Identification of streetscape compositions: a deep learning approach

Ana Luiza Favarão Leão[a], **Hugo Queiroz Abonizio**[b], **Sylvio Barbon Júnior**[c], **Milena Kanashiro**[d]

[a] Londrina State University, Department of Architecture and Urbanism, Graduate Program in Architecture and Urbanism, Londrina, PR, Brasil. E-mail: analuiza.favarao@uel.br

[b] Londrina State University, Department of Computing, Graduate Program in Computer Science, Londrina, PR, Brasil. E-mail: hugo.abonizio@uel.br

[c] Londrina State University, Department of Computing, Graduate Program in Computer Science, Londrina, PR, Brasil. E-mail: barbon@uel.br

[d] Londrina State University, Department of Architecture and Urbanism, Graduate Program in Architecture and Urbanism, Londrina, PR, Brasil. E-mail: milena@uel.br

https://doi.org/10.47235/rmu.v8i1.140
Submitted on March 8, 2020. Accepted on May 21, 2020.

Abstract. *The environment's composition can have an impact on human behavior; however, this relationship remains uncertain until the cities' qualities and landscape can be analyzed empirically. Images obtained through Google Street View (GSV) enable a large volume of data for automated assessment of environmental characteristics. Deep learning techniques have advanced in the identification of compositional elements of the built environment. In this sense, this study seeks to investigate and test a procedure for identifying the configuration and composition of the urban landscape, classifying images obtained from GSV through a deep learning approach. From an image dataset of three different neighborhoods in Londrina-PR, a deep learning model for image classification was proposed. The model achieved a good performance, correctly attributing 87.6% of the samples to the corresponding neighborhoods in the case study. Compositional characteristics were empirically identified, considering the distribution of the samples in the obtained feature space. The proposed model contributes to the definition of spatial units as well as in the measurement of environmental qualities, optimizing data collection, expanding sample sizes, and providing objectivity to results. This approach contributes to the expansion of city's analytical scales, identifying compositional and relational patterns in the understanding of elements influent in human behavior.*

Keywords: u*rban morphology, built environment, deep learning, image classification, Google Street View.*

Introduction

The city as an anthropic artifact, with different configurations, reflects the complexity of socio-spatial relationships. Thus, the composition and shape of the city, the result of natural and built characteristics, are essential parameters in analyzes focused on Urban Planning and Design. The composition of the built environment has been associated with urban dwellers' physical activity (Sallis et al., 2015), their satisfaction with the environment (Lee et al., 2017), safety against crime (Kamalipour, Faizi e Memarian, 2014) and even happiness of urban individuals (Kent, Ma e Mulley, 2017; Seresinhe *et al.*, 2019). In this sense, urban design can influence people's choices and behavior. However, this relationship remains uncertain until the compositions of the urban landscape and its spatial qualities can be defined, quantified, measured, and tested empirically (Ewing e Handy, 2009).

Among the various theoretical and methodological approaches to the analysis of the urban environment, from historical or morphological readings to appropriations, perceptions, or temporalities, one of the strategies is the definition of spatial units, based on the recognition of patterns. For example, the delimitation of Landscape Units allows for the identification of attributes responsible for landscape dynamics (Amorim e Oliveira, 2008); the definition of sectors/districts/neighborhoods establishes limits for data aggregation in walkability studies (Gehrke e Wang, 2020); the design of homogeneous urban units for grouping areas with the same environmental or physical-spatial characteristics in the formation of urban land use classes (Medeiros e Grigio, 2019). These processes seek to define spatial patterns based on predefined parameters. For such, analyzes of individual characteristics, and subsequent information overlapping for the delimitation of areas have been conducted through field data gathering, secondary data, and criteria definition.

In recent years, online urban images of products like Google Street View (GSV) have become increasingly available, depicting the user's eye-level perspective (Middel et al., 2019). Through the use of these tools, big data and machine learning approaches become possible for studies on the visual characterization of typological elements (Doersch *et al.*, 2012; Yin e Wang, 2016; Liu *et al.*, 2017; Moosavi, 2017; Zhang *et al.*, 2017; Shen *et al.*, 2018), the measurement of urban qualities that possibly influence street scale behavior and well-being (Yin *et al.*, 2015; Yin e Wang, 2016; Liu *et al.*, 2017), as well as the use of high-resolution images for urban morphology studies (Moosavi, 2017; Shen et al., 2018; Zhang et al., 2017). This type of procedure can be a more efficient alternative (Ben-Joseph *et al.*, 2015) when compared to field surveys or audits, that in addition to possibly being prone to subjectivity, are more costly and not always safe for the researcher. (Badland et al., 2010).

On the other hand, urban image acquisition methods such as GSV make it possible to obtain a large data volume for analyses. In this context, approaches using machine learning have gained space in the literature. Such objective methods of obtaining data from the urban environment reduce the time spend on in locus surveys while allowing a larger volume of data (Zhang et al., 2018).

However, traditional machine learning is limited to hand-crafted features to describe a problem and its ability to deal with high dimensional data (e.g., pixels of an image obtained through GSV), resulting in the problem known as Curse of Dimensionality (Poggio et al., 2017). In contrast, deep learning methods present multiple layers of data processing and representation, through the composition of non-linear modules that transform the initial representation (e.g., pixels of an image), into elements of a higher level of abstraction (Lecun, Bengio e Hinton, 2015). Deep learning techniques, such as Convolutional Neural Networks, have advanced with computer vision applications in many research domains, including the identification of compositional elements of the built environment such as aesthetics (Tan et al., 2017) and its landscapes (Zhou et al., 2016).

Considering such methodological and conceptual aspects, this research has the main objective to test a procedure for identifying the configuration and composition of the urban landscape through an image classification approach based on a deep learning method. Considering the hypothesis that the evaluated images, using a classification model, will allow for the recognition of built environment patterns in images obtained through the GSV. This work explores and discusses the application of deep learning models as a tool for the analysis of the urban landscape. In short, we seek to group similar geo-informative landscape compositions (Doersch et al., 2012). From an image dataset, compositional characteristics were identified, considering the distribution of the samples in the feature space constructed through a classification model.

The results of this research contribute to the promotion of deep learning applications for the understanding of the urban landscape in an objective and evidence-based manner. The created model obtained an accuracy of 87.6%, indicating a good generalization capacity. This result demonstrates excellent model performance, suggesting that compositional characteristics that vary between neighborhoods were correctly abstracted, learned, and recognized.

Methods

The adopted research strategy was the case study, as the phenomenon analyzed is contemporary and contextual, therefore inseparable from reality (Yin, 2001). As a case, three sectors defined by the municipality were from the city of Londrina-PR, mainly considering their socioeconomic differences, developmental timeframes, urban layout, and geographic location: *Centro Histórico*, *Cinco Conjuntos* and *Gleba Palhano* (Figure 01).

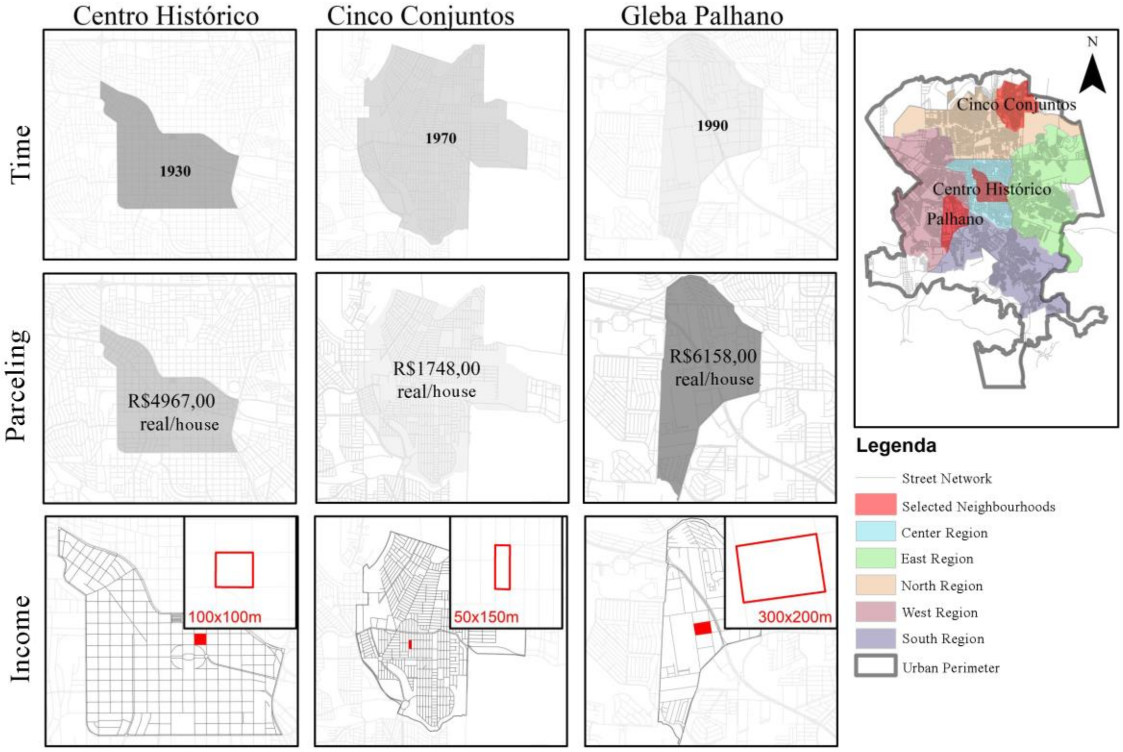


Figure 1. Selected Neighborhoods (source: the authors, 2020).

Londrina has an estimated population of 569,733 inhabitants (IBGE, 2018). The city was built from a pre-project in 1932, as part of the colonization process of the *Companhia de Terras Norte do Paraná*, in the historical context of coffee production. The following years were increasingly marked by urban growth around a planned nucleus (Töws, Mendes e Vercezi, 2010), which today configures the neighborhood called *Centro Histórico*. Orthogonality characteristics of the grid, intense mixed use and vertical residential density mark the assigned value of the city's initial nucleus.

Gleba Palhano, a sector located in the Southwest Zone of Londrina, started its consolidation in the 90's. Numerous factors conditioned the development of this area, but mainly the appeal to the real estate market due to its privileged location within the urban network. Today it is the object of great land speculation. It is currently the most valued region of the city, surpassing the historic center (Oura, 2006). Characterized by an intense vertical occupation, the pattern of buildings allocated in large lots is directed to the high-income population.

The development of the Southwest Zone of Londrina is opposed to that of the North Zone, characterized, above all, by social interest housing projects for the lower class and slum areas (Töws, Mendes e Vercezi, 2010). *Cinco Conjuntos*, the third selected area, is located in the North Zone and has its origin in housing projects from the 1970s. This development was mediated by the logistics of house construction of the Housing Company - COHAB / LDA, responsible for subsidizing the construction of low-income housing throughout the outskirts of Londrina (Beidack e Fresca, 2011). From the 1990s, with a consolidated population density, a new centrality was established throughout the Saul Elkind avenue, with commercial and service land uses.

To create the neighborhood identification model, a supervised machine learning approach was used, in which a set of samples labeled with the corresponding class is used to model the conditional distribution of the classes that make up the problem (Bishop, 2006). In this approach, a model is induced by a set of input-output pairs that are used to learn existing patterns in the data. In this work, an image of the urban landscape is used as an input to the model, and the neighborhood to which the image belongs is the output.

The development of the model and analysis of the results were divided into four stages: (1) data collection, in which the acquisition of the images available in the GSV was performed; (2) pre-processing, in which the images go through a treatment phase that aims to improve the quality of the data; (3) the model induction stage, for model training using the samples that were pre-processed; and, finally, (4) the analysis of the results through model interpretation techniques. Figure 02 illustrates these steps in detail.

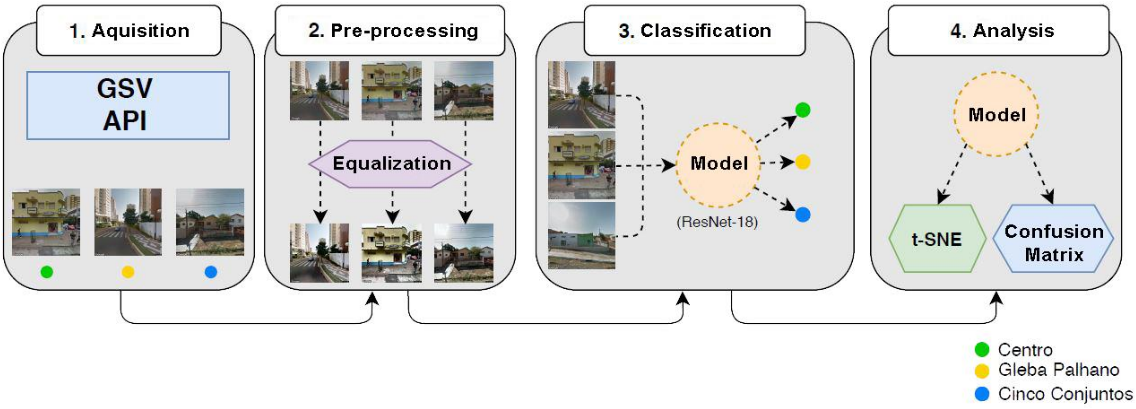


Figure 2. Methodological development steps of the research (source: the authors, 2020).

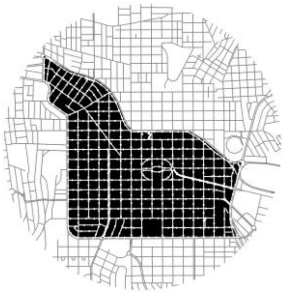
Acquisition

A sample of points distributed every 100 meters through the street network was conducted using the ArcGis 10.6 software. The Data Management toolbox was used, with the Generate Points Along Lines tool, which allocates points along lines at fixed intervals. Considering street centerline data provided by the City of Londrina (Prefeitura do Município de Londrina, 2020), such selected metric interval was adopted as it represents the approximate block sizes of the initial city plan (Yamaki, 2017, p. 64).

Through the GSV API (Application Programming Interface), images were obtained from using geographic location of the sample points. The following parameters were used for the requests: camera oriented to the north, neutral vertical rotation in relation to the ground, and the definition of the field of view at 90 degrees. The samples that resulted in an acquisition error by the API, due to the lack of available images for the given coordinate, were discarded.

A total of 2,017 samples was obtained, forming the image dataset used in this research (Table 01). The sampling totaled n = 554 points in the *Centro Histórico*, n = 368 points in *Gleba Palhano*, and n = 1095 in *Cinco Conjuntos*. There was an imbalance in the sample number by neighborhood, mainly due to differences in the extent of the street network. It is pointed out that *Cinco Conjuntos*, in addition to having the largest dimension in area, has a more extensive street network (Table 01) due to macro-parceling marked by long rectangular blocks with lots of minimum dimensions.

Table 1. Street network characteristics and sampling (source: the authors, 2020).

	Cinco Conjuntos	*Centro Histórico*	*Gleba Palhano*
Neighborhood			
Extent	129,76 km	49,68 km	61,16 km
Street network	6,80km²	4,26km²	3,26km²
Sample	1095	554	368

Pre-Processing

After the samples were collected, the next step was the pre-processing of the images, aiming to improve the quality of the predictions produced by the model and avoid overfitting due to uninformative characteristics. Considering that the GSV images were acquired at different times of the day and in different climatic conditions, an image processing was conducted in order to prevent characteristics such as lighting and saturation from being highlighted by the model and recognized as information related to the neighborhood.

To remove the lighting differences, a color space was used where, unlike the standard RGB, the lighting is isolated from other components of the image. Therefore, the images were converted to the YCbCr color space, using the OpenCV[1] library. YCbCr is a color space in which the Y channel represents the luminance, the Cb channel the blue chrominance and Cr the red chrominance (Gowda e Yuan, 2019). The histogram of the luminance channel was used to equalize the images. Thus, a broader and more uniform distribution of the Y channel intensity values was created, attenuating the differences caused by lighting contrasts.

Classification

After pre-processing, with the images prepared for modeling, the next step was the creation of the image classification model. In order to assess the performance of the model without any inductive bias, the images were divided into a training set and a test set, where the first is used to induce the model and the second is used to assess the model's performance. Therefore, the test set serves as a parameter to quantify the model's ability to generalize to new samples, i.e., samples not used during the training phase (Bishop, 2006). Thus, the samples were randomly separated and stratified in relation to the class in a set containing 80% of the images used for training the model, and the remaining 20% were used to measure performance.

Then, for the creation of the image classification model, the ResNet architecture was selected (He e Sun, 2016), a Convolutional Neural Network that uses residual blocks and shortcut connections to avoid the vanishing/exploding gradients problem (Glorot e Bengio, 2010). This architecture has proven effective in a wide range of problems, including applications in medical image analysis (Litjens et al., 2017), agriculture (Kamilaris e Prenafeta-Boldú, 2018), automatic scene description (Anderson et al., 2018) and pedestrian identification (Fan et al., 2018). Therefore, the ResNet architecture proves to be robust and suitable for several applications in image analysis.

An 18-layer ResNet (ResNet-18) previously trained on the ImageNet dataset was used (Deng et al., 2009). Such an approach aims to utilize the patterns learned from large datasets - such as ImageNet, which contains millions of samples - and transfer this learning to leverage the identification of characteristics with a smaller number of samples. These patterns range from simple features such as lines and borders to more complex features such as polygonal shapes and textures in deeper layers of the network.

From the network setup with the previously trained weights, the model was fine-tuned for 30 epochs using the Adam algorithm (*Adaptive Moment Estimation*) (Kingma e Ba, 2014) to optimize the parameters. By doing so, the parameters learned in the previous training are adjusted in the investigated dataset, learning to classify the images in their respective neighborhood.

Since there is an imbalance between classes, i.e., different amounts of samples in each neighborhood, where one is significantly larger than the others, different weights between classes were used to compensate for the representativeness of each class in the total set. Traditional data augmentation techniques were also applied, such as horizontal rotation, which doubles the variation of images by mirroring the horizontal axis, and random clipping, which creates several variations of the image while maintaining the same center (Takahashi, Matsubara e Uehara, 2019; Shorten e Khoshgoftaar, 2019). These procedures were implemented through the machine learning library *PyTorch*[2] and its default parameters.

Analysis

In order to understand the patterns recognized by the model and interpret the modeled characteristics, a visualization in two dimensions of the test set samples was performed using a dimensionality reduction technique. The selected technique was t-SNE (t-Distributed Stochastic Neighbor Embedding), a non-linear method of data visualization with high dimensionality that assigns to each point of an N-dimensional space a location in a new 2D or 3D space (Maaten e Hinton, 2008).

The algorithm works by creating a Gaussian probability distribution based on the similarity between each pair of samples in the original space. Then, this distribution is recreated using the t-Student distribution in the new low-dimensional space. The optimization of the distribution in the new projection is done through gradient descent on the Kullback-Leibler divergence between the two distributions. Unlike other dimensionality reduction techniques that take into account only linear relationships in the feature space, such as Principal Component Analysis (Burges, 2010), t-SNE also maps complex non-linear relationships to the local and global data structure.

To visualize the samples across the feature space learned by the model, the representation resulting from the last convolution layer of the network was extracted, before being passed to the classification layer. Thus, vectors with 512 values were extracted for each image, representing the features learned by the model. From those feature vectors, the samples were then projected in two dimensions using t-SNE (Chan et al., 2018).

In addition to the visualization of the samples in the feature space, the confusion matrix was also used (Powers, 2011), indicating the sample count in relation to its original label and the prediction made by the classification model. Therefore, it is possible to analyze the performance of the model in relation to each class and the similarities learned between them, in the comparison of the original label of the samples and the label obtained by the classification.

Results

The t-SNE technique was applied for the visualization of the samples learned by the model. In Figure 03, two graphs are presented with the sample distribution from the test set, differentiating by color the neighborhood to which the sample belongs and highlighting the samples that were misclassified. From this projection, it was possible to identify cluster patterns in the structure of the learned feature space.

The graph presented on the left, Figure 03(a), depicts two large groups of samples, one clustering together *Centro* samples and other from *Cinco Conjuntos* samples, indicating a clear division of the characteristics detected by the model in these two areas. *Gleba Palhano*, in an intermediary position, is divided into two clusters, one closer to samples from *Cinco Conjuntos* and another closer to *Centro*. This division is due to shared characteristics between *Gleba Palhano* and the other neighborhoods. A section of the area is recently verticalized – resembling traces found in the city center – whereas other areas are still empty and in construction.

The graph on the right, Figure 03(b), identifies the samples that were labeled differently from the actual neighborhood that it belongs. It is essential to highlight that there is an area between the clusters presents a high overlap, indicating that there is no linear distinction among them. Therefore, it is possible to identify samples from the *Centro* neighborhoods, allocated in the *Cinco Conjuntos* cluster, with the same being recognized in combinations of other classes. Such a result is derived from the similar urban landscape characteristics existing in these areas. Therefore, precisely the samples in the transition areas between classes, locus of more considerable disturbance, were those that the model wrongly predicted its neighborhood.

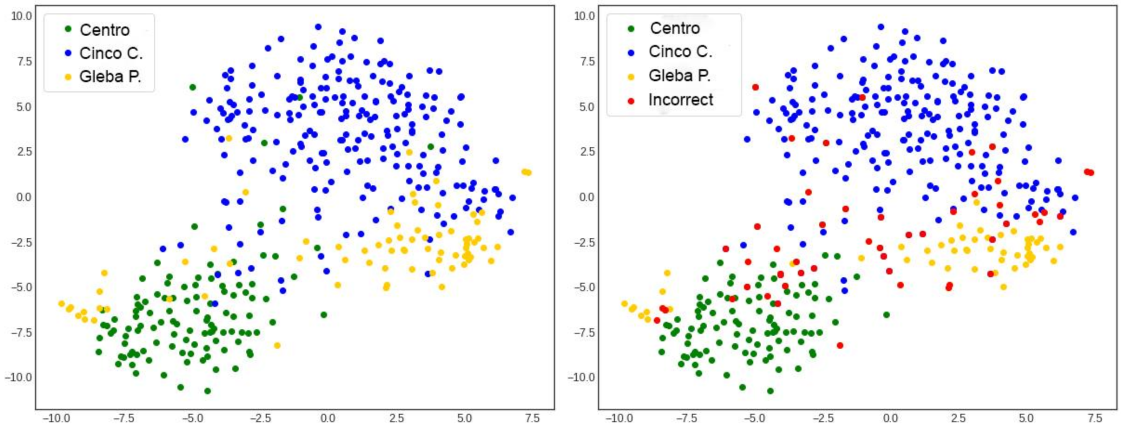


Figure 3. Feature space - Projection in two dimensions of the samples using t-SNE; (a) left: formation of sample clusters of the same class; (b) on the right: highlight of samples misclassified by the model (source: the authors, 2020).

In addition to visualizing the division of the classes in the feature space, the two-dimensional projection of the samples with the original images at each point in the two-dimensional space was generated (Figure 04). By doing so, it is possible to observe the similarity of the closer samples in the space, in comparison with the samples in more distant groups. The clusters recognized in Figure 03 are easily identified in Figure 04, indicating the difference of characteristics in the composition of the landscape of *Cinco Conjuntos*, at the top, and *Centro*, at the bottom. The two separate groups of samples from *Gleba Palhano* are also identifiable in the areas next to *Cinco Conjunto*, on the right part of the graph, and in *Centro*, in the lower-left corner.

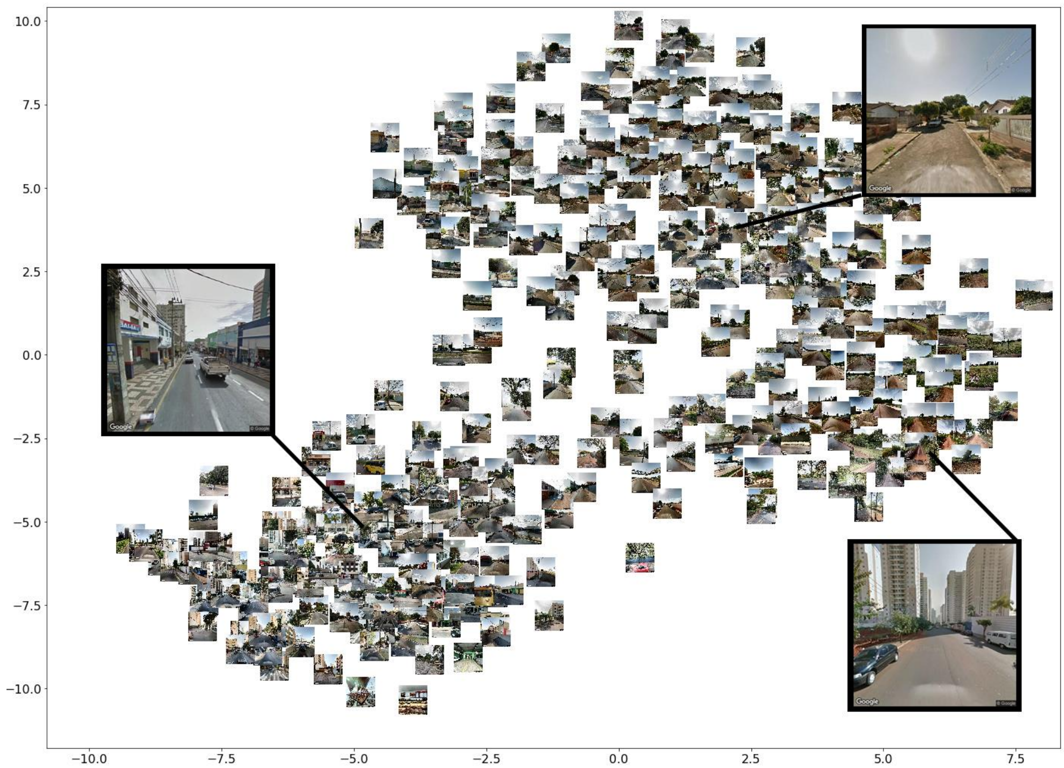


Figure 4. Two-dimensional projection of the samples using t-SNE showing the images collected in each point of the two-dimensional space (source: the authors, 2020).

The confusion matrix compares the actual label of the samples, i.e., the neighborhood to which the sample originally belongs, and the label predicted by the classification model in the test set (Table 02). It is possible to verify which were the model's main errors concerning to each class. When analyzing the number of correct and incorrect samples, the sample imbalance between classes must be considered.

The neighborhood in which the classification model achieved the best performance was *Cinco Conjuntos*, where only 6% of the samples in the test set were misclassified. This fact can be explained by two arguments: this class has the largest number of samples, representing more significant variability of compositions to be learned by the model in the training set and generalized for the test set; and such performance can be explained by the homogeneity of the landscape, result of the prevalence of social interest housing projects.

The *Centro Histórico* obtained a medium performance, with misclassifications in 12% of the samples in the test set. Despite presenting less dispersion in the feature space, compared to samples from *Cinco Conjuntos*, the model identified samples from other classes with similar morphological characteristics. This probably occurs due to the temporal aspects that characterize the area with elements inherent to the sedimentation of the central portion of the city.

The worst performance was for *Gleba Palhano*, with 31% of the samples classified incorrectly. The class is mainly clustered in two regions. However, there are samples dispersed in the intermediate area between the classes. This occurrence may be associated with the smaller number of samples in the class. This fact makes it difficult to learn the characteristics during the training phase due to the reduced variability in space. Figure 03 depicts that these samples were the ones where points are less easily distinguished between classes.

Table 2. Confusion Matrix (source: the authors, 2020).

		Predicted Label		
		Centro	*Cinco Conjuntos*	*Gleba Palhano*
Real Label	*Centro*	98	8	5
	Cinco Conjuntos	6	205	8
	Gleba Palhano	10	13	51

Figure 05 exemplifies situations in which the model misclassified the samples. Regarding predictions of the *Centro Histórico* neighborhood, images incorrectly predicted as belonging to the *Cinco Conjuntos* neighborhood can be portrayed by the example that contemplates characteristics of a residential landscape, in this case, with aesthetic elements of the city's oldest residential typologies such as the plateau. An example of a prediction of the *Centro Histórico* as belonging to *Gleba Palhano* portrays a mixed landscape, with elements of verticalization and units of only story. The samples wrongly predicted from *Cinco Conjuntos* as belonging to the *Centro Histórico* were mainly images depicting typically commercial typologies, quite common in the city center. In contrast, representative samples from *Cinco Conjuntos* were interpreted as belonging to *Gleba Palhano,* in many cases, when urban voids were portrayed. *Gleba Palhano* was misclassified as *Centro Histórico,* mostly in images portraying commercial typologies and active façades. On the other hand, Gleba Palhano was predicted as *Cinco Conjuntos* in images of urban voids.

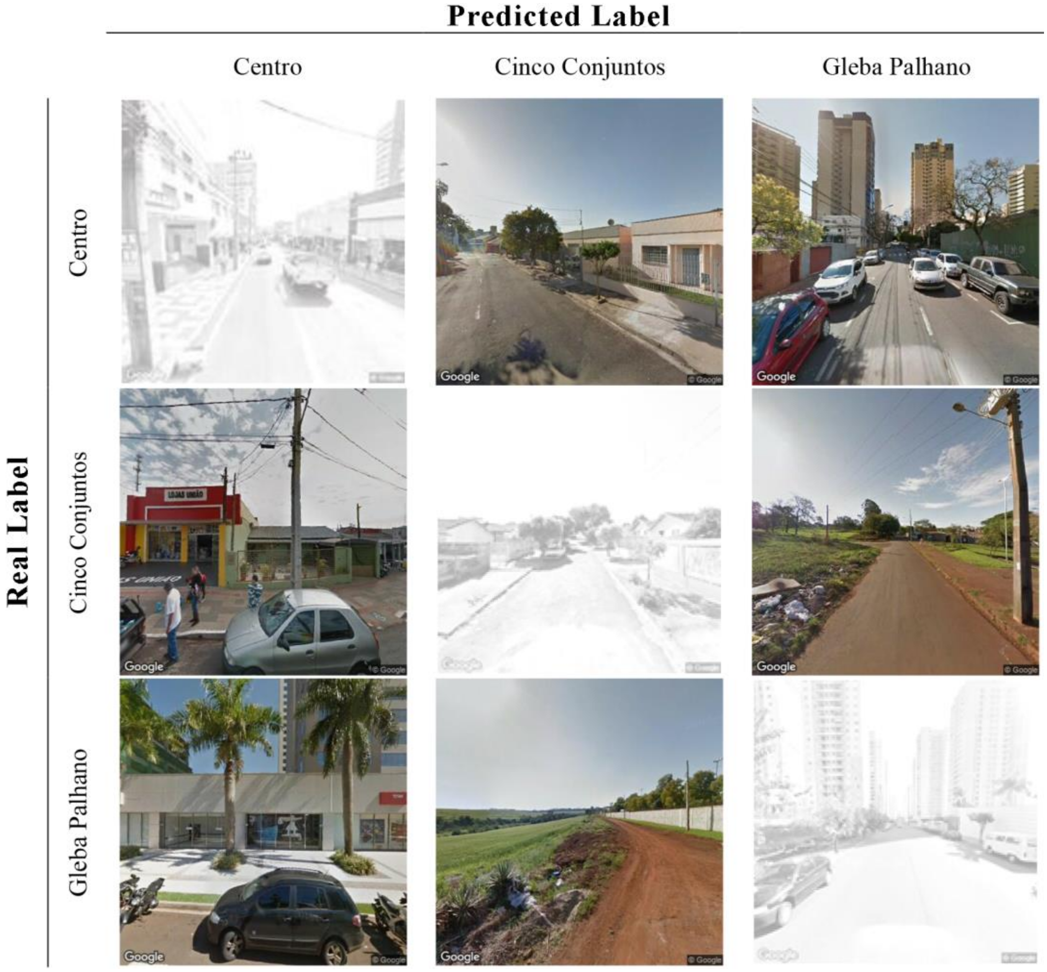


Figure 5. Images classified incorrectly (source: the authors, 2020).

Discussion

The results demonstrate the logicality in the use of machine learning to identify urban landscape characteristics. The model presented a higher predictive capacity when extracting the main components of North Zone of the city – an area of social interest housing projects – marked by the homogeneity of horizontal one-story houses. Such characteristics increase the proportion of sky in the images, and due to the peripheral location, the horizon line is constant.

On the other hand, in the city center, vertical areas are of more a compact configuration with land-use patterns of commercial or service buildings with the consequent presence of active facades. Such areas were identified in a coherently. Elements such as the number of cars, the presence of street parking spaces, and intense traffic signs seem to have contributed to the correct identification.

It can be pointed out that the samples that represent commercial buildings in the center and north areas were mostly correctly differentiated. This accurate insertion in their contexts indicates typological patterns of different commercial categories (vicinal or not) and a relational interpretation of the landscape composition between sectors.

Vegetation seems to be another important element for identifying urban landscapes when using computer vision. It is assumed that factors related not only to existence but also the types of vegetation, maturation, and maintenance of vegetated areas were relevant in the correct association of the analyzed areas.

Incorrect classifications were detected mainly in the Southwest sector of the city, the *Gleba Palhano* neighborhood. It is conjectured that this result was due to the region's recent growth, which is still under consolidation, thus sharing relevant characteristics with the other two neighborhoods. The existence of empty lots and old remnants of isolated buildings found in *Gleba Palhano* are also characteristic of *Cinco Conjuntos*. Further, the recent verticalization is a similarity to the *Centro* area. Therefore, it is understood that the inaccuracies resulted from the sector's consolidation process.

The passage of time and the maintenance of the urban environment’s physical elements are not always objectively accounted for in landscape evaluations, as they depend on the subjective perception/cognition of the researcher. However, when applying an image classification methodology, it is conjectured that temporality was effectively interpreted in image processing, both in vegetation and in the apparent detrition of building materials.

Classic readings of the environment, such as Christopher Alexander’s (1979), propose that elements such as buildings, walls, streets, and fences, form interrelated patterns. However, the identification of these patterns is marked by subjectivity. The identification of the landscape configuration and composition through the methodology applied here advances in the use of machine learning, through the classification of images as an innovative strategy that is parallel to traditional urban analyses. Computer vision implemented with the use of deep learning will allow expanding the analytical scales of the city, identifying its compositional and relational patterns.

Conclusions

Through the application of an identification procedure of configuration and composition of the urban landscape, through deep learning for image classification, the potential of this approach for identifying characteristics of the built environment uncovered. A practical approach was proposed to objectively identify and classify urban composition of neighborhoods with different socioeconomic levels, developmental periods, and, consequently, different landscapes. The model created obtained an accuracy of 87.6%, indicating a good generalization capacity. Thus, it is understood that the compositional characteristics of the different neighborhoods were effectively identified.

Study limitations include mainly, elements related to GSV images: (1) the camera perspective differs from that of a pedestrian since the images are obtained from the centerline of the street network. Thus, specific landscape elements can be obstructed, mostly by vehicles; (2) the images are not always recent; therefore, the possibility that changes in the landscape may have occurred cannot be ruled out. Nevertheless, the ease, speed, and low costs of using GSV imagery considerably overcome these limitations. It is also pointed out that, despite the selection of neighborhoods with different characteristics in different areas, the classification of neighborhoods in a single case study can reduce the possibility of generalizing the results. Future studies should seek applications for the identification of characteristics of the urban landscape in different scenarios

Despite these limitations, the obtained results are encouraging as proof of concept. The initial practical implication of models, such as proposed in this study, is to reduce the time required for data collection. From a theoretical perspective, scientific advances are made in the use of computational techniques for the objective understanding of urban qualities, which can influence human behavior, as well as social and economic elements. Future studies may explore more advanced techniques for interpreting the model's results. Further, the application of similar methodologies for the recognition of environmental characteristics related to urban meta-qualities such as walkability and vitality can be explored.

Notes

[1] https://opencv.org/
[2] https://pytorch.org/

References

Alexander, C. (1979). *The Timeless Way of Building*, 1º ed. Oxford University Press, New York.

Amorim, R. R., Oliveira, R. C. de. (2008). As unidades de paisagem como uma categoria de análise geográfica: o exemplo do município de São Vicente-SP. *Soc. Nat*. 20, 177–198.

Anderson, P., He, X., Buehler, C., Teney, D., Johnson, M., Gould, S. (2018). Bottom-Up and Top-Down Attention for Image Captioning and Visual Question Answering, in: *IEEE Conference on Computer Vision and Pattern Recognition* (CVPR). pp. 6077–6086.

Badland, H. M., Opit, S., Witten, K., Kearns, R. A., Mavoa, S. (2010). Can Virtual Streetscape Audits Reliably Replace Physical Streetscape Audits? J. *Urban Heal.* 87, 1007–1016. Available from: https://doi.org/10.1007/s11524-010-9505-x

Beidack, A. R. dos S., Fresca, T. M. (2011). Urban Restructuring and new centralities: a study about the north zone of Londrina – PR. *Bol. Geográfico* 29, 147–163. Available from: https://doi.org/10.4025/bolgeogr.v29i2.9898

Ben-joseph, E., Lee, J. S., Seoul, M., Cromley, E. K., Laden, F., Troped, P. J. (2015). Virtual and Actual: Relative Accuracy of On-Site and Web-based Instruments in Auditing the Environment for Physical Activity. *Heal. Place*. 19, 138–150. Available from: https://doi.org/10.1016/ j.healthplace.2012.11.001.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*, 1º ed. Springer Science + Business Media, New York, NY.

Burges, C. J. C. (2010). Geometric Methods for Feature Extraction and Dimensional Reduction - A Guided Tour, in: Maimon, Oded, Rokach, L. (Eds.), *Data mining and knowledge discovery handbook: a complete guide for practitioners and researchers*. Springer Science + Business Media, pp. 53–82. Available from: https://doi.org/10.1007/978-0-387-09823-4

Chan, D. M., Rao, R., Huang, F., Canny, J. F. (2018). t-SNE-CUDA: GPU-Accelerated t-SNE and its Applications to Modern Data. 2018 30th *Int. Symp. Comput. Archit. High Perform*. Comput. 330–338. Available from: https://doi.org/10.1109/ SBAC-PAD.2018.00060

Deng, J., Dong, W., Socher, R., Li, L., Li, K., Fei-fei, L. (2009). ImageNet: A large-scale hierarchical image database, in: *IEEE Conference on Computer Vision and Pattern Recognition.* Miami, FL, pp. 248–255. Available from: https://doi.org/10.1109/CVPR.2009.5206848

Doersch, C., Singh, S., Gupta, A., Sivic, J., Efros, A., Doersch, C., Singh, S., Gupta, A., Sivic, J., Efros, A., Makes, W., Look, P. (2012). What Makes Paris Look like Paris? *ACM Trans. Graph. (SIGGRAPH 2012)* 31. Available from: https://doi.org/10.1145/2185520.2185597

Ewing, R., Handy, S. (2009). Measuring the Unmeasurable: Urban Design Qualities Related to Walkability. *J. Urban Des.* 14, 65–84. Available from: https://doi.org/10.1080/13574800802451155

Fan, H., Zheng, L., Yan, C., Yang, Y. (2018). Unsupervised Person Re-identification: Clustering and Fine-tuning. *arXiv Preprint*. Available from: https://arxiv.org/abs/1705.10444

Gehrke, S. R., Wang, L. (2020). Operationalizing the neighborhood effects of the built environment on travel behavior. *J. Transp. Geogr.* 82, 12. Available from: https://doi.org/10.1016/j.jtrangeo.2019. 102561

Glorot, X., Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks, in: *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics (AISTATS)* 2010, Chia La- guna Resort, Sardinia, Italy., pp.249–256.

Gowda, S. N., Yuan, C. (2019). *ColorNet: Investigating the Importance of Color Spaces for Image Classification*, in: Jawahar C., Li H., Mori G., S.K. (Ed.), Computer Vision – ACCV 2018. Springer International Publishing, pp.581–596. Available from: https://doi.org/10.1007/978-3-030-20870-7

He, K., Sun, J. (2016). Deep Residual Learning for Image Recognition, in: *EEE Conference on Computer Vision and Pattern Recognition (CVPR)*. pp.1–9.

IBGE (2018). Panorama municipal: Londrina-Paraná [WWW Document]. URL https://cidades.ibge.gov.br/brasil/pr/londrina/panorama (accessed 3.26.18).

Kamalipour, H., Faizi, M., Memarian, G. (2014). Safe place by design: Urban crime in relation to spatiality and sociality. *Curr. Urban Stud.* 2, pp.152–162. Available from: https://doi.org/10.4236/cus.2014.22015

Kamilaris, A., Prenafeta-Boldú, F. X. (2018). Deep Learning in Agriculture: A Survey. *Comput. Electron. Agric.* 147, pp.70–90. Available from: https://doi.org/10.1016/j.compag.2018.02.016

Kent, J. L., Ma, L., Mulley, C. (2017). The objective and perceived built environment: What matters for happiness? *Cities Heal.* 8834, pp.1–13. Available from: https://doi.org/10.1080/23748834.2017.1371456

Kingma, D. P., Ba, J. L. (2014). Adam: A method for stochastic optimization. arXiv Preprint. pp.1–15. Available from: https://arxiv.org/abs/ 1412.6980

Lecun, Y., Bengio, Y., Hinton, G. (2015). Deep learning. *Nature* 521, pp.436–444. Available from: https://doi.org/10.1038/nature14539

Lee, S. M., Conway, T. L., Frank, L. D., Saelens, B. E., Cain, K. L., Sallis, J. F. (2017). The Relation of Perceived and Objective Environment Attributes to Neighborhood Satisfaction. *Environ. Behav.* 49, pp.136–160. Available from: https://doi.org/10.1177/0013916515623823

Litjens, G., Kooi, T., Bejnordi, B. E., Arindra, A., Setio, A., Ciompi, F., Ghafoorian, M., Laak, J. A. W. M., Van Der, Ginneken, B. Van, Sánchez, C. I. (2017). A survey on deep learning. *Medical image analysis 42*, pp.60–88. Available from: https://doi.org/10.1016/j.media.2017.07.005

Liu, L., Silva, E. A., Wu, C., Wang, H. (2017). A machine learning-based method for the large-scale evaluation of the qualities of the urban environment. *Comput. Environ. Urban Syst.* 65, pp.113–125. Available from: https://doi.org/10.1016/j.compenvurbsys.2017.06.003

Maaten, L. V. D., Hinton, G. (2008). Visualizing data using t-SNE. *Journal of machine learning research, 9*, pp.2579-2605.

Medeiros, F. F., Grigio, A. M. (2019). Identificação das Unidades Homogêneas e Padrão da Ocupação Urbana como subsídio ao ordenamento territorial em Mossoró, RN – Brasil. *EURE (Santiago)*, 45, pp.245–270.

Middel, A., Lukasczyk, J., Zakrzewski, S., Arnold, M., Maciejewski, R. (2019). Urban form and composition of street canyons: A human-centric big data and deep learning approach. *Landsc. Urban Plan.* 183, pp.122–132. Available from:https://doi.org/10.1016/j.landurbplan.2018.12.001

Moosavi, V. (2017). Urban morphology meets deep learning: Exploring urban forms in one million cities, town and villages across the planet. *arXiv Preprint.* arXiv:1709, 1–10. Available from: https://arxiv.org/abs/1709.02939.

Oura, K. (2006). Verticalização em Londrina - Paraná (1950-2005): A produção do espaço urbano e seu desenvolvimento pelos edifícios verticais. *Dissertação de Mestrado - Universidade* Presbiteriana São Paulo.

Poggio, T., Mhaskar, H., Rosasco, L., Miranda, B., Liao, Q. (2017). Why and When Can Deep – but Not Shallow – Networks Avoid the Curse of Dimensionality: a Review. *Int. J. Autom. Comput.* 14, pp.503–519. Available from: https://doi.org/10.1007/s11633-017-1054-2

Powers, D. M. W. (2011). Evaluation: from Precision, Recall and F-measure to ROC, Informedness, Markedness and Correlation. *J. Mach. Learn. Technol.* 2, pp.37–63.

Prefeitura do Município de Londrina (2020). Sistema de Informação Geográfica de Londrina – SIGLON [WWW Document]. Available from: http://www1.londrina.pr.gov.br/index.php?option=com_content&view=article&id=20114&Itemid=1988 (Acessado em 15 de janeiro de 2020).

Sallis, J. F., Cain, K. L., Conway, T. L., Gavand, K. A., Millstein, R. A., Geremia, C. M., Frank, L. D., Saelens, B. E., Glanz, K., King, A. C. (2015). Is Your Neighborhood Designed to Support Physical Activity? A Brief Streetscape Audit Tool. *Prev. Chronic*

Dis. Available from: https://doi.org/10.5888/pcd12.150098

Seresinhe, C. I., Preis, T., Mackerron, G., Moat, H. S. (2019). Happiness is Greater in More Scenic Locations. *Sci. Rep.* pp.1–11. Available from: https://doi.org/10.1038/s41598-019-40854-6

Shen, Q., Member, S., Zeng, W., Ye, Y., Stefan, M., Schubiger, S., Burkhard, R., Qu, H. (2018). StreetVizor : Visual Exploration of Human-Scale Urban Forms Based on Street Views. *IEEE Trans. Vis. Comput. Graph.* 24, pp.1004–1013. Available from: https://doi.org/10.1109/TVCG.2017.2744159

Shorten, C., Khoshgoftaar, T. M. (2019). A survey on Image Data Augmentation for Deep Learning. *J. Big Data.* Available from: https://doi.org/10.1186/s40537-019-0197-0

Takahashi, R., Matsubara, T., Uehara, K. (2019). Data Augmentation using Random Image Cropping and Patching for Deep CNNs. *ArXiv Preprint.* abs/1811.0, pp.1–16. Available from: https://arxiv.org/abs/1811.09030.

Tan, Y., Tang, P., Zhou, Y., Luo, W., Kang, Y., Li, G. (2017). Neurocomputing Photograph aesthetical evaluation and classi fication with deep convolutional neural networks. *Neurocomputing* 228, pp.165–175. Available from: https://doi.org/10.1016/j.neucom.2016.08.098

Töws, R. L., Mendes, C. M., Vercezi, J. T. (2010). The city as a business: the case from Londrina-PR and from Maringá-PR. *Bol. Geográfico* 28, pp.91–103.

Yamaki, H. T. (2017). *Terras do Norte: paisagem e morfologia*, 1 ed. Ed. H. Yamaki e UEL, Londrina.

Yin, L., Cheng, Q., Wang, Z., Shao, Z. (2015). 'Big data' for pedestrian volume: Exploring the use of Google Street View images for pedestrian counts. *Appl. Geogr.* 63, pp.337–345. Available from: https://doi.org/10.1016/j.apgeog.2015.07.010

Yin, L., Wang, Z. (2016). Measuring visual enclosure for street walkability: Using machine learning algorithms and Google Street View imagery. *Appl. Geogr.* 76, pp.147–153. Available from: https://doi.org/10.1016/j.apgeog.2016.09.024

Yin, R. K. (2001). E*studo de caso: Planejamento e Métodos*, 2º. ed. Bookman Companhia Editora, São Paulo.

Zhang, F., Zhou, B., Liu, L., Liu, Y., Fung, H. H., Lin, H., Ratti, C. (2018). Landscape and Urban Planning Measuring human perceptions of a large-scale urban region using machine learning. *Landsc. Urban Plan.* 180, pp.148–160. Available from: https://doi.org/10.1016/j.landurbplan.2018.08.020

Zhang, W., Li, W., Zhang, C., Hanink, D. M., Li, X., Wang, W. (2017). Parcel feature data derived from Google Street View images for urban land use classification in Brooklyn, New York City. *Data in Brief.* 12, pp.175–179. Available from: https://doi.org/10.1016/j.dib.2017.04.002

Zhou, B., Khosla, A., Lapedriza, A., Torralba, A., Oliva, A. (2016). *Places: An Image Database for Deep Scene Understanding*. J. Vis. 17, pp.1–12.

Tradução do título, resumo e palavras-chave

Identificação de composições da paisagem urbana: uma abordagem de deep learning.

Resumo. *A composição do ambiente pode exercer impactos sobre seus usuários, no entanto, esta relação permanece incerta até que as composições da paisagem urbana e suas qualidades espaciais possam ser analisadas empiricamente. Imagens obtidas através do Google Street View (GSV) possibilitam um grande volume de dados para avaliação automatizada das características ambientais. Técnicas de deep learning têm avançado na identificação de elementos compositivos do ambiente construído. Neste sentido, este estudo busca investigar e testar um procedimento de identificação da conFigureção e composição da paisagem urbana, por meio da classificação de imagens obtidas pelo GSV. A partir de um banco de imagens de três bairros de Londrina-PR, um modelo de deep learning para classificação de imagens foi proposto. O modelo obteve um bom desempenho, atribuindo corretamente 87,6% das amostras dos respectivos bairros do estudo de caso. Características compositivas foram empiricamente identificadas, considerando a distribuição das amostras no espaço de busca obtido. O modelo proposto contribui na definição de recortes espaciais bem como na mensuração de qualidades ambientais, otimizando coletas de dados, ampliando amostras e conferindo objetividade aos resultados. Esta abordagem contribui na expansão das escalas analíticas da cidade, identificando padrões compositivos e relacionais para o entendimento de elementos influentes no comportamento humano.*

Palavras-chave. *morfologia urbana, ambiente construído, aprendizado profundo, classificação de imagens, Google Street View.*

Editor responsável pela submissão: Julio Celso Borello Vargas.