

# Enhancing Real Estate Listings Through Image Classification and Enhancement: A Comparative Study <sup>†</sup>

Eyüp Tolunay Küp <sup>1,2,\*</sup>, Melih Sözdinler <sup>1,3</sup> , Ali Hakan Işık <sup>4</sup> , Yalçın Doksanbir <sup>1</sup> and Gökhan Akpınar <sup>1</sup>

<sup>1</sup> Emlakjet, R&D Center, R&D, 34764 İstanbul, Turkey; melih.sozdinler@emlakjet.com (M.S.); yalcin.doksanbir@emlakjet.com (Y.D.); gokhan.akpinar@emlakjet.com (G.A.)

<sup>2</sup> Industrial Engineering, Faculty of Engineering and Natural Sciences, Kadir Has University, 34083 İstanbul, Turkey

<sup>3</sup> Computer Engineering, Faculty of Engineering and Natural Sciences, Işık University, 34980 İstanbul, Turkey

<sup>4</sup> Computer Engineering, Faculty of Engineering and Architecture, Burdur Mehmet Akif Ersoy, 15030 Burdur, Turkey; ahakan@mehmetakif.edu.tr

\* Correspondence: tolunay.kup@emlakjet.com

<sup>†</sup> Presented at the 2024 IEEE 6th Eurasia Conference on IoT, Communication and Engineering, Yunlin, Taiwan, 15–17 November 2024.

**Abstract:** We extended real estate property listings on the online prop-tech platform. On the platform, the images were classified into the specified classes according to quality criteria. The necessary interventions were made by measuring the platform's appropriateness level and increasing the advertisements' visual appeal. A dataset of 3000 labeled images was utilized to compare different image classification models, including convolutional neural networks (CNNs), VGG16, residual networks (ResNets), and the LLaVA large language model (LLM). Each model's performance and benchmark results were measured to identify the most effective method. In addition, the classification pipeline was expanded using image enhancement with contrastive unsupervised representation learning (CURL). This method assessed the impact of improved image quality on classification accuracy and the overall attractiveness of property listings. For each classification model, the performance was evaluated in binary conditions, with and without the application of CURL. The results showed that applying image enhancement with CURL enhances image quality and improves classification performance, particularly in models such as CNN and ResNet. The study results enable a better visual representation of real estate properties, resulting in higher-quality and engaging user listings. They also underscore the importance of combining advanced image processing techniques with classification models to optimize image presentation and categorization in the real estate industry. The extended platform offers information on the role of machine learning models and image enhancement methods in technology for the real estate industry. Also, an alternative solution that can be integrated into intelligent listing systems is proposed in this study to improve user experience and information accuracy. The platform proves that artificial intelligence and machine learning can be integrated for cloud-distributed services, paving the way for future innovations in the real estate sector and intelligent marketplace platforms.



Academic Editors: Teen-Hang Meen,  
Chi-Ting Ho and Cheng-Fu Yang

Published: 22 May 2025

**Citation:** Küp, E.T.; Sözdinler, M.; Işık, A.H.; Doksanbir, Y.; Akpınar, G. Enhancing Real Estate Listings Through Image Classification and Enhancement: A Comparative Study. *Eng. Proc.* **2025**, *92*, 80. <https://doi.org/10.3390/engproc2025092080>

**Copyright:** © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** image classification; image enhancement; prop-tech; real estate; room classification; convolutional neural networks; vgg16; Resnet; large language models; contrastive unsupervised representation learning

## 1. Introduction

Online real-estate platforms provide important information for those planning to buy or rent real estate and affect customer behaviors and processes in property searches and purchases. As of 2023, the Turkish real estate industry has grown to USD 46 billion [1]. With digitalization and online real-estate platforms, real estate sales increase up to 15% annually. For instance, one-third of Turkey's real estate sale transactions were performed online [2].

Properties are advertised on digital platforms a hundred times more often than they are advertised offline and such advertisements encompass the entire property market. Visual elements in digital real estate portfolios, such as photos and virtual tours, are paramount for attracting customers. On these platforms, the photographs viewed by potential buyers or tenants often determine whether a property is worth buying, highlighting how important the photo quality is. On the other hand, the quality and classification of images are essential for attracting customers' attention and are critical in reducing editing time and increasing listing review time. Well-organized and presented images enhance users' experience and make the search for properties efficient. The more accurately these images appear, the more effectively and efficiently users utilize them. Therefore, estate agencies need to optimize images in their listings.

We examined image classification and enhancement methods to improve property listings on an online real estate platform. Using a dataset consisting of 3000 labeled images, different image classification models such as convolutional neural networks (CNNs), VGG16, residual networks (ResNets), and the LLaVA large language model (LLM) were compared to evaluate the performance of each model and select the most effective one for real estate image classification. In addition, the image enhancement process was integrated using the contrastive unsupervised representation learning (CURL) method prior to image classification. This method evaluates the effect of improved image quality on classification accuracy and overall attractiveness of property listings. The CURL method makes low-quality real estate images explicit and detailed. The performance of each classification model was examined separately for cases where CURL was applied and not applied.

Image enhancement with CURL increased image quality and improved classification performance, especially when models such as CNNs and ResNets were involved. The accuracy of the CNN model using CURL increased by 1.2% compared with that without CURL. This result supports attractive and informative listings for online users by providing an accurate and outstanding representation of real estate properties. The results of this study highlight how advanced image processing techniques are integrated with classification models and how image presentation and categorization are optimized in the online real estate industry.

## 2. Literature Review

Digitalization has significantly changed the real estate industry and has created a competitive environment in the digital property marketplace. Many real estate agents are using innovative solutions such as predictive machine learning applications, large language models, and generative artificial intelligence to survive in this competitive environment. Scientific studies in this field have also been conducted to meet the practical needs of the sector.

Yuan et al. compared different approaches using regression and classification models for property price prediction [3]. They determined the most effective method by evaluating different machine learning techniques. The result proved the usability and the performance of machine learning models for the real estate industry in forecasting property prices. Lemeš and Akagic predicted the cost of properties and assessed the performances of the algorithms. The prediction accuracy was increased by centering on feature engineering

and model optimization [4]. Law et al. built prediction models for housing prices using deep learning models, street and satellite images, and economic factors [5]. According to the previous results, exterior and top-view real estate images are used to effectively evaluate properties.

Semnani and Rezaei studied real estate price prediction using satellite images and deep learning techniques [6]. Liu et al. applied deep learning convolutional neural networks to automatically extract building structures from high-resolution satellite images and enhanced the accuracy of land and urban mapping solutions [7]. Zhu et al. explored how deep learning algorithms classify house images and their consequences in real estate businesses [8] and improve the effectiveness of real estate management systems using real estate images. The results proved that image processing improvements positively affect estate valuation, marketing, and user experience processes in digital real estate marketplaces—particularly with regard to satellite images.

Bappy et al. used convolutional neural networks (CNNs) to classify the images of real estate properties based on the different rooms available [9]. This makes it easier for real estate agents or companies to organize property listings and significantly facilitates property searching. Råhlén and Sjöqvist examined image classification methods for assessing fundamental properties [10]. In this study, they examined the characteristics of images of real estate properties using several machine learning methodologies. The visual geometry group (VGG) plays a vital role in image classification. He et al. built complex neural networks for image recognition using the proposed ResNet, resulting in better performance [11,12]. Simonyan and Zisserman designed the VGG network with deeper networks and smaller convolutional filters for better classification results [13].

With the acceleration in improving multimodal models, the contrastive language-image pretraining (CLIP) model is one of the most recent significant models. The model demonstrated the ability to learn image concepts through natural language [14]. As a result of that, CLIP shows an improved capability of classifying images in a zero-shot fashion. Similarly, Jia et al. developed another model named ALIGN for visual image and text representation learning in a broader sense and added content-based noisy text supervision to state the further developments in the area [15].

For image enhancement, Moran et al. used client URL (CURL) with neural curve layers to enable global adjustment effectively [16]. In a similar context, Hu et al. put forth 'Exposure', an all-encompassing enhancement of photographs through neural networks but with explicit control over the enhancement parameters, the so-called white box approach [17]. Gharbi et al. developed a comprehensive image enhancement model that employs bilateral grid processing with neural networks for global modifications comparable to CURL [18]. Such models and methods allow features derived from images to significantly influence the valuation of properties. Also, various computer vision problems are solved to highly benefit from the performance and flexibility of neural networks.

### 3. Methodology

#### 3.1. Dataset

The dataset used in this research incorporated 3000 labeled images from an online real estate platform. These images included lands and residential and commercial properties in various room categories (e.g., bedrooms, living rooms, kitchens, bathrooms) and furnishing statuses (e.g., furnished, unfurnished, partially furnished). Quality control expertise in the company manually labeled the images to ensure that the classifications correctly reflected the room functionalities. The dataset was split into a training set (80%) and a testing set (20%) for the training and validation of the models. All images were resized to a constant size of  $224 \times 224$  pixels, confirming the input dimensions used for different models. Dataset

I contained 1000 images labeled with two labels, while Dataset II contained 3000 images labeled with four.

### 3.2. Image Classification Models

#### 3.2.1. CNNs

A baseline CNN architecture was used in this research as it is well suited for image classification and learns spatial patterns and the hierarchy of features through convolutional layers. This baseline model enables a fundamental comparison with complex architectures. CNNs have presented excellent results in image-based tasks at different levels of abstraction. The CNN architecture used in this research had 12 convolutional layers, and a total of 22 layers, including MaxPooling, FullyConnectedLayer, ReLULayer, DropOutLayer, and SoftmaxLayer layers. The image in the input layer is  $224 \times 224 \times 3$  in size, so it aligns with other models. It uses  $3 \times 3$  and  $6 \times 6$  filters and convolution-ReLU layers on top of each other in its architecture layers.

#### 3.2.2. VGG16

The VGG16 model, introduced by Simonyan and Zisserman, is known for its deep architecture, which consists of 16 primarily convolutional layers [13]. VGG16 is highly effective at extracting detailed features from images but at the cost of computational complexity. VGGNet architecture has two different types, with 16 and 19 layers: VGG16 and VGG19. The number of weight layers determines the overall number of layers. VGG16 architecture consists of 13 convolutions and three fully connected layers. There are 41 layers, including MaxPooling, FullyConnectedLayer, ReLULayer, DropOutLayer, and SoftmaxLayer layers. The image to be included in the input layer is  $224 \times 224 \times 3$ . VGGNet architecture uses  $3 \times 3$  filters in all layers and uses convolution-ReLU layers before the pooling layer. In the real estate images, VGG16 captures fine-grained details, such as furniture arrangements and room structures, which are crucial for property listings.

#### 3.2.3. ResNET

We adopted ResNet architecture for image classification to address the limitations of vanishing gradients in deep learning [11]. Using ResNet's skip connections, deeper networks were trained by maintaining learned information across layers. This model shows state-of-the-art effectiveness in large-scale image classification and recognition tasks in real estate image classification. ResNet has many variations, and ResNets18, ResNet50, and ResNet101 are frequently used. In this research, pretrained ResNet50 models were used. The ResNet152 model consisted of 152 main layers. We divided its architecture into four main parts: the convolutional layers, the identity block, the convolutional block, and the fully connected layers.

#### 3.2.4. LLaVA Large Language Model (LLM)

The LLaVA model was used to explore the benefits of integrating textual and visual understanding in classification tasks [19]. Unlike traditional models focusing on visual data, LLaVA leverages a multimodal approach by combining images with associated text, such as property descriptions or titles. This model enhanced classification performance by utilizing semantic boundaries and rules derived from user input in this research.

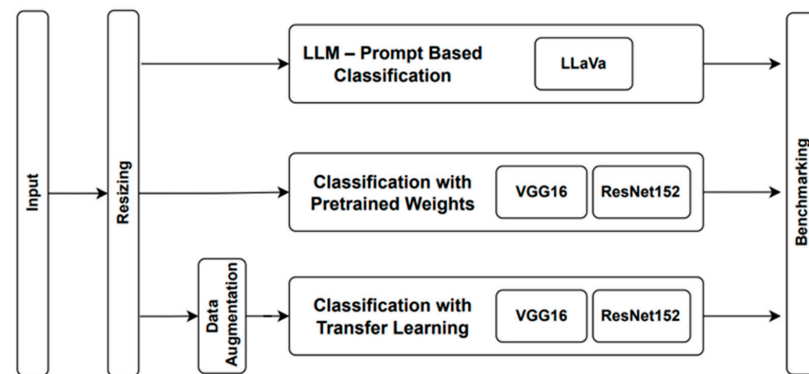
### 3.3. Training

VGG, Resnet, and personalized CNN were trained using the same image datasets, curated from diverse real estate listings, to ensure that they learned information about a wide range of property types. The hyperparameters were optimized individually for each

architecture to ensure valid performance. Learning rate annealing, batch normalization, and early stopping were employed to prevent overfitting and improve model generalization.

#### 4. Experiment

We compared the performances of CNN, VGG16, ResNet152, and Lava Vision on Dataset I with 1000 images labeled with two labels and Dataset II with 3000 images labeled with four labels (Figure 1). The models were trained with an 80:20 split for both datasets, fine-tuned for two models, and zero-shot for Lava Vision. We evaluated the classification capability and the extent of an image-based model to another larger and more varied population.



**Figure 1.** Experiment process.

The accuracy (94.08%) was the highest when using the Lava Vision model. The trained ResNet152 and VGG16 models showed accuracies of 91.89 and 91.14%, respectively. The models adapted to less learning in the dataset. The pretrained models did not show the same level of performance, which showed the necessity of pretraining for image classification and time and machine resources are constraints. When Dataset II was used, the performance of the models was weaker because of the additional intricacy. Nonetheless, the Lava Vision model proved its efficacy with an average accuracy of 91.27% and a receiver operating characteristic (ROC)–area under the curve (AUC) of higher than 95%. The effectiveness of prompt-based modeling was verified in diverse and larger data settings (Table 1). The performance of the pretrained and trained ResNet152 and VGG16 models proved effective for complicated real estate image processes. Image enhancement affected the classification algorithm performance, especially for ResNet and VGG. In Resnet152, the performance was improved by 1.23% on Dataset 1 with transfer learning.

**Table 1.** Accuracy of models.

Model	Experiment Type	Accuracy (%)	
		Dataset I <sup>a</sup>	Dataset II <sup>b</sup> (Average over Labels)
CNN <sup>c</sup>	Trained (80-20)	84.12	81.72
	Trained (80-20) with Image Enhancement	84.29	81.91
VGG16	Pretrained	87.28	82.64
	Pretrained with Enhancement	87.58	82.94
	Trained (80-20)	91.14	86.44
	Trained (80-20) with Enhancement	92.09	86.81

Table 1. Cont.

Model	Accuracy (%)		
	Experiment Type	Dataset I <sup>a</sup>	Dataset II <sup>b</sup> (Average over Labels)
CNN <sup>c</sup>	Trained (80-20)	84.12	81.72
	Pretrained	89.37	83.58
ResNet152	Pretrained with Enhancement	90.01	84.12
	Trained (80-20)	91.89	87.22
	Trained (80-20) with Enhancement	93.20	87.94
LLaVa Vision	Prompt-based	94.08	91.27
	Prompt-based with Enhancement	94.22	91.53

<sup>a</sup> Dataset I has a 50:50 balance ratio. <sup>b</sup> Dataset II has a 25:25:25:25 label balance ratio. <sup>c</sup> The CNN model is given in the III.a.CNN Model.

## 5. Conclusions and Future Works

AI and ML models in the real estate industry significantly enhance the user experience, particularly in cloud services, while supporting user and customer satisfaction. Classifying and enhancing images were used in higher-quality real estate listings. Promising results were obtained by integrating large language models (LLMs). When using the LLaVA model with other models, such as LLama and ChatGPT-4o, and preprocessing techniques such as image enhancement before classification, improved classification results were obtained. Various image enhancement models can be used to improve the performance. Combining different methods improved the efficiency of uploading images and classifying them according to specific quality criteria, thus increasing the visual appeal of real estate listings. In addition, the automated process significantly reduced the time and effort for quality control, increasing operational efficiency. Faster and more accurate results with less human intervention accelerated the overall processes. As the overall image quality of the real estate platform increases and becomes more appealing, an increase in user satisfaction and the number of customers is expected. These developments ensure the platform's reliability and professionalism and positively impact customer loyalty and revenue growth in the long term.

**Author Contributions:** Conceptualization, E.T.K., G.A. and M.S. methodology, E.T.K., M.S. and G.A.; software, E.T.K., M.S., Y.D. and G.A. validation, E.T.K., A.H.I., M.S. and G.A. formal analysis, A.H.I., E.T.K., M.S. and G.A.; investigation, E.T.K., A.H.I., M.S., G.A. and Y.D. resources, G.A. and Y.D.; data curation, E.T.K., M.S. and G.A.; writing—original draft preparation, E.T.K., M.S. and G.A.; writing—review and editing, E.T.K., A.H.I., M.S., G.A. and Y.D. visualization, E.T.K. and M.S.; supervision, A.H.I., M.S., G.A. and Y.D.; project administration, Y.D.; funding acquisition, G.A. and Y.D. All authors have read and agreed to the published version of the manuscript.

**Funding:** The research by Emlakjet (Emlakjet İnternet Hizmetleri ve Gayrimenkul Danışmanlığı Anonim Şirketi) was carried out at the Emlakjet Research and Development Center with financial support from The Scientific and Technological Research Council of Türkiye (TÜBİTAK) (Grant No: 7220634).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** No new data were created.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Statista. Turkey: Annual Growth Rate of the Real Estate Industry 2023. Available online: <https://www.statista.com/statistics/1374914/turkey-annual-growth-rate-of-the-real-estate-industry/> (accessed on 18 October 2024).
2. Statista. Turkey: Residential Real Estate Transactions 2028. Available online: <https://www.statista.com/forecasts/1427261/residential-real-estate-transactions-value-turkey> (accessed on 18 October 2024).
3. Yuan, F.; Wu, J.; Wei, Y.D.; Wang, L. Policy change, amenity, and spatiotemporal dynamics of housing prices in Nanjing, China. *Land Use Policy* **2018**, *75*, 225–236. [\[CrossRef\]](#)
4. Lemeš, L.; Akagic, A. *Prediction of Real Estate Market Prices with Regression Algorithms*; Lecture Notes in Networks and Systems; Springer: Cham, Switzerland, 2023; Volume 539, pp. 401–411. [\[CrossRef\]](#)
5. Law, S.; Paige, B.; Russell, C. Take a look around: Using street view and satellite images to estimate house prices. *ACM Trans. Intell. Syst. Technol.* **2019**, *10*, 54. [\[CrossRef\]](#)
6. Semnani, S.J.; Rezaei, H. House Price Prediction using Satellite Imagery. *arXiv* **2021**, arXiv:2105.06060.
7. Liu, Y.; Gross, L.; Li, Z.; Li, X.; Fan, X.; Qi, W. Automatic Building Extraction on High-Resolution Remote Sensing Imagery Using Deep Convolutional Encoder-Decoder with Spatial Pyramid Pooling. *IEEE Access* **2019**, *7*, 128774–128786. [\[CrossRef\]](#)
8. Zhu, X.X.; Tuia, D.; Mou, L.; Xia, G.S.; Zhang, L.; Xu, F.; Fraundorfer, F. Deep learning in remote sensing: A review. *IEEE Geosci. Remote Sens. Mag.* **2017**, *5*, 8–36. [\[CrossRef\]](#)
9. Bappy, J.H.; Barr, J.R.; Srinivasan, N.; Roy-Chowdhury, A.K. Real estate image classification. In Proceedings of the 2017 IEEE Winter Conference on Applications of Computer Vision, WACV 2017, Santa Rosa, CA, USA, 24–31 March 2017; pp. 373–381. [\[CrossRef\]](#)
10. Råhlén, O.; Sjöqvist, S. Image Classification of Real Estate Images with Transfer Learning. 2019. Available online: <https://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-259759> (accessed on 11 September 2024).
11. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [\[CrossRef\]](#)
12. He, K.; Zhang, X.; Ren, S.; Sun, J. *Identity Mappings in Deep Residual Networks*; Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Springer: Cham, Switzerland, 2016; Volume 9908, pp. 630–645. [\[CrossRef\]](#)
13. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. In Proceedings of the 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, 7–9 May 2015. Conference Track Proceedings.
14. Radford, A.; Kim, J.W.; Hallacy, C.; Ramesh, A.; Goh, G.; Agarwal, S.; Sastry, G.; Askell, A.; Mishkin, P.; Clark, J.; et al. Learning Transferable Visual Models From Natural Language Supervision. *Proc. Mach. Learn. Res.* **2021**, *139*, 8748–8763.
15. Jia, C.; Yang, Y.; Xia, Y.; Chen, Y.T.; Parekh, Z.; Pham, H.; Le, Q.V.; Sung, Y.; Li, Z.; Duerig, T. Scaling Up Visual and Vision-Language Representation Learning With Noisy Text Supervision. *Proc. Mach. Learn. Res.* **2021**, *139*, 4904–4916.
16. Moran, S.; McDonagh, S.; Slabaugh, G. CuRL: Neural curve layers for global image enhancement. In Proceedings of the International Conference on Pattern Recognition, Milan, Italy, 10–15 January 2021; pp. 9796–9803. [\[CrossRef\]](#)
17. Hu, Y.; He, H.; Xu, C.; Wang, B.; Lin, S. Exposure: A White-Box Photo Post-Processing Framework. *ACM Trans. Graph.* **2017**, *37*, 26. [\[CrossRef\]](#)
18. Gharbi, M.; Chen, J.; Barron, J.T.; Hasinoff, S.W.; Durand, F. Deep bilateral learning for real-time image enhancement. *ACM Trans. Graph.* **2017**, *36*, 118. [\[CrossRef\]](#)
19. Liu, H.; Li, C.; Wu, Q.; Lee, Y.J. Visual Instruction Tuning. *Adv. Neural Inf. Process. Syst.* **2023**, *36*, 34892–34916.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.