Intelligent Systems with Applications 14 (2022) 200081

Contents lists available at ScienceDirect

Intelligent Systems with Applications

journal homepage: www.elsevier.com/locate/iswa

Vision-based housing price estimation using interior, exterior & satellite images

Ali Nouriani*, Lance Lemke

University of Minnesota, Twin Cities Minneapolis, MN, USA

ARTICLE INFO

Article history:
Received 10 January 2022
Revised 7 March 2022
Accepted 15 April 2022
Available online 29 April 2022

Keywords:
Computer Vision
Automatic price estimation
Housing price
Satellite view
CNN
Luxury level

ABSTRACT

Real estate price estimation has been an interesting subject in the literature from the appearance of online real estate services like Zillow and Redfin. These websites and many other works in the literature have proposed their methods for evaluation and pricing of the real estate. However, these methods fail to consider important information about the appearance and the neighborhood of the house which leads to occasional incorrect estimations. The novel proposed method in this paper tries to estimate housing price by considering attributes of the home as well as interior, exterior, and satellite visual features of the house. Deep convolutional neural networks on a large dataset of images of interior, exterior and satellite images of houses are trained to extract visual features of the houses. These features along with house attributes are fed to another system to automatically estimate the value of the house. Finally, the performance of the system is compared to Zestimate and some vision-based methods in the literature on a new dataset.

1. Introduction

Real estate price estimation is a challenging objective since it depends on many factors like number of rooms, area, age and location. Housing pricing is of practical importance in everyday real estate transactions and is effective in estimating housing market value. One important factor in pricing of real estate which is usually ignored by current estimators is the aesthetics and visual appearance of the house and the neighborhood. The traditional method of estimating real estate prices relies on statistical analysis of textual information of the houses which has been commercialized in online real estate database companies like Zillow, Trulia and Redfin. Although some research works tried to fill this gap, most of them fail to consider both interior and exterior and the neighborhood appeal in their estimates. In this project, we propose a novel method which considers both textual attributes of the house and interior, exterior, and satellite visual features of the real estate.

The large online database of real state has led to many studies on automatic housing pricing in recent years, unfortunately most of these works estimate the apartment prices based on statistical analysis of text information of the houses, hence in this paper, we propose a novel method to include both text information and interior, exterior, and satellite visual features of the house. A room classifier is trained on the data to automatically label room types. Another network is trained to classify the images to different levels based on their beauty and aesthetics to different luxury levels. Finally a regressor trained to predict the price based on the text data along with the luxury levels extracted from images. We have compared different classification methods and regression models to get the best result. Fig. 1 shows an example of the collected data.

Many articles have been published providing information on how machine learning can help improve the accuracy of automated real estate pricing methods. Some methods use a combination of certain images and metadata to estimate housing prices, while other methods use only text data. Image, outdoor image, satellite image or a combination of two types of images to evaluate the value of the house. The most commonly used text data are square feet, age, and number of bedrooms (Naser, Serte, & Turjman, 2020). Some methods only check the exterior image of the house. These methods use the extracted features of AlexNet, GoogLeNet or VGG16/19 (Bessinger & Jacobs, 2016). These attributes may have a positive or negative impact on the house prices. For example, an internal courtyard is a positive features and an industrial scene is a negative environment. In Bessinger & Jacobs (2016), Bessinger reported an RMSE error of 28281$ and median error of 10% but in

* Corresponding author.
E-mail addresses: nouri011@umn.edu (A. Nouriani), lemke203@umn.edu (L. Lemke).

https://doi.org/10.1016/j.iswa.2022.200081

Fig. 1. Examples of internal, external, satellite and text attributes of a house.

comparison with text data-only metods showed 4% improvement. Some works could identify certain features in the picture, such as window bars, hedges, gable roofs and tropical plants, to predict the characteristics of the city and their relation with the price of the house (Arietta & Sean, 2014). Another document uses street images to predict the perceived characteristics of places in the city (Ordonez & Berg, 2014).Some of these articles indicate that these models are powerful and flexible enough to be used in different cities (You, Pang, Cao, & Luo, 2017). In Peng et al. (2021) and Shin, Kim, & Hong (2021), studied and compared LASSO, Deep Walk and LSTM-based methods for the prediction real-estate. They reported average errors as 24.83%, 23.28% and 9.94% respectively. Other methods examine internal images and use image classification to determine the degree of luxury (Poursaeed, Matera, & Belongie, 2017). This method uses luxury ratings from multiple sources to rate images of living spaces with different levels of luxury (Ahmed & Moustafa, 2016).

There are papers that solely use satellite imagery to extract the details of a neighborhood. One method enables CNN to extract spatial features in order to study the impact of regional information on real estate prices (Bency, Rallapalli, Ganti, Srivatsa, & Manjunath, 2017). Another method is to use satellite images and street view images to characterize the urban quality of the area (Law, Paige, & Russell, 2019). They found that correct characterization of regional quality can improve the prediction of house prices. Wang, Chen, Su, Wang, & Huang (2021) proposed a novel end-to-end self-joint model for housing price estimation. They tried to include information about the neighborhodd amenities including the parks, schools and transit using satellite images. They reported minimum error of 12.11% as their best result (Wang et al., 2021).

Some methods only consider the textual information of the house. These methods use text data as well as statistical and machine learning methods to estimate the value of a house (Afonso, Melo, Oliveira, Sousa, & Berton, 2019; Benjamin, 2004; Khamis & Kamarudin, 2014; Li, 2009; Li, Huang, & Li., 2019; Rafiei & Adeli, 2016; Shim & Hwang, 2018; Truong, Nguyen, Dang, & Mei, 2020; Wang, Zou, Zhang, & Shi, 2019; Zhao, Chetty, & Tran, 2019). The features that text methods use include square footage, zip code, number of rooms, and market data to estimate prices. The text-based method is the most popular method used by real estate agents and websites to estimate the value of a property. Both Zillow (Redfin, 2022) and Redfin (Zestimate, 2022) use this method.

There are multiple base works for housing price estimation some of which use textual methods and some use vision-based methods. The textual methods have progressed so that commercialized in Zillow's "Zestimate" (Redfin, 2022) and Redfin's "Redfin estimate" (Zestimate, 2022). Both the Zestimate and Redfin estimates utilize statistical and machine learning models using textual data to make predictions of home prices. Zestimate and Redfin Estimate have a national median error percentage of 7.7% (Redfin, 2022) and 5.92% (Zestimate, 2022) respectively. On the other hand, there are also vision-based research works like (Bessinger & Jacobs, 2016) that studies only exterior images of the apartments which extracts features using AlexNet, GoogLeNet, and VGG16/19. Another paper uses four pictures from each house passed through a neural network to estimate the price of the home (Poursaeed et al., 2017). This method was shown to realize a median error percentage of 5.6% (Poursaeed et al., 2017).

The limitation of these baseline works is that they do not include all aspects of the house, such as: Text information, street view, interior view and satellite images. Although the text-only methods work well but various valuable features of the house are still missing in these methods. Certain aspects will affect the value of the house that are not reflected in the text information such as style, architecture, neighborhood aesthetics, etc. None of the baseline methods include both external and the impact of internal appeals on the price. Therefore the advantage of our method is to include all the important factors in evaluation of a house. In addition, We have shown that our method increases the accuracy and outperforms the baseline methods. However, we should mention the limitation of this method is the need for more information of the house like satellite and interior views which might not always be accessible for evaluation.

2. Data collection and annotation

We have collected a suitable data set of about 25 thousand images of different rooms of houses with an almost normally distributed attributes such as price, number of rooms, lot size, and age. The images are collected from different websites like Zillow, Redfin, and Realtor. There are few data-sets available especially that contain house images from all rooms and satellite view. The house satellite images are collected from bing maps for each house using the address of the house. The collected dataset includes only one story houses and apartments are not discussed in this paper.

A reasonable method for evaluation of the luxury levels of the rooms is to compare each room category (like bedroom, bathroom, etc.) individually, i.e. better results could be generated by comparing room images from the same type instead of comparing rooms from different types. Therefor we have trained a deep network to classify the images to 7 room categories of bathroom, bedroom, kitchen, living room, dining room, front image, and satellite image. The room classifier is trained on labeled images collected from LSUN (Yu et al., 2015) and google images search and Places datasets (Zhou, Lapedriza, Xiao, Torralba, & Oliva, 2014) with more than 1 Million images. LSUN benchmark includes 1,315,802 images of living rooms, 3,033,042 images of bedrooms, 657,571 images of dining rooms and 2,212,277 images of kitchens and no images of exterior, bathroom and satellite views. Places datset has about 28,000 images of each category of living room, bedroom, dining room, kitchen, bathroom, exterior and no images of satellite view. While we used the aforementioned benchmarks for training our room classifies, in comparison, we have collected 4 thousands of images for each category including satellite views. Fig. 2 shows an example of correctly classified room types.

Each room category was labeled to different luxury levels using crowd-source. We have randomly shown each participant a set of images for each room type along with the samples as the benchmark for luxury levels and asked crowd workers to select images with similar luxury level compared to the samples. We have

Fig. 2. An example of correctly classified room types.

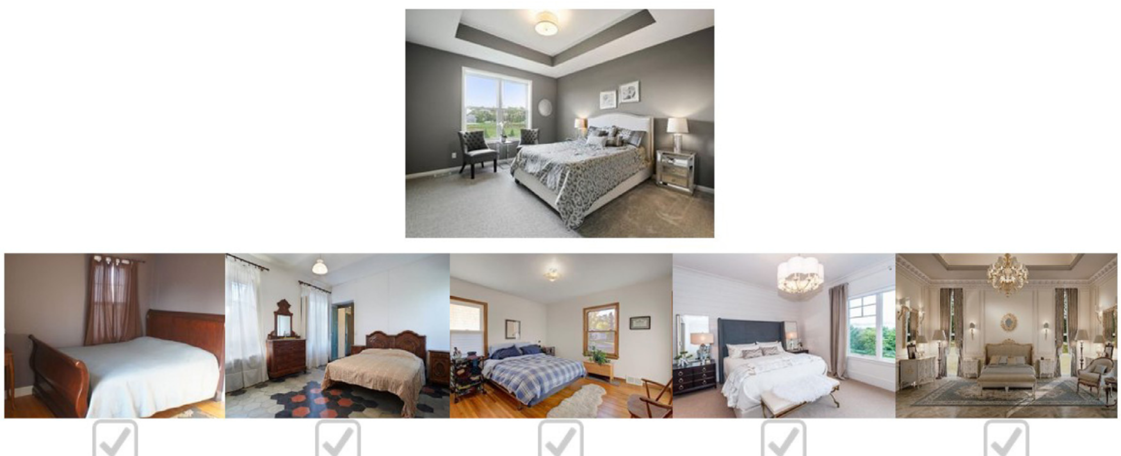

Fig. 3. Example of the crowd-source luxury level interface.

collected the labels through MTurk (Amazon, 2022). An example of the labeled luxury levels are shown in Fig. 3.

3. Model

The implemented procedure is shown in Fig. 4 as follows:

1. Classify images to room categories (bathroom, bedroom, kitchen, living room, dining room, front, satellite images)
2. Classify each category of images to luxury levels
3. Take "Mode" of luxury levels of all room types as level feature
4. Combine with text data
5. Input to regression model
6. Compare the performance to with the literature methods.

3.1. Image classification by room type

We trained VGG16, DenseNet and GoogLeNet for classification of the images to the aforementioned 7 room types. The accuracy of the methods were 93%, 92.3%, and 91.7% respectively. Hence we used VGG16 as our room classifier. The classifier is then applied to the large dataset of real estate collected from online resources such as Zillow and Redfin. An example of the classified room types are shown in Fig. 2.

A fine-tuned VGG16 network used on the collected data to classify room categories of bathroom, bedroom, kitchen, living room, dining room, external front, and satellite image. We have tuned the last three layers of VGG network and added a flatten layer, two dense layers and a normalization layer to increase the accuracy and reliability of the network.

3.2. Image classification by luxury level

We have classified each category of images to different luxury levels from the low to high adapted from Poursaeed et al. (2017). This paper used a crowd-source method for labeling images to different luxury levels. Since the scope of luxury levels, the number of levels and the benchmark for luxury levels are not clear, a crowd-source system is created to ask people which images they think are in the same level. The luxury level is related to aesthetics of the house, is subjective and could not be directly measured by the price. Therefore we used the t-STE algorithm (Maaten & Weinberger, 2012) to compute a two-dimensional embedding of the images to form clusters of similar images. Using this method we can determine the required clusters that adequately represent the luxury levels. We have concluded with some trial and errors on our dataset, using elbow method, that 5 luxury levels is enough to accurately divide the dataset. Then randomly selected images of each cluster could represent the corresponding luxury level. We have used randomly selected 80% of the dataset from Poursaeed et al. (2017) as the benchmark and asked crowd workers to select images with similar luxury levels compared to the benchmarks.

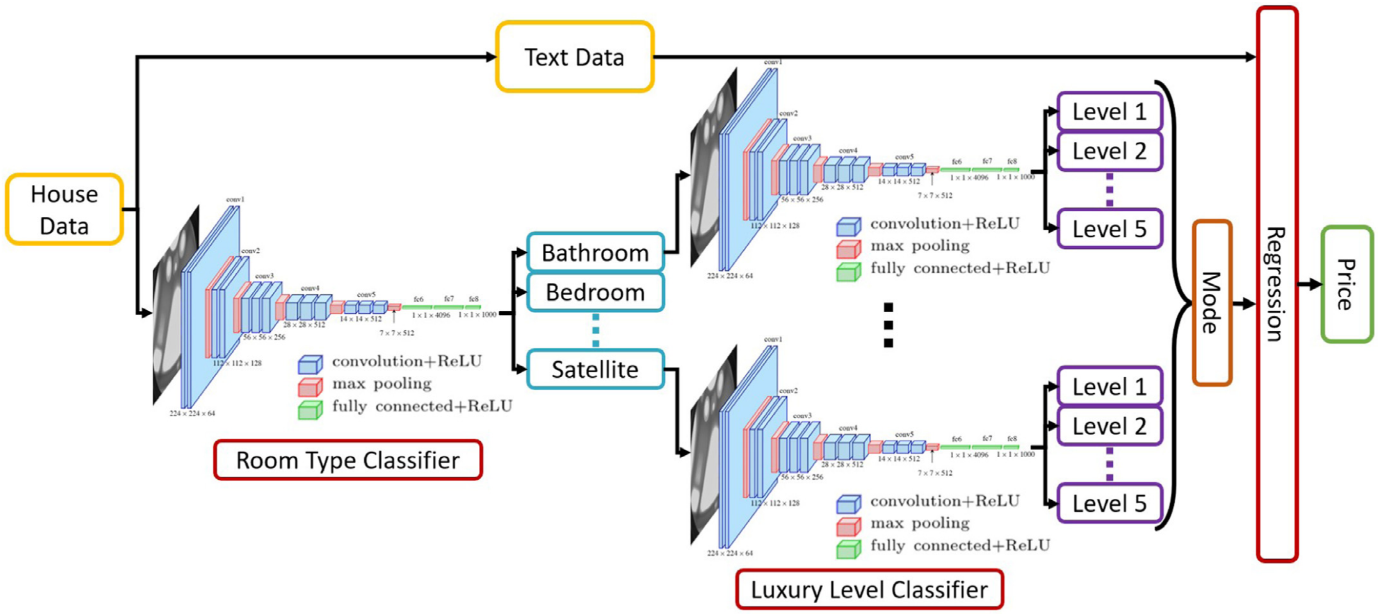

Fig. 4. Schematic of network. The photos are classified to their room category, each room type is classified to their luxury level. The mode of luxury levels is then concatenated with the normalized metadata and passes through a regression model to estimate the price.

Then to automate the process of luxury level classification we offered and compared 3 networks:

1. GoogLeNet plus KNN model
2. GoogLeNet plus SVM model
3. Fine-tuned and modified VGG network as a CNN model

We train the model for each category of rooms and then took the mode of each category as the level feature.

3.2.1. Luxury level mode

The estimated luxury levels using the aforementioned methods are for each room type of bedroom, bathroom, kitchen, etc. We have found it helpful to take "mode" of luxury levels from different room types as the general luxury level of the house in terms of matching luxury levels with price levels accuracy. The mode is calculated simply using the following equation Nouriani et al., 2021:

$$\text{luxury level} = \text{mode}\{P_i \mid i \in [1, 7]\}$$

In which, P_i shows the predicted luxury level for room type i which are assigned alphabetically as bathroom $= 1$, bedroom $= 2$, dining room $= 3$, front view $= 4$, kitchen $= 5$, living room $= 6$, satellite view $= 7$. For example the system classifies the luxury level from each room type of bathroom, bedroom, kitchen, ... as $\{1, 1, 2, 1, 3, 1, 1\}$ respectively, we assign the mode from this set as the predicted class (luxury level) hence the level is predicted as 1. This procedure is repeated with each of the following methods to get the best result.

For example, using SVM method, the confusion matrices for luxury level classification of each room type and the confusion matrix of the mode of the levels is shown in Fig. 5. The mode shows an average of 22% jump in the accuracy of luxury level classification. We have seen improved accuracy for the other two methods of KNN, and CNN similarly.

3.2.2. GoogLeNet plus KNN

We used a GooLeNet since it has been trained on a large dataset of visual data and extracted features have meaning in terms of explaining the visual data. For example, we have found that similarity between a house pictures and a church reduces the value of the house. Or detection of a scene as a yard increases the luxury level of the house (Bessinger & Jacobs, 2016). Therefore in this model we have combined a GoogLeNet with other models to classify luxury levels. In the first method for luxury classification we used a combination of the GoogLeNet and a k-nearest neighbors (KNN) model. The GoogLeNet details could be found here (Szegedy et al., 2015). The KNN model is fitted on the features extracted by the last layer of the GoogLeNet. The optimal value of K is selected using a grid search with the target of best accuracy as $K = 5$. Fig. 6(a) shows the confusion matrix of the first method with the overall accuracy of 59%. This method preforms worse in the higher luxury levels.

3.2.3. GoogLeNet plus SVM

In the second method for luxury classification similarly we used a combination of the GoogLeNet and a Support Vector Machine (SVM). The SVM model is trained on the features extracted by the last layer of the GoogLeNet. The SVM model uses Radial Basis Function (RBF) kernel with balanced class weight and C = 1e3, and gamma = 0.1. The optimal value of hyper parameters are selected through a grid search. Fig. 6(b) shows the confusion matrix of this method in luxury level classification. The overall accuracy is improved to about 74%, moreover shows a more balanced result across different levels.

3.2.4. CNN model

In the third method, instead of combining the GoogLeNet extracted features with another method, we directly fine-tuned a VGG16 network as a Convolutional Neural Network (CNN) model. With some try and errors, we could get the best results with adding two dense layers and a normalization layer to VGG16. The details of VGG16 network are brought here (Zhang, Zou, He, & Sun, 2015). As shown in Fig. 6(c), the accuracy is about 89.6% and the confusion matrix of CNN method shows better overall result compared to the previous methods.

3.3. Regression model

The extracted features including luxury levels and meta data of the houses are concatenated to form as the features of each data point. Then We applied and compared the performance of various

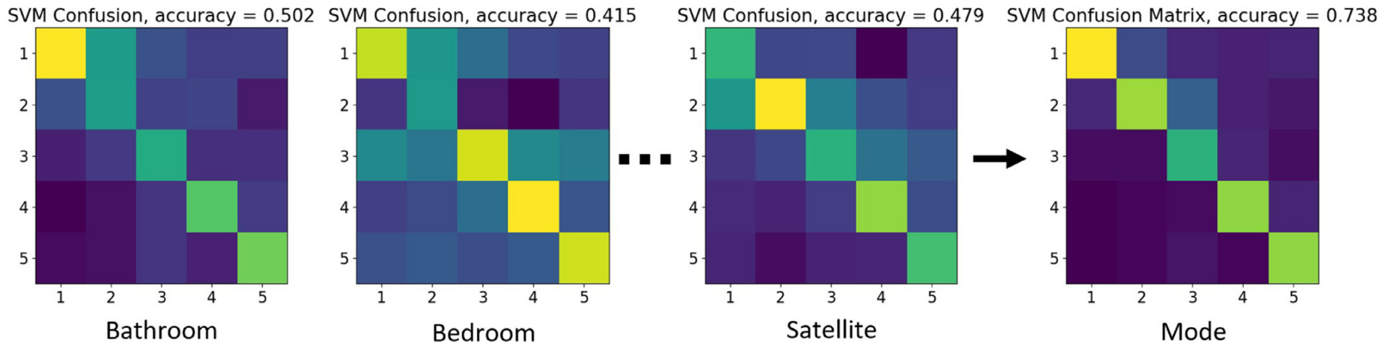

Fig. 5. Taking mode of the luxury levels of image categories (room types + satellite) for SVM method.

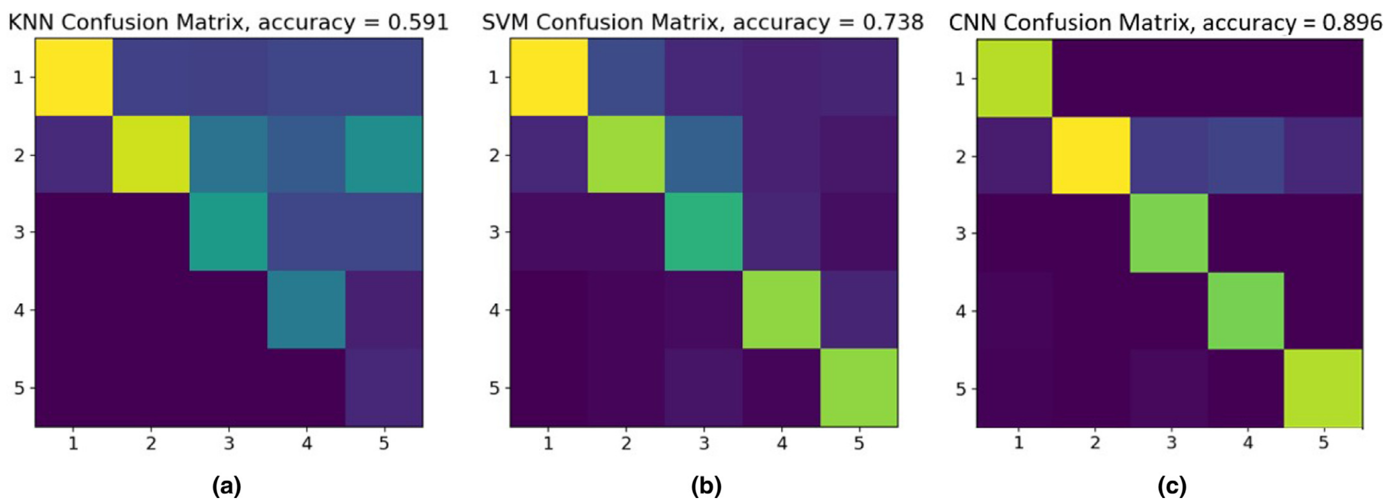

Fig. 6. Confusion matrix and accuracy luxury level classification methods.

Table 1
Cross-validation score of different regression methods.

Regression method	Mean squared error	Standard deviation
Linear Regression	0.0214	0.0132
LASSO	0.2562	0.0485
Elastic Net	0.2560	0.0481
K-Nearest Neighbors	0.0355	0.0222
Decision Tree	0.0228	0.0154
Gradient Boosting Machine	0.0175	0.0127

regression models like Linear, Lasso, Elastic net, KNN, Decision tree and Gradient boosting to get the best result. Table 1 shows the cross-validation score of these methods resulted from training on the standardized 80% randomly-selected training data.

The Gradient boosting has shown the best performance compared to the other methods, so the data is trained using Gradient boosting method. The n-estimator is then chosen by a grid search. The correlation matrix of the features of the data is shown in Fig. 7. The features with higher correlation with the house price are then picked as the most meaningful features for the price estimation and training the regression model. As shown in Fig. 7, some features have a positive effect on the price of the house, for instance the square footage of the house, the number of rooms and the luxury level have a direction relation with the price of the house in all models. This result is reasonable and intuitive since higher luxury level should increases the overall value of the house. On the other hand, some features like the age has a negative effect on the price as expected. The effect of luxury level in relation with location in the city and other features is out of scope of this paper and should be discussed in the future studies.

4. Results

The performance of regression models are compared for these scenarios:

1. Text only data
2. Text plus luxury level from KNN model
3. Text plus luxury level from SVM model
4. Text plus luxury level from CNN model

A summary of the median error of these methods is shown in Fig. 8. The CNN model outperforms the other methods. This result is expected since it gives a better correlation with the price tag compared to the other methods as shown in Fig. 7. Therefore we can see the superior performance of CNN compared to the SVM. Likewise SVM performs better than KNN and KNN is better than only text data due to the better correlation of luxury level. This result shows the importance and effect of visual information in housing price estimation and certifies the added details.

As qualitative and validation examples, we have shown some of our randomly selected data instances and compared our estimation with the results of Zillow as seen in Table 2. Fig. 9 shows sample pictures of the homes with their predicted prices and ground truth prices. This examples show superior accuracy of our method compared to Zillow which again certifies the importance of visual data in price estimation. For the sake of comparison to baselines, we have calculated the median error rate of the

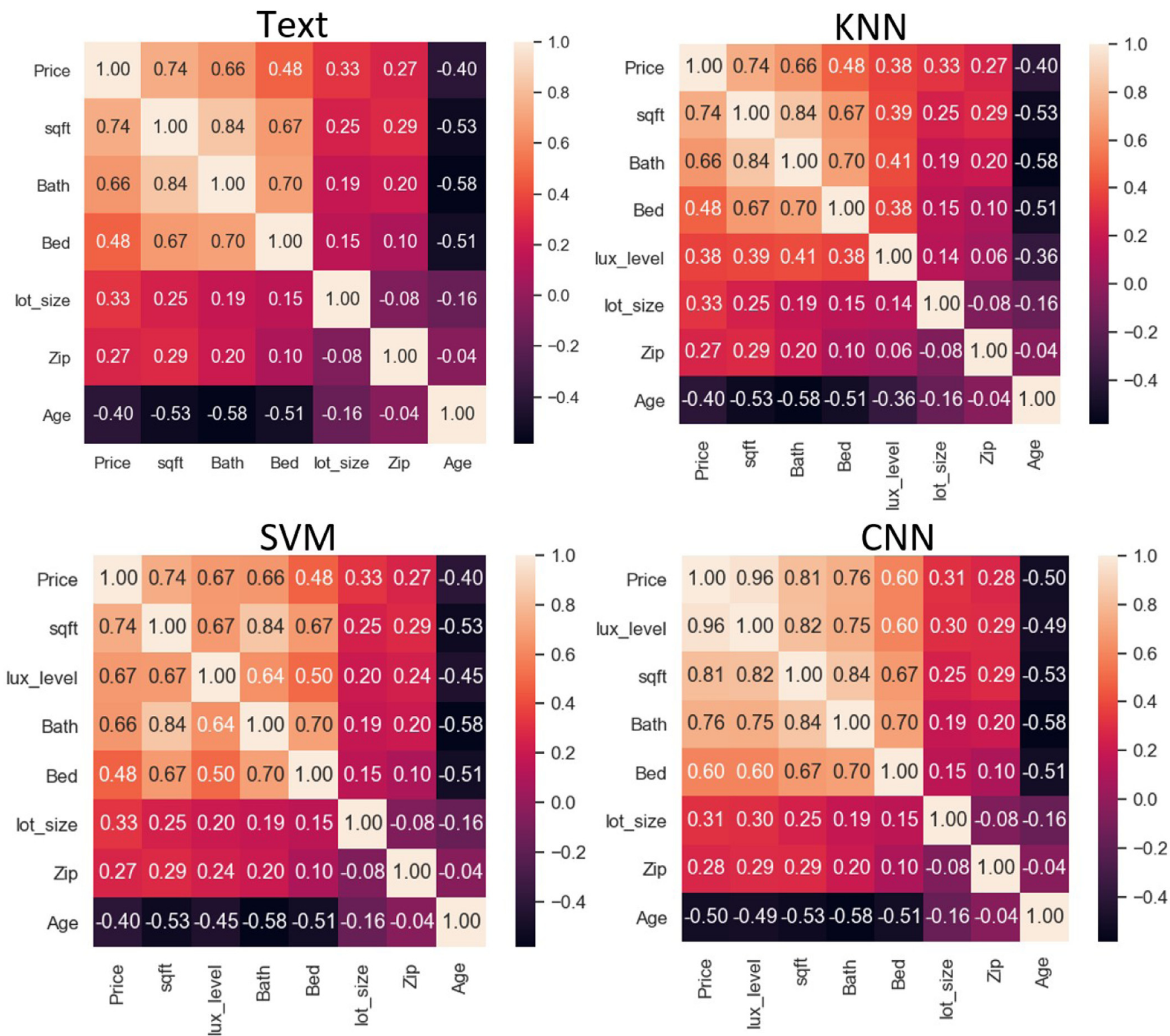

Fig. 7. Correlation of between features of the data for different lux level classifiers.

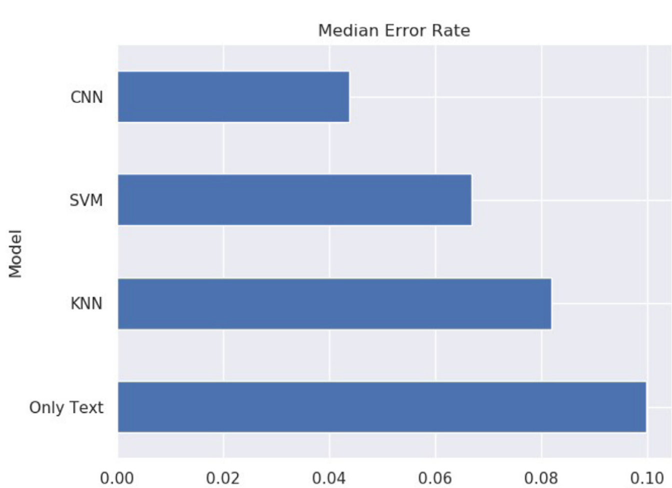

Fig. 8. Median error rate of different models.

baseline paper and Zestimate on part of our data with 1800 instances which was collected from Zillow. Our method using CNN model shows a median error rate of 4.98% on this dataset while the baseline paper's method by Poursaeed et al. using luxury levels (Poursaeed et al., 2017) we could get 5.6% and with Zestimate we have got median error of better than 7.3%. This clearly shows better performance of our method compared to the baseline estimates. Also we have tested LSTM method by Shin et al. (2021) on our dataset. We could get the median error of 10.94%. In addition, we used Bessinger's dataset in Bessinger & Jacobs (2016) which reported an RMSE error of 28281$ while we could get an RMSE error of 24137$. You should note that this dataset only included exterior images and we did not have access to satellite or interior views of these houses, hence we might have got even better results if we had all views.

The reason for the better accuracy of our method lies in two facts. First, our proposed method incorporates all visual features of a house including interior design, exterior view and neighborhood information from satellite image when the data is accessible. Second, our method incorporates the images using luxury levels

Fig. 9. Sample images of houses with their address, value predicted by our model, and the ground truth price.

Table 2
Compare result on randomly selected data instances.

Address	Ground truth	Zestimate predicted price	Our error %	Zestimate error %	
9421 Larges Ct	925,000	852,762	898354.1	2.88	7.81
129 E 68th St	318,500	342,207	335132.6	5.22	7.44
4372 Lexington	390,000	355,133	359557.1	7.80	8.94
8106 N 127th Ave	600,000	641,132	563456.7	6.09	6.85
9120 Grey Cloud	447,500	481,821	414,141	7.45	7.67
1528 S 110th St	565,000	600,983	575750.9	1.90	6.37

instead of direct visual features in the regression model, as for example Bessinger did in Bessinger & Jacobs (2016). This reduces the dimension of the data for the regression model and hence is superior in terms of both computation cost and accuracy.

5. Conclusion

Housing price estimation is a well-known and studied subject in the literature and also the industry. Most of these researches emphasize on textual attributes of the houses. We have shown that the visual aspects of a house could help us to have a more accurate estimate of the price. In order to incorporate visual data of the houses, we proposed luxury level classification of the house images for miscellaneous rooms, exterior view and even satellite view of the houses for the first time.

The existing visual-based housing price estimation methods usually limited to exterior view or only one of these various room categories. We have tried to get the best result by including all of these external and internal images in our network and classified each image category of external, rooms, satellite, etc. images to different luxury levels. Then to decide the final luxury level of the house, we found the statistical mode to be accurate and reliable in terms of price estimation.

For this project we collected a rare data collection of the exterior, interior and satellite images of houses around Minnesota along with several textual attributes such as address, sell date, price, square footage etc. and manually labeled some of the images. We trained a modified VGG16 network on part of our manually labeled data and used that network to classify house images into different rooms. Next, in order to classify images to different luxury levels, we have tested and compared several methods using KNN, SVM and CNN to find the best result in image level classifications by their final median error rate. Finally, these information along with the textual attributes are fed to different regression models to find the best regression model by their negative mean squared error.

In summary, our method using visual features from home images luxury level in conjunction with textual data to get an estimate of the price of the house has shown to be more accurate than the other baseline models such as Zillow's Zestimate. The strength of our method is in that it takes into account more visual information than other model which allows for more accurate price prediction. Since our model is trained on solely data from the Midwest, the next step would be training and validation on a bigger data set for other markets.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Ali Nouriani: Conceptualization, Methodology, Writing – original draft. **Lance Lemke:** Data curation, Validation, Writing – original draft.

References

Afonso, B., Melo, L., Oliveira, W., Sousa, S., & Berton, L. (2019). Housing prices prediction with a deep learning and random forest ensemble. In *Anais do XVI encontro nacional de inteligência artificial e computacional* (pp. 389–400).

Ahmed, E., & Moustafa, M. (2016). House price estimation from visual and textual features. ArXiv preprint arXiv:1609.08399.

Amazon Mechanical Turk https://www.mturk.com.

Arietta, S. M., et al., (2014). City forensics: Using visual elements to predict non-visual city attributes. *IEEE Transactions on Visualization and Computer Graphics, 20*(12), 2624–2633.

Bency, A. J., Rallapalli, S., Ganti, R. K., Srivatsa, M., & Manjunath, B. S. (2017). Beyond spatial auto-regressive models: Predicting housing prices with satellite imagery. In *2017 IEEE winter conference on applications of computer vision (WACV)* (pp. 320–329).

Benjamin, J. D., et al., (2004). Mass appraisal. *Journal of Real Estate Practice and Education, 7*(1), 65–78.

Bessinger, Z., & Jacobs, N. (2016). Quantifying curb appeal. In *2016 IEEE international conference on image processing (ICIP)* (pp. 4388–4392).

Khamis, A. B., & Kamarudin, N. K. K. B. (2014). Comparative study on estimate house price using statistical and neural network model. *International Journal of Scientific & Technology Research, 3*(12), 126–131.

Law, S., Paige, B., & Russell, C. (2019). Take a look around. In *ACM transactions on intelligent systems and technology (TIST): vol. 10* (pp. 1–19).

Li, Da-Ying, Wei Xu, Hong Zhao, and Rong-Qiu Chen. “A SVR based forecasting approach for real estate price prediction.” In 2009 International conference on machine learning and cybernetics, vol. 2, pp. 970–974. IEEE, 2009.

Li, H., Huang, X., & Li, X. (2019). Urban land price assessment based on GIS and deep learning. In *IGARSS 2019 - 2019 IEEE international geoscience and remote sensing symposium, Yokohama, Japan* (pp. 935–938).

Maaten, L. V. D., & Weinberger, K. (2012). Stochastic triplet embedding. In *Machine learning for signal processing (MLSP), 2012 IEEE international workshop on* (pp. 1–6).

Naser, N., Serte, S., & Al-Turjman, F. (2020). From traditional house price appraisal to computer vision-based: A survey. In *International conference on forthcoming networks and sustainability in the IoT era* (pp. 1–10). Cham: Springer.

Nouriani, A., McGovern, R. A., & Rajamani, R. (2020). Step Length Estimation Using Inertial Measurements Units. In *In 2021 American Control Conference (ACC)* (pp. 666–671). Cham: Springer. IEEE.

Ordonez, V., & Berg, T. L. (2014). Learning high-level judgments of urban perception. In *European conference on computer vision: vol. 8694* (pp. 494–510). Springer.

Peng, H., Li, J., Wang, Z., Yang, R., Liu, M., Zhang, M., et al., (2021). Lifelong property price prediction: A case study for the toronto real estate market. *IEEE Transactions on Knowledge and Data Engineering*, 1–16.

Poursaeed, O., Matera, T., & Belongie, S. (2017). Vision-based real estate price estimation. *Machine Vision and Applications*. https://doi.org/10.1007/s00138-018-0922-2.

Rafiei, M. H., & Adeli, H. (2016). A novel machine learning model for estimation of sale prices of real estate units. *Journal of Construction Engineering and Management, 142*(2), 04015066.

Redfin Estimate, https://www.redfin.com/redfin-estimate.

Shim, J., & Hwang, C. (2018). Kernel-based geographically and temporally weighted autoregressive model for house price estimation. *PLoS One, 13*(10), E0205063.

Shin, E. K., Kim, E. M., & Hong, T. H. (2021). Real estate price forecasting by exploiting the regional analysis based on SOM and LSTM. *The Journal of Information Systems, 30*(2), 147–163.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., et al., (2015). Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1–9).

Truong, Q., Nguyen, M., Dang, H., & Mei, B. (2020). Housing price prediction via improved machine learning techniques. *Procedia Computer Science, 174*, 433–442.

Wang, F., Zou, Y., Zhang, H., & Shi, H. (2019). House price prediction approach based on deep learning and ARIMA model. In *IEEE 7th international conference on computer science and network technology (ICCSNT), Dalian, China* (pp. 303–307).

Wang, P.-Y., Chen, C.-T., Su, J.-W., Wang, T.-Y., & Huang, S.-H. (2021). Deep learning model for house price prediction using heterogeneous data analysis along with joint self-attention mechanism. *IEEE Access, 9*, 55244–55259.

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

Yu, F., Seff, A., Zhang, Y., Song, S., Funkhouser, T., & Xiao, J. (2015). Lsun: Construction of a large-scale image dataset using deep learning with humans in the loop. ArXiv preprint arXiv:1506.03365.

Zhang, X., Zou, J., He, K., & Sun, J. (2015). Accelerating very deep convolutional networks for classification and detection. In *IEEE transactions on pattern analysis and machine intelligence: vol. 38* (pp. 1943–1955).

Zhao, Y., Chetty, G., & Tran, D. (2019). Deep learning with XGBoost for real estate appraisal. In 2019 IEEE symposium series on computational intelligence (SSCI), Xiamen, China.

Zhou, B., Lapedriza, A., Xiao, J., Torralba, A., & Oliva, A. (2014). Learning deep features for scene recognition using places database. In *Advances in neural information processing systems* (pp. 487–495).

Zestimate https://www.zillow.com/zestimate.