Comparative Study of Deep Learning Neural Networks for Image Classification

Lanqin Huang
College of Electrical Engineering, Sichuan University, Chengdu, China
2020141500192@stu.scu.edu.cn

Abstract. Image classification plays a crucial role in image recognition in present times. Researchers have developed numerous methods for achieving image classification through long-term research. Deep Learning, as one of the most important methodologies, is used to clarify the contents of an image. This paper provides an overview of four typical deep learning neural networks used for image classification, including conventional neural networks (CNN), recurrent neural networks (RNN), generative adversarial networks (GAN), and deep neural networks (DNN). The architecture, function, and applications of each network are discussed and compared. CNN is commonly used for image recognition, while RNN is suitable for speech recognition, natural language processing, and time series forecasting. DNN is ideal for handling high-dimensional data, and GAN can generate new data samples. The StyleGAN is introduced as an application of GAN, which can produce high-quality images. Finally, the future work and challenges in image recognition are discussed.

Keywords: Image classification, Conventional neural networks, Recurrent neural networks, Generative adversarial networks, Deep neural networks.

1. Introduction

Deep learning has become a popular methodology and is extensively used in various applications like optical character recognition and image classification. Hence, studying and reviewing deep learning is important and necessary.

Deep learning utilizes non-linear modules to transform representations from one level to a higher, more abstract level. By combining multiple levels of representation, it becomes possible to learn complex functions [1]. Image classification algorithms are utilized to determine the category to which an input image belongs [2]. The accuracy of image classification can be enhanced by utilizing deep learning, which involves creating machine learning models with multiple hidden layers and training them with extensive amounts of data. A common strategy to realize image classification is to use neural networks, such as conventional neural networks, recurrent neural networks, generative adversarial networks, and deep neural networks. It is likely that deep learning will continue to develop increasingly sophisticated models to explain its exceptional performance. Moreover, the unsupervised learning capabilities of deep learning are expected to improve further as the world generates an ever-increasing amount of data, providing ample opportunities for training and refining these models [3].

This paper reviews several typical neural networks used in image classification: conventional neural networks (CNN), recurrent neural networks (RNN), generative adversarial networks (GAN), and deep neural networks (DNN). Meanwhile, a few models are introduced and compared. Finally, the future work and challenges in the Image recognition field are prospected.

2. Neural Networks

2.1. Conventional Neural Networks

Conventional neural networks are an effective model in the field of deep learning. CNNs are a type of artificial neural network. The typical architecture of a convolutional neural network consists of three main layers: the convolutional layer, the pooling layer, and the fully connected layer. Fig. 1 illustrates the structure of a CNN. The pooling layer and convolutional layer are two distinct types of layers found in the CNN [4]. The primary role of the convolutional layer is to capture distinct characteristics in the input data, such as edges, shapes, and colors, using filters. The pooling layer reduces computation and optimization effort through down sampling, while retaining important information. The fully connected layer is responsible for taking the feature vectors extracted by the convolutional and pooling layers and feeding them into a classifier or regressor to produce the ultimate output.

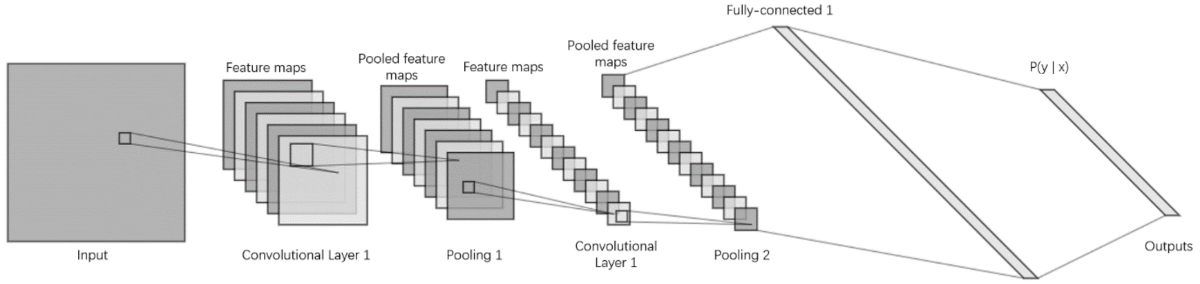

Figure 1. Standard convolutional neural network architecture

2.2. Recurrent Neural Networks

RNN is an artificial neural network that is designed to process sequential or time series data, such as language, speech, and images. In contrast to feed-forward networks like CNNs, recurrent neural networks have recurrent connections, where the previous hidden state is used as an input for the current state [5]. The basic component of an RNN is a cyclic unit that takes input from the current time step and the previous time step's state, generating the current time step's state and output. Several time steps can be unfolded to train and infer RNNs. The framework of a recurrent neural network is illustrated in Fig. 2.

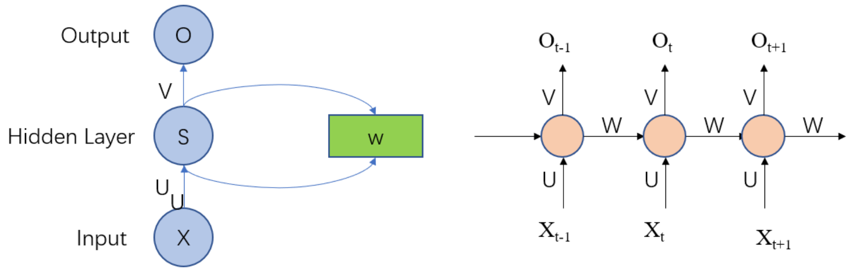

Figure 2. Framework of a recurrent neural networks

2.3. Generative Adversarial Networks

GANs typically consist of two neural networks: a generator and a discriminator. The primary structure of a GAN model is depicted in Fig. 3. The objective of the generator is to produce synthetic data that appears genuine. The primary goal of the discriminator is to distinguish between actual and generated input data. The two networks compete with each other to continuously improve their performance, ultimately reaching an equilibrium point that makes it difficult for the discriminator to identify the data generated by the generator. By training them in an adversarial manner, it is possible to construct a generator that can convincingly produce a real image from a random noise input and a discriminator that performs better at differentiating between real and fake images [6].

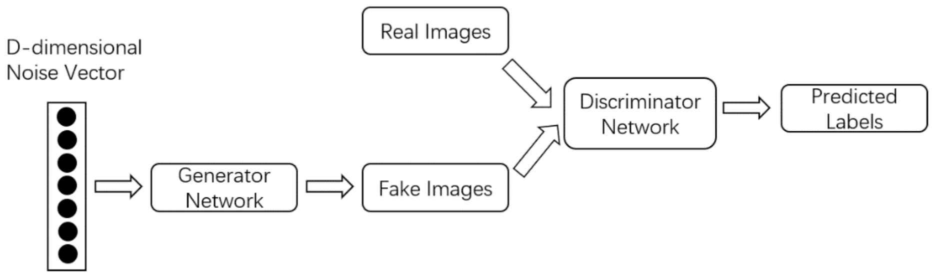

Figure 3. The main structure of GAN model

2.4. Deep Neural Networks

DNNs are composed of multiple layers of neurons, including input, hidden, and output layers. Each layer contains numerous neurons, which are interconnected with weights and biases, forming a complex nonlinear mapping function. A typical structure of DNN model is shown in Fig. 4. DNNs are capable of handling high-dimensional, complex data, such as images, speech, and text.

The input layer receives raw data, such as pixel values in images, waveform values in speech, or word vectors in text. The hidden layers apply nonlinear transformations to the input data to extract higher-level features, such as image edges, phonemes in speech, or semantics in text. Multiple hidden layers can learn different levels of features. The output layer uses the characteristics learned by the hidden layers to produce the final output, such as classification labels, regression values, or generated samples. The output layer typically employs activation functions, such as softmax or sigmoid, to map output values between 0 and 1.

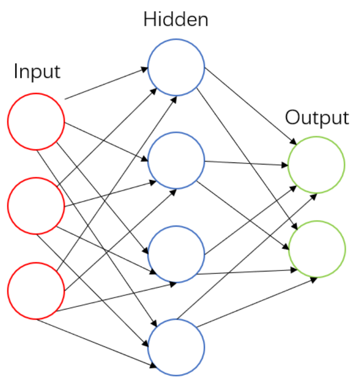

Figure 4. A typical structure about deep neural networks

2.5. Comparison

CNN, RNN, DNN, and GAN are four distinct deep learning algorithms, each with unique features and practical applications.

CNN uses convolutional and pooling layers to extract local image features, making it suitable for tasks such as image recognition, object detection, and facial recognition.

RNN processes sequence data using recurrent units, making it capable of capturing temporal dependencies, and is well-suited for tasks such as speech recognition, natural language processing, and time series forecasting.

DNN uses multiple fully connected layers to create complex nonlinear maps that can handle high-dimensional and complicated data for tasks such as classification, regression, and generation.

GAN utilizes two opposing neural networks to generate new data samples, allowing it to achieve tasks such as image conversion, image repair, and image synthesis.

3. Applications

3.1. StyleGAN

StyleGAN is a type of generative adversarial network with a distinctive architecture consisting of three key components: a mapping network, style modulation, and a synthesis network [7]. The mapping network converts a random latent vector z into an intermediate latent space w, which provides finer control over the image style. Style modulation is achieved by using w to parameterize the adaptive instance normalization (AdaIN) layer of each layer in the synthesis network, allowing for greater control over image details. The synthesis network is a progressively growing convolutional neural network that generates images from low to high resolution while also adding noise at each layer to increase diversity. The style-based generator is shown in Fig. 5.

StyleGAN2 was developed to address a flaw in StyleGAN, where a small number of generated images had obvious water droplets that were also present on the feature map. StyleGAN2 improves and solves this issue.

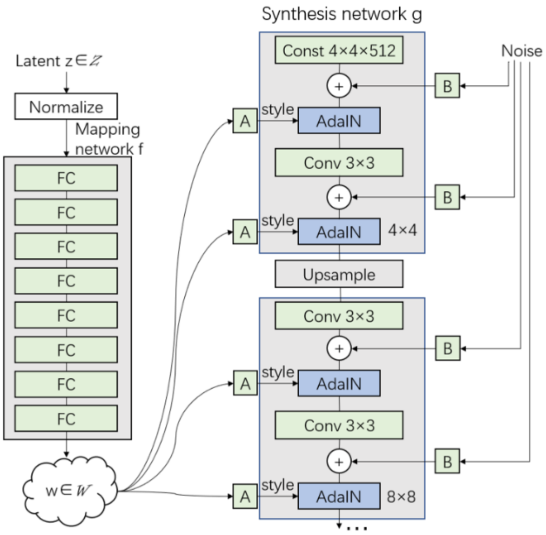

Figure 5. Style-based generator

3.2. IndRNN

Independently recurrent neural network (IndRNN) is characterized by neurons in a layer that function independently of each other [5]. IndRNNs can be stacked to form a deep neural network, which can be further enhanced by adding residual connections between layers to improve training robustness [5]. The independence of neurons within each layer also allows for a more straightforward interpretation of the neural network's behavior [4]. The basic IndRNN architecture is shown in Fig. 6.

IndRNN is a flexible neural network that can handle various types of sequential data, including speech recognition, human activity recognition, and image classification. IndRNNs are designed to handle input sequences of different lengths and can capture long-term dependencies over time. They can also be combined with other neural network models, such as convolutional neural networks and attention mechanisms, to improve overall model performance.

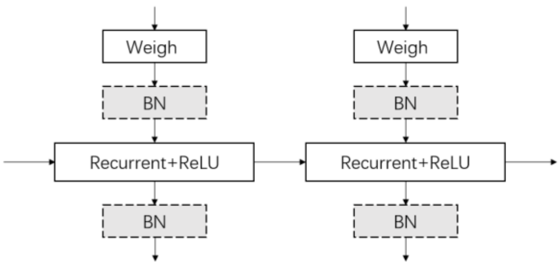

Figure 6. Basic IndRNN architecture

3.3. Identify Tables in Document Images

A deep neural network can be used to identify tables in document images. Table comprehension involves a three-step process: detecting table boundaries, specifying the table's structure via row and column analysis, and extracting information from table cells for a thorough understanding [8]. Comparing traditional methods and deep learning methods, we can see that traditional methods require manual feature and rule design, while deep learning methods can learn both automatically. Traditional methods work better for structured tables, while deep learning methods are better for unstructured tables. Traditional methods may struggle with noise, rotation, and distortion, while deep learning methods are more robust and can generalize better. Therefore, using deep learning in this field has significant implications.

3.4. Handwritten Digit Recognition

Handwritten digit recognition is an important pattern recognition technique used to process certain data types, such as financial statements, postal codes, and various invoices. It is a subset of character recognition technology [9]. Neural networks can be used to efficiently solve handwritten digit recognition. For example, a CNN model can be developed that includes two convolutional layers, two pooling layers, a fully connected layer, and an output layer. Each convolutional layer uses the ReLU activation function and Dropout regularization technique. The output layer implements the softmax activation function and the cross-entropy loss function [10]. By using the convolutional neural network model, the accuracy of handwritten digit recognition tasks can reach 98.8%, surpassing some traditional methods.

4. Future Prospect

Neural networks have been one of the most significant developments in the field of image recognition. They have shown tremendous success in a variety of tasks, including image classification, object detection, and segmentation. With the continued advancements in machine learning and computer vision, the future of neural networks in image recognition looks very promising.

In recent years, deep learning has become a prominent research area that entails utilizing neural networks with multiple hidden layers. These deep neural networks have demonstrated substantial advancements in image recognition activities, particularly in object recognition.

Despite their usefulness in image recognition, Neural Networks confront various challenges such as overfitting, insufficient data, and limited computational resources. Overfitting arises when the model performs exceptionally well on the training set but poorly on the test set or new data. This phenomenon can result from the model's complexity or lack of diversity in the training data. Inadequate data refers to the difficulty in obtaining a sufficient amount of labeled data required to train a neural network, which can lead to the model not learning generalizable features or being sensitive to noise and outliers. Limited computing resources can also hinder the training and inference of neural networks due to the high demand for computing power and memory space, which can affect the model's performance and restrict its deployment to platforms with limited hardware and cost constraints, such as mobile devices.

In the future, we can expect further improvements in the accuracy and efficiency of image recognition systems based on neural networks. Researchers are exploring new architectures and training methods that can further optimize these networks for image recognition tasks.

Overall, it is clear that neural networks will continue to play a significant role in the field of image recognition, and we can expect continued progress and advancements in this area in the years to come.

5. Conclusion

Deep learning is a continuously evolving and useful methodology in the field of image recognition. Neural networks are highly effective tools for completing various tasks in this field. Researchers can choose from a variety of options, such as CNNs, RNNs, or GANs, to achieve their objectives. However, numerous challenges remain in this field, and it is critical to continue to work diligently to address them.

References

[1] LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. Nature 521, 2015, 436–444.

[2] L. Peng, B. Qiang and J. Wu, A Survey: Image Classification Models Based on Convolutional Neural Networks, 2022 14th International Conference on Computer Research and Development (ICCRD), Shenzhen, China, 2022, 291-298.

[3] X. Du, Y. Cai, S. Wang and L. Zhang, Overview of deep learning, 2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC), Wuhan, China, 2016, 159-164.

[4] Z. He, Deep Learning in Image Classification: A Survey Report, 2020 2nd International Conference on Information Technology and Computer Application (ITCA), Guangzhou, China, 2020, 174-177.

[5] S. Li, W. Li, C. Cook, C. Zhu and Y. Gao, Independently Recurrent Neural Network (IndRNN): Building A Longer and Deeper RNN, 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018, 5457-5466.

[6] H. Meng and F. Guo, Image Classification and Generation Based on GAN Model, 2021 3rd International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI), Taiyuan, China, 2021, 180-183.

[7] T. Karras, S. Laine and T. Aila, A Style-Based Generator Architecture for Generative Adversarial Networks, 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 2019, 4396-4405.

[8] K. A. Hashmi, M. Liwicki, D. Stricker, M. A. Afzal, M. A. Afzal and M. Z. Afzal, Current Status and Performance Analysis of Table Recognition in Document Images with Deep Neural Networks, in IEEE Access, 2021, 9: 87663-87685.

[9] B. Charbuty and A. Abdulazeez, Classification based on decision tree algorithm for machine learning, Journal of Applied Science and Technology Trends, 2021, 2(1): 20–28.

[10] S. S. Rajput and Y. Choi, Handwritten Digit Recognition using Convolution Neural Networks, 2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 2022, 0163-0168.