Application of Deep learning in computer vision

Sun Xin *

School of Henan University of Technology, Henan450001, China

* Corresponding Author Email: 2503889413@qq.com

Abstract. The application of artificial intelligence is deep learning which is one of the current topics in the computer field as well as for the application of computer vision. With the continuous enhancement of deep learning, the algorithm performance is constantly updated. This review paper provides a brief overview of the basic concepts of computer vision and deep learning. Image classification, semantic segmentation and object detection are introduced in this paper followed by a description of their real world applications in various computer vision tasks, such as smart transportation and face recognition. Afterwards, the main applications of deep learning in the research field are demonstrated in this paper.

Keywords: Computer vision; Deep Learning; Image classification; Semantic Segmentation; Object detection.

1. Introduction

Computer vision technology uses geometry, physics, statistics, and learning theory to build models, and then decomposes discriminant information from image data. Computer vision is an interdisciplinary field that enables the system and computer to derive resourceful information from videos, digital images and other forms of visual data [16] it imitates the human eye and is used to train models to perform various functions with the help of related tools, algorithms, and data. This field was revolutionized in the past years due to deep learning techniques. Computer vision usually involves the evaluation of images or videos, simulates the human visual system, captures visual information, to achieve the interpretation of image content with no or little human intervention.

As a branch of machine learning, deep learning simulates the human brain to realize intelligent processing of data, enables machines to learn independently, and makes continuous progress and development. Deep learning technology has been applied to almost every field of computer vision, such as object detection, image classification, and semantic segmentation. And in the image search, automatic driving, medical detection equipment and other products have immeasurable commercial value and broad application prospects.

This paper aimed to introduce the techniques of deep learning in the Computer Vision domain and the recent achievements in application scenarios. The rest of the paper is organized as below. We will first review the techniques and the application of deep learning in computer vision in section 2 and then review two application scenarios in Section 3. Section 4 concludes this paper.

2. Computer vision techniques and applications

2.1. Image classification

In recent years, deep learning performed superior in the field of computer vision to traditional machine learning technology. Indeed, the image classification issue drew great attention as a prominent research topic. The classification system consists of a database containing predefined patterns, and the classification system compares with the detected objects to classify into proper categories. The main process of image classification includes 6 steps: (1) the first is to capture digital data to acquire the data we need; (2) pre-processing of image data, this step is for cleaning data; (3) feature extraction, this is the process of measuring or calculating or detecting the features from image samples; (4) selection of training data, selection of the particular attribute which best describes the pattern, compare the image patterns with the target patterns and classification; (5)

classification output; (6) accuracy assessment, identify possible sources of errors and as an indicator used in comparisons.[1]

Several techniques involving complex algorithms are presented for image classification purposes. Technology is mainly divided into several categories, Artificial Neural Network and Support vector machine are introduced below.

2.1.1 ANN

Artificial Neural Network intends to simulate the behavior of the biological system composed of "neurons", and it is capable of machine learning as well as pattern recognition. ANN is an information processing technique which works like the human brain processes information. A neural network consists of layers and basic units, also known as its "architecture" or "topology," and includes an interchangeable weight adjustment mechanism.

The classic network normally contains three layers: (1) Input layer: numerous neurons receive a large number of non-linear Input messages. The input message is called an input vector; (2) the hidden layer is the layer composed of many neurons and links between the input layer and the output layer. The hidden layer may have one or more layers. The number of nodes in the hidden layer is variable, but the more the number, the more significant the nonlinearity of the neural network, thus the stronger the neural network; (3) Output layer: messages are transmitted, analyzed and weighed in neuronal links to form output results. The output message is called the output vector [2].

Neural networks offer a number of advantages, including performing well with linear and nonlinear data. It has a self-learning function. For example, when realizing image recognition, many different image samples and corresponding recognition results are input into the artificial neural network first, and the network will gradually learn to recognize similar images through a self-learning function. But it also comes with some limitations, the common criticism of neural networks is robotics, it requires a large diversity of data to train. It provides any information about the relative significance of the various parameters. It is very difficult to interpret the overall structure of the network as it may be stuck at local minima and also have over-fitting problems.

2.1.2 SVM

Support Vector Machine (SVM) is a relatively simple Supervised Machine Learning Algorithm used for classification and/or regression. Compared with other algorithms, the outstanding feature of the SVM algorithm is that it can use the classifier to deal with classification problems and SVM regressor to deal with regression problems.

As a linear algorithm at its core can be imagined almost like a Linear or Logistic Regression. For example, an SVM classifier creates a line (plane or hyper-plane, depending upon the dimensionality of the data) in an N-dimensional space to classify data points that belong to two separate classes. It is also noteworthy that the original SVM classifier had this objective and was originally designed to solve binary classification problems. However, linear regression that uses the concept of line of best fit, which is the predictive line that gives the minimum Sum of Squared Error (if using OLS Regression), or Logistic Regression that uses Maximum Likelihood Estimation to find the best fitting sigmoid curve, support Vector Machines uses the concept of Margins to come up with predictions.

In terms of how does Support Vector Machine algorithm work, it can be understood in the following ways:

Step 1: SVM algorithm predicts the classes. One of the classes is identified as 1 while the other is identified as -1.

Step 2: Convert the business problem into a mathematical equation involving unknowns. These unknowns are then found by converting the problem into an optimization problem. In the case of the SVM classifier, a loss function known as the hinge loss function is used and tweaked to find the maximum margin.

Step 3: This loss function can also be called a cost function whose cost is 0 when no class is incorrectly predicted. However, if this is not the case, then error/loss is calculated.

Step 4: As is the case with most optimization problems, weights are optimized by calculating the gradients using advanced mathematical concepts of calculus viz. partial derivatives.

Each machine learning algorithm has its own set of advantages and disadvantages that makes it unique. For the advantage of SVM, the kernel function can be used to map nonlinear samples to high-dimensional space to make them linearly separable, and then the optimal segmentation hyperplane can be found by maximizing the classification interval between data sets, which shows many unique advantages in classification problems. For example, for image classification. Experimental results show that after three to four rounds of feedback, the support vector machine can obtain significantly higher search accuracy than the traditional query optimization scheme. For the disadvantage, all the samples are calculated to see which the support vectors are. The computational complexity is high and the computational speed is slow. Also support vector machines are only suitable for tasks with small batch samples, but not for tasks with millions or even hundreds of millions of samples [3].

2.2. Semantic Segmentation

Semantic Segmentation is a computer vision task that involves grouping together similar parts of the image that belong to the same class. Generally, it can be divided into three steps:

(1). Classifying: classifying a certain object in the image.

(2). Localizing: finding the object and drawing a bounding box around it.

(3). Segmentation: grouping the pixels in a localized image by creating a segmentation mask [6].

Autonomous driving is a common application of Semantic Segmentation. The autonomous driving application is based on the Deep Labv3+ network structure, which uses the encoder-decoder and ASPP combined structure, and the ResNet50 with extended convolution is constructed by the base network. In the coding part, Deep Labv3 is used as the encoder, and the ResNet50 structure with extended convolution is applied to gradually reduce the resolution of feature images and capture high-level semantic information. Then through the ASPP module, the information of context is mined by pooling features of different resolutions. In the decoding process, a simple but effective module is used to recover the target boundary details: a 3×3 convolution is applied to refine the features, followed by a simple 4-fold bilinear up-sampling. In semantic segmentation of traffic scenes, the core goal of attention mechanism is to select the information that is more critical to the current segmentation task target from numerous information, so the attention mechanism module is added. The dual-attention mechanism module used in this application is a lightweight attention module, which uses the attention module in both space and channel, and can learn key target features and regions of interest in channel and space dimensions [16].

This is based on the Deep Labv3 + take attention modules optimized traffic scene image segmentation model, without any increase in number and not change each stage convolution layer under the premise of receptive field retain more high-level semantic information, through the study of channel information and location information, the restore boundary information can obtain more abundant features of high accuracy convolution neural network.

2.3. Object detection

Object detection is a profound computer vision technique that focuses on identifying and labeling objects within images, videos, and even live footage. Object detection models are trained with a surplus of annotated visuals in order to carry out this process with new data. These object detection models are trained with hundreds of thousands of visual content to optimize the detection accuracy on an automatic basis later on.

The first largely successful family of methods was R-CNN (Region-Based Convolutional Neural Network), which was proposed in 2014. It surpassed its predecessors by extracting merely 2,000 regions from the image, which were referred to as region proposals, instead of an exceedingly large number of regions prior to this. The flowchart of R-CNN is the following: the input image is selected, of which 2,000 region proposals are extracted.

Fast R-CNN significantly cut down on process times by inputting the image to the pre-trained CNN in order to generate a convolutional feature map, eliminating the process of breaking the image down into 2,000 region proposals. Training hours for the CNN drop from around 84 hours to around 9 hours with Fast R-CNN. Additionally, the test time drops from around 50 seconds to around 2.5 seconds. A third and even more upgraded model was later introduced that would be known as Faster RCNN. It brought down the around 2.5-second test speed from Fast R-CNN to an unparalleled around 0.2 seconds, making it the fastest of its predecessors and optimal for real-time object detection.

3. Applications of deep learning and computer vision

Techniques and the application of deep learning in computer vision have been introduced above. This part mainly shows the main applications of deep learning of computer vision in the real world.

Computer vision focuses on image and video understanding. It involves tasks such as object detection, image classification, and segmentation. In medicine, computer vision focuses on research in medical imaging, medical video, and real clinical deployment. The clinical tasks are suitable for CV span many categories, such as screening, diagnosis, detecting conditions, predicting future outcomes, segmenting pathologies from organs to cells, monitoring disease, and clinical research. First of all, in applications of deep learning and computer vision can be used in clinical diagnosis.

Remarkably, deep-learning computer vision models have achieved physician-level accuracy at clinical diagnostic tasks. The key clinical tasks for DL in dermatology include lesion-specific differential diagnostics, finding concerning lesions amongst many benign lesions, and helping track lesion growth over time. These studies were largely restricted to the binary classification task of discerning benign vs malignant cutaneous lesions, classifying either melanomas from nevi or carcinomas from seborrheic keratoses. Research has also identified the numerous advantages of using computer vision and deep learning applications to diagnose breast cancer. Trained with a vast database of images consisting of both healthy and cancerous tissue, it can help automate the identification process and reduce the chances of human error. Incorporating these algorithms into clinical workflows would allow their utility to support other key tasks, including large-scale detection of malignancies on patients with many lesions, and tracking lesions across images in order to capture temporal features, such as growth and color changes. This area remains fairly unexplored, with initial works that jointly train CNNs to detect and track lesions. With the rapid improvements in technology, healthcare computer vision systems may be used for diagnosing other types of cancer, including bone and lung cancer, in the near future.

There are also many applications in smart transportation [22].

Intelligent transportation system can recognize the physical and physiological characteristics of running objects through a variety of intelligent equipment, can quickly locate the target, fully show the advantages of computer vision technology in intelligent transportation system computer vision technology can be divided into four applications.[4] First, real-time traffic monitoring, the technology used in the real traffic is the use of the existing in the operation of the high cameras to capture target vehicle, and through the analysis of the data analysis and image processing technology background from image sequence, motion parameters and information related to gain data, and then to transfer the data to the database for intelligent recognition, using intelligent transportation monitoring system.[24] The second is the route navigation service, the use of computer vision technology in automobile navigation has an important application space, significant performance in the full use of road extraction technology and vehicle detection technology, so as to provide routes to create full convenience, Finally, the safe driving of the vehicle is effectively realized. The third is intelligent charging, which is reflected in the capture of the image of the vehicle in high-speed driving, the selection and extraction of vehicle parameters, and the precise geographical location of the positioning image taken. Finally, intelligent characters are fully recognized and segmented. Finally, vehicle assisted driving mainly refers to the surrounding environment of a vehicle driving through computer vision technology, including the road conditions before and after road signs Passers-by,

etc., to identify the data and analysis, and then take advantage of screening and filtering of information systems, finally provide forecasts and guide the user's driving, which can better ensure the safety of users and the entire traffic safety, to avoid the danger of traffic accidents and the occurrence of unnecessary traffic violations.[26]

4. Conclusion

In general, this paper provides a comprehensive analysis of computer vision and deep learning. For the introduction of image classification, an introduction to two techniques, ANN and SVM, is provided. Semantic segmentation and image classification are also introduced in this paper, and their applications in smart transportation and clinical diagnosis are mentioned after that. Now computer vision model is the main core of the information technology, and the information is based on the international automation technology and services, Therefore, the application of computer visualization graphics patterns to computer technology and automation technology has important significance and far-reaching implications, so it is now necessary to develop computer image technology to promote future research.) At the same time, we should strengthen the theoretical research on the deep learning model. The deep learning model can reflect the features of the image perfectly after training with a large amount of data. However, there is still a lot of unfinished work on the theoretical research of the deep learning model, and the theoretical basis of the model needs to be further strengthened. How to find the minimum value? What information is lost when the image information enters the next layer of the model? Compared with static image recognition, deep learning is not applied in video image recognition. On the one hand, due to the lack of a large number of video image data sets, On the other hand, the training of video images requires a deeper model. The model structure is more complex, and the calculation is larger, so the training time will be longer. How to improve these problems is one of the key issues in the research of video image recognition.

References

[1] Kamavisdar, P., Saluja, S., & Agrawal, S. (2013). A survey on image classification approaches and techniques. International Journal of Advanced Research in Computer and Communication Engineering, 2(1), 1005-1009.

[2] Sheetal Sharma. (August 8, 2017). Artificial Neural Network (ANN) in Machine Learning.

[3] SUMEET BANSAL. (n.d.). Introduction to SVM – Support Vector Machine Algorithm of Machine Learning.

[4] WU Nizhen (2021).Computer Vision Technology Research and Development Trend Analysis. Science and Technology Innovation and Application (34), 58-61.

[5] WANG Longfei & Yan Chunman (2021).Review on Semantic Segmentation of Road Scenes. Laser & Optoelectronics Progress (12), 44-66.

[6] Nilesh, Barla. (June 20, 2022). The Beginner’s Guide to Semantic Segmentation.

[7] Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. A. (2017, February). Inception-v4, inception-resnet and the impact of residual connections on learning. In Thirty-first AAAI conference on artificial intelligence.

[8] Simon, P., & Uma, V. (2020). Deep learning based feature extraction for texture classification. Procedia Computer Science, 171, 1680-1687.

[9] Esteva, A., Chou, K., Yeung, S., Naik, N., Madani, A., Mottaghi, A. ... & Socher, R. (2021). Deep learning-enabled medical computer vision. NPJ digital medicine, 4(1), 1-9.

[10] ZHENG Yuanpan, LI Guangyang & LI Ye. (2019). Survey of Application of Deep Learning in Image Recognition. Computer Engineering and Applications, 55(12), 20-36.

[11] WU Jialin (2021). Research on Image Classification Mwthod of Skin Diseases Based on Enhanced Deep Learning (Master’s thesis, XiJing University).

[12] LIU Bo (2019). A Review of Computer Vision Research. The World of Digital Communication, 000(012), 97.

[13] WANG Shu. (2016). The Research on Deep Learning algorithm and its application in Image Classification. (Doctoral dissertation, Nanjing University of Posts and Telecommunications).

[14] LU Hongtao & Luo Mukun (2022). Survey on New Progress of Deep Learning Based Computer Vision. Data acquisition and processing (02), 247-278.

[15] LIANG Tianfen, ZHANG Nanfeng, ZHANG Yanxi, YUAN Jinhao & GAO Xiangdong (2021). Summary of Research Progress on Application of Prohibited Item Detection in X-Ray Images. Computer Engineering and Applications (16), 74-82.

[16] HE Miaoying & CUI Yuchao (2021). Semantic segmentation of traffic scenes for autonomous driving. Computer application (S1), 25-30.

[17] LIU Zhe. On Computer Vision Technology [J]. Game Day Magazine, 2019, 25(8):159.

[18] CHEN Chuan, CHEN Zhe & DING Shuanghui (2020). Innovation of Computer Vision Teaching Contents under Development of Deep Learning. Computers and Modernization (06), 107-113.

[19] LI Guohe, QIAO Yinghan, WU Weijiang, ZHENG Yifeng, HONG Yunfeng & ZHOU Xiaoming.(2019).Review of deep learning and its application in computer vision. Application Research of Computers (12), 3521-3529+3564.

[20] Viola, P., & Jones, M. (2001, December). Rapid object detection using a boosted cascade of simple features. In Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001 (Vol. 1, pp. I-I). Ieee.

[21] WANG Longfei & Yan Chunman (2021).Review on Semantic Segmentation of Road Scenes. Laser & Optoelectronics Progress (12), 44-66.

[22] Gaudenz Boesch. (n.d.). Top 10 Applications Of Deep Learning and Computer Vision In Healthcare.

[23] LIU Wenting & LU Xinming (2022).Research Progress of Transformer Based on Computer Vision. Computer Engineering and Applications (06), 1-16.

[24] HE Chongchong (2022).Analyze the Application of Computer Vision Technology in Intelligent Transportation Systems. China Plant Engineering (05), 38-39.

[25] Liu, Y., & Zheng, Y. F. (2005, July). One-against-all multi-class SVM classification using reliability measures. In Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005. (Vol. 2, pp. 849-854). IEEE.

[26] DUAN Jian & GuoJian-min (2018).Application of Computer Vision Technology in Intelligent Transportation System. Journal of Jiamusi Vocational Institute (09), 396-397.