Detection of Art created by AI

Aryan Sahu
MS in Computer Science
asahu27@asu.edu
Arizona State University, Tempe, AZ, USA

Deep Zaveri
MS in Computer Science
dzaveri1@asu.edu
Arizona State University, Tempe, AZ, USA

Anvita Lingampalli
MS in Computer Science
alingam1@asu.edu
Arizona State University, Tempe, AZ, USA

Lakshmi Yogitha Puppala
MS in Computer Engineering
lpuppala@asu.edu
Arizona State University, Tempe, AZ, USA

ABSTRACT

Artificial intelligence has transformed the production of art, however it can be difficult to distinguish AI-generated art from real art. The need for trustworthy detection techniques in a variety of situations is addressed in this work. In order to appropriately recognize AI-generated photos, we suggest a revolutionary method that makes use of cutting-edge algorithms, helping to maintain the integrity of art in the digital age.

KEYWORDS

Artificial Intelligence, Art, Resnet50, Inception, VGG-16, VGG-19, Classification

1 INTRODUCTION

With the introduction of Artificial Intelligence, there has been a downfall in the market of art. There are a lot of Techniques in which real-looking art can be generated using Artificial Intelligence. As a result, art enthusiasts are having a very hard time telling the difference between real and artificial intelligence-generated art.

Due to this difficulty, there is an increasing need for reliable techniques to determine an image's authenticity in a variety of settings, such as internet platforms and art exhibitions. As a result, scholars have been delving into the development of complex algorithms and techniques meant to distinguish between works of art produced by AI systems and those that are truly artistic.

These detection techniques look for features including pixel patterns, stylistic inconsistencies, and underlying algorithms in an effort to offer trustworthy instruments for safeguarding transparency and the integrity of art in the digital era. In this study, we make a contribution to this effort by suggesting a method for precisely identifying AI-generated photos.

2 RELATED WORK

Baraheem and Nguyen (2023) [1] proposed a method for detecting GAN-generated images leveraging transfer learning and multiple classifiers. Their comprehensive dataset compilation covering various synthesis models and input modalities enhanced the model's generalization capability. However, limitations include a focus primarily on GAN-generated images and potential struggles with classifying certain authentic images with specific characteristics. This technique gives an accuracy of 97.8%.

Hong and Zhang (2024) [5] introduced the WildFake dataset tailored for deepfake detection, addressing a crucial gap in the field. While serving as a standardized benchmark, limitations lie in its real-world deployment and effectiveness, along with potential limitations in capturing all nuances present in real-world deepfake images.

Ha et al. (2024) [4] explored distinguishing human art from AI-generated images, highlighting the effectiveness of human artists in detection. However, the reliance on specific detectors and limited exploration of potential solutions for automated detectors were notable limitations.

Martin-Rodriguez et al. (2023) [6] investigated pixel-wise feature extraction techniques for AI-generated image detection, achieving high accuracy rates. Limitations included the lack of discussion on potential implications of misclassifications and scalability challenges.

Castellano and Vessio (2021) [3] provided a comprehensive overview of deep learning approaches to pattern extraction and recognition in visual art. While valuable, limitations include potential overlook of recent developments and bias towards certain approaches or applications.

Nguyen et al. (2023) [7] aimed to discriminate human-drawn and AI-generated human face art through facial feature analysis, with potential implications for art authentication. Limitations included tailored methodologies and challenges in establishing definitive ground truth. The model has an accuracy of 71.43%.

Xi et al. (2023) [8] introduced a novel approach for detecting AI-generated images using a dual-stream network enhanced with cross-attention mechanisms. While promising, limitations included reliance on standard evaluation metrics and potential computational costs.

Bianco et al. (Year) [2] explored Deep Learning models to identify AI-generated art, offering insights into model decision-making processes. Limitations included potential biases in the dataset and challenges in interpretability. Their Resnet50 model has an accuracy of 96.54%.

Overall, recent literature underscores the importance of AI-generated image detection, with advancements in methodologies and datasets. However, limitations persist in the generalization of detection models, real-world applicability, interpretability of results, and ethical considerations. Future research should address these limitations to advance the field towards robust and ethically sound AI-generated image detection systems.

3 DATASET

The AI Generated Images vs Real Images dataset has been used for this project. It consists of 539 AI-generated Art-Images and 436 Real Art-Images. This dataset was selected to enable thorough study and comparison since the photos were representative of a range of subjects and styles frequently found in AI-generated artwork.

4 METHODOLOGY/IMPLEMENTATION

We have used the four models and compared their results and found the most appropriate model for detecting Ai generated art. We have used the following models - Resnet50, Inception, VGG-16, VGG-19 pre-trained models for the classification of Ai generated and real art images.

4.1 Data Preprocessing

A fair representation of both AI-generated and real images were a priority in the preparation of the dataset for our study, as this is essential for efficiently training and assessing models. Three subsets of the dataset were created: test, validation, and training sets. Ten percent of the data were set aside for the test and validation sets, and the remaining eighty percent were allocated to the training set in accordance with normal practice.

In order to ensure uniformity and streamline the processing process, every image was resized to a standard 224 x 224 pixel size in the case of Resnet50, VGG-16, VGG-19 models and for the Inception model, the images were resized to 299 x 299 pixel size. Additionally, all picture pixel values were standardized to lie between 0 and 1 in order to improve computing performance and guarantee consistent behavior across all systems. By lowering the dynamic range of pixel values, normalization facilitates learning and convergence of the model during training.

It is crucial to decide how to balance the dataset so that each subset contains an equal amount of genuine and AI-generated photos. This method ensures fair representation and strong generalization by reducing bias towards any specific class during model training and evaluation. We sought to establish a consistent and trustworthy dataset that would support the creation and assessment of machine learning models for image classification tasks, specifically in the area of differentiating between AI-generated and real images. To this end, we followed these pre-processing steps and dataset partitioning strategies.

4.2 Model Preparation

First, the ResNet50, VGG-16, VGG-19 and Inception models were loaded from the Tensorflow.Keras library. The final fully linked layers, which handled the initial classification task, were eliminated in order to modify the model for the particular goal of differentiating between AI-generated and actual photos. Then, a new fully linked layer was introduced. It had two output nodes and was designed to distinguish between AI-generated and genuine images. With this adjustment, the model was able to produce binary classifications that were appropriate for the goals of the study.

The original model layers' weights of the respective models were locked once the architectural changes were made to guarantee that the learnt representations would not change while being fine-tuned. These layers were frozen so that the newly added fully connected layer could be tailored specifically for the identification of AI-generated images, while the model was still able to use the useful features that were taken during the pre-training phase.

This method made it easier to fine-tune the pre-trained models to distinguish between AI-generated and actual images—an important task in a variety of fields, including fraud detection, picture forensics, and content moderation. This methodology comprises a methodical process of model adaption that is customized to the particular needs of the study with the goal of improving the model's performance in the intended classification task.

4.3 Model Training

Prior to beginning the training process, the optimizer and loss function—two essential factors for a successful training—were configured. Because of its ability to handle big datasets and varying learning rates, the Adam optimizer was chosen. In comparison to other optimizers, Adam's method allows for faster convergence by computing adaptive learning rates for every parameter. For this multi-class classification assignment, the categorical cross-entropy loss function was appropriate as it quantified the discrepancy between the actual and anticipated class distributions. The degree to which a model's probabilistic outputs during training agree with the true labels is measured by cross-entropy loss.

The model started training on the designated dataset after compilation. The model was able to keep its ability to generalize while making the best use of the memory that was available. The entire dataset was processed in mini-batches of 32 samples (batch size) throughout each epoch. In order to provide outputs that could be compared to the genuine labels, the model ran forward passes. The designated optimizer was then used to backpropagate errors in order to update weights and minimize the loss function.

Iteratively, over 100 epochs, the model improved its parameters through this forward and backward propagation procedure. The model has enough time to converge and learn the training data distributions without overfitting with the number of epochs was set to 100.

The model's performance was simultaneously tracked on an unrelated validation dataset that was not used for training. This holding set was an approximation of the model's generalization performance to unknown data.Training was stopped to prevent overfitting once validation metrics peaked or began to decline. Thus, during the entire training process, the validation set was essential for model selection, hyperparameter adjustment, and minimizing overfitting.

4.4 Model Evaluation

Following a successful training process, the model's efficacy was determined by carefully analyzing its performance on a test set using a variety of criteria. F1-score, recall, accuracy, and precision were among the metrics used. A number of phases were included in the evaluation process, such as calculating testing and training accuracy's and losses. To give a thorough grasp of the model's behavior, additional computations for the training and testing stages included precision, recall, and F1-score. Through a methodical examination of these indicators, a comprehensive evaluation of the

model's abilities was attained, illuminating both its advantages and disadvantages.

5 RESULTS

A quantitative assessment of the models VGG16, VGG19, ResNet50, and Inception performance was conducted using measures such as test accuracy, F1 score, recall, and precision. The results are shown in table 1, 2, 3, and 4. The training and testing accuracies for VGG16 model was the highest and got around % Training accuracy and 83.33% Testing accuracy.

Table 1: VGG16 Model Results

Metrics	Values
Test Accuracy	0.8333333134651184
Precision	0.8160835762876579
Recall	0.8163265306122449
F1- Score	0.8158605514087242

Table 2: VGG19 Model Results

Metrics	Values
Test Accuracy	0.703125
Precision	0.7843379212149976
Recall	0.7653061224489796
F1- Score	0.7650372863922201

Table 3: Inception Model Results

Metrics	Values
Test Accuracy	0.6875
Precision	0.7035392787251887
Recall	0.6836734693877551
F1- Score	0.6637834866914928

Table 4: ResNet50 Model Results

Metrics	Values
Test Accuracy	0.6224489808082581
Precision	0.6307692307692307
Recall	0.7592592592592592
F1- Score	0.6890756302521008

6 TESTING RESULTS

The results obtained by the four models are shown in figures below.

Once the models were successfully trained, they were used to forecast image authenticity and distinguish between real and artificial intelligence (AI)-generated images. The possibility that an image will fall into either group is indicated by the probability ratings for each class in these forecasts.

Out of the four models that were trained and tested, VGG16 produced the best test accuracy of 83.33%, whereas VGG19, Inception, and ResNet50 models produced the accuracy of 70.3%, 68.75%, 62.24% respectively.

6.1 Results for VGG16

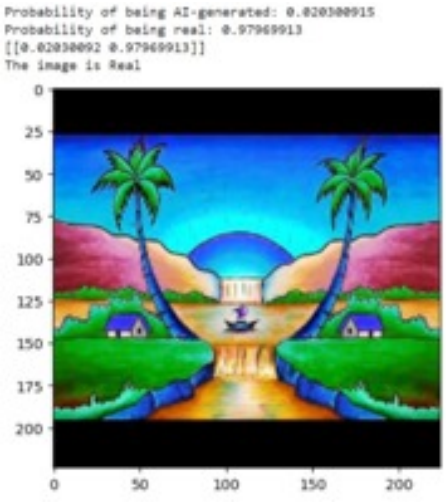

Figure 6.1.1: Real Image

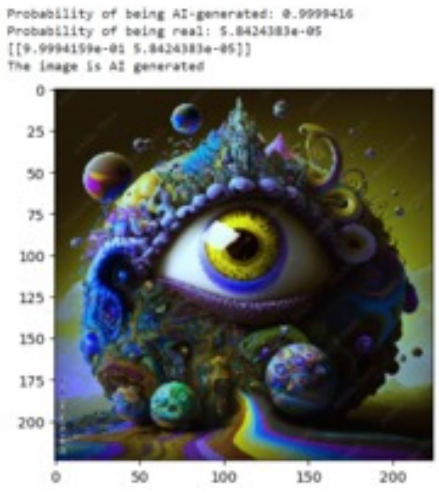

Figure 6.1.2: AI Generated Image

Figure 6.1.3: Real Image

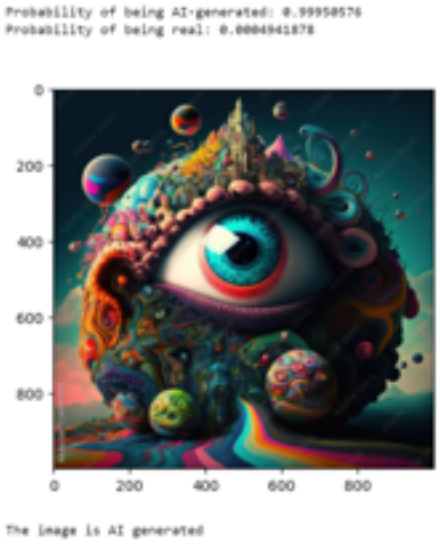

Figure 6.2.2: AI Generated Image

The VGG16 model, when presented with a real image, it predicts its authenticity with a probability of 0.97 for realness and a 0.02 for AI generation. When analyzing an AI-generated image, it identifies it as such with a probability of 0.99 for AI generation and a 5.84e-05 for realness. The overall accuracy of the VGG16 model across numerous testing images is standing at 0.83. This indicates that the model performs well in 97% of the cases, thus being the best model out of all.

6.2 Results for VGG19

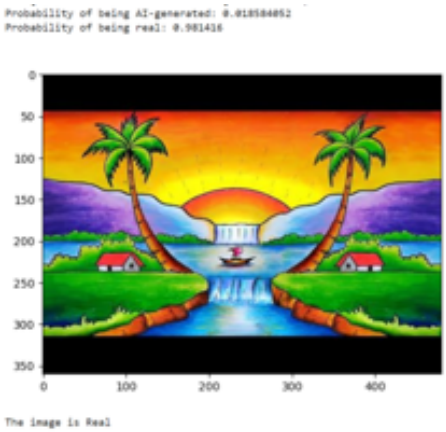

Figure 6.2.1: Real Image

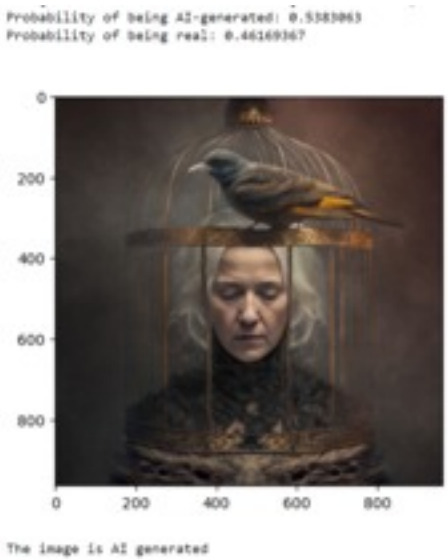

Figure 6.2.3: AI Generated Image

The VGG19 model, when presented with a real image, it predicts its authenticity with a probability of 0.98 for realness and a 0.01 for AI generation. When analyzing an AI-generated image, it identifies it as such with a probability of 0.99 for AI generation and a 0.00049 for realness. The overall accuracy of the VGG19 model across numerous testing images is standing at 0.703. Though the model's accuracy is not as best as that of VGG16 it still is the second best model.

6.3 Results for Inception

Figure 6.3.1: Real Image

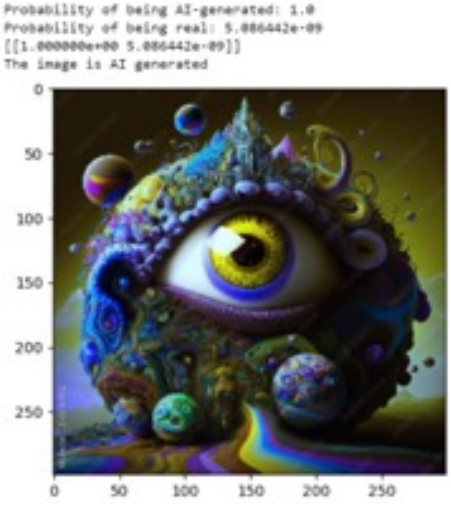

Figure 6.3.2: AI Generated Image

Figure 6.3.3: AI Generated Image

The Inception model, when presented with a real image, it predicts its authenticity with a probability of 0.53 for realness and a 0.46 for AI generation. When analyzing an AI-generated image, it identifies it as such with a probability of 1.0 for AI generation and a 05.08e-09 for realness. The overall accuracy of the Inception model across numerous testing images is standing at 0.688. This indicates that while the model performs well in many cases, there remains room for improvement in accurately classifying a broader range of images.

6.4 Results for ResNet50

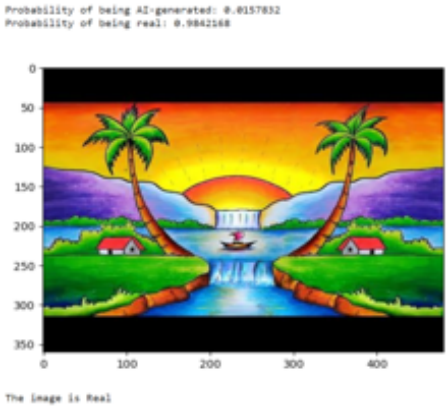

Figure 6.4.1: Real Image

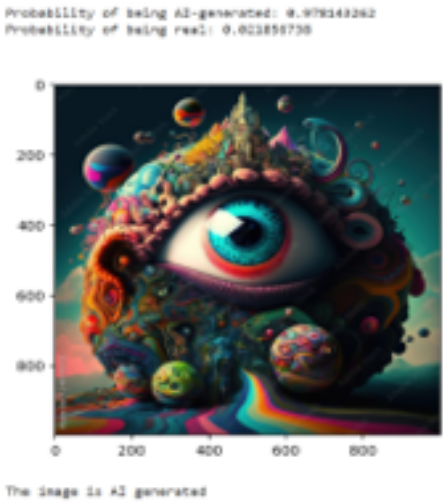

Figure 6.4.2: AI Generated Image

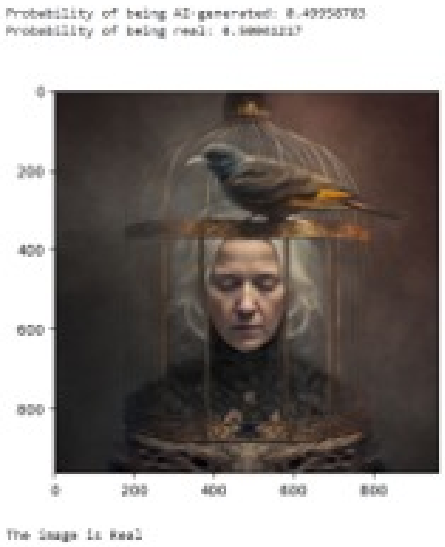

Figure 6.4.3: Real Image

The ResNet50 model, when presented with a real image, it predicts its authenticity with a probability of 0.98 for realness and a 0.015 for AI generation. When analyzing an AI-generated image, it identifies it as such with a probability of 0.97 for AI generation and a 0.02 for realness. The overall accuracy of the ResNet50 model across numerous testing images is standing at 0.622, making it the worst model compared to the other models.

7 CONCLUSION

To sum up, our project has become able to distinguish between artificial intelligence (AI)-generated and authentic art, which is essential to maintaining the integrity of art in the digital age. By training and testing four well-known models, VGG16, VGG19, ResNet50, and Inception, we have identified VGG-16 as the best model and created a framework for identifying images created by artificial intelligence. We improved our accuracy by thoroughly training the models using a vast and diverse dataset.

REFERENCES

[1] S.S. Baraheem and T.V. Nguyen. 2023. Ai vs. AI: Can ai detect AI-generated images?'. *Journal of Imaging* 9, 10 (Sept. 2023), 199. https://doi.org/10.3390/jimaging9100199

[2] Tommaso Bianco, Giovanna Castellano, Raffaele Scaringi, and Gennaro Vessio. 2023. Identifying AI-Generated Art with Deep Learning.

[3] Giovanna Castellano and Gennaro Vessio. 2021. Deep learning approaches to pattern extraction and recognition in paintings and drawings: an overview. *Neural Computing and Applications* 33 (10 2021). https://doi.org/10.1007/s00521-021-05893-z

[4] Anna Yoo Jeong Ha, Josephine Passananti, Ronik Bhaskar, Shawn Shan, Reid Southen, Haitao Zheng, and Ben Y. Zhao. 2024. Organic or Diffused: Can We Distinguish Human Art from AI-generated Images? arXiv:2402.03214 [cs.CV]

[5] Yan Hong and Jianfu Zhang. 2024. WildFake: A Large-scale Challenging Dataset for AI-Generated Images Detection. arXiv:2402.11843 [cs.CV]

[6] Fernando Martin-Rodriguez, Rocio Garcia-Mojon, and Monica Fernandez-Barciela. 2023. Detection of AI-Created Images Using Pixel-Wise Feature Extraction and Convolutional Neural Networks. *Sensors* 23, 22 (2023). https://doi.org/10.3390/s23229037

[7] Minh-Quang Nguyen, Khanh Ho, Hoang-Minh Nguyen, Canh-Minh Tu, Minh-Triet Tran, and Trong-Le Do. 2023. Unmasking The Artist: Discriminating Human-Drawn And AI-Generated Human Face Art Through Facial Feature Analysis. 1–6. https://doi.org/10.1109/MAPR59823.2023.10289113

[8] Ziyi Xi, Wenmin Huang, Kangkang Wei, Weiqi Luo, and Peijia Zheng. 2023. AI-Generated Image Detection using a Cross-Attention Enhanced Dual-Stream Network. arXiv:2306.07005 [cs.CV]