



Semi-Automatic Satellite Image Segmentation for Urban Expansion Analysis in Santa Cruz Island, Galápagos

Joan Daniel Rivas-Andrade¹, Paulina Vizcaino-Imacaña², Ivan Reyes-Chacón²,
and Manuel Eugenio Morocho-Cayamcela^{*1,2}

¹ School of Mathematical and Computational Sciences, DeepARC Research Group, Yachay Tech University, Had. San José s/n y Proyecto Yachay, Urcuquí 100119, Ecuador
{joan.rivas, mmorocho}@yachaytech.edu.ec

² Faculty of Digital Engineering and Emerging Technologies, International University of Ecuador (UIDE), Quito 170411, Ecuador
{pvizcaino, ireyes, mamorochoca}@uide.edu.ec

Abstract. This paper presents a semi-automatic workflow for transforming multitemporal satellite imagery into interpretable geospatial communication outputs about urban expansion in Santa Cruz Island, Galápagos. The work is positioned as an applied communication technology because it converts visually complex remote-sensing data into binary masks, comparative maps, and quantitative indicators that can support environmental reporting, public communication, and decision-making in a sensitive island territory. Urban masks in the first iteration were produced with HSV-based segmentation, manually post-processed using a Python correction tool, and then fine-tuned over an ImageNet-pretrained EfficientNet-B0 encoder. For the rapid experiment, five image-mask pairs were selected and split into four training pairs and one validation pair. The best validation IoU: 0.6544, and F1 measure score: 0.7911 Based on the comparisons of urban pixel areas between 2013 and 2024, the predicted net increases in urban pixel area grew from a mere 5.66 million pixels to a substantial total of 14.43 million pixels, implying a relative net change of 154.76%. Results are reported as model-based pixel-area indicators because the current images are not georeferenced orthomosaics. Beyond the technical implementation, the study contributes a reproducible pipeline for visually communicating urban transformation in contexts where official spatial datasets may be limited.

Keywords: Geospatial communication, Satellite imagery, Urban expansion, Semantic segmentation, U-Net, Santa Cruz Island, Galápagos

1 Introduction

Urban expansion in environmentally sensitive territories is not only a spatial process but also a communication challenge. Local authorities, researchers and residents as well as conservation stakeholders usually require intuitive visual proof on where settlements

are rising and how this expansion relates to protected areas, road infrastructure or tourism pressure. This challenge is crucial, for example in Santa Cruz Island (Galápagos) where urban growth occurs close to protected ecosystems and agricultural areas as well as touristic services. Rather than directly detecting population from images, this work approximates urban expansion by segmenting visible buildings and built structures in satellite captures.

Within communication and applied technologies, maps, overlays, and geovisualizations function as mediated representations that translate complex territorial processes into interpretable visual messages. Cartographic and geovisualization theory has long emphasized that maps are not neutral pictures of space; they are designed representations that structure perception, attention, and interpretation [1, 2]. In that sense, the present study treats satellite-based segmentation outputs as geospatial communication artifacts: they condense complex image information into masks, comparative growth maps, and indicators that can be discussed by non-specialist audiences.

Two main reasons lie behind our interest in using deep learning methods for this task. The second is, when the images contain clouds, vegetation, shadows, bright soil or road fragments and map labels fully automatic segmentation may be incorrect. We accommodate established image processing, human correction and neural network fine-tuning to ensure both reproduction of outputs and intuitivity of human visualisation. The main objective is to compare predicted urban masks from 2013 and 2024 and quantify the increase in detected built-up area. The contributions are threefold: (i) a compact workflow that uses HSV-based pseudo-labeling, manual refinement, and U-Net fine-tuning; (ii) a communication-oriented output design based on comparative masks and urban-growth overlays; and (iii) a pixel-based quantification of urban expansion in Santa Cruz Island that can be used as a preliminary indicator for further geospatial analysis.

2 Related Work and Theoretical Foundation

2.1 Geospatial Communication and Visualization

The relevance of this study to ICOMTA lies in the use of image-processing and learning-based methods to produce communicable spatial evidence. Geospatial communication refers here to the design and transmission of spatial information through maps, overlays, indicators, and interactive or reproducible workflows. MacEachren argues that map design involves cognitive and semiotic decisions that influence how spatial information is understood [1]. Kraak similarly describes geovisualization as an integration of cartography, image analysis, information visualization, exploratory data analysis, and GIS for the visual exploration and presentation of geospatial data [2].

The significance of this perspective lies in the fact that urban growth cannot be fully represented by a segmentation model alone. Communication arises when the model output is converted into a visual quantitative form that facilitates interpretation. For that reason, the binary urban masks comparison 2013 to 2024 and the colored growth visualization presented in this paper are not only outputs of technical nature; they are products of visual communication about how territory has changed.

The rise of web-based geographic information and user-generated geographic data also shows that geospatial technologies have become part of contemporary communication ecosystems. Goodchild describes citizens and digital platforms as new producers and disseminators of geographic information [3]. Although the present work does not use volunteered geographic information, it follows the same broader logic: digital tools can make geographic evidence easier to produce, share, and interpret.

2.2 Urban Remote Sensing and Semantic Segmentation

Remote sensing has been widely used to measure urban growth, built-up areas, and impervious surfaces. Herold et al. combine remote sensing, spatial metrics, and urban-growth modeling to analyze the spatiotemporal form of urban change [4]. Weng reviews requirements, methods, and trends for remote sensing of urban impervious surfaces, highlighting the importance of image quality, scale, and classification strategy [5]. Seto et al. demonstrate that urban expansion has direct implications for biodiversity and carbon pools, which is especially relevant for ecologically sensitive regions [6].

Recent advances in convolutional neural networks have improved dense image labeling and object extraction from aerial and satellite images. U-Net introduced an encoder-decoder architecture for pixel-level segmentation with limited annotated data [7], while EfficientNet proposed a systematic model-scaling strategy for efficient convolutional backbones [8]. In remote sensing, fully convolutional approaches have been used for building extraction from aerial and satellite imagery [9]. The present paper adapts these ideas to a small-data scenario by first creating pseudo-labels, then manually refining them, and finally fine-tuning a U-Net model.

The literature suggests that urban remote sensing becomes more useful when it is connected to interpretation and communication. For this reason, the proposed workflow emphasizes not only model performance but also the creation of outputs that can be read as maps: stable urban area, new urban growth, and uncertain loss or detection changes.

3 Methodology

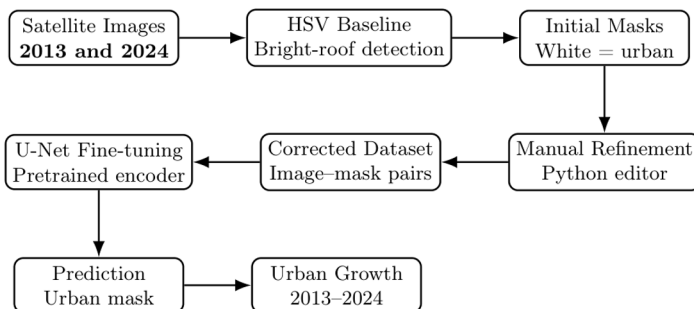


Fig. 1. Workflow for satellite-based urban expansion analysis and geospatial communication of urban growth in Santa Cruz Island, Galápagos.

3.1 Satellite Image Collection and Annotation Goal

Input data Satellite screenshot of Santa Cruz Island Overall, the input to the model came as paired captures of the year 2013 and more specifically, the year is for 2024. Each image was paired with a binary mask where white pixels indicated buildings or other visible built structures in the dataset and black pixels denoted all other areas classified as non-urban, including vegetation, roads, soil, shadows and clouds. The dataset for rapid tuning contained five image-mask pairs. The dataset consisted of four training pairs and one validation pair. Here, the annotation rule was kept constant for both years in order to compare growth. The purpose of the annotation was to provide a clear signal for manually annotated buildings with an application above detection, but not as rigorous as a formal official cadastral feedback product. Of course, this distinction matters for communication: outputs should be interpreted as a kind of proof of models that we produce when trained.

3.2 HSV-Based Initial Segmentation

The first segmentation stage used the HSV color space. The baseline assumed that many roofs and built structures appear as bright and low-saturation regions. Therefore, pixels with high value and relatively low saturation were selected as urban candidates. Morphological filtering and connected-component filtering were then applied to reduce noise.

This method rapidly produced initial masks, but it also missed dark roofs, roofs under vegetation, and low-contrast buildings. It also confused bright soil, clouds, map labels, and some road fragments with buildings. Therefore, the HSV result was treated as a pseudo-label rather than a final ground truth.

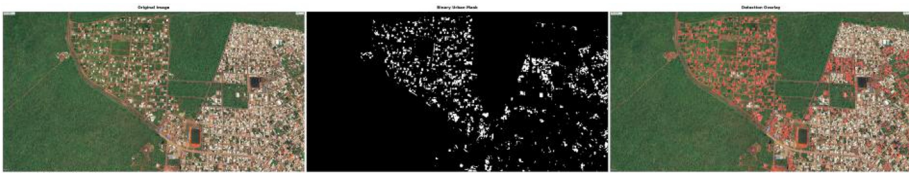


Fig. 2. Limitations of HSV-based segmentation: missed structures and false positives.

3.3 Mask Refinement Tool

To avoid redrawing masks manually, a Python-based correction tool was implemented. The tool loads the original satellite image and its preliminary binary mask, displays the mask as a transparent red overlay, and allows the user to add or remove mask regions with the mouse. This step improves the quality of the labels used for neural-network training and makes the annotation process easier to audit.

From a communication-technology perspective, the manual refinement interface also serves as a verification layer. It allows the analyst to inspect the visual relationship

between the image and the mask before the mask becomes training data, reducing the risk of communicating obvious segmentation errors as if they were urban growth.



Fig. 3. Manual mask refinement using the Python-based correction tool.

3.4 Learning-Based Segmentation

After mask correction, a U-Net model was fine-tuned to classify each pixel as urban or non-urban. The model used an EfficientNet-B0 encoder initialized with ImageNet weights and was implemented with segmentation-models-pytorch [10]. Since the source images were large, training was performed using random 512 x 512 patches instead of full images. To compensate for the small proportion of urban pixels, the sampler favored patches containing positive mask pixels with a positive_patch_fraction of 0.75.

Table 1. Dataset split and main fine-tuning parameters.

Setting	Value	Setting	Value
Model	U-Net	Encoder	EfficientNet-B0
Initial weights	ImageNet	Library	SMP
Training pairs	4	Validation pairs	1
Patch size	512 x 512	Positive patch frac- tion	0.75
Epochs	10	Samples per epoch	300
Validation patches	100	Batch size	1
Learning rate	3×10^{-4}	Loss	BCE + Dice
Positive weight	3.0	Best checkpoint	epoch 7

Lightweight augmentation included flips, rotations, brightness changes, saturation changes, and blur. During inference, large images were predicted by tiles, probability maps were merged, a threshold was applied, and small connected components were removed.

3.5 Urban Growth Measurement

Urban growth was estimated by counting the white pixels in the binary masks predicted for 2013 and 2024. Pixels detected in both years were interpreted as stable urban area, pixels detected only in 2024 as new urban growth, and pixels detected only in 2013 as loss or uncertain detection. Due to the way in which these images correspond to screenshots and not mosaic based georeferenced orthomosaics, results are reported as model based pixel-area indicators in lieu of official hectare measurements.

3.6 Communication-Oriented Output Design

The end result at a finer scale is to convey the change in three orthogonal pieces of information: (i) binary masks on urban and non-urban pixels, (ii) summary tables by year and zone indicating the quantity of change and (iii) a colorized overlay indicating stable urban area, new urban growth, or while exact loss or uncertain detections. This design choice responds to the need to make technical segmentation results understandable for audiences that may not be familiar with neural networks or remote-sensing metrics.

4 Experimental Results

4.1 Segmentation Performance

The 10-epoch fine-tuning experiment reached its best validation checkpoint at epoch 7, with an IoU of 0.6544 and an F1-score of 0.7911. The precision of 0.8819 indicates that most pixels predicted as urban were correct, while the recall of 0.7173 shows that some urban regions were still missed. The later decrease in validation performance motivated keeping the best checkpoint rather than the final epoch.

Table 2. Best validation checkpoint from the 10-epoch fine-tuning experiment.

Epoch	Val. loss	IoU	F1-score	Precision	Recall
7	0.4192	0.6544	0.7911	0.8819	0.7173

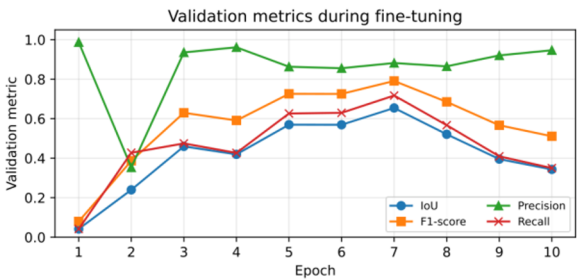


Fig. 4. Validation metrics during fine-tuning.

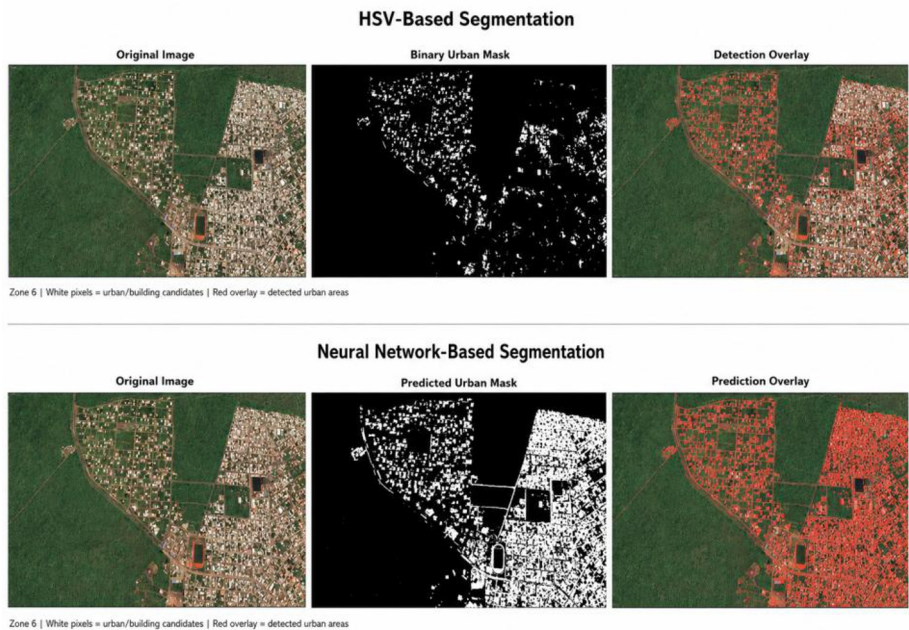


Fig. 5. Qualitative comparison between HSV-based and neural-network segmentation.

4.2 Urban Expansion Results

The predicted masks show a clear increase in detected urban pixel area. Across all analyzed Santa Cruz zones, the urban area increased from 5,663,312 pixels in 2013 to 14,428,135 pixels in 2024, producing a net increase of 8,764,823 pixels and a relative net growth of 154.76%.

Table 3. Pixel-based urban expansion summary from the predicted masks.

Metric	Total zones	Base zone
Urban area 2013	5,663,312 px	2,725,212 px
Urban area 2024	14,428,135 px	7,829,338 px
Stable urban area	3,911,249 px	2,296,545 px
New urban growth	10,516,886 px	5,532,793 px
Loss / uncertain area	1,752,063 px	428,667 px
Net growth	8,764,823 px	5,104,126 px
Relative growth vs. 2013	154.76%	187.29%

The base zone shown in Fig. 6 increased from 2,725,212 pixels in 2013 to 7,829,338 pixels in 2024, equivalent to a net increase of 5,104,126 pixels and an estimated 2.87x increase.

Table 4. Relative net growth by analyzed zone.

Zone	Relative net growth vs. 2013
Zone 5	503.93%
Zone 6	379.12%
Base zone	187.29%
Zone 4	170.91%
Zone 3	158.99%
Zone 8	76.97%
Zone 2	22.52%

The largest relative growth was observed in Zone 5, followed by Zone 6 and the base zone. However, relative percentages must be interpreted carefully: when a zone has a small detected urban area in 2013, even a moderate increase in 2024 can produce a very large percentage change.

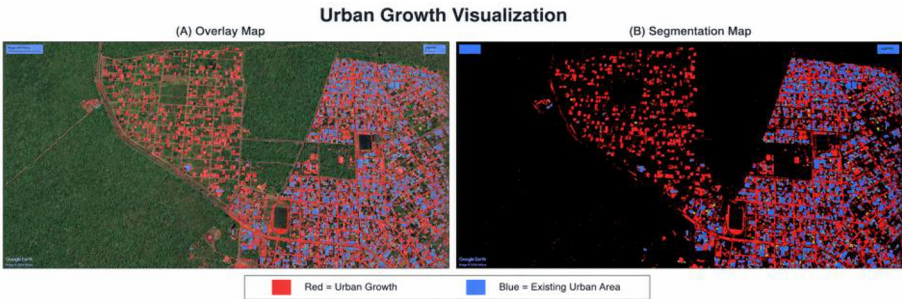


Fig. 6. Predicted urban-growth visualization for Santa Cruz Island. Red indicates new urban growth, blue stable urban area, and yellow loss or uncertain detections.

5 Discussion

The proposed workflow combines speed and adaptability: HSV segmentation provides fast pseudo-labels, manual correction improves label quality, and U-Net fine-tuning adapts the detector to the visual conditions of Santa Cruz Island. Results indicate a sharp estimated built-up area increase from 2013 to 2024 but should be interpreted as model derived pixel-area indicators and not official land-area estimates. The central axis of the revised framing of the paper is connection to communication technologies. A segmentation metric driving this workflow is realizing a visual interpretation that can be modeled on the surfaces, overlays, and tabulations. The method thus aids environmental communication in translating satellite impression into a more perception accessible visual vocabulary: stable urban classes, newly detected urban expansion and uncertain transformations. The study also shows how applied artificial intelligence can support communication processes without replacing human interpretation. The manual refinement step recognizes that segmentation errors can change the message conveyed by a map. Therefore, the analyst remains part of the communication pipeline, validating

masks before they are used to train the model and before results are transformed into public-facing visualizations.

Several limitations must be considered. Clouds, roads, shadows, map labels, bright soil and segmentation errors can all result in false positives. This small test set contained only five corrected image and mask pairs, therefore the model is intended as a quick proof of concept rather than widely generalizable detector. Also the images were screenshots and not georeferenced orthomosaics meaning the pixel counts cannot yet be accurately converted to hectare-level estimates. Future work should include georeferenced imagery, more corrected masks, external validation, threshold calibration, and user-centered evaluation of how different audiences interpret the resulting maps.

6 Conclusion

This work presents a reproducible pipeline for estimating and communicating urban expansion in Santa Cruz Island, Galápagos. HSV-based pseudo-labeling, manual mask refinement, and U-Net fine-tuning were combined to produce urban masks from satellite screenshots. The best checkpoint achieved an IoU of 0.6544 and an F1-score of 0.7911. The comparison between 2013 and 2024 estimated an increase in detected urban area from 5.66 million to 14.43 million pixels, corresponding to a net increase of 154.76%. This paper views segmentation as a geospatial communication technology that opens new means of interpretation through visualization and quantitative evidence from, often spectrally complex, satellite imagery. The masks, tables, and growth maps can be useful to show urban transformation to researchers as well as institutional actors or non-specialist audiences.

Nevertheless, hectare-level measurement requires georeferenced imagery or a reliable meters-per-pixel scale, and future work should validate both the technical accuracy and the communicative usefulness of the outputs.

7 Acknowledgements

The authors acknowledge the support provided by the International University of Ecuador (UIDE) through the research project: Use of Emerging Technologies for the Sustainable Management of Natural Resources in the Galápagos Islands, code: UIDE-GI-PPI-PRY-CI-26-09.

References

1. MacEachren, A.M.: *How Maps Work: Representation, Visualization, and Design*. Guilford Press, New York (1995).
2. Kraak, M.J.: Geovisualization illustrated. *ISPRS Journal of Photogrammetry and Remote Sensing* 57(5-6), 390-399 (2003).
3. Goodchild, M.F.: Citizens as sensors: the world of volunteered geography. *GeoJournal* 69, 211-221 (2007).

4. Herold, M., Goldstein, N.C., Clarke, K.C.: The spatiotemporal form of urban growth: measurement, analysis and modeling. *Remote Sensing of Environment* 86(3), 286-302 (2003).
5. Weng, Q.: Remote sensing of impervious surfaces in the urban areas: requirements, methods, and trends. *Remote Sensing of Environment* 117, 34-49 (2012).
6. Seto, K.C., Güneralp, B., Hutya, L.R.: Global forecasts of urban expansion to 2030 and direct impacts on biodiversity and carbon pools. *Proceedings of the National Academy of Sciences* 109(40), 16083-16088 (2012).
7. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation. In: *MICCAI 2015*, pp. 234-241. Springer (2015).
8. Tan, M., Le, Q.V.: EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In: *Proceedings of the 36th International Conference on Machine Learning, PMLR* 97, pp. 6105-6114 (2019).
9. Ji, S., Wei, S., Lu, M.: Fully Convolutional Networks for Multisource Building Extraction from an Open Aerial and Satellite Imagery Data Set. *IEEE Transactions on Geoscience and Remote Sensing* 57(1), 574-586 (2019).
10. Yakubovskiy, P.: Segmentation Models PyTorch. Software repository (2020).

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

