MULTI-LABEL OBJECT CLASSIFICATION IN POINT CLOUDS USING GRAPH CONVOLUTIONAL NETWORK

Md. Nahid Hasan[1] *Md Sohag Mia*[2] *Nafis Sadeq*[3] *Muhammad Abdullah Adnan*[4]

[1, 4]Bangladesh University of Engineering and Technology, Dhaka - 1000, Bangladesh
[2]Nanjing University of Information Science and Technology, Nanjing, China
[3]University of California, San Diego, United States

2026 IEEE International Conference on Image Processing (ICIP) | 979-8-3315-5151-3/26/$31.00 ©2026 IEEE | DOI: 10.1109/ICIP61757.2026.11630534

ABSTRACT

Point clouds are a fundamental representation for 3D perception in robotics, autonomous systems, and augmented reality; however, despite progress in single-label classification, multi-label object recognition in point clouds—where multiple object categories inherently co-exist—remains underexplored. This paper proposes *Multi-ContexPoint-GCN*, a context-aware framework that formulates point cloud classification as a multi-label problem by employing Binary Cross-Entropy loss with sigmoid activation and integrating a Graph Convolutional Network (GCN) branch — chosen for its ability to exploit sparse, structured label co-occurrence through localized message passing over a data-driven label graph — to model inter-label dependencies arising from the physical layout constraints of indoor 3D scenes. Extensive experiments on the S3DIS and ScanNet datasets demonstrate the effectiveness and generalizability of the proposed approach across different backbone architectures. On S3DIS, the GCN branch improves mean Average Precision (mAP) by 3.18% with PointNet and by 1.39% with PointNet++, while on ScanNet, gains of 2.05% and 1.33% mAP are achieved using PointNet and PointNet++ backbones, respectively. These results confirm that explicitly modeling label correlations consistently enhances multi-label point cloud recognition, establishing a robust paradigm for context-aware 3D scene understanding.

***Index Terms*—** Graph Convolutional Network, Multi-Label Object Classification, Point Cloud, Point Cloud Classification

1 Introduction

Point clouds are a fundamental representation for three-dimensional (3D) scene understanding, with applications in robotics, autonomous systems, augmented reality, and architectural modeling. Acquired via LiDAR or photogrammetry, point clouds provide rich geometric and semantic information but pose challenges due to their irregular and unordered nature [1,2]. Accurate interpretation of point clouds is therefore essential for enabling machines to perceive complex 3D environments.

Recent advances in deep learning have driven significant progress in 3D object classification, semantic segmentation, and detection. Point-based architectures such as PointNet and PointNet++ learn directly from raw point sets, while Graph Convolutional Networks (GCNs) capture contextual relationships by modeling spatial or semantic dependencies through graph structures [3,4]. Despite these advances, most existing methods formulate point cloud understanding as a *single-label* classification problem, assigning each point or instance to one category (e.g., "wall" or "chair") [5]. In real-world indoor scenes, however, objects and regions often exhibit *multiple co-existing semantic labels*, such as "wall" and "painted" or "chair" and "wooden." While multi-label learning has been explored in point cloud *segmentation* and *detection*—where labels are predicted per point or bounding box—multi-label point cloud classification remains largely unexplored. Existing approaches primarily target dense prediction and do not address unified classification with multiple semantic labels, leaving a clear gap between real-world complexity and current classification paradigms. Moreover, semantic labels in 3D scenes are strongly interdependent, with frequent co-occurrence patterns (e.g., "table" and "chair," "ceiling" and "wall"). Although GCNs have been used to model label correlations [4, 6, 7], such efforts are mostly limited to image-based or single-label settings and are not directly applicable to multi-label point cloud classification. GCNs are well-suited for this role: label co-occurrence in 3D scenes is sparse and physically constrained, forming a natural graph structure for efficient message passing. Unlike recurrent chains or dense attention, GCNs use a data-driven adjacency matrix and capture higher-order dependencies (e.g., 'table'→'chair'→'floor') without quadratic complexity. [8,9].

To address this gap, we propose *Multi-ContexPoint-GCN*, a context-aware framework for *multi-label point cloud classification*. The framework integrates point-based backbones (PointNet and PointNet++) with a Graph Convolutional Network (GCN) branch to model semantic label dependencies. Multi-label prediction is enabled via sigmoid activation and Binary Cross-Entropy (BCE) loss, allowing independent

Authorized licensed use limited to: IEEE Xplore. Downloaded on September 03,2026 at 17:12:04 UTC from IEEE Xplore. Restrictions apply.

probability estimation per label. We evaluate our approach on the S3DIS [5] and ScanNet [10] benchmarks and demonstrate consistent improvements across datasets and backbone architectures, highlighting the robustness and generalizability of structured label-dependency modeling.

In summary, this paper makes the following contributions:

- We introduce **Multi-ContexPoint-GCN**, a unified framework that extends point cloud classification to the multi-label setting.
- We develop a **GCN-based label correlation module** that captures semantic co-occurrence relationships and enhances multi-label reasoning.
- We validate the effectiveness and generalizability of the proposed approach on **S3DIS and ScanNet** using **PointNet and PointNet++** backbones, achieving consistent mAP improvements of up to **3.18%**.

2 Literature Review

The field of 3D point cloud analysis has advanced rapidly with deep learning. Key breakthroughs include PointNet [11], which first enabled direct, permutation-invariant neural processing of point clouds, followed by PointNet++ [12] that introduced hierarchical local neighborhood aggregation. Subsequent point-based methods proposed learned convolutions in PointCNN [13], dynamic graphs in Dynamic Graph CNN [3], and recent progress in RandLA-Net [8] and Point Transformer [14]. Graph-based approaches model point clouds as graphs with points as nodes and spatial relations as edges. Graph Convolutional Networks (GCNs) [15] aggregate neighbor information via message passing. Early work focused on RGB-D segmentation [16], while later advances in attention-based graph convolution [7] and kernel-based KPConv [17] demonstrated greater flexibility. Nevertheless, most graph-based methods remain limited to single-label prediction.

Multi-label classification has grown in importance for objects with multiple co-occurring attributes. Dependency modeling now dominates over independent classifiers. Notable examples include the GCN approach of Chen et al. [4], classifier chains [18], large-scale benchmarks [19], adaptive medical models [20], and Word2Vec-based metadata methods [21]. However, multi-label point cloud classification remains underexplored. Major benchmarks (Semantic3D, ScanNet, ModelNet40) focus on single-label tasks or per-point segmentation. Even multi-view methods [22, 23] have not addressed the multi-label setting. Thus, a significant gap exists between real-world semantic complexity and current point cloud classification paradigms. Graph-based representations have been introduced to aggregate multi-channel features by modeling spectral correlations as graph structures, enabling GCNs [24] to propagate contextual information across modalities iteratively.

In conclusion, although point-based and graph-based methods excel in single-label tasks and GCNs effectively capture spatial relationships, multi-label point cloud classification is a key research gap. The present work fills this gap with a GCN-based framework that explicitly models label correlations.

3 Proposed Method

Our approach consists of two main components. The first component is a point cloud backbone that extracts and encodes geometric features from the input. The second component integrates a GCN to capture label correlations, thereby guiding the backbone toward improved accuracy. The overall architecture is shown in Figure 1.

3.1 Multi-ContextPoint-GCN

In multi-object datasets, label correlations naturally arise, as observed in large-scale indoor benchmarks such as S3DIS [5] and ScanNet [10], which contain 13 and 18 categories, respectively. For example, the presence of a table often implies nearby chairs, while walls commonly co-occur with ceilings and floors, reflecting practical spatial dependencies in indoor environments. To capture these relationships, we propose a dual-pipeline *Multi-ContexPoint-GCN*: (1) a primary pipeline that performs multi-label prediction directly from point cloud geometry, and (2) a secondary GCN-based pipeline that models semantic label correlations to refine the predictions. By integrating geometric feature learning with explicit modeling of label dependencies, the proposed design effectively leverages contextual information and leads to consistent performance improvements across datasets.

PointNet Path. PointNet and its hierarchical extension, PointNet++ [11, 12], were originally designed for single-label point cloud classification and segmentation tasks. While PointNet learns global features directly from unordered point sets, PointNet++ extends this capability by capturing local geometric structures through hierarchical sampling and neighborhood aggregation, making it more effective for complex indoor scenes. However, both architectures require adaptation to support multi-label classification.

Inspired by multi-label learning in image analysis [4, 7] and tabular data [25], where multiple labels naturally co-exist, we modify both backbones to enable multi-label prediction. Specifically, we: (1) replace the Negative Log-Likelihood (NLL) loss with Binary Cross-Entropy (BCE) loss (Equation 1) to independently estimate label probabilities; (2) substitute the Log Softmax layer with Sigmoid activation to allow simultaneous prediction of multiple labels; and (3) adopt mean Average Precision (mAP) instead of accuracy as the primary evaluation metric to balance precision and recall in multi-label settings better. These adaptations enable both PointNet- and PointNet++-based pipelines to model complex scenes where multiple semantic categories co-occur effectively.

$$-\frac{1}{N}\sum_{i=1}^{N}\{y_i \log(p_i) + (1-y_i)\log(1-p_i)\} \qquad (1)$$

Authorized licensed use limited to: IEEE Xplore. Downloaded on September 03,2026 at 17:12:04 UTC from IEEE Xplore. Restrictions apply.

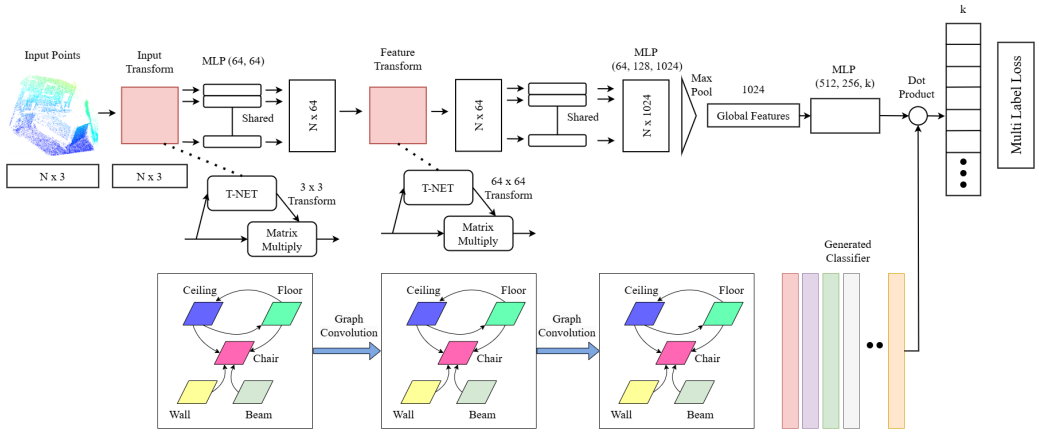

Fig. 1: Multi-ContexPoint-GCN architecture - PointNet: (i) The Upper part of the Multi-ContexPoint-GCN architecture is the modified PointNet. (ii) The lower part is the GCN module producing the generated classifier. (iii) The last result is produced by merging both outputs using the dot product.

Graph Convolutional Path. The GCN path models and propagates label correlations across classes. A GCN is chosen over alternatives such as classifier chains or dense attention because label dependencies in 3D scenes are sparse and physically structured, reflecting room layout constraints, making localized graph message passing both principled and efficient. We construct a label graph where nodes are semantic classes, initialized with 300-dimensional GloVe-6B embeddings [26], providing semantic priors that capture functional relationships between object categories (e.g., 'sofa' and 'bookcase' as co-occurring furniture), complementing the geometric features from the point cloud backbone. Instead of relying on a fixed adjacency, a learned correlation matrix encodes probabilistic label relationships. Stacked GCN layers refine these representations, capturing higher-order co-occurrences.

Correlation Matrix. A co-occurrence matrix M records label pair frequencies in the S3DIS dataset. Each entry M_{ij} represents how often labels L_i and L_j co-occur. From this, a conditional probability matrix P is derived, defined as $P_i = \frac{M_i}{N_i}$, where $P_{ij} = P(L_j|L_i)$. Because rare labels introduce noise, weak connections are filtered by binarizing P with a threshold τ (Equation 2), yielding the final adjacency matrix Ad.

$$Ad_{ij} = \begin{cases} 0, & \text{if } p_{ij} < \tau, \\ 1, & \text{if } p_{ij} \geq \tau. \end{cases} \tag{2}$$

$$\grave{Ad}_{ij} = \begin{cases} \dfrac{p}{\sum_{j=1}^{n} Ad_{ij}}, & i \neq j, \\ 1-p, & i = j. \end{cases} \tag{3}$$

Here, $\grave{Ad}_{ij}$ is the reweighted correlation matrix. $p = 1$ ignores self-features, while $p = 0$ ignores neighbors. Intermediate values preserve both contributions, improving feature propagation.

Over-Smoothing Problem. GCNs aggregate neighborhood features with those of a node itself, which can cause over-smoothing, making distinct clusters indistinguishable. To mitigate this, we apply a reweighted scheme [4], where p controls the balance between self-features and neighbor features (Equation 3).

3.2 Graph Attention Transformer Network

In [7], a graph attention transformer layer was proposed to enhance node representations by dynamically adapting the network structure. Unlike methods that rely on fixed adjacency matrices, this approach extends the correlation matrix A to uncover subgraph patterns and multi-hop relationships. Each subgraph G encodes transformed structural information derived from A. Inspired by vision transformers, multi-head self-attention is applied to identify subgraphs. For each head i, query, key, and value matrices (Q_i, K_i, V_i) are obtained via linear transformations of A (Equation 4a). Scaled dot-product attention estimates node relevance, producing outputs that are concatenated and projected to form the subgraph G_j (Equation 4b). Each head uses non-shared parameters for diversity. Multiple subgraphs $G_1, G_2, \ldots, G_k$ are then integrated into the final adjacency matrix $\acute{A}$ through element-wise multiplication, $\acute{A} = \prod_{j=1}^{k} G_j$. In our framework, we adopt this attention mechanism, passing the resulting $\acute{A}$ to the GCN module for label correlation modeling.

$$[Q_i, K_i, V_i] = A[W_i^Q, W_i^K, W_i^V] \tag{4a}$$

$$G_j = Conc.(Atten.(Q_1, K_1, V_1), ...Atten.(Q_n, K_n, V_n))W \tag{4b}$$

4 Experiments

4.1 Dataset

Experiments were conducted on two large-scale indoor point cloud benchmarks: S3DIS and ScanNet. S3DIS con-

Authorized licensed use limited to: IEEE Xplore. Downloaded on September 03,2026 at 17:12:04 UTC from IEEE Xplore. Restrictions apply.

Table 1: Comparison of different models on the S3DIS dataset.

Method	ceiling	floor	wall	beam	column	window	door	table	chair	sofa	bookcase	board	clutter	mAP
PointNet [11]	99.72	99.93	95.65	00.36	31.54	63.12	57.58	89.81	73.45	**13.16**	76.04	15.94	80.20	61.27
Multi-ContexPoint-GCN (Binary) - PointNet	99.30	98.36	89.89	00.27	25.38	61.63	57.86	89.20	78.21	5.79	70.93	16.14	63.70	58.21
PointNet Attention [7]	**99.89**	**99.97**	95.36	00.38	26.60	63.29	61.11	90.26	69.22	11.85	74.70	20.51	77.86	60.85
PointNet++ [12]	99.95	99.99	**98.56**	00.58	42.40	72.63	85.85	**95.05**	**92.00**	50.92	**90.77**	49.85	88.91	74.42
Multi-ContexPoint-GCN (Re-Weighted) - PointNet	99.86	99.95	**97.18**	**00.47**	**34.39**	**67.37**	**71.49**	**92.18**	**82.08**	6.23	**81.12**	**20.73**	**84.78**	**64.45**
Multi-ContexPoint-GCN (Re-Weighted) - PointNet++	**99.98**	**99.99**	98.19	**00.72**	**45.35**	**76.02**	**85.90**	94.97	91.24	**58.57**	89.83	**54.08**	**90.71**	**75.81**

Table 2: Comparison of different models on the Scannet dataset.

Method	Bath	Bed	BkShf	Cab	Chair	Cntr	Curt	Desk	Door	GrbgBn	Pic	frig	Shrowr	Sink	Sofa	Tabl	Toil	Wind	mAP
PointNet [11]	37.71	22.42	**18.81**	**53.31**	**85.55**	17.14	15.89	24.45	72.45	79.53	37.04	21.14	**35.02**	48.22	**29.33**	83.30	46.45	54.30	43.45
Multi-ContexPoint-GCN (Re-Weighted) - PointNet	**40.23**	**29.78**	15.59	53.23	83.74	**18.42**	**17.65**	**40.01**	**77.93**	**83.47**	**37.58**	**22.82**	22.38	**48.82**	29.20	**85.85**	**47.97**	**64.33**	**45.50**
PointNet++ [12]	**56.35**	**66.89**	20.32	57.22	91.29	34.08	**41.17**	63.80	85.96	87.37	42.85	42.34	**37.82**	64.34	51.55	89.10	72.78	**68.77**	59.66
Multi-ContexPoint-GCN (Re-Weighted) - PointNet++	46.71	61.70	**21.38**	**64.74**	**91.95**	**40.36**	41.14	**68.62**	**88.91**	**87.89**	**44.59**	**54.12**	32.91	**67.29**	**53.95**	**90.55**	**73.63**	67.44	**60.99**

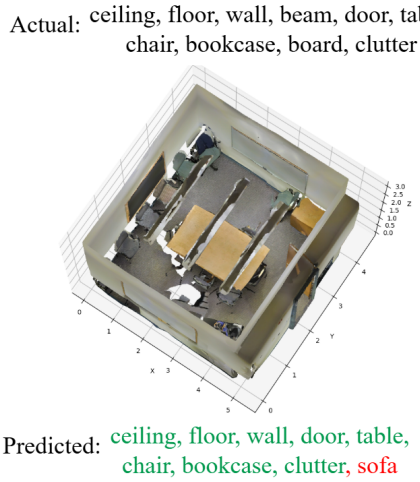

Fig. 2: Multi-ContexPoint-GCN Model Prediction of a 3D point cloud scene.

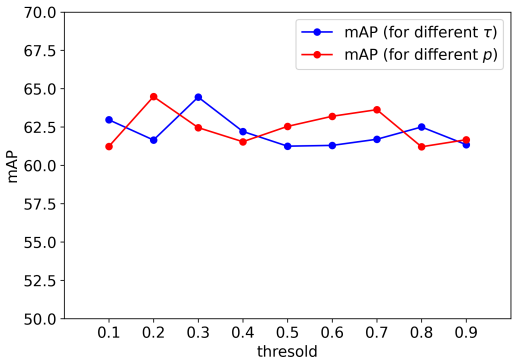

Fig. 3: Impact of varying τ and p thresholds on mAP. (S3DIS Dataset)

tains point clouds from six indoor areas (offices, conference rooms, hallways, etc.) with 3D coordinates, RGB values, and 13 semantic labels, including ceiling, floor, wall, beam, column, window, door, table, chair, sofa, bookcase, board, and clutter. Area 5 was used for testing, and the rest for training. The dataset is challenging due to occlusions, varying point densities, and complex layouts. ScanNet provides 3D reconstructions from diverse real-world indoor environments with 3D coordinates, color, and 18 semantic categories, featuring sensor noise and incomplete geometry, making it suitable for evaluating the robustness and generalizability of multi-label classification across datasets with varying scene structures and label distributions.

4.2 Experimental Setup

The multi-label setting is handled using sigmoid activation with BCE loss in the point cloud backbone. Each input has 50k sampled points (x, y, z). To improve robustness, Gaussian noise (mean 0, std 0.02) and random vertical rotations are applied. The GCN pathway has two layers: the first maps 300-dimensional GloVe embeddings [26] to 1,024 features, and the second reduces them to 128. Label correlations are encoded via a learned matrix, binarized at $\tau = 0.3$, with aggregation weight $p = 0.2$. LeakyReLU (slope 0.2) is used throughout.

Optimization uses SGD with learning rate 0.01, momentum 0.9, and a scheduler reducing the rate by 0.7 every 20 epochs. A matrix difference loss, scaled by 0.001, regularizes the learned feature transformation matrix to encourage near-orthogonality, improving stability and generalization. Models are trained for 150 epochs on S3DIS. The same configuration is used for ScanNet, but training is limited to 50 epochs, as longer training did not improve performance.

Authorized licensed use limited to: IEEE Xplore. Downloaded on September 03,2026 at 17:12:04 UTC from IEEE Xplore. Restrictions apply.

4.3 Results

We evaluate our model on the S3DIS dataset using the standard split, with Areas 1–4 and 6 for training (46,623 samples) and Area 5 for testing (18,923 samples). As reported in Table 1, the PointNet baseline achieves 61.27% mAP, performing strongly on dominant planar categories such as ceiling (99.72%) and floor (99.93%), but struggling on geometrically complex or underrepresented classes, including beam (0.36%) and board (15.94%). The Multi-ContexPoint-GCN (Binary)–PointNet variant reduces performance to 58.21% mAP, with notable drops for wall (95.65% $\rightarrow$ 89.89%) and clutter (80.20% $\rightarrow$ 63.70%). This indicates that hard binarization of label correlations discards useful probabilistic information, an effect amplified by class imbalance where frequent categories such as table and chair dominate rare ones like sofa and beam.

Our proposed Multi-ContexPoint-GCN (Re-Weighted) yields the best PointNet-based performance at 64.45% mAP (+3.18). The largest gains are observed for door (57.58% $\rightarrow$ 71.49%), chair (73.45% $\rightarrow$ 82.08%), and bookcase (76.04% $\rightarrow$ 81.12%), highlighting the benefit of re-weighted label dependency modeling, particularly for functionally related categories. Using PointNet++ as a stronger backbone, the baseline reaches 74.42% mAP, with high accuracy on ceiling (99.95%), floor (99.99%), and wall (98.56%), but weaker results for beam (0.58%) and column (42.40%). Integrating the proposed framework further improves performance to 75.81% mAP (+1.39), with consistent gains for sofa (50.92% $\rightarrow$ 58.57%), board (49.85% $\rightarrow$ 54.08%), and clutter (88.91% $\rightarrow$ 90.71%), demonstrating that explicit label correlation modeling remains effective even with hierarchical geometric features.

We additionally evaluate on the ScanNet dataset (Table 2). With PointNet, the baseline achieves 43.45% mAP, performing well on chair (85.55%), table (83.30%), and door (72.45%), but poorly on smaller or less frequent objects such as bookshelf (18.81%) and curtain (15.89%). The proposed re-weighted model improves performance to 45.50% mAP, with notable gains for desk (24.45% $\rightarrow$ 40.01%), door (72.45% $\rightarrow$ 77.93%), and window (54.30% $\rightarrow$ 64.33%). With PointNet++, ScanNet performance increases from 59.66% to 60.99% mAP (+1.33) when incorporating the proposed framework, yielding improvements for cabinet (57.22% $\rightarrow$ 64.74%), desk (63.80% $\rightarrow$ 68.62%), and refrigerator (42.85% $\rightarrow$ 44.59%). Across both datasets, slender or texture-poor categories such as beams and sofas remain challenging, reflecting inherent geometric ambiguity and data sparsity.

5 Ablation Study

Adding GCN slightly increases model size but minimally affects computation: PointNet++ goes from 1.74 M params / 8.14 G FLOPs to 2.21 M / 8.15 G, and PointNet from 3.47 M / 3.58 G FLOPs to 3.94 M / 3.60 G. This shows GCN adds little overhead while improving feature learning.

Table 3: Multi-ContexPoint-GCN performance with varying GCN layers (S3DIS Dataset).

#Layers	Dimensions	mAP
1	128	62.25
2	1024, 128	64.45
3	1024, 1024, 128	62.5

Effects of τ and p. The threshold τ determines which label correlations influence predictions. Sensitivity analysis over $\tau \in [0.1, 0.9]$ (Figure 3) shows peak performance at $\tau = 0.3$ (64.45% mAP), balancing signal and noise. $\tau > 0.5$ removes meaningful relationships, dropping mAP to 61.2%, while $\tau < 0.2$ introduces noise, reducing mAP to 62.8%. The critical region is $\tau = 0.2 - 0.4$, where 0.1 changes yield >1.5% mAP variation. At $\tau = 0.3$, the label graph has an average degree of 4.2, preserving small-world properties for effective message passing. The re-weighting p balances node features with neighborhood context (Equation 3); too small ($p = 0.1$) reduces correlation usage, while too large ($p > 0.5$) oversmooths features. Optimal $p = 0.2$ preserves 80% of original features while capturing context for medium-frequency objects (Figure 3).

Optimizing GCN Layer Depth. Each GCN layer aggregates neighbor information. Deeper networks capture higher-order neighbors but may oversmooth. Ablation (Table 3) shows two layers (1024 $\rightarrow$ 128) achieve the best balance between context aggregation and discriminative representation, outperforming deeper setups.

6 Conclusion

Understanding three-dimensional (3D) scenes is fundamental to computer vision, with applications in robotics, augmented reality, and autonomous systems. Although significant progress has been made in single-label classification and semantic segmentation on benchmarks such as S3DIS and ScanNet, multi-label point cloud recognition remains underexplored. We address this gap with Multi-ContextPoint-GCN, which integrates a GCN-based label correlation module into point cloud backbones. On S3DIS, the proposed framework improves PointNet performance by 3.18% mAP, with pronounced gains for co-occurring object categories. Consistent improvements are also observed on ScanNet and with stronger backbones such as PointNet++, demonstrating robust generalization. Graph attention mechanisms did not yield additional gains; accordingly, comparisons with Point Transformer-based models are reserved for future work. These results show that modeling semantic dependencies complements geometric feature learning and enables more context-aware multi-label 3D scene understanding. Future work will explore multi-modal fusion and lightweight designs for real-world deployment.

Authorized licensed use limited to: IEEE Xplore. Downloaded on September 03,2026 at 17:12:04 UTC from IEEE Xplore. Restrictions apply.

7 References

[1] Yulan Guo, Hanyun Wang, Qingyong Hu, Hao Liu, Li Liu, and Mohammed Bennamoun, "Deep learning for 3d point clouds: A survey," 2020.

[2] Huang Zhang, Changshuo Wang, Shengwei Tian, Baoli Lu, Liping Zhang, Xin Ning, and Xiao Bai, "Deep learning-based 3d point cloud classification: A systematic survey and outlook," 2023.

[3] Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E. Sarma, Michael M. Bronstein, and Justin M. Solomon, "Dynamic graph cnn for learning on point clouds," 2019.

[4] Zhao-Min Chen, Xiu-Shen Wei, Peng Wang, and Yanwen Guo, "Multi-label image recognition with graph convolutional networks," 2019.

[5] Iro Armeni, Ozan Sener, Amir R. Zamir, Helen Jiang, Ioannis Brilakis, Martin Fischer, and Silvio Savarese., "3d semantic parsing of large-scale indoor spaces," 2016.

[6] Thomas N. Kipf and Max Welling, "Semi-supervised classification with graph convolutional networks," 2017.

[7] Jin Yuan, Shikai Chen, Yao Zhang, Zhongchao Shi, Xin Geng, Jianping Fan, and Yong Rui, "Graph attention transformer network for multi-label image classification," 2024.

[8] Timo Hackel, Nikolay Savinov, Lubor Ladicky, Jan Dirk Wegner, Konrad Schindler, and Marc Pollefeys, "Semantic3d.net: A new large-scale point cloud classification benchmark," 2017.

[9] Lin Q Pan X and and Ye S, "Deep learning based approaches from semantic point clouds to semantic bim models for heritage digital twin," 2024.

[10] Dai A, Chang A, Savva M, Halber M, Funkhouser T, and Nießner M, "Scannet: Richly-annotated 3d reconstructions of indoor scenes," 2017.

[11] Charles R. Qi, Hao Su, Kaichun Mo, and Leonidas J. Guibas, "Pointnet: Deep learning on point sets for 3d classification and segmentation," 2017.

[12] Charles R. Qi, Li Yi, Hao Su, and Leonidas J. Guibas, "Pointnet++: Deep hierarchical feature learning on point sets in a metric space," 2017.

[13] Yangyan Li, Rui Bu, Mingchao Sun, Wei Wu, Xinhan Di, and Baoquan Chen, "Pointcnn: convolution on x-transformed points," in *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, Red Hook, NY, USA, 2018, NIPS'18, p. 828–838, Curran Associates Inc.

[14] Hengshuang Zhao, Li Jiang, Jiaya Jia, Philip Torr, and Vladlen Koltun, "Point transformer," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021, pp. 16239–16248.

[15] Yoonhyuk Choi, Jiho Choi, Taewook Ko, and Chong-Kwon Kim, "Mitigating overfitting in graph neural networks via feature and hyperplane perturbation," in *Proceedings of the Eighteenth ACM International Conference on Web Search and Data Mining*, New York, NY, USA, 2025, WSDM '25, p. 50–59, Association for Computing Machinery.

[16] Qi C, R Su, H Niessner, M Dai, A Yan, and M Guibas, "Volumetric and multi-view cnns for object classification on 3d data," 2017.

[17] H. Thomas, C. R. Qi, J.-E. Deschaud, B. Marcotegui, F. Goulette, and L. J. Guibas, "Kpconv: Flexible and deformable convolution for point clouds," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, pp. 6411–6420.

[18] Read J, Pfahringer B, Holmes G, and Frank E, "Classifier chains for multi-label classification," 2011.

[19] J. M. Moyano, E. L. Gibaja, and S. Ventura, "Comprehensive comparative study of multi-label classification methods," *Expert Systems with Applications*, vol. 203, pp. 117215, 2022.

[20] M. M. Islam, A. Alqahtani, R. U. Khan, N. R. Aljohani, A. Alharbi, and M. Alshahrani, "Hybrid multi-label classification model for medical applications based on adaptive synthetic data and ensemble learning," *Applied Sciences*, vol. 13, no. 17, pp. 9904, 2023.

[21] Y. Shen, Y. Rehman, M. S. Khan, F. Khan, and S. Rahman, "Multi-label classification of research articles using word2vec and identification of similarity threshold," *Scientific Reports*, vol. 11, pp. 21409, 2021.

[22] H. You, Y. Feng, R. Ji, and Y. Zhao, "Multi-view features joint learning with label and local distribution consistency for point cloud classification," *Remote Sensing*, vol. 12, no. 1, pp. 135, 2019.

[23] Y. Lin, G. Vosselman, Y. Cao, and M. Y Yang, "Semantic segmentation of mobile mapping point clouds via multi-view label transfer," *ISPRS Journal of Photogrammetry and Remote Sensing*, pp. 45–61, 2023.

[24] Junting Deng, Ethan Chen, and Vanessa Chen, "Graph convolutional network aggregation for broad-spectral object detection," in *2025 IEEE International Conference on Image Processing (ICIP)*, 2025, pp. 2288–2293.

[25] Sajjad Kamali Siahroudi, Zahra Ahmadi, and Daniel Kudenko, "Effectively capturing label correlation for tabular multi-label classification," in *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, New York, NY, USA, 2024, CIKM '24, p. 1060–1069, Association for Computing Machinery.

[26] Jeffrey Pennington, Richard Socher, and Christopher Manning, "GloVe: Global vectors for word representation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Alessandro Moschitti, Bo Pang, and Walter Daelemans, Eds., Doha, Qatar, Oct. 2014, pp. 1532–1543, Association for Computational Linguistics.

Authorized licensed use limited to: IEEE Xplore. Downloaded on September 03,2026 at 17:12:04 UTC from IEEE Xplore. Restrictions apply.