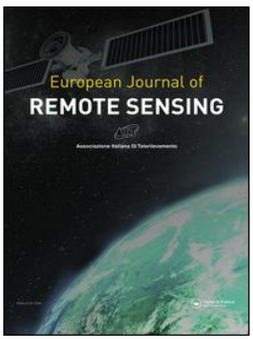

European Journal of Remote Sensing

ISSN: (Print) 2279-7254 (Online) Journal homepage: https://www.tandfonline.com/loi/tejr20

A new road extraction method using Sentinel-1 SAR images based on the deep fully convolutional neural network

Qianqian Zhang, Qingling Kong, Chao Zhang, Shucheng You, Hai Wei, Ruizhi Sun & Li Li

To cite this article: Qianqian Zhang, Qingling Kong, Chao Zhang, Shucheng You, Hai Wei, Ruizhi Sun & Li Li (2019) A new road extraction method using Sentinel-1 SAR images based on the deep fully convolutional neural network, European Journal of Remote Sensing, 52:1, 572-582, DOI: 10.1080/22797254.2019.1694447

To link to this article: https://doi.org/10.1080/22797254.2019.1694447

Published online: 28 Nov 2019.

EUROPEAN JOURNAL OF REMOTE SENSING
2019, VOL. 52, NO. 1, 572–582
https://doi.org/10.1080/22797254.2019.1694447

Taylor & Francis
Taylor & Francis Group

A new road extraction method using Sentinel-1 SAR images based on the deep fully convolutional neural network

Qianqian Zhang[a,b], Qingling Kong[a,b], Chao Zhang[b,c], Shucheng You[d], Hai Wei[d], Ruizhi Sun[a] and Li Li[b,c]

[a]College of Information and Electrical Engineering, China Agricultural University, Beijing, China; [b]Key Laboratory of Remote Sensing for Agri-Hazards, Ministry of Agriculture, Beijing, China; [c]College of Land Science and Technology, China Agricultural University, Beijing, China; [d]Land Remote Sensing, China Land Surveying and Planning Institute, Beijing, China

ABSTRACT

There is a pressing need for an automatic road extraction method due to the continuous development of transportation networks. Free from the influence of weather, satellite-mounted synthetic-aperture radar (SAR) opens the way for such a road detection application. This study introduces an automatic discrimination method based on a deep neural network (DNN) adjusted for roads from dual-polarimetric (VV and VH) Sentinel-1 SAR imagery. In this proposed method, the U-Net extended the convolutional neural network (CNN), is adjusted for road extraction from SAR images. To investigate the potential of using the U-Net based fully convolutional neural network (FCN) for road extraction, LeNet-5, is applied as a preliminary DNN model. Additionally, several training optimizations are introduced to improve accuracy. To assess the performance of the different polarization modes used in road extraction, both single-polarimetric and dual-polarimetric data were investigated. Moreover, four machine learning algorithms have been compared for accuracy and speed. As a result, the outcome evaluation of Precision, Recall, and F_1 obtained by FCN is better than the original CNN, and the training time has been significantly reduced. The DNN model (CNN and FCN) is superior to machine learning methods in accuracy and elapsed computation time.

ARTICLE HISTORY
Received 17 October 2018
Revised 14 September 2019
Accepted 14 November 2019

KEYWORDS
Deep learning; Sentinel-1; SAR; DNN; FCN; road extraction

Introduction

Roads and transportation networks have developed extensively over the past couple of decades in China; thus, there is a pressing need for an accurate road extraction method to keep pace with the necessity of continuously updating road maps. Traditional surveying methods are laborious and time-consuming. However, the considerable development of remote sensing technology in recent years has opened the way to an automatic road detection application supporting a wide coverage area. The advantages of satellite remote sensing technology are: it is macroscopical, dynamic, and provides real-time data. Taken into consideration different images (optical or SAR), different areas (suburb or rural) and different types of roads (highways, rural roads or streets), Wang et al. (2016) have classified road extraction algorithms into some examples: classification (feature)-based, knowledge-based, mathematical morphology, active contour model, and dynamic programming, etc. Among these methods (Long & Zhao, 2005; Miao, Shi, Samat, Lisini, & Gamba, 2016; Naouai, Hamouda, & Weber, 2010), supervised classification-based methods performed the best, with respect to precision. However, most of them are applied for optical satellite imagery with higher pixel resolution (in contrast to SAR imagery). Moreover, optical remote sensing data are susceptible to cloud cover and rain, which may restrict data accessibility and therefore limit its usability. On the other hand, SAR is not affected by weather conditions and therefore provides rich detail for terrain recognition along with amplitude, intensity, scattering coefficients, etc. When using medium pixel- resolution image, optical image based extraction is far from satisfactory because of the misunderstanding between road area and other spectrally similar terrains. In practice, we recognised road area that consists of at least one pixel. Additionally, with an increase in the number of high-resolution SAR satellites, such as the Advanced Land Observing Satellite (ALOS), the Advanced Synthetic Aperture Radar (ASAR), Sentinel-1, and Gaofen (GF), high-resolution road extraction using SAR image (HRRES) is more accessible and can be applied at a lower cost or even free.

There have progressively been more studies being published on SAR image information extraction (Du & Lee, 1996; Niu & Ban, 2013; Saatchi & Moghaddam, 2000; Yu & Acton, 2004). Common methods, such as random forest (RF)(Du, Samat, Waske, Liu, & Li, 2015), support vector machines (SVM)(Dumitru & Datcu, 2013), and information entropy (IE) (Kussul, Skakun, Shelestov, & Kussul, 2014) have been proven to perform acceptably. All of these extraction

CONTACT Li Li lilixch@cau.edu.cn China Agricultural University, East Campus, Qinghua P.O. Box 62, East Road No. 17, Haidian District, Beijing 100083, PR China

techniques use the feature-based approach. Feature extraction is a key stage in object extraction, but it still requires considerable development. Various methods have been proposed for image processing frameworks whereby property-oriented or structural features are selected manually. For SAR image classification, many features have been introduced to date. For example, Markov random fields (MRF) extension features (Voisin, Krylov, Moser, Serpico, & Zerubia, 2013) have been proposed to extract objects in SAR images. A gray-level co-occurrence matrix (GLCM) features based on deep learning model was developed (Geng et al., 2015). Spatial features, such as local indicators of spatial association (LISA) region covariance (RC), can also be applied to extract humanistic environmental information(Stasolla & Gamba, 2008). However, as typical linear objects, road features are difficult to extract because of their similarity in backscatter characteristics to other linear objects, such as rivers, highways, railroads, bridges, etc. Thus, the extraction of features is critical to road detection. Nevertheless, it remains a challenge to extract road features in SAR images both automatically and on a large scale. This is because accuracy is not dependent only on image resolution, but is also affected by the combination of limited features and labels. In addition, the increased computational complexity and scenario types (i.e. platform, sensor, and task) may make it more difficult for those who have less experience using SAR.

Recently, the advancement in deep learning and computer vision (Erhan, Szegedy, Toshev, & Anguelov, 2014; LeCun, Bengio, & Hinton, 2015; Ren, He, Girshick, & Sun, 2015) has set off a wave of object recognition techniques. Being a time-saving method, deep learning can implicitly learn features and generate abstractions from inputs, mirroring what humans trained in specialized fields do, which does not require the end- user to design a proper set of features for a specific field, making its application relatively easy to road extraction. The deep neural network (DNN), a popular deep learning application (Sghaier, Foucher, & Lepage, 2017; Xu, Wang, Zhang, Li, & Zhang, 2017; Zhou, Wang, Xu, & Jin, 2016), exhibits particularly high efficiency in SAR image classification. Feature extraction processes have been integrated into DNN. Among these, convolutional neural networks (CNN) (LeCun et al., 2015) and fully convolutional neural networks (FCN)(Long, Shelhamer, & Darrell, 2015; Wu, Gao, Li, Xue, & Li, 2018), including their variations, are powerful tools for image classification. Ongoing efforts in addressing problems of object identification in DNN include those changes currently (Xu et al., 2017). The performance assessment of the classic CNN model on data retrieved from the TerraSAR-X satellite, which is a type of high-resolution commercial SAR product, reported an accuracy rate of greater than 97% for building extraction. Although this accuracy metric differs from the pixel-wise metric commonly used for semantic segmentation, it can still be seen as a high accuracy method. On the other hand, there is almost no literature on road extraction. Most previous studies have been based on morphological operations, accompanied by a threshold to improve result accuracy. Currently, the neural network is not yet intelligent enough to reliably segment road information, particularly for SAR images of lower resolutions. A novel cascaded end-to-end convolutional neural network (CasNet) (Cheng et al., 2017) is used to identify roads, obtaining good results. However, this method only applies to very high resolution (VHR) satellite remote sensing imagery. Additionally, CasNet cannot effectually detect roads for continuous and large areas of occlusions. To overcome poor localization for object classification, the probabilistic graphical models are incorporated in the form of fully conditional random fields (CRF) (Chen, Papandreou, Kokkinos, Murphy, & Yuille, 2018).

The Sentinel-1 SAR satellite provides a cost-effective alternative to obtaining high-resolution data, free of cost, for road extraction. However, Sentinel-1 SAR images provide no road labels themselves. Although there are many public datasets available that provide optical images, a standard road extraction dataset from SAR images is still urgently needed. Fortunately, FCNs can extract objects using fewer samples, which makes the standard dataset accessible. Thus, this study proposes a universal classification method based on FCNs. To make this case, we first extracted road information from dual-polarimetric Sentinel-1 SAR images based on FCN and then compared its performance to CNN results. Extensive validations from direct observations (ground truth) have been used to verify the accuracy and efficiency of this proposed method. The objective of this study is to evaluate the potential of FCN in extracting road information from Sentinel-1 SAR images. The remaining article is organized as follows: The proposed methodology is described in Chapter 2; results and discussion are provided in Chapter 3; Our conclusions are provided in Chapter 4.

Proposed methodology

Figure 1 provides a schematic flowchart of the proposed method which we used to extract road networks from Sentinel-1 SAR images. This proposed method is divided into four parts: Sentinel-1 SAR pre-processing and sample preparation, model training and, training optimization, extraction based on the pre-trained model, and model validation and comparison.

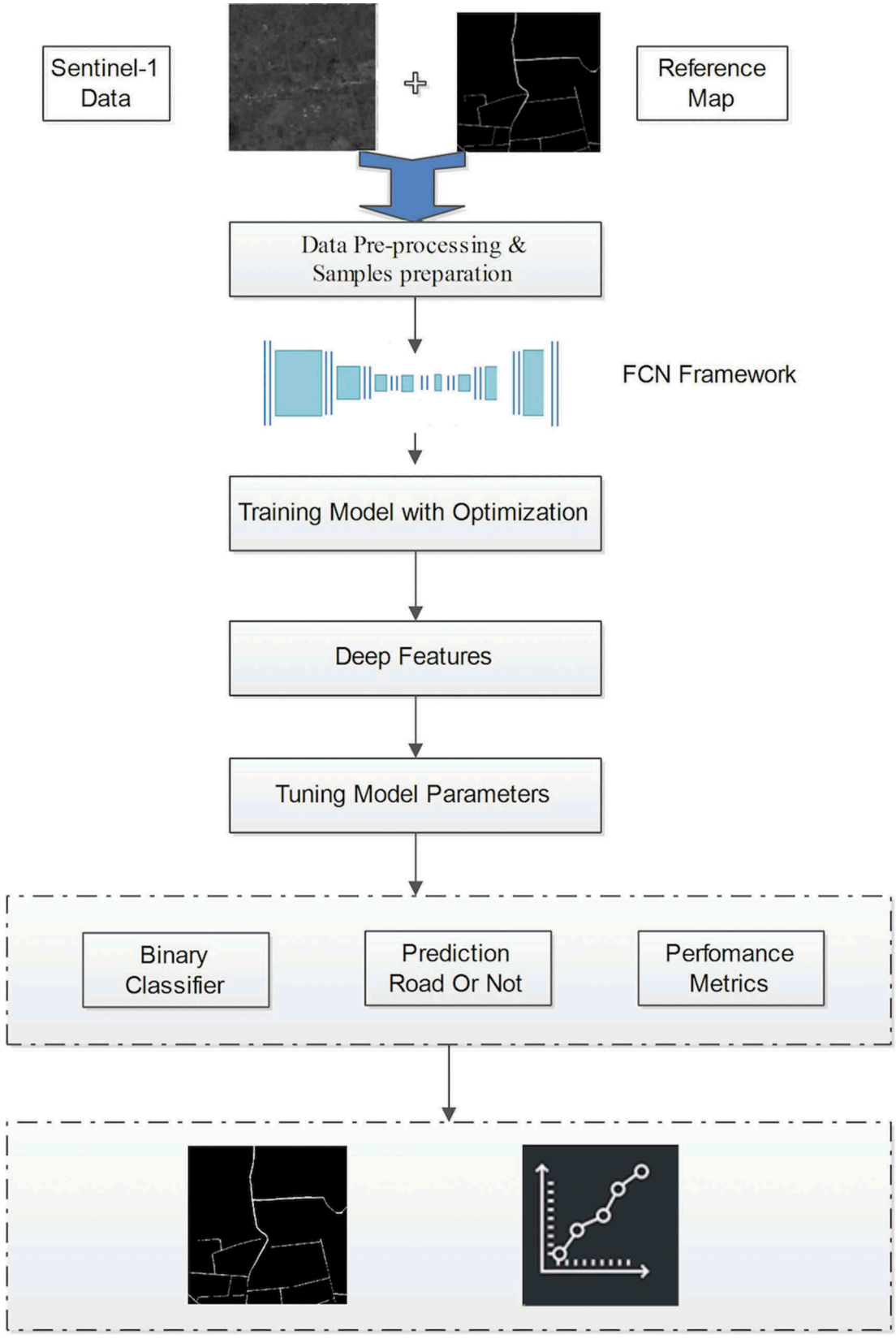

Figure 1. Flowchart of the extraction of road using the proposed FCN.

Sentinel-1 SAR image pre-processing and sample preparation

In this study, we used one piece of dual-polarimetric (VV and VH) Ground Range Detected (GRD) C-band SAR image obtained from Sentinel-1 sensors using the interferometric wide (IW) swath mode. Sentinel-1, launched by the European Space Agency, is a constellation with two C-Band SAR satellites and has a 12-day revisit time for each satellite (Torres et al. 2012). Using Interferometric Wide (IW) swath

acquisition mode, the geometric resolution is 20 m × 22 m and a pixel space range of 10 m × 10 m with a swath width of 250 km. We selected an area between 39.76–39.71° N and 116.38–116.44° E, covered a small area in Beijing, China, with complex road networks. The scene of the image was taken in descending orbital mode on 15 April 2017, and can be obtained from Copernicus Open Access Hub (https://scihub.coperni cus.eu/dhus/#/home). The SAR image was first pre-processed with the following steps (Geudtner, Torres, Snoeij, Davidson, & Rommen, 2014): radiation correction, speckle filter, terrain correction, and decibel conversion. For speckle filter, Refined Lee filter (Lee, Grunes, & Grandi, 1999) was selected after many attempts and the window size was set to 7*7 pixels. For terrain correction, a 90-m digital elevation model (DEM) in the form of Shuttle Radar Topography Mission (SRTM) DEMs was used in view of the relatively flat topography of the study area. For decibel conversion, the radar backscattering coefficients (σ° also known as sigma0 or sigma nought) in decibels (dB) were obtained from digital number (DN) values using the following equation:

$$\sigma^0 = log_{10}\left[DN^2/A^2\right] \quad (1)$$

where A is the calibration factor and can be looked up from sigma naught values in the lookup table (LUT).

According to rules of manual interpretation of SAR image overlaid by Google optical map of the corresponding period and road cover information in the field survey,we masked road areas in different regions online directly as a vector layer. The vector layer will be covert into a raster layer with the same pixel space with Sentinel-1. According deep learning theory, the size of each sampled tile affects the result. In this study, we used a common size of 256 × 256 (pixel). In this way, the labelled dataset of the study area was cropped into 1330 smaller tiles, and the best 500 tiles were selected as samples. Among these, 300 images were randomly selected as training samples while the other 200 images were preserved as testing samples. The samples were converted into a general format. All sample dimensions remain the same, which allowed for their immediate usage in the deep learning model.

Road extraction using DNN

The core of high-performance classification methods is the selection of a suitable DNN model. In this study, we used an extended CNN model – FCN, which has been proven to be able to extract cells in biomedical images (Ronneberger, Fischer, & Brox, 2015), adjusted for road extraction from Sentinel-1 SAR data.

CNN architecture

We first applied the DNN model of CNN architecture to the SAR data. The CNN architecture used in this study was inspired by the traditional LeNet-5, which is good at classifying digits in images. In this study, we kept one fully connected layer to perform road extraction. The CNN designed for SAR road extraction was obtained by stacking two convolution layers interleaved with two max-pooling layers as well as one fully connected layer that was replaced by the convolution operation in FCN. Finally, the road is classified by a sigmoid classifier. Figure 2 illustrates the CNN architecture used.

In Figure 2, C represents the convolution functions, while S refers to the down-sampling operation, which means that C1 and C3 are both convolutions, while S2 and S4 perform the down-sampling. F5 is the only fully connected layer. Sig6 represents the sigmoid function used. The convolutional layers compute the convolution of the input images with a bank of filters to obtain local features. A suitable convolution kernel size must be defined due to its effect on extraction performance. Effective information cannot be extracted using too small a kernel size; however, a large kernel size could result from overfitting errors, which makes training difficult. In this study, we adjusted the kernel size according to the different tasks. We also used the dropout strategy in the CNN structure to avoid overfitting. Moreover, rectified linear units (ReLU) were selected as an activation function in the fully connected layer. The hyperparameter was part tuned as described by an earlier study(Xu et al., 2017). The C1 layer includes four kernels while the C3 includes eight kernels. For both layers, we selected 5 × 5 (unit: pixel) as the kernel size. For the S2 and S4 layers, the 3 × 3 (unit: pixel) window was selected for max pooling. In this study, max pooling

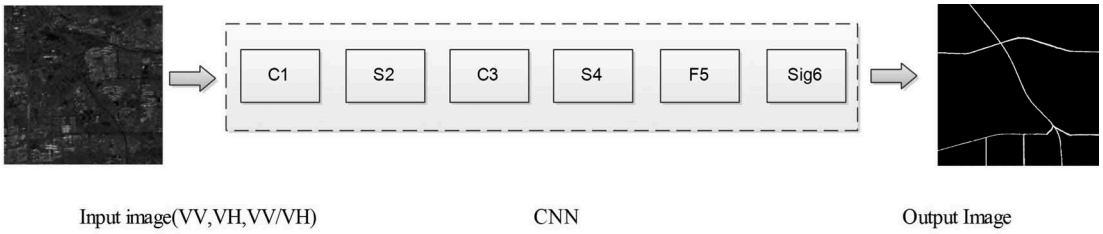

Figure 2. The architecture of CNN.

was used as a sub-sampling procedure to eliminate the deviation and distortion of images. The F5 layer contained 128 neurons. Sigmoid functions are often used as threshold functions in neural networks due to their singular and monotonic nature, which allows variables to be mapped between 0 and 1. In this study, the Sig6 layer is a sigmoid function defined by Equation (2):

$$S(x) = 1/\left(1 + e^{-1}\right) \quad (2)$$

The proposed FCN

FCN is an extension of CNN, developed by replacing a convolutional layer with a fully connected layer (discussed in section 2.2.1). To bypass the input size bottleneck of CNN, a fully connected layer is used to replace convolution operations (globally). FCN is widely considered the best at performing pixel-wise road extraction. This study used FCN architecture based on recent investigations of U-Net (Xu et al., 2017). The unit of CNN and FCN in this section is pixel. Being a more complex implementation of FCN architecture, U-Net adds up-sampling layers to reconstruct the image.

Through a selection of the multiple tests conducted, the convolutional layers of the original U-Net were reduced, designing a simpler U-Net structure that has a higher road extraction accuracy. See Figure 3 for a pictorial representation of the architecture.

In Figure 3, dark blue arrows represent convolution and activation functions; grey arrows represent replication; red arrows represent down-sampling; green arrows represent up-sampling followed by convolution; conv 1 × 1 represents a convolution operation with a core of 1 × 1. All unit of measure in CNN and FCN is pixel. Using only convolution and down-sampling, which also represents an end-to-end image architecture, our proposed FCN includes 20 convolution layers throughout the entire network, namely, four down-sampling steps and four up-sampling steps. The FCN network is connected by a convolutional layer instead of a fully connected layer. The detailed layer shape is shown in Table 1.

Optimization technique

Although deep learning technology has achieved great success and dominance in various fields, it is still susceptible to shortcomings, such as local minima (Srivastava, Hinton, Krizhevsky, Sutskever, & Salakhutdinov, 2014; Tirumala, 2014). Therefore, appropriate optimization techniques should be used to both avoid such shortcomings and achieve ideal results. In this study, we accessed four optimization methods for the proposed FCN, namely, stochastic gradient descent (SGD), Momentum, root mean square propagation (RMSProp), and adaptive moment estimation (Adam).

Being one of the basic optimization algorithms, SGD has been used as the primary algorithm for the success of deep learning (Bengio, 2009). To optimize iteration times, SGD selects a mean gradient as a cost function. Assuming m small batch (that are independent and identically distributed) samples are used for

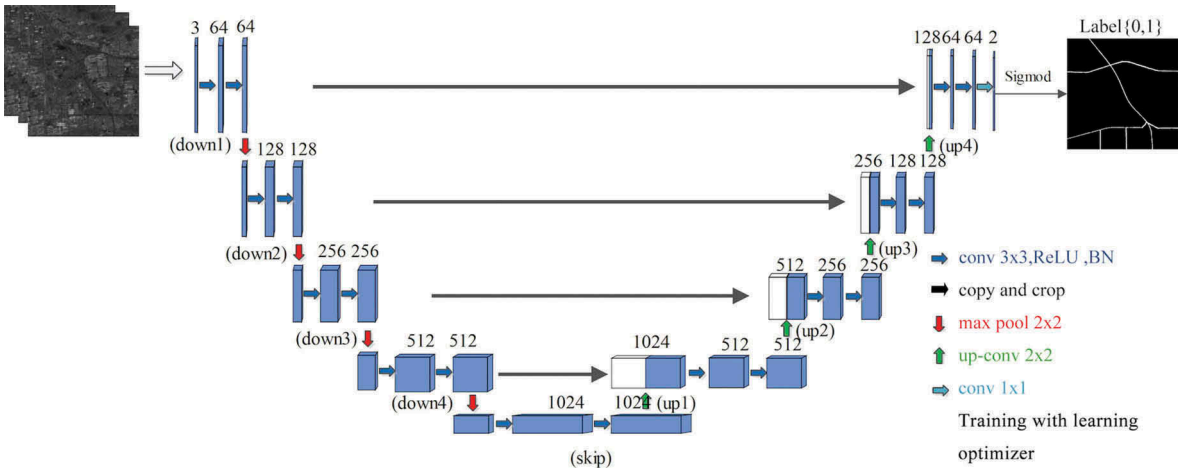

Figure 3. Illustration of how the FCN(U-Net) is used to extract road automatedly.

Table 1. Parameters of the proposed FCN.

Layer Name	Feature Size	Kernel	Layer Name	Feature Size	Kernel
input	−2,56,256	-	pooling4	(16,16)	2 × 2
Conv2d (1)	−1,28,128	(16,16), stride = 2	skip_connection	(16,16)	2 × 2
pooling1	−1,28,128	(3,3), maxpool, stride = 2	Up-sampling1	(32,32)	2 × 2
Conv2d (2)	(64,64)	1 × 1	Up-sampling2	(64,64)	2 × 2
pooling2	(64,64)	2 × 2 average pool	Up-sampling3	−1,28,128	2 × 2
Conv2d (3)	(32,32)	1 × 1	Up-sampling4	−2,56,256	1 × 1
pooling3	(32,32)	2 × 2 average pool	Conv2d (5)	(256,1,1)	256,1 × 1
Conv2d (4)	(16,16)	1 × 1	output		-

iteration, the mean gradient is defined as the mean gradient of the m batches. As shown in Equation (3), this algorithm calculates the mini-batch gradient for each iteration, then updates the parameters.

$$g_t = \nabla_{\theta_{t-1}} f(\theta_{t-1}) \text{ , where } \Delta\theta_t = -\eta \times g_t \quad (3)$$

In this study, η is the learning rate; t is the iteration time; g_t is the gradient; and SGD relies entirely on the gradient of the current batch.

Momentum uses the concept of momentum in physics simulations as a reference, and it accumulates the previous momentum instead of a true gradient (Sutskever, Martens, Dahl, & Hinton, 2013). The momentum for information updates is given as follows:

$$v_{t+1} = \mu v_t - \varepsilon \nabla f(\theta_t + \mu v_t) \text{ , where } \theta_{t+1} = \theta + v_{t+1} \quad (4)$$

In this study, v_t is the current velocity vector that has the same dimension as the parameter vector θ. The learning rate ε is similar to η (described above). Using momentum ε may be smaller than η for larger magnitude values of the gradient. Finally, $\mu \in (0,1)$ determines how many iterations the previous gradients have incorporated into the current update.

Similar to Momentum, RMSProp allows the cumulative gradient to decay by a fraction at every iteration by introducing an attenuation factor. Adam is essentially an RMSProp as it reflects Momentum, dynamically adjusting the learning rate of each parameter using first and second order moment estimates of the gradient. The main advantage of Adam is that the learning rate of each iteration has a certain range with the offset correction, which makes parameters relatively stable.

Model training

The proposed FCN was implemented using the TensorFlow open source library (Abadi et al., 2016). The library provides a user-friendly API for the Python language. Moreover, we adopted the mini-batch SGD with GPU acceleration. For road extraction, Huber loss was calculated pixel-wise using different weights for each pixel class according to Equation (5). To avoid overfitting, parameters stop updating when loss stops decreasing,

$$\text{loss}(x,y) = \frac{1}{N} \begin{cases} \frac{1}{2}(x_i - y_i), \text{if} |x_i - y_i| < 1 \\ |x_i - y_i| - \frac{1}{2}, \text{othewise} \end{cases} \quad (5)$$

where N is the dimension of observation y_i and variable x_i. To balance the learning process, we used the gradient and learning rate. The learning rate was initialized at 0.00005, and four different optimizers (SGD, Momentum, RMSProp, and Adam) were used to accelerate convergence for comparison.

Error metric selection during the test phase

At present, there are no unified evaluation standard for road extraction (Wang et al., 2016). Precision (correctness) and Recall (completeness) (Powers, 2011) being the most common evaluation metrics in deep learning-based classifications. Thus, Precision and Recall are selected as the evaluation criteria for this road extraction accuracy assessment, defined as follows:

$$\text{Precision} = TP/(TP + FP) \quad (6)$$

$$\text{Recall} = TP/(TP + FN) \quad (7)$$

where *TP*, *FP*, and *FN* represent true positive, false positive, and false negative of road extraction, respectively.

From Equation 6 and 7, the precision is a ratio that describes the proportion of correct or incorrect predictions made by the model. Technically, this method is called the Bernoulli trial. Fortunately, for large sample sizes (e.g. $n > 30$), the Gaussian distribution can be used to approximate. The confidence interval could be easily calculate using the Gaussian distribution hypothesis of the extraction precision.

$$\text{CI} = \text{z} \times \sqrt{\frac{\text{Precision} \times (1 - \text{Precision})}{\text{n}}} \quad (8)$$

where CI is the confidence interval, z is a significant level of 95% of the Gaussian distribution corresponding to a critical value of 1.96. In our experiment, n takes the value 200.

When these two indicators (*Precision* and *Recall*) conflict with each other, a comparison between the different models is difficult. Thus, another parameter (F_1 score) was introduced, which is an indicator used in statistics to measure the accuracy of dichotomous models. The F_1 score can be considered as a combination of Precision and Recall with a maximum of 1 and a minimum of 0. Therefore, the F_1 score was also used to measure accuracy, which is calculated as follows:

$$F_1 = (2 * \text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (9)$$

Results and analysis

In this section, the detailed performance evaluation of the two parallel schemes for HRRES. These two individual results and the final fused dual-polarimetric results are also compared.

Optimization algorithm selection for training strategy

To select a suitable optimization method, four optimization algorithms in the proposed FCN were used. Figure 4 provides the loss curves of the four optimizers.

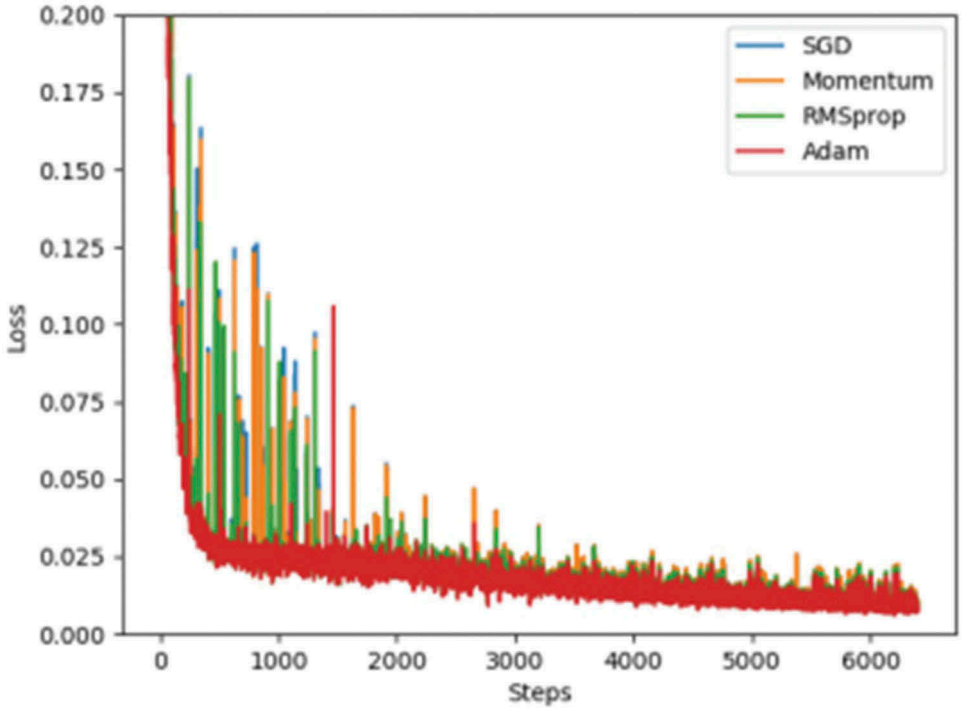

Figure 4. Training curves using different optimizers.

Among the four optimization methods, the Adam algorithm outperformed the others in computational efficiency and convergence rates. Thus, the Adam optimizer was selected for the training procedure.

First scheme: HRRES with labelled single-polarization SAR

CNN and FCN are used to test the road extraction accuracies of different SAR band. Table 2 provides results from two polarization bands.

Results from FCN and CNN architecture showed that VV mode outperformed VH mode. For the FCN road extraction results, the F_1 score for VV polarization was greater than 90%, while identification results under VH polarization was less than 90%. Results showed that co-polarization (VV) Sentinel-1 SAR images were more suitable than cross-polarization Sentinel-1 SAR images (VH) for road extraction. This is primarily due to the higher energy required for co-polarization SAR systems compared to cross-polarization SAR systems (Van Zyl, Zebker, & Elachi, 1987). The extracted roads are shown in Figure 5. The research area combines both urban and rural areas where both city and township roads exist as well as dense vegetation, namely, a complex road network that makes road information difficult to extract. The red labels on the map are used to make the contrast more obvious. It can be found that the highlight parts are quite different. As expected, single-polarization SAR images for road extraction, the performance of the proposed FCN architecture exceeded CNN.

Second scheme: HRRES by with labelled dual-polarization SAR

For Sentinel-1, we defined dual-polarization SAR band as VH/VV. The accuracy was shown in Table 4. According to Table 3, the Precision index of VV polarization was only 3% higher than that of dual polarization, and the other indexes under dual-polarization were all higher than those of single polarization by at least 2% to 5%. Therefore, we concluded that the use of dual-polarization Sentinel-1 SAR images was better than single-polarization Sentinel-1 SAR images for road extraction. Figure 6 showed the road maps generated using FCN and CNN as well as the ground truth. In Figure 6, some roads that were not detected by the CNN model were detected by the FCN model. Moreover, some gaps in roads have now been filled, which provides superior complexity of the road network in this combined urban and rural area. Additionally, road network monitoring results are smooth and clear.

Table 2. Accuracy and confidence interval of the two DNN models of the first scheme.

	Precision ± CI		Recall		F_1	
Name	VH	VV	VH	VV	VH	VV
CNN	70.32%±0.39%	**78.46%**±0.36%	72.43%	79.21%	71.36%	78.83%
FCN	89.13%±0.29%	90.46%±0.27%	88.34%	91.11%	88.73%	90.78%

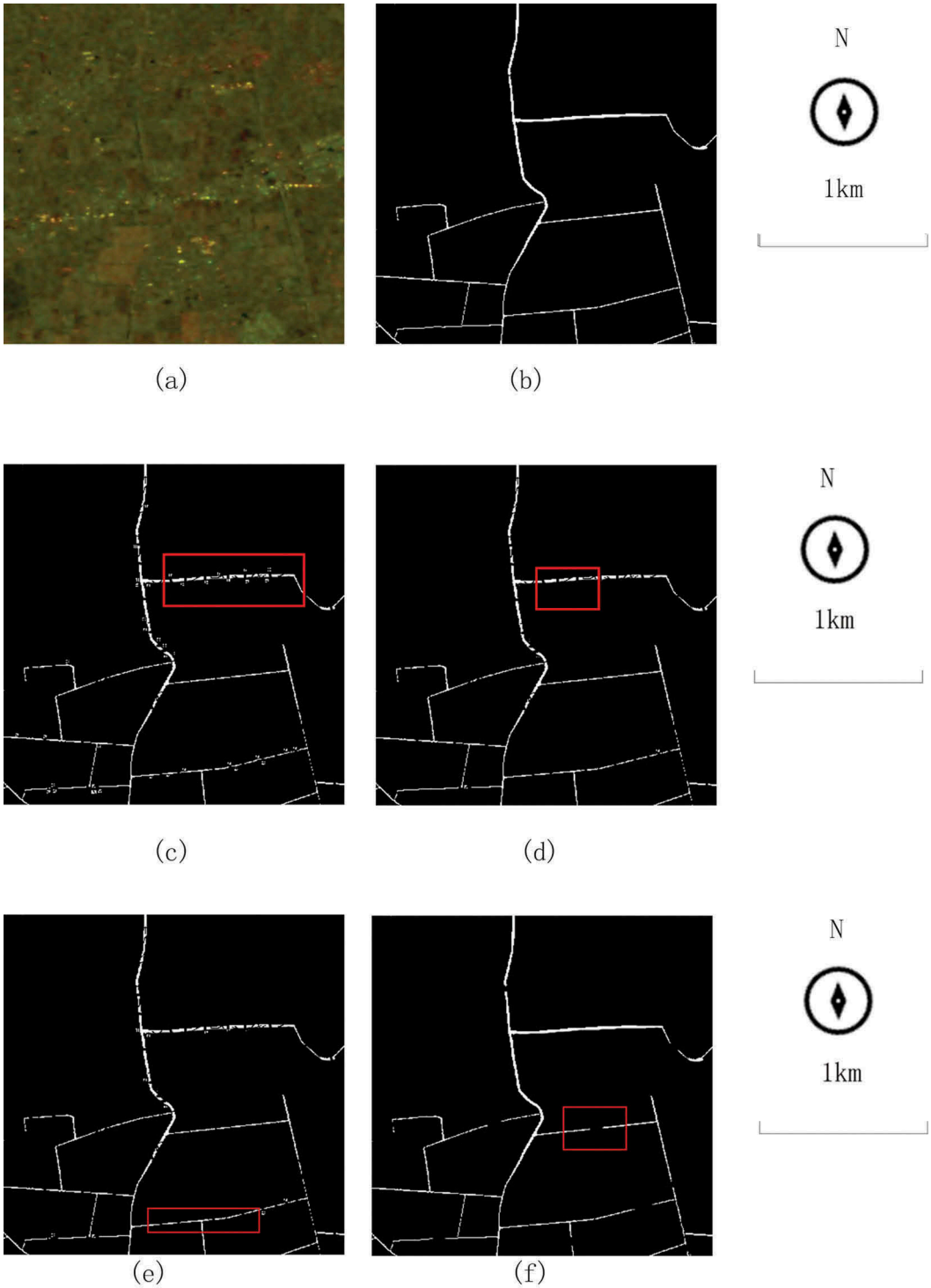

Figure 5. Extracted road map for single- polarization and dual-polarization SAR using CNN and FCN.

(a)Sentinel1 image of selected patch R:VV;G:VH;B:VV/VH. (b) Ground truths of road from Google Earth. (c)Roads extracted from VV mode using CNN. (d) Roads extracted from VV mode using FCN. (e)Roads extracted from VH mode using CNN. (f)Roads extracted from VH mode using FCN

Table 3. Accuracy and confidence interval of the two DNN models of the second scheme.

DNN Name	Precision ± CI	Recall	F_1
CNN	**75.32%**±0.37%	83.81%	79.34%
FCN	92.13%±0.25%	95.17%	93.63%

Table 4. Execution times for training models of two DNN models (ms).

Dataset Size (Train/Total*100%)	10	30	50	70	90	100
CNN	10.34	19.6	26.22%	30.22	39.43	56.22
FCN	9.62	15.54	21.32	25.09	30.09	36.85

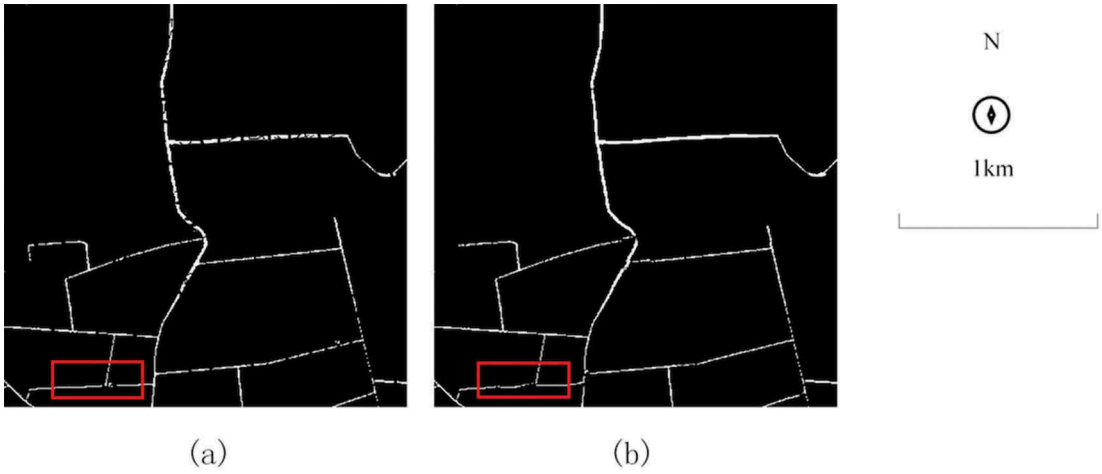

Figure 6. Extracted road map for dual-polarization SAR VV/VH using CNN(a) and FCN(b).

Table 5. Accuracy and confidence interval of the four machine methods using different polarization.

	Precision±CI		Recall		F_1	
Name	VH	VV	VH	VV	VH	VV
SVM	70.32%±6.33%	78.45%±5.70%	72.43%	78.98%	71.36%	78.71%
KNN	69.28%±6.39%	73.82%±6.09%	76.62%	84.21%	72.77%	78.67%
DT	70.13%±6.19%	72.52%±6.19%	76.38%	89.45%	73.12%	80.10%
RF	79.13%±4.43%	88.46%±4.43%	88.34%	90.02%	83.48%	89.23%

Comparison with state-of-the-art

A comparison of computing time between two DNN models

To quantitatively evaluate the data computing efficiency of the proposed FCN method (compared to CNN), these two deep learning networks were evaluated with different sizes of the training datasets. The final result takes the average time of 10 times.

VV band dataset was taken as test input. The experiments were conducted using Python 3.5.2 on a server equipped with an Intel Xeon CPU (12 cores and 64 GB memory) and an NVIDIA Tesla K40 GPU. Table 4 provides the computing time evaluation results of these two road extraction methods.

Table 4 clearly shows that the FCN is much faster to CNN in training time especially for larger dataset.

Comparison with other methods

The FCN is seldom used in road extraction using Sentinel-1 SAR data. However, we compare the proposal with machine learning methods based on our data with different polarizations. The accuracy of the different polarization for each methods can be summaried in Table 5.

Computing time for model training was also evaluated using the VV band in Table 6. For fair comparison, total dataset was used as input. The two DNN modes take a long time for training, while the accuracy is promising. But DNN can achieve a generally fitted model than machine learning methods. Different methods are suitable for different application purposes. Based on the optimization model, we generated regional road map using FCN. Figure.7 shows the result.

Conclusions

In this study, FCN(U-Net), a novel DNN for pixel-level road extraction was proposed. To verify the performance of the method, a set of dual-polarimetric SAR images from Sentinel-1 product were collected, and manually generated ground truth data (feature identification and labelling) with the help of Google Maps. Optimization techniques are also used in the two DNNs process. For comparison, CNN, the basic DNN, along with four other machine learning methods were used to verify the performance of the collected SAR data. The experiment results demonstrated that the proposed network (FCN) significantly outperformed traditional machine learning methods in accuracy and less computational time than CNN.

In the future, we intend to introduce the transfer learning model to our architecture to further improve accuracy. Additionally, we will explore the performance and sensitivity in different topographic areas and grouping other images.

Table 6. Execution times for training total data set of different machine learning methods(ms).

Method	SVM	KNN	DT	RF
Computing Time	23.15	24.89	27.58	22.78

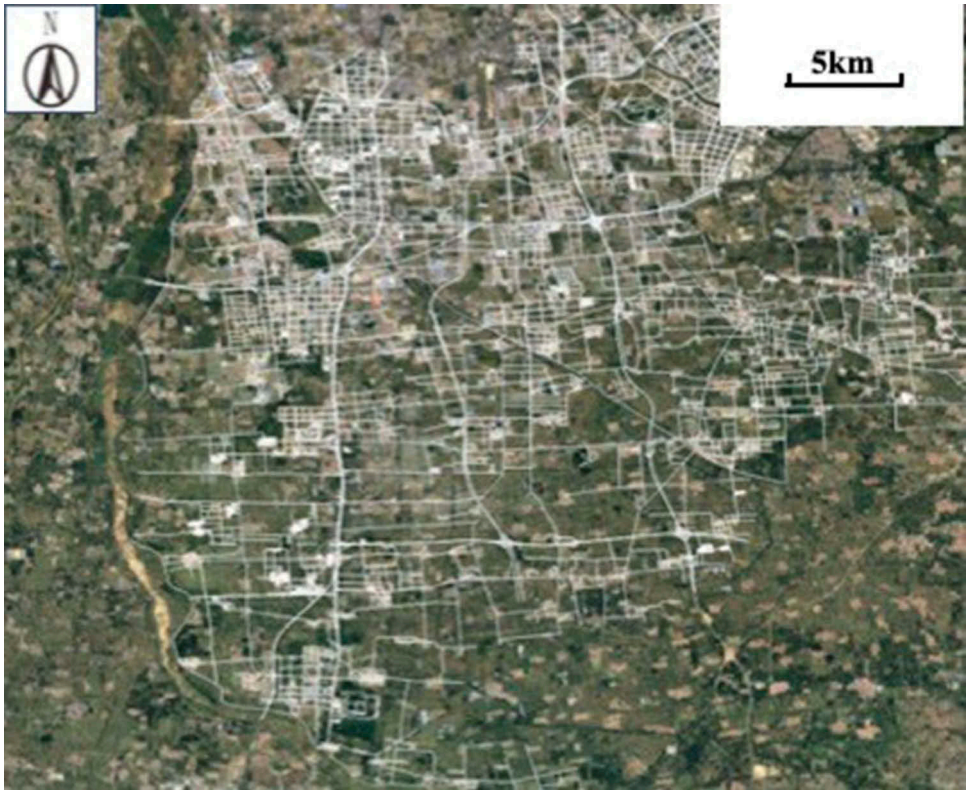

Figure 7. The final road map (white) overlay google satellite map in Daxing district.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This study was supported by the National Key R&D Program of China [2017YFB0502700].

ORCID

Qianqian Zhang http://orcid.org/0000-0003-2825-6238

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., … Isard, M. (2016). *Tensorflow: A system for large-scale machine learning*. Paper presented at the OSDI, Savannah, GA, USA.

Bengio, Y. (2009). Learning deep architectures for AI. *Foundations & Trends® in Machine Learning*, *2*(1), 1–127. doi:10.1561/2200000006

Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A.L. (2018). DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, *40*(4), 834–848. doi:10.1109/TPAMI.2017.2699184

Cheng, G., Wang, Y., Xu, S., Wang, H., Xiang, S., & Pan, C. (2017). Automatic road detection and centerline extraction via cascaded end-to-end convolutional neural network. *IEEE Transactions on Geoscience and Remote Sensing*, *55*(6), 3322–3337. doi:10.1109/TGRS.2017.2669341

Du, L., & Lee, J.S. (1996). Fuzzy classification of earth terrain covers using complex polarimetric SAR data. *International Journal of Remote Sensing*, *17*(4), 809–826. doi:10.1080/01431169608949047

Du, P., Samat, A., Waske, B., Liu, S., & Li, Z. (2015). Random forest and rotation forest for fully polarized SAR image classification using polarimetric and spatial features. *ISPRS Journal of Photogrammetry and Remote Sensing*, *105*, 38–53. doi:10.1016/j.isprsjprs.2015.03.002

Dumitru, C.O., & Datcu, M. (2013). Information content of very high resolution SAR images: Study of feature extraction and imaging parameters. *IEEE Transactions on Geoscience and Remote Sensing*, *51*(8), 4591–4610. doi:10.1109/TGRS.2013.2265413

Erhan, D., Szegedy, C., Toshev, A., & Anguelov, D. (2014). *Scalable object detection using deep neural networks*. Paper presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA.

Geng, J., Fan, J., Wang, H., Ma, X., Li, B., & Chen, F. (2015). High-resolution SAR image classification via deep convolutional autoencoders. *IEEE Geoscience and Remote Sensing Letters*, *12*(11), 2351–2355. doi:10.1109/LGRS.2015.2478256

Geudtner, D., Torres, R., Snoeij, P., Davidson, M., & Rommen, B. (2014). *Sentinel-1 System capabilities and applications*. Paper presented at the Geoscience and Remote Sensing Symposium, Quebec City, QC, Canada.

Kussul, N., Skakun, S., Shelestov, A., & Kussul, O. (2014). *The use of satellite SAR imagery to crop classification in Ukraine within JECAM project*. Paper presented at the Geoscience and Remote Sensing Symposium (IGARSS), 2014 IEEE International, Quebec City, QC, Canada.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. doi:10.1038/nature14539

Lee, J.S., Grunes, M.R., & Grandi, G.D. (1999). Polarimetric SAR speckle filtering and its implication for classification[J]. *IEEE Transactions on Geoscience & Remote Sensing*, *37*(5), 2363–2373.

Long, H., & Zhao, Z. (2005). Urban road extraction from high-resolution optical satellite images. *International Journal of Remote Sensing*, *26*(22), 4907–4921. doi:10.1080/01431160500258966

Long, J., Shelhamer, E., & Darrell, T. (2015). *Fully convolutional networks for semantic segmentation*. Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition, Boston, MA, USA.

Miao, Z., Shi, W., Samat, A., Lisini, G., & Gamba, P. (2016). Information fusion for urban road extraction from VHR optical satellite images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *9*, 1817–1829. doi:10.1109/JSTARS.2015.2498663

Naouai, M., Hamouda, A., & Weber, C. (2010). *Urban road extraction from high-resolution optical satellite images*. Paper presented at the International Conference Image Analysis and Recognition, Póvoa de Varzim, Portugal.

Niu, X., & Ban, Y. (2013). Multi-temporal RADARSAT-2 polarimetric SAR data for urban land-cover classification using an object-based support vector machine and a rule-based approach. *International Journal of Remote Sensing*, *34*(1), 1–26. doi:10.1080/01431161.2012.700133

Powers, D.M.W. (2011). Evaluation: From precision, recall and F-Factor to ROC, informedness, markedness & correlation. *Journal of Machine Learning Technologies*, *2*, 2229–3981.

Ren, S., He, K., Girshick, R., & Sun, J. (2015). *Faster r-cnn: Towards real-time object detection with region proposal networks*. Paper presented at the Advances in neural information processing systems, Montreal, Quebec, Canada.

Ronneberger, O., Fischer, P., & Brox, T. (2015). *U-Net: Convolutional networks for biomedical image segmentation*. Paper presented at the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany.

Saatchi, S.S., & Moghaddam, M. (2000). Estimation of crown and stem water content and biomass of boreal forest using polarimetric SAR imagery. *IEEE Transactions on Geoscience and Remote Sensing*, *38*(2), 697–709. doi:10.1109/36.841999

Sghaier, M.O., Foucher, S., & Lepage, R. (2017). River extraction from high-resolution sar images combining a structural feature set and mathematical morphology. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *10*(3), 1025–1038. doi:10.1109/JSTARS.2016.2609804

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, *15*(1), 1929–1958. doi:10.1214/12-AOS1000

Stasolla, M., & Gamba, P. (2008). Spatial indexes for the extraction of formal and informal human settlements from high-resolution SAR images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *1*(2), 98–106. doi:10.1109/JSTARS.2008.921099

Sutskever, I., Martens, J., Dahl, G., & Hinton, G. (2013). *On the importance of initialization and momentum in deep learning*. Paper presented at the International Conference on International Conference on Machine Learning, Atlanta, GA,USA.

Tirumala, S.S. (2014). *Implementation of evolutionary algorithms for deep architectures*. Paper presented at the Aic 2014 - International Workshop on Artificial Intelligence and Cognition, Torino, Italy.

Torres, R., Snoeij, P., Geudtner, D., Bibby, D., Davidson, M., & Attema, E., ... & Traver, I. *N*. (2012). GMES Sentinel-1 mission. Remote Sensing of Environment, 120, 9–24

Van Zyl, J.J., Zebker, H.A., & Elachi, C. (1987). Imaging radar polarization signatures: Theory and observation. *Radio Science*, *22*(04), 529–543.

Voisin, A., Krylov, V.A., Moser, G., Serpico, S.B., & Zerubia, J. (2013). Classification of very high resolution SAR images of urban areas using copulas and texture in a hierarchical Markov random field model. *IEEE Geoscience and Remote Sensing Letters*, *10*(1), 96–100. doi:10.1109/LGRS.2012.2193869

Wang, W., Yang, N., Zhang, Y., Wang, F., Cao, T., & Eklund, P. (2016). A review of road extraction from remote sensing images. *Journal of Traffic and Transportation Engineering (English edition)*, *3*(3), 271–282. doi:10.1016/j.jtte.2016.05.005

Wu, Z., Gao, Y., Li, L., Xue, J., & Li, Y. (2018). Semantic segmentation of high-resolution remote sensing images using fully convolutional network with adaptive threshold. *Connection Science*, 1–16. doi:10.1080/09540091.2018.1510902

Xu, Z., Wang, R., Zhang, H., Li, N., & Zhang, L. (2017). Building extraction from high-resolution SAR imagery based on deep neural networks. *Remote Sensing Letters*, *8*(9), 888–896. doi:10.1080/2150704X.2017.1335906

Yu, Y., & Acton, S. (2004). Automated delineation of coastline from polarimetric SAR imagery. *International Journal of Remote Sensing*, *25*(17), 3423–3438. doi:10.1080/0143116032000160444

Zhou, Y., Wang, H., Xu, F., & Jin, Y.-Q. (2016). Polarimetric SAR image classification using deep convolutional neural networks. *IEEE Geoscience and Remote Sensing Letters*, *13*(12), 1935–1939. doi:10.1109/LGRS.2016.2618840