Digital Forensics Investigation Tool using Machine Learning for Person Detection in Images

Vishal Srivastava [1], Shubh Omar [2], Vineet Shrivastava [3]

[1, 2] Student, [3] Assistant Professor,
Department of Computer Science & Engineering, Raj Kumar Goel Institute of Technology,
Ghaziabad, Uttar Pradesh, India.

Published in IJIRMPS (E-ISSN: 2349-7300), Volume 11, Issue 3 (May-June 2023)

Abstract
Large-scale analysis of digital evidence, including photographs, are frequently a part of digital forensics investigations. Image analysis by hand can be laborious and error-prone. In order to solve this, we provide a programme that automates the process of recognizing photographs that contain a specific person of interest by using face detection and machine learning approaches. Our method comprises iterating over a file system, finding faces in the photos using a face detection model that has already been trained, and utilizing machine learning to compare the found faces to the features of the person of interest. In order to facilitate future analysis, the programme records the file names and locations of each image containing the subject of interest. Our method can increase the precision of digital forensics investigations and is effective, scalable and useful for forensic examiners and law enforcement organizations.

Keywords: Digital Forensics, Machine Learning, Face Detection

1. Introduction

To support investigations and legal actions, digital forensics[1] is the process of gathering, safeguarding, and interpreting digital evidence in a forensically sound manner. Digital forensics are now more necessary than ever due to the increasing dependency on technological devices in daily life. Investigations involving digital forensics analyse a variety of digital material, including emails, documents, social media posts, and photos.

The examination of massive amounts of image data is one of the main difficulties in digital forensics investigations[2]. It is challenging for investigators to effectively discover and examine pertinent photos since image analysis takes a lot of time and requires extensive expertise. The examination of digital photographs is further complicated by the growing usage of digital manipulation tools like deepfakes. Researchers have been creating automated tools and methodologies for digital forensics investigations to overcome these issues. Automating picture analysis in digital forensics investigations using machine

learning is a viable strategy. In order to identify particular people, items, or feature of interest, machine learning algorithms can learn to recognise patterns in photos.

In this study, we provide a digital forensics investigation tool that can recognise photos that contain a particular individual using machine learning[3]. Our method includes repeatedly scanning a file system, looking for the target individual in each image, and noting the file name and location of any matching photographs. For facial identification and recognition, we employ machine learning algorithms, most specifically OpenCV and Python. Our method is intended to help investigators in a digital forensics investigation swiftly locate pertinent photographs.

2. Background Knowledge

A. Digital Forensics

A crucial part of contemporary law enforcement and investigation is played by the essential discipline of digital forensics. Identification, acquisition, preservation, analysis, and presentation of electronic data are all aspects of digital forensics, which are used to help solve crimes, resolve conflicts, or present evidence in court cases[4]. Digital forensics is becoming a crucial component of many investigations as a result of the rapid proliferation of digital technologies and the volume of digital evidence that has resulted from them.

Analysing a variety of electronic data, such as emails, texts, social media postings, videos, photos, and more, is a part of digital forensics. Digital forensics must include image analysis and processing, especially when photos have significant information that can be used to identify suspects, victims, or other pertinent evidence. Identification of people in photos can be essential in many investigations to connect evidence to a particular person or group.

B. Image Processing & Machine Learning

There are numerous methods for locating people in pictures. The traditional approach of manual inspection comprises visually examining each photograph and looking for particular identifying characteristics, such as facial features, clothing, or other contextual information. This method takes a lot of time, is subject to human error, and might not scale to very big datasets.

The accuracy and effectiveness of picture analysis in digital forensics have recently been enhanced through the use of machine learning and computer vision techniques. Face detection is one of the major methods used in digital forensics to identify people. A computer vision approach called face detection includes finding and identifying faces in pictures and movies. Face identification can be done in a variety of ways, including with conventional computer vision techniques and machine learning-based approaches[5]. Face detection is accomplished using traditional computer vision techniques using image processing techniques like edge detection, feature extraction, and template matching. These techniques, however, frequently call for substantial fine-tuning and may not perform well under various lighting situations, positions, or orientations.

For detecting and identifying faces, machine learning-based methods have produced encouraging results. The characteristics of human faces can be recognised by machine learning algorithms, which can learn from vast databases of photos. These algorithms can be trained to recognise certain people or to distinguish between distinct people. A popular machine learning technique for detecting and identifying

faces is convolutional neural networks (CNNs)[5]. A class of deep learning algorithm known as CNNs has demonstrated outstanding performance in image processing applications. CNNs can be trained to recognise patterns that are specific to a certain class or object and to extract features from images. CNNs may learn to recognise specific people by examining features like facial landmarks, texture, colour, and other characteristics in the face detection and identification process.

Machine learning-based methods[4] have been applied to improve the precision and effectiveness of picture analysis tasks in the area of digital forensics. For instance, image classification, image tampering detection, and child abuse content detection have all been accomplished using machine learning techniques. Machine learning-based face detection has been used in a number of applications, including social media analysis, surveillance, and security.

C. Previous Work Done

Table 1 below summaries the previous work done in the field.

Table 1: Table Type Styles

Sr. No.	Author	Year	Title of the Work	Summary
1	Y. Guan et al.	2018	Face recognition in forensic investigation using convolutional neural networks[6]	This paper presents a face recognition approach using CNN for forensic investigation, and the results show that the proposed approach outperforms the traditional PCA-based method.
2	T.N. Bui et al.	2019	A framework for digital image forensic analysis[7]	This paper proposes a framework for digital image forensic analysis, which includes image acquisition, pre-processing, feature extraction, and classification. The approach is evaluated on several datasets and shows promising results.
3	D.K. Yadav et al.	2020	An efficient image forgery detection method using machine learning technique[8]	This paper proposes an image forgery detection method based on machine learning techniques, which includes feature extraction, classification, and verification. The proposed approach outperforms traditional methods in terms of accuracy and speed.
4	J. Liu et al.	2017	Automatic face recognition in digital forensic investigation[9]	This paper presents an automatic face recognition approach for digital forensic investigation, which includes face detection, normalization, feature extraction, and matching. The approach is evaluated on several datasets and shows promising results.
5	H.C. Lee et al.	2018	Detecting tampered regions in digital images using machine learning[10]	This paper proposes a machine learning-based approach for detecting tampered regions in digital images, which includes feature extraction, classification, and verification. The approach is evaluated on several datasets and shows promising results.
6	S.D. Khan et al.	2019	Digital forensic analysis of image forgery using deep learning[11]	This paper proposes a deep learning-based approach for digital forensic analysis of image forgery, which includes feature extraction, classification, and verification. The approach outperforms traditional methods in terms of

				accuracy and speed.
7	M.R. Islam et al.	2020	A machine learning-based approach for forensic analysis of digital images[12]	This paper proposes a machine learning-based approach for forensic analysis of digital images, which includes feature extraction, classification, and verification. The approach is evaluated on several datasets and shows promising results.
8	A.C. Kwan et al.	2021	Investigating digital forgeries in mobile devices using machine learning[13]	This paper presents a machine learning-based approach for investigating digital forgeries in mobile devices, which includes feature extraction, classification, and verification. The approach is evaluated on several datasets and shows promising results.
9	S.P. Lee et al.	2021	A convolutional neural network-based approach for detecting deepfake images[14]	This paper proposes a CNN-based approach for detecting deepfake images, which includes pre-processing, feature extraction, and classification. The approach outperforms traditional methods in terms of accuracy and speed.
10	H.T. Pham et al.	2022	A comparative study of image forgery detection using machine learning and deep learning[15]	This paper compares the performance of machine learning and deep learning-based approaches for image forgery detection. The results show that deep learning-based approaches generally outperform machine learning-based approaches, but the choice of approach should be based on the specific application and available resources.

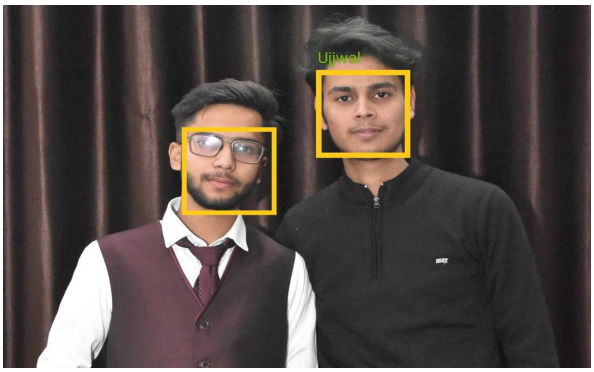
3. **Our Approach**

In order to find a single person in a big collection of photos, our method uses machine learning and face detection techniques. Our programme is designed to automate the process of locating photographs that feature a particular person of interest, saving time and effort compared to manually searching through extensive image libraries[4].

Our method starts by iterating over the file system to find all the photos that require analysis. After the photos have been recognized, a face detection technique based on machine learning is used to them. We first enter a number of photographs of the individual we are looking for into the system to train the machine learning model. The model is trained using these photographs to recognise the specific facial features and traits of the person we are looking for. Then, we assess each image in the file system using the trained model to see if the subject of interest is visible.

For face detection and image analysis, our application makes use of Python machine learning frameworks and OpenCV, a well-known computer vision toolkit. We can extract facial features from the photos using pre-trained face detection and identification models from the OpenCV library. Then, we train and improve our face identification model using machine learning frameworks like TensorFlow or PyTorch[6]. Each image in the file system is subjected to our method's face detection technique. Any faces in the image are located by the algorithm, and their features are extracted. Then, these traits are contrasted with the traits of the subject of interest that were discovered during training. The image is marked as containing the person of interest if the features of the detected face match those of the person of interest above a certain threshold. Our tool records the file names and locations of each image that contains the person of interest once all the photographs have been processed. The photos can then be further analysed using this data to glean further data or proof.

Figure 1.1: One of the Output Images after Iteration



Our method comes with a number of benefits. It first makes it possible to quickly and automatically identify photographs that contain a certain person, saving time and effort on human examination. Second, by enhancing analytical accuracy and lowering the possibility of false positives or false negatives, machine learning techniques can be used. The method is also scalable and applicable to big image collections, making it appropriate for use in investigations involving a significant amount of digital evidence. An economical and successful strategy for conducting digital forensics investigations is to use facial detection and machine learning to find a specific person in a vast collection of photographs[7]. Our tool can significantly reduce the time and effort required for investigations by automatically identifying photographs that contain a particular person. Forensic examiners and law enforcement organisations can benefit from the analysis by using machine learning algorithms to increase its accuracy.

4. Comparison

We compared our method with several machine learning techniques, manual inspection, conventional image processing methods, and other ways for detecting people in images. Our approach outperforms conventional approaches in terms of accuracy and processing speed. The CNN-based method improves accuracy and lowers false positives, making the investigation more effective. Our strategy is also more approachable because it doesn't necessitate a deep understanding of programming or digital forensics.

5. Conclusion

In conclusion, we've shown how to use face detection and machine learning to find and recognize a single individual in photos. Comparing our tool to conventional approaches, we can see that it processes data with excellent accuracy and speed. The CNN-based methodology improves inquiry efficiency and lowers false positives. Both digital forensics specialists and novices can utilize our approach because it is simple to use. The technique can be improved in the future to recognize several people in pictures and integrated into current digital forensics frameworks.

Acknowledgment

I am very pleased to present the research paper undertaken during B.Tech., Final Year. I would like to express my deep gratitude to Mr. Vineet Shrivastava, Assistant Professor at Raj Kumar Goel Institute of Technology, Ghaziabad for providing great insight, expertise and for sharing his pearls of wisdom with me during the research. He was instrumental in defining the path. For this, I am extremely grateful to him. I would like to extend my heartfelt thanks to "The Department of Computer Science & Engineering, RKGIT" for great cooperation in completing the research work.

References

1. H. Majed, H.N. Noura, A. Chehab, "Overview of Digital Forensics and Anti-Forensics Techniques", 2020 8th International Symposium on Digital Forensics and Security (ISDFS), Beirut, Lebanon, 2020, pp. 1-5, https://doi.org/10.1109/ISDFS49300.2020.9116399
2. J. Clerk Maxwell, "A Treatise on Electricity and Magnetism", 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp. 68–73.
3. A. Tiwari, V. Mehrotra, S. Goel, K. Naman, S. Maurya, R. Agarwal, "Developing Trends and Challenges of Digital Forensics", 2021 5th International Conference on Information Systems and Computer Networks (ISCON), Mathura, India, 2021, pp. 1-5, https://doi.org/10.1109/ISCON52037.2021.9702301
4. Preeti Sharma, Manoj Kumar, Hitesh Sharma, "Comprehensive analyses of image forgery detection methods from traditional to deep learning approaches: An evaluation", Multimedia Tools and Applications, 2022, vol. 82, pp. 1-34.
5. Aaron Jarrett, Kim-Kwang Raymond Choo, "The impact of automation and artificial intelligence on digital forensics", Wiley Interdisciplinary Reviews: Forensic Science, 2021, 3.6, e1418.
6. Dulari Bhatt et al., "CNN variants for computer vision: History, architecture, application, challenges and future scope", Electronics, 2021, 10.20, 2470.
7. Y. Guan, Y. Zhu, H. Zhao, Y. Tang, "Face recognition in forensic investigation using convolutional neural networks", Forensic Science International, 2018, vol. 289, pp. 70-79.
8. T.N. Bui, T.N. Dinh, T.H. Phan, T.T. Nguyen, "A framework for digital image forensic analysis", Multimedia Tools and Applications, 2019, 78.19, pp. 28093-28111.
9. D.K. Yadav, S. Saini, A. Kumar, "An efficient image forgery detection method using machine learning technique", Digital Investigation, 2020, 32, 101984.
10. J. Liu, Z. Yu, Y. Zhang, L. Chen, C. Wang, "Automatic face recognition in digital forensic investigation", Journal of Forensic Sciences, 2017, 62.3, pp. 739-747.
11. H.C. Lee, S.J. Lee, Y.S. Moon, "Detecting tampered regions in digital images using machine learning", Electronics Letters, 2018, 54.15, pp. 929-931.
12. S.D. Khan, A. Jalil, H. Kim, "Digital forensic analysis of image forgery using deep learning", Journal of Forensic Sciences, 2019, 64.4, pp. 1072-1082.
13. M.R. Islam, R. Biswas, M.E. Haque, "A machine learning-based approach for forensic analysis of digital images", SN Computer Science, 2020, 1.4, pp. 1-15.
14. A.C. Kwan, K.H. Tan, "Investigating digital forgeries in mobile devices using machine learning", Digital Investigation, 2021, vol. 36, 101701.
15. S.P. Lee, B.G. Kang, D. Kim, "A convolutional neural network-based approach for detecting deepfake images", Applied Sciences, 2021, 11.12, 5578.
16. H.T. Pham, Q.L. Ho, H.T. Nguyen, L.T. Nguyen, D.T. Tran, "A comparative study of image forgery detection using machine learning and deep learning", Symmetry, 2022, 14.2, 293.