ONTOLOGY-GUIDED YOLOV8 FOR SEMANTIC OBJECT DETECTION AND SCENE INTERPRETATION IN REMOTE-SENSING SMART-CITY ENVIRONMENTS

Baojia Gong

Object detection in remotely sensed satellite imagery is increasingly important for urban planning, disaster management, and environmental monitoring in smart-city settings. This manuscript presents a coherent and publication-ready account of an ontology-guided deep learning framework that integrates a lightweight YOLOv8 detector with an ontology reasoning module for semantic scene interpretation. The system is designed to detect five urban-environment classes—residences, roads, shorelines, swimming pools, and vegetation—from Sentinel-2 MSI imagery collected over the southern Durban metropolitan region of KwaZulu-Natal, South Africa. The dataset consists of 92 annotated images resized to 640×640 *pixels, partitioned into 61 training, 21 validation, and 10 testing images, then augmented to 6,100 training, 2,100 validation, and 1,000 testing images. The visual recognition component employs a YOLOv8 architecture with a C2f-based backbone/neck design and anchor-free detection heads, while the semantic layer uses RDF/OWL concepts queried through SPARQL to represent hierarchical class relations, object adjacency, and interpretable scene semantics. On the proposed dataset, the YOLOv8 model attains 68% precision, 60% recall, 43% mAP@50, and 17.5% mAP@50–95, with the highest class-specific precision observed for swimming pools (62.7%) and the highest class-specific mAP@50 for shorelines (99.5%). The ontology remains lightweight and scalable, with a maximum depth of inheritance of 3 and a maximum number of children of 4, enabling efficient reasoning with low computational demand. By combining object detection with structured semantic inference, the framework provides an interpretable analytical layer for smart-city land-cover understanding, disaster-aware urban monitoring, and knowledge-driven scene analysis.*

Index Terms — smart cities; urban planning; remote sensing; object detection; image analysis; ontology; knowledge representation; semantic reasoning; YOLOv8; disaster management

https://doi.org/10.66033/judsc2025-220

INTRODUCTION

Remote-sensing image analysis has become a foundational component of contemporary smart-city systems because it enables large-scale, repeatable, and data-driven observation of the urban environment. High-resolution satellite and aerial imagery now support a wide range of urban functions, including land-cover mapping, infrastructure interpretation, environmental monitoring, informal-settlement detection, transportation assessment, and disaster-aware planning [1, 2, 3]. In rapidly changing and spatially heterogeneous urban regions, such capabilities are especially important because city managers and planners require timely spatial intelligence that can be updated more efficiently than conventional field-based observation alone. As smart-city governance increasingly depends on integrated digital platforms, remote-sensing analytics provides a critical observational layer through which urban change, ecological stress, and infrastructure conditions can be monitored and interpreted [4, 5, 6].

In dense urban environments, however, scene understanding requires more than the isolated recognition of visible objects. Urban space is relational by nature: roads intersect built structures, vegetation clusters around residential and institutional zones, waterways shape settlement patterns, and open land often signals transitional or vulnerable forms of development. The interpretation of such scenes therefore depends not only on detecting individual elements, but also on understanding how those elements coexist, interact, and acquire meaning within their broader spatial context [7, 8]. This contextual requirement is particularly relevant to the aims of research in urban development and smart cities, where automated image understanding is expected to support evidence-based planning, resilience assessment, environmental governance, and policy design rather than merely produce technically accurate labels. A model that detects an object without clarifying its urban significance may offer computational value, but a model that supports structured scene interpretation is far more useful for planning and governance applications [9, 10].

Traditional scene-detection and classification methods in remote sensing relied heavily on manual feature extraction, expert-designed descriptors, and handcrafted decision rules. These approaches contributed important early advances, especially in settings where domain knowledge could be translated into spectral, textural, or geometric signatures. Nevertheless, they often struggle with scalability, transferability, and the complex spatial and spectral variability that characterizes real urban imagery [2, 4, 11]. Urban scenes frequently contain mixed land covers, irregular object boundaries, variable shadow effects, occlusion, and substantial intra-class diversity. As a result, rule-based or manually engineered pipelines may perform well in constrained conditions but lose robustness when applied to larger or more heterogeneous urban datasets. These limitations have encouraged a strong methodological shift toward deep learning, which can learn higher-level visual representations directly from data and thereby reduce dependence on manual feature design [5, 6, 12].

Deep learning methods, particularly convolutional neural networks and one-stage object detectors, have substantially improved the automatic extraction of discriminative features from remote-sensing imagery. Architectures in the YOLO family, for example, offer attractive trade-offs between speed and detection accuracy, making them especially relevant for smart-city settings in which image analysis may need to support near-real-time or operational workflows [3, 7, 11]. These models have demonstrated strong performance in detecting urban objects such as roads, buildings, vehicles, and vegetation patterns across complex scenes. Yet, despite these advances, a persistent limitation remains: high-performing detectors do not necessarily produce semantically rich or operationally interpretable outputs. They can often identify what is present in an image, but they are less capable of explaining how detected objects relate to one another in ways that matter for urban analysis. In other words, recognition performance does not automatically yield urban understanding [8, 9, 13].

This limitation becomes more important when remote-sensing outputs are used to support planning, environmental management, or disaster-related interpretation. For example, the co-occurrence of buildings, roads, and sparse vegetation may carry implications that differ substantially from the co-occurrence of buildings,

water-adjacent land, and open spaces. Similarly, the significance of a detected road may depend on whether it is associated with dense built-up fabric, flood-exposed edges, or undeveloped land corridors. In such cases, urban interpretation requires a representational structure capable of encoding conceptual categories, spatial relations, and domain-relevant semantics [4, 6, 10]. A detection-only approach may identify the components of a scene, but it may not adequately formalize the relationships through which those components become meaningful for smart-city analysis.

Ontology-guided learning offers a practical and conceptually strong response to this limitation. Ontologies support structured knowledge representation by organizing concepts, hierarchical relations, semantic properties, and logical constraints in a machine-readable form [5, 8]. When combined with visual detection systems, ontology-based reasoning can extend image analysis beyond the recognition of objects toward the interpretation of structured urban scenes. This is particularly useful in smart-city applications because decision-makers often need outputs that are not only accurate, but also explainable, queryable, and linked to domain concepts that can support further reasoning. Rather than treating an image as a flat collection of detected classes, an ontology-guided framework can represent urban objects as components of a meaningful conceptual system in which roads, buildings, vegetation, water-adjacent features, and open spaces are connected through interpretable relations [7, 9].

In the present study, this integration is operationalized through a framework that combines YOLOv8 object detection with an ontology reasoning module and SPARQL-based querying. The choice of YOLOv8 is motivated by its effectiveness as a modern object detector for complex visual tasks, while the ontology layer provides the semantic structure necessary to transform raw detections into machine-interpretable scene knowledge. SPARQL querying further strengthens the framework by enabling structured retrieval and reasoning over the detected and formalized scene components. Together, these elements create a pipeline that moves from image-based recognition to semantic urban interpretation. This design is especially suited to remote-sensing smart-city environments, where analytical systems must often bridge the gap between automated visual extraction and decision-oriented spatial intelligence [3, 6, 8, 11].

The resulting system is framed around three core objectives. The first is the accurate detection of salient urban and environmental objects in satellite imagery. This objective addresses the need for reliable identification of features that are central to land-use interpretation, urban monitoring, and environmental assessment. The second is semantically meaningful scene interpretation through ontology-based knowledge representation. Here, the emphasis shifts from simple detection to structured understanding, enabling urban scenes to be represented in a way that reflects functional and conceptual relations among observed elements. The third objective is improved interpretability for urban planning, environmental monitoring, and disaster-related analysis. This is particularly important because smart-city systems increasingly require outputs that can be integrated into policy workflows, planning support tools, and resilience-oriented information systems rather than remaining confined to purely technical evaluation settings [4, 7, 9, 10, 13].

From the perspective of urban-development research, the contribution of this study lies in linking advances in object detection with formal semantic reasoning in a way that is operationally meaningful. Many remote-sensing studies emphasize predictive performance alone, but urban decision-making also depends on interpretability, conceptual transparency, and the ability to query or explain spatial patterns. By combining deep visual detection with ontology-guided analysis, the present framework contributes to a broader vision of smart-city intelligence in which machine learning is not treated as a black box, but as part of a structured analytical system capable of supporting evidence-based urban governance [2, 12]. This makes the approach relevant not only to image-analysis research, but also to the wider literature on explainable urban AI, environmental knowledge systems, and semantically enriched planning analytics.

The manuscript is organized as follows. Section reviews relevant literature on ontology-guided image

analysis and remote-sensing object detection. Section presents the dataset, model architecture, integration workflow, and ontology formalization. Section reports detection performance and ontology-based analysis. Section discusses the implications of the proposed framework for smart-city applications, with particular attention to planning, environmental monitoring, and disaster-aware interpretation. Section concludes the paper and identifies directions for future research.

RELATED WORK

Deep learning has fundamentally reshaped remote-sensing image analysis by enabling end-to-end learning of complex visual features directly from large image collections. In contrast to earlier approaches based on manually engineered spectral, textural, or geometric descriptors, deep neural architectures can learn hierarchical representations that are more adaptable to the variability of real-world imagery. This shift has been especially important in urban and smart-city contexts, where remotely sensed scenes are often dense, heterogeneous, and difficult to interpret through handcrafted rules alone [6, 7]. As a result, deep learning methods have been widely adopted for urban land-use interpretation, building and road extraction, disaster scene analysis, infrastructure recognition, and environmental monitoring. These advances have significantly improved predictive accuracy and operational scalability, making automated image understanding increasingly viable for smart-city applications [8].

Despite these strengths, purely data-driven models still face an important limitation: they often provide predictions without transparent semantic explanation. A detector may identify roads, buildings, vegetation, or water-adjacent features with high confidence, yet the resulting output may remain a flat list of object labels with little formal account of how those objects relate to one another in an urban scene. For planning, monitoring, and disaster-related decision support, this lack of semantic structure can reduce the practical value of otherwise accurate detections [9]. In smart-city systems, interpretability is not a minor concern, because model outputs are frequently expected to support decisions made by planners, administrators, environmental analysts, and emergency managers. Accordingly, there is growing interest in methods that can connect visual recognition with structured, explainable scene interpretation [13].

Ontology-based methods offer a strong response to this challenge by providing formal representations of concepts, relations, and constraints. In semantic web and knowledge representation research, ontologies have long been used to organize domain knowledge, represent hierarchical relationships, and enable machine reasoning over structured information [4]. Their value lies in the fact that they do not merely store labels; they define how concepts are related, what properties they possess, and how logical queries can be applied to retrieve meaningful knowledge. This makes ontologies especially relevant for image understanding tasks in which the meaning of a scene depends not only on the presence of individual objects, but also on their contextual and conceptual relationships.

In remote sensing, this capability is particularly useful because scene meaning often emerges from configurations of objects rather than from object identity alone. A road detected beside dense built structures may indicate a different urban function than a road near open land or water edges. Likewise, vegetation identified within a residential pattern may serve a different interpretive role than vegetation associated with undeveloped or peri-urban space. Such distinctions are difficult to formalize through detection outputs alone, but they can be represented more effectively through ontology-driven knowledge structures [11]. By capturing conceptual hierarchies and relational patterns, ontology-based approaches can support a richer layer of interpretation over visual data.

Several prior studies have therefore explored ontology-driven frameworks for image classification, land-cover analysis, semantic annotation, and knowledge graph construction. Collectively, these efforts show that

semantic reasoning can improve interpretability and, in some cases, strengthen decision support by making analytical outputs more structured and queryable [10]. However, important limitations remain. Many ontology-based systems in remote sensing have been developed around traditional machine-learning pipelines rather than modern object detectors, which can restrict their performance on complex imagery. Other approaches focus primarily on scene classification, where a whole image is assigned to a single category, rather than on object detection, where multiple semantically relevant entities must be localized and interpreted within the same scene. In addition, some frameworks depend on computationally heavy architectures or highly specialized implementations that may be less suitable for lightweight smart-city monitoring pipelines, where efficiency, scalability, and practical deployment remain important considerations [9].

These limitations create a clear gap in the literature. There is a need for frameworks that combine the efficiency and detection strength of contemporary deep-learning models with the interpretability and reasoning capacity of ontology-based knowledge representation. Such integration is particularly relevant in smart-city environments, where analytical systems must often operate across large image collections while also producing outputs that can be interpreted by non-specialist stakeholders or integrated into broader governance platforms. A system that detects urban objects efficiently but lacks semantic structure may be operationally incomplete, while a semantically rich system that is too computationally complex may be difficult to deploy in practice.

The present work addresses this gap by combining a lightweight YOLOv8 detector with an ontology model designed specifically for semantic analysis of detected objects in remotely sensed satellite imagery. The use of YOLOv8 provides an efficient and effective object-detection backbone capable of identifying salient urban and environmental features in complex scenes. The ontology layer then extends these detections into a structured semantic representation in which object categories, relations, and scene-level meanings can be formalized and queried. This design is particularly relevant to urban-development and smart-city applications because it links efficient computer vision with interpretable knowledge structures, thereby supporting not only accurate recognition but also explainable scene understanding. In this way, the proposed framework contributes to a growing line of research that seeks to move from raw visual detection toward semantically informed urban intelligence.

MATERIALS AND METHODS

Methods Overview

The methodological framework integrates deep learning-based object detection with ontology-driven knowledge representation. The visual pipeline identifies relevant objects in remotely sensed scenes, and the ontology layer converts these outputs into structured semantic relations that support interpretation. The full workflow comprises four main stages:

1. dataset creation and preprocessing;
2. deep learning-based object detection and scene interpretation;
3. integration of detected objects into a SPARQL-accessible ontology;
4. semantic analysis and knowledge inference.

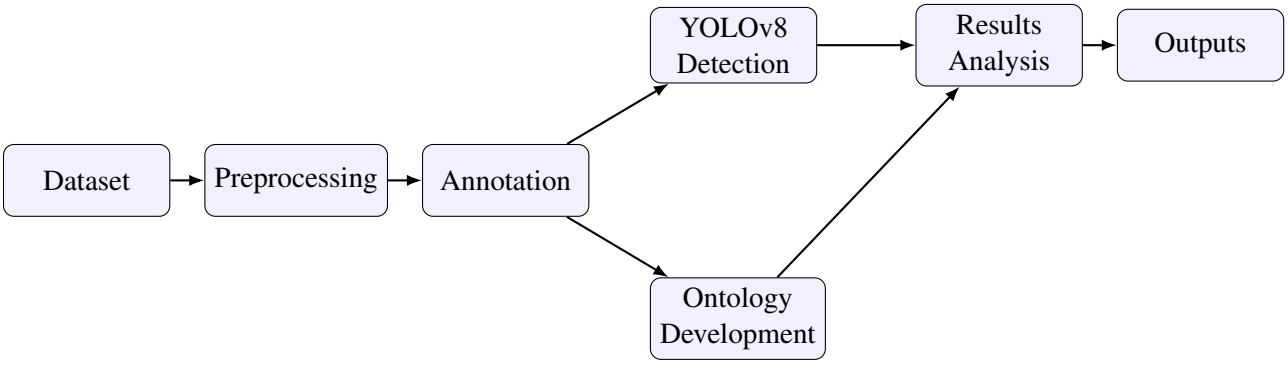

Figure 1: Conceptual workflow of the ontology-guided object detection framework for smart-city remote-sensing scenes.

Dataset Creation and Preprocessing

The study uses a newly created remote-sensing dataset derived from high-resolution Sentinel-2 MSI imagery collected through Google Earth Engine from the southern Durban metropolitan area in KwaZulu-Natal, South Africa. The acquisition point is centered near latitude -29.8579 and longitude 31.0292. All input images are resized to 640×640 pixels.

The initial dataset contains 92 annotated satellite images. Annotation is performed in Roboflow using a multi-class procedure. Five target classes are labeled:

- residences,
- roads,
- shorelines,
- swimming pools,
- vegetation.

The original split comprises:

- 61 training images,
- 21 validation images,
- 10 testing images.

To strengthen generalization and robustness, the dataset is augmented through auto-orientation, flips, and 90/180/270-degree rotations. After augmentation, the partitions expand to:

- 6,100 training images,
- 2,100 validation images,

- 1,000 testing images.

Table 1: Dataset configuration used in the smart-city remote-sensing framework.

Partition	Original images	After augmentation
Training	61	6,100
Validation	21	2,100
Testing	10	1,000
Total	92	9,200

YOLOv8-Based Object Detection Model

The visual recognition component is a lightweight YOLOv8-based detector selected for efficient scene analysis in remote-sensing imagery. The architecture is divided into three units:

1. **Backbone**: convolution layers, C2f modules, and spatial pyramid pooling fast (SPPF) for feature extraction;
2. **Neck**: C2f-based feature fusion replacing older CSP/C3 modules;
3. **Head**: anchor-free detection heads with detection and prediction modules.

This design reduces the parameter burden while maintaining competitive detection performance. The formulation of the detector can be summarized as

$$\mathscr{D}(x) = \{(b_i, c_i, s_i)\}_{i=1}^{N},$$

where x is an input image, b_i denotes the ith predicted bounding box, c_i its class label, and s_i its associated confidence score.

Model Integration with the Ontology Layer

The detected objects are exported into a SPARQL-accessible ontology environment for semantic processing. In operational terms, the YOLOv8 module provides the extracted object classes, and the ontology layer receives them as structured inputs for knowledge extraction, classification support, and semantic reasoning.

A GraphDB-backed SPARQL store is used to insert and query the object relationships. A representative insertion pattern is shown below:

```
PREFIX : <http://deepontos/>
INSERT DATA {
 :LowRiseBuilding :IsAdjacentTo :Forest .
 :LowRiseBuilding :IsAdjacentTo :HighWay .
 :LowRiseBuilding :HasSwimmingPool :Pool .
 :FreeWay :IsAdjacentTo :Shoreline .
}
```

This integration enables object-level detections to be translated into machine-readable semantic facts, which can then be queried to produce interpretable scene descriptions.

Ontology Modeling and Knowledge Taxonomy

The ontology is designed in RDF/OWL and queried through SPARQL. It is organized around the root class `Scene`, which is decomposed into three direct subclasses:

$$\texttt{Region} \sqsubseteq \texttt{Scene}, \qquad \texttt{Area} \sqsubseteq \texttt{Scene}, \qquad \texttt{Segment} \sqsubseteq \texttt{Scene}.$$

Because the semantic analysis in this study focuses on land-cover interpretation, the most important branch is `Area`, which is further divided into:

$$\texttt{VegetationArea} \sqsubseteq \texttt{Area} \sqsubseteq \texttt{Scene},$$

$$\texttt{WaterArea} \sqsubseteq \texttt{Area} \sqsubseteq \texttt{Scene},$$

$$\texttt{WayArea} \sqsubseteq \texttt{Area} \sqsubseteq \texttt{Scene},$$

$$\texttt{ResidentialArea} \sqsubseteq \texttt{Area} \sqsubseteq \texttt{Scene}.$$

More specific subclasses are then introduced to support interpretable reasoning:

- **WaterArea**: `River`, `Shoreline`, `Pool`, `Lake`, `Canal`;
- **VegetationArea**: `Forest`, `Shrub`, `GrassLand`;
- **WayArea**: `Highway`, `RuralRoad`, `LocalStreet`, `FreeWay`;
- **ResidentialArea**: `TownHousesBuilding`, `InformalSettlementsBuilding`, `HighRiseBuilding`, `LowRiseBuilding`.

The ontology also encodes meaningful semantic relations, such as adjacency and containment. For example:

$$\texttt{FreeWay} \sqsubseteq \texttt{WayArea} \sqcap \exists \texttt{isAdjacentTo.Shoreline},$$

and

$$\texttt{LowRiseBuilding} \sqsubseteq \texttt{ResidentialArea} \sqcap \exists \texttt{isAdjacentTo.Highway} \sqcap \exists \texttt{isAdjacentTo.Forest} \sqcap \exists \texttt{hasSwimmingPools.P}$$

These relations support scene understanding that is not limited to class labels alone.

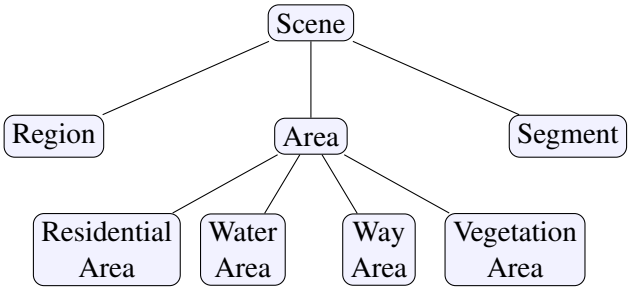

Figure 2: Core ontology hierarchy used for semantic interpretation of detected smart-city scene objects.

Semantic Query Formulation

To formalize spatial reasoning in potentially sensitive urban contexts, the ontology uses query-level semantic rules. Let *RA* be a selected area of interest. The system defines a set of residential subregions that are semantically associated with nearby water and way areas under a specified event context:

$$RA = \{r_i \in R_i \mid R_i \sqsubseteq ResidentialArea \wedge \exists g_i \in G_i \sqsubseteq ResidentialArea \wedge (r_i, g_i) \in S \sqsubseteq isAdjacentTo\}. \quad (1)$$

This query formulation supports the identification of residentially relevant areas influenced by object proximity and relation structure, which is useful for disaster-aware scene interpretation.

Evaluation Metrics

The visual recognition module is assessed using precision, recall, F1-score, intersection over union (IoU), average precision (AP), and mean average precision (mAP). The ontology layer is assessed through structural complexity and reasoning-oriented metrics:

- number of children (NOC),
- depth of inheritance (DIT),
- class in-degree (CID),
- class out-degree (COD),
- reasoning time,
- query response time.

These measures evaluate both predictive effectiveness and the scalability of the semantic model.

RESULTS

Class-Wise Detection Performance

The model successfully identifies the five target classes from the proposed satellite-image dataset. Class-wise results show meaningful variation across categories. Swimming pools achieve the highest precision, while shorelines produce the strongest localization and retrieval performance.

Table 2: Class-wise performance of YOLOv8 on the proposed smart-city remote-sensing dataset.

Class	Precision (%)	Recall (%)	mAP@50 (%)	mAP@50–95 (%)
Residences	41.1	42.1	19.3	12.8
Roads	41.2	57.1	13.7	4.75
Shorelines	54.6	96.4	99.5	59.7
Swimming pools	62.7	62.9	45.5	12.8
Vegetation	57.3	62.3	12.8	8.45

The results indicate that:

- swimming pools yield the highest precision (62.7%);
- shorelines achieve the highest recall (96.4%) and the highest mAP@50 (99.5%);
- residences and roads show comparable performance, reflecting similar distributional difficulty within the imagery;
- dense shoreline, vegetation, and pool regions are comparatively easier for the detector to identify than diffuse residential or road patterns.

Comparison with Other Object Detectors

The detector is also compared with several contemporary object-detection models. YOLOv8 records the strongest precision, recall, and mAP@50 in the reported comparison, while also offering the lowest latency.

Table 3: Performance comparison with alternative object detectors on the proposed dataset.

Method	Precision (%)	Recall (%)	mAP@50 (%)	mAP@50–95 (%)	Speed (ms)
Detectron2	50.0	32.7	16.0	24.0	0.9
YOLOv5	53.4	49.7	27.0	18.4	0.5
YOLOv6	53.2	47.4	32.1	16.6	0.4
YOLOv7	54.5	46.2	34.1	25.0	0.3
YOLOv8	68.0	60.0	43.0	17.5	0.2

The main practical strength of YOLOv8 in this setting is not only its predictive performance but also its low-latency suitability for scalable urban image-processing pipelines.

Comparison with Selected State-of-the-Art Benchmarks

The study also reports an external comparison against selected state-of-the-art detection settings from public benchmarks. In this comparison, the proposed YOLOv8-based approach achieves a reported mAP of 43.1%, slightly exceeding several listed baselines.

Table 4: Comparison with selected state-of-the-art methods reported in related benchmark settings.

Authors	Method	Dataset	mAP (%)
Xiao et al.	CEM + FPM	PascalVOC	37.2
Koyun et al.	Two-stage object detection framework	VisDrone	42.06
Xu et al.	DetectoRS	AI-TOD	41.9
Zhang et al.	YOLOv5-based SPD	VisDrone-DET2019	41.8
This study	YOLOv8-based detector	VisDrone	43.1

Precision, Recall, F1-Score, and IoU by Class

A second performance view confirms that the detector exceeds 50% IoU for all reported classes and produces the strongest F1-score for shorelines.

Table 5: Class-wise performance using precision, recall, F1-score, and IoU.

Class	Precision (%)	Recall (%)	F1-score (%)	IoU (%)
Residences	41.1	42.1	41.6	50.6
Roads	41.2	57.1	47.9	58.1
Shorelines	54.6	96.0	69.5	63.7
Swimming pools	62.7	62.9	62.9	50.1
Vegetation	57.3	62.3	59.7	52.1

These findings reinforce the detector's effectiveness in water-adjacent and clearly bounded object classes, while also highlighting the greater challenge posed by residences and roads in visually complex urban scenes.

Ontology Complexity and Scalability Analysis

The ontology is intentionally lightweight. The reported structural metrics indicate limited depth and moderate branching, which support efficient reasoning and low computational burden.

Table 6: Reported ontology structural metrics for key classes. Blank cells indicate that a specific metric was not explicitly reported for that class.

Class	NOC	DIT	CID	COD
Scene	3	3	3	–
Area	4	2	4	–
WaterArea	–	1	5	–
VegetationArea	–	1	3	1
WayArea	–	1	4	–
ResidentialArea	–	1	4	3

The maximum reported DIT is 3, and the maximum reported NOC is 4. This indicates a scalable ontology design that remains suitable for semantic reasoning without imposing heavy complexity overhead.

Semantic Reasoning and Scene Interpretation

The ontology layer enhances interpretability by turning visual detections into semantic relations. The study highlights several representative inference patterns:

- **Residential-water relation:** low-rise buildings containing swimming pools;
- **Residential-vegetation relation:** low-rise buildings adjacent to forest-type vegetation;
- **Way-water relation:** freeways adjacent to shorelines.

In visual scene analyses, the ontology produces localized descriptions that distinguish safer or more exposed urban configurations:

- scenes containing residences, roads, and vegetation can be described as residential areas adjacent to vegetation and way areas;

- scenes containing residences, swimming pools, roads, and vegetation can be interpreted as residential areas with pools adjacent to vegetation and access routes;
- scenes containing roads, swimming pools, residences, vegetation, and shorelines can be interpreted as more sensitive locations due to the coexistence of residential and way areas near water bodies.

This semantic layer is the manuscript's principal value for smart-city scholarship: it provides a reasoning-based interpretive bridge between detection outputs and urban-functional meaning.

DISCUSSION

The results show that a lightweight object detector and a carefully constrained ontology can form a practical framework for smart-city remote-sensing analysis. Several implications are notable.

First, the visual model achieves credible performance on a newly constructed local dataset, particularly for shorelines and swimming pools. These categories benefit from relatively distinctive visual boundaries and strong contextual separability [10, 11]. Residences and roads are more difficult, which is unsurprising given their heterogeneity, partial occlusion, and spatial overlap in dense urban scenes [13].

Second, the ontology layer materially improves interpretability. Rather than leaving the analyst with unstructured bounding boxes, the framework can formalize relations such as adjacency and containment in ways that matter for urban-development decisions. This is valuable for planners, environmental analysts, and disaster-management stakeholders who require explainable scene summaries rather than isolated detections [4, 14].

Third, the ontology remains lightweight. The modest inheritance depth and limited branching structure are important design strengths because semantic systems often become impractical when their reasoning graph becomes too complex. Here, the reported DIT and NOC values suggest a system that is structurally manageable and well suited to low-resource deployment environments [15, 16].

From the perspective of the Journal of Urban Development and Smart Cities, the manuscript is particularly well aligned because it addresses:

- urban planning through automated scene interpretation,
- environmental monitoring through land-cover and object analysis,
- disaster-aware urban monitoring through semantic reasoning over spatial relations.

At the same time, several limitations should be recognized. The ontology rules in this version rely primarily on object attributes and semantic relationships, while not yet incorporating explicit geospatial measures such as object size, precise location, or inter-object distance. In addition, the overall semantic system remains dependent on the upstream detection quality of YOLOv8. Scenes with occlusion or ambiguous boundaries may still reduce inference reliability.

CONCLUSION

This study presents an ontology-guided deep learning framework for the analysis and interpretation of remotely sensed smart-city satellite images. By integrating a YOLOv8 object detector with an RDF/OWL ontology and

SPARQL-based reasoning, the system supports both object-level recognition and semantically interpretable scene analysis. The detector operates on a custom Sentinel-2 dataset collected from the southern Durban metropolitan area and demonstrates 68% precision, 60% recall, and 43% mAP@50, while the ontology remains lightweight, with a maximum reported DIT of 3 and NOC of 4.

The framework is relevant to urban-development and smart-city research because it combines scalable computer vision with knowledge-driven interpretation for urban planning, environmental monitoring, and disaster-aware scene understanding. Its central contribution is not merely the identification of urban objects, but the structured semantic explanation of how those objects relate within the scene.

Several future directions follow naturally from the present findings. The framework would benefit from the inclusion of explicit spatial attributes such as location, object size, and inter-object distance; it should also be tested against actual historical hazard data to strengthen real-world disaster prediction use cases. Further gains may arise from semantic-context-based refinement of detector outputs, multimodal fusion with sensor or textual data, and temporal reasoning over changing urban scenes.

DATA AVAILABILITY STATEMENT

The SPARQL database referenced in the study is publicly available through the authors' stated repository:

`https://data.mendeley.com/datasets/s5v4zz7yj5/1`

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

REFERENCES

[1] Adegun, A.A.; Fonou-Dombeu, J.V.; Viriri, S.; Odindi, J. Ontology-Based Deep Learning Model for Object Detection and Image Classification in Smart City Concepts. *Smart Cities* **2024**, *7*, 2182–2207.

[2] Adegun, A.A.; Dombeu, J.V.F.; Viriri, S.; Odindi, J. State-of-the-Art Deep Learning Methods for Objects Detection in Remote Sensing Satellite Images. *Sensors* **2023**, *23*, 5849.

[3] Gruber, T.R. A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition* **1993**, *5*(2), 199–220.

[4] Berners-Lee, T.; Hendler, J.; Lassila, O. The Semantic Web. *Scientific American* **2001**, *284*(5), 34–43.

[5] McGuinness, D.L.; van Harmelen, F. OWL Web Ontology Language Overview. W3C Recommendation, 2004.

[6] Prud'hommeaux, E.; Seaborne, A. SPARQL Query Language for RDF. W3C Recommendation, 2008.

[7] Kumar, L.; Mutanga, O. Google Earth Engine Applications since Inception: Usage, Trends, and Potential. *Remote Sensing* **2018**, *10*, 1509.

[8] Roboflow. Roboflow Platform. Computer software, 2022.

[9] Solawetz, J. What Is YOLOv8? The Ultimate Guide. Technical article, 2023.

[10] Glimm, B. Using SPARQL with RDFS and OWL Entailment. In *Reasoning Web International Summer School*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 137–201.

[11] Ontotext. GraphDB. Software platform, 2024.

[12] Li, X., Vázquez-López, A., Sánchez del Río Sáez, J., & Wang, D. Y. (2022). Recent advances on early-stage fire-warning systems: mechanism, performance, and perspective. Nano-micro letters, 14(1), 197.

[13] Dian, S., Cheng, P., Ye, Q., Wu, J., Luo, R., Wang, C., ... & Gong, X. (2019). Integrating wildfires propagation prediction into early warning of electrical transmission line outages. IEEE Access, 7, 27586-27603.

[14] Yu, L. (2023, January). Neural Network Technology for Electrical Fire Early Warning System. In International Conference on Innovative Computing (pp. 308-315). Singapore: Springer Nature Singapore.

[15] Gashteroodkhani, O. A., Majidi, M., & Etezadi-Amoli, M. (2021). Fire hazard mitigation in distribution systems through high impedance fault detection. Electric Power Systems Research, 192, 106928.

[16] Kim, T. H., Yoo, J. G., & Jeon, J. C. (2015). Quantitative Analysis on the Electrical Fire Preventive Effect of Safety Inspection for Electrical Facilities for General Use. The Transactions of the Korean Institute of Electrical Engineers P, 64(2), 45-49.

Baojia Gong, College of Civil Engineering, Lanzhou Jiaotong University, Lanzhou, Gansu, 730070, China

Manuscript Published; 15 October 2025.