FSDL Report on Aerial Segmentation

Fabian Berressem Jonathan Schaust

May 14, 2021

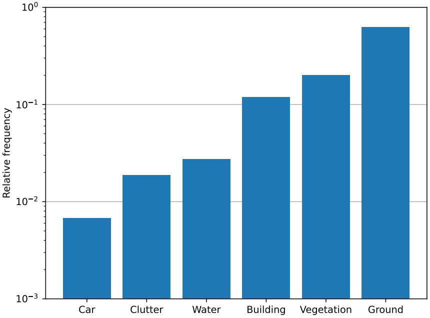
Project Description

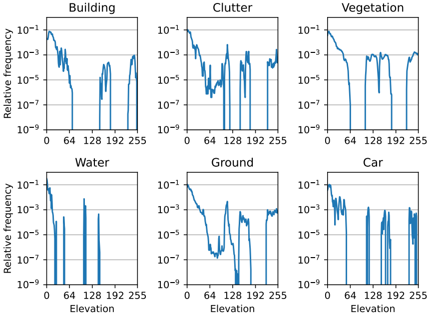
- Task of this project: Segmentation of aerial orthomosaics into six different classes (building, clutter, vegetation, water, ground, car)
- Given data: 55 aerial photographs with corresponding segmentation and elevation maps (total of 9.7 GB data)
- Example:

Object class	Color
Building	Red
Clutter	Purple
Vegetation	Green
Water	Orange
Ground	White
Car	Blue
Ignore	Magenta

Data Exploration

- Found strong class imbalance (two orders of magnitude)
  $\Rightarrow$ Possibly have to account for this in loss function
- Distribution of elevations very similar
  $\Rightarrow$ Hints at poor quality of labeling, limited information in elevation maps



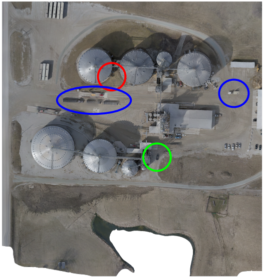


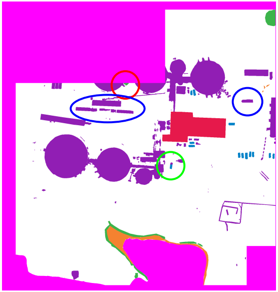
Approaches

- Limited model search to fully convolutional models to allow for different input sizes
- Implemented additional loss functions such as F_1-score, Focal Tversky Loss because of class imbalance
- Joint learning of segmentation mask and elevation map
- Implemented learning on adversarial examples using Fast Gradient Signed Method

Encountered Problems

- Moving objects are “sliced” (rather insignificant for prediction)
- Multiple occurrences of wrong or inconsistent labels
- Some objects larger than receptive field, context can not propagate properly
- Images very large, have to be cut into subimages for inference
- Sometimes blank spots in the pictures, have to be padded
  $\Rightarrow$ Results in “artificial” edges





Solution Strategies

- Use a smoothed strided inference to avoid artifacts at boundaries due to cutting images into subimages
  $\Rightarrow$ Predictions on subimages overlap such that boundary artifacts can be accounted for
- Use cascading inference to allow for propagation of long-distance context (such as large buildings)
  $\Rightarrow$ Inference is done successively at different resolutions of the image: first predict on whole image downsampled to input-size of network, then predict on whole image downsampled with increased resolution, merge predictions, *etc.*
  $\Rightarrow$ Complexity scales as $N^{CI}(\epsilon) \leq N\frac{\epsilon^2}{\epsilon^2 - 1}$ with increase of resolution ϵ and complexity of (strided) prediction N

Evaluation

- Used F_1-score on validation set to choose model
- Performances varied only slightly
  $\Rightarrow$ Performance is limited more by the task itself rather than model capabilities
- Chose MobileNetV2 pretrained on ImageNet as encoder with a cross-entropy loss and $\alpha_{elev} = 10$
- Model was trained on subimages of size 300×300 while inference was done with subimages of size 4000×4000 with stride 7 with smoothing and cascading inference with $\epsilon = 2$
- Scored $\langle F_1 \rangle_{test} = 0.8496$, which is up 1.6% from the baseline-value of $\langle F_1^{base} \rangle_{test} = 0.8361$

Conclusion

- Were able to improve on the task of aerial segmentation compared to the baseline
- Implemented multiple including additional loss functions, different inferences, adversarial training, *etc*.
- Found severe problems with dataset
  $\Rightarrow$ To improve on task significantly, quality of dataset <u>has</u> to be improved