Image Processing Method of a Visual Communication System Based on Convolutional Neural Network

Liang Sun, Shijiazhuang Institute of Railway Technology, China
Pengsheng Wang, Shijiazhuang Institute of Railway Technology, China*
Paiying Liu, Shijiazhuang Institute of Railway Technology, China
Zhengang Nie, Shijiazhuang Institute of Railway Technology, China

ABSTRACT

Unmanned motion platforms are being used in a wide range of industries. Unmanned motion platforms must have an autonomous and intelligent navigation procedure in order to carry out their system functions. Traditional inertial navigation and radio navigation have poor accuracy and autonomy when not dependent on satellite circumstances. The accuracy of image recognition algorithms must meet strict standards. This study and exploration of the high-precision scene image recognition system is based on convolutional neural network structure optimization. To demonstrate the viability of the approach, simulation experiments are carried out on the NUC dataset using the recognition technique based on a convolutional neural network that is proposed. The fundamental network architecture of a convolutional neural network is optimized using the L2 regularization technique. The experimental findings demonstrate that the NUC dataset now has better recognition accuracy. In terms of recognition accuracy, the suggested method satisfies the predetermined requirements.

KEYWORDS
Brain-Like Navigation, Convolutional Neural Network, Image Recognition, Saliency Detection, Visual Communication

INTRODUCTION

Visual media are used to provide a variety of information to individuals in the Internet era. Images have a significant role in the transmission of information in visual communication. High-quality photos can enhance the transmission effect and provide a more comprehensive and useful message. However, a number of externally undesirable elements cause the image quality to decrease during the formation and transmission of the image. Ensuring that the photographs are aesthetically pleasing requires the adoption of image processing techniques for image repair and enhancement.
Convolutional neural networks can significantly increase image identification rates allowing for more effective data mining of visual information. As artificial intelligence technology develops, people are becoming increasingly concerned about deep learning. The recognition rate is low, and the typical picture recognition technique is comparatively out-of-date. In light of the vast amount of image data available, it is clear that the classic recognition method cannot satisfy the current needs. This essay examines how image processing technology is used in the context of visual communication. The picture recognition algorithm is optimized using convolutional neural network theory.


Xu et al. (2018) studied deep convolutional neural network based autonomous marine vehicle maneuvering. To improve it, an autonomous collision avoidance method based on vision technology as a human visual system was proposed (Xu et al., 2018). Zhao et al. (2018) proposed a method for input image data processing and autonomous motion estimation using convolutional neural networks. Gavali et al. (2019) study deep convolutional neural network for image classification on cuda platform. The image classification process is performed using LeNet network model (Gavali et al., 2019). Bosse et al. (2019) derive the concept of distortion sensitivity as a property of the reference image which compensates for the lack of perceptual relevance of a given computational quality model potential: Convolutional neural network (CNN) for image feature extraction combined with recurrent neural network (RNN) based on attentional mechanism for automatic lip-reading recognition (Bosse et al., 2019).

The use of unmanned motion platforms in the military and in daily life has gradually grown into a new research hotspot thanks to the ongoing advancements in artificial intelligence and computer science technologies. Additionally, it applies to more complex scenarios, such as those in the military, the aerospace industry, the medical field, and daily life (Lu & Li, 2019). Unmanned motion platforms must be capable of performing self-positioning, map construction, and path planning in the working environment to be able to operate autonomously. Examples of such platforms include unmanned vehicles, unmanned aerial vehicles, and mobile robots (Peng et al., 2020).

The inertial navigation system is the most popular autonomous navigation and positioning technique for unmanned motion platforms, but it has a dead reckoning offset error, and its position estimation error will keep growing over time, making it impossible to perform high-precision, long-duration autonomy. Relevant researchers concentrate on animals with autonomous navigation abilities to investigate higher-precision and more dependable navigation methods and means and improve the autonomy and intelligence of the navigation system. John O’Keefe, an American researcher, initially looked into the relationship between the hippocampus and spatial localization cognition in the animal brain in 1971 (Trnovszky et al., 2017), revealing place cells (location cells) for localization.

Grid cells, another crucial component in the process of autonomous navigation, were first made public in 2005 by the Norwegian scientist Moser and his wife. The principles of bionic and brain-inspired navigation in simulated animal navigation modes were codified with the discovery of these two navigational cells that make up the brain’s localization system, and the technology of bionic brain-inspired navigation advanced quickly. In contrast to conventional navigation, brain-inspired navigation technology has the traits of “self-learning and self-evolution” to the environment, i.e., navigation planning has the capacity for “memory” and “learning,” and its navigation system is an intelligent system, capable of realizing the carrier’s autonomous navigation and trajectory planning functions.

Traditional navigation technology primarily relies on sensor measurement data to determine the carrier’s position, attitude, and speed in the current environment. The primary goal of brain-like navigation technology is to enable autonomous navigation and localization in the workplace by reducing reliance on sensor measurements and enhancing the carrier’s own “cognitive” and “memory”
aptitude to the environment. The precise steps of brain-like navigation include using visual sensors to initially gather environmental data, extracting depth information through algorithms (simulating the firing effect of the equine body in the biological brain), enhancing the cognitive ability of the environment, and creating an adaptive process of “self-learning and self-evolution.”

Finally, it completes the carrier’s autonomous control, trajectory planning, and navigation functions (Huang et al., 2015). The animal brain navigation cell model served as the foundation for the brain-like navigation technique presented in this paper. The issue of the inertial navigation system’s subpar long-term accuracy is effectively resolved by setting the position unit node in the trajectory of the autonomous motion platform and executing position recognition at the node. Whether the system position inaccuracy may be properly repaired in this procedure depends on whether high-precision scene identification can be completed at the position unit node.

Robots can imitate human analysis and learn to tackle complicated picture recognition problems using deep learning since they have the capacity to learn the intrinsic rules and expression level of samples throughout the training process. The subject of picture and text recognition, in particular, has a wide range of potential applications for deep learning-based image recognition technology, which can input sample data immediately for processing and train basic network models to learn image attributes. Convolutional neural network (CNN), a crucial deep learning model, can learn the relevant useful features from a huge number of picture samples without having to go through a laborious feature extraction process (Jiang et al., 2020). It is very suitable for image recognition in complex environments. Therefore, combining the image recognition technology based on deep learning with the brain-like navigation method can improve the scene recognition accuracy of the position unit nodes, thereby realizing the correction of the position error of the inertial navigation system, and improving the autonomy and intelligence of the unmanned motion platform. In image recognition, each neuron in the input layer may represent the grey value of a pixel. However, there are several problems in the application of this neural network to image recognition. One is that the spatial structure of the image is not considered, and the recognition performance will be limited; second, the neurons of each adjacent two layers are all connected, and the training speed is limited due to too many parameters. Convolutional neural networks can solve these problems. The convolutional neural network uses a special structure for image recognition and can be trained quickly. Because of the high speed, it is easy to use the multi-layer neural network, and the multi-layer structure has great advantages in recognition accuracy.

This paper provides a systematic description of image recognition algorithms and introduces the characteristics of different recognition algorithms. In order to achieve high-precision scene image recognition, the structure of the convolutional neural network is optimized and combined with the saliency detection algorithm. Simulation experiments were conducted on the NUC dataset, which basically proved the feasibility of the algorithm. However, the practical application of scene recognition algorithms is complex and requires denser real data for testing. Therefore, the algorithm proposed in this paper has room for further improvement.

Compared with ordinary neural networks, convolutional neural networks are more suitable for processing images because they simplify the preliminary image processing and can directly input unprocessed images, which are widely used in the fields of healthcare, transportation, and security. Convolutional neural networks consist of learning weight neurons and bias neurons. Each neuron can receive data input and perform a dot product. The final fully connected layer has a loss function. Each layer of the CNN can convert one value to another through an activation function.

Combining deep learning-based image recognition technology and brain-like navigation methods to improve the scene recognition accuracy of positional cell nodes, and then realizing the correction of positional error of inertial guidance system, has an important role in promoting the autonomy and intelligence of unmanned motion platforms. Therefore, this paper combines the image recognition technology based on deep learning and brain-like navigation method to improve the scene recognition accuracy of place cell nodes, and then realize the correction of the position error of inertial navigation.
system, which plays an important role in promoting the autonomy and intelligence of unmanned motion platform.

The innovation of conducting simulation experiments on the NUC dataset in conjunction with saliency detection algorithms is that new domain crossovers are explored through this fusion, providing new perspectives on the application of saliency detection algorithms. At the same time, this may also introduce new algorithm designs for specific domain requirements, bringing innovative approaches to domain development. The final experimental results are also expected to provide valuable insights into the related fields, revealing the potential impact of the application of saliency detection in the nuclear energy field, and thus advancing the related fields and the field of saliency detection.

**MATERIALS AND METHODS**

In 1965, studies on digital picture processing and recognition were begun. Digital images have a number of significant advantages over analog ones, including simple processing, simple storage, simple transmission, and simple compression. These benefits act as a powerful catalyst for the advancement of image recognition technologies. However, it was not widely employed at the time due to the limitations of technology and the requirement for hardware development. The development of image recognition technology has been made possible by the ongoing advancements in computer performance. Template matching and deep learning are the techniques for picture recognition that are most frequently utilized (Wang et al., 2017).

The simplest picture recognition technique is the template matching algorithm. By establishing a set of common templates and then matching them on the target image using a sliding window, depending on the degree of matching, algorithm, achieves target categorization. To locate the sections that are substantially comparable to the target image, including the direction and size, the specific procedure is to first create a set of templates and then compare these templates with the target image. The precise location of the target image can be identified after the matching is finished. However, the template matching algorithm has some drawbacks and limitations. First, it needs to design templates of similar targets in advance, thus requiring high research experience of the designer. Second, during the matching process, the detection window can only be moved in parallel, and it cannot deal with the situation of target image rotation or size change, which leads to the failure of the algorithm. In order to solve the limitations of the template matching algorithm, scholars at home and abroad have conducted a series of studies and introduced more complex image recognition algorithms, such as convolutional neural networks based on deep learning. These methods can better deal with the target image rotation, size change, and other complex situations, significantly improving the accuracy and robustness of image recognition.

The sequential similarity detection algorithm (SSDA) and the traditional mean absolute difference algorithm (MAD) were consecutively proposed by Leese and Barnea (Lou & Shi, 2020), which set the groundwork for the later development of template matching algorithms. A brand new diffraction template matching technique was put forth by Atsushi et al. of Kyoto University in Japan in 2020 (Yang et al., 2021). Assuming the candidate structure, a reasonable three-dimensional structural model is extracted based on the diffraction pattern of a single noise target, which improves the speed of diffraction template matching.

In the same year, Rodriguez-Rodriguez (Yahya et al., 2021) of the University of Las Palmas, Canary Islands first segmented the image to obtain the shape of the target character and the detected character and used the template matching scheme based on bit operations for identification. The results showed that the error rate of this method was only 0.21%.

Domestic scholars have also paid a great deal of attention to how to strengthen the picture recognition algorithm based on template matching and increase its applicability in difficult contexts. For the issue that closed-loop detection in the conventional Rat SLAM algorithm is easily influenced by illumination, Soochow University proposed an FT model in 2020. This model converts RGB
images into saliency maps and extracts images containing more original image feature information from the saliency map (Lee et al., 2021). The visual scene recognition in closed-loop detection is more reliable thanks to the visualization template.

In the same year, Yang Yang (Miao, 2022) of Xi’an Jiaotong University proposed a template matching method based on large-angle positioning, which iteratively searches for the corresponding feature blocks, updates the template position through rotation transformation, and finds the corresponding features between the template image and the target image. The detection accuracy is improved when the rotation in the target plane is large.

In 2021, the Shenyang Institute of Automation, Chinese Academy of Sciences (Su et al., 2022) adopted an initial free template matching method based on the principal direction Fourier transform operator, which avoided the allocation of initial values and expanded the capture range of the sliding window, with a matching accuracy of 0.72mm. The single matching time is 16s, which greatly improves the matching efficiency.

Deep learning has a lot of potential in many domains and is one of the key algorithms in the field of image recognition. Deep learning can explain the traits and characteristics of objects more deeply and abstractly. The goal of machine learning is to identify a function during the model-training process and obtain a specified output from a given input. There are numerous similar functions, though. Finding a function with generalizability requires using the deep learning training model for trial and error. Deep learning has a multi-layer perceptron structure, which combines low-level characteristics to create more abstract high-level features that are then represented by distributed data features.

Since the concept of deep learning was first proposed by professors such as Hinton (Xie & Lv, 2021) in 2006, it has triggered a wave of research by foreign scholars. In 2019, Andong et al. (Ahn, 2020) of Texas A&M University proposed a single image-based long-short-term memory artificial neural network (LSTM) sequence feature construction strategy, using pixel matching (PM) and block matching (BM). Two pixel-based similarity measurement methods select candidate sequences from the whole image and use matching pixels with high similarity as the input of LSTM, and the overall accuracy of final image classification is up to 96.20%.

In 2020, Ghaza (Anwar et al., 2018) of the University of Southern Queensland designed a hybrid deep learning neural network EDLM, which uses convolutional neural networks to extract image features and then fuses CNN and recurrent neural networks to generate image classification features, which have better recognition accuracy and generalization, transformation ability.

A face recognition framework based on CNN and RNN was proposed in 2021 by Rizhao (Kumar, 2020) of Nanyang Technological University in Singapore. It uses deep learning to model face information, a cyclic mechanism to learn local image information, and finally fuses local and global information to complete high-precision classification. The University of So Paulo’s Mendes et al. (Zhang et al., 2018) introduced a new lightweight and quick supervised CNN architecture together with a new feature extraction model in 2021 and used it for both indoor and outdoor robot navigation. In 2019, Zhang Ke of North China Electric Power University (Lu & Li, 2019) proposed the idea of introducing the squeeze excitation module into the DenseNet framework and constructed a new MFR-Dense Net structural model, which can more accurately extract image features and complete image classification (Valipour et al., 2016). In 2020, Xu Xiaowei of Huazhong Agricultural University proposed an enhanced classification method based on recurrent neural network (RNN) and random forest (RF), and the accuracy of remote sensing scene image classification reached 87%. In 2020, Nanjing University of Aeronautics and Astronautics used a short-time Fourier transform to process data, and finally used a convolutional neural network combined with RNN to classify data to achieve stable and high-precision recognition (Huang et al., 2015).

Recent years have seen significant changes in image recognition technology as a result of the quick development of related theories and technologies. Training data sets have grown in size, feature extraction algorithms have gradually tended to learn visual features, and training models have grown in complexity (Peng et al., 2020). Theoretical technology has eventually given way to real-world
application scenarios. Deep learning approaches in particular have made enormous strides in the field of image recognition, but they still have a long way to go. More study is still needed to understand how to construct networks for quick training of high-performance models (Rui et al., 2019).

**DISCUSSION**

The core strategy of brain-like navigation adopted in this paper is to build a place cell node model and give each node accurate location information; the next strategy is to perform scene matching at nodes. If the recognition is successful, the accumulated error of the system can be corrected, and navigation commands can be adjusted. The image recognition algorithm based on deep learning can adapt well to the changes in lighting and angle of the scene and can quickly and robustly learn image features and complete accurate recognition. This article uses a deep learning algorithm based on convolutional neural networks to complete position recognition.

Convolutional neural networks initially entered the public’s field of vision thanks to Yann Lecun in 1989. Convolutional neural networks have been popular thanks to the development of deep learning algorithms and advancements in computer hardware performance (Yang et al., 2020). Convolutional neural networks are seen as having a unique hierarchical structure, much like conventional network models, with the exception of the types and functions of the layers. The convolutional neural network is primarily composed of a number of fundamental layer structures, including the data input layer, convolution layer, activation layer, pooling layer, and the fully connected layer (Anwar et al., 2018). The convolution layer is made up of several feature planes, and each feature plane is made up of many identically connected neurons (Pang et al., 2023). A back-propagation algorithm can train it to improve the image’s ability to be recognized. In Figure 1, the processing impact is displayed (Zhang et al., 2023).

The operation process of convolution of image and convolution kernel can be expressed as:

\[
S(i,j) = (K^*I)(i,j) = \sum_{m} \sum_{n} I(i+m, j+n)K(m, n)
\]

(1)

Among them, I represents a two-dimensional matrix composed of image pixel values (coordinates are (m, n)), K is the weight parameter optimized by the algorithm, and s is the feature map output after convolution (coordinates are (i, j)).

Animals perceive external things from local to global and always receive information from a small area first. Similarly, each image has a certain organizational structure. While the link to the

![Figure 1. Image preprocessing effect diagram](image-url)
global point, which is further away, is somewhat weak, the connection to the local point is very strong. This serves as inspiration for the idea that neurons extracting image characteristics need not detect every point in the sample data, but rather, only those pixels with significant connection, which they can then integrate to gain the overall information of the image.

As seen in Figure 2, the local connection strategy drastically reduces the number of parameters, considerably enhancing the network’s training time, which is crucial for advancing neural networks’ use in image recognition.

Although each layer of the neural network transforms images in a linear manner, not all of the data samples used to train the model are linearly separable. Therefore, a nonlinear function (also known as an activation function) is introduced to the network structure in order to improve the expressiveness of the network model, eliminate redundant information, and prevent overfitting. The output of each layer is a linear function of the input of the upper layer if the activation function is not utilized. The neural network’s output is always a linear combination of inputs, regardless of how many layers it contains. The most basic perceptron is this one. In artificial neural network models, activation functions are crucial for learning and comprehending extremely complex and nonlinear functions.

The Sigmoid function’s output range is between 0 and 1, and formula (2) shows its mathematical expression. However, a vanishing gradient issue could result from the network model using the sigmoid activation function. Additionally, the function’s curve does not pass through the origin, a condition that lowers the effectiveness of updating weights, and the function’s exponential operation uses a lot of computer power. The Sigmoid function gradually left the historical stage in practical applications for a number of reasons:

\[
\sigma(x) = \frac{1}{1 + e^{-x}}
\]  

(2)

The Tanh activation function is a hyperbolic tangent function, similar to the Sigmoid function. When the input is close to infinitely small and infinite, the weight update efficiency of the Tanh

Figure 2. Schematic diagram of full connection mode and partial connection mode
function is still not high. In addition, the Tanh function still includes exponential operations, and the problem of high computational complexity remains unsolved. However, its output range is changed to (-1,1), which solves the disadvantage that the Sigmoid function is not centered at the origin and is generally used in binary classification problems. Its mathematical expression and function image are shown in formula (3):

\[
\tanh(x) = \frac{\sinh(x)}{\cosh(x)} = \frac{e^x - e^{-x}}{e^x + e^{-x}}
\]  

(3)

The linear correction unit, or ReLU function, is a segmented curve with the origin as the limit. Additionally, the function does not use an exponential operation, which simplifies the calculation and increases calculation speed. Additionally, the linear and unsaturated design results in a substantially faster rate of convergence. It should be noted that a suitable learning rate must be supplied while using the ReLU function. The training procedure could end with the irreversible death of the neuron node if the learning rate value is set too high and the parameter update speed is set too quickly. Although the ReLU function has certain defects, it has still become the most commonly used activation function. The mathematical expression and function image are shown in equation (4):

\[
\text{ReLU}(x) = \max(0, x)
\]  

(4)

The exponential linear unit (ELU) function is an improvement over the ReLU function, whose output average is close to zero, speeding up learning. At the same time, ELU increases the non-negative output when the input is negative, reduces the possibility of gradient disappearance, and avoids the problem of irreversible neuron death. However, it still needs to perform exponential operation, which is not efficient in calculation, and has the problem of gradient saturation. Its mathematical expression and function image are shown in formula (5):

\[
\begin{align*}
    &x, x > 0 \\
    &\alpha (e^x - 1), x \leq 0
\end{align*}
\]  

(5)

The mean square error loss function can complete the prediction of specific values, and its expression form is shown in formula (6):

\[
MSE = \frac{1}{N} \sum_{i=1}^{N} \| y_i - a \|^2
\]  

(6)

In the application of linear support vector machines, for some linearly inseparable data, the hinge loss function introduces slack variables, and its mathematical expression is shown in equation (7):

\[
L(y(w \cdot x + b)) = [1 - y(w \cdot x + b)]^+
\]  

(7)

Among them, “+” represents a function that takes a positive value, as shown in formula (8):
Cross-inheritance can measure the difference between the real probability distribution and the predicted probability distribution. SoftMax loss is the most used cross-inheritance loss function, which can output the conditional probability of each category of the image set: the SoftMax function and the calculation of the cross-inheritance loss. The formulas are shown in formula (9) and formula (10), respectively:

\[
p_i = \frac{e^{z_i}}{\sum_{j=1}^{k} e^{z_j}}
\]  

Among them, \(z\) represents the network output (prediction score) of class \(i\) and \(p_i\) represents the probability of class \(i\) (\(k\) is the number of classes):

\[
L = \frac{1}{N} \sum_{i=1}^{N} \left( \sum_{j=1}^{k} y_{ij} \log p_i \right)
\]  

The value of iteration can also represent the number of times of model training. The meaning of an iteration is that the convolutional neural network uses a batch of image samples to perform a complete parameter update. The calculation method is shown in formula (11):

\[
\text{number of batches} = \text{iteration} = \frac{\text{training set size}}{\text{batch size}}
\]

The enhancement algorithm based on color space transformation can increase the illumination difference, brightness difference, and color difference of the training set. Common methods include brightness/contrast transformation, histogram equalization, and HSV color space transformation. H in the HSV color space represents hue, S represents saturation, and V represents brightness. The conversion formula from RGB space to HSV space is shown in formula (11). However, based on the research background of this paper, the image set used for convolutional neural network training All are typical landmark scenes and the effect of inputting the HSV color space transformed image data into the network is not obvious, so only a very small amount of data adopts this algorithm:

\[
s = \begin{cases} 
0, & \text{if } \max = 0 \\
\frac{\max - \min}{\max}, & \text{otherwise}
\end{cases}
\]

\[
v = \max
\]

Distance measurement: For each pixel that is searched, calculate its color distance and spatial distance in the labxy color image plane space and the cluster center, respectively. Finally, the two distances are fused to obtain the final metric distance, and the calculation formula is shown in formulas (13) to (15):

\[
[z]_i = \begin{cases} 
z, & z > 0 \\
0, & z \leq 0
\end{cases}
\]  

(8)
\[
d_c = \sqrt{(l_j - l_i)^2 + (a_j - a_i)^2 + (b_j - b_i)^2}
\]
\[
d_s = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2}
\]
\[
D' = \sqrt{\left(\frac{d_c}{N_c}\right)^2 + \left(\frac{d_s}{N_s}\right)^2}
\]

RESULT ANALYSIS

This article selects VGGNet as the basic convolutional neural network model. TensorFlow builds the basic framework and optimizes and improves the structure based on practical applications. TensorFlow is a second-generation machine learning system developed by the Google team, which is very suitable for the field of deep learning. The infrastructure is shown in Figure 3.

The TensorFlow system follows a good hierarchical architecture, including two parts: the front-end and back-end. The front-end part is responsible for building data flow diagrams, while the back-end is responsible for performing various calculations. The client is an important component of the front-end system and is used in the TensorFlow framework to construct computational diagrams using programming language as a channel. The back-end system runs based on different layer structures. Among them, the computing layer is responsible for implementing specific mathematical operations. The networking layer enables data exchange between components and supports RDMA communication between nodes. The device layer is responsible for supporting multiple computing devices.

Figure 3. TensorFlow infrastructure
In image recognition problems, the most important thing is the predictive ability of the model. In order to prevent overfitting and improve the generalization ability of the model, L2 regularization algorithm is introduced in this paper. L2 regularization method prevents the network model from fitting interference information, such as complex background, in the training set and reduces the complexity of the model by limiting the weight value of the largest number of parameters in the model. The principle is shown in Figure 4.

The elliptical region represents the loss function that needs to be optimized, while the perfectly circular region represents the limiting condition. When there are no limiting conditions, use the gradient descent method to find the optimal solution along the negative direction of the gradient (blue arrow). When there are constraints, it is necessary to iterate within the yellow area, which can only be located at the top edge of the circle at most in order to minimize the training sample error.

To verify the effectiveness of the convolutional neural network model proposed in this paper, a series of comparative experiments were conducted to compare the network model with typical LeNet5 models, AlexNet models, and SeqSLAM algorithms for loss curves and training accuracy curves. The LeNet5 model is a shallow convolutional neural network with a seven-layer network structure, which has achieved excellent results in smaller and simpler feature sets of images. AlexNet has deepened the network layers, made the model more complex, and opened up the application of CNN on large-scale datasets. SeqSLAM is a location recognition algorithm based on appearance, which completes sequence image matching under harsh environmental conditions and can maintain high accuracy and low time complexity. Due to the fact that the SeqSLAM algorithm does not involve the training process of CNN models, this article only uses it as a comparative algorithm in terms of image recognition accuracy.

The NUC dataset is an image set captured by drones in a typical scene of North Central University, including 10 scenes and a total of 3670 original images, which have been augmented to a certain extent by data enhancement algorithms.

This experiment is carried out in the PyCharm simulation environment and library functions such as TensorFlow 1.14.0, NumPy 1.16.2, and matplotlib 3.0.3 are installed. Before the formal training starts, first set the parameters, set the epoch to 20 and the batch to 9, and obtain the model loss curve and accuracy curve, as shown in Figures 5 and 6.

The NordLand dataset is a train journey captured in four different seasons. To verify the adaptability and robustness of the model under the changing lighting environment, this paper uses

Figure 4. L2 regularization method
the images of spring, autumn, and winter in the NordLand dataset as the training set, and the summer images as the test set, and trains different CNN models respectively, as shown in Figure 7, for the resulting loss curve.
By training different network models, the loss curve and the accuracy curve can be obtained, as shown in Figure 8 among them, because the initial value of the loss curve is too large, the similarity between the curves is slightly higher. Therefore, this article additionally shows the curve of the training number of 7 to distinguish.

It can be seen from a series of simulation experiments that the CNN model in this paper performs very well on different datasets. Among them, the loss curve of each model on the NUC dataset basically converges after 8 times of training, and the model in this paper performs best on the accuracy curve. The images of the same scene in the NordL and dataset do not have much angle change, and the difference in features is mainly reflected in the changes of illumination and seasons. The loss curves of the three models are not very different. In terms of the accuracy curve, the model in this paper is still the best, followed by AlexNet. The image characteristics of different scenes in the round point data set are obviously different, and each model can converge in a short time. Although there are different degrees of fluctuations in the early stage, the difference between curves is generally small, and the convergence of loss curve and accuracy curve can achieve better values.

After the completion of the experiments, the different recognition algorithms are tested for accuracy using the test image set for each dataset. The results are shown in Table 1.

As can be seen from Table 1, the recognition accuracy of this paper’s algorithm on different datasets can basically reach more than 90%, and it can adapt to the changes in image angle and illumination in complex environments. Among them, the images in the NordLand dataset have obvious seasonal changes, and the SeqSLAM algorithm has stronger robustness in recognizing them, but the algorithm proposed in this paper also achieves good results. For the NUC dataset and Garden Point dataset, the algorithm in this paper shows great advantages, and the accuracy is higher than the other three algorithms.

To further illustrate the applicability and superiority of this paper’s algorithm in practical applications, flight experiments are set up. The UAV is utilized to make several flights at an altitude of 300 meters, and seven nodes are set up on its flight trajectory to collect enough scene images to train the CNN model.
Figure 9 demonstrates that the model finds convergence and experiences no significant fluctuations during the fourth training. The drone took 640x480 photographs, of which seven scenes were included. A total of 999 images were used to train the model, 104 images were utilized for prediction, just two images had errors, and the test’s recognition accuracy rate reached 98.08%. This research convincingly proves the algorithm’s high precision properties. Additionally, the average time spent predicting images is 0.03798s/frame, and the UAV’s flight speed is roughly 4m/s, so the target scene recognition may be essentially finished in time and fulfill the demands for real-time performance.

This chapter mainly includes two parts: an image recognition algorithm and experimental verification based on VGG-16. The algorithm part first introduces the overall framework of the basic neural network and expounds the construction method of the convolutional neural network model in the TensorFlow environment. Secondly, the L2 regularization method is used to optimize the original network model to enhance the generalization ability of the algorithm and avoid overfitting. Finally, the parameters of the model in the training process are explained.
In addition to recognition accuracy, the real-time algorithm is often considered in the process of image recognition, so this section sets up a real-time validation experiment. When training a Convolutional Neural Network (CNN), a large number of image samples need to be provided as input data in order for the network to capture more details and features of the image.

The model calculation process will also call a large number of parameters, and thus will consume a lot of time, but the training time can be counted as a pre-preparation time. In this study, only data related to the time spent in the prediction process were calculated and recorded, and specific values and details of these data can be found in Table 2. The experimental results show that this paper’s algorithms in the guarantee of accurate recognition of the images of each scene under the premise of the algorithm also maintains high real-time performance.

The experimental part is divided into three parts: CNN model test, recognition algorithm test, and UAV experiment. First, different model training experiments were designed for the NUC dataset, NordL dataset, and Garden Point dataset, and compared from the loss curve and the accuracy curve, respectively. It converges faster and is more stable. Secondly, to verify the effectiveness of the image recognition algorithm in this paper, the accuracy of each algorithm is tested by using the test set images. The experiments show that the model can achieve better recognition accuracy in complex environments. The time consumption of the image test set in the model prediction process verifies the real-time performance of the algorithm. Finally, the algorithm was verified by flight experiments using the training set and test set collected on the spot by the drone, and the recognition accuracy reached 98.08%.

Table 2. Length of time sent on different test sets of the prediction process

<table>
<thead>
<tr>
<th></th>
<th>NUC Test Set</th>
<th>NordLand Test</th>
<th>Garden Point Test Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of test set samples (sheets)</td>
<td>3760</td>
<td>4000</td>
<td>1150</td>
</tr>
<tr>
<td>Sample size (pixels)</td>
<td>256*256</td>
<td>224*224</td>
<td>960*540</td>
</tr>
<tr>
<td>Average elapsed time (s/sheet)</td>
<td>0.02454</td>
<td>0.02332</td>
<td>0.05410</td>
</tr>
</tbody>
</table>
CONCLUSION

Based on the effective identification of target images in the unmanned motion platform and the completion of the research goal of position error correction, this paper studies the high-precision scene image recognition algorithm. This paper conducts a series of innovative research on image recognition algorithms. To verify its reliability and superiority, the typical LeNet5 model, AlexNet model, and Seq SLAM algorithm are used for comparison and verification of real data. The performance is remarkable. Among them, the recognition accuracy rates on the NUC dataset, Nord Land dataset, and Garden Point dataset are 95.3%, 90.24%, and 98.27%, respectively. In addition, the real image training set and test set are collected during the flight by the drone. The experimental verification results show that the recognition accuracy rate is 98.08%.

The method proposed in this article can accurately complete high-precision image recognition in complex environments. However, the practical application environment of scene recognition algorithms is very complex. The network model still needs to use more dense real data for self-learning and validation testing. The algorithm proposed in this paper has only completed the simulation experiments and hardware transplantation and has not been applied to the actual position error correction system. How to integrate the high-precision image recognition algorithm into the intelligent navigation to complete the error correction is the content for the future and should be the main research. This paper focuses on image processing of visual communication systems based on convolutional neural networks. It can be used as a basic technology in various fields of society. But the tests we have done on the types of images are not enough, and we can also expand the training set. Secondly, the research on the recognition of complex features is not enough, and it is still worth further experiments.

Through this paper, practical knowledge will be gained on how to utilize convolutional neural networks in the field of visual communication to improve image quality and transmission efficiency. An opportunity to understand the principles of deep learning and its application in communication systems will be provided. In addition, the image processing methods and technical principles presented in this paper will expand research ideas, inspire innovative thinking, and provide useful guidance for the implementation of practical engineering projects.

In addition, the application environment of the scene recognition algorithms is very complex, and the network models still need to be self-learning and validation tested using denser real data.

This paper first introduces the basic structure and working mechanism of the convolutional neural network and systematically explains the VGGNet model, which lays the foundation for the subsequent work. A solution is given for the image recognition scheme based on convolutional neural network in terms of hardware. Improved CNN algorithm based on saliency detection, most importantly, the improved algorithm for image recognition is explained. The basic principle of the algorithm is explained. A comparison with other saliency detection algorithms is made to verify the superiority of the proposed algorithm in this paper.

The future research direction of this model can be to apply the image processing method of convolutional neural networks to image communication and processing in the military field. The research can focus on improving the covertness, target detection and recognition capability, image quality recovery, transmission stability, etc. of military image communication, providing innovative solutions for military intelligence analysis, reconnaissance, and communication, and enhancing the effectiveness and security of military communication.

DATA AVAILABILITY

The figures used to support the findings of this study are included in the article.
CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

FUNDING STATEMENT

This work was not supported by any funds.

ACKNOWLEDGMENT

The authors would like to show sincere thanks to those techniques who have contributed to this research.
REFERENCES


PMID:35669654


