Comparison of Four Convolutional Neural Network-Based Algorithms for Sports Image Classification

Xiangchen Liu

Liangxin College, China Jiliang University, Zhejiang, Hangzhou, 310018, China
2100201105@cjlu.edu.cn

Abstract. Sports image classification is a challenging task that involves multiple types of sports, with difficulties in feature recognition and suboptimal detection results. This study employs four pretrained models, namely Residual Network 50 (ResNet-50), EfficientNet B7, Densely Connected Convolutional Network 121 (DenseNet-121), and You Only Look Once version 8 (YOLOv8), to address the problem of classifying 100 different sports image categories. The dataset contains 12,200 sports images, which serves as a robust experimental foundation of this research. By comparison their performances it could be found that ResNet-50 exhibited outstanding performance on the training set, achieving an accuracy of 90.80%, and 88.75% on validation set. The EfficientNet B7 model, achieved an accuracy of 37.45% for training and 62.42% for inference. The less impressive performance possibly due to its limited representation capabilities when dealing with specific sports image classification tasks. DenseNet-121 attained an accuracy of 71.791% on the training and 86.211% on the validation set. Compared to EfficientNet B7, its performance is better, suggesting the dense connectivity architecture is more suitable for extracting image features. Furthermore, YOLOv8n model delivered exceptional performance with an average accuracy of 94.90% on the training set, 96.60% on the validation set. These results showcase the robust performance of YOLOv8n in sports image classification and detection. In conclusion, this study provides valuable insights into addressing complex image classification problems by comparing the performance of different algorithms in sports image classification. Understanding the strengths and weaknesses of these various algorithms is crucial for a deeper comprehension of image classification tasks and guiding future research endeavors.

Keywords: Neural Network, Sports Image Classification, Deep Learning.

1 Introduction

Sports image classification is an important field in computer vision and machine learning with widespread applications [1]. It can be applied to athlete motion analysis, sports event video analysis, intelligent fitness monitoring, and more. Accurate classification of different sports images can assist coaches in analyzing athlete techniques more effectively, aid referees in making more precise judgments of fouls in sports matches, and enable intelligent devices to monitor training outcomes and provide recommendations.

B. H. Ahmad (ed.), *Proceedings of the 2023 International Conference on Data Science, Advanced Algorithm and Intelligent Computing (DAI 2023)*, Advances in Intelligent Systems Research 180,
https://doi.org/10.2991/978-94-6463-370-2_20

Therefore, research in sports image classification algorithms holds significant importance in advancing sports training and the sports industry [2,3].

Traditional methods for sports image classification mainly rely on manually extracted visual features combined with machine learning algorithms. However, these methods are not robust to image variations and struggle to adapt to changes in position, angle, and environmental conditions [4]. In recent years, methods based on deep neural networks have made remarkable progress in image classification tasks. Convolutional neural networks (CNNs) can automatically learn feature representations from a large amount of labeled data, exhibiting greater adaptability to image variations. Currently, deep learning has become the mainstream approach in image classification [5].

This study focuses on a dataset containing 100 sports categories and employ four typical deep neural network models, including Residual Network 50 (ResNet-50), Densely Connected Convolutional Network 121 (DenseNet-121), EfficientNet-B7, and You Only Look Once version 8 (YOLOv8), to address the multi-class sports image classification problem. The classification performance of different models is compared and their respective strengths, providing insights for the development of improved sports image classification systems are analyzed. The structure of this paper is as follows: the first part introduces the research background and significance; the second part describes the dataset; the third part provides detailed descriptions of the deep learning models used; the fourth part outlines the model training and evaluation processes; and finally, the experimental results are summarized and discussed. This research can serve as a valuable reference for the development of sports image processing and intelligent analysis technologies.

2 Method

2.1 Dataset

This experiment utilized a dataset named "100 Sports Image Classification." The dataset was collected by the authors from the internet and underwent a process of duplicate image detection to avoid any image duplication, ensuring the independence of the training, validation, and test datasets [6]. The dataset comprises 100 categories of sports images, with each category containing 112 images, resulting in a total of 11,200 images. During the dataset loading process, batch processing was employed to enhance training efficiency. The dataset was split into training, validation, and test sets in a ratio of 70%, 15%, and 15%, respectively.

Simultaneously, the authors applied a series of data preprocessing steps to the collected internet images to ensure data quality and consistency. For example, image dimensions were resized to 224x224 to ensure that the models were trained on images of the same size, preventing biases in input data dimensions. To enhance the robustness of the models, data augmentation techniques were applied during the training phase. Random cropping, horizontal flipping, color transformations, and other operations were employed to generate a more diverse set of training samples. This helped improve the models' adaptability to images under different angles and lighting conditions.

2.2 Models

In this study, four major deep learning architectures, namely YOLO, ResNet, DenseNet, and EfficientNet, were implemented for the image classification task using the PyTorch framework.

ResNet-50. ResNet-50 is one of the representative models in the residual network series, introduced by researchers including Kaiming He from Microsoft Research in 2015 [7]. Its primary innovation lies in the introduction of residual structures, which successfully addressed the problem of vanishing gradients when deepening neural networks, making training deep networks possible. ResNet-50 consists of a total of 50 network layers, with a significant usage of 3-layer residual blocks. Each residual block contains multiple convolutional layers, and it employs a "skip connection" by directly adding the input to the output after two convolutional layers. This skip connection allows the input to bypass these layers, preventing the issue of vanishing gradients. This residual learning framework significantly extends the network's depth, making ResNet-50 much deeper compared to earlier popular models.

In image classification tasks, ResNet-50's depth of up to 50 layers enables it to extract richer feature representations and be more sensitive to subtle image details. It exhibits excellent performance on large-scale image classification datasets like ImageNet, with a Top-5 error rate as low as 5.25%. Many studies have also shown that ResNet-50's deep architecture is well-suited for fine-grained image classification tasks, such as the 100-class sports classification problem in this research. Additionally, ResNet-50 provides pre-trained weights obtained on a large amount of image data, which serves as a good initialization for model parameters, leading to faster and better convergence during training. As a result, ResNet-50 has become one of the most widely used models in computer vision.

However, due to its depth, ResNet-50 does have drawbacks such as longer training times and the potential for overfitting. In summary, ResNet-50, as a classic deep residual network, excels in fine-grained image classification tasks, providing strong support for the problem addressed in this research.

DenseNet-121. DenseNet is an advanced convolutional neural network architecture introduced in 2017 by researchers from Cornell University, Facebook AI Research, and others [8]. Its innovative feature is the reinforcement of connections between network layers, creating a densely connected network structure. In contrast to the residual connections in ResNet, DenseNet establishes direct connections between each layer and all subsequent layers in the network. For example, the output of the l-th layer serves as additional input to the l+1 layer, and the output of the l+1 layer is not only input to the l+2 layer but also directly passed on to subsequent layers like l+3, l+4, and so on. This dense connectivity facilitates feature reuse, allowing more layers' parameters to participate in both forward and backward propagation, enhancing effective feature fusion and gradient propagation.

DenseNet-121, as a lightweight model in the DenseNet series, comprises 121 convolutional layers. Its dense connectivity design keeps the model's parameter count low, with only around 12 million parameters, while achieving outstanding performance. Particularly in fine-grained classification tasks, DenseNet-121's multi-level feature fusion capability is well-suited for capturing subtle differences in images, leading to high precision. For instance, in this research's 100-class sports classification task, DenseNet-121 is capable of distinguishing fine-grained action features across different sports.

Additionally, DenseNet-121 provides pre-trained weights on large-scale image data, making it easier for the model to converge. However, its dense connectivity structure does result in higher computational requirements and relatively longer inference times. Overall, DenseNet-121, as an efficient deep network, demonstrates good performance even under limited computational resources, making it suitable for the multi-class image classification task addressed in this research.

EfficientNet-B7. EfficientNet is a novel series of convolutional neural networks introduced by the Google Brain team in 2019 [9]. Its optimization objective is to create models with the best performance under constraints on parameters and computations. The innovation behind EfficientNet lies in the introduction of a strategy known as Compound Scaling. This strategy balances the adjustment of network depth, width, and resolution, allowing the model to scale to arbitrary sizes while maintaining a balance between performance and efficiency.

EfficientNet-B7 is the largest model in this series, containing over 72 million parameters, which is equivalent to more than 350 layers in the network. It achieves an impressive top-1 accuracy of 84.3% on the ImageNet dataset, making it the top-performing model in the EfficientNet series. Despite its large model size, EfficientNet-B7 maintains high computational efficiency. Compared to other similarly sized models, it delivers over 5 times the performance improvement while only increasing the number of floating-point operations by 50%. This efficiency is primarily attributed to the unique Compound Scaling mechanism of EfficientNet, which effectively reduces parameter redundancy by carefully balancing growth across various dimensions.

As a result, EfficientNet-B7 efficiently utilizes parameters and computational resources, translating them into model performance. This is of significant importance for the multi-class image classification task addressed in this research. Despite longer training times, EfficientNet-B7 offers exceptionally high classification accuracy, enabling the recognition of subtle differences within the 100-class sports images and achieving fine-grained classification.

YOLOv8n. YOLO, introduced by Redmon et al. in 2016, is a one-stage object detection model known for its speed and real-time object detection capabilities [10]. YOLO divides an image into a grid and directly predicts the probability of the presence of objects, their classes, and coordinates within each grid cell. The entire detection process is completed in a single forward pass, without the need for region proposals or post-processing stages, making it highly efficient. Through continuous version iterations and optimizations, the latest YOLOv8 has achieved significant improvements in both detection performance and speed.

YOLOv8n is a lightweight version of YOLOv8, where 'n' stands for nano, indicating a more lightweight model. YOLOv8n retains some of the modules and design principles of YOLOv8 but undergoes model pruning and simplification to reduce the number of parameters and computations. Additionally, YOLOv8n-cls has been optimized by adjusting the head module for image classification tasks, as opposed to general object detection. As a result, YOLOv8n-cls has a smaller model size and faster speed, making it highly suitable for real-time classification in resource-constrained environments.

Compared to more complex models, YOLOv8n-cls has fewer parameters, allowing for rapid convergence and prediction. While its classification accuracy may be slightly lower than the first three models, YOLOv8n-cls still achieves excellent results. In summary, YOLOv8n-cls offers a high-speed and efficient image classification solution, making it well-suited for the classification of a large number of samples in this research.

2.3 Evaluation Metrics

Accuracy, as a fundamental metric for assessing model classification performance, was extensively utilized in this study for four different algorithms: ResNet-50, EfficientNet B7, DenseNet-121, and YOLOv8n, to address the challenge of classifying 100 categories of sports images.

3 Result

From the experimental results, it is evident that the four models exhibit varying performance on the task of classifying 100 sports images, as shown in Table 1.

Table 1. Performance comparison.

	ResNet-50	EfficientNet-B7	DenseNet-121	YOLOv8n-cls
Accuracy	90.80%	62.42%	86.21%	96.60%

The performance of the four models on the task of classifying 100 sports images varies as follows:

After 25 training epochs, ResNet-50 achieved the highest accuracy on the test set, reaching 90.80%. For the validation dataset, the model achieved an accuracy of 88.75%. This can be attributed to its deep residual structure, which allows it to extract richer image features. However, one of its drawbacks is the longer training time. It's worth noting that the accuracy on certain categories may be lower. This could be attributed to potential class imbalance in the dataset, where some categories have more samples than others, causing the model to lean towards predicting the more numerous categories.

After 25 training epochs, EfficientNet-B7 exhibited relatively lower accuracy, with a score of only 62.42% and 37.45% on the training dataset. This could be due to the model's limited parameter count, which hinders its ability to represent complex images effectively. Compared to ResNet-50, EfficientNet B7's performance is relatively lower, which may be attributed to its limited representational capacity for handling this specific task of sports image classification. This limitation could have hindered its ability to effectively capture key features in the images.

After 25 training epochs, enseNet121 achieved an accuracy of 71.79% on the training dataset and an accuracy of 86.21% on the validation dataset. Compared to EfficientNet B7, DenseNet-121 performs better in this task; however, it still falls short of ResNet-50's performance. This observation suggests that, for this specific classification task, the dense connectivity architecture may be more suitable for feature extraction. Its dense connectivity design enhances feature propagation, but it comes with higher computational requirements.

The YOLOv8n model achieved significant accuracy results after 30 training epochs. On the training set, the average accuracy was 0.949, and on the validation set, the average accuracy was 0.966. This indicates that YOLOv8n performs exceptionally well in detecting and classifying sports images. Furthermore, the model achieved a top1 accuracy of 96.9% and a top5 accuracy of 99.8%, strongly suggesting its powerful performance in classification tasks.

In summary, ResNet-50 performs the best due to its deep architecture. DenseNet-121 and YOLOv8n-cls also achieved good results, with the former emphasizing feature fusion and the latter being optimized for classification tasks. Therefore, for the task of classifying 100 sports images, classic deep CNNs like ResNet-50 are the top choice, while DenseNet-121 and YOLOv8n-cls are strong alternatives. Further adjustments are needed for EfficientNet-B7 to improve its performance. These findings provide valuable insights for selecting different models in similar classification tasks.

4 Discussion

The experimental results demonstrate that ResNet-50, with its deep residual structure, achieved the highest accuracy of 90.8% in the task of classifying 100 sports images. This highlights the advantage of classic deep convolutional neural networks in feature extraction from images. EfficientNet-B7's performance was relatively disappointing, with an accuracy of only 62.422%. This may be attributed to its design of compound scaling strategy, which did not yield the expected results on this specific dataset and task. DenseNet-121 and YOLOv8n-cls achieved accuracies of 86.211% and 96.6%, respectively, showcasing their effectiveness in feature fusion and model optimization. Particularly, YOLOv8n-cls achieved better classification results through lightweight design. In conclusion, considering the strengths of each model, ResNet-50, DenseNet-121, and YOLOv8n-cls all deserve to be considered as preferred models for sports image classification. EfficientNet-B7 may require further adjustments to enhance its generalization capability.

From the experimental results, it is evident that different models exhibit varying degrees of sensitivity to the key hyperparameter, the learning rate.

Taking YOLOv8n-cls as an example, when the learning rate is set to 0.0005, the accuracy reaches 96.6%. Increasing the learning rate to 0.001 maintains the accuracy at around 96.6%. However, when further increasing the learning rate to 0.01 and 1, there is a noticeable fluctuation and decline in accuracy. This suggests that YOLOv8n-cls is not highly sensitive to changes within a smaller range of learning rates but is adversely affected by excessively large learning rates that go beyond a certain threshold. Therefore, the choice of learning rate should be made carefully, and a systematic search within an appropriate range may be necessary.

Other models like ResNet-50 also exhibit similar sensitivity to learning rates, and it's crucial to determine the optimal hyperparameter range for each specific model. Additionally, factors such as batch size and weight decay can also impact the results, necessitating a comprehensive consideration of various hyperparameters.

This study has explored the classification performance of various deep learning models, laying the foundation for the development of more robust sports analysis systems. Future research can expand in several directions to further advance the field of sports image classification and analysis:

1. Collecting Larger and More Comprehensive Datasets: Researchers can attempt to gather larger datasets that encompass a broader range of sports categories, aiming to cover various sports disciplines comprehensively. This will enhance the model's ability to train and perform well in a wider variety of sporting contexts, improving robustness and generalization.

2. Exploring Advanced Computer Vision Models: The research can explore the use of more advanced computer vision models such as EfficientNetV2, ResNeXt, and others to further enhance classification performance. With the continuous development of deep learning models, adopting the latest model architectures may yield better results.

3. Incorporating Temporal Information for Video Analysis: Researchers can consider expanding their focus to include video-based sports classification and action recognition, incorporating the temporal dimension. This would enable models to analyze dynamic features of sports more accurately, providing additional context information for sports video recognition.

4. Building a Knowledge Graph for Sports: Based on classification results, researchers can construct a sports knowledge graph to facilitate more intelligent sports analysis. This will help in gaining deeper insights into the interconnections between different sports, offering valuable insights and recommendations for sports science and training.

5. Practical Applications in Sports Training and Assessment: Exploring the practical applications of classification results in real-world scenarios such as sports training and athlete assessment. Integrating sports image classification technology with practical settings can provide useful tools for coaches, athletes, and sports scientists, helping them better understand and improve athletic performance.

As models and datasets continue to evolve, the expectation is that sports image classification will expand into the realm of intelligent sports analysis, offering greater impact and potential for the fields of sports science, sports training, and health management. Research in this domain will continue to provide strong support and innovation in sports-related areas.

5 Conclusion

This study conducted comparative experiments on image classification tasks involving 100 sports categories, utilizing four classic deep learning models: ResNet-50, EfficientNet-B7, DenseNet-121, and YOLOv8n-cls. In terms of methodology, these models were implemented using the PyTorch framework, and similar training parameters were employed to ensure comparability. The training and validation sets were divided in a 7:3 ratio, and Adam optimizer, cross-entropy loss function, as well as data augmentation techniques including flips, rotations, and brightness adjustments were applied.

The results showed that in terms of classification accuracy on the test set, ResNet-50 achieved the highest accuracy at 90.8%, performing the best. EfficientNet-B7 lagged behind with only 62.422% accuracy, DenseNet-121 achieved 86.211%, and YOLOv8n-cls achieved the highest accuracy at 96.6%. Considering the design characteristics of each model, it can be observed that ResNet-50 excelled due to its deep residual structure, enabling it to extract richer features. EfficientNet-B7's compound scaling strategy performed poorly in this specific task. DenseNet-121's dense connections enhanced feature fusion, while YOLOv8n-cls underwent specialized lightweight design.

In summary, this series of comparative experiments validates the strong expressive capability of classic deep convolutional neural network models in fine-grained image classification tasks. It is evident that rational model selection is crucial for achieving optimal results. This study provides valuable model analysis for building a more accurate sports image classification system. Future work will involve expanding the dataset size and exploring methods such as model ensemble to further enhance classification performance.

Acknowledgment

This work is supported by Zhejiang Province College Student Research and Innovation Program Funding Project (Project Title: Internal Damage Detection and Intelligent Recognition of High-Pressure Composite Hydrogen Storage Gas Cylinders)

References

1. Podgorelec, V., Pečnik, Š., & Vrbančič, G. Classification of similar sports images using convolutional neural network with hyper-parameter optimization. Applied Sciences, 10(23), 8494 (2020).
2. Rangasamy, K., As'ari, M. A., Rahmad, N. A., Ghazali, N. F., & Ismail, S. Deep learning in sport video analysis: a review. Telecommunication Computing Electronics and Control, 18(4), 1926-1933 (2020).
3. Naik, B. T., Hashmi, M. F., & Bokde, N. D. A comprehensive review of computer vision in sports: Open issues, future trends and research directions. Applied Sciences, 12(9), 4429 (2022).

4. Loussaief, S., & Abdelkrim, A. Machine learning framework for image classification. In 2016 7th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications, 58-61 (2016).
5. Rawat, W., & Wang, Z. Deep convolutional neural networks for image classification: A comprehensive review. Neural computation, 29(9), 2352-2449 (2017).
6. 100 Sports Image Classification, URL: https://www.kaggle.com/datasets/gpiosenka/sports-classification, Last Accessed 2023/09/04
7. He, K., Zhang, X., Ren, S., & Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, 770-778 (2016).
8. Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition, 4700-4708 (2017).
9. Tan, M., & Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In International conference on machine learning, 6105-6114 (2019).
10. Terven, J., & Cordova-Esparza, D. A comprehensive review of YOLO: From YOLOv1 to YOLOv8 and beyond. arXiv preprint arXiv:2304.00501 (2023).