MACHINE LEARNING FOR IMAGE-BASED RECOGNITION OF BUILDING AGE FOR URBAN ENERGY SIMULATION – TESTING AND VALIDATION ON AN EXEMPLARY CITY QUARTER

Alexander Benz, Mara Geske, Conrad Voelker[1]

[1] Bauhaus-Universität Weimar, Deutschland, E-Mail: alexander.benz@uni-weimar.de

Abstract

Data acquisition for urban energy simulations is time consuming and usually associated with incomplete data sets which enhanced by subjective approaches (e.g., estimation of building age and a probable construction). The analysis of exterior views based on machine learning algorithms enables an automated and reproducible recognition of building age classes. Therefore, convolutional neural networks (CNNs) were trained in a supervised process with approximately 3,700 images of residential buildings from different German building age classes. The highest accuracy obtained (56 % correct) exceeds the human prediction accuracy (37 % correct). A real life test and validation of the trained CNNs is conducted on an inhomogeneous city quarter with predominantly residential use.

Introduction

Energy simulations continuously focus urban environments, such as city quarters or cities. Hence, various approaches towards modelling and simulation within the urban context have been published in the past (Alaia et al., 2020; Malhotra, A. et al., 2022). One major drawback of such investigations usually is the compromise between the required computation time, aspired resolution (e.g., building, quarter or city) and the achieved accuracy (e.g., deviation between calculated and measured energy demand). With the aim to reduce the workload, researchers propose automated workflows, including simulations and a pre-processing of the applied data sets (Remmen et al., 2016). Figure 1 depicts the illustration of an exemplary automated workflow. Figure 1 does not only show the workflow of an automated building simulation but also indicates the application of appropriate data formats such as CityGML or GeoJSON, both of which are able to represent urban environments (e.g., buildings). Even though published by state authorities, such as (TLBG, 2022) and commonly used, these data sets usually do not include information on building age or details on construction (Biljecki, 2017). Hence, statistical data (see area highlighted in red and included in Figure 1) are being accessed. On the other hand, alternative city models, based on voluntary user inputs, such as OpenStreetMap, often contain incorrect information (Hecht et al., 2015). Consequently, various approaches have been published on methods to obtain a building stock. To achieve an objective as well as automatable approach towards the data acquisition in a large building stock, the authors suggest an image-based recognition of the building age using methods of Machine Learning.

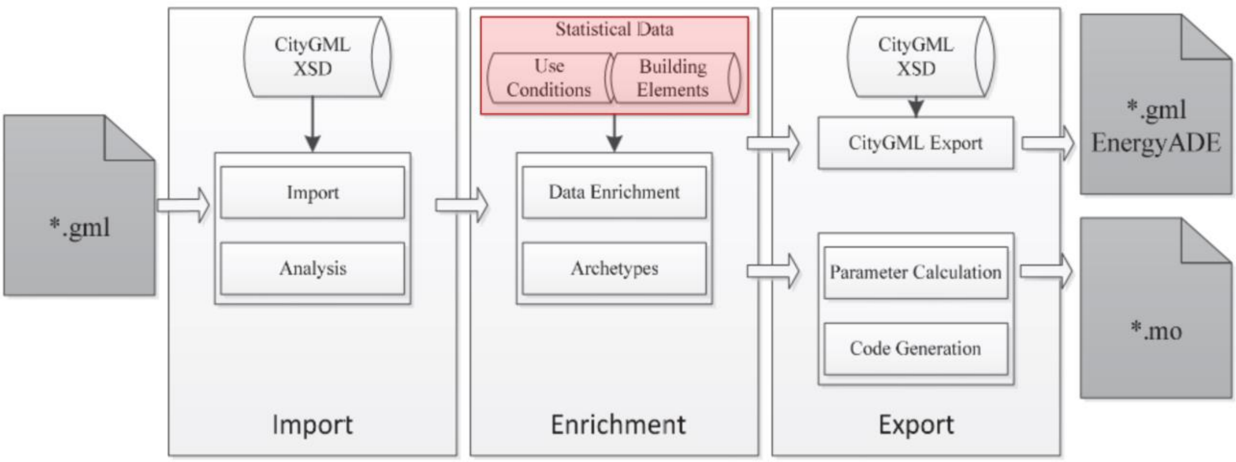

Figure 1: Workflow for an automated building simulation, adapted from (Remmen et al., 2016)

As a reference for the classification of building images, the IWU-typology as a subset of the TABULA project is applied (Loga et al., 2015). Within this typology, the building age is presumably the most significant criterion for classification and subsequent information (e.g., U-value, construction materials) can be derived as inputs for urban energy simulations (Benz, A. et al., 2018; Benz, A. et al., 2021).

State of the art

Automated classifications of building styles based on images have been carried out by several authors. (Chu, 2012) retrieved visual patterns from images by referring to the set of Scale-invariant feature transform (SIFT) descriptors in architecture images. The authors identified four different architectural styles and achieved an accuracy of up to 82 %. (Goel et al., 2012) classified various data sets (such as Oxford 5k and 100k (University of Oxford, 2022)) in different architectural categories, such as: Art Noveau, Baroque, Gothic Renaissance, Romanesque. The study achieved an average accuracy of 95 %. (Doersch et al., 2012) investigated Histogram of Oriented Gradients (HOG) descriptors in image patches and identified different cities based on this images. A similar approach was studied by (Lee et al., 2015). The authors of this study discriminate 10 building epochs, using Google Street View and cadastre maps.

As architectural styles only refer to a style period and most approaches focus on acient epochs, more elaborated approaches using a higher resoultion of epochs are required for energy simulations. Hence, further studies on the automated classification of the building age were presented by (Zeppelzauer et al., 2018). Based on image patches, extracted from 6,450 building images as a training set, the authors used CNNs to achieve a classification of six epochs. Within this study two different CNN-architectures (AlexNet and ResNet50, both pretrained on ImageNet) were investigated, yielding an accuracy of 61 % on the testing data set. (Despotovic et al., 2018) applied a pretrained AlexNet for the prediction of the heating demand of single family dwellings. The authors obtained an accuracy of 53 %.

In contrast to the classification of buildings into discrete epochs, an alternative approach is the application of regressions. As presented by (Li et al., 2018) a continous resolution of building ages can be achieved. The study analysed 520,000 images as inputs for a number of pre-trained CNNs (ResNet18, ResNet50, DenseNet161). Using the test images, the authors achieved an error of less than 5 years for 15 % and less than 10 years for 25 % of the data used.

Eventhough various approaches on the estimation of building age have been published, no coupling to existing building typologies has been investigated. Furthermore no practical application in an urban environment was presented in literature. To address this issue, a test and selective validation on an urban environment is presented within this conference contribution.

Method

This study uses 3,714 images, acquired solely in the city of Weimar, for the training of CNNs. The CNNs are trained to classify building images, obtained at street-level, into 10 predefined epochs. In contrast to previous approaches which apply patch-extraction strategies (Zeppelzauer et al., 2018) to obtain an enhanced ground-truth, this study follows a naïve approach based on images in an unedited state. This results in a massive reduction in computation time, as the data set is kept at a minimum and no additional algorithms need to be implemented. Within this study, two different CNN-architectures (AlexNet and ResNet50) are investigated. The open source library PyTorch is used to create the CNNs. Based on previous observations (Benz, 2021; Zeppelzauer et al., 2018), the best results for training procedures on building images are expected with ResNet50 (pretrained on ImageNet) as CNN-architecture.

Eventhough the ResNet50-architecture includes 50 layers and therefore significantly more layers than AlexNet (8 layers), it promises a compromise between computation time and accuracy, whereas AlexNet tends towards longer training times and lower accuracies. The significantly longer training time of AlexNet is a result of three fully connected layers, which are attached to five convolutional layers. In contrast ResNet50 utilizes skip connections, which bypass individual layers and accelarate training processes.

Table 1: Distribution of image data within the ground-truth acquired in the city of Weimar

	...-1859	**1860-1918**	**1919-1948**	**1949-1957**	**1958-1968**	**1969-1978**	**1979-1983**	**1984-1994**	**1995-2001**	**2002-...**
training	325	1,248	877	87	121	144	93	170	427	222
validation	47	178	125	12	17	21	13	24	61	32
test	93	356	251	25	35	41	27	49	122	64

With respect to existing typologies, the image ground-truth was split into ten classes, representing the ten epochs depicted by (Loga et al., 2015). Table 1 represents the distribution of the entire image ground-truth.

The deliberate limitation on the city of Weimar leeds to a strongly imbalanced ground-truth. As this circumstance might cause a negative influence on the training process and hence a higher potential for missclassifications, this study also investigates different balancing strategies (reduction of elements in majority class and weighting factors for the training process) in order to reduce potential misclassifications.

These weight factors f_i are calculated as follows:

$$f_i = \frac{n_{max}}{n_i}, \qquad \textbf{(1)}$$

with n_{max} representing the quantity of images in the majority class and n_i the number of images in the corresponding class.

To enhance the estimation accuracy, extensive data augmentation was implemented (random horizontal flipping, random perspective distortion, random alteration of brightness, contrast as well as saturation). Training was conducted over 20 epochs. To take into account the progress of training, an adaptive learning rate was chosen. Within the training process, two different optimizers (Adaptive moment estimation (Adam) and Stochastic Gradient Descent (SGD)) were investigated. Whereas SGD has been applied in various other studies (e.g., Zeppelzauer et al., 2018), to best of our knowledge no studies on the application of Adam as an optimizer have been published in the context of image-based building age estimation. However, according to (Kingma, 2015), Adam works well on empirical optimization problems and has less computation costs as well as better opimization results. Therefore, higher accuracies and a faster convergence towards an optimum are expected from this optimizer.

In accordance with the IWU-typology (Loga et al., 2015), the final layer of each CNN is costumized to 10 output neurons, representing the 10 epochs of building ages. As depicted in Figure 2, the accuracy declines if all classes are balanced by reducing the majority classes. The authors therefore do not consider this approach to be benefitial. Eventhough the application of weight factors reduces the prediction accuracies of AlexNet, the performance of ResNet50 remains almost constant with approximately 51 % correct predictions. Based on this conclusion, AlexNet is neglected for further investigations.

An analysis of the corresponding confusion matrices (Figure 3) visualizes the impacts introduced by the weight balancing. The cluster of incorrect predictions in the lower left corner (new buildings predicted as old) is reduced. Predictions of the periods until 1948 are also shifted towards the main diagonal. On the other hand the cluster on the top right corner is slightly amplified and predictions in the lower right corner shift towards a fuzzy state.

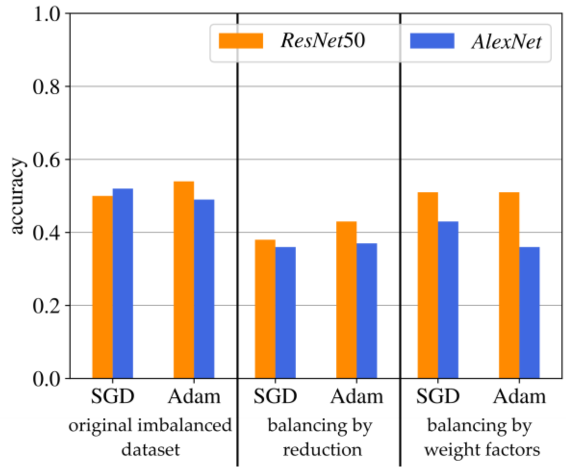

Figure 2: Prediction accuracy on test data (1.0 being 100 % correct predictions)

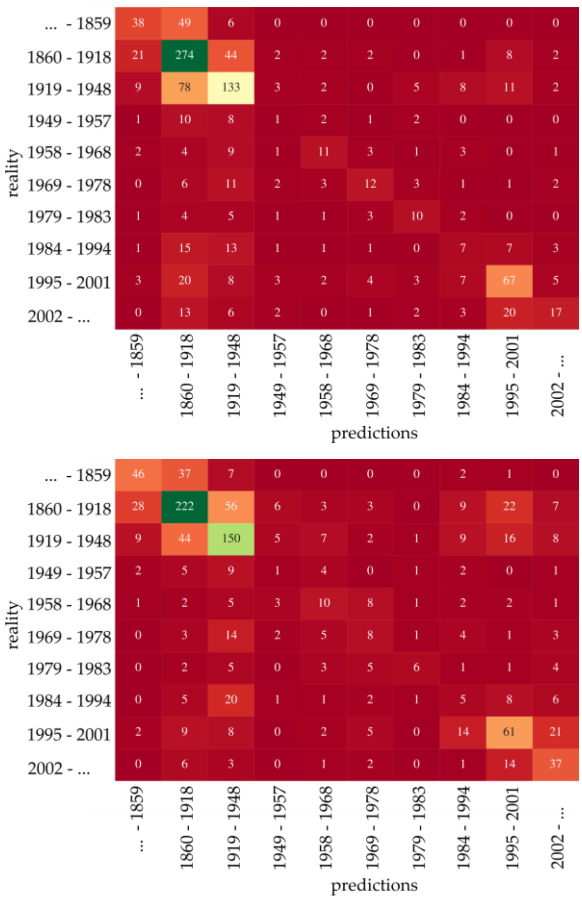

Figure 3: Confusion matrix of imbalanced original data set using ResNet50 with Adam (top) and weight balanced data set using ResNet50 and Adam (bottom)

As a reference for the trained CNNs, a human baseline is established. Therefore, 400 images were chosen randomly from the ground-truth and classified by 8 experts, yielding an accuracy of 37 %. Both approaches, training on imbalanced as well as on weight-balanced data, outperform this baseline.

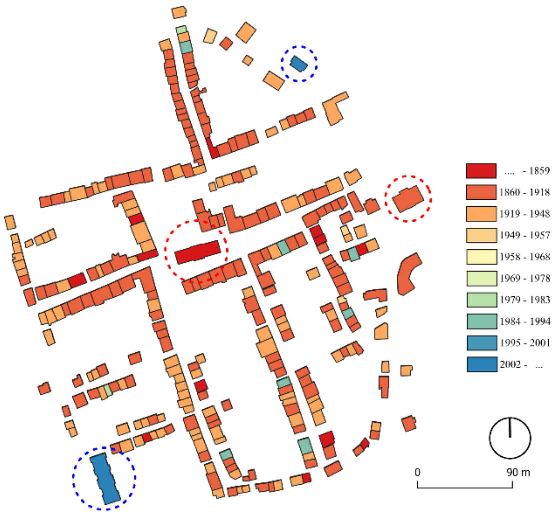

Figure 4: Results of CNN-estimation of building age within the town center of Königsee

Test of the proposed method on a city quarter

To proof the capability of the proposed method, the CNNs are applied on imagery acquired in Königsee, a town located in the southern part of the federal state of Thuringia, Germany. Data acquisition was restricted to the center of Königsee, which includes approximately 300 buildings. Each building located in this area, is represented by one image of the building façade (handheld camera at street-level). Within a first stage, a suitable data set containing the building geometries at LoD 2 (representing footprint, height and roof type of each building) was obtained from regional authorities (TLBG, 2022). From these data set unnecessary entries (e.g. unoccupied buildings, sheds and garages) were removed. As the proposed method only applies to residential buildings, industrial estates (e.g. workshops or factories) were excluded from the data set as well.

Estimation of building age

The obtained image files are used as input for the previously trained CNNs. For insights on the training process, we refer to the chapter *Method*. Herefore all image tensors undergoe a normalization (including mean and standard deviation) as well as a resizing in order to meet the required input size of the trained CNNs. The output of each CNN is represented by a fully-connected layer, consisting of 10 neurons. As an estimator for the probability of each class, softmax was used. This rescales the tensor of all output neurons in a way, so that all elements of each tensor are in the range [0,1] and sum to 1. Figure 4 depicts the building footprints within the center of Königsee, including the CNN-estimations of building epochs. For this estimation, the best performing CNN (ResNet50, Adam optimizer, trained on the imbalanced dataset) was applied.

Following the image-based recognition of building ages, it can be concluded that a majority of the building stock was constructed within the periods of 1860-1918 and 1918-1948 respectively. According to the CNN-estimation, a minority was constructed in the period of 1984-1994 and only two buildings were constructed after 2002.

The data set now represents an enriched state and contains the building footprints, the corresponding heights as well as the building age based on CNN-estimations and therefore can be used as input for automated building simulations. This can be accomplished by coupling the enhanced data set with suitable building typologies, from which corresponding information on construction can be derived.

Validation of the proposed method

As the previously presented visualization in Figure 4 depicts only the epoch of highest probability for each building, the question of uncertainty arises. In this specific case, the uncertainty reffers to a potential fuzziness in estimations. Figures 5 and 6 display the buildings (highlighted in red circles in Figure 4) and the corresponding predicted epochs. The class with the highest likelihood (also visualized in Figure 4) is highlighted in dark red.

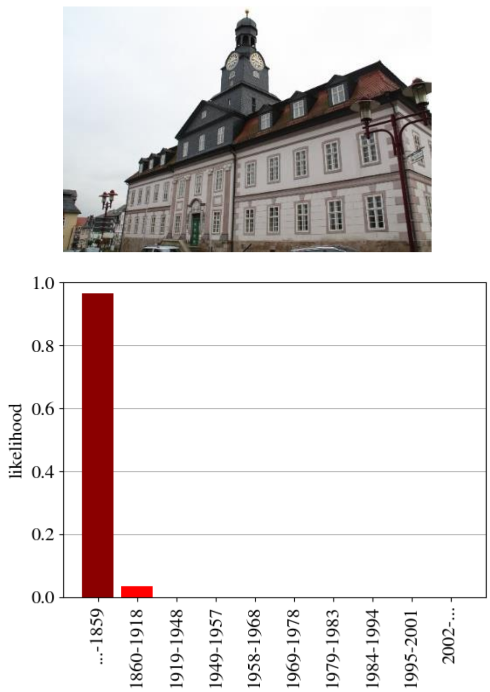

Figure 5: Town hall (top) and corresponding predicted likelihoods of building epochs (bottom)

Figure 5 shows a distinct estimation of the building epochs for the town hall in Königsee. The highest likelihood of approximately 97 % for the epoch before 1859 corresponds to the correct epoch (year of construction: 1719).

Figure 6 depicts the analysis of the city’s elementary school building. Eventhough the class 1860-1918 represents the majority vote (approximately 77 %), a relatively high probability is predicted for the first class (…-1859). Despite this increased fuzziness, a clear voting for the correct epoch is given (1887, stated by (Wikimedia Foundation, 2022)).

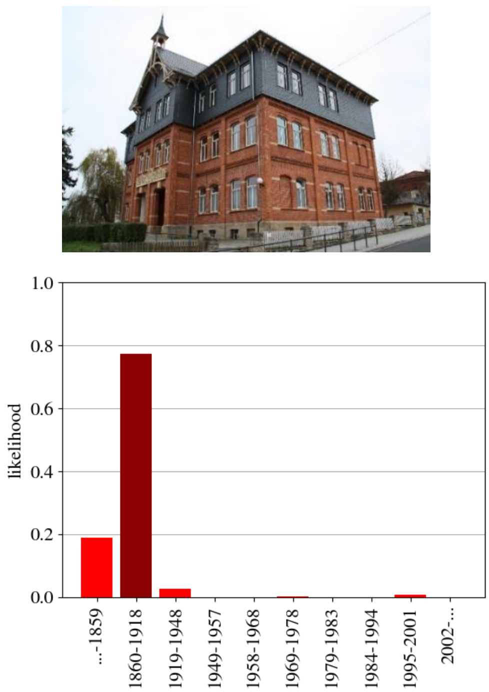

Figure 6: Elementary school building (top) with corresponding predicted likelihoods of building epochs (bottom)

Figure 7 displays the predicted epochs for two recently constructed buildings (highlighted in blue circles in Figure 4). The class with the highest likelihood (also visualized in Figure 4) is highlighted in dark blue. With respect to data privacy, no images of these buildings are presented.

In both cases, the epoch of highest likelihood corresponds to the correct year of construction (2013 and 2015 respectively). An increased fuzziness is present for the analysis of the retirement home. A significant shift towards the epochs of 1984-1994 and 1995-2001 results in a likelihood of 55 % for the correct epoch of later than 2002. A potential reason for this increased fuzziness is a restricted view on the building façade. Due to limited space around the building, the image obtained by the authors does not cover the entire façade.

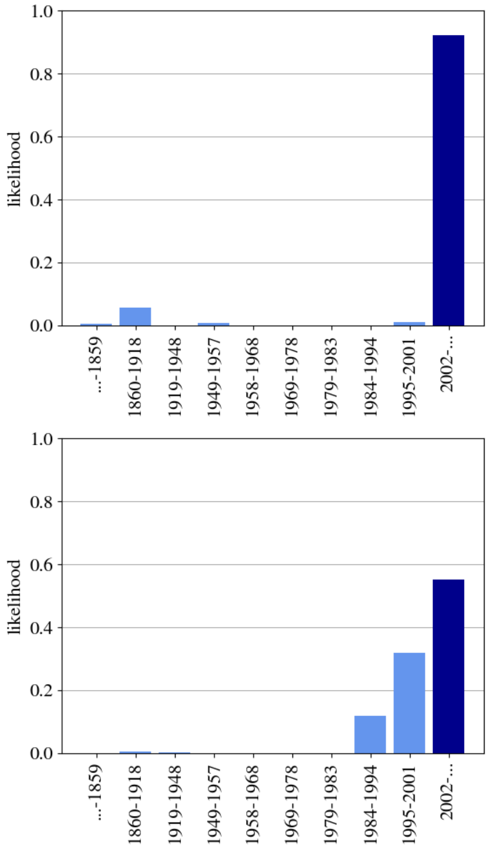

Figure 7: Predicted likelihoods of building epochs, single family dwelling (top), retirement home (bottom)

Discussion

The presented validation of the proposed methodology verifies the generally good performance of the trained CNNs. A good perfomance for the periods earlier than 1948 and later than 2002 is also obtained during the training process. However, it is still unclear how the CNNs perform in the midfield (1949-2001) of the applied building typology. Based on the training process, the authors assume a lower accuracy for these periods.

The presented study focused on the region of Thuringia and therefore might be subjected to a data set bias. This assumption is affirmed in the investigations conducted by (Benz, 2021; Zeppelzauer et al., 2018) Hence, further investigations should be carried out focusing the regional influence on the performance of image analysis and including a more comprehensive building stock for validation, as this study only includes a partial validation on a relatively small number of buildings (4 buildings).

Conclusion and Outlook

Within this study, a naïve approach for an image-based estimation was investigated. The authors used pretrained CNNs for a training on a relatively sparse ground-truth of 3,714 images, obtained solely in the city of Weimar. Two CNN-architectures as well as different training strategies were investigated, yielding different prediction accuracies and distributions in the corresbonding confusion matrices. The test of the trained CNNs achieved a maximum accuracy of about 55 %, which outperforms the human baseline of 37 %. With the application of the trained CNNs in the town of Königsee, a proof of concept was presented, illustrating the potential of large-scale applications for the estimation of building ages and therefore a useful extension to current workflows for automated building simulations.

Following the promising results of this proof of concept, further studies need to be conducted. The most important aspect is a potential data-set bias and the handling of it. Future studies should also focuse on a more sophisticated approach based on image patches, obtained from the original imagery.

Acknowledgements

We genuinely thank Mr. Thomas Lichtenheld (head of building authority of Königsee) for the support during acquisition of building images within the center of Königsee.

References

Alaia, S. et al. 2020. Multi-domain urban-scale energy modelling tools: A review, Sustainable Cities and Society 54, p 101872.

Benz, A. 2021. Automatisierte bildbasierte Erkennung des Gebäudealters, 31. Hanseatische Sanierungstage 2021, Lübeck, Deutschland.

Benz, A. et al. 2018. Thermisch-energetische Gebäudesimulation auf Basis eines Bauwerksinformationsmodells, Bauphysik 40, p. 61-67.

Benz, A. et al. 2021. Framework for a UAS-based assessment of energy performance of buildings, Energy and Buildings 250, p 111266.

Biljecki, F.; Sindram, M. 2017. Estimating building age with 3D GIS, 12th 3D Geoinfo Conference 2017, Melbourne, Australia.

Chu, W. T.; Tsai, M. H. 2012. Visual pattern discovery for architecture image classification and product image search, International Conference on Multimedia Retrieval 2012, Hong Kong, PRC.

Despotovic, M. et al. 2018. Predicting heating energy demand by computer vision, Computer science – Research and Development 33, p. 231-232.

Doersch, C. et al. 2012. What makes Paris look like Paris, ACM Transactions on Graphics, p. 1-9.

Goel, A et al. 2012. Are buildings only instances? Eighth Indian Conference on Computer Vision 2012, Bombay, India.

Hecht, R. et al. 2015. Automatic identification of building types based on topographic databases – a comparison of different data sources, International Journal of Cartography 1, p. 18-31.

Kingma, D. P.; Lei Ba, Jimmy 2015. Adam: A method for stochastic optimization, International Conference on Learning Representations 2015, San Diego, USA.

Lee, S.; Maisonneuve, N. et al. 2015. Linking past to present: Discovering style in two centuries of architecture. IEEE International Conference on Computational Photography (ICCP) (2015).

Li, C. et al. 2018. Estimation of building age from Google street view images using deep learning, 10th International Conference on Geographic Information Science 2018, Melbourne, Australia.

Loga, T. et al. 2015. Deutsche Wohngebäudetypologie: Beispielhafte Maßnahmen zur Verbesserung der Energieeffizienz von typischen Wohngebäuden, IWU, 2. Auflage, Darmstadt 2015.

Malhotra, A. et al. 2022. Information modelling for urban building energy simulation – A taxonomic review, Building and Environment 208, p. 108552.

Remmen, P. et al. 2016. Import and export for dynamic building performance simulation in Modelica, Building Simulation and Optimization Conference 2016, Newcastle upon Tyne, United Kingdom

TLBG, 2022. Thüringer Landesamt für Bodenmanagement und Geoinformation, 2022. https://tlbg.thueringen.de/ (last access: 28.02.2022).

University of Oxford 2022. Department of Engineering Science, The oxford buildings dataset, https://www.robots.ox.ac.uk/~vgg/data/oxbuildings/ (last access: 28.02.2022)

Wikimedia Foundation, 2022. Liste der Kulturdenkmale in Königsee. https://de.wikipedia.org/wiki/Liste_der_Kulturdenkmale_in_K%C3%B6nigsee (last access: 28.02.2022).

Zeppelzauer, M. et al. 2018. Automatic prediction of building age from photographs, International Conference on Multimedia Retrieval 2018, Yokohama, Japan.