Semantic Segmentation of Poverty from Satellite Imagery with U-Net

Abstract. Accurate mapping of poverty distribution is crucial for informed policymaking and effective allocation of resources. This study addresses the challenge of mapping poverty distribution by leveraging high-resolution satellite imagery and deep learning techniques. Specifically, we employ U-Net and a hybrid U-Net with VGG16 architectures for semantic segmentation to achieve poverty mapping, utilizing annotations based on building characteristics, road infrastructure, green spaces, water bodies, beaches, informal settlements (commonly referred to as slums), and residential villas as indicators of poverty levels in Moroccan cities. Extensive experimentation with K-means and fuzzy c-means clustering techniques did not yield meaningful delineation results. The visual criteria serve as fundamental indicators for initial annotation delineation. The 745 annotated satellite images underwent preprocessing steps, including normalization and patch creation, to enhance model training. Focal loss functions were implemented to mitigate class imbalance issues inherent in the data. The U-Net-VGG16 model achieved an accuracy of 82.58%, an Intersection over Union (IoU) of 70.28%, a recall of 67.95%, a precision of 83.65%, and an F1-score of 70.66%, outperforming the baseline U-Net model. These results underscore the potential of deep learning-based semantic segmentation in effectively mapping poverty through high-resolution satellite imagery. The findings contribute to developing data-driven tools for policymakers aimed at poverty alleviation and urban planning.

Keywords: Deep learning, Poverty mapping, Satellite imagery, Semantic segmentation, U-Net (Unet), VGG16, Focal loss, CNNs

1 Introduction

Factual and timely poverty data are essential for effective policymaking and the implementation of poverty alleviation strategies. However, traditional data collection methods, such as national censuses, are typically conducted every ten years and are both time-consuming and costly, leading to outdated information that hinders responsive decision-making [1]. The definition of poverty remains a complex challenge. While absolute poverty is characterized by insufficient financial resources to meet basic needs, as defined by the World Bank [2], multidimensional poverty indices offer a more comprehensive view by incorporating a multi-factorial aspect through the use of criteria such as education, health, and living standards [3]. According to Britannica, poverty is said to exist when people lack the means to satisfy their basic needs. In this context, identifying poor populations first requires determining what constitutes basic needs. Traditional poverty measurements often rely on outdated or incomplete data, hindering effective decision-making [4]. Recent advancements in remote sensing and Artificial Intelligence (AI) provide promising alternatives. The adoption of AI has seen a significant surge in recent years. Driven by breakthroughs in data science and deep learning, AI now enables the resolution of complex, data-intensive challenges and supports more informed decision-making processes. High-resolution satellite imagery, combined with deep learning techniques [5], offers a cost-effective and scalable alternative for mapping poverty distribution, enables the analysis of physical indicators such as building densities, the state of road infrastructure, the presence or absence of verdant parks and open spaces, the course of water, the presence of beaches, the existence of informal settlements, and the delineation of residential villas, and land use patterns, which can serve as proxies for socioeconomic status. Studies have demonstrated the effectiveness of convolutional neural networks (CNNs) [6] in predicting poverty levels from satellite images, offering a scalable and cost-effective solution for poverty mapping [7]. In this study, we apply semantic segmentation using U-Net and a hybrid U-Net with VGG16 architectures to analyze satellite imagery of Moroccan cities. This approach aims to provide a novel, low-cost method for assessing poverty distribution. The insights gained from this analysis can inform data-driven policymaking and contribute to more effective poverty mitigation strategies.

2 Related work

Over the past two decades, satellite imagery has been extensively utilized in machine learning and deep learning applications to address socio-economic challenges, particularly in poverty mapping. Jean et al. [7] demonstrated the use of satellite imagery combined with machine learning to predict poverty levels in African countries. Their study highlighted the effectiveness of high-resolution satellite images in estimating consumption expenditure and asset wealth. Xie et al. [5] discussed the application of transfer learning using deep features from satellite images to map poverty. They showcased the benefits of employing pre-trained models and transfer learning techniques to enhance poverty prediction accuracy. Shakeel et al. [8] introduced a deep learning model for counting built structures in satellite images using attention-based re-weighting. Their model exhibited high accuracy in counting buildings across various geographical regions. Ayush et al. [7] explored the use of high-resolution remote sensing images for efficient poverty mapping. They addressed the challenges and opportunities associated with using satellite imagery for poverty prediction and mapping.

Fig. 1. Nighttime lights highlighting disparities between affluent and impoverished regions [9]

The illustration in Figure 1 underscores the stark disparities between wealthy and impoverished regions, discernible through nighttime illumination analysis. Prosperous areas exhibit abundant luminous sources, contrasting sharply with the dimly lit impoverished zones. Regions devoid of nighttime illumination often indicate extreme impoverishment. Such maps serve as pivotal resources for initiating the annotation process of satellite images, facilitating the creation of segmentation masks that accurately capture and depict distinct realities. However, in our project, nighttime light data for Moroccan cities were unavailable. Nighttime lights serve as a fundamental indicator of economic prosperity. Global nighttime maps reveal that numerous developing nations exhibit limited illumination. Jean et al. [10] integrated these nighttime maps with detailed daytime satellite images. Through advanced machine learning techniques, this amalgamation yielded precise estimations of household consumption and asset levels, both challenging to quantify in economically disadvantaged regions. In particular, data from both nighttime and daytime sources are publicly accessible and not subject to proprietary ownership. Their study demonstrated an accurate, inexpensive, and scalable method for estimating consumption expenditure and asset wealth from high-resolution satellite imagery. Utilizing survey and satellite data from five African countries including Nigeria, Tanzania, Uganda, Malawi, and Rwanda, they showcased how powerful machine-learning techniques can be applied in settings with limited training data. Given the complexity and diversity inherent in daytime satellite imagery, their findings revealed that the CNN model surpassed other methodologies. The use of transfer learning proved invaluable in feature extraction, subsequently channeled into a Ridge regression model applied to model daily consumption or assets. It's important to note that the poverty line is defined at \$1.90 USD, a threshold subject to fluctuations based on economic conditions. A notable correlation exists between nighttime luminosity and daily consumption patterns within the studied countries. Figure 2 provides evidence of the potential of utilizing satellite imagery in poverty modeling. It underscores the added value of integrating nighttime intensity maps to augment satellite image annotations, even though data for Morocco's nighttime intensity remains unavailable. In summary, studies have explored methodologies for mapping and modeling poverty using satellite imagery. Notably, approaches such as those presented in [5] and [10] have been influential. Building upon these foundational works, this study employs semantic segmentation techniques to further investigate poverty mapping through satellite imagery.

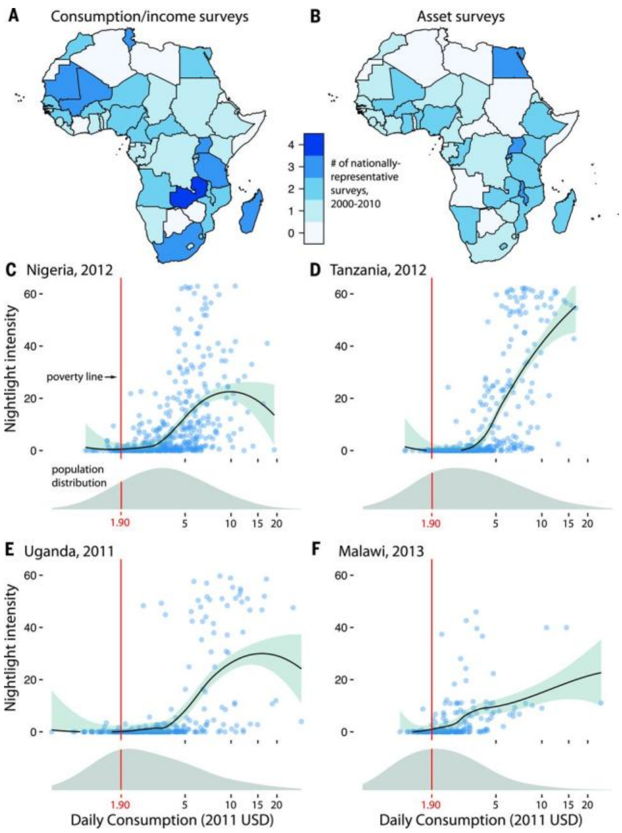

Fig. 2. Poverty data gaps and their correlation with nighttime illumination [10]

3 Methodology

3.1 Empirical Design

The empirical design for this study involves several key steps. It begins with a review of related work on poverty and satellite imagery to establish a foundational understanding of existing methods. This is followed by an exploration of potential statistical data sources suitable for annotating satellite images with socio-economic indicators. Criteria for manual annotation are then defined, including the main annotation criteria such as building density and infrastructure types. An attempt is made to replicate the findings of a previous study [8] to benchmark and validate the methodology. The Pseudo-Mercator 3857 projection is used for accurate area calculation, which is critical for computing and thresholding building density. Satellite images are manually annotated using an offline setup of CVAT (Computer Vision Annotation Tool), and image quality is enhanced using QGIS to improve feature visibility. The dataset is then pre-processed to standardize image dimensions and formats, followed by training and evaluation of machine learning models to assess their ability.

3.2 Satellite Imagery Analysis

Quantum Geographic Information System (QGIS) was utilized to calculate the areas covered by satellite images, which contributed to computing the building density annotation criteria. Additionally, QGIS was used to enhance the brightness and quality of the satellite images, including adjustments to contrast and luminosity.

3.3 Annotation Criteria

Criterion Based on Statistical Data
To ensure the robustness and accuracy of the annotations, statistical data related to poverty, daily per capita income, daily consumption, and nighttime lights were considered. These data sources have been shown to be effective in previous studies [10]. However, efforts to acquire this data were unsuccessful. Hence, visual cues such as the condition and structure of buildings, road infrastructure, presence of greenery, water bodies, informal settlements, and residential villas were used as annotation criteria.

Unsupervised Learning Techniques.
Unsupervised learning techniques, including K-means clustering and Fuzzy c-means clustering [11], [12], were explored to identify patterns in the satellite images. Despite extensive experimentation, these approaches did not yield significant outcomes for our analysis; however, they provided insightful points to identify the delineation of potential classes for the annotation process. The Figures 3 and 4 showcase the results of the K-means clustering analysis applied to a satellite image of Casablanca. The range of potenti3 and clusters spanned from 3 to 5, with 4 (see Figure 6, 5) clusters appearing to align closely with discernible features within the satellite image of Figure 5.

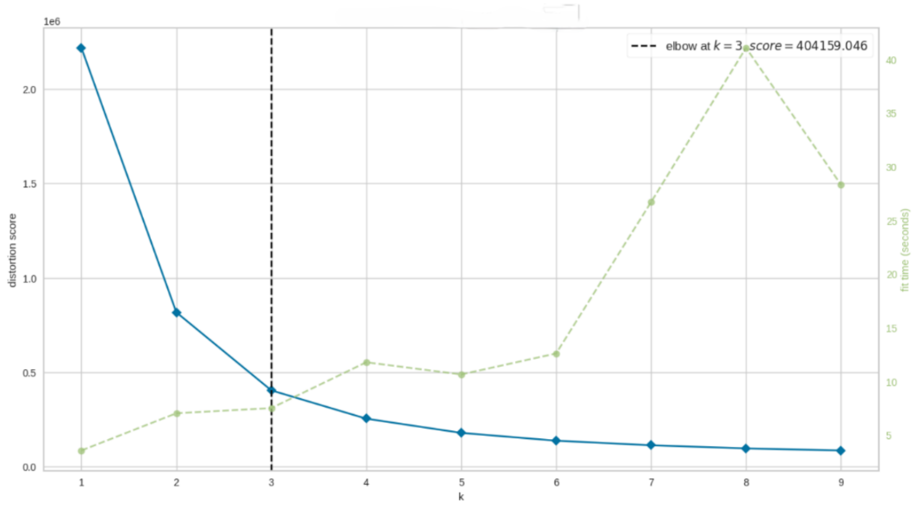

Fig. 3. K-means distortion score elbow

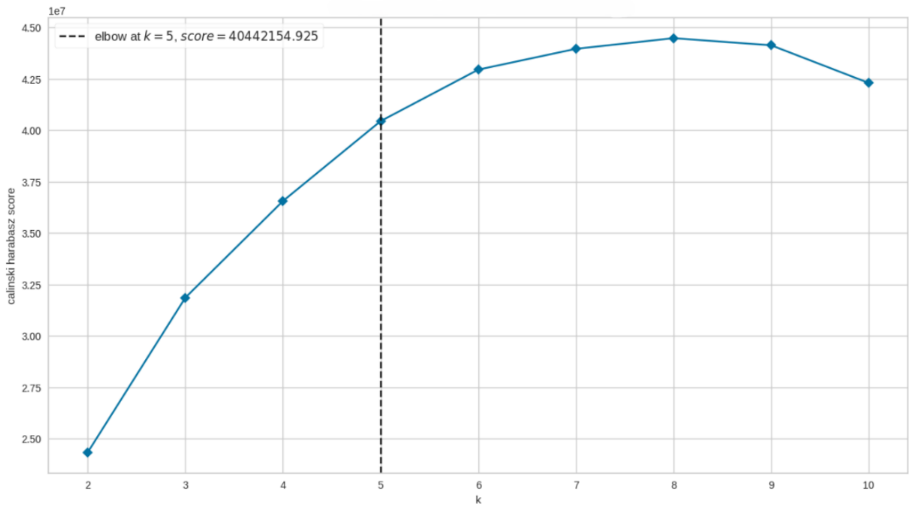

Fig. 4. K-means Calinski-Harabasz score

3.4 Deep Built-Structure Counting Model

The deep built-structure counting model, as presented in the [8] article, utilizes a Fusion network to integrate features from three distinct branches. Among these, two branches employ attention mechanisms, while the third relies on a fully connected network trained using DenseNet features. The attention-based branches include the Cross-Channel Parametric Pooling (CCPP) and Global Weighted Average Pooling (GWAP), both designed to filter DenseNet features through the SS-NET (Satellite Segmentation Network). Initially, the attention-based branches are trained independently before being fused into the FusionNet architecture. The resulting architecture supports regression tasks for building counting in satellite imagery. The dataset used for training this model consists of satellite imagery from diverse geographical regions, ensuring a comprehensive scope of terrains, such as urban centers, plains, deserts, and more. The dataset spans continents, including Asia, Europe, North America, and Africa, capturing a wide variety of built structures. This global scope of imagery enhances the flexibility of the model and allows its application to Moroccan cities. Specifically, the attributes present in the satellite imagery of Moroccan cities align well with those in the global dataset, further strengthening the model's applicability and reliability. For the purpose of replicating the deep built-structure counting model, the SSNet Base was utilized for feature extraction. After an extensive training and testing phase, a Mean Absolute Error (MAE) of 3.76 was achieved, while the original model reported an MAE of 3.65. In the training phase, the model produced a Root Mean Squared Error (RMSE) of 4.260, which was in close agreement with the original GitHub code's RMSE of 4.253.

The Annotation Criterion Based on Building Density

To establish a consistent and uniform annotation process for the satellite images, the findings of the "Analyse multidimensionnelle des classes moyennes au Maroc (Multidimensional analysis of the middle classes in Morocco)" study were considered, which outlines the socio-economic distribution of the Moroccan population. This classification divides the population into three categories: 50.1% middle class, 34.4% modest class, and 15.5% wealthy class. These proportions served as the foundation for setting thresholds based on building density, helping in the annotation of satellite images for different urban categories. For each Moroccan city under consideration, density thresholds were defined based on the building density within satellite image patches. In the case of Casablanca, for example, the threshold ranges were as presented in Table 1.

Table 1. Classes labels and definition (according to buildings density)

Classes	Label	Buildings density (buildings/Km²)
RICH	Green	< 362.64
MEDICUM	Yellow	[362.64, 540.856]
POOR	Red	> 540.856

These thresholds were calculated based on the first and third quartiles of the density distribution, adjusted by half of the standard deviation. This density-based criterion aimed to ensure homogeneity in the annotation process across different cities.

Satellite Image Annotation Process

The annotation of satellite imagery was carried out using CVAT. The segmentation of each satellite image was based on the annotation criteria previously mentioned. Each satellite image was partitioned into meaningful segments that were analyzed for building density. Using the deep built-structure counting model, the number of buildings in each patch was computed, followed by calculating the density of buildings within the segment. The patches were partitioned into standardized sizes of (336, 336, 3) for model [8] input, ensuring a consistent approach to annotating each segment and calculating the building density.

Data Preprocessing

QGIS was utilized to enhance the satellite images by adjusting brightness, contrast, and overall visual quality, ensuring that features relevant to building counting were clearly visible. Due to varying image resolutions, the images were divided into smaller patches of size (512, 512, 3) to maintain consistency and quality during segmentation. This patching approach facilitated effective model training. Additionally, min-max normalization was applied to each patch, scaling pixel values to a range of 0 to 1, which is essential for preparing the data for deep learning models.

Fig. 5. A satellite image of Casablanca

Fig. 6. An annotated satellite image of Casablanca corresponding to Figure 5

4 Results and Discussion

On the test dataset, the loss function converged to a value of 0.117, while the accuracy reached 82.58%. Simultaneously, the Jaccard Index (also known as Intersection-over-Union) [13] achieved a value of 70.28%. These results are further supported by additional evaluation metrics: the recall was 67.95%, indicating the model's effectiveness in capturing relevant structures, and the precision reached 83.65%, reflecting its ability to avoid false positives. The F1-score, a harmonic mean of precision and recall, attained a robust value of 70.66%. These performance metrics were obtained using the Unet-VGG16 base model trained on a carefully balanced dataset. This dataset was enhanced using data augmentation techniques, specifically RandomRotation and RandomFlip [14], applied to both rich and poor segmentation masks. Additionally, the use of the focal loss function helped address class imbalance, which likely contributed to the improved model performance over the balanced dataset and earlier stopping parameter. Table 2 presents a comparative summary of performance metrics for both the Unet and Unet-VGG16 models. Both models were trained on the same balanced dataset using oversampling strategies to mitigate disparities between rich and poor segmentation masks. The results clearly demonstrate the superior performance of the Unet-VGG16 model across all key evaluation metrics, highlighting its suitability for accurate segmentation tasks. On the other hand, over the unbalanced data, both U-net and U-net VGG16 models face overfitting, even across all loss functions tried and implemented, including asymmetric focal, dice focal, and balanced cross-entropy loss functions. This overfitting issue has been widely discussed in previous studies, which highlight the challenges posed by class imbalance in neural networks [15]. Additionally, the use of focal loss to address this imbalance has been shown to be effective, although it can still result in suboptimal performance [16]. In the sample prediction showcases (see Figures 7, 8, and 9), the first column displays the original image patch, the second column presents the ground truth, and the third column reveals the predicted segmentation mask.

Table 2. Performance Comparison of Unet and Unet-VGG16 Base Models

Model	Accuracy	IoU	Recall	Precision	F_1-score
Unet base	**70.32%**	**55.42%**	**67.03%**	**83.91%**	**70.51%**
Unet-VGG16	82.58%	70.28%	67.95%	83.65%	70.66%

Fig. 7. Sample prediction for the city of Casablanca

Fig. 8. Sample prediction for the city of Agadir

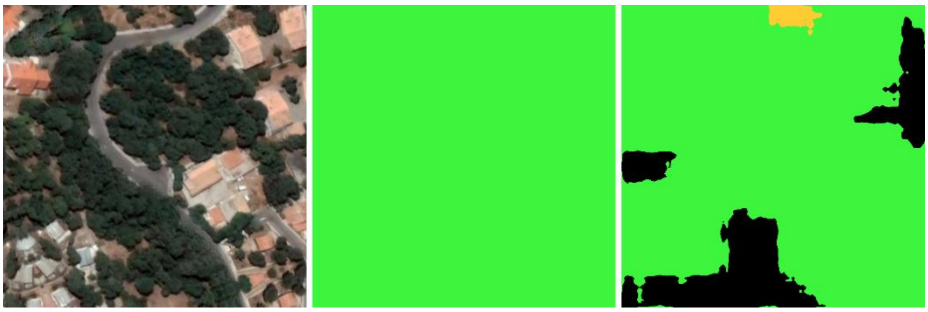

Fig. 9. Sample prediction for the city of Ifrane

5 Conclusion

Poverty mapping through semantic segmentation, high-resolution satellite imagery, and deep learning offers a low-cost and scalable solution. Despite its potential, challenges like data annotation quality and class imbalance persist. While the U-Net model provides faster convergence, the U-Net VGG16 demonstrates better capability in capturing complex features and performances, especially when trained on a balanced and augmented dataset. However, prediction accuracy remains limited by the quality of annotations. Integrating verified statistical indicators such as household income and nighttime light intensity [9] can improve annotation reliability, robustness, and model outcomes.

Future work should explore transfer learning with models like Segment Anything [17], expand the dataset, and refine annotation criteria. Additionally, incorporating auxiliary data sources such as socioeconomic surveys and geospatial statistics can further enhance the robustness and accuracy of poverty mapping systems. These advancements can support data-driven policy-making and more effective poverty alleviation strategies.

References

1. M. Karen Tinsley, “Poverty and population density: Implications for economic development policy,” Journal of Higher Education Outreach and Engagement, Volume 11, Number 1, p. 195, (2006)
2. W. B. Group, “Poverty measurement from noise to signal. . . and how the media can help.”
3. A. H. H. S. H. Houssine Ihnach, Abdenbi El Ansary, “Les effets de la politique budg´etaire sur la pauvret´e et les in´egalit´es au maroc,” AFD ´Editions, November 2020.
4. F. DAOUDIM, “Analyse multidimensionnelle des classes moyennes au maroc,” 2-4 March 2020
5. M. D. S. Michael Xie, Neal Jean, “Transfer learning from deep features for remote sensing and poverty mapping,” ArXiv, Cornell University, 2016
6. H. Lei, N. Bao, M. Yu, and Y. Cao, “Estimating and mapping tailings properties of the largest iron cluster in china for resource potential and reuse: A new perspective from interpretable cnn model and proposed spectral index based on hyperspectral satellite imagery,” International Journal of Applied Earth Observation and Geoinformation, vol. 139, p. 104512,2025.Available:https://www.sciencedirect.com/science/article/pii/S1569843225001591
7. K. M. D. S. Kumar Ayush, Burak Uzkent, “Efficient poverty mapping from high resolution remote sensing images,” The Thirty-Fifth AAAI Conference on Artificial Intelligence (AAAI-21), 2021-05-18
8. A. Shakeel, W. Sultani, and M. Ali, “Deep built-structure counting in satellite imagery using attention based re-weighting,” ISPRS Journal of Photogrammetry and Remote Sensing, vol. 151, pp. 313–321, 2019.

9. M. Charpe, “Night lights as a proxy for economic growth? an application to structural transformation,” 3d January 2022
10. M. W. D. S. Neal Jean, Marshall Burke, “Combining satellite imagery and machine learning to predict poverty,” American Association for the Advancement of Science, 2016.
11. N. Dhanachandra, K. Manglem, and Y. J. Chanu, “Image segmentation using k -means clustering algorithm and subtractive clustering algorithm,” Procedia Computer Science, vol. 54, pp. 764–771, 2015, eleventh International Conference on Communication Networks, ICCN 2015, August 21-23, 2015, Bangalore, India Eleventh International Conference on Data Mining and Warehousing, ICDMW 2015, August 21-23, 2015, Bangalore, India Eleventh International Conference on Image and Signal Processing, ICISP 2015, August 21-23, 2015, Bangalore, India.
12. K.-S. Chuang, H.-L. Tzeng, S. Chen, J. Wu, and T.-J. Chen, “Fuzzy c-means clustering with spatial information for image segmentation,” Computerized Medical Imaging and Graphics, vol. 30, no. 1, pp. 9–15, 2006
13. H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, “Generalized intersection over union: A metric and a loss for bounding box regression,” 2019.
14. F. Chollet et al., “Keras,” 2015, https://www.tensorflow.org/api_docs/python/tf/keras/layers/RandomRotation
15. Z. Li, K. Kamnitsas, and B. Glocker, “Overfitting of neural nets under class imbalance: Analysis and improvements for segmentation,” arXiv preprint arXiv:1907.10982, 2019.
16. M. Yeung, E. Sala, C.-B. Sch¨onlieb, and L. Rundo, “Unified focal loss: Generalising dice and cross entropy-based losses to handle class imbalanced medical image segmentation,” arXiv preprint arXiv:2102.04525, 2021
17. A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, P. Doll´ar, and R. Girshick, “Segment anything,” 2023