Problem Chosen	2021 MCM/ICM Summary Sheet	Team Control Number
C		2102554

Collision of Inspiration and Experience: A System to Lock Vespa Mandarinia

Summary

The acceleration of globalization brings convenience to human beings, but it also brings a major threat species invasion. In Washington state, a wasp called Vespa mandarinia is invading as an alien species. If it is not controlled, it will cause serious damage to the local ecological environment. However, how to design an efficient monitoring mechanism to detect this organism with limited resources is an urgent problem. Based on this, according to the data and information provided in the material, we designed the relevant mathematical model, and tested and analyzed our model.

For question 1, we did not directly predict the time series of the location. We consider the necessary conditions for the survival of Vespa mandarinia. By introducing three intermediate variables of temperature, precipitation and altitude, we first use the **Spatial-Time Series Interpolation (STSI) Model** to fill in the data of the intermediate variables, and then use our **Probability Density Model** to find out the time and place where the target is most likely to appear. This mechanism has a certain theoretical basis by means of geography, biological knowledge and probability theory.

For problem 2, we think that we can use deep learning method to complete the task of image classification on a given image data set. According to the literature, using **Convolutional Neural Network (CNN)** training model is an effective method for image classification, so we choose two novel and effective architectures: Se-RESNET and WS-DAN. After comparing the results, we finally choose to use **WS-DAN** which is based on attention mechanism for image classification. After the training of the model, the indicators show that the model is effective.

For problem 3, we combine problem 1 and problem 2 to design a comprehensive model combining model 1 and model 2. The comprehensive evaluation report can help the relevant personnel to find the target more efficiently.

For question 4, we discussed "when the model needs to be updated", "how to update" and "how often to update". In addition, we also discussed the influence of population reproduction and the distance from the activity site to the beehive in the updating method, and proposed an idea to improve the model. Similarly, for the model trained by WS-DAN neural network, when more positive samples are collected, we need to retrain to ensure the accuracy.

Finally, for question 5, according to our model, we think that during the breeding season (Autumn), if only a few Vespa mandarinia are found in high-risk areas and there are no Vespa mandarinia in low-risk areas, then Vespa mandarinia can be considered to have been controlled.

At the end of the paper, we wrote a memorandum to the U.S. Department of Agriculture.

Keywords:STSI, Possibility Density Model, Convolutional Neural Network, WS-DAN

Contents

1 Introduction

1.1 Problem Background

With the continuous development of human society, globalization has brought many advantages, but also brought about the destruction of the ecological environment: the phenomenon of species invasion. Since invasive species are not constrained by natural enemies and local organisms generally have no resistance to invasive species, invasive species often seriously endanger the diversity of local organisms, leading to further degradation of ecosystems, loss of ecosystem functions and services.And some exotic species can directly harm human health.Therefore, human intervention in invasive species is important and necessary.Therefore, a set of systems that can effectively monitor, screen and predict the trend of invasive species is very important to deal with invasive species.

Vespa mandarinia is the largest species of hornet in the world, and the occurrence of the nest was alarming. A small number of Vespa Mandarinia are capable of destroying a whole colony of European honeybees in a short time. It's a powerful predator of local bees and it may cause great agricultural loss. So building up models to predict their movement and control them is necessary.

1.2 Restatement of Problems

As Vespa Mandarinia, an endemic bumblebee in Asia, has been found in Washington State for many times, given the local hazards of the invasive species, the Washington state government has set up helplines and websites and called for Vespa Mandarinia to be reported. Each uploaded report contains a unique label (GlobalID), the date of the witness, the official judgment of the lab, the opinion of the laboratory, the date of submission, the longitude and latitude of the witness and the bee pictures taken by the witness.Analyze the submitted Vespa Mandarinia data that has been witnessed to determine "how to handle a large number of report data" and "With limited human resources in government agencies, how to filter large amounts of data submitted so that genuine Vespa Mandarinia reports are as close as possible to missing and further investigate based on the screened reports".

To achieve our goals, we need to:

- Judge whether the spread of Vespa mandarini can be predicted and its accuracy based on the existing data
- Using the determined data (image, location), create a model to judge whether the future report is a positive report, and analyze and discuss the model.
- Discuss the significance of the judgment of the model to screen out the true occurrence of Vespa mandarinia
- How to update the model and update frequency after getting more reports over time
- Based on the constructed model, what will be the evidence that "Vespa mandarinia has been eliminated"

We consider "limited government resources" as limited human, material and financial resources. That requires us to design a effective model or strategy to predict effectively. And in our model we will use probability or probability density to describe pest spread prediction.

1.3 Analysis and Our Work

1. For Problem 1, the datasets provide only 14 cases with the latitude, longitude and time. So this is a typical Spatial-Temporal Series Interpolation problem. But straightly using the position and time to predict the occurrence of Vespa Mandarinia is not suitable. According to biological knowledge, a lovely living space with fine temperature, water and trees is suitable for insects living. And height infects the number of trees, which is necessary for Vespa Mandarinia to live. Based on this, we establish a **STSI(Spatial-Temporal Series Interpolation) model** to evaluate temperature and precipitation data. Then, we test if tempreture, precipitation and height has normality. If yes, we can use normal distribute and established a **Possibility-Density model** to evaluate the possibility of appearance for one place.

2. For Problem 2, the datasets has a lot of pictures and Problem 2 requires us to build up a classifier. With rapid development of deep learning techniques, we can use **Convolutional Neural Network** to do image-classification. We compared two CNN architectures(SE-ResNet and **WS-DAN**) and chose WS-DAN as our model of Problem 2. Because the number of positive samples is far less than negative, we also used preprocessing techniques like data augmentation to make the classifier performs better.

3. For Problem 3, we combined our model in Problem 1 and Problem 2, and established a prediction mechanism. Using strategy we designed, we could process cases that have high probability to make our work effective.

4. For Problem 4, we considered the effection of the population growth of Vespa Mandarinia, the number of wasps discovered per time and distance's affect on their activity intensity. Then we discussed if number of positive samples becomes larger, how our model should be renewed and how long for our model to renew.

5. For Problem 5, we discussed in what cases we can asure that the Vespa Mandarinia has been controlled.

2 Preparation of the Models

2.1 Assumptions

To simplify our model and eliminate the complexity, we make the following main assumptions in this literature. All assumptions will be re-emphasized once they are used in the construction of our model:

1. We assume that in Washington, temperature and precipitation do not change abruptly with geographic location, that is, temperature and precipitation change continuously with geographic location, which is the basis of our problem modeling.

2. We assume that only a small number of Vespa Mandarinia enter Washington at the beginning, otherwise there will be many Vespa Mandarinia groups and the model will be too complex.

3. We assume that if the predicted data can pass the normality test, we can fit the data with a normal distribution.This makes our probability density model more concise and persuasive.

2.2 Notations

The symbols in the table are for reference only. If there are symbols not marked in the table, the meaning shall be subject to the text.

Table 1: Notations

Symbol	Definition
T	The Average Temperature of a Day
W	The Average precipitation of a Day
H	The Height
R	The Radium of the earth
$latA$	The Latitude of A
$longA$	The Longitude of A
$d(A, B)$	The Distance between A and B
x	x=[T,W,H]
$f(x)$	The Probability Density Function of x
$F1$	F1-score of our neural network
Acc	Accuracy of our neural network
Pre	Precision
Rec	Recall
AAA	Average Application Accuracy
$N(t)$	The Number of Vespa Mandarinia at time t
$A(d, N)$	The Activity Intensity of Distance d and Population N
p	The Probablity Density
λ	The Increase Rate of Vespa Mandarinia
w	Weight
a	A Constant
k	A Constant
K	A Threshold we designed

3 Model Establishing and Solving

3.1 Problem 1: Prediction of Vespa Mandarinia.

Because Vespa Mandarinia is a rare occurrence, Therefore, we believe that a general model cannot accurately process such a small amount of data. And also, it is not enough to consider only the time and place factors. We should consider why they occur at these times and places.

We think that for any insect, a suitable living environment is required. So after referring to the University of Pennsylvania literature describing Vespa Mandarinia's behavior, we used temperature, humidity, and height as intermediate variables in the prediction model. That is, In our model, temperature, precipitation, and height change further with location and time (independent variables).

Vespa Mandarinia's colony is usually underneath trees, so the presence of forests can also greatly affect Vespa Mandarinia's existence. At the same time, the height of the elevation affects the distribution of forests, such as the Rocky Mountains, where there are many trees, and Vespa Mandarinia is likely to nest in the Rocky Mountains. But where it occurs is still related to temperature, precipitation, and altitude.

To make problem easier, we use normality test to check if these attributes has normality, if yes, we could use normal distribution to fit data. For the normal distribution, the closer to the expectation, the greater the probability density function value, and the greater the probability value in the vicinity. So,from probability density value we can judge the probability of occurrence. If we insist on using probability instead of probability density, our computing complexity will be exceedingly large because we need to compute interval for each variable.

So, we first set up a temperature and precipitation prediction model, and then build a probability density model based on the above analysis. Finally, we draw a probability-density-map to see where Vespa Mandarinia is most likely to appear.

3.1.1 Spatial-Temporal Series Interpolation(STSI) The location of Vespa Mandarinia did not have accurate temperature and water details, so we performed a data search.We took monthly averages from five sites in Washington, D.C., and then expanded the data in both spatial and temporal dimensions using space-time series interpolation.

Table 2: Source of Data

City	Source
Olympia	http://www.worldweather.cn/zh/city.html?cityId=2019
Seattle	http://www.worldweather.cn/zh/city.html?cityId=277
Yakima	http://www.worldweather.cn/zh/city.html?cityId=846
Tacoma	http://www.worldweather.cn/zh/city.html?cityId=2040
Spokane	http://www.worldweather.cn/zh/city.html?cityId=833

Time-part In the time dimension, we use cubic spline interpolation method to make up the daily average precipitation and temperature of each station. We have collected the monthly average temperature and precipitation.

Cubic spline interpolation requires each function $S_i(x)$ between two data points is a cubic function:

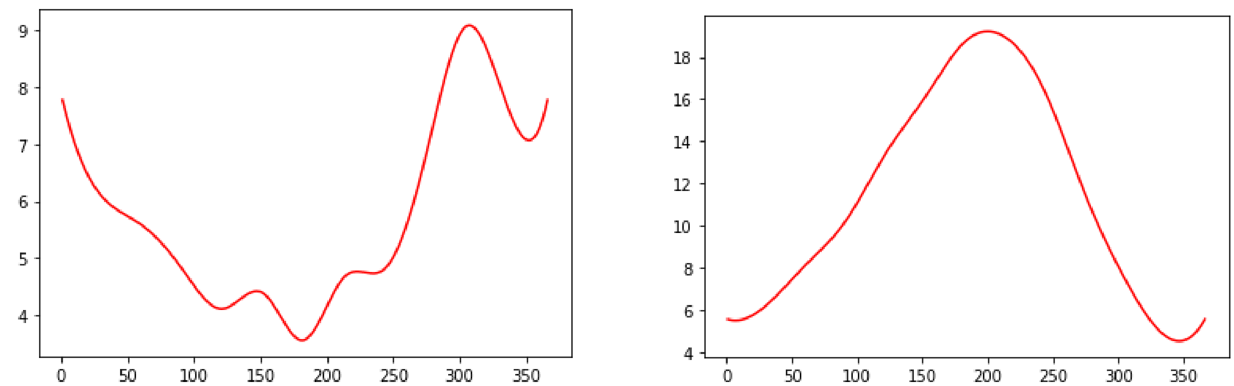
$$S_i(x) = a_i x^3 + b_i x^2 + c_i x + d_i \tag{1}$$

And it should follows:

$$\begin{cases} S_i(x_{i+1}) = S_{i+1}(x_{i+1}) \\ S_i'(x_{i+1}) = S_{i+1}'(x_{i+1}) \\ S_i''(x_{i+1}) = S_{i+1}''(x_{i+1}) \end{cases} \tag{2}$$

The advantage of cubic spline interpolation is that the predicted function is continuous, and the derivative is also continuous. Compared with linear interpolation, cubic spline interpolation is more accurate. We have collected data for the last 30 years locally, and after data processing, we can generally assume that the data is cyclical, with a one-year cycle. But, in fact, there will be some deviation between the data of the current year and the average data collected, which will also lead to some deviation in the data of the forecast date. We will discuss the error after our STSI-model established.

Temperature and precipitation interpolations for the Seattle Station are shown in Figure 1. The horizontal coordinate represents date, The vertical coordinate represents temperature or precipitation.



(a) precipitation interpolation. (b) temperature interpolation.

Figure 1: Interpolation of Temperature and Precipitation of Seattle

Space-part We choose Inverse Distance Weight method (IDW) to evaluate any point's temperature and precipitation[1]. According to the first law of geography, the principle of similar climate at close points. For one city's temperature T and precipitation W, we can get IDW equations below:

$$\begin{aligned} T &= \frac{\sum_{i=1}^{5} T_i/d_i^2}{\sum_{i=1}^{5} 1/d_i^2} \\ W &= \frac{\sum_{i=1}^{5} W_i/d_i^2}{\sum_{i=1}^{5} 1/d_i^2} \end{aligned} \tag{3}$$

There d_i means the distance from this city to the i-th station. We took meteorological data from five stations so that we could theoretically assess temperature and precipitation at each target location. The problem, however, is how to calculate the distance between two locations given their latitude and longitude. We think of the earth as a sine sphere, and according to the spherical cosine formula, we have the following equation:

$$C = \sin(latA)\sin(latB) + \cos(latA)\cos(latB)\cos(longA - longB) \tag{4}$$

There $longA$, $latA$, $longB$, $latB$ mean longitude of A, latitude of A, longitude of B, latitude of B. Then we have:

$$d(A, B) = \frac{\pi}{180} R \arccos C \tag{5}$$

So we can compute the temperature and precipitation at the appearing time and in the area.

Height we can collect height data on Google map and we can analyze it. And we don't need to predict height data.

3.1.2 Probability Density Model For T, W, H, we do Kolmogorov-Smirnov test on them, which is a common method for normality test. If p-value>0.05, we think they follows normality. The result is below:

Table 3: normality test

attribute	p-value
T	0.40945787548446855
W	0.7682164859783146
H	0.6192154398651544

Fortunately, they all follow normal distribution. This means we can use normal distribution to fit data. And considering their relevance and covariance matrix Σ, we can get distribution function:

$$\begin{gathered} f(x) = \frac{1}{\sqrt{8\pi^3|\Sigma|}} e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)} \\ x = [T, W, H], \mu = [\bar{T}, \bar{W}, \bar{H}] \end{gathered} \tag{6}$$

Usually we use probability more than probability density to evaluate. If we know that the measurement errors of temperature, precipitation and height are ΔT, ΔW and ΔH respectively, then our probability is equal to a triple integral in region Ω:

$$\begin{gathered} \Omega = \begin{cases} T - \Delta T < t < T + \Delta T \\ W - \Delta W < w < W + \Delta W \\ H - \Delta H < h < H + \Delta H \end{cases} \\ P(T, W, H) = \iiint_\Omega f(t, w, h) dtdwdh \end{gathered} \tag{7}$$

As we can see, our computational complexity increases a lot with the calculation of the triple integral. However, we know that if the error of each variable is constant, the closer to the center of the sample, the greater the probability density and the greater the probability value.So we just need to calculate the probability density. We will use function (6) as our probability density model.

3.1.3 Prediction Our prediction is based on our model and unverified data:

Based on model Using K-Means to analysis Vespa Mandarinia's colonies' positions, and we get the cluster result in Figure 2. We think the cluster centroids are very likely to be the location of the honeycomb.

Studies have shown that bees do not move 30 kilometers farther than their nests.And the last four occurrences are very close, let's assume that the centroid of the last four samples is the honeycomb. At sample centroid C, the latitude for 30km is $\frac{30}{R}$, and the longitude for 30km is $\frac{30}{(R\cos(latA))}$, Therefore, we identified the survey area, started collecting data from the area and predicted temperature and precipitation. We chose to predict September 1 because September and October were the most active

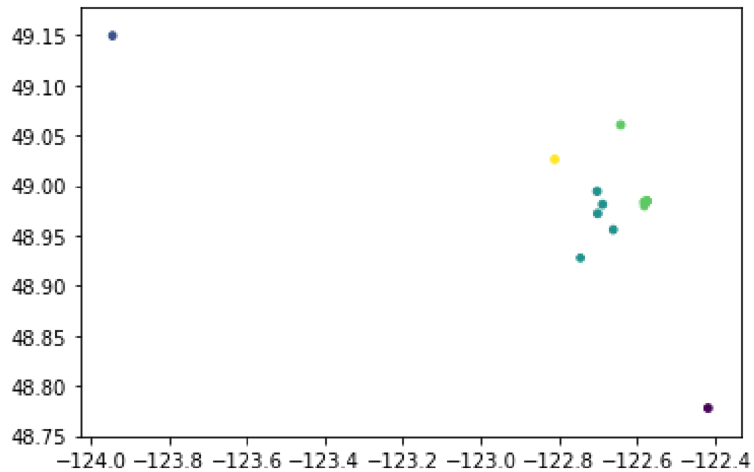
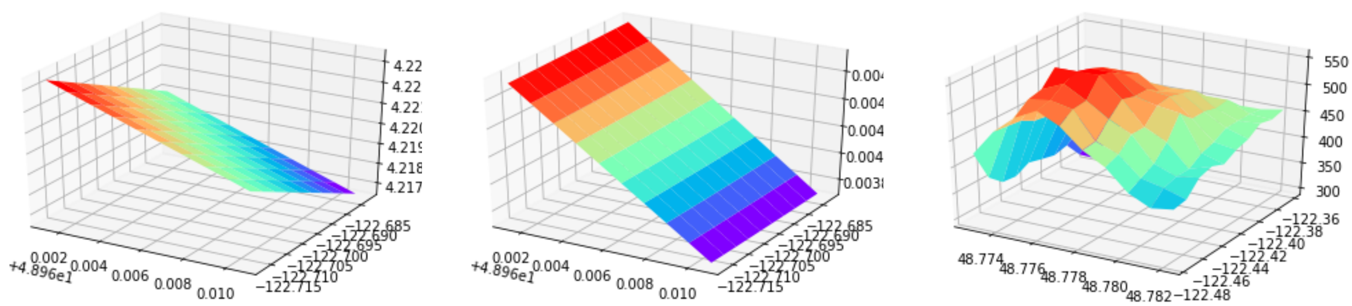


Figure 2: K-means

periods for the wasps, during which their numbers increased dramatically. We've mapped the predicted temperature and precipitation data for this area, and we've also mapped the height data for this area that we've collected on Google Maps, as shown in Figure 3. Then we calculate the probability density and get Figure 4.



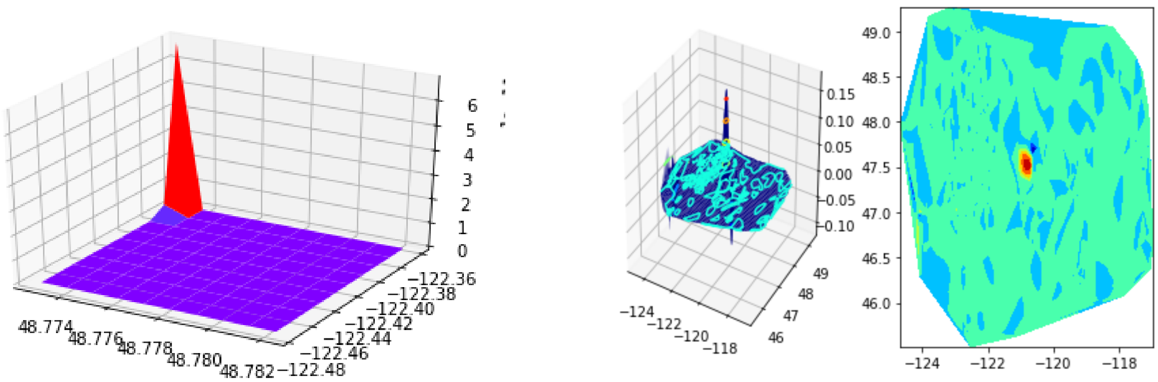
(a) precipitation. (b) temperature. (c) height.

Figure 3: predict temperature, precipitation and collected height data of target area

The higher probability density is, the more likely for Vespa Mandarinia to appear. From Figure 4,it's obvious that, at (-122.40788,48.77282), Vespa Mandarinia is most likely to appear.

Then we will search around this point. If we find Vespa mandarinia or its nest, then our model prediction is accurate. Next, we can use this model to update the prediction data after the data information is updated.

Based on unvarified data For the undetermined data in the provided datasets, we collected when and where they occurred. Because the exact target elevation is difficult to obtain, we use $\bar{H}$to calculate it.Similarly, we draw a picture to describe the possibilities: In Figure 5, we can clearly see that there is a very high possibility of about one point. In this contour map, color represents the value of probability density.Obviously, for Figure 5, there is a small area with a high probability density (the red



(a) Based on Model Prediction (b) Based on Unvarified Data

Figure 4: Probability Density Value.

part of figure a). After searching, the highest point appeared in (-122.494965, 48.890159) ,2020-9-25. Based on the probability density values, 100 unconfirmed samples are predicted as possible.

3.1.4 Error Analysis If more and more positive samples appear, especially in areas with low probability, we should update the model.Since our prediction of air temperature and precipitation is based on the assumption that one year is a data cycle, and Vespa Mandarinia's honeycomb changes once a year, the valid prediction time of the model should be less than one year.

In addition, the measurement errors of temperature, precipitation, and elevation also affect the prediction errors. By estimating the distances between each station and the other four stations and comparing them with the observed values, the average errors of temperature and precipitation are 6.354653453216351 and 3.7448414907407754, respectively. The errors of T and W are mainly in STSI part instead of probability-density part.

3.2 Problem 2: Image Classification

At the same time, we think it is important and necessary to introduce a picture classification model because the website set up by government agencies has a mechanism for submitting pictures and pictures are very intuitive and accurate as one of the criteria for judging.

In recent years, with the application of deep learning technology, especially convolutional neural network (CNN) in image recognition, the computer can quickly and accurately complete a series of image classification tasks [2]. Therefore, we think that using convolutional neural network for image-classification is advantageous. There are many classical architectures widely used in image classification such as LeNet-5[3] and AlexNet[4]. However, due to the limitation of the number of positive samples in the dataset, we decided to use more complex network to solve problems. The training results show that the neural network we choose is quite competent for this task.

In section 3.2.1, we analyze the problems and objectives of this image recognition. In section 3.2.2, we compare two networks (SE-resnt, WS-DAN) that were initially selected, and finally select WS-DAN network. In section 3.2.3, we propose evaluation criteria based on the characteristics of this picture dataset and test the model WSDAN-VM (WSDAN-Vespa Mandarinia)built by the neural network with the new evaluation criteria. In section 3.2.4, we used the model to classify the Unverified pictures in the dataset in preparation for the next comprehensive analysis.

3.2.1 Analysis

Analysis of Data We classify all the provided image information into four categories according to the original positive ID, negative ID, unverified and unprocessed tags. We think that as a two category (yes or no Vespa mania) task, only positive ID and negative ID are used The two image sets of ID are reasonable, and the information (unverified and unprocessed) that is not clear for the time being should not be considered. Therefore, the image data we get have the following characteristics:

- Small amount of data: only 3032 pictures have been proved to be suitable for use
- There is a serious imbalance in the amount of information between the two categories (positive:negative is nearly 1:232)
- The requirements of fine-grained are high:

  (1) there are great differences within the class. The pictures of the same species of bees in different positions and perspectives are often very different.

  (2) Low inter class variance. Bees belonging to different classes may be very similar except for some slight differences. For example, the only difference is the color of bees, the shape of abdominal stripes, the distance between eyes and so on.

  (3) Limited training data.

Analysis of Model The classification results we expect have the following characteristics:

- high accuracy;
- because our ultimate goal is to find the true Vespa mandarinia as much as possible, if we predict positive as negative, we think the error is fatal (low acceptance);
- if we predict negative as positive, we think the error is acceptable (high acceptance).

In Figure 5, we can see the difference between a bee and other bees.Both images come from the dataset provided. They differ very slightly in shape and are very similar in size, color, and shape, making it difficult to classify them.

We think that the conventional convolutional neural network is not enough to support the above requirements, so we need to find a reasonable model to solve the above problems. After literature search and comparison, we think that the introduction of "attention mechanism" in the conventional CNN is reasonable and useful. After comparison, we choose the "SE-ResNet" and "WS-DAN" models. Both of them are excellent convolutional neural network models in recent years, and they are far ahead in terms of test accuracy. Therefore, we expect that the use of these two models will bring great help to our classification task.

3.2.2 Architecture

(a) possitive. (b) negative.

Figure 5: Vespa Mandarinia or not.

SE-ResNet Convolution operator is the core building block of convolutional neural networks (CNNs), which enables the network to construct information features by fusing the spatial and channel information in the local receptive field of each layer.[5] A large number of previous studies have investigated the spatial components of this relationship, trying to enhance the representativeness of CNN by improving the spatial coding quality of the whole feature level. A new architecture unit, which is called "squeeze and excite" (SE) block, can adaptively recalibrates the channel feature response by explicitly modeling the interdependencies between channels. These blocks can be stacked together to form a senet architecture, which can be effectively promoted on different datasets. And the SE block can significantly improve the performance of the most advanced CNN with a slight increase in computing cost[6].

SE-ResNet is based on ResNet, which introduced Shortcut Connections and Residual Learning[7]. This model's greatest contribution is adding SE-blocks to make network performs better.

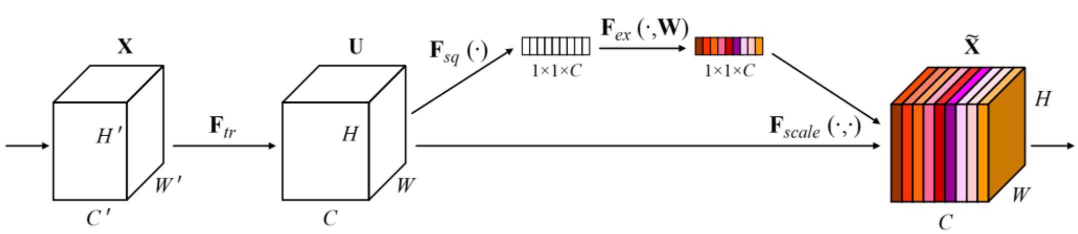


Figure 6: Architecture of SE-ResNet

For data characteristic 1 (small amount): it has not been solved theoretically.

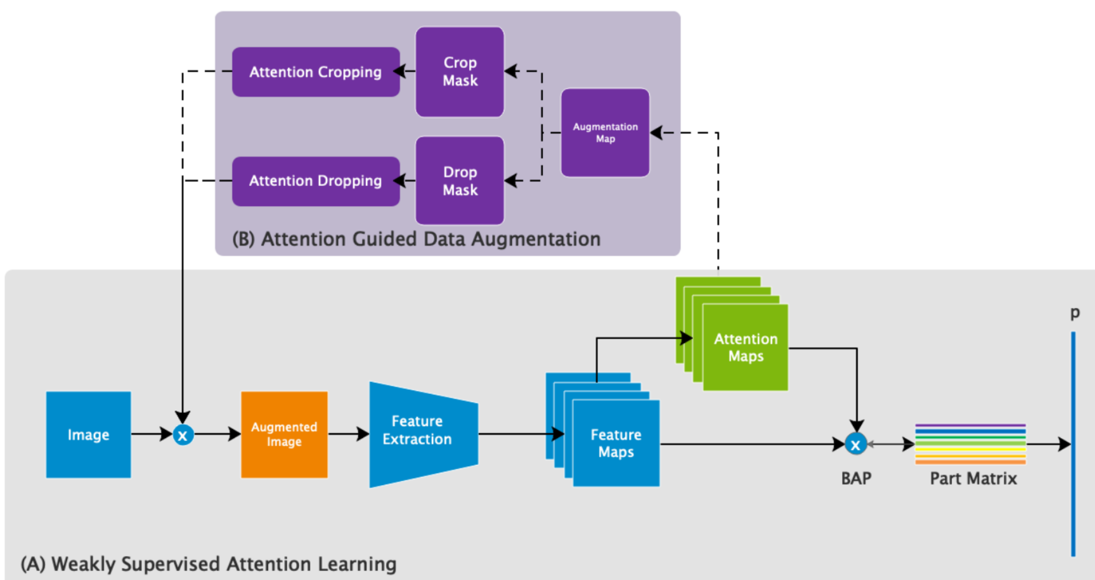
For data feature 2 (imbalance): this model has not been effectively solved in principle, so we use the method of artificial data enhancement to translate, zoom and rotate the original 13 images (because the identification of bees needs color characteristics, we think that the "change color" of data enhancement method is unreasonable) and get 208 images. To a certain extent, it solves the problem of "extremely unbalanced data". After consulting the literature, data enhancement does not make the model over fitting, on the contrary, data enhancement is the way to avoid over fitting.

For data feature 3 (fine-grained): the model significantly improves the performance by adaptively recalibrating the channel feature response. We think the problem has been solved in theory.

We use 80% data for training and 20% data for evaluation. The training set does not intersect the test set. Batch size is 32, SGD is used as optimizer, learning rate is initialized to 1e-3, LR scheduler uses steplr equal interval gradient descent method to make LR 0.9 times of the original every 5 epoch, so as to better improve the function fitting ability. Due to time constraints, we trained a total of 20 epochs. In the 20th epoch, the accuracy of the final test set reaches 99.84%. Therefore, we think that this model solves part of the problems caused by the original data.

WS-DAN WS-DAN introduced attention mechanism. For each training image, firstly generate an attention map through weak supervised learning to represent the discriminating part of the object. Next, enhance these attention map guided images, including attention clipping and attention dropping. The proposed WS-DAN improves the classification accuracy by two times. Comprehensive experiments on ordinary fine-grained visual classification datasets show that WS-DAN outperforms the most advanced methods, which proves its effectiveness[8].

In the architecture of WS-DAN (as shown in Figure 7), we can see that the training process and testing process of this network are different, and the main difference lies in the attention map. During the training process, after augmentation map processed data, it will be divided into two parts(Crop Mask and Drop Mask), and combine them in training period. And when WS-DAN is testing, it just use object map and bounding box for object localization and refinement.

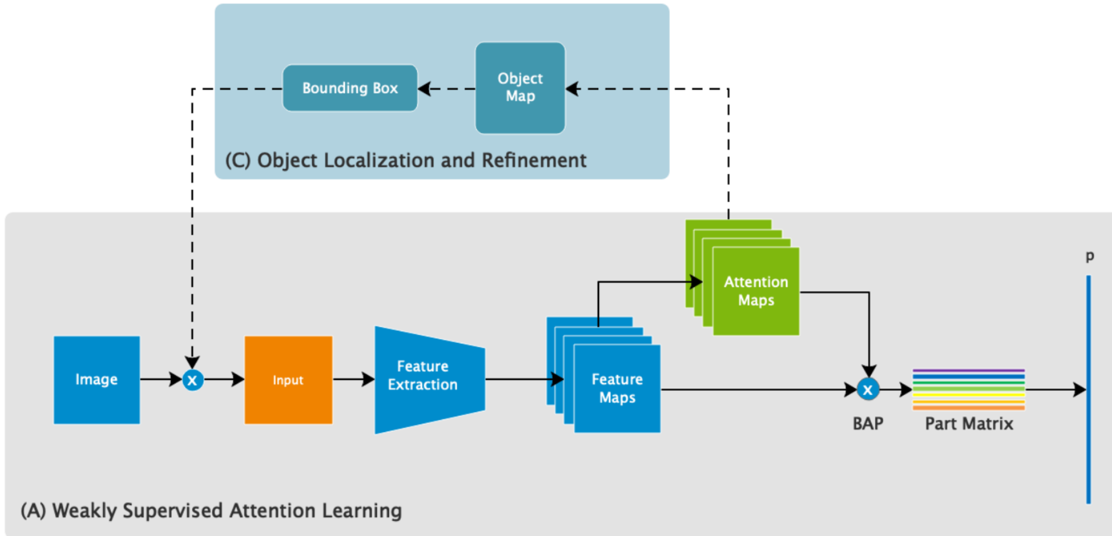


(a) training process.

For feature 1 (small amount of data): this model has the potential of data expansion, we think that this problem has been solved.

For problem 2 (imbalance): this model has the potential of data expansion.

For problem 3 (fine-grained): by referring to literature[4], we know that the comprehensive experiments on ordinary fine-grained visual classification data sets show that ws-dan is superior to the



(b) testing process.

Figure 7: Overview training and testing process of WSDAN.

most advanced methods, which proves its effectiveness.

For fair comparison, our training and testing parameters are the same as Se RESNET model. In the 20th epoch, the accuracy of the final test set reaches 99.92% Therefore, we think that this model solves the problem brought by the original data.

3.2.3 Evaluate of Model In theory, WS-DAN is more suitable for the original image data set, and the accuracy is higher. In addition, due to the time factor, we only choose ws-dan with better performance for further evaluation. We realized that in the positive:negative In the case of positive:negative = 1:232, even if all the pictures in the test set are predicted to be negative, the accuracy of prediction will reach 99.57%. Based on this situation, we decided to increase F1 score, precision and recall values as more convincing measures. Related definition is listed in equation (8)-(11)

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

$$Pre = \frac{TP}{TP + FP} \tag{9}$$

$$Rec = \frac{TP}{TP + FN} \tag{10}$$

$$F1 = \frac{2}{\frac{1}{Rec} + \frac{1}{Pre}} \tag{11}$$

From the meaning of the formula, we can see that precision represents the degree of predicting negative as positive, and we think that the error is acceptable. Recall indicates the degree to which positive is predicted to be negative, and we believe that mistakes are fatal (low acceptance). That is, we are most concerned about whether recall tends to be 1.

3.2.4 Running results

training period During our training period, we recorded related data and show a part of it below:

epoch	accuracy	F1	Pre	Rec
1	0.9951	0.0000	0.0000	0.0000
20	0.9992	0.9231	0.8571	1.0000
100	1.0000	1.0000	1.0000	1.0000

Table 4: Index in Training Period

It can be seen that recall gradually approaches to 100% with the increase of epoch.

Result: We recorded some indexes in the training process and drew them into the curve shown in Figure 8. In Figure 8 we can see that, although the F1 scores, precision and recall are relatively low at the beginning of a period of time, after a period of oscillation, they can achieve nearly 100% effect. This is a good result of the experiment.

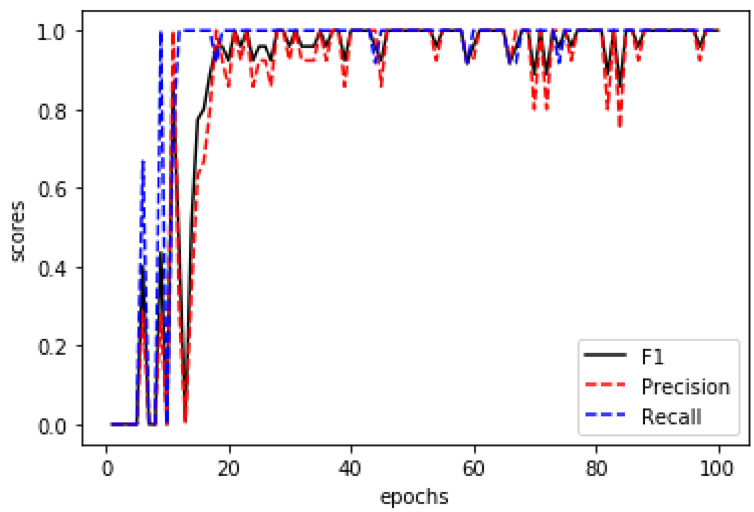


Figure 8: Architecture of WS-DAN

Therefore, we have reason to think that this model can solve the problems caused by the original data with excellent test results.

On the Test Dataset We selected the Unverified pictures from the picture set and evaluated them using the model that was successfully trained above. We use floating point numbers in the [0,1] interval as the evaluation result.The closer the result is to 1, the more likely it is to be positive, and the closer the result is to 0, the more likely it is to be negative.

3.3 Problem 3: A Mechanism to Predict

We present two models in Problem 1 and Problem 2, respectively. We think that combining them is an effective way to solve the problems 3. Due to the small amount of data, all known data has been used to construct the model.Data used should not be treated as test data. So we put forward the following proposals:

Process:

1. First, It needs n unprocessed data to be judged
2. Then, Comparing the spatial and temporal information and picture information of these verified data with our classification, we can get the application accuracy of Model 1($Acc_{1.i}$) and Model 2($Acc_{2.i}$).
3. Use the equation(12) to calculate the average application accuracy of Model 1 and Model 2(AAA_1;AAA_2), and assign the calculation weight(w_1;w_2) by the average accuracy of the two models (the weight size is proportional to the accuracy).

$$AAA_1 = \frac{\sum_{i=1}^{n} Acc_{1.i}}{n}; AAA_2 = \frac{\sum_{i=1}^{n} Acc_{2.i}}{n} \tag{12}$$

4. Target report information(L_{pred}) is processed using Model 1 and Model 2 by equation(13), and weighted averaging is done after two possibilities are obtained to get weighted probability.

$$L_{pred} = w_1 \times AAA_1 + w_2 \times AAA_2 \tag{13}$$

5. Reports with the highest weighted likelihood(AAA_1;AAA_2) should be the first to notice

3.4 Problem 4: Update of Model

3.4.1 Under what circumstances should we update our model? We think in these situations the model should be updated:

1. In the case of few positive samples in the initial stage, as long as we collect positive samples, we should update the model. When the positive samples reached 50 cases, every 5 positive samples should be updated; when the positive samples reached 100 cases, every 10 positive samples should be updated
2. Several confirmed cases show that there burst many Vespa Mandarinia.
3. The next burst places are far from the previous cases.
4. Someone witnessed a Vespa mandarinia hive. This means that Vespa mandarinia is acting very positively around this location.
5. If we haven't clear the colonies in one year, the Vespa Mandarinia will breed very quickly and the number of investigation position will increase since now there are only few groups of wasps.
6. Now we consider that all the Vespa Mandarinias are from the same family. But after a long period they might breed into several families, so at this time model should be updated.

3.4.2 How to update our model? if there bursts more Vespa Mandarinia, the centroid of wasps should be updated. And with more appearance data collected, the average temperature and precipitation should be renewed to get a new probability density function. The best expectation is that we can collect detailed data.

If there appears a branch of wasps, we should take the acting strength into consideration. For example, at 2019/9/19, someone witnessed a colony in which there are 150 live Vespa Mandarinia and 600 eggs. So when we consider the number of Vespa Mandarinia, we consider their number and establish two changes.

From distribution's aspect,Each Vespa mandarinia is predicted to appear, it must have the corresponding temperature data and precipitation data records. For example, if predict data is $t = [1, 2, 3, 4, 5]$, and at the place which correspond to number 2 there occurs 5 wasps. Our data will be changed to $t = [1, 2, 2, 2, 2, 2, 3, 4, 5]$. And for new data, we recheck if the data satisfies the normal distribution or not. If still satisfying, we compute new average temperature, new average precipitation and compute a new distribution function.

From colony positions' aspect, if we have witnessed a branch of Vespa Mandarinia, when we clustering the occurrence data, we use weighted summary instead of simple K-means. So that the strength of activation can be described and infect the cluster centroids' position, making them more accurate.

We haven't take breeding of Vespa Mandarinias into consideration. If take it into consideration, in each breeding season, the active strength is higher which is positively relevant to the number of wasps. If there are many branches of Vespa Mandarinia, for one place the possibility of occurrence is a addition of several branches. Because of the lack of natural enemies, suppose that the number of Vespa Mandarinia follows:

$$N(t) = N(0)(1 + \lambda)^t \tag{14}$$

This is a typical Exponential growth. And then these Vespa Mandarinia will be divided into several branches to find their own living place.

We consider the effect of activity intensity on the frequency of occurrence. However, the activity intensity was related to the number of Vespa Mandarinia and the distance between them and their nests. Generally speaking, the closer to the hive, the greater the intensity of activity. We assume that the intensity of activity is inversely proportional to the distance from the hive and directly proportional to the number of colonies. The function of activity intensity can be expressed as equation (15)

$$A(d, N) = a\frac{N(t)}{d} \tag{15}$$

where a is a constant.

We assume that there are n branches, and taking the distance into consideration, for example, we can use exponential distribution to describe the probability of occurrence of Vespa Mandarinia.(Of course, it may also be other forms of distribution, and the specific type depends on the data)

$$p(d, N) = k \exp{(-\frac{A(d, N)}{k})} \tag{16}$$

But we still need to evaluate the $N_i(t)$. Assume that, if $N(t) > K$ The killer bees will be divided into two parts as new $N(t)$ So the updated probability density F is related to $d, N, T, W, H.$ and we can write it as:

$$p = F(T, W, H, d, N(t)) \tag{17}$$

and if there are many colonies, we add their p and get P_{sum} then we draw figure. To make integral of P_{sum} equals to 1, we use a weighted summary:

$$P_{sum} = \frac{\sum_{i=1}^{n}(\frac{N_i(t)}{d_i})p_i}{\sum_{i=1}^{n}(\frac{N_i(t)}{d_i})} \tag{18}$$

For the wsdan model, we only need to put the new verified classified images into datasets and train the network.

3.4.3 How long should our model be updated? When there is only a small amount of data at the beginning, we should retrain our classification ws-dan model every day as long as there is a positive image update. (after testing, one night is enough for wsdan network to train the new model)

When the positive samples reached 50 cases, every 5 positive samples should be updated; when the positive samples reached 100 cases, every 10 positive samples should be updated. Without more cases to observe, we can't asure the update time in neumeric evaluation. Even so, we still suggest that the update time should not exceed one year. As mentioned in the modeling of question 1, our estimated temperature data and precipitation data are based on the average of the previous 30 years' experience, and it is assumed that the estimated data have the characteristics of one year cycle. Moreover, this kind of bee will change the nest location once a year, so it is better to re estimate the model within a year.

But in some special cases we should adjust at once. For example, once a colony witnessed, our model should be updated at once because the Vespa Mandarinia acts strongly in this area. After estimating the total population according to the ecological method, if $N(t)$ is more than a threshold K we define previously according to biological knowledge, our model should be updated.

3.5 Problem 5: Controlled it or not?

In Problem 1 we established a model and we can predict the target space and get each place's probability density. And similarly, we can also draw a figure to show where is the most possible for Vespa Mandarinia to appear. In this places, if we collected bees, we can use our trained WS-DAN model in Problem 2 to classify. And we can combine the two methods. For these places, during their most active period(spring and autumn) there are three conditions:

Condition 1:there has no case of Vespa Mandarinia.

Condition 2:there occurred bees, but our classifier in Problem 2 predict that they are not Vespa Mandarinia.

Condition 3:In low-risk area there is no Vespa Mandarinia, and in high-risk area there are only few targets.

By combining the two methods, we can effectively judge if the Vespa Mandarinia has been cleaned out or not. If the high risk area has only small or no increases and low-risk area has no increases we can say that the Vespa Mandarinia has been cleaned up.

4 Senitivity Analysis

In problem 1 we established STSI-model and probability-density model, and in problem 2 we established WSDAN-classifier. So our Senitivity is based on Problem 1 and Problem 2:

4.1 Model of Problem 1. Our senitivity is carried on 4 parameters: $\bar{T}, \bar{W}, \bar{H}, \Sigma$. For 3 average value, take temperature for example, if $\bar{T}$ increases ΔT, our variable T gets further or closer to $\bar{T}$. How far and how close is related to ΔT, but one thing we are sure is that, the closer to $\bar{T}$, the greater probability-density changes. If $\bar{T}$ increased ΔT, for changing rate r we have equations below:

$$r = \frac{|f(T, W, H) - f(T - \Delta T, W, H)|}{f(T, W, H)} \tag{19}$$

For covariance matrix Σ, the maximum of $f(T, W, H)$ is related to $|\Sigma|$. If $|\Sigma|$ increases 10%, our maximum value will decrease 9.09%.

From our analysis above, with changes of four parameters, our probability density value also changes, which means our model is sensitive. However, the magnitude of numerical change is similar to that of parameter change, which indicates that our model has good stability.

4.2 Model of Problem 2. There seems not many parameters for us to carry out senitivity analysis. We just consider the number of epochs. We find that if the number of epochs is more than a threshold K(about 80), the accuracy, F1 score, precision and recall won't increase too much. So similarly, if the epoch is far more than K, a little change on the number of epochs won't cause the performance of model.

But when the epoch is less than K, the performance will be greatly infected. And the more you change epochs, the more changes performance has. Accurate numerical effect need to further experiments. This shows that our neural network is not so sensitive to epoch in some cases, but it has very good stability and robustness.

5 Strengths and Weaknesses

5.1 Strengths

- More convincing. Our model is based on the nature of Vespa mandarinia. Compared with time series analysis, this model considers the related properties of wasps. And we use normal distribution to fit the data, which makes our method more consistent with the statistical law. So our model is more convincing than direct prediction of location and time.

- Higher accuracy. Our neural network architecture is wsdan, which uses attention enhancement mechanism and is highly suitable for ordinary fine-grained visual classification, so its performance is better than other architectures such as se-resnet50. Compared with the traditional CNN structure, it has higher accuracy and robustness, especially when the positive samples are too small.

  Our model combines the two methods to form a set of effective strategies, which helps the staff to effectively detect the submitted reports.

- The application time is long. For our probability density model, we discuss its update time and method. Our model is based on the average data collected in the past 30 years, considering the universality, so we can use our model to predict the data within one year. At the same time, we designed a scientific method to update the model. With the increase of data, the updated model using our method will be more accurate.

- Easy to use. Only time and place image data is needed to complete the assessment. And the model sorts the reports according to the possibility, so that the relevant government personnel can obtain the most likely reports only through simple operation.
- Strong generalization ability.Our model is based on the conditions under which organisms survive and is therefore expected to be suitable for a variety of invasive species.

5.2 Weaknesses

- The results were biased. Our model in question 1 is a rough estimate, and there is a certain deviation in the prediction results.
- It may be over fitting. If more positive data is provided, the model in question 2 will perform better. Without more possible samples, our good results in training are likely to be over fitting.
- Uncertainty. Due to the lack of positive data, all the valid data are used to build the model, and there is no more confirmed data for us to test the whole model, so the actual accuracy needs to be confirmed.

6 Conclusion

For Problem 1 *We established a probability-density model and tested it. According to one place's longitude, latitude and occur time, we can estimate the temperature, precipitation by using our STSI-model and calculate probability-density value. The closer to the average it is, the more possible for Vespa Mandarinia to occur.*

For Problem 2 *We tested two architectures for image-classification named SE-ResNet and WS-DAN, and finally choose WS-DAN. After training, our model performs a nice result.*

For Problem 3 *A good strategy to predict is to combine two models. Use probability-density model to predict first and according to the result we use WS-DAN to predict effectively. On one hand we handle high-risk area previously; on the other hand, we can check our model if it should be updated.*

For Problem 4 *Within a year our model should be updated. And with more and more cases collected our dataset will be larger so we also need to retrain our WS-DAN network. The update method of probability-density model can investigate influence of the population increase and each colonies' distance*

For Problem 5 *By using our strategy in Problem 3, if the high risk area has small increases and low-risk area has no increases we can say that the Vespa Mandarinia has been cleaned up.*

Memo

TO: The Washington State Department of Agriculture
From: Team 2102554
SUBJECT: Methods to prevent the spread of Vespa mandarinia
DATE: February ,8, 2021

We think the invasion of Asian bumblebees in Washington state is predictable.Based on this, we established a comprehensive model to predict the occurrence probability of Vespa mandarinia in the region and the accuracy of residents' reported cases, and put forward the coping strategies. Our work, conclusions and suggestions are as follows:

Our work

First, we use the meteorological data of five weather stations in Washington state to estimate the temperature and precipitation by STSI method. Based on our research on the living habits of Apis mellifera, combined with the previous climate information and altitude information, we model and calculate the probability density function of the occurrence of bumblebees in different places and times.

Secondly, we use the existing hornet image information and wsdan network to train a model that can classify the images of public report. Our model performs well in the training process, which shows that it can effectively predict the authenticity of eyewitness events according to the pictures provided by eyewitnesses.

Thirdly, we combine the two models to make it a comprehensive model, and give a comprehensive evaluation index the probability of real discovery.

Fourth, we propose an effective method to update the model for data increase.

Conclusion

Combined with the information of public report and considering the population growth pattern of Vespa mandarinia, we believe that Vespa mandarinia is still concentrated in northwest Washington state. More specifically, for those with higher probability in the calculation results of probability model, we think that there is a greater possibility of Vespa mandarinia disaster.

Recomendation

The effective way to deal with Vespa mandarinia disaster is to find the nest of bees and destroy it. In order to achieve more significant results, the final model provided by us will give an accurate prediction of the authenticity probability of the witness event. With the help of the model, the relevant personnel will take effective investigation actions to find and destroy the nest of Vespa mandarinia as soon as possible under the condition of limited resources, so as to effectively curb the spread of Vespa mandarinia until the elimination of Vespa mandarinia. The specific action strategies are as follows:

Authenticity detection

The eyewitness events reported by the public need to be dealt with in time, which can alleviate the anxiety of the public to a certain extent. More importantly, it is very important for the control of pests. The solutions are as follows:

For eyewitness events reported by the public, we should combine our probability model and picture recognition model (if pictures are provided), make timely response to eyewitnesses, and organize personnel to detect them (priority should be given to the reports with high real probability).

Due to the limited public resources, we should refer to our probability model and give priority to the detection of areas with high probability of occurrence and the confirmation of real report areas.

Destroy Vespa mandarinia nest immediately after discovery and update information.

Active detection

Due to the uneven population distribution, in sparsely populated areas, people can rarely report the emergence of bees. Therefore, even if there is no witness event in the high probability areas predicted by the model, it is necessary to organize personnel to investigate.

Update information

After the test, the test results should be used to update the model to improve the accuracy of prediction and evaluation.

Through the use of our model, we can efficiently complete the task of report identification, and save the cost of resources. At the same time, the high accuracy and prediction function of the model can help control the invasion of honeybee to a certain extent.

we will be very honored if our model is adopted.

Yours Sincerely:

Team 2102554

References

[1] Fu Jing., Yi Zhen-zhao., Zhao Sha., Comparision of GIS-based Spatial Interpolation on the Distribution of Precipitation.*Rural Water Conservancy and Hydropower in China*, 2021.

[2] Li CE, Chen Haixia (in Chinese), et al. Overview of convolutional neural networks in deep learning algorithms [J]. *Electronic testing*, 2018 (23): 61-62

[3] Yann LeCun.,Leon Bottou., Yoshua Bengio., Patrick Haffener., Gradient-Based Learning Applied to Document Recognition.[J]. *Proceedings of IEEE*, 1998.

[4] Alex Krizhevsky., IIya Sutskever., Geoffery E.Hinton., ImageNet Classification with Deep Convolutional Neural Networks.2012.

[5] Jie, Hu, Li, et al. Squeeze-and-Excitation Networks.[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2019.

[6] Lin Jingdong, Wu Xinyi, Chai Yi, et al. Overview of convolutional neural network structure optimization [J]. *Acta automatica Sinica*, 2020,46 (1): 24-37 DOI:10.16383/j.aas.c180275.

[7] He Kaiming., Zhang Xiangyu., Ren Shaoqing., Sun Jian., Deep Residual Learning for Image Recognition.2015.

[8] Hu, Tao, et al., See Better Before Looking Closer: Weakly Supervised Data Augmentation Network for Fine-Grained Visual Classification. 2019.

Appendices

Appendix1: A small part of python code in Problem 1

Environment: Python 3.8.2 and Pytorch 1.0.2, GPU:NVIDIA GeForce GTX1070 ,CPU: intel i7, OS: Win10.

```
import numpy as np
R=6371.004
def dist(longA,latA,longB,latB):
    C = np.sin(latA)*np.sin(latB) + np.cos(latA)*np.cos(latB)*np.cos(longA-longB)
    Distance = R*np.arccos(C)*np.pi/180
    return Distance
import pandas as pd
df1=pd.read_excel("pos/Seattle.xlsx")
avtemp1=(np.array(df1.lowtemp)+np.array(df1.hightemp))/2
avtemp1=np.append(avtemp1,avtemp1[0])
month=np.array([1,32,60,91,121,152,182,213,244,274,305,335,366])
import matplotlib.pyplot as plt
from scipy.interpolate import interp1d
x=np.arange(1,367,1)
f1=interp1d(month,avtemp1,kind='cubic')
y1=f1(x)
avwater1=np.array(df1['average total'])/np.array(df1['average days'])
avwater1=np.append(avwater1,avwater1[0])
f_1=interp1d(month,avwater1,kind='cubic')
y_1=f_1(x)
lat=np.array(df0.lat)
long=np.array(df0.long)
D=[]
for i in range(14):
    d=[]
    d.append(dist(long[i],lat[i],longS1,latS1))
    d.append(dist(long[i],lat[i],longS2,latS2))
    d.append(dist(long[i],lat[i],longS3,latS3))
    d.append(dist(long[i],lat[i],longS4,latS4))
    d.append(dist(long[i],lat[i],longS5,latS5))
    D.append(d)
D=np.array(D)
w=1/D
W=[]
for i in range(14):
    weight=[]
    for j in range(5):
        weight.append(((w[i][j])**2)/((w[i][0])**2+(w[i][1])**2
        +(w[i][2])**2+(w[i][3])**2+(w[i][4])**2))
    W.append(weight)
W=np.array(W)
```