International Journal of Science and Engineering Applications
Volume 12-Issue 11, 09 – 11, 2023, ISSN:- 2319 - 7560
DOI: 10.7753/IJSEA1211.1003

Exploring the Application Path of Machine Learning in Computer Vision Processing

Zhu Zeyu
Guangling College of Yangzhou University
225000, Jiangsu, China

Abstract: In the current economic development of society, with the application and development of machine learning algorithms, computer vision processing technology based on this has also become a key form of technology in the field of artificial intelligence. Reasonably applying machine learning algorithms to computer vision processing can make it more suitable for human thinking and meet practical visual processing needs. Has won good honors in many large-scale recognition studies. This article mainly studies the main applications of deep convolutional neural networks in computer vision. Analyze the pooling operation and image classification object detection of deep convolutional networks and promote the application and development of deep convolutional neural networks in computer vision.

Keywords: Machine Learning, Application Path, Computer Vision Processing

1. INTRODUCTION

Convolutional Neural Network (CNN) is a well-known deep learning architecture inspired by the natural visual perception mechanisms of organisms. In 1959, Hubel et al. discovered that cells in the visual cortex of animals were responsible for detecting light in the receptive field. Inspired by this discovery, Japanese scientist Fukushima proposed a hierarchical multi-layer artificial neural network called the neurocognitive machine around 1980. The neurocognitive machine model is composed of various types of cell units, with the most important two being called "S-type cells" and "C-type cells".

Since 2014, many machine learning frameworks have been applied to image detection in computer vision processing, such as R-CNN framework, FastR-CNN framework, FastR-CNN framework, YOLO framework, and SSD framework. Among the machine learning image detection frameworks mentioned above, YOLO framework has the highest detection speed. Through practical research, it has been found that its detection speed can reach 155fps/s, but its detection accuracy is the lowest, only 52.7; Although the Faster CNN framework has the highest detection accuracy, its detection speed is very slow. Compared to other detection frameworks, the SSD framework has advantages in both detection accuracy and detection speed.

Therefore, in specific computer vision processing, the SSD framework can be used as its image detection framework. Essentially, this convolutional neural network is the first successfully developed multi-layer neural network, and this algorithm model is more conducive to network input of multiple micro signals. At the same time as learning gradually deepens, there is a wave of information learning. Currently, convolutional neural networks have been preliminarily applied to large-scale and different machine learning applications such as natural speech processing, image recognition, and speech recognition.

The application of bionics and engineering methods. In practical applications of computer vision processing, machine learning mainly simulates human learning behavior to obtain new knowledge and skills, and summarizes and organizes existing knowledge structures, thereby continuously improving its performance in the processing process. Artificial intelligence is a key part of the combination of machine learning and computer vision processing and is one of the important means to achieve intelligent computer vision processing. In specific combinations, machine learning techniques of bionics and engineering can be used to effectively implement various functions of computer vision processing.

2. THE PROPOSED METHODOLOGY

2.1 The main applications of mechanical learning in computer vision processing

The use of biomimetic technology can effectively simulate human visual and learning abilities. A systematic introduction to the basic components and principles of CNN is provided. Section 3 elaborates and discusses the latest research progress in various aspects of CNN, such as convolutional layers, pooling layers, activation functions, etc. in recent years. Section 4 summarizes representative CNN architectures since 1998; Section 5 introduces the application of CNN in image classification/localization, object detection, object segmentation, object tracking, behavior recognition, and image super-resolution reconstruction. Finally, prospects are made for the future research directions of CNN. In the process of processing photos, computers can use corresponding algorithms to segment semantic graphics, while also making reasonable distinctions between various main elements.

To achieve this goal, a sufficiently powerful building block is needed, which is to predict the pixel distribution in various classified images by training classifiers. This task poses many computational challenges for machine learning, especially in computers with large pixel counts, where image classification tasks require over a million training and testing sessions. The features collected from the convolutional layer can be input into the classifier for training, in theory, inputting the various information features collected by the convolutional layer into the classifier requires a lot of calculation, especially in larger image resolutions, to obtain the final calculation classification results.

However, due to the certain characteristics of local areas in the image, it is likely to be used in another field. Therefore, it is necessary to perform feature aggregation statistics on the local positions of the image, which is a pooling operation. Artificial intelligence is a key part of the combination of machine learning and computer vision processing and is one of the important means to achieve intelligent computer vision processing. In specific combinations, machine learning techniques of bionics and engineering can be used to effectively implement various functions of computer vision processing. The use of biomimetic technology can effectively simulate human visual and learning abilities. Typical pooling operations include average pooling and maximum pooling.

The maximum pooling function takes the maximum value of elements in a block as the output of the function, extracting the local maximum response of the feature plane. It is usually used for extracting low-level features and selecting the most prominent features from the input feature map. The mean pooling function takes the arithmetic mean of all elements in the calculated block as the output of the function and extracts the mean of the local response of the feature plane. The so-called artistic style transfer refers to extracting style from an existing image, such as in Van Gogh's "Night Sky", and then importing another image with other content and styles, such as a city's architectural complex. Then let the system draw the urban architectural complex again in the style of 'Night Sky'.

2.2 Art Style Transfer and Introduction of Machine Learning Algorithms

Although humans can easily recognize the style features in images, for computers, how to convert the style of one image into the style of another is an equally complex and abstract problem. Traditional image art style transfer methods are difficult to meet the requirements of practical applications in terms of visual effects. In pooling operations, if a continuous range of images is selected as the pooling position, the same neural network will appear between the two, resulting in the application of convolutional features. Therefore, these pooling works have a certain degree of translation invariance and can consistently output the same classification results within the same features and classifiers.

Compared with the convolutional features, these classification results can effectively reduce the working dimensions of the feature vectors and reduce the computational workload, enabling effective expansion and supplementation of the training data and avoiding its strong fitting effect. Convolutional layers are an indispensable part of convolutional neural network architecture, mainly used for learning feature representations of input images. Therefore, researchers are constantly trying to improve the convolutional layer in CNN architecture to improve network performance. Below are some key innovative measures in this regard. After importing an image P into the VGG (Convolutional Layer) machine learning network, a series of vectors are obtained in the first layer of the network, and intermediate vectors are obtained in each subsequent network layer. Each pixel in the network is composed of three values: red, green, and blue, representing image features.

Because VGG19 belongs to a machine learning network that has completed a series of simulated human visual system training, and the parameters have been determined, the intermediate vector obtained through parameter calculation can be used to represent this image. In this case, the feature map within a certain convolutional layer can be defined as the content of the image. Most convolutional neural network models require image size data input, but it is easy to lose the original data in the image information during image cropping; Or adjust the aspect ratio and size of the image to avoid deformation and distortion. And pay attention to whether the roll base in the input image size has a constraint effect, ensuring that the dimensionality is fixed during the input process.

Since the application of deep learning in the ILSVRC2012 image classification competition and achieving good results, this deep learning model method has gradually been promoted in image recognition. Moreover, the emergence of new neural network models is constantly comforting their performance, promoting the rapid improvement of image feature learning in this network model. Spatial pyramid pooling (SPP) was proposed by The et al. in 2014. The key advantage of SPP is that it can generate fixed size feature vectors regardless of the size of the input feature map, and then input them into the fully connected layer. SPP will perform pooling operations on local areas in the input feature map that are proportional to the image size, in order to obtain fixed size feature vectors.

This is different from the pooling of sliding windows in previous deep networks, where the number of sliding windows depends on the size of the input image. By replacing the last Spooling layer with SPP, He aiming et al. proposed a new SPP-Net that can handle images of different sizes. Compared to the definition of image content, the definition of image style has higher difficulty. In style definition, it is not possible to randomly select a feature map within a certain layer as a style layer. Instead, it is necessary to take all feature maps within a layer and multiply them in pairs to obtain a Gram matrix, which mainly includes image color information and texture information. This matrix is the image style.

3. CONCLUSION

In the specific application of computer vision processing technology, machine learning algorithms have very good application advantages. One solution to early computer vision problems was through mathematical modeling and analysis methods. However, with the rapid development of machine learning in recent years, the combination of computer vision and machine learning has attracted more widespread attention from researchers, achieving a significant leap in the field of computer vision. Currently, people's use of deep learning is only limited to the application of simple reasoning calculations, and good research results have been achieved in the field of image and speech. This also indicates that with the in-depth research and feature extraction of convolutional neural networks, they can more effectively represent some of their features in other fields, and with the development of complex reasoning, they will delve into more aspects of artificial intelligence operations.

4. REFERENCES

[1] Wang Song. Scene Analysis and Reconstruction Based on Machine Learning [D]. Jiangnan University [2023-10-22]. DOI: CNKI: CDMD: 21018.063584

[2] Lu Hongtao, Zhang Qinchuan. A Review of the Application of Deep Convolutional Neural Networks in Computer Vision [J]. Data Acquisition and Processing, 2016, 31 (1): 17. DOI: 10.16337/j.1004-9037.2016.01.001

[3] Liu Chenyi, Xu Mingwei, Geng Nan, et al. Overview of Intelligent Routing Algorithms Based on Machine Learning [J]. Computer Research and Development, 2020, 57 (4): 17. DOI: 10.7544

[4] Xu Zhijie, Wang Jing, Liu Ying, et al. Current Status and Prospects of Computer Vision Core Technology [J]. Journal of Xi'an University of Posts and Telecommunications, 2012. DOI: CNKI: SUN: XAYD.0.2012-06-003

[5] Yang Lu, Wang Hui, Wei Min. A Review of Machine Learning Based Unreferenced Image Quality Evaluation [J]. Computer Engineering and Applications, 2018, 54 (19): 9. DOI: 10.3778/j.issn.1002-8331.1807-0169

[6] Wei Long, Liu Le, Liu Jiji, et al. Non destructive testing method for solid rocket motors based on machine learning and computer vision: CN202011584416.4 [P]. CN112529899A [2023-10-22]

[7] Xiao Jun, Zhuang Yueting, Wu Fei. Overview of the Application of Computer Vision and Machine Learning Technology in 3D Human Animation [J]. Journal of Computer Aided Design and Graphics, 2008, 20 (3): 10. DOI: CNKI: SUN: JSJF.02008-03-002

[8] Li Xianpeng, Guan Yanling. Research on the Application of Computer Vision and Machine Learning Technology in 3D Human Animation [J]. Tomorrow Fashion, 2018 (2): 1

[9] Alwaseela Abdalla Mohamed Hassan. Phenotypic study of rapeseed based on RGB images and machine learning algorithms [D]. Zhejiang University [2023-10-22]

[10] Li Ce, Shang Xinyu, Liu Yuwei, et al. A pipeline defect recognition method based on computer vision and machine learning: CN201910136101.4 [P]. CN109800824A [2023-10-22]

[11] Sun Yujian, Xu Zili, Xin Cun, et al. A crack identification method based on machine learning algorithms and computer vision: CN202111157916.4 [P]. CN202111157916.4 [2023-10-22]

[12] Fu Yongmin, Fan Lei, Li Changjin, et al. A cascaded detection method for tobacco impurity images based on computer vision and machine learning [J]. Journal of Light Industry, 2023

[13] Zhuang Ying. Potential Applications of Computer Vision and Machine Learning in Museum Collection Research and Exhibition [J]. Southeast Culture, 2018 (S01): 5. DOI: CNKI: SUN: DNWH.0.2018-S1-015

[14] Billie, F. Spencer, Jr., et al. Progress in Research on Inspection and Monitoring of Civil Infrastructure Based on Computer Vision [J]. Engineering, 2019, v.5 (02): 41-90. DOI: CNKI: SUN: GOCH.0.2019-02-006