Identification of Artificially Generated Images

Visual Recognition

1st Rahul Mahto
Dept. of Information Technology
IIT Allahabad
Prayagraj, India
iit2020022@iiita.ac.in

2nd Rohit Chowdhury
Dept. of Information Technology
IIT Allahabad
Prayagraj, India
iit2020043@iiita.ac.in

3rd Shashikant Thankur
Dept. of Information Technology
IIT Allahabad
Prayagraj, India
iit2020024@iiita.ac.in

4th Mohit Kumar
Dept. of Information Technology
IIT Allahabad
Prayagraj, India
iit2020220@iiita.ac.in

5th Shubham Kumar Bhokta
Dept. of Information Technology
IIT Allahabad
Prayagraj, India
iit2020007@iiita.ac.in

***Abstract*—In our proposed approach, we use Deep Convolutional Neural Networks (DNNs), in particular the ResNet structure, to distinguish between real and fake images. We concentrate on using unique patterns, and features found in the pixel level and structural properties of the generated images.**

***Keywords*—Visual Recognition, Convolutional Neural Networks, tensorflow, ResNet, Pixel, fake images.**

I. Introduction

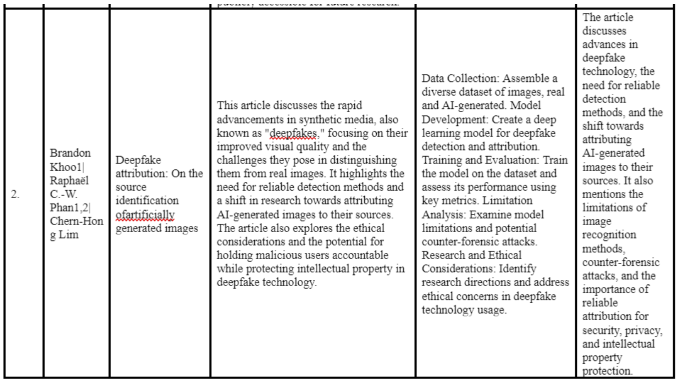
In the fields of image processing and computer vision, the spread of artificial intelligence technology has led to the creation of convincingly deceptive artificially created images. These synthetic images, often created using deep learning-based models, challenge the authenticity and reliability of visual content. In this paper, we present a comprehensive approach to artificially generated image recognition using a ResNet-based framework.

We have used the ResNet model. The ResNet model, known for its exceptional feature extraction capabilities, enables us to construct an effective classifier that is capable of detecting the signs of artificial image generation.

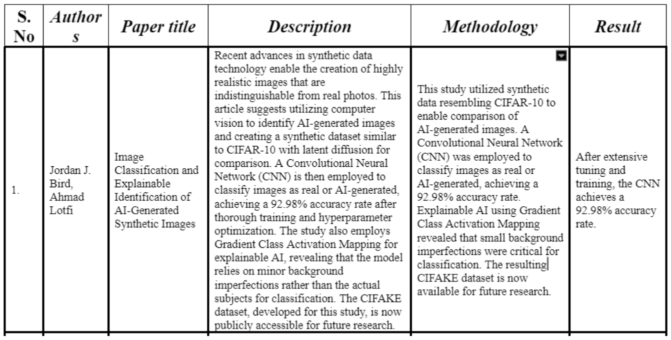
Our results demonstrate the model and its ability to accurately distinguish between genuine and manipulated images, contributing to the critical task of content authentication and ensuring the integrity of visual information in the digital age. The proposed method has a wide range of applications, from combating disinformation and image forgery to improving the credibility of digital media. As image processing techniques become increasingly sophisticated, our ResNet-based solution provides a reliable way to distinguish between real and artificially created. It is a valuable tool to ensure the authenticity and credibility of visual content in various fields.

II. Literature Review

A literature review for the detection of AI-generated images can provide a comprehensive overview of the current state of research, methodologies, and advancements in this field. Provide an overview of the development of AI-generated images, starting from early techniques like deep learning-based image generation.



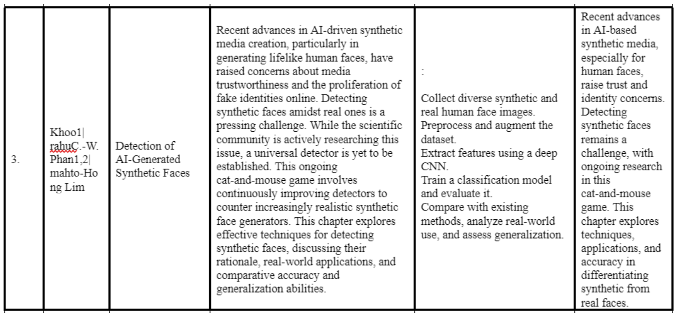
2.	Brandon Khoo1| Raphaël C.-W. Phan1,2| Chern-Hong Lim	Deepfake attribution: On the source identification ofartificially generated images	This article discusses the rapid advancements in synthetic media, also known as "deepfakes," focusing on their improved visual quality and the challenges they pose in distinguishing them from real images. It highlights the need for reliable detection methods and a shift in research towards attributing AI-generated images to their sources. The article also explores the ethical considerations and the potential for holding malicious users accountable while protecting intellectual property in deepfake technology.	Data Collection: Assemble a diverse dataset of images, real and AI-generated. Model Development: Create a deep learning model for deepfake detection and attribution. Training and Evaluation: Train the model on the dataset and assess its performance using key metrics. Limitation Analysis: Examine model limitations and potential counter-forensic attacks. Research and Ethical Considerations: Identify research directions and address ethical concerns in deepfake technology usage.	The article discusses advances in deepfake technology, the need for reliable detection methods, and the shift towards attributing AI-generated images to their sources. It also mentions the limitations of image recognition methods, counter-forensic attacks, and the importance of reliable attribution for security, privacy, and intellectual property protection.



S. No	Authors	Paper title	Description	Methodology	Result
1.	Jordan J. Bird, Ahmad Lotfi	Image Classification and Explainable Identification of AI-Generated Synthetic Images	Recent advances in synthetic data technology enable the creation of highly realistic images that are indistinguishable from real photos. This article suggests utilizing computer vision to identify AI-generated images and creating a synthetic dataset similar to CIFAR-10 with latent diffusion for comparison. A Convolutional Neural Network (CNN) is then employed to classify images as real or AI-generated, achieving a 92.98% accuracy rate after thorough training and hyperparameter optimization. The study also employs Gradient Class Activation Mapping for explainable AI, revealing that the model relies on minor background imperfections rather than the actual subjects for classification. The CIFAKE dataset, developed for this study, is now publicly accessible for future research.	This study utilized synthetic data resembling CIFAR-10 to enable comparison of AI-generated images. A Convolutional Neural Network (CNN) was employed to classify images as real or AI-generated, achieving a 92.98% accuracy rate. Explainable AI using Gradient Class Activation Mapping revealed that small background imperfections were critical for classification. The resulting CIFAKE dataset is now available for future research.	After extensive tuning and training, the CNN achieves a 92.98% accuracy rate.

III. Objective

The objective of identifying artificially generated images is preventing the misuse of synthetic content, safeguarding people's security and privacy, and upholding trust, authenticity, and credibility across a range of areas are the goals of detecting



3.	Khoo1| rahuC.-W. Phan1,2| mahto-Hong Lim	Detection of AI-Generated Synthetic Faces	Recent advances in AI-driven synthetic media creation, particularly in generating lifelike human faces, have raised concerns about media trustworthiness and the proliferation of fake identities online. Detecting synthetic faces amidst real ones is a pressing challenge. While the scientific community is actively researching this issue, a universal detector is yet to be established. This ongoing cat-and-mouse game involves continuously improving detectors to counter increasingly realistic synthetic face generators. This chapter explores effective techniques for detecting synthetic faces, discussing their rationale, real-world applications, and comparative accuracy and generalization abilities.	: Collect diverse synthetic and real human face images. Preprocess and augment the dataset. Extract features using a deep CNN. Train a classification model and evaluate it. Compare with existing methods, analyze real-world use, and assess generalization.	Recent advances in AI-based synthetic media, especially for human faces, raise trust and identity concerns. Detecting synthetic faces remains a challenge, with ongoing research in this cat-and-mouse game. This chapter explores techniques, applications, and accuracy in differentiating synthetic from real faces.

artificially made images. It is essential to maintaining the accuracy and consistency of information and digital media.

The following are the main goals of identifying photos that have been intentionally generated:

1. Authentication and Trustworthiness: The main objective is to confirm the legitimacy of photos in order to make sure that they haven't been maliciously edited or distorted. This is critical in domains where maintaining the integrity of visual evidence is critical, such as journalism, forensics, and legal procedures.

2. DeepFake Image Detection: Since deep learning and generative adversarial networks (GANs) have gained popularity, it has gotten easier and easier to produce convincingly fake images and movies. Recognising deepfake photos is crucial to stopping the dissemination of incorrect or misleading information.

3. Preventing Misinformation: Recognising artificially produced photographs can aid in the fight against the dissemination of false information, especially on social media and other internet forums. It guarantees that modified content won't trick users.

4. Privacy Protection: Individuals' security and privacy may be jeopardised by the unauthorised use of synthetic photographs, such as those used to create false profile pictures. Recognising these photos contributes to the security of personal data.

IV. Deep Convolutional Neural Network

A deep convolutional neural network (CNN) is an artificial neural network designed to process and analyze visual information, including images and videos. Deep convolutional networks (DCNNs) have revolutionized the computer vision process and have become the basis of many modern image processing and recognition systems. Some of the key characteristics of a deep CNN include:

1. Convolutional Layers: CNNs' fundamental building components are these. Convolution operations are used by convolutional layers to identify features and patterns in the input image. From straightforward edges and textures to more intricate structures like shapes and objects, these patterns can be found.

2. Pooling Layers: To decrease the spatial dimensions of the feature maps and increase the computational efficiency of the network, pooling layers are frequently employed after convolutional layers. Max-pooling and average-pooling are two common pooling operations.

3. Fully Connected Layers: CNNs usually contain one or more fully connected layers following a number of convolutional and pooling layers. These layers are in charge of predicting things by using the features that the earlier layers were able to extract.

4. Activation Functions: To add non-linearity to the model, fully connected layers and convolution are applied before non-linear activation functions like Rectified Linear Unit (ReLU). CNNs can now understand intricate relationships in the data as a result.

5. Depth: CNNs are "deep" because they have many layers, frequently with a large number of filters in each layer. They can learn hierarchical features with this depth, working their way up from basic edges to intricate object representations.

6. Weight Sharing: One of the core ideas of CNNs is weight sharing. Different regions of the input image are subjected to the same set of learnable filters in convolutional layers. The network is able to learn translation-invariant features thanks to this weight sharing.

7. Dropout: To avoid overfitting, dropout is a regularisation technique frequently employed in CNNs. Random neurons are "dropped out" of the training process by having their outputs set to zero. The network is forced to pick up more robust features as a result.

8. Batch Normalisation: Another method for accelerating and stabilising training is batch normalisation. By standardising the inputs to every layer in the network, it improves gradient flow and increases training stability.

V. Residual Neural Network

Residual neural network (resNet), also known as Residual neural network, is a deep learning (DNN) architecture that is designed to solve the vanishing gradient (VNG) problem in extremely deep networks. Residual neural networks (resNets) were first introduced in 2015 in the paper "deep Residual learning for image recognition" by Xiangyu Zhang (author of the book "Residual Neural Networks for Deep Learning") and Jian Sun. Since then, ResNets have had a significant impact on the fields of computer vision (CV) and deep learning (DL).

Key characteristics and components of ResNets include:

1. Residual Block: Consisting of two or more convolutional layers, the residual block is the fundamental building block of a ResNet. A shortcut connection adds the input directly to the output of one or more convolutional layers, omitting one or more of them. The residual connection is the name given to this connection.

2. Skip Connections: The vanishing gradient issue is avoided during training thanks to the skip or residual connections, which facilitate gradient flow. Deep networks can be trained more effectively with the aid of these connections.

3. Identity Mapping: In certain instances, no further convolutional layers are added; instead, the input is added directly to the output. When the dimensions of the input and output are the same, this is known as an identity mapping.

4. Bottleneck Architectures: In order to reduce the number of parameters and computational load and make deep ResNets computationally efficient, bottleneck architectures often use 1x1 convolutions to reduce the dimensionality before applying 3x3 convolutions.

5. Architecture Depth: With hundreds or even thousands of layers, ResNets can have extremely deep architectures. For a variety of computer vision tasks, deep networks are essential for capturing hierarchical features and producing state-of-the-art results.

6. Batch Normalisation: By normalising the inputs to each layer, batch normalisation is frequently used in ResNets to stabilise and speed up training.

ResNets have achieved remarkable results in computer vision tasks such as object detection, image segmentation, and image classification. They are often utilised as the core architecture in many deep learning models and have won numerous image recognition competitions. ResNet models that have already been trained are also accessible and can be adjusted for particular purposes, which makes them an important resource for deep learning.

VI. METHODOLOGY

Here are the steps for detecting ai generated images:

1. Data Collection: Both real images and images produced and altered by AI are included in our dataset. Fake images are pictures of faces that have been produced through a variety of techniques. https://zenodo.org/record/5528418#.YpdlS2hBzDd is the dataset's source. Images of real or fake human faces in the 256 x 256 JPG format can be found in this dataset.

2. Model Selection: To accurately classify images, our model makes use of ResNet50 and Convolutional Neural Networks (CNNs).

3. Data Splitting: Divide the dataset into test, validation, and training sets. This aids in model training, hyperparameter adjustment, and performance assessment.

4. Training: Using a categorical crossentropy as the loss function and accuracy as a metric, train your chosen model on the training set. Our model has been trained over fifteen epochs.

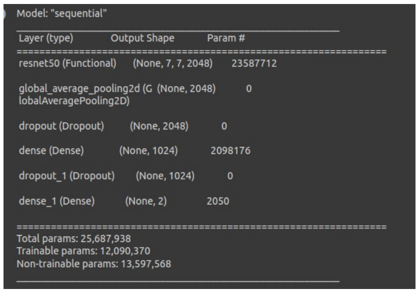
5. Evaluation Metrics: Accuracy, precision, recall, and F1 score are the metrics used to evaluate the performance of the model.

6. Model Fine-Tuning: To increase detection accuracy, make necessary adjustments to hyperparameters, model architecture, and feature extraction methods.

7. Testing: During testing, we discovered that our model's accuracy is 77%, recall is 75%, and f1 score is 74%.

VII. RESULTS

We have trained our model for 15 epoch, in which we got the following results while testing : The accuracy of our model is 77%, and recall is 75%, and f1 score is 74%.



```
Model: "sequential"
_________________________________________________________________
 Layer (type)            Output Shape           Param #
=================================================================
 resnet50 (Functional)      (None, 7, 7, 2048)      23587712

 global_average_pooling2d (G  (None, 2048)          0
 lobalAveragePooling2D)

 dropout (Dropout)          (None, 2048)            0

 dense (Dense)              (None, 1024)            2098176

 dropout_1 (Dropout)        (None, 1024)            0

 dense_1 (Dense)            (None, 2)               2050

=================================================================
Total params: 25,687,938
Trainable params: 12,090,370
Non-trainable params: 13,597,568
_________________________________________________________________
```

Fig. 1. Model Parameters

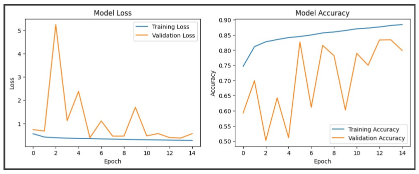
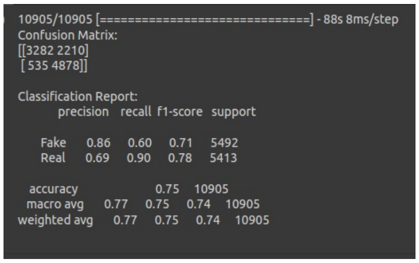


Fig. 2. Validation Loss and Accuracy

VIII. CONCLUSION

In conclusion, identifying artificially generated images is increasingly critical in the digital age due to advancements in AI and deep learning, making it easier to create realistic fake



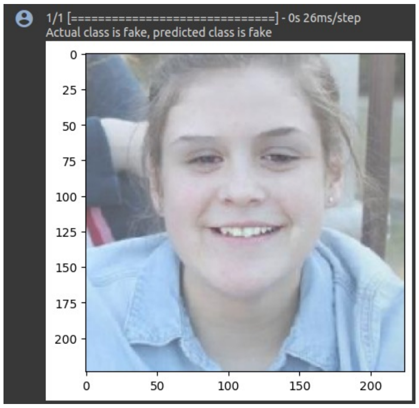
```
10905/10905 [==============================] - 88s 8ms/step
Confusion Matrix:
[[3282 2210]
 [ 535 4878]]

Classification Report:
          precision   recall  f1-score   support

    Fake      0.86      0.60      0.71      5492
    Real      0.69      0.90      0.78      5413

  accuracy                        0.75     10905
 macro avg      0.77      0.75      0.74     10905
weighted avg    0.77      0.75      0.74     10905
```

Fig. 3. Confusion matrix and results



```
1/1 [==============================] - 0s 26ms/step
Actual class is fake, predicted class is fake
```

Fig. 4. Testing on image

images with potential for misuse. Researchers and technologists have developed innovative techniques, from traditional forensics to cutting-edge deep learning, to detect such images. Collaboration among experts in computer vision, machine learning, and digital forensics is essential. Educational efforts and public awareness are also vital in recognizing and countering fake imagery. As technology evolves, our detection and prevention strategies must adapt. Through vigilance, collaboration, and digital literacy, we can protect the authenticity and trustworthiness of the images that shape our world.

IX. Acknowledgment

I am pleased to extend my heartfelt appreciation and gratitude to Prof Shiv Ram Dubey Sir, who served as our teacher and mentor throughout the duration of my project. His invaluable guidance, support, and feedback played a crucial role in shaping my understanding of the subject matter and helped me achieve my goals.

I would also like to express my sincere thanks to the Teaching Assistants, for their exceptional support and assistance during the course of my project. Their expertise and dedication were instrumental in helping me develop my skills and achieve my objectives.

References

[1] https://arxiv.org
[2] www.researchgate.net
[3] link.springer.com
[4] https://zenodo.org/records/
[5] CIFAKE: IMAGE CLASSIFICATION AND EXPLAINABLE IDENTIFICATION OF AI-GENERATED SYNTHETIC IMAGES
[6] Exposing computer generated images by using deep convolutional neural networks
[7] Distinguish computer generated and digital images: A CNN solution