MAPPING GATED COMMUNITIES IN BRAZIL USING GEOSPATIAL FEATURES AND A DEEP-LEARNING APPROACH

Author:
Julia Michalski Linhares
Master's Student, Graduate Program in Economics (PPGE/FEARP-USP)

Submission Area:
Spatial Data Analysis and Spatial Econometrics

JEL Classification:
R14

Ribeirão Preto, SP – July 25, 2025

RESUMO

A rápida expansão de condomínios horizontais fechados no Brasil superou a disponibilidade de dados espacialmente explícitos, dificultando a avaliação rigorosa de sua distribuição e tendências tipológicas. Este estudo apresenta o primeiro banco de dados georreferenciado desses empreendimentos na região Sudeste do país e propõe uma metodologia em duas etapas que combina extração de características vetoriais com redes neurais convolucionais (CNNs). Na primeira etapa, foram coletados e filtrados polígonos do OpenStreetMap — aplicando tags de barreira e uso do solo, exclusões baseadas em nomes, limiares de densidade e inferência de vias privadas via envoltória convexa — para delinear 7955 possíveis condomínios fechados. Um CNN treinado com 300 amostras rotuladas manualmente alcançou 85% de acurácia na validação do mapeamento. Na segunda etapa, sobrepomos uma grade de 50m × 50m sobre esses polígonos, foi extraído recortes de imagens de alta resolução, pré-processamos e aumentamos o conjunto de dados para 300 imagens, e realizado fine-tunning de modelos pré-treinados (EfficientNetB0, ResNet50, MobileNetV2, VGG16, InceptionV3). A CNN preliminar atingiu 97% de acurácia ao distinguir “casa” de “prédio”, mas ainda é necessário treinamentos futuros com maior base de dados. Esse fluxo de trabalho reprodutível e o produto geoespacial resultante fornecem a planejadores urbanos, formuladores de políticas públicas e pesquisadores uma ferramenta essencial para monitorar o crescimento de condomínios fechados.

KEY-WORDS:
Condomínios Horizontais Fechados. Rede Neural Convolucional. Classificação de imagens.

ABSTRACT

The rapid expansion of gated horizontal condominiums in Brazil has outstripped the availability of spatially explicit data, impeding rigorous assessment of their distribution and typological trends. This study delivers the first georeferenced database of such developments for the country's Southeast region and proposes a two-stage methodology that melds vector-based feature extraction with convolutional neural networks (CNNs). In the first stage, we harvest and filter OpenStreetMap polygons—applying barrier and land-use tags, name-based exclusions, density thresholds, and private-road inference via convex hulls—to delineate 7955 candidate gated communities. A CNN trained on 300 manually labeled samples achieves an 85% mapping accuracy upon validation. In the second stage, we overlay a 50m × 50m grid on these parcels, extract high-resolution image patches, preprocess and augment the dataset to 300 images, and fine-tune pre-trained backbones (EfficientNetB0, ResNet50, MobileNetV2, VGG16, InceptionV3). The preliminary CNN attains 97% accuracy in distinguishing "house" versus "building," but in need for further tests with bigger datasets. This reproducible pipeline and resulting geospatial product furnish planners, policymakers, and researchers with a vital tool for monitoring gated community growth, infrastructure demands, and socio-spatial equity.

KEY-WORDS:
Gated communities. Convolutional Neural Network. Image-level Classification.

SUMÁRIO

1 Introduction

Urban planning has become an increasingly relevant issue due to its central role in mitigating the effects of climate change and promoting balanced socio-economic indicators. In this context, it has emerged as a key subject of public policies and a matter of interest in global governance. Since 1976, the United Nations (UN) has held, on a vicennial basis, the Habitat Conference, aimed at discussing urban transformations in the context of globalization, as well as forms of housing and development. The third edition of the event, Habitat III, held in 2016 in the city of Quito, culminated in the adoption of the New Urban Agenda (NAU) (United Nations, 2017), a guiding instrument for sustainable urban development over the subsequent two decades.

In Brazil, the issue of urban planning acquired normative contours with the 1988 Federal Constitution, which, in Article 182, established the mandatory implementation of urban development policies, to be carried out through master plans elaborated by municipalities. This guideline was later regulated by the Estatuto da Cidade (City Statute, Law No. 10.257/2001) (Brasil, 2001), which defined principles, guidelines, and instruments of urban policy, including the obligation to adopt master plans in municipalities with populations exceeding twenty thousand inhabitants. The creation of the Ministry of Cities in 2003 represented another significant institutional milestone, as it centralized the coordination of public policies directed toward urban development.

In this context, between 2015 and 2019, there was a 19% increase in urbanized areas in Brazil (IBGE, 2022), as well as a 6.5% population growth according to the most recent census, compared to the 2010 census (IBGE, 2011), highlighting a period of intense demographic and spatial transformation.

Among the recent urban trends, gated communities, characterized by low-density private residential layouts, are an increasingly widespread urban phenomenon worldwide, particularly in Latin America (Kostenwein, 2021). Demographic Census data show that the number of houses located in gated communities or residential complexes in Brazil exceeded 1.7 million in 2022, accounting for 2.45% of all dwellings in the country. In 2010 this figure was just over 1 million, representing 1.78% of residences. These numbers indicate that such developments have gained a larger share of Brazil's housing stock over the period (IBGE, 2011, 2023). Despite the availability of Census data on the number of dwellings located within gated villages or condominium complexes, and notwithstanding several case-specific studies that have manually quantified these developments in individual municipalities, there exists no comprehensive geolocated database of such residential estates.

Such a geolocated database of gated residential developments is indispensable for both urban research and public policy. By systematically mapping the location, extent, and typology of these low-density housing complexes, we provide a foundational resource for analyzing their spatial distribution and growth dynamics. Planners and policymakers can leverage this dataset to assess infrastructure needs, evaluate the impacts of gated communities on social equity and urban form, and guide land-use regulations. Moreover, researchers in geography and urban studies will benefit from a standardized, reproducible reference for examining how these developments interact with transportation networks, service provision, and environmental constraints. Finally, by coupling this dataset with machine-learning methods such as convolutional neural networks, we enable scalable and automated classification of new areas—thus ensuring that decision-makers have up-to-date, high-resolution information to inform sustainable and inclusive urban development.

Thus, the primary objective of this study is to create a pioneering geolocated dataset of these housing developments. Accordingly, the work is structured around two main pillars:

1. **OSM features classification**: To identify the location and size of existing gated communities in Brazil.

2. **Convolutional neural network**: To enhance accuracy of mapping using image-level classification.

2 Literature Review

Gated communities emerged in the United States in the late twentieth century, particularly in suburban areas, and similar urban patterns soon appeared in developing countries in Asia and Africa (Bagaeen & Uduku, 2010). In Brazil, gated communities are likewise a comparatively recent phenomenon, beginning in 1970 with Alphaville São Paulo (do Rio Caldeira, 2000), despite the lack of comprehensive national data, localized studies have documented substantial growth since then (Ivo, 2012; Manhães & Arruda, 2018; Silva & Lopes, 2022).

Understanding why such developments continue to proliferate also requires an examination of the motivations behind residential location choices, particularly those that drive demand for gated communities. According to (Nechyba & Walsh, 2004), the factors influencing residential location choices can be classified into those that push people away from city centers and those that attract them to suburban areas. Applied to gated communities, the primary "push" factor is crime (Ivo, 2012; Manhães & Arruda, 2018; Silva & Lopes, 2022), as these developments offer a perceived sense of security. Another relevant factor is the desire to escape unsustainable urban growth (Silva & Lopes, 2022), often associated with the search for a better quality of life. Among the "pull" factors, the desire for green spaces stands out (Manhães & Arruda, 2018), since these residential areas often provide greater contact with nature.

However, this expansion is not without consequences. As several authors have noted, the proliferation of gated communities has significant implications for urban form and equity. The arise of gated communities contributes to urban dispersion, resulting in cities that are functionally fragmented and socially segregated (Rueda, 1997). Sposito, 2016 discusses the transportation challenges in dispersed Brazilian cities, emphasizing the poor quality of urban and interurban roads, which disproportionately affects low-income populations dependent on public transportation and limits their access to urban cores. In the same vein, Basso et al., 2017 in a study of Cascavel, Paraná, concludes that these gated communities harm the road network by fostering urban segregation and breaking up the landscape, and argues that urban planning should be reoriented toward the construction of more democratic cities.

Manhães e Arruda, 2017 examines the rise of gated communities in Campos dos Goytacazes, Rio de Janeiro, and how they reshape urban spatial configurations, intensifying sociospatial segregation and fragmentation in a setting where the absence of targeted legislation and effective oversight allows interested actors to manipulate municipal actions. The study draws upon a data collection of the Municipal Secretariat of Works and Urban Planning of the Campos dos Goytacazes City Hall to establish the condominium mapping, data that are not available across other municipalities.

3 Data and Methodology

The Brazilian Census includes a variable indicating dwelling type, one of whose categories is “villa house or condominium.” Although this information permits analysis of the growth of such developments nationwide, the Census does not report their total number or geographic locations—only the count of households falling into this category. As shown in Figure 1, these developments have increased over time, particularly in medium- and large-sized municipalities, as noted in the literature reviewed in Section 2.

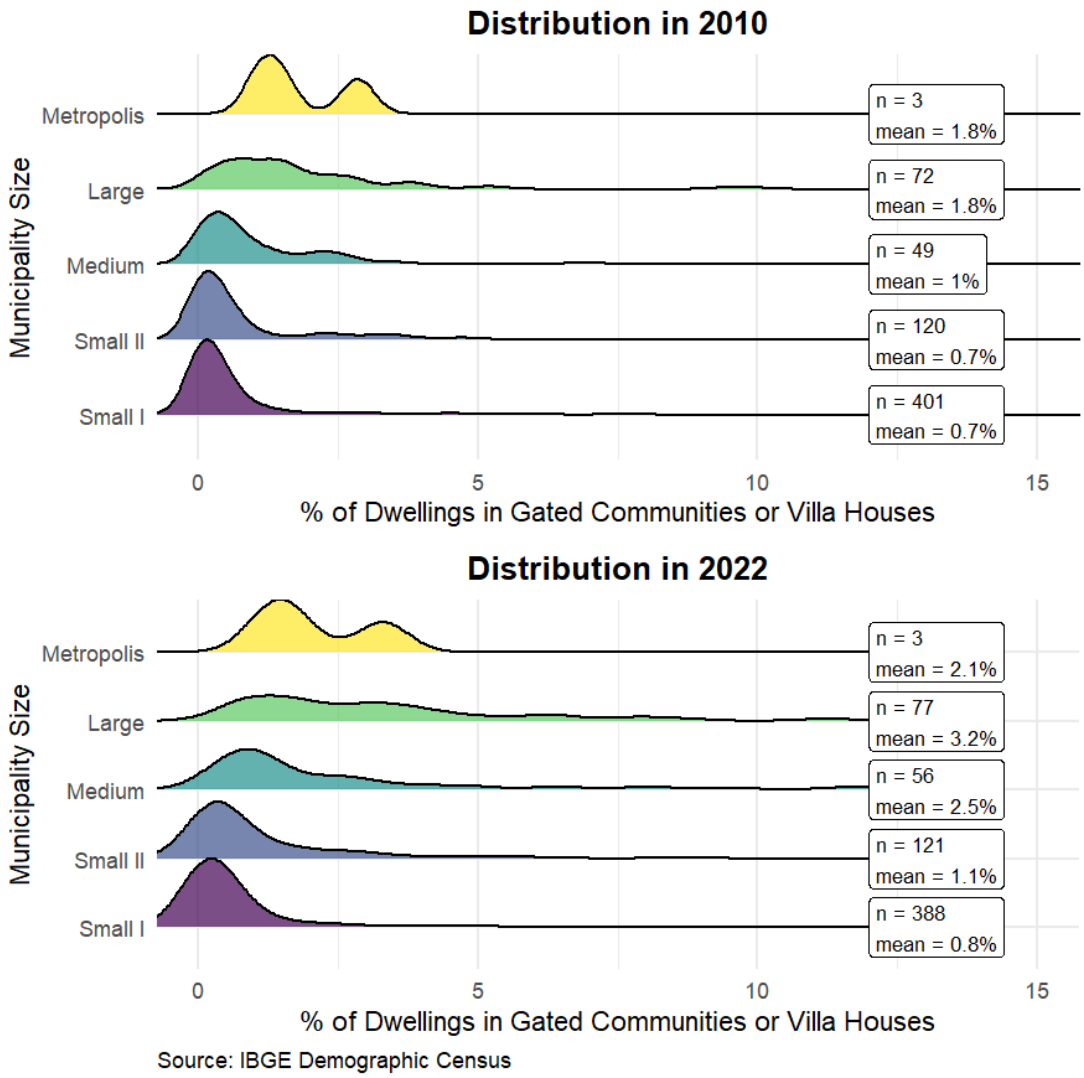

Figura 1: Evolution of Gated Community Dwelling Share by Municipality Size (2010 vs. 2022)

As outlined in Section 1, the aim of this study is to develop a pioneering database on gated horizontal communities.We chose to focus exclusively on horizontal condominiums because they exhibit a higher incidence of the issues discussed in Section 2. Currently, no national database exists that provides the number and spatial distribution of these developments, although some localized studies have mapped them in selected municipalities. As a result, the most critical component of the project is the accurate spatial and temporal mapping of gated communities throughout the Southeastern region, with a pioneering dataset. To this end, the work has been structured into two parts, each involving specific technical challenges and corresponding methodological solutions. The following subsections describe the tools and methods employed at each stage of the study.

3.1 OSM FEATURES CLASSIFICATION

Gatted communities typically exhibit a distinctive combination of spatial characteristics, like low building density, curvilinear street layouts, perimeter walls, limited connections to surrounding road networks, and a single main entrance gate—that facilitate their systematic identification and mapping. These hallmarks serve as the foundational criteria for our initial gated-community delineation.

To construct the dataset of gated communities, the initial step involved extracting polygon features from OpenStreetMap (OSM) that represent residential areas. The use of OSM is supported by a growing body of literature Chicombo, 2021; De Andrade et al., 2019; Pereira et al., 2022, particularly due to its open-access geospatial data and its comprehensive classification of territorial elements such as road types, land use, and buildings. These attributes enable a broad and consistent identification of gated communities across all urban areas represented and regularly updated on the platform. To retain only relevant features, multiple filters were applied: selected polygons were required to be tagged with land use = residential and barrier = wall, and to lack conflicting tags indicating public amenities or non-residential functions. Further refinement was based on the polygon names, excluding those that contained terms such as "Flat," "Building," or "Favela," which typically characterize vertical or informal developments not consistent with the definition of gated communities.

Subsequently, the dataset was refined to include only horizontal residential developments. This entailed removing polygons that were located within dense clusters of similarly sized structures, a spatial configuration that generally indicates apartment buildings. A minimum area threshold was also applied to eliminate small parcels typically associated with vertical developments, thereby ensuring that the dataset focused on spatial units with sufficient territorial extension.

To complement this filtered dataset, additional candidates for gated communities were identified by detecting clusters of connected private roads, using lines shapefile data extracted from OpenStreetMaps. These road segments, tagged with restricted access classifications that are caracteristic of gatted communities, such as access = private or access = destination, were grouped based on spatial connectivity and converted into estimated residential polygons using convex hull operations. This approach allowed the inclusion of gated communities that may not have been directly mapped as polygons in OSM but whose spatial layout could be inferred from road infrastructure.

3.2 CONVOLUTIONAL NEURAL NETWORK

Even after filtering based on vector-derived features, the mapping accuracy remains moderate due to the persistent misclassification of high-rise buildings. To address this limitation, we employ satellite image classification using convolutional neural networks (CNNs) to distinguish between low-density housing developments and vertical apartment complexes. CNNs are a specialized class of deep learning architectures designed to process grid-structured data (Goodfellow et al., 2016), making them particularly well-suited for image and video recognition tasks by automatically learning hierarchical feature representations from raw pixel values.

The capacity of convolutional neural networks to autonomously learn salient contextual features constitutes their principal advantage. Furthermore, CNNs frequently outperform traditional classifiers, such as Random Forests and Support Vector Machines, in both accuracy and scalability when applied to large geospatial datasets (Maggiori et al., 2016). However, these benefits are tempered by significant drawbacks: CNNs demand substantial computational resources and are prone to overfitting, whereby models achieve high accuracy on training data yet fail to generalize to novel samples (Pritt & Chern, 2017).

When applied to satellite imagery, deep learning can encounter challenges related to the standardization of input tiles, where important features may be lost during resizing and normalization, and the need for very large, labeled datasets to effectively train complex models (Pritt & Chern, 2017). Nonetheless, deep-learning approaches have demonstrated considerable success in a variety of remote-sensing tasks, face detection (Schroff et al., 2015) and object recognition (Redmon & Farhadi, 2017).

The underlying premise is that, if a human can inspect a satellite image and determine whether a given structure is a house or a multi-story building, then the same discriminative visual cues can be encoded and learned by a deep-learning model, thereby automating the classification process.

In this second stage, as illustrated in Figure 2, a regular 50m × 50m grid of square polygons was overlaid onto the initial mapped layer, and only the grid cells intersecting the previously delineated condominium polygons were retained. Each grid cell was exported as a raster image, thereby producing a comprehensive database of house- and building-labeled image patches. These selected image patches were then used as input to the CNN classifier to determine whether each patch corresponded to a “house” or a “building".

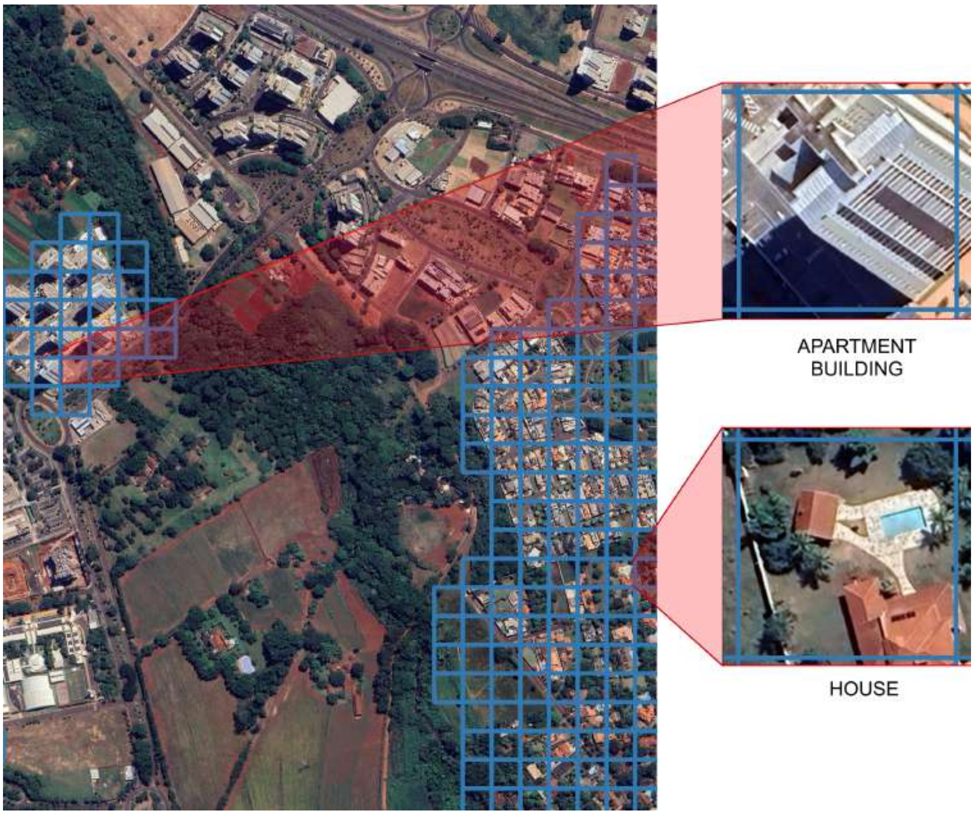

Figura 2: 50m x 50m grid atop mapped gatted communities in the city of Ribeirão Preto - São Paulo

Prior to commencing the analysis, it is essential to perform image preprocessing to remove noise and normalize the data, this step is indispensable to optimize model performance and enhance robustness (Maharana et al., 2022). The preprocessing pipeline comprises the following steps:

- **Image resizing and normalization**: all patches are resized to a fixed resolution (e.g. 256 × 256px) and pixel values scaled to a common range (e.g. [0,1] or [–1,1]).
- **Data augmentation**: generate additional training samples by applying random rotations, flips, brightness and contrast adjustments, and other perturbations to increase model robustness and help prevent overfitting.
- **Contrast enhancement**: apply histogram equalization to improve local contrast and highlight structural details, for this is necessary to apply a gray scale.

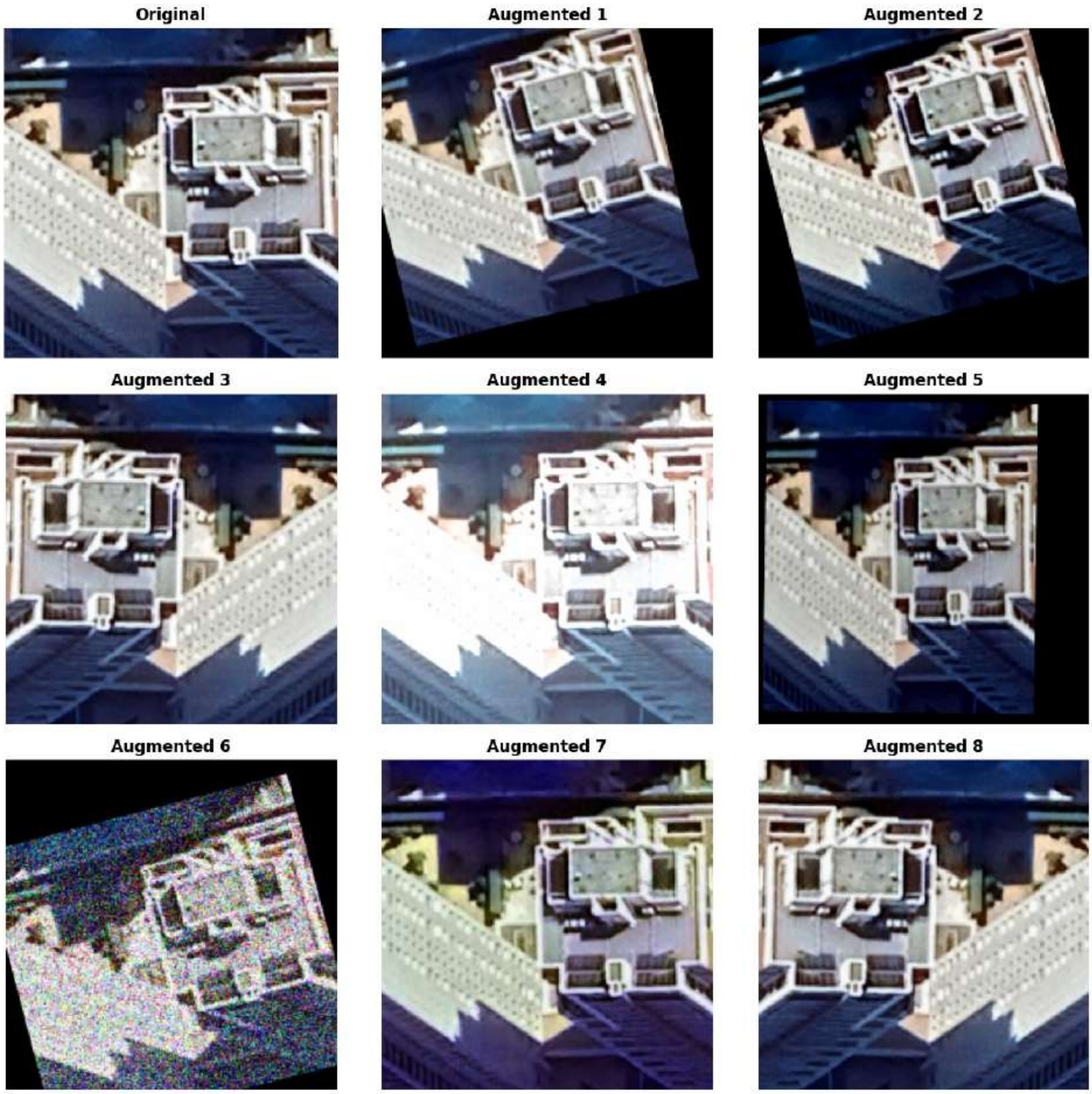

Figura 3: Example of data augmentation

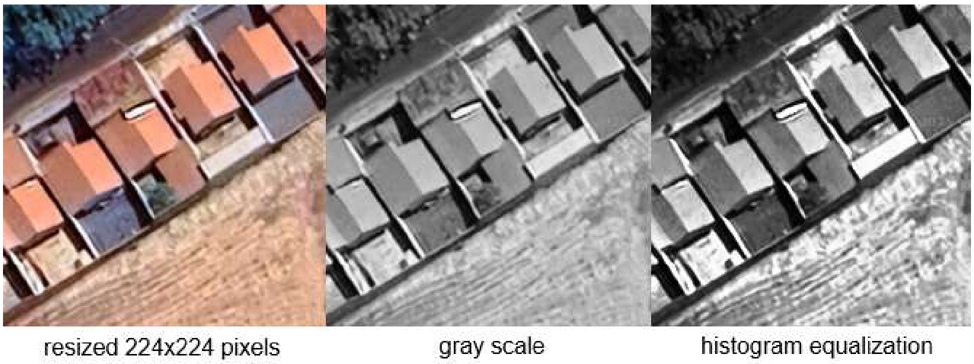

Figura 4: Preprocessing steps

Following preprocessing, the resulting image patches must be partitioned into distinct training and testing subsets to ensure an unbiased evaluation of our convolutional neural network. We employ a stratified random split reserving 70% of the data for model training and 30% for testing—so that each subset maintains the same proportion of “house” and “building” classes.

Our CNN fine-tune existing architectures: EfficientNetB0, ResNet50, MobileNetV2, VGG16, and InceptionV3, each pre-trained on large benchmark datasets. This approach is essential because it capitalizes on representations already learned by these models, allowing us to adapt them to the specific task of distinguishing “house” versus “building” with far less labeled data and computational effort than would be required to train a network from scratch. By initializing our networks with pre-existing weights, we accelerate convergence and reduce training time.

For the initial analysis, we used 150 image patches, with 84 showing houses and 66 showing buildings. After preprocessing and data augmentation, which doubled each class, the final set contained 300 images. All stages of the pipeline were implemented in Python, employing the Keras library for construction and training of the convolutional neural networks and the scikit-learn framework for data preprocessing, model evaluation, and utility functions (Chollet, 2015; Pedregosa et al., 2011).

4 Preliminary Results

At the conclusion of the mapping exercise, 7,955 gated communities were identified across the four states. To evaluate the accuracy of this mapping, a manual validation was undertaken: 100 mapped gated communities were randomly selected and verified, with the aid of Google Maps satellite imagery, to confirm both their existence and their correct classification. The validation showed that 85 communities were accurately mapped, while 15 were false positives—mainly vertical condominiums and neighborhoods in small towns—yielding an overall accuracy rate of 85 %. Figure 5 represents an example of the mapping section for the city of Ribeirão Preto, in São Paulo.

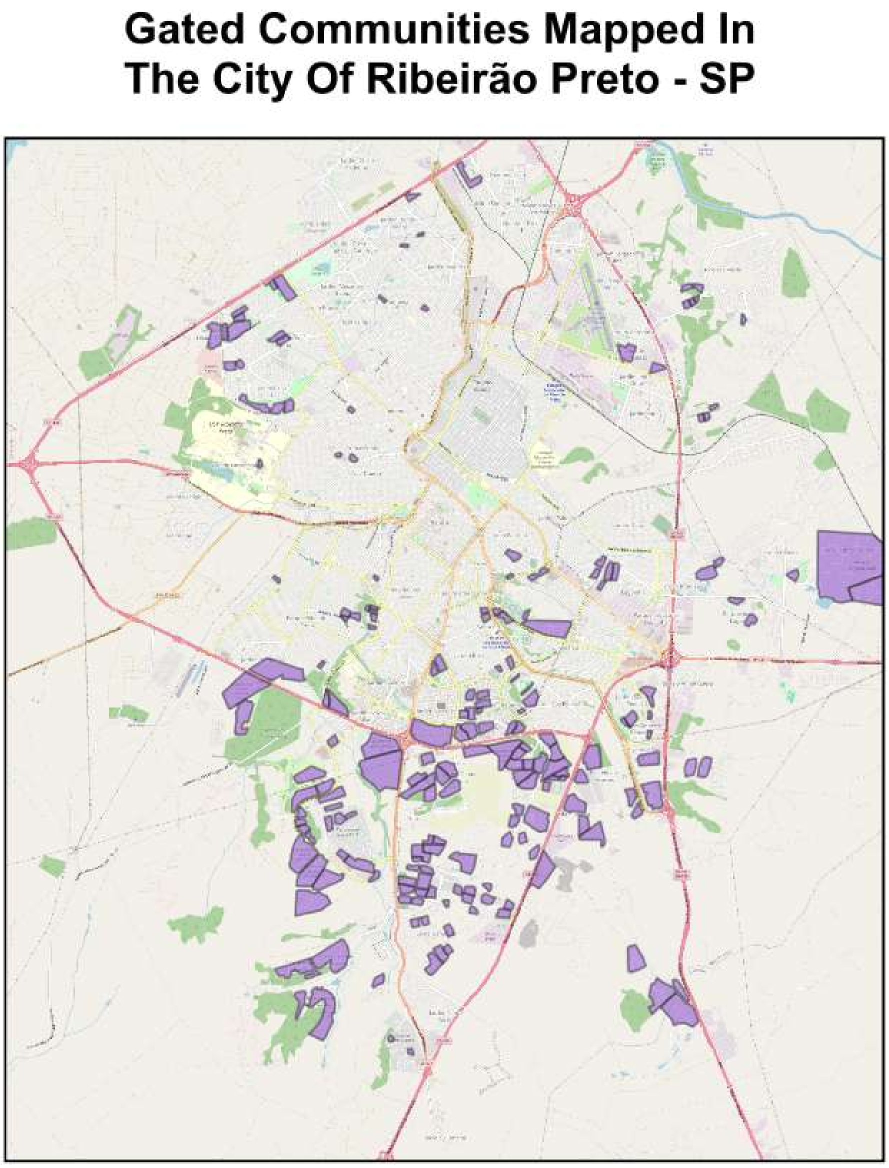

Figura 5: Gated Communities in Ribeirão Preto

Note: Figure produced using mapped dataset and QGIS software. The figure shows, for the city of Ribeirão Preto, the constructed gatted communities in purple polygons.

This misclassification of vertical condominiums was anticipated, as only a small subset of polygons carried explicit labels denoting “house” or “building.” Given that both horizontal and vertical developments share common attributes, such as enclosing walls and internal private-

Figura 6: Model Validation

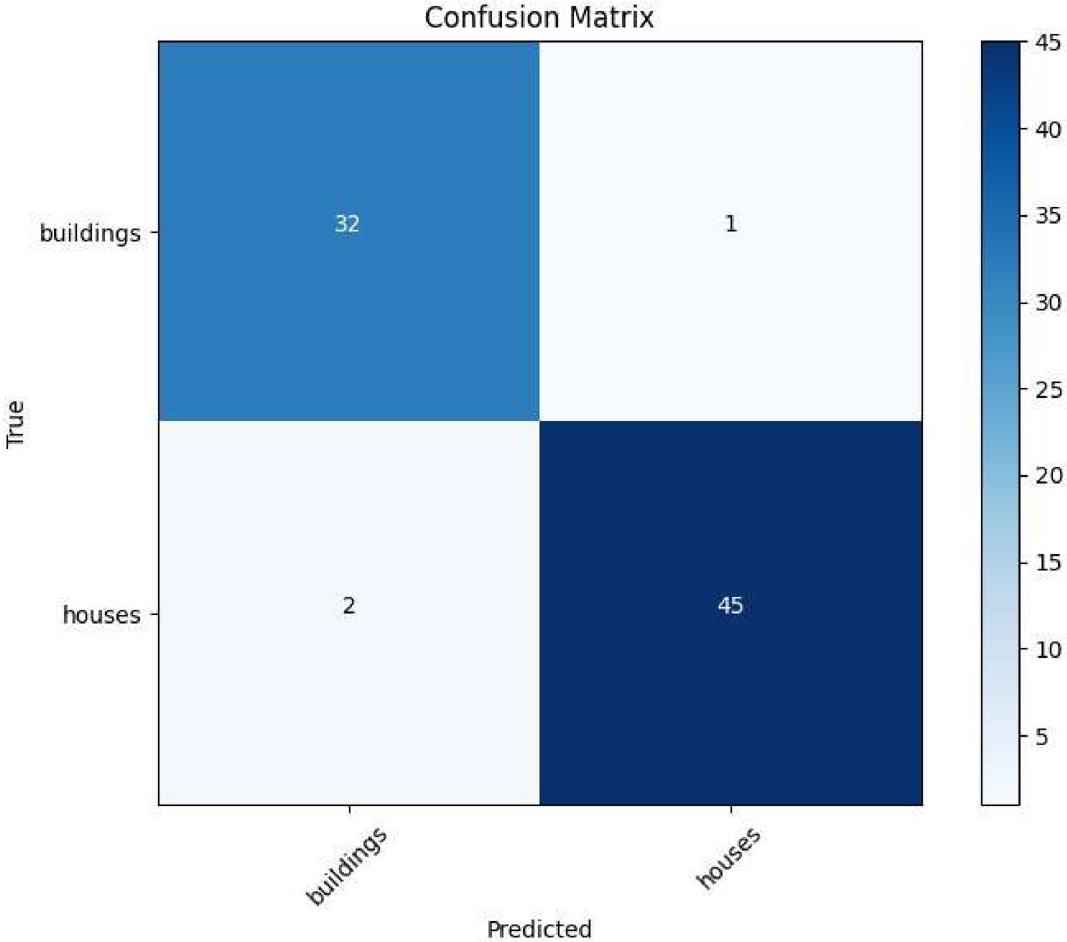

Figura 7: Confusion Matrix done by the CNN

access roads, relying solely on these labels inevitably leads to ambiguity. To overcome this limitation, we therefore employed the CNN, which can learn more nuanced visual and contextual features to distinguish between low-rise and multi-story residential complexes.

The preliminary CNN results indicate an accuracy of 97% in image classification, which, despite being high, is based on a small sample size and therefore may not be fully reliable. Nonetheless, the model demonstrates strong performance in identifying houses, achieving especially high precision for those with distinctive roofing features, such as tiled roofs, as shown in Figure 6

5 Conclusion and Next Steps

The initial geoprocessing workflow achieved an overall mapping accuracy of 85%, but horizontal condominiums were often misclassified as vertical developments. To address this limitation, our first CNN, trained on a relatively small image dataset, attained a classification accuracy of 97% when distinguishing houses from multi-story buildings. This preliminary performance, albeit good, is in need to expand the training set, a goal that can be met by leveraging the extensive polygon inventory generated during initial mapping. Once the CNN has been retrained on a larger and more diverse collection of image patches, and higher accuracies are achieved, a post-processing stage will be introduced. In this phase, all patches that the CNN confidently labels as buildings will be cross-referenced against the original condominium layer to exclude any parcels with a high probability of containing vertical structures. By filtering out these false positives, we anticipate improving the overall mapping accuracy beyond the initial 85% threshold.

Furthermore, we plan to scale the project to encompass gated communities across the entire Brazilian territory, thereby enhancing the CNN's accuracy and effectively establishing the first national database of this residential development type.

This newly constructed database represents the most comprehensive and systematic effort to date to document the presence and expansion of gated communities across the Southeastern region. It includes georeferenced polygons for each identified development. The dataset not only fills a significant empirical gap in the national urban data landscape but also creates opportunities for further research and informed policymaking in areas such as land use regulation, infrastructure planning, and socio-spatial segregation.

References

Bagaeen, S., & Uduku, O. (2010). Gated Histories: An Introduction to Themes and Concepts. Em *Gated communities: Social sustainability in contemporary and historical gated developments* (pp. 1–8). Routledge.

Basso, M. H., Simonetto, F. G., Lima, J. G. d., Ferrarini, L. C., & Müller, P. (2017). A influência de condomínios horizontais fechados no planejamento municipal e dinâmica urbana. *Anais do 15º Encontro Científico Cultural Interinstitucional e 1º Encontro Internacional*.

Brasil. (2001). Lei n.º 10.257, de 10 de julho de 2001 — Estatuto da Cidade [Regulamenta os arts. 182 e 183 da Constituição Federal, estabelece diretrizes gerais da política urbana e dá outras providências.].

Chicombo, T. F. (2021). Proposta de Identificação de Subcentros Urbanos com Base na Distribuição Espacial das Funções: caso da Cidade de Maputo-Moçambique. *Sociedade & Natureza*, *33*, e57722.

Chollet, F. (2015). *Keras: The Python Deep Learning API* [Accessed: 2025-07-26]. Keras Team. https://keras.io/

De Andrade, D. J., De Souza, A. A. M., & Gomes, C. (2019). Análise temporal da expansão urbana nos municípios do Vale do Paraíba Paulista. *Mercator (Fortaleza)*, *18*, e18005.

do Rio Caldeira, T. P. (2000). *Cidade de muros: crime, segregação e cidadania em São Paulo*. Editora 34.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning* [Using Chapter 9: Convolutional Networks]. MIT Press. http://alvarestech.com/temp/deep/Deep%20Learning%20by%20Ian%20Goodfellow,%20Yoshua%20Bengio,%20Aaron%20Courville%20(z-lib.org).pdf

IBGE. (2011). Censo Demográfico 2010: Resultados do Universo - Características da População e dos Domicílios [Acesso em: 25 maio 2025].

IBGE. (2022). *Áreas urbanizadas do Brasil : 2019* (IBGE, Ed.) [Disponível somente em meio digital.].

IBGE. (2023). Censo Demográfico 2022: Características dos Domicílios - Resultados preliminares da Amostra [Acesso em: 25 maio 2025].

Ivo, A. B. L. (2012). Jardins do Éden: Salvador, uma cidade global-dual. *Caderno CRH*, *25*(64), 131–146. https://doi.org/10.1590/s0103-49792012000100010

Kostenwein, D. (2021). Between walls and fences: How different types of gated communities shape the streets around them. *Urban Studies*, *58*(16), 3230–3246.

Maggiori, E., Tarabalka, Y., Charpiat, G., & Alliez, P. (2016). Convolutional neural networks for large-scale remote-sensing image classification. *IEEE Transactions on geoscience and remote sensing*, *55*(2), 645–657.

Maharana, K., Mondal, S., & Nemade, B. (2022). A review: Data pre-processing and data augmentation techniques. *Global Transitions Proceedings*, *3*(1), 91–99.

Manhães, V. R. R., & Arruda, A. P. S. N. d. (2017). Os condomínios horizontais fechados sob a ótica da legislação Federal e Municipal [Artigo disponível em PDF. Cidade de estudo: Campos dos Goytacazes/RJ]. *GT2 - Cidades, Política Urbana e Processos Sociais*.

Manhães, V. R. R., & Arruda, A. P. S. N. d. (2018). A expansão dos condomínios fechados: o caso de Campos dos Goytacazes, RJ. *Cadernos do Desenvolvimento Fluminense*, (12), 23–36. https://doi.org/10.12957/cdf.2017.36733

Nechyba, T. J., & Walsh, R. P. (2004). Urban Sprawl. *Journal of Economic Perspectives*, *18*(4), 177–200. https://doi.org/10.1257/0895330042632681

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python [Accessed: 2025-07-26]. *Journal of Machine Learning Research*, *12*, 2825–2830. https://scikit-learn.org/

Pereira, R. H. M., Parga, J. P., Saraiva, M., Bazzo, J. P., Tomasiello, D. B., Silva, L. P., Nadalin, V. G., & Barbosa, R. (2022). TD 2802 - Forma urbana e mobilidade sustentável: evidências de cidades brasileiras. *Texto para Discussão*, 1–66. https://doi.org/10.38116/td2802

Pritt, M., & Chern, G. (2017). Satellite image classification with deep learning. *2017 IEEE applied imagery pattern recognition workshop (AIPR)*, 1–7.

Redmon, J., & Farhadi, A. (2017). YOLO9000: better, faster, stronger. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 7263–7271.

Rueda, S. (1997). La ciudad compacta y diversa frente a la conurbación difusa. *Ciudades para un futuro más sostenible*, *19*(01), 69–83.

Schroff, F., Kalenichenko, D., & Philbin, J. (2015). Facenet: A unified embedding for face recognition and clustering. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 815–823.

Silva, G. C., & Lopes, W. G. R. (2022). Aspectos Gerais sobre Condomínios Horizontais e Loteamentos Fechados. Em *Os desafios das políticas sociais, inclusão e o trabalho em rede na promoção da qualidade de vida* (pp. 190–210). EPITAYA. https://doi.org/10.47879/ed.ep.2022380p190

Sposito, M. E. B. (2016). Urbanização difusa, cidades médias e novos habitats urbanos. Em *Espaços fechados e cidades-Insegurança urbana e fragmentação socioespacial* (pp. 41–60). Editora Unesp.

United Nations. (2017). New Urban Agenda [Adopted at the United Nations Conference on Housing and Sustainable Urban Development (Habitat III), Quito, Ecuador, 17–20 October 2016].