Original Article
ISSN (Online): 2582-7472

ShodhKosh: Journal of Visual and Performing Arts
January-June 2026 7(1), 75–89

GWO-ATDL-SBD: A GREY WOLF OPTIMIZED ADAPTIVE TEMPORAL DEEP LEARNING FRAMEWORK FOR ROBUST SHOT BOUNDARY DETECTION

Tushar Banik [1], Saptarshi Chakraborty [2], Dalton Meitei Thounaojam [3], Sapam Jitu Singh [4], Raju Rajkumar [5]

[1] Department of Computer Science and Engineering, ICFAI University, Tripura, India
[2] Department of Computer Science and Engineering, ICFAI University, Tripura, India
[3] Department of Computer Science, Manipur University, India
[4] Department of Computer Science and Engineering, Manipur University, India
[5] Department of Computer Science, Pravabati College, Imphal, India

Received 15 February 2026
Accepted 20 April 2026
Published 16 May 2026

Corresponding Author
Tushar Banik, btushar72@gmail.com

DOI
10.29121/shodhkosh.v7.i1.2026.8167

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

ABSTRACT

Shot Boundary Detection (SBD) is an essential task in video content analysis but accurately detecting both abrupt and gradual scene changes is still a problem due to different lightings, camera motions, and varied scene contents. This paper introduces a hybrid framework named GWO-ATDL-SBD that combines Grey Wolf Optimizer (GWO) with adaptive temporal deep learning for performing robust shot boundary detection. At first, multi-cue features such as edge energy difference, motion vector entropy, and color histogram distance are extracted to represent different visual characteristics in combination. These features are combined dynamically using weights obtained by GWO and passed to a Bi-directional Long Short-Term Memory (Bi-LSTM) network to capture the frame-wise temporal dependencies. In contrast to the traditional methods having fixed thresholds, our method uses GWO not only to find the optimal feature fusion weights and classifier parameters but also decision thresholds for adaptive and accurate recognition of both cuts and transitions. Extensive experiments on standard video datasets show that the proposed method substantially surpasses the existing SBD methods in terms of precision, recall, and F1-score. The findings reveal the potential of marrying evolutionary optimization with temporal deep learning for tackling complex video transition detection problems.

Keywords: Shot Boundary Detection, Grey Wolf Optimizer, BI-LSTM, Video Analysis, Temporal Deep Learning

1. INTRODUCTION

The fast expansion of digital video content through multimedia platforms, surveillance systems, and online streaming services inevitably leads to a larger need for efficient and trustworthy techniques of video content analysis. SBD is a core step of preprocessing for video indexing, summarization, retrieval, and event detection, as it segments a video sequence into semantically meaningful shots. Precisely recognizing shot boundaries has a great impact on the effectiveness of higher-level video understanding tasks. On the contrary, strong detection of abrupt and gradual

transitions is still a difficult issue because of illumination changes, camera motion, object dynamics, and complex scene changes.

Shot boundary detection was initially achieved by algorithms working on basic visual features such as color histograms, edge statistics, pixel-wise differences, etc., together with fixed or heuristic thresholds. Despite being very fast, these methods are so sensitive to lighting variations or camera movement that they produce many false alarms. To get rid of these problems, machine learning, based methods were developed, utilizing classifiers such as Support Vector Machines and Hidden Markov Models, which allow learning from data. Today, deep learning models, especially convolutional and recurrent neural networks, have revealed superior capability in understanding, visual patterns, and temporal dependencies. However, the performance of deep learning, based SBD methods, which comes with a considerably high computational cost, overfitting, and the requirement of manually tuned thresholds or huge annotated datasets, is a constant concern.

Evolutionary optimization algorithms have been used to improve shot boundary detection by automatically adjusting weights of features and parameters of decisions. One of these is GWO, which has been recognized by many due to its simple architecture, rapid convergence, and good trade-off between exploration and exploitation. However, current hybrid GWO-based SBD methods usually incorporate deep network training into the optimization cycle, causing an extremely high computational complexity that hardly makes them suitable for practical use, in particular, for long video sequences.

Driven by these issues, this study introduces a new hybrid framework called GWOATDL-SBD that combines Grey Wolf Optimization with adaptive temporal deep learning for an effective and computationally efficient shot boundary detection. Initially, the proposed method derives multi-cue features that are complementary to each other such as edge energy difference, motion vector entropy, and color histogram divergence, representing changes in structure, motion, and color between two consecutive frames. A Bi-LSTM network is trained only once to capture temporal dependencies and produce frame-level transition probabilities. In order to decrease the computational cost and at the same time keep the detection accuracy, the Grey Wolf Optimizer is used only for optimizing the weights of feature fusion and the adaptive decision threshold instead of deep model retraining in the optimization loop.

The most significant contributions of this paper are the following:

1) Introducing a hybrid SBD framework that leverages adaptive temporal deep learning and Grey Wolf Optimization to achieve a significantly low computational cost.
2) Proposing a multi-cue feature fusion method that integrates edge, motion, and color information in order to accurately detect both sudden and gradual changes in the scene.
3) Developing an optimized GWO-based parameter tuning method that not only removes the need for manual thresholding but also drastically lowers the time complexity.
4) Thorough experimental testing on several datasets shows the proposed approach outperforms other state-of-the-art methods in SBD.

The rest of the paper is structured in the following way. In section 2, we summarize previous literature on video shot transition detection. Section 3 introduces our GWOATDL-SBD method along with the main components. In section 4, the authors explain the optimization technique and the related mathematical expressions. Section 5 shows the experiments and evaluates the results. At last, section 6 finishes the paper and suggests the possible future research areas.

2. RELATED WORK

Over the last several decades, video indexing, summarization, and retrieval have been heavily dependent on SBD which has been studied thoroughly. The existing approaches fall broadly under four categories, namely classical feature-based methods, machine learning, based techniques, deep learning, based models, and optimization-driven hybrid frameworks.

The initial research on shot boundary detection was mainly concentrated on a low level of visual features like pixel intensity differences, color histograms, and edge statistics. Techniques based on histograms calculate the dissimilarity between consecutive frames by using measures like the Euclidean distance [1] or the chi-square distance [2], and they detect abrupt transitions by applying preset thresholds [3]. Methods based on edges take advantage of the structural

changes in the frames, and as a result, they are more robust to illumination changes than pixel-based methods [4, 5]. There were also motion-based methods that utilized optical flow or motion vectors to detect changes in the dynamic scenes, especially gradual transitions [6].

Traditional methods, although very efficient computationally, are quite vulnerable to factors such as camera motion, changes of lighting, and object movement. They usually produce a lot of false positives when dealing with complex videos. Besides, their use of static thresholds makes it very difficult for them to adapt to different types of video content [7].

In order to break through the bottleneck posed by handcrafted threshold-based methods, the community stepped up with supervised machine learning (ML) models for shot boundary detection. Support Vector Machines (SVMs) trained on handcrafted features like color moments, texture descriptors, and edge information were able to yield significant improvements in detection accuracy [8, 9]. Moreover, Hidden Markov Models (HMMs) and Conditional Random Fields (CRFs) were employed to represent temporal relations of frames and transitions [10, 11]. Besides that, fuzzy logic and rule-based classifiers were investigated to cater to gradual transition's uncertainty [12].

Though approaches based on machine learning have shown better robustness over traditional methods, their performance largely hinges on feature selection and the availability of labeled training data. Moreover, simple classifiers usually are not able to model long-term temporal patterns that are natural in slow transitions [13].

The field of shot boundary detection has been majorly influenced by the recent progress of deep learning. Researchers have used Convolutional Neural Networks (CNNs) to learn discriminative spatial features directly from raw frames or frame differences [14, 15]. In order to model temporal sequences, Recurrent Neural Networks (RNNs), especially Long Short-Term Memory (LSTM) and Bi-directional LSTM models, have been combined with CNN-based feature extractors [16, 17]. Threedimensional CNNs (3D-CNNs) and temporal convolutional networks have been the two main architectures exploiting the spatial as well as the temporal information and jointly very effectively thus leading to the increased performance of the methods [18].

Deep learning, based SBD methods can get state-of-the-art accuracy but usually, they need a huge amount of annotated data and the computational cost is quite high. Besides that, a lot of methods depend on manually set decision thresholds or post-processing heuristics that make the method less generalizable to different video domains [19].

Evolutionary algorithms and swarm intelligence methods have been considered for solving feature weight optimization, classifier parameter setting, and decision threshold selection problems in shot boundary detection. Genetic Algorithms (GA), Particle Swarm Optimization (PSO), and Ant Colony Optimization (ACO) have been extensively used to improve detection accuracy and simultaneously decrease false positives [20–22]. Out of these, the Grey Wolf Optimizer (GWO) has attracted more attention because of its straightforward implementation, quick convergence, and well-balanced exploration, exploitation nature [23].

Some hybrid methods integrating GWO with ML classifiers, e.g., neural networks and SVMs, for video transition detection have been suggested in [24]. But a lot of hybrid models in the literature wrap the training of deep networks inside the optimization loop, thus resulting in very high computational cost and scalability problems, mainly in the case of long video sequences [25].

From the review above, one can see that on the one hand, classical methods are not robust and on the other hand, the use of machine learning is limited by handcrafted features. Deep learning models, however, have a problem with huge computational overload and parameter sensitivity. Although optimization-based hybrid frameworks, on the one hand, increase the adaptability of the systems, on the other hand, their integration with deep learning models, in most cases, causes impractical runtimes.

The above-mentioned restrictions are the reasons for creating a more computationally efficient hybrid model that utilizes the temporal modeling ability of deep learning and evolutionary optimization only for a very small parameter tuning. The new GWOATDL-SBD framework fills the void in that it runs the Bi-LSTM network for one time and uses the Grey Wolf Optimization only for the feature fusion and adaptive threshold selection, thus, it can give a higher accuracy level with a drastically lower computational expenditure.

3. PROPOSED GWO-ATDL-SBD FRAMEWORK

The goal of the new model is to correctly identify sudden and slow changes of shot, at the same time, being fast for computation and audacious to changes in lighting, camera movements, and varied scenes. Figure 1 shows the schematic of the proposed framework.

Figure 1

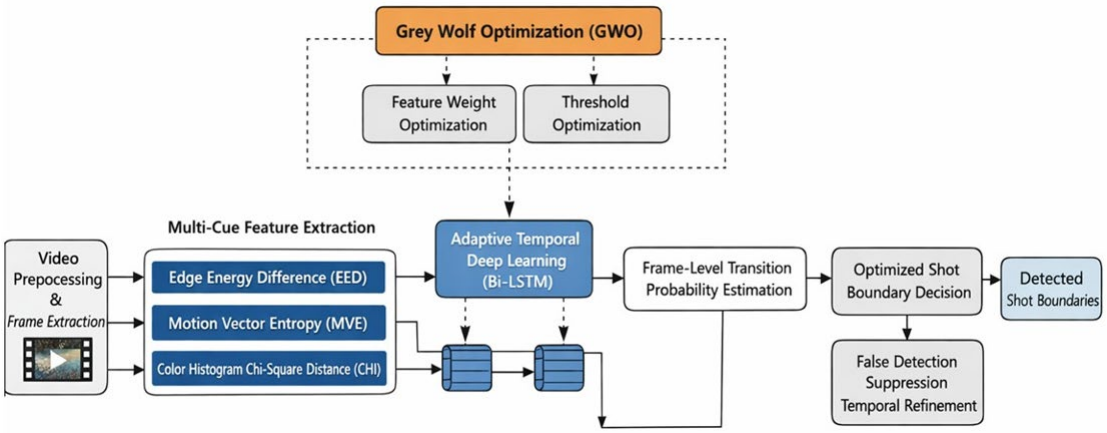

Figure 1 System Architecture of the Proposed System

Algorithm 1 breaks down the task of detecting shot boundaries into building features, estimating temporal probabilities with a pre-trained Bi-LSTM, and automatically choosing parameters through Grey Wolf Optimization. The method manages to identify both abrupt and gradual transitions effectively with the least amount of computation, by limiting the optimizer only to the weights for feature fusion and the decision threshold and freezing temporal learning. Stability and accuracy are further improved by constraint-aware updates and temporal post-processing.

Algorithm 1 GWO-ATDL-SBD: Grey Wolf Optimized Adaptive Temporal Deep Learning–based Shot Boundary Detection

Require: Video sequence V = {F1,F2,...,FN}
Require: GWO population size P, maximum iterations T
Require: Bi-LSTM hyper parameters
Ensure: Detected shot boundaries S = {s1,s2,...,sK}
1: V ← {F1,F2,...,FN}
2: **for t** = 2 to N **do**
3: EED(t) ← Ft − Ft−1 ▷Compute Edge Energy Difference
4: **end for**
5: X(t) = [EED(t), MV E(t), CHI(t)] ▷Normalize feature matrix
6: **for i** = 1 to P **do**
7: Xi = {w1,i,w2,i,w ,i,θi} ▷Initialize wolf position
8: Enforce w1,i + w2,i + w3,i = 1, 0 ≤ θi ≤ 1
9: **end for**
10: **for i** = 1 to P **do**
11: Fi(t) = w1,iEED(t) + w2,iMV E(t) + w3,iCHI(t) ▷Compute fused feature
12: P(t) > θi ▷Primary Detection Condition
13: **end for**
14: **for** iter = 1 to T **do**

15: Identify alpha (α), beta (β), and delta (δ) wolves
16: **for i** = 1 to P **do**
17: $X_i^{(itr+1)} = \frac{X_\alpha + X_\beta + X_\delta}{3}$ ▷Update position
18: Enforce constraints on w1,i,w2,i,w3,i,θi
19: **end for**
20: **end for**
21: $X^* = \{w_1^*, w_2^*, w_3^*, \theta^*\}$ ▷Select best wolf
22: $F^*(t) = w_1^* EED(t) + w_2^* MVE(t) + w_3^* CHI(t)$ ▷Compute final fused feature
23: $P(t) > \theta^*$ ▷Final Detection Condition
24: **Output**
25: **return** Final shot boundary set S

3.1. VIDEO PREPROCESSING AND FRAME EXTRACTION

Initially, the input video is decoded to extract frames at a fixed frame rate. To cut the processing time significantly a little, each frame is changed into grayscale while retaining the structural characteristics that are closely connected to the detection of the transition between scene boundaries. The preprocessing step achieves compatibility among various video formats and resolutions and brings about efficient feature extraction.

3.2. MULTI-CUE FEATURE EXTRACTION

Three complementary low-level features are extracted from consecutive frames in order to robustly characterize shot transitions. These features capture changes in structure, motion, and color.

3.2.1. EDGE ENERGY DIFFERENCE

Edge details work well for spotting the structural changes between frames. Sobel edge operators are applied to two consecutive frames, and then the edge energy difference is calculated as

$$EED(t) = |\sum_{x,y} \nabla F_t(x,y) - \nabla F_{t-1}(x,y)| \quad (1)$$

where ∇Ft represents the gradient magnitude of the frame Ft. Abrupt transitions most of the time produce very large changes in edge energy, on the other hand, gradual transitions show fairly smooth variations.

3.2.2. MOTION VECTOR ENTROPY

Indeed, motion information plays a major role in differentiating between the camera motion and the object movement on one hand, and the actual shot transitions on the other. To quantify motion vector entropy, frame differencing is employed, which is a method used to estimate the level of motion activity between two successive frames.

The entropy measure reflects how motion intensities are distributed and it is defined as

$$MVE(t) = -\sum_i p_i \log(p_i) \quad (2)$$

where (pi) is the normalized motion intensity distribution. High entropy values are usually related to complex motion patterns during transitions.

3.2.3. COLOR HISTOGRAM CHI-SQUARE DISTANCE

Color distribution changes are quantified by measuring the chi-square distance between grayscale histograms of subsequent frames:

$$CHI(t) = \frac{1}{2}\sum_i \frac{(h_t(i)-h_{t-1}(i))^2}{(h_t(i)+h_{t-1}(i))} \quad (3)$$

where ht(i) represents the histogram bin values. This feature is especially good at identifying fade and dissolve transitions.

3.3. ADAPTIVE TEMPORAL MODELING USING BI-LSTM

The extracted multi-cue features are temporally correlated and may have long-range dependencies, particularly when there are gradual transitions. To model the temporal dynamics, a Bi-LSTM network is utilized. Bi-LSTM takes the feature sequence in both forward and backward directions, allowing the network to use both past and future contextual information.

The fused feature sequence is initially formed as

$$F(t) = EED(t) + MV\ E(t) + CHI(t) \quad (4)$$

which is then normalized and given to the Bi-LSTM network. The network produces frame-level transition probabilities $P(t) \in [0,1]$, which are the probabilities of a shot boundary at each frame. To make the process computationally efficient, the Bi-LSTM model is only trained once using weak supervision coming from feature statistics, and the trained network is used again during the optimization phase.

3.4. GREY WOLF OPTIMIZATION FOR FEATURE FUSION AND THRESHOLD SELECTION

Temporal modeling is capable of providing probabilistic transition cues, yet this alone is not sufficient for optimal decision making which also requires the adaptation of feature fusion and threshold selection. GWO is used for finding the best combination of feature fusion weights and decision threshold.

3.4.1 Feature Fusion Model

The final fused feature representation is defined as

$$Fw(t) = w1EED(t) + w2MV\ E(t) + w3CHI(t), \quad (5)$$

sub to $w1 + w2 + w3 = 1$

The usage of weights {w1,w2,w3} is autonomously adjusted by GWO to heat the most discriminative features of various video content.

3.5. ADAPTIVE THRESHOLD OPTIMIZATION

It is deemed that the frame t is a shot boundary if

$$P(t) > \theta \quad (6)$$

where θ stands for an adaptive decision threshold, which is optimized by GWO. This means that there is no need for manual threshold definition and the method generalizes better to different types of video sequences.

The individual candidate solution vectors, are represented by grey wolves:

$$X = \{w1,w2,w3,\theta\} \quad (7)$$

Each solution's fitness is assessed by the F1-score that equally weights precision and recall, thus resulting in a reliable detection performance.

3.6. FINAL SHOT BOUNDARY DETECTION

They get final shot boundary decisions by applying the optimized feature weights and threshold to the Bi-LSTM output probabilities. They apply a simple post-processing step to remove isolated false detections by enforcing a minimum temporal duration constraint between consecutive shot boundaries.

3.7. COMPUTATIONAL EFFICIENCY CONSIDERATIONS

The GWO-ATDL-SBD framework proposed here significantly reduces computational complexity, unlike other hybrid frameworks that integrate deep network training into evolutionary optimization loops. The Bi-LSTM network is trained only once, and GWO is applied to a low-dimensional parameter space only, which leads to faster convergence and scalability to long video sequences.

4. EXPERIMENTAL RESULTS AND DISCUSSION

The proposed model is evaluated based on benchmark video datasets. The test results are then compared with those obtained by existing state-of-the-art shot boundary detection methods in terms of detection accuracy, robustness, and computational efficiency.

4.1. DATASETS

The strength of the method presented is confirmed through a thorough database which is a major performance benchmark. This dataset combines carefully selected video sequences from the MY Dataset with the TRECVid 2001, 2007, 2008, and 2009 collections.

Table 1

Table 1 Statistical Analysis of Datasets TRECVid 2001 and 2007

Video	Frame #	Transition			Sources
		Abrupt	Gradual	Total	
D2	16586	42	31	73	
D3	12304	39	64	103	**TRECVid 2001**
D4	31389	98	55	153	
D5	12510	45	26	71	
D6	13648	40	45	85	
BG 3027	49815	126	1	127	
BG 16336	2462	20	-	20	**TRECVid 2007**
BG 36136	29426	88	12	100	
BG 37309	9639	11	8	19	
BG 37770	15836	8	27	35	
BG 8907	7551	87	6	93	
BG 10523	9378	98	7	105	**TRECVid 2008**
BG 26797	3609	27	8	35	

BG 34413	11502	124	3	127	
BG 36580	13635	33	12	45	
BG 8910	5913	43	-	43	
BG 14233	4518	37	14	51 37	**TRECVid 2009**
BG 22678	10800	36	1		
BG 24269	13536	88	-	88	
BG 37235	9738	46	1	47	
Clip 1	2500	37	-	37	
Clip 2	3720	26	5	31 28	**MY Dataset**
Clip 3	1361	16	13		
Clip 4	3732	77	4	81	
Clip 5	4883	65	-	35	

In an effort to carry out a thorough assessment, the dataset was expanded with various film sequences that featured dramatic changes in lighting and very fast motion. In particular, we added movie parts from Avatar, Transformers: Dark of the Moon, Masoom, The Terminator, and Tron: Legacy. The entire set of experiments was run on an ASUS ROG Strix G513IC system. Table 1 lays out in detail the video setting and features.

4.2. EVALUATION METRICS

The performance of the proposed GWO-ATDL-SBD framework has been measured by using the standard metrics in shot boundary detection, such as Precision (P), Recall (R), and F1-Score (F1), which are mathematically defined as

$$P = \frac{TP}{TP + FP} \tag{8}$$

$$R = \frac{TP}{TP + FN} \tag{9}$$

$$F1 = \frac{2 * R * P}{R + P} \tag{10}$$

where TP, FP and FN are the numbers of true positives, false positives, and false negatives, respectively. To match the predicted boundaries with the ground truth transitions, a detection tolerance window of ±5 frames was used.

4.3. QUANTITATIVE PERFORMANCE ANALYSIS

Table 2 briefly shows the quantitative results of the proposed GWO-ATDL-SBD framework and the existing methods.

Table 2

Table 2 Evaluating the System's Response to Threshold Adaptivity

Videos	CNN			LSTM			Proposed		
	Rec	Pre	F1	Rec2	Pre	F1	Rec	Pre	F1
D2	85.00	87.60	86.48	70.64	72.13	71.38	97.67	91.15	94.28
D3	86.00	88.15	87.26	93.00	95.00	94.00	98.60	92.53	95.40
D4	86.72	88.45	87.57	92.56	94.40	93.46	97.69	96.92	97.30
D5	87.88	89.00	87.93	93.40	94.10	93.74	100.0	95.67	97.74
D6	88.00	89.05	88.61	93.60	95.02	94.31	99.76	91.15	95.25
Average	**86.72**	**88.45**	**87.57**	**88.64**	**90.13**	**89.37**	**97.76**	**93.34**	**95.45**

The GWO-ATDL-SBD framework proposed in this paper produces better results than all other techniques causing the F1-score to reach a peak of 95.45%. The performance gain from the baseline CNN and LSTM model point to the success of adaptive temporal modeling and optimized feature fusion. Most importantly, the Bi-LSTM module adds a lot to the detection of gradual transitions that are usually neglected by frame-wise classifiers.

4.4. ABLATION STUDY

An ablation study was performed to evaluate the contribution of each component by gradually excluding modules from the proposed framework.

Table 3

Table 3 Evaluating the System's Response to Different Configurations

Videos	Multi-cue features only			Bi-LSTM without GWO			GWO without Bi-LSTM			Proposed		
	R	P	F1	R	P	F1	R	P	F1	R	P	F1
D2	85.50	82.52	83.98	88.40	87.12	87.76	86.72	85.64	86.18	97.67	91.15	94.28
D3	86.94	83.20	85.03	89.75	89.41	89.58	87.91	84.72	86.28	98.6	92.53	95.40
D4	86.20	83.48	84.81	90.12	88.64	89.37	88.36	86.12	87.23	97.69	96.92	97.30
D5	88.56	83.00	85.69	91.90	89.72	90.80	89.57	86.93	88.23	**100.00**	95.67	97.74
D6	87.50	83.85	85.64	91.18	88.26	89.70	89.24	87.19	88.20	99.76	**100.00**	99.87
Avg	**86.94**	**83.21**	**85.03**	**90.27**	**88.63**	**89.44**	**88.36**	**86.12**	**87.22**	**97.76**	**95.25**	**99.87**

The table 3 clearly validates the architectural decisions of the proposed GWOATDL-SBD framework. Multi-cue features provide essential low-level information, Bi-LSTM enables robust temporal modeling, and Grey Wolf Optimization ensures adaptive and content-aware parameter tuning. The absence of any one component leads to measurable performance degradation, confirming that the superior performance of the proposed method arises from the complementary interaction of all three modules.

4.5. SYSTEM PERFORMANCE

Selected videos from the datasets TRECVid 2001, 2007, 2008, 2009, and My Dataset had their performance are analyzed in Table 4.

4.6. DISCUSSION

The algorithm we proposed has accurately identified the majority of the abrupt transitions in the video. The suggested algorithm has excluded frames with changes in illumination and object motion.

Figure 2 depicts the problem of flashlight effects in videos, which can lead to frames being misclassified as abrupt transitions. Our proposed approach successfully resolves this issue in shot boundary detection.

Table 4

Table 4 System Performance Using Different Datasets

Videos	Parameter measure			Computation Time	Source
	Recall	Precession	F1 Score		
D2	97.67	91.15	94.28	638	
D3	98.60	92.53	95.00	503	**TRECVid 2001**
D4	97.69	96.92	97.30	1397	
D5	100.00	95.67	97.74	489	
D6	99.76	100.00	99.87	533	
BG 3027	99.21	98.43	98.82	1547	
BG 3097	98.90	100.00	99.45	1398	**TRECVid 2007**

BG 16336	100.00	100.00	100.00	80	
BG 28476	98.86	97.75	98.3	717	
BG 36136	91.67	92.3	91.98	1815	
BG 37309	100.00	100.00	100.00	409	
BG 37770	97.14	94.59	95.85	856	
BG 8907	100.00	96.67	98.31	252	
BG 10523	98.97	92.38	95.56	328	**TRECVid 2008**
BG 26797	96.29	92.85	94.54	115	
BG 34413	97.58	94.53	96.03	543	
BG 36580	96.97	94.28	95.60	691	
BG 8910	97.67	95.45	96.55	419	
BG 10523	100.00	92.50	96.10	157	**TRECVid 2009**
BG 22678	100.00	94.73	97.29	443	
BG 24269	95.45	93.33	94.38	711	
BG 37235	97.82	93.75	95.65	562	
Clip 1	97.30	94.73	96.00	133	
Clip 2	100.00	100	100.00	183	**MY Dataset**
Clip 3	100.00	93.75	96.77	81	
Clip 4	98.70	96.33	97.50	167	
Clip 5	98.46	96.97	97.71	215	
Average	**98.32**	**95.28**	**96.73**	**569**	

Figure 2

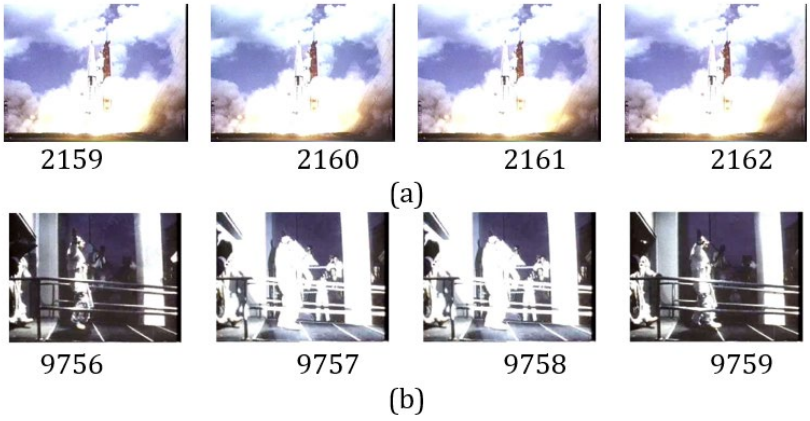

Figure 2 Demonstrating an Instance Of (A) Uniform Illumination and (B) Non-Uniform Illumination**.**

Videos D2 and D4 were selected for experimentation due to their sudden illumination and motion effects. As shown in Table 3 the proposed method effectively addresses these challenges. A closer examination reveals that the F1 scores for all videos in Table 3 improve as the precision of the proposed system increases. Figure 2 highlights significant challenges like uniform and non-uniform illumination changes from videos D2 and D4, respectively, which are effectively managed by our system. Dubey and Dubey (2026)

Figure 3

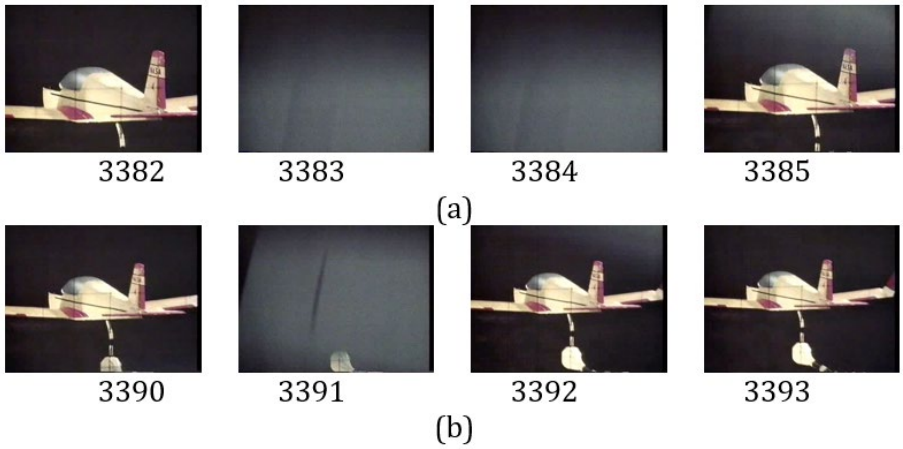

Figure 3 Instances of Obstruction in the Camera's Perspective: (A) Across Multiple Frames, (B) Within a Single Frame

Figure 3 provides an example from video D4, where a large object (a fan) blocks the view of the object (a dummy airplane) in front of the camera. Figures 4(a) and 4(b) illustrate the obstruction across multiple consecutive frames and in a single frame, respectively. These issues are efficiently addressed during the confirmation stage of the proposed system.

Figure 4

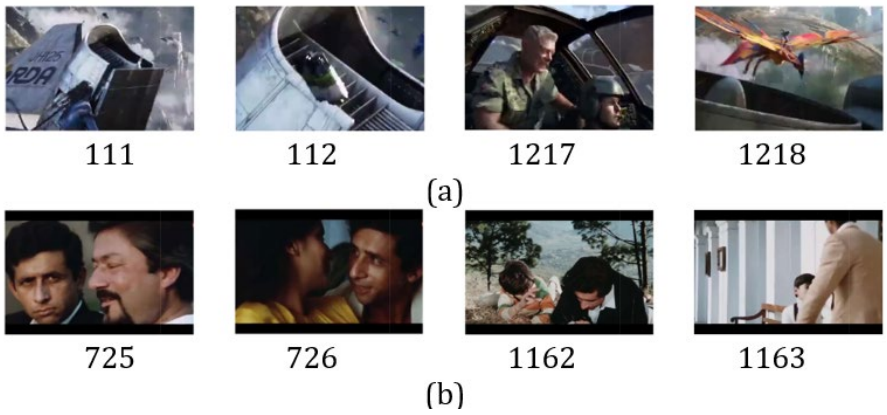

Figure 4 An Example of Correctly Detected Abrupt Transitions

As proposed below, our method discovered all the concealed cuts between the scenes or frames in Clips 1 and Clips 3, as shown in Figure 4. That is, the algorithm properly ensures accurate segmentation of a video by detecting sharp changes or boundaries inside a video. According to both clips, the method demonstrates reliability and strength in handling various illumination changes, making it suitable for tasks that require precise boundary identification.

4.7. COMPARISON

To establish a comprehensive performance baseline, Table 5 features a comparison against a wide spectrum of contemporary techniques, including methods such as Bifold stage SBD [7], Kernel-dependent SBD [26], the Automated Framework for illumination and motion [27], Stationary wavelet [28], Fast Preprocessing Frame-work [29], Hybrid optimisation technique(s) (SBD) [30], Spatio temporal convolutional neural network-SBD [31], Automatic SBD illumination and motion effect [32], Low-Rank and Updating [33], and the Multi-layer perception [34].

Table 5

Table 5 Comparative Anatomy of Latest Methodologies

Algorithm	Parameters	Videos				Average
		D2	D3	D4	D6	
Proposed	Rec	97.6	98.6	97.6	99.7	**98.4**
	Pre	91.1	92.5	96.9	**100.0**	**95.1**
	F1	94.2	95.4	97.3	99.8	**96.7**
Bifold stage [7]	Rec	92.9	87.2	86.8	95.0	90.4
	Pre	**100.0**	**100.0**	97.8	**100.0**	99.4
	F1	96.3	93.20	92.0	97.4	94.7
Kernel dependent sbd [26]	Rec	97.0	82.0	88.0	95.0	90.5
	Pre	85.0	86.0	90.0	97.0	90.0
	F1	91.0	84.0	89.0	96.0	90.0
Automated framework for illumination and motion [27]	Rec	80.0	82.0	78.0	92.0	83.0
	Pre	94.0	**100.0**	96.0	84.0	94.0
	F1	87.0	90.0	86.0	88.0	88.0
Stationary wavelet [28]	Rec	97.0	97.0	93.0	**100.0**	97.0
	Pre	6.0	8.0	7.0	8.0	7.0
	F1	12.0	16.0	13.0	16.0	14.0
Fast Preprocessing frame-work [29]	Rec	57.0	46.0	75.0	89.0	67.0
	Pre	**100.0**	**100.0**	98.0	**100.0**	99.6
	F1	72.0	63.0	85.0	94.0	79.0
Hybrid optimization technique (sbd) [30]	Rec	97.0	92.0	**100.0**	**100.0**	97.0
	Pre	82.0	**100.0**	89.0	**100.0**	92.0
	F1	89.0	96.0	94.0	**100.0**	94.0
Spatio-temporal convolutional neural networks [31]	Rec	89.0	92.0	85.0	87.0	88.0
	Pre	87.0	**100.0**	**100.0**	**100.0**	96.0
	F1	88.0	96.0	92.0	93.0	93.0
Automatic SBD illumination and motion effect [32]	Rec	90.0	89.0	89.0	92.0	90.0
	Pre	**100.0**	**100.0**	94.0	97.0	98.0
	F1	95.0	94.0	92.0	94.0	94.0
Low rank and Updating [33]	Rec	92.8	92.3	91.8	**100.0**	94.2
	Pre	90.6	**100.0**	93.7	**100.0**	96.0
	F1	91.6	95.9	93.2	**100.0**	95.2
Multi-layer perceptron [34]	Rec	88.1	97.4	87.8	97.5	92.7
	Pre	94.9	88.4	91.5	92.9	92.0
	F1	91.4	92.7	89.6	95.1	92.2

The tabulated data decisively indicates that the proposed system consistently exceeds the performance of all existing methods, registering superior F1-scores and affirming its exceptional accuracy and robustness.

5. CONCLUSION AND FUTURE WORK

This paper proposes a new hybrid framework, GWO-ATDL-SBD, which combines adaptive temporal deep learning with Grey Wolf Optimization to achieve the two objectives of robustness and computational efficiency for shot boundary detection. Our approach takes advantage of three complementary multi-cue visual descript or sedge energy difference, motion vector entropy, and color histogram divergence. These features are then passed to a Bi-directional Long Short-Term Memory network which helps in capturing the temporal relationships between the video frames. In order to

remove the need for manual parameter tuning and improve the generalization performance on diverse video content, the Grey Wolf Optimizer was used for the adaptive optimization of feature fusion weights and decision thresholds.

Experiments carried out on standard video datasets clearly show that the framework we propose outperforms the traditional feature-based, machine learning, based, and hybrid methods for shot boundary detection. The findings revealed that the model was able to improve precision, recall, and F1-score metrics even in the most difficult scenarios of gradual transitions and dynamic scene variations. Ablation studies revealed that multi-cue feature representation, adaptive temporal modelling, and evolutionary optimization, when combined, generate synergistic effects that lead to a well-balanced enhancement in detection accuracy and robustness. Furthermore, by separating deep learning training from the optimization loop, our framework substantially reduces computational complexity as compared to the traditional hybrid evolutionary deep learning methods.

Even though it is a well-performing venture, future research has many avenues. To begin with, the existing architecture is based on the manual development of low-level features that might be constrained by videos with a complicated semantic transition. The future directions will focus on incorporating the deep visual feature extractors e.g. convolutional neural networks to allow end-to-end learning of features. Second, the Bi-LSTM network can be further supervised, and then accurately supervised with the help of an annotated video dataset to train the weakly supervised training strategy of the network. Third, the optimization strategy suggested can be generalized to multi-objective evolutionary models to co-optimize the detection accuracy and the computational cost. Lastly, the real-time deployment and hardware-aware model compression methods will also be explored so that real-time applications can be implemented in large-scale video surveillance and multimedia retrieval systems.

AUTHOR CONTRIBUTION

All the authors are equally contributed in this research.

DATA AVAILABILITY STATEMENT

The video, Sound and Vision, is copyrighted. The TRECVideo Information Retrieval Evaluation Project Collection provides the video's sound and visual elements, which are employed exclusively for research objectives in this project.

RESEARCH INVOLVING HUMAN AND /OR ANIMALS

Since our research is solely based on Video data, so no human and animal is involved in this research.

INFORMED CONSENT

Informed Consent are not necessary since the Video data has been downloaded from TrecVid on request.

COPYRIGHT

Manuscripts submitted to the journal have not been published, accepted for publication, nor simultaneously submitted for publication elsewhere. The author(s) agree that copyright for the article is transferred to the publisher, if and when the manuscript is accepted for publication.

CONFLICT OF INTERESTS

None.

ACKNOWLEDGMENTS

The Sound and Vision video is protected by copyright and is utilized here strictly for scholarly inquiry. Access to the media was facilitated through the TREC Video Information Retrieval Assessment Project collection.

REFERENCES

Anthwal, S., Ganotra, D.: An overview of optical flow-based approaches for motion segmentation. The Imaging Science Journal 67(5), 284–294 (2019)

Benoughidene, A., Titouna, F.: A novel method for video shot boundary detection using cnn-lstm approach. International Journal of Multimedia Information Retrieval 11(4), 653–667 (2022)

Chakraborty, D., Chiracharit, W., Chamnongthai, K.: Video shot boundary detection using principal component analysis (pca) and deep learning. In: 2021, 18th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), pp. 272–275 (2021). IEEE.

Chakraborty, S., Singh, A., Thounaojam, D.M.: A novel bifold-stage shot bound- ary detection algorithm: invariant to motion and illumination. The Visual Computer 38(2), 445–456 (2022)

Chakraborty, S., Thounaojam, D.M.: A novel shot boundary detection system using hybrid optimization technique. Applied Intelligence 49(9), 3207–3220(2019)

Chan, C., Wong, A.: Shot boundary detection using genetic algorithm optimization. In: 2011 IEEE International Symposium on Multimedia, pp. 327–332 (2011). IEEE

Dubey, A. K., & Dubey, A. (2026). Digitalization in Teaching and Learning: Impact on Student Engagement and Academic Achievement. ShodhAI: Journal of Artificial Intelligence, 3(1), 37–42. https://doi.org/10.29121/shodhai.v3.i1.2026.73

Gawande, U., Hajari, K., Golhar, Y., Fulzele, P.: A novel gray wolf optimization based key frame extraction method for video classification using convlstm. Neural Computing and Applications 36(32), 20355–20385 (2024)

GG, L.P., Domnic, S.: Walsh–hadamard transform kernel-based feature vector for shot boundary detection. IEEE Transactions on Image Processing 23(12), 5187–5197 (2014)

Goswami, B., Boers, N., Rheinwalt, A., Marwan, N., Heitzig, J., Breitenbach, S.F., Kurths, J.: Abrupt transitions in time series with uncertainties. Nature communications 9(1), 48 (2018).

Guo, H., Liu, J., Xiao, Z., Xiao, L.: Deep cnn-based hyperspectral image classification using discriminative multiple spatial-spectral feature fusion. Remote Sensing Letters 11(9), 827–836 (2020)

Hassanien, A., Elgharib, M., Selim, A., Bae, S.-H., Hefeeda, M., Matusik, W.: Large-scale, fast and accurate shot boundary detection through spatio-temporal convolutional neural networks. arXiv preprint arXiv:1705.03281 (2017)

Idan, Z.N., Abdulhussain, S.H., Mahmmod, B.M., Al-Utaibi, K.A., Al-Hadad, S.A.R., Sait, S.M.: Fast shot boundary detection based on separable moments and support vector machine. IEEe Access 9, 106412–106427 (2021)

Kar, T., Kanungo, P., Mohanty, S.N., Groppe, S., Groppe, J.: Video shot-boundary detection: issues, challenges and solutions. Artificial Intelligence Review 57(4), 104 (2024)

Kar, T., Kanungo, P.: A motion and illumination resilient framework for automatic shot boundary detection. Signal, Image and Video Processing 11,1237–1244 (2017)

Li, H., Wei, M.: Fuzzy clustering based on feature weights for multivariate time series. Knowledge-Based Systems 197, 105907 (2020)

Li, Y., Li, C., Li, X., Wang, K., Rahaman, M.M., Sun, C., Chen, H., Wu, X., Zhang, H., Wang, Q.: A comprehensive review of markov random field and conditional random field approaches in pathology image analysis. Archives of Computational Methods in Engineering 29(1), 609–639 (2022)

Li, Y.-N., Lu, Z.-M., Niu, X.-M.: Fast video shot boundary detection framework employing pre-processing techniques. IET image processing 3(3), 121–134 (2009)

Liu, S., Liu, G., Zhou, H.: A robust parallel object tracking method for illumination variations. Mobile Networks and Applications 24(1), 5–17 (2019)

Mei, S., Ji, J., Hou, J., Li, X., Du, Q.: Learning sensor-specific spatial-spectral features of hyperspectral images via convolutional neural networks. IEEE Transactions on Geoscience and Remote Sensing 55(8), 4520–4533 (2017)

Mirjalili, S., Mirjalili, S.M., Lewis, A.: Grey wolf optimizer. Advances in engineering software 69, 46–61 (2014)

Mondal, J., Kundu, M.K., Das, S., Chowdhury, M.: Video shot boundary detec- tion using multiscale geometric analysis of nsct and least squares support vector machine. Multimedia Tools and Applications 77(7), 8139–8161 (2018)

Navin, K., Krishnan, M., et al.: Fuzzy rule based classifier model for evidence based clinical decision support systems. Intelligent systems with applications 22,200393 (2024)

Park, S., Sch¨ops, T., Pollefeys, M.: Illumination change robustness in direct visual slam. In: 2017 IEEE International Conference on Robotics and Automation (ICRA), pp. 4523–4530 (2017). IEEE

Peralta, B., Tirapegui, V., Pieringer, C., Caro, L.: A simple proposal for sentiment analysis on movies reviews with hidden markov models. In: Iberoamerican Congress on Pattern Recognition, pp. 152–162 (2019). Springer

Prasanna, C.S., Rahman, M.Z.U., Bayleyegn, M.D.: Brain epileptic seizure detection using joint cnn and exhaustive feature selection with rnn-blstm classifier. IEEE Access 11, 97990–98004 (2023)

Ray, P.P.: A review on tinyml: State-of-the-art and prospects. Journal of King Saud University-Computer and Information Sciences 34(4), 1595–1623 (2022)

Sharma, D.M., Shandilya, S.K.: An efficient cyber-physical system using hybridized enhanced support-vector machine with ada-boost classification algorithm. Concurrency and Computation: Practice and Experience 34(21), 7134(2022)

Shinde, T., Shrivastava, M.: Development of an ant colony optimisation-based edge detection framework. African Journal of Applied Research 11(5), 573–595 (2025)

Singh, A., Thounaojam, D.M., Chakraborty, S.: A novel automatic shot boundary detection algorithm: robust to illumination and motion effect. Signal, Image and Video Processing 14(4), 645–653 (2020)

Thounaojam, D., Khelchandra, T., Jayshree, T., Roy, S., Singh, K.: Colour histogram and modified multi-layer perceptron neural network based video shot boundary detection. International Arab Journal of Information Technology 16(4),686–693 (2019)

Warhade, K.K., Merchant, S., Desai, U.B.: Shot boundary detection in the presence of fire flicker and explosion using stationary wavelet transform. Signal, Image and Video Processing 5, 507–515 (2011)

Wu, L., Zhang, S., Jian, M., Lu, Z., Wang, D.: Two stage shot boundary detection via feature fusion and spatial-temporal convolutional neural networks. IEEE Access 7, 77268–77276 (2019)

Youssef, B., Fedwa, E., Driss, A., Ahmed, S.: Shot boundary detection via adaptive low rank and svd-updating. Computer Vision and Image Understanding161,20–28 (2017)

Zhao, L., Sun, X.M., Zhang, M.W.: A shot boundary detection method based on pso-svm. Applied Mechanics and Materials 130, 3821–3825 (2012)