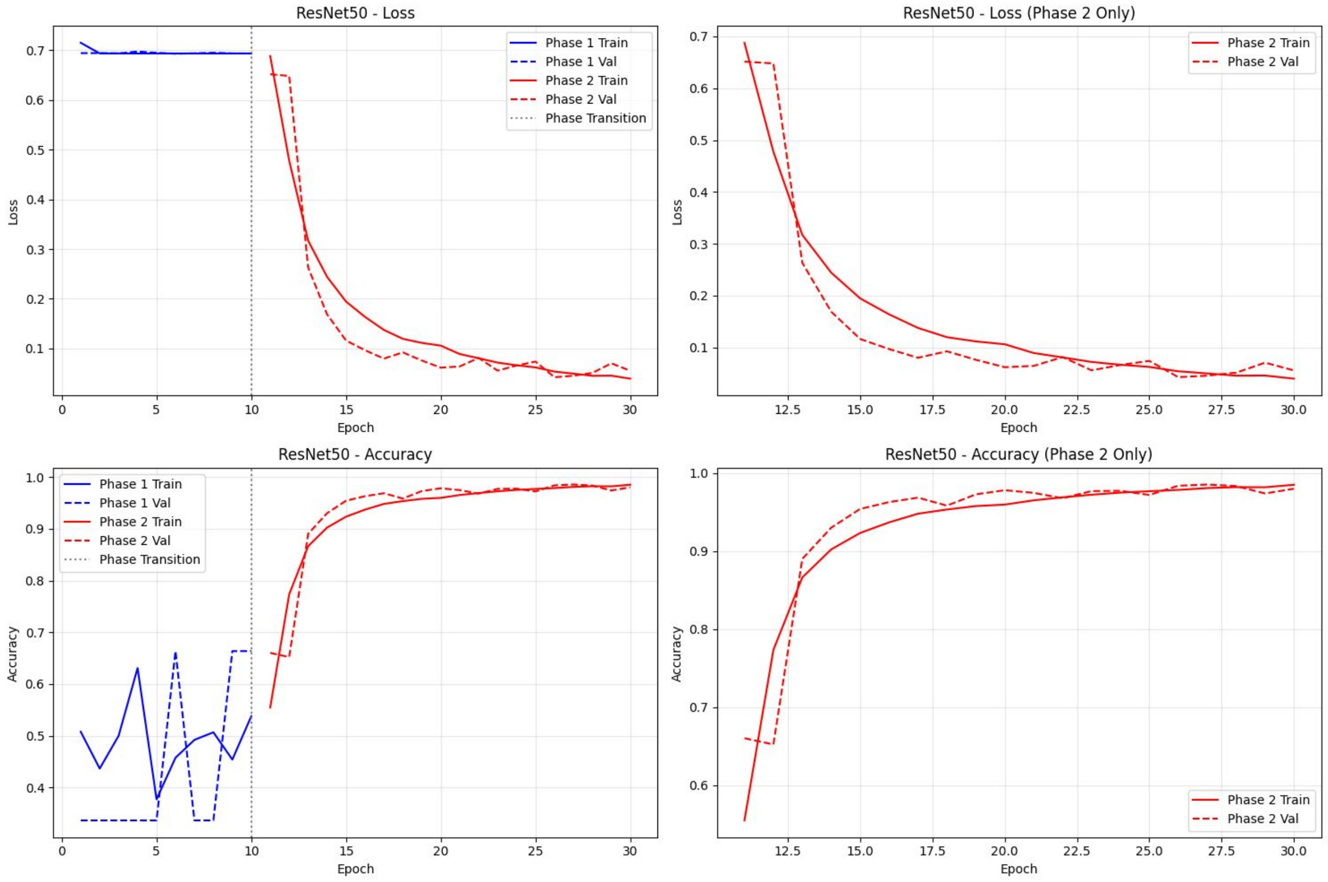
$$F1\ score = \frac{2\ x\ Precision\ x\ Recall}{Precision + Recall}$$

With:

- TP = True Positive
- TN = True Negative
- FP = False Positive
- FN = False Negative

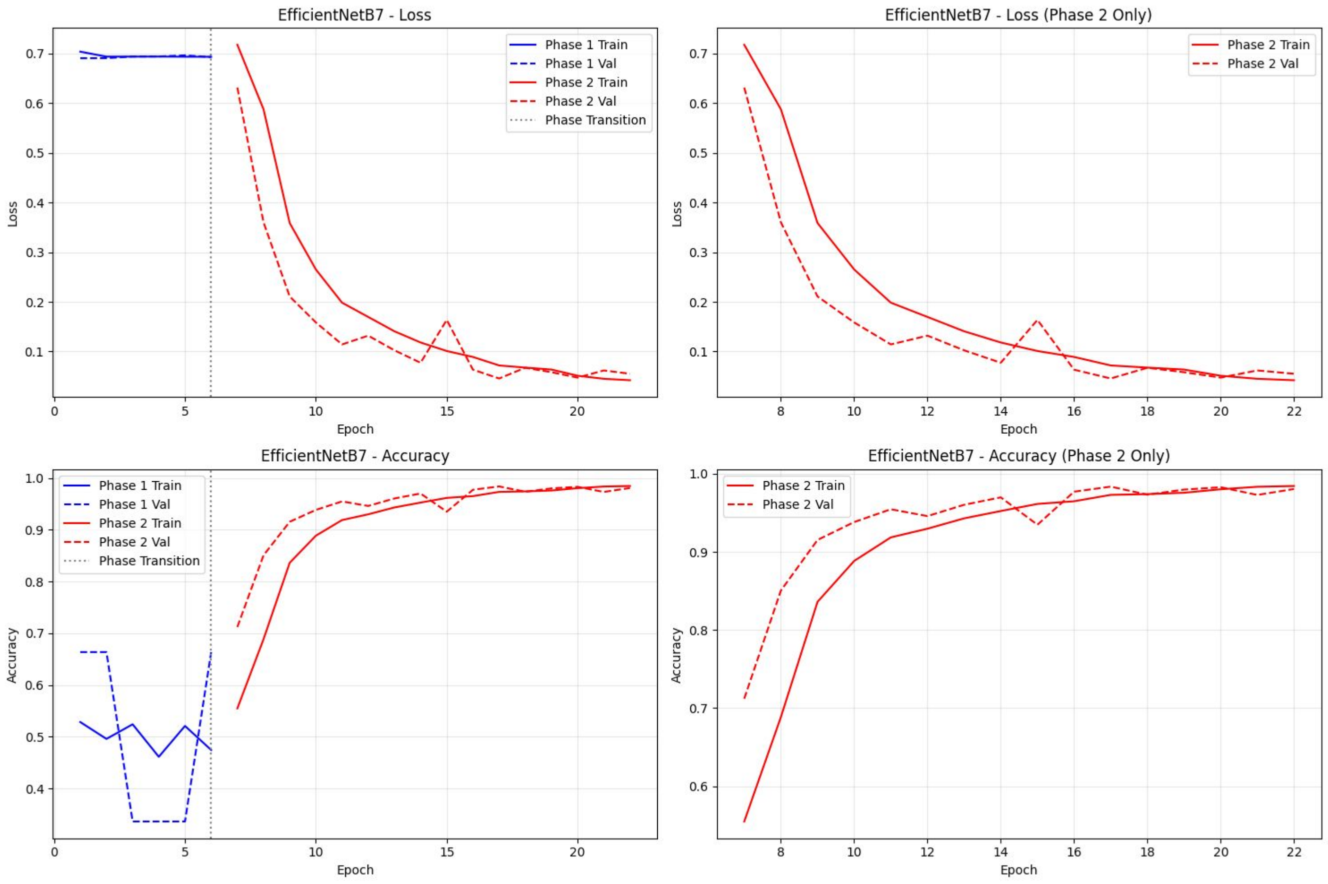
Result and Discussion

ResNet50 model across two-phase training on Celeb-DF-v2



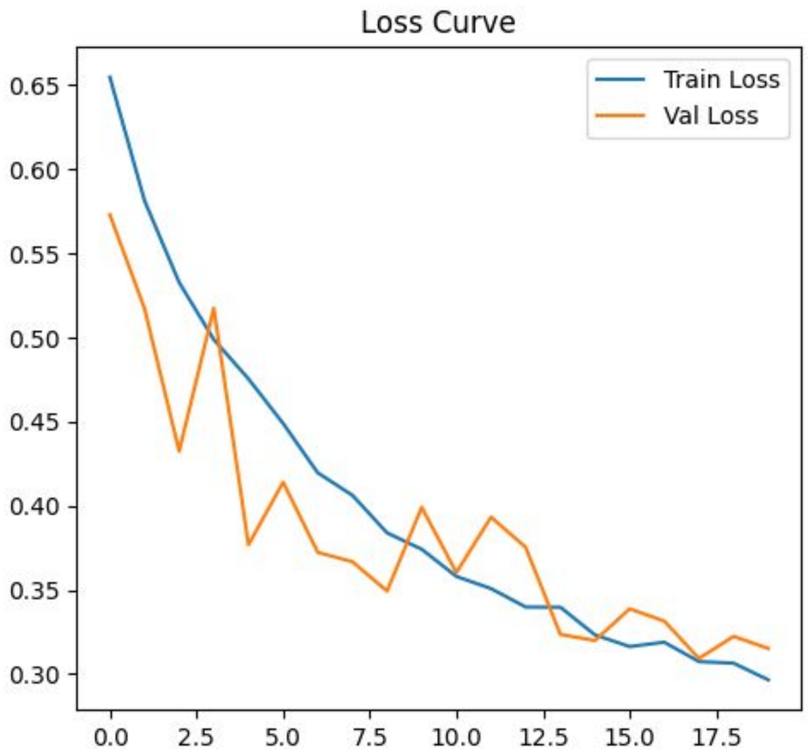
Result and Discussion

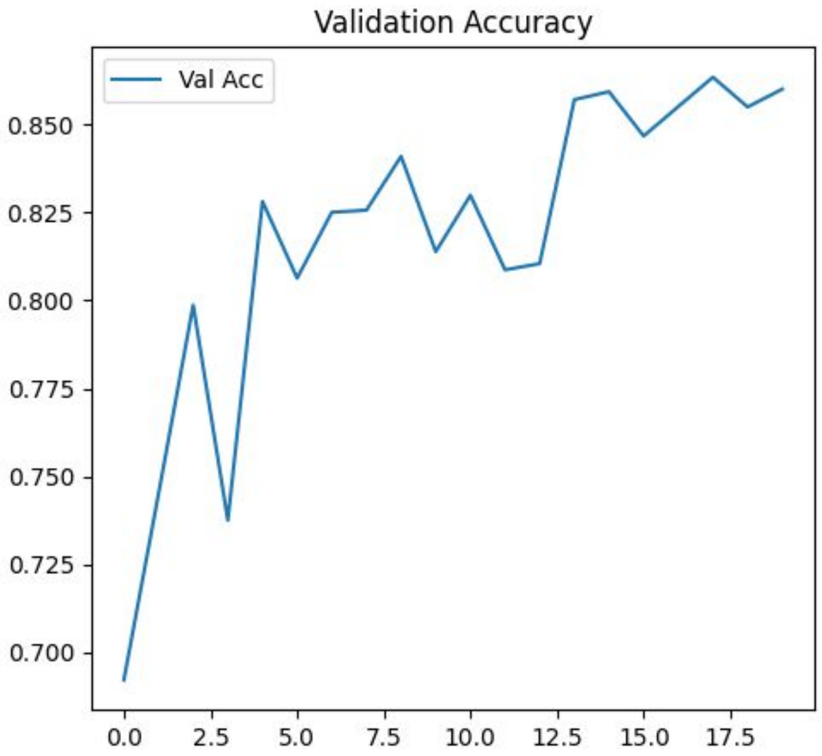
EfficientNetB7 model across two-phase training on Celeb-DF-v2



Result and Discussion

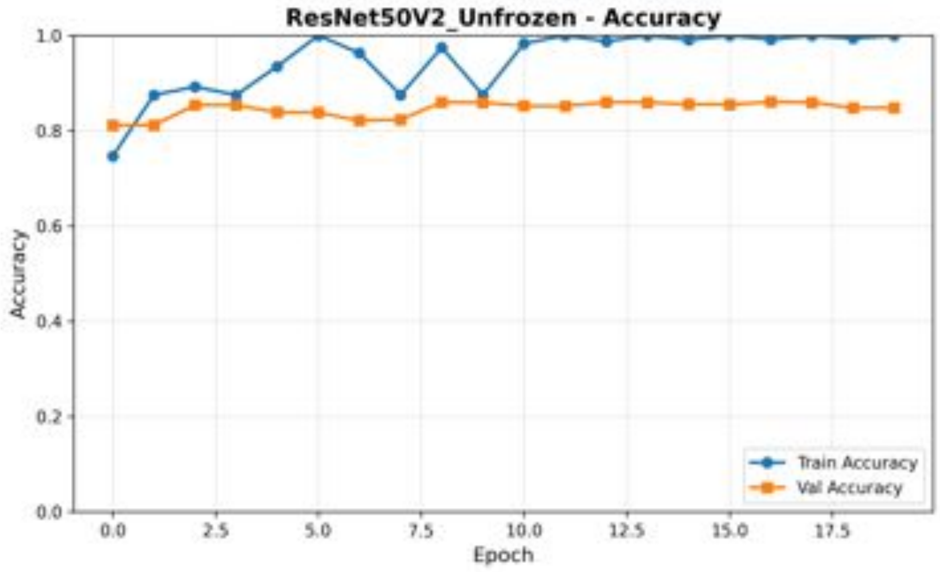
EfficientNetB7 model training on FaceForensics++

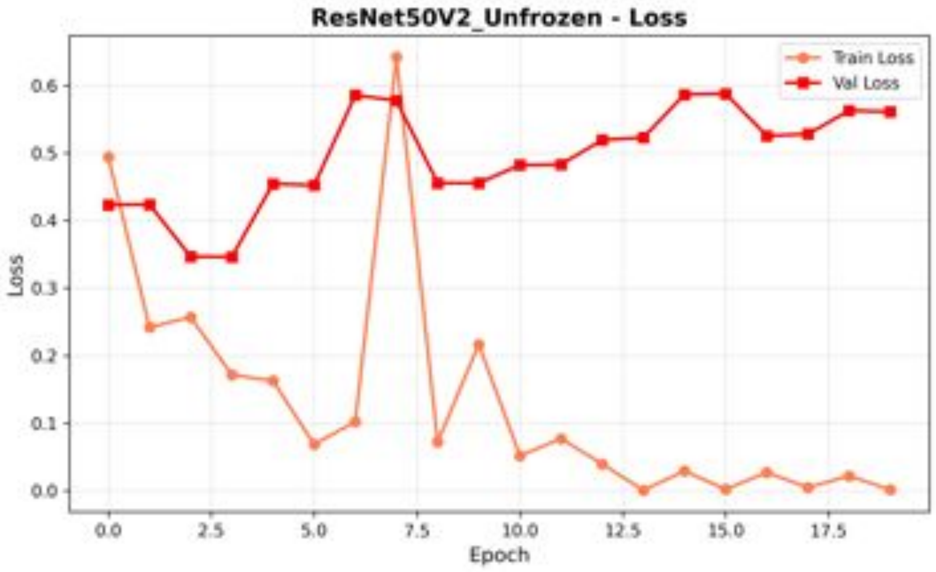




Result and Discussion

ResNet50 model training on FaceForensics++





UNIVERSITAS GADJAH MADA

Result and Discussion (Evaluation)

Evaluation on Validation Set

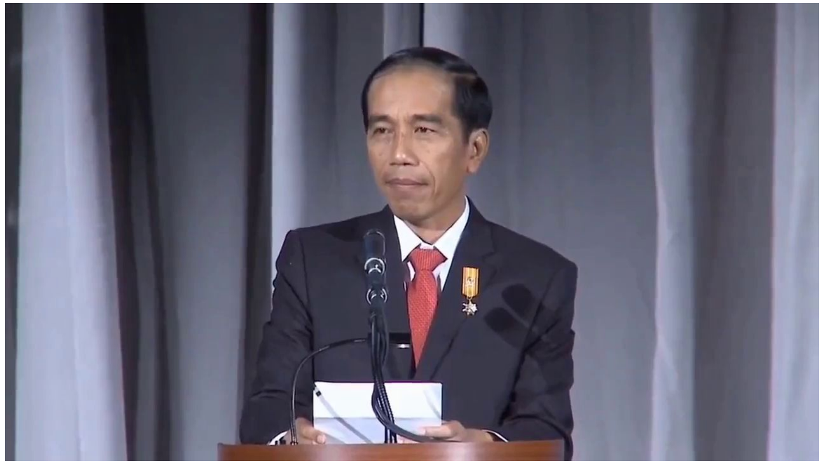
Dataset	Model	Accuracy	Precision	Recall	F1-Score
CelebDFv2	EfficientNetB7	**97.5%**	**99.6%**	**93.4%**	**96.4%**
	ResNet50	93.5%	95.3%	85.9%	90.3%
FaceForensic++	EfficientNetB7	**90.1%**	83.5%	**89.1%**	**86.2%**
	ResNet50	81.98%	**90.17%**	54.1%	67.63%

Result and Discussion

UNIVERSITAS GADJAH MADA

Evaluation on Unseen Data

Model	Actual	Predicted	Confidence
EfficientNet-FF++	Real	Real	59.16%
	Fake	Real	85.11%
ResNet-FF++	Real	Fake	65.40%
	Fake	Fake	55.35%
EfficientNet-Celeb-DF-v2	Real	Fake	98.63%
	Fake	Real	66.44%
ResNet-Celeb-DF-v2	Real	Real	50.57%
	Fake	Real	54.22%



Conclusion

UNIVERSITAS GADJAH MADA

1. EfficientNetB7 achieved the **highest accuracy on both dataset**, with 90.1% accuracy on FaceForensics++ and 97.5% accuracy on CelebDFv2.

2. **Generalization problem**: Models develop narrow, dataset-specific notions of authenticity, leading to unacceptable false positive rates on unseen authentic content.

Future Works

1. Multi-dataset or progressive training.
2. Data augmentation.
3. Incorporating temporal information.
4. Fusion of additional modalities.
5. Consideration to evaluate model with recently developed deepfake methods.

Thank You

UNIVERSITAS
GADJAH MADA

LOCALLY ROOTED, GLOBALLY RESPECTED

ugm.ac.id