

---

# Towards Automated Real Estate Assessment from Satellite Images with CNNs

Valentin Muhr

Kufstein University of Applied Sciences  
Email: v.muhr@a1.net

Miroslav Despotovic

Kufstein University of Applied Sciences  
Email: miroslav.despotovic@fh-kufstein.ac.at

David Koch

Kufstein University of Applied Sciences  
Email: david.koch@fh-kufstein.ac.at

Mario Döller

Kufstein University of Applied Sciences  
Email: mario.doeller@fh-kufstein.ac.at

Matthias Zeppelzauer

St.Pölten University of Applied Sciences  
Email: matthias.zeppelzauer@fhstp.ac.at

**Abstract**—A driving factor for real estate prices is the location quality. Models for location quality are usually built from available price information and distinct GIS information. In this paper, we present a first approach towards the automated assessment of location quality from satellite images using computer vision. For this purpose, we first introduce a novel dataset generated from publicly available data sources with suitable ground-truth annotations for location assessment. Next, we adapt a state-of-the-art convolutional neural network (CNN) and adapt it to predict different land covers and objects from satellite images. Finally, we feed information derived from the recognized land covers into a regression-based price model which acts as a proxy for the assessment of location quality. Our results show that (i) land cover classification can be performed with high accuracy and demonstrates that automatic classification could further be used in the future for the detection of mis-aligned and erroneous GIS data; (ii) our adapted network reaches state-of-the-art performance in much less training time compared to our reference network; (iii) the automatically extracted visual information improves the prediction of real estate prices and thereby shows clear potential for the description of location quality.

## I. INTRODUCTION

One of the most important criteria for assessing real estate is its location and its neighborhood. The by the authors of this paper specified research objective is to assess the *quality and livability* of urban geographical locations to support the automatic assessment of real-estate objects and properties. To realize this goal, we aim at leveraging the capabilities of automated land cover analysis in satellite images.

Land Cover Analysis focuses on the extraction of information from remote sensing satellite image data and GIS data with the goal to study and monitor geological resources and their dynamic changes [1]. In this aspect, remote sensing has become more and more important due to the increasing amount of available geographical data [2], provided through satellite recordings world-wide and comprehensive descriptive information collections from GIS. These data allow better understanding and development of the earths global and physical processes [3] and in combination with GIS have fostered several applications and meaningful results in the fields of agriculture, environment, and eco-environment assessment [4],

[5], [6], [7]. Due to these comprehensive data former limitations to recognize fine-grained structural patterns of objects (e.g. buildings) and background are dissolved [8], and thereby enable a number of novel approaches, as the one presented in this work.

A rich literature on satellite image analysis and land cover classification exists [9], [10] and numerous datasets have been introduced [11], [12] for developing automatic classification and detection algorithms. Furthermore, recently, deep learning has made a significant impact on the field, especially Convolutional Neural Networks (CNNs) [13] have improved classification performance of satellite imagery significantly [9], [14], [15], [16], [17]. Although there is work on many different tasks in the context of satellite image analysis, such as building detection, road network analysis, terrain classification, flood detection, and even poverty detection, the detection of location quality has, to the knowledge of the authors, not been investigated so far [18], [19], [20], [21], [22].

For a proper development and implementation of our research objectives, we had to build our own satellite image dataset, which suits the requirements for location quality assessment. For this purpose, we used publicly available satellite image data from the Tyrolean Tiris Database<sup>1</sup> and the freely available Open Government Data<sup>2</sup> (OGD). For location quality assessment we have defined a set of eight relevant land covers, which reflect geological characteristics and also considerable infrastructure in Austria. Based on this novel dataset, we designed a CNN for location assessment and compared it to an existing network (VGG network model) [23]. The networks were designed (and adapted respectively) to classify the land covers and objects in our dataset.

This work represents a first step towards our main research objective to assess location quality by providing the necessary basis information for this task, namely pre-classified land

<sup>1</sup>Orthofoto Tirol - data.gv.at. [Online] Available: <https://www.data.gv.at/katalog/dataset/35691b6c-9ed7-4517-b4b3-688b0569729a>. Accessed on: Aug. 30, 2017

<sup>2</sup>Datenauftritt Land Tirol — data.gv.at. [Online] Available: <https://www.data.gv.at/auftritte/?organisation=land-tirol&katFilter=geographie-und-planung#showresults>. Accessed on: Aug 30, 2017

covers, terrains and objects in RGB satellite images. The training of regression models for location quality based on this analysis data is currently under investigation. The hedonic pricing method, provides the theoretical background in real estate and is widely used in housing research and appraisals [24]. Originally developed for automobiles by [25], hedonic price models describe how the quantity and quality of these characteristics determine its price in a particular market. In a formal way, the general hedonic price function takes the form

$$P_i = f(S_i, L_i, N_i) \quad (1)$$

where  $P_i$  is the *log* of the price or rent of house  $i$ ,  $S_i$  is a vector of structural housing characteristics,  $L_i$  is a vector of location variables and  $N_i$  is the neighborhood characteristics. In the field of hedonic pricing, only the aspects that are available as quantifiable data, e.g. for location variables, such as distances to schools, public transport, etc., are considered. For instance, if one wants to evaluate sociodemographic data within the scope of hedonic pricing on a certain geographical level, an appropriate statement can be developed only on that particular aggregated level. Therefore, the exact location or the immediate neighborhood can be valued only conditionally in a fully automatic way. Furthermore, the demarcation of micro-residential areas based on e.g. predefined spatial units, and price similarity cannot always lead to an adequate qualitative distinction. Pattern recognition from satellite images circumvents this problem, as the factual delineation of the macro-location is ignored and only the features within selected micro-location are being investigated.

## II. RELATED WORK

The major objective of image analysis is to extract discriminative visual features from images [26]. An early approach, which is sometimes still applied today, is the application of manually designed visual features based on color- and edge-descriptors [27], [28], e.g. histograms and correlograms. Later local features (e.g. SIFT, SURF, ORB) were introduced and became popular within the Bag-of-Visual-Word model for image classification [29]. During the last years, increasingly more automated methods for feature extraction and entirely end-to-end learned approaches based on neural networks (deep learning, DL) have been introduced and achieved remarkable improvements in image analysis and classification [30], [31], [32].

In remote sensing different image types are utilized, ranging from grayscale and RGB images, to multi-spectral images, and Synthetic Aperture Radar (SAR) images, as well as measuring procedures such as Light Detection and Ranging (LiDAR) [33] enabling 3-dimensional geographical surface scanning. RGB satellite images are the most widely distributed resources in remote sensing, since region-wide captures are broadly available.

In RGB satellite image analysis we can differentiate between low-level tasks and (more sophisticated) higher-level tasks, which often build upon low-level tasks. Low-level tasks focus on detection of certain objects, such as buildings [34],

[35], [36], [37], road networks [38], [39], [40], and the segmentation or classification of different land covers, such as grassland, forest, and water [41], [42], [43], [44]. Satellite images are usually cut into smaller patches before analysis, in order to provide local spatial information (at a particular location) together with its surrounding context (neighborhood) to the subsequent analysis. Afterwards, traditionally feature extraction is employed followed by machine learning [28]. Deep Learning, in contrast, combines feature extraction and classification into one process, which is demonstrated by an early work by Mnih and Hinton [45]. This work is based on a neural network with multiple local and fully-connected layers, designed to detect roads and road networks. In recent years many different methods for feature learning based on neural networks have been introduced [46], [47], [41], [31], [48]. In these methods, the extracted features (neural activations at intermediate layers of the networks) are further processed by other classifiers, such as Support Vector Machines (SVM) to obtain a final classification. In other works the networks combine feature extraction and learning into one supervised learning process by adding a classification layer on top of the network [45], [22], [44], [49], [50], [51], [32], [43].

Higher-level tasks often build upon a basic land cover classification or segmentation and include, for example building type classification, i.e., the classification of residential buildings, apartment buildings, and industrial and factory building [52], [53] to support urban development and proper planning. Higher-level tasks exist for different scaling levels. An example for a very fine-grained analysis is the segmentation of rooftops of buildings into areas of different angles [35] and thereby the evaluation of the suitability of these rooftops for photo-voltaic systems [54], [55]. Other works focus on higher-level tasks at a much larger scale, such as the analysis of abstract demographic and socio-economic parameters from satellite images. An example is the work of [56] where the authors try to derive regions of poverty through satellite image data. They exploit the light intensity in areas of settlement in night satellite images as an indicator. Based on this information they apply transfer learning to identify characteristic visual features which indicate poverty. In a follow-up work, the authors were able to explain 75% of the variation of economic outcomes from satellite images with a CNN [19].

The assessment of location quality is a new high-level task for satellite image analysis introduced in this paper, which builds upon low-level land cover classification. To facilitate the characterization of location quality from land covers, quality-relevant land covers need to be selected and recognized in the images. We have performed a survey on satellite image datasets to identify suitable annotated data collections for location assessment.

One of the most recent dataset is the AID dataset. It has been introduced in early 2017 by [57] and covers 10.000 object images divided into 30 individual categories. The set entails high intra-class variations, since visible lighting conditions and time or season differ from sample to sample. Also due

to multi-resolution collections, the images range widely in their spatial resolution of 50 cm and up to eight meters. For our initial experiments on location assessment, we aim at a more controlled dataset with less degrees of freedom (i.e. different scales and different seasons) to reduce the number of influencing factors. We consider the AID dataset as a challenging dataset for future experiments, once a first location assessment method has been established.

The RSSCN7 dataset, introduced by [58], contains 2800 satellite images with eight different annotated land cover classes. With 400 images in the categories of grassland, forest, farmland, parking lot, residential region, industrial region, and river and lake, these classes are closely resembled to the introduced dataset of current reading. Challenging is the fact that the samples are captured under different seasons, weather conditions and resolutions. The overall size, however, is much smaller than our dataset (10k images per class vs. 400 images per class).

The UC Merced Land Use Dataset by [59], is one of the most utilized datasets [60], [57], [61], [62], [63] and has been introduced in 2010. This set covers 21 object classes with 100 images each, in a spatial resolution of 30 cm. Although, this dataset covers mostly all of the, for this research determined, land covers and objects, the research project of this reading requires far more image samples per category.

Due to the reasons mentioned above, we have created a new satellite dataset from publicly available satellite and GIS data to better address our research objectives.

### III. DATASET

The satellite image dataset for our investigation has been retrieved from the orthophoto map of the Tyrolean Tiris Database in combination with Open Government Data (OGD). Images together with the OGD have been loaded into the GIS software QGIS (version 2.18.2), which enables to align the satellite images and the vector layers of the OGD and to process them together. The OGD covers multiple different land cover and object categories. From the rich available data, we have selected those categories that are most relevant for our research purpose of location quality assessment. A total of eight object and land cover categories has been selected:

Residential Buildings	Local Roads
Non-Residential Buildings	Grassland
Highway	Water
Federal Highway	Forest

These eight classes are assumed to be most descriptive and provisioning for the visual representation of local land covers and infrastructures in Austria. Together they represent the ground truth classes of our dataset. By using the scripting capabilities of QGIS we have exported large-scale image tiles from the satellite images at the highest available resolution. The area from which tiles were extracted covers the entire region of Tyrol in Austria. Since large areas in Tyrol are

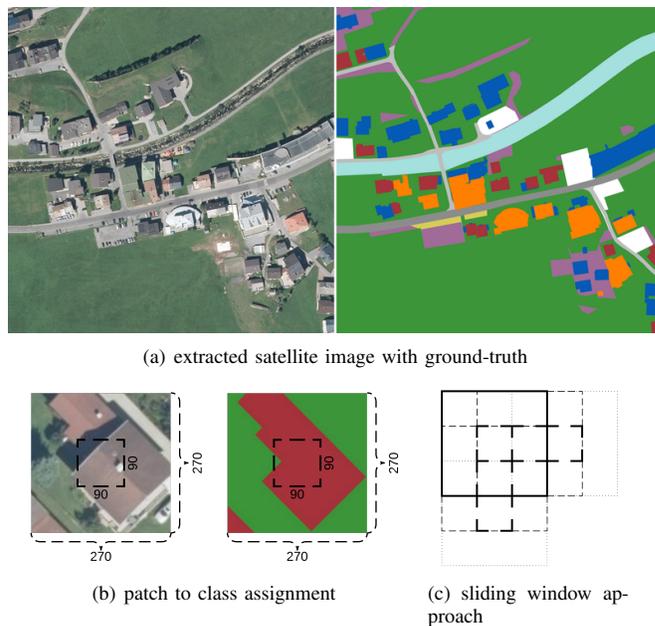


Fig. 1. (a) (left) extracted satellite image; (right) extracted ground truth. (b) Patch extraction algorithm, localizing the centered object and identifying its corresponding class. (c) Sliding window approach, visualizing the right and downward movement by 90px.; Credit: Land Tirol - data.tirol.gv.at

hardly or non-settled (covered settlement area 11.8% of whole Tyrol [64]) we have selected those tiles which cover settlement areas, resulting in a total of 21,076 tiles. Each obtained tile (4,050x4,050 px) covers 500x500 m<sup>2</sup>. The tiles have a ground resolution of 8.1 cm<sup>2</sup> per pixel. The vector layers were rasterized with the same resolution as the satellite image tiles for further processing. Figure 1(a) shows a part of an image tile with the corresponding ground-truth. The dataset has then been partitioned randomly into a training set (80%), including a hold-out validation set (10%), and a test set (20%). This initial split guarantees a clear separation of training and test data in our experiments.

Next, every satellite image and ground truth layer has been processed in parallel by a sliding window operation to cut out patches from the images. The patch size is determined by  $4050px/15 = 270px$ . A single patch thus covers 478 m<sup>2</sup> of landscape and thereby provides a significant amount of visual context for the detection of land covers and objects.

For each image patch a ground-truth label has been computed. The ground-truth label is derived from the center area of a patch (90x90 px) and is specified as the label of the most frequently occurring land cover in this area. The remaining area of the patch is considered as contextual information and is not used to define the class label. Figures 1(b) illustrates the labeling process of an image patch, showing a residential building.

The patch extraction is performed with a step size of 90 px for the sliding window operation. In this way a dense coverage of the input tiles is obtained, see Figure 1(c). Thereby, each tile is cut into 1,849 patches.

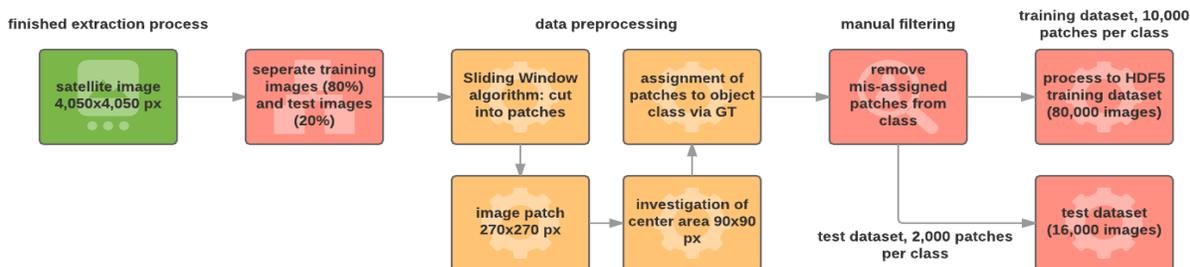


Fig. 2. The process workflow of database construction.



Fig. 3. An OGD polygon which is indicating the shape of a federal highway.; Credit: Land Tirol - data.tirol.gv.at

During construction of the dataset it showed that the OGD shape files (vector layers) frequently exhibit poor accuracy. Distorted and inaccurate polygons occur frequently, which do not match with the satellite image. Figure 3, demonstrates an example of an inaccurate OGD polygon. The visualized polygon should indicate the shape of a federal highway but actually overlaps to a large degree with the nearby forest. Similar observations have been made in all other categories. To clean the dataset, we have manually investigated the input tiles and the patches and removed those with inaccurate OGD polygons. For each category 15,000 patches have been manually investigated. The amount of patches with inaccurate ground-truth are listed in Table I. The percentage reaches from 6% for local roads to 57% for highway. The only exception are the two building classes with a tolerable deviation of only 1%. Since category of highway and water consisted of so many wrong images, additional images to the previous 15,000 have been investigated. The final training set consists of 10,000 patches per category and the test set contains another 2,000 patches per category. Thus, there are 80,000 patches in the training set and 16,000 patches in the. The complete process of dataset generation is summarized in Figure 2. Finally, the training dataset is converted to HDF5 format for easier processing.

#### IV. METHOD

For the classification of the different land covers and objects in our dataset we propose - in accordance with the current state of the art - end-to-end learning by deep convolutional

TABLE I  
NUMBERS AND PERCENTAGES OF MIS-ASSIGNED PATCHES INSIDE EACH CATEGORY DUE TO INACCURATE OGD.

Category	% mis-assigned
Residential Buildings	1% (132/15,000)
Non-Residential Buildings	1% (147/15,000)
Highway	57% (8,578/15,000)
Federal Highway	20% (2,991/15,000)
Local Road	6% (913/15,000)
Grassland	22.5% (3390/15,000)
Water	35.5% (5,332/15,000)
Forest	10% (1,489/15,000)

neural networks (CNNs). As described in Section III, the satellite images have been pre-processed, i.e. cut into patches by applying a sliding window operation to the large-scale image tiles. The patches are stored in a HDF5 image database and fed into the employed networks

##### A. Center Patch Approach

The center patch approach is inspired by the work of [22] and has been considered already during database construction. The idea of this approach is that the networks should learn to classify only the center area of a given patch and to use the surrounding context information as additional input. For this reason, we label each patch according to the majority class in the center of the patch (see Section III). The visual context is used for making more accurate decisions and is not being directly predicted. By following this approach the networks are tuned to classify the image center by using additional information from the neighborhood. This is reasonable since some objects are more frequently surrounded by certain land covers than others, e.g. gardens around residential buildings, woods and bushes next to highways and water. Additionally, context information helps to disambiguate larger objects, that cover the entire center area. For the network training this means that the entire patch is fed in as input and the label for the center region is learned, independent of the land covers present in the neighborhood.

##### B. Network Models

For the training process, two network models have been employed for land cover classification. First, the VGG Network [23], which acts as a baseline and second our own model, called SatNet-8 in the following. The VGG model is used

with its original architecture. Only minor adaptations have been made, i.e. increase image input size to 270x270, set number of output neurons in the last fully-connected layer to the number of classes in our dataset and application of the 'adam' optimizer instead of 'RMSProp'. SatNet-8 is a variation of VGG with the same input and output dimensions. For training, the same parameters and number of epochs are used to enable a fair performance comparison.

1) *VGG Network*: The VGG network is a CNN with 16 layers, i.e. 13 convolutional (conv) layers and three fully-connected (fc) layers, which have been formed to logical groups. The first group represents two conv layers and one max pooling operation. The conv layers operate with 64 filters and a filter size of 3 and ReLU as activation function. Max pooling is set to a stride of 2. In the second group the number of filters per conv layer increases to 128. The third group consists of three conv layers with 256 filters each and one max pooling operation. The fourth and the fifth group exhibit 512 filters for every conv layer. After the conv layers, two fc layers follow. These fc layers consist of 4096 neurons each and employ again ReLU for activation. Followed by a dropout function with a 50% dropout rate after the first and second fc layer. Dropout helps to avoid overfitting by randomly suppressing weight updates during learning [65]. The third and final fc layer defines the output and has been changed to a number of  $K = 8$  neurons to be compatible to our dataset. This layer uses a softmax activation function. See Table II for a overview of the architecture. Finally a logistic regression function is applied.

2) *SatNet-8*: During the development of SatNet-8 the VGG Network has been used as a reference model. Multiple structural modifications have been made and each training session has been investigated in relation to its previous performance and structural design. Our basic assumption for the optimization is that we assume satellite images to have a lower overall complexity than arbitrary (object-related or scene-related) images, which were originally used to train VGG and similar networks. Thus, we expect that the complexity of the network can be reduced without losing discriminative power. Therefore, most modifications have been made with the goal to reduce the models complexity and to keep training performance at a high level.

After evaluating several different modifications and strategies for the reduction of complexity, we decided for the following architecture. The SatNet-8 consists of only 10 conv layers (instead of 13 as in VGG) and three smaller fc layers, which have been arranged into five groups. Every group represents two conv layers and a max pooling operation, followed by a batch normalization (bn) layer. This layer acts as a regularizer for the distribution of the inputs to a given layer. It normalizes the layers input for each batch during the training process. As a result, it allows higher learning rates and improves accuracy. The first two groups of conv layers operate with 128 filters with a filter size of 3 and ReLU activation. Since the first conv layers of a network model recognize simple edge and shape information, the number of filters has been increased because

TABLE II  
COMPARISON OF ARCHITECTURE BETWEEN VGG AND SATNET-8

Layer Grp.	VGG	SatNet-8
Group 1	conv1 - 64 filters conv2 - 64 filters max pool	conv1 - 128 filters conv2 - 128 filters max pool batch normalization
Group 2	conv3 - 128 filters conv4 - 128 filters max pool	conv3 - 128 filters conv4 - 128 filters max pool batch normalization
Group 3	conv5 - 256 filters conv6 - 256 filters conv7 - 256 filters max pool	conv5 - 256 filters conv6 - 256 filters  max pool batch normalization
Group 4	conv8 - 512 filters conv9 - 512 filters conv10 - 512 filters max pool	conv7 - 256 filters conv8 - 256 filters  max pool batch normalization
Group 5	conv11 - 512 filters conv12 - 512 filters conv13 - 512 filters max pool	conv9 - 512 filters conv10 - 512 filters  max pool batch normalization
	fc1 - 4096 nodes dropout - 50% rate	fc1 - 1024 nodes dropout 50% rate
	fc2 - 4096 nodes dropout - 50% rate	fc2 - 1024 nodes dropout 50% rate
	fc3 - 8 nodes regression function	fc3 - 8 nodes regression function
	<b># parameters: 180.8 M</b>	<b># parameters: 43.8 M</b>

many object types in our dataset are characterized by simple shapes, such as lines (e.g. borders of buildings and roads). The max pooling is set to a stride of 2 and thereby always halves the input dimension. The following two groups exhibit 256 filters, which is again doubled to a total of 512 in the fifth group. The first and the second fc layers consist of only 1048 neurons with ReLU activation. Since, the number of classes in our dataset is much smaller than that used for the original VGG network, this reduction in complexity is reasonable. To avoid overfitting dropout functions with a 50% dropout rate are employed for the fc layers. The final output layer is similar to VGG Networks output layer, including the logistic regression function. The adaptations performed for the SatNet-8 network lead to a reduction of parameters by approximately a factor of four. Both networks are initialized with random weights and trained from scratch in our experiments.

### C. Implementation

We employ Tensorflow with GPU support by CUDA Toolkit (version 7.5) and cuDNN (version 5.1) for training. On top of Tensorflow, the higher-level API TFLearn<sup>3</sup> has been employed, which allows for rapid prototyping of experimental setups while still retaining full access to Tensorflows capacities. The training hardware has been an Intel Core i7-7700K with an Nvidia GeForce GTX 1080 Ti.

<sup>3</sup>A. Damien, TFLearn — TensorFlow Deep Learning Library. [Online] Available: <http://tflearn.org>. Accessed on: Aug 30, 2017

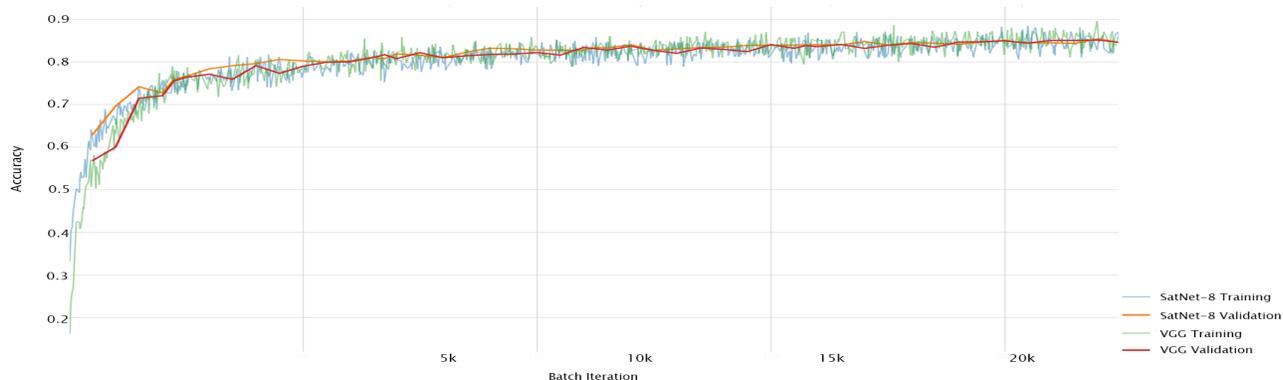


Fig. 4. Training and validation accuracy over the number of batch iterations.

## V. EXPERIMENTS & RESULTS

### A. Training Setup

Prior to training, the networks’ input layers have been adjusted to an input dimension of 270x270 px (the size of the image patches in our dataset) and three color channels. We performed different pre-processing steps on the input images. First, we zero-center the images to normalize them regarding illumination and contrast variations. We zero center each image by the measured mean across all three RGB color channels only. This normalization facilitates the networks to converge faster. Second, we apply augmentation to the training images. Augmentation adds additional variation to the training set by performing image transformations on the training patches. We employ two transforms to the patches: rotations by 0, 90, 180, and 270 degree and vertical flipping. While the training data is loaded into the system, the methods randomly rotate every image and subsequently perform vertical flipping.

The learning rate is a key parameter in training and influences the convergence of the training significantly. A too high rate can lead to an unstable training while a too low rate unnecessarily increases the required training time. We employ stochastic gradient descent (adam optimizer) for learning with an initial learning rate of 0.0001. This optimizer computes adaptive learning rates for each parameter during the training process[66] and yields good and stable results in practice. To measure classification loss, we utilize the categorical cross-entropy loss function.

To monitor the training process, we have employed a validation set. This set has contained 10% of all training samples and has been chosen randomly thereof. These samples serve the network to validate its current performance after a training iteration and are never used for training itself. The training batch size has been adjusted to 32 images. A value of 32 represents a good performance trade-off for the training hardware employed in our experiments. We train both the networks from scratch for a number of 10 epochs.

Figure 4 visualizes the learning performance of both networks over the entire training. In direct comparison, both

network models reach a similar performance level after 10 epochs. Interestingly, however, SatNet-8 faster learns, especially in early epochs. The VGG takes longer and more iterations to reach the same accuracy as the SatNet-8. Also the SatNet-8 seems to be more robust on the validation set. The main reason for the faster learning of SatNet-8 is the significantly lower number of parameters. This is also reflected by the training times. The training of the VGG network has taken 122.6 hours, whereas the SatNet-8 finished training after approx. 79.6 hours. As shown in Figure 4, both networks have potentially not reached their best performance capabilities after 10 epochs as the validation accuracy continuously increases which further shows that the network does not overfit on the data. Anyway, due to the total amount of training time, we stopped training after epoch 10. The final result is an overall training accuracy of VGG of 86.55% and SatNet-8 of 84.98%. The validation accuracy of VGG is 84.51% and that of SatNet-8 is 84.48%. There is no significant difference between the performance measures, except for the significantly lower training time of SatNet-8.

### B. Classification Results

After both networks have been trained on our dataset, we have evaluated them on the so far unseen test set. Therefore, every test image patch has been passed through the network and the most probable class from the eight possible classes has been taken as the final prediction for the patch. To assess the performance, we have computed the overall accuracy as well as a confusion matrix. The confusion matrix provides insight into the quality of classification as it displays mis-classifications between all categories and thereby reveals the networks weaknesses and strengths. Figure 5 and Figure 6 present the confusion matrices with the percentages of correctly and incorrectly classified test image patches. The top row indicates the predicted categories (pc) and the left row provides the true label (tl). The diagonal contains the achieved classification accuracy in percent (in bold letters). The off-diagonal values represent mis-classifications. The column ‘false positive’ contains the percentage of false predictions between the two building-related categories (residential building

t\pc	residential building	non-residential building	highway	federal highway	local road	grassland	water	forest	false positive	% mis-classification
residential building	<b>74</b>	<b>23.7</b>	0.05	0.05	1.45	0.25	0.45	0.05	23.7	26
non-residential building	<b>18.6</b>	<b>75.4</b>	0.35	0.3	2.15	0.15	2.9	0.15	18.6	24.6
highway	0	1.2	<b>91.05</b>	<b>6.3</b>	0.7	0.05	0.6	0.1	7	8.95
federal highway	0.4	1.6	6.8	<b>73.35</b>	<b>14.7</b>	0.5	1.95	0.7	21.5	26.65
local road	2	6.25	1	<b>9.45</b>	<b>75.9</b>	1.45	3.25	0.7	10.45	24.1
grassland	0.55	0.2	0	0.35	<b>1.75</b>	<b>94.75</b>	1.4	1		5.25
water	0.05	0.45	0	0.1	0.5	<b>1.05</b>	<b>95.95</b>	1.9		4.05
forest	0	0	0.05	0.15	0.35	0.3	<b>1.55</b>	<b>97.6</b>		2.4

Fig. 5. Confusion Matrix of the VGG Network. Values in percentage terms. 'false positive' describes percentage of false predictions between related object classes. '% mis-classification' is the sum of all wrongly predicted images.

t\pc	residential building	non-residential building	highway	federal highway	local road	grassland	water	forest	false positive	% mis-classification
residential building	<b>81.7</b>	<b>15.65</b>	0.05	0.1	1.65	0.15	0.65	0.05	15.65	18.3
non-residential building	<b>26</b>	<b>69.05</b>	0.4	0.15	2.35	0.3	1.45	0.3	26	30.95
highway	0	0.6	<b>92.65</b>	<b>5.1</b>	1.1	0	0.45	0.1	6.2	7.35
federal highway	0.5	0.95	4.95	<b>61.85</b>	<b>29.45</b>	0.85	0.7	0.75	34.35	38.15
local road	2.55	3.9	0.25	5.25	<b>83.15</b>	1.6	2.15	1.15	5.5	16.85
grassland	0.45	0.2	0	0.15	1.15	<b>95.75</b>	1	<b>1.3</b>		4.25
water	0.15	0.45	0.05	0.5	0.8	0.85	<b>94.75</b>	<b>2.45</b>		5.25
forest	0	0	0.05	0.05	0.25	0.1	<b>0.6</b>	<b>98.95</b>		1.05

Fig. 6. Confusion Matrix of the SatNet-8. Values in percentage terms. 'false positive' describes percentage of false predictions between related object classes. '% mis-classification' is the sum of all wrongly predicted images.

and non-residential building, first two rows) and the percentage of false predictions among the three street categories (highway, federal highway, and local road, rows 3-5). Column 'percentage of mis-classifications' sums up the percentages of wrongly predicted images per class. The VGG Network achieved its best results on the classes grassland, water, and forest. With an accuracy of 94.75% to 97.6%. VGG is capable of predicting 1,895 / 2,000 grassland samples, 1,918 / 2,000 water samples, and 1,952 / 2,000 forest samples correctly. When it comes to the prediction of street types, the category of highway achieved 91.05% accuracy, where only some samples tend to be mis-classified as federal highway. Since in some areas these two street types look quiet similar, these errors are comprehensible. A similar pattern are observed for federal highway and local road. The building-related classes can be detected with a lower accuracy of 74% and 75.4% for residential building and non-residential building, respectively. A total amount of 1,480 / 2,000 and 1,508 / 2,000 samples have been correctly allocated to their corresponding class. Mis-classifications of buildings, however, mostly occur between the two building categories. If we join both categories, an overall accuracy for building detection of approximately 96% would be reached.

The SatNet-8 has achieved similar results to the VGG Network with slightly different prediction accuracies. The biggest difference is the class of federal highway, which is only predicted correctly in 61.85% of all cases, whereas the VGG

TABLE III  
PERFORMANCE COMPARISON BETWEEN VGG AND SATNET-8

	VGG	SatNet-8
average accuracy	84.75%	84.73%
building detection	95.85%	96.2%
building classification	74.7%	75.375%
road detection	93.08%	94.58%
road classification	80.1%	79.21%

Network achieved 73.35%. However, the SatNet-8 performs notable better in the categories of residential building 81.7% and local road 83.15%. The overall prediction accuracy and predicted values of related object groups reveal that both networks perform similar and at a very high performance level, see Table III.

For location quality assessment the building-related categories have an especially high relevance. Thus, we want to compare the performance level of our building type classification (see Table III) to related works. An approach for building classification is presented in [53]. The authors classify buildings from Light Detection and Ranging (LiDAR) data [67] and consider three building types: single-family houses, multiple-family houses, and non-residential buildings. A overall accuracy of 70% is achieved for all three classes which is slightly lower than our accuracy of 74.7% to 75.4%. This may on the one hand be due to the larger number of classes and thus the more complex tasks but on the other hand LIDAR data provides 3D information which is not available for our method. Thus, the performance level achieved by our method can be considered similar. Another approach for building classification is presented in [52]. The authors investigate the classification of buildings into three main classes: Residential/Small Buildings, Apartment Buildings, and Industrial and Factory Building by a Random Forest (RF) classifier and report an average F1 score of 69.56% over all three classes. Although F1 score cannot be directly compared to classification accuracy, it shows that the performance level is similar.

### C. Detection of Erroneous GIS Data

As mentioned already in Section III, we have observed partly poor accuracy of the OGD polygons with the visual content in the satellite images, which has led to some serious complications during dataset generation. The detection of wrong OGD polygons could be automated by applying a reliable land cover classification or segmentation algorithm to satellite image and by finding inconsistencies between the prediction and the polygons. Thereby, mis-aligned polygons of buildings, roads, rivers, lakes, forest, and other object categories could be identified and in a further step corrected (semi-)automatically.

Figure 7 provides an example image with the ground-truth information derived from the available OGD polygons. It can be observed that a number of buildings are missing in the ground-truth which is maybe due to the fact that the polygons



Fig. 7. (a) Original satellite image from the Tiris Database; (b) Ground-truth derived from the original GIS Information (OGD polygons); (c) Predicted patch-wise labels via SatNet-8. (d) Color legend for land cover classes. (The yellow boundaries indicate buildings and land cover objects, which have been detected correctly by the SatNet-8, but have not been displayed accurately in the OGD GT.); Credit: Land Tirol - data.tirol.gv.at

TABLE IV  
 HEDONIC PRICE MODEL: FOR CONVENIENCE WE SHOW ONLY PARAMETERS FROM THE SATNET-8 PREDICTIONS. COLUMN TWO SHOWS THE EFFECT OF THE COEFFICIENTS ON THE LOGARITHMUS OF THE PRICE.  $R^2$  IS THE ADJUSTED COEFFICIENT OF DETERMINATION AND RMSE MEANS THE ROOT-MEAN-SQUARE ERROR

	Hedonic Price Model
Residential Buildings	0.004*
Non-Residential Buildings	0.005
Highway	-0.012**
Federal Highway	-0.007**
Local Road	-0.001
Water	0.001
Grassland	0.002***
Forest	0.002**
Adj. $R^2$	0.66
Num. observations	2739
RMSE	0.34

\*\*\* $p < 0.001$ , \*\* $p < 0.01$ , \* $p < 0.05$

are rather old. In Figure 7(c) the patch-wise prediction of the SatNet-8 is depicted. We can see that several buildings that were not present in the ground-truth were correctly detected (highlighted in yellow boundaries). We thus, see a great potential for the automatic detection of inaccurate GIS data by automated land cover classification.

#### D. Hedonic Pricing

Location quality and livability are usually approximated by the price or price-level of a location. This means that by predicting the price of a location, conclusions about the location quality can be drawn. A popular approach for the modeling of prices of real estate is hedonic pricing (which we briefly described in the introduction). We evaluated the utility of our land cover classifications for hedonic price modeling by feeding parameters, derived from the SatNet-8 predictions, to the hedonic (regression) model as additional inputs. Therefore, we used 2739 valuations of residential buildings, as we focus only on the location and neighborhood characteristics. In our model, for the location variables, we used only the municipalities in Tyrol, Austria. Additional to this information, we used the information from our eight land cover clusters. Therefore, we calculated the share (in terms

of covered area) of each category within a 100 meter radius from each residential building. The result of the regression is displayed in Table IV. For convenience we cut out the location coefficients. All municipality dummies are significant and reflect the expected magnitude, but are not in the focus of this paper. The categories from the SatNet-8 reflect the expected magnitude and most are statistically significant. For example the category grassland shows that a higher portion of grassland results in a higher price. On the other side, the presence of highway or federal highway in the near neighborhood indicates a negative impact which is related to a decline in price. This is in line with real estate theory. Our investigation provides first insights that confirm a positive contribution of the automatically extracted visual parameters and demonstrates that this novel type of modeling location quality has promising potential.

## VI. CONCLUSION

In this paper, we presented a first approach towards the automated assessment of location quality from satellite images. We adapted the SatNet-8 to predict different land covers and objects from our novel dataset. There out, we have fed information derived from the recognized land covers into a regression-based price model which acts as a proxy for the assessment of location quality. Our results show that (i) land cover classification can be performed with high accuracy and demonstrate that automatic classification could further be used in the future for the detection of mis-aligned and erroneous GIS data; (ii) our adapted SatNet-8 reaches state-of-the-art performance in much less training time compared to the reference VGG Network; (iii) the automatically extracted visual information improves the prediction of real estate prices and thereby shows clear potential for the description of location quality. Future work will include the extension of the proposed approach to larger areas including urban areas like cities and additional land covers.

## ACKNOWLEDGMENT

This work was supported by the Austrian Research Promotion Agency (FFG), Project No. 855784 and Project No. 856333.

## REFERENCES

- [1] C. Berlanga-Robles and A. Ruiz-Luna, "Land use mapping and change detection in the coastal zone of northwest Mexico using remote sensing techniques," vol. 18, pp. 514–522, 06 2002.
- [2] Z. Wang and L. Liu, "Assessment of coarse-resolution land cover products using CASI hyperspectral data in an arid zone in northwestern China," *Remote Sensing*, vol. 6, no. 4, pp. 2864–2883, 2014. [Online]. Available: <http://www.mdpi.com/2072-4292/6/4/2864>
- [3] A. T. Hudak and C. A. Wessman, "Textural analysis of historical aerial photography to characterize woody plant encroachment in South African savanna," *Remote Sensing of Environment*, vol. 66, no. 3, pp. 317–330, 1998.
- [4] X. Li and A. G. O. Yeh, "Principal component analysis of stacked multi-temporal images for the monitoring of rapid urban expansion in the Pearl River Delta," vol. 19, pp. 1501–1518, 05 1998.
- [5] H. Chang and W. S. Yoon, "Improving the classification of Landsat data using standardized principal components analysis," *KSCIE Journal of Civil Engineering*, vol. 7, no. 4, pp. 469–474, Jul 2003. [Online]. Available: <https://doi.org/10.1007/BF02895842>
- [6] H. Long, X. Wu, W. Wang, and G. Dong, "Analysis of urban-rural land-use change during 1995–2006 and its policy dimensional driving forces in Chongqing, China," *Sensors*, vol. 8, no. 2, pp. 681–699, 2008. [Online]. Available: <http://www.mdpi.com/1424-8220/8/2/681>
- [7] P. K. Mallupattu and J. R. S. Reddy, "Analysis of land use/land cover changes using remote sensing data and GIS at an urban area, Tirupati, India," vol. 2013, p. 268623, 05 2013.
- [8] J. Han, D. Zhang, G. Cheng, L. Guo, and J. Ren, "Object detection in optical remote sensing images based on weakly supervised learning and high-level feature learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 6, pp. 3325–3337, June 2015.
- [9] G. Cheng and J. Han, "A survey on object detection in optical remote sensing images," *ISPRS Journal of Photogrammetry & Remote Sensing*, 2016.
- [10] D. Lu, Q. Weng, E. Moran, G. Li, and S. Hetrick, *Remote sensing image classification*. CRC Press/Taylor and Francis: Boca Raton, FL, USA, 2011.
- [11] C. Homer, J. Dewitz, L. Yang, S. Jin, P. Danielson, G. Xian, J. Coulston, N. Herold, J. Wickham, and K. Megown, "Completion of the 2011 national land cover database for the conterminous United States—representing a decade of land cover change information," *Photogrammetric Engineering & Remote Sensing*, vol. 81, no. 5, pp. 345–354, 2015.
- [12] G.-S. Xia, J. Hu, F. Hu, B. Shi, X. Bai, Y. Zhong, L. Zhang, and X. Lu, "Aid: A benchmark data set for performance evaluation of aerial scene classification," *IEEE Transactions on Geoscience and Remote Sensing*, 2017.
- [13] Y. LeCun, Y. Bengio *et al.*, "Convolutional networks for images, speech, and time series," *The handbook of brain theory and neural networks*, vol. 3361, no. 10, p. 1995, 1995.
- [14] T. Ishii, R. Nakamura, H. Nakada, Y. Mochizuki, and H. Ishikawa, "Surface object recognition with CNN and SVM in Landsat 8 images," in *Proceedings of the 14th IAPR International Conference on Machine Vision Applications, MVA 2015*, 2015.
- [15] P. N. Druzhkov and V. D. Kustikova, "A survey of deep learning methods and software tools for image classification and object detection," *Pattern Recognition and Image Analysis*, vol. 26, no. 1, pp. 9–15, 2016. [Online]. Available: <http://link.springer.com/10.1134/S1054661816010065>
- [16] D. Tuia, C. Persello, and L. Bruzzone, "Domain adaptation for the classification of remote sensing data: An overview of recent advances," 2016.
- [17] G. Cheng, P. Zhou, and J. Han, "Learning Rotation-Invariant Convolutional Neural Networks for Object Detection in VHR Optical Remote Sensing Images," *IEEE Transactions on Geoscience and Remote Sensing*, 2016.
- [18] S. Ahmed, M. Liwicki, M. Weber, and A. Dengel, "Improved Automatic Analysis of Architectural Floor Plans," in *2011 International Conference on Document Analysis and Recognition*. IEEE, Sep 2011, pp. 864–869. [Online]. Available: <http://ieeexplore.ieee.org/articleDetails.jsp?arnumber=6065434>
- [19] N. Jean, M. Burke, M. Xie, W. M. Davis, D. B. Lobell, and S. Ermon, "Combining satellite imagery and machine learning to predict poverty," *Science*, vol. 353, no. 6301, 2016. [Online]. Available: <https://pdfs.semanticscholar.org/1b3a/c4b4187a3dbc9373869e7774b1dcf3f748d2.pdf>
- [20] L. J. Zhao, P. Tang, and L. Z. Huo, "Land-use scene classification using a concentric circle-structured multiscale bag-of-visual-words model," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 12, 2014.
- [21] Y. Cao, X. Niu, and Y. Dou, "Region-based convolutional neural networks for object detection in very high resolution remote sensing images," in *2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery, ICNC-FSKD 2016*, 2016.
- [22] S. Saito and Y. Aoki, "Building and road detection from large aerial imagery," vol. 9405, 02 2015.
- [23] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2014. [Online]. Available: <http://arxiv.org/abs/1409.1556>
- [24] W.-C. Liao and X. Wang, "Hedonic house prices and spatial quantile regression," *Journal of Housing Economics*, vol. 21, no. 1, pp. 16–27, 2012.
- [25] A. T. Court, "Hedonic Price Indexes with Automotive Examples," in *The Dynamics of Automobile Demand*, C. F. Roos, Ed. New York: General Motors, 1939, pp. 99–117.
- [26] G. Kumar and P. K. Bhatia, "A detailed review of feature extraction in image processing systems," in *2014 Fourth International Conference on Advanced Computing Communication Technologies*, Feb 2014, pp. 5–12.
- [27] J. A. dos Santos, O. A. B. Penatti, and R. da Silva Torres, "Evaluating the potential of texture and color descriptors for remote sensing image retrieval and classification," in *VISAPP (2)*, 2010, pp. 203–208.
- [28] Y.-W. Seo, C. Urmson, and D. Wettergreen, "Exploiting Publicly Available Cartographic Resources for Aerial Image Analysis."
- [29] J. Sivic and A. Zisserman, "Video google: A text retrieval approach to object matching in videos," in *null*. IEEE, 2003, p. 1470.
- [30] K. Nogueira, O. A. B. Penatti, and J. A. Dos Santos, "Towards Better Exploiting Convolutional Neural Networks for Remote Sensing Scene Classification," 2016.
- [31] O. A. B. Penatti, K. Nogueira, and J. A. dos Santos, "Do deep features generalize from everyday objects to remote sensing and aerial scenes domains?" in *2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, June 2015, pp. 44–51.
- [32] K. Nogueira, W. O. Miranda, and J. A. D. Santos, "Improving spatial feature representation from aerial scenes by using convolutional networks," in *2015 28th SIBGRAPI Conference on Graphics, Patterns and Images*, Aug 2015, pp. 289–296.
- [33] J. Zhao and S. You, "Road network extraction from airborne lidar data using scene context," in *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, June 2012, pp. 9–16.
- [34] J. P. Cohen, W. Ding, C. Kuhlman, A. Chen, and L. Di, "Rapid building detection using machine learning," *CoRR*, vol. abs/1603.04392, 2016. [Online]. Available: <http://arxiv.org/abs/1603.04392>
- [35] Y. E. Merabet, C. Meurie, Y. Ruichek, A. Sbihi, and R. Touahni, "Building roof segmentation from aerial images using a line-and region-based watershed segmentation technique," in *Sensors*, 2015.
- [36] A. O. Ok, C. Senaras, and B. Yuksel, "Automated detection of arbitrarily shaped buildings in complex environments from monocular vhr optical satellite imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 3, pp. 1701–1717, March 2013.
- [37] F. Dornaika, A. Moujahid, Y. E. Merabet, and Y. Ruichek, "Building detection from orthophotos using a machine learning approach: An empirical study on image segmentation and descriptors," *Expert Systems with Applications*, vol. 58, pp. 130 – 142, 2016. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0957417416301154>
- [38] J. D. Wegner, J. A. Montoya-Zegarra, and K. Schindler, "A higher-order crf model for road network extraction," in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, June 2013, pp. 1698–1705.
- [39] J. A. Montoya-Zegarra, J. D. Wegner, L. Ladický, and K. Schindler, "Mind the gap: modeling local and global context in (road) networks," in *German Conference on Pattern Recognition*. Springer, 2014, pp. 212–223.
- [40] C. Poullis, "Tensor-cuts: A simultaneous multi-type feature extractor and classifier and its application to road extraction from satellite images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 95, pp. 93 – 108, 2014. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0924271614001543>

- [41] F. Hu, G.-S. Xia, J. Hu, and L. Zhang, "Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery," *Remote Sensing*, vol. 7, no. 11, pp. 14 680–14 707, 2015. [Online]. Available: <http://www.mdpi.com/2072-4292/7/11/14680>
- [42] S. Basu, S. Ganguly, S. Mukhopadhyay, R. DiBiano, M. Karki, and R. R. Nemani, "DeepSAT - A learning framework for satellite imagery," *CoRR*, vol. abs/1509.03602, 2015. [Online]. Available: <http://arxiv.org/abs/1509.03602>
- [43] M. Castelluccio, G. Poggi, C. Sansone, and L. Verdoliva, "Land use classification in remote sensing images by convolutional neural networks," *CoRR*, vol. abs/1508.00092, 2015. [Online]. Available: <http://arxiv.org/abs/1508.00092>
- [44] M. Längkvist, A. Kiselev, M. Alirezaie, and A. Loutfi, "Classification and segmentation of satellite orthoimagery using convolutional neural networks," vol. 8, p. 329, 04 2016.
- [45] V. Mnih and G. E. Hinton, *Learning to Detect Roads in High-Resolution Aerial Images*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 210–223.
- [46] J. Wang, J. Song, M. Chen, and Z. Yang, "Road network extraction: A neural-dynamic framework based on deep learning and a finite state machine," *International Journal of Remote Sensing*, vol. 36, no. 12, pp. 3144–3169, 2015.
- [47] A. Lagrange and B. Le Saux, "Convolutional neural networks for semantic labeling," Tech. Rep., Onera–The French Aerospace Lab, Tech. Rep., 2015.
- [48] A. Lagrange, B. L. Saux, A. Beaupère, A. Boulch, A. Chan-Hon-Tong, S. Herbin, H. Randrianarivo, and M. Ferecatu, "Benchmarking classification of earth-observation data: From learning explicit features to convolutional networks," in *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, July 2015, pp. 4173–4176.
- [49] S. Paisitkriangkrai, J. Sherrah, P. Janney, and A. V.-D. Hengel, "Effective semantic pixel labelling with convolutional networks and conditional random fields," in *2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, June 2015, pp. 36–43.
- [50] O. Firat, G. Can, and F. T. Y. Vural, "Representation learning for contextual object and region detection in remote sensing," in *2014 22nd International Conference on Pattern Recognition*, Aug 2014, pp. 3708–3713.
- [51] Q. Lv, Y. Dou, X. Niu, J. Xu, and B. Li, "Classification of land cover based on deep belief networks using polarimetric radarsat-2 data," in *2014 IEEE Geoscience and Remote Sensing Symposium*, July 2014, pp. 4679–4682.
- [52] M. Belgiu, I. Tomljenovic, T. J. Lampoltshammer, T. Blaschke, and B. Höfle, "Ontology-based classification of building types detected from airborne laser scanning data," *Remote Sensing*, vol. 6, no. 2, pp. 1347–1366, 2014. [Online]. Available: <http://www.mdpi.com/2072-4292/6/2/1347>
- [53] Z. Lu, J. Im, J. Rhee, and M. Hodgson, "Building type classification using spatial and landscape attributes derived from lidar remote sensing data," *Landscape and Urban Planning*, vol. 130, pp. 134 – 148, 2014. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0169204614001601>
- [54] N. Srećković, N. Lukač, B. Žalik, and G. Štumberger, "Determining roof surfaces suitable for the installation of PV (photovoltaic) systems, based on LiDAR (Light Detection And Ranging) data, pyranometer measurements, and distribution network configuration," *Energy*, vol. 96, pp. 404–414, feb 2016. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0360544215017235>
- [55] S. Szabó, P. Enyedi, M. Horváth, Z. Kovács, P. Burai, T. Csoknyai, and G. Szabó, "Automated registration of potential locations for solar energy production with Light Detection And Ranging (LiDAR) and small format photogrammetry," *Journal of Cleaner Production*, vol. 112, pp. 3820–3829, jan 2016. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0959652615010483>
- [56] M. Xie, N. Jean, M. Burke, D. Lobell, and S. Ermon, "Transfer Learning from Deep Features for Remote Sensing and Poverty Mapping," p. 16, 2015. [Online]. Available: <http://arxiv.org/abs/1510.00098>
- [57] G. S. Xia, J. Hu, F. Hu, B. Shi, X. Bai, Y. Zhong, L. Zhang, and X. Lu, "AID: A benchmark data set for performance evaluation of aerial scene classification," *IEEE Transactions on Geoscience and Remote Sensing*, 2017. [Online]. Available: <https://arxiv.org/pdf/1608.05167.pdf>
- [58] Q. Zou, L. Ni, T. Zhang, and Q. Wang, "Deep learning based feature selection for remote sensing scene classification," vol. 12, pp. 1–5, 11 2015.
- [59] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, ser. GIS '10. New York, NY, USA: ACM, 2010, pp. 270–279. [Online]. Available: <http://doi.acm.org/10.1145/1869790.1869829>
- [60] G. Cheng, J. Han, and X. Lu, "Remote sensing image scene classification: Benchmark and state of the art," *Proceedings of the IEEE*, vol. PP, no. 99, pp. 1–19, 2017.
- [61] O. A. B. Penatti, K. Nogueira, and J. A. Dos Santos, "Do Deep Features Generalize from Everyday Objects to Remote Sensing and Aerial Scenes Domains?"
- [62] K. Nogueira, W. O. Miranda, and J. A. D. Santos, "Improving Spatial Feature Representation from Aerial Scenes by Using Convolutional Networks," in *Brazilian Symposium of Computer Graphic and Image Processing*, 2015.
- [63] M. Castelluccio, G. Poggi, C. Sansone, and L. Verdoliva, "Land Use Classification in Remote Sensing Images by Convolutional Neural Networks," *Arxiv preprint arXiv:1508.00092*, aug 2015. [Online]. Available: <http://arxiv.org/abs/1508.00092>
- [64] Amt der Tiroler Landesregierung. Flächennutzung. Accessed: 2017-08-29. [Online]. Available: <https://www.tirol.gv.at/statistik-budget/statistik/flaechennutzung/>
- [65] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *Journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014.
- [66] S. Ruder, "An overview of gradient descent optimization algorithms," *CoRR*, vol. abs/1609.04747, 2016. [Online]. Available: <http://arxiv.org/abs/1609.04747>
- [67] S. Reutebuch, H. Andersen, and R. Mcgaughey, "Light detection and ranging (lidar): An emerging tool for multiple resource inventory," vol. 103, pp. 286–292, 09 2005.