Hybrid Model for Robust Digital Image Authenticity Detection

Research proposal

Submitted in partial fulfilment of the requirement for the degree of Bachelor of Science Honours in Information Technology

By:
M. G. S. K. Jayarathne
2019/ICT/108

Department of Physical Science
Faculty of Applied Science
University of Vavuniya
Sri Lanka

January, 2025

1 Introduction

1.1 Background

Today in the post-mobile era, digital images are far and wide-spread. Digital image can be defined as ”a representation of a two-dimensional image as a finite set of digital values, known as picture elements or pixels”[1]. The Explosion of digital images also led to the development of digital image modification methods[1][2]. Image modification can be defined as “altering or modifying an image to achieve a desired result” an idea that dates back to 1840. It reproduces images with different parameter values [2].

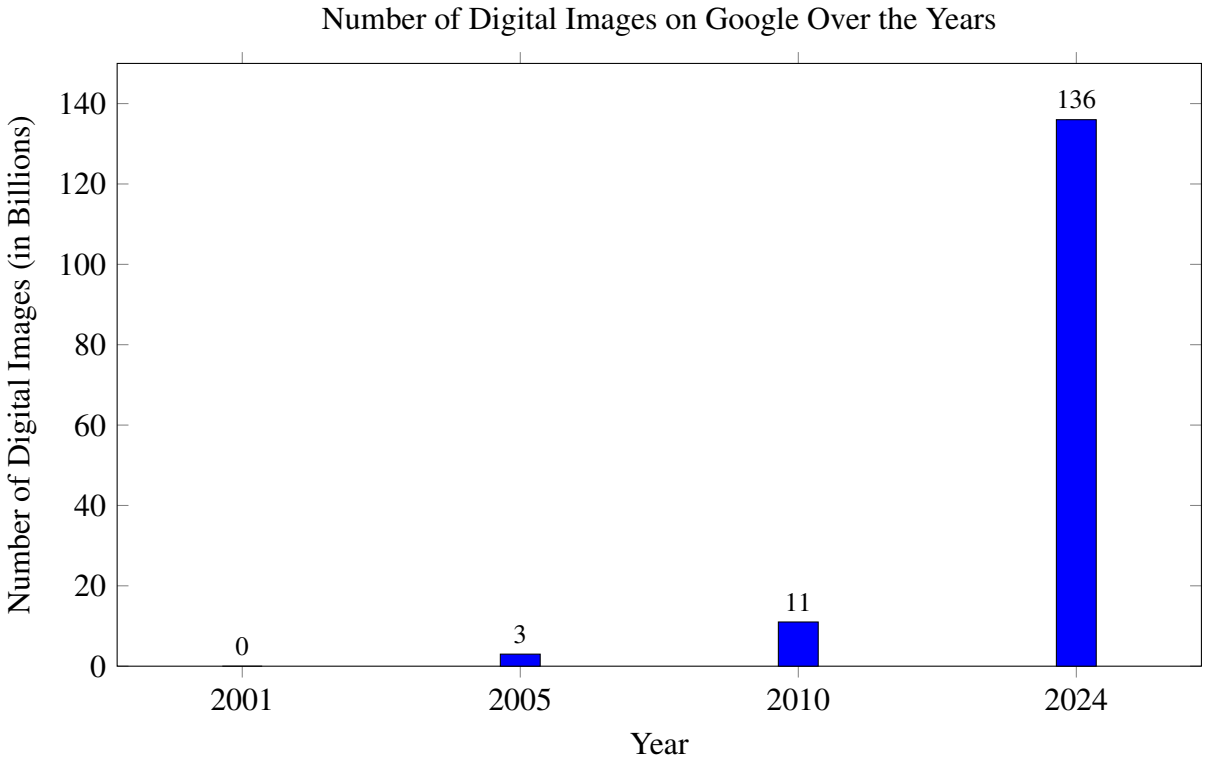


Figure 1:

Resource: ”How Many Photos are Taken Every Day?” by Matic Broz, December 9, 2024.

There are an estimated 136 billion images on Google Image Search in 2024. The number of images could reach 382 trillion by 2030. Once a digital image is uploaded to cyberspace, there’s no guarantee that it’ll be not modified[1]. Due to that reason, image forgery detection has become an increasingly popular research field[5]. When image manipulation is tangled with unethical intentions it would have a huge negative impact on the society.

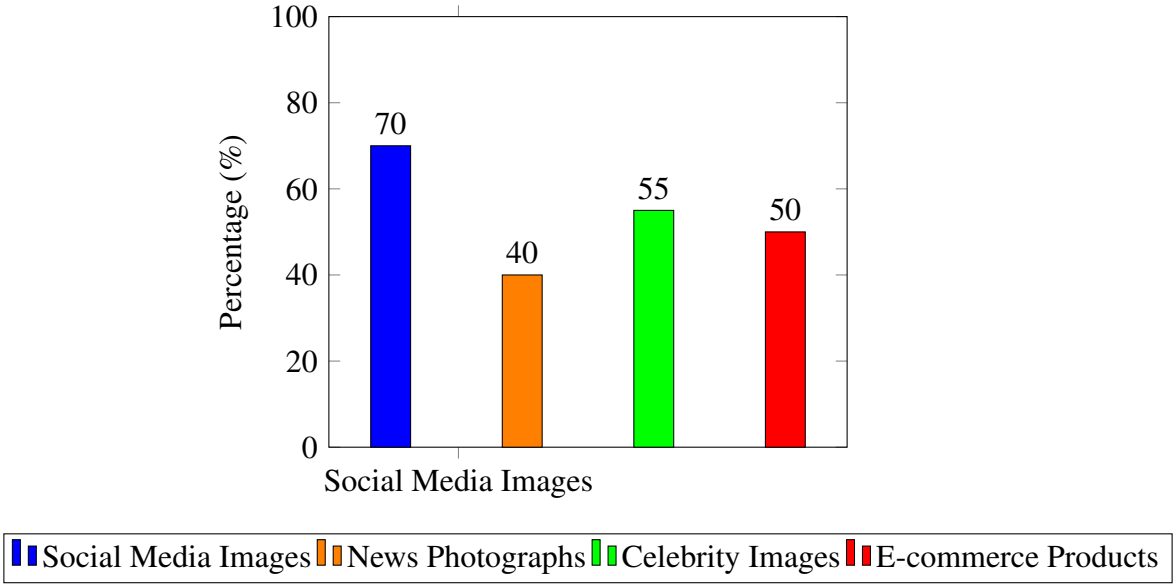


Figure 2: Distribution of Tampered Image Types (Digital Image Association (2024))

A well-known example is the Kerry Fonda 2004 election controversy [4]. Where in the election of 2004 John Kerry was libeled by a photo showing Jane Fonda giving a speech to support the cause. This enraged the general public and caused negative sentiments towards John Kerry. Even though it was later proven that it had been created by combining 2 images, the impact of the published photo had been irreversible.

Figure 3: High profile example of a image modification thar circulated online in 2004. The figure of Jane Fonda, captured in an unrelated 1972 photo, added into pre-existing John Kerry picture taken in 1970 (image taken from Reuters 2004).

Digital images are associated with a wide range of industries social media, forensic investigation, criminal investigation, intelligence systems, medical imaging, insurance claims, and journalism[5]. It's an essential and important content in the social media and the internet. There

are 2 main types of image authenticity detection. The first one is the active method and the passive method [5]. Active techniques also known as non-blind image forgery detection involve using attributes embedded in the image at the time of image acquisition, the main types are digital signature and digital watermarking. Passive techniques or blind image forgery detection rely on characteristics of the image itself to detect the modifications, the main types are pixel-based, compression-based, camera-based, physics-based, and geometric-based[5][6]. Due to the need to identify digital image tampering, the field of image authentication techniques is seeing rapid technological transformation[1]. Currently, existing tools for image authenticity detection restrict generalizability across broader use cases, an accuracy rate can differ according to the types of forgery detected, and it may have blind spots with newer generative techniques [19][7]. Integration of advanced techniques that use different image attributes can enhance the detection of image authenticity more effectively and accurately[10].

1.2 Problem Definition

Current limitations in image authenticity verification include:

1. Hybrid model for image authenticity detection combining both active passive techniques does not exist.
2. Difficulty in handling complex types of image tampering techniques.
3. Lack of generalizability, meaning they often perform well only under specific type of manipulations.

Some computational issues need to be resolved to surmount these limits. This means developing a robust mode for image feature extraction and analysis, ensuring that optimal attributes are used to assert image authenticity. A great deal of care and in-depth scrutiny of the literature so far was taken to ensure that the best techniques were selected to find the best result from the proposed system.(different search strings including specific keywords were used to filter the best publications existing, and research questions were asked in order to review selected papers according to a scoring system[12].)
Search strings used:

((”image authenticity” OR ”digital image verification”) AND (”forensic techniques” OR ”image analysis”) AND (”tampering detection”))

((”image forgery detection” OR ”digital image manipulation” OR ”image tampering”) AND (”statistical methods” OR ”feature fusion” OR ”image forensics” OR ”hybrid architecture”) AND (”image modification” OR ”tampering identification”))

((”image forgery detection” OR ”digital image manipulation” OR ”image tampering”) AND (”statistical methods” OR ”forensic analysis” OR ”image forensics” OR ”multi modal”)

AND ("image modification" OR "tampering identification"))

Research questions asked:

1. What are the existing techniques for image authenticity verification?
2. How effective are current single, multi-modal, and hybrid approaches in detecting image tampering?
3. What are their primary strengths and limitations?
4. Which image attributes (metadata, noise patterns, shadow inconsistencies) are most reliable for detecting digital manipulation?
5. How do different attributes correlate in determining image authenticity?
6. What is their individual and combined effectiveness in identifying tampering?
7. What are the current technological limitations in image authenticity verification?
8. How do emerging technologies like AI and machine learning enhance image authentication methods?
9. What innovative approaches can address existing detection challenges?
10. How do image authentication needs vary across digital forensics, social media, and medical imaging?
11. What unique challenges exist in different domain-specific tampering detection?
12. How can a multi-modal approach provide flexible, adaptable solutions?
13. What are the emerging trends in digital image authentication?
14. What potential technological developments could revolutionize multi-modal authentication?
15. How can research bridge current gaps in image tampering detection?

By asking these questions and after careful review, the research proposed the development of an adaptive feature fusion system that integrates multiple detection techniques. This unified decision-making mechanism will enhance the capability to detect modern image manipulation methods by leveraging various image attributes.

1.3 Research Objectives

The aim of this research is to improve trust in digital media by ensuring the integrity of visual content, thereby aiding users in making informed decisions based on tampered images or by comparing both original and tampered images. This research seeks to significantly enhance the accuracy and the efficiency of image authentication process.

1. **Develop a hybrid model:** To create an integrated hybrid model for detecting image authenticity by combining various techniques that focus on different attributes of an image. This will also fuse machine learning and deep learning methodologies to enhance the identification of whether an image is unique or has undergone digital tampering, ultimately optimizing the results in image authenticity verification.

2. **Enhance Image Authenticity Verification:** To improve image authenticity detection accuracy, efficiency, and reliability through a combination of various techniques.

3. **Probabilistic authenticity scoring mechanism:** Aims to implement a mechanism that quantifies how far an image has deviated from its original state and which regions have been tampered with.

1.4 Motivation

Due to the development in computational power, and the increase in mobile devices, machine learning has increased image manipulation for malicious purposes, such as misinformation, fraud, entertainment, and propaganda which disrupted the lives of the majority of the world population[5][6][7][8]. In safeguarding the visual content here digital images to be specific by using robust methods can be used to prevent digital manipulation. Creating a versatile system that can be applied to a broader number of domains can decrease the specific needs of each of them. Image authentication is a rich stream that trades between the four features which are robustness, security, versatility, and efficiency[9]. Currently existing tools used lack the power of combating the newly rising technologies like deep learning algorithms in safeguarding the originality of the image. By Combining techniques that are chosen by a careful literature review will contribute to protecting the integrity of digital images. Moreover, this approach would make sure the proposed system is much more accurate and efficient than the current existing ones in use. It would support ethical standings in legal proceedings and safeguard people would would fall prey to digital image forgery. It would help to empower the users in every corner of the world which would help to create a well-informed society restoring public trust in visual content.

2 Related Works

Ref	Year	Authors	Title	Methodology	Key Findings
1	2024	A. Nagm, M. Moussa, R. Shoitan, A. Ali, M. Mashhour, A. S. Salama	Detecting image manipulation with ELA-CNN integration: a powerful framework for authenticity verification[16]	CNN with error level analysis	copy-move and splicing forgeries
2	2023	Yasir Hamid, Sanaa Elyassami, Yonis Gulzar	An improvised CNN model for fake image detection[6]	CNN with error level analysis	Splicing forgeries
3	2022	Rachna Mehta, Karan Kumar, Adi Alhudhaif	Ensemble Deep Learning for Universal Forgery Detection[15]	Pixels analysis and edge information	Resampling forgery
4	2017	Bor-Chun Chen, Pallabi Ghosh, Vlad I. Morariu	Detection of Metadata Tampering through Discrepancy between Image Content and Metadata[3]	Metadata, Image content	simple feature fusion
5	2016	XiuLi Bi, Chi-Man Pun, Xiao-Chen Yuan	Multi-scale feature extraction and adaptive matching for image forgery detection[1]	JPEG compression and noise addition	copy-move tamper detection, geometric transforms

6	2014	Bo Liu, Chi-Man Pun, Xiao-Chen Yuan	Digital Image Forgery Detection Using JPEG Features and Local Noise Discrepancies[11]	JPEG compression and noise analysis	valid for highly compressed images
7	2010	Girija Chetty, Monica Singh	Image Tamper Detection Based on Multimodal Fusion[5]	pixel features and correlation analysis features fusion	copy-move tamper detection only

Figure 4: Overview of Image Forgery Detection Studies (2010-2024)

3 Research Gap

In image authentication, there exists a significant research gap that underscores the need for an integrated approach to effectively combat digital forgeries. Most of the existing systems tend to focus on singular detection methods, lacking a unified multi-model strategy that could enhance detection accuracy across various types of manipulations. This limitation highlights the necessity for an interdisciplinary approach that combines techniques from different domains to create a more robust image authentication framework. Additionally, current methodologies often suffer from high computational complexity, which can hinder their practical application in real-time scenarios. Therefore, there is a pressing need to explore novel attribute combination strategies that have not been thoroughly investigated, particularly adaptive feature fusion techniques that can dynamically select and prioritize the most relevant attributes based on the characteristics of the image being analyzed. By addressing these gaps, this research aims to improve trust in digital media and safeguard users against misinformation and fraud.

4 Materials and Methods

4.1 Materials/Tools

- Data source - COLUMBIA, CASIA-v1, IMD
- Programming Environment- Python 3.9+
- Python imaging library
- Flask or Django for web application development

4.2 Procedures

4.2.1 Data Preparation

Dataset Collection The benchmark dataset used consists of three main components:

1. Original image
2. Manipulated image
3. Binary mask

It is crucial to maintain correct file-to-file associations in the dataset.

Split into Training and Testing Dataset The dataset is divided into training and testing subsets to evaluate the model's performance.

Data Augmentation Data augmentation techniques, such as rotation, resizing, flipping, splitting, and slight color changes, are applied to create variations in the dataset. Overlaps of modified images are avoided to ensure proper training.

4.2.2 Feature Engineering

To detect image manipulation, pre-computed feature maps are employed to assist in accurately identifying tampered regions:

1. **Metadata:** Metadata anomalies are embedded as a separate feature channel or as an auxiliary input to the segmentation network. This helps in the classification of manipulated versus non-manipulated images[17][13].
2. **Noise Analysis:** Generated noise level maps can analyze and represent images on multiple scales and are used as additional input channels along with RGB images and enhance the model's ability to detect inconsistencies[9][21].

3. **Error Level Analysis (ELA):** ELA images are generated to highlight compression artifacts, which are then fed as an additional input channel[14].

4. **Pixel Analysis:** Generated pixel intensity maps can highlight the variations within an image, enabling the identification of tampered or manipulated areas. By focusing on pixel-level differences and irregularities[21].

5. **Histogram analysis** Employed to identify the unusual patterns in pixel intensity distribution. [8]

6. **Edge and Border Detection:** Highlights suspicious boundary transitions, and pixel distributions providing insights about potential manipulations by identifying unusual patterns. [18].

The dataset now includes both the original images and feature-engineered datasets. Reports for binary classification (manipulated or not manipulated) are generated using various models.

4.2.3 Model Selection

A strong model is required for effectively detecting manipulated pixels in images. The proposed methodology embodies multiple paradigms for enhancing detection accuracy and reliability. These are some of the major approaches that will be applied:

Patch-Based Classification-Classical ML Approach
Traditional approach where it segments the image into small patches[20] and then extracts features above mentioned. Each patch will be classified as manipulated or not.

Deep Learning Models
Other architectures that will be used DeepLab[4], and Mask R-CNN[7]. These kinds of architectures have state-of-the-art performances for image segmentation. So, they are capable of identifying modified regions with high accuracy while ensuring superiority in performance over different datasets.

Hybrid Model Development
The integration of above five approaches will result in the development of a hybrid model that embodies strengths from both machine learning and deep learning techniques. This hybrid model will utilize the outputs of all methods through a unified decision-making mechanism that synthesizes results for better detection accuracy.
The hybrid model has come to complement the deficiencies in different methods by incorporating the strengths of various single methods. The machine-learning-based methods, depending as they do on predefined features, could feel quite limited in cases that are complex image manipulations. This integrated model now targets specific manipulation types with increased power and provides a far more robust solution in determining the authenticity of an image. Notable points:

Feature Fusion Use the outputs (feature maps or predictions) from all five methods as inputs to a hybrid architecture.

Design the Hybrid Model Combine feature maps from each methodology in a unified network, such as a multi-stream network where each stream corresponds to one method.

4.2.4 Model Training

The input consists of images (with extra feature channels) fed into the selected segmentation network. The output consists of predicted masks with the same spatial dimensions, performing pixel-level classification. The predicted mask is compared with the ground truth mask using loss functions like Binary Cross-Entropy (BCE) and Dice Loss.

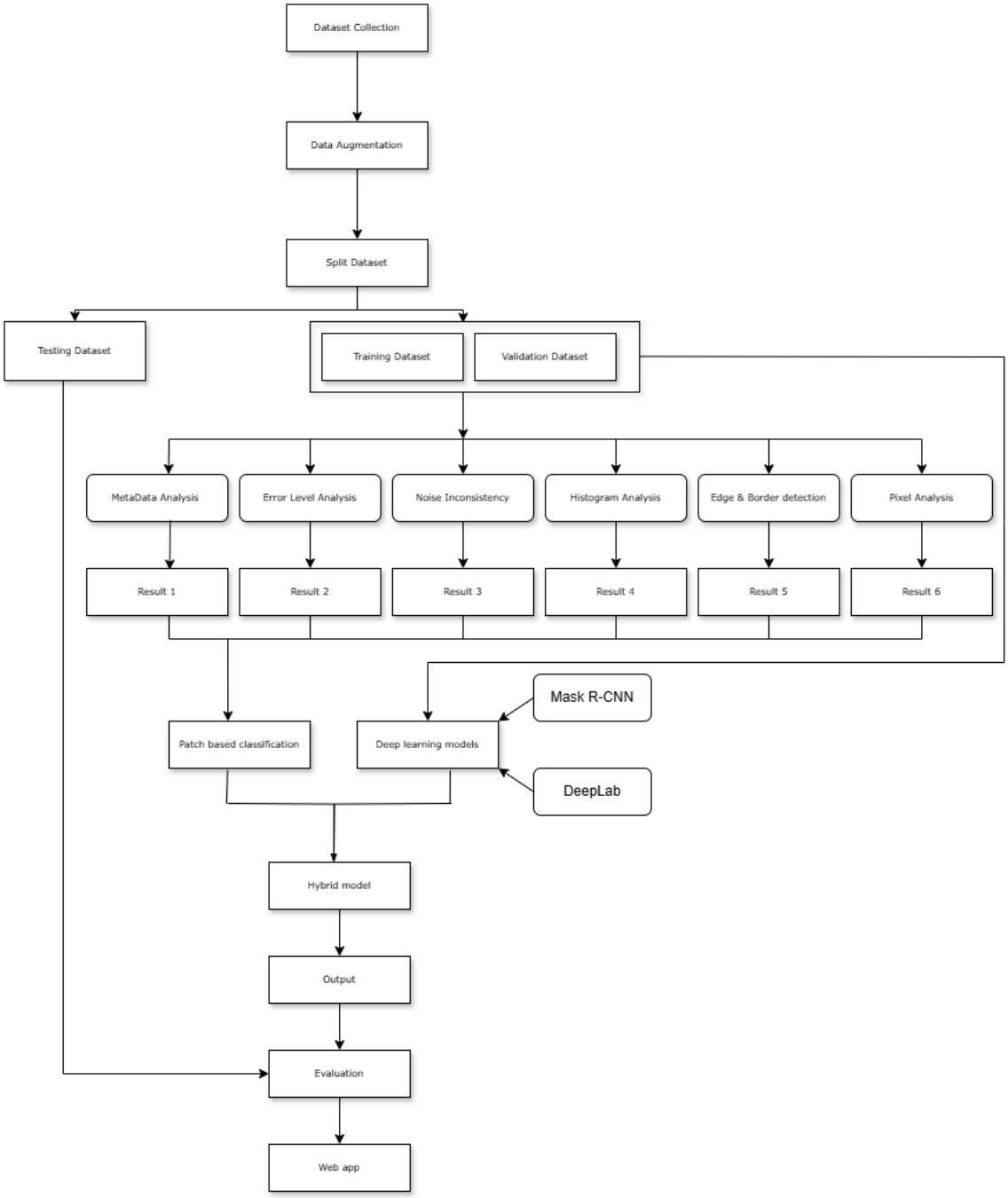
4.2.5 Model Evaluation

The following metrics are used to evaluate the model:

- **Intersection-over-Union (IoU):** Measures the overlap between predicted and ground truth masks.
- **Dice Coefficient (F1 Score):** Evaluates the similarity between predicted and ground truth masks.

4.3 Data Analysis

- Comparative evaluation of detection techniques
  Above mentioned traditional models would be fed with the dataset and same procedure would be done with hybrid model and accuracy and results would be analyzed and compared for better insights.
- Visualization of results
  A web application would be developed to visualize the final result of the hybrid architecture.



5 Expected Outcome

1. Development of a versatile hybrid image authenticity detection model that integrates various techniques.

2. Creating a probabilistic scoring system that quantifies the extent of modifications, providing users with clear insights into the regions which have been altered.

3. Significant enhancement in the accuracy of detecting different types of digital image manipulations, addressing current limitations in existing systems.

4. Increased trust in digital media by ensuring the integrity of digital images, empowering users to make informed decisions based on authenticity.

6 Gantt Chart

Work/Time (Months)	Nov-24	Jan-2	Jan-25	Feb-25	Mar-25	Apr-25	May-25	Jun-25
Title selection								
Proposal								
Literature Review								
Resource gathering								
Tools and Techniques								
Implementation								
Writing								
Publication / Conferences								
Presentation and Submission								

References

[1] X. Bi, C. M. Pun, and X. C. Yuan. "Multi-scale feature extraction and adaptive matching for copy-move forgery detection". In: *Multimedia Tools and Applications* 77 (Jan. 2018), pp. 363–385. DOI: 10.1007/s11042-016-4276-3. URL: https://doi.org/10.1007/s11042-016-4276-3.

[2] Gajanan K. Birajdar and Vijay H. Mankar. “Digital image forgery detection using passive techniques: A survey”. In: *Digital Investigation* 10.3 (2013), pp. 226–245. ISSN: 1742-2876. DOI: https://doi.org/10.1016/j.diin.2013.04.007. URL: https://www.sciencedirect.com/science/article/pii/S1742287613000364.

[3] Bor-Chun Chen et al. “Detection of Metadata Tampering Through Discrepancy Between Image Content and Metadata Using Multi-task Deep Learning”. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. 2017, pp. 1872–1880. DOI: 10.1109/CVPRW.2017.234.

[4] Liang-Chieh Chen et al. “DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 40.4 (2018), pp. 834–848. DOI: 10.1109/TPAMI.2017.2699184.

[5] Girija Chetty, Monica Singh, and Matthew White. “Blind Image Tamper Detection Based on Multimodal Fusion”. In: *Neural Information Processing. Models and Applications*. Ed. by Kok Wai Wong, B. Sumudu U. Mendis, and Abdesselam Bouzerdoum. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 557–564. ISBN: 978-3-642-17534-3. DOI: 10.1007/978-3-642-17534-3_69.

[6] Y. Hamid, S. Elyassami, Y. Gulzar, et al. “An improvised CNN model for fake image detection”. In: *International Journal of Information Technology* 15 (Jan. 2023), pp. 5–15. DOI: 10.1007/s41870-022-01130-5. URL: https://doi.org/10.1007/s41870-022-01130-5.

[7] Kaiming He et al. “Mask R-CNN”. In: *2017 IEEE International Conference on Computer Vision (ICCV)*. 2017, pp. 2980–2988. DOI: 10.1109/ICCV.2017.322.

[8] Dirk Hölscher et al. “Exploring the Efficacy and Limitations of Histogram-Based Fake Image Detection”. In: *Procedia Computer Science* 246 (2024). 28th International Conference on Knowledge Based and Intelligent information and Engineering Systems (KES 2024), pp. 2882–2891. ISSN: 1877-0509. DOI: https://doi.org/10.1016/j.procs.2024.09.382. URL: https://www.sciencedirect.com/science/article/pii/S1877050924024165.

[9] A. K. Jaiswal and R. Srivastava. “Forensic image analysis using inconsistent noise pattern”. In: *Pattern Analysis and Applications* 24 (May 2021), pp. 655–667. DOI: 10.1007/s10044-020-00930-4. URL: https://doi.org/10.1007/s10044-020-00930-4.

[10] Enji Liang et al. “Multi-Scale Feature Attention Fusion for Image Splicing Forgery Detection”. In: 21.1 (Dec. 2024). ISSN: 1551-6857. DOI: 10.1145/3698770. URL: https://doi.org/10.1145/3698770.

[11] Bo Liu, Chi-Man Pun, and Xiao-Chen Yuan. “Digital Image Forgery Detection Using JPEG Features and Local Noise Discrepancies”. In: *The Scientific World Journal* 2014 (2014). Research Article, Open Access, First published: 16 March 2014, p. 230425. DOI: 10.1155/2014/230425. URL: https://doi.org/10.1155/2014/230425.

[12] Ruchika Malhotra. “A systematic review of machine learning techniques for software fault prediction”. In: *Applied Soft Computing* 27 (2015), pp. 504–518. ISSN: 1568-4946. DOI: https://doi.org/10.1016/j.asoc.2014.11.023. URL: https://www.sciencedirect.com/science/article/pii/S1568494614005857.

[13] Renu Gopal Mani et al. “A Survey on Digital Image Forensics: Metadata and Image forgeries”. In: *Workshop on Autonomic Communication*. 2022. URL: https://api.semanticscholar.org/CorpusID:249335675.

[14] Ocha Maulidya and Imam Riadi. “Image Forensics to Detect Image Authenticity using Error Level Analysis and Noise Analysis Methods”. In: *International Journal of Computer Applications* 185.28 (Aug. 2023), pp. 6–11. DOI: 10.5120/ijca2023923026. URL: https://doi.org/10.5120/ijca2023923026.

[15] Rachna Mehta et al. “An ensemble learning approach for resampling forgery detection using Markov process”. In: *Applied Soft Computing* 147 (2023), p. 110734. ISSN: 1568-4946. DOI: https://doi.org/10.1016/j.asoc.2023.110734. URL: https://www.sciencedirect.com/science/article/pii/S1568494623007524.

[16] A. M. Nagm et al. “Detecting image manipulation with ELA-CNN integration: a powerful framework for authenticity verification”. In: *PeerJ Computer Science* 10 (2024), e2205. DOI: 10.7717/peerj-cs.2205. URL: https://doi.org/10.7717/peerj-cs.2205.

[17] Rafael Padilha et al. “Content-Aware Detection of Temporal Metadata Manipulation”. In: *IEEE Transactions on Information Forensics and Security* 17 (2022), pp. 1316–1327. DOI: 10.1109/TIFS.2022.3159154.

[18] Giuseppe Papari and Nicolai Petkov. “Edge and line oriented contour detection: State of the art”. In: *Image and Vision Computing* 29.2 (2011), pp. 79–103. ISSN: 0262-8856. DOI: https://doi.org/10.1016/j.imavis.2010.08.009. URL: https://www.sciencedirect.com/science/article/pii/S0262885610001253.

[19] U. Samariya, S. D. Kamble, S. Singh, et al. “A survey on copy-move image forgery detection based on deep-learning techniques”. In: *Multimedia Tools and Applications* (2024). DOI: 10.1007/s11042-024-20323-7. URL: https://doi.org/10.1007/s11042-024-20323-7.

[20] Xinyi Wang, He Wang, and Shaozhang Niu. “An Intelligent Forensics Approach for Detecting Patch-Based Image Inpainting”. In: *Mathematical Problems in Engineering* 2020.8 (2020), pp. 1–10. DOI: 10.1155/2020/8892989. URL: https://doi.org/10.1155/2020/8892989.

[21] Peng Zhou et al. *Learning Rich Features for Image Manipulation Detection*. 2018. arXiv: 1805.04953 [cs.CV]. URL: https://arxiv.org/abs/1805.04953.