From platforms to price: the impact of condition prediction using computer vision on real estate pricing models

Article

Accepted Version

Koch, D., Despotovic, M. ORCID: https://orcid.org/0000-0001-6282-9672, Thaler, S. ORCID: https://orcid.org/0000-0001-7743-3451 and Zhang, Z. (2025) From platforms to price: the impact of condition prediction using computer vision on real estate pricing models. Journal of European Real Estate Research. ISSN 1753-9277 doi: 10.1108/JERER-03-2024-0013 Available at https://centaur.reading.ac.uk/125424/

To link to this article DOI: http://dx.doi.org/10.1108/JERER-03-2024-0013

Publisher: Emerald

From Platforms to Price: The Impact of Condition Prediction using Computer Vision on Real Estate Pricing Models

Abstract

This study investigates the use of images from online real estate platforms in Central European real estate pricing models, exploring both the challenges and advantages of utilizing such data. It specifically focuses on the impact of manually and algorithmically classifying building conditions on the accuracy of price predictions. The research demonstrates that incorporating image-based condition assessments can significantly improve model quality in both, the manual and algorithmical condition assessment, even with limited data. Additionally, the article identifies and addresses obstacles encountered during the study using a computer vision classification algorithm, providing guidance for future research in real estate and computer vision to refine pricing models further.

Keywords: Real Estate Pricing, Image-based Condition Assessment, Computer Vision Obstructions

1. Introduction

In numerous European countries, the opacity of real estate market data starkly contrasts with the transparency found in Anglo-American markets, where multiple listing services (MLS) facilitate a comprehensive database for automated valuation models. Due to the absence of such data sources, many European models rely on data from online platforms, which, while readily available and cost-effective, often present biases such as listing only offer prices rather than transaction prices and overly favorable property presentations. Despite these drawbacks, the inclusion of property images offers a rich, albeit underutilised, source of information.

Existing research, as highlighted by [32], [42], and [58], acknowledges the use of images in real estate analysis. Yet, the exploration of their impact on pricing models remains limited, particularly due to the lack of standardisation in image quality utilised in studies, such as the uniform images from Google Street View mentioned in [32]. This article seeks to bridge this gap by focusing on the practical application of real estate images from online platforms, despite their inherent imperfections, to enhance the accuracy and relevance of valuation models and price indices. Thus, our work specifically focuses on the derivation of a building condition variable that can be integrated in a hedonic pricing model and consquently be employed in Automated Valution Models (AVMs).

By employing Convolutional Neural Networks (CNNs) to analyze and categorize real estate images based on shapes, colors, and brightness, this study aims to integrate this visual data into pricing models. This approach not only acknowledges the diversity of real estate properties but also aims to improve the predictive accuracy of pricing models. The use of CNNs in real estate valuation could significantly reduce appraisal times and costs, offering immediate property valuations as suggested by [3].

Focusing on single-family homes, including their images, this article manually classifies the condition of properties using a standardized scheme. The significant impact of property maintenance on value, as investigated by [18], [56], and [9], underscores the importance of condition in valuation models. [34] further supports this, noting that incorporating condition variables enhances price estimation accuracy. Thus, we posit a positive impact on the pricing model's performance, aiming to improve data quality for automated valuation models, especially in regions heavily reliant on online platform data. This research addresses three key aspects:

- Does the standardised condition classification provide added value in a pricing model?

- Can computer vision replicate the condition assessment of a standardised classification?[1]
- Do the images contain any information such that it can explain parts of the regression residuals of the baseline pricing model?

Our study demonstrates that incorporating condition information from images into real estate pricing models significantly enhances their accuracy, challenging the widespread issue of inaccurate valuations identified within the sector by [1] and [5]. Accurate valuations are vital, as highlighted by [39], because house prices impact the quality of life for residents and influence various economic cycles. Additionally, as [30] notes, they play a crucial role in forecasting key macroeconomic indicators, including inflation and economic output.

We also found that CNNs are effective in classifying images for this purpose. Our analysis has identified significant challenges inherent to this approach, providing valuable insights that could aid researchers in refining their methodologies. The use of images from real estate platforms, coupled with established classification schemes, clearly leads to a more accurate pricing model. The results of our article and the literature contribute to an improvement of the quality of AVM models, especially in the banking industry and the construction of real estate price indices for evaluating economic developments.The integration of CNNs for analyzing real estate images not only represents a methodological extension of existing valuation approaches but also enables an objective classification of condition attributes, which is important for traditional hedonic models by providing interpretable and standardized features.

This article includes a literature review on the use of computer vision for real estate valuation, followed by a detailed explanation of our methodology. We then present our findings, discuss the challenges encountered, and suggest ways to navigate these in future research as well as in practice. The article concludes with a summary of our critical discoveries and their implications for advancing the field.

2. Computer Vision and Pricing Models - A Literature Review

The hedonic pricing model, introduced by [25] and [46], stands as the benchmark in real estate pricing, attributing specific value contributions to distinct property characteristics (see [13] for a detailed review). This model has historically emphasised factors like neighbourhood attributes, environmental and structural

[1] It is not the objective to compare subjective (human) and objective (machine learning algorithm) classification accuracy.

amenities, and social influences on property values, with a thorough discussion on variable usage and model composition found in [48]. Despite the critical role of visual property inspection in valuation, the integration of images into pricing models remained overlooked until the advent of machine learning and computer vision technologies enabled effective prediction using digital real estate imagery.

CNNs, while not substituting human judgement in property valuation, provide a robust approximation of an image's attributes. Deep learning models, including CNNs, excel in analysing large datasets for pattern recognition, speech processing, computer vision, language understanding [27], and enhancing recommendation systems [33]. The methodology's growing significance in research is attributed to its capacity for extracting multiple features from data, particularly images [16].

A first broad overview of real estate and computer vision is given in [19], a more systematic approach review of current approaches in this field is given by [49]. Early research intersecting real estate and computer vision, such as the study by [29], utilize Google Street View images to evaluate neighborhood safety perceptions, identifying greenery as a significant safety indicator. Additionally, [57] show that pictures together with deep learning improve pricing . Further, [59] demonstrated that visible greenery significantly boosts Beijing's property prices by up to 10%.

[42]'s innovative approach, predicting property luxury levels through CNN analysis of interior and exterior images, showcased the potential of visual data to enhance valuation accuracy beyond traditional automated methods. Architectural style also influences property values significantly, as evidenced by [4] and [2], who noted a price premium associated with iconic architecture. Moreover, building age has been effectively predicted using exterior images through machine learning [58]. Also related to visual stimuli and property valuation [41] demonstrate that combining image and text data with traditional features enhances the accuracy of hedonic price models, achieving up to 25% improvement using advanced machine learning techniques. Additionally, the model's "black box" nature is clarified through agnostic methods, revealing that some housing attributes exhibit nonmonotonic and nonlinear relationships with prices.

The recent study by [32] incorporated architectural styles into a hedonic model, aligning machine learning assessments with expert evaluations to identify price premiums across various styles. However, this article shifts focus to the building condition—a factor less influenced by age and more by ongoing maintenance and renovation efforts, providing a distinct yet subjective assessment criterion that enhances automated valuation methods with image integration. Notably, the images utilised in this study are sourced directly from real estate platforms, mirroring real-world presentation rather than curated research datasets. [55] expands the existing methodologies to specifically analyse the factors on which an ML model focuses. In this process, building elements such as doors, windows, etc., are classified in advance. The focus here is on architectural style of the building and residual value of an AVM model. Unlike our area, in comparison to, for example, [55] and [57], the emphasis is on the explicit modelling of a property's condition derived through computer vision and consequently its impact on the price.

The challenge in image-based real estate assessment lies in the inherent bias of presentation styles, as every presentation format can influence valuation decisions [37]. For instance, [22] found that showcasing an apartment in disarray significantly impacts valuation negatively, regardless of its structural value, highlighting the presentation's effect on information processing and decision-making.

3. Methodology

This article concentrates on the valuation of single-family homes, emphasising the role of a condition variable derived from images on real estate platforms. This condition variable is assessed both manually by real estate professionals, according to standardised classifications, and automatically by computer vision algorithms. Our investigation extends to analysing whether the information extracted from images can reduce the regression residuals in a pricing model, thus potentially enhancing its accuracy (see also [55]). We also discuss the challenges associated with using computer vision for image assessment in the context of real estate economics.

Particularly in markets with lower transparency, such as the Austrian market which this study focuses on, the cost approach often underpins single-family house valuations [44, 36]. In general, the Austrian market is by far less transparent. For example, property listings do not provide specific addresses, etc., as is common in markets like the U.S. or U.K. In the U.S. and U.K. markets, the comparison method for single-family houses can be applied. However, in Austria, due to the limited availability and quality of data, this is currently not possible, and thus the cost approach must be used. Building upon this foundation, our research modifies the traditional hedonic pricing model into a hedonic pricing cost approach model, as similarly employed in [7, 58, 21]. Notably, the conventional hedonic pricing model incorporates 30-40 variables. In contrast, our adapted cost approach model simplifies this by using fewer predefined variables, focusing on the aggregate value of the land and the building, including its components. The valuation process involves estimating the building's replacement cost, considering factors like amenities and location, and then adjusting for depreciation related to age, quality, and other factors such as construction deficiencies.

3.1. Variable *Real Estate Condition* in images

In assessing a building's condition, images serve as a pivotal yet contentious variable in property valuation due to its subjective nature, especially when evaluated by an appraiser. This subjectivity often casts doubt on the reliability of the condition variable as used on real estate platforms. Despite this, the condition variable is crucial, highlighting the necessity for careful consideration in its application to valuation. Another variable, the year of construction, contrasts with the condition by being an objectively quantifiable factor, obtainable from a building's construction records, thus providing an unbiased element in our pricing model analysis.

The traditional method for assessing a building's condition involves a physical inspection by a broker or an appraiser, relying heavily on their expertise and experience. While this expert evaluation is generally regarded as accurate, it is not immune to subjective biases influenced by personal opinion.

When deriving a building's condition from images, we primarily analyse the exterior visual data. However, assessing both the exterior and interior conditions can offer a more comprehensive understanding of the building's overall state. The exterior facade's condition may not fully represent the interior's state, which could have undergone renovations, such as updated plumbing or flooring, thus affecting the property's valuation. The exterior appearance, however, remains a critical factor in the valuation process, underscoring the complexity of accurately assessing a property's condition from images alone.

Hence, we have the following different variables that can be considered in a pricing model:

- objective - directly quantifiable variables: examples from the literature of such variables, which also are predictable by a CNN, are the year of construction [58], the Heat Energy Demand [7] or architectural style [32].
- "objective" - not directly quantifiable variables which have to be assessed by an expert:
  - On site inspection: As a rule, during the inspection of the object, the expert assesses the condition and quality of the built structure. Then, a conclusion is drawn about the variable to be assessed. This assessment is necessary within the framework of a market value appraisal. Only an inspection can obtain a comprehensive classification of the property, however, this requires a relevant effort.
  - Assessment on the basis of an image: Only an image is presented and the assessment of the condition or quality is made on the basis of the image. This can lead to a distortion, as the view only documents parts and not the entirety of the object. The advantage, however, is that this

assessment is simple and fast and hence well suited for automated pricing models.

Based on the structural differences that constitute the different condition classes, these structural differences shall also be identifiable visually, which supports our hypothesis that computer vision with CNNs should make the (objective) condition assessment via images explicit. In our article we have made the assessment on the basis of an image of the exterior to see to what extent an effect on the price can be observed. This results in the following procedure:

1. Take images with features of the real estate from the platform.
2. Classify the front image into three condition categories: *bad*, *average* and *good* using a predefined classification scheme.
3. Include classification in the pricing model.
4. Train a CNN according to the three categories.
5. Include the CNN classification in the pricing model.

3.2. Used Data Set

We collected images and key housing features (such as the local average plot price in the municipality, plot size, building size, and year of construction) for 2,500 properties listed on Austrian online real estate platforms in 2019 (before Corona). This dataset, comprising images and detailed attributes of each property, was refined by selecting houses with plot sizes ranging from 200 to 4,000 square meters and constructed before 2018. After this filtering, we were left with 1,482 properties. Further narrowing our focus to those with suitable front views of the houses resulted in a final dataset of 960 properties, summary statistics for these properties can be found in Table 1 together with some example images for the three condition classes in Figure 1. As mentioned in the introduction, this dataset contains the typical disadvantages found on online real estate platforms.

The images underwent manual inspection to classify the condition of each building - good, average and poor - to determine the impact of subjective assessment on the accuracy of our pricing model. This classification process was conducted by researchers trained in real estate valuation, following a standardised scheme developed by the Association of Austrian Appraisers to minimise subjectivity [40] outlined in Table 2. This scheme assesses twelve different components of a building, encompassing both exterior (such as the roof and facade) and interior elements (including doors and housing technology). Given our dataset comprised solely exterior images, our classification focused on visible outdoor elements: the construction, roof, facade, windows, and exterior doors. Sample images representing each condition category are presented for reference.

Table 1

Descriptive Statistics of the Full Data Set for the Pricing Model: The table shows descriptive statistics for the variables with the number of observations used in the pricing model. The statistics are: minimum value, 1st quantile value, median, mean, 3rd quantile value, maximum value and the standard deviation of the variabless price, floor size, plot size, plot price and year of construction (Source: Authors own work).

Statistic	N	Min	1st Qu.	Median	Mean	3rd Qu.	Max	Std.
price	694	72,000	278,000	420,000	492,657	652,250	1,490,000	301,292
floor_size	694	45.0	124.0	160.0	176.7	216.8	399.0	72.3
plot_size	694	206.0	582.2	835.0	990.8	1204.0	3794.0	619.3
plot_price	694	8.7	56.4	112.4	190.4	254.2	1416.3	203.6
y_of_const	694	1570	1962	1979	1974	1999	2017	-

Of the 960 houses 290 buildings are considered to be in a good condition (class *good*), 561 in average condition (class *average*) and 109 buildings in bad condition (class *bad*). From that data set we obtained a sub data set where we excluded all real estate where the year of construction was missing. That data set then contains 694 houses where 233 where in condition class *good*, 395 in class *average* and 66 in class *bad*. This means that for training the CNN model, we use the full dataset with 960 houses, which contains all images for training the condition classification. For validation of the AVM model, we used the dataset containing all variables (especially the year of construction), resulting in a smaller dataset of 694 houses as the year of construction was missing for the excluded buildings.

Regarding the size of the data set for a hedonic model [10] state that many models in the literature for environmental studies are applied to small sample sizes, despite having a huge data base on house transactions. They name e.g. [6] where the authors investigate the evaluation of costs and benefits that come along with land use planning and which is based on 433 observations and 206 observations. Another example for a small sample size is [38] where they look at the valuation of national parks in urban areas with 641 observations. For additional examples see [10].

The article of [8] uses a hedonic model to estimate office rents in UK. The authors use a large number of independent variables but the article lacks a considerable sample size with a sample of 29. The authors state that it is a universal problem in early studies in the field. This is due to counteracting heterogeneity and, thus focusing on a specific geographically delineated area, consequently, focusing on a very homogeneous data set.

Another justification for a small sample size comes from the book of [45] and is delineated in [14].

(a) Class *good* (b) Class *average* (c) Class *bad*

Figure 1
Examples of the three classes that were assess by experts based on a standardised scheme (Source: immobilienscout24.at).

According to [45] and his rule of thumb, the appropriate samples size in behavioural studies should be larger than 30 and smaller than 500. More specifically, the sample should be ten times (or more) larger than the count of variables in the model. Another rule of thumb comes from [11] where the statement is that the sample size N should be greater than 50 + 8m where m is the number of independent variables (for testing multiple correlations) or 104 + m for testing predictors on the individual level. Both rules are fulfilled with our sample size. Since we use an adapted hedonic pricing cost approach model, we have fewer independent variables because e.g. we do not explain the location in terms of one variable but map it in the variable ”plot price”.

3.3. Hedonic Pricing Cost Approach Model

We adapted the Cost Approach to real estate valuation, transforming it into a hedonic pricing model. This approach values a property by adding the land’s value to the building’s value, adjusting for any applicable discounts. To account for geographical influences, we incorporated the average plot prices within the municipality into our regression-based pricing model. By including the average land price, we were able to bypass specific location variables, as these are implicitly reflected in the land price, with our analysis focusing primarily on the attributes of the building.

Our foundational pricing model employs a semi-log multiple regression model (model 1.a), predicting the logarithm of the price using independent variables such as *plot_price*, *plot_size*, *floor_size*, and *year* (the year of construction). We recognise a non-linear depreciation pattern in property values over time, hence the inclusion of both first and second-degree polynomials for the *year* variable in our model.

To assess the impact of a building’s visual condition on pricing accuracy, we introduced a dummy vari-

Table 2
TABLE OF THE THREE CONDITION CLASSES WITH THE CORRESPONDING FEATURES: THE TABLE IS BASED ON THE STANDARDS OF THE ASSOCIATION OF AUSTRIAN APPRAISERS [40] (SOURCE: AUTHORS OWN WORK).

Weight	%	***Bad***	***Average***	***Good***
Construction	50	Solid construction, contemporary building technology	good quality of materials, contemporary technology (thermal and sound insulation)	solid, quality materials, close to passive house technology, very good building physical properties
Roof	16	ventilated roof (cold roof), simple covering (sheet metal, clay roof tiles), foil sealing for flat roofs	ventilated roof (cold roof), with vapour barrier, thermal insulation, good covering (tiles, plastic-bonded roofing tiles, metal covering), bituminous waterproofing for flat roofs	like "average", but high-quality materials, elaborate construction, copper sheeting, green roofs, etc.
Facades	18	rubbed plaster, simple thermal insulating plaster, sheet metal sills	External thermal insulation composite system, plastic-bonded plaster, flashing, cladding, stone window sills, etc.	like "average", but noble materials and artistic design, curtain wall elements with rear ventilation, special thermal insulation
Windows and Exterior doors	16	Wood or plastic standard version	Hardwood, plastic, combination fittings, sunshade	Wood/aluminium windows, triple insulating glazing, sound insulation, sun protection, roller shutters (automatic operation), burglar protection

able for the condition based on (subjective) standards (*Condition_Standards*) into our pricing model (model 1.b). Furthermore, we incorporated a dummy variable for the condition assessed by a computer vision algorithm (*Condition_CNN*) (Model 1.c), exploring how these visual inspections enhance the predictive capability of our model.

$$\begin{aligned}
&(a)\ log(Price) = f(poly(year,2)1, poly(year,2)2, plot_size, floor_size, plot_price) \\
&(b)\ log(Price) = f(poly(year,2)1, poly(year,2)2, plot_size, floor_size, plot_price + Condition_Standards) \\
&(c)\ log(Price) = f(poly(year,2)1, poly(year,2)2, plot_size, floor_size, plot_price + Condition_CNN)
\end{aligned} \tag{1}$$

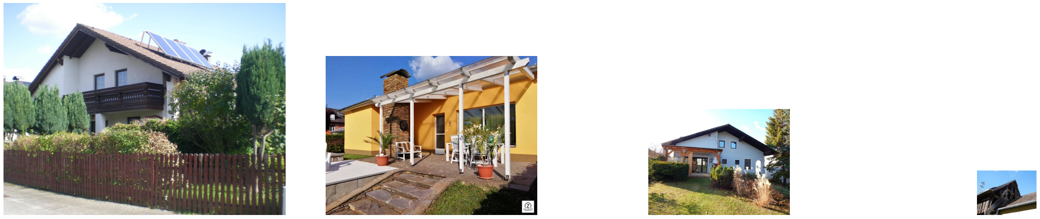
To draw conclusions on the precision of the models, we compare the adjusted R-squared of the models and the effect size in the models with the different condition variables. The main focus in this analysis was on the differences between our reference models from model 1.a) and the model 1.b) and the differences between the standards assessment and the assessment of the CNN of the building condition model 1.c). It is to note that in the three main pricing models only data that included the year of construction is considered.

3.4. CNN predictions

To train our machine learning algorithm we take the original data set including the buildings where the year of construction is missing and cut the corresponding images in patches of size 224 x 224 pixels (size of the input vectors of our CNN) with an overlap of 33.3%[2]. We included the incomplete data where the year of construction is missing as a CNN is highly dependent on the size of the data set for training (the ground truth) [35]. A higher number of training data increases the likelihood of learning specific features that a certain class has in common.

Due to the variance in image sizes, our patching process generated over 40,000 image patches, a method depicted in figure 4. Unlike standard practices in applications like facial recognition, where subjects are positioned uniformly relative to the camera [26], our approach does not regulate for angles, distances, image sizes, or quality. Instead, we utilised images directly from online real estate listings, aiming to create a training and prediction environment that mirrors real-world conditions. The impact of different pixel sizes on the number of generated patches is detailed in figure 2, illustrating a departure from the controlled image acquisition methods used in studies such as [32], where a specific algorithm extracts standardised 640x640

[2]That more specifically means that adjacent patches contain 33.3% of each other. An example can be seen in figure 4.



(a) Size: 1920 x 1440 pixels (b) Size: 1440 x 1080 pixels (c) Size: 960 x 720 pixels (d) Size: 400 x 300 pixels

Figure 2
Size comparison of different images in the data set. Ordered from largest image with regard to pixels to the smallest image. The images are scaled down to 5.5% of their original size (Source: immobilienscout24.at).

pixel front-view images from Google Street View.

Our methodology involved a manual inspection of the image patches, excluding those that failed to display significant features of the house (only patches displaying 50% or more of a key housing feature were retained) or were significantly obscured by elements like vegetation. The distribution of patches per building varied widely, ranging from 1 to 231 patches, with a median of 13 patches and an average of 22. This variance is reflected in the distribution of patches per image, as shown in figure 3, indicating the influence of the original image's pixel size and the house's relative size within the image. Following a meticulous review for flaws and inconsistencies, we proceeded to train a CNN on the remaining approximately 12,000 image patches to predict their designated class.

To rigorously evaluate our CNN predictions, we meticulously structured our dataset into training and test sets. Initially, we segregated 25% of the buildings from the smallest class (*bad*, encompassing 109 buildings) to form multiple test sets, averaging 28 houses per set. The remaining buildings from this class (82 in total) set the benchmark for a balanced training set size, ensuring each building in the smallest class is represented at least once in the test sets. This methodology allows for the creation of 20 balanced test and training set combinations, affording us 20 distinct "ground truths" and subsequently, 20 uniquely trained CNNs. The distribution of buildings across these 20 sets, inclusive of unbalanced, balanced, and test configurations, is detailed in Table 3.

In addition to these balanced sets, we crafted 20 unbalanced training sets to leverage the full breadth of our dataset. This step facilitates the implementation of a transfer learning and fine-tuning strategy, as recommended by [43], where the CNN—pre-trained on the extensive ImageNet dataset—undergoes further training on our unbalanced dataset. Initially, we freeze all but the back-end layers of the network, adjusting the model to our specific dataset while retaining the pre-trained weights. Subsequent to this initial phase,

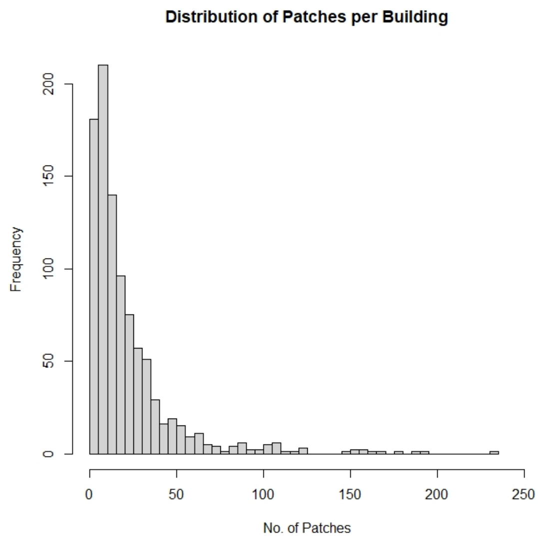


Figure 3
Histogram of the patch count per image for the 960 houses in the CNN ground truth (Source: Authors own work).

(a) Original image of the house

(b) 224 x 224 patches of the image

Figure 4
Patched image of the house (224 x 224 pixels) with an overlap of 33.3%. The patches marked in red are considered as containing enough building features to be used in the CNN training (Source: immobilienscout24.at).

all layers are unfrozen, and the entire network undergoes retraining on the balanced dataset to fine-tune its ability to discern between different image classes.

Utilising this refined model, we proceed to predict the classes of over 22,000 image patches. Given the necessity of multiple predictions for some buildings due to oversampling, we aggregate these predictions to assign a final condition class to each of the 960 buildings. This determination is made through majority voting among the predicted classes for all patches associated with a building, ensuring the most representative condition class is selected for inclusion in the pricing model.

Table 3
FIVE EXEMPLARY GROUND TRUTH OF THE TWENTY GROUND TRUTHS THAT ARE USED IN THIS ARTICLE. THE UPPER THIRD DEPICTS THE COUNT OF BUILDINGS OF FIVE TEST SETS FOR THE THREE CLASSES, THE MIDDLE THIRD THE COUNT OF THE UNBALANCED TRAINING SETS AND THE LOWER THIRD THE COUNT OF THE BUILDINGS IN THE BALANCED TRAINING SETS (SOURCE: AUTHORS OWN WORK).

Ground Truth Composition						
TEST SETS	**Total No IMG**	**1**	**2**	**3**	**4**	**5**
Class 1	290	29	29	29	29	29
Class 2	561	28	28	28	28	28
Class 3	109	27	27	27	28	27
UNBALANCED TRAINING SETS	**Total No IMG**	**1**	**2**	**3**	**4**	**5**
Class 1	290	261	261	261	261	261
Class 2	561	533	533	533	533	533
Class 3	109	82	82	82	81	82
BALANCED TRAINING SETS	**Total No IMG**	**1**	**2**	**3**	**4**	**5**
Class 1	290	82	82	82	82	82
Class 2	561	82	82	82	82	82
Class 3	109	82	82	82	82	82

The CNNs are based on the EfficientNet0 architecture [54] which is pre-trained on ImageNet data. We fed the images with three colour channels (RGB) into the network for training and normalised the image arrays to a range from zero to one by dividing the pixel values by 255. The tuning of the CNNs lasted for 50 epochs with a learning rate of 0.0001 using the ADAM optimiser function. The batch size in every epoch was 16. We used a drop out rate of 0.5, with no image augmentation and also no early-stopping. The experiments have been performed on a server host with Windows 10 Enterprise OS, 64 GB RAM and an NVIDIA RTX 2080 Ti GPU.

4. Results

4.1. Pricing Model

Table 4 presents the outcomes of five regression analyses based on our dataset, which includes buildings that come with the year of construction (thus the dataset for the hedonic pricing models is smaller than the dataset used to train the CNN). The baseline regression model (1) utilises readily accessible data from online real estate platforms, specifically building area, plot size, and construction year, yet neglecting property images to assess the often optimistically skewed and visually observable building condition. Regression model (2) then introduces human assessments of condition based on established standards using images, which enhances the model's accuracy. This addition reveals expected trends and discounts across the condition classification dummies, with the *good* condition class serving as the reference.

Model (3) explores the scenario where the year of construction variable is omitted. The quality of this model dips below that of model (2) but remains superior to the baseline, suggesting that excluding construction year while incorporating expert-based condition classifications marginally improves model performance. In contrast, model (4) incorporates the condition variable as predicted by a CNN, and also surpasses the baseline model in performance, demonstrating appropriate coefficient signs and an overall improvement in model quality.

A comparison between models incorporating standard assessments (model (2)) and those using CNN-based classifications (model (4)) against the baseline (model (1)) indicates that traditional assessments of condition yield a higher model quality than those derived from computer vision algorithms. Notably, the influence of the condition dummy variables appears to be diminished in the models reliant on CNN assessments, particularly for the *bad* condition class. Human evaluations are prone to highlighting the stark differences between 'average' and 'bad,' whereas the algorithm exhibits a finer sensitivity, detecting not only the extremes but also a fluid progression between these categories.

The coefficients for the four key parameters in the baseline pricing model exhibit significant predictive power in all five models, with the exception of plot size in models (3) and (5), indicating a positive influence on property prices. Specifically, the *year_of_construction* regressor demonstrates high significance with the anticipated positive direction in models (1), (2), and (4), underscoring its importance in the valuation. Similarly, the plot price variable, acting as a proxy for location quality within a municipality, is significant across all models, highlighting its relevance in price determination.

The inclusion of CNN-predicted condition variables in both the comprehensive model (4), which retains

Table 4

REGRESSION RESULTS PRICE (FULL DATA SET): REGRESSION TABLE WITH THE BASELINE PRICING MODEL WITH PLOT SIZE, PLOT PRICE, FLOOR SIZE AND A FIRST AND SECOND DEGREE POLYNOMIAL FOR THE YEAR OF CONSTRUCTION (1), THE BASELINE HEDONIC MODEL INCLUDING THE CONDITION ASSESSMENT BY STANDARDS (2), THE BASELINE HEDONIC MODEL EXCLUDING THE YEAR OF CONSTRUCTION BUT INCLUDING THE CONDITION ASSESSMENT (3), THE BASELINE HEDONIC MODEL INCLUDING THE CNN CONDITION ASSESSMENT (4) AND THE BASELINE HEDONIC MODEL EXCLUDING THE YEAR OF CONSTRUCTION BUT INCLUDING THE CNN CONDITION ASSESSMENT (SOURCE: AUTHORS OWN WORK).

	Dependent variable:				
	log(price)				
	(1)	(2)	(3)	(4)	(5)
plot_size	0.0001***	0.0001***	0.00003	0.0001***	0.00004
	(0.00003)	(0.00002)	(0.00003)	(0.00003)	(0.00003)
plot_price	0.002***	0.002***	0.002***	0.002***	0.002***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
floor_size	0.004***	0.004***	0.004***	0.004***	0.004***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
poly(year_of_construction, 2)1	4.538***	2.760***		3.867***	
	(0.401)	(0.402)		(0.400)	
poly(year_of_ construction, 2)2	3.520***	2.310***		2.803***	
	(0.398)	(0.382)		(0.396)	
Condition_Standards_Average		−0.266***	−0.370***		
		(0.033)	(0.032)		
Condition_Standards_Bad		−0.618***	−0.823***		
		(0.058)	(0.055)		
Condition_CNN_Average				−0.200***	−0.283***
				(0.033)	(0.035)
Condition_CNN_Bad				−0.293***	−0.459***
				(0.046)	(0.047)
Constant	11.784***	12.114***	12.252***	11.961***	12.069***
	(0.047)	(0.054)	(0.055)	(0.052)	(0.056)
Observations	694	694	694	694	694
R^2	0.633	0.690	0.656	0.659	0.593
Adjusted R^2	0.630	0.686	0.654	0.656	0.590
Residual Std. Error	0.393 (df = 688)	0.362 (df = 686)	0.380 (df = 688)	0.379 (df = 686)	0.414 (df = 688)
F Statistic	237.005*** (df = 5; 688)	217.745*** (df = 7; 686)	262.829*** (df = 5; 688)	189.580*** (df = 7; 686)	200.607*** (df = 5; 688)
Note:					*p<0.1; **p<0.05; ***p<0.01

the year of construction, and the simplified model (5), which omits this variable, significantly impacts price predictions. The sign of these impacts aligns with expectations, illustrating the predictive value of the condition classifications derived from computer vision. A direct comparison between model (4), which integrates the assessment variable, and model (1), which lacks this regressor, reveals an improvement in the pricing model upon including the condition assessment. However, model (5) shows a decline in quality compared to the baseline, suggesting two possible interpretations: first, that the addition of computer vision assessments might diminish the year of construction's predictive value due to potential confounding effects; and second, that the assessment variable's contribution could be considered marginal, as indicated by a lower adjusted R-squared when substituting the year of construction with computer vision classification outcomes.

4.2. CNN Prediction Results

Table 5 contains the aggregate confusion matrix for the predictions of the 20 independent CNNs, which also is the aggregate result of the cross validation. 62% of the true class *good* images are predicted into the correct class. In contrast 58% of the class *average* house images get the correct prediction. In class *bad* 60% are predicted into the correct class. This states that the CNN comes to similar accuracy for the three classes. But when observing the absolute count of predictions we see that the predicted number of class *average* images decreased while the count of class *good* and class *bad* images increased. One explanation can be a mean reversion property of the CNN such that the overall predictions per class tend to be balanced or that the delineation of classes is not clear enough. The second explanation can be further argued with the fact that the algorithm has to learn many features that are not clearly distinct. Therefore, the network struggles to assign some of the true features to the true class and instead finds a correlation with one of the other adjacent classes.

The confusion matrix in table 5 shows that the class *good* images have much more false negatives than in class *average*. Analogue, class *average* image predictions show much more false positives in class *good* predictions. The situation is very similar when looking at the wrong predictions in class *bad* and class *average*. This makes us conclude that the class *good* and class *bad* images cause problems due to the fact, that the images are not clearly delineated from the class *average* images. That can also be seen when comparing the precision and recall metrics for the three classes in the table 5. Once again, the possible explanation is that the multitude of characteristics that have to be considered, hence, the probability is low that the CNN always detects the difference in the characteristics that are delineating the classes.

Table 5
AGGREGATE CONFUSION MATRIX OF THE 20 CNN TRAINING ATTEMPTS. THE TABLE SHOWS THE ABSOLUTE NUMBER OF PREDICTIONS AND THE PERCENTAGE OF THE TOTAL NUMBER OF IMAGES (TRUE CLASS) PER CONDITION CLASS *good*, *average*, AND *bad*. THE SUM COLUMN SHOWS THE TOTAL NUMBER OF THE RESPECTIVE PREDICTED CLASS AND THE PERCENTAGE CHANGE COMPARED TO THE TRUE NUMBER OF THAT CLASS. THE ROW SUM DEPICTS THE NUMBER OF TRUE IMAGES PER CLASS. PRECISION AND RECALL ROWS INDICATE THE TWO EVALUATION CLASSIFIERS FOR THE OUTPUT QUALITY OF THE MODEL. THEY ARE FREQUENTLY USED WHEN DEALING WITH IMBALANCED DATA (SOURCE: AUTHORS OWN WORK).

		True			
	Class	***good***	***average***	***bad***	**Sum**
Prediction	***good***	179 (61.7%)	152 (27.1%)	6 (5.5%)	337 (+16.2%)
	average	87 (30.0%)	326 (58.1%)	38 (34.9%)	451 (-19.6%)
	bad	24 (8.3%)	83 (14.8%)	65 (59.6%)	172 (+57.8%)
	Sum	290	561	109	
Precision		0.53	0.72	0.38	
Recall		0.62	0.58	0.60	

4.3. Qualitative Analysis

The investigation of the images that show the best fit in CNN predictions (95% of patches in a image are predicted into the correct, expert-assessed class; examples are depicted in figure 5) suggest that the brighter a feature of a house the more likely it is to be classified in class *good*. That e.g. constitutes that the depreciation by the weather shifts colour tones of the building surfaces. It also seams apparent that less wooden components are correlated with a better condition classification. Class *average* buildings (or patches) seem to be correlated with less brightness and also less colourful building surfaces and advanced roof deterioration. The images also contain wooden components and on average smaller windows compared to the majority of class *good* images. Class *bad* examples mainly depict wooden constructions with rough surfaces and generally a very used look of materials. The colour tone is in the darkest section of the three classes and gives a less promising impression to the assessor.

Regarding wrong predictions we investigate the images where the algorithm predicted 5% or less of the patches into the correct class. There, class *good* has 15 images that are wrongly predicted, in class *average* we find 24 images in a wrong class and in class *bad* only 3 images are predicted wrong. Generally, the example images of the CNN prediction do not show any obvious hints in a way that the algorithm learned ”wrong” features that would not correlate with the corresponding class of the images. The network learns features for the classification irrespective of the size, form and quality of the images (also including the brokerage firm’s logo seems not to effect a wrong prediction).

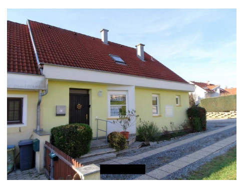
The wrong predictions (true class *good* gets predicted into class *bad*) range from images where the majority relative voting is indefinite (e.g. figure 6 (a)) to images where the percentage of predicted patches

(a) Class *good* with 8 patches

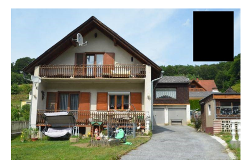
(b) Class *good* with 3 patches

(c) Class *good* with 9 patches

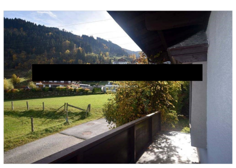
(d) Class *average* with 1 patch



(e) Class *average* with 19 patches



(f) Class *average* with 7 patches



(g) Class *bad* with 12 patches

(h) Class *bad* with 20 patches

(i) Class *bad* with 9 patches

Figure 5

Examples of condition class predictions where 95% or more of the patches are predicted into the correct class. the logos of the brokerage firms were blacked out (Source: immobilienscout24.at).

in majority assigns a wrong class. In the true class *good* images which are predicted in class *bad* wood is a dominant component of the houses. Furthermore, rough building surfaces dominate four of the wrongly predicted images. Observe the wrongly predicted class *bad* houses in figure 7 (d). Here the indecisiveness of the network is more obvious than in the previous wrong predictions.

Regarding the class *good* buildings one caveat has to be mentioned: some of the images show a swimming pool. This is definitely not relevant for the condition classification as the patches containing (parts of) the pool area were excluded.

5. Regression Residuals Classification

Real estate images can convey critical information about a property, such as the quality of construction, estimated construction year, and the building's condition, which reflects its maintenance history. These visual cues, discernible from property images, may encapsulate factors influencing property valuation [20]. Our research aims to explore whether these images can account for some of the variation in our baseline pricing model's residuals, suggesting that the images might implicitly contain aggregated data on these variables. This is in line with the suggested approach of [55], which shows an increase in price prediction accuracy. We posit that the images hold consolidated information, which, when classified, could elucidate portions of the regression residuals. This investigation is conducted in two scenarios: one including the year of construction (Model A) and the other excluding it (Model B) from the baseline regression, from which we derive the residuals. Subsequently, we categorise these residuals into three distinct classes—low, medium, and high—to thoroughly examine their relationship with the visual information contained in the images. The methodology and findings of this approach are elaborated in the subsequent sections.

1. Baseline models to obtain the regression residuals:
   (a) Model A: $log\_price \sim plot\_size + plot\_price + floor\_size + year\_of\_construction$
   (b) Model B: $log\_price \sim plot\_size + plot\_price + floor\_size$
2. Calculation of the residuals of the baseline models of the price (not the logarithmic price) to obtain the percentage share of the residuals which is used to assign the buildings to three classes[3].
3. Split of the relative residuals in tertiles (three classes) for the two baseline models:
   - Class LOW: Houses in the lowest 33.3% tertile of relative residuals.

[3]This is crucial as also a small residual (a small monetary amount) can cause a significant deviation on a low-price property.

(a) True Class: 1 Predicted Class: 3, 42 patches, 34% C1, 21% C2, 45% C3, uncertain class prediction.

(b) True Class: 1 Predicted Class: 3, 19 patches, 16% C1, 19% C2, 65% C3, majority class prediction. same wrongly predicted image as in first network

(c) True Class: 1 Predicted Class: 3, 6 patches, 33% C1, 0% C2, 67% C3, majority class prediction.

(d) True Class: 1 Predicted Class: 3, 72 patches, 37% C1, 18% C2, 45% C3, uncertain class prediction.

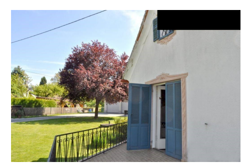
(e) True Class: 1 Predicted Class: 3, 9 patches, 23% C1, 23% C2, 54% C3, majority class prediction.

(f) True Class: 1 Predicted Class: 3, 16 patches, 31% C1, 25% C2, 44% C3, uncertain class prediction.

Figure 6
images where the CNN predicted the lowest class despite the human expert classified it into the highest class based on the standards (Source: immobilienscout24.at).

(a) True Class: 3 Predicted Class: 1, 36 patches, 44% C1, 8% C2, 48% C3, uncertain class prediction.

(b) True Class: 3 Predicted Class: 1, 80 patches, 51% C1, 32% C2, 17% C3, majority class prediction.



(c) True Class: 3 Predicted Class: 1, 96 patches, 55% C1, 37% C2, 8% C3, majority class prediction.

(d) True Class: 3 Predicted Class: 1, 18 patches, 58% C1, 1% C2, 41% C3, majority class prediction.

(e) True Class: 3 Predicted Class: 1, 8 patches, 30% C1, 51% C2, 19% C3, majority class prediction.

(f) True Class: 3 Predicted Class: 1, 25 patches, 34% C1, 27% C2, 39% C3, uncertain class prediction.

Figure 7
images where the CNN predicted the lowest class despite the human expert classified it into the highest class based on the standards (Source: immobilienscout24.at).

- Class AVERAGE: Houses in the medium 33.3% tertile of relative residuals.
- Class HIGH: Houses in the highest 33.% tertile of relative residuals.

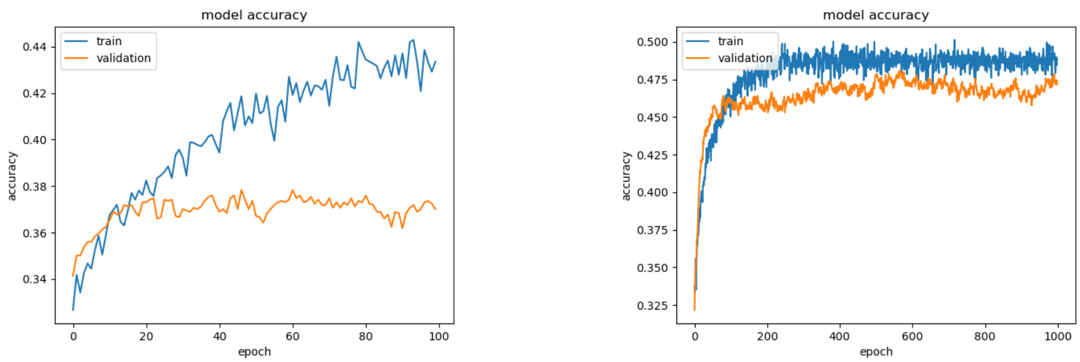
4. Training of the CNN. Here the same methodology is applied as in the main condition classification section.
5. If the CNN is learning specific patterns in the pictures, the predicted classes are included into the baseline regression to test their effect on the coefficient of determination.

Subsequently, we employed the same CNN used for condition classification to categorise three different classes of residuals, utilising five-fold cross validation to ensure each building received a prediction. However, the training process and subsequent predictions indicated that the network struggled to identify specific patterns within the images that correlate with the residual classes. The highest accuracy achieved among the five models was only 38.0% (as detailed in figure 8.a), essentially mirroring the success rate of random chance, which would be approximately 33.3%. This suggests that the model's predictive capability was not meaningfully better than guessing.

Furthermore, when we removed the year of construction from the baseline regression analysis, the CNN's accuracy slightly improved in the context of five-fold cross-validation, reaching an average peak of 45% in the most accurate model. However, this improvement still falls short of demonstrating substantial precision (refer to figure 8.b). It appears that the CNN may have adapted to recognise aspects of the residuals attributable to the exclusion of the construction year variable, rather than identifying broader, more meaningful patterns within the data.

In contrast to [55] who are focusing on especially architectural style, and who find a positive impact on pricing accuracy when incorporating a CNN predicted residuals variable, we find that when including the age variable in the determination of the regression residuals, our CNN is not picking up any detectable pattern related to the residuals. When neglecting the age variable in the residuals determination, our CNN model is detecting a relationship between residuals and the images, thus we see further research opportunities regarding the use of regression residuals coming from AVMs.

The main caveat of our investigation of the residuals is that we used a classification algorithm instead of a CNN that outputs a regression and, therefore, outputs continuous variables instead of classes. Furthermore, the patching of images might significantly influence predictability of the images. Still, we believe that the pictures contain information that can positively impact the accuracy of a hedonic pricing model but we leave a deeper investigation open for future research.



(a) CNN training on the regression residuals including the year of construction.

(b) CNN training on the regression residuals excluding the year of construction.

Figure 8
Comparison of the development of the two model accuracies over the training epochs of the best two models in the five-fold cross validation (Source: Authors own work).

6. Obstructions of Computer Vision in Real Estate Valuation

We've identified numerous challenges that warrant careful consideration in both future research and practical applications. Those challenges are based on our own experiences and solutions from this project and other related projects. Thus, the ensuing section is intended to function as a guide, highlighting critical areas for focus when integrating computer vision technologies into real estate valuation processes.

6.1. Images

The selection of ground truth data and images is pivotal for effective classification by computer vision algorithms. A substantial dataset is essential for training and testing a CNN, as emphasised by [35]. Furthermore, achieving balance across classes within the dataset is highly recommended to enhance model performance [15]. For datasets with a limited number of images, creating patches from original images presents a viable solution to augment the dataset [53]. In our approach, images were segmented into multiple patches based on the CNN's input pixel dimensions. This technique inherently shifts the CNN's attention to finer details within the images, such as windows, doors, and roofing materials, mirroring the human eye's foveal vision focus on specific areas [37].

The process of image patching also necessitated decisions regarding the dimensions of each patch and the degree of overlap between patches. These choices significantly affect the dataset's size and the granularity of features captured, with greater overlap leading to redundancy of features across patches. A considerable portion of the patched dataset included areas not directly relevant to property valuation, such as surrounding land or extraneous nearby elements. Thus, we faced a decision on whether to exclude certain patches from the training dataset or to adopt a more comprehensive approach by including entire images.

Image capture parameters pose additional challenges, including the optimal distance, angle, and environmental conditions for photographing properties. While utilising readily available online platform images offers a pragmatic approach, adhering to standardised parameters, such as specific angles or weather conditions, could potentially enhance consistency and realism. This can be achieved by sourcing images from platforms like Google Street View, which allows for some standardisation, as suggested by [32]. However, this method requires access to up-to-date property addresses and assumes that the built environment has not significantly changed, limitations inherent to relying on such platforms for current imagery.

6.2. Classification

When training a CNN on classification tasks, determining the number of classes and establishing a precise classification scheme is essential. For instance, in predicting a property's condition, one can either adopt predefined criteria for class distinctions or devise custom classes. If custom classes are defined, it's critical to ensure they are distinctly separable to avoid ambiguities or overlaps, necessitating a well-defined standard for class differentiation.

A significant challenge in using images for training is the subjective nature of human classification. Even experienced real estate appraisers may have divergent opinions on the same image, introducing an element of bias. To mitigate this, the consensus among a substantial group of appraisers or evaluators is vital to affirm the classifications' relative objectivity.

The issue of class distribution poses another challenge; data sourced from public platforms often exhibit uneven class representation. Such imbalance can hinder a CNN's training effectiveness, as equal frequency across classes is preferable for optimal model performance [15]. Oversampling, or augmenting the dataset by replicating images from underrepresented classes, is a strategy to address this [28]. However, this method introduces its own set of problems, notably the repetition of images, which could bias the model. Image augmentation techniques, such as applying random transformations to the training images, can help diminish these concerns by diversifying the visual data, thereby reducing the impact of repeated images.

An alternative strategy for managing imbalanced datasets involves adapting the training methodology of the CNN. Initially, leveraging a network pre-trained on a comprehensive dataset, such as ImageNet, where only the back-end layers are adjusted, allows for initial learning from the unbalanced dataset. Subsequently, fine-tuning the entire network on a balanced dataset ensures that the model adjusts to the nuanced features across all classes [43]. This two-step process, incorporating both pre-training and fine-tuning, enhances the model's ability to generalise across diverse data, offering a sophisticated solution to the challenges posed by imbalanced datasets.

6.3. Neural Network

The choice of network architecture significantly influences the accuracy of CNN classifications, with various architectures like ResNet [12], AlexNet [24], GoogLeNet [51], VGG16, VGG19 [47], and EfficientNet [54] offering diverse capabilities. EfficientNet, for instance, features eight scalable network types (EfficientNet0-7) tailored for different image sizes, such as EfficientNet7, which is optimised for 600 x 600 pixel images, potentially enhancing the detail and information captured. This nuanced approach to handling image sizes within specific architectures, and its implications on predictive accuracy in real estate economics, remains an explorable area.

The architecture's depth, or the number of trainable layers within the CNN, can also affect prediction outcomes. Tailoring a custom architecture to adjust the layer count could further refine performance [17]. Equally crucial is the choice of optimiser, which adjusts the neural network's weights to minimise losses and solve optimisation problems [50]. Popular options include SGD (Stochastic Gradient Descent) and the ADAM optimiser, each requiring different hyper parameter settings.

Selecting the optimal hyper parameters (like learning rate, training epochs, and batch size) often involves manual adjustment through trial and error, a time-intensive process. Tools such as Early Stopping can automate the determination of ideal training epochs. For more complex hyper parameter tuning, Auto Tuning packages offer a systematic approach by testing a range of parameter combinations and selecting the optimal model configuration [31].

Finally, when working with limited datasets, choosing the appropriate size for training and testing sets is crucial. Cross-validation is particularly effective for small datasets, providing a comprehensive prediction across all data points. However, this method is time-consuming, especially when identifying the best training parameters. An alternative approach, involving distinct training, validation, and test sets, offers a streamlined methodology suitable for proof-of-concept studies or initial parameter determination in CNNs, especially with larger datasets. This method allows for efficient evaluation and parameter tuning, supporting preliminary research efforts.

7. Implications and Future Directions for AVMs

Building on the main objective of this article, we see significant potential for further research and practical applications. The ability to extract information, particularly the structural condition, from real estate images opens new avenues for improving AVMs. These additional data sources enable AVMs to be trained and optimized with greater precision, which is especially beneficial for banks, institutional real estate investors but also public institutions [52]. Banks and institutional real estate funds are often required to conduct continuous revaluations of their real estate portfolios due to regulatory requirements while public institutions need valuations to calculate tax burden. By utilizing automated image analyses, the condition of properties can be assessed efficiently and in a standardized manner, saving both time and resources.

Moreover, the standardized classification of property condition offers the opportunity to establish uniform evaluation frameworks. This is particularly relevant for existing AVM models, which are frequently used as a basis for valuations in real estate apps due to their advanced algorithms and straight-forward accessible data can offer more precise, timely, and cost-effective property valuations[23]. Automating the assessment of condition variables could enable laypeople to conduct more objective property valuations without having to evaluate subjective or complex variables themselves. Users would only need to provide clearly defined, objectively measurable parameters such as size, year of construction, or location. This reduces room for interpretation and minimizes potential errors in valuation processes.

Additionally, in traditional valuation methods such as the Cost Approach or the Comparison Method, the objectivity enabled by computer-assisted image analysis supports a reliable foundation and facilitates objective comparisons between properties. This approach also provides economic and commercial benefits by accelerating valuation processes and reducing costs.

From a societal perspective, this research contributes to greater transparency in the real estate market by establishing a comprehensible and consistent basis for property valuations. This could positively influence public attitudes toward real estate valuation, strengthen trust in valuation processes[23]. In education, these methods could serve as practical examples of how modern technologies can be applied in real estate valuation, bridging the gap between theory and practice.

In summary, the findings of this article make an important contribution to the advancement of data-driven valuation models, provide concrete approaches to improving efficiency in practice, and promote the standardization and objectivity of valuation methods, with far-reaching implications for research, practice, and society.

8. Conclusion

This study examines the impact of building condition classifications derived from images on a straightforward real estate pricing model, with all data sourced from online real estate platforms. Utilising a three-class condition classification system provided by the Association of Austrian Appraisers, real estate professionals classify buildings according to this standardised framework, aiming to mitigate the subjectivity inherent in such evaluations, despite the unavoidable influence of personal biases. Our objective is not to dissect the perceptions of these experts but to assess how these classifications affect a simple hedonic pricing model tailored for real estate valuation, specifically employing the cost approach.

Additionally, we explore whether a CNN leveraging computer vision can accurately reflect these expert-derived condition classifications and, by extension, serve effectively within a pricing model to automate the assessment of single-family homes.

Our findings indicate that human-based condition classifications significantly enhance the pricing model for 694 properties on the Austrian market, as evidenced by data from an online real estate platform. Incorporating condition assessments from the CNN yields an improved adjusted R-squared for the model, though not to the same extent as the expert-based classifications. This suggests a refinement in automated valuation models through the application of computer vision, albeit with a noted decline in model accuracy when excluding the year of construction.

Employing images from 960 buildings, we trained a CNN to predict the three condition classes, achieving an overall accuracy of approximately 60%. This indicates the CNN's capability to identify features correlative with the condition classes, despite the challenges highlighted throughout our study.

We also delved into whether building images could account for variations in regression residuals from our baseline pricing model. By categorizing residuals into three classes from high to low and employing five-fold cross-validation, we found that the CNN failed to detect any features directly associated with the regression residuals. Still, exploring the impact of building condition assessment using computer vision to account for regression residuals in simple pricing models, especially when leveraging data from online real estate platforms, presents an intriguing avenue for future research.

In conclusion, this article demonstrates that building condition classification plays a crucial role in the precision of automated valuation models and that computational methods like CNNs hold potential for enhancing real estate evaluations using both data and visual content from online platforms. We implemented our methods in a small market like Austria, with its limited quantity and quality of data. Therefore, we see even greater potential in other countries. Furthermore, it underscores the various challenges and considerations inherent in adopting a data-driven approach with computer vision for real estate assessment.

References

[1] Abidoye, R. B. and Chan, A. P. [2018], 'Improving property valuation accuracy: A comparison of hedonic pricing model and artificial neural network', *Pacific Rim Property Research Journal* **24**(1), 71–83.
URL: *https://doi.org/10.1080/14445921.2018.1436306*

[2] Ahlfeldt, G. and Mastro, A. [2012], 'Valuing Iconic Design: Frank Lloyd Wright Architecture in Oak Park, Illinois', *Housing Studies* **27**(8), 1079–1099.

[3] Bergadano, F., Bertilone, R., Paolotti, D. and Ruffo, G. [2019], 'Learning real estate automated valuation models from heterogeneous data sources', *arXiv preprint arXiv:1909.00704* .

[4] Buitelaar, E. and Schilder, F. [2017], 'The Economics of Style: Measuring the Price Effect of Neo-Traditional Architecture in Housing', *Real Estate Economics* **45**(1), 7–27.

[5] Bulut, B., Allahverdi, N., Kahramanli, H. and Yalpir, S. [2011], 'A residential real-estate valuation model with reduced attributes', *International Journal of Mathematical Models and Methods in Applied Sciences* **5**(3), 586–593.

[6] Cheshire, P. and Sheppard, S. [2002], 'The welfare economics of land use planning', *Journal of Urban economics* **52**(2), 242–269.

[7] Despotovic, M., Sakeena, M., Koch, D., Döller, M. and Zeppelzauer, M. [2018], 'predicting heating energy demand by computer vision', *Computer Science - Research and Development* **33**(1-2), 231–232.

[8] Dunse, N. and Jones, C. [1998], 'A hedonic price model of office rents', *Journal of property valuation and investment* .

[9] Francke, M. K. and van de Minne, A. M. [2017], 'Land, Structure and Depreciation', *Real Estate Economics* **45**(2), 415–451.

[10] Gibbons, S., Mourato, S. and Resende, G. M. [2014], 'The amenity value of english nature: a hedonic price approach', *Environmental and Resource Economics* **57**(2), 175–196.

[11] Green, S. B. [1991], 'How many subjects does it take to do a regression analysis', *Multivariate behavioral research* **26**(3), 499–510.

[12] He, K., Zhang, X., Ren, S. and Sun, J. [2016], Deep residual learning for image recognition, *in* 'Proceedings of the IEEE conference on computer vision and pattern recognition', pp. 770–778.

[13] Herath, S., Maier, G. et al. [2010], 'The hedonic price method in real estate and housing market research. a review of the literature', *SRE-Discussion Papers* **2010/03**.

[14] Hill, R. [1998], 'What sample size is "enough" in internet survey research', *Interpersonal Computing and Technology: An electronic journal for the 21st century* **6**(3-4), 1–12.

[15] Johnson, J. M. and Khoshgoftaar, T. M. [2019], 'Survey on deep learning with class imbalance', *Journal of Big Data* **6**(1), 1–54.

[16] Khan, A., Sohail, A., Zahoora, U. and Qureshi, A. S. [2020], 'A survey of the recent architectures of deep convolutional neural networks', *Artificial Intelligence Review* pp. 1–70.
URL: *http://link.springer.com/10.1007/s10462-020-09825-6*

[17] Khan, S. H., Abbas, Z., Rizvi, S. D. et al. [2019], Classification of diabetic retinopathy images based on customised cnn architecture, *in* '2019 Amity International conference on artificial intelligence (AICAI)', IEEE, pp. 244–248.

[18] Knight, J. R. and Sirmans, C. [1996], 'Depreciation, Maintenance, and Housing Prices', *Journal of Housing Economics* **5**(4), 369–389.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S1051137796900194*

[19] Koch, D., Despotovic, M., Leiber, S., Sakeena, M., Döller, M. and Zeppelzauer, M. [2019*a*], 'Real estate image analysis: A literature review', *Journal of Real Estate Literature* **27**(2), 269–300.

[20] Koch, D., Despotovic, M., Leiber, S., Sakeena, M., Döller, M. and Zeppelzauer, M. [2019*b*], 'Real Estate Image Analysis: A Literature Review', *Journal of Real Estate Literature* **27**(2), 269–300.

[21] Koch, D., Despotovic, M., Sakeena, M., Döller, M. and Zeppelzauer, M. [2018], Visual estimation of building condition with patch-level convnets, *in* 'Proceedings of the 2018 ACM Workshop on Multimedia for Real Estate Tech', pp. 12–17.

[22] Koch, D., Thaler, S. and Mayr, R. [2021], 'Order versus disorder–the impact on value estimates of durable consumption goods', *Applied Economics Letters* **28**(19), 1635–1640.

[23] Kok, N., Koponen, E.-L. and Martínez-Barbosa, C. A. [2017], 'Big data in real estate? from manual appraisal to automated valuation', *The Journal of Portfolio Management* **43**(6), 202–211.

[24] Krizhevsky, A., Sutskever, I. and Hinton, G. E. [2012], ImageNet classification with deep convolutional neural networks, *in* T. F. Gonzalez, ed., 'Advances in neural information processing systems', Chapman and Hall/CRC, pp. 1097–1105.
URL: *https://www.taylorfrancis.com/books/9781420010749 http://dl.acm.org/citation.cfm?doid=3098997.3065386*

[25] Lancaster, K. J. [1966], 'A New Approach to Consumer Theory', *Journal of Political Economy* **74**(2), 132–157.
URL: *https://www.journals.uchicago.edu/doi/10.1086/259131*

[26] Lawrence, S., Giles, C. L., Tsoi, A. C. and Back, A. D. [1997], 'Face recognition: A convolutional neural-network approach', *IEEE transactions on neural networks* **8**(1), 98–113.

[27] Lecun, Y., Bengio, Y. and Hinton, G. [2015], 'Deep learning', *Nature* **521**(7553), 436–444.

[28] Lei, Y., Yang, B., Jiang, X., Jia, F., Li, N. and Nandi, A. K. [2020], 'Applications of machine learning to machine fault diagnosis: A review and roadmap', *Mechanical Systems and Signal Processing* **138**, 106587.

[29] Li, X., Zhang, C. and Li, W. [2015], 'Does the visibility of greenery increase perceived safety in urban areas? Evidence from the place pulse 1.0 dataset', *ISPRS International Journal of Geo-Information* **4**(3), 1166–1183.

[30] Li, Y. and Leatham, D. J. [2011], Forcasting Housing Prices: Dynamic Factor Model vs. LBVAR Model, *in* 'Annual Meeting of Agricultural and Applied Economics Association'.

[31] Liang, H., Zhang, S., Sun, J., He, X., Huang, W., Zhuang, K. and Li, Z. [2019], 'Darts+: Improved differentiable architecture search with early stopping', *arXiv preprint arXiv:1909.06035* .

[32] Lindenthal, T. and Johnson, E. B. [2021], *Machine Learning, Architectural Styles and Property Values*, The Journal of Real Estate Finance and Economics.

[33] Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y. and Alsaadi, F. E. [2017], 'A Survey of Deep Neural Network Architectures and Their Applications', *Neurocomputing* **234**, 11–26.
URL: *https://www.sciencedirect.com/science/article/pii/S0925231216315533*

[34] Miller, N., Sah, V. and Sklarz, M. [2018], 'Estimating Property Condition Effect on Residential Property Value: Evidence from U.S. Home Sales Data', *Journal of Real Estate Research* **40**(2), 179–198.

[35] Ning, H., Li, Z., Wang, C. and Yang, L. [2020], 'Choosing an appropriate training set size when using existing data to train neural networks for land cover segmentation', *Annals of GIS* **26**(4), 329–342.

[36] of Valuers' Association, T. E. G. [2020], *European Valuation Standards (EVS 2020)*, The European Group of Valuers' Association.

[37] Orquin, J. L., Perkovic, S. and Grunert, K. G. [2018], 'Visual biases in decision making', *Applied Economic Perspectives and Policy* **40**(4), 523–537.

[38] Pearson, L., Tisdell, C. and Lisle, A. [2002], 'The impact of noosa national park on surrounding property values: An application of the hedonic price method', *Economic Analysis and Policy* **32**(2), 155–171.

[39] Piao, Y., Chen, A. and Shang, Z. [2019], 'Housing Price Prediction Based on CNN', *9th International Conference on Information Science and Technology, ICIST 2019* pp. 491–495.

[40] Popp, R. [2020], 'Empfehlungen für Herstellungskosten 2020', *Sachverständige* **2**, 89–96.

[41] Potrawa, T. and Tetereva, A. [2022], 'How much is the view from the window worth? machine learning-driven hedonic pricing model of the real estate market', *Journal of Business Research* **144**, 50–65.

[42] Poursaeed, O., Matera, T. and Belongie, S. [2018], 'Vision-based real estate price estimation', *Machine Vision and Applications* **29**(4), 667–676.
URL: *https://doi.org/10.1007/s00138-018-0922-2*

[43] Radenović, F., Tolias, G. and Chum, O. [2018], 'Fine-tuning cnn image retrieval with no human annotation', *IEEE transactions on pattern analysis and machine intelligence* **41**(7), 1655–1668.

[44] RICS [2019], 'Rics valuation–global standards, incorporating the ivsc international valuation standards as issued november 2019 and effective from 31 january 2020'.

[45] Roscoe, J. T. [1975], *Fundamental research statistics for the behavioral sciences [by] John T. Roscoe*, Holt, Rinehart and Winston.

[46] Rosen, S. [1974], ‘Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition’, *Journal of Political Economy* **82**(1), 34–55.
URL: *https://www.journals.uchicago.edu/doi/10.1086/260169*

[47] Simonyan, K. and Zisserman, A. [2015], Very Deep Convolutional Networks for Large-Scale Image Recognition, *in* ‘International Conference on Learning Representations’, pp. 1–14.

[48] Sirmans, G. S., Macpherson, D. A. and Zietz, E. N. [2005], ‘The Composition of Hedonic Pricing Models’, *Journal of Real Estate Literature* **13**(1), 1–44.

[49] Starzyńska-Grześ, M. B., Roussel, R., Jacoby, S. and Asadipour, A. [2023], ‘Computer vision-based analysis of buildings and built environments: A systematic review of current approaches’, *ACM Computing Surveys* **55**(13s), 1–25.

[50] Suresh, G., Gnanaprakash, V. and Santhiya, R. [2019], Performance analysis of different cnn architecture with different optimisers for plant disease classification, *in* ‘2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS)’, IEEE, pp. 916–921.

[51] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. and Rabinovich, A. [2015], Going deeper with convolutions, *in* ‘Proceedings of the IEEE conference on computer vision and pattern recognition’, pp. 1–9.

[52] Tajani, F., Morano, P. and Ntalianis, K. [2018], ‘Automated valuation models for real estate portfolios: A method for the value updates of the property assets’, *Journal of Property Investment & Finance* **36**(4), 324–347.

[53] Takahashi, R., Matsubara, T. and Uehara, K. [2019], ‘Data augmentation using random image cropping and patching for deep cnns’, *IEEE Transactions on Circuits and Systems for Video Technology* **30**(9), 2917–2931.

[54] Tan, M. and Le, Q. [2019], Efficientnet: Rethinking model scaling for convolutional neural networks, *in* ‘International conference on machine learning’, PMLR, pp. 6105–6114.

[55] Wan, W. X. and Lindenthal, T. [2023], ‘Testing machine learning systems in real estate’, *Real estate economics* **51**(3), 754–778.

[56] Wilhelmsson, M. [2008], 'House price depreciation rates and level of maintenance', *Journal of Housing Economics* **17**(1), 88–101.
URL: *https://linkinghub.elsevier.com/retrieve/pii/S1051137707000435*

[57] You, Q., Pang, R., Cao, L. and Luo, J. [2017], 'Image-based appraisal of real estate properties', *IEEE transactions on multimedia* **19**(12), 2751–2759.

[58] Zeppelzauer, M., Despotovic, M., Sakeena, M., Koch, D. and Döller, M. [2018], 'Automatic prediction of building age from photographs', *ICMR 2018 - Proceedings of the 2018 ACM International Conference on Multimedia Retrieval* **3206060**(c), 126–134.

[59] Zhang, Y. and Dong, R. [2018], 'Impacts of street-visible greenery on housing prices: Evidence from a hedonic price model and a massive street view image dataset in Beijing', *ISPRS International Journal of Geo-Information* **7**(3).