Research Article

Analysis of Sports Video Intelligent Classification Technology Based on Neural Network Algorithm and Transfer Learning

Han Guangyu [ID]

Physical Education Institute, Xinxiang Medical University, Xinxiang 453003, Henan, China

Correspondence should be addressed to Han Guangyu; 031013@xxmu.edu.cn

With the rapid development of information technology, digital content shows an explosive growth trend. Sports video classification is of great significance for digital content archiving in the server. Therefore, the accurate classification of sports video categories is realized by using deep neural network algorithm (DNN), convolutional neural network (CNN), and transfer learning. Block brightness comparison coding (BICC) and block color histogram are proposed, which reflect the brightness relationship between different regions in video and the color information in the region. The maximum mean difference (MMD) algorithm is adopted to achieve the purpose of transfer learning. On the basis of obtaining the features of sports video images, the sports video image classification method based on deep learning coding model is adopted to realize sports video classification. The results show that, for different types of sports videos, the overall classification effect of this method is obviously better than other current sports video classification methods, which greatly improves the classification effect of sports videos.

1. Introduction

As a branch of artificial intelligence, more and more different methods are emerging with the increasing development of theory, and they play an important role. Deep learning is one of the more popular areas. The development of DNN is driven by the continuous breakthroughs and innovations of deep learning. DNN is the most commonly used architecture in deep learning [1]. The concept of deep learning was first proposed by Hilton et al. in 2006. In recent years, it has shined in various fields, making it a hot topic in the AI field [2]. The main idea of deep learning is to simulate the neural structure of the human brain. It is hoped that the machine can process the received data like a human. The realization process is to extract features from the data through multiple nonlinear stages [3]. However, when we actually want to solve a problem by using machine learning algorithm, whether we use shallow machine learning algorithm or deep learning algorithm, we are faced with some practical conditions, such as the lack of data and information and the difference between training data and practical application data, resulting in poor model construction and poor practical application effect [4]. This is because traditional machine learning requires that the data of learning and the data of actual application scenarios should meet the same statistical characteristics, and it means that whenever there is a new application scenario, it is necessary to collect enough annotation information to make the traditional machine learning methods play an effective role, which will consume a lot of manpower, material resources, and time cost [5]. With the rapid development of machine learning in recent years, people pay more and more attention to this problem, and transfer learning, one of the learning methods to solve this problem, has become a very active research field [6].

The goal of transfer learning is to use existing knowledge to deal with problems in different but related fields, so as to realize the transfer of knowledge between related fields. It is an intelligent learning method similar to human basic learning ability. Just like in real life, we humans can judge whether the new problems we face are related to the previous accumulation and whether we can adopt the knowledge learned before through the knowledge we have learned or accumulated before [7]. Psychologists have found that, in

order to realize transfer learning, it is certainly necessary for the two fields or learning materials to have common factors, and the transfer effect is proportional to the common factors, and the more the common factors, the better the transfer effect. Of course, there are common factors and different factors, which may also cause negative transfer and affect learning [8]. Sports video is a very important resource in video. Sports programs have hundreds of millions of loyal viewers all over the world. The classification of sports video has become the focus of many researchers. Sports video intelligent classification technology can not only intelligently classify and sort out massive sports video data and reduce people's workload, but also provide people with better spiritual enjoyment in daily life. It is also the basis of intelligent radio and television intelligent classification. Therefore, sports video intelligent classification technology can be widely used in sports video management, retrieval, query, and other fields and has broad development prospects and great value [9]. At present, the traditional transfer learning method is mainly to find a common feature space and map the data in two domains into the feature space by mapping function, so as to make the data distribution difference between the two domains as small as possible while maintaining the original characteristics of the data as much as possible and then train the classifier on these mapped data [10]. Existing sports video classification technologies pay more attention to the extraction of video features. It can be seen that feature extraction is the key to sports video classification. Researchers have proposed different feature models for classification in terms of motion, color, edge, etc., but different features. The role played in sports videos is different. Some features have a good classification effect. Obviously, the features extracted by these methods are not more expressive than those extracted by deep learning, and the deep transfer learning technology can better meet the end-to-end needs in practical applications. [11].

Video features mainly focus on motion, color, brightness, edge, texture, audio, etc. These features are mainly used in the classifier in a separate way, or in a simple linear fusion (horizontal); that is, different feature vectors are used in a linear combination according to the law. This fusion method improves the classification performance to a certain extent. However, most of them separate different features and ignore the semantic relationship between features [12]. In order to establish more connection and fusion between features, this paper proposes an intelligent sports video classification technology based on DNN and transfer learning and makes an in-depth study on the field of sports video classification. The third paragraph introduces the video region feature extraction and processing and migration learning video classification system architecture. Block brightness comparison coding and block color histogram are proposed, which reflect the brightness relationship between different regions in the video and the color information in the region. The fourth paragraph introduces the transfer learning of DNN. The algorithm of maximum mean difference (MMD) is proposed for migration learning. The fifth paragraph introduces the classification and experimental

results of migration learning sports video images based on deep learning coding model. On the basis of obtaining the features of sports video images, the sports video image classification method based on deep learning coding model is adopted to realize sports video classification. In order to analyze the effectiveness of the method classification in the article, the sports videos are set to be figure skating, badminton, and yoga in order. Due to the supervised fine-tuning, the parameters of each layer of the deep learning network can be adjusted in the form of error back-propagation to optimize the classification effect.

Through the analysis of traditional transfer learning algorithm and DNN transfer learning algorithm, this paper studies a video classification transfer learning algorithm based on DNN. Innovation contributions include the following: (1) It solves the problem that the target domain data has no label and domain adaptation. (2) It provides different suggestions for automatic classification for the rapidly increasing amount of sports video data. (3) To improve the effect of sports video classification, a sports video classification method based on DNN and transfer learning is proposed. The classification accuracy, recall, and maximum value of this method are better than the comparison method, and a better classification effect of sports video is achieved.

2. Related Work

Many data mining and machine learning algorithms have achieved great success in many fields (such as classification, regression, and clustering). However, most of these algorithms have a common assumption: the training set and the test set are in the same feature space and obey the same distribution. Zhang et al. [13] point out that, with the passage of time and the adjustment of application scenarios, the training data collected before may be out of date. It is a pity to discard these out-of-date data, but the adjustment of scenarios leads to differences in statistical characteristics between training data and application data, resulting in poor model effect. Ma et al. [14] indicate that traditional machine learning does not have the ability to transfer learning. Whenever a new application scenario is encountered, the process of data collection, training, and verification will be carried out again. For example, every time we learn something from "nothing," then every time we learn, we will spend a lot of energy. Migration learning can make full use of outdated data to ensure that the target model has better effect, thus reducing the cost of data collection in new target tasks. Moradzadeh et al. [15] proposed a method called joint distributed adaptation. They assume that the marginal distributions of the source and target domains are different from the conditional probability distributions, and the target domain has no data annotations. This method looks for a linear transformation, so that the edge distribution of the transformed data can be as close as possible, and the conditional probability distribution difference is also as small as possible. The method to measure the degree of the difference in the data distribution is still the maximum mean difference. Compared with TCA, this method requires iteration and the source domain needs to be labeled. However, the

method proposed in [16] has a hyperparameter, that is, the number of subspaces, which means that the user cannot know how many intermediate points should be found. Mahmoud et al.[17] proposed a geodesic stream-core method to solve the problem of choosing several intermediate points.

Qiu et al. [18] show that principal components are used to reduce the dimensionality of video visual and audio features to describe video content, and time series of motion features are used to distinguish motion event classifications in football sports videos. Chen et al. [19] studied the classification of simple sports in sports videos by detecting some motion modes in video frames, such as running, jumping, translation, and zooming of the serve lens. Liu et al. [20] show that some knowledge in the source domain is obtained by training the source convolution neural network on the ImageNet dataset beforehand, and then the network is migrated and studied on the image classification datasets Calteech-101 and Catech-256, respectively, which improves the classification accuracy by 40% compared with traditional training methods. Zhang et al.[21] show that, by studying the transfer learning effect of different layer features of convolutional neural networks, it is found that the transfer learning ability of lower layer features is weaker than that of higher layer features. Wang et al. [22] studied the combination of video and audio features to classify videos and experimentally studied the effect of principal component analysis (PCA) on feature dimension reduction but failed to solve the correlation between features, and the fusion scheme of two types of features was slightly insufficient. Migration learning solves the problem of insufficient training of neural networks with small samples, while greatly reducing the training cost of the network. This paper takes sports video classification as the focus of research and applies convolutional neural networks to a broader field. Therefore, the research of transfer learning is one of the current development trends. Researchers are committed to finding better video features or using multifeature fusion methods to classify videos, while improving the performance of the classifier and the accuracy of classification.

3. Video Classification

3.1. Extraction and Processing of Video Area Features.
Video area features are divided into (BICC) and block color histograms, which reflect the brightness relationship between different areas in the video and the color information within the area. The algorithm framework is shown in Figure 1.

For each type of video, the brightness distribution of the frame is relatively uniform and there is a certain brightness difference between the target in the frame and its surrounding environment. The human visual system distinguishes the object by the brightness difference between the target in the frame and its surrounding environment, and BICC is obtained by comparing the average brightness between blocks [23]. BICC reflects the brightness relationship between intraframe blocks and has strong anti-interference ability for

different video segments of the same kind of video due to factors such as brightness and darkness.

Assuming that the frame size is $M \times N$, each frame is divided into log×seven blocks, and the size of each block is $h \times u$, where $h = M/k$, $u = N/k$, and x_i represents the i-th pixel in the block The brightness value of each block is $\overline{X(l)}, l \in [1, k \times k]$; then

$$\overline{X(l)} = \sum_{i=1}^{h \times u} \frac{x_i}{(h \times u)}. \tag{1}$$

If the block brightness comparison coded value is expressed as y, the result of the brightness comparison between the m-th block and the nth block in a frame can be expressed as (2), where $1 < m < K \times K$, $2 < n < K \times K-1$.

$$y\left[(m-1) \times (k \times k) + n + \frac{m(m+1)}{2}\right] = \begin{cases} 1 & \text{if } \overline{X(m)} > \overline{X(n)}, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

According to formula (2), the frame can be coded by 0 and 1 according to the block brightness mean value comparison.

For some videos, there is no obvious difference in brightness, but there is a significant difference in color distribution, such as volleyball and table tennis videos, which can be used for video classification by dividing video frames into blocks and counting their color histograms. The extraction process of block color histogram is as follows.

Assuming that the frame size is $M \times N$, the frame is divided into $k \times k$ blocks, and the size of each block is $h \times u$, where $h = M/k$, $u = N/k$, and x_{ir}, x_{ig}, and x_{ib}, respectively, represent the i-th block in the block. The value of the R, G, and B components corresponds to the pixel point, P represents the number of straight squares, and $\text{HIST}_{m,n,p}$ represents the color histogram of the nth block in the frame, where $m \in \{r, g, b\}, n \in [1, k \times k], p \in [1, P]$, p_m represent the quantization range of the color component m corresponding to the p square, but

$$\text{HIST}_{m,n,p} = \sum_{i=1}^{h \times u} \left(x_{i,m,n} \in p_m\right). \tag{3}$$

Features are extracted based on video frame blocks. When there are more block partitions and more histograms in the BlockHist histogram, the dimension of the extracted feature vector is higher. Therefore, PCA is used to reduce the dimension of the extracted features to speed up the processing.

3.2. Migration Learning Video Classification System Architecture.
Convolutional neural network (CNN) is a neural network architecture for image classification, which usually includes convolution layer and pooling layer. The loss function and accuracy are reported on the test set of each CNN architecture. Loss function: how far is the difference between the prediction and the actual results? The greater the predicted value, the worse the accuracy of the
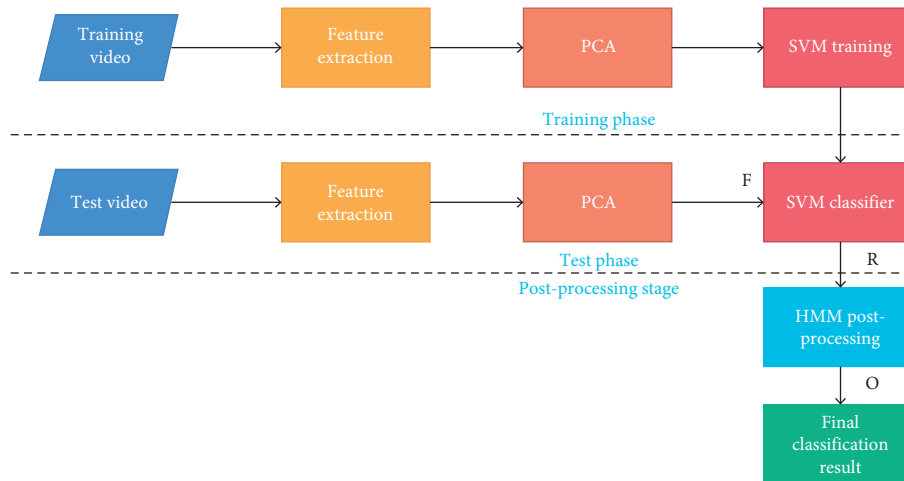
FIGURE 1: Video classification frame diagram.

model fitting data points. Accuracy of test set is the accuracy of the model's prediction of test set data.

With the development of artificial intelligence technology, video classification technology is needed in various fields, such as face recognition, video location, and so on. On the one hand, with the advent of the Internet age, although the available target domain data that can fully meet the target tasks cannot be collected, enough original digital videos that are similar or related can be collected, and it is easier to train a source convolutional neural network that can better complete the video classification tasks [24]. In addition, DNN has gradually become a research hotspot in recent years, and the convolutional neural network, as a star in the field of video classification, has also made progress that cannot be ignored. Compared with other traditional video classification methods, it has a better classification accuracy. Therefore, most of the current fields are keen to complete video classification tasks through deep convolutional neural networks from many aspects such as processing speed, classification accuracy, and user experience. According to the functions of video classification related systems, the realized video classification system can be divided into three parts, namely, network training, network migration, and network testing [25]. The overall functional framework of the system is shown in Figure 2.

Network migration module is the core part of the whole system. For convolution models that are already available, the network training module can be almost unused, loaded directly with existing parameters, and then used. The migration module also has superparameter settings, which include not only the learning rate and iteration rate, but also the ratio of convolution layer and full-connection layer parameter changes and the K selection function. Task mapping is the mapping of the source and target tasks when changing the top-level classifier structure of the source domain model according to the target domain tasks, which makes the modification of the network more reasonable. The process of retraining the new network by the fine-tuning function is similar to that of the network training model, except that K selection algorithm is added in the process, and
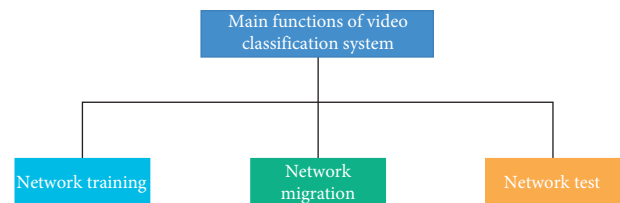

FIGURE 2: The overall framework of the system.

only some parameters are adjusted. The network test module has an evaluation method for the constructed convolutional neural network, so as to know whether the model meets the task research and how to modify the network.

The main workflow of migration learning video classification system is the training, migration, and testing process of convolutional neural network, in which whether the training process is carried out depends on whether there are trained available model parameters. The following describes the workflow of the system in the form of a flowchart, as shown in Figure 3.

4. DNN Transfer Learning

Fully connected neural network model (DNN) is also called deep neural network in some cases. Different from traditional perceptron, each node has an operation relationship with all nodes in the next layer, which is the meaning of "fully connected" in the name. The middle layer in the above figure also becomes a hidden layer. Fully connected neural network usually has multiple hidden layers. Adding a hidden layer can better separate data. However, too many hidden layers will also increase the training time and produce overfitting.

On the other hand, since the last feature of a neural network, such as the last layer of a convolution layer, must be related to a particular dataset, assuming that the last layer of the network is the softmax layer and that the entire network has been trained on a dataset, the output of the softmax layer must be closely related to that dataset, so make this feature special [26].
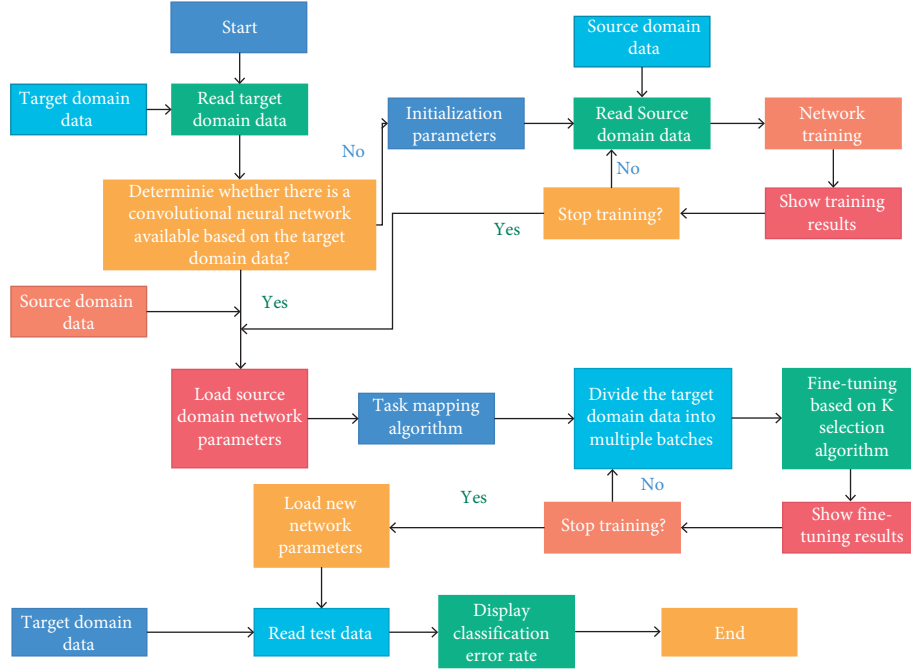
FIGURE 3: Work flowchart of video classification system.

The first two modules of VGG13 network can be fixed (not counting pooling layer, but the first four layers) because the first few layers of DNN trained on image data set are all common features. These network layers have been trained on ImageNet data set and can be used to extract common features on source domain and target domain [27]. The domain adaptation problem is difficult to solve because the target domain data has no label. In order to solve this problem, many methods try to limit the error value of the target domain to the error of the source domain plus a difference value. MMD is one of the methods, and its formula is

$$\text{MMD}[F, p, q] = \left[\sup_{f \in F}\left(E_p[f(x)] - tE_p n[f(y)]\right)\right]. \quad (4)$$

Among them, F is the space where the function f is located, the distribution of x is p, and the distribution of y is q. Equation (4) means the upper bound of the difference between the expected value $E_p[f(x)]$ and $E_q[f(y)]$ after random projection of the data x on the source domain and the data y on the target domain through the function f. Obviously, when $p = q$, that is, when the distributions of x and y are the same, no matter how the function f is projected, the expectation will be the same, so in the case of $p = q$, MMD is 0. However, if p_q, and the function space F is rich enough, the MMD will never be zero. Its value is determined by the function in function space F which makes the distribution difference between the X and Y projections the greatest. It is important to note that if the function space F is too rich and f has too many possibilities to take values, MMD can easily be infinite, so the function space F needs to be constrained. F is proved to be the best when F is the unit ball in the reproducing kernel Hilbert space (i.e., space less than or

equal to 1 from the origin). At this point, the formula for MMD^2 becomes

$$
\begin{aligned}
\text{MMD}^2[F, p, q] &= \left[\sup_{\|f\|_H \leq 1}\left(E_x[f(x)]\right) - E_Y[f(y)]\right], \\
&= \left[\sup_{\|f\|_H \leq 1} \langle \mu_p - \mu_q, f \rangle_H\right]^2, \quad (5) \\
&= \left\|\mu_p - \mu_q\right\|_H^2.
\end{aligned}
$$

An important property of (5) is that $p \neq q$ if and only if $\text{MMD}^2 = 0$. Because a single kernel function may not be optimal, if a single kernel is used, it can be Gaussian kernel or linear kernel, but the disadvantage is obvious: it is impossible to know which kernel is good. Therefore, multikernel maximum mean error (MK-MMD) is used for feature mapping, and the kernel is weighted by m different Gaussian kernels, and the weight is β_u, and the formula is

$$K = \left\{ k = \sum_{u=1}^{m} \beta_u k_u : \beta_u \geq 0, \forall u \right\}. \quad (6)$$

The restriction on the coefficient β_u ensures that the obtained k is unique, so that the measurement method for the difference between the source domain and the target domain is determined to be MK-MMD.

In general convolution networks, the characteristics will eventually change from general to special before the last layer, and this transformation will become larger and larger as the number of layers of the network goes deeper to the full-connection layer. In other words, deep networks are almost specialized by specific tasks in the full-connection layer. If there is not enough data in the target domain for

supervised learning, the full-connection layer cannot be trained directly in the target domain. Since the source domain is labeled, train the entire network on the source domain to improve its accuracy on the source domain. At the same time, use MK-MMD to make the data on the source domain be on the output of the fully connected layer and the target domain. The output difference of the data on the fully connected layer is as small as possible, so as to reduce the distribution difference between the data of the source domain and the data of the target domain after feature mapping, so as to achieve the purpose of migration learning. The experimental results are shown in Figures 4 and 5.

As shown in Figure 6, it can be seen that, in different stages of training, the accuracy of models with different fine-tuning depths is ranked consistent.

After carefully observing and analyzing the fine-tuning results at different depths, we find that the highest accuracy will first increase with the deepening of fine-tuning layers, then reach the highest at a certain depth, and then gradually decrease, showing a single peak curve. As shown in Figures 7 and 8, it can be observed that each curve has only one peak. For example, the peak of the GoogLeNet model of SUN397 is 4inceptions, and the peak of the GoogLeNet model of FOOD101 is 8inceptions. The peak of the AlexNet model of FoOD101 is 3 FCs+3convs. This shows that when starting the comparison from the smallest depth, once the highest accuracy rate of the next larger depth is not higher than the previous one, then the previous smaller depth is the best fine-tuning training depth. This kind of accuracy does not fluctuate with the change of the fine-tuning depth and is very suitable for iteratively finding the optimal solution without falling into the local optimum.

So you can think of an iterative approach to finding the best fine-tuned training depth. However, this method does not utilize the model weights that have been previously trained for each additional depth but instead retrains a model for a new fine-tuning depth.

According to the experimental results in Figures 9 and 10, using Places365_GoogLeNetfeature on SUN397 data set has better fitting effect on training set than ImageNet_GoogLeNetfeature and has higher accuracy on test set. Places365 is a scene dataset with 1.8 million pictures, SUN397 is also a scene dataset, and ImageNet is a large image dataset related to objects. Therefore, the more similar the new data is to the data used in the pretraining model, the higher the accuracy of the model after transfer learning.

5. Sports Video Image Classification and Experimental Results

5.1. Transfer Learning Sports Video Image Classification Based on Deep Learning Coding Model. On the basis of acquiring the characteristics of the sports video image's key chain frame, the sports video image classification method based on the deep learning coding model is adopted to realize the sports video classification. The specific detailed process is as follows:
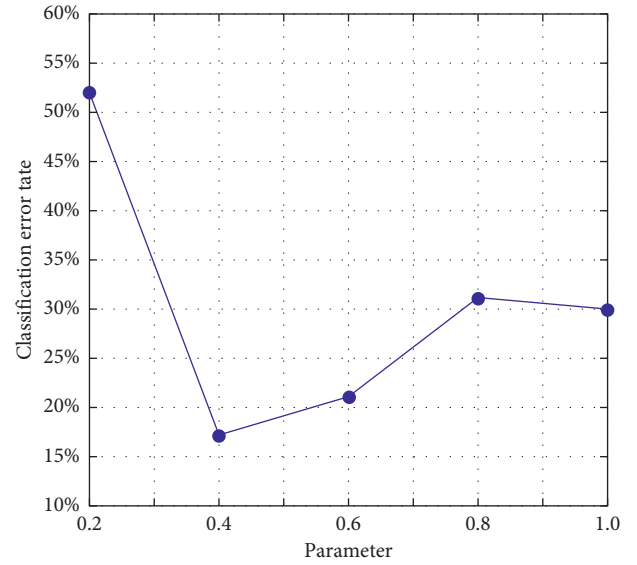

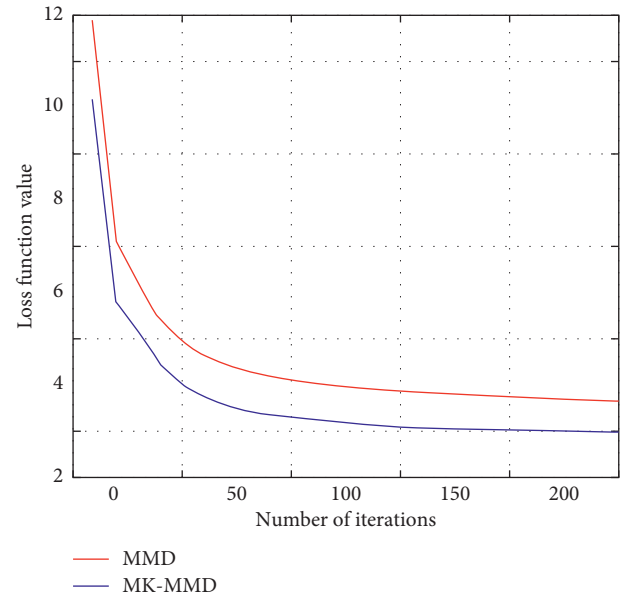FIGURE 4: The impact of K selection on migration results.

FIGURE 5: Two transfer learning loss functions.

(1) Set the feature library of sports video images to be linked to $G = \{g_1, g_2, \cdots, g_m\}$, and M represents the number of features

(2) Use unsupervised restricted Boltzmann machine (RBM) to encode G; use CD algorithm to obtain convergence parameter codebook and become a complete dictionary

(3) Use the kind label information of training data to spread the errors of the dictionary acquired by learning nonforward, supervise the learning of RBM network, and then optimize the acquisition of G-code to get the optimized visual dictionary and description coefficient
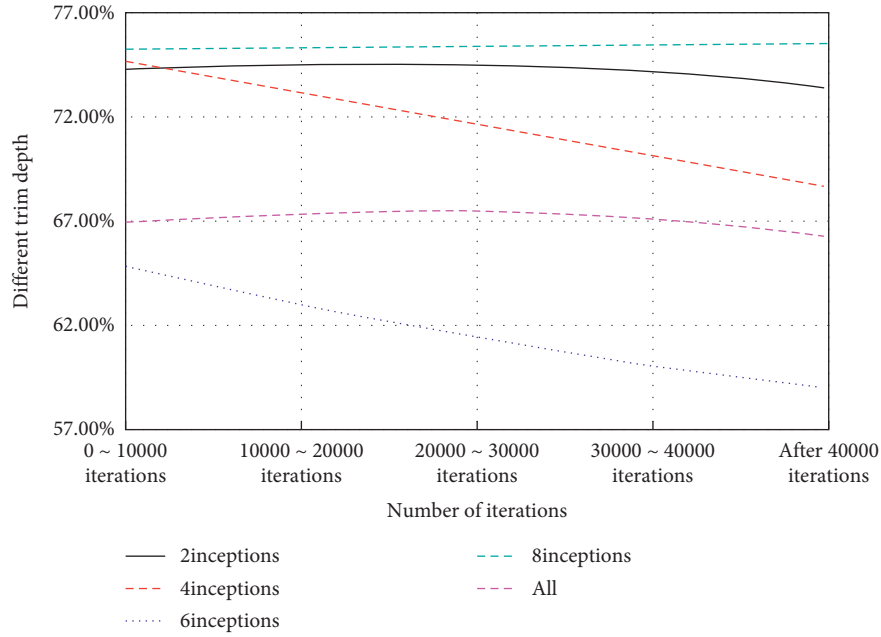
FIGURE 6: The accuracy of the FOOD101 GoogLeNet model with different fine-tuning depths in different training stages.
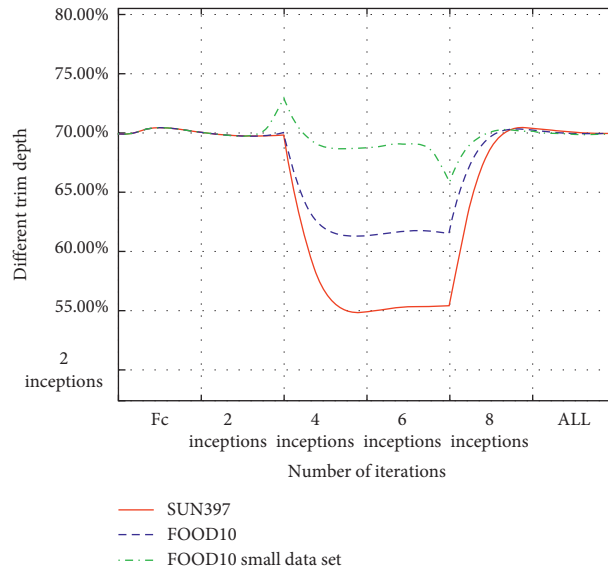

FIGURE 7: The accuracy of the three data sets under different fine-tuning depths of GoogLeNet.

(4) Rub the support vector machine classifier with the deep learning coding vector of the training sports video image to realize sports video classification

G is coded by using multilayer restricted Boltzmann machine and becomes an exemplary visual dictionary. According to the spatial information of G, the neighboring G features are set as the input of RBM, and the RBM is trained by CD fast algorithm to obtain the hidden stack features. Then, the adjacent hidden layer features are regarded as the input of the lower RBM, and the output dictionary is obtained. Among them are ϖ_1 and ϖ_2. Belonging to the connection weight of RBM, RBM has a visible layer and a

hidden layer, and neurons based on the same level in RBM do not have a connection relationship. During network training, the hidden layer and the explicit layer of the RBM are connected according to the conditional probability distribution. The conditional probability of the explicit layer and the hidden layer is

$$q(s_i|y) = \text{sigmoid}\left(c_i + \sum_{j=1}^{i} \varpi_{ji} y_i\right),$$

$$q(y_i|s) = \text{sigmoid}\left(b_i + \sum_{j=1}^{i} \varpi_{ji} s_i\right). \tag{7}$$
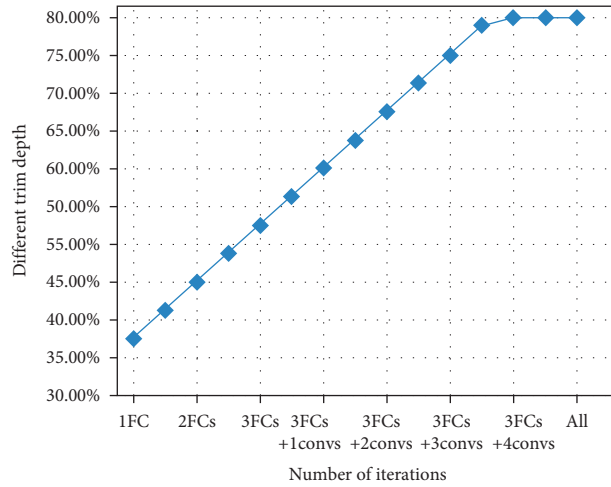
FIGURE 8: The accuracy of FOOD101 under different fine-tuning depths of Alex Net.

——— Sun397_googlenet_imagenet_finetuning_fc_1r0001

- - - Sun397_googlenet_places365_finetuning_fc_1r0001
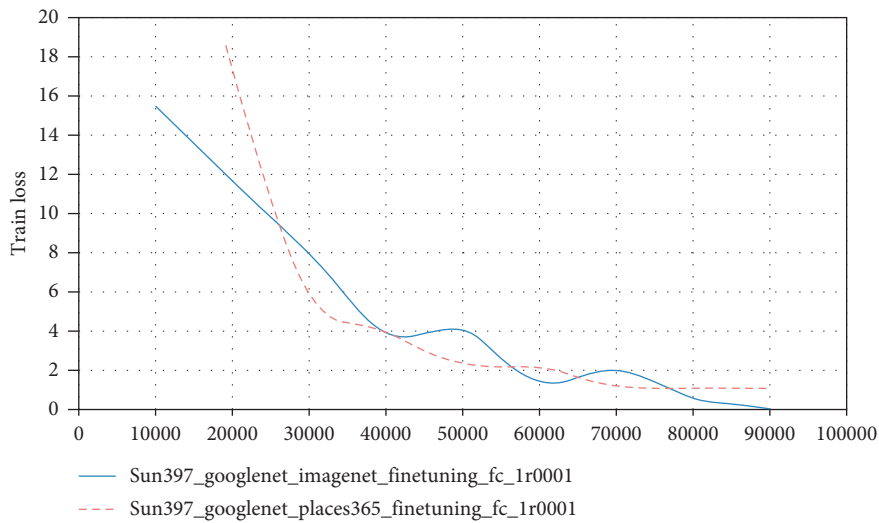
FIGURE 9: Train loss change curve during training.

——— Sun397_googlenet_imagenet_finetuning_fc_1r0001

- - - Sun397_googlenet_places365_finetuning_fc_1r0001
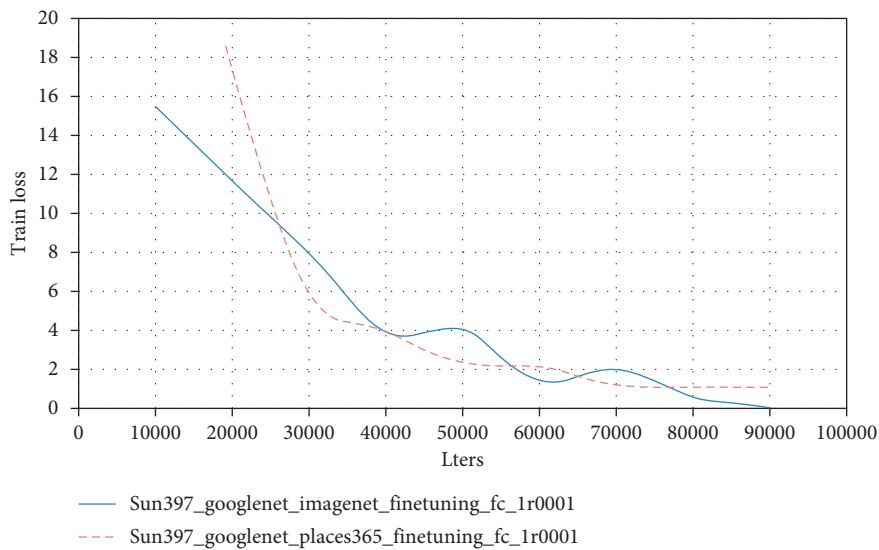
FIGURE 10: Test accuracy change curve during training.

TABLE 1: The key frame miss rate of sports video images of the method in the article.

Number of extractions	Figure skating	Badminton	Yoga
1	0.03	0.03	0.02
2	0.03	0.03	0.02
3	0.02	0.03	0.02
4	0.02	0.03	0.03
5	0.02	0.04	0.03
6	0.03	0.04	0.03
7	0.02	0.03	0.04
Mean value	0.02	0.03	0.03

In the formula, y_i and s_i describe the feature layer and encoding layer of sports video in turn, that is, the explicit layer and the hidden layer in RBM.

Using CD algorithm, RBM is quickly learned to improve the convergence efficiency of parameters, and the amount of update to obtain weights is ϖ_{ji}:

$$\Delta\varpi_{ji} = \phi\left(\langle y_i s_i \rangle_{\text{data}} - \langle y_i s_i \rangle_{\text{model}}\right). \tag{8}$$

Among them, ϕ describes the learning speed. Using CD algorithm can get the latest parameters of sports video features until the parameters converge and get the initial visual dictionary.

5.2. Experimental Results. In order to analyze the effectiveness of the method classification in the article, the sports videos are set to be figure skating, badminton, and yoga in order.

When the method in the test article recognizes the three types of sports videos of figure skating, badminton, and yoga, the comprehensiveness of the key frames of the video images is extracted, and the missed recognition rate is the evaluation index. The results are shown in Table 1.

The data in Table 1 shows that when the method is used to identify the three types of sports videos: figure skating, badminton, and yoga, the average miss rate of extracting video image keyframes is less than 0.03, which indicates that the method in this paper can extract the key frame features from sports video images in an all-round way. Three sports videos were imported into the Caltech image collection to test the interference of different visual point sizes on the methods in this paper. With the increasing scale of sports video visual dictionary, the classification accuracy of the proposed method also increases. When the scale of sports video visual dictionary reaches 2000 MB, the classification accuracy of the proposed method is stable at 0.98. In order to test the influence of supervised fine-tuning on the classification effect of the proposed method, there are fine-tuning and no fine-tuning in the test. There is a significant difference between supervised fine-tuning and unsupervised fine-tuning. Supervised fine-tuning can improve the classification accuracy of methods in the text. This is because supervised fine-tuning can use the form of error backpropagation to adjust the parameters of each layer of the deep learning network to optimize the classification effect.

6. Conclusions

By analyzing the traditional transfer learning algorithm and DNN transfer learning algorithm, this paper studies a DNN based video classification transfer learning algorithm and designs it as a video classification system based on convolutional neural network. The first two modules of vgg13 network can be fixed because the first several layers of DNN trained on the image data set are common features. These network layers have been trained on ImageNet datasets and can be used to extract common features on source and target domains. Because the target domain data has no label, the domain adaptation problem is difficult to solve. How to automatically classify these sports videos has become a hotspot and difficulty in current research. In order to improve the effect of sports video classification, a sports video classification method based on DNN and transfer learning is proposed. In the classification process of this method, supervised fine-tuning is introduced to improve the classification accuracy of this method. The experimental results show that, for figure skating, badminton, and yoga sports videos, the classification accuracy, recall, and maximum value of this method are better than the comparison method, and a better sports video classification effect is achieved.

Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The author declares that there are no conflicts of interest regarding this work.

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.

[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.

[3] Y. Shen, Z. Huang, S. Jiang, and B. Hu, "A combined recommendation algorithm of MF and DNN with transfer learning," *Journal of Air Force Early Warning Academy*, vol. 33, no. 02, pp. 62–67, 2019.

[4] H. Tong, J. Chu, and J. Shen, "Image classification technology based on convolutional neural network," *Science and Technology Vision*, vol. 000, no. 032, pp. 36-37, 2017.

[5] Y. Yang, J. Sun, and S. Yu, "Aircraft target recognition based on transfer learning of convolutional neural network," *Modern Radar*, vol. 41, no. 12, pp. 35–39, 2019.

[6] 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.

[7] 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. JUL.15, pp. 133–146, 2019.

[8] 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.

[9] 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.

[10] Z. Zhang, "Research on sports video content classification based on artificial intelligence technology," *Modern Electronic Technology*, vol. 043, no. 009, pp. 58–61, 2020.

[11] D. Pan and W. Guan, "Sports video classification method based on hidden Markov model," *Journal of Natural Science of Xiangtan University*, vol. 39, no. 001, pp. 73–77, 2017.

[12] Li Chen, H. Fei, and H. Li, "Analysis of information geometry measurement of internal migration of DNN," *Journal of Hunan University: Natural Science Edition*, vol. 046, no. 002, pp. 97–104, 2019.

[13] L. Zhang and L. Cao, "Research on video description method based on deep transfer learning," *Journal of Testing Technology*, vol. 32, no. 05, pp. 80–86, 2018.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] N. Qiu, X. Wang, P. Wang, S. Zhou, and Y. Wang, "Research on convolutional neural network algorithm combined with transfer learning model," *Computer Engineering and Applications*, vol. 56, no. 5, pp. 43–48, 2020.

[19] C. Lifu, W. Hong, C. Xianliang, G. Zhenghua, and J. Zhiwei, "Convolutional neural network SAR image target recognition based on transfer learning," *Chinese Space Science Technology*, vol. 038, no. 006, pp. 45–51, 2018.

[20] L. Xiaodong, C. Xinbao, Y. Guanghui, and L. Peng, "Exploration of ball sports target tracking methods in sports video," *Popular Science and Technology*, vol. 20, no. 01, pp. 114–116, 2018.

[21] Y. Zhang, "Research on the classification of sports video types based on machine learning technology," *Microcomputer Applications*, vol. 36, no. 11, pp. 46–48, 2020.

[22] Y. Wang, M. Liu, and W. Yan, "Ant colony optimization algorithm to optimize support vector machine video classification," *Modern Electronic Technology*, vol. 043, no. 001, pp. 56–58, 2020.

[23] D. Liu, X. Wang, and M. Liu, "Research on olympic video classification mechanism based on Lagrangian gauss transform," *Optoelectronics Letters*, vol. 30, no. 10, pp. 104–109, 2019.

[24] Y. Zhou and B. Yang, "Sports video athlete detection using convolutional neural network," *Journal of Natural Science of Xiangtan University*, vol. 39, no. 001, pp. 95–98, 2017.

[25] E. Deniz, A. Sengür, Z. Kadiroglu, Y. Guo, V. Bajaj, and U. Budak, "Transfer learning based histopathologic image classification for breast cancer detection," *Health Information Science and Systems*, vol. 6, no. 1, pp. 1–7, 2018.

[26] G. Zhou, Z. Zeng, J. X. Huang, and T. He, "Transfer learning for cross-lingual sentiment classification with weakly shared deep neural networks," in *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in information Retrival*, vol. 2016, pp. 245–254, Pisa, Italy, July 2016.

[27] G. Chen, J. Qi, C. Tang, Y. Wang, Y. Wu, and X. Shi, "Analysis and research of key genes in gene expression network based on complex network," *Complexity*, vol. 2020, pp. 1–12, Article ID 8816633, 2020.