

Presentations “Machine learning”

Active Learning to Assist Annotation of Aerial Images in Environmental Surveys

Mathieu Laroze, mathieu.laroze@irisa.fr, University of Rennes 1 September 28, 2018

Nowadays, remote sensing technologies greatly ease environmental assessment over large study areas using aerial images, e.g. for monitoring human environmental pressure. Such data are most often analyzed by a manual operator, leading to costly and non scalable solutions for large area.

Object detection algorithms are used to fasten and automate the counting processes. In the fields of both machine learning and image processing, many algorithms have been developed to tackle this complex task. We propose a method to assist the annotation process in aerial images by introducing an active learning process. This method is evaluated on a real case of monitoring seashell gatherer pressure on the seashore of Morbihan (France)

Adapting a Deep Learning-based Animal Detector to Unseen Data

Benjamin Kellenberger

We study the feasibility of transferring a deep convolutional neural network designed to detect animals in drone-based imagery to a new, unseen dataset. This process is known as *domain adaptation* and is particularly challenging, since the underlying dataset distributions vary both in image space (e.g., images may be acquired under different weather conditions) and label space (the ratio of animals to background might be different in the new dataset). We investigate distribution-matching methods and show how to combine it with deep learning approaches.

BUILDING CLASSIFICATION OF VHR AIRBORNE STEREO IMAGES USING FULLY CONVOLUTIONAL NETWORKS AND FREE TRAINING SAMPLES

Name : (Equal Contributors) **Chen Yizi**, Weixiao Gao, Elyta Widyaningrum, Mingxue Zheng, Kaixuan Zhou

E-mail : Y.Chen-35@student.tudelft.nl, W.Gao-1@tudelft.nl, E.Widyaningrum@tudelft.nl, M.Zheng-1@tudelft.nl, K-Zhou-1@tudelft.nl

Semantic segmentation, especially for buildings, from the very high resolution (VHR) airborne images is an important task in urban mapping applications. Nowadays, the deep learning has significantly improved and applied in computer vision applications. Fully Convolutional Networks (FCN) is one of the tops voted method due to their good performance and high computational efficiency. However, the state-of-art results of deep nets depend on the training on large-scale benchmark datasets. Unfortunately, the benchmarks of VHR images are limited and have less generalization capability to another area of interest. As existing high precision base maps are easily available and objects are not changed dramatically in an urban area, the map information can be used to label images for training samples. Apart from object changes between maps and images due to time differences, the maps often cannot perfectly match with images. In this study, the main mislabeling sources are considered and addressed by utilizing stereo images, such as relief displacement, different representation between the base map and the image, and occlusion areas in the image. These free training samples are then fed to a pre-trained FCN. To find the better result, we applied

fine-tuning with different learning rates and freezing different layers. We further improved the results by introducing atrous convolution. By using free training samples, we achieve a promising building classification with 85.6% overall accuracy and 83.77% F1 score, while the result from ISPRS benchmark by using manual labels has 92.02% overall accuracy and 84.06% F1 score, due to the building complexities in our study area.

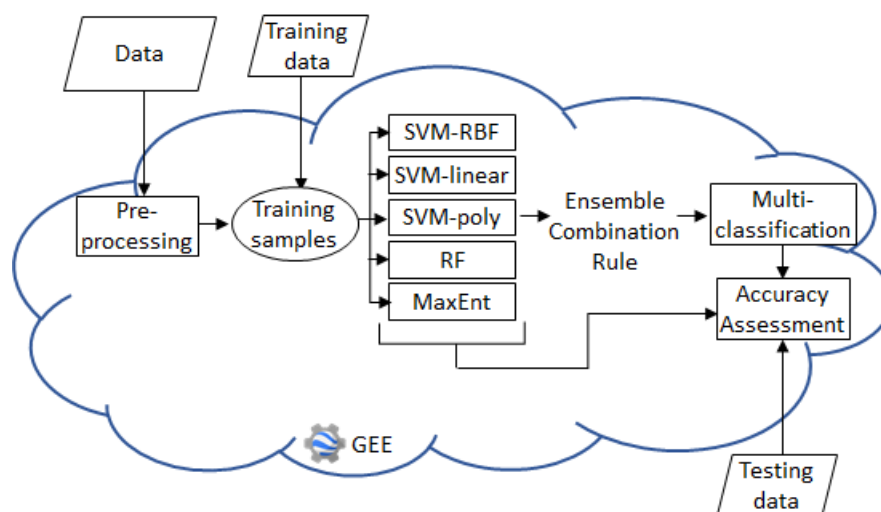
Cloud-Based Multi-Temporal Ensemble Classifier to Map Smallholder Farming Systems

Name: **Rosa Aguilar**

Affiliation: University of Twente

E-mail: r.m.aguilardearchila@utwente.nl

Abstract (~100 words and optionally 1-2 figures): Our ensemble combines a selection of spatial and spectral features derived from multi-spectral Worldview-2 images, field data, and five machine learning classifiers to produce a map of the most dominant crops in our study area. Different ensemble sizes were evaluated using two combination rules, namely majority voting and weighted majority voting. Both strategies outperform any of the tested single classifiers. This means an overall accuracy improvement of up to 4.65%. Our results demonstrate the potential of ensemble classifiers to map crops grown by West African smallholders. The use of ensembles demands high computational capability, but the increasing availability of cloud computing solutions allows their efficient implementation and may accommodate the data processing needs of local organizations.



Comparison of machine learning algorithms for large-scale land cover fraction Estimation

Name: **Dainius Masiliūnas**, Nandin-Erdene Tsendbazar, Martin Herold, Jan Verbesselt

Affiliation: Wageningen University & Research

E-mail: dainius.masiliunas@wur.nl

Most current global land cover maps assign a single class for each mapped pixel and disregard mixed pixels. In contrast, by assigning fractions of each land cover class to each pixel, land cover fraction mapping is more precise. However, its studies have been limited to a small

scale, with few classes and algorithms compared. We compared the classification accuracy for five machine learning algorithms, trained and tested on over 26000 sample points throughout Africa, with ten land cover classes. Results showed that random forest regression achieved the highest classification accuracy. The results are a milestone for creating a global land cover fraction map with a user-definable flexible legend.

Crop Classification of Worldview-2 Time Series using Support Vector Machine (SVM) and Random Forest (RF)

Name: **Azar Zafari** (Co-authors: Raul Zurita-Milla and Emma Izquierdo-Verdiguier)

Affiliation: Faculty of Geo-Information Science and Earth Observation (ITC). University of Twente, Enschede, The Netherlands

E-mail: a.zafari@utwente.nl

Land cover mapping using high dimensional data is a common task in remote sensing. Random Forest (RF) and Support Vector Machine (SVM) are often reported in the literature as efficient classifiers for land cover mapping, particularly, in dealing with high-dimensional data. In this research, the possibility of crop classification on time series of Worldview2 images is evaluated in an integrated approach using two most acknowledged supervised learner including random forest (RF) and support vector machine (SVM).

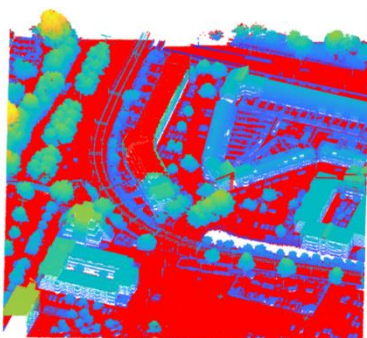
Deep learning approaches for ground classification of 3D point clouds

Name: **Mario Soilán**

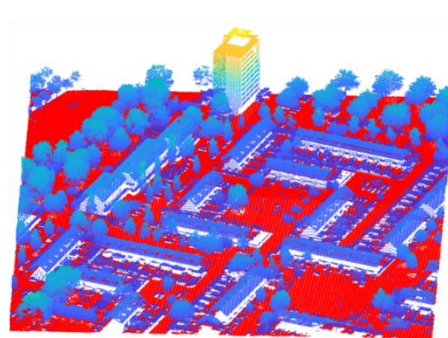
Affiliation: Department of Geoscience and Remote Sensing, Faculty of Civil Engineering and Geosciences, TU Delft / Department of Materials Engineering, Applied Mechanics and Construction, School of Industrial Engineering, University of Vigo, 36310, Spain

E-mail: msoilan@uvigo.es

This work presents two different methods for ground classification using 3D point cloud data from the *Actueel Hoogtebestand Nederland* dataset, both of them based in different deep learning approaches, thus minimizing the heuristic-based processing of the 3D data. First, a shallow version of the SegNet architecture is trained using 2D images that are generated from the 3D point cloud. Second, the PointNet model, which had been initially designed for indoor data, is employed for ground classification using directly the 3D data as input to the model. The classification results on a test set of around 25 million points show a remarkable performance of both methods, with F-score metrics above 96%.



s-SegNet



PointNet

Deep Learning Models to Count Buildings in High-Resolution Overhead Images

Name: **Sylvain Lobry**, Devis Tuia

Affiliation: Wageningen University, The Netherlands

E-mail: sylvain.lobry@wur.nl

This talk addresses the problem of automatically counting buildings in high-resolution overhead imagery. We formulate the problem as a regression task and study the relevance of deep-learning based methods. Two convolutional neural networks architectures are proposed. We show that a model enforcing equivariance to rotations is beneficial for the task of counting in remotely sensed images. We also compare two loss functions for the training of the model. From this, we recommend guidelines for the choice of an appropriate loss function depending on the reliability of the ground truth and the targeted application.

Rotation and scale equivariant neural networks for earth observation

Name: **Diego Marcos** (PhD, supervised by Devis Tuia)

Affiliation: Wageningen University

Email: diego.marcos@wur.nl

Deep learning, and in particular Convolutional Neural Networks (CNN), have been a game changer for the automatic understanding of earth observation imagery. Given sufficient training examples, deep CNNs can obtain substantially better performance than competing methods at a variety of tasks, including land-cover and land-use classification. On the other hand, CNNs often require large models that are notoriously difficult to design and train. We propose to reduce the size of CNN models by internally encoding the knowledge that the same object at different orientations and different scales should be assigned the same output value. We show how this results in better and lighter models for land-cover and land-use classification using very high resolution imagery.

Semantic façade segmentation from airborne oblique images

Name: **Yaping Lin**, Francesco Nex, Michael Yang

Affiliation: University of Twente, Faculty ITC, Department of Earth Observation Science (EOS)

E-mail: y.lin-1@student.utwente.nl

Current photogrammetric techniques and airborne oblique camera systems allow the generation of high-resolution 2D and 3D data in urban areas. Traditionally, façade segmentation has been performed from terrestrial view. In our work, high-resolution images from aerial views are used to address the problem in urban areas. Random forests are compared with state-of-art FCNs. Random forests use hand-craft features derived from images and point clouds. In contrast, FCNs learn features from RGB bands and the third components of normal vectors. In both cases, 3D features are projected back into the image space to support the façade interpretation. Fully connected conditional random field (CRF) is taken as a post-processing of the FCN to refine the segmentation results. Results show that the models embedding the 3D component outperform the solution using only images. FCNs significantly outperformed random forests, especially for the balcony delineation

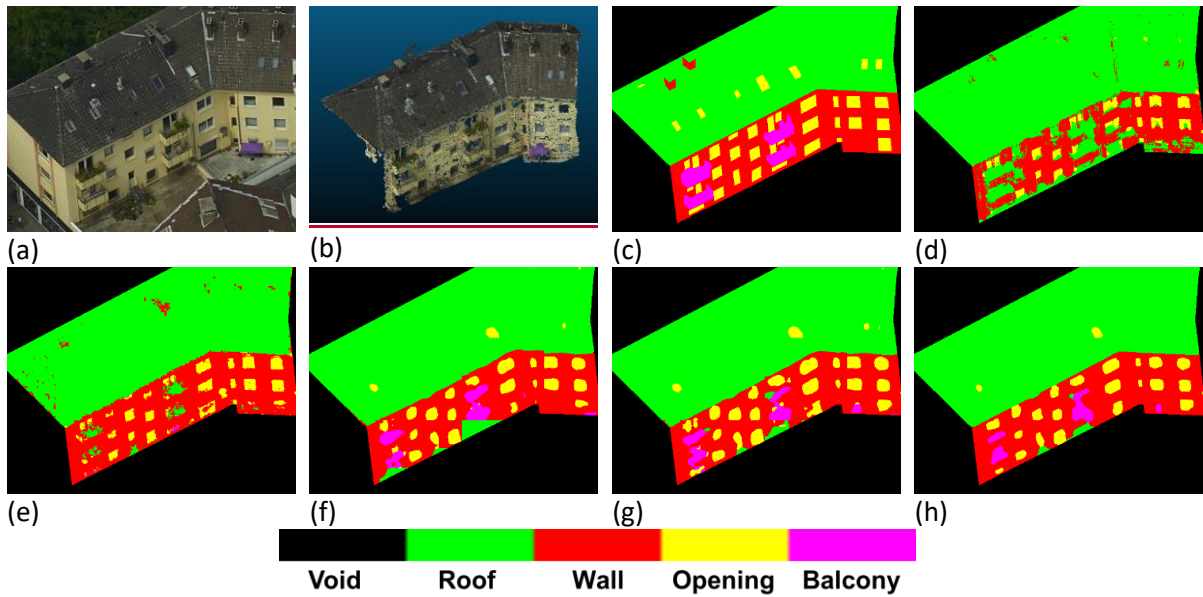


Figure An example from our dataset. (a) cropped façade image from oblique aerial images. (b) cropped façade point cloud (c) ground truth. (d) result from random forest using hand-craft image features. (e) result from random forest using hand-craft image and point cloud features. (f) result from the vgg16 net fine-tuned by RGB images. (g) result from the vgg-16 net fine-tuned by both RGB images and point cloud features. (h) result achieved refining the result of (g) with fully connected CRF.

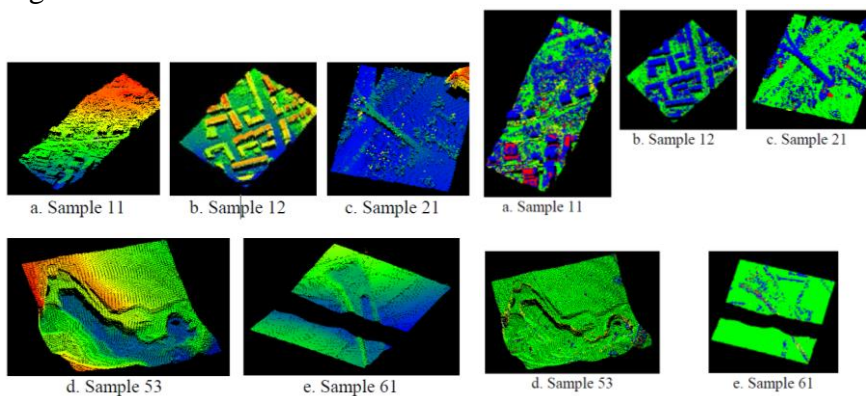
Fully Convolutional Networks for Ground Classification from Airborne Laser Scanner data

Name: **Sander Oude Elberink**

Affiliation: University of Twente, Faculty of Geo-Information Science and Earth Observation

E-mail: s.j.oudeelberink@utwente.nl

We present an efficient procedure to classify airborne laser scanner data into ground and non-ground. The classification is performed by a Fully Convolutional Network (FCN), a modified version of CNN designed for pixel-wise image classification. Compared to the previous CNN-based technique and LAsTools software, the proposed method reduces the total error and type I error (while type II error is slightly higher). The method was also tested on AHN-3 data resulting in 4.02% of total error, 2.15% of type I error and 6.14% of type II error. We show that this method can be extended to further classify the point cloud into buildings and vegetation.



Five samples of airborne laser scanner data colored by height (left), plus our ground classification results (right).

The detailed interpretation of pole-like street furniture in mobile laser scanning data

Name: **Fashuai Li**, Sander Oude Elberink, George Vosselman

Affiliation: University of Twente

E-mail: f.li@utwente.nl

The interpretation of pole-like road furniture in mobile laser scanning data has received much attention in recent years. Most current studies interpret road furniture as a single object, which is infeasible for road furniture with multiple classes. In order to tackle this problem, we propose a framework using machine learning classifiers to interpret road furniture into detailed classes based on their functionalities such as street lights and traffic signs connected with poles (Figure 1). The overall accuracy of the interpretation in one test site is higher than 90%. A screenshot of our result is as shown in Figure 2. To conclude, our framework well interprets road furniture at a detailed level, which is of great importance for 3D precise mapping.

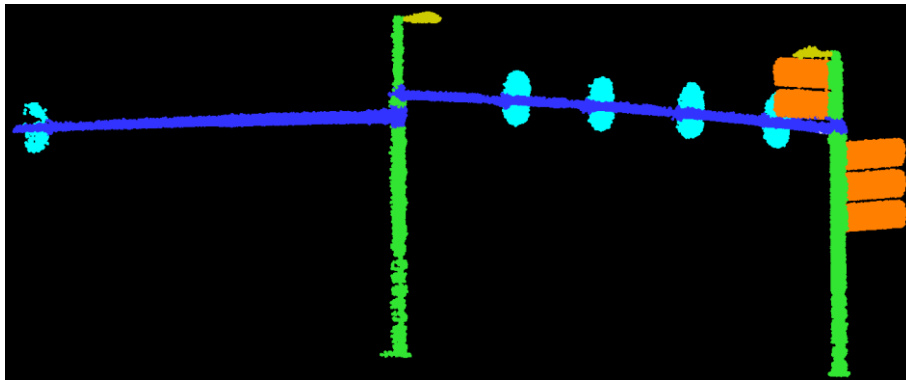


Figure 1. The interpreted road furniture (Orange: Street signs, Yellow: Street lights, Cyan: Traffic lights, Green: vertical poles, Blue: Horizontal poles)

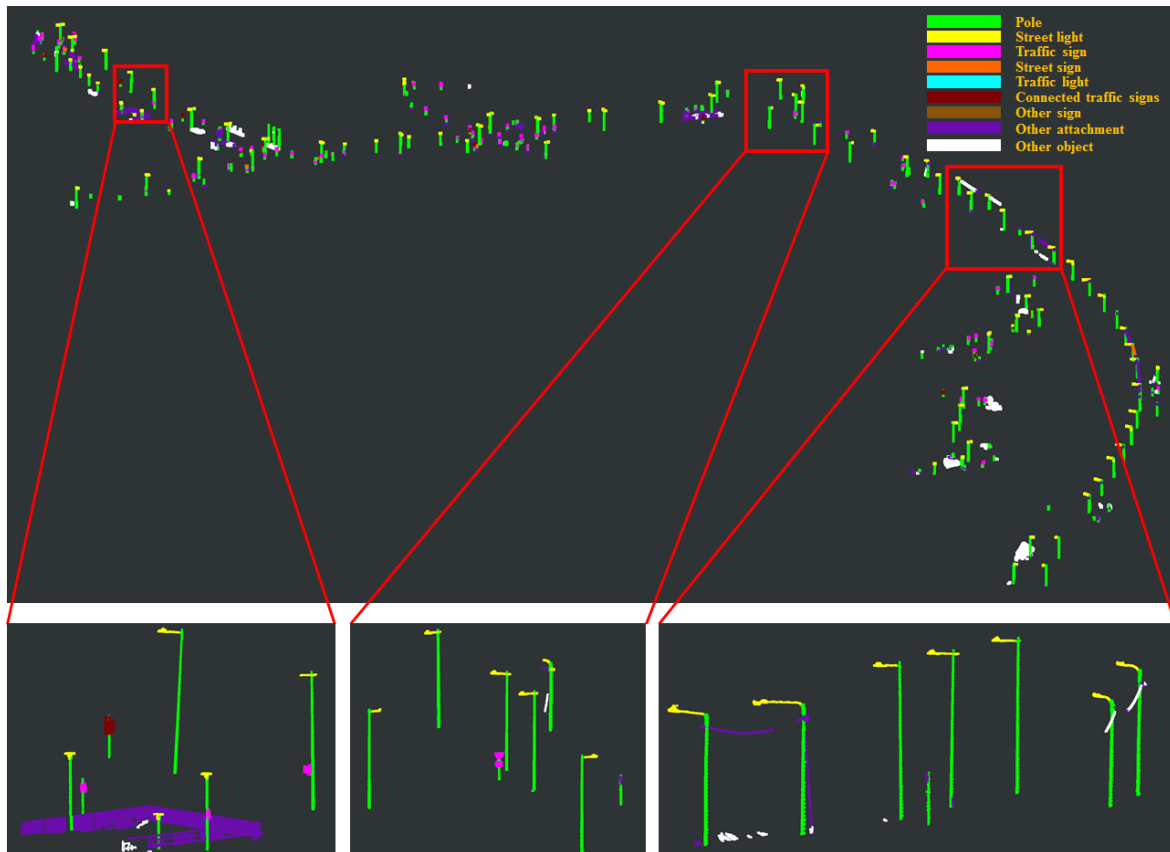


Figure 2. The interpretation of road furniture in Saunalahti dataset

Contextual Classification of 3D Textured Meshes for Urban Scene Interpretation

Name: Weixiao GAO

Affiliation: Technische Universiteit Delft

E-mail: W.Gao-1@tudelft.nl

Abstract:

The rapid development of multiple view geometry techniques has facilitated quick generation of dense urban 3D point clouds and meshes. The generated data can serve as base information for the semantic interpretation of 3D scenes and further automatic reconstruction of highly detailed 3D city models. Although textured meshes, which include both, geometric and radiometric features, contain more information than 3D point clouds, their classification has been so far barely explored. This paper introduces a novel approach for interpreting urban 3D scene composed of textured meshes acquired by multi-view stereo techniques. Our classification process applies a multi-feature region growing and Markov Random Field (MRF). This enables to consider the spatial relation between manifold unsupervised classified objects. The performance of the presented procedure is evaluated using 3D textured meshes generated from images provided by the EuroSDR/ISPRS benchmark on multi-view photogrammetry and by the Bentley/Acute3D. The test sample is selected from the area of Dortmund and manually labelled as ground truth. According to the evaluation result, the overall accuracy of our algorithm achieves 94.0%.

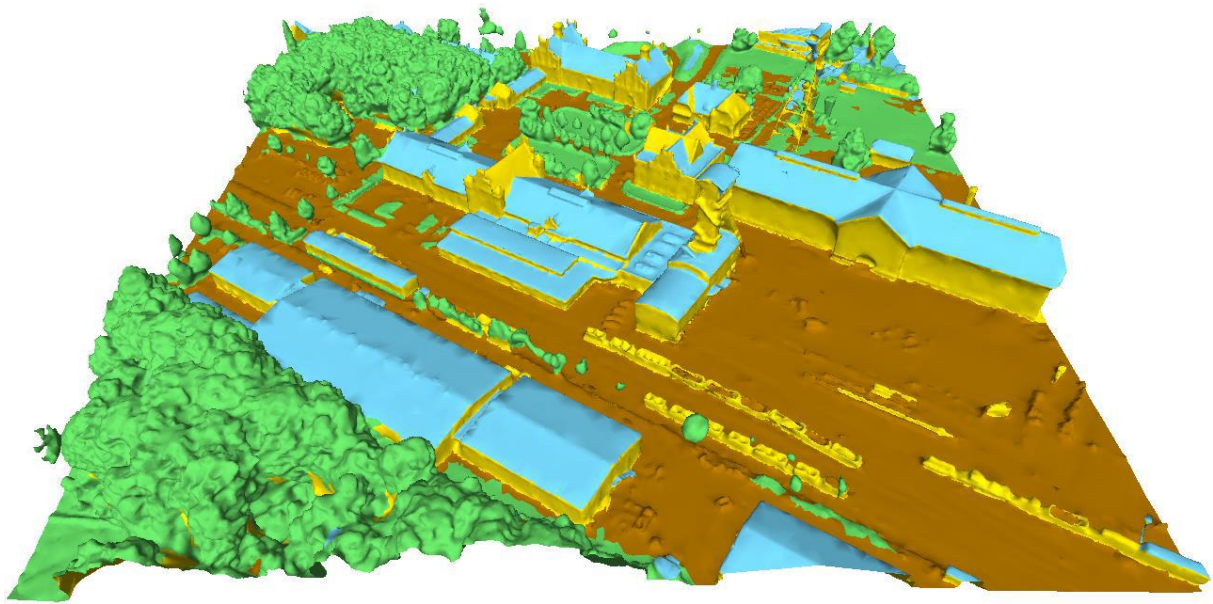


Figure1: Textured Mesh Classification Result