FAKE IMAGE DETECTION USING MACHINE LEARNING

Vaibhav Kishore
Department of Computer Science and Engineering,
Integral University, Lucknow.

Mohammad Suaib
Assistant Professor Department of Computer Science and Engineering
Integral University, Lucknow.

Abstract

Image editing is now so widespread because to the availability of image processing tools like Adobe Photoshop or GIMP. Detecting such phone photos is unavoidable if image-based cybercrime is to be exposed. Because of its ubiquity, a photograph captured with a digital camera or smartphone is frequently saved in the JPEG format. The JPEG method works with 8x8 pixel picture grids that are compressed individually. Unmodified photos have a comparable amount of inaccuracy. Due to a comparable quantity of faults over the whole picture, each block should deteriorate at about the same pace during the resaving procedure. Error Level Analysis detects that the compression ratio of this false picture differs from that of the genuine image.

Our paper's goal is to create a picture forensics programme that can detect any type of photo modification. The vertical and horizontal histograms of the error level analysis image were then used to determine the site of the alteration. The suggested method was able to recognize the changed picture while also displaying the specific position of the adjustments, according to the results.

Keywords— *Fake Image detector, Photoshop Edited, Fake Image Processing, Detection, Meta Data Analysis*.

I. INTRODUCTION

The usage of technology in today's world has exploded, and one of the most prevalent forms of communication is the use of photographs. Images are now widely utilized in newspapers, magazines, websites, and ads, and they convey a wealth of information. Because of their widespread use, people's confidence in pictures is growing every day. Picture forging is the act of modifying or manipulating an image by changing some information inside it, and Image Forgery Detection is the process of determining whether the image is authentic or not.

In today's world, an enormous number of individuals have been victims of picture fraud. Many individuals modify photographs with image manipulation software and use them as evidence to deceive the court or numerous other people on social networking sites or applications. As a result, every image uploaded on social media should be assessed and classified as either authentic or fraudulent. Social media is one of the finest tools for socializing, sharing, and spreading knowledge, but it can also mislead individuals, causing mayhem due to unintended misleading propaganda.

This paper will then break down into three suggested methodology for evaluating the original ideas of an image, with the first section focusing on metadata analysis, the second on image error level analysis, and the third section focusing on developing a machine learning model to evaluate the image.

II. LITERATURE REVIEW

Today's forensic techniques for manipulating photographs necessitate the use of an expert to assess the

image's trustworthiness. This method may work for a limited number of photographs, but it is not suggested for a big number of images, such as those found on a social networking platform. As a result, we need to develop a system that can employ existing machine learning techniques to assess whether a picture is real or false, and then make it available to the general public for usage.

- S. Beram and colleagues devised a method for detecting doctoring in digital photographs [4]. Doctoring normally entails a number of processes, which are usually performed in the order of initial picture operations like scaling, rotation, brightness shift, smoothing, and so on. Binary similarity, picture quality, and wavelet statistics are among the statistical aspects that these approaches are dedicated to. The following are the three types of forensic facilities:

1. Image quality metrics: They look at the difference between both the doctored image and the original image. If the actual picture is not accessible, a hazy rendition of the image is used to imitate the test. [2,5]
2. Higher - level wavelet statistics: These statistics are produced from the image's multi-level decomposition.
3. Binary similarities measurements: These measurements capture the texture and correlation within both the Bit planes of lesser relevance, which are more vulnerable to manipulation.

First, single tools are designed to determine the essential image-processing functions in order to affect the identification of doctorate effects. Then, these individual "weak" detectors assembled together to determine the presence of a doctorate in an expert fusion scheme.

4. Enhance the meet the individual needs contrast picture: Contrast enhancement can be used to disguise the visual proof of image manipulation. Evidence of cut-and-paste forgeries may be discovered if these operations are tracked down. Cut-and-paste forgeries can be detected with the use of forensic tools.

5. Detecting histogram equalization in images: The Histogram Equalization Operation, like other contrast enhancement operations, creates spontaneous peaks and gaps in the picture histogram. The methods for detecting picture histogram equalization have been improved.

III. OBJECTIVE OF THE PROJECT

In essence, a metadata analyzer is a tag selection and search algorithm. If keywords like Photoshop, Gimp, Adobe, and other similar terms appear in the text, the likelihood of it being tampered with increases. Fakeness and realness are two distinct characteristics that are kept separate.

These qualities are already added to photographs by cameras and photo modification software if they are utilized, but they may be readily tampered with or modified, therefore they should only be used as a rough guide.

Fault Level Thinking resaves a specific image at a specific error rate, such as 96 percent, and then looks for a virtual change; if one is found, it signifies the cells have hit their global minimum for loss at that quality level.

Machine learning is comparable to data mining in terms of how it works. Both systems sift through data in order to find patterns. Machine learning, on the other hand, analyses data to find patterns in data and change programme operations appropriately, rather than harvesting data for human interpretation as in data mining applications.

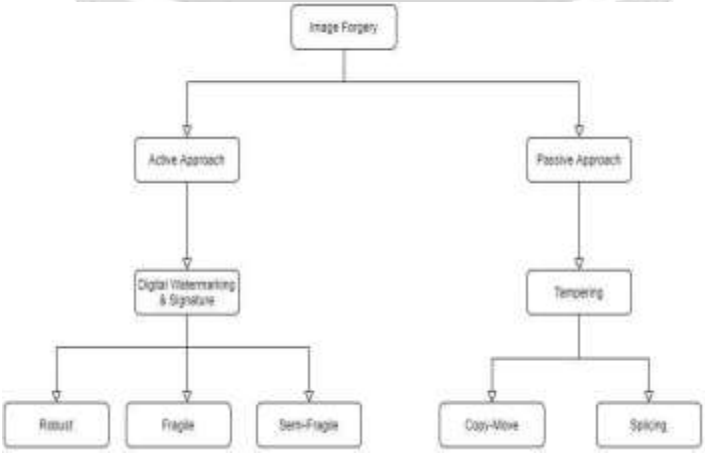

Fig-1: Types of image tampering techniques

IV. PROBLEM DEFINITION

The technology stores an image at 100 percent quality initially, then converts it to a 90 percent quality image. The distinction between the actual can be discovered using the difference approach. The output picture is the input image's needed accuracy level analysis (ELA) image. This image is now saved as a buffering image and delivered to the human brain to be processed further.

- 1) Make, Model, and Software
- 2) Image size
- 3) Timestamps
- 4) Types of metadata
- 5) Descriptions
- 6) Missing metadata
- 7) Altered Metadata

V. PROPOSED METHODOLOGY

1. Metadata Evaluation

In essence, a metadata analyzer is a tag selection and search algorithm. If keywords like Photoshop, Gimp, Adobe, and other similar terms appear in the text, the likelihood of it being tampered with increases.

2. Analysis of Error Levels

Error Level Analysis resaves a specific image at a specific error rate, such as 96 percent, and then looks for a virtual change; if one is found, it signifies the cells have hit their local minima for error at that quality level.

3. The Convolutional Neural Network (CNN) is a type of neural network that

A multilayer perceptron neural network with a few hidden layers on both the input and output levels. When an image is selected for review, it is first transformed from the Compression and Error Level Analysis stage to an ELA representation.

4. Transfer learning: improves the learner by transferring information from one domain to the necessary domain. It's a method of creating a model for one activity and then using it as a starting point for another.

5. Model VGG16 is a convolutional neural network design that focuses on having a stride 1 Convolution layer with the same padding and a stride 2 max pool layer.

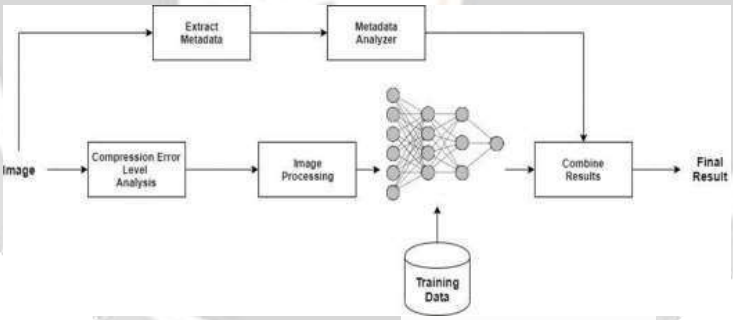

Fig-2: Proposed method architecture.

Spark is a fast and general-purpose cluster computing system for large-scale in-memory data processing. Spark has a similar programming model to Map Reduce but extends it with a data-sharing abstraction called Resilient Distributed Datasets or RDD. A Spark was designed to be fast for iterative algorithms, support for in-memory storage and efficient fault recovery. Spark Core consists of two APIs which are the unstructured and structured APIs. The unstructured API is RDDs, Accumulators, and Broadcast variables.

Processing: Large-scale datasets are frequently noisy, duplicated, and contain a variety of data kinds, posing significant hurdles to knowledge discovery and data modelling. In general, intrusion detection algorithms work with one or more forms of raw input data, such as the SVM algorithm, which exclusively works with numerical data. As a result, we prepare the data and transform the dataset's categorical data to numerical data.

VI. REFERENCE

[1] Luo, Weiqi, Jiwu Huang, and Guoping Qiu. "Robust detection of region-duplication forgery in digital image." Pattern Recognition, 2006. ICPR 2006. 18th International Conference on. Vol. 4. IEEE, 2006.

[2] S. Gholap and P. K. Bora, Illuminant colour based image forensics, in Proc. IEEE Region 10 Conf. 2008.

[3] Leida Li, Shushang Li, Hancheng Z -Journal of Information Hiding and Multimedia Signal Processing, Vol. 4, No. 1, pp. 46-56, January 2013.

[4] Tiago and Christian et al Exposing Digital Image Forgeries by Illumination Color Classification. IEEE Transactions on Information Forensics and Security (Page: 1182 1194)Year of Publication: 2013.

[5] Reshma P.D and Arunvinodh C IMAGE FORGERY DETECTION USING SVM CLASSIFIER Conference on Innovations in Information, Embedde and Communication Systems (ICIIECS), 2015.

[6] S.Shaid."TypesofImageForgery."Internet:http://csc.fsksm.utm.my/syed/research/image-forensics/11-types-of-mageforgery.html, Feb.08, 2010 12:17 [Dec. 4, 2012].

[7] Z. He, W. Sun, W. Lu, and H. Lu. "Digital image splicing detection based on approximate run length," Pattern Recogn .Lett., vol. 32, pp. 1591-1597, 2011.

[8] A picture's worth, Digital Image Analysis and Forensics, N Krawetz - 2007 Ph D, Hacker Factor Solutions.

[9]] http://imagej.net/Welcome ImageJ is an open source image processing program designed for scientific multidimensional images.J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[10]] http://forensics.idealtest.org/ CASIA v2.0 CASIA V2.0 is with larger size and with more realistic and challenged fake images by using post-processing of tampered regions. It contains 7491 authentic and 5123 tampered color images.

[11]] http://neuroph.sourceforge.net/ Neuroph Framework Neuroph is lightweight Java neural network framework to develop common neural network architectures. It contains well designed, open source Java library with small number of basic classes which correspond to basic NN concepts.

[12] https://github.com/drewnoakes/metadata-extractor Metadata-extractor is a straightforward Java library for reading metadata from image files.

[13]] https://www.github.com/afsalashyana/FakeImageDetection GitHub repositor for fake image detector desktop application written in javafx.

[14] Bellemare, M. G.; Danihelka, I.; Dabney, W.; Mohamed, S.; Lakshminarayanan, B.; Hoyer, S.; and Munos, R. 2017. The cramer distance as a solution to biased wasserstein gradients. arXiv preprint arXiv:1705.10743.

[15] Binkowski, M.; Sutherland, D. J.; Arbel, M.; and Gretton, A. 2018. ´ Demystifying MMD GANs. In ICLR.

[16] Chai, L.; Bau, D.; Lim, S.-N.; and Isola, P. 2020. What makes fake images detectable? Understanding properties that generalize. In ECCV, 103–120. Springer.

[17] Chandrasegaran, K.; Tran, N.-T.; and Cheung, N.-M. 2021. A Closer Look at Fourier Spectrum Discrepancies for CNNgenerated Images Detection. In CVPR, 7200–7209. Chen, T.; Zhai, X.; Ritter, M.; Lucic, M.; and Houlsby, N. 2019. Self-supervised gans via auxiliary rotation loss. In CVPR, 12154– 12163.

[18] Durall, R.; Keuper, M.; and Keuper, J. 2020. Watch your upconvolution: Cnn based generative deep neural networks are failing to reproduce spectral distributions. In CVPR, 7890–7899.

[19] Frank, J.; Eisenhofer, T.; Schonherr, L.; Fischer, A.; Kolossa, D.; ¨ and Holz, T. 2020. Leveraging frequency analysis for deep fake image recognition. In ICML, 3247–3258. PMLR.

[20] Haliassos, A.; Vougioukas, K.; Petridis, S.; and Pantic, M. 2021. Lips Don't Lie: A Generalisable and Robust Approach To Face Forgery Detection. In CVPR, 5039–5049.

[21] Huh, M.; Liu, A.; Owens, A.; and Efros, A. A. 2018. Fighting fake news: Image splice detection via learned self-consistency. In ECCV, 101–117.

[22] Jeon, H.; Bang, Y. O.; Kim, J.; and Woo, S. 2020. T-GD: Transferable GAN-generated Images Detection Framework. In ICML, 4746–4761. PMLR.

[23] Joslin, M.; and Hao, S. 2020. Attributing and Detecting Fake Images Generated by Known GANs. In 2020 IEEE Security and Privacy Workshops (SPW), 8–14. IEEE.

[24] Karras, T.; Aila, T.; Laine, S.; and Lehtinen, J. 2017. Progressive growing of gans for improved quality, stability, and variation. arXiv preprint arXiv:1710.10196.

[25] Karras, T.; Laine, S.; and Aila, T. 2019. A style-based generator architecture for generative adversarial networks. In CVPR, 4401– 4410.

[26] Karras, T.; Laine, S.; Aittala, M.; Hellsten, J.; Lehtinen, J.; and Aila, T. 2020. Analyzing and improving the image quality of stylegan. In CVPR, 8110–8119.
[27] Kendall, A.; Gal, Y.; and Cipolla, R. 2018. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In CVPR, 7482–7491.
[28] Khosla, P.; Teterwak, P.; Wang, C.; Sarna, A.; Tian, Y.; Isola, P.; Maschinot, A.; Liu, C.; and Krishnan, D. 2020. Supervised contrastive learning. arXiv preprint arXiv:2004.11362.
[29] Kim, C.; Ren, Y.; and Yang, Y. 2020. Decentralized Attribution of Generative Models. arXiv preprint arXiv:2010.13974.
[30] Lee, K. S.; Tran, N.-T.; and Cheung, N.-M. 2021. InfoMax-GAN: Improved adversarial image generation via information maximization and contrastive learning. In WACV, 3942–3952.
[31] Li, H.; Li, B.; Tan, S.; and Huang, J. 2020. Identification of deep network generated images using disparities in color components. Signal Processing, 174: 107616.
[32] Liu, H.; Li, X.; Zhou, W.; Chen, Y.; He, Y.; Xue, H.; Zhang, W.; and Yu, N. 2021. Spatial-phase shallow learning: rethinking face forgery detection in frequency domain. In CVPR, 772–781.
[33] Liu, Z.; Luo, P.; Wang, X.; and Tang, X. 2015. Deep learning face attributes in the wild. In ICCV, 3730–3738.
[34] Liu, Z.; Qi, X.; and Torr, P. H. 2020. Global texture enhancement for fake face detection in the wild. In CVPR, 8060–8069.