Mõttus, Matti; Molinier, Matthieu; Halme, Eelis; Cu, The; Laaksonen, Jorma

Patch size selection for analysis of sub-meter resolution hyperspectral imagery of forests

Published in:
Proceedings of 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS

DOI:
10.1109/IGARSS47720.2021.9554257

Published: 16/07/2021

Document Version
Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

PATCH SIZE SELECTION FOR ANALYSIS OF SUB-METER RESOLUTION HYPERSPECTRAL IMAGERY OF FORESTS

Matti Mõttus[*]*, Matthieu Molinier, Eelis Halme*

VTT Technical Research Centre of Finland Ltd
P.O. BOX 1000, FI-02044 VTT
Finland

Hai Cu, Jorma Laaksonen[*]

Aalto University School of Science
P.O. BOX 15400, FI-00076 Aalto
Finland

ABSTRACT

Very high resolution remote sensing data of forests, where individual tree crowns are separable, contains structural information on tree size and density. Such information is complementary to the spectral signatures currently used in forestry applications. Advanced machine learning methods, e.g. convolutional neural networks (CNNs), offer an automated and standardized way of retrieving both spectral and structural information from imagery. A key characteristic in CNNs is patch size, which should be large enough to include dominant structural scale, yet as small as possible to avoid unnecessary averaging. Our results show that the patch should be larger than one tree, but increasing it excessively reduces retrieval accuracy. Furthermore, large patch sizes can cause loss of independence between training and validation data, leading to overestimating model performance.

Index Terms— Very high-resolution imagery, Hyperspectral data, Deep learning, Convolutional Neural Networks, Patch size, forest variable prediction, TAIGA dataset

1. INTRODUCTION

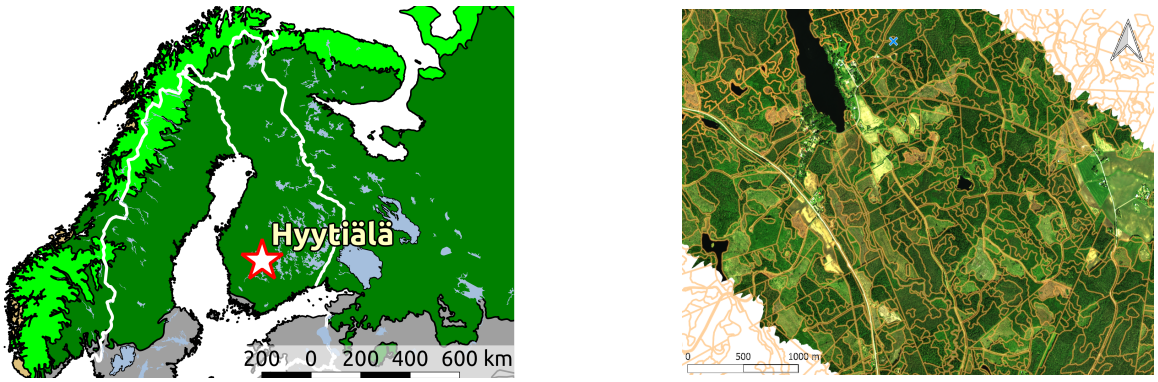
Machine learning based methods have the potential to make use of the complex indirect and nonlinear relationships in modern Earth Observation (EO) data and reveal their full information content [1]. Deep learning methods have their background in speech recognition and computer vision fields, and are hence recognized for their ability to process image data. These methods, increasing their popularity in EO applications, have been adapted to the spectral and spatial characteristics of remote sensing imagery [2]. Various deep learning architectures have been successfully applied: Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNN), hybrid ConvRNN, attention modules, capsule networks, etc. CNNs were used in earlier works [3] and are still widely used [4].

EO involves data of varying resolutions. EO data from aircraft or satellites with spatial resolution of one meter or less (known as Very High Resolution, VHR, imagery) is becoming increasingly available for extensive geographic areas. Hyperspectral Imaging (HSI, also known as imaging spectroscopy) captures reflectance over hundreds of bands forming a continuous spectrum. When acquired from aircraft, HSI is often VHR data, and can be expected to contain more information compared to most other optical EO data due to their high spatial and spectral resolutions. Current hyperspectral EO algorithms, however, characterize targets based on their spectral properties. Ignoring the rich spatial detail present in VHR images inevitably leads to underutilization of their information content. CNNs have been found to be powerful tools [4] for such high-dimensional data and several review articles are available on the topic [5, 4]. CNNs can extract discriminatory features from the data cube and exploit the spectral and spatial information of the data which, in HSI classifications, improves retrieval accuracy [6, 7, 8].

Many specific CNN architectures can be devised for analyzing (hyper)spectral RS imagery. A fundamental physical question which needs to be answered for any model is the choice of the size of the window, known as patch size, given as an input to the neural network. The patch size should be selected to correspond to the characteristic features of the object while still keeping the patch homogeneous: i.e., the studied characteristic could be considered constant across the patch. Patch size, typically specified in pixels, in imagery of forested areas therefore depends on the dimensions of the typical objects, tree crowns, and on landscape structure, i.e., the size of a homogeneous forest stand. Patch size also limits the maximum number of independent samples which can be collected from an image. Unfortunately, many datasets used as benchmarks in deep learning are not large enough to avoid overlap between training and testing data at large patch sizes.

The aim of this paper is to study the effects of the image patch size, used as an input to a CNN, on the ability of the CNN to retrieve forest characteristics from VHR hyperspectral imagery of a boreal forest. To achieve this, we present a new HSI dataset that allows using patches up to 45×45 pixels without spatial overlap between training and test sets.

[*]Thanks to Academy of Finland for funding AIROBEST (grant 317387).



(a) Location of the test site in Finland

(b) Zoom-in on stand boundaries

Fig. 1. Geographic context of the study: (a) Location of the test site in the European boreal forest zone (dark green color) and (b) RGB image of a subset of the mosaicked hyperspectral data, overlaid by the stand boundaries of the region (orange).

2. DATA

For the study, we used the The Artificial Intelligence dataset for forest Geographical Applications (TAIGA) dataset openly available via the Fairdata IDA service (https://ida.fairdata.fi). Data were collected in the southern boreal zone of Finland, in the vicinity of Hyytiälä forestry field station (61°50'44"N, 24°17'10"E, Fig. 1a). The undulating landscape (mean elevation approx. 150 m above sea level) is covered by managed boreal forests, agricultural fields and wetlands. Scots pine (*Pinus sylvestris*), Norway spruce (*Picea abies*) and Silver birch (*Betula pendula*), either as mixed or pure stands, are the main overstory species. Forest floor is covered by different shrubs, lichens and mosses with no visible bare soil.

Airborne HSI was collected with the AISA Eagle II scanner in 128 bands between 400 and 950 nm from a height of approximately 1 km above ground. The spatial resolution and pixel size of the data were 0.7 m. Seven flight lines were flown before noon (mean solar zenith angle 41.3°) along the direction of sunrays under a clear sky on June 15, 2017 and mosaicked to obtain a hyperspectral image of an area of approximately 9 km by 3 km. Software packages by Rese Applications (Switzerland), Parge and Atcor, were used for georectification and atmospheric correction of the data, respectively, using a digital elevation model based on nationwide laser scanning data and sun photometer-measured aerosol optical properties provided by the AErosol RObotic NETwork (AERONET, https://aeronet.gsfc.nasa.gov). A comprehensive description of data acquisition and preprocessing is reported by Markiet et al. [9].

In Finland, forest resource data and forests are managed on the level of a stand (Fig. 1b), an aggregation of trees that are sufficiently uniform in species composition, size, arrangement and age. In the inventory process, forestry data are collected using airborne laser scanning and imaging supported by nationwide field measurements. The forestry parameters are computed for inventory units, 16×16 m^2 cells. Finally, stand level data are aggregated from the grid cells using vector files with stand geometries.

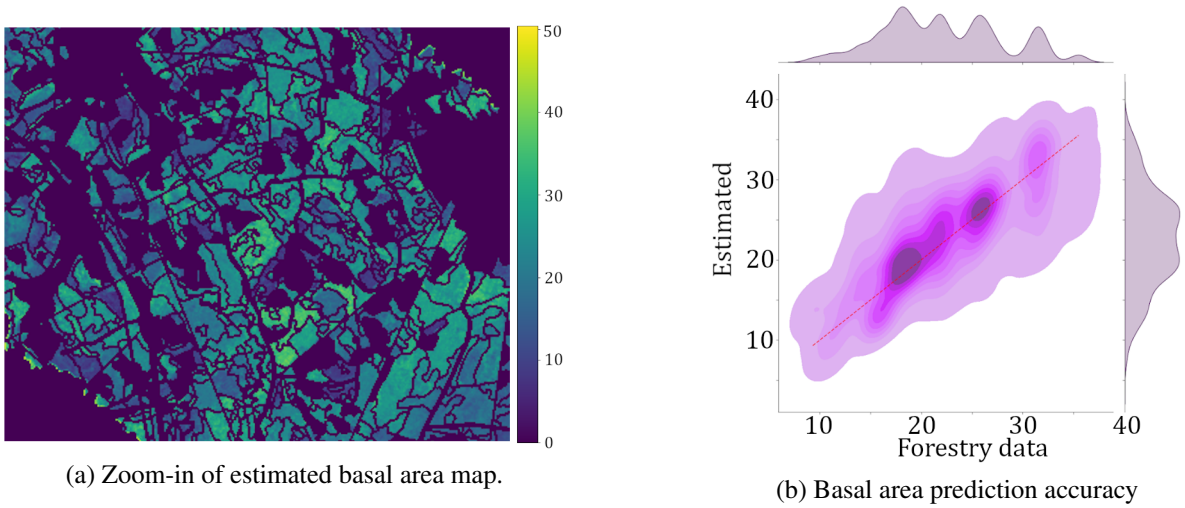
In line with the forestry practices, we assumed stands to represent homogeneous forest units. To avoid the influence of geolocation errors and neighborhood effects, we downsized the stands by 10 meters before extracting pixels from the HSI. Thus, all data outside forest stands or within 10 m of a stand border were excluded from analysis.

We selected two forest parameters for further analysis. Basal area, expressed in units of m^2/ha, is the cross-sectional area of tree trunks at breast height, 1.3 m, per unit ground area. This key forestry variable correlates with the growing stock of forests and their biomass. Second, we computed the leaf area index, (LAI, half of the two-sided leaf area per unit ground area), a key biophysical variable, using the forestry data and biometric regressions [9]. We calculated the effective LAI, which is corrected for the effect of the clumping of conifer needles into shoots on the transmission of incident sunlight and is thus a predictor of the reflective properties of forests. For birch, no clumping correction was applied.

We normalized the two continuous forest variables between 0 and 1 before feeding them to a CNN. To remove outliers (unphysically large values due to errors in the data), we clipped the upper limit to the 98$^{\text{th}}$ percentile.

3. CNN FOR PREDICTING FORESTRY VARIABLES

We used a CNN pipeline based on a multi-task learning model consisting of four 3D convolutional layers, introduced by Phu Pham [10] and inspired by Chen et al. [6]. The 3D input kernel's size was $3\times3\times48$ and it could adapt in both the spatial and hyperspectral dimensions of the input data. The first layer had 128-dimensional output and was followed by $2\times2\times2$ pooling. The second convolutional layer consisted of parallel $1\times1\times1$, $1\times1\times3$, $1\times1\times5$ and $1\times1\times11$ networks; its output



(a) Zoom-in of estimated basal area map.

(b) Basal area prediction accuracy

Fig. 2. Prediction results for basal area [m^2 ha^{-1}] using a window of 45 pixels.

Table 1. The numbers of stands and sizes of training, validation and test sets.

set	stands	samples
training	〉560〈	37749
validation		4195
test	111	20263

was also 128-dimensional. The last two CNN layers both had $3\times3\times32$ kernels, whereas their output dimensionalities were 64 and 32, respectively. Between these two layers, $2\times2\times2$ pooling was performed. Batch normalization and ReLU nonlinearity were used after the convolution layers, followed by the optional pooling. Dropout regularization with a probability of 0.5 was used for all layers.

The convolutional layers were followed by one fully-connected layer shared between all learning tasks, each learning task corresponding to one output variable. For each output variable, a separate stack of two fully-connected layers was included in the model. For the continuous variables, mean square error loss was used at the model output. The final stand-wise predictions for the forest variables of the test set were obtained by averaging the predictions for the image samples located in each stand of the test set.

Different stands were used for training and testing: 17% of the stands were selected randomly for testing and the remaining 83% were split between training and validation (Table 1). The HSI samples of the sets were then extracted from inside each stand with the stride of 13×13 pixels. The size of these image samples was varied as 27×27, 33×33, ..., 91×91 pixels in order to progressively increase image area for the training of the CNN model. While yielding more training data, increasing the patch size leads to the danger of including the same pixels in both the training and testing sets.

Table 2. Forest variable prediction results with varying input patch size to the CNN model.

patch size	over-lap	Basal area [m^2/ha]		LAI [–]	
		RMSE	rRMSE	RMSE	rRMSE
27×27	0.00%	3.69	16.35%	0.74	22.68%
33×33	0.00%	3.48	15.39%	0.72	22.08%
39×39	0.00%	3.54	15.67%	0.73	22.45%
45×45	0.59%	3.60	15.95%	0.69	21.29%
75×75	17.64%	3.87	17.12%	0.75	23.02%
91×91	28.58%	3.85	17.05%	0.74	22.91%

4. SPATIAL OVERLAP ANALYSIS

Spatial overlap between test and training pixels leads to overly optimistic accuracy assessment of patch-based deep models. This issue was known earlier for spatial-spectral methods applied to popular small-scale HSI datasets, and is especially severe when using random sampling and larger patches [4]. As a mitigation measure, the IEEE Geoscience and Remote Sensing Society (GRSS) designed fixed training and testing sets for these datasets, available on the data and algorithm evaluation website (DASE) - http://dase.grss-ieee.org.

In order to demonstrate the utility of the TAIGA dataset for safely applying deep models with large patches, we ran a simple spatial overlap analysis. First, training and testing 'ground truths' were binarized into masks. Then, training and testing masks were iteratively grown by applying a patch of increasing size over all original pixels, simulating the receptive field of a patch-based network in training and testing phases. Finally, the number of pixels overlapping between the training and test masks was counted and normalized by the total number of test pixels covered by the receptive field. In addition to TAIGA, we analyzed Indian Pines, Pavia Uni-

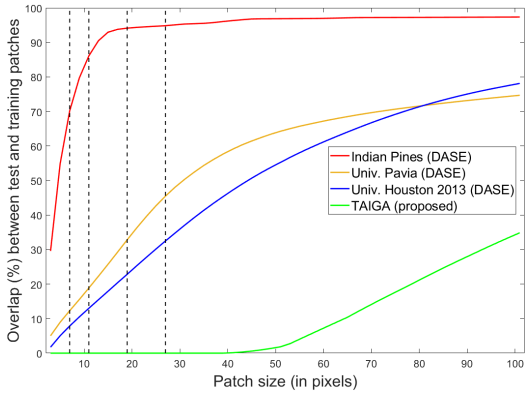


Fig. 3. Overlap between test patches and training patches as a function of patch size, for different hyperspectral datasets. Dashed lines show typical patch sizes used in HSI studies [4].

versity and University of Houston 2013 datasets, using the DASE disjoint training and testing samples.

5. RESULTS AND DISCUSSION

We calculated root-mean-square error (RMSE) and relative root-mean-square error (rRMSE) as the prediction accuracy. The prediction accuracy for basal area (rRMSE = 15–21%) is better than for LAI (rRMSE = 21–28%, Table 2). The prediction accuracy generally increased with patch size for both variables up to 45 $\times$ 45 pixels, then decreased with higher patch sizes as significant spatial overlap occurred (75 $\times$ 75 pixels and 91 $\times$ 91 pixels). As an example, the predicted basal area map and prediciton accuracy for the optimal patch size (45$\times$45 pixels) is shown in Fig. 2.

The spatial overlap analysis shows that the selected DASE datasets have some spatial overlap already at the lowest patch size used here, 3$\times$3 pixels, whereas overlap for TAIGA starts above 43$\times$43 pixels (Fig. 3). The DASE datasets have more than 10% spatial overlap for patches larger than 11$\times$11 pixels with the Indian Pines dataset saturating at 15$\times$15 pixels. In contrast, TAIGA has only 10% overlap at 65$\times$65 pixels, allowing robust prediction accuracy estimates for patch sizes corresponding to mid-resolution EO satellite pixels and quantifying the improved information content of VHR imagery.

As we increased patch size beyond the optimal value, the accuracy of forest variable predictions decreased, which is usually not observed in studies on small-scale HSI datasets [4]. This phenomenon, probably caused by landscape heterogeneity, calls for further research. We hope TAIGA, a reference dataset with continuous spatial coverage, will help to understand and solve the remaining issues in the use of VHR HSI data for vegetation analysis, especially those caused by the small size imagery currently available for testing.

6. REFERENCES

[1] J. Verrelst, Z. Malenovský, C. Van der Tol, G. Camps-Valls, J.-P. Gastellu-Etchegorry, P. Lewis, P. North, and J. Moreno, "Quantifying vegetation biophysical variables from imaging spectroscopy data: a review on retrieval methods," *Surveys in Geophysics*, vol. 40, no. 3, pp. 589–629, 2019.

[2] X. X. Zhu, D. Tuia, L. Mou, G. Xia, L. Zhang, F. Xu, and F. Fraundorfer, "Deep learning in remote sensing: A comprehensive review and list of resources," *IEEE Geoscience and Remote Sensing Magazine*, vol. 5, no. 4, pp. 8–36, Dec 2017.

[3] G. Fu, C. Liu, R. Zhou, T. Sun, and Q. Zhang, "Classification for high resolution remote sensing imagery using a fully convolutional network," *Remote Sensing*, vol. 9, no. 5, 2017.

[4] M.E. Paoletti, J.M. Haut, J. Plaza, and A. Plaza, "Deep learning classifiers for hyperspectral imaging: A review," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 158, pp. 279–317, 2019.

[5] S. Li, W. Song, L. Fang, Y. Chen, P. Ghamisi, and J. A. Benediktsson, "Deep learning for hyperspectral image classification: An overview," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 9, pp. 6690–6709, 2019.

[6] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 10, pp. 6232–6251, 2016.

[7] S. Mei, J. Ji, J. Hou, X. Li, and Q. Du, "Learning sensor-specific spatial-spectral features of hyperspectral images via convolutional neural networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 8, pp. 4520–4533, 2017.

[8] M.E. Paoletti, J.M. Haut, J. Plaza, and A. Plaza, "A new deep convolutional neural network for fast hyperspectral image classification," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 145, pp. 120–147, 2018.

[9] V. Markiet, R. Hernández-Clemente, and M. Mõttus, "Spectral similarity and PRI variations for a boreal forest stand using multi-angular airborne imagery," *Remote Sensing*, vol. 9, no. 10, pp. 1005, 2017.

[10] P. Pham, "Deep learning methods for modelling forest biomass and structures from hyperspectral imagery," Master's thesis, Aalto University School of Science, 2019.