Original Manuscript ID : VT-2024-03829

Original Correspondence Title: "AAG -Net: Angle-Attention Graph Neural Network for Point Cloud Sampling "

To: Editor, IEEE Transactions on Vehicular Technology

Re: Response to reviewers

Dear Editor,

Thank you for allowing a resubmission of our manuscript, with an opportunity to address the reviewers' comments.

We are uploading (a) our point-by-point response to the comments (below) (response to reviewers, under Author's Response Files"), (b) an updated manuscript with yellow highlighting indicating changes (as Highlighted PDF"), and (c) a clean updated manuscript without highlights ("Main Manuscript").

Best regards,

<SEOKJIN HONG> et al.

Reviewer#1, Concern # 1 :

There are multiple occurrences of "in this letter", do you mean "in this section"?.

Author response:

We sincerely appreciate your careful review and valuable comments regarding the terminology used in our manuscript. Consistency and clarity in terminology are essential for the readability and professional presentation of research work. As you correctly pointed out, the repeated usage of "in this letter" may be ambiguous or inappropriate given the specific context of the sections in our manuscript. Therefore, we fully agree that the phrase should be revised to more precisely reflect the content of each section.

Author action:

We updated the manuscript at the two identified locations:

(Update # 1)

Section II.B (Global-Local Graph Point Attention)

Before Revision:

In this letter, we adopt the hierarchical point set feature learning process proposed by PointNet++ [2] as the basis for hierarchical learning of point clouds

After Revision:

In this section, we adopt the hierarchical point set feature learning process proposed by PointNet++ [2] as the basis for hierarchical learning of point clouds

(Update # 2)

Section II.B (Global-Local Graph Point Attention)

Before Revision:

In this letter, the points containing hierarchically learned spatial structure information are connected to global information feature vectors and graphs to combine local and global information.

After Revision:

In this section, the points containing hierarchically learned spatial structure information are connected to global information feature vectors and graphs to combine local and global information.

Reviewer#1, Concern # 2 :

On page 2, "AGG-Net" is not consistent with the "AAG-Net" stated earlier

Author response:

We sincerely appreciate your careful review and valuable comments regarding the typographical error identified in our manuscript. Accurate naming conventions are critical for clarity and precise communication of research. As you pointed out, the typographical error "AGG-Net" should correctly be "AAG-Net," consistent with the naming used throughout the manuscript.

Author action:

We thoroughly reviewed the manuscript and corrected all occurrences of the typographical error from "AGG-Net" to "AAG-Net." This ensures consistency and correctness in referencing our proposed model.

Abstract

Incorrect: "we propose AGG-Net"

Corrected: "we propose AAG-Net"

Section II (Methods) introduction

Incorrect: "overview of our proposed AGG-Net"

Corrected: "overview of our proposed AAG-Net"

Section III.D (Ablation Study)

Incorrect: "accuracy of AGG-Net trained on GNNs"

Corrected: "accuracy of AAG-Net trained on GNNs"

Section III.D (Ablation Study)

Incorrect: "sampling of AGG-Net"

Corrected: "sampling of AAG-Net"

Conclusion

Incorrect: "we proposed AGG-Net"

Corrected: "we proposed AAG-Net"

Conclusion

Incorrect: "AGG-Net method achieves"

Corrected: "AAG-Net method achieves"

Reviewer#1, Concern # 3 :

In equation 4, why not use quaternion representation instead of decomposing the angles? That can lead to more concise form and potentially save some computation

Author response:

We sincerely appreciate the reviewer's insightful comment. Indeed, quaternion representations are compact and effective in encoding 3D rotations, and they are particularly well suited for tasks involving object pose, orientation tracking, or rigid body transformations.

In our case, however, the goal is not to rotate vectors but to help the model **learn meaningful local geometric relationships** between a center point and its neighbors. From a deep learning input perspective, **decomposing the directional information into axis-specific angles (i.e., θ_x, θ_y, θ_z)** offers the advantage of **explicitly providing interpretable cues** about where each neighbor lies in 3D space.

While quaternions implicitly contain rotational information, their components do not directly indicate the degree to which a neighbor point is aligned with each axis. In contrast, our approach allows the network to receive **explicit per-axis orientation signals**, which facilitates attention learning by making the local spatial distribution more transparent to the model. This helps suppress redundant directions and emphasize diverse neighbor orientations during feature aggregation.

We fully agree that quaternion-based modeling has strong merits and may offer advantages in tasks with global rotation invariance or transformation learning. However, for our specific objective—local structure modeling for point cloud sampling—**axis-wise angle decomposition provided more direct and interpretable input features**, which proved effective in both performance and learning stability.

We appreciate the reviewer's perspective and will consider quaternion-based designs for future work where rotational dynamics play a larger role.

Reviewer#1, Concern # 4 :

On page 3, "and 2×S edges is generated", please explain why there are 2 x S edges, is the graph directed?

Author response:

We sincerely appreciate your insightful feedback and careful attention to detail regarding the clarification needed on the structure and directionality of the graph edges in our manuscript. Clearly defining these aspects is indeed crucial for ensuring that readers fully understand our methodology and can accurately reproduce our work. We acknowledge that the previous description was insufficient and could lead to ambiguity.

Author action:

To address your valuable comment, we have revised the manuscript to clearly specify that our graph structure is a directed graph. Specifically, in the Methods > Global-Local Graph Point Attention section (Page 3), we explicitly mentioned that the edges are directed from local points to the global point. Consequently, this configuration results in a total of "2 × S" edges, as each local point individually connects to the global feature point, thereby clearly reflecting the directional relationship established between local and global features.

(Update # 1)

Section II.B (Global-Local Graph Point Attention)

Before Revision:

By connecting the global and local information with edges, a graph consisting of (S +1) × C point features and 2 × S edges is generated.

After Revision:

By connecting the global and local information with directed edges from each local point to the global point, we construct a directed graph consisting of (S+1)×C point features and S directed edges. Each local point connects individually to the global point, establishing a unidirectional relationship from local to global features.

Reviewer#1, Concern # 5 :

In table 1, what are the precision and recall? Any false positives? How much data in the dataset is used for training and how much for testing and validation?

Author response:

We sincerely appreciate your thoughtful and constructive suggestion to include additional evaluation metrics such as Precision and Recall in our manuscript. Incorporating these metrics indeed provides a more comprehensive and insightful evaluation of our method's performance. Your recommendation is very helpful, as Accuracy alone may not sufficiently reflect all critical aspects of our model's performance, especially in classification tasks where class imbalance and false positives or negatives could significantly impact the overall evaluation. Furthermore, clearly specifying the dataset splits ensures reproducibility and enhances transparency, allowing readers and reviewers to better understand and validate our experimental results.

Author action:

(Update # 1)

Section III.C (Experimental Results)

In response to your valuable feedback, we have performed additional experiments specifically for our proposed method, AAG-Net, to compute Precision and Recall. These metrics have now been separately presented in the newly added Table II of our manuscript, providing a detailed performance analysis of our model.

TABLE II
PRECISION(%) AND RECALL (%) ON MODELNET40

Sample size	8	16	32	64
Precision	8.7	24.87	54.53	79.26
Recall	24.31	55.12	76.92	87.53

(Update # 2)

Section III.A (Datasets)

Before Revision:

To evaluate the performance of the proposed approach on the 3D point cloud sampling task, the ModelNet40 dataset was used. The ModelNet40 classification dataset consists of computer-generated 3D CAD models. It is composed of 12,311 data instances divided into 40 different categories. The dataset includes carefully labeled models that are standardized in size and orientation. Therefore, ModelNet40 is primarily used to evaluate the performance of models on classification tasks.

After Revision:

To evaluate the performance of the proposed approach on the 3D point cloud sampling task, we utilized the ModelNet40 dataset, which consists of computer-generated 3D CAD models. The dataset contains a total of 12,311 instances categorized into 40 different classes, with carefully labeled models standardized in size and orientation. The ModelNet40 dataset is provided with a predefined split consisting of 9,843 instances for training and 2,468 instances for testing, following official train-test split. ModelNet40 is widely used as a benchmark for evaluating the performance of classification models on point cloud data.

Reviewer#1, Concern # 6 :

In table 2, given that there is tiny drop in "accuracy" when removing the "Angle MLP", is it really useful? Is the slight increase in accuracy more caused by the increase in the network parameters?

Author response:

Thank you very much for your insightful comment regarding the role and utility of the Angle MLP in our model. Although the observed accuracy difference is small when the Angle MLP is removed, we believe its role extends beyond simply adding additional parameters.

The primary motivation for incorporating the Angle MLP is to preprocess angular information before feeding it into the attention mechanism. By embedding the raw angular data into a more expressive representation, the Angle MLP enables the model to more effectively capture subtle but critical spatial relationships among points. Direct concatenation of raw angles to the attention mechanism would result in less meaningful representations, potentially limiting the model's capacity to accurately learn from angular information.

Moreover, even small accuracy improvements can be highly significant in 3D point cloud classification tasks, especially when dealing with sparse or limited sampled points. Thus, the slight improvement observed with the Angle MLP is indicative of its effective preprocessing role rather than merely increased model complexity.

To further support this argument, we explicitly clarified this preprocessing role of the Angle MLP in our revised manuscript .

Author action:

Section II.A (Angle Attention Mechanism)

we explicitly clarified this preprocessing role of the Angle MLP in our revised manuscript

"Specifically, rather than directly concatenating raw angle values to the attention mechanism, an MLP (Angle MLP) is used to preprocess and embed these angular values into more expressive representations. This allows the attention mechanism to better capture and leverage meaningful geometric relationships, enhancing the overall representational capacity of the model."

Reviewer#2, Concern # 1 :

The angle-based attention mechanism and hierarchical learning in AAG-Net contribute to increased computational complexity compared to simpler sampling methods like Farthest Point Sampling (FPS). To improve efficiency, the authors could discuss lightweight versions of AAG-Net, such as by reducing the depth of Graph Neural Networks (GNNs) or employing sparse attention mechanisms to minimize computation without sacrificing performance.

Author response:

Thank you very much for your insightful suggestion regarding computational complexity and potential lightweight strategies for AAG-Net. We fully agree with your observation that the angle-based attention mechanism and hierarchical structure inherently increase computational complexity, making it necessary to explore lighter variants of our model.

Author action:

Section III.D (Ablation Study)

To address your valuable suggestion, we conducted additional experiments by reducing the depth of our GNN structure (reducing the number of layers from 4 to 3) and evaluated its impact on inference time and accuracy. These additional experimental results have been clearly summarized in a newly added comparison table within the revised manuscript.

TABLE IV
EFFICIENCY COMPARISON BETWEEN ORIGINAL AND LIGHTWEIGHT AAG-NET

Sample size	Method	Layers	Time(m/s)	Accuracy
32	SampleNet	-	8.61	80.32
	Ours	4	14.28	88.25
	Ours(Light)	3	11.62	83.95

Specifically, the results and analysis of this lightweight variant of AAG-Net are now explicitly presented in the revised manuscript, within the Ablation Study section (III.D), immediately following the existing ablation experiments. We discuss that although reducing the number of layers resulted in some decrease in accuracy, the lightweight variant of AAG-Net still achieves competitive performance compared to state-of-the-art sampling methods, highlighting its practical applicability under reduced computational complexity.

We have added a new paragraph in Section III.D (Ablation Study) discussing a lightweight version of AAG-Net with a reduced number of GNN layers.

“To evaluate the efficiency-performance tradeoff, we also tested a lightweight version of AAG-Net with fewer GNN layers. As summarized in Table IV, this variant maintained competitive performance with reduced inference time.”

Reviewer#2, Concern # 2 :

Additionally, the reliance on k-NN for graph construction may introduce scalability challenges when dealing with large-scale point clouds. The authors could consider alternative graph-building strategies, such as approximate nearest neighbor (ANN) search or adaptive sampling techniques, to enhance computational efficiency while maintaining accuracy.

Author response:

We sincerely appreciate your valuable feedback regarding the potential scalability issues associated with using k-NN for graph construction, particularly in the context of large-scale point clouds. We agree with your insightful suggestion that approximate nearest neighbor (ANN) methods can significantly improve computational efficiency without substantially compromising accuracy. Given that our method considers not only spatial coordinates but also additional high-dimensional feature vectors H = $\{h_1, h_2, ..., h_n\} \in \mathbb{R}^K$, using ANN methods becomes particularly advantageous and relevant, ensuring that our neighbor selection remains efficient and scalable.

Author action:

To directly address your recommendation, we replaced the conventional k-NN approach with an approximate nearest neighbor (ANN) search method. Specifically, we adopted the ANNOY (Approximate Nearest Neighbors Oh Yeah) algorithm due to its efficiency and effectiveness in handling high-dimensional data and large-scale point clouds. As an illustrative example, we performed experiments under conditions of 1024 points, 256-dimensional feature vectors, and 16 neighbors per point. The results demonstrated that at higher dimensions (starting from 256-dimensional feature vectors), ANN significantly outperforms k-NN in terms of computational efficiency:

- KNN graph construction time: 1.7 ms
- ANN graph construction time (using ANNOY): 1.1 ms

(Update # 1)

Section II.A (Angle Attention Mechanism)

We carefully revised figures and text that previously referenced k-NN, replacing them explicitly with ANN.
In the revised manuscript, we explicitly integrated your suggestion by modifying the Methods section (II.A Angle Attention Mechanism) to clearly justify the adoption of ANN:

"From the coordinates and features H = $\{h_1, h_2, ..., h_n\} \in \mathbb{R}^K$, previous methods normally utilize k-NN for neighbor selection. However, due to scalability concerns arising from the high-dimensional nature of the features and increased dataset size, we adopted approximate nearest neighbor (ANN) search, specifically using the ANNOY(Approximate Nearest Neighbors Oh Yeah) algorithm, to efficiently select neighbors."

(Update # 2)

Section III.C (Experimental Results)

Section III.D (Ablation Study)

Due to the adoption of the ANN method for neighbor selection, we updated the results reported in **Table I,II.** The revised accuracy values in Table now reflect the performance achieved using ANN instead of the previously employed k-NN, ensuring the results presented are consistent with the revised method throughout the manuscript.

TABLE I
CLASSIFICATION ACCURACY (%) ON MODELNET40

Sample size	8	16	32	64
RS	8.7	24.87	54.53	79.26
FPS	24.31	55.12	76.92	87.53
SampleNet	78.36	80.60	80.32	79.36
APSNet	81.42	83.89	88.15	88.38
AAG-Net(ours)	87.24	87.93	88.25	88.70

TABLE III
CLASSIFICATION ACCURACY OF ABLATION MODELS ON MODELNET40

Exp	Ablation model	Accuracy
1	AAG-Net(ours)	87.24
2	Remove Angle MLP	87.01
3	Remove Angle	84.06
4	Remove GL-GPA	83.21

Reviewer#2, Concern # 3 :

Finally, real-world factors like sensor noise, missing data, and occlusions could impact the robustness of AAG-Net. The authors are encouraged to briefly discuss these potential challenges and suggest possible solutions, such as data augmentation, denoising techniques, or uncertainty-aware sampling strategies, to ensure the model remains effective in diverse real-world scenarios.

Author response:

We sincerely appreciate your thoughtful and valuable feedback regarding the robustness of our proposed AAG-Net in real-world scenarios. We recognize that issues such as sensor noise, missing data, and occlusions significantly influence model performance and reliability, especially in practical autonomous driving applications. Addressing these real-world challenges is indeed critical for ensuring that our model remains robust, effective, and applicable across diverse operational conditions. Your suggestions on including discussions about possible solutions such as data augmentation, denoising techniques, and uncertainty-aware sampling strategies are particularly beneficial and relevant.

Author action:

Section IV (Conclusion section)

In response to your valuable recommendation, we have explicitly discussed strategies to enhance robustness in the revised manuscript. Specifically, we have included the following discussion in the Conclusion section:

" In practical autonomous driving scenarios, however, the robustness of AAG-Net may be affected by sensor noise, missing data, and occlusions. To address these challenges, future research could explore data augmentation strategies—such as injecting simulated noise, point dropout, or artificial occlusions during training—to enhance resilience. Furthermore, geometric denoising techniques, including Statistical Outlier Removal and bilateral filtering, could be employed as preprocessing steps to improve input quality. These directions could further increase the applicability of AAG-Net in real-world environments."

Reviewer#3, Concern # 1 :

In the introduction or conclusion, it would be valuable to discuss the potential of AAG-Net for tasks beyond classification in vehicular technology. Specifically, how could the angle attention mechanism be adapted to enhance other critical tasks, such as object detection or semantic segmentation, in autonomous driving scenarios

Author response:

We sincerely appreciate your insightful suggestion regarding the broader applicability of our angle-based attention mechanism, particularly emphasizing its potential role in enhancing 3D object detection tasks within autonomous driving scenarios. Your suggestion is highly valuable, as recent research has shown that effective learned sampling of points significantly improves object detection performance.

Author action:

Section IV (Conclusion section)

To clearly address your recommendation, we expanded our discussion in the **Conclusion section** of the manuscript, explicitly referencing relevant recent research such as LSNet, a learned sampling network developed specifically for 3D object detection tasks. The revised manuscript now includes the following detailed explanation:

"Furthermore, although our proposed AAG-Net was validated specifically for classification tasks, the angle-based attention mechanism could also significantly contribute to improving 3D object detection tasks in autonomous driving scenarios. Specifically, recent studies such as LSNet [Ref] demonstrate the critical importance of selecting representative and informative points through learned sampling methods for enhanced object detection performance. Incorporating our angle-based attention into such learned sampling frameworks could potentially provide superior point selections by effectively leveraging angular relationships among neighboring points."

Reviewer#3, Concern # 2 :

Elaborate on the process of selecting the hyperparameters (β, γ, δ, λ). Were they tuned using a validation set, or were they determined empirically?

Author response:

Thank you very much for your insightful question regarding the hyperparameter selection process (β, γ, δ, λ) in our manuscript. Clearly outlining the rationale behind hyperparameter choices significantly enhances the reproducibility and transparency of our experimental approach. We agree that providing clear and detailed reasoning for these selections is essential for readers to understand and replicate our results accurately.

Author action:

Section III.B (Implementation Details)

We clarify that the hyperparameters (β, γ, δ, λ) were determined empirically, guided by established practices from prior research, specifically the validated hyperparameter settings provided by SampleNet [Ref].

In the revised manuscript, we explicitly included this clarification in Section III.B (Implementation Details) as follows:

"The hyperparameters β, γ, and δ were empirically set based on values provided by SampleNet [Ref] (β=1, γ=1, δ=0), and the hyperparameter λ=30 was empirically determined through preliminary experiments to optimally balance the task-specific and sampling losses."
