International Cartographic Association
Association Cartographique Internationale

Recognition and Semantic Information Extraction for Map Based on Deep Learning

Yong Wang[a,*], Kaixuan Du[b], Xianghong Che[a], Ruiyuan Ma[a] and Fu Ren[b]

[a] *Chinese Academy of Surveying and Mapping, wangyong@casm.ac.cn, chexh@casm.ac.cn, 3218630753@qq.com*
[b] *School of Resources and Environmental Science, Wuhan University, dukaixuan@whu.edu.cn, 603676001@qq.com*

* Corresponding author

Abstract: Geospatial information contained in maps plays an important role in geographic information data acquisition, map understanding, intelligent mapping and other applications. In terms of map recognition and geospatial information extraction from maps, traditional methods that heavily rely on human or human-computer interaction for semantic recognition can no longer meet the real-time needs. In this paper, we first analysed the composition and characteristics of maps, and then systematically illustrated the semantic understanding methods of map image recognition, target detection of geographic features and semantic segmentation of geographic features based on deep learning architecture, which is crucial to intelligent map recognition and mapping.

Keywords: CNN, Map Recognition, Geographic Feature Detection, Semantic Segmentation

1. Introduction

As an important language, maps express physical space with graphical symbols, which are rich in geospatial data that cannot be matched by the ability of expressions such as text. Maps occupy an important position in the expression and application of geospatial information. A large amount of current research work in cartography is on the process of mapping geospatial objects into maps, only little research has been done on the extraction of geospatial objects from maps and then map semantic understanding. The future of the world is humans and machines coexist, and map semantic understanding in machine holds significant promise.

Map image semantic deconstruction is the recognition and extraction of various basic symbolic contents in the map or the content of the organization structure on it based on the composition of the map. It is the inverse process of map content mapping by realizing vectorization, objectification and structuring of map images through computer image processing and artificial intelligence and other technologies.

Map semantic understanding is a process of intelligent map recognition using algorithms, which focuses on map content recognition and extraction. The coarse-grained understanding of map content is achieved by target detection of interest areas such as key targets and name notes. Due to the diversity of map application requirements and the variety of map types as shown in Figure 1, the understanding of map content requires strong generalization ability, and deep learning methods have obvious advantages in this regard. This paper will realize different granularity of map semantic understanding based on deep learning methods from recognition of map images recognition, key targets extraction in the map and important line objects in maps.

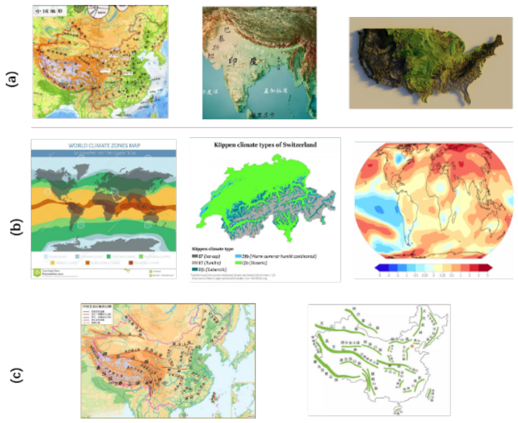

Figure 1. Different thematic maps: (a) topographic; (b) climate; (c) mountain terrain.

2. Map Recognition

2.1 Introduction to Image Classification

In the face of massive map image data, using manual or human-machine interaction to filter map images is an unrealistic method. With the improvement of computer arithmetic power such as distributed computing and cloud computing, a good arithmetic environment is provided for machine recognition of map images . In this context, the deep learning method based on convolutional neural network has significantly better accuracy than the traditional methods based on manual features and classifiers without arithmetic power bottleneck.

Le Net model One of the earliest proposed convolutional neural network models, mainly used for MNIST handwriting classification, is significantly less capable in the face of complex graphic classification tasks. Alex Net As an early deep convolutional neural network model, its

Proceedings of the International Cartographic Association, 5, 25, 2023.
31st International Cartographic Conference (ICC 2023), 13–18 August 2023, Cape Town, South Africa. This contribution underwent single-blind peer review based on submitted abstracts. https://doi.org/10.5194/ica-proc-5-25-2023

structure includes 5 convolutional layers and 3 fully connected layers, which had excellent picture recognition effect at that time, but the network model used a large convolutional kernel on the initial layers of the model, which led to its large number of parameters. VGG net is an inherited framework from Le Net and Alex Net, and is especially similar to the Alex Net framework. In addition, there is the Rest Net family that introduces the residual structure, which mainly addresses the problem that the training error increases with the number of layers in the network depth. Lightweight convolutional neural networks have obvious advantages over the previously mentioned network models in terms of storage space, running time, and computational resource consumption. Squeeze Net network proposed by Iandola F N, Han S, Moskewicz M W, et al contains a total of 11 layers, including one convolutional layer, nine Fire Module layers and one fully connected layer, with about 1.2x106 parameters, and the actual size of the whole network model is 4.8 MB.

2.2 Automatic Recognition of Maps

For the task of map image recognition, this paper selects two network models, Resnet50 and Squeeze Net, to start with, and compares the model effects in terms of both image recognition accuracy and model training and testing efficiency, and concludes that both deep learning methods can be used for map image recognition under different demand conditions.

The method flow of map ®mage recognition using Resnet50 convolutional neural network model is shown in Figure 2.

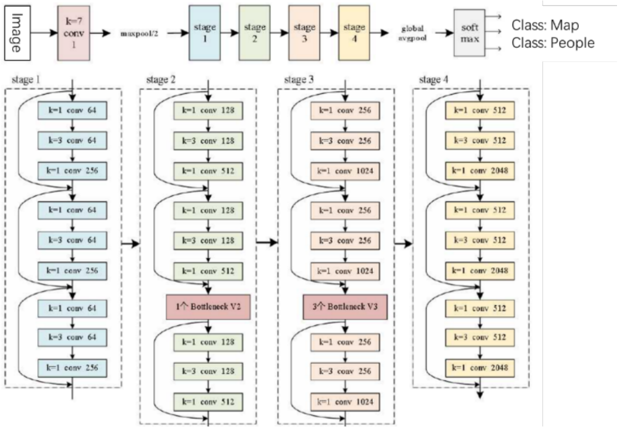

Figure 2 structure of Resnet50 for map classification

The flow of the method for map image recognition using Squeeze Net convolutional neural network model is shown in Figure 3 .

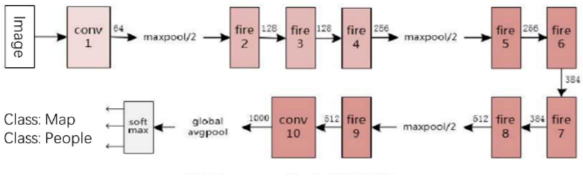

Figure 3. Structure of Squeeze Net for map classification

From the above two figures, we can see that the model parameters of the Squeeze Net network model are much smaller than those of the Resnet50 network model.

2.3 Data and experimental results

We have prepared 92,543 images for map image classification, and then randomly divided 73,933 images into training set, and 14,787 images into test set.The hardware and software environment for the model training is i9-10900X CPU @ 3.70GHz, GeForce RTX 3090 graphics card, Ubuntu 18.04 operating system, Python 3.6, and Tensorflow 2.5 development language. Python 3.6, and the deep learning framework Tensorflow 2.5.

In order to evaluate the accuracy of the model effectively, this paper uses recall, correctness and f1 measure values to evaluate the prediction results of the model. In the map image classification study, by constructing the sample library independently and under the same conditions by comparing the ResNet50 model and Squeeze Net model under the same conditions, the correct rate and recall rate of ResNet50 are 2.01% and 0.32% higher than those of Squeeze Net in map classification; however, the ResNet50 network model is much larger than the Squeeze Net network model because the number of parameters reaches 25.5x106 and the size of the whole network model is about 98 MB. However, the ResNet50 network model is much larger than the Squeeze Net network model because the number of parameters reaches 25.5x10, and the size of the whole network model is about 98MB, so it needs longer training and testing time, and the training and testing time of ResNet50 is 2.51 times and 6.43 times of Squeeze Net under the same conditions. The final experiment shows that the training and testing time of ResNet50 is 2.51 times and 6.43 times of Squeeze Net under the same conditions.

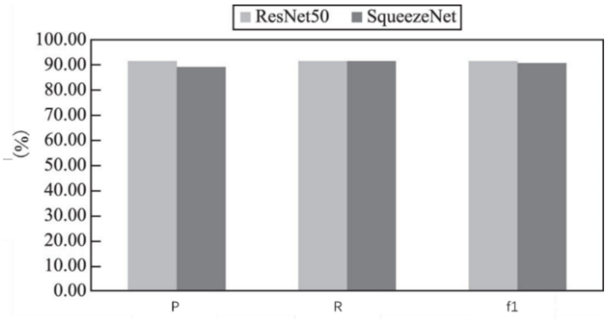

Figure 4 Comparison of Test Accuracy of the Two Networks

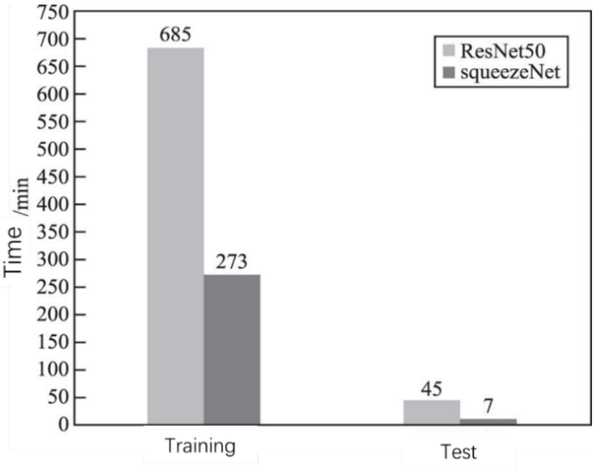

Figure 5 Time Comparison of the Two Networks

Through the analysis of this result, the Squeeze Net network model has a more obvious advantage in systems where resources are insufficient and efficiency and storage become limiting conditions; for systems where resources are sufficient and efficiency and storage capacity do not pose any limitations, the use of ResNet50 is a more preferable choice to ensure more substantial accuracy of classification results. The deep convolutional neural network model approach can achieve map recognition under different computational conditions and demands.

3. Geographic Feature Detection for Map Image

Introduction to Geographic Feature Detection Convolutional neural network models for image classification have achieved considerable success and reached a high level of usability. As for geographic feature detection, the process is more difficult than image classification which requires target localization in addition to identifying the type, which is more difficult for the ordinary deep convolutional neural network model.

geographic feature detection, as a fundamental problem in computer vision, can be classified into two aspects of research according to the focus of related studies: first, geographic feature detection of general objects; second, geographic feature detection research established for specific applications. In this paper, the research on regional geographic feature detection belongs to both the second type of geographic feature detection research, where we develop target-specific recognition and localization research for geographic area targets in maps, and introduce geographic feature detection technology into the research on spatial semantic extraction and recognition of map images.

Compared with the concrete objects in the natural environment, map area targets have higher abstraction and uncertainty, and the same area geographic feature objects in the map have greater differences due to different projection methods, scales, drawing styles and drawing habits, which makes map area geographic feature detection more challenging than the detection of objects in the natural environment.

In the geographic feature detection process, geographic feature detection can be divided into single-stage geographic feature detection methods and two-stage geographic feature detection methods based on the presence or absence of a candidate region generation process in the detection process. The dual-stage geographic feature detection method has a candidate region generation process, so usually, the dual-stage geographic feature detection model takes longer time and is not as fast as the single-stage geographic feature detection, but it is this process that makes the dual-stage geographic feature detection better than the single-stage geographic feature detection in terms of accuracy.

R-CNN, as the pioneer of two-stage geographic feature detection model, applies the convolutional neural network to the geographic feature detection task of image, which has greatly improved the performance in terms of accuracy compared with the traditional method based on manual feature extraction. geographic feature detection model, its first regional proposal network (RPN) and share convolutional features with the backbone network, breaking the time bottleneck of R-CNN and Faster R-CNN models on top of selective search. The literature achieves the detection of multiple specified problem regions in a map by using Faster R-CNN as the base network, which further stimulates the feasibility of deep learning methods in map target recognition by fusing multi-scale feature pyramids and fixed-size features of problem regions.

3.1 Feature detection methods

Map images have obvious big data characteristics, and the traditional manual or human-machine interaction-based approach has obviously failed to meet the needs of recognition and spatial semantic extraction of map images data. In order to carry out the detection of regions of interest in map images as fast as possible with reliable accuracy, the advantages of single-stage geographic feature detection model in terms of time are highlighted.

The single-stage detection model has a high detection speed because there is no candidate frame generation process, but its detection accuracy has obvious disadvantages compared with the two-stage geographic feature detection. loss function, which reduces the contribution of negative samples in model training by multiplying a weakening exponent on the basis of the original cross-entropy loss function, solves the problem of unbalanced training samples to some extent, and improves the detection accuracy of the model while ensuring the detection speed of the single-stage geographic feature detection model.

The commonly used cross-entropy loss function is shown in Equation (1).

$$CE(p, y) = CE(p_t) = -\log(p_t) \tag{1}$$

In order to make the focal loss more balanced for different categories, α coefficients are introduced to obtain the focal loss with better effect.

$$p_t = \begin{cases} p & if\ y = 1 \\ 1 - p & otherwise, \end{cases} \tag{2}$$

Considering the huge difference in the number of positive and negative samples, in order to equalize the guiding effect of positive and negative samples on the direction of convergence of the model parameters, a coefficient $(1-p_t)^{\gamma}$ is introduced. The Focal loss loss function is obtained as shown in Equation (3):

$$FL(p_t) = -(1-p_t)^{\gamma}\log(p_t) \tag{3}$$

In order to make the focal loss more balanced for different categories, α coefficients are introduced to obtain the focal loss with better effect.

$$FL(p_t) = -\alpha_t(1-p_t)^{\gamma}\log(p_t) \tag{4}$$

The RetinaNet network model with focal loss is used in the research process of geographic feature detection, and its network structure is shown in Figure 6. The network mainly consists of three parts: feature extraction network, feature pyramid network and classification regression full convolutional network.

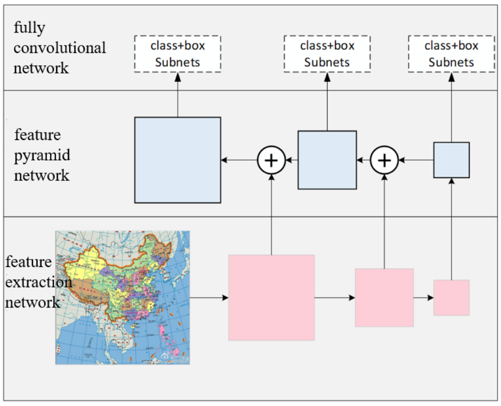

Figure 6. Schematic Diagram of Retina Net

The structure of the full convolutional network in the above figure consists of two sub-networks, classification, and position regression, which are shown in Figure 7.

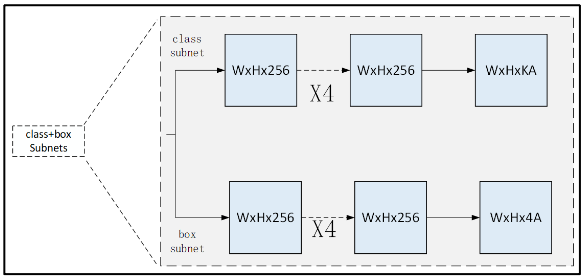

Figure 7. Schematic Diagram of class and box subnets

3.2 Data and Experimental results

Three region of interest in map selected for geographic feature detection in our paper are Taiwan, Tibet and Chinese mainland. The test set was intended to provide an unbiased evaluation of a trained model using the training set. The specific distribution is shown in Table 1. The three targets in the map images were manually annotated using a web-based image annotation tool. The tool outputs an annotation file with an interactive drawing of a bounding box containing all the pixels of the target, which include the directory of each image, the coordinates of the top left corner for the annotated bounding box, the width and height of the annotated bounding box, and the name of the target (Table 2). The principle of manual annotation is to use the smallest possible box to completely cover the targets but get rid of the useless background.

	ROI	Training Dataset	Test Dataset	Total
Target 1	Taiwan	2151	538	2689
Target 2	Tibet	582	146	728
Target 3	Chinese mainland	459	115	574
Total		3192	799	3991

Table 1. Sample distribution of different targets.

path_img_file	box_x	box_y	Width	Height	Label
image_0001.jpg	890	659	944	743	Taiwan
image_0002.jpg	775	631	845	721	Taiwan
image_0003.jpg	36	57	762	535	Tibet

Table 2. Target annotation format.

The hardware and software environment for the model training is i9-10900X CPU @ 3.70GHz, GeForce RTX 3090 graphics card, Ubuntu 18.04 operating system, Python 3.6, and Tensorflow 2.5 development language. Python 3.6, and the deep learning framework Tensorflow 2.5.

We used four evaluation metrics, including intersection over union (IOU), precision, recall, and harmonic mean of precision and recall. IOU is used to measure how much our predicted boundary overlapped with the ground truth (the target's real boundary), which calculated the coincidence degree between the predicted box and the ground truth box. IOU is defined by Equation (5), where B_p represents the predicted bounding box and B_{gt} represents the ground truth bounding box. The threshold of IOU indicates whether the detection is valid or not.

$$IOU = {area(B_{gt} \cap B_p)}\Big/{area(B_{gt} \cup B_p)} \tag{5}$$

The models are evaluated with our test samples, the results are as shown in Table3.

	Taiwan	Tibet	Chinese Mainland
Precision (P)	0.92	0.77	0.52
Recall®)	0.91	0.96	0.94
f1_socre	0.92	0. 86	0.67

Table 3. Accuracy statistics of different targets detection model.

By using a target detection method based on a deep convolutional neural network model, we achieve the identification and extraction of important targets in the map.

4. Semantic segmentation of Geographic Feature in Map

4.1 Introduction to image semantic segmentation

The geographic feature detection for map image eventually only gets a rectangular box to frame out the range of the target area, which has a high usability in the problems of spatial semantic extraction for location and topological relations, but does not provide effective support in the research problems involving finer boundary shapes. In order to obtain finer elements in maps, we introduce the solution to the semantic segmentation problem in the field of deep learning into the recognition and spatial semantic advance of map images.

The map elements described in this subsection refer to the semantic segmentation of symbols with actual geographical meaning in the map, and the identification of the target elements of interest in the map from the pixel level. Internet maps, as a form of map existence, have large differences in Internet maps due to the uneven professional quality of the participants, and the use of traditional semantic extraction of map elements based on manual features has been unable to effectively extract the target elements in Internet map images. In this subsection, we investigate the effect of semantic extraction of this element by using the U-net semantic segmentation network model for a line object in the map.

4.2 Semantic Segmentation methods for Geographic Feature

The network model of U-Net consists of the encoding and decoding processes of the reduced path on the left and the expanded path on the right, which are combined to resemble the shape of a letter "U", as shown in Figure 8, where the blue and white boxes indicate the different stages of feature maps; the blue arrow indicates the 3x3 convolution process for feature extraction; the gray arrow is the jump connection, which is used for feature fusion by copying the clipped feature maps; the red arrow indicates the pooling process for dimensionality reduction; green arrows indicate up sampling for dimensionality recovery; cyan arrows indicate 1x1 convolution for outputting results. In the reduction path of the model, the classical structure of convolutional networks is used, and two 3x3 convolutional kernels are repeatedly used for convolutional operations, followed by activation by ReLU activation function and a 2x2 maximum pooling operation for downsampling. After each downsampling, the number of feature maps is doubled. In the expansion path, each stage contains an upsampling of the feature map, an up-adoption using 2x2 convolution, and the total number of feature channels is reduced by half at the end, followed by a convolution operation with two 3x3 convolution kernels, followed by a ReLU activation function activation.

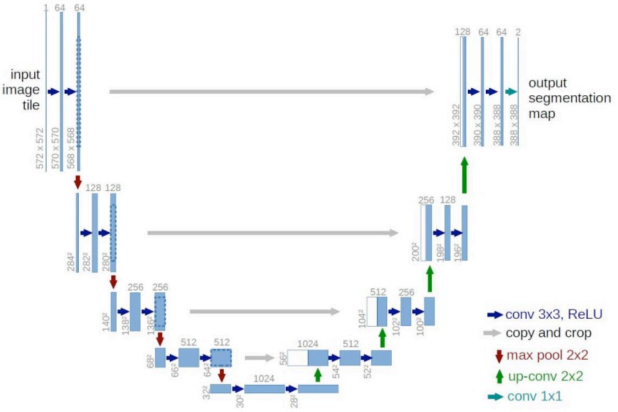

Figure 8. Schematic Diagram of U-net

4.3 Data and experimental results

Deep learning methods cannot be carried out without the preparation of training samples. In this subsection, in order to perform sample annotation with faster speed and better results, we introduce image spectral transform to extract high frequency information in images, which are usually line features. Then the images are manually annotated to obtain fine annotated samples after the extraction of high frequency information. In this subsection, the spectral transform of the map sample is realized by Laplace transform, and then the manual adjustment of the sample labels is performed on the basis of Laplace transform, and the final semantic segmentation model training sample is shown in Figure 9.

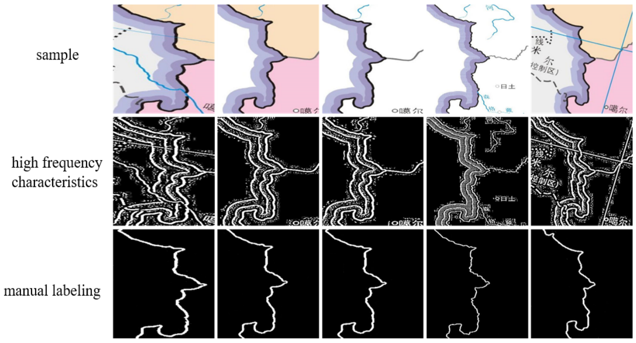

Figure 9. Generation of Sample Label

With the sample labeling method described above, a total of 48 training samples and 17 test samples were labeled in this subsection of the study. The hardware and software environment for the model training is i9-10900X CPU @ 3.70GHz, GeForce RTX 3090 graphics card, Ubuntu 18.04 operating system, Python 3.6, and Tensorflow 2.5 development language. Python 3.6, and the deep learning framework Tensorflow 2.5.

And the semantic segmentation results are evaluated by two metrics, pixel accuracy PA and pixel intersection IOU ratio, which are shown in Equation (6) and Equation (7). The pixel accuracy is a more basic metric of semantic segmentation, which indicates the proportion of correctly segmented pixels to the total pixels in the whole image, and can be understood as the percentage of correctly classified pixels in the image. The intersection and merge ratio is the intersection of the true and predicted values of pixels divided by the merge of the true and predicted values of pixels.

$$PA = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

$$IoU = \frac{TP}{TP + FP + FN} \quad (7)$$

where TP denotes the number of pixels that are correctly segmented, TN the number of correctly non-targeted segmented pixels, FP the number of incorrectly segmented pixels, and FN the number of pixels in which the target object is not segmented.

Using the model to identify the test data, the results obtained are shown in Figure 10.

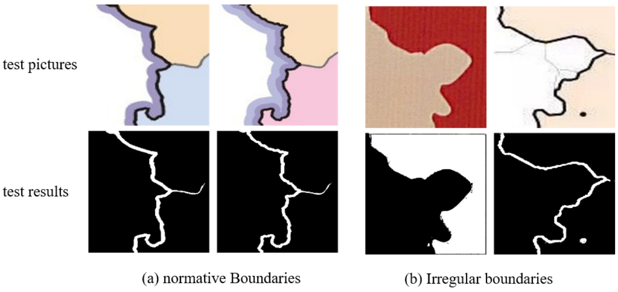

Figure 10. Results of Image Semantic Segmentation

The results of the following Table 4 were obtained by counting the test data.

metrics	test-img1	test-img2	test-img3	test-img4	AVG
PA	90.75	65.24	86.70	99.62	85.58
IoU	85.70	61.92	83.31	98.33	82.32

Table.4 Typical Geographic Target Boundary Line extraction result with U-net

From the above table, the pixel accuracy PA of the U-net network model for the test data can reach up to 0.99 with a mean value of 0.85, and the intersection ratio IoU can reach up to 0.98 with a mean value of 0.82. The mean values of both indexes are above 0.8. Through the analysis of the results of this experiment, the typical geographic elements in the map can be carried out by the deep learning semantic segmentation method.

5. Conclusion

Facing the massive map picture data in the Internet environment or historical paper atlas, map picture recognition and spatial semantics serve as an effective means to achieve geospatial data mining for this type of data. This paper discussed the specific of big data faced by Internet images, and pointed out the importance of geospatial information contained in Internet map images. Then, the importance and feasibility of Internet map recognition and spatial semantic extraction are discussed from various aspects such as classification recognition of map images based on deep learning, target detection of interest regions in map images, semantic segmentation at the pixel level of line symbols, and recognition of annotated content in maps. Although, the extraction of spatio-temporal semantics from Internet map images has been achieved to a certain extent and certain applications have been made, further Internet map data mining is needed in the face of the uncertainty brought by the influence of multiple medium factors such as the massive Internet map data, projection method, scale, and drawing style.

6. References

Abend P, Harvey F. Maps as geomedial action spaces: considering the shift from logocentric to egocentric engagements[J]. GeoJournal, 2017, 82(1): 171-183.

Ai Tinghua. Some Thoughts on Deep Learning Enabling Cartography[J]. Acta Geodaetica et Cartographica Sinica, 2021,50(09):1170-1182.

Bai Youda, Liu Jiping, Huang Long, et al. Analysis of two convolutional neural networks for map image recognition [J]. Science of Surveying and Mapping, 2021, 46(11):126-134.

Cao, G.; Xie, X.; Yang, W.; Liao, Q.; Shi, G.; Wu, J. Feature-Fused SSD: Fast Detection for Small Objects. In Proceedings of the Ninth International Conference on Graphic and Image Processing (ICGIP 2017), Qingdao, China, 10 April 2018; Volume 10615, pp. 381–388.

Chen, L.-C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected Crfs. *arXiv* **2014**, arXiv:1412.7062.

Cho, S.M.; Kim, Y.-G.; Jeong, J.; Kim, I.; Lee, H.; Kim, N. Automatic Tip Detection of Surgical Instruments in Biportal Endoscopic Spine Surgery. *Comput. Biol. Med.* **2021**, *133*, 104384.

CUI Tengteng, LIU Jiping, LUO An. Intelligent identification method of network map images based on convolutional neural network[J]. Science of Surveying and Mapping, 2019, 44(1): 118-123.

Dai, J.; Li, Y.; He, K.; Sun, J. R-Fcn: Object Detection via Region-Based Fully Convolutional Networks. *Adv. Neural Inf. Process. Syst.* **2016**, *29 , 379–387*.

Du Kaixuan, Wang Liang, Wang Yong, et al. An Identification Method of Map Images Based on Activate Learning and Convolutional Neural Network[J]. Science of Surveying and Mapping, 2020, 45(7): 139-147.

Du K, Che X, Wang Y, et al. Comparison of RetinaNet-Based Single-Target Cascading and Multi-Target Detection Models for Administrative Regions in Network Map Pictures[Z]. 2022: 22.

Felzenszwalb, P.; McAllester, D.; Ramanan, D. A Discriminatively Trained, Multiscale, Deformable Part Model. In Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, USA, 23–28 June 2008; pp. 1–8.

Guo Renzhong, Chen Yebin, Ying Shen, et al. Geographic Visualization of Pan-Map with the Context of Ternary Spaces[J/OL]. Wuhan Daxue Xuebao (Xinxi Kexue

Ban)/Geomatics and Information Science of Wuhan University, 2018, 43(11): 1603-1610.

HE Haiwei, QIAN Haizhong, XIE Limin, DUAN Peixiang. Interchange recognition method based on CNN[J]. Acta Geodaetica et Cartographica Sinica, 2018,47(03):385-395.

He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

HE Xiaofei, ZOU Zhengrong, TAO Chao, ZHANG Jiaxing. Combined saliency with multi-convolutional neural network for high resolution remote sensing[J]. Acta Geodaetica et Cartographica Sinica,2016,45(09):1073-1080.

Hsu, C.-W.; Lin, C.-J. A Comparison of Methods for Multiclass Support Vector Machines. *IEEE Trans. Neural Netw.* **2002**, *13*, 415–425.

Iandola F N, Han S, Moskewicz M W, et al. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5 MB model size [J/OL]. arxiv.org [2022-08-18]. https://arxiv.org/abs/1602.07360, 2016.

Jiao, L.; Zhang, F.; Liu, F.; Yang, S.; Li, L.; Feng, Z.; Qu, R. A Survey of Deep Learning-Based Object Detection. *IEEE Access* **2019**, *7*, 128837–128868.

Jiayao, W.; Yi, C. Discussions on the Attributes of Cartography and the Value of Map. Acta Geod. Cartogr. Sin. 2015, 44, 237.

Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[J]. Advances in neural information processing systems, 2012, 25.

LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

LI Chaoqi. Age classification methods based on deep convolutional network and HAMX Model[D]. School of Computer Science and Technology Donghua University, 2017.

Lin T Y , Goyal P , Girshick R , et al. Focal Loss for Dense Object Detection[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, PP(99):2999-3007.

LIU Kang, QIAN Xu, WANG Ziqiang. Survey on active learning algorithms[J]. Computer Engineering andApplications, 2012, 48（34）：1-4..

Liu, S.; Cai, T.; Tang, X.; Zhang, Y.; Wang, C. Visual Recognition of Traffic Signs in Natural Scenes Based on Improved RetinaNet. *Entropy* **2022**, *24*, 112.

Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. Ssd: Single Shot Multibox Detector. In *Proceedings of the European Conference on Computer Vision, Amsterdam, Netherlands, 8-16 October 2016*; pp. 21–37; DOI: doi.org/10.1007/978-3-319-46448-0_2.

MA Fei. Scanning contour map recognition system based on Windows[J]. Geomatics and Information Science of Wuhan University[J], 1995(3):228-233.

Ren Fu, Hou Wanyue. Identification Method of Map Name Annotation Category for Machine Reading[J]. Geomatics and Information Science of Wuhan University, 2020, 45(2): 273-280.

Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-Cnn: Towards Real-Time Object Detection with Region Proposal Networks. *Adv. Neural Inf. Process. Syst.* **2017**, *39,* 1137–1149; DOI:10.1109/TPAMI.2016.2577031 .

Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. *IEEE Conf. Comput. Vis. Pattern Recognit.* **2016**, Seattle, WA, USA, 27-30 June 2016; 779–788; DOI:10.1109/CVPR.2016.91.

Schneiderman, H.; Kanade, T. Object Detection Using the Statistics of Parts. *Int. J. Comput. Vis.* **2004**, *56*, 151–177.

Sharma, V.; Mir, R.N. A Comprehensive and Systematic Look up into Deep Learning Based Object Detection Techniques: A Review. *Comput. Sci. Rev.* **2020**, *38*, 100301.

Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.

Schneiderman, H.; Kanade, T. Object Detection Using the Statistics of Parts. Int. J. Comput. Vis. 2004, 56, 151–177.

Sharma, V.; Mir, R.N. A Comprehensive and Systematic Look up into Deep Learning Based Object Detection Techniques: A Review. *Comput. Sci. Rev.* **2020**, *38*, 100301.

TIAN Yan-ling, ZHANG Wei-tong, ZHANG Qie-shi, LU Gang, Review on image scene classification technology[J]. Acta Electronica Sinica, 2019,47(4):915-926.

Tripathi, S.; Dane, G.; Kang, B.; Bhaskaran, V.; Nguyen, T. Lcdet: Low-Complexity Fully-Convolutional Neural Networks for Object Detection in Embedded Systems. *IEEE Conf. Comput. Vis. Pattern Recognit. Workshops* **2017**, 411-420 ; DOI: 10.1109/CVPRW.2017.56.

Wang, N.; Gao, X.; Tao, D.; Yang, H.; Li, X. Facial Feature Point Detection: A Comprehensive Survey. *Neurocomputing* **2018**, *275*, 50–65.

WANG Xuebing, GUO Qingsheng, WANG Yong, et al.Feature extraction and automatic recognition method for map images[J]. Geomatics & Spatial Information Technology,2019,42(9):28-32.

WANG Tian-heng, ZHANG Yi. Research on an active learning algorithm based on multi-application scenes[J]. Modern Computer, 2018(10):40-43.

Wu, X.; Sahoo, D.; Hoi, S.C.H. Recent Advances in Deep Learning for Object Detection. *Neurocomputing* **2020**, *396*, 39–64.

YANG Yun. Geographic information extraction based on maps and remote sensing images[D]. Information Engineering University, 2008.

YANG Qihe, ZHU Wenzhong, HUANG Wenjian. The introduction and prospect of map pattern recognition[J]. Geomatics Technology and Equipment, 1996(03):16-20.

Ye, Q.; Doermann, D. Text Detection and Recognition in Imagery: A Survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *37*, 1480–1500.

Zhang Ke, Feng Xiaohan, Guo Yurong, et al. Overview of deep convolutional neural networks for image classification[J]. Journal of Image and Graphics, 2021, 26(10): 2305-2325.

ZHOU Fei-Yan, JIN Lin-Peng, DONG Jun. Review of convolutional neural network[J]. Chinese Journal of Computers, 2019, 44(1): 118-123.

Zou Z, Shi Z, Guo Y, et al. Object detection in 20 years: A survey. arXiv[J]. arXiv preprint arXiv:1905.05055, 2019.

31st International Cartographic Conference (ICC 2023), 13–18 August 2023, Cape Town, South Africa. This contribution underwent single-blind peer review based on submitted abstracts. https://doi.org/10.5194/ica-proc-5-25-2023