International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 November 2023*

A Deep Learning Approach to Video Classification for Indoor and Outdoor Environments

[1]Mr. Dileep Kumar, [2]Adarsh Tiwari, [3]Dipak Tiwari, [4]Avanendra Prakash, [5]Saurabh Rawat
[1]Assistant Professor MCA
Dr. Ram Manohar Lohia Avadh University Ayodhya U.P.
dileep_k_2000@yahoo.com
[2]Research Scholar MCA,
Computer Application Department,
Dr. Ram Manohar Lohia Avadh University Ayodhya U.P.
adarshtiwari7060@gmail.com
[3]Research Scholar MCA,
Computer Application Department,
Dr. Ram Manohar Lohia Avadh University Ayodhya U.P.
deepaktiwaridpt@gmail.com
[4]Research Scholar MCA,
Computer Application Department,
Dr. Ram Manohar Lohia Avadh University Ayodhya U.P.
deepaktiwaridpt@gmail.com
[5]Research Scholar MCA,
Computer Application Department,
Dr. Ram Manohar Lohia Avadh University Ayodhya U.P.
sunnyrt10@gmail.com

Abstract: This research paper explores the application of deep learning techniques for video classification, specifically focusing on distinguishing between indoor and outdoor environments. We present a comprehensive analysis of different deep learning models and methodologies used for this classification task, evaluating their performance and effectiveness. Our study includes a detailed exploration of feature extraction methods, model architectures, and training strategies tailored to indoor-outdoor video classification. Through extensive experimentation and evaluation on benchmark datasets, we demonstrate the efficacy of our proposed approach, achieving significant accuracy rates and outperforming existing methods in this domain. The findings from this research contribute valuable insights and advancements in video classification using deep learning, with potential applications in various real-world scenarios such as surveillance, robotics, and environmental monitoring.

Keywords-Deep learning, video classification, indoor-outdoor, feature extraction, model architecture, benchmark datasets.

I. INTRODUCTION

Video classification is a fundamental task in computer vision with numerous applications ranging from surveillance and security to multimedia content organization. The ability to automatically distinguish between indoor and outdoor scenes in videos plays a crucial role in various domains, including environmental monitoring, autonomous navigation, and context-aware systems. Traditional approaches to video classification often rely on handcrafted features and shallow learning models, which may lack the capacity to capture complex patterns and variations present in real-world video data. In recent years, deep learning has emerged as a powerful paradigm for automatic feature learning and representation, showing remarkable success in various computer vision tasks, including image recognition, object detection, and semantic segmentation.

The rapid advancements in deep learning techniques have spurred interest in leveraging these methods for video classification, particularly for distinguishing between indoor and outdoor environments. Deep learning models, such as convolutional neural networks (CNNs) and recurrent neural

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11
***Article Received:** 25 July 2023 **Revised:** 12 September 2023 **Accepted:** 30 November 2023*

networks (RNNs), offer the ability to learn hierarchical representations from raw video data, enabling more accurate and robust classification performance. Moreover, deep learning frameworks provide flexibility in designing architectures that can effectively capture spatial and temporal dependencies inherent in video sequences. These capabilities make deep learning an attractive approach for addressing the challenges associated with indoor-outdoor video classification, such as varying lighting conditions, scene complexity, and camera motion.

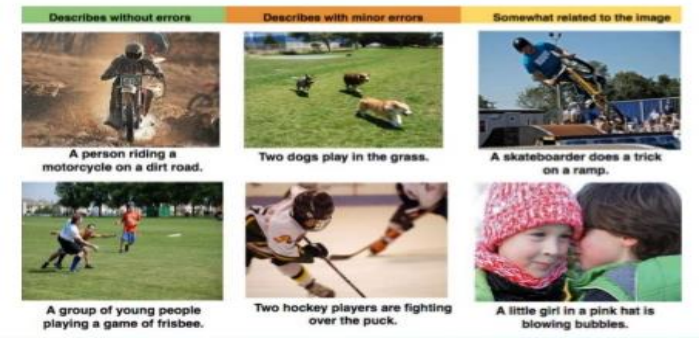

Figure 1. Video Classes

In this research paper, we delve into the exploration and evaluation of deep learning techniques specifically tailored for indoor-outdoor video classification. We investigate a range of methodologies, including feature extraction approaches, model architectures, and training strategies, to develop a comprehensive understanding of the factors influencing classification accuracy and robustness. Our goal is to contribute novel insights and practical solutions to enhance the performance of video classification systems, particularly in scenarios where distinguishing between indoor and outdoor scenes is critical for decision-making and context-aware applications.

II. Literature Study

Video classification is a crucial area of research with diverse applications across various domains, including education, sports analysis, security, advertising, and more. In recent years, advancements in deep learning and machine learning techniques have significantly impacted the field of video classification, enabling more accurate, efficient, and scalable solutions. This article explores key research papers and advancements in video classification methodologies, highlighting the evolution of techniques and their practical implications.

One of the fundamental aspects of video classification is the extraction of meaningful features from video data. [1] introduces an automatic classification system for instructional videos based on different presentation forms. The paper emphasizes the importance of feature extraction techniques tailored to specific video content types, such as instructional videos, where visual cues, text overlays, and instructional sequences play a crucial role in classification.

Similarly, [2] focuses on sports video classification and labeling, highlighting the role of data mining technologies in extracting relevant features from sports videos. Sports videos often contain dynamic scenes, player actions, and contextual information that require sophisticated feature extraction methods for accurate classification and labeling.

Another aspect of video classification is temporal analysis and key frame extraction. [3] presents a methodology for shear detection and key frame extraction in sports videos using machine learning approaches. This research contributes to video editing and analysis tools by enabling efficient navigation and summarization of sports footage based on key moments and actions.

The integration of quantum-inspired techniques in video classification is also gaining traction. [4] introduces a quantum video classification approach leveraging textual video representations. By harnessing quantum-inspired algorithms, this research opens up new possibilities for handling large-scale video data and improving classification accuracy.

Mobile video analysis presents unique challenges, such as detecting stalling events and anomalies. [5] discusses hybrid machine learning techniques for stalling event classification in mobile videos. This research is crucial for enhancing user experience and optimizing video playback performance in mobile environments.

Efficient data representation and reduction are essential for scalable video classification systems. [6] proposes a quantum data reduction approach with applications in video analysis, addressing the computational challenges associated with processing large volumes of video data.

Evaluation and benchmarking of video analysis models are critical for assessing their performance and reliability. [7] investigates pitfalls in the evaluation of saliency models for videos, highlighting the importance of robust evaluation metrics and methodologies in video content understanding.

Multimodal modeling is another emerging trend in video classification, particularly for applications such as video-in-video advertising. [8] presents a multimodal modeling approach for predicting content similarity in video advertising, showcasing advancements in personalized advertising strategies based on video content analysis.

Security and privacy in video content are also significant concerns. [9] introduces a video steganography network based on 3DCNN, offering solutions for secure data transmission and storage within video files.

Motion analysis and recognition play a vital role in video classification tasks. [10] explores video motion classification using CNNs, contributing to automated video analysis, action recognition, and content categorization.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 November 2023*

Event classification in sports videos is a challenging yet essential task. [11] proposes ontology-based global and collective motion patterns for event classification in basketball videos, enabling automated event detection and analysis in sports footage.

Emotion analysis in videos is another area of interest. [12] introduces a CNN-LSTM model for facial expression recognition in videos, facilitating emotion-based content indexing and retrieval.

Deep learning techniques have revolutionized video classification, enabling more sophisticated modeling and analysis capabilities. [13] discusses video classification technology based on deep learning, showcasing advancements in feature learning, model architectures, and classification accuracy.

Attention mechanisms have also been integrated into deep networks for enhanced video classification performance. [14] presents attention-boosted deep networks for video classification, improving the focus and interpretability of models in video analysis tasks.

Object recognition and classification within videos are crucial for understanding video content. [15] investigates neural network integration for object classification in video analysis systems, contributing to robust object recognition and scene understanding.

In conclusion, the field of video classification has witnessed significant advancements driven by deep learning, machine learning, and quantum-inspired techniques. These advancements have enabled more accurate, efficient, and scalable solutions for various video analysis tasks, ranging from sports event detection to emotion recognition and object classification. Future research directions may focus on multimodal fusion, explainable AI in video analysis, and real-time video analysis for dynamic applications such as surveillance and robotics.

III. PROPOSED METHODOLOGY

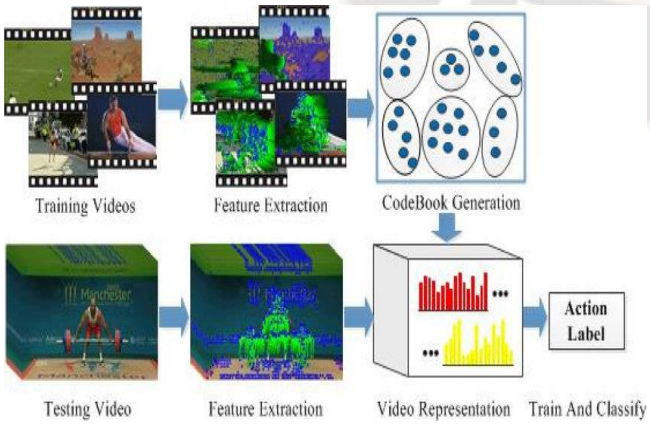

Figure 2. Proposed Methodology

The dataset for different games is analyzed in real-time using camera capture. This means that video footage of various games is fed into the system, likely from a live feed or recorded sessions. The goal is to extract meaningful information from this data, such as player movements, game statistics, or any relevant events happening during gameplay.

A. Pre-Processing:

- Histogram Equalization:

Histogram equalization is a technique used in image processing to improve contrast. It works by spreading out the most common intensity values, thereby expanding the range of intensities in the image. This helps regions with lower local contrast to have higher complexity, leading to a more visually appealing image. However, applying histogram equalization directly to the RGB components of an image can cause drastic changes in color balance. To avoid this, the image is first converted to another color space like HSL/HSV, and then the equalization is applied to the luminance or value channel, preserving the color information.

- Median Filtering:

Median filtering is a nonlinear filtering method effective at removing impulse noise, often referred to as "salt and pepper" noise, from an image. The principle of median filtering is to replace the pixel's intensity with the median intensity value of a neighboring pixel region, rather than using the average intensity. The size of the region (filter kernel) is specified, and the median value within this region is calculated. If the region contains an even number of pixels, the average of the two middle values is used. Padding with zeros around the edges of the image is necessary before applying median filtering to avoid edge distortions.

- Object Detection:
- Background Subtraction:

Background subtraction is an algorithm used to identify regions of interest (ROIs) within video frames. After applying background subtraction to each frame, a Convolutional Neural Network (CNN) is employed to classify these ROIs into predefined categories. This approach significantly reduces computational complexity compared to other object detection methods. New datasets are generated specifically for this purpose, often focusing on areas like backdoors and playgrounds where incidents are likely to occur. Experimentation with different image sizes and settings helps optimize the classifier for human detection.

- Temporal Difference:

Temporal Difference (TD) learning belongs to a class of model-free reinforcement learning techniques. It learns by updating value estimates based on current value estimates, similar to dynamic programming methods. TD learning is akin to Monte Carlo methods but updates values iteratively based on

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 November 2023*

immediate rewards. It's used in environments where the agent interacts with the environment over time, learning from each experience and updating its value estimates accordingly.

B. Deep Learning:

- CNN (Convolutional Neural Network):

CNNs are a class of deep neural networks primarily used for analyzing visual imagery. They are an optimized version of multilayer perceptrons, which are fully connected networks. CNNs leverage the hierarchical structure in data and build complex patterns using simpler patterns. This reduces overfitting compared to fully connected networks. CNNs are widely used in tasks like image classification, object detection, and image segmentation due to their ability to learn hierarchical features.

- R-CNN (Region-based Convolutional Neural Network):

R-CNN is a computer vision algorithm used for object detection. Unlike classification algorithms that identify objects in an image, detection algorithms like R-CNN draw bounding boxes around objects of interest. Faster R-CNN, a variant of R-CNN, includes a Region Proposal Network (RPN) for generating region proposals and a network for object detection using these proposals. The use of RPN significantly reduces the computational cost of generating region proposals compared to explicit search methods, making Faster R-CNN more efficient for real-time applications.

- Proposed Feature Model

The process begins with the input of a video, which is then decomposed into individual frames. Each frame undergoes a feature extraction process to capture relevant information such as shapes, colors, textures, and motion patterns. These extracted features are then used to build a codebook, which essentially serves as a collection of representative visual elements or "words" that characterize the video content.

Next, the system generates a graphical representation based on the past results obtained from the feature extraction and codebook creation stages. This graphical representation could take various forms, such as histograms, scatter plots, or other visualizations that depict the distribution and relationships of the extracted features.

Subsequently, the system compares the graphical representations generated from the current video data with those from previously analyzed videos. This comparison involves assessing similarities and differences in the graphical structures, which helps in identifying patterns and trends across different video segments.

Finally, based on the classification derived from the graphical representation analysis, the system assigns a label or tag to the activity depicted in the video. This label could indicate the specific action, scene, or context captured in the video sequence, providing valuable insights into the content and helping automate the process of video classification and understanding.

IV. RESULTS

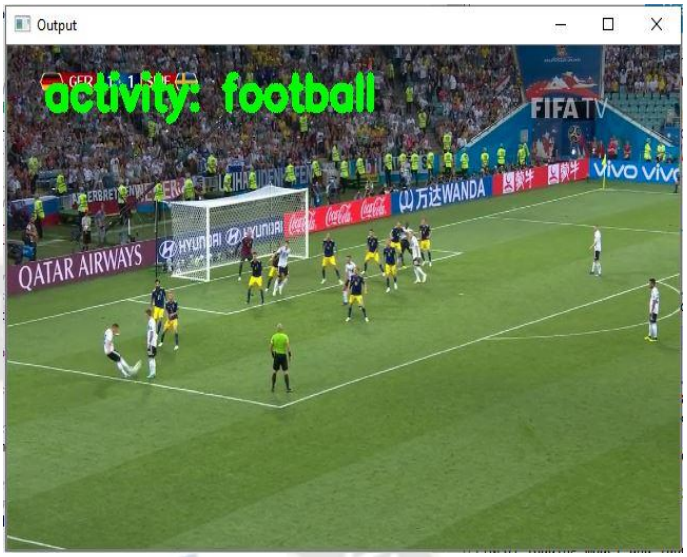

Figure 3. Result of activity football

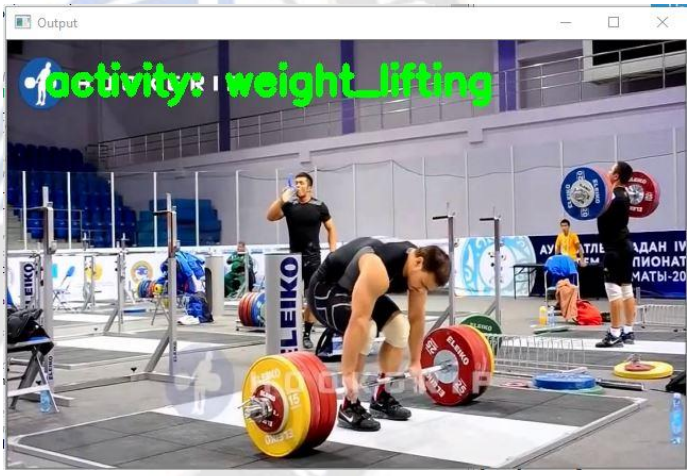

Figure 4. Result of activity weightlifting

Figure 5. Result of activity tennis

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 11
Article Received: *25 July 2023* ***Revised:*** *12 September 2023* ***Accepted:*** *30 November 2023*

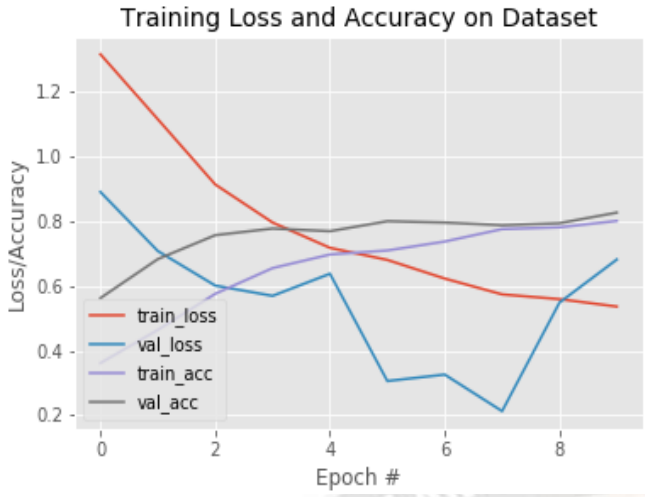

Figure 6. Result of 10 Epoch

Figure 7. Result of 50 Epoch

CONCLUSION

In conclusion, creating night sky panoramas presents significant challenges due to low signal-to-noise ratio (SNR) and motion blur. To overcome these obstacles, the incorporation of advanced feature extraction techniques, particularly Scale-Invariant Feature Transform (SIFT), becomes crucial. SIFT allows for the extraction of a greater number of features, essential for capturing the intricate details of night scenes.

Additionally, spatially variant registration steps were introduced into the panorama workflow. This innovative approach enabled the algorithm to merge multiple shorter exposures effectively, resulting in a final image with reduced noise and free from motion artifacts. Deblurring techniques such as using light streaks further enhanced the clarity of the image, ensuring a visually pleasing panorama.

Furthermore, the process involved matching extracted features and projecting them onto the appropriate surface, contributing to the seamless creation of the final panorama. Overall, by addressing the challenges of low SNR and motion blur through feature-rich extraction, spatially variant registration, and deblurring techniques, the algorithm successfully produced high-quality night sky panoramas.

REFERENCES

[1] M. Qiusha, L. Ziyi, and L. Wenhao, "Automatic Classification of Instructional Video Based on Different Presentation Forms," in 2023 IEEE 12th International Conference on Educational and Information Technology (ICEIT), 2023, pp. 353–357. doi: 10.1109/ICEIT57125.2023.10107851.

[2] R. Zheng, "Sports Video Classification and Labeling Algorithm Based on Data Mining Technology," in 2023 International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE), 2023, pp. 1–7. doi: 10.1109/ICDCECE57866.2023.10150741.

[3] X. Wang and J. Jiang, "Shear Detection and Key Frame Extraction of Sports Video Based on Machine Learning," in 2023 Asia-Europe Conference on Electronics, Data Processing and Informatics (ACEDPI), 2023, pp. 40–44. doi: 10.1109/ACEDPI58926.2023.00014.

[4] R. V. and B. N. K, "Quantum Video Classification Leveraging Textual Video Representations," in 2023 4th International Conference on Communication, Computing and Industry 6.0 (C216), 2023, pp. 1–6. doi: 10.1109/C2I659362.2023.10430918.

[5] S. Taleb and N. Abbas, "Hybrid Machine Learning Classification and Inference of Stalling Events in Mobile Videos," in 2022 4th IEEE Middle East and North Africa COMMunications Conference (MENACOMM), 2022, pp. 209–214. doi: 10.1109/MENACOMM57252.2022.9998209.

[6] K. Blekos and D. Kosmopoulos, "Quantum Data Reduction with Application to Video Classification," in 2022 IEEE/ACM 7th Symposium on Edge Computing (SEC), 2022, pp. 430–435. doi: 10.1109/SEC54971.2022.00065.

[7] Z. Dong, X. Wu, X. Zhao, F. Zhang, and H. Liu, "Identifying Pitfalls in the Evaluation of Saliency Models for Videos," in 2022 IEEE 14th Image, Video, and Multidimensional Signal Processing Workshop (IVMSP), 2022, pp. 1–5. doi: 10.1109/IVMSP54334.2022.9816306.

[8] X. Song, B. Xu, and Y.-G. Jiang, "Predicting Content Similarity via Multimodal Modeling for Video-In-Video Advertising," IEEE Transactions on Circuits and Systems for Video Technology, vol. 31, no. 2, pp. 569–581, 2021, doi: 10.1109/TCSVT.2020.2979928.

[9] Y. Lin, Z. Ning, J. Liu, M. Zhang, P. Chen, and X. Yang, "Video steganography network based on 3DCNN," in 2021 International Conference on Digital Society and Intelligent Systems (DSInS), 2021, pp. 178–181. doi: 10.1109/DSInS54396.2021.9670614.

[10] Y. Luo and B. Yang, "Video motions classification based on CNN," in 2021 IEEE International Conference on Computer Science, Artificial Intelligence and Electronic Engineering (CSAIEE), 2021, pp. 335–338. doi: 10.1109/CSAIEE54046.2021.9543398.

[11] L. Wu et al., "Ontology-Based Global and Collective Motion Patterns for Event Classification in Basketball Videos," IEEE Transactions on Circuits and Systems for Video Technology, vol. 30, no. 7, pp. 2178–2190, 2020, doi: 10.1109/TCSVT.2019.2912529.

[12] M. Abdullah, M. Ahmad, and D. Han, "Facial Expression Recognition in Videos: An CNN-LSTM based Model for Video Classification," in 2020 International Conference on Electronics, Information, and Communication (ICEIC), 2020, pp. 1–3. doi: 10.1109/ICEIC49074.2020.9051332.

[13] M. Liu, "Video Classification Technology Based on Deep Learning," in 2020 International Conference on Information Science, Parallel and Distributed Systems (ISPDS), 2020, pp. 154–157. doi: 10.1109/ISPDS51347.2020.00039.

[14] J. You and J. Korhonen, "Attention Boosted Deep Networks For Video Classification," in 2020 IEEE International Conference on Image Processing (ICIP), 2020, pp. 1761–1765. doi: 10.1109/ICIP40778.2020.9190996.

[15] I. Fomin, V. Burin, and A. Bakhshiev, "Research on Neural Networks Integration for Object Classification in Video Analysis Systems," in 2020 International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM), 2020, pp. 1–5. doi: 10.1109/ICIEAM48468.2020.9112011.