Baghdad Science Journal

Volume 23 | Issue 4 Article 14

4-23-2026

# A New Reinforcement Learning Agent Framework for Digital Image Forgery Detection Using Prioritized Experience Replay

Muthana S. Mahdi
*Department of Computer Science, College of Science, Mustansiriyah University, Baghdad, Iraq*,
muthanasalih@uomustansiriyah.edu.iq

Saad N. Alsaad
*Department of Computer Science, College of Science, Mustansiriyah University, Baghdad, Iraq*,
dr.alsaadcs@uomustansiriyah.edu.iq

Hasanen S. Abdullah
*Department of Computer Science, University of Technology, Baghdad, Iraq*,
Hasanen.S.Abdullah@uotechnology.edu.iq



## How to Cite this Article

Mahdi, Muthana S.; Alsaad, Saad N.; and Abdullah, Hasanen S. (2026) "A New Reinforcement Learning Agent Framework for Digital Image Forgery Detection Using Prioritized Experience Replay," *Baghdad Science Journal*: Vol. 23: Iss. 4, Article 14.
DOI: https://doi.org/10.21123/2411-7986.5270

BAGHDAD SCIENCE JOURNAL 2026;23(4):1316–1332



RESEARCH ARTICLE

# A New Reinforcement Learning Agent Framework for Digital Image Forgery Detection Using Prioritized Experience Replay

**Muthana S. Mahdi** [1,*], **Saad N. Alsaad** [1], **Hasanen S. Abdullah** [2]

[1] Department of Computer Science, College of Science, Mustansiriyah University, Baghdad, Iraq
[2] Department of Computer Science, University of Technology, Baghdad, Iraq

## ABSTRACT

Digital image forgery detection has become an urgent and complex problem in an age when powerful editing tools can easily alter photographs. The familiar maxim "a picture is worth a thousand words" can no longer be taken at face value, since even subtle manipulations may conceal or fabricate critical details. Conventional detection methods frequently suffer from highly imbalanced datasets and narrow feature representations that fail to capture the diverse artifacts introduced by modern editing techniques. In response, this work presents a novel framework that casts the forgery detection task as a reinforcement learning problem, enabling an agent to learn a sequence of analysis steps rather than relying on a fixed classification pipeline. This is the first study to deploy a Deep Q-Network (DQN) agent specifically for image forgery detection. The proposed architecture features a dual-branch feature extractor: one branch applies Error Level Analysis to reveal compression inconsistencies. At the same time, the other performs fine-grained noise analysis to detect manipulation traces. These features are supplied to a custom Gym environment designed with balanced sampling to mitigate class imbalance and augmented by a prioritized experience replay mechanism that biases learning toward the agent's most challenging examples. Extensive experiments across multiple datasets, including testing on the latest image manipulation model, Gemini 2.0 Flash, show impressive accuracy rates between 97% and 98%. These results highlight the robustness and efficiency of the proposed approach, offering a promising new direction for digital forensic investigations.

**Keywords:** Balanced gym environment, Deep q-network, Deepfake, Dual-branch feature extraction, Prioritized experience replay, Reinforcement learning agent, Splice image forgery detection

## Introduction

Digital authenticity is one of the central issues due to recent rapid technological progress in the field of image manipulation, such as deepfakes and advanced editing programs. The current technologies now exploit social media networks to disseminate persuasive, false information, decreasing the credibility that people have in the digital sources. [1] Forensic experts, law enforcement agencies, and media organizations need image authenticity identification to determine scientifically if an image has been tampered with. This is increasingly crucial in maintaining visual evidence integrity. [2] Digital image splicing is one of the most popular tampering methods that combines segments cut from numerous images to generate false composites. [3] The modifications during manipulations affect fundamental documents, including face

Received 28 April 2025; revised 23 August 2025; accepted 5 September 2025.
Available online 23 April 2026

* Corresponding author.
E-mail addresses: muthanasalih@uomustansiriyah.edu.iq (M. S. Mahdi), dr.alsaadcs@uomustansiriyah.edu.iq (S. N. Alsaad), Hasanen.S.Abdullah@uotechnology.edu.iq (H. S. Abdullah).

https://doi.org/10.21123/2411-7986.5270

billing statements and official paperwork, making digital verification and trust increasingly difficult to establish.

Multiple tamper detection systems were developed through recent attempts to tackle this problem.[4] Traditional forensic investigation depends on detecting irregularities between forensic traces that include freak edges, double JPEG compression effects, inconsistent light elements, noise patterns, and camera processing inconsistencies.[5] The current detection methods function properly for common cases, yet struggle to detect the complex modern forgeries that have been developed by sophisticated techniques.[6] Digital image forensics problems are mainly solved through deep learning-based approaches that utilize convolutional neural networks (CNNs).[7]

These methods, composed of multiple neural networks, perform image preprocessing as well as feature extraction before classification through stochastic gradient descent optimization.[8] However, existing databases show a fundamental flaw because they contain imbalanced sets of authentic and tampered images, which leads to biased models that fail to perform correctly in real-world use.[9]

This paper proposes a new reinforcement learning framework for splice image forgery detection, presenting several notable contributions. Firstly, the study pioneers the formulation of forgery detection as a reinforcement learning problem, employing a Deep Q-network (DQN) agent for the first time in this domain. Such a new method changes the traditional classification problem to a dynamic decision-making problem, making adaptive learning possible even in the face of complicated image manipulations.

Secondly, a balanced Gym-based setting is presented to overcome the widespread problem of dataset imbalance since authentic and forged images are equally represented during training. Thirdly, a prioritized experience replays mechanism sampling 80 percent of misclassified and 20 percent of correctly classified cases directs the learning process on the difficult cases, reducing overfitting, and speeding up convergence. It suggests a dual-branch feature extraction architecture, which consists of high-level semantic features produced by the Error Level Analysis on a pre-trained ResNet50 and low-level noise patterns produced by a custom CNN. This complementary combination increases the model's robustness and accuracy with a wide range of datasets.

## Related work

Detection of image forgery has improved tremendously in recent years, especially in the problem of image splicing. This section is a discussion of a group of related papers.

In 2020, a noise level-based method of detection was developed by Alshwely and Alsaad.[10] The quantity results were good using the detection method. It did not test it on different data sets, however, and this led to performance issues of skewed data and limited applicability. In 2020, too, Hussien et al.[11] proposed an automated method of forgery detection, which relies on Colour Filter Array (CFA) artifacts. This approach was a novel approach based on CFA features. However, it has challenges in adjusting to the complexities of image manipulation and results in a weakness in feature representation. The same year, Islam et al.[12] came up with a methodology that was based on the fusion of Discrete Cosine Transform (DCT) and Local Binary Patterns (LBP). They also used a combination of two detection strategies, and thus, the accuracy of their findings was increased. Nevertheless, the method has difficulties with the appropriate representation of features and complicated manipulations.

Siddiqui et al.[13] suggested a technique that relies on the Discrete Wavelet Transform (DWT) and edge-weighted LBP. Their technique increased the edge detection of forged pictures. Nonetheless, it is also prone to the imbalanced representation of data and the inability to deal with complex manipulations.

Ali et al.[14] designed a deep-learning algorithm based on image recompression in 2022. The methodology enhanced the percentage of detection, but it is overfitting. Another study by Xu et al.[15] dealt with the detection of the restricted image splicing wherein a scale-adaptive deep matching network is proposed to address the issue of correctly identifying the forged regions in the suspected images. It is also limited in terms of feature representation, although it enjoys the power of statistical analysis. The approach of Baumy et al.[16] is a combination of statistical features and Principal Component Analysis (PCA) to further analyze features, but it is associated with the problem of poor computational efficiency.

In 2023, Sedeeq[17] investigated an image-splicing detection method with a discrete wavelet transform and a co-occurrence matrix. In spite of the fact that the other analytic perspective of pixel value dependencies increased the detection reliability, it faced challenges of complex manipulations. Khalil et al.[18] improved detection with the help of transfer learning. They are computationally efficient and are not very good in terms of dataset imbalance.

Ramirez-Rodriguez et al.[19] introduced a Siamese network-based system that identifies splice manipulations by recognizing patterns and solves the issues of certain image editing applications despite the feature

**Table 1.** The summary of the related works, with the strengths and weaknesses of each.

| Ref. | Approach (Method Used) | Strength Points | Weak Points or Restrictions |
|---|---|---|---|
| [10] | Noise level-based detection | Solid quantitative performance | Spreads different network evaluation, dataset imbalance, and restricted feature representation. |
| [11] | CFA artifact analysis for forgery detection | Strong retrieval and extraction of camera-based features. | Problems of adapting to complicated manipulations; poor representation of features. |
| [12] | DCT and LBP framework for copy-move and splicing detection | Dual-technique for better detection. | Poor ability to represent features, difficulties in difficult manipulations. |
| [13] | DWT and edge-weighted LBP for image splicing detection | Increased forged region edge recognition. | Has the problem of imbalance in datasets and the inability to process complex manipulations. |
| [14] | Deep learning via image recompression | Enhanced accuracy in detection. | Prone to overfitting |
| [15] | Scale-adaptive deep matching network | A strong standard statistical analysis. | Poor in feature representation. |
| [16] | Combination of statistical features with PCA | Intensive studying of the feature space. | Issues of computational efficiency. |
| [17] | Detection technique using the discrete wavelet transform | Promoting the reliability of detection in another way of analysis. | Problems in adjusting to complicated manipulations. |
| [18] | Transfer learning for forgery detection | Computationally efficient | Problems related to dataset imbalance. |
| [19] | Siamese network-based framework | Surmounting the problems of certain image editing applications. | Limitations in Feature Representation |
| [20] | Integration of Error Level Analysis with CNN Architecture | Improved image manipulation detecting system. | Vulnerability to overfitting |
| [21] | RIFD-Net: Multi-channel CNN for robust detection | Better detection efficiency; Strong feature extraction. | Demands large computing time. |

representation weaknesses. Nagm et al.,[20] in another study, combined the Error Level Analysis with a Convolutional Neural Network architecture to improve the detection system of manipulated images, which can successfully verify authenticity with a great degree of accuracy. The model, however, is susceptible to overfitting. Lastly, Shan et al.[21] unveiled RIFD-Net, a multiplexed CNN with greater efficiency. Nonetheless, it involves high computing power.

This paper presents a splice image forgery detection technique that is a new proposal that provides a solution to the limitations identified in the related literature. First, it converts the traditional classification problem into a reinforcement learning problem, which allows a dynamic process of decision-making to be adapted to the challenges of image manipulation. The given approach directly addresses the problem of imbalance in datasets because it uses a balanced Gym environment, according to which authentic and tampered images are represented equally throughout the training process. Second, the dual-branch feature representation extractor is able to overcome limitations in feature representation with a dual-branch feature representation extractor that simultaneously acquires high-level semantic representation (by using a pre-trained ResNet50) and low-level noise patterns (through a custom CNN). This combination has a profound effect on the discovery of subtle forgery artifacts. Third, a prioritized experience replay process in which 80 percent of the training instances are sampled among challenging experiments helps in decreasing overfitting and speeds up learning by concentrating the agent on updating on challenging instances. Lastly, image processing in memory (without using temporary disk memory) enhances the efficiency of computational speed and the latency, and adds to a more hearty and scalable detection structure. The related works are summarized with their strengths and weaknesses in Table 1.

As far as we know, there is no prior effort that attempts to recast the splice image forgery detection problem as a reinforcement learning problem based on a Deep Q-Network. The currently used approaches are based on fixed-point classifiers like CNNs, Siamese networks, or manual models based on features, which are used in a fixed pipeline. On the contrary, our approach presents an interactive agent, which learns adaptive strategies of analysis by trial and error. This change of methodology indicates a fresh perspective in the domain and the possibility of reinforcement learning to deal with complicated and dynamic challenges of forgery detection.

## Proposed methodology

This work presents a novel learning-based reinforcement (RL) system that can detect digital image forgeries with great accuracy and resiliency. The suggested methodology is a synergetic combination of sophisticated methods of image preprocessing, a dual-branch deep feature extractor, and a DQN agent. It


**Fig. 1.** Main phases of the proposed method.

has an advantageous experience replay mechanism, which focuses on the learning of various experiences. The proposed method phases are shown in Fig. 1.

### *A. Environment preparation phase*

The flow of data into the reinforcement learning agent is engineered in a strictly controlled process as part of the environment preparation stage. The environment was constructed on the OpenAI Gym interface[22] to partition the dataset into two subsets, namely authentic and tampered images, in a systematically organized manner. In every training episode, 100 images, 50 genuine and 50 tampered, are randomly sampled in a balanced way in the previously created datasets. Such images are randomly matched to make sure that they are not repeated and re-indexed so that once all the available images have been used, the index is re-established. Each image is also processed by extensive preprocessing prior to being processed by the agent to increase forensic artifacts, such as error-level inconsistencies and fine noise patterns. This is then followed by the use of a reward system where positive reinforcement is given for the right classifications and negative punishment is issued for incorrect classifications. This is a powerful and dynamic structure that guarantees a balanced, effective, and replicable data pipeline, which is optimal to the learning efficacy and total digital image forgery detection performance of the agent.

#### *Image preprocessing*

There are two different pre-processing methods used. The details of the preprocessing operations are illustrated in Fig. 2.

**Error-Level Analysis (ELA):** This method is used to compress the image at a 98 percent quality to produce a compressed image. Calculation of the difference between the original and compressed images is made to reveal areas having inconsistent compression artifacts. The ELA image is then scaled to 256 × 256 pixels and made normalized to become an input to the feature extractor.

**Noise Analysis:** This is an enhanced multi-scale algorithm that transforms images into grayscale and uses three median filters (3 × 3, 5 × 5, and 7 × 7) to create three baseline pictures. Then, the difference between the original image and all the median-filtered images is calculated to get three noise residual images at various scales. These multi-scale details allow different characteristics by capturing the change of noise on the various kernel sizes, and enable the identification of subtle traces through editing and much better tolerate forgery identification accuracy. The noise images obtained are resized and normalized as in the case of the ELA image.

These are important preprocessing stages that allow the feature extractor to pick up higher-level discrepancies and lower-level noise levels that are indicative of digital forgeries.

### *B. DQN agent preparation phase*

The classification problem is formulated as a series of decision-making tasks where the RL agent interacts with the environment by classifying images. The agent is applied by the DQN algorithm, whose methodology is outlined in reference.[23] The policy of the agent is represented by a convolutional neural network that has the dual-branch feature extractor in the Dual-Branch Feature Extraction Section.

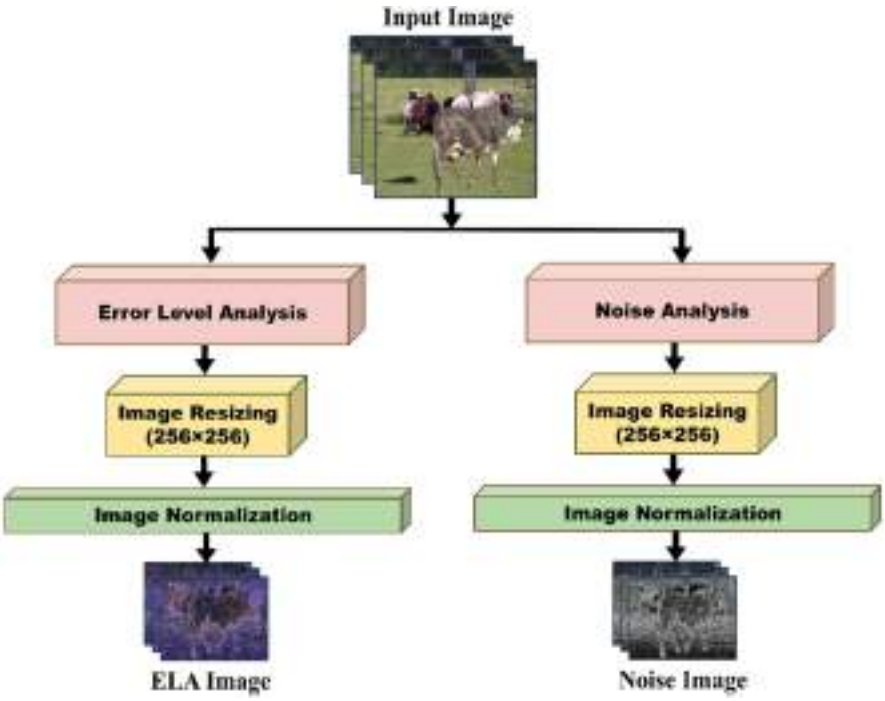


**Fig. 2.** The details of the preprocessing operations.

*Dual-branch feature extraction*

The key feature of the proposed method is the dual-branch feature extractor, which works on the image, taking two complementary paths. The two complementary branches chosen in the feature extraction module, Error Level Analysis (ELA) and noise analysis, have been inspired by their different yet complementary forensic properties. Recent investigations have indicated that ELA, in combination with deep neural networks, can be of significant benefit in the detection of compression-related anomalies in tampered images, especially those in the JPEG format.[24] ELA works since manipulated areas can be quantized by different JPEG quantization values than their immediate neighbours, so that they can be spotted in the transformed space. Instead, noise analysis has been proven to detect the cause of the irregularities of spatial noise patterns due to resampling, blending, or inpainting, even where compression artifacts are so insignificant or absent.[25] This fullness is a complement to ELA that aims at another group of forensic indicators. As forgeries in reality may have either form of trace or both, combining these two fields of study allows more forms of manipulation to be covered, enhancing strength and generalisation.

**ELA Branch:** This branch leverages a pre-trained ResNet50 network. The network is truncated by removing the final two layers, thereby retaining only the convolutional layers that capture high-level semantic features. An adaptive global average pooling layer is then applied to transform the resulting feature maps into a 64-dimensional feature vector. This branch is adept at identifying structural inconsistencies and subtle compression artifacts typical of tampered images. Fig. 3 shows the architecture of the ELA Branch network. Table 2 summarises the structural parameters and hyperparameters of the ELA-branch network.

**Noise Analysis Branch:** This branch is designed as a custom convolutional neural network (CNN) tailored to capture fine-grained noise patterns. It consists of two convolutional layers with kernel sizes of 5. The first convolutional layer outputs 128 feature maps, while the second outputs 256. Each convolution is followed by a ReLU activation and an average pooling layer to reduce spatial dimensions. After these layers, an adaptive average pooling operation is applied to generate a fixed-size feature map. The flattened output is then passed through two fully connected layers with dropout and batch normalization, yielding a 64-dimensional feature vector. Fig. 4 shows the architecture of the Noise Branch network. Table 3 summarises the structural parameters and hyperparameters of the dual-branch network.

The outputs from the two branches are concatenated, resulting in a merged feature vector of dimension 128. This combined representation is refined through several fully connected layers to obtain a compact and discriminative representation that the

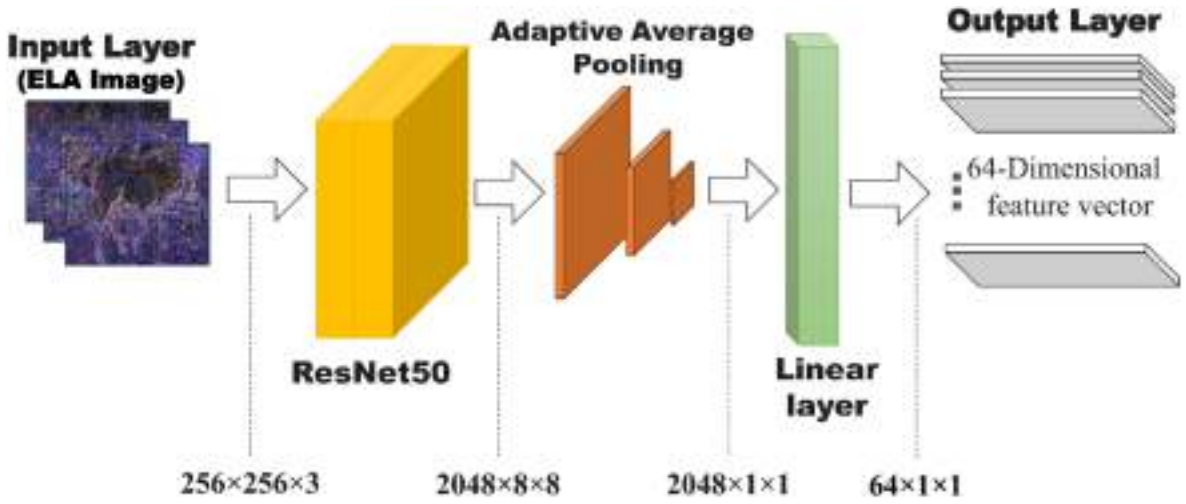


**Fig. 3.** ELA branch network architecture.

**Table 2.** Summary of network structural parameters and hyperparameters for the ELA branch.

| Parameter | Value/Configuration |
|---|---|
| Input Size | 256 × 256 × 3 |
| Base Network | Pre-trained ResNet50 |
| Retained Layers | All convolutional layers except the final two |
| Pooling Mechanism | AdaptiveAvgPool2d((1, 1)) |
| Output Dimension | 64 (after passing through the Linear layer) |

RL agent subsequently uses to make classification decisions. Fig. 5 shows the structure of the neural network after feature fusion.

### *Prioritized experience replay*

The idea of the incorporation of a prioritized experience replay buffer into the DQN framework is one of the innovative aspects of the proposed methodology. This element promotes generalization and reduces overfitting since it adds more variability to the training experiences of the agent. The replay buffer classifies experiences into two: hard experiences, where the agent made bad classifications (which yield negative rewards), and easy experiences, where the agent made correct classification choices (where this yields positive rewards). It uses a targeted sampling, which takes 80 percent of the mini-batch of the difficult cases and 20 percent of the easy cases. This strategy guarantees the agent trains on difficult samples, speeds up convergence, and also boosts the detection results to a large extent. Unlike the traditional replay buffers, which are sampled at equal intervals, the prioritized mechanism focuses on the most informative experiences. Consequently, the generalization power of the agent will increase, as well as his or her efficiency in identifying complex patterns of fraud, which will result in a higher and more efficient detection of forgery.

## *C. Learning phase*

In the mentioned model of reinforcement learning, the agent makes a step by being presented with a processed image that is the present state of the environment. The agent decides the proper course of action by using its neural network architecture, a classification that is forged or authentic. An appropriate reward is then given according to the correctness of this classification, which in effect rewards the correct decision and punishes the mistake. Reward (+1) and punishment (–1) are given as a reward and penalty, respectively. The agent seeks to achieve the sum total of reward accumulated within an episode, thus acquiring a policy that minimizes the misclassification errors. This is followed by another pre-processed image being shown to the agent, and the state of the environment is updated. This process repeats itself through a few training steps as the agent upgrades its ability to make decisions as it goes through training. This type of cyclical interaction makes the agent more efficient in learning and adds a lot of value to the overall effectiveness and accuracy of the forgery detection system.

Unlike the traditional supervised classifiers, including CNNs or Transformers, that train on uniformly sampled data to minimise a surrogate loss, the DQN agent directly maximises the operating reward signal based on making the correct decision. Such off-policy formulation permits the application of the prioritized experience replay to allow the agent to concentrate on informative and challenging samples, thus enhancing the efficiency of the samples and the convergence

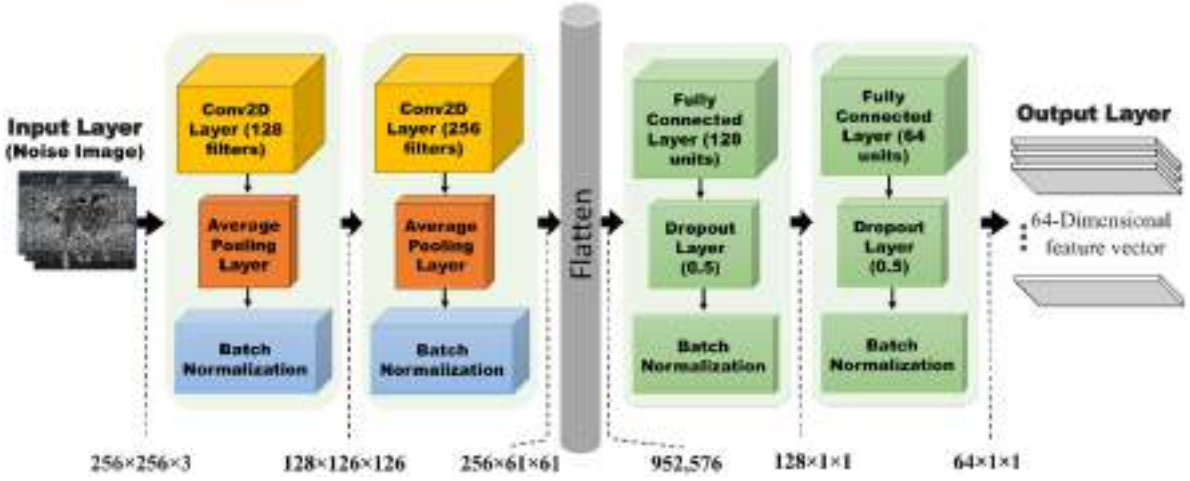


**Fig. 4.** The architecture of the noise branch network.

**Table 3.** Summary of network structural parameters and hyperparameters for the noise analysis branch.

| Parameter | Value/Configuration |
|---|---|
| Input Size | 256 × 256 × 3 |
| Convolutional Layers | 2 layers with kernel size = 5 (filters: 128 and 256, respectively), Stride is '1', and Padding is 'valid'. |
| Activation & Pooling | ReLU activation followed by Average Pooling (kernel size = 2) |
| Adaptive Pooling | AdaptiveAvgPool2d((1, 1)) |
| Dense Layers | 2 layers: 256→128 and 128→64 with dropout (0.5) and batch normalization (BN) |
| Output Dimension | 64 |

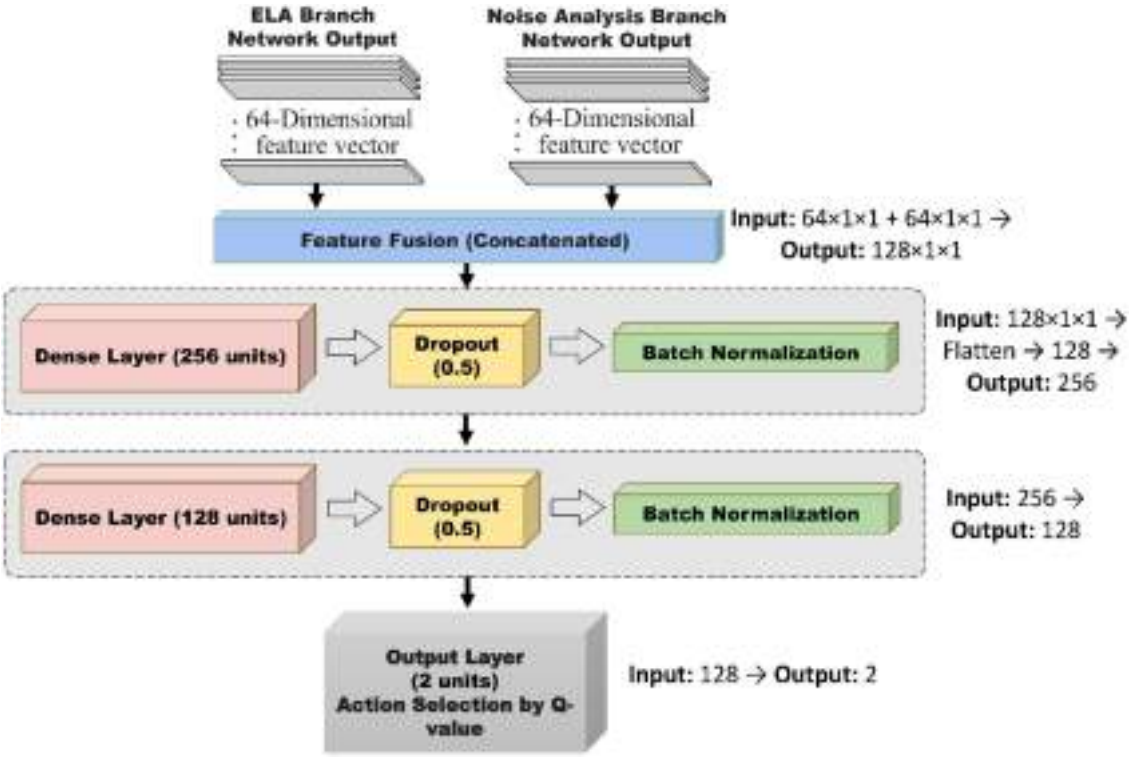


**Fig. 5.** The structure of the neural network after feature fusion.

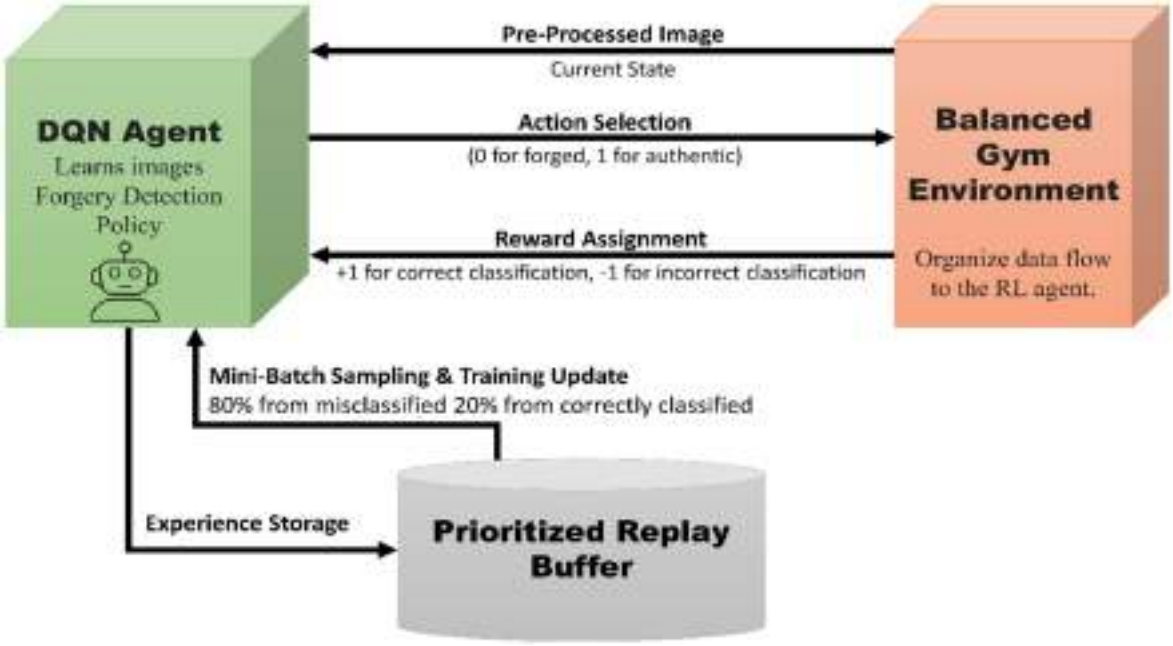


**Fig. 6.** The agent's interaction with the environment.

stability. They are especially useful in image forgery detection when the manipulations are not necessarily of the same complexity in terms of their distribution.

An important element of the training step is exploration-exploitation trade-off management. First, the agent uses a high exploration strategy with Exploration Initial Epsilon of 1.0, i.e., actions are randomly selected. Such randomness prompts the agent to search the state-action space in a broad way. The rate of exploration declines slowly with time, with an empirically defined Exploration Fraction until it finally reaches an Exploration Final Epsilon of $10^{-10}$. The slower the rate of exploration, the stronger is the tendency of the agent to draw upon the knowledge that it has acquired (exploitation) to make informed choices. Fig. 6 indicates the interaction of the agent with the environment.

A prioritized replay buffer is also used to further improve the training process, and in this case, experiences (state, action, reward, next state) are stored and sampled unevenly during the learning phase. Namely, the agent must concentrate on challenging trials and not easy cases. Thus, the buffer is geared towards sampling of 80 percent of hard experiences and 20 percent of easy experiences. This narrowing down of sampling enables the agent to quickly rectify its errors, leading to accelerated convergence and accuracy. Every single step of training, one new image is presented by the environment, 32 experiences are selected by the buffer, around 26 are erroneously classified, and 6 are correctly classified. This is a focused strategy whereby the agent strengthens the corrective behaviours but does not ignore the less complex ones of which it is already well trained. Correct and incorrect classifications receive binary rewards of $+1$ and $-1$, respectively and this is provided right after every action. This direct feedback enhances stable and effective learning in the suggested environment.

An additional test of the significance of the prioritized experience replay mechanism was done by a second training run where we turned off this feature, and the agent instead trusted uniform random sampling. In this condition, there was no difference in the likelihood of all experiences, both easy and difficult, being repeated in training. Because of this, the agent could not reliably concentrate on difficult or wrongly classified samples, causing slower convergence and a significant decline in performance. To be more precise, the agent was unable to effectively correct its errors, and thus its accuracy, precision, recall, and F1 scores were lower than in the initial configuration. The present findings underscore the importance of prioritization in directing the agent in focusing his/her attention on the most informative experiences as well as enhancing overall learning performance.

Table 4 details key hyperparameters for the DQN training. Algorithm 1 illustrates the general structure of the proposed approach.

**Algorithm 1** The general structure of the proposed approach.

Input: Pre-processed image

Output: Classification decision (forged vs. authentic)

1. The agent receives the current state (image pre-processed with ELA and noise analysis).
2. The image is processed by a dual-branch feature extractor: the ELA branch to capture high-level semantic features (64-Dimension). and the Noise Analysis branch to extract fine-grained noise patterns (64-Dimension).
3. Concatenate the two vectors (128-D) and refine via fully connected layers to obtain the state embedding.
4. The agent selects an action (0 for forged, 1 for authentic) using the current policy.
5. The environment assigns a reward (+1 for correct classification, –1 for misclassification) and returns the next state.
6. Store transition (state, action, reward, next state) in the prioritized replay buffer.
7. A mini-batch of 32 experiences is sampled (80% from misclassified cases and 20% from correctly classified ones) to update the DQN via gradient descent.
8. The training process iterates over multiple episodes until the agent's policy converges.

**Table 4.** Hyperparameter Values Used for DQN.

| Parameter | Value/Configuration |
|---|---|
| Buffer Size | 20,000 experiences |
| Exploration Initial Epsilon | 1.0 |
| Exploration Fraction | 0.09 |
| Exploration Final Epsilon | $10^{-10}$ |
| Learning Rate | 0.0001 |
| Device | Cuda |
| Accelerator | GPU T4 × 2 |

## Results and discussion

The methodology uses PyTorch and the Stable Baselines3 library on the Kaggle platform with a T4 × 2 GPU. A pipeline containing the environment setup, image preprocessing routines, dual-branch feature extractor, and DQN agent is all based on a cohesive pipeline. All datasets have been divided into 80 training and 20 testing. To reduce the bias in large datasets and balance the training process in various data collections, the training was done in two phases. Fig. 7 depicts the training curve of accuracy, reward, precision, and recall, respectively, in the initial training phase. The initial step entailed 1,000 training instances, which required about four hours using three huge datasets, and the specifications can be found in Table 5.

The figures indicate that initially, the performance of the agent is random, as the agent has a high exploration rate at the start of training. The more the training, the less the exploration and more exploitation, which allows the agent to benefit on the basis of the accumulated knowledge. This leads to a high level of performance, and performance progresses to almost total dependence on the strategies learned in the last episodes.

The second stage was followed by the complete training that had another 300 episodes and lasted about 1.5 hours. It uses three smaller datasets as described in Table 6, with the addition of 5 percent of the data set of the first phase to avoid overfitting to the data in the second phase. Fig. 8 depicts the training curves of accuracy, reward, precision, and recall, respectively, in the second training phase.

The curves of accuracy, reward, precision, and recall show the same general trends during the training process as shown in Figs. 7 and 8. This tendency is not surprising since the given task of classification is binary, and a balanced dataset is used, with an equal number of authentic and tampered images shown every episode. As all right predictions effectively add to both the reward signal and evaluation measures, and the model is consistent across both classes, the resultant curves plot on close-to-parallel curves.

Each training episode consisted of 100 training steps, which guaranteed the regular advancement of the training process. The application of varied sources of data and the maximization of utility to the agent in relation to input condition balances this two-stage training process. A customized mechanism of the callback monitors the important performance measures (e.g., total reward, accuracy, precision, and recall) during each episode. Monitoring of performance in real-time enables hyperparameters and network structure to be fine-tuned to achieve the best performance. The suggested system operates in memory image processing without using temporary disk files, enhancing the efficiency of feature extraction. Moreover, we also use a prioritized experience replay sampling mechanism that samples 80 percent of their hard experiences, guiding the agent to hard cases, thereby optimizing the rate of learning and

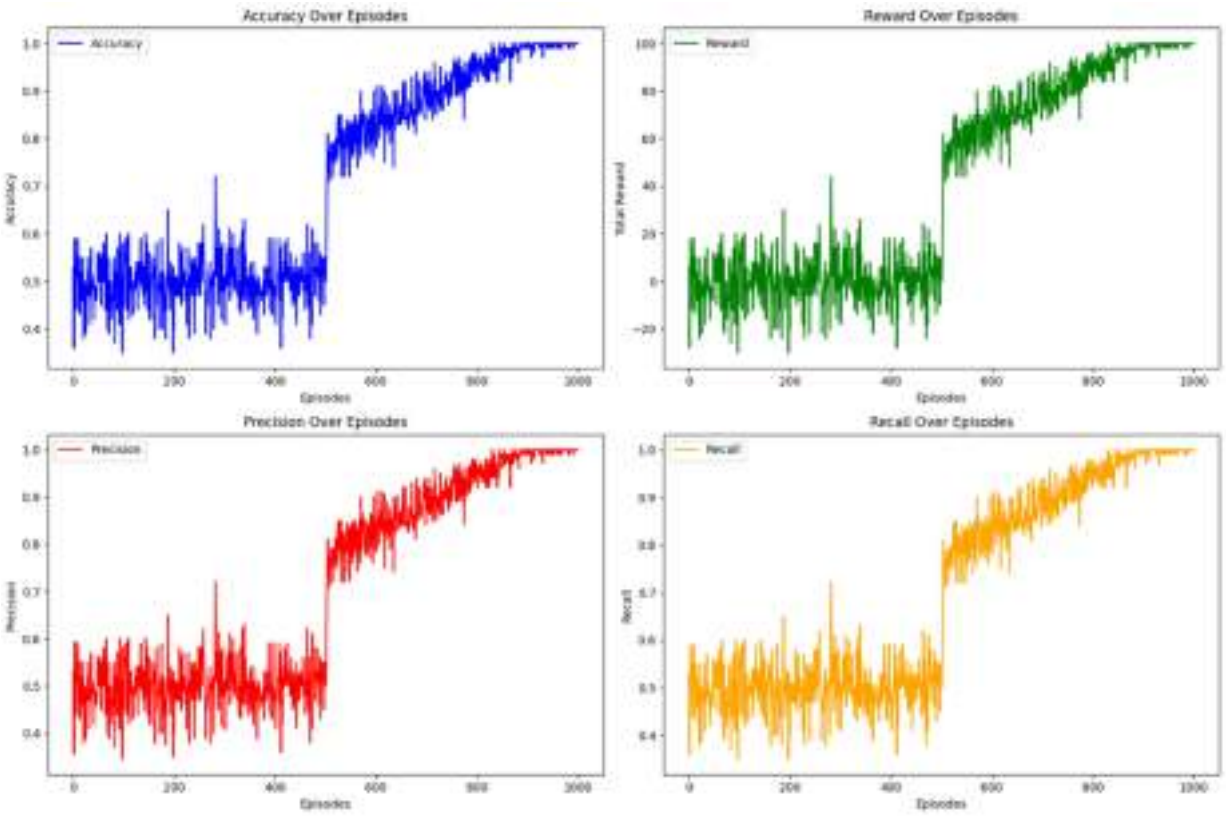


**Fig. 7.** Training curves for accuracy, reward, precision, and recall, respectively, in the first training phase.

**Table 5.** The details of the datasets used in the first training phase.

| Dataset | Number of Images | Format | Notes |
|---|---|---|---|
| DIS25K[26] | 24,964 | JPG | Deepfake Image Splices |
| DIS100K[27] | 100,000 | JPG | Deepfake Image Splices |
| MCCOCO[28] | 118,287 | JPG | Original images |

agent improvement. Tests were conducted to test the functionality of the proposed framework on the image forgery task. The DQN-Agent Model was evaluated using accuracy, precision, recall, and F1 scores across multiple datasets.

Total evaluation metrics determining the proposed framework are visible from Eqs. (1) to (4).[2]

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \tag{1}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{3}$$

$$\text{F1 Scoe} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives, respectively. Table 7 and Fig. 9 show the results achieved across different datasets in the testing phase, respectively.

**Table 6.** The details of the datasets used in the second training phase.

| Dataset | Number of Images (Original/Forged) | Format | Notes |
|---|---|---|---|
| CASIAv2[29] | 12,614 (7491/5123) | JPG, TIFF, BMP | Manual Fake Image Splices |
| IFD[30] | 13000 (6500/6500) | JPG, TIFF, BMP | Manual Fake Image Splices |
| MISD[31] | 922 (619/303) | JPG | Manual Fake Image Splices |

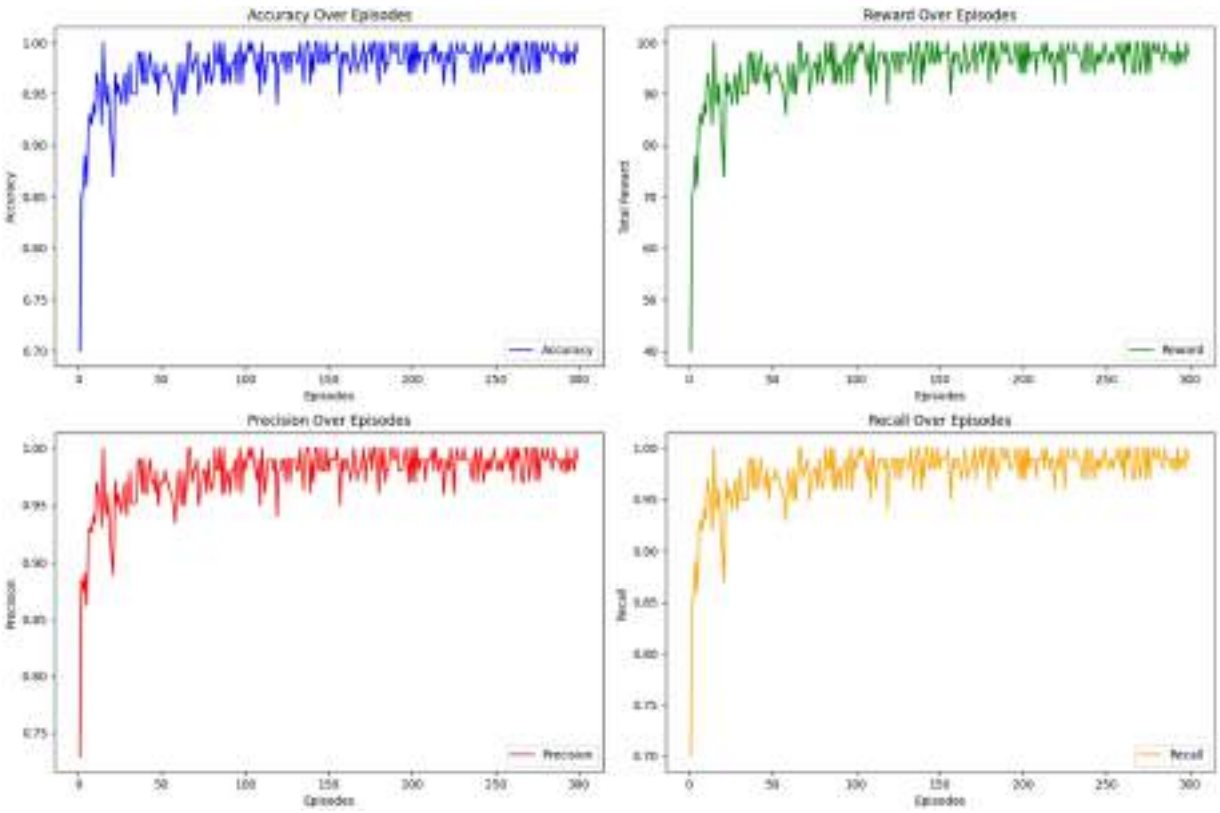


**Fig. 8.** Training curves for accuracy, reward, precision, and recall, respectively, in the second training phase.

**Table 7.** The performance metrics across the different datasets.

| Dataset | Accuracy | Precision | Recall | F1 Score |
|---|---|---|---|---|
| DIS25K + MCCOCO | 98.1% | 98.3% | 97.9% | 98.1% |
| DIS100K + MCCOCO | 98.4% | 98.6% | 98.5% | 98.5% |
| CASIAv2 | 97.4% | 97.3% | 97.6% | 97.4% |
| IFD | 98.3% | 98.6% | 98.2% | 98.4% |
| MISD | 97.8% | 97.6% | 98.0% | 97.8% |

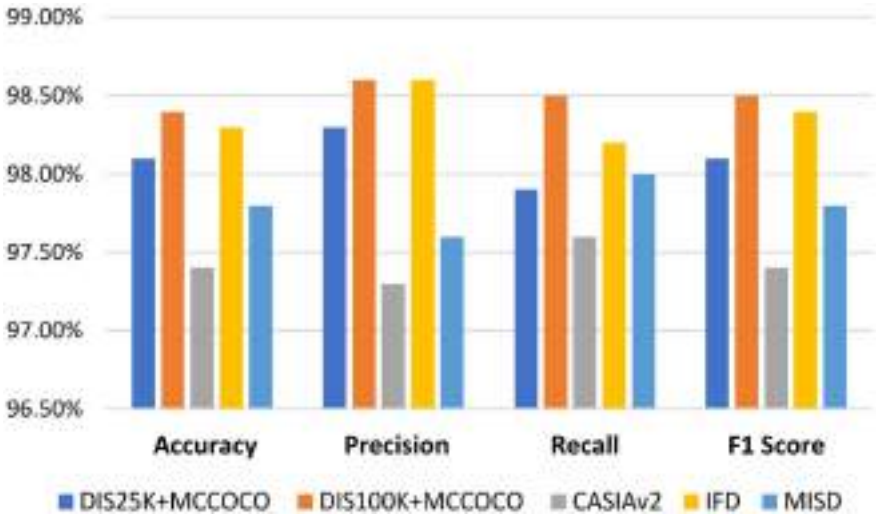


**Fig. 9.** The performance metrics across the different datasets.

**Fig. 10.** Some sample detection results on Casia V2: original images in the top row, spliced fakes in the bottom row.

To conduct a comprehensive and genuine evaluation of the proposed framework, the CASIA V2 dataset was selected to compare with related works. It is considered one of the best benchmark datasets and presents a significant challenge. To create realistic-looking images, the spliced regions in the forged images have undergone various processing operations, including distortion, rotation, and scaling. Additionally, the detection process became more complex due to the blurred edges of connected regions in the images.

The dataset offered by CASIA V2 is an ideal one to test the algorithms of forgery detection since its complexity corresponds to the effect of real life. The benchmark enables comparison with new methods to test the proposed approaches because of its wide level of acceptance among researchers. Fig. 10 presents a sample of the successful detection of original and fake images using the Casia V2 dataset.

As can be seen in Fig. 10, the agent proves to be able to distinguish between original and spliced images in the Casia V2 dataset: the upper row represents

**Table 8.** Comparison of the proposed method to related works on the CasiaV2 dataset.

| Method | Accuracy | Precision | Recall | F1-score |
|---|---|---|---|---|
| [14] | 92.2% | 85.9% | 97.0% | 91.1% |
| [17] | 94.5% | 94.7% | 94.6% | 94.7% |
| [18] | 94.7% | 94.8% | 94.7% | 95.0% |
| [20] | 94.14% | 94.1% | 94.0% | 94.0% |
| **Proposed** | **97.4%** | **97.3%** | **97.6%** | **97.4%** |

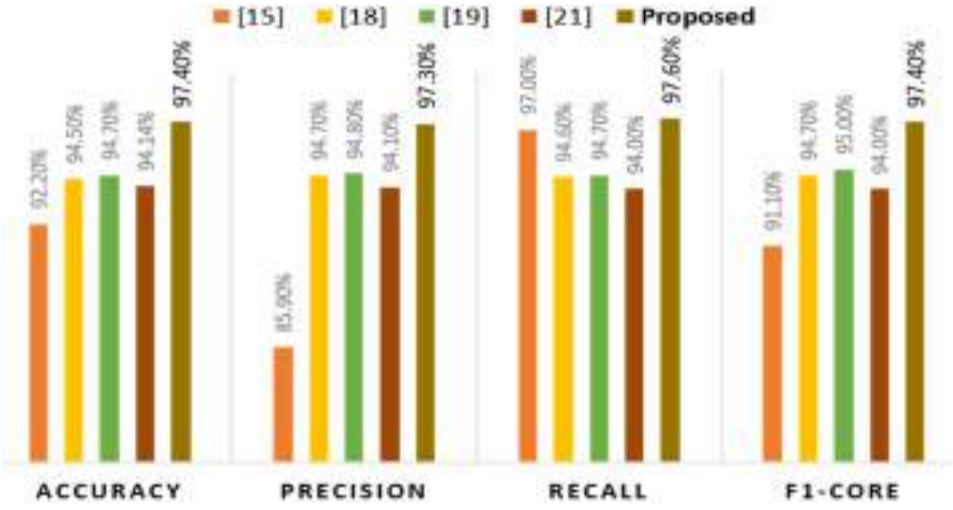


**Fig. 11.** The comparison of the proposed method with related works on the CasiaV2 dataset.

**Fig. 12.** Exemplary detection results on the custom Gemini 2.0 Flash data: original images (top row; left three images of the Casia v2 test set, right two images of in-house collected images), and spliced fakes (bottom row).

the actual images and the lower row represents the identified forgeries.

Table 8 and Fig. 11 compare the proposed method and a number of related methods on the Casiav2 data. The suggested structure is more accurate and has better precision, recall, and F1 scores than those revealed in previous research. These comparisons highlight the strength and applicability of the proposed approach.

To further evaluate the agent's generalization capabilities, a custom dataset of 200 forged images was created by manipulating 200 original images comprising 100 images from the Casia v2 test set (unseen during training) and an additional 100 images collected by our team. This was achieved using the state-of-the-art Gemini 2.0 Flash (Image Generation Experimental) model,[32] which was released recently this year. The agent demonstrated impressive performance on this dataset, achieving a detection accuracy of 97.50%, a precision of 97.97%, a recall of 97.0%, and an F1 score of 97.48%. Fig. 12 provides a qualitative illustration of the agent's behaviour on the Gemini 2.0 Flash test set. The three left-most columns originate from the Casia v2 test set, whereas the two right-most columns were collected in-house to ensure additional scene diversity. Each pair has the original image at the top and the image of the splice generated by Gemini on the bottom. The agent always places a label of a real one on the pristine samples and a fake on the manipulated ones. These illustrations affirm that the dual-branch ELA-plus-noise extractor catches the minor artifacts added by the state-of-the-art diffusion models and confirm the obtained findings.

Expansive feature extraction schemes, reinforcement learning, and prioritized experience replay rounds are superior in digital image forgery detection, which is an attractive procedure to other associated papers. The findings validate the idea that the suggested approach is effective at addressing the long-standing issues in the field of forensic applications, such as imbalance in the datasets, overfitting, or computational inefficiencies, to open up new research and practical implementation on the forensic application front.

In order to assess the effect of every substantial element in the proposed framework, we performed a focused ablation analysis in three dimensions. First, we have taken out the prioritization mechanism of the experience replay buffer and returned to uniform random sampling. This caused a perceptible decline in the performance since the agent no longer targeted challenging or informative samples, which caused slower convergence and poorer overall performance. Second, we modeled the model with the ELA feature branch only. Although the performance was also relatively high, a minor accuracy decline was noticed because the agent was unable to identify noise inconsistencies in tampered areas. Third, we tested the model with the noise analysis branch only. In this instance, the error was even bigger since the agent did not have access to compression-related artifacts that are normally displayed by ELA. These findings are important in highlighting the complementary effect of the dual-branch feature extractor and the relevance of prioritized sampling in strengthening important learning features. These experiments are summarized by the results shown in Tables 9 to 11, respectively.

## Limitations and future work

Although the proposed framework has demonstrated a steady performance both on standard

**Table 9.** The performance measures in the absence of prioritized experience replay in the various datasets.

| Dataset | Accuracy (%) | Precision (%) | Recall (%) | F1 Score (%) |
|---|---|---|---|---|
| DIS25K + MCCOCO | 94.20 | 94.75 | 93.60 | 94.17 |
| DIS100K + MCCOCO | 93.85 | 94.10 | 93.00 | 93.55 |
| CASIAv2 | 92.40 | 92.10 | 91.80 | 91.95 |
| IFD | 93.90 | 94.20 | 92.80 | 93.49 |
| MISD | 92.60 | 93.10 | 91.70 | 92.39 |

**Table 10.** Performance measures using only the ELA analysis branch on the various datasets.

| Dataset | Accuracy (%) | Precision (%) | Recall (%) | F1 Score (%) |
|---|---|---|---|---|
| DIS25K + MCCOCO | 96.40 | 96.80 | 95.70 | 96.24 |
| DIS100K + MCCOCO | 96.10 | 96.50 | 95.30 | 95.89 |
| CASIAv2 | 95.70 | 95.30 | 95.00 | 95.15 |
| IFD | 95.80 | 96.10 | 95.20 | 95.65 |
| MISD | 94.90 | 95.40 | 94.30 | 94.84 |

**Table 11.** Performance measures using only the noise analysis branch on the various datasets.

| Dataset | Accuracy (%) | Precision (%) | Recall (%) | F1 Score (%) |
|---|---|---|---|---|
| DIS25K + MCCOCO | 93.70 | 93.90 | 92.80 | 93.35 |
| DIS100K + MCCOCO | 93.30 | 93.50 | 92.10 | 92.80 |
| CASIAv2 | 93.90 | 93.40 | 93.00 | 93.19 |
| IFD | 93.10 | 93.50 | 92.00 | 92.74 |
| MISD | 93.20 | 93.50 | 92.10 | 92.80 |

datasets and on a custom diffusion-based test set, some areas can still be explored. Specifically, the agent is not evaluated yet against inputs that are far out of the training distribution, e.g. grossly distorted images, unknown synthetic forgeries, or content that has unusual semantic structure. Though preliminary tests on different samples did not show any significant performance decrease, we admit that the task of processing extreme distribution shifts is an overall problem of data-driven models. The future research will focus on the domain generalization methods and improvements in the training strategies to enhance the adaptability of the model.

## Conclusion

This paper introduces a digital image forgery detection model that is revolutionary in that it incorporates the concept of reinforcement learning in the forensic model. The suggested approach manages to resolve the long-lasting problems of imbalance in the dataset, overfitting, and lack of feature representations by redefining the detection problem as a sequential decision-making process. The dual-branch feature extractor effectively integrates the high-level semantic information in a pre-trained ResNet50 and low-level noise features based on a custom CNN, and increases the total discriminative capabilities of the system. The experience replay mechanism, which is the most important, also makes sure that the agent learns efficiently by its most challenging misclassifications. As far as we know, this is the first implementation of a DQN agent for this purpose, and it has demonstrated excellent accuracy rates of 97% to 98% on a wide range of datasets. The scalability and efficiency of the suggested framework highlight its possible use in the real-world sphere of digital forensics. The next step in the future is to tighten hyperparameter optimization and experiment with deeper learning architectures to counter new forgery techniques, which will lead to more resilient detection systems.

## Authors' declaration

- Conflicts of Interest: None.
- We hereby confirm that all the Figures and Tables in the manuscript are ours. Furthermore, any Figures and images that are not ours have been included with the necessary permission for republication, which is attached to the manuscript.
- No animal studies are present in the manuscript.
- No human studies are present in the manuscript.
- Ethical Clearance: The project was approved by the local ethical committee at Mustansiriyah University.

## Acknowledgment

The authors acknowledge the Department of Computer Science, College of Science, Mustansiriyah

University, and the Department of Computer Science, University of Technology, to sponsor this work.

## Authors' contributions statement

Conceptualization, MSM and SNA; methodology, HSA; software, MSM; validation, SNA, HSA, and MSM; formal analysis, MSM; investigation, HSA; resources, SNA; data curation, MSM; writing—original draft preparation, MSM; writing—review and editing, SNA; visualization, HSA; supervision, SNA; project administration, HSA; funding acquisition, MSM.

## Data availability

The datasets generated and analysed during the current study are available in the list of references.

## References


1. Farhan MH, Shaker K, Al-Janabi S. Efficient approach for the localization of copy-move forgeries using PointRend with RegNetX. Baghdad Sci J. 2024;21(4):1416–1426. https://doi.org/10.21123/bsj.2023.8304.
2. Mahdi MS, Alsaad SN. Detection of copy-move forgery in digital image based on SIFT features and automatic matching thresholds. In: Proc. Int. Conf. Appl. Comput. Support Ind.: Innov Technol Cham: Springer; 2019;17–31. https://doi.org/10.1007/978-3-030-38752-5_2.
3. Kumari R, Garg H. Image splicing forgery detection: a review. Multim Tools Appl. 2024;84(8):4163–4201. https://doi.org/10.1007/s11042-024-18801-z.
4. Sharma P, Kumar M, Sharma H. Comprehensive analyses of image forgery detection methods from traditional to deep learning approaches: an evaluation. Multim. Tools Appl. 2023;82(12):18117–18150. https://doi.org/10.1007/s11042-022-13808-w.
5. Choudhary RR, Paliwal S, Meena G. Image forgery detection system using VGG16 UNET model. Procedia Comput Sci. 2024;235:735–744. https://doi.org/10.1016/j.procs.2024.04.070.
6. Mahdi MS, Alsaad SN. False matches removing in copy-move forgery detection algorithms. Al-Mustansiriyah J Sci. 2020;31(1):47–53. https://doi.org/10.23851/mjs.v31i1.748.
7. Al-Tai MH, Nema BM, Al-Sherbaz A. Deep learning for fake news detection: literature review. Al-Mustansiriyah J Sci. 2023;34(2):70–81. https://doi.org/10.23851/mjs.v34i2.1292.
8. Mahdi MS, Ali ZL, Rashid AR, Ibrahim NK, Abdulghafour AWA. A hybrid deep learning model for facial emotion recognition: combining multi-scale features, dynamic attention, and residual connections. In: Proc. Int. Conf. Appl. Innov. IT. 2025;13(2):69–77. https://doi.org/10.25673/120395.
9. Shi C, Chen L, Wang C, Zhou X, Qin Z. Review of image forensic techniques based on deep learning. Mathematics. 2023;11:3134. https://doi.org/10.3390/math11143134.
10. Alshwely MK, Alsaad SN. Image splicing detection based on noise level approach. Al-Mustansiriyah J Sci 2020;31(4):55–61. https://doi.org/10.23851/mjs.v31i4.899.
11. Hussien NY, Mahmoud RO, Zayed HH. Deep learning on digital image splicing detection using CFA artifacts. Int J Sociotech Knowl Dev. 2020;12(2):31–44. https://doi.org/10.4018/IJSKD.2020040102.
12. Islam MM, Karmakar G, Kamruzzaman J, Murshed M. A robust forgery detection method for copy–move and splicing attacks in images. Electronics. 2020;9(9):1500. https://doi.org/10.3390/electronics9091500.
13. Siddiqi MH, Asghar K, Draz U, Ali A, Alruwaili M, Alhwaiti Y, *et al.* Image splicing based forgery detection using discrete wavelet transform and edge weighted local binary patterns. Secur. Commun. Networks. 2021;2021:4270776. https://doi.org/10.1155/2021/4270776.
14. Ali SS, Ganapathi II, Vu NS, Ali SD, Saxena N, Werghi N. Image forgery detection using deep learning by recompressing images. Electronics. 2022;11(3):403. https://doi.org/10.3390/electronics11030403.
15. Xu S, Lv S, Liu Y, Xia C, Gan N. Scale-adaptive deep matching network for constrained image splicing detection and localization. Appl. Sci. 2022;12(13):6480. https://doi.org/10.3390/app12136480.
16. Baumy A, Algarni AD, Abdalla M, El-Shafai W, Abd El-Samie FE, Soliman NF. Efficient forgery detection approaches for digital color images. Comput Mater Continua. 2022;71(2):3257–3276. https://doi.org/10.32604/cmc.2022.021047.
17. Sedeeq I. Image splicing detection based on discrete wavelet transform and co-occurrence matrix. Iraqi J Sci. 2023;64(11):5940–5951. https://doi.org/10.24996/ijs.2023.64.11.38.
18. Khalil AH, Ghalwash AZ, Elsayed HAG, Salama GI, Ghalwash HAG. Enhancing digital image forgery detection using transfer learning. IEEE Access. 2023;11:91583–91594. https://doi.org/10.1109/ACCESS.2023.3307357.
19. Ramirez-Rodriguez AE, Arevalo-Ancona RE, Perez-Meana H, Cedillo-Hernandez M, Nakano-Miyatake M. AISMSNet: advanced image splicing manipulation identification based on Siamese networks. Appl Sci. 2024;14(13):5545. https://doi.org/10.3390/app14135545.
20. Nagm AM, Moussa MM, Shoitan R, Ali A, Mashhour M, Salama AS, *et al.* Detecting image manipulation with ELA-CNN integration: a powerful framework for authenticity verification. PeerJ Comput. Sci. 2024;10:e2205. https://doi.org/10.7717/peerj-cs.2205.
21. Shan W, Zou D, Wang P, Yue J, Liu A, Li J. RIFD-Net: a robust image forgery detection network. IEEE Access. 2024;12:20326–20340. https://doi.org/10.1109/ACCESS.2024.3359991.
22. Gymnasium. An API standard for reinforcement learning with a diverse collection of reference environments [Internet]. 2025.
23. Wang X, Wang S, Liang X, Zhao D, Huang J, Xu X, *et al.* Deep reinforcement learning: a survey. IEEE Trans. Neural Netw Learn Syst. 2022;35(4):5064–5078. https://doi.org/10.1109/TNNLS.2022.3207346.
24. Joshi R, Gupta A, Kanvinde N, Ghonge P. Forged image detection using SOTA image classification deep learning methods for image forensics with error level analysis. In: Proc. 2022 13th Int Conf Comput Commun Netw Technol. (ICCCNT). Piscataway, NJ: IEEE; 2022;1–6. https://doi.org/10.1109/ICCCNT54827.2022.9984489.
25. Zeng P, Tong L, Liang Y, Zhou N, Wu J. Multitask image splicing tampering detection based on attention mechanism. Mathematics. 2022;10(20):3852. https://doi.org/10.3390/math10203852.
26. Tahir E. DIS25k: Diverse Image Splicing 25 000 dataset [Internet]. 2024.

27. Tahir E. DIS100k: Diverse Image Splicing 100 000 dataset [Internet]. 2024.
28. Hesaraki S. COCO Dataset 2017 [Internet]. 2023.
29. Dong J, Wang W, Tan T. CASIA image tampering detection evaluation database. In: 2013 IEEE China Summit Int Conf Signal Inf Process. IEEE; 2013 Jul.;422–426. https://doi.org/10.1109/ChinaSIP.2013.6625374.
30. Labid93. Image forgery detection dataset [Internet]. 2023.
31. Kadam KD, Ahirrao S, Kotecha K. Multiple image splicing dataset (MISD): A dataset for multiple splicing. Data. 2021;6(10):102. https://doi.org/10.3390/data6100102.
32. Gemini 2.0 Flash (Image Generation Experimental). Google AI Studio [Internet].

# إطار عمل جديد لوكيل التعلم التعزيزي للكشف عن تزوير الصور الرقمية باستخدام إعادة تشغيل التجربة ذات الأولوية

**مثنى ص. مهدي[1]، سعد ن. السعد[1]، حسنين س. عبدالله[2]**

[1]قسم علوم الحاسوب، كلية العلوم، الجامعة المستنصرية، بغداد، العراق.
[2]قسم علوم الحاسوب، الجامعة التكنولوجية، بغداد، العراق.

## الخلاصة

أصبح كشف تزوير الصور الرقمية مشكلةً مُلحة ومعقدة في عصرٍ تُمكن فيه أدوات التحرير القوية من تعديل الصور بسهولة. لم يعد من الممكن الأخذ بالمقولة الشائعة "الصورة تساوي ألف كلمة" على ظاهرها، فحتى التلاعبات الدقيقة قد تُخفي أو تُزيّف تفاصيل بالغة الأهمية. غالبًا ما تُعاني أساليب الكشف التقليدية من اختلال شديد في مجموعات البيانات وتمثيلاتٍ محدودة للخصائص تفشل في التقاط الآثار المتنوعة التي تُقدمها تقنيات التحرير الحديثة. استجابةً لذلك، يُقدم هذا العمل إطارًا جديدًا يُصوّر مهمة كشف التزوير كمشكلة تعلم مُعزز، مما يُمكّن الوكيل من تعلم سلسلة من خطوات التحليل بدلاً من الاعتماد على مسار تصنيف ثابت. هذه هي أول دراسة تستخدم وكيل Deep Q-Network (DQN) مُخصصًا لكشف تزوير الصور. تتميز البنية المُقترحة بمستخرج خصائص ثنائي الفروع: يُطبق أحد الفروع تحليل مستوى الخطأ للكشف عن تناقضات الضغط. وفي الوقت نفسه، يُجري الفرع الآخر تحليلًا دقيقًا للضوضاء للكشف عن آثار التلاعب. تُقدّم هذه الميزات لبيئة Gym مُخصّصة، مُصمّمة بأخذ عينات متوازن للحدّ من اختلال توازن الفئات، ومُعزّزة بآلية إعادة تشغيل للتجربة ذات أولوية، تُوجّه عملية التعلّم نحو أمثلة الوكيل الأكثر صعوبة. تُظهر التجارب المُكثّفة على مجموعات بيانات مُتعددة، بما في ذلك الاختبار على أحدث نموذج لمعالجة الصور، Gemini 2.0 Flash، مُعدّلات دقة مُذهلة تتراوح بين 97% و98%. تُسلّط هذه النتائج الضوء على متانة وكفاءة النهج المُقترح، مُقدّمةً اتجاهًا جديدًا واعدًا لتحقيقات الطب الشرعي الرقمي.

**الكلمات المفتاحية:** بيئة متوازنة، شبكة كيو العميقة، تقنية التزييف العميق، استخراج الميزات ثنائية الفروع، إعادة تشغيل التجارب ذات الأولوية، وكيل التعلم المعزز، كشف تزوير الصور المدمجة.