Attention-based CNN for Cloud Segmentation

Babur Nawyan
University of Toronto
babur.nawyan@mail.utoronto.ca

Humza Iqbal
University of Toronto
humza.iqbal@mail.utoronto.ca

Ekenedilichukwu Jideofor Akuneme
University of Toronto
ekene.akuneme@mail.utoronto.ca

Abstract

Satellite imagery, such as scene images from the Landsat 8 satellite, is vital in monitoring and understanding Earth's systems. These images are valuable data for various applications such as environmental monitoring, agriculture and urban planning, climate change studies, and disaster management. However, the presence of clouds in these images presents a significant challenge. Clouds can obscure important surface information, reducing data quality and introducing analysis errors. Accurately detecting and handling cloud-covered regions in satellite images is essential to improve image clarity, ensure accurate surface monitoring, and enable reliable time-series analysis. Effective cloud detection enables researchers and practitioners to harness the full potential of satellite image data, ensuring that decisions based on these images are reasonable and logical. This project aims to develop a deep learning-based cloud segmentation model using the 95-Cloud dataset from Landsat 8 satellite imagery. The proposed model seeks to improve segmentation accuracy in complex cloud environments by leveraging attention mechanisms within convolutional neural networks.

1 Introduction

Given input satellite image patches with multiple spectral bands, cloud detection algorithms perform binary segmentation and output a binary mask image where each pixel is predicted to be either part of a cloud-covered region or not part of a cloud-covered region. We propose a fully convolutional neural network inspired by the Cloud-Net [1], Cloud-Net+ [2], and Attention U-Net [3] algorithms paired with the the combined loss function of Binary Cross-Entropy (BCE) Loss and Dice Loss to achieve an effective cloud detection/segmentation model. The choice of the loss function above is based on our observation that the ground truth image patches in the 95-Cloud dataset are slightly imbalanced in terms of the number of positive pixels (pixels that are part of a cloud-covered region) that the image patches contain. This observation was made based on the exploratory data analysis in section 3. Given that the 95-Cloud dataset consists of highly accurate ground truth image patches and has been used to train and test the Cloud-Net+ algorithm, we will also make use of the 95-Cloud dataset for both the training and testing process. In addition, we will apply data augmentation techniques such as image patch rotation and flipping (horizontal and vertical) to help improve the robustness and generalization of the model. Lastly, the Jaccard Index, Precision, Recall, and Accuracy will be used as the evaluation metrics for the model as the baseline model (Cloud- Net+) is also evaluated on these four metrics. The formula for the Jaccard Index is given in section 2.

Cloud detection, along with estimation of cloud coverage, are among the most critical processes in the analysis of satellite imagery. Transferring remotely sensed data, such as satellite images from air/space-borne sensors to ground stations is an expensive process in terms of time, bandwidth,

storage, and computational resources. In addition, clouds can obscure important surface information, making it difficult to extract valuable data from optical images that are heavily cloud-covered. Since, on average, 67% of the Earth's surface is covered by clouds at any given time [4], a considerable amount of resources can be optimized and conserved by only transferring images with minimal cloud and shadow coverage. Consequently, an effective and accurate cloud detection the algorithm will enable researchers and practitioners in various fields such as environmental monitoring, agriculture, urban planning, climate change studies, etc to use satellite images more effectively by having a clear understanding of satellite images and what each pixel represents.

Additionally, an accurate understanding and detection of cloud-covered region in satellite images can provide useful information about climate and atmospheric parameters [5], as well as natural disasters such as hurricanes [6] and volcanic eruptions [7]. Thus, the detection of clouds and cloud shadows in satellite images are an essential pre-processing task for many applications.

However, cloud and cloud shadow detection is challenging when only a limited number of spectral bands are available, as there is less information for the model to learn how to detect clouds accurately. Many air/space-borne systems, such as ZY-3, HJ-1, and GF-2, are equipped only with visible (RGB) and near-infrared bands [8]. Thus, algorithms and models capable of performing accurate cloud detection using only a few spectral bands are highly valuable, which we aim to achieve by creating a neural network model. This model is also promising since traditional algorithms struggle with dynamic atmosphere conditions, neural networks can learn and make accurate predictions even with this variability.

2 Background and related work

Threshold-Based Approaches: Threshold-based techniques, like the commonly used FMask (Function of Mask), utilize decision trees with various spectral band thresholds to categorize pixels into clouds, shadows, or clear regions. Over the years, this approach has been refined, with Zhu et al. enhancing it for compatibility with Landsat and Sentinel datasets, enabling better handling of snow and cloud shadows [8]. While these methods are computationally efficient, they often struggle in diverse atmospheric conditions, particularly when it comes to differentiating thin clouds or haze [9].

Handcrafted Algorithms: Algorithms based on handcrafted features aim to utilize spectral information in a more targeted manner. For example, Haze Optimized Transformation (HOT) relies on correlations between spectral bands to separate clouds from clear areas [10]. However, these methods frequently require manual parameter adjustments and can be sensitive to variations in scenes, limiting their scalability for automated large-scale applications [11].

Deep Learning Approaches: The rise of deep learning techniques has significantly transformed cloud detection, offering higher accuracy and more automation. Convolutional Neural Networks (CNNs) have shown great success in cloud classification tasks, with models like those by Xie et al. utilizing multi-level CNN architectures to distinguish between thin clouds, thick clouds, and clear skies [12].

Fully Convolutional Networks (FCNs), especially those modeled after U- Net [13], have been particularly impactful in advancing semantic segmentation. For instance, the Cloud-Net model developed by Mohajerani and Saeedi employs an FCN framework to capture both localized and broad features essential for cloud detection [1, 14]. One of the advantages of Cloud-Net is that it bypasses the need for extensive pre-processing, such as super-pixel segmentation, making it well-suited for real-time applications.

Building on the Cloud-Net foundation, the improved Cloud-Net+ introduces a specialized Filtered Jaccard Loss function along with Sunlight Direction-Aware Augmentation (SDAA) to enhance cloud shadow segmentation and address diverse atmospheric conditions [2]. This custom loss function is designed to boost detection accuracy, especially when dealing with sparse cloud formations or shadows that are difficult to differentiate from the surrounding elements [2].

A further enhancement of the U-Net design is the Attention U-Net, which incorporates Attention Gates to selectively focus on the most pertinent features while reducing background noise [3]. This enhancement has proven successful in complex medical segmentation tasks, such as pancreas and liver segmentation in CT scans, where it has demonstrated higher Dice Similarity Coefficients (DSC) compared to the standard U-Net [3]. These attention gates refine the features transmitted

through skip connections, ensuring that the model emphasizes regions of interest without needing additional object localization steps. Incorporating attention mechanisms have been shown to greatly enhance segmentation accuracy, especially in scenarios where the target regions are small or have low contrast [3]. Given the proven effectiveness of attention-based models in medical imaging, our work extends this idea to cloud detection.

Cloud detection in remote sensing imagery has been approached using a variety of methods over the years, including threshold-based techniques, handcrafted algorithms, and, more recently, deep learning models. Our model builds on these deep learning techniques and stands out due to:

- **Attention-Gated Skip Connections:** Unlike Cloud-Net+, we integrate attention mechanisms to focus on ambiguous cloud regions, improving the detection of thin clouds and shadows.
- **Combined Loss Functions:** By combining BCE and Dice Loss, our model addresses class imbalance, which existing models often struggle with.

Loss function: Our approach includes the use of a loss function that combines Binary Cross-Entropy (BCE) Loss and Dice Loss which we denote as CL (Combined Loss). Let y be the model's prediction while t is the corresponding ground truth and let N denote the total number of pixels in t. In addition, let ϵ be a small positive number (such as 10^{-7}) to avoid division by zero in the Dice Loss and let $\lambda \in (0, 1)$ which is weight of the BCE Loss in the combined loss function (CL). Consequently, $1 - \lambda$ is the weight of the Dice Loss in the combined loss function (CL). The larger the value of λ, the closer the combined loss function gets to the BCE Loss while the smaller the value of λ, the closer the combined function gets to the Dice Loss. Lastly, let tp, tn, fp, fn be the numbers of true positive, true negative, false positive, false negative pixels for each class in each test set scene while M is the total number of scenes in the test data. Then, the Binary Cross-Entropy Loss, Dice Loss, Combined Loss, and the Jaccard Index is defined as follows.

$$\text{BCE}(y,t) = -\frac{\sum_{i=1}^{N}\left[t_i \log(y_i) + (1-t_i)\log(1-y_i)\right]}{N}, \qquad \text{Dice}(y,t) = 1 - \frac{2\sum_{i=1}^{N} y_i t_i}{\sum_{i=1}^{N} y_i + \sum_{i=1}^{N} t_i + \epsilon}$$

$$\text{CL}(y,t) = \lambda \text{BCE}(y,t) + (1-\lambda)\text{Dice}(y,t), \qquad \text{Jaccard}(tp,fp,fn) = \frac{\sum_{i=1}^{M} tp_i}{\sum_{i=1}^{M}(tp_i + fp_i + fn_i)}$$

3 Data

The 98-Cloud dataset's training set, which is an extension of the 38-Cloud dataset's training set consists of 34701 non-overlapping patches of 384×384 pixels extracted from 75 Landsat 8 Collection 1 Level-1 scenes. The scene images are mostly located in North America. The test set of the 95-Cloud dataset includes 9201 patches of 20 scenes which is exactly the same as the 38-Cloud dataset's test set. There is no image patch in the dataset with an invalid pixel value such as a pixel value less 0 or NaN (Not a Number) value.

Input Data Each input image patch in the dataset consists of 4 corresponding spectral channels which are Red (band 4), Green (band 3), Blue (band 2), and Near Infrared (band 5) where these channels are not combined into a single image patch. Instead, there are 4 separate image patches for each of the spectral channels above. Each input image patch is a 384×384 grayscale TIFF file with a 16-bit depth, meaning that the pixel values range from 0 to 65535. The 16-bit depth format for TIFF files is prevalent in scientific and medical imaging where greater grayscale precision is needed. Our program combines these four different corresponding input image patches into a single input image patch with the shape of $4 \times 384 \times 384$ where 4 is the number of the channels as a first step before applying any other data processing or augmentation techniques.

Ground Truth Data The ground truth image patches are 384×384 grayscale TIFF files with 8-bit depth, meaning that the pixel values range from 0 to 255. In the 95-Cloud dataset, the ground truth patches only contain pixel values of 0 and 255. The pixel value 255 indicates that the pixel is part of a cloud-covered region in the image while the pixel value 0 indicates that the pixel is not part of a cloud-covered region.

Dataset Features Table 1 below contains some numerical features of the folders containing the 4 spectral channel image patches. The table contains the mean pixel value and the distribution of

the pixel values of the image patches for each of the 4 spectral channels. Table 1 below contains the distribution of the pixel value 0 in the ground truth image patches. It is important to note that the table contains information regarding the ground truth of the training set and the ground truth image patches have a pixel value of either 0 or 255. We can observe that approximately 48% of all 34701 ground truth image patches contain solely the pixel value 0 (which are the black pixels). It is important to note that if we get rid of these image patches, then the number of files in the 90% to 100% range will be $21056 - 16772 = 4284$. Then, the distribution in Table 1 will be approximately symmetric which implies that the dataset will become more balanced. Additionally, these patches make up approximately 48% of the entire training set (which gets split into the actual training and validation set). Thus, we removed these patches to decrease the training time significantly.

Table 1: Distribution of pixel value 0 in ground truth (GT) image patch files

Percentage of pixel value 0 in GT images (%)	**Number of files**	**Percentage of total files (%)**
0% to 10%	6251	18.01%
10% to 20%	901	2.59%
20% to 30%	787	2.26%
30% to 40%	752	2.16%
40% to 50%	847	2.44%
50% to 60%	816	2.35%
60% to 70%	950	2.73%
70% to 80%	1091	3.14%
80% to 90%	1250	3.60%
90% to 100%	21056	60.67%
Exactly 100%	16772	48.33%

Data Augmentation In an attempt to improve the robustness and generalization of the model, we will apply data augmentation techniques to the training set. The input and ground truth image patches will be augmented with horizontal flips with a probability of 0.4, vertical flips with a probability of 0.4, and random rotations by 90-degree increments with a probability of 0.4. Additionally, the input image patches will also be augmented with random brightness and contract adjustment of up to $\pm 20\%$ with a probability of 0.1. The input image patches will also be augmented with randomly added Gaussian noise to the image pixel values. The range of the variance of the random Gaussian noise is 3 to 10 and this augmentation technique is applied with a probability of 0.2. To ensure that the pixel values of the input image patches are valid after its pixel values are modified through the data augmentation techniques, the pixel values are clamped to the range of 0 and 65535 which is valid pixel value range for 16-bit depth TIFF images. Lastly, the pixel values of both the input and ground truth image patches are scaled down to the range of $[0, 1]$ by dividing every pixel value of the input image patches by 65535 and the ground truth image patches by 255. All of the data augmentation techniques were implemented with Albumentations.

4 Model architecture

Integrating Attention Mechanisms with Cloud Detection: Building upon the successes of both Cloud-Net and Attention U-Net, our proposed model aims to leverage the strengths of these architectures for cloud detection in satellite imagery. The model incorporates elements from Cloud-Net, which efficiently captures multi-scale features using a Fully Convolutional Network (FCN) structure, while also integrating attention mechanisms inspired by the Attention U-Net.

Architecture Overview: Our model adopts an encoder-decoder structure similar to the original U-Net, enhanced with attention gates to focus on critical cloud regions while ignoring irrelevant background noise. Consult Figure 1 for an illustration of the main elements of the model's architecture. We used PyTorch to implement our model. The key components of our architecture are:

1. **Encoder**: The encoder extracts hierarchical features using convolutional layers and max-pooling. Each block includes two convolutional layers followed by batch normalization and ReLU activation. These layers capture both high-level contextual information and low-level spatial details. In the model, we have 4 encoder blocks.

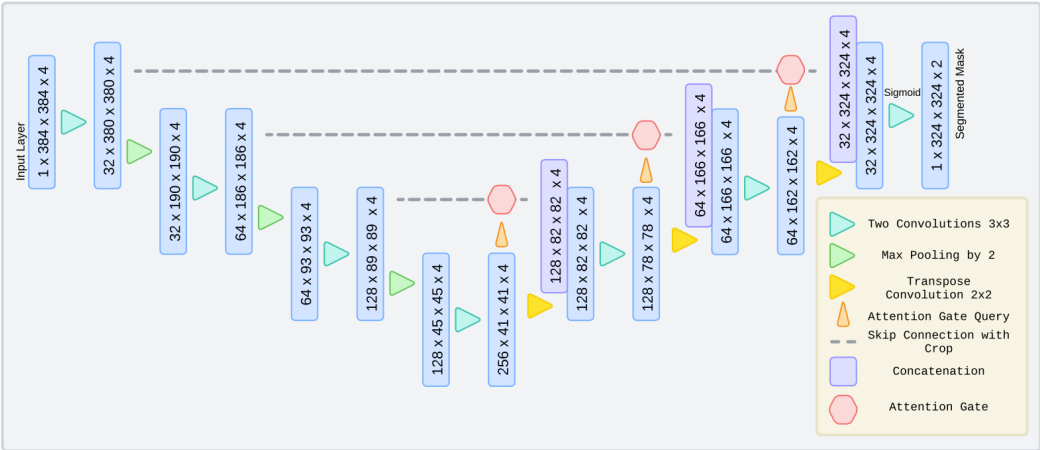

Figure 1: Proposed model architecture for cloud detection. The architecture integrates attention gates with the encoder-decoder structure of U-Net.

2. **Attention-Gated Skip Connections:** Attention gates filter the skip connections before feeding them into the decoder. The attention gates refine features based on context from the decoder and spatial information from the encoder. We explain the attention computation below. Skip connections themselves also prevent information from being lost in down-sampling. We also crop these connections so the shapes match. We have 3 attention blocks.
3. **Decoder:** The decoder reconstructs the high-resolution segmentation mask using transpose convolutions and concatenates the attention-filtered features from the encoder. Each decoder block includes a transpose convolution layer followed by a convolutional block. We have 4 decoder blocks.
4. **Final Convolution:** After the last decoder block, a 1x1 convolution layer maps the output to a single-channel segmentation mask.
5. **Loss Function**: The model will use a combination of Binary Cross-Entropy (BCE) and Dice Loss which we denote as CL (Combined Loss). BCE Loss is often used as baseline loss for binary segmentation. However, it can struggle with an imbalanced dataset where the number of positive and negative pixels are quite different. On the other hand, Dice Loss performs well when the dataset is imbalanced as it focuses on overlapping regions, improving boundary segmentation. Since we have observed that the ground truth image patches in the 95-Cloud dataset are slightly imbalanced in terms of how many image patches contain negative pixels (pixels that are not part of a cloud-covered region) in certain ranges, the combination of BCE and Dice Loss is the most appropriate given the 95-Cloud dataset and our model architecture.

Attention Gate Computation: The encoder features and decoder features are passed through 1x1 convolutional layers and batch normalization to reduce their dimensionality. The transformed features are element-wise added and passed through a ReLU activation to compute intermediate relevance scores. A 1x1 convolution followed by a sigmoid activation generates an attention map that highlights relevant spatial regions. The encoder features are multiplied by the attention map, passing only context-relevant information to the decoder.

Hyperparameter Settings:

- **Total Layers:** We have a total of 13 convolution layers, 4 transpose convolution layers, 4 batch norm layers, and 3 attention/concatenation blocks.
- **Optimizer & Learning Rate Scheduler:** Adam optimizer with learning rate `1e-6`. We also tried `1e-1`, `1e-2`, `1e-3`, `1e-4`, `1e-5` for the learning rate. We use `ReduceLROnPlateau` with factor=0.1, patience=4, and mode='min'. Every time we ran a new session of training, we changed the factor and patience parameters.

- **Batch Size:** 16
- **Shuffling:** Training data shuffled; validation and test sets not shuffled.

Expected Benefits: By integrating attention mechanisms with the efficient FCN backbone of Cloud-Net, our model aims to:

- Improve segmentation accuracy by focusing on complex cloud patterns and reducing false positives.
- Enhance the model's ability to detect cloud shadows, which are often confused with other landscape features.
- The use of a combination of Binary Cross-Entropy (BCE) Loss and Dice Loss helps address the challenge of imbalanced datasets, ensuring that the model performs well even when cloud coverage is sparse.

Our approach combines the strengths of attention-based segmentation models in the medical imaging domain with specialized cloud detection techniques, creating a robust solution for satellite imagery analysis.

5 Results

Performance on 95-cloud test set Our model was trained entirely on the training set of the 95-cloud dataset where the original training set was split into the actual training set and validation set. Table 2 summarizes and compares the performance of our model and the baseline models Cloud-Net and Cloud-Net+ paired with the loss functions FJL_1 and FJL_2 which are proposed in the Cloud-Net+ paper. The proposed model's performance below is based on the threshold 0.7 where the model's output is binarized into 0 (for negative) and 1 (for positive) using this threshold. Note that the Cloud-Net+ model uses 0.5 as its threshold to binarize predictions.

Table 2: Comparison of the proposed model with baseline models on the 98-Cloud dataset (in $\%$)

Model	Jaccard Index	Precision	Recall	Accuracy
Our proposed model	81.56	89.29	89.58	96.72
Cloud-Net	90.83	**97.67**	92.84	97.00
Cloud-Net+ + FJL_1	**91.57**	96.64	**94.28**	**97.23**
Cloud-Net+ + FJL_2	91.01	97.49	93.19	97.06

Evaluation metrics We evaluated our model's performance on the four metrics of Jaccard Index, Precision, Recall, and Accuracy since these metrics are used in the evaluation of the Cloud-Net and Cloud-Net+ models. Since the Jaccard Index directly measures the overlap between the predicted and ground truth, it is robust against class imbalance while Precision gives insight into the quality of the predicted positive regions in the image. Recall emphasizes how much of the ground truth region is correctly captured by the model and accuracy measures the proportion of correctly classified pixels which makes it a simple and intuitive metric. These four metrics combined gives us a good insight into the model's performance on the task of binary segmentation.

Training loss and validation loss The proposed model was trained with mini-batches of size 16. The value of the combined loss function of BCE and Dice Loss has been averaged for each epoch. During each epoch, after the model has been trained on every mini-batch, the model was evaluation with the combined loss function where the $\lambda = 0.5$. Hence, the BCE Loss and the Dice Loss contributed equally to the combined loss function. Figure 2 is the plot of the train and validation loss of the proposed model over epochs.

6 Discussion

First, it can be observed from that the proposed model has performed worse than the baseline models in terms of the score of Jaccard Index.

Figure 2: Average loss value for the training and validation set over epochs

The results of our model, as shown in Table 2, provide valuable insights when compared to baseline models like Cloud-Net and Cloud-Net+ (with F JL1 and F JL2 loss functions). Below is a summary of the key findings:

Interpretation of the test set performance: Our model achieved a Jaccard Index of 81.56%, lower than Cloud-Net and Cloud-Net+ (both above 91%). This metric, which measures overlap between predicted and ground truth regions, highlights that our model struggles with fine-grained segmentation. This may be due to differences in architecture or optimization strategies. Improvements such as advanced loss functions or enhanced feature extraction could help address this gap.

Our Precision score of 89.29% is high but below the Precision score of the baseline models (96.64%-97.67%), suggesting some mis-classification of non-cloud regions. However, the Recall of 89.58% demonstrates strong sensitivity to true cloud regions, capturing most cloud pixels but occasionally overestimating their presence. This tradeoff indicates a slight bias toward capturing all possible clouds, which could be useful in applications where missing clouds is costly.

With an Accuracy of 96.72%, our model performs close to the baselines (97.00%-97.23%). While this shows reliable overall classification, accuracy alone does not address the challenges of segmenting nuanced cloud boundaries or handling class imbalances.

Interpretation of training and validation loss: The combined loss function (BCE and Dice Loss with $\lambda = 0.5$) ensured stable learning, as evidenced by the convergence of training and validation loss. However, the modest Jaccard Index in contrast to the extremely high accuracy suggests that the current loss formulation could be improved to enhance overlap accuracy, potentially by using a smaller weight λ for BCE and hence a larger weight for Dice Loss.

Model complexity and inference speed It is important to note that the added complexity of the attention mechanism increases the model complexity and size, potentially leading to longer inference time. However, since satellite images are not frequently captured and processed through neural network models in real-time, the added complexity of the attention mechanism can possibly can allow the model to adapt well to diverse datasets, potentially outweighing the disadvantages of the model complexity and inference speed.

While our model demonstrates solid performance, it falls short of the baselines, highlighting the benefits of the architectural innovations and specialized loss functions in Cloud-Net+. Their higher scores suggest better handling of complex spatial and spectral features in satellite imagery.

7 Limitations

Dataset Bias: The 95-Cloud dataset mainly consists of images from North America. This limited geographic diversity can introduce bias because the dataset doesn't represent cloud patterns from other parts of the world. For example, cloud formations in tropical regions, like the Amazon, are quite different from those in temperate regions like North America. If the model is trained only on this dataset, it may not perform well when applied to satellite imagery from regions with drastically different atmospheric conditions. This limitation reduces its global applicability, making it less effective in real-world, diverse environments.

Generalization: Deep learning models often learn features specific to the training dataset. If the 95-Cloud dataset doesn't include a wide variety of atmospheric conditions (e.g., dust storms, high-altitude cirrus clouds, or volcanic ash clouds), the model might struggle to detect clouds in such novel

situations. Essentially, the model becomes too reliant on patterns it has seen before, and without exposure to diverse conditions, it may underperform when deployed in different real-world scenarios. This lack of generalization can significantly impact the reliability of the model.

Computational Cost: Attention mechanisms are great for improving segmentation accuracy because they help the model focus on the most relevant parts of an image. However, this comes with a downside: increased computational cost. Calculating attention weights adds complexity to the model, making it slower and more resource-intensive. This is especially problematic when dealing with high-resolution satellite images, where the computational and memory requirements can become overwhelming. For real-time applications or deployments on devices with limited resources, this added complexity could be a significant bottleneck.

8 Ethical considerations

There are various ethical considerations in developing a neural network model that performs cloud detection on satellite images. These ethical considerations include privacy concerns, bias in the neural network model, repurposed usage of the model, etc.

Privacy Concerns Even though the 95-Cloud dataset is created from the Landsat 8 scenes which are publicly available, misuse of the satellite images in the dataset could lead to unintended surveillance of private areas and/or identification of certain locations, potentially revealing sensitive information about individuals, communities, businesses, government agencies, etc.

Bias in Model The 95-Cloud dataset is created from Landsat 8 scene images. These scene images are captured by the Landsat 8 satellite which is part of a program by NASA and the U.S. Geological Survey. Since the scene images in the 95-Cloud dataset are mainly from North America, a model trained on this dataset will likely contain some bias. Thus, the usage of this model with input satellite image patches that are captured in North America might result in a different behavior than when the model is used with input image patches that were captured outside of North America.

Repurposed Usage of the Model Cloud detection models could be repurposed for military applications, corporate surveillance, and criminal activity. Given a model that can accurately detect cloud-covered region in satellite image patches, it can be repurposed for surveillance of certain countries, monitoring competitor facilities in the same industry, or planning large-scale criminal activities that might be preventable by surveillance performed through satellite images.

9 Conclusion

We presented a deep learning-based cloud detection model leveraging attention mechanisms and a U-Net-inspired architecture. Evaluated on the 95-Cloud dataset, our approach demonstrated promising performance, though it fell slightly short of established models like Cloud-Net+. The integration of attention gates improved the model's ability to focus on complex cloud structures, indicating the potential of attention-based segmentation in remote sensing applications.

Despite its strengths, the model faces challenges related to geographic generalization and computational demands. Expanding the dataset to include diverse atmospheric conditions from different regions could mitigate dataset bias and improve performance in real-world scenarios. Additionally, optimizing the computational efficiency of attention gate implementations could enable faster processing and real-time deployment.

Future work will focus on enhancing the model by incorporating advanced loss functions, experimenting with alternative attention mechanisms, and exploring transfer learning techniques. These enhancements aim to create a more robust, scalable, and globally applicable cloud detection system suitable for various remote sensing tasks.

References

[1] Sorour Mohajerani and Parvaneh Saeedi. Cloud-net: An end-to-end cloud detection algorithm for landsat 8 imagery. *arXiv preprint arXiv:1901.10077*, 2019.

[2] Sorour Mohajerani and Parvaneh Saeedi. Cloud and cloud shadow segmentation via filtered jaccard loss and parametric augmentation. *arXiv preprint arXiv:2001.08768*, 2021.

[3] O. Oktay, J. Schlemper, L. Le Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N.Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert. Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*, 2018.

[4] Michael D King, Steven Platnick, W Paul Menzel, Steven A Ackerman, and Paul A Hubanks. Spatial and temporal distribution of clouds observed by modis onboard the terra and aqua satellites. *IEEE transactions on geoscience and remote sensing*, 51(7):3826–3852, 2013.

[5] Cyrus Raza Mirza, Toshio Koike, Kun Yang, and Tobias Graf. The development of 1-d ice cloud microphysics data assimilation system (imdas) for cloud parameter retrievals by integrating satellite data. In *IGARSS 2008-2008 IEEE International Geoscience and Remote Sensing Symposium*, volume 2, pages II–501. IEEE, 2008.

[6] R.S. Reddy, D. Lu, F. Tuluri, and M. Fadavi. Simulation and prediction of hurricane lili during landfall over the central gulf states using mm5 modeling system and satellite data. In *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 36–39, 2017.

[7] L. Zhu, M. Wang, J. Shao, C. Liu, C. Zhao, and Y. Zhao. Remote sensing of global volcanic eruptions using fengyun series satellites. In *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pages 4797–4800, 2015.

[8] Zhe Zhu, Sheng Wang, and Curtis E. Woodcock. Improvement and expansion of the fmask algorithm: Cloud, cloud shadow, and snow detection for landsats 4-8 and sentinel-2 images. *Remote Sensing of Environment*, 159:269–277, 2015.

[9] Zhe Zhu and Curtis Woodcock. Object-based cloud and cloud shadow detection in landsat imagery. *Remote Sensing of Environment*, 118:83–94, 2012.

[10] Y. Zhang, B. Guindon, and J. Cihlar. An image transform to characterize and compensate for spatial variations in thin cloud contamination of landsat images. *Remote Sensing of Environment*, 82:173–187, 2002.

[11] Yifang Yuan and Xiaoqiang Hu. Bag-of-words and object-based classification for cloud extraction from satellite imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8:4197–4205, 2015.

[12] Fang Xie, Min Shi, Zhiwei Shi, Jianbo Yin, and Dongbin Zhao. Multilevel cloud detection in remote sensing images based on deep learning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10:3631–3640, 2017.

[13] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *arXiv preprint arXiv:1505.04597*, 2015.

[14] Sorour Mohajerani, T. A. Krammer, and Parvaneh Saeedi. A cloud detection algorithm for remote sensing images using fully convolutional neural networks. In *IEEE International Workshop on Multimedia Signal Processing (MMSP)*, pages 1–5, 2018.