# Visual Estimation of Building Condition with Patch-level ConvNets

David Koch
Kufstein University of Applied Sciences
david.koch@fh-kufstein.ac.at

Miroslav Despotovic
Kufstein University of Applied Sciences
miroslav.despotovic@fh-kufstein.ac.at

Muntaha Sakeena
St.Pölten University of Applied Sciences
msakeena@fh-kufstein.ac.at

Mario Döller
Kufstein University of Applied Sciences
mario.doeller@fh-kufstein.ac.at

Matthias Zeppelzauer
St.Pölten University of Applied Sciences
m.zeppelzauerh@fhstp.ac.at



## ABSTRACT

The condition of a building is an important factor for real estate valuation. Currently, the estimation of condition is determined by real estate appraisers which makes it subjective to a certain degree. We propose a novel vision-based approach for the assessment of the building condition from exterior views of the building. To this end, we develop a multi-scale patch-based pattern extraction approach and combine it with convolutional neural networks to estimate building condition from visual clues. Our evaluation shows that visually estimated building condition can serve as a proxy for condition estimates by appraisers.

## CCS CONCEPTS

• **Information systems → Information retrieval**; • **Computing methodologies → Visual content-based indexing and retrieval**; *Supervised learning*; *Neural networks*;

## KEYWORDS

Content-based image retrieval, visual pattern extraction, image classification, visual building analysis, building condition estimation, single-family-housing, deep learning, regression models.

**ACM Reference Format:**
David Koch, Miroslav Despotovic, Muntaha Sakeena, Mario Döller, and Matthias Zeppelzauer. 2018. Visual Estimation of Building Condition with Patch-level ConvNets. In *Multimedia for RETech'18: Workshop on Multimedia for Real Estate Tech, June 11, 2018, Yokohama, Japan.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3210499.3210526

## 1 INTRODUCTION

A house is made up of many characteristics, all of which may affect its value. Hedonic regression analysis is typically used to estimate the marginal contribution of these characteristics and to predict real estate prices [27]. There exist numerous studies on hedonic pricing in theoretical and empirical work, see for example [22, 26, 27]. In general the hedonic price function takes the form

$$P_i = f(S_i, L_i, N_i) \tag{1}$$

where $P_i$ is the logarithm 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 this paper, we focus on structural housing characteristics (e.g. square footage, number of bathrooms, age) and in particular the *condition* of a building with the aim to assess building condition automatically. The main research question of our work is: *can the condition of the building be estimated reliably in an automated fashion from an unconstrained photograph of the building by computer vision algorithms*? This focus is motivated by the following two observations: Firstly, the condition of a property has a significant influence on its value, which is reflected also in hedonic price models, see for example [13, 19]. There are a number of studies, particularly in connection with age and depreciation, see [2, 3, 6, 10, 11, 14, 18, 23, 31], which show that the condition of a building is an important factor for real estate valuation. Secondly, the condition is usually assessed by appraisers or brokers subjectively. This makes condition differ substantially from other variables like, e.g. year of construction, which can usually be estimated exactly and do not offer room for interpretation. The same applies to the variables like the presence of a balcony and the number of rooms, etc. The condition, however, is a variable that is not clearly and objectively defined and that often has a subjective bias. We hypothesize that real estate image analysis (REIA) is a promising means to capture the condition of a building in a more standardized and objective way.

Based on the work in [32], we present an approach for the estimation of building condition from image patches. Experiments with a large dataset show that useful visual clues for the estimation of condition can be extracted automatically and that the estimated condition has a positive impact on price estimation.

## 2 RELATED WORK

Real estate prices can only be estimated based on a proper valuation model and highly depends on the selection of characteristics which are likely to reflect the real value. Price estimation usually takes


*Multimedia for RETech'18, June 11, 2018, Yokohama, Japan*

ACM ISBN 978-1-4503-5797-5/18/06…$15.00
https://doi.org/10.1145/3210499.3210526

building parameters in the form of readily available metadata into account, such as building age, number of stories and number of units [30]. Rich literature exists that focuses on price estimation based on metadata [15–17].

Artificial neural networks are capable of learning complex regression and classification models based on labeled training data and have shown to generalize well to unseen data. Thus, they have been increasingly used for price estimation of properties based on building metadata in the past and provide a satisfactory proxy for hedonic models [7, 8]. See [25] for a comparison between hedonic models and artificial neural networks. Neural networks have further shown to be extremely powerful for the extraction of visual features and image classification [20]. Only little work has been done so far on leveraging visual information for real estate evaluations (real estate image analysis, REIA). First REIA approaches include Eman et al. [1] who combine visual features extracted from photographs with metadata provided by real estate companies. They extract SURF features [4] from indoor and outdoor pictures of buildings, combine them with traditional building parameters and train a regression-based neural network from these inputs for price estimation. Omid et al. [24] analyze interior and exterior photos to assess the luxury level of a building and further use this estimate as additional parameter in traditional house price estimation. In comparison with Eman et al. they used a much larger dataset with multiple interior and exterior photos of specific houses. Despotovic et al. [9] leveraged visual building features derived from outdoor building pictures to predict the approximate heating energy demand of a building by using convolutional neural networks (CNNs).

Our goal is to predict the condition of a house automatically from an unconstrained exterior view. This is based on the assumption that the actual condition of a building is reflected in its visual appearance (e.g. impurities of the facade, state of windows, etc.). Contrary to previous works, we employ only a single exterior image for our analysis and no interior images (often not available in practice). This challenges the analysis because weather and lighting conditions strongly influence the visual appearance. Furthermore, instead of predicting the price directly, we assess the relative *discount* of price given by the building condition. Thus, the actual real estate prices (also difficult to get in practice) are not necessary for our method.

## 3 APPROACH

We assume that the condition of a building can be detected by means of various visual clues, such as uneven structure and cracks in the facade, damaged painting of doors and windows, weathering of the roof, etc. We have a priori carried out a qualitative analysis of building images in order to examine which visual features potentially relate to the condition of a building. To this end, we have studied building images in detail to determine representative patterns. Figure 1 shows patches representing three different condition classes (excellent, good, poor) for different building elements (e.g. roofs and windows). We observe that aside from the facade of the building architectural elements, such as windows, doors, stairs, plinths, and roofs have potential to indicate its condition. The identified clues are rather at a local than a global level of a building. We thus design our approach to analyze the images locally in a patch-wise manner. We first extract patches from the images


**Figure 1: Visual appearance of the building elements roof (top row), facade (middle row), and window (bottom row) for the conditions very good, good, and poor.**

in a dense manner and then select those which are most likely to contain meaningful structure (Section 3.1). Next, we train a CNN from the models that predicts different condition categories from the individual patches. Finally, to obtain an estimate for an entire building, we classify all its patches and perform majority voting on the patch-level predictions (Section 3.2).

### 3.1 Patch Extraction and Selection

The input to our approach are unconstrained images of the exterior of a building. The buildings may appear in different scales, perspectives, lighting condition, and parts of the buildings may be cropped. In a first step, we extract patches from the images. We apply regular sampling (as in Dense Scale Invariant Feature Transform, DSIFT [21]). At each location we extract overlapping patches of different scales to account for scale variations, similarly to PHOW features [5]. This scheme results in a large set of partially redundant patches.

To reduce the number of patches and to increase the heterogeneity of captured visual patterns, we apply the following patch selection strategy. We extract the SIFT descriptor for each patch and apply k-means clustering for all patches of a given input image. From the resulting $k$ clusters we extract the patch which is closest to the respective cluster centroid as a representative of the cluster. Only these $k$ representative patches are considered further. To remove patches with low information content, i.e. low contrast (e.g. in areas of sky, roads) we take the norm of the SIFT descriptor as an indicator. Since the SIFT descriptor represents a gradient histogram, descriptors with small norm contain less edge-like structures than descriptors with larger norm. We select a fixed percentage $t$ of those representative patches with the largest norm for further processing.

The resulting set of patches capture heterogeneous patterns with high contrast. Using patches with multiple scales facilitates capturing building elements of different sizes. Overlap between the patches increases the spatial coverage of the facades. Patch extraction shows that most parts of the facades are captures well and densely. Homogeneous regions (e.g. sky but also homogeneous regions on the facade or on the roof) are widely omitted. Nevertheless, patterns not related to buildings may still be contained in this set, such as parts of cars, traffic signs, trees and bushes. To better filter out such irrelevant patches, we train a classifier that categorizes patches into 12 different classes of non-related objects and one class of building-related patterns. Non-related objects include e.g. cars, trees, people, asphalt, and poles. For classification we employ a Convolutional Neural Network (CNN), i.e., Alexnet [20] which was

pre-trained on ImageNet. We modify the output layer to fit our target classes and fine tune the network with a training dataset containing 8000 labeled patches. The fine tuned network is applied to all candidate patches extracted by our approach.

### 3.2 Building Condition Prediction

The prediction of building condition is a two-stage approach. In the first stage we classify each input patch that passes the patch selection from Section 3.1 into a set of pre-defined condition classes. For this purpose, we train a network to classify individual patches. We employ residual networks (ResNet50 [12]) due to their lower number of parameters (as e.g. compared to AlexNet) to speed up training. We adapt the output layer to the building condition classes and iteratively fine tune all layers of the network. To increase the heterogeneity of the training patches, we apply data augmentation (e.g. flipping, cropping, brightness variations) to the patches.

In the second stage, we infer the condition of an entire building from the predictions at patch level. First, each patch extracted for a building is classified into a distinct condition category. Second, all patch-wise predictions are aggregated to obtain a final prediction for the entire building. Aggregation is either performed by majority voting (MV) on the patch-wise predictions or by averaging the class likelihoods (LH), i.e. the outputs of the softmax layer of the individual patches. To make aggregation more selective we filter out ambiguous patches, i.e. patches for which the class likelihoods do not show a clear winner. A patch is considered ambiguous if the difference between the highest and the second highest class likelihood is below a certain threshold (0.25 in our experiments).

## 4 EXPERIMENTAL SETUP

### 4.1 Dataset

For our experiments, we use data provided by a renowned Austrian real estate software provider. The dataset consists of RGB images of single-family houses with different resolutions and qualities as well as JSON files containing house characteristics (metadata)[1]. The images show the exterior of the houses and were taken from different angles, distances, and perspectives by different real estate experts. Weather and lighting conditions vary across the images. The available metadata include among others condition scores for all buildings ranging from c1 (best) to c9 (worst).

The first two categories c1 and c2 refer to buildings that are as good as new and free of defects. Category c3 and c4 means normal conditions, i.e. only usual maintenance works are necessary. Categories from c5 onwards refer to buildings that need different amounts of repairs. Due to the semantic similarity of the above mentioned classes and the imbalanced class cardinalities (see also 1) we group the nine categories into three classes: "A" (good condition), "B" (normal condition), and "C" (needs repairs) and use these three condition classes for our experiments.

We partition the dataset randomly into a training, validation, and test set. Table 1 shows the distribution of the different partitions and categories as well as the aggregation of condition categories to the three target classes ("A", "B", and "C"). Table 2 shows the number of individual houses for each partition and target class of the dataset. Although for some houses more than one image exists, all images are independently processed and assessed.

[1]The images used in this work are not under creative commons license and thus cannot be shared, we will, however, publicly share extracted features and trained networks under https://phaidra.fhstp.ac.at/detail_object/o:2960.

| Category | Training | Validation | Test | Target Class |
|---|---|---|---|---|
| c1 | 2416 | 387 | 702 | A |
| c2 | 83 | 24 | 35 | A |
| c3 | 3212 | 433 | 975 | B |
| c4 | 121 | 23 | 33 | B |
| c5 | 1298 | 184 | 355 | C |
| c6 | 102 | 15 | 34 | C |
| c7 | 120 | 15 | 44 | C |
| c8 | 2 | 2 | 2 | C |
| c9 | 13 | 3 | 3 | C |

**Table 1: Employed dataset: distribution of images across the three partitions of the dataset (train, validation, test) and the condition categories. The last column shows the aggregation of condition categories to target classes for classification.**

| Partition | Target Class | Unique Houses | Available Images |
|---|---|---|---|
| Training | A | 1272 | 2499 |
| | B | 1931 | 3333 |
| | C | 1040 | 1535 |
| Validation | A | 206 | 411 |
| | B | 276 | 456 |
| | C | 205 | 219 |
| Test | A | 376 | 737 |
| | B | 576 | 1008 |
| | C | 391 | 438 |
| **Sum** | | **6273** | **10636** |

**Table 2: Dataset characteristics: number of unique houses for each partition and target class of the dataset.**

### 4.2 Setup & Training

For training, we extract multiple patches of different sizes from the images in the training set and select the $t = 21\%$ highest-contrast patches from $k = 50$ patch clusters, obtained by the patch selection strategy from Section 3.1. These parameters were found to represent a good tradeoff between the amount of resulting data, redundancy in the patches and spatial coverage. Experiments with smaller datasets (lower percentage, less clusters) lead to weaker results. Larger datasets were not tested so far due to computational costs. For classification, we adapt the ResNet50 model (pre-trained on ImageNet) [12] and shrink the output layer to three neurons, referring to classes "A", "B" and "C" (see Section 4.1). From the training patches we re-train all layers of the ResNet50 model. Training is performed for 30 epochs with a learning rate of 0.0001, a momentum of 0.9 and a decay of 0.0005. We apply a comprehensive data augmentation on all patches including cropping, flipping, scaling as well as brightness and color transformations. The total number of training patches is 153,356. The same procedure is applied to the validation images resulting in a set of 21,995 validation patches. We implemented the approach in Matlab and used the MatCovNet framework [29] for training the network. The experiments have

| | | Predicted | | |
|---|---|---|---|---|
| | | A | B | C |
| True | A | 505 | 205 | 25 |
| | B | 227 | 713 | 67 |
| | C | 67 | 163 | 206 |

| | | Predicted | | |
|---|---|---|---|---|
| | | A | B | C |
| True | A | 491 | 211 | 33 |
| | B | 214 | 711 | 82 |
| | C | 64 | 166 | 206 |

**Table 3: Test results: confusion matrices showing true positive and false classifications on the test images with left: majority voting (MV), right: average likelihood (LH).**


**Figure 2: Patches with a clear prediction for a certain building condition (likelihood > 0.99). The network successfully captures patterns related to the visually apparent condition.**

been performed on a workstation with Ubuntu 14 OS, 64 GB RAM and an NVIDIA GTX 1080Ti.

### 4.3 Evaluation & Research Questions

Our evaluation comprises three basic research questions:

(1) How accurately can the three conditions captured by the target classes ("A": good condition, "B": normal condition, and "C": needs repairs) be differentiated automatically?
(2) Does the network learn meaningful visual patterns that correlate with the visually assessable condition?
(3) How does the automatically extracted building condition compare to the condition provided by experts in the prediction of cost-related parameters?

To account for the first question, we compute the classification accuracy and analyze the classification confusions in Section 5.1. To investigate the second question, we investigate patches classified with different confidences by the network to analyze on which visual information its decisions are based on (see Section 5.2). To investigate the third question, we design a regression model to predict the discount of the building (which is closely related to its cost) and compare regression performance between the model based on automatically extracted condition and the model based on expert assessments (see Section 5.3).

## 5 RESULTS

### 5.1 Classification Results

We compute the confusion matrix and classification accuracy to examine the performance of our approach in predicting building condition. Table 3 shows the confusion matrices based on majority voting (MV) and average likelihood (LH), see Section 3.2 for the three target classes. Confusion matrices contain the true labels along the vertical axis and the predicted labels along horizontal axis. Along the diagonal are the correct predictions (true positives) while off-diagonal elements show the incorrect predictions[2].

Results from Table 3 show that MV is the slightly better aggregation strategy compared to LH, which may be due to the fact that MV is less prone to noise and outliers. The confusion matrix of MV shows that 205 images of class A are wrongly predicted as class B while just 25 images are wrongly predicted as class C. Neighboring classes (A&B, B&C) are more often confused then non-neighboring classes (A&C, C&A). Misclassifications between neighboring classes are to a large part due to fuzzy class boundaries leading to ambiguities. The rather low number of confusions between non-neighboring classes (A&C), however, shows that when ambiguities along the class boundaries can be neglected our approach yields a strong discriminative abilities. The overall accuracy based on MV is 65.38% and for LH is 64.65%. These results are significantly higher than the random baseline for the test set of 46% (according to the zero rule). These results show the our approach is able to extract discriminative visual patterns that are relevant to distinguish different building conditions.

### 5.2 Qualitative Analysis

To investigate which characteristic visual patterns are learned by the network for each building condition class, we investigate patches with different likelihoods and confidences. Figure 2 shows examples of correctly predicted patches from the test set with highest likelihoods for their class (likelihood > 0.99). The patches show typical characteristics for the respective condition class, whereby the best building condition is reflected by rather newer and more modern architecture. Similarly, patches indicative for the poorest building condition belong to rather old looking and little maintained buildings. These results show that the network learns characteristic patterns that reflect the visual appearance of different building conditions. We further compute the correlation between building age and predicted condition and find a strong positive correlation of 0.616. This is reasonable, as older buildings are more likely to exhibit a poorer condition than new buildings. Another observation is that the correctly predicted patches tend to capture rather large parts of the buildings. This shows that a certain amount of contextual information is beneficial to predict the condition.

Next, we select the most ambiguous patches from the test set, i.e. those patches which cannot be assigned clearly to a class and from which no information about the condition of the house can be deduced. See Figure 3 for example patches. These patches capture less expressive patches and often do not represent facade elements. Furthermore, for some patches the facade elements are occluded by trees and other objects making them less useful.

In addition to the correctly predicted and ambiguous patches, we also analyze the most confident misclassifications of our approach. Figure 4 shows examples for the two non-neighboring and thus

[2]Note that the total number of images per class may vary from the numbers in Table 1 and 2 because in some cases patch filtering (see Section 3.1) removes all patches from a given input image. Such images are not included in the confusion matrices.

**Figure 3: Ambiguous patches represent less expressive structures from which the condition can hardly be estimated.**


**Figure 4: Misclassification between non-neighboring classes (likelihood >0.99). Although predictions are wrong, they are to a great extent visually plausible which demonstrates that the visual patterns learned by the network are meaningful.**

most distant classes ("A" and "C'"). From the visual examination of the patches we can understand the appraiser's assessment only to a limited extent. The prediction of the network seems to be even more plausible than the appraiser's assessment. There are two insights from this observation. Firstly, appraisers usually estimate the condition not only on the basis of the visual appearance of the external facade, but also on the basis of the overall impression of the building (interior space, technical equipment, etc.) which may make a difference in the assessment. Secondly, the evaluation of the condition is determined individually by each appraiser and thus, different assessments are not necessarily comparable and may live on different subjective scales. This fact becomes apparent in the context of automatic classification, for which this individual bias is reduced due to the training from a large dataset.

## 5.3 Cost Prediction of Buildings

In addition to the attributes of the property, our data set also contains a value indication. All properties were valued by appraisers using the "cost approach". The cost approach is the standard method used in Austria to determine the market value of single-family homes. The cost approach consists of an estimation of the land value plus the replacement cost of the building in relation to a comparable property. Replacement cost represents the estimated costs to construct (construction cost, ancillary costs, etc.) for a new building. These costs have to be reduced in consideration of the age, condition and functional and economic obsolescence of the building by a certain *discount* [28].

Figure 5 shows the discount in percent of replacement cost (of a new building) according to condition for the true model and the predicted conditions (with majority voting, MV and average class likelihoods, LH). Both approaches are almost identical but more importantly, both models are very similar to the true assessment.

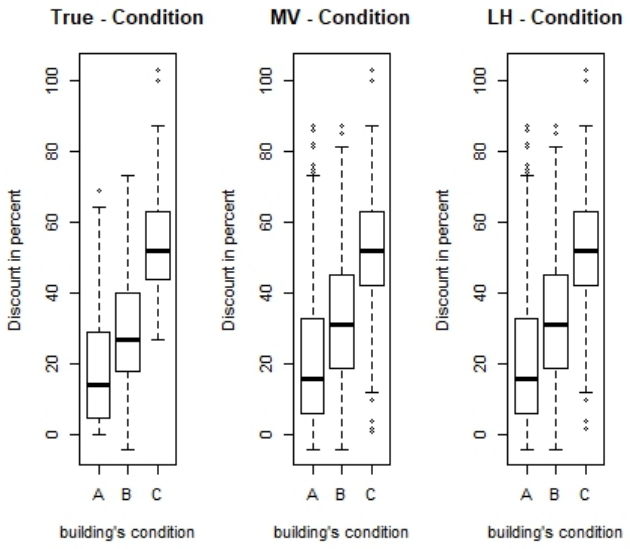


**Figure 5: Discount in percent of replacement cost (for new buildings) according to condition for the true condition by appraisers and the predicted conditions from MV and LH.**

This simple descriptive analysis already shows that the true model and our prediction from the images are highly correlated.

Finally, we build a regression model to predict the discount of a building and feed it with the true and the predicted condition. As an additional parameter the year of construction is added in all experiments. Table 4 shows the results of the regression for the three models ("True" refers to the regression result from the manually assessed condition and "MV" and "LH" refer to the predicted conditions. In all three models we have modeled the percentage discount as a function of condition and year of construction. The MV and LH models are almost identical to the model built upon the appraisers' assessments. The adjusted $R^2$ is in all three models around 0.6, which means that about 60% of the deviation can be explained by the models. The variable condition is coded as a dummy variable. The starting point is the state "A" and integrated in the intercept. Condition "B" has a coefficient of -0.049 in the "True" model (see row starting with "true: B/A"), which means that the discount for condition B is 4.9% higher than for condition A. In condition C, this discount is even 9.0 percent higher (see: "true: B/A": -0.090). This reflects well the actual relation that poorer conditions lead to larger discounts. In all models, the variable condition is significant and has the correct sign. It can be seen that the same results are obtained by image recognition and the appraiser's assessment with respect to discount. This shows that the automatically extracted condition may be a creditable substitute for the appraiser's assessments.

# 6 CONCLUSION

We have presented a first method for the estimation of building condition from unconstrained photographs. Our approach operates on a patch-level at different scales and tries to learn characteristic visual patterns related to different condition categories. Experiments

| | True | MV | LH |
|---|---|---|---|
| (Intercept) | -11.471*** | -12.132*** | -12.149*** |
| | (0.314) | (0.265) | (0.264) |
| year of construction | 0.006*** | 0.006*** | 0.006*** |
| | (0.000) | (0.000) | (0.000) |
| true: B/A | -0.049*** | | |
| | (0.007) | | |
| true: C/A | -0.090*** | | |
| | (0.011) | | |
| predictedMV: B/A | | -0.043*** | |
| | | (0.007) | |
| predictedMV: C/A | | -0.077*** | |
| | | (0.011) | |
| predictedLH: B/A | | | -0.045*** |
| | | | (0.007) |
| predictedLH: C/A | | | -0.077*** |
| | | | (0.011) |
| adj. $R^2$ | 0.602 | 0.599 | 0.599 |
| sigma | 0.133 | 0.133 | 0.133 |
| F | 1021.258 | 1009.438 | 1010.674 |
| p | 0.000 | 0.000 | 0.000 |

$^{***}p < 0.001, ^{**}p < 0.01, ^{*}p < 0.05$

**Table 4: Regression results: Column two to four show the effect of the coefficients on the discount in percent of replacement cost (new building). Adj. $R^2$ is the adjusted coefficient of determination.**

yield a classification accuracy of 65% and show that the patterns learned by the network are visually meaningful. Image-based prediction delivers equally good results in discount estimation as the assessments of the appraisers and thus predicted condition can serve as a proxy for condition estimates of appraisers. Future work will include the extension of the approach from patch-level to a global image level and to leverage network visualization techniques [33] to identify characteristic patterns as well as multi-task learning of age and condition to exploit mutual relations.

## ACKNOWLEDGMENTS

We thank Sprengnetter Austria GmbH for providing real estate images and meta-data for our experiments. This work was supported by the Austrian Research Promotion Agency (FFG), Project No. 855784 and Project No. 856333.

## REFERENCES

[1] Eman H. Ahmed and Mohamed Moustafa. 2016. House Price Estimation from Visual and Textual Features. (2016).
[2] Marcus Allen. 1997. Measuring the Effects of "Adults Only" Age Restrictions on Condominium Prices. *Journal of Real Estate Research* 14 (1997), 339–346.
[3] Andrew E Baum. 1993. Quality, Depreciation, and Property Performance. *Journal of Real Estate Research* 8, 4 (1993), 541–566.
[4] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. 2008. Speeded-up robust features (SURF). *Computer vision and image understanding* 110, 3 (2008), 346–359.
[5] Anna Bosch, Andrew Zisserman, and Xavier Munoz. 2007. Image classification using random forests and ferns. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*. IEEE, 1–8.
[6] Charles Carter, Zhenguo Lin, Marcus Allen, and William Haloupek. 2011. Another Look at Effects of "Adults-Only" Age Restrictions on Housing Prices. *The Journal of Real Estate Finance and Economics* 46 (2011), 1–16.
[7] Xiaochen Chen, Lai Wei, and Jiaxin Xu. 2017. House Price Prediction Using LSTM. *The Computing Research Repository (CoRR)* (2017).
[8] Vincenza Chiarazzoa, Leonardo Caggiania, Mario Marinellia, and Michele Ottomanellia. 2014. A Neural Network based model for real estate price estimation considering environmental quality of property location. *Transportation Research Procedia* 3 (4 2014), 810–817.
[9] Miroslav Despotovic, Muntaha Sakeena, David Koch, Mario Döller, and Matthias Zeppelzauer. 2018. Poster abstract: predicting heating energy demand by computer vision. *Computer Science - Research and Development* 33, 1 (01 Feb 2018), 231–232.
[10] Malcolm Harrison. 2004. Defining Housing Quality and Environment:Disability, Standards and Social Factors. *Housing Studies* (2004).
[11] Jia He and Jing Wu. 2016. Doing well by doing good? The case of housing construction quality in China. *Regional Science and Urban Economics* 57 (2016), 46–53.
[12] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep Residual Learning for Image Recognition. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
[13] Shanaka Herath, Johanna Choumert, and Gunther Maier. 2015. The value of the greenbelt in Vienna: a spatial hedonic analysis. *The Annals of Regional Science* 54, 2 (2015), 349–374.
[14] Shinichiro Iwata and Hisaki Yamaga. 2008. Rental externality, tenure security, and housing quality. *Journal of Housing Economics* 17, 3 (sep 2008), 201–211.
[15] Michal Juszczyk. 2015. Application of committees of neural networks for conceptual cost estimation of residential buildings. *AIP Conference Proceedings* 1648 (2015).
[16] Michal Juszczyk. 2017. The Challenges of Nonparametric Cost Estimation of Construction Works With the Use of Artificial Intelligence Tools. *Creative Construction Conference 2017.* (2017).
[17] Michal Juszczyk, Theodore Simos, and Charalambos Tsitouras. 2016. Application of PCA-based data compression in the ANN-supported conceptual cost estimation of residential buildings. *AIP Conference Proceedings*. 1738, 1 (2016).
[18] John R. Knight and C.F. Sirmans. 1996. Depreciation, Maintenance, and Housing Prices. *Journal of Housing Economics* 5, 4 (dec 1996), 369–389. https://doi.org/10.1006/jhec.1996.0019
[19] David Koch and Gunther Maier. 2015. The influence of estate agencies' location and time on Internet: An empirical application for flats in Vienna. *Jahrbuch fur Regionalwissenschaft* 35, 2 (2015), 147–171.
[20] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*. 1097–1105.
[21] David G Lowe. 1999. Object recognition from local scale-invariant features. In *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, Vol. 2. Ieee, 1150–1157.
[22] Stephen Malpezzi. 2003. Hedonic Pricing Models: A Selective and Applied Review. In *Housing economics and public policy: Essays in Honour of Duncan Maclennan* (1 ed.), Tony O O'Sullivan and Kenneth Gibb (Eds.). Blackwell Science Ltd, Oxford, 67–89.
[23] Benedetto Manganelli. 2013. Maintenance, Building Depreciation and Land Rent. *Architecture, Building Materials and Engineering Management* 357 (2013), 2207–2214.
[24] Omid Poursaeed, Tomas Matera, and Serge J. Belongie. 2017. Vision-based Real Estate Price Estimation. *CoRR* abs/1707.05489 (2017).
[25] H. Selim. 2009. Determinants of house prices in Turkey: Hedonic regression versus artificial neural network. *Expert Systems with Applications* 36 (2009), 2843–2852.
[26] G Stacy Sirmans, Lynn MacDonald, and David A Macpherson. 2010. A Meta-analysis of Selling Price and Time-on-the-Market. *Journal of Housing Research* 19, 2 (2010), 139–152.
[27] G Stacy Sirmans, David A Macpherson, and Emily N Zietz. 2005. The Composition of Hedonic Pricing Models. *Journal of Real Estate Literature* 13, 1 (2005), 3–43.
[28] TEGoVA. 2016. European Valuation Standards 2016. *The European Group of Valuers Associations* 8 (2016), 1–378.
[29] Andrea Vedaldi and Karel Lenc. 2015. Matconvnet: Convolutional neural networks for matlab. In *Proceedings of the 23rd ACM international conference on Multimedia*. ACM, 689–692.
[30] Yung Yau. 2008. Building conditions in Yau Tsim Mong, Hong Kong: Appraisal, exploration and estimation. *Journal of Building Appraisel* 3 (2008), 319–329.
[31] Velma Zahirovich-Herbert and Karen M. Gibler. 2014. The effect of new residential construction on housing prices. *Journal of Housing Economics* 26 (2014), 1–18.
[32] M. Zeppelzauer, M. Despotovic, M. Sakeena, D. Koch, and M. Döller. 2018. Automatic Prediction of Building Age from Photographs. In *ACM International Conference on Multimedia Retrieval (ICMR)*. ACM. https://arxiv.org/abs/1804.02205
[33] Bolei Zhou, Aditya Khosla, Agata Lapedriza, Aude Oliva, and Antonio Torralba. 2016. Learning deep features for discriminative localization. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2921–2929.