Analysis of The Role of Deep Learning Models in Image Classification Applications

Xiang Li

Shenzhen College of International Education, 518043 Shenzhen, China

Abstract. Image classification is a fundamental task in computer science, underpinning various applications such as object detection, face recognition, and object interaction analysis. The concept holds significant value due to its wide-ranging applications across multiple fields. Traditional methods for image classification, however, have been limited by their slow processing speed, rigidity, and high costs. The integration of deep learning models, particularly Convolutional Neural Networks (CNNs) and Support Vector Machines (SVMs), has revolutionized this process, enabling the development of automated, fast, and practical systems. These advanced models are now employed in diverse areas, including biomedical science, remote sensing, and business management, thanks to their ability to achieve high accuracy across a broad spectrum of scenarios. Training these models involves the use of well-known datasets like Canadian Institute for Advanced Research (CIFAR) and Modified National Institute of Standards and Technology (MNIST), which provide the necessary data for optimization and validation. The paper examines the structure, functionality, advantages, and limitations of CNNs and SVMs in the context of image classification, demonstrating that deep learning-driven classification is now a mainstream research focus. This study highlights the transformative impact of these models and provides insights into their future potential.

1 Introduction

Image classification, which is presented as the basic task of computer vision related to visual identification. It includes some predefined categories so that images are sent to correct classes according to its content [1, 2]. The concept of classifying is needed in various aspects and fields, and derived tasks are like object detection, pose recognition, and segmentation [2]. For humans, in the case of knowing the basic knowledge of some things, classification is done unconsciously in mind. (E.g. people will coordinate shops to different types when they are walking in a shopping mall.) However, simple and basic image classification is not that simple to computers or other systems [1], as machines need to deal with images with high diversity. The focus of the research is about the extraction of features in images, which means that computers are required to react like human actions such as understanding, analysing and judging [3]. The study about this aspect is called computer vision (CV). The mainstream of

Corresponding author: s23461.li@stu.scie.com.cn

image classification is combining the linear process of dealing with features with deep learning models so that programs are more flexible and accurate in various scenarios.

The image classification aided by computer has been applied in medical science, as computers have high potential to achieve the goal of accurate recognition [4]. Moreover, it is expected to work out concrete status of human body to reduce the risk of therapy. The whole matured process includes three steps, the first is information extraction, the second is filtration of detected features and the last is classification [4]. Technology also has a huge amount of potential value in other aspects like Internet applications [3]. With strong ability of analyzing data, precise recommendation of commodity can be achieved by computer, with high quality of image classification. In conclusion, when artificial intelligence (AI) combines with the basic task of computers, huge amounts of improvements can be made in almost all types of industry.

Different techniques of image classification are Computed Tomography (CT), Magnetic Resonance Imaging (MRI), which is in biomedical machines [4]. Internet of Things (IoT) that is tested in smart vehicles is linked with image classification in remote sensing (RS) to gain large scale of data when driving [5]. RS data usually requires higher efficiency and quality of dealing with features in images, especially in those shot by astronomical devices such as satellite and ground sensors with high-resolution [6]. In these cases, the speed of classification that needs high manpower cost was no longer acceptable, that is, classification still relied on human's aim. Moreover, the method of transforming data into different features regarding pictures such as texture was not that effective. Researchers have set the goal of automated and quantitative [6]. Recent computer-based methods included algorithms designed to analyse pictures in a more reliable way. Traditionally, it takes time to manually preset the image features for classification algorithms, which loses efficiency and portability. A recent improvement was a biological neural network called Artificial Neural Network (ANN) that simulated neuron activity in organisms [2]. Further research also focused on the simulation of neural networks; convolutional neural network (CNN) was a deep learning model with satisfactory accuracy. By processing data from images layer by layer, CNN model showed remarkable ability in competition with other classifiers [6].

Given the unique significance of image classification, this paper explores its various applications, outlines the working processes, and examines potential enhancements to existing methods. It also identifies key challenges, such as the limitations of certain deep learning models. The primary goal of this paper is to provide a comprehensive understanding of the current state of image classification. The paper begins by discussing the methodologies commonly employed in image classification, followed by an analysis of results from recent studies. Finally, it concludes with a summary of findings, highlighting both the advancements and the ongoing challenges in the field. By doing so, the paper aims to offer valuable insights for researchers and practitioners looking to deepen their knowledge of image classification technology and its future potential.

2 Methodology

The paper includes the concept of image classification, deep learning models that applied in it, some developed methods related to image classification around some fields, and results with conclusions. The paper mainly focuses on giving information so that it may be helpful for other researchers. The structure of the essay is shown in Fig. 1.

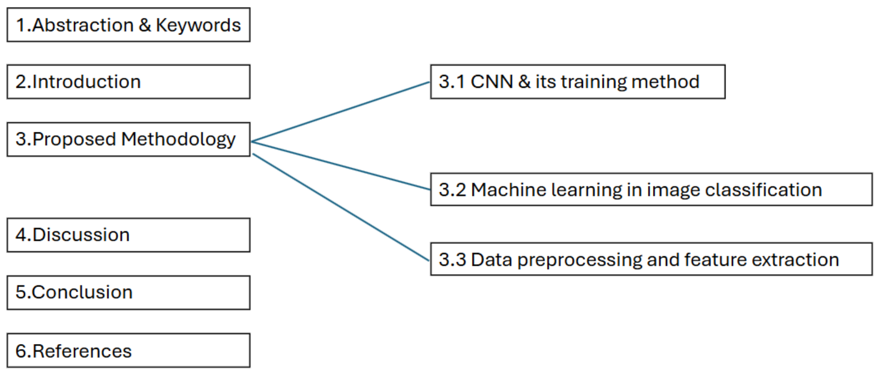

Fig. 1. The framework and structure (Picture credit: Original).

2.1 Dataset

For mentioned CNN and other classifiers, to check out the ability when doing some tasks like image classification, object detection and relation detection in images [7], datasets with diverse kinds of photos were used. Canadian Institute for Advanced Research (CIFAR) 10 and 100 were two datasets used to do so, with hundreds of training images and testing images that divided into different classes by predefined labels [8]. Both two datasets had 60,000 images with the size of 32*32, and the number at the end of the name indicated the number of classes the dataset had. (E.g. "10" indicates 10 classes.) [9,10] Modified National Institute of Standards and Technology (MNIST) dataset was another popular dataset used in CNN and deep convolutional neural network (DCNN), including 70,000 handwritten digit images [1]. The recent dataset introduced in [7] was Open Images V4, which advocated tests with nature, that is, images in dataset were not labelled aforehand, and the dataset integrated images, annotations and boxes in vast numbers.

2.2 Proposed approach

2.2.1 CNN structure and training

To build an artificial neural network, a large amount of information processing unit is needed. Take CNN as an example, CNNs are designed for input of two-dimensional data, and its name expresses the concept: CNN is an artificial neural network, with neurons on each layer [3]. As mentioned before, CNN had a better performance than other classifiers using shallow structure models. A main characteristic of CNN was the high acceptability of data. Convolutional operations helped CNN to deal with each pixel of an image, so CNN was available for many scenarios, even with some abstract cases [6]. Layers in CNN are convolutional, polling and fully connected layers, with another output layer. The important part is convolutional layer, as it acts like both classifier and feature generator. Matrices generated by convolutional process compose the layer. The task of pooling layer is to collect all features and filtrate out the key part of each image [8]. During the process of classifying, neurons are connected to each other on non-linear convolutional layers so that data in images can be extracted, and, in turn, images are classified [2]. The specific method of training CNN can be divided into four steps. First is to choose a dataset for training and validation information. Secondly, debugging and modifying CNN, making sure the fully connected layer has the same size as the testing data in the dataset. Third is to start training. In this step, it is necessary to set related options and factors such as learning rare and validation data to a suitable number or range. The last step is to calculate the accuracy of the network [8]. CNN has another important concept called parameter sharing. In convolutional network, convolutional kernel is used to extract information from images. Features collected by kernels are shared to other kernels so that the training process is much less time costly [11].

2.2.2 Machine learning applied in training of model

As the core of AI, machine learning is a subject that includes different disciplines such as probability theory and approximation theory [11]. The goal of machine learning is to simulate human behaviour of learning, and, in turn, understand laws by support of large amount of data through designed algorithms. The learning process can be multi-dimensional, as AI based on machine learning can receive data like texts, sounds, and images, which makes the combination between it and image classification become possible. In remote sensing area, support vector machine (SVM) is one of the classifiers built with machine learning. SVM is applied in image classification process in order to create a new learning method using strategy of margin sampling (MS). The new learning process was developed because the efficiency of training needed to be improved. The principle was to rank samples in training sets in an iterative process according to a specific criterion, so that the most useful samples were selected to help improve the model [9]. An active learning network can be built through this method. In datasets, there are two types of information: spatial and spectral. Respectively, two criterions are used to rank the significance of samples by sorting algorithms. However, remote sensing images often contain information in numerous classes, so a new strategy called one-against-all (OAA) has been introduced to meet the requirement.

2.2.3 Data preprocessing and feature extraction

Before the process of image, there is usually a preprocessing of data, including restoration of bad lines, image registration and geometric rectification. The last two operations are necessary when remotely sensed data need processing. Based on different types of images, different operations are required to help the later classification. For example, for multisensory data, atmospheric and radiometric correction is needed [11]. For the section of feature extraction, the classification system usually converts the image into two patterns: histogram of oriented gradients (HOG) and local binary patterns (LBP) to reflect different factors of the image. When the system executes some tasks like object detection and face recognition, two types of patterns are both feasible [12].

3 Result and Discussion

Undoubtedly, CNN has become a critical success in the orientation of combining machine learning and image classification, and for Bag of Words which manually deal with images, CNN made breakthroughs in many aspects such as efficiency, cost and flexibility. The most advantageous part of CNN was that the work did not require manual operation, the bottleneck of all past classifiers. All people need to do is to input images and wait for the result, as CNN was based on machine learning, the process was automated. Simulating neurons in organisms was the key to the success of CNN. After the emergence, CNN attracted attention not only from the engineering community, but also from famous enterprises like Microsoft. CNN was applied in recognition of Arabic handwriting and East Asian handwriting of different languages [1]. Although CNN was outstanding enough, there was still space of improvement and optimization. In CNN, convolutional layer uses linear filters, and this has caused that when facing abstract information extraction, CNN is not that capable. A possible improvement is to replace the original linear filters with perceptron that has a multi-layer structure. Moreover, a theory about CNN reinforcement is doubly convolutional neural network, which improves the accuracy of image classification, and at the same time, saves more storage space as the architecture of the network uses parameter efficiently [1].

SVM is a method introduced by Vapnik and the team, and it is one of the most popular supervised classification methods. Binary SVM uses linear process, and it divides data into two classes if the data is linearly divisible. SVM is composed of support vectors and a hyperplane, a plane structure that separates SVM into predefined categories. Hyperplane helps to reduce the incidence of errors and optimize the process of classification. As classes are divided physically, SVM cannot deal with non-linear data, which is a disadvantage. The solution given by the research team is to convert data so that the data is separable by using kernel trick method that related to different functions used in classifiers. Thus, by replacing the kernel of SVM, it is available in most circumstances [13]. Compared with CNN, SVM is also an outstanding classifier that can reach the goal of face detection, object and text recognition, and different identifications in various scenarios. One of the advantages of SVM is that the solutions provided by it contain a few samples, which means that SVM is suitable for the case that requires a large amount of data input [14].

4 Conclusion

This paper explores key aspects of image classification, with a particular focus on the applications and comparisons of advanced models and methods such as CNNs and SVMs. The primary objective is to clarify the progression of research in the field, providing readers with a solid understanding of current trends and mainstream approaches in image classification. The paper highlights how image classification has evolved from a labour-intensive task into an automated, efficient system that is now widely adopted across various industries. Looking ahead, the future of image classification research is promising. Many researchers are focusing on enhancing model flexibility and feasibility, with a growing interest in developing specialized AI systems for this purpose. The emergence of AI technologies like Chat Generative Pre-Trained Transformer (ChatGPT), which excels in learning from human data, suggests that the next frontier in image classification may involve creating models that combine advanced learning capabilities with higher accuracy. This could lead to more sophisticated and adaptive image classification systems, further revolutionizing industries that rely on this technology. Ultimately, the paper underscores the potential for AI-driven advancements to continue shaping the future of image classification.

References

1. W. Rawat & Z. Wang, Deep convolutional neural networks for image classification: A comprehensive review. Neural computation, 29(9), 2352-2449 (2017)
2. L. Chen, S. Li, Q. Bai, J. Yang, S. Jiang & Y. Miao, Review of image classification algorithms based on convolutional neural networks. Remote Sensing, 13(22), 4712 (2021)
3. A.A.M. Al-Saffar, H. Tao & M.A. Talab, Review of deep convolution neural network in image classification. In 2017 International conference on radar, antenna, microwave, electronics, and telecommunications, 26-31 (2017)

4. E. Miranda, M. Aryuni & E. Irwansyah, A survey of medical image classification techniques. International conference on information management and technology, 56-61 (2016)
5. M. Li, S. Zang, B. Zhang, S. Li & C. Wu, A review of remote sensing image classification techniques: The role of spatio-contextual information. European Journal of Remote Sensing, 47(1), 389-411 (2014)
6. J. Song, S. Gao, Y. Zhu & C. Ma, A survey of remote sensing image classification based on CNNs. Big earth data, 3(3), 232-254 (2019)
7. A. Kuznetsova, H. Rom, N. Alldrin, J. Uijlings, I. Krasin, J. Pont-Tuset & V. Ferrari, The open images dataset v4: Unified image classification, object detection, and visual relationship detection at scale. International journal of computer vision, 128(7), 1956-1981 (2020)
8. N. Sharma, V. Jain & A. Mishra, An analysis of convolutional neural networks for image classification. Procedia computer science, 132, 377-384 (2018)
9. E. Pasolli, F. Melgani, D. Tuia, F. Pacifici & W.J. Emery, SVM active learning approach for image classification using spatial information. IEEE Transactions on Geoscience and Remote Sensing, 52(4), 2217-2233 (2013)
10. H. Wu, Q. Liu & X. Liu, A review on deep learning approaches to image classification and object segmentation. Computers, Materials & Continua, 60(2) (2019)
11. P. Wang, E. Fan & P. Wang, Comparative analysis of image classification algorithms based on traditional machine learning and deep learning. Pattern recognition letters, 141, 61-67 (2021)
12. Y. Lin, F. Lv, S. Zhu, M. Yang, T. Cour, K. Yu & T. Huang, Large-scale image classification: Fast feature extraction and SVM training. In CVPR, 1689-1696 (2011)
13. A. Kaul & S. Raina, Support vector machine versus convolutional neural network for hyperspectral image classification: A systematic review. Concurrency and Computation: Practice and Experience, 34(15), e6945 (2022)
14. D.M. Abdullah & A.M. Abdulazeez, Machine learning applications based on SVM classification a review. Qubahan Academic Journal, 1(2), 81-90 (2021)