Image Classification Based on Deep Learning

Hanyang Tan
FEIT, UTS, Sydney, Australia

Keywords: DDN, CNN, CIFAR 10.

Abstract: Image classification technology, as a core research direction in the field of computer vision, has become the focus of widespread attention among researchers with the development of deep learning technology. Although convolutional neural networks (CNN) have made revolutionary progress in image processing, there are still problems such as overfitting and the complexity of handling diverse data sets. This paper presents a hybrid model composed of a Convolutional Neural Network (CNN) module and a time-frequency composite weighting module. The CNN module effectively performs deep feature extraction, while the time-frequency composite weighting module is capable of achieving better performance. Through experimental verification on CIFAR 10, this paper demonstrates the excellent performance of the hybrid model on image classification tasks, with an accuracy of 90%. The results of this paper not only prove the effectiveness of combining different deep learning architectures to improve image classification accuracy, but also provide new ideas and methods for the development of future image processing technology.

1 INTRODUCTION

In recent years, machine learning-based data analysis methods have achieved notable results in tasks involving text, video, and audio. Image classification, a fundamental technique within data analysis, plays a crucial role in diverse applications spanning business, military, and everyday life scenarios. In the early stages of image classification research, the process required the design of manual features based on the characteristics of images, followed by classification using machine learning models. For instance, features such as color histograms and texture information were extracted and then classified using machine learning models like Support Vector Machines (SVMs) and Decision Trees. Traditional machine learning methods, characterized by a limited number of parameters and a heavy reliance on the results of manual feature extraction, significantly increased the difficulty of model optimization. Fortunately, the advent of deep learning has enabled the joint optimization of feature extraction and classification modules, representing a significant leap forward in the field.

Image preprocessing mainly includes image clipping, scaling and normalization, which ensures the consistency of input data. In the feature extraction stage, deep learning model is used to extract the deep features in the image, and the image is encoded as a feature vector. According to the input image features, the classification module predicts the probability distribution of the categories, which is usually processed using the softmax function.

In the field of computer vision, deep learning has emerged as a pivotal technology for advancing image classification techniques. This research aims to further enhance image classification performance by integrating Convolutional Neural Networks (CNNs) with Deep Decision Networks (DDNs). While CNNs have revolutionized image processing with their ability to autonomously extract hierarchical features, this paper introduces an innovative hybrid model designed to improve image classification accuracy. The hybrid model combines the powerful feature extraction capabilities of CNNs with the unique decision-making perspective offered by DDNs, aiming to create a more robust and adaptive image classification system.

Specifically, the model proposed in this paper is comprised of key components including convolutional layers, fully connected layers, activation functions, and a softmax layer. These components work in concert to enhance the model's ability to recognize various features in images, while

Tan, H.
Image Classification Based on Deep Learning.
DOI: 10.5220/0012835500004547

In *Proceedings of the 1st International Conference on Data Science and Engineering (ICDSE 2024)*, pages 313-317
ISBN: 978-989-758-690-3

the decision-making mechanism is optimized for the classification process. In the experimental section, this model was tested on the CIFAR-10 dataset, a standard benchmark in the field. The results demonstrate that the model achieved an accuracy rate of 90% in the image classification task, showcasing the effectiveness of our model.

2 RELATED WORK

Early image recognition relied on traditional feature descriptor design, which usually required manual feature extraction. These methods require domain expertise to design and extract relevant features of images, which greatly limits the efficiency of image extraction. In addition, because the feature extraction and classification models cannot perform parameter optimization at the same time, the classification effect is poor.

Deep learning has had a profound impact on image recognition. It has significantly improved the accuracy of recognition systems by automatically learning relevant features from raw image data, eliminating the need for handcrafted feature engineering (Murthy et al. 2016). These models are particularly effective in performing complex tasks by optimizing parameters across both feature extraction and downstream tasks simultaneously. Additionally, the deep network structures are capable of extracting high-level semantic features, which is fundamental in understanding and interpreting complex image content. Such capabilities have propelled deep learning to the forefront of advancing technologies in computer vision, enabling significant progress in object detection, localization, semantic segmentation, and image generation.

Venkatesh N. Murthy and Vivek Singh's (2013) research introduces Deep Decision Networks (DDNs) as a novel solution for image classification. DNNS mainly alleviates the problem of gradient disappearance or gradient explosion in deep networks through phased training. By merging the straightforward structure of decision trees with the capabilities of deep learning, DDNs use decision stumps at each node for initial classification and allocate specialized nodes for more complex scenarios. This strategy enhances the efficiency and accuracy in handling large, varied datasets, showcasing a significant advancement in the approach to image classification challenges.

3 METHDOLOGY

In paper proposes an innovative hybrid deep learning model that combines a convolutional neural network (CNN) and a deep decision network (DDN), with the aim of improving the performance of traditional models in image recognition accuracy. First, all emerging images were normalized and necessary preprocessed. Then, the normalized data is used to process it through DDN, and a new loss function is introduced to improve the classification accuracy. Then, CIFAR 10 was used for experimental verification. In aligning with the pioneering approaches for feature extraction, this study leverages the insights from Su (2015), who demonstrated the efficacy of multi-view convolutional neural networks in recognizing complex 3D shapes by extracting nuanced features that capture the essence of the objects from various angles (Ciresan et al. 2012). This principle of extracting deep, meaningful features forms the cornerstone of our methodology, where the CNN component of our hybrid model meticulously learns to identify intricate patterns within the CIFAR-10 dataset images. The ability of CNNs to discern and learn from the dataset's diversity not only underscores the adaptability of our model but also its potential to generalize across different image classification tasks, drawing from the foundation laid by the referenced work in enhancing model performance through sophisticated feature extraction techniques.

3.1 Normalizing Images

Normalization of images is an important preprocessing step. The main purpose is to convert image data into a more consistent range to facilitate the training of neural networks. Normalization usually involves two key steps: adjusting the mean and standard deviation of the data. Normalization refers to converting image data from the original pixel value range (usually 0 to 255) to a smaller range (such as -1 to 1 or 0 to 1) (Su et al. 2015). The main propose to normalizing is to improve the stability and convergence speed of model training. The method used in the paper is based on "Z-score standardization". Building upon the foundational work of Simonyan and Vedaldi (Simonyan & Vedaldi 2013), who emphasized the critical role of deep feature extraction in enhancing image classification models, this study adheres to a rigorous normalization process to ensure the consistency and reliability of input data for neural network training (Su et al. 2015).

3.2 Mixed Model

The methodology section introduces an innovative approach by combining Convolutional Neural Networks (CNNs) and Deep Decision Networks (DDNs) to analyze the CIFAR-10 dataset. The model employs CNN layers for robust feature extraction from images, leveraging their capability to autonomously learn and identify intricate patterns.

Next, the model simulates the process of a deep decision network (DDN) through dense layers, a step designed to make decisions based on features extracted by the CNN. Inspired by the groundbreaking work of Cireșan et al. and Zheng et al., who showcased the significant improvements in image classification accuracy through the use of multi-column deep neural networks, our research adopts a similar philosophy in enhancing the robustness and accuracy of our hybrid model (Zheng et al. 2021). This integration method not only takes advantage of the powerful capabilities of CNN in feature extraction, but also attempts to simulate the advantages of DDN in decision-making efficiency, thereby improving the classification accuracy of the model. This hybrid architecture aims to enhance classification accuracy by utilizing CNN's strength in feature extraction and approximating DDN's decision-making efficiency. The model is compiled and trained with categorical cross-entropy loss and Adam optimizer, evaluated to demonstrate its effectiveness in image classification tasks.

4 EXPERIMENT AND RESULT

4.1 Datasets

The CIFAR-10 database (Krizhevsky), developed by the Canadian Institute for Advanced Research, is a standard test set widely used in computer vision research. It contains 60,000 32x32 pixel color images divided into 10 categories with 6,000 images in each category. These categories include common objects such as airplanes, cars, birds, cats, deer, dogs, frogs, horses, boats, and trucks. Images in the database are carefully selected and annotated to ensure an even distribution of images within each category.

he CIFAR-10 images are filtered from the larger 80 million tiny images dataset, which contains about 80 million small images of 32x32 pixels. Each category in the CIFAR-10 dataset is filtered from this large dataset to ensure image quality and category balance. In addition, the diversity and realism of images in CIFAR-10 make it ideal for testing image processing algorithms, especially when dealing with common problems in real-world images, such as changing lighting conditions, different viewing angles, and background noise.

CIFAR-10 was originally designed to provide a benchmark testing platform for computer vision algorithms, especially for evaluating the performance of image recognition and classification algorithms. The dataset is divided into 50,000 training images and 10,000 test images to help researchers train and validate their models.

To ensure data diversity and practicality, CIFAR-10 images are collected from a variety of scenes and backgrounds, covering a variety of lighting conditions and postures. This database is not only highly respected in academia, but also widely used in industry, providing important data support for improving the accuracy and robustness of image processing technology.

The use of CIFAR-10 has greatly promoted the development of the field of computer vision, especially in the research of deep learning and convolutional neural networks. It provides researchers with a standardized platform to compare the effects of different algorithms and inspires innovation and progress in image recognition technology by researchers around the world.

4.2 Results

When evaluated on the CIFAR-10 dataset, a standard benchmark for computer vision, the model showed a significant classification accuracy of 90%. This performance metric emphasizes the model's ability to accurately process and classify images across different categories of the dataset. Our method integrates the convolutional neural network with the deep decision network, and makes full use of the feature extraction capability of CNN and the decision-making capability of DNN to improve the accuracy of image classification.

5 DISCUSSION

In the current study, we successfully developed a hybrid model combining convolutional neural networks (CNN) and deep decision networks (DDN) for image recognition tasks. This innovative attempt not only marks the advancement of the application of deep learning technology in the field of image processing, but also demonstrates the huge potential of cross-domain fusion technology. However, although our model demonstrates excellent performance on multiple datasets, there are still a series of challenges and opportunities to further improve model efficiency, accuracy, and interpretability. This chapter will discuss these challenges in depth and explore possible future research directions and technological improvement paths, to promote scientific research and technological innovation in this field.

As author contemplate the future trajectory and potential enhancements for our hybrid model, the integration of advanced rendering techniques and contrastive learning principles, exemplified by the works of Lassner and Zollhofer (2021) and Wang et al. (2019), respectively, presents a compelling avenue for innovation. The application of efficient sphere-based neural rendering can significantly enrich the visual representation and interpretability of images, while adopting contrastive learning strategies from the domain of long-tailed image classification promises to address data imbalance and improve classification accuracy across diverse datasets. Moving forward, the exploration of these methodologies, alongside the innovative strategies suggested in References (Hinton et al. 2015) and (Alzubaidi 2021), will be instrumental in overcoming the current limitations of our model. By harnessing these cutting-edge approaches, author aims to enhance the model's robustness, adaptability, and performance, ensuring its applicability to a broader spectrum of image classification challenges and setting a new benchmark for future research in the field. Secondly, this paper also exposed the interpretability shortcomings of deep learning models. Although the model performed well on the classification task, it was difficult to understand why the model made the classification decision it did. This lack of interpretability may limit the usefulness of the model in certain application scenarios, especially those that require a high degree of transparency and interpretability.

Besides, In the pursuit of enhancing the efficiency of our hybrid model, recent studies offer promising methodologies that could be directly applicable. For instance, leveraging advanced model compression techniques, as discussed by (Hinton et al. 2015), can significantly reduce the computational footprint of deep learning models without compromising their performance. This approach is critical for deploying sophisticated models in resource-constrained environments. Concurrently, the application of Neural Architecture Search (NAS) methodologies, exemplified in (Alzubaidi 2021), presents a strategic pathway to automatically discover optimal model architectures that balance accuracy with computational efficiency. Integrating these cutting-edge techniques promises not only to elevate the operational efficiency of our hybrid model but also to extend its applicability across a broader spectrum of real-world scenarios, where computational resources are often limited. Future iterations of our research will explore these avenues, aiming to harness the potential of (Hinton et al. 2015) and (Alzubaidi 2021) to surmount current efficiency constraints, thereby enhancing the model's viability for extensive deployment.

In this paper, author explored the application of deep learning technologies in image recognition by integrating Convolutional Neural Networks (CNN) and Deep Decision Networks (DDN). Recent literature demonstrates the immense potential of deep learning in handling complex tasks such as image recognition and image caption generation. Specifically, a review article (Hossain 2019) delves into the challenges of deep learning, such as data imbalance and model compression, as well as its applications in fields like medical imaging.

Through continuous research and technological innovation, we look forward to achieving broader and more profound impacts in the fields of deep learning and image recognition.

6 CONCLUSION

In this paper, we employ deep learning techniques for image classification, specifically an architecture that combines convolutional neural networks (CNN) and deep decision networks (DDN). The experimental results show that this hybrid model significantly improves the accuracy and performance of image recognition. However, in discussing these results, we also recognize some key challenges and limitations.

First, although this model performs well on the CIFAR-10 dataset, this does not mean that it can be effective on all types of image recognition tasks. For example, this model may have difficulty processing more complex or irregular image data sets. Therefore, future work may need to explore how to adapt and optimize the model so that it can better handle various types of image data.

Secondly, this paper also exposed the interpretability shortcomings of deep learning models. Although the model performed well on the classification task, it was difficult to understand why the model made the classification decision it did. This lack of interpretability may limit the usefulness of the model in certain application scenarios, especially those that require a high degree of transparency and interpretability.

Finally, this paper also raises the issue of data dependence of deep learning models. Although we used CIFAR-10, a widely used standard dataset, this also means that author's results depend heavily on this specific dataset. If the quality or representativeness of the data set is insufficient, this model may not achieve the same performance.

In summary, although author's research has achieved certain results in the field of image classification, there are still many challenges and issues that need to be addressed in future work. Through further research and improvement, we believe that the application of deep learning technology in image recognition and other fields will be more widespread and effective.

REFERENCES

V. Murthy, S. Maji, and R. Manmatha, "Deep Decision Network for Multi-Class Image Classification" in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2016), 2240-2248.

K. Simonyan, A. Vedaldi. Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps. arXiv preprint arXiv:1312.6034, (2013).

H. Su, S. Maji, E. Kalogerakis, and E. Learned-Miller, "Multi-view Convolutional Neural Networks for 3D Shape Recognition" in *Proceedings of the IEEE International Conference on Computer Vision* (2015) 945-953.

D. C. Cireșan, U. Meier, and J. Schmidhuber, Multi-column Deep Neural Networks for Image Classification, arXiv preprint arXiv:1202.2745.

Y. Zheng, J. Wu, Y. Qin, F. Zhang, and L. Cui, "Zero-Shot Instance Segmentation" in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2021), 2593-2602.

Krizhevsky, Learning Multiple Layers of Features from Tiny Images. Retrieved from https://www.cs.toronto.edu/~kriz/cifar.html.

P. Wang, K. Han, X.-S. Wei, L. Zhang, and L. Wang, "Contrastive Learning Based Hybrid Networks for Long-Tailed Image Classification" in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2021), 943-952.

M. Tan and Q. V. Le, EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks, arXiv preprint arXiv:1905.11946 (2019).

G. Hinton, O. Vinyals, and J. Dean, "Distilling the Knowledge in a Neural Network", in *Proceedings of the NIPS Deep Learning and Representation Learning Workshop* (2015).

Alzubaidi, Laith, Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *Journal of big Data,* 8, 1-74 (2021).

Hossain, MD Zakir, A comprehensive survey of deep learning for image captioning. *ACM Computing Surveys* 51(6). 1-36 (2019).