TeMA

Journal of Land Use, Mobility and Environment

print ISSN 1970-9889 e-ISSN 1970-9870
FedOA press - University of Naples Federico II

DOAJ   Rivista scientifica di classe A - 08/F1   Scopus   WEB OF SCIENCE

NEW CHALLENGES FOR XXI CENTURY CITIES

Global warming, ageing of population, reduction of energy consumption, immigration flows, optimization of land use, technological innovation

Vol.17 n.3
Dicember 2024

TeMA Journal was established with the primary objective of fostering and strengthening the integration between urban transformation studies and those focused on mobility governance, in all their aspects, with a view to environmental sustainability. The three issues of the 2024 volume of TeMA Journal propose articles that deal the effects of global warming, the ageing of population, the reduction of energy consumption from fossil fuels, the immigration flows from disadvantaged regions, the technological innovation and the optimization of land use.

TeMA is the Journal of Land Use, Mobility and Environment and offers papers with a unified approach to planning, mobility and environmental sustainability. With ANVUR resolution of April 2020, TeMA journal and the articles published from 2016 are included in the A category of scientific journals. The articles are included in main scientific database as Scopus (from 2023), Web of Science (from 2015) and the Directory of Open Access Journals (DOAJ). It is included in Sparc Europe Seal of Open Access Journals, and the Directory of Open Access Journals.

TeMA

Journal of
Land Use, Mobility and Environment

NEW CHALLENGES FOR XXI CENTURY CITIES:

Global warming, ageing of population, reduction of energy consumption, immigration flows, optimization of land use, technological innovation

3 (2024)

Published by
Laboratory of Land Use Mobility and Environment
DICEA - Department of Civil, Architectural and Environmental Engineering
University of Naples "Federico II"

TeMA is realized by CAB - Center for Libraries at “Federico II” University of Naples using Open Journal System

Editor-in-Chief: Rocco Papa
print ISSN 1970-9889 | online ISSN 1970-9870
Licence: Cancelleria del Tribunale di Napoli, n° 6 of 29/01/2008

Editorial correspondence
Laboratory of Land Use, Mobility and Environment
DICEA - Department of Civil, Building and Environmental Engineering
University of Naples "Federico II"
Piazzale Tecchio, 80
80125 Naples

www.serena.unina.it/index.php/tema
e-mail: redazione.tema@unina.it

The cover image shows heavy damage from floods in the Valencia, eastern Spain in October 2024
(Source: t.me/beholdisraelchannel/43470?single). The cover image was elaborated on to improve the resolution quality using free AI software.

TeMA. Journal of Land Use, Mobility and Environment offers researches, applications and contributions with a unified approach to planning and mobility and publishes original inter-disciplinary papers on the interaction of transport, land use and environment. Domains include: engineering, planning, modeling, behavior, economics, geography, regional science, sociology, architecture and design, network science and complex systems.

With ANVUR resolution of April 2020, TeMA Journal and the articles published from 2016 are included in A category of scientific journals. The articles published on TeMA are included in main international scientific database as Scopus (from 2023), Web of Science (from 2015) and the *Directory of Open Access Journals* (DOAJ). TeMA Journal has also received the *Sparc Europe Seal* for Open Access Journals released by *Scholarly Publishing and Academic Resources Coalition* (SPARC Europe). TeMA is published under a Creative Commons Attribution 4.0 License and is blind peer reviewed at least by two referees selected among high-profile scientists. TeMA has been published since 2007 and is indexed in the main bibliographical databases and it is present in the catalogues of hundreds of academic and research libraries worldwide.

TeMA Journal of
Land Use, Mobility and Environment

NEW CHALLENGES FOR XXI CENTURY CITIES:
Global warming, ageing of population, reduction of energy consumption, immigration flows, optimization of land use, technological innovation

3 (2024)

Contents

REVIEW NOTES

TeMA

Journal of
Land Use, Mobility and Environment

TeMA 3 (2024) 397- 410
print ISSN 1970-9889, e-ISSN 1970-9870
DOI: 10.6093/1970-9870/10729
Received 26th January 2024, Accepted 20th November 2024, Available online 31st December 2024

Building type classification using deep learning for transport planning

Aniruddha Khatua [a], Arkopal K. Goswami [b], Bharath H. Aithal [c*]

[a] Ranbir and Chitra Gupta School of Infrastructure Design and Management
Indian Insitute of Technology Kharagpur, Kharagpur, West Bengal, India
e-mail: aaniruddha4@kgpian.iitkgp.ac.in
ORCID: https://orcid.org/0000-0002-4308-0595

[b] Ranbir and Chitra Gupta School of Infrastructure Design and Management
Indian Insitute of Technology Kharagpur, Kharagpur, West Bengal, India
e-mail: akgoswami@infra.iitkgp.ac.in
ORCID: https://orcid.org/0000-0003-1369-215X

[c] Ranbir and Chitra Gupta School of Infrastructure Design and Management
Indian Insitute of Technology Kharagpur, Kharagpur, West Bengal, India
e-mail: bharath@infra.iitkgp.ac.in
ORCID: https://orcid.org/0000-0002-4323-6254
* Corresponding author

Abstract

The transportation and land use sectors are closely interdependent, and real-life circumstances often exhibit substantial reciprocal influences. Currently, efforts are being made to enhance transportation and land use sustainably; hence to achieve sustainability, it is necessary to have well-optimized plans and implementations for the advancements, which consequently leads to an increased demand for vast amounts of data. Conducting manual surveys to collect data on various types of buildings is considerably costly, requires much labor, and is time-consuming. Remote sensing technology has demonstrated significant promise to encompass a greater extent in a reduced timeframe, as well as to engage in thorough data collection and effectively manage and analyze the acquired data. This work centers on constructing a classification system that categorizes buildings depending on their use, specifically distinguishing between residential and non-residential structures. The classification challenge is accomplished through instance segmentation using the state-of-the-art YOLOV8 model architecture and remotely sensed images. The main objective of this project is to create base maps for travel analysis zones (TAZs) using identified buildings. To properly align the output images generated by the model, geographical data is appended to the output images derived from the original input images.

Keywords

Image segmentation; Land use classification; Transportation planning; YOLOV8.

How to cite item in APA format

Khatua, A., Goswami, A. K., Aithal, B. H. (2024). Building type classification using deep learning for transport planning. *TeMA - Journal of Land Use, Mobility and Environment*, *17* (3), 397-410. http://dx.doi.org/10.6093/1970-9870/10729

1. Introduction

Over the years, Indian cities have grown substantially, prompting administrative bodies to institute various measures to contend with the challenges arising from this demographic surge (Aithal & Ramachandra, 2016). The ramifications of this population expansion are intricate, with urban development becoming a pivotal aspect of addressing the needs of the burgeoning populace. One crucial aspect of this development lies in the simultaneous growth of multiple industries within urban areas. A notable interdependence exists between land use and transportation sectors, implying that advancements in one sector invariably necessitate corresponding developments in the other. This interlinkage is particularly evident in the context of urban planning and development. For instance, envisioning the expansion of a business district requires not only an augmentation of physical space but also necessitates well-developed transportation amenities. The term "amenities" in this context encompasses a comprehensive spectrum, ranging from infrastructure like roads and public transit networks to considerations of accessibility and service quality. The quality and efficiency of transportation infrastructure play a crucial role in facilitating the smooth functioning of urban centers and supporting the growth of various industries.

In practical terms, this might involve establishing efficient road networks, implementing robust public transportation systems, and enhancing overall accessibility within the urban fabric. The relationship between land use and transportation is symbiotic; improved transportation facilitates the accessibility of different areas (Raju et al., 2020; Galderisi et al., 2021), decision-making (Tira, 2021; Gazzola et al., 2021), and influencing land-use patterns, while changes in land use necessitate adjustments and improvements in transportation systems. Therefore, understanding and managing this intricate interplay is vital for sustainable and well-planned urban development, ensuring that the needs of a growing population are met effectively.

The role of land use features in the realm of transportation planning is widely recognized for its significance. However, the traditional method of collecting information on buildings over expansive areas faces practical challenges such as a time-consuming process, reduced reliability, high costs, and substantial manpower requirements. In response to these challenges, remote sensing emerges as a powerful solution, demonstrating its efficacy in resource management by efficiently covering vast areas while minimizing labor and time demands (Ramachandra et al., 2012; Carpentieri & Favo, 2017; Zullo et al., 2015). The escalating demand for transportation often results in the need for urban expansion, leading to significant sprawls on the outskirts. Delhi, for instance, has witnessed a remarkable increase of over 800% in its built-up area within four decades (Ramachandra et al., 2015; Bharath et al., 2017). Conducting spatial studies encompassing diverse land use features necessitates using effective models for extracting information. Adopting machine learning or deep learning-based models is gaining traction in the geospatial domain and various research areas. Extracting features from satellite or aerial photography is integral to remote sensing and GIS, with building extraction standing out as a substantial task. Researchers have developed diverse methodologies, including training fully convolutional networks on multi-source datasets to exploit variations in targeted objects (Ji et al., 2018), detecting buildings from low-contrast images (Amir et al., 2018; Chen et al., 2018; Li & Cao, 2019), utilizing images of different scales (Yang et al., 2018; Chen et al., 2022), leveraging building shadows at different sun angles (Sirmacek & Unsalan, 2008), and incorporating spatial feature-based metrics (Pavlidis & Liow, 1989). Segmentation is the most efficient method for extracting buildings from remotely sensed images. Numerous deep learning architectures have been devised for this purpose, including U-Net with various modifications (Wang & Miao, 2022; Prakash et al., 2022; Madhumita et al., 2023), VGGNet (Simonyan & Zisserman, 2014), GoogleNet (Szegedy et al., 2015), ResNet (He et al., 2016), ImageNet (Krizhevsky et al., 2017), and Segnet (Badrinarayanan et al., 2017). The categorization of land use features operates at multiple levels, with building extraction falling under level-1 classification, involving the isolation of buildings from the total image. Beyond this, level-2 classification (Which means after classifying and extracting the buildings form the entire image, those buildings are againg classified into different classes) encompasses the categorization of building types based on utility, distinguishing between residential and non-residential structures. Achieving this classification level requires viewing buildings as entities with distinct architectural configurations, characteristics, dimensions, and scales. Object identification or instance segmentation becomes essential to surpass level-1 classification. Various model designs have emerged within the computer vision sector to harness substantial advancements. Notable models include Fast-CNN (Girshick, 2015), You Only Look Once (YOLO) (Redmon et al., 2016), SSD: Single Shot Multibox Detector (Liu et al., 2016), and Spatial Pyramid Pooling (He et al., 2015). These models showcase the evolving landscape of techniques and approaches in extracting and categorising land use features, playing a crucial role in integrating remote sensing and GIS into transportation planning and spatial studies. The allocation of land in urban areas has far-reaching implications for various aspects of transportation planning, exerting significant influence on key factors such as the identification of congestion zones, the evaluation of traffic flow dynamics, and the construction of origin-destination matrices. An integral component of transportation analyses involves considering land use issues, encompassing factors like land use mix and zoning, which offer valuable insights into the diverse activities occurring in different regions of an area throughout the day (Levine, 2010). Understanding the spatial layout becomes crucial in comprehending the current arrangement of streets (Zhang et al., 2019). The existence of densely populated commercial zones, for example, can have a profound impact on traffic flow and congestion levels (Waloejo, 2020; Yap et al., 2022). The geographical distribution of such commercial hubs directly influences the movement patterns of vehicles, contributing to the formulation of effective congestion mitigation strategies. Beyond commercial zones, residential areas also play a pivotal role in shaping transportation patterns. The significant increase in demand of urbanization due to rapidly increasing population asks for development of residential land uses (Mirzahossein et al., 2022). Suburban residential zones, characterized by lower population densities, often necessitate increased travel for work-related purposes, consequently elevating trip generation rates. This dynamic interaction between land use and travel behavior underscores the importance of accurately characterizing the types of buildings and land uses within an urban area. Moreover, land use intricately affects the accessibility and connectivity of different regions within an urban landscape. The densification of transport nodes orginating form various facilities comprising work, educational, recreational, etc are influenced by the transit oriented developments (TOD) (Agyemang et al., 2020). The configuration of roads, proximity of transportation hubs, and the distribution of various land uses collectively contribute to the overall traffic patterns and transportation network efficiency. Accurate and detailed classification of building types and their associated land uses becomes instrumental in conducting transportation analyses with added efficiency. It facilitates a more comprehensive understanding of the complex interplay between land use and transportation dynamics, enabling urban planners and policymakers to make informed decisions that contribute to optimising transportation systems and improving overall urban mobility. Travel Analysis Zones (TAZs) are fundamental transportation modeling and analysis building blocks. Their design significantly impacts the quality and accuracy of insights derived from such models. A common approach utilizes established Census geographies like blocks, block groups, and tracts as base units for TAZ delineation. This ensures data availability and compatibility with other sources (You et al., 1998; Moghaddam et al., 2022). Another method emphasizes accessibility to transportation infrastructure like roads, rail lines, and transit stations. TAZs are formed based on travel times, distance to key network elements (Ducruet & Lugo, 2013). TAZs can be generated focusing on the area's homogeneity in landuses and socioeconomic characteristics. This helps to understand the travel patterns associated with residential areas, commercial districts and industrial zones (Adams & Tiesdell, 2012). Factoring in multiple determinants like geography, networks, and landuse, a more nuanced representation of travel behaviour can be achieved (Yang et al., 2022). The primary goal of this study is to illustrate how outputs generated by a deep learning-based approach, as described here, can be effectively applied in land-use and transportation planning. By leveraging advanced deep learning techniques, this study aims to produce high-quality building classification data that has information about residential and non-residential structures.

Accurate building classification is crucial for informed decision-making in urban planning, as it provides valuable insights into the distribution and types of buildings within an area. This information can then support various applications, such as optimizing public transportation routes, identifying areas suitable for commercial development, managing zoning regulations, and planning infrastructure improvements. Using a deep learning-based model, the study seeks to offer a more efficient and automated way to generate detailed, up-to-date information, which is often challenging and time-consuming to collect using traditional methods.

2. Methodology

This study aims to generate high-quality building classification data capable of distinguishing between residential and non-residential structures. Accurate building classification is essential for urban planning and management, as it offers valuable insights into the spatial distribution and types of buildings in a given area. This detailed information plays a pivotal role in supporting various applications, such as optimizing public transportation routes, identifying areas suitable for commercial development, managing zoning regulations, and planning for infrastructure improvements. We aim to develop a more efficient and automated approach to produce up-to-date building utility data using a deep learning-based model. This method addresses the challenges of traditional data collection, which is often labor-intensive, time-consuming, and prone to inaccuracies. The deep learning model employed in this study processes satellite or aerial images of a given area to identify all buildings in the image. Once detected, the model classifies each identified building into residential or non-residential categories. The resulting output is a detailed image highlighting the buildings and their respective classifications. This provides a clear and easy-to-interpret representation of the area. This classified building data can then generate zone-specific maps and be helpful for more informed decision-making for urban planners, policymakers, and transportation authorities. Additionally, to provide a comprehensive understanding of the deep learning model used in this study, we have included a brief discussion of its architecture, working principles, performance metrics, and potential limitations. This insight offers a holistic view of how the model functions and the extent to which it can contribute to urban and transportation planning initiatives. The YOLOV8 (Jocher et al., 2023) model undergoes comprehensive training utilizing datasets sourced from SpaceNet-3 (Fig.2) and ISPRS Potsdam (Fig.3), the latter provided by BSF Swissphoto. This mixing of diverse datasets enriches the model's learning with a broader range of real-world scenarios and geospatial contexts. To facilitate the training process, the complete dataset is partitioned into three subsets: Train, Test, and Validation. Notably, modifications are systematically introduced to both the model architecture and the training dataset to optimize performance. These modifications aim to refine the model’s ability to accurately identify and categorize objects, particularly in the context of geospatial imagery. The iterative nature of the training process involves continuous evaluations using the Validation dataset after each epoch. This approach ensures that the model’s learning progresses controlled, allowing for adjusting parameters and fine-tuning to achieve optimal performance. The convergence of training and validation accuracy to predefined criteria is a crucial checkpoint. Once the model consistently meets the desired performance standards on both the training and validation sets, it undergoes a rigorous evaluation using previously unseen images from the Test dataset. This final assessment gauges the model's generalization capabilities and effectiveness in accurately detecting and classifying objects in real-world scenarios outside the training data. To showcase the model performance and the utilization of the model, two datasets are used such as, SpaceNet 3 Vegas and ISPRS Potsdam. As the model is trained on images from different sensors (both aerial and satellite sensors), the learning of the model becomes robust due to different scales, aspects, and pixel values. The SpaceNet 3 dataset (Fig.2), originally meant for extracting road networks from satellite images, contains a diverse range of buildings. This diversity is highlighted as an important feature for improving the training of deep learning models. The size of the images (1300x1300 pixels) and the high spatial resolution (30 cm per pixel) are provided, emphasizing the data quality for detailed analysis. The ISPRS Potsdam dataset

(Fig.3) is a widely recognized benchmark in remote sensing and aerial image analysis. This dataset provides high-resolution aerial imagery of the city of Potsdam, Germany. It is specifically designed to facilitate research in tasks such as semantic segmentation, object detection, and classification within urban environments. The dataset includes true orthophotos with a ground sampling distance (GSD) of 5 cm, allowing for the precise identification and analysis of objects. The diversity and high level of detail in the ISPRS Potsdam dataset can be a resource for researchers and practitioners to use in advanced image analysis algorithms for urban area mapping. In essence, the training and evaluation methodology outlined underscores a meticulous and iterative process involving diverse datasets and continuous assessments to ensure the YOLOV8 model is finely tuned and proficient in its geospatial object detection tasks. The utilization of separate Train, Test, and Validation sets, along with the adoption of performance criteria, establishes a robust framework for achieving and validating the desired level of accuracy and reliability in the model's predictions. The work demonstrated in this research is to execute a nuanced classification of building rooftops into two discrete categories: residential and non-residential. This classification is pivotal for the subsequent allocation of appropriate zones, a process contingent on the density of structures within a specific geographical area. The zoning strategy is straightforward: if a given region predominantly features residential structures, it is designated as a residential zone; conversely, areas not intended for residential purposes are classified accordingly. The designation of a mixed zone pertains to areas showcasing a relatively equitable distribution of various building types.

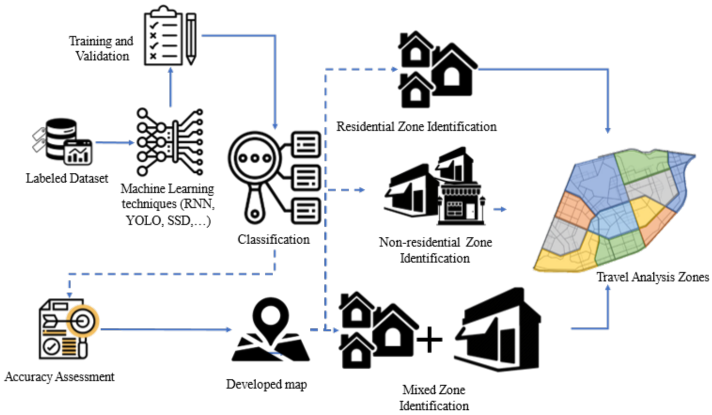

Fig.1 Methodology diagram

The utilization of zonal data forms the cornerstone for generating a foundational map of the targeted region, serving as a crucial resource for diverse transportation planning analyses. Fig.1 encapsulates the entire methodology, visually representing the comprehensive process. The validation accuracy, a crucial metric for assessing the model's performance, is presented in Tab.2, offering insights into the reliability and eff'ctiveness of the classification and segmentation techniques employed. Additionally, Figure 5 offers illustrative instances of the test findings, providing visual validation of the model's capabilities. The resultant images undergo a georeferencing step following the classification and segmentation processes. This georeferencing procedure aligns the images geographically, establishing the groundwork for the construction of Transportation Analysis Zones (TAZ) base-maps. Notably, all photos are standardized by downsizing them to square images with a resolution of 960 pixels, a preparatory step to enhance training efficiency and accuracy. The execution of the entire model pipeline, from classification to georeferencing, is conducted on the Python platform. This implementation is supported by an array of libraries, including but not limited to numpy, opencv-python, pytorch, CUDNN, and others. This technological framework ensures the seamless execution of the research methodology, leveraging advanced computational capabilities to achieve accurate and insightful results. In summary, the research encompasses a multifaceted approach, integrating image processing, geospatial analysis, and deep learning techniques, all executed on a robust computational platform to accomplish the study's overarching objective of refined building rooftop classification for transportation planning.

(a) (b)

Fig.2 (a) Examples of residential buildings (SpaceNet-3); (b) Examples of non-residential buildings (SpaceNet-3)

(a) (b)

Fig.3 (a) Examples of residential buildings (ISPRS); (b) Examples of non-residential buildings (ISPRS)

YOLOV8, as illustrated in Fig.4, inherits a foundational architecture akin to its precursor, YOLOV5, while introducing augmentations that contribute to the refinement of its training and prediction capabilities. Notably, a pivotal enhancement incorporated into this model is integrating a C2F module (Terven & Cordova-Esparza, 2023). The C2F module functions as a cross-stage partial bottleneck featuring two convolution modules. This architectural addition enables YOLOV8 to amalgamate high-level features containing critical information, thereby facilitating an augmentation in detection accuracy and a reduction in box loss.

One of the distinctive attributes of the model architecture lies in incorporating a dual decoupled head, coupled with the absence of anchors. This unique configuration empowers the model to autonomously handle diverse tasks, encompassing detection, classification, and regression. Eliminating anchors enhances the model's adaptability in addressing myriad scenarios, improving overall performance. Moreover, the architecture embraces various design branches, each assigned specific roles to improve overall correctness. A noteworthy aspect of YOLOV8 is its use of the sigmoid function to predict the probability of a bounding box surrounding a targeted object. Simultaneously, the softmax function is employed to compute the probability of a detected object belonging to a designated class. This dual-functionality approach enhances the model's precision and reliability in both localizing and categorizing objects within the given context. The meticulously crafted architecture, including these features, collectively positions YOLOV8 as an advanced and versatile model proficient in multifaceted object detection tasks.

The YOLOV8 model is highly adaptable and capable of training and making predictions on images or videos of any size. While it comes with a default frame size of 640x640 pixels for training and testing, users can adjust this frame size according to their specific requirements. Generally, high-resolution images involve larger frame sizes, which can enhance the accuracy and performance of the model during both the training and inference stages. However, this improvement comes at the cost of increased computational resource usage. The YOLOV8 model is available in multiple variants, ranging from YOLOV8n (nano) to YOLOv8x (extra-large), each with different numbers of parameters. These variants determine how complex and detailed the model's learning process is. Models with more parameters, such as YOLOv8x, tend to achieve better learning and produce more accurate results, but they also demand significantly more computational resources and longer training time. It's recommended to use a GPU-enabled environment for optimal performance, as this significantly speeds up the training and inference processes. Although the model can also run on CPUs, using a GPU accelerates the computations, making it more efficient, especially for larger datasets or higher-resolution images.

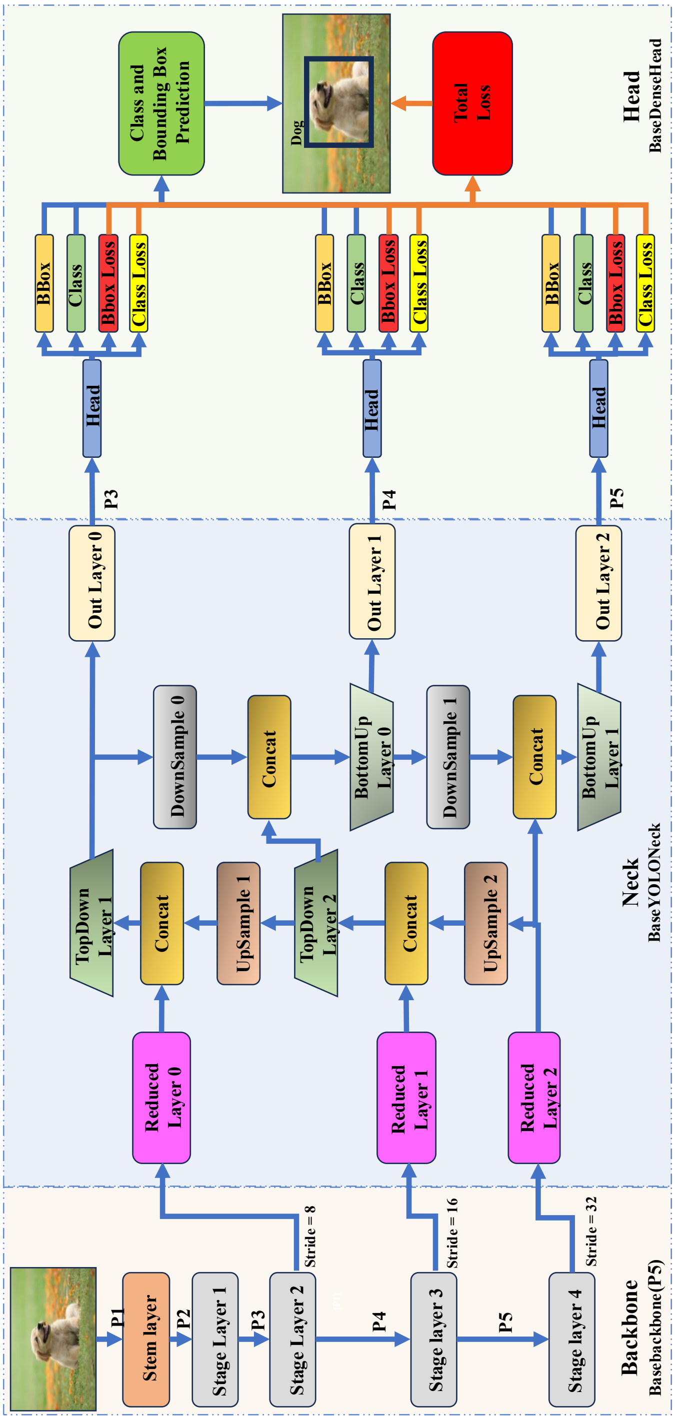

Fig.4 YOLOV8 Base architecture

SL No.	Metric	Formula
1	Mean Avearage Precision (mAP)	$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i$
2	Precision (P)	$\frac{TP}{TP + FP}$
3	Recall (R)	$\frac{TP}{TP + FN}$
4	Accuracy	$\frac{TP + TN}{TP + FN + TN + FN}$

Tab.1 Perfromance metrics details

3. Results and discussion

The YOLOV8 model architecture stands out for its versatility, demonstrating proficiency in a spectrum of computer vision tasks, including but not limited to segmentation, object identification, picture classification, object tracking, and posture estimation. This broad applicability has contributed significantly to the model's popularity, attributed to its simplistic design, user-friendly interface, and adaptability to various data input formats.

In the context of this study, the YOLOV8 model architecture is leveraged for the specific task of building classification. The dataset employed encompasses over 500 images, with approximately 300 sourced from the SpaceNet 3 dataset and the remaining images derived from the Potsdam dataset. In preparation for the training phase, all original images undergo a meticulous image augmentation process to enhance the model's training accuracy by introducing diverse variations to the dataset. The optimization process is facilitated by the implementation of Stochastic Gradient Descent (SGD) as the optimizer, with an initial learning rate set at 0.007. To ensure stability and effective convergence during training, the momentum is finely tuned to a value of 0.937. A judicious choice of hyperparameters is also made, with the weight decay parameter set at 0.0005. These optimization parameters collectively contribute to the model's ability to learn and generalize effectively from the training dataset. Tab.2 is enlisted to provide a comprehensive overview of the performance metrics achieved by the YOLOV8 model in the course of its application to the building classification task. This table serves as a valuable reference point for evaluating the efficacy and accuracy of the model in relation to the specific objectives outlined in the study.

Class	Precision	Recall	mAP(50)
Overall	0.944	0.892	0.945
CB	0.936	0.805	0.903
RB	0.953	0.979	0.986

Tab.2 Perfromance metrics. CB = Non-residential Building; RB = Residential Building

The primary objective of this study is the nuanced categorization of building rooftops into distinct classes, specifically differentiating between residential and non-residential structures. This categorization serves as a pivotal step in the subsequent assignment of appropriate zones based on the density of buildings within a given geographic area. The zoning process operates on the premise that areas predominantly characterized by residential structures are designated as residential zones, while those dominated by non-residential structures are categorized as non-residential zones. A mixed zone designation is applied to areas where the proportion of each building type is relatively balanced. This zoning methodology lays the foundation for the creation of a detailed and informative map of the targeted area. By systematically incorporating the zonal data into the mapping process, the resulting cartographic representation provides a comprehensive overview of the spatial distribution of residential and non-residential structures. This map, enriched with zonal insights, serves as a valuable tool for conducting various transportation planning studies. The utility of this approach extends beyond mere categorization, as the zoning information encapsulates the socio-economic and land-use characteristics of the area. Such a detailed understanding of the spatial composition becomes instrumental in making informed decisions regarding transportation planning, urban development, and resource allocation. The methodological framework outlined in this study not only contributes to building classification but also facilitates a broader understanding of the built environment, thereby enhancing the applicability of the generated maps for strategic urban planning initiatives.

In order to create a zone-dispersed map, it is very essential to incorporate geographic information. This will enable the alignment of images with a spatial reference. In order to include spatial information in the resulting segmented image, the geographical information is extracted from the input images and used as a reference for each individual image. Fig.6 illustrates scenes derived from the categorized and segmented photos reflecting the datasets utilized in this investigation.

Fig.5 Examples of detection results (Cyan: Non-residential buildings; Yellow: Residential buildings)

Instance segmentation is used in this study as the desired objects are identified, classified, and segmented. Any instance segmentation model requires polygon annotations for training, and the polygons are made meticulously to produce better annotations, yield better training accuracy. Manual preparation of annotation in terms of the polygons is labour-intensive and time-consuming, hence automatic or semi-automatic mask generation algorithms are gaining popularity. Most of the instance segmentation models' work pace is also contingent on the hardware capabilities, so dealing with more pixel density calls for better computational power.

The model used in this study relies on pixel-based information, and to enhance the learning, size, shape, aspect, and scale of the targeted objects (here, rooftops of the buildings) are also considered. To confirm building types, every building from every image in the training dataset was checked by overlaying the images on the Googe map. To ensure better learning and minimal occurrence of classification as well as segmentation error, image augmentations such as rotation (90˚), flip (both vertical and horizontal) are applied. While testing the model's performance, it was found that residential buildings with slender appearance are mostly misclassified as non-residential buildings; to mitigate that, reannotation of such buildings is done.

Fig.6 Examples of detection results and land use map preparation

The presented illustration (Fig.7) servs as a visual representation of the intricate process of generating pertinent cartographic information. The focal point of this process involves the careful selection of scenes from the SpaceNet 3 Vegas dataset. These chosen scenes are strategically employed to form a composite image of larger dimensions. This composite image, representative of a broader spatial context, is subsequently introduced into the model pipeline. Within the model pipeline, the composite image undergoes a series of computational steps and analyses, culminating in generation of output. This output, derived encapsulates valuable cartographic information with enhanced contextual insights. The resulting image, now enriched with the extracted data, is purposefully employed in delineating and categorizing distinct zones within the geographic are under consideration.

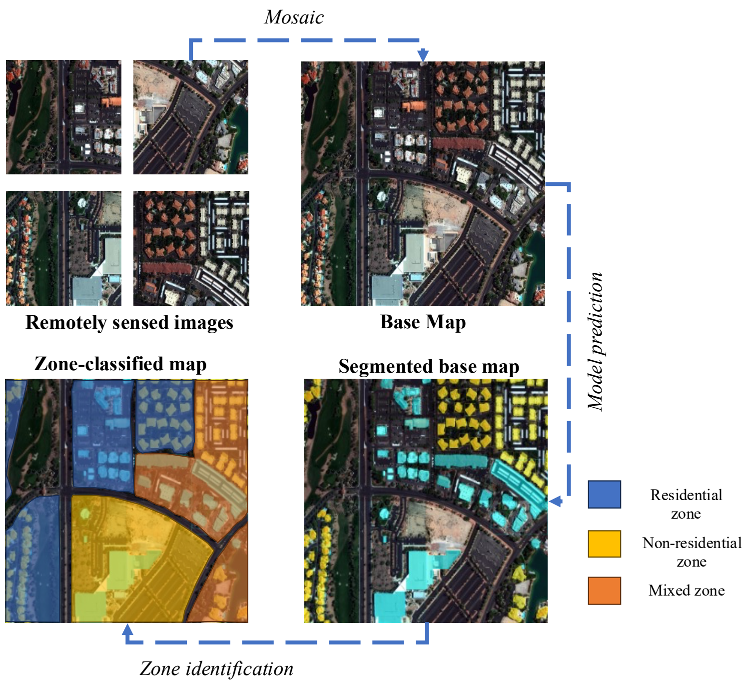

Fig.7 Insert here image description

The zones identified in the resulting image are further classified into residential, non-residential, and mixed zones. This classification process contributes to the creation of comprehensive cartographic representation, effectively conveying the spatial distribution characteristics of different structures.

The study's initial phase involved using georeferenced images from both datasets. The inherent georeferencing of these images facilitated a seamless alignment process, enabling the synthesis of accurate maps. Georeferencing is a crucial step in geospatial analysis, as it establishes the spatial location and positioning of features within an image in relation to real-world coordinates. It is imperative to note that standard instance segmentation models inherently lack the capability to transfer geospatial information from the input image to the output image. Recognizing this limitation, a subsequent step was introduced to address the geospatial integrity of the output. Consequently, all output images underwent an additional georeferencing process, aligning them with respect to the geospatial information present in the input images (Khatua et al., 2024). This meticulous georeferencing of the output images is paramount for maintaining spatial accuracy and ensuring that the generated maps faithfully represent the geographical features depicted in the input imagery. The alignment process contributes to the preservation of geospatial relationships, allowing for meaningful comparisons and analyses across different datasets and facilitating further geospatial operations with precision.

The YOLOV8 model's segmentation effectiveness relies on its detection capability. This means the segmentation quality is affected by the detector models. It outlines three key challenges related to using bounding boxes: overlap, dataset size, and unnecessary pixels. Overlapping occurs when te objects and close together, leading to potential confusion during training. Narrow objects might results in bounding boxes containing many irrelevant pixels., which can interfere the model's learning process by blending the target object with nearby pixels.

In summary, the entire workflow underscores the critical role of georeferencing in our study. The georeferenced images serve as the foundation for accurate mapping, and the subsequent alignment of output images ensures the preservation of geospatial context, enhancing the reliability and utility of the generated cartographic information.

4. Conclusion

The study delves into the intricate realm of instance segmentation, leveraging its capabilities for identifying, classifying, and segmenting desired objects, with a specific focus on rooftops in building structures. The process involves meticulous polygon annotations for model training, a traditionally labour-intensive and time-consuming task. To alleviate this, this study explores automatic or semi-automatic mask generation algorithms. The model adopted relies on pixel-based information, factoring in the targeted objects' size, shape, aspect and scale, and incorporates image augmentation for enhanced learning. However, challenges surface during testing, particularly in the misclassification of slender residential buildings as non-residential buildings, prompting the need for reannotation to mitigate errors.

Currently, this model has been tested only on planned urban environments where buildings exhibit distinct and easily recognizable features. These well-defined attributes make it easier for the model to learn and accurately detect objects, resulting in enhanced performance. However, the model's capabilities and additional limitations may become more apparent when applied to rural areas or urban settings characterized by organic and irregular development. In such environments, the architectural features are less structured, and buildings vary widely in shape, size, and arrangement, posing a greater challenge for accurate detection and segmentation. To thoroughly evaluate the model's adaptability and robustness in these complex scenarios, it will be necessary to train and test it on additional datasets that capture the diversity of rural and organically developed urban landscapes. Expanding this evaluation to include such datasets represents an important direction for future research, as it will help identify potential shortcomings and guide further improvements in the model's performance across different geographic settings.

The study addresses the critical aspect of georeferencing, acknowledging the inherent limitation of instance segmentation models in transferring geospatial information. The georeferencing of images, both input and output, is meticulously executed to ensure accurate alignment, fostering meaningful comparisons and analyses. This process is pivotal for maintaining spatial accuracy, enhancing the reliability of the generated maps, and enabling effective geospatial operations.

In summary, this research contributes to the advancement of instance segmentation methodologies and showcases the integration of geospatial considerations in cartographic information generation. The proposed workflow emphasizes the synergy between advanced computer vision techniques, spatial analysis, and georeferencing for robust and accurate results in diverse applications, from object identification to cartographic mapping.

References

Aamir, M., Pu, Y. F., Rahman, Z., Tahir, M., Naeem, H. & Dai, Q. (2018). A framework for automatic building detection from low-contrast satellite images. *Symmetry, 11* (1), 3. https://doi.org/10.3390/sym11010003

Adams, D. & Tiesdell, S. (2012). *Shaping places: Urban planning, design and development*. Routledge. https://doi.org/10.4324/9780203105665

Agyemang, K. K., Amoako-Sakyi, R. O., Antwi, K. B., Mensah, C. A. & Abane, A. M. (2020). Transit-oriented development: Theory and implementation challenges in Ghana. *TeMA - Journal of Land Use, Mobility and Environment, 13* (3), 409-425. https://doi.org/10.6092/1970-9870/6800

Aithal, B. H. & Ramachandra, T. V. (2016). Visualization of urban growth pattern in Chennai using geoinformatics and spatial metrics. *Journal of the Indian Society of Remote Sensing, 44*, 617-633. https://doi.org/10.1007/s12524-015-0535-0

Badrinarayanan, V., Kendall, A. & Cipolla, R. (2017). SegNet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 39* (12), 2481-2495. https://doi.org/10.1109/TPAMI.2016.2644615

Bharath, H. A., Chandan, M. C., Vinay, S. & Ramachandra, T. V. (2017). Modelling the growth of two rapidly urbanizing Indian cities. *Journal of Geomatics, 11* (12), 149-166.

Carpentieri, G. & Favo, F. (2017). The end-use electric energy consumption in urban areas: A GIS-based methodology. An application in the city of Naples. *TeMA - Journal of Land Use, Mobility and Environment, 10* (2), 139-156. https://doi.org/10.6092/1970-9870/5173

Chen, C., Gong, W., Chen, Y. & Li, W. (2018). Learning a two-stage CNN model for multi-sized building detection in remote sensing images. *Remote Sensing Letters, 10* (2), 103-110. https://doi.org/10.1080/2150704X.2018.1528398

Chen, X., Qiu, C., Guo, W., Yu, A., Tong, X. & Schmitt, M. (2022). Multiscale feature learning by transformer for building extraction from satellite images. *IEEE Geoscience and Remote Sensing Letters, 19*, 1-5. https://doi.org/10.1109/LGRS.2022.3150085

Ducruet, C. & Lugo, I. (2013). Structure and dynamics of transportation networks: Models, methods and applications. In *The SAGE handbook of transport studies*, 347-364.

Galderisi, A., Guida, G. & Limongi, G. (2021). Emergency and spatial planning towards cooperative approaches. *TeMA - Journal of Land Use, Mobility and Environment*, 73-92. https://doi.org/10.6092/1970-9870/7466

Girshick, R. & Fast, R. C. N. N. (2015). Microsoft Research. *Fast R-CNN, 27.* https://doi.org/10.13140/RG.2.2.30558.38729

Gazzola, V., Menoni, S., Belloni, A. & Zuliani, C. (2021). Territorial aspects of emergency plans for dams. The case study of Lombardia Region. *TeMA - Journal of Land Use, Mobility and Environment*, 93-108. https://doi.org/10.6092/1970-9870/7467

He, K., Zhang, X., Ren, S. & Sun, J. (2015). Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 37* (9), 1904-1916. https://doi.org/10.1109/TPAMI.2015.2389824

He, K., Zhang, X., Ren, S. & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770-778. https://doi.org/10.1109/CVPR.2016.90

Ji, S., Wei, S. & Lu, M. (2018). 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. https://doi.org/10.1109/TGRS.2018.2859200

Jocher, G., Chaurasia, A. & Qiu, J. (2023). YOLO by Ultralytics (Version 8.0.0) [Computer software]. https://github.com/ultralytics/ultralytics

Khatua, A., Bhattacharya, A. & Aithal, B. H. (2024). Automated georeferencing and extraction of building footprints from remotely sensed imagery using deep learning. In *GISTAM*, 128-135.

Krizhevsky, A., Sutskever, I. & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM, 60* (6), 84-90. https://doi.org/10.1145/3065386

Levine, J. (2010). *Zoned out: Regulation, markets, and choices in transportation and metropolitan land use*. Routledge.

Li, J. & Cao, J. (2019). A framework for automatic building detection from low-contrast VHR satellite imagery. In *Proceedings of the 3rd International Conference on Video and Image Processing*, 52-56. https://doi.org/10.1145/3377325.3377329

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y. & Berg, A. C. (2016). SSD: Single shot multibox detector. In *Computer Vision-ECCV 2016*, 21-37. Springer. https://doi.org/10.1007/978-3-319-46448-0_2

Madhumita, D., Bharath, H. A., Devendra, V. P. & Shivam, B. (2023). Road segmentation: Exploiting the efficiency of skip connections for efficient semantic segmentation. *Journal of South Asian Logistics and Transport, 3* (1), 59-73. https://doi.org/10.4038/jsalt.v3i1.60

Moghaddam, S. H. M., Ameli, M., Rao, K. R. & Tiwari, G. (2022, November). Delineation of traffic analysis zone for public transportation OD matrix estimation based on socio-spatial practices. In *The 4th Symposium on Management of Future Motorway and Urban Traffic Systems 2022 (MFTS2022)*.

Mirzahossein, H., Noferesti, V. & Jin, X. (2022). Residential development simulation based on learning by agent-based model. *TeMA - Journal of Land Use, Mobility and Environment, 15* (2), 193-207. https://doi.org/10.6092/1970-9870/7435

Pavlidis, T. & Liow, Y. T. (1989). Integrating methodologies in image analysis. In *Computer Vision and Shape Recognition*, 429-437.

Prakash, P. S., Soni, J. & Bharath, H. A. (2022). Building extraction from remote sensing images using deep learning and transfer learning. In *IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium* (pp. 3079-3082). IEEE. https://doi.org/10.1109/IGARSS46834.2022.9884704

Raju, S. R., Wajid, S., Radhakrishnan, N. & Mathew, S. (2020). Accessibility analysis for healthcare centers using gravity model and geospatial techniques. *TeMA - Journal of Land Use, Mobility and Environment, 13* (1), 6-20. https://doi.org/10.6092/1970-9870/6860

Ramachandra, T. V., Aithal, B. H. & Sanna, D. D. (2012). Insights to urban dynamics through landscape spatial pattern analysis. *International Journal of Applied Earth Observation and Geoinformation, 18*, 329-343. https://doi.org/10.1016/j.jag.2012.01.017

Ramachandra, T. V., Bharath, A. H. & Sowmyashree, M. V. (2015). Monitoring urbanization and its implications in a mega city from space: Spatiotemporal patterns and its indicators. *Journal of Environmental Management, 148*, 67-81. https://doi.org/10.1016/j.jenvman.2014.03.015

Redmon, J., Divvala, S., Girshick, R. & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 779-788. https://doi.org/10.1109/CVPR.2016.91

Simonyan, K. & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint* arXiv:1409.1556. https://doi.org/10.48550/arXiv.1409.1556

Sirmacek, B. & Unsalan, C. (2008). Building detection from aerial images using invariant color features and shadow information. In *2008 23rd International Symposium on Computer and Information Sciences*, 1-5. IEEE. https://doi.org/10.1109/ISCIS.2008.4717927

SpaceNet on Amazon Web Services (AWS). (2018). "Datasets." *The SpaceNet Catalog*. https://spacenet.ai

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan,D., Vanhoucke, V. & Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1-9. https://doi.org/10.1109/CVPR.2015.7298594

Terven, J. & Cordova-Esparza, D. (2023). A comprehensive review of YOLO: From YOLOv1 and beyond. *arXiv preprint* arXiv:2304.00501. https://doi.org/10.48550/arXiv.2304.00501

Tira, M. (2021). Planning to prevent disasters. *TeMA - Journal of Land Use, Mobility and Environment*, 191-202. https://doi.org/10.6092/1970-9870/7468

Waloejo, B. S. (2020). The effect of commercial areas and industrial zone improvement on road service levels between city Surabaya-Sidoarjo. In *2nd International Symposium on Transportation Studies in Developing Countries (ISTSDC 2019)*, 31-35. Atlantis Press. https://doi.org/10.2991/aebmr.k.200220.007

Wang, H. & Miao, F. (2022). Building extraction from remote sensing images using deep residual U-Net. *European Journal of Remote Sensing, 55* (1), 71-85. https://doi.org/10.1080/22797254.2022.2034778

Yang, B., Tian, Y., Wang, J., Hu, X. & An, S. (2022). How to improve urban transportation planning in big data era? A practice in the study of traffic analysis zone delineation. *Transport Policy, 127*, 1-14. https://doi.org/10.1016/j.tranpol.2022.07.009

Yang, H. L., Yuan, J., Lunga, D., Laverdiere, M., Rose, A. & Bhaduri, B. (2018). Building extraction at scale using convolutional neural networks: Mapping of the United States. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 11* (8), 2600-2614. https://doi.org/10.1109/JSTARS.2018.2837673

Yap, J. Y., Omar, N. & Ismail, I. (2022). A study of traffic congestion influenced by the pattern of land use. In *IOP Conference Series: Earth and Environmental Science* (Vol. 1022, No. 1, p. 012035). IOP Publishing. https://doi.org/10.1088/ 1755-1315/1022/1/012035

You, J., Nedović-Budić, Z. & Kim, T. J. (1998). A GIS-based traffic analysis zone design: Technique. *Transportation Planning and Technology, 21* (1-2), 45-68. https://doi.org/10.1080/03081069808717596

Zhang, N., Zhao, X., Liu, T., Lei, M., Wang, C. & Wang, Y. (2019). Layout planning of highway transportation environment monitoring network: The case of Xinjiang, China. *Sustainability, 12* (1), 290. https://doi.org/10.3390/su12010290

Zullo, F., Paolinelli, G., Fiordigigli, V., Fiorini, L. & Romano, B. (2015). Urban development in Tuscany: Land uptake and landscapes changes. *TeMA - Journal of Land Use, Mobility and Environment, 8* (2), 183-202. https://doi.org/10.6092/1970-9870/2864

Image Sources

SpaceNet on Amazon Web Services (AWS). "Datasets." The SpaceNet Catalog. Last modified April 30, 2018. Accessed on Februrary 2023. https://SpaceNetchallenge.github.io/datasets/datasetHomePage.html

Fig.2: Examples of residential buildings (SpaceNet-3) (b) Examples of non-residential buildings (SpaceNet-3). 2d semantic

labeling contest - potsdam. International Society for Photogrammetry and Remote Sensing.

Fig.3: Examples of residential buildings (ISPRS Potsdam) (b) Examples of non-residential buildings (ISPRS Potsdam)

Author's profile

Aniruddha Khatua

Research Scholar at Ranbir and Chitra Gupta School of Infrastructure Design and Management, Indian Institute of Technology Kharagpur. His research is mainly oriented towards feature extraction from remotely sensed images.

Arkopal Kishore Goswami

Professor at Ranbir and Chitra Gupta School of Infrastructure Design and Management, Indian Institute of Technology Kharagpur. Dr. Goswami is leading the Multimodal Urban Sustainable Transportation (MUST) research group at IIT Kharagpur. The primary research areas of the group include, Urban Transport Sustainability, Urban Travel Behavior, Seamless Integration of Urban Travel Modes, Active Transportation and Impact on Human Health, and Smart Management of Transport Infrastructure. His research program aims to develop the next generation of transport infrastructure professionals, who would be equipped with a holistic understanding of the impacts of transportation on liveability.

Bharath Haridas Aithal

Professor at Ranbir and Chitra Gupta School of Infrastructure Design and Management, Indian Institute of Technology Kharagpur. Dr. Aithal's method and theoretical research as well as a considerable portion of his applied work addresses Spatial and longitudinal data. An innovative contribution of his work is to show that agents that change the urban aspects of a urban social life are major players that have to be considered, while modelling in urban domain. This has established a new perspective to the analysis of urban spatial data and modelling. This perspective creates a new class of functional estimation procedures that non-parametrically account for the within-subject correlation in an intuitive and efficient manner. Thus providing a space for more efficient algorithmic modelling of urban developments and urban sociology.