Deep Neural Network Based on Sparse Auto-Encoder for Road Extraction

Sheng Liu, Shuxiao Chang, Ting Cao(✉), and Xinyue Li

Department of Computer Science and Engineering, Xi'an University of Technology, Xi'an, Shaanxi, China
caoting@xaut.edu.cn

Abstract. Road extraction from aerial image has realistic significance for GIS data updating. In view of the complexity challenging for acquiring road information, this paper proposes supervised model that combines Convolutional Neural Network (CNN) with Sparse Auto-Encoder (SAE) to cope with the road extraction task. First, the road features are extracted from the amount of non-annotated data using SAE model that aim to train the road features using CNN principle with implementing convolution and pooling to reduce model complexity. Second, the encoder network completes the operation, and after the deep pooling and deconvolution operations, the intermediate features are extracted by the decoder network and sampled back to the input image of the same size on the map. Third, the soft-max classifier categorizes images into roads and non-roads. Finally, the experiments verify that the proposed method outperforms the traditional methods and could achieve the satisfy result.

Keywords: Road extraction · aerial image · Deep learning · Convolutional Neural Network · Sparse Auto-encoder

1 Introduction

Road extraction from aerial images has vital usage in many applications including geographic information system, intelligent transportation system, environmental security and protection [1]. Various road extraction approaches can achieve road extraction successfully when the road exhibit obvious contrast respect with the non-road areas [2]. However, when the road with complex situation, such as road vehicles, buildings, tree occlusion cases, road extraction often appears discontinuous or gaps [3]. It is still challenging to deal with shadow or occlusion, geospatial information (urban, suburban or rural), and image scales, and obtain full and smooth road network automatically [4].

With the rapid development of deep learning in recent years [5], road extraction can be regarded as a classification task to distinguish aerial image into the road areas and the background areas [6, 7]. The state-of-the-art Convolutional Neural Network (CNN) is viewed as a successful deep learning model. CNN has advantages in hierarchical learning that makes it more efficient in feature extraction and image classification.

V. C. M. Leung et al. (Eds.): Qshine 2023, LNICST 573, pp. 120–127, 2024.
https://doi.org/10.1007/978-3-031-65126-7_12

Therefore, due to the aerial images have more complex backgrounds and targets. In this paper, a semi-automatic method combined Deep CNN with SAE (Sparse Auto-Encoder) is proposed to detect the road information from aerial image. First, the SAE model is carried out to learn the relationships and features of complex data and extract concise expressions from them automatically. Second, the encoder network completes the operation, and after the deep pooling and deconvolution operations, the intermediate features are extracted by the decoder network and sampled back to the input image of the same size on the map. Both convolution and pooling are implemented to reduce model complexity and boost distance calculation. Third, the final output is obtained by using the classifier, which is the probability distribution in the image representing the likelihood that the pixels in each region belong to the road and the non-road.

2 Related Work

In recent years, many methods have studied on road extraction from aerial image. Pradhan [8] proposed an automatic road extraction method by the neural network, which was superior to many methods of previous studies due to their ability to incorporate both multi-source information. Soni [9] presented a neural network to extract roads by a variety of texture parameters, and followed by the road vectorization stage. Experiments were carried out on different IKONOS and Quick Bird sample images to prove the road extraction capability of the proposed method.

Moreover, Nguyen [10] proposed a road extraction scheme based on feature learning, using convolutional neural network to capture the local structure of the road network. Due to the powerful learning ability of CNN, the road extraction method that we proposed can obtain high quality results. Wei [11] introduced a concise CNN for road extraction in aerial image. The paper proposed a new loss function which integrates the road geometry information into the cross-entropy loss. Experimental results showed that the model could perform well in accuracy, recall, F-score and accuracy.

Also, Wang [12] adopted a single patch architecture to extract roads from high-resolution images. Alshehhi [13] proposed a CNN network with integrated structure based on Alex-Net and VGG-net. Due to the large network structure, Alex-Net paid attention to the information of the large area. VGG networks focused on local details because of their small size. In this work, the training, verification and testing of the current popular deep learning models under different parameters have a good foundation for the identify and extraction of large geological and scientific data such as roads and buildings [14]. The accuracy of the road extraction is significantly improved.

3 Methodology

In our work, a semi-supervised based deep learning method was proposed, which combines Deep CNN (Convolutional Neural Network) with SAE (Sparse Auto-Encoder) to detect the road information from aerial image. In this part, the detail description of the concrete algorithms applied in our network is shown at first, and the overall framework and the algorithm execution process are elaborated on the follow.

3.1 SAE Model

The performance of image classification is largely depended on the pros and cons of extracted features. SAE model is more suitable for unsupervised learning, which does not need a large number of tags during training massive aerial images. It can avoid the annotation of massive remote sensing images, and greatly improve the automation of the method. The unnecessary of annotation work can greatly improve the automation and efficiency of the algorithm [16].

The classic structure of SAE usually includes an input layer, in Fig. 1, a hidden layer, and an output layer, where +1 is the offset term.

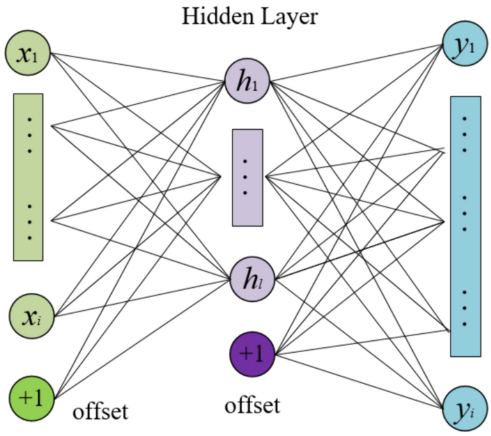


Fig. 1. SAE model

The loss function for neural network can be denoted as in Eq. (1):

$$J(W,b)=\left[\frac{1}{m}\sum_{i=1}^{m}\left(\frac{1}{2}\left\|h_{W,b}(x^i)-y^i\right\|^2\right)\right]+\frac{\lambda}{2}\sum_{l=1}^{n_l-1}\sum_{i=1}^{s_l}\sum_{j=1}^{s_l+1}(W_{ji}^l)^2 \tag{1}$$

where, m is the amount of input samples, $(W,\ b)$ is the network parameter, n_l stands for the layers amount, s_l denotes the node amount in L layer, λ means the regularization and $h_{W,b}(x^i)$ means the output sample.

The SAE algorithm constrains the output of the hidden layer, so that the average could be high as 0. The loss function of SAE algorithm can be denoted as in Eq. (2):

$$J_{sparse}(W,b)=J(W,b)+\beta\sum_{j=1}^{s2}KL(\rho\|\hat{\rho})\tag{2}$$

where, ρ stands for the sparse parameter, $KL(\rho||\hat{\rho})$ measures the distributions.

3.2 Deep CNN

Methods based on deep learning have aroused widespread concerns, which establish a high-level semantic mapping relation by extracting the features. As a kind of feed-forward deep learning network, the Deep CNN is suitable for image feature extraction and recognition [15]. Usually, CNN architecture includes convolutional, mapping, pooling, fully connected, and output layers, that can be formed by stacking multiple underlying network structures.

First, feature extraction is performed in convolutional layer, and the formula can be denoted as in Eq. (3):

$$y_i = b_i + \sum\nolimits_i k_{ij} \ddot{A} x_i \tag{3}$$

where, y_i means the output image, x_i means the input image, $\otimes$ denotes convolution operator and k_{ij} is kernel function, finally, b_i is deviation value.

Second, the mapping layer employs a nonlinear activation function to obtain the feature map from the convolutional layer. The commonly used activation function is ReLU, sigmoid, tanh and softplus. Usually, the ReLU (Rectified Linear Units) function is employed as the activation function because the output will be zero, which could reduce the network and smooth the over-fitting problem.

Then, the pooling layer could avoid over-fitting phenomenon and maintain spatial invariance. And the full connection layer connects to all the previous layers including convolution layer or another full connection layer.

To train the network as a better performance, some operators, such as the local response normalization and dropout regularization method, are added to optimize results and speed up the training process. It randomly reduces the output of some neurons and reduces the neurons in the network that are no longer involved in the computation.

Finally, the classifier layer with full link is used to output in probabilistic form for each category. The most used loss function output in is the softmax function.

3.3 Framework

Therefore, a semi-supervised based deep learning method was proposed. The specific steps are as follows: Feature extraction part adopts the SAE model to study and find out the relationship between the optimal, get a concise expression, DCNN decoder of network from the encoder on the extraction of feature mapping samples back to the same size of the input image, and finally, at the end of the DCNN network using softmax classifier for the probability of road pixels in the final output.

Figure 2 shows the framework of our proposed method. The features learned by SAE, are applied to the convolution of a large number of training sets and test sets. The proposed DCNN network layers include one input, five conventional, two pooling, and one output. A max-pooling operation is performed between layer 1 and Layer 2. The average pooling layer follows the convolution of the five layers.

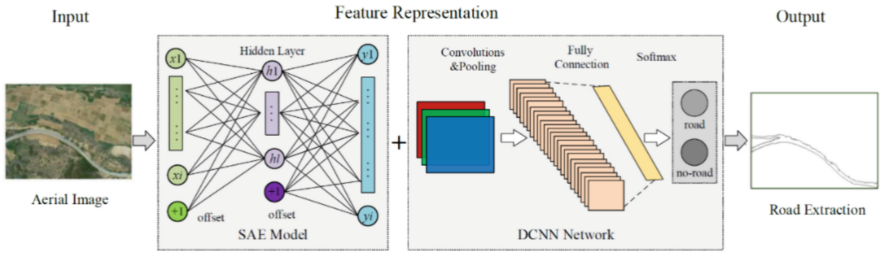


Fig. 2. Framework

4 Experimental Result and Analysis

4.1 Dataset Description

The experimental results of the above network framework are as follows. The dataset consists of two categories (urban roads and rural roads) with 900 images per category, where 400 images for training and 50 images for each group. For each image, the classification of ground truth is annotated by manual with the advice of experts carefully.

Through a large number of experiments, different initial values are selected, and the parameters with the highest performance in the cluster are selected to complete the network design. The evaluation system including Completeness, Correctness and F1 is used to test the road extraction performance. The formula can be denoted as in Eq. (4):

$$Com = \frac{TP}{TP + FN} \; Cor = \frac{TP}{TP + FP} \; F_1 = 2\frac{Com \times Cor}{Com{+}Cor} \tag{4}$$

where, *Com* means the completeness of matched with *GT* (ground truth) calculated by *TP* (truth positive) and *FN* (false negative), and *Cor* is correctness of matched with ground truth by *TP* and *FP* (false positive). F_1 is an overall that combines *Com* and *Cor*.

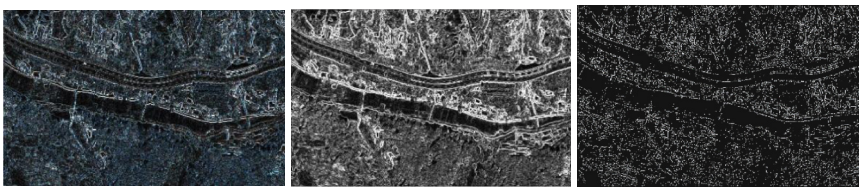
4.2 Result and Analysis

The proposed method is compared and to test robustness and flexibility of the related methods. The showing example from the testing dataset are shown in Fig. 3. In our testing images, the images were numbered as follows.

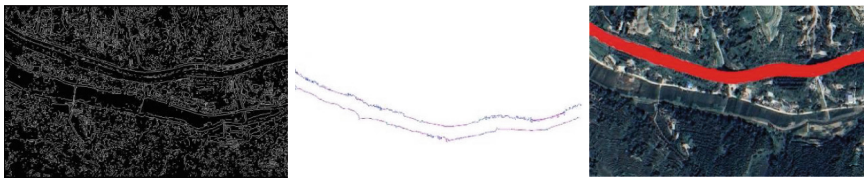
Figure 3 shows the different ways to achieve the road extraction, Table 1 gives the objective comparison of the results using Completeness, Correctness and F_1. Figure 4 is the line chart, which could exhibit the objective comparison more intuitively. In terms of Completeness, Correctness, and F1-score, the proposed method gives the best result in general.

The test can verify that the proposed method has some advantages compared to some existing methods in this field, and the results by our method is very close to ground truth, which are higher than the other methods. But for the Correctness, it is Performance is a bit poor which is caused by the almost indistinguishable gray level between the roads and the background in the bottom of the image.

(a)Input (b)binary (c)Otsu



(d)Prewitt (e)Sobel (f)Valley

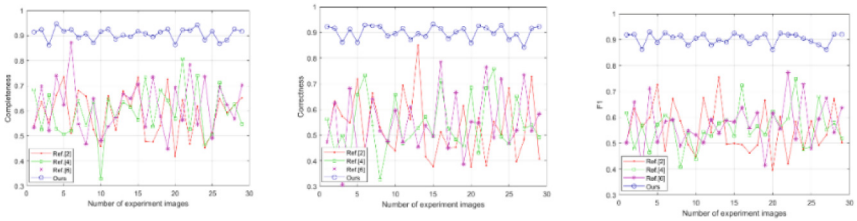


(g)Canny (h)Our method (i)Ground truth

Fig. 3. Comparison of different methods

Table 1. Objective Comparison

Methods	Completeness			Correctness			F1		
	aver	max	min	aver	max	min	aver	max	min
Ref. [4]	0.587	0.735	0.418	0.534	0.851	0.376	0.553	0.755	0.396
Ref. [6]	0.596	0.807	0.329	0.544	0.758	0.327	0.560	0.749	0.408
Ref. [8]	0.619	0.871	0.448	0.549	0.786	0.308	0.574	0.773	0.387
Our method	0.906	0.926	0.864	0.901	0.929	0.859	0.903	0.926	0.861



(a)Completeness (b)Correctness (c)F1

Fig. 4. Curve comparison for different methods.

5 Conclusion

Target detection in aerial images has been widely applied in many fields, including agriculture, forestry, electric power, land resources, urban planning, etc. In the acquisition process of aviation data, aircraft or UAV are constrained by the external environment, stability, wind resistance ability and clarity are limited, jitter phenomenon often occurs, camera Angle changes, etc. These uncertain factors will directly lead to the difficulty of road extraction.

In this paper, a semi-automatic framework combining DCNN and SAE is studied to extract road information from aerial images. SAE model is used to learn the correlation between complex data, and the brief expression is found from the feature perspective. The decoder network samples the feature map extracted from the encoder network back to the input image of the same size, and finally the correct classification output is obtained by softmax classifier. Experimental results show that the proposed algorithm reduces the complexity of the model and improves the speed of calculation.

Acknowledgment. This work was supported in part by the Special Project of Technological Innovation and Guidance in Shaanxi Province under Grant 2022QFY01-03, in part by the Natural Science Foundation in Shaanxi Province under Grant 2022JQ-476, and in part by the Natural Science Foundation of Deduction Department in Shaanxi Province under Grant 2022JK0474, and by Science and Technology Program in Xi'an city under Grant 21XJZZ0055.

References

1. Eerapu, K.K., Lal, S., Narasimhadhan, A.V.: O-Seg-Net: robust encoder and decoder architecture for objects segmentation from aerial imagery data. IEEE Trans. Emerg. Top. Comput. Intell. **PP**(99), 1–12 (2021). https://doi.org/10.1109/TETCI.2020.3045485
2. Abdollahi, A., Pradhan, B.: Road extraction from open-source remote sensing dataset based on the modified deep convolutional autoencoders model. In: 43rd COSPAR Scientific Assembly Sydney, Australia, 28 January–04 February 2021. 2021
3. Tabibi, Z., Schwebel, D.C., Zolfaghari, H.: Road-crossing behavior in complex traffic situations: a comparison of children with and without ADHD. Child Psychiatry Hum. Dev. 1–8 (2021). https://doi.org/10.1007/s10578-021-01200-y

4. Sebasco, N.P., Sevil, H.E.: Graph-based image segmentation for road extraction from post-disaster aerial footage. Drones **6**(11), 315 (2022). https://doi.org/10.3390/drones6110315
5. Vigneshwaran, S.A., Panneer, S.: Situational Analysis of Road Traffic Accidents-Acase of Madurai District rural areas (2020)
6. Zhang, X., Ma, W., Li, C., et al.: Fully convolutional network-based ensemble method for road extraction from aerial images. IEEE Geosci. Remote Sens. Lett. **PP**(99), 1–5 (2019). https://doi.org/10.1109/LGRS.2019.2953523
7. Cheng, G., Wang, Y., Xu, S., et al.: Automatic road detection and centerline extraction via cascaded end-to-end convolutional neural network. IEEE Trans. Geosci. Remote Sens. **55**(6), 3322–3337 (2017). https://doi.org/10.1109/TGRS.2017.2669341
8. Alamri, A.M.: RoadVec-Net: a new approach for simultaneous road network segmentation and vectorization from aerial and google earth imagery in a complex urban set-up. GISci. Remote Sens. (2021). https://doi.org/10.1080/15481603.2021.1972713
9. Soni, P.K., Rajpal, N., Mehta, R.: Road network extraction using multi-layered filtering and tensor voting from aerial images. Egypt. J. Remote Sens. Space Sci. **24**(2), 211–219 (2021). https://doi.org/10.1016/j.ejrs.2021.01.004
10. Nguyen, T.L., Han, D.: Detection of road surface changes from multi-temporal unmanned aerial vehicle images using a convolutional Siamese network. Sustainability **12**(6), 2482 (2020). https://doi.org/10.3390/su12062482
11. Wei, Y., Wang, Z., Xu, M.: Road structure refined CNN for road extraction in aerial image. IEEE Geosci. Remote Sens. Lett. **14**(5), 709–713. 1027. https://doi.org/10.1109/LGRS.2017.2672734
12. Wang, S., Mu, X., Yang, D., et al.: Road extraction from remote sensing images using the inner convolution integrated encoder-decoder network and directional conditional random fields. Remote Sens. (2021).https://doi.org/10.3390/rs13030465
13. Alshehhi, R., Marpu, P.R., Woon, W.L., et al.: Simultaneous extraction of roads and buildings in remote sensing imagery with convolutional neural networks. Isprs J. Photogramm. Remote Sens. **130**(aug.), 139–149 (2017). https://doi.org/10.1016/j.isprsjprs.2017.05.002
14. Ganapathy, P., Skipper, J.A.: A novel ROC approach for performance evaluation of target detection algorithms. In: Conference on Automatic Target Recognition XVII. Department of Biomedical, Industrial and Human Factors Engineering, Wright State University, 207 Russ Engineering Center, 3640 Colonel Glenn Hwy, Dayton, OH 45435 (2007)
15. Alshaikhli, T., Liu, W., Maruyama, Y.: Simultaneous extraction of road and centerline from aerial images using a deep convolutional neural network. Int. J. Geo-Inf. (3) (2021). https://doi.org/10.3390/IJGI10030147
16. Pereg, D., Cohen, I., Vassiliou, A.A.: Sparse seismic deconvolution via recurrent neural network. J. Appl. Geophys. **175**, 103979 (2020). https://doi.org/10.1016/j.jappgeo.2020.103979