DOI: 10.53469/jtpes.2024.04(01).06

Integrating Advanced Computer Vision and AI Algorithms for Autonomous Driving Systems

Kai Tan[1,*], Jiang Wu[2], Hong Zhou[3], Yixu Wang[4], Jianfeng Chen[5]

[1]Electrical & Computer Engineering ,University of Washington, San jose, CA, USA
[2]Computer science ,University of Southern California, Los Angeles, CA, USA
[3]Computer Technology, Peking University, Beijing, CN
[4]Computer Technology, Independent Researcher, Beijing, CN
[5]Statistics ,Independent Researcher, Fairfax,Va,USA
Correspondence Author, tank5@uw.edu

Abstract: ***Autonomous vehicle is a typical high-tech comprehensive application, including scene perception, optimization calculation, multi-level assisted driving and other functions, using computer vision, sensors, information fusion, information communication, high-performance computing, artificial intelligence and automatic control and other technologies. In these technologies, computer vision, as a direct entry point to data processing, is an integral part of autonomous driving. Secondly, it brings revolutionary changes to the future transportation system. The application of image processing and computer vision in autonomous driving plays a key role in enabling vehicles to perceive and understand the surrounding environment and achieve intelligent decision-making and control. Therefore, in combination with the application of computer vision and artificial intelligence in automatic driving, this paper expounds the image processing technology in automatic driving, including camera and sensor technology, image acquisition and preprocessing, feature extraction and object detection, so as to discuss the application of computer vision algorithm in automatic driving. The research on lane keeping and recognition, obstacle detection and avoidance, traffic signal and sign recognition is of great practical significance.***

Keywords: Computer vision technology; Image processing; Autonomous driving; Road detection.

1. INTRODUCTION

Vision-based object detection plays an indispensable role in environmental perception system. As one of the research hotspots of image processing and computer vision direction, it can help autonomous driving system detect target objects such as vehicles, pedestrians and traffic signs, which is one of the important technologies to realize autonomous navigation and improve traffic safety. Therefore, it is of great significance to optimize and improve the research and application of vision-based object detection technology. In the current field of autonomous driving, image-based object detection technology is one of the core and most challenging problems. Effectively solving this problem will promote the development of the entire autonomous driving technology and further improve the safety and intelligence of autonomous vehicles on the road[1]. At the same time, thanks to the higher accuracy and reliability of autonomous vehicles in following traffic rules and reducing accidents, it is expected that millions of lives will be saved and billions of yuan will be saved as a result, and the realization of such a dream represents a major advance in the transportation field. With the continuous innovation and efforts of researchers and industry, it may soon become a reality.

As the direct entry of data processing, computer vision is an indispensable part of automatic driving. In addition, most transportation authorities around the world advocate the concept of "defensive driving." Defensive driving is a mechanism to predict and help avoid crises, requiring drivers to prevent other traffic accidents caused by their own negligence or violations in addition to obeying traffic rules. Assisted autonomous driving is based on a number of highly innovative integrated applications, and its key modules can be summarized as environment perception, behavior decision making, path planning and motion control[2]. The primary problem facing assisted automatic driving is how to effectively collect and quickly process the surrounding environmental data and the internal data of the vehicle, which is also the basic data support of automatic driving. Therefore, the research of this paper is of great significance.

2. RELATED WORK

Vision-based object detection is an indispensable part of environment perception system and an important direction of computer vision and image processing research. This technology is widely used in many fields such as object detection, intelligent transportation, robot path planning and video surveillance in autonomous driving systems [5]. In recent years, the development of big data and the improvement of computing performance have made deep neural networks achieve unprecedented success[3], surpassing traditional methods in various fields and ranking in the devil's position, especially in image classification, object detection and semantic segmentation. In recent years, the development of deep learning technology has promoted the global Internet giants to take autonomous driving as an important strategic direction for future development, which has pushed the autonomous driving industry into a new historical period. Waymo and Uber and other foreign companies began to work on self-driving technology as early as 2017. Waymo tested self-driving cars in the Phoenix area, and announced in November 2022 that it would open self-driving ride-hailing services to the public in downtown Phoenix.

2.1 Autonomous Driving Camera and Sensor Technology

Image sensors generally adopt certain modes to collect image data, BGR mode and CFA mode are commonly used. BGR mode is a kind of image data mode that can be directly displayed and compressed. It is determined by R(red), G(green), B(blue) primary color values to determine 1 pixel[4]. For example, the SUPER CCD image sensor used by Fusu digital camera adopts this mode. Its advantage is that the image data generated by the image sensor can be directly displayed and other subsequent processing without interpolation, and the image effect is the best, but the cost is high, and it is often used in professional cameras. General digital camera sensor (CCD or CMOS) accounts for about 10% to 25% of the total cost of the whole machine, in order to reduce the cost, reduce the size, most of the digital cameras on the market use CFA mode[5], that is, the surface of the pixel Array covered with a layer of Color Filter Array (CFA), There are a variety of color wave arrays, and the most widely used is the Baver format filter array, which meets the GRBG law, and the number of green pixels is twice the number of red or blue pixels, because the peak sensitivity of the human eye to visible light spectrum is located in the middle band, which exactly corresponds to the green spectral component.

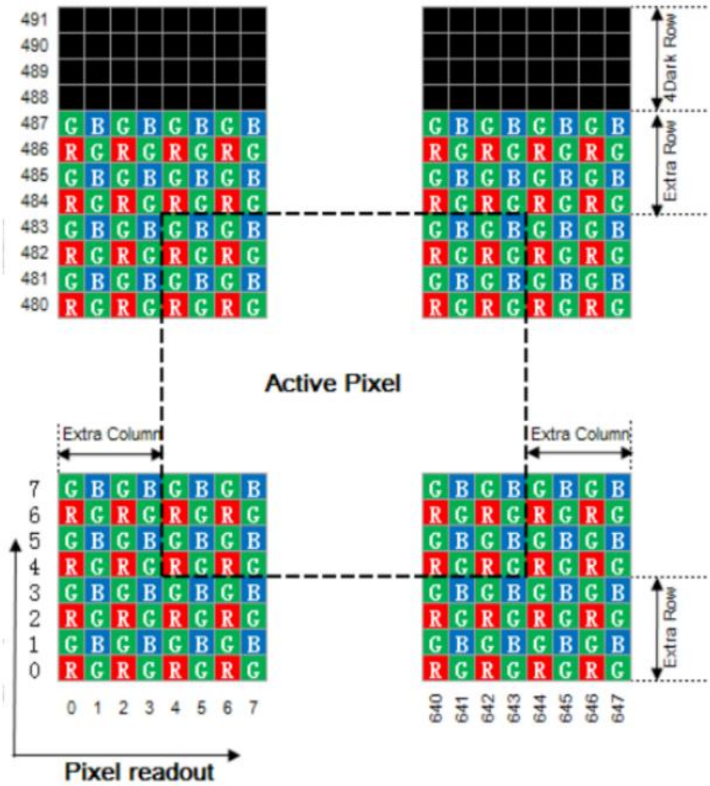

Figure 1: CAF mode image sensor

The above image is an image sensor using CFA mode, with an effective resolution of 640 480. The image data in this mode is represented by only one of the 3 values G and B[6]. In this way, each pixel can only capture one of the three primary colors R, G, and B, while missing the other two color values, and the result is a Mosaic image. In order to get a full-color image, it is necessary to use the color information of its surrounding pixels to estimate the missing other two colors, this process is called color interpolation, also known as color interpolation or mosaics.

2.2 Image Acquisition and Preprocessing

Image sensors generally use specific modes to collect image data, with BGR mode and CFA mode being common choices.To achieve this, modern autonomous vehicles are often equipped with multiple sensors, including cameras, liDAR, millimeter-wave radar, and ultrasonic sensors. These sensors are distributed at different locations in the vehicle to obtain a full range of views and diverse data[7]. Cameras are one of the most commonly used sensors. They can capture high-resolution images that provide rich visual information for tasks such as road sign recognition, vehicle detection, pedestrian detection, and lane keeping. However, the image data captured by the camera usually needs to be corrected and processed to correct lens distortion, adjust brightness and contrast, and remove noise from the image. Therefore, it is necessary to use image acquisition and processing technology.

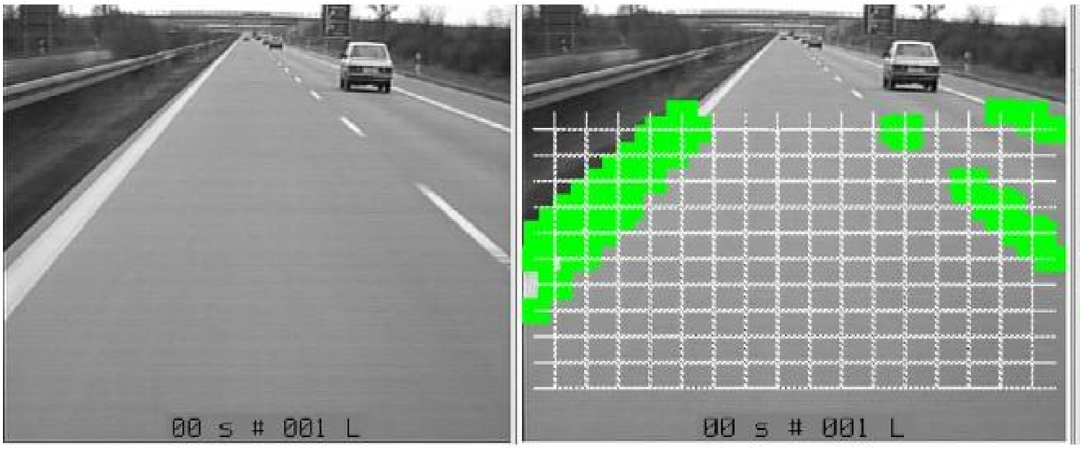

Figure 2: Image segmentation to extract lane lines

In the real problem, the image processing can extract the lane line by image segmentation, and judge whether the vehicle position deviates. Detect and recognize signal lights, speed limit traffic signs, adjust speed; Detect and identify pedestrians, electric vehicles, etc., and maintain a safe distance; Mature automatic driving often needs multiple sensors to work together, [8]Lidar plus vision and the like, LiDAR for large object recognition effect is very good, but the recognition effect of small volume is not good, such as debris rolling on the road, not removed roadblocks, this situation still needs to be visual recognition, timely adjustment of the driving state.

2.3 Image Feature Extraction and Object Detection

In feature extraction and object detection, (1) useful features need to be extracted from images captured by vehicle cameras. These features, which may include edges, corners, textures, etc., are unique in the image and help distinguish between different objects. Typically, computer vision algorithms preprocess images to enhance the recognizability of these features, such as removing noise and adjusting contrast and brightness. (2) Once the feature is extracted, the next step is object detection. This means identifying and locating different types of objects in an image. To achieve this goal, commonly used methods include classical object detection techniques such as Haar cascade classifiers[9], HOG features, and convolutional neural net endings (CNNS), which allow systems to be trained to recognize specific types of objects such as vehicles, pedestrians, or traffic signs. The two-dimensional convolution operation is as follows:

$$\mathrm{Conv}(I,K)(x,y)=\sum_{u=1}^{m}\sum_{v=1}^{n} I(x-u,y-v)\cdot K(u,v) \tag{1}$$

It can be used to illustrate how the CNN convolves with the input image (I) through the convolution kernel to extract features. Where the Input Image is represented by I(x,y), where x and y represent the horizontal and vertical pixel coordinate filters or Convolution Kernel of the image, respectively: Expressed as K(u,v), where u and v represent the horizontal and vertical dimensions of the convolution kernel

Convolution Operation: denoted by (IXK) (x,y), convolution operation between input image and filter K.

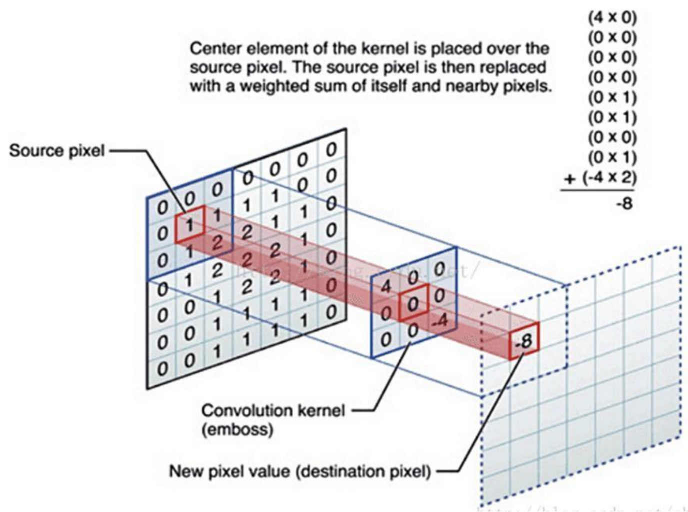

Figure 3: CNN operates on the principle of convolution with input images through the convolution kernel

The results of feature extraction and object detection are used for decision making and path planning in autonomous driving systems[10]. By recognizing other vehicles and pedestrians, the self-driving system can take appropriate actions, such as slowing down, changing lanes, or stopping to ensure driving safety. At the same time, this information can also be used to plan the path of the vehicle to avoid obstacles or comply with traffic rules.

3. METHODOLOGY

The automatic driving system of the car is a very important part of the driving process of the car. In recent years, with the advancement and research of automotive technology, autonomous driving systems have also faced new challenges. From target detection, feature extraction, path planning, sensor fusion, vehicle control and other aspects, the automatic driving system is being transformed[11]. In this paper, the automatic driving system combining object detection, image processing, path planning and control is implemented on the basis of computer vision technology.

3.1 Object Detection

Target detection is the process by which a computer vision system identifies and locates various objects in an image, often used in road target recognition within the automatic driving system. Common object detection techniques include:Bounding box detection: Locate the target by drawing a rectangular box in the image, such as YOLO, SSD, etc.Segmentation detection: The image is divided into multiple regions to represent the spatial relationship of different targets, such as FCN, Mask R-CNN, etc.
Mathematical model formula:

$$P(C|x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(c_i-\mu)^2}{2\sigma^2}} \tag{2}$$

Where P(C｜x) represents the probability that the target class is C, x represents the image pixel value, μ represents the mean of the target class, and σ represents the variance of the target class.

Secondly, target tracking is a process in which the computer vision system tracks and tracks the identified target to provide real-time position and speed information. Common target tracking technologies include: feature-based tracking: tracking is carried out by extracting the features of the target, such as color, shape, edge, etc.

State-based tracking[12]: By establishing the state model of the target, such as Kalman filter, tracking.

Mathematical model formula:

$$\hat{x}_{k|k} = K_k(y_k - \hat{y}_{k|k-1}) \tag{3}$$

Where x k|k represents the estimate of the target, y - k represents the actual value, k - k - 1 represents the estimate at the previous time, and K - k represents the Kalman gain.

3.2 Image Segmentation

Segmentation is the task of classifying pixels in an image.Segmentation is initially divided into semantic segmentation and instance segmentation. Semantic segmentation is to give a category for each pixel in the image, such as ground, tree, car, person, etc. Instance segmentation is similar to object detection, but instance segmentation is to give all pixels of each target, and different targets of the same category should give different ids, that is, each target can be clearly distinguished. This year, there was research to unify semantic segmentation and instance segmentation together, called panoramic segmentation, as shown in the following figure:

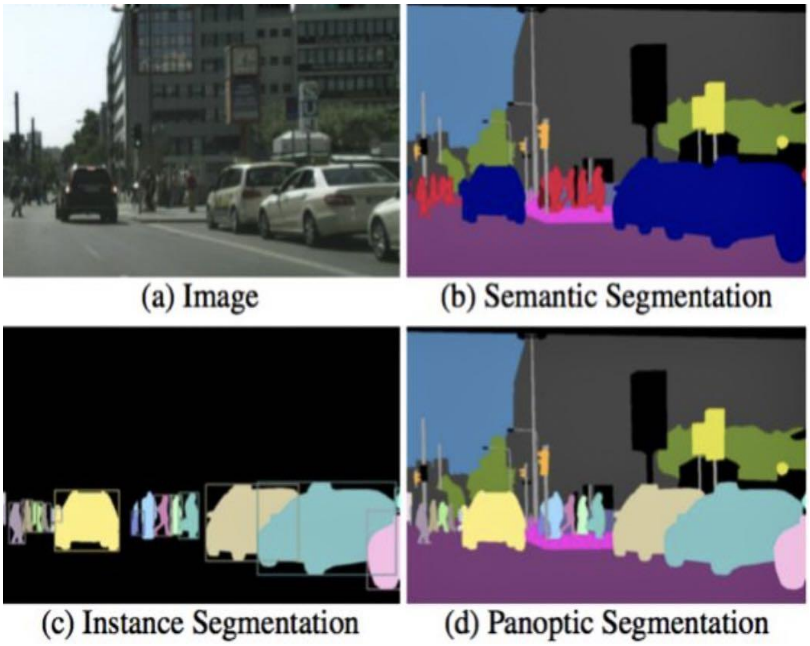

Figure 4: Driving section image segmentation program

Semantic segmentation is widely used in unmanned driving. For example, pavement segmentation, crosswalk segmentation and so on[13]. One of the earlier and classic models of semantic segmentation is FCN.

FCN has several classic improvements. First, the full convolution layer is replaced by the full convolution layer; second, the small-resolution Feature Map after convolution is sampled from the upper layer to obtain the result of the original resolution size; finally, FCN uses the cross-layer connection mode. Cross-layer connection can combine the high-level semantic features with the low-level location features, which makes the segmentation result more accurate. The FCN structure diagram is as follows:

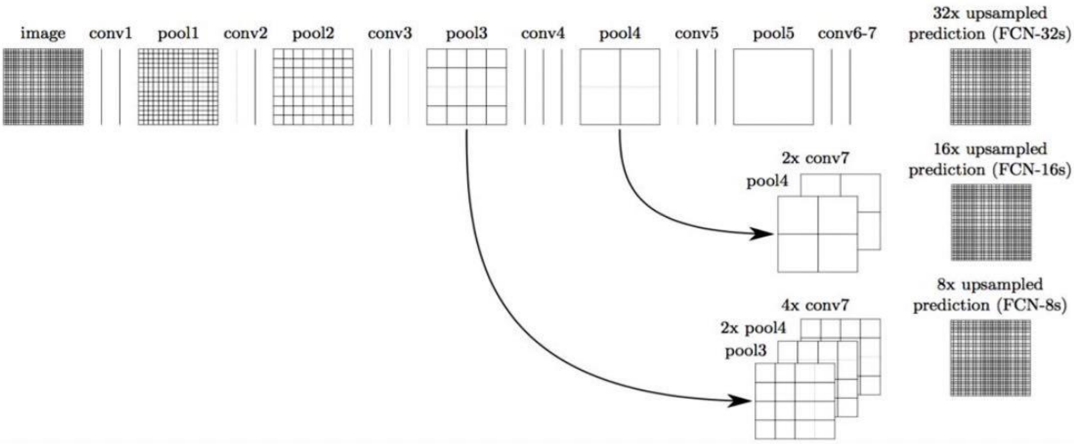

Figure 4: FCN driverless road section segmentation principle

3.3 Section Distance Estimation

There are many ways to calculate distance information:

1) Laser ranging in automatic driving, the principle is to calculate the distance according to the time of laser reflection. The distance calculated in this way is the most accurate, but the output frequency calculated depends on the frequency of the laser itself, and the general laser is 10Hz.

2) Monocular depth estimation, the principle is that the input is the monocular camera picture, and then the CNN model of depth estimation is used to predict, and the depth of each pixel is output[14]. The advantage of this method is that the frequency can be higher, and the disadvantage is that the depth error is relatively large.

3) Structured light ranging, the principle is that the camera emits a unique structure of structured light, according to the polarization of the returned light and other characteristics, calculate the distance of each pixel. The main disadvantage of this method is that the structural light is greatly affected by natural light, so it is difficult to use outdoors.

4) binocular ranging, the principle is based on the small difference seen by the two lenses, according to the distance between the two lenses, calculate the distance of the object. The disadvantage of this method is that the error of calculating the distance of distant objects is large.

According to the internal calculation of the camera, the principle is similar to pinhole imaging. Each point in the picture can be converted into a line in space according to the camera parameters, so for a plane with a fixed height, the intersection point can be found to calculate the distance[15]. Usually the fixed plane uses the ground, that is, we can know the exact distance of each point on the ground in the picture. This calculation method is extremely accurate in the case of accurate internal parameters in the camera, but only for a fixed height plane.

This step includes the scene understanding part of image processing for automatic driving. The so-called scene understanding is the process in which the computer vision system understands and interprets various objects and relationships in the image to provide high-level information and decision support. Common scene understanding techniques include:[16] (1) Relationship extraction: by analyzing the relationship between objects, such as the distance between vehicles, speed and so on. (2) Event identification: by analyzing the dynamic changes of the target, such as vehicle collision, pedestrian crossing, etc.

Mathematical model formula:

$$P(S|E) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(s_i-\mu)^2}{2\sigma^2}} \tag{4}$$

Where P(S｜E) represents the probability of scenario S giving event E, Si represents the characteristics of the scenario,μ represents the mean of the events, and σ represents the variance of the events.

4. CONCLUSION

This study comprehensively discusses the application and importance of computer vision in automatic driving technology, and emphasizes that computer vision is an indispensable part of automatic driving, which enables the vehicle to perceive and understand the surrounding environment, and make intelligent decisions and controls. Especially in object detection, image segmentation and feature extraction, computer vision technology is very important for improving traffic safety and realizing automatic navigation. With the development of deep learning and big data technologies, the accuracy and reliability of these systems continue to improve, heralding major advances in the future of autonomous driving. Building on the principles of computer vision detection and segmentation of the direction of computer vision, this paper also introduces the principle of correlative algorithm of automatic driving and its practical application in automatic driving[17]. After that, some algorithms of distance estimation and some visual solutions in the industry are introduced. At present, the mainstream unmanned vehicle obstacle perception in the industry relies on laser, and the visual scheme is relatively not very mature[18-20]. However, we still prefer visual solutions because of their low cost, which can reduce the dependence on highly refined maps.

In conclusion, the future development directions of autonomous driving technology primarily encompass the following aspects:

1) Algorithm optimization: With the continuous development of deep learning, computer vision, machine learning and other technologies, the algorithm of automatic driving system will be continuously optimized to improve its accuracy, real-time and reliability.

2) sensor technology: Future autonomous driving systems will rely more on multi-modal sensor technology, such as LiDAR, LiDAR, ultrasonic, etc., to provide more accurate environmental understanding and decision support.

3) safety and reliability: the safety and reliability of autonomous driving systems will become a key issue for future development, requiring more rigorous testing and verification.

4) Laws and regulations: With the development of autonomous driving technology, relevant laws and regulations will also be continuously improved to regulate the use and management of autonomous driving systems.

5) social Acceptance: The popularization of autonomous driving technology will require social acceptance and support, and need to address issues such as safety, road traffic, occupational structure and so on.

ACKNOWLEDGEMENT

This paper is largely inspired and guided by the research results of Xu, J., Pan, L., Zeng, Q., Sun, W., & Wan, W. (2023). Their paper, "Based on TPUGRAPHS Predicting Model Runtimes Using Graph," published in the journal Frontiers in Computing and Intelligent Systems The artificial intelligence related algorithms proposed in "Neural Networks" (Vol. 6, No. 1, pp. 66-69) and the application space of deep learning in the field of autonomous driving and other core technologies provide an important perspective and methodology basis for my research. I hereby acknowledge Xu, J., Pan, L., Zeng, Q., Sun, W., and Wan, W. Express my deep gratitude and respect.

REFERENCES

[1] Pan, Linying, et al. “Research Progress of Diabetic Disease Prediction Model in Deep Learning”. Journal of Theory and Practice of Engineering Science, vol. 3, no. 12, Dec. 2023, pp. 15-21, doi:10.53469/jtpes.2023.03(12).03.

[2] Apollo. Promote open capability-share resource-accelearte innovation-sustain mutual benefit EB/0Lhttp://apollo. author/,2018-01-17.

[3] Kim S, Hwang Y. A survey on deep learning based methods and datasets for monocular 3D object detection(JElectronics,2021,10(4): 517.

[4] Xinyu Zhao, et al. “Effective Combination of 3D-DenseNet’s Artificial Intelligence Technology and Gallbladder Cancer Diagnosis Model”. Frontiers in Computing and Intelligent Systems, vol. 6, no. 3, Jan. 2024, pp. 81-84, https://doi.org/10.54097/iMKyFavE.

[5] Shulin Li, et al. “Application Analysis of AI Technology Combined With Spiral CT Scanning in Early Lung Cancer Screening”. Frontiers in Computing and Intelligent Systems, vol. 6, no. 3, Jan. 2024, pp. 52-55, https://doi.org/10.54097/LAwfJzEA.

[6] Liu, Bo & Zhao, Xinyu & Hu, Hao & Lin, Qunwei & Huang, Jiaxin. . Detection of Esophageal Cancer Lesions Based on CBAM Faster R-CNN. Journal of Theory and Practice of Engineering Science. 2023,3: 36-42. 10.53469/jtpes.2023.03(12).06.

[7] 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.

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

[9] Zhou, H., Lou, Y., Xiong, J., Wang, Y., & Liu, Y. . Improvement of Deep Learning Model for Gastrointestinal Tract Segmentation Surgery. Frontiers in Computing and Intelligent Systems, 2023, 6(1): 103-106.6

[10] Implementation of an AI-based MRD Evaluation and Prediction Model for Multiple Myeloma. . Frontiers in Computing and Intelligent Systems, 2024,6(3): 127-131. https://doi.org/10.54097/zJ4MnbWW.

[11] Zhang, Q., Cai, G., Cai, M., Qian, J., & Song, T. . Deep Learning Model Aids Breast Cancer Detection. Frontiers in Computing and Intelligent Systems, 2023,6(1): 99-102.3
[12] Xu, J., Pan, L., Zeng, Q., Sun, W., & Wan, W. . Based on TPUGRAPHS Predicting Model Runtimes Using Graph Neural Networks. Frontiers in Computing and Intelligent Systems, 2023,6(1): 66-69.7
[13] Wan,Weixiang, et al. "Development and Evaluation of Intelligent Medical Decision Support Systems." Academic Journal of Science and Technology 2023,8(2) : 22-25.
[14] Tian, M., Shen, Z., Wu, X., Wei, K., & Liu, Y. . The Application of Artificial Intelligence in Medical Diagnostics: A New Frontier. Academic Journal of Science and Technology,2023, 8(2): 57-61.7
[15] Shen, Z., Wei, K., Zang, H., Li, L., & Wang, G. . The Application of Artificial Intelligence to The Bayesian Model Algorithm for Combining Genome Data. Academic Journal of Science and Technology, 2023,8(3): 132-135.2
[16] Zheng He, et al. “The Importance of AI Algorithm Combined With Tunable LCST Smart Polymers in Biomedical Applications”. Frontiers in Computing and Intelligent Systems, vol. 6, no. 3, Jan. 2024, pp. 92-95, https://doi.org/10.54097/d30EoLHw.
[17] Prediction of Atmospheric Carbon Dioxide Radiative Transfer Model based on Machine Learning. . Frontiers in Computing and Intelligent Systems,2024, 6(3): 132-136. https://doi.org/10.54097/ObMPjw5n
[18] Liu, Y., Duan, S., Shen, Z., He, Z., & Li, L. . Grasp and Inspection of Mechanical Parts based on Visual Image Recognition Technology. Journal of Theory and Practice of Engineering Science, 2023,3(12): 22-28.1
[19] Simonelli A, Bulo S R, Porzi L, et al. Disentangling monocular 3d object detection/CI/Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 1991-1999.
[20] Xinyu Zhao, et al. “Effective Combination of 3D-DenseNet’s Artificial Intelligence Technology and Gallbladder Cancer Diagnosis Model”. Frontiers in Computing and Intelligent Systems, vol. 6, no. 3, Jan. 2024, pp. 81-84, https://doi.org/10.54097/iMKyFavE.
[21] Liu, B. . Based on intelligent advertising recommendation and abnormal advertising monitoring system in the field of machine learning. International Journal of Computer Science and Information Technology, 2023,1(1): 17-23.