UCLOUDNET: A RESIDUAL U-NET WITH DEEP SUPERVISION FOR CLOUD IMAGE SEGMENTATION

Yijie Li[1], *Hewei Wang*[1], *Shaofan Wang*[2], *Yee Hui Lee*[4], *Muhammad Salman Pathan*[5], *Soumyabrata Dev*[1,3]

[1]School of Computer Science, University College Dublin, Dublin D04V1W8, Ireland
[2]Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China
[3]The ADAPT SFI Research Centre, Dublin D04V1W8, Ireland
[4]School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore
[5]Innovation Value Institute, Maynooth University, Ireland

ABSTRACT

Recent advancements in meteorology involve the use of ground-based sky cameras for cloud observation. Analyzing images from these cameras helps in calculating cloud coverage and understanding atmospheric phenomena. Traditionally, cloud image segmentation relied on conventional computer vision techniques. However, with the advent of deep learning, convolutional neural networks (CNNs) are increasingly applied for this purpose. Despite their effectiveness, CNNs often require many epochs to converge, posing challenges for real-time processing in sky camera systems. In this paper, we introduce a residual U-Net with deep supervision for cloud segmentation which provides better accuracy than previous approaches, and with less training consumption. By utilizing residual connection in encoders of UCloudNet, the feature extraction ability is further improved. In the spirit of reproducible research, the model code, dataset, and results of the experiments in this paper are available at: https://github.com/Att100/UCloudNet.

Index Terms— deep learning, cloud segmentation, deep supervision, U-Net, residual network.

1. INTRODUCTION

Cloud information analysis plays a crucial role in the field of meteorological research, offering valuable insights into weather patterns and facilitating enhanced forecasting methods. As computer vision and machine learning advance, they have expanded into various interdisciplinary fields such as meteorology estimation [1–3] and weather variable prediction [4–6]. While traditionally cloud imagery has been captured via meteorological satellites in near-earth orbit, the recent utilization of ground-based sky cameras [7, 8] has gained prominence due to their enhanced temporal and spatial resolution. With the advent of these cameras, several datasets of optical RGB images, such as SWIMSEG [9], SWINSEG [10], and SWINySEG [11]. To better extract the cloud information from those images, the deep learning method with a fully convolution network (FCN) is widely used for cloud segmentation which consists of a series of encoders and decoders. However, this early design without an extra modified structure makes it hard to aggregate features from the first few layers. Additionally, training deep convolutional neural networks remains challenging without a shortcut skip connection.

In this paper, we propose the residual U-Net model with deep supervision for the cloud segmentation task, called UCloudNet. UCloudNet consists of a series of convolution blocks with residual connections which enhance the feature aggregation ability of the original U-Net by including more feature map fusion operations. The experiments prove that our proposed UCloudNet can achieve better performance with less training time and iterations than previous approaches.

The main contributions of our work are threefold:

- We propose UCloudNet, a novel U-Net based model enhanced with residual connections in encoder, significantly improving capability for feature extraction.

- We adopt deep supervision in our UCloudNet which substantially reduces the training time consumption.

- We conduct extensive experiments on three different benchmark datasets and validate the effectiveness of our proposed modules and training strategy.

2. RELATED WORKS

In addressing the challenge of segmenting sky/cloud images, a variety of methods have been developed. These methods are generally categorized into traditional computer vision techniques, as outlined in visual model related studies like [12–
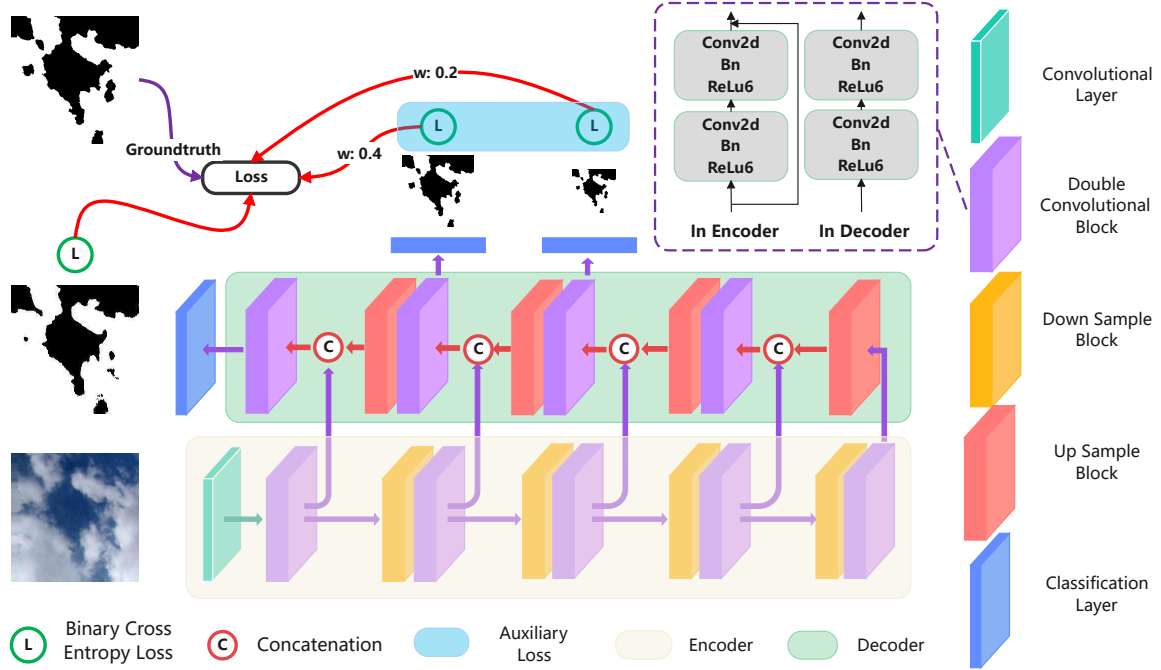

Fig. 1: The architecture of the UCloudNet model. The procedure between the output of model and the segmentation mask has been omitted in this figure.
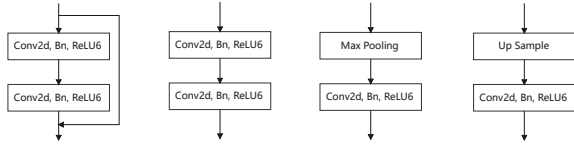

Fig. 2: The structure of 'Double Convolution Block' in encoder, decoder, 'Down Sample Block', and 'Up Sample Block' (from left to right).

14], and more recent deep learning approaches, as demonstrated in [11,15] multi-label image segmentation. Traditional techniques often rely on color feature analysis, static convolution filters, and pixel gradients. For instance, Dev *et al.* [13] use principal component analysis (PCA) and fuzzy cluster to evaluate the color model aims to capture the greatest color variance between cloud and sky in 2014. While these methods are adept at capturing the general sky distribution, they tend to lose finer details, resulting in suboptimal segmentation accuracy and imprecise binary masks that don't align well with the image boundaries. In contrast, the introduction of deep learning techniques has notably enhanced the quality of binary cloud mask generation. The encoder-decoder-based network is the most popular approach for segmentation-related tasks, for example, [16] adopts a network with dual attention to perform salient object detection. Dev *et al.* [11] CloudSegNet introduced an innovative approach in 2019, utilizing a standard Fully Convolutional Network (FCN) framework.

This method compresses information into high-dimensional feature maps through downsampling, followed by upsampling procedures to refine segmentation results. This significantly improves boundary detail and overall segmentation accuracy. Additionally, Dev et al. [15] presented a novel multi-label sky/cloud segmentation strategy in 2019, utilizing a multi-class U-Net model to classify images into three categories, enabling more precise analysis via three-class segmentation.

3. ARCHITECTURE

Our UCloudNet is based on the U-Net [17] structure which contains a series of decoders and encoders with channels concatenation in each stage. To compare with the original U-Net structure, we use a hyper-parameter k to control the parameters amount and inspired by ResNet, we add residual connection in each convolution block in encoder which is helpful for training the deeper layers. As for the training strategy, we use deep supervision to support the training process. The architecture of our proposed model is shown in Fig. 1. Our model contains a series of 'Double Convolution' blocks, 'Down Sample' blocks, and 'Up Sample' blocks. We explain these blocks in the following sections.

3.1. Double Convolution Block (DCB)

Fig. 2 describes the structure of the double convolution block. The 'Double Convolution Block' contains two groups of layers which include an original 3x3 convolution layer, a ReLU6
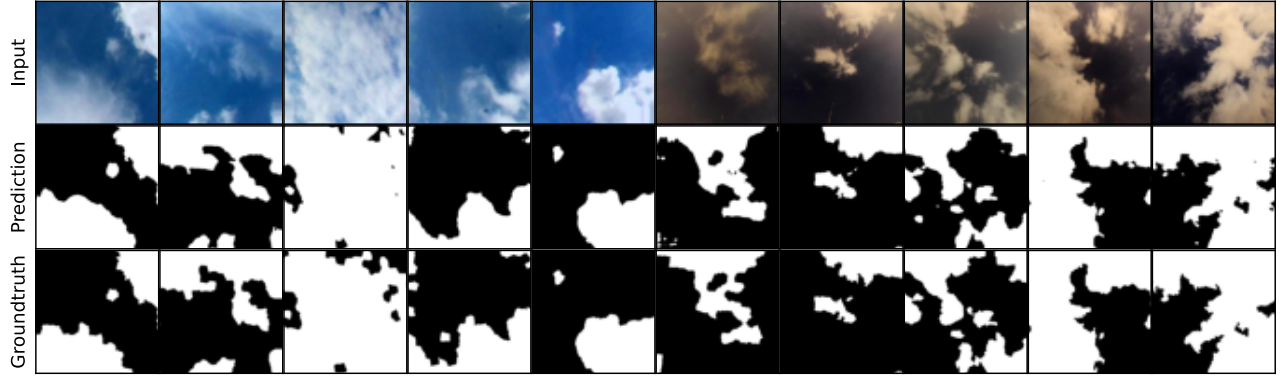
5554

Fig. 3: Results of cloud segmentation for day-time (1-6 columns) and night-time (7-12 columns).

activation layer, and a batch-normalization layer. This group of layers is called 'BasicConv2d' in our implementation. The structure of the DCB module is different in encoder and decoder, we only apply residual connection in the DCB in encoder while the structure of DCB in decoder is just a simple stack of 'Conv-Bn-ReLu6' groups. The different designs of DCB can result from the short-cut concatenation in the U-Net structure. The U-Net structure already has a short connection between encoder and decoder, and therefore there is no need to apply a short-cut in the decoder again.

3.2. Down Sample Block and Up Sample Block

The 'Down Sample Block (DSB)' includes a Max-pooling layer and a 'BasicConv2d' block. The 'Up Sample Block (UPB)' is a combination of an Up-sample layer and a 'BasicConv2d' block. The UPB module receives the output of a previous DCB and the output of this block will be concatenated with the output of another DCB in encoder with the same feature map size.

3.3. Deep supervision

The deep supervision training strategy can effectively support the training of deep neural networks and improve the regularization ability. In our implementation, we use two additional auxiliary loss branches to enable deep supervision. These two loss branches are located at stages with $1/2$ and $1/4$ output size, as shown at the top left of Fig. 1. In order to perform the binary cross entropy loss, we interpolate the target map into $1/2$ and $1/4$ of the original size. Additionally, a group of weight values is considered in our total loss so that the significance of a high-resolution prediction map is larger than a low-resolution prediction map.

3.4. Parameters control and loss function

The number of parameters of our model can be controlled by a hyper-parameter k. The number of the filters of DCB in encoder can be specified by $k * 2^s$ while the configuration of DCB in decoder is $k * 2^{3-s}$. The filters of the convolution layer in the "Down Sample" block can be calculated by $k * 2^{s+1}$ and the filter number of the "Up Sample" block is $k * 2^{4-s}$, while $s = 0, 1, 2, 3$.

We use binary cross entropy as the loss function and the total loss function can be represented as follows:

$$L(p,y) = -\frac{1}{N} * \sum_{i=0}^{N} y_i * \log p_i + (1 - y_i) * \log (1 - p_i) \quad (1)$$

$$L_{total}(p,y) = L(p,y) + 0.4 * L(p_2, y_2) + 0.2 * L(p_4, y_4) \quad (2)$$

4. EXPERIMENTS & RESULTS

We conduct experiments under different configurations, including the size of the model (controlled by k), usage of learning rate decay, and the deep supervision strategy.

4.1. Dataset and Training Configurations

The cloud segmentation dataset used in our study was sourced from the Singapore Whole Sky Nychthemeron Image SEGmentation Database (SWINySEG), including a total of 6078 daytime and 690 nighttime cloud images. Our proposed UCloudNet was developed using the PaddlePaddle framework and trained on a single NVIDIA Tesla V100-SXM2 16GB GPU. We partitioned the dataset into training and testing sets with an 8:2 split, and trained on three subsets: daytime, nighttime, and the complete SWINySEG dataset. The training was conducted with a batch size of 16 for a total of 100 epochs. We use Adam optimizer, employed with an initial learning rate of 0.001. We incorporated an exponential decay in the learning rate, with a gamma value of 0.95 applied after each epoch.

Table 1: Comparison of UCloudNet with other cloud segmentation methods

| Dataset	Method	Precision	Recall	F-measure	Error-rate
SWINySEG (day)	Dev et al. 2014 [13]	0.89	0.92	0.89	0.09
(augmented SWIMSEG)	Long et al. [12]	0.89	0.82	0.81	0.14
	Li et al. [18]	0.81	**0.97**	0.86	0.12
	CloudSegNet [11]	0.92	0.88	0.89	0.07
	Souza et al. [19]	**0.99**	0.53	0.63	0.18
	UCloudNet(k=2)+aux+lr-decay	0.90	0.93	0.92	0.07
	UCloudNet(k=4)+aux+lr-decay	0.92	0.94	**0.93**	**0.06**
SWINySEG (night)	Dev et al. 2017 [10]	0.94	0.74	0.82	0.13
(augmented SWINSEG)	Yang et al. 2009 [20]	**0.98**	0.65	0.76	0.16
	Yang et al. 2010 [14]	0.73	0.33	0.41	0.37
	Gacal et al. [21]	0.48	**0.99**	0.62	0.50
	CloudSegNet [11]	0.88	0.91	0.89	0.08
	UCloudNet(k=2)+aux+lr-decay	0.92	0.94	0.93	0.06
	UCloudNet(k=4)+aux+lr-decay	0.95	0.95	**0.95**	**0.04**
SWINySEG (day+night)	CloudSegNet [11]	0.92	0.87	0.89	0.08
	UCloudNet(k=2)+aux+lr-decay	0.90	0.92	0.91	0.08
	UCloudNet(k=4)	0.92	0.90	0.91	0.08
	UCloudNet(k=4)+lr-decay	0.91	0.94	0.92	0.07
	UCloudNet(k=4)+aux+lr-decay	**0.92**	**0.94**	**0.93**	**0.06**

4.2. Metrics

We assessed the effectiveness of our model using four commonly adopted metrics: precision, recall, F-measure, and error rate. The F-measure is represented by $\frac{2 \times Precision \times Recall}{Precision + Recall}$. Precision is defined as $\frac{TP}{TP+FP}$, while recall is $\frac{TP}{TP+FN}$, and the error rate is calculated using $\frac{FP+FN}{P+N}$. We employed the PR curve across 256 different thresholds.

4.3. Quantitative Analysis

Quantitative evaluation results of our methods are shown in Table 1, which shows the precision, recall, F-measure, and error-rate of our proposed UCloudNet with other methods on different data sets. On the full SWINySEG data set, UCloudNet (k=4) with deep supervision and learning rate decay has the best performance on all the four metrics while UCloudNet (k=4) with only learning rate decay has the second best performance which can prove that deep supervision with auxiliary loss can improve our model performance. On day-time images, the overall performance is better than others. As for the night-time images, our model has the lowest error rate.

4.4. Qualitative Analysis

Qualitatively, we use several day-time and night-time images as our evaluation samples. We use these images as input and perform a threshold with p=0.5 on the sigmoid output, shown in Fig. 3. We observed our binary prediction maps and compared them with input images and ground truths, it is proved that our proposed model can accurately extract the overall feature and generate binary prediction correctly. However, there
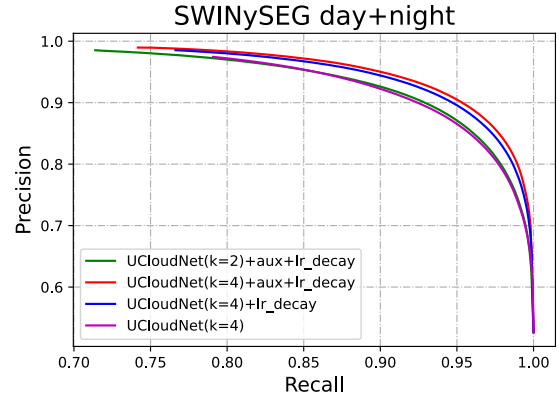
Fig. 4: PR curve of UCloudNet with different training configurations on full SWINySEG ground-based cloud segmentation dataset.

still exists some negative phenomenons, for example, the lack of accuracy on segmentation on edges. Besides, we evaluate our proposed model with the PR curve, shown in Fig. 4. In order to receive clearer and more precise results, we set 256 thresholds to generate the curve, the area under the curve is bigger, and the overall performance of the model is better.

5. CONCLUSION

In this paper, we introduce a residual U-Net with deep supervision for cloud-sky segmentation. We train our model with different configurations on various splits of SWINySEG dataset. Our proposed method achieves better performance as compared to the other methods and our experiments prove that deep supervision with auxiliary loss can gain better

performance. Additionally, our model only needs less than 17500 iterations (100 epochs with batch-size 16) to converge which can significantly save the training consumption, as compared to other deep learning methods.

6. REFERENCES

[1] P. Dey, B. P. Das, Y. H. Lee, and S. Dev, "NeSNet: A deep network for estimating near-surface pollutant concentrations," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2023.

[2] S. Dev, F. M. Savoy, Y. H. Lee, and S. Winkler, "Estimating Solar Irradiance Using Sky Imagers," *Atmospheric Measurement Techniques*, vol. 12, no. 10, pp. 5417–5429, 2019.

[3] N. Akrami, K. Ziarati, and S. Dev, "Graph-based Local Climate Classification in Iran," *International Journal of Climatology*, vol. 42, no. 3, pp. 1337–1353, 2022.

[4] S. Manandhar, S. Dev, Y. H. Lee, Y. S. Meng, and S. Winkler, "A Data-Driven Approach for Accurate Rainfall Prediction," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 11, pp. 9323–9331, 2019.

[5] H. Wang, M. S. Pathan, Y. H. Lee, and S. Dev, "Day-ahead Forecasts of Air Temperature," in *2021 IEEE USNC-URSI Radio Science Meeting (Joint with AP-S Symposium)*. IEEE, 2021, pp. 94–95.

[6] M. Jain, P. Yadav, and S. Dev, "Holistic and lightweight approach for solar irradiance forecasting," *IEEE Transactions on Geoscience and Remote Sensing*, 2024.

[7] M. Jain, I. Gollini, M. Bertolotto, G. McArdle, and S. Dev, "An extremely-low cost ground-based whole sky imager," in *Proc. IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE, 2021, pp. 8209–8212.

[8] S. Dev, F. M. Savoy, Y. H. Lee, and S. Winkler, "Design of low-cost, compact and weather-proof whole sky imagers for high-dynamic-range captures," in *Proc. IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE, 2015, pp. 5359–5362.

[9] S. Dev, Y. H. Lee, and S. Winkler, "Color-based segmentation of sky/cloud images from ground-based cameras," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 1, pp. 231–242, 2016.

[10] S. Dev, F. M. Savoy, Y. H. Lee, and S. Winkler, "Night-time sky/cloud image segmentation," in *Proc. IEEE International Conference on Image Processing (ICIP)*. IEEE, 2017, pp. 345–349.

[11] S. Dev, A. Nautiyal, Y. H. Lee, and S. Winkler, "Cloud-segnet: A deep network for nychthemeron cloud image segmentation," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 12, pp. 1814–1818, 2019.

[12] C. N. Long, J. M. Sabburg, J. Calbó, and D. Pagès, "Retrieving cloud characteristics from ground-based daytime color all-sky images," *Journal of Atmospheric and Oceanic Technology*, vol. 23, no. 5, pp. 633–652, 2006.

[13] S. Dev, Y. H. Lee, and S. Winkler, "Systematic study of color spaces and components for the segmentation of sky/cloud images," in *Proc. IEEE International Conference on Image Processing (ICIP)*. IEEE, 2014, pp. 5102–5106.

[14] J. Yang, W. Lv, Y. Ma, W. Yao, and Q. Li, "An automatic groundbased cloud detection method based on local threshold interpolation," *Acta Meteorologica Sinica*, vol. 68, no. 6, pp. 1007–1017, 2010.

[15] S. Dev, S. Manandhar, Y. H. Lee, and S. Winkler, "Multi-label cloud segmentation using a deep network," in *2019 USNC-URSI Radio Science Meeting (Joint with AP-S Symposium)*. IEEE, 2019, pp. 113–114.

[16] Y. Li, H. Wang, Z. Li, S. Wang, S. Dev, and G. Zuo, "DAANet: Dual Attention Aggregating Network for Salient Object Detection," in *IEEE International Conference on Robotics and Biomimetics (ROBIO)*. IEEE, 2023, pp. 1–7.

[17] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[18] Q. Li, W. Lu, and J. Yang, "A hybrid thresholding algorithm for cloud detection on ground-based color images," *Journal of atmospheric and oceanic technology*, vol. 28, no. 10, pp. 1286–1296, 2011.

[19] M. P. Souza-Echer, E. B. Pereira, L. Bins, and M. Andrade, "A simple method for the assessment of the cloud cover state in high-latitude regions by a ground-based digital camera," *Journal of Atmospheric and Oceanic Technology*, vol. 23, no. 3, pp. 437–447, 2006.

[20] Q. Yang, L. Tang, W. Dong, and Y. Sun, "Image edge detecting based on gap statistic model and relative entropy," in *Proc. Sixth International Conference on Fuzzy Systems and Knowledge Discovery*, vol. 5. IEEE, 2009, pp. 384–387.

[21] G. F. B. Gacal, C. Antioquia, and N. Lagrosas, "Ground-based detection of nighttime clouds above Manila observatory (14.64° n, 121.07° e) using a digital camera," *Applied Optics*, vol. 55, no. 22, pp. 6040–6045, 2016.