AI Art Detection and Methods of Classification using Deep Learning

Tharrun S
School of Computer Science (SCOPE)
Vellore Institute of Technology
Amaravati, India
tharrunsatish@gmail.com

Nitin Kausik Remella
School of Computer Science (SCOPE)
Vellore Institute of Technology
Amaravati, India
kausikremella@gmail.com

Smit Kavani
School of Computer Science (SCOPE)
Vellore Institute of Technology
Amaravati, India
smitkavani21@gmail.com

***Abstract*—In the context of AI-powered art discovery, this research focuses on the investigation of CNN architecture, ResNet, and MobileNetV2. The main goal is to create an accurate and effective system that can recognize AI art from real art within enormous art collections. This is accomplished by compiling a vast dataset of artworks from diverse historical eras and artistic movements. Metrics like accuracy, precision, and recall are used to systematically assess the performance of the aforementioned CNN architectures. The variance in accuracy displayed by these architectures across many artistic styles is particularly intriguing, indicating both their unique aptitudes and limitations in catching subtle aspects of different types of art. Notably, this work forgoes the use of transfer learning strategies in favor of concentrating on the built-in capacities of these models without the use of external pre-trained weights. This method contributes to a better understanding of these structures' intrinsic capacities for detecting art by revealing insights into their unadulterated potential. The results of this study provide useful direction for choosing appropriate CNN architectures for particular art identification tasks, assisting in the creation of AI-driven systems for art analysis, curation, and preservation. This study contributes to the advancement of AI-powered creativity and understanding by shining light on the intricacies of architectural performance in a field as complex as art.**

Keywords: Deep Learning, AI Art Detection, Classification.

I. Introduction

AI art is an exciting and quickly developing topic that has emerged in recent years from the fusion of artificial intelligence (AI) with the art world. This field includes the creation, manipulation, and replication of artistic content using cutting-edge machine-learning methods. It is essential to develop techniques for reliably detecting and differentiating between AI-generated artworks and those produced by human artists as AI-generated art becomes more prominent in galleries, internet platforms, and cultural conversations. This study explores the difficulties, approaches, and effects of detecting artworks created by artificial intelligence. It focuses on the crucial topic of AI art detection.

For the art world and society in general, the development of AI-generated art brings both potential and challenges. On the one hand, AI provides creative and experimental tools for artists, allowing them to foray into new expressional spheres. On the other hand, the quick development of AI techniques has resulted in the creation of works that question conventional ideas of originality, authorship, and the function of human creativity. There are concerns about the moral ramifications of giving artificial intelligence-generated art a monetary value as it enters galleries and the art market.

There are different types of art, each with its own distinct styles and techniques. Robust training data and sophisticated feature extraction approaches are needed to build models that can correctly categorize and distinguish between these various types. Large and superior-quality datasets are essential for developing precise AI algorithms for art detection. However, because of factors like variances in lighting, image quality, and artistic interpretation, many artworks may have minimal data accessible or the data may be noisy and inconsistent.

This study's main objective is to address the demand for trustworthy and effective AI art-detecting techniques. Examining the numerous methods and approaches designed to discern between art produced by AI and art produced by humans. Analyzing brushstrokes, texture, color schemes, and other visual elements may be necessary for this, along with taking metadata and the context of the image into account. The ability of AI models to accurately mimic a variety of artistic styles makes it hard to distinguish between AI-generated art and human-made art. Taking into account a variety of visual and contextual indicators that could offer insights into an artwork's origin, this paper aims to investigate the technological developments and tactics used to recognize AI-generated art.

II. Literature Survey

Angus Forbes et al [1] explores the development of Artificial Intelligence and Machine Learning in most recent years and proposes implementations of these aforementioned concepts to produce Creative AI. The research emphasizes the use of algorithms that identify, manipulate, or replicate data in order to facilitate generative AIs or multi-modal mapping of the user input to media output to learn feature detection at a deeper level.

Eva Cetinic's and James She's [2] research gives a comprehensive examination of the two aspects of artificial intelligence that can be used in art: one being for the specific of analysis of art and how it can be applied to collections of digital artworks; and the other being how AI can be utilized for creative purposes and production of new artwork. In the context of AI training, various practical and theoretical aspects of AI art are taken into consideration for a concise projection of AI art's future progression with respect to detection, classification, and multimodal mapping.

Michael Mateas [3] talks about expressive AI which as quoted by the authors is the "inter-discipline of AI-based cultural production, combining art practice and AI-research practice" exploring the possibilities of creative AI and applications of Creative AI also by giving some context to the influence of such AI on the life of people either positive or negative. The research also sheds light on human-AI interactions and finally tops it off with how Expressive AI that can include generational functions can be more than just an application or a tool in a far wider scope.

Mateja Culjak et al [4] paper explains and details an approach to an automatic art genre classification and how to go about fine-tuning the model of such caliber to give efficient speeds and efficient feature extraction. Provides a large and articulate data set with over *15,000 images* over six classes of various types of paintings on which using image detection extensions of CNN have had different success rates for each class providing us some idea on the working of CNN in learning features of art work.

Maciej Wiatrak et al [5] provides an overview of the challenges faced in training Generative Adversarial Networks (GANs) and introduces the purpose of the survey, which is to comprehensively review the methods proposed in the literature for stabilizing GAN training. It mentions the instability problems encountered during GAN training, such as non-convergence, vanishing or exploding gradients, and mode collapse, and highlights the importance of addressing these issues for effective GAN training

Ahmed Elgamma et al [6] introduces a new system called Creative Adversarial Networks (CAN) for generating art. The system learns about different artistic styles and enhances creativity by deviating from those styles. The authors conducted experiments showing that human subjects couldn't distinguish art generated by CAN from art created by contemporary artists displayed in top art fairs.

Vivek Kanji malam [7] is a Neural network research that aims to classify digits and doodles similar to the functioning of the mnist dataset. The model used in this research is trained using Cifar10 dataset also in adddition with the Minst dataset to identify over 10 classes of objects.

Goodfellow et al [8] introduces the concepts of a Generational Adversarial Network. The text discusses the phenomenon of adversarial examples, which are inputs that are intentionally perturbed to cause machine learning models, including neural networks, to misclassify them with high confidence. Early explanations focused on nonlinearity and overfitting, but the authors argue that the primary reason for neural networks' vulnerability to adversarial perturbations is their linear nature. They propose a fast method of generating adversarial examples and show that adversarial training can provide additional regularization benefits.

Alzantot et al [9] is a brief study on the use of Synthetic data and model architectures that are able to create synthetic data capable of providing more data for models to train more adequately similar to a kin of augmented data. Explains sophisticated models that use generators and discriminators and their purposes.

Castellano et al [10] This paper gives an overview of deep learning approaches used in the field of visual arts, specifically for pattern extraction and recognition in painting and drawing. Recent advancements in deep learning and computer vision, along with the availability of large digitized art collections, have provided opportunities for computer science researchers to develop automatic tools for analyzing and understanding visual arts. This deeper understanding can make visual arts more accessible to a wider audience, contributing to the dissemination of culture.

Goldberg [11] textbook on natural language processing explains in great detail the subsetted concept of deep learning. Provides the implementation of Deep learning models in the fields of NLP and LLM. The book provides a in-depth text on RNN and its variation such as LSTMs and their application and limitations.

Eva Cetinic's [12] studies the digitization of fine art collections that has led to an increase in the availability and preservation of artworks, making them accessible to a wider audience. This paper explores different methods of extracting image features to classify paintings by genre. By using a pre-trained deep convolutional neural network, the authors achieved an accuracy of 77.57% in genre classification, highlighting the potential of computer vision techniques in automating the identification of painting characteristics and generating metadata.

Cetnic et al [13] studies the fine-tuning of Convolutional Neural networks of which are implemented in the context of fine art classification. With the increase of the digital art space, the research paper proposes a CNN network named CaffeNet a modified CNN with five conv2d layers and three fully connected layers, using Relu activation functions. The research paper delves into a fine-tuning setup for the proposed CNN architecture and also analyses the impact that domain-specific weight initialization [14] has on it.

Qian Xiang et al [15] discusses fruit image classification as an important technology for profitable fruit-picking robots and increasing competitiveness in the global fruit market. Although deep learning, especially DCNN, excels at classifying images, the resource requirements make them unsuitable for resource-constrained contexts such as automated harvesting robots. To balance resource constraints and accuracy, the study uses MobileNetV2 lightweight neural network with transfer learning method. This approach replaces the upper layer of the pre-trained MobileNetV2 network with a convolution layer and a Softmax classifier, incorporating dropout to reduce overfitting. Through a two-step training process with the Adam optimizer, the proposed method achieves a classification accuracy of 85.12% on a fruit dataset of 3,670 images. Compared to other networks such as MobileNetV1, InceptionV3 and DenseNet121, this hybrid network offers a favorable balance between accuracy and speed, making it possible for the network to be deployed on low power devices such as mobile phones

Dhananjay Theckedath and RR Sedamkar [16] discusses the importance of effect sensing in human-computer interface system and presents a study using convolutional neural network (CNN) with transfer learning method to detect 7 affect states basic. The article compares three pre-trained networks: VGG16, ResNet50 and SE-ResNet50, integrating the new architecture block. The networks were trained and evaluated on the image dataset, achieving validation accuracy of 96.8%, 99.47%, and 97.34% for VGG16, ResNet50 and SE-ResNet50, respectively. The rating also takes into account accuracy and recall, indicating the ability to accurately detect effects across all networks, with ResNet50 being the most accurate.

Eva Cetnic et al [13] discusses the application of Convolutional Neural Networks (CNN) for fine art classification. The authors explore the use of CNNs for various art-related image classification tasks, including artist, genre, style, time period, and national artistic context classification. They also investigate the transferability of deep representations across different domains and demonstrate the practical applicability of their results in enhancing search systems for online art collections.

Gaozhong Tang et al [17] addresses the challenge of predicting crowd flows in urban areas by leveraging both spatial and temporal features. The authors propose a novel approach that extracts spatial features from city maps using convolutional neural networks and incorporates a sequence feature fusion mechanism to merge spatial and temporal features for accurate crowd flow prediction.

Jia Deng et al [18] discusses the introduction of a new database called ”ImageNet,” which is a large-scale ontology of images built upon the WordNet structure. ImageNet aims to provide a vast collection of annotated images organized according to the semantic hierarchy of WordNet, offering opportunities for research in computer vision and beyond. The paper highlights the scale, accuracy, diversity, and hierarchical structure of ImageNet and demonstrates its usefulness through applications in object recognition, image classification, and automatic object clustering.

Mondal et al [19]Implies Artworks and paintings have been an important part of human civilization since ancient times, providing valuable insights into various subjects. Archiving digital versions of paintings helps preserve the works of different painters. In this study, a conventional Neural Network is used to classify artworks, focusing on both foreign and Indian painters, with an average accuracy of 85.05%. Provides more proof that CNN's are able to feature extract learn art in a more atomic level.

Johnson et al [20] explores image processing tools that are capable of assisting historians and such. Research has been conducted in accordance with the Van Gogh and Kroller-Muller museums to create a data set consisting of 101 high-resolution gray-scaled images that can be used for image processing.

Mark Sandler et al [21] introduces a new mobile architecture called MobileNetV2, which improves the performance of mobile models across various tasks and model sizes. It also discusses the application of these models to object detection and semantic segmentation.

kannan K [22] is a project on Kaggle that is a use case for the current problem that this paper tries to delve into and start. The project has achieved an accuracy score of 91.0% with the assistance of the transfer learning mechanisms of MobileNetV2 architecture.

shen et al [23] This paper explores how with the advancement of artificial intelligence (AI) technology, art creation is becoming more diverse and interactive, driven by intelligent, data-driven content expression. AI aims to replicate human cognitive abilities, enabling natural responses, emotion decoding, and recognition of human traits. This has led to a shift in interactive art, focusing on integrated, interactive, and emotionally engaging artistic expressions that study natural human behavior and combine it with intelligent systems. In the context of the research article, the authors explore the intersection of AI technology and interactive art, analyzing their historical development and proposing the impact of AI on creative thinking, creative modes, and artistic experiences in interactive art.

III. Proposed Methodology

A. Construction of Dataset

The goal of this study is to train a model to learn the differences in the finer details of art generated by AI tools, such as Dall - E, with that of human-drawn art. This requires the construction of a balanced and robust dataset that consists of two classes “AI Generated Art” and “Real Human Art”. The dataset constructed is a compilation of various datasets that are previously used for the same or any similar test cases, such as that in art genre classification [12], doodle classification [7], and a prior AI art classifier that uses MobileNetV2 [22]. Overall in conclusion the dataset acquired consists of around 32,000 images of classified art. The dataset comprises around 14,500 real images, while around 16,000 AI Generated images. In this scenario, the difference in sample images seems to assist in the prevention of overfitting of the model.

B. Data Pipeline and Preprocessing

For creating a data pipeline the ”image dataset from directory” function that is provided under the keras preprocessing module is used. This helps in creating a dataset whilst preprocessing the data helping to reduce the burden of doing the same at later stages of the model's creation.

The image size is set to a standard 256 also automatically resizing images that are of different dimensions. The image dataset from directory function converts images into a Numpy array format which is iterated through using a NumPy iterator.

A 70:20:10 split is used to split the dataset into train, test and validation respectively.

Fig. 1. Sample Images from Dataset

C. Deep learning model

1) Generational Adversarial Network: Generational Adversarial Neural Network (GANS) which was proposed by Ian Goodfellow et al. [8] is an unsupervised deep learning method which uses two networks, a generator, and a discriminator network. The generator network is capable of generating synthetic data [9] which is used to train the discriminator network along with real data. The goal of the generator is to maximize the discriminator loss while the discriminator's goal is to reduce generator loss that is when the discriminator is not able to differentiate between real samples and the synthetic samples. In more detailed and technical senses a GAN that can produce artwork is referred to as Creative Adversarial Network (CAN) [6].

Hence GANS could be considered as a suitable candidate for the classification of images. The major limitation of a GAN is that it faces mode collapse, where the model starts to explore a small set of patterns that will produce similar and recurrent shapes [5]. Thus making it much more difficult to quantitatively evaluate and compare GAN architecture [10] making it less viable for this context.

2) Recurrent Neural Networks: A type of neural network that works well with the use of sequential data such a time series data making it most apt for natural language processing purposes [11]. That being said there is not much potential for RNNs when it comes to visual art implementations. Data is rather spatial in nature in this context and not temporal in nature [17] which makes RNNs and its variation an unviable method to solve the problem. For a decent result, it is more often than not used along with Convolutional Neural Networks.

3) CNN: For image classification problems, the Convolutional Neural Network is the most conventional model to implement. There have been pleather of such use cases before where a CNN is used to classify art. Such is the case of using AI to classify art by genres [13], Art painting identification [19], or Artist identification [20]. These confirm that CNNs are capable of learning art datasets and distinguishing them upon the differences in what can be called as "fingerprints" [23] which could include the learning of color usage and brush strokes which impacts AI art in a significant way. The theory is that the model is able to pick up on the erratic and more randomly off-put strokes and the methodical usage of colors done by AI-generated art models.

4) MobileNetV2: MobileNetV2 is a pre-trained transfer learning Convolutional Neural Network model where the pre-trained model which is trained on over a million images from the ImageNet database [18]. It is predominantly used for mobile applications or any sort of embedded applications [21]. MobileNetV2 uses ReLu activation since ReLu has high information-retaining capabilities [21].

Input	Operator	t	c	n	s
$224^2 \times 3$	conv2d	-	32	1	2
$112^2 \times 32$	bottleneck	1	16	1	1
$112^2 \times 16$	bottleneck	6	24	2	2
$56^2 \times 24$	bottleneck	6	32	3	2
$28^2 \times 32$	bottleneck	6	64	4	2
$14^2 \times 64$	bottleneck	6	96	3	1
$14^2 \times 96$	bottleneck	6	160	3	2
$7^2 \times 160$	bottleneck	6	320	1	1
$7^2 \times 320$	conv2d 1x1	-	1280	1	1
$7^2 \times 1280$	avgpool 7x7	-	-	1	-
$1 \times 1 \times 1280$	conv2d 1x1	-	k	-	

Fig. 2. MobileNetV2 Architectue

MobileNetV2 with its complex architecture can be highly power-consuming and computational process to train. It takes an exponentially longer time to process and hence may seem a non-viable option. There exists a prior approach in trying to solve this problem using a CNN but with the use of a MobileNetV2 architecture [22] but for this study the goal is to formulate a model that as much as possible tries to reduce the use of transfer learning.

D. Proposed model

Hence the final model that is proposed is a CNN The base working model consists of Convolutional layers, Max Pooling layers, Flatten Layer, Dense Layers, and various Batch Normalization Models. Most of which are mentioned are standard for a Convolutional Neural Network borrowing Batch Normalization layers. The convolutional layers as known are used for feature extraction. Combined with max pooling we get a standard feature learning layer but the problem arises while training the model. While training it was noticed that after each epoch, a complete iteration of that data does through the network, the validation accuracy, and the validation loss estimated by the model was largely inconsistent when compared to the training accuracy and training loss, which is calculated by binary cross entropy with it being a binary classification problem. This is countered by the use of a data

$$H_p(q) = -\frac{1}{N}\sum_{i=1}^{N} y_i \cdot log(p(y_i)) + (1 - y_i) \cdot log(1 - p(y_i))$$

Binary Cross-Entropy / Log Loss

Fig. 3. Binary Cross Entropy Calculation

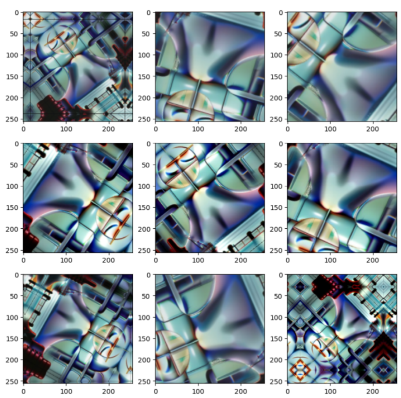

Fig. 4. Data Augmentation Sample

augmentation layer at the beginning of the model and batch normalization after each Convolutional/Max Pooling layer. Along with getting more consistent Validation accuracies and Validation losses, which proves that the model is learning, it also gives a comparatively better accuracy yield compared to the previous methods. The best results were observed when Random zoom, flips, rotations and contrast data augmentations were added. These additions provide the model with additional data to work with giving us better yields.

The model as shown in Figure 5 comprises 6 Conv2d layers which are combined with corresponding Max Pooling layers. These layers handle the feature extraction of given inputs. These are followed by dropout layers and batch normalization layers alternatively. These help in regularization of the data, i.e. prevention of overfitting, and normalizing training values in times of inconsistency. Following the flatten layer there are three dense layers with two dropout layers which compute for a result with the extracted features as input. Overall there are 651,009 parameters in which 650,625 parameters are trainable, whilst 384 parameters are nontrainable parameters provided by dropout, batch normalization, and max pooling layers. For the Convolutional Layers the activation function used is ReLu activation [24], which in this context is an exceptionally well-performing activation function as it is most resistant to the vanishing gradient problem [25] during the training phase of the model [26]. Figure 6 shows the activation graph for ReLu. ReLu is suitable during the training of the model [27] since it

Layer (type)	Output Shape	Param #
sequential (Sequential)	(None, 256, 256, 3)	0
conv2d (Conv2D)	(None, 254, 254, 64)	1792
max_pooling2d (MaxPooling2D)	(None, 127, 127, 64)	0
conv2d_1 (Conv2D)	(None, 125, 125, 128)	73856
max_pooling2d_1 (MaxPooling2	(None, 62, 62, 128)	0
dropout (Dropout)	(None, 62, 62, 128)	0
conv2d_2 (Conv2D)	(None, 60, 60, 64)	73792
max_pooling2d_2 (MaxPooling2	(None, 30, 30, 64)	0
batch_normalization (BatchNo	(None, 30, 30, 64)	256
conv2d_3 (Conv2D)	(None, 28, 28, 128)	73856
max_pooling2d_3 (MaxPooling2	(None, 14, 14, 128)	0
dropout_1 (Dropout)	(None, 14, 14, 128)	0
conv2d_4 (Conv2D)	(None, 12, 12, 128)	147584
max_pooling2d_4 (MaxPooling2	(None, 6, 6, 128)	0
batch_normalization_1 (Batch	(None, 6, 6, 128)	512
conv2d_5 (Conv2D)	(None, 4, 4, 64)	73792
max_pooling2d_5 (MaxPooling2	(None, 2, 2, 64)	0
flatten (Flatten)	(None, 256)	0
dense (Dense)	(None, 512)	131584
dropout_2 (Dropout)	(None, 512)	0
dense_1 (Dense)	(None, 128)	65664
dropout_3 (Dropout)	(None, 128)	0
dense_2 (Dense)	(None, 64)	8256
dense_3 (Dense)	(None, 1)	65

Fig. 5. Model Architecture

prevents the vanishing gradient problem but it is not as suitable to deduce the final output that the model is to give. In general, for this scenario softmax functions or sigmoid functions are more preferred [27]. Softmax function [28] is more widely used since it is more resistant to vanishing gradient than the sigmoid function [29]. The softmax function as given in Fig.7 is much more suitable for multi-class classification problems than binary classification [28].

Leaving us with the best activation function in this scenario which is sigmoid activation [30]. Sigmoid produces a function value in between 0 - 1 which is precisely what is required. Sigmoid function here is given by Fig 8.

The final output received is a number ranging from 0-1 in which any image with output above 0.5 is classified as a real image and an image with output less than 0.5 is classified as an AI-generated image.

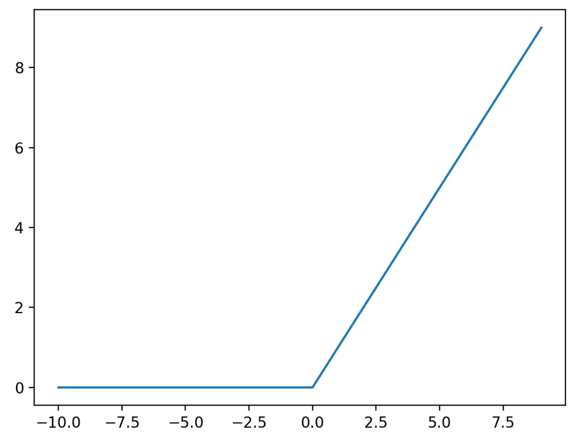

Fig. 6. Relu Activation graph

$$\text{softmax}(z_j) = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}} \text{ for j = 1,...,K}$$

Fig. 7. Softmax Activation function

IV. Results

The goal of the model is to be able to identify the probability of the Image fed to the model to be an AI-generated image. The model is trained on 620 batches per epoch with a total of 20 epochs. The validation accuracy of the model in the range of 2 – 10 is hovering around the range of 75 to 80 accuracy which indicates that the model is learning the data slowly but preventing overfitting of the data. Around the 10 – 19 epoch range we can see a significant rise in the rate of learning important data points as the accuracy of the validation is in the range of 80 to 87 with 87 being the best performance of the model. The blue line indicates the model when is it is training on unseen data i.e. validation data whereas the red line indicates the model's accuracy in the proper classification of data that it is previously trained on i.e. training data.

Empirically, accuracy seems like quite a limited measure of the quality of predictions. To predict whether a sample belongs to a class or not, our model outputs a number in the range of [0,1] where 0 or 1 being ground truth and ground false. To calculate accuracy, we take some arbitrary threshold (the most common being 0.5) to calculate whether the prediction in True or False. Since this threshold can be subject to change, accuracy can't be considered a correct or true indicator of the dataset

An accurate measure to check if the model is being trained properly is by observing the loss and validation loss or the

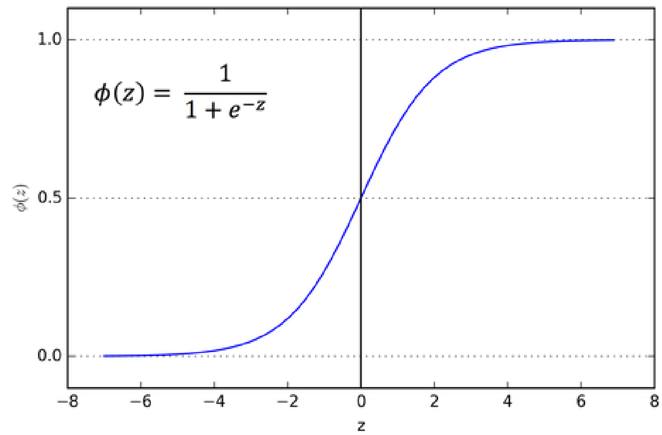

Fig. 8. Sigmoid Activation function

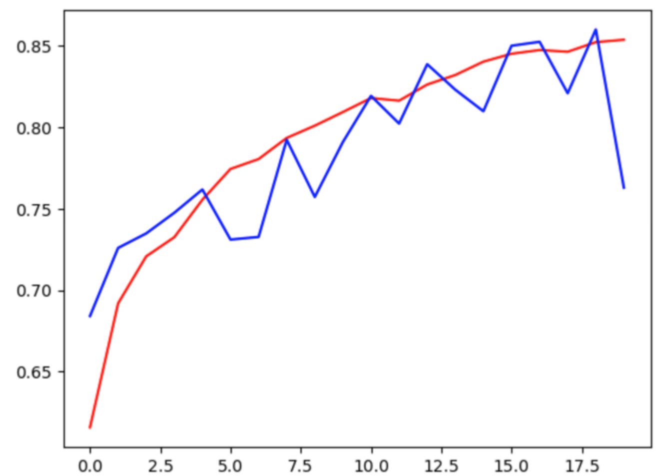

Fig. 9. Validation Accuracy and Training Accuracy

loss that occurs due to predictions made on unseen data. If the Validation loss is increasing the there can be two major reasons for that. Either the model is cramming values(validation accuracy is decreasing) and not learning the data or there is some overfitting present(validation accuracy is increasing).

The model is trained on 620 batches per epoch with a total of 20 epochs. The validation loss of the model is not consistent and results in peaks during the training of the model. This can be extrapolated for two reasons. The data set is not consistent such as unclear images or the model failing to see some major patterns in the dataset. The model Loss is a continuously decreasing slope from left to right indicating that the model is performing very well on the data that it can see and learn from. Since the problem does not lie in finding major patterns we can conclude that the spikes in the Validation loss is due to inconsistent data fed to the model.

In order to present a comparative understanding of the models performance, we have performed a comparative analysis based on the binary precision of the different models. The

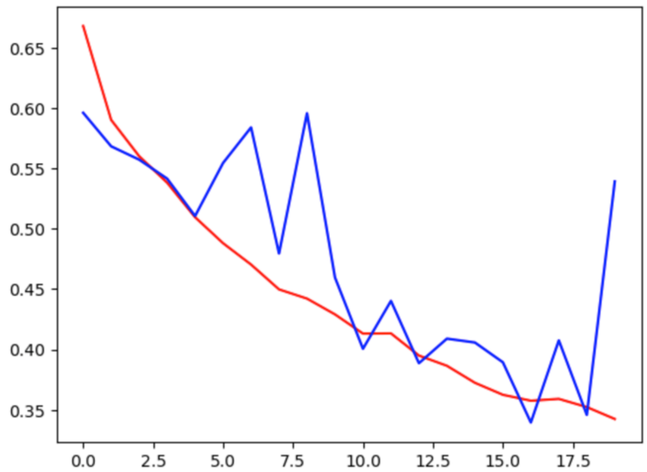

Fig. 10. Validation Loss and Training Loss

Models that are used for this are the MobileNetV2, ResNet50, No Data Augmentation, No Batch Normalization

TABLE I
COMPARITIVE ANALYSIS

Model	Accuracy
MobileNetV2	0.91
CNN with Augmentation	0.87
No Data Augmentation	0.81
No Batch Normalization	0.65
Resnet50	0.55

V. DISCUSSION AND LIMITATIONS

A. *Strengths and Limitations of Each Model*

MobileNetV2 [15] is trained on the ImageNet Dataset which has over 1 million images with depthwise separable Convolutionals which is more efficient than the general CNN models. MobileNetV2 replaces the CNN with a factorized version which splits the convolutional layers into two separable layers Depthwise and Pointwise convolutional layers. This results in a Network that is k^{21} or 8 to 9 times faster than the traditional CNN models. Since it splits the layers into more layers the computation time increase and the complexity of the model also increase which leads to longer training times and more complex calculations.

Resnet50 [16] is a Deep Learning model that uses an extensive 50 layers-deep CNN architecture that leads to improved accuracy. It uses the concept of residual blocks which helps in mitigating the vanishing gradient descent model. Pre-trained ResNet models have shown exceptional performance in generalizing when fine-tuned to suit a particular task. However, due to the large number of layers present in the model, the computational requirement also increases which results in a slower model. In some cases, it might lead to overfitting in the model which needs to be countered with careful regularization and implementing data augmentation.

CNN with augmentation is one of the most basic and most configurable types of Model present today. The model is completely open to customization to fit the need for a particular task. The Number of Convolutional layers are not specified and can be increased or decreased with regard to the performance of the model. Data augmentation is used before being fed to the model for evaluation. This prevents the model from overfitting and underfitting the data leading to improved accuracy. However, Due to its very customizable nature, a very deep understanding of the dataset is required in order to avoid overfitting or unnecessary computation. In some cases, it may require a very carefully implemented regularization technique.

CNN is a very robust mechanism to perform object detection from a given set of images and can be very powerfull when the right parameters are set. Since the number of Convolutional Neural Networks are not predefined the outcome is very heavily dependent on the layers present. However this can sometimes cause the model to overfit and lose its accuracy. without the help of methods such as Data augmentation the only way to reduce overfitting will be using regularization techniques which can cause a loss of Data.

B. *Implication for AI Art Detection*

Deep Learning holds major significance in the field of Arts as it can be used to make Art or identify Arts that are generated by such AI. The robust capabilities of these models in extracting intricate patterns and features from images have transformative potential in various aspects of identification of Art that is generated by AI

Deep learning models offer the promise of enhancing detection accuracy in Generated image analysis. By training on vast datasets and learning from a multitude of examples, these models can recognize subtle anomalies that might elude human observers. This heightened sensitivity could lead to identification of real art versus fake art.

The deployment of deep learning models allows for the automation of image analysis, reducing the burden on art curators and inspectors. This automation accelerates the process of allows curators and inspectors to focus more on getting high valued art for the right price to prevent fraudulent art pieces to be donated or sold.

Deep learning models provide a level of consistency and standardization in art detection. Unlike human observers who might be influenced by factors such as fatigue or familiarity with specific cases, these models apply the same analytical approach across all images, reducing inter-observer variability.

The insights extracted by deep learning models from art images can inform personalized detection pathways. The model is able to identify the nuances that a human might overlook such as the number of stokes in the painting or the similarity in art styles.

While the potential benefits of deep learning in medical image analysis are substantial, ethical considerations should be carefully addressed. Since the model is trained on various art styles there might be cases where the model might show a false positive or a false negative resulting in wrong classification of the art. The tool must be accompanied by a through check via humans to prevent wrong classification and update the model accordingly

VI. Conclusion and Future Scope

In this study, a comparative analysis of 3 prominent deep learning models regular CNN, ResNet and MobileNetV2 was conducted for the detection and classification of Images based on if they were AI-generated or Real. Our investigation revealed nuanced differences in the performance of these models, shedding light on their strengths and limitations. Convolutional Neural Networks (CNN) have exhibited amazing performance in AI art detection, with an impressive accuracy rate of 87.5. The ability of this technology to accurately identify and categorize art offers up fascinating possibilities for the art world and beyond. Artificial intelligence and art analysis together have wider applicability in many different fields and improve our knowledge of creative nuances.

The accomplishment of 87.5 accuracy denotes a substantial advance in automating the challenging process of art recognition. CNNs have shown to be a reliable option for this task thanks to their hierarchical feature extraction and pattern recognition skills. These networks have learned to discern the minute characteristics that set one artistic style apart from another by being trained on enormous archives of art images from various genres and eras. As a result, an automated tool that quickly recognizes and classifies artworks will be useful to art specialists, fans, and scholars, speeding up the study and categorization process.

Looking ahead, the potential and diversity of AI art identification utilizing CNN architecture are both encouraging. One way to advance is by making the precision even better. Even while 87.5 percent is an impressive accomplishment, the field is still developing, and researchers can work toward greater accuracy, possibly getting to the point where AI systems can match or even outperform human discernment. Additionally, the extension of the training dataset could result in more comprehensive art style detection, accepting lesser-known or underrepresented regional styles. The potential for AI-driven art history study is another fascinating direction. As CNN-based art identification algorithms develop, they may provide insightful information about the development of artistic styles and the factors that influenced them.

Additionally, the incorporation of AI art recognition into apps for virtual and augmented reality may completely alter how viewers experience art. Visitors to museums may have access to mobile devices that offer historical background, artist biographies, and stylistic analyses of the artwork they are witnessing. The effective implementation of CNN architecture in AI art detection with a 90 accuracy rate, in conclusion, marks a crucial turning point at the nexus of technology and art. For people all throughout the world, the future of this field holds the promise of ever-improving accuracy, improved art historical study, and life-changing artistic experiences. We set out on a journey that not only re-imagines how we see and understand art but also highlights the limitless potential of human-machine collaboration as we continue to harness the power of artificial intelligence.

VII. Acknowledgement

We would like to extend our sincere gratitude to our professor Dr. Misha Chandar for their guidance and support throughout the course of this research. Their expertise and encouragement has been crucial in the writing of this work.

We are also thankful to the Vellore Institute of Technology for providing us with a productive environment for this research. For which we are truly privileged to be a part of.

References

[1] Angus Forbes et al. Creative ai: From expressive mimicry to critical inquiry. *Artnodes*, (26):1–10, 2020.

[2] Eva Cetinic and James She. Understanding and creating art with ai: Review and outlook. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 18(2):1–22, 2022.

[3] Michael Mateas. Expressive ai: A hybrid art and science practice. *Leonardo*, 34(2):147–153, 2001.

[4] Mateja Čuljak, Bruno Mikuš, Karlo Jež, and Stjepan Hadjić. Classification of art paintings by genre. In *2011 Proceedings of the 34th International Convention MIPRO*, pages 1634–1639. IEEE, 2011.

[5] Maciej Wiatrak, Stefano V Albrecht, and Andrew Nystrom. Stabilizing generative adversarial networks: A survey. *arXiv preprint arXiv:1910.00927*, 2019.

[6] Ahmed Elgammal, Bingchen Liu, Mohamed Elhoseiny, and Marian Mazzone. Can: Creative adversarial networks, generating" art" by learning about styles and deviating from style norms. *arXiv preprint arXiv:1706.07068*, 2017.

[7] Vivek Kanji Malam. Digits & doodle identification using neural network.

[8] Ian J Goodfellow, Jonathon Shlens, and Christian Szegedy. Explaining and harnessing adversarial examples. *arXiv preprint arXiv:1412.6572*, 2014.

[9] Moustafa Alzantot, Supriyo Chakraborty, and Mani Srivastava. Sensegen: A deep learning architecture for synthetic sensor data generation. In *2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, pages 188–193. IEEE, 2017.

[10] Giovanna Castellano and Gennaro Vessio. Deep learning approaches to pattern extraction and recognition in paintings and drawings: An overview. *Neural Computing and Applications*, 33(19):12263–12282, 2021.

[11] Yoav Goldberg. *Neural network methods for natural language processing*. Springer Nature, 2022.

[12] Eva Cetinic and Sonja Grgic. Genre classification of paintings. In *2016 international symposium elmar*, pages 201–204. IEEE, 2016.

[13] Eva Cetinic, Tomislav Lipic, and Sonja Grgic. Fine-tuning convolutional neural networks for fine art classification. *Expert Systems with Applications*, 114:107–118, 2018.

[14] Ameet Deshpande and Vedant Somani. Weight initialization in neural language models. *arXiv preprint arXiv:1805.06503*, 2018.

[15] Qian Xiang, Xiaodan Wang, Rui Li, Guoling Zhang, Jie Lai, and Qingshuang Hu. Fruit image classification based on mobilenetv2 with transfer learning technique. In *Proceedings of the 3rd international conference on computer science and application engineering*, pages 1–7, 2019.

[16] Dhananjay Theckedath and RR Sedamkar. Detecting affect states using vgg16, resnet50 and se-resnet50 networks. *SN Computer Science*, 1:1–7, 2020.

[17] Gaozhong Tang, Bo Li, Hong-Ning Dai, and Xi Zheng. Sprnn: A spatial–temporal recurrent neural network for crowd flow prediction. *Information Sciences*, 614:19–34, 2022.

[18] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[19] Kaustav Mondal and HB Anita. Categorization of artwork images based on painters using cnn. In *Journal of Physics: Conference Series*, volume 1818, page 012223. IOP Publishing, 2021.

[20] C Richard Johnson, Ella Hendriks, Igor J Berezhnoy, Eugene Brevdo, Shannon M Hughes, Ingrid Daubechies, Jia Li, Eric Postma, and James Z Wang. Image processing for artist identification. *IEEE Signal Processing Magazine*, 25(4):37–48, 2008.

[21] Mark Sandler, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4510–4520, 2018.

[22] Kausthu b kannan. Ai-human art classification - mobilenetv2 - 91

[23] Yan Shen and Fang Yu. The influence of artificial intelligence on art design in the digital age. *Scientific Programming*, 2021:1–10, 2021.

[24] Hidenori Ide and Takio Kurita. Improvement of learning for cnn with relu activation by sparse regularization. In *2017 international joint conference on neural networks (IJCNN)*, pages 2684–2691. IEEE, 2017.

[25] Muye Chen et al. Vanishing gradient problem in training neural networks. 2022.

[26] Abien Fred Agarap. Deep learning using rectified linear units (relu). *arXiv preprint arXiv:1803.08375*, 2018.

[27] Prajit Ramachandran, Barret Zoph, and Quoc V Le. Searching for activation functions. *arXiv preprint arXiv:1710.05941*, 2017.

[28] Hao Peng, Jianxin Li, Yangqiu Song, and Yaopeng Liu. Incrementally learning the hierarchical softmax function for neural language models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31, 2017.

[29] Anil Menon, Kishan Mehrotra, Chilukuri K Mohan, and Sanjay Ranka. Characterization of a class of sigmoid functions with applications to neural networks. *Neural networks*, 9(5):819–835, 1996.

[30] Sagar Sharma, Simone Sharma, and Anidhya Athaiya. Activation functions in neural networks. *Towards Data Sci*, 6(12):310–316, 2017.