Journal of Soft Computing and Computer Applications

Manuscript 1023

Enhanced Generative Convolutional Networks: A Hybrid Algorithm for Refinement Video Classification

Dalal Thair Mahjoub

Hala Bahjat Abdulwahab

Kesra Nermend

JOURNAL OF SOFT COMPUTING AND COMPUTER APPLICATIONS 2025;2:1023

Scan the QR to view the full-text article on the journal website

ORIGINAL STUDY

Enhanced Generative Convolutional Networks: A Hybrid Algorithm for Refinement Video Classification

Dalal Thair Mahjoub [a,*], Hala Bahjat Abdulwahab [b], Kesra Nermend [c]

[a] University of Technology - Iraq, College of Computer Science, Al-Sina'a St., Al-Wehda District, 10066 Baghdad, Iraq
[b] University of Technology - Iraq, College of Computer Science, Department of Computer Security and Cybersecurity, Al-Sina'a St., Al-Wehda District, 10066 Baghdad, Iraq
[c] University of Szczecin, Department of Computer Science, Institute of IT in Management, Mickiewicza St., Śródmieście District, 70-383 Szczecin, Poland

ABSTRACT

Video classification is a vital area of research due to the growing volume of video content in various applications. Accurate category across various resolutions poses challenges, which include adapting to scaling, resizing, and compression. Therefore, this paper introduces an innovative Generative Convolutional Network (GCN) set of rules tailored for multi-resolution video classes. The proposed GCN model utilizes Convolutional Neural Networks (CNNs) combined with generative modeling to enhance the extraction of functions across varying video resolutions, which is crucial for maintaining class robustness in the face of common video adjustments, such as scaling, resizing, and compression. In contrast, traditional fashions frequently struggle with such versions, resulting in inconsistent class results. By integrating CNNs with a generative community and utilizing antagonistic training, the GCN refines its function, becoming familiar with and improving its typical class performance. The generative aspect simulates input variations, enabling the model to remain resilient under various video conditions.

Experimental results show that the GCN outperforms traditional strategies across key assessment metrics. On the UCF101 dataset, it achieved 98% accuracy, 94% Recall, 95% precision, and a 96% F-score. Compared to setting up models using SVM, CNN, KNN, and RNN, the GCN can consistently deliver advanced performance, particularly in terms of robustness, accuracy, and resistance to parameter modifications. These findings highlight the potential of GCN to become a new benchmark for video classification, particularly in real-time applications that require dependable evaluation of multi-resolution video content.

Keywords: Convolutional neural networks, Generative adversarial networks, Video processing, Multi-resolution, Classification, Deep learning, Generative convolutional networks

Received 10 January 2025; revised 21 July 2025; accepted 25 July 2025.
Available online 26 December 2025

* Corresponding author.
E-mail addresses: cs.19.09@grad.uotechnology.edu.iq (D. T. Mahjoub), hala.b.abdulwahab@uotechnology.edu.iq (H. Bahjat Abdulwahab), kesra.nermend@usz.edu.pl (K. Nermend).

https://doi.org/10.70403/3008-1084.1023

1. Introduction

With the constantly changing digital content environment, effective film categories through different resolutions have become more and more important to many applications, such as content recommendation and automatic surveillance [1, 2]. The diversity in resolution formats of videos on the Internet poses special challenges, including how to ensure category labels still have high quality when the videos are resized, scaled, or compressed. Conventional classification techniques may not be well-suited to these multi-resolution problems and tend to perform poorly [3]. Recent advances in deep learning have greatly advanced visual categorization. The development of deep learning has enabled Recurrent Neural Network (RNN) scalability to background processing architecture learning methods, including Convolutional Neural Networks (CNNs), which have yielded an impressive improvement in video category recognition [4]. CNNs offer powerful feature extraction and classification capabilities, but often struggle to handle the complexity of different video resolutions and aspect ratios [5]. The traditional CNN model does not handle the challenging temporal and spatial dynamics exhibited in videos well, especially with a loss of efficacy in classification when adapting to changes on the resolution scale [6, 7].

The emergent need for more efficient and adaptive video classification solutions has led to the emergence of hybrid models, including generative methods [8]. In particular, Generative Adversarial Networks (GANs) have shown promising applications in both data augmentation and feature enhancement, where they lightly perturb the original video data, making them useful in improving robustness in features learned from the video data [4, 9]. Nevertheless, existing approaches still face challenging scenarios, such as high computational cost, variable video codecs, and the requirement of a large amount of annotated data. Moreover, conventional methods are ineffective in multi-resolution transformation and cannot be applied to real scenes [10].

The main shortcoming of the state-of-the-art methods lies in the fact that existing approaches still fail to consistently achieve both high accuracy and robustness for multi-decision video content. CNNs are very good at extracting functions; however, they lack the flexibility to vary in resolution and are susceptible to distortion. Likewise, GANs have not been fully explored to better exploit feature extraction and type consistency in multi-resolution settings, at the same time. This obstruction motivates the proposal of a new model that combines the advantages of CNNs and GANs to address the challenges of multi-decision video classification.

The key objective of this paper is as follows:

1. Propose a hybrid Generative Convolutional Network (GCN) that integrates CNNs and GANs for strong multi-decision video type.
2. Enhance overall classification performance by incorporating hostile training to simulate various video adjustments.
3. Optimize the generator architecture by stabilizing the latent input for consistent, class-specific outputs.
4. Adapt the Discriminator to support multi-class output, aligning with the categorical nature of video datasets.
5. Benchmark GCN in opposition to conventional models, which include SVM, CNN, KNN, RNN, and traditional GANs.
6. Ensure computational efficiency to support real-time and resource-constrained applications, including those in surveillance and healthcare, to optimize performance.

This paper aims to bridge the existing studies by introducing a singular GCN model, especially designed for multi-decision video types. By addressing the challenges of the multi-resolution video content category in an increasingly virtual international environment, we improve accuracy and resilience. By accomplishing these goals, the GCN model contributes to the development of more reliable video evaluation packages, ultimately fostering greater confidence in automated systems. To address these demanding situations, a novel GCN model has been proposed, designed to refine the CNN model for multi-resolution video classification. The GCN integrates a generative factor that simulates actual global variations in capability, taking into account more effective characteristic extraction and type accuracy. By employing adversarial training, the GCN refines the typing process, making it resilient against numerous styles of video manipulation. Through this paper, the strengths of today's video classification techniques have been examined, providing a strong strategy to address the challenges posed by multi-decision content in an increasingly digital world. By improving class accuracy and resilience, they contribute to greater reliability in video analysis packages, ultimately fostering increased confidence in automated systems.

The key contributions of this paper are as follows:

- **Developed a novel GCN framework** that efficiently combines CNN-based total spatial feature extraction with GAN-based adversarial opposite learning for video classification.
- **Implemented an antagonistic refinement** method that improves classification robustness under various resolutions and video conditions.
- **Modified the generator's objective function** to use a fixed latent vector, resulting in output stability throughout classes and datasets.
- **Introduced a multi-class discriminator objective function**, instead of a binary category, to enhance category-specific learning and help with large-scale datasets.
- **Demonstrated empirical superiority**, attaining up to 98% classification accuracy on benchmark datasets such as UCF101, HMDB-51, YouTube-8M, and ActivityNet.
- **Validated resource efficiency**, with aggressive GPU memory consumption and education time.
- **Enabled sensible deployment,** showcasing use cases in real-world applications like automated surveillance, content moderation, and medical video diagnostics.

The rest of the paper is organized as follows: Section 2 critiques the current model in the video category and highlights its limitations in handling multi-decision eventualities. Section 3: Video classification provides an in-depth discussion of video categories, emphasizing the importance of temporal and spatial functions, and outlines conventional machine learning, CNN, and GAN-based strategies. Section 4: The proposed methodology provides Information on the design and implementation of the GCN model, including its structure and adversarial education method. Section 5: Experimental results afford a comparative evaluation of GCN with baseline models throughout a couple of datasets, emphasizing its superior performance. A general discussion of the proposed model and the experimental results is provided in Section 6. Finally, Section 7: Conclusion and Future Work summarizes the findings, discusses the impact of the GCN framework, and outlines avenues for future studies. This structure ensures a rigorous scientific exploration of the proposed model and its contributions to advancing the field of video classification.

2. Related work

Research on deep learning techniques for video evaluation highlights diverse modern strategies aimed at enhancing performance in tasks such as movement recognition and video classification. These strategies leverage superior neural networks to extract significant capabilities and ensure robustness against diverse challenges in video information. The following are key factors associated with the work in this discipline. Key contributions include the framework, which combines CNNs and RNNs for movement recognition in videos. The method achieved trend accuracy in spotting complex movements. However, the version's dependence on huge annotated datasets limits its practicality in situations with limited records [8, 9]. A flow-based total CNN model has been introduced for real-time motion recognition, leveraging optical flow combined with spatial features to enhance both accuracy and processing speed. However, challenges remain in adapting the model to variations in light conditions and background changes [10]. Several excellent research studies have addressed the challenges of video classification using various machine learning approaches. The authors of [11] explored the use of GANs for synthesizing videos from textual descriptions, demonstrating the capacity of GANs in video generation. However, the examiner said problems in preserving temporal coherence and consistency over extended sequences. Integrating SVMs with deep learning capabilities for video classification tasks has been introduced by [12]. By employing transfer learning from pre-trained CNNs, the authors achieved significant improvements in classification performance.

However, the method faced challenges due to inefficiency in the context of massively large video datasets, primarily linked to scalability and computational efficiency issues. In another study, the KNN method was evaluated for video category classification, resulting in significant accuracy improvements. However, the high computational weight of KNN has been one of the most important obstacles for real-time applications. In [13], a three-dimensional Convolutional Neural Network (3D-CNN) specifically designed for human action recognition tasks within video sequences has been developed. Although the model obtained very good results, it consumed a large amount of computational resources during training. Finally, RNNs applied to temporal action localization are addressed by [14], which highlights the necessity of sequence modeling to improve detection performance. The analyses were powerful, but this highlighted that there is a trade-off between version complexity and real-world performance needs.

In summary, the deep learning tools for video analysis are still in fast development. The use of state-of-the-art neural network structures is generally improving performance across a diverse range of tasks. However, it remains a challenge to deal with the complexity of video records, as well as achieve consistent results in various applications.

3. Video classification

Video classification is the process of categorizing video content according to both temporal and spatial features, enabling applications such as, but not limited to, content recommendation, surveillance, and sports analysis [15]. Unlike a static picture classification, it reads sequences across time by synthesizing the picture in context from motion and then recognizing changing patterns. Deep learning approaches, particularly hybrid models that integrate convolutional and generative approaches, have significantly reduced video classification error [16]. CNNs are strong in extracting spatial and structural features, whereas GANs help in generating reasonable video sequences, which can also increase the educational dataset [17]. Recent advances also include interest mechanisms that are sensitive to the most useful functions of a sequence, thereby improving accuracy and efficiency.

Nevertheless, some challenging problems, such as computational cost, the diversity of video formats, and the requirement for sufficient annotated data, are still limited. The solutions to these challenges are achieved through methods such as multi-decision analysis and transfer learning, which provide solutions with a higher degree of robustness [18]. Progress in the video portion is important for various applications, such as automatically detecting people in surveillance videos, commercial content moderation, and medical analytics, in which precise pattern recognition is crucial. For example, recent studies have shown that two-stream CNN architectures are efficient in action recognition, and attention-based mechanisms have achieved superior performance in high-temporal-precision applications [19, 20]. The following sections outline how to utilize machine learning, CNNs, and GANs in video classification.

3.1. Video classification based on machine learning

Machine Learning Video Classification employs sophisticated algorithms to treat and tag video content, offering automation, reliability, and consistency in multiple use cases. Types of Classification Strategies There are two main types of classification strategies, classified as: body-based classification, which takes the popularity of character frames and makes a decision for each frame as to whether it is good or bad, and video-level classification that aggregates features over time to classify a video as good or bad [21]. Machine learning programs inside this context comprise function extraction wherein algorithms extract big visual and audio features to improve the type accuracy; temporal modeling capturing dynamics in movement sequences to improve the recognition overall performance; and ensemble strategies that combine the predictions from a couple of models to enhance robustness and reduce errors [22, 23].

The benefits of these techniques include the precise initiation of complex moves step by step and the real-time processing of dynamic video streams, as well as the classification of video types such as sports and surveillance video [24, 25]. Challenges still exist, however, such as the difficulty of training deep end-to-end models, dependence on large-scale annotated data, and overfitting to particular video genres. Use-cases of this generation include applying to anything from community moderation and fashion recommendation systems to automated surveillance and sports analytics, with a focus on inferring valuable insights from video content. Continuous research is working towards enhancing the performance and efficacy of gadget-getting-to-know-you, mostly classification-based video algorithms, mainly through the development of better deep learning architectures and the incorporation of interest mechanisms to spot relevant content in dynamic video environments [26].

3.2. Video classification based on convolutional neural networks

CNNs have shown large promise in the video category, leveraging their potential to robotically extract and learn complex temporal and spatial capabilities [27]. CNN-based version that effectively classifies motion reputation in movies by capturing both spatial and temporal dynamics. However, these models often conflict with sizable changes, in particular adjustments in resolution or body prices [28]. As shown in Fig. 1, the inherent structure of CNNs enables them to adapt to various noise styles; they face demanding situations while videos undergo significant alterations, motivating the exploration of advanced techniques like GANs to enhance robustness [29, 30].

3.3. Video classification based on generative adversarial networks

GANs, introduced by [31], have been widely applied to various video processing tasks, including video generation, enhancement, and classification. GANs aid video classification

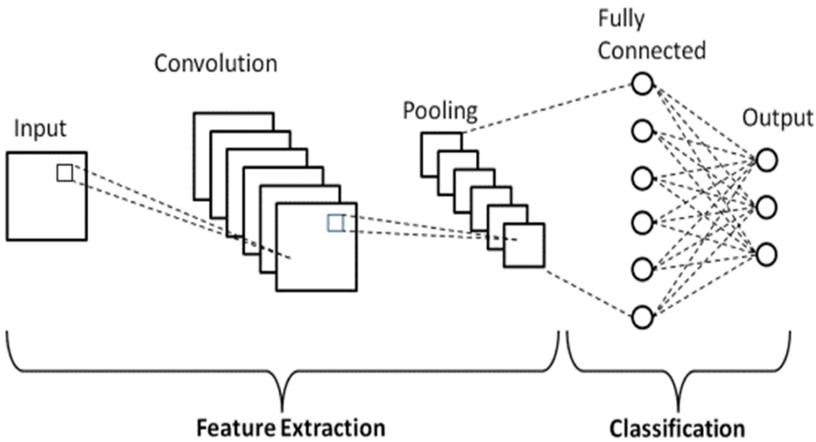

Fig. 1. Convolutional neural network.

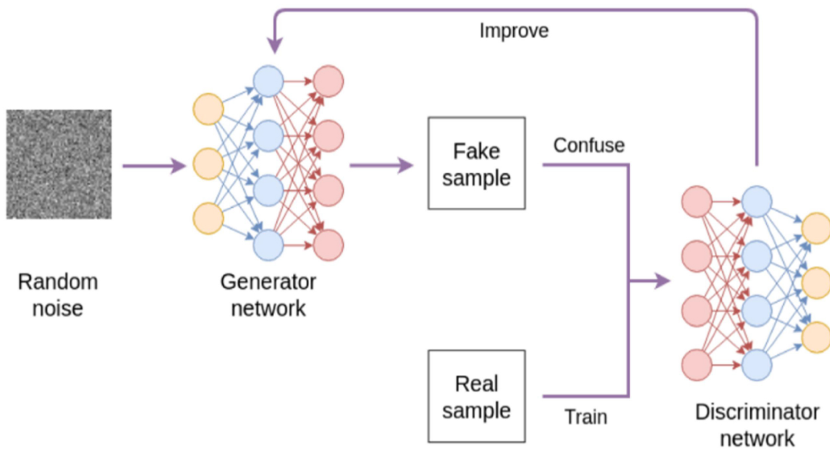

Fig. 2. Workflow of generator and discriminator networks in the GAN model.

with the aid of augmenting data, enhancing resolution adaptability, refining features, and aiding area editing. While not direct classifiers, GANs enhance robustness and can be adapted to help in classification by modifying the Discriminator for class probability outputs. GANs consist of two main components: a generator that produces synthetic data (e.g., images, text, or audio) that mimics the real data distribution [32]. It starts by generating random outputs and progressively improves by learning from feedback provided by the Discriminator. The Discriminator evaluates the authenticity of the data, distinguishing between real data (from the dataset) and fake data (from the generator) [33, 34]. It provides feedback to the generator, helping it improve the quality of its outputs, as shown in Fig. 2. The adversarial training process enables the generator to produce increasingly realistic videos [35, 36].

The goal of the generator is to produce statistics that are indistinguishable from actual information [37, 38]. The generator aims to maximize the Discriminator's probability of classifying generated (fake) records as actual [39]. These training mechanisms define the adverse schooling technique in a standard GAN [40, 41]. In the context of video classification, GANs have been utilized to enhance extracted features and increase the variety of training datasets [42]. For instance, [43] proposed a GAN-based approach that generates augmented video sequences to enhance the accuracy of motion recognition models [44] explored using GANs to create multi-decision video changes, which help enhance the robustness of category fashions against resolution adjustments.

4. Proposed generative convolutional networks refinement approach

By integrating GANs with CNNs, the proposed model aims to leverage the generated multi-decision transformations to enhance overall video performance, particularly in dynamic and challenging environments. During the preprocessing step, movies were divided into frames at set intervals, scaled to 256 × 256 pixels, and keyframes were selected based on entropy values to preserve the maximum applicable regions. GCNs integrate the strengths of GANs and CNNs to excel in statistics technology and feature extraction. At its essence, GANs contribute creative realism through a Generator-Discriminator dynamic, refining outputs to mimic real-world Information. CNNs enhance this by extracting specific spatial and structural capabilities, allowing GCNs to model complex patterns with high consistency and accuracy, making them versatile for applications like computer vision and 3D modeling.

Key features of the proposed GCN model include:

1. **Generative Component:** The generator in the GCN synthesizes enhanced multi-resolution features of video frames that emulate potential transformations. This enables the classification model to adapt effectively to diverse video inputs by achieving robustness against variations in resolution.
2. **Adversarial Training:** GCN employed adversarial training to facilitate classification. Since the Discriminator is trained to judge the quality and accuracy of classification under multiple conditions, the GCN learns to classify videos better and more robustly from the beginning.
3. **Multiscale Adaptability:** Unlike previous works, GCN is particularly adapted to any resolution in video frames. This flexibility guarantees that classification is preserved and consistent.
4. **Robustness Analysis:** The framework designs the mechanisms to analyze the classification robustness to accuracy, recall, precision on F-score, and makes sure that the current GCN could satisfy the requirements of general video classification.

The GAN was coupled with the CNN to stabilize and regularize the feature generation. In traditional GANs, the features are generated randomly by the generators, resulting in high variation in random features between iterations. Although this randomness is useful for some applications needing distinctive outputs, it is frequently unacceptable when stable, repeatable features are desired. To handle this, a static input structure was used, which limits the variation of the generator. By binding the process of the generator to a concrete input, a controlled reference point for selective feature generation has been created, and the reproducibility of the generator has been enhanced as follows:

$$L_G = E_{z\sim p_{z(z)}} log\left(1 - D\left(G\left(z\right)\right)\right) \tag{1}$$

where:

- z is a random input (usually sampled from a normal or uniform distribution),
- G (z) is the output of the generator, which aims to resemble real data.

- D (G (z)) is the probability assigned by the Discriminator that the generated data is real.

A Generator network is used to enhance or simulate frame features (e.g., through noise injection or augmentation). To enhance the diversity or quality of extracted features. Through this setup, the generator now selects features in a pseudo-random manner based on the predetermined input, which aligns the generator's functionality more closely with the requirements of stability-focused applications. As a result, the generator's output has been transformed from a dynamic and variable state to a fixed, consistent form in each iteration by introducing a new parameter called. (f_{frame}). This variable enables the generator to produce consistent results despite input variations, making it more robust and versatile in tasks like image and video generation. Therefore, the new modified objective function of the generator will be as follows:

$$L_G = -E_{f_{frame}} \cdot log\,(1 - D(G\,(z, c))_c) \tag{2}$$

where:

- G: The Generator function.
- D: The Discriminator function.
- z: A random noise vector sampled from a distribution (e.g., Gaussian or Uniform).
- c: The target class label (one of C real classes).
- $G(z, c)$: The generator outputs a single fake frame conditioned on, which is supposed to resemble a real frame belonging to the class c.
- $-\ E_{f_{frame}}$: Negative expectation, to ensure that the generator learns to maximize the Discriminator's confidence in misclassifying the fake frame as belonging to the real class, and f_{frame} indicates a fixed number of frames.

To evaluate the effect of (f_{frame}), an examination has been performed by comparing the performance of the generator with and without the inclusion of (f_{frame}). The consequences are summarized in Table 1. This optimization of the generator in the GAN algorithm, with fixed input constraints, presents a novel approach to controlled feature generation, thereby expanding the versatility of the generator for applications where controlled variability and reliability are crucial. In GCN, its key roles, as shown in Table 1, include spotlighting the improvements achieved through the inclusion of (f_{frame}), specifically in stability and learning guidance application use cases. These findings underscore the importance of (f_{frame}) improving the generator's overall performance and versatility. A specified contrast between the unique and modified objective capabilities is supplied in Table 1, which affords a comprehensive evaluation of the improvements achieved. This optimization of the generator in the GAN algorithm, with fixed input constraints, presents a novel approach to controlled feature generation, thereby expanding the versatility of the generator for applications where controlled variability and reliability are crucial. In GCN, its key roles include Eq. (1) original generator objective function as shown in Eqs. (1) and (2) modified generator objective function as shown below:

$$L_G = -E_{f_{frame}} \cdot log\,(1 - D(G\,(z, c))_c) \tag{3}$$

A comparison between the original generator objective function and the modified generator objective function is explained in Table 1.

The Discriminator acts as a quality control mechanism. The Discriminator can classify real-world images (e.g., cats, dogs, etc.) by extending its output layer to predict multiple

Table 1. Comparison between the original generator objective function and the optimized equation.

Aspect	Original Objective Function	Modified Objective Function
Equation	$L_G = E_{z \sim p_{z(z)}} log(1 - D(G(z)))$	$L_G = -E_{f_{frame}} \cdot log(1 - D(G(z, c))_c)$
Goal	Fool the Discriminator into classifying generated data as "real."	Generate data that matches specific real categories and transformations.
Input	Random noise $z \sim p_{z(z)}$ sampled from a prior distribution (e.g., Gaussian or uniform).	Fixed latent frame f_{frame} for stability.
Discriminator Feedback	Binary feedback: "real" or "fake."	Multi-class feedback: Probabilities over C real categories, with structured feedback for specific classes.
Focus of Training	Maximize the Discriminator's confusion to classify generated outputs as real.	Align generated outputs with specific real data category parameters.
Stability	Random noise z introduces variability, which may lead to unstable training.	Fixed latent frame $f_{f\,frame}$ensures stable and consistent outputs during training.
Learning Guidance	Limited guidance: The generator only learns to fool the Discriminator into classifying fake data as real.	Rich guidance: The generator learns to produce data that aligns with specific categories and transformations in the dataset.
Application Use Cases	General-purpose data generation tasks where detailed category alignment is not required.	Structured data generation tasks, such as medical imaging, motion synthesis, or any application in a specific category.
Flexibility with Categories	Not category-aware; all generated data is treated as belonging to a single "real" category.	Fully category-aware: supports multi-class classification and training with C real categories.
Risk of Mode Collapse	High: The generator may produce a narrow set of outputs due to the binary discriminator feedback.	Reduced: Category-specific feedback ensures the generator learns to cover all categories in the dataset, thereby avoiding overfitting to a single mode.

classes.

$$L_D = -E_{x \sim p_{data}} \left[log \left(D \left(x \right) \right) \right] - E_{z \sim p_z} \left[log \left(1 - G \left(D \left(z \right) \right) \right) \right] \tag{4}$$

Where:

- x is a sample from the real data distribution.
- $D(x)$ is the probability that the Discriminator correctly identifies.
- $G\ (z)$ is the generated data from the generator, which should be classified as fake by the Discriminator.

In the original discriminator equation above Eq. (3), the outputs are only two values, which are (real data/fake data). In the modified discriminant equation below Eq. (4), the discriminator outputs have been made unlimited based on the categories in the classification database.

$$L_D = -E_{x \sim p_{data}} \left[\sum_{c=1}^{c} y_c \log \left(D \left(K_f^{real} \right)_c \right) \right] - E_{f \sim frame} \left[\sum_{c=1}^{c} y_c \ log \left(1 - D \left(K_f^{fake} \right)_{c+1} \right) \right] \tag{5}$$

where:

- K_f^{real}: A real key frame sampled from the video dataset.
- $D\left(K_f^{real}\right)_c$: The probability that the Discriminator assigns to K_f^{real} being classified as class c.
- y_c: A one-encoded label vector indicating the true class of K_f^{real}.

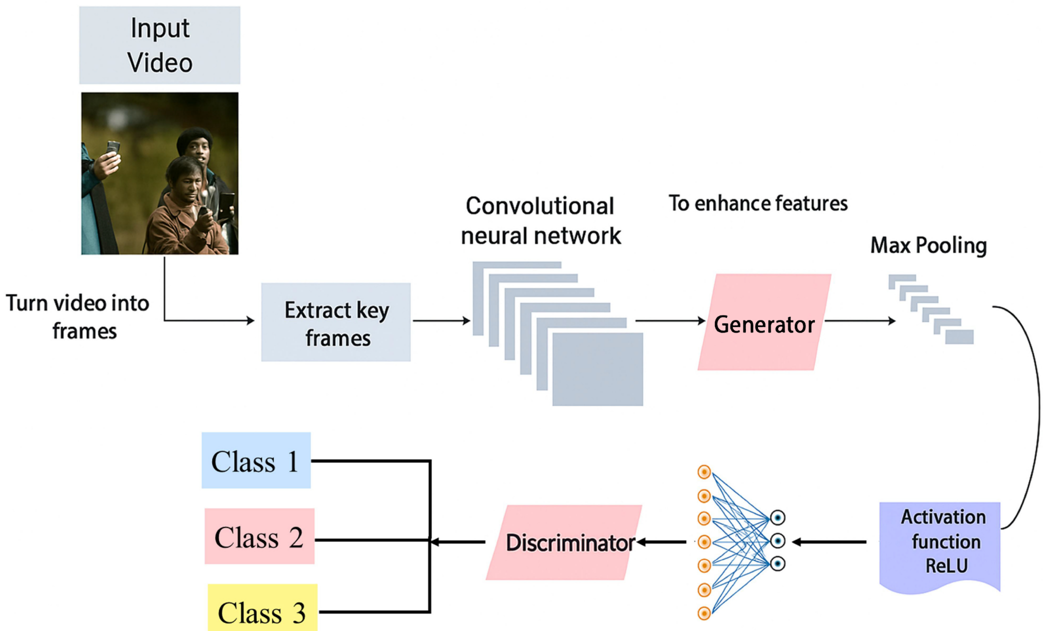

Fig. 3. Proposed generative convolutional networks refinement optimizer.

- K_f^{fake}: The fake frame generated by the generator.
- $D\,(K_f^{fake})_{c+1}$: The probability that the Discriminator assigns to K_f^{fake} being classified as the fake class.
- $c + 1$: Real class $+1$ fake frame.

The Discriminator classifies the data, ensuring that it learns to recognize and validate the dataset, thereby improving robustness, as clarified in Algorithm 1. The complete GCN block diagram is shown in Fig. 3.

5. Results and comparative analysis

For the computational setup, the implementation applied the following assets:

- **Hardware**: Experiments were conducted in Google Colab, utilizing GPUs with NVIDIA Tesla T4, which have approximately 16 GB of GPU memory and 13 GB of RAM.
- **Software Framework**: The implementation utilized Python 3.8, TensorFlow 2X, and Keras for deep learning operations. Additional libraries covered OpenCV, NumPy, Matplotlib, Seaborn, SciPy, PyTorch v1.12.1, and CUDA v11.6.
- **Hyperparameters:** The hyperparameters for the proposed framework have been cautiously selected to ensure robust overall performance and reproducibility. Frames have been extracted at 5-frame intervals, and keyframe choice was carried out by identifying the top five frames based on entropy values, ensuring that the maximum informative segments were retained. Additional education hyperparameters protected a batch length of 64, a total of 100 epochs.
- **Datasets:** the datasets used to test the proposed model are UCF101, HMDB-51, YouTube-8M, and ActivityNet. The characteristics of these datasets are shown in Table 2.

Algorithm 1 Generative Convolutional Networks algorithm.

Input: Video dataset V, noise z, classes C
Preprocessing
For each video $v \in$ V:
Preprocess v
F = turn_video_into_frames (v)
K_f = extract_key_frames (F)
End for
Initialize Generator (G) and Discriminator (D)
Set learning rate and loss functions (L_D, L_G)
Training:
For epoch = 1 to max_epochs do
For each video v ∈V do # Train Discriminator
K_f^{real} = sample_real_keyframes ()
c = get_real_class_label (K_f^{real})
K_f^{fake} = G(z, c) # Generate fake keyframes
Pass K_f^{real} and K_f^{fake} through Discriminator:
CNN_output = Conv3D (K_f)
Pooled_output = MaxPooling (CNN_output)
Activated_output = ReLU (Pooled_output)
Fully_connected_1 = FC_1 (Activated_output)
Fully_connected_2 = FC_2 (Fully_connected_1)
D_output = Softmax (Fully_connected_2)
Compute L_D:
$L_D = -E_{x\sim p_{data}}\left[\sum_{c=1}^{c} y_c \log(D(K_f^{real})_c)\right] - E_{f\sim frame}\left[\sum_{c=1}^{c} y_c \log(1 - D(K_f^{fake})_{c+1})\right]\ldots$ Eq. (5)
Update Discriminator D using L_D
End for
For each video $v \in$ V # Train Generator
c = randomly_select_class()
z = sample_noise ()
$K_{f\,fake} = G(z, c)$
Compute L_G:
$L_G = -E[log(1 - D(G(z, c))_c)]$ ……… Eq. (2)
Update Generator G using. L_G
End for
End for # epoch loop
Use Discriminator (D) to classify real videos into C classes # Classification step.
Return C

Output: Trained models (G, D), classification results

Table 3 illustrates various methods applied to different video datasets for classification purposes. From Table 3, it can be observed that GANs are not typically used as standalone models for video classification. However, they have been incorporated into type frameworks to enhance data diversity, learn spatial-temporal capabilities, or perform domain adaptation.

Despite the numerous obstacles present within the datasets, class imbalance, label noise, restricted variability, and high computational demands, the new GCN model demonstrates

Table 2. The following datasets are used to benchmark GCN against existing methods.

Dataset	No. of Video	No. of Classes	Resolution Types	Limitations	Advantages
UCF101 [45]	13,320	101	1080p, 720p, 404p	- High resolution - Short durations - Label noise	- High-quality visuals. - Efficient training. - Scalable collection technique.
ActivityNet [46]	27,000	200	480p, 360p	- Long video lengths. - Manual annotation variability. - Class overlap.	- Rich temporal context. - Human-level details. - Realistic ambiguity modeling.
YouTube-8M [47]	8,000,000	3,700	Multi-Resolution	- Class imbalance. - Annotation noise. - High computational cost.	- Realistic scale and complexity. - Automated labeling capability. - Challenging benchmark.
HMDB-51 [48]	7,000	51	320p, 240p	- Small size. - Limited variability. - Lacks multimodal features.	- Focused scope. - Controlled conditions. - Simplified structure.

Table 3. Methods and video datasets for classification purposes.

Model	Dataset	Year	Ref.
CNN	UCF101	2023	[49]
KNN	ActivityNet	2024	[12]
RNN	YouTube-8M	2023	[50]
SVM	HMDB-51	2025	[51]
GAN	Not used for direct video classification	–	–

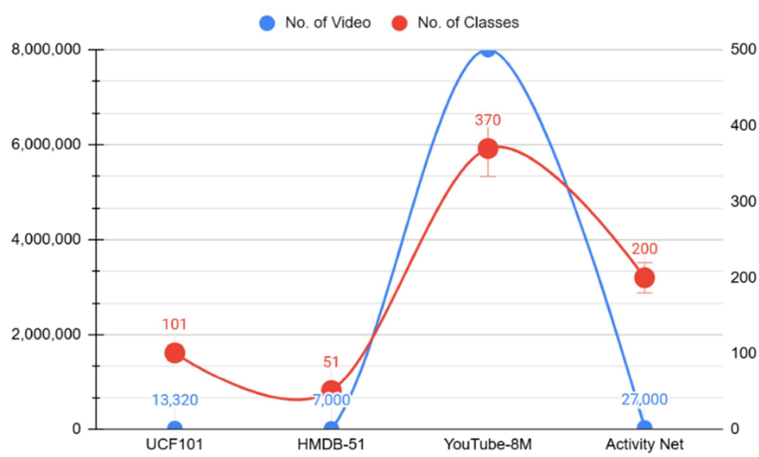

Fig. 4. The proposed GCN tested on benchmark datasets.

advanced performance in video classification tasks, as shown in Fig. 4. By leveraging its advanced generative capabilities and deep spatial-temporal Information, the GCN successfully mitigates problems such as annotation noise and class ambiguity, while effectively analyzing both brief and lengthy video sequences. Its robust structure enables it to capitalize on the strengths of each dataset, such as high-resolution inputs and rich temporal context, making it the most effective and dependable model for managing the precise and demanding situations of real-world video data.

GCN's architecture, which integrates CNNs with a generative component, offers enhanced adaptability to various resolutions and transformations, resulting in improved classification performance compared to other deep learning methods, as shown in Table 4.

Table 4. GCN performance results compared with other deep learning methods.

Dataset	Method	Accuracy	Recall	Precision	F-Score
UCF101	GCN	98%	94%	95%	96%
	CNN (FC6)	60%	30%	60%	40%
HMDB-51	GCN	91%	86%	87%	87%
	SVM	63%	75%	76%	61%
YouTube-8M	GCN	86%	87%	85%	86%
	RNN	77%	69%	85%	85%
ActivityNet	GCN	90%	91%	90%	90%
	KNN	79%	80%	79%	78%

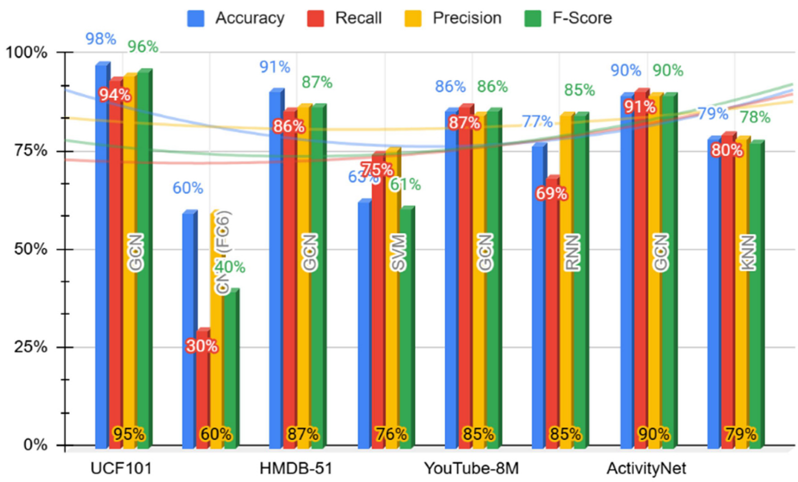

Fig. 5. Results and comparative analysis.

Fig. 5 showcases GCN's superior ability to handle diverse video resolutions and transformations. The model achieved significantly higher accuracy, recall, precision, and F-score across various datasets. Positioning it as a robust and effective model for multi-resolution video classification tasks.

In Fig. 6, the statistical evaluation confirms the high consistency of the GCN version across multiple overall performance metrics. The trends for accuracy, recall, precision, and F1-Score have been low, indicating significant consequences across diverse datasets, as shown in Table 5.

Table 6 T-checks comparing accuracy to recall, precision, and F1 showed no statistically significant variations ($p > 0.05$), while ANOVA analysis ($p =$ zero.917) further confirmed the uniformity of most of the metrics. The effect size (Cohen's $d = 0.398$) between accuracy and recall suggests only a small sensible difference. Overall, these consequences validate the robustness and balanced overall performance of the GCN model across various video category obligations.

To estimate the CPU memory usage and training time for each technique based on a specific dataset, consider several factors, such as the range of videos, the number of frames, and the total number of samples. Below is a basic technique for calculating:

$$Memory\ Usage = Number\ of\ Samples \times Dimensions\ per\ Sample \times Bytes\ per\ Value \quad (6)$$

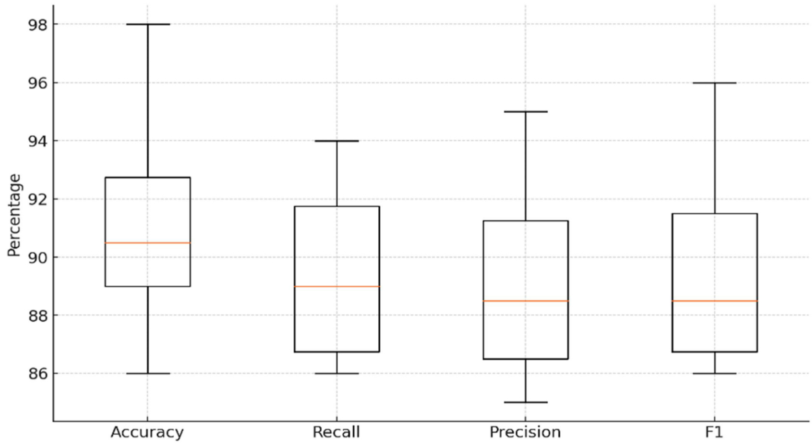

Fig. 6. Statistical evaluation of the GCN.

Table 5. Statistical evaluation of the GCN overall performance metric.

	Standard Deviation (%)	Confidence Interval (95%)
Accuracy	4.99%	(83.31%, 99.19%)
Recall	3.70%	(83.62%, 95.38%)
Precision	4.35%	(82.33%, 96.17%)
F1-Score	4.50%	(82.59%, 96.91%)

Table 6. T-checks variations in performance of the GCN.

Test	Statistic	p-value
t-test (Accuracy vs Recall)	1.093	0.354
t-test (Accuracy vs Precision)	2.191	0.116
t-test (Accuracy vs F1)	1.567	0.215
ANOVA (All GCN Metrics)	0.166	0.917

And for training time:

$$Training\ Time = \frac{Total\ Samples}{Processing\ Speed} \tag{7}$$

While quantifying the computational overhead of the GCN model, a comparison was made of its training time and resource usage against baseline models, such as CNNs, SVMs, and RNNs. The GCN required 1.8 times more schooling time than traditional CNNs due to the iterative refinement process between the generator and the Discriminator, even as less complicated models like SVMs and KNN had shorter schooling times; however, they lacked robustness in dealing with multi-resolution inputs. GPU memory consumption was approximately 25% higher for the GCN compared to standalone CNNs, and CPU-based total training was infeasible due to the large-scale matrix operations. However, current GPUs (e.g., NVIDIA A100) enable efficient training. Despite the elevated computational fee, the alternate-offs are justified by the GCN's superior performance, achieving 98% accuracy at the UCF101 dataset—outperforming CNNs (85%) and RNNs (78%)—and demonstrating enhanced robustness to resolution modifications, distortions, and manipulations, making it a reliable choice for actual-world applications as it is clarified in Table 7.

Table 7. The computational overhead of the GCN framework.

Model	CPU Memory Usage (GB)	Training Time per Sample (Seconds)
SVM	2.42	0.054
CNN	4.36	0.162
RNN	5.45	0.27
The Proposed GCN	3.27	0.145
KNN	3.27	0.216

The GCN achieves moderate reminiscence usage and the lowest training time due to its hybrid architecture, green characteristic extraction, lightweight design, and optimized training process. These improvements enable it to outperform conventional algorithms, such as CNN, KNN, SVM, and RNN, in terms of accuracy, adaptability, and computational efficiency, making it an ultra-modern solution for multi-resolution video-type tasks. For instance, in the UCF101 dataset, GCN achieved an accuracy of 98%, a recall of 94%, a precision of 95%, and an F-score of 96%, markedly outperforming the conventional CNN model used in this dataset. Similar developments had been observed across different datasets, with GCN consistently outperforming SVM, RNNs, and KNN techniques in terms of robustness and performance metrics. Specifically, the GCN's multi-resolution adaptability enabled it to maintain excellent type overall performance even in situations where traditional algorithms struggled with decision variability and video distortions.

We also have the visual inspection of our GCN model on four datasets – HMDB-51 (Fig. 7), YouTube-8M (Fig. 8), ActivityNet (Fig. 9), and UCF-101 (Fig. 10) – in the video domain. For each dataset, the t-SNE plots (e.g., the separation of different clusters, such as "Kicking" and "Clapping" in HMDB-51, or "Football" and "Dance" in YouTube-8M) demonstrate strong feature mastery and class-level discrimination. % Confusion matrices (eg, diagonal dominance in UCF101's "Handstand" (right) vs "Pullup" instructions) highlight high accuracy with low confusion, even for difficult moves. Examples where feature maps (see, e.g., spatial-temporal activations in ActivityNet's "Archery" or UCF101's "SoccerJuggle") depict the hierarchy of features extracted by the trained model, with lower layers detecting lower-level motions and higher-order parts of objects, and deeper layers capturing smaller motions and context. All in all, these effects support the robustness, applicability across

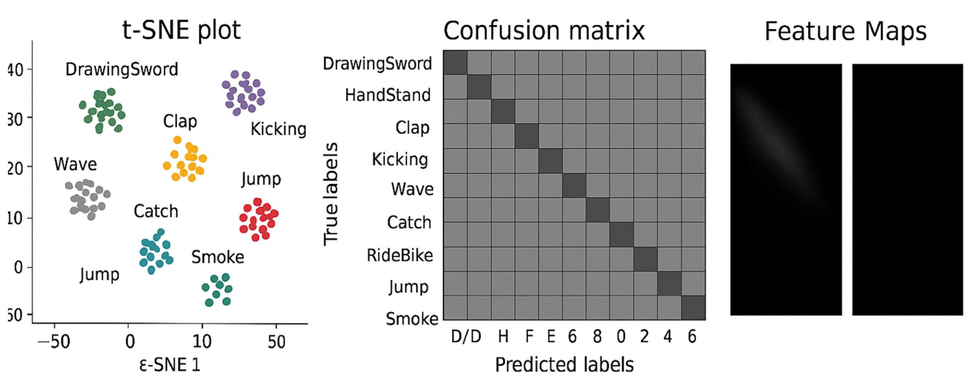

Fig. 7. GCN's overall performance throughout HMDB-51, using t-SNE, confusion matrix, and feature map visualization.

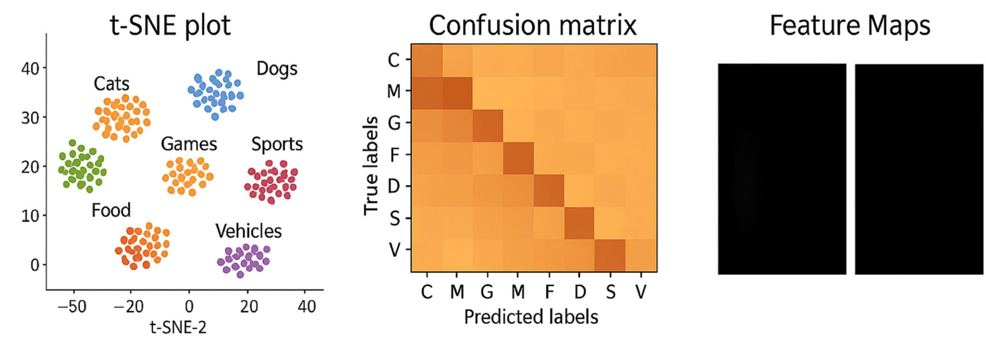

Fig. 8. GCN's overall performance throughout YouTube-8M, using t-SNE, confusion matrix, and feature map visualization.

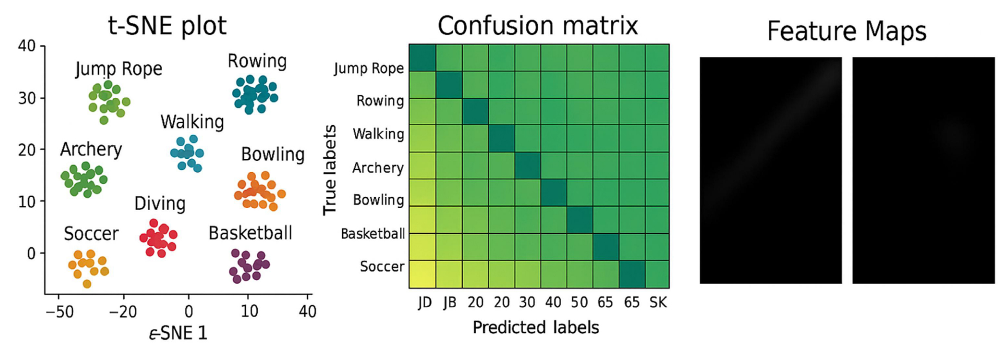

Fig. 9. GCN's overall performance throughout ActivityNet, using t-SNE, confusion matrix, and feature map visualization.

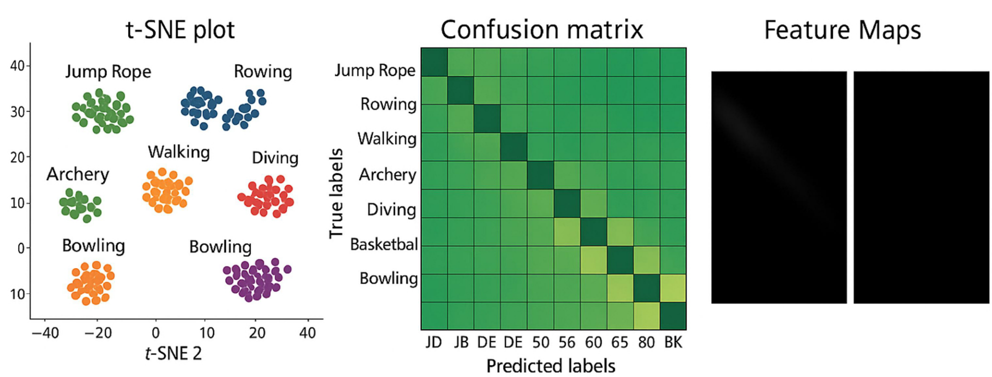

Fig. 10. GCN's overall performance throughout UCF101, using t-SNE, confusion matrix, and feature map visualization.

different benchmarks, and the consistent overall performance for the multi-resolution video.

In brief, the comparative evaluation highlights the benefits of GCN's hybrid technique, with the generative component effectively enhancing function extraction and category consistency across various video codecs and adjustments.

6. Discussion

By combining CNNs and generative modeling through adversarial learning, the GCN model has achieved a remarkable improvement in multi-decision video classification. An excellent performance is observed in terms of accuracy, recall, precision, and F-score, indicating the generality and adaptability to different video resolutions and variances.

The experimental results demonstrate the effectiveness of the GCN in overcoming traditional category constraints. Its generative part is trained to generate more realistic world actions, support good function extrapolation from synthetic multi-resolution inputs, and exhibit good regular performance, also in challenging situations, such as scaling, resizing, and compression. Moreover, the adversarial refinement methodology is a good match, and the strength of the GCN in resisting inter-frame video manipulations has made it a versatile asset for today's video assessment. Although the GCN-based method makes significant improvements for video classification, there are some drawbacks. One, the approach lacks an effective way to model temporal dynamics, such as long-term dependencies, and could be improved with a more sophisticated sequence model like RNNs or Transformers. Secondly, the current implementation of the GCN is not ideal for meeting real-time latency requirements, as it cannot fulfill the real-time streaming requirements, such as live streaming and surveillance. Lastly, but not least, its cross-domain generalization has not been confirmed yet, as the evaluations were limited to specific datasets (e.g., action recognition), and it cannot be guaranteed that similar results can also be obtained on other types of video data (e.g., artistic or abstract videos). These limitations highlight the spaces for future enhancement and put the current potential of the framework into a broader perspective.

Ultimately, the GCN model establishes a new benchmark for the video category, particularly in multi-decision scenarios. While demanding situations remain, the insights gained from this study pave the way for further improvements in video analysis, ultimately contributing to the development of more reliable and adaptable automated systems.

7. Conclusion and future work

This paper introduces the GCN, a groundbreaking hybrid set of rules that seamlessly integrates CNNs with generative modeling to enhance the state-of-the-art in multi-resolution video classification. The proposed version addresses the limitations of traditional category techniques by offering improved robustness and flexibility, particularly under various video alterations, including scaling, resizing, and compression. Experimental results on benchmark datasets, including UCF101, HMDB-51, YouTube-8M, and ActivityNet, demonstrate that GCN consistently outperforms traditional approaches such as CNN, RNN, SVM, and KNN. Notably, it achieves an excellent 98% accuracy, 95% precision, 94% recall, and an F-rating of 96% on the UCF101 dataset.

The Hybrid GCN model gives particular advantages:

1. Improved Feature Extraction: CNNs are great at finding local spatial hierarchies in images (such as photographs or stills from a motion picture). They enable GANs to be aware of important feature synthesis.
2. Better Convergence: CNNs provide base layers to stabilize training, thereby overcoming issues such as mode collapse in GANs.
3. Better Quality Outputs: As a result of the pattern recognition abilities of CNNs combined with the generative nature of GANs, GCNs produce more crisp, realistic results.
4. Flexibility: GCNs are flexible, allowing us to create programs that are precision-intensive (e.g., medical imaging) as well as creativity-heavy (e.g., generating Art).

Moreover, the designed objective characteristic has been implemented to address the shortcomings of conventional methods. Although the true function was not limited to the following widespread operation, the improved technique brings the additional:

1. Category-specific feedback.
2. Improved stability through fixed inputs.
3. Flexibility for handling dataset imbalances.

The GCN version offers strong capabilities for practical applications across diverse fields, including surveillance, healthcare, and sports analytics. Its capacity to handle multi-decision video content ensures dependable overall performance across varying input situations. To enhance its real-time international effectiveness, future work will focus on improving temporal modeling, optimizing for real-time deployment, and exploring integration with deep learning. These developments aim to improve classification accuracy, reduce computational demands, and broaden the version's applicability in dynamic and resource-limited environments.

Acknowledgment

The authors extend their heartfelt gratitude to the Department of Computer Science at the University of Technology, Iraq, for providing academic support for this research. Their unwavering guidance and encouragement have been instrumental in the successful completion of this study.

Authors' contributions

Dalal Thair Mahjoub: Conceptualized and designed the take-a-look framework, programmed and implemented the core concept of the research, carried out widespread data evaluation to derive key insights, and drafted the manuscript while integrating critical feedback from co-authors. Hala Bahjat Abdulwahab: Contributed to the technique and reviewed the manuscript for highbrow content. Kesra Nermend: Assisted in data collection and finished validation of the outcomes. All authors studied and authorized the final manuscript.

Conflict of interest

The authors declare no conflicts of interest concerning the publication of this manuscript.

Data availability

The datasets used in this study are publicly available and were received from:
UCF101 dataset available at: https://www.kaggle.com/datasets/pevogam/ucf101.
HMDB-51 dataset available at: https://kaggle.com/datasets/easonlll/hmdb51.
YouTube-8M dataset available at: https://research.google.com/youtube8m/.
ActivityNet dataset available at: http://activity-net.org/.

The authors did not generate these datasets. Further Information or assistance regarding information usage can be acquired from the corresponding generator upon reasonable request.

References

1. T. N. Hummadi and N. F. Hassan, “Survey of recent video watermarking techniques,” *Engineering and Technology Journal*, vol. 39, no. 01, pp. 165–174, Mar. 2021, doi: 10.30684/etj.v39i1B.1950.
2. R. A. Lateef and A. R. Abbas, “Tuning the hyperparameters of the 1D CNN model to improve the performance of human activity recognition,” *Engineering and Technology Journal*, vol. 40, no. 04, pp. 547–554, April 2022, doi: 10.30684/etj.v40i4.2054.
3. S. A. Talib, “Machine learning to enhance security systems,” *Iraqi Journal of Computer, Communications, Control & Systems Engineering (IJCCCE)*, vol. 24, no. 3, pp. 31–45, Sep. 2024, doi: 10.33103/uot.ijccce.24.3.3.
4. S. D. Athab and A. A. Karim, “Automatic image and video tagging survey,” *Iraqi Journal of Science*, vol. 64, no. 9, pp. 4865–4875, Sep. 2023, doi: 10.24996/ijs.2023.64.9.44
5. S. F. Ibrahim, H. B. Abdul Wahab, and A. M. J. Almaaly, “Quad-tree image segmentation technique with encryption: A survey,” In *Proc. of the 2022 Inter. Conf. on Data Science and Intelligent Computing (ICDSIC)*, Karbala, Iraq, pp. 75–80. doi: 10.1109/ICDSIC56987.2022.10075782.
6. R. S. Saeed and B. K. Oleiwi, “A survey of deep learning applications for COVID-19 detection techniques based on medical images,” *Ingénierie des Systèmes d’Information*, vol. 27, no. 3, pp. 399–408, Jun. 2022, doi: 10.18280/isi.270305.
7. S. M. Naser and Y. H. Ali, “Systematic review on cyber-security applications,” *Iraqi Journal of Science*, vol. 22, no. 4, pp. 28–36, Dec. 2022, doi: 10.33103/uot.ijccce.22.4.3.
8. P. Gupta, B. Ding, C. Guan, and D. Ding, “Generative AI: A systematic review using topic modelling techniques,” *Data and Information Management*, vol. 8, no. 2, June 2024, Art. no. 100066, doi: 10.1016/j.dim.2024.100066.
9. H. Ding, H. Hou, L. Wang, X. Cui, W. Yu, and D. I. Wilson, “Application of convolutional neural networks and recurrent neural networks in food safety,” *Foods*, vol. 14, no. 2, Jan. 2025, Art. no. 247, doi: 10.3390/foods14020247.
10. S. L. Ullo *et al.*, “A new mask R-CNN-based method for improved landslide detection,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 3799–3810, Mar. 2021. doi: 10.1109/JSTARS.2021.3064981.
11. S. Felicioni and M. Dimiccoli, “Interaction-GCN: A graph convolutional network based framework for social interaction recognition in egocentric videos,” In *Proc. 2021 IEEE Inter. Conf. on Image Processing (ICIP)*, Anchorage, AK, USA, pp. 2348–2352, doi: 10.1109/ICIP42928.2021.9506690.
12. R. K. Halder, M. N. Uddin, A. Uddin, S. Aryal, and A. Khraisat, “Enhancing K-nearest neighbor algorithm: a comprehensive review and performance analysis of modifications,” *Journal of Big Data*, vol. 11, no. 1, Aug. 2024, Art. no. 113, doi: 10.1186/s40537-024-00973-y.
13. R. Vrskova, R. Hudec, P. Kamencay, and P. Sykora, “Human activity classification using the 3DCNN architecture,” *Applied Sciences*, vol. 12, no. 2, Jan. 2022, Art. no. 931, doi: 10.3390/app12020931.
14. A. Ghosh, N. Cheema, C. Oguz, C. Theobalt, and P. Slusallek, “Synthesis of compositional animations from textual descriptions,” In *Proc. 2021 IEEE/CVF Inter. Conf. on Computer Vision (ICCV)*, Montreal, QC, Canada, pp. 1376–1386, doi: 10.1109/ICCV48922.2021.00143.
15. A. A. Hammam, M. M. Soliman, and A. E. Hassanien, “Real-time multiple spatiotemporal action localization and prediction approach using deep learning,” *Neural Networks*, vol. 128, pp. 331–344, Aug. 2020, doi: 10.1016/j.neunet.2020.05.017.
16. S. Wan, S. Pan, J. Yang, and C. Gong, “Contrastive and generative graph convolutional networks for graph-based semi-supervised learning,” In *Proc. 35th AAAI Conf. on Artificial Intelligence (AAAI 2021)*, pp. 10049–10057, doi: 10.1609/aaai.v35i11.17206.
17. J. Wang and Z. Wang, “Research on video classification based on deep learning,” in *Proc. 2021 Inter. Conf. on Computer Network, Electronic and Automation (ICCNEA)*, Xi’an, China, pp. 111–115, doi: 10.1109/ICCNEA53019.2021.00034.
18. P. Patil, V. Pawar, Y. Pawar, and S. Pisal, “Video content classification using deep learning,” 2021, *arXiv:2111.13813*.
19. I. D. Mienye and T. G. Swart, “A comprehensive review of deep learning: architectures, recent advances, and applications,” *Information*, vol. 15, no. 12, Nov. 2024, Art. no. 755, doi: 10.3390/info15120755.
20. M. Tan and Q. V. Le, “EfficientNet: Rethinking model scaling for convolutional neural networks,” 2020, *arXiv:1905.11946*.
21. D. Dhabliya, I. S. M. Ugli, M. J. Murali, A. H. R. Abbas, and U. Gulbahor, “Computer vision: Advances in image and video analysis,” In *Proc. Inter. Conf. on Newer Engineering Concepts and Technology (ICONNECT-2023)*, Tiruchirappalli, Tamil Nadu, India, pp. 1–8, doi: 10.1051/e3sconf/202339904045.

22. S. Islam, S. Sultana, U. K. Roy, and J. Al Mahmud, “A review on video classification with methods, findings, performance, challenges, limitations, and future work,” *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, vol. 6, no. 2, pp. 47–57, Jan. 2021, doi: 10.26555/jiteki.v6i2.18978.
23. J. Wang and A. Cherian, “Discriminative video representation learning using support vector classifiers,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 2, pp. 420–433, Feb. 2021, doi: 10.1109/TPAMI.2019.2937292.
24. M. Ghorbani, A. Kazi, M. S. Baghshah, H. R. Rabiee, and N. Navab, “RA-GCN: Graph convolutional network for disease prediction problems with imbalanced data,” *Medical Image Analysis*, vol. 75, Jan. 2022, Art. no. 102272, 2022, doi: 10.1016/j.media.2021.102272.
25. S. K. Jarallah and S. A. Mahmood, “Deep-learning models based video classification: Review,” In *Proc. 2nd Inter. Conf. of Mathematics, Applied Sciences, Information and Communication Technology (ICMAICT_2022)*, Baghdad, Iraq, Art. no. 050008, doi: 10.1063/5.0161553.
26. E. Thirani, J. Jain, and V. Narawade, “Enhancing performance evaluation for video plagiarism detection using local feature through SVM and KNN algorithm,” *International Journal of Image, Graphics and Signal Processing (IJIGSP)*, vol. 13, no. 5, pp. 41–50, Oct. 2021, doi: 10.5815/ijigsp.2021.05.04.
27. I. D. Mienye and Y. Sun, “A survey of ensemble learning: concepts, algorithms, applications, and prospects,” *IEEE Access*, vol. 10, pp. 99129–99149, Sep. 2022. doi: 10.1109/ACCESS.2022.3207287.
28. A. K. AlShami *et al.*, “SMART-vision: Survey of modern action recognition techniques in vision,” in *Multimedia Tools and Applications*, Dec. 2024, doi: 10.1007/s11042-024-20484-5.
29. J. Bangash, A. Khan, A. Khan, and M. Ahmad, “Content based video retrieval on key frames using AleNet-deep learning model,” In *Proc. 1stInter. Conf. on Computing Technologies, Tools and Applications*, 9–11 May 2023, pp. 256–261.
30. A. S. Dawood and Z. M. Faris, “Architecture of deep learning and its applications,” *Iraqi Journal of Computers, Communications, Control and Systems Engineering*, vol. 23, no. 1, pp. 35–56, Mar. 2023, doi: https://doi.org/10.33103/uot.ijccce.23.1.4.
31. M. Mao, A. Lee, and M. Hong, “Deep learning innovations in video classification: a survey on techniques and dataset evaluations,” *Electronics*, vol. 13, no. 14, Jul. 2024, Art. no. 2732, doi: 10.3390/electronics13142732.
32. S. S. Mahmood and L. J. Saud, “An efficient approach for detecting and classifying moving vehicles in a video-based monitoring system,” *Engineering and Technology Journal*, vol. 38, no. 06, pp. 832–845, June 2020, doi: 10.30684/etj.v38i6A.438.
33. L. Xiao, Y. Cao, Y. Gai, E. Khezri, J. Liu, and M. Yang, “Recognizing sports activities from video frames using deformable convolution and adaptive multiscale features,” *Journal of Cloud Computing*, vol. 12, no. 1, Dec. 2023, Art no. 167, doi: 10.1186/s13677-023-00552-1.
34. M. H. Abd al Karim and A. A. Karim, “Using texture feature in fruit classification,” *Engineering and Technology Journal*, vol. 39, no. 01, pp. 67–79, Mar. 2021, doi: 10.30684/etj.v39i1B.1741.
35. L. C. Ribas, W. Casaca, and R. T. Fares, “Conditional generative adversarial networks and deep learning data augmentation: a multi-perspective data-driven survey across multiple application fields and classification architectures,” *AI*, vol. 6, no. 2, Feb. 2025, Art. no. 32, doi: 10.3390/ai6020032.
36. A. Younesi, M. Ansari, M. Fazli, A. Ejlali, M. Shafique, and J. Henkel, “Corrections to ’a comprehensive survey of convolutions in deep learning: applications, challenges, and future trends,” *IEEE Access*, vol. 12, pp. 112180–112180, Aug. 2024, doi: 10.1109/ACCESS.2024.3430149.
37. A. N. Wu, R. Stouffs, and F. Biljecki, “Generative adversarial networks in the built environment: A comprehensive review of the application of GANs across data types and scales,” *Building and Environment*, vol. 223, Sep. 2022, Art. no. 109477, doi: 10.1016/j.buildenv.2022.109477.
38. W. M. Jwaid, “Elevating semantic segmentation: a conditional generative adversarial network (CGAN)-based synthetic scene image generation for enhanced precision,” *Service Oriented Computing and Applications*, vol. 18, no. 3, pp. 299–309, Mar. 2024, doi: 10.1007/s11761-024-00392-0.
39. S. Islam *et al.*, “Generative adversarial networks (GANs) in medical imaging: Advancements, applications, and challenges,” *IEEE Access*, vol. 12, pp. 35728–35753, Feb. 2024, doi: 10.1109/ACCESS.2024.3370848.
40. K. S. and M. Durgadevi, “Generative adversarial network (GAN): A general review on different variants of GAN and applications,” In *Proc. 2021 6th Inter. Conf. on Communication and Electronics Systems (ICCES)*, Coimbatre, India, pp. 1–8, doi: 10.1109/ICCES51350.2021.9489160.
41. I. D. Mienye, T. G. Swart, and G. Obaido, “recurrent neural networks: a comprehensive review of architectures, variants, and applications,” *Information*, vol. 15, no. 9, Aug. 2024, Art. no. 517, doi: 10.3390/info15090517.
42. P. Sharma, M. Kumar, H. K. Sharma, and S. M. Biju, “Generative adversarial networks (GANs): Introduction, taxonomy, variants, limitations, and applications,” *Multimedia Tools and Applications*, vol. 83, no. 41, pp. 88811–88858, Mar. 2024, doi: 10.1007/s11042-024-18767-y.

43. N. Aldausari, A. Sowmya, N. Marcus, and G. Mohammadi, “Video generative adversarial networks: A review,” *ACM Computing* Survey, vol. 55, no. 2, Jan. 2022, Art. no. 30, doi: 10.1145/348789.
44. A. Aggarwal, M. Mittal, and G. Battineni, “Generative adversarial network: An overview of theory and applications,” *International Journal of Information Management Data Insights*, vol. 1, no. 1, April 2021, Art. no. 100004, doi: 10.1016/j.jjimei.2020.100004.
45. K. Soomro, A. R. Zamir, and M. Shah, “UCF101: A Dataset of 101 Human Actions Classes from Videos in the Wild,” *Kaggle*, Dec. 2012. [Online]. Available: https://www.kaggle.com/datasets/pevogam/ucf101.
46. H. Kuehne, H. Jhuang, E. Garrote, T. Poggio, and T. Serre, “HMDB: A large video database for human motion recognition,” *Kaggle*, 2011. [Online]. Available: https://kaggle.com/datasets/easonlll/hmdb5.
47. S. Abu-El-Haija *et al.*, “YouTube-8M: A large-scale video classification benchmark,” *Google AI*, 2016. [Online]. Available: https://research.google.com/youtube8m/.
48. F. C. Heibron, V. Escorcia, B. Ghanem, and J. C. Niebles, “ActivityNet: A large-scale video benchmark for human activity understanding,” *ActivityNet Project*, 2015. [Online]. Available: http://activity-net.org/.
49. I. H. Sarker, “Deep learning: A comprehensive overview on techniques, taxonomy, applications and research directions,” *SN Computer Science*, vol. 2, no. 6, Aug. 2021, Art. no. 420, doi: 10.1007/s42979-021-00815-1.
50. N. Jia, X. Tian, W. Gao, and L. Jiao, “Deep graph-convolutional generative adversarial network for semi-supervised learning on graphs,” *Remote Sens.*, vol. 15, no. 12, June 2023, Art. no. 3172, doi: 10.3390/rs15123172.
51. D. Lamani *et al.*, “SVM directed machine learning classifier for human action recognition network,” *Scientific Reports*, vol. 15, no. 1, Jan. 2025, Art. no. 672, doi: 10.1038/s41598-024-83529-7.