Journal of IoT in Social, Mobile, Analytics, and Cloud (ISSN: 2582-1369)
www.irojournals.com/iroismac/

Hybrid Graph-Kernel Fusion Network for Heritage Point Cloud Segmentation

Geetha Ramani R.[1], Mirsha A K.[2], Piruthviraj V S.[3]

[1]Department of Information Technology, Anna University, Chennai, India.

[2,3]Department of Artificial Intelligence and Data Science, Anna University, Chennai, India.

E-mail: [1]rgeetha@yahoo.com, [2]a.k.mirsha9@gmail.com, [3]harinirajamsairam@gmail.com

Abstract

Heritage point cloud semantic segmentation is generally used for digital heritage preservation, Heritage Building Information Modeling (H-BIM) and intelligent architectural analysis. However, it is difficult to identify various architectural elements because of the non-uniformity of the point clouds, complexity in geometry and structural occlusions. This research work proposes a novel architecture called the Hybrid Graph-Kernel Fusion Network (HKGFNet), which combines the graph and kernel approaches in terms of contextual information and feature extraction respectively, via a bi-directional cross-attention and gating technique for performing discriminative segmentation. The proposed framework is tested using the Images & PointClouds Cultural Heritage dataset containing labeled 3D point clouds of heritage buildings. The proposed architecture is designed using a systematic procedure consisting of data preprocessing, training and evaluation with a 70:15:15 dataset split. The experiment results yielded a mean Intersection over Union (mIoU) of 90.79% and overall accuracy of 96.15%, while the ablation study validated the efficacy of the hybrid fusion method proposed. The resulting segmentation revealed accurate identification of the various architectural elements and maintained the structural boundaries of the building, implying the suitability of HKGFNet in heritage documentation and Scan-to-BIM processes.

Keywords: Semantic Segmentation, Heritage Building Information Modeling (H-BIM), Point Cloud Processing, Hybrid Graph–Kernel Fusion, 3D Cultural Heritage Analysis.

DOI: https://doi.org/10.36548/jismac.2026.3.005
Extended version of article received from **International Conference on Smart Innovations in Engineering & Technology (ICSIET 2026)**
Received: 18.05.2026, received in revised form: 22.06.2026, accepted: 12.07.2026, published: 23.07.2026

1. Introduction

The fast evolution of 3D sensing techniques, including LiDAR, photogrammetry, and laser scanning, has made it possible to create dense point clouds, which can be used to accurately describe real-life environments. Point clouds have become an important source of information in such applications as urban mapping, digital twin creation, robotic applications, and cultural heritage conservation. There are many point cloud analysis methods, among which semantic segmentation is very important because it allows assigning semantic labels to every single point, which helps to recognize architectural parts and creates the foundation for H-BIM. Automated semantic segmentation of point clouds significantly saves time in conservation projects as compared to traditional manual methods.

Modern developments in deep learning have resulted in the development of point cloud semantic segmentation algorithms that use point-based, graph-based, kernel-based, and transformer-based models. While the graph-based model is adept at capturing the relationships between neighboring points, the kernel-based convolutions perform exceptionally well in capturing local geometric details of irregularly distributed points. Nevertheless, the task becomes increasingly difficult in case of heritage point clouds because of their geometric complexity, ornamental nature, varied densities of points, occlusions, and imbalance in classes. The architectural parts of the structures such as vaults and arches have different geometric characteristics.

In order to overcome the aforementioned challenges, this study proposed an approach that leverages a novel Hybrid Graph-Kernel Fusion Network (HKGFNet). The main idea of this framework is to combine dynamic graph-based neighborhood relationships with kernel-based extraction of local geometrical properties of objects to jointly train on local and structural features. A bidirectional cross-attention scheme is used to merge features obtained by both branches, thus allowing the network to learn detailed architectural structures while considering global context information. The proposed model is experimentally validated using Images & PointClouds Cultural Heritage Dataset, wherein the proposed hybrid architecture outperforms other models in terms of accuracy.

2. Literature Survey

Semantic segmentation of point clouds are recently employed in various 3D sensing modalities, such as LiDAR, photogrammetry, and 3D scanning in applications like urban mapping, autonomous vehicles, and heritage documentation. The extraction of geometric features and their subsequent accurate classification is vital for scene understanding and modeling. Recent works on advanced imaging and characterization techniques have stressed the necessity of feature extraction and representation learning for the analysis of complex structures at different scales, showing the importance of computational techniques in spatial data analysis [1]. Despite the hype, the inherent irregularity and lack of order in point cloud datasets pose challenges to feature extraction and semantic classification processes. Deep learning methods have greatly benefited the task of point cloud segmentation by performing geometric representation learning on point clouds [2].

The early point-based methods used hierarchical feature learning methods that learned local geometric structures from neighborhood aggregations, which were able to perform semantic understanding of 3D scenes [2]. To further improve modeling of spatial relationships between points, graph-based models were introduced to dynamically build neighborhood graphs and learn interactions of points within those graphs. The methods successfully captured the structural dependencies and context, and consequently performed better segmentation in complex scenes [3]. Kernel-based convolutions have also proven to be very efficient due to direct usage of learnable kernels on non-uniform point clouds. Such methods are good at capturing local geometric structures and preserving object boundaries, therefore are applicable in architectural and heritage use cases [4]. However, the graph-based methods mainly concentrate on relation information, and kernel-based methods on local geometry.

The recent development includes the introduction of transformer architecture using self-attention technique to utilize local and global context. The stratified attention technique improves the segmentation accuracy in large scale scenes. In addition, the use of self-supervised pre-training method boosts the representation learning process in feature extraction from unlabeled point cloud [5], [6]. Despite their success, these models still suffer from computational complexity and large training datasets requirement. The efficient architecture models have been designed to provide the trade-off between computational efficiency and segmentation accuracy in large-scale point clouds processing [7]. The random sampling and efficient feature aggregation can make it possible to perform practical semantic segmentation in massive 3D point clouds with millions of points.

In the area of cultural heritage, semantic segmentation is highly important for H-BIM, restoration and digital preservation. It has been proven through comparative analysis that deep learning methods are better than machine learning methods in identifying the elements of the architecture of the historic building [8]. Moreover, multi-view deep learning systems have provided enhanced results for segmentation due to the inclusion of information from different perspectives [9], [10].

In spite of considerable advances in the field of point cloud semantic segmentation, existing techniques primarily concentrate either on graph-based relational learning or on kernel-based feature extraction from geometry. Graph-based techniques are effective in capturing structure information, but they can neglect the local geometric information. On the other hand, kernel-based techniques are very effective in capturing local geometric information but they are lacking global context information. These limitations are even overstated when applied to heritage datasets, which contain complex architectural components, complex geometry, occlusions, and class imbalance. Thereby, there is an increasing need for proposing a new method that can effectively learn both global structure information and local geometry information.

3. Proposed Methodology

The proposed Hybrid Graph-Kernel Fusion Network (HKGFNet) model is meant for segmenting heritage point cloud data into semantically meaningful classes. The structure of the network incorporates the strengths of graph-based neighborhood reasoning as well as kernel-based geometric feature extraction to learn the relationship and geometric properties at the same time. Such combination provides the successful description of various architectural forms often found in heritage data.

3.1 System Architecture

The HKGFNet model aims at performing semantic segmentation of cultural heritage point clouds by combining graph-based contextual analysis with kernel-based geometric feature extraction in one deep learning model. The proposed framework involves various stages, which include data preprocessing, data preparation, feature embedding, hybrid graph–

kernel representation learning, cross-attention fusion, adaptive gating, model training, and semantic segmentation. To begin with, the point clouds obtained from heritage objects are preprocessed in order to enhance their quality and normalize the geometrical representations used in the network. After that, the hybrid model extracts complementary structural and geometrical features using graph and kernel branches respectively, and the representations are combined by the cross-attention mechanism and adaptive gating module to generate discriminative semantic features. Finally, the optimized feature representations are classified into predefined H-BIM semantic categories using a point-wise classification head.

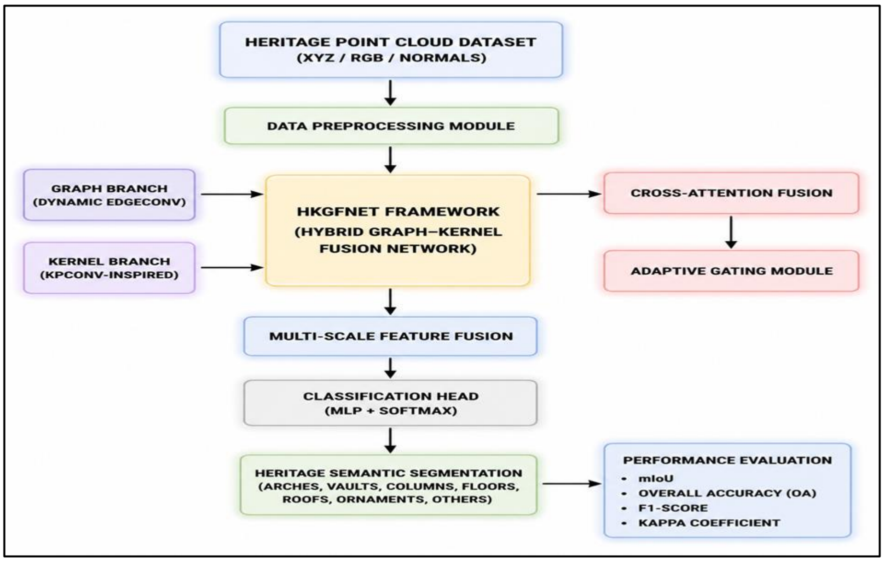

Figure 1. Hybrid graph-kernel fusion network for semantic segmentation framework

The overall workflow of the proposed HKGFNet framework is illustrated in Figure 1, and the individual stages are described in the following subsections.

3.2 Input Representation and Feature Engineering

The input consists of three-dimensional point clouds represented by spatial coordinates (x, y, z). Before the process of feature learning, the point clouds are normalized to mitigate any scale differences. This is then followed by projecting the normalized coordinates to a high-dimensional feature space via an embedding layer.

$$\tilde{p}_i = \frac{p_i - \mathrm{mean}(P)}{\max(|p_i - \mathrm{mean}(P)|)} \tag{1}$$

where $\tilde{p}_i$ denotes the normalized value of the i-th element, p_i represents the original value of the corresponding element, and $\mathrm{mean}(P)$ is the arithmetic mean of all elements in the set P. The term $|p_i - \text{mean}(P)|$ measures the absolute deviation of each value from the mean, while $\max(\cdot)$ identifies the largest such deviation in the entire set. As a result, each value around zero by subtracting the mean and then scales it using the maximum absolute deviation, which makes the transformed values easier to compare across the dataset. Normalization is commonly used to place values on a comparable scale so that no single feature dominates purely because of its magnitude.

3.3 Data Preprocessing

Heritage point cloud data were first processed by preprocessing to improve the quality of the dataset and provide consistent input. For the removal of noise data caused by data capturing, the use of statistical outlier removal (SOR) method was adopted while the voxel down-sampling method was used to reduce the complexity of computation without much impact on the geometry. To enhance robustness and generalizability, all the point clouds were normalized before training. Data augmentation methods that include random rotation, random scaling, Gaussian jittering, and structural point dropout were used in training.

3.4 Dataset Description

The HKGFNet framework has been tested based on the Images & PointClouds Cultural Heritage Dataset [11]. This publicly accessible benchmark has been designed for semantic segmentation and H-BIM application purposes. The dataset is made up of different heritage buildings with annotated semantic points. The main features of the dataset are presented in Table 1 below.

Table 1. Characteristics of the images and PointClouds cultural heritage dataset

Attribute	Specification
Dataset Name	Images & PointClouds Cultural Heritage Dataset
Source	Zenodo Repository [11]
Heritage Buildings	5
Data Modalities	Photogrammetric Images and 3D Point Clouds
Input Features	XYZ Coordinates, RGB Values, Surface Normals
Annotation Type	Point-wise Semantic Labels
Number of Semantic Classes	10

Semantic Categories	Wall, Roof, Vault, Column, Moulding, Floor, Door/Window, Stair, Arch, Other
Camera Information	Intrinsic and Extrinsic Calibration Parameters
Primary Application	Heritage Semantic Segmentation and H-BIM
Experimental Dataset Split	Training (70%), Validation (15%), Testing (15%)

3.5 Multi-Scale Hybrid Processing Stages

The key component of the HKGFNet architecture involves various processing layers that are arranged in sequence, wherein both the graph and kernel-based feature extraction components function simultaneously. This dual channel design allows the model to capture complementary information from the same input data.

Graph Module (Dynamic EdgeConv): The graph module uses the Dynamic Edge Convolution (EdgeConv) approach that can capture the interaction between adjacent points through a Dynamic Edge Convolution (EdgeConv) approach. At each point, the graph is built dynamically through k nearest neighbors (k-NN) according to the current representation of the features. The relative positional information between the center point and adjacent points is captured using a positional encoding network which makes it easier for the model to learn local dependencies.

Kernel Module (Geometric Field Branch): Similar to the graph branch, the kernel branch extracts geometric features by means of a KPConv-based approach. Kernel points, which are a group of learnable points in the neighborhood, allow the model to extract geometric features from irregular point clouds. The kernel branch captures local surface features, including curvatures and boundaries, through weighted aggregation of point features in the neighborhood. These properties are vital to detect complex architectural features such as moulding, decorations, and column capitals.

3.6 Cross-Attention Fusion and Gating Mechanism

In order to integrate the complementary features learned through the graph and kernel branches, HKGFNet uses a two-way cross-attention fusion strategy after each processing step.

Cross-Attention: In this system, features produced from the graph branch will be leveraged in improving kernel features, whereas features produced from kernel features will serve to provide geometry-based information for the graph branch. Through this process of information sharing, the network will gain better insight into heritage architecture elements.

Adaptive Gating: After the process of feature interaction, there comes an adaptive gating mechanism that helps decide the importance of the two kinds of representations – graph-based and kernel-based. This is done using a lightweight MLP that estimates feature weights based on aggregated context. The fused representation obtained after this stage dynamically adjusts itself depending on the properties of the input region and achieves a balance between local geometry and global structure. This way, the model can focus on geometry for decorative components and on structure for larger surfaces.

$$F(s) = \mathrm{Linear}\left(\sigma \cdot F_{\mathrm{att}_{\mathrm{G}}} + (1-\sigma) \cdot F_{\mathrm{att}_{\mathrm{K}}}\right) \tag{2}$$

where, $F_{\mathrm{att\_G}}$ and $F_{\mathrm{att\_K}}$ denote the graph-attended and kernel-attended feature representations, respectively. The learnable gating coefficient σ controls the relative contribution of each feature branch, while the $\mathrm{Linear}(\cdot)$ operation projects the fused representation into a unified feature space for subsequent classification.

3.7 Dataset Preparation and Model Training Strategy

The preprocessed point clouds were fed into the HKGFNet architecture as inputs. While training, the two branches with graph and kernel representations learned complementary features from the same input data. The cross-attention fusion allowed for information exchange among the two types of features, whereas the adaptive gating helped to balance their contribution in the process of semantic feature learning. The Adam optimizer was used for optimizing the model parameters since Adam exhibited stable convergence behavior when training deep models. The cross-entropy loss function was utilized in guiding the parameter learning in the course of training. The process of training went on iteratively until meeting the convergence criterion. Finally, the trained model was tested on the testing dataset for evaluating its performance in semantic segmentation.

3.8 Loss Function

Semantic segmentation was formulated as a multi-class classification problem in which each point was assigned to a predefined H-BIM category. The objective function was based on categorical cross-entropy loss, which can be expressed as

$$L = -\sum_{i=1}^{N} \sum_{c=1}^{C} y_{ic} \log(\hat{y}_{ic}) \tag{3}$$

where, N represents the total number of points, C denotes the number of semantic classes, y_{ic} is the ground-truth label, and $\hat{y}_{ic}$ represents the predicted probability for class c. Minimization of this loss encourages accurate point-wise classification across all heritage architectural categories.

3.9 Training Algorithm

The overall training procedure of the proposed HKGFNet framework is summarized in Algorithm 1.

Algorithm 1. HKGFNet training and evaluation procedure

Input: Preprocessed heritage point cloud dataset D
Output: Trained HKGFNet model M and semantic segmentation results
Begin
1: Load heritage point cloud dataset D
2: Preprocess D using SOR, voxel downsampling, and normalization
3: Perform data augmentation
4: Split D into training, validation, and testing subsets
5: Initialize HKGFNet parameters
6: while convergence criterion is not satisfied do
7: Generate point feature embeddings
8: Extract graph features using Dynamic EdgeConv
9: Extract kernel features using KPConv-inspired module
10: Fuse both features through bidirectional cross-attention
11: Refine fused features using adaptive gating
12: Predict point-wise semantic labels
13: Compute categorical cross-entropy loss
14: Update network parameters using Adam optimizer
15: end while
16: Evaluate the trained model on the testing dataset
17: Compute mIoU, Overall Accuracy, Precision, Recall, F1-score, and Kappa
18: Generate semantic segmentation map S
End

4. Results and Discussion

The proposed HKGFNet was evaluated by using the Images and PointClouds Cultural Heritage Dataset [11]. The performance evaluation was carried out based on mIoU and overall accuracy, which have been commonly used for performance measurement in semantic segmentation.

4.1 Quantitative Performance Analysis

Quantitative analysis of the proposed HKGFNet with state-of-the-art point cloud semantic segmentation techniques is shown in Table 2.

Table 2. Comparative performance analysis of semantic segmentation models

Model	mIoU (%)	Overall Accuracy (%)
KPConv [4]	85.1	86.4
MP-DGCNN [3]	85.2	84.1
HKGFNet (Proposed model)	90.79	96.15

Table 2 shows the segmentation capabilities of HKGFNet in comparison to the representative point cloud semantic segmentation architectures such as KPConv and MP-DGCNN. While KPConv concentrates mostly on extracting geometric features, MP-DGCNN relies on neighborhood-based structure learning. Through hybrid graph-kernel fusion that provides a representation of both geometric and contextual features, HKGFNet is capable of generating more discriminative semantic features and consistency in segmentation of complicated heritage point clouds.

4.2 Final Semantic Segmentation Output

The resultant qualitative semantic segmentation outputs achieved via the suggested HKGFNet framework are shown in Figure 2 below. This figure shows the process of transforming the input heritage point cloud to a semantically labeled one, thereby illustrating the ability of the framework to provide semantically labeled outputs for intricate architecture without changing the geometry.

Figure 2. Semantic segmentation results generated by the proposed HKGFNet

These qualitative results are in line with the quantitative and ablation study presented above. Through the use of the hybrid approach of using graph context and kernel geometrical features, the framework is able to accurately distinguish between different parts of architecture while maintaining the geometry intact. The bidirectional cross-attention and adaptive gate contribute significantly to feature discrimination and generate accurate semantics that can be used in heritage building information modeling, Scan to BIM workflows, documentation and cultural heritage preservation. The visual output presented here is in line with the quantitative and ablation studies performed above. The clear separation between different parts of architecture indicates the success of the fusion in providing reliable semantic labeling of heritage architecture.

5. Conclusion

This study addressed the challenge of semantic segmentation for cultural heritage point clouds by developing the Hybrid Graph–Kernel Fusion Network (HKGFNet), which integrates graph-based contextual learning and kernel-based geometric feature extraction through a bidirectional cross-attention mechanism and adaptive gating strategy. The developed architecture is capable of capturing both geometric properties and contextual relations at different scales. The performance of the proposed framework was evaluated on the Images & PointClouds Cultural Heritage Dataset, where HKGFNet managed to achieve 90.79% mean Intersection over Union (mIoU) and 96.15% overall accuracy. Furthermore, the ablative study proved the benefit of the hybrid graph-kernel fusion architecture regarding the learning of more informative representations. The obtained results demonstrate that the presented framework can provide consistent and reliable segmentation of architectural elements while preserving their structural coherence, which makes the method applicable to the tasks of H-BIM generation, Scan-to-BIM processing, digital documentation, and cultural heritage preservation. As part of future work, efforts will be directed towards the development of HKGFNet by implementing transformer-based feature learning, multimodal fusion, self-supervised and domain adaptation approaches.

References

[1] Aziz, Abbas, Huma Shaikh, Amna Abbas, Kissa E. Zehra, and Bakhtawar Javed. "Microscopic Techniques for Nanomaterials Characterization: A Concise Review." Microscopy Research and Technique 2025, vol 88, no. 5: 1599-1614.

[2] Qi, Charles Ruizhongtai, Li Yi, Hao Su, and Leonidas J. Guibas. "Pointnet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space." Advances in Neural Information Processing Systems 2017:30.

[3] Wang, Yue, Yongbin Sun, Ziwei Liu, Sanjay E. Sarma, Michael M. Bronstein, and Justin M. Solomon. "Dynamic Graph CNN for Learning on Point Clouds." ACM Transactions on Graphics (ToG) 2019, vol 38, no. 5: 1-12.

[4] Thomas, Hugues, Charles R. Qi, Jean-Emmanuel Deschaud, Beatriz Marcotegui, François Goulette, and Leonidas J. Guibas. "Kpconv: Flexible and Deformable Convolution for Point Clouds." In Proceedings of the IEEE/CVF International Conference on Computer Vision 2019: 6411-6420.

[5] Lai, Xin, Jianhui Liu, Li Jiang, Liwei Wang, Hengshuang Zhao, Shu Liu, Xiaojuan Qi, and Jiaya Jia. "Stratified Transformer for 3d Point Cloud Segmentation." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2022: 8500-8509.

[6] Yu, Xumin, Lulu Tang, Yongming Rao, Tiejun Huang, Jie Zhou, and Jiwen Lu. "Point-bert: Pre-Training 3d Point Cloud Transformers with Masked Point Modeling." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2022: 19313-19322.

[7] Hu, Qingyong, Bo Yang, Linhai Xie, Stefano Rosa, Yulan Guo, Zhihua Wang, Niki Trigoni, and Andrew Markham. "Randla-net: Efficient Semantic Segmentation of Large-Scale Point Clouds." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2020: 11108-11117.

[8] Matrone, Francesca, Eleonora Grilli, Massimo Martini, Marina Paolanti, Roberto Pierdicca, and Fabio Remondino. "Comparing Machine and Deep Learning Methods for Large 3D Heritage Semantic Segmentation." ISPRS International Journal of Geo-Information 2020, vol 9, no. 9: 535.

[9] Pellis, Eugenio, Andrea Masiero, Michele Betti, Grazia Tucci, and Pierre Grussenmeyer. "A Deep Learning Multiview Approach for the Semantic Segmentation of Heritage Building Point Clouds." International Journal of Architectural Heritage 2025, vol 19, no. 12: 3117-3139.

[10] Lu, Dening, Jun Zhou, Kyle Yilin Gao, Jing Du, Linlin Xu, and Jonathan Li. "Dynamic Clustering Transformer Network for Point Cloud Segmentation." International Journal of Applied Earth Observation and Geoinformation 2024, vol 128: 103791.

[11] G. Pellis, A. Masiero, M. Betti, G. Tucci, and P. Grussenmeyer, Images & PointClouds Cultural Heritage Dataset, Zenodo, 2025. Available - https://zenodo.org/records/14626611