SWINySeg:
Singapore Whole sky Nychthemeron
Image SEGmentation Database

April 16, 2019

Contact:
Stefan Winkler
winkler@comp.nus.edu.sg

1 Introduction

The SWINySeg database is introduced and used in the following paper:

S. Dev, A. Nautiyal, Y. H. Lee, and S. Winkler: "CloudSegNet: A Deep Network for Nychthemeron Cloud Image Segmentation" *IEEE Geoscience and Remote Sensing Letters*, 2019.

If you use or adapt any part of this dataset, please cite the following paper:

- S. Dev, A. Nautiyal, Y. H. Lee, and S. Winkler: "CloudSegNet: A Deep Network for Nychthemeron Cloud Image Segmentation" *IEEE Geoscience and Remote Sensing Letters*, 2019.

2 Description

The SWINySeg dataset contains 6768 images of sky/cloud patches along with their corresponding binary ground truth maps, which were generated in consultation with cloud experts. Representative sample images from the SWINySeg database are shown below in Fig. 1.
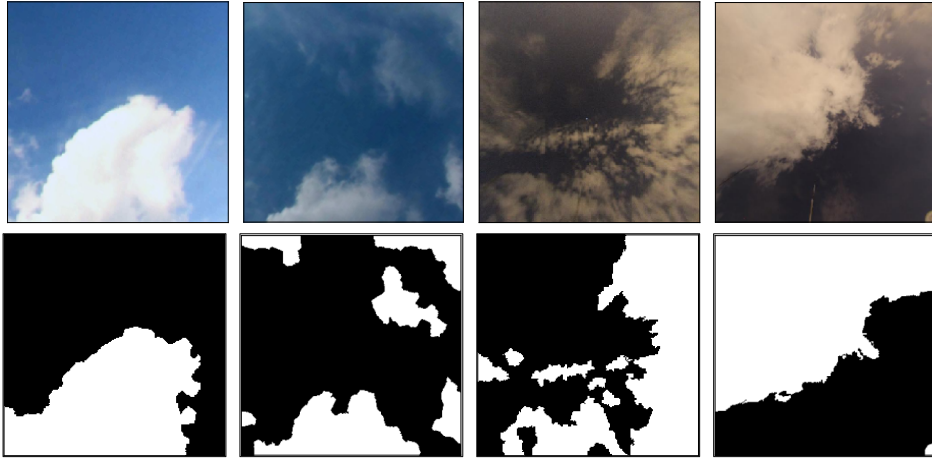

Figure 1: Sample images from the SWINySeg database (top row) along with corresponding sky/cloud segmentation ground truth (bottom row).

The images from SWINySeg dataset [1] are from two publicly available sky/cloud image segmentation datasets – SWIMSEG (Singapore Whole Sky IMaging SEGmentation dataset) [2] and SWINSEG (Singapore Whole sky Nighttime Image SEGmentation Database) [3] – to create a composite dataset of nychthemeron (day-and-night) images. The SWIMSEG dataset consists of 1013 daytime sky/cloud images, along with manually annotated ground-truth maps. The SWINSEG dataset consists of 115 diverse nighttime sky/cloud images, along with the corresponding ground-truth maps. All images were captured using WAHRSIS, a calibrated ground-based whole sky imager, located at Nanyang Technological University Singapore at the location $(1.34N, 103.68E)$. The

corresponding binary ground truth maps were annotated after consultation with experts from Singapore Meteorological Services.

The images from the SWIMSEG and SWINSEG databases were undistorted using a process described in [2, 3]. All daytime- and nighttime- images were resized to the dimension of 300×300 in the SWINySeg dataset.

In order to increase the number of images in the composite dataset of nychthemeron images, we perform image augmentation techniques for both SWIMSEG- and SWINSEG-datasets. We combine a number of image-based operations with different degrees of magnitude. We include several such operations, such as flipping, rotation, shifting, zooming, and shearing. The total number of augmented daytime- and nighttime-images we obtain are 5065 and 575, respectively. This results in a total of 6768 nychthemeron sky/cloud images in the SWINySeg dataset.

3 Database Content

The database consists of two folders, "images" and "GTmaps", and a comma delimited file 'metadata.csv'. The "images" folder contains all the image patches, while the folder "GTmaps" contains the corresponding binary ground truth maps.

The original sky/cloud image files are named <ImageType><ImageNumber>.jpg, and the corresponding ground truth maps are named <ImageType><ImageNumber>.png. The augmented versions are appended with _<AugmentationNumber> in the file name.

For example, `./images/d0003.jpg` and `./GTmaps/d0003.png` are a daytime image and its corresponding binary ground-truth map. `./images/d0003_1.jpg` is an augmented image obtained from this daytime image, and `./GTmaps/d0003_1.png` is its corresponding binary ground-truth map.

The `metadata.csv` file contains all the related metadata information pertaining to the individual input image, as follows.

1. **Name**: Name of the nychthemeron sky/cloud image file.

2. **Date**: Capture date of the image. It is represented in YYYYMMDD format, where YYYY, MM, and DD represent the year, month, and day respectively.

3. **Time**: Capture time of the image. It is represented in 24-hour hhmmss format, where hh, mm, and ss represent the hour, minute, and second respectively.

4. **Fnumber**: F-number of the camera.

5. **ExposureTime**: Exposure time (in seconds) of the camera.

6. **ISO**: ISO setting of the camera.

4 Licensing Information

The dataset is released under a Creative Commons license (https://creativecommons.org/licenses/by-nc/4.0/). You are free to:

5 Additional information

- license.html: This is the licensing file that we recommend you read before using or sharing this dataset.

- README.pdf: This file.

References

[1] S. Dev, A. Nautiyal, Y. H. Lee, and S. Winkler, "CloudSegNet: A deep network for nychthemeron cloud image segmentation," *IEEE Geoscience and Remote Sensing Letters*, 2019.

[2] S. Dev, Y. H. Lee, and S. Winkler, "Color-based segmentation of sky/cloud images from ground-based cameras," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 1, pp. 231–242, 2017.

[3] S. Dev, F. M. Savoy, Y. H. Lee, and S. Winkler, "Nighttime sky/cloud image segmentation," in *Proc. International Conference on Image Processing (ICIP)*, 2017.