• **Article** •

FDCPNet: Feature Discrimination and Context Propagation Network for 3D Shape Representation

Weimin Shi [1], Yuan Xiong [1], Qianwen Wang [1], Han Jiang[1], Zhong Zhou [1, 2*]

1. State Key Laboratory of Virtual Reality Technology and Systems, School of Computer Science and Engineering, Beihang University, Beijing, 100191, P.R.China.

2. Zhongguancun Laboratory, Beijing, P.R.China.

*** Corresponding author**：zz@buaa.edu.cn

Received: XX December 2023 **Accepted:** XX December 2023

Supported by National Key R&D Program of China under Grant (2022YFC3803600)

Abstract 3D shape representation using mesh data is essential in various applications, such as virtual reality and simulation technologies. Current methods extracting features from mesh edges or faces struggle with complex 3D models due to edge-based approaches missing global context and face-based methods overlooking variations in adjacent areas, which affects overall precision. To address these issues, we propose the Feature Discrimination and Context Propagation Network (FDCPNet), a novel approach that synergistically integrates local and global features in mesh datasets. FDCPNet is composed of two modules: 1) Feature Discrimination Module (FDM), which employs an attention mechanism to enhance the identification of key local features, 2) Context Propagation Module (CPM) enriches key local features by integrating global contextual information, facilitating a more detailed and comprehensive representation of crucial areas within the mesh model. Experiments on the popular datasets validate FDCPNet's effectiveness, showing a improvement in classification accuracy over the baseline MeshNet. Furthermore, even with reduced mesh face numbers and limited training data, FDCPNet achieves promising results, showing its robustness in variable complexity scenarios.

Keywords 3D shape representation; Mesh model; Meshnet; Feature discrimination; Context propagation

1 Introduction

In computer vision and graphics, analyzing and representing 3D shapes is critical for applications in virtual reality[1-3], simulation technology[4-6], and model simplification[7-9]. In recent years, deep learning has made significant progress in 3D shape analysis and representation[14, 15], enhancing the accuracy and efficiency of processing complex 3D models. It has been effectively applied to various types of 3D data, including multi-view images[10, 11], voxel grid[12, 13], point cloud[16, 19], and mesh[17, 18].

In this study, we consider 3D shape representation methods based on mesh data. Compared to other types of 3D data, mesh offers a more comprehensive and detailed 3D shape representation. This enhanced representation results from mesh data constructing a structural framework and continuous surfaces for 3D objects by defining vertices, edges, and faces. Additionally, mesh data incorporates extra attributes like textures, further enhancing the rendering effects.

Existing 3D shape representation methods based on mesh data typically extract features from mesh edges or faces to facilitate effective geometric and topological feature learning[28]. However, edge-based approaches like MeshCNN[17] and FPCNN[31] tend to focus on local features within mesh regions, often missing the hierarchical global shape context and struggling with complex topologies and abnormal geometric shapes. Conversely, face-based approaches, exemplified by MeshNet[18] and SubdivNet[21], may overlook subtle variations between faces, leading to less precise feature representation.

To address these issues, this paper proposes the Feature Discrimination and Context Propagation Network (FDCPNet), a novel method that effectively represents key regional features in 3D mesh models, as illustrated in Figure 1. FDCPNet identifies variations between different regions by enhancing key local features and further enriches their representation through the integration of contextual information.

In the implementation of FDCPNet, we adopt MeshNet as our baseline, leveraging its proficiency in extracting and processing face features from mesh data. FDCPNet consists of two modules: Feature Discrimination Module (FDM) and Context Propagation Module (CPM). Specifically, FDM employs a multi-head attention mechanism to identify and enhance key local features in the mesh model. This module initially divides input face features into multiple subsets, each corresponding to a different attention head. This division facilitates the capture of the mesh's local features across multiple dimensions. Subsequently, each subset is independently mapped to a specific subspace through linear transformation layers, allowing FDM to understand and represent the mesh's local geometric and topological features.

CPM plays a critical role in integrating local features, extracted by FDM, with the global contextual features in the mesh model. Utilizing residual connections and feature averaging scaling techniques, CPM computes average feature representations to effectively encapsulate the mesh model's global features. This computation of global feature effectively reflects the model's primary geometric and topological characteristics. By integrating global features with local features, CPM aligns fine details with the hierarchical structure, enriching the representation of key features. This results in a more accurate and comprehensive representation of the 3D shape.

Experimental results have shown the effectiveness of our FDCPNet. On the Manifold40 dataset, it surpassed the baseline model with a 1.9% increase in classification accuracy, reaching 90.3%. Notably, when the number of mesh faces was reduced to one-tenth of the training set (50 faces), the network maintained a classification accuracy of 89.7%. Furthermore, even when trained on datasets with limited sample sizes, comprising only four samples per category, it still achieved an accuracy of 60.5%. These results indicate its applicability and robustness in practical applications with meshes of varying complexities or limited data. In summary, our work contributes as follows:

- We propose FDCPNet, which integrates FDM and CPM to enhance the identification of key local features and enrich their representation through contextual information in 3D mesh models.

- We evaluate FDCPNet on popular datasets, achieving an effective improvement in classification accuracy over the baseline.
- We evidence our network's robustness under specific conditions, such as reduced numbers of mesh faces and limited training data, confirming its potential in practical mesh processing applications.

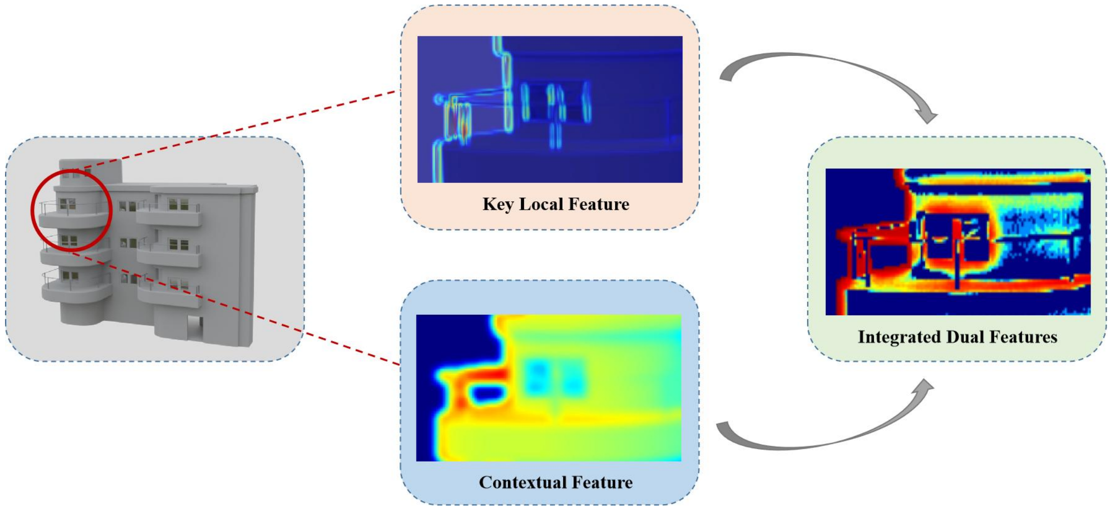

Figure 1 Method Overview. Integrating key local features with contextual features enriches the representation of crucial areas in 3D mesh models, facilitating a more comprehensive understanding of the 3D shape.

2 Related work

Recent advancements in deep learning have significantly advanced 3D shape representation, introducing innovative and efficient methods for modeling complex geometries. This section provides a comprehensive overview of these approaches and evaluates their strengths and limitations.

Multi-view images. This strategy involves representing 3D objects through several 2D perspectives, effectively translating them into a 2D format. Hang et al.[10] utilized conventional CNN architectures to integrate information from various 2D views into compact descriptors, enhancing the classification of 3D objects. Kalogerakis et al.[11] applied multi-view and multi-scale projections on different segments of 3D objects, subsequently employing Conditional Random Fields for region characterization. Kanezaki et al.[29] investigated fixed multi-view angles to identify essential views that accurately depict model features for classification. Nonetheless, this method can introduce redundant information and might not entirely capture all aspects of 3D details.

Voxel grid. In these methods, 3D shapes are transformed into binary voxel grids. Wu et al.[12] represented these shapes as probability distributions in subdivided voxels, applying a deep belief network for feature learning. Sedaghat et al.[13] included model orientation as a variable in classification, refining the voxel-based representation of 3D shapes. Ren et al.[32] aimed to solve the inefficiency issue in voxel convolution by representing models in multi-layer 2D formats, and extracting classification features through 3D deep shape descriptors. Despite their effectiveness in structured feature representation, these methods face limitations due to high memory requirements at increased resolutions, restricting their practicality in large-scale and intricate 3D models.

3D Point Cloud. These methods conceptualize 3D shapes as collections of unordered 3D points. Qi et al.[16] introduced PointNet, a framework for classifying and segmenting these point clouds, achieving efficient classification of sparse point cloud data. Subsequently, Qi et al.[19] further extracted multi-scale features from point clouds, improving 3D shape representation. Ma et al.[30] developed descriptors that aggregate progressively extracted local features, employing residual MLPs for the final shape representation. However, point cloud data lacks surface structure information and is inherently sparse and unordered, complicating the understanding and reconstruction of complex 3D shapes. However, the inherent sparsity and lack of surface structure in point cloud data pose challenges in comprehending and reconstructing complex 3D geometries[20].

3D Mesh. Mesh data excels in precisely representing complex geometries and intricate surface attributes in 3D shape representation. Current methods primarily focus on mesh model edges or faces. Hanocka et al.[17] developed MeshCNN, a method that learns mesh geometric and topological structures through edge-based convolutions and pooling. Liang et al.[31] focused on the information from neighboring edges, providing an intuitive description of 3D shapes by adjusting edge folding sequences. In contrast, Feng et al.[18] converted mesh data into face lists, adopting a face-unit feature learning approach for more precise and efficient 3D shape analysis and representation. Hu et al.[21] proposed an approach that accumulates local features from adjacent faces, building a structured hierarchy for representing 3D object shapes.

In this paper, we adopt MeshNet[18] as our baseline, capitalizing on its efficiency in extracting and processing face features from mesh data. Our approach aims to enhance the network's representation of 3D object shapes, by accurately and comprehensively identifying key features in mesh model.

3 Methods

3.1 Method Overview

In this study, the Feature Discrimination and Context Propagation Network (FDCPNet) is proposed, a novel method for 3D shape representation, as illustrated in Figure 2. This method includes two modules: the Feature Discrimination Module (FDM) and the Context Propagation Module (CPM). FDM utilizes a multi-head attention mechanism to precisely identify and emphasize key local features in the mesh. This process involves dividing input features into multiple subsets, each associated with a distinct attention head, thus enabling a thorough multi-dimensional assessment of local features, resulting in a rich representation of the mesh's local geometric and topological intricacies. Concurrently, CPM effectively combines these local features with global contextual information through the use of residual connections and feature averaging. The integration of the two modules in FDCPNet, emphasizing key local features and merging with global context, enhances its capability to accurately represent the shapes of 3D objects.

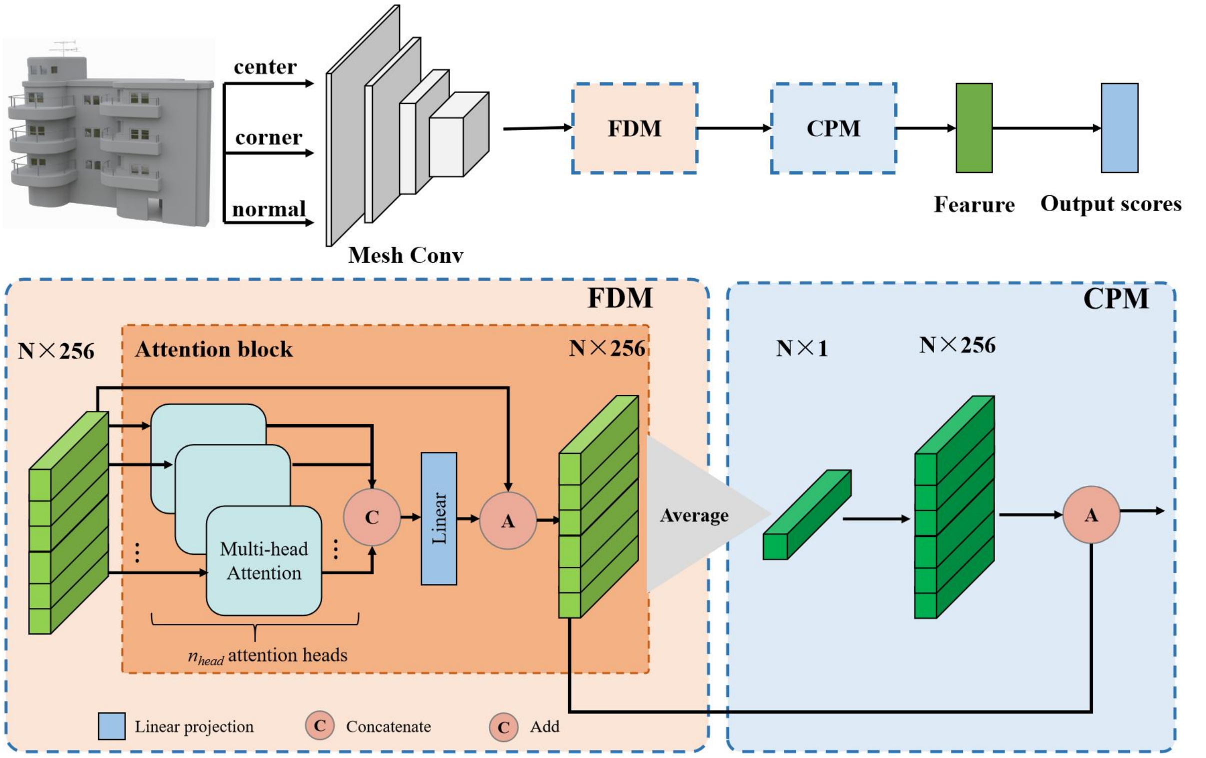

Figure 2 The proposed FDCPNet includes two modules: Feature Discrimination Module (Sec 3.2) and Context Propagation Module (Sec 3.3).

3.2 Feature Discrimination Module

We adopt MeshNet[18] as our baseline. MeshNet interprets mesh data as a sequence of faces, comprising centers, corners, and normals. MeshNet employs spatial descriptors to process face center values related to spatial positions. These descriptors are composed of multilayer perceptrons, used to output initial spatial features. Structural descriptors apply face rotation convolutions to capture internal and external structural information. This process effectively generates initial structural features of the faces. Subsequently, these features are processed through mesh convolution blocks, aggregating information from adjacent faces, resulting in a consolidated feature representation denoted as F_{agg}. Although MeshNet efficiently captures the basic structure of mesh models, its averaged face processing approach does not adequately emphasize key areas critical for shape representation. This can lead to omitting crucial local features, diminishing the network's overall effectiveness.

To address this, we propose the Feature Discrimination Module (FDM), aimed at enhancing the network's ability to identify and highlight key features in mesh models. In our technical implementation, FDM employs a multi-head attention mechanism to discern variations among the model's faces. It receives F_{agg}, a blend of spatial and structural features, as input. This mechanism divides the F_{agg} along feature dimensions d into multiple subsets, each aligned to a specific head in the multi-head framework. Assuming the total number of heads is n, F_{agg} is divided into multiple subsets:

$$\mathrm{F_{agg}} = \left[\mathrm{F_{split}^{0}}, \mathrm{F_{split}^{1}}, \cdots, \mathrm{F_{split}^{n-1}}\right],\ \mathrm{F_{split}^{n}} \in \mathrm{R}^{\mathrm{N}\times\frac{\mathrm{d}}{\mathrm{n}}} \qquad (1)$$

This strategy allows the network to capture diversified feature information across different dimensions. The subdivided feature subsets F_{split}^{i} are individually processed through linear transformation layers in each head, projecting each subset into a unique subspace. This projection allows for a multi-perspective analysis of the input features, thus equipping the network with the ability to discern critical geometric and topological attributes of mesh models from various dimensions. The computation in each subspace is formulated as:

$$F_{head}^{i} = W_i \cdot F_{split}^{i} + b_i \tag{2}$$

Where W_i and b_i are the weight and bias of the i_{th} head, respectively. By processing multiple heads in parallel, the network is enabled to simultaneously focus on various areas and local features in the mesh model. Each head independently captures different geometric or topological features in the mesh model. For instance, some heads may focus on the relative positioning of facets and topological connectivity, while others might concentrate on aspects like the curvature of surfaces or surface textures. This integration of diversified perspectives enables the network to detect changes and differences in local features in the model, thereby enhancing its capability to discriminate between distinct regions.

Following each head completes its independent feature transformation, an attention mechanism is applied to weight the features of various regions in the mesh model:

$$\text{Attention}(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{3}$$

Here, Q, K, V represent the Query, Key, and Value, respectively, derived from the transformation of F_{dead}^{i}. By computing the dot product between Q (Query) and K (Key), the network quantifies the correlation or similarity between features. For mesh models, this approach allows the model to discern which local features are geometrically or topologically significant, emphasizing those features vital to the model's overall structure and form. Additionally, the introduction of the multi-head attention mechanism effectively processes and analyzes noise and redundant information in mesh models. By optimizing the focus of the different heads in FDM, the network can identify and filter out non-essential local features.

3.3 Context Propagation Module

The Context Propagation Module (CPM) in the FDCPNet plays a crucial role in integrating local and global context representation in 3D mesh models. Its primary objective is to integrate the detailed local features, extracted by FDM, with comprehensive global contextual information. Such an integration strategy is crucial for achieving a thorough and nuanced representation of 3D mesh models.

Inspired by the research of Woo et al.[33], CPM first utilizes average feature scaling techniques to obtain global features:

$$\bar{x}_j = \frac{1}{c}\sum_{k=1}^{c} x_{j,k},\ j \in [0, N-1] \tag{4}$$

Where c represents the dimension of the feature. By calculating the global features $\bar{x}_j$, which reflect the model's predominant geometric shapes and essential topological structures, CPM provides a fast and effective method for the network to encapsulate global features. This method allows the network to intensify its focus on these key trends in subsequent processing phases, thereby improving the recognition of the model's comprehensive geometric and topological attributes.

These global features $\bar{x}_j$ are then integrated with the initially extracted local features. This fusion process is represented as:

$$x_j^{'} = x_j + \lambda \cdot \bar{x}_j \qquad (5)$$

Here, λ represents a learnable scaling factor designed to balance the impact of local and global features. By adjusting λ, the network can retain essential local features while incorporating global features to enhance understanding of the mesh model's overall structure. This fusion strategy improves the network's ability to capture features in mesh models that are significant at both local and global scales, consequently improving the accuracy in identifying key shape features of the model.

4 Experiments

4.1 Datasets

In experiments, we evaluate the effectiveness of FDCPNet using two datasets: the Manifold40[21] and the Cube Engraving[17]. The Manifold40 is an extension of ModelNet40[13], and includes 12,311 mesh models across 40 categories. The Cube Engraving includes 4,381 samples distributed across 22 categories. Each sample comprises small objects randomly embedded on the surface of a cube.

In particularly, Manifold40 is an optimized version of ModelNet40, it retains the same data structure and classification rules as ModelNet40. To conduct an in-depth comparison of representation methods across different data modalities, we directly compared FDCPNet with several methods implemented on ModelNet40, similar to comparisons previously conducted by Hu et al.[21].

4.2 Implementation Details

In our technical implementation for classification tasks, we employed a series of fully connected layers to perform the classification. Before the last two fully connected layers, we introduced dropout layers with a drop probability of 0.5. The model optimization relied on cross-entropy loss, and we used Accuracy (Acc) to measure the classification performance of our proposed method.

Our network was implemented on the PyTorch 1.8.0 platform and utilized a Stochastic Gradient Descent (SGD) optimizer for updating the weights and biases of the neural network. The training batch size was set to 32, with an initial learning rate of 5e−3, momentum of 0.9, and weight decay of 0.0005. During training, we also implemented a learning rate adjustment strategy. Specifically, if there was no significant improvement in performance on the validation set over several epochs, the learning rate would decrease at a fixed ratio. This approach allowed for fine-tuning of the model parameters, further optimizing network performance. To ensure effective processing of mesh data of varying scales, we standardized the input data by normalizing all mesh models to a uniform scale. Additionally, Gaussian noise was applied to the vertex positions for data augmentation, enhancing the network's generalization capabilities and robustness.

Table I Performance comparison of different methods in model classification.

Method	Modality	ModelNet40 Acc	Manifold40 Acc
3DShapeNets[22]	volume	77.3%	-
VoxNet[23]	volume	83.0%	-
FPNN[24]	volume	88.4%	-
LFD[25]	view	75.5%	-
MVCNN[26]	view	90.1%	-
Pairwise[27]	view	90.7%	-
PointNet[16]	point	89.2%	-
PointNet++[19]	point	90.7%	87.9%
SPH[31]	mesh	68.2%	-
MeshCNN[17]	mesh	-	79.3%
FPCNN[31]	mesh	-	83.1%
MeshNet[18]	mesh	91.9%	88.4%
FDCPNet (ours)	mesh	-	**90.3%**

Note: ModelNet40 Acc and Manifold40 Acc represent test results on the respective datasets. Manifold40, an improved refinement of ModelNet40, shares its classification rules, enabling direct comparability of method results across both datasets[21]. Comparative method results listed are from the Manifold40 benchmark[21], with "-" indicating that results for the corresponding dataset were not provided. As data in the ModelNet40 dataset have been updated to Manifold40, FDCPNet is exclusively evaluated on Manifold40.

4.3 Comparative Results

In the tasks of 3D model classification, our FDCPNet has shown performance advantages. As shown in Table I, FDCPNet achieved a classification accuracy of 90.3% on the Manifold40 dataset, significantly outperforming other mainstream 3D shape representation methods. Notably, compared to MeshNet, FDCPNet achieved a 1.9% increase in classification accuracy. These results not only validate the effectiveness of our method in processing mesh data but also reflect its advancement in capturing and expressing complex 3D shape features.

In a comprehensive comparison of 3D shape representation techniques, mesh models stand out for their ability to handle complex structures and intricate details. Compared to representations based on volume, point clouds, and other forms, mesh models can more accurately reflect the geometric features and topological structures of objects. FDCPNet effectively exploits these intrinsic benefits of mesh models by incorporating feature discrimination module and context propagation module. The experiment results indicate that this integration not only shows the potential of mesh models in 3D shape analysis but also shows that our method can further enhance the representational performance of mesh models for 3D shapes by comprehensively representing key features in mesh models.

In addition, we validate the performance of FDCPNet on the Cube Engraving dataset[21]. As shown in Table II, FDCPNet achieves a classification accuracy of 93.5%, which is a 1.3% improvement over MeshCNN. These results highlight FDCPNet's effectiveness in representing various 3D shape categories. Due to the unique characteristics of the Cube Engraving dataset, which includes large flat areas and subtle texture variations, these results further indicate that FDCPNet can learn key local features, thereby improving classification performance.

Table II Classification results of different methods on Cube Engraving.

Method	Modality	Acc
PointNet++[19]	point	64.3%
MeshCNN[17]	mesh	92.2%
FDCPNet (ours)	mesh	93.5%

4.4 Performance Analysis of FDM and CPM

Table III Analysis of the impact of each components in FDCPNet.

Method	Acc
MeshNet	88.4%
MeshNet + FDM	89.7%
MeshNet + CPM	88.8%
MeshNet + FDM + CPM (FDCPNet)	**90.3**%

This section evaluates the impact of FDM and CPM on the performance of network, as shown in Table III. When only FDM was added, there was a 1.3% increase in classification accuracy compared to MeshNet, achieving 89.7%. The improved network performance indicates that FDM effectively enhances the network's ability to identify key features in mesh models, confirming the importance of exploring and emphasizing local features in 3D shape representation.

Compared to MeshNet, adding only CPM to the network resulted in a 0.4% improvement in classification accuracy, a smaller increment compared to the network augmented with only FDM. This is attributed to the lack of sufficient local feature enhancement in the absence of FDM, resulting in less effective global information integration.

The most significant improvement in the network's performance was observed when both FDM and CPM were integrated. This indicates that the enhancement of local features and the integration of global features contribute to 3D shape representation. The precise processing of local features provides the network with rich detailed information. Simultaneously, the integration of global contextual information enhances the expression of key features, leading to an improved understanding and representation of the mesh model's overall geometric and topological structure.

4.5 Impact of Different Attention Heads' Number

Table IV Performance of Different Attention Heads' Number in FDM.

Method	Acc
FDCPNet (n=4)	89.5%
FDCPNet (n=8)	90.0%
FDCPNet (n=16)	**90.3**%
FDCPNet (n=32)	89.1%

Note: In FDCPNet (n=x), 'n' represents the number of attention heads in the Feature Discrimination Module (FDM).

Table IV illustrates the variation in network performance with different numbers of attention heads in FDM. An increase in the number of attention heads from 4 to 16 significantly improved the classification accuracy of the network. Specifically, with 4 attention heads, FDCPNet achieved a classification accuracy of 89.5%. The accuracy improved to 90.0% when the number of attention heads was increased to 8. Further increasing the attention heads to 16, we observed optimal network performance, with a classification accuracy of 90.3%. These results indicate that increasing the number of attention heads in an appropriate range can effectively enhance FDCPNet's ability to process 3D mesh data, as more heads contribute to better representation and emphasis of the shape features in mesh models.

However, an increase to 32 attention heads resulted in a decline in network performance, with the classification accuracy dropping to 89.1%. This change suggests that exceeding a certain threshold in the number of attention heads may lead to excessive information integration and redundancy, adversely affecting network performance. To summarize, choosing the right number of attention heads in FDM is critical for enhancing FDCPNet's performance in 3D shape classification. Future research could focus on identifying the optimal number of attention heads for various 3D shape representation tasks and devising strategies to mitigate the negative effects of excessive parameterization.

4.6 Robustness Evaluation

Table V Classification results of different numbers of faces on Manifold40.

Number of Faces	MeshNet[18]	FDCPNet (ours)
500	88.4%	**90.3%**
300	87.9%	90.1%
100	86.7%	89.8%
50	86.1%	89.7%

To validate the robustness of our proposed method, we tested the FDCPNet using mesh models with different numbers of faces, as depicted in Figure 3. The experimental results, detailed in Table V, show FDCPNet's adaptability and robustness to changes in the face numbers of mesh models. Specifically, with the 500 faces, FDCPNet achieved a classification accuracy of 90.3%. When the number of faces was reduced to 300, the accuracy decreased only slightly to 90.1%. Even with a further reduction to 50 faces, the accuracy moderately dropped to 89.7%, getting promising results.

In contrast, when the number of faces was reduced from 500 to 50, the performance of MeshNet declined by 2.3%. Whereas, FDCPNet's performance only decrease 0.6%. This disparity demonstrates that FDCPNet offers an effective solution that can reduce model complexity while still preserving an accurate representation of the shape.

Table VI Classification results of different numbers of training samples on Manifold40.

Samples of category	MeshCNN[17]	MeshNet[18]	FDCPNet
Full	79.3%	88.4%	**90.3%**
16	37.4%	77.8%	**79.1%**
4	24.5%	57.1%	**60.5%**

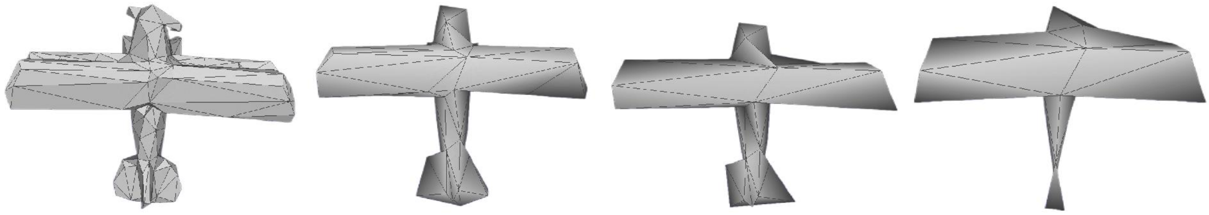

Figure 3 Examples of mesh models with different number of faces

Given the high acquisition cost of 3D models and the resulting scarcity of 3D mesh data in real-world applications, we focused on evaluating FDCPNet's classification performance under conditions of sparse training samples to assess its robustness. The experimental results reveal that FDCPNet maintains superior performance across different training sample sizes. As shown in Table VI, when trained with a full set of samples from each category, FDCPNet achieved its best performance, attaining a classification accuracy of 90.3%. This result notably surpassed MeshNet's 88.4% and MeshCNN's 79.3%. Moreover, even when the number of samples per category was reduced to just 16 and 4, FDCPNet exhibited a reasonable decline in performance, still outperforming the other methods. Notably, with only 4 samples per category—about 5% of the full dataset—the classification accuracy of FDCPNet increased by 36.0% and 3.4% over MeshCNN and MeshNet, respectively.

This robust performance is due to the effective collaboration of the FDM and the CPM in FDCPNet. The FDM introduces a multi-head attention mechanism that analyzes the input face features from multiple perspectives, enabling the network to capture more detailed and abundant local features and effectively extract key information from limited data. Meanwhile, the CPM calculates the global average features to effectively capture and represent the main structural features of the mesh models. By integrating local and global features, the CPM further enhances the network's accurate representation of shapes, ensuring that the network can precisely depict the model shape, even under conditions with limited numbers of faces or samples. These strategies enable FDCPNet to excel, compared to other methods, not only in standard 3D shape representation tasks, but also in demonstrating good stability and robustness under conditions of limited data availability.

In summary, FDCPNet demonstrates effective learning capabilities and robustness across datasets of various complexities and sizes. This makes FDCPNet a promising solution in environments with limited data, such as in the fields of cultural heritage preservation and remote sensing exploration for 3D shape representation. For instance, in situations where repeated scanning to acquire detailed 3D data is challenging due to the fragility or inaccessibility of artifacts, FDCPNet still maintains good performance.

4.7 Time And Space Complexity Comparison

Table VII Comparison results of time and space complexity for different methods in classification tasks.

Method	#params (M)	FLOPs / sample (M)
MVCNN[26]	60.0	62057
MeshNet[18]	4.25	509
FDCPNet (ours)	10.17	673

Note: #params" indicates the total number of parameters in the network, representing the space complexity; "FLOPs/sample" denotes the number of floating-point operations per input sample, reflecting the time complexity.

To delve deeper into the performance of FDCPNet in practical scenarios, we compare its time and space complexity with several methods in classification tasks. As shown in Table VII, compared to the baseline model MeshNet, FDCPNet has an increase of 5.92M params and 164M FLOPs. Despite the increased time and space complexity, the enhanced complexity of the network architecture allows FDCPNet to capture and represent key regional features of 3D models more effectively. Particularly, it shows improved robustness in data-limited application scenarios. To further optimize the potential of our method, future work will consider employing techniques such as knowledge distillation or more efficient convolutional operators to meet the demands of actual deployment.

5 Conclusion

This study proposes FDCPNet, an innovative approach that integrates FDM and CPM. FDCPNet not only identifies feature representations of various key areas but also effectively enriches key feature information, thereby offering a more comprehensive understanding of complex 3D shapes. Importantly, FDCPNet shows robustness, maintaining promising performance even under specific conditions with reduced mesh faces and limited training data. Future research will extend its validation to other baseline models. This broader evaluation will ascertain its versatility and effectiveness across a wider range of 3D data and model types, thereby solidifying its utility in the field of 3D shape representation.

References

1 Bahirat K, Lai C, Mcmahan R P, Prabhakaran B. Designing and evaluating a mesh simplification algorithm for virtual reality. ACM Transactions on Multimedia Computing, Communications, and Applications, 2018, 14(3s): 1–26.

2 Tong Q, Wei W, Zhang Y, Wang D. Survey on Hand-Based Haptic Interaction for Virtual Reality. IEEE Transactions on Haptics, 2023.

3 Xu W, Wang Y, Huang W, Duan Y. An efficient nonlinear mass-spring model for anatomical virtual reality. IEEE Transactions on Instrumentation and Measurement, 2022, 71: 1-10.

4 Mani M, Dorgan A J. A perspective on the state of aerospace computational fluid dynamics technology. Annual Review of Fluid Mechanics, 2023, 55: 431–457.

5 Sadeghalvaad M, Razavi S R, Sabbaghi S, Rasouli K. Heating performance of a large-scale line heater by adding synthesized carbon-nanodots to the heater bath fluid: CFD simulation and experimental study. Advanced Powder Technology, 2023, 34(3): 103960.

6 Liu Q, Nie W, Hua Y, Jia L, Li C, Ma H. Peng H. A study on the dust control effect of the dust extraction system in TBM construction tunnels based on CFD computer simulation technology. Advanced Powder Technology, 2019, 30(10): 2059–2075.

7 Liu X, Jia J, Liu C. Survey of lightweighting methods of huge 3D models for online Web3D visualization. Virtual Reality & Intelligent Hardware, 2023, 5(5): 395–406.

8 Munkberg J, Hasselgren J, Shen T, Gao J, Chen W, Evans A, Fidler S. Extracting triangular 3d models, materials, and lighting from images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 8280–8290.

9 Pan J, Han X, Chen W, Tang J, Jia K. Deep mesh reconstruction from single rgb images via topology modification networks. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 9964–9973.

10 Hang S, Subhransu K, Erik L M. Multiview convolutional neural networks for 3D shape recognition. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. 2015: 945–953.

11 Kalogerakis E, Averkiou M, Maji S, Chaudhuri S. 3D shape segmentation with projective convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017, 6630–6639.

12 Wu Z, Song S, Khosla A, Yu F, Zhang L, Tang X, Xiao J. 3D ShapeNets: A deep representation for volumetric shapes. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015, 1912–1920.

13 Sedaghat N, Zolfaghari M, Amiri E, et al. Orientation-boosted voxel nets for 3D object recognition. arXiv preprint arXiv:1604.03351, 2016.

14 Hanocka R, Fish N, Wang Z, Giryes R, Fleishman S, Cohenor D. ALIGNet: Partial-shape agnostic alignment via unsupervised learning. ACM Transactions on Graphics, 2019, 38(1): 1–14.

15 Graham B, Engelcke M, Der Maaten LV. 3D shape segmentation with projective convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018, 9224–9232.

16 Qi C R, Su H, Mo K, Guibas L J. Pointnet: Deep learning on point sets for 3d classification and segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 652–660.

17 Hanocka R, Hertz A, Fish N, Giryes R, Fleishman S, Cohen-Or D. Meshcnn: a network with an edge. ACM Transactions on Graphics, 2019, 38(4): 1–12.

18 Feng Y, Feng Y, You H, Zhao X, Gao Y. Meshnet: Mesh neural network for 3d shape representation. In: Proceedings of the AAAI Conference on Artificial Intelligence. 2019, 33(01): 8279–8286.

19 Qi C R, Yi L, Su H, Guibas L J. PointNet++ : Deep hierarchical feature learning on point sets in a metric space. In: Advances in neural information processing systems. 2017, 5100–5109.

20 Li Y, Bu R, Sun M, Wu W, Di X, Chen B. PointCNN: Convolution on X -transformed points. In: Advances in neural information processing systems. 2018, 820–830.

21 Hu S M, Liu Z N, Guo M H, Cai J X, Huang J, Mu T J, Martin R R. Subdivision-based mesh convolution networks. ACM Transactions on Graphics, 2022, 41(3): 1-16.

22 Wu Z R, Song S, Khosla A, Yu F S, Zhang L G, Tang X O, Xiao J X. 3d shapenets: A deep representation for volumetric shapes. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015, 1912–1920.

23 Maturana D, Scherer S. Voxnet: A 3d convolutional neural network for real-time object recognition. In: 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems. 2015, 922–928.

24 Li Y, Pirk S, Su H, Qi C R, Guibas L J. Fpnn: Field probing neural networks for 3d data. In: Advances in Neural Information Processing Systems, 2016, 29.

25 Chen D Y, Tian X P, Shen Y T, Ouhyoung M. On visual similarity based 3D model retrieval. Computer graphics forum. Oxford, UK: Blackwell Publishing, Inc, 2003, 22(3): 223–232.

26 Su H, Maji S, Kalogerakis E, et al. Multi-view convolutional neural networks for 3d shape recognition. In: Proceedings of the IEEE International Conference on Computer Vision. 2015, 945–953.

27 Johns E, Leutenegger S, Davison A J. Pairwise decomposition of image sequences for active multi-view recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016, 3813–3822.

28 Dai J, Fan R, Song Y, et al. MEAN: An attention-based approach for 3D mesh shape classification[J]. The Visual Computer, 2023: 1-14.

29 Kanezaki A, Matsushita Y, Nishida Y. Rotationnet: Joint object categorization and pose estimation using multiviews from unsupervised viewpoints. In Proceedings of the IEEE conference on computer vision and pattern recognition. 2018, 5010-5019.

30 Ma X, Qin C, You H, et al. Rethinking network design and local geometry in point cloud: A simple residual MLP framework. arXiv preprint arXiv:2202.07123, 2022.

31 Liang Y, He F, Zeng X, et al. Feature-preserved convolutional neural network for 3D mesh recognition. Applied Soft Computing, 2022, 128: 109500.

32 Ren M, Niu L, Fang Y. 3d-a-nets: 3d deep dense descriptor for volumetric shapes with adversarial networks[J]. arXiv preprint arXiv:1711.10108, 2017.

33 Hyeon-Woo N, Yu-Ji K, Heo B, et al. Scratching Visual Transformer's Back with Uniform Attention[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023: 5807-5818.

34 Milano F, Loquercio A, Rosinol A, et al. Primal-dual mesh convolutional neural networks[J]. Advances in Neural Information Processing Systems, 2020, 33: 952-963.