K-CLOUDGAN: WEAKLY SUPERVISED CLOUD REMOVAL TOWARDS AUTOMATED SATELLITE SURVEILLANCE

Xiaofan Wu
Lancaster University
Lancaster, United Kingdom
xiaofan.wu@lancaster.ac.uk

Changtsun Li
Deakin University
Waurn Ponds, Australia
changtsun.li@deakin.edu.au

Peter M. Atkinson
Lancaster University
Lancaster, United Kingdom
pma@lancaster.ac.uk

Ce Zhang
Bristol University
Bristol, United Kingdom
dr23902@bristol.ac.uk

Plamen P. Angelov
Lancaster University
Lancaster, United Kingdom
p.angelov@lancaster.ac.uk

Richard Jiang
Lancaster University
Lancaster, United Kingdom
r.jiang2@lancaster.ac.uk

***Abstract*—Cloud occlusion significantly impairs the quality and usability of remote sensing imagery. This work tackles the problem in two stages: cloud segmentation and cloud removal. For segmentation, we employ transfer learning with three advanced models—U-Net, DeepLabV3+, and PSPNet—achieving approximately 90% accuracy on real-world datasets, thereby demonstrating the effectiveness of transfer learning for this task. For cloud removal, we introduce K-CloudGAN, a weakly supervised inpainting framework that integrates K-means-based pseudo-labeling with SN-PatchGAN and transfer learning. This design reduces reliance on large-scale labeled datasets while enhancing restoration quality. Compared to existing approaches like CloudGAN, our method achieves a 10% improvement in terms of SSIM and PSNR metrics, confirming its superior performance in restoring cloud-obscured satellite imagery.**

***Index Terms*—Satellite Surveillance, *K*-CloudGAN, Cloud Removal**

I. Introduction

Remote sensing is critical for Earth resource exploration, disaster prediction, and ecological monitoring [1–4]. However, cloud coverage—averaging 67% globally [5, 6]—often obscures surface features, leading to significant information loss in satellite imagery [7]. While cloud removal can restore hidden information, it remains constrained by the high cost and effort required for accurate dataset annotation.

This research consists of two stages: cloud segmentation and removal. Traditional cloud segmentation methods — such as thresholding [8], texture-based classification [9], and clustering [10]—leverage spectral and topographical cues but face scalability and adaptability challenges. Recent advances like Fmask [11] have improved accuracy, yet complex atmospheric conditions remain a bottleneck. As for Cloud Removal, deep learning models such as SpaGAN [12] and CloudGAN [13] adopt generative adversarial networks (GANs) for end-to-end cloud removal but often struggle with thin clouds and boundary preservation.

In our work, to address these challenges, we evaluate state-of-the-art segmentation models — U-Net [14], DeepLabV3+ [15], and PSPNet [16] — to improve cloud segmentation accuracy while reducing annotation burden via transfer learning. We further introduce *K*-CloudGAN, a weakly supervised cloud removal framework combining *K*-means-based pseudo-labeling, transfer learning, and SN-PatchGAN to achieve better performance. Model performance is evaluated using metrics including mIoU, mPA, Precision, Recall, SSIM, and PSNR.

II. Cloud Segmentation

A. Typical Cloud Segmentation Methods

This research consists of two stages: cloud segmentation and removal, utilizing the U-Net, DeepLabV3+ and PSPNet models. The RICE and 38-Cloud datasets were trained for 100 and 10 epochs, respectively, to predict cloud masks. The evaluation showed that U-Net and DeepLabV3+ perform well with minor deviations in the predictions.

U-Net. The U-Net structure extracts and upsamples features for semantic segmentation [14]. Feature extraction combines convolution and max pooling, with VGG16 or VGG19 backbones. Predictions adjust feature layer channels using 1 x 1 convolutions.

DeepLabV3+. DeepLabV3+ employs an encoder-decoder structure with dilated convolutions for improved perceptual fields [15]. MobileNetV2 [17] serves as the lightweight backbone, using separable convolutions for feature extraction and 1 x 1 convolutions for channel adjustment. Resizing ensures consistent output dimensions.

PSPNet. PSPNet enhances global information capture using a pyramidal pooling module, combining characteristics into grids of varying sizes [16]. MobileNetV2 is the backbone. Predictions integrate features using 3 x 3 and 1 x 1 convolutions, followed by resizing for consistent dimensions.

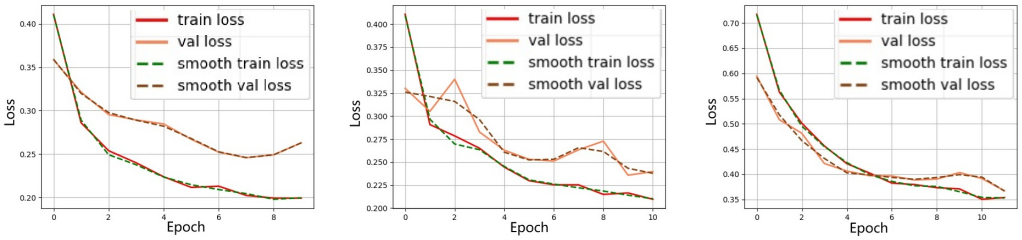

Fig. 1. 38-Cloud Loss (U-Net, DeepLabV3+, PSPNet)

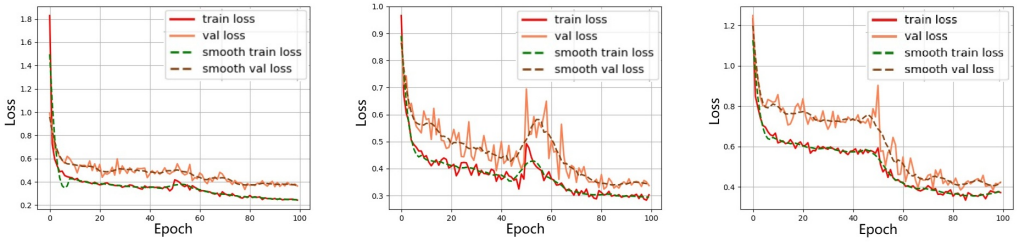

Fig. 2. RICE Loss (U-Net, DeepLabV3+, PSPNet)

Loss Function. Dice loss, calculated as:

$$Dice\ loss = \frac{2|X \bigcap Y|}{|X| + |Y|} \tag{1}$$

is used for all three networks, providing semantic segmentation evaluation metrics. Here, X and Y are the model prediction set and the ground reference set. Figures 1 - 2 shows the training loss over different datasets.

TABLE I
CLOUD SEGMENTATION RESULTS ON DATASETS

	mIoU	mPA	Precision	Recall
Unet: 38-Cloud	88.83%	94.19%	94.06%	94.19%
DeepLabV3+: 38-Cloud	**89.28%**	**94.43%**	**94.31%**	**94.43%**
PSPnet: 38-Cloud	82.95%	90.82%	90.71%	90.82%
Unet: RICE	78.21%	**87.94%**	87.63%	**87.94%**
DeepLabV3+: RICE	**78.21%**	87.30%	**88.56%**	87.30%
PSPnet: RICE	76.57%	86.40%	87.14%	86.40%

B. Cloud Segmentation Tests

Test Environment. All experiments were conducted using Python 3.7, PyTorch 1.12.1, TensorFlow 2.8.2, on a Tesla K80 GPU with CUDA 11.2 to accelerate training.

Model Training & Validation. Models were trained on the 38-Cloud and RICE datasets with a 90:10 train-test split. Training on 38-Cloud spanned 10 epochs (6–8 hours), while RICE took 1 minute per epoch. U-Net, DeepLabV3+, and PSPNet used VGG16 as the backbone, trained for 10–100 iterations. Input images were resized to 512×512 for U-Net and DeepLabV3+, and 473×473 for PSPNet. Labels were binarized into background and cloud.

Transfer learning was applied by freezing backbone parts. U-Net used Adam, while DeepLabV3+ and PSPNet used SGD with cosine-decay learning rates. The RICE dataset (736 samples) was trained for 100 epochs, treating all cloud types as a single target class.

C. Evaluation Results and Analysis

Table I summarizes the performance of all three models on the 38-Cloud and RICE datasets, with DeepLabV3+ achieving the highest average performance, followed by U-Net.

III. K-CLOUDGAN FOR CLOUD REMOVAL

A. Gated Convolutions and SN-PatchGAN

While cloud segmentation is a relatively simple binary segmentation task, cloud removal is often way more complicated as it is a generative task. In this work, we engage gated convolutions and SN-PatchGAN [18] as our in-painting tool, retoring the cloud-removed images based on the cloud masks from the cloud segmentation.

SN-PatchGAN and gated convolution is an effective tool for free-form image inpainting. Gated convolution improves partial convolution with a soft gating mechanism The output is calculated using:

$$\text{Gating}_{y,x} = \sum W_g \cdot I, \quad \text{Feature}_{y,x} = \sum W_f \cdot I \tag{2}$$

$$O_{y,x} = \phi(\text{Feature}_{y,x}) \odot \sigma(\text{Gating}_{y,x}) \tag{3}$$

where, W_g and W_f are weights and I is the input. The loss function combines pixel-level L1 reconstruction and SN-PatchGAN loss:

$$\mathcal{L}_G = -\mathbb{E}_{z \sim \mathbb{P}_z(z)}[D^{sn}(G(z))] \tag{4}$$

where, D^{sn} represents the spectral-normalized discriminator, and G is the image inpainting network that takes incomplete image z.

B. The Proposed K-CloudGAN Method

Supervised cloud segmentation often suffers from overfitting—models trained on specific datasets typically perform well on those datasets but degrade significantly on unseen data. Unsupervised segmentation can mitigate this by leveraging mechanisms such as saliency and attention [19, 20]. In this work, we adopt a simple *K*-means approach and propose a new framework, *K*-CloudGAN.

In our framework, we employ *K*-mean clustering, an unsupervised segmentation method, to improve the generalizability of our models on cloud removal. In our *K*-CloudGAN, the original AE-based detection model was replaced with *K*-means clustering to generate pseudo-ground labels. U-Net, selected for its superior performance, was used for cloud segmentation. By leveraging SN-PatchGAN, we have two schemes:

1) ***K*-CloudGAN-v1**. In this scheme, we combine the masks from both K-mean clustering and UNet as the input to SN-PatchGAN.
2) ***K*-CloudGAN-v2**. In this scheme, we simply use the masks from K-mean clustering as the input to SN-PatchGAN.

We will then examine which one can work better in our experiments.

***K*-Segmentation.** *K*-means clustering was applied in the HSV color space to group image pixels into clusters based on Euclidean distance. The optimal number of clusters was determined to be three using the elbow method. Figure3 shows the results for K = 3, 4, and 5, with the final cloud mask generated from the K = 3 model.

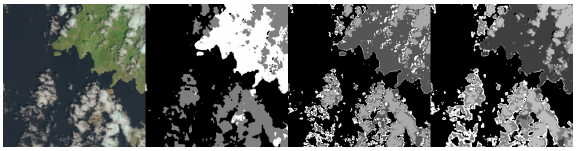

Fig. 3. *K*-Means Segmentation (K = 3, K = 4, K = 5)

C. Cloud Removal Experiments

Experimental Environment. Experiments were run using Python 3.7, PyTorch 1.12.1, TensorFlow 2.8.2 and CUDA 11.2 on a Tesla K80 GPU. Code modifications were necessary for TensorFlow 2.x compatibility.

Metrics. Testing was done with 74 images from the RICE2 dataset, using a pre-trained SN-PatchGAN checkpoint [13]. SSIM and PSNR were calculated for image restoration.

$$SSIM(\mathbf{x},\mathbf{y}) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (5)$$

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j) - K(i,j)]^2 \quad (6)$$

$$PSNR = 10\log_{10}\left(\frac{MAX_I^2}{MSE}\right) \quad (7)$$

where, x and y are the predicted results and the ground truth, respectively, and μ represents the mean value of a set. I and K denote the generated image and the ground truth, respectively.

TABLE II
K-CLOUDGAN & CLOUDGAN

	K-CloudGAN(1)(our)	CloudGAN(1) [13]
SSIM	**0.7149±0.306**	0.629±0.248
PSNR	**21.0587±4.3926**	18.630±5.942

TABLE III
K-CLOUDGAN(1) (WITH K-MEANS+UNET MASK) & K-CLOUDGAN(2) (WITH K-MEANS MASK ONLY)

	K-CloudGAN(1)	K-CloudGAN(2)
SSIM	0.7149±0.306	**0.7492±0.925**
PSNR	21.0587±4.3926	**23.6719±5.008**

D. Evaluation and Analysis

Cloud removal testing on the RICE dataset showed *K*-CloudGAN-v1 model achieved in image restoration with SSIM = 0.7149 ± 0.306 and PSNR = 21.0587 ± 4.3926 dB, with ab increased accuracy by 10% over CloudGAN, the SOTA method on cloud removal, highlighting the potential of combining unsupervised, transfer, and weakly supervised learning for remote sensing imagery. The results compared to CloudGAN are shown in Table II and Figure 4.

K-CloudGAN-v2 leveraging pseudo-masks from *K*-means segmentation yielded better restoration than *K*-CloudGAN-v1, as shown in Table III and Figure 5. The *K*-means generated mask (Figure 5, right) better matched the edges.

IV. CONCLUSION

This work presents an effective two-stage framework to address cloud occlusion in remote sensing imagery. First, we demonstrate that transfer learning applied to U-Net, DeepLabV3+, and PSPNet achieves around 90% segmentation accuracy on real-world datasets, validating its potential to reduce annotation costs. Second, we propose *K*-CloudGAN, a weakly supervised cloud removal method that leverages K-means pseudo-labeling and SN-PatchGAN, outperforming prior methods such as CloudGAN by 10% in terms of SSIM and PSNR. Together, these contributions enhance the robustness and scalability of cloud removal pipelines for satellite surveillance.

REFERENCES

[1] C.-Y. Chiang, C. Barnes, P. Angelov, and R. Jiang, "Deep learning-based automated forest health diagnosis from aerial images," *IEEE Access*, vol. 8, pp. 144 064–144 076, 2020.

[2] R. Dinakaran, L. Zhang, C.-T. Li, A. Bouridane, and R. Jiang, "Robust and fair undersea target detection with automated underwater vehicles for biodiversity data collection," *Remote Sensing*, vol. 14, no. 15, 2022.

[3] A. Shen, Y. Zhu, P. Angelov, and R. Jiang, "Marine debris detection in satellite surveillance using attention mechanisms,"

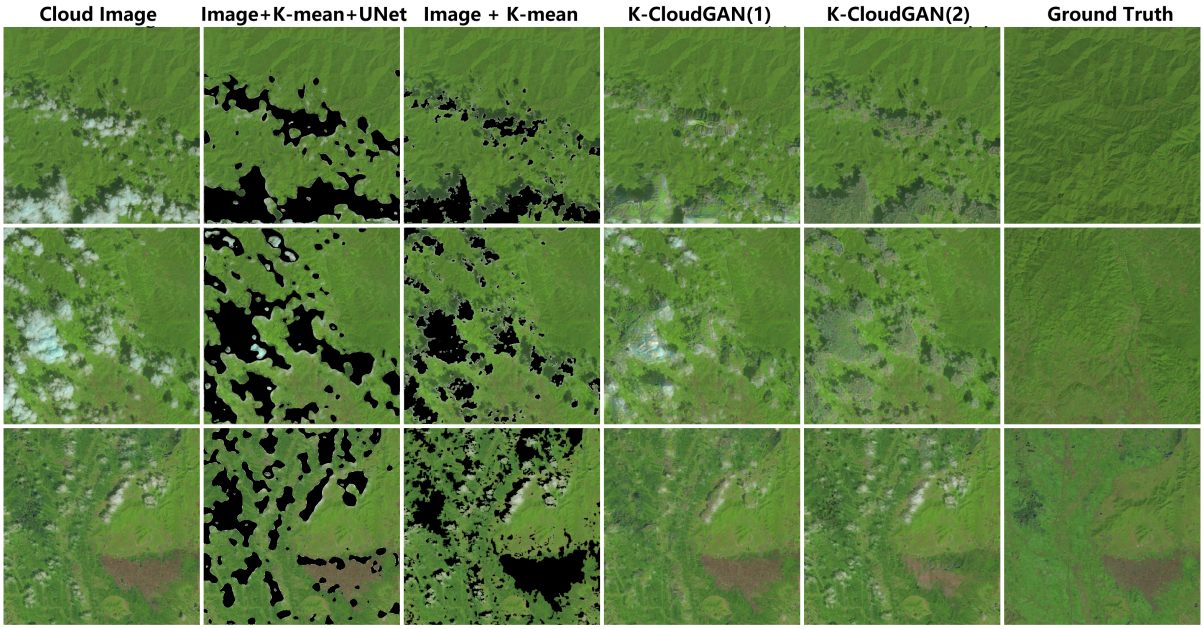

Fig. 4. SN-PatchGAN repair results (cloud, in-painting, ground reference).

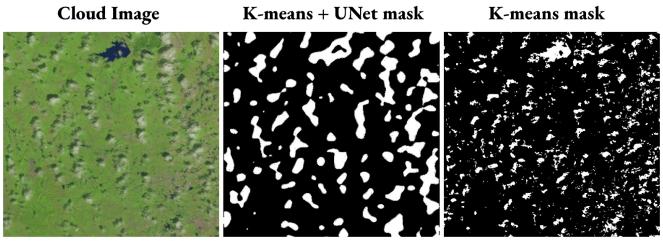

Fig. 5. Comparison of mask contours: *K*-means vs. UNet segmentation.

IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 17, pp. 4320–4330, 2024.

[4] X. Lin, C.-T. Li, S. Adams, A. Z. Kouzani, R. Jiang, L. He, Y. Hu, M. Vernon, E. Doeven, L. Webb, T. Mcclellan, and A. Guskich, "Self-supervised leaf segmentation under complex lighting conditions," *Pattern Recognition*, vol. 135, p. 109021, 2023.

[5] R. A. Schiffer and W. B. Rossow, "The international satellite cloud climatology project (isccp): The first project of the world climate research programme," *Bulletin of the American Meteorological Society*, vol. 64, no. 7, pp. 779–784, 1983.

[6] M. D. King, S. Platnick, W. P. Menzel, S. A. Ackerman, and P. A. Hubanks, "Spatial and temporal distribution of clouds observed by modis onboard the terra and aqua satellites," *IEEE Trans. Geoscience Remote Sensing*, vol. 51, no. 7, pp. 3826–3852, 2013.

[7] G. Jedlovec, "Automated detection of clouds in satellite imagery." IntechOpen, 2009.

[8] R. Ren, L. Gu, and H. Wang, "Clouds and clouds shadows detection and matching in modis multispectral satellite images," in *2012 International Conference on Industrial Control and Electronics Engineering*. IEEE, 2012, pp. 71–74.

[9] C. Christodoulou, S. Michaelides, and C. Pattichis, "Multifeature texture analysis for the classification of clouds in satellite imagery," *IEEE transactions on geoscience and remote sensing*, vol. 41, no. 11, pp. 2662–2668, 2003.

[10] A. Golzari Oskouei, M. Hashemzadeh, B. Asheghi, and M. A. Balafar, "Cgffcm: Cluster-weight and group-local feature-weight learning in fuzzy c-means clustering algorithm for color image segmentation," *Applied soft computing*, vol. 113, p. 108005, 2021.

[11] S. Qiu, Z. Zhu, and B. He, "Fmask 4.0: Improved cloud and cloud shadow detection in landsats 4–8 and sentinel-2 imagery," *Remote sensing of environment*, vol. 231, p. 111205, 2019.

[12] H. Pan, "Cloud removal for remote sensing imagery via spatial attention generative adversarial network," 2020.

[13] S. K. Jerry Schonenberg, "Cloudgan," https://github.com/JerrySchonenberg/CloudGAN.

[14] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, ser. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2015, pp. 234–241.

[15] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017.

[16] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," 2016.

[17] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," 2018.

[18] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. Huang, "Free-form image inpainting with gated convolution," in *2019 IEEE/CVF Int. Conf Computer Vision (ICCV)*. IEEE, 2019, p. 4470.

[19] R. Jiang and D. Crookes, "Shallow unorganized neural networks using smart neuron model for visual perception," *IEEE Access*, vol. 7, pp. 152 701–152 714, 2019.

[20] D. Konar, S. Bhattacharyya, T. K. Gandhi, B. K. Panigrahi, and R. Jiang, "3-d quantum-inspired self-supervised tensor network for volumetric segmentation of medical images," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 8, pp. 10 312–10 325, 2024.