Relational Reasoning in Remote Sensing: A Review of Current Paradigms and the Case for a Co-Evolutionary Framework

Ahmed Qusay Subhe [a,1], Muntadher Aidi Shareef [a,2,*]

[a] Northern Technical University, Technical Engineering College / Kirkuk, Department of Surveying Technical Engineering, Kirkuk 36001, Iraq
[1] ahmedchakmakchi90@gmail.com; [2] muntadher.a.shareef@ntu.edu.iq
* corresponding author

ARTICLE INFO

Article history
Received January 21, 2026
Revised March 7, 2026
Accepted May 21, 2026

Keywords
Remote Sensing (RS),
Graph Neural Networks (GNNs),
Adaptive spatial primitives
Learnable graph topology
Relational reasoning

ABSTRACT

Remote sensing activities entail modelling long distance dependencies beyond the neighbourhood. The classic CNNs, which are founded on fixed-grid Euclidean representation, suffer the inherent shortcoming of being unable to capture non-local dependency, and an uneven structure of geographical things. Despite the fact that Object-Based Image Analysis (OBIA) have proposed the context-dependent spatial primitives, they are nevertheless constrained by the fixed segmentation boundaries and the use of heuristic graph creation that makes them not adaptable. Graph Neural Networks (GNNs) provide a framework of non-Euclidean relational reasoning, however the current approaches presuppose the fixed graphs and fixed semantics of nodes, and restrict dynamic representation of spatial-temporal changes in remote sensing data. It is proposed in this paper that an end-to-end shift to common frameworks between adaptive spatial primitives and graph topologies should be undertaken. These structures recursively co-construct spatial entities and their association with task specific loss cues to allocate new differentiable layouts of groupings and the consequent emergence of edges modulated by the expanding node representation. This multi-modal and multiscale representation is unified in a way that directly facilitates easy inference, and it can also adapt to temporal change. The shortcomings of the lack of labels and computational access to scaling to feedback loops in both feature extraction and relational reasoning are addressed by these systems and finally lead to the enhancement of semantic coherence and representable form. We present the long-desired paradigm that combines UQ, encourages lifelong learning, and overcome the drawbacks of the current fixed-grid or heuristic-graph-based cycles of RS analytics to move to adaptable, and comparatively minimized spatial and relational representations.

1. Introduction

The more advanced the surroundings of the RS applications are, as a global monitor up to disaster response and city planning, the analytical models are likely to reason about the spatial context far beyond the immediate local contiguous neighborhoods . Despite the fact that deep learning dominates the RS images analysis field, present paradigms have been stretched to the limits of the relations of the worlds that display complicated, non-local and irregular processes.

An example is the CNNs; hierarchies of features constructed on individual pixels result in the very good performance in most image processing tasks . The primary component in their architecture, however, the elementary call it the convolutational operator, is given intrinsically on a fixed grid and on Euclidean distances in Rn. Kernels scan standard and fixed pixel windows, making CNNs hard to comprehend structures on spatially distributed portions or model geographic objects with discontinuous edges [1] [2]. The mismatch between geometrico-topological data of real world objects and inflexible structure of model raster creates a structural bottleneck, and stops the models at higher order relational reasoning [3].

Graph Neural Networks (GNNs) are quite an appealing alternative since they enable to learn on non-Euclidean data in a very general manner, representing the elements of space as nodes and creating the connections between them to describe the relationships between them [4]. Such a way of information dissemination through this mechanism throughout the message-passing algorithm, suggests that GNNs must have the capability of capturing interactions between arbitrary nodes within a graph even in the event they are not socially distant [5]. That is to say that they are best adapted to RS work, whereby functional or thematic connectivity can sometimes go beyond adjacency. But a closer consideration of how they are applied reveals that there is a silence bottleneck in the world of RS: they are based on heuristically acquired, predefined, and static graphs. Most of the work flows may be distilled to the following form, with the graph structure (nodes (primitives) and edges(the links between nodes)) being either fixed before training and not able to change during training. The drawback is that the GNN is less expressive and, consequently, the ability of its quality reasoning system is being utilized on a fixed and inaccurate skeleton. This carries on the rigidity which this architecture was intended to prevent [4].

This would entail radical rethink in general strategy. The fact that the spatial representation was not necessarily uncoupled with the relational learning is a huge issue that should be addressed. We suggest that there be simultaneous acquisition of spatially adaptive primitives and graph topology. Such co-evolutionary systems learn inputs (primitive do not have a relational structure) not provided to them, but instead learned via data. With task-based loss-indications, coarse boundaries can be narrowed down and graph connections can also be tightened to allow the system to learn how to organize most useful spatial and relational structures concerning a particular task alone. Such a direction would probably allow richer, more adaptive, and more semantically founded relational reasoning in remote sensing.

The historical overview in terms of the critical theoretical systems applicable to spatial representation of RS will first be provided in this essay and it will be demonstrated that the key theme that emerges in other spatial theories is the structural rigidity that cuts across the history of RS. It will also expound on the existing roadblock in the use of GNN, which is the constraint of the topology being fixed. Such a co-evolutionary framework and architectural constituents and complementarity with other AI paradigms will then be outlined (at very high level of abstraction) in the paper. We will comment about how the new paradigm comes in with tremendous research opportunities which can tackle tremendous challenges in PGM like scarcity of labels, dynamic modeling of the environment and reliability of models.

2. Materials and Methods

Remote sensing classification history traces back to the pursuit of seeking an adequate method of coding the spatial information. Each new paradigm has tried to address the problem of the previous one and when critically reflected on this evolution, this is the only thing which becomes constant is that they still (wherever possible) are based on the primitives of space which are at rest. This is the systematic propensity to structural rigidity, the aesthetic decision of canonical unit of analysis is pre-programmed a priori and unlearned, has caused the accomplishment of relational reasoning to be a challenge (Figure 1).

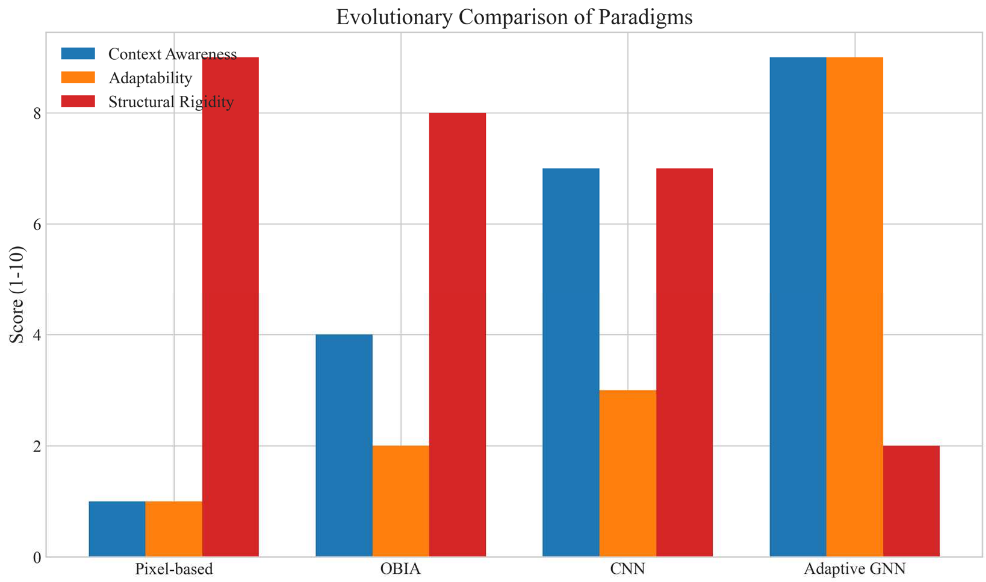

Figure 1. Evolutionary Comparison of Paradigms

2.1 The Pixel-Based Era the Presupposition of Independence.

2.1.1 Premises of Early Statistical Classifiers.

Pixel-Based Approaches The first pixel based method of automating RS classification was the pixel-based method where each pixel was modeled as a single data point in the form of a spectral samples vector [6]. High dimensional statistical classifiers in spectral feature space presented above [7]. High dimensional statistical classifiers in spectral feature space presented above [8]. In this manner, a pixel was set to that class having a maximum posterior probability [9]. Such techniques could be directly observed mathematically and computed upon calculable technology of the time and produced repeatable conformational permits given a comparable set of teaching information [9].

2.1.2 Rejection: The Space-Lessness of context.

And that brings us back to the idea of pixel independence that turned out to be also the most serious weakness of the said paradigm. These classifiers were context blind by virtue of their ability to remove all the spatial information, adjacency, texture and shape [10][11]. This was already a harsh limitation in spectrally heterogeneous scenes (or high resolution data objects with large intra-class spectro-variance). As an example, certain categories of urban elements (asphalt road and specific roofing material) may create spectral features of other urban ones leading to false classification. In fact, however, lacking the factual knowledge that pixels along a road fall within a building, being not used in predicting linear roof pans themselves, they are a part of one in a planar sense. (2022). The absence of relational reasoning brought the knit-and-pepper distortion into classification maps2480, where erroneously classified pixels are not directly interacting and hence gestures to areas that would otherwise be homogenous [12]. This inadequacy was demonstrated by the effect of sensor resolution: further spatial resolution corrects the mixed pixel problem, but further spectral resolution increases spectral variability within classes (e.g. shadows on a building, texture in forest canopy) which, ironically, destroys spectral homogeneity based classifiers [13]. By extension, any prior efforts to deal with the decontextualization of land cover mapping by post-processing heuristics (i.e. the use of a Minimum Mapping Unit (MMU) to discard pixels less than a size threshold), reclassification of small isolated patches, and the like, are just a series of compromises aimed at addressing what is fundamentally a representational problem [14].

2.2 The Object-Based Paradigm (OBIA): A Step Towards Context.

2.2.1 Innovation Pixels to Image Objects.

The strong solution to the problems posed by pixel-based techniques is OBIA (Object-Based Image Analysis) or GEOBIA (Geographic OBIA) [15]. The strong solution to the problems posed by pixel-based techniques is OBIA (Object-Based Image Analysis) or GEOBIA (Geographic OBIA) [16]. This has two major advantages. First, it can incorporate local spatial context in the model and can eliminate the "salt-and-pepper" effect. Second, it enabled each primitive to be defined by a wider range of features. Besides average spectra, shape (e.g., area and compactness), texture (e.g., statistics based on Grey-Level Co-occurrence Matrices) and contextual (e.g., nearest-neighbor relation to other objects) features may be further used to characterise each object [17].

2.2.2 Censure: The Inflexibility of Heuristic Segmentation.

But the effect of OBIA has brought about some form of rigidity within the structure of developing the idea. The segmentation is controlled by user-defined heuristic parameters (size, shape, and compactness) that determine this step to identify the primitives which is achieved as a static preprocessing that is not relevant to the classification process [18][19]. Consequently, the resulting object boundaries are predetermined and not set up in accordance with semantic needs connected to the task. And that is where the problem lies: if one object in the initial segmentation is mistakenly combined (for example, if a shadow makes one building into two objects, or if a therapy donut field and a planning target volume are defaulted into the same object), it will be cursed to fail throughout the structural data analysis downstream pipeline. Still more recent techniques, like the hierarchical segmentation, which form nested groupings of objects at different scales, are still founded on predefined, fixed layers that cannot be adjusted to feedback in the classifiers [20]. This way, the contextual instrumentality promised was limited.

2.3 Convolutional Paradigm: Implicit Local Context.

2.3.1 Innovation: Learning End-to-End Features.

Another paradigm shift was caused by the creation of the Convolutional Neural Networks (CNNs) that automated feature engineering, which is an important part of OBIA [21]. CNNs use raw pixel data to directly learn hierarchical spatial-spectral properties in an end-to-end fashion [22]. By producing maps that preserved spatial information without requiring distinct ideas of object or region-based systems, Full pixelwise semantic segmentation was also made possible by architectures such as Fully Convolutional Networks (FCNs) [23]. The column (barrel/strip) in the neocortex replaced the receptive field the gridded patch of pixels that a neuron can see as the primary unit of analysis.

2.3.2 Criticism: The Euclidean Tyranny.

CNNs offer an answer to the grid-based restrictive form of (Euclidean) grids, despite the fact that their multi-layered convolution layers implicitly learn local structure [24]. By design, the convolution operator aggregates information from nearby, predetermined neighbors. Because of this, it is terrificly well composed with standard data models such as photos, yet can in principle never probe geographic objects, the edges of which are locally non-grid aligned, or whose edges are explicitly intended to mark geographically discontinuous domains [25]. Under CNN's paradigm, a road network, a river system, or an ecologically related but fragmented collection of habitat patches cannot be depicted as a whole in their natural state [26]. The receptive field is a fixed-shape and topology primitive even though it has learning properties. This is the contemporary version of the same structural stiffness issue that has existed from the start.

The same can be said of either the machine symbols, programmable systems, or other symbolic paradigms, which all share a common denominator: that at the core of all lies the spatial primitive bearing no relation to the learning process. This makes one of these historical paths worth researching. Because a pixel was a static, context-free unit, pixel-based techniques were ineffective. Although OBIA created more sophisticated primitives, the method of fixing them remained static and heuristic, meaning that they were set up prior to learning about its patterns. With the exception of features that are automatically learned, CNN does away with the requirement for human feature design. But neurons The main, historical justification for the paradigm that is being promoted in this study is this common error: separating the unit of analysis from the learning process (Figure 2) (Table 1).

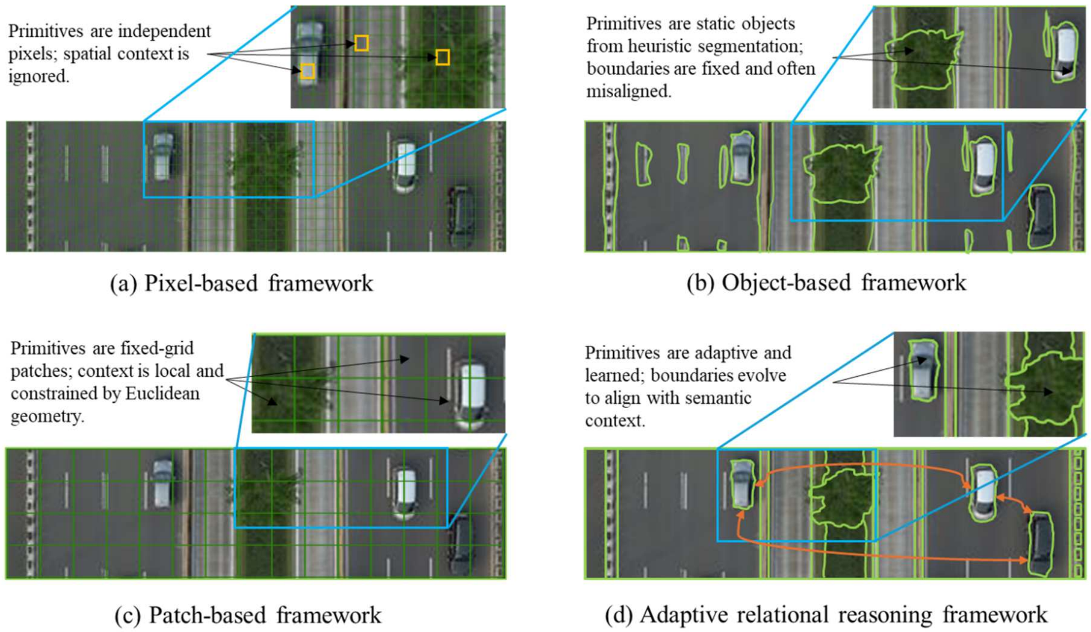

Figure 2. The remote sensing classification of the conceptual development of spatial primitives.

The performance of the algorithm on the samples is shown in the images below: (a) Independent pixel grid: This illustrates the pixel based age, in which every unit is treated individually. (b) A group of non-uniform and fixed fragments of OBIA where the step is guided on relevant objects, but have fixed limits. (c) The superimposing grid based receptor fields of a CNN with the local context and implicitly trained on a fixed architecture. (d) a set of evolving, adaptive primitives with dynamically defined boundaries and which evolve alongside the learning task on which the current paradigm is trying to abstract.

Table 1. Evolutionary Comparison of Remote Sensing Classification Paradigms.

Paradigm	The simplest element of the analysis (Spatial Primitive).	Attitude to Relational Reasoning.	Key Advantages	Persistent Limitations
Pixel-Based Classification (PBC)	The pixel is considered as an independent spectral sample.	Implicit / Discarded. The relational context is disregarded, classification is just made based on the spectral signature of each pixel in isolation.	Mathematically clear and computationally efficient on limited hardware. Reproducible outputs given the same training statistics.	Context blind, by definition, resulting in salt-and-pepper noise. Bases on powerful assumptions of spectral separability that break down on complicated topography.

Paradigm	The simplest element of the analysis (Spatial Primitive).	Attitude to Relational Reasoning.	Key Advantages	Persistent Limitations
Object-Based Image Analysis (OBIA)	Prior to the classification, the image object or super pixs are created through heuristic segmentation algorithms.	Heuristic / Constrained. Context is represented by objects characteristics (shape, texture, adjacency), though the graph of objects is built on the basis of the static segmentation lines.	Eliminates the salt and pepper effects through working with coherent objects. It is possible to combine spectral, spatial, geometric, and textural characteristics.	Static Representation. Flow boundaries are hard before learning and unadaptable. Segmentation is heuristic and therefore lacks flexibility in various scenes.
Convolutional Neural Networks (CNNs)	The receptive field or patch which is fixed. Primitives are overlapping regions of the Euclidean image grid, which are homogeneous.	Implicit / Local. Hierarchical feature aggregation in local and adjacent receptive fields is an implicit way of learning context. Lacks in defining long-range dependencies.	Learning of hierarchies end-to-end. Excellent local pattern recognition and object detection and have adequate data.	Static Representation. The inflexible Euclidean grid does not coincide with non-geometric geographic objects. The fixed topology is a problem because it does not allow modeling of non-local relationships.
Adaptive GNN Framework	Adaptation Spatial Primitives. The boundaries and composition of spatial units are learned to change dynamically as part of the training and are based on task specific loss signals.	Explicit / Adaptive. Relational reasoning is graphically carried out on a graph that can be learned. End to end The nodes (primitives) and edges (connectivity) are co-evolved in an end-to-end structure.	Great adaptability and seminal coherence. Breaks down grid constraint and non-relational reasoning by combining feature extraction and relational reasoning.	Overcomes fundamental constraints but causes new issues in scalability, computational effectiveness, and interpretability of models that need to be addressed in future studies.

3. Results and Discussion

3.1 The Static Topology Constraint of Bottleneck in Graph-Based Remote Sensing.

Graph Neural Networks (GNNs), by offering a natural method to handle non-Euclidean data along with capturing relational structure explicitly, have become the optimal solution for several of the limitations of grid-based models [27]. Nevertheless, despite the immense potential these methods hold, they have not yet seen widespread application in remote sensing due to an implementation gap that presents a silent bottleneck when seeking to leverage them. This bottleneck can thus be attributed to the use of static, heuristic-defined graph topologies that should have been retired by the basics of end-to-end learning and flexibility (of GNNs).

3.1.2 The Prospect of GNNs of Non-Euclidean Reasoning.

GNNs are based on the graph structures. G=(V,E) with V and E set of nodes and edges respectively. [28]. In RS, nodes can represent spatial primitives (such as pixels, superpixels, or objects), and edges may depict any type of relation based either on the spatial proximity of the pixels of the underlying anatomy, or even on the functional connectivity of the underlying anatomy (relations of any kind, such as spectral similarity or fMRI-based functional connectivity, are also considered that may lead to a graph arrangement) [29]. GNN uses the message passing mechanism, in which each node iteratively gathers signals from neighboring nodes to recalculate its feature representation (embedding) [30]. Unlike CNNs, this approach captures both local and long-range dependencies and permits information to flow throughout the graph [31]. Due to this capability, GNNs are best positioned to conceptually describe the multi-scale, haphazard, and complex interactions that are characteristic of geographical phenomena.

3.1.3 The Implementation Gap: The Unspoken Bottleneck.

Regretfully, using GNNs on predefined/static graphs is still the standard practice in RS [32]. By addressing the static data structure with a dynamic, adaptable learning model, they have committed a fundamental contradiction. There are two important on-site manifestations of this separation in the graph from graph learning.

Initially, the nodes on the graph are often constructed using conventional static primitives that were employed in earlier paradigms. When pixels are produced using an algorithm like SLIC, for example, we refer to them as superpixels [33] or sections from an OBIA pipeline, function as the graph's fundamental building block. For a GNN, heuristic procedures for picture segmentation are developed outside of the training loop [34]. These primitives sometimes propagate segmentation faults and deformable representation biases into the graph, as discussed in Section 2, and do not correspond to true semantic entities.

Second, extremely basic static heuristics are used to construct the connections between these nodes. Most of those are for node connection based on initial feature similarity or spatial proximity (e.g., k-nearest neighbors or distance threshold) [35]. More complex (functional) relationships, such the hydrological interaction between nonadjacent areas of a watershed or the relationship between an industrial sector and a distant transportation hub, cannot be explained by such heuristics. After that, this graph topology is set and maintained during training. The GNN can be trained on weights of these predefined edges (as in a Graph Attention Network, or GAT), but cannot remove negative edges or add with it any new edges to represent discovered associations. [36]. Because of this, GNN's powerful reasoning ability is limited to acting on possibly inaccurate scaffolds, which results in a performance and flexibility barrier.

3.1.4 A Taxonomy of Static GNN Applications in RS

All of the current GNN application scenarios on RS exhibit this stagnant character. This limit persists in several forms, as demonstrated by a brief taxonomy:

a. Prescribed Fixed Graphs with GNNs: GNN is typically used on a clearly specified graph structure that remains constant throughout training. Heuristics are used to create edges, and nodes are constructed over static primitives. The chance that the graph could rearrange in accordance with actual, the structural inelasticity should remove task-specific relations that develop with a finer level of detail/content in case of feature refinements [37].
b. GNNs in Feature Refinement: This algorithm applies a GNN as a plug-in on the obtained CNN model feature vectors [38]. Strengthening the neighbors of features with relationship context is the goal. However, the flow of information is one-way. The initial feature extractor (CNN and/or rudimentary mining) cannot receive feedback or information from the GNN; it only updates the features over a fixed graph. A static topology that is not ideal limits the shape.
c. Spatio-Temporal GNNs (STGNNs): These models combine the temporal modeling unit (like RNN) with spatial graph propagation to extend GNNs for temporal time-series data [39][40]. A static spatial graph, on the other hand, is often described from the initial time step and is repeated

for each time step. Since relational structural dynamics are crucial for modeling seasonal variation, disaster progression, and land use change, the scope does not take this into account.

A major paradox occurs when powerful non-Euclidean models such as GNNs are used directly on data structures which are a product of conventional Euclidean and heuristic paradigms. Although the model has been given a static, fixed data representation, it is intended to be dynamic and contain arbitrary connections. Rather than GNN, the issue is the separation of graph creation and learning. This important difference necessitates a completely end-to-end solution with the representation and the logic working together to evolve.

3.2. The Co-Evolutionary Framework: Jointly Learning Primitives and Topologies

A new architectural paradigm is required to overcome the static topology constraint and unlock the full potential of graph-based reasoning for remote sensing. In contrast to chain-like disconnected workflows, this paradigm directly creates a single, coherent framework from beginning to end: spatial primitives and graph topologies are embedded learnable components rather than strictly predetermined inputs. This allows you to teach both the world representation (primitives and their relations) in this way as well as what is known about it (learning features), making for a powerful feedback loop that makes a system more semantically coherent and precise.

3.2.1 Adaptive Primitive Generation: Learning the Units of Analysis

The first pillar of this system is adopting adaptive spatial primitives. These techniques use the data properties themselves to determine the optimal spatial units for a particular task, as opposed to predetermined, heuristic units like pixels, patches, or superpixels [41]. This is accomplished by incorporating a segmentation or grouping mechanism that can be learned into the network itself [42].

Using a CNN's intermediate feature maps, which make up a rich, multi-scale descriptor space, is one such method. These attributes can then be used by a learnable module (such as a superpixels generation network or differentiable clustering layer) to produce an initial set of primitives [43]. The parameters are immediately updated via backpropagation for the reason that this module is a section of the end-to-end graph. The feedback signal is the result of the loss of a past task. Such boundaries will be adjusted by the grouping module in case the current primitive border would result in a major classification error or semantic poorness (such as a single primitive has several-class features). This may be a merge operation that combines the primitives that have very similar learnt embeddings, or a split operation that subdivides a mixed-up primitive into more unified fragments. This repeated propagation enhances the general semantic consistency in that the picture primitives which are the nodes of the final graph adapt to the semantic structure that is changing with the scene in the course of the iterations

3.2.2 Constructable Relational Construction: Learn the Relational Structure.

The second element is trainable graph building where the topology of the graph, which consists of edges and edge weights, is no longer considered as fixed with hyperparameters [44]. This overcomes the limitations of basic heuristics (such spatial proximity) by allowing the model to investigate and encode the most crucial links for a task [36].

Graph formation is learnable through a number of techniques. One method is to start with a minimally linked graph and determine whether edges should be added or removed using a straightforward heuristic. For instance, you may forecast whether a new link will exist based on the closeness of node embeddings following each message passing loop. The model has the ability to draw an additional edge between two nodes that belong to different parts of the space provided that they always map to a similar visible part, meaning that they are linked invisibly. Another great way is to use attention scores of a GAT. Non-adjacent node scores that are strongly high or strongly low between nodes can however lead to the addition or removal of an edge respectively than just changing the weight of existing edges [39]. The graph structure is perfected in terms of functional/semantic representation of data by exposing the existence of the edges and the weight to whether they contribute to the minimization of the ultimate loss/cost incurred in the task.

3.2.3 The Unified Architecture: Co-Evolution and Feedback Loops.

The combination of the two elements into a unified image under the control of both overlapping (i.e. bidirectional) feedback loops is the real force behind the framework (Figure 3). But it is co-evolutionary and cyclic as opposed to linear.

a. Feature extraction: A CNN backbone is used to extract rich features maps using the images as input.
b. Primitive & Graph Generation: This property is inputted into an adaptive module which proposes the topology of the graph (i.e., edges) and an initial set of primitives (i.e., nodes).
c. Relational Reasoning: By communicating relational context, a GNN uses this graph to carry out message transmission and change node embeddings.
d. Prediction and Loss Calculation: A task-specific loss is computed and predictions (like classification) are made using the most recent node embeddings.
e. Backpropagation: The architecture is used to propagate the loss gradient. This gradient concurrently modifies (1) GNN settings to improve reasoning, They are (2) the topology refinement (graph adaptive graph) module (adding or removing an edge), and (3) the refinement (adding or removing a node) module in the spatial representation (Compiler primitive).
f. The CNN parameters of discovering features which are helpful in efficient relational reasoning and more close to good primitives.

This feedback loop ensures that global relational reasoning (in the GNN) and local feature extraction (in the CNN) are informed by one another rather than being separated. Primitives' quality dictates how to connect them in a graph, while the learned graph's quality dictates how primitives should be formed. This is a more abstract design, to allow models to assemble dynamic, emergent representations of the underlying geospatial system: this is because recent work on learning systems introductions and methods has shifted its focus by no longer seeking features that work on fixed units, but rather on the units themselves and interactions among them (Table 2). Also, we can pretrain on large-scale, unlabeled datasets of RS, unsupervised, and then start the whole developmental course where we form powerful priors on FE and SO and then fine-tune in task-specific manners [45].

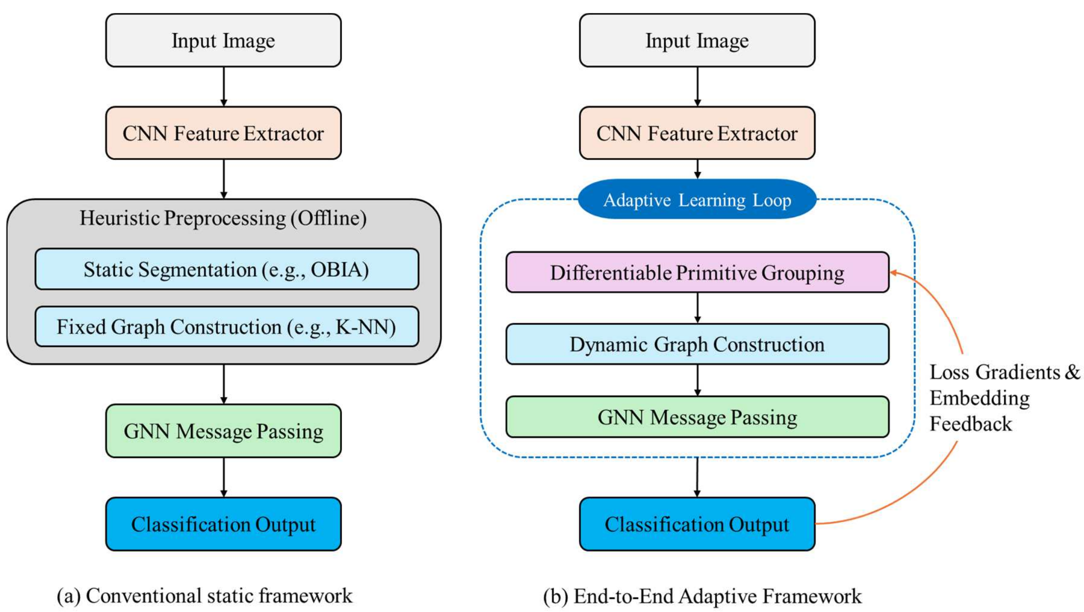

Figure 3. Similarity in Architecture between Static and Adaptive Graph Learning Structures.

Vol. 7, No. 1, June 2026, pp. 1-18

Table 2. Heuristic vs. Learnable Graph Construction Comparative Analysis

Criterion	**Heuristic Construction of Graphs.**	**Constructable and Constructible Graphs.**
Methodology	The given preprocessing step, which does not belong to the learning loop, is referred to as graph topology. Connection is founded on fixed rules such as proximity of space or similarity of spectral.	One of the end-to-end model components that can be learned is graph topology. Under the influence of task-specific loss, and changing node embedding, connectivity changes over time during training.
Adaptability	Low. The graph is structural and is fixed. It is incapable of adapting to new patterns, rectifying initial errors and reacting to dynamic changes in the data.	High. The structure of the graph is dynamic and plastic. It can add, prune or re-weight edges to constantly optimize the relational pathways of the certain task and scene.
Primitive Dependency	Based on primitives (pixels, superpixels, OBIA segments) which are (or must be) static and predefined. The definitions of the nodes are not linked to the process of relational reasoning.	Combined with the generation of adaptive primitives. The definition of nodes can be (refined) split/merged depending upon feedback on the graph learning module, so that primitives and relations develop co-evolutionally.
Handling of Multi-Modal Data	Bases on heuristic fusion and cross-modality links which are usually simple geometric overlaps. Lacks dynamic and context specific relationships.	Learning of cross-modal relationships is facilitated. The connections between modalities may be established and rated depending on their combined contribution to the task and be adjusted to the changing circumstances.
Handling of Long-Range Dependencies	Frequently incapable of forming genuine long-range dependencies, heuristics usually being local proximity biased. The important non-local connections are often not included.	Is capable of explicitly finding and modelling useful long-range dependencies. Relationships may also be established across any two nodes in case their association is useful in training.
Computational Overhead	Lower initial overhead. Nonetheless, it may be inefficient when training the graph when the graph is somewhat poorly organized (e.g., too dense) to the extent that superfluous message flow occurs.	The complexity of the computations may be increased because of the necessity to review and revise the graph structure. Needs effective implementation and sampling in order to be scaled.

3.3. The Architectural Synergies and Extended Uses.

The joint learning of graph topologies and adaptive primitives which has been presented based on co-evolutionary method provides a rich platform to incorporate more complex AI paradigms and tackle a broader variety of challenging remote sensing tasks. The framework is a sort of a universal adapter, eliminating the necessity of ad ad hoc heuristic fusion answers to each new model or datatype by applying an organized method to dynamic representation.

3.3.1 Combination with Other AI Paradigms.

The original design The First True Hybrid CNN-GNN design: The next study is the first to propose and implement a true hybrid CNN-GNN design. Both are closely connected in a co-optimization cycle as compared to the naive traditional pipelined setup in which a CNN serves as a feature extractor to a fixed GNN. Besides local discrimination, the CNN is also trained to learn features that are suitable to construct coherent primitives and smoothly propagate along the learned graph. On the other hand, the CNN features are made better by the global information obtained through the relational thinking of GNN [46].

Synergy with Vision Transformers (ViTs): Despite the observation that Vision Transformers are great at integrating long-range dependencies using self-attention, they tend to cleanse a grid of fixed static patches, or tokens [47]. CNN rigidity is comparable to this "static token" requirement. By replacing the static patching with an adaptable primitive module, the adaptive framework can solve this problem. Instead of random square patches, this would have generated semantically significant, unevenly shaped tokens to the attention mechanism of the transformer that may be capable of computing attention across coherent geographical objects (Figure 4). It is a hybrid paradigm according to which the GNN can introduce a learned, sparse structure, which will affect the dense attention of the transformer, becoming more effective and interpretable despite the fact that integrating ViTs and GNNs can be somewhat challenging in the sense of computational cost and modeling higher-order relationships [39].

Attention: The adaptive perspective reconsiders attention as a dynamic agent of structure seeking as opposed to an impaled straightforward weighting strategy. As mentioned earlier, the attention weights of a GAT can give add or delete edge indications to the learnable graph. This takes the focus off as a means of merely varying the transmission of information, and puts it as a means of defining not just the way in which information could diffuse through the system, but also the way in which it must, dynamically, and on task.

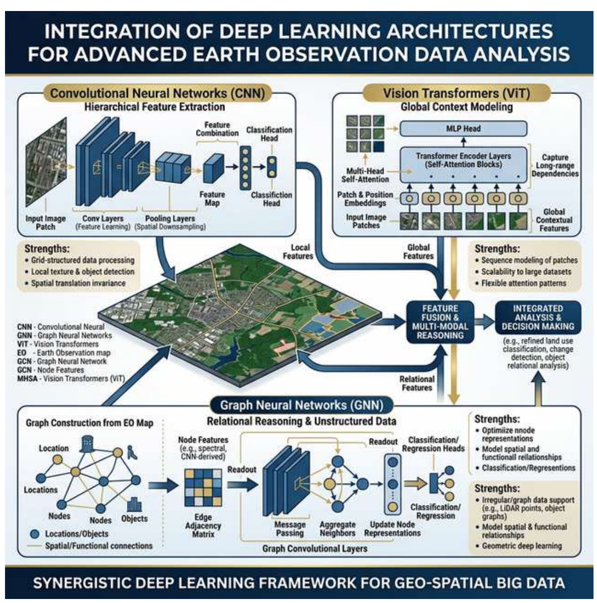

Figure 4. Integration with AI Paradigms

3.3.2 A Unified Substrate for Multi-Modal Data Fusion

One of the biggest issues in RS is still heterogeneous data source fusion. A natural remedy for grid-based (or "crisp") co-registration is provided by the adaptive technique [48]. The adaptive primitive module can take information in a number of modalities and learn to form meso-primitives that

Vol. 7, No. 1, June 2026, pp. 1-18

adaptively obey both the native geometries and the power of each source of data, rather than trying to jam everything into one raster [8].

a. Structural, Smoothness and Moisture: To classify land cover and the position of floods, the structure, roughness and moisture data of SAR data may be integrated with the fine spectral and textural features of optical data [49].
b. Use of LiDAR-based Structure Information: The model is able to make topographical and vertical reasoning using direct integration of 3D-structure data (e.g., elevation and height) obtained with LiDAR into the primitive and graph learning process. In a graph like this, edges simply indicate physical proximity; they can also indicate structural connectedness, such as adjacent roofs or hydrological flow paths created by terrain gradients [50].
c. Add the socioeconomic data, as well as administrative areas and vector types of GIS data (e.g. road networks, etc) that can be inserted naturally into the platform. This non-raster information can as well be fed directly to condition the generation of primitives and graph structure, allowing for the learning of both physical properties and relationships defined by human systems (like zoning laws or transportation connectivity) within a single model.

3.3.3 Scalability and Efficiency Considerations

The computational cost of dynamic adaptive designs is a significant concern. Compared to static structures, optimizing primitives and graph topology can be more difficult. However, there are a number of targeted solutions to this problem:

In each training step, the model may learn to sample the most informative nodes and edges and selective attention allows the model to tie up its computing resources in what is important or uncertain in a graph. Consequently, methodical scaling is smarter and efficient [51].

Distributed and Parallel Processing: While data parallelism is common, managing dynamic graphs presents difficulties that can be resolved by hybrid parallelization techniques. For instance, graph-based components can make use of more robust partitioning and synchronization algorithms that are designed to incorporate small changes in the graph's structure, and feature extraction over the raw picture may parallelize across devices [52].

Energy Efficiency and Memory The computational overhead of adaptive designs also influences memory and energy consumption. Due to its size, practical large-scale deployments will rely on techniques such as incremental model updates avoiding full graph reconstruction, sparse representations in RAM and hardware-aware implementations [53].

It offers a sounder, more powerful basis for addressing the next generation of smart multimodal AI in geospatial as an adaptable substrate on which one can lay down new techniques and ideas.

3.4 Open Research Directions and Grand Challenges

By eliminating this structural constraint, the newly developed co-evolutionary framework instead characterizes many important remote sensing problems as opportunities for innovation rather than intractable problems. This adaptive approach offers novel toolkits to address problems such as model trust, dynamic environments and uncertain information.

3.4.1 Great Label Scarcity and Data Imbalance.

Few-Shot and Zero-Shot Learning: In cases of limited data, models are required to extrapolate on solid information in the past. Even though they operate over underlying fixed spatial representations, recent few-shot approaches to RS often seek to acquire transferable feature embeddings. 81. The adaptive model can also learn the transferable structural priors. Besides training the model to learn to split a scene efficiently and construct a relational graph by using sparse samples, unary/binary potential can be inferred in specific instances by meta-learning on a variety of tasks [54]. Consequently, with a single or few examples, the model can be able to generalize structural learning of features and relationships to form coherent primitives in a new class.

Synthetic Data Generation and Augmentation: Current techniques for augmenting data yield samples that are contextually naive yet spectrally realistic. The adaptive architecture can be used to train generative models (like GANs) to produce artificial data that is both relationally consistent and visually realistic. Prior to enabling the placing of the generated items in a contextually sound manner, the learning graph is used as a potent framework, resulting in more effective training samples [55].

3.4.2 Dynamic and Evolving Spatial Relationships Modeling.

Modeling of Land Use and Cover alter (LUCC): Object boundaries and the functional relationships will change based on the dynamic process of land use change. A static model is unable to account for this. Because primitives are allowed to change over time to fit changing land cover patches, the adaptive framework is naturally suited for this application, while the learnable graph can record the formation of new connections (such as a new road connecting two previously distinct regions) and the breakdown of existing connections [56].

The adaptive setup can effectively distinguish between those transient changes in appearance and actual variations in land cover. The stability of geometric qualities in terms of open-close trends can be ensured by primitives being invariant to changes in time of the graphs even when their eigenvalues vary.

Incremental and Continual Learning Approaches: New models should continually adapt whenever new data regarding the same classes becomes available, in order to avoid catastrophic interference on previously learned knowledge. The structural flexibility of the adaptive framework is a potent ally for continued learning. This permits the model to incrementally modify its primitives and graph topology when new data arrives, thus recycling and learnt parameters occur at learning time as defined by [51].

Sudden Event Detection and Model Updating: The model's perception of a scene must be quickly reconfigured in response to sudden events like earthquakes, wildfires, or floods. Near real-time disaster response and monitoring can benefit greatly from the adaptive framework's ability to quickly redefine primitives (for example, to surround a flood) and update graph connections (e.g. to represent a broken transport connection).

3.4.3 Trust and explainability of GNN-Based RS.

Graph-Based Decisions Interpretation: The black box and mystery surrounding most deep learning methods are crippling to trust. The interpretability is also another new opportunity presented by the adaptive framework since it learns clear and dynamic representations of primitives and their relationships. It is possible to exhaustively explore the graph structure (relationships which are significant by a decision) learned and to investigate the primitives (sub-structures of a graph) learned by the model to determine what it finds interesting as an object. Unlike full implicit models, this gives an insight into the decision making process of the model [57].

Forecast Uncertainty: Reputable models will provide both a forecast and its level of confidence. Uncertainty may be structurally modeled thanks to this adaptive framework. As opposed to the uncertainty of pixels, the model can tell how probable it is that it has a particular edge on the graph or that it is not confident about its primitive boundaries. This is a more detailed and interpretive confidence measure of models.

Implications for Ethics and Policy of Automated RS Analysis: Automated analysis may reinforce biases in the training data. The auditors can verify that the framework learns the representations because it is visible. This enables the review of whether discovered primitives and linkages reflect societal biases (e.g., biased crowdsourced labeling) or are based in real-world phenomena (Table 3). Thus, it becomes simplified to develop and deploy more unbiased and accountable models and this is necessary particularly where RS data are used to inform societal policy (Figure 5).

Vol. 7, No. 1, June 2026, pp. 1-18

Table 3. Challenges and Open Research Directions made possible by Adaptive Frameworks.

Research Area	Challenge Description	Intervention of Adaptive Primitives & Graphs to the Challenge.
Label Scarcity and Data Imbalance	Few-Shot and Zero-Shot Learning: Learners need to be trained to generalize with a very limited number of labelled examples or none whatsoever, and need to discover very general and powerful relational patterns.	A flexible architecture is able to learn to produce the best, context-bound primitives and graph structures using only a limited number of examples and concentrate relational reasoning on the most salient features to get the most information.
	Synthetic Data Generation: Generating synthetic data may reduce the data scarcity, but does not have realistic spatial context, and relational consistency.	Generative models can be equipped with adaptive graphs, which can be used to generate synthetically realistic data, and to do so in a manner that is contextually and relationally coherent.
Dynamic and Evolving Spatial Relationships	Simulating Land Use Change: Tracking changes necessitates models to apprehend the combination of the perimeter of objects and their working relationships.	Adaptive primitives have capabilities to change their boundary dynamically to follow the changing land cover forms. Graphs can be learned that can represent the formation and breaking of relationships over time.
	Rapid Event Detection: In the instantaneous events such as floods or wild fires, models need to re-examine their perception of a scene in near real-time.	The capacity of the framework to reconfigure the primitives and graph topology speedily permits it subsequent to adjust the novel condition of the surrounding, defining the limits of events and simulating new relational effects.
	Continual and Incremental Learning: Models should be able to learn continuously on a new flow of data without losing the previous learned information disastrously, which necessitates plasticity in structures.	The model offers a system of structural adaptation. Primitives and graphs may develop over time in order to assimilate new information, and this means that the representational space of the model may change.
Explainability and Trust in GNN-Based RS	Interpretability of Decisions: Deep learning in its black box form is not trustworthy because it is hard to fathom why a model has made a specific decision.	The framework puts primitives and graph structure explicit and learnable, which gives it a place to be interpreted. The spatial units learned may be pictured and the connections between them may be imagined, which allows one to see the reasoning of the model.
	Uncertainty Quantification: Models must indicate their level of confidence in a prediction which is essential in high stakes applications such as disaster response.	Uncertainty may be represented both on a feature and a structural level. The framework is able to realize uncertainty in its primitive definitions and graphical relationships to give a truer measure of confidence.
	Ethical and Policy Implications: Computer analysis tends to be biased. Models are supposed to be neutral, transparent and accountable to eradicate negative impacts to the society.	The auditing of representations which are learned can be done through a transparent adaptive framework. One can further examine whether educated primitives and associations are representations of phenomena in the world or biases.

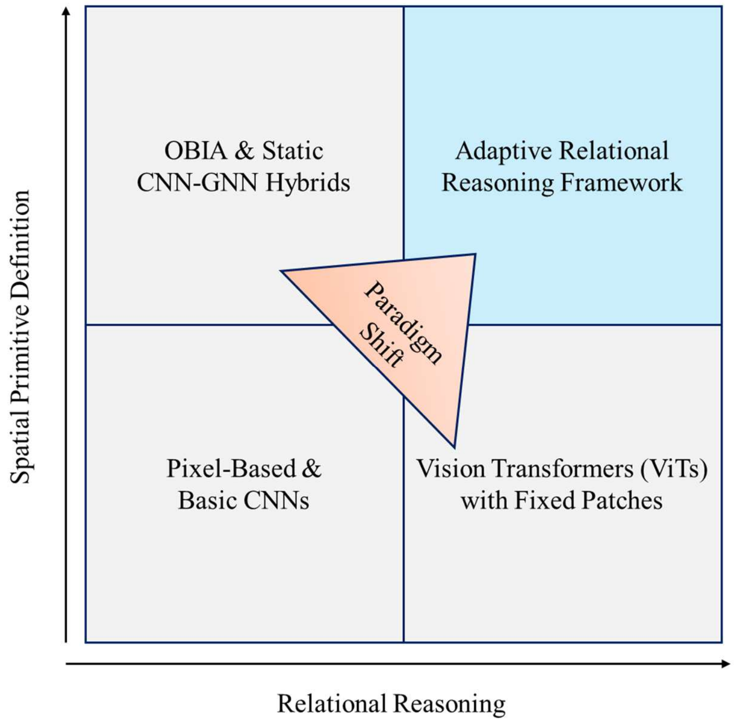

Figure 5. Graphical View of the Trade-off between Model Adaptability and Representational Rigidity.

4. Conclusion

Remote sensing categorization and its evolution are carefully reviewed in this article, which demonstrates that regime continuity is substantiated by the rigidity of a structure in terms of restricting processes. All these methods are motivated by a wide category of classifiers where spatial information is not represented by features that are formed by hand, and none of them recognize the importance of spatial context in obtaining high-resolution picture analysis. Due to a persistent reliance on predetermined, static graph structures, the introduction of Graph Neural Networks, which promised non-Euclidean reasoning, has yet to yield many benefits. This paper's primary contention is that in order to overcome these shortcomings, a change in emphasis is necessary. We have advocated the employment of end-to-end, combined structures that both discover graph topologies and adaptive spatial primitives. A global relational reasoning process is developed during model training in order to develop an isomorphic space that can naturally adjust its topology in order to represent the topological arrangements of the data. This co-evolution depicts the mutual feed-forward between the two processes. The creation of learnable graphs allows the explicit learning and representation of complex, non-local, multi-modal interactions that are unavailable to methods built on heuristics, and adaptive primitives can make sure that the graph nodes are coherent and relevant to a task. An example of an architectural innovation that allows us to shift from a pipeline (decouple, train CNN, fusion module) model to the reality of a strong synergy is the integration of CNNs and GNNs within the same joint optimization loop. Such a cohesive connection can offer a sound system for dealing with problems that have been proven to be present in the remote sensing community, like efficient functioning under conditions of scarcity, powerful multimodal data integration, and response to dynamics in time. They can provide flexibility, which is difficult with fixed representations, since they can expand and change the primitives and graphs to better represent the seasonal changes, alterations in land use, or any unforeseen occurrences. Moreover, such a paradigm shift could play a critical role in the explainability and reliability of AI for geospatial analysis. The framework facilitates the interpretation of model decisions, quantification of structural uncertainty in data, and verification of potential biases in XFZ by learning the representational structures of the model—primitives and links. As an example of plasticity, we also train morphological components using variational EMs. This moves the field closer to transparent and responsible analytics and away from opaque, black-box systems. Although adaptive structures are computationally challenging to scale, technical issues are memory management and performance. There are too many advantages, which have to do with conceptual matters and pragmatic ones, in aligning the representational structure of a model to the data it is meant to simulate. The idea of advanced remote sensing analytics is intelligent, as it looks at intelligent systems, which can learn what to observe, how to feel on the side, and not simply implementing more and more powerful models on more and more archaic and rigid representations. We can make models that reflect the complexity, dynamism, and relationality of the environment we live in by adopting architectures in which spatial primitives are used to develop graph topologies, and vice versa.

References

[1] Z. Xue, Q. Xu, and M. Zhang, "Local transformer with spatial partition restore for hyperspectral image classification," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 15, pp. 4307-4325, 2022, doi:10.1109/jstars.2022.3174135.

[2] Z. Du, H. Ye, and F. Cao, "A novel local–global graph convolutional method for point cloud semantic segmentation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 35, no. 4, pp. 4798-4812, Apr. 2024, doi:10.1109/tnnls.2022.3155282.

[3] S. Prasad, L. M. Bruce, and J. Chanussot, "Introduction," in *Optical Remote Sensing*, Springer, 2011, pp. 1-8, doi: 10.1007/978-3-642-14212-3_1.

[4] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436-444, May 2015, doi:10.1038/nature14539.

[5] J. Guo, K. Huang, X. Yi, Z. Su, and R. Zhang, "Rethinking spectral graph neural networks with spatially adaptive filtering," *IEEE Trans. Neural Netw. Learn. Syst.*, pp. 1-15, 2026, doi: 10.1109/tnnls.2026.3673098.

[6] M. Hussain, D. Chen, A. Cheng, H. Wei, and D. Stanley, "Change detection from remotely sensed images: From pixel-based to object-based approaches," *ISPRS J. Photogramm. Remote Sens.*, vol. 80, pp. 91-106, Jun. 2013, doi:10.1016/j.isprsjprs.2013.03.006.

[7] A. Knudby, "Remote sensing," *Open Library*, 2021. [Online]. Available: https://ecampusontario.pressbooks.pub/remotesensing/front-matter/welcome-to-remote-sensing/. [Accessed: Mar. 18, 2026].

[8] W. Yuan, "Graph neural network-based multimodal sensor fusion for robust autonomous driving perception," in *Second Int. Conf. Intell. Transp. Smart Cities (ICITSC 2025)*, SPIE, Jun. 2025, p. 44, doi: 10.1117/12.3073578.

[9] J. Ju, E. D. Kolaczyk, and S. Gopal, "Gaussian mixture discriminant analysis and sub-pixel land cover characterization in remote sensing," *Remote Sens. Environ.*, vol. 84, no. 4, pp. 550-560, Apr. 2003, doi:10.1016/s0034-4257(02)00172-4.

[10] R. Haralick and H. Joo, "A context classifier," *IEEE Trans. Geosci. Remote Sens.*, vol. GE-24, no. 6, pp. 997-1007, Nov. 1986, doi: 10.1109/tgrs.1986.289563.

[11] B. Jeon and D. A. Landgrebe, "Classification with spatio-temporal interpixel class dependency contexts," *IEEE Trans. Geosci. Remote Sens.*, vol. 30, no. 4, pp. 663-672, Jul. 1992, doi: 10.1109/36.158859.

[12] Y. Chen, Y. Zhou, Y. Ge, R. An, and Y. Chen, "Enhancing land cover mapping through integration of pixel-based and object-based classifications from remotely sensed imagery," *Remote Sens.*, vol. 10, no. 1, p. 77, Jan. 2018, doi:10.3390/rs10010077.

[13] M. Dalponte, L. Bruzzone, L. Vescovo, and D. Gianelle, "The role of spectral resolution and classifier complexity in the analysis of hyperspectral images of forest areas," *Remote Sens. Environ.*, vol. 113, no. 11, pp. 2345-2355, Nov. 2009, doi: 10.1016/j.rse.2009.06.013.

[14] E. Gromny, S. Lewiński, M. Rybicki, R. Malinowski, M. Krupiński, and A. Nowakowski, "Post-processing tools for land cover classification of Sentinel-2," in *Photonics Appl. Astron. Commun. Ind. High-Energy Phys. Exp. 2019*, SPIE, Nov. 2019, p. 232, doi: 10.1117/12.2537325.

[15] S. Yang, M. Hou, and S. Li, "Three-dimensional point cloud semantic segmentation for cultural heritage: A comprehensive review," *Remote Sens.*, vol. 15, no. 3, p. 548, Jan. 2023, doi: 10.3390/rs15030548.

[16] G. Chen, Q. Weng, G. J. Hay, and Y. He, "Geographic object-based image analysis (GEOBIA): Emerging trends and future opportunities," *GISci. Remote Sens.*, vol. 55, no. 2, pp. 159-182, Jan. 2018, doi:10.1080/15481603.2018.1426092.

[17] S. T. M. Ataky, "Image classification methods based on texture analysis and characterization," Ph.D. dissertation, École de technologie supérieure, 2022.

[18] B. Ez-zahouani, A. Teodoro, O. El Kharki, L. Jianhua, I. Kotaridis, X. Yuan, and L. Ma, "Remote sensing imagery segmentation in object-based analysis: A review of methods, optimization, and quality evaluation over the past 20 years," *Remote Sens. Appl. Soc. Environ.*, vol. 32, p. 101031, Nov. 2023, doi: 10.1016/j.rsase.2023.101031.
[19] M. D. Hossain and D. Chen, "Segmentation for object-based image analysis (OBIA): A review of algorithms and challenges from remote sensing perspective," *ISPRS J. Photogramm. Remote Sens.*, vol. 150, pp. 115-134, Apr. 2019, doi: 10.1016/j.isprsjprs.2019.02.009.
[20] M. Seyedhosseini and T. Tasdizen, "Semantic image segmentation with contextual hierarchical models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 5, pp. 951-964, May 2016, doi: 10.1109/tpami.2015.2473846.
[21] A. Thapa, T. Horanont, B. Neupane, and J. Aryal, "Deep learning for remote sensing image scene classification: A review and meta-analysis," *Remote Sens.*, vol. 15, no. 19, p. 4804, Oct. 2023, doi: 10.3390/rs15194804.
[22] D. Yu, Q. Xu, H. Guo, C. Zhao, Y. Lin, and D. Li, "An efficient and lightweight convolutional neural network for remote sensing image scene classification," *Sensors*, vol. 20, no. 7, p. 1999, Apr. 2020, doi: 10.3390/s20071999.
[23] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *2015 IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431-3440, doi: 10.1109/cvpr.2015.7298965.
[24] U. A. Bhatti, H. Tang, G. Wu, S. Marjan, and A. Hussain, "Deep learning with graph convolutional networks: An overview and latest applications in computational intelligence," *Int. J. Intell. Syst.*, vol. 2023, no. 1, Jan. 2023, doi:10.1155/2023/8342104.
[25] J. Ghosh and S. Gupta, "ADAM optimizer and categorical crossentropy loss function-based CNN method for diagnosing colorectal cancer," in *2023 Int. Conf. Comput. Intell. Sustain. Eng. Solutions (CISES)*, Apr. 2023, pp. 470-474, doi: 10.1109/cises58720.2023.10183491.
[26] S. Stabinger, D. Peer, and A. Rodríguez-Sánchez, "Arguments for the unsuitability of convolutional neural networks for non-local tasks," *Neural Netw.*, vol. 142, pp. 171-179, Oct. 2021, doi: 10.1016/j.neunet.2021.05.001.
[27] N. Rusk, "Deep learning," *Nat. Methods*, vol. 13, p. 35, 2016, doi: 10.1038/nmeth.3707.
[28] X.-M. Zhang, L. Liang, L. Liu, and M.-J. Tang, "Graph neural networks and their current applications in bioinformatics," *Front. Genet.*, vol. 12, Jul. 2021, doi: 10.3389/fgene.2021.690049.
[29] J. R. Stevens, R. G. Resmini, and D. W. Messinger, "Spectral-density-based graph construction techniques for hyperspectral image analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 10, pp. 5966-5983, Oct. 2017, doi:10.1109/tgrs.2017.2718547.
[30] D. Sarkar, S. Roy, S. Malakar, and R. Sarkar, "A modified GNN architecture with enhanced aggregator and message passing functions," *Eng. Appl. Artif. Intell.*, vol. 122, p. 106077, Jun. 2023, doi: 10.1016/j.engappai.2023.106077.
[31] N. Sheikh and S. Saha, "Graph neural networks for multi-sensor Earth observation," in *Deep Learning for Multi-Sensor Earth Observation*, Elsevier, 2025, pp. 211-230, doi: 10.1016/b978-0-44-326484-9.00020-8.
[32] T. R. Harsha, V. P. Bhat, K. Prabhanjan, S. K. Vinay, and L. Thomas, "Graph neural networks for scene understanding: A review and future directions," in *Graph Neural Networks: Essentials and Use Cases*, Springer, 2025, pp. 277-339, doi: 10.1007/978-3-031-88538-9_13.
[33] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint arXiv:1609.02907*, 2016.
[34] S. Khoshraftar and A. An, "A survey on graph representation learning methods," *ACM Trans. Intell. Syst. Technol.*, vol. 15, no. 1, pp. 1-55, Jan. 2024, doi: 10.1145/3633518.
[35] J. Zhang, K. A. Ehinger, J. Ding, and J. Yang, "A prior-based graph for salient object detection," in *2014 IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 1175-1178, doi: 10.1109/icip.2014.7025234.
[36] L. Wang, W. Huang, M. Zhang, S. Pan, X. Chang, and S. W. Su, "Pruning graph neural networks by evaluating edge properties," *Knowl.-Based Syst.*, vol. 256, p. 109847, Nov. 2022, doi: 10.1016/j.knosys.2022.109847.
[37] W. Bi, L. Du, Q. Fu, Y. Wang, S. Han, and D. Zhang, "MM-GNN: Mix-moment graph neural network towards modeling neighborhood feature distribution," in *Proc. Sixteenth ACM Int. Conf. Web Search Data Min.*, Feb. 2023, pp. 132-140, doi: 10.1145/3539597.3570457.
[38] A. Bera, Z. Wharton, Y. Liu, N. Bessis, and A. Behera, "SR-GNN: Spatial relation-aware graph neural network for fine-grained image categorization," *IEEE Trans. Image Process.*, vol. 31, pp. 6017-6031, 2022, doi:10.1109/tip.2022.3205215.
[39] L. Wang et al., "NTIRE 2024 challenge on stereo image super-resolution: Methods and results," in *2024 IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2024, pp. 6198-6207, doi:10.1109/cvprw63382.2024.00624.
[40] Z. Li, J. Yu, G. Zhang, and L. Xu, "Dynamic spatio-temporal graph network with adaptive propagation mechanism for multivariate time series forecasting," *Expert Syst. Appl.*, vol. 216, p. 119374, Apr. 2023, doi:10.1016/j.eswa.2022.119374.
[41] W. Kim, A. Kanezaki, and M. Tanaka, "Unsupervised learning of image segmentation based on differentiable feature clustering," *IEEE Trans. Image Process.*, vol. 29, pp. 8055-8068, 2020, doi: 10.1109/tip.2020.3011269.
[42] H. Shen, H. Ding, Y. Zhang, X. Cong, Z.-Q. Zhao, and X. Jiang, "Spatial-frequency adaptive remote sensing image dehazing with mixture of experts," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, pp. 1-14, 2024, doi:10.1109/tgrs.2024.3458986.
[43] Z. Wang, J. Wang, K. Yang, L. Wang, F. Su, and X. Chen, "Semantic segmentation of high-resolution remote sensing images based on a class feature attention mechanism fused with Deeplabv3+," *Comput. Geosci.*, vol. 158, p. 104969, Jan. 2022, doi: 10.1016/j.cageo.2021.104969.
[44] V.-A. Darvariu, S. Hailes, and M. Musolesi, "Goal-directed graph construction using reinforcement learning," *Proc. R. Soc. A Math. Phys. Eng. Sci.*, vol. 477, no. 2254, Oct. 2021, doi: 10.1098/rspa.2021.0168.

[45] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," *arXiv preprint arXiv:1710.10903*, 2017.

[46] F. Rastakhiz, "GNN-CNN: An efficient hybrid model of convolutional and graph neural networks for text representation," SSRN, 2025, doi: 10.2139/ssrn.5346208.

[47] A. A. Aleissaee et al., "Transformers in remote sensing: A survey," *Remote Sens.*, vol. 15, no. 7, p. 1860, Mar. 2023, doi: 10.3390/rs15071860.

[48] S. Aboagye-Ntow and X. Huang, "A survey of multimodal data fusion in Earth observation-remote sensing," *Preprints*, Oct. 2025, doi: 10.20944/preprints202510.0743.v1.

[49] A. Irfan, Y. Li, X. E, and G. Sun, "Land use and land cover classification with deep learning-based fusion of SAR and optical data," *Remote Sens.*, vol. 17, no. 7, p. 1298, Apr. 2025, doi: 10.3390/rs17071298.

[50] C. Wen, X. Li, X. Yao, L. Peng, and T. Chi, "Airborne LiDAR point cloud classification with global-local graph attention convolution neural network," *ISPRS J. Photogramm. Remote Sens.*, vol. 173, pp. 181-194, Mar. 2021, doi: 10.1016/j.isprsjprs.2021.01.007.

[51] K. Wang, "An adaptive graph sampling framework for graph analytics," *Soc. Netw. Anal. Min.*, vol. 14, no. 1, Dec. 2023, doi: 10.1007/s13278-023-01157-x.

[52] L. Meng et al., "A survey of distributed graph algorithms on massive graphs," *ACM Comput. Surv.*, vol. 57, no. 2, pp. 1-39, Oct. 2024, doi: 10.1145/3694966.

[53] N. Rashid, B. U. Demirel, and M. Abdullah Al Faruque, "AHAR: Adaptive CNN for energy-efficient human activity recognition in low-power edge devices," *IEEE Internet Things J.*, vol. 9, no. 15, pp. 13041-13051, Aug. 2022, doi: 10.1109/jiot.2022.3140465.

[54] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, "A comprehensive survey on graph neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 1, pp. 4-24, Jan. 2021, doi:10.1109/tnnls.2020.2978386.

[55] A. Mumuni and F. Mumuni, "Data augmentation: A comprehensive survey of modern approaches," *Array*, vol. 16, p. 100258, Dec. 2022, doi: 10.1016/j.array.2022.100258.

[56] S. Gaur and R. Singh, "A comprehensive review on land use/land cover (LULC) change modeling for urban development: Current status and future prospects," *Sustainability*, vol. 15, no. 2, p. 903, Jan. 2023, doi:10.3390/su15020903.

[57] L. Zhong, Z. Chen, Z. Wu, S. Du, Z. Chen, and S. Wang, "Learnable graph convolutional network with semisupervised graph information bottleneck," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 36, no. 1, pp. 433-446, Jan. 2025, doi: 10.1109/tnnls.2023.3322739.