ISSN: 2414-1895 DOI: 10.6919/ICJE.202309_9(9).0010

An Improved ResNet101 Network based Pavement Classification Method

Tingting Zhang*, Shuyan Ren, Hailong Duan, and Shuoyi Wen

Tianjin University of Technology and Education, School of Automation and Electrical ngineering, Tianjin 300222, China

*2652990100@qq.com

Abstract

To address the problems of the current traditional pavement classification and recognition algorithms such as long time consumption, low accuracy and poor feature extraction ability of the original ResNet101 network, an image classification method based on the improved ResNet101 network model is proposed. Based on the original ResNet101 network, the final average pooling layer and classification layer are deleted, and the model is used as the backbone network, and global average pooling layer, Dropout regularization and batch regularization are added after the backbone network to reduce the model overfitting. The experimental results show that the method achieves 95% accuracy for pavement image recognition, which is 5.7% more accurate than the original ResNet101 model and 24% more accurate than the support vector machine classifier, which is more effective in improving the model recognition accuracy and has certain guiding significance.

Keywords

Pavement Classification Recognition; ResNet101; Backbone Network; Regularization; Support Vector Machine.

1. Introduction

The application value of road classification recognition is great, especially to be able to analyze the road surface is there is no obstacle, but also to be able to analyze the road that can be chosen. In recent years, the scope of application of road surface intelligent robots has been expanding, and they can be applied to be used for minesweeping and obstacle breaking, armed patrol, nuclear, biological and chemical detection, hazardous materials transportation, fire guidance, communication relay and rear-load protection, etc. They are necessary equipment for the future change of the Army's combat mode to non-contact, non-linear, asymmetric and zero-casualty. As the application of pavement intelligent robots in non-structural environments continues to increase, especially the requirement for complex ground perception capabilities also increases. The environment perception system is one of the cores of the autonomous navigation control system of a pavement intelligent robot, and the pavement classification recognition is an important part of the perception system. The purpose of pavement classification recognition is to automatically distinguish different pavement types such as grass and mud using cameras or vision sensors. However, it is very complicated to separate the different pavement types because the field of view angle has far and near and the camera quality is not standard. In the study of pavement type classification, researchers have proposed various solutions using different vision sensors.In 1997, Lorigo proposed to achieve functions such as obstacle crossing by multiple sensors including monocular and stereo camera synergy, using brightness, RGB and HSV feature extraction algorithms based on luminance, RGB and HSV feature extraction algorithms

ISSN: 2414-1895 DOI: 10.6919/ICJE.202309_9(9).0010

independently of each other and pattern recognition, and the research related to classification and recognition of ground types came into being[1] Blas et al. used LBP [2]feature extraction and K-mean clustering algorithm for fast online segmentation to identify ground types and passable paths, and achieved prediction of new pavement types, but the shortcomings are that it requires a high position relationship between the robot and the identified pavement and is prone to multi-classification leading to identification failure.[3] In an attempt to solve the problem of robustness of recognition due to environmental changes, Broten used an adaptive learning approach, which distinguishes between different types of pavements by learning from test samples.Jean-Frabcois uses a 3D point cloud statistics-based classification method with LIDAR and monocular cameras for real-time pavement recognition.[4] Vernaza et al. overturned the traditional algorithmic idea of pixel-by-pixel computation and learning and proposed a new algorithm for real-time ground type recognition using a Markov random field framework, showing that the new algorithm is on average 10% more accurate than the traditional algorithm[5]. In 2010, T. Y. Kim used a wavelet feature extraction method and selected a neural network machine learning method to train the classifier with an average accuracy of 80% .[6].

This paper proposes an improved ResNet101 network-based pavement classification method, which reduces model overfitting and improves model recognition accuracy by removing the final average pooling layer and classification layer on the basis of the traditional ResNet101 network, setting the model as the backbone network, and adding techniques such as global average pooling layer and regularization after the backbone network.

2. Traditional Machine Learning Algorithms

2.1 Feature Extraction and Processing Methods

The image signal is the basis and prerequisite for pavement classification[7] and recognition, so the features need to be extracted and processed.[8] extraction and processing. Texture features can represent the texture direction of the target unit, and can distinguish different objects by the difference of fine texture. There are many methods to extract texture features, and in this paper, LBP texture features are selected to distinguish different pavement types.

Since the extracted LBP features are two-dimensional features, and the subsequent classification requires one-dimensional features, the LBP features need to be expanded by rows to form new one-dimensional features.

2.2 KNN Algorithm

After the features are extracted and processed, the images are then classified by the classification algorithm. Among the traditional machine learning algorithms, KNN classification algorithm is the easiest to understand and it is one of the simplest algorithms in data mining classification techniques. It is shown in equation (1).

$$d(X,Y)=\sum_{i=1}^{n}\left\|x_i-y_i\right\|^2 \tag{1}$$

The basic steps of the algorithm are.

Step1: Calculate the distance between the test set feature vector and the training set feature vector.

Step2:Sort by distance.

Step3:Select the nearest K points.

Step4:Calculate the occurrence frequency of the category in which the first K points are located.

Step5:Count the categories with the highest frequency in the first K points as the classification categories of the test set.

Volume 9 Issue 9, 2023
DOI: 10.6919/ICJE.202309_9(9).0010

2.3 Support Vector Machines

Support vector machine (SVM) implements nonlinear decisions in the original space by constructing hyperplanes in a high-dimensional space. Its basic idea is to solve the separation hyperplane that can correctly partition the training data set and has the largest geometric interval. The schematic diagram of the SVM algorithm is shown in Fig.1, where $w \cdot x + b = 0$ is the separation hyperplane[9] There are infinitely many such hyperplanes for linearly divisible data sets (i.e., perceptron), but the geometrically spaced separation hyperplane is the only one.

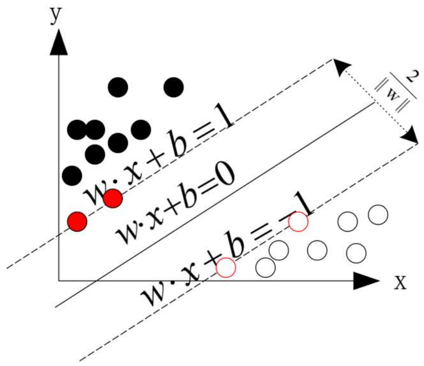


Fig. 1 Principle diagram of SVM algorithm

The algorithm steps are.

Step1: Turn the multiclassification problem into a convex optimization problem.

Find the optimal solution for $\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{R}\zeta_i$, where C is the penalty factor and the default value of 1.0 is used.

Step2: Define the Lagrangian objective function. Where 1 is the penalty factor.

$$L(w,b,a,\zeta) = \frac{1}{2}\|w\|^2 \tag{2}$$

$$-\sum_{i=1}^{n} a_i (w \cdot x_i + b) - 1 + \zeta_i \tag{3}$$

Step3: Find the optimal solution.

Step4: Get the optimal classification function.

3. Deep Learning

3.1 Support Vector Machines

The literature finds that among the classification algorithms, convolutional neural network (CNN) has better autonomous learning ability and robustness.[10] CNN is composed of an input layer, a convolutional layer, a pooling layer, a fully connected layer, and an output layer. By increasing the number of convolutional layers, more complex features can be extracted. Convolutional neural networks perform better than other deep neural network architectures due to their unique way of processing. instead of processing one pixel at a time, CNNs combine several pixels together, so they can find temporal patterns. In recent years, it has become a trend to use convolutional neural networks to process and analyze data, and many classical network models such as VGGNet, ResNet, DenseNet and GooleLeNet have been proposed and widely used in computer vision and natural language processing.[11]They are widely used in computer vision and natural language processing.

3.2 ResNet101 Network

In the process of convolutional neural network development and research, the depth of convolutional neural network layers has been increasingly required due to the increasing number of classification problems and recognition difficulties. 2015 saw the birth of the residual network-ResNet[12] The residual network, ResNet, which was born in 2015, uses the residual module for training and establishes an effective connection between input and output, so that the neural network can be widened in depth[13]The problem of gradient disappearance or gradient explosion caused by the deepening of layers is cleverly solved by using the residual module for training, which establishes an effective connection between input and output, enabling the neural network to maintain its ability of feature expression while widening its depth. The introduction of the residual module is a crucial part of the development process of convolutional neural networks, and the structure of this module is shown in Fig. 2.

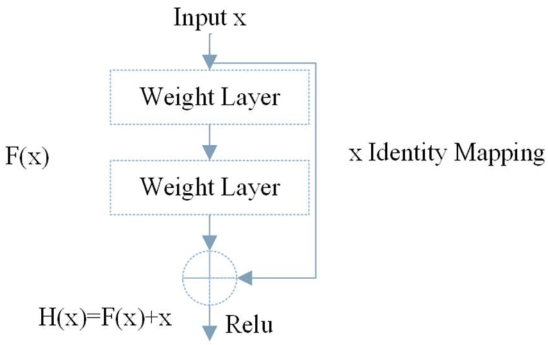


Fig. 2 Residual structure of ResNet network

ResNet101 is based on the structure of VGG network, and the residual learning block is added to it based on the short-circuit mechanism. The network has up to 101 layers, and the residual block consists of three convolutional layers of sizes 1×1, 3×3 and 1×1, as shown in Fig. 3,which are sequentially concatenated and merged with the input, and the Rule activation function is added after each layer.

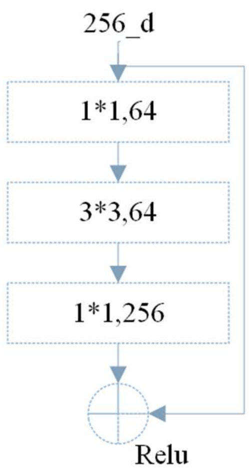


Fig. 3 3-layer residual learning module

Volume 9 Issue 9, 2023
DOI: 10.6919/ICJE.202309_9(9).0010

The ResNet101 network construction is shown in Fig.4.

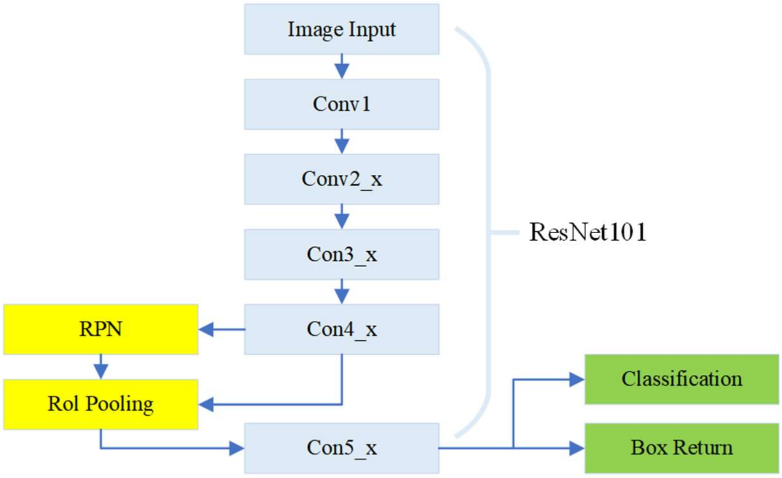


Fig. 4 ResNet101 network structure

4. Network Model Construction and Optimization

4.1 Transfer Learning

Migration learning is a method proposed to solve the problem of overfitting of datasets during the training process of neural network learning. By saving the feature parameters pre-trained in a large network and then applying them to a completely new task, the efficiency and accuracy of the data classification problem is improved by exploiting the portability of the feature model weights between different classification data. Since the training is done on top of the pre-trained model, the training time is greatly reduced and the results are generally satisfactory.

Based on the advantages of migration learning on sample datasets, this paper uses the ResNet101 network as the base architecture, sets the new sample parameters as those already trained on the ImageNet dataset, and the migration learning method is used to train the pavement image dataset.

4.2 Model Fine-tuning

The main improvements are as follows:

1) First, the migration model is fine-tuned to remove the final average pooling and classification layers, and the model is used as the backbone network.

2) Set the global average pooling layer After ResNet101 backbone network, the network parameters are reduced to avoid overfitting. After that, two fully connected layers are added and the number of neurons are both set to 512.

3) activation function selection Relu. activation function is an extremely important feature of artificial neural networks. It determines whether a neuron should be activated, and the activation represents the information received by the neuron with respect to the given information. A neural network without an activation function is essentially just a linear regression model. The activation function applies a nonlinear transformation to the input, allowing it to learn and perform more complex tasks.

4) Add Dropout regularization after the activation function processing by setting the value of the dropout probability to 0.5, i.e., discarding half of the parameters as they are trained in the network, in order to reduce the impact of parameter changes on the output results.

5) Build BN layer to Batch Normalization technique The BN layer algorithm is shown in Eqs. (4) and (5), which represent the regularization and scale transformation and translation of the data, respectively. In Eqs.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202309_9(9).0010

$$\hat{\alpha}_i = \frac{\alpha_i - \mu_B}{\sqrt{{\sigma_\beta}^2 + \varepsilon}} \tag{4}$$

$$b_i = \gamma \cdot \hat{\alpha}_i + \varepsilon \tag{5}$$

μ_B indicates the mean value of the batch data; $\sigma^2{}_\beta$ indicates the deviation of the batch data.

6) Add the classification layer Soft-max with the number of output neurons as 4, so that the output meets the requirements of the data set in this paper.

5. Experimental Results and Analysis

5.1 Pavement Image Dataset and its Pre-processing

A perfect dataset is indispensable for classification recognition. At present, there is no unified classification dataset with sufficient richness for pavement type classification, so in order to better accomplish the classification task, a pavement image dataset needs to be established in advance. The dataset contains grass pavement, mud pavement, gravel pavement and snow pavement, with a total of 1600 images. Some of the pavement images are shown in Fig.5.

a. Grass road b. Dirt road

c. Stone pavement d. Snow pavement

Fig. 5 Images of the pavement in part of the dataset

Due to the relatively small size of this dataset, the model generalization ability is poor, and it is easy to overfit when directly fed into the network operation, which seriously affects the accuracy of image classification recognition. Therefore, in order to make the model training better, data set expansion is needed, and data set expansion is also called data augmentation. In this paper, the original 1600 images are expanded to 7540 images by first exaggerating and randomly disordering the training set samples using data enhancement methods such as pan, flip and shear transform.

The experiment divides the dataset into two parts, the training set and the test set, in a ratio of 9:1, where the number of training set is 6786 and the number of test set is 754.

This experiment is based on Python scripting language, Tensorflow 2.0.0 deep learning framework and pycharm platform, with test set accuracy as the model evaluation metric.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202309_9(9).0010

5.2 Effect of Improvement Strategy on Experimental Results

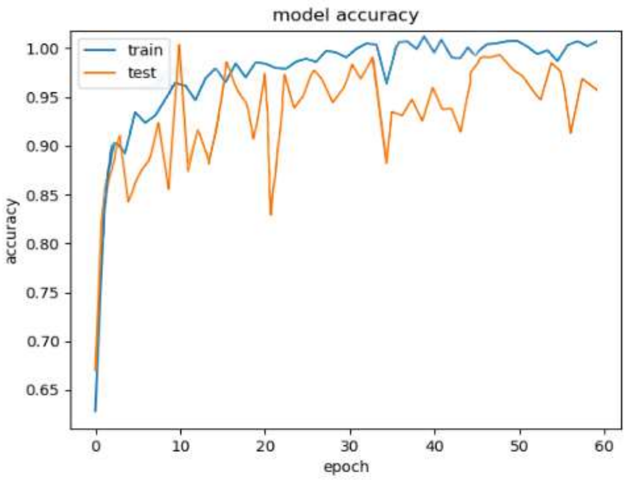
To verify the effectiveness of the improved model proposed in this paper on the classification and recognition problem of pavement dataset, the improved model is compared with the KNN and SVM in traditional machine learning and the original model of ResNet101, respectively, with the optimizer chosen as Adam and the loss function as cross entropy function. After all iterations of 60 times, the maximum value of test accuracy is used as the basis for judgment. Compared with the improved model in this paper, the final obtained model accuracy and loss comparison is shown in Table 1.

Table 1. Comparison of training results of fine-tuned models

Models	Test accuracy	Test Losses
KNN	0.69	-
SVM	0.71	-
ResNet101	0.893(epoch=60)	0.17
ResNet101 Improved Model	0.95(epoch=60)	0.08

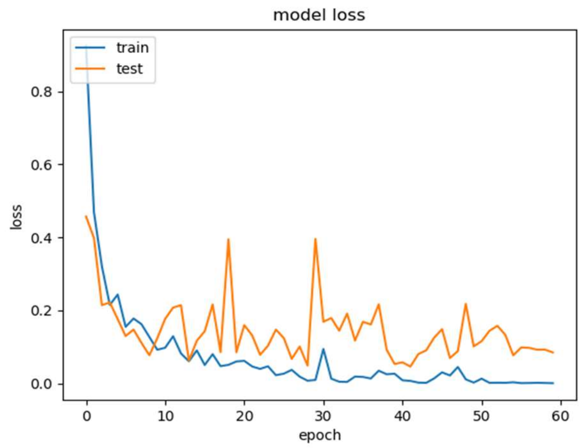
As can be seen from Table 1, deep learning outperforms traditional machine learning for classification. In the traditional machine learning module, the test accuracy of SVM classifier is higher than that of KNN classifier. In deep learning, the improved model of ResNet101 has improved the test accuracy by 5.7% and reduced the test loss by 0.09 compared with the original ResNet101 model, which proves that the improved method in this paper can alleviate the problem of model overfitting to a certain extent and is feasible in the pavement classification problem. The accuracy and loss curves of the model training process before and after the improvement are shown in Fig. 6 and Fig. 7.

It can be seen that the original ResNet101 network performs better on the training set, and the recognition accuracy increases steadily with the number of iterations, finally reaching an accuracy of nearly 96%, but the accuracy of the test set fluctuates in a wide range, which shows that the model has produced overfitting problems. The improved network model was able to stabilize the test accuracy at 95% after about 40 iterations of training, which was better than the original model in terms of generalization ability.



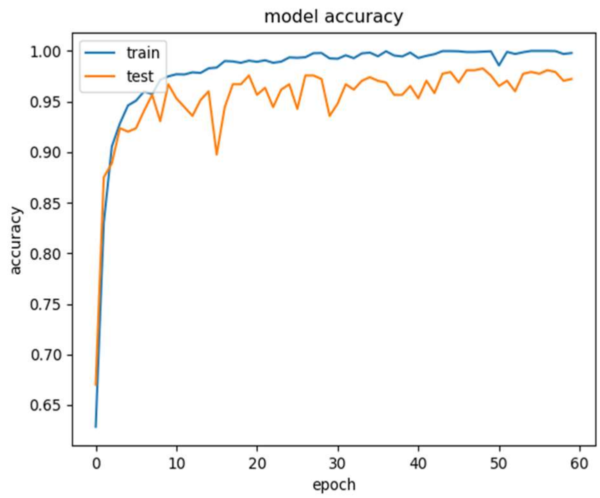
a. Accuracy of training set and test set before improvement

Volume 9 Issue 9, 2023
DOI: 10.6919/ICJE.202309_9(9).0010

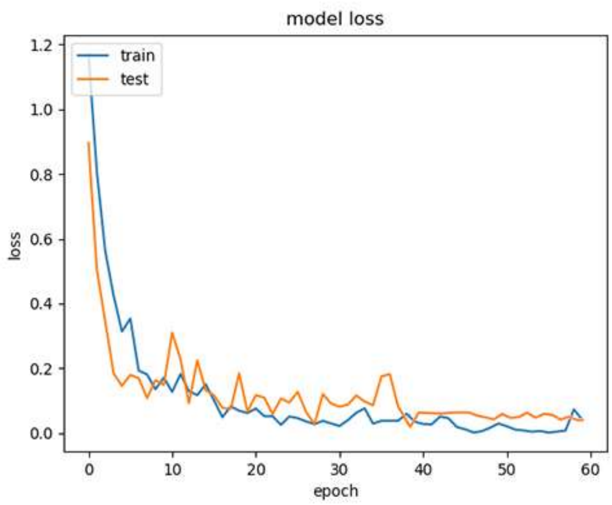


b. Loss of training set and test set before improvement

Fig. 6 Accuracy and loss curve of the model before improvement



a. Accuracy of the improved training set and test set



b. Improved training set and test set loss

Fig. 7 Accuracy and loss curve of the model before improvement

5.3 Effect of Different Optimizers and Learning Rates on Experimental Results

Many optimization methods can solve the problem of optimal model solution in deep learning, and the choice of different optimizers will also make the training appear different effects. If the learning rate is chosen too small during training, the model convergence will be too slow; while choosing too large will lead to curve oscillation, which will easily skip the optimal value.

In order to verify the effects of different optimizers and learning rates on the classification and recognition results of the pavement dataset, the optimizers SGDM, Adam and RMSprop were chosen to conduct experiments on the improved model, and the initial learning rates were set to 0.01, 0.001 and 0.000 1, respectively, and the test results obtained are shown in Table 2.

Table 2. Comparison of recognition accuracy with different optimizers and learning rates

Learning Rate	SGDM	Adam	RMSprop
0.01	0.931	0.942	0.929
0.001	0.941	0.95	0.931
0.0001	0.907	0.942	0.938

As can be seen in Table 2, with Adam as the optimizer and 0.001 as the learning rate parameter, the model was able to converge quickly and tested with the highest accuracy.

6. Conclusion

In this paper, by analyzing the requirements of a pavement intelligent robot for complex ground perception capability, the pavement data set is divided into four categories: grass pavement, muddy pavement, gravel pavement and snow pavement, and a pavement image classification method based on ResNet101 network is proposed. It is pointed out that for different pavements, the robot can develop the optimal walking method including the optimal walking speed and the maximum turning speed by autonomous trajectory planning. The original ResNet101 network is improved by removing the final average pooling layer and classification layer, and fine-tuning its model structure parameters, introducing regularization techniques to improve the model generalization ability, and selecting the Adam optimizer to accelerate the convergence speed of the model to make it show better performance. The improved ResNet101 network model improves the accuracy by 24% compared with the SVM classification method in traditional machine learning and 5.7% compared with the original ResNet101 network model, and the model generalization ability becomes better, which is feasible for solving the classification and recognition problem of road images.

References

[1] Lorigo L M, Brooks R A, Grimsou W E L. Visually- guided obstacle avoidance in unstructured environments[C]//Intelligent Robots and Systems, 1997. IROS'97., Proceedings of the 1997 IEEE/RSJ International Conference on. IEEE, 1997,1: 373-379.

[2] Blas M R, Agrawal M, Sundaresan A, et al. Fast color/texture segmentation for outdoor robots[C]// Intelligent Robots and Systems, 2008. iros 2008. IEEE/ RSJ International Conference on. RSJ International Conference on. IEEE, 2008: 4078-4085.

[3] G.S. Broten, B.L. Digney. A learning system approach for terrain perception using Eigen images[C]. Defense R&D. Canada. 2008.

[4] Strom J,Richardson A,Olson E.Graph- based segmentation for colored 3d laser point clouds[C]// Intelligent Robots and Systems (IROS),2010 IEEE/RSJ International Conference on. IEEE, 2010: 2131-2136.

[5] Vernaza P,Taskar B,Lee D D.Online, self- supervised terrain classification via discriminatively trained submodular Markov random fields[C]// Robotics and Automation, 2008.ICRA 2008. IEEE International Conference on.IEEE, 2008: 2750-2757.

[6] Kim T Y,Sung G Y, Lyou J. Robust terrain classification by introducing environmental sensors[C]//Safety Security and Rescue Robotics (SSRR), 2010 IEEE International Workshop on.IEEE, 2010: 1-6.

[7] Hu Xiaojuan. Research on text classification method based on feature selection[D]. Jilin University,2018.

[8] Bian Zhaoqi, Zhang Xegong. Pattern Recognition (2nd ed.) [M]. Beijing:Tsinghua University Press,2000.

[9] Wu, Zhengwen. Research on the application of convolutional neural network in image classification [D]. University of Electronic Science and Technology, 2015.

[10] He L, Zheng Zaoxian, Xiang Fengtao, et al. Research progress of text classification technology based on deep learning [J]. Computer Engineering, 2021, 47(2) : 1-11.

[11] HE K M, ZHANG X Y, REN S Q, et al. Deep Residual Learning for Image Recogniti [C]// IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas:IEEE, 2016: 770-778.

[12] SELESNICK I W. Wavelet transform with tunalble Q-factor[J]. IEEE Transactions on Signal Processing, 2011,59(8): 3560-3575.

[13] PAN S J, QIANG Y. A Survey on Transfer Learning [J].IEEE Transactions on Knowledge and Data Engineering, 2010, 22(10) : 1345-1359.