Computers, Materials & Continua CMC, vol.65, no.3, pp.2455-2474, 2020

Detection of Precipitation Cloud over the Tibet Based on the Improved U-Net

Runzhe Tao[1, *], Yonghong Zhang[1], Lihua Wang[1], Pengyan Cai[1] and Haowen Tan[2]

Abstract: Aiming at the problem of radar base and ground observation stations on the Tibet is sparsely distributed and cannot achieve large-scale precipitation monitoring. U-Net, an advanced machine learning (ML) method, is used to develop a robust and rapid algorithm for precipitating cloud detection based on the new-generation geostationary satellite of FengYun-4A (FY-4A). First, in this algorithm, the real-time multi-band infrared brightness temperature from FY-4A combined with the data of Digital Elevation Model (DEM) has been used as predictor variables for our model. Second, the efficiency of the feature was improved by changing the traditional convolution layer serial connection method of U-Net to residual mapping. Then, in order to solve the problem of the network that would produce semantic differences when directly concentrated with low-level and high-level features, we use dense skip pathways to reuse feature maps of different layers as inputs for concatenate neural networks feature layers from different depths. Finally, according to the characteristics of precipitation clouds, the pooling layer of U-Net was replaced by a convolution operation to realize the detection of small precipitation clouds. It was experimentally concluded that the Pixel Accuracy (PA) and Mean Intersection over Union (MIoU) of the improved U-Net on the test set could reach 0.916 and 0.928, the detection of precipitation clouds over Tibet were well actualized.

Keywords: U-net, fy-4a, precipitation cloud, dense skip connections, residual network.

1 Introduction

As the most common cause of mountain disasters, rainfall requires expeditious monitoring and forecasting. By extension, the detection of rainfall clouds and the prediction of rainfall in alpine regions, such as Tibet, is critical. Many techniques supporting these requirements have been developed due to the wide distribution of observation base stations and abundant observation methods in the plain area. Existing methods mainly include Numerical Weather Prediction (NWP) [Forbes, Haiden and Magnusson (2015); Zsoter, Pappenberger and Richardson (2015)] and extrapolation methods based on radar echoes [Otsuka, Tuerhong and Richardson (2016); Shi, Li, Gu et

[1] School of Automation, Nanjing University of Information Science and Technology, Nanjing, 210044, China.

[2] Department of Computer Engineering, Chosun University, Gwangju, 501759, Korea.

[*] Corresponding Author: Runzhe Tao. Email: taorunzhe@sina.com.

Received: 13 May 2020; Accepted: 21 July 2020

al. (2018)], but they have limitations in the plateau areas. Due to the complex dynamic processes, physical processes and sparse observations on the plateau area, NWP has huge model uncertainty [He and Li (2013)]. Although rainfall forecasting by radar echo map is an effective method, limitations imposed by tough natural environment, complex terrain and high average elevation, the presence of only three radar base stations on Tibet. So, it is challenging to use radar for the estimation of precipitation over a wide range. As a solution, remote sensing images with high temporal resolution, which are not restricted by geographical environment, could be used in the Plateau areas for weather monitoring.

Most of the methods previously proposed to monitor precipitation via satellite images employ two distinctive features of precipitation clouds: brightness temperature and texture features. These methods study the relationship between brightness temperature and precipitation intensity and explore the frequency and spatial characteristics of images based on the texture information of the precipitation cloud, realizing the recognition of the precipitation cloud boundary [Biswas, Farrar, Gopalan et al. (2013); Zhou and Wu (2019)]. Although these methods are simple and efficient, the threshold of brightness temperature and frequency are easy to be interfered by the complex underlying surface. We see three drawbacks of the existing satellite image-based methods. First, many methods focus on the independent classification of individual pixels and rarely use spatial information that extends beyond adjacent pixels. Thus, these pixel-based approaches ignore the fact that clouds are a spatially continuous and highly dynamic phenomenon. Second, the process of identifying precipitation clouds is laborious, necessitating many experiments to deduce the relationship between precipitation and cloud clusters for determining the corresponding threshold. Furthermore, the detection results of precipitation cloud are sensitive to these thresholds, and are susceptible to interference from the underlying surface. With an average elevation is 4000 m, covered with ice and snow all year round, the threshold result determined by previous authors may not be applicable to the Tibetan area [Kan, Zhang, Zhu et al. (2018)]. Third, products derived from satellite data are required either in a timely manner for nowcasting or in the form of large time series for analysis. Rapid reaction to a situation demands processing time, which is particularly important to identify precipitation clouds, and this must be significantly shorter than the production rate of the corresponding satellite [Drönner, Korfhage, Egli et al. (2018)].

To address these three problems, we introduce the method of deep learning to detect and predict precipitation clouds in satellite images. Since 2014, with the rapid development of deep learning algorithms, great achievements have been made in image recognition, target detection, and the classification of remote sensing images [Long, Shelhamer and Darrell (2014); Yue, Zhao, Mao et al. (2015); Zhang, Duan, Sun et al. (2019)]. Compared with the traditional methods of pattern recognition, deep learning has a more powerful ability to perform feature learning and expression. It is, therefore, necessary to select the appropriate network from deep learning for the semantic segmentation of precipitation clouds over Tibet with two following considerations. On the one hand, FengYun-4(FY-4), as the only geostationary satellite with high temporal resolution covering the whole Tibetan area, was successfully launched on 11th December, 2016, and the observation data began to be released on 23rd April, 2018. However, GPM product was only updated to the end of 2018. Only a small number of sample data sets for model has be established

with the satellite images from April to December in 2018 of FY-4A. The semantic segmentation of remote sensing images is predominantly based on RGB three-channel image analysis. However, according to our previous precipitation statistics, the rainfall in Tibet mostly occurs at night, when the RGB three-channel cannot be used. Therefore, exploration of the relationship between the infrared channels of satellite images and precipitation clouds is critical.

Consequently, U-Net, with its concise structure and outstanding performance, shows great promise in image segmentation with small sample is selected as the backbone for our study [Ronneberger, Fischer and Brox (2015)]. Here, we have improved the structure of U-Net to detect precipitation clouds over Tibet with specific work being summarized as follows: (1) We use Global Precipitation Measurement (GPM) Mission product to label FY-4A satellite cloud images from the previous moment, so that the trained model can not only detect, but also predict precipitation clouds; (2) In view of the complex topographical conditions of Tibetan area, the real-time multi-band infrared of brightness temperature from FY-4A, combined with the data of Digital Elevation Model (DEM), has been used as predictor variables for our model; (3) Because of the size and multi-channel input satellite images, to make a full use of the features proposed by each convolution layer and to reduce the burden of deep network training, it is necessary to combine the U-Net with the residual network and change the serial connection mode of convolution layer into a form of residual mapping; (4) To solve the problem of the network would produce semantic differences that result in poor segmentation when directly concatenating low-level and high-level features, this paper designs standard convolution modules and a series of redesigned skip pathways between the encoder and decoder sub-networks instead of normal skip connections; (5) The pooling layer of down-sampling in the U-Net is improved to down-convolution, which reduces the loss of detailed information and improves the detection ability of small precipitation clouds.

Through experiments, we have concluded that the Pixel Accuracy (PA) and Mean Intersection over Union (MIoU) of the improved U-Net on the test set could reach 0.916 and 0.928, which is superior to the U-Net, SegNet [Badrinarayanan, Kendall and Cipolla (2017)] and DeeplabV3, suggesting that the improved U-Net can detect and forecast precipitation cloud over Tibetan area.

2 Study area

As shown in Fig. 1, the average altitude is high and the difference of altitude is great in the Tibetan area. Thus, the data of Digital Elevation Model (DEM) is integrated into the input data of network to consider the impact of different altitudes on the infrared band.

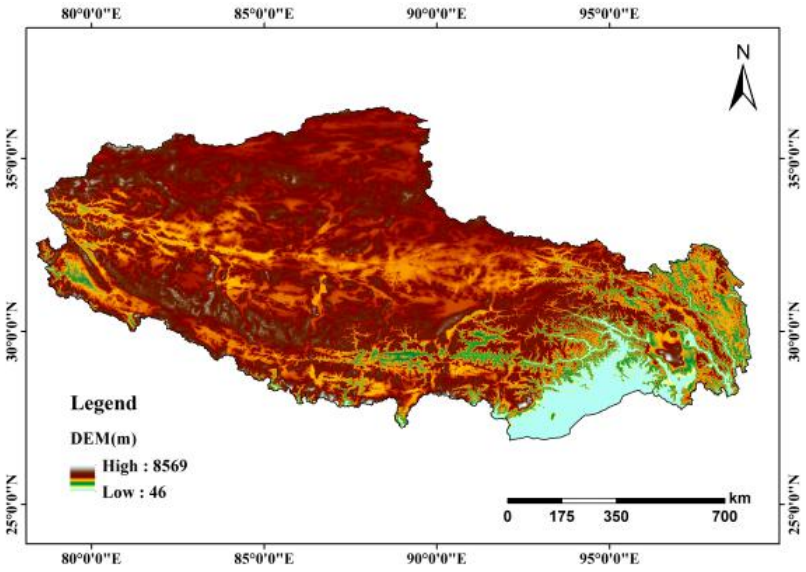


Figure 1: DEM data of Tibet

3 Method

Traditional CNNs processed computer vision tasks by taking a small area around the pixel as CNN's input [Zeng, Dai, Li et al. (2019); Liu, Yang, Lv et al. (2019)], and context information was ignored in feature extraction leading to low accuracy. To solve the above problems, FCN, SegNet and U-Net were applied in semantic segmentation of deep learning. Compared with the traditional CNNs, the FCN fused the multi-scale information by connecting and adding different feature maps. Although FCN implemented end-to-end training and learning, there were still many shortcomings, such as numerous network parameters, time-consuming, and low precision. The proposal of the U-Net had solved the above problems.

3.1 U-Net

U-Net was a deep learning algorithm for semantic segmentation based on the pixel level. The main network of this paper is improved on the basis of U-Net network, and the structure of modified U-Net proposed in this work is shown in Fig. 2.

As shown in Fig. 2, the modified U-Net network consists of a contracting path (left side) that extracts features and an expansive path (right side) that fuses features. The contracting path includes a plurality of standard convolution modules and down-sampling parts, where in each standard convolution module consists of the repeated application of two 3×3 convolution layers. One convolution layer is followed by a ReLU activation function, and a Batch Normalization layer. In each down-sampling part, there is a 3×3 convolution operation with stride 2 for double the number of feature channels at each down-sampling step. Therefore, the number of channels increases from input image is 32 to feature maps of the lowest network is 512. The feature map passing through one down-sampling layer is a scale, and modified U-Net network has 5 scale feature maps including

input image. Every step in the expansive is path consisted of an up-sampling layer, a concatenation with the correspondingly cropped feature map from the contracting path, and a standard convolution module. In up-sampling layer, an 3×3 up-convolution is used to extend the number of feature maps and improve resolution. Feature maps must be cropped due to the loss of border pixels of precipitation cloud and background in continuous convolution layers. At the final layer, 1×1 convolution followed by a sigmoid activation function is used to map each 16-component feature vector to the desired classes. The part of precipitation cloud feature extraction is considered to be an encoder, and the process of obtaining the score map is regarded as a decoder by pixel-level classification between precipitation cloud and background.

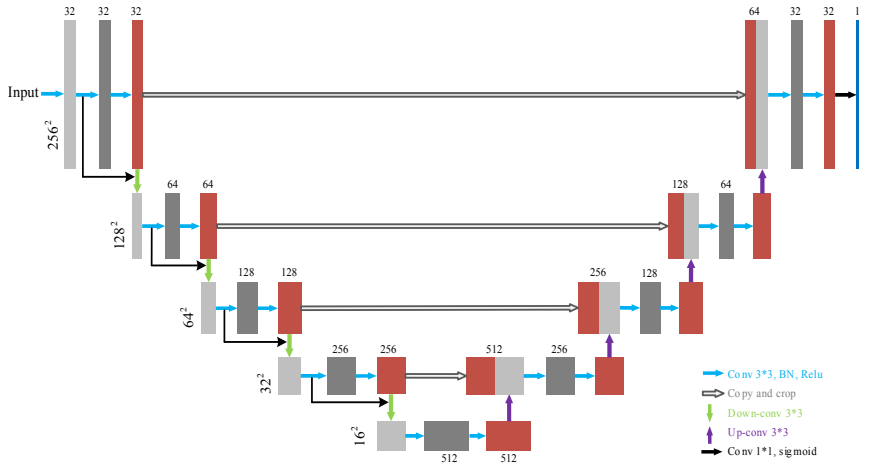


Figure 2: Modified U-Net structure

Continuous convolution layers in the contracting path are used to extract features of remote sensing imageries, and 4 down-sampling layers are able to realize multi-scale feature recognition of areas image with precipitation cloud. The expansive path fuses the output of up-sampling and the same-scale cropped feature maps of contracting path, allowing that the network replenishes context information lost due to convolution and pooling in the contracting path. Thus, the expansive path is able to propagate feature maps to higher resolution. Compared with the direct addition of features in FCN, which results in insufficient information of feature maps, a special feature fusion method is used in modified U-Net: concatenates features together in the channel dimension to form multi-scale feature maps with more low-level information.

3.2 Residual networks

Due to the input image size is fixed at 256×256 pixels, and the FY-4A satellite images contain multiple infrared channels, a deep U-Net is required to express the features accurately. However, with the deepening of the network structure produces two problems. The first is a vanishing/exploding gradient, which causes a fact that the training model is not easy to fit. Such problems can be solved by normalized initialization

and intermediate normalization layers. The second is the degradation phenomenon known as 'degradation'. With the development of residual network, 'degradation' has been effectively solved, residual learning (Fig. 3) has significantly reduced the burden of deep network training and has increased the number of network layers [He, Zhang, Ren et al. (2016); Wang, Jiang, Luo et al. (2019)].

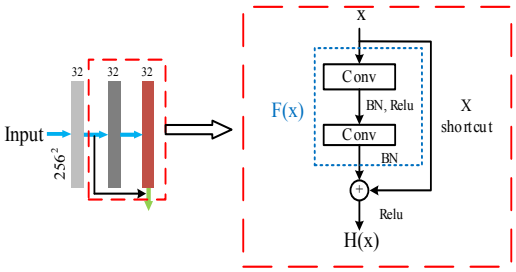


Figure 3: Shortcut connection

Residual structure is used to introduce a shortcut connection based on the traditional linear network structure and skip the connection of some layers, as the shortcut converges with the principal path through the method of additive fusion. After added shortcuts, the bottom errors in the training process can be propagated to the upper layer through the shortcut, which reduces the phenomenon of gradient disappearance caused by too many layers and improves the training accuracy.

3.3 Dense skip connections

In the U-Net, skip-connected feature fusion directly combines low-level features of the encoder with high-level features of the decoder [Huang, Wan, Liu et al. (2018)].

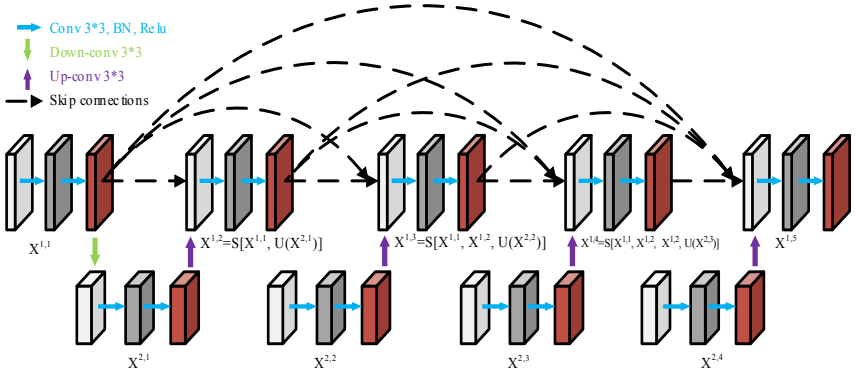


Figure 4: Dense skip connection structure

In the feature extraction stage, the underlying structure after multiple down sampling can capture low-level features, such as brightness temperature and texture of the areas image with precipitation cloud. These features are suitable for classification. After the concatenate operation, features in the same layer are passed directly from the encoder to the decoder, thus expanding the receptive fields to capture more abstract advanced

features, such as contours, shapes and other spatial position features. Such features provide accurate segmentation and positioning. Both low-level and high-level features have their own advantages. However, high-level features have weaker ability to reconstruct the binary predictions of original resolutions. It's quite possible that the network would produces semantic differences that result in poor segmentation when directly concatenating low-level and high-level features. In order to solve this problem, this paper designs standard convolution modules and a series of redesigned skip pathways between the encoder and decoder sub-networks instead of normal skip connections. Hence, this design is to capture fine-grained details of the background object before encoder's high-resolution feature maps merged with the corresponding semantically rich feature maps of the decoder. At the same time, dense skip pathways reuse feature maps of different layers as inputs to concatenate neural networks feature layers of different depths, which can reduce the negative impact of semantic differences and improve optimizer efficiency. The segmentation result is improved when the received encoder feature maps and the corresponding decoder feature maps are semantically similar. Dense skip connections consist of multiple standard convolution modules and up-sampling modules, where standard convolution modules and skip pathways are one of the cores in the entire improved U-Net. The nested dense skip connections (a scheme that connects long and short pathways) and previous information of input images are used together to restore information loss caused by down-sampling, and make all layers information fully disseminated in the network. Fig. 4 shows the structure of dense skip connections, and it can be seen that three standard convolution modules are added to the dense connection pathways between nodes $x^{1,1}$ and $x^{1,5}$. In front of each convolution layer, there is a concatenation layer that fuses outputs from the last convolution layers and the corresponding up-sampling outputs from standard convolution modules in the next layer.

Combining dense skip pathways with standard convolution modules, Fig. 4 shows that $x^{i,j}$ represents the output of node i, j, where i denotes the down-sampling layer index along the encoder and j indexes the convolution layer of the standard convolution module along the skip pathway. The formula of feature maps $x^{i,j}$ can be described as:

$$x^{i,j} = \begin{cases} S\left(x^{i,j}\right), & j = 1 \\ S\left(\left[x^{i,k}\right]_{k=1}^{j-1}, U\left(x^{i+1,j-1}\right)\right), & j \geq 2 \end{cases} \tag{1}$$

where [.] denotes a concatenation layer, $U(\cdot)$ represents a 3×3 deconvolution layer of up-sampling, function $S(\cdot)$ is a standard convolution module. Specifically, nodes at level $j = 1$ only receive one input from the down-sampling output of the previous layer in the encoder; nodes at level $j \geq 2$ receive a total of j inputs. In addition, j-1 inputs are from all node outputs of the same layer in the previous skip pathways, and the last remaining input is the up-sampling output of the node from the next skip pathways. Thus, the prior feature maps merge into the same concatenation layer by channels along the designedly dense skip pathways.

There are two main advantages in dense skip connections as follows: (1) During the backpropagation process, dense skip connections allow convolution block of each layer

to receive the gradient signal of the output back transfer for updating parameters, which inhibition the gradient disappearance; (2) After dense skip connections, a large number of features in the network can be reused, so that fewer standard convolution modules can be used to improve feature utilization and speed up training.

3.4 Removal of pooling

The pooling technique in U-Net aims to reducing computational complexity by keeping the statistics of a group of features instead of their original values. However, the pooling layer may not be effective for the extraction of cloud contour and the identification of small precipitation cloud clusters in this study [Chen, Chen and Lin (2018)]. Next, we illustrate the situation in more detail and provide simple examples to explain why the pooling layer may also not be effective here. Consider the following three matrices in Eq. (2), whose values represent the cloud density level (cloud thickness) of cloud clusters at certain locations. We know that the thicker the cloud is, the lower the brightness temperature will be, the more likely it is to produce rain. In the example, 0 is cloudless, 1 is a thin cloud without precipitation, 2 is cloud that may produce precipitation, and 4 is cumulonimbus with a thick layer of cloud.

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 4 & 4 & 1 \\ 1 & 4 & 4 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 2 & 2 & 0 \\ 2 & 4 & 4 & 2 \\ 2 & 4 & 4 & 2 \\ 0 & 2 & 2 & 0 \end{bmatrix} \begin{bmatrix} 2 & 2 & 2 & 2 \\ 2 & 2 & 2 & 2 \\ 2 & 2 & 2 & 2 \\ 2 & 2 & 2 & 2 \end{bmatrix} \tag{2}$$

Arguably the first two ‘Cloud’ are different because the latter comes with clouds may also produce precipitation on the periphery of the cloud. But if we apply a 2×2 maximum-pooling on the matrices, the first two matrices become indistinguishable, as follows in Eq. (3). The toy example demonstrates the potential harmfulness of applying maximum-pooling in terms of dropping some secondary ‘magnitude’ information.

$$\begin{bmatrix} 4 & 4 \\ 4 & 4 \end{bmatrix} \begin{bmatrix} 4 & 4 \\ 4 & 4 \end{bmatrix} \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix} \tag{3}$$

On the other hand, the last two ‘Cloud’ are also different as one has a larger cloud density near the center. If we apply a 2×2 average-pooling on the matrices, the last two become indistinguishable, as can be seen in Eq. (4). The toy example demonstrates the potential harmfulness of applying average-pooling in terms of dropping the ‘contrast’ information.

$$\begin{bmatrix} 1.75 & 1.75 \\ 1.75 & 1.75 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \tag{4}$$

Because of the potential harm of the pooling techniques on our recognition of precipitation cloud, we decide to replaced all pooling layers of the U-Net while using 3×3 convolution with stride 2, so the operation of down-sampling can be realized.

3.5 Improved U-Net

The optimization of network mainly depends on experience, which requires a lot of artificial simulation experiments. Thus, in this paper, the number of network structure layers in the improved U-Net must be optimally selected. If the number of network layers is too small, effective features such as non-raining clouds, snow and land surface may not

be extracted. However, too many network layers will result in redundant feature information. As the number of layers increases, effective information in each layer will lose. Through the experiments, the model of precipitation cloud segmentation in our paper selects a network structure with five-layer, and its basic structure-codec undergoes four down-samples and four up-samples.

Fig. 5 shows the details of the improved U-Net in this paper. The encoder includes 5 nodes that $x^{1,1}$, $x^{2,1}$, $x^{3,1}$, $x^{4,1}$, $x^{5,1}$, and the decoder includes 5 nodes that $x^{5,1}$, $x^{4,2}$, $x^{3,3}$, $x^{2,4}$, $x^{1,5}$, where the number of channels corresponding to $x^{1,1}$, $x^{2,1}$, $x^{3,1}$, $x^{4,1}$, $x^{5,1}$, is 16, 32, 64, 128, 256 respectively, and the number of channels in the same layer is the same. Moreover, all continuous convolution layers are outputted according to the standard convolution module, and for the convenience of drawing, the $x^{i,j}$ is used instead of the convolution module in Fig. 3.

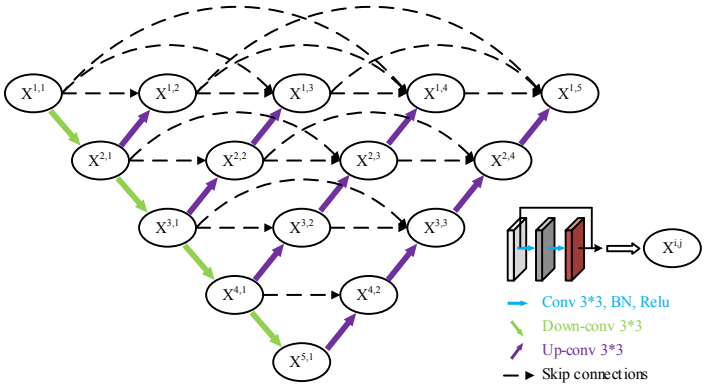


Figure 5: Improved U-Net network structure

Taking node $x^{1,1}$ in Fig. 5 as an example, we explain the specific process of improved U-Net to realize precipitation cloud segmentation. In our paper, the remote sensing satellite image with a size of 256×256×11 as the input. First of all, the original image passes through a standard convolution module to output $x^{1,1}$, and the number of convolution kernels is 32, so the size of the feature map is 256×256×32. In order to make full use of the feature information extracted from each convolution layer in the network and improve the training effect of the network, we change the serial connection of convolution layer into residual mapping superposition, and the final result of superposition is the output of standard convolution module. Secondly, the output $x^{1,1}$ is down-sampled by a 3×3 convolution with stride 2, thus, the feature map of pool1 with the size is 256×256×32. Then, we perform the operation of standard convolution module on the pool1 to obtain the output node $x^{2,1}$, which is a feature map with a size of 128×128×64. Finally, $x^{1,1}$ is used as the low-level feature input and $x^{2,1}$ is used as the high-level feature input to output $x^{1,2}$ with a size of 256×256×32 feature map.

4 Experiments

4.1 Datasets and label making

4.1.1 FY-4A data

Due to the rapid change of precipitation clouds, satellite images with high temporal resolution are required for monitoring. Currently, only the FY-4A satellite can provide high temporal resolution data over the Tibetan area. The specific indicators are shown in Tab. 1. Amassed statistics show that precipitation in Tibet occurs mostly at night, and therefore, infrared channels of FY-4A satellite are selected to use in this paper. Data preprocessing for these data includes radiometric calibration, projection, subset, etc.

Table 1: Characteristics of the FY-4A bands

Channel	Band No.	Central wavelength	Spatial resolutio	Primary Application
Visible & Near-Infrared	1	0.46	1	Aerosol
	2	0.64	0.5~1	Vegetation
	3	0.86	1	Cirrus
	4	1.38	2	Cloud
	5	1.61	2	Cloud, Snow
	6	2.25	2~4	Cloud, Aerosol
Shortwave Infrared	7	3.8	2	Fire, Land and surface
	8	3.8	4	Fire, Land and surface
Water Vapor	9	6.5	4	WV
	10	7.2	4	WV
Longwave Infrared	11	8.5	4	WV, Cloud
	12	10.8	4	Cloud
	13	12.0	4	SST, Cloud
	14	13.3	4	Cloud

4.1.2 The Global Precipitation Measurement (GPM) mission

Actual precipitation data is needed to label precipitation clouds on satellite images. In this paper, GPM IMERG (Integrated Multi-satellite Retrievals for GPM), a precipitation product that combines sensor with geosynchronous satellite data carried by GPM is used. This product is revised by the Rainfall Statistics of the International Meteorological and the time resolution is 30 minutes, but, the time of release is usually delayed by 3-4 months.

4.1.3 Label making

June, July and August are three months with heavy precipitation in Tibet. 150 moments of large-scale rainfall in Tibet from these three months in 2018 are randomly selected, and GPM data are downloaded at the matched moment. The selected GPM IMERG data represents the cumulative precipitation in half an hour. To recognize and predict the precipitation cloud clusters, we use GPM IMERG to label precipitation clouds in FY-4A images from the previous moment. The specific process is shown in Fig. 6. Fig. 6(a)

shows the GPM data of cumulative precipitation within 30 minutes at UTC: 201807231500-201807231530. Fig. 6(b) is the satellite image from FY-4A at UTC: 201807231500. We labelled 200 images of size is 512×256 pixels, and divided them into 400 images of size is 256×256 pixels. 300 images for training, and 100 for testing.

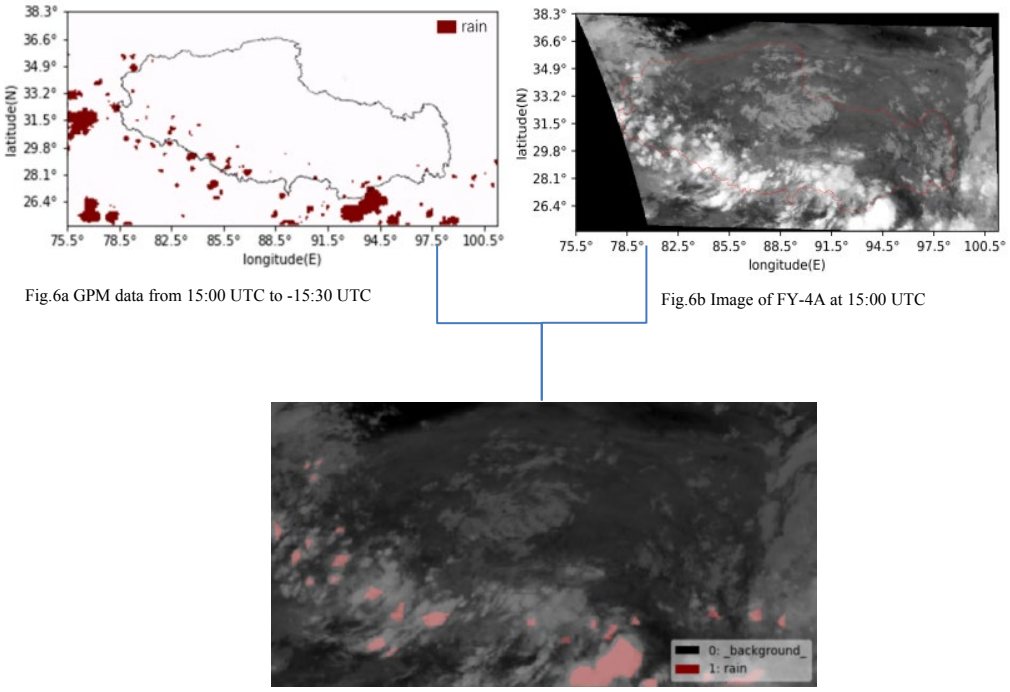


Fig.6a GPM data from 15:00 UTC to -15:30 UTC

Fig.6b Image of FY-4A at 15:00 UTC

Fig. 6c Label image

Figure 6: The process of label making

4.2 Implementation details

The experiments use the python language and Keras framework on NVIDIA GeForce GTX Titan V GPU, and our paper adopts Adam optimizer. Adam optimizer controls the weight distribution (momentum and current gradient) using an exponential decay rate β_1 with the coefficient 0.9, and controls the effect of the previous gradient's square by an exponential decay rate β_2 with the coefficient 0.999. The initial learning rate of the network is set to 0.01. In the process of training, the auto attenuation of learning rate is used. If the loss value of five successive epochs does not decrease, the learning rate will decrease by half, and the minimum value is 0.0001. Each training process includes a forward process and a back-propagation process. The former infers the prediction results that are compared with the real labels to generate loss values by combined dice-ce-loss function, while the latter updates the network weights based on the loss values by Adam optimizer. The combined-dice-ce-loss function [Zhou, Siddiquee, Tajbakhsh et al. (2018)] in this paper is a combination of binary cross-entropy loss function and dice-

coefficient-loss function, it is used as the loss function for the multi-level Deep Supervision. It is defined as follows:

$$L\left(Y,\hat{Y}\right)=-\frac{1}{N}\sum_{b=1}^{N}\left(\frac{1}{2}\cdot Y_b\cdot\log\hat{Y}_b+\frac{2\cdot\hat{Y}_b\cdot Y_b}{\hat{Y}_b+Y_b}\right) \tag{5}$$

where $\hat{Y}_b$ and Y_b denote the predicted probability and the ground truths of b_{th} image respectively, and N represents the batch size.

Metrics evaluation: To evaluate the performance of each semantic segmentation model in the cloud datasets of our paper, we adopt Pixel Accuracy (PA) and Mean Intersection over Union (MIoU). The corresponding formulas for the two different evaluation indexes are as follows:

$$PA=\frac{\sum_{i=0}^{k}P_{ii}}{\sum_{i=0}^{k}\sum_{j=0}^{k}P_{ij}} \tag{6}$$

$$MIOU=\frac{1}{k+1}\sum_{i=0}^{k}\frac{P_{ii}}{\sum_{j=0}^{k}P_{ij}+\sum_{j=0}^{k}P_{ji}-P_{ii}} \tag{7}$$

where i denotes the true value, and j is the predicted value, and k denotes the number of categories (excluding the background). p_{ij} denotes the number of pixels that originally belonged to class i but was predicted as class j, and p_{ii} denotes the number of pixels predicted to be true, thus, p_{ij} and p_{ji} denote false positives and false negatives, respectively. Since this article is a two-class segmentation, the MIoU calculates the intersection and union of real and predicted values of each category, then makes the ratio, and finally averages the results according to all categories (background).

4.3 Ablation experiment

Currently, only the FY-4A has high time resolution and 14 observation channels cover the whole area of Tibet in China. If three visible channels are removed, the remaining 11 channels of FY-4A are not all useful for the recognition and prediction of rainfall, so we performed ablation experiments on the input features to determine the optimal variable of model and the results are shown in Tab. 2. Tab. 2 indicates that the near-infrared band is useless for the prediction of precipitation clouds, and the shortwave infrared is useful. The main channels for recognition and prediction precipitation cloud are water vapor and long-wave infrared. Elevation data were used as a one-dimensional variable because the channel of 11.2μm was susceptible to surface and altitude. When we combined real-time multi-band infrared brightness temperature from FY-4A and DEM, we get the highest value of MIoU which is 0.916. Thus No. 7 in Tab. 2 could be used as predictor variables for network.

Table 2: Different variables of the input

No.	Channel	Variable (Unit)	MIoU in test set
1	Near-Infrared	$T_{1.38}$, $T_{1.61}$, $T_{2.25}$ (K)	0.48
2	Near+Shortwave Infrared	$T_{1.61}$, $T_{2.25}$, $T_{3.8}$ (K)	0.591
3	Water Vapor	$T_{6.5}$, $T_{7.2}$ (K)	0.795
4	Longwave Infrared	$T_{8.5}$, $T_{10.8}$, $T_{12.0}$, $T_{13.3}$ (K)	0.809
5	Bright temperature difference (BTD)	$\Delta T_{12.0\text{-}10.8}$, $\Delta T_{12.0\text{-}7.2}$, $\Delta T_{13.3\text{-}10.8}$ (K)	0.801
6	7 infrared channels +BTD	$T_{3.8}$, $T_{6.5}$, $T_{7.2}$, $T_{8.5}$, $T_{10.8}$, $T_{12.0}$, $T_{13.3}$ (K) $\Delta T_{12.0\text{-}10.8}$, $\Delta T_{12.0\text{-}7.2}$, $\Delta T_{3.8\text{-}10.8}$ (K)	0.907
7	7 infrared channels +BTD+ (DEM)	$T_{3.8}$, $T_{6.5}$, $T_{7.2}$, $T_{8.5}$, $T_{10.8}$, $T_{12.0}$, $T_{13.3}$ (K) $\Delta T_{12.0\text{-}10.8}$, $\Delta T_{12.0\text{-}7.2}$, $\Delta T_{3.8\text{-}10.8}$ (K) DEM (m)	0.916

4.4 Feature Visualization

To fully illustrate the role of residual structure in the improved U-Net, this paper uses feature visualization to carry out specific analyses. Fig. 7 is an input image containing precipitation clouds. The input image undergoes three convolutions in $x^{1,1}$ of the original U-Net. The results of convolution are shown in Fig. 8. The Fig. 9 illustrates the results extracted by $x^{1,1}$ after adding residual structure to the U-Net.

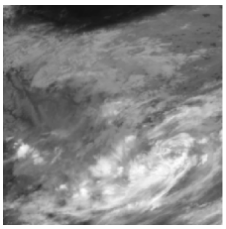


Figure 7: The image of input

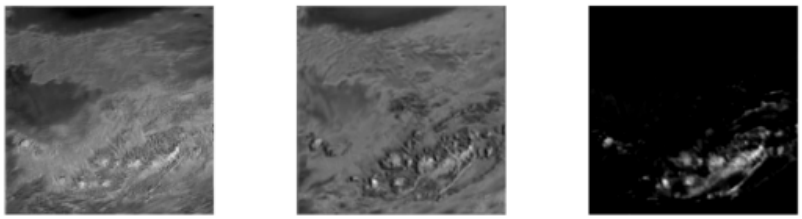


Figure 8: $x^{1,1}$ of the original U-Net

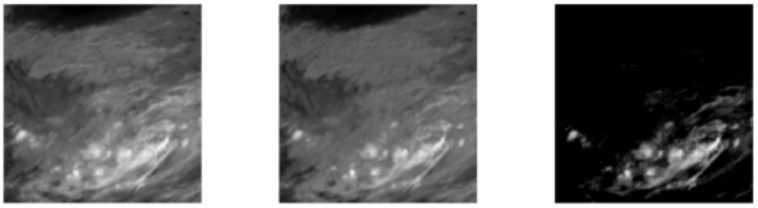
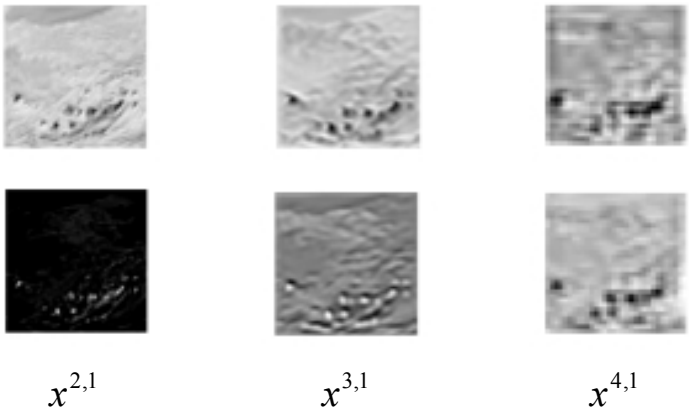


Figure 9: $x^{1,1}$ of the improved U-Net

Comparing Fig. 8 with Fig. 9, the input image undergoes three convolution operations in $x^{1,1}$, it can be seen that information loss has occurred in the original U-Net (many details are lost). This situation becomes more obvious with increase in network layers, which ultimately leads to the unsatisfactory effect of the model being trained by the U-Net. Because of the residual structure, the improved U-Net can automatically enhance the feature information after each convolution. Fig. 9 shows that the $x^{1,1}$ of the improved U-Net extracts more cloud information. Therefore, the use of residual structure to improve the U-Net significantly increases its feature extraction and network generalization ability. The features extracted from $x^{i,j}$ of the improved U-Net are analyzed and have the following characteristics.

From Fig. 10 we can see that the features learned from $x^{1,1}$ and $x^{2,1}$ are low-level features such as the edges of cloud clusters. The results of the $x^{3,1}$ extraction, however, are slightly more complex where texture features, such as some grid texture and layout features on the cloud, have been learned. $x^{4,1}$ learns complete and distinctive features, such as the outline of precipitation clouds. Through feature visualization, we can see that the feature extracted after convolution ignores the background and extracts only the key information. With an increase in network depth, superfluous information is lost as the results become more focused on the core features of classification and extracte precipitation clouds. The visualization of deconvolution is no longer shown in this paper.



$x^{2,1}$ $x^{3,1}$ $x^{4,1}$

Figure 10: Features extracted from each $x^{i,j}$

4.5 Analysis of result

As is shown in Tab. 3, firstly, we input the same dataset into the original U-Net, and the Miou and PA of U-Net are 85.7% and 87.3%, respectively. Secondly, compared with U-Net, the U-Net with residual structure has a certain improvement in accuracy. It can be seen that the MIoU and PA increasd to 87.4% and 88%. Then, we continue adding the structure of dense skip connections to the network. The recognition accuracy is significantly improved. MIoU and PA of U-Net are 90.2% and 91.1%. Finally, we further optimize the network structure and replace all pooling layers with down-convolution, and the final recognition of improved U-Net is the most accurate. We have analysed the effects of improved U-Net in the detection of precipitation clouds in a specific case on the test set. Fig. 11 is the label image at the 20180721200000UTC.

Table 3: Comparison of evaluation metrics on different models

Method	U-Net	U-Net+Residual structure	U-Net+Residual structure+Dense skip connections	U-Net+Residual structure+Dense skip connections+Down-conv (Improved U-Net)	SegNet	DeeplabV3
MIoU	0.857	0.874	0.902	0.916	0.861	0.891
PA	0.873	0.880	0.911	0.928	0.868	0.903

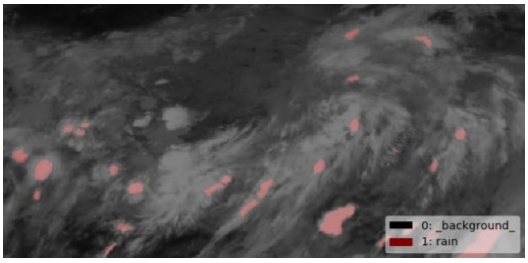


Figure 11: Label image at the 20180721200000UTC

The detection results of precipitation cloud using the trained original U-Net for the FY-4A image at UTC:20180721200000 are shown in Fig. 12.

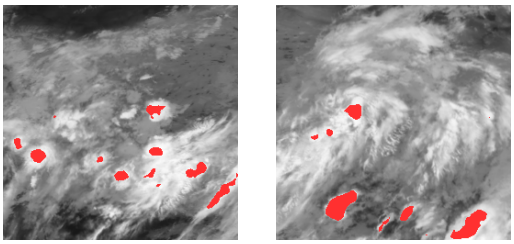


Figure 12: Segmentation result of using the original U-Net

We have found that many details were lost with increasing convolution network depth in original U-Net by feature visualization, and it is difficult or even impossible to recognize small or unclear targets. From the results in Fig. 12, it can be seen that only some large precipitation clouds can be extracted, many small precipitation clouds cannot be detected,

and the edge information of cloud clusters is poorly extracted. To solve the problem of unsatisfactory feature extraction in the U-Net and make a full use of the feature information extracted from the network, residual structure is added to each layer of the convolution block in the U-Net. The results of extraction are shown in Fig. 13.

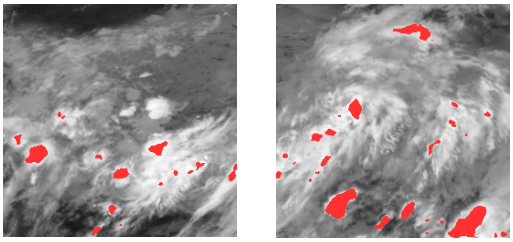


Figure 13: Segmentation result of U-Net with residual structure

From Fig. 13, we can see that the detection of precipitation clouds using U-Net with residual structure is improved, and some small precipitation cloud are identified. However, high-level features have weaker ability to reconstruct the binary predictions of original resolutions. It is quite possible that the network would produce semantic differences that result in poor segmentation when directly concatenating low-level and high-level features. The final recognition results have shown that the model is not ideal for extracting the contour information of precipitation cloud and there are many pixels misdetected as precipitation cloud.

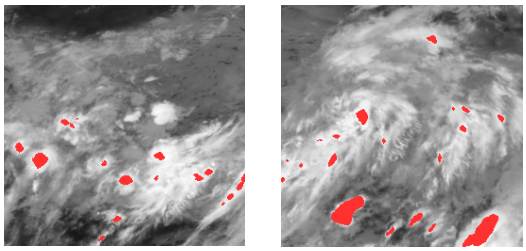


Figure 14: Segmentation result of U-Net+Residual structure+Dense skip connections

When we continue adding the structure of dense skip connections to the network, false detection pixels are significantly reduced in Fig. 14. However, the pooling layer will lose valuable details in the process of down-sampling, resulting in some small size precipitation clouds which only occupy a few pixels not being detected by the network.

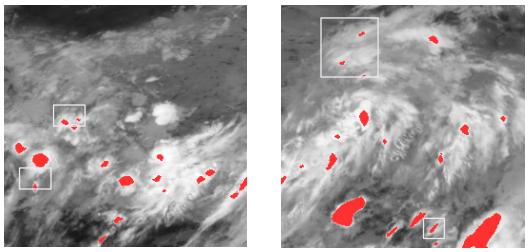
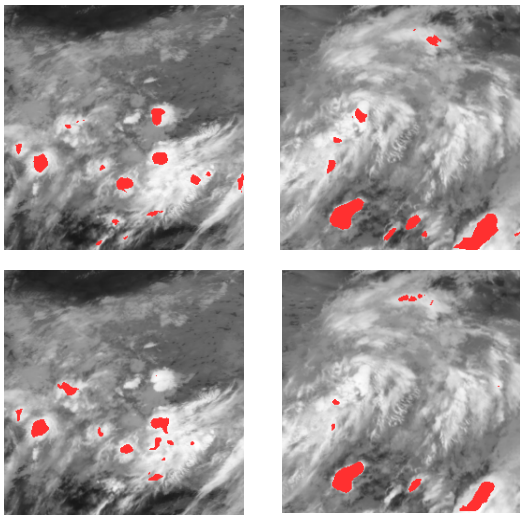


Figure 15: Segmentation result of the improved U-Net

Therefore, the pooling layer in the down-sampling process is changed to a 3×3 convolution with stride 2. From the results in Fig. 15, it can be seen that the improved model is more accurate for detection of small size precipitation clouds in detail (marked in white box). Overall, the extraction of edge information for precipitation clouds is significantly increased, and the extracted contour information of cloud clusters is more ideal. From the MIoU and PA on the test set, it can be seen that the improved U-Net proposed herein can effectively accomplish the detection of precipitation cloud over Tibetan area.

To test the segmentation performance of this study, different methods were used to segment the precipitation clouds, as shown in Figs. 16 and 17. As a new segmentation model, DeepLabV3 has achieved excellent performance on public data sets such as PASCAL VOC. SegNet as a classic segmentation model has a wide range of applications in many industries. Therefore, the above two kinds of segmentation models are selected for segmentation comparison on the test set.

Due to the bilinear interpolation method used in the up-sampling of DeeplabV3, the training time is short, but the accuracy of the model is low and the generalization ability is poor. Bilinear interpolation results in many inaccurate details, so the segmentation for contour of precipitation cloud and the small cloud are not ideal in Fig. 16(a).



(a) DeeplabV3

(b) SegNet

Figure 16: Segmentation result of DeeplabV3 and SegNet

The detection results of SegNet are shown in Fig. 16(b), although SegNet improves the segmentation effect and training efficiency compared with FCN, the segmentation of details is relatively poor. The indicators of MIoU and PA are similar to the original U-Net. In summary, through qualitative comparison, improved U-Net is superior to other models for detection of precipitation cloud based on satellite image.

5 Conclusions

This paper aims at investigating and developing an algorithm for recognition and prediction of precipitation cloud over Tibet by combining real-time FY-4A observation

data and DEM data. Firstly, the GPM IMERG of cumulative precipitation statistics is used to label the satellite cloud image of FY-4A at the previous moment, so that the established model can detect and forecast precipitation clouds. Secondly, U-Net, a semantic segmentation method suitable for small samples, is used as a basic model. Residual network is combined with U-Net to make a full use of the feature information extracted from each convolution layer, and enhance the ability of feature extraction and generalization of the network. Next, to solve the problem of the network would produce semantic differences when directly concatenating low-level and high-level features, dense skip pathways reuse feature maps of different layers as inputs to concatenate neural networks feature layers of different depths. Finally, to overcome U-Net's limitation of detection for small size targets from background, according to the characteristics of precipitation clouds, the pooling layer in the down-sampling of U-Net is modified to Down-convolution. This enhances the model's ability to recognize small precipitation clouds and increases the accuracy of cloud contour information extraction. The evaluation index of MIoU and PA shows that the improved U-NET is significantly advanced than the original U-Net.

Funding Statement: The authors would like to acknowledge the financial support from the National Science Foundation of China (Grant No. 41875027).

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

Badrinarayanan, V.; Kendall, A.; Cipolla, R. (2017): Segnet: a deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 39, no. 12, pp. 2481- 2495.

Biswas, S. K.; Farrar, S.; Gopalan, K.; Santos-Garcia, A.; Jones, W. L. et al. (2013): Intercalibration of microwave radiometer brightness temperatures for the global precipitation measurement mission. *IEEE Transactions on Geoence & Remote Sensing*, vol. 51, no. 3, pp. 1465-1477.

Chen, B.; Chen, B. F.; Lin, H. T. (2018): Rotation-blended cnns on a new open dataset for tropical cyclone image-to-intensity regression. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, ACM*, pp. 90-99.

Drönner, J.; Korfhage, N.; Egli, S.; Mühling, M.; Thies, B. et al. (2018): Fast cloud segmentation using convolutional neural networks. *Remote Sensing*, vol. 1782, no. 10, pp. 1-24.

Forbes, R.; Haiden, T.; Magnusson, L. (2015): Improvements in ifs forecasts of heavy precipitation. *ECMWF Newsletter*, no. 144, pp. 21-26.

He, K. M.; Zhang, X. Y.; Ren, S. Q.; Sun, J. (2016): Deep residual learning for image recognition. *Proceedings of The IEEE Conference on Computer Vision & Pattern Recognition*, pp. 770-778.

He, Y.; Li, G. P. (2013): Numerical experiments on influence of Tibetan plateau on persistent heavy rain in south China. *Chinese Journal of Atmospheric Sciences*, vol. 37, no. 4, pp. 933-944.

Huang, G.; Wan, Z.; Liu, X.; Hui, J.; Zhang, Z. (2019): Ship detection based on squeeze excitation skip-connection path networks for optical remote sensing images. *Neurocomputing*, vol. 3327, pp. 215-223.

Kan, X.; Zhang, Y. H.; Zhu, L. L.; Xiao, L. M.; Wang, J. G. et al. (2018): Snow cover mapping for mountainous areas by fusion of modis l1b and geographic data based on stacked denoising auto-encoders. *Computers, Materials & Continua*, vol. 57, no. 1, pp. 49-68.

Liu, J.; Yang, Y. H.; Lv, S. Q.; Wang, J.; Chen, H. (2019): Attention-based biglu-cnn for Chinese question classification. *Journal of Ambient Intelligence & Humanized Computing*, https://doi.org/10.1007/s12652-019-01344-9.

Long, J.; Shelhamer, E.; Darrell, T. (2015): Fully convolutional networks for semantic segmentation. *Proceedings of The IEEE Conference on Computer Vision & Pattern Recognition*, pp. 3431-3440.

Otsuka, S.; Tuerhong, G.; Kikuchi, R. (2016): Precipitation nowcasting with three-dimensional space-time extrapolation of dense and frequent phased-array weather radar observations. *Weather & Forecasting*, vol. 31, no. 1, pp. 329-340.

Ronneberger, O.; Fischer, P.; Brox, T. (2015): U-net: convolutional networks for biomedical image segmentation. *International Conference on Medical Image Computing & Computer-assisted Intervention, Cham*, pp. 234-241.

Shi, E.; Li, Q.; Gu, D. Q.; Zhao, Z. M. (2018): Weather radar echo extrapolation method based on convolutional neural networks. *Journal of Computer Applications*, vol. 38, no. 3, pp. 661-665.

Wang, W.; Jiang, Y. B.; Luo, Y. H.; Li, J.; Wang, X. et al. (2019): An advanced deep residual dense network approach for image super-resolution. *International Journal of Computational Intelligence Systems*, vol. 12, no. 2, pp. 1592-1601.

Yue, J.; Zhao, W.; Mao, S.; Liu, H. (2015): Spectral-spatial classification of hyperspectral images using deep convolutional neural networks. *Remote Sensing Letters*, vol. 6, no. 6, pp. 468-477.

Zeng, D. J.; Dai, Y.; Li, F.; Wang, J.; Sangaiah, A. K. (2019): Aspect based sentiment analysis by a linguistically regularized cnn with gated mechanism. *Journal of Intelligent & Fuzzy Systems*, vol. 36, no. 5, pp. 3971-3980.

Zhang, X.; Duan, J.; Sun, W.; Jha, S. (2019): A tumor perception system based on a multi-layer mass-spring model. *International Journal of Sensor Networks*, vol. 31, no. 1, pp. 24-32.

Zhou, Y.; Wu, T. (2019): Composite analysis of precipitation intensity and distribution characteristics of western track landfall typhoons over china under strong and weak monsoon conditions. *Atmospheric Research*, vol. 225, pp. 131-143.

Zhou, Z.; Siddiquee, M.; Tajbakhsh, N.; Liang, J. (2018): UNet++: a nested U-Net architecture for medical image segmentation. *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support Springer*, pp. 3-11.

Zsoter, E.; Pappenberger, F.; Richardson, D. (2015): Sensitivity of model climate to sampling configurations and the impact on the extreme forecast index. *Meteorological Applications*, vol. 22, no. 2, pp. 236-247.