Proceedings of the Third International Conference on Inventive Research in Computing Applications (ICIRCA-2021)
IEEE Xplore Part Number: CFP21N67-ART; ISBN: 978-0-7381-4627-0

Application Analysis of Image Enhancement Method in Deep Learning Image Recognition Scene

Lian Ding*, Wei Du
Chengdu Agricultural College, Wenjiang, Sichuan, China, 611130
Dinglianwenjiangca@sohu.com

2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA) | 978-1-6654-3877-3/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICIRCA51532.2021.9544905

***Abstract*— Application analysis of the image enhancement method in deep learning image recognition scene is conducted in this paper. Generally speaking, scene recognition of natural scenes is relatively difficult due to the more complex and diverse environment. It is usually done through two steps: text detection and text recognition. To enhance the traditional methods, this paper integrates the deep learning models to construct the core efficient framework for dealing with the complex data. The text method uses a sequence recognition network based on a two-way decoder based on adjacent attention weights to recognize text images and predict the output. For the further analysis, the core systematic modelling is demonstrated. The proposed model is tested on the public data sets as a reference. The experimental verification has shown the result that the proposed model is efficient.**

Keywords*— *Deep Learning; Image Recognition; Image Enhancement; Data Understanding

I. INTRODUCTION

In recent years, due to the importance of text recognition in natural scenes in a wide range of applications, it has attracted widespread attention from academia and industry. So far, the regular word recognition has been remarkably successful. The method based on the convolutional neural network has been widely used [1, 2, 3]. Many research methods combine the recursive neural network and also attention mechanism into recognition model and achieve good results [4].

Natural scene text detection is a key technology to realize intelligent scene perception and then has important research significance. However, due to the complex and also diverse backgrounds of text in natural scenes, inconsistent text fonts, inconsistent sizes, and the uncertain directions, the current processing of this task as the desired effect that has not been achieved yet. With the development of the augmented reality technology, more and more augmented reality devices have entered people's lives [5, 6, 7]. Many applications of these devices require an understanding of the scene.

The application scenarios of general low-quality character recognition include license plate recognition, the basic prefix number recognition, and value-added tax invoice recognition. Low-quality characters have the characteristics of small font size, blurred characters, and compactness.

The external reasons for low character quality include low resolution of the capture device and uneven external lighting. Hence, the figure 1 shows the traditional framework.

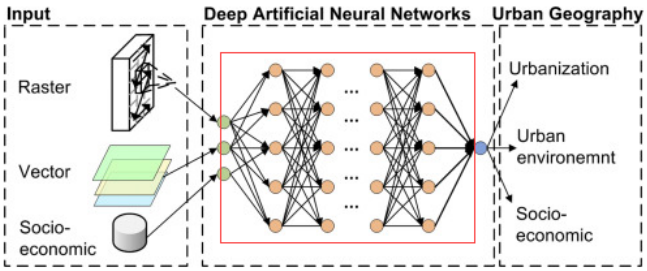

Fig. 1. The Image Recognition Scene Framework: Traditional

The brightness transformation enhancement of the image includes linear enhancement and non-linear enhancement, and the non-linear enhancement can be divided into the non-linear exponential enhancement, non-linear logarithmic enhancement, threshold function enhancement, histogram transformation enhancement, and spatial filtering enhancement. Therefore, the general study of exponential enhancement and logarithmic enhancement in nonlinear enhancement, the improvement of exponential enhancement algorithm and also the logarithmic enhancement algorithm, which can improve their reliability and practicability in actual image processing, and the further promote the application of image brightness transformation in other fields. In the next sections, the designed framework will be discussed for further analysis.

II. THE PROPOSED METHODOLOGY

A. The Deep Learning Model as Basis

In the complete data set, all the data have complete object attribute values. In an incomplete data set, all data objects have missing attribute values [8, 9, 10].

Under this condition, the automatic encoding machine is constructed using partial data of the complete data subset, and then the missing data objects are simulated, which can realize the random setting of the attribute values of the data objects of each instance to 0. By inputting incomplete objects to simulate, data can be minimized and reconstructed. In the equation 1, the basic model as reference is presented.

$$\Re(f) = \sum_{x \in X} |f(x)|^2 \ln\left(\frac{|f(x)|^2}{\|f\|_2^2}\right)\pi(x) \quad (1)$$

The goal of the core reinforcement learning algorithm is to calculate a strategy to then maximize the return from the environment. The environment defines the action space A and the state space S. Agent starts with the initial strategy, uses

multiple rollouts to evaluate it, and then updates it based on the results of these rollouts. The figure 2 shows the content.

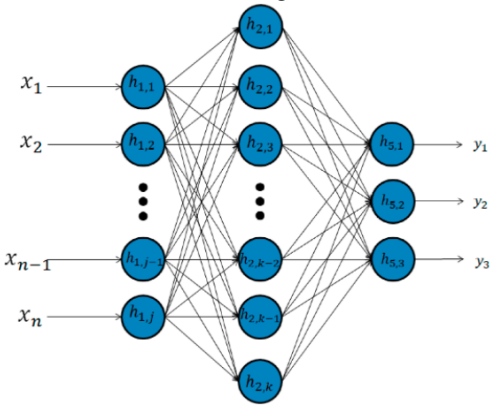

Fig. 2. The Deep Structure Model

CNN is one of first structures to successfully implement multi-layer neural networks, and has been then continuously developed to then form a variety of architectures suitable for various scenarios. In order to solve the problem of "gradient disappearance" caused by the depth of the neural network and further improve the accuracy of the training set of the network model, the ResNet structure was first proposed. However, the training of the ResNet model usually requires a huge data set to support and requires high hardware performance, and it is not easy to be widely replicated and verified. To test the further performance, the formula 2 shows the steps [11, 12, 13].

$$P(O|\lambda) = \sum_{all\ Q} P(Q|\lambda)P(O,Q|\lambda)$$
$$= \sum_{all\ Q} \pi_{q1} b_{q1}(O_1) a_{q1q2} b_{q2}(O_2) \ldots a_{qT-1qT} b_{qT}(O_T) \quad (2)$$

The actions of the tree nodes depend only on the nodes themselves, so the entire decision tree does not have to be coded in the environment state. Decisions need to be then made based on the global state, but this does not mean that the state representation needs to encode the entire decision tree.

Assembly inspection based on mixed reality technology has the following characteristics: ①Virtual information and real scenes are seamlessly superimposed, interdependent and context-sensitive, real-time interaction, no vertigo, and can effectively enhance the assembly experience of the assembler; ②No fixed camera sampling and fixed screen display, better assembly freedom, that can ensure accurate assembly while simplifying the assembly line layout [14, 15].

Hence, the figure 3 shows the selected model as the core references of the core further studies. This is because the environment builds the decision tree on a node-by-node basis, and the node only needs to consider its own state to select the corresponding action, and each node contains the rule subset of its parent node. Combining the above ideas, in the actual algorithm application process, it is necessary to complete the establishment of a three-layer network model. Each layer of the network must output parameters to the upper layer, and the uppermost layer is responsible for the extraction of feature output. After we build the decision tree, we reset the gradient, and then the algorithm iterates over all tree nodes to aggregate the gradient. Finally, DeepCut uses gradients to update the parameters of the Actor-Critic network, and then proceed to the next rollout. Figure 3 shows the model.

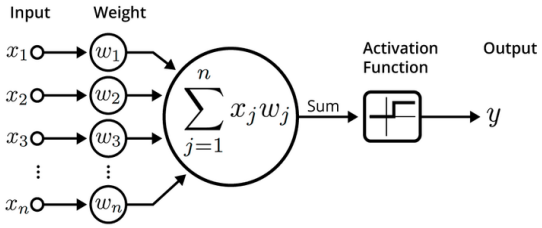

Fig. 3. The Selected Deep Learning Model

B. The Image Enhancement Method

The global night image enhancement algorithm does not consider the spatial distribution of brightness and details in the image, and uses uniform parameters and modes to enhance the image brightness, which always brings inevitable side effects.

Scholars have set their sights on local image processing methods. Sliding window overlapping algorithm is convenient for all pixels of the image and can effectively avoid the block effect, but the algorithm has low efficiency and also consumes much computing resources [16, 17, 18].

The sliding window partial overlap algorithm combines the above two characteristics, adopts larger moving step size, and takes the average pixel gray value of multiple equalization as the gray value of the final output image. In the figure 4, we show the sliding framework as the basis.

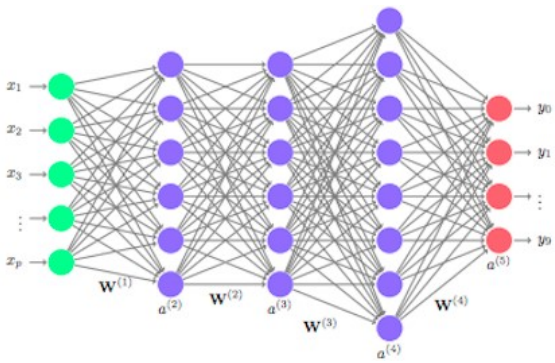

Fig. 4. The Sliding Framework for Analysis

Different from the traditional natural image denoising, the remote sensing image contains more detailed information such as the texture and edge of the ground feature. These details are very important for the remote sensing task [19].

Therefore, the difficulty of the core remote sensing image denoising lies in solving the problem of blurring the denoising result. Optimization is shown as formula 3.

$$\min x \quad \begin{cases} s.t. \sum_{j=1}^{N} x_j \lambda_j \le x \\ \sum_{j=1}^{N} y_j \lambda_j \le y_k \\ \lambda' e = 1 \end{cases} \tag{3}$$

This network uses a series of sampling operations to learn end-to-end nonlinear mapping in a core multi-scale space to directly reconstruct the denoised image.

At the same time, it also uses two techniques of high-frequency layer decomposition and learning residual mapping to compress the mapping range and simplify the difficulty of training. When using a spatial filter template to process an image, align the center of the template with a certain pixel in the image, the template is all then covered on the image, the template coefficient is multiplied by the gray value of the image pixels covered by the template, and then all The multiplication results are added together, and the calculation result is used as the new gray value of the pixels in the center of the template to cover the image. Then, the template is traversed from left to right and top to bottom sequentially on the entire image. In the end, the gray values of image pixels will be updated except for the edges of the image that cannot be completely covered by the template. The figure 5 shows the optimized structure.

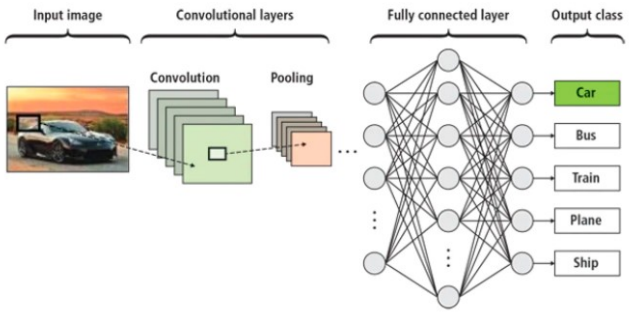

Fig. 5. Image Enhancement Structure: Optimized

C. *The Image Recognition Scene*

Multi-modal feature fusion will well combine the multiple biological features to obtain a fusion feature, which can obtain richer target identity information, thereby achieving the more reliable and accurate identification. Hololens has the small viewing angle and limited computing speed. If too many resources are consumed when performing positioning, it will affect the efficiency of some other applications and the scene reconstruction. Therefore, the signs used should have the characteristics of distinctiveness, easy identification, and fast calculation. In order to calculate the depth information, the artificial markers are pasted on some positions in the scene in advance, and the corresponding positions of these markers in the world coordinate system are obtained. When the markers are scanned, the pose can be calculated immediately.

When the user walks in the scene, the device captures these marks and immediately calculates the position of the user in the scene. The formula 4 shows the optimal.

$$\min\{\theta\}, s.t. \sum_{j=1}^{n} X_j \lambda_j \le \theta X_{j_0} \tag{4}$$

The component-based human behavior recognition method requires precise positioning of components and scene-related areas. Here we use the human body key node prediction network to effectively locate multiple body parts. There are two reasons for choosing this network: on the one hand, the key nodes basically show the position of the parts, and the bounding boxes of the parts can be then generated through processing; on the other hand, the human body key node prediction task has a very rich annotation and data set, and the human body is critical node prediction is also called pose estimation. The distance is shown as formula 5.

$$\overline{Bin}\_(n,k,\delta) def \min\{r : 1 - Bin(n,k,r) \ge \delta\} \tag{5}$$

The decoder can enhance the ability to select the correct attention area through correct attention feedback. However, there are various types of noise in natural scene images.

In practical applications, the decoder may be tricked into focusing on the blurred background area. If the decoder generates the wrong areas of the attention and selects non-corresponding features, this will result in prediction failure. In the figure 6, the optimal scene is presented.

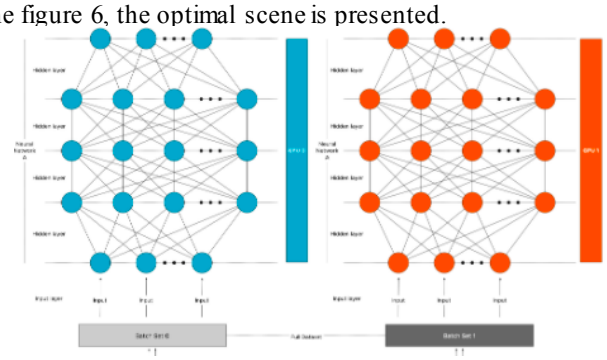

Fig. 6. Optimal Image Recognition Scene

III. VERIFICATION

Using the mean square error method to compare the errors of single-modal features has a certain statistical value, but it is difficult to describe the degree of deviation between features by calculating the mean square error of multi-modal fusion features. In this paper, the absolute error is used to calculate the statistical deviation of the fusion feature, which directly reflects the oscillation amplitude of the error feature. In the table 1, the testing on different lighting scenarios is shown.

TABLE I. TESTING ON THE LIGHTING SCENARIOS

Test Times	*Light Condition*	*Accuracy (%)*
30	general	93.55
30	general	94.28
50	general	94.12
50	general	93.97
70	general	95.12

Test Times	*Light Condition*	*Accuracy (%)*
70	general	95.53
100	general	96.85

All network layer parameters of the model are initialized randomly. The training is carried out by the stochastic gradient descent method, and the gradient is calculated by the back propagation algorithm. Since both the convolutional neural network and the recurrent neural network can be then back-propagated, the test is carried out in a complex environment. The figure 7 shows the results for references.

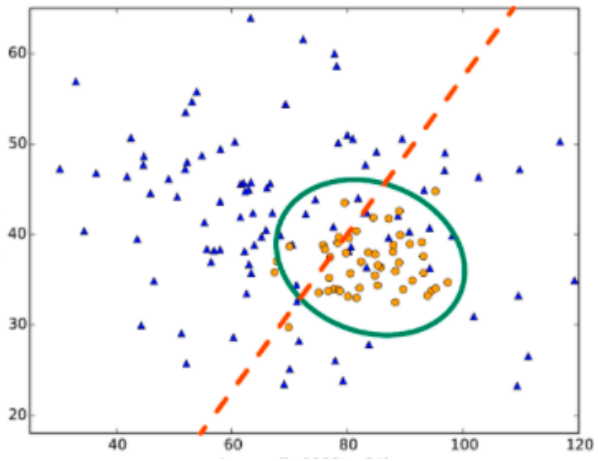

Fig. 7. Recognition Data Analysis Result

IV. CONCLUSION

Application analysis of the image enhancement method in deep learning image recognition scene is conducted in this paper. This article continues the traditional recognition ideas reflecting the core features. Based on the component-based behavior method, a general scene-component-based behavior recognition method is proposed, and the scenes are added to further improve the recognition. The model is tested on the public platforms. In the next stage, the comprehensive analysis will be considered.

V. ACKNOWLEDGEMENT

This research has been financed by The Modern design and culture research center of Sichuan philosophy and Social Sciences Key Research Base in 2018 “Application Research on information visualization design of smart campus in the context of big data -- Taking Chengdu agricultural science and Technology Vocational College as an example” （MD18C003）.

REFERENCES

[1] Hua, Y., Moua, L., Lin, J., Heidler, K. and Zhu, X.X., 2021. Aerial Scene Understanding in The Wild: Multi-Scene Recognition via Prototype-based Memory Networks. arXiv preprint arXiv:2104.11200.

[2] Wan, Z., He, M., Chen, H., Bai, X. and Yao, C., 2020, April. Textscanner: Reading characters in order for robust scene text recognition. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 34, No. 07, pp. 12120-12127).

[3] Wang, Z., Wang, L., Wang, Y., Zhang, B. and Qiao, Y., 2017. Weakly supervised patchnets: Describing and aggregating local patches for scene recognition. IEEE Transactions on Image Processing, 26(4), pp.2028-2041.

[4] Nascimento, G., Laranjeira, C., Braz, V., Lacerda, A. and Nascimento, E.R., 2017, November. A robust indoor scene recognition method based on sparse representation. In Iberoamerican Congress on Pattern Recognition (pp. 408-415). Springer, Cham.

[5] Li, T., Lei, S., Wang, W. and Wang, Q., 2020, August. Research on MR virtual scene location method based on image recognition. In 2020 International Conference on Information Science, Parallel and Distributed Systems (ISPDS) (pp. 109-113). IEEE.

[6] Anwer, R.M., Khan, F.S., van de Weijer, J., Molinier, M. and Laaksonen, J., 2018. Binary patterns encoded convolutional neural networks for texture recognition and remote sensing scene classification. ISPRS journal of photogrammetry and remote sensing, 138, pp.74-85.

[7] Zuo, L.Q., Sun, H.M., Mao, Q.C., Qi, R. and Jia, R.S., 2019. Natural scene text recognition based on encoder-decoder framework. IEEE Access, 7, pp.62616-62623.

[8] Horiguchi, S., Amano, S., Ogawa, M. and Aizawa, K., 2018. Personalized classifier for food image recognition. IEEE Transactions on Multimedia, 20(10), pp.2836-2848.

[9] Wang, S., Wang, R., Yao, Z., Shan, S. and Chen, X., 2020. Cross-modal scene graph matching for relationship-aware image-text retrieval. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (pp. 1508-1517).

[10] Xu, S., Mu, X., Chai, D. and Zhang, X., 2018. Remote sensing image scene classification based on generative adversarial networks. Remote sensing letters, 9(7), pp.617-626.

[11] Zhou, B., Lapedriza, A., Khosla, A., Oliva, A. and Torralba, A., 2017. Places: A 10 million image database for scene recognition. IEEE transactions on pattern analysis and machine intelligence, 40(6), pp.1452-1464.

[12] Luo, C., Jin, L. and Sun, Z., 2019. Moran: A multi-object rectified attention network for scene text recognition. Pattern Recognition, 90, pp.109-118.

[13] Liu, Y., Jin, L., Zhang, S., Luo, C. and Zhang, S., 2019. Curved scene text detection via transverse and longitudinal sequence connection. Pattern Recognition, 90, pp.337-345.

[14] López-Cifuentes, A., Escudero-Viñolo, M., Bescós, J. and García-Martín, Á., 2020. Semantic-aware scene recognition. Pattern Recognition, 102, p.107256.

[15] Xinyu, L., Xiaochun, L., Rongfeng, C., Yizhou, F., Tianmin, X. and Junyan, C., 2019, December. Application of the faster R-CNN algorithm in scene recognition function design. In 2019 15th International Conference on Computational Intelligence and Security (CIS) (pp. 16-19). IEEE.

[16] Cheng, B., Chen, L.C., Wei, Y., Zhu, Y., Huang, Z., Xiong, J., Huang, T.S., Hwu, W.M. and Shi, H., 2019. Spgnet: Semantic prediction guidance for scene parsing. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 5218-5228).

[17] Ji, J., Krishna, R., Fei-Fei, L. and Niebles, J.C., 2020. Action genome: Actions as compositions of spatio-temporal scene graphs. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 10236-10247).

[18] Bhunia, A.K., Kumar, G., Roy, P.P., Balasubramanian, R. and Pal, U., 2018. Text recognition in scene image and video frame using Color Channel selection. Multimedia tools and applications, 77(7), pp.8551-8578.

[19] Shi, B., Yang, M., Wang, X., Lyu, P., Yao, C. and Bai, X., 2018. Aster: An attentional scene text recognizer with flexible rectification. IEEE transactions on pattern analysis and machine intelligence, 41(9), pp.2035-2048.