Efficient Cloud Detection in Remote Sensing Images using Edge-aware Segmentation Network and Easy-to-hard Training Strategy

Kun Yuan, Gaofeng Meng, Dongcai Cheng, Jun Bai, Shiming Xiang and Chunhong Pan

National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences
School of Computer and Control Engineering, University of Chinese Academy of Sciences
Research Center for Brain-inspired Intelligence, Institute of Automation, Chinese Academy of Sciences

Outline

1. Overview and motivation

2. Edge-aware network and Easy-to-hard training strategy

3. Experiments and analysis

4. Conclusion and future work

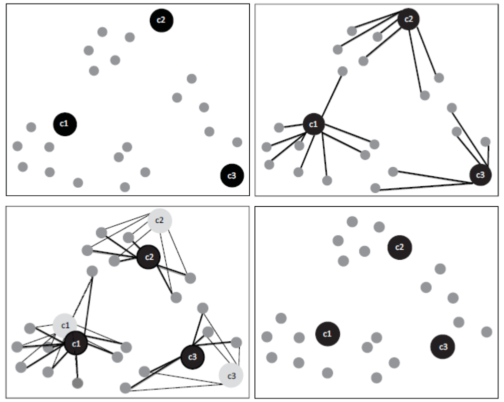
Overview and motivation

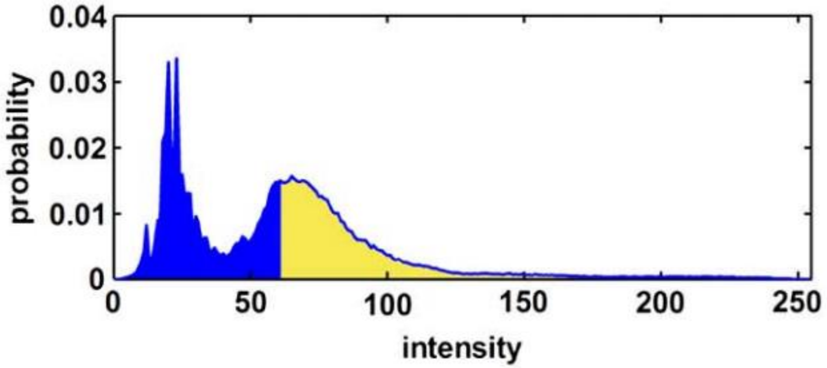
- Motivation: detecting cloud regions in remote sensing images is of great importance in weather forecasting, cloud removal and other applications.

- Challenges: bright non-cloud regions, semitransparent cloud regions, ambiguous boundaries of cloud layer and uneven distribution make this issue intractable.

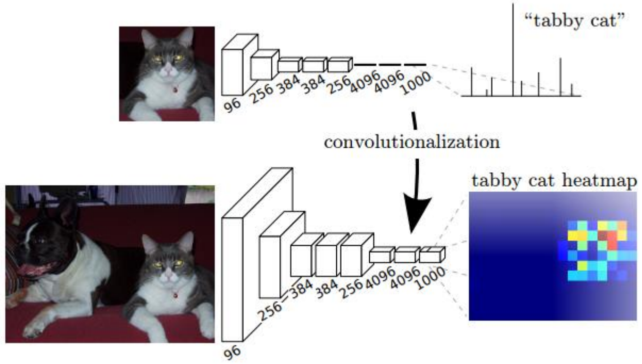
Overview and motivation

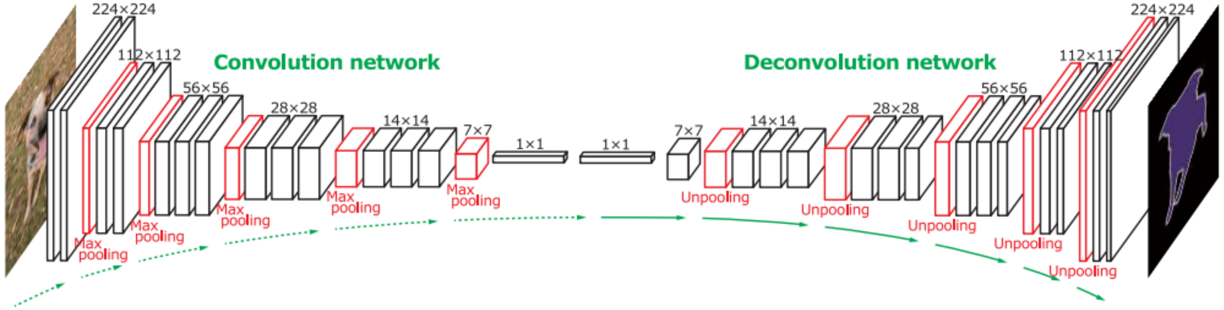
- Traditional methods: clustering based on intensity or handcrafted features





- Deep learning methods: segmentation networks, training through large amount of cloud images





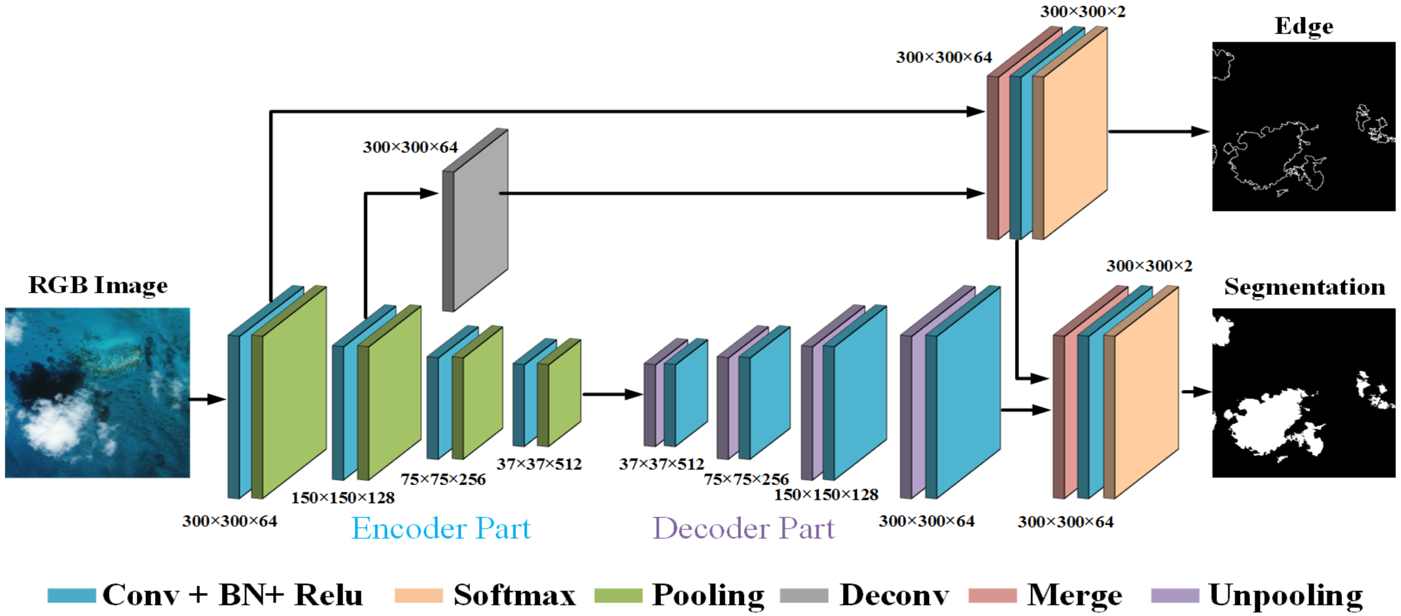
Overview and motivation

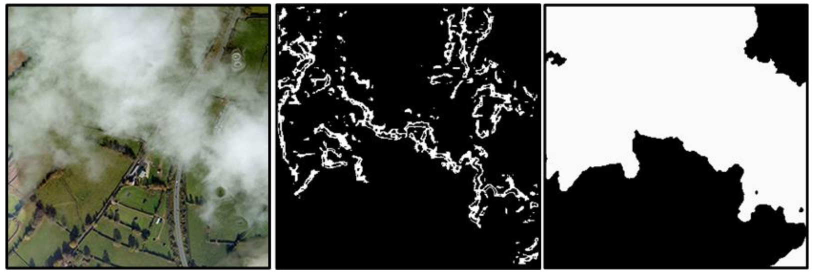
Our main contributions:

- An edge-aware network is proposed which combines ***cloud segmentation*** and ***cloud edge detection*** together to encourage better detection results near cloud boundaries, resulting in an end-to-end approach for accurate cloud detection.

- A training strategy based on ***easy-to-hard*** sample selection is proposed for efficient training of the network. Selected samples is governed by a weight that is annealed until entire are considered.

Edge-aware network



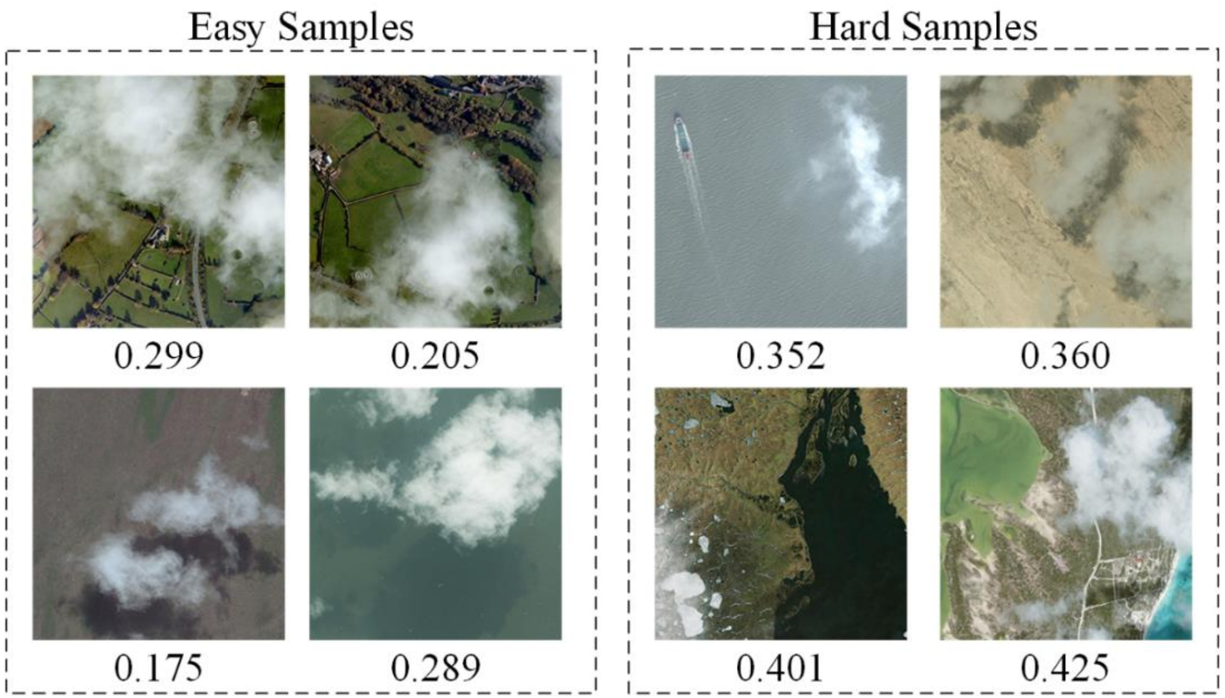


Cloud Image　Edge Detection　Segmentation

- The edge-aware network is developed for cloud detection to ***regularize the former convolutional layers*** as well as ***enhancing segmentation results near boundary regions***.
- The main network is consists of ***encoder*** and ***decoder*** parts, which automatically learns features for cloud segmentation.
- Experiments demonstrate that these implementation preferences are efficient and effectual.

Easy-to-hard training strategy

- Due to the uneven distribution of cloud, in which types of the backgrounds vary from sample to sample, the difficulty for learning samples is different.
- we start with learning easier samples of cloud data, and then gradually take more complex samples into consideration. This simple idea demonstrated to be beneficial in avoiding bad local minima and in achieving a better generalization result.



Loss

- Given cloud dataset $D = \{(x_1, y_1, z_1), \ldots, (x_k, y_k, z_k), \ldots, (x_n, y_n, z_n)\}$
- The segmentation loss is calculated by binary cross-entropy:

$$L_{seg} = \sum_{i=1}^{N} -\frac{1}{N} \{y_i ln f_{seg_i}(x_k, w) + (1 - y_i)\ln(1 - f_{seg_i}(x_k, w))\}$$

- Total loss can be written as:

$$L(x_k, y_k, z_k, f, w; \alpha, \beta) = L_{seg} + \frac{\beta}{\alpha} L_{edge}$$

- In our network, the goal is to jointly learn the model parameter w and the latent weight variable v $= [v_1, \ldots, v_n]$ by minimizing:

$$min \sum_{k=1}^{n} v_k L(x_k, y_k, z_k, f, w; \alpha, \beta) - \lambda \sum_{k=1}^{n} v_k \, , s.t. v \in \{0,1\}.$$

- In each iteration, samples in the selected block are employed to train the network and obtain the optimal w$*$, while the other block is fixed. With the fixed w, the global optimum v $= [v_1, \ldots, v_n]$ can be easily calculated by:

$$v_k^* = \begin{cases} 1, & L_{seg} < \lambda \\ 0, & otherwise. \end{cases}$$

Data augmentation

- Data description

Collected from Google Earth with a spatial resolution of 3.0−5.0 m. There are 118 training images, 6 validation images, and 30 testing images, each with size around 688×488. We label the ground truths of all the images by hands. The ground truth of the edge is automatically calculated using 8-direction edge detection filters. Background scenes in the cloud images are diverse, including desert, sea, forest, island, city, ice land, etc.

- Data augmentation

1.Corner crop 2.center crop 3.random crop 4.rotate 5.mirror

There are totally 29119 training samples and 3900 validation samples, each with size 300×300.

Before training, histogram equalization is adapted to samples instead of removing the mean value.

Comparisons evaluation metrics

- Five approaches are compared with our method: k-means [23], mean-shift [24], Progressive Refinement Scheme (PRS) [1], fully convolutional networks (FCNs) [16] and SegNet [13].
- Evaluation indexes are defined to compare and analyze results. The right rate (RR), error rate (ER), false alarm rate (FAR), ratio of RR to ER (RER) and intersection over union (IOU) are used to evaluate the cloud detection results. They can be computed as:

$$RR = \frac{TP}{TP+FN}, \quad IOU = \frac{TP}{TP+FP+FN},$$

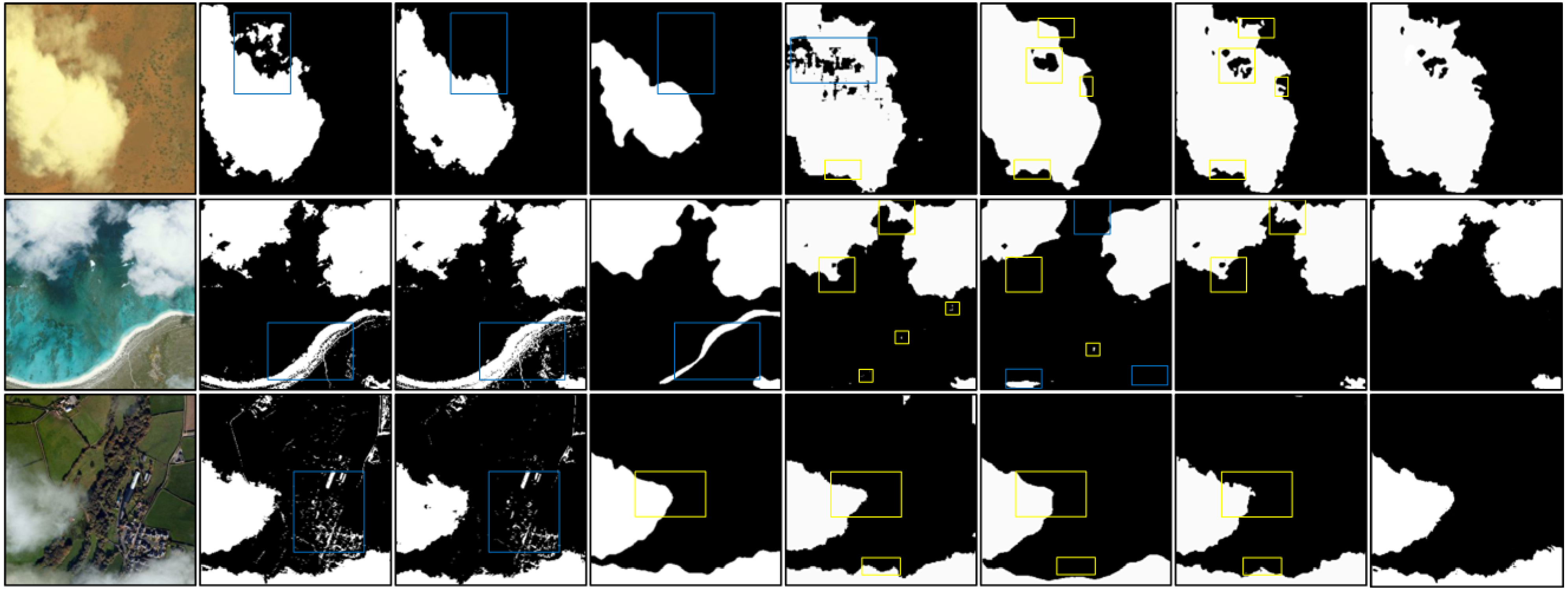
$$ER = \frac{FP+TN}{TP+FP+TN+FN}, \quad RER = \frac{RR}{ER},$$

$$FAR = \frac{FP}{TP+FP+TN+FN},$$

	RR	ER	FAR	RER	IOU
K-means	0.7644	0.1863	0.1196	9.1004	0.5454
Mean-shift	0.6475	0.1803	0.0981	7.2370	0.5018
PRS	0.8020	0.1222	0.0772	14.5679	0.6201
FCNs	0.8676	0.1323	0.0661	10.2407	0.7755
SegNet	0.9191	0.0809	0.0405	20.7698	0.8562
Our method	**0.9334**	**0.0666**	**0.0333**	**24.9407**	**0.8786**

Table 1: Segmentation results on five evaluation indexes.

Visualization



Input K-means Mean-shift PRS FCNs SegNet Our Method Ground Truth

Fig. 4: Visual comparison of cloud detection results with six methods. In the blue rectangle regions, mislabeled pixels can be seen in semitransparent cloud, sandbeach, streets and houses regions when traditional methods are used. In the yellow rectangle regions, detection accuracy around edge regions has been distinctly improved. Compared with FCNs, spatial consistent can be maintained by our method.

Ablation Analysis

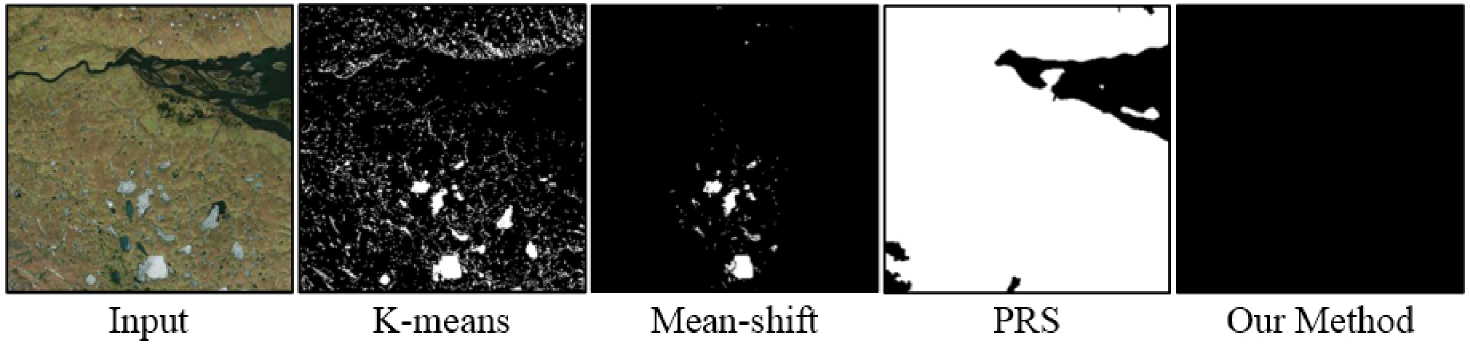


Fig. 5: Visual comparison of cloud detection for ice-covered regions. Due to the very similar brightness of ices and cloud regions, Kmeans, Mean-shift and PRS fail to distinguish them correctly. In comparison, our method can successfully distinguish these regions.

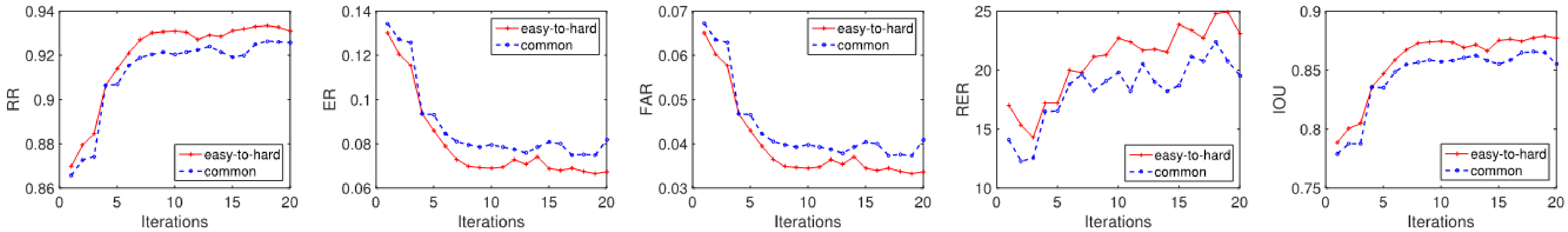


Fig. 6: Comparisons on Five Evaluation Indexes. Red lines denotes evaluations by our training strategy, while blue one represents non-strategy training method. Distinctly, better results can be achieved on all evaluation indexes by easy-to-hard strategy.

Conclusion

- In this paper, an edge-aware deep neural network has been proposed for cloud detection in RSIs. Compared with handcrafted or low-level features based cloud detection methods, our network combines cloud segmentation with cloud edge detection to encourage a better detection result near cloud boundaries. An easy-to-hard training strategy based on sample selection is also proposed to speed up the convergence of the network and improve the final segmentation results. Both visual and quantitative comparisons show that our method can yield superior results over the state-of-the-art methods.