www.cadanda.com Computer-Aided Design AND APPLICATIONS

Construction of Intelligent Visual Communication System Based on Deep Learning and Computer Aided Driving

Yue Wu[1], Kun Zhao[2,*] and Cong Peng[3]

[1]Academy of Fine Arts and Design, Hebei Institute of Communications, ShiJiaZhuang, HeBei 050000, China, wuyue@hebic.edu.cn
[2]Academy of Fine Arts and Design, Hebei Institute Of Communications, ShiJiaZhuang, HeBei 050000, China, zhaokun@hebic.edu.cn
[3]Academy of Fine Arts and Design, Hebei Institute Of Communications, ShiJiaZhuang, HeBei 050000, China, pengcong@hebic.edu.cn

Corresponding author: Kun Zhao, zhaokun@hebic.edu.cn

Abstract. Because of the great development of computer at present, it is natural to apply computer vision technology of assistant driving to it. The auxiliary driving system can effectively help the driver understand the environmental information around the vehicle. The work can bring strong visual impact to people after being processed by the image processing technology. In addition, the computer graphics and image technology also provides more inspiration for the design staff and improves their innovation ability. Through the system hardware, the acquisition, processing, transmission, storage and control of visual information are realized, which provides hardware support for the digital system. In terms of software, the CAD auxiliary technology is introduced, and the application principle and role of CAD auxiliary technology in the system are explained. In addition, the software flow is also given. Based on SSD target detection algorithm, an improved SSD license plate detection algorithm suitable for complex scene is proposed. The improved SSD algorithm improves targets such as license plates in complex scenes by fusing feature information of different scales, deepening the depth of the prediction layer, and setting the default frame proportion matching the size of the license plates.

Keywords: Deep Learning; Computer-Aided; Visual Communication Design; SSD License Plate Detection Algorithm.
DOI: https://doi.org/10.14733/cadaps.2023.S8.55-65

1 INTRODUCTION

Artificial intelligence technology has brought earth-shaking. Benefiting from the improvement of computer hardware level, deep learning has shone brilliantly in a variety of AI algorithms, and has brought unprecedented breakthroughs [1]. In the deep learning algorithm, the convolutional neural network adopting the feature learning strategy has achieved performance beyond the traditional algorithm based on manual features in many areas of computer vision, such as target detection, target recognition and image semantic segmentation. Most of the manual features are shallow features, which are vulnerable to noise interference, which shows strong stability and spatial invariance [2].

People can understand different types of information in the process of watching. At the same time, with the help of relevant media, they can obtain content that cannot be obtained, such as listening, and then conduct all-round communication, which can effectively reduce the impact of communication caused by poor words. On this basis, we can interact reasonably through different understanding of images. People can find the relevance in the process of ingenious changes of points, lines and surfaces in view of the characteristics of the image and artistic conception of related objects. On this basis, we can form a more comprehensive understanding of things. This is the most direct feeling that the visual communication design can bring to the audience, and can help the audience in the process of watching [3]. Understand the deeper connotation of things. People's spiritual world will also be sublimated in the process of watching and visual transmission [4].

Visual communication with CAD technology has three-dimensional and dynamic special effects, changes the traditional planarization and static characteristics of images, transforms image products into virtual information images, expands the extension of traditional visual communication design, and makes a qualitative leap in visual communication [5]. Autonomous vehicle is a typical high-tech integrated application. Including scene perception, optimization calculation, multi-level assisted driving, etc. Song et al. [6] uses computer vision, taking sensors, information fusion and computer vision as the direct entry points for data processing. This technology is based on many new and advanced comprehensive applications. The system collects road image information through camera, and detects and identifies road image information based on computer vision. Its key modules can be summarized as environment perception, behavior decision-making, path planning and motion control. At the same time, the surrounding environment data and internal data of the vehicle are the basic data to support the auxiliary driver. Wang et al. [7] uses computer vision as the main means of perception. The driving strategy can be continuously changed according to the driver's participation. However, the effectiveness of the sensing system is limited. Once the system fails, we must consider whether there is an alternative. This is the purpose of high-resolution map. Based on artificial intelligence technology, a perception analysis model of the dynamic environment around the vehicle is established. So as to assist the decision-making of self-driving vehicles. Most sensing systems adopt a sensing scheme combining computer vision and radar. Wu and Li [8] filters other contours according to the contour area and the length-to-width ratio of the rectangle through the contour and calculates the area to improve the calculation efficiency. In the color extraction phase, you need to convert the color space, and then set the HSV threshold of the color. Finally, the image is normalized to match the image in the model, and the image is processed by Gaussian blur to reduce the noise in the image. Yan et al. [9] communicates through the level change of GPIO to complete the corresponding functions, and makes corresponding decisions immediately on the collected road condition information. Due to the real-time detection of the system, this function requires high software performance, so different data monitoring methods are used for data processing. Zhang [10] believes that vehicle recognition is based on license plate detection and vehicle cutting based on license plate. The road image is compressed, filtered and binarized by the system graphics system and then input into the lane recognition model to process the collected road image. The size and shape of the cropped photos are different at different distances and on different vehicles, so the cropped photos must be compressed and processed. Adjust the number of layers of neural network to ensure the reliability of recognition accuracy.

2 STATE OF THE ART

2.1 Overview of Visual Communication System

In other words, visual communication technology has also been greatly expanded. For most people, one of the important ways to obtain information is vision. In the process of daily life and work, people can obtain the information contained in words and pictures by observing and reading them. It is precisely because of the existence of this way of communication that can help relevant staff to build a unique visual effect by designing words and pictures, And embed some main data information according to relevant requirements. For the visual communication system, the picture can be further beautified and designed to enhance the viewing value contained in the picture.

2.2 Characteristic Description of Computer-Aided Model

The data volume of CAD models is also increasing. The traditional 3D model feature extraction technology is no longer applicable to CAD models. Industrial CAD 3D models have many shapes that ordinary 3D models do not have, such as holes, cavities and other structures. Using traditional 3D model feature extraction technology to propose CAD models is not ideal, and the accuracy is low. Therefore, many scholars continue to explore the feature extraction of such models. Converting 3D CAD model into attribute adjacency graph is the core idea of most CAD model feature extraction methods.

3D CAD models are basically represented by patches or boundaries. The data structure is very complex and cannot be directly used for similar evaluation. A good model feature descriptor is an important condition for 3D CAD model similarity evaluation and retrieval. An effective and reasonable feature descriptor should have the following characteristics: ⊖ integrity: feature descriptor can well interpret the topological structure and shape features of the model itself; The geometric structure of 3D CAD model is very complex. Its geometric elements include not only simple faces, lines and points, but also complex conics, free surfaces and Trimed surfaces. Therefore, a good 3D CAD feature descriptor should have a wide range of representation capabilities. Uniqueness: there is one-to-one correspondence between the 3D CAD model and its descriptor; Uniqueness can ensure that the model will not have serious over-fitting phenomenon. ⊛ Stability: the feature descriptor will not have a large impact when it encounters slight fluctuations on the model surface; The surface of the model will sometimes produce slight fluctuations, and the feature descriptor will not be greatly disturbed when describing the model features, so as to ensure the accuracy of the similarity evaluation. ④ Hierarchical: The feature descriptors of 3D CAD models should meet the representation of each level; Lock the local details until they are needed, which is the desired effect of the hierarchical representation of the feature descriptor. An excellent feature representation algorithm should have the effect of representing the multi-granularity characteristics of the model at multiple levels, which is more conducive to the reasonable consistency matching of 3D CAD models.

2.3 Deep Learning Algorithm for 3D CAD Model

According to the different input information of the depth learning model, it for 3D CAD model can be divided into: it based on incomplete representation of 3D model; Deep learning algorithm based on original representation of 3D model. If the feature descriptor is extracted from the 3D model view information or the model itself. The input descriptor is an incomplete 3D model feature, which over-extracts the model information. The specific process is as follows: first, build a deep learning model; Next, select Light Field Descriptor (LFD) to extract the view; Then select 35 moments of Zernike to extract 3500-dimensional feature vectors from the view. The deep learning model built by Qin is shown in Figure 1.

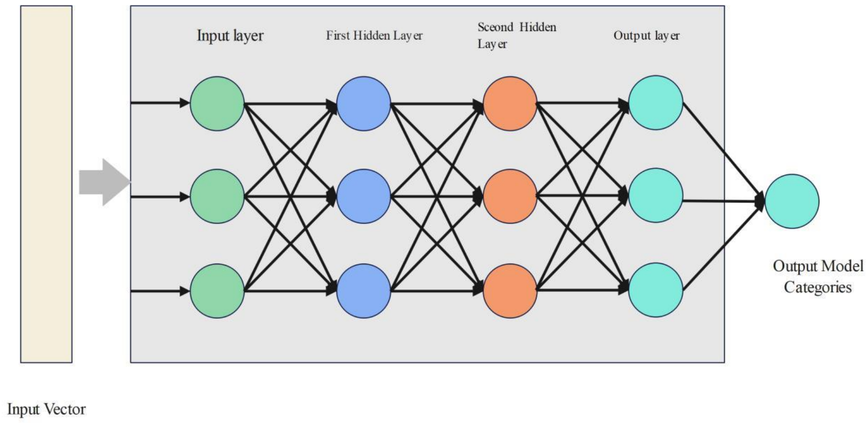

Figure 1: Deep learning model built by Qin.

After the view is extracted, the feature vector needs to be extracted. Select 35 moments of Zernike to obtain 3500-dimensional feature vectors as the input of the depth learning model. Zernike defines a set of polynomials set as Z (x, y) (the parameters satisfy {x2+y2 ≤ 1}). The Zernike feature of an image refers to the projection of the image on the orthogonal polynomial Zmn (x, y). The following formula (1) defines an m-order Zernike moment with a repetition rate of n:

$$L_{mn} = \frac{n+1}{\pi} \iint_{x^2 + y^2 \leq 1} f(x, y) Z_{mn} \, dxdy \tag{1}$$

The integral can be used to solve the Zernike moment characteristics of digital images by using summation instead. For example, formula 2:

$$L_{mn} = \frac{n+1}{\pi} \sum_{x} \sum_{y} f(x, y) z_{mn} \quad x^2 + y^2 \leq 1 \tag{2}$$

When calculating the Zernike moment feature, gravity must be moved to the coordinate origin to map the image pixel points into the unit circle. Because Zernike moments can arbitrarily construct higher-order moments containing more information on the image, Zernike moments have good recognition effect.

3 METHODOLOGY

3.1 Deep Learning Model and Principle

In the 1990s, LeCun et al. proposed due to the lack of computer hardware and training data at that time, the early CNN model did not perform well in dealing with complex problems. In 2012, the Alexnet model of convolutional neural network won the championship in the mageNet competition. Since then, CNN has officially entered the public view. VGGNet, GoogleNed ResNet and other powerful CNN models have been proposed. Although these models are different, they all follow the basic principles of CNN model design: local connection, weight sharing and down-sampling. Cyclic neural network (RNN) increases the horizontal connection between neurons on the basis of general neural network, making the neural network can be used to deal with sequence problems. Now RNN is widely used in machine translation, speech recognition, image generation and other application scenarios. The calculation formulas of RNN output layer and hidden layer are as follows:

$$o_t = g\left(Vs_t\right) \tag{3}$$

$$s_t = f\left(Ux_t + Ws_{t-1}\right) \tag{4}$$

RNN can learn sequence information through such hidden layer design, thus capturing the correlation between contexts. It greatly limits the ability of RNN and can only deal with short sequence information features.

Convolution neural network is a mapping of input and output, and the whole network is the mapping function of this relationship. Compared with general functions, this mapping relationship is complex. However, because the same convolution kernel at each layer shares parameters, compared with other deep learning models, the parameters are few, which can improve the learning efficiency. Convolution neural network has been commercialized in digital recognition. In that year, Bank of America adopted a Le Net-5 neural convolution network to recognize the handwritten digits of customers' checks. Le Net-5 is shown in Figure 2. The size of the input image is 32 * 32. After the last full connection layers, it becomes a vector. In the figure, feature maps refer to the features extracted by convolution filter. A filter extracts a feature map.

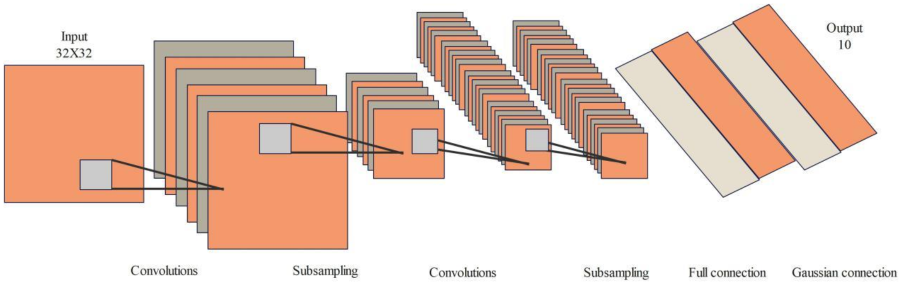

Figure 2: Le Net-5 structure diagram.

However, a layer of SoftMax classification layer is often required during classification. As shown in Figure 2, a convolutional neural network is used for classification: after the convolution layer, it passes through the pooling layer; Then pass through several of these layers; Then it passes through the full connection layer; Classified by the Soft Max classification layer. After learning layer by layer, the convolution neural network extracts the implicit features in the sample data. Compared with other deep learning models, the advantages of convolution neural network are summarized as follows: the reason for parameter sharing is that it can adopt parallel training mode, reducing the cost and complexity of network training; The two-dimensional structure information of the image can be used for combining with the view to classify the three-dimensional model.

3.2 Target Detection Based on Deep Learning

YOLO is the first phase target detection algorithm proposed. This algorithm has simple structure, fast detection speed, but low positioning accuracy. SSD algorithm, another typical representative of the first-stage algorithm. The combination of these two strategies makes SSD not only maintain the advantage of fast operation of the first-stage algorithm, but also achieve the positioning accuracy comparable to the second-stage target detection algorithm. The simulation training diagram of Fast R – CNN, SSD and YOLO algorithm is shown in Figure 3.

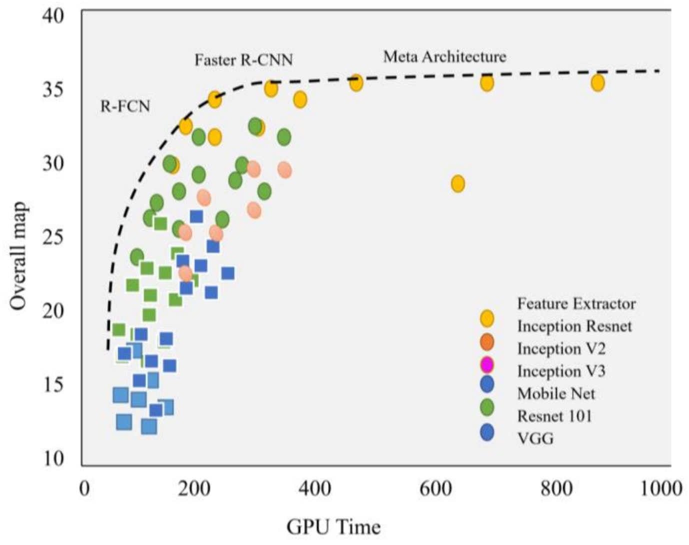

Figure 3: Simulation training diagram of deep learning algorithms.

The calculation formula for the size of other layers sm is:

$$s_m = s_{\min} + \frac{s_{\max} - s_{\min}}{M-1}(m-1), m \in [1, M] \tag{5}$$

Where M=6. Further, five different scales are set, including:

$$w_m^a = s_m \sqrt{a_r} \tag{6}$$

$$h_m^a = \frac{s_m}{\sqrt{a_r}} \tag{7}$$

In addition, for scale 1, set an additional width and height as:

$$s_m' = \sqrt{s_m s_{m+1}} \tag{8}$$

Loss function of SSD algorithm

$$L(x,c,l,g) = \frac{1}{N}\left(L_{conf}(x,c) + \alpha L_{llc}(x,l,g)\right) \tag{9}$$

Classification loss is a typical softmax loss function:

$$L_{\text{conf}}(x,c) = -\sum_{i \in P_{ax}}^{N} x_{i,j}^{p} \log\left(\hat{c}_i^p\right) - \sum_{i \in Neg} \log\left(\hat{c}_i^0\right) \tag{10}$$

$$\hat{c}_i^p = \frac{e^{c_i^p}}{\sum_p \hat{c}_i^p} \tag{11}$$

The location loss function uses SmoothLI between the predicted coordinate offset (/) and the tag offset (#).

$$\text{l o s s :} \square L_{\text{loc}}(x,l,g) = \sum_{i \in P_{os}}^{N} \sum_{m \in \{x,y,w,h\}} x_{i,j}^{k} \ \text{smooth}_{L1}\left(l_i^m - \hat{g}_j^m\right) \tag{12}$$

The calculation formula of Smooth LI "loss" is:

$$smooth_{L1}(x) = \begin{cases} 0.5x^2 \\ |\ x\ | -0.5 \end{cases} \tag{13}$$

4 RESULT ANALYSIS AND DISCUSSION

4.1 Design of Digital Visual Communication System Based on CAD

In this system, the user layer is mainly responsible for interactive operation, and the communication service module is responsible for transmitting information between the server and the client; The collaborative agent module is mainly responsible for group awareness and cooperation with other functional modules; The visual communication design retrieval module connects the data management module in the application layer in the background to realize the classification retrieval of visual communication design materials and parts and preview the retrieval results; The CAD graphics module creates a human-computer interactive operating environment, connects the file I/O module and data management module in the background, and converts the I/O of standard files to each other.

The communication coordinates the communication distribution when designing the visual communication design, realizes the directional forwarding and sending of information on the server, and can establish or remove user connections; The collaboration management module implements concurrency control, session management, communication forwarding and consistency maintenance among members when modeling tasks occur; The data management module is connected with multiple databases, and a realistic visual communication design model is built by folding algorithm, and the constructed model is directly transferred to the user layer; The CAD auxiliary module uses three CAD auxiliary technologies, namely CAD visualization technology, 3D visual modeling technology and HO visualization algorithm, to realize visual communication design.

CAD obtains the input data and initialization graphics required for structural analysis through the pre-processing module, transfers the values calculated by the calculation module to the post-processing module to output the relevant data, and constructs a three-dimensional visual model. On this basis, HO visualization algorithm is used to adjust the transparency and gray values to enhance the visual transmission effect. Combine digitization and human-computer interaction to collect geometric information of license plate materials and parts. Firstly, collect the edge corner coordinates of each license plate material, confirm the vertex and bottom points, connect the adjacent vertex and bottom points to form the side of the visual communication design material, confirm the vertex and bottom points in order to form the top and bottom surfaces, connect the points, lines and surfaces. The texture features of visual communication design materials and parts are extracted through the contribution equation:

$$A = A_0 - k\frac{\lambda_1\left(a-a_r\right)+\theta_1\left(b-b_r\right)+\varphi_1\left(c-c_r\right)}{\lambda_3\left(a-a_r\right)+\theta_3\left(b-b_r\right)+\varphi_3\left(c-c_r\right)} \tag{14}$$

$$B = B_0 - k\frac{\lambda_2\left(a-a_r\right)+\theta_2\left(b-b_r\right)+\varphi_2\left(c-c_r\right)}{\lambda_3\left(a-a_r\right)+\theta_3\left(b-b_r\right)+\varphi_3\left(c-c_r\right)} \tag{15}$$

4.2 Analysis of Experimental Results

For each GT box, as long as the IOU of the two is greater than 0.5, matching can be achieved, that is, one GT box can match multiple default boxes. Among them, IOU is calculated by Jaccard Overlap, and the Jaccard Overlap calculation method is used for Area A and Area B. Therefore, in this paper, negative samples are sorted according to the confidence level, and negative samples are selected from high to low to ensure that the proportion of negative samples: positive sample size is close to 3:1. The whole network was trained on the PC of NVIDIA's GeForce GTX965M graphics card (5.2 computing power). After more than 7 hours of training, after 6000 iterations, the loss function on the test set was basically stable at about 1.3%, and the network convergence completed the training. The change of the total loss function of network training is shown in Figure 4.

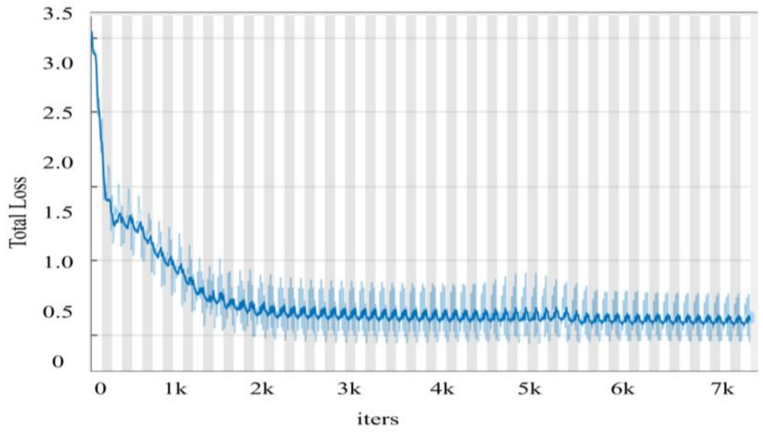

Figure 4: Training process of license plate retrieval classification network.

The binary image and its corresponding vertical projection are shown in Figure 5. It can be found that there are peaks and valleys in the vertical projection corresponding to a continuous character. The license plate was obtained by intercepting the corresponding area of each wave peak from left to right. However, when there are rivets, left and right boundaries or stains, they are slightly skewed, and the Chinese characters are not connected, it is difficult to define the peaks and valleys, and the segmentation effect of the algorithm is reduced.

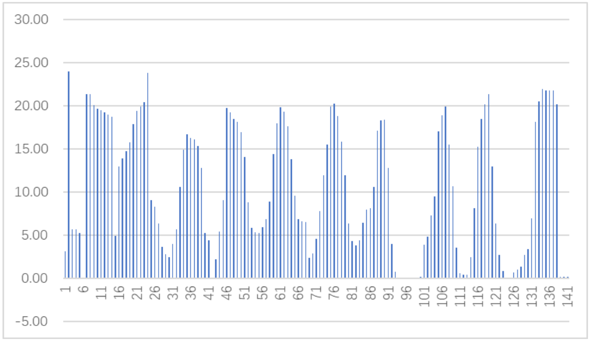

Figure 5: Binary license plate image and its vertical projection.

The system is verified from the perspective of visualization. The system is compared with the digital system of visual communication based on sensors. Sensor-based visual communication digital system and feature-based visual communication digital system draw models with high gray values, poor transparency, poor clarity and poor visual communication effect. The drawing time gradually increases, but the system in this paper always keeps within 4 s, and the drawing time is fast, while the other two systems continue to increase, and the longest drawing time has exceeded 10 s, and improve the working efficiency in the actual use process.

The processed image information is designed in plane and 3D using CAD technology. The effectiveness of this system is verified by comparing with the traditional system. The traditional system cannot realize the multi-dimensional and dynamic image without CAD technology, and the visual communication effect is poor. In order to compare the visual communication effect of this system and the traditional system, a comparative experiment was carried out, and in order to make the experimental results prominent and representative, a numerical comparison method was used for comparison.

There are also differences in the energy consumed by the system when drawing images, with the size of 1600 × For example, the energy consumed by this system and the other two systems in drawing images is shown in Figure 6. The energy consumption always remains below 0.45 J, and the trend is relatively stable. The energy consumption of the other two systems is on the rise with the increase of system use time. Among them, sensor-based vision

The transmission digital system fluctuates greatly during the working time. The system consumes the least energy and can bring better efficiency to enterprises in the actual use process.

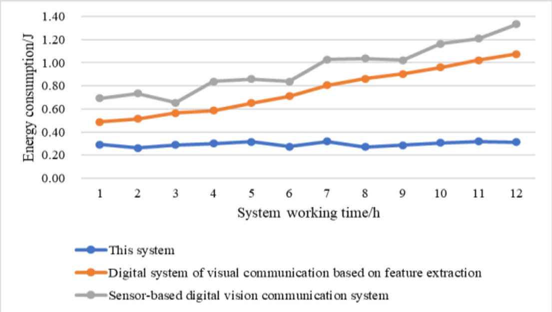

Figure 6: Comparison of energy consumption.

The recognition accuracy of the two-stage system is directly affected by the detection accuracy. When testing the recognition performance of the two-stage system, and then use the license plate image after the detection and cutting processing to carry out the recognition test. In this way, the influence of the detector on the recognition performance can be taken into account (the average edit distance calculation method is similar). For the system at one stage, the recognition result has been generated by the joint action of the recognizer and the detector. In addition, this paper points out that the recognition module in the first stage does not use STN to correct the license plate, which may lead to a decline in the recognition performance of the system in the first stage. Therefore, we have trained 7 additional license plate recognition modules with STN removed on different training sets to build a two-stage system for comparative analysis. The test method is the same as the standard two-stage system.

This section will test the whole system. The data set used here is a verification set composed of 1216 data images collected previously (the resolution is about 800 * 800, and the specific situation is described in the license plate data set in Chapter 3). Experiments are carried out on PC and Jetson TX1 embedded platform respectively. The criteria for determining the correct recognition based on the experimental results are that the IoU of the detected target prediction frame and the previously calibrated true value GT frame is more than 0.8, and the license plate classification is correct, and the license plate character recognition is correct. The experimental results of the visual communication design system using various depth learning algorithms are shown in Figure 7. It makes the extracted features more powerful, effectively filters out the background interference, and also benefits from the simplified process to avoid information loss in image processing and transmission, The accuracy has been greatly improved.

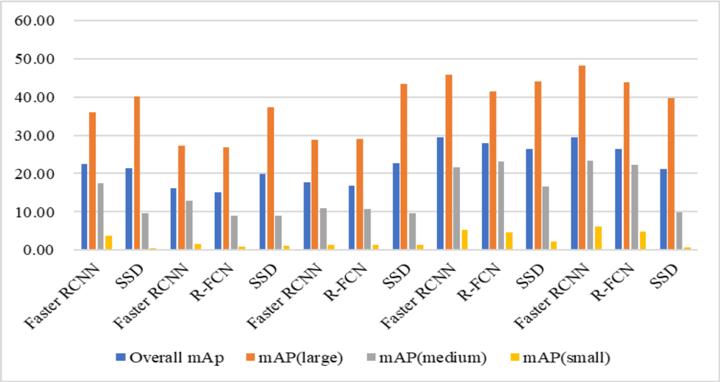

Figure 7: Experimental results of visual communication system design for each depth learning algorithm.

5 CONCLUSION

This paper designs a new automatic digital system for visual communication based on CAD technology. It sets information through CAD technology to realize image analysis, better process image feature points and complete automatic processing. The system proposed in this paper has higher processing efficiency, higher consistency between the transmitted results and the actual results, and is more suitable for practical applications. Based on the classic SSD deep learning target detection algorithm, this paper proposes an improved CAD based visual communication system design, which is applicable to the improved SSD algorithm for license plate detection. By setting the aspect ratio of the default frame that matches the license plate better, the improved SSD algorithm improves the effectiveness and efficiency of the calculation. Through experimental tests, the improved SSD algorithm shows better performance than other target detection and license plate detection algorithms in license plate detection tasks.

Yue Wu, https://orcid.org/0000-0002-8765-8192
Kun Zhao, https://orcid.org/0000-0003-2039-541X
Cong Peng, https://orcid.org/0000-0003-1944-6988

REFERENCES

[1] Ho, G.-T.-S.; Tsang, Y.-P.; Wu, C.-H.; Wong, W.-H.; Choy, K.-L.: A computer vision-based roadside occupation surveillance system for intelligent transport in smart cities, Sensors, 19(8), 2019, 1796. https://doi.org/10.3390/s19081796

[2] El-Shazli, A.-M.-A.; Youssef, S.-M.; Soliman, A.-H.: Intelligent Computer-Aided Model for Efficient Diagnosis of Digital Breast Tomosynthesis 3D Imaging Using Deep Learning, Applied Sciences, 12(11), 2022, 5736. https://doi.org/10.3390/app12115736

[3] Fan, M.; Li, Y.: The application of computer graphics processing in visual communication design, Journal of Intelligent & Fuzzy Systems, 39(4), 2020, 5183-5191. https://doi.org/10.3233/JIFS-189003

[4] Lee, J.; Lee, H.; Mun, D.: 3D convolutional neural network for machining feature recognition with gradient-based visual explanations from 3D CAD models, Scientific Reports, 12(1), 2022, 1-14. https://doi.org/10.1038/s41598-022-19212-6

[5] Mohammed, Z.-F.; Abdulla, A.-A.: An efficient CAD system for ALL cell identification from microscopic blood images, Multimedia Tools and Applications, 80(4), 2021, 6355-6368. https://doi.org/10.1007/s11042-020-10066-6

[6] Song, H.; Nguyen, A.-D.; Gong, M.; Lee, S.: A review of computer vision methods for purpose on computer-aided diagnosis, Journal of International Society for Simulation Surgery, 3(1), 2016, 1-8. https://doi.org/10.18204/JISSiS.2016.3.1.001

[7] Wang, T.; Gong, L.; Wang, C.; Yang, Y.; Gao, Y.; Zhou, X.; Chen, H.: Via: A novel vision-transformer accelerator based on fpga, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 41(11), 2022, 4088-4099. https://doi.org/10.1109/TCAD.2022.3197489

[8] Wu, H.; Li, G.: Visual communication design elements of Internet of Things based on cloud computing applied in graffiti art schema, Soft Computing, 24(11), 2020, 8077-8086. https://doi.org/10.1007/s00500-019-04171-4

[9] Yan, X.; Zhang, H.; Li, H.: Computer vision‐based recognition of 3D relationship between construction entities for monitoring struck‐by accidents, Computer‐Aided Civil and Infrastructure Engineering, 35(9), 2020, 1023-1038. https://doi.org/10.1111/mice.12536

[10] Zhang, Y.: Computer-Assisted Human-Computer Interaction in Visual Communication, Computer-Aided Design and Applications, 18(S1), 2020, 109-119. https://doi.org/10.14733/cadaps.2021.s1.109-119