



HAL
open science

Complementarity deep learning features and normalized differences index for LTZ estimation

Adrien Chan-Hon-Tong

► **To cite this version:**

Adrien Chan-Hon-Tong. Complementarity deep learning features and normalized differences index for LTZ estimation. 2017. hal-01480726v2

HAL Id: hal-01480726

<https://hal.science/hal-01480726v2>

Preprint submitted on 3 Apr 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Complementarity deep learning features and normalized differences index for LTZ estimation

Adrien CHAN-HON-TONG

April 3, 2017

Abstract

We describe some experimentations based on the data of the 2017 data fusion contest (IEEE-IGARSS). These experiments highlights the complementarity between both deep learning features and normalized differences index for LTZ estimation

1 Introduction

Data fusion contest are a set of challenges of the remote sensing community. The 2017 challenge is about predicting LTZ zones [7] from training cities to unknown cities.

See <http://www.grss-ieee.org/community/technical-committees/data-fusion/data-fusion-contest/>

Provided inputs for doing this prediction include remote sensing images: landsat 100m images (multiple images per city), sentinel2 100m images, sentinel2 5m images and openstreetmap information available for these cities during challenge duration.

However, both these modality are very different and complementary. In one hand, openstreetmap provides spatially resolved, symbolic and semantic informations about the city but these informations may be biased (bias include difference of quality from different annotators, missing annotations, erroneous semantic, erroneous location, semantic misunderstanding between different annotators). On the other hand, remote sensing images may be low resolution, blurred, distorted by atmospheric events. Thus, taking advantage of those two modality may lead to better accuracy than using one modality only - even in a context where only few ground truth annotations (assumed not noisy) are available.

2 Experiment

We evaluate several pipelines which use raw images, normalized differences, and osm. Following the rules of the 2017 data fusion contest, we evaluate the quality of the LTZ estimation by measuring the pixelwise accuracy. The current experiment are done both on the testing data and

with a leave one city out protocol allowing more deeper evaluation (this last protocol is: all training cities except one are used to train the model which is then applied to the excluded city, this operation being done for all cities).

2.1 raw images

Raw images are used directly as features. For the landsat, we transform the provided multi date set of images into a mean image (we compute, for each pixel coordinate, the mean of the value of this pixel in all the image) and a variance image (we compute, for each pixel coordinate, the variance of the value of this pixel in all the image). Of course, this *variance* should not be interpreted as the real variance in a statistical term as there are often only two images. however, this is a simple but robust way to transform a set of value into a fixed number of values that are invariant from any permutation of the set.

Then, we concatenate all provided bands for both landsat and sentinel leading to a 27 channels images of each cities.

Then, we perform a 1 pixel classification (we try to add spatial context but due to the few number of annotated pixels, this result in a dramatic overfitting). Thus, each pixel is classified based only on the 27 values providing from the 27 bands (9 sentinel bands and 9 means and 9 variances for the 9 landsat bands).

2.2 osm

We use pretrained convolutional deep learning to extract features from osm. We use both landuse and building provided rasters and road vector (we rasterize the road vector on the same grid than the 2 other rasters). From the landuse, we extract *green* areas (vegetation, forest, park, ...) only. All 3 rasters are binarize (value is either 0 or 255) to form a classic 8bit RGB image (Road Green Building for the channel).

We apply a segnet like neural network [1] on these images. More precisely, we use a vgg16 [6] initialised by imagenet weight up to the conv3_3 layer following to a simple average deconvolution.

The predominance of the encoding part is due to the fact that images are 20m resolved against 100m for the images and LTZ annotations.

Also, as the images are known to be biased, we add a 2x2 pooling layer after each convolution to make the network robust to error and to increase context information.

2.3 normalized difference

We also extract well known normalized difference (we will call it *ndi*) features from each image. We use NDVI, NDWI, MNDWI, NDBI and WRI (see [5] for all, [3, 2] for historic description of the first ones).

method	berlin	HK	paris	rome	sao paulo	average	test
images (1vsall)	42%	25%	74%	22%	43%	38%	-
osm (1vsall)	47%	43%	59%	33%	21%	36%	-
images + osm (1vsall)	53%	51%	72%	30%	54%	48%	56%
images + osm	50%	52%	73%	33%	68%	51%	58%
images + ndi + osm	51%	53%	73%	34%	68%	52%	-
ndi + osm	57%	53%	67%	48%	52%	54%	57%

Table 1: Results of the leave one city out

2.4 results

Each pipeline is built by extracting a set of feature maps and using svm (e.g. [4]) to perform pixelwise classification. Precisely liblinear or libsvm (with linear kernel) have been used: liblinear uses 1vsall and libsvm uses 1vs1 which will be the default classifier in the following.

Leave one city out results are detailed in table 1 with *average* result and test result when available. In order to penalize, very unstable results across cities we weight the worse accuracy by a factor 2.

3 Discussion

Even if the challenge have ended with much more high score than the 57% reached by our method, we still think that this method is interesting for two reason. First, we believe that *ndi + osm* will be more stable on new city (neither from train or test - even if testing data where hidden, feedback from evaluation server may result in a little overfitting on *test* data). Then, our pipeline is kind of generic regarding the task (which is here LTZ estimation but could be another kind of geographic index for an other utilisation). In addition, the different performed experiments are interesting, in our opinion.

3.1 osm

OSM only performs poorly but it can be explained by the lack of some classes: there is no way to predict water from our osm raster. More deeply, our osm raster is able to be processed by a pretrained network but is very poor regarding all the osm information and regarding the information needed to perform correct LTZ estimation (e.g. difference between middle rise and low rise).

3.2 raw images as features

Raw images do not perform bad whereas it is very hard to understand how raw images can be used without any kind of normalization. A closer look to intensity across cities (per class) clearly show that avera intensity value of different images are not the same (probably due to illumination variance). However, svm still manadge to understand something into this feature vector. Currently, we try to investigate this point but we could

not due to bilinear numerical weakness: R-IR-B from sentinel perform better than sentinel only even on train data which is not acceptable with svm. Temporal part of landsat seems to be the informative part whereas we just compute the means/variance of different landsat without taking into account that its are not equally temporally distant across cities (nul variance could corresponds to unvariant area or close acquisition).

3.3 ndi

This is one of the reason, we consider ndi. Normalized difference index are more robust to illumination, thus both raw value of ndi and means/variance from multiple acquisition are relevant. Very surprisingly, adding ndi to the other features do not deeply increase the average accuracy. Currently, as we rely on 1 vs 1 svm, it is possible that some pair of classes are separated only by the sign of some band difference. In other words, the svm seems to have found by itself relevant band differences.

However, using ndi only highly increases stability of the results across cities. We feel that temporal information is much more stable but removing variance part leads to ambiguous results. Deep features from both images or ndi have been considered but lead to dramatic overfitting.

3.4 conclusion

In our opinion, the main result of this work is the important complementarity between *traditionnal* normalized difference index and pretrained deep learning features for adhoc geographic index estimation.

References

- [1] Vijay Badrinarayanan, Ankur Handa, and Roberto Cipolla. Segnet a deep convolutional encoder decoder architecture for robust semantic pixelwise labelling. In *arXiv preprint*, 2015.
- [2] Xiao-Ling Chen, Hong-Mei Zhao, Ping-Xiang Li, and Zhi-Yong Yin. Remote sensing image-based analysis of the relationship between urban heat island and land use/cover changes. In *Remote sensing of environment*, 2006.
- [3] Bo-Cai Gao. Ndwia normalized difference water index for remote sensing of vegetation liquid water from space. In *Remote sensing of environment*, 1996.
- [4] T. Joachims, T. Finley, and Chun-Nam J. Yu. Cutting-plane training of structural svms. In *Machine Learning*, 2009.
- [5] Komeil Rokni, Anuar Ahmad, Ali Selamat, and Sharifeh Hazini. Water feature extraction and change detection using multitemporal landsat imagery. In *Remote Sensing*, 2014.
- [6] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In *CoRR*, 2014.
- [7] I.D. Stewart and T.R. Oke. Local climate zones for urban temperature studies. In *Bulletin of the American Meteorological Society*, 2012.