International Journal of Engineering and Information Systems (IJEAIS)
ISSN: 2643-640X
Vol. 7 Issue 11, November - 2023, Pages: 15-26

Models of Video Data Classification Using Convolutional Neural Networks

Iryna Tvoroshenko, Viktor Pomazan, Volodymyr Gorokhovatskyi, Oleg Kobylin

Department of Informatics
Kharkiv National University of Radio Electronics,
Kharkiv, Ukraine
e-mail: iryna.tvoroshenko@nure.ua; viktor.pomazan@nure.ua; volodymyr.horokhovatskyi@nure.ua; oleg.kobylin@nure.ua

Abstract*—This paper explores the classification of video data with convolutional neural networks. It discusses how convolutional neural network architecture can do this task with fastness and maximum efficiency. By analyzing the different convolutional neural network models model was proposed that showed great results in this video classification task. The strengths and weaknesses of the neural network model for video classification have been identified, and the prospects for further work have also been outlined.*

Keywords—artificial intelligence; classification; convolutional neural networks (CNN); neural network; video classification

1. Introduction

Video classification is the intricate process of categorizing video content automatically into distinct classes or categories based on their inherent content, attributes, and features. This intricate procedure encompasses the analysis of audio, visual, and textual data, coupled with the application of machine learning and deep learning algorithms to identify discernible patterns and trends within the data.

The significance of video classification is multifaceted and extends across various dimensions:

- Huge Volumes of Video: with the emergence of streaming platforms, video blogs, social networks, and other online resources, the volume of video content is growing exponentially. Classification helps users quickly find content that interests them.
- Personalization: classification algorithms enable platforms to provide users with content that aligns with their preferences and interests, enhancing the user experience.
- Recommendations and Advertising: video classification allows recommendation systems to suggest content that users may find interesting. This also contributes to more effective and relevant advertising.
- Content Analysis: classification enables the analysis and evaluation of content across various aspects, such as emotional tone, and the presence of specific objects, scenes, or individuals.
- Monitoring and Safety: automated video classification can be used to track unwanted content, such as violence, extremism, or misinformation, enhancing user safety.
- Education and Medicine: in education, classification aids students in quickly accessing educational materials. In medicine, video data analysis can be used for diagnosis and patient monitoring.
- Robotics and Automation: video classification can be a crucial part of visual perception systems in robots and autonomous devices.

The overarching goal of video classification is to strike the right balance between automation and accuracy, to ensure the best experience for users, and effective management of video material.

This paper aims to research and develop artificial intelligence methods based on neural networks for automatic video data classification. Additionally, we will explore contemporary neural network architectures and video data preprocessing methods to determine the most effective approaches for classification.

Overall deep learning methods have significantly enhanced the effectiveness of video classification models, enabling them to attain remarkable precision and resilience.

International Journal of Engineering and Information Systems (IJEAIS)
ISSN: 2643-640X
Vol. 7 Issue 11, November - 2023, Pages: 15-26

2. RELATED WORKS

The analysis of the available literature shows a general trend of application and improvement of deep neural networks, namely convolutional neural networks (CNNs), which are constantly used to solve various computer vision problems. Provides an overview of the research on deep neural networks and their application to video data classification.

Papers [1-3] address the task of recognizing human actions in web videos, emphasizing the application of computer vision and machine learning methods for video action classification. Article [4] discusses the application of convolutional neural networks in medical imaging, particularly in radiology. The authors provide an overview of convolutional neural networks and their application in image analysis for diagnostic purposes. The works [5, 6] describe the capabilities of image and video classification using convolutional neural networks, along with their prospects in this direction.

In the paper [7], an informational resource created by MathWorks. It provides information on applying deep learning using the MATLAB programming environment. This resource offers an overview of deep learning methods, instructions on using MATLAB tools for creating and training neural networks, code examples, and practical guidance. In the article [8] authors discuss various deep learning architectures, data preprocessing methods, and key trends in the field of video classification using neural networks.

An analysis of recent publications [1-8] has shown that the issues under consideration have not been sufficiently researched and that scientific results are not in all cases brought to practical implementation and need further research, which confirms the relevance and importance of theoretical and practical results obtained in the work.

3. MATERIALS AND MODELS

The template is used to format your paper and style the text. All margins, column widths, line spaces, and text fonts are prescribed; please do not alter them. You may note peculiarities. For example, the head margin in this template measures proportionately more than is customary.

This measurement and others are deliberate, using specifications that anticipate your paper as one part of the entire proceedings, and not as an independent document. Please do not revise any of the current designations.

3.1 Step 1

The dataset that was used has 50 video categories with 6676 videos, with a minimum of 100 videos for each class. Also, this dataset is an improvement of UCF11 dataset which was much less than UCF50 [1]. UCF50 data set's 50 action categories were collected from YouTube. The examples of data from the dataset UCF are shown in Fig. 1 and Fig. 2.

Fig. 1. *Videos of "PushUps" activity class.*

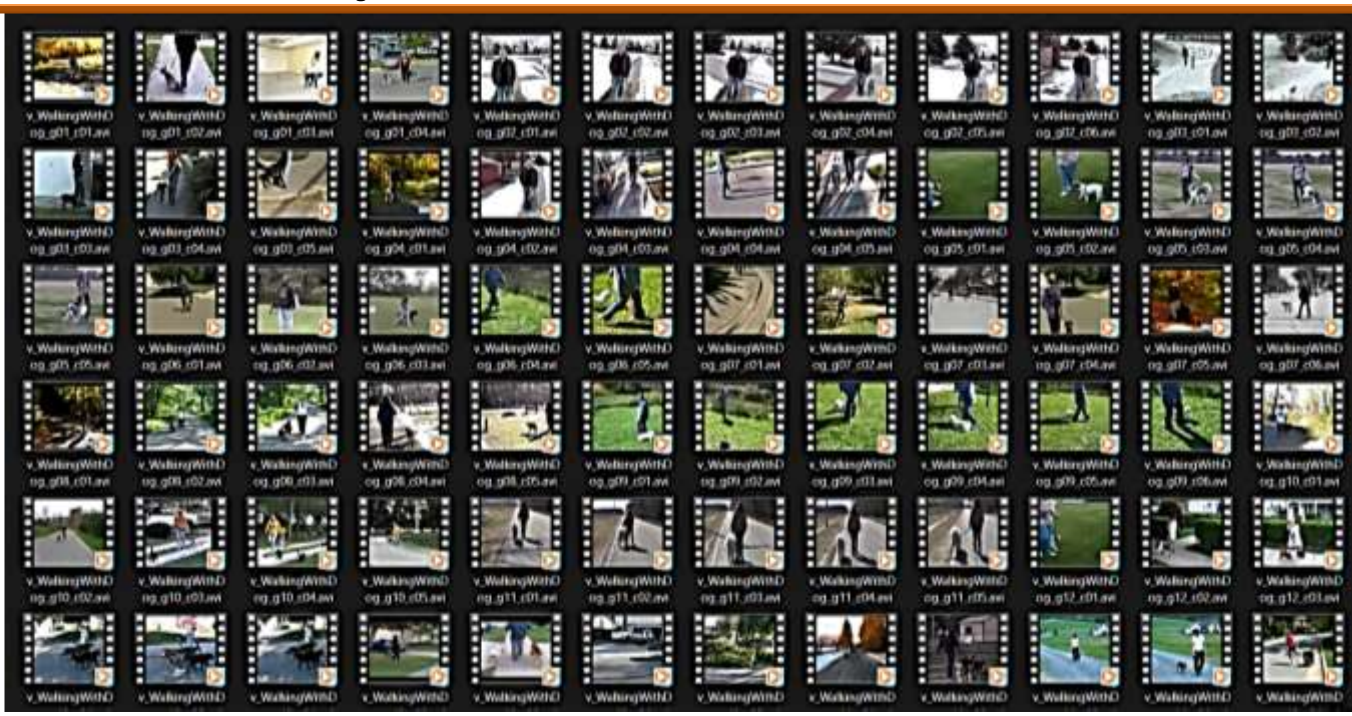

Fig. 2. *Videos of "Walking with dog" activity class*.

3.2 Step 2

CNNs are a subset of artificial neural networks (ANNs) that are very good at solving various visual tasks: image search, image segmentation, image classification, face recognition, road sign recognition, and much more.

The inception of CNNs dates back to the 1960s and marked a significant milestone in the field of computer vision, solidifying their status as the foremost neural network architecture in the realm of deep learning. CNNs have exhibited their prowess in addressing intricate visual challenges that demand substantial computational resources. Their primary applications extend to fields like video processing, segmentation, object detection [9-11], image classification [12-14], speech recognition [15-17], and natural language processing.

To illustrate, researchers such as Shri have harnessed CNNs for video analysis [5], while Roncancio has applied them to image analysis [6]. CNNs function as feedforward networks, meaning that data flows unidirectionally from inputs to outputs. Inspired by the structure of the brain's visual cortex, notably the layered organization of simple and complex cells as elucidated by Hubel and Wiesel in their groundbreaking work from 1959 and 1962, CNN architectures have evolved. These architectures come in various forms but generally comprise layers of convolutional and pooling operations organized into modular components [2-6]. CNNs have revolutionized the field of IT, proving to be indispensable tools for a wide array of visual and perceptual tasks. Their ability to mimic certain aspects of human visual processing has led to significant advancements across numerous domains, making them a cornerstone technology in the realm of artificial intelligence.

For the classification of video data using architecture [7], a similar approach can be followed as with image classification, but with some additional steps related to video processing [8].

Video data preparation: the video dataset should be divided into training, validation, and testing sets. The video frame should be labeled with a corresponding class [8].

Frame extraction: videos can be considered as sequences of frames. Before utilizing the architecture for classification, frames need to be extracted from the videos. This was achieved using the video processing library OpenCV.

Applying the model to frames: apply the same architecture you described to classify each extracted frame. This allows obtaining classifications for each video frame.

Results aggregation: various approaches can be used to obtain the final video classification [8]. For instance, probabilities or predicted classes for all frames can be averaged, and a decision can be made based on this aggregated information. Another approach involves using recurrent or convolutional architectures that can account for the sequence of frames and treat them as temporal data.

International Journal of Engineering and Information Systems (IJEAIS)
ISSN: 2643-640X
Vol. 7 Issue 11, November - 2023, Pages: 15-26

Training and evaluation: similar to training models for images, video data can be used for training and testing your model. Assess the model's performance on validation and testing sets.

Inference (application): after successful training and testing, you can apply the trained model to classify new videos. Pass each frame through the model and aggregate the results to obtain the overall video classification.

Overall, this process involves several key steps, including data preparation, frame extraction, model application to frames, and aggregation of results, training, and evaluation, and finally, the deployment of the trained model for video classification tasks. By considering the temporal aspects of video data and utilizing deep learning architectures, this approach allows for accurate and effective classification of video content.

The CNN design is shown in Fig. 3.

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 62, 62, 64)	1792
conv2d_1 (Conv2D)	(None, 60, 60, 64)	36928
batch_normalization (BatchNormalization)	(None, 60, 60, 64)	256
max_pooling2d (MaxPooling2D)	(None, 30, 30, 64)	0
global_average_pooling2d (GlobalAveragePooling2D)	(None, 64)	0
dense (Dense)	(None, 256)	16640
batch_normalization_1 (BatchNormalization)	(None, 256)	1024
dense_1 (Dense)	(None, 50)	12850

Total params: 69490 (271.45 KB)
Trainable params: 68850 (268.95 KB)
Non-trainable params: 640 (2.50 KB)

Fig. 3. *Design of custom CNN for video classification*.

In general, the architecture closely follows a widely used framework in deep learning specifically designed for processing images and video data. The process starts with convolutional layers, which are responsible for detecting complex patterns from raw pixel data. After that, normalization and pooling layers come into play, reducing spatial dimensions and addressing the issue of overfitting. The inclusion of normalization and pooling layers helps manage data complexity and prevents the model from learning noise in the training dataset. To further condense spatial information, a global average pooling layer is introduced, which compresses spatial data into a compact vector representation. This vector captures the most important information, discarding less significant details, and allowing subsequent layers to work more efficiently. The final dense layers, which are fully connected, handle classification based on high-level features extracted from earlier layers. These features encompass distinguishing traits and characteristics of the input data. To enhance training stability and expedite convergence, batch normalization is employed.

This technique standardizes input data for each layer, helping to prevent internal coordinate shifts and promoting more consistent and faster learning. With a total of 50 output classes, the architecture is particularly well-suited for the accurate categorization of videos across various human actions and activities. This broad classification potential demonstrates the model's versatility and its ability to recognize complex visual patterns, making it a valuable tool in video analysis and recognition tasks.

3.3 Step 3

To evaluate the performance of the video data classification application, the UCF-50 dataset was obtained. This dataset comprises a wide array of video data encompassing various themes and scenarios.

The initial step involved downloading and preparing this dataset for training and testing purposes. Following the dataset preparation, a convolutional neural network (CNN) was selected as the model architecture for this task. The CNN is well-suited for tasks involving visual data, as it can effectively capture spatial patterns and features from images or frames within videos.

The training phase involved feeding the prepared video data into the CNN. The network learned to recognize intricate patterns and features present in the videos, gradually adjusting its internal parameters to optimize the classification accuracy. During training, the network iteratively adjusted its weights based on the differences between predicted and actual class labels. Once the CNN was trained on the UCF-50 dataset, the next step was to evaluate its performance on the test dataset. The test dataset was preprocessed by breaking down the videos into individual frames and resizing them to a consistent size. This step ensured that all input data had the same dimensions, a requirement for CNN architectures.

Subsequently, the trained CNN was deployed to process and classify the individual frames from the test videos. The network's learned features were used to differentiate between different video classes. As the frames passed through the network's layers, the model assigned each frame to a specific class based on the patterns it recognized. The outcome of this classification process was the accurate categorization of the test videos into their respective classes. This process demonstrated the effectiveness of the trained CNN in discerning and classifying diverse video content.

The utilization of deep learning techniques and the CNN architecture allowed the application to achieve reliable and accurate video data classification results.

The results of the testing application are shown in Fig. 4 – Fig. 10.

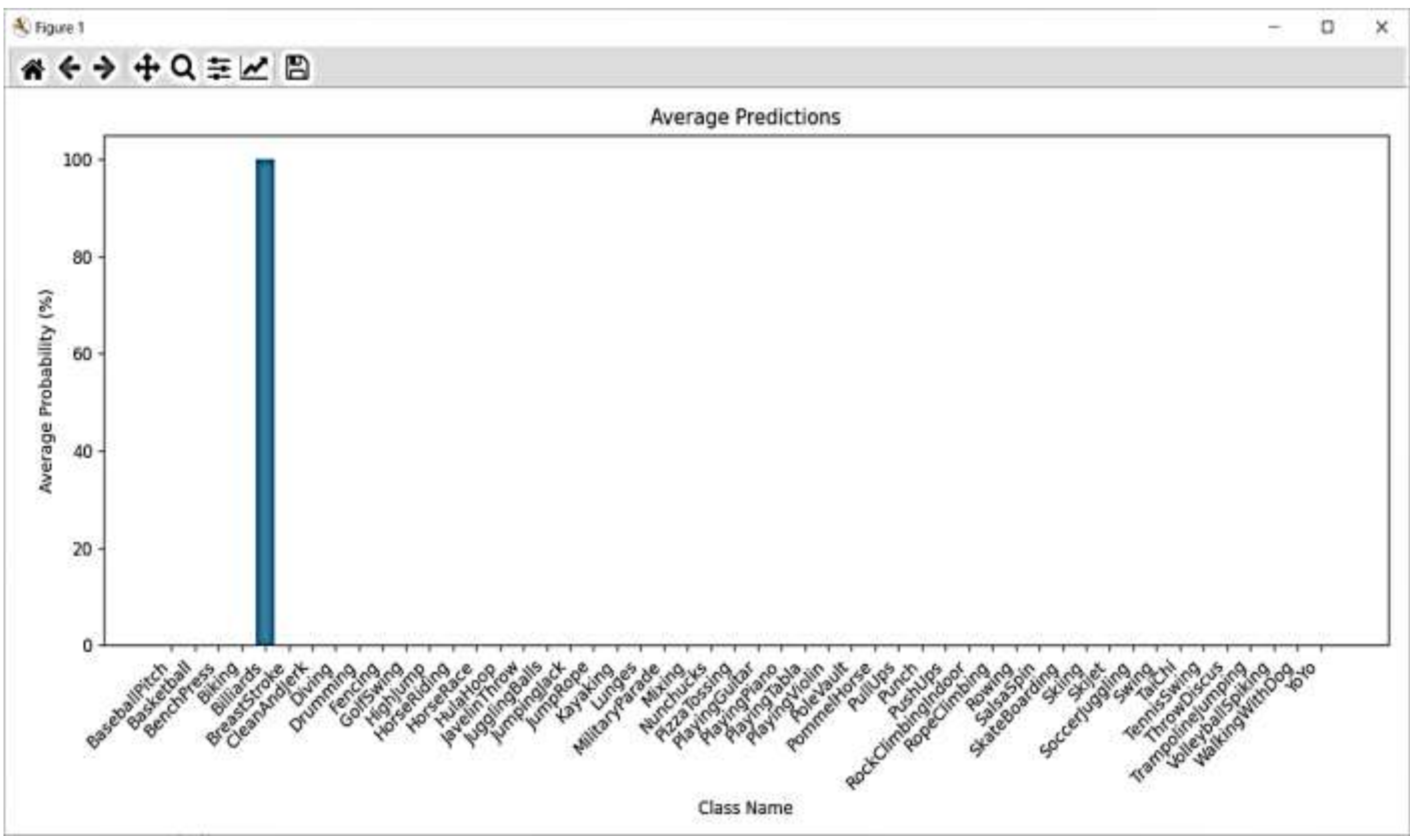

Fig. 4. *Result of classification video with billiards activity.*

International Journal of Engineering and Information Systems (IJEAIS)
ISSN: 2643-640X
Vol. 7 Issue 11, November - 2023, Pages: 15-26

```
CLASS NAME: Billiards   AVERAGED PROBABILITY: 100.0%
CLASS NAME: HighJump   AVERAGED PROBABILITY: 0.0%
CLASS NAME: Punch   AVERAGED PROBABILITY: 0.0%
CLASS NAME: CleanAndJerk   AVERAGED PROBABILITY: 0.0%
CLASS NAME: Fencing   AVERAGED PROBABILITY: 0.0%
CLASS NAME: Lunges   AVERAGED PROBABILITY: 0.0%
CLASS NAME: PullUps   AVERAGED PROBABILITY: 0.0%
CLASS NAME: PoleVault   AVERAGED PROBABILITY: 0.0%
CLASS NAME: MilitaryParade   AVERAGED PROBABILITY: 0.0%
CLASS NAME: PlayingViolin   AVERAGED PROBABILITY: 0.0%
CLASS NAME: WalkingWithDog   AVERAGED PROBABILITY: 0.0%
CLASS NAME: Diving   AVERAGED PROBABILITY: 0.0%
CLASS NAME: RockClimbingIndoor   AVERAGED PROBABILITY: 0.0%
CLASS NAME: JugglingBalls   AVERAGED PROBABILITY: 0.0%
CLASS NAME: SalsaSpin   AVERAGED PROBABILITY: 0.0%
CLASS NAME: Drumming   AVERAGED PROBABILITY: 0.0%
CLASS NAME: PommelHorse   AVERAGED PROBABILITY: 0.0%
CLASS NAME: Kayaking   AVERAGED PROBABILITY: 0.0%
CLASS NAME: Nunchucks   AVERAGED PROBABILITY: 0.0%
CLASS NAME: BenchPress   AVERAGED PROBABILITY: 0.0%
CLASS NAME: HorseRace   AVERAGED PROBABILITY: 0.0%
CLASS NAME: SkateBoarding   AVERAGED PROBABILITY: 0.0%
CLASS NAME: Skijet   AVERAGED PROBABILITY: 0.0%
CLASS NAME: TennisSwing   AVERAGED PROBABILITY: 0.0%
CLASS NAME: BreastStroke   AVERAGED PROBABILITY: 0.0%
CLASS NAME: Skiing   AVERAGED PROBABILITY: 0.0%
CLASS NAME: YoYo   AVERAGED PROBABILITY: 0.0%
CLASS NAME: RopeClimbing   AVERAGED PROBABILITY: 0.0%
CLASS NAME: PlayingGuitar   AVERAGED PROBABILITY: 0.0%
```

Fig. 5. *Calculation of average probability for all classes after classification "Billiards".*

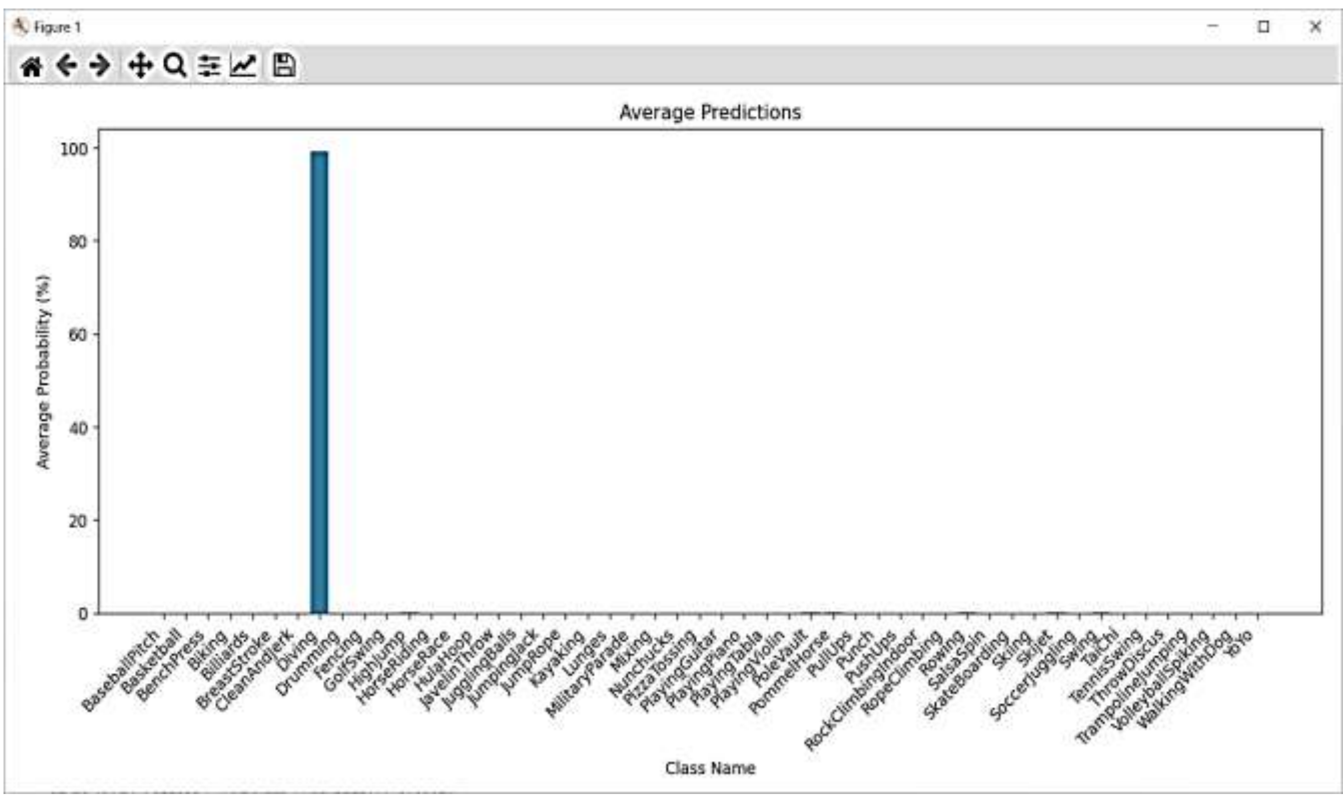

Fig. 6. *Result of classification video with diving activity.*

International Journal of Engineering and Information Systems (IJEAIS)
ISSN: 2643-640X
Vol. 7 Issue 11, November - 2023, Pages: 15-26

```
CLASS NAME: Diving   AVERAGED PROBABILITY: 99.145%
CLASS NAME: Swing   AVERAGED PROBABILITY: 0.1918%
CLASS NAME: Rowing   AVERAGED PROBABILITY: 0.1321%
CLASS NAME: HighJump   AVERAGED PROBABILITY: 0.1174%
CLASS NAME: PoleVault   AVERAGED PROBABILITY: 0.1091%
CLASS NAME: PommelHorse   AVERAGED PROBABILITY: 0.0809%
CLASS NAME: Skijet   AVERAGED PROBABILITY: 0.0778%
CLASS NAME: SkateBoarding   AVERAGED PROBABILITY: 0.0326%
CLASS NAME: Lunges   AVERAGED PROBABILITY: 0.0226%
CLASS NAME: Skiing   AVERAGED PROBABILITY: 0.0194%
CLASS NAME: Biking   AVERAGED PROBABILITY: 0.0147%
CLASS NAME: Punch   AVERAGED PROBABILITY: 0.0133%
CLASS NAME: JavelinThrow   AVERAGED PROBABILITY: 0.0112%
CLASS NAME: MilitaryParade   AVERAGED PROBABILITY: 0.011%
CLASS NAME: HorseRiding   AVERAGED PROBABILITY: 0.0037%
CLASS NAME: Billiards   AVERAGED PROBABILITY: 0.003%
CLASS NAME: Basketball   AVERAGED PROBABILITY: 0.0023%
CLASS NAME: CleanAndJerk   AVERAGED PROBABILITY: 0.0022%
CLASS NAME: ThrowDiscus   AVERAGED PROBABILITY: 0.0018%
CLASS NAME: RockClimbingIndoor   AVERAGED PROBABILITY: 0.0015%
CLASS NAME: PushUps   AVERAGED PROBABILITY: 0.0012%
CLASS NAME: PlayingPiano   AVERAGED PROBABILITY: 0.001%
CLASS NAME: RopeClimbing   AVERAGED PROBABILITY: 0.0009%
CLASS NAME: PlayingTabla   AVERAGED PROBABILITY: 0.0007%
CLASS NAME: TaiChi   AVERAGED PROBABILITY: 0.0005%
CLASS NAME: Kayaking   AVERAGED PROBABILITY: 0.0005%
CLASS NAME: TrampolineJumping   AVERAGED PROBABILITY: 0.0003%
CLASS NAME: PlayingGuitar   AVERAGED PROBABILITY: 0.0003%
CLASS NAME: YoYo   AVERAGED PROBABILITY: 0.0003%
CLASS NAME: BreastStroke   AVERAGED PROBABILITY: 0.0003%
CLASS NAME: VolleyballSpiking   AVERAGED PROBABILITY: 0.0002%
CLASS NAME: SalsaSpin   AVERAGED PROBABILITY: 0.0002%
CLASS NAME: PullUps   AVERAGED PROBABILITY: 0.0002%
CLASS NAME: Nunchucks   AVERAGED PROBABILITY: 0.0001%
CLASS NAME: Mixing   AVERAGED PROBABILITY: 0.0001%
CLASS NAME: JumpingJack   AVERAGED PROBABILITY: 0.0001%
CLASS NAME: WalkingWithDog   AVERAGED PROBABILITY: 0.0001%
```

Fig. 7. *Calculation of average probability for all classes after classification "Diving".*

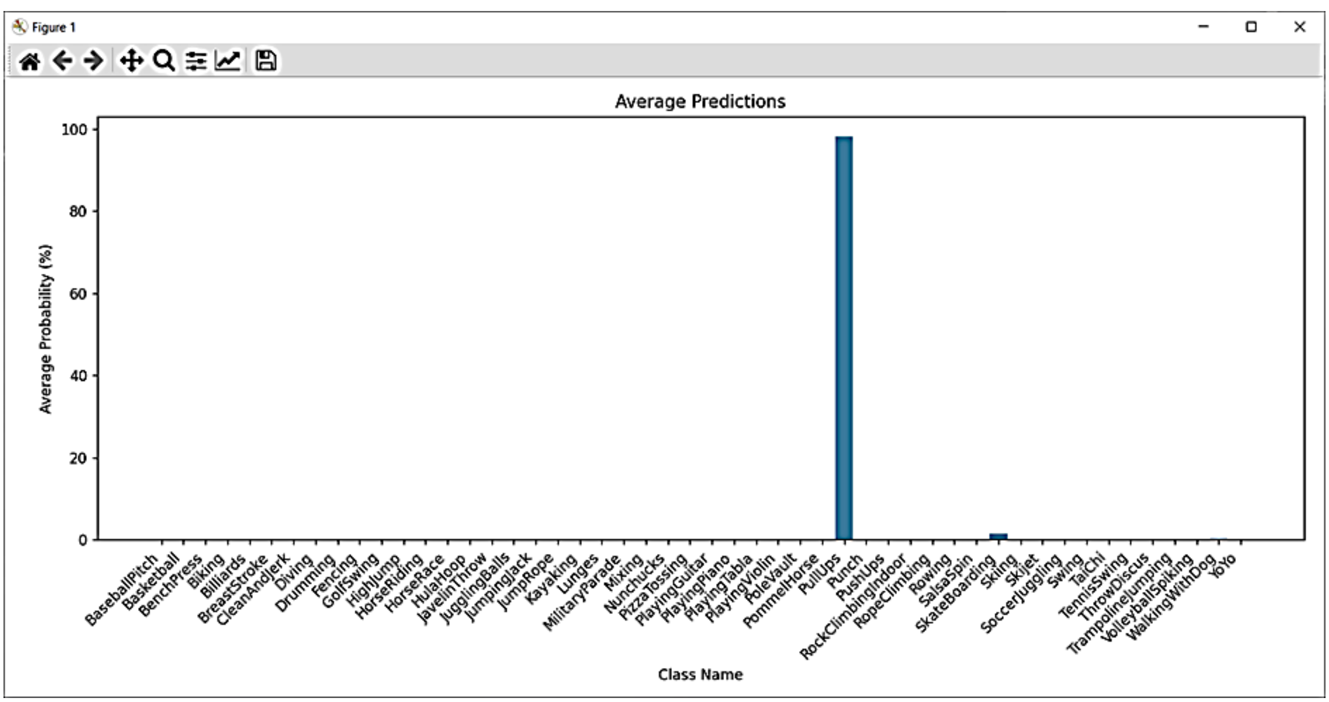

Fig. 8. *Result of classification video with pull-up activity.*

```
CLASS NAME: PullUps    AVERAGED PROBABILITY: 98.1533%
CLASS NAME: SkateBoarding    AVERAGED PROBABILITY: 1.3983%
CLASS NAME: WalkingWithDog    AVERAGED PROBABILITY: 0.2214%
CLASS NAME: Drumming    AVERAGED PROBABILITY: 0.0963%
CLASS NAME: JumpRope    AVERAGED PROBABILITY: 0.0442%
CLASS NAME: Skiing    AVERAGED PROBABILITY: 0.0237%
CLASS NAME: PlayingViolin    AVERAGED PROBABILITY: 0.022%
CLASS NAME: Biking    AVERAGED PROBABILITY: 0.0086%
CLASS NAME: PommelHorse    AVERAGED PROBABILITY: 0.0059%
CLASS NAME: JumpingJack    AVERAGED PROBABILITY: 0.0058%
CLASS NAME: RopeClimbing    AVERAGED PROBABILITY: 0.0044%
CLASS NAME: Fencing    AVERAGED PROBABILITY: 0.0031%
CLASS NAME: PushUps    AVERAGED PROBABILITY: 0.0024%
CLASS NAME: Billiards    AVERAGED PROBABILITY: 0.0%
CLASS NAME: CleanAndJerk    AVERAGED PROBABILITY: 0.0%
CLASS NAME: TennisSwing    AVERAGED PROBABILITY: 0.0%
CLASS NAME: MilitaryParade    AVERAGED PROBABILITY: 0.0%
CLASS NAME: HorseRace    AVERAGED PROBABILITY: 0.0%
CLASS NAME: BreastStroke    AVERAGED PROBABILITY: 0.0%
CLASS NAME: Diving    AVERAGED PROBABILITY: 0.0%
CLASS NAME: ThrowDiscus    AVERAGED PROBABILITY: 0.0%
CLASS NAME: VolleyballSpiking    AVERAGED PROBABILITY: 0.0%
CLASS NAME: BaseballPitch    AVERAGED PROBABILITY: 0.0%
CLASS NAME: JavelinThrow    AVERAGED PROBABILITY: 0.0%
```

Fig. 9. *Calculation of average probability for all classes after classification "PullUps".*

```
300/300 [==============================] - 339s 1s/step - loss: 0.3990 - accuracy: 0.8916 - val_loss: 1.0204 - val_accuracy: 0.7050
Epoch 12/50
300/300 [==============================] - 304s 1s/step - loss: 0.3475 - accuracy: 0.9036 - val_loss: 1.0461 - val_accuracy: 0.7073
Epoch 13/50
300/300 [==============================] - 290s 965ms/step - loss: 0.3218 - accuracy: 0.9113 - val_loss: 1.4863 - val_accuracy: 0.6233
Epoch 14/50
300/300 [==============================] - 286s 951ms/step - loss: 0.2860 - accuracy: 0.9202 - val_loss: 0.7414 - val_accuracy: 0.7829
Epoch 15/50
300/300 [==============================] - 298s 995ms/step - loss: 0.2613 - accuracy: 0.9296 - val_loss: 1.1029 - val_accuracy: 0.7033
Epoch 16/50
300/300 [==============================] - 289s 962ms/step - loss: 0.2439 - accuracy: 0.9327 - val_loss: 1.2924 - val_accuracy: 0.6929
Epoch 17/50
300/300 [==============================] - 292s 973ms/step - loss: 0.2291 - accuracy: 0.9359 - val_loss: 0.6503 - val_accuracy: 0.8050
Epoch 18/50
300/300 [==============================] - 282s 938ms/step - loss: 0.2094 - accuracy: 0.9430 - val_loss: 0.8512 - val_accuracy: 0.7625
Epoch 19/50
Epoch 34/50
300/300 [==============================] - 2833s 9s/step - loss: 0.0918 - accuracy: 0.9720 - val_loss: 1.2472 - val_accuracy: 0.7179
Epoch 35/50
300/300 [==============================] - 194s 647ms/step - loss: 0.0967 - accuracy: 0.9718 - val_loss: 0.5122 - val_accuracy: 0.8587
Epoch 36/50
300/300 [==============================] - 238s 792ms/step - loss: 0.0856 - accuracy: 0.9749 - val_loss: 0.4340 - val_accuracy: 0.8798
188/188 [==============================] - 28s 148ms/step - loss: 0.3277 - accuracy: 0.9073
D:\Python\Lib\site-packages\keras\src\engine\training.py:3000: UserWarning: You are saving your model as an HDF5 file via `model.save()`. Thi
```

Fig. 10. *Result of CNN learning.*

4. RESULTS OF TESTING

The training results of the implemented model exhibit a notable level of quality. During the initial epochs of training, it's customary to observe relatively modest accuracy scores and higher loss values. This is unsurprising as the network is in the process of familiarizing itself with the intricate details present within the dataset. However, as training progresses, a discernible trend emerges: the accuracy metric gradually ascends while the associated losses progressively decline. These developments indicate the network's growing proficiency in recognizing the inherent characteristics of the data and subsequently learning from them. Furthermore, the model's performance is assessed on a separate test dataset. The accuracy metrics for this evaluation also experience an encouraging uptick, reaching an impressive rate of around 90.71%.

The result of classification by every class shown in Table 1 – Table 5.

International Journal of Engineering and Information Systems (IJEAIS)
ISSN: 2643-640X
Vol. 7 Issue 11, November - 2023, Pages: 15-26

Classes for video classification	Baseball Pitch	Basket Ball	Bench Press	Biking	Billiards	Breast Stroke	Clean and Jerk	Diving	Drumming	Fencing
Baseball Pitch	69.64%									
Basket Ball		38.86%								
Bench Press			81.34%							
Biking				63.13%						
Billiards					100%					
Breast Stroke						99.99%				
Clean and Jerk							83.69%			
Diving								99.14%		
Drumming									94.83%	
Fencing										72.68%

Table 1: Result of classification for class Baseball Pitch – Fencing

Classes for video classification	Golf Swing	High Jump	Horse Riding	Horse race	Hula hoop	Javelin Throw	Juggling balls	Jumping jack	Jump Rope	Kayaking
Golf Swing	89.03%									
High Jump		97.06%								
Horse Riding			55.34%							
Horse race				99.82%						
Hula hoop					92.36%					
Javelin Throw						99.62%				
Juggling balls							99.27%			
Jumping jack								99.95%		
Jump Rope									55.20%	
Kayaking										99.58%

Table 2: Result of classification for class Golf Swing – Kayaking

Table 3: Result of classification for class Lunges – Pole vault

Classes for video classification	Lunges	Millitary Parade	Mixing	Nunchucks	Pizza Tossing	Playing guitar	Playing piano	Playing table	Playing violin	Pole vault
Lunges	57.89%									
Millitary Parade		99.84%								
Mixing			99.24%							
Nunchucks				99.92%						
Pizza Tossing					93.51%					
Playing guitar						99.32%				
Playing piano							100%			
Playing table								93.52%		
Playing violin									94.89%	
Pole vault										94.26%

Classes for video classification	Pommel horse	Pull ups	Punch	Push ups	Rock climbing Indoor	Rope climbing	Rowing	Salsa spin	Skate-boarding	Skiing
Pommel horse	99.99%									
Pull ups		95.96%								
Punch			97.61%							
Push ups				99.79%						
Rock climbing Indoor					94.08%					
Rope climbing						66.89%				
Rowing							60.36%			
Salsa spin								97.60%		
Skateboarding									40.75%	
Skiing										81.22%

Table 4: Result of classification for class Pommel horse – Skiing

Table 5: Result of classification for class Skyjet – Yoyo

Classes for video classification	Skyjet	Soccer juggling	Swing	Tai chi	Tennis swing	Throw discuss	Trampling Jumping	Volleyball	Walking With Dog	Yoyo
Skyjet	92.62%									
Soccer juggling		92.16%								
Swing			99.58%							
Tai chi				98.02%						
Tennis swing					99.09%					
Throw discuss						85.48%				
Trampling Jumping							99.59%			
Volleyball								34.37%		
Walking With Dog									99.48%	
Yoyo										79.94%

From the provided Table 1 – Table 5, was observed that the video classification algorithm performs with varying degrees of accuracy for different types of actions. Billiards achieves a 100% accuracy rate, indicating that this class is easily recognizable in video sequences. High Jump and Horse race also exhibit high accuracy rates, at 97.06% and 99.82%, respectively. Jumping jack and Pommel horse demonstrate exceptionally high accuracy at 99.95% and 99.99%, highlighting the algorithm's proficiency in correctly classifying these actions. Conversely, Volleyball exhibits the lowest accuracy percentage among all classes, standing at 34.37%. This suggests challenges in recognizing this particular action. Basket Ball, Biking, Playing guitar, Jump Rope, and Skateboarding also display low classification accuracy, indicating that additional enhancements to the algorithm may be required for effective recognition of these actions. In conclusion, the classification algorithm exhibits varying levels of effectiveness across different action classes, and improving accuracy for specific classes may necessitate further tuning and model training.

These results demonstrate the model's ability to generalize, as it manages to achieve high accuracy even on previously unseen data. Analysis of the accuracy trend over each subsequent training epoch reveals a consistent and steady upward trajectory. This phenomenon indicates a reliable and stable convergence of the model, which confirms the thesis that the network learning process is steadily progressing without excessive fluctuations. Accompanying the increase in accuracy, the corresponding loss values show a synchronized decrease with each epoch. This behavior is expected, as a well-trained model seeks to minimize its loss function, reflecting its ability to understand and capture the underlying patterns in the data set.

In summary, we can confidently say that the model has effectively demonstrated a significant improvement in accuracy. At the same time, the reduction in loss values maintains a reliable consistency. This combined performance emphasizes the model's ability to learn, adapt, and extract significant information from the data. The model's progress in improving accuracy and reducing loss underscores its potential for successful application in real-world environments, where it can classify and analyze video data with commendable accuracy.

5. Conclusion

This paper describes the successful development and training of a convolutional neural network (CNN) for video data classification. Based on the obtained results, it can be confidently stated that the application of CNNs in video analysis demonstrates a high potential. At the beginning of the experiment, during the initial training epochs, low accuracy and loss values were observed, which is a typical phenomenon as the network starts to learn the peculiarities of the provided data. However, over time, the network's accuracy increased and the losses decreased, indicating its ability to recognize video data characteristics and learn from them. An essential aspect of training the CNN at this stage was the utilization of data augmentation methods, which improved the network's generalization capability and made it more robust to diverse video conditions. This played a pivotal role in achieving a stable and high accuracy on the test dataset, reaching approximately 90.73%.

Analysis of the accuracy and loss graphs confirmed that the network exhibits stable convergence toward optimal values. The gradual accuracy improvement and consistent loss reduction during training are strong indicators of the neural network's successful performance. The developed CNN effectively tackled the video data classification task, achieving high accuracy on the test dataset. Its ability to recognize and generalize video data characteristics makes this approach highly promising for various applications related to video analysis and processing.

6. Acknowledgment

The authors acknowledge the support of the Department of Informatics, Kharkiv National University of Radio Electronics, Ukraine, in numerous help and support to complete this paper.

7. References

[1] Reddy, K. K., & Shah, M. (2013). Recognizing 50 human action categories of web videos, Mach. Vis. Appl., vol. 24, pp. 971-981.

[2] Bezdan, T., & Bacanin, N. (2019). Convolutional neural network layers and architectures, in Sinteza 2019 – Int. Sci. Conf. Inf. Technol. Data Relat. Res., Belgrade, Singidunum University, Serbia, pp. 445-451.

[3] Yamashita, R., Nishio, M., Do, R. K. G., & Togashi, K. (2018). Convolutional neural networks: An overview and application in radiology, Insights Imaging, vol. 9, pp. 611-629.

[4] Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition, Proc. IEEE, vol. 86, no. 11, pp. 2278-2324.

[5] Shri, S. J., & Jothilakshmi, S. (2019). Crowd video event classification using convolutional neural network, Comput. Commun., vol. 147, pp. 35-39.

[6] Roncancio, R., Gamal, A. E., & Gore, J. P. (2022). Turbulent flame image classification using convolutionalneural networks, Energy AI, vol. 10, p. 100193.

[7] The Mathworks, Introducing Deep Learning with MATLAB, 2018.

[8] Rehman, A., & Belhaouari, S. B. (2021). Deep learning for video classification: A review, TechRxiv. Preprint.

[9] Pomazan, V., Tvoroshenko, I., & Gorokhovatskyi, V. (2023). Development of an application for recognizing emotions using convolutional neural networks, Int. J. Acad. Inf. Syst. Res., vol. 7, no. 7, pp. 25-36.

[10] Gorokhovatskyi, V., Peredrii, O., Tvoroshenko, I., & Markov, T. (2023). Distance matrix for a set of structural description components as a tool for image classifier creating, Adv. Inf. Syst., vol. 7, no. 1, pp. 5-13. Гороховатський, В. О., Передрій, О. О., Творошенко, І. С., & Марков, Т. Є. (2023). Матриця відстаней для множини компонентів структурного опису як інструмент для створення класифікатора зображень, Сучасні інформаційні системи, 7(1), С. 5-13.

[11] Gadetska, S. V., Gorokhovatskyi, V. O., Stiahlyk, N. I., & Vlasenko, N. V. (2021). Statistical data analysis tools in image classification methods based on the description as a set of binary descriptors of key points, Radio Electron. Comput. Sci. Control, vol. 4, pp. 58-68.

[12] Gorokhovatskyi, V., Tvoroshenko, I., Kobylin, O., & Vlasenko, N. (2023). Search for visual objects by request in the form of a cluster representation for the structural image description, Adv. Electr. Electron. Eng., vol. 21, no. 1, pp. 19-27.

[13] Daradkeh, Y. I., Gorokhovatskyi, V., Tvoroshenko, I., & Zeghid, M. (2022). Tools for fast metric data search in structural methods for image classification, IEEE Access, vol. 10, pp. 124738-124746.

[14] Pomazan V., Tvoroshenko I., & Gorokhovatskyi V. (2023). Handwritten character recognition models based on convolutional neural networks, Int. J. Acad. Eng. Res., vol. 7, no. 9, pp. 64-72.

[15] Tvoroshenko I., Gorokhovatskyi V., Kobylin O., & Tvoroshenko A. (2023). Application of deep learning methods for recognizing and classifying culinary dishes in images, Int. J. Acad. Appl. Res., vol. 7, no. 9, pp. 57-70.

[16] Gadetska, S., Gorokhovatskyi, V., Stiahlyk, N., & Vlasenko, N. (2022). Aggregate Parametric Representation of Image Structural Description in Statistical Classification Methods.

[17] Gorokhovatskyi, V., Gadetska, S., & Stiahlyk, N. (2020). Image structural classification technologies based on statistical analysis of descriptions in the form of bit descriptor set.