Available online at https://ijmras.com/
INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH AND STUDIES
ISSN: 2640 7272
Volume:06; Issue:06 (2023)
doi: 10.33826/ijmras/v06i06.5

An analysis of Video Categorization, including its Approaches, Results, Performance, Problems, Solutions, and Future Directions

LIN ZHENQUAN

Research Scholar Lincoln University College Malaysia

Abstract

Internet accessibility and bandwidth have improved dramatically in recent years. Since connecting to the Internet is now so cheap, it has facilitated the widespread and rapid dissemination of information in the forms of text, audio, and video. Predicting the appropriate category for this video footage is necessary for a variety of uses. For the sake of human efficiency, several machine learning approaches have been created for video categorization. Existing review articles on video classification have a number of drawbacks, including limited analysis, poor organisation, failure to disclose research gaps or conclusions, and inadequate description of benefits, drawbacks, and future directions for investigation. However, we believe that our review article comes close to surpassing these constraints.

This research aims to provide a comprehensive overview of the current state of video categorization by analysing and comparing the many approaches now in use and recommending the way that has shown to be the most successful and efficient. First, we look at how films are categorised using taxonomy, current applications, processes, and datasets. Second, the current connection in science, deep learning, and the model of machine learning, as well as the associated inconveniences, challenges, flaws, and possible work, data, and performance assessments. The study of video classification systems, including their characteristics, tools, advantages, and disadvantages, for the purpose of comparing the methods they have used, is a significant part of this review. Finally, we provide a tabular overview of key aspects. The RNN, CNN, and combination technique outperforms the CNN dependent approach in terms of accuracy and independence extraction functions.

Corresponding Author: LIN ZHENQUAN[1]. *Research Scholar Lincoln University College Malaysia .*

Keywords: Video Classification, Machine learning, Deep learning, Video, Video classification.

INTRODUCTION

Today, almost everyone in the globe uses the internet. When it comes to content dissemination (including audio, video, text, and picture files), social media platforms are indispensable [1]. At around the same time, people express their feelings on social media about that same feature so that other users may rapidly learn the whole story; for this reason, user opinions are utilised to gauge public sentiment. However, it is difficult and time intensive for consumers to hire a person to analyse the opinions of individuals via the vast amounts of material available. Scientists describe a machine learning strategy for data mining in order to gauge public opinion. For the purpose of discovering people's perspectives by collecting and analysing social and other resources of subjective information, video classification is a subfield of mining that applies NLP to text and machine linguistics to video. The deep learning technique has shown to be the most trustworthy and productive of the available options.

The methods for video categorization are compared and contrasted in this research. Several review and survey papers have been published on the topic of video categorization. Here we detail the works of a few recent review articles with their strengths and weaknesses. [2] Provides a great overview of deep learning with applications to video categorization and captioning. Although the deep model, data, and feature extraction techniques are all well-described, research gaps, benefits, and performance are all left out of this assessment of deep learning-based approaches to video categorization. Q. Ren presented a short overview of video categorization techniques in 2019. This is a fairly brief overview, since it only describes a method for classifying videos. Approach, dataset, performance indicators, research needs, and current method constraints are not described. In 2020, Anusya has provided a brief overview of methods for categorising videos [4]. This article only introduces the topic of video categorization for tagging and describes a few of the most up-to-date available methods in the field. This review is deficient in many key areas, including coverage, discussion of study constraints, and use of appropriate tools and methodologies. Rani [5] just conducted an analysis on video tagging in 2020. New methods for categorising videos are discussed, as well as a brief synopsis of related studies, in this analysis. Lacking a more in-depth explanation, as well as an analysis of the most current job to identify research output, gaps, and findings, are some of the work's major flaws. By 2020, a new, comprehensive evaluation [6] on the topic of classifying live sports videos will have been completed. This article provides a thorough presentation of the most up-to-date research on live sports video categorization, including related tools, feature extraction, video interaction characteristics, etc. This is a more in-depth analysis, and there is no table summarising the review's findings, research gaps, or the benefits and drawbacks of already used approaches.

The above justification suggests that most reviewers have conducted historical research reviews in the past. Recent reviews often include an overview of the methods together with descriptions of relevant findings. Similar patterns of categorization may be seen in the

Corresponding Author: LIN ZHENQUAN[1]. *Research Scholar Lincoln University College Malaysia* .

conventional survey paper for video research, whether we're talking about comparative or associated studies. Our research study paper is different from the others since we used a variety of critical research methods. The following are some of our most significant contributions to this review article:

First, we'll have a roundtable discussion on the classification taxonomy, current applications, tools, and data stores for videos.

Include future projects, data, performance measurements, and pertinent contemporary references to science, deep learning, and a model of machine learning in your discussion of the overall drawbacks, problems, and obstacles.

Third, by comparing the various methods, we can objectively examine the tools, benefits, and drawbacks of various video categorization systems. Display a summary table of chosen features.

Here's how the remainder of the paper is laid out. Background information on the study of video categorization techniques is provided in Section 1. In Section 2, we provide an in-depth evaluation of the most recent studies, including an outline of their key characteristics, limitations, and suggestions for further study. The research technique used in this overview is shown in Figure 1.

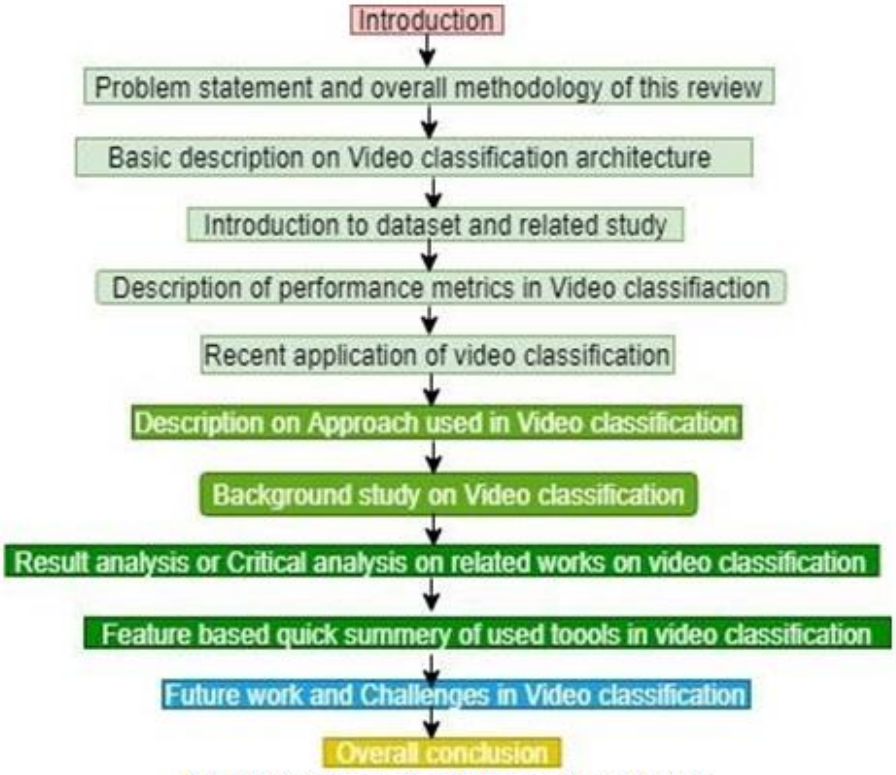

Fig. 1. Overall methodology of this review

Corresponding Author: LIN ZHENQUAN[1]. *Research Scholar Lincoln University College Malaysia .*

Engineering for video categorization

It is important to follow the prescribed sequence while using a video categorization algorithm. The foundational procedures for video labelling are shown in Figure 2. The data collecting phase is followed by the feature extraction and feature matching and classification technique execution phases. Video, text, audio, and picture data may all be reviewed as part of the data collecting process. The preprocessing stage of video processing involves a number of crucial steps, including the aforementioned video conversion, segmentation, and analysis, all of which are necessary for further feature or information extraction. The core of the video classification process is the application of an algorithm to the steps of feature extraction, feature matching, and feature classification.

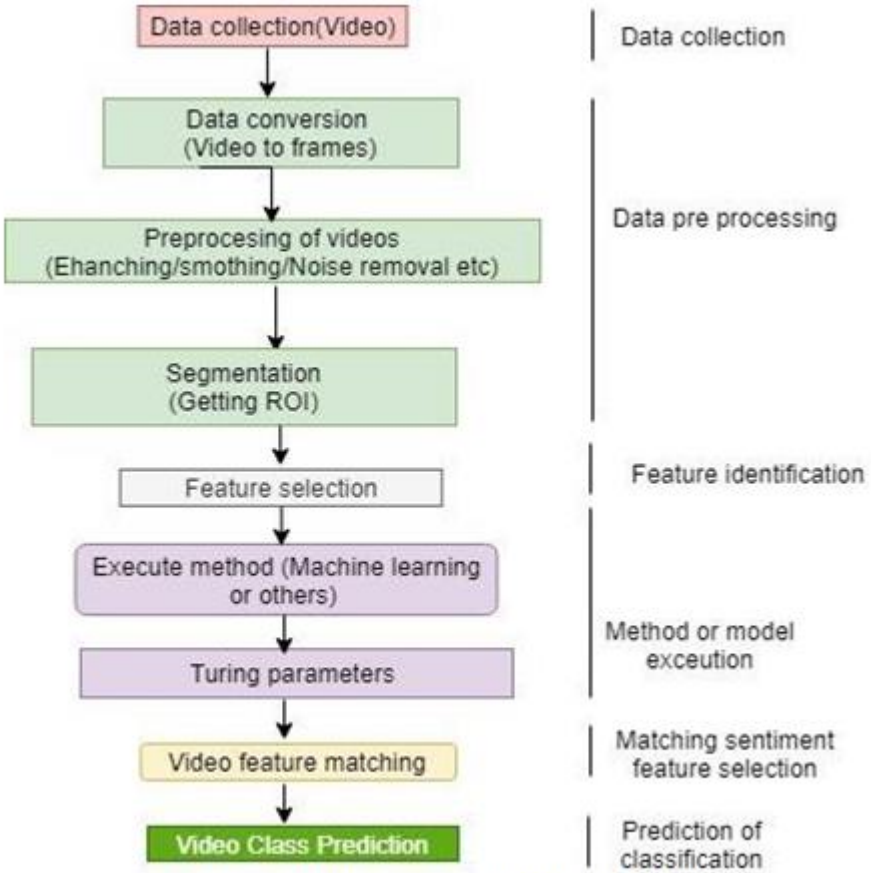

Fig. 2. Basic steps in video classification

Videos are classified using a data set

There has been a substantial time and effort investment in media-related study domains from several scientific and research institutions in the collection and annotation of video data sets. The most popular datasets are YouTube-8M, HCF-50, HCF-101, HMDB51, and many more. Weizmann, KTH, and Hollywood are examples of the smaller data sets; although their total number and video formats may be lower, they are very well-labeled. In addition, the medium set data has over 50 pictures including UCF101, Thumbos'14, and HMDB51. Google's massive data set from YouTube (8M videos), the Sports 1M database, ActivityNet, the Kinetics

Corresponding Author: LIN ZHENQUAN[1]. *Research Scholar Lincoln University College Malaysia* .

database, and others. Table 1 provides a summary of the data. Here, only the Weizmann and KTH datasets remain unchanged throughout time.

Table 1. Summary of Video task dataset

Name of the dataset	Year	Number of video categories	Amount of video
Weizmann	2005	9	81
KTH	2004	6	2361
Hollywood	2008	8	430
UCF50	2012	50	6676
HMDB51	2013	51	6474
UCF101	2012	101	13320
Thumos'14	2014	101	18394
Youtube-8M	2016	4800	8264650
Sports-1M	2014	87	1133158
ActivityNet	2015	203	27901
Kinetics	2017	400	306245

Video categorization performance metrics

In this part, we outline the most widely used criteria for evaluating the effectiveness of video categorization systems. Effectiveness of a dataset method may be shown with the use of performance metrics. Precession (Precision measurements are undertaking positive meaning determination), Recall, and other performance analysis processes all fall within the purview of video categorization (Precision tests are percentage tests for productive detection of positive result of the classifier). However, some of the performance metrics used to evaluate video classification studies are included in Table 2 below. The results of similar studies, measured in Table 2, are presented below.

Table 2. List of Performance Metrics used in video classification

Performance metrics	Reference and Year
Accuracy	[7]-2020, [8]-2020, [9]-2020
Precession	[8] [9]-2020
Recall	[8] [9]-2020
F1 Score	[8] [9]-2020
Micro F1	[10] [11]-2020
K-Fold:3,5,10-fold	[12]-2019

Classification of videos and their uses

Video categorization has a wide variety of uses. In Table 3, I include a few of them together with references to their most recent works. When using video for a firewall duty, the user must be certain that only the specified forms of video are permitted. Video analysis may be used for a wide variety of purposes, including but not limited to the following: live streaming prediction, action recognition, violence detection, character identification, traffic management, social media analysis, emotion analysis, movie review, and event prediction.

Corresponding Author: LIN ZHENQUAN[1]. *Research Scholar Lincoln University College Malaysia* .

Table 3. Summary of Application of Video classification

Name of Applications of Video classification	Authors and Years
Violence detection from video of real time game	[7] - 2020
Video Scene classification	[13] - 2020
Event prediction	[9] - 2020
Animation movie video classification	[8] - 2020
Sport player action recognition	[14] - 2020
Twitter video classification	[15] - 2020
Stock Market prediction	[16] - 2020
Movie video trailer classification	[17] - 2020

METHODS FOR FILTERING VIDEO CONTENT

Since there are so many videos out there, it's crucial to have a reliable system for categorising them. The primary goal of the video classification technique is to categorise videos according to their intended use (e.g., for sports, movies, comedic videos, educational purposes, etc.). Video may be broken down into three categories: audio, visual, and textual. There is also the option of utilising a hybrid strategy (combining two or more algorithms) to classify the videos. Video classification methods are shown in a taxonomy in Figure 3.

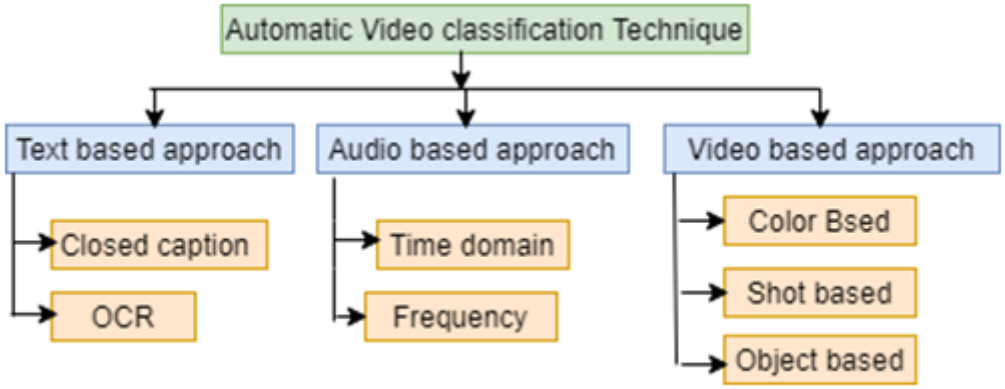

Fig. 3. Taxonomy of Video Classification Approaches

Strategy based on Textual Analysis

During this procedure, we create video texts and assess their classifiability. It's possible that this writing is either plainly legible, or that it was transcribed from spoken words. In the first group, we find the process that produces digital text. Details such as the scoreboard, player's jersey number, on-screen text, and so on are all examples. OCR could be used to get the text out of images like those [18, 19]. Throughout the second grouping, the text is extracted from speech by voice recognition. Subtitles and closed captions are the most common use of this method. Songs and pet noises are two examples of the many types of audio that benefit from closed captioning. Videos often have subtitles added for clarity.

Corresponding Author: LIN ZHENQUAN[1]. *Research Scholar Lincoln University College Malaysia .*

Method based on sound waves

The fact that audio processing requires less time and effort explains why it is being employed more often than text-based analysis. The data associated with audio takes up far less space than that of video or text. Audio processing involves sampling a single sound and then inspecting each sample for a set of attributes. The samples may overlap in certain cases. Functions might be based on either the time domain or the frequency domain.

Method based on videos

Since most vision-dependent information is understood by humans, the method was employed by the vast majority of academics. Some writers have used visual elements like music and text when it was essential. Ability to see is mostly acquired from collections of still images or motion videos. One of video's defining characteristics is its resemblance to a series of still images. Alternatively, we may call what we see on a video a series of still images. Color, motion, and length of time in frame are common visual features. These aspects must transmit information about the video's illumination, movement, backdrop, or frame rate.

Methods for classifying videos: a comparison

Based on the explanation, it is clear that each method has its own unique workings and success outcomes; thus, we give a comparison table that ranks the methods according to how well they fit into the context of current applications. The benefits and drawbacks of every approach are laid forth in Table 4.

Table 4. Comparison among video classification approach.

Classification Method	Feature list	Advantages	Disadvantages
Text Based Video Classification	Optical Character Recognition Closed Captions features Speech Recognition	Higher accuracy Higher dimensionality	Expensive in computation Higher error rate Works of text-based format only.
Audio Based Video Classification	Physical Features as well as Perceptual Features	Short length, Computationally cheaper	Difficult to differentiate multiple and similar sounds
Video Based Video Classification	Color Based Features, Shot-Based Features and Object-Based Features	Easily implemented. Not in converted representation.	Large size Computation is expensive Pre-processing is needed Identification of shots, track is difficult

Video Classification Methodology: A Historical Perspective

Common techniques include the supervised SVM and CNN as well as the unsupervised. The video categorization system provides a wide range of answers (LSTM, GRU etc). In this part, we will look at the most popular approaches to video categorization, highlighting their underlying mechanisms, practical benefits, and drawbacks. Dictionary-based Naive Bayes for video categorization [8]. It has been shown that a classifier based on Naive Bayes functions performs better than other models if the assertion of independent predictors is true. Inference via independent predictors is the basic model that Naive Bayes attempts to mimic. Similarly,

Corresponding Author: LIN ZHENQUAN[1]. *Research Scholar Lincoln University College Malaysia* .

support vector machines (SVM) is a popular detection approach used in video classification [20]. The categorization of videos on the internet is another method used to detect hate speech [14]. Working with a support vector machine (SVM) tool to categorise the pilot and to weight production to enhance classification accuracy was another task related to Twitter video classification [9]. When data is noisy or there is overlap between target classes, the SVM algorithm struggles to perform effectively.

Video labelling may be accomplished in a number of ways using K techniques. Recently, Peng [13] completed another another video categorization assignment. To segment videos and recover their visual characteristics, this method is utilised. The regular k-means aggregation approach improves the original grouping values of labialized video samples. When we have k smalls, k-means calculation is often quicker than hierarchical clusters. The HMM (Hidden Markov Model) is used for classification and extraction, however the K-proximate neighbour (KNN) approach is simple and quick to implement, with the downside being that K-value is hard to predict. Real-time video surveillance using R-CNN and HMM, with a focus on the research of children's facial speech [21]. Advances in HMM Methodology Different-length inputs are the simplest generalisation for sequence data, and there is a strong theoretical foundation for quick learning algorithms using raw sequence information.

There are a few problems with HMM. There are a large number of loosely defined criteria in HMMs, and they cannot depend on any hidden states. An effective deep pipeline template-based designs to accelerate the full 2-D and 3-D CNNs on FPGA [22] is shown in a work that demonstrates that 3D CNN is better suited to the categorization of the video. Three-dimensional deep convolutional neural networks were employed for action recognition [23]. In 3D convolutions, both spatial and temporal data may be properly combined. Plan for the future Recurrent neural networks directly translate temporal dynamics to video frames of varying lengths. An RNN does this by creating knowledge-preserving looping networks [24]. This iterative loop structure will be used by the neural network to store the input sequence. This is how the RNN works. RNN uses the past feedback to help us find meaning in every situation.

There are two different LSTM varieties in RNN, in addition to the other kind, GRU. A recurrent neural network (RNN) with long short-term memory (LSTM) neurons is being trained on SIFT characteristics in sports videos [25]. The reliability of Baccouche work has been well admired. With the development of deep learning algorithms and architecture, function extraction can now be done mechanically. RNN might be fine-tuned through back propagation. The 2D Gated Bidirectional Neural Networks were used on new, higher-quality footage to detect aggressive behaviour. Gated Return Units (GRUs) are an existing CNN feature that Kyunghyun created [26]. When compared to other methods, deep literacy has been shown to be more dependable and effective [27]. The method of instruction is more efficient. The video categorization system using deep learning is compared in detail in Table 5.

Corresponding Author: LIN ZHENQUAN[1]. *Research Scholar Lincoln University College Malaysia* .

Table 5. Overall comparison among deep learning-based method for video classification.

Model	Advantage	Drawback
2D CNN	Can capture spatial feature.	Can capture spatial feature from video data
3D CNN	Can capture both spatial and temporal feature.	Expensive for its 3D structure for working
RNN	Can capture both spatial and temporal feature from sequence data.	Has short memory ability, could not be able in real situation.
LSTM	Can capture both spatial and temporal feature from sequence data.	Gradient explosion, Takes more training time.
GRU	Can capture both spatial and temporal feature from sequence data in a faster time	The reset gate of GRU controls if the previous hidden state needs to be ignored.

Discussion of findings, including an evaluation of relevant literature, on the topic of Video Categorization

To overcome the shortcomings of currently applicable methodologies, several deep learning systems may use extensive data collecting and processing power to achieve higher levels of precision and accuracy. In this section, we examine the most up-to-date developments in video categorization and offer our findings. Data, techniques, model, kind, advantage, and disadvantage of the most up-to-date video classification methods in 2020 are presented in analytical style columns in Table 6. There are a number of related papers in the literature on video categorization. The two main categories of conventional techniques are classical machine learning and deep learning. Video categorization often employs the support vector machine (SVM), a technique that employs Naive Bayes and Dictionary [8]. Multiple approaches to video categorization using the letter K. Peng [14] just conducted some work to classify videos. The most recent study [28] employed the Random Forest algorithm to classify videos found on YouTube. Video categorization using K-nearest neighbour diagrams and the end-to-end information diagrams [29].

Real-time video surveillance using R-CNN and HMM, with a focus on the research of children's facial speech [21]. Automatically identifying certain input data or vine frames allows a deep learning structure to learn then represent data across multiple processing layers [30]. There's no need for unique IDs or extractors that are useful in most architectural designs. In the deep learning approach, for instance, it is the local properties of an image that are first learnt, rather than the global context [31]. Methods of deep learning that can detect subtle or nuanced behaviour are the subject of intensive study [32]. The convolutional neural network (CNN), the repeated neural network (RNN), and the long-term memory are all examples of popular deep learning models (LSTM). Success with a high accuracy of deep learning technology in such a visual job inspired its application to video data processing. At initially, CNN runs independently for still-image data extraction [23]. Despite this, 2D-CNN is unable to collect transient information from video streams. The study [33] evaluates CNN with LSTM-RNN and identifies the possibility for a stronger development of Recurrent Convolutional Neural Networks; it employs CNN for the huge video categorization and demonstrates that the slow melting system performs better than the normally early fusion model. In [35], a CNN with a two-stream structure is being employed, with one stream responsible for spatial functionality and the other for temporal.

Corresponding Author: LIN ZHENQUAN[1]. *Research Scholar Lincoln University College Malaysia* .

Activity and behaviour recognition are achieved by the use of description and CNN in [36] research. Grade bundling encrypts information about certain time periods by combining video frames together. The method for learning will use the synergies between neural network convolutions to achieve Bilevel optimization. Batch standardisation and the CNN extractor Optimizing performance with an LSTM function extractor is another possibility [37]. To represent the interdependencies between features, [38] proposed non-linear context gating, which was then utilised to categorise films. The solution to this issue for 2D CNN is 3D-CNN, which is built to recover both spatial and temporal information from video frames [39]. This RNN then proceeded to the behaviour identification phase. Knowledge of time, both present and past, may be recorded effectively using the RNN-based method [40]. The data we have now allows us to make this prediction. However, in practise, the short-term memory of RNN design cannot be increased. To help with this issue, the LSTM model was proposed. The time sequence of a movie may be determined by using this model. When and why a secret state is stored in the LSTM model's memory device is up for grabs [41]. Due of its superiority, the LSTM model is widely used in computer vision applications like action recognition. Recent methods for video categorization are summarised in Table 6.

Corresponding Author: LIN ZHENQUAN[1]. *Research Scholar Lincoln University College Malaysia* .

Table 6. Video classification recent approach

Author and Year	Type	Task	Lexicon or dataset	Performance and Data domain	Approach	Video analysis Features	Type of data	Advantages or findings	Disadvantages or limitations
[17] - 2020	Deep learning	Video movie analysis	LMTD-9, MMTF-14, and ML-25M	The combined accuracy of ILDNet for LMTD-9, MMTF-14, and ML-25M 86.15%, 83.06%, 85.3% respectively	Bi-LSTM, and LSTM	Can acquire discriminative and compre-hensive higher level features with a unique combination of Inception V4, Bi-LSTM, and also LSTM layers.	Video	Can recognize six types of emotion from video trailers	Performance limited predefined and compared EmoGDB dataset content
[7]- 2020	Deep learning	Video violence detection	Hockey game dataset, Violent Flow dataset and Video Real Life Violence Situations dataset	Accuracy:98%	2D CNN, BiGRU	A simple end-to-end deep learning method to find violence in video sequences with CNN and RNN.	Video	Can detect violence in video sequences.	Works with small dataset, Higher training time.
[9]- 2020	Machine learning	Video analysis	Chilean earthquake and Catalan independence referendum	Dataset 1: Highest accuracy with SVM 0.812±0.067, BAF TAN got highest Precession 0.898±0.003 F1 Score with SVM of 0.899±0.042, BFTAN got highest Recall 0.898±0.003. Dataset-2 :highest : RF got highest precession 0.922±0.002, RF highest accuracy 0.858±0.008, Highest Recall 0.985±0.008 by SVM, RF got highest F1 Score 0.908±0.00.	Bayesian network classifiers, Uses TAN and BF TAN	BOW, Term document matrix (TDM) for tweet to vector representation	Video	Provides good result by using Bayes classifier with Directed acyclic graph (DAG) for twitter comments emotion.	Biased to hashtag, Heavily time and event dependent
[13] - 2020	Hybrid	Video analysis	Haberman sub-library data, German Credit Data sub-library, Heart sub-library dataset, UCI data.	Accuray: Haberman sub-library 72%, German Credit Data sub-library 75%, Heart sub-library data 92%	CNN, K means	Clustering video using k means algorithm.	Video	Handle multi-visual features	Dataset is small, Accuracy is not high.
[8] - 2020	Hybrid	Emotion analysis from Video	Danmaku reviews	The F1 scores by SD-NB is 82.3% for positive class and also 93.6% for the negative class. Also shows good result for precession and recall.	Sentiment diction ary and and naïve bayes	Can classify video sentiment and opinion	Text and Video	Can classify seven video sentiment	Domain depended
[21] - 2020	Deep learning	Video analysis of infant	Clinical dataset for infant expression	Mean average precision is 81.9% and also 84.8% for four infant expressions and three states evaluated with both clinical and daily datasets. Precision for discomfort detection 90%.	HMM, CNN	Does infant expressions and also the states detection, object tracking and detection compensation with HMM and R-CNN.	Image	Nicely do expression detection, Utilization of HMM with CNN is good.	Unable to handle temporal data.

Various approaches to video categorization using text, audio, and video feature extraction may be found in the literature. Each algorithm, such as HMM, ANN, SVM, or RNN, has its own strengths and weaknesses. By combining two or more of these strategies, you may get the benefits of both plans at once.

Corresponding Author: LIN ZHENQUAN[1]. *Research Scholar Lincoln University College Malaysia* .

Brief overview of feature-based video categorization methods

In order to make an educated guess as to the final result of video categorization, several techniques are used. The parameters of a multilayer neural network are learned and fine-tuned, and the network is normalised and supported by a deep learning and machine learning based video categorization system. In this part, you'll find a concise table summarising the specifics of the techniques used to categorise the videos. To get optimal results, mixed or deep learning makes use of both traditional machine learning and underlying neural network models. Video categorization makes use of a wide variety of techniques and algorithms. In this article, we provide a high-level summary of the systems and methods currently in use for video categorization. Based on the approaches, resources, and algorithms used, we selected the twenty characteristics shown in Table 7.

Table 7. Features used in video classification method.

F1:3D CNN	F8:CNN	F15:Dependency tree
F2:Random Forest	F9:Naive Bayes	F16:Machine learning
F3:HMM	F10:Postag	F17: Deep learning
F4:GMM	F11: KNN	F18: Hybrid
F5:2D CNN	F12: Capsule network	F19:Tfidf
F6:RNN	F13:LSTM	F20:K means
F7:Support Vector Machine	F14:GRU	

Table 8 provides a concise summary of modern machine learning-based categorization algorithms. The twenty characteristics chosen from Table 7 are the subject of this table. With this presentation, we are introducing recent work and use a sign to illustrate the marching features with twenty characteristics.

Table 8. Overview of feature-based video classification.

Authors and Year	F_1	F2	F3	F4	F5	F6	F7	F8	F9	F9	F11	F12	F13	F14	F15	F16	F17	F18	F19	F20
[7]-2020	✓				✓		✓		✓							✓	✓			
[9]-2020							✓	✓								✓				
[13]-2020								✓								✓	✓	✓		✓
[8]-2020									✓							✓	✓	✓		
[14]-2020							✓									✓	✓	✓		
[21]-2020			✓						✓							✓				
[15]-2020									✓	✓		✓				✓				
[28]-2020		✓														✓				
[42]-2020				✓												✓				
[39]-2017	✓															✓				
[40]-2016						✓										✓				
[41]-2015						✓							✓	✓		✓				
[43]-2009																✓			✓	
[29]-2017											✓					✓				
[16]-2018							✓									✓				
[17]-2020						✓							✓			✓				

Corresponding Author: LIN ZHENQUAN[1]. *Research Scholar Lincoln University College Malaysia .*

The next steps and difficulties in video categorization

In the long run, this study might be expanded to include more techniques. However, efficient video categorization relies heavily on the properties of frames as well as frame retrieval. Additionally, patterns can improve the accuracy of classifications. An further obstacle may arise from the incorporation of increasingly diverse forms of video into the dataset with improved efficiency and general functionality, calling for investigation of techniques that explicitly define camera motion. We need this so that we can categorise lengthier films, detect several actions inside a video, discover relationships between videos, and categorise the actions of various objects within a video. Trending work in video categorization is the prediction of live-streamed video games.

CONCLUSION

This article provides a comprehensive analysis of many video categorization approaches and methods, discussing their merits, shortcomings, difficulties, data summaries, research voids, and overall performance. Based on this paper's analysis. It has been determined that a video-based technique is superior than textual or aural approaches for classifying videos. Text extraction has replaced video categorization as the least used procedure. Although audio and video feature extractions have their uses, it is clear that the performance of classification tasks may be improved if equal weight is given to the gathering of visual and aural information in the video itself. In contrast, audio-based solutions need less processing power. In addition, we have the option of using several techniques to recognise films, allowing us to circumvent the shortcomings of currently used approaches. To develop novel methods, we will first segment photos, then apply the threshold to distinguish between them, and last categorise them. It's possible that we'll utilise a categorization system that predicts how movies, games, and events will do. This is your opportunity to categorise films based on their content, such as the movie's theme, the music used in it, the intensity of the battle scenes, or the comedy of the comedic moments. Numerous approaches have been developed and shown effectiveness for video categorization. In addition to highlighting the benefits of new approaches, this review article highlights the drawbacks of existing methods, such as their inability to process multiple features simultaneously, their comparatively lengthy training periods for deep learning and traditional machine learning, and their subpar accuracy when dealing with multilevel videos. Researchers may capitalise on current tendencies and open doors by working to improve video categorization. To categorise lengthier movies, to identify various actions inside a video, to discover relationships between videos, and to classify the behaviour of several objects within a video. Video categorization is a growing field, and predictions for live-streamed video games are on the rise.

Corresponding Author: LIN ZHENQUAN[1]. *Research Scholar Lincoln University College Malaysia* .

REFERENCES

[1] Brezeale, D. and D.J. Cook, Automatic video classification: A survey of the literature. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 38, no. 3, p. 416-430, 2008. DOI: https://doi.org/10.1109/TSMCC.2008.919173

[2] Wu, Z., et al., Deep learning for video classification and captioning, in Frontiers of multimedia research, 3122867 p. 3-29, 2017. DOI: https://doi.org/10.1145/3122865.3122867

[3] Ren, Q., et al., A Survey on Video Classification Methods Based on Deep Learning. DEStech Transactions on Computer Science and Engineering, cisnrc, 33301 .p. 1-7, 2019. DOI: https://doi.org/10.12783/dtcse/cisnrc2019/33301

[4] Anushya, A., VIDEO TAGGING USING DEEP LEARNING: A SURVEY, International Journal of Computer Science and Mobile Computing,Vol.9 Issue.2,pg. 49-55,2020.

[5] Rani, P., J. Kaur, and S. Kaswan, Automatic Video Classification: A Review. EAI Endorsed Transactions on Creative Technologies, ,7(24), p. 163996,2020). DOI: https://doi.org/10.4108/eai.13-7-2018.163996

[6] Li, Y., C. Wang, and J. Liu, A Systematic Review of Literature on User Behavior in Video Game Live Streaming. International Journal of Environmental Research and Public Health, vol. 17, no. 9, p. 3328,2020. DOI: https://doi.org/10.3390/ijerph17093328

[7] Zhen, M., et al. Learning Discriminative Feature with CRF for Unsupervised Video Object Segmentation. in European Conference on Computer Vision. Springer, LNCS, volume 12372,pp 445-46,2020. DOI: https://doi.org/10.1007/978-3-030-58583-9_27

[8] Li, Z., R. Li, and G. Jin, Sentiment Analysis of Danmaku Videos Based on Naïve Bayes and Sentiment Dictionary. IEEE Access, vol. 8, p. 75073-75084,2020. DOI: https://doi.org/10.1109/ACCESS.2020.2986582

[9] Ruz, G.A., P.A. Henríquez, and A. Mascareño, Sentiment analysis of Twitter data during critical events through Bayesian networks classifiers. Future Generation Computer Systems, 106: p. 92-104,2020. DOI: https://doi.org/10.1016/j.future.2020.01.005

[10] Xu, Q., et al., Aspect-based sentiment classification with multi-attention network. Neurocomputing, vol. 388, p. 135- 143, 2020. DOI: https://doi.org/10.1016/j.future.2020.01.005

[11] Bibi, M., et al., A Cooperative Binary-Clustering Framework Based on Majority Voting for Twitter Sentiment Analysis. IEEE Access, Vol. 8, p. 68580 - 68592,2020. DOI: https://doi.org/10.1109/ACCESS.2020.2983859

[12] Sailunaz, K. and R. Alhajj, Emotion and sentiment analysis from Twitter text. Journal of Computational Science, vol. 36, p. 101003, 2020. DOI: https://doi.org/10.1016/j.jocs.2019.05.009

[13] Peng, T., et al., Video Classification Based On the Improved K-Means Clustering Algorithm. E&ES, vol. 440, no. 3, p. 032060,2020. DOI: https://doi.org/10.1088/1755-1315/440/3/032060

[14] Li, X. and S. Geng, Research on sports retrieval recognition of action based on feature

Corresponding Author: LIN ZHENQUAN[1]. *Research Scholar Lincoln University College Malaysia* .

extraction and SVM classification algorithm. Journal of Intelligent & Fuzzy Systems, vol. 39, no. 4, pp. 5797-5808, 2020. DOI: https://doi.org/10.3233/JIFS-189056

[15] Alomari, E., R. Mehmood, and I. Katib, Sentiment Analysis of Arabic Tweets for Road Traffic Congestion and Event Detection, in Smart Infrastructure and Applications, Springer. p. 37-54, 2020. DOI: https://doi.org/10.1007/978-3- 030-13705-2_2

[16] Ren, R., D.D. Wu, and T. Liu, Forecasting stock market movement direction using sentiment analysis and support vector machine. IEEE Systems Journal, vol. 13, no. 1, p. 760-770, 2020.DOI: https://doi.org/10.1109/JSYST.2018.2794462

[17] Yadav, A. and D.K. Vishwakarma, A unified framework of deep networks for genre classification using movie trailer.
Applied Soft Computing, vol. 96: p. 106624, 2020. DOI: https://doi.org/10.1016/j.asoc.2020.106624

[18] Parameswaran, S., et al., Exploring Various Aspects of Gabor Filter in Classifying Facial Expression, in Advances in Communication Systems and Networks, Springer. p. 487-500, 2020. DOI: https://doi.org/10.1007/978-981-15- 3992-3_41

[19] Hauptmann, A., et al., with the Informedia Digital Video Library System, MULTIMEDIA '94,Pages 480–481, 1994.

[20] Warner, W. and J. Hirschberg. Detecting hate speech on the world wide web. in Proceedings of the second workshop on language in social media. 2012. Association for Computational Linguistics. (LSM 2012), pages 19–26, 2012.

[21] Li, C., et al., Infant Facial Expression Analysis: Towards A Real-time Video Monitoring System Using R-CNN and HMM. IEEE Journal of Biomedical and Health Informatics, 9254091, pp 1-12, 2020. DOI: https://doi.org/10.1109/JBHI.2020.3037031

[22] Shen, J., et al., Towards an efficient deep pipelined template-based architecture for accelerating the entire 2D and 3D CNNs on FPGA. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 2019. 1442
- 1455,Vol. 39, no. 7, July 2020. DOI: https://doi.org/10.1109/TCAD.2019.2912894

[23] Meng, B., X. Liu, and X. Wang, Human action recognition based on quaternion spatial-temporal convolutional neural network and LSTM in RGB videos. Multimedia Tools and Applications, vol. 77, no. 20, p. 26901-26918,2018.
DOI: https://doi.org/10.1007/s11042-018-5893-9

[24] Yang, H., et al., Asymmetric 3d convolutional neural networks for action recognition. Pattern recognition, vol. 85, p. 1-12, 2019. DOI: https://doi.org/10.1016/j.patcog.2018.07.028

[25] Kar, A., et al. Adascan: Adaptive scan pooling in deep convolutional neural networks for human action recognition in videos. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. (CVPR), pp. 3376-3385,2017. DOI: https://doi.org/10.1109/CVPR.2017.604

[26] Cho, K., et al., Learning phrase representations using RNN encoder-decoder for statistical machine translation.
arXiv preprint arXiv:1406.1078, p. 1-45, 2014. DOI: https://doi.org/10.3115/v1/D14-1179

[27] Shofiqul, M.S.I., N. Ab Ghani, and M.M. Ahmed, A review on recent advances in Deep learning for Sentiment Analysis: Performances, Challenges and Limitations. COMPUSOFT: An International Journal of Advanced Computer Technology, vol. 9, no. 7, p. 3768-3776, 2020.

[28] Kalra, G.S., R.S. Kathuria, and A. Kumar. YouTube Video Classification based on Title and Description Text. in

Corresponding Author: LIN ZHENQUAN[1]. *Research Scholar Lincoln University College Malaysia* .

2019 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS). 2019. IEEE.

ICCCIS48478,p. 8974514,2019. DOI: https://doi.org/10.1109/ICCCIS48478.2019.8974514

[29] Yuan, F., et al., End-to-end video classification with knowledge graphs. arXiv preprint arXiv:1711.01714, 2017. 1711.01714, pp 1-9, 2017.

[30] Voulodimos, A., et al., *Deep learning for computer vision: A brief review.* Computational intelligence and neuroscience, 7068349, pp 1-13, 2019. DOI: https://doi.org/10.1155/2018/7068349

[31] Sargano, A.B., P. Angelov, and Z. Habib, *A comprehensive review on handcrafted and learning-based action representation approaches for human activity recognition.* applied sciences, vol. 7, no. 1, p. 110,2017. DOI: https://doi.org/10.3390/app7010110

[32] Elboushaki, A., et al., *MultiD-CNN: A multi-dimensional feature learning approach based on deep convolutional networks for gesture recognition in RGB-D image sequences.* Expert Systems with Applications, vol. 139: p. 112829, 2020. DOI: https://doi.org/10.1016/j.eswa.2019.112829

[33] Huiqun, Z., W. Hui, and W. Xiaoling. *Application research of video annotation in sports video analysis*. in *2011 International Conference on Future Computer Science and Education*.IEEE, 6041660, p. 1-5, 2011. DOI: https://doi.org/10.1109/ICFCSE.2011.24

[34] Herath, S., M. Harandi, and F. Porikli, *Going deeper into action recognition: A survey*. Image and vision computing, vol. 60, p. 4-21, 2017. DOI: https://doi.org/10.1016/j.imavis.2017.01.010

[35] Chen, H., et al., *Action recognition with temporal scale-invariant deep learning framework.* China Communications, vol. 14, no. 2, p. 163-172, 2017. DOI: https://doi.org/10.1109/CC.2017.7868164

[36] Peng, X., et al. *Action recognition with stacked fisher vectors*. in *European Conference on Computer Vision*, Springer.
ECCV,2014,pp 581-595, 2014. DOI: https://doi.org/10.1007/978-3-319-10602-1_38

[37] Lan, Z., et al. *Beyond gaussian pyramid: Multi-skip feature stacking for action recognition*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*, (CVPR), pp. 204-212, 2015.

[38] Dalal, N., B. Triggs, and C. Schmid. *Human detection using oriented histograms of flow and appearance*. in *European conference on computer vision*, Springer. ECCV, p. 428-441, 2006. DOI: https://doi.org/10.1007/11744047_33

[39] Asadi-Aghbolaghi, M., et al. *A survey on deep learning based approaches for action and gesture recognition in image sequences.* in *2017 12th IEEE international conference on automatic face & gesture recognition (FG 2017)*,
IEEE. 7961779, p. 1-8, 2017. DOI: https://doi.org/10.1109/FG.2017.150

[40] Yang, X., P. Molchanov, and J. Kautz. *Multilayer and multimodal fusion of deep neural networks for video classification*. in *Proceedings of the 24th ACM international conference on Multimedia*, 2964297, p. 978–987. 2016.
DOI: https://doi.org/10.1145/2964284.2964297

[41] Yue-Hei Ng, J., et al. *Beyond short snippets: Deep networks for video classification*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*,(CVPR), p. 4694-4702, 2015.

Corresponding Author: LIN ZHENQUAN[1]. *Research Scholar Lincoln University College Malaysia* .

[42 Dvir, A., et al., *Encrypted Video Traffic Clustering Demystified.* Computers & Security, Volume 96, p. 101917, 2020.

DOI: https://doi.org/10.1016/j.cose.2020.101917

[43] Yin, D., et al., *Detection of harassment on web 2.0.* Proceedings of the Content Analysis in the WEB, 2: p. 1-7, 2009.

Corresponding Author: LIN ZHENQUAN[1]. *Research Scholar Lincoln University College Malaysia* .