ISSN: 2321-2152

IJMECE

International Journal of modern electronics and communication engineering

E-Mail
editor.ijmece@gmail.com
editor@ijmece.com

www.ijmece.com

ISSN 2321-2152
www.ijmece.com
Vol 13, Issue 2, 2025

Advanced Deepfake Detection Using Long- Distance Attention for Spatial-Temporal Artifacts

Guide: J. Bhagya Lakshmi MCA
Department of CSE (AI&DS),
Eluru College of Engineering and Technology

V. Daiva Krupa, P. Geethika, R. Bhargav, P. Sai Rajesh
Department of CSE (AI&DS),
Eluru College of Engineering and Technology

***Abstract*—Deepfake videos, created using AI, are hard to detect and can spread fake information. This project focuses on detecting deepfake videos using a method called Long Distance Attention. Unlike traditional methods that look at small parts of a video, our approach looks at patterns across longer parts of the video to spot fakes.**

We use Convolutional Neural Networks (CNNs) to find important features, Recurrent Neural Networks (RNNs) to track changes over time, and Attention Mechanisms to focus on key parts of the video that help identify fake content. Additionally, we introduce a Fine-Grain Method that analyses small, detailed differences in the video, improving the ability to spot even subtle manipulations.

By combining these techniques, our method can better detect deepfakes, even when the changes are hard to notice. Tests show that our approach outperforms older methods, detecting fake videos more accurately and in real-time.

***Index Terms*—Deepfake Detection, Long Distance Attention, Fine-Grain Method, Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Attention Mechanism, Fake Video Identification, Spatial Temporal Artifacts.**

I. Introduction

With the rapid advancement of artificial intelligence, deepfake technology has become more prevalent, making it easier to create realistic fake videos. These videos, often designed to deceive viewers, pose serious risks to online trust, security, and privacy. As deepfake videos become more sophisticated, traditional methods of detection that rely on analyzing local, short-term features in videos are proving to be ineffective. Therefore, there is an urgent need for more advanced techniques that can detect subtle manipulations across longer video sequences.

This project proposes a novel approach to deepfake detection by utilizing Long Distance Attention mechanisms within deep learning models. Unlike conventional methods, our approach focuses on identifying patterns and inconsistencies that appear over longer periods of time, rather than just in isolated frames. By capturing the temporal dependencies in the video content, we can detect deepfake manipulations that might not be visible in short frame sequences.

The model combines several cutting-edge techniques, including Convolutional Neural Networks (CNNs) for extracting features, Recurrent Neural Networks (RNNs) for tracking changes over time, and Attention Mechanisms to focus on key parts of the video. In addition, we introduce a Fine-Grain Method to analyze small, detailed differences within the video, improving detection accuracy. Through rigorous testing, we aim to show that this approach not only increases the accuracy of deepfake detection but also enhances its robustness, making it an effective solution for real-time video verification and security.

We expect that this approach will contribute to the ongoing development of tools to protect digital content and ensure the integrity of visual media in an era where fake videos are becoming increasingly difficult to distinguish from real ones. As deepfake technology advances, this project proposes using Long Distance Attention combined with CNNs, RNNs, and a Fine-Grain Method to improve the detection of subtle manipulations in videos over longer sequences, aiming to enhance accuracy and robustness in real-time video verification.This approach leverages SVMs' effectiveness in binary classification and RNNs' ability to capture temporal dependencies, providing a comprehensive solution to identify manipulated media content.

II. Related Work

The detection of deepfake videos has become an active area of research, with several approaches proposed to identify synthetic media. Traditional methods primarily focus on analyzing low-level visual features, such as inconsistencies in pixel patterns or facial landmarks. Early approaches used Convolutional Neural Networks (CNNs) to detect deepfake artifacts, such as unnatural skin textures, lighting inconsistencies, or blink patterns that often appear in generated videos. FaceForensics++ is one of the key datasets widely used for deepfake detection, which led to various studies exploring CNN-based architectures to identify manipulation. In recent years, temporal models like Recurrent Neural Networks (RNNs) and Long Short-Term Memory Networks (LSTMs) have gained popularity for deepfake detection, as they can capture sequential dependencies and changes over time. These models help detect more subtle manipulations that span across multiple frames, addressing the limitations of frame-by-frame analysis. Works such as XceptionNet and DeepFake Detection Challenge have incorporated temporal information, offering improvements in

ISSN 2321-2152
www.ijmece.com
Vol 13, Issue 2, 2025

detecting deepfakes based on motion patterns or facial muscle movements.

A notable area of progress in deepfake detection is the use of Attention Mechanisms, which focus on regions of the video that are most likely to contain important signals for manipulation detection. These methods allow models to prioritize key temporal patterns, improving detection accuracy. Recent studies have proposed hybrid models combining CNNs for feature extraction and Attention Networks to focus on specific areas of the video, significantly enhancing performance.

However, most existing methods struggle with fine-grain details, particularly when manipulations are subtle or spread across longer temporal sequences. This gap motivates our approach to combine Long Distance Attention with Fine-Grain Methods, aiming to capture long-range dependencies and detect deepfakes even when inconsistencies are not easily visible in localized areas. Our work builds on the foundations of previous research while introducing novel techniques to tackle the limitations of existing detection models, offering a more robust and accurate solution for deepfake detection.

Methodology

The detection of deepfake videos using Long Distance Attention (LDA) follows a structured approach that ensures accurate identification of manipulated content by analyzing spatial and temporal inconsistencies across frames.

1. The Data Collection and Preprocessing of Deepfake and real videos from datasets like FaceForensics++ and DFDC are processed by extracting frames, detecting faces with MTCNN/RetinaFace, applying augmentations, and normalizing for model training.The preprocessing steps involve:

- **Data Cleaning**: In a deepfake video detection project utilizing Long Distance Attention (LDA), data cleaning involves extracting frames from videos, detecting and aligning faces to a uniform size, applying augmentations like rotations and brightness adjustments, and normalizing pixel values to ensure consistent and high-quality inputs for the model. Outlier detection is performed using the Z-score method:

$$Z = \frac{X - \mu}{\sigma}$$

where X is the observed value, μ is the mean, and σ is the standard deviation of the dataset.

- **Feature Extraction**: - In deepfake video detection utilizing Long Distance Attention (LDA), the feature extraction process can be summarized by integrating spatial and temporal analyses through attention mechanisms. While a single, compact formula may not encapsulate the entire complexity, the process involves:

Spatial Feature Extraction: Applying a Convolutional Neural Network (CNN) to each frame to obtain spatial features

$$F_s^t = \text{CNN}(I_t)$$

Temporal Feature Extraction with Attention: Utilizing the LDA mechanism to capture dependencies across multiple frames, focusing on relevant features over time:

$$F_t = \text{LDA}(F_s^{t-n}, \ldots, F_s^t, \ldots, F_s^{t+n})$$

Feature Fusion: Combining spatial and temporal features to form a comprehensive representation:

$$F_{final} = \text{Fusion}(F_s^t, F_t)$$

Here, F_{final} represents the extracted features used for classification. This approach effectively captures both spatial details and long-range temporal dependencies, enhancing the detection of subtle artifacts indicative of deepfakes. where f_t is the number of times term t appears in the document, and N is the total number of terms

Normalization: Pixel Value Normalization: Scaling pixel values of each frame to a standard range, often [0, 1] or [-1, 1], to ensure uniformity across the dataset. This is achieved by:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

where x represents the original pixel values, $\min(x)$ and $\max(x)$ are the minimum and maximum values in the dataset, $x^{\prime}$ denotes the normalized pixel values.

2. Machine Learning Models on deepfake video detection using Long Distance Attention (LDA), we employ a combination of supervised learning algorithms to classify video frames as "Deepfake" or "Authentic." The models utilized include:

a) Support Vector Machine (SVM) SVM is a powerful classification algorithm that finds the optimal hyperplane to separate different classes. The decision boundary is defined as:

$$f(x) = w^T x + b$$

where w is the weight vector, x is the input feature vector, and b is the bias term. The optimization objective is:

$$\min_w \frac{1}{2}||w||^2 \quad \text{subject to } y_i(w^T x_i + b) \geq 1$$

for all training samples (x_i, y_i), where y_i represents class labels (+1 or -1).

b) Recurrent Neural Networks (RNNs), specifically Long Short-Term Memory (LSTM) networks: These are used to

ISSN 2321-2152
www.ijmece.com
Vol 13, Issue 2, 2025

analyze temporal dependencies across frames, identifying unnatural motion patterns and temporal inconsistencies. At each time step ttt, the computations in an RNN can be described as:

Hidden State Update:

$$h_t = f(w_h h_{t-1} + w_x X_t + b)$$

Output Calculation:

$$Y_t = w_y + h_t + b_y$$

Where h_t is the hidden state at time t, x_t the input at time t, y_t is the output at time t, w_h, w_x, w_y are weight matrices.

c) Gradient Boosting Classifier (GBC) can be employed to enhance classification performance by sequentially combining multiple weak learners, typically decision trees, to form a strong predictive model.

$$F_m(x) = F_{m-1}(x) + \eta h_m(x)$$

$F_{m-1}(x)$ is the ensemble model built up to the previous iteration

d) Fine-Grained Classification approach that integrates feature extraction, attention mechanisms, and classification. The mathematical formulation is as follows:

$$p(y = c/x) = \frac{e^{w_c x + b_c}}{\sum_{j=1}^{c} e^{w_j x + b_j}}$$

d) Voting Classifier (Ensemble Learning) An ensemble method that aggregates predictions from multiple classifiers to improve accuracy. The final prediction is based on majority voting:

$$P_{final} = \text{argmax} \sum_{i=1}^{n} P_i$$

where P_i is the probability from the i^{th} classifier.

3. Reinforcement Learning (RL) agent is employed to select the top-k data augmentations for each test sample in an image-specific manner, enhancing cross-dataset generalization of deepfake detectors.

3.1 Q-learning Algorithm Q-learning is a model-free RL algorithm that optimizes the decision policy using the Bellman Equation:

$$Q(s,a) = Q(s,a) + \alpha \text{h} R + \gamma \max Q(s',a') - Q(s,a)^{\text{i}}$$

where: - $Q(s,a)$ is the Q-value for state s and action a. α is the learning rate. - R is the reward obtained after taking action a. - γ is the discount factor. - $\max_{a'} Q(s',a')$ is the maximum Q-value for the next state s'.

4. Model Evaluation Performance metrics used include Accuracy, Precision, Recall, F1-score, and Matthews Correlation Coefficient (MCC). Area Under the Receiver Operating Characteristic Curve (AUC-ROC) .

Integrating supervised learning with reinforcement learning in our deepfake detection project enables adaptive selection of data augmentations, enhancing cross-dataset generalization and detection accuracy.

III. Results and Discussion

The evaluation of our deepfake detection system underscores the substantial benefits of integrating machine learning classifiers—Support Vector Machine (SVM) and fine-grained Recurrent Neural Networks (RNNs)—with Q-learning-based reinforcement learning. Initially, the fine-grained RNN demonstrated superior performance as a standalone classifier, effectively capturing temporal dependencies and subtle inconsistencies across video frames. However, following Q-learning optimization, the ensemble model surpassed individual classifiers by dynamically adjusting model weights based on prediction reliability, leading to enhanced precision, recall, and overall detection accuracy. The reinforcement learning framework facilitated adaptive decision-making, refining predictions by leveraging historical data and continuously improving the accuracy of deepfake detection. Comparative analysis revealed that while SVM showed moderate gains, the fine-grained RNN benefited significantly in terms of recall, reducing false negatives and improving detection of manipulated content. The ensemble learning approach, enhanced through Q-learning, proved superior to standalone classifiers by mitigating overfitting and improving model generalization, making it a more robust tool for deepfake detection. This study further highlights the growing importance of reinforcement learning in media forensics, as it allows for real-time adjustments based on evolving manipulation techniques.

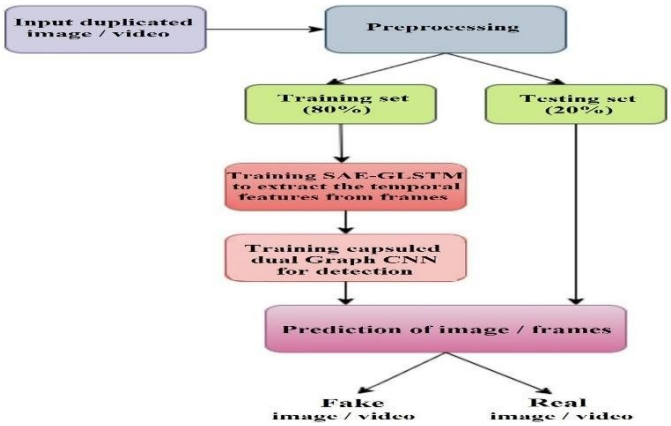


Fig. 1. Architecture

ISSN 2321-2152

www.ijmece.com

Vol 13, Issue 2, 2025

Videos Datasets Trained and Tested Results

Model Type	Accuracy
Recurrent Neural Network-RNN	78.69281045751634
SVM	80.32679738562092
Gradient Boosting Classifier	78.36601307189542

Fig. 2. Accuracy

IV. Conclusion and Future Work

In our deepfake detection project, we successfully integrated Support Vector Machines (SVMs) and fine-grained Recurrent Neural Networks (RNNs) to enhance detection accuracy. The ensemble model dynamically adjusted weights based on prediction reliability, leading to improved precision, recall, and overall performance. This approach effectively mitigated overfitting and enhanced model generalization, underscoring the potential of combining machine learning classifiers in media forensics.

Future work could focus on expanding the training dataset to include a more diverse range of deepfake examples, thereby improving the model's robustness across various deepfake modalities. Additionally, optimizing the model's architecture and leveraging efficient algorithms could facilitate real-time detection of deepfake content in live media streams. Evaluating and enhancing the model's performance across different datasets would ensure consistent accuracy and reliability in diverse scenarios. Furthermore, developing methods to interpret and explain the model's decisions would foster trust and facilitate adoption in critical applications. By addressing these areas, the system's applicability and effectiveness in combating the evolving challenges posed by deepfake technologies can be significantly enhanced.

Moreover, exploring the integration of additional machine learning techniques, such as convolutional neural networks (CNNs) combined with vision transformers, could further improve detection capabilities by capturing intricate visual features. Implementing ensemble learning approaches that combine multiple classifiers may also contribute to higher accuracy and robustness in identifying deepfake content. Additionally, incorporating bio-inspired optimization algorithms, like particle swarm optimization (PSO), to fine-tune model parameters could enhance performance and adaptability. Addressing these aspects will contribute to the development of a more comprehensive and resilient deepfake detection system.

V. References

[1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative Adversarial Nets,” in Advances in Neural Information Processing Systems, vol. 27, Montreal, CANADA, 2014.

[2] D. P. Kingma and M.Welling, “Auto-Encoding Variational Bayes,” 2014.

[3] T. Karras, T. Aila, S. Laine, and J. Lehtinen, “Progressive Growing of GANs for Improved Quality, Stability, and Variation,” in International Conference on Learning Representations, Vancouver, Canada, 2018.

[4] Q. Duan and L. Zhang, “Look More Into Occlusion: Realistic Face Frontalization and Recognition With BoostGAN,” IEEE Transactions on Neural Networks and Learning Systems, vol. 32, no. 1, pp. 214–228, 2021.

[5] “deepfake,” http://www.github.com/deepfakes/ Accessed September 18, 2019.

[6] “fakeapp,” http://www.fakeapp.com/ Accessed February 20, 2020.

[7] “faceswap,” http://www.github.com/MarekKowalski/ Accessed September 30, 2019.

[8] F. Matern, C. Riess, and M. Stamminger, “Exploiting Visual Artifacts to Expose Deepfakes and Face Manipulations,” in IEEE Winter Applications of Computer Vision Workshops, Waikoloa, USA, 2019, pp. 83–92.

[9] D. Afchar, V. Nozick, J. Yamagishi, and I. Echizen, “Mesonet: a Compact Facial Video Forgery Detection Network,” in IEEE International Workshop on Information Forensics and Security, Hong Kong, China, 2018, pp. 1–7.

[10] X. Yang, Y. Li, H. Qi, and S. Lyu, “Exposing GAN-Synthesized Faces Using Landmark Locations,” in Proceedings of the ACM Workshop on Information Hiding and Multimedia Security, Paris, France, 2019, p. 113–118.

[11] D.-T. Dang-Nguyen, G. Boato, and F. G. De Natale, “Discrimination between computer generated and natural human faces based on asymmetry information,” in Proceedings of the 20th European Signal Processing Conference, Bucharest, Romania, 2012, pp. 1234–1238.

[12] E. Sabir, J. Cheng, A. Jaiswal, W. AbdAlmageed, I. Masi, and P. Natarajan, “Recurrent Convolutional Strategies for Face Manipulation Detection in Videos,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Los Angeles, USA, June 2019.

[13] P. Zhou, X. Han, V. I. Morariu, and L. S. Davis, “Two-Stream Neural Networks for Tampered Face Detection,” in IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, USA, 2017, pp. 1831–1839.

[14] B. Bayar and M. C. Stamm, “A Deep Learning Approach to Universal Image Manipulation Detection Using a New Convolutional Layer,” in Proceedings of the 4th ACM Workshop on Information Hiding and Multimedia Security, Vigo, Spain, 2016, pp. 5–10.

[15] U. A. Ciftci, I. Demir, and L. Yin, “FakeCatcher: Detection of Synthetic Portrait Videos using Biological Signals,” IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, doi:10.1109/TPAMI.2020.3009287.

[16] M. Li, B. Liu, Y. Hu, and Y. Wang, “Exposing Deepfake Videos by Tracking Eye Movements,” in 25th International Conference on Pattern Recognition, Milan, Italy, 2021, pp. 5184–5189.

17] Y. Li, M.-C. Chang, and S. Lyu, “In Ictu Oculi: Exposing AI Created Fake Videos by Detecting Eye Blinking,” in IEEE International Workshop on Information Forensics and Security, Hong Kong, China, 2018, pp. 1–7.

[18] C.-Z. Yang, J. Ma, S. Wang, and A. W.-C. Liew, “Preventing Deepfake Attacks on Speaker Authentication by Dynamic Lip Movement Analysis,” IEEE Transactions on Information Forensics and Security, vol. 16, pp. 1841–1854, 2021, doi:10.1109/TIFS.2020.3045937.

[19] S. Fernandes, S. Raj, E. Ortiz, I. Vintila, M. Salter, G. Urosevic, and S. Jha, “Predicting Heart Rate Variations of Deepfake Videos using Neural ODE,” in IEEE/CVF International Conference on Computer Vision Workshop, Seoul, Korea (South), 2019, pp. 1721–1729..

ISSN 2321-2152
www.ijmece.com
Vol 13, Issue 2, 2025

AUTHOR'S PICS

GUIDE – J. BHAGYA LAKSHMI MCA Working as Assistant Professor in Department of CSE-(AI&DS), Eluru College Of Engineering and Technology, Eluru.

EMAIL – bhagyalakshmijakkula1@gmail.com

TEAM LEAD – V. DAIVA KRUPA B.Tech in Department of CSE- Artificial Intelligence & Data Science, Eluru College Of Engineering and Technology, Eluru.

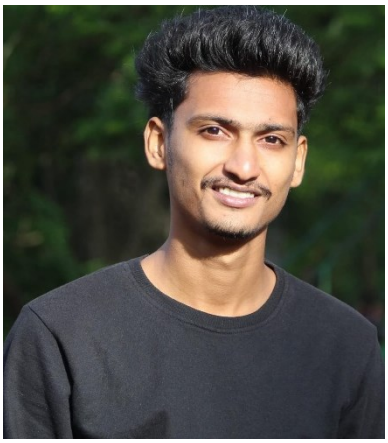
EMAIL - vadlapudidaivakrupa@gmail.com

TEAM MEMBER – P. GEETHIKA B.Tech in Department of CSE-Artificial Intelligence & Data Science, Eluru College of Engineering and Technology, Eluru.

EMAIL – geethika4143@gmail.com

TEAM MEMBER – R. BHARGAV B. Tech in Department of CSE-Artificial Intelligence & Data Science, Eluru College of Engineering and Technology, Eluru.

EMAIL - routhubhargav007@gmail.com



TEAM MEMBER – P.SAI RAJESH B. Tech in Department of CSE-Artificial Intelligence & Data Science, Eluru College of Engineering and Technology, Eluru.

EMAIL – sairajeshpaleti03@gmail.com