International Journal for Housing Science and Its Applications
Publish August 10, 2025. Volume 46, Issue 4 Pages 6132-6141

https://doi.org/10.70517/ijhsa464449

Image Data Acquisition and Deep Learning and Data OCR Recognition Algorithms

Yubao Zhang[1,*]
[1] School of Design and Communication, Zhejiang Fashion Institute of Technology, Ningbo, Zhejiang, 315211, China
Corresponding authors: (e-mail: zbaob@163.com).

Abstract With the rapid development of information technology, the collection and analysis of image data play an increasingly important role in various fields. Deep learning (DL) and Optical Character Recognition (OCR) algorithms, as cutting-edge technologies in artificial intelligence and machine learning, have greatly promoted the progress of image data processing. In order to further understand the performance of different DL models in data OCR recognition image data acquisition, this article iteratively trains different models on the same dataset (COCO-Text dataset) and collects OCR image data. Finally, different models can be analyzed, and the accuracy, precision rate, recall, and F1 scores of GAN (Generative Adversarial Network) are 0.94, 0.93, 0.92, and 0.93, respectively. The analysis shows that when the number of iterations is sufficient, GAN has better OCR image data acquisition performance than other deep learning models; when the number of iterations is insufficient, the OCR image data acquisition performance of GAN decreases significantly. When the number of iterations is sufficient, CNN has better OCR image data acquisition performance; when the number of iterations is insufficient, CNN can still maintain good OCR image data acquisition performance.

Index Terms Deep Learning, Optical Character Recognition, Image Acquisition, Image Analysis, Generative Adversarial Network

I. Introduction

Image data acquisition is a key step in the field of deep learning and computer vision. High-quality and diverse image data is acquired through a variety of equipment and technical means to support subsequent algorithm training and model optimization. As technology continues to advance, image data acquisition is also improving [1]. Deep learning is an important branch of machine learning. It simulates the working mechanism of the human brain through multi-layer neural networks, automatically learns features from large amounts of data, and performs prediction and classification [2]. It consists of multiple layers, including an input layer, multiple hidden layers, and an output layer. Each layer is connected by weights to extract and transform data features layer by layer. The basic principle of deep learning is to pass the input data layer by layer through forward propagation, calculate the gap between the predicted result and the true value through the loss function, and then calculate the gradient and update the network weight through back propagation to make the loss function gradually smaller [3].

Common deep learning models include CNN, RNN, GAN, etc. As an important technology in the field of artificial intelligence, deep learning is constantly promoting technological progress in various industries [4]. With the improvement of algorithms and the increase of computing power, deep learning has a wider application prospect in the future [5]. Facing the challenges of data and computing, it can continue to explore and innovate to achieve more efficient and intelligent deep learning models, bringing more changes and progress to all areas of society. The data OCR (Optical Character Recognition) recognition algorithm converts the text information in the image into machine-readable text [6]. Traditional OCR technology mainly relies on image processing and pattern recognition, which is divided into steps such as character segmentation, feature extraction and classification. Traditional OCR works well when processing formatted text, but performs poorly when processing complex backgrounds and irregular text [7]. The emergence of deep learning has brought new directions to OCR technology [8]. Through deep learning models such as CNN and recurrent neural networks, the OCR algorithm can automatically learn complex features in images, achieve end-to-end character recognition, and greatly improve recognition accuracy and robustness.

This article mainly discusses the image data collection and analysis based on deep learning and data OCR recognition algorithm, and studies the whole process from image data acquisition, preprocessing to deep learning model construction and optimization. It studies the image data collection and analysis based on deep learning and data OCR recognition algorithm, with basic principles, model structure, training methods and performance evaluation as the main research contents [9]. Through relevant experiments, the effects of deep learning and image data collection of data OCR recognition algorithms are analyzed [10]. This article looks forward to the future development trend of OCR recognition image data collection and deep learning, and provide more powerful technical support and effective solutions for all walks of life. Through the research and discussion in this article, the article hopes to provide readers with some understanding of image data acquisition and data OCR recognition, stimulate more research interests and innovative ideas in this field, and contribute to the development of image data processing technology.

II. Related Work

In the past, image data collection mainly relied on traditional film cameras and early digital cameras. Although these devices could provide a certain quality of image data, due to technological limitations, the resolution, clarity, and collection efficiency of images were relatively low. With the advancement of digital technology, especially the popularity of digital cameras, smartphones, and video cameras, the efficiency and quality of image data collection have been significantly improved. There are various types of modern image data acquisition devices, including high-resolution SLR cameras, smartphones with powerful camera capabilities, high-definition cameras carried by drones, 3D scanners, and satellite imaging devices used for Earth observation. These devices can not only collect image data in various environments, but also meet the needs of different application scenarios, such as indoor and outdoor monitoring, aerial inspection, industrial detection, and remote sensing. In the future, real-time data acquisition can become a trend. By combining 5G technology and edge computing, real-time image data acquisition and processing can be realized. In addition, intelligent data collection can further improve the efficiency and coverage of data collection, achieving efficient and intelligent data collection through the use of automated devices such as robots and drones. Multimodal data fusion can also be an important direction in the future, integrating image data with other types of data (such as text, audio, sensor data) to provide more comprehensive and in-depth analysis.

DL has a wide range of applications in various fields [11]. In computer vision, DL is used for tasks such as image classification, object detection, face recognition, and image segmentation, significantly improving the accuracy and efficiency of image processing [12]. In natural language processing, DL is used for machine translation, sentiment analysis, text generation, and language modeling, and can handle complex language data [13]. In speech recognition, DL achieves efficient conversion from speech to text and is applied to speech assistants and speech recognition systems. In medical diagnosis, DL assists doctors in disease diagnosis and prediction by analyzing medical images. In the field of financial services, DL is used for risk assessment, fraud detection, and algorithmic trading, improving the precision rate and speed of financial data analysis. In automatic driving, DL is used for environment awareness, path planning and behavior prediction, which improves the safety and reliability of the auto drive system [14].

The data OCR recognition algorithm has a wide range of application scenarios in practical applications [15], as shown in Figure 1. For example, in document management systems, OCR technology can automatically recognize and extract text information from documents, achieving electronic and automated management of documents [16]. In bill processing systems, OCR technology can efficiently recognize textual information on bills, perform automatic classification and data entry [17]. In intelligent customer service systems, OCR technology can recognize and process image information uploaded by users, providing intelligent services and support [18].

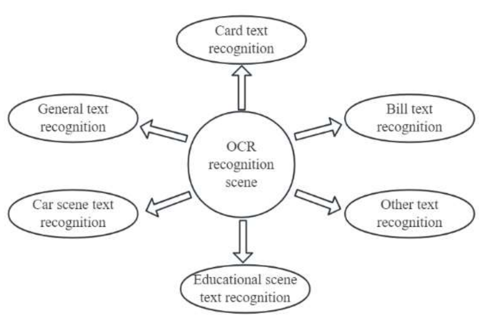


Figure 1: OCR recognition scenario

III. Selection of DL models

In image data acquisition and data OCR recognition tasks, the selection of DL models directly affects the performance and accuracy of the system [19]. DL automatically extracts advanced features from images through multi-layer neural networks, no longer relying on manually designed feature extraction methods, greatly improving the efficiency and accuracy of feature extraction [20]. Deep learning can process complex, high-dimensional image data, recognize and analyze subtle differences and complex patterns in images, and improve the quality of image data acquisition [21]. It can complete tasks such as image classification, target detection, image generation and image segmentation, greatly enhancing the ability to analyze and process image data; it can analyze the collected image data through deep learning models, identify deficiencies and deviations in the data, optimize data collection, ensure data diversity and representativeness, and improve the generalization ability of the model.

III. A. CNN

CNN is a deep learning model for image data processing. It extracts local features of the image through the convolution layer, reduces the feature map size through the pooling layer sampling, and combines all features through the fully connected layer for final classification. The CNN hierarchical structure can gradually extract and combine image features. With the increase of network layers and parameter optimization, it can continuously learn more complex and abstract features, greatly improving the processing performance and generalization ability of image recognition and classification, as shown in Figure 2.

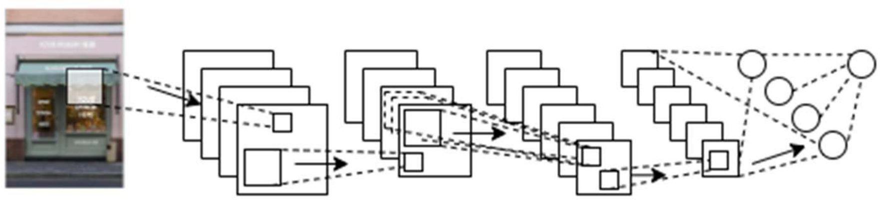


Figure 2: CNN Neural Network

III. B. Recurrent Neural Network

Recurrent Neural Network (RNN) is a neural network that processes sequence data. It can process input sequences through internal states (memory) and is widely used in natural language processing, time series prediction, and speech recognition. The RNN hidden layer receives the input of the current time step and the hidden state of the previous time step, which can capture the dependencies in the time series and is suitable for processing language models, time series prediction, and machine translation.

III. C. GAN

GAN is a deep learning model consisting of a generator and a discriminator, which generates realistic data samples through adversarial training. The generator is responsible for synthesizing data from random noise vectors, while the discriminator is used to distinguish between generated data and real data. The generator and the discriminator compete with each other. The generator strives to generate realistic data to deceive the discriminator, while the discriminator tries its best to distinguish between generated data and real data. Through adversarial training, the model is continuously optimized until the data generated by the generator cannot be accurately distinguished by the discriminator. Using different deep learning models can effectively strengthen the data OCR recognition algorithm and improve the accuracy and robustness in different application scenarios [22].

IV. Data Collection

IV. A. Selecting Data Sources and Collection Methods

Before data collection begins, it is necessary to clarify the purpose of data collection, including determining the data application scenario, data type, resolution requirements, quantity requirements, etc. [23]. Optical Character Recognition (OCR) is a technology that converts text in an image into computer-editable text. For the research of OCR algorithm, it is necessary to collect text image data of various fonts, sizes, languages and scenes to ensure the generalization ability and accuracy of the model [24]. Choosing the appropriate device in data collection is a crucial step [25]. According to project requirements, combined with practicality and universality, smartphones can be selected as data collection devices.

Based on the device and application location, this article chooses to use a combination of static and dynamic capture methods for data collection, using smartphones to capture images in fixed positions to ensure image clarity and stability [26]. Static texts such as books, documents, and billboards can be collected by taking photos [27]. Keyframe images in videos can be extracted by recording videos, and text in dynamic scenes such as traffic signs and mobile advertisements can be collected [28]. The basic principle of OCR recognition algorithm is to extract text regions in the image through image processing technology, and then convert these regions into corresponding text information [29].

IV. B. Data Preprocessing

Image data preprocessing is a crucial step to improve image quality, standardize data formats, remove noise, and enhance data diversity, laying a solid foundation for subsequent algorithm processing and model training. Image quality checks can be used to evaluate whether input data meets expected standards. Sharpness detection is one of the commonly used methods, which uses Laplace transform to detect the sharpness of an image and evaluate whether the image is clear enough, as shown in formula (1). The degree of blurriness can be evaluated by the variance of the image after Laplace transform. If the variance is small, the image may be blurry and require further processing, as shown in formula (2).

$$L(x,y)=\frac{\partial^2 I(x,y)}{\partial x^2}+\frac{\partial^2 I(x,y)}{\partial y^2} \tag{1}$$

$$\sigma^2=\frac{1}{MN}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1}(L(x,y)-\mu)^2 \tag{2}$$

Among them, M and N are the width and height of the image, respectively, and μ is the mean of the image after Laplace transform.

IV. B. 1) Brightness and Contrast Adjustment

Brightness and contrast adjustment is another important step in preprocessing. Brightness adjustment increases or decreases the overall brightness of an image through addition operations, as shown in formula (3). Contrast adjustment enhances or reduces the contrast of the image through linear transformation, as shown in formula (4). Typically, when the adjustment coefficient is greater than 1, the contrast is enhanced, and when it is lower than 1, the contrast is reduced.

$$\mathrm{I}'(\mathrm{x},\mathrm{y})=\mathrm{I}(\mathrm{x},\mathrm{y})+\beta \tag{3}$$

$$\mathrm{I}'(\mathrm{x},\mathrm{y})=\alpha\cdot\mathrm{I}(\mathrm{x},\mathrm{y})+\beta \tag{4}$$

α is the contrast adjustment coefficient, usually $\alpha>1$ enhances contrast, $\alpha<1$ reduces contrast, and β is the brightness adjustment value used to adjust brightness.

IV. B. 2) Noise Reduction Processing

The denoising process uses a Gaussian filter to smooth the image, reduce noise, and maintain edge details. The Gaussian filter adjusts the smoothness level based on the standard deviation, as shown in formulas (5) and (6).

$$G(x,y)=\frac{1}{2\pi\sigma^2}e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{5}$$

$$I'(x,y)=I(x,y)*G(x,y) \tag{6}$$

Among them, $*$ represents convolution operation, and σ is the standard deviation.

IV. B. 3) Grayscale and Binary Processing

Grayscale conversion is the process of converting color images into grayscale images, reducing computational complexity. The commonly used method is to convert the pixel values of the red, green, and blue channels into a single grayscale value through the weighted average method, as shown in formula (7). The next step is binarization, as shown in formula (8), to convert grayscale images into black and white images for subsequent processing and analysis. The results are shown in Figure 3.

$$\mathrm{I_{gray}}=0.299\cdot\mathrm{R}+0.587\cdot\mathrm{G}+0.114\cdot\mathrm{B} \tag{7}$$

Among them, R, G and B represent the pixel values of the red, green, and blue channels, respectively.

$$I_{binary}(x,y)=\begin{cases}1 & \text{if } I_{gray}(x,y)>T\\ 0 & \text{otherwise}\end{cases} \tag{8}$$

T is the threshold and I_{gray} is the grayscale image.

Figure 3: Grayscale and Binary Processing

IV. B. 4) Image Enhancement

Image enhancement improves the visual effect of images through histogram equalization and gamma correction. Histogram equalization redistributes pixel values and enhances image contrast by calculating the cumulative distribution function. Gamma correction regularization adjusts the brightness of an image through nonlinear transformations, using gamma values to increase or decrease the brightness of the image. Data augmentation increases the diversity of data and enhances the robustness of models in different directions and poses of images through methods such as rotation, flipping, scaling, translation, and adding random noise. It can randomly rotate and flip images to simulate different shooting conditions, increasing data diversity.

Through quality inspection and image adjustment, the quality of the image can be effectively improved, providing reliable data for subsequent image processing and analysis, ensuring the high accuracy of deep learning and data OCR recognition algorithms [30].

V. Design Experiments

By designing and implementing a series of experiments, the effectiveness of DL and OCR recognition algorithms in image data acquisition and analysis is verified [31]. The specific objectives include evaluating data preprocessing, model training, model evaluation, performance comparison, and result analysis, as shown in Table 1.

Table 1: Experimental Steps

Step	Details
Data preparation	Collecting, labeling and classifying image data, preprocessing and data enhancement
Model select	Select the appropriate DL models (CNN, RNN, GAN, etc.) according to the experimental requirements
Hyperparameter settings	Set the hyperparameters of the model training (learning rate, batch size, regularization parameters, etc.)
Model training	The model was trained on the training set, and the performance was monitored and the hyperparameters were adjusted on the validation set
Model evaluation	The final performance of the model was evaluated on the test set and the indicators were recorded (accuracy, precision, recall, F1 score)
Performance contrast	Compare the performance of different models and analyze their advantages and disadvantages
Confusion matrix	Draw the confusion matrix and analyze the classification error types
Interpretation of result	Combine the experimental results to analyze, summarize and make suggestions for improvement

V. A. Dataset Selection

This article selects COCO-Text as the basic dataset for the experiment. COCO-Text contains text images from various scenes, such as streets, shops, billboards, and other scenes. Each image has corresponding text annotation information, including the position of the text, bounding box information, and text content. The COCO-Text dataset contains a large number of images, covering text images from various scenarios. The amount of text in the dataset is relatively large, involving various lengths and types of text. The dataset, including images and annotation files, can be downloaded from the official COCO-Text website. Different classification ratios of datasets can be used to train deep learning models and understand model stability.

It ensures stable performance of the model on datasets of different scales. Parsable JSON (JavaScript Object Notation) files to obtain the text regions and corresponding labels in each image.

V. B. Model Design

This article selects CNN, RNN, and GAN models for experiments. TensorFlow can be used to construct each model, defining input shapes, convolutional layers, pooling layers, LSTM (Long Short-Term Memory) layers, etc. It can compile models, set optimizers, loss functions, and evaluation metrics. The learning rate, batch size, and regularization parameters can be set, as shown in Table 2. Training iterated 100 times.

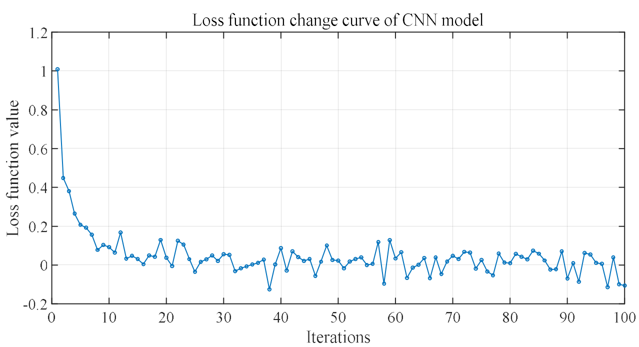
Table 2: Parameter Settings

Model	Learning rate	Batch size	Regularization parameter
CNN	0.001	64	0.0001
RNN	0.001	32	0.0001
GAN	0.0002	128	0.0001

V. C. Performance Situation

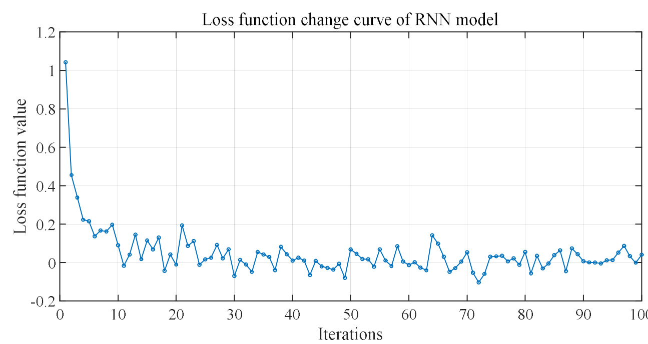
The test set can be used to evaluate model performance, recording metrics such as accuracy, precision rate, recall, F1 score, etc. This article compares the performance indicators of CNN, RNN, and GAN models and visualizes the results using a bar chart. It can draw confusion matrices for three models and analyze the true positive, false positive, true negative, and false negative situations in the results.

VI. Result Analysis



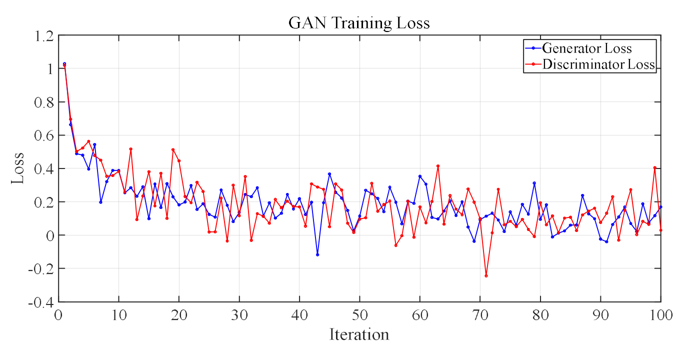
(a)

(a) CNN loss function



(b)

(b) RNN loss function



(c)

(c) GAN loss function

Figure 4: Loss function

The loss function provides a quantitative criterion to evaluate model performance. During the training and validation process, a smaller loss function value usually indicates better performance of the model on the current dataset, which can help identify overfitting or underfitting issues in the model.

According to the analysis of Figure 4, after 100 iterations, CNN has the smallest loss function value and the highest prediction accuracy. This indicates that CNN can effectively capture the features of the data and achieve good performance in relatively short iterations when processing this dataset. In contrast, within the same number of iterations, GAN did not achieve the minimum loss function value and had lower prediction accuracy. This may be due to the complex training process of GAN, which involves the game between the generator and discriminator, resulting in slow convergence speed and even instability in early iterations.

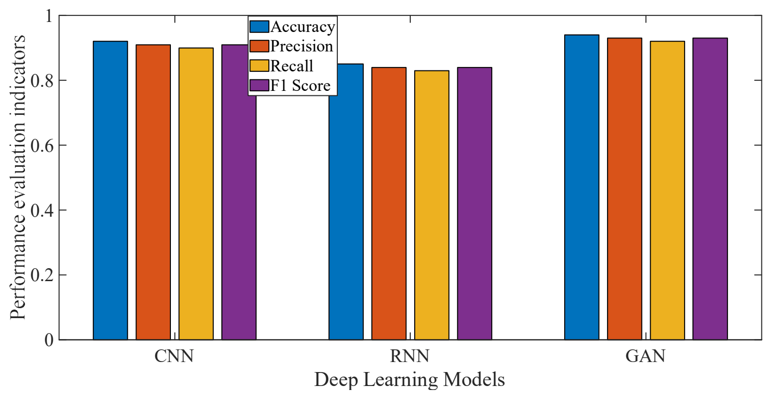
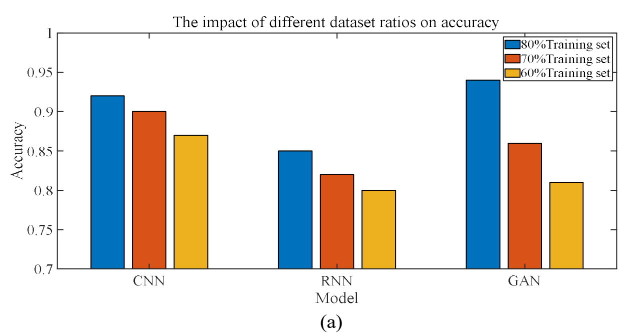
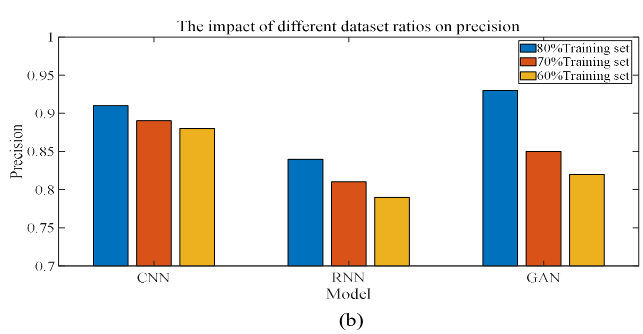


Figure 5: Performance indicators of CNN, RNN, and GAN

According to the results in Figure 5, it can be concluded that the accuracy, precision rate, recall, and F1 score of GAN are all at a high level, while RNN performs poorly among the three models. CNN has superior performance in image processing tasks; RNN is weak in image processing; GAN performs the best among the three models with sufficient iterations.



(a) Accuracy under different training sets



(b) Precision rate under different training sets

Figure 6: Performance under different training sets

According to Figure 6, it can be concluded that GAN has better OCR image data acquisition performance compared to other DL models when the training set is sufficient; When the training set is insufficient, the OCR image data acquisition performance of GAN significantly decreases. CNN maintains good OCR image data acquisition performance in various proportions of training sets, without being significantly affected. RNN performs poorly in various proportions of training sets.

VI. A. CNN Model Analysis

Through the structure of convolutional layers and pooling layers, local information can be effectively captured, preserving the importance of spatial structure. The use of parameter sharing in convolutional layers greatly reduces the number of parameters that need to be trained, improves the efficiency and generalization ability of the model. It performs excellently in processing large-scale datasets and can handle thousands or even more data samples. However, there are some issues that require high data quality, such as noise or poor data that may affect the training and prediction performance of the model. Usually, a large number of labeled data samples are required for supervised learning, which may become a bottleneck in data collection in certain fields. Usually, high computing resources are required for training and inference, especially in deep networks or large-scale datasets, where the rational utilization of computing resources needs to be considered. Due to the complexity and hierarchical structure of CNN, its internal feature extraction and decision-making process are difficult to explain, which may affect the understanding and trust of the model.

VI. B. RNN Model Analysis

RNN is suitable for processing sequential data, such as text, speech, etc. In OCR recognition, it can process character sequences or word sequences; The memory mechanism can capture contextual information, which helps improve recognition accuracy; It can handle indefinite length sequences and adapt to input of text or image sequences of different lengths. Compared to traditional rule-based or feature engineering methods, RNN can learn more complex patterns and features. End-to-end learning can be carried out, from raw data to the final output of recognition results, simplifying the process and improving efficiency. Traditional RNN cannot analyze dependencies when processing long sequences, resulting in poor performance. RNN needs to calculate the sequence step by step, which has high computational complexity and is inefficient for large-scale data sets and complex models. It requires step-by-step processing of sequence data, large amounts of computing resources and long analysis time, and long training time. When the amount of data is small or the model complexity is high, overfitting problems are likely to occur during training.

VI. C. GAN Model Analysis

The advantage of GAN lies in its ability to generate realistic image data samples, which helps to expand the dataset and improve the performance and generalization ability of OCR models. The generated data samples have diversity and can cover different image features and scenes, making the OCR model more comprehensive without the need for manual annotation, saving a lot of labor costs, especially for large-scale datasets. The data samples generated through GAN can be used for data augmentation, improving the robustness and generalization ability of the model. However, the training process of GAN is relatively complex, requiring the maintenance of dynamic balance between the generator and discriminator, as well as high requirements for parameter tuning and training skills. Sometimes generators may fall into pattern crashes, generating duplicate or similar images, lacking diversity. The generated data samples may not fully cover the potential distribution of real data, leading to performance degradation of the model in certain specific situations. Samples often lack labels of real data, making it difficult to evaluate the quality and applicability of generated data. The GAN training process requires a lot of computing resources and a long analysis time, and has high requirements for hardware equipment and training environment.

VII. Conclusions

In this study, the image data collection and analysis based on deep learning and data OCR recognition algorithm were studied and analyzed. By integrating advanced deep learning technology and efficient OCR recognition algorithms, efficient collection and accurate analysis of image data are achieved. The experimental results show that different deep learning models have improved the ability of OCR image data collection and analysis, proving the effectiveness of integrating deep learning models into image data analysis. OCR recognition algorithms show extremely high accuracy when processing text image data. Combined with the feature extraction capabilities of the deep learning model, the OCR algorithm can more quickly and accurately identify character data and text information in images, greatly improving data processing efficiency and data recognition accuracy. Experiments show that the GAN in the deep learning model has relatively good versatility and robustness in OCR recognition algorithm. When the number of iterations is sufficient, GAN has better OCR image data acquisition performance than other deep learning models; when the number of iterations is insufficient, GAN's OCR image data acquisition performance decreases significantly. When the number of iterations is sufficient, CNN has better OCR image data acquisition performance; when the number of iterations is insufficient, CNN can still maintain good OCR image data acquisition performance. Whether on different types of image data or in different application scenarios, deep learning models have shown relatively stable performance, proving their broad application prospects. This study has achieved certain results in image data acquisition and analysis, but there is still much room for improvement. Future work can further optimize the model structure to improve the performance of the model in processing complex image data. By studying more types of deep learning models and OCR recognition technologies, combined with more diverse data sets, the generalization ability and practicality of the system can be improved in many aspects.

References

[1] Liu J, Zhu F, Chai C, Yuyu Luo,Nan Tang. Automatic data acquisition for deep learning[J]. Proceedings of the VLDB Endowment, 2021, 14(12): 2739-2742.
[2] Serre T. Deep learning: the good, the bad, and the ugly[J]. Annual review of vision science, 2019, 5: 399-426.
[3] Janiesch C, Zschech P, Heinrich K. Machine learning and deep learning[J]. Electronic Markets, 2021, 31(3): 685-695.
[4] Bengio Y, Lecun Y, Hinton G. Deep learning for AI[J]. Communications of the ACM, 2021, 64(7): 58-65.
[5] Mathew A, Amudha P, Sivakumari S. Deep learning techniques: an overview[J]. Advanced Machine Learning Technologies and Applications: Proceedings of AMLTA 2020, 2021: 599-608.
[6] Memon J, Sami M, Khan R A, et al. Handwritten optical character recognition (OCR): A comprehensive systematic literature review (SLR)[J]. IEEE access, 2020, 8: 142642-142668.
[7] Nguyen Q D, Le D A, Phan N M, et al. OCR error correction using correction patterns and self-organizing migrating algorithm[J]. Pattern Analysis and Applications, 2021, 24(2): 701-721.
[8] Jain P, Taneja K, Taneja H. Which OCR toolset is good and why: A comparative study[J]. Kuwait Journal of Science, 2021, 48(2).
[9] Rahmati M, Fateh M, Rezvani M, et al. Printed Persian OCR system using deep learning[J]. IET Image Processing, 2020, 14(15): 3920-3931.
[10] Sampath A K, Gomathi N. Handwritten optical character recognition by hybrid neural network training algorithm[J]. The Imaging Science Journal, 2019, 67(7): 359-373.
[11] Dong S, Wang P, Abbas K. A survey on deep learning and its applications[J]. Computer Science Review, 2021, 40: 100379.
[12] Wang X, Zhao Y, Pourpanah F. Recent advances in deep learning[J]. International Journal of Machine Learning and Cybernetics, 2020, 11: 747-750.
[13] Alzubaidi L, Zhang J, Humaidi A J, et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions[J]. Journal of big Data, 2021, 8: 1-74.
[14] Khan M, Jan B, Farman H, et al. Deep learning methods and applications[J]. Deep learning: convergence to big data analytics, 2019: 31-42.
[15] Thompson N C, Greenewald K, Lee K, et al. The computational limits of deep learning[J]. arXiv preprint arXiv:2007.05558, 2020.
[16] Liu Y, Li Z, Li H, et al. On the hidden mystery of ocr in large multimodal models[J]. arXiv preprint arXiv:2305.07895, 2023.
[17] Hu S. Research on data acquisition algorithms based on image processing and artificial intelligence[J]. International Journal of Pattern Recognition and Artificial Intelligence, 2020, 34(06): 2054016.
[18] Glasenapp L A, Hoppe A F, Wisintainer M A, et al. OCR applied for identification of vehicles with irregular documentation using IoT[J]. Electronics, 2023, 12(5): 1083.
[19] Reul C, Christ D, Hartelt A, et al. OCR4all—An open-source tool providing a (semi-) automatic OCR workflow for historical printings[J]. Applied Sciences, 2019, 9(22): 4853.
[20] Michalak H, Okarma K. Robust combined binarization method of non-uniformly illuminated document images for alphanumerical character recognition[J]. Sensors, 2020, 20(10): 2914.
[21] Ranjan A, Behera V N J, Reza M. Ocr using computer vision and machine learning[J]. Machine Learning Algorithms for Industrial Applications, 2021: 83-105.
[22] Boros E, Nguyen N K, Lejeune G, et al. Assessing the impact of OCR noise on multilingual event detection over digitised documents[J]. International Journal on Digital Libraries, 2022, 23(3): 241-266.
[23] Martínek J, Lenc L, Král P. Building an efficient OCR system for historical documents with little training data[J]. Neural Computing and Applications, 2020, 32: 17209-17227.
[24] Moudgil A, Singh S, Gautam V. An overview of recent trends in OCR systems for manuscripts[J]. Cyber Intelligence and Information Retrieval: Proceedings of CIIR 2021, 2022: 525-533.
[25] Shambharkar Y, Salagrama S, Sharma K, et al. An automatic framework for number plate detection using ocr and deep learning approach[J]. International Journal of Advanced Computer Science and Applications, 2023, 14(4).
[26] Brady D J, Fang L, Ma Z. Deep learning for camera data acquisition, control, and image estimation[J]. Advances in Optics and Photonics, 2020, 12(4): 787-846.
[27] Kettunen K, Koistinen M, Kervinen J. Ground truth OCR sample data of Finnish historical newspapers and journals in data improvement validation of a re-OCRing process[J]. LIBER Quarterly: The Journal of the Association of European Research Libraries, 2020, 30(1): 1-20.
[28] Irimia C, Harbuzariu F, Hazi I, et al. Official Document Identification and Data Extraction using Templates and OCR[J]. Procedia Computer Science, 2022, 207: 1571-1580.
[29] Nguyen T T H, Jatowt A, Coustaty M, et al. Survey of post-OCR processing approaches[J]. ACM Computing Surveys (CSUR), 2021, 54(6): 1-37.

[30] Ströbel P B, Clematide S, Volk M. How much data do you need? about the creation of a ground truth for black letter and the effectiveness of neural OCR[J]. 2020.
[31] Clausner C, Antonacopoulos A, Pletschacher S. Efficient and effective OCR engine training[J]. International Journal on Document Analysis and Recognition (IJDAR), 2020, 23: 73-88.