Comparative Analysis of Convolutional Neural Networks in Regards to Deepfake Detection

By

Jeet Kumar

B.A. Business Technology Administration (University of Maryland: Baltimore County)
2019

Capstone

Submitted in partial satisfaction of the requirements for the degree of

MASTER OF SCIENCE

In

CYBERSECURITY

In the

GRADUATE SCHOOL

Of

HOOD COLLEGE

December 2024

Accepted:

Ahmed Salem, Ph.D.
Committee Member

George Dimitoglou, Ph.D
Program Director

Carol Jim, Ph.D.
Committee Member

George Dimitoglou, Ph.D
Capstone Advisor

TABLE OF CONTENTS

1.1 LIST OF TABLES

1.2 LIST OF FIGURES

2. ABSTRACT

Deepfake technologies pose a significant threat in the field of cybersecurity, requiring advanced deepfake detection methods to mitigate potentially malicious risks. This thesis conducts a comparative analysis of convolutional neural networks (CNNs) for deepfake detection, focusing on the performance of the EfficientNetV2B0 and ResNet50 models. Using the FaceForensics++ dataset as input, this research will examine several key evaluation metrics including training and validation accuracy, training and validation loss, and training time per epoch, to assess each model's efficacy and effectiveness. ResNet50 achieved a higher final validation accuracy of 88.41%, demonstrating a superior performance in identifying manipulated content, at the cost of longer training times per epoch. Alternatively, EfficientNetV2B0 reached a lower final validation accuracy of 79.88%, while exhibiting faster learning times and a reduction in demand for computational resources. This study stresses the trade-offs between achieving a higher detection accuracy and the optimization of computational efficiency. Furthermore, this will provide a broader comprehension into the practical considerations for the production and deployment of deepfake detection models for realistic environments. With the contribution of these findings, a broader understanding of CNN-based deepfake detection can be put to use and establish a foundation for future research and development that can focus on refining model scalability, improving evaluation metrics and augmenting learning patterns to overcome evolving deepfake techniques.

3. INTRODUCTION

AI-generated content represents a growing threat in the digital world, specifically deepfakes. Manipulated audio and video that is fabricated to appear as deceptive, yet hyper-realistic media, has become a prevalent adversary for not only cybersecurity professionals but also society. It has created substantial implications for the general public such as the proliferation of misinformation, identity theft and diminution of trust in media. This threat has established a need for a scalable and efficient deepfake detection method, which can be procured by being based on convolutional neural networks (CNNs). This technology has shown outstanding success in resolving image classification problems, which makes it a practical candidate for deepfake detection [1]. Within this framework, modern CNN architectures like ResNet 50 and EfficientNetV2 offer unique approaches and propitious advantages in the act of detecting manipulated content.

ResNet50, a revised version of ResNet, improves on its predecessor by incorporating optimized scaling strategies and enhanced training techniques. This allows for deeper learning architectures without an increase in computational complexity and a drain on resources [2]. EfficientNetV2 has been designed to maximize efficiency by meticulously balancing network depth, width, and resolution while maintaining a resource-constrained approach. This study will evaluate how these specific architectures will perform on a deepfake dataset, FaceForensics++, and examine their detection capabilities and efficacy in terms of training and inference time. Therefore, conducting this comparative analysis will contribute to the development of reliable deepfake detection methods.

An emerging proponent in deepfake detection is discerning the level of sophistication the manipulation technique utilizes to alter the authentic media [3]. Many deepfake generative models have become so sophisticated that traditional detection methods have become ineffectual in identifying anomalies or subtle inconsistencies within the media and are unable to differentiate between the real and fake media. Additionally, deepfake detection systems must be able to balance efficacy with expediency, especially in time-restricted environments. This generates challenges related to complexity, computational efficiency and scalability. An understanding of these difficulties establishes an urgency in the need for more robust and innovative detection methods as this technology continues to proliferate.

While many studies implement CNNs for image classification, few have specifically compared ResNet50 and EfficientNetV2 for deepfake detection. Most of the existing literature primarily focuses on their performance in image recognition tasks like ImageNet without providing a comprehensive assessment of their performance in distinguishing deepfakes. The scalability of these models in regards to deepfake detection is something these studies fail to address. This work will address the gaps in the literature by conducting an extensive examination of these CNNs which will evaluate accuracy, computational efficiency and efficacy. By examining how cutting-edge architectures like ResNet50 and EfficientNetV2 perform in the detection of deepfakes, this study will provide new insights in their suitability for this arduous task.

The significance of this study lies in the evolution of AI and its potential impact on cybersecurity and society. As deepfake generated media becomes harder to detect, the

capability to deploy resilient detection systems will become progressively critical. The aim of this research is to encourage the inception of more efficient solutions to deepfakes which will help safeguard organizations and individuals from the precarious nature of deepfake media. Furthermore, this study can advance the development of deepfake detection systems in real-time deployment and real-world scenarios.

4. LITERATURE REVIEW

The rise of artificial intelligence in the technology industry has produced many marvels in advancement. However, it has also led to the generation of malicious content like the deepfake technology. This method uses deep learning techniques to generate audio and video content which is difficult to distinguish from genuine media. Deepfakes create a unique and consequential threat to the privacy and security of organizations and people alike [4], [5]. This technology creates a unique challenge of identifying, recognizing, and discerning real and fake content and has initiated a development of effective detection mechanisms for this task [3]. The process of accurately detecting deepfakes has prompted academics and researchers to investigate various techniques of classification including convolutional neural networks (CNNs) to address this issue [6], [7], [8].

Deepfake technology is an image classification area where CNNs have shown considerable success. At the core of the CNNs' capability is the unique architecture it employs. This ability to mimic the visual system of a human to identify patterns within an image gives it the capacity to solve complex issues. The technique of extracting and learning from distinctive features within the pixels of an image makes it a suitable candidate for locating the subtle manipulations within deepfakes [9]. As the data propagates through deeper layers, the CNNs learn more abstract features by grouping particular sets of pixels which gives an unparalleled, progressive approach in capturing inconspicuous variations in images.

Among prominent CNNs is ResNet or Residual Networks [9], which introduced the concept of Residual Learning [2], [10]. This facilitates the training of substantially deeper learning networks by establishing a solution for the vanishing gradient problem. This problem occurs when the depth of a network increases, the model's weights diminish considerably as the gradient passes through the neural network layers. This prevents the network from further optimization leading to inferior performance overall. A recent version of ResNet, ResNet50, balances computational efficiency and performance for image analysis which makes it a clear candidate for the task of deepfake detection [2].

Another CNN that will be used is EfficientNet, which is notable for its scaling of depth, width, and resolution. The major benefit of using this CNN model will be its use of a restrained approach to computational resources. This will become the main metric of performance which contrasts fundamentally to ResNet50, which primarily values accuracy [11]. An improved version of EfficientNet, EfficientNetV2B0, utilizes a more sophisticated scaling system and moderate training time for its implementation, making it modular for any scenario [12].

These are promising candidates when it comes to determining the validity of deepfakes, however, their differences in architecture, computational efficiency and resource

consumption may lead to a distinction of performance in relation to their detection efficacy.

4.1 DEEPFAKE DETECTION

Deepfake technology refers to artificial intelligence that masquerades as real by manipulating original content and manufacturing it to appear realistic. By using deep learning techniques, specifically generative adversarial networks (GANs), deepfakes can create realistic audio and video and synthesize new media entirely [13]. Over the past decade, this technology has created a massive trend, driven by open-source software and trends in the social media industry, making it easier for amateurs and professionals alike to manufacture hyper-realistic content [14], [15]. This has created a threat to the cybersecurity industry as deepfakes establish themselves as a unique tool for any would-be attackers. By generating and synthesizing lifelike content, attacks such as impersonations, social engineering, and fraud are all viable options for anyone willing to use this technology for malicious purposes. It can also lead to widespread misinformation by creating falsehoods in news and influence public opinion, especially during important times like elections or political conflicts. As an answer to this evolving challenge, effective deepfake detection can help identify content that has been forged, protect organizations, people, and society from deceitful misconduct, and assist with the reduction of the impact of identity theft and misinformation. The development of increasingly robust algorithms and neural network technologies will require advancement as the deepfake technology continues to evolve.

4.2 BACKGROUND OF NEURAL NETWORKS

Neural networks are the foundation of the modern deepfake detection technology we know today. These models are inspired by the learning patterns of the human brain, which uses interconnected layers of structured data to create the capability of learning complicated information based on patterns recognized in data. A significant innovation within the neural network field came with the introduction of CNNs [1], [9], [16]. During its inception, this technology was used for rudimentary image analysis by extracting key features within an image, like edges or textures, in initial layers and recognizing different models and objects within deeper layers [17]. This level of cognizance makes CNNs suitable for tasks within computer vision, like deepfake detection, where the recognition of subtle manipulations between forged and authentic content is critical [1].

4.3 BACKGROUND OF CONVOLUTIONAL NEURAL NETWORKS

CNNs are a category of deep learning models that created a foundation for various applications within image analysis [17], [18]. The architecture of a CNN uses several types of layers to extract features from an input image like textures or patterns. Convolutional and pooling layers are used in tandem to train the filters to highlight areas of importance and reduce the computational requirements. These layers create a feature map, which is passed through the layers to establish a link between the translations. This assists in substantiating a final prediction for the image or video being tested. Notable CNNs like Xception [19], ResNet [9] and EfficientNet [11] have been used extensively for the classification of images and have achieved remarkable benchmarks. Due to their algorithmic approach to recognizing complex visual patterns, any subtle shifts or transformations in an image or video seldom affects the analysis that is done by the CNN model.

EfficientNetV2 is an advanced CNN which enhances the performance of its predecessor, EfficientNet [12]. It uses a model scaling approach to balance network depth, width, and resolution, which allows it to achieve and maintain state-of-the-art efficacy. A notable architectural improvement that is utilized is the fused-MBConv layer, which combines the usage of depthwise and pointwise convolutions and transitions this into a single operation, which decreases training time [12]. The EfficientNetV2 has been tested

rigorously on various datasets where the model showcased an improvement in efficacy and training speed compared to other CNNs like ResNet.

ResNet50 is an improved version of the original ResNet (Residual Networks) architecture, which was created to train deep neural networks at a faster pace [20]. One of its notable features is the introduction of residual connections, or “skip connections,” which allow the network to train properly without suffering from the vanishing gradient issue [9]. By bypassing layers within the network, ResNet is able to performance degradation that afflicts other models and also makes it easier for the network to train efficiently. ResNet50 has added more layers to the model, but it compensates for this intensive usage of resources by ensuring the network expands in a more balanced approach. This technique grants the model a more computationally efficient advantage than its predecessor and ensures that an increase in efficacy is not achieved at the cost of resources, making it more practical for deepfake detection applications [2].

4.4 COMPARATIVE METRICS

EfficientNetV2B0 and ResNet50 are designed with different principles and priorities in terms of complexity, training speed, and scalability. While there are multiple studies corroborating their efficacy in real-world scenarios, there are few examples of literature that compare the robustness of ResNet50 and EfficientNetV2B0 for deepfake detection in images or videos [21], [22]. By using a combination of performance metrics, it is possible to gain a holistic understanding of which model is more applicable to the challenges presented by deepfake detection. The total amount of time that is required to train a model can provide insight into the computational efficiency of the models. The precision and accuracy in identifying deepfakes can be used as metrics to gauge the validity of each model, especially if applied to the same dataset. Measuring the resource usage of each model will also provide an explanation into the rationality of each model for this particular issue. Using these metrics, a comprehensive analysis can serve as evidence for the plausibility of these models and their application to this generative issue.

The literature encompassing deepfake detection techniques using convolutional networks (CNNs) like ResNet50 and EfficientNetV2B0, reflects a significant development in image and video analysis with both architectural frameworks indicating a notable performance in comprehensive classification tasks. While there are gaps in the existing research comparing these two CNNs in the domain of deepfake detection, each excels in its own different manner [23], [24]. The focus on efficiency through residual connections that ResNet50 demonstrates and the focus on resource-efficient training that EfficientNetV2B0 offers, make them unique tools for establishing more dependable and scalable deepfake detection systems. There are various characteristics of these two architectures that remain understudied, including computational efficiency and detection efficacy, especially for environments with constrained resources. These areas of comprehensive research will need to be studied further to provide beneficial insights to the cybersecurity community and industry as a whole and to establish a future safeguard against deepfake technology.

5. PROBLEM/SOLUTION

This thesis addresses the research questions of how effective deepfake detection models, like ResNet50 and EfficientNetV2 convolutional neural networks (CNNs) can be when identifying synthetic media. The analysis performed will focus on their accuracy, efficacy, and validity in practical application. Using the FaceForensics++ dataset, this research will determine which model will fulfill the expectations required when dealing with rapidly evolving variables. This, in turn, will contribute to advancements made in automated detection methodology.

5.1 METHODOLOGY

To address this issue, an empirical comparison of these two CNN architectures involving training and curating of proper parameters must take place. This will involve employing the FaceForensics++ database, a well-tested and comprehensive resource frequently used by many researchers for image recognition applications. Since both models are used for image recognition and have been optimized for large datasets, this comparative study will provide the insight necessary to identify which CNN is best suited for this specific task.

This methodology will include preprocessing of the data to establish a base of consistency by controlling and standardizing the input into each model. This process will include generating images from each of the video files within the FaceForensics++ database, resizing all of the images to the same pixel dimensions and equalizing their color channels. Each model will then be trained and validated using the same training/validation split and every parameter will be recorded and normalized to ensure optimal model performance. This will minimize variables unrelated to model performance and will create a non-biased analysis.

This observational design will take a qualitative approach to the comparison of the ResNet50 and EfficientNetV2 regarding deepfake detection. The purpose of this is to

emphasize and validate the differences between the two CNNs in terms of capability and efficacy.

5.2 ANALYSIS

The primary objective is to evaluate the performance differences between ResNet50 and EfficientNetV2B0 in terms of accuracy and efficiency. This phase will include several evaluation metrics, mainly the training and validation accuracies and losses as well as the training time per epoch. The metrics will provide a comprehensive view of each model's performance while providing an evaluation of their use case. Training and validation accuracy will measure how well each model will classify as real or fake. Monitoring accuracy over the course of 50 training epochs allows for a clear assessment of the performance improvement over time. The training and validation loss are analyzed to identify potential issues with the models like underfitting or overfitting. A model exhibiting a lower training loss compared to the validation loss would indicate overfitting, which might prompt further mitigation techniques like augmentation of data.

Using training time per epoch as a metric is important as it provides insight into the computational cost of using each architecture. An epoch time of 50 was chosen to emphasize the rate of change in accuracy that is gained over time of each CNN model. A lower epoch count will not showcase this change as the rate of change is sporadic during the earlier epochs. A higher epoch count will show an increase in training and validation accuracy, but this change becomes gradual and becomes repetitive as the executions run by the model reach a certain threshold. While training time per epoch should not be

considered as an impactive metric as training and validation accuracy as computational costs decrease, it becomes relevant in scenarios where time will be a known hindrance.

The approach began with preparing the FaceForensics++ dataset, which contains real and manipulated facial images. The data was organized into ‘real’ and ‘fake’ folders to facilitate an efficient data loading process. The dataset was then fed into ResNet50 and EfficientNetV2B0, with each model utilizing pre-trained ImageNet weights for transfer learning. This strategy provided an equalized starting point, thereby reducing the overall training time needed for convergence. Both of the CNN models were optimized on the deepfake dataset using a consistent set of hyperparameters including the batch size of 32, and the number of epochs, to ensure a fair comparison and assessment.

During training, the training and validation accuracy, as well as the training and validation loss, were recorded at each epoch as input into a graph which will help visualize the models’ learning curves and will help demonstrate training dynamics. This will highlight the tradeoffs between achieving higher training and validation accuracy or the optimization of the model for faster training times.

In summary, the methodology mentioned above integrates a comprehensive set of metrics – training and validation accuracy, training and validation loss and training time per

epoch – to thoroughly assess the performance of Resnet50 and EfficientNetV2B0 for deepfake detection. This structured approach involves consistent data preparation, transfer learning and detailed monitoring of dynamic metrics, which will ensure the findings are both replicable and robust. This methodology provides a coherent blueprint for future researchers looking to build upon or validate this work, thereby emphasizing the importance of balancing accuracy and efficiency in the design of deepfake detection systems.

5.3 RESULTS

The first set of results is displayed as graphs illustrating the training and validation accuracy of 50 epochs for ResNet50 (Figure 1) and EfficientNetV2B0 (Figure 2). These graphs provide a visual representation of how the models' accuracy evolves over time during training, with the x-axis indicating the number of epochs and the y-axis representing the accuracy percentage. For ResNet50, the training accuracy exhibits a gradual but steady increase, eventually stabilizing around 90%, while the validation accuracy plateaus at 88.41%. EfficientNetv2B0 reaches its peak accuracy of 79.57% at a much quicker pace, while failing to match the higher accuracy level of ResNet50. This outcome highlights the trade-offs between training time and final accuracy.

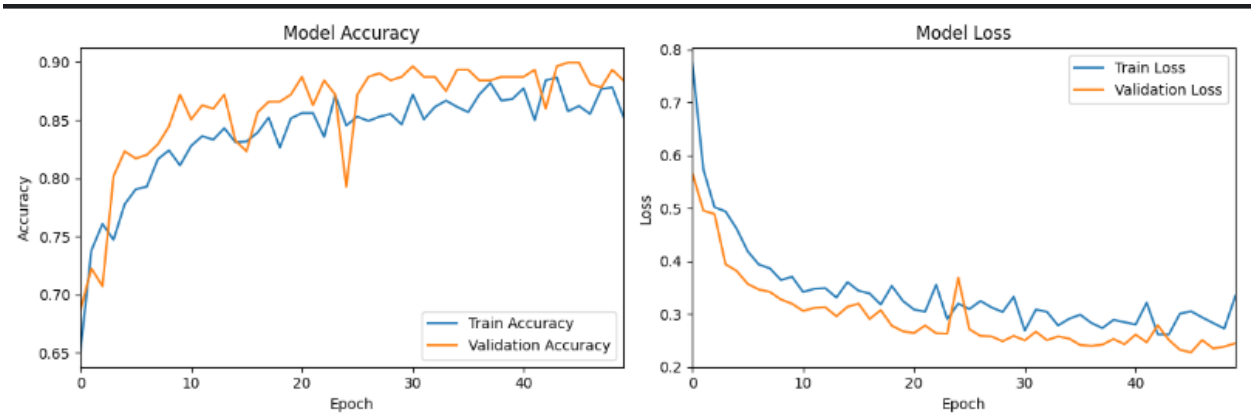

Figure 1: Training time and Validation Accuracy for ResNet50 for 50 Epochs. This graph shows this model's training and validation accuracy and loss over time.

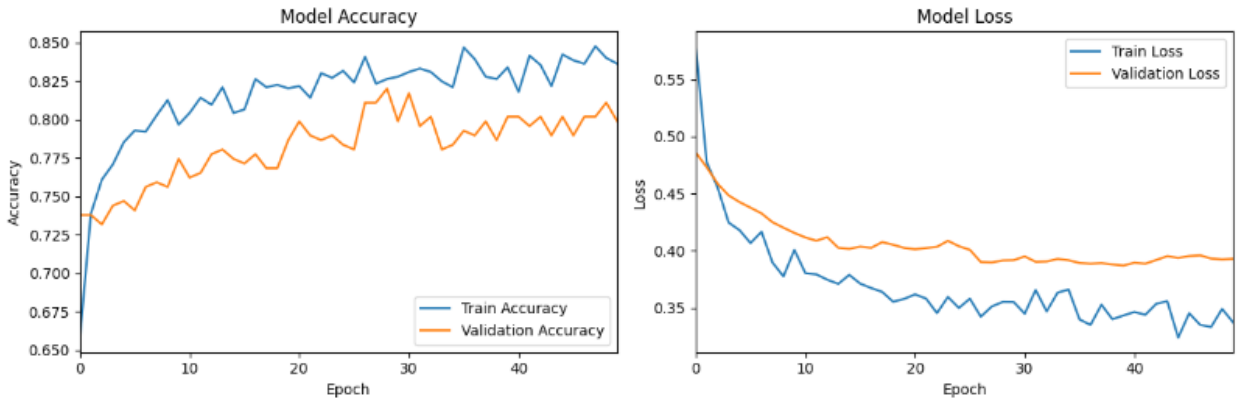

Figure 2: Training time and Validation Accuracy for EfficientNetV2B0 for 50 Epochs. This graph shows this model's training and validation accuracy and loss over time.

The second pair of graphs displays the training and validation loss for both models throughout the 50 epochs. While epochs are represented on the x-axis, the loss values of the models are represented on the y-axis. This indicates the competency the model is with learning, with lower values corresponding to an increase in efficacy. ResNet50 (Figure 1) exhibits a training loss value that decreases more gradually, which indicates a slower but steadier optimization process. The validation loss follows a similar trend with minor fluctuations. EfficientNetV2B0 (Figure 2) shows a quicker reduction in training loss, reflecting efficiency; however, its validation loss stabilizes at a higher level compared to ResNet50. These graphs show the difference between each model in their capabilities in the generalization to new and unseen data, specifically the validation set.

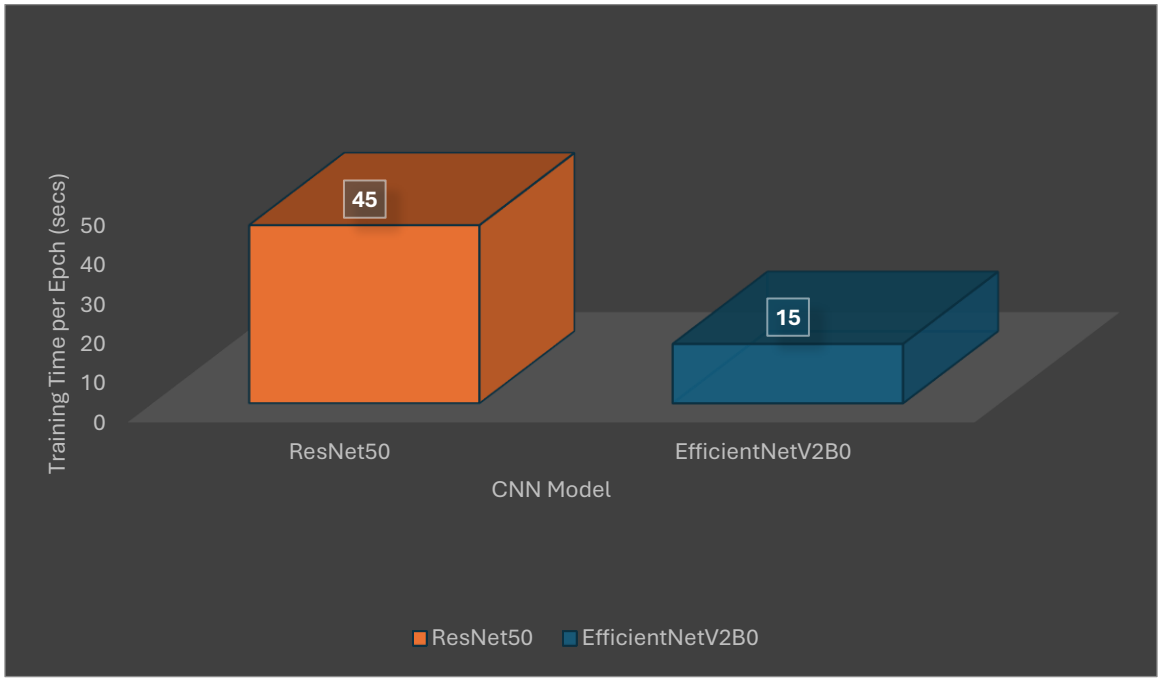

Table 1: Average Training Time per Epoch (secs) for ResNet50 and EfficientNetV2B0. This graph demonstrates the computational efficiency of each model.

The analysis of the training per time per epoch is crucial to understanding the computational efficiency of each model. The results, presented in Figure 3, illustrate the average time each model requires to complete one epoch. ResNet50 consistently shows a higher time per epoch compared to EfficientNetv2B0, emphasizing an increase in computational demand.

5.4 DISCUSSION

The results presented in the previous section provide an in-depth overview of the performance metrics obtained from training and evaluation ResNet50 and EfficientNetV2B0 using the FaceForensics++ deepfake dataset. The aim is to correlate these findings to the research questions and offer further insight into how the models' behavior and performance can inform future approaches to deepfake detection.

The trends found in training and validation accuracy indicate that ResNet50 achieved a higher final validation accuracy of 88.41% compared to EfficientNetV2B0, which reached up to 79.88%. The higher accuracy suggests that ResNet50 is better suited for deepfake detection in scenarios where the priority is achieving a high classification performance. The apparent higher performance can be considered as a bit of a misleading indicator because of the slower rate of learning that ResNet50 has, as evidenced by the gradual increase in training accuracy. This behavior implies that while ResNet50 is an effective model, it requires more epochs to fully optimize, which could become a restriction in time-sensitive applications.

In contrast, EfficientNetV2B0 demonstrates a lower peak validation accuracy while showcasing a rapid increase in training accuracy and a quicker reduction in training loss (Figure 2). This indicates of the fact that this model is significantly more efficient than

ResNet50 in terms of training time and learning speed. The model's quicker convergence suggests that this would be more suitable for scenarios where training time and computational resources are constrained. However, this accentuates the tradeoff of this model: it sacrifices a certain degree of accuracy, which would impact its reliability factor in the application where the precision of deepfake detection is a crucial element.

The training and validation loss curves in both graphs further emphasize the generalization capabilities of both models. ResNet50's consistent reduction in training and validation loss indicates that it generalizes adequately, with fewer signs of overfitting. This finding supports the interpretation that ResNet50 will maintain a substantial performance on unseen data, an essential attribute for deepfake detection tasks that will often involve sporadic and diverse manipulations. On the other hand, the validation loss of EfficientNetV2B0 stabilizes at a much higher level. This suggests that while the model learns quicker, it does not generalize as effectively as ResNet50. This limiting factor must be considered in applications where the detection model will be applied to different situations.

The training time per epoch results reveal an interesting aspect of model performance. ResNet50 requires significantly more time per epoch, reflecting a higher complexity and demand for resources. For environments where processing speed and resource efficiency are paramount, EfficientNetV2B0 will be an essential contender. However, the lower

final validation accuracy must be weighed against this efficiency. For example, in real-time detection systems, the quicker training and validation times of EfficientNetV2B0 would be favorable but there would be a risk of reduced detection reliability.

The interpretations of these results involve acknowledging several assumptions and constraints that may impact their broader applicability. ResNet50's higher accuracy and consistent loss reduction makes it viable candidate for high-accuracy requirement scenarios while EfficientNetV2B0's efficiency and speed offer advantages in resource-limited environments. These interpretations contribute to answering the research questions by delineating when and why each model will be preferred over the other depending on specific metrics and the constraints of the applications. The findings also exhibit the importance of continued research to optimize deepfake detection model further, balancing efficacy and adaptability, while contemplating the evolution of deepfake technology.

6. CONCLUSION

The findings presented by this thesis suggests that applying CNNs for deepfake detection, through administering a comparative analysis of specific architectures like ResNet50 and EfficientNetV2B0, provides valuable insight into addressing the evolving challenges posed by malicious deepfake technology. This topic should be considered significant due to the rapid proliferation of deepfakes and the implications for cybersecurity, societal trust and media authenticity. The focus of this work lies in contributing to the advancement of detection methodologies, specifically refining performance metrics and improving model scalability to develop and enhance the practical implementation of deepfake detection systems.

The foremost problem addressed was how to develop a proper comparison and evaluate CNNs for their efficacy in detecting deepfakes. Contrary to existing approaches that rely on a single model or use multiple models without addressing their performance, this research explores the relative strengths and weaknesses of two state-of-the-art CNNs in tandem with the FaceForensics++ dataset. This methodology employs real-world considerations such as computational efficiency and resource consumption. By examining these factors, this research contributes to the reconciliation of theoretical advancements and practical applicability in deepfake detection.

The main findings of this research demonstrated that while ResNet50 exceptionally outperformed EfficientNetV2B0 in terms of accuracy, the latter was more efficient in terms of time and resources. These results highlight the importance of tailoring and refining CNN architectures based on the specific requirements of detection systems. Furthermore, the methodology of this study involved meticulous training, validation and an evaluation of CNNs with multiple performance metrics to provide a replicable framework for future projects.

The work accomplished is demonstrably significant as it identifies important questions about model scalability, evaluation consistency and implementational practicality. These are challenges that will hinder real world deployment and by providing this comparative analysis, researchers will be informed enough to conduct higher quality research and employ superior decision making when deploying deepfake detection systems. The recognition of these areas for improvement ensures that this research will serve as a foundation for future innovations within this field.

7. APPENDIX

I. Installing a proper IDE like PyCharm IDE

Step 1: Download PyCharm

a) Visit the official PyCharm website: https://www.jetbrains.com/pycharm/.

b) Click Download and select the appropriate version for your operating system:

c) Community (free, sufficient for this project).

Step 2: Install PyCharm

a) Run the downloaded installer file.

b) Follow the on-screen instructions to complete the installation:

c) Select installation path.

d) Opt for adding PyCharm to the system path (recommended).

e) Check the box for creating desktop and taskbar shortcuts.

Step 3: Configure PyCharm

a) Launch PyCharm and select "New Project".

b) Choose Python Interpreter during project creation. If Python is not installed, PyCharm will prompt you to download and configure it. Select Python version 3.8 or later.

II. Installing Python Libraries

Step 1: Install TensorFlow

a) EfficientNetV2B0 and ResNet50 both rely on TensorFlow. To install it:

b) Open the terminal within PyCharm or your system terminal.

c) Run the following command:

d) bash

e) pip install tensorflow

f) Verify installation by running:

g) bash

h) python -c "import tensorflow as tf; print(tf.__version__)"

Step 2: Install EfficientNetV2 Library

a) EfficientNetV2 is included in TensorFlow's Keras Applications, but you may need an additional library for pre-trained weights:

b) Install the EfficientNet package:

c) bash

d) pip install -U efficientnet

Step 3: Install Required Libraries for ResNet50

a) ResNet50 is also supported through TensorFlow's Keras Applications. Ensure that you install the following:

b) NumPy and other essential packages:

c) bash

d) pip install numpy matplotlib

III. Setting Up the Project

Step 1: Clone or Download Project Files

a) If you are using a pre-configured repository for your project (e.g., ResNet50 GitHub repository), clone it using:

b) bash

c) git clone https://github.com/nachiket273/pytorch_resnet_rs

Step 2: Create a Virtual Environment (Optional but Recommended)

a) In PyCharm, go to File > Settings > Project > Python Interpreter.

b) Click Add Interpreter > New Virtual Environment.

c) Specify the path for the virtual environment and click Create.

d) Install the required libraries within this environment

e) bash

f) pip install tensorflow efficientnet numpy matplotlib

IV. Running the Models

Step 1: Set Up Scripts

a) Ensure the main training and evaluation scripts are present in your project folder.

b) Install the FaceForensics++ database from the proper website: https://www.kaggle.com/datasets/hungle3401/faceforensics

c) Unzip the folder and record where it saved.

d) Modify paths in the correct files to point to the FaceForensics++ dataset (EX: stored at C:\Users\Jeet\Documents\FF++).

Step 2: Run the Scripts

a) Open the desired Python file in PyCharm.

b) Click the green play button in the top-right corner to execute the script.

c) Monitor outputs in the PyCharm terminal to ensure successful execution.

V. Troubleshooting

"Module not found": Ensure all libraries are installed in the active Python environment.

a) bash

b) pip install <missing_library>

TensorFlow errors due to GPU: If using a GPU, install GPU-compatible TensorFlow and drivers:

a) bash

b) pip install tensorflow-gpu

8. REFERENCES

1] L. B. Y. B. a. P. H. Yann LeCun, "Gradient Based Learning Applied to Document Recognition," *Proceedings of the IEEE,* vol. 86, no. 11, pp. pp. 2278-2324, 1998.

[2] W. F. X. D. E. D. C. A. S. T.-Y. L. J. S. B. Z. Irwan Bello, "Revisiting ResNets: Improved Training and Scaling Strategies," *Advances in Neural Information Processing Systems,* 2021.

[3] S. T. E. R. L. Bismi Fathima Nasar, "A Survey on Deepfake Detection Techniques," *International Journal of Computer Engineering in Research Trends,* vol. 7, no. 8, pp. 49-55, 2020.

[4] T. Hwang, "Deepfakes: A Grounded Threat Assessment," Center for Security and Emerging Technology, Georgetown, 2020.

[5] A. Naitali, M. Ridouani, F. Salahdine and Kaabouch, "Deepfake Attacks: Generation, Detection, Datasets, Challenges, and Research Directions," *Computers,* vol. 10, no. 12, p. 216, 2023.

[6] C. R. a. M. S. M. Matern, "Exploiting visual artifacts to expose deepfakes and face manipulations," *EEE Winter Conference on Applications of Computer Vision (WACV),* pp. 83-92, 2019.

[7] Y. L. a. S. L. X. Yang, "Exposing deep fakes using inconsistent head poses," ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 8261-8265, 2019.

[8] J. C. Nela Petrzelkova, "Detection of Synthetic Face Images: Accuracy, Robustness, Generalization," arXiv, 2024.

[9] K. Z. X. R. S. &. S. J. He, "Deep residual learning for image recognition," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,* pp. 770-778, 2016.

[10] C. W. Z. Z. Y. Z. H. L. Z. Z. Y. S. T. H. J. M. R. M. M. L. A. S. Hang Zhang, "ResNeSt: Split-Attention Networks," arXiv preprint arXiv:2004.08955, 2020.

[11] M. Tan, "Efficientnet: Rethinking model scaling for convolutional neural networks," *International Conference on Machine Learning,* pp. 1-11, 2019.

[12] M. a. Q. L. Tan, "Efficientnetv2: Smaller models and faster training," *International conference on machine learning,* pp. pp. 10096-10106, 2021.

[13] J. P.-A. ,. M. M. B. X. D. W.-F. S. O. ,. A. C. Y. B. Ian J. Goodfellow∗, "Generative Adverserial Nets," *Advances in Neural Information Processing Systems,* vol. 27, 2014.

[14] P. K. a. S. Marcel, "DeepFakes: A New Threat to Face Recognition? Assessment and Detection," *IEEE Signal Processing Magazine,* vol. 36, no. 1, pp. 20-27, 2019.

[15] H. H. S. A. I. K. I. A. M. Samer Hussain Al-Khazraji, "Impact of Deepfake Technology on Social Media: Detection, Misinformation and Societal Implications," *The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM),* vol. 23, pp. 429-411, 2023.

[16] I. S. a. G. E. H. A. Krizhevsky, "ImageNet classification with deep convolutional neural networks," *Proc. Advances in Neural Information Processing Systems (NIPS),* pp. pp. 1097-1105, 2012.

[17] Y. B. a. G. H. Y. LeCun, "Deep learning," *Nature,* vol. 521, pp. 436-444, 2015.

[18] F. L. W. Y. S. P. a. J. Z. Zewen Li, "A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects," *IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS,* vol. 99, pp. 1-21, 2021.

[19] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* pp. 1251-1258, 2017.

[20] V. D. R. R. A. A. S. S. C. Y. P. N. S. A. Sanjay A. Agrawal, "Advancements in NSFW Content Detection: A Comprehensive Review of ResNet-50 Based Approaches," ResearchGate, 2023.

[21] M. Al-Gaashani, N. Samee, R. Alnashwan, M. Khayyat and M. Muthanna, "Using a Resnet50 with a Kernel Attention Mechanism for Rice Disease Diagnosis," *Life,* vol. 13, p. 1277, 2023.

[22] S. Tummala, V. Thadikemalla, S. Kadry, M. Sharaf and H. Rauf, "EfficientNetV2 Based Ensemble Model for Quality Estimation of Diabetic Retinopathy Images from DeepDRiD," *Diagnostics,* vol. 13, p. 622, 2023.

[23] N. J. B. P. Shwetambari Borade, "ResNet50 DeepFake Detector: Unmasking Reality," *iNDIAN JOURNALOFSCIENCEANDTECHNOLOGY,* vol. 13, no. 12, pp. 1263-1271, 2024.

[24] A. P. Alexey Egorov, "EfficientNets for DeepFake Detection: Comparison of Pretrained Models," *IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus),* pp. 598-600, 2021.