Aerial Image Similarity Estimation Using Cloud Removal Methods

Dominykas Kaminskas, Vytautas Valaitis

Vilnius University

Map-based aerial vehicle localization: images taken by on-board cameras during flight are compared to an aerial map to find similarities between them.
However, the accuracy of a map-based approach decreases during cloudy weather conditions. Removing clouds using neural networks and generative image inpainting can increase the amount of information found in aerial images.

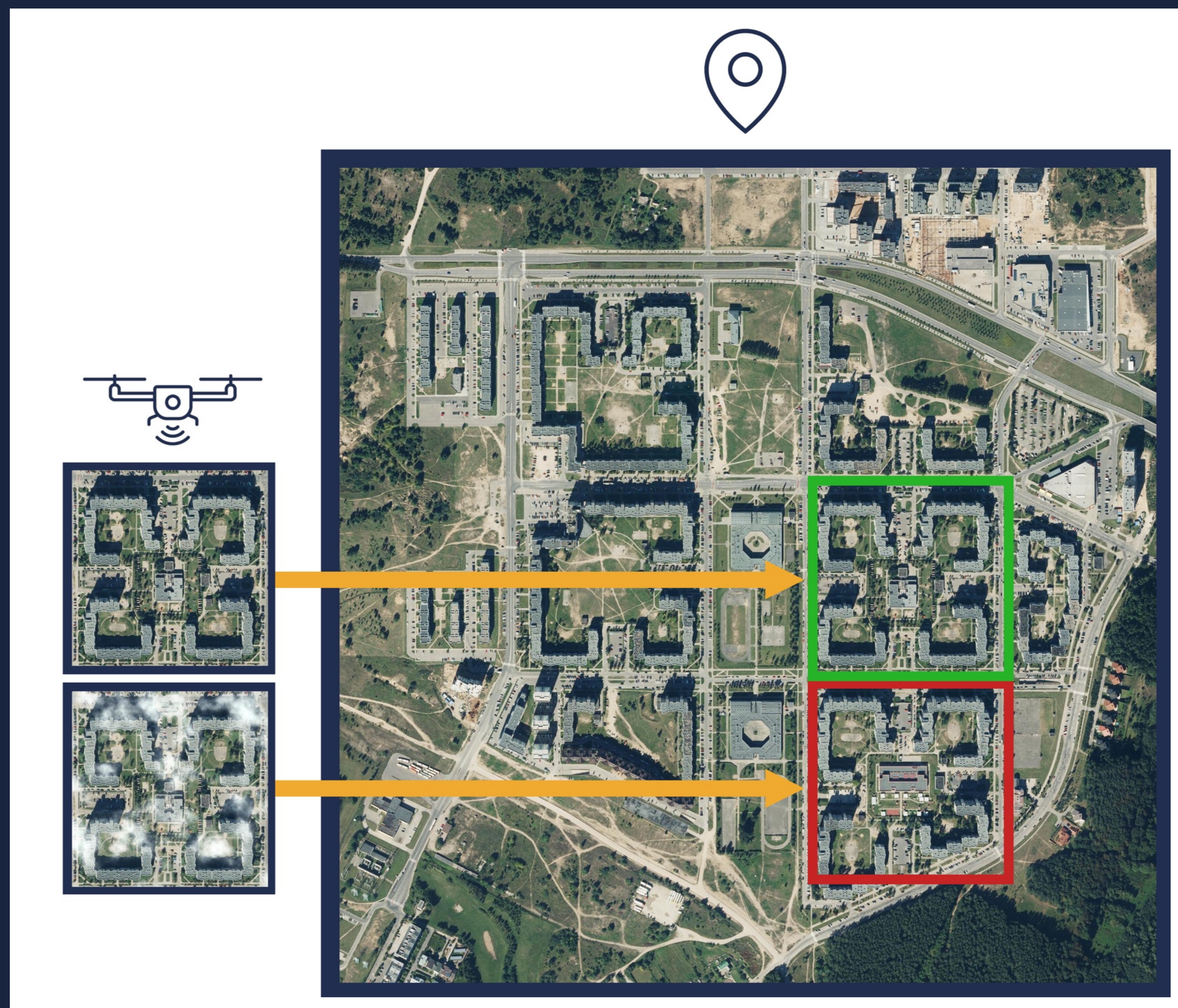

Figure 1. The same image with generated clouds is misidentified by the neural network

Research stages:

1. Collect dataset

- Part of Vilnius aerial map (25 km^2) was cut into smaller images (224px x 224px) with overlap.
- Images were grouped into triplets for training: anchor images were taken from a 2013 map, while positive and negative images were taken from a 2016 map (Fig 3).
- Dataset contained more than 20,000 images.

2. Generate clouds

- Since it is difficult to find a pair of identical images with different weather conditions it was decided to generate clouds
- SatteliteCloudGenerator[2] tool was used to create realistic clouds.
- It is based on structural noise and has various parameters such as colour, density and transparency for different cloud formations.
- Given an original image it returns a cloudy image, a cloud mask and a shadow mask as a result.

3. Remove clouds

- CloudGAN model was used to remove clouds from images.
- It is a modified version of DeepFill network[3] that was additionally trained on datasets intended for cloud detection and cloud removal tasks.
- Cloud mask has to be supplied for a cloud removal method to work.
- Models fail to provide realistic results when > 40% of the input image is covered by clouds.

4. Calculate accuracy

- Precision@K recommendation system was implemented to determine the accuracy of each network: 1000 random points were chosen on a map and cut out to serve as input images; for each input image the model had to return 5 most similar images from a dataset; an image was considered accurate if the shift between input and output images was less than 15m.
- The same experiment was repeated 3 times with different datasets: original, cloudy and cloud-removed (Tab 1).

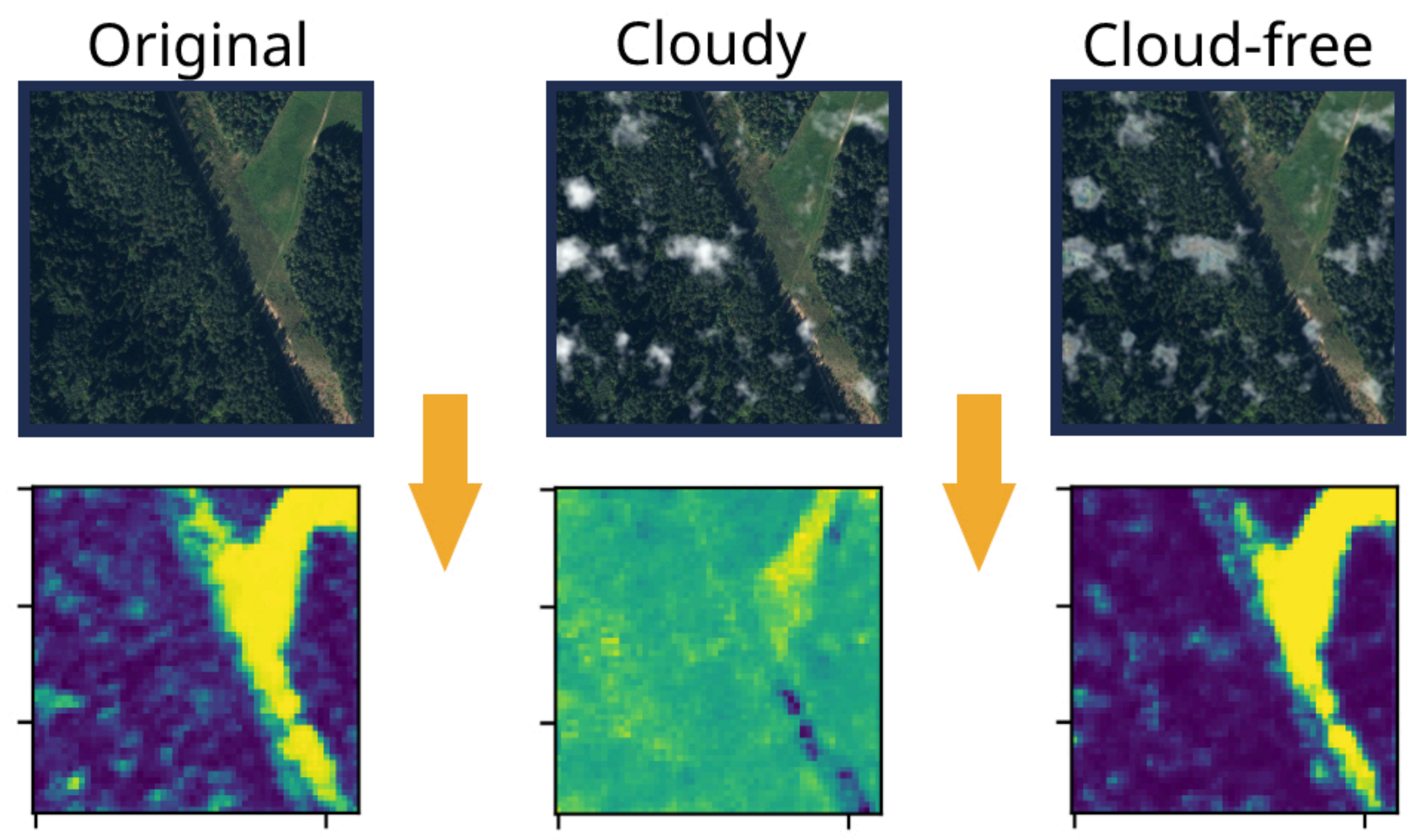

Figure 2. EfficientNet-B2 (59) (0) network extracted feature maps (bottom) for given aerial images (top). The feature map of a cloud-free image closely resembles the feature map of an original image

$$loss = \sum_{i=1}^{N} [\|f(a_i) - f(p_i)\|_2^2 - \|f(a_i) - f(n_i)\|_2^2 + \alpha]$$

Figure 3. Triplet loss function used additionally train each network with a custom aerial image dataset. Here a_i represents anchor image, p_i - positive image, p_i - negative image, α - margin value

Table 1. Accuracy of different convolutional neural network configurations

Network	Removed layers	Frozen layers	Original	Cloudy	Cloud-free
MobileNet	**35**	**0**	97.8%	87.7%	**94.2%**
	54	**0**	99.2%	96.5%	**97%**
EfficientNet-V2-B0	30	30	95%	84.8%	83.2%
	71	30	97.5%	94.5%	83.6%
VGG-16	10	10	99.2%	88.6%	87.2%
	14	**14**	97.6%	90.5%	**90.8%**
ResNet-50	50	38	99.7%	92.1%	85.8%
	80	38	93.7%	92.9%	90.6%
EfficientNet-B2	**59**	**0**	88.8%	67.2%	**93.6%**
	68	**0**	84.8%	80.4%	**87.8%**

References

[1] V. Valaitis, V. Marcinkevičius, R. Jurevičius. *Learning Aerial Image Similarity Using Triplet Networks*. 2020-02, p. 195–207. ISBN 978-3-030-40615-8.
[2] Czerkawski, M.; Atkinson, R.; Michie, C.; Tachtatzis, C. *SatelliteCloudGenerator: Controllable Cloud and Shadow Synthesis for Multi-Spectral Optical Satellite Images*. Remote Sensing. 2023, 15, 4138.
[3] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, T. S. Huang. *Free-form image inpainting with gated convolution*. Proceedings of the IEEE/CVF international conference on computer vision. 2019, p. 4471–4480.

Contacts

dominykas.kaminskas@mif.stud.vu.lt
vytautas.valaitis@mif.vu.lt