AI based Tamper Detection for Digital Media

Deepika Borgaonkar[1*], *Ginnaram* Saikishore[2], *Hinduri* Thanmai[2], *Guttala* Jyothsya[2], and *Kothapalli* Lasya[2]

[1]Chaitanya Bharathi Institute of Technology, Hyderabad, India
[2]VNR Vignana Jyothi Institute of Engineering & Technology, Hyderabad, India

Abstract. Image and video forgeries are becoming more advanced with the latest developments of artificial intelligence and other digital editing software that have highly developed and take the artificial intelligence and the digital editing software to another level that becomes impossible to identify without the need to intervene in the data. This becomes a serious challenge to the forensic investigation department, journalism and cyber security. Although they are very popular, traditional convolution neural networks (CNNs) have serious limitations in identifying the most delicate manipulations because they do not allow the preservation of spatial hierarchies. To fill in this hole, we introduce a hybrid tamper detection system based on combining Capsule Networks and Error Level Analysis (ELA). CapsNet preserves spatial relation to detect the fine-grained anomalous areas, whereas ELA focuses on non uniformities throughout the compression artifacts to determine the areas where tampering has occurred. The suggested approach was tested on the common forensic datasets and showed higher accuracy and resilience than conventional CNN-based models in both the detection of image and video forgery. The study presents a stable and understandable method of authenticating digital media that can be used to provide a robust defense against misinformation, cybercrime, and deepfake filtering.

1 Introduction

The acceleration of editing technology and AI-created synthetic media has exponentially raised the scale and sophistication of counterfeit content. Splicing, copy-move forgery, and deepfakes can now perform flawless manipulation of images and videos, raising serious challenges in areas such as digital forensics, journalism, legal evidence validation, and cybersecurity. With visual content dominating public opinion and decision-making, media authentication capability is more urgently needed than ever before.

Conventional forgery detection methods, including metadata inspection or statistical noise analysis, have lost their effectiveness against contemporary manipulation methods. Consequently, deep learning models, especially Convolutional Neural Networks (CNNs), have been extensively used because they can learn visual patterns and identify anomalies. But CNNs also have one critical weakness: they tend to be based on pooling layers that ignore significant spatial context, and thus are worse at detecting subtle or localized forgeries—particularly on high-resolution or well-edited material.

To counter this, we introduce a hybrid forgery detection system that combines Capsule Networks (CapsNet) and Error Level Analysis (ELA). CapsNet retains spatial hierarchies by dynamic routing between capsules, enabling it to identify fine-grained structural anomalies in images and video frames. In contrast to traditional CNNs, it retains positional context, making it particularly suitable for identifying minute changes that define high-quality forgeries.

Among forensic techniques ELA operates as an add-on that detects compression variation patterns that occur only during tampering events. An analysis using the Enhancement Local Area method web service displays artifactual changes in JPEG compression which allows analysts to pinpoint tampering zones more accurately. When integrating ELA outputs with CapsNet through an add-on input scheme the system receives both elementary forensic features and advanced semantic processing capabilities.

The strategy aims to detect various forms of manipulation with standardized success rates on both image and video data. The deep learning and forensic analysis fusion powers this framework to produce reliable and diversified and interpretable solutions against synthetic content fraud detection problems.

2 Literature survey

As the manipulation technologies in the sphere of digital technologies develop dramatically, it has become more difficult to find out the image and video forgeries. Old-fashioned forensic techniques are not enough and the heavy movement toward deep learning-based techniques has occurred.

Sabour et al. (2017) were the pioneers in proposing Capsule Networks (CapsNet) [1], where routing and the implementation of the CNNs itself are performed with replacement of scalar neurons by vectorized capsules.

* Corresponding author: deepika.borgaonkar12@gmail.com

The design maintains the hierarchies of space, e.g., orientations, scales, and partwholes, and thus is very useful in identifying subtle manipulations that CNNs tend to overlook. Based on their work, our strategy will be based on it since CapsNet is able to detect fine-grained structural anomalies that are fragile and indicative of sophisticated forgeries.

The datasets have played a major role in advancing the forgery detection. The FaceForensics++ created by Rössler et al. (2019) [2] demonstrated the difficulties with identifying high-quality manipulations in video sequences. In a similar fashion, Zhou et al. (2017) [3] introduced a two-stream neural network, which separates the spatial and temporal features and showed that the integration of motion-related cues with stationary image features enhances resistance to video forgeries. Afchar et al. (2018) [4] also proposed MesoNet, the small-size deep network in detecting forgeries in facial videos thereby emphasizing the need of such simple but efficient architecture in practical settings.

Preprocessing methods are also very important. Error Level Analysis (ELA) was proposed by Cozzolino et al. (2017) [5] as the method to identify the area of tampering within JPEG images by taking advantage of the recompression error cases. ELA per se cannot semantically interpret manipulations; nevertheless, it gives useful low-level forensic hints, which can be used to enhance deeper models. Simultaneously, Yu et al. (2020) [6] demonstrated that attention to suspicious areas is more effective than homogenous image processing and that it affects how we combine routing-by-agreement of CapsNet with localized forensic indicators.

In the case of video forgery detection, Ganguly et al. (2022) [7] exploited the concept of visual attention on dynamic frames to overcome the temporal disparities generated during the deepfake generation. On the same note, Nguyen et al. (2019) [8] showed the prospects of Capsule Networks in forensic implementation, which directly motivated our choice to integrate CapsNet into forgery detection. Verdoliva (2020) [9] also insisted on the need to have versatile forensic systems, which will be resistant to manipulation methods that change with time. Amerini et al. (2017) [10] were complementary to it, providing an approach to localization of image forgeries, with specific emphasis on the fact that accurate spatial classification can increase forensic reliability and interpretability.

Big data has also contributed to accelerations. In the study of DeepFake forensics, Li et al. (2020) [11] proposed a new challenging dataset, Celeb-DF, and made it possible to benchmark it realistically. This line of work was expanded by Dang et al. (2020) [12], who performed a comparison of different digital face manipulations, justifying the importance of strong approaches in relation to different types of manipulation. The DeepFake Detection Challenge (DFDC) dataset has been published by Dolhansky et al. (2019) [13], and it has been used to date as a major point of reference during comparative evaluations. A survey of deep learning-based image forgery detection was carried out by Jain and Jain (2020) [14], which puts our decision to combine CapsNet with preprocessing in context. The representation of a reliability perspective by T. Wang et al. (2023) [15] offered a full survey in terms of designing systems of forgery detection that generalize in an adversarial environment.

In addition to datasets and surveys, new detection methods have developed. Dang et al. (2022) [16] showed that resampling features with deep learning are effective in forgery localization, which has a direct impact on our focus on fine-grained forensic features. Zhou et al. (2018) [17] concentrated on rich features to manipulation detection, which agrees with the fact that CapsNet naturally retains spatial hierarchies. Bappy et al. (2017) [18] examined noise-based splicing detection based on camera attributes and camera features as forensic indicators that integrate with other methods such as ELA.

Deepfakes threat has also been researched in terms of security. Korshunov and Marcel (2018) [19] examined the threat of deepfakes in connection to face recognition with emphasis on the need to detect properly. Lastly, Barni and Tondi (2018) [20] examined the transferability of adversarial attacks in image forensics and found it to be vulnerable to researching a more robust architecture like our hybrid framework.

To conclude, CNN-based methods and related datasets, architectures, and preprocessing have given rise to forgery detection studies. Nonetheless, the current approaches have a hard time in capturing spatial correlations and fine-grained manipulations. We work around this gap with a more competent and accurate scheme of detection of forging of images and videos, combining the spatial hierarchy preservation of CapsNet and the detection of artifacts used in ELA to provide a more robust, accurate and interpretable architecture.

3 Methodology

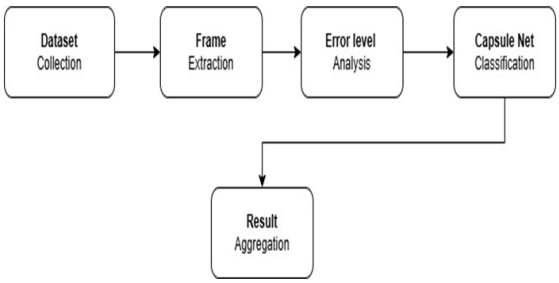

Fig. 1. Methodology.

First, the general approach of the proposed framework is outlined in Figure 1, where it is shown how Error Level Analysis (ELA) preprocessing is integrated with Capsule Networks (CapsNet) to show how two low-level forensic cues can be combined with high-level reasoning focused on spatial factors.

3.1 Preprocessing of data

Convert all images and videos into fixed file formats which should be JPEG or PNG. The objective of Error Level Analysis (ELA) is to produce pixel-level

inconsistencies for revealing manipulation areas in possible tampered content.

3.2 CapsNet + ELA architecture

The Structural forgery detection requires the use of Capsule Network (CapsNet) that preserves spatial arrangements within hierarchies.

ELA-improved inputs guide the detection of tampered local areas through an analysis of compression features.

When CapsNet operates with ELA functionality it boosts both interpretability levels while enhancing the option to identify forged elements.

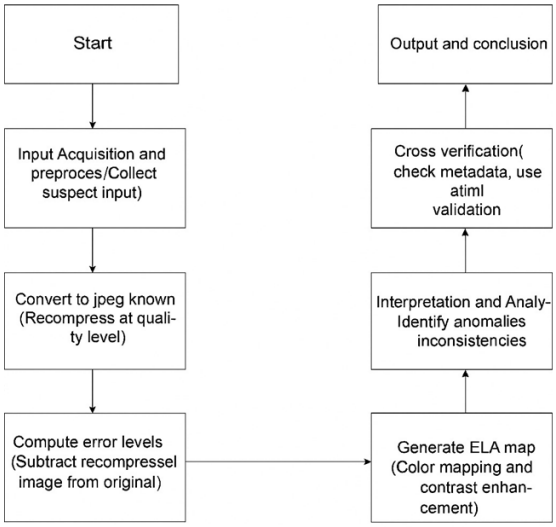

Fig. 2. ELA image/video preprocessing.

Figure 2 illustrates the ELA contribution in the preprocessing step, where pixel-resolution inconsistencies and compression noises are derived in tampered images/videos to act as important features in further classification.

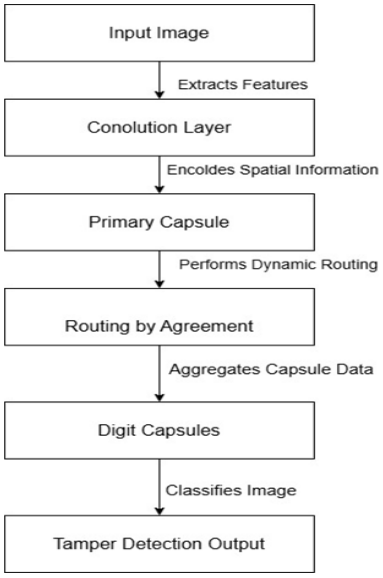

Fig. 3. Capsule net architecture.

The capsule Network architecture, as depicted in Figure 3 is used to maintain spatial hierarchies by employing the dynamic routing procedure between and among the capsules thus allowing a fine-grained detection of the manipulated content to be undertaken.

3.2.1 Capsule net algorithm overview

3.2.1.1 Motivation

Classic CNNs are good at image classification but not at handling spatial hierarchies and pose information. Sabour et al. (2017) tackled this by maintaining spatial relationships between features and capturing more intricate patterns like orientation, scale, and deformation — particularly helpful in the detection of image/video forgeries.

3.2.1.2 The Primary Elements of CapsNet

- Capsules: A capsule is a set of neurons whose output is a vector (not a scalar as in CNNs). The magnitude of the vector is the probability of the presence of a feature, and the direction is its pose information (position, rotation, scale, etc.).
- Dynamic Routing (Between Capsules): Rather than employing max pooling, CapsNet utilizes dynamic routing to determine which higher-level capsule a lower-level capsule must route its output to. This iterative method reinforces connections among capsules that concur on the pose of an object.

3.2.2 CapsNet architecture (standard flow)

- Input Layer: Takes raw or ELA-processed images/video frames.
- Convolutional Layer: Extracts low-level features (such as edges and textures).
- Primary Capsules: Reshape convolutional outputs into vectors (capsules)Each capsule detects simple spatial features.
- Digit Capsules (or Class Capsules): Take inputs through dynamic routing from primary capsules. Encode complex object-level features and relations.
- Output: The length of each output vector represents the probability of each class (such as tampered or untampered). The capsule with the maximum vector length makes the prediction.

3.3 Model training and loss function

The model is trained as a binary classifier to identify tampered vs. untampered images/videos.

3.3.1 Loss function

Capsule Nets use margin loss to establish their classification capabilities.

The addition of reconstruction loss increases interpretability for the model.

3.3.2 Optimization techniques

The model uses SGD or Adam optimizer for efficient learning.

3.4 Deployment

Multiple real-world applications can implement the trained model which has been developed from the dataset.

Web applications for digital forensics.

Mobile applications for real-time verification.

Forensic tools for law enforcement and media verification.

4 Results

Figure 4 shows the outcome of real-time validation which is the visual evidence of the system reliability distinguishing between original and forged media via deployment.

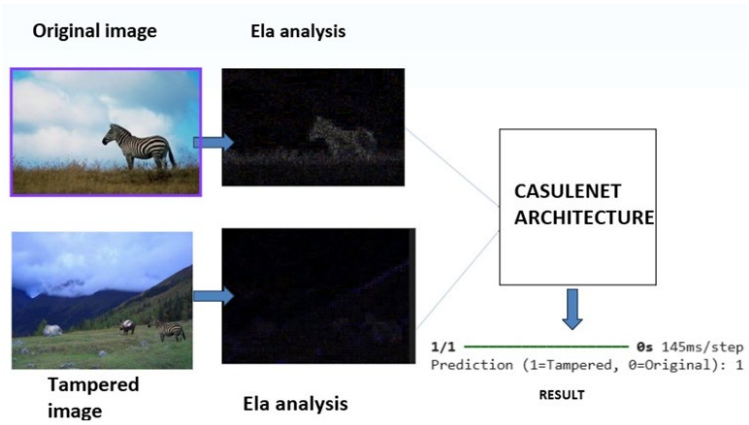

Fig. 4. Result of real-time verification.

This paper presents an approach for detecting tampered image and video content through a combination of Error Level Analysis (ELA) and Capsule Networks. The CG-1050 dataset was employed for image forgery detection, while a customized video dataset reflecting real-world forgery scenarios was used for video analysis. The evaluation process involved testing with both images and videos, followed by result analysis to determine the optimal detection model.

ELA was utilized to identify compression inconsistencies that arise due to image editing, while CapsNet classified the processed inputs to distinguish between authentic and tampered content. Data augmentation techniques were applied to expand the dataset and improve generalization of the model. The proposed implementation achieved an accuracy of 97% in differentiating tampered images from authentic ones, highlighting its robustness and reliability.

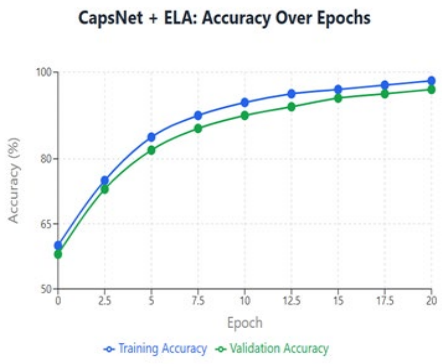

Fig. 5. Training and validation accuracy of capsnet with error level analysis (ELA).

Figure 5 illustrates the training and validation accuracy of the proposed model. The results show a consistent improvement across iterations, ultimately achieving nearly 97% accuracy. This highlights the effectiveness of integrating CapsNet with ELA, emphasizing the strength of this combined approach in enhancing model performance.

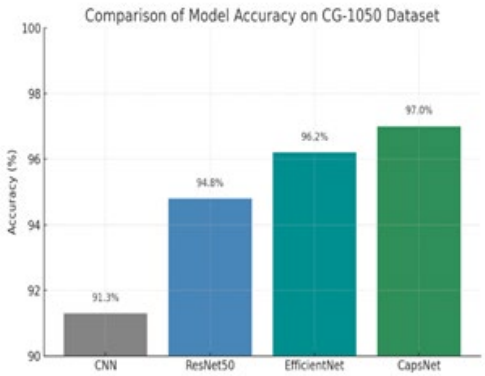

Fig. 6. Accuracy achieved for different models.

A comparative assessment of the model accuracy is represented in Figure 6 displaying the superiority of the proposed CapsNet + ELA approach to traditional CNN architecture, ResNet50 and EfficientNet models with the highest accuracy of 97 percent.

5 Conclusion

This paper introduces a new method of image and video forgery detection through the combination of Capsule Networks (CapsNet) and Error Level Analysis (ELA). Existing CNN-based techniques tend to falter with fine-

grained tamper localization since they do not maintain spatial hierarchies. CapsNet transcends this drawback because it retains spatial relationships, thus enhancing the accuracy of tamper detection. ELA augments this method by pointing out anomalies in compression artifacts so that tampered area detection becomes more accurate.

The proposed method underwent tests on standard forensic data samples to exhibit better performance than typical deep learning network structures. standard deep learning architectures. By combining CapsNet with ELA the system achieves improved results. CapsNet along with ELA provides protection against compression deformations and adversarial manipulations in addition to preventing deepfake types of artificial forgeries. The model demonstrates improved accuracy through its ability to locally identify tampering with fine precision. interpretability during forensic analysis

Future scope

A combination of CapsNet and ELA-based forgery detection framework shows great potential yet several domains convey potential for improvement.

A real-time optimization framework becomes essential when implementing a forgery detection system in journalism as well as for social media applications and digital forensics processes. The system needs to minimize both processing time and system delays to enable quick and immediate forgery detection responses.

For digital media detection systems to remain effective the capability to grow with increasing media file volumes must exist. The system must maintain strong accuracy together with high performance when processing big mixed data sets that incorporate various manipulation types.

The framework becomes more effective through multi-modal expansion since it expands into text and audio and cross-media detection to help fight misinformation on various platforms.

The adoption of Explainable AI (XAI) enables forensic analysts to grasp model outputs because of visual or semantic explanations which improve transparency for legal and investigative work.

The model must undergo modifications to establish higher resistance against adversarial attacks to achieve better robustness. Through adversarial training and defense methods the model gains better dependability when deployed in actual operational conditions.

The creation of minimalistic versions of the model will establish possibilities for real-time utilization throughout forensic software and social networks as well as law enforcement systems for large-scale automated content assessment.

Author contributions

Conceptualization, GS and HT; methodology, GS; software, GJ; validation, HT, KL, and GJ; formal analysis, HT; investigation, GJ; resources, BD; data curation, BD; writing—original draft preparation, HT; writing—review and editing, BD; visualization, HT; supervision, BD; project administration, BD.

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors

References

1. S. Sabour, N. Frosst, G.E. Hinton, Dynamic routing between capsules, NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems, Curran Associates Inc, Red Hook, NY, United States, December 04 (2017), 3859-3869
2. A. Rössler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, M. Nießner, FaceForensics++: Learning to detect manipulated facial images, In Proceedings of the IEEE/CVF international conference on computer vision, IEEE, Seoul, South Korea, February 27 (2019), 1-11
3. P. Zhou, X. Han, V.I. Morariu, L.S. Davis, Two-stream neural networks for tampered face detection, In 2017 IEEE conference on computer vision and pattern recognition workshops (CVPRW), IEEE, Honolulu, HI, USA, August 24 (2017), 1831-1839
4. D. Afchar, V. Nozick, J. Yamagishi, I. Echizen, MesoNet: A compact facial video forgery detection network, In 2018 IEEE international workshop on information forensics and security (WIFS), IEEE, Hong Kong, China, January 31 (2019), 1-7
5. D. Cozzolino, G. Poggi, L. Verdoliva, Recasting residual-based local descriptors as convolutional neural networks: An application to image forgery detection, In Proceedings of the 5th ACM workshop on information hiding and multimedia security, Association for Computing Machinery, New York, NY, United States, June 20 (2017), 159-164
6. C. Yu, J. Jiang, X. Zhou, Attention-based neural networks for image forgery detection. Pattern Recogn. Lett, (2020).
7. S. Ganguly, S. Mohiuddin, S. Malakar, E. Cuevas, R. Sarkar, Visual attention-based deepfake video forgery detection. Pattern Anal. Appl. **25**, 981-992 (2022). https://doi.org/10.1007/s10044-022-01083-2
8. H.H. Nguyen, J. Yamagishi, I. Echizen, Capsule-Forensics: Using capsule networks to detect forged images and videos, In ICASSP 2019-2019 IEEE international conference on acoustics, speech and signal processing (ICASSP), IEEE, Brighton, UK, May 12-17 (2019), 2307-2311
9. L. Verdoliva, Media forensics and DeepFakes: An overview. IEEE J. Sel. Top Signal Process. **14**, 910-932 (2020). https://doi.org/10.1109/JSTSP.2020.3002101
10. I. Amerini, T. Uricchio, L. Ballan, R. Caldelli, A deep learning-based approach for image forgery detection and localization. Pattern Recogn. Lett.

113, 430-437 (2017). https://doi.org/10.1016/j.patrec.2017.04.016

11. Y. Li, X. Yang, P. Sun, H. Qi, S. Lyu, Celeb-DF: A large-scale challenging dataset for DeepFake forensics, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Seattle, WA, USA, August 05 (2020), 3204-3213
12. Dang, H., Liu, F., Stehouwer, J., Liu, X., & Jain, A. K., On the detection of digital face manipulation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern recognition, IEEE, Seattle, WA, USA, August 05 (2020), 5781-5790
13. B. Dolhansky, R. Howes, B. Pflaum, N. Baram, C.C. Ferrer, The DeepFake Detection Challenge (DFDC) Preview Dataset. arXiv preprint, (2019). https://doi.org/10.48550/arXiv.1910.08854
14. A. Jain, S. Jain, Deep learning based image forgery detection: A survey. arXiv preprint, (2020). https://doi.org/10.48550/arXiv.2005.04296
15. T. Wang, X. Liao, K.P. Chow, X. Lin, Y. Wang, Deepfake detection: A comprehensive survey from the reliability perspective, ACM Computing Surveys, Association for Computing Machinery, New York, NY, United States, November 11 (2024), 1-35
16. H. Dang, F. Liu, J. Stehouwer, X. Liu, A.K. Jain, Detection and localization of image forgeries using resampling features and deep learning. IEEE Trans. Inf. Forensics Secur. **17**, 167-182 (2022).
17. Zhou, P., Han, X., Morariu, V. I., & Davis, L. S., Learning rich features for image manipulation detection, In Proceedings of the IEEE conference on computer vision and pattern recognition, IEEE, Salt Lake City, UT, USA, December 16 (2018), 1053-1061
18. J.H. Bappy, C. Simons, A.K. Roy-Chowdhury, A. A. Rad, Exposing splicing forgery using noise features and camera characteristics, IEEE Winter Conference on Applications of Computer Vision (WACV), (2017).
19. P. Korshunov, S. Marcel, DeepFakes: A new threat to face recognition? Assessment and detection. arXiv preprint, (2018). https://doi.org/10.48550/arXiv.1812.08685
20. M. Barni, B. Tondi, Transferability of Adversarial Attacks in Image Forensics. IEEE International Conference on Image Processing (ICIP), (2018).