AN EFFECTIVE ROAD EXTRACTION METHOD from REMOTE SENSING IMAGES BASED on SELF-ADAPTIVE THRESHOLD FUNCTION

Zhuozheng Wang[1, 2], Meng Zhang[1, 2,*] and Wei Liu [1, 2]
[1]Faculty of Information Technology, Beijing, China.
[2]Intelligent Signal Processing Laboratory, Beijing University of Technology, Beijing, China.
E-mail: wangzhuozheng@bjut.edu.cn
E-mail: zhangmeng@emails.bjut.edu.cn

***Abstract*—In the field of remote sensing imagery, road extraction is one of the key technologies supporting for Landuse Landcover classification. In this paper, a new semantic segmentation neural network named SAT U-Net is proposed for road extraction from remote sensing imagery. The new improved network replaces the sigmoid layer in the U-Net with a self-adaptive threshold method proposed to self-adaptively adjust the road thresholds for segmentation results of U-Net. The proposed method is combined with the strength of U-Net architecture to retain the complete road spatial features, thus overcomes the problem of unconnected and blurry roads in the segmentation results. To prove the effectiveness and utility of the proposed network, it was experimented on the test set of a public road dataset and compared with U-Net in five different environments. Experimental results demonstrate that the proposed method is superior to U-Net and presents clearer and more complete road structures.**

I. INTRODUCTION

Remote sensing image technology is a comprehensive emerging technology that provides the basis for human decision-making and planning .It refers that using various sensors in high-altitude and outer space obtains remote sensing data and extracts feature information. As the main land cover references, road is the major part of the modern transportation system. In the urban or rural regions, the general terrain features of the entire area can be obtained by extracting the road information, which is of significant interest for geographical, political, economic and military. Therefore, road extraction from remote sensing images has become a necessary step for many modern applications, such as urban planning[1], vehicle navigation[2], intelligent transportation [2], land use detection [3], military strikes[1], etc. Nowadays, a great deal of remote sensing data have been derived with the rapid development of remote sensing technology. Road is one of the most basic and important geographic information among numerous remote sensing data that people are paying attention to. Therefore, the research on the road information extraction from remote sensing imagery has shown a great potential of becoming a new science and technology.

However, there are various types of interference in actual imaging: The inherent multiplicative noise in Synthetic Aperture Radar (SAR) image causes the edge of the road and the contrast with the surrounding environment to be blurry; The high buildings and trees will form shadows on the road and destroy the continuity of the road lines or regional features, which result in some roads with a single edge or even no edge; The striated interference with a certain gray value formed by the green belt on both sides of the roads results in the misidentified road regions; The roads are adhere to the surrounding buildings in order that some the double edges of the original continuous road is labeled as single edge and the width of different areas are changed greatly on the same road [1]. In addition, the different background of road will result in different degree of complexity for road extraction from remote sensing images. For instance, in the urban regions [4], due to obvious parallel edges of the buildings which are similar to the two sides of the road, it is easy to misjudge the road; in the mountainous [4], the complexity of the terrain results in blurry geometric features of the shape of road. In view of the existence of above interference, the diverse road types and complex environmental background, cumbersome manual interpretation methods will undoubtedly consume excessive manpower and material resources. And the accuracy of extraction cannot be satisfied all the time. Therefore, it has been a hot research topic that how to achieve efficient and automatic road extraction.

The theoretical basis of deep learning is artificial neural network, which retains the essence of neural network, uses multi-layer network to learn abstract concepts and join self-learning, self-feedback, understanding and summarization. One of the most significant specialties is the ability to automatically learn features from large amounts of data without manually extracting features, which is highly compatible with the extraction of road remote sensing information from a large number of remote sensing data. In this paper, a self-adaptive algorithm based on deep learning method is proposed to realize road extraction automatically and effectively from remote sensing images.

The remainder of this paper is organized as follows. Section II reviews some related works of road extraction. The details of SAT U-Net, loss function and evaluation metrics are presented in Section III. Section IV introduces experiments, including dataset, performances and results. Conclusion is provided in Section V. Section VI describes our future work.

II. RELATED WORK

A variety of methods for extracting roads from remote sensing images have been proposed in recent years. Most of these methods can be divided into two categories: road area extraction and road centerline extraction. Road area extraction [5-8] can generate pixel-level labeling of roads, while road centerline extraction [9] aims at detecting skeletons of a road. This article focuses on extracting road regions from remote sensing images. The task of extracting roads from satellite images is formulated as a binary classification problem: marking the road or non-road of each pixel. In this paper, the road extraction task is treated as a binary semantic segmentation task to generate pixel-level labels for the road.

In 2006, reference [10] firstly proposed deep learning based on artificial neural network. Reference [11] firstly exploited deep learning techniques to extract road information. They employed restricted Boltzmann machines (RBM) to segment road from high resolution remote sensing images. Convolution neural network (CNN) [12], as the most widely used deep learning framework, is a deep artificial neural network with convolution kernels which has been successfully applied to remote sensing image classification [13], dense semantic labeling of aerial images [14], and object detection in remote sensing imagery [15]. Reference [8] utilized CNN to assign each patch extracted from the whole image as road, building or background. This method achieves better results than Mnih and Hinton's method [11] on the Massachusetts roads dataset. Fully Convolutional Networks (FCN) [16] is an improvement on convolutional neural networks that replaces the fully connected layer in CNN with a convolutional layer and adds a deconvolution layer for upsampling. Reference [17] proposed a road structure refined convolutional neural network (RSRCNN) approach for road extraction in aerial images based on FCN and road-structure-based loss function, which outperform other deep learning methods. However, the low resolution of center feature map and its corresponding large receptive field result in the loss of detail spatial information in FCN architecture. Reference [18] proposed the U-Net that concatenate feature maps from different levels to improve segmentation accuracy. Low level detail information is combined with high level semantic information to achieve good performance in biomedical image segmentation. The necessary details are included while upsampling with skip-connections and the accurate segmentation boundary is reconstructed, solving the problem of losing too much spatial information. However, the road threshold of each road segmentation result is quite different because of different complexity of environmental background. And they cannot be determined by the output of the sigmoid layer in U-Net causing some blurry road structures and dissatisfying accuracy. Taking consideration of this problem, the improved approach replaces the sigmoid function with a self-adaptive road threshold method proposed in this paper to determine the road threshold of each road segmentation result of U-Net and improves the segmentation results of U-Net on visual and accuracy.

III. METHOD

A. SAT U-Net

The proposed approach is based on an improved FCN architecture, U-Net [17]. Figure 1 shows the overall architecture of the entire SAT U-Net preserving the classic U-Net structure: the architecture consists of a contracting path that captures contextual information (left side in Fig. 1) and a symmetric expanding path that allows precise positioning (right side in Fig. 1). The contracting path follows the typical architecture of CNN, consisting of the repeated application of two 3×3 stride 1 and padding 1 convolutions , each followed by batch normalization (BN) and the rectification linear unit (ReLU). BN is used for convergence acceleration and ReLU is adopted as the primary activation function during training. A 2×2 max pooling with stride 2 for downsampling followed by them. At each downsampling step, the number of feature channels is doubled. The expansive path consists of nearest neighbor upsampling on the feature map, then completing a 3×3 upconvolution that halves the number of feature channels, a concatenation with the correspondingly copied feature map from the contracting path, and two 3×3 stride 1 convolutions, in which each is also followed by BN and ReLU. At the final level, each 64 component feature vector is mapped to the two classes with a 1x1 convolution.

In the U-Net structure, the contracting network is followed by a fully symmetrical extending network where the pooling is replaced by the upsampling, thus increasing the resolution of the output. The high-resolution feature of the contraction path is skip-connected with the upsampled output, leading to the necessary detail features contained while upsampling. Combined the local information with the global information and preserving the complete spatial features of the road generate road segmentation results thus achieve road extraction from remote sensing images. However, road thresholds are quite different because of different distribution of roads and complexity of environmental background. And in U-Net architecture, the road threshold of each prediction map is a stationary value which cannot be controlled by the outputs from sigmoid layer of the U-Net. The excessively large predefined threshold will result in the road disconnection and blurry road structures in each prediction map. And the extremely small predefined threshold will result in mistaking for labeling the background as roads. The accuracy of segmentation will be influenced by these factors. Therefore, the self-adaptive threshold function proposed is used to replace the sigmoid layer. The ground truth's road histogram applied as a guide self-adaptively and dynamically adjusts the road threshold with the histogram absolute distance between each prediction map and its corresponding

ground truth. It is described in detail in the following subsection. According to road thresholds, the segmentation result presents clearer and more complete road structures.

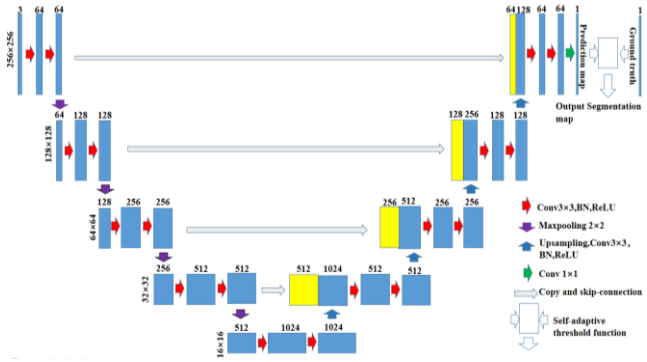

Fig. 1 Architecture of the proposed SAT U-Net.

B. Self-adaptive threshold function

In SAT U-Net architecture, the self-adaptive threshold function proposed in this paper is used to determine the road threshold for each prediction map. The specific implementation process is as follows: Firstly, the roads in each predicted grayscale image are presented according to the initial threshold. R_p and R_g denote the proportion of roads pixel in the prediction map and its corresponding ground truth respectively. And they are defined as:

$$R_p = \frac{n_p}{N} \tag{1}$$

$$R_g = \frac{n_g}{N} \tag{2}$$

where n_p and n_g are the number of pixels which denote the location of roads in prediction map and ground truth respectively, while N denotes the total number of pixels. Then d denotes the pixel histogram absolute distance between the prediction map and ground truth, and it is defined as:

$$d = \left| R_g - R_p \right| \tag{3}$$

the road threshold for each prediction map is adjusted according to d. The specific adjustment rules are as follows: The minimum distance is predefined to $d_{\min}$. If the d is more than $d_{\min}$, threshold will be decayed or enhanced in the following equation:

$$t_{i+1} = \lambda t_i + \xi \qquad n = 0,1,\ldots,i_{\max} \tag{4}$$

where t_i denotes the threshold for decay or enhancement, which λ is defined as the coefficient of decay or enhancement, which ξ is given as bias. The coefficient of decay or enhancement depends on the relationship of R_p and R_g, and defined as :

$$\begin{cases} \lambda > 1, R_p > R_g \\ \lambda < 1, R_p < R_g \end{cases} \tag{5}$$

In the process of threshold adjustment, the change of n_p will also take place with the adjustment of t_i. After multiple threshold adjustments, the d gradually decreases. If the d is less than $d_{\min}$ or the times of adjustments exceed the $i_{\max}$,the threshold adjustment is stopped, and the road threshold of the predicted segmentation map is determined. Finally, the fine segmentation of the road is achieved according to the self-adaptive t_i.

C. Loss function

In the training phase, given a set of training images and the corresponding ground truth as input to the U-Net, the one-dimensional vectors are obtained with a 1×1 convolution. The sigmoid function is adopted to normalize one-dimensional vector to logits between 0 and 1, being given as follow:

$$S(x) = \frac{1}{1+e^{-x}} \tag{6}$$

where x denotes the one-dimensional vector. The task of extracting roads from satellite images is formulated as a binary classification problem: marking the road or non-road of each pixel. Pixel-wise cross entropy is adopted as loss function to optimize the model, being defined as:

$$L = -[y \log \hat{y} + (1-y)\log(1-\hat{y})] \tag{7}$$

where y denotes the ground truth and $\hat{y}$ denotes the prediction map.

D. Evaluation metrics

Pixel accuracy (Pixel Accuracy, PA), Dice coefficient [19, 20] and Jaccard index [21] are used as the standard quality metrics to evaluate proposed model. PA is a common algorithm metric in the field of semantic segmentation. The index is obtained by calculating the confusion matrix which is defined as follows:

$$P = \begin{bmatrix} p_{1,1} & \cdots & \\ \vdots & \ddots & \vdots \\ p_{i,1} & \cdots & \end{bmatrix} \tag{8}$$

where $p_{1,j}$ denotes the number of pixels in the dataset, which should be labeled as i instead of being predicted to be the total number of pixels in class j. n denotes the number of categories. PA denotes the ratio of the number of pixels marked correctly to the total pixels, which is defined as:

$$PA=\frac{\sum_{i=1}^{n} p_{i,i}}{\sum_{j=1}^{n}\sum_{i=1}^{n} p_{i,j}} \tag{9}$$

The Dice coefficient is a statistic used to gauge the similarity of two samples, which is defined as:

$$s(X,Y)=\frac{2|X\cap}{|X|+|Y|} \tag{10}$$

where X and Y denote two samples respectively.

The Jaccard index, also known as the Jaccard similarity coefficient is a statistic used for comparing the similarity and diversity of sample sets, which is defined as:

$$J(X,Y)=\frac{|X\cap}{|X\cup}=\frac{\cap}{\cap} \tag{11}$$

where X and Y denote two samples respectively.

IV. EXPERIMENT

To demonstrate the accuracy and efficiency of the proposed SAT U-Net, it was test on Massachusetts roads dataset and compared with U-Net.

A. Dataset

The Massachusetts Roads Dataset consists of 1171 aerial images of the state of Massachusetts. The dataset includes a wide variety of urban, suburban, and rural regions and covers an area of over 2600 km^2. As with the building data, each image is 1500×1500 pixels in size with a resolution of 1.2 meter per pixel, covering an area of 3.24 km^2. In the experiment, the images with blank part in the original dataset were removed, and the rest were randomly divided into the training set (532 images), the validation set (152 images) and the test set (76 images) according to the ratio of 7:2:1. The U-Net on the training set of the dataset were trained and the SAT U-Net proposed in the test set were used to self-adaptively adjust the road threshold of prediction map, thus improve the road segmentation results.

B. Experiment details

In this experiment, the pytorch deep learning framework was used and all models were trained on two NIVIDA GTX1080ti GPUs. In the training phase, the U-Net model was trained on all of 532 labeled images with a size of 1500×1500.To minimize overheads and maximize the GPU memory, the batch sizes of the training and validation data-sets were condensed into a single image and fixed-sized images were utilized to 256×256 as the input to the network model. In addition, in order to avoid over fitting, data augmentation was done in an ambitious way, including image rotation (random angle), center cropping, image shifting, brightness adjustment, color adjustment, contrast adjustment, vertical and horizontal flip. The learning rate was originally set 2e-4 and the decay iteration was set to 68.There were in total 350 iterations for training the model. When the number of iteration exceeds 282 epochs, the learning rate was reduced by a factor of 0.02 in every epoch. Since the task is binary classification, BCE (binary cross entropy) loss was used as the loss function and Adam was chose as the optimizer. It took about 316 epochs for the proposed network to converge.

After every training iteration, the validation set was input into our proposed SAT U-Net to select the best model and save it which is the highest score of the sum of Dice coefficient and Jaccard similarity. The road segmentation threshold was empirically initialized to 0.5. The decay and enhancement coefficient was set to 0.3 and 1.7 respectively. The bias was set to 1e-10.To obtain more accurate road segmentation thresholds, the histogram distance minimum was set to 1e-3 and the upper limit of times was set to 10. In the test phases, the test set was input into the best model and the Pixel accuracy, Dice coefficient and Jaccard index were adopted as the standard quality metrics to evaluate the road segmentation results.

C. Result

Fig. 2 illustrates two example results of U-Net [17] and the proposed SAT U-Net. After observation reveal that results are clearer and sharper in the proposed method as compared to U-Net. It is clear from the red window of the first two flows that the roads are clearer in the result of SAT U-Net while the roads are not successfully labeled in the result of U-Net inversely. Meanwhile, there are some cases where the background is labeled as roads in the U-Net. For example, it is evident from the blue window of the last two flows that some dirt roads incorrectly is labeled in the result of U-Net, which is not labeled as road in the ground truth. The primary reason for this may be a similarity between the roads and background. By contrast, the proposed method is based on the histogram absolute distance between prediction map and ground truth to ensure the road threshold so that these roads are less labeled in the SAT U-Net. And the segmentation accuracy is further improved.

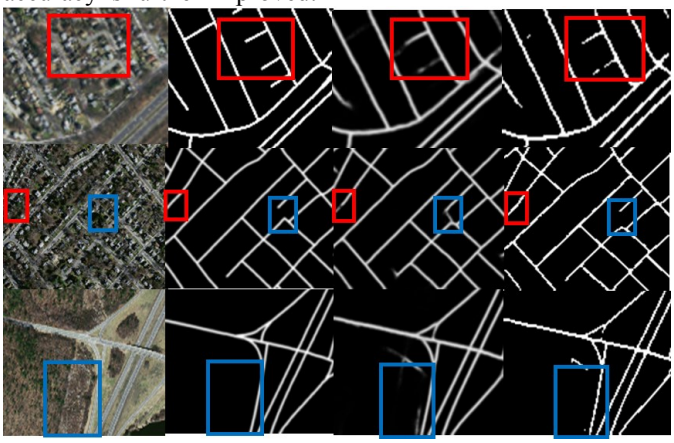

Fig.2 Example results of two models.(a)Input image;(b)Ground truth;(c)U-Net;(d)SAT U-Net

Table 1 presents the results of two models in the Massachusetts roads test set. It is clear that the improved approach significantly outperform U-Net in terms of PA, Jaccard index and Dice coefficient respectively. Moreover, road segmentation performance was compared with two methods in urban regions, grassland, woodland, cultivated land and airports. Table 2 to 6 show the comparison of the road segmentation results of the two methods for five different environments. Overall the proposed method is superior to U-net. The result of grassland provides the best result on the performance of segmentation. In case of the airport and woodland, the similarity between background and roads such as airport runway and land results in the poorer performance of the two models in these two environments.

Table 1 Results on test set of two models in the Massachusetts roads dataset

Model	Accuracy	Dice Coefficient	Jaccard index
U-Net	0.9596	0.5456	0.3928
SAT U-Net	**0.9658**	**0.5960**	**0.4498**

Table 2 Results on two methods in city

Model	Accuracy	Dice Coefficient	Jaccard index
U-Net	0.8882	0.4954	0.3328
SAT U-Net	0.9028	0.5368	0.3726

Table 3 Results on two methods in grassland

Model	Accuracy	Dice Coefficient	Jaccard index
U-Net	0.9719	0.5535	0.3850
SAT U-Net	0.9791	0.6313	0.4653

Table 4 Results on two methods in woodland

Model	Accuracy	Dice Coefficient	Jaccard index
U-Net	0.9902	0.1929	0.1227
SAT U-Net	0.9909	0.2016	0.1335

Table 5 Results on two methods in cultivated land

Model	Accuracy	Dice Coefficient	Jaccard index
U-Net	0.9799	0.5242	0.3573
SAT U-Net	0.9841	0.5706	0.4052

Table 6 Results on two methods in airport

Model	Accuracy	Dice Coefficient	Jaccard index
U-Net	0.9553	0.2647	0.1552
SAT U-Net	0.9552	0.2920	0.1743

V. CONCLUSIONS

In this paper, the SAT U-Net have been proposed for road extraction from remote sensing images. The proposed network replaces the sigmoid layer in the U-Net with a self-adaptive threshold method proposed to self-adaptively adjust the road thresholds for segmentation results of U-Net. Incorporated the advantages of U-Net, the complete road spatial features are preserved and clearer road boundaries are reconstructed. Experimental results demonstrate that the proposed method present better segmentation result than U-Net, in terms of PA, Dice coefficient, Jaccard index. The improved approach solves the problem of unconnected and blurry roads due to the uncertain road threshold and presents a clearer and more complete road structure, providing the accurate and efficient data support in the field of remote sensing data analysis.

VI. FUTURE WORK

A self-adaptive threshold function in the SAT U-Net is proposed to determine the thresholds for road prediction maps. However, in the process of determining the road thresholds, it takes too much time for all images on validation set and test set to calculate the histogram absolute distance. And calculating the histogram distance relies on the histogram results of ground truth to be a reference, which considerably increases the complexity of the road extraction work. Therefore, we will strive to research a new method to determine the road thresholds without the demands of ground truth. In addition, the experimental results show that the proposed method has poor results for road segmentation which is similar to background features, such as dirt roads, forest trails, airport runways, etc. Consequently, we will try our best to research a road segmentation method with higher accuracy, better robustness and versatility to estimate the limitation of road extraction work in any environment and provide more accurate and reliable data support for human beings.

ACKNOWLEDGMENT

This work was supported by Beijing Municipal Natural Science Foundation (No. 4192005).

REFERENCES

[1] J. Cheng, G. Gao, X. Ku and J. Sun, “Review of road network extraction from SAR images,” *Journal of Image and Graphics*, vol.18, no. 1, pp. 11-23, 2013.

[2] G. Cheng, F. Zhu, S. Xiang and C. Pan., “Road centerline extraction Semisupervised Segmentation and Multidirection Nonmaximum Suppression,” *IEEE Geoscience & Remote Sensing Letters*, vol. 13, no. 4 , pp. 545-549, 2017.

[3] Feng. P and Gao. F, “Method of road information extraction in high resolution remote sensing images,” *Modern Electronic Technology*, vol. 38, no. 17, pp. 53-57, 2015.

[4] X. Zhao, “Research on Urban Road Extraction Method of High Resolution Remote Sensing Image,” Huazhong University of Science and Technology ,Wuhan, Hubei, China, 2010.

[5] X. Huang and L. Zhang,“Road centreline extraction from high-resolution imagery based on multiscale structural features and support vector machines,” *Journal of remote sensing,* vol. 30, no. 8, pp. 1977-1987, 2009 .

[6] C. Unsalan and B. Sirmacek, “Road Network Detection Using Probabilistic and Graph Theoretical Methods,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 11, pp. 4441-4453, 2012.

[7] G. Cheng, Y. Wang, Y. Gong, F Zhu and C. Pan, "Urban road extraction via graph cuts based probability propagation", in i*nternational conference on image processing*,2013, pp. 5072-5076.
[8] S. Saito, T. Yamashita, and Y. Aoki, "Multiple object extraction from aerial imagery with convolutional neural networks," *J. ELECTRON IMAGING*, vol. 2016, no. 10, pp. 1–9, 2016.
[9] C. Sujatha and D. Selvathi, "Connected component-based technique for automatic extraction of road centerline in high resolution satellite images," *Eurasip Journal on Image and Video Processing,* vol. 2015, no. 1, pp. 1-8, 2015.
[10] G. Hinton, S. Osindero and Y. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation,* vol. 18, no. 7, pp. 1527-1554, 2006.
[11] V. Mnih and G. Hinton, "Learning to detect roads in high-resolution aerial images," in *ECCV*, pp. 210–223, 2010.
[12] Y. LeCun, L. Bottou, Y. Bengio and P. Haffner, "Gradient-based learning applied to document recognition", *Proceedings of the IEEE,* vol. 86, no. 11, pp. 2278-2324, 1998.
[13] E. Maggiori, Y. Tarabalka, G. Charpiat and P. Alliez, "Convolutional neural networks for large-scale remote-sensing image classification," *IEEE Trans. Geosci. Remote Sens*, vol. 55, no. 2, pp. 645-657, 2017.
[14] M. Volpi and D. Tuia, "Dense Semantic Labeling of Subdecimeter Resolution Images With Convolutional Neural Networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 2, pp. 881-893, 2017.
[15] I. Sevo and A. Avramovic, "Convolutional Neural Network Based Automatic Object Detection on Aerial Images," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 5, pp. 740-744, 2016.
[16] J. Long, E. Shelhamer and T. Darrell, "Fully convolutional networks for semantic segmentation" *computer vision and pattern recognition*, 2015, pp. 3431-3440.
[17] Y. Wei, Z. Wang and M. Xu, "Road Structure Refined CNN for Road Extraction in Aerial Image", *IEEE Geoscience and Remote Sensing Letters,* vol. 14, no. 5, pp. 709-713, 2017.
[18] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *MICCAI*, 2015, pp. 234–241.
[19] T. Sørensen, "A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons," *Kongelige Danske Videnskabernes Selskab,* vol. 5, no. 1, pp. 1-34, 1948.
[20] L. Dice, "Measures of the Amount of Ecologic Association Between Species," *Ecology* ,vol. 26, no. 3, pp. 297-302, 1945
[21] P. Jaccard, "Distribution de la flore alpine dans le bassin des Dranses et dans quelques régions voisines," *Bulletin de la Société Vaudoise des Sciences Naturelles,* vol. 37, pp. 241-272, 1901.