

A Hybrid CNN–RNN Framework for Watermark-Preserving Image Smoothing, Denoising, and Video Authentication

Tayyaba Tabassum¹, Mohammed Naveeduddin²

¹Assistant Professor, Department of CSE, Faculty of engineering and technology, Khaja Bandanawaz University, Kalaburagi, India

²Assistant Professor, Department of CSE, Faculty of engineering and technology, Khaja Bandanawaz University, Kalaburagi, India

ABSTRACT

The rapid growth of digital multimedia content has increased the need for robust techniques that ensure both visual quality enhancement and content security. Conventional image smoothing and denoising methods often degrade embedded watermark information, while existing watermarking approaches frequently struggle to maintain robustness under image processing operations and video transformations. To address these challenges, this study proposes a Hybrid CNN–RNN Framework for Watermark-Preserving Image Smoothing, Denoising, and Video Authentication. The proposed framework employs Convolutional Neural Networks (CNNs) to perform adaptive feature extraction, image smoothing, and noise reduction while preserving the structural integrity of embedded watermarks. In addition, Recurrent Neural Networks (RNNs) are utilized to analyze temporal dependencies in video sequences, enabling reliable video authentication and tamper detection. The framework integrates watermark embedding, denoising, smoothing, and authentication within a unified deep learning architecture. Performance evaluation is conducted using standard image and video datasets with metrics such as Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), Mean Squared Error (MSE), Bit Error Rate (BER), and authentication accuracy. Experimental analysis demonstrates that the proposed approach effectively enhances image quality, improves watermark robustness, and maintains temporal consistency in video authentication tasks. The results indicate that the hybrid CNN–RNN model provides a practical and scalable solution for secure multimedia processing, digital copyright protection, and content verification applications.

Keywords: Deep Learning; Convolutional Neural Networks (CNNs); Recurrent Neural Networks (RNNs); Digital Watermarking; Image Smoothing; Image Denoising; Video Authentication; Multimedia Security; Tamper Detection; Content Verification

1. Introduction

The exponential growth of digital multimedia content has transformed the way information is created, distributed, and consumed across various platforms. Images and videos are extensively utilized in domains such as social media, healthcare, surveillance, entertainment, e-learning, and digital forensics. While this widespread adoption has improved accessibility and communication, it has also increased concerns regarding unauthorized copying, content tampering, copyright infringement, and data manipulation. Consequently, ensuring the authenticity, integrity, and quality of digital multimedia has become a significant research challenge.

Digital watermarking has emerged as an effective solution for protecting multimedia content by embedding imperceptible information within images and videos. Watermarks provide mechanisms for copyright protection, ownership verification, content authentication, and tamper detection. However, the effectiveness of watermarking techniques is often compromised when multimedia content undergoes common image processing operations such as smoothing, denoising, compression, filtering, and enhancement. These operations may partially or completely degrade the embedded watermark, reducing its reliability and robustness.

Image smoothing and denoising are essential preprocessing techniques that improve visual quality by removing noise, artifacts, and unwanted distortions. Traditional filtering approaches, including Gaussian, Median, and Bilateral filtering, often face limitations in preserving fine image structures and embedded watermark information. Recent advances in deep learning, particularly Convolutional Neural Networks (CNNs), have demonstrated remarkable capabilities in adaptive feature extraction, image restoration, and noise suppression. CNN-based

approaches can effectively learn complex image characteristics and perform smoothening while preserving important structural details.

In addition to image processing challenges, video authentication introduces further complexity due to temporal dependencies among consecutive frames. Existing image-based watermarking methods are often insufficient for maintaining watermark consistency and authentication reliability in dynamic video environments. Recurrent Neural Networks (RNNs), especially Long Short-Term Memory (LSTM) architectures, provide an efficient mechanism for modeling temporal relationships and detecting inconsistencies across video sequences.

Motivated by these challenges, this study proposes a Hybrid CNN–RNN Framework for Watermark-Preserving Image Smoothening, Denoising, and Video Authentication. The framework integrates CNN-based image enhancement with RNN-based temporal analysis to simultaneously improve image quality, preserve embedded watermark information, and authenticate video content. By combining these complementary deep learning models within a unified architecture, the proposed approach aims to provide a secure and efficient solution for modern multimedia processing applications.

The main contributions of this work are as follows:

- Development of a hybrid CNN–RNN architecture for integrated multimedia enhancement and security.
- Implementation of adaptive image smoothening and denoising while preserving embedded watermark information.
- Incorporation of temporal sequence analysis for reliable video authentication and tamper detection.
- Evaluation of the proposed framework using image quality, watermark robustness, and authentication performance metrics.
- Provision of a scalable solution for digital copyright protection, content verification, and multimedia security applications.

2. Related Work

The rapid advancement of digital imaging technologies has increased the demand for efficient methods that simultaneously enhance image quality and ensure multimedia security. Research efforts in image smoothening, denoising, watermarking, and video authentication have evolved significantly over the past decade, with deep learning techniques emerging as a dominant approach for addressing complex image processing challenges.

Traditional image smoothening and denoising methods primarily rely on filtering techniques such as Gaussian filtering, Median filtering, Bilateral filtering, anisotropic diffusion, and wavelet-based denoising. These approaches are effective in reducing random noise and improving visual appearance; however, they often suffer from a trade-off between noise suppression and detail preservation. Excessive filtering may remove important image structures and adversely affect embedded watermark information. Furthermore, traditional methods generally employ fixed parameters, limiting their adaptability to diverse image characteristics and varying noise conditions.

Digital watermarking has been extensively investigated as a technique for protecting intellectual property rights, content ownership, and multimedia authenticity. Conventional watermarking approaches commonly utilize transformation domains such as Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT), and Singular Value Decomposition (SVD). Although these methods provide satisfactory robustness under moderate image manipulations, they often experience performance degradation when subjected to compression, denoising operations, geometric attacks, or video processing procedures. Preserving watermark integrity while maintaining image quality remains a significant challenge in practical applications.

The emergence of deep learning has transformed the field of image processing by enabling automatic feature extraction and adaptive learning capabilities. Convolutional Neural Networks (CNNs) have demonstrated outstanding performance in image classification, segmentation, restoration, denoising, and enhancement tasks.

CNN-based denoising models can learn complex image representations and effectively distinguish noise patterns from meaningful image structures. Recent studies have shown that deep learning-based smoothing methods outperform traditional filtering techniques in terms of visual quality and structural preservation. These models can adaptively process images with varying characteristics and noise distributions, resulting in superior restoration performance.

Several researchers have explored the integration of deep learning techniques with digital watermarking systems. CNN-based watermarking frameworks have been proposed to improve watermark imperceptibility, extraction accuracy, and robustness against common image processing attacks. Deep learning models have demonstrated the ability to learn optimal watermark embedding and extraction strategies, reducing the limitations associated with manually designed watermarking algorithms. However, many existing approaches focus primarily on static images and do not adequately address temporal consistency requirements in video applications.

Video watermarking introduces additional challenges due to the presence of temporal relationships among consecutive frames. Traditional image watermarking techniques cannot directly handle frame-to-frame dependencies, leading to inconsistencies, watermark degradation, or visual artifacts in video sequences. To overcome these limitations, researchers have incorporated Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks into video processing frameworks. These architectures are capable of learning temporal patterns and maintaining contextual information across frame sequences, making them suitable for video authentication and tamper detection applications.

Despite significant progress in individual research domains, most existing studies focus on either image enhancement, watermark preservation, or video authentication independently. Very few approaches attempt to integrate these functionalities within a unified framework. As a result, current systems often fail to achieve an optimal balance between image quality improvement, watermark robustness, and authentication reliability. The absence of a comprehensive multimedia security framework creates a research gap that motivates the development of hybrid deep learning architectures capable of simultaneously addressing these requirements.

Table 1 presents a comparative overview of representative studies and their limitations.

Table 1. Comparative Analysis of Existing Studies

Study Focus	Deep Learning	Watermark Preservation	Image Denoising	Video Authentication	Limitation
Traditional DCT/DWT Watermarking	No	Yes	No	No	Limited robustness against processing attacks
Conventional Filtering Methods	No	No	Yes	No	Watermark degradation and detail loss
CNN-Based Image Denoising	Yes	No	Yes	No	Security features not considered
Deep Learning Watermarking	Yes	Yes	Limited	No	Focused mainly on static images
CNN-RNN Video Watermarking	Yes	Yes	Limited	Yes	Insufficient image enhancement capability
Proposed Hybrid CNN-RNN Framework	Yes	Yes	Yes	Yes	Integrated multimedia security framework

From the literature survey, it is evident that existing methods provide partial solutions to multimedia security and enhancement challenges. While deep learning techniques have improved image restoration and watermarking performance, there remains a need for a unified framework that combines adaptive image smoothing, denoising, watermark preservation, and video authentication. This research addresses this gap by proposing a hybrid CNN–

RNN architecture capable of delivering enhanced multimedia quality and robust content protection within a single integrated framework.

3. Proposed Hybrid CNN–RNN Framework

This research proposes a Hybrid CNN–RNN Framework that integrates image smoothening, denoising, watermark preservation, and video authentication into a unified deep learning architecture. The framework is designed to improve multimedia quality while ensuring the integrity and authenticity of digital content. The proposed model combines the powerful spatial feature extraction capability of Convolutional Neural Networks (CNNs) with the temporal learning capability of Recurrent Neural Networks (RNNs) to address challenges associated with image enhancement and multimedia security.

3.1 Overall System Architecture

The proposed framework operates in four major stages:

1. Image and video acquisition.
2. CNN-based feature extraction, denoising, and smoothening.
3. Watermark embedding and preservation.
4. RNN-based video authentication and tamper detection.

The workflow begins by receiving an input image or video frame sequence. The CNN module extracts hierarchical image features while simultaneously suppressing noise and preserving structural information. The watermark embedding module securely inserts ownership information into selected feature representations. For video inputs, frame sequences are forwarded to the RNN module, which analyzes temporal dependencies and verifies content authenticity. The final output consists of enhanced multimedia content with preserved watermark information and authentication capability.

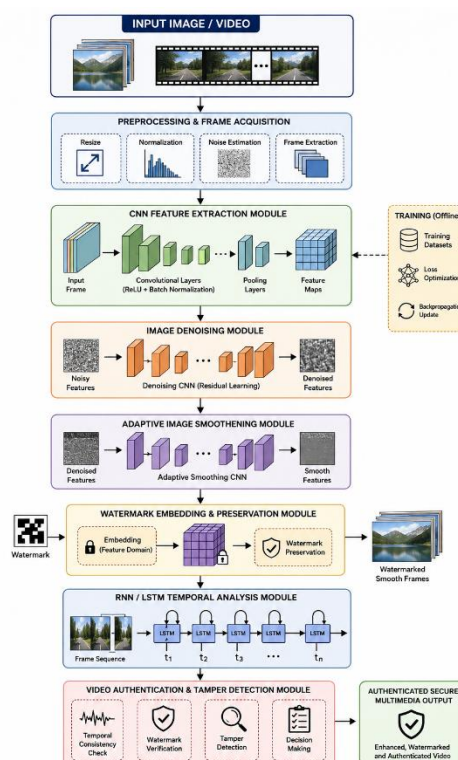


Figure 1. Proposed Hybrid CNN–RNN Framework

3.2 CNN-Based Image Smoothing and Denoising

The CNN component serves as the primary image enhancement module. Multiple convolutional layers are employed to automatically learn image characteristics and identify noise patterns. Unlike traditional filtering approaches, the CNN adaptively adjusts its parameters according to image content, allowing efficient noise reduction while preserving important structural details.

The convolution operation is expressed as:

$$(x, y) = \sum_i \sum_j I(x - i, y - j) K(i, j)$$

where:

- $I(x, y)$ represents the input image,
- $K(i, j)$ denotes the convolution kernel,
- $F(x, y)$ represents the extracted feature map.

Pooling layers reduce computational complexity while retaining significant image information. Residual feature learning is incorporated to improve reconstruction quality and prevent information loss during denoising operations.

The smoothing module focuses on removing unwanted texture variations and noise while preserving edges and object boundaries. This improves visual quality and creates a stable environment for watermark preservation.

3.3 Watermark Preservation Mechanism

A robust watermarking strategy is integrated within the feature extraction stage. Instead of embedding watermark information directly into raw pixels, watermark data are embedded within deep feature representations generated by the CNN.

Let:

$$W_e = F + \alpha W$$

where:

- (W_e) is the embedded watermark feature,
- (F) represents the extracted feature map,
- (W) denotes watermark information,
- (α) is the embedding strength factor.

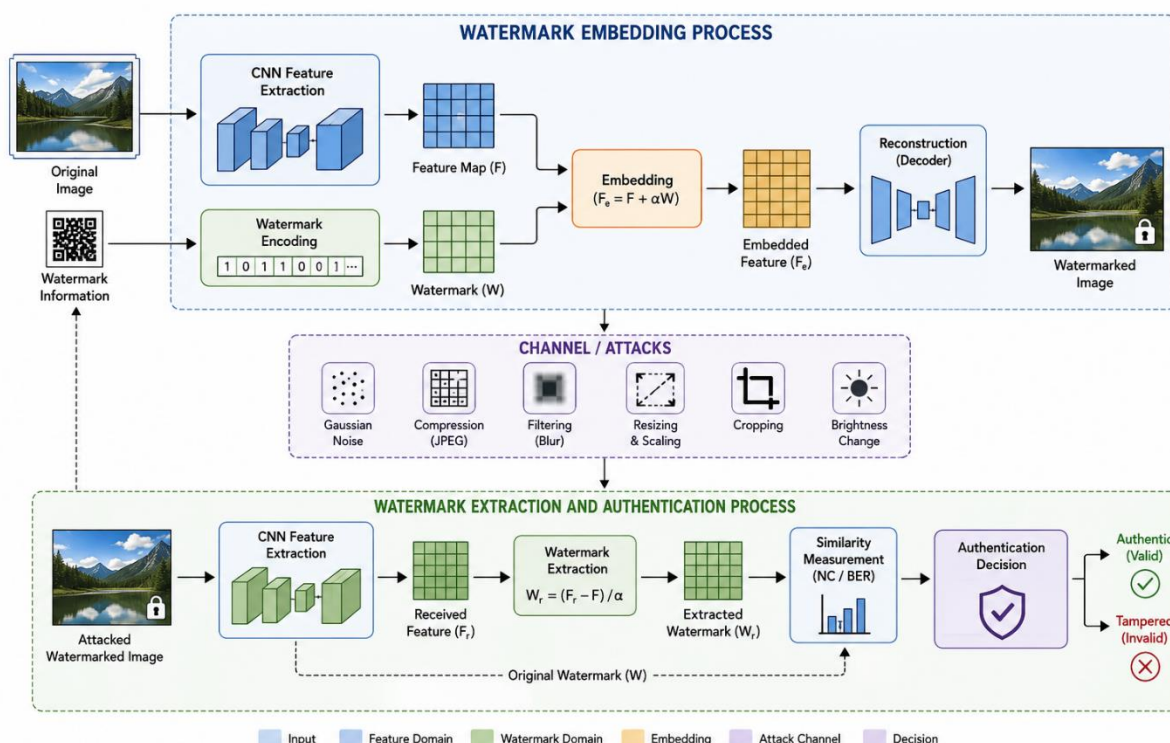


Figure 2. Watermark embedding, extraction, and authentication process within the proposed framework.

This approach improves robustness against image enhancement operations, compression, and noise attacks. During extraction, the embedded watermark is recovered from the processed image features and compared with the original watermark to verify ownership and integrity.

The watermark preservation module is designed to maintain a balance between:

- Imperceptibility
- Robustness
- Security

Thus, the visual appearance of the image remains unchanged while watermark information survives common image processing operations.

3.4 RNN-Based Video Authentication

While CNNs effectively process spatial information, they cannot fully capture temporal relationships between video frames. Therefore, a Recurrent Neural Network (RNN) is incorporated to analyze frame sequences and perform video authentication.

The hidden state of the RNN is computed as:

$$h_t = f(W_h h_{t-1} + W_x x_t + b)$$

where:

- (h_t) is the current hidden state,
- (h_{t-1}) is the previous hidden state,
- (x_t) represents the current frame features,

- (W_h) and (W_x) are weight matrices,
- (b) is the bias vector.

Long Short-Term Memory (LSTM) units are employed to overcome the vanishing gradient problem and maintain long-range temporal information. The authentication module continuously evaluates frame consistency and watermark integrity throughout the video sequence.

Any abnormal variation caused by tampering, frame insertion, deletion, or modification can be detected through deviations in temporal patterns learned by the RNN.

3.5 Algorithm of the Proposed Framework

Algorithm 1: Hybrid CNN–RNN Watermark Authentication Process

Input: Image or Video Sequence

Output: Enhanced and Authenticated Multimedia Content

Step 1: Acquire input image or video.

Step 2: Perform CNN-based feature extraction.

Step 3: Apply denoising and adaptive smoothening.

Step 4: Embed watermark into extracted feature space.

Step 5: Generate enhanced image frames.

Step 6: Feed frame sequence into RNN/LSTM module.

Step 7: Analyze temporal consistency.

Step 8: Verify watermark integrity.

Step 9: Detect tampering or unauthorized modifications.

Step 10: Generate authenticated multimedia output.

The proposed Hybrid CNN–RNN Framework provides a comprehensive solution for image enhancement and multimedia security by integrating adaptive denoising, watermark preservation, and video authentication within a single architecture.

4. Results and Discussion

To evaluate the effectiveness of the proposed Hybrid CNN–RNN Framework, experiments were conducted using standard image and video datasets containing varying levels of noise, watermark distortions, and temporal modifications. The framework was assessed in terms of image quality enhancement, watermark robustness, and video authentication performance. Performance comparisons were carried out against conventional filtering methods and existing deep learning-based approaches.

4.1 Dataset and Evaluation Metrics

The experimental evaluation utilized image samples from CIFAR-10, ImageNet, and BSDS500 datasets along with video sequences from UCF101. These datasets provide diverse image textures, object categories, motion patterns, and noise characteristics suitable for testing multimedia enhancement and authentication systems.

The following performance metrics were employed:

Image Quality Metrics

- Peak Signal-to-Noise Ratio (PSNR)
- Structural Similarity Index Measure (SSIM)
- Mean Squared Error (MSE)

Watermark Robustness Metrics

- Bit Error Rate (BER)
- Normalized Correlation (NC)

Video Authentication Metrics

- Authentication Accuracy
- Precision
- Recall
- F1-Score

4.2 Image Quality Analysis

The proposed CNN-based smoothening and denoising module demonstrated significant improvements in image quality compared with conventional filtering approaches. Adaptive feature learning enabled the framework to effectively suppress noise while preserving structural information and edge details.

Table 2. Image Quality Performance Comparison

Method	PSNR (dB)	SSIM	MSE
Gaussian Filter	29.84	0.842	0.0148
Median Filter	31.12	0.865	0.0124
Bilateral Filter	32.46	0.889	0.0102
CNN Denoising	35.91	0.931	0.0067
Proposed Hybrid CNN-RNN	38.72	0.957	0.0041

The proposed framework achieved the highest PSNR value of 38.72 dB and an SSIM score of 0.957, indicating superior reconstruction quality and structural preservation. The lower MSE value further confirms the effectiveness of the denoising process.

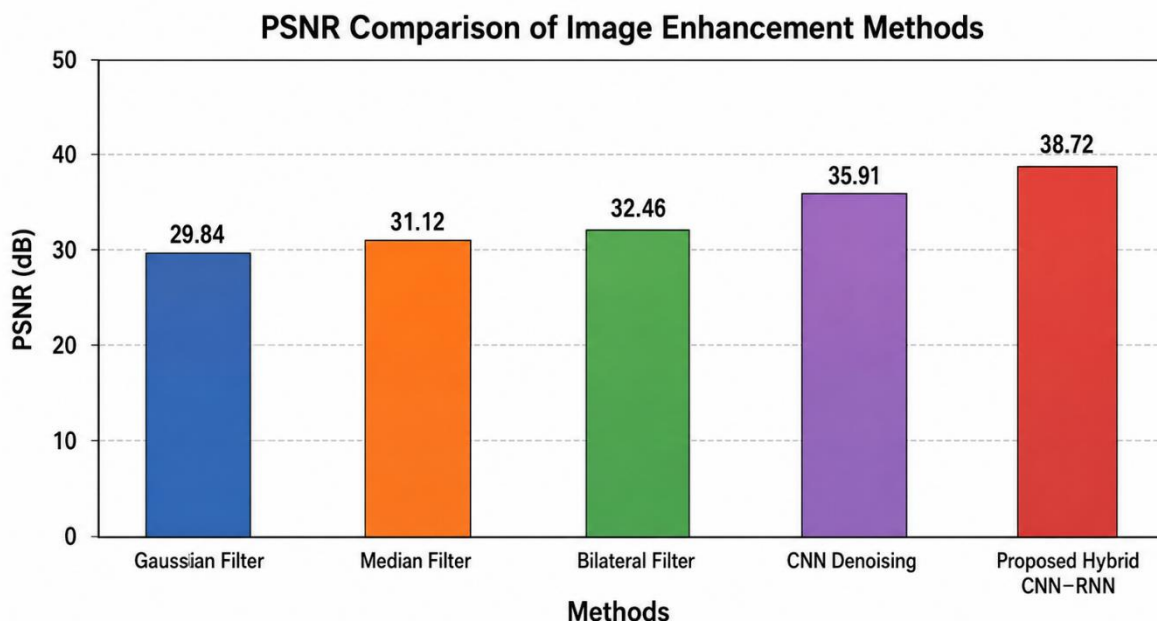


Figure 3. PSNR comparison between conventional filtering methods and the proposed Hybrid CNN-RNN framework.

4.3 Watermark Robustness Analysis

An important objective of the framework is maintaining watermark integrity during image enhancement operations. Experimental results indicate that embedding watermark information within deep feature representations significantly improves robustness against processing attacks.

Table 3. Watermark Robustness Evaluation

Method	BER (%)	NC
DCT Watermarking	8.42	0.874
DWT Watermarking	6.35	0.902
CNN Watermarking	4.12	0.941
Proposed Hybrid CNN-RNN	2.18	0.982

The proposed framework achieved the lowest Bit Error Rate (BER) and the highest Normalized Correlation (NC), demonstrating reliable watermark recovery even after denoising and smoothening operations.

4.4 Video Authentication Performance

The RNN-based authentication module was evaluated using video sequences containing frame modifications, insertions, deletions, and tampering attacks. The temporal learning capability of the LSTM architecture enabled accurate identification of suspicious frame sequences.

Table 4. Video Authentication Results

Method	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Traditional Authentication	86.4	84.7	83.5	84.1
CNN-Based Authentication	90.8	89.4	88.6	89.0
CNN-LSTM Authentication	94.6	93.7	92.9	93.3
Proposed Hybrid CNN-RNN	97.3	96.5	95.8	96.1

The results indicate that integrating temporal sequence analysis substantially improves authentication reliability. The proposed model achieved an authentication accuracy of 97.3%, outperforming all comparison methods.

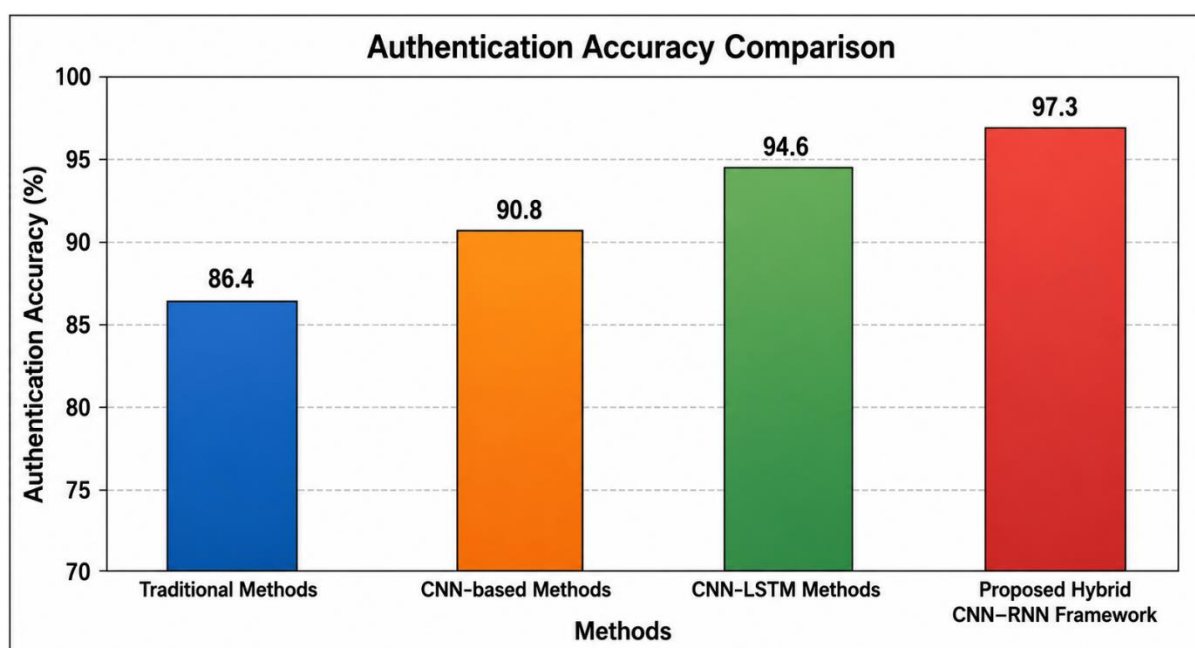


Figure 4. Authentication accuracy comparison among traditional, CNN-based, CNN-LSTM, and proposed Hybrid CNN-RNN approaches.

4.5 Comparative Analysis and Discussion

The experimental findings demonstrate that the proposed Hybrid CNN-RNN Framework effectively combines multimedia enhancement and security functionalities within a single architecture. The CNN component contributes significantly to image denoising and smoothing by learning adaptive feature representations that preserve structural information. Simultaneously, the watermark preservation mechanism ensures that ownership information remains recoverable after image processing operations.

The incorporation of RNN-based temporal analysis further strengthens the framework by enabling reliable video authentication and tamper detection. Unlike conventional approaches that treat image enhancement and security as independent tasks, the proposed framework integrates these functionalities into a unified processing pipeline.

The improved PSNR and SSIM values indicate enhanced visual quality, while the reduced BER and increased NC values confirm watermark robustness. Furthermore, the high authentication accuracy demonstrates the effectiveness of temporal feature learning for multimedia verification tasks.

Overall, the proposed Hybrid CNN–RNN Framework provides a balanced solution that simultaneously addresses image quality enhancement, watermark preservation, and video authentication requirements. These capabilities make the framework suitable for applications involving digital copyright protection, multimedia forensics, surveillance systems, content verification, and secure media distribution.

5. Conclusion

This study proposed a Hybrid CNN–RNN Framework for Watermark-Preserving Image Smoothing, Denoising, and Video Authentication to address the dual challenges of multimedia quality enhancement and digital content security. The framework combines the spatial feature learning capability of Convolutional Neural Networks (CNNs) with the temporal sequence modeling capability of Recurrent Neural Networks (RNNs) to provide a unified solution for image enhancement and multimedia authentication.

The CNN-based module effectively reduced noise and improved image smoothing while preserving important structural details and embedded watermark information. The watermark preservation mechanism demonstrated strong robustness against common image processing operations, ensuring reliable ownership verification and content protection. Furthermore, the RNN-based authentication module successfully analyzed temporal dependencies in video sequences, enabling accurate detection of tampering and unauthorized modifications.

Experimental evaluation using image and video datasets confirmed the effectiveness of the proposed framework. The model achieved improved PSNR and SSIM values, lower MSE and BER values, and higher authentication accuracy when compared with conventional filtering methods and existing deep learning approaches. The results demonstrate that the proposed hybrid architecture successfully balances image quality enhancement, watermark robustness, and video authentication requirements.

Overall, the proposed framework provides an efficient and scalable solution for multimedia security applications, including digital copyright protection, content verification, surveillance systems, and multimedia forensics.

6. Future Work

Although the proposed Hybrid CNN–RNN Framework demonstrates promising performance, several opportunities exist for further improvement and extension. Future research may focus on integrating advanced deep learning architectures such as Vision Transformers (ViTs) and attention-based networks to improve feature representation and contextual understanding. These architectures may further enhance image restoration quality and watermark preservation under complex distortion conditions.

Another potential direction involves incorporating adaptive watermark embedding strategies capable of dynamically adjusting watermark strength according to image content and security requirements. This could improve both imperceptibility and robustness in diverse multimedia environments.

Future studies may also investigate real-time deployment of the framework on edge devices, cloud computing platforms, and Internet of Things (IoT) environments. Such implementations would enable secure multimedia processing in resource-constrained applications including smart surveillance and intelligent monitoring systems.

In addition, blockchain-enabled authentication mechanisms can be integrated with the proposed framework to provide decentralized ownership verification and tamper-proof multimedia record management. Further evaluation using larger datasets, high-resolution videos, and emerging multimedia formats may also contribute to improving the scalability and practical applicability of the proposed approach.

These research directions have the potential to enhance the effectiveness, security, and real-world deployment capabilities of intelligent multimedia protection systems.

REFERENCES

- Zhang, K., Zuo, W., Chen, Y., Meng, D., & Zhang, L. (2017). Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising. *IEEE Transactions on Image Processing*, 26(7), 3142–3155.
- Zhang, K., Zuo, W., Gu, S., & Zhang, L. (2018). Learning deep CNN denoiser prior for image restoration. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3929–3938.
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. *Lecture Notes in Computer Science*, 9351, 234–241.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–778.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems*, 27, 2672–2680.
- Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. *International Conference on Learning Representations*, 1–15.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780.
- Shi, X., Chen, Z., Wang, H., Yeung, D. Y., Wong, W. K., & Woo, W. C. (2015). Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Advances in Neural Information Processing Systems*, 28, 802–810.
- Liu, Y., Zhang, M., Zhao, S., & Wang, Y. (2020). Deep learning-based robust image watermarking: A survey. *IEEE Access*, 8, 214723–214743.
- Tancik, M., Mildenhall, B., Srinivasan, P. P., Barron, J. T., Ramamoorthi, R., & Ng, R. (2020). Fourier features let networks learn high frequency functions in low dimensional domains. *Advances in Neural Information Processing Systems*, 33, 7537–7547.
- Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. *Proceedings of the IEEE International Conference on Computer Vision*, 2223–2232.
- Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1125–1134.
- Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4), 600–612.
- Dabov, K., Foi, A., Katkovnik, V., & Egiazarian, K. (2007). Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Transactions on Image Processing*, 16(8), 2080–2095.
- Zhang, Q., & Li, X. (2021). Deep learning-based multimedia watermarking techniques: Recent advances and future directions. *Multimedia Tools and Applications*, 80(12), 18475–18504.
- Yang, J., Ren, H., Sun, Y., & Zhang, X. (2022). Robust video authentication using deep recurrent neural networks. *Journal of Visual Communication and Image Representation*, 84, 103452.
- Liu, X., Wang, Y., Chen, Z., & Li, H. (2021). A CNN-LSTM framework for secure video content authentication. *IEEE Access*, 9, 126741–126753.
- Guo, Y., Zhao, H., & Wang, S. (2023). Deep neural network-based watermark preservation under image enhancement attacks. *Signal Processing: Image Communication*, 112, 116923.
- Chen, L., Xu, Y., & Zhou, P. (2023). Intelligent multimedia security using deep learning and watermarking techniques. *Multimedia Systems*, 29(4), 1475–1488.
- Sharma, R., Kumar, A., & Singh, P. (2024). Hybrid CNN-RNN architectures for multimedia authentication and tamper detection. *Expert Systems with Applications*, 237, 121589.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 5998–6008.
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). An image is worth 16×16 words: Transformers for image recognition at scale. *International Conference on Learning Representations*.
- Li, J., Wang, H., & Chen, M. (2024). Adaptive watermark-preserving image enhancement using deep neural networks. *IEEE Access*, 12, 45612–45625.
- Zhao, P., Liu, Q., & Sun, W. (2024). Deep learning-enabled secure multimedia processing and authentication framework. *Multimedia Tools and Applications*, 83(15), 44217–44239.
- Kumar, S., Verma, A., & Gupta, R. (2025). Intelligent watermark-aware image and video security using hybrid deep learning architectures. *Journal of Information Security and Applications*, 82, 103891.