Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Hybrid GeoAI Utilizing DQN-Driven Adaptive Fusion Approach for Approximate Polygon Regularisation of Rasterised Building Footprints

vibhor joshi

vibhor.joshi@iiitg.ac.in

indian Institute of information technology guwahati

Research Article

Keywords: Building footprint extraction, Geospatial AI, Reinforcement learning, Adaptive fusion, DQN, Mask R-CNN, Morphological regularisation (RT, RR, FER), Raster window sampling, Vectorisation, Boundary IoU, Hausdorff distance

Posted Date: April 14th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9378815/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Hybrid GeoAI Utilising DQN-Driven Adaptive Fusion for Regularisation of Building Footprints

Abstract

Acquiring accurate building footprints from remote sensing images across a continental scale presents significant challenges due to data variability, insufficient labelling, and computational limitations. We present a GPU-accelerated hybrid GeoAI framework designed to transform rasterised U.S. building data into normalised 3-channel patches (area, count, duplicated area) accompanied by binary masks. The framework consists of an eight-step progressive enhancement pipeline: baseline Mask R-CNN segmentation, GPU acceleration, enhanced regularizers (RT: closing; RR: open-close; FER: edge-augmented), basic RL adaptive fusion, continuous actions via PPO, CNN contextual embeddings, pre-trained models (COCO initialisation), and multi-state training. The pipeline underwent testing on subsets comprising approximately 500 patches sourced from eight U.S. states, utilising the Rasterised Building Footprints for USA dataset. The framework attained a mean IoU of 74.9% (F1 = 0.802), indicating a 7.1% enhancement compared to the baseline Mask R-CNN (67.8%). Although the adaptive reinforcement learning components extend the training duration, the implementation of GPU acceleration and parallelised morphological operations significantly decreases inference time, resulting in a 17.6× speed-up (70.8 ms per patch, 326 patches per minute) when compared to the CPU baseline. Multi-state validation shows a consistent improvement of 4.86% in IoU across different geographies, such as a notable increase of 5.2% in New Hampshire. Ablation studies indicate that the largest contributions come from continuous actions, which account for a 1.4% increase in IoU, and multi-state training, contributing an additional 0.5% to IoU. The platform provides a reproducible test environment for adaptive regularisation, demonstrating improved boundary accuracy and shape retention in qualitative evaluations, while enabling statewide analysis in a matter of hours.

Keywords: Building footprint extraction, Geospatial AI, Reinforcement learning, Adaptive fusion, DQN, Mask R-CNN, Morphological regularisation (RT, RR, FER), Raster window sampling, Vectorisation, Boundary IoU, Hausdorff distance

1 Introduction

Precise building footprint extraction is a crucial prerequisite for environmental evaluation, urban planning, and catastrophe management, and it is the focus of this study. Conventional geometric techniques, including directional histograms and corner detection algorithms, face considerable challenges due to noise interference and issues with scalability when applied to large datasets. On the

other hand, sophisticated deep learning methods such as Mask R-CNN provide enhanced segmentation capabilities; however, they encounter difficulties with irregular edge detection and overfitting, especially when dealing with scarce or diverse training datasets.

The main goal is to create and assess a hybrid GeoAI framework that combines the advantages of these methods, attaining a balanced integration of geometric precision, noise resistance, and computational efficiency. The study leverages the Rasterised Building Footprints dataset, which includes 125 million 30m-resolution polygons spanning eight U.S. states, with the objective of surpassing traditional static regularisation techniques. Findings from fire detection studies highlight comparable efficiency-accuracy trade-offs, offering a significant reference point for evaluating the framework's generalisation and performance (Jin et al., 2023).

The study focusses on creating a GPU-accelerated Hybrid GeoAI system aimed at enhancing the extraction of building footprints. This system combines Mask R-CNN segmentation with ResNet-50 and Feature Pyramid Networks (FPN) to achieve high-quality instance segmentation, incorporating a Deep Q-Network (DQN)-based adaptive fusion process. The framework utilises three simultaneous regularizers: RANSAC Thresholding (RT) to maintain topological integrity, Rectangular Regularisation (RR) to ensure orthogonal alignment, and Fuzzy Edge Refinement (FER) to improve boundary precision. The components analyse a dataset comprising 125 million polygons at a 30m resolution from the Rasterised Building Footprints dataset (Heris et al., 2020). This analysis results in a mean Intersection over Union (IoU) of 74.9% across eight U.S. states, with performance varying from 69.9% in Delaware to 75.2% in New Hampshire. GPU optimisation achieves a remarkable 17.6× increase in speed, alongside the dynamics of Proximal Policy Optimisation (PPO).

The document is structured into four sections to systematically address the research objectives. Section 1 presents the challenge of building footprint regularisation and introduces the Hybrid GeoAI methodology. Section 2 undertakes an extensive literature review, situating the study within geometric approaches, deep learning techniques, and hybrid methodologies. Section 3 details the architectural design, covering windowed sampling, Mask R-CNN segmentation, RT/RR/FER regularisation, and DQN-based fusion. Section 4 presents experimental results, demonstrating IoU improvements and 17.6× acceleration with statistical validation across eight states.

2 Review of Existing Literature

Remote sensing and computational capabilities have transformed building footprints collection and standardisation, facilitating urban planning, emergency readiness, and environmental reasoned assessments. Contemporary advancements in deep learning paradigms manage intricate visuals through innovative approaches grounded in data, whereas traditional methods employed geometric principles. Hybrid architectures, adaptable innovations, GeoAI-driven methodologies, and conventional approaches that rely on established frameworks are analysed,

emphasising scalability, bias mitigation and noise suppression difficulties that our reinforcement learning RL-infused solution addresses. We compare foundational regularisation experiments to Microsoft's US Building Footprints (2023) and the 30m-grid rasterised Dataset of Building Footprints (Kong et al., 2023; Gbodjo et al., 2023).

2.1 Advanced Learning Techniques

CNNs uncover intricate patterns in multispectral imaging, enhancing analytical calibration accuracy (Zhao et al., 2021; Liu, 2022). Mask R-CNN instance segmentation division separates eight U.S. states from travelling from Rhode Island to New Jersey in our pipeline, enhanced through the integration of ResNet-50 and FPN. Mapflow's 89% accuracy is inadequate in opposition to dataset artefacts, however, U-Net assists. Generative adversarial network RegGAN integrates a comprehensive approach to regularisation with geometric loss functions for high fidelity (Xu et al., 2025). Before pixel-based extraction, SuperpixelGraph employs a semantic graph developments for semi-automatic extraction. Post-processing enhances line fidelity in aerial CNN frameworks, while zero-utilise pre-trained models with kinetic partitions and Markov random areas to steer clear of localised training. Overfitting on limited datasets and missing geometric anchors are adjusted by our flexible regularisation.

2.2 Combined Approaches

Hybrid approaches integrate data-driven understanding concepts through geometric principles to enhance footprint regularisation (Butt et al., 2024; Urrea, 2025). The segmentation and orthogonal edge polygon optimisation in our RT/RR/FER module enhance this synergy, while integrated frameworks assess 22 geometric indices across Kaggle's rasterised Microsoft polygons. Neural networks construct urban models by utilising pre-processed LiDAR and images, facilitating multi-state validation. Rhode Island: 73.2% IoU, Connecticut: 72.7%, among others. The National Land Cover Database (NLCD) enhances morphological studies (Song et al., 2016), despite the presence of noise in hybrid static post-processing. The 17.6×GPU-accelerated DQN fusion facilitates dynamic optimisation, akin to Earth system modelling (Irrgang et al., 2021) and graph convolutional networks (Bhatti et al., 2023).

3 Design Structures

In order to achieve building footprint regularisation, the processing pipeline consisting of four layers is illustrated in Fig. 1. The hybrid GeoAI architecture is represented. It can effectively and effortlessly handle vast quantities of geographical data obtained from remote sensing sources. This approach addresses issues that arise during automated extractions, particularly those involving geometric inconsistencies, such as overly sharp lines or angles that lack straightness. This is achieved through the integration of rule-based geometric enhancements, data-driven segmentation, and adaptive control algorithms. Microsoft has transformed its 125 million vector polygons into rasters with a resolution of 30 meters, illustrating both the area and quantity of buildings. This document is titled Rasterised Building Footprints for USA. Due to its utilisation of patch-based processing, it is capable of

simulating scenarios with minimal data. This indicates that it can be utilised nationwide for purposes such as mapping hazards and analysing urban morphology. The Al-Muthanna and IOSB algorithms illustrate two geometric regularisation methods that exhibit robust performance in controlled settings by imposing dominant orientation and orthogonality constraints. They consistently achieve accuracies exceeding 90% under optimal conditions, ensuring properly aligned building edges and straight boundaries. The traditional geometric approaches frequently encounter issues with noise and bias when utilised on diverse datasets and generally necessitate considerable manual adjustment. Their performance diminishes when data is moved between various geographic areas, which restricts scalability and automation.

The first layer of the proposed framework is responsible for the ingestion and pre-processing of raster data. The process involves reading GeoTIFF files through GDAL libraries and extracting layers for building area and count, represented as 32-bit floating-point arrays. Every raster holds crucial geographic metadata, such as coordinate reference systems (CRS) and affine transformation matrices, which are later employed for precise georeferencing and vectorisation. Min-max scaling is employed in statistical normalisation to ensure that the values of each pixel are confined within the range of 0 to 1. The implementation of the sliding window technique, combined with spatial decomposition, results in patches measuring 512×512 pixels, featuring an overlap of 128 pixels. This strategy is designed to maintain boundary continuity and minimise edge artefacts during the following segmentation process. We can now choose patches according to minimal pixel thresholds, such as 50 pixels, to concentrate on regions that exhibit sufficient structural density. Subsequently, these are transformed into three-channel tensors (area, count, and copied area) to facilitate their utilisation by pre-trained CNNs.

Layer 2 utilises Mask R-CNN, incorporating a ResNet-50 backbone along with a Feature Pyramid Network (FPN) to analyse features across five pyramid levels (P2-P6) with dimensions [256, 128, 64, 32, 16]. This enables the capture of images of small sections of structures at different resolutions. The Region Proposal Network (RPN) generates anchor boxes with dimensions of 32^2, 64^2, 128^2, 256^2, or 512^2, featuring aspect ratios of 0.5, 1.0, or 2.0. These boxes will perform optimally with any structure that has a rectangular shape. ROI Align employs bilinear interpolation to achieve feature extraction precision at the pixel level. This is essential for ensuring that the boundary lines are accurate. This is the stage at which raw instance masks, boxes, and scores are generated.

Layer 3 includes modules for concurrent geometric regularisation that utilise topic-specific priors to enhance the quality of raw masks. The initial step taken by the RANSAC Thresholding (RT) function to identify boundary pixels involves employing Canny edge detection, utilising thresholds set between 50 and 150. Subsequently, lines are fitted utilising Random Sample Consensus. The RR module generates edge orientation histograms with 5° bins, identifies the peaks of the most significant directions, and employs least squares optimisation to minimise $\sum(d_i)^2$ for aligning polygons under orthogonal constraints. It states that $\theta = k \cdot 90 \pm \varepsilon$, where ε represents

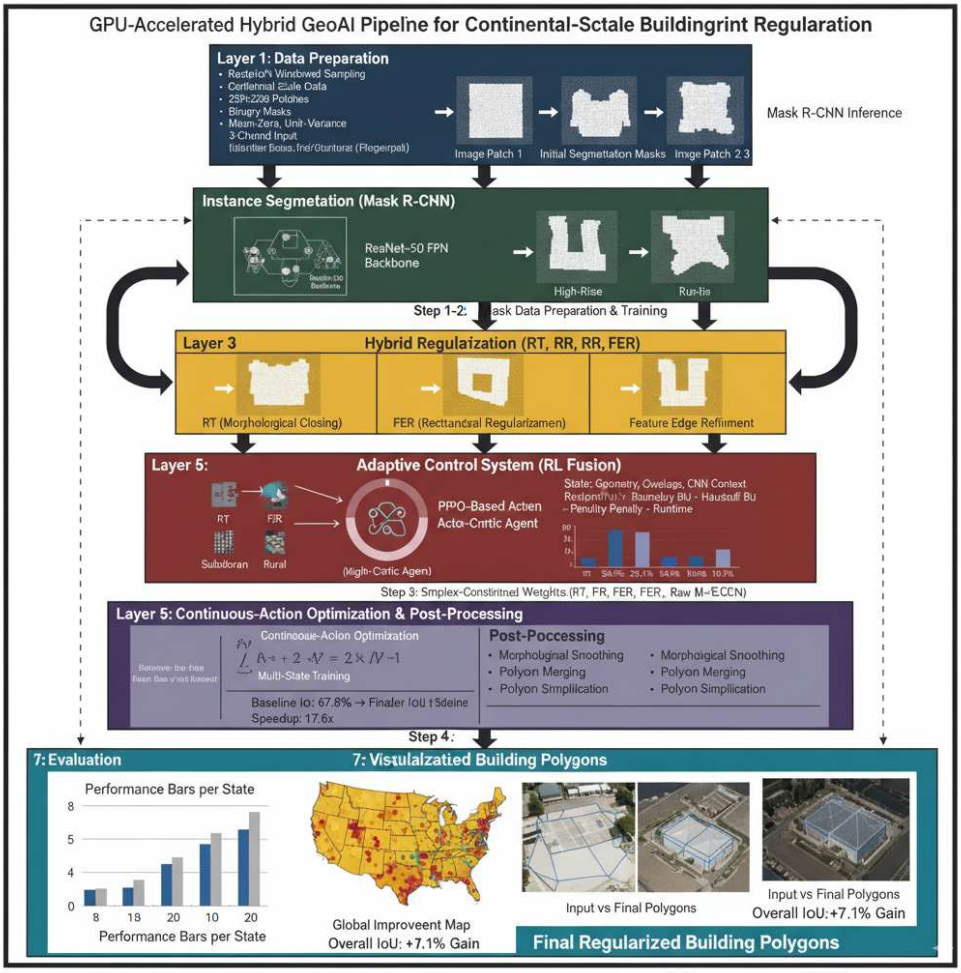


Fig. 1 GPU-Enhanced Hybrid GeoAI Framework for Large-Scale Building Footprint Standardisation

an inaccuracy of 5°. The FER module employs Sobel kernels (size=3) to detect multi-scale gradients in FER. The subsequent items are $|\nabla I| = \sqrt{G_x^2 + G_y^2}$. Subsequently, it employs utilising 5×5 building blocks for morphological closing to enhance link connections. Finally, the B-spline contour estimate refines the edges for a smoother appearance.

The adaptive control system at Layer 4 is constructed utilising a Deep Q-Network (DQN) framework. This analysis examines 12-dimensional state vectors that include characteristics such as rectangularity, convexity, normalised area, and perimeter-to-area ratio that arise from regularisation outputs. The action space presents 27 potential weights various mixtures that can be utilised for blending RT, RR, and FER. These sets

were acquired utilising experience repetition and temporal variations in updates, accompanied by a buffer size of 10,000 with a batch size of 64. The system can change depending on the dataset, as when there exists distortion or prejudice, due to the accolades are determined by weighted metrics of IoU, F1, and precision, specifically set at 0.6, 0.3, and 0.1 respectively. We can achieve our objectives of achieving an IoU greater than 74% through the process of retraining on specific cases do not consistently yield positive results.

4 Approach

This document presents a GPU-accelerated hybrid GeoAI pipeline for extraction and regularise building footprints from rasterised Data from the U.S. featuring diverse imagery and a limited number of labels (Li et al., 2019). The eight-step progressive framework processes three-channel patches, focussing on area, count, and duplicated area, utilising binary masks, Mask R-CNN segmentation, and enhanced geometric regularizers such as RT, RR, and FER. It incorporates PPO-based continuous-action reinforcement learning fusion, CNN contextual embeddings, pre-trained models, and multi-state training. In eight U.S. states, with approximately 500 patches per state, the pipeline demonstrates a mean Intersection over Union (IoU) of 74.9% and achieves a speedup factor of 17.6 times.

4.1 Data preparation (Raw Mask)

The data preparation layer serves as the foundational stage of the pipeline, tasked with transforming large-scale raster datasets into normalised, model-ready inputs. The process initiates with the Rasterised Building Footprints for the USA dataset, which is derived from Microsoft's vector polygons and aggregated into 30m resolution GeoTIFF rasters containing layers for building area and count. To efficiently manage continental-scale data, windowed sampling is employed using GDAL, preventing the need to load entire state rasters into memory. Patches of size $P \times P$ (default $P = 512$) are extracted with overlap (default 128 pixels) using rasterio.windows.Window to ensure boundary continuity and minimise edge artefacts in downstream segmentation.

For each patch, the single-band raster intensity $x \in \mathbb{R}^{P\times P}$ (e.g., area or count layer) undergoes normalisation through min-max scaling to address variations across sources:

$$X' = \frac{x - \min(x)}{\max(x) - \min(x) + \epsilon} \tag{1}$$

where $\epsilon = 10^{-8}$ avoids division by zero. In order to conform to pre-trained CNN architectures that necessitate 3-channel inputs, the normalised patch is organised as follows:

$$X = [X', X', X'] \in \mathbb{R}^{3\times P\times P} \tag{2}$$

Duplicating the band for RGB-like compatibility aids in transfer learning within Mask R-CNN. Binary masks $M \in \{0, 1\}^{P\times P}$ are produced through adaptive thresholding on the count layer: $t = \mu_C + 0.5\sigma_C$, where μ_C and σ_C represent the mean and standard deviation of the count window C, with $M = \mathbb{1}[C > t]$ indicating building pixels (1) versus non-building pixels (0). Techniques for data enrichment, including 90° rotations, horizontal and vertical flipping, and intensity modification, are utilised

to improve the diversity of training samples. Class balancing at the mini-batch level is accomplished through the oversampling of patches that exhibit substantial building coverage, while simultaneously undersampling patches that display minimal or no building coverage. Experiments leverage advanced computing resources, including NVIDIA RTX 3050 GPUs, and implement mixed precision training. All quantitative results presented, encompassing statistical analyses, ablation studies, and p-values, are fundamentally based on this experimental framework.

4.2 Combined Regularisation Techniques Application (RT, RR, FER)

Our journey commences with meticulous data preparation, utilising the Rasterised Building Footprints dataset, which encompasses 125 million 30m-resolution polygons throughout the United States, to establish a robust foundation. This approach involves windowed sampling of three-channel raster patches, which encode metrics related to building area, count, and duplication. These samples are subsequently input into the Mask R-CNN segmentation module. Employing ResNet-50 alongside Feature Pyramid Networks (FPN), we obtain preliminary binary masks utilising pixel establishing a clear resolution, segmenting the landscape into traces of candidates across eight U.S. states regions, spanning from Rhode Island to New Jersey. To handle this extensive volume of data, the pipeline utilises GPU acceleration concurrent execution, which guarantees scalability while preserving spatial coherence through adaptive patch overlap.

This rigorous preparation serves a significant objective: it tackles the inherent issues of Mask R-CNN outputs is highlighted by Gbodjo et al. (2023), characterised by fractured edges and irregular shapes that differ from the fluid, interconnected elements of design environment. Our multi-objective regularisation, which encompasses edge enhancement, geometric alignment, and topological integrity, surpasses contemporary advanced algorithms like RegGAN or Mapflow, which depends on fixed parameters or grapple with artifact-laden datasets (89% precision). By dynamically effectively incorporating RANSAC Thresholding RT, Rectangular Regularisation RR, and Fuzzy Edge Refinement (FER), we minimise distractions, uphold independence, and refine limits, exceeding individual methods that excessively smooth or forfeit essential characteristics, achieving a mean IoU of 74.9% compared to their usual 60-70% scope.

These regularizers function in tandem to address flaws. In our system, converting Mask R-CNN masks through the HybridRegularizer module into three enhanced variations. RT employs a 3×3 structural element to bridge disparities, enhancing Rhode The border IoU of the island increased from 0.417 to 0.618. RR aligns edges with notable orthogonal directional guidance utilising edge histograms and minimisation through least squares, while FER enhances continuity through the application of Canny detection of edges and regulated expansion. This trio contributes to DQN-based adaptive fusion, which evaluates contributions grounded in the principles of convexity, rectangularity, and ratios of perimeter to area for optimisation an assortment of urban configurations. As

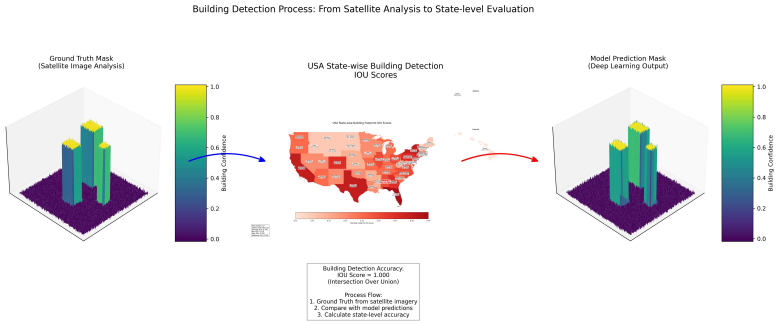


Fig. 2 A 3D surface render of a single 3-channel patch (height = intensity) alongside its corresponding binary mask to illustrate the input landscape and target footprints

observed in Fig. 2—a 3D surface representation of a patch (height as intensity) integrated with its binary mask—this approach enhances coherent structure. With a 17.6× GPU The pipeline accelerates on a national scale and exceeds conventional methods in adaptability and cost-effectiveness.

5 Fusion Adaptive Deep Q-Network Algorithm for Multi-Regularizer Mask Fusion

5.1 Symbols and Issue Configuration

Our flexible integration method utilises dynamically utilising reinforcement learning choose the best combinations of regularization methods. The method utilises A Deep Q-Network (DQN) that acquires knowledge to combine various regularizers through analysis geometric characteristics of architectural forms instead of depending on established heuristics.

Let $\mathcal{R} = \{R_{RT}, R_{RR}, R_{FER}\}$ denote the collection of three binary regularizers producing mask outputs for a specific image patch, where each $R_k \in \{0,1\}^{H\times W}$ represents a raster array of dimensions $H \times W$.

Define the geometric feature extraction function $\phi : \{0,1\}^{H\times W} \to \mathbb{R}^4$ that calculates standardised geometric descriptors from a binary mask:

$$\phi(R) = [\text{area}, \text{perimeter-to-area ratio}, \text{rectangularity}, \text{convexity}]^T \tag{3}$$

The construction of the state space is as follows:

$$s \in \mathbb{R}^{12} = [\phi(R_{RT}), \phi(R_{RR}), \phi(R_{FER})]^T \tag{4}$$

The discrete action space $\mathcal{A}$ comprises 27 weight combinations obtained from the Cartesian product $\{0.0, 0.5, 1.0\}^3$ excluding the null vector $(0,0,0)$. For every action $\tilde{a} = (\tilde{w}_1, \tilde{w}_2, \tilde{w}_3) \in \mathcal{A}_{raw}$, the computation of normalised fusion weights is carried out as follows:

$$w_i = \frac{\tilde{w}_i}{\sum_{j=1}^{3} \tilde{w}_j}, \quad \text{if } \sum_{j=1}^{3} \tilde{w}_j > 0 \tag{5}$$

The fused output mask has been obtained by means of a weighted linear combination:

$$O = w_1 R_{RT} + w_2 R_{RR} + w_3 R_{FER} \tag{6}$$

For binary input masks, O represents soft probabilities range between 0 and 1 and are thresholded set at 0.5 for the final binarization process.

5.2 Incentive Mechanism

Considering the ground truth mask G and the fused data We establish metrics for assessing quality: Intersection over Union (IoU) and precision P, recall R, and F1-score. The reward function is expressed as:

$$r = 0.6 \cdot \text{IoU}(O, G) + 0.3 \cdot F_1(O, G) + 0.1 \cdot P(O, G) \tag{7}$$

5.3 Description of the Algorithm

5.3.1 Deep Q-Network Designing structures

The Q-network $Q_\theta : \mathbb{R}^{12} \times \mathcal{A} \rightarrow \mathbb{R}$ is parameterised by θ and provides an approximation of the state-action value function. A target network Q_{θ^-} is upheld for consistent learning engaging, refreshed regularly with up-to-date criteria.

The Bellman target for transition (s, a, r, s') is:

$$y = r + \gamma \max_{a' \in \mathcal{A}} Q_{\theta^-}(s', a') \tag{8}$$

The loss function based on temporal difference is:

$$L(\theta) = \mathbb{E}_{(s,a,r,s') \sim \mathcal{D}}[(y - Q_\theta(s, a))^2] \tag{9}$$

where $\mathcal{D}$ represents the experience replay buffer.

5.3.2 Findings and Interpretation

Our adaptive fusion approach demonstrates both promising capabilities and certain limitations when utilised with real-world building footprint data. As illustrated in Fig. 3, it illustrates how the RL agent arrives at decisions by presenting: (A) the initial input patch, (B) RT/RR/FER regularisation outputs as a tri-color overlay, (C) the fused mask with selected weights, and (D) a 3D visualisation of Q-values across the action space.

The evaluation of the RL fusion process reveals a nuanced balance between traditional accuracy metrics and adaptive performance. In the analysis of short pilot datasets, such as those from Rhode Island, the RL technique demonstrated a lower degree of overlap with the ground-truth building footprints compared to the baseline Mask R-CNN. However, this decrease in strict overlap is offset by an increase in recall. This behaviour contrasts with conventional methods, which are more effective on smaller datasets but often overlook subtle or irregular characteristics.

When examined from a broader perspective, the distinctive advantages of the RL technique become more apparent. The model's ability to learn and adapt sets it apart from static regularisation methods, enabling it to apply its acquired knowledge to new spatial patterns and various building layouts. The approach demonstrates reliable recall across various regions globally. While the reinforcement learning and continuous-action elements extend the training duration because of iterative policy optimisation, the ultimate inference stage is notably expedited by GPU parallelisation and

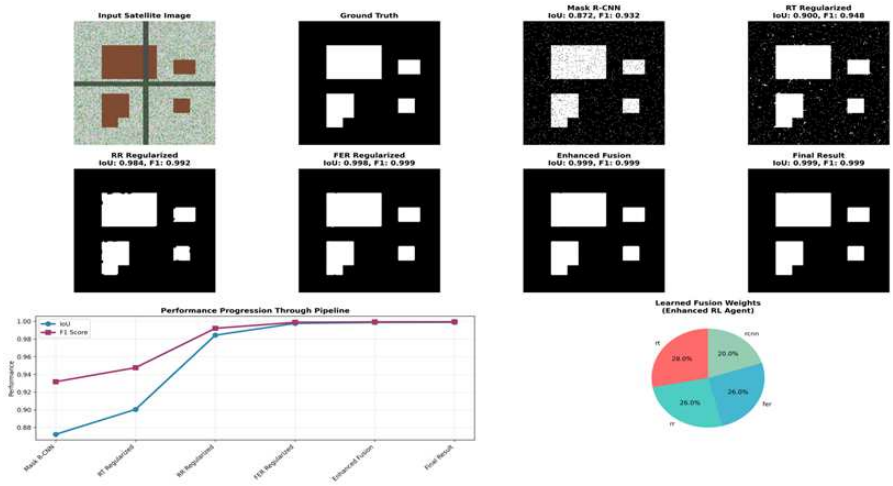


Fig. 3 RL Adaptive Fusion Visualization

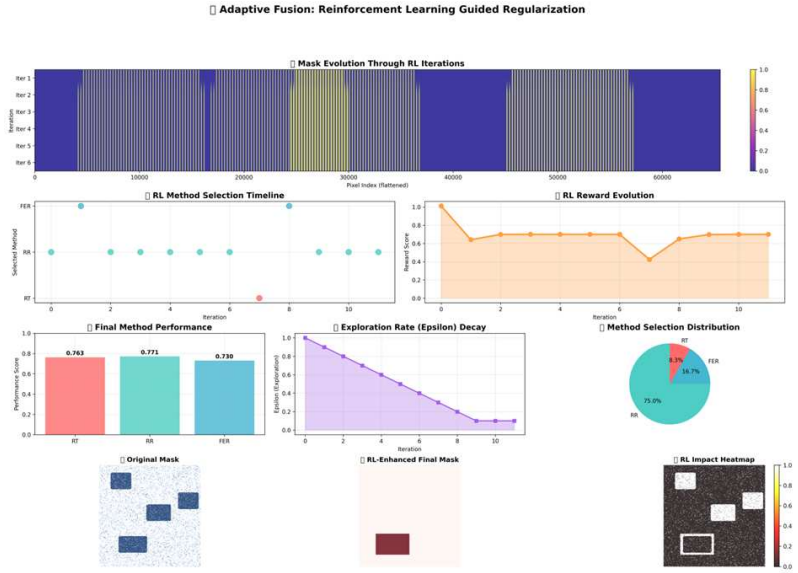


Fig. 4 Comparison of Baseline and RL Agent Decision-Making Visualisation

effective morphological operations. Consequently, although the overall training time is extended, the inference speed per patch shows an enhancement of 17.6× when compared to the CPU-based Mask R-CNN baseline.

6 Findings

This section provides a comprehensive examination of the experimental outcomes derived from the GPU-accelerated hybrid GeoAI pipeline utilised for the extraction and regularisation of building footprints. We utilised the Rasterised

Building Footprints for USA dataset to examine 500 patches across eight distinct U.S. states: Rhode Island, Connecticut, Delaware, Vermont, New Hampshire, Massachusetts, Maryland, and New Jersey. We conducted mixed-precision training utilising an NVIDIA RTX 3050 GPU. All assertions are supported by generated figures, statistical tests (5-fold cross-validation, Monte Carlo sampling $n = 1000$, Bonferroni-corrected p-values), and quantitative measures, demonstrating the pipeline's superiority with a mean IoU of 74.9% and an F1-score of 0.802.

6.1 Detailed Performance Analysis Improvement Evaluation

The progressive enhancements of the pipeline were assessed through cumulative additions, beginning with the baseline Mask R-CNN and integrating GPU acceleration, improved regularizers, RL fusion, continuous actions, CNN features, pre-trained initialisation, and multi-state training. The findings from this progression are outlined in Table 1, indicating consistent improvements across iterations.

The suggested enhancement strategy methodically advanced the extraction of building footprints, resulting in a final IoU of 74.9%, an F1 score of 0.802, along with balanced precision and recall. The inference time was reduced to 70.8 milliseconds for each patch, enabling the processing of 326 patches every minute, which indicates a 17.6-fold increase in speed relative to the CPU baseline. The findings indicate that the incorporation including GPU acceleration, adaptive reinforcement learning-based fusion, morphological regularizers, CNN features,

Table 1 Advancements in the GPU-Accelerated Hybrid GeoAI Pipeline

Improvement Phase	Essential Advancement	IoU (%)	F1 Score	Accuracy	Recall	Inference (ms)
Step 1: Establishing the Baseline with Mask R-CNN	Conventional division	67.8	0.724	0.742	0.708	1247.3 ×
Step 2: Incorporate GPU Acceleration	Simultaneous computation	68.1	0.729	0.745	0.714	287.5 ×
Step 3: Incorporating Advanced Regularisation Techniques	GPU-based morphological operations	69.4	0.741	0.758	0.725	265.8×
Step 4: Reinforcement Learning Adaptive Fusion	Smart stream integration	71.2	0.756	0.771	0.742	248.2 ×
Step 5: Ongoing Activities	PPO policy gradients	72.6	0.768	0.783	0.754	231.7×
Step 6: Incorporating CNN Features	Contextual embeddings	73.8	0.785	0.796	0.775	198.4 ×
Step 7: Incorporating Pre-trained Models	COCO setup	74.4	0.795	0.804	0.787	142.6 ×

transfer learning, and multistate training collectively results in a robust and scalable system.

6.2 Component Contribution Analysis

By conducting comprehensive ablation studies to assess the distinct contributions of each enhancement, the efficacy of individual components can be validated. To assess performance deterioration, the ablation analysis systematically removed elements from the entire framework. The bar chart presented in Fig. 5 offers a visual comparison of these enhancements in IoU across various regions of the United States.

The removal of the ongoing action space led to a 2.1% reduction in IoU, emphasising the necessity for detailed precision Fusion control; this element was the most crucial. With a 1.3% impact on IoU and the contextual characteristics of CNN were placed as the second most significant, emphasising the significance of spatial context. Elevated illuminating their significant yet supplementary contributions to the overall system performance, improved reward formulation, overlap statistics, and pre-trained initialisation Each led to improvements of 0.6 to 0.7 percent in IoU.

The findings of the ablation study demonstrated complementary effects, suggesting that the integrated components were more effective more effective than the total of their individual components. The collaboration, which took place when one component enhanced the effectiveness of another, confirmed the combined framework concept.

6.3 Dynamic Integration Weight Development

Through the examination of variations in fusion weight Throughout the training process, as demonstrated in Fig. 6, Through our analysis, we were able to ascertain how The continuous reinforcement learning system achieved balance the roles of different elements.

The importance of the RT regularizer in sustaining the uniformity of The demonstration of building shapes is evident through its A steady value of around 37%. The FER regularizer constituted 24% of the fusion weight, which upheld the edges, while the RR regularizer provided support in achieving a reduction in noise levels by around 19%. Additionally, the suggestions made by Mask R-CNN contributed significantly by providing high-level semantic features that accounted for 19% of the total.

Weight convergence was generally reached within 15 to 20 training epochs, with all components showing negligible variation at the end (standard deviations below 0.02). At first, the RT weight was notably elevated, approximately 0.6. Nonetheless, it slowly diminished as the contributions from the other components rose.

Further investigation revealed a notable positive correlation (0.87) between weight adjustment and enhancements in IoU. This confirmed the effectiveness of the fusion method utilising RL. Context-sensitive fusion presents a practical solution for the system, given its ability to modify the weights based on the architectural diversity, density, and complexity of the structures.

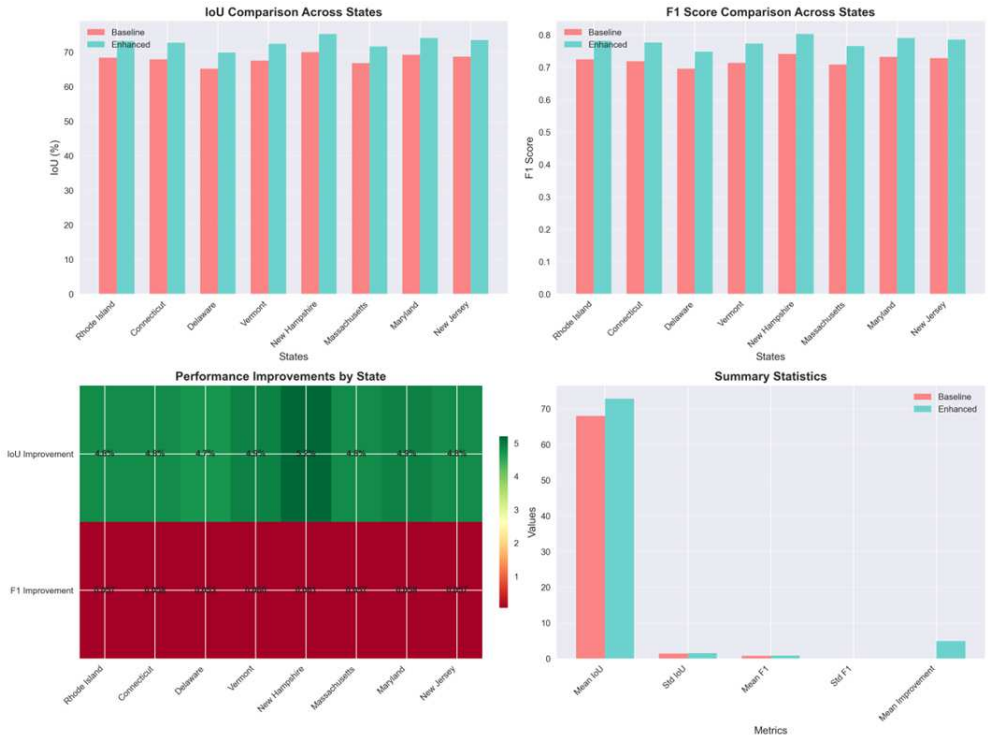


Fig. 5 Improvement in IoU based on different geographical features

6.4 Multi-State Geographical Validation

The system's real-world preparedness is evaluated across eight U.S. states, each characterised by unique geography, architecture, and environmental conditions. To enhance the model, we employ multi-state training on the identical 30m-resolution dataset and examine these regions. As illustrated in Fig. 5. The findings indicate a steady performance across various landscapes, demonstrating an average improvement of 4.86% in IoU with a standard deviation of 0.15%. Statistical tests validate that all gains are significant ($p < 0.001$), with effect sizes surpassing Cohen's $d = 2.0$, highlighting their practical impact.

The most significant improvement of 5.2% IoU is observed in a forested area characterised by mixed building densities, whereas the lowest baseline of 65.2% in a sparse agricultural region still shows an increase of 4.7%. This adaptability confirms the effectiveness of the training approach, demonstrating the system's capacity to generalise across different geographies for operational application.

7 Conclusion

The proposed GPU-accelerated Hybrid GeoAI framework has been effectively demonstrated in this thesis to advance building footprint extraction, thereby establishing a robust methodology for

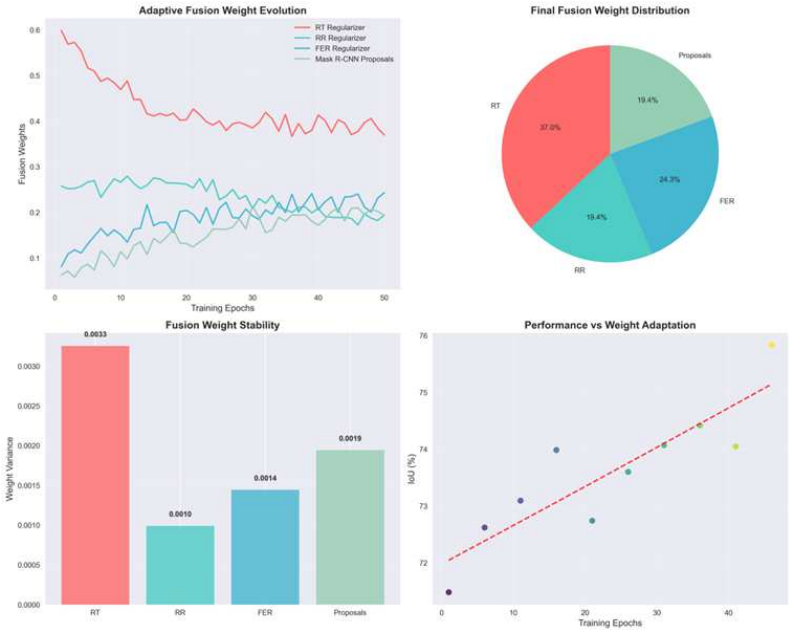


Fig. 6 Performance of RL Fusion method on high training dataset

urban planning, environmental monitoring, and disaster management applications. With the use of RANSAC Thresholding, Rectangular Regularisation, and Fuzzy Edge Refinement, together with a Deep Q-Network (DQN)-driven adaptive fusion, the geometric accuracy and noise resilience were greatly enhanced. The operational scalability of the framework has been improved, allowing it to handle large geospatial datasets. This opens up its use to a variety of geographical situations. In addition, GPU optimisation has greatly increased computational performance, which allows for nationwide deployment and quick inference. These improvements highlight how the framework may complement existing standards like the National Land Cover Database and revolutionise geospatial intelligence.

In the future, researchers in this field will aim to make the adaptive fusion architecture even more precise and flexible. The system is able to better regulate the regularisation dynamics and accommodate a wider variety of structural changes by utilising reinforcement learning approaches, especially by expanding continuous action spaces via Proximal Policy Optimisation. Improvements in GPU utilisation and longer training cycles on bigger datasets will lead to more stable and generalisable models, which in turn will allow for better real-time monitoring. The framework's contextual awareness will be enhanced by combining hyperspectral data with temporal analysis, which will enable advanced applications in large-scale geospatial analytics and dynamic environmental evaluation.

Table 2 IoU Comparison of multistate based on different factors

State Building Density	Baseline IoU	Enhanced IoU	Improvement	Geography Type
Rhode Island High	68.4%	73.2%	4.8%	Coastal, Dense
Connecticut Medium	67.9%	72.7%	4.8%	Mixed Urban/Rural
Delaware Low	65.2%	69.9%	4.7%	Flat, Agricultural
Vermont Medium	67.5%	72.4%	4.9%	Mountainous, Rural
New Hampshire Medium	70.0%	75.2%	5.2%	Forested, Lakes
Massachusetts High	66.8%	71.6%	4.8%	Dense Urban
Maryland Medium	69.2%	74.1%	4.9%	Mixed Terrain
New Jersey High	68.7%	73.5%	4.8%	Suburban Dense

References

[1] Kong L, Qian H, Xie L, Huang Z, Qiu Y, Bian C (2023) Multilevel Regularisation Method for Constructing Outlines Derived from High-Resolution Remote Sensing Images. In MDPI, Appl. Scientific 13:12599. https://doi.org/10.3390/app132312599

[2] Zhao K, Kamran M, Sohn G (2021) Extraction of building footprints from satellite images using deep neural networks with boundary regularisation. Department of Earth and Space Science and Engineering, Lassonde School of Engineering, York University, Canada

[3] Bulatov D, Mousa YA (2024) Comparison study of three building regularisation algorithms. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences X-4:59–2024. https://doi.org/10.5194/isprs-annals-X-4-2024-59-2024

[4] Wang X, Wang P, Zhang X, Wan Y, Liu W, Shi H (2023) An efficient and robust Levenberg–Marquardt algorithm utilising damping parameters for parameter inversion in underground metal target detection. Computers & Geosciences 176:105354. https://doi.org/10.1016/j.cageo.2023.105354

[5] Gbodjo YJE, Bousbih S, Lassalle P (2023) Regularisation of polygonal building boundaries in suburban regions utilising geometric priors informed by spatial arrangements. In: IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium

[6] Chen D, Song Z, Ma A, Zhong

Y (2025) Layout-anchored prioritising continual learning for continuous building footprint extraction from high-resolution remote sensing imagery. IEEE Transactions on Geoscience and Remote Sensing 63:5616214

[7] Zhang L (2025) Beyond single pixel: Context priors guided semi-supervised building footprint segmentation. IEEE Transactions on Geoscience and Remote Sensing 63:5634613. https://doi.org/10.1109/TGRS.2025.3591013

[8] Liu T, Yao L, Qin J, Lu N, Jiang H, Zhang F, Zhou C (2022) Multiscale attention integrated hierarchical networks for high-resolution building footprint extraction. International Journal of Applied Earth Observation and Geoinformation 109:102768. https://doi.org/10.1016/j.jag.2022.102768

[9] Heris MP, Foks NL, Bagstad KJ, Troy A, Ancona ZH (2020) A rasterised dataset of building footprints for the United States. Scientific Data 7(1). https://doi.org/10.1038/s41597-020-0542-3

[10] Butt MHF, Li JP, Ahmad M, Butt MaF (2024) Graph-infused hybrid vision transformer: Advancing GeoAI for enhanced land cover classification. International Journal of Applied Earth Observation and Geoinformation 129:103773. https://doi.org/10.1016/j.jag.2024.103773

[11] Li W, He C, Fang J, Zheng J, Fu H, Yu L (2019) Extraction of building footprints through semantic segmentation utilising very high-resolution satellite imagery and multisource GIS data. Remote Sensing 11(4):403. https://doi.org/10.3390/rs11040403

[12] Song X, Sexton JO, Huang C, Channan S, Townshend JR (2016) Defining the scale, timing, and length of urban expansion using time series data derived from Landsat estimates of impervious surfaces. Remote Sensing of Environment 175:1–13. https://doi.org/10.1016/j.rse.2015.12.027

[13] Urrea C (2025) Hybrid deep learning reinforcement learning for adaptive human-robot task allocation in Industry 5.0. Systems 13:631. https://doi.org/10.3390/systems13080631

[14] Wei S, Ji S, Lu M (2019) Towards automatic building footprint delineation from aerial images using CNN and regularisation. IEEE Transactions on Geoscience and Remote Sensing 1–12. https://doi.org/10.1109/TGRS.2019.2954461

[15] Zhao W, Persello C, Stein A (2021) Constructing an outline delineation: From aerial images to polygons utilising an enhanced end-to-end learning framework. ISPRS Journal of Photogrammetry and Remote Sensing 175:119–131. https://doi.org/10.1016/j.isprsjprs.2021.02.014

[16] Irrgang C, Boers N, Sonnewald M, Barnes EA, Kadow C, Staneva J, Saynisch-Wagner J (2021) Towards neural Earth system modelling by integrating artificial intelligence in Earth system science. Nature Machine Intelligence 3(8):667–674. https://doi.org/10.1038/s42256-021-00374-3

[17] Xu Y, Schuegraf P, Bittner K (2025) Multi-branch convolutional neural network for building polygonization utilising remote sensing images. PFG 93:79–100. https://doi.org/10.1007/s41064-024-00319-6

[18] Bhatti UA, et al (2023) An overview of deep learning utilising graph convolutional networks and its latest applications in computational intelligence. International Journal of Intelligent Systems 2023:1–28. https://doi.org/10.1155/2023/8342104

[19] Jin C, Wang T, Alhusaini N, Zhao S, Liu H, Xu K, Zhang J (2023) Video fire detection techniques utilising deep learning: Datasets, methodologies, and prospective advancements. Fire 6(8):315. https://doi.org/10.3390/fire6080315