Contents

Identify Sandstorm Using Deep Learning

Abdallah Elsayed
Abdulrahman Ali
Abdulrahman Magdy
Abdulrahman Ashraf
Abdulrahman Yasser
Abdulrahman Abdelnasser
Supervisiors:
Dr. Marwa
Dr. Sayed
Dr. Salwa

May 2023

Abstract

Sandstorms are natural phenomena that affect the environment and human activities common in arid and semi-arid regions[1]. Identifying sandstorms from satellite images is challenging due to the complex and variable characteristics of sand, land surface and clouds. In this study, we propose a novel method for sandstorm identification using deep learning (DL). We tried to identify different sources of sandstorms by applying two different Convolution Neural Network models (U-NET and DeepLapV3). These CNN models are tested over a benchmark dataset from Moderate Resolution Imaging spectroradiometer (MODIS) that has the characteristics of high resolution and provides measurements in large-scale global dynamics including changes in Earth's cloud cover[2]. The performance that the two models achieve is that the accuracy of DeepLabV3 was 94%, compared to U-Net's 91% Similarly, the MIOU of DeepLabV3 was 88%, while U-Net's was 85%.

1 Introduction

Sandstorms, also known as dust storms, occur when strong winds pick up and carry large amounts of sand and dust particles, often resulting in reduced visibility and air quality[3]. These natural phenomena can have severe impacts on agriculture, infrastructure, and human health, particularly in arid and semi-arid regions. Therefore, early detection and monitoring of sandstorms are critical for effective mitigation measures and disaster management[4].

Traditionally, sandstorm detection and monitoring have relied on ground-based sensors and meteorological data. However, these methods have limitations in terms of spatial coverage and accuracy. With the advancement of remote sensing technologies, satellite or aerial images have become a valuable source of information for sandstorm detection and monitoring. Semantic image segmentation is a computer vision technique that involves assigning a label to each pixel in an image based on its semantic meaning[5]. This technique has shown promising results in detecting and segmenting sandstorms from satellite or aerial images. Among the various semantic image segmentation models, U-Net and DeepLabV3 are two popular choices[6]. In this research paper, we compare the performance of U-Net and DeepLabV3 in accurately detecting and segmenting sandstorms. We evaluate the models using mean Intersection over Union (MIOU) and accuracy metrics and visualize the segmentation results to assess their performance.

Our findings indicate that DeepLabV3 outperforms U-Net in terms of accuracy and robustness due to its more advanced architecture and optimization techniques. DeepLabV3 employs an encoder-decoder architecture with an atrous spatial pyramid pooling module to capture multi-scale contextual information effectively[7]. Additionally, it uses a combination of cross-entropy loss and Lovasz-Softmax loss to improve segmentation performance[8].

However, depending on the specific use case and available resources, U-Net can also be a viable option. U-Net is a simpler architecture that can provide fast and accurate segmentation results with less computational cost[9]. Overall, our research contributes to the ongoing efforts to develop accurate and efficient methods for sandstorm detection and monitoring using semantic image segmentation models. These models have the potential to improve the understanding and management of sandstorms, ultimately reducing their impacts on the environment and human health.

2 Related work

2.1 Sandstorm

Several studies have investigated the causes and effects of sandstorms:

- One study analyzed the impact of sandstorms on air quality in Riyadh, Saudi Arabia. The study found that the concentration of particulate matter (PM10) increased significantly during sandstorms, which can cause respiratory problems and other health issues.[10]
- Another study investigated the impact of sandstorms on agriculture in China. The study found that sandstorms can reduce crop yields and damage crops by burying them in sand or causing abrasion to their leaves.[11]
- A study analyzed the impact of sandstorms on power transmission lines in China. The study found that sandstorms can cause power outages by covering insulators with sand, reducing their insulation performance.[12]

2.2 U-Net

Several studies have explored the use of U-Net for image segmentation in various fields, including medical imaging, remote sensing, and computer vision. However, there are limited studies that investigate the application of U-Net for sandstorm detection and segmentation:

- In one study, researchers proposed a deep learning-based approach for sandstorm detection using U-Net. The model was trained on a dataset of satellite images and achieved 87% accuracy in detecting sandstorms.[13]
- Another study utilized U-Net for sandstorm segmentation in satellite images. The authors applied transfer learning to fine-tune the pre-trained U-Net model on their dataset and achieved promising results.[14]
- In a different study, U-Net was used for sandstorm detection in videos captured by a ground-based camera. The authors applied a modified version of U-Net that included temporal information to improve the model's performance.[15]

2.3 DeepLabV3

DeepLabV3 has been widely used in various fields for semantic segmentation tasks, including medical imaging, autonomous driving, and remote sensing:

- In one study, researchers proposed a modified version of DeepLabV3 for semantic segmentation of high-resolution satellite images. The authors introduced a novel multi-scale fusion module to improve the model's accuracy and efficiency.[16]
- Another study utilized DeepLabV3 for semantic segmentation of urban scenes in autonomous driving scenarios. The authors applied atrous spatial pyramid pooling and conditional random fields to refine the segmentation results.[17]
- In a different study, DeepLabV3 was used for vegetation segmentation in remote sensing images. The authors applied transfer learning to fine-tune the pre-trained DeepLabV3 model on their dataset and achieved state-of-the-art results.[18]

3 Methodology

In this section, we talk about U-Net architecture and DeepLabV3 approaches for sandstorm segmentation and explain each part of the two architectures.

3.1 U-Net

U-Net is a state-of-the-art convolutional auto-encoder network which is originally developed for biomedical image segmentation [19]. The main reason that we use U-Net instead of classical auto-encoder is in auto-encoder we use downsampling (encoder) and upsampling (decoder) but it loses some information about mapping the low-level features into a high-level feature. U-Net overcome this issue by using concatenation between the encoder (low-level or contraction path) and decoder(high-level or expansive path). The model is divided into three parts: the encoder, bridge, and decoder units. The encoder extracts low-level features, and the decoder extracts the corresponding high-level features, The expansive path is more or less symmetric to the contracting path, and yields a U-shaped architecture. Several feature channels are added to create a path between the low-level features and the corresponding high-level features to facilitate backward propagation during training [20].

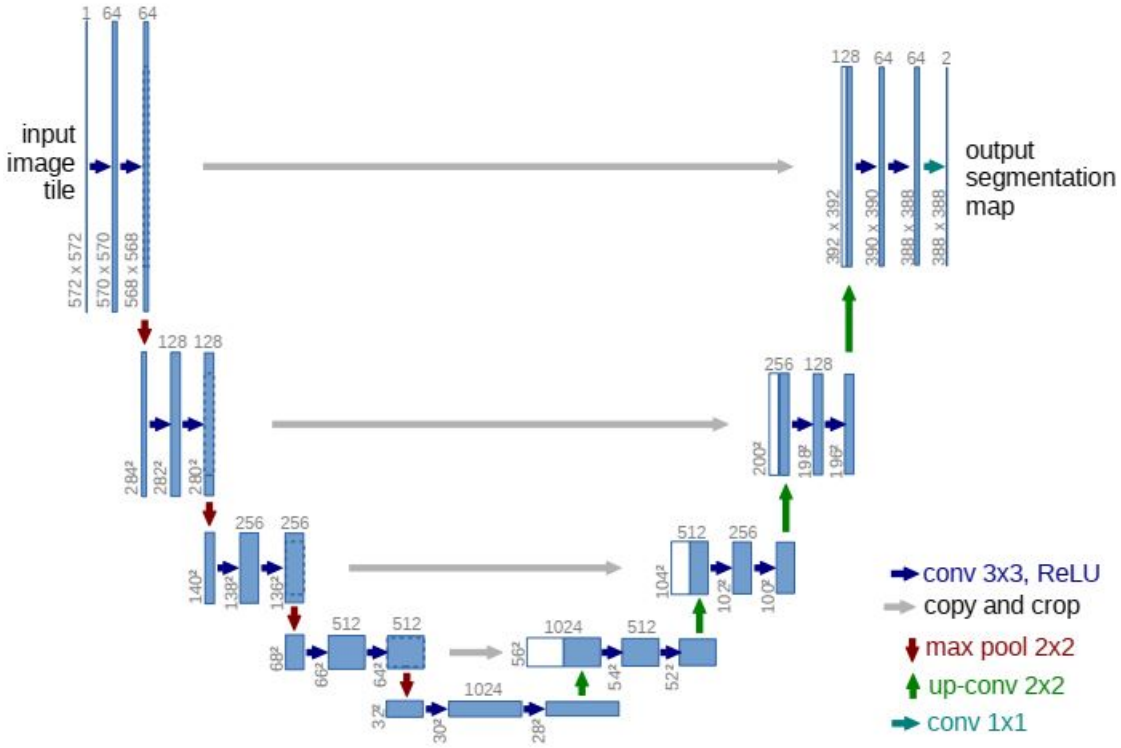

Figure 1: U-net architecture.[19]

The network architecture is illustrated in figure 1. It consists of a contraction path (left side), an expansive path (right side), and a Concatenation path (arrows between two paths).

- The contraction path follows the typical architecture of a convolutional network, it consists of the repeated application of two 3x3 convolutions (downsampling or unpadded convolutions), each followed by an activation function rectified linear unit (Relu) and a 2x2 max pooling operation with stride 2 for downsampling. At each downsampling step, we double the number of feature channels.
  During the contraction, the spatial information is reduced while feature information is increased. The contraction path follows the formula:
  $convlayer1 -> convlayer2 -> maxpooling -> dropout(optional)$

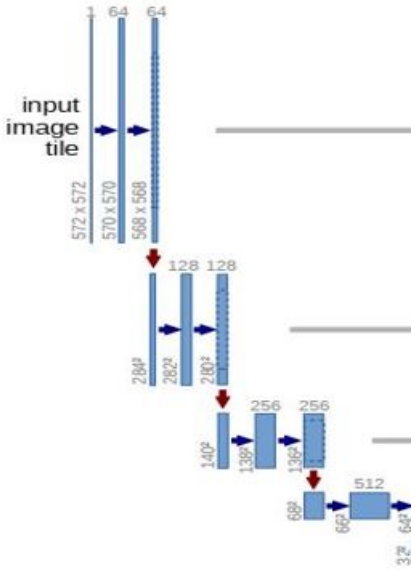

Figure 2: Contraction path

- The expansive path consists of an upsampling of the feature map followed by a 2 X 2 convolution that halves the number of feature channels. It consists of 4 decoder blocks(upsampling, conv-block) as in figure 3. The image is going to be upsized to its original size. The formula used in expansive path follows: Conv2dtranspose $->$ *concatenate* $->$ *convlayer*1 $->$ *convlayer*2

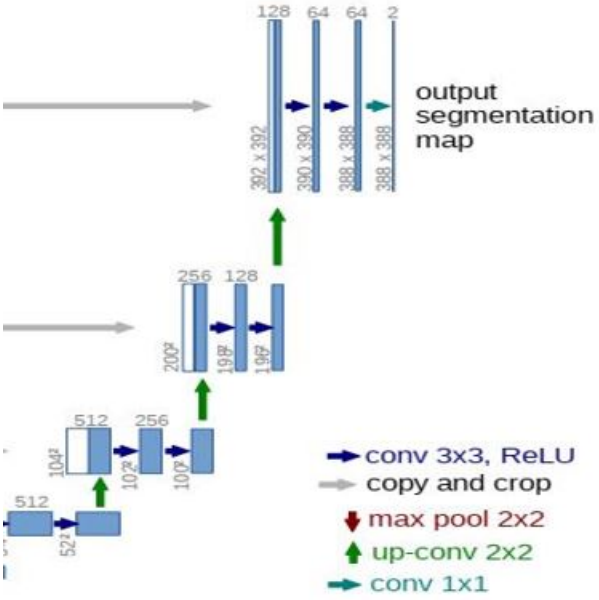

Figure 3: Expansive path

- A concatenation path with the correspondingly cropped feature map from the contracting path and two 3x3 convolutions, each followed by ReLU. The cropping is necessary due to the loss of border pixels in every convolution.[19]
- At the final layer, a 1x1 convolution is used to map each 64-component feature vector to the number of classes

3.2 DeepLabV3

DeepLabV3 is introduced for semantic segmentation [21] to acquire dense feature extraction at multiple scales using atrous convolution with up-sampled filters in conjunction with spatial pyramid clustering. DeepLabv3 is a semantic segmentation architecture that improves upon DeepLabv2 with several modifications. To handle the problem of segmenting objects at multiple scales, modules are designed which employ atrous convolution in cascade or in parallel to capture multi-scale context by adopting multiple atrous rates. Atrous convolution exists to handle the issue of the repeated combination of max-pooling and striding at consecutive layers in DCCN that reduces significantly the spatial resolution of the resulting feature maps. It offers a simple yet powerful alternative to using deconvolutional. It also allows to effectively enlarge the field of view of filters without increasing the number of parameters or the amount of computation. Atrous convolution as a shorthand for convolution with up-sampled filters. Filter up-sampling amounts to inserting holes ('trous' in French) between nonzero filter taps[22]. Furthermore, the Atrous Spatial Pyramid Pooling module from DeepLabv2 is augmented with image-level features encoding global context and further boosting performance. Apart from using Atrous Convolution, DeepLabV3 uses an improved ASPP module by including batch normalization and image-level features. It gets rid of CRF (Conditional Random Field) as used in V1 and V2 [23]. The original DeepLab architecture is utilized along with the encoder-decoder structure in order to acquire advantages of both approaches: multi-scale contextual information at multiple pooling rates and multiple effective fields-of-view from spatial pyramid pooling and sharper object boundaries to a gradual recovery of spatial information from autoencoder[22]. figure 4 will demonstrate the idea of DeepLabV3 architecture.

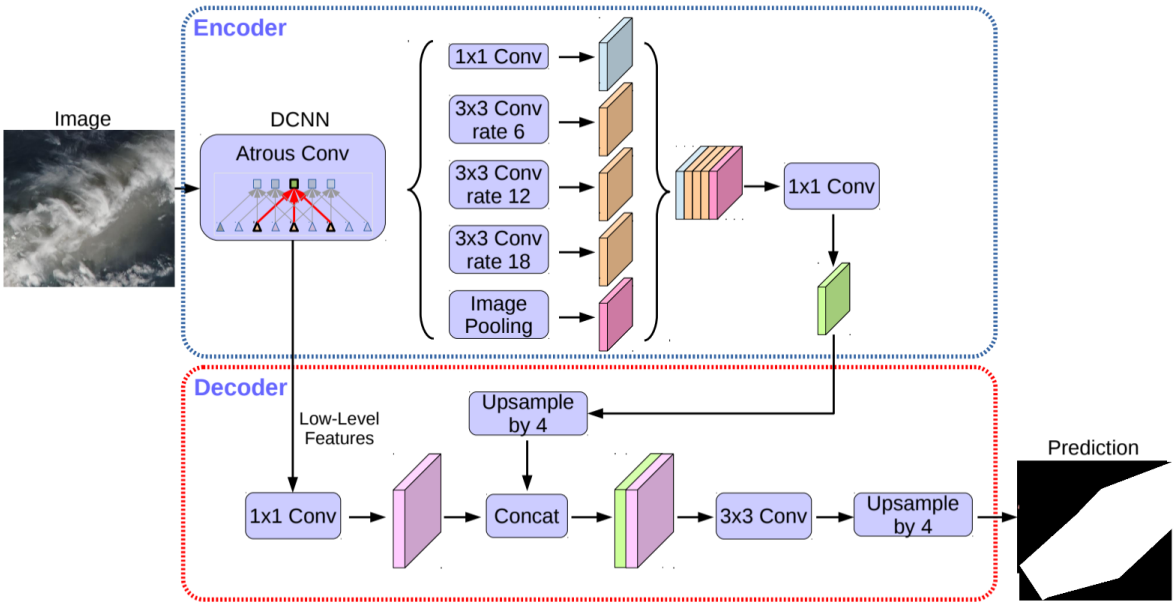

Figure 4: DeepLabV3 Architecture [22]

Deeplabv3 architecture adopts a novel encoder-decoder with atrous separable convolution. The encoder-decoder model is able to obtain sharp object boundaries. The general encoder-decoder networks have been successfully applied to many computer vision tasks, including object detection, human pose estimation, and also semantic segmentation[22].

4 Experiment Details

In this section, the dataset and the experiments to build and evaluate the models are discussed in more detail.

4.1 Dataset collection

The data used are real-color images of airborne dust in several sandstorm events in several regions of the land and ocean from 2003 to 2017. This data was collected through the MODIS dataset.

Figure 5: Sample of dataset

The Moderate Resolution Imaging Spectroradiometer (MODIS) is a key instrument aboard two NASA Earth Observing System (EOS) satellites that have been in operation since 1999. MODIS captures data in 36 spectral bands, ranging in wavelength from 0.4 to 14.4 micrometers, and produces imagery at a resolution of up to 250 meters per pixel[24].

This data includes 80 RGB images each with visually-recognizable sandstorm incidents in high latitudes over land and ocean throughout the world. The collected images are in different shapes and each image in the collected dataset is in JPEG format. All collected images are manually annotated for semantic segmentation using the CVAT tool[13], where the pixels that have a storm are marked as white and the pixels that do not have a storm are marked as black. Hence, each true-color image has a bitmap image that represents the ground-truth labels with respect to the sandstorm incident.

4.2 Experiment setup

To prevent overfitting due to a limited number of samples, the collected dataset is augmented using statistical affine and elastic transformations. All sandstorm images are also resized to (256, 256, 3) using bilinear interpolation before being inputted into the base models, which have dimensions set to 256. The dataset is split into training, validation, and test sets in 80:10:10 ratios for all base models.

4.2.1 U-Net setup

As we know, U-Net consists of two paths encoder path and decoder path with skip features between the two paths. In the encoder path, we use 4 encoder blocks each of them consisting of 2 conv2d with constant kernel size (3, 3) and activation Relu with batch-normalization after each conv layer and max-pooling2d layer to downsampling the size of the image with increasing the features by increasing number of filters in each block. We keep the version before downsampling for use as skip features in the decoder path. Then we use 2 conv2d as a bridge between two paths. In the decoder path, we use 4 decoder blocks each of them consisting of conv2d-transpose to upsampling the size of the image to return to its original size and 2 conv2d after concatenating with the skip features from the encoder path. Finally, we use conv2d as the output layer with kernel size (1, 1) and activation sigmoid because of binary segmentation.

4.2.2 DeepLabV3 setup

The model is based on the ResNet50 architecture and uses an Atrous Spatial Pyramid Pooling (ASPP) module to capture multi-scale context information. The input images are passed to a pre-trained ResNet50 model with the last two layers removed. The output of the ResNet50 model is then fed into the ASPP module to extract multi-scale features. The ASPP module consists of several parallel convolutional layers with different dilation rates to capture

different scales of context information. The output of the ASPP module is upsampled and concatenated with low-level features from the ResNet50 model. The concatenated features are then passed through two convolutional layers with batch normalization and ReLU activation functions. The output of the final convolutional layer is upsampled and passed through a final convolutional layer with a sigmoid activation function to produce a binary segmentation mask.

4.3 Evaluation metrics

To evaluate the performance of a sandstorm image segmentation algorithm, several performance measures can be used. These measures help to quantify the accuracy and efficiency of the algorithm in segmenting sandstorms from the background.

4.3.1 Accuracy

The accuracy measures the percentage of correctly classified pixels in the image. It is calculated as the ratio of the number of correctly classified pixels to the total number of pixels in the image. The accuracy is a useful performance measure for sandstorm image segmentation because it provides a simple and intuitive measure of the algorithm's ability to accurately classify each pixel in the image[25].

4.3.2 MIOU

Mean Intersection over Union (MIOU) is a commonly used performance measure for evaluating the accuracy of image segmentation models, including sandstorm segmentation models. The MIOU measures the average overlap between the segmented sand particles and the ground truth across all classes in the image. It is calculated as the mean of the IOU for each class in the image. The MIOU is a useful performance measure for sandstorm image segmentation because it takes into account the overlap between different classes in the image, which can be important in accurately segmenting sand particles from other objects in the scene[26].

5 Result and Discussion

In this study, we aimed to evaluate the impact of data augmentation, dataset dependency, and model architecture on the performance of image segmentation models.

5.1 Comparison between real dataset and augmented dataset

We conducted experiments using two datasets, one real and one artificially augmented and compared the performance of two state-of-the-art models, U-Net and DeepLabV3.

Dataset	Performance	U-Net	DeepLapV3
Real	Accuracy	77%	81%
	MIOU	72%	76%
Augmented	Accuracy	91%	94%
	MIOU	85%	89%

Table 1: Impact of data augmentation

Table 1 shows that both U-Net and DeepLabV3 achieved high accuracy on the real dataset, with DeepLabV3 performing slightly better than U-Net. On the augmented dataset, both models achieved even higher accuracy, with DeepLabV3 outperforming U-Net. This suggests that data augmentation can significantly improve the performance of image segmentation models.

5.2 Comparison between dependent dataset and independent dataset

We compared the performance of the models on the dependent dataset and the independent dataset. The dependent dataset was generated by randomly splitting the original dataset into training and validation sets, while the independent dataset was obtained from a different source.

Dataset	Performance	U-Net	DeepLapV3
Dependent	Accuracy	91%	94%
	MIOU	85%	89%
Independent	Accuracy	91%	93%
	MIOU	84%	88%

Table 2: Impact of data dependency

Table 2 shows that both U-Net and DeepLabV3 achieved high accuracy on both the dependent and independent datasets, with DeepLabV3 performing slightly better than U-Net on both datasets. In terms of MIOU, both models achieved high scores on the dependent dataset, with DeepLabV3 outperforming U-Net. However, on the independent dataset, the performance of both models dropped slightly, with U-Net performing slightly worse than DeepLabV3.

5.3 Comparison between U-Net and DeepLabV3

We compared the performance of U-Net and DeepLabV3 on same dataset and hyper parameters.

Models	Accuracy	MIOU
U-Net	91%	85%
DeepLabV3	94%	88%

Table 3: Impact of model architecture

This table shows that DeepLabV3 achieved higher accuracy than U-Net on the same dataset and hyper parameters, with a difference of 3 percentage points. This suggests that DeepLabV3 may be a more effective model for image segmentation tasks that require high-precision segmentation.

5.4 Visualization

5.4.1 U-Net Output

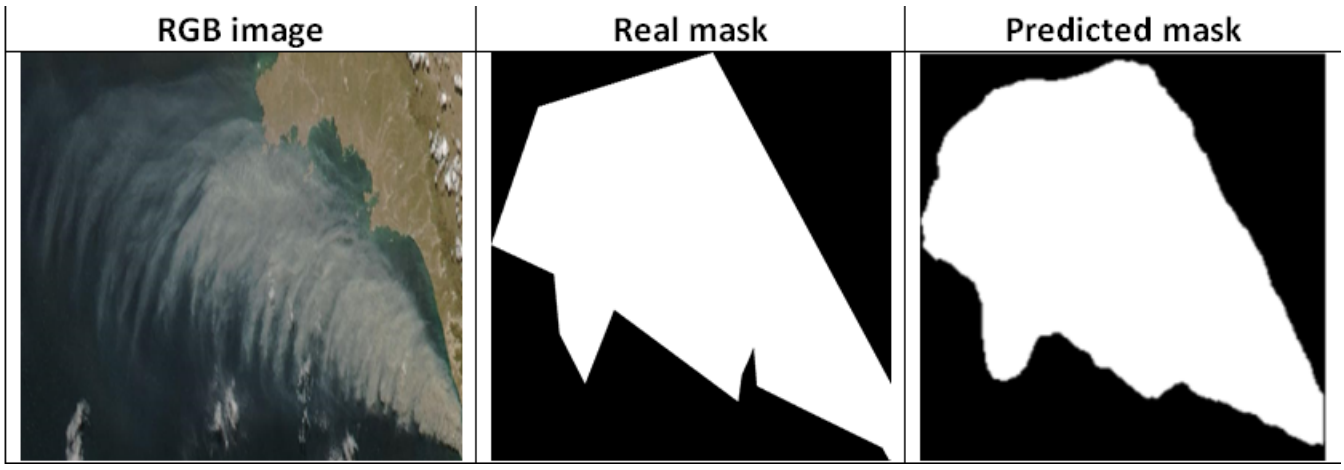

Figure 6: U-Net output

.

5.4.2 DeepLabV3 Output

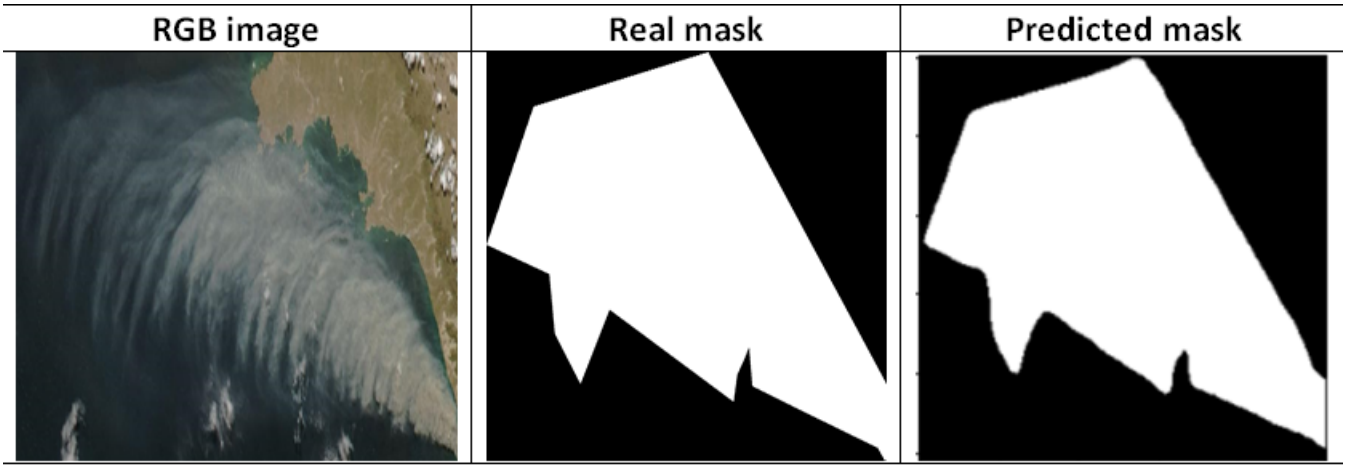

Figure 7: DeepLabV3 output

6 Conclusion

In this paper, deep learning approaches are presented to semantically segment and detect sandstorms from the satellite true-color images. The used models are U-Net and DeepLabv3. Both U-Net and DeepLabV3 were found to perform well in detecting sandstorm areas in the images, with DeepLabV3 showing slightly better results in terms of accuracy and speed. However, it is important to note that the performance of these models can vary depending on the specific characteristics of the images and the parameters used for training.

References

[1] R. Indoitu, L. Orlovsky, and N. Orlovsky. Dust storms in central asia: Spatial and temporal variations. *Journal of Arid Environments*, 85:62–70, 2012.

[2] Sina Muster, Moritz Langer, Anna Abnizova, Kathy L Young, and Julia Boike. Spatio-temporal sensitivity of modis land surface temperature anomalies indicates high potential for large-scale land cover change detection in arctic permafrost landscapes. *Remote sensing of environment*, 168:1–12, 2015.

[3] Sherwood B Idso. Dust storms. *Scientific American*, 235(4):108–115, 1976.

[4] Bernard Gomez. Degradation of vegetation and agricultural productivity due to natural disasters and land use strategies to mitigate their impacts on agriculture, rangelands and forestry. *Natural Disasters and Extreme Events in Agriculture: Impacts and Mitigation*, pages 259–276, 2005.

[5] George Papandreou, Liang-Chieh Chen, Kevin P Murphy, and Alan L Yuille. Weakly-and semi-supervised learning of a deep convolutional network for semantic image segmentation. In *Proceedings of the IEEE international conference on computer vision*, pages 1742–1750, 2015.

[6] Yu Qu, Lin Sun, Nan Ma, et al. Prior land surface reflectance-based sandstorm detection from space using deep learning. *Frontiers in Earth Science*, 10:1–19, 2022.

[7] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 801–818, 2018.

[8] ZC Men, J Jiang, Xian Guo, LJ Chen, and DS Liu. Airport runway semantic segmentation based on dcnn in high spatial resolution remote sensing images. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 42:361–366, 2020.

[9] Jason Kugelman, Joseph Allman, Scott A Read, Stephen J Vincent, Janelle Tong, Michael Kalloniatis, Fred K Chen, Michael J Collins, and David Alonso-Caneiro. A comparison of deep learning u-net architectures for posterior segment oct retinal layer segmentation. *Scientific Reports*, 12(1):14888, 2022.

[10] Badr Alharbi, Mohammed Mujtaba Shareef, and Tahir Husain. Study of chemical characteristics of particulate matter concentrations in riyadh, saudi arabia. *Atmospheric Pollution Research*, 6(1):88–98, 2015.

[11] Jingmin Gao, Dongxu Chen, and Maimaitijiang Saitiniyazi. Vulnerability of oasis agriculture to sandstorm disasters in xinjiang, china. *Bangladesh Journal of Botany*, 49(4):981–988, 2020.

[12] Zhuoqun Zhang, Jiashu Liu, Kangjie Shao, and Peng Zhang. Analysis of wind-sand-load-induced dynamic response of transmission tower-line systems. *Shock and Vibration*, 2022, 2022.

[13] NS Bandara. Ensemble deep learning for automated dust storm detection using satellite images. In *2022 International Research Conference on Smart Computing and Systems Engineering (SCSE)*, volume 5, pages 178–183. IEEE, 2022.

[14] Haibo Huang, Haobo Chen, Haohao Xu, Ying Chen, Qihui Yu, Yehua Cai, and Qi Zhang. Cross-tissue/organ transfer learning for the segmentation of ultrasound images using deep residual u-net. *Journal of Medical and Biological Engineering*, 41:137–145, 2021.

[15] Reza Azad, Maryam Asadi-Aghbolaghi, Mahmood Fathy, and Sergio Escalera. Bi-directional convlstm u-net with densley connected convolutions. In *Proceedings of the IEEE/CVF international conference on computer vision workshops*, pages 0–0, 2019.

[16] Zhanpeng Zhang and Kaipeng Zhang. Farsee-net: Real-time semantic segmentation by efficient multi-scale context aggregation and feature space super-resolution. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 8411–8417. IEEE, 2020.

[17] Bhakti Baheti, Shubham Innani, Suhas Gajre, and Sanjay Talbar. Semantic scene segmentation in unstructured environment with modified deeplabv3+. *Pattern Recognition Letters*, 138:223–229, 2020.

[18] Eduardo Assunção, Pedro D Gaspar, Ricardo Mesquita, Maria P Simões, Khadijeh Alibabaei, André Veiros, and Hugo Proença. Real-time weed control application using a jetson nano edge device and a spray mechanism. *Remote Sensing*, 14(17):4217, 2022.

[19] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015.

[20] Haiying Wang and Fang Miao. Building extraction from remote sensing images using deep residual u-net. *European Journal of Remote Sensing*, 55(1):71–85, 2022.

[21] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848, 2017.

[22] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 801–818, 2018.

[23] Thorsten Hoeser and Claudia Kuenzer. Object detection and image segmentation with deep learning on earth observation data: A review-part i: Evolution and recent trends. *Remote Sensing*, 12(10):1667, 2020.

[24] Xiaoxiong Xiong, Junqiang Sun, William Barnes, Vincent Salomonson, Joseph Esposito, Hector Erives, and Bruce Guenther. Multiyear on-orbit calibration and performance of terra modis reflective solar bands. *IEEE Transactions on Geoscience and Remote Sensing*, 45(4):879–889, 2007.

[25] Bert Guindon and Ying Zhang. Application of the dice coefficient to accuracy assessment of object-based image classification. *Canadian Journal of Remote Sensing*, 43(1):48–61, 2017.

[26] Zhufeng Pan, Jian Yang, Xing-er Wang, Feiliang Wang, Iftikhar Azim, and Chenyu Wang. Image-based surface scratch detection on architectural glass panels using deep learning approach. *Construction and Building Materials*, 282:122717, 2021.