MaskRCNN and UNet on Understanding Clouds from Satellite Images

Jiaqi Tang, Jingjing Lin, Xinyi Ye, Yeqing Liu

Abstract

This project compared two methods for cloud satellite image identification, MaskRCNN, and UNet. The dataset is from NASA Worldview, including 22184 images, and there can have one or more types of clouds in each image, which are Fish, Flower, Gravel, and Sugar. The process consists of be divided into 2 parts, detection, and identification. In detection, the clouds were segmented from the remote sensing imagine for classification with the pre-trained COCO dataset. In identification, multitask classification was conducted. We found that UNet is faster than MaskCNN, and it also has a lower log loss at its best epoch compared to MaskRCNN. The accuracy of the best model is 0.550.

Introduction

Climate change has been a serious topic in recent years and even influenced political decision making. As increasing emissions of greenhouse gas results in the rise in global temperatures, atmospheric scientists need to know 'How will these shallow clouds react?'. Finding better ways to capture clouds in climate models is one of the fastest-growing priorities among climate scientists (Chelsea Harvey, 2019). Clouds have different shapes and different features. Small changes in the way clouds organize matter a lot for the climate. In particular, shallow clouds reflect a significant portion of the incoming solar radiation back into space and thereby help to keep the planet cool (Diane Toomey, 2017). Therefore, a better physical understanding of clouds could help build better climate models. Although human eyes are pretty good at detecting features of different clouds, it is challenging to build traditional rule-based algorithms to separate cloud features. The true-color images were taken from two polar-orbiting satellites by NASA, and each image may contain more than one type of clouds. Hence, the most challenging task in this deep learning project is to segment the picture and detect the location of clouds in pictures and we selected MaskRCNN and UNet as two deep learning algorithms to solve the challenge and to classify clouds.

Related Work

We will discuss the related work about these two algorithms, MaskRCNN and UNet separately since they are definitely different methods and both of them were widely used for image segmentation.

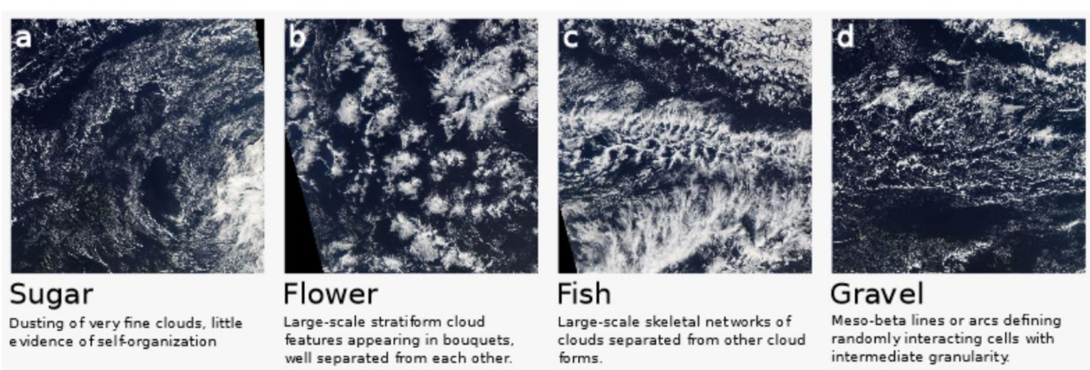
MaskRCNN:

A small team at UC Berkeley, led by Professor Jitendra Malik focused on the topic 'generalize to object detection' and figured out that a CNN model can lead to dramatically higher object detection performance on images than systems based on simpler HOG-like features. The result of the R-CNN comes out with bounding boxes and labels for each object in the image (Dhruv Parthasarathy, 2017). In contrast to faster RCNN, the researcher at Facebook AI, Kaiming He and his team went one step further and locate exact pixels of each object instead of just bounding boxes, which is known as MaskRCNN. Mask R-CNN outperforms all existing, single-model entries on every task (Kaiming He et al, 2017).

UNet:

Researchers from BIOSS Centre for Biological Signalling Studies at the University of Freiburg utilized UNet to do biomedical image segmentation and concluded that the u-net architecture achieves very good performance on very different biomedical segmentation applications (O. Ronneberger et al, 2015). UNet also works efficiently in segmenting high-resolution satellite images, the article (Guillaume Chhor, 2017) stated that the accuracy reached 94.5% by using U-net on two-dimensional satellite images.

Datasets



The dataset includes 22184 satellite images of 1400 x 2100 px that contain certain cloud formations, with label names: Fish, Flower, Gravel, Sugar from NASA Worldview. Each image has at least one cloud formation, and can possibly contain up to all four. The labels were created in a crowd-sourcing activity at the Max-Planck-Institute for Meteorology in Hamburg, Germany, and the Laboratoire de météorologie Dynamique in Paris, France.

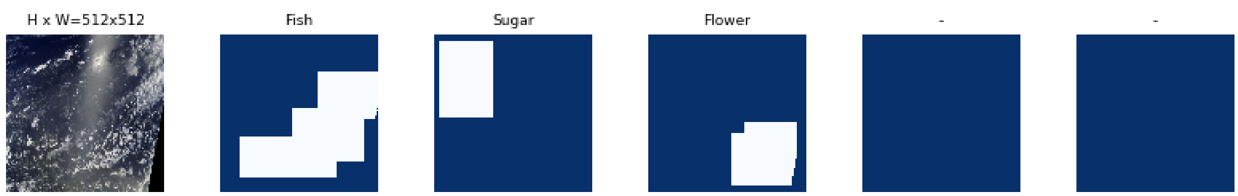
Methods

MaskRCNN:

Detection: Pre-trained model with MS COCO to segment objects. Ideally, we would follow the following steps to achieve the detection function:

- Anchor sorting and filtering
- Bounding box refinement
- Mask generation
- Layer activations
- Weight histograms
- Logging to tensor board
- Composing different results to a final result (we only need to detect cloud)

After the detection step, we will segment the cloud which is with a higher probability from the remote sensing imagine for classification. The images below which contains ‘white parts’ are the results of training set after segmentation detection. After segmentation detection, full connected CNN was utilized for classification in order to classify this segmentation into different types of clouds.

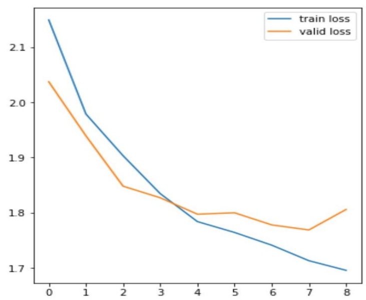


UNet:

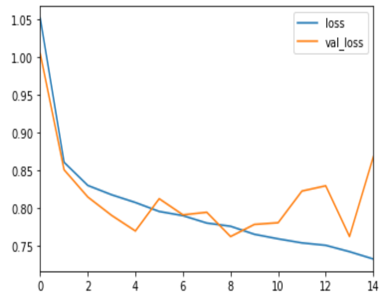
Unlike Mask RCNN, which is a pixel to pixel method and really useful for vehicle detection. **UNet, which** can produce only one mask, is more appropriate for medical images and remote sensing imagery. The unit contains two paths. The first path is the contraction path (also called the encoder) which is used to capture the context in the image. The encoder is a stack of convolutional and max-pooling layers. The second path (decoder) is used to enable precise localization using transposed convolutions. Since the satellite imagines don’t have an obvious boundary for cloud segmentation, unlike human, vehicle photos. The UNet method has been used for comparing the results with the counterpart from Mask RCNN.

Results

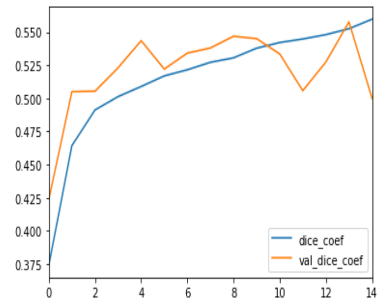
Loss vs. Epoch in MaskRCNN



Loss vs. Epoch in Unet



Accuracy VS Epoch in Unet



Discussion of Results

The best epoch for MaskRCNN is 8, with a validation loss of 1.768. The best epoch for UNet is 12, with a validation loss of 0.77. The performance of UNet has less loss than the counterpart of Mask RCNN. In this case, UNet tends to have better performance and higher accuracy. The accuracy of the best UNet model is 0.550.

Conclusions and Future Work

According to the results, UNet tends to perform better in segmentation detection and classification of satellite imagery, compared to MaskRCNN. And these results are able to be used for future research for cloud classification and other remote sensing image studies.
However, there are still some improvement should be made in future. First, Cross-validation will be supposed to be utilized for obtaining average accuracy in future work. Secondly, more optimizers should be used in the future. There are other Semantic Segmentation methods for satellite imagery classification, it is supposed to be utilized for future work.

References

[1] Chelsea Harvey, 2019, *Clouds May Hold the Key to Future Warming*, E&E NEWS.
[2] Diane Toomey, 2017, *Investigating the Enigma of Clouds and Climate Change.*

[3] Dhruv Parthasarathy, 2017, *A Brief History of CNNs in Image Segmentation: From R-CNN to Mask R-CNN.*
[4] Kaiming He et al, 2017, *MaskRCNN*, International Conference on Computer Vision (ICCV).
[5] O. Ronneberger, P. Fischer, T. Brox, 2015, *U-Net: Convolutional Networks for Biomedical Image Segmentation,* Computer Science Department and BIOSS Centre for Biological Signalling Studies, University of Freiburg, Germany.
[6] Guillaume Chhor, 2017, *Satellite Image Segmentation for Building Detection using U-net*