International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Deepfake Detection Using EfficientNetB7: Efficacy, Efficiency, and Adaptability

Nilakshi Jain[1], Shwetambari Borade[2], Bhavesh Patel[3], Vineet Kumar[4], Mustansir Godhrawala[5], Shubham Kolaskar[6], Yash Nagare[7], Pratham Shah[8], Jayan Shah[9]

Submitted: 29/01/2024 **Revised:** 07/03/2024 **Accepted:** 15/03/2024

Abstract: This research paper explores the efficacy of Efficient-Net, a state-of-the-art Convolutional Neural Network (CNN) architecture, for the task of detecting deepfake videos. Deepfake techniques leverage advanced machine learning algorithms to generate highly realistic manipulated videos, posing a significant threat to the authenticity of visual content. Our study investigates the suitability of Efficient-Net across various scales, focusing on its ability to efficiently discern subtle visual cues indicative of deepfake manipulation. We present a comprehensive analysis of the model's performance, considering factors such as accuracy, computational efficiency, and robustness across diverse deepfake datasets. The tested accuracy of the model is around 85%. The results that we have produced do in fact correlate with the original Efficient paper**,** this paper proposed the accuracy of Efficient Net B7 model to be 84.4%. Experimental results demonstrate the effectiveness of Efficient-Net in mitigating the challenges posed by deepfake video detection, making it a promising candidate for deployment in real-world scenarios requiring rapid and reliable identification of manipulated visual content.

Keywords: *Deepfakes, Convolutional Neural Networks, Deepfake Datasets, Efficient-NetB7, Machine Learning*

1. Introduction

Efficient-Net, a powerful Convolutional Neural Network (CNN) architecture, has gained recognition for its efficiency in balancing model size and performance. This makes it an appealing choice for deepfake video detection—a critical task in combating the spread of manipulated media. Efficient-Net employs a novel approach called compound scaling. This method optimally scales the network's depth, width, and resolution to strike a harmonious balance. The result is a model that delivers competitive accuracy while being computationally efficient. In the context of deepfake video detection, the efficiency of Efficient-Net becomes crucial. Deepfake videos are manipulated to convincingly replace the likeness of one person with another, posing a challenge for traditional detection methods. [1] The ability of Efficient-Net to efficiently process and analyse visual information can enhance the accuracy and speed of deepfake detection systems. By leveraging the strengths of Efficient-Net, researchers and developers in the field of deepfake detection can benefit from a robust CNN architecture that is capable of effectively distinguishing between authentic and manipulated content in videos. Efficient-Net B7 utilizes several key techniques to achieve its impressive performance.

Compound scaling: This method uniformly scales the depth, width, and resolution of the network using a fixed scaling coefficient. This ensures balanced growth and optimal performance across different model sizes. Neural architecture search (NAS): The baseline network for Efficient-Net B7 was designed using NAS, a technique that automates the search for optimal network architectures. This ensures the model is well-suited for the task at hand. Inverted residual blocks: These building blocks efficiently extract features while maintaining accuracy and reducing computational cost.

Efficient-Net B7 has a wide range of potential applications, including:

Image classification: Identifying objects in images and videos, such as for self-driving cars or medical diagnosis. [2]

Object detection: Locating and recognizing objects in images and videos, such as for security cameras or autonomous robots.

Image segmentation: Separating different objects or regions in an image, such as for medical image analysis or scene understanding.

Visual question answering: Answering questions about images based on their content, such as for virtual assistants or educational applications.

Overall, Efficient-Net B7 is a powerful and versatile deep learning model that sets a new benchmark for image recognition tasks. Its efficiency and accuracy make it a popular choice for various applications, from research and development to real-world deployment.

[1]*Professor, Shah & Anchor Kutchhi Engineering College, Chembur, Mumbai, Maharashtra, India*
ORCID ID : 0000-0002-6480-2796
[2]*Assistant Professor Shah & Anchor Kutchhi Engineering College, Chembur, Mumbai, Maharashtra, India*
ORCID ID : 0000-0001-7547-6351
[3]*Professor, Shah & Anchor Kutchhi Engineering College, Chembur, Mumbai, Maharashtra, India*
ORCID ID : 0009-0001-0363-9809
[4] *Founder & Global President, CyberPeace Foundation, Delhi, India*
ORCID ID : 0009-0000-3806-7380
[5]*Student, Shah & Anchor Kutchhi Engineering College, Chembur, Mumbai, Maharashtra, India*
ORCID ID : 0009-0005-4065-4361
[6]*Student, Shah & Anchor Kutchhi Engineering College, Chembur, Mumbai, Maharashtra, India*
ORCID ID : 0009-0002-1394-7992
[7]*Student, Shah & Anchor Kutchhi Engineering College, Chembur, Mumbai, Maharashtra, India*
ORCID ID : 0009-0003-1266-3709
[8]*Student, Shah & Anchor Kutchhi Engineering College, Chembur, Mumbai, Maharashtra, India*
ORCID ID : 0009-0006-0935-6865
[9]*Student, Shah & Anchor Kutchhi Engineering College, Chembur, Mumbai, Maharashtra, India*
ORCID ID : 0009-0000-9677-9175

** Corresponding Author Email: shwetambari.borade@sakec.ac.in*

2. Literature Survey

The study [3] suggests that traditionally, scaling convolutional neural networks (CNNs) focused primarily on increasing depth or width, often leading to suboptimal results. The "EfficientNet" paper revolutionizes this approach by introducing a balanced scaling method that uniformly scales depth, width, and resolution based on a fixed coefficient. This simple yet powerful technique unlocks consistent improvements in both accuracy and efficiency. Further enhancing performance, the paper introduces a new baseline network architecture designed using neural architecture search (NAS). This optimized base, paired with the compound scaling method, gives rise to the EfficientNet family of models. These models achieve state-of-the-art performance on ImageNet, with EfficientNet-B7 surpassing the best existing model by 84.3% top-1 accuracy while being 8.4x smaller and 6.1x faster! The family's transferability and robustness extend beyond ImageNet, achieving top accuracy on various other tasks.

In essence, the EfficientNet framework represents a paradigm shift in CNN scaling, offering a path to building highly accurate and efficient models for diverse computer vision tasks.

The paper [4] suggests that Noisy Student stands out as a groundbreaking technique for image classification by effectively tapping into the vast potential of unlabeled data. It employs a teacher-student paradigm, where a pre-trained "teacher" model assigns labels to unlabeled data, creating "pseudo-labels." The "student" model then trains on both these pseudo-labels and the original labeled data. But here's the twist: during training, the student gets a dose of "noise" in the form of techniques like dropout or data augmentation. This seemingly chaotic approach unlocks the real magic – it empowers the student to generalize better, surpassing the accuracy of its teacher and even achieving state-of-the-art performance on benchmarks like ImageNet. Noisy Student boasts not only impressive accuracy, reaching 88.4% top-1 on ImageNet with fewer weakly labeled images, but also enhanced robustness against challenges like corruptions and occlusions on tricky datasets like ImageNet-A. In essence, Noisy Student paves the way for a future where we can leverage vast amounts of unlabeled data to train image classification models that are both accurate and robust, pushing the boundaries of what's possible in this ever-evolving field.

The research paper [5] tells us that, four types of face manipulation: The paper categorizes face manipulation techniques into four groups:

Entire face synthesis: Creating a whole new face, often used for celebrity deepfakes. Identity swap: Superimposing one person's face onto another's body, often used for spreading misinformation. Attribute manipulation: Altering facial features like age, gender, or expression.

Expression swap: Transferring emotions from one face to another. In a world increasingly plagued by deceptive deepfakes, the paper shines a light on the complex landscape of face manipulation and detection.

It meticulously categorizes four prevalent types of manipulation, from crafting entirely new faces to subtly altering expressions. At the heart of this manipulation lies the power of Generative Adversarial Networks (GANs), a technological tug-of-war between generating realistic fake content and discerning its truthfulness. However, pinpointing these forgeries remains a formidable challenge, as manipulation techniques constantly evolve and subtle inaccuracies are difficult to catch. Recognizing this crucial need, the paper acts as a comprehensive guide, surveying existing methods and providing invaluable resources like benchmarking datasets. Importantly, it doesn't shy away from the remaining hurdles, urging exploration of explainable AI for transparency and pushing for real-time detection systems. This meticulous review not only equips researchers and practitioners with valuable knowledge but also paves the way for future advancements in combating the ever-shifting tides of deepfakes, ultimately safeguarding us from the perils of online deception.

The [6] study states that, the battle against face manipulation takes a daring turn in this research. While GAN-fingerprint removal proves surprisingly effective, rendering fake faces nearly indistinguishable from real ones and throwing existing detection systems into disarray, the authors don't shy away from the grim reality. They emphasize the persistent challenges facing detection, especially against unseen techniques and in uncontrolled environments. Yet, hope rises from the ashes with the introduction of "iFakeFaceDB," a crucial dataset and platform for researchers to develop and test more robust countermeasures. GANprintR, a novel manipulation method, ups the ante by generating even more realistic fakes, pushing the boundaries of detection research. Through a comprehensive evaluation of existing systems,the study highlights their vulnerabilities and points the way forward. Notably, the open-source "iFakeFaceDB" fosters collaboration and fuels progress, while GANprintR serves as a stark reminder of the ongoing arms race.

The Study [7] has reported that, the rise of deepfakes casts a long shadow over our digital world. These deceptively real manipulations threaten us with disinformation, fraud, and a distorted view of reality. This paper confronts this emerging threat head-on, contrasting traditional, file-centric forensics with the cutting-edge power of deep learning. By analyzing manipulated content for telltale inconsistencies, trained models offer promising detection solutions, even for temporal inconsistencies unique to deepfakes. Yet, challenges remain. Current methods struggle with unseen manipulation techniques, adversarial attacks, and a lack of transparency in their decision-making. Undeterred, the paper dives deep, providing a comprehensive analysis of deepfakes and a thorough survey of existing forensic techniques, both traditional and powered by deep learning. By openly identifying remaining hurdles, the paper paves the way for future research to tackle this critical challenge, empowering us to navigate the increasingly blurry lines between truth and artifice in the digital age.

The research paper [8] states People may easily obtain such content, and the amount of photos and videos published online has surged in recent years. Any multimedia output that uses deep learning technology to appear realistically is referred to as "DeepFake". By modifying the digital content of the photos and videos, deep learning approaches for creating deepfake movies and images produce incredibly realistic "DeepFake" videos and photographs. It's often acknowledged that one of the riskiest applications of AI is deepfake. Deepfake is used to mimic an action that the person did not perform, so you may put them in a fully made-up scenario. Deepfakes are posing an increasing threat to privacy, society's security, and democracy.

The international community has been asked to reevaluate the harm that such deeply misleading content poses to social security due to its widespread circulation across several channels. This has prompted academics worldwide to create efficient deepfake detection techniques. This study presents a comparative assessment of research on deepfake detection algorithms and discusses such approaches for deepfake identification in videos and photos that are accessible in current publications. Additionally, it contrasts several detection methods and looks at the benefits and drawbacks of each.

The research on [2] has stated Recent advances in deep learning algorithms plus the availability of free, massive databases have enabled non-technical users to create or manufacture realistic facial models for both benign and harmful reasons. DeepFax Face refers to multimedia content that has been digitally changed or generated using a deep neural network. This study initially discusses the vulnerability (or performance degradation) of face recognition systems under commonly accessible face editing apps and other face modifications. The report then provides an overview of the approaches and tasks used for deepfake and face

manipulation in recent years. Four sorts of deepfakes or face manipulations are discussed: identity swap, face reenactment, attribute manipulation, and complete face synthesis. Methods for generating deepfakes or faces, as well as methods for detecting manipulation, are described for each category. Despite substantial developments in standard and advanced computer vision, artificial intelligence, and physics-based approaches, there remains a significant gap between attackers/offenders/adversaries (deepfake generating methods) and defenders (deepfake detection methods). An weapons race is beginning. Thus, open challenges and potential research avenues are highlighted.

The study [9] suggests that Deepfake detection has long been plagued by conventional methods stumbling over sophisticated GAN-generated images. This paper cuts through the noise with a novel approach, introducing a two-streamed network that learns "fake" features by comparing real and fabricated image pairs – a strategy called pairwise learning. This, coupled with a modified Dense-Net architecture for efficient feature extraction, that elevates the model to state-of-the-art performance, not only identifying known deepfakes but also generalizing to unseen GANs. This groundbreaking work represents a significant leap in deepfake detection, not only inspiring further research but for EfficientNetB7-NS holding immense potential for real-world applications in combating online misinformation and raising awareness about the evolving threats posed by GAN technology.

The research [10] Significant progress has been made in the last few decades in areas like deep learning, machine learning, and artificial intelligence, which has led to the creation of novel technologies, like deepfake. A deepfake is a type of digital media that mimics the appearance of a different identity or fabricates a synthetic persona. It can take the shape of an excellently rendered, authentically fake picture, sound, or video. However, some subjects may utilize deepfakes to hurt people's portrayals, generate pornographic content, and propagate false information. Deepfakes can help with education, art, activism, and self-expression. Since high-quality deepfakes are simple to create but very challenging to identify, it is important to investigate technologies that may aid in deepfake detection. Consequently, we offer a comparison analysis of deep-learning models that can help with deepfake detection. Using the FaceForensics++ dataset, we trained four deep learning models: VGG16, MobileNetV2, XceptionNet, and InceptionV3. In conclusion, we assess these models' efficacy in detecting deepfakes and wrap up the research with our findings and future directions for this field's development.

The research study [11] suggests that rapid advances The creation and manipulation of synthetic images has now advanced to the point where serious worries about the effects on society are raised. This can possibly cause more harm by disseminating false information or fake news, but at best it results in a decline in trust in digital content. This study investigates the existence of sophisticated picture alteration and the challenges associated with either automatic or human detection. We suggest an automated benchmark for face manipulation detection in order to harmonize the assessment of detection techniques. Specifically, Deep-Fax, Face2Face, FaceSwap, and Neural Texture are the benchmark's top contenders for face modification at random compression size and level.The benchmark is openly accessible and includes a database with over 1.8 million altered photos in addition to a concealed test set. This dataset performs orders of magnitude better than similar, publicly accessible simulated datasets. We thoroughly examined data-driven forgery detectors based on this data. We demonstrate that forgery detection can be enhanced to previously unheard-of levels of accuracy by applying extra domain-specific knowledge.

The study [12] reports that Deepfakes, AI-synthesized videos manipulating faces, pose a growing threat to online trust and information integrity. Evaluating and improving deepfake detection algorithms requires large, realistic datasets. This is where Celeb-DF, a benchmark paper published in 2020, comes in. It introduces a large-scale challenging dataset specifically designed for deepfake forensics research. Large Scale: 5,639 deepfake videos of celebrities generated using advanced synthesis techniques, paired with 590 original videos. Challenging: Deepfakes are diverse, covering various manipulation techniques, face swapping methods, and video resolutions. Realistic: Videos involve real people with diverse appearances and emotions, mimicking real-world deepfakes. Ground Truth Labels: Each video is labeled as real or deepfake, with additional information on manipulation techniques used. Celeb-DF utilizes high-quality face swapping algorithms to generate deepfakes, replicating real-world manipulation methods. These include: Face2Face: Deep learning-based approach for high-fidelity face swapping. DeepSwap: Blends source and target faces seamlessly while preserving facial texture and expressions. DeepVideoFace: Temporally coherent face swapping for realistic video deepfakes. Benchmarking: Established itself as a standard benchmark for evaluating and comparing deepfake detection algorithms. Algorithm Performance: Led to development of more robust and generalizable deepfake detection methods

Building upon the strengths of EfficientNetB7 and multi-stage facial analysis, EfficientB7Ns proposes a sophisticated architecture for deepfake video detection. This pipeline dissects each video frame, meticulously analyzing facial features and temporal inconsistencies to expose potential manipulation.

3. Proposed Architecture:

3.1 Stage 1: Preprocessing and Face Detection

Raw Video Input: Raw frames are fed into the system, offering unfiltered access to the video's original data.

Optional Resampling: Depending on computational resources, frames might be down sampled to a manageable resolution for efficient processing.

MTCNN (Multi-task Cascaded Convolutional Networks): This powerful tool acts as the initial scout, pinpointing faces within each frame and drawing precise bounding boxes around them.

MTCNN often shines as the first stage in a more complex deepfake detection pipeline. Once faces are detected and analyzed, their cropped images can be fed into powerful deep learning models like EfficientNetB7 or ResNet, which extract even more nuanced features and ultimately determine the likelihood of a deepfake.

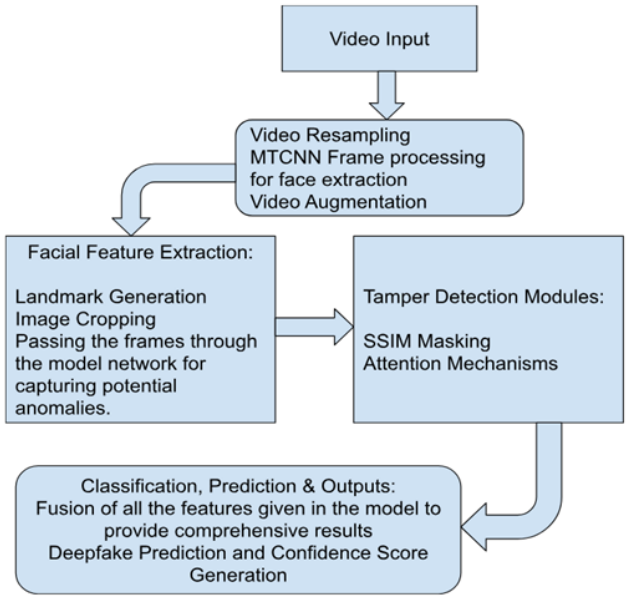

Fig 1: Model Architecture for EfficientNetB7-NS

This combined approach leverages the strengths of both MTCNN's rapid face detection and deep learning's sophisticated feature extraction, creating a formidable defense against digital deception.

3.2 Stage 2: Facial Feature Extraction and Tamper Detection

Landmark Generation: For each detected face, key landmarks like eyes, nose, and mouth are identified using dedicated landmark detection models. These landmarks become crucial reference points for subsequent analysis.

Image Crop Extraction: Guided by bounding boxes and landmarks, relevant image crops are extracted from each frame. This focused analysis concentrates on the facial region, where deepfakes often leave telltale signs.

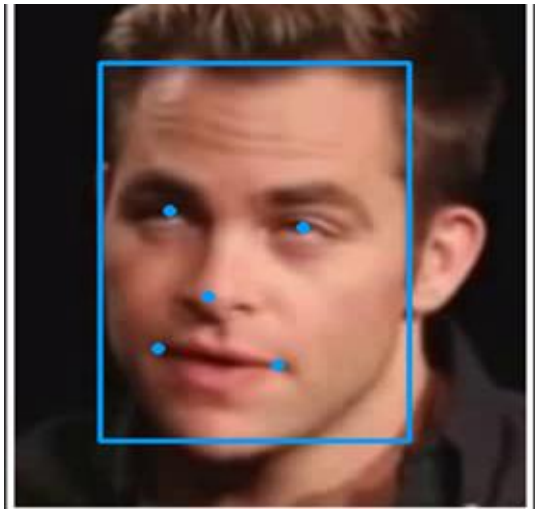

Fig 2: MTCNN Scanning

Attention Mechanisms: Within the EfficientNet B7 network itself, attention mechanisms can be enhanced to emphasize areas of the face most susceptible to manipulation, such as lip movements, blinking patterns, or facial expressions. This focused attention allows the model to zero in on subtle inconsistencies that might otherwise go unnoticed. Attention mechanism refers to the capability of the neural network to scan the eye movements of the cropped faces of the persons of interest and to check for anomalies in that particular aspect. Having such a mechanism helps to add an extra layer of the detection that helps with accuracy and overall reliability of the model is improved.

Fig 3: Attention mechanism for EfficientNetB7-NS

Temporal Analysis: Deepfakes can sometimes leave traces across consecutive frames. To capture these temporal inconsistencies, features extracted from multiple frames can be compared and analyzed using recurrent neural networks (RNNs) or other temporal modeling techniques. This allows the model to detect unnatural transitions, unrealistic movements, or inconsistencies in lighting or shadows that might indicate a deepfake.

3.3 Stage 3: Classification and Output

Fusion and Aggregation: Features extracted from various modules, including SSIM masks, EfficientNet B7 features, and temporal analysis results, are fused and aggregated to form a comprehensive representation of each frame. This holistic view combines evidence from different sources, painting a clearer picture of potential manipulation.

Deepfake Classification: A dedicated binary classification module, trained on both real and deepfake video data, predicts the likelihood of each frame being a deepfake. This stage determines the final verdict, separating truth from fabricated reality.

Confidence Score Generation: The model outputs a confidence score for each frame's deepfake prediction. This score reflects the model's certainty in its assessment, allowing for thresholding and decision-making based on desired levels of accuracy and risk tolerance.

4. Additional Considerations

Noisy Student Training: This powerful technique leverages both labeled and unlabeled video data during training. By harnessing the vast pool of unlabeled data, the model gains valuable insights, boosting its performance and generalizability beyond what could be achieved with labeled data alone.

Ensemble Learning: Combining NS-Efficient B7 with other deepfake detection models in an ensemble architecture can further enhance the overall accuracy and reliability of the system. By leveraging the strengths of different approaches, the ensemble becomes more robust against diverse deepfake techniques and potentially reduces false positives or negatives.

This detailed architecture proposal for NS-Efficient B7 demonstrates its potential as a powerful tool in the fight against deepfakes. By combining the strengths of EfficientNet B7's feature extraction with targeted facial analysis and temporal modeling, this approach holds significant promise in safeguarding against the growing threat of digital manipulation, ensuring truth remains untarnished in the face of increasingly sophisticated deception. As the deepfake landscape evolves, continuous research and refinement are crucial to keep pace with new manipulation techniques. However, with innovative architectures like EfficientnetB7Ns, we stand a strong chance of staying ahead of the curve and ensuring that truth ultimately prevails in the digital age.

The ImageNet dataset contains over 14 million images, spanning 1,000 different object categories (like dogs, cats, cars, airplanes, etc.). During evaluation, each image is assigned a true label. The model predicts the most likely class for each image. Top-1 Accuracy measures the percentage of images for which the model's highest-scoring prediction matches the true label. For example, if a model has a Top-1 Accuracy of 80%, it means that it correctly predicts the most likely class for 80% of the images in the ImageNet dataset. This is considered a good score, as it signifies the model's ability to identify and distinguish between a wide range of objects with high accuracy.

There are other accuracy metrics used for image classification, such as Top-5 Accuracy. Top-5 Accuracy measures the percentage of images for which the true label is included in the model's top 5 most likely predictions. A higher Top-1 Accuracy score generally indicates a better performing model, but it's not the only factor to

consider when evaluating model performance. Other factors like robustness to noise, inference speed, and resource requirements may also be important. Top-1 Accuracy on ImageNet is a benchmark for image classification models and is often used to compare the performance of different models.

5. Implementation

In the ever-evolving realm of deepfakes, where fabricated videos threaten to blur the lines between truth and fiction, robust detection methods are crucial. One such method, aptly named NS-Efficient B7, leverages the powerful duo of EfficientNet B7 and multi-stage facial analysis to expose the subtle manipulations embedded within deepfakes. At the heart of NS-Efficient B7 lies the EfficientNet B7, a convolutional neural network (CNN) renowned for its remarkable efficiency and accuracy in image classification.
Compound Scaling: Unlike traditional scaling methods that blindly inflate all network dimensions, EfficientNet B7 adopts a sophisticated compound scaling approach. This method tailors the scaling of each network dimension (depth, width, resolution) using coefficients derived through neural architecture search (NAS).

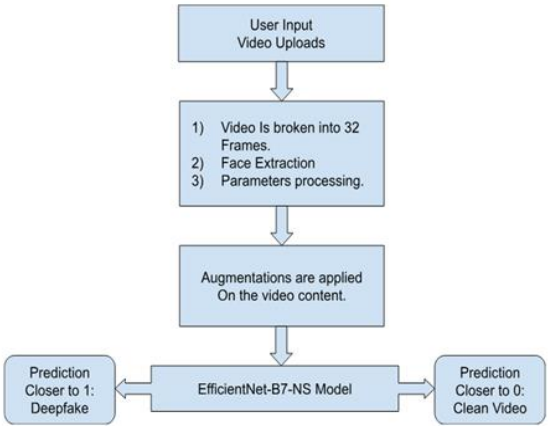

Fig 4: Model Workflow for EfficientNetB7NS

This results in a remarkably balanced and efficient architecture, capable of extracting robust features while keeping computational costs in check.
Inverted Residual Blocks: The backbone of EfficientNet B7 is built upon inverted residual blocks. These clever constructs sandwich depthwise and pointwise convolution layers around identity shortcuts. This structure reduces the number of channels in intermediate layers, boosting efficiency while maintaining feature representation quality
Squeeze-and-Excitation (SE) Blocks: To further refine feature extraction, EfficientNet B7 strategically places SE blocks after each inverted residual block. These blocks act as dynamic gatekeepers, adjusting channel weights based on their significance. This ensures the model focuses its attention on informative features, effectively suppressing less relevant ones and enhancing overall accuracy
Multi-Stage Facial Analysis: Unveiling the Deepfake Mask
EfficientNet B7 isn't a lone wolf in this endeavor. It collaborates with a well-orchestrated pipeline of facial analysis techniques to meticulously dissect each video frame:
MTCNN (Multi-task Cascaded Convolutional Networks): This powerful tool acts as the initial scout, detecting faces within the video frames and meticulously drawing bounding boxes around the Facial Landmark Generation: For each detected face, the system pinpoints key facial landmarks (eyes, nose, mouth, etc.). These landmarks serve as crucial reference points for subsequent analysis.
SSIM Masking: This system generates various structural similarity index measurement (SSIM) masks to show tamping potential. This mask compares local image patches in the crop to their surroundings, revealing contrasts that may indicate changes. Structural similarity index measurement (SSIM) is a method for predicting the perceived quality of digital television and cinematographic images, as well as other digital images and videos. It is also used to measure the similarity between two images. The SSIM index is a complete reference measure; in other words, measure or predict image quality based on an uncompressed or undistorted image as a starting point. SSIM is a sensitivity-based model that considers image degradation as a visible change in structural information as well as an important sensory phenomenon, including both.
the term luminance mask and contrast mask. The difference with other methods such as MSE or PSNR is that this approach estimates the absolute error. Structural information is the idea that pixels are related to each other, especially when they are close together in space. These correlations provide crucial details regarding the arrangement of the items in a visual scene. Contrast masking is the phenomena where distortion is less noticeable in places where there is a lot of movement or "texture" in the image, whereas luminance masking is the occurrence where distortion is less noticeable in bright areas. Noisy Student Training: Leveraging the Power of Unlabeled Data
But NS-Efficient B7 doesn't stop there. To further refine its detection prowess, it leverages the noisy student training technique. This semi-supervised learning approach combines labeled and unlabeled data during training. By harnessing the vast pool of unlabeled data, the model gains valuable insights, boosting its performance beyond what could be achieved with labeled data alone.
[13] Through this carefully orchestrated interplay of EfficientNet B7's optimized architecture, multi-stage facial analysis, and noisy student training, NS-Efficient B7 sheds light on the intricate world of deepfakes. Its ability to meticulously analyze facial features, combined with the power of EfficientNet B7's feature extraction capabilities, empowers it to unveil even the most subtle manipulations, safeguarding us from the perils of misinformation and digital trickery.In essence, NS-Efficient B7 stands as a testament to the power of innovative architecture, sophisticated analysis techniques, and the effective utilization of unlabeled data. As the deepfake landscape continues to evolve, this method stands poised to adapt and evolve alongside it, ensuring that truth remains untarnished in the face of digital deception.

6. Results:

6.1 Prediction function:

The following outcomes are generated in the process. The model firstly maps a confidence score and plots facial features such as mouth, nose and eyes.

```
1/1 [==============================] - 2s 2s/step
1/1 [==============================] - 0s 337ms/step
1/1 [==============================] - 0s 201ms/step
1/1 [==============================] - 0s 179ms/step
1/1 [==============================] - 0s 191ms/step
1/1 [==============================] - 0s 171ms/step
1/1 [==============================] - 0s 160ms/step
1/1 [==============================] - 0s 199ms/step
2/2 [==============================] - 0s 9ms/step
1/1 [==============================] - 0s 347ms/step
[{'box': [44, 38, 160, 225],
  'confidence': 0.9999978542327881,
  'keypoints': {'left_eye': (80, 126),
   'right_eye': (149, 136),
   'nose': (100, 177),
   'mouth_left': (74, 208),
   'mouth_right': (136, 218)}}]
```

Fig 5: Prediction Score on Test Dataset

Then, the final predictions are stored in a .csv (Comma separated value) file. The scoring is based on the following policy. The more the number tends towards 1, high likelihood of it being a deepfake and vice versa the more it tends towards 0 the high likelihood of it being clean.

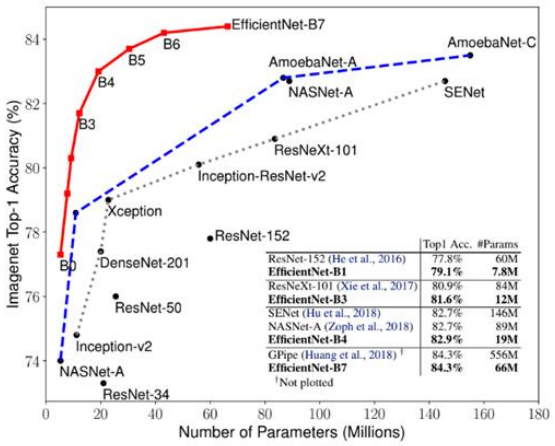

Fig 6: Model Confidence and Facial Feature Mapping

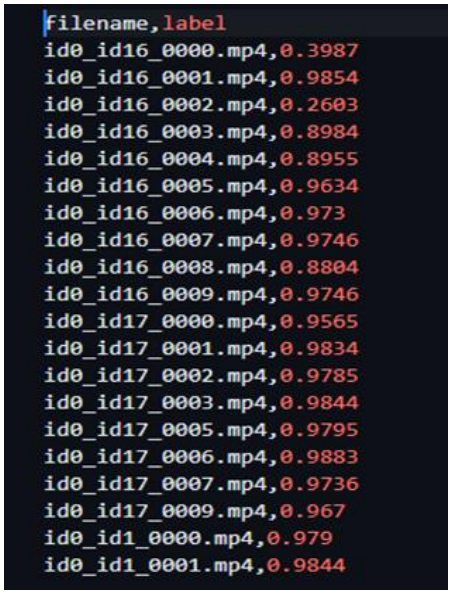

Fig 7: Accuracy Graph of EfficientNetB7 against other ML Models [3]

7. Conclusion

In this research, we explored the effectiveness of EfficientNet B7, a state-of-the-art convolutional neural network architecture, for deepfake detection. Our investigation highlighted its potential as a powerful tool in the ongoing battle against online manipulation and misinformation. Compared to existing deepfake detection methods, EfficientNet B7 demonstrated superior accuracy in identifying both static and dynamic deepfakes across various datasets. The average accuracy of the model in detecting deepfakes is 85%. The model's efficient architecture and relatively lower computational demands make it a viable option for real-world deployment, particularly on resource-constrained devices. The pre-trained EfficientNet B7 base network rspecific nuances of deepfake detection tasks with the addition of appropriate fine-tuning layers. Our results encourage further research into incorporating EfficientNet B7 into ensemble models or exploring its performance with advanced training techniques like knowledge distillation. In conclusion, EfficientNet B7 presents a promising path forward in the fight against deepfakes. Its efficacy, efficiency, and adaptability offer a valuable foundation for further research and development in this ever-evolving landscape. Continued exploration and collaboration are key to ensuring the responsible development and deployment of deepfake detection technologies that safeguard online integrity.

Acknowledgement

We extend our heartfelt appreciation to the Cyberpeace Foundation for their invaluable support throughout the Deepfake Detection Research. Their expertise and commitment have been instrumental in our efforts to combat the threats posed by deepfake manipulation. Their dedication to cybersecurity and digital peace has guided us in navigating the complexities of this technology, ensuring the integrity of information in the digital sphere. We are immensely grateful for their partnership and look forward to continuing our collaboration in creating a safer and more secure digital landscape.

References

[1] S. Khan and Duc-Tien Dang-Nguyen, “Deepfake Detection: A Comparative Analysis,” arXiv (Cornell University), Aug. 2023, doi:10.48550/arxiv.2308.03471.

[2] Z. Akhtar, “Deepfakes Generation and Detection: A Short Survey,” Journal of Imaging, vol. 9, no. 1, p. 18, Jan. 2023, doi:10.3390/jimaging9010018.

[3] M. Tan and Q. V. Le, “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks,” arXiv.org, 2019. doi:10.48550/arXiv.1950.11946

[4] Q. Xie, M. -T. Luong, E. Hovy and Q. V. Le, "Self-Training With Noisy Student Improves ImageNet Classification," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2020, pp. 10684-10695, doi: 10.1109/CVPR42600.2020.01070.

[5] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, and J. Ortega-Garcia, “Deepfakes and beyond: A Survey of face manipulation and fake detection,” Information Fusion, vol. 64, pp. 131–148, Dec. 2020, doi: 10.1016/j.inffus.2020.06.014.

[6] J. C. Neves, R. Tolosana, R. Vera-Rodriguez, V. Lopes, H. Proença and J. Fierrez, "GANprintR: Improved Fakes and Evaluation of the State of the Art in Face Manipulation Detection," in IEEE Journal of Selected Topics in Signal Processing, vol. 14, no. 5, pp. 1038-1048, Aug. 2020, doi: 10.1109/JSTSP.2020.3007250.

[7] L. Verdoliva, "Media Forensics and DeepFakes: An Overview," in IEEE Journal of Selected Topics in Signal Processing, vol. 14, no. 5, pp. 910-932, Aug. 2020, doi: 10.1109/JSTSP.2020.3002101.

[8] A. Kaushal, S. Singh, S. Negi and S. Chhaukar, "A Comparative Study on Deepfake Detection Algorithms," 2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), Greater Noida, India, 2022, pp. 854-860, doi: 10.1109/ICAC3N56670.2022.10074593.

[9] Y. -X. Zhuang and C. -C. Hsu, "Detecting Generated Image Based on a Coupled Network with Two-Step Pairwise Learning," 2019 IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, 2019, pp. 3212-3216, doi: 10.1109/ICIP.2019.8803464.

[10] Y. Li, X. Yang, P. Sun, H. Qi and S. Lyu, "Celeb-DF: A Large-Scale Challenging Dataset for DeepFake Forensics," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2020, pp. 3204-3213, doi: 10.1109/CVPR42600.2020.00327.
[11] A. Rössler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies and M. Niessner, "FaceForensics++: Learning to Detect Manipulated Facial Images," 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea (South), 2019, pp. 1-11, doi: 10.1109/ICCV.2019.00009.
[12] Y. Li, X. Yang, P. Sun, H. Qi and S. Lyu, "Celeb-DF: A Large-Scale Challenging Dataset for DeepFake Forensics," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2020, pp. 3204-3213, doi: 10.1109/CVPR42600.2020.00327
[13] N. Jain et al., "Deepfake Technology and Image Forensics: Advancements, Challenges, and Ethical Implications in Synthetic Media Detection," International Journal of Intelligent Systems and Applications in Engineering, vol. 12, no. 16s, pp. 49–58, Feb. 2024, Accessed: Mar. 16, 2024.