

Point-GN: A Non-Parametric Network Using Gaussian Positional Encoding for Point Cloud Classification

Marzieh Mohammadi* Amir Salarpour*
 Sirjan University of Technology, Iran

mrziehmohamadi@gmail.com, salarpour@sirjantech.ac.ir

Abstract

This paper introduces **Point-GN**, a novel non-parametric network for efficient and accurate 3D point cloud classification. Unlike conventional deep learning models that rely on a large number of trainable parameters, **Point-GN** leverages non-learnable components—specifically, **Farthest Point Sampling (FPS)**, **k-Nearest Neighbors (k-NN)**, and **Gaussian Positional Encoding (GPE)**—to extract both local and global geometric features. This design eliminates the need for additional training while maintaining high performance, making **Point-GN** particularly suited for real-time, resource-constrained applications. We evaluate **Point-GN** on two benchmark datasets, **ModelNet40** and **ScanObjectNN**, achieving classification accuracies of **85.29%** and **85.89%**, respectively, while significantly reducing computational complexity. **Point-GN** outperforms existing non-parametric methods and matches the performance of fully trained models, all with zero learnable parameters. Our results demonstrate that **Point-GN** is a promising solution for 3D point cloud classification in practical, real-time environments. For more details, see the code at: https://github.com/asalarpour/Point_GN.

1. Introduction

Point cloud classification is a critical task in 3D data analysis and has been widely employed in various fields, including object detection [45, 47], 3D reconstruction [36], robotics [29], and medicine [5, 39]. Unlike 2D images that are structured in regular grids, point clouds consist of unordered and irregular sets of points, presenting unique challenges for efficient and accurate analysis. The unordered nature and high dimensionality of point clouds make traditional 2D image processing techniques unsuitable for this task, thereby necessitating specialized algorithms that can handle the unique structure of 3D data.

Deep learning has significantly advanced point cloud

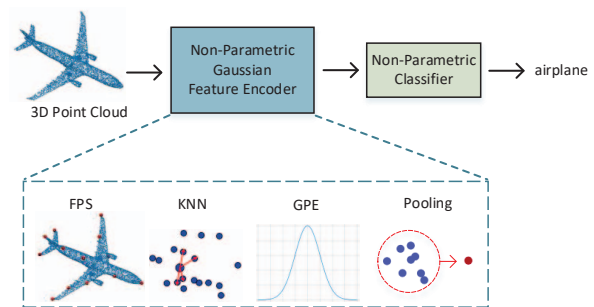


Figure 1. Illustration of the proposed non-parametric network for point cloud classification.

classification, with models like PointNet [20] and PointNet++ [21] enabling direct processing of point clouds without requiring 3D voxelization or mesh generation. By treating point clouds as unordered sets of points, these models learn permutation-invariant features. However, their reliance on large numbers of parameters leads to higher memory usage and longer training times. For example, PointNet++ extends PointNet with hierarchical structures and local feature capture but also increases computational complexity, which can hinder scalability and real-time performance.

The trend toward more complex models further compounds the issue. For instance, CurveNet [15], which focuses on curve-based feature extraction, increases training times by up to 10 times compared to PointNet++, with little improvement in performance. Similarly, PointMLP [11] adds 11.9 million parameters over PointNet++, achieving only a modest 0.5% gain in precision. This growing complexity underscores the need for more efficient models that balance accuracy and computational cost, especially for real-time applications like autonomous driving [19], AR/VR, and medical diagnostics.

To address the high computational and memory demands

*These authors contributed equally to this work.

of parametric models, recent research has explored non-parametric methods. These approaches, such as PointClip [42], which uses pre-trained 2D models for point cloud classification, and Point-NN [43], which avoids learnable weights, aim to reduce parameter sets and improve efficiency. However, these methods often trade off computational efficiency for classification accuracy, highlighting the need for further improvement.

In this paper, we propose a novel non-parametric method for point cloud classification that introduces a Gaussian function for positional embedding, as shown in Fig. 1. This Gaussian embedding enhances classification accuracy by capturing local geometric structures without requiring extra parameters or retraining. Integrated into the non-parametric framework of PointNet++ with Farthest Point Sampling (FPS) and k-Nearest Neighbors (k-NN), it improves performance while maintaining computational efficiency.

Our contributions are threefold:

- We introduce a **Gaussian embedding function** to the non-parametric framework, which significantly improves classification accuracy without adding computational overhead.
- We **simplify the network design**, eliminating unnecessary complexities while maintaining performance comparable to state-of-the-art parametric models, offering a more efficient alternative for resource-constrained environments.
- We demonstrate the **efficiency and scalability** of our approach through extensive experiments on popular benchmarks, including ModelNet40 [34] and ScanObjectNN [28], achieving competitive performance without additional costs in terms of memory or computational requirements.

This method offers a practical solution for efficient, accurate point cloud classification, ideal for real-time applications with limited resources, such as robotics and autonomous systems [1, 14], where both accuracy and efficiency are essential.

2. Related works

In this section, we provide an overview of prior approaches to 3D point cloud classification, focusing on both projection-based and point-based methods, and introduce recent advancements in positional encoding techniques, which have significantly impacted model performance and efficiency.

2.1. 3D Point Cloud Classification

3D point cloud classification methods can be divided into projection-based and point-based approaches.

Projection-based methods convert 3D point clouds into 2D representations, such as depth maps [3, 4, 26] or voxel grids [9, 12, 31], enabling the use of 2D image processing techniques. However, they often lose spatial details due to the sparsity and incompleteness of point clouds.

Point-based methods process raw 3D point clouds directly, preserving geometric information. PointNet [20] processes points independently and aggregates global features using max pooling but struggles with local geometric details. PointNet++ [21] improves this by introducing a hierarchical architecture for capturing local features. Other advancements include convolutional [7, 18, 24, 27] and graph-based models [10, 32], as well as attention and transformer-based methods [37, 40, 46] that model long-range dependencies.

However, these models are computationally intensive, limiting their use in real-time applications. Efficient methods like ShellNet [44] and RandLA-Net [6] reduce memory usage, but still face challenges with large-scale data. Convolutional models like KPConv [27] reduce memory overhead through sparse convolutions but still require significant resources.

Inspired by Point-NN [43], we propose a novel non-parametric model for point cloud classification that improves feature extraction without introducing additional trainable parameters, thus enhancing both efficiency and scalability.

2.2. Positional Encodings

Positional encoding was first introduced in the Transformer architecture [30] to inject positional information into input sequences, such as words in a sentence, using sinusoidal functions. This method has since been widely adopted in natural language processing and computer vision, where capturing spatial relationships is critical. In point cloud processing, positional encoding allows the model to retain spatial awareness in 3D space, particularly for applications involving coordinate-based models, such as 2D image synthesis [16] or 3D scene reconstruction [17].

In 3D point cloud processing, positional encoding plays a crucial role in capturing the underlying geometric relationships between points. One notable application is in Neural Radiance Fields (NeRF) [13], where sinusoidal encoding transforms input coordinates into higher-dimensional feature spaces, enabling the accurate reconstruction of fine-grained details in 3D scenes. Such approaches demonstrate the importance of encoding spatial information when dealing with high-frequency signals, as it accelerates network convergence and enhances the model's ability to capture complex geometric structures.

Building on these advancements, we propose a novel positional encoding scheme based on Gaussian functions,

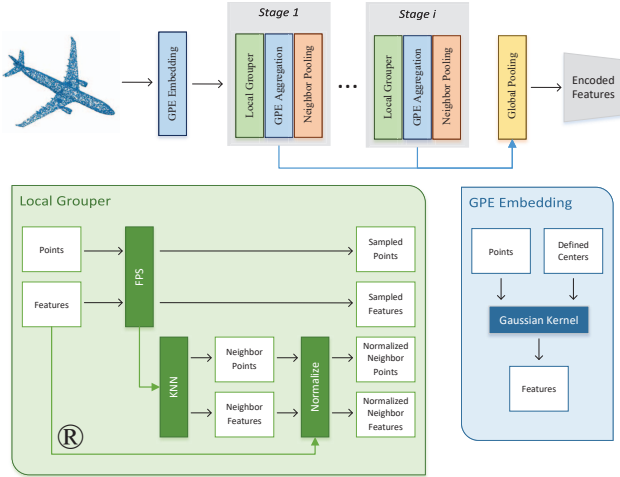


Figure 2. Overview of the architecture of our Non-Parametric Feature Encoder. The figure illustrates the main components of Point-GN, including Gaussian Positional Encoding (GPE), local grouping and feature aggregation. Each stage of the network is designed to efficiently capture spatial relationships within the point cloud without the need for learnable parameters.

specifically tailored for 3D point cloud classification. Unlike sinusoidal encodings, Gaussian embeddings allow for more flexible representation of spatial relationships in 3D space, offering improved feature extraction and classification performance. Our approach demonstrates that Gaussian embeddings can match, and in some cases surpass, the performance of traditional sinusoidal encodings, particularly when applied to point clouds in non-parametric frameworks.

3. The Proposed Method

In this section, we present the details of **Point-GN**, our Non-Parametric Network that utilizes Gaussian Positional Encoding (GPE) for point cloud classification. Fig. 2 provides an overview of the Point-GN framework, which highlights the key stages in processing point cloud data. To ground the discussion, we first revisit the fundamental concepts behind 3D point clouds and how they are typically classified. We then elaborate on the design of our non-parametric feature encoder and classifier, which are integral components of Point-GN.

3.1. Background

A 3D point cloud is a collection of points in 3D space representing the shape or structure of an object or scene. Each point $\mathbf{p}_i = (x_i, y_i, z_i) \in \mathbb{R}^3$ is defined by its coordinates and may have additional attributes, such as color or surface normal vectors. Given a point cloud $\mathcal{P} = \{\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_N\}$, where N is the number of points, each point \mathbf{p}_i is represented by its coordinates (x_i, y_i, z_i) . An en-

coder extracts meaningful information from the point cloud into a compact representation:

$$\text{Encoder}(\mathcal{P}) = \mathbf{F} \in \mathbb{R}^d \quad (1)$$

where \mathbf{F} is the feature vector that encodes the essential characteristics of the point cloud into a d -dimensional space.

For classification, the feature vector \mathbf{F} is fed into a classifier that maps the feature vector to C classes, producing a vector of logits $\mathbf{y} \in \mathbb{R}^C$:

$$\text{Classifier}(\mathbf{F}) = \mathbf{y} = (y_1, y_2, \dots, y_C) \quad (2)$$

The predicted class c is determined by selecting the class with the highest score:

$$c = \arg \max_i y_i \quad (3)$$

3.2. Gaussian Positional Encoding (GPE)

Gaussian Positional Encoding (GPE) embeds spatial information into the feature representation of individual points in the 3D point cloud. By transforming raw point coordinates into a higher-dimensional space, GPE provides the model with richer spatial context without introducing learnable parameters. The encoding is formulated as:

$$\gamma_x(x_i, v_j) = \exp\left(-\frac{\|x_i - v_j\|^2}{2\sigma^2}\right) \quad (4)$$

$$\gamma_y(y_i, v_j) = \exp\left(-\frac{\|y_i - v_j\|^2}{2\sigma^2}\right) \quad (5)$$

$$\gamma_z(z_i, v_j) = \exp\left(-\frac{\|z_i - v_j\|^2}{2\sigma^2}\right) \quad (6)$$

where v_j are predefined reference points, and σ is the standard deviation that controls the focus on local vs. global spatial information. A smaller σ captures local detail, while a larger σ captures broader spatial patterns.

The encoded feature vector for each point is:

$$\gamma(\mathbf{p}_i) = [\gamma_x(x_i, v_j), \gamma_y(y_i, v_j), \gamma_z(z_i, v_j)]_{j=1}^V \quad (7)$$

where V is the number of reference points along each axis.

3.3. Non-Parametric Feature Encoder

In our approach, the non-parametric feature encoder leverages Gaussian Positional Encoding (GPE) to capture and aggregate spatial information from 3D point clouds. This hierarchical encoder adapts to various input configurations without relying heavily on learned parameters, making it flexible across different tasks.

3.3.1 GPE Embedding

The embedding process begins by applying GPE to each point \mathbf{p}_i in the point cloud \mathcal{P} . This transformation maps the raw 3D coordinates into a higher-dimensional feature space, enhancing the model’s ability to understand spatial relationships. As a result, each point \mathbf{p}_i is represented by a richer feature vector $\gamma(\mathbf{p}_i)$, capturing its spatial relationships more effectively.

This transformation expands the original 3D coordinates into a $V \times 3$ -dimensional space, where V represents the number of reference points v_j used in the encoding. These reference values are strategically chosen or learned by the model, often distributed uniformly between -1 and 1 . The parameter σ , controlling the width of the Gaussian function, determines how spatial information is captured. A smaller σ focuses on local details, while a larger σ captures more global spatial relationships. This flexibility enables the GPE to balance the capture of both local and global spatial information effectively.

The GPE embedding serves as the foundational step in our non-parametric feature encoder. By transforming raw 3D point clouds into a feature space rich in spatial context, $\gamma(\mathbf{p}_i)$ enhances the model’s ability to recognize and utilize the underlying geometric structures of the data. This process is integral to the effectiveness of the encoder, leading to improved performance in tasks such as object recognition, segmentation, and classification, where a deep understanding of both local and global spatial relationships is essential.

3.3.2 Local Grouper

After GPE embedding, feature extraction proceeds through multiple stages, each involving a local grouper, GPE aggregation, and neighbor pooling. At each stage, the input point cloud from the previous stage is represented as $\{\mathbf{p}_i, \gamma(\mathbf{p}_i)\}_{i=1}^N$, where $\mathbf{p}_i \in \mathbb{R}^3$ denotes the coordinates of point i and $\gamma(\mathbf{p}_i) \in \mathbb{R}^{V \times 3}$ represents the GPE-embedded features of that point.

The process begins with Farthest Point Sampling (FPS) to downsample the number of points from N to $N/2$, selecting a subset of local center points:

$$\{\mathbf{p}_j, \gamma(\mathbf{p}_j)\}_{j=1}^{N/2} = \text{FPS}(\{\mathbf{p}_i, \gamma(\mathbf{p}_i)\}_{i=1}^N) \quad (8)$$

Next, the downsampled point coordinates \mathbf{p}_j and the original point coordinates \mathbf{p}_i are used by the K-Nearest Neighbors (KNN) algorithm to find the K nearest neighbors for each downsampled point \mathbf{p}_j . The indices of these nearest neighbors are used to retrieve the corresponding coordinates and features:

$$\text{id}x_j = \text{KNN}(\mathbf{p}_j, \mathbf{p}_i) \quad (9)$$

The retrieved coordinates and features are:

$$\mathbf{P}_j = \text{retrieve}(\{\mathbf{p}_i\}_{i=1}^N, \text{id}x_j) \in \mathbb{R}^{K \times 3} \quad (10)$$

$$\Gamma_j = \text{retrieve}(\{\gamma(\mathbf{p}_i)\}_{i=1}^N, \text{id}x_j) \in \mathbb{R}^{K \times (V \times 3)} \quad (11)$$

Here, \mathbf{P}_j represents the gathered coordinates, and Γ_j represents the gathered features for the point \mathbf{p}_j . The retrieved coordinates \mathbf{P}_j and features Γ_j are then normalized using the mean and standard deviation of each point’s neighbors. These normalized coordinates and features are then passed to the next stage for further processing.

3.3.3 GPE Aggregation

The features from the Local Grouper are then fed into the GPE Aggregation module. Here, GPE is applied to the retrieved coordinates \mathbf{P}_j to encode spatial information. These encoded features are then weighted and combined with retrieved features Γ_j , emphasizing points closer to the center. The updated feature representation is:

$$\Gamma_j \leftarrow \Gamma_j + \gamma(\mathbf{P}_j) \odot \gamma(\mathbf{P}_j) \quad (12)$$

In this formulation, $\gamma(\mathbf{P}_j)$ represents the encoded spatial information of the retrieved neighbors, and Γ_j represents the features retrieved from the nearest neighbors. The element-wise multiplication \odot ensures that the final aggregated features are influenced by both local features and spatial encoding, effectively capturing detailed local geometry while preserving spatial relationships.

3.3.4 Neighbor Pooling

Following GPE Aggregation, the neighbor pooling process aggregates features using both mean and max operations across the neighbor dimension. For each point, the pooled features are calculated as:

$$\Phi_j = \text{Mean}(\Gamma_j) + \text{Max}(\Gamma_j), \quad \forall j \in \{1, \dots, N/2\} \quad (13)$$

Here, $\text{Mean}(\Gamma_j)$ and $\text{Max}(\Gamma_j)$ are permutation-invariant operations, ensuring that the order of neighbors does not affect the pooled features.

3.3.5 Aggregation Across Stages

The non-parametric feature encoder includes four stages, each producing pooled features Φ_j^s . After processing through all stages, global pooling is applied to the results from each stage. The final feature vector \mathbf{F} for the non-parametric feature encoder is obtained by concatenating the global mean and max features from all four stages:

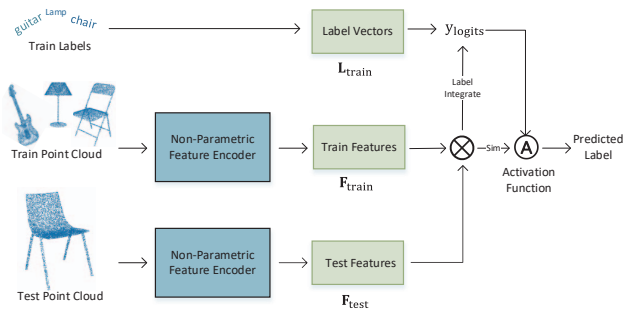


Figure 3. Illustration of the non-parametric classifier pipeline. The test feature is compared with the stored feature embeddings from the training set, and similarity scores are computed to assign the most likely class label based on proximity in feature space.

$$\mathbf{F} = \bigoplus_{s=1}^4 [\text{Mean}(\Phi_j^s) + \text{Max}(\Phi_j^s)] \quad (14)$$

This formulation captures and aggregates spatial and feature information across multiple levels by combining the mean and max features from each stage.

3.4. Non-Parametric Classifier

To preserve the non-parametric nature of Point-GN, we adopt a similarity-based classification approach [43]. Given a test point cloud, its feature vector \mathbf{F}_{test} is compared to the feature embeddings $\mathbf{F}_{\text{train}}$ from the training set. The classifier computes a similarity score between the test and training embeddings, which is then used to assign a class label based on the closest matching features. This similarity-driven mechanism avoids the need for traditional parametric models, maintaining the flexibility and efficiency of the non-parametric framework.

$$\mathbf{F}_{\text{train}} = \bigoplus_{m=1}^M \mathbf{F}_m \quad (15)$$

$$\mathbf{L}_{\text{train}} = \bigoplus_{m=1}^M \mathbf{L}_m \quad (16)$$

3.4.1 Feature Representation and Label Embedding

Given a training set of M point clouds $\{\mathcal{P}_m\}_{m=1}^M$, each belonging to one of C categories, we first extract a global feature vector \mathbf{F}_m for each point cloud \mathcal{P}_m using the non-parametric encoder. The corresponding labels $\{y_m\}_{m=1}^M$ are transformed into embedded label vectors \mathbf{L}_m .

The process of storing these feature and label embeddings is shown in Fig. 3. The feature embeddings are stored

in the global feature matrix $\mathbf{F}_{\text{train}}$, and the label embeddings are stored in the label matrix $\mathbf{L}_{\text{train}}$, both defined as follows:

3.4.2 Similarity-Based Classification

For a test point cloud $\mathcal{P}_{\text{test}}$, the non-parametric encoder generates the feature vector \mathbf{F}_{test} . We compute the similarity between \mathbf{F}_{test} and the stored training features in $\mathbf{F}_{\text{train}}$ using the following equation:

$$\text{Sim} = \mathbf{F}_{\text{test}} \cdot \mathbf{F}_{\text{train}}^T \quad (17)$$

The similarity scores in Sim are used to weight the corresponding label embeddings from $\mathbf{L}_{\text{train}}$. The final classification logits are computed using the following activation function:

$$\mathbf{y}_{\text{logits}} = \exp(-\gamma \cdot (1 - \text{Sim})) \cdot \mathbf{L}_{\text{train}} \quad (18)$$

Here, γ is a scaling factor, and $\exp(-\gamma \cdot (1 - \text{Sim}))$ serves as the activation function adapted from Tip-Adapter [41], where higher similarity scores result in stronger contributions from the corresponding labels in $\mathbf{L}_{\text{train}}$.

3.4.3 Classification Decision

The predicted class label is determined by applying an activation function to the logits, selecting the category with the highest value. In our case, we use a softmax activation for this final step:

$$c = \arg \max(\text{softmax}(\mathbf{y}_{\text{logits}})) \quad (19)$$

Through this similarity-based label integration, the classifier is able to effectively differentiate between various point cloud instances using a simple and efficient mechanism.

4. Experiments

In this section, we benchmark the performance of **Point-GN** against state-of-the-art methods for 3D shape classification. We conduct experiments on two widely recognized datasets: **ModelNet40** [34] and **ScanObjectNN** [28]. These datasets were selected for their complementary characteristics: ModelNet40 comprises clean, synthetic 3D models, while ScanObjectNN presents more challenging real-world scenarios with occlusions and background noise. By evaluating on these datasets, we aim to demonstrate the robustness and versatility of our approach across both synthetic and real-world data. Additionally, we compare Point-GN to existing non-parametric methods to showcase its efficiency and competitiveness.

Method	Acc. (%)	Param.
PointNet [20]	89.2	3.5 M
PointNet++ [21]	90.7	1.5 M
PointCNN [7]	92.2	0.6 M
DGCNN [32]	92.9	1.8 M
GBNet [23]	93.8	8.4 M
CurveNet [35]	93.8	2.0 M
PointNext-S [22]	93.2	1.4 M
PointMLP [11]	94.1	12.6 M
Point-NN [43]	81.8	0.0 M
Point-GN (ours)	85.3	0.0 M

Table 1. **Shape Classification on Synthetic ModelNet40 [34].** All compared methods take 1,024 points as input. We report the accuracy without the voting strategy.

4.1. Experimental Setup

We evaluate the performance of our **Point-GN** method on a system equipped with an NVIDIA RTX 4090 GPU. Although **Point-GN** is a non-parametric method and does not require traditional model training, the GPU significantly accelerates the inference process, allowing for efficient evaluation across the large and diverse **ModelNet40 [34]** and **ScanObjectNN [28]** datasets. This high-performance hardware ensures that our approach can handle the complexity and size of real-world 3D data, providing rapid evaluations during the benchmarking process.

4.2. Dataset Details

The **ModelNet40 [34]** dataset consists of 12,311 CAD models across 40 object categories, split into 9,843 samples for training and 2,468 for testing. This dataset is widely used for point cloud classification due to its clean, synthetic nature, providing a controlled environment for benchmarking.

In contrast, the **ScanObjectNN [28]** dataset presents a more challenging real-world scenario, with 2,902 samples across 15 object categories. Objects in ScanObjectNN are often occluded, cluttered, or contain background noise, providing a closer simulation to real-world 3D data. The dataset is divided into three official subsets: **OBJ-BG**, which contains objects with background noise, **OBJ-ONLY**, with objects without background, and **PB-T50-RS**, featuring partial occlusions and transformations. These subsets test the robustness of models under various degrees of complexity.

For both datasets, we follow the common practice of sampling 1,024 points from each object, as used in prior works (e.g., PointNet++ [21], DGCNN [32]). Our model combines **maximum pooling** and **average pooling** to enhance feature aggregation, inspired by DGCNN [32].

Method	OBJ-BG	OBJ-ONLY	PB-T50-RS	Param.
3DmFV [2]	68.2	73.8	63.0	-
PointNet [20]	73.3	79.2	68.2	3.5 M
PointNet++ [21]	82.3	84.3	77.9	1.5 M
DGCNN [32]	82.8	86.2	78.1	1.8 M
PointCNN [7]	86.1	85.5	78.5	-
GBNet [23]	-	-	80.5	8.4 M
PointMLP [11]	-	-	85.4	12.6 M
PointNeXt-S [22]	-	-	87.7	1.5 M
PointMetaBase-S [8]	-	-	87.9	0.6 M
Point-NN [43]	71.1	74.9	64.9	0.0 M
Point-GN (ours)	85.2	86.0	86.4	0.0 M

Table 2. **Shape Classification on the Real-world ScanObjectNN [28].** We report the accuracy (%) on three official splits of ScanObjectNN: OBJ-BG, OBJ-ONLY, and PB-T50-RS. The results in blue correspond to fully trained models, which show lower accuracy than our train-free method, Point-GN, which outperforms all others.

4.3. Shape Classification on ModelNet40

We evaluate the performance of Point-GN on the **ModelNet40 [34]** dataset in Tab. 1. Point-GN achieves an accuracy of **85.3%**, demonstrating strong performance in synthetic 3D shape classification. This result highlights Point-GN’s ability to effectively capture both local and global geometric features, all while maintaining a minimal model complexity.

When compared to the non-parametric Point-NN [43], Point-GN shows a **+3.5%** improvement in accuracy, despite having zero trainable parameters. This demonstrates the effectiveness of our approach in extracting meaningful features without relying on large parameter counts. Furthermore, Point-GN achieves an inference speed of **301 samples/second** (measured on our system), ensuring high efficiency for real-time applications. This is especially notable when compared to parametric models like PointMLP [11], which requires 12.6M parameters to achieve a slightly higher accuracy of 94.1%.

The combination of competitive accuracy and exceptional computational efficiency makes Point-GN an attractive choice for resource-constrained environments, where real-time performance and minimal model complexity are crucial.

4.4. Shape Classification on ScanObjectNN

On the **ScanObjectNN [28]** dataset (Tab. 2), Point-GN demonstrates superior performance in real-world scenarios, outperforming most existing methods across all three official splits: OBJ-BG, OBJ-ONLY, and PB-T50-RS. Notably, **Point-NN [43]**, the only other non-parametric method for comparison, is significantly outperformed by Point-GN. In

Method	5-way		10-way	
	10-shot	20-shot	10-shot	20-shot
DGCNN [32]	31.6	40.8	19.9	16.9
FoldingNet [38]	33.4	35.8	18.6	15.4
PointNet++ [21]	38.5	42.4	23.0	18.8
PointNet [20]	52.0	57.8	46.6	35.2
3D-GAN [33]	55.8	65.8	40.3	48.4
PointCNN [7]	65.4	68.6	46.6	50.0
Point-NN [43]	88.8	90.9	79.9	84.9
Point-GN (ours)	90.7	90.9	81.6	86.4

Table 3. **Few-shot Classification on ModelNet40 [34]**. We compute the mean accuracy (%) across 10 separate runs. The presented results of existing methods are sourced from [25].

the most challenging split, **PB-T50-RS**, Point-GN achieves a **+21.5%** improvement in accuracy over Point-NN, highlighting its robustness to occlusions and background noise in real-world data.

Rather than using different setups for each split, we adopted a single configuration for all three splits and aimed to find the best average performance. The **average accuracy** across the three splits for Point-GN is **85.89%**, demonstrating its consistency across varying conditions. This approach ensures we identify a setup that works best on average, rather than optimizing separately for each split.

When compared to fully trained models (shown in blue in Tab. 2), Point-GN consistently outperforms most, achieving higher accuracy across all splits. In particular, it surpasses models like PointNet [20], PointNet++ [21], and PointMLP [11] by substantial margins. The difference between Point-GN and the top-performing model, PointMetaBase-S [8], is less than 2%, indicating that Point-GN competes closely with state-of-the-art methods despite having zero trainable parameters.

These results underscore the power of Point-GN’s non-parametric design, which offers competitive performance while maintaining computational efficiency and avoiding the complexity of parameter-heavy models.

4.5. Few-shot Classification on ModelNet40

In the few-shot classification task on ModelNet40 [34] (Tab. 3), Point-GN outperforms existing methods, demonstrating the best performance in both the 5-way and 10-way settings. Notably, non-parametric methods, including Point-NN [43] and Point-GN, significantly outperform parametric models in this scenario. While traditional deep learning models with learnable parameters often struggle with overfitting when only a small number of training samples are available, both Point-NN and Point-GN manage to avoid this pitfall, achieving superior generalization.

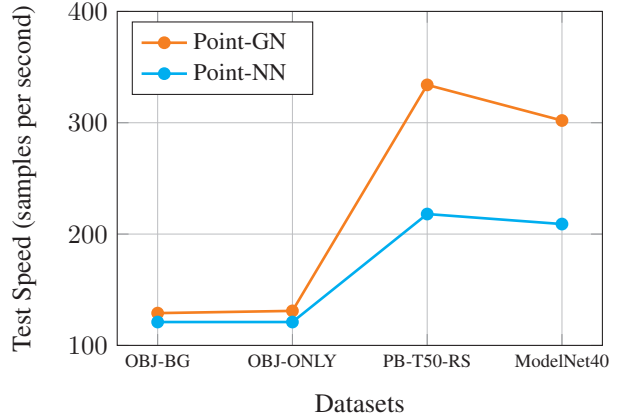


Figure 4. **Test Speed (samples per second) on ScanObjectNN [28] and ModelNet40 [34] datasets**. The plot compares the inference speed of Point-NN [43] and Point-GN on four different datasets. Point-GN shows significant improvements in inference speed across all datasets.

Point-GN, with its non-parametric design, achieves higher accuracy than Point-NN [43] in 3 out of 4 configurations (both the 5-way and 10-way 10-shot and 20-shot settings). In the remaining configuration (5-way 20-shot), Point-GN matches the performance of Point-NN, underscoring the consistency and robustness of our approach. Compared to traditional parametric methods, such as PointNet [20] and PointNet++ [21], which require more complex training and a larger number of parameters, Point-GN excels despite having zero trainable parameters, proving its effectiveness in few-shot learning scenarios.

These results highlight the power of our non-parametric method, which can achieve high accuracy with limited data and fewer resources, making it particularly suitable for applications with constrained training data.

4.6. Computational Complexity Analysis

We evaluate the computational efficiency of Point-GN in Fig. 4, where we compare its inference speed to Point-NN [43] by running both models on our system. Despite having zero trainable parameters, Point-GN demonstrates significantly faster inference across both the ScanObjectNN [28] and ModelNet40 [34] datasets. This efficiency makes Point-GN ideal for real-time applications, such as autonomous driving and robotic perception.

In contrast, Point-NN [43], while also non-parametric, achieves slower inference speeds, highlighting Point-GN as the more suitable choice for time-sensitive tasks that require both speed and high performance.

4.7. Ablation Study

We conducted an ablation study to evaluate the impact of different model configurations on the performance of Point-

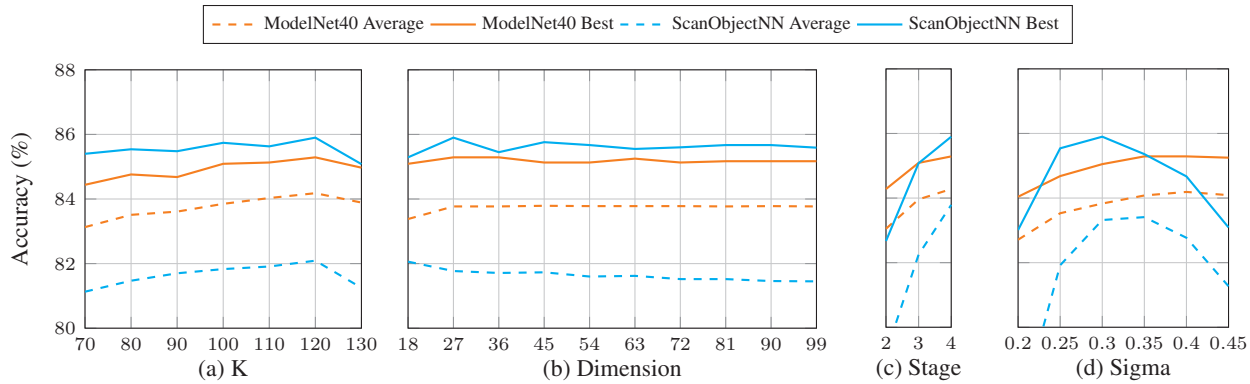


Figure 5. Ablation study results showing the sensitivity of Point-GN’s performance to key hyperparameters: (a) Number of neighbors (K), (b) Dimension of Gaussian Positional Encoding (GPE), (c) Number of stages, and (d) Sigma (σ). We compare the performance of the model on ModelNet40 [34] (orange) and ScanObjectNN [28] (cyan) datasets, showing both the average and best performances.

GN. Specifically, we examined the effect of four key factors: the number of neighbors k in the k -Nearest Neighbors algorithm, the dimension of Gaussian Positional Encoding (GPE), the number of stages in the model and the standard deviation σ of the Gaussian function.

Effect of Number of Neighbors (k). The number of neighbors k used in the k -NN algorithm also plays a critical role in model performance. Our experiments show that $k = 120$ offers the best trade-off between computational efficiency and accuracy. Smaller values, such as $k = 70$, fail to capture sufficient local context, while larger values, like $k = 130$, introduce irrelevant neighbors that confuse the model, hindering its ability to discern fine-grained geometric features. The performance comparison is shown in Fig. 5 (a).

Effect of GPE Dimension. The dimensionality of the Gaussian Positional Encoding (GPE) significantly affects model performance. As shown in Fig. 5 (b), increasing the dimension up to 27 improves accuracy, achieving the best performance in both ModelNet40 and ScanObjectNN datasets. However, beyond 45, further increases in dimension result in diminishing returns, with accuracy slightly declining. This suggests that while higher dimensions can capture more complex features, excessively high dimensions add unnecessary complexity without substantial performance gains.

Effect of Number of Stages. Increasing the number of stages in the model generally improves accuracy. The **4-stage configuration** achieves the highest accuracy of **85.3%** on the ModelNet40 dataset and **85.9%** on the ScanObjectNN dataset, suggesting that deeper models are better equipped to capture complex spatial relationships and improve classification performance. Refer to Fig. 5 (c) for a detailed comparison of performance across different stage configurations.

Impact of Sigma (σ). The σ parameter of the Gaus-

sian kernel (σ) determines the degree of locality in feature aggregation. For the ModelNet40 dataset, $\sigma = 0.35$ and $\sigma = 0.4$ yield the highest accuracy, effectively preserving both local and broader spatial contexts. On the other hand, for the ScanObjectNN dataset, we find that $\sigma = 0.3$ is the most effective, achieving the best accuracy by maintaining a balance between capturing fine-grained details and minimizing noise. Smaller values ($\sigma < 0.3$) fail to aggregate sufficient local features, while larger values ($\sigma > 0.4$) introduce excessive smoothing, negatively impacting performance. These trends are illustrated in Fig. 5 (d).

5. Conclusion

In this paper, we introduced Point-GN a novel non-parametric network for 3D point cloud classification that combines Gaussian Positional Encoding (GPE) with non-learnable components such as FPS and KNN to efficiently capture both local and global geometric structures. By eliminating the need for learnable parameters, Point-GN provides a highly efficient and lightweight model suitable for real-time and resource-constrained environments. Experimental results on ModelNet40 and ScanObjectNN demonstrate that Point-GN achieves competitive accuracy, outperforming existing non-parametric methods while requiring zero trainable parameters and delivering fast inference speeds.

For future work, we plan to extend the non-parametric framework of Point-GN by incorporating additional point cloud features and exploring its potential in more complex 3D tasks, such as semantic segmentation and object detection, further enhancing the model’s versatility and real-world applicability.

References

- [1] Mohammed Aldeen, Pedram MohajerAnsari, Jin Ma, Mashrur Chowdhury, Long Cheng, and Mert D. Pesé. An initial exploration of employing large multimodal models in defending against autonomous vehicles attacks. In *2024 IEEE Intelligent Vehicles Symposium (IV)*, pages 3334–3341, 2024. 2
- [2] Yizhak Ben-Shabat, Michael Lindenbaum, and Anath Fischer. 3dmfv: Three-dimensional point cloud classification in real-time using convolutional neural networks. *IEEE Robotics and Automation Letters*, 3(4):3145–3152, 2018. 6
- [3] Yifan Feng, Zizhao Zhang, Xibin Zhao, Rongrong Ji, and Yue Gao. Gvcnn: Group-view convolutional neural networks for 3d shape recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 264–272, 2018. 2
- [4] Ankit Goyal, Hei Law, Bowei Liu, Alejandro Newell, and Jia Deng. Revisiting point cloud shape classification with a simple and effective baseline. In *International Conference on Machine Learning*, pages 3809–3820. PMLR, 2021. 2
- [5] Jia Guo, Xuanxia Yao, Mengyu Shen, Jiafei Wang, and Wanyou Liao. A deep learning network for point cloud of medicine structure. In *2018 9th International Conference on Information Technology in Medicine and Education (ITME)*, pages 683–687. IEEE, 2018. 1
- [6] Qingyong Hu, 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*, pages 11108–11117, 2020. 2
- [7] Yangyan Li, Rui Bu, Mingchao Sun, Wei Wu, Xinhan Di, and Baoquan Chen. Pointcnn: Convolution on x-transformed points. *Advances in neural information processing systems*, 31, 2018. 2, 6, 7
- [8] Haojia Lin, Xiawu Zheng, Lijiang Li, Fei Chao, Shanshan Wang, Yan Wang, Yonghong Tian, and Rongrong Ji. Meta architecture for point cloud analysis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17682–17691, 2023. 6, 7
- [9] Zhijian Liu, Haotian Tang, Yujun Lin, and Song Han. Point-voxel cnn for efficient 3d deep learning. *Advances in neural information processing systems*, 32, 2019. 2
- [10] Qiang Lu, Chao Chen, Wenjun Xie, and Yuetong Luo. Point-ngcnn: Deep convolutional networks on 3d point clouds with neighborhood graph filters. *Computers & Graphics*, 86:42–51, 2020. 2
- [11] Xu Ma, Can Qin, Haoxuan You, Haoxi Ran, and Yun Fu. Rethinking network design and local geometry in point cloud: A simple residual mlp framework. *arXiv preprint arXiv:2202.07123*, 2022. 1, 6, 7
- [12] Daniel Maturana and Sebastian Scherer. Voxnet: A 3d convolutional neural network for real-time object recognition. In *2015 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pages 922–928. IEEE, 2015. 2
- [13] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021. 2
- [14] Pedram MohajerAnsari, Alkim Domeke, Jan de Voor, Arkajyoti Mitra, Grace Johnson, Amir Salarpour, Habeeb Olu-fowobi, Mohammad Hamad, and Mert D Pesé. Discovering new shadow patterns for black-box attacks on lane detection of autonomous vehicles. *arXiv preprint arXiv:2409.18248*, 2024. 2
- [15] AAM Muzahid, Wanggen Wan, Ferdous Sohel, Lianyao Wu, and Li Hou. Curvenet: Curvature-based multitask learning deep networks for 3d object recognition. *IEEE/CAA Journal of Automatica Sinica*, 8(6):1177–1187, 2020. 1
- [16] Anh Nguyen, Jason Yosinski, and Jeff Clune. Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 427–436, 2015. 2
- [17] Michael Niemeyer, Lars Mescheder, Michael Oechsle, and Andreas Geiger. Differentiable volumetric rendering: Learning implicit 3d representations without 3d supervision. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3504–3515, 2020. 2
- [18] SM Moein Peyghambarzadeh, Fatemeh Azizmalayeri, Hassan Khotanlou, and Amir Salarpour. Point-planenet: Plane kernel based convolutional neural network for point clouds analysis. *Digital Signal Processing*, 98:102633, 2020. 2
- [19] Charles R Qi, Wei Liu, Chenxia Wu, Hao Su, and Leonidas J Guibas. Frustum pointnets for 3d object detection from rgb-d data. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 918–927, 2018. 1
- [20] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 652–660, 2017. 1, 2, 6, 7
- [21] Charles Ruizhongtai Qi, 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*, 30, 2017. 1, 2, 6, 7
- [22] Guocheng Qian, Yuchen Li, Houwen Peng, Jinjie Mai, Hasan Hammoud, Mohamed Elhoseiny, and Bernard Ghanem. Pointnext: Revisiting pointnet++ with improved training and scaling strategies. *Advances in neural information processing systems*, 35:23192–23204, 2022. 6
- [23] Shi Qiu, Saeed Anwar, and Nick Barnes. Geometric back-projection network for point cloud classification. *IEEE Transactions on Multimedia*, 24:1943–1955, 2021. 6
- [24] Amir Salarpour, Hassan Khotanlou, and Nikolaos Mavridis. Long-term estimation of human spatial interactions through multiple laser ranging sensors. In *2014 International Conference on Robotics and Emerging Allied Technologies in Engineering (iCREATE)*, pages 109–114. IEEE, 2014. 2
- [25] Charu Sharma and Manohar Kaul. Self-supervised few-shot learning on point clouds. *Advances in Neural Information Processing Systems*, 33:7212–7221, 2020. 7
- [26] Hang Su, Subhransu Maji, Evangelos Kalogerakis, and Erik Learned-Miller. Multi-view convolutional neural networks

- for 3d shape recognition. In *Proceedings of the IEEE international conference on computer vision*, pages 945–953, 2015. 2
- [27] Hugues Thomas, 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*, pages 6411–6420, 2019. 2
- [28] Mikaela Angelina Uy, Quang-Hieu Pham, Binh-Son Hua, Thanh Nguyen, and Sai-Kit Yeung. Revisiting point cloud classification: A new benchmark dataset and classification model on real-world data. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1588–1597, 2019. 2, 5, 6, 7, 8
- [29] Jacob Varley, Chad DeChant, Adam Richardson, Joaquín Ruales, and Peter Allen. Shape completion enabled robotic grasping. In *2017 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pages 2442–2447. IEEE, 2017. 1
- [30] A Vaswani. Attention is all you need. *Advances in Neural Information Processing Systems*, 2017. 2
- [31] Lei Wang, Yuchun Huang, Jie Shan, and Liu He. Msnet: Multi-scale convolutional network for point cloud classification. *Remote Sensing*, 10(4):612, 2018. 2
- [32] Yue Wang, 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)*, 38(5):1–12, 2019. 2, 6, 7
- [33] Jiajun Wu, Chengkai Zhang, Tianfan Xue, Bill Freeman, and Josh Tenenbaum. Learning a probabilistic latent space of object shapes via 3d generative-adversarial modeling. *Advances in neural information processing systems*, 29, 2016. 7
- [34] Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 3d shapenets: A deep representation for volumetric shapes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1912–1920, 2015. 2, 5, 6, 7, 8
- [35] Tiange Xiang, Chaoyi Zhang, Yang Song, Jianhui Yu, and Weidong Cai. Walk in the cloud: Learning curves for point clouds shape analysis. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 915–924, 2021. 6
- [36] Tianxu Xu, Dong An, Yuetong Jia, and Yang Yue. A review: Point cloud-based 3d human joints estimation. *Sensors*, 21(5):1684, 2021. 1
- [37] Jiancheng Yang, Qiang Zhang, Bingbing Ni, Linguo Li, Jinxian Liu, Mengdie Zhou, and Qi Tian. Modeling point clouds with self-attention and gumbel subset sampling. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3323–3332, 2019. 2
- [38] Yaoqing Yang, Chen Feng, Yiru Shen, and Dong Tian. Foldingnet: Point cloud auto-encoder via deep grid deformation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 206–215, 2018. 7
- [39] Jianhui Yu, Chaoyi Zhang, Heng Wang, Dingxin Zhang, Yang Song, Tiange Xiang, Dongnan Liu, and Weidong Cai. 3d medical point transformer: Introducing convolution to attention networks for medical point cloud analysis. *arXiv preprint arXiv:2112.04863*, 2021. 1
- [40] Xumin Yu, 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*, pages 19313–19322, 2022. 2
- [41] Renrui Zhang, Rongyao Fang, Wei Zhang, Peng Gao, Kunchang Li, Jifeng Dai, Yu Qiao, and Hongsheng Li. Tip-adapter: Training-free clip-adapter for better vision-language modeling. *arXiv preprint arXiv:2111.03930*, 2021. 5
- [42] Renrui Zhang, Ziyu Guo, Wei Zhang, Kunchang Li, Xupeng Miao, Bin Cui, Yu Qiao, Peng Gao, and Hongsheng Li. Pointclip: Point cloud understanding by clip. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8552–8562, 2022. 2
- [43] Renrui Zhang, Lihui Wang, Ziyu Guo, Yali Wang, Peng Gao, Hongsheng Li, and Jianbo Shi. Parameter is not all you need: Starting from non-parametric networks for 3d point cloud analysis. *arXiv preprint arXiv:2303.08134*, 2023. 2, 5, 6, 7
- [44] Zhiyuan Zhang, Binh-Son Hua, and Sai-Kit Yeung. Shellnet: Efficient point cloud convolutional neural networks using concentric shells statistics. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1607–1616, 2019. 2
- [45] Wang Zhangyu, Yu Guizhen, Wu Xinkai, Li Haoran, and Li Da. A camera and lidar data fusion method for railway object detection. *IEEE Sensors Journal*, 21(12):13442–13454, 2021. 1
- [46] Hengshuang Zhao, Li Jiang, Jiaya Jia, Philip HS Torr, and Vladlen Koltun. Point transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 16259–16268, 2021. 2
- [47] Yin Zhou and Oncel Tuzel. Voxelnet: End-to-end learning for point cloud based 3d object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4490–4499, 2018. 1