# Image Tamper Detection Using Error Level Analysis and Convolutional Neural Networks

# Abstract:

Fake image detection has become an urgent task. Advanced technology benefits us but it also poses a threat to us when used in Cyber Crime. In reality, images are often considered as solid evidence to prove something concrete and thus image falsification or tampering in any way that benefits a party at fault, results in misleading. To detect this manipulation, it takes a substantial amount of image data and a model that can process every pixel and still provide results that are accurate and efficient enough to be used in real time. Therefore, a model that uses Convolutional Neural Networks in combination with Error Level Analysis has been developed that achieved an accuracy of 94\% and convergence with 40 epochs. Using this model, a website has been developed that provides users with the facility to check if an image is original or has been tampered with and gives them the option to store authentic images in a secure environment.

Keyword - Fake Image, Deep Learning, Convolutional Neural Network, Error Level Analysis

# Introduction:

People frequently accept what they can see, which impairs their judgement and causes a number of undesirable reactions. As forgeries have grown, there is a far greater need than ever to identify them. The main motivations behind image fabrication are evil ones. Information is distorted, immorality and fake news are spread, money is fraudulently obtained from an unknowing audience, the reputation of a well-known celebrity or other public figure is ruined, and there is a spread of unfavourable political influence among users of a digital platform. Since digital image editing tools have become more widely available and simpler to use, there has been an increase in the use of altered photos. As a result, more phoney or altered photographs are being posted online, which can be utilised for a range of nefarious activities like disseminating propaganda or conveying false information. The issue has been made worse by the development of deepfake technology, which makes it possible to produce difficult-to-detect fake images and movies that are incredibly convincing. As it diminishes confidence in digital media and can be used to disseminate false information, damage reputations, and meddle in political processes, this tendency has important socioeconomic consequences. For the aforementioned reasons, it is essential to provide techniques for determining if an image is real or altered.

A photograph typically implies the reality of what occurred. However, due to the abundance of digital images in our daily lives and the ease and simplicity with which they can be altered using readily accessible digital processing tools, seeing is sometimes no longer believed in the digital era. Images that have been altered frequently appear in reports, experiments, and even in court documents. As a result, we can no longer assume that photos are legitimate. Image tampering detection tools and methods are greatly desired. The majority of photos taken nowadays do not have digital watermarks, despite the fact that they can be used as a tool to add legitimacy to images. And it's likely that things will stay this way for the foreseeable future. A third party is also required to licence watermarks, and the usefulness and reliability of digital watermarks for image manipulation detection have not yet been proven. Passive image manipulation detection is hence more useful and crucial. By embedding watermarks in the original photographs, it tries to instantly confirm the authenticity of digital photos. Researchers have been concentrating more and more on picture tampering detection in recent years, particularly on passive techniques.

Numerous methods for detecting image manipulation have been proposed over time, including statistical, feature-based, and learning-based techniques. These techniques' sensitivity to particular tampering tactics and their applicability to novel tampering scenarios, however, are frequently their limitations. Additionally, the rapid development of image manipulation tools has made it more challenging to distinguish between altered and authentic photographs

# Program Statement:

Today, digital images have completely replaced the conventional photographs from every sphere of life but unfortunately, they seldom enjoy the credibility of their conventional counterparts, thanks to the rapid advancements in the field of digital image processing.

To identify a tampered image is a challenging task. Reliability of an image being shared which is tampered is lost. This project aims to develop an application that stores genuine images, protecting users against fake images and the impact that they may have on unsuspecting users.

# Scope:

The scope of the project is to detect tampered images and to encrypt and store only authentic images for a user account. The user would be given an option to upload an image. The uploaded image would be passed onto the pre-trained model which has been trained on CASIA 2 dataset and has an accuracy of upto 93\%. The model would return a confidence score stating whether the uploaded image is real or tampered. Tampering may include Image Retouching, Image Splicing, and Copy-Move Attack, Morphing. If the image is authentic, the image would get encrypted to protect user privacy and security and would be stored onto the backend. The user will be able to retrieve the image at their discretion.

# Objectives:

Users can input images that need to be tested.

The input image is resized, flattened and converted to an ela image.

This image is then tested against the CNN Model to detect any tampering.

As a result, the image will be classified as Authentic or Tampered.

Authentic Images are encrypted and stored in the cloud storage.

Tampered Images are rejected.

Users can input images that need to be tested.

The input image is resized, flattened and converted to an ela image.

This image is then tested against the CNN Model to detect any tampering.

As a result, the image will be classified as Authentic or Tampered.

Authentic Images are encrypted and stored in the cloud storage.

Tampered Images are rejected.

# Literature Survey:

We have conducted literature surveys for various techniques for detecting image tampering. We tried to find research papers which worked on a wide variety of image tampering methods such as Image Retouching, Image Splicing, and Copy-Move Attack Morphing. So far, we have collected 11 papers. We have decided to include papers for the purpose of the report which talk about detecting tampered images using the above mentioned techniques.

In paper [1] titled “Detecting fake images by identifying potential texture difference”, the authors create a GAN classification network that efficiently learns the exposed difference and realises the real and fake detection of face images. Faceforensics++ dataset is used as the verification dataset of the proposed method. Deepfake, and FaceSwap were used for manipulation to obtain Deep Fake images. The authors randomly took 14,000 images from 140 videos as the test set, including 2800 real images and 11,200 fake images and conducted

guided filter on them and tested the accuracy in the model. The inference drawn was that the subtle texture differences that exist between the real and fake face image and manifest them through the image saliency method. The advantages of the model was that there was high detection accuracy in both full image training and face image training. The enlarged texture features help the network to quickly and accurately capture the differences. The limitations of the model was that the network model training requires large-scale data to achieve good results. For the unknown tampering method, it is still necessary to train a new authentication network.

In paper [2] titled “Fake Image Detection in Twitter using Flood Fill Algorithm and Deep Neural Networks”, the authors utilise the flood fill algorithm to highlight the forged object in the image and a Deep Learning based solution is proposed to detect whether the image is fake or genuine. The dataset is collected from publicly available image verification corpus consisting of over 800 fictitious and genuine images. The framework was evaluated on the basis of robustness, accuracy, precision, recall and F1 score. The accuracy achieved was 96%.

The literature compared the proposed methodology with Naive-Bayes, SVM and Random Forest which show that using Deep Neural Networks is significantly better than existing models. The major advantage of the model was that no model exists to detect diffused images. The limitation of the model was that it took a lot of time and space for the model to detect efficiently.

In paper [3] titled “Intelligent Morphed Image Identification using Error Level Analysis and Deep learning”, the authors describe how Convolutional Neural Networks are a type of neural network that are used effectively in image recognition classification and applications. The authors use CASIA 2.0 which contains 7491 original images and 5123 tampered images. The size of the dataset is changed to 224x224 pixels.The proposed model achieved an accuracy of 91.33%. The presented model predicted 536 out of 574 original images and 321 out of 369 tampered images accurately. The main advantage of CNN compared to its predecessors is that it automatically detects the important features without any human supervision ELA works by re-saving the image at 95% compression, and evaluating the difference with the original. Modified areas are easily seen due their characteristic aspects in the ELA representation. The limitation of the model was that for each false image, we have a corresponding mask. We use this mask to sample false images along the edges of the sewing area to ensure that the false and non-false parts of the image contribute at least 25%.

In paper [4] titled “Image forgery detection using ELA and Deep Learning”, the authors divide the dataset into 2 categories, normalise the image, perform analysis on the level of compression error image, from the compression result use VGG 16 for CNN in recognizing the image according to the ELA. VGG 16 is used for training because VGG is perfect for training with minimal datasets. The authors use CASIA 2.0 which contains 7491 original images and 5123 tampered images. The size of the dataset is changed to 224x224 pixels. The training accuracy of the model is achieved up to 92.2% and for validation 88.46% using 100 epoch. VGG16 has gained recognition for precision, they used VGG16 as a pre-training model for original image and forgery images recognition. The limitations of using ELA technique is that image quality deteriorates after using ELA.

In paper [5] titled “Fake-image detection with Robust Hashing”, the authors investigate whether robust hashing has a possibility to robustly detect fake-images even when multiple manipulation techniques such as JPEG compression are applied to images for the first time. In an experiment, the proposed fake detection with robust hashing is demonstrated to

outperform state-of-the-art one under the use of various datasets including fake images generated with GANs.In the framework, robust hash value is computed from easy reference

image by using a robust hash method, and stored in a database. Similar to reference images, a robust hash value is computed from a query one by using the same hash method. The hash value of the query is compared with those stored in the database. Finally, the query image is judged whether it is real or fake in accordance with the distance between two hash values.The proposed method was confirmed to still maintain a high accuracy even under the additional manipulation.

In paper [6] titled “ A survey on image tampering and its detection in real-world photos”, the authors provide an overview on typical image tampering types, released image tampering datasets and recent tampering detection approaches. The authors have compared multiple image datasets like Columbia Gray dataset, CASIA dataset, MICC dataset, IMD dataset, COVERAGE dataset etc. Then, they have discovered the clues left by image tampering like edge anomaly. JPEG compression inconsistency, lighting inconsistency and inconsistencies of camera traces. Further, they discussed different detection methods using image processing and deep learning models. They obtained an accuracy of 94.8% using the Columbia Color dataset and an accuracy of 95.6% on the CASIA v2.0 dataset. A model using Markov features obtained an accuracy over 90% on the Columiba Gray dataset.

In paper [7] titled “ Detection of Fake Images via the Ensemble of deep representations from Multi-Color Spaces”, the authors state that current methods remain vulnerable when testing samples undergo post-processing attacks. The authors employ residual signals of chrominance components from multi-color spaces, including YCbCr, HSV and Lab, to learn robust deep representations via the well-designed shallow convolutional neural network (CNN). Then, the learned deep representations from different colour spaces are concatenated and then fed into the Random Forest (RF), which is the widely used en-semble classifier, to obtain final detection results. The authors use “Align & Cropped” PNG images in CelebFaces Attributes dataset to construct our dataset. All images in CelebA are preprocessed by cropping the 128×128 central region and then resized into 64 × 64. For real images, they randomly selected 10000 preprocessed “Align & Cropped” PNG images. Then, they employed PG-GAN to generate 10000 fake images. PG-GAN is able to generate realistic images by growing both the generator and discriminator progressively. Experimental results demonstrated that the proposed method outperforms state-of-the-art methods and has more robust detection accuracies (above 99\% in most cases) against post-processing attacks, especially for image blurring operations.

In paper [8] titled “Deteksi Pemalsuan Gambar dengan ELA dan Deep Learning”, the authors propose a new method that uses Convolutional Neural Networks and ELA to reduce computational costs and increase efficiency. The authors used a dataset of 1771 tampered images and 2940 real images that were converted into ELA images and resized to a size of 128x128 pixels. The ELA images were then normalised and split into a 80:20 ratio for training and for validation, respectively. An RMSProp optimizer was used for optimization. The first layer of the CNN consisted of 32 filters and a 5x5 kernel. The second layer of the CNN consisted of 32 filters, a 5x5 kernel and a Max Pooling layer of size 2x2. The two CNN layers used the glorot uniform kernel initializer and the RELU activation function. The next layer used was fully connected with 256 neurons and a RELU activation function. The proposed method reached a max accuracy of 91.83% with 9 epochs.

In paper [9] titled “No One Can Escape: A General Approach to Detect Tampered and Generated Image” the authors propose a method to detect general fake images and also GAN images. They convert RGB image to be detected into YCrCb colour space and extract the image edge information of the Cr component and Cb component. Then, convert image edge features into GLCM in order to do image scaling without losing the image tempering information. Finally, GLCM is fed into the designed deep neural network based on depthwise

separable convolution for training and detection. The edge feature extraction method and the deep neural network model designed can identify tampered images and GANs generated

images with a high macro average of F1 score of 0.9865.

\section{Datasets}

# Research Methodology:

The dataset with which the model has been trained is the CASIA 2.0 dataset.The Institute of Automation, Chinese Academy of Sciences, has compiled and annotated a sizable face picture dataset known as CASIA 2.0 (CASIA WebFace Database 2.0). It is among the largest freely accessible face recognition datasets with over 500,000 photos and over 10,000 subjects. The pictures in CASIA 2.0 are from the internet and come in a wide variety of poses, expressions, lighting, and occlusions. The dataset is commonly used in computer vision research and development, particularly for face recognition and face detection. Both real and phoney (manipulated) facial photos are included. The fake images are made by mixing or modifying existing images, for example by changing the facial expression or substituting one person's face for another, whereas the original images are actual shots of real people. The fictitious photos in CASIA 2.0 are intended to imitate actual situations in which face images may be purposefully altered to trick face recognition software. Researchers may test the robustness and generalizability of face recognition algorithms to these kinds of modifications using the dataset, which offers a difficult test environment. It is crucial to remember that the proportion of phoney photos in CASIA 2.0 may change and that detailed information about their distribution and variety is not made available to the public.

The investigation of compression artefacts in digital data that has lossy compression, like JPEG, is known as error level analysis (ELA). When used, lossy compression typically produces a uniform level of compression artefacts since it is typically applied uniformly to a set of data, such as an image. Alternatively, the data may be divided into segments with various degrees of compression artefacts. This variation may result from the different portions having undergone varying numbers of times of the same lossy compression or from the different parts having undergone various types of lossy compression. It may consequently be possible to tell if the data has been changed if there is a difference in the level of compression artefacts in various portions of the data. With JPEG, even a composite made up of components that underwent identical compressions would exhibit different compression artefacts. The data to be evaluated is subjected to an extra round of lossy compression, this time at a known, consistent level, and the result is subtracted from the original data under inquiry in order to make the generally subtle compression artefacts more clearly obvious. The difference image that results is then manually examined for any varying levels of compression artefacts. Additionally, metadata detailing the precise lossy compression utilised is occasionally included in digital data formats like JPEG. If the observed compression artefacts in such data are different from those predicted by the given metadata description, the metadata may not accurately reflect the compressed data and may therefore be a sign that the data have been altered. For our model, we have saved the image to 90% of its original quality, enhanced the brightness of the image.



Example of Real Image

Example of ELA of Real Image



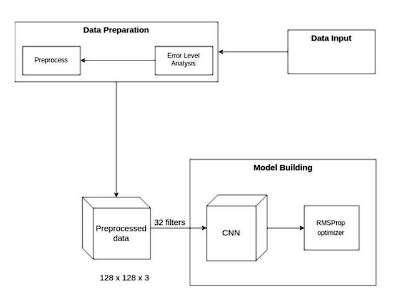
Example of Fake Image



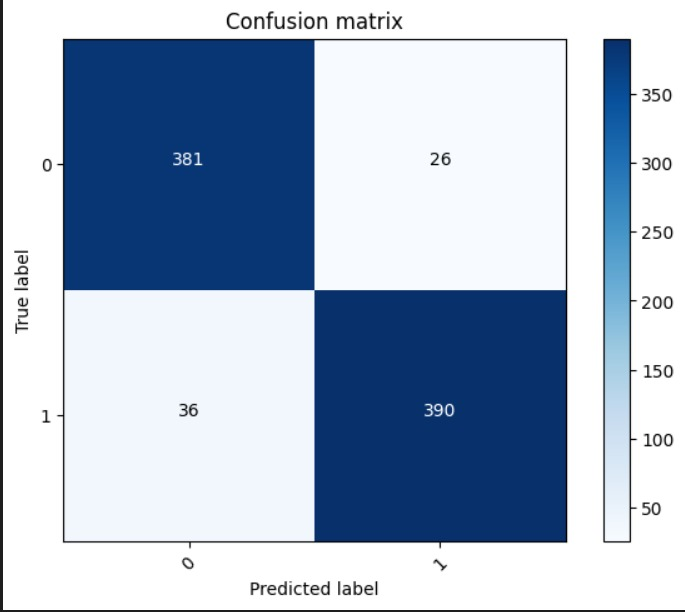
Example of ELA of Fake Image



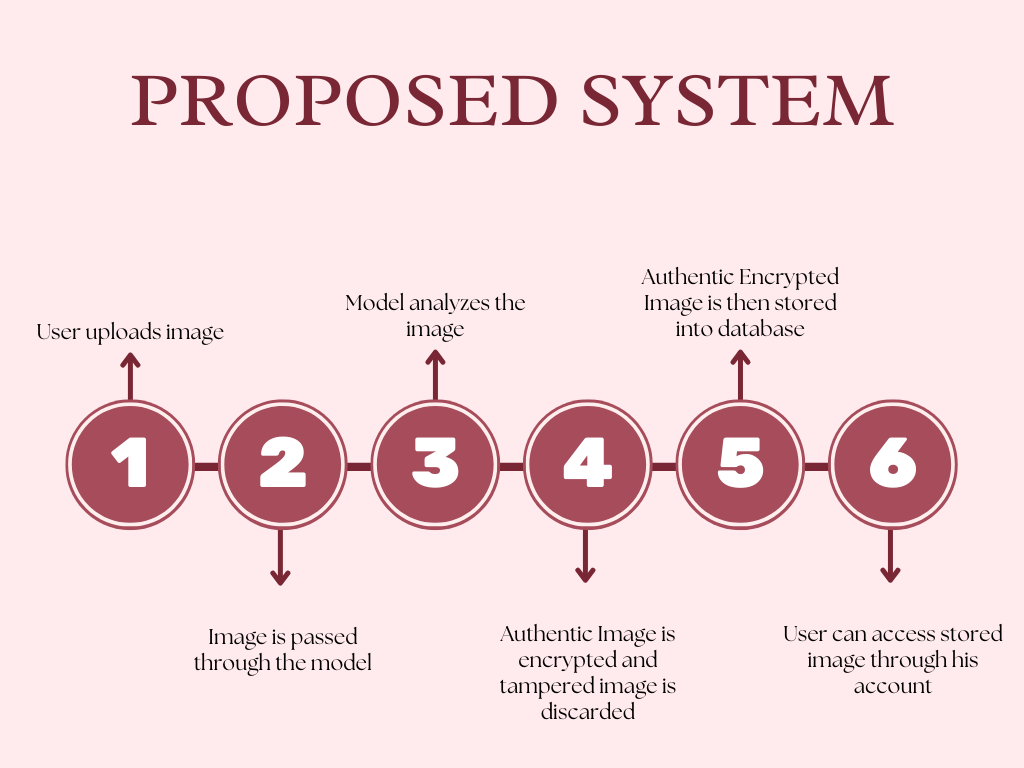
CNN is an example of a feedforward network, meaning that information only flows in one direction, from input to output. Despite the fact that there are several CNN architectures, most CNNs consist of a pooling layer, numerous convolutional layers, and other layers. One or more completely connected layers will come next. CNN receives an image as input for image classification so that each pixel can be processed. In essence, the convolutional layer serves as a feature extractor that uses the input image to the CNN to learn how to represent these features. The pooling layer's job is to lower the spatial resolution of feature maps in the meantime. Typically, there is a stack of numerous convolutional and pooling layers before the fully connected layer that are used to extract more abstract feature representations.



Confusion Matrix:



# Analysis and Design:



The above diagram represents our proposed system. The user would be given an option to upload an image which can either be authentic or tampered. The provided image would be passed through our pre-trained model which has been trained on CASIA 2 dataset and has an accuracy of upto 93%. The model will analyse the image and give a confidence score. The tampered image would be rejected and the authentic image would be encrypted and stored into the database. The user can then access stored images through the account which the user creates.

An image to be operated on is acquired from the user. The image undergoes some processes like flattening, brightening and splicing before it can be passed through the model.

ELA is then applied to the image. Image is once reverified and converted into a pixel array for the image to be passed through the neural network. Before all of this the model is already trained and retrained as and when new images are passed through it. After processing, if the image is detected as tampered/fake, a report is generated automatically. The user has to choose from deciding whether to print the report or not.

# 