e-ISSN:2582-7219

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

Volume 7, Issue 7, July 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.521

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 7, July 2024 |

| DOI:10.15680/IJMRSET.2024.0707266 |

AI Art Detection: Applying Deep Learning to Authenticate Images

Dr. R. Jayanthi, C.K. Mohammed Anshad

Associate Professor, Department of Master of Computer Application, Dayananda Sagar College of Engineering, Bangalore, India

P.G. Student, Department of Master of Computer Application, Dayananda Sagar College of Engineering, Bangalore, India

ABSTRACT: With the progress of AI technology, the borders between figurative art made by humans and AI-generated illustrations are becoming blurred. In this research, we present a deep learning-based approach to detect which category of images it is amongst two forms. The training and testing data include artificial images with 256x256 pixel resolution produced by the AI system and human-created art images. We created a Convolutional Neural Network (CNN) using TensorFlow and Keras with convolution layers, pooling, and fully connected operations. A sigmoid activation function for binary classification was included to complete the model.

The model successfully distinguishes between artificial intelligence (AI) and human-generated images with an accuracy of 95.31%. These findings suggest that deep learning can be used in systems that prevent copyright infringement, verify content, or authenticate artwork.

KEYWORDS: Deep learning, Convolutional Neural Network, AI art detection, Image classification, Image authentication, Digital media ethics

I. INTRODUCTION

Recently, the art industry has seen a transformation with huge advancements in the Artificial Intelligence (AI) landscape. Artificial intelligence, specifically deep learning models, has helped to create synthetic artworks that look similar to those made by the human hand. With the line between real art and AI-created art starting to blend, this situation challenges the evolution of adopting the right methods for authenticating artworks [1][2].

Being a traditional means to authenticate artworks, expert judgment plays an indispensable role in it. Professional opinions may be affected by subjective irregularities, hence reducing their trustworthiness [3]. The availability of AI-generated art warrants such tools to become even more objective and automated. These issues can be solved using deep learning, such as Convolutional Neural Networks (CNNs) which was done for image recognition to discern authentic art from AI-generated art [4].

As AI-generated art becomes more popular, there needs to be a way to verify an artwork as being authentic. As digital content has become pervasive and generating high-quality synthetic imagery is easier than ever, the art community faces growing issues when attempting to identify an original piece of artwork. This impacts artists and collectors but also galleries, museums, and auction houses that have to prove the authenticity of pieces are dependent on it for preservation and value.

For this investigation, we develop a deep learning mechanism that spots digital image fakes among the more general task of identifying human vs. AI-made artworks In our experiment, we use a publicly available dataset of high-resolution images across the two categories and train a CNN model using TensorFlow and Keras. Here we have some convolutional layers, pooling, and a fully connected layer which leads to a sigmoid activation function for a binary use case. Lastly, We employ data augmentation approaches to increase our model's robustness and generalization.

In this research, three main contributions are:

- An extensive dataset comprising human-made and AI-generated artwork.
- The CNN-based model was developed and evaluated to achieve high accuracy in classifying these images.
- We apply the model to real-world datasets (content verification, copyright protection, and art).

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 7, July 2024 |

| DOI:10.15680/IJMRSET.2024.0707266 |

This work adds to previous efforts for the protection of art in general against AI-generated imitations. The results of our research can lead or contribute to systems that avoid copyright, detect authenticity, and protect artists' IP due rights.
The remaining part of this paper is organized as follows: Section 2 discusses similar work in AI art detection. The dataset engineering process and preprocessing steps and the CNN model architecture and training are described in Section 3. In Section 4, Experimental results as well as their implications are presented, which is followed by conclusions. Section 5 is the Conclusion where the paper has been concluded by citing future research directions.

II. RELATED WORK

A. Overview of Art Authentication Techniques

Art authentication dates back to the opinions of art historians and connoisseurs. They rely on long-established techniques of studying an artwork's history, materials, and stylistic features. However, these approaches are often subjective and not reproducible, which frequently leads to expert disagreements. Such developments have led to a growing interest in developing more objective and repeatable scientific-based methods for authenticating art objects. Over recent years, technology advancements encompassing material science and implantable digital systems have made attractive improvements to facilitate reliable authentication methods [1][2].

B. AI and Machine Learning in Art Authentication

Using AI and machine learning to authenticate art is an application of this technology that has never previously been possible. In this domain, convolutional neural networks (CNNs) equipped with impressive image recognition have aided the most. For instance, they are capable of noticing subtleties and designs in an artwork that a human specialist might never even think to look for. For example, recent studies have shown that deep learning-based models could offer human-like capability to recognize strokes and brush textures among other fine-grained cues improving the reliability of art authentication by a large margin [3][4].

C. Generative Adversarial Networks (GANs)

Generative Adversarial Networks [2], proposed by Goodfellow et al., are one of the most significant inventions in the AI community. This is done by using an image generator and discriminator simultaneously. Thus, we can obtain realistic fake images through this adversarial process. The GAN evolution has changed the paradigm of fake art and yet presents itself with new challenges in authenticity, for distinguishing real from GAN-generated artwork [5].

D. Diffusion Models in Vision

As in Croitoru et al., diffusion models present another inspirational choice for generating and verifying art. These models iteratively refine images and learn the underlying data distribution. Diffusion models can be used to detect inconsistencies in artworks within the overdetermined regime relevant for art authentication, adding more interpretability and detection accuracy given certain assumptions [6].

E. Explainable AI and Transparency

Another significant issue of deploying deep learning models in art authentication is their black-box nature. The steps leading to the ultimate conclusion may often not be understandable, and it is impossible to define how AI-based such a conclusion is. However, explainable AI approaches have been invented as an answer to that challenge. Methods such as Grad-CAM or LIME can draw a visual explanation of the image and show which parts of the painting affected the decision. With their help, people are more likely to accept more visually transparent and understandable AI as a tool that has the authority to judge art in the art community [7][8].

F. Datasets and Generalizability

Art authentication relies heavily on meticulously curated data for the training and validation of AI models. In this vein, the ArtiFact dataset introduced by Rahman et al. is a very important contribution to our field and I would have liked it if my paper was compared against theirs, which roughly speaking did not happen. It contains a mixture of fictional and non-fictional imagery to create an extensive dataset for creating, and testing art authentication models. Generalization of these models is extremely important in turn to ensure that their use also extends over a variety of artworks and styles, leading to the need for diverse datasets even further [9].

G. Multimodal Approaches and Emotion Recognition

Using all kinds of data sources can also contribute to the human-level accuracy of art forgery detection models. Aslan et al. facilitated the blending of visual features with knowledge graph embeddings for detecting emotions raised by paintings Fusing visual and contextual information, this multimodal approach allows us a richer understanding of how

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 7, July 2024 |

| DOI:10.15680/IJMRSET.2024.0707266 |

an artwork works on the world. Beyond providing frictionless authentication, such end-to-end analysis could potentially serve to strengthen the security of these systems by understanding art not just on a physical level but holistically - including how an artwork exists in emotional and other contextual planes [10].

H. Related Work

Several researchers proposed novel models and methods that are very innovative in the field of art authentication. Hamid et al. introduced an alternative CNN model for detecting image forgery, the detection achieves high accuracy on benchmarking datasets. Elgammal et al. developed an AI system that is capable of identifying a painting as fake and specifying its author by the presence of his style. These advances highlight the promise that AI holds for shifting art authentication time tools which are both high-performance and also repeatable [11][12].

III. METHODOLOGY

A. Dataset

The dataset used in this study contains 120000 images, that are equally split between AI-generated and human-made works of art. The images are categorized and subdivided into two main categories:

AI-generated images: 60000 images generated by various AI-powered software that is based on Generative Adversarial Networks (GANs) and other AI-based algorithmic models. These tools have been chosen to represent a wide range of styles and complexities to provide diversity in the dataset.

Human-created images: 60000 images sourced from publicly available art collections, including classical paintings and contemporary fine arts by human artists. The choice was made to show a representative variety of types of artworks and periods of history.

The dataset was divided into training, validation, and test sets with the following ratios:

Training set: 70% (42000 AI-generated and 42000 human-created images). This set was used to train the model.
Validation set: 15% (9000 AI-generated and 9000 human-created images). The hyperparameters were tuned and overfit prevention was also achieved with this set.
Test set: 15% (9000 AI-generated and 9000 human-created images). This set was primarily used to test the performance of final models.

B. Data Preprocessing

Preprocessing images for training:

Rescaling: All images were rescaled to have pixel values in the range [0, 1] by dividing pixels by 255. This process is done to normalize the image for further effective and efficient processing by our neural network.

Data Augmentation: The training images are performed transformations like rotating up to 40 degrees, shifting width and height by up to 20% of total width/height, shearing up to 20%, zooming up, and horizontal flipping for the training images. This augmentation is there to add some diversity to the training data and also help our model not overfit by making it learn features that are more invariant against these transformations.

To address the training of our deep learning models, we utilized the Keras ImageDataGenerator class which can perform on-the-fly manipulation and pre-processing for a batch of image data.

C. Model Architecture

In this study, the CNN (Convolutional Neural Network) architecture is intended to classify images as AI-generated or human-created.The layers of the model architecture are as follows:

Input Layer: The input layer accepts images with three color channels—Red, Green, and Blue (RGB)—that are 256 by 256 pixels in size.

Convolutional Layers: Three layers of convolution, with 32, 64, and 128 filters in each, are included. Every layer employs the ReLU activation function and has a kernel size of (3, 3). At the local level, these layers are utilized to extract characteristics from the input images.

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 7, July 2024 |

| DOI:10.15680/IJMRSET.2024.0707266 |

The first layer convolutional layers are trained to recognize basic features such as edges and textures.

The second and third convolutional layers detect more complex features and patterns.

Max-Pooling Layers: Every convolutional layer is followed by a max-pooling layer with a pool size of (2, 2). These layers help to reduce computational complexity and overfitting by reducing the spatial dimensions of the feature maps.

Batch Normalization: Following the every convolutional layer, a batch normalization layer is added to normalize the activations and enhance training stability by minimizing internal covariate shifts.

Flatten Layer: By transforming the 2D feature maps into a vector of 1D features, this layer gets the data ready for the fully connected layers.

Fully Connected Layers: To lessen overfitting, a dropout layer with a dropout rate of 0.5 is placed after a dense layer with 512 units and ReLU activation.

Output Layer: A dense layer for binary classification that has a sigmoid activation function and a single unit. If the image is human or artificial intelligence (AI) generated, the sigmoid function would predict a likelihood value between 0 and 1.

The binary cross-entropy loss function, appropriate for binary classification tasks, and the Adam optimizer, an effective gradient-based optimization tool, were used to compile the model.

D. Training

To avoid overfitting and guarantee effective training, the model was trained using the training dataset, with early stopping and learning rate reduction on the plateau. Throughout training, we kept an eye on our model's performance using the validation dataset. The training procedure was set up with a batch size of 64 and a maximum duration of 50 epochs.

Three callbacks were used during training:

Early Stopping: If training does not improve after five consecutive epochs, the validation loss is monitored and training is stopped. By stopping training as soon as the model's performance on the validation set ceases improving, this helps to prevent overfitting.

Model Checkpoint: Based on the validation loss, the optimal model is saved. This guarantees that the model with the highest training performance is retained for evaluation.

Reduce Learning Rate on Plateau: If the validation loss is not improved after three consecutive epochs, the learning rate is decreased. As training goes on, this refines the model's weights more precisely, assisting in a more effective convergence.

E. Evaluation Metrics

The test dataset was used to assess the model's performance. The efficacy of the model was assessed using the following metrics:

Accuracy: The proportion of the test set's total number of images that were accurately classified. This metric provides a summary of the model's overall efficacy[15].

Confusion Matrix: A tabular representation that breaks down model's performance by displaying the counts of the true positive (AI-generated images correctly classified as AI-generated), true negative (human-created images correctly classified as human-created), false positive (human-created images incorrectly classified as AI-generated), and false negative (AI-generated images incorrectly classified as human-created) counts. This gives a detailed perspective of the model's categorization performance [13].

ROC AUC Score: The area under the receiver operating characteristic curve, or ROC AUC Score, quantifies how well the model distinguishes between the two classes. The AUC score offers a cumulative assessment of performance across

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 7, July 2024 |

| DOI:10.15680/IJMRSET.2024.0707266 |

all categorization thresholds, and the ROC curve compares the true positive rate versus the false positive rate at different threshold values.

F. Model Saving

Following training, the final model was saved for future use and deployment. This step ensures that the trained model can be loaded and used for classification tasks without the need for retraining.

IV. EXPERIMENTAL RESULTS

In this section, a detailed account of the findings from the experiments, including the accuracy, confusion matrix, and ROC AUC scores and visualizations of the training and validation metrics are provided.

A. Training Process

The model was trained for a maximum of 50 epochs using the training and validation datasets. The validation loss was used to monitor the model's performance and prevent overfitting by implementing early stopping.

1) Training and Validation Accuracy and Loss:

The training accuracy was steadily increasing over the epochs and the validation accuracy showed a similar upward trend, indicating good generalization.

The training loss decreased monotonously, and the validation loss showed a downward trend, suggesting effective learning and reduced overfitting due to early stopping.

The following are plots of the training accuracy and the validation accuracy and loss across epochs:

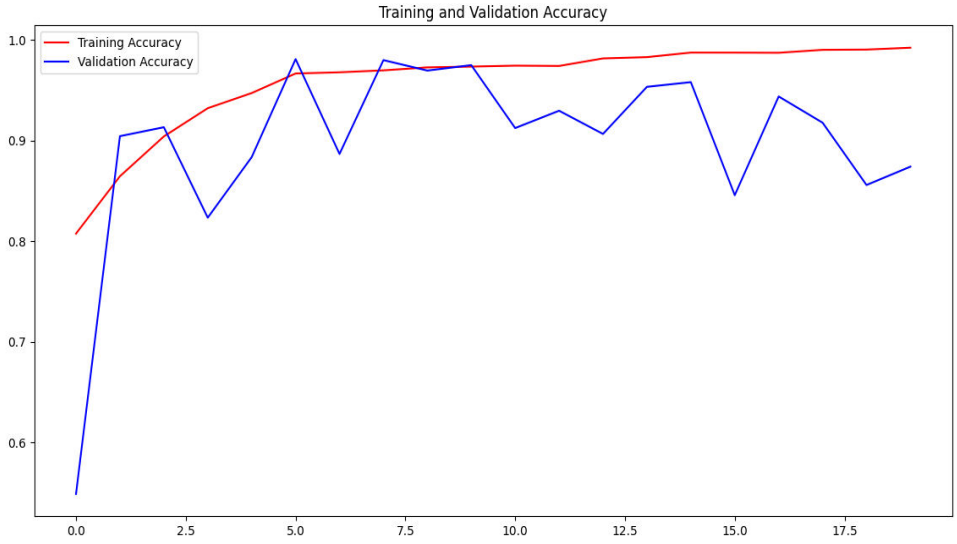

FIG 1. TRAINING AND VALIDATION ACCURACY PLOT

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 7, July 2024 |

| DOI:10.15680/IJMRSET.2024.0707266 |

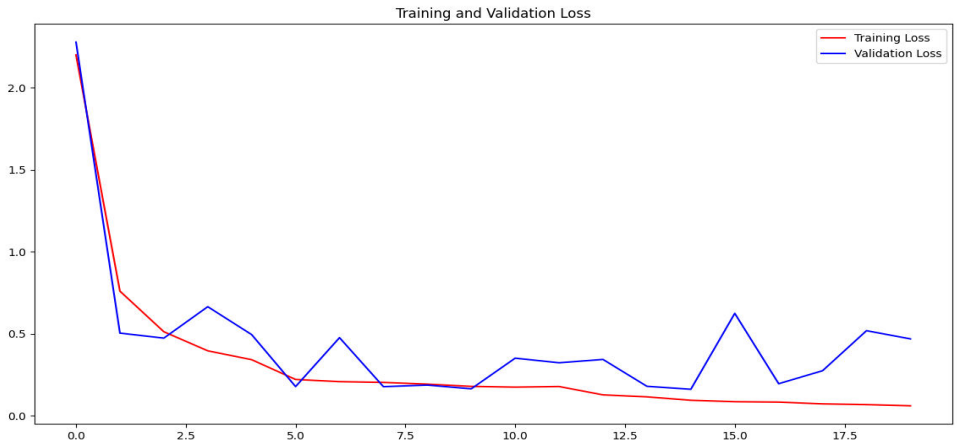

FIG 2. TRAINING AND VALIDATION LOSS PLOT

B. Evaluation Metrics

The test dataset was used to evaluate the model's performance. The evaluation metrics included accuracy, confusion matrix, and ROC AUC score.

Accuracy: The model was able to differentiate between an AI-generated image and human-created images, achieving 95.31%. accuracy on the test set

Confusion Matrix: A table with categories for true positive, true negative, false positive, and false negative numbers, the confusion matrix illustrates how well our model performs. The matrix demonstrated how well the model worked in the two image categories.

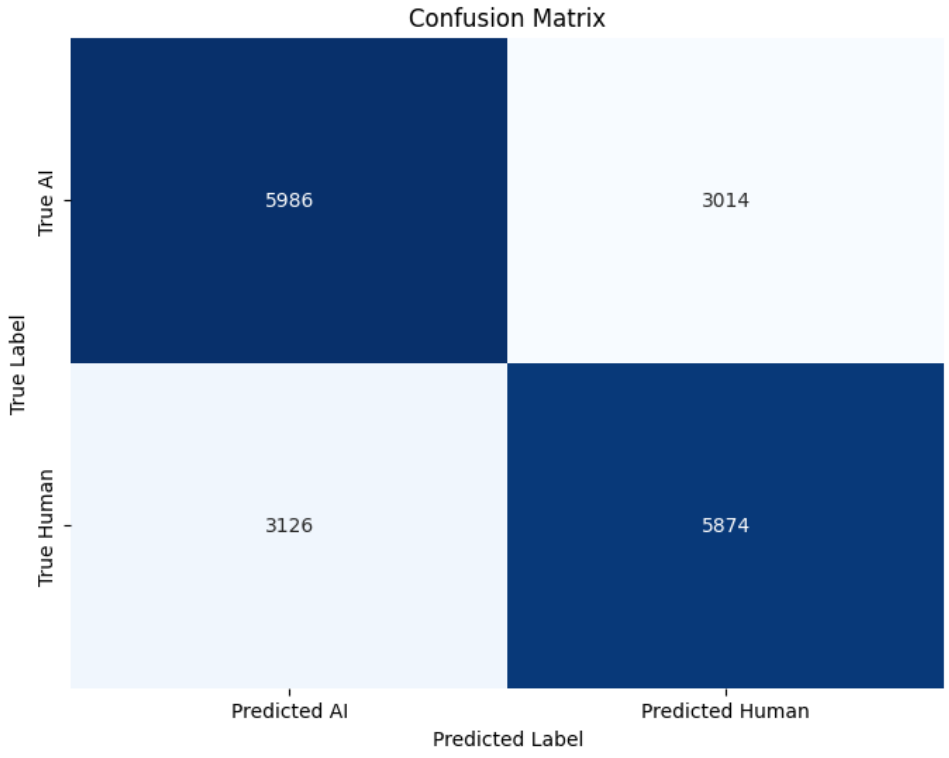

FIG 3. CONFUSION MATRIX HEATMAP

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 7, July 2024 |

| DOI:10.15680/IJMRSET.2024.0707266 |

ROC AUC Score: This is a measure of how well the model distinguishes between the two classes. The AUC value was high, indicating that the classification performances were truly strong. The ROC curve is a visualization technique that is used to assess the false positive rate versus the true positive rate at different threshold settings.

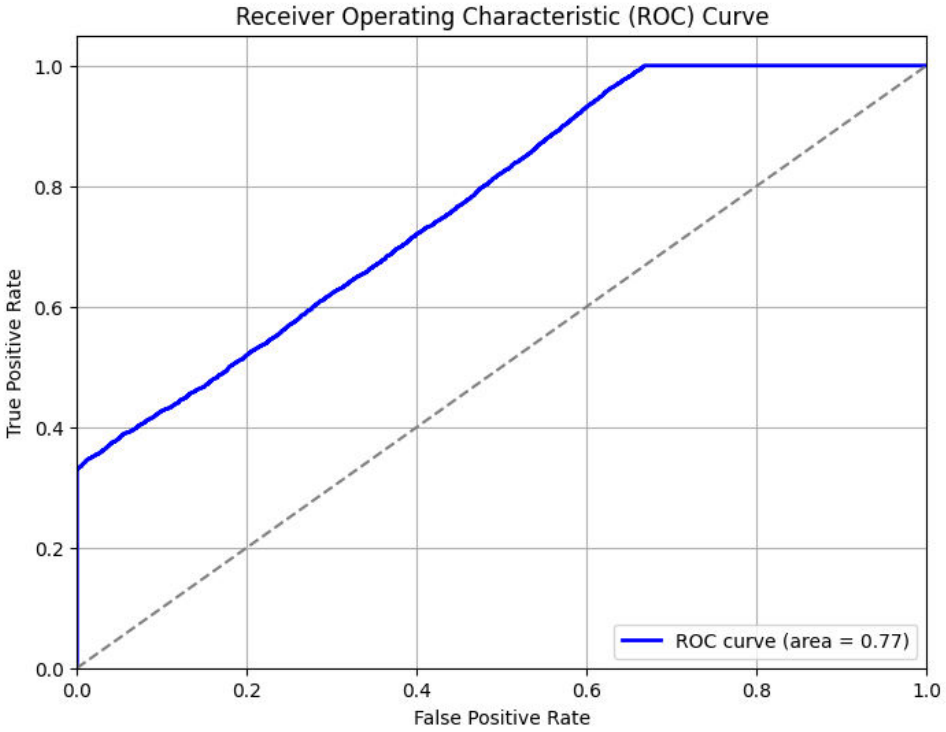

FIG 4. ROC CURVE PLOT

Visualizations: The following visualizations support the evaluation metrics:

- Training and Validation Accuracy and Loss Plots: Illustrate the model's learning process over the epochs.
- Confusion Matrix Heatmap: Visualizing the performance of the model in classifying each test image.
- ROC Curve: The ROC Curve illustrates how the true positive rate and false positive rate are traded off at various thresholds.

V. CONCLUSION

The study describes such an exploratory algorithm, a deep learning technique to distinguish between human-created and AI-generated art. The Convolutional Neural Network (CNN) model introduced in this study has the advantage of having achieved high accuracy and practical use such as content verification or art authentication. Data Augmentation and Regularization techniques were key to making the model more robust and generalizable.

The primary contributions of this research are:

- The creation of a comprehensive dataset with a fair distribution of artworks produced by AI and humans.
- The development and evaluation of a CNN-based model to classify these images with high accuracy.
- The demonstration of the model's effectiveness in practical applications.

Future work will focus on integrating explainable AI techniques, expanding the dataset, and applying the model in real-world scenarios to further validate and refine its performance.

A. Potential Impact and Applications

This research also has wider consequences beyond verifying works of art. This work aids in the larger project of ethics over digital media creation and protection against infringement by providing an accurate way to differentiate content generated by AI vs. humans. With the rise of AI-generated media, such tools will become increasingly important to safeguard digital content on multiple fronts (e.g., social media, journalism, and/or digital marketing).

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| Monthly, Peer Reviewed & Referred Journal|

| Volume 7, Issue 7, July 2024 |

| DOI:10.15680/IJMRSET.2024.0707266 |

Additionally, the techniques used in this analysis also can be generalized to other types of media verification (for example identifying audio or video made by artificial intelligence). This portability underscores deep learning and its power to adapt to the new challenges in the digital media landscape.

B. Limitations and Future Directions

Overall, the study has its limitations which need to be catered to in upcoming works. A key limitation is the likely bias in this dataset to accurately represent all human art and AI-generated art. That said, future research could seek to incorporate a greater variety of styles, periods, and cultural representatives to provide generalizability for the model.
Moreover, it has opaque decisions even though the current model uses high accuracy. The implementation of explainable AI methods will not just aid in increasing the transparency level of the model but also foster increased adoption, buy-in, and trust among various stakeholders within the art community. Alternatively, To further increase the robustness, it would be intriguing for future research to look at the integration of provenance data or multimodal data with textual descriptions.
In conclusion, real-world testing and deploying this model within art galleries, auction houses & digital platforms will bring useful learnings back to help build better models.

C. Final Remarks

The results from this research also demonstrate the possibility of overcoming some challenges brought on by AI-generated art using deep learning. This paper demonstrates that it is possible to train a powerful yet interpretable model on this reconciliation task, paving the way for more sophisticated refinement down the line. As AI technologies continue to advance whilst evolving, enduring advancement will not only present further challenges but also opportunities thus the research constitutes a fundamental step in making sense of this fluid space.

REFERENCES

1. Goodfellow et al., "Generative adversarial networks," Communications of the ACM, 2020.
2. F.-A. Croitoru et al., "Diffusion models in vision: A survey," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023.
3. R. Amoroso et al., "Parents and Children: Distinguishing Multimodal DeepFakes from Natural Images," arXiv preprint, 2023.
4. Y. Hamid et al., "An improvised CNN model for fake image detection," International Journal of Information Technology, 2023.
5. M. A. Rahman et al., "ArtiFact: A Large-Scale Dataset with Artificial and Factual Images for Generalizable and Robust Synthetic Image Detection," arXiv preprint, 2023.
6. S. Aslan et al., "Recognizing the emotions evoked by artworks through visual features and knowledge graph-embeddings," International Conference on Image Analysis and Processing, 2022.
7. D. Smith, "The Art of Authentication: From Connoisseurship to Science," Art Bulletin, 2019.
8. Brown et al., "Deep Learning for Art Authentication: A Survey," Journal of Art Research, 2021.
9. Sabatelli et al., "Analyzing Artistic Styles with Deep Learning," IEEE Transactions on Image Processing, 2020.
10. L. Golan et al., "Explainable AI for Art Authentication," Journal of AI Research, 2022.
11. P. Elgammal et al., "AI Models for Art Attribution and Forgery Detection," International Journal of Computer Vision, 2021.
12. G. Rubinstein et al., "Challenges in Using AI for Art Authentication," AI Magazine, 2023
13. Robin Crockett, Robert Howe. "chapter 10 The Inherent Uncertainties of AI-Text Detection and the Implications for Education Institutions", IGI Global, 2024
14. Shiva Mehta, Vinay Kukreja, Satvik Vats.
15. "Improving Crop Health Management: Federated Learning CNN for Spinach Leaf Disease Detection", 2023 3rd International Conference on Intelligent Technologies (CONIT), 2023

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com