International Journal of Applied Earth Observations and Geoinformation 103 (2021) 102498

Contents lists available at ScienceDirect

International Journal of Applied Earth Observations and Geoinformation

journal homepage: www.elsevier.com/locate/jag

Leveraging optical and SAR data with a UU-Net for large-scale road extraction

Yinyi Lin [a], Luoma Wan [a], Hongsheng Zhang [b,*], Shan Wei [b], Peifeng Ma [a,c,*], Yu Li [d], Zhuoyi Zhao [a]

[a] Institute of Space and Earth Information Science, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong
[b] Department of Geography, The University of Hong Kong, Pokfulam, Hong Kong
[c] Shenzhen Research Institute, The Chinese University of Hong Kong, Shenzhen, China
[d] Faculty of Information Technology, Beijing University of Technology, Beijing, China

ARTICLE INFO

Keywords:
Road
Optical
SAR
OSM
U-Net

ABSTRACT

Road datasets are fundamental and imperative for traffic management and urban planning. Different high-resolution optical remote sensing images are widely used for automatic road extraction but the results are usually limited to local scale and spectral confusions in barren and cropland, while accurate large-scale road extraction remains challenging. In this study, we incorporated medium resolution optical and SAR data, i.e., 10-meter resolution Sentinel-1 and Sentinel-2, for road extraction at a large scale and evaluated the contribution of different data sources. We developed a United U-Net (UU-Net) to fuse optical and SAR data for road extraction, which was trained and evaluated on a large-scale multisource road extraction dataset. The UU-Net achieved better accuracy than traditional deep convolutional networks with optical or SAR data alone, which obtained an average F1 of 0.5502 and an average IoU of 0.4021, outperforming in 160 out of 200 (80%) 0.5-by-0.5 degree evaluation grids. The results indicated that SAR contributes more to road extraction in barren land, while optical data contributes more to large slope areas. The road accuracy is positively related to elevation and urban percentage, which distributes higher in eastern China and lower in western. The road centerline from 10 m road showed comparable results with that from Open Street Map (OSM), indicating its promising applications to support large-scale urban transportation studies.

1. Introduction

The road is a line that serves as the stabilized base for urban communication. Road network works as the indicator of urban sprawl and development (Zhao 2010), which is closely related to energy consumption (Chai et al. 2016) and environmental issues (Xie et al. 2017). According to Intergovernmental Panel on Climate Change (IPCC) report, the greenhouse gas (GHG) emissions from the transportation sector was doubled since 1970, which was responsible for about 23% of the total energy-related CO2 emission in 2010. The road vehicles accounted for 80% of the increased emissions and the construction of the road was responsible for the indirect GHG emissions. In addition, the transport infrastructure demand is closely linked with environmental factors like land use (Cervero 2013; Kasraian et al. 2016), air pollution (Sun et al. 2018; Zhao 2010), noise (Oltean-Dumbrava et al. 2013) and etc.

The existing large-scale road data sets include the global roads open access data set (gROADS) (Center for International Earth Science Information Network - CIESIN - Columbia University and Information Technology Outreach Services - ITOS - University of Georgia 2013) from 1980 to 2010 and open street map (OSM) highway data. gROADS has quality control management but the time is limited to the year of 2010, which requires road updating process. OSM highway data is up to date but has uncertainties from the crowdsourcing data and lacks of width information. Road extraction from remote sensing satellite images is efficient and fundamental for traffic management, road monitoring, city updating (Shi et al. 2014; Wang et al. 2016), which can provide reliable road information. The road extraction methods can be categorized into the following two parts: morphological method and segmentation method. The morphological method considers the geometric and topological features including the edge detection (Zeng et al. 2019), hough

* Corresponding authors.
E-mail addresses: yinyilin@link.cuhk.edu.hk (Y. Lin), roma@link.cuhk.edu.hk (L. Wan), zhanghs@hku.hk (H. Zhang), shanw@connect.hku.hk (S. Wei), mapeifeng@cuhk.edu.hk (P. Ma), yuli@bjut.edu.cn (Y. Li), zhao.zhuoyi@link.cuhk.edu.hk (Z. Zhao).

https://doi.org/10.1016/j.jag.2021.102498
Received 7 May 2021; Received in revised form 31 July 2021; Accepted 9 August 2021
Available online 19 August 2021

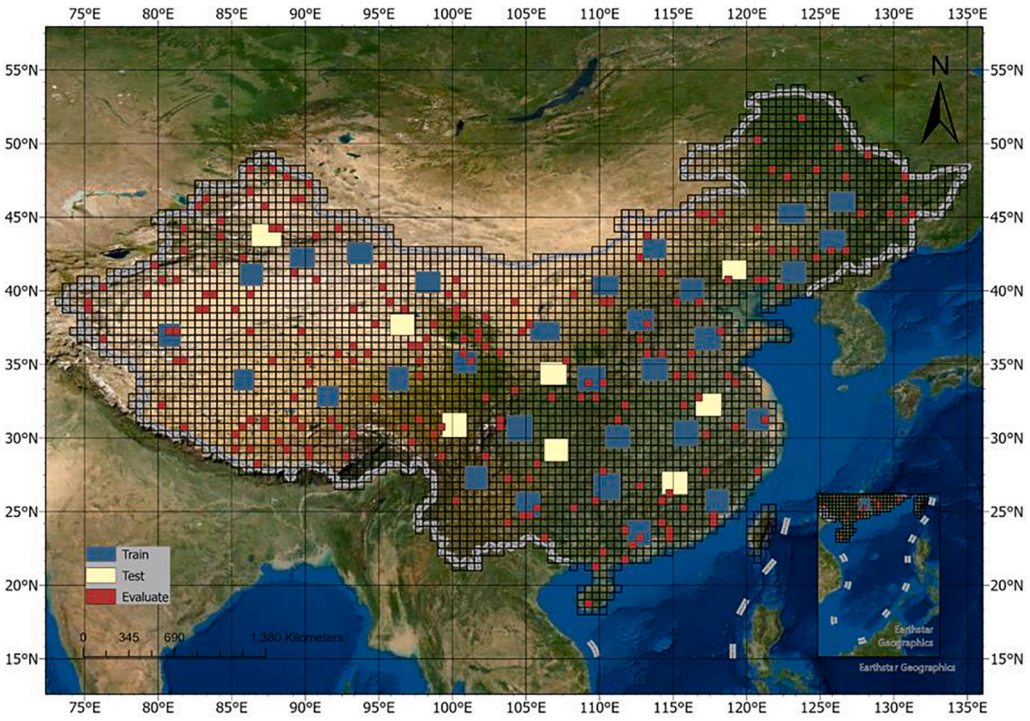

Fig. 1. Study area; study unit division; training plots, testing plots and evaluation plots distribution.

transformation (Jia et al. 2005) and etc. The segmentation method usually works in a supervised way, like support vector machine, Markov Random Field, and deep learning (Kearney et al. 2020; Song and Civco 2004; Wang et al. 2013; Zhang et al. 2018; Zhou et al. 2018). To date, deep learning shows its great potential in object detection and image segmentation (Badrinarayanan et al. 2017; Minaee et al. 2020; Zhao et al. 2019), which is also applicable for road extraction (Liu et al., 2019; Lu et al. 2019; Mnih and Hinton 2010; Saito et al. 2016; Wei et al. 2020). Among the state-of-the-art deep learning method, U-Net (Ronneberger et al. 2015) performs well in image segmentation, from which there are several road detection methods developed, for example, D-LinkNet (Zhou et al. 2018), deep residual U-Net (Zhang et al. 2018), weakly supervised ResUNet (Wu et al. 2019). The above deep learning networks usually work well in the single data source and high resolution remote sensing data while the application to large scale study is underexplored.

In terms of the different data sources, both the optical dataset and synthetic aperture radar (SAR) dataset are usually used for road extraction (Gamba et al. 2006), and the combination of both data sources outperforms a single data source. The spectral and shape information of the optical dataset benefit for road extraction while the geometric and physical information of SAR also shows its potential. Optical data for road extraction is limited to cloud contamination, shadows, and intra-/inter- class confusions (Cheng et al. 2019). With all-day and all-weather characteristics, SAR performs well in different weather conditions, which provides a fundamental data source for earth observation archives. Zhang et al. proposed the dual-polarimetric (VV and VH) of Sentinel-1 based on a deep fully convolutional neural network (Zhang et al. 2019). Generally, the road targets in SAR data are easily confused with the rivers and railways (Henry et al. 2018), and desert regions (Stewart et al. 2020). Previous methods usually use single-source data for road extraction and are constraint by spectral or backscattering confusions. Multisource data provide better performance comparing with single-source data in the research of impervious surfaces (Lin et al. 2020; Zhang et al. 2014), forest (Qin et al. 2019), etc. Combining different features for road extraction is promising (Lisini et al. 2011) and the effectiveness should be further explored.

The existing road extraction datasets generally focus on high resolution data and urban areas, the application of which to large scale study is still uncertain. Currently, available road extraction datasets include the Massachusetts roads data set (Mnih 2013) and DeepGlobe Road Extraction dataset (Demir et al. 2018), which generally focus on the urban area with consistent road widths and large background differences. Zhou et al. proposed the Wuhan University (WHU) Road Dataset that considers the challenging scenes of shadows and trees occlusion (Zhou et al. 2020). The connectivity and completeness of road extraction are easily affected by the occlusion of trees and buildings, as well as the discriminability of road and background (Lu et al. 2021; Wei et al. 2020). In addition, rural roads with diverse materials, large curvature changes and severe shelter problems (Liu et al. 2017) and roads in forested mountainous areas (Ferraz et al. 2016) are not well addressed. Road extraction at a large scale is usually challenging in the situation of variant topology, diverse spectrum and complex scenarios (Ren et al. 2020). The influences of different data sources, landscape patterns, and land cover types are yet to be explored for large-scale road extraction.

Overall, the combined use of optical and SAR demonstrates surpassed performance than single data source, while the contribution of each data source is underexplored. Second, previous study usually limits to dataset, or small plots, and focus on high resolution data sets. The road extraction performance at large scale is more trivial for the large geographic context, climate span, terrain influences and land cover diversities, the impacts of which should be further analysed. With relatively higher resolution, free availability and global coverage, Sentinel-1 and Sentinel-2 show great potential in large-scale research. In this study, we proposed the United U-Net (UU-Net) integrating Sentinel-1 and Sentinel-2 data for large-scale road extraction. The following two research objectives will be addressed: (1) To develop a multisource deep learning model for large scale road extraction (2) To analyse the SAR and optical data contributions for large scale road extraction. The performance and accuracy of the 10 m road extraction result will be analysed, which can serve as the pilot research for large-scale road extraction.

Table 1
Data information of Sentinel-1, Sentinel-2 and OSM.

Sensors	Date	Bands	Spatial Resolution (m)	Region & Scene Num.
Sentinel-1	20190101–20191231	VV, VH	10	Global Coverage, 31,732
Sentinel-2	20190101–20191231	Blue (496.6 nm (S2A) / 492.1 nm (S2B)), Green (560 nm (S2A) / 559 nm (S2B)), Red (664.5 nm (S2A) / 665 nm (S2B)), NIR (835.1 nm (S2A) / 833 nm (S2B))	10	Global Coverage, 371,394
OSM	20,200,101	Highway	–	Global Coverage

Table 2
Training and testing patches in terms of different climate regions.

Climate	Training	Testing
Warm Temperature	24,000	9600
Boreal	14,000	4800
Arid	16,000	2400
Polar	6000	2400

2. Large-scale road extraction dataset

2.1. Study area and dataset

With longitude spanning from around 73°E to 135°E and latitude from around 18°N to 53°N, China was chosen as the study area, which contained different landforms and complex climate zones. According to the World Factbook (https://www.cia.gov/library/publications/the-world-factbook/geos/ch.html), the geographic zones are mostly comprised of mountains, hills, basins, plains, and plateaus. According to the Köppen climate classification (Kottek et al. 2006) A1F1 scenario, there are 4 major climate zones including warm temperature, boreal, arid and polar in China, which consists of 8 climate types: Cfa (warm temperature, fully humid, hot summer), Cwa (warm temperature, winter dry, hot summer), Cwb (warm temperature, winter dry, warm summer), Dwa (snow, winter dry, hot summer), Dwb (snow, winter dry, warm summer), BSk (arid, steppe, cold arid), BWk (arid, desert, cold arid), and ET (polar, polar tundra). The landforms, topography and climate shape the versatile environment in different landscapes (Zheng 2006). The density and pattern of the road network are closely related to the different social-environmental statuses and land cover (Hawbaker et al. 2005). The study area was divided into 0.5-by-0.5 degree grids for processing as Fig. 1.

The multisource road extraction data sets included global coverage Sentinel-1, Sentinel-2 as in Table 1. The satellite data were collected and preprocessed from google earth engine (GEE) (Gorelick et al. 2017). The VV and VH polarization intensity band of Sentinel-1 GRD products were used. In addition, 4 bands covering visible and near-infrared spectrum with 10 m resolution of Sentinel-2 Level-1C products were selected. As for the Sentinel-1 and Sentinel-2 satellite data, the annual images ranging from Jan 1st 2019 to Dec 31st, 2019 were used. The road label data were from OSM (https://www.openstreetmap.org/), which was assumed as reference data for model training and accuracy assessment. The time snapshot of Jan 1st, 2020 was used as the representation of the road of the year 2019. Different from most previous studies on small scales, we trained the deep learning model across the whole study area and evaluate in a large-scale manner.

2.2. Data preprocessing

2.2.1. Sentinel-2 multispectral data preprocessing

Sentinel-2 multispectral instrument level-2A surface reflectance was used in this study, which has been atmospheric, terrain and cirrus corrected using the Sen2cor module from European Space Agency. First, the product with cloud-pixel percentage larger than 20 was filtered out. Then, at pixel dimension, the QA60 band was evaluated to filter out the pixels contaminated by dense cloud and cirrus cloud. After that, we rescaled the reflectance from 0 to 1 using the scale factor of 10000. Assuming that the road changes within a year are relatively small, the median value for each pixel was used to generate the full coverage multispectral data of China in the year 2019. The median value can also avoid the oversaturated and lower value of vegetation with large phenological changes (Venkatappa et al. 2019).

2.2.2. Sentinel-1 VV/VH backscattering coefficient preprocessing

Sentinel-1 ground range detection (GRD) level 1 products with IW mode were used in this study. The preprocessing steps including - apply the orbit file, thermal noise removal, radiometric calibration, radiometric terrain flattening and terrain correction - were applied to the GRD data using the Sentinel-1 Toolbox. After that, the terrain-corrected VV and VH polarization intensity decibels were generated. The temporal mean value per pixel dimension was used for mosaic, which considered the time series effect and noise reduction (Quin et al. 2014).

2.2.3. OSM preprocessing

The OSM highway datasets were in vector format. We assumed the road width of the dataset was 10 m and we buffered 5 m around each road. After that, we transformed the road from vector to raster with road pixel of the assigned value of 1 and the background of 0 with 10 m resolution. For the OSM data, different kinds of roads were used in this study, including principle road of motorway, trunk, primary road, secondary road, tertiary road, unclassified road, residential road and some special road. It is arbitrary that we used a 5 m buffer for different kinds of roads. It is a simplification for large-scale analysis, which we assume that our deep learning model can tolerate the noise caused by the OSM data.

2.3. Large-scale multisource road extraction dataset generation

The polarization intensity features of Sentinel-1, spectral features of Sentinel-2 and OSM highway data were stacked and registered in 10-meter resolution with a geographical system of EPSG:4326. Different from the previous road extraction dataset, the proposed dataset includes optical and SAR data and is in a large-scale manner, i.e., entire China. We treated the optical, SAR data and the OSM highway data as a data stack and applied sampling in our study area.

Stratified random sampling methods were used for samples generation. In the model training process, 30 training plots and 8 testing plots were randomly chosen across the study area according to the Köppen climate classification. For each climate region, we randomly selected several plots. At each plot, 2000 patches were randomly selected as training samples and 2400 patches were chosen as testing (validation and evaluation) patches. A total of 60,000 training patches and 19,200 testing patches were first selected, which distributes as Table 2. Warm temperature dominated a large part of the study area, which accounts for the higher proportion of the sample size. The training and testing samples follow the same spatial distribution, which ensures the representativeness of the sampling process. Pixel level annotations of remote sensing road extraction datasets are unavailable in large-scale regions. Weakly supervised learning (Papandreou et al. 2015; Wu et al. 2019) was used in this study. We used a fixed width of the road instead of the

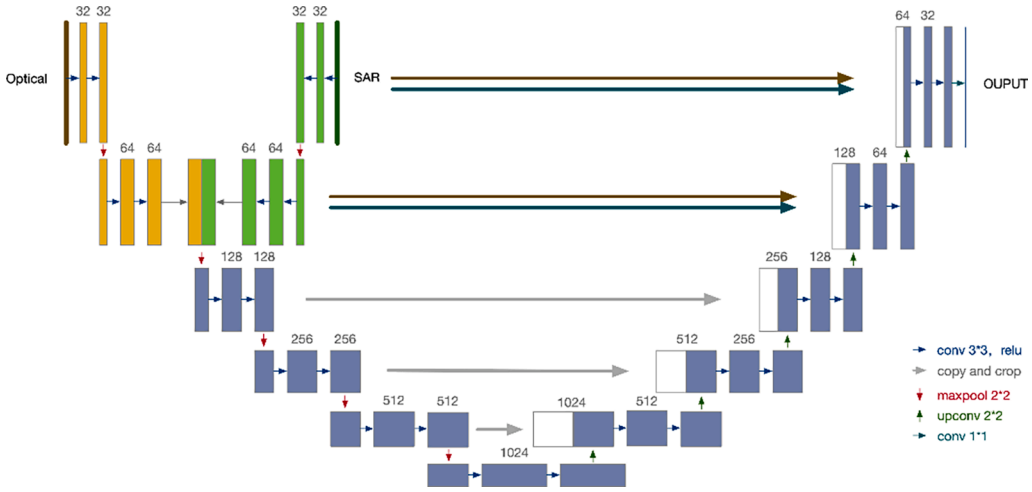

Fig. 2. UU-Net structure.

Table 3
Architecture, size of feature maps, and total model parameters of UU-Net.

Layer	Type	Patch size / stride	Output shape	Depth	Params
Input	Input Layer	–	256*256*6	0	
Input1 (opt)	Input Layer	–	256*256*4	4	
Input2 (SAR)	Input Layer	–	256*256*2	2	
Encoder1 (opt)	Conv2D	3x3	128x128x32	3	10,688
	BN				
	ReLu				
	Conv2D	3x3			
	BN				
	ReLu				
	MaxPooling	2x2/2			
Encoder2 (opt)			64x64x64	6	55,936
Encoder1 (SAR)			128x128x32		10,112
Encoder2 (SAR)			64x64x64		55,936
concatenate (opt + SAR)	Concatenate		64x64x128	1	0
Encoder3			32x32x128	1	296,192
Encoder4			16x16x256	2	887,296
Encoder5			8x8x512	5	3,544,064
Conv	Conv2D		8x8x1024	1	4,723,712
	BN				
	ReLu				
Conv			8x8x1024		9,442,304
Decoder1	Conv2DTranspose	2x2/2	16x16x512	5	9,184,768
	Concatenate				
	BN				
	ReLu				
	Conv2D	3x3			
	BN				
	ReLu				
	Conv2D	3x3			
	BN				
	ReLu				
Decoder2			32x32x256	2	2,298,624
Decoder3			64x64x128	1	575,872
Decoder4			128x128x64	6	181,696
Decoder5			256x256x32	3	45,792
Conv2D	Conv2D	1x1	256x256x1	1	33
Sigmoid			256x256x1		
Total parameters: 31,313,025					

exact width. To balance the positive and negative samples in the training process (Bonafilia et al. 2019), we filtered out the patches that the road pixel percentage was lower than 1%. Overall, 18,283 training patches and 5322 testing patches were used in the model training and evaluation. The testing patches were further split into 865 validation patches and 4457 evaluation patches. In the large-scale evaluation process, 200 plots as in Fig. 1 were chosen for evaluation, which were distributed randomly across the whole study area.

2.4. Auxiliary data for large-scale analysis

For the evaluation grids, we considered the elevation, slope, and land cover (croplands, barren and urban) for large-scale spatial autocorrelation analysis. Elevation and slope are usually considered as the limited natural factors since flatter land with lower elevation are easier for urban development (Feng et al. 2021). The elevation is from Shutter Radar Topography Mission (SRTM) digital elevation data (Farr et al.

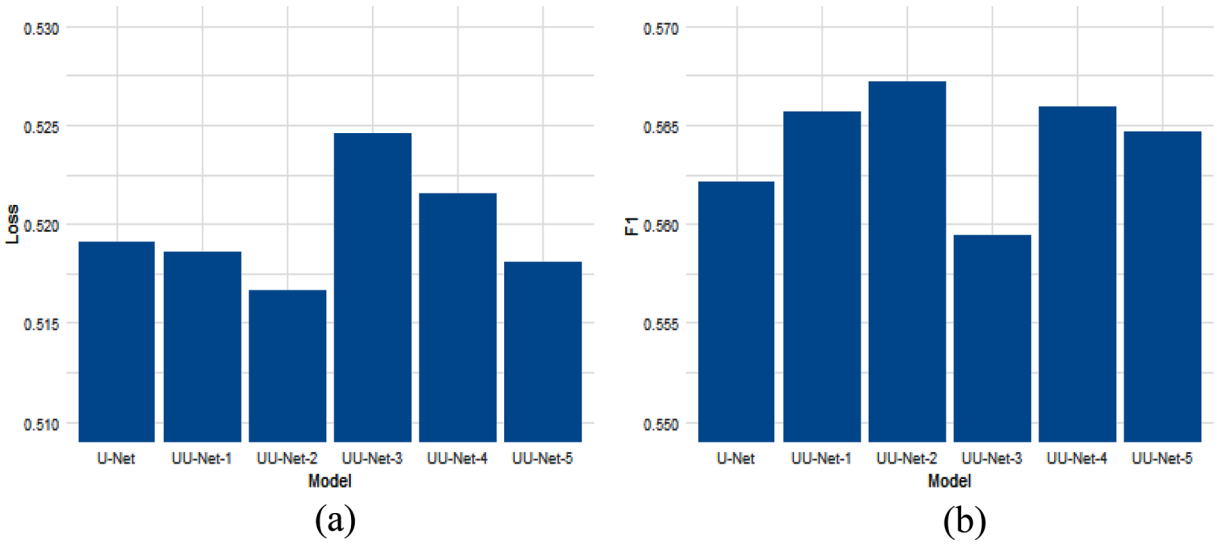

Fig. 3. Effectiveness of different fusion levels of UU-Net.

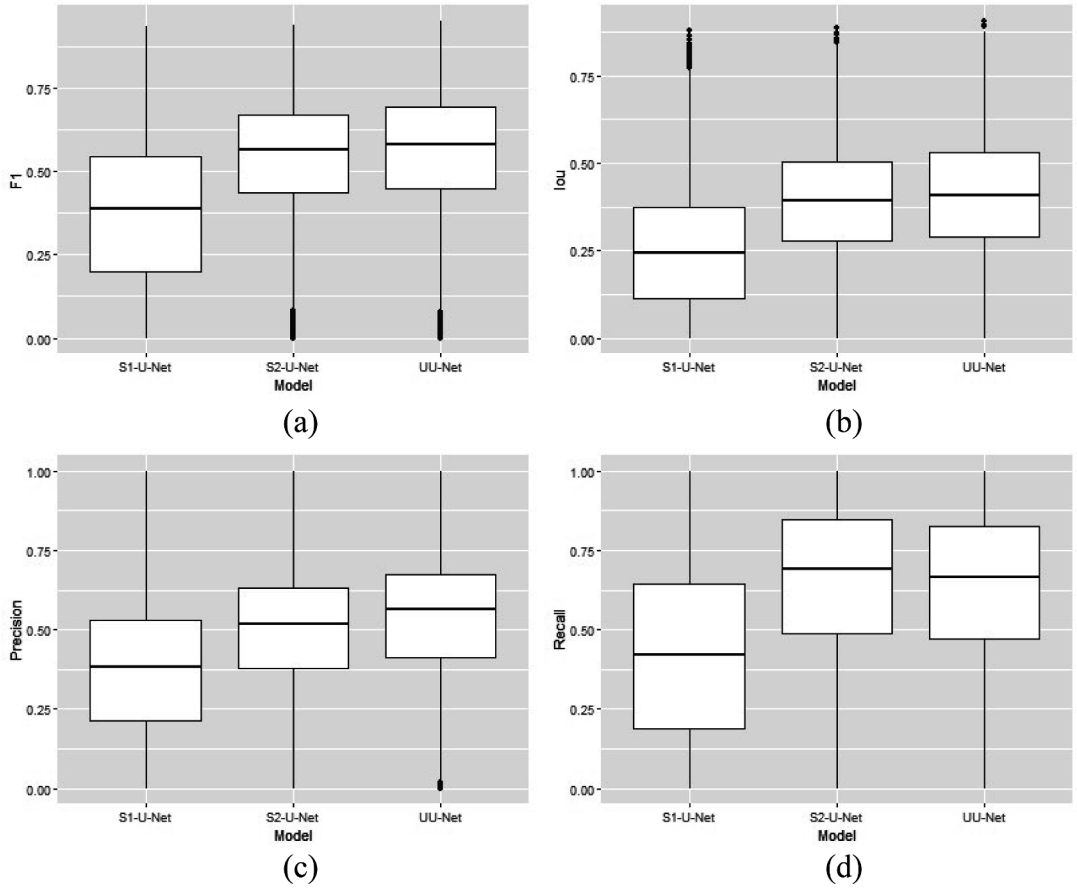

Fig. 4. F1, IoU, precision and recall comparison between the S1-U-Net, S2-U-Net and UU-Net.

2007), which is processed at a resolution of 1 arc-second (approximately 30 m) and has undergone void-filling. The slope was calculated from elevation and ranges from 0 to 90 degrees. Land cover data is from the Terra and Aqua combined Moderate Resolution Imaging Spectroradiometer (MODIS) Land Cover Type (MCD12Q1) Version 6 product (Friedl and Sulla-Menashe 2015) of International Geosphere-Biosphere Programme (IGBP) classification scheme, which is of the year 2019 with 500 m resolution. For the evaluation grid in Fig. 1, zonal statistics of elevation and slope per grid were attained, and group zonal statistics of major land cover types including urban, cropland, forest, grassland and barren were attained. The per grid elevation and slope were further normalized from 0 to 1 while the land cover was shown as the percentage of different land cover types.

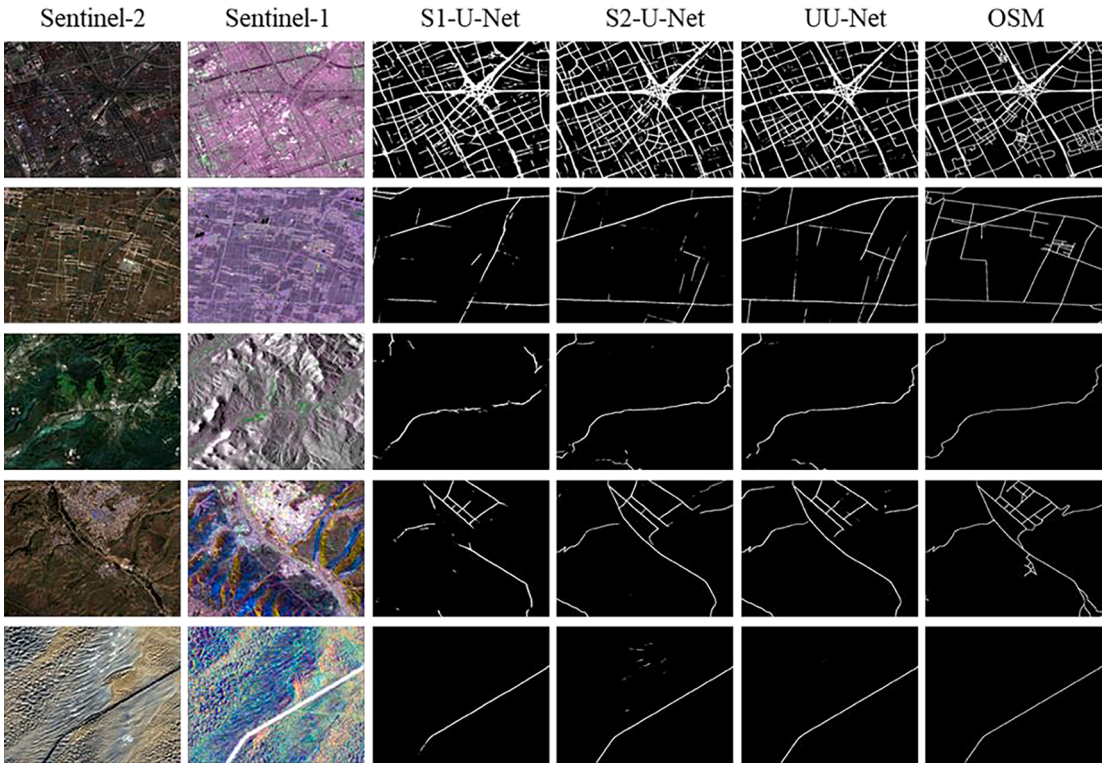

Fig. 5. Road extraction result comparison in different landscapes: Sentinel-2, Sentinel-1, S1-U-Net, S2-U-Net, UU-Net and OSM. Landscape types from top to bottom: urban, cropland, forest, grassland and barren.

3. Methodology

Stratified random sampling was performed on the large-scale VV and VH backscattering features of Sentinel-1, spectral features of Sentinel-2 and OSM map to generate the training sets, validation sets and evaluation sets. The training sets and validation sets were used to train the UU-Net model. With the well-trained model, we performed the large-scale road prediction and finally got the large-scale road segmentation map. After that, we analyzed the large-scale results on the evaluation sets.

3.1. UU-Net structure and implementation details

U-Net (Ronneberger et al. 2015) worked as an encoder and decoder structure. On one hand, the symmetric U-shape structure can capture the context information, on the other hand, the expansive path from the high resolution feature helps the precise localization. In the contracting path, it consisted of the repeated encoders of two 3×3 convolution layers followed by a rectified linear unit (ReLU), then a 2×2 max pooling with a stride of 2 for downsampling. In the expansive path, the feature map included the 2×2 up-convolution and the cropped feature map from the contracting path. For the final layer, the 1×1 convolution layer was used to generate the final segmentation map.

We proposed the United U-Net (UU-Net) to combine two inputs of optical (4 spectral bands) and SAR (2 polarization intensity features) as in Fig. 2, which looks like two U shapes. Different fusion levels were tested to find out the optimal fusion level for optical and SAR, i.e., fusion in feature level in the 1st encoder, 2nd encoder, 3rd encoder, … and 5th encoder. Fig. 2 indicates fusion in the 2nd encoder, that is, after two encoders, the optical and SAR encoded features are concatenated together and there are two cropped features in the corresponding expansive path. Binary cross entropy dice loss (Equation (3)) includes the binary cross entropy loss (Equation (1)) and dice loss (Equation (2)). In binary cross entropy loss, the N denoted the sample size, and $log(p(y_i))$ represented the log probability of the predicted class. The dice loss was used as the loss function to consider the imbalanced road and background situation, which can be calculated from the true positive (TP), false positive (FP) and false negative (FN) generating from the ground truth and predicted labels. Adam (Kingma and Ba 2014) was chosen as the optimizer for its computationally efficient capability.

$$L(y,y_i)_{BCE} = -\frac{1}{N}\sum_{i=1}^{N} y_i \bullet \log(\mathrm{p}(\mathrm{y_i})) + (1-\mathrm{y_i})\bullet\log(1-\mathrm{p}(\mathrm{y_i})) \quad (1)$$

$$L_{Dice} = 1 - \frac{2TP}{2TP+FP+FN} \quad (2)$$

$$L = L(y,y_i)_{BCE} + L_{Dice} \quad (3)$$

The UU-Net network architecture, size of feature maps can be referred to Table 3. U-Net using SAR (S1-U-Net) and U-Net using optical data (S2-U-Net) were used for comparison. S1-U-Net and S2-U-Net follow the basic U-Net structure. The total parameters of UU-Net, S1-U-Net and S2-U-Net are 31,313,025, 31,126,209, and 31,126,785 respectively.

3.2. Experiment setup and evaluation metrics

Since the large physical variance of optical and SAR, the direct stack of the features is unfair. The optical and SAR features in the data stack were first normalized to a mean value of 0 and standard deviation of 1 to keep its comparability among feature dimensions. The input patch size of the optical and SAR stack was empirically set as spatial dimension of 256×256 and feature dimension of 6 (4 optical features and 2 SAR features). Then, data augmentations including flipping and rotation were performed on 25% of the training patches to increase the randomness. Flipping included random flipping from left to right and random flipping from up to down. The rotation included random rotation of 90, 180, 270 and 360 degrees.

Four matrices were used for validation including precision, recall, F1

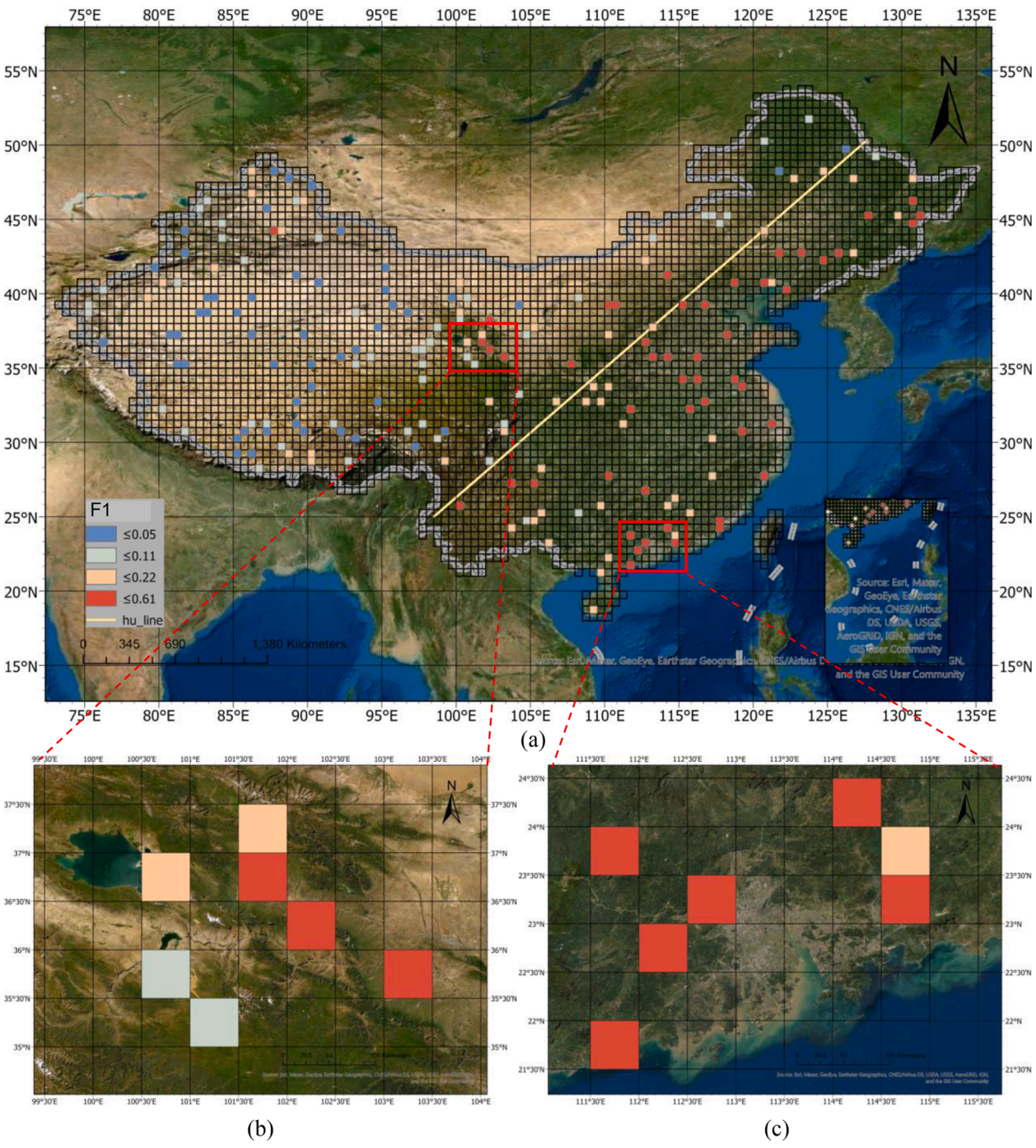

Fig. 6. Large-scale road extraction accuracy distribution of UU-Net.

score, mean intersection over union (mIoU). Based on the confusion matrix, TP, TN, FP and FN can be generated from the ground truth and predicted labels. The precision (correctness) and recall (completeness) (Sujatha and Selvathi 2015) can be calculated as:

$$precision = \frac{TP}{TP + FP} \tag{4}$$

$$recall = \frac{TP}{TP + FN} \tag{5}$$

F1 score was calculated from precision and recall:

$$F1 = 2 \bullet \frac{\text{precision} \bullet \text{recall}}{\text{precision} + \text{recall}} \tag{6}$$

IoU considered the union and intersection of the targets, which can be computed from:

$$IoU = \frac{TP}{FN + FP + TP} \tag{7}$$

3.3. Large-scale road extraction accuracy evaluation

Global Moran's I (Getis and Ord 1992) was used for the spatial autocorrelation analysis for the road extraction pattern, which considered the road extraction accuracy value and its location simultaneously. It is usually used in social and environmental geospatial analysis like population or temperature. The z-score (Equation (10)) will represent the patterns of clustered, dispersed, or random and p-value will

Table 4
Summary of OLS results. OLS: ordinary least square. VIF: variance inflation factor. AICc: Akaike's information criterion.

UUNet OLS	R2	0.70	AICc	−493.99
Variable	Coefficient	StdError	p-value	VIF
Elevation	−0.08	0.02	0.000*	2.46
Slope	−0.01	0.03	0.672	1.76
Forest	−0.03	0.04	0.505	2.34
Grass	−0.06	0.02	0.017*	3.51
Cropland	0.13	0.03	0.000*	2.27
Urban	0.62	0.08	0.000*	1.14
Barren	−0.14	0.02	0.000*	2.84
S2-U-Net OLS	**R2**	0.70	**AICc**	−501.17
Variable	**Coefficient**	**StdError**	**p-value**	**VIF**
Elevation	−0.08	0.02	0.001*	2.46
Slope	−0.01	0.03	0.799	1.76
Forest	−0.03	0.04	0.455	2.34
Grass	−0.07	0.02	0.011*	3.51
Cropland	0.12	0.03	0.000*	2.27
Urban	0.60	0.08	0.000*	1.14
Barren	−0.14	0.02	0.000*	2.84
S1-U-Net OLS	**R2**	0.75	**AICc**	−659.65
Variable	**Coefficient**	**StdError**	**p-value**	**VIF**
Elevation	−0.04	0.02	0.000*	2.46
Slope	−0.04	0.02	0.013*	1.76
Forest	−0.05	0.03	0.077	2.34
Grass	−0.05	0.02	0.010*	3.51
Cropland	0.11	0.02	0.000*	2.27
Urban	0.53	0.06	0.000*	1.14
Barren	−0.10	0.02	0.000*	2.84

determine the significance of the index. In this study, we used the global Moran's I to test the road extraction performance on large-scale road extraction accuracy with the hypothesis that the road extraction accuracy will be affected by its geographical location. The global Moran's I was calculated as

$$\mathrm{I} = \frac{n}{S_0} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{i,j} z_i z_j}{\sum_{i=1}^{n} z_i^2} \tag{8}$$

Where i and j were two features in the geographic context. $n$is the feature number. The z_i was the deviation of an attribute for feature i from its means $(x_i - \widehat{X})$, which is similar for z_j. $w_{i,j}$was the weight between i and j. S_0 was the aggregation of all the spatial weights,

$$S_0 = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{i,j} \tag{9}$$

The Z_I score can be calculated as,

$$Z_I = \frac{I - E[I]}{\sqrt{V[I]}} \tag{10}$$

Where

$$E[I] = -1(\mathrm{n}-1) \tag{11}$$

$$V[I] = E\left[I^2\right] - E[I]^2 \tag{12}$$

To examine the impact factors that influence the road extraction accuracy y, the ordinary least square regression (Goodchild et al. 1993) considered the factors of elevation ($x_{elevation}$), slope (x_{slope}), and land cover types percentage (including cropland ($x_{cropland}$), barren (x_{barren}), urban (x_{urban}), forest (x_{forest}) and grassland ($x_{grassland}$)). The β_i denotes the coefficient and the ε indicates the residual. For large scale impact analysis, all patches in 200 grids were considered to reflect the actual landscape situation.

$$y = \beta_0 + \beta_1 x_{elevation} + \beta_2 x_{slope} + \beta_3 x_{urban} + \beta_4 x_{cropland} + \beta_5 x_{forest} + \beta_6 x_{grassland} + \beta_7 x_{barren} + \varepsilon \tag{13}$$

4. Results

4.1. Fusion level analysis of the UU-Net

To explore the effective fusion level of optical and SAR for large-scale road extraction, the accuracy of different encoders' fusion levels are compared in Fig. 3. When the optical and SAR features were fused in the second encoder, it achieved the lowest loss of 0.5166 and the highest F1 score of 0.5672. The direct layer stacking of the optical and SAR features is not the most optimal fusion strategy for their different imaging mechanism. After features are encoded, both features are at an abstract level, which is more suitable for optical and SAR feature-level fusion. In addition, the upper two streams of expansive paths can keep the spatial context information of both optical and SAR respectively, which makes it more competitive in road discrimination than just layer stacking. In addition, the result does not indicate that a deeper fusion level is better, instead, a gradual decrease in accuracy can be found after fusion in the 3rd encoder. In this study, we chose the UU-Net-2, i.e., optical and SAR features fusing in the second encoders as our final model and further explored its large-scale performance.

4.2. Contribution of optical and SAR for road extraction

The road extraction accuracy was compared among S1-U-Net, S2-U-Net and UU-Net as in Fig. 4. The UU-Net got the best performance among the three groups, with an average F1 of 0.5502 and IoU of 0.4021. The performance of S2-U-Net ranked second, which was slightly lower than the UU-Net, with an average F1 of 0.5357 and average IoU of 0.3863. Concerning S1-U-Net, only 2 features, i.e., VV and VH backscattering coefficients intensity were used for road extraction, comparing with 4 features of S2-U-Net, and 6 features of UU-Net. The road extraction accuracy of S1-U-Net was relatively lower, with F1 of 0.3749 and IoU of 0.2536. Large variations can be found in a certain network for complex landscapes in the study area. Concerning the precision, UU-Net obtained higher correctness than the other two methods, followed by S2-U-Net and S1-U-Net. In terms of the recall, S2-U-Net and UU-Net showed higher completeness than S1-U-Net. The improvement from UU-Net comes from improved correctness and completeness.

Detailed road extraction performances are compared in Fig. 5. In the urban area (row 1), the road detection result in three methods kept good performance in major roads but showed less road in residential blocks, i.e., road within the park or residential block, and pedestrian. It's worth to mention that the 10 m road extraction result is not so complete as the OSM data, especially for the residential blocks that are less than 100 m width. As in satellite images, the road detection in residential blocks was limited by the image resolution, which showed spectral confusions between buildings and roads in Sentinel-2 and backscattering coefficients confusions in Sentinel-1. In the cropland dominated land cover (row 2) and the small settlements with low road density, UU-Net was more reliable than S2-U-Net by reducing the false detection errors. Row 3–5 indicated the mountainous area with forest cover, agricultural land and desert. The omission error of S1-U-Net was obvious than the two other methods. The terrain effect was a major problem for road extraction using SAR data. S2-U-Net tended to have a higher false detection rate in farmland. The synthetic result of UU-Net was more consistent with OSM road with more complete road detection results in the mountainous area and more correct road results in cropland. The discontinuous phenomenon also exists in the road extraction result, which may result from the image quality like shadow and occlusion, or specific functions like tunnel and underground.

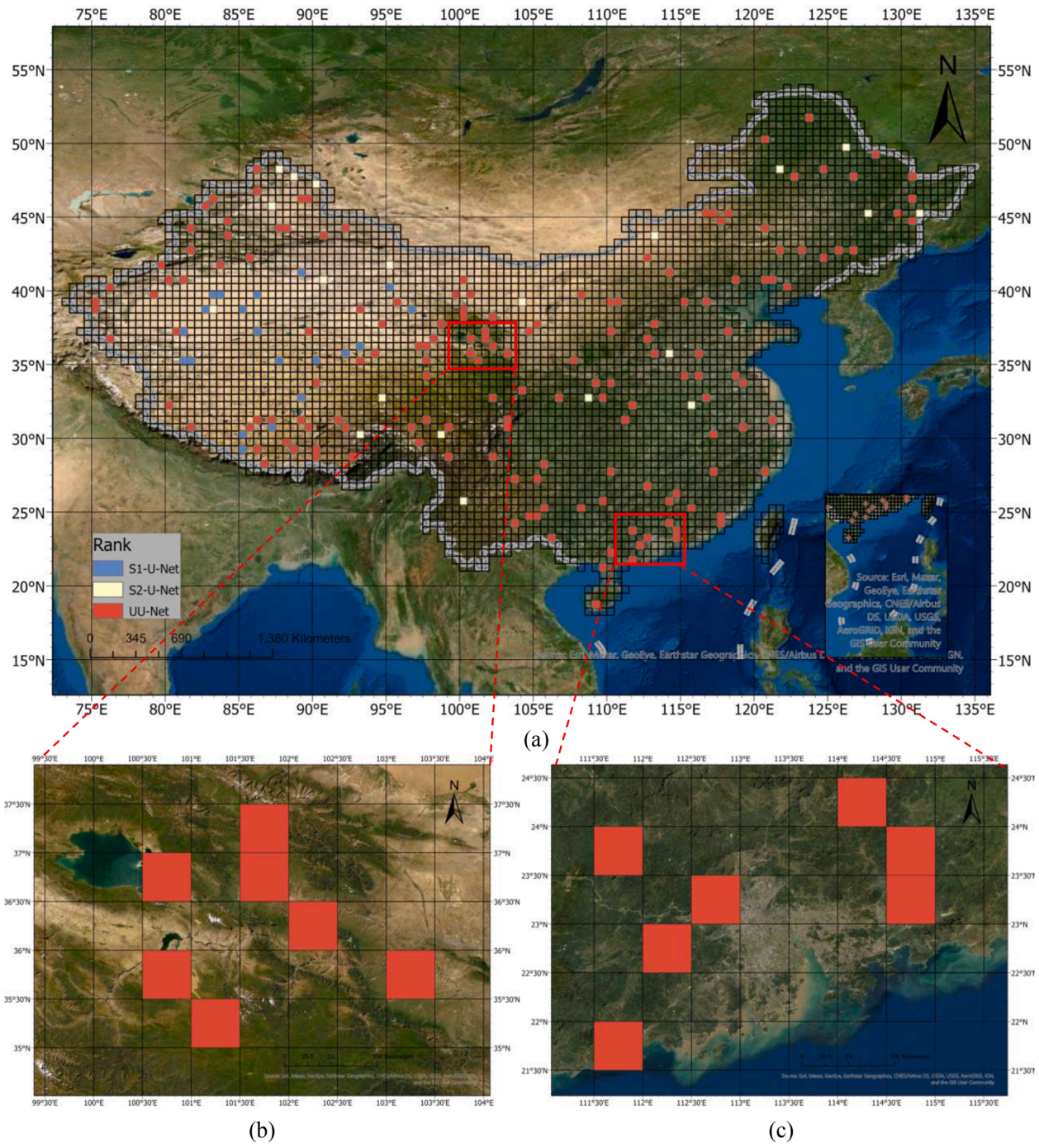

Fig. 7. Best accuracy distribution in road extraction result.

4.3. Road extraction application at large scale

Fig. 6 demonstrated the large-scale road extraction accuracy results using UU-Net, which showed lower accuracy in the south and higher accuracy in the east. The accuracy distribution of S1-U-Net and S2-U-Net also showed similar patterns. The pattern is also consistent with the Huhuanyong line (400 mm equivalent precipitation line) (Hu 1935), which infers that more than 90% population distributed in the southeast part of China. The southeast part constitutes of plain, watershed, hills, Karst landform and Danxia landform while the northwest part constitutes of prairie, desert, and plateau. It is surprising that the model accuracy distribution also conforms with the population geographical patterns, which reflects the climate and terrain restrictions on the model accuracy. Different precipitation and soil types also account for the accuracy disparity in different regions.

The Z-score were 19.55, 19.82, 18.72 in UU-Net, S2-U-Net and S1-U-Net respectively with significant p value. Given the Z-score higher than 2.58, there was less than 1% likelihood that this cluster pattern could be the result of random chance. The results indicated that the result of one model for large-scale prediction showed a spatial autocorrelation pattern. Table. 4 shows the ordinary least square regression results considering different factors. Forest did not significantly impact the accuracy of the three models. Slope did not significantly affect the accuracy of S2-U-Net and UU-Net but affected the accuracy of S1-U-Net, which indicated the terrain effects of SAR in the mountainous area. As a result, when incorporating SAR to the road extraction model as in UU-

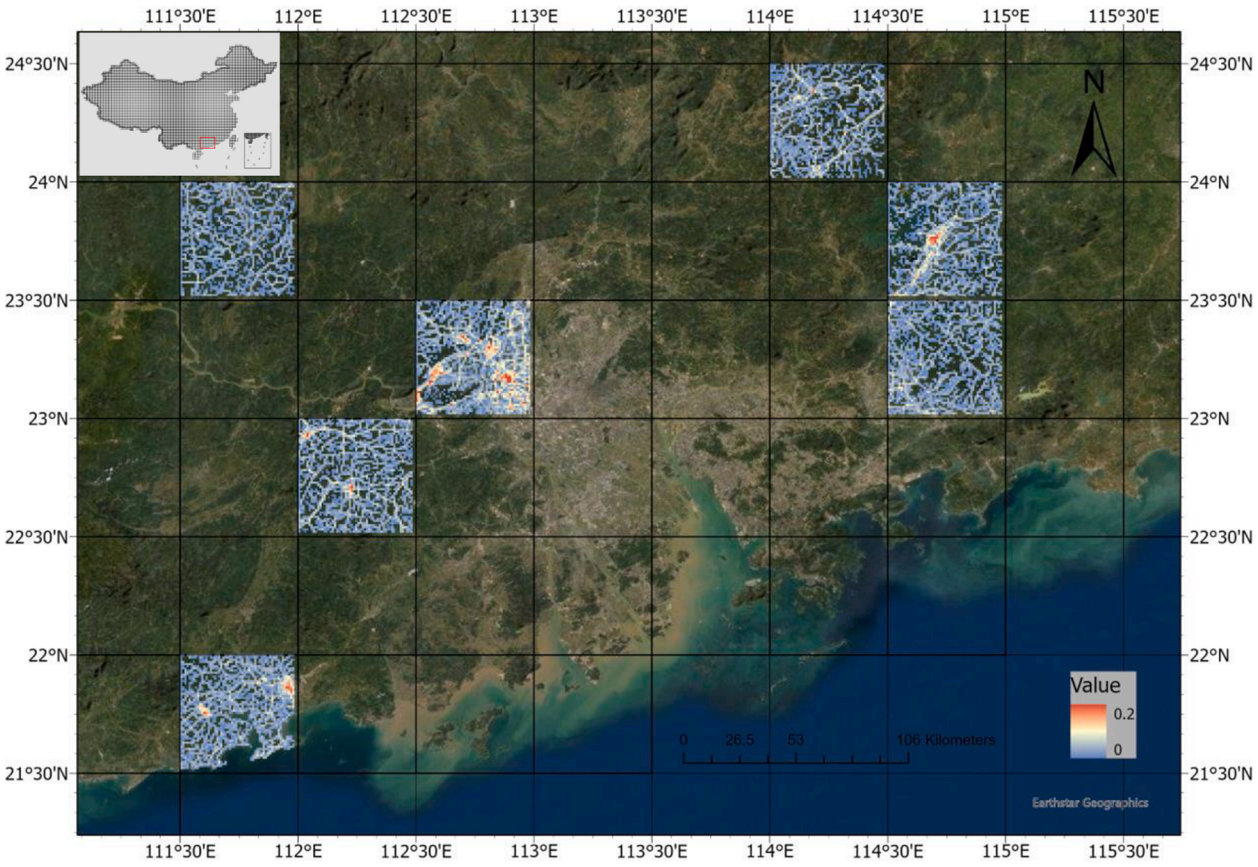

Fig. 8. Road density of 1 km^2 using UU-Net in eastern China.

Net, the model will be less affected by the slope. The three models were all affected by the elevation, cropland, grass, urban and barren. Generally, the road extraction accuracy has a negative relationship with the elevation, a positive relationship with the urban and a negative relationship with barren. This study first provides a quantitative analysis of the impact factors of optical and SAR data across the large-scale region.

The best accuracy distribution of the road extraction models was demonstrated as Fig. 7 with UU-Net outperformed in 160 grids, S2-U-Net outperformed in 20 grids, and S1-U-Net outperformed in 20 grids. The result indicates the fusion of SAR and optical data works better than single source data in large-scale road extraction. Higher accuracy can be obtained from UU-Net in large scale road extraction. S2-U-Net performs well in the large slope area, croplands, and grassland, and the high accuracy distributes sparsely across the whole China. S1-U-Net performs well in western part of China where is dominated by high elevation and barren land cover.

The final road extraction results for the western and eastern selected regions were as Fig. 8 and Fig. 9 respectively. In 1 km^2 grid, the road density varied from 0 to 20 percent. The road density was high in the urban center and decreased along the urban gradient. The road density patterns of urban agglomerations were in ring and radial shape. The eastern regions usually have flatter slopes and higher urbanization levels while in the western regions, mountains are dominated. So the road density was relatively lower and distributed sparsely.

5. Discussion

5.1. Large-scale optical and SAR fusion

The spectral features of optical data and backscattering coefficients of SAR data are physically different. The variance becomes larger in a large-scale region with diverse climate and complex terrain effects. Simple layer stacking of the optical and SAR data does not present the best results. In this study, the integration of these two kinds of features in the second encoders of the U-Net achieved the lowest loss and highest F1 score. Different deep learning models can result in different fusion strategies, U-Net works well in remote sensing segmentation and can be a baseline for the fusion of optical and SAR for road extraction. Still, the UU-Net fusion of optical and SAR in the second encoders is implicit. To this stage, we can only conclude the results from the accuracy matrix and cannot explicitly demonstrate the fusion effects.

Variant data characteristics of different data sources generate different uncertainties for road extraction. In this study, we compared the road extraction results among S1-U-Net, S2-U-Net and UU-Net. The performance of multisource data (UU-Net) was not the sum of single-source data (S1-U-Net and S2-U-Net) for the reason that we used the feature-based fusion method. We compared the road extraction results and found that there is randomness that some roads may be detected by S1-U-Net or S2-U-Net but cannot be detected by UU-Net. Still, we visually found some similar patterns when the single or multi-source data were used. SAR data helped to delineate the road near bare land, road in farmland, road under trees or shadows, and road made of bare lands. But road extraction using SAR data sometimes will generate errors in small streams which has similar low backscattering characteristics. For the roads in the mountainous area, it is hard to detect by SAR for the terrain influences. Concerning the road extraction using optical data, it achieved good performance in the concrete/asphalt roads, and mountainous roads. Also, it has false detection in small streams. When we consider the multisource method (UU-Net), the proposed method can improve the road detection in the road near bare land, road in cropland, road in the residential area and reduce the misclassification in small streams.

In this study, we fused the optical and SAR based on the U-Net. Other deep learning networks, such as D-LinkNet and ResUNet, are potential

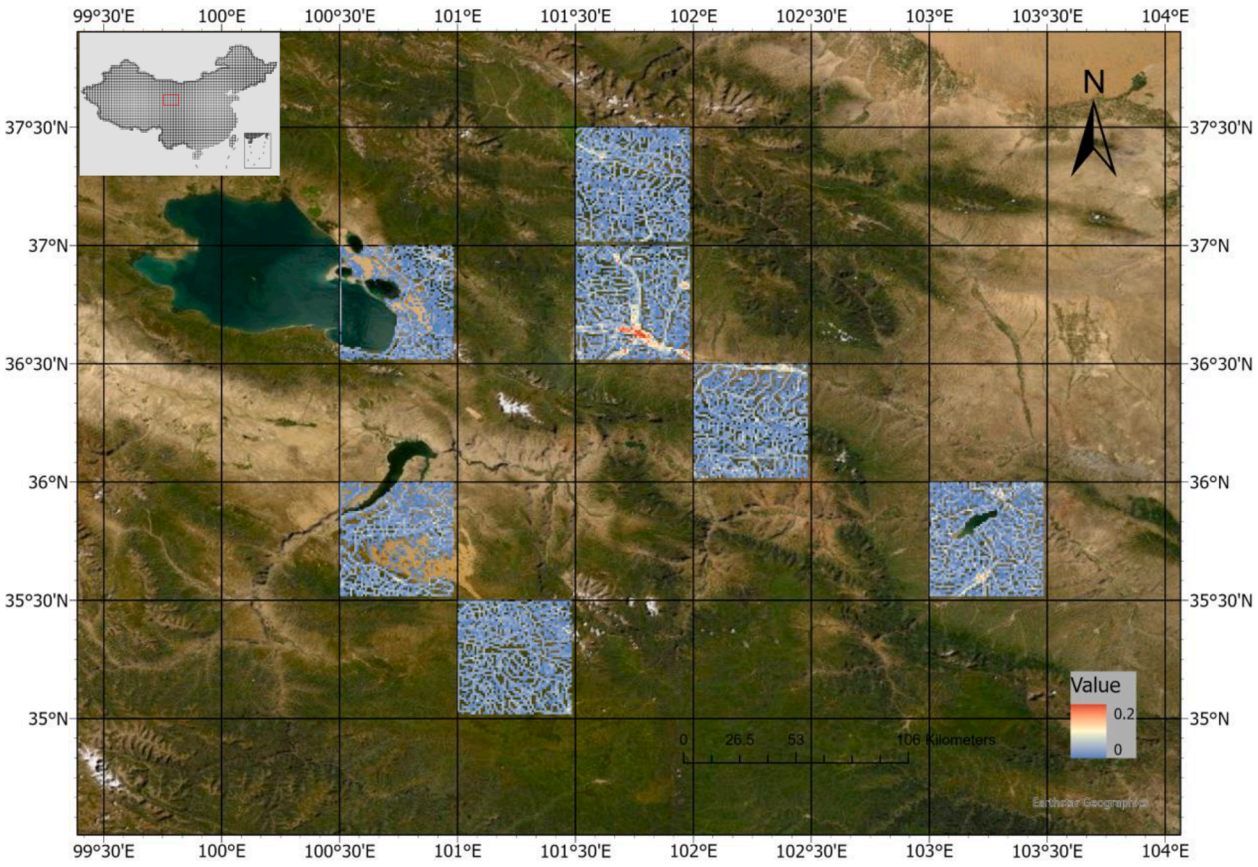

Fig. 9. Road density of 1 km^2 using UU-Net in western China.

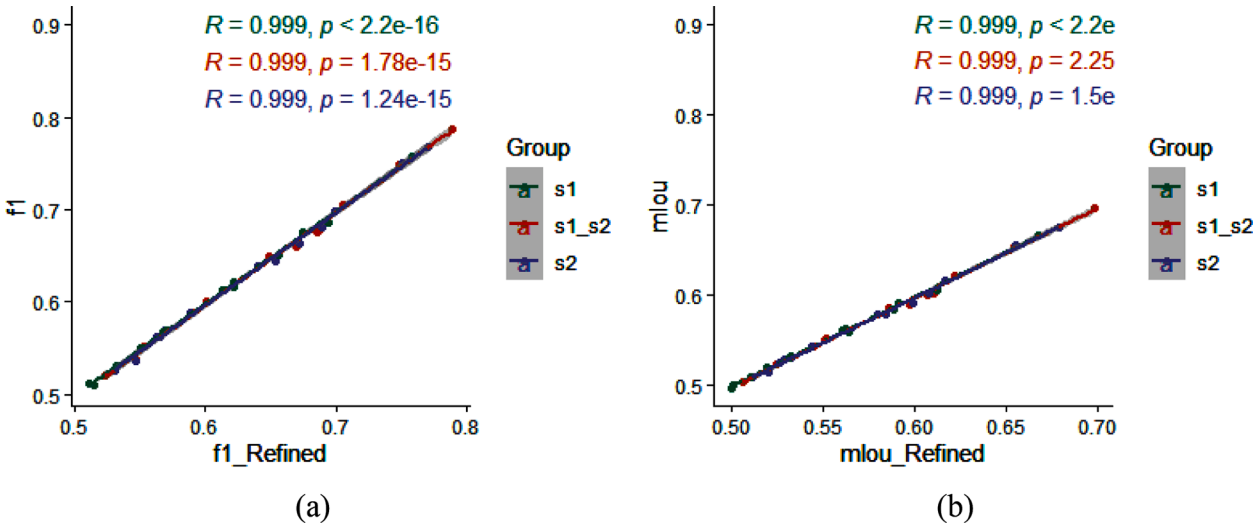

Fig. 10. Scatter plots of (a). F1 and refined F1, (b). mIoU and refined mIoU.

for accurate road extraction using optical and SAR at a large scale. Still, the optimal fusion strategy, network structure and parameter settings can be explored in the future study.

5.2. Large-scale road extraction issues

It's trivial to train a deep learning model for large-scale road extraction in terms of data availability and computation resources. The average accuracy, with the average F1 of 0.5502 and IoU of 0.4021 in 200 0.5-by-0.5 degree grids, is relatively lower as 1) in a large scale study, the landscapes have more diversities than the dataset level, which is reflected in the results that the road extraction has higher accuracy in the urban area, which is consistent with the previous study; and 2) the 10 m resolution is relatively coarser than the other research with meter level resolution. Coarser resolution may lead to more uncertainties for road extraction.

OSM road centreline data were used as the reference training labels for large-scale road extraction. The label conversion was arbitrary since we used fixed road width for different kinds of roads. Instead, the roads are of variant road widths in the actual environment. In addition, there

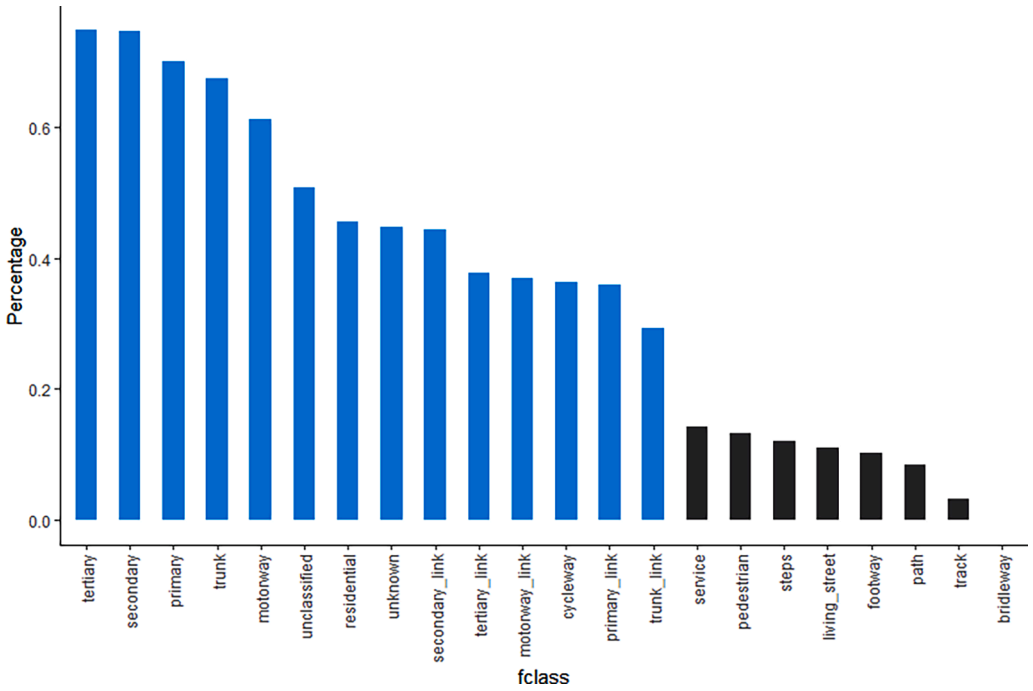

Fig. 11. 10 m resolution road centreline percentage in OSM, blue: OSM principle road, black: OSM special road. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

are also some uncertainties and quality issues of the OSM road dataset. The road location and road width are the noise when we use the OSM road dataset for training and testing. The large-scale labelling data are unavailable in a large-scale study. It's an alternative that we use the voluntary data and a simplification road width for model training. To compare the OSM and the ground truth road, we refined the OSM road centerline by visual inspection across the 13 grids (random select from around 7 % of the 200 grids) to generate the ground truth data. For the road that OSM has a low recall rate, we linked the road segment and added the missing road. For the incorrect roads, we corrected their location. The refined OSM road dataset served as a more complete and correct subset of the evaluation dataset. Then, we followed section 2.2.3 to generate the raster map for pixel-based matrix calculation. We compared the accuracy using the OSM and the refined OSM as in Fig. 10. The accuracy using OSM and refined OSM were highly consistent with an R of 0.999 and a significant P value, which confirmed the OSM capability for large-scale study.

The 10 m Sentinel-1 and Sentinel-2 satellite data sets are available and applicable for large-scale road extraction with relatively higher resolution. For the narrow roads that widths are less than 10 m, it is still discernible when the spatial context information is considered, but it is hard to delineate that in the area with high road density. A higher resolution like meter level may perform better results but limit the data coverage. The 10 m resolution dataset is available and efficient for large-scale road extraction. We further extracted the centerline from the 10 m road extraction result. The centerline results (Fig. 11) were consistent with the OSM principle road levels including the motorway, trunk, primary, secondary, tertiary road, and the residential road while has a low recall rate in special road types including the service, pedestrian, step, footway, path, track and bridleway. Some detailed roads in the residential blocks are missing for their resolution constraint in complex urban components. The results suggested the 10 m resolution centerline showed comparable performance with the OSM centerline, which confirmed its capability in a large-scale study.

5.3. *Large-scale road extraction application*

Spatial autocorrelation of accuracy can be observed in large-scale road extraction. For a large geographical context, the climate and terrain are two important dimensions that influence the road extraction result. As in the case of China, different terrain and climates shape the diverse landforms. There are spectral confusions and backscattering confusions in different land covers, like the agricultural land and mountainous area. In addition, for different urban gradients, the road density is different along the gradient.

As we can see, no single model outperforms in all the evaluation grids and matrices. We find the UU-Net outperforms in the majority of grids, which means it showed overall and improved performance regardless of the complex geographical environment. In some cases, S2-U-Net or S1-U-Net also show comparable performance. George Box stated that "All models are wrong, but some are useful" (Box 1979), which also hold in a large-scale study. Though the final UU-Net cannot outperform in all cases, the result indicates UU-Net, S1-U-Net and S2-U-Net have their advantages. It's suggested that for different geographical regions, different models are integrated to achieve better large-scale accuracy. In our study, S2-U-Net can be used in the mountainous area while S1-U-Net can be used in barren.

6. Conclusion

In this study, we used Sentinel-1 and Sentinel-2 with UU-Net for large-scale road extraction in terms of different social-economic and geographic regions. First, the synergistic use of SAR and optical data was better than the single sources in different accuracy matrices. UU-Net achieved the average F1 score of0.5502, comparing with S1-U-Net of 0.3749 and S2-U-Net of 0.5357. The road extraction accuracy of the three models showed similar patterns with higher accuracy in the southeastern part and lower accuracy in the northwestern part, indicating the spatial autocorrelation effects in large-scale analysis. Ordinary least square regression showed the road extraction accuracy has a positive relationship with the urban land cover type and a negative relationship with the elevation and barren land cover type. SAR

contributes to the road extraction in barren and cropland areas and optical data contributes to the large slope regions. The proposed UU-Net works effectively in large-scale road extraction, which can serve as the data support for road updating and traffic management.

CRediT authorship contribution statement

Yinyi Lin: Methodology, Investigation, Visualization, Writing - original draft. **Luoma Wan:** Methodology, Writing - review & editing. **Hongsheng Zhang:** Conceptualization, Supervision, Funding acquisition, Writing - review & editing. **Shan Wei:** Data curation, Writing - review & editing. **Peifeng Ma:** Supervision, Writing - review & editing. **Yu Li:** Writing - review & editing. **Zhuoyi Zhao:** Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This study was jointly supported by the National Natural Science Foundation of China (42022061, 42071390 and 41971278), the Research Grants Council (RGC) of Hong Kong (HKU27602020, HKU14605917 and CUHK14504219), and The University of Hong Kong (201909185015 and 202011159112).

References

Badrinarayanan, V., Kendall, A., Cipolla, R., 2017. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 39 (12), 2481–2495.

Bonafilia, D., Gill, J., Basu, S., Yang, D., 2019. Building high resolution maps for humanitarian aid and development with weakly-and semi-supervised learning. In: In, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 1–9.

Box, G.E.P., 1979. In: Robustness in Statistics. Elsevier, pp. 201–236. https://doi.org/10.1016/B978-0-12-438150-6.50018-2.

Center for International Earth Science Information Network - CIESIN - Columbia University, & Information Technology Outreach Services - ITOS - University of Georgia (2013). Global Roads Open Access Data Set, Version 1 (gROADSv1). In. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC).

Cervero, R.B., 2013. Linking urban transport and land use in developing countries. Journal of Transport and Land Use 6 (1), 7. https://doi.org/10.5198/jtlu.v6i110.5198/jtlu.v6i1.425.

Chai, J., Lu, Q.-Y., Wang, S.-Y., Lai, K.K., 2016. Analysis of road transportation energy consumption demand in China. Transportation Research Part D-Transport and Environment 48, 112–124.

Cheng, G., Wu, C., Huang, Q., Meng, Y.u., Shi, J., Chen, J., Yan, D., 2019. Recognizing road from satellite images by structured neural network. Neurocomputing 356, 131–141.

Demir, I., Koperski, K., Lindenbaum, D., Pang, G., Huang, J., Bast, S., Hughes, F., Tuia, D., Raskar, R., 2018. DeepGlobe 2018: A Challenge to Parse the Earth through Satellite Images. In: Proceedings 2018 Ieee/Cvf Conference on Computer Vision and Pattern Recognition Workshops (Cvprw), pp. 172–181.

Farr, T.G., Rosen, P.A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., Alsdorf, D., 2007. The shuttle radar topography mission. Rev. Geophys. 45 (2) https://doi.org/10.1029/2005RG000183.

Feng, R.D., Wang, F.Y., Wang, K.Y., Xu, S.J., 2021. Quantifying influences of anthropogenic-natural factors on ecological land evolution in mega-urban agglomeration: A case study of Guangdong-Hong Kong-Macao greater Bay area. J. Cleaner Prod. 283.

Ferraz, A., Mallet, C., Chehata, N., 2016. Large-scale road detection in forested mountainous areas using airborne topographic lidar data. ISPRS J. Photogramm. Remote Sens. 112, 23–36.

Friedl, M., & Sulla-Menashe, D. (2015). MCD12Q1 MODIS/Terra+ aqua land cover type yearly L3 global 500m SIN grid V006 . NASA EOSDIS Land Processes DAAC, 10.

Gamba, P., Dell'Acqua, F., Lisini, G., 2006. Improving urban road extraction in high-resolution images exploiting directional filtering, perceptual grouping, and simple topological concepts. IEEE Geosci. Remote Sens. Lett. 3 (3), 387–391.

Getis, A., & Ord, J.K. (1992). The Analysis of Spatial Association by Use of Distance Statistics. Geographical Analysis, 24, 189-206.

Goodchild, M.F., Parks, B.O., & Steyaert, L.T. (1993). Environmental modeling with GIS.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sens. Environ. 202, 18–27.

Hawbaker, T.J., Radeloff, V.C., Hammer, R.B., Clayton, M.K., 2005. Road density and landscape pattern in relation to housing density, and ownership, land cover, and soils. Landscape Ecol. 20 (5), 609–625.

Henry, C., Azimi, S.M., Merkle, N., 2018. Road Segmentation in SAR Satellite Images With Deep Fully Convolutional Neural Networks. IEEE Geosci. Remote Sens. Lett. 15 (12), 1867–1871.

Hu, H., 1935. Distribution of China's population. Acta Geol. Sin. 2, 33–74.

Jia, C.L., Ji, K.F., Jiang, Y.M., Kuang, G.Y., 2005. Road extraction from high-resolution SAR imagery using hough transform. IGARSS 2005: IEEE International Geoscience and Remote Sensing Symposium.

Kasraian, D., Maat, K., Stead, D., van Wee, B., 2016. Long-term impacts of transport infrastructure networks on land-use change: an international review of empirical studies. Transport Reviews 36 (6), 772–792.

Kearney, S.P., Coops, N.C., Sethi, S., Stenhouse, G.B., 2020. Maintaining accurate, current, rural road network data: An extraction and updating routine using RapidEye, participatory GIS and deep learning. Int. J. Appl. Earth Obs. Geoinf. 87, 102031. https://doi.org/10.1016/j.jag.2019.102031.

Kingma, D.P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Kottek, M., Grieser, J., Beck, C., Rudolf, B., & Rubel, F. (2006). World map of the Köppen-Geiger climate classification updated.

Lin, Y., Zhang, H., Lin, H., Gamba, P.E., Liu, X., 2020. Incorporating synthetic aperture radar and optical images to investigate the annual dynamics of anthropogenic impervious surface at large scale. Remote Sens. Environ. 242, 111757. https://doi.org/10.1016/j.rse.2020.111757.

Lisini, G., Gamba, P., Dell'Acqua, F., Holecz, F., 2011. First results on road network extraction and fusion on optical and SAR images using a multi-scale adaptive approach. International Journal of Image and Data Fusion 2 (4), 363–375.

Liu, J., Qin, Q.M., Li, J., Li, Y.P., 2017. Rural Road Extraction from High-Resolution Remote Sensing Images Based on Geometric Feature Inference. ISPRS Int. J. Geo-Inf. 6.

Liu, Y., Yao, J., Lu, X., Xia, M., Wang, X., Liu, Y., 2019. RoadNet: Learning to comprehensively analyze road networks in complex urban scenes from high-resolution remotely sensed images. IEEE Trans. Geosci. Remote Sens. 57 (4), 2043–2056.

Lu, X., Zhong, Y., Zheng, Z., Zhang, L., 2021. GAMSNet: Globally aware road detection network with multi-scale residual learning. ISPRS J. Photogramm. Remote Sens. 175, 340–352.

Lu, X., Zhong, Y., Zheng, Z., Liu, Y., Zhao, J.i., Ma, A., Yang, J., 2019. Multi-Scale and Multi-Task Deep Learning Framework for Automatic Road Extraction. IEEE Trans. Geosci. Remote Sens. 57 (11), 9362–9377.

Minaee, S., Boykov, Y., Porikli, F., Plaza, A., Kehtarnavaz, N., & Terzopoulos, D. (2020). Image segmentation using deep learning: A survey. arXiv preprint arXiv: 2001.05566.

Mnih, V., 2013. Machine learning for aerial image labeling. Citeseer.

Mnih, V., Hinton, G.E., 2010. Learning to Detect Roads in High-Resolution Aerial Images. *Computer Vision - Eccv 2010*. Pt Vi 6316, 210–223.

Oltean-Dumbrava, C., Watts, G., Miah, A., 2013. Transport infrastructure: making more sustainable decisions for noise reduction. J. Cleaner Prod. 42, 58–68.

Papandreou, G., Chen, L.-C., Murphy, K.P., Yuille, A.L., 2015. Weakly-and semi-supervised learning of a deep convolutional network for semantic image segmentation. In: In, *Proceedings of the IEEE international conference on computer vision*, pp. 1742–1750.

Qin, Y., Xiao, X., Dong, J., Zhang, Y., Wu, X., Shimabukuro, Y., Arai, E., Biradar, C., Wang, J., Zou, Z., Liu, F., Shi, Z., Doughty, R., Moore, B., 2019. Improved estimates of forest cover and loss in the Brazilian Amazon in 2000–2017. Nat. Sustainability 2 (8), 764–772.

Quin, G., Pinel-Puyssegur, B., Nicolas, J.-M., Loreaux, P., 2014. MIMOSA: An Automatic Change Detection Method for SAR Time Series. IEEE Trans. Geosci. Remote Sens. 52 (9), 5349–5363.

Ren, Y., Yu, Y., Guan, H., 2020. DA-CapsUNet: A Dual-Attention Capsule U-Net for Road Extraction from Remote Sensing Imagery. Remote Sensing 12 (18), 2866. https://doi.org/10.3390/rs12182866.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation. Medical Image Computing and Computer-Assisted Intervention, Pt Iii 9351, 234–241.

Saito, S., Yamashita, T., Aoki, Y., 2016. Multiple Object Extraction from Aerial imagery with Convolutional Neural Networks. J. Imaging Sci. Technol. 60.

Shi, W., Miao, Z., Debayle, J., 2014. An Integrated Method for Urban Main-Road Centerline Extraction From Optical Remotely Sensed Imagery. IEEE Trans. Geosci. Remote Sens. 52 (6), 3359–3372.

Song, M., Civco, D., 2004. Road extraction using SVM and image segmentation. Photogramm. Eng. Remote Sens. 70 (12), 1365–1371.

Stewart, C., Lazzarini, M., Luna, A., Albani, S., 2020. Deep Learning with Open Data for Desert Road Mapping. Remote Sensing 12 (14), 2274. https://doi.org/10.3390/rs12142274.

Sujatha, C., Selvathi, D., 2015. Connected component-based technique for automatic extraction of road centerline in high resolution satellite images. EURASIP Journal on Image and Video Processing 2015, 1–16.

Sun, C., Luo, Y., Li, J., 2018. Urban traffic infrastructure investment and air pollution: Evidence from the 83 cities in China. J. Cleaner Prod. 172, 488–496.

Venkatappa, M., Sasaki, N., Shrestha, R.P., Tripathi, N.K., Ma, H.O., 2019. Determination of Vegetation Thresholds for Assessing Land Use and Land Use Changes in Cambodia using the Google Earth Engine Cloud-Computing Platform. Remote Sensing 11.
Wang, J.L., Qian, J.H., Ma, R.B., 2013. Urban Road Information Extraction from High Resolution Remotely Sensed Image based on Semantic Model. 2013 21st International Conference on Geoinformatics (Geoinformatics).
Wang, W., Yang, N., Zhang, Y.i., 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.
Wei, Y., Zhang, K., Ji, S., 2020. Simultaneous Road Surface and Centerline Extraction From Large-Scale Remote Sensing Images Using CNN-Based Segmentation and Tracing. IEEE Trans. Geosci. Remote Sens. 58 (12), 8919–8931.
Wu, S.B., Du, C., Chen, H., Xu, Y.X., Guo, N., Jing, N., 2019. Road Extraction from Very High Resolution Images Using Weakly labeled OpenStreetMap Centerline. ISPRS Int. J. Geo-Inf. 8.
Xie, R., Fang, J., Liu, C., 2017. The effects of transportation infrastructure on urban carbon emissions. Appl. Energy 196, 199–207.
Zeng, T., Gao, Q., Ding, Z., Chen, J., Li, G., 2019. Road Network Extraction From Low-Contrast SAR Images. IEEE Geosci. Remote Sens. Lett. 16 (6), 907–911.
Zhang, Q., Kong, Q., Zhang, C., You, S., Wei, H., Sun, R., Li, L.i., 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.
Zhang, Y., Zhang, H., Lin, H., 2014. Improving the impervious surface estimation with combined use of optical and SAR remote sensing images. Remote Sens. Environ. 141, 155–167.
Zhang, Z., Liu, Q., Wang, Y., 2018. Road Extraction by Deep Residual U-Net. IEEE Geosci. Remote Sens. Lett. 15 (5), 749–753.
Zhao, P., 2010. Sustainable urban expansion and transportation in a growing megacity: Consequences of urban sprawl for mobility on the urban fringe of Beijing. Habitat International 34 (2), 236–243.
Zhao, Z.-Q., Zheng, P., Xu, S.-T., Wu, X., 2019. Object Detection With Deep Learning: A Review. IEEE Trans. Neural Networks Learn. Syst. 30 (11), 3212–3232.
Zheng, P., 2006. China's Geography. 五洲传播出版社.
Zhou, L.C., Zhang, C., Wu, M., 2018. D-LinkNet: LinkNet with Pretrained Encoder and Dilated Convolution for High Resolution Satellite Imagery Road Extraction. In: Proceedings 2018 Ieee/Cvf Conference on Computer Vision and Pattern Recognition Workshops (Cvprw), pp. 192–196.
Zhou, M., Sui, H., Chen, S., Wang, J., Chen, X.u., 2020. BT-RoadNet: A boundary and topologically-aware neural network for road extraction from high-resolution remote sensing imagery. ISPRS J. Photogramm. Remote Sens. 168, 288–306.