WEP Warwick Evans Publishing

International Journal of Computer Science and Information Technology

ISSN: 3005-9682 (Print), ISSN: 3005-7140 (Online) | Volume 2, Number 1, Year 2024
DOI: https://doi.org/10.62051/ijcsit.v2n1.06
Journal homepage: https://wepub.org/index.php/IJCSIT/index

Computer Vision and Deep Learning Transforming Image Recognition and Beyond

Yizhi Chen[1,*], Sihao Wang[2], Luqi Lin[3], Zhengrong Cui[4], Yanqi Zong[5]

[1]Information Studies, Trine University, Allen Park, MI, USA;

[2]Mathematics, Southern Methodist University, Dallas, USA;

[3]Software Engineering, Sun Yat-sen University, Shanghai, China;

[4]Software Engineering, Northeastern University, Shanghai, China;

[5]Information Studies, Trine University, Phoenix, AZ, USA.

*Corresponding Author: eizyc66@gmail.com

ABSTRACT

Computer vision is a cutting-edge information processing technology that seeks to mimic the human visual nervous system. Its primary aim is to emulate the psychological processes of human vision to interpret and depict objective scenery. This revolutionary field encompasses a wide range of applications, including life sciences, medical diagnosis, military operations, scientific research, and many others. At the heart of computer vision lies the theoretical core, which includes deep learning, image recognition, target detection, and target tracking These elements combine to enable computers to process, analyze, and understand images, allowing for the classification of objects based on various patterns One of the standout advantages of deep learning techniques, when compared to traditional methods, is their ability to automatically learn and adapt to the specific features required for a given problem. This adaptive nature of deep learning networks has opened up new possibilities and paved the way for remarkable breakthroughs in the field of computer vision. This paper examines the practical application of computer vision processing technology and convolutional neural networks (CNNs) and elucidates the advancements in artificial intelligence within the field of computer vision image recognition. It does so by showcasing the tangible benefits and functionalities of these technologies.

KEYWORDS

Computer Vision; Deep Learning; I mage Recognition; Applications

1. INTRODUCTION

Image recognition stands out as a crucial application within the realm of computer vision. It empowers computers to swiftly identify and categorize objects in images, with real-world implications spanning multiple industries. For instance, in the medical field, image recognition can rapidly pinpoint and classify diseases within human body scans. In the realm of security monitoring, it can swiftly detect and identify abnormal behaviors by analyzing real-time surveillance footage. Furthermore, in the emerging field of autonomous vehicles, image recognition plays a pivotal role in enabling self-driving cars to perceive and comprehend their surroundings.The advent of deep learning has significantly advanced the capabilities of computer vision. Specifically, supervised Convolutional Neural Networks (CNNs) have proven to be immensely successful in tasks like image classification

and object detection. These networks have also sparked exploration into pixel-level markup, notably in the area of semantic segmentation.

In conclusion, computer vision, driven by deep learning and image recognition technologies, has revolutionized how we process, analyze, and understand visual information. Its wide-ranging applications hold promise for numerous industries, making it a transformative force in the realms of science, healthcare, security, and transportation. As research in this field continues to advance, we can expect even greater breakthroughs that will reshape the way we interact with and interpret the visual world around us.

2. RELATED WORK

2.1. Computer vision

The principle of computer vision mainly includes image acquisition, image preprocessing, feature extraction, image recognition and image understanding. Image acquisition is the basis of computer vision, which uses cameras, sensors and other devices to convert images in the real world into digital signals for computer processing. The quality of image acquisition directly affects the effect of subsequent processing, so it is very important to select the appropriate equipment and acquisition parameters. Image preprocessing is to process the acquired image, including denoising, enhancing, sizing and so on. Denoising can be realized by filtering algorithms, such as median filtering, Gaussian filtering, etc. Enhancement can improve image quality by adjusting contrast, degree, etc. Sizing allows you to scale an image to the right size. The purpose of image preprocessing is to reduce the noise, highlight the key information in the image, and prepare for the subsequent feature extraction.

Therefore, feature extraction is the core of computer vision. Features are distinctive and descriptive information in an image that can be used to distinguish between different objects or scenes. Commonly used feature extraction methods include edge detection, corner detection, texture analysis and so on. Edge detection can extract edge information by looking for pixels with large changes in light and dark in the image. Corner detection can extract corner information by detecting extreme value points in images. Texture analysis can extract texture features by statistical distribution of pixels in the image. The purpose of feature extraction is to convert complex image information into simple feature vectors, which is convenient for subsequent image recognition and understanding.

2.2. Application of computer vision image recognition

Image recognition is one of the important applications of computer vision. Image recognition refers to the analysis and comparison of images by computers to find out the objects or scenes in the images. Common image recognition methods include template matching, machine learning and deep learning. Template matching is a simple recognition method, which is used to calculate the similarity between the image and the known template. Machine learning is a recognition method based on statistical methods. It builds a classification model by training samples, and then compares the image to the model. Deep learning is a recognition method based on neural network, which learns the features and patterns in images through multi-level network structure, so as to achieve high-precision image recognition.

Image understanding is the highest level of computer vision, which is the process of semantic analysis and reasoning of images. Image understanding includes object detection, scene understanding and behavior analysis. Day mark detection is to find out the position and boundary of a specific object in the image; Scene understanding refers to the analysis and description of the scene in the image. Behavior analysis refers to the action recognition and behavior analysis of people or objects in images. The main purpose of image understanding is to convert the content of the image into a form that the computer can understand and process, which provides the basis for the computer to make decisions

and interactions. The principle of computer vision mainly includes image acquisition, image preprocessing, feature extraction, image recognition and image understanding. Through the acquisition, processing, analysis and understanding of images, computer vision can realize the ability to automatically recognize, analyze and understand images, providing strong support and applications for artificial intelligence and robotics.

2.3. Image recognition implementation principle

Image classification is one of the most classic tasks in the field of computer vision, the purpose of which is to correspond the input image to a predefined semantic category, that is, to label the category. The traditional image classification method consists of basic feature learning, feature coding, space constraint, classifier design, model fusion and so on.

First, extract features from the image, Classic feature extraction methods include HOG(Histogram of Oriented Gradient) and LBP(Local Bianray Pattern). Local binary mode), SIFT(Scale-Invariant Feature Transform), etc., can also fuse multiple features to retain more useful information. Then, after the feature is encoded, the redundancy and noise are removed, and the feature coding is generated. The classical methods include sparse coding, local linear constraint coding, Fisher vector coding, etc. Then, the feature aggregation is realized after the spatial feature constraint, such as the classic pyramid feature matching method. Finally, classifiers are used for classification. Classical classifiers include SVM, random forest and so on.

The pre-activation Residual Unit, ResNeXt, and Inception were used as the basic units of the residual attention network to construct the attention module. Given the output T(x) and input x of the trunk branch, the mask branch uses a bottom-up, top-down structure to learn the same size mask M(x). The bottom-up, top-down structure mimics rapid feedforward and feedback attention processes. The output mask is used as a control gate for neurons in the main branch. The output of the Attention module H is:

$$H_{i,c}(x) = M_{i,c}(x) * T_{i,c}(x) \tag{1}$$

Where the range of i is all spatial locations, $c \in \{1,... C\}$ is the index of the channel. The entire structure can be trained end-to-end. In the attention module, the attention mask can be used not only as a feature selector in the forward inference process, but also as a gradient update filter in the backpropagation process. In the soft mask branch, the mask gradient of the input feature is:

$$\frac{\partial M(x,\theta)T(x,\phi)}{\partial \phi} = M(x,\theta)\frac{\partial T(x,\phi)}{\partial \phi} \tag{2}$$

Where θ is the mask branch parameter and φ is the trunk branch parameter. This property makes the attention module robust to noise labels. Mask branching prevents false gradients (from noise labels) to update backbone parameters.

Unlike the stacked attention modules we designed, there is an easy way to generate soft weight masks using a single network branch, similar to the spatial transformer layer. However, these methods have several drawbacks on challenging data sets such as ImageNet. First, images with chaotic background images, complex scenes, and large changes in appearance need to be modeled with different types of attention.

$$H_{i,c}(x) = (1 + M_{i,c}(x)) * F_{i,c}(x) \tag{3}$$

The range of M(x) is [0, 1], M(x) approximates 0, and H(x) will approximate the original feature F(x). The authors call this approach attentional residual learning.

The mask branch consists of fast feedforward scanning and top-down feedback steps. The former operation quickly gathers global information about the entire image, while the latter operation combines the global information with the original feature map. In a convolutional neural network,

these two steps unfold into a bottom-up, top-down, fully convolutional structure. From the input, perform a few max-pooling sessions to rapidly increase the receptive field after a small Residual Units. Once the minimum resolution is reached, the global information is expanded through a symmetrical top-down architecture to guide the input features at each location. The output is upsampled with linear interpolation after Residual Units. The amount of bilinear interpolation is the same as that of max-pooling so that the output size is the same as the input feature map. Then, after two successive 1x1 convolution layers, the SIGMOid-type layer normalizes the output range to [0,1].

3. METHODOLOGY

Image-based systems are becoming popular for collecting pavement condition data for pavement management activities, with pavement engineers defining various distress categories depending on the type of pavement. However, today's software solutions have limitations in automatically identifying pavement types correctly from the collected images.

In this paper, the road recognition system PvmtTPNet based on convolutional neural network (CNN) has acceptable consistency, accuracy and high efficiency in automatic recognition of road types.

3.1. Training data

There are three types of pavement commonly assessed and measured in PMS: asphalt concrete pavement (AC), jointed concrete pavement (JPCP), and continuous reinforced concrete pavement (CRCP). A total of 21,000 two-dimensional (2D) images were collected, covering 84,000 meters (52.20 miles) of long road slices. 80% of the prepared images were randomly selected for the training of the proposed network, and the remaining 20% were used for testing. In the training process, the prepared two-dimensional image is reduced to 475×512 two-dimensional image to improve the computational efficiency. Figure 2 is a graphical example of a pre-prepared data set.

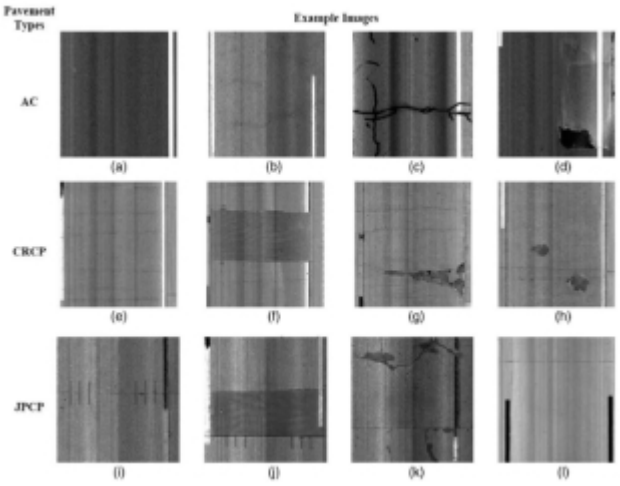

Figure 1: A graph of a dataset is an example

3.2. Architecture of Convolutional PvmtTPNet

PvmtTPNet consists of six layers: three convolutional layers, two fully connected layers, and an output layer. The input side of PvmtTPNet is the prepared two-dimensional pavement image, and the output layer calculates the probability distribution of the predicted pavement type. In each convolutional layer, eight cores of size 13×13 are used to extract features of the input image, such as edges and shapes. For these two fully connected layers, we implemented 32 nodes and 16 nodes, respectively, to preserve the features of the most important pavement images.

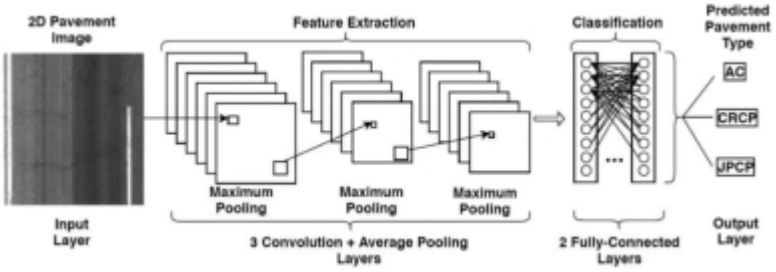

Figure 2: Convolutional architecture for image recognition

In the network training process, a combination of different techniques is used to adjust the hyperparameters in PvmtTPNet according to the prepared two-dimensional images. The parameters of the network are gradually adjusted to reduce the error between the output score and the expected score pattern, so as to reduce training losses and improve training accuracy (LeCunetal.2015). After extensive training, PvmtTPNet is able to predict the pavement type for a given two-dimensional image based on a scoring vector, where the highest scores for all categories will correspond to the pavement type.

3.3. Training result

The classification accuracy and cross entropy losses of network training and testing are shown in Figure 3. With the increase of the number of training cycles, the classification accuracy increases and the cross entropy loss decreases. Training on PvmtTPNet's 100 era readiness dataset takes 28 hours to complete on a NVIDIAtitanVGPU card. Through the selection and combination of training techniques, the classification accuracy of test data is still close to that of training data, which indicates that there are few overfitting problems in the network.

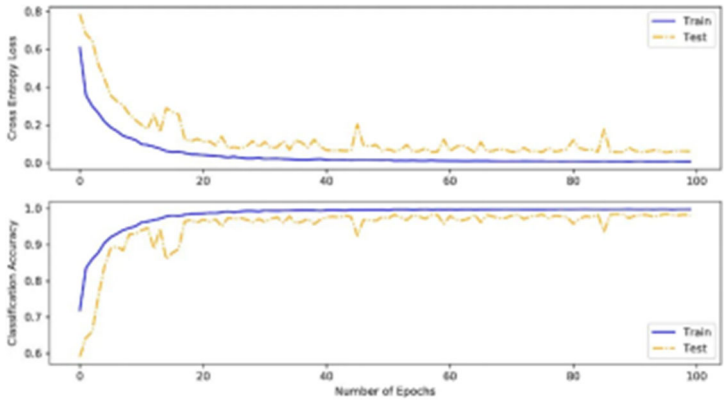

Figure 3: ROC curve of the training model

Meanwhile, the cross-entropy loss of training data and test data is 0.0067 and 0.054, respectively. Therefore, the parameters derived in stage 96 are considered to be the optimal parameters for PvmtTPNet. The classification accuracy of training data in the optimal period reaches 99.83%.

The conclusion of this experiment:

1. This study develops a deep learning (DL) based network, called PvmtTPNet, that can automatically identify pavement types from images to facilitate fully automated pavement condition surveys. PvmtTPNet implements an architecture based on convolutional neural networks to learn the features of images from the pavement class.

2. The current calibration linear unit (ReLUs) is used as the activation function of the convolutional layer and the fully connected layer, which can be trained quickly and effectively, and has become the default activation function of modern deep learning neural networks.

4. CONCLUSION

This paper explores in depth the importance and application of computer vision and deep learning techniques in image recognition and other fields. Computer vision is not just an information processing technology, it is also a revolutionary evolution of the field, designed to mimic the function of the human visual nervous system, to achieve the interpretation and presentation of objective scenes. In various fields, from life sciences to medical diagnostics, from military operations to scientific research, computer vision plays an important role. The successful application of deep learning and convolutional neural networks in image recognition, especially in the fields of image classification and object detection, has brought great advances in computer vision. Compared with traditional methods, deep learning techniques are characterized by automatic learning and adaptive problems, opening up new possibilities for breakthroughs in the field of computer vision.

In the practical application of computer vision, image recognition is a key field, which can be widely used in medicine, security monitoring, automatic driving and other fields. This paper mentions that image recognition in the medical field can quickly identify diseases in human body images, image recognition in safety monitoring can quickly detect abnormal behaviors, and image recognition in automatic driving can realize environmental perception and understanding. These practical cases demonstrate the wide application and importance of computer vision technology.

In terms of the future outlook, computer vision and deep learning technologies are expected to continue to evolve, bringing more innovation and progress to various fields. As research continues, we can expect more breakthroughs that will further change the way we understand and interact with the visual world. Computer vision has become an important part of the modern science and technology field, and its potential and prospects are exciting, and will continue to promote the development of artificial intelligence and deep learning, bringing more convenience and innovation to our life and work.

ACKNOWLEDGEMENT

In this article about artificial intelligence and computer vision technology, I have learned a lot of Ms. Yulu Gong's research results and theories, and these contributions have provided valuable resources and inspiration for the writing of this paper. Here, I would like to express my heartfelt thanks to Ms. Yulu Gong. Ms. Yulu Gong's research has had a profound impact on the development of the field of artificial intelligence, and her results are not only widely recognized in the academic community, but also have great potential in practical applications. The experimental and theoretical part of this paper benefits from her core views and research results, which provide me with profound inspiration and help me better understand and elaborate the topic of this paper.

I would like to thank Ms. Yulu Gong for her selfless dedication and outstanding work. I deeply admire her for her contribution to the advancement of the field of artificial intelligence. At the same time, I would like to thank all the scholars and researchers who have supported and inspired me in this paper. Your work has provided important support for the improvement and deepening of this paper. Finally, I would like to express my heartfelt thanks again to Ms. Yulu Gong and all relevant researchers for their contributions. We hope that our efforts can make a modest contribution to the future development of artificial intelligence.

REFERENCES

[1] “Based on Intelligent Advertising Recommendation and Abnormal Advertising Monitoring System in the Field of Machine Learning”. International Journal of Computer Science and Information Technology, vol. 1, no. 1, Dec. 2023, pp. 17-23, https://doi.org/10.62051/ijcsit.v1n1.03.

[2] Yu, Liqiang, et al. “Research on Machine Learning With Algorithms and Development”. Journal of Theory and Practice of Engineering Science, vol. 3, no. 12, Dec. 2023, pp. 7-14, doi:10.53469/jtpes.2023.03(12).02.

[3] Huang, J., Zhao, X., Che, C., Lin, Q., & Liu, B. (2024). Enhancing Essay Scoring with Adversarial Weights Perturbation and Metric-specific AttentionPooling. arXiv preprint arXiv:2401.05433.

[4] Tan, Kai, et al. “Integrating Advanced Computer Vision and AI Algorithms for Autonomous Driving Systems”. Journal of Theory and Practice of Engineering Science, vol. 4, no. 01, Jan. 2024, pp. 41-48, doi:10.53469/jtpes.2024.04(01).06.

[5] Tianbo, Song, Hu Weijun, Cai Jiangfeng, Liu Weijia, Yuan Quan, and He Kun. "Bio-inspired Swarm Intelligence: a Flocking Project With Group Object Recognition." In 2023 3rd International Conference on Consumer Electronics and Computer Engineering (ICCECE), pp. 834-837. IEEE, 2023.DOI: 10.1109/mce.2022.3206678.

[6] “The Application of Artificial Intelligence in Medical Diagnostics: A New Frontier”. Academic Journal of Science and Technology, vol. 8, no. 2, Dec. 2023, pp. 57-61, https://doi.org/10.54097/ajst.v8i2.14945.

[7] Pan, Yiming, et al. “Application of Three-Dimensional Coding Network in Screening and Diagnosis of Cervical Precancerous Lesions”. Frontiers in Computing and Intelligent Systems, vol. 6, no. 3, Jan. 2024, pp. 61-64, https://doi.org/10.54097/mi3VM0yB.

[8] Liu, B., Zhao, X., Hu, H., Lin, Q., & Huang, J. (2023). Detection of Esophageal Cancer Lesions Based on CBAM Faster R-CNN. Journal of Theory and Practice of Engineering Science, 3(12), 36–42. https://doi.org/10.53469/jtpes.2023.03(12).06.

[9] K. Jin, Z. Z. Zhong and E. Y. Zhao, "Sustainable Digital Marketing Under Big Data: An AI Random Forest Model Approach," in IEEE Transactions on Engineering Management, vol. 71, pp. 3566-3579, 2024, doi: 10.1109/TEM.2023.3348991.

[10] Liu, Bo, et al. "Integration and Performance Analysis of Artificial Intelligence and Computer Vision Based on Deep Learning Algorithms." arXiv preprint arXiv:2312.12872 (2023).

[11] Jin, Keyan. "Impacts of Word of Mouth (WOM) on E-Business Online Pricing." JGIM vol.31, no.3 2023: pp.1-17. http://doi.org/10.4018/JGIM.324813.

[12] Yu, L., Liu, B., Lin, Q., Zhao, X., & Che, C. (2024). Semantic Similarity Matching for Patent Documents Using Ensemble BERT-related Model and Novel Text Processing Method. arXiv preprint arXiv:2401.06782.

[13] “The Application of Artificial Intelligence to The Bayesian Model Algorithm for Combining Genome Data”. Academic Journal of Science and Technology, vol. 8, no. 3, Dec. 2023, pp. 132-5, https://doi.org/10.54097/ykhccb53.

[14] Wei, Kuo, et al. “Strategic Application of AI Intelligent Algorithm in Network Threat Detection and Defense”. Journal of Theory and Practice of Engineering Science, vol. 4, no. 01, Jan. 2024, pp. 49-57, doi:10.53469/jtpes.2024.04(01).07.

[15] Du, Shuqian, et al. “Application of HPV-16 in Liquid-Based Thin Layer Cytology of Host Genetic Lesions Based on AI Diagnostic Technology Presentation of Liquid”. Journal of Theory and Practice of Engineering Science, vol. 3, no. 12, Dec. 2023, pp. 1-6, doi:10.53469/jtpes.2023.03(12).01.

[16] Xin, Q., He, Y., Pan, Y., Wang, Y., & Du, S. (2023). The implementation of an AI-driven advertising push system based on a NLP algorithm. International Journal of Computer Science and Information Technology, 1(1), 30-37.0.

[17] Pan, Yiming, et al. “Application of Three-Dimensional Coding Network in Screening and Diagnosis of Cervical Precancerous Lesions”. Frontiers in Computing and Intelligent Systems, vol. 6, no. 3, Jan. 2024, pp. 61-64, https://doi.org/10.54097/mi3VM0yB.

[18] “Enhancing Computer Digital Signal Processing through the Utilization of RNN Sequence Algorithms”. International Journal of Computer Science and Information Technology, vol. 1, no. 1, Dec. 2023, pp. 60-68, https://doi.org/10.62051/ijcsit.v1n1.09.

[19] “Implementation of Computer Vision Technology Based on Artificial Intelligence for Medical Image Analysis”. International Journal of Computer Science and Information Technology, vol. 1, no. 1, Dec. 2023, pp. 69-76, https://doi.org/10.62051/ijcsit.v1n1.10.

[20] Dong, Xinqi, et al. “The Prediction Trend of Enterprise Financial Risk Based on Machine Learning ARIMA Model”. Journal of Theory and Practice of Engineering Science, vol. 4, no. 01, Jan. 2024, pp. 65-71, doi:10.53469/jtpes.2024.04(01).09.

[21] “A Deep Learning-Based Algorithm for Crop Disease Identification Positioning Using Computer Vision”. International Journal of Computer Science and Information Technology, vol. 1, no. 1, Dec. 2023, pp. 85-92, https://doi.org/10.62051/ijcsit.v1n1.12.

[22] Wang, Sihao, et al. “Diabetes Risk Analysis Based on Machine Learning LASSO Regression Model”. Journal of Theory and Practice of Engineering Science, vol. 4, no. 01, Jan. 2024, pp. 58-64, doi:10.53469/jtpes.2024.04(01).08.