International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 04 April 2025
Page No: 1111 - 1116
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0160

Hybrid Deep CapNet-VGG19 Model for Detecting Forged Images and Videos

Mr. Gowtham[1], Ms. Jayakrishana Bharathi S [2], Ms.Dharani M [3], Ms.Prasika S [4]
[1]Assistant Professor, Department of CSE, Jai Shriram Engineering College, Tirupur, India.
[2,3,4] UG Scholar Department of CSE, Jai Shriram Engineering College, Tirupur, India.
Emails: *jaibharathi345@gmail.com[1], dharanimuthusamy04@gmail.com [2], prasikasaravanan03@gmail.com[3]*

Abstract
In recent days, Video and image forgery detection becomes significant aspect due to modern editing software which offers powerful and easy to handle tools for the manipulation of videos. a technique of video detection faces much more challenges like video dataset, post-processing issues, noise, computational time/complexity, deepfake detection approach and insufficient anti-forensic. Consequently, this concern is addressed in this work by presenting a deep capsule network model based forensic detection. For this purpose, this paper presented a Deep learning (DL) based model for the detection of forged images and videos. This model presents a technique which uses a Deep capsule network (CapsNet) and VGG19 model for detecting forged videos and images in a huge range of forgery scenarios, which includes detection of replay attack and computer-generated video/image detection (both fully and partially). The performance outcome is estimated for deepfake dataset and outcomes attained are compared with existing models in terms of accuracy at both frame level and video level. The analysis shows that proposed model is effective and offers enhanced outcome than traditional models.

Keywords: *Forensic detection, Video and image forgery, deepfake, Deep learning, Deep CapsNet, VGG19.*

1. Introduction
A rise of deep fake technology opens a new frontier in the world of digital communication, thus creating the convincing content of synthetic video [1]. Though, this advancement provides potential for the positive application, this in turn poses important risks for the integrity and security of information. For creating fake news media and bypassing facial authentication, forged videos and images could be used. A manipulated videos and image quality have seen a significant enhancement with the growth of advanced networking frameworks [2]. And this in turn simplified the facial forgeries creation more dramatically. At this age of social networks which serves as major source of information, fake news with the multimedia manipulated could spread quickly and have some noteworthy effects. A phenomenon of deep fake is a better example of this kind of threat- any such people having computer could create videos thus incorporating the facial image of any celebrity using the technique of artificial intelligence based human image synthesis [3]. Various countermeasures were presented for dealing with manipulated videos and images with the use of machine learning (ML) schemes. Moreover, most of them intends at specified kind of attacks and might fails in detecting replay attack when the displayed video is of the actual targeted person. Deep learning models shows huge range of success in varied applications [4-7]. Existing convolutional neural network (CNNs), though with powerful video and image analysis tool have some limitations on applying to deep fake detection. However, CNN are more likely susceptible to overfitting issue, particularly when trained on the limited data. This might lead to poor generalization thus causing CNN to fail on encountering different or new kind of deepfakes. The overfitting issue might be problematic highly in the detection of deepfake context because of relative scarcity of deepfake video for the training purposes. To overcome the limitations of existing CNN models, with the introduction of dynamic routing algorithm “capsule network” architecture model was presented with a primary outcome remarkably [8]. Several recent studies prove that, in agreement among capsules computed by dynamic routing model, hierarchical relationship is posed among the object’s parts. This in turn

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 04 April 2025
Page No: 1111 - 1116
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0160

improves the vision tasks accuracy. The application of capsule network to the task of forensic is the main focus of this work and is regarded a challenging concern. Moreover, the agreement among capsules attained on using dynamic routing model could boost the performance of detection on a complex and closely flawless forged videos and images.

1.1.Research gap/Problem identification

Several forms of multimedia fraud detection are complex in detection, specifically certain evidence is not used, hence a technique of video detection faces much more challenges like video dataset, post-processing issues, noise, computational time/complexity, deepfake detection approach and insufficient anti-forensic. Consequently, this concern is addressed in this work by presenting a deep capsule network model based forensic detection.

1.2.Objective/Contribution

This work presents a scheme which uses a capsule network for detecting forged videos and images in a huge range of forgery scenarios, which includes detection of replay attack and computer-generated video/image detection (both fully and partially). It is a pioneering work using capsule network that are designed for the issue of computer vision and to resolve the concern of digital forensic.

2. Related Works

A new enhanced fake video detection scheme is presented in the work [9] for addressing several swapping threats and the problem of low generalization. The stage of preprocessing was employed for minimizing noise in data thus to enhance its quality. A suggested framework uses modified application of capsule neural network (CapsNet) having enhanced routing approach. A capsule network model was suggested in the work [10] for detecting object dependent forensic model in the surveillance videos. In presented model, motion residual is used which was computed from each vide frame for extracting intra- and inter-frame inherent statistical characteristics of video sequence as the capsule network input. The outcome shows that the proposed model attains significant performance for authentic, double compressed and detection of forged frame, irrespective of picture groups degree and length of compression in the videos. A prominent digital manipulation was suggested in [11] having special emphasis on facial content because of its huge number of probable applications. In specific, principles of six kinds of digital face manipulation was covered like synthesis of entire face, swapping identity, morphing face, manipulation of attribute, swapping expression, and audio & text-to-video. In the work [12], a new deepfake detection model iCaps-Dfake was proposed which competes with existing models of deepfake video detection thus addressing the issue of low generalization. Local binary pattern (LBP) and CNN were the two models of feature extraction employed based on modified High-resolution network (HRNet) together with the use of CapsNet model implementing concurrent routing approach. The author in the work [13], a novel deep learning strategy was employed using efficient-capsule network (E-Cap Net) to classify facial images generated over varied deepfake generative models. Specifically, low-cost max-feature map (MFM) activation function was introduced at each primary capsule of suggested E-Cap net. MFM activation use enables E-Cap Net becomes robust and light since it suppresses the lower activation neurons in every primary capsule. The major intention of the work [14] was to use CNN and CapsNet along with LSTM for distinguishing which video frames were generated by deepfake approach and which was the real one. It was also necessary to identify predicted model output thus analyzing the patterns with the use of Explainable AI. The model presented in [15] implemented two designs which reaches a progress discovery accuracy, thus addressing the temporal and spatial disorders, inherent to deepfake media. This is discovered from result that this scheme was effectual on varied datasets thus suggesting better ability of generalization in various kinds of Deepfake media. A design approach was capable for the task of real-time detection in its present form.

3. Proposed Work

The proposed working module of proposed methodology is narrated briefly in this section. The suggested scheme is depicted in figure 1. For input of video, video will be split as frames at the preprocessing stage. The results of classification

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 04 April 2025
Page No: 1111 - 1116
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0160

(posterior probabilities) will then be attained in post-processing stage for getting final outcome. The remaining sections are identical to input image. In the phase of preprocessing, faces are detected and thus scaled to 128×128. As like existing model [1], Vgg-19 model is employed for extracting the latent features, that are considered input to the capsule network. But, the proposed work differs by taking the output of third maxpooling layer instead of three outputs from ReLU layer. This is done to reduce the reduce input size to capsule network. (Figure 1)

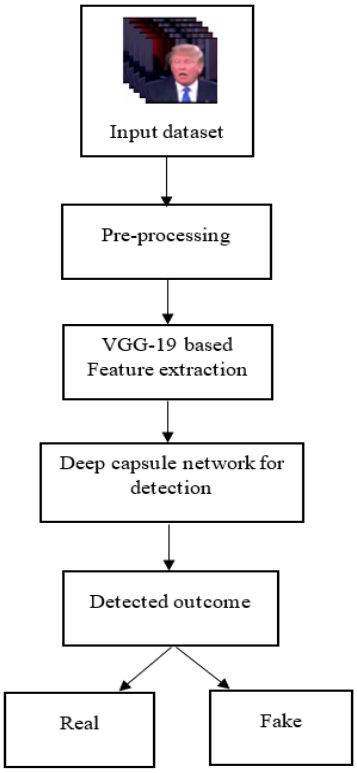

Figure 1 Illustration of Proposed Working Module Flow

3.1.Preprocessing

Pre-processing of data is the initial step on designing the scheme. This in turn diverts the raw data that were gathered from several sources to more suitable information. The preprocessing step is needed as raw data comprises of inconsistent or missing values and redundant information, that could inhibit the scheme from getting desired outcome thus causing substantial loss of data. However, most of DL model have conventional targeted form input. Before the training process, dataset should be in that form. The dataset comprises of real and fake videos along with labelling information. Initially, library is used for extracting one frame per second rom each video. A facial frame is needed for training scheme, such that proposed detection process is employed. The resolution of data is corrected by cropping face frames from full frames to train. After data preprocessing, size of (5×128×128×3) are chosen as targeted size, at which frame is 5 with height 128, width 128 and 3 channels.

3.2.VGG19 based feature extraction

After the stage of preprocessing, VGG-19 model is employed as feature extractor so as to extract features from preprocessed data. The latent features are extracted and is given as input to capsule network. In the pre-processing phase, faces are detected and scaled to 128 128. As like existing model [1], Vgg-19 model is employed for extracting the latent features, that are considered input to the capsule network. But, the proposed work differs by taking the output of third maxpooling layer instead of three outputs from ReLU layer. This is done to reduce the reduce input size to capsule network. The Vgg-19 network model comprises of five blocks every one comprises of few convolution layers which is followed by max pooling layer. For the function of convolution, Vgg-19 uses kernel size of 3 × 3 with channels 64, 128, 256, 512, and 512 at its convolution blocks correspondingly.

3.3.Deep Capsule Network for detection

The proposed deep capsule network comprises of three primary capsules with two output capsules, one for real and another one for fake images. (Figure 2) shows the overall design of capsule network.

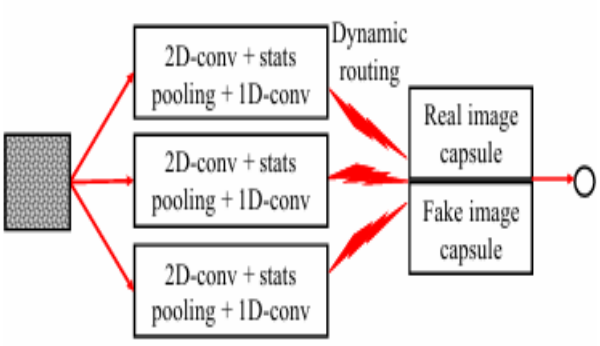

Figure 2 Deep Capsule Forensic Model Design

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 04 April 2025
Page No: 1111 - 1116
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0160

An extraction of latent feature carried by VGG-19 model is the input that are distributed to three primary capsules. The statistical pooling is used since it is important for the detection of forgery. Three capsules outputs ($u_{j|i}$) are routed dynamically to the output capsule (v_j) for r iteration with the use of proposed dynamic routing algorithm 1. A network comprises of 2.8 million parameters, which is relatively small for such network. The existing dynamic routing approach is improved slightly by adding the random Gaussian noise to 3-D weighted tensor W thereby employing one such additional squash as like equation 1 before routing on iterating. (Figure 3) shows the architecture of deep capsule network framework.

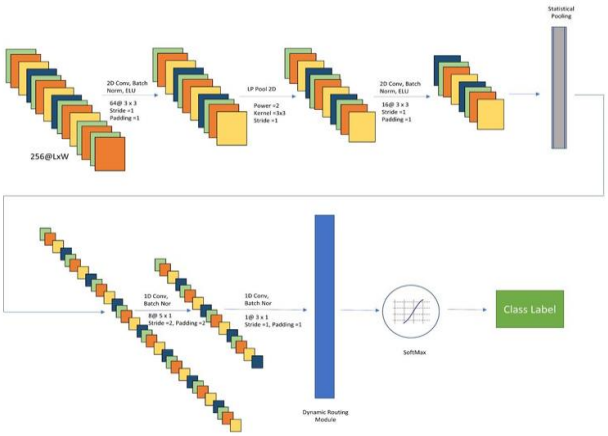

Figure 3 Architecture of Proposed Deep Capsule Network Model

The noise added aids in reducing overfitting issue whereas, the additional equation makes the network a stable one. The primary and output capsule outcome will be illustrated in (Figure 4) shown below.

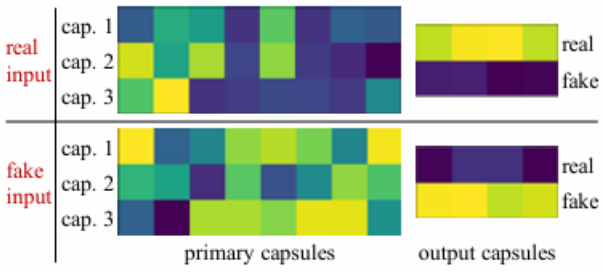

Figure 4 Average Outcome Attained

v_j=squash(s_j)=‖s_j ‖^2/(1+‖s_j ‖^2) s_j/‖s_j ‖ (1)

The cross-entropy function is used and is expressed as follows:

L=-(ylog(yˆ)+(1-y) log(1-yˆ)) (2)

Here, y denotes ground truth label and yˆ denotes predicted label computed by following equation at which m denotes the dimension number of output capsule v_j.

yˆ=1/m ∑_i softmax([■(v_1^T@v_2^T)]_(1,i)) (3)

This equation is used for using the length of output capsule which promotes the separation among 2 output capsules at each dimension. Thus, from this the forged video and image is detected and classified as real and fake.

4. Performance Analysis

The performance assessment is carried in this work and the outcomes attained are projected to represent the effectiveness of proposed model.

4.1.Hardware and Software Setup

The hardware specification (AMD Ryzen 5, 8GB RAM, 256GB SSD, and HD camera) will meet desired specification needs for the testing and development. Moreover, the training of DL models on huge dataset, upgrading of 16GB RAM thus adding GPU (example NVIDIA RTX) series for accelerating computation is needed. For the real-time performance, optimization model will be the key. The Software requirement is given in Table 1.

Table 1 Software Setup

OS	**Ubuntu 20.04 (stable and highly supported for DL)**
Language	Python (versatile and well suited with DL)
Framework	Tensorflow or PyTorch (both supports implementation of capsule network)
Libraries	OpenCV, NumPy/SciPy, Matplotlib
Database	SQLite
IDE	Jupyter Notebook (prototyping), VS code.PyCharm (development)

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 04 April 2025
Page No: 1111 - 1116
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0160

4.2.Performance Evaluation of Proposed Methodology

Table 2 shows the accuracy rate performance comparison with existing models and proposed strategy at frame level on deepfake dataset. The analysis shows that proposed model is offering enhanced outcome than existing models [1].

Table 2 Comparative Estimation of Accuracy at Frame Level

Methods	Values
Meso-4	89.1
Meso-Inception-4	91.7
CNN	92.36
Capsule Forensic	94.47
Capule Forensic noise	95.93
Proposed (Deep CapsNet+VGG19)	97.8

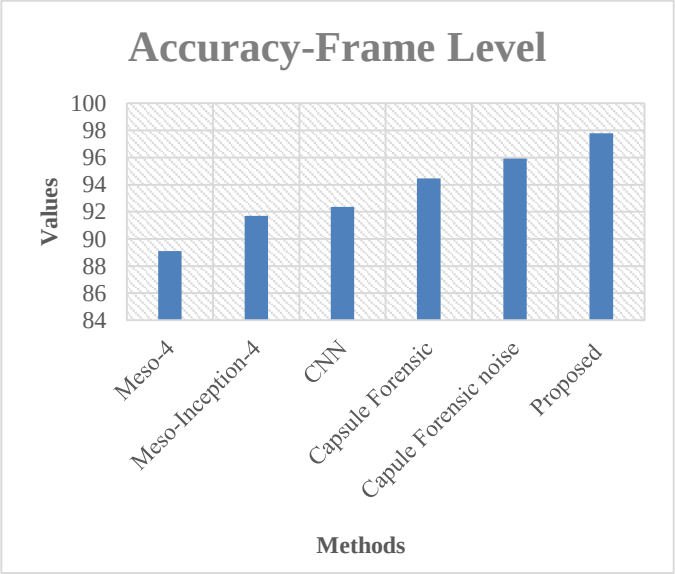

Figure 5 Performance Comparison of Accuracy on Face Swapping Detection-Frame Level

Figure 5 shows the accuracy rate performance comparison with existing models and proposed strategy at frame level on deepfake dataset. The analysis shows that proposed model is offering enhanced outcome than existing models [1]. Table 3 represents the accuracy rate performance comparison with existing models and proposed strategy at video level on deepfake dataset. The analysis shows that proposed model is offering enhanced outcome than traditional models [1].

Table 3 Comparative Estimation of Accuracy at Video Level

Methods	Values
Meso-4	96.9
Meso-Inception-4	98.4
Capsule Forensic	97.69
Capule Forensic noise	99.23
Proposed	99.56

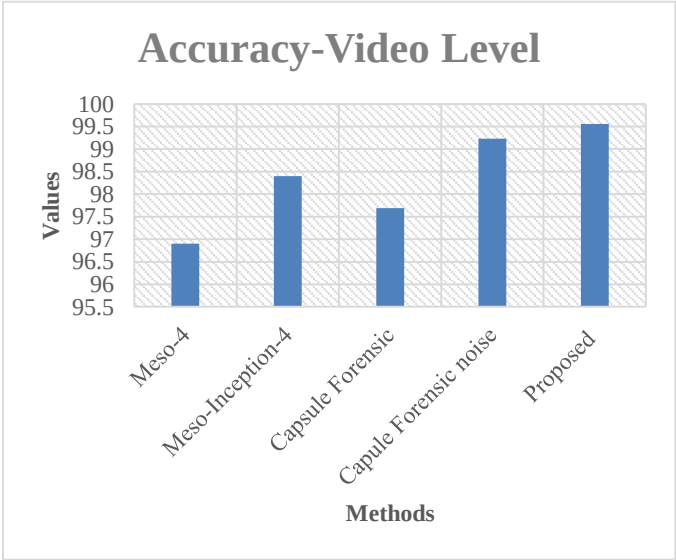

Figure 6 Comparative Estimation of Accuracy at Video Level

Figure 6 represents the accuracy rate performance comparison with existing models and proposed strategy at video level on deepfake dataset. The analysis shows that proposed model is offering enhanced outcome than traditional models [1].

Conclusion

A DL based model for the detection of forged images and videos was presented in this paper. This model presents a technique which uses a Deep capsule network (CapsNet) and VGG19 model for detecting forged videos and images in a huge range of forgery scenarios, which includes detection of replay attack and computer-generated video/image detection (both fully and partially). The performance outcome is estimated for deepfake dataset and outcomes attained

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 04 April 2025
Page No: 1111 - 1116
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0160

are compared with existing models in terms of accuracy at both frame level and video level. The analysis shows that proposed model is effective and offers enhanced outcome than existing methods.

References

[1]. Nguyen, H. H., Yamagishi, J., & Echizen, I. (2019, May). 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) (pp. 2307-2311). IEEE.

[2]. Munawar, M., Noreen, I., Alharthi, R. S., & Sarwar, N. (2023). Forged video detection using deep learning: A slr. Applied Computational Intelligence and Soft Computing, 2023(1), 6661192.

[3]. Mohiuddin, S., Malakar, S., Kumar, M., & Sarkar, R. (2023). A comprehensive survey on state-of-the-art video forgery detection techniques. Multimedia Tools and Applications, 82(22), 33499-33539.

[4]. Kandasamy, V., Hubálovský, Š., & Trojovský, P. (2022). Deep fake detection using a sparse auto encoder with a graph capsule dual graph CNN. PeerJ Computer Science, 8, e953.

[5]. Tao, Y., & Chang, Y. (2024, September). Research on Deep Forgery Defense Technology Based on Artificial Intelligence Security. In Proceedings of the 2024 9th International Conference on Cyber Security and Information Engineering (pp. 7-11).

[6]. Dincer, S., Ustubioglu, B., Ulutas, G., Tahaoglu, G., & Ustubioglu, A. (2023, September). Robust Audio Forgery Detection Method Based on Capsule Network. In 2023 International Conference on Electrical and Information Technology (IEIT) (pp. 243-247). IEEE.

[7]. Lu, T., Bao, Y., & Li, L. (2023). Deepfake Video Detection Based on Improved CapsNet and Temporal–Spatial Features. Computers, Materials and Continua, 75(1), 715-740.

[8]. Liu, Y., Qin, Q., Yang, W., Wu, A., Ma, W., & Zhang, J. (2022, May). Forgery Face Image Detection Based on Improved Capsule Network. In 2022 5th International Conference on Artificial Intelligence and Big Data (ICAIBD) (pp. 321-325). IEEE.

[9]. Khalil, S. S., Youssef, S. M., & Saleh, S. N. (2021, March). A multi-layer capsule-based forensics model for fake detection of digital visual media. In 2020 International Conference on Communications, Signal Processing, and their Applications (ICCSPA) (pp. 1-6). IEEE.

[10]. Bakas, J., Naskar, R., Nappi, M., & Bakshi, S. (2023). Object-based forgery detection in surveillance video using capsule network. Journal of Ambient Intelligence and Humanized Computing, 1-11.

[11]. Nguyen, H. H., Yamagishi, J., & Echizen, I. (2022). Capsule-forensics networks for deepfake detection. In Handbook of Digital Face Manipulation and Detection: From DeepFakes to Morphing Attacks (pp. 275-301). Cham: Springer International Publishing.

[12]. Khalil, S. S., Youssef, S. M., & Saleh, S. N. (2021). iCaps-Dfake: An integrated capsule-based model for deepfake image and video detection. Future Internet, 13(4), 93.

[13]. Ilyas, H., Javed, A., Malik, K. M., & Irtaza, A. (2023). E-Cap Net: an efficient-capsule network for shallow and deepfakes forgery detection. Multimedia Systems, 29(4), 2165-2180.

[14]. Ishrak, G. H., Mahmud, Z., Farabe, M. D., Tinni, T. K., Reza, T., & Parvez, M. Z. (2024). Explainable Deepfake Video Detection using Convolutional Neural Network and CapsuleNet. arXiv preprint arXiv:2404.12841.

[15]. Abisha, M. B., Kathrine, J. W., & Kushmitha, S. (2024, December). Capsule Networks and LSTM Models for Robust Deepfake Detection in Audio and Video. In 2024 International Conference on IoT Based Control Networks and Intelligent Systems (ICICNIS) (pp. 1569-1576). IEEE.