The Development and Technical Research and Analysis of Image Classification

Sihan Meng[a]
School of Software, Dalian Foreign Language University, Dalian, China

Keywords: Intricate Evolution, Sophisticated Techniques, Image Classification.

Abstract: This study endeavours to offer a comprehensive overview of the intricate evolution and sophisticated techniques utilized in image classification. Commencing with a meticulous examination of the developmental trajectory, it meticulously delineates the fundamental concepts and principles that underpin this dynamic field. Subsequently, it delves deep into an exhaustive analysis of pivotal models, meticulously dissecting their intricate mechanisms and shedding light on their nuanced functionalities. Image classification has been applied in various fields to improve various aspects. This paper analyses and studies the core network model technology and performance. Analysing the representational and generalization performance of key models through core analysis techniques Moreover, the study meticulously evaluates the performance metrics of these models, meticulously spotlighting significant technological advancements and breakthroughs. Furthermore, it critically examines the nuanced interplay of strengths, weaknesses, and future potentials inherent within these models, providing valuable insights for future research and development endeavours. Finally, the paper synthesizes and consolidates the entire image classification landscape, offering profound insights into emerging trends, paving the way for enhanced accuracy and efficacy in real-world applications.

1 INTRODUCTION

In the era of information explosion, the growth rate of image data is very fast, so the study of image classification is of great significance. Image classification is a kind of major research direction in the computer vision domain. It replaces human visual interpretation by using computer quantitative analysis of images, dividing each image or region in an image or multiple images into one of multiple categories. The improvement of image classification technology contributed to the growth of the computer vision. It mainly processes include image data preprocessing (Bhattacharyya, 2011) and the like. Computer vision tasks are based on image classification, such as localization, detection, and segmentation. The task is considered secondary to human nature, but it is much more challenging for automated systems. Scientists and practitioners have been working on developing advanced classification methods and techniques to improve classification accuracy. If the performance of the classification network is improved, its corresponding application level will also improve, for instance target detection, segmentation and so on.

From the 1960s to the 1970s, the image classification technology was still in its infancy. The research in this period is mainly focused on simple pattern recognition and feature extraction technology. In the 1980s, image processing technology took a huge leap forward. In the 1990s, the introduction of neural network brought a new perspective to the image classification technology. In particular, the convolutional neural network (CNN) model is suitable for handwritten digit recognition, proposed by LeCun et al. (Lecun, 1998) It has had a profound impact in the territory of image classification. On account of the lack of large-scale training data of the first CNN models-LeNet-5 (Lecun, 1998), as well as the limitations of the basic theory and computer computing capacity, LeNet-5 could not handle complex image recognition (Zhou, 2017) and only showed excellent performance in handwriting recognition tasks. However, it has also had a profound impact in the territory of image classification. CNN is a deep learning model

[a] https://orcid.org/0009-0007-3460-1620

Meng, S.
The Development and Technical Research and Analysis of Image Classification.
DOI: 10.5220/0012972500004508

In *Proceedings of the 1st International Conference on Engineering Management, Information Technology and Intelligence (EMITI 2024)*, pages 770-774
ISBN: 978-989-758-713-9

dedicated to process data with grid structure. In 1998, Yan LeCun et al. first applied convolutional neural networks to on the image classification task, with great success in the handwritten digit recognition task. Simonyan and Zisserman proposed the Visual Geometry Group (VGG) network structure in 2014, which is one of the most popular convolutional neural networks for now. It is welcomed by the majority of researchers because of its simple structure and strong application. While Christian Szegedy et al proposed GoogLeNet in 2014 and won the ImageNet competition in 2014.Scientists have proposed some landmark models, such as VGG, GoogLeNet, ResNet, EfficientNet and so on. Before 2020, CNN technology was used in the vast majority of image classification models, with relatively fixed network mechanisms, including basic modules such as convolution cores, residuals, pooling units, and linear layers. From 2017 to now, more and more models with excellent performance have appeared, but CNN still has irreplaceable advantages in image classification.

The primary aim of this study is to offer a comprehensive overview of the evolution and techniques employed in image classification. Initially, it delineates the developmental trajectory of image classification, elucidating fundamental concepts and principles. Subsequently, it delves into an analysis of pivotal models in image classification, dissecting their underlying mechanisms. Furthermore, the study evaluates the performance of these models, spotlighting key technological advancements. Additionally, it conducts a critical examination of the strengths, weaknesses, and future potentials of these models. Finally, the paper consolidates and synthesizes the entire image classification system, providing insights into future directions and prospects.

2 METHODOLOGIES

2.1 Dataset Description and Preprocessing

Currently, prominent datasets utilized for image classification include MNIST, CIFAR-10, ImageNet, COCO, Open Image, and YouTube-8M. MNIST, introduced by LeCun et al. (LeCun, 2012) in 1998, serves as a fundamental learning framework with 70,000 instances across different classes,60,000 training images and 10,000 testing images. CIFAR-10, crafted by Alex Krizhevskyp, features 10 classes (Krizhevsky, 2009). ImageNet, a project led by Professor Fei Fei Li's team from Stanford University, boasts over 14 million images (Deng, 2009), while the Microsoft COCO dataset comprises more than 300,000 images and over 2 million instances, useful for classification and recognition tasks (Lin, 2014). Open Images, a vast dataset provided by Google, offers about 9 million images annotated with labels and bounding boxes, with its fourth version being the largest to date. YouTube-8M, a massive video dataset, includes over 7 million labelled videos spanning 4,716 classes and 8 billion YouTube links, featuring training, validation, and test sets. These datasets serve diverse purposes, from training classifiers to object detection and relationship detection.

2.2 Proposed Approach

First, this study first introduces the development of the field of image classification, and then introduces several data sets commonly used for image classification: MNIST, CIFAR10, ImageNet, COCO, Open Image, Youtube-8M. Then it analysed the structure of its dataset, and the development of each dataset. Then it introduces several commonly used models in the territory of image classification: CNN, RNN and VGG, and analyses their main performance deeply. The pipeline of this study is shown in the Figure 1.

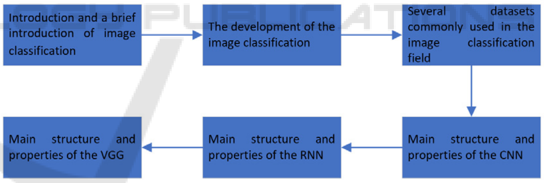
Introduction and a brief introduction of image classification
The development of the image classification
Several datasets commonly used in the image classification field
Main structure and properties of the CNN
Main structure and properties of the RNN
Main structure and properties of the VGG

Figure 1: The pipeline of study (Photo/Picture credit: Original).

2.2.1 CNN

Convolutional neural networks make a huge leap forward in image classification through supervised learning, while the CNN design roots in the structure of the visual system. CNN is short for cellular neural networks. It consists of a stack of alternating convolutional and pooling layers. In short, the CNN can be trained by the backpropagation error signal. In the past few years, the neural network (LeCun, 1990) has proven to be effective for simple recognition tasks. Convolutional layer is the most important component of CNN. On each convolution layer, the input cube is convolved with multiple learnable fit graphs to generate multiple feature maps. Currently, relevant workers have proposed several effective methods to help CNN to quickly converge and prevent overfitting, such as ReLU activation (Krizhevsky, 2012).

The advantageous performance of deep CNN usually derives from its deeper and broader architecture. The CNN outperforms many strategies (Henaff, 2015) in image analysis visibly. Overall, CNN is a useful tool for machine learning and many correlated domains, and the use of supervised learning combined with powerful CNN (LeCun, 2015) models is arguably a huge improvement in the field of image classification.

2.2.2 RNN

Although CNN is very popular in the computer vision domain, CNN cannot handle the data and sequences of complex structures. Sot this study proposes a new model the RNN. The full name of the RNN is the recursive neural network. RNN is a neural network of a specified type with a specific mathematical description that can be specialized to processing sequence information. For example, local image features, explicit detection, sliding window methods, or some recently proposed target detection methods, all employ different mathematical forms of attention mechanisms. RNN can not only operate on the input vector sequence, but also generate the output vector sequence. Generally, first input, through the hidden layer, final output. The nodes are unconnected from each layer. In practice, the RNN is able to conduct any length of the sequence data.

2.2.3 VGG

VGG is a kind of CNN model which is used extensively proposed by Kelencionyan and Andrew zi Selman. The VGG is short for the Visual Geometry Group at Oxford University. The VGG network explores the importance of improving the depth of the network to the final image recognition accuracy, while trying to use small convolution kernels to build deep convolutional networks in VGG. On the basis of the size and size of the convolution kernel, the VGG network can be divided into six configurations, among which the most famous configurations are two configurations: VGG 16 and VGG 19. The VGG 16 and VGG19 (LeCun, 2015) models are the accumulation of the convolutional layers. The VGG network (Simonyan, 2015) is constructed from extremely tiny convolutional filters. Studies show that the traditional VGG model can make the entire network to replicate the features of the front layer. Hosny et al. conducted the use of the several different experimental models (e. g. Alex-net, ResNet, VGG, and Google Net) and concluded that the VGG model required a lot of internal storage and high-allocation hardware. Although the whole network structure of VGG is relatively simple, the operation of large convolution is realized through overstocking small convolution kernel, deepening the structure of the network, and thus improving the overall identification effect.

3 RESULTS AND DISCUSSION

3.1 Comparative Analysis Performance

In the realm of third-generation neural networks, three commonly employed models are CNNs, recurrent neural networks (RNNs), and deep neural networks (DNNs). CNNs, structured for feature extraction. Convolutional layers extract spatial features like edges and angles from images, with subsequent pooling layers reducing feature dimensionality and computational load. Fully connected layers integrate extracted features for final output. Despite efficient training and spatial feature extraction, CNNs incur high memory consumption and computing constraints, particularly with large-scale data, as shown in the Table 1.

RNNs, built on recursive layers, comprise input, hidden, and output layers. Their recurrent nature allows the hidden layer to receive input not only from the previous layer but also from its own past states, enabling temporal feature extraction suitable for sequence data like text and time series. However, RNNs suffer from high computational complexity and low training efficiency due to their recurrent structure, and they struggle with adapting to varying input sequence lengths.

DNNs, the feature was by absolutely connected layers where each neuron links to every neuron in the preceding layer, excel in multilevel feature extraction. The model typically includes input, hidden, and output layers, with the hidden layer(s) playing a central role. While increasing hidden layers enhances data feature separation, excessive layers can lead to overfitting and prolonged training times. Notably, DNNs lack the ability to model temporal changes in time series data.

Table 1: The comparison of different models.

Model	Model Structure	Feature representation ability	Training efficiency	Model complexity	Robustness
CNN	Local links, and the convolution structure	Strong ability to extract local features.	The training efficiency is relatively high.	The model structure is relatively simple.	Robustness to data noise, deformation, etc
RNN	It has a cyclic structure	Suitable for the processing of the sequence data.	The training efficiency is relatively low.	The model structure is relatively complex.	Some robustness to data noise, deformation, etc
DNN	With a fully connected structure	handle the task of multilevel feature extraction.	The training efficiency is relatively low.	The model structure is relatively simple.	Robustness to data noise, deformation, etc

3.2 Analysis of Advantages and Disadvantages

The CNN adopts the local connection, which can extract many features in the data through the convolution kernel. In addition, CNN has good classification accuracy, and can also handle multiple images and other information for translation invariance, with good model generalization ability. However, CNN has strong dependence on hyperparameters, complex calculation process, and high requirements on data quality, which has some difficulty in practical application. In many image classification tasks, CNN is more sensitive to the label attributes, and the network performance will be affected. The RNN can enter the message at the last moment, and the input of any length can be processed, and the shape of its model does not change the shape with the increase of the input length. However, RNN is slow to calculate, hard to get information once upon a time, unable to think about the future input to the current state, and gradient vanishing and explosion problems are always easy to occur, as shown in the Table 2.

The DNN model is relatively simple, less time-consuming, and has high learning efficiency. However, because DNN adopts the form of full connection, it is undemanding to lead to overfitting, and also undemanding to fall into the local optima. Moreover, with the increase of the number of neural network layers, the gradient will decay, and the underlying gradient is basically 0, which is out of the question to model the changes on the time series. Moreover, with the increase of the number of neural network layers, the gradient will decay, and the underlying gradient is basically 0, which is out of the question to model the changes on the time series.

Table 2: The advantages and disadvantages of models.

Model	Advantages	Disadvantages
CNN	It can extract many features, can be translation invariant processing, has good model generalization ability and reduce the computation of the model.	Too strong dependence on hyper parameters, sensitive label properties, complex computation process, and high data quality requirements.
RNN	It can remember the last input information and process any length of the input,And the shape generally does not change.	Easy to occur gradient disappearance and explosion phenomenon, the calculation speed is slow, it is hard to gain the information for ages.
DNN	Simple structure, high learning efficiency, short time-consuming	It is undemanding to overfit and fall into local optima and cannot perform complex nonlinear structure.

3.3 Application Prospects and Future Prospects

Image classification can be applied in many fields. Image classification can be managed and classified by image library. By labelling or classifying images, it is convenient for users to retrieve and manage large-scale images and automatically identify and classify commodity images, which is widely used in libraries, museums and other scenes. Image classification can also be used in e-commerce, social media and other fields. Through visual search engines, people can input images to find similar or related images to facilitate people to find interesting content and goods. Image classification cannot be used for visual search engines, the management and indexing of image libraries. This is of great significance in libraries, museums and other scenes.

The application of image classification in the medical field is also significantly prominent, and image classification can assist in diagnosis, can help doctors to automatically identify and locate diseases to improve the early diagnosis and treatment effect of diseases. Not only does the computer understanding and interpretation of images need to be solved through image classification, but also the development and progress of various fields need to be promoted by image classification. Deep learning in the field has made significant progress of image classification, the future will through continuously research and innovation, research more efficient algorithm and hardware architecture to improve the computing efficiency of the model, further improve the image classification performance and robustness, to solve the effect of dealing with complex scenarios and data imbalance and vulnerabilities, image classification will usher in better performance and more widely used.

4 CONCLUSIONS

In conclusion, this study has delved into several foundational models crucial for image classification applications, meticulously scrutinizing their performance metrics and delineating their respective strengths and weaknesses. The dominance of the CNN model in the domain of image classification has been reaffirmed once again. Looking forward, the research trajectory of this study is poised to shift towards the exploration of the RNN model in the ensuing research phase. Emphasis will be placed on augmenting its capability to forecast forthcoming sequential data, thereby propelling the advancement of neural network models in processing sequential information. This pursuit holds the promise of further refining image classification methodologies, paving the way for enhanced accuracy and efficacy in real-world applications.

REFERENCES

Bhattacharyya, S. A., 2011. Brief Survey of Color Image Preprocessing and Segmentation Techniques. J. Pattern Recognit. Res. vol, 1, pp: 120–129.

Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. 2009. Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition, pp. 248-255.

Henaff, M., Bruna, J., and LeCun, Y. 2015. Deep convolutional networks on graph-structured data. arXiv:1506.05163.

Krizhevsky, A., and Hinton, G. 2009. Learning Multiple Layers of Features from Tiny Images. Technical Report, University of Toronto, vol, 1(4), p:7.

Krizhevsky, A., Sutskever, I., and Hinton, G. 2012. ImageNet Classification with Deep Convolutional Neural Networks. Advances in Neural Information Processing Systems.

LeCun, Y., Bengio, Y., and Hinton, G. 2015. Deep learning. Nature, vol. 521(7553), pp: 436–444.

Lecun, Y.; Bottou, L. Bengio, Y.; Haffner, P. 1998. Gradient-Based Learning Applied to Document Recognition. Proc. vol, 86, pp: 2278–2324.

LeCun, Y.; Cortes, C.; Burges, C.J.C. 2012. The MNIST Database of Handwritten Digits. 2012.

LeCun, Y. Denker, J., Henderson, D., Howard, R., Hubbard, W., and Jackel, L. 1990. Handwritten Digit Recognition with a Back-Propagation Network. Advances in Neural Information Processing Systems.

Lin, T. Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P. et al. 2014. Microsoft COCO: common objects in context. European Conference on Computer Vision, pp. 740-755.

Simonyan, K., and Zisserman, A. 2015. Deep Convolutional Networks for Large-Scale Image Recognition. ICLR, pp: 1-14.

Zhou, J.; Zhao, Y. 2017. Application of convolution neural network in image classification and object detection. Comput. Eng. Appl. vol, 53, pp: 34–41.