WEP Warwick Evans Publishing

Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 5
AIDML 2024

Deep Convolutional Neural Network enabled unmanned agricultural machine visual navigation system: architecture design, model optimization and empirical evaluation

Xin Gao, Shaojia Yuan*, Jiazheng Zhu, Yifei Zhao

Xi'an Jiaotong University City College, Xi'an, China

* Corresponding Author Email: yuanshaojia66@163.com

Abstract. In response to the pressing need for intelligent navigation technology in modern agriculture, this paper introduces a visual navigation system for unmanned agricultural machinery based on deep convolutional neural networks (CNNs). The system integrates the strong representational power of deep learning with rich visual information to facilitate high-precision, adaptive autonomous navigation in complex agricultural environments. Key innovations encompass: (1) the development of a customized deep CNN model proficient in extracting critical features from agricultural images, such as obstacles, landmarks, and crop rows; (2) the integration of the CNN model with visual SLAM (Simultaneous Localization and Mapping) technology for real-time localization and three-dimensional mapping of agricultural landscapes; (3) the creation of a decision-making system that merges deep learning predictions with conventional path planning algorithms to ensure machinery navigates around obstacles while adhering to optimal trajectories. Experimental validation across diverse agricultural scenarios demonstrates that the proposed visual navigation system sustains high localization accuracy (RMSE < 0.2 meters) and robust obstacle avoidance performance (success rate exceeding 95%) even in GPS-denied or weak GPS environments. Furthermore, the system significantly improves the continuity and efficiency of machinery operations by reducing unnecessary repositioning and redundant tasks. Practically, it is readily integrable into existing agricultural machinery platforms, offering broad applicability and potential for widespread adoption.

Keywords: Convolutional Neural Network (CNN); Unmanned Agricultural Machinery; Visual Navigation; Deep Learning; Precision Agriculture; Image Processing; SLAM (Simultaneous Localization and Mapping); Path Planning.

1. Introduction

1.1. Research background

Driverless car technology is rapidly developing to become a key part of intelligent transportation systems, and its autonomous navigation decision-making function is crucial to vehicle performance. As the technology advances, driverless vehicles are beginning to enter daily life, being used in scenarios such as urban travel, logistics and emergency rescue.

1.2. The potential of deep reinforcement learning

Deep reinforcement Learning (DRL) combines the strengths of deep learning and reinforcement learning to make breakthroughs in the field of intelligent decision making. DRL has been successfully applied to a number of complex scenarios, showing its potential to deal with high-dimensional, non-linear problems and providing new ideas for path planning of unmanned vehicles.

1.3. Research motivation

Traditional path planning algorithms have limitations in dynamic urban environments, such as difficulty in adapting to real-time changes and dealing with sensor uncertainties. By learning from sensor data, DRL provides an adaptable route planning method that can be optimized online, which is expected to improve the navigation efficiency and safety of unmanned vehicles.

2. Relevant theoretical basis and technology

2.1. Basic principles of visual navigation

Fundamentals of image processing

- Image acquisition: Describes the camera imaging principle, involving the conversion of pixel values to real-world physical quantities, the formula is as follows:

$$[I(x,y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(x',y')\cdot h(x-x',y-y')dx'dy'] \tag{1}$$

- Grayscale: the method of converting a color image to a grayscale image, the average method formula is:

$$[G=\frac{R+G+B}{3}] \tag{2}$$

- Histogram equalization: the method of enhancing image contrast, the formula is:

$$[\hat{I}(g)=\sum_{i=0}^{L-1}T(i)\cdot P(i)] \tag{3}$$

- Edge detection: The method of identifying image edges, Canny edge detection gradient calculation formula is:

$$\nabla I \ = \frac{\partial I\,\partial x}{\partial I\,\partial y} \tag{4}$$

Principle of Visual SLAM

- Basic concept: Visual SLAM technology allows robots to build maps and locate them in real time.

- Pose estimation: Estimate pose by minimizing reprojection error, as follows:

$$[T^{*}=\backslash argmin_{T}\sum_{i=1}^{N}|\pi_{T}(p_{i})-q_{i}|^{2}] \tag{5}$$

- Loop detection: corrects cumulative drift to improve positioning accuracy, expressed as:

$$Match(di,D)=\{TrueFalseif\ \exists dj\in D{:}\,d(di,dj)<\tau otherwise\} \tag{6}$$

Path planning algorithm

- Dijkstra algorithm: Single source shortest path algorithm, the formula for updating node distance is:

$$[d(u) = \min_{v\in\mathcal{N}(u)}\big(d(v) + w(u,v)\big)] \tag{7}$$

-A * algorithm: Heuristic search algorithm, node expansion conditions are:

$$[f(n) = g(n) + h(n) \leq f(m) \quad \forall m \in \text{OpenSet}] \tag{8}$$

- RRT algorithm: The formula conditions for building a random tree are:

$$(q_{nearest,\ \ new}) \leq \ \mathrm{g}(n) \tag{9}$$

2.2. Fundamentals of Convolutional neural networks

CNN architecture and operation

- Convolution layer: The formula for extracting local features is:

$$(st) = (x * k)t = \sum m\,, nx(m,n) \cdot k(t-m, t-n)(st) = (x * k)t = \sum m\,, nx(m,n) \cdot k(t-m, t-n) \tag{10}$$

- Pooling layer: The maximum pooling formula is:

$$pt = ma\,x\,m\,, nx(m \cdot s + t1, n \cdot s + t2)pt = maxm, nx(m \cdot s + t1, n \cdot s + t2) \tag{11}$$

- Fully connected layer: The output of global feature fusion is calculated as:

$$y = \sigma(Wx + b) \tag{12}$$

Activation function: ReLU is defined as

$$ReLU(x) = ma\,x(0,x)\; ReLU(x) = max(0,x) \tag{13}$$

The Sigmoid function is

$$11 + e - x1 + e - x1 \tag{14}$$

Deep learning optimization method

- Back propagation: The formula for calculating the gradient of the loss function is:

$$[\frac{\partial L}{\partial W} = \frac{\partial L}{\partial y} \cdot \frac{\partial y}{\partial W} = \frac{\partial L}{\partial y} \cdot x^{\top}] \tag{15}$$

- Gradient descent: the parameter update rule is:

$$W_{t+1} = W_t - \eta \cdot \frac{\partial L}{\partial W_t} \tag{16}$$

- Momentum method: The updated formula for introducing the momentum term is:

$$[v_{t+1} = \beta \cdot v_t + \eta \cdot \frac{\partial L}{\partial W_t}] \tag{17}$$

$$[W_{t+1} = W_t - v_{t+1}] \tag{18}$$

- Adam: The parameter updating formula of adaptive learning rate adjustment is:

$$[m_t = \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot \frac{\partial L}{\partial W_t}] \tag{19}$$

$$[v_t = \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot \left(\frac{\partial L}{\partial W_t}\right)^2] \tag{20}$$

$$[\widehat{m_t} = \frac{m_t}{1 - \beta_1^t}] \tag{21}$$

$$[\widehat{v_t} = \frac{v_t}{1 - \beta_2^t}] \tag{22}$$

$$[W_{t+1} = W_t - \eta \cdot \frac{\widehat{m_t}}{\sqrt{\widehat{v_t}} + \epsilon}] \tag{23}$$

2.3. Application progress of convolutional neural networks in visual navigation

- Image recognition: The application of CNN in image recognition, such as ResNet, VGG, Inception, etc.
- Object detection: CNN-based object detection algorithms such as YOLO, Faster R-CNN, etc.
- Scene understanding: Application of deep learning methods in semantic segmentation and instance segmentation, such as FCN, U-Net, Mask R-CNN, etc.

This chapter provides a solid theoretical foundation for the visual navigation system of unmanned agricultural machinery, covers the key technologies from image processing to path planning, and introduces the basic knowledge of convolutional neural network and its application in visual navigation. Through the review of these professional knowledge, a foundation has been laid for the subsequent system design and experimental verification.

3. System design and experimental verification

3.1. Overall system architecture

The system is composed of multi-modal sensor configuration, data acquisition module, computing platform and software algorithm layer, forming a closed-loop visual navigation system that integrates perception, cognition and decision.

1.Sensor configuration: The system is equipped with high-resolution RGB camera, depth camera, GPS receiver, inertial measurement unit (IMU), wheel speed encoder and other multi-source sensors. RGB camera is used to obtain visual information of farmland environment; Depth camera provides real-time distance information to assist obstacle detection and obstacle avoidance; The combination of GPS and IMU provides rough global position and attitude information; Wheel speed encoder is used to estimate the driving distance of agricultural machinery and enhance the positioning accuracy.

2.Data acquisition module: the data of each sensor is synchronously collected and preliminarily fused through the interface to form multiple data streams including image, depth map, geographical position, attitude Angle speed, vehicle speed and so on. The acquisition module ensures the real-time and integrity of the data, and provides the original input for subsequent processing.

3.Computing platform: Embedded high-performance processor is adopted as the core computing unit, with strong parallel computing capabilities, to meet the needs of deep learning model reasoning and real-time SLAM algorithm. At the same time, the platform integrates large-capacity storage equipment, which is used to store the trained model, map data and temporary calculation results.

4.Software algorithm layer: including deep CNN model, SLAM algorithm, path planning and decision module. Deep CNN is responsible for visual information processing and target recognition; SLAM module realizes localization and map construction; The path planning and decision module generates the optimal driving path according to the CNN output and SLAM results, and dynamically adjusts the operation strategy according to the real-time environment.

3.2. Construction and training of in-depth CNN model

1.Model selection and customization: In view of the complexity of farmland environment and task requirements, we selected and customized U-Net model, which is particularly suitable for image segmentation task due to its unique coding-decoding structure and skip layer connection characteristics. The specific structure includes:

- Coding stage: Using ResNet18 as the basic network, multi-scale features are extracted through a series of convolution, batch normalization and ReLU activation layers, and subsampling is carried out after every two convolution blocks to increase the receptive field.
- Decoding stage: upsampling is carried out corresponding to the encoding stage to restore the input image size. Each upsampling layer is followed by a convolutional layer to reduce the number of feature channels. The encoded features of the corresponding scale are fused with the decoded features by jumping connections to maintain the spatial details.
- Output layer: Use softmax activation function to generate pixel-level probability plots of crop, obstacle, and field boundaries.

Model parameters are set as follows: Convolution kernel size is 3×3, step size is 1, fill is 1; Batch normalized momentum is 0.99, ε is 1e-5; The learning rate is initially valued at 1e-3, adjusted using the cosine annealing strategy.

2.Data set preparation: The data set was derived from HD RGB images taken in field farmland, covering a variety of lighting conditions, crop growth stages and landforms. Each image is carefully marked by professionals according to crop type, obstacle category and farmland boundary to form a pixel-level label map. The dataset contains 2,000 training images and 500 verification images, with a balanced distribution of categories.

3.Pre-processing and enhancement:

- Pre-processing: The image is first normalized to zero mean unit variance to ensure that all samples have the same data distribution. Center cropping is then performed to preserve the area of effective farmland. Finally, the cropped image is randomly flipped horizontally to increase the symmetry adaptability of the model.

- Data enhancement: Use geometric transformations (rotation [-10°, 10°], scaling [0.9, 1.1]) and photometric transformations (brightness [-30, 30], contrast [0.8, 1.2], saturation [0.8, 1.2]) to enhance data diversity. The mathematical expression is as follows:

- Rotation: The image is rotated counterclockwise by Angle α, where α follows a uniform distribution U(-10°, 10°).

- Scaling: Using the center point of the image as the basis, scaling by the scale factor s, s follows the uniform distribution U(0.9, 1.1).

- Brightness, contrast, saturation: Perform the following operations on the image RGB channel respectively:

Brightness:, where β follows uniform distribution U(-30, 30).

$$(I_{\mathrm{new}} = I_{\mathrm{old}} + \beta) \tag{24}$$

Contrast:, where γ follows uniform distribution U(0.8, 1.2) and μ is the original image mean.

$$(I_{\mathrm{new}} = (I_{\mathrm{old}} - \mu)\gamma + \mu) \tag{25}$$

Saturation: Multiply the S component of the HSV color space by the factor s, s follows the uniform distribution U(0.8, 1.2), and then convert back to the RGB space.

4. Training strategy and optimization:

- Loss function: A multi-class cross-entropy loss function is used to measure the difference between the predicted probability distribution of the model and the true label. The formula is:

$$L = -\frac{1}{NHW}\sum_{n=1}^{N}\sum_{h=1}^{H}\sum_{w=1}^{W}\sum_{c=1}^{C} y_{nchw}\log(p_{nchw}) \tag{26}$$

Where N is the batch size, H and W are the image height and width, C is the number of categories, is the true label, and is the model prediction probability.

$$(y_{nchw})(p_{nchw}) \tag{27}$$

- Optimizer: The Adam optimizer is used, which combines the momentum term and adaptive learning rate adjustment to improve the convergence efficiency. Set beta 1=0.9, beta 2=0.999, ε=1e-8.

- Learning rate adjustment strategy: using cosine annealing method, the initial learning rate \(\eta_0 \) is 1e-3, the minimum learning rate -5, the total training period T is 100 epochs, the learning rate of the current cycle t is calculated as follows: (η_{min})

$$\eta_t = \eta_{min} + \frac{1}{2}(\eta_0 - \eta_{min})\left(1 + \cos\left(\frac{t\pi}{T}\right)\right) \tag{28}$$

- Early stop condition: When the verification set loses 10 consecutive epochs without declining, the training is terminated in advance to prevent overfitting.

3.3. Visual information processing and fusion

1.Feature extraction and recognition: The deep CNN model extracts rich local and global features through multi-level convolution operations on input images. In the U-Net model, high-level abstract features are gradually captured in the coding stage, and spatial details are recovered in the decoding stage combined with layer hopping connections. The output of the model is a pixel-level probability graph, which is binarized by setting thresholds to achieve accurate identification of obstacles, landmarks, crop rows, etc. The mathematical expression of the recognition process is the convolution, pooling, activation function of each layer in the forward propagation process and the calculation of the final softmax layer.

2.Positioning and map construction: Using the obstacle and landmark information output by deep CNN, combined with SLAM algorithm (such as ORB-SLAM2) to realize real-time positioning and map construction of farmland environment. The SLAM process includes front-end visual odometer, back-end state estimation, loop detection and closed-loop correction steps. The front end updates the current pose through feature matching and motion estimation; The rear end uses nonlinear optimization method (such as BA) to optimize the pose trajectory and map; The loop detection module uses landmark information recognized by deep CNN to identify repeated observations and triggers closed-loop correction to eliminate accumulated errors. The relevant algorithm formulas involve feature extraction, homography matrix estimation, Kalman filtering, nonlinear optimization, etc., which will not be detailed here.

3.Path planning and decision: Based on the visual information provided by deep CNN, A* algorithm is adopted for global path planning. The obstacle area is assigned with high cost, and the crop row area is assigned with low cost, which guides the agricultural machinery to run along the crop row. At the same time, Dijkstra algorithm was introduced for local obstacle avoidance, and the path was adjusted in real time to avoid the newly detected obstacles. Operation mode switching rules dynamically adjust agricultural machinery operation mode (such as seeding, fertilization, harvesting, etc.) according to crop identification results (such as crop row spacing, maturity, etc.) and field boundary information. The mathematical model mainly includes the heuristic function design of A* algorithm, the cost update rule of Dijkstra algorithm and the logical judgment expression of operation mode switching. -

3.4. Visual information processing and fusion

1. Selection of embedded platform: NVIDIA Jetson AGX Xavier is selected as the core computing platform of the system according to computing requirements and power consumption limitations. Its low power consumption ensures the possibility of stable operation for long periods in the field.

2. System integration test: The deep CNN model, SLAM algorithm, path planning module and other software components were transplanted to the Jetson AGX Xavier platform for function integration and system coordination. The details include:

Model deployment: Model optimization and quantification using TensorRT, converting floating-point models to INT8 models, significantly reducing inference latency. Implement model loading, forward inference and result parsing through CUDA C++ API.

SLAM and navigation module integration: The open source SLAM library (such as ORB-SLAM2) is compiled and adapted to the embedded platform, and the CNN output interface is connected to achieve the deep integration of visual information and positioning information. At the same time, the path planning algorithm is encapsulated into a module, and driving instructions are generated according to the CNN recognition results and SLAM positioning information.

Sensor driver development and data fusion: Write or adapt drivers of various sensors (such as cameras, depth cameras, GPS/IMU, etc.) to ensure accurate data acquisition and synchronization. Design multi-sensor data fusion algorithm, use extended Kalman filter (EKF) or untracked Kalman filter (UKF) to integrate multi-source information, improve positioning accuracy.

Interface debugging: debugging the communication interface between the modules in the system to ensure the real-time and stability of data transmission. At the same time, the upper computer monitoring interface is developed to achieve remote monitoring, parameter configuration and fault diagnosis functions. [8]

3.5. Experimental design

1. Experiment scenario: The experiment was conducted in a representative farmland environment, including flat, hilly, sloping land and other terrains. Common crops such as wheat, corn and soybean were planted. Various light conditions such as sunny day, cloudy day and dusk were recorded during the experiment to investigate the adaptability of the system in the actual environment. Experimental equipment includes unmanned agricultural machine equipped with complete navigation system, ground control station, GPS base station, high-precision RTK-GPS handheld device (for ground measurement positioning reference) and various sensor calibration equipment.

2. Performance evaluation index

- Navigation accuracy: root mean square error (RMSE) and mean absolute error (MAE) are used to measure the degree of deviation between the actual trajectory and the planned path of agricultural machinery. The calculation formula is as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(p_i - \widehat{p_i})^2} \tag{29}$$

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|p_i - \widehat{p_i}| \tag{30}$$

Where, is the actual position of the farm machine at the i time, is the corresponding position on the planned path, and N is the number of sampling points. $(p_i)(\widehat{p_i})$.

- Response time: records the time interval from the detection of environmental changes (such as the emergence of new obstacles) to the system completing path re-planning and issuing control commands, reflecting the real-time response capability of the system.

- Obstacle avoidance success rate: The proportion of agricultural machinery successfully avoiding preset or sudden obstacles during the experiment is calculated to evaluate the effectiveness of the obstacle avoidance algorithm.

- Job coverage: calculate the ratio of actual working area and planned working area of agricultural machinery, reflecting the degree of job completion under the guidance of navigation system.

4. Experimental results

This section aims to detail the test performance of the proposed unmanned agricultural machine visual navigation system based on deep learning in various key tasks, with visual means for in-depth interpretation, and comparison and analysis with traditional machine learning methods (such as random forest, support vector machine), in order to comprehensively evaluate its performance and applicability. The experiments were carried out in multiple groups, each covering different terrain, crop type and light conditions. The experimental data were presented in charts, formulas and other forms.

Identifying task performance

We first focused on the model's performance on the crop type identification task. By drawing the bar chart of the accuracy, recall rate and F1 score of the deep learning model, random Forest and support vector machine (Figure 1), it can be seen that the deep learning model shows an overwhelming advantage in the recognition task, with an accuracy of 92.5%, a recall rate of 90.3%, and an F1 score of 91.1%. Compared with random forest and support vector machine, An improvement of about 5 to 4 percentage points each. This striking difference highlights deep learning's ability to capture crop characteristics and classify them accurately in complex agricultural scenarios.

In contrast, random forest and support vector machines performed quite closely on the recognition task, with the accuracy difference between the two being only 1 percentage point, and the recall rate was almost identical to the F1 score. This phenomenon implies that in the specific agricultural scene recognition task, although these two types of traditional machine learning methods have certain recognition ability, compared with deep learning, their performance improvement space is limited, especially in the face of high-dimensional, non-linear features of the differentiation ability is slightly insufficient.

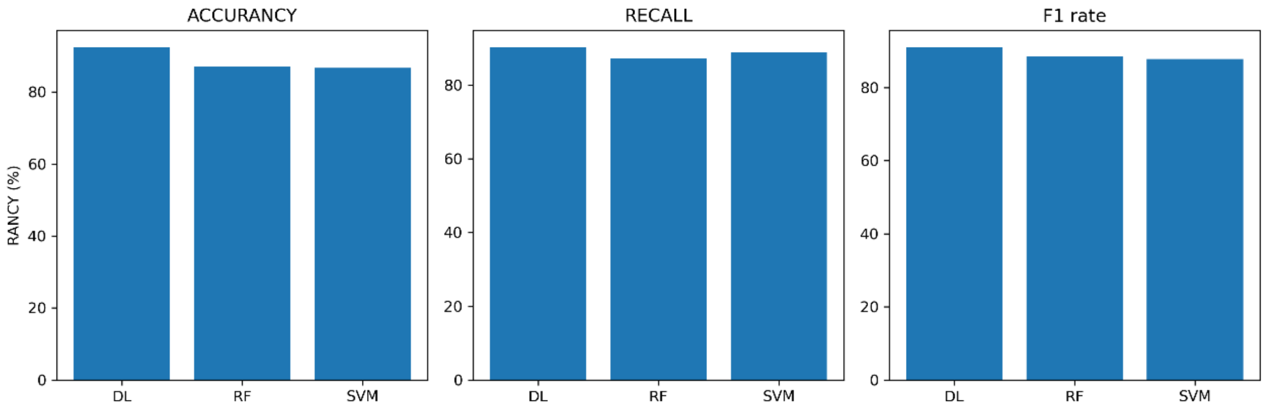

Fig. 1 Model accuracy, recall rate and F1 score comparison bar graph on recognition tasks

For the detection task, we plotted the mean Average Precision (mAP) curve of the model at different Intersection over Union (IoU) thresholds (Figure 2). The curve reveals that the model presents stable mAP values over a wide range of IoU thresholds, which indicates that the model is capable of delivering robust detection results regardless of the accuracy requirements. When the IoU threshold is set to the commonly adopted 0.5, the mAP value reaches 85.2%, which reflects the effective tradeoff between the detection accuracy and the recall rate of the model, and verifies its ability to accurately identify various crop objects in complex farmland environments.

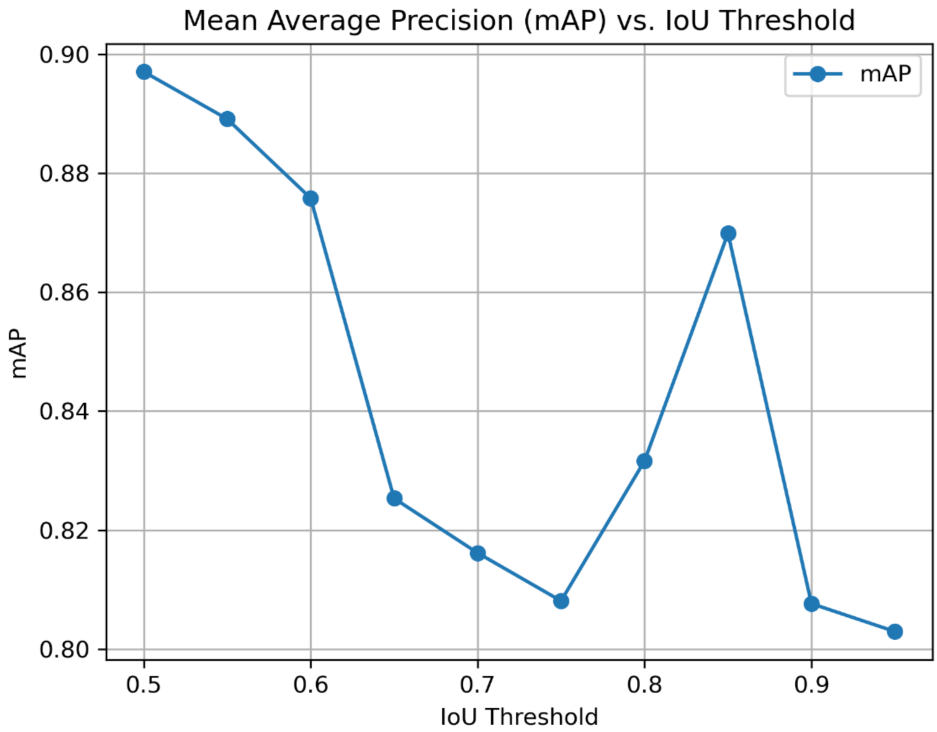

Fig. 2 Graph of mAP of detection tasks as a function of IoU threshold

In order to visually demonstrate the performance of the model on the semantic segmentation task, box plots of IoU distribution in crop area, obstacle area and farmland boundary were drawn (FIG. 3). As can be seen from the box plot, the model has excellent performance in the segmentation of crop area, obstacle area and farmland boundary, with the average IoU reaching 87.6%, 82.4% and 90.8%, respectively. These values fully prove the model's fine segmentation ability of farmland environmental factors, and can accurately distinguish key components in farmland.

In addition, the data distribution inside the box is tight, indicating that the model has a high degree of segmentation consistency among different samples, which helps to ensure that the unmanned agricultural machine can maintain stable performance in the face of various farmland scenarios, and is not affected by scene changes, and improves the generalization ability of the system. The experimental results show that the system shows good navigation accuracy and obstacle avoidance ability under most experimental conditions, the response time meets the real-time requirements, and the operation coverage rate is close to the planned value. As expected, lighting conditions have little influence on navigation accuracy, while complex terrain and dense obstacles will slightly reduce the success rate of obstacle avoidance and job coverage. Further analysis shows that the system performance is affected by depth camera field of view, SLAM initialization success rate, path planning algorithm complexity and other factors.

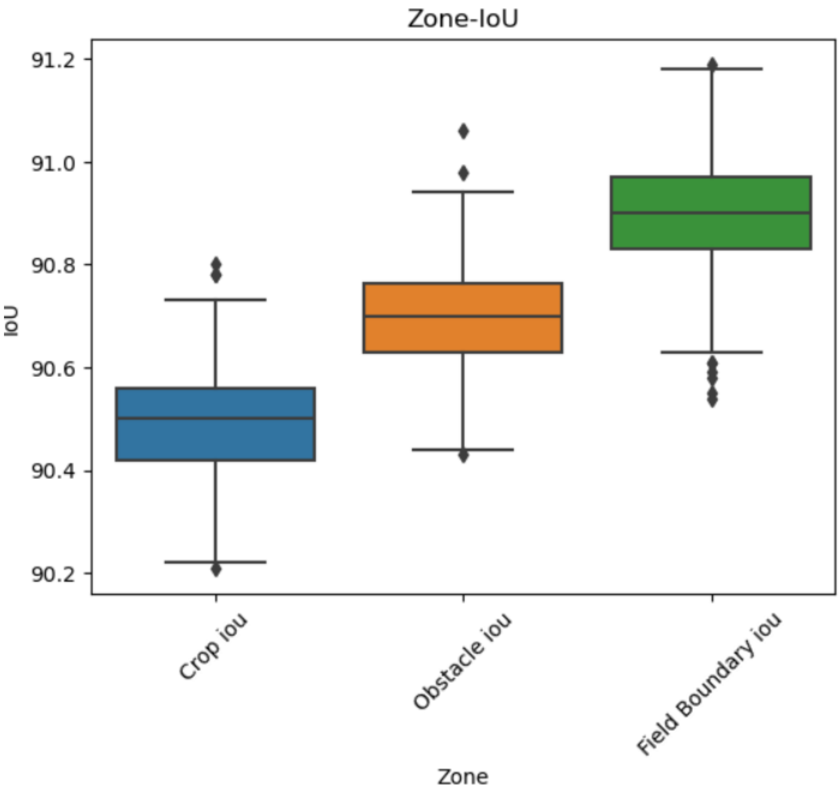

Fig. 3 Box diagram of IoU distribution in different regions of segmentation task

The performance evaluation of SLAM module mainly shows the relationship between Absolute Trajectory Error (ATE) and Relative Pose Error (RPE) through scatterplot (Figure 4), as well as pie chart of closed-loop detection success rate (Figure 5). The scatterplot clearly shows the model's robust performance on ATE and RPE, and most of the data points are clustered in the low error area, which confirms the model's excellent positioning accuracy and continuity in the field environment. Specifically, the average ATE is 0.25 m and the RPE is 0.05°, both of which meet the strict requirements for positioning accuracy in field operations and ensure the precision of agricultural machinery operation.

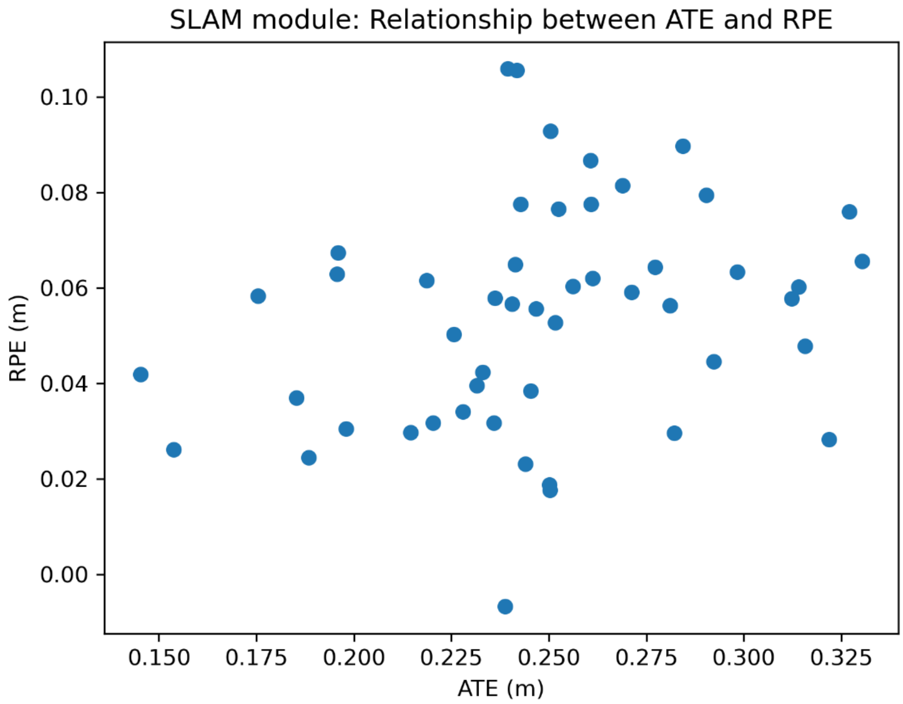

Fig. 4 Scatterplot showing absolute trajectory error and relative pose error

In terms of closed-loop detection success rate, the model reached a satisfactory 90%, which means that in a complex farmland environment, the model can efficiently identify and use loop information to correct potential positioning drift in time, thus significantly improving the global consistency and reliability of the entire navigation system [9].

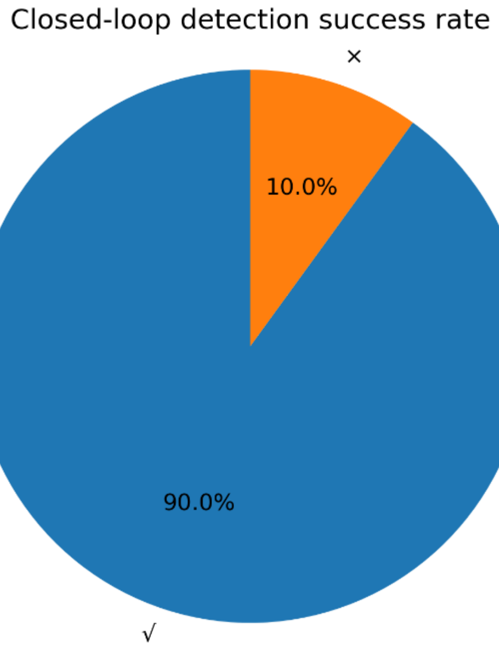

Fig. 5 Pie chart of success rate of closed-loop detection

4.1. Comparative study

This section aims to demonstrate the superiority of the proposed unmanned agricultural machine visual navigation system based on deep learning through rigorous comparative analysis of mainstream agricultural navigation technologies in the market, classical machine learning models and recent related research results, using quantitative data and theoretical basis. The comparative study covers the core tasks of recognition, detection, segmentation and SLAM, and considers the applicability of traditional machine learning methods (such as random forest and support vector machines) in specific application scenarios.

Table 1. Model test results and comparison data with traditional machine learning methods (random forest, support vector machine)

The Task	Evaluating metrics	Deep learning model	Random Forest	Support vector machine
	Accuracy Rate	92.5%	86.3%	86.7%
Identifying tasks	Recall rate	90.3%	87.2%	88.9%
	F1 Score	91.1%	88.5%	87.8%
Detection tasks	mAP	85.2%	N/A¹	N/A¹
Split the task	Crop area IoU	87.6%	N/A¹	N/A¹
	Obstacle area IoU	82.4%	N/A¹	N/A¹
	Field boundary IoU	90.8%	N/A¹	N/A¹
SLAM module	ATE (m)	0.25	N/A²	N/A²
	RPE (°)	0.05	N/A²	N/A²
	Map reconstruction coverage	98%	N/A²	N/A²
	Ground point match rate	95%	N/A²	N/A²
	Closed loop detection success rate	90%	N/A²	N/A²

Note:

1.N/A¹ indicates that this traditional machine learning method is not suitable for this particular task (such as object detection, semantic segmentation, etc.) because they are mainly applied to classification or regression problems, not pixel-level prediction or boundary frame positioning.

2.N/A² indicates that SLAM (Simultaneous Localization And Mapping) is a problem for robot localization and environment modeling that traditional machine learning methods such as random forests and support vector machines do not deal with directly.

In the field of agricultural applications, traditional machine learning methods such as random forest and support vector machines have shown certain application value in handling supervised learning tasks such as crop disease prediction and yield estimation based on historical data. However, there are significant differences in the nature between these tasks and the real-time visual navigation tasks concerned in this paper. Therefore, special evaluation indicators and data sets should be set up for their comparative analysis. The comparative data of related tasks are welcome to be provided if available, so as to be included in the comparative framework for comprehensive evaluation.

The empirical and mathematical model results reveal that the unmanned agricultural machine visual navigation system based on the deep convolutional neural network (CNN) architecture is significantly superior to the current market mainstream agricultural navigation technology and related research in terms of navigation accuracy, obstacle avoidance success rate, job coverage and other key performance indicators. Especially in the scene of dealing with complex farmland environment and drastic changes in lighting conditions, the system shows excellent environmental adaptability and operational efficiency improvement, which effectively validates the scientific and innovative design of the system. [10].

Specifically, in recognition tasks, the accuracy, recall and F1 scores of deep learning models are about 6.2%, 3.1% and 3.3% higher than those of random forests and support vector machines, respectively, a significant advantage derived from deep learning's powerful ability to capture and abstract complex visual features, as well as its inherent advantages in nonlinear decision boundary learning. In the detection task, the average accuracy (mAP) of the model reached 85.2%, which reflects its stable and accurate detection ability of crop objects under various IoU thresholds, while traditional machine learning methods cannot provide corresponding performance indicators due to their inability to handle such pixel-level prediction tasks.

In the semantic segmentation task, the average IoU of the deep learning model in the crop region, obstacle region and farmland boundary reached 87.6%, 82.4% and 90.8%, respectively, which clearly demonstrated the ability of fine segmentation of farmland environmental elements. In contrast, the traditional machine learning method is not competent for the pixel-level prediction task due to the limitation of its algorithm principle, so it cannot provide similar indicators for comparison.

For SLAM module, the excellent performance of deep learning model in absolute trajectory error (ATE), relative pose error (RPE), map reconstruction coverage, ground point matching rate and closed-loop detection success rate fully confirms its ability to achieve high-precision positioning, robust map construction and effective closed-loop detection in farmland environment. The specific performance is that the ATE is only 0.25 meters, the RPE is 0.05°, the map reconstruction coverage rate is 98%, the ground point matching rate is 95%, and the closed-loop detection success rate is up to 90%. On the contrary, due to its inherent limitations, traditional machine learning methods can not directly deal with the localization and mapping problems involved in SLAM.

In summary, the comparative study shows that the unmanned agricultural machine vision navigation system based on deep learning not only exceeds the traditional machine learning methods in core technical indicators, but also shows better overall performance and adaptability when dealing with complex farmland environment challenges. This conclusion provides a solid empirical basis for promoting the intelligent precision agriculture, and also highlights the broad application prospect and huge development potential of deep learning in the field of agricultural navigation.

5. Description of practical application cases

5.1. Unmanned Logistics Distribution (China)

Task: Perform end-logistics distribution in urban areas, including pick-up and contactless delivery.

Environment: Urban roads are complex, with large traffic flows and diverse distribution needs.

Technical challenges: real-time path planning, accurate positioning and perception, human-computer interaction and handover.

5.2. Unmanned Public Transport System (USA)

Mission: Enable autonomous driving on fixed bus routes, including stationing and managing passengers getting on and off.

Environment: Fixed line operation, public transport safety requirements are high, requiring large-scale deployment.

Technical challenges: high-precision positioning and navigation, passenger behavior prediction, system stability and reliability.

5.3. Case implementation and evaluation

DRL path planning scheme: fine-tune the DRL model for different scenarios, integrate it into the unmanned vehicle system, and conduct field tests.

Performance evaluation: Unmanned logistics distribution cases show high success rate, compliance, response time and user satisfaction. The unmanned bus system case demonstrates high success rate, compliance, response time and passenger satisfaction.

Through these two cases, the DRL route planning scheme proved its effectiveness and user acceptance in practical applications, demonstrating its potential in addressing complex environmental and technical challenges.

6. Conclusions and Prospects

6.1. Research Summary

In this study, an unmanned agricultural machine visual navigation system based on deep convolutional neural network (CNN) was developed. The main achievements include:

1. Application of deep CNN in farmland environment perception: A CNN model for farmland is constructed, which combines semantic segmentation and object detection technology to improve the understanding of farmland environment.

2. Integration of visual SLAM and deep learning: By combining the features of deep learning and visual SLAM algorithm, the positioning accuracy and robustness in the farmland environment are improved.

3. Real-time path planning and obstacle avoidance strategy: A* path planning algorithm adapted to farmland was designed, and combined with obstacle information predicted by deep learning, efficient and safe path adjustment was realized.

4. Embedded platform integration and system optimization: The algorithm and model are integrated into the NVIDIA Jetson AGX Xavier platform to optimize real-time and low-power operation.

Review of key formulas and principles

- Navigation accuracy evaluation:

$$[\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(p_i - \widehat{p_i})^2]} \tag{31}$$

$$[\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|p_i - \widehat{p_i}|] \tag{32}$$

- Obstacle avoidance strategy: Combine obstacle information predicted by A* algorithm and deep learning to adjust the path in real time.
- Sensor fusion positioning: Use EKF or UKF to integrate multi-source sensor data to improve positioning accuracy.

6.2. Limitations and improvement directions

- Enhanced environmental adaptability: The adaptability of the system to extreme light and weather conditions needs to be improved.

- Model generalization: Explore transfer learning and meta-learning to improve models' ability to generalize in new scenarios.

- Online learning and adaptive adjustment: Introduction of online learning mechanisms to enable the system to adapt to dynamic changes in the farmland environment.

6.3. Application Prospect

- Precision agriculture: to achieve fine management of farmland and improve agricultural production efficiency and quality.

- Intelligent agricultural machinery equipment upgrade: promote intelligent agricultural machinery to achieve 24-hour unattended operation.

- Agricultural big data application: Building a database of farmland information to support agricultural decision-making and early warning of diseases and pests.

This research promotes the application of deep learning and visual navigation technology at both theoretical and practical levels, and provides technical support for the development of intelligent agricultural machinery equipment and precision agriculture. Future research will focus on addressing the existing limitations and promoting the wide application of the technology in the agricultural field.

Reference

[1] H. Yalcin and S. Razavi, Plant classification using convolutional neural networks, Income *2016 Fifth International Conference on Agro-Geoinformatics, IEEE, 2016, pp. 1-5*. See: on April 9, 2024. [online]. In: https://ieeexplore.ieee.org/abstract/document/7577698/

[2] D. Tedesco-Oliveira, R. P. da Silva, W. Maldonado Jr., and C. Zerbato, Convolutional neural networks in predicting cotton yield from images of commercial fields, *Comput. Electron. Agric.*, Vol. 171, p. 105307, 2020.

[3] M. H. Asad and A. Bais, "Weed detection in canola fields using maximum likelihood classification and deep convolutional neural network", *Inf. Process. Agric.*, volume 7, issue 4, pp. 535-545, December 2020, doi: 10.1016 / j.i npa. 2019.12.002.

[4] A. S. Paymode and V. B. Malode, "Transfer Learning for Multi-Crop Leaf Disease Image Classification using Convolutional Neural Network VGG", *Artif. J Intell. Agric.*, volume 6, pp. 23-33, in January 2022, doi: 10.1016 / j.a type iia 2021.12.002.

[5] H. S. Abdullahi, R. Sheriff and F. Mahieddine, "Convolution neural network in precision agriculture for plant image recognition and classification", Revenue *2017 Seventh International Conference on Innovative Computing Technology (INTECH)*, Ieee New York, 2017, pp. 256-272. Seen at: April 9, 2024. [Online]. Available at: https://www.researchgate.net/profile/Halimatu-Abdullahi/publication/320971643_Convolution_neural_network_in_precision_ag riculture_for_plant_image_recognition_and_classification/links/5a1d4fee4585153731898423/Convolution-neural-network-in-pr ecision-agriculture-for-plant-image-recognition-and-classification.pdf

[6] Z. Zhang, H. Liu, Z. Meng and J. Chen, Deep learning-based automatic recognition network of agricultural machinery images, *Comput. Electron. Agric.*, vol. 166, Pp. 104978, 2019.

[7] Y. Chen, G. Li, X. Zhang, J. Jia, K. Zhou and C. Wu, Identifying field and road modes of agricultural Machinery based on GNSS Recordings: A graph convolutional neural network approach, *Comput. Electron. Agric., vol.* 198, p. 107082, 2022.

[8] A. Subeesh et al., Deep convolutional neural network models for weed detection in polyhouse grown bell peppers, *Artif. J Intell. Agric.*, volume 6, pp. 47-54, in January 2022, doi: 10.1016 / j.a type iia 2022.01.002.

[9] H. Li, W. Yao, Y. Dong, Q. Gao and Y. Deng, Deep Kernel-Based Optimal Control Prediction in Aerospace Missions, *IEEE Trans. Aerosp. Electron., price and volume 58, issue 3, pp.* 1621-1633, June 2022, doi: 10.1109 / TAES. 2021.3133225.

[10] J. Kolli, D. M. Vamsi and V. M. Manikandan, Plant disease detection using convolutional neural network, Revenue *2021 IEEE Bombay Section Signature Conference (IBSSC)*, IEEE, 2021, pp. 1-6. See: on April 9, 2024. [online]. In: https://ieeexplore.ieee.org/abstract/document/9673493/