A PROJECT REPORT ON

“Synthetic Image Detection”

SUBMITTED TO

DEPARTMENT OF STATISTICS

SHIVAJI UNIVERSITY

KOLHAPUR

2023-24

FOR THE PARTIAL FULFILLMENT OF THE DEGREE

M.Sc. Statistics and Applied Statistics and Informatics

SUBMITTED BY,

Miss. Shinde Utkarsha Ananda

Miss. Shendage Akshada Jagannath

Under the guidance of

Mr. S. K. Ganjave

CERTIFICATE

This is to certify that the project report entitled **"Synthetic Image Detection"** being submitted by **Miss. Shinde Utkarsha Ananda**, **Miss. Shendage Akshada Jagannath** as a partial fulfillment for the award of degree of M.Sc. II (Sem-IV) in Statistics/Applied Statistics and Informatics of Shivaji University, Kolhapur, is a record of bonafide work carried out by them under my supervision and guidance. To the best my knowledge the matter presented in the project has not been submitted earlier.

Mr. S. K. Ganjave
Project Guide,
Department of Statistics,
Shivaji University,
Kolhapur.

Prof. Dr. S.B. Mahadik
Head of the Department,
Department of Statistics,
Shivaji University,
Kolhapur.

Date:

Place: Kolhapur

PREFACE

We're excited to share our project report on "Synthetic Image Detection" as a part of our Master of Science program at Shivaji University, Kolhapur. Despite facing time and resource constraints, we've put our best efforts into thoroughly studying the topic. Our project relies solely on existing data, and we've strived to present our findings clearly and concisely. We hope this report proves helpful to anyone interested in this subject.

ACKNOWLEDGEMENT

We're delighted to submit this project report on "Synthetic Image Detection" as part of our M.Sc. II program. We extend our sincere thanks to Prof. Dr. S. B. Mahadik, Head of the Statistics Department, for granting us permission to undertake this project.

We're immensely grateful to Mr. S. K. Ganjave for his invaluable guidance and timely support throughout the project. We also express our appreciation to the faculty members of the Statistics Department, including Dr. S. D. Pawar, Mr. S. V. Rajguru, and Dr. S. M. Patil.

Special thanks to the non-teaching staff for their assistance and cooperation. Lastly, we extend our gratitude to our friends for their encouragement and support throughout this endeavor.

ABSTRACT

This project aims to develop a deep learning model from scratch for detecting artificially generated images. In recent years, generative artificial intelligence (AI) has demonstrated considerable promise in generating images based on human descriptions. Techniques such as convolutional neural networks (CNNs) have been employed to train on extensive image datasets, while natural language processing (NLP) models have been utilized for generating images. However, this project emphasizes the creation of a deep learning model entirely from scratch.

The proliferation of synthetic images poses challenges for various applications, including image verification and content moderation. Addressing this, our project focuses on training a robust deep learning model capable of discerning between camera-captured and artificially generated images. By leveraging foundational principles of deep learning and avoiding reliance on pre-trained models, we aim to develop a novel solution tailored specifically to the task of synthetic image detection.

The project will involve extensive experimentation with different architectures, hyperparameters, and training strategies to optimize the model's performance. Additionally, data preprocessing techniques, including augmentation and normalization, will be employed to enhance the model's generalization ability and robustness to variations in input images.

The ultimate goal of this project is to contribute to the advancement of image classification and detection technologies, particularly in the context of identifying synthetic content. The developed deep learning model has the potential to find applications in various domains, including social media moderation, digital forensics, and content authenticity verification.

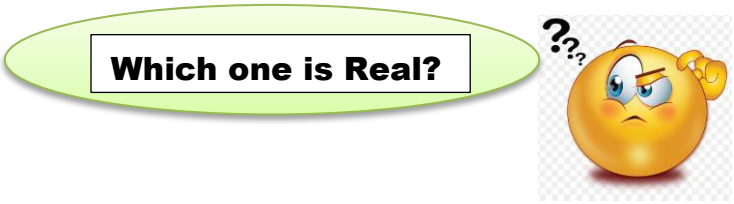
Through this endeavor, we seek to not only address the challenges posed by artificially generated images but also contribute to the broader field of deep learning research by exploring new methodologies for model development and training.

INDEX

Sr. No.	Title	Page No.
1	Introduction	7-8
2	Objectives	9
3	Data Collection / Data Description	10
4	Methodology	11-14
5	Implementation	15-16
6	Limitations And Future Scopes of the study	17
7	References	18

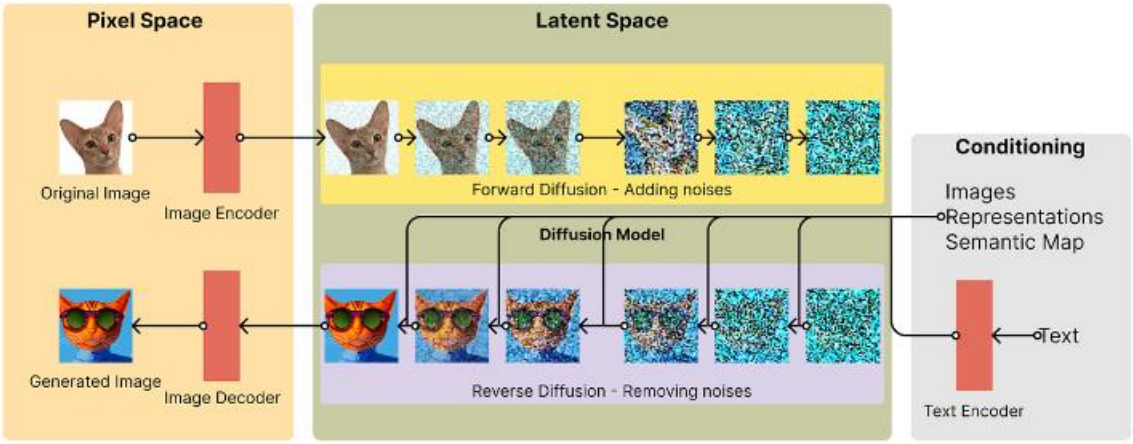
INTRODUCTION

In today's digital age, telling apart real images from computer-made ones is incredibly important. Our project focuses on creating a smart system that can do just that. By teaching computers to distinguish between pictures taken by cameras and those created by computers, we're helping to keep digital content honest and prevent the spread of false information. This is really important because fake images can mislead people and have serious consequences for society. By focusing on making sure images are genuine, we're making the internet a safer and more reliable place for everyone.



Among the images we're working with, some are genuine photographs captured by cameras, while others are artificially generated using a Diffusion model. These models employ complex algorithms to simulate real-world scenes, creating images that look remarkably similar to authentic photographs. By specifically focusing on images generated by the Diffusion model, we can study the characteristics and features unique to these synthetic images. This allows us to develop a system capable of accurately distinguishing between real and artificially generated images, enhancing our ability to combat misinformation and ensure the integrity of digital content.

Pipeline of generating images using diffusion model:

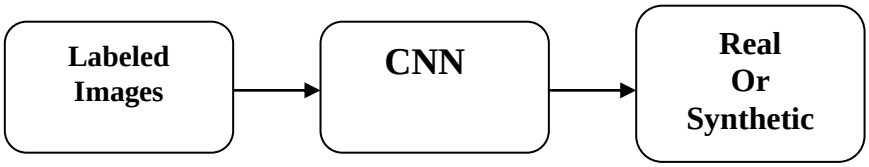


This is only one approach called as "Image-to-image synthesis". We can generate images using different approaches like sketch-to-image synthesis, Text-to-image synthesis, etc.

Project Overview:

Our project revolves around creating a clever system that can classify images into two categories: those captured by cameras and those generated by computers. We're employing advanced computer techniques to closely examine images and understand what distinguishes real ones from synthetic ones. Our aim is to develop a tool that assists people in verifying the authenticity of images, which could be handy for checking online posts and unraveling digital puzzles.

To achieve this goal, we have used Convolutional Neural Networks (CNNs), which extract meaningful features from the images.



Motivation:

The motivation behind our project stems from the growing prevalence of fake and manipulated images circulating online. These images can trick people, sway public opinion, and make it hard to trust what we see on the internet. By developing a strong system that can spot fake images, we're giving users the power to see through deception and make smarter choices. This helps lessen the impact of false information on society.

OBJECTIVES

- To find a key feature or characteristic that can reliably distinguish between authentic and synthetic images.
- To develop a model with the ability to accurately detect whether the given image is artificially generated or authentic.

DATA DESCRIPTION

The data we're using for our project is crucial, but it's not the main focus. Our dataset is basically images labeled with their respective categories as Real or synthetic.

Data format: The dataset consist of image files and each image file represents a specific real life objects. The dataset is split in half: one part contains 700 real images taken by cameras, while the other part contains 700 synthetic images which are generated by Generative models like Diffusion models, Generative Adverserial Network (GANs), VAE etc. Initially, we focused on using only synthetic images generated by Diffusion models. These models create images using complex algorithms, mimicking real-world scenes. And further included the images generated by, Generative Adverserial Network (GANs). From total dataset we used 1161 images for training and 239 images for testing.

Data source: For real images **Common Objects in COntext-stuff** (COCO-stuff) And for synthetic images we are using the subset of a dataset called "**ArtiFact: Real and Fake Image Dataset**"[link], which includes a variety of image files. These images are all sized at 200×200 pixels and cover a range of subjects, including people, animals, objects, food, and nature scenes. Eventually, we included digital arts also for our analysis.

METHODOLOGY

The trained model is based on architecture of Convolutional Neural Network (CNN). The utilization of CNN for visual feature extraction plays crucial role. By employing CNNs, we extract essential features from images, which are subsequently utilized to generate descriptive captions that accurately depict the visual content.

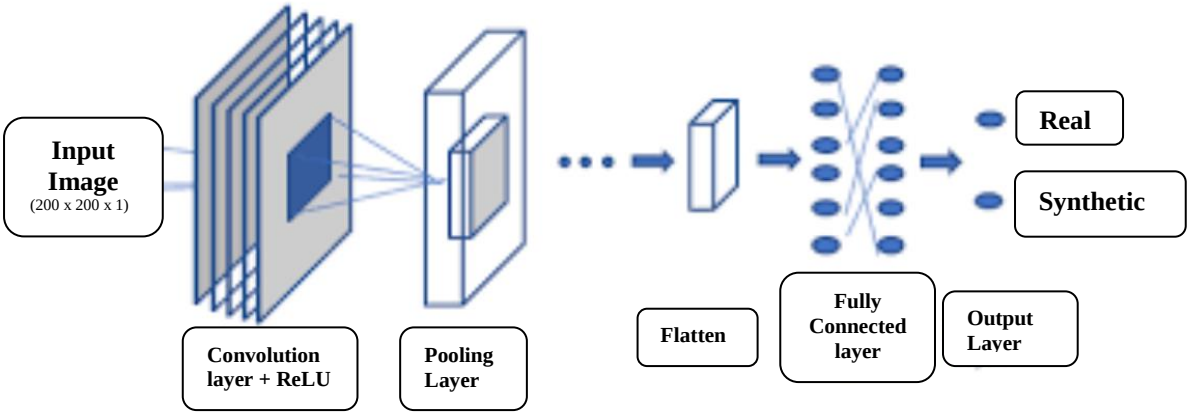


Fig. CNN Architecture

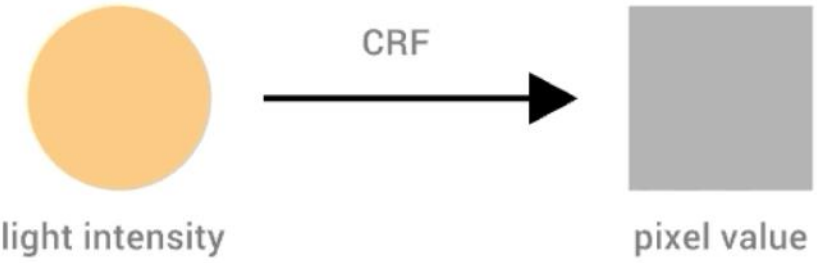
This project introduces a Convolutional Neural Network architecture model consisting of 16 layers, which include convolutional, pooling, and fully connected layers. The model starts with a convolutional layer, followed by another convolutional layer, a dropout layer, and a max pooling layer. This pattern repeats twice before connecting to a flatten layer, which feeds into a fully connected layer and finally, the output layer.

The aim of this study is to create a reliable model capable of distinguishing between real and artificially generated images. By combining different types of layers, the model learns to recognize important patterns in both types of images. This capability is crucial for tasks like image verification and content moderation. To achieve this, we experimented with different layer configurations and training techniques. We applied preprocessing methods like normalization and **Fast Fourier transformation** to improve the model's performance across various datasets.

By using convolutional neural networks, we aim to develop a practical solution for detecting synthetic images and ensuring the authenticity of digital content.

Fast Fourier Transform (FFT):

In general, a camera captures scene Radiance (light emitted by a source) and converts it to a Digital Image (array of pixels). However there is a mid-point in this process in which the digital camera converts irradiance (light arriving at it's sensor) to the Digital Image, this process is described by the Camera Response Function (CRF).



During the process of capturing images, light passing through the sensor can introduce errors known as "noise." These noise patterns vary depending on the specific camera sensor being used. According to "• On the detection of synthetic images generated by diffusion models", this noise patterns are also present in the synthetic images.

One method of analyzing these patterns is by using Fast Fourier Transform (FFT). FFT allows us to extract features from the images that can help distinguish between them. By applying FFT to the images, we can identify and quantify the unique noise patterns introduced by different camera sensors. These extracted features serve as valuable indicators for discriminating between images captured by different cameras, aiding in the development of our detection system.

Any periodic function can be rewritten as the weighted sum of infinite sinusoids of different frequencies. In image processing, Discrete Fourier Transformation (DFT) decomposes images into its sine and cosine components.

For a square image of size (N×N) the 2D DFT is,

$$F(u,v) = \sum_{x=0}^{N-1}\sum_{y=0}^{N-1} f(x,y)e^{-i2\pi\left(\frac{xu}{N}+\frac{vy}{N}\right)}$$

Where,

$$e^{-j\theta} = cos\theta - j\,sin\theta$$

Example:

$$f(x,y) = \begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix} \quad and \quad kerenel = \begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}$$

$$f(u,v) = [kerenel] * f(x,y) * [kerenel]^T$$

$$= \begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix} * \begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix} * \begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}$$

$$f(u,v) = \begin{bmatrix}4 & 0\\ 0 & 0\end{bmatrix}$$

In DFT, we transform the spatial domain of image in the frequency domain. The **FFT is an algorithm for computing the DFT** of a sequence of N complex numbers more efficiently than the standard DFT computation.

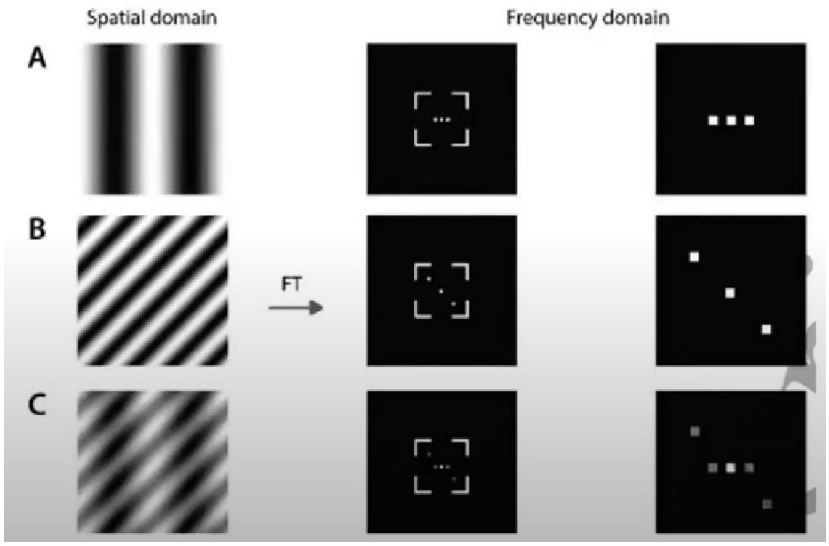
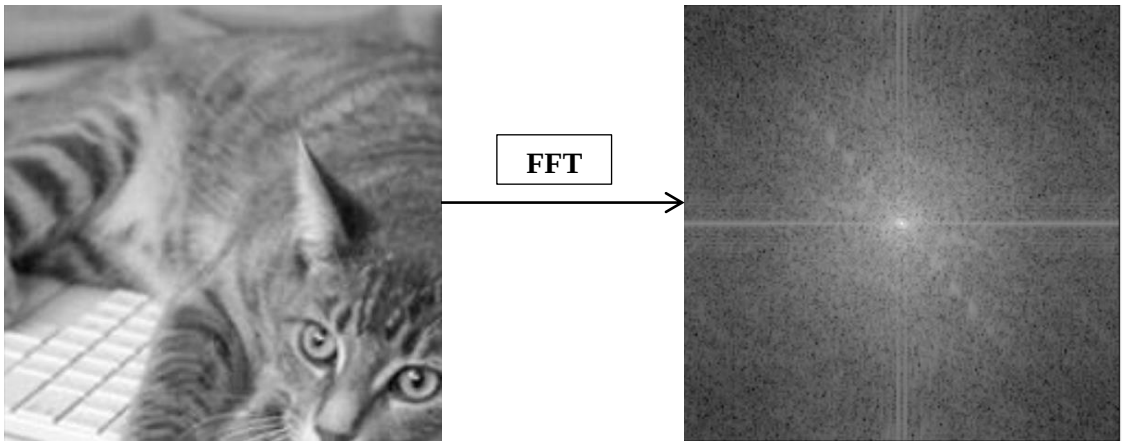


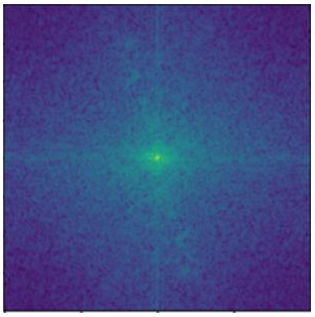
Fig. FFT transformation from Spatial to Frequency Domain

FFT on the actual image:

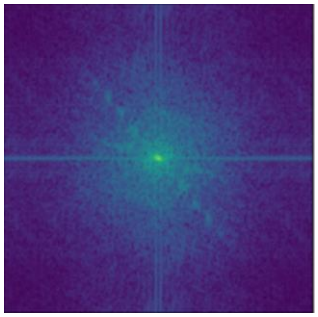


(This is shifted FFT in which the middle part has high intensity and corner points have low intensity.)

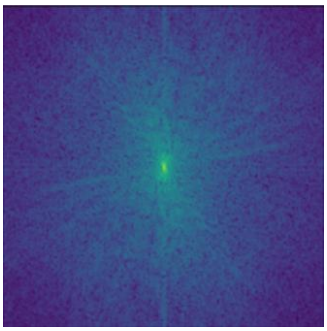
Some FFTs of Real and Synthetic images:



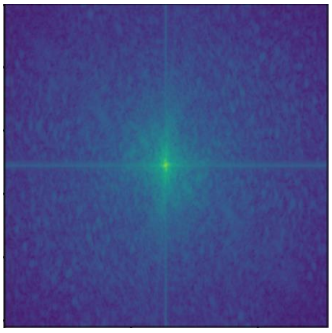
Real Image



Real image

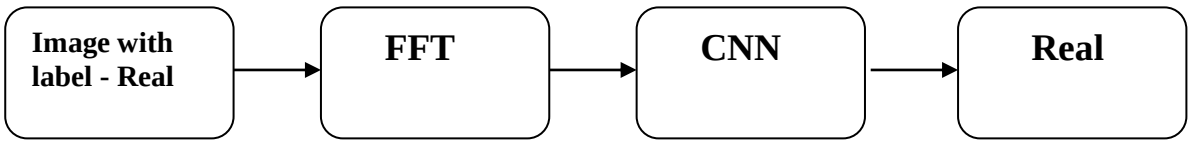


Synthetic image



Synthetic image

From above FFTs we can see that there is always a slight difference in the FFT of real images and Synthetic images and our aim is to capture this difference by using the deep learning model we have created. Then our model will be as follows,



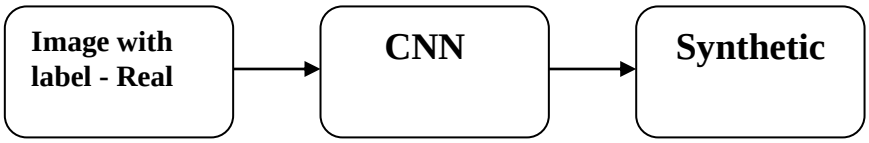
In essence, we converted input images from their original spatial domain to the frequency domain. In the frequency domain, images are represented by sine and cosine waves with varying frequencies. Two-dimensional Fast Fourier Transforms (FFTs) reveal variations in black and white patterns along specific directions in the image.

IMPLEMENTATION

We imported required libraries for image reading, text preprocessing, optimizers and to create CNNs. We split training data into model train dataset and model validate dataset in proportion as 85% train dataset and 15% validation dataset. We convert training images to 3D array and give them as input to Convolutional layers; these layers extract features from the input arrays using kernels. The convolutional layers and their corresponding kernels, pooling layers are added at multiple levels to extract meaningful features. The output is then converted to 1D array and given as input to fully connected layers of CNN

1. Using Raw Images:

Initially, we input the training images into the CNN and assess whether it could effectively discern the disparities between them.

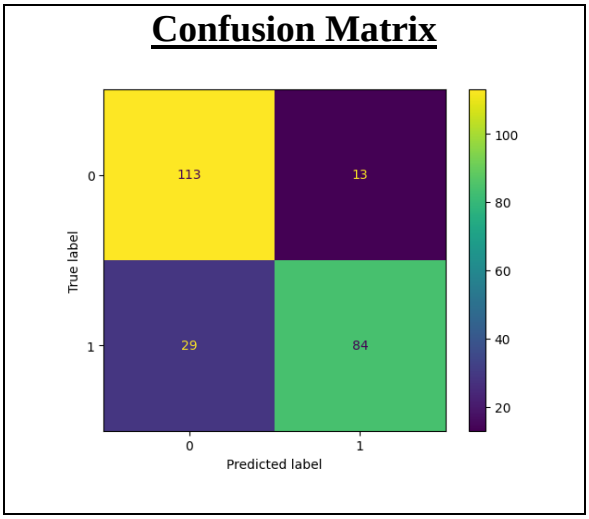


However, our findings revealed that the CNN struggled to differentiate between the images, yielding only a 48% accuracy rate. Upon analysis, we determined that the model's performance was hindered by the diverse nature of the images within the dataset. The wide-ranging subjects, encompassing humans, animals, objects, and various scenes, posed a challenge for the CNN to effectively learn and distinguish between real and synthetic images. Consequently, the model's accuracy remained stagnant and failed to improve over time.

2. Using Fast Fourier Transformation (FFT):

In our project, we explored the integration of Fast Fourier Transforms (FFTs) with Convolutional Neural Networks (CNNs) to enhance accuracy in image classification tasks. By passing FFTs through CNNs instead of raw images, we achieved a significant improvement in accuracy, reaching an impressive 82.84%. To accomplish this integration, we employed the "Adam" optimizer and measured loss using binary cross-entropy. Adam optimizer combines the strengths of AdaGrad and momentum optimizers, making it well-suited for our task. Additionally, binary cross-entropy is utilized for its effectiveness in binary classification tasks.

The integration of FFTs and CNNs represents a novel approach to image classification, offering improved accuracy compared to traditional methods. By leveraging the frequency domain information captured by FFTs, our model gains deeper insights into image features, leading to enhanced performance in distinguishing between different classes.



Here, the confusion matrix reveals that out of 239 test images, 197 were accurately classified, while 42 were misclassified. This indicates that **our model demonstrated an effective classification rate of 82%**.

Further, we experiment with our model using 20 images – 10 real and 10 synthetic. Initially, we selected synthetic images with noticeable differences from real ones and evaluated our model performance and it was 90%. Then, we swapped these synthetic images with ones that closely resembled real images, making it challenging to spot the difference. Despite similar FFT patterns in both types of synthetic images, our model yield 85% accuracy. This highlights that **when visual inspection becomes crucial, the FFTs also become crucial to classify**.

In conclusion, our study demonstrates the effectiveness of integrating FFTs with CNNs for image classification, showcasing the potential for improved accuracy and performance in real-world applications. Although FFTs are noise patterns of the images but they still affect the adequacy of the model according to visual difference of real and synthetic images.

Limitations of the study

Achieving the desired accuracy in our project hinges on proper hyperparameter tuning, which necessitates access to sufficient computational power. Our approach to analysis has centered on a single transformation method, yet it's important to note that there are numerous other techniques available for distinguishing between real and synthetic images.

Future scope of the study

Our primary focus throughout this project has been on constructing a model entirely from scratch. While this approach reflects our dedication to tailoring the model to our specific requirements, we recognize the potential for further accuracy improvements through the adoption of more advanced methods.

Exploring alternative transformation methods could offer valuable insights into enhancing the accuracy of our model. By considering a broader range of techniques, such as transfer learning or ensemble learning, we can potentially refine our model's performance and achieve greater accuracy in identifying synthetic images.

REFERENCES

- AI vs. AI: Can AI Detect AI-Generated Images.
- On the detection of synthetic images generated by diffusion models.
- Noiseprint: a CNN-based camera model fingerprint
- Do GAN leaves Fingerprints.

APPENDIX