CDD & CloudNet: A Benchmark Dataset & Model for Object Detection Performance

Mohd Ariful haque
Cyber Physical Systems
Clark Atlanta University
Atlanta, GA
mohdariful.haque@students.cau.edu

Rakib Hossain Rifat
Dept. of Computer Science
Texas Tech University
Lubbock, TX 79409
rrifat@ttu.edu

Marufa Kamal
Dept. of CSE
BRAC University
Dhaka, Bangladesh
marufa.kamal1@g.bracu.ac.bd

Roy George
Cyber Physical Systems
Clark Atlanta University
Atlanta, GA
rgeorge@cau.edu

Kishor Datta Gupta
Cyber Physical Systems
Clark Atlanta University
Atlanta, GA
kgupta@cau.edu

Khalil Shujaee
Cyber Physical Systems
Clark Atlanta University
Atlanta, GA
kshujaee@cau.edu

***Abstract*—Aerial imagery obtained through remote sensing is extensively utilized across diverse industries, particularly for object detection applications where it has demonstrated considerable efficacy. However, clouds in these images can obstruct evaluation and detection tasks. This study therefore involved the compilation of a cloud dataset, which categorized images into two classes: those containing clouds and those without. These images were sourced from the publicly available Maxar 'Hurricane Ian' repository, which contains images from various natural events. We demonstrated the impact of cloud removal during pre-processing on object detection using this dataset and employed six CNN models, including a custom model, for cloud detection benchmarking. These models were used to detect objects in aerial images from two other events in the Maxar dataset. Our results show significant improvements in precision, recall, and F1-score for CNN models, along with optimized training times for object detection in the CloudNet+YOLO combination. The findings demonstrate the effectiveness of our approach in improving object detection accuracy and efficiency in remote sensing imagery, particularly in challenging cloud-covered scenarios.**

***Index Terms*—Cloud Detection, Dataset, Deep Learning, CNN, ResNet, Vgg16, DenseNet169, EfficientNet, MobileNet**

I. Introduction

As remote sensing technology continues to advance at a rapid pace, the utilization of remote sensing imagery has expanded significantly across a diverse range of fields. Aerial images can benefit different sectors such as military target identification, environmental monitoring, meteorology, mineral exploration, and geographical mapping. In the context of aerial imagery, clouds are significant visual obstructions that can obscure smaller ground objects. According to a study conducted by NASA [1], 67% of the Earth's surface is covered by clouds. When detecting ground objects from aerial images, it is crucial to exclude images in which a substantial portion is blocked by cloud cover. This helps in maintaining the quality and validity of the data used for ground object detection and analysis. Thus, the identification of cloud objects plays a vital role in the pre-processing of the images for such applications with object recognition, image recognition, 3-D surface generation [2].

Over the years various algorithms [3]–[6] have been proposed for the detection of clouds in images. These algorithms utilize processing techniques such as analyzing the hue, high and low concentration features, thermal infrared information, texture, and geometry of the clouds. Additionally, various thresholding techniques have been implemented to enhance cloud detection accuracy. More recently, deep learning models have shown promise in cloud detection by learning features directly from the images. These models have the potential to improve cloud detection accuracy and robustness through their ability to automatically learn and adapt to complex patterns in the data. Researchers have utilized different custom and existing CNN models, such as Feature Pyramid Networks (FPN), multi-scale convolutional feature fusion (MSCFF) [7], and supervised learning techniques [8], for cloud detection in images. These techniques leverage advanced neural network architectures and supervised learning methods to improve the accuracy and robustness of cloud detection from images containing various objects.

In this paper, we introduce a new cloud image dataset comprising manually segregated images containing only clouds. These images have been resized to focus solely on the cloud formations. To create this dataset, we utilized the Maxar dataset and specifically selected images related to the "Hurricane Ian" event [9]. These images depict both pre and post-event aerial views, with a focus on the ground-level impact of Hurricane Ian. We applied techniques to detect and separate the cloud-containing images from the dataset. For object detection, we employed various CNN models, including a custom CNN designed with the smallest number of parameters. In our results section, we evaluate the benefits of this preprocessing step by demonstrating improved time optimization and the enhanced performance of object detection.

979-8-3503-5475-1/24/$31.00 ©2024 IEEE

II. Related Work

In this section, we will look into some of the available aerial benchmark datasets and papers that discuss methodologies to detect clouds in aerial images. Remote sensing images have proven to be quite handy in different cases such as Agriculture, Disaster Management, Urban Planning, Climate Studies, etc. AID Dataset [10] is quite popular for its large number of images and wide range of classes, 30 to be exact. It contains RGB images and is proficient in exploring different scenarios. Another popular dataset from the United States Geological Survey National Map covers [11] which covers 20 regions in the USA. The dataset comprises 21 categories of land-use scenes, each with single-label annotation. There are 100 images in each category, all sized 256×256 pixels. Shallow clouds [12] is another popular dataset that focuses on 4 different types of cloud formation types, fish, flower, gravel, and sugar. The images in this dataset were sourced from the National Aeronautics and Space Administration (NASA) Worldview. They cover three regions, spanning 21 degrees of latitude and 14 degrees of longitude, and crowd-sourcing activity was used for labeling them. Besides these using Google Earth images various other RS datasets were introduced The WHU-RS19 [13] dataset is a well-known resource in remote sensing, commonly employed for tasks like land-use and land-cover classification. It contains about 19 categories in total. Another is RSSCN7 [14], which is collected from 7 scenes containing a total of 2,400 images, with 400 images for each of the seven land-use scene categories.

With all these RS image datasets available researchers have tried to look more into the cloud detection techniques over the years for better results. In the paper, [15], a novel cloud detection method for remote sensing images using deep learning is presented. It enhances the SLIC method for superpixel segmentation and employs a deep CNN with two branches to detect clouds, distinguish between thick and thin clouds, and identify non-cloud regions. The CNN architecture consists of 4 convolutional layers and 2 fully connected layers, with the same structure in both branches. Experimental results demonstrate its superior accuracy and recall, achieving a recall of 0.97 for thick clouds and 0.84 for thin clouds, the highest compared to existing methods and single-scale CNN models. A novel CNN model(CDnet) has also been used by Jingyu Yang et al. [16] in their paper for cloud detection utilizing the capabilities of deep learning, for segmenting cloud regions from thumbnails of Remote Sensing Images. It employs a Feature Pyramid Module (FPM) to extract both multiscale and global contextual information with boundary refinement. The CDnet utilizes an encoder-decoder network architecture. Their results outperform other models with 96.47, 91.70, 85.06 and 90.41 in terms of Overall Accuracy, mean Intersection over Union, Kappa, and Class-wise Accuracy. In another cloud detection paper [4], the authors propose a method that starts by creating a significance map to highlight cloud and non-cloud differences. They then use an optimal threshold to get a rough cloud detection result. A robust detail map helps eliminate non-cloud regions from cloud candidates. Finally, guided feathering detects semitransparent cloud pixels around cloud boundaries. Although this method has limitations with bright non-cloud regions lacking detail, it shows good visual and quantitative performance with time optimization. Zhiwei Li [7] and co-authors have shared a global high-resolution cloud detection validation dataset online. Their paper introduces MSCFF, a deep-learning method for cloud detection in remote sensing images. MSCFF uses an encoder-decoder module to extract multi-scale features and a fusion module to combine them. MSCFF outperforms traditional methods and state-of-the-art deep learning models, especially in bright surface areas. Among all methods compared, MSCFF achieved the highest mIoU and F-score values, surpassing traditional methods like Fmask, MFC, and PRS, as well as newer methods like DeepLab and DCN.

III. Dataset

The dataset was generated using satellite images of Hurricane Ian obtained from the Maxar dataset [9]. A total of 746 images were collected, comprising images captured before and after the hurricane. Due to the satellite's positioning, some images had dark portions and significant cloud cover. To isolate only the cloud portions, each image was divided into 1024 square-shaped images. Subsequently, a black frame elimination filter was applied to eliminate black frames. The images were then provided to two annotators for annotation. Each annotator independently annotated the images into two classes: one class was for cloud and another class was for not cloud, and a third annotator conducted a final examination. The third annotator resolved any annotation errors by addressing conflicting decisions made by the initial annotators. Each image in the dataset has dimensions of 640 x 640 pixels. Finally, the resultant dataset consisted of **2,476** images, with **1342** (54.2%) labeled as Cloud and **1134** (45.8%) as Not Cloud images. The flowchart of the whole dataset creation process is shown in Figure 1

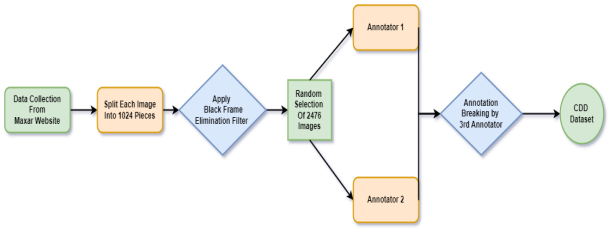

Fig. 1. Flowchart of Dataset Creation

IV. Methodology

To assess the performance of various transfer learning architectures, we employed five different models, namely vgg16, densenet169, resnet18, efficientnet_b0, and mobilenetv3_small_075. The selection of these models was based on their size and parameter counts, with the respective number of parameters for each model provided in V. Our CDD dataset

Fig. 2. Examples of CDD Dataset

was divided into three splits, and to ensure a fair benchmark, we maintained a consistent split ratio of 80:10:10 for all six models. To validate the performance of the dataset and model we have also evaluated the model learning in two different events of the Maxar dataset, we have used kalehe DRC flooding [17] and Indonesia Earthquake and Tsunami [18] from Maxar dataset.

A. CloudNet

To assess the efficacy of our dataset and identify clouds, we have developed CloudNet, a lightweight convolutional neural network (CNN). CloudNet is specifically designed for RGB images, prioritizing efficiency and simplicity. It implements a two-layer convolutional architecture, subsequently followed by completely connected layers. The initial convolutional layer processes three-channel RGB images into two by employing a 2x2 Max Pooling operation, a 3x3 kernel, and Rectified Linear Unit (ReLU) activation. Following this, the features are refined even further in the second convolutional layer, which converts the two channels into four. The flattened output from the convolutional layers is processed by the fully connected layers, which consist of a first fully connected layer with an input size of 4 * 56 * 56 and an output size of 8, which is activated by ReLU. Subsequently, for binary classification, a second fully connected layer is employed, which has an input size of 8 and an output size of 2. The efficient and uncomplicated architecture of CloudNet facilitates the detection of clouds in real time and offers valuable insights into the quality of the dataset via performance evaluation of binary classification.

V. Experimental Analysis

A. Experimental Setup

In this experiment, we use Google Colab Pro Plus. To be specific, we used the power of the Tesla T4 GPU for deep learning models. Our deep learning models were constructed using Python 3, PyTorch, and Poutyne, complemented by PyTorch Image Models (Timm). The table I presents the hyperparameter settings utilized for this research set for all the models to ensure consistency.

B. Results

In our study, we evaluated our cloud dataset using six CNN based models. Upon examining the detailed results

TABLE I
Hyperparameters for Model Training

Hyperparameter	Value
Image Dimension	(224, 224)
Number of Epochs	25
Batch Size	32
Learning Rate	1e-3
Mean	(0.485, 0.456, 0.406)
Standard Deviation (STD)	(0.229, 0.224, 0.225)

TABLE II
Volume of image and time comparison on Two Events

	Volume(A1)	Time(A1)	Volume(A2)	Time(A2)
Event-1	16384	711.41	5355	**444.89**
Event-2	38912	1685.24	8917	**801.31**

Note: A1 and A2 refer to volume and time without pre-processing and after pre-processing is done. Here, time is measured in seconds and Event-1 and 2 refer to Kalehe DRC flooding and Indonesia Earthquake and Tsunami event from the Maxar dataset.

of the pre-trained models in table III after 25 epochs of training, we observed that MobileNetV3 achieved the highest accuracy among the models. It demonstrated a training loss of 0.0033 and an impressive test accuracy of 98.38% on the test dataset. ResNet18 also demonstrated strong performance in both the validation set and training accuracy. Overall, all models performed well, exhibiting similar ranges of training accuracy between 98-99% and loss, with the exception of VGG16. Despite having the highest number of parameters, VGG16 performed poorly in detecting cloud and non-cloud images, achieving a training accuracy of 47%. It is noteworthy that despite having significantly fewer parameters than the other models, our custom CNN model, CloudNet, stood out. With only 100,510 parameters which is 10 time smaller compared to MobileNet's 1,018,922 parameters, CloudNet achieved outstanding results, boasting a training accuracy of 99% and a minimal training loss of 0.017. Furthermore, its test accuracy of 97.58% was competitive with that of the other models. In our evaluation of model performance for cloud and non-cloud images using Precision, Recall, and F1-score (as shown in table IV), MobileNetv3 achieved the highest F1-score of 99% with good precision and recall of 99% and 98% for the cloud class. In contrast, VGG16's complex network and higher number of parameters resulted in poorer performance across all metrics, yielding an F1-score of 72% on our smaller dataset. Additionally, recall for cloud classes was notably high across all models, with four achieving 100%. Our CloudNet model performed well in the cloud class, with a recall of 99% and an F1-score of 98%, despite having fewer parameters. For non-cloud classes, the F1-score remained stable at 97%.

Our cloud dataset helped optimize processing time, as shown in table II, where the time taken for training the dataset with CloudNet and YOLO is shown on two different datasets. Post-pre-processing, the training duration was nearly halved compared to the initial time, highlighting the dataset's impact on optimization.

TABLE III
MODEL PERFORMANCE METRICS

Model Name	Train Acc.	Train Loss	Validation Acc.	Validation Loss	Test Acc.	Test Loss
resnet18	99.54	0.010	99.193	0.0612	98.38	0.063
mobilenetv3_small_075	99.898	0.0033	98.790	0.0350	98.38	0.075
densenet169	99.191	0.037	99.19	0.040	98.79	0.03
efficientnet_b0	98.93	0.030	98.38	0.120	97.98	0.16
vgg16	47.272	0.694	48.387	0.697	55.64	0.693
CloudNet	99.29	0.017	98.387	0.060	97.580	0.080

TABLE IV
MODEL EVALUATION METRICS FOR CLOUD AND NOT CLOUD CLASSES

Model Name	Cloud			Not Cloud		
	Precision	Recall	F1-Score	Precision	Recall	F1-Score
resnet18	0.97	1.0	0.99	1.00	0.96	0.98
mobilenetv3_small_075	0.99	0.98	0.99	0.97	0.99	0.98
densenet169	0.98	1.0	0.99	1.00	0.97	0.99
efficientnet_b0	0.97	1.0	0.98	1.0	0.95	0.98
vgg16	0.56	1.0	0.72	0.0	0.0	0.0
CloudNet	0.96	0.99	0.98	0.99	0.95	0.97

TABLE V
NUMBER OF PARAMETERS IN DIFFERENT MODELS

Model	Number of Parameters
vgg16	134,268,738
densenet169	12,487,810
resnet18	11,177,538
efficientnet_b0	4,010,110
mobilenetv3_small_075	1,018,922
CloudNet	100,510

VI. CONCLUSION AND FUTURE WORK

In conclusion, we have presented a dataset of 2476 cloud and non-cloud images which can be used to detect the pattern of the clouds in aerial images. This detection can be necessary during the pre-processing phase for different aerial image tasks. In our work, we have demonstrated how after pre-processing different CNN models perform for object detection from aerial images. We have also observed that the overall time has been optimized by removing cloud images during the pre-processing stage. In the future, the dataset can be improved by increasing the number of images and adding domain-specific features related to cloud formation, movement, or other meteorological factors.

VII. ACKNOWLEDGEMENT

This research is funded in part by NSF Grant No. 2306109, and DOEd Grant P116Z220008. Any opinions, findings, and conclusions expressed here are those of the author(s) and do not reflect the views of the sponsor(s).

REFERENCES

[1] NASA, "Cloudy earth," https://earthobservatory.nasa.gov/images/85843/cloudy-earth, 2015, accessed: January 2, 2024.

[2] W. Xiao, H. Cao, M. Tang, Z. Zhang, and N. Chen, "3d urban object change detection from aerial and terrestrial point clouds: A review," *International Journal of Applied Earth Observation and Geoinformation*, vol. 118, p. 103258, 2023.

[3] C. Wang, L. Huang, and A. Rosenfeld, "Detecting clouds and cloud shadows on aerial photographs," *Pattern Recognition Letters*, vol. 12, no. 1, pp. 55–64, 1991.

[4] Q. Zhang and C. Xiao, "Cloud detection of rgb color aerial photographs by progressive refinement scheme," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 11, pp. 7264–7275, 2014.

[5] W. Yi, Z. Jing, and G. Shuang, "Hue–saturation–intensity and texture feature-based cloud detection algorithm for unmanned aerial vehicle images," *International Journal of Advanced Robotic Systems*, vol. 17, no. 3, p. 1729881420903532, 2020.

[6] Z. Zhang, Q. Liu, and Y. Wang, "Road extraction by deep residual u-net," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 5, pp. 749–753, 2018.

[7] Z. Li, H. Shen, Q. Cheng, Y. Liu, S. You, and Z. He, "Deep learning based cloud detection for medium and high resolution remote sensing images of different sensors," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 150, pp. 197–212, 2019.

[8] Y. Li, W. Chen, Y. Zhang, C. Tao, R. Xiao, and Y. Tan, "Accurate cloud detection in high-resolution remote sensing imagery by weakly supervised deep learning," *Remote Sensing of Environment*, vol. 250, p. 112045, 2020.

[9] Maxar, "Hurricane ian," https://www.maxar.com/open-data/hurricane-ian, 2021, accessed: January 2, 2024.

[10] G.-S. Xia, J. Hu, F. Hu, B. Shi, X. Bai, Y. Zhong, L. Zhang, and X. Lu, "Aid: A benchmark data set for performance evaluation of aerial scene classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 7, pp. 3965–3981, 2017.

[11] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems*, 2010, pp. 270–279.

[12] W. R. M. D. Stephan Rasp, Hauke Schulz, "Understanding clouds from satellite images," 2019. [Online]. Available: https://kaggle.com/competitions/understanding_cloud_organization

[13] G. Sheng, W. Yang, T. Xu, and H. Sun, "High-resolution satellite scene classification using a sparse coding based multiple feature combination," *International journal of remote sensing*, vol. 33, no. 8, pp. 2395–2412, 2012.

[14] Q. Zou, L. Ni, T. Zhang, and Q. Wang, "Deep learning based feature selection for remote sensing scene classification," *IEEE Geoscience and remote sensing letters*, vol. 12, no. 11, pp. 2321–2325, 2015.

[15] F. Xie, M. Shi, Z. Shi, J. Yin, and D. Zhao, "Multilevel cloud detection in remote sensing images based on deep learning," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 8, pp. 3631–3640, 2017.

[16] J. Yang, J. Guo, H. Yue, Z. Liu, H. Hu, and K. Li, "Cdnet: Cnn-based cloud detection for remote sensing imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 8, pp. 6195–6211, 2019.

[17] Maxar, “kalehe drc flooding,” https://www.maxar.com/open-data/kalehe-drc-flooding-2023, 2021, accessed: January 2, 2024.
[18] ——, “indonesia earthquake tsunami,” https://www.maxar.com/open-data/indonesia-earthquake-tsunami, 2021, accessed: January 2, 2024.