International Journal of Scientific Research in Science, Engineering and Technology

Print ISSN - 2395-1990
Online ISSN : 2394-4099

OPEN ACCESS

Available Online at : www.ijsrset.com
doi : https://doi.org/10.32628/IJSRSET

Artificial Intelligence Techniques for Identifying and Detecting Objects

Pushpa G

Senior Scale Lecturer, Department of Computer Science and Engineering, Government Polytechnic, Channasandra, Kadugodi, Bangalore, India

ARTICLE INFO

Article History:
Accepted: 01 June 2019
Published: 20 June 2019

Publication Issue :
Volume 6, Issue 3
May-June-2019

Page Number :
538-545

ABSTRACT

AI has become necessary due to the use of deep learning and machine learning techniques. AI seeks to reduce human intervention in order to automate tasks. IT, healthcare, banking, and agriculture all make extensive use of it. Several deep learning algorithms that mimic the intelligence of the human brain are used to do this. It is possible to adjust these AI algorithms in accordance with evolving requirements and increased effectiveness. In order to classify the photographs and identify the things they include, this research attempts to make use of the advancements in AI technology. CNN (Convolutional Neural Networks) is a popular AI method. The CNN is a multi-layered deep learning system that extracts and filters the parameters found in the pictures. To increase the accuracy of picture recognition, certain extra layers of the CNN algorithm and ResNet50 are utilized to extract the parameters. ImageNet is the picture dataset used to train and evaluate the suggested model. Prior to being sent to the suggested model, the photos are first processed. The photos retrieved following the first processing are used to train, validate, and test the suggested model. Until the highest level of precision is achieved, the same procedure is carried out several times. It is noted how well the suggested model is in recognizing images. A comparison is made between the results achieved and various picture classification methods, such as VGG16 and VGG19. In terms of accuracy, it is determined that the suggested model performs better than other conventional techniques.

Keywords : AI Methods, Classification, CNN, ResNet50, Image Recognition, Object Detection, Feature Extraction.

I.INTRODUCTION

Computer vision, a general phrase referring to the process of training computers on visual data and producing outcomes that are comparable to those of the human method, is synonymous with image recognition. This machine-based technology can detect and classify things, people, and locations, among other visual tasks. Computers must be taught and educated to do a certain activity, in contrast to the human brain, which can identify and react to items with ease. An image recognition algorithm makes it simple for the computer to start collecting data. After this data collection, the typically formed data is interpreted as a single entity. The picture may be either raster or vector depending on how the computer understands it. While computers require programming and processing to interpret and understand pictures and tasks, human brains do it practically automatically. Images cannot be distinguished from one another by a computer. Raster pictures, often known as bitmaps, are made up of individual pixels that are arranged collectively to form a grid. Vector graphics are composed of a collection of polygons of various colours, each with an explanation, and are composed of line art based on algorithms rather than pixels. The accumulating and interpretation processes are crucial since any compromise in these phases might lead to ineffective recognition later on.

AI in the process of picture recognition

AI is at the forefront of image recognition, which includes text, object, and facial recognition. This sophisticated biometric authentication method uses software to recognize and validate an individual. Using facial traits and instantly compare them to millions of others that are similar. To locate its match, the algorithm consults the deep learning database. Digital marketing, social networking sites, and the smartphone business all employ facial recognition technologies. Object recognition is a computer vision technique that locates and detects things in a picture. Image detection is not the same as object detection. Item detection identifies each item by drawing a box around it, whereas image detection labels each image. This technique aids in tracking the object's location and movement. In this regard, object detection has special benefits and capabilities and offers more information than image detection. The process of providing photos to the model in order to identify the textual region is known as text detection or reading a text in a natural image. The area is plotted with a bounding box. Text in the image is recognized using AI. In plain language, it involves identifying and removing text from a picture.

II. RELATED WORKS

The effectiveness of artificial neural networks (ANN) in the healthcare industry was examined by M. Kim et al. (2019). Additionally covered were the evolution of ANN applications in healthcare and the history of medical image processing.

In order to improve the effectiveness and classification of billing recognition, M. W. Tian et al. (2019) offered a number of test and particle applications for picture recognition in the banking industry.

I. B. Abbasor (2019) talked about the idea of using computer vision and deep learning techniques to recognize images of different plants and crops. The primary goal of image recognition in orchards, woods, and agriculture is to use hyperspectral analysis of satellite photos to identify and categorize different types of forest, fruit, vegetable, and leaf images based on plants. According to J. Chen (2020), the usual algorithm produced the facial photos' important points from the typical day. In the current case, the precise point of the facial photographs is designed using the CNN approach to provide a stunning result. The use of deep learning and machine learning in the mining industry for drilling, blasting, mine planning, equipment selection, mineral processing, etc. was examined and reported by D. Ali and S. Frimpong (2020). And difficulties and problems in the present methods are also examined. Lastly, in order to establish a mining industry that is safe, effective, and efficient, the author favored AI, ML, and Dl. A review of deep learning techniques for medical image processing and recognition was conducted by M. M. Nair et al. in 2021. The author also covered the difficulties, characteristics, problems, and advancements in deep learning medical picture identification.

The artificial intelligence-based picture identification system was described by author Z. Zang (2021). He gave a thorough analysis of artificial intelligence, picture recognition, and the creation of AI applications for image identification.

A unique deep learning technique for locating or identifying items in e-business platforms was put out by P. Zhang in 2021. And use past sales data from e-commerce sites to develop the TensorFlow framework's LSTM network model technique. To obtain better results, the LSTM network model and AR models were then compared. The outcomes demonstrate that the LSTM approach is more straightforward and accessible. The quality of product recognition is improved when LSTM and innovative deep learning are used in e-business for picture recognition.

For picture recognition, Y. Li (2022) examined deep learning techniques such as CNN (Convolution Neural Network), RNN (Recurrent Neural Network), and GAN (generative adversarial network). Additionally, he talked about how well such algorithms work in image recognition applications across a range of domains, including medical image recognition, face recognition, remote sensing, and image recognition. Additionally, he examined how deep learning has advanced in picture identification. It demonstrates that deep learning for image identification will provide high-quality pictures. H. Li and colleagues (2022) have written on the players' motion effect recognition. Image recognition using artificial intelligence (AI) technology is used for this investigation. AIT yields more accurate ratio results than the prior technique, with 98.8% for recognition, 97.7% for performance, and 95.9% for accuracy.

III. DESIGN AND IMPLEMENTATION

A. IMAGE RECOGNITION

Deep learning algorithms are typically used for this. These deep learning algorithms identify subsequent photos by using the knowledge they have gained from the images. But traditional deep-learning algorithms also have trouble identifying the items in the picture. The following actions are often taken to address this problem:

1) Image Recognition
2) Verification of images
3) Identifying objects
4) Understanding scenes
5) Recognition of certain objects.

B. VERIFICATION OF IMAGES

It is a crucial stage in image processing since the existence of irrelevant pictures can impact prediction accuracy and lengthen processing times. A reference pattern containing feature vectors that aid in image verification makes up the image verification procedure. Any photos that do not fit that pattern will not be processed further. A distance value that indicates how comparable the captured photos are serves as the basis for this rejection. Rejected photos are those that fall inside the distance value. Animal, face, and fingerprint pictures are a few of the image types employed for identification. A loss function may be used to automatically identify photos in a deep learning algorithm. The loss function is used to determine which pictures have the given distance, and it is computed for each image.

C. DETECTING OBJECTS

It may be characterized as identifying an object's placement inside a picture. It is mostly employed in algorithms for video processing that identify an object's motion inside a video frame. Face and pedestrian detection are examples of object detection.

Since there won't be more than two or three moving objects in the video, these jobs also fall under object detection. Deep learning or machine learning techniques are used to detect objects. AdaBoost, SVM, and HOG features are a few of the popular object detection techniques. Conventional deep learning algorithms typically need at least two classifiers for training. This is done in order to scan photos more quickly. Even with a single network, deep learning-based algorithms are able to recognize more objects in a single frame.

D. CLASSIFICATION OF IMAGES

This is done in order to categorize the items in the pictures according to their respective categories. In order to assist categorize the objects, classification algorithms typically use predetermined categorization. This method may be used to classify and categorize a wide range of things. Conventional deep learning methods take a collection of characteristics from the training photos in order to identify the objects. A histogram is created from this collection of characteristics. It is necessary to detect the item represented by these histograms. The objects in the picture are categorized using the items whose characteristics fit the histogram.

Since 2015, this technique has been employed by a variety of classification algorithms to categorize photos and apply labels or tags.

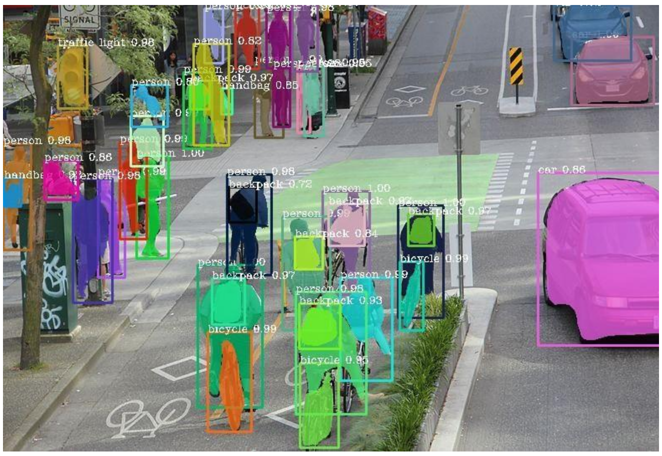


Figure 1: Segmenting Semantic Objects (Scene Understanding)

SCENE UNDERSTANDING (SEMANTIC SEGMENTATION)

In this process, the middle layers extracts more usable parameters as the input layers consisting of more parameters are filtered initially. The initial filtering loads the middle layers with usable parameters. So, the middle layer is used for the segmentation. It can be seen in the following **Figure-1**.

SPECIFIC OBJECT RECOGNITION

A variety of histograms with feature sets for locating certain items are used to give characteristics to each object. These feature sets are used by the SIFT to identify the objects. SIFT uses a voting procedure to identify these things. It calculates the distance between the feature sets and the real picture. Positive votes are given to the photos that fall inside the designated range. As the pictures adhere to the same reference pattern, the SIFT will have no trouble identifying the items. LIFT (learned invariant feature transform) is utilized to enhance the process's performance rather than machine learning techniques for object recognition. Through deep learning, every SIFT process is converted to LIFT.

CNN-BASED IMAGE IDENTIFICATION

Prior to analysing images using deep learning algorithms, the image recognition process often included identifying the sorts of images. Usually, experts segment photos manually or semi-manually as part of this processing. These procedures are not automated as doing so degrades the algorithms' learning quality. CNN is a popular deep learning method that can identify pictures by learning their properties during training. This work uses CNN for scene detection, which is the process of identifying objects in photographs.

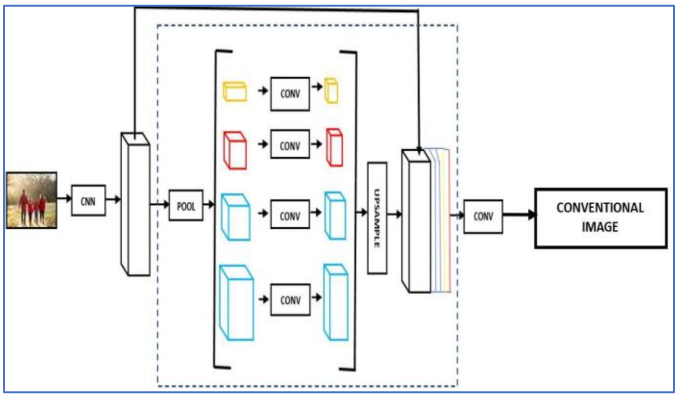


Figure-2. The architecture of the Convolutional Neural networks

For the CNN to categorize the photos, it must be taught. The CNN algorithms are first trained on a training dataset. During the training phase, the algorithms take the parameters out of the training pictures and store them in the convolutional layers. The backpropagation method is used to do this; the weights and kernel that are acquired are then passed into several algorithm layers. **Figure 2** explains it. Each layer processes the pictures, and real-time outcomes are compared with the predictions made by the CNN layers. After noting the inaccuracies, the convolutional layers are once more fed the settings that yield accurate results. Until the highest level of precision is achieved, this procedure is repeated.

IV. PROPOSED CNN ARCHITECTURE

CNN and ResNet50 layers make up the suggested deep learning model, which is used to determine the trainable parameters. For processing efficiency, the photos' size is changed to a standard 28 by 28 pixel format. To convert these photos to the desired resolution, they are first processed using the previously described processes. Training and testing sets of these photos are separated. The following figures demonstrate the results of feeding the training set into the suggested model. Two-dimensional convolutional layers make up the first layers. Every parameter that is accessible is extracted. After that, batch normalization is applied to the photos. Without changing the crucial parameters, the batch normalization procedure normalizes the photos to eliminate extraneous features. Once more, the pictures' useful parameters are extracted using a convolutional 2-dimensional layer, and batch normalization is carried out. The photos go through max-pooling after the procedure is repeated twice. By eliminating the undesirable parameters from the total parameters acquired, the pooling layers lower the number of parameters. The useable parameters are extracted from the photos by repeating all of the previous stages. Following this procedure, the resulting pictures are passed to the ResNet50 and dropout layers while remaining at the default pixel size of 28 by 28. The dropout layers prevent overfitting of the pictures, while the ResNet layers gather the parameters using the residual approach. Prior to being flattened by going via the flattening layer, the pictures are sent to three ResNet50 and dropout layers. In order to improve prediction accuracy, the flattening layer softens the picture and produces high contrast images. The following table shows the parameters that were acquired after going through the ResNet50 and dropout layers. The photos are recognized using the parameters that were acquired. The predictions of our suggested model are displayed in **Figure 3** below. Since it is still in the training stage, it is first verified using simple pictures of animals. The next sections address the prediction accuracy using the parameters that were acquired.

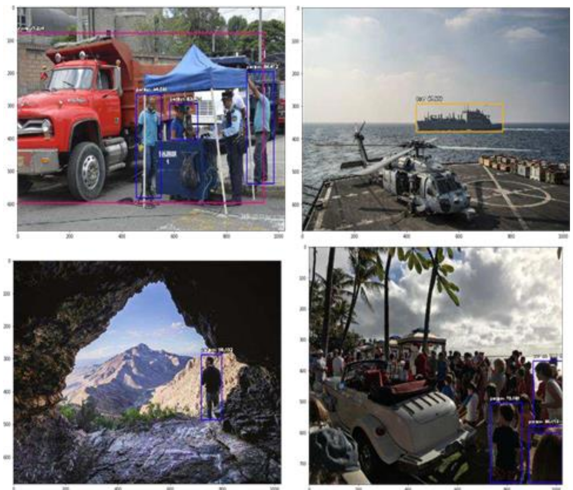


Figure-3. Classified Images (Trained Phase)

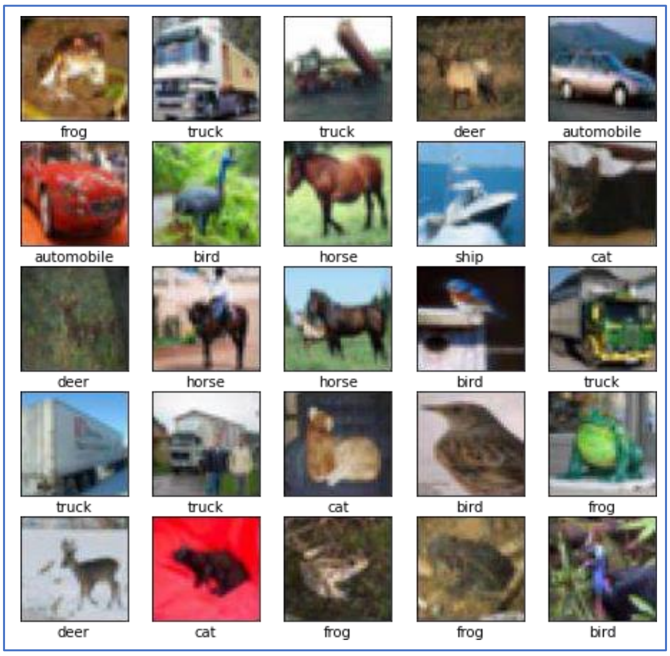
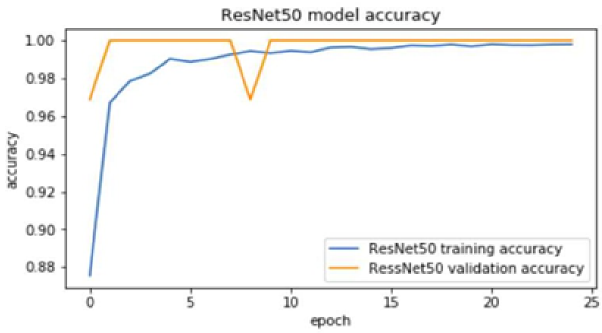


Figure-4. Classified Images (Testing Phase)

The following pictures, which are displayed in Figure 4, are fed into the suggested model during the testing phase. The figure displays the prediction accuracy. It recognizes a variety of items in the figure, such as ships, cars, people, and kids. The accuracy of the suggested model's ability to identify things in the photos will be covered in the parts that follow.

V. RESULTS AND DISCUSSION

The proposed CNN and ResNet architectures implemented in Python software and the experimental results are verified.



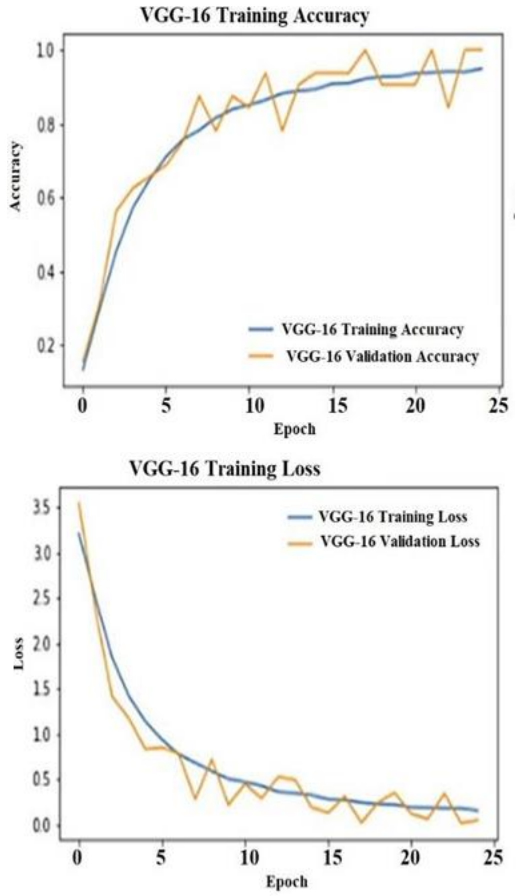


Figure-5. Accuracy and loss of VGG-16 algorithm

The training loss and validation loss also decrease with the increasing epochs. The accuracy of the VGG16 widely fluctuates even after ten epochs both in the training and validation process. The VGG16 shows wide fluctuations. Even after complete training, there is no consistent accuracy.

The accuracy is consistent only after 10 to 15 epochs. The accuracy also faces wide fluctuation after the training process. The loss of the VGG19 also takes more than 10 to 15 epochs for getting consistency. However, the VGG19 model provides better accuracy than the VGG16 model.

The Following figure shows the accuracy of the predictions during training and validation. During the training process, the accuracy improves faster during the initial stage and reaches a constant level after five epochs, and after ten epochs, the accuracy obtained is consistent. The ResNet50 functions effectively above 5 to 10 epochs. Even during the validation process, there is a significant dip in the accuracy within ten epochs. Thus, the proposed model provides better accuracy than the other image recognition models.

Figure-6. Accuracy of the Proposed ResNet50 model

The following figure shows the error loss from the proposed ResNet model. It can be seen from the following figure that the loss significantly reduces within five epochs, and the loss is consistent after ten epochs. Even during the validation process, the fluctuation during a loss are very less after 10 epochs. The proposed model provides consistent results after 10 epochs. The previous models considered took nearly 15 epochs for getting consistent results. Even after 15 epochs, the accuracy and loss saw minor fluctuations. But the proposed model provides better results with minimum fluctuations compared to the previous models.

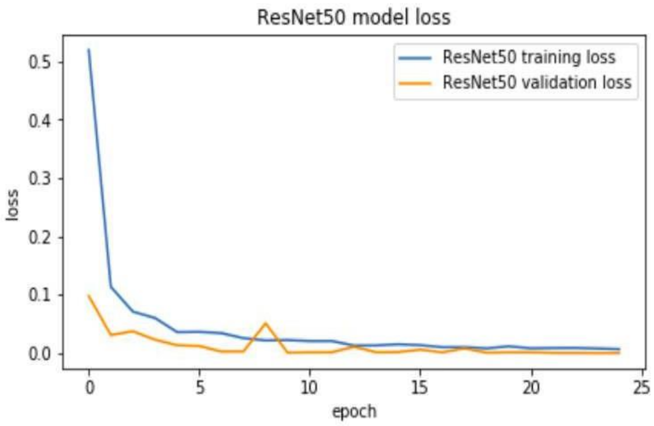


TABLE I

FIGURE-7. LOSS OF THE PROPOSED RESNET50 MODEL

Comparison of Model Performance

We evaluated the performance of three convolutional neural network (CNN) architectures—VGG19, VGG16, and ResNet50—based on their training accuracy, validation accuracy, time consumption per step, and loss.

Model	Training accuracy	Validation accuracy	Time used	Loss
VGG19	0.8476	0.8764	184s 335ms/step	0.5500
VGG16	0.8225	0.8567	273s 496ms/step	0.2528
ResNet50	0.9789	1	79s 167ms/step	0.0080

Table 1. Performance comparison of algorithms considered

Accuracy Analysis

ResNet50 achieved the highest training accuracy (97.89%) and perfect validation accuracy (100%), suggesting it effectively learned the dataset patterns.
VGG19 attained a training accuracy of 84.76% and validation accuracy of 87.64%, which is slightly better than VGG16, which had a training accuracy of 82.25% and validation accuracy of 85.67%.
The difference in accuracy between VGG models and ResNet50 indicates that ResNet50 generalizes better to unseen data.

Computational Efficiency

ResNet50 was the fastest model, completing training in 79s (167ms/step), making it the most efficient in computation.
VGG19 required 184s (335ms/step), while VGG16 took the longest at 273s (496ms/step).
The higher computational time of VGG models is due to their deeper architecture and large number of parameters, leading to increased processing requirements.

Loss Analysis

ResNet50 had the lowest loss (0.0080), indicating minimal errors in predictions.
VGG16 had a significantly higher loss (0.2528) compared to ResNet50, while VGG19 had an even higher loss of 0.5500.
The low loss in ResNet50 suggests better optimization and convergence compared to VGG models.

VI. CONCLUSION

The picture identification technique makes use of a number of AI-based algorithms. These algorithms are hybrids of AI. These methods address the goal and include a number of machine learning, deep learning, and classification techniques. Additionally, a hybrid deep learning method combining CNN and ResNet50 has been suggested in this study. The suggested CNN the ResNet50 optimizes and chooses the necessary features, while the CNN algorithms, which are made up of several layers, extract features from the pictures. Repeating the procedure several times improves the pictures' accuracy, and the accuracy achieved grows significantly with each epoch. Until the highest level of accuracy is consistently achieved, this iterative process is repeated. The imagenet dataset is used to train the suggested model. The pictures are divided into sets for testing, validation, and training. The suggested model is applied to the photos, and the resulting output is thoroughly examined. The accuracy and loss of the findings are compared with those of other conventional techniques, such as VGG16 and VGG19. Additionally compared is the picture recognition efficiency. The comparisons demonstrate that the suggested model performs better in terms of accuracy, loss, and efficiency than the other models taken into consideration. Additionally, the suggested model requires fewer picture recognition epochs.

REFERENCES

[1]. M. Kim, J. Yun, Y. Cho, K. Shin, R. Jang, H. J. Bae, and N. Kim, “Deep learning in medical imaging,” Neurospine, vol. 16, no. 4, pp. 657, 2017.

[2]. M. W. Tian, S. R. Yan, X. X. Tian, and J. A. Liu, “Research on image recognition method of bank financing bill based on binary tree decision,” Journal of Visual Communication and Image Representation, vol. 60, pp. 123-128, 2017.

[3]. M. M. Nair, S. Kumari, A. K. Tyagi, and K. Sravanthi, “Deep learning for medical image recognition: open issues and a way to forward,” in Proceedings of the Second International Conference on Information Management and Machine Intelligence, pp. 349-365, Springer, Singapore, 2012.

[4]. I. B. Abbasov, Image Recognition in Agriculture and Landscape Protection.

[5]. D. Ali and S. Frimpong, “Artificial intelligence, machine learning and process automation: existing knowledge frontier and way forward for mining sector,” Artificial Intelligence Review, vol. 53, no. 8, pp. 6025-6042, 2017.

[6]. J. Chen, “Image recognition technology based on neural network,” IEEE Access, vol. 8, pp. 157161-157167, 2015.

[7]. P. Zhang, “E-commerce products recognition based on a deep learning architecture: Theory and implementation,” Future Generation Computer Systems, vol. 125, pp. 672-676, 2018.

[8]. Y. Li, “Research and application of deep learning in image recognition,” in 2022 IEEE 2nd International Conference on Power, Electronics and Computer Applications (ICPECA), pp. 994-999, IEEE, January 2012.

[9]. Z. Zhang, L. Zhao, and T. Yang, “Research on the application of artificial intelligence in image recognition technology,” in Journal of Physics: Conference Series, vol. 1992, no. 3, p. 032118, IOP Publishing, August 2021.