AIRPORT RUNWAY SEMANTIC SEGMENTATION BASED ON DCNN IN HIGH SPATIAL RESOLUTION REMOTE SENSING IMAGES

Men Zecheng[1,2,3], Jiang Jie[2,*], Guo Xian[2], Chen Lijun[3], Liu Dongsheng[4,5]

[1] College of Geoscience and Surveying Engineering, China University of Mining and Technology (Beijing), Beijing, China - menzec@qq.com
[2]School of Geomatics and Urban Spatial Informatics, Beijing University of Civil Engineering and Architecture, Beijing, China – (jiangjie, guoxian)@bucea.edu.cn
[3]National Geomatics Center of China, Beijing, China - chenlj@ngcc.cn
[4]School of Geological and Surveying & Mapping Engineering, Chang'an University, ShaanXi Province, China - liudongsheng@piesat.cn
[5]Beijing PIESAT Information Technology Co., Ltd. Beijing, China

KEY WORDS: Airport runway, Semantic Segmentation, DeepLabv3, Lovasz-softmax, Remote Sensing

ABSTRACT:

Due to the diverse structure and complex background of airports, fast and accurate airport detection in remote sensing images is challenging. Currently, airport detection method is mostly based on boxes, but pixel-based detection method which identifies airport runway outline has been merely reported. In this paper, a framework using deep convolutional neural network is proposed to accurately identify runway contour from high resolution remote sensing images. Firstly, we make a large and medium airport runway semantic segmentation data set (excluding the south Korean region) including 1,464 airport runways. Then DeepLabv3 semantic segmentation network with cross-entropy loss is trained using airport runway dataset. After the training using cross-entropy loss, lovasz-softmax loss function is used to train network and improve the intersection-over-union (IoU) score by 5.9%. The IoU score 0.75 is selected as the threshold of whether the runway is detected and we get accuracy and recall are 96.64% and 94.32% respectively. Compared with the state-of-the-art method, our method improves 1.3% and 1.6% of accuracy and recall respectively. We extract the number of airport runway as well as their basic contours of all the Korean large and medium airports from the remote sensing images across South Korea. The results show that our method can effectively detect the runway contour from the remote sensing images of a large range of complex scenes, and can provide a reference for the detection of the airport.

1. INTRODUCTION

Remote sensing image object detection technology has attracted massive attention, especially in the fields of urban management, agriculture and military. As one of the most important facilities, the accurate detection of airports has attracted widespread concerns. However, it is challenging to accurately detect airport in remote sensing images with the diverse structure and complex background.

Due to the characteristics of large aspect ratio and internal gray uniformity of the airport, the runway is the most discriminating feature of airport. Many methods have been proposed based on airport runway to detect airport from remote sensing images in recent years. According to the characteristics used, they can be divided into two categories: 1. using features designed by prior knowledge to extract airports(Tang, 2015; Zhu, 2015), 2. using features automatically extracted from convolutional neural networks(Xiao, 2017; Zhu, 2018).

The methods with features of artificial design are mainly divided into two kinds: line detectors (Tang, 2015) and saliency models (Zhu, 2015). The advantages of line detectors based methods are fast and low complexity, but they are easily to be disturbed by complex background. Compared with methods based line detectors, the ways using saliency models are robust. But their sliding windows will introduce extra overlap redundancy. line detectors and saliency models employ artificial features that heavily depend on prior knowledge. Instead of using human-designed features, deep convolutional neural networks(DCNNs) is designed to extract low-level and high-level features and have been applied to detecting airports. Xiao etc. (Xiao, 2017) extract the multi-scale fusion features of airport using GoogleNet-LF model. But GoogleNet-LF model repeatedly calculates features of the inner area of the airports, resulting in massive additional computation. Xu etc.(Zhu, 2018) design a cascade region proposal networks that locate airport directly, which is an end-to-end way to detect airport. However, these methods using features extracted by human and DCNNs are box-based to locate airport currently, which could not identify outline of airport.

In this paper, we propose a framework that the precise outline of airport runway can be identified from remote sensing images using DeepLabv3. Then we use lovasz-softmax loss instead of cross cross-entropy to improve our accuracy by nearly 6% during training DeepLabv3 network. In order to validate effectiveness of our method, we extract the number of Korean large and medium airports runway as well as their basic contours using remote sensing images across South Korea.

The remainder of this paper is organized as follows. In Section 2, we briefly introduce our recognition framework including DeepLabv3 and lovasz-softmax loss. In Section 3, we evaluate the multiple segmentation networks and lovasz-softmax loss with DeepLabv3 as well as present experimental results. This paper is concluded in Section 4.

[*]Corresponding author

https://doi.org/10.5194/isprs-archives-XLII-3-W10-361-2020 | © Authors 2020. CC BY 4.0 License.

2. METHOD

2.1 DeepLabv3 Semantic Segmentation Network

The Fully Convolutional Network(FCN) developed by Long (Long, 2014) takes a natural image as input and predicts a segmentation map of the same size as input image. Based on FCN, many semantic segmentation networks with different net-structures were designed in recent years. Among them, the DeepLabv3 is one of the well performance networks, which can get rich contextual information by employing large receptive field. It is beneficial to detection airport runway using DeepLabv3, since airport runways have various scale and they are similar to texture feature with freeways.

2.1.1 Overview of DeepLabv3: DeepLabv3 (Chen, 2017) network takes an entire images as input and as pixel probability value as output. The structure of DeepLabv3 is shown in Figure 1. The network is mainly divided into two parts: features extracted networks and Atrous Spatial Pyramid Pooling(ASPP) module. As shown in Figure 1, we select Resnet50 to extract features. After inputting features into ASPP, 1*1 convolution layer contact multiscale features from ASPP and the pixel probability value with the same size as the input image is obtained. DeepLabv3 design a new module that adopts atrous convolution in cascade or in parallel to capture multi-scale context by adopting multiple atrous rates in features networks, which could handle the problem of segmenting objects at multiple scales.

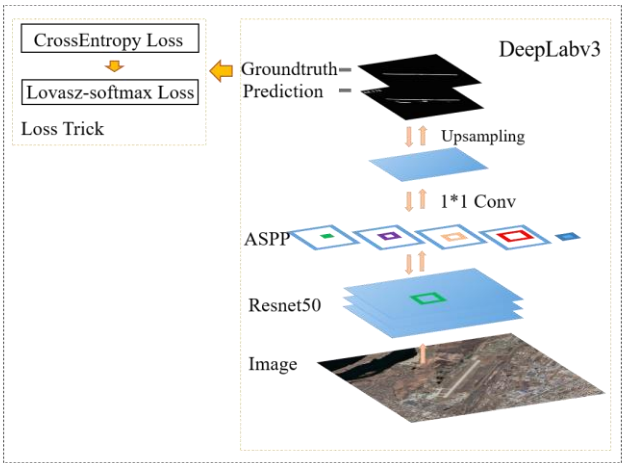


Figure 1. The structure of DeepLabv3

2.1.2 ASPP: ASPP use multiple parallel atrous convolutional layers with different sampling rates at the incoming convolutional feature layer. Thus, ASPP can capture objects as well as image context at multiple scales. ASPP are commonly take four rates (r = {6,12,18,24}) to obtain different scales features. In order to encode global context, image-level features are added to ASPP in deeplabv3.

Although accuracy of DeepLabv3+ (Chen, 2018) is higher than DeepLabv3, its model parameters size is much larger than Deeplabv3, as is shown in Table 1, which greatly increases the training and detection time. Since the recognition accuracy of DeepLabv3 can meet the requirements, this paper selects DeepLabv3 to identify the airport runway.

Model Name	DeepLabv3	DeepLabv3+
Backbone	Resnet50	Xception
Input Image Size (Pixel)	3*512*512	3*512*512
Params size (MB)	151.66	208.85
Forward/backward pass size (MB):	3916.15	12169.77
Estimated Total Size (MB):	4070.81	12381.62

Table 1 Parameters size of deeplabv3 and deeplabv3 plus

2.2 Lovasz-softmax Loss

Semantic segmentation networks classify pixel i of image into an object class $c \in C$. They commonly rely on logistic regression, optimizing cross entropy loss during training:

$$\text{loss}(\mathbf{f}) = -\frac{1}{p}\sum_{i=1}^{p} log f_i(y_i^*) \tag{1}$$

where p the number of cells of the image, y_i^* the real class of cell i, and $f_i(y_i^*)$ the probability that the network predicts that cell i belongs to the real class, and $\mathbf{f}$ a vector of all network outputs $f_i(c)$.

Because background pixels occupy most of the pixels in our training images, our dataset is a class unbalanced dataset. The cross entropy loss is based on integrals over the segmentation regions and may affect training performance and stability. Lovasz-softmax loss proposed by Berman (Berman, 2017) is based on the Jaccard index and tackle the problem of class unbalanced well. The Jaccard index, also known as the intersection-over-union (IoU) score, is used in our method.

$$\mathrm{J}_c(Y, y) = \frac{(Y=C)\cap(y=C)}{(Y=C)\cup(y=C)} \tag{2}$$

where Y the ground truth, y the predicted result, and the Jaccard index of category C $\mathrm{J}_c(Y, y)$. Since J_c is the ratio of the intersection of the prediction result and the ground truth result and the union, $\mathrm{J}_c \in [0,\ 1]$. Berman proposes a new loss function: lovasz-softmax

$$\Delta\mathrm{J}_c \ = 1 - \mathrm{J}_c(Y, y) \tag{3}$$

According to Berman's experiments, lovasz-softmax loss is superior to cross-entropy loss during training semantic segmentation networks. The performance of cross entropy loss and lovasz-softmax loss will be shown in chapter 3.4.1.

2.3 IoP and IoG Definitions

Different from the existing boxes-based airport detecting and locating methods, the semantic segmentation method used in this paper focus on extracting the runway in pixel-level. In order to compare with other methods conveniently, this paper defines two indices：IoP (Intersection-over-Prediction) and IoG (Intersection-over-GroundTruth)：

$$\mathrm{IoP}_o = \frac{(Y=O)\cap(y=O)}{(Y=O)} \tag{4}$$

$$\mathrm{IoG}_o = \frac{(Y=O)\cap(y=O)}{(y=O)} \tag{5}$$

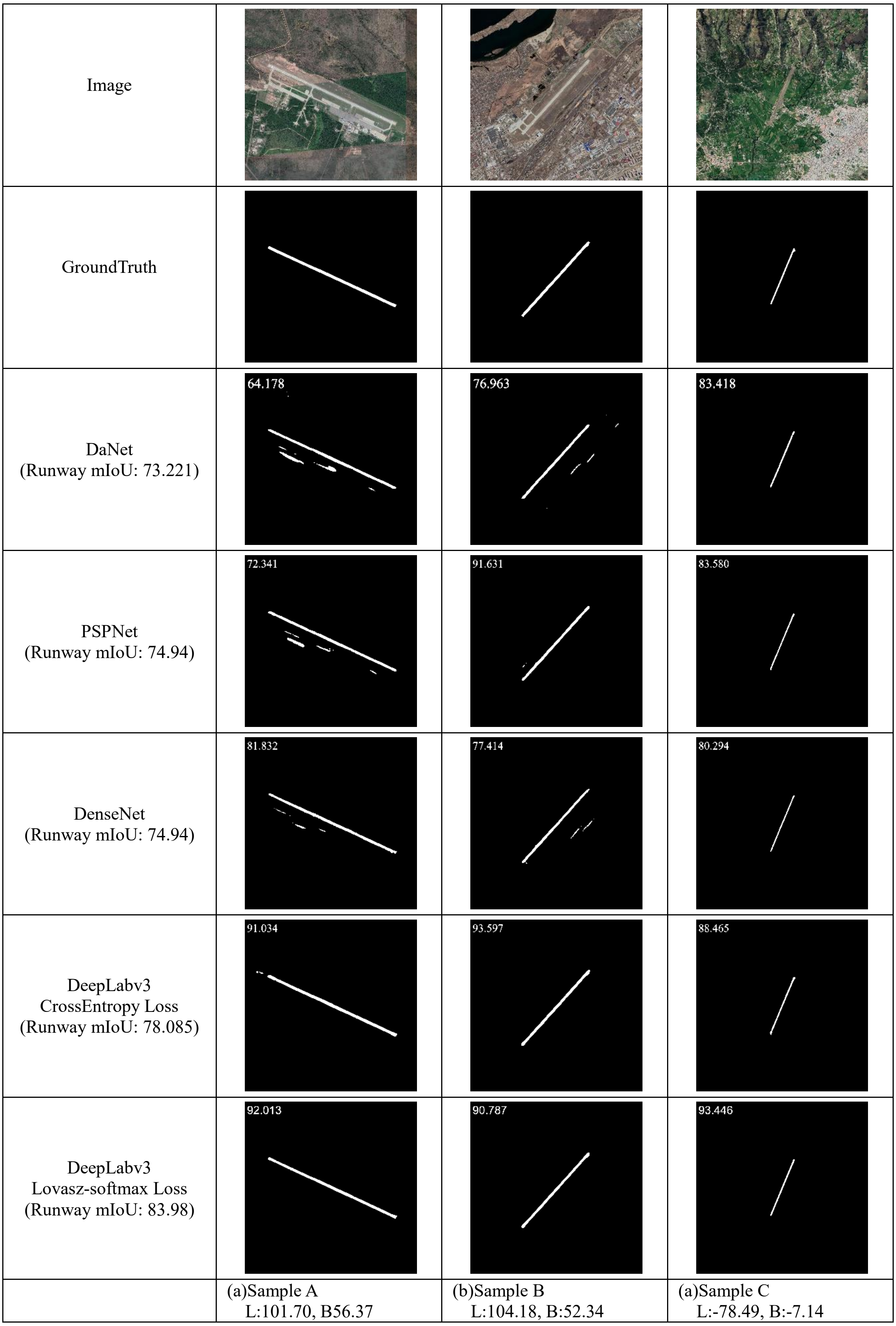


Figure 3 Results of different CNN networks (Runway mIoU is the mean IoU of result on validation set, the number in the upper left corner of the result image is the current result)

Where O the object of image, Y the GroundTruth of the airport object O and y the predicted result. We detect an airport runway from remote sensing images by most of runway pixels instead of all runway pixels. When defining 0.75 is the threshold whether the runway is detected:

$$R_{\mathrm{IoP}} = \begin{cases} \mathrm{IoP}_o > 0.75:\ 1, & Found \\ else:\ 0, & Not\ Found \end{cases} \tag{6}$$

$$R_{IoG} = \begin{cases} \mathrm{IoG}_o > 0.75:\ 1, & Found \\ else:\ 0, & Not\ Found \end{cases} \tag{7}$$

Based on above indices, the precision and recall rate can be calculated as:

$$\mathrm{Precision} = \sum R_{IoP} / N \tag{8}$$

$$\mathrm{Recall} = \sum R_{IoG} / N \tag{9}$$

Where *N* is the number of validation samples.

3. EXPERIMENT

3.1 Data

According to the global airport information database of OurAirport website including airport type as well as coordinates and so on, we collect 1300 remote sensing images containing large and medium airports from Google Earth (excluding South Korea for performance test in 3.4.3). The image size is 1536*1536 in spatial. Locations of airport samples are shown in Figure 2. Among them, 900 images are randomly chosen to train network and the rest are used for validation. The experimental environment included a Xeon Gold 5118 CPU, 64G of memory, and an NVIDIA Quadro P5000 graphics card (16 GB memory).

The entire networks were implemented using the Pytorch framework. We adopt Resnet50 as the basic feature extraction network (loading pre-training weights by ImageNet). The parameter of network training: batch size 2, learning rate 0.005 and learning rate is reduced to 10e-5 by polynomial decrement, the number of iterations 60,000, and the Loss function cross-entropy loss.

In order to avoid network overfitting, images are randomly flipped left or right, flipped up or down, and rotated by 0~45 degrees in advance. Afterwards, images are randomly scaled with a ratio of 0.75-1.25 as well as cut to 1120*1120.Finally images are normalized.

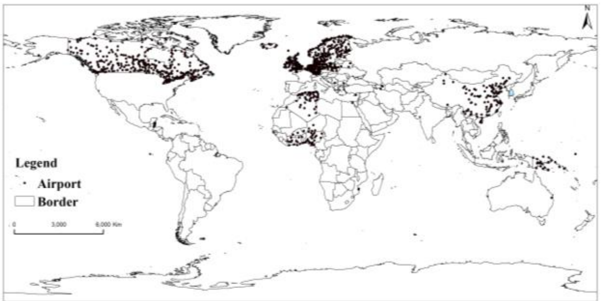


Figure 2 Location of airport sample

3.2 Results of Different Semantic Segmentation Models

In order to verify the effectiveness of the selected networks, this paper selects three widely-used networks to compare with DeepLabv3, 1. PSPNet: using the pyramid pooling module to aggregate context information of different regions; 2. DENSE_ASPP NET: connecting a set of dilated convolution in a dense manner; 3. DANET using a positional attention mechanism and a channel attention mechanism to enhance global feature fusion. These three networks were trained and validated on the same sample sets, and the data pre-processing ways same as DeepLabv3 are used to avoid network overfitting during training. The mean IoU (mIoU) of the three networks is compared with DeepLabv3 on the validation set and shown in Table 2. According to the results, DeepLabv3 is higher than PSPNet, DENSE_ASPP NET, DANET, respectively.

Model	Airport Runway mIoU (%)
DANET	73.22
PSPNet	74.80
DENSE_ASPP NET	74.94
DeepLabv3	78.09

Table 2 Performance of different networks

3.3 Analyze result of different network:

Figure 3 shows some typical runway detection results of the above semantic segmentation networks. In these experiments, mountains, trees, buildings, roads, waters, and other features in images are grouped in Background. Results of DaNet, PSPNet, DenseNet, DeepLabv3 with cross entropy loss are row 3 to 6 respectively. It is obvious that deeplabv3 result sample A and B have less false-alarm pixels than other networks. Performance of different sample A, B, C is shown in Table 3. Analyzing the structure of the different networks, the advantage of DeepLabv3 is that DeepLabv3 adds image-level features into ASPP and adopts atrous convolution in cascade in features extracted networks. Thus, DeepLabv3 captures more multiscale context to identify the pixel of airport runway.

Model	Sample A	Sample B	Sample C
DANET	64.18 %	76.96 %	83.42 %
PSPNet	72.34 %	91.63 %	83.58 %
DENSE_ASPP	81.83 %	77.41 %	80.29 %
DeepLabv3	91.03 %	93.60 %	88.47 %

Table 3 Performance of different Sample A, B, C

3.4 Discussion

3.4.1 Results of Different Loss Functions: The direct use of lovasz-softmax loss to train network will cause sharp gradient in the network and converge difficultly. Therefore, in this paper, lovasz-softmax loss was used to train network when the training of cross-entropy loss finished. The batch size, learning rate, and the number of iterations are 2, 0.0002, 30000 respectively. Data preprocessing ways are the same as before. The result of airport runway IoU in the validation set is shown in Table 4, which depicts that the use of lovasz-softmax loss improves the accuracy of the results by nearly 6% compared with that using cross entropy loss.

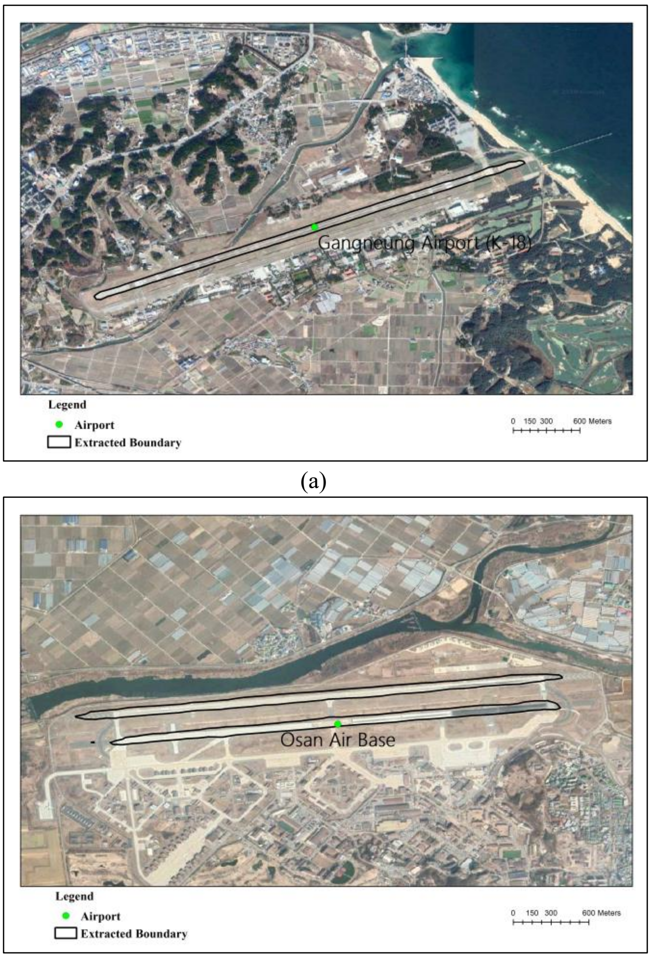
Loss Function	Airport Runway IoU(%)
cross-entropy Loss	78.09
Lovasz-softmax Loss	83.98

Table 4 Airport runway IoU of Cross-entropy Loss and Lovasz-softmax loss

3.4.2 Compared to Other Methods: In order to demonstrate the superiority of our method in airport detection, this paper compares the proposed DeepLabv3(with lovasz-softmax loss) with three state-of-the-art methods: a method based on geometry and texture(Tang, 2015), a Google-LF method that combines multi-scale information(Xiao, 2017), and an improved neural network object detection method(Xu, 2018). Since authors do not disclose the source code, it is hard to determine a large number of parameters used by their experiments through the papers. The accuracy of different methods is shown in Table 4, and the precision and recall values are from corresponding original papers. Compared with Xu, the accuracy and recall of this paper are increased by 1.6% and 1.3% respectively. The reason is that the ASPP module of DeepLabv3 uses multiple rates atrous convolutional layers at the incoming convolutional feature layer and encodes multi-scale features of image. The number of airport in dataset of Tang, Xiao, Xu is 170, 403, 400 respectively. Therefore, the accuracy improvement also benefits from our large dataset containing 1300 airports.

Method	Precision(%)	Recall(%)
Tang (2015)	93.65	83.07
Xiao (2017)	92.29	91.56
Xu (2018)	95.00	93.10
Ours	96.64	94.32

Table 5 Comparison of airport detection results



(a)

(b)

Figure 4 (a) and (b) are the extraction results of airport Gangneung airport and Osan air base respectively

3.4.3 Performance: In order to test the performance of the model when applied to a wide range of remote sensing images, this paper identifies the airport runways throughout Korea (Our training data does not include Korean airports). The remote sensing images of South Korea were downloaded from Google Earth. We successfully detected all 32 large and medium airports of global airport information database in Korea according to the result, which further demonstrates the practicality of the proposed method. Results of Gangneung airport and Osan air-base are taken as representative and shown in Figure 4. The runway results are converted from the binary map to the airport runway vector profile as the final results. Model also identifies 8 small airport runways of global airport information database at a resolution of 4 m. However, due to the narrow runways of many small airports, it is difficult to extract all the airport runway whit 4m resolution images. If we extract small airport runway, we use remote sensing images with a resolution higher than 4m.

4. CONCLUSION

Based on the worldwide large and medium airport runway semantic segmentation dataset, this paper uses the DeepLabv3 semantic segmentation network to identify the quantity of the airport runways from single remote sensing image, as well as to extract their basic contours. Accuracy of DeepLabv3 is improved by nearly 6% by using the lovasz-softmax loss function instead of cross-entropy loss during training. Results also show that our model identifies all the large and medium airport runways in the images acquired across the whole South Korea, indicating the effectiveness of our model.

Due to the diversity of the airport's appearance, it is difficult to determine the limits of the two ends of the airport runway when we mark samples, which increases the difficulty of training models. Future research can be based on two aspects: refine airport runway labeling rules and reduce model parameters.

REFERENCES

Berman, M., Triki, A.R., B. Blaschko, M., 2017. The Lov´asz-Softmax loss A tractable surrogate for the optimization of the intersection-over-union measure in neural networks.

Chen, L.-C., Papandreou, G., Schroff, F., Adam, H., 2017. Rethinking Atrous Convolution for Semantic Image Segmentation.

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H., 2018. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation.

Long, J., Shelhamer, E., Darrell, T., 2014. Fully convolutional networks for semantic segmentation. *IEEE Transactions on Pattern Analysis & Machine Intelligence* 39 (4), 640–651.

Tang, G., Xiao, Z., Liu, Q., Liu, H., 2015. A Novel Airport Detection Method via Line Segment Classification and Texture Classification. *IEEE Geoscience and Remote Sensing Letters* 12 (12), 2408–2412.

Xiao, Z., Gong, Y., Long, Y., Li, D., Wang, X., Liu, H., 2017. Airport Detection Based on a Multiscale Fusion Feature for Optical Remote Sensing Images. *IEEE Geoscience and Remote Sensing Letters* 14 (9), 1469–1473.

Xu, Y., Zhu, M., Li, S., Feng, H., Ma, S., Che, J., 2018. End-to-

End Airport Detection in Remote Sensing Images Combining Cascade Region Proposal Networks and Multi-Threshold Detection Networks. *Remote Sensing* 10 (10), 1516.

Zhu, D., Wang, B., Zhang, L., 2015. Airport Target Detection in Remote Sensing Images: A New Method Based on Two-Way Saliency. *IEEE Geoscience and Remote Sensing Letters* 12 (5), 1096–1100.

Zhu, M.M., Xu, Y.L., Ma, S.P., Ma, H.Q., 2018. Airport Detection on Remote Sensing Images Using Fater Region-based Convolutional Neural Network. *Journal of Physics: Conference Series* 1060, 12037.