Leveraging supplementary modalities in automated real estate valuation using comparative judgments and deep learning

Miroslav Despotovic and David Koch
University of Applied Sciences Kufstein, Kufstein, Austria

Eric Stumpe
St. Poelten University of Applied Sciences, St. Poelten, Austria

Wolfgang A. Brunauer
DataScience Service GmbH, Vienna, Austria, and

Matthias Zeppelzauer
St. Poelten University of Applied Sciences, St. Poelten, Austria

Abstract

Purpose – In this study the authors aim to outline new ways of information extraction for automated valuation models, which in turn would help to increase transparency in valuation procedures and thus contribute to more reliable statements about the value of real estate.

Design/methodology/approach – The authors hypothesize that empirical error in the interpretation and qualitative assessment of visual content can be minimized by collating the assessments of multiple individuals and through use of repeated trials. Motivated by this problem, the authors developed an experimental approach for semi-automatic extraction of qualitative real estate metadata based on Comparative Judgments and Deep Learning. The authors evaluate the feasibility of our approach with the help of Hedonic Models.

Findings – The results show that the collated assessments of qualitative features of interior images show a notable effect on the price models and thus over potential for further research within this paradigm.

Originality/value – To the best of the authors' knowledge, this is the first approach that combines and collates the subjective ratings of visual features and deep learning for real estate use cases.

Keywords AVM, Automated valuation models, Deep learning, Computer vision, Hedonic pricing, Subjective judgments, Media content analysis, Elo rating

Paper type Research paper

1. Introduction
In addition to established and standardized real estate valuation methods, which in many jurisdictions are legally binding in certain contexts, Automated Valuation Models (AVMs) are increasingly being used in the profession. Their areas of application range from property and wealth taxation, to the verification of the value of real estate portfolios, to the calculation of insurance and credit risks (Brunauer *et al.*, 2017). The premiums and discounts calculated by an AVM on the price of a piece of real estate in a given market are a function of a dynamic interaction of temporal and geographic factors, but also of intrinsic characteristics, both quantitative and qualitative, of the property itself. In consequence, the quality of an AVM depends on the amount of viable input data describing such characteristics. In general, standardized data is used in these models, which is coded into a readable format and usually represents general information about the property like, for example, purchase price, year of construction, floor area, or the street address, from which in turn further data points (distance to schools, the city center, etc.) can be computed. To supplement this observable objective data, subjective expert assessments of both the property and the wider market are usually gathered, too. These could concern, for example, the condition or location quality of the given piece of real estate.

Given the complex composition of this input, and the large amount of representative data required for Machine Learning (ML)-based AVMs to be effective, data enrichment and data pre-processing represent two of the most important tasks in the entire analytical process chain of AVMs. First, the researcher has to make informed choices regarding the data and variables to draw on in designing their AVM. Needless to say, these will be variables that the analyst assumes will have a marginal effect (in statistical terms) on the price of a property. Hedonic-based AVMs, in particular, use pooled information derived from multiple sources like real market transactions, offer data, or other proprietary and non-proprietary data sources (Glumac and Des Rosiers, 2020). Due to the continuous development in the field of Machine Learning and in order to maximize potential data sources, companies and data analysts increasingly extract information from additional modalities, such as visual or textual data (e.g. from real estate listings), so that researchers can integrate more qualitative data into their AVMs (Desai, 2019). In the case of condominiums and rental properties, interior photos are widely available as part of listings, which makes them easily accessible and promising resource in this regard.

2. Problem statement
As part of the conventional real estate valuation and marketing process, the assessment of qualitative real estate characteristics, as recorded in visual or textual data (images and descriptions in real estate listings, for instance) is an important but demanding task that is largely based on individual expertise. What is more, the individual assessment of criteria such as location quality, interior design, floor plan design, or the general condition of the apartment is often influenced by subjective perception, personal preferences, or self-interest. Unfortunately, not all of these problems are instantly resolved by moving to ML-based automated valuation methods, as manual annotation of qualitative features is a prerequisite for successful Machine Learning. The influence of the annotator's subjective interpretation is

usually not a problem when it comes to simple categorization, such as the detection and classification of objects within an image, e.g. the presence of a shower or a bathtub in a bathroom. This is why the annotation in such case is usually performed by a single person. Subjectivity becomes much more of an issue, however, when the semantic significance of multiple objects within an entire scene needs to be established.

The problem can be explained more clearly with the help of Figure 1. The assessment of the quality of the shown bathrooms can easily suffer under the vagaries of subjective perception, personal preferences, or a lack of expertise on the part of the annotating person. The individual qualitative assessment of new interior items, such as a new kitchen, can be even more complex. The new kitchens shown in the image have a price range between EUR 700–5,000, yet their difference in quality is difficult to assess based on visual data alone. Consequently, qualitative interpretations by a single annotator can lead to biases in ML-based assessment model. To the extent that ML-based systems are designed to improve on these conventional methods, the question remains how such a bias can be minimized or even eradicated.

To this end, we hypothesize that the empirical error in quality estimation in general, and in the interpretation of semantic cues in visual media in particular, can be minimized by combining and collating the assessments of multiple independent individuals. Furthermore, we assume that by employing repeated trials when respondents rate the images, the total error variance within the ratings of an individual and thus overall range of collated ratings of the given visual content can be reduced. Thus, the challenging task in this context is to develop a research design that allows repeated independent assessments by multiple individuals, has acceptable manual and computational effort, and provides reliable results. Motivated by this problem, we developed an experimental approach for semi-automatic extraction of qualitative real estate assessments (i.e. quality ratings of real estate interiors) based on comparative judgments of visual content and computer vision. We investigate our assumptions in a concrete case study by having respondents rate the quality of bathrooms using interior views of residential apartments, by analyzing the effect of their ratings with the rental price of a unit, and by automatically generated ratings using human estimates. Thus, the present study has two main objectives:

(1) To present an approach that efficiently combines and collates the subjective ratings of visual features by multiple individuals and to show how such a method could be easily adapted to Machine Learning approaches in a broader sense.

(2) To evaluate the effect of image-based assessments of the quality of interior spaces and the rental price of apartments, and thus to check the applicability of the approach to real estate use cases.

More broadly, we aim to outline new ways of information extraction for automated valuation models, which in turn would help to increase transparency in valuation procedures and thus contribute to more reliable statements about the value of real estate.

Figure 1.
Exemplary representation of new kitchens (top) and bathrooms (bottom)

Source(s): Justimmo (2021)

3. Literature review

To achieve the objectives outlined in the previous section, we combine approaches from different disciplines and mix quantitative and qualitative paradigm. In the following section, we would like to explain the theoretical background of the mixed methods that have been addressed in this paper. The literature review is structured into three sub-sections that cover important research areas that contributed to our own methodological design: media content analysis, hedonic valuation models, and deep learning.

3.1 Media content analysis and comparative judgments

Measuring visual stimuli based on emotions and subjective judgment is a widely used methodology in many research areas, e.g. by Baveye *et al.* (2018) and Greenwald *et al.* (1989). Despite advances in ML and current software systems' ability to recognize lower-level visual content, humans are still much more capable when it comes to perceiving semantic cues at a high cognitive and affective level. The basic idea for the proposed qualitative assessment is based on work by Hoffmann *et al.* (2012) and Thurstone (1994). It has been widely established that an observer often makes different comparative judgments about the same pair of stimuli on successive occasions. In other words, the observer is inconsistent in his or her comparative judgments from one occasion to the next (Melinger and Schulte im Walde, 2005). According to Thurstone (1994), any such phenomenon is referred to as a fluctuating discriminative process. Following Robinson (2005), we assume that emotions may potentially influence the evaluative judgments of the participants in our own research design. The analysis of emotions that are triggered by audio-visual stimuli is the focus of Affective Content Analysis (ACA). In the respective literature, various approaches to emotion mapping have been proposed, though discrete and dimensional mapping are the predominant ones. In the current study, we deliberately rely on discrete emotion models. Within this framework, there are 22 possible types of discrete emotions that could be triggered when participants view the visual content, and they are usually expressed as binary scales, e.g. pleased/displeased, approve/disapprove, like/dislike, etc (Hoffmann *et al.*, 2012).

Experiments involving ACA have a long tradition in broadcasting research (ITU, 2012). As early as the 1970s, the first concepts were developed to incorporate an automatism into the assessment process, where in the beginning the inferences were achieved by aggregating test results from objective computer-based and subjective human-based experiments (Webster *et al.*, 1993). Affective Media Content Analysis has wide applications to this day, ranging from Sentiment Modeling (Chen *et al.*, 2014) to Human-Computer Interaction and Affective Computing (Bee *et al.*, 2006; Hoffmann *et al.*, 2012) and Image Retrieval (Zhao *et al.*, 2018). A highly complementary method to ACA for more effectively eliciting human responses from interaction with visual media content is Alternative Forced Choice (AFC) (ASTM International, 2009), which is essentially based on the concept of multi-alternative perceptual decision (Ditterich, 2010). The method allows for a scene or an object from the scene to be tested, with the scene or object sharing a common conceptual category and properties but nevertheless differing visually. The forced choice to be performed happens, as the name implies, from multiple but pre-determined alternatives. In the context of ACA, AFC has been used, e.g. for emotional face recognition (Thomas *et al.*, 2007) or to determine individual color preference (Yu *et al.*, 2020).

3.2 Hedonic theory in the context of automated real estate valuation

Computer-aided valuation of properties can be traced back to the 1970s (Carbone and Longini, 1977) and was initially introduced under the term Automated Assessment System (Case, 1978). Today, an Automated Valuation System (AVS) is defined as data analysis software consisting of single or multiple AVMs and a user interface; these elements combined

are used to establish a price estimate for an individual property or parcel of land through a structured decision-making process (Glumac and Des Rosiers, 2020). In its application, an AVM is essentially reliant on the data, the approach, and the method used. In terms of methods, the Hedonic Price Method (HPM) has dominated automated valuation for decades due to its versatility and flexibility. It is compatible with a wide variety of data and can be used along with different automated valuation approaches (e.g. probabilistic, non-probabilistic, market or income approaches). Hedonic theory holds that a good is composed of many characteristics, all of which can affect its value (Rosen, 1974). An HPM analyses the marginal effects of these characteristics on the price of the good. The use of HPMs as an empirical valuation method can be traced back to 1939, originally focusing on the estimation of hedonic price indexes for automobiles (Court, 1939; Goodman, 1998). Research in this area was revived in 1961 in the work of Griliches (1961). Over the past decades, countless theoretical and empirical studies on hedonic pricing in the real estate and housing market have been performed. Good reviews of the literature in this area can be found in the work of Herath and Maier (2010) and Malpezzi (2008), amongst others. For on overview of the underlying functioning of an HPM, see 4.2.

3.3 Deep learning for visual pattern recognition

In certain scientific problems, such as in the one at hand, one is confronted with the application of different data modalities (e.g. tabular and visual data). Convolutional Neural Networks (ConvNets) are used in this study to extract intrinsic information from complex visual features. Like many artificial intelligence (AI) applications, the functionality of ConvNets is based on the theoretical foundations of deep learning, with a dominant focus on the theories of approximation and optimization, as well on the paradigm of representation learning (Bengio *et al.*, 2013). ConvNets enable deep learning of highly representative image features from training data in a layered hierarchical fashion. An effective technique that successfully uses ConvNets for image classification and regression is transfer learning, i.e. fine-tuning the ConvNets models while maintaining criteria of domain adaptation (Shin *et al.*, 2016; Robinson, 2005).

In order to arrive at a holistic price valuation of real estate using various data modalities, it is necessary to automatically include features in the price formula, for which the use of deep learning algorithms is currently the most promising approach. So far, ConvNets have been mainly used for classification of visual data (Koch *et al.*, 2020). A real estate related example is the article of Renigier-Biłozor *et al.* (2022), where human emotions generated by looking at real estate images are classified in order to incorporate the detected emotions into a valuation model. In the article of Glaeser *et al.* (2018) ConvNet is used to evaluate the impact of the exterior and, in Poursaeed *et al.* (2018), the interior visual appearance of a building on prices. The use of ConvNets for regression is not as widespread as for classification problems, but is increasingly gaining application such as for position recognition in buildings (Ballesta *et al.*, 2021). The regression ConvNet methodology is also used for predicting stock prices via annual reports and text analysis (Dereli and Saraclar, 2019) and using historical data (Mehtab and Sen, 2020). Other applications include prediction of angles (Fischer *et al.*, 2015), prediction of distances for 3D position estimates (Mahendran *et al.*, 2017), or age estimation (Rothe *et al.*, 2016). In the real estate field, Solovev and Pröllochs (2021) choose a pretrained ConvNet to predict apartment rent prices using pictures of the floor plans as the input. Shen *et al.* (2022) also use a regression ConvNet to predict rental prices in Wuhan neighborhoods based on the spatial density of points of interest. The authors find that a regression ConvNet outperforms other prediction methodologies. Regression-based ConvNets are of special importance for our approach, where we aim to predict subjective quality assessments from images.

4. Approach

The basis of our approach is the subjective estimation of the quality of bathrooms, where the obtained estimates are integrated into a hedonic model to evaluate their effect on the rental price. In a separate experimental stage, the same estimates are learned from a neural network to evaluate the feasibility of annotating new data. To better illustrate the proposed methods, we have summarized their theoretical foundation in this section. Figure 2 shows the methodological steps in the proposed order.

4.1 Elo rating

In any qualitative assessment that results in a classification, the main problem lies in the unavoidable fuzziness of the decision boundaries between individual classes. Consequently, the implementation of subjectively rated quality classes in the price prediction of a unit of real estate can have in certain scenarios a significant impact on the estimated sales or rental price. This is problematic, because it implies that detailed description lists and substantial experience on the part of the judging person are required to adequately assess the quality of a given real estate characteristic. To overcome this problem, we apply a system in which respondents are successively presented with pairs of randomly selected images from a pool with representative images and only have to decide which of the two images depicts the higher quality interior. By using such subjective serial pairwise comparisons of image features, we can assign images a metric score (and not a class!) that is based on simultaneous as well as successive contrasts as a result of repeated trials. In other words, respondents perform an Alternative Forced Choice between three alternatives (viz. win-lose-draw) whereby the responses are recorded by an automatic backend scoring system. The proposed approach is expandable and requires a sufficient number of respondents, representative images, as well as repeating of trials in order to decrease the error variance continuously and thus to generate useable data.

To measure the quality of real estate interior using loop wise direct comparison of image pairs in a continually updated manner (assuming that estimated quality scores will vary sufficiently in the data distribution) we apply the Elo formula. The Elo rating system is used in practice to quantify the relative abilities of chess players (Elo, 1978). In an Elo rating, each player starts with an initial score and depending on how he/she plays against players with higher or lower Elo, his/her score is being updated according to equations 1 and 3.

Following equation represents expected Elo value of a win for the problem at hand (Tsang *et al.*, 2016): If image A has rating R_A and image B has rating R_B, then the expected value of image A beating image B is given by

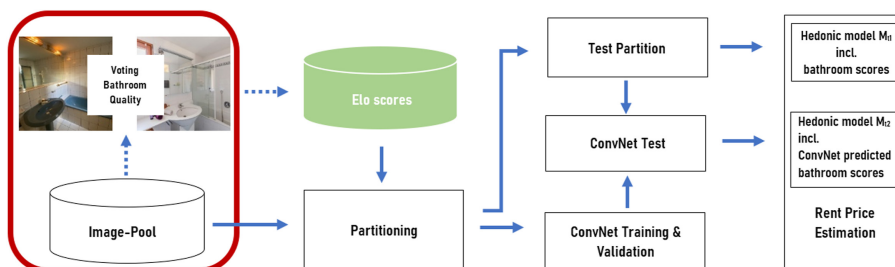$$Pr(A > B) = E_{AB} = \frac{1}{1 + 10^{\frac{R_A - R_B}{400}}} \tag{1}$$

Figure 2.
Overview of the
methodological steps
in the proposed order,
starting with
subjective assessment
based on voting (left),
computation of quality
ratings, data
partitioning and
automatic generation
of ratings (middle) and
concluding with
evaluation using
hedonic models (right)

where R_A and R_B are initialized with 1500. Hence, note the logistic property:

$$E_{AB} + E_{BA} = 1 \qquad (2)$$

The parameter 400 controls the different probabilities of the possible outcomes in favor of either the higher or the lower rated image (Elo, 1978). Having an expected score for image A when voting against image B and three possible outcomes for A, namely win, lose, or draw, which correspond to values of 1, 0, and 0.5, respectively, we calculate the Elo score as follows:

$$R'_A = \begin{cases} R_A + K_A(1 - E_{AB}), if \ A\ wins \\ R_A + K_A\left(\dfrac{1}{2} - E_{AB}\right), if \ A\ draws \\ R_A + K_A(0 - E_{AB}), if \ A\ loses \end{cases} \qquad (3)$$

The constant K (K = 10) is used to adjust the weighting sensitivity of the score update. It is assumed that the Elo algorithm is sufficiently robust to map the scores with appropriate proportionality, provided that player performance resp. Player skills (in the present case the participants) follow a normal distribution, i.e. remain constant over time, which is not always the case in reality (Glickman and Jones, 1999).

4.2 ConvNet regression

Elo scores can be obtained by pair-wise comparisons of human raters as described in Section 4.1. This is, however, time-intensive and thus expensive. We propose to estimate Elo scores automatically using a regression-based ConvNet. Thus, in this experimental phase, we aim to verify whether human-estimated scores can be learned from a ConvNet and generalized to new images unknown to the trained network. For training, we use the scalable EfficientNet network (Tan and Le, 2019). The network uses optimized constants for width, depth, and resolution, as well as the coefficient for available computational resources, to allow adaptive model scaling to set up models for different input sizes and different numbers of floating point operations (FLOPS) - (EfficientNet models B0-B7). The underlying structure of the base model EfficientNet-B0 consists of seven building blocks, each with inverted residual blocks (Sandler *et al.*, 2018) and stem layers (see Figure 3). Stem layers act as a compression mechanism that leads to a rapid reduction in the spatial size of activations, reducing storage and computational costs. Residual blocks are inverted blocks with depthwise separable convolution (Chollet, 2017), which in turn significantly reduces the number of parameters. Furthermore, each residual block contains a squeeze and excitation sub-block that dynamically assigns high weights to more important channels, thus mapping channel dependence while providing access to global spatial information of the input signal.

If the model is going to be trained for a regression task, some modifications of the network are required. This includes adding a batch normalization layer, a dropout layer, and a regression output layer at the top, and changing the last dense layer to 1 neuron (see Figure 3).

Figure 3.
Simplified representation of EfficientNet-B0 architecture with modified top layers (outlined in the figure with dotted lines) for regression task
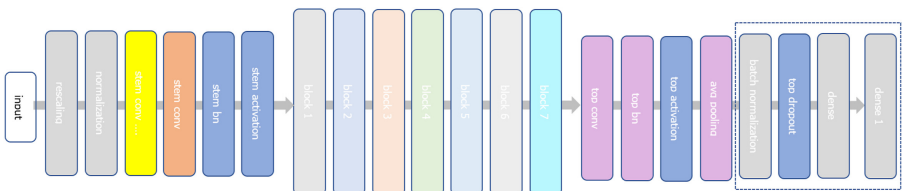
4.3 Hedonic Regression
To evaluate the effect of estimated and predicted Elo scores on the real estate price, we use the Hedonic Price Method (HPM). HPM is also known as Hedonic Regression and is a commonly used method to predict real estate prices by estimating the marginal contribution of real estate characteristics to the price. In general, the hedonic price function has the form

$$P_i = f(Z_i) \tag{4}$$

where P is the price of the unit of real estate and f is a function of the vectorized values Z, which describe the characteristics of the property. The basic assumption in Hedonic Pricing is that the relevant determinants of the dependent variable (price or index) are known in advance. In practice, different variables are used depending on the research question, the preferences of the researchers, or the availability of data (Herath and Maier, 2010). Sirmans *et al.* (2009) summarize 470 possible variables for Hedonic Pricing that have been used in the scientific literature to date. If we divide real estate characteristics into three main subcategories, then the price function has the form:

$$P_i = f(S_i; L_i; N_i) \tag{5}$$

where S_i is a vector of structural real estate characteristics, L_i is a vector of location variables, and N_i represents neighborhood characteristics. In the present study, we include the human estimates in the structural variables *S* and omit the neighborhood characteristics *N*. For the variable setup in our hedonic model, see Section 5.3.2.

5. Experimental setup
In line with the proposed methodological steps (see Figure 2 in Section 4), the following section outlines criteria for evaluating the performance of our model in relation to the research questions, provides a detailed description of the data, and lists individual experimental steps.

5.1 Research questions
Our objective is to answer the following research questions:

(1) RQ1: To what extent can qualitative characteristics of real estate interiors be derived from photographs by means of repeated comparative judgments of multiple individuals and how significant is their effect on the rental price of apartments?

(2) Beyond this, we want to answer RQ2: To what extent is a ConvNet regressor able to generate plausible quality judgments by learning these human estimates from associated images and are these predicted judgments beneficial for real estate price estimation?

The research questions are discussed in Sections 6 and 7.

5.2 Data
For the human-based estimation of the bathroom quality, we use the initial pool D_i with manually selected 1,000 representative images of bathrooms. The images originate from the real estate listings published in the year 2020 (Justimmo, 2021) which include also structural and location characteristics. Of these, 250 images correspond to instances in test dataset P_1 which includes structural and location characteristics of 250 rental apartments (one image per apartment). The test data set P_1 is then used to derive the effect of human-estimated as well as of ConvNet-predicted bathroom quality scores on the apartment rental price (see Sections 5.3.3 and 5.3.4). The remaining 750 images from D_i were used for the training and

validation of the ConvNet (see Section 5.3.3). For the human estimation, the images were not processed, i.e. they were used in their original shape and size.

5.3 Steps of the experiment
In the sub-sections below, the individual experimental steps are described in detail.

5.3.1 Elo rating: setup for human judgments. We customized a web browser application (Gerneth, 2014) based on the Python modules Flask and Sqlite. Using the application's user interface (see Figure 4), the participant can vote between two displayed images according to Three-Alternative Forced Choice (3AFC) paradigm between a win for either image or a draw for the displayed image pair. For each vote a quality score is calculated backend based on Elo rating (see Equations 1 and 3). For each voting round with a respondent, 500 pairs of images are randomly regenerated out of the image pool D_i with 1000 representative images. Note that not all possible combinations of image pairs can be tested in the procedure, since this would result in 499,500 possible combinations for 1,000 images (n(n-1)/2). The voting incorporates 16 voting rounds with 8 participants and 2 repeated trials per participant whereas each person could repeat the trial only after the remaining 7 participants have voted once. After each voting round all estimated Elo scores are saved into a separate database table resulting finally in a final data set S_c with 1000 scores for each image from D_i.

5.3.2 Evaluation of the effect of human estimated scores on the rental price. This experimental setup is designed to address research question RQ1 (see Section 5.1). We union test data set P_i with associated scores from S_c as test data T_{s1}. We use test set T_{s1} to estimate target variable apartment *gross rent price* in a hedonic model M_{t1} by including estimated scores along with the structural apartment characteristics *living area, year of construction, floor, garden,* and overall *condition* of apartment as predictor variables. Thereby, we infer the explanatory power of different settings of the model and the marginal
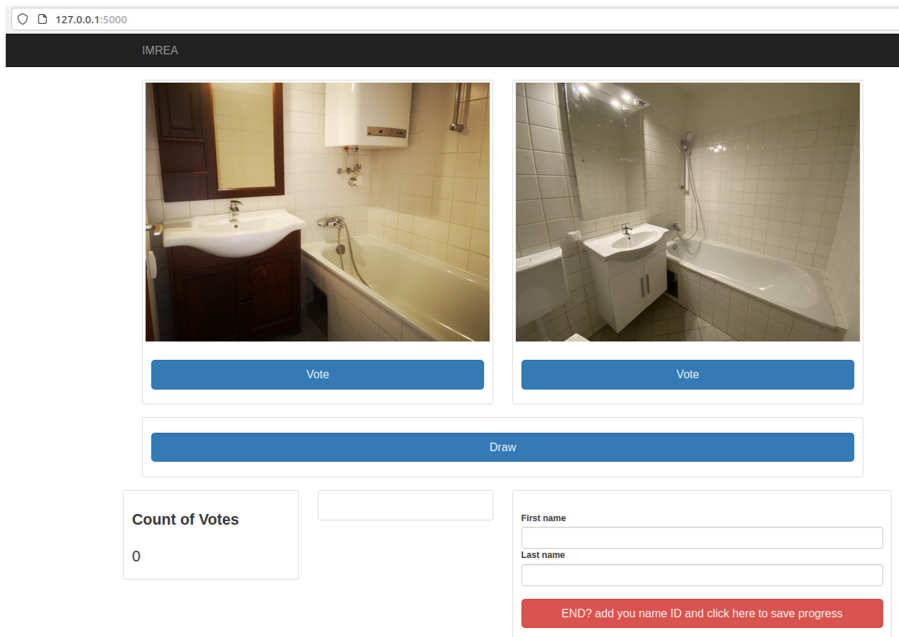

Figure 4.
Graphical user interface of the voting application

effects of selected predictors. Hereby, we control model for the apartments' overall condition, which permits a more conclusive inference on the effect of estimated scores.

5.3.3 Training of ConvNet with human estimated scores. The following three experimental steps aim to answer research question RQ2. (see Section 5.1). Using the initial set of 1,000 images from D_i and human estimated scores from S_c we partition the data in a ratio of 65:25:10 as training set Tr_{650}, test set T_{250} and validation set V_{100} whereas the 250 test images from T_{250} correspond to instances in the test dataset P_1. Thus, we assign from Tr_{650} human estimated scores as the response variable and associated images as the predictor variable and train the ConvNet for the regression task. For training, we first crop the images to a square shape, starting from the center of the image maximizing the image dimension by either the original height or width. Then the images are scaled to 224×224 according to the dimension of the EfficientNet0 input. For the training we set training parameters as shown in Table 1. We limit the image augmentation to horizontal flipping only, as additional augmentation showed a negative impact on the network performance. We decide to let the ConvNet converge more slowly, thus reducing the learning rate on plateau when the loss stops decreasing, starting from 0.001 through 0.0002 to a minimum learning rate of 0.0001. We do not apply early stopping for regularization, but post select the training stage with the best performance. As model metric for validation and training we apply mean-squared-error (MSE) loss and root-mean-squared-error (RMSE).

5.3.4 Evaluation of the effect of ConvNet predicted quality scores on rental price. We evaluate the scores predicted by ConvNet in a hedonic model. Thus, we use the trained ConvNet, to predict scores for images in the data set T_{250}. Subsequently, we union test data set P_i with predicted scores as test data T_{s2}. We use test set T_{s2} to estimate target variable *gross rent price* in a hedonic model M_{t2} having exactly the same evaluation settings and having the same data for the remaining features as for model Mt_1.

5.3.5 Training and evaluation of convnet using training sets of different sizes. Final experimental step aims to evaluate to what extent the training of ConvNet can benefit from the size of the training set resp. From additional training data and how this affects the price estimation. We quantify the degree of improvement using root mean squared error (RMSE) to weight the cost of creating additional annotated training data against the expected benefit, e.g., if the expected benefit of additional training data is small, it may not justify the cost of creating additional annotations. We randomly subset training set Tr_{650} into two smaller training sets Tr_{450}, and Tr_{250}. Subsequently we train the network with training sets Tr_{650}, Tr_{450}, and Tr_{250} and evaluate all three models on the test images in T_{250}. We then merge test data set P_i with the predicted values of the three trained networks and test the effect applying the 2nd setting of the hedonic model M_{t2}.

6. Results

In the following section, we first analyze the results for human estimates and their ConvNet predictions. We then evaluate their impact on the rental price in two hedonic models. In the last step, we elaborate on the results on Convnet training performance with respect to training data of different sizes. Section 7 discusses the results in relation to the research questions

Optimizer	Adam	Learning rate	0.001
Batch size	16	Epochs	1000
Augmentation	Horizontal flip	Reduce learning rate on plateau	min lr 0.0001
Top dropout rate	0.3		factor 0.2
			patience 21

Table 1.
Settings applied for the ConvNet training

posed. For a better understanding, the estimated values for the bathroom quality will henceforth mainly be referred to as human estimated scores.

6.1 Analyses of human estimated scores and their ConvNet predictions

We examine in the first step selected statistics of interest for the human estimated and predicted scores. Starting from an initial Elo score (R = 1500), the min-max range and the variance of the human estimated scores have increased with each subsequent voting round and revealed expectedly a slightly double-humped data distribution. The final min-max range is eventually 129 Elo scores. The plot in Figure 5 shows positive correlation with moderate relationship between the human estimated and predicted scores which is also confirmed by Pearson correlation coefficient (0.426).

It is assumed that the variables *estimated scores* and overall *condition* could have both, a shared and an independent effect on the apartment rental price, which could also be observed. The Kruskal–Wallis rank sum test indicates small p value (0.001826). Therefore, it is expected that there are noticeable differences between the two variables in their central tendency. If we incorporate a smoothing conditional function into the scatter plot (Figure 5), we observe a tendency that bathrooms with higher scores are close to new and fully renovated apartments. On the other hand, bathrooms with lower scores tend to cover apartments with less attractive quality states such as *well-kept*, *like-new* and *partly renovated*, whereas the confidence band for partially renovated apartments shows obvious uncertainties in the estimates. This is mainly due to an imbalance of the condition classes, respectively the ratio of the cardinalities of the condition classes *partly renovated* and *well-kept* is 1:12.5. We can also find the weakest correlation between the scores and apartments' overall condition *like-new*. The frequency analysis for the variable *year of construction* and apartments' condition *like-new* shows distribution of instances across all construction periods. This calls into question the relevance of the feature expression *like-new*.

D'Agostino's skewness test applied to human estimated scores shows certain skewness in the data distribution ($p = 0.01216$). However, the modeling showed no improvement by incorporating the logarithmic transformation of this variable. The same applies to the predicted scores.
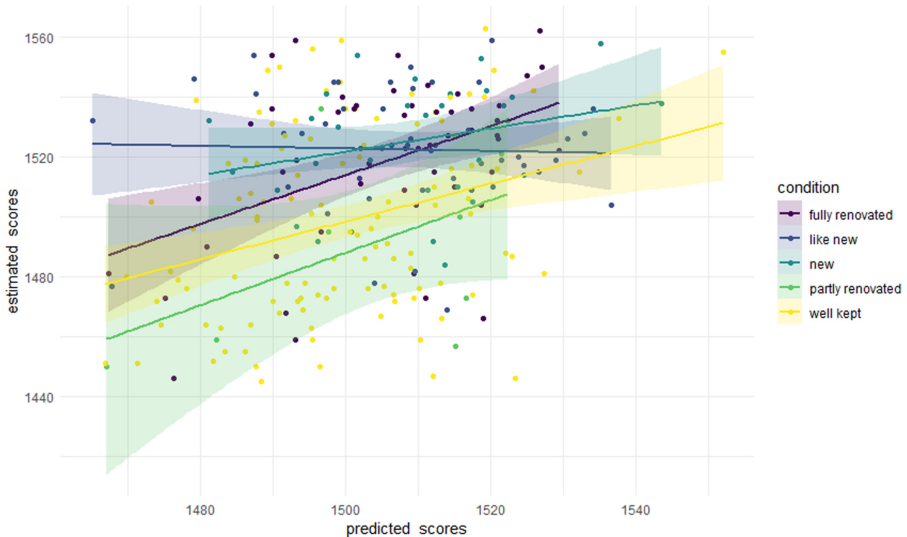

Figure 5.
The scatter plot shows the relationship between the human estimated and predicted values aggregated across the classes of the overall condition of the apartment

The visual comparison of the images and estimated scores reflects the subjects' subjective perception of what is seen in Figure 6, which shows the same bathroom photos as in Figure 1, this time with human-estimated scores in the upper right and ConvNet predicted scores in the lower right. The given images were selected for the illustration because the qualitative differences of the interiors and the deviations between the ground truth and the predictions are obvious in them.

We also examine the change in the empirical error of the estimated bathroom scores per voting round by setting up the hedonic model M_{t1} for the estimation of log-transformed response variable gross rent. Table 2 shows the model performance after voting rounds 8, 10, 12, and 16, indicating the standard errors, the coefficients, the T- and p-values of the scores, as well as the adjusted R^2 of the model and the p-value of the F-test. The decreasing standard error, p-value, price premium as well as p-value of the F-test per voting round are noticeable. On the other hand, increasing T-values and Adj. R^2 can be observed, whereby the increase of the Adj. R^2 is not excessive in view of the low price premium imposed by the bathroom scores.

6.2 Evaluation of the effect of estimated quality scores on rental price

In this stage, we examine the extent to which the estimated scores have an impact on the rental price of the apartments if we control for additional variables of interest. Table 3 shows summary of four different settings of the hedonic model M_{t1}. In the table, the standard errors are given below the estimators, rounded to 3 decimal places for clarity and the significance of the predictors can be easily derived from the indicated asterisks. In the 1st setting, the response variable is regressed without any of two predictors for condition quality. In the 2nd setting, estimated scores are added, in the 3rd setting the model is controlled for the variable overall condition of the apartments (estimated by experts or real estate agents). The variable living area shows pronounced skewness ($p = 4.331 \times 10^{-16}$) and is therefore log-transformed in the 4th setting expecting an additional model improvement.

In all 4 settings, the response variable gross rent is log-transformed. Reference levels for predictors city, overall condition and floor (not shown in regression outputs in Tables 3 and 4) are city of Klagenfurt, new apartment and 1st floor.

In the model M_{t1}, the values of the variable bathroom scores in the 2nd setting indicate a significant effect on the response variable. As we already stated in Section 6.1, it cannot necessarily be inferred from the condition of an individual room that the entire apartment is in

Source(s): Justimmo (2021)

Voting round	Standard error	T Value	P Value	F Test	adjR^2	Coefficient
8	0.0007188	4.538	0.000008983	4.7797E-85	0.8194	0.003262
10	0.0005741	4.711	0.000004184	2.3087E-85	0.8205	0.002705
12	0.0004195	4.78	0.00000306	1.7133E-85	0.821	0.002005
16	0.0003696	4.852	0.000002208	1.2544E-85	0.8215	0.001793

	Setting 1	Setting 2	Setting 3	Setting 4
(Intercept)	5.755 (***)	3.074 (***)	3.806 (***)	1.048 (*)
	0.044	0.554	0.586	0.515
city Salzburg	0.446 (***)	0.438 (***)	0.440 (***)	0.470 (***)
	0.034	0.033	0.032	0.029
city St. Pölten	0.120 (**)	0.128 (***)	0.086 (*)	0.099 (**)
	0.039	0.037	0.037	0.033
living area	0.010 (***)	0.010 (***)	0.010 (***)	
	0	0	0	
poly(yoc, degree = 2)1	0.519 (**)	0.348 (*)	0.232	0.244
	0.168	0.165	0.175	0.157
poly(yoc, degree = 2)2	1.294 (***)	1.000 (***)	0.775 (***)	0.673 (***)
	0.169	0.173	0.189	0.169
floor high	−0.038	−0.024	−0.015	0.009
	0.046	0.044	0.042	0.038
floor midd	−0.016	−0.015	−0.009	0.001
	0.024	0.023	0.022	0.02
floor parterre	−0.005	−0.004	0.001	−0.003
	0.04	0.039	0.038	0.034
garden	0.04	0.048	0.062	0.080 (**)
	0.036	0.035	0.034	0.03
bathroom scores		0.002 (***)	0.001 (***)	0.002 (***)
		0	0	0
condition like new			−0.095 (**)	−0.092 (**)
			0.036	0.032
condition well kept			−0.164 (***)	−0.150 (***)
			0.038	0.034
condition partly renovated			−0.141 (*)	−0.142 (*)
			0.065	0.058
condition fully renovated			−0.112 (**)	−0.097 (*)
			0.042	0.038
log(living area)				0.774 (***)
				0.024
Adj. R 2	0.805	0.826	0.832	0.866
Adj. R2 w/o bathroom scores		0.805	0.8237	0.8558
Num. Obs	250	250	250	250

Table 3.
Performance of the M_{t1} model with log-transformed response variable gross rent and four settings: without (1) and with (2) estimated scores, controlled for apartment's overall condition (3), and with log transformed living area (4)

an equivalent general condition what has been previously proved using Kruskal–Wallis rank sum test. Analogously, a strong relationship between overall condition and the response variable was found using the same statistical test as well as moderate correlation of 0.359 between estimated scores and the response variable was also observed. Further support for the effect of estimated scores on rental price is found in the 3rd model setting, in which we control for the condition of the apartment. If we include solely the condition of the apartment without scores, the model improves to 0.8237 compared to the baseline model setting and, as expected, shows a stronger effect than estimated scores in the 2nd setting. However, this effect is significantly smaller in the 3rd setting if we include estimated scores. In addition, the scores in the 3rd setting remain significantly high.

This confirms that there is additional informational content in the images that our human respondents were able to extract. In addition to these effects of central concern, additional relevant observation can be made based on the 4th experimental setting of the model. The adjusted coefficient of determination in the 4th setting shows notable improvement in the model's goodness of fit due to effect of the log-transformed living area. Since the variable year of construction shows a non-linearity, as expected, the application of orthogonal polynomials is superfluous, whereby the significance of the variable has increased noticeably. Also, in this

	Setting 1	Setting 2	Setting 3	Setting 4
(Intercept)	5.755 (***)	2.309 (*)	3.295 (***)	1.195
	0.044	1	0.983	0.894
city Salzburg	0.446 (***)	0.434 (***)	0.439 (***)	0.473 (***)
	0.034	0.034	0.032	0.03
city St. Pölten	0.120 (**)	0.114 (**)	0.071	0.084 (*)
	0.039	0.038	0.038	0.034
living area	0.010 (***)	0.010 (***)	0.010 (***)	
	0	0	0	
poly(yoc, degree = 2)1	0.519 (**)	0.482 (**)	0.351 (*)	0.382 (*)
	0.168	0.165	0.173	0.157
poly(yoc, degree = 2)2	1.294 (***)	1.174 (***)	0.892 (***)	0.819 (***)
	0.169	0.169	0.185	0.168
floor high	−0.038	−0.027	−0.017	0.006
	0.046	0.045	0.043	0.039
floor midd	−0.016	−0.017	−0.011	−0.001
	0.024	0.023	0.022	0.02
floor parterre	−0.005	0.004	0.007	0.001
	0.04	0.04	0.038	0.035
garden	0.04	0.036	0.057	0.077 (*)
	0.036	0.036	0.035	0.031
predicted bathroom scores		0.002 (***)	0.002 (**)	0.001 (*)
		0.001	0.001	0.001
condition like new			−0.093 (*)	−0.091 (**)
			0.037	0.033
condition well kept			−0.176 (***)	−0.167 (***)
			0.039	0.035
condition partly renovated			−0.166 (*)	−0.170 (**)
			0.065	0.059
condition fully renovated			−0.100 (*)	−0.085 (*)
			0.042	0.039
log(living area)				0.793 (***)
				0.024
Adj. R 2	0.805	0.814	0.828	0.858
Adj. R2 w/o predicted scores		0.805	0.8237	0.8558
Num. Obs	250	250	250	250

Table 4.
Performance of the M_{t2} model with log-transformed response variable gross rent and four settings: without (1) and with (2) estimated scores, controlled for apartments' condition (3), and with log-transformed living area (4)

4th setting, the scores' significance remains high. Based on the present results, it can be confirmed that there is a notable influence of the human estimated scores on the response variable. An in-depth analysis of these results is provided in Section 7.

6.3 Evaluation of the effect of predicted scores on the rental price
In this subsection, we examine the effect of scores for the quality of bathrooms on the rental price predicted by ConvNet.

Table 4 shows the four settings of the model M_{t2}, which includes network's predictions in place of the human estimated scores. Except for this key change, the data and the variable setup is exactly the same as in model M_{t1}. The integration of the ConvNet-predicted scores into the model (2nd setting), improves the goodness of fit of the model by solely 0.9%. We note the high significance of the predicted scores in this setting of the model. Incorporating the variable general condition of the apartment in the 3rd setting increases the goodness of fit of the model by 1.4%, while the significance of the predicted scores remains high, but with a clear decrease of p value from 0.000664 to 0.00889. We can observe the same effect in the 4th setting if we include log-transformed living area, predicted values remain in the significant

range though with a decrease in p value. Besides, it can be observed that the goodness of fit in the settings 3 and 4 in the Model M_{t2} is slightly lower than in the Model M_{t1} due to the weaker effect of the predicted scores. This confirms the more significant explanatory power of human-estimated scores over network predictions. Finally, we note that the correlation between the predicted scores and the response variable (0.243) is weaker compared to the human estimates and the response variable (0.358).

6.4 Training and evaluation of convnet using training data of different size

With the following experiments we want to evaluate how important the size of the training set is for the task of scores prediction and if the evaluated network architecture can effectively take benefit of additional training data. Knowing the impact of additional training data is important in practice, as it helps to weigh the cost of creating additional labeled training data against the expected benefit. Thus, with the help of this experiment, we investigate the explanatory power of the network using training data of different size and infer in Section 7 if this in practice eventually can impact the effect of predicted scores in an applicable hedonic model.

To answer this question, we perform ConvNet training on differently large training sets Tr_{650}, Tr_{450}, and Tr_{250} to see if additional training data improves test performance. The evaluation is performed in on the same test set, i.e. for all training partitions Tr_{650}, Tr_{450}, and Tr_{250} there is one test set T_{250} which contains unseen independent data. Quantitative results are summarized in Table 5. To investigate the impact of the dataset size on scores extraction performance we always train EfficientNet-B0 using same pre-trained weights, first with the largest training set Tr_{650} and then with the remaining two smaller training sets. The results show that the generalizability of the network decreases with each smaller dataset used for training. The RMSE, for instance, shows decreasing of 9.5 and 4.8 prediction errors for models with training sets Tr_{450} and Tr_{650}. Also, a slight improvement in the model goodness of fit in the M_{t2} in 2nd Setting is noted when we use more data for training.

7. Discussions
7.1 Limitations of the study
There are several limitations to this study that should be acknowledged.

(1) The sample size for experiments is not large. The amount of the test and training data for ConvNet as well as data for hedonic models depends predominantly on the overall number of voted images. We had to design the subjective comparative experiments in such a way that the duration of a voting round would not be excessively long and would not strongly affect the validity of the responses.

(2) There are no reference data sets in the scientific community that could be used for the present problem at hand.

(3) Lastly, there are no research studies on this topic within the real estate domain to be compared.

Training data	Best training epoch	Training RMSE	Best validation epoch	Validation RMSE	Test RMSE	Adj.R^2 model M_{t2}
650	998	36.560	455	25.721	29.578	0.814
450	916	42.084	676	26.825	33.482	0.8105
250	898	46.872	710	28.777	43.736	0.8085

Table 5.
Performance of the ConvNet model using different amounts of training data

Despite mentioned limitations, this study provides valuable insights into the potential benefits of the approach presented. The results support the need for further research to confirm and extend these findings, and to address the limitations of this study.

7.2 Interpretation of results and key findings
There is a large amount of available multimodal, semi-structured and unstructured useful information that can be incorporated into AVMs and thus can provide deep and comprehensive insights to appraisers and help to draw further conclusions in price estimation. By integrating alternative data extraction approaches, as well as the ability to process larger amount of data using Machine Learning, modeling difficulties can be significantly reduced and data from new sources and modalities can be obtained (Lahat *et al.*, 2015). As for the problem at hand, the usefulness of human-based qualitative estimation of individual apartment rooms and the transfer of the estimates to a Machine Learning model for prediction on unknown data is justified if an effect of the human estimates and predictions on property price can be confirmed within a potentially applicable model.

The data obtained in this study through human subjective judgments demonstrate the desired effect in two ways: (1) empirical error in quality estimation is being minimized by repeated interpretation of visual semantic clues by multiple subjects (Table 2), (2) the human-estimated quality scores in this study showed a well-correlated relationship and a discernible stable effect on the response variable in the hedonic model (Table 3). This is also validated in model settings where we control the effect for the overall condition of the apartment. This confirms that, despite the limitation mentioned, our methodology for eliciting and computing qualitative ratings (i.e. the Elo scores) provided a sound approximation of subjects' affective responses.

We tried to answer the question to what extent it is possible to generate synthetic subjective quality judgments, i.e. to annotate new data in an automatic fashion using information gained from the previous experiment. For this we could observe the following. (1) The price premium of the scores in both hedonic models is small, which is to be expected (Tables 3 and 4), and from Table 4 that (2) the predicted scores in the second model in all settings are within significant range, (3) and the predicted scores in the 2nd model lose significance once stronger marginal effects of additional predictors occur. However, the observed results indicate that the network can detect additional information content in the images resp. Demonstrates a recognizable approximation. The loss of variable significance and goodness of fit in M_{t2} reinforces the assumption that with a larger amount of training data, the approximation power of the network should be improved and thus the effect of predicted scores on the response variable in the model.

The values for the M_{t2} model goodness of fit in Table 5 are attributable to the use of more training data. The results in regard to generalizability of the neural network confirm that additional training data contribute to higher prediction accuracy. These results are consistent with previous findings on computer vision methods in the literature (Luo *et al.*, 2018; Li *et al.*, 2016) that network architectures successfully benefit from additional training data and thus generalize better to independent test data.

8. Conclusions
We presented a study on incorporating repeated subjective human judgments about bathroom quality using visual information for the integration into AVMs. We have also tested the robustness and generalization capability of a popular convolutional network architecture for a regression task to predict estimated human judgments. Thereby, we would like to point out that our study is intentionally open-ended and primarily designed to establish potential

applicability of the approach, i.e., the results were obtained without specific optimization of any of the methods and thus used to establish a baseline. Our goal was to show how coherently the presented methods work on the existing dataset in the featured scenario without further tuning them and thereby potentially overfitting them to the data.

With this study and our findings, we want to extend the existing information extraction methods for automated valuation models, which in turn would contribute to a higher transparency of valuation procedures and thus to more reliable statements about the value of real estate. We see several promising directions for extending our study in future: (1) The possibility to process more information by controlling the number of inputs for rating, (2) applicability of the proposed approach in different scenarios (different rooms, or location quality), (3) combining visual data from additional context (e.g. floor plans) to extend the information content and thus improve the generalization of the estimates.

References

ASTM International (2009), "Standard terminology relating to sensory evaluations of materials and products, e253-09a", ASTM International, West Conshohocken, PA. E253-09a.

Ballesta, M., Payá, L., Cebollada, S., Reinoso, O. and Murcia, F. (2021), "A cnn regression approach to mobile robot localization using omnidirectional images", *Applied Sciences*, Vol. 11, p. 75.

Baveye, Y., Chamert, C., Dellandréa, E. and Chen, L. (2018), "Affective video content analysis: a multidisciplinary insight", *IEEE Transactions on Affective Computing*, Vol. 9, pp. 396-409.

Bee, N., Prendinger, H., Nakasone, A., André, E. and Ishizuka, M. (2006), "AutoSelect: what you want is what you get: real-time processing of visual attention and affect", in *Perception and Interactive Technologies. PIT 2006. Lecture Notes in Computer Science*, Springer, Berlin, Heidelberg, Vol. 4021.

Bengio, Y., Courville, A.C. and Vincent, P. (2013), "Representation learning: a review and new perspectives", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 35, pp. 1798-1828.

Brunauer, W.A., Weberndorfer, R. and Feilmayr, W. (2017), "A statistically founded sales comparison approach", 24th Annual European Real Estate Society Conference. *ERES*, Delft, Netherlands.

Carbone, R. and Longini, R. (1977), "A feedback model for automated real estate assessment", *Management Science*, Vol. 24, pp. 241-248.

Case, K.E. (1978), *Property Taxation: The Need for Reform*, Ballinger Publishing Company, Cambridge, Massachusetts, ISBN 139780884104858.

Chen, T., Yu, F.X., Chen, J., Cui, Y., Chen, Y. and Chang, S. (2014), "Object-based visual sentiment concept analysis and application", *Proceedings of the 22nd ACM international conference on Multimedia*.

Chollet, F. (2017), "Xception: deep learning with depthwise separable convolutions", *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1800-1807.

Court, A.T. (1939), "Hedonic price indexes with automotive examples", *The Dynamics of Automotive Demand*, pp. 98-119.

Dereli, N. and Saraclar, M. (2019), "Convolutional neural networks for financial text regression", *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, pp. 331-337.

Desai, J. (2019), "Attribute recognition in real estate listings", available at: https://www.zillow.com/tech/attribute-recognition-in-real-estate listings/ (accessed 30 June 2022).

Ditterich, J. (2010), "A comparison between mechanisms of multi-alternative perceptual decision making: ability to explain human behavior, predictions for neurophysiology, and relationship with decision theory", *Frontiers in Neuroscience*, Vol. 4 No. 184, doi: 10.3389/fnins.2010.00184.

Elo, A.E. (1978), *The Rating of Chessplayers, Past and Present*, Arco Publishing, New York.

Fischer, P., Dosovitskiy, A. and Brox, T. (2015), "Image orientation estimation with convolutional networks", *German Conference on Pattern Recognition*.

Gerneth, C. (2014), "Facemash", available at: https://github.com/c7h/facemash (accessed 22 November 2021).

Glaeser, E., Kincaid, M.S. and Naik, N. (2018), "Computer vision and real estate: do looks matter and do incentives determine looks", *National Bureau of Economic Research Working Paper Series*, Supp. 25174, doi: 10.3386/w25174.

Glickman, M. and Jones, A. (1999), *Rating the Chess Rating System*, Chance Berlin Then New York, New York, Vol. 12, pp. 21-28.

Glumac, B. and Des Rosiers, F. (2020), "Towards a taxonomy for real estate and land automated valuation systems", *Journal of Property Investment and Finance*, Vol. 39 No. 5, pp. 450-463, doi: 10.1108/JPIF-07-2020-0087.

Goodman, A. (1998), "Andrew court and the invention of hedonic price analysis", *Journal of Urban Economics*, Vol. 44, pp. 291-298.

Greenwald, M., Cook, E. and Lang, P. (1989), "Affective judgment and psychophysiological response: dimensional covariation in the evaluation of pictorial stimuli", *Journal of Psychophysiology*, Vol. 3, pp. 51-64.

Griliches, Z. (1961), "Hedonic price indexes for automobiles: an econometric of quality change", *The Price Statistics of the Federal Goverment*, National Bureau of Economic Research, Cambridge, Massachusetts, pp. 173-196.

Herath, S. and Maier, G. (2010), "The hedonic price method in real estate and housing market research. A review of the literature", *SRE-Discussion Papers 2010/03*, WU Vienna University of Economics and Business.

Hoffmann, H., Scheck, A., Schuster, T., Walter, S., Limbrecht, K., Traue, H. and Kessler, H. (2012), "Mapping discrete emotions into the dimensional space: an empirical approach", *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 3316-3320.

ITU (2012), "Recommendation BT.500: Methodology for the subjective assessment of the quality of television pictures", International Telecommunication Union (ITU), available at: https://www.itu.int/rec/R-REC-BT.500 (accessed 12 April 2022).

Justimmo (2021), *Customized Data for Austrian Real Estate Market*, B&G Consulting & Commerce GmbH, Vienna.

Koch, D., Despotovic, M., Leiber, S., Sakeena, M., Döller, M. and Zeppelzauer, M. (2020), "Real estate image analysis: a literature review", *Journal of Real Estate Literature*, Vol. 27, pp. 269-300.

Lahat, D., Adali, T. and Jutten, C. (2015), "Multimodal data fusion: an overview of methods, challenges, and prospects", *Proceedings of the IEEE*, Vol. 103 No. 9, pp. 1449-1477, doi: 10.1109/JPROC.2015.2460697.

Li, M., Ma, L., Blaschke, T., Cheng, L. and Tiede, D.A. (2016), "Systematic comparison of different object-based classification techniques using high spatial resolution imagery in agricultural environments", *International Journal of Applied Earth Observation and Geoinformation*, Vol. 49, pp. 87-98.

Luo, C., Li, X., Wang, L., He, J., Li, D. and Zhou, J. (2018), "How does the data set affect CNN-based image classification performance?", *5th International Conference on Systems and Informatics (ICSAI), Nanjing, China, 2018*, pp. 361-366.

Mahendran, S., Ali, H. and Vidal, R. (2017), "3D pose regression using convolutional neural networks", *IEEE International Conference on Computer Vision Workshops (ICCVW)*, pp. 2174-2182.

Malpezzi, S. (2008), "Hedonic pricing models: a selective and applied review", in O'Sullivan, T. and Gibb, K. (Eds), *Housing Economics and Public Policy*, Wiley-Blackwell, pp. 67-89.

Mehtab, S. and Sen, J. (2020), "Stock Price Prediction Using Convolutional Neural Networks on a Multivariate Time Series", *Proceedings of the 3rd National Conference on Machine Learning and Artificial Intelligence*, February, 2020, New Delhi, India, available at SSRN: https://ssrn.com/abstract=3665363

Melinger, A. and Schulte im Walde, S. (2005), "Accumulation of visual memory for natural scenes: a medium-term memory?", *Proceedings of the Annual Meeting of the Cognitive Science Society*, Vol. 27.

Poursaeed, O., Matera, T. and Belongie, S. (2018), "Vision-based real estate price estimation", *Machine Vision and Applications*, Vol. 29, pp. 667-676.

Renigier-Biłozor, M., Janowski, M., Walacik, M. and Chmielewska, A. (2022), "Human emotion recognition in the significance assessment of property attributes", *Journal of Housing and the Built Environment*, Vol. 37, pp. 23-56.

Robinson, J. (2005), *Deeper than Reason: Emotion and its Role in Literature*, Music, and Art, Press, Oxford, ISBN-10: 0199263655.

Rosen, S. (1974), "Hedonic prices and implicit markets: product differentiation in pure competition", *Journal of Political Economy*, Vol. 82, pp. 34-55.

Rothe, R., Timofte, R. and Van Gool, L. (2016), "Deep expectation of real and apparent age from a single image without facial landmarks", *International Journal of Computer Vision*, Vol. 126, pp. 144-157.

Sandler, M.H., Zhu, A.G., Zhmoginov, M. and Chen, L. (2018), "MobileNetV2: inverted residuals and linear bottlenecks", *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4510-4520.

Shen, H., Li, L., Zhu, H. and Li, F. (2022), "A pricing model for urban rental housing based on convolutional neural networks and spatial density: a case study of wuhan, China", *ISPRS International Journal of Geo-Information*, Vol. 11, p. 53.

Shin, H., Roth, H.R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D.J. and Summers, R.M. (2016), "Deep convolutional neural networks for computer aided detection: CNN architectures, dataset characteristics and transfer learning", *IEEE Transactions on Medical Imaging*, Vol. 35, pp. 1285-1298.

Sirmans, G.S., Macpherson, D.A. and Norman, Z.E. (2009), "The composition of hedonic pricing models", *Journal of Real Estate Literature*, Vol. 13, p. 3.

Solovev, K. and Pröllochs, N. (2021), "Integrating floor plans into hedonic models for rent price appraisal", *Proceedings of the Web Conference*, pp. 2838-2847.

Tan, M. and Le, Q.V. (2019), "EfficientNet: rethinking model scaling for convolutional neural networks", *Proceedings of the 36th International Conference on Machine Learning*, ICM, Long Beach, 9-15 June 2019, 6105-6114.

Thomas, L.A., De Bellis, M.D., Graham, R. and LaBar, K.S. (2007), "Development of emotional facial recognition in late childhood and adolescence", *Developmental Science*, Vol. 10 No. 5, pp. 547-558.

Thurstone, L.L. (1994), "A law of comparative judgment", *Psychological Review*, Vol. 34, pp. 273-286.

Tsang, S.C., Ngan, Y.T.H. and Pang, G. (2016), "Fabric inspection based on the Elo rating method", *Pattern Recognition*, Vol. 51, pp. 378-394.

Webster, A.A., Jones, C., Pinson, M.H., Voran, S.D. and Wolf, S. (1993), "Objective video quality assessment system based on human perception", *SPIE Human Vision, Visual Processing, and Digital Display*, Vol. 4, pp. 15-26.

Yu, L., Westland, S. and Li, Z. (2020), "Analysis of experiments to determine individual colour preference", *Color Research and Application*, Vol. 46 No. 1, pp. 155-167.

Zhao, S., Ding, G., Huang, Q., Chua, T., Schuller, B. and Keutzer, K. (2018), "Affective image content analysis: a comprehensive survey", *International Joint Conference on Artificial Intelligence*.

Further reading

Capizzi, M.T. and Ferguson, R. (2005), "Loyalty trends for the twenty-first century", *Journal of Consumer Marketing*, Vol. 22 No. 2, pp. 72-80.

Carvalho, D.V., Pereira, E.M. and Cardoso, J.S. (2019), "Machine learning interpretability: a survey on methods and metrics", *Electronics*, Vol. 8, p. 832.

Chougule, S., Kozonek, N., Ismail, A., Adam, G., Narayan, V. and Schulze, M. (2018), "Reliable multilane detection and classification by utilizing cnn as a regression network", *Proceedings of the european conference on computer vision (ECCV) workshops*.

Davis, S.E., Cremaschi, S. and Eden, M.R. (2017), "Efficient surrogate model development: optimum model form based on input function characteristics", *Computer-aided Chemical Engineering*, Vol. 40, pp. 457-462.

Escalante, H.J., Escalera, S., Guyon, I., Bar, X., Güçlütür, Y., Güçlü, U. and van Gerven, M.A.J. (2018), *Explainable and Interpretable Models in Computer Vision and Machine Learning*, Springer Nature Switzerland AG.

Hamad, E.O., Savundranayagam, M.Y., Holmes, J.D., Kinsella, E.A. and Johnson, A.M. (2016), "Toward a mixed-methods research approach to content analysis in the digital age: the combined content-analysis model and its applications to health care twitter feeds", *Journal of Medical Internet Research*, Vol. 18 No. 3, doi: 10.2196/jmir.5391.e60.

Johnson, R.B. and Onwuegbuzie, A.J. (2004), "Mixed methods research: a research paradigm whose time has come", *Educational Researcher*, Vol. 33, pp. 14-26.

Law, S., Paige, B. and Russell, C. (2019), "Take a look around", *ACM Transactions on Intelligent Systems and Technology (TIST)*, Vol. 10, pp. 1-19.

Molnar, C., Casalicchio, G. and Bischl, B. (2020), "Interpretable Machine Learning – A Brief History, State-of-the-Art and Challenges", *ECML PKDD 2020 Workshops Communications in Computer and Information Science*, Vol. 1323. Springer, Cham, available at: https://doi.org/10.1007/978-3-030-65965-3_28

Shanaka, H. and Maier, G. (2010), "The hedonic price method in real estate and housing market research. A review of the literature", *SRE-Discussion Papers 2010/03*, WU Vienna University of Economics and Business.

Symonds, J. and Stephen, G. (2009), "The death of mixed methods: research labels and their casualties", *The British Educational Research Association Annual Conference*, Edinburgh, Heriot Watt University, pp. 3-6, September.

Corresponding author
Miroslav Despotovic can be contacted at: Miroslav.Despotovic@fh-kufstein.ac.at