Applied Intelligence (2021) 51:8088–8105
https://doi.org/10.1007/s10489-021-02268-8

# Where do university graduates live? – A computer vision approach using satellite images

David Koch[1] · Miroslav Despotovic[1] · Simon Thaler[1] · Matthias Zeppelzauer[2]

Accepted: 9 February 2021 / Published online: 25 March 2021


**Abstract**
In this article, we examine to what extent the settlement of university graduates can be derived from satellite images. We apply a convolutional neural network (CNN) to grid images of a city and predict five density classes of university graduates at a micro level (250 m × 250 m grid size). The CNN reaches an accuracy rate of 40.5% (random approach: 20%). Furthermore, the accuracy increases to 78.3% when considering a one-class deviation compared to the true class. We also examine the predictability of inhabited and uninhabited grid cells, where we achieve an accuracy of 95.3% using the same CNN. From this, we conclude that there is information that correlates with graduate density that can be derived by analysing only satellite images. The findings show the high potential of computer vision for urban and regional economics. Particularly in data-poor regions, the approach utilised facilitates comparative analytics and provides a possible solution for the modifiable aerial unit (MAU) problem. The MAU problem is a statistical bias that can influence the results of a spatial data analysis of point-estimate data that is aggregated in districts of different shapes and sizes, distorting the results.

**Keywords** Satellite images · Demographic structure · Machine learning · Urban areas

## 1 Introduction

The prediction of urban areas and the corresponding growth [4, 9] utilising machine learning [1, 12, 22, 23] is a vital topic in built environments. This work, as a first pilot study, focuses on the prediction of demographic structures, more specifically the distribution of university graduates, in urban areas, as they can be used as a proxy for the wealth distribution of cities as well as for the attractiveness of neighbourhoods [6, 13, 20]. Satellite images can assist in deriving visual characteristics—whether a residential area is generally attractive—without the assistance of any other (demographic) statistical data, distance measures, points of interest (POIs) or any other relevant GIS data. Such visual characteristics include the material of the buildings in a residential area and the immediate neighbourhood or vegetation in the area, as the presence of vegetation, street trees, parks, forests, open spaces, and bodies of water usually enhances residential areas [24, 30]. We assume that the presence of the aforementioned types of land covers are indicators of a wealthy neighbourhood and thus also correlate with the settlement of university graduates.

Satellite images and the capabilities of today's computer vision techniques, in combination with machine learning, play an increasingly important role in economic evaluation. Some examples of statistical economic variables that can be predicted using computer vision are the gross domestic product (GDP) [5], economic growth [17] and poverty [51]]; computer vision can also be used in detecting, estimating and monitoring socioeconomic dynamics such as urbanisation, population and economic activity [3] using nighttime luminosity. The focus of our research is on predicting the university graduate ratio (GR) in an urban population (the city of Vienna, Austria) by the use of satellite images depicting a 250 m × 250 m area, which is the smallest population grid data available in Austria. Figure 1 shows example satellite images from our dataset with different densities of university graduates and the related visual characteristics.

Our approach enables us to constantly monitor economic data in a populated area and to circumvent the modifiable areal unit (MAU) problem, which arises due to administra-

✉ Simon Thaler
simon.thaler@fh-kufstein.ac.at

[1] Fachhochschule Kufstein Tirol Bildungs GmbH, Andreas Hofer-Straße 7, Kufstein, 6330, Austria

[2] Fachhochschule St. Pölten GmbH, Matthias Corvinus-Straße 15, St. Pölten, 3100, Austria

(a) (b) (c) (d)

**Fig. 1** Where do university graduates live? Four example areas from our dataset are represented by high-resolution satellite images (4285 × 4285 pixels). From left to right: Image (**a**), with a high graduate ratio of 51.2%, in the city centre of Vienna. Image (**b**), with a high graduate ratio of 32.4%, in the outskirts of the city. Image (**c**), with a graduate ratio of 10.0%, near an industrial zone. Image (**d**), which displays panel buildings with a graduate ratio of 3.3%

tive borders and other spatial limitations [41, 47]. The MAU problem arises when spatial aggregated data are analysed, as the size and scale of the aggregation district can lead to statistical bias. The MAU problem is resolved when the trained neural network can be applied to arbitrary locations, independent of the size and scale of any predefined grid. Part of the increased accuracy needed to circumvent MAU problem distortions is the use of the smallest possible spatial area when predicting sociodemographic data [50]. For a review of the MAU problem and suggested solutions see [8, 40].

Additionally, most evaluations conducted in the field are executed at a highly aggregated level (e.g., 1 km × 1 km satellite image grid cells [21]) to predict economic variables. Our approach differs in the sense that we focus on a much more fine-grained analysis of the images, which can enable a more precise economic analysis in the future. This small-scale analysis together with the free positioning of the satellite image grid can address problems and provide possible solutions (e.g., varying the shape and size of the investigated area) [8, 40, 52]. Additionally, we do not make any assumptions, and our work is fully data driven. This means that characteristic visual patterns are discovered and learned autonomously and do not have to be predefined.

The overall aim of this article is to combine methodology from the disciplines of urban economics and computer vision to realise innovative services for urban analysis. We enable the analysis of an important economic variable, namely the settlement of university graduates, with a low-resource technology, assisting in urban policy analysis. An advantage of this approach is that a properly trained computer vision model can be applied to any other raster of satellite images. This makes the proposed methodology applicable to any size and position of a corresponding raster and therefore relaxes the dependency from the existance of suitable GIS data. As a computer vision-based prediction with a previously trained model is a time-saving procedure, the efficiency of urban planning and development processes can be increased.

## 2 Related Work

Computer vision has gained great importance in recent years due to the increasing availability of geo-data [49]. It includes various techniques for detecting and monitoring the physical properties of an area by photographing or measuring its reflected and emitted radiation at a large distance from the Earth's surface [43]. Computer vision in this area has thus far been used for land-use classification and segmentation (e.g., detecting and segmenting meadows, forests, and roads) [11, 35, 38, 53], building (footprint) detection [34, 42], building detection and classification (e.g., differentiating residential, commercial, single-family houses, and apartments) [27, 32, 45], building roof analysis (e.g., to estimate the potential for solar power systems) [10, 33] and 3D city modelling [15, 25]. For a comprehensive review of computer vision and neural network applications in the real estate sector, urban systems and built environments, see [26] and [14]. We contribute to this research in the sense that we establish a novel, visually grounded approach to predict the spatial distribution of residents according to their education level from satellite images.

The relevance of satellite images in economic prediction was assessed in [7], where the authors provided proof of a relationship between economic development and deforestation utilising satellite images of forest cover. They found that the key determinant of per capita income differences among countries is determined by the relation between forest cover and GDP.

The authors of [21] utilised daytime satellite images to predict socioeconomic variables for consumption expenditure and asset wealth by employing a CNN. Trained separately for each country, the model explains 37 to 55% of the variation in the average household consumption of the four countries and 55 to 75% of the household asset wealth

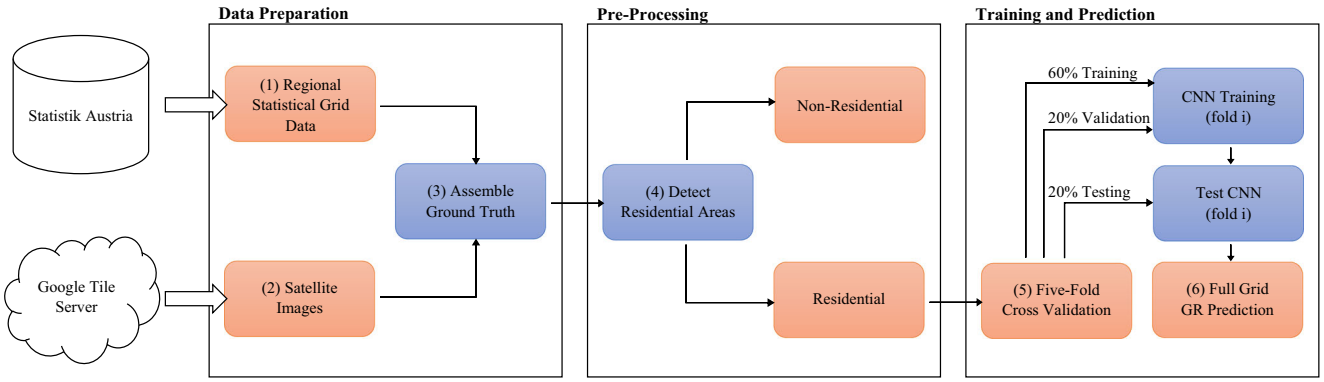



**Fig. 2** Flow chart summarising the proposed method. Our approach consists of three main parts: "Data Preparation" (alignment of meta-data and satellite images), "Pre-Processing" (detection of inhabited areas) and "Training and Prediction" (prediction of graduate ratios)

variation across all five compared countries. For their investigation, the researchers used 1 km × 1 km daytime satellite images with up to 10 km of noise in the ground-truth data to protect the survey respondents. In our work, we focus on much smaller areas for a more fine-grained analysis and employ an accurate high-resolution ground truth. Similar to their methodology, we employ transfer learning to increase the speed of learning meaningful visual features from the satellite images. Research in urban economics leveraging computer vision and machine learning does not only focus on the prediction of economic variables. The authors of [39], e.g., predicted the perceived safety of US cities by analysing street view images. They found that the visual appearance of a neighbourhood can influence the liveability for the neighbourhoods' inhabitants. This task is related to our article, as university graduates tend to agglomerate in areas with higher quality of life and thus in safer neighbourhoods [6, 13, 20].

The cost of living plays a negligible role in the location choice, in contrast to the prevalent wage levels [31]. This is in line with research that suggests that rural population growth is reduced by schooling, as highly educated individuals will migrate into urban areas, where they can expect a higher return on their education [19, 36].

The agglomeration of human capital or knowledge has been addressed in the theoretical foundations of the new economic geography, with its core-periphery model focusing on the spatial concentration and specialisation of production factors [28]. A derivation of the core-periphery model yields theoretical proof of the agglomeration of skilled workers [37], where education is among the determinants of skill.[1]

[1]In this regard the reader might bear in mind that according to [2] ], education is not equivalent to skill, as education is part of the development process that determines skill. Other factors that determine skill are the abilities and traits of a labourer. For a detailed introduction to recent economic geography, the core-periphery-model and the agglomeration of production factors, see [29].

Reference [46] follows a similar idea as our work, as they predict the socioeconomic profile of a city using satellite images and a neural network. They find proof of visual patterns that correlate with the socioeconomic classes of the inhabitants. Nonetheless, the methodology employed differs significantly from our approach, as the authors use a different social group and a more complex preparation of the ground truth; they focus their predictions on the presence of certain objects (e.g., swimming pools). Thus, our approach has broader applicability, as it is not dependent on the presence of predefined objects.

Recently, satellite images have become an increasingly popular source of data in the field of urban economics. The migration and agglomeration of production factors, in this case education or skill, is a prevalent topic in the field of economics that has been examined in different ways (e.g., [19, 36] as well as [37] on a theoretical level). Overall, the question of whether academic agglomeration is reflected visually in satellite images is currently open. We examine the agglomeration of graduates on a fine scale by analysing small grids of satellite images together with population statistics.

## 3 Methodology

For our investigation, we employ population data together with satellite images of the city of Vienna in Austria. The objective is to predict the spatial distribution of university graduates on a grid cell level using a convolutional neural network (CNN). Our approach is based on the assumption that satellite images of residential areas contain visual indicators that allow an estimation of the proportion of university graduates in a defined residential area.

An overview of our approach is depicted in Fig. 2 and described below (the numbers in the description are linked to the figure). A detailed description is given in Sections 3.1, "Data Preparation and Pre-Processing", and 3.2 "Training and Prediction".

First, we obtain the regional statistical grid data (1) for the study location. Next, we extract the grid coordinates to collect the respective satellite images (2). Then, we align the statistical grid data (population count and university graduate count per grid cell) with the satellite images. As a result, we obtain cell-based satellite images together with the computed graduate ratios that serve as the target variables (the ground truth) for our experiment (3). The ground truth is necessary to train and evaluate our model, which aims at learning the relationship between the visual modality (images) and the statistical measure (density of graduates).

In the next step, we identify grid cells with no or very little population density. This can be performed automatically by training a model for the detection of low-population-density areas or by using information from the ground truth, i.e., using a certain threshold (20 inhabitants, in this paper) to differentiate residential and non-residential cells (4). The non-residential cells are excluded from further analysis in our approach.

For the training of our density prediction network, we employ a five-fold cross validation protocol (5), where we obtain five independent predictions of graduate ratios covering the populated area of the city. After predicting all five folds, we obtain a prediction of the graduate ratios for the entire city. Combining the predictions with the ground truth, we compute confusion matrices to analyse misclassifications and to estimate the overall performance (6).

### 3.1 Data Preparation and Pre-Processing

**Demographic data** For our investigation, we use 250 m × 250 m regional statistical grid data[2], which are laid out across the entire federal territory and are made publicly available by *Statistik Austria*, the statistical office of the Republic of Austria. The grid is independent of administrative boundaries and therefore allows for a more subject-related delimitation of territories, which also solves the aforementioned MAU problem [47]. In future research, the determination of grids can be done independently on the basis of size. The statistical grid data with the corresponding satellite images are needed only for training and not for prediction. The dataset includes the grid cell coordinates with the population count as well as the count of university graduates.

[2]Statistik Austria offers a Europe-wide grid on the area-true Lambert azimuthal projection (ETRS-LAEA grid) according to the EU directive INSPIRE. A uniform European projection system is particularly advantageous for the exchange of geo-data in Europe, since the geo-data no longer has to be transformed in a time-consuming way. This also makes it easier for small-scale, cross-border presentations, evaluations and research in Europe.

**Satellite data** To obtain suitable image data, we need high-resolution satellite images for the study location. We retrieve satellite images from the Google Tile Server[3] of size 4285 × 4285 pixels (at a resolution of 1 pixel ≈ 5.8 centimetres) and resize them to 224 × 224 pixel images that match the regional statistical grid data (at a resolution of 1 pixel ≈ 111.6 centimetres). Resizing is necessary, as the employed neural network processes 224 × 224 pixel images.

**Population and graduate ratio** The statistical grid data with the matching satellite images consist of 6,632 grid cell data points. The dataset contains absolute numbers for the overall population as well as absolute numbers of university graduates for every cell. For machine learning, however, a normalised value range of the target variable is beneficial (see Fig. 3 for the distribution of the absolute and relative data). Furthermore, absolute numbers add a population size bias; thus, using proportions results in less biased results. To convert the data to relative numbers, we calculate the graduate ratio (GR) for every grid cell by dividing the absolute number of university graduates by the total population of the grid cell. This ratio facilitates the interpretation and comparability of the distribution of graduates in the investigated areas. To obtain the classes for prediction, we separate the dataset using the GR-20% percentile. Reference [46] also uses five classes in their prediction model based on a neural network. Their approach differs in the sense that the determination of classes is made according to the presence or absence of certain objects in the images, which is decided a priori by the researchers.. An equal distribution of classes is beneficial in machine learning. The result is a set of five classes, from class 1, containing the lowest graduate ratio grid cells, up to class 5, with the highest graduate ratio grid cells. In Appendix A, we show sample satellite images for all the classes, from low (class 1) to high (class 5) graduate ratios.

**Pre-filtering** The employed categorisation scheme is to some degree sensitive to changes in absolute numbers, especially in sparsely populated grid cells, where small changes in the absolute numbers can strongly impact the resulting ratios. To mitigate this sensitivity, we filter out the sparsely inhabited and completely uninhabited grid cells. For our experiments on graduate density prediction, we define a threshold of 20 inhabitants. This (i) excludes sparsely populated cells that are counterproductive for our analysis (artificially increase the accuracy) and (ii) counteracts unstable class assignments for sparsely

[3]for a detailed description see the following link: https://stackoverflow.com/questions/58846393/how-to-apply-api-key-to-google-maps-tile-server-url

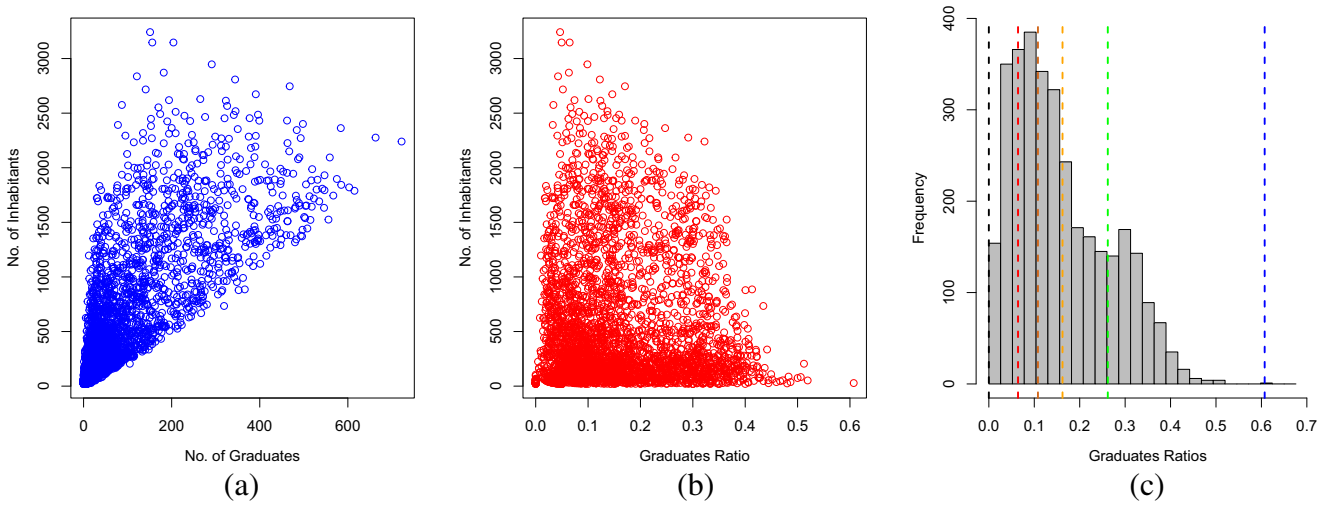



**Fig. 3** Graph (**a**) depicts the absolute graduate count per grid cell on the x-axis over the population count per grid cell on the y-axis. Graph (**b**) depicts the calculated graduate ratio per grid cell on the x-axis and the population count per grid cell on the y-axis. Graph (**c**) depicts the grid cell-based graduate ratio on the x-axis and the frequency of the corresponding ratio on the y-axis. The coloured lines depict the upper 20% percentile boundaries for the five graduate ratio classes in ascending order, with the numerical boundaries in brackets (black: null ratio [0.0%], red: class 1 [6.4%], orange: class 2 [10.8%], yellow: class 3 [16.2%], green: class 4 [26.2%], blue: class 5 [60.7%])

populated cells and thus improves the robustness of classes. After filtering, our ground truth consists of 3,314 grid cells. For the filtering of the dataset, we tested an automatic approach. For further details, see Section 4.1.

### 3.2 Training and Prediction

Once the uninhabited areas have been filtered out (via our population threshold), we train a CNN for the prediction of graduate ratio classes using satellite images. To model the relationships between the visual information from the satellite images and the five target classes, we build VGG-16 [44], which is a pre-trained CNN, and apply transfer learning to adapt it to our requirements. Prior to the selection of VGG-16 as the network model, we have evaluated a number of alternative promising CNN architectures, namely DenseNet201 [18] (a CNN that is 201 layers deep with connections between each layer and subsequent layers, preserving features in previous layers, giving the model more flexibility in multi-scale modeling) and VGG-19 [44] (a deeper version of VGG-16). We trained all networks on the graduates density data from Vienna where VGG-16 showed the best accuracy on the test set compared to VGG-19 and DenseNet201 and was thus selected for all further experiments.

**Network architecture** VGG-16 is a feed-forward neural network architecture that builds upon a stack of convolutional filters followed by several dense (fully-connected) layers (see Fig. 4 for an overview and Table 1 for details on all hyperparameters of the architecture).

The inputs to the network are three-channel RGB images of size 224 × 224 pixels (i.e. 224×224×3 tensors) covering one grid cell of 250 m × 250 m. The network consists of two major parts. The first part is a stack of convolutional layers that aims at learning a hierarchical (multi-scale) visual representation from satellite images. In each of the five convolutional layer groups, image filters are learned for different image scales. The pooling layers after each layer group reduce the resolution of the representation by half. The filters in the early layers (e.g., Conv 1-1 and Conv 1-2) represent very basic and generic small-scale image structures (usually edges). The intermediate layers (e.g., Conv 3-1 to 3-3) represent larger-scale structures (e.g., image textures). The higher layers (Conv 5-1 to 5-3) capture visual structures at the largest scale, related to buildings and building parts. The hierarchical representation stack is followed by a stack of dense layers, which can be considered a non-linear classifier. We employ two dense layers, as in the original VGG implementation, and replace the third dense layer (i.e., the output layer) by a smaller layer with five nodes, where each node corresponds to one density class. The neuron activation functions throughout the entire network are rectified linear unit (ReLU) functions of the form:

$$f(x) = max(0, x)$$

where $x$ is the current activation fed into the activation function. After the last dense layer, we position a softmax layer. The softmax layer re-scales the outputs $x_j$ of the

**Table 1** VGG-16 network architecture: The adapted network architecture for graduate density estimation with satellite images input, which is zero-center normalised, based on the VGG-16 network (138M network parameters)

| Shape: | Layer type: | Activation: | Pooling: | Dropout: |
|---|---|---|---|---|
| 224×224×3 | input | – | – | – |
| 64 3×3×3 | conv | ReLU | – | – |
| 64 3×3×64 | conv | ReLU | max pooling | – |
| 128 3×3×64 | conv | ReLU | – | – |
| 128 3×3×128 | conv | ReLU | max pooling | – |
| 256 3×3×128 | conv | ReLU | – | – |
| 256 3×3×256 | conv | ReLU | – | – |
| 256 3×3×256 | conv | ReLU | max pooling | – |
| 512 3×3×256 | conv | ReLU | – | – |
| 512 3×3×512 | conv | ReLU | – | – |
| 512 3×3×512 | conv | ReLU | max pooling | – |
| 512 3×3×512 | conv | ReLU | – | – |
| 512 3×3×512 | conv | ReLU | – | – |
| 512 3×3×512 | conv | ReLU | -max pooling | – |
| 4096 | fully connected | – | – | 50% dropout |
| 4096 | fully connected | – | – | 50% dropout |
| 5 | fully connected | – | – | – |
| | softmax | – | – | – |
| loss function: | categorical cross-entropy | | | |

Convolutional layers with stride [1 1] and padding [1 1 1 1]. 2x2 max pooling with stride [2 2] and padding [0 0 0 0]

network to obtain class probabilities (for the five density classes) that sum to one overall:

$$y_j = \frac{e^{x_j}}{\sum_{i=1}^{N} e^{x_i}},$$

where $y_j$ represents the normalised output for the respective un-scaled network output $x_j$ ($y_j > 0$) and $N$ is the number of network outputs: $x_1, ..., x_N$ ($N = 5$ in our case).

**Training** Prior to training, we normalise the input images by subtracting the average red, green and blue values from each image. As a result, all the images become zero-centred colour channels. Normalisation is recommended to accelerate the optimisation process during training (gradient descent). To estimate the model quality, we employ categorical cross-entropy loss to measure the match between the network predictions and the ground truth during training. Categorical Cross-entropy is commonly used for multi-class classification problems and defined as follows:

where $\hat{y}_i$ is the i-th scalar value in the model output, $y_i$ is the corresponding target value, and N is the number of scalar values in the model output ($N = 5$ classes in our case).

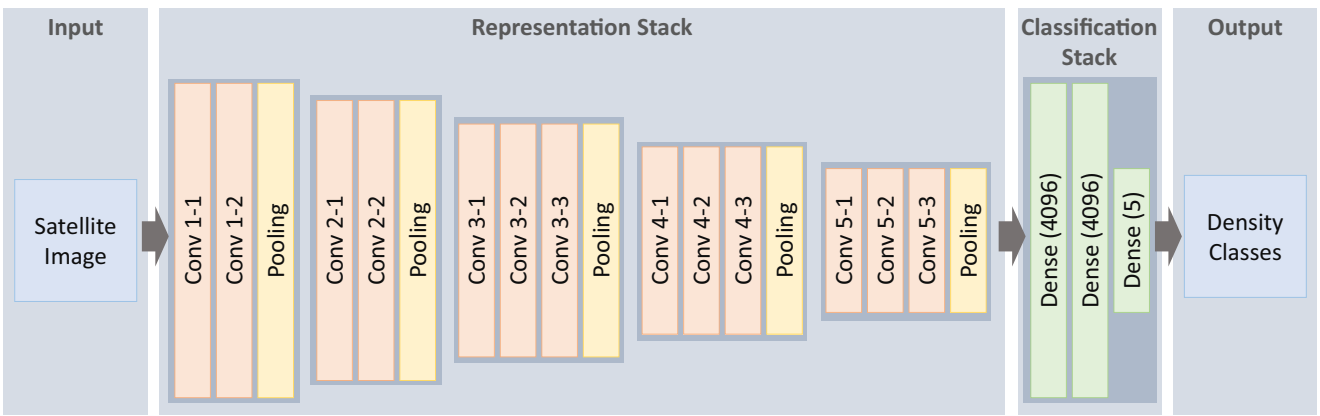



**Fig. 4** The adapted network architecture for graduate density estimation from satellite images based on the VGG-16 network

The loss function estimates how well the network outputs correspond to the desired target outputs and is used as a target function during training which is minimised.

The percentile-based classes obtained from the statistical data serve as the target variables for training. Training is performed via mini-batch gradient descent with a batch-size of 32 images. The SGD optimizer is used to minimise the loss function. To avoid over-fitting the network, we freeze the first 10 network layers during training. This means that the network weights for those layers remain unchanged. Only the higher layers are fine-tuned and adapted to the current task. For training the neural network, we use the MATLAB framework MatConvNet from [48]. The experiments were performed on a workstation with an Ubuntu 18 OS, 64 GB RAM and an NVIDIA GTX 1080Ti. Retraining (fine-tuning) is performed for 50 epochs with a learning rate of 0.0001, momentum of 0.9 and weight decay of 0.0005. We use classification accuracy as performance measure in our experiments. To monitor overfitting during training a validation set is employed (see below)

**Prediction and evaluation** For training, we employ five-fold cross-validation. The motivation for using cross-validation is to make the best use of the limited amount of data (number of grid cells) that is available for our experiment. We train five networks using five independent training partitions from the ground-truth dataset. The partitions are composed of randomly assigned datasets to avoid location dependencies. To avoid bias from different locations in the city or from similar characteristics of neighbouring cells, three parts (60%) of the available data are used for training, one part (20%) of the available data is used for validation and the fifth part (20%) of the available data is used for predicting and testing. The assignments of the parts to the training, validation and test data vary across all five iterations.

After the application of all five networks, the result is a prediction of the entire study area, which can subsequently be evaluated with the five test sets, which are composed of 20% of the ground truth each, thus yielding the entire study area and the aggregated confusion matrix in Section 4.2 "Prediction of Graduate Ratios". Since all the networks are trained independently from different data and there is no optimisation of a hyperparameter over all networks, their results can be considered independent. Thus, their joint predictions provide a reasonable performance estimate for the prediction of the target variable (the graduate ratio distribution) over the entire study area (i.e., the whole grid of Vienna). As a performance measure, we employ the accuracy rate (the portion of correctly classified grid cells), which is justified due to the balanced class sizes in the experiment.

Note that we do not use the cross-validation approach to select the model or training parameters (e.g., the network architecture, learning rate or loss function), i.e., to optimise our approach. This is important to avoid over-fitting and biased (i.e., overly optimistic) results. For all five folds, the same hyperparameters are used.

## 4 Results

Below, we first state the results for the automated differentiation of inhabited and non-inhabited areas and then present the results of the graduate ratio prediction. We conclude the result presentation with the analysis of graduates class deviations.

### 4.1 Prediction of Inhabited Grid Cells

In our analysis, we focus on inhabited areas only. In a preliminary study, we investigate whether we can automatically differentiate inhabited from uninhabited grid cells to provide automatic pre-filtering.[4] The results show that the trained CNN correctly predicts 95.3% of the inhabited and uninhabited areas in the test dataset. This shows that data pre-processing can be almost fully automated in future work.

Nevertheless, for the following experiments, we manually split the inhabited and non-inhabited areas using ground-truth information with a manually defined threshold of 20 residents per cell. The reason for enforcing this separation is to assure that the subsequent experiment is completely based on error-free data.

### 4.2 Prediction of Graduate Ratios

After filtering out all uninhabited areas, 3,313 populated grid cells remain for the subsequent analysis. In what follows, we investigate the prediction accuracy of the distribution of graduates obtained by our approach. For each fold, we compute the accuracy rate on the independent test set and identify false detections. The aggregated (summed) confusion matrix of all five cross-validation runs in Table 2 shows the correctly predicted grid cells on its main diagonal (absolute numbers and percentages of the respective classes). With 245 an overall accuracy rate (AR) of 40.5%, we are able to correctly predict twice as many grid cells as a random approach (which would yield a classification accuracy of 20% due to the five equally likely

[4] We fine-tune a pre-trained network (Resnet50 pre-trained on ImageNet [16]) for 30 epochs, a learning rate of 0.0001, a momentum of 0.9 and a weight decay of 0.0005 for the study area of Vienna. We split the data into 470 validation and 5750 training samples and evaluate the network on a set with 470 independent test images

**Table 2** Aggregated confusion matrix for Vienna

| | | True class | | | | | |
|---|---|---|---|---|---|---|---|
| | | $c_1$ | $c_2$ | $c_3$ | $c_4$ | $c_5$ | |
| Prediction | $c_1$ | **400 (60.3%)** | 228 (34.1%) | 134 (20.4%) | 72 (10.9%) | 17 (2.6%) | |
| | $c_2$ | 121 (18.3%) | **136 (20.3%)** | 109 (16.6%) | 51 (7.7%) | 19 (2.9%) | |
| | $c_3$ | 101 (15.2%) | 166 (24.8%) | **201 (30.6%)** | 153 (23.2%) | 61 (9.2%) | |
| | $c_4$ | 32 (4.8%) | 96 (14.4%) | 130 (19.8%) | **174 (26.3%)** | 134 (20.2%) | |
| | $c_5$ | 9 (1.4%) | 43 (6.4%) | 83 (12.6%) | 211 (31.9%) | **432 (65.1%)** | **AR: 40.5%** |

The main diagonal of the matrix shows the correct predictions of our classifier (CNN) summed over all five folds. The accuracy rate of the predictions is shown at the bottom right of the table (abbreviated by AR). *All correct predictions per class as well as the overall accuracy (AR) are printed in bold letters*. The percentages in brackets show the distribution of predictions for each class (which sum to one for each column)

classes), and we obtain an overall accuracy rate that is 10.5% higher than that in reference [46] (30.0% overall accuracy) which employs a similar prediction model for a sociodemographic variable. Furthermore, 78.3% of the predicted density estimates deviate by no more than one class from the true class (random approach: 52%). This indicates that the model learned visual patterns that correlate with the graduate density in the grid cells.

Comparing the prediction accuracy for the individual graduate density classes (the diagonal of Table 2), we observe that the model performs best for the lowest and highest graduate ratio classes (i.e., class 1 and class 5; the same evidence is used as in reference [46]).The weakest performance is achieved for class 2, where 34.1% (true class 2: 669 observations) of the grid cells belonging to class 2 are assigned to class 1 (compared to 20.3% or 136 grid cells of the data that are correctly predicted as class 2). This may be due to the narrow width of the corresponding class boundaries of classes 1 and 2. The upper boundary of class 1 is a graduate ratio of 6.4%, and class 2 exhibits an upper boundary of 10.8%, resulting in a margin of only 4.4 percentage points. Thus, class 2 spans a low range of ratios, which can explain the difficulties in robustly predicting the grid cells.

Next, we investigate whether there is a bias in the misclassifications towards higher or lower densities. To this end, we sum the misclassifications above the diagonal in the confusion matrix (the sum of upper diagonals, SUD) and the sum below the diagonal (the sum of lower diagonals, SLD). The similar values for the SUD, 29.5%, and SLD, 30.0%, indicate that there is no bias towards higher or lower densities.

To further analyse a potential bias, we investigate the five confusion matrices obtained for the five folds. In Table 3, the accuracy rates of the five CNN runs and the respective SUDs and SLDs are listed. The confusion matrices for all five runs can be found in Appendix B. From Table 3, we can conclude that the performance is at a comparable level across the whole study area (approx. 37-45% accuracy), and thus, the dependency on the training data selection is rather low. Looking at the accuracy rates as well as the SUDs and SLDs, we do not see significant outliers in the deviations from the aggregated results in Table 2 which indicates that our models perform equally well over all five test datasets.

After analysing the aggregated results, we examine the data on the grid level by generating a heat map that depicts the true and predicted class distributions. In Fig. 5, we show the city area partitioned into the statistical regional grid employed. Utilising the ground truth data, we are able to construct a grid map of the distribution of the true graduate ratio for the city area depicted in the image, and we plot the predicted grid cells of our model in image (b). Both maps show similar trends and patterns. The differences are mostly

**Table 3** Accuracy rates (ACR) and sums of upper diagonals (SUDs) as well as sums of lower diagonals (SLDs) for all five runs of cross-validation

| Network run | ACR | SUD | SLD | Dev. ACR |
|---|---|---|---|---|
| Network run 1 | 37.6% | 36.5% | 25.9% | \| 2.9 \| |
| Network run 2 | 41.3% | 34.3% | 24.4% | \| 0.8 \| |
| Network run 3 | 44.9% | 26.5% | 28.6% | \| 4.4 \| |
| Network run 4 | 39.4% | 18.2% | 42.4% | \| 1.1 \| |
| Network run 5 | 39.4% | 32.3% | 28.3% | \| 1.1 \| |

In the column "Dev. ACR", the absolute deviations of the five runs from the aggregated model are listed

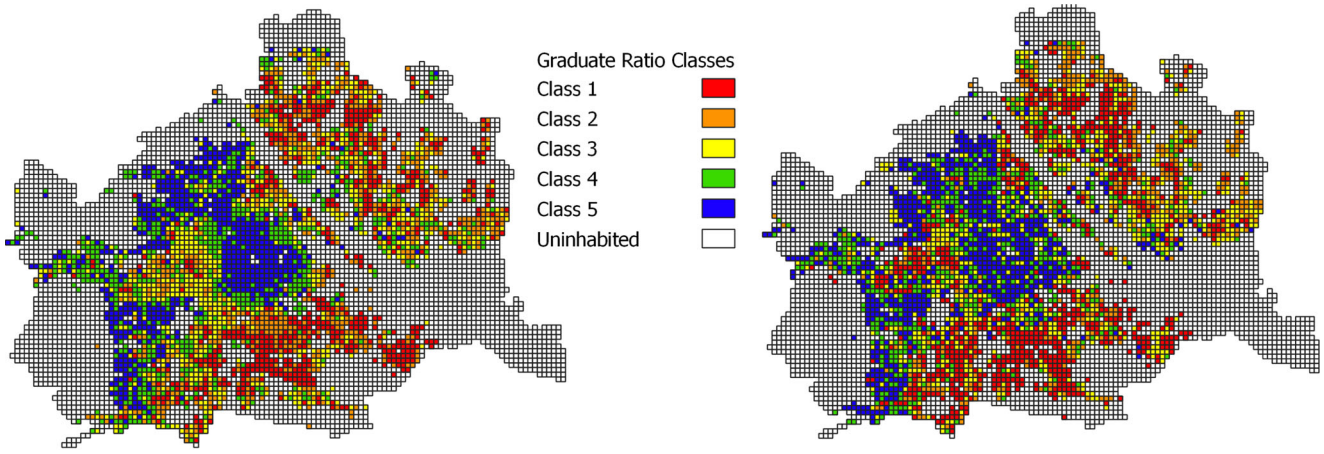



**Fig. 5** True distribution of graduates in Vienna according to the ground truth (**a**) and the prediction of our model (**b**). The red cells indicate a grid cell with a low graduate ratio. The blue grid cells show a high graduate ratio. The white cells are uninhabited areas (fewer than 20 inhabitants)

on a local level. Therefore, the predictions of our model are consistent.

To evaluate our results, we sensitise our approach in two ways:

i) i) To evaluate the generalisation ability of our methodology, we evaluate our approach on other (yet unseen) cities for which adequate ground-truth or reference data are available. We select the Austrian cities of Graz, Linz and Salzburg for testing and predict the distribution of the graduate ratios over the space.Our network, when trained on the Vienna ground truth, achieves 28.7% accuracy in Graz, 32.7% accuracy in Linz and 27.8% accuracy in Salzburg. The lower accuracy rates are mainly because the populations of these three cities are considerably smaller than that of Vienna (by a factor of approx. 10).

ii) Furthermore, we tested the prediction accuracy for Vienna with a different number of graduate ratio classes. Using two graduate ratio classes, the CNN achieves a prediction accuracy of 74.1%; three classes yield an accuracy of 48.7%, and four classes yield a prediction accuracy of 43.9%. In summary, these accuracy rates are substantially higher than those of random guessing (50%, 33% and 25%, respectively), which further confirms that the CNN finds patterns in the satellite images that correlate with graduate density.

### 4.3 Analysis of Class Deviations

To further investigate the performance of our machine learning approach, we create an additional heat map, shown in Fig. 6, which depicts the *deviation* of the predicted classes from the true class. This enables us to analyse the spatial distribution of the misclassifications, which in turn can help to better understand the performance of the model. We compute the signed deviation between the prediction and ground truth for each cell. Blue indicates an underestimation of graduate density, while red indicates an overestimation of the density compared to the true value per cell. All grid cells with correct predictions are coloured in green.

Figure 6 shows that large areas are correctly predicted (green). False predictions are distributed across the entire area, and no large spatial clusters can be observed. A certain tendency can be observed towards false predictions as we move into suburban regions away from the city centre. When considering the overall population density (see Appendix C for the heat map), there may be a link between the prediction accuracy and population density. As the population count decreases towards the periphery, this may explain the difficulties of the prediction model in suburban areas, leading to a higher number of misclassifications.

Examples of matching ("Match") and mismatching ("Mismatch") predictions of our CNN are marked on the city map in Fig. 6. Satellite images of the four example matches and mismatches (a)-(d) are shown in Figs. 7 and 8. Pairs (a) and (b) show strongly deviating images, where the CNN predicts a much higher or lower GR class. Pairs (c) and (d) show minor deviations compared to the true class.

Match (a) is located in the suburban area of Vienna and corresponds to class 4 (a graduate ratio of 21.6%). The CNN correctly predicts the satellite image as class 4. Mismatch (a) is in the same suburban area and neighbours the cell of Match (a). The CNN assigns it to the incorrect class; i.e., the ground truth indicates class 1 (a graduate ratio of 5%), while the CNN classifies it as class 5 (highest graduate ratio). By human visual judgement, both pictures

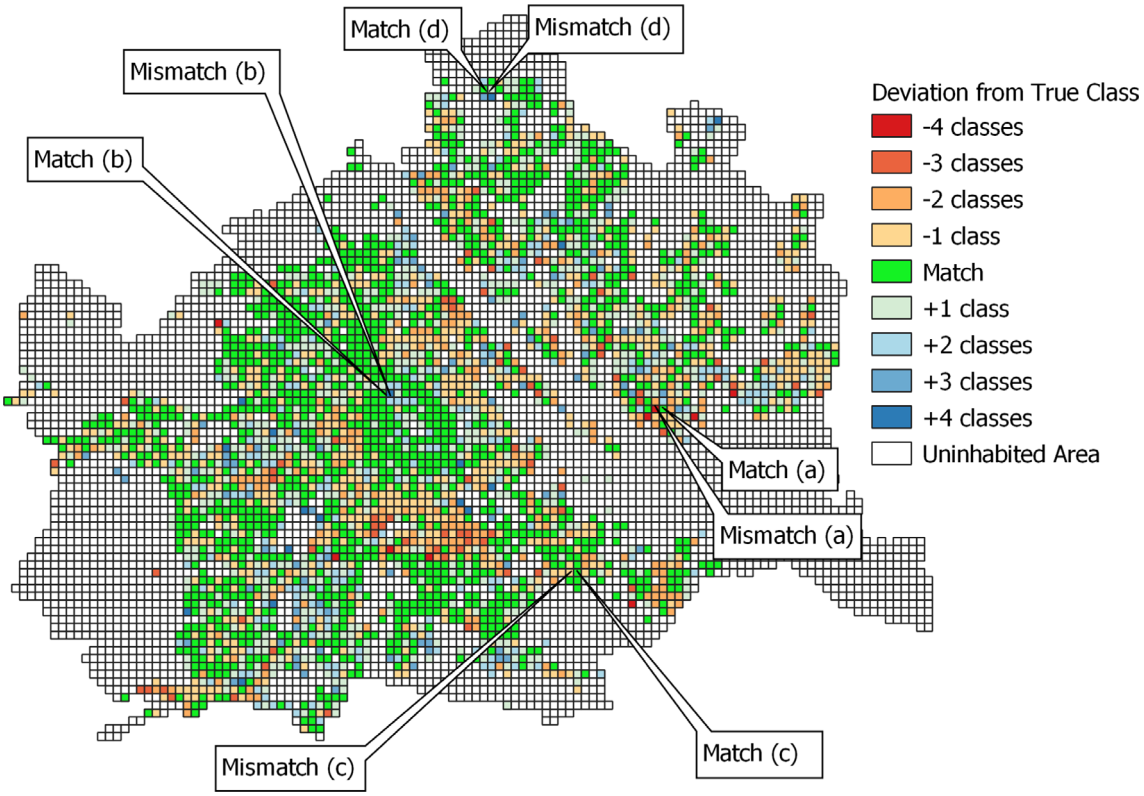



**Fig. 6** Heat map showing the deviation of the model predictions from the ground truth. Green indicates a match of the predicted and true graduate ratios. The reddish and blueish colours indicate an underestimation and overestimation of the density, respectively. More saturated colours indicate larger errors in estimation

generally look similar,; one difference is that Mismatch (a) depicts fewer streets and buildings than Match (a). Both pictures show vegetation and generally look like attractive residential areas. Thus, due to the attractiveness of the neighbourhood, it is reasonable to us that the CNN would predict Mismatch (a) as an incorrect class.

Match (b) is located in the city centre of Vienna. The true graduate ratio is 28.1%, i.e., class 5. Our prediction model correctly predicts this class. The neighbouring Mismatch (b) has a graduate ratio of 32.4% according to ground truth and thus would fall into class 5 as well. The model, however, fails to predict the true class and assigns class 1. Comparing the two pictures, we can see similar images with residential housing of different densities. An obvious difference is the square-shaped houses (panel buildings) in the wrongly predicted cell. We observe that panel buildings frequently correlate with lower graduate ratios (see, e.g., Match (c)). We assume that the network has recognised that this pattern

(a) (b) (c) (d)

**Fig. 7** Matches. Example satellite images (in the 224 × 224 pixel resolution employed for the CNN) of correctly matched predictions. The images correspond to Match (**a**) (GR: 21.6% with a population of 97), Match (**b**) (GR: 28.1% with a population of 839), Match (**c**) (GR: 3.0% with a population of 985) and Match (**d**) (GR: 8.1% with a population of 86)

(a) (b) (c) (d)

**Fig. 8** Mismatches. Example satellite images (in the 224 × 224 pixel resolution employed for the CNN) of false predictions. The images correspond to Mismatch (**a**) (true class: 1, predicted class: 5, GR: 5.0% with a population of 40), Mismatch (**b**) (true class: 5, predicted class: 1, GR: 32.4% with a population of 173), Mismatch (**c**) (true class: 1, predicted class: 2, GR: 5.7% with a population of 456), and Mismatch (**d**) (true class: 3, predicted class: 2, GR: 15.5% with a population of 181)

frequently accompanies a low graduate density and thus assigns the wrong GR class to the cell.

Match (c) is located on the outskirts of Vienna and shows numerous panel buildings. The true graduate ratio of the grid cell is 3.0% and is correctly predicted as class 1 by our approach. The neighbouring Mismatch (c) is falsely predicted as class 2 but actually belongs to class 1 (a graduate ratio of 5.7%). The two cells do not significantly deviate from each other visually, as both are only sparsely populated and show considerable areas covered by vegetation, especially trees. One difference is that match (c) contains more panel buildings and mismatch (c) contains more individual buildings surrounded by greenery. This may incline the model towards predicting a higher graduate density class, which is generally in line with the ground truth (there is a higher graduate ratio in Mismatch (c), 5.7%, than in Match (c), 3.0%). The network seems to overestimate the ratio, and thus, the result falls into the higher density class.

Match (d) is again in a suburban area and corresponds to a grid cell with a graduate ratio of 8.1% (correctly classified as class 2). The neighbouring cell, labelled as Mismatch (d), has a graduate ratio of 15.5% according to the ground truth and thus falls into class 3. Our model predicts it as a class 2 image. Match (d) depicts a rural neighbourhood with areas of farmland. Mismatch (d) displays denser settlement with less single family housing. This might be the reason why our approach underestimates the graduate density.

Overall, when looking at the example matching and mismatching predictions, we can draw the conclusion that a high proportion of the chosen pictures are difficult to assess even with human judgement. This may be a reason for the difficulties of the CNN in accurately modelling the classes and can also explain class confusions (especially between neighbouring classes). Regarding the prediction of graduate settlement, we are aware that the CNN could also predict variables such as rent or housing/apartment prices instead of the intended variable of graduate class. In that regard, one could claim that graduates are agglomerating in desirable areas [6, 13, 20] and therefore increasing rents or vice versa. Such correlations in the data are worth closer examination and represent an important direction of our follow-up research.

## 5 Conclusion

In this article, we show that a CNN can predict the spatial distribution of university graduates in a city using only satellite images. Our research hypothesis is that visual features exist in satellite images that correlate with the settlement of graduates. To investigate this hypothesis, we leverage the rich capabilities of machine learning to extract useful data from satellite images (250 m × 250 m small-scale city grid cells) and to link it to statistical population data. We split the statistical population data into five equally balanced classes with a wide range of graduate ratios. We train five neural networks on the ground-truth data and achieve an overall accuracy rate of 40.5% (random baseline: 20%) in predicting the five graduate density classes for the study site of the Austrian city of Vienna. We also show that we can differentiate inhabited and uninhabited areas with a probability of 96% by purely visual features using machine learning.

Our findings show that computer vision (i) has great potential for future examinations in urban economics (socio-economic and demographic studies), (ii) can mitigate the MAU problem or can serve as the basis for a solution and (iii) can be used in economic fields where no (statistical reference) data are available or the data are outdated, as computer vision can be deployed independently of the availability of statistical data and the metadata derived

from them. Computer vision therefore opens up a so far underestimated but extremely useful information source for economic analyses.

Future research will analyse how the network recognises the distribution of graduates in detail. Colours, contours, textures, arrangements of buildings, etc. can play a role. Another step will be to investigate to what extent a network trained on one site (e.g., Vienna) generalises to another site and whether the findings and patterns are consistent.

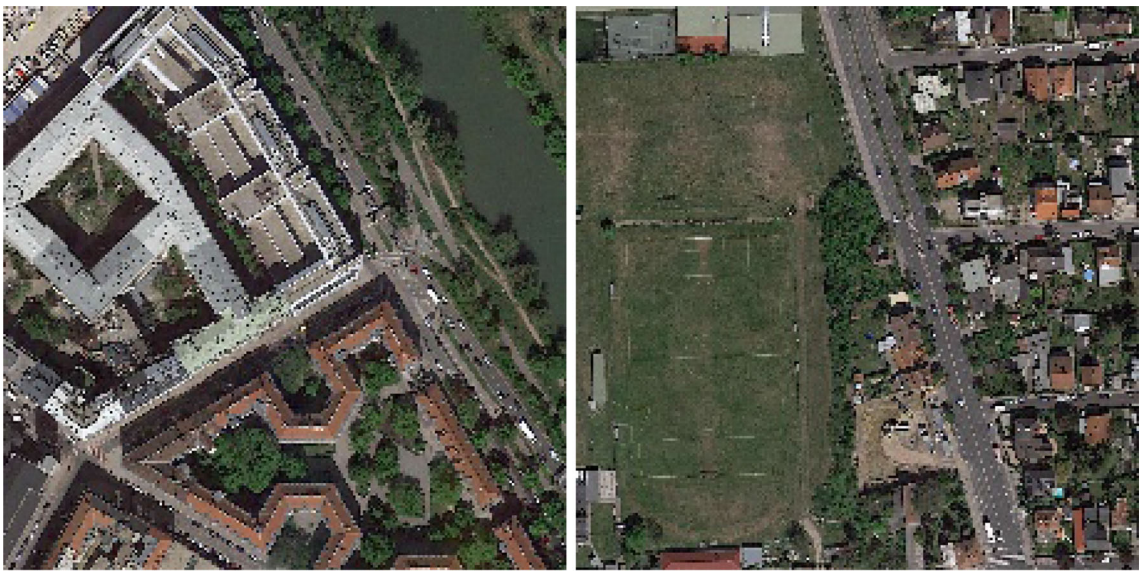
Overall, we see a broad applicability for our prediction model in future research and practice. The investigation of the predictability of graduate settlement in a metropolitan area could enhance future urban planning and guide urban development in the sense that it is controlled for human capital agglomeration. Especially in regard to urban governance, our findings can add a new dimension to city planning if future research is able to extract the visual characteristics that increase graduate agglomeration in certain city areas.

## Appendix A: Example Images for Different Graduate Classes

Example images of the five graduate density classes employed in our study, from the top, "class 1" (with a low graduateratio), to the bottom, "class 5" (with a high graduate ratio)

(a) (b)

**Fig. 9** : Class 1 examples: Image (**a**) depicts a location near the city centre of Vienna, with a graduate ratio of 5.8% and a population of 956. Image (**b**) is located in the suburban area of Vienna, with a graduate ratio of 1.6% and a population of 81



(a) (b)

**Fig. 10** Class 2 examples: Image (**a**) depicts a grid cell near the city centre of Vienna, with a graduate ratio of 8.1% and a population of 2,024. Image (**b**) is located in the suburban area of Vienna with a graduate ratio of 10.3% and a population of 107

(a) (b)

**Fig. 11** Class 3 examples: Image (**a**) depicts a grid cell near the city centre of Vienna, with a graduate ratio of 16.1% and a population of 1,516. Image (**b**) is located in the suburban area of Vienna with a graduate ratio of 14.4% and a population of 181 it falls into class 3

(a) (b)

**Fig. 12** Class 4 examples: Image (**a**) depicts a grid cell near the city centre of Vienna, with a graduate ratio of 18.1% and a population of 1,598. Image (**b**) is located in the suburban area of Vienna with a graduate ratio of 22.6% and a population of 62 it falls into class 4

(a) (b)

**Fig. 13** Class 5 examples: Image (**a**) depicts a grid cell near the city centre of Vienna, with a graduate ratio of 41.3% and a population of 886. Image (**b**) is located in the suburban area of Vienna with a graduate ratio of 35.0% and a population of 20 it falls into class 5

## Appendix B: Confusion Matrices for All Network Folds

**Table 4** Confusion matrix of network run 1

| | | Prediction | | | | | |
|---|---|---|---|---|---|---|---|
| | | c1 | c2 | c3 | c4 | c5 | |
| True class | c1 | 45 (34.3%) | 16 (12.0%) | 10 (7.7%) | 10 (7.6%) | 3 (2.3%) | |
| | c2 | 63 (48.1%) | 73 (54.9%) | 61 (46.9%) | 29 (22.1%) | 17 (13.1%) | |
| | c3 | 20 (15.3%) | 30 (22.6%) | 37 (28.5%) | 29 (22.1%) | 21 (16.1%) | |
| | c4 | 3 (2.3%) | 12 (9.0%) | 19 (14.6%) | 45 (34.4%) | 43 (33.1%) | |
| | c5 | – | 2 (1.5%) | 3 (2.3%) | 18 (13.8%) | 46 (35.4%) | AR: 37.6% |

**Table 5** Confusion matrix of network run 2 for Vienna

| | | Prediction | | | | | |
|---|---|---|---|---|---|---|---|
| | | c1 | c2 | c3 | c4 | c5 | |
| True class | c1 | 103 (76.9%) | 60 (44.5%) | 46 (34.9%) | 22 (16.5%) | 5 (3.7%) | |
| | c2 | 16 (11.9%) | 18 (13.3%) | 23 (17.4%) | 8 (6.0%) | 1 (0.8%) | |
| | c3 | 11 (8.2%) | 27 (20.0%) | 25 (18.9%) | 23 (17.3%) | 9 (6.7%) | |
| | c4 | 4 (3.0%) | 22 (16.3%) | 28 (21.2%) | 43 (32.3%) | 32 (23.9%) | |
| | c5 | – | 8 (5.9%) | 10 (7.6%) | 37 (27.8%) | 87 (64.9%) | AR: 41.3% |

**Table 6** Confusion matrix of network run 3 for Vienna

| | | Prediction | | | | | |
|---|---|---|---|---|---|---|---|
| | | c1 | c2 | c3 | c4 | c5 | |
| True class | c1 | 93 (69.4%) | 50 (37.0%) | 25 (18.9%) | 14 (10.5%) | 5 (3.7%) | |
| | c2 | 8 (6.0%) | 17 (12.6%) | 15 (11.4%) | 8 (6.0%) | – | |
| | c3 | 27 (20.1%) | 31 (23.0%) | 55 (41.7%) | 44 (33.1%) | 7 (5.2%) | |
| | c4 | 2 (1.5%) | 17 (12.6%) | 21 (15.9%) | 22 (16.6%) | 9 (6.7%) | |
| | c5 | 4 (3.0%) | 20 (14.8%) | 16 (12.1%) | 45 (33.8%) | 113 (84.4%) | AR: 44.9% |

**Table 7** Confusion matrix of network run 4 for Vienna

| | | Prediction | | | | | |
|---|---|---|---|---|---|---|---|
| | | c1 | c2 | c3 | c4 | c5 | |
| True class | c1 | 71 (53.0%) | 42 (31.1%) | 16 (12.1%) | 3 (2.3%) | – | |
| | c2 | 33 (24.6%) | 28 (20.7%) | 10 (7.6%) | 6 (4.5%) | 1 (0.8%) | |
| | c3 | 12 (9.0%) | 20 (14.8%) | 21 (15.9%) | 5 (3.8%) | 6 (4.5%) | |
| | c4 | 18 (13.4%) | 41 (30.4%) | 50 (37.9%) | 49 (36.8%) | 33 (24.6%) | |
| | c5 | – | 4 (3.0%) | 35 (26.5%) | 70 (52.6%) | 94 (70.1%) | AR: 39.4% |

**Table 8** Confusion matrix of network run 5 for Vienna

| | | Prediction | | | | | |
|---|---|---|---|---|---|---|---|
| | | c1 | c2 | c3 | c4 | c5 | |
| True class | c1 | 88 (67.7%) | 60 (45.8%) | 37 (28.2%) | 23 (17.5%) | 4 (3.1%) | |
| | c2 | 1 (0.8%) | – | – | – | – | |
| | c3 | 31 (23.9%) | 58 (44.3%) | 63 (48.1%) | 52 (39.7%) | 18 (13.7%) | |
| | c4 | 5 (3.8%) | 4 (3.0%) | 12 (9.2%) | 15 (11.5%) | 17 (13.0%) | |
| | c5 | 5 (3.8%) | 9 (6.9%) | 19 (14.5%) | 41 (31.3%) | 92 (70.2%) | AR: 39.4% |

## Appendix C: Population Distribution in the Study Area

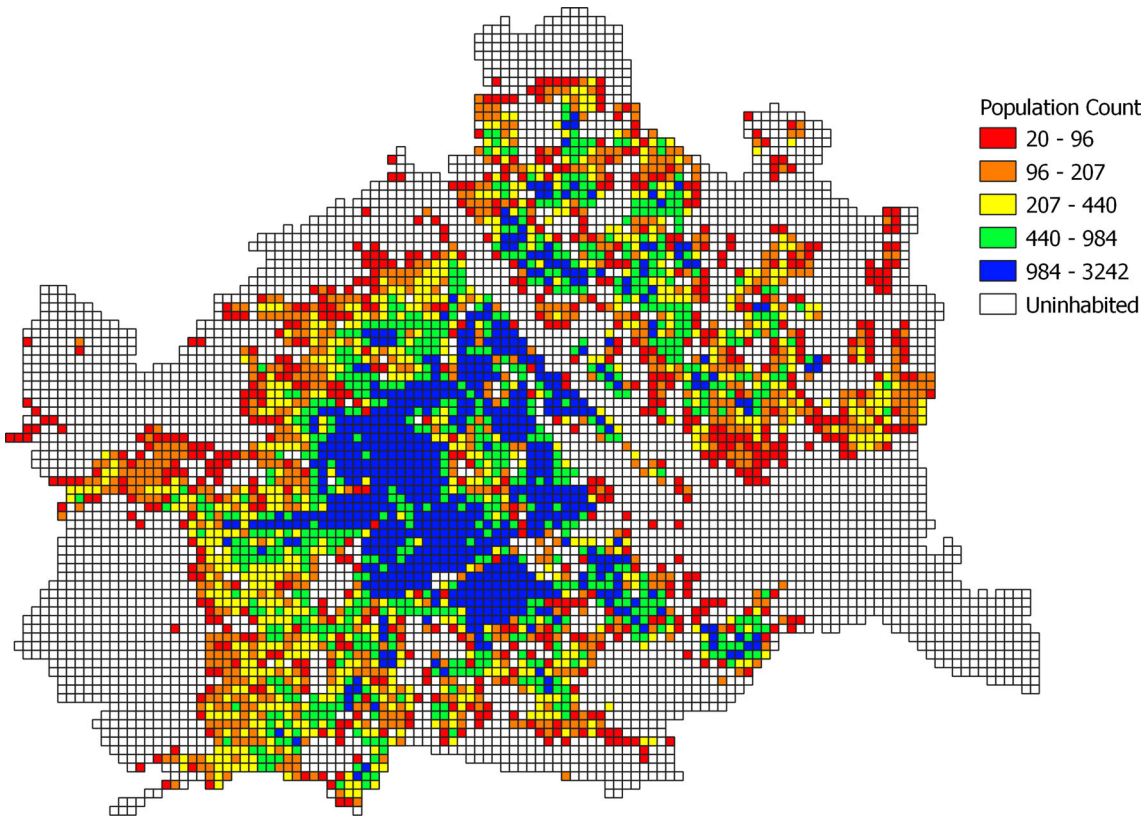



**Fig. 14** Heat map depicting the distribution of the population density in the 250 m × 250 m regional statistical grid data for Vienna. The red cells indicate a low population density, and the blue cells indicate a high population density. The white cells indicate a population of fewer than 20 people

**Funding** Open access funding provided by FH Kufstein Tirol - University of Applied Sciences.



## References

1. Arribas-Bel D, Garcia-López M-À, Viladecans-Marsal E (2019) Building(s and) cities: Delineating urban areas with a machine learning algorithm. Journal of Urban Economics, pp 103217
2. Bacolod M, Blum BS, Strange WC (2010) Elements of Skill: Traits, Intelligence, Education and Agglomeration. J Reg Sci 50(1):245–280
3. Bennett MM, Smith LC (2017) Advances in using multitemporal night-time lights satellite imagery to detect, estimate, and monitor socioeconomic dynamics. Remote Sens Environ 192(2019):176–197
4. Rafael Ch, Martin DA, Vargas JF (2020) Measuring the Size and Growth of Cities Using Nighttime Light. Journal of Urban Economics, pp 103254
5. Xi C, Nordhaus WD (2011) Using luminosity data as a proxy for economic statistics. Proc Natl Acad Sci USA 108(21):8589–8594
6. Costa DL, Kahn ME (2000) Power Couples: Changes in the Locational Choice of the College Educated, 1940-1990. Q J Econ 115(4):1287–1315
7. Cuaresma JC, Danylo O, Fritz S, McCallum I, Obersteiner M, See L, Walsh B (2017) Economic development and forest cover: Evidence from satellite data. Sci Rep 7(2016):1–8
8. Dark SJ, Bram D (2007) The modifiable areal unit problem (MAUP) in physical geography. Prog Phys Geogr 31(5):471–479
9. de Bellefon M-P, Combes P-p, Duranton G, Gobillon L, Gorin C (2019) Delineating urban areas using building density. Journal of Urban Economics, pp 103226
10. El Merabet Y, Meurie C, Ruichek Y, Sbihi A, Touahni R (2015) Building roof segmentation from aerial images using a line-and region-based watershed segmentation technique. Sensors 15(2):3172–3203
11. Fröhlich B, Bach E, Walde I, Hese S, Schmullius C, Denzler J (2013) Land cover classification of satellite images using contextual information. ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences II-3/W1(3w1):1–6
12. Galdo V, Li Y, Rama M (2019) Identifying urban areas by combining human judgment and machine learning: An application to India. Journal of Urban Economics, (February):103229 dec
13. Glaeser EL, Kolko J, Saiz A (2001) Consumer city. J Econ Geogr 1(1):27–50
14. Grekousis G (2019) Artificial neural networks and deep learning in urban geography: a systematic review and meta-analysis. Comput Environ Urban Syst 74(October 2018):244–256
15. Hammoudi K, Dornaika F (2011) A featureless approach to 3D polyhedral building modeling from aerial images. Sensors 11(1):228–259
16. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2016, pp 770–778
17. Vernon Henderson J, Storeygard A, Weil DN (2012) Measuring economic growth from outer space. Am Econ Rev 102(2):994–1028
18. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ (2017) Densely connected convolutional networks. In: Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition
19. Huang T-L, Orazem PF, Wohlgemuth D (2002) Rural Population Growth, 1950–1990: The Roles of Human Capital, Industry Structure, and Government Policy. Am J Agric Econ 84(3):615–627
20. Signe (University of Jyväskylä) Jauhiainen (2005) Regional concentration of highly educated couples. In: ERSA conference papers, pp 15
21. Jean N, Burke M, Xie M, Matthew DW, Lobell DB, Ermon S (2016) Combining satellite imagery and machine learning to predict poverty. Science 353(6301):790–794
22. Jeawak SS, Jones CB, Schockaert S (2020) Predicting the environment from social media: a collective classification approach. Comput Environ Urban Syst 82(2015):101487
23. Jiang S, Alves A, Rodrigues F, Ferreira J, Pereira FC (2015) Mining point-of-interest data from social networks for urban land use classification and disaggregation. Comput Environ Urban Syst 53:36–46
24. Kadish J, Netusil NR (2012) Valuing vegetation in an urban watershed. Landsc Urban Plan 104(1):59–65
25. Kobayashi Y (2006) Photogrammetry and 3D city modelling. In: Digital architecture and construction, volume 1 of wit transactions on the built environment, vol 90. WIT Press, Southampton, pp 209–218
26. Koch D, Despotovic M, Leiber S, Sakeena M, Doeller M, Zeppelzauer M (2019) Real estate image analysis: a literature review. J Real Estate Lit 27(2):269–300
27. Kodors S, Rausis A, Ratkevics A, Zvirgzds J, Teilans A, Ansone I (2017) Real estate monitoring system based on remote sensing and image recognition technologies. Procedia Comp Sci 104(December):460–467
28. Krugman P (1991) Increasing returns and economic geography. J Polit Econ 99(3):483–499
29. Krugman P (1992) Geography and Trade, 1st edn. The MIT Press, Cambridge
30. Li X, Zhang C, Li W, Kuzovkina YA, Weiner D (2015) Who lives in greener neighborhoods? The distribution of street greenery and its association with residents' socioeconomic conditions in Hartford, Connecticut, USA. Urban Forestry and Urban Greening 14(4):751–759
31. Liu Y, Shen J, Xu W, Wang G (2017) From school to university to work: migration of highly educated youths in China. Ann Reg Sci 59(3):651–676
32. Lu Z, Im J, Rhee J, Hodgson M (2014) Building type classification using spatial and landscape attributes derived from liDAR remote sensing data, vol 130
33. Lukač N, Seme S, Žlaus D, Štumberger G, Žalik B (2014) Buildings roofs photovoltaic potential assessment based on liDAR (Light Detection And Ranging) data. Energy 66:598–609
34. Lukashevich P, Zalessky B, Belotserkovsky A (2017) Building detection on aerial and space images. In: International conference on information and digital technologies (IDT), IEEE, pp 246–251

35. Marmanis D, Datcu M, Esch T, Stilla U (2016) Deep learning earth observation classification using ImageNet pretrained networks. IEEE Geosci Remote Sens Lett 13(1):105–109
36. Mills B, Hazarika G (2001) The migration of young adults from non-metropolitan counties. Am J Agric Econ 83(2):329–340
37. Mori T, Turrini A (2005) Skills, agglomeration and segmentation. Eur Econ Rev 49(1):201–225
38. Muhr V, Despotovic M, Koch D, Döller M, Zeppelzauer M (2017) Towards automated real estate assessment from satellite images with CNNs. Forum Media Technology 10:14–23
39. Naik N, Philipoom J, Raskar R, Hidalgo C (2014) Streetscore – predicting the perceived safety of one million streetscapes. In: 2014 IEEE Conference on computer vision and pattern recognition workshops, vol 55, IEEE, pp 793–799
40. Openshaw S. (1984) Ecological Fallacies and the Analysis of Areal Census Data. Environment and Planning A: Economy and Space 16(1):17–31
41. Openshaw S, Taylor PJ (1979) A Million or so Correlation coefficients: Three Experiments on the Modifiable Areal Unit Problem. Wrigley N. Publishers, London
42. Raikar A, Hanji G (2016) Automatic building detection from satellite images using internal gray variance and digital surface model. Int J Comput Appl 145(3):25–33
43. Richards JA, Xiuping J (2006) Remote Sensing Digital Image Analysis, 4th edn. Springer, Berlin
44. Simonyan K, Zisserman A (2015) Very deep convolutional networks for Large-Scale image recognition. In: International conference on learning representations, pp 1–14
45. Sumer E, Turker M (2013) An adaptive fuzzy-genetic algorithm approach for building detection using high-resolution satellite images Computers. Environ Urban Syst 39:48–62
46. Tapiador FJ, Avelar S, Tavares-Corrêa C, Zah R (2011) Deriving fine-scale socioeconomic information of urban areas using very high-resolution satellite imagery. Int J Remote Sens 32(21):6437–6456
47. Unwin DJ (1996) GIS, spatial analysis and spatial statistics. Prog Hum Geogr 20(4):540–551
48. Vedaldi A, Lenc K (2015) Matconvnet: Convolutional neural networks for MATLAB. In: Proceedings of the ACM Multimedia Conference, pp 689–692
49. Wang Z, Liu L (2014) Assessment of coarse-resolution land cover products using CASI hyperspectral data in an arid zone in Northwestern China. Remote Sens 6(4):2864–2883
50. Weir-Smith G (2016) Changing boundaries: Overcoming modifiable areal unit problems related to unemployment data in South Africa. S Afr J Sci 112(3-4):1–8
51. Xie M, Jean N, Burke M, Lobell D, Ermon S (2016) Transfer learning from deep features for remote sensing and poverty mapping. In: 30th AAAI Conference on Artificial Intelligence, pp 3929–3935
52. Zhang M, Kukadia N (2005) Metrics of urban form and the modifiable areal unit problem. Transportation Research Record, (1902):71–79
53. Zhang W, Li W, Zhang C, Hanink DM, Li X, Wang W (2017) Parcel-based urban land use classification in megacity using airborne LiDAR, high resolution orthoimagery, and Google Street View. Comput Environ Urban Syst 64:215–228