Real vs. Fake Face Classification Using Deep Learning-Based Convolutional Neural Networks

1st Ashen Shanuka
Undergraduate of Department of Industrial management
University of Kelaniya
Sri Lanka
ashenshanuka1222@gmail.com
IM/2019/008

2nd Malshan Weerasinghe
Undergraduate of Department of Industrial management
University of Kelaniya
Sri Lanka
malshanweerasinghe99@gmail.com
IM/2019/009

3rd Tharindu Adikari
Undergraduate of Department of Industrial management
University of Kelaniya
Sri Lanka
tsaadhkari@gmail.com
IM/2019/052

4th Dishan Sandasara
Undergraduate of Department of Industrial management
University of Kelaniya
Sri Lanka
sandasarad@gmail.com
IM/2019/054

***Abstract*—The rapid growth of artificial intelligence (AI) technologies has led to the creation of hyper realistic synthetic images, commonly known as deepfakes. These images pose significant challenges in various domains, including security, media, and personal privacy. The ability to accurately distinguish between real and AI-generated images is crucial to mitigating the potential misuse of deepfake technology. This study explores the application of deep learning-based Convolutional Neural Networks (CNNs) for the classification of real versus fake faces, leveraging a dataset of real and AI-generated images. This research utilizes a CNN model to classify images as either real or AI-generated. The dataset, sourced from Kaggle, includes a balanced collection of real and synthetic faces. Data augmentation techniques, such as rescaling, shearing, and zooming, are employed to enhance the diversity of the training set. The CNN architecture comprises multiple convolutional and pooling layers, followed by dense layers for classification. The model is trained using the Adam optimizer and binary cross-entropy loss function, with accuracy as the primary performance metric. The CNN model achieved a test accuracy of approximately 82demonstrating its effectiveness in distinguishing between real and AI-generated images. The confusion matrix and classification report indicate high precision and recall values, underscoring the model's reliability. The use of data augmentation contributed significantly to the model's ability to generalize new data, as evidenced by the high validation accuracy. The results highlight the potential of CNNs in addressing the challenges posed by deepfake technology. The model's high accuracy and robust evaluation metrics suggest that it can serve as a reliable tool for real-time image classification. However, the study also identifies areas for improvement, such as addressing occasional fluctuations in validation loss and exploring alternative architectures to further enhance performance. This research underscores the efficacy of CNNs in classifying real versus AI generated images, offering a promising approach to combat the growing threat of deepfakes. Future work will focus on refining the model and expanding its application to video data, thereby broadening its utility in various real-world scenarios.**

***Index Terms*—AI-generated faces, Classification, CNN, Deepfake, Neural Network**

I. Introduction

The advent of deep learning and AI technologies has revolutionized the field of image synthesis, enabling the creation of highly realistic synthetic images, often referred to as deepfakes. These images, generated using advanced algorithms such as Generative Adversarial Networks (GANs), have become increasingly prevalent across various platforms, raising significant concerns about their potential misuse. The ability to distinguish between real and AI-generated images is critical, particularly in contexts where authenticity is paramount, such as journalism, law enforcement, and social media. The need for effective deepfake detection mechanisms is underscored by the rapid growth of deepfake content. According to a report by Deeptrace, the number of deepfake videos online doubled from 7,964 in December 2018 to 14,678 in June 2019, with a significant portion of these videos being used for malicious purposes, including misinformation and identity theft. This alarming trend highlights the urgent need for robust detection systems capable of identifying synthetic media with high accuracy.[1]

Recent advancements in deep learning, particularly Convolutional Neural Networks (CNNs), have shown promise in addressing the challenges associated with deepfake detection. CNNs are well suited for image classification tasks due to their ability to automatically learn hierarchical feature representations from raw pixel data. Studies have demonstrated the efficacy of CNNs in various image recognition tasks, achieving state-of-the-art performance in domains such as facial recognition and object detection. This research aims to leverage the power of CNNs to develop a reliable and accurate method for classifying real versus AI-generated images. By employing a dataset comprising both real and synthetic faces, this study seeks to explore the potential of CNNs in discerning subtle differences between authentic and fabricated images. The methodology involves the use of data augmentation techniques to enhance the diversity of the training set, thereby improving the model's generalization capabilities. The significance of this research lies in its potential to provide a scalable solution to the deepfake detection problem, offering a tool that can be integrated into various applications to safeguard against the misuse of synthetic media. By advancing the understanding of how CNNs can be applied to this domain, this study contributes to the broader effort to develop effective countermeasures against the growing threat of deepfakes.

A. Research Questions

This research aims to explore the capabilities and limitations of deep neural networks in distinguishing between real and AI-generated faces. The study seeks to address the following key questions. The first research question seeks to identify the distinguishing features between real and AI-generated faces using deep neural network analysis. This involves uncovering subtle patterns and characteristics that differentiate genuine images from those created by AI, which may not be immediately apparent through traditional analysis methods. The second research question focuses on evaluating the effectiveness of the developed deep neural network model in accurately classifying real and AI-generated faces. It aims to assess the reliability and robustness of the model in distinguishing between these two types of images. The third research question explores the impact of various preprocessing techniques on the performance of the deep neural network model. It investigates how methods like data augmentation and normalization can enhance the model's ability to generalize across different datasets and improve classification accuracy. Finally, the fourth research question addresses how the performance of the deep neural network model in detecting AI-generated faces can be evaluated and further improved. This involves a comprehensive analysis of the model's strengths and weaknesses, with the goal of identifying areas for enhancement to increase accuracy and reliability in real-world applications.

B. Research Objectives

The objective of this research is to identify the unique features that distinguish real faces from AI-generated ones using deep neural network analysis. This involves leveraging the capabilities of neural networks to detect subtle patterns and characteristics that may be challenging to discern through conventional methods. Another objective is to assess the effectiveness of the developed deep neural network model in classifying real and AI-generated faces. This includes evaluating the model's reliability and robustness in making accurate distinctions between genuine and synthetic images. Additionally, this research aims to explore and implement preprocessing techniques that enhance the performance of the deep neural network model. By employing methods such as data augmentation and normalization, the objective is to improve the model's ability to generalize across diverse datasets and achieve higher accuracy in classification tasks. Finally, the research is committed to evaluating the overall performance of the deep neural network model in detecting AI-generated faces. This involves identifying potential areas for improvement and refining the model to ensure greater accuracy and effectiveness in practical applications.

II. Literature review

A. Advances in Deep Learning for Image Classification

The realm of deep learning has seen significant advancements, particularly in the domain of image classification. These advancements are crucial for the development of sophisticated models capable of distinguishing between real and AI-generated faces.[2] One notable development is the use of deep convolutional neural networks (DCNNs), which have shown remarkable efficacy in face recognition tasks. Deep learning-based networks excel in identifying and learning specific facial patterns by converting face-related data into mathematical representations, resulting in high-accuracy face recognition. [3] Deep learning methods, such as those involving generative adversarial networks (GANs), have revolutionized image synthesis.[4]. The use of advanced GANs like Progressive Growth GANs (PGGAN) and BigGAN, which produce highly photorealistic images indistinguishable from real ones in a limited time frame.[5], [6] These models have set new benchmarks in generating realistic facial images, complicating the task of distinguishing between genuine and synthetic faces.

However, these advancements come with their limitations. Boudníková and Kleisner noted that their study focused solely on male faces of European origin due to the consistent smiling feature of female faces generated by GANs.[7] This limitation underscores the need for more diverse datasets that include various ethnicities and genders to improve the generalizability of deep learning models.[8] Perception study also emphasized the importance of using diverse facial datasets to ensure ecological validity in artificial face perception research. [9] Moreover, the concept of shared representations between artificial and biological systems has been explored. Marr's (1982) three-level framework, suggests that despite differences in physical implementations, both DCNNs and the human brain may use similar representations to achieve face gender classification. This hypothesis was supported by findings that showed highly similar class-invariant representations (CIs) between VGG-Face models and human perception. [10], [11] The application of deep learning extends beyond face recognition to deepfake detection. The effectiveness of using deep learning methodologies, such as the XGBoost approach combined with CNN and Inception Res-Net techniques, achieving a 90% accuracy score in video deepfake detection using datasets like CelebDF and FaceForensics++. Similarly, MobileNet and Xception models have shown high accuracy in automatic deepfake video classification, with scores ranging from 91% to 98%. [3]

B. AI-Generated Content and Its Detection

The detection of AI-generated content, particularly synthetic faces, has become a significant focus within the field of deep learning. [7] AI-generated faces differ statistically from natural faces, exhibiting lower variability in facial shape, reduced morphological disparity, and lower levels of facial asymmetry. These differences, while quantifiable through geometric morphometrics, are not easily perceptible to the human eye.[12]

One interesting aspect is the lower shape variation in AI-generated faces, which suggests a higher level of averageness. This concept aligns with findings from AI-Synthe sized Faces Are Indistinguishable from Real Faces and More Trustworthy , indicating that objects closer to the average are perceived as more typical and natural.[13] Thus, AI-generated faces, being more average, are perceived as more trustworthy and natural compared to the more variable natural faces [7]. This phenomenon also ties into the ”uncanny valley” effect, where objects that are almost human-like but not quite can evoke discomfort. However, AI-generated faces have seemingly transcended this valley by appearing sufficiently human-like and, paradoxically, by being less distinctive[7], [13].

The potential for AI-generated faces to be used in various industries, such as marketing and entertainment, underscores the importance of understanding and improving the algorithms behind their creation. The aim is to generate content that is more indistinguishable from natural faces, thereby increasing their utility and acceptance.[7]

Further exploration into AI-generated content reveals ethical concerns, particularly regarding biases in deep learning models. Research by Tian demonstrated that biases, such as the Own-Race Effect (ORE), can emerge from unbalanced training datasets.[14] This bias occurs when models trained predominantly on one race's faces perform better on those faces compared to others. The study found that faces of the majority race were more sparsely distributed in the representational space, leading to better performance, a phenomenon observed in both artificial and human systems [15]

Addressing AI biases involves more than just balancing datasets; it requires innovative algorithmic approaches. For instance, multi-task adversarial learning methods have been proposed to manipulate biased representational subspaces and mitigate model biases in natural language processing and computer vision [16], [17]. These methods involve designing algorithms that can modulate internal representations within DCNNs, thus ensuring fairer and more accurate model performance. [18]

The proliferation of deepfake technology, which leverages GANs to create realistic yet fake multimedia content, presents further challenges.[19] Deepfakes are used in various malicious activities, including identity theft, cyber extortion, and misinformation campaigns [3]. Studies have shown that deepfake detection can be achieved through various deep learning techniques, such as examining human eye blinking patterns or using multiple CNN architectures for robust identification [3], [20]. These detection methods are crucial in mitigating the risks posed by synthetic media, emphasizing the need for continuous advancements in detection technologies to keep pace with the evolving capabilities of AI-generated content. [21]

C. Evaluation Metrics and Techniques for Face Detection Models

Evaluating the performance of face detection models is crucial for determining their effectiveness and reliability. Various metrics and techniques have been developed to assess these models, with particular focus on deep learning approaches such as Generative Adversarial Networks (GANs) and Convolutional Neural Networks (CNNs)[22].GANs have become a cornerstone for image processing, particularly for generating realistic images.[23] They operate through a dual network structure consisting of a generator and a discriminator, both typically implemented as deep convolutional networks (DCGANs).[24] This architecture allows for the creation of high-quality images by optimizing functions from f-divergences, such as the Kullback-Leibler and Jensen-Shannon divergences.[23] DCGANs have demonstrated their effectiveness across various applications, including pixel graphics generation, image resolution enhancement, and style transfer .[25]

Evaluating GAN-generated content requires robust metrics to quantify the realism and accuracy of the generated images. Commonly used metrics include the Inception Score (IS) and Fréchet Inception Distance (FID).[26] These metrics assess the quality of generated images by comparing their distribution to real images, providing a measure of how closely the synthetic images mimic the real ones. Additionally, metrics such as precision, recall, and Area Under the Receiver Operating Characteristic Curve (AUC) are employed to evaluate detection models' performance in distinguishing between real and fake images[3].

Deep learning-based face detection models also utilize sophisticated evaluation techniques. For instance, VGG-Face, a model pretrained on face datasets, has been used to study the Face Inversion Effect (FIE), which shows how upright and inverted faces are represented differently[14]. The performance of these models is often validated through confusion matrix analysis, which helps visualize the model's accuracy and error rates in classifying different face types. Models like VGG16 and MobileNet achieve high accuracy and low error rates, with VGG16 showing an accuracy of 90% and the proposed model reaching 94%. [3], [27]

In forensic image analysis, methods to detect manipulated images include the use of radon transform and deep learning techniques. CNN-based approaches, such as those proposes have shown high accuracy in identifying tampered regions by learning important features that distinguish real images from manipulated ones.[28] Techniques like multi-task learning with Fully Convolutional Networks (FCNs) further enhance the detection capabilities by simultaneously learning surface labels and boundaries of spliced regions [29]

Another advanced technique involves using a combination of deep learning models for improved performance. For example, combining CNNs with ResNet50 and VGG16 through ensemble learning can significantly enhance detection accuracy[20]. This approach leverages the strengths of different models, resulting in higher overall performance metrics. [30] Despite significant advancements in deep learning for image classification and deepfake detection, several research gaps remain. Current models often lack generalizability due to the limited diversity in training datasets, leading to potential biases like the Own-Race Effect (ORE). Additionally, while deepfake detection has seen progress, there is a need for more robust methods that perform reliably across diverse, real-world conditions, particularly in video formats. Finally, the ethical implications and biases in AI-generated content detection require further exploration to ensure fair and equitable model performance across all demographic groups. Addressing these gaps is crucial for the future development of more inclusive and reliable deep learning applications.

III. Methodology

The methodology for classifying real vs AI-generated images using the provided Google Colab notebook involves several key steps, each employing specific techniques to ensure effective model training and evaluation. Below is a detailed explanation of the techniques and flow used in each step, along with the rationale behind using them

A. Data Loading and Preparation

The initial step involves loading and preparing the dataset. The dataset comprises images of real and AI-generated faces sourced from Kaggle. It consists of a total of 975 images where the sample size of the AI-generated images is 700 and real images is 589. The ImageDataGenerator utility from Keras is used to handle this task. This utility is essential because it allows for real-time data augmentation and efficient loading of images in batches, which is crucial for handling large datasets.

The dataset is divided into training and testing directories. The training data is augmented using transformations such as rescaling, shearing, zooming, and horizontal flipping. These transformations help in increasing the diversity of the training set without actually increasing the number of training images, which is vital for preventing overfitting and improving the model's generalization ability. The pixel values of images are rescaled to the range [0, 1] for both training and testing datasets to standardize the input data, which helps in faster convergence during model training.

B. Data Augmentation

Data augmentation is a critical technique used to artificially expand the size of the training dataset by creating modified versions of images in the dataset. In this methodology, transformations such as shear, zoom, and horizontal flip are applied to the training data. These augmentations introduce variability in the training images, which helps the model to generalize better to new, unseen data. By rescaling the pixel values, the data is normalized, which is a common practice in deep learning to ensure that the model trains efficiently and effectively.

C. Model Building

The core of this methodology is the construction of a Convolutional Neural Network (CNN). CNNs are particularly well-suited for image classification tasks due to their ability to automatically and adaptively learn spatial hierarchies of features from input images. The model is built using the Sequential API from Keras, which allows for the easy stacking of layers. The CNN model consists of 4 Convolutional layers, 4 Max Pooling layers, 1 Batch Normalisation layer, 1 Flatten layer, 2 Dense layers.

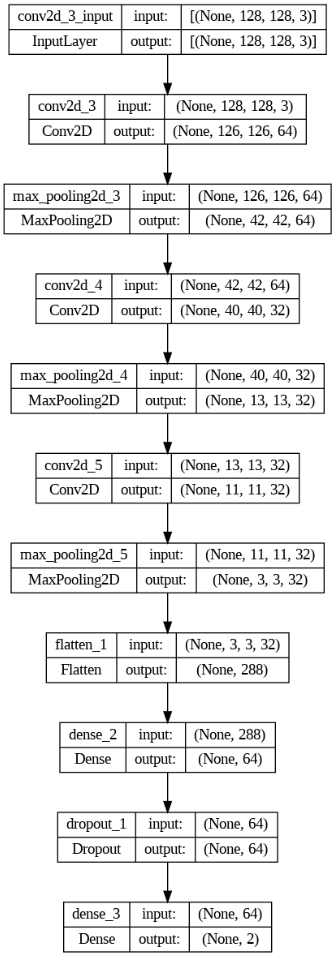
conv2d_3_input | input: | [(None, 128, 128, 3)]
InputLayer | output: | [(None, 128, 128, 3)]
conv2d_3 | input: | (None, 128, 128, 3)
Conv2D | output: | (None, 126, 126, 64)
max_pooling2d_3 | input: | (None, 126, 126, 64)
MaxPooling2D | output: | (None, 42, 42, 64)
conv2d_4 | input: | (None, 42, 42, 64)
Conv2D | output: | (None, 40, 40, 32)
max_pooling2d_4 | input: | (None, 40, 40, 32)
MaxPooling2D | output: | (None, 13, 13, 32)
conv2d_5 | input: | (None, 13, 13, 32)
Conv2D | output: | (None, 11, 11, 32)
max_pooling2d_5 | input: | (None, 11, 11, 32)
MaxPooling2D | output: | (None, 3, 3, 32)
flatten_1 | input: | (None, 3, 3, 32)
Flatten | output: | (None, 288)
dense_2 | input: | (None, 288)
Dense | output: | (None, 64)
dropout_1 | input: | (None, 64)
Dropout | output: | (None, 64)
dense_3 | input: | (None, 64)
Dense | output: | (None, 2)

Fig. 1. Architecture of the Neural Network layers

Convolutional Layers: These layers apply convolution operations to the input images, extracting features such as edges, textures, and patterns. Multiple convolutional layers are stacked to capture complex features at different levels of abstraction. The activation function used for the convolutional layers is Relu.

Max Pooling Layers: These layers reduce the spatial dimensions of the feature maps, which helps in reducing the computational cost and controlling overfitting by retaining only the most important features.

Flatten Layer: This layer converts the 2D feature maps into a 1D feature vector, which can then be fed into fully connected (dense) layers

Dense Layers: These layers perform classification based on the features extracted by the convolutional layers. The final dense layer uses a sigmoid activation function to output a probability score for binary classification (real vs AI-generated). Dropout Layer: This layer helps in preventing overfitting by randomly setting a fraction of input units to 0 during training, which forces the model to learn more robust features.

The feature map visualizations depict the output of convolutional and max-pooling layers from a CNN model. Each row represents a different layer, showing how the network processes and transforms the input image. Convolutional layers (conv2d) extract features like edges and textures, while max-pooling layers (max-pooling2d) reduce dimensionality, highlighting the most significant features. As layers progress, the feature maps become more abstract, focusing on complex patterns. This visualization helps understand how the model learns and identifies distinguishing features in images.

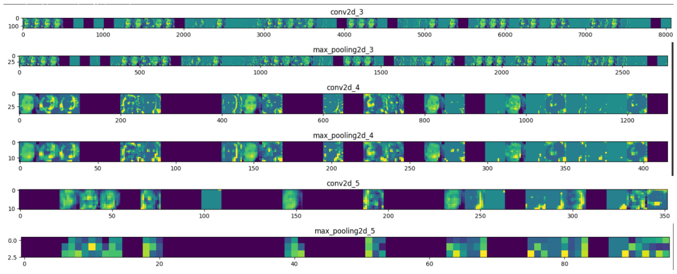

Fig. 2. Feature Map Visualizations of Convolutional and Max-Pooling Layers

D. Model Compilation

Once the model architecture is defined, it is compiled using the Adam optimizer and binary cross-entropy loss function. The Adam optimizer is chosen for its adaptive learning rate capabilities, which help in efficiently handling sparse gradients and noisy data. The binary cross-entropy loss function is appropriate for binary classification tasks, as it measures the performance of the model by comparing the predicted probabilities to the actual class labels. Accuracy is set as the metric to evaluate the model's performance during training, providing a straightforward measure of how well the model is performing.

E. Model Training

The model is trained using the training data generator, which feeds batches of augmented images to the model. This batch training approach improves computational efficiency and helps in faster convergence. The training process involves multiple epochs, where the model iteratively learns from the training data. Validation data is used to monitor the model's performance and prevent overfitting. Accuracy is used as the monitoring metric. By evaluating the model on validation data after each epoch, we can ensure that the model is not just memorizing the training data but is also generalizing well to new data.

F. Model Evaluation

After training, the model's performance is evaluated on the test data. This involves calculating the test accuracy and generating a confusion matrix and classification report. The confusion matrix provides a detailed breakdown of the model's performance by showing the number of true positives, true negatives, false positives, and false negatives. The classification report summarizes the precision, recall, F1-score, and support for each class, offering a comprehensive view of the model's strengths and weaknesses in distinguishing between real and AI-generated images.

G. Visualization

Finally, the training history is visualized to assess the model's performance over time. Plotting the training and validation accuracy and loss over epochs helps in understanding how well the model is learning and whether it is overfitting or underfitting. These visualizations provide valuable insights into the model's training dynamics and can guide further tuning and optimization of the model. In summary, this methodology leverages data augmentation, CNN architecture, and comprehensive evaluation techniques to build a robust model for classifying real vs AI-generated images. Each step is carefully designed to ensure efficient data handling, effective feature extraction, and thorough performance assessment, resulting in a model that can accurately distinguish between real and AI-generated faces.

IV. Results and Discussion

The model was trained using the augmented training dataset and validated on a separate validation set. The training process involved 25 epochs, and the model's performance was monitored using accuracy and loss metrics for both training and validation sets.

A. Training and Validation Accuracy

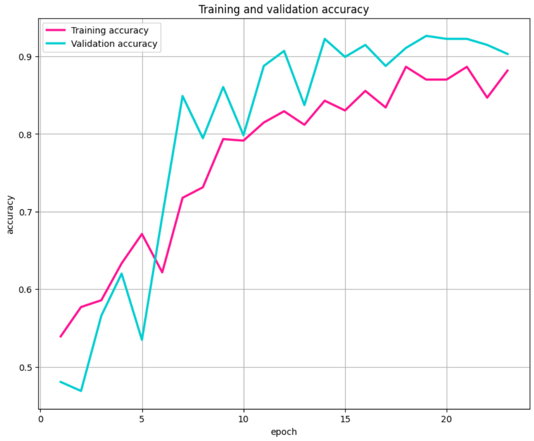

Fig. 3. Training and Validation Accuracy Over Epochs

The accuracy plot shows that the training accuracy steadily increased over the epochs, reaching a high value by the end of the training period. The validation accuracy also showed an upward trend, indicating that the model was learning effectively and generalizing well to the validation data.

B. Training and Validation Loss

The loss plot indicates that the training loss decreased consistently over the epochs, which is a sign of the model learning and fitting the training data. The validation loss also decreased, though not as smoothly as the training loss, suggesting some fluctuations but overall improvement.

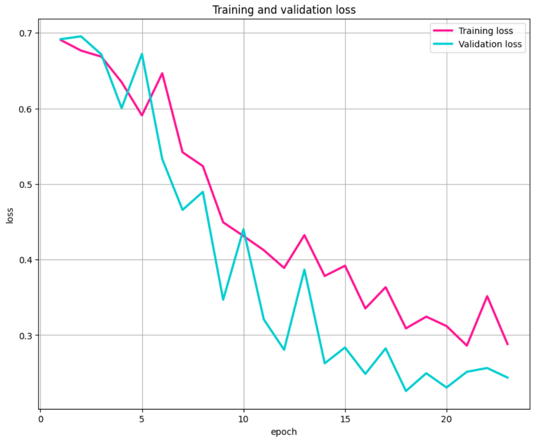

Fig. 4. Training and Validation Loss Over Epochs

C. Test Accuracy

The model achieved a test accuracy of approximately 82%, which demonstrates its ability to distinguish between real and AI-generated images with high accuracy.

D. Confusion Matrix

The confusion matrix provided a detailed breakdown of the model's performance

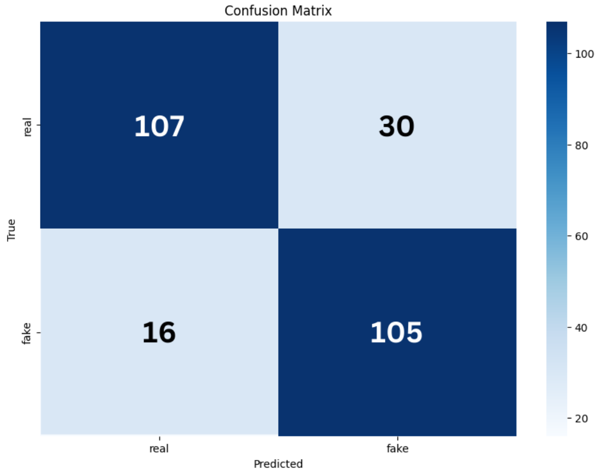

Fig. 5. Confusion Matrix

- True Positives (Real images correctly classified as real)
- True Negatives (AI-generated images correctly classified as AI-generated)
- False Positives (Real images incorrectly classified as AI-generated)
- False Negatives (AI-generated images incorrectly classified as real)

The high number of true positives and true negatives, coupled with a low number of false positives and false negatives, indicates that the model is highly effective in classifying the images correctly.

E. Classification Report

The classification report summarized the precision, recall, F1-score, and support for each class (Real and AI-Generated). The precision and recall values were high for both classes, resulting in high F1-scores. This indicates that the model performs well in both detecting real images and identifying AI-generated images.

Accuracy	0.8218
Recall	0.8217
Precision	0.8267
F1 Score	0.8217

V. Conclusion

The results indicate that the convolutional neural network (CNN) model is highly effective in distinguishing between real and AI-generated images. The high training and validation accuracy, along with the low training and validation loss, suggest that the model has learned to extract relevant features from the images that are indicative of their class (real or AI-generated). One of the key factors contributing to the model's success is the use of data augmentation. By applying transformations such as shear, zoom, and horizontal flip, the training set was diversified, which helped the model generalize better to new, unseen data. This is evident from the high validation accuracy and the model's performance on the test set. The confusion matrix and classification report further validate the model's robustness. The high precision and recall values for both classes indicate that the model is not only accurate but also reliable in its predictions. The low number of false positives and false negatives suggests that the model has a strong ability to correctly classify images, minimizing misclassification. However, there are some fluctuations observed in the validation loss, which could be indicative of occasional overfitting or the inherent variability in the validation set. This could be addressed by further fine-tuning the model, such as adjusting the learning rate, adding more regularization, or using a larger and more diverse dataset.

A. Future Work

Future research will focus on refining the current CNN model to enhance its generalization capabilities and accuracy in distinguishing between real and AI-generated images. This includes exploring alternative architectures, such as deeper networks or hybrid models, to capture more complex patterns. Additionally, expanding the dataset to include a wider variety of faces across different ethnicities and conditions will improve model robustness. Incorporating temporal data analysis could extend the model's application to video deepfake detection, addressing the dynamic nature of video content. Finally, addressing ethical considerations and biases in AI-generated content detection will be crucial to ensure fair and equitable performance across diverse demographic groups.

References

[1] E. Altuncu, V. N. L. Franqueira, and S. Li, 'Deepfake: Definitions, Performance Metrics and Standards, Datasets and Benchmarks, and a Meta-Review', Aug. 21, 2022, arXiv: arXiv:2208.10913. Accessed: Aug. 11, 2024. [Online]. Available: http://arxiv.org/abs/2208.10913

[2] J. Ricker, D. Assenmacher, T. Holz, A. Fischer, and E. Quiring, 'AI-Generated Faces in the Real World: A Large-Scale Case Study of Twitter Profile Images', Aug. 06, 2024, arXiv: arXiv:2404.14244. Accessed: Aug. 08, 2024. [Online]. Available: http://arxiv.org/abs/2404.14244

[3] A. Raza, K. Munir, and M. Almutairi, 'A Novel Deep Learning Approach for Deepfake Image Detection', Applied Sciences, vol. 12, no. 19, p. 9820, Sep. 2022, doi: 10.3390/app12199820

[4] D. M. Montserrat et al., 'Deepfakes Detection with Automatic Face Weighting', in 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA: IEEE, Jun. 2020, pp. 2851–2859. doi: 10.1109/CVPRW50498.2020.00342

[5] C.-C. Hsu, Y.-X. Zhuang, and C.-Y. Lee, 'Deep Fake Image Detection Based on Pairwise Learning', Applied Sciences, vol. 10, no. 1, p. 370, Jan. 2020, doi: 10.3390/app10010370

[6] Y. Said, M. Barr, and H. E. Ahmed, 'Design of a Face Recognition System based on Convolutional Neural Network (CNN)', Eng. Technol. Appl. Sci. Res., vol. 10, no. 3, pp. 5608–5612, Jun. 2020, doi: 10.48084/etasr.3490

[7] O. Boudníková and K. Kleisner, 'AI-generated faces show lower morphological diversity than real faces do', AR, vol. 87, no. 1, pp. 81–91, Jun. 2024, doi: 10.18778/1898-6773.87.1.06

[8] J. J. Bird and A. Lotfi, 'CIFAKE: Image Classification and Explainable Identification of AI-Generated Synthetic Images', IEEE Access, vol. 12, pp. 15642–15650, 2024, doi: 10.1109/ACCESS.2024.3356122

[9] S. St et al., 'Deep learning model for deep fake face recognition and detection', PeerJ Computer Science, vol. 8, p. e881, Feb. 2022, doi: 10.7717/peerj-cs.881

[10] Y. Song, Y. Qu, S. Xu, and J. Liu, 'Implementation-Independent Representation for Deep Convolutional Neural Networks and Humans in Processing Faces', Front. Comput. Neurosci., vol. 14, p. 601314, Jan. 2021, doi: 10.3389/fncom.2020.601314

[11] T. Kanade, 'Neural Network-Based Face Detection.pdf'

[12] J. M. Alrikabi and K. H. Alibraheemi, 'Deep Learning-Based Face Detection and Recognition System', 2021

[13] A. Kumar, A. Kaur, and M. Kumar, 'Face detection techniques: a review', Artif Intell Rev, vol. 52, no. 2, pp.

[14] J. Tian, H. Xie, S. Hu, and J. Liu, 'Multidimensional Face Representation in a Deep Convolutional Neural Network Reveals the Mechanism Underlying AI Racism', Front. Comput. Neurosci., vol. 15, p. 620281, Mar. 2021, doi: 10.3389/fncom.2021.620281.

[15] F. Marra, D. Gragnaniello, D. Cozzolino, and L. Verdoliva, 'Detection of GAN-Generated Fake Images over Social Networks', in 2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), Miami, FL: IEEE, Apr. 2018, pp. 384–389. doi: 10.1109/MIPR.2018.00084.

[16] S. Zheng, Z. Zeng, Z. Li, W. Kong, and W. Zhong, 'A photovoltaic resource evaluation method based on meteorological and geospatial data', E3S Web of Conf., vol. 536, p. 02001, 2024, doi: 10.1051/e3sconf/202453602001.

[17] P. Patil, V. Deshpande, V. Malge, and A. Bevinmanchi, 'Fake Face Detection Using CNN', IJRASET, vol. 10, no. 9, pp. 519–522, Sep. 2022, doi: 10.22214/ijraset.2022.45829.

[18] R. Chauhan, D. M. Sethi, and D. S. Ahuja, 'Deep Learning-Based Methods for Detecting Generated Fake Faces'.

[19] E. Şafak and N. Barışçı, 'Detection of fake face images using lightweight convolutional neural networks with stacking ensemble learning method', PeerJ Computer Science, vol. 10, p. e2103, Jun. 2024, doi: 10.7717/peerj-cs.2103.

[20] F. M. Salman and S. S. Abu-Naser, 'Classification of Real and Fake Human Faces Using Deep Learning', vol. 6, no. 3, 2022.

[21] D. U. Eswar, 'FAKE FACE DETECTION USING ARTIFICIAL INTELLIGENCE', IRJMETS.

[22] C. Rathgeb, R. Tolosana, R. Vera-Rodriguez, and C. Busch, Eds., Handbook of Digital Face Manipulation and Detection: From DeepFakes to Morphing Attacks. in Advances in Computer Vision and Pattern Recognition. Cham: Springer International Publishing, 2022. doi: 10.1007/978-3-030-87664-7.

[23] S. C. Leonov, A. Vasilyev, A. Makovetskii, and J. Diaz-Escobar, 'An algorithm of face recognition based on generative adversarial networks', in Applications of Digital Image Processing XLI, A. G. Tescher, Ed., San Diego, United States: SPIE, Sep. 2018, p. 94. doi: 10.1117/12.2321039.

[24] M. Favorskaya, 'Fake Face Image Detection Using Deep Learning-Based Local and Global Matching', in Short Paper Proceedings of the 2nd Siberian Scientific Workshop on Data Analysis Technologies with Applications 2021, CEUR-WS.org, 2021. doi: 10.47813/sibdata-2-2021-20.

[25] D. Park, H. Na, and D. Choi, 'Performance Comparison and Visualization of AI-Generated-Image Detection Methods', IEEE Access, vol. 12, pp. 62609–62627, 2024, doi: 10.1109/ACCESS.2024.3394250.

[26] D. Cozzolino, G. Poggi, R. Corvi, M. Niessner, and L. Verdoliva, 'Raising the Bar of AI-generated Image Detection with CLIP'.

[27] Kavi B. Obaid, S. R. M. Zeebaree, and O. M. Ahmed, 'Deep Learning Models Based on Image Classification: A Review', Oct. 2020, doi: 10.5281/ZENODO.4108433.

[28] J. Sharma, S. Sharma, V. Kumar, H. S. Hussein, and H. Alshazly, 'Deepfakes Classification of Faces Using Convolutional Neural Networks', TS, vol. 39, no. 3, pp. 1027–1037, Jun. 2022, doi: 10.18280/ts.390330.

[29] L. M. Dang, S. I. Hassan, S. Im, J. Lee, S. Lee, and H. Moon, 'Deep Learning Based Computer Generated Face Identification Using Convolutional Neural Network', Applied Sciences, vol. 8, no. 12, p. 2610, Dec. 2018, doi: 10.3390/app8122610.

[30] J. Atwan, M. Wedyan, D. Albashish, E. Aljaafrah, R. Alturki, and B. Alshawi, 'Using Deep Learning to Recognize Fake Faces', IJACSA, vol. 15, no. 1, 2024, doi: 10.14569/IJACSA.2024.01501113.