International Transactions on Electrical Engineering and Computer Science
G. Syam Kumar, Vol: 3, No: 2, pp: 92-100, June 2024

Sports Videos Classification using Advanced Deep Neural Networks

G. Syam Kumar

Abstract: The field of digital content is experiencing a meteoric rise in popularity as a direct result of the rapid development of information technology. When it comes to the archiving of digital content on the assistant, the segregation in sports videos is of an extremely important part. Consequently, the utilization of deep-neural-network algorithm (DNN), convolutional-neural-network (CNN), and deliver learning allows for the correct segregation of sports video classification to be achieved. There are two methods that have been proposed: block-brightness-comparison-coding (BICC) cum block colour histogram. Both of these methods analyze the contrast relationship among various parts of a video cum the colour matter that is present in a sector. In order to accomplish the goal of transfer learning, the maximum-mean-difference (MMD) procedure is utilized. Obtaining characteristics in sports video pictures is the foundation for the sports video image segregation approach that is dependent on deep-learning-coding model. This method is utilized in order that accomplish task of sports video segregation. As a consequence of the findings, it is clear that the overall segregation reaction of this procedure is significantly superior to that of other sports video classification methods that are currently in use. This results in a significant improvement in the classification effect of sports videos.

Keywords: Sports video, Convolution neural networks, Deep learning, Maximum mean difference

1. Introduction

In the field of artificial intelligence, a growing number of distinct approaches are being developed with the expansion of theoretical frameworks, and these approaches play a significant part in the field. There is a growing interest in deep learning as a field of study. The ongoing advancements and developments in deep learning are the driving force behind the creation of deep neural networks (DNN). The DNN architecture is the one that is utilized the most frequently in deep learning [1].

Article History
Received: 16-03-2024; Revised: 17-05-2024;
Accepted: 25-05-2024

✉ G. Syam Kumar
pdjntukakinada@gmail.com

Department of Physical Education, Jawaharlal Nehru Technological University Kakinada, Kakinada-533003, India

Within the year 2006, Hilton et al. were the ones who initially put forward the idea of deep learning. In recent years, it has excelled in a variety of domains, which has made it a topic of discussion in the field of artificial intelligence [2]. The primary objective of deep learning is to create a model that is analogous to the neuronal makeup of the brain of human. Its envisaged machine would be able to process the data that it has received in the same manner as a human. During the realization phase, features are extracted from the data by going through a number of nonlinear steps repeatedly [3]. Anyways if we prefer to solve the problem considering MN Algorithm, then we go use either DL algorithm or ML Algorithm then its confronted on few real time conditions. These conditions include a lack of data and information, as well as a contrast among trained data cum data used in real applications [4]. It is necessary to collect sufficient annotation information whenever there is a new application scenario. This is due to the fact that classical machine learning demands that the learning data and the data from real-world application scenarios have comparable statistical properties.

https://doi.org/10.62760/iteecs.3.2.2024.92

This implies that in order for typical machine learning techniques to be useful in a new application environment, sufficient data must be gathered. This will require a significant investment of time, money, and labor [5]. People are paying an increasing amount of attention to this issue as a result of the speedy growth of machine learning in current years. Transfer learning, which is one of the learning methods that can be used to resolve the issue, has become active study topic [6]. Purpose of transfer learning to make utilize of previously acquired information in order for solve issues that are not only linked to one another but also to other subjects, with the end objective of achieving knowledge transfer between fields that are related. It is a method of intelligent learning that is comparable to the fundamental learning ability of humans. As is the case in real life, we humans are able to determine whether the new challenges we are confronted with are connected to the accumulation of knowledge that we have acquired in the past and whether we are able to incorporate the knowledge that we have acquired through the knowledge that we have acquired or accumulated in the past [7]. It has been discovered by psychologists in order for achieve transfer learning, its absolutely required for the 2 areas or learning materials that share same elements. Furthermore, The fraction of the transfer impact to the common components, and the extra same factors there are the best that the transfer source is. Obviously, more elements that are shared by many people as well as factors that are unique to each individual that can also have a detrimental impact on learning [8]. In the realm of video, sports video is an extremely valuable resource. All across the world, there are hundreds of millions of devoted spectators who watch sports broadcasts continuously. The categorization of sports videos has emerged as a primary topic for a significant number of scholars. The intelligent classification technology for sports videos not alone intelligently categorize cum sort out enormous amounts of sports video information, hence reducing the amount of work that people have to do, but it may also lower the amount of spiritual satisfaction that people have in their daily lives. In addition to that, it serves as the foundation for intelligent radio and television classification technologies. As a result, the intelligent categorization technology for sports videos has the potential to be very useful in a variety of disciplines, including sports video management, retrieval, and querying, and it also offers a wide range of development possibilities and a great deal of value [9]. Currently, the traditional method of transfer learning consists primarily of locating a shared feature space cum mapping the information from 2 different sources with the feature space using a function called mapping. The goal of this mapping function is to minimize the difference in data distribution between the two domains while preserving the basic characteristics in the information to the greatest extent possible. After this, the classifier is trained using the mapped data [10]. The extraction of video features is given a greater amount of attention by the sports video categorization systems that are currently available. It is clear that the extraction of features is the most important step in the segregation in sports videos. In terms of motion, color, edge, and other properties, researchers have developed a variety of feature models for classification. However, these models are distinct from one another. Sports videos take on a distinct role than other types of videos. Particular characteristics have a strong impact on classification. It should come as no surprise that the characteristics These methods' recoveries aren't much more expressive than deep learning's. Furthermore, in practical applications, deep transfer learning technology can more closely fit the end-to-end needs [11].

Movement, shade, luminosity, border, feel, and sound, and several other aspects are the primary focuses of video features. The majority of the time, these features are utilized in the classifier in a manner that is either distinct or straightforward linear fusion-horizontal. In other words, various featured vectors were utilized by a linear mixed in accordance with the regulation. In a certain sense, the classification performance is enhanced by the utilization of this fusion method. The majority of them, on the other hand, isolate the various features cum disregard the semantic bonding that exists among the features. A merit sports video segregation technology that is dependent on deep neural networks cum transfer learning is proposed in this paper. Additionally, an in-depth subject of sports video segregation is taken into consideration [12]. The purpose of this paper is to develop more connections and fusions between characteristics. It is proposed that block contrast comparing coding cum block colour histogram be utilized in order to replicate the brightness bonding that exists among various parts in video as well as the

colour data that is existing inside the region. The concept of transfer learning is presented in the fourth paragraph of the article. For the cause of migration learning maximum mean difference (MMD) method has been conceptualized. The categorization and practical results in migrating learning sports video-images dependent with deep learning coding algorithm are shown in the fifth paragraph in the article. Getting the characteristics of sports-video pictures is the foundation for the sports-video image segregation approach that is dependent on the deep learning coding algorithm. This method is utilizes to accomplish the task of sports video segregation. The sports videos are arranged in the order of form figure-skating-badminton cum yoga so that the efficiency in technique categorization described in the article may be evaluated to determine how well it works. As a result of the supervised finetuning, the constraints in each layer in the deep learning network also be modified through the use called error backpropagation in order to achieve the best possible segregation effect.

Purpose of this study is to investigate a video segregation transfer learning technique that is dependent on DNN. This is accomplished by employing an examination of both regular transfer learning algorithms cum DNN transfer learning algorithms. Some examples of contributions to innovation include the following: (1) It provides a solution to the issue that the data from the target domain does not have a label and domain adaptation. (2) It offers a variety of ideas for automatic classification, which makes it suitable for the continuously growing volume of data pertaining to sports videos. (3) A method for the classification of sports videos that is based on deep-neural-networks cum transfer learning is proposed to enhance the impactness of the segregation process. In comparison to the comparison approach, this method achieves superior results in terms of segregation accuracy, recall cum peak value. Additionally, it brings about a more favorable categorization effect for sports videos.

The remaining paper is organized as follows. Section 2 discusses about the recent literature survey, section 3 about materials and methods, section 4 about results and section 5 about conclusion.

2. Recent literature

A significant number of data mining and machine learning algorithms been well implemented many of domains, adding in classification, regression, cum clustering, among others. On the other hand, the majority of these algorithms are based on the assumption that the trained set cum the test set are simultaneously locates in similar feature space and adhere to the similar distribution. Reference [10] explains the training data that was obtained in the past could not be relevant anymore because of the passage of time and the modifications that have been made to the application situations. It is unfortunate with obsolete data are discarded, but in retrofit of scenarios resembles to these disparities [8]. This will ensure that the procedure is conducted correctly. As an illustration, each time we acquire knowledge from "nothing," in addition to each time we acquire knowledge, we will expend a significant amount of energy. In order to guarantee thus the aim model has a greater response, migration learning also make use of information that been obsolete. This decreases the amount of money that is spent on data gathering for new target activities. It was proposed by Moradzadeh and colleagues [5] that a mechanism known as cooperative distributed adaptation be used. This technique seeks to achieve a linear transformation in order to ensure the edge distribution of converted information is similar to the original as possible, and that the difference between conditional probability distribution and the original is low in possibility. The highest mean difference is still the method that is used to measure the degree to which the data distribution differs from one dataset to another. However, the approach that is said in [6] contains a hyper parameter, where the number in subspaces. This says the user is unable to determine the number of intermediate points that should be identified. An approach known as the geodesic stream-core method was suggested by [11] as a solution to the problem of selecting several intermediate sites.

It is demonstrated principle elements are utilized in minimize the dimensionality in video visual cum audio constraints in order to more accurately characterize video information. Additionally, time series in motion constraints are utilized in order to differentiate between motion context categories in football sports films. [6] conducted a study on the segregation in simple sports in sports films. They did

this by identifying certain motion modes in the video frames. These motion modes included fast moving, high jumps, translation, and closeness of the camera lens. Liu et al. demonstrate in some wisdom in the main domain can be achieved in training the main CNN on the Image Net dataset in advance. This results in an improvement of forty percent in classification accuracy when compared to the conventional training methods. In their investigation of transfer learning impact of various layer constraints in CNN, Zhang et al. demonstrate that the capacity of lower layer features to transfer learning is less robust than that of high layer constraints. This was discovered from the understanding of the transfer learning response. The researchers Wang et al. investigated the use in video cum audio constraints in conjunction with one another in order to classify videos. They also conducted an experimental investigation into the impact in principal component analysis - PCA in the reduction of constraint dimensions. However, they were unable to solve the problem of the correlation among constraints cum the fusion plan of 2 different models of constraint was slightly inadequate. Not sufficient training of neural networks with limited samples is a problem that can be solved by migration learning, which also significantly reduces the amount of money that is required to train the network. Within the scope of this study, the categorization of sports videos serves as the primary research topic, and convolutional neural networks are applied to a more general domain. The investigation in transfer learning, therefore one of the markable improvements that all now taking place. In order to improve the presentation in the classifier cum the accurate of segregation, researchers dedicated to working toward the discovery of improved video features or the utilization of multifeatured fusion approaches for the purpose of classifying movies.

3. Materials and Methods

In the first part of this section, we will present the dataset that was utilized and then proceed to describe the preparation methods. Following that, we will provide our recommended model, together with the specifics of the training and the implementation. In conclusion, we will next present a synopsis of the evaluation metrics that were utilized in our research. The steps involved in this process are shown in Fig. 1.

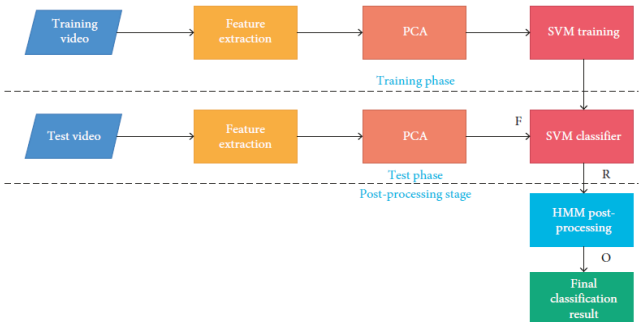

Fig. 1: Procedural steps in the video classification

3.1 DATA Set

Specifically designed for classification jobs in high-definition sports video content, the Deep Sports-VDS dataset has been painstakingly curated to meet specific requirements. It combines recordings obtained from high-definition sports broadcasts, high-speed cameras used during live matches, and selected parts from well-known datasets such as UCF101 and Sports-1M (as illustrated in Fig. 2 and Fig. 3). As a result of the key selection criteria being greater video resolution and clarity, this dataset is now in a position to serve as a benchmark for analyzing the influence of video quality on the categorization of low-resolution sports videos. Each film, which is comprised of over 200 movies spanning twenty different sports disciplines, has been meticulously annotated by seasoned sports specialists, ensuring that the label and content precisely coincide with one another. In order to facilitate the process of categorization, the clips are limited to a duration of fifty seconds, which results in a dataset size of around twenty gigabytes.

In order to train and analyze our model in an efficient manner, we have partitioned the Deep Sports-VDS dataset into three distinct sets: the training set, the validation set, and the testing set. To be more specific, eighty percent of the dataset has been designated as the training set. This set will be utilized to train our model and ensure that its parameters are optimized. Each of the validation and testing sets has been given ten percent of the data, and the remaining twenty percent has been distributed evenly between the two sets. During the training phase, the recognition set will utilizes to monitor performance of the design cum make any necessary improvements. On the other hand, the testing set will be utilized to evaluate the model's final performance on data that it has not before encountered.

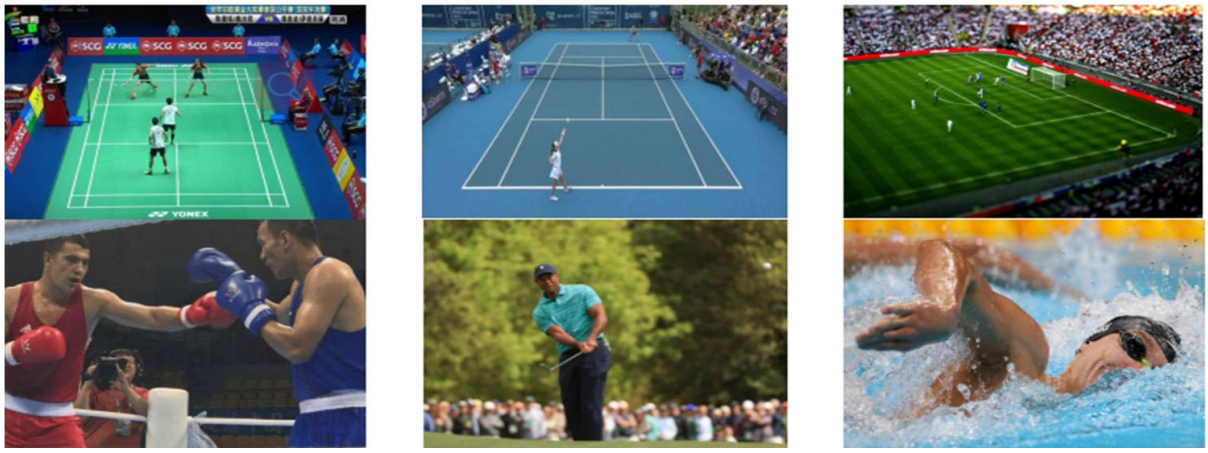

Fig. 2: Image frames from a partial video dataset.

Fig. 3: Images from self-collected sports video dataset (contains artifacts and low resolution).

For the purpose of ensuring that our model is well-generalized and capable of accurately classifying sports films from a wide variety of sources and conditions, our division is responsible. In order to evaluate the effectiveness of our proposed model across a wide range of sports video quality, we compiled a sports video dataset that was characterized by poor resolution and blurriness. This collection encompasses twenty of the most popular sports disciplines, and it was compiled from various web platforms and low-resolution cameras. It is a reflection of the issues that are inherent to original sports video processing in the real world. Our findings provide further evidence that the model is both resilient and capable of managing a wide range of video qualities. Each of the one hundred video clips that make up the dataset has been post-processed and normalized to approximately fifty seconds. With an average resolution that is close to 480 pixels per inch, file sizes range anywhere from 5 to 10 megabytes. Considering the diverse and authentic origins of the dataset, there is a significant variation in the quality of the video. This variation encompasses a variety of aspects, including resolution, frame rate, clarity, and noise, amongst others. Our dataset becomes more demanding and pragmatic as a result of these inequalities, which is a fitting reflection of the complexities involved in interpreting actual sports footage. During the process of collecting the dataset, we selected representative sports such as basketball, soccer, and table tennis to ensure that it would be diverse and applicable to a wide range of situations. When it came to labeling, we placed a high priority on

precision, utilizing professionals, and placing an emphasis on teamwork in order to achieve consistent outcomes.

3.2 Preprocessing

Because of the intrinsic nature of sports videos, they frequently contain motion artifacts as a result of the quick motions that are captured in them. Our research requires preprocessing steps to be completed before classification, which includes the removal of artifacts and noise. This is done in order to overcome this issue. To be more specific, frames are collected from the sports videos and then scaled to a consistent resolution of 520 pixels by 520 pixels. In order to identify high-frequency components in the images, we make use of the Laplacian second-order differential linear operator. It is possible that blurriness and motion distortions are present because there are not enough of these components. We determine that images that fall below this variance are likely to be blurry and filled with artifacts by computing the image variance using the Laplacian operator filter and setting a threshold of 0.3. Both of these methods are used to determine the image variance.

3.3. Model and Training

We apply center cropping to these frames in order to extract the first 80 frames from each video and reduce the possibility of noise caused by backgrounds that are not important to the subject of the video. The side length of the central square is determined by the shorter dimension of the image, which is then expanded to a size of 299 × 299 pixels through the use of the OpenCV library. Because of this, every single sports video input is represented as a tensor that is 80 × 299 × 299 × 3 in size. The process starts with the extraction of fine-grained features from the down sampled image, which ultimately results in the development of Feature 1. This is the initial branch of the process. Fig. 4 provides a full explanation of how this extraction begins. To begin, the image is partitioned into four blocks. Every single block is subjected to a customized convolution process that is 1 × 1, employing a total of 32 filters. Additionally, a ReLU activation function and a stride of 1 are utilized in conjunction with these filters in order to optimize the extraction process. Each block produces a feature map as its output, which is then concatenated along the spatial dimension with the other feature maps. In order to successfully restore the dimensions of the combined feature map, this concatenated map is subjected to an additional transformation. This transformation involves a 1 × 1 convolution with 128 filters, the same padding, and a stride of 1. The map that has been restored is next subjected to processing by means of two consecutive Conv2D layers. Each of these layers utilizes two convolution kernels that are 3 × 3 in size and have a stride of 1 and 32 filters.

The second branch begins with F0 going through the process of feature extraction using two Conv2D layers, which is then followed by a phase of down sampling in order to create feature map F2, as shown in Fig. 5. After then, the network pathway that is responsible for F2 divides into two branches: the main branch and the auxiliary branch. In the main branch, the F2 variable undergoes processing by means of a fine-grained feature layer, yielding the F2′ variable as a result. Following the concatenation of this new feature map, F2′, with F1 in the channel dimension, the latter is then subjected to additional processing using a Conv2D layer, which ultimately results in F2″. In the auxiliary branch, F2 first goes through a Conv2D layer, and then it goes through a down sampling layer. This happens simultaneously. Following the completion of the skip concatenation with F1, it is subjected to an additional level of fine-grained feature extraction. After that comes an up sampling layer that makes use of bilinear interpolation, which ultimately results in the feature F3 being produced. Last but not least, the final feature, F_final, is achieved by concatenating F3 with F2″. This sophisticated process, which involves the fusion of features from various convolutional depths, incorporates a more comprehensive collection of information that is both local and global in content.

3.4. Implementation Details

For the purpose of this work, our framework was constructed by utilizing the PyTorch library, and for computing, we made use of two NVIDIA GeForce RTX 3080 GPUs (NVIDIA, California, USA). The batch size was set to sixteen, and the Adam optimizer was utilized for the training process. A total of two hundred training epochs were set up in the configuration.

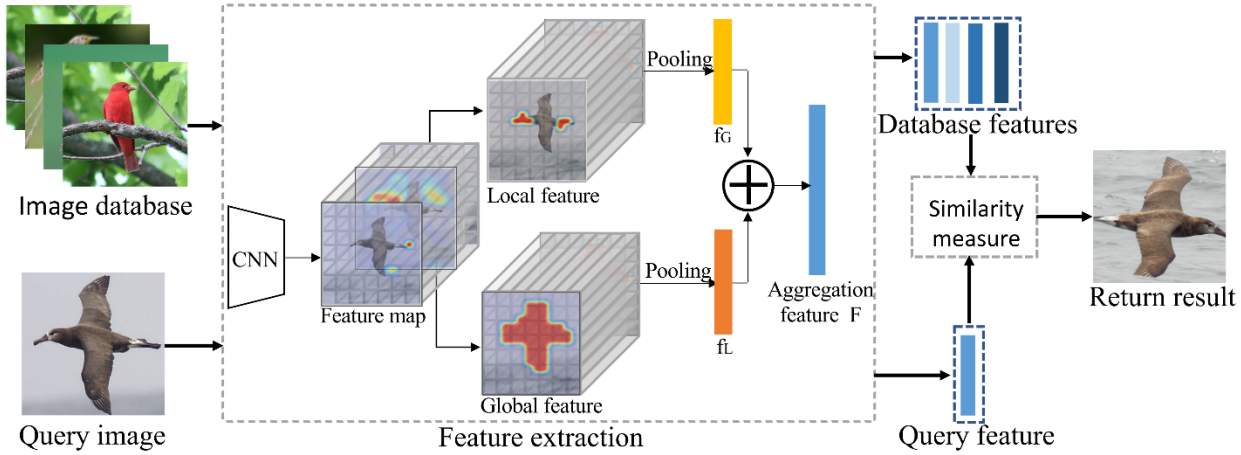

Fig. 4: The fine-grained feature extraction architecture

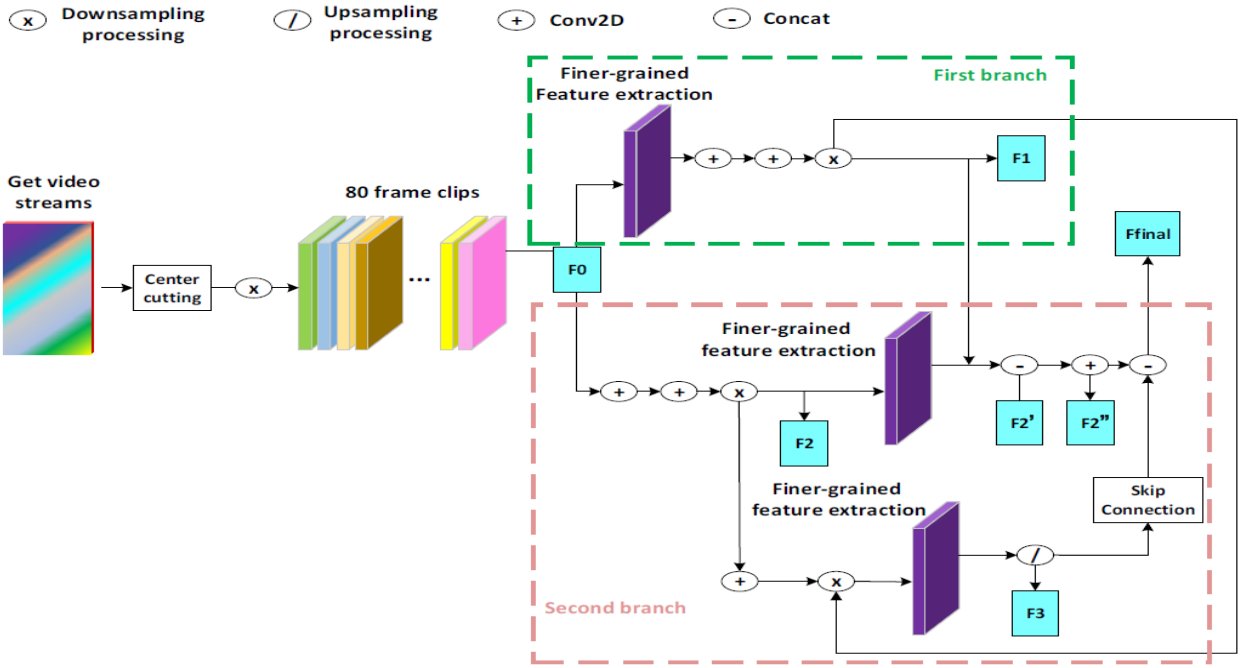

Fig. 5: Two–branch Neural network model architecture for sports video classification

The use of early halting was implemented in order to prevent overfitting. Additionally, we utilized the cosine annealing process in order to implement a dynamic learning rate modification. In order to facilitate effective navigation across the loss landscape of the model and to assist in the convergence to an ideal set of weights, the learning rate was adjusted to vary within the range of 0.001 to 0.00001, which gave rise to the aforementioned range.

4. Results

Fig. 6 illustrates several instances of video data classification that we offer in this part. These examples are based on the model that was proposed in our research report. In this study, we apply both high-resolution and low-resolution datasets to demonstrate the classification results for inputs, and then we compare these results to those obtained by other standard methods. Last but not least, in order to demonstrate the comprehensive generalization capabilities of our classification algorithms, we make use of a confusion matrix.

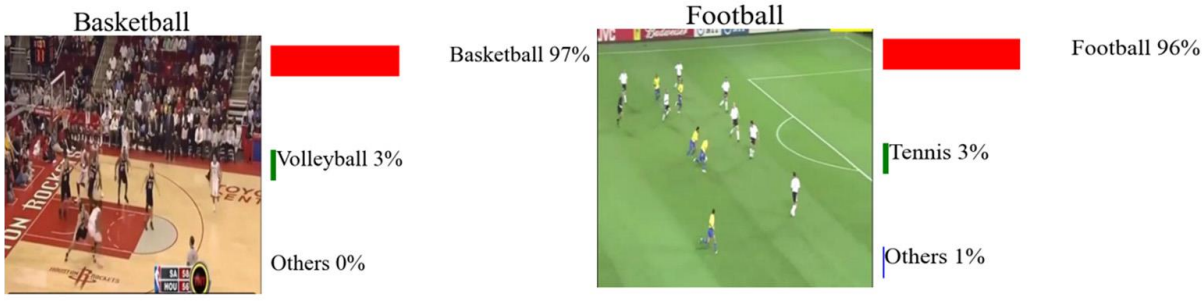

Fig. 6. Example of classification results of proposed algorithm.

Table. 1: Classification results of different classification algorithms on the high-resolution dataset.

Model	Precision	Accuracy	F-Score	Recall
Dense Net	0.96900	0.97300	0.96700	0.96510
VggNet-16	0.93380	0.96310	0.92380	0.91400
Inception v3	0.95310	0.97810	0.96540	0.96770
ResNet-50	0.93800	0.96450	0.95280	0.96810
Proposed method	0.97180	0.98040	0.97610	0.97230

The purpose of this investigation was to evaluate our suggested method in comparison to other well-known models that are included in the SPORTS1 m dataset as well as a bespoke dataset that contains sports films of varied resolutions. Using this information, we were able to demonstrate that the technique that we proposed is effective. The evaluation was initially carried out using the SPORTS1 m dataset at the beginning stages. Both Table. 1, include the results that correspond to the respective parameters. In light of the empirical findings that have been reported, it is clear that the method that we have proposed displays higher performance in comparison to other models when it is evaluated on a dataset that has a high resolution. To be more specific, the approach was able to attain a recall rate of 0.9723, an F-score of 0.9761, an accuracy rate of 0.9804, and a precision of 0.9718.

5. Conclusion

The purpose of this research is to offer a revolutionary deep learning framework that is specifically built for the classification of sports videos. We make use of approaches such as key frame selection, fuzzy noise reduction algorithm, and dual-branch neural network in order to overcome the constraints of traditional methods when it comes to dealing with videos that have a poor resolution and a fast-paced nature. The results of our experiments show that our method performs exceptionally well on two distinct sports video datasets, exceeding existing methodologies that are considered to be the industry standard. This indicates that our method is capable of handling sports video classification jobs under a variety of settings with a high degree of accuracy and durability. In addition, the findings of this research will be put to use in the field of sports training and video analysis.

Conflict of Interest

The authors declared "No conflict of interest"

References

[1] R. Duan, T. Kawahara, M. Dantsuji, and J. Zhang, "Articulatory modeling for pronunciation error detection without non-native training data based on DNN transfer learning," *IEICE - Transactions on Info and Systems*, Vol. E100.D, No. 9, pp. 2174–2182, 2017. https://doi.org/10.1587/transinf.2017EDP7019

[2] A. H. Vo, L. Hoang Son, M. T. Vo, and T. Le, "A novel framework for trash classification using deep transfer learning," *IEEE Access*, Vol. 7, pp. 178631–178639, 2019. https://doi.org/10.1109/ACCESS.2019.2959033

[3] Z. Kastrati, A. S. Imran, and A. Kurti, "Integrating word embeddings and document

topics with deep learning in a video classification framework," *Pattern Recognition Letters*, Vol. 128, No. Dec, pp. 85–92, 2019. https://doi.org/10.1016/j.patrec.2019.08.019

[4] Z. Kang, B. Yang, Z. Li, and P. Wang, "OTLAMC: an online transfer learning algorithm for multi-class classification," *Knowledge-Based Systems*, Vol. 176, No.15, pp. 133–146, 2019. https://doi.org/10.1016/j.knosys.2019.03.024

[5] N. Dif, M. O. Attaoui, Z. Elberrichi, M. Lebbah, and H. Azzag, "Transfer learning from synthetic labels for histopathological images classification," *Applied Intelligence*, Vol. 52, No. 2, pp. 1–20, 2022. https://doi.org/10.1007/s10489-021-02425-z

[6] F. Zhang and J. Yan, "Cloud image classification method based on deep convolutional neural network," *Xibei Gongye Daxue Xuebao/Journal of Northwestern Polytechnical University*, Vol. 38, No. 4, pp. 740–746, 2020. https://doi.org/10.1051/jnwpu/20203840740

[7] Z. Ma, H. Yu, W. Chen, and J. Guo, "Short utterance based speech language identification in intelligent vehicles with time-scale modifications and deep bottleneck features," *IEEE Transactions on Vehicular Technology*, Vol. 68, No. 1, pp. 121–128, 2019. https://doi.org/10.1109/TVT.2018.2879361

[8] A. Moradzadeh and N. R. Aluru, "Transfer-learning-based coarse-graining method for simple fluids: toward deep inverse liquid-state theory," *Journal of Physical Chemistry Letters*, Vol. 10, No. 6, pp. 1242–1250, 2019. http://dx.doi.org/10.1021/acs.jpclett.8b03872

[9] R. K. Samala, H. P. Chan, L. Hadjiiski, M. A. Helvie, C. D. Richter, and K. H. Cha, "Breast cancer diagnosis in digital breast tomosynthesis: effects of training sample size on multi-stage transfer learning using deep neural nets," *IEEE Transactions on Medical Imaging*, Vol. 38, No. 3, pp. 686–696, 2019. https://doi.org/10.1109/TMI.2018.2870343

[10] H Mahmoud, A Alharbi, and D Khafga, "Breast cancer classification using deep convolution neural network with transfer learning," *Intelligent Automation & Soft Computing*, Vol. 29, No. 3, pp. 803–814, 2021. http://dx.doi.org/10.32604/iasc.2021.018607

[11] Sarma, M. Sen, K. Deb, P. K. Dhar, and T. Koshiba "Traditional Bangladeshi sports video classification using deep learning method", *Applied Sciences*, Vol. 11, No. 5, art.no. 2149, 2021. https://doi.org/10.3390/app11052149

[12] L. Wenming "Deep Learning Based Sports Video Classification Research", *Applied Mathematics and Nonlinear Sciences*, 2021. https://doi.org/10.2478/amns.2023.2.00029

***