DOI: 10.36558/rsc.v13i2.8251

An Approach Based On CNN To Residential Environment Classification Focused On Real Estate Business

Ramon Brignoli
Department of Metal-Mechanics (DAMM)
Federal Institute of Santa Catarina
Florianópolis (SC) - Brazil
ramon.brignoli@gmail.com

Maurício Edgar Stivanello
Department of Metal-Mechanics (DAMM)
Federal Institute of Santa Catarina
Florianópolis (SC) - Brazil
mauricio.stivanello@ifsc.edu.br

ABSTRACT

Few studies available in the literature address the classification of images of home environments using computer vision. The demand for such a task focused on real estate business applications has not yet been studied, and the requirements differ from those existing in applications such as robotics for which the problem has already been previously explored. In the present work, a classification system for furnished or unfurnished domestic environments is proposed. A convolutional neural network is used to classify images as belonging to the façade, kitchen, bathroom, bedroom, living room or external area environments. The results obtained by evaluating different architectures and using real images of the application domain show that the proposed approach allows to achieve accuracy of approximately 97%, being even superior to previous work focused on other application domains.

Keywords

Classification of Environments, CNN, Computer Vision, Convolutional Neural Networks, Image Analysis, Real Estate Business.

1. INTRODUCTION

The Real estate sector of buying, selling and leasing real estate has increasingly adapted to new technologies. Today, internet portals have been used by companies in this field in order to digitize their business [4]. Among the digitized steps is the real estate announcement, which demands the organization and publication of descriptive content, usually textual and illustrative.

The organization and categorization of the images to be used in the dissemination of ads is important since the orderly presentation of the environments facilitates the understanding of spaces by customers. Such structuring of the images, however, demands great manual work due to the large volume of real estate and images generated and by consequence it is often not used.

Few studies available in the literature address the classification of images of environments using computational vision. Still, the works found are focused on different applications of the proposal. For example, in robotics applications, the classification of spaces can help in obtaining semantic information from mobile robots [8, 9]. Depending on the type of application described, the images only include furnished spaces, and thus challenges ranking in the context of ads, such as empty environments, were not addressed.

Considering these challenges, in the present work a classification system based on convolutional neural networks that automatically categorizes environments present in the images used in real estate advertisements is proposed. The categories used in the classification are façade, kitchen, bathroom, bedroom, living room, external area, selected in order to meet requirements present in such field of application.

In section 2, related papers that have already addressed the theme of classification of environments for other applications are described. In section 3, details of the architecture of the proposed classification system are presented. In section 4, the experimental results obtained with the use of the developed system are presented. Finally, in Section 5, the conclusions and perspectives for future work are discussed.

2. THE PROBLEM OF CLASSIFICATION OF ENVIRONMENTS AND PREVIOUSLY EXPLORED APPROACHES

The classification of environments through images is of interest to different application domains. In addition to the problem of image classification in the context of application of the real estate sector, a theme not yet explored or not explored enough, tasks associated with the application of robotics are reported in studies in the literature.

[9] proposes the use of environment classification to assist in locating mobile robots during operation in cooperative environments. The laboratory, hall, auditorium, office and kitchen environments are listed as relevant in such application. To perform the classification, a localization system that uses descriptors extracted from the images provided by a camera and also from the distance data provided by a laser sensor with 360 degree field of view is proposed. Such descriptors are used for the classification of environments statistically using Hidden Markov Models. The results achieved vary among the categories of environments used, presenting a maximum error of 2.67%.

In a more recent study, the problem of classifying household rooms was considered, focusing on the integration of

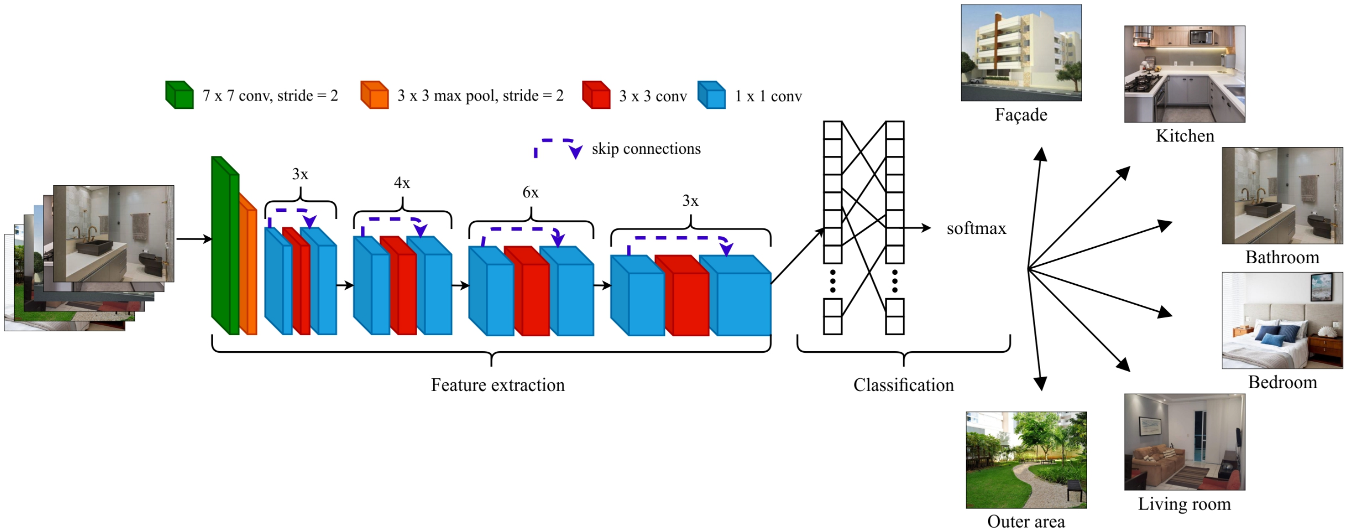


Figure 1: Simplified description of the classifier architecture

such capacity to a social robot [8]. Internal environments such as bathroom, bedroom, dining room, kitchen and living room were listed. The proposed method for classification involves the use of Convolutional Neural Networks (CNNs) whose input are images acquired by conventional cameras. The proposed system was able to classify the environments presenting an error lower than 6.39%.

Although in these studies the results have allowed promising results in robotics applications, some aspects are quite different from those found in an application scenario for the real estate sector. In this area, for example, it would not be feasible to use data provided by three-dimensional sensors since the acquisition of the materials is usually done by non-specialized equipment, such as digital cameras or even mobile phones. Another important aspect is that some environments, or even characteristics of them, are not considered in the studies found. For example, none of them listed external environments, important for the application mentioned. Still, environments or unfurnished spaces were not considered in such works, also relevant.

In relation to the different techniques used to create a classifier in the correlated studies, the use of CNNs stands out. The interest is justified, among other aspects, by its simplicity of use when compared to classical image processing techniques, since once its architecture has been defined, it is only necessary training for results to be obtained. Such characteristics have allowed applications of different areas, including quality control and recognition of pieces and parts in the industry [10, 13, 12, 12], diagnosis of imaging diseases in medicine [1], segmentation of forest images in biology [14], among others.

3. CLASSIFICATION OF HOUSEHOLD ENVIRONMENTS BASED ON CNNS

The main functional requirement that guided the development of the proposed system was the need for automatic classification of environments present in the images used in real estate listings. The categories considered in the study are façade, kitchen, bathroom, bedroom, living room, external area, defined as important according to experts.

A proposal based on CNNs was used in the development of the classification system for the mentioned application. Its use was motivated by the fact that they have provided superior results than those obtained when using classical computer vision techniques in different application scenarios, as described in Section 2.

The used classifier corresponds to a Residual Artificial Neural Network (ResNet). This type of network has been selected since it corresponds to a well-established architecture, in addition to the fact that there are general purpose models previously trained with large sets of images and thus prepared to recognize various visual characteristics. Thereby, when the training of a new model is carried out aimed at the classification of new categories, such as those that categorize real estate environments, the use of previously trained models improves accuracy and reduces the training time required. It happens since the image descriptors in the previous training can be transferred or reused in the new application, without having to be learned again [2].

Three ResNet configurations have been tested, with 18, 34 and 50 layers. An overview of the architecture adapted to the classification of the listed environments is presented in Figure 1, using the ResNet-50 setting as an example.

As presented in the architecture figure, blocks with different dimensions and a step equal to 2 are used in the characteristic extraction layers. Repetitions of 3, 4, 6 and 3 times are used from the first to the last block, respectively. The dotted arrows present over each second layer indicate the shortcut connections that connect blocks of different dimensions. Such connections allow the training of deeper networks without performance loss. Since these shortcut connections are only validated in cases of dimension matching, and this will not always be the case, zero padding is used to allow residual operations over the entire process. Batch normalization is used to avoid the vanishing gradient problem. After the sequence of convolutions, the last layer is flattened and passed on to a fully connected layer. Finally, the softmax function is used to generate the probability of the 6 classes corresponding to the façade, kitchen, bathroom, bedroom, living room and external area.

The illustrated architecture is valid for the ResNet-50.

The architectures of the ResNet-18 and 34 configurations are similar, but instead of groups of three convolutions (1x1, 3x3 and 1x1), the repeating blocks in the extraction of characteristics have only two 3x3 convolutions and the number of repetitions is also different.

4. EXPERIMENTAL RESULTS

4.1 Implementation and Training of Classifiers

The software that implements the classifiers described in section 3 was developed using the Python programming language together with the Fastai library [7, 6]. A proprietary image dataset containing 19.581 actual samples of the application domain distributed in the six classes to be considered by the classifiers, as exemplified in Table 1, was used in the training. The computing environment used during the training task was Google Colab.

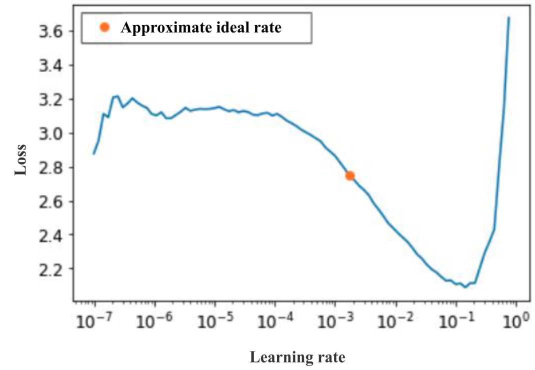
The procedure adopted for model training for each ResNet configuration was the same. Transfer learning was used to reduce the time required for the training task. For this purpose, the pre-trained ResNet model on the ImageNet database was used as a starting point. To adjust the CNN model so that it performed the specific classification of the desired classes, the previously mentioned image dataset was used. 30% of the images representing samples from different classes of environments were randomly selected to make up the validation group, with the remaining 70% used for training. Regarding transformations, a resizing of 128x128 was applied, and a random data augmentation was also defined on the image dataset.

The group of images described and the models whose weights were previously trained were used to train the CNNs. A first training stage where the convolutional layers remained fixed and where only the weights of the classification layers could be modified was used to obtain a first model. A second training stage where both the weights of the convolution layers and the classification layers could be modified was performed, in order to obtain a final improved model for each of the three ResNet configurations evaluated. The number of times and the learning rate applied to the training were obtained experimentally, from tests performed on a small batch of images. As can be seen in Figure 2, a learning rate equal to 0.005 proved to be adequate for training. Likewise, a number of 10 epochs proved sufficient, and higher values did not result in gains for the models.

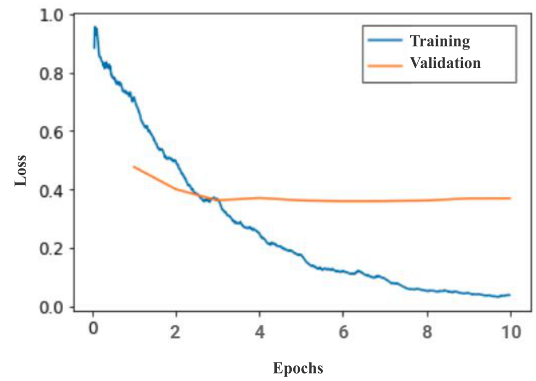
4.2 Evaluation of Classification Results

The evaluation of the classification results of the models obtained was performed using 30% of the images of the dataset used in the work and which were not used in the training stage [11]. The performance of the models can be observed based on Precision (P_r), Recall (R_c), F1 score (F_1) and Accuracy (A_{cc}) [5]. Such metrics are calculated considering results where both the value indicated by the classifier and indicated by the specialist are positive (TP), where both the value indicated by the classifier and indicated by the specialist are negative (TN), where the value indicated by the specialist is negative but the indicated by the classifier is positive (FP), and where the value indicated by the specialist is positive but the one indicated by the classifier is negative (FN).

Considering such results, we have that the value of precision can be calculated by:



(a) Learning rate evaluation



(b) Training epochs evaluation

Figure 2: Parameters selection

$$P_r = \frac{TP}{TP + FP}, \tag{1}$$

the recall can be calculated by:

$$R_c = \frac{TP}{TP + FN}, \tag{2}$$

the F1 score is given by:

$$F_1 = 2 \cdot \frac{P_r \cdot R_c}{P_r + R_c}, \tag{3}$$

and in turn, the accuracy is calculated by:

$$A_{cc} = \frac{TP + TN}{TP + TN + FP + FN}, \tag{4}$$

corresponding to the percentage of correct ratings result. It is desirable that the values of such indexes be as close to 100% as possible.

The performance statistics of the evaluated classification models for the different architectures obtained based on the indexes described are presented in Table 2.

As can be seen, the architectures achieved very approximate results, and the ResNet-50 obtained slightly higher average performance. We can evaluate in more detail the situa-

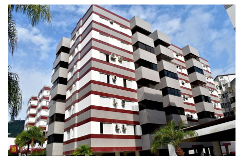
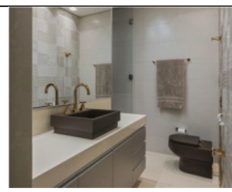
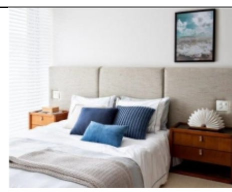
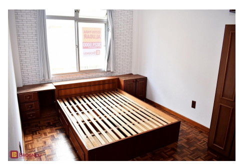
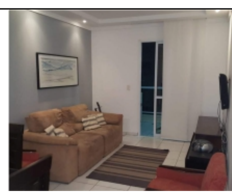
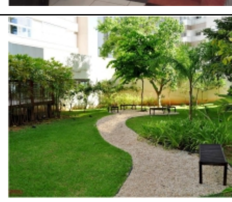
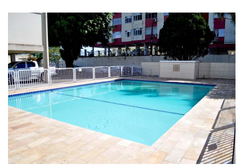
Environment	Example 1	Example 2	Example 3
Façade			
Kitchen			
Bathroom			
Bedroom			
Living room			
Outer area			

Table 1: Image dataset examples

Table 2: Models performance

Architecture	Accuracy	Precision	Recall	F1
ResNet-18	**0,976**	0,965	0,963	0,964
ResNet-34	0,967	0,966	0,965	0,965
ResNet-50	0,969	**0,968**	**0,970**	**0,969**

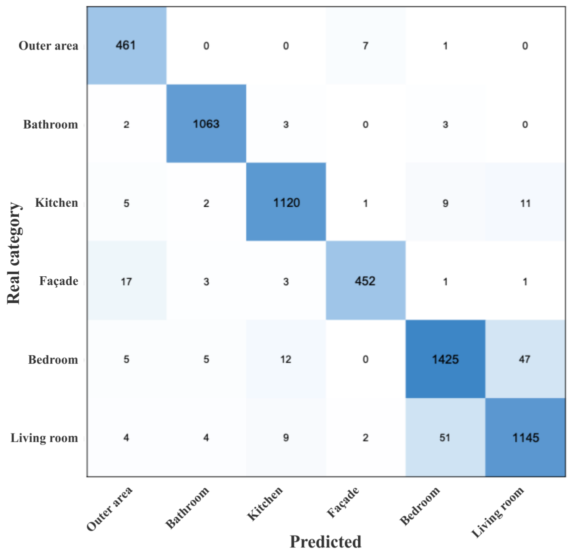


Figure 3: Confusion matrix - ResNet-50

tions that contributed to the existence of an error of approximately 0.031% associated with this architecture through the confusion matrix presented in Figure 3.

It is possible to observe from the matrix that there is a predominance of classification errors associated with the bedroom and living room categories. In the case of bedroom images, 1425 were correctly classified. Among the incorrectly classified images, most were erroneously indicated as room. The same is observed in the opposite direction, where 51 of the living room photos were erroneously classified as bedrooms.

The difficulty of the model in classifying some living room and bedroom images is linked to the fact that there are several unfurnished rooms in the catalogs from where the database was generated, as well as for the fact that these two environments have very similar visual characteristics in this condition. This fact can be observed in Table 3, which presents some of the images that generated the greatest loss when classified. As can be seen, the absence of furniture or any kind of evident and distinguishable feature prevents the classifier from performing well.

Another source of classification errors concerns the classification of façade images. 17 images of this category were incorrectly pointed out as belonging to the outer area category. By looking at examples of images of the two categories in Table 4, we can conclude that the incorrect indications are the result of the fact that some of the images of the two categories of environments have great structural similarity.

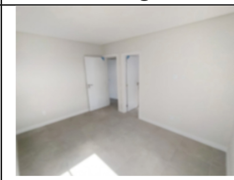
Real category	**Prediction**	**Image**
Bedroom	Living room	
Living room	Bedroom	

Table 3: Comparison of classification between living room and bedroom

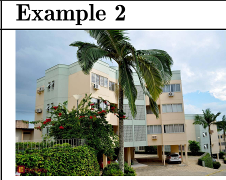
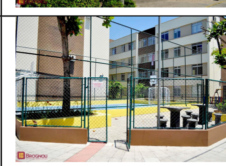
Env.	**Example 1**	**Example 2**
Façade		
External area		

Table 4: Examples of similar images of the Façade and External Area categories

Table 5: Average processing time

Architecture	Average running time (ms)
ResNet-18	72
ResNet-34	71
ResNet-50	80

4.3 Computational Cost Assessment

The average inference processing time for the different evaluated architectures using Google Colaboratory infrastructure [3] is presented in Table 5.

The computational cost observed is considered adequate for the non-functional requirements associated with a system aimed at the real estate sector, where real-time restrictions are not present.

5. CONCLUSIONS AND FUTURE WORK

In the present work, a proposal for a method of classification of images of residential environments using convolutional neural networks is described. Different architectures were evaluated.

Among them, the **ResNet-50** network presented the best results for most of the calculated performance indexes. This architecture reached an accuracy of 96.9% and requires an inference time of 80 ms. It is considered that the accuracy and computational cost observed meet the requirements of a classification system focused on the application domain considered.

As future work, it is intended to create a classifier that includes new categories, such as swimming pool and parking, and also investigate a solution to reduce the occurrence of incorrect indications associated with the living room and bedroom categories whenever there are images of unfurnished environments, as well as with the façade and external area categories. It is also considered the availability of the classifier in an API so that it can be used online.

6. REFERENCES

[1] S. Bansal, A. Rustagi, and A. Kumar. Alzheimer's disease diagnosis based on feature extraction using optimised crow search algorithm and deep learning. *International Journal of Computer Applications in Technology*, 65(4):325–333, 2021.

[2] K. K. Bressem, L. C. Adams, C. Erxleben, B. Hamm, N. S. M., L. Vahldiek, Janis L. Zhu, J. Wang, and K. Li. Comparing different deep learning architectures for classification of chest radiographs. *Scientific Reports*, 10(1):13590, 2020.

[3] T. Carneiro, R. V. Medeiros Da NóBrega, T. Nepomuceno, G.-B. Bian, V. H. C. De Albuquerque, and P. P. R. Filho. Performance analysis of google colaboratory as a tool for accelerating deep learning applications. *IEEE Access*, 6:61677–61685, 2018.

[4] N. d. S. Dias. Adaptação organizacional: A influência das novas tecnologias na estratégia de empresas do setor imobiliário da grande florianópolis, 2019.

[5] M. Grandini, E. Bagli, and G. Visani. Metrics for multi-class classification: an overview, 2020.

[6] J. Howard. *Deep Learning for Coders with Fastai and PyTorch*. O'Reilly Media, 2018.

[7] J. Howard and S. Gugger. Fastai: A layered api for deep learning. *Information*, 11(2):108, Feb 2020.

[8] K. M. Othman and A. B. Rad. An indoor room classification system for social robots via integration of cnn and ecoc. *Applied Sciences*, 9(3), 2019.

[9] A. Rottmann, O. M. Mozos, C. Stachniss, and W. Burgard. Semantic place classification of indoor environments with mobile robots using boosting. In *Proceedings of the 20th National Conference on Artificial Intelligence - Volume 3*, AAAI'05, page 1306–1311. AAAI Press, 2005.

[10] M. E. Stivanello, J. E. N. Masson, and M. R. Stemmer. A cnn approach for online metal can end rivet inspection. *International Journal of Computer Applications in Technology*, 69(3):282–290, 2022.

[11] V. K. Vemuri. The hundred-page machine learning book. *Journal of Information Technology Case and Application Research*, 22(2):136–138, 2020.

[12] D. Vriesman, A. Zimmer, A. S. Britto Jr, and A. L. Koerich. Texture cnn for thermoelectric metal pipe image classification. *arXiv preprint arXiv:1905.12003*, 2019.

[13] H. Wang, T. Pan, and M. K. Ahsan. Hand-drawn electronic component recognition using deep learning algorithm. *Int. J. Computer Applications in Technology*, 62(1):13–19, 2020.

[14] L. Zhu, J. Wang, and K. Li. Computer image analysis for various shading factors segmentation in forest canopy using convolutional neural networks. *International Journal of Computer Applications in Technology*, 64(4):415–428, 2020.