



ARTICLE

# MMNet: Integration Multi-Attention and Multi-Strategy Network for Feature Recognition

Shuai Ma<sup>1</sup>, Xiang Fang<sup>1,2,\*</sup> and Liya Han<sup>1</sup>

<sup>1</sup>Chengdu Aircraft Industrial (Group) Co., Ltd., Chengdu, China

<sup>2</sup>National College for Excellent Engineers, Beihang University, Beijing, China

\*Corresponding Author: Xiang Fang. Email: fangxiangbuaa@buaa.edu.cn

Received: 23 December 2025; Accepted: 03 April 2026; Published: 27 May 2026

**ABSTRACT:** Automated feature recognition (AFR) plays an important role in automated measurement path planning and metrological data processing in the manufacturing industry. Existing AFR methods face critical limitations, such as the loss of geometric-topological fidelity during Computer-aided design (CAD) model conversion and inadequate instance segmentation for dimensional metrology. To address these challenges, we propose an integrated multi-attention and multi-strategy network (MMNet) for feature recognition, which mainly includes the multi-attention geometric and attribute fusion module (MGAM) and the multi-strategy semantic and instance segmentation module (MSIM). Specifically, MGAM employs multi-attention mechanisms to synergize local geometric features with global attributes of the boundary representation (B-rep) to enhance recognition precision. Subsequently, MSIM integrates a graph neural network for face-level semantic segmentation with rule-based instance clustering to achieve robust feature recognition. Experimental results demonstrate that our proposed method outperforms other state-of-the-art methods on MFCAD, MFCAD++, and our real-world datasets.

**KEYWORDS:** Feature recognition; multi-attention; multi-strategy; instance segmentation

## 1 Introduction

In the domain of aerospace manufacturing, frame-beam components serve as critical elements within aircraft primary load-bearing structures, whose mechanical performance directly governs the overall structural integrity of airframes. For example, a representative frame-beam component commonly incorporates 200 to 400 sets of precision-distributed stiffening ribs and geometrically complex transitional fillets, whose dimensional accuracy critically influence stress distribution optimization and fatigue crack nucleation suppression. The contemporary industrial measurement paradigm confronts regulatory mandates that require feature-coverage detection of rib thicknesses and fillet transition zones, while conventional 2D drawing-based measurement planning exhibits severe efficiency constraints—A single component demands 4–6 labor-hours for manual feature annotation. Manual measurement planning has become increasingly inadequate to meet the real-time data requirements of modern pulsed assembly lines. Consequently, AFR enables intelligent measurement planning that eliminates human intervention, thereby revolutionizing quality assurance workflows to align with Industry 4.0 operational tempos.

Current AFR methods for mechanical components can be categorized into rule-based and learning-based paradigms [1]. Rule-based AFR methods [2] rely on explicitly defined geometric attributes and topological relations to establish deterministic recognition rules. These systems iteratively traverse B-rep

elements—including faces, edges, and loops—through graph traversal algorithms, subsequently applying pattern matching against predefined rules. Representative implementations encompass expert systems [3], subgraph isomorphism [4], volume decomposition [5] and hybrid strategies [6]. However, such approaches face inherent limitations of escalating combinatorial complexity in rule definition for intersecting features and difficulties in defining complete rules for all features. In contrast, learning-based AFR methods leverage machine learning or deep learning to capture latent space representations of features from datasets, eliminating reliance on rigid rule-based definitions while demonstrating enhanced accuracy and robustness compared to conventional approaches. However, learning-based AFR methods still face challenges regarding geometric and topological detail loss and limitations in instance segmentation. Conversion of native CAD B-rep models into triangular meshes [7], point clouds [8,9], or multiview images [10,11] discards geometric details and topological adjacencies. While existing approaches have successfully converted B-rep models into neural-network-friendly formats, their conversion pipelines inevitably lead to the loss of critical topological information, specifically the hierarchical structural connectivity and parametric continuity constraints inherent to B-rep representations. Moreover, current AFR methods predominantly focus on semantic segmentation, yet fail to address the critical industrial requirement of in-feature dimensional metrology. While AAGNet pioneers multi-task learning through parallel semantic and instance segmentation heads, its architecture struggles to overcome task imbalance. Semantic segmentation requires face-level precision, while instance segmentation demands combinatorial grouping logic. This divergence causes conflicting gradient updates during optimization. To address the aforementioned challenges, this paper proposes MMNet for feature recognition, which integrates local geometric features with global attributes of the B-rep boundary by the MGAM and achieves robust feature recognition through MSIM. In summary, our work has the following contributions:

- We propose MGAM to fuse local geometric features with global attributes of B-rep boundary to enhance recognition precision. The module firstly converts a B-rep model into a face adjacency graph, where graph nodes and edges respectively correspond to B-rep faces and edges. Then the local geometric features of the faces and edges from sampled points on UV parametric domains are combined with their global attributes by a cross attention and channel-wise attention mechanism.
- We propose MSIM which couples a graph neural network for face-level semantic segmentation with rule-based instance clustering to achieve robust feature recognition. The module firstly uses the graph neural network to classify B-rep face types. Then a rule-based clustering algorithm aggregates faces depending on the face-type classification results and predefined rules, realizing robust feature instance segmentation.
- On two public CAD model datasets, MFCAD [12], MFCAD++ [13] as well as our real-world dataset, we utilize two quantitative metrics, Accuracy and mean Intersection over Union (mIoU) to evaluate our performance of semantic segmentation. Meanwhile, our results are compared with six state-of-the-art methods, which includes PointNet++ [14], Hierarchical CADNet [13], UV-Net [15], MeshCNN [16], DGCNN [17] and AAGNet [18]. Experimental results show that our method outperforms these methods in the above metrics.

## 2 Related Work

### 2.1 Rule-Based Methods

**Logic rule and expert systems** [3,19–21]. Traditional feature recognition methods rely on manually defined logic rules and expert systems to identify potential features in B-rep models. The application of such approaches is well developed. However, these methods require exhaustive rule specification for all possible features. Complex models necessitate intricate rule hierarchies, leading to cumbersome systems [8].

**Graph-based methods** [2,4,6,22]. Such methods convert B-rep models into attribute adjacency graphs, where feature recognition becomes subgraph isomorphism detection. These methods exhaustively leverage the topological attributes inherent in B-rep models but they suffer from NP-hard computational complexity.

**Volume decomposition** [23–27]. Such methods segment B-rep models into primitive volumes via convex hull decomposition or Constructive Solid Geometry (CSG) tree reconstruction and then recombine them into recognizable volumes. These methods involve computationally intensive Boolean operations, rendering real-time processing infeasible for complex models.

**Hint-based methods** [24,28–30]. Geometric and topological hints are extracted from B-rep models to probabilistically infer features. Such methods can partially address the challenge of recognizing intersecting features. However, they suffer from incomplete coverage of diverse feature hints and reasoning rules.

## 2.2 Learning-Based Methods

To address the limitations of rule-based methods, learning-based approaches are emerging as advanced solutions. Recent research employs neural networks to recognize features directly from B-rep models. Unlike rule-based methods that rely on manually defined feature patterns, these neural networks autonomously learn feature representations from labeled B-rep datasets. Since B-rep data cannot be directly processed by neural networks, current deep learning methods can be classified based on their converted data formats.

**Voxel network.** Voxel-based networks [31–34] discretize B-rep models into regular voxel grids for feature recognition through deep learning. Researchers like Zhang et al. [31] and Peddireddy et al. [32] utilized 3D convolutional neural networks (CNNs) to learn geometric and topological patterns from voxelized B-rep data. Ning et al. [35] proposed a hybrid approach. Firstly, they perform minimum subgraph segmentation to partition the model into disjoint sub-regions, and then each sub-region is voxelized and classified via 3D CNNs. There are some limitations within voxelization. Voxel discretization erodes precise parametric surfaces, inducing geometric information loss. Face adjacency and edge connectivity are poorly preserved, causing topology disruption. Furthermore, the high memory footprint for complex models and massive grids contributes to computational inefficiency.

**Multi-view image network.** This approach [10,11,36,37] projects B-rep models into 2D images from multiple viewpoints and processes them using image-based neural networks. Shi et al. [10] proposed MsvNet, a two-stage workflow. First, individual features are isolated via geometric saliency detection. Then, orthographic projections are generated for each feature. SsdNet [38], an end-to-end network, integrates feature localization and classification using a Single Shot Multibox Detector (SSD). Although these methods can partially address the challenge of recognizing intersecting features, the conversion from 3D B-rep models to 2D images introduces geometric and topological distortions, while mapping 2D recognition results back to 3D faces presents significant challenges.

**Point cloud network.** Recent advancements in deep learning have driven innovative methods for feature recognition using 3D point cloud data. Ma et al. [39] developed a hybrid framework integrating convolutional neural networks (CNNs) with PointNet to automate feature extraction from sparse point cloud representations. Building upon this foundation, Yao et al. [40] introduced a hierarchical neural network architecture that addresses overlapping feature complexities through an unsupervised segmentation strategy for multi-feature decomposition, coupled with PointNet++ [14] for individual feature classification. Further advancing detection efficiency, Zhang et al. [8] proposed Associatively Segmenting and Identifying Network (ASIN), a pioneering single-stage framework enabling concurrent semantic segmentation, instance segmentation, and base face detection within intersecting feature scenarios, drawing inspiration from

contemporary point cloud instance segmentation techniques [41,42]. In parallel, Colligan et al. [43] established a standardized pipeline for generating annotated point clouds from B-rep models while preserving bidirectional geometric correspondences, addressing critical data preparation challenges. Expanding into assembly analysis, Worner et al. [44] implemented a dual-stream PointNet architecture that simultaneously identifies form features and infers their mechanical interconnections as joint relationships. Despite these computational advantages, point cloud-based approaches exhibit inherent limitations. The potential under-sampling of fine geometric details during point cloud generation [45], particularly for micro-features, and the absence of explicit topological relationships between fundamental B-rep elements (vertices, edges, faces), constrain high-level geometric reasoning capabilities. These constraints necessitate careful consideration when deploying such methods in precision-critical manufacturing applications.

**Mesh network** [7,16,46,47]. Mesh-based approaches predominantly operate on discretized mesh representations converted from B-rep models. Jia et al. [7] proposed a cascaded framework integrating MeshCNN's topology-aware filters with Faster R-CNN's region proposal mechanism. Böhm et al. [47] developed a graph neural network (GNN) that directly learns hierarchical geometric patterns from raw mesh graphs through edge-conditioned convolution. Although mesh-based representations provide simplified structural regularity, their conversion pipelines compromise geometric fidelity.

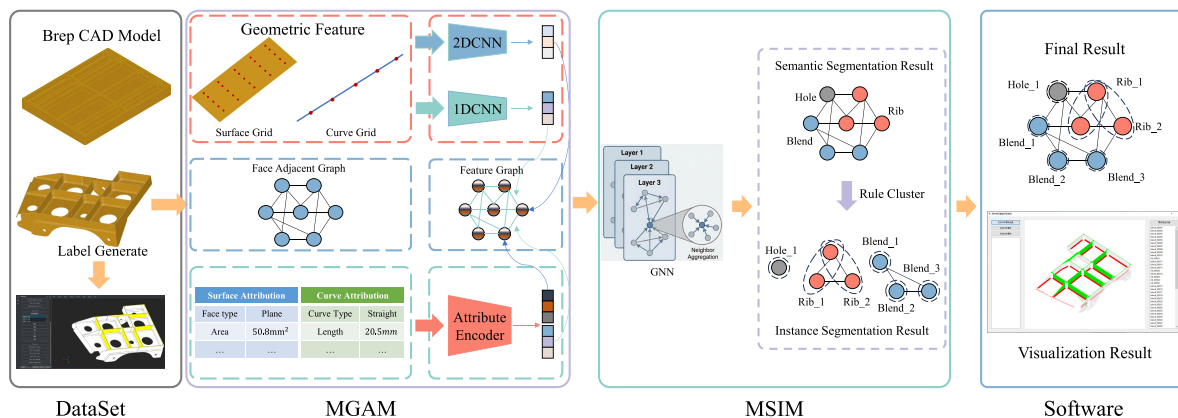
**Graph network** [48–52]. Graph-based representations have become prevalent in processing B-rep models due to their inherent structural compatibility with topological graphs. Cao et al. [12] constructed a dataset of synthetic CAD models annotated with diverse manufacturing features, where each face carries feature labels, and implemented feature recognition through graph neural networks for face-level classification. UV-Net [15] introduced uniform sampling in the UV parametric domain of faces and edges to extract geometric signatures via convolutional neural networks. BrepNet [45] further developed specialized convolutional kernels to enable information propagation between adjacent faces and edges. Colligan et al. [13] proposed Hierarchical CADNet by integrating B-rep graph representations with mesh-based geometric embeddings, achieving enhanced recognition of manufacturing features. Wang et al. [53] adopted a hybrid strategy combining subgraph extraction algorithms for feature isolation with GNN-based classification. Building upon UV-Net's sampling paradigm, Wu et al. [18] incorporated both local sampled-point geometry and global face and edge attributes to construct comprehensive graph embeddings for feature recognition. Park et al. [48] implemented attention-enhanced GNNs to capture design intent semantics, while Li et al. [54] established edge adjacency graph representations directly from B-rep topology for GNN feature recognition.

While the aforementioned learning-based methods have significantly advanced the AFR field, two critical gaps remain: the superficial fusion of geometric and topological properties, and the architectural conflicts in concurrent semantic and instance segmentation tasks. Specifically, state-of-the-art graph methods extract rich features but lack adaptive fusion mechanisms to weigh the relative importance of continuous shapes vs. discrete attributes. Meanwhile, attempts at unified instance segmentation struggle with the inherent divergence between pixel/face-level classification and combinatorial grouping logic. Our proposed MMNet directly tackles these limitations. By integrating the MGAM for highly interpretable feature alignment and the MSIM to mathematically decouple the segmentation tasks, MMNet achieves superior geometric fidelity and deployment robustness to existing paradigms.

### 3 Method

B-rep is the predominant method in contemporary CAD modeling for describing solid geometry, which defines the geometric shape and topological structure of an object through explicit representation of its surface boundaries. A B-rep model encapsulates both geometric and topological information of each entity in a CAD part. **Geometric attributes** include positional coordinates, type specifications and parametric

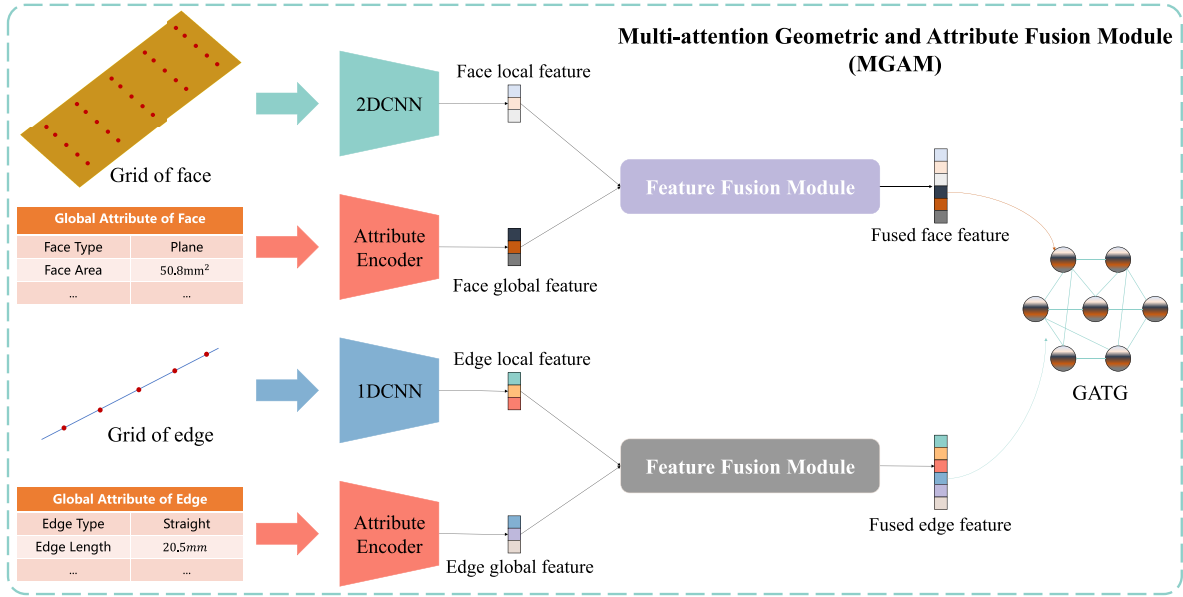
definitions for faces and edges. **Topological relationships** formalize connectivity relations among the entities. B-rep can precisely describe a component but it is difficult to be learned in a neural network due to its complexity and non-uniform structure. To address the challenge, we must first convert the B-rep to a learning-friendly format. As mentioned in related works, many methods have converted the B-rep into point clouds, meshes, multi-view images, which can't completely encode the topological and geometric information of B-rep, leading to a disadvantage for feature recognition. So Geometry-attributed Topological Graph (GATG), an adjacency graph encoding the geometric and topological information of B-rep, is proposed in Fig. 1. GATG fundamentally constitutes a face adjacency graph, where each node corresponds to a face in the B-rep model. An undirected edge exists between two adjacent face nodes, explicitly representing their shared boundary edge in the original B-rep structure. This representation effectively encapsulates topological connectivity between adjacent faces in B-rep. The parametric representations of faces and edges in B-rep models, while mathematically precise, present fundamental incompatibilities with neural network processing due to their continuous analytic formulations. Current discretization approaches (triangulated meshes, voxel grids, or unstructured point clouds) inevitably degrade topological fidelity. Drawing inspiration from UV-Net [15], we discretize B-rep faces within their UV parametric domains. As CAD surfaces are 2-manifolds, their parametric representations inherently reside in a 2D coordinate system. By sampling structured grids within this UV space, each grid point maps to a 3D point on the surface, enabling extraction of local geometric information such as coordinates and normals. Similarly, we apply a similar sampling strategy to each edge of the B-rep. To enhance the geometric expressiveness of B-rep representations, inspired by the attributed adjacency graph [4], we enrich both faces and edges with global attributes in the GATG through the following descriptors. Face attributes include surface types (plane, cylinder, sphere, non-uniform rational B-spline (NURBS) surface, etc.), areas and centroid coordinates. Edge attributes include edge type (line, circle, spline, etc.), length, convexity and dihedral angle between adjacent faces.



**Figure 1:** The architecture of our proposed MMNet operates through three sequential stages. First, the B-rep model is converted into a face adjacency graph, where nodes represent topological faces and edges correspond to shared boundary edges. Each face and edge undergoes parametric UV-grid sampling to generate structured points, which are then encoded through convolutional networks and fused with attribute information to enrich the graph's geometric-topological representations. This enhanced face adjacency graph is subsequently processed by graph neural networks for semantic segmentation of face feature types. Finally, a rule-based clustering algorithm performs instance segmentation by aggregating topologically coherent faces into feature instances.

### 3.1 MGAM: Multi-Attention Geometric and Attribute Fusion Module

In the B-rep model, the geometric representation of each face integrates both local and global attributes. Local face features are derived by processing regular parametric grid samples through a 2D convolutional neural network. Global face attributes are encoded via channel-attention-weighted fully-connected layers to prioritize critical surface properties. Similarly, edge geometry is characterized by local and global features. Local edge features are extracted from parameterized curve samples using a 1D CNN. Global edge attributes enhance representation via descriptors processed by the same attention mechanism. Then two cross-attention-based feature fusion modules are proposed to fuse the global and local features for faces and edges. Finally, the fused face features and edge features are bound to the nodes and edges in the GATG. The workflow is shown in Fig. 2.



**Figure 2:** MGAM integrates local geometric features and global attributes of B-rep models through a hierarchical multi-attention framework, generating an information-enhanced face adjacency graph that preserves native CAD semantics.

**Geometric local attribute encoder.** The geometric local attribute encoder processes the UV grid points of B-rep faces and edges. For each face, its geometric local attributes  $f_l \in \mathbb{R}^{n \times n \times 7}$  denote a structured grid sampled from the UV parametric domain, encoding per-point 3D coordinates, normals, and a trimming status flag (1 for valid, 0 for trimmed regions). The geometric local attributes of each face are processed by a 2D convolutional neural network. The 2D CNN processes the local attributes of surfaces,  $f_l \in \mathbb{R}^{n \times n \times 7}$ , to yield the dense feature representation  $H_l$ . It consists of sequential layers:  $Conv2D(7, 32, 3) \rightarrow Conv2D(32, 64, 3) \rightarrow Conv2D(64, 128, 3) \rightarrow Pool(1, 1) \rightarrow FC(128, 128)$ , where  $Conv2D(i, o, k)$  is a 2D convolutional layer with  $i$  input channels,  $o$  output channels and a kernel size of  $k$ ,  $Pool(m, n)$  is an adaptive average pooling layer that yields an  $m \times n$  feature map and  $FC(i, o)$  is a fully connected layer that encodes an  $i - D$  vector to an  $o - D$  vector. For each edge, its geometric attributes  $e_l \in \mathbb{R}^{n \times 6}$  denote sampled points along the 1D parametric domain, capturing 3D coordinates and tangential vectors. Similarly, the 1D CNN processes the local attributes of edges,  $e_l \in \mathbb{R}^{n \times 6}$ . It consists of sequential layers:  $Conv1D(6, 32, 3) \rightarrow Conv1D(32, 64, 3) \rightarrow Conv1D(64, 128, 3) \rightarrow Pool(1, 1) \rightarrow FC(128, 128)$ . In summary, the embeddings of the face geometric local attribute  $H_l$  and edge geometric local attributes  $h_l$  can be formulated as:

$$\begin{aligned} H_l &= \mathbf{F}_{2D}(f_l) \in \mathbb{R}^{128} \\ h_l &= \mathbf{F}_{1D}(e_l) \in \mathbb{R}^{128} \end{aligned} \quad (1)$$

**Geometric global attribute encoder.** For each face, the geometric global attributes  $f_g \in \mathbb{R}^{n_f}$  denote high-level geometric descriptors such as surface types (plane, cylinder, sphere, NURBS surface, etc.), areas and centroid coordinates. For each edge, the geometric global attributes  $e_g \in \mathbb{R}^{n_e}$  denotes the edge-specific properties including edge types (line, circle, spline, etc.), length, convexity and dihedral angle between adjacent faces. The geometric global attributes of faces and edges contain both categorical labels (face/edge type indicators with binary 0/1 values) and continuous geometric data (coordinate information tied to B-rep model positioning). To ensure numerical stability during network training, we implement selective channel normalization. Let the geometric global attributes of faces in a B-rep model be denoted by  $\mathbf{F}_g \in \mathbb{R}^{C \times N}$ , where  $C$  denotes the number of attribute channels and  $N$  denotes the number of faces. Define the index set  $S = \{c_1, c_2, \dots, c_k\} \subseteq \{1, 2, \dots, C\}$  for coordinate-related channels. The per-channel statistics are computed as follows:

$$\begin{aligned} \mu_c &= \frac{1}{N} \sum_{i=1}^N \mathbf{F}_g^{(c,i)} \\ \sigma_c^2 &= \frac{1}{N} \sum_{i=1}^N (\mathbf{F}_g^{(c,i)} - \mu_c)^2 \end{aligned} \quad (2)$$

Subsequently, a normalization transformation is applied using learnable scale and shift parameters,  $\gamma_c$  and  $\beta_c$ :

$$\hat{\mathbf{F}}_g^{(c,i)} = \gamma_c \frac{\mathbf{F}_g^{(c,i)} - \mu_c}{\sqrt{\sigma_c^2 + \varepsilon}} + \beta_c, \forall c \in S \quad (3)$$

Similarly, the edge geometric global attributes will be normalized as  $\hat{\mathbf{E}}_g$ . Then the attributes will be encoded by the fully connected network (FCN), formulated as:

$$\begin{aligned} \mathbf{H}_g &= \text{FCN}(\hat{\mathbf{F}}_g) \\ \mathbf{h}_g &= \text{FCN}(\hat{\mathbf{E}}_g) \end{aligned} \quad (4)$$

As the varying significance of feature channels in global attributes and the inherent redundancy in face and edge type indicators, all surface can be represented as NURBS surface, a channel attention mechanism is introduced to dynamically recalibrate feature channel importance. The channel-wise weight can be formulated as:

$$\begin{aligned} s_f &= \sigma(\mathbf{W}_2 \cdot \text{ReLU}(\mathbf{W}_1 \cdot \mathbf{H}_g)) \in [0, 1] \\ s_e &= \sigma(\mathbf{W}_4 \cdot \text{ReLU}(\mathbf{W}_3 \cdot \mathbf{h}_g)) \in [0, 1] \end{aligned} \quad (5)$$

where  $\sigma(\cdot)$  denotes the Sigmoid function,  $\text{ReLU}(\cdot)$  denotes the activation layer,  $\mathbf{W}_i$  denotes the learnable matrix. Then apply channel-wise attention weights on  $\mathbf{H}_g$  and  $\mathbf{h}_g$ :

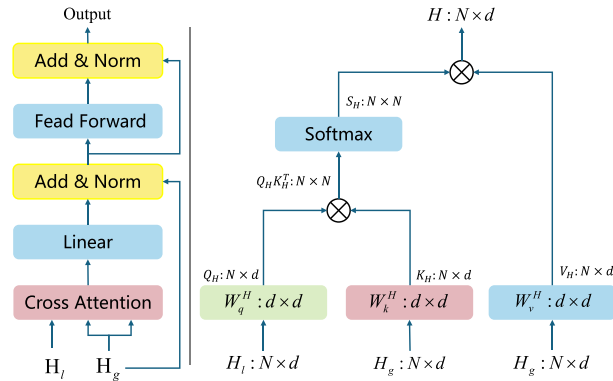
$$\begin{aligned} \mathbf{H}_g &= s_f \odot \mathbf{H}_g \\ \mathbf{h}_g &= s_e \odot \mathbf{h}_g \end{aligned} \quad (6)$$

The design of the channel-wise attention mechanism is specifically targeted at the heterogeneous nature of B-rep global attributes. These attributes combine discrete categorical indicators with continuous geometric metrics. In the context of AFR, different attributes carry varying discriminative weights depending on the target feature. For example, identifying a transitional fillet relies heavily on the cylindrical or NURBS surface type attribute, while the absolute surface area is less critical. Furthermore, B-rep structures inherently contain mathematical redundancies: all surfaces can be theoretically represented as NURBS surface. The channel-wise attention acts as an adaptive geometric feature selector, computing weights  $s_f$  and  $s_e$  to dynamically amplify the most determinative attributes for a specific face or edge while suppressing redundant representations, thereby stabilizing the learning process.

**Attention-based global and local feature fusion.** Each face and edge in the B-rep model is thus represented by two complementary feature components. Local features,  $\mathbf{H}_l$  for faces and  $\mathbf{h}_l$  for edges, are derived through convolutional processing of regular sample points from parametric grid sampling, capturing coordinate and normal vector information. Global features,  $\mathbf{H}_g$  for faces and  $\mathbf{h}_g$  for edges, are obtained by processing global attributes through fully connected layers followed by channel-attention-based feature weighting. The faces and edges of a B-rep model are represented by capturing both global features,  $\mathbf{H}_g, \mathbf{h}_g$  and local features  $\mathbf{H}_l, \mathbf{h}_l$ . To effectively integrate these complementary representations, a cross-attention-based feature fusion module [55] is proposed in Fig. 3. This module leverages the inherent correlation between global and local features. For instance, if a face is planar, the normals of all its sampled points remain consistent. The cross-attention mechanism dynamically adjusts the weights of global and local features by computing their similarity scores, implicitly aligning them to enhance fusion performance. This approach bridges the gap between macro-geometric properties and micro-local sampling patterns, significantly improving downstream tasks like feature recognition. The fused features can be formulated as:

$$\begin{aligned} \mathbf{H} &= \text{Attn}(\mathbf{Q}_H, \mathbf{K}_H, \mathbf{V}_H) = \text{Softmax}\left(\frac{\mathbf{Q}_H \mathbf{K}_H^T}{\sqrt{d_H}}\right) \mathbf{V}_H \\ \mathbf{h} &= \text{Attn}(\mathbf{Q}_h, \mathbf{K}_h, \mathbf{V}_h) = \text{Softmax}\left(\frac{\mathbf{Q}_h \mathbf{K}_h^T}{\sqrt{d_h}}\right) \mathbf{V}_h \end{aligned} \quad (7)$$

where  $\mathbf{Q}_H = \mathbf{H}_l \mathbf{W}_q^H$ ,  $\mathbf{K}_H = \mathbf{H}_g \mathbf{W}_k^H$ ,  $\mathbf{V}_H = \mathbf{H}_g \mathbf{W}_v^H$  are the query, key and values matrices of the attention operation for face feature while  $\mathbf{Q}_h = \mathbf{h}_l \mathbf{W}_q^H$ ,  $\mathbf{K}_h = \mathbf{h}_g \mathbf{W}_k^H$ ,  $\mathbf{V}_h = \mathbf{h}_g \mathbf{W}_v^H$  for the edges.  $\mathbf{W}_q^H, \mathbf{W}_k^H, \mathbf{W}_v^H \dots$  are the weight matrices of the trainable linear projection layers. Finally, the face embeddings  $H$  are bound with the nodes, while the edge embeddings  $h$  are bound with the edges in GATG.

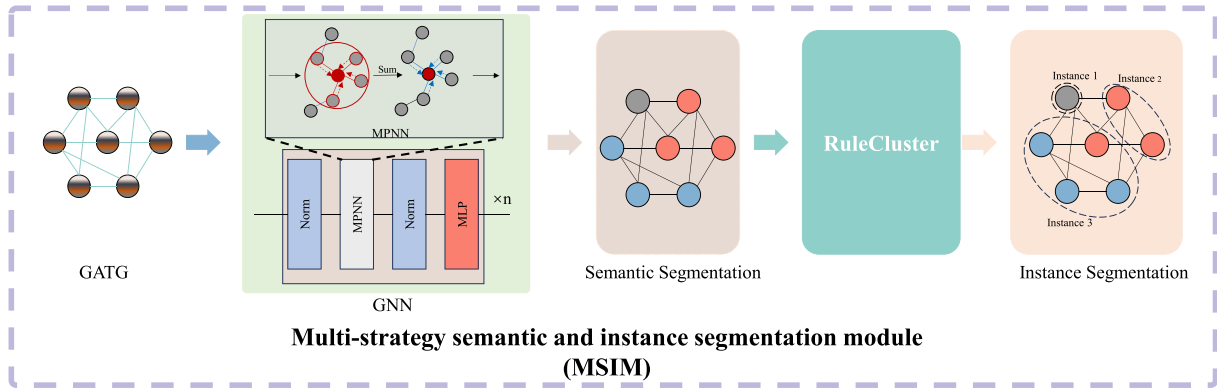


**Figure 3:** Left: The structure of cross-attention based feature fusion module. Right: The computation graph of cross-attention.

The deployment of cross-attention is deeply rooted in the fundamental characteristic of B-rep data: the decoupling of topological boundaries and underlying geometric surfaces. Here, local features ( $\mathbf{H}_l, \mathbf{h}_l$ ) encode the micro-level geometric topography through spatial coordinates and normal vectors, whereas global features ( $\mathbf{H}_g, \mathbf{h}_g$ ) dictate the macro-level geometric semantics. The cross-attention module functions as a macro-micro alignment mechanism. The physical intuition is that global attributes should guide the aggregation of local geometric details, and vice versa. For instance, if a face is planar, the normals of all its sampled points remain consistent. In this scenario, the cross-attention mechanism dynamically computes high similarity scores between the global planar embedding and the local normal descriptors, effectively forcing the network to attend to this geometric consistency. By bridging the micro-geometric variations and macro-topological definitions, this attention design explicitly prevents the loss of geometric-topological fidelity, making the fused representations highly robust for the downstream instance segmentation tasks.

### 3.2 MSIM: Multi-Strategy Semantic and Instance Segmentation Module

Graph neural networks constitute a specialized deep learning framework for processing graph-structured data, employing iterative message passing between adjacent nodes to progressively refine feature representations at nodes, edges and the whole graph. While traditionally applied in recommender systems and molecular property prediction, this paper pioneers GNN deployment for CAD model feature recognition through our proposed MMNet, which innovatively integrates B-rep geometric attributes with topological information to accomplish instance segmentation. The MSIM consists of a semantic segmentation network and a clustering algorithm, as illustrated in Fig. 4.



**Figure 4:** MSIM operates through dual phases. GNN-based semantic feature learning that classifies each B-rep face into predefined categories followed by knowledge-guided face cluster employing geometric-topological rules to achieve instance segmentation.

**Semantic segmentation.** The semantic segmentation module employs a GNN and a node-wise classifier. The GNN can learn the topological connectivity of the GATG and refines node features by propagating and fusing geometric and topological information from adjacent nodes and edges. The classifier can semantically segments each node of the GATG, which corresponds to an individual face in the B-rep model. The GNN comprises  $n$  blocks, each of which can be formulated as follows:

$$\begin{aligned} H'_i &= H_i + MPNN(Norm(H_i)) \\ H_{i+1} &= H'_i + W\sigma(Norm(H'_i)) \end{aligned} \quad (8)$$

where  $H_i$  denotes the output of each block, MPNN denotes the message passing neural network,  $W$  denotes a Multi-Layer Perceptron (MLP), Norm indicates layer normalization, and  $\sigma$  denotes a non-linear activation function.

The MPNN constitutes the core computational mechanism, operating through three rigorously defined stages. **Message:** Each node  $v_i$  computes a message  $m_{ij}$  for every neighbor node  $v_j$ , where  $m_{ij} = \phi(H_i^l, h_e^{ij})$ . Here,  $H_i^l$  denotes the current embedding of node  $v_i$  at layer  $l$  and  $h_e^{ij}$  denotes the edge embedding between  $v_i$  and  $v_j$ . The message function  $\phi(\cdot)$  is a learnable neural layer. **Aggregation:** Each node  $v_i$  leverages a permutation-invariant function to aggregate all the messages passed from its neighbors  $N(i)$ , formulated as  $a_i^l = \oplus m_{ji}$ , where  $\oplus$  denotes a permutation-invariant operator (sum, mean or max), ensuring the invariance to the neighbor orders. **Update:** The node embedding is updated through all the messages from its neighbors and itself. Finally, the MPNN can be formulated as follows:

$$H_i^{l+1} = M(H_i^l \otimes A(H_i^l, m_{ij})) \quad (9)$$

where  $H_i^l$  denotes the output of  $l$ -th layer of MPNN,  $A(\cdot)$  denotes the aggregation operator,  $\otimes$  denotes the concatenation operators. Through the topology encoder, we can get the feature map  $\mathbf{H} \in \mathbb{R}^{n \times 128}$  of all the nodes. In our implementation, the GNN is configured with  $L = 4$  sequential message-passing blocks to ensure a sufficient receptive field across the face adjacency graph. Specifically, the learnable message function  $\phi(\cdot)$  is implemented as a two-layer MLP with intermediate ReLU activations. For the permutation-invariant aggregation operator  $\oplus$ , we employ a max-pooling function. Empirical evaluations indicate that max-pooling is highly effective in isolating the most salient and discriminative geometric and topological signals from neighboring B-rep entities, outperforming mean or sum operations for feature recognition tasks. Subsequently, the feature map is passed into the classifier to achieve semantic segmentation. The classifier consists of sequential linear layers designed to classify geometric and semantic category of each B-rep face (rib side face, rib top or fillet). It generates an  $N \times C$  probability matrix indicating the confidence of each face belonging to a specific class, where  $C$  denotes the number of predefined feature types. The semantic segmentation head is supervised via a cross-entropy loss between predicted probabilities and the ground-truth labels, which can be formulated as follows:

$$\mathcal{L} = -\frac{1}{N} \sum_{n=1}^N \sum_{i=1}^C y_i^{(n)} \cdot \log(\hat{y}_i^{(n)}) \quad (10)$$

where  $N$  denotes the number of faces,  $C$  represents the number of predefined semantic categories,  $y_i^{(n)} \in \{0, 1\}$  is the ground-truth indicator for class  $i$  of the  $n$ -th face.  $\hat{y}_i^{(n)}$  is the predicted probability of class  $i$  for the  $n_{th}$  face.

**Rule-based cluster.** Existing feature recognition methods predominantly perform semantic segmentation on individual faces of mechanical parts. However, in practical quality measurement scenarios, determining whether feature dimensions meet specifications requires establishing feature instances and analyzing the geometric relationships between the constituent faces within each instance. For example, when inspecting rib features, engineers must measure the rib thickness (the distance between two side faces) and the fillet radius (the curvature of transitional surfaces). Consequently, relying solely on face-level semantic segmentation proves insufficient for such metrological requirements. Our proposed rule-based clustering algorithm extends semantic segmentation to instance segmentation through predefined geometric-constraint principles. Specifically, the clustering rules for ribs are defined as follows:

Principle 1: The rib consists of one top face and two side faces.

Principle 2: Antiparallel normals. The normals of the two side faces within one rib should satisfy:

$$\mathbf{n}_1 \cdot \mathbf{n}_2 \leq -\tau_{\text{parallel}}, \quad \tau_{\text{parallel}} \in [0.9, 1.0] \quad (11)$$

where  $\mathbf{n}_1$  and  $\mathbf{n}_2$  denotes the normals of the two faces in candidates. This ensures opposing orientation of adjacent faces.

Principle 3: Projective coincidence. The orthogonal projections of side faces within a rib onto their normal-aligned planes must satisfy:

$$\begin{cases} \frac{|\Omega_1 \cap \Omega_2|}{\min(|\Omega_1|, |\Omega_2|)} \geq \tau_{\text{overlap}} \\ \|\mathbf{c}_1 - \mathbf{c}_2\|_{\text{parallel}} \leq \varepsilon_{\text{offset}} \end{cases} \quad (12)$$

where  $\Omega_i$  denotes the projected region of face  $i$ ,  $\mathbf{c}_i$  represents the centroid coordinate of  $\Omega_i$ ,  $\tau_{\text{overlap}}$  indicates the overlap ratio threshold and offset defines the positional tolerance parallel to the faces.

Principle 4: The two side faces of a rib share a common top face that is topologically adjacent to both and each side wall has an adjacent fillet.

Following the aforementioned principles, all candidate pairs of rib side faces within the semantic segmentation results are filtered. Successfully matched candidate pairs are automatically assigned to the same rib instance. Subsequently, the top face and fillets are detected by the topological relations. The robustness and generalization of this rule-based clustering strategy are underpinned by two key factors. First, robustness is explicitly engineered through tolerance-controlled thresholds, which effectively absorb minor CAD conversion artifacts and geometric imperfections without disrupting the grouping logic. Second, while the demonstrated principles focus on specific features, the fundamental methodology exhibits strong generalization. By utilizing universal B-rep geometric invariants rather than dataset-specific data distributions, this strategy can be systematically extended to construct robust recognition logic for diverse mechanical instances across various industrial applications. For example, extending this framework to recognize instances of machining holes or slots would merely require formulating rules based on the coaxiality of adjacent cylindrical faces or the orthogonality of intersecting planar boundaries. This highly interpretable, tolerance-controlled approach ensures that the framework can be robustly adapted to diverse mechanical instances across broader engineering metrology applications.

## 4 Experiments

### 4.1 Datasets

We utilize three datasets in our experiments: MFCAD [12], MFCAD++ [13] and our real-world dataset. The CAD models in these datasets are represented in B-rep with labels for each face.

**MFCAD**, a synthetic dataset consists of 15488 synthetic mechanical parts with 16 distinct feature categories. It is partitioned into training, test, and validation sets following a 60/20/20% split ratio.

**MFCAD++**, a synthetic dataset serves as an enhanced version of MFCAD, expanding the feature categories from 16 to 24 types. This upgraded dataset contains 59,655 synthetic CAD models with both planar and non-planar surfaces, incorporating more complex intersecting geometric features. The dataset is partitioned into training (70%), test (15%), and validation (15%) subset.

**Real-world Dataset.** While the MFCAD and MFCAD++ datasets utilize synthetic CAD parts that exhibit discrepancies from real industrial components, we propose a domain-specific aerospace CAD dataset

to enable reliable feature recognition for geometric measurement. This real-world benchmark contains 400 part models annotated with measurement-critical features, including ribs and fillets.

To establish the credibility and robustness of this dataset, we elaborate on its statistical distribution and annotation protocols. Regarding model complexity, the dataset encompasses a diverse spectrum of topological structures, ranging from local connection brackets to highly complex primary load-bearing frame-beam components. The number of B-rep faces per model spans from 41 to 2555. The number of feature instances (ribs and fillets) per model averages at 40, adequately reflecting the high-density feature distribution typical in modern aerospace manufacturing. The dataset is partitioned into training (60%), test (20%), and validation (20%) subset.

#### 4.2 Evaluation Metrics

The proposed MSIM comprises two cascaded modules, a semantic segmentation network that classifies each topological face in B-rep models into predefined manufacturing feature categories and a rule-based instance clustering module that aggregates adjacent faces into feature-level instance through predefined constraints. The segmentation accuracy of the first module is quantitatively validated using two established metrics, per-face classification accuracy and mIoU, which can be formulated as:

$$Acc = \frac{|\mathcal{F}_{correct}|}{|\mathcal{F}_{total}|} \times 100\% \quad (13)$$

$$mIoU = \frac{1}{C} \sum_{c=1}^C \frac{L_c \cap P_c}{L_c \cup P_c}$$

where  $\mathcal{F}_{correct}$  denotes correctly classified faces,  $\mathcal{F}_{total}$  denotes the total faces,  $C$  denotes the number of classes,  $L_c$  denotes the faces of class  $c$  in ground truth labels,  $P_c$  denotes the faces of class  $c$  in prediction result.

#### 4.3 Results on MFCAD and MFCAD++

The results on the MFCAD and MFCAD++ test dataset are shown in Table 1. The result of PointNet++ [14] and Hierarchical CADNet [13] are reported in [13] while UV-Net [15], MeshCNN [16], DGCNN [17] are reported in [15].

**Table 1:** Performance comparison on MFCAD and MFCAD++ datasets.

Method	MFCAD		MFCAD++	
	Acc (%)	mIoU (%)	Acc (%)	mIoU (%)
PointNet++	91.35	–	85.88	–
DGCNN	82.50	67.70	85.98	–
Hierarchical CADNet	99.90	–	97.37	–
UV-Net	99.95	99.87	–	–
MeshCNN	99.89	99.70	–	–
AAGNet	99.99	99.99	99.24	98.64
MMNet(Ours)	<b>99.99</b>	<b>99.99</b>	<b>99.38</b>	<b>98.86</b>

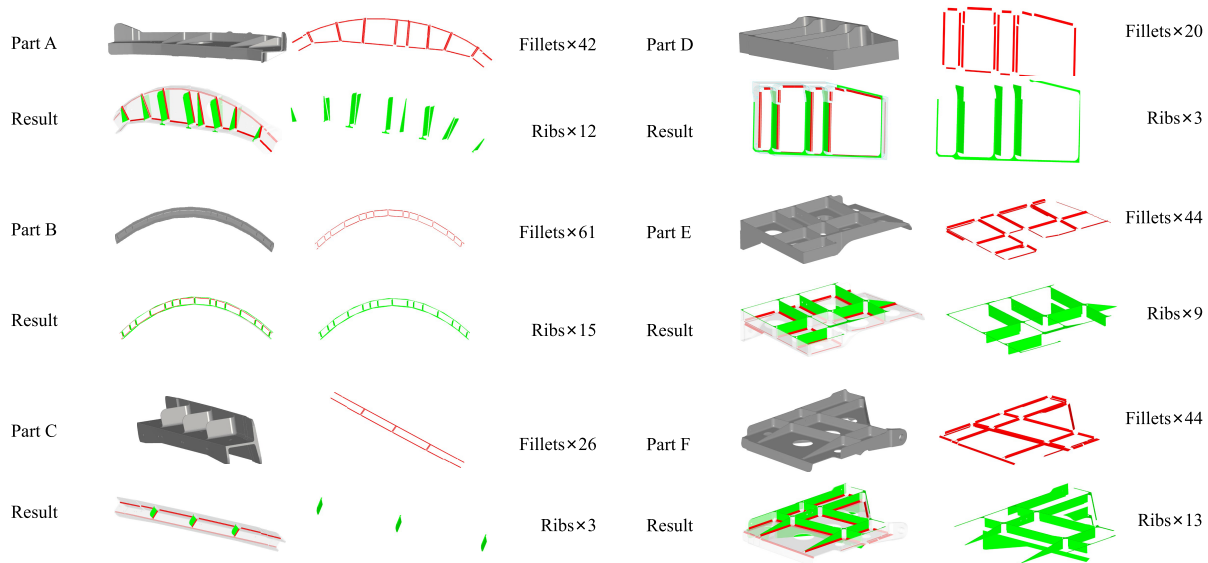
Note: The best results are highlighted in **bold**.

The results demonstrate that MMNet outperforms previous methods on the two metrics. MMNet achieves a highest accuracy than the state-of-the-art results. Furthermore, as shown in Table 1, PointNet++

and DGCNN, that convert B-rep model into point clouds and MeshCNN, that converts B-rep model into meshes achieve inferior semantic segmentation accuracy compared to graph-based networks on the MFCAD and MFCAD++ datasets. The primary cause resides is that while processing discrete surface samples, point-cloud and mesh-based methods exhibit fundamental limitations in extracting persistent topological features critical for rare category differentiation, concurrently incurring irreversible geometric fidelity degradation due to inherent structural discretization artifacts.

#### 4.4 Results on Real-World Dataset

The proposed MMNet focuses on automatic detection of ribs and fillets, two predominant measurement features in aerospace structural components. Accordingly, the annotation schema of our Real-world dataset exclusively includes these two features. We compare the semantic segmentation accuracy with Hierarchical CADNet, UV-Net, AAGNet and demonstrate the visual results of instance segmentation. We set the  $\tau_{\text{overlap}} = 0.9$ ,  $\epsilon_{\text{offset}} = 0.5$  in the rule-based cluster. The instance segmentation results of six CAD models in real-world dataset are shown in Fig. 5. The feature recognition results for each part are visualized through a quad-view layout. The top-left is the original B-rep model of the aerospace part. The bottom-left is the semantic segmentation results, where the green surfaces are the identified ribs and the red surfaces are the fillets. The top-right is the instance segmentation results for the fillets.



**Figure 5:** The visual results of feature recognition for six parts in real-world dataset.

The bottom-right is the instance segmentation results of ribs. The quantitative result of semantic and instance segmentation on Real-world dataset is shown in Table 2. The results demonstrate that our proposed MMNet outperforms other graph-based networks on semantic segmentation on this real-world dataset and the rule-based clustering algorithm successfully groups faces belonging to identical feature instances through predefined heuristics, thereby enabling precise feature instance segmentation.

**Table 2:** Performance comparison on real-world dataset.

Method	Semantic Segmentation	
	Acc (%)	mIoU (%)
Hierarchical CADNet	98.29	98.12
UV-Net	99.60	99.52
AAGNet	99.77	99.09
MMNet (Ours)	<b>99.95</b>	<b>99.86</b>

Note: The best results are highlighted in **bold**.

#### 4.5 Ablation Study

To systematically evaluate the individual contributions of local and global features to semantic segmentation accuracy, we conducted ablation studies. The result is shown in Table 3. Comparative ablation results demonstrate that this channel-wise reweighting strategy improves face classification accuracy compared to the baseline. The results conclusively demonstrate that removing any core component degrades MMNet’s semantic segmentation accuracy on the Real-world dataset. The most severe performance drop occurs when ablating face global topological attributes. This phenomenon likely stems from the dataset’s inherent feature characteristics. The rib features are dominated by large parallel planar faces where global planar type flags provide critical discriminative cues. Beyond the numerical verification of each component in Table 3, it is essential to analyze the underlying physical causes of these degradations. The most severe drop (−8.88% in mIoU) occurs when ablating face global topological attributes. This is fundamentally because structural features like ribs are predominantly composed of large parallel planar faces; without explicitly feeding the global “planar” indicator, the network confuses these faces with other flat-like topological entities. Conversely, removing local features ( $H_l, h_l$ ) results in a −1.86% mIoU drop. The physical cause here is boundary ambiguity. In precision manufacturing, distinguishing a transitional fillet with a large bending radius from an adjacent flat wall relies intrinsically on the micro-level coordinate variations and normal vector shifts captured by the local UV-grid points. Without local features, the network suffers from geometric blindness at topological boundaries. Although ablating the cross-attention fusion module yields a minor numerical drop (−0.01% in Acc, −0.11% in mIoU), its qualitative role is indispensable. This minimal variance is due to a performance ceiling effect: our baseline accuracy already exceeds 99%, as the dataset is statistically dominated by unambiguous topological faces where simple feature concatenation suffices. Beyond the quantitative metrics, the cross-attention fusion module provides critical qualitative contributions, particularly in complex geometric scenarios. For instance, when dealing with micro-fillets bridging highly curved adjacent walls or intersecting features with subtle boundary transitions, simple feature concatenation often struggles to resolve geometric ambiguity, leading to localized semantic noise. In these hard cases, the cross-attention mechanism adaptively aligns the micro-level geometric topography with the macro-level topological definitions, effectively acting as a semantic stabilizer. This alignment is highly critical, as it prevents misclassifications in ambiguous boundary regions, thereby guaranteeing the high-fidelity semantic priors required by the downstream zero-tolerance rule-based clustering.

**Table 3:** Ablation study on key components (%).

Method	Acc	mIoU
<b>Full</b>	<b>99.95</b>	<b>99.86</b>
w/o Face global features	91.07 (−8.88)	90.94 (−8.92)
w/o Edge global features	99.56 (−0.39)	98.35 (−1.51)
w/o Local features	99.25 (−0.7)	98.00 (−1.86)
w/o Channel-wise reweight	99.60 (−0.35)	98.68 (−1.18)
w/o Cross-attention fusion	99.94 (−0.01)	99.75 (−0.11)

Note: The best results are highlighted in **bold**.

#### 4.6 Computational Efficiency Analysis

To thoroughly evaluate the deployment feasibility of the proposed MMNet in real-world industrial measurement scenarios, we conducted a comparative analysis of computational efficiency. The evaluation metrics encompass the number of network parameters (Params), GPU memory usage during inference, and the average single-model processing time (Inference Time). All efficiency benchmark experiments were executed on an identical hardware environment equipped with a single NVIDIA RTX 3090 GPU and an Intel Core i7-12700k CPU.

As presented in [Table 4](#), the proposed MMNet achieves an average inference time of 0.6 s per B-rep model. While the integration of the MGAM inherently introduces a slight increase in network parameters and GPU memory consumption compared to lighter baseline architectures, this computational overhead is strictly bounded and highly manageable.

**Table 4:** Computational efficiency comparison of different methods.

Method	Params (M)	Inference Time (s)
UV-Net	1.23	15.3
Hierarchical CADNet	6.60	7.2
AAGNet	<b>0.44</b>	<b>0.36</b>
MMNet (Ours)	0.73	0.6

Note: The best results are highlighted in **bold**.

Crucially, in contemporary industrial metrology, conventional manual feature annotation for a representative aerospace frame-beam component rigorously demands 4–6 labor-hours. In stark contrast, MMNet accomplishes automated semantic and instance segmentation in mere milliseconds. This extraordinary acceleration decisively confirms that the proposed method not only meets but substantially exceeds the real-time processing requirements of Industry 4.0 automated inspection lines, demonstrating robust industrial deployment feasibility.

## 5 Conclusions

In this paper, we propose MMNet, an integrated multi-attention and multi-strategy network for feature recognition. The framework converts B-rep models into face adjacency graphs for face-level semantic segmentation (via MGAM) and leverages geometric-topological constraints for robust rule-based instance

clustering (via MSIM). Experimental results on MFCAD, MFCAD++, and our real-world dataset demonstrate that MMNet achieves state-of-the-art performance, particularly in extracting measurement-critical features like ribs and fillets on aerospace components.

Despite these promising results, two main limitations warrant future investigation. First, the framework's performance intrinsically relies on the topological integrity of the input B-rep models. Second, while MSIM's core paradigm exhibits strong scalability to other manufacturing features, manually formulating explicit clustering rules for an exhaustive feature library remains a labor-intensive process.

To address these challenges, our future work will explore robust graph completion algorithms to handle defective CAD models. Furthermore, to enhance the framework's generality and autonomy, we plan to investigate learning-based 3D instance segmentation architectures (e.g., metric-learning or transformer-based mask generation) and leverage large language models (LLMs) or data-driven approaches to automatically synthesize clustering rules.

**Acknowledgement:** The authors sincerely thank Chengdu Aircraft Industrial (Group) Co., Ltd. for providing the real-world CAD models and technical expertise for dataset construction.

**Funding Statement:** This research was funded by National Science and Technology Major Project, grant number 2025ZD1601305.

**Author Contributions:** The authors confirm contribution to the paper as follows: conceptualization, Shuai Ma; methodology, Xiang Fang; resources, Liya Han; writing—original draft preparation, Shuai Ma, Xiang Fang; writing—review and editing, Liya Han. All authors reviewed and approved the final version of the manuscript.

**Availability of Data and Materials:** The MFCAD are openly available at [<https://github.com/hducg/MFCAD>]. The MFCAD++ are openly available at [[https://gitlab.com/qub\\_femg/machine-learning/mfcad2-dataset](https://gitlab.com/qub_femg/machine-learning/mfcad2-dataset)].

**Ethics Approval:** Not applicable. This study does not involve human or animal subjects.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Shi Y, Zhang Y, Xia K, Harik R. A critical review of feature recognition techniques. *Comput Aided Des Appl.* 2020;17(5):861–99.
2. Xu T, Li J, Chen Z. Automatic machining feature recognition based on MBD and process semantics. *Comput Ind.* 2022;142(3):103736. doi:10.1016/j.compind.2022.103736.
3. Henderson MR, Anderson DC. Computer recognition and extraction of form features: a CAD/CAM link. *Comput Ind.* 1984;5(4):329–39. doi:10.1016/0166-3615(84)90056-3.
4. Joshi S, Chang TC. Graph-based heuristics for recognition of machined features from a 3D solid model. *Comput Aided Des.* 1988;20(2):58–66. doi:10.1016/0010-4485(88)90050-4.
5. Kim YS. Volumetric feature recognition using convex decomposition. *Manuf Res Technol.* 1994;20(2):39–63. doi:10.1016/b978-0-444-81600-9.50008-0.
6. Gao S, Shah JJ. Automatic recognition of interacting machining features based on minimal condition subgraph. *Comput Aided Des.* 1998;30(9):727–39. doi:10.1016/s0010-4485(98)00033-5.
7. Jia JL, Zhang SW, Cao YR, Qi XL, Zhu W. Machining feature recognition method based on improved mesh neural network. *Iran J Sci Technol Trans Mech Eng.* 2023;47(4):2045–58. doi:10.1007/s40997-023-00610-8.
8. Zhang H, Zhang S, Zhang Y, Liang J, Wang Z. Machining feature recognition based on a novel multi-task deep learning network. *Robot Comput Integr Manuf.* 2022;77(4):102369. doi:10.1016/j.rcim.2022.102369.
9. Zhang H, Wang W, Zhang S, Wang Z, Zhang Y, Zhou J, et al. Point cloud self-supervised learning for machining feature recognition. *J Manuf Syst.* 2024;77(5):78–95. doi:10.1016/j.jmsy.2024.08.029.

10. Shi P, Qi Q, Qin Y, Scott PJ, Jiang X. A novel learning-based feature recognition method using multiple sectional view representation. *J Intell Manuf.* 2020;31:1291–309.
11. Shi P, Qi Q, Qin Y, Scott PJ, Jiang X. Highly interacting machining feature recognition via small sample learning. *Robot Comput Integr Manuf.* 2022;73:102260.
12. Cao W, Robinson T, Hua Y, Boussuge F, Colligan AR, Pan W. Graph representation of 3D CAD models for machining feature recognition with deep learning. In: *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference.* New York, NY, USA: American Society of Mechanical Engineers; 2020. Vol. 84003.
13. Colligan AR, Robinson TT, Nolan DC, Hua Y, Cao W. Hierarchical CADNet: learning from B-reps for machining feature recognition. *Comput Aided Des.* 2022;147:103226.
14. Qi CR, Yi L, Su H, Guibas LJ. Pointnet++: deep hierarchical feature learning on point sets in a metric space. *Adv Neural Inf Process Syst.* 2017;30:1–72.
15. Jayaraman PK, Sanghi A, Lambourne JG, Willis KD, Davies T, Shayani H, et al. UV-Net: learning from boundary representations. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2021 Jun 20–25; Nashville, TN, USA.* p. 11703–12.
16. Hanocka R, Hertz A, Fish N, Giryes R, Fleishman S, Cohen-Or D. MeshCNN: a network with an edge. *ACM Trans Graph.* 2019;38(4):1–12.
17. Phan AV, Le Nguyen M, Nguyen YLH, Bui LT. DGCNN: a convolutional neural network over large-scale labeled graphs. *Neural Netw.* 2018;108:533–43.
18. Wu H, Lei R, Peng Y, Gao L. AAGNet: a graph neural network towards multi-task machining feature recognition. *Robot Comput Integr Manuf.* 2024;86:102661.
19. Donaldson IA, Corney JR. Rule-based feature recognition for 2.5D machined components. *Int J Comput Integr Manuf.* 1993;6(1–2):51–64. doi:10.1080/09511929308944555.
20. Al-wswasi M, Ivanov A. A novel and smart interactive feature recognition system for rotational parts using a STEP file. *Int J Adv Manuf Technol.* 2019;104:261–84. doi:10.1007/s00170-019-03849-1.
21. Zhang Y, Wang F, Chen T. User-defined machining feature recognition based on semantic reasoning for B-rep models. *Comput-Aided Des Appl.* 2023;20(4):763–85. doi:10.14733/cadaps.2023.763-785.
22. Li Y, Ding Y, Mou W, Guo H. Feature recognition technology for aircraft structural parts based on a holistic attribute adjacency graph. *Proc Inst Mech Eng Part B J Eng Manuf.* 2010;224(2):271–8. doi:10.1243/09544054jem1634.
23. Sakurai H. Volume decomposition and feature recognition: part 1—polyhedral objects. *Comput Aided Des.* 1995;27(11):833–43.
24. Nau DS, Gupta SK, Kramer TR, Regli WC, Zhang G. Development of machining alternatives, based on MRSEVs. In: *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference.* New York, NY, USA: American Society of Mechanical Engineers; 1993. Vol. 97645, p. 47–57.
25. Ferreira J, Hinduja S. Convex hull-based feature-recognition method for 2.5D components. *Comput Aided Des.* 1990;22(1):41–9.
26. Woo Y, Sakurai H. Recognition of maximal features by volume decomposition. *Comput Aided Des.* 2002;34(3):195–207. doi:10.1016/s0010-4485(01)00080-x.
27. Woo Y. Fast cell-based decomposition and applications to solid modeling. *Comput-Aided Des.* 2003;35(11):969–77. doi:10.1016/s0010-4485(02)00144-6.
28. Vandenbrande JH, Requicha AA. Spatial reasoning for the automatic recognition of machinable features in solid models. *IEEE Trans Pattern Anal Mach Intell.* 1993;15(12):1269–85. doi:10.1109/34.250845.
29. Han J, Requicha AA. Integration of feature based design and feature recognition. *Comput Aided Des.* 1997;29(5):393–403. doi:10.1016/s0010-4485(96)00079-6.
30. Li H, Huang Y, Sun Y, Chen L. Hint-based generic shape feature recognition from three-dimensional B-rep models. *Adv Mech Eng.* 2015;7(4):1687814015582082. doi:10.1177/1687814015582082.
31. Zhang Z, Jaiswal P, Rai R. FeatureNet: machining feature recognition based on 3D convolution neural network. *Comput Aided Des.* 2018;101(6):12–22. doi:10.1016/j.cad.2018.03.006.

32. Peddireddy D, Fu X, Shankar A, Wang H, Joung BG, Aggarwal V, et al. Identifying manufacturability and machining processes using deep 3D convolutional networks. *J Manuf Process*. 2021;64:1336–48. doi:10.1016/j.jmapro.2021.02.034.
33. Vatandoust F, Yan X, Rosen D, Melkote SN. Manufacturing feature recognition with a sparse voxel-based convolutional neural network. *J Comput Inf Sci Eng*. 2025;25(3):031002. doi:10.1115/1.4067334.
34. Huang X, Wang L, Chen M, Huang D, He B, Jiang C. Automated recognition of machining features in mechanical components utilizing 3D convolutional neural networks. In: *Proceedings of the 2024 8th International Conference on Electrical, Mechanical and Computer Engineering (ICEMCE)*; 2024 Oct 24–27; Xi'an, China. p. 1101–5.
35. Ning F, Shi Y, Cai M, Xu W. Part machining feature recognition based on a deep learning method. *J Intell Manuf*. 2023;34(2):809–21. doi:10.1007/s10845-021-01827-7.
36. Shi P, Qi Q, Qin Y, Scott PJ, Jiang X. Intersecting machining feature localization and recognition via single shot multibox detector. *IEEE Trans Ind Inform*. 2020;17(5):3292–302. doi:10.1109/tii.2020.3030620.
37. Deng J, Sun C, Liu X, Du G, Jiang L, Yang X. High-frequency workpiece image recognition model based on hybrid attention mechanism. *Appl Sci*. 2025;15(1):94. doi:10.3390/app15010094.
38. Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu CY, et al. SSD: single shot multibox detector. In: *Proceedings of the Computer Vision—ECCV 2016: 14th European Conference*; 2016 Oct 11–14; Amsterdam, The Netherlands. p. 21–37.
39. Ma Y, Zhang Y, Luo X. Automatic recognition of machining features based on point cloud data using convolution neural networks. In: *Proceedings of the 2019 International Conference on Artificial Intelligence and Computer Science*; 2019 Jul 12–13; Wuhan, China. p. 229–35.
40. Yao X, Wang D, Yu T, Luan C, Fu J. A machining feature recognition approach based on hierarchical neural network for multi-feature point cloud models. *J Intell Manuf*. 2023;34(6):2599–610.
41. Wang W, Yu R, Huang Q, Neumann U. SGPN: similarity group proposal network for 3D point cloud instance segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; 2018 Jun 18–23; Salt Lake City, UT, USA. p. 2569–78.
42. Wang X, Liu S, Shen X, Shen C, Jia J. Associatively segmenting instances and semantics in point clouds. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*; 2019 Jun 15–20; Long Beach, CA, USA. p. 4096–105.
43. Colligan AR, Robinson TT, Nolan DC, Hua Y. Point cloud dataset creation for machine learning on cad models. *Comput Aided Des Appl*. 2021;18(4):760–71. doi:10.14733/cadaps.2021.760-771.
44. Worner JM, Brovkina D, Riedel O. Feature recognition for graph-based assembly product representation using machine learning. In: *Proceedings of the 2021 21st International Conference on Control, Automation and Systems (ICCAS)*; 2021 Oct 12–15; Jeju, Republic of Korea. p. 629–35.
45. Lambourne JG, Willis KD, Jayaraman PK, Sanghi A, Meltzer P, Shayani H. BRepNet: a topological message passing system for solid models. In: *Proceedings of the IEEE/CVF Conference on Computer Vision And Pattern Recognition*; 2021 Jun 20–25; Nashville, TN, USA. p. 12773–82.
46. 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 Jan 28–Feb 1; Honolulu, HI, USA. p. 8279–86.
47. Böhm SA, Neumayer M, Zagar BL, Riß F, Kortüm C, Knoll A. Advancing 3D mesh analysis: a graph learning approach for intersecting 3D geometry classification. In: *International Conference on Pattern Recognition*. Berlin/Heidelberg, Germany: Springer; 2024. p. 143–59.
48. Park JH, Lim S, Yeo C, Joung YK, Mun D. DFGAT for recognizing design features from a B-rep model for mechanical parts. *Robot Comput Integr Manuf*. 2025;93(8):102938. doi:10.1016/j.rcim.2024.102938.
49. Su Z, Bai J, Ma X, Qin F. UVConv: a UV grid-based convolutional neural network for classification and segmentation of 3D CAD models. *Comput Aided Geom Des*. 2026;125:102516.
50. Van Camp T, Benoit D. A lightweight graph representation for scalable deep learning on 3D CAD data. In: *Proceedings of the ORBEL 40: 40th Annual Conference of the Belgian Operational Research Society*; 2026 Feb 5–6; Leuven Belgium. p. 69–70.

51. Wang F, Cao Y, Wang H, Yin S, Hu X, Cui X. AAGATNet: an attentive graph network for machining feature recognition on free-form surfaces. *Comput-Aided Des.* 2026;193:104041.
52. Lai W, Xu T, Niu B, Wang H. GraphBRep: explicit graph diffusion of B-Rep topology for efficient CAD generation. *J Comput Des Eng.* 2026;13(1):259–74. doi:10.1093/jcde/qwaf136.
53. Wang P, Yang WA, You Y. A hybrid learning framework for manufacturing feature recognition using graph neural networks. *J Manuf Process.* 2023;85(2):387–404. doi:10.1016/j.jmapro.2022.10.075.
54. Li Y, Li E, Lenover M, Mann S, Bedi S. Edge adjacency graph and neural network architecture for machining feature recognition. *Int J Adv Manuf Technol.* 2025;136(2):897–908. doi:10.1007/s00170-024-14903-y.
55. Gheini M, Ren X, May J. Cross-attention is all you need: adapting pretrained transformers for machine translation. arXiv:2104.08771. 2021.