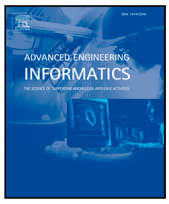
Advanced Engineering Informatics 62 (2024) 102900

Contents lists available at ScienceDirect

Advanced Engineering Informatics

journal homepage: www.elsevier.com/locate/aei



Full length article

Spatially embedded transformer: A point cloud deep learning model for aero-engine coaxiality prediction based on virtual measurement

Wu Tianyi [a,1], Shang Ke [b,1], Jin Xin [b,d,*], Zhang Zhijing [b], Li Chaojiang [b], Steven Wang [a], Liu Jun [c,*]

[a] Department of Mechanical Engineering, City University of Hong Kong, 999077, Hong Kong, China
[b] Department of Mechanical Engineering, Beijing Institute of Technology, Beijing, 100081, China
[c] Department of Data and Systems Engineering, The University of Hong Kong, 999077, Hong Kong, China
[d] Yangtze Delta Region Academy of Beijing Institute of Technology, Jiaxing, 314000, China

ARTICLE INFO

Keywords:
Aero-engine assembly
Coaxiality prediction
Point cloud transformer
Virtual measurement

ABSTRACT

Coaxiality is a critical indicator of assembly accuracy in aero-engines, directly impacting the device's operational performance and lifespan. Due to the enclosed nature of the aero-engine casing system, measuring the coaxiality of assembled components presents significant challenges. This paper introduces a novel deep learning architecture, the spatially embedded transformer (SETrans), designed to predict coaxiality from unassembled part data by correlating it with the contact surface points of assembled components. Additionally, a virtual measurement model is developed to collect micron-scale point cloud data, facilitating the fine-tuning of the deep learning model. The SETrans utilizes the transformer's capability for global information aggregation to process point cloud inputs, capturing the comprehensive relationships across assembled surfaces. A newly designed module, the spatial bias, integrates distance and angular information between neighboring point clouds into the transformer block, enhancing the model's ability to capture fine-grained local details. Experimental validation is conducted using two distinct datasets representing different assembly scenarios: the aero-engine casing, sampled using contact-based coordinate measuring machines, and the rotor, sampled using non-contact optical gaging products. These specific sampling methods test the generalizability of the SETrans across diverse measurement techniques. Comparative analysis with other point cloud deep learning benchmarks shows that the proposed approach achieves top prediction accuracies of 93.65% and 94.31% with a coaxiality precision of 0.01 mm across different data domains. These results confirm the effectiveness of the SETrans and demonstrate its adaptability to real-world assembly conditions involving various components.

1. Introduction

Coaxiality is a fundamental measure of assembly accuracy in aero-engines, influencing key operational aspects such as failure rates, reliability, dynamic balance, and vibration. Errors in coaxiality, which often arise during the assembly process, directly affect the engine's performance [1]. The assembly of casings and rotors, as illustrated in Fig. 1, plays a pivotal role in ensuring engine coaxiality [2]. Accurate predictions of coaxiality between these components prior to assembly are essential for maintaining the engine's operational stability.

The intricate and enclosed structure of aero-engines makes the direct measurement of coaxiality for internal components, such as shaft holes, impractical. Various methods have been explored to predict coaxiality. Notably, digital twin methods [3] and virtual reality methods [4] have been employed in aero-engine assembly. However, these digital models are typically low-precision and require substantial computational resources to ensure visual and operational effectiveness. The dimensional chain calculation method simplifies coaxiality prediction by transforming three-dimensional problems into one-dimensional issues, gaining popularity for its straightforward approach [5]. The Jacobian-Torsor model [6] combines the Jacobian matrix with tolerance zone representations, reducing prediction errors that arise from the non-rigidity of assembled parts. Zhang et al. [7] further enhances this approach by integrating error propagation with homogeneous transformation matrices (HTM) to analyze linear and rotational motion effects on coaxiality errors. These methods, represented by HTM and Jacobian-Torsor, significantly streamline complex assembly process calculations and minimize redundant computations. Nevertheless, they often overlook the effects of non-uniform machining errors, or geometric distribution errors (GDE), of contact surfaces on assembly, which can compromise the realism of the contact model and the precision of

* Corresponding authors.
E-mail addresses: goldking@bit.edu.cn (J. Xin), dr.jun.liu@hku.hk (L. Jun).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.aei.2024.102900
Received 26 July 2024; Received in revised form 27 September 2024; Accepted 19 October 2024
Available online 30 October 2024

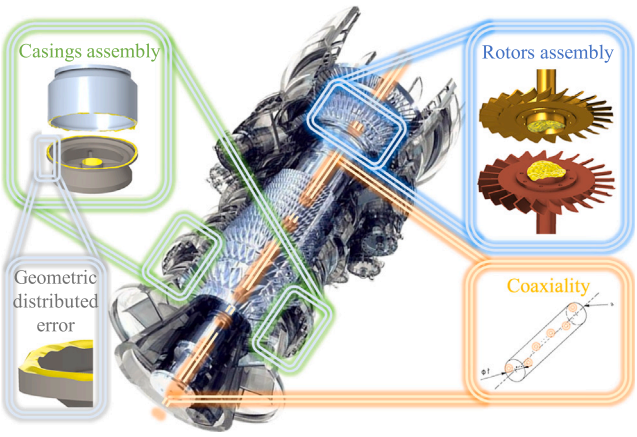


Fig. 1. Coaxiality in aero-engine assembly.

the assembly model. To account for GDE in assembly accuracy predictions, skin model shapes are constructed using either dense point clouds or triangular meshes [8,9]. However, the demand for dense (over 1000 points) and high-precision (measured in microns or sub-microns) point clouds for accurate coaxiality prediction is cost-prohibitive on aero-engine production lines. In response, this paper introduces a virtual measurement method that merges real unassembled part measurements with virtual coaxiality calculations. This approach enhances point density through GDE reconstruction while preserving the accuracy of initial high-precision point clouds, thus reducing the cost of individual surface measurements. It establishes a correlation between point clouds and coaxiality within assembly datasets through virtual assembly and the Monte Carlo method.

Virtual measurement methods have made strides in predicting coaxiality but often necessitate frequent adjustments to model parameters based on contact states during each prediction cycle. Moreover, introducing a new pair of surfaces requires remodeling, significantly increasing labor and computational costs. To bypass these drawbacks and enable automatic and efficient coaxiality prediction, establishing a direct link between aero-engine point clouds and assembly coaxiality is essential. Deep learning, a framework capable of managing nonlinear equations through millions of parameters, effectively establishes end-to-end mapping relationships between discrete data sets and corresponding physical labels [10]. With point cloud deep learning, getting the prediction coaxiality directly from a trained network and input point cloud is possible. According to the input data format, point cloud deep learning methods are categorized into three types: view-based, volumetric, and point-based [11]. As for the view-based methods, the point cloud data are transferred to images via different angles projection [12–14]. The neural network processes the images to obtain the objects' features. Avoiding the information loss inherent in converting 3D points to 2D images, the volumetric method voxelizes the unstructured point cloud to a structured 3D grid. Subsequently, 3D convolutional neural networks (CNN) are applied to the integrated data format for 3D shape classification [15–17]. However, both methods involve transforming point cloud data, leading to high computational and memory costs, and they may not scale efficiently with dense 3D data. To end this, PointNet [18], the pioneer of the point-based method, directly processes the original 3D point cloud and realizes the label prediction. The disorder and unstructured point clouds are processed by symmetric multi-layer perception (MLP) to guarantee invariance under permutations and rotation. The dominance of the point-based method is established due to the high computational efficiency and accuracy, which benefits from the use of original point clouds. Based on PointNet, PointNet++ [19] incorporates hierarchical MLP structures to capture local point cloud features efficiently and robustly. PointCNN [20] redefines the convolutional operator to process irregular point clouds, enabling the application of CNNs to raw point clouds. InterpCNN [21] interpolates the CNN weights with the point cloud coordinates, achieving the integration of physical properties and deep networks. DGCNN [22] feeds the point set into the graph convolutions and densely connects the local features. DeepGCNs [23] expands the complexity of the point cloud baseline and analyzes the relationship between the network depth and 3D scene understanding. Despite the efficiency in handling raw point cloud data, point-based deep learning methods face challenges in prediction performance due to the potential loss of global information. Current point cloud deep learning architecture, including the MLP and CNN, primarily focuses on aggregating neighboring points. However, due to the inherent disorder within point clouds, each point's position crucially impacts the object's pose and, consequently, network performance. Thus, the global relationships among different points cannot be overlooked, emphasizing the need for methods that can effectively capture both local and global structures within point clouds.

Recently, the transformer has become dominant in natural language processing [24,25] and image analysis [26,27] due to the global information extraction capability via the self-attention mechanism [28]. It outperforms traditional MLP and CNN and realizes top performance across a variety of downstream tasks. The self-attention mechanism's invariance to the permutation and cardinality of inputs [29] aligns well with the nature of point clouds as sets in 3D space, making transformers an ideal choice for point cloud deep learning. In this work, the transformer block is integrated into coaxiality prediction tasks, introducing a new point cloud deep learning backbone: SETrans. SETrans enables end-to-end coaxiality prediction and is evaluated alongside established point cloud deep learning baselines. To verify the generalization of SETrans, two datasets are derived from distinct aero-engine components: the casing and the rotor, sourced from simulated real parts. These datasets are created using different sampling techniques, specifically a contact-based coordinate measuring machine (CMM) and non-contact optical gaging products (OGP), to establish unique data domains for each set. The virtual measurement method is employed to augment the raw data, ensuring it meets the training specifications of the deep learning model while also reducing the costs associated with data collection. SETrans is then tested across these domains and benchmarked against other point cloud deep learning methods to validate its effectiveness and generalizability.

The rest of this paper is organized as follows: Section 2 provides a review of point cloud transformers. Section 3 briefly introduces the total architecture of the end-to-end coaxiality prediction system. Section 4 elaborates on virtual measurement method and dataset construction processes. Section 5 shows the total architecture of SETrans and the sub-module of this new backbone. Section 6 compares SETrans with other point cloud deep learning baselines and tests it in two different assembly parts. Section 7 draws the main conclusions.

2. Review of point cloud transformer

Due to the invariance to permutation and cardinality of input elements, transformer family models are well-suited for point cloud inputs [30]. Point2sequence [31] was a pioneer in applying global attention to entire point clouds, achieving effective global feature extraction. However, it lacks applicability for large-scale 3D scene understanding. The PCT [32] hybridizes the self-attention design of the Vision Transformer [33] with point cloud data formats, creating a pure point cloud transformer architecture. This approach introduces the $l_1 Norm$ to the attention feature map, ensuring the model's feasibility on large datasets. Similarly, the contemporaneous point cloud transformer aims to handle size-varying inputs. The Set Transformer [34] models interactions among input elements for better feature aggregation, while PATs [35] introduce shuffle attention to achieve a larger receptive field. These innovations enhance prediction performance for large-size point

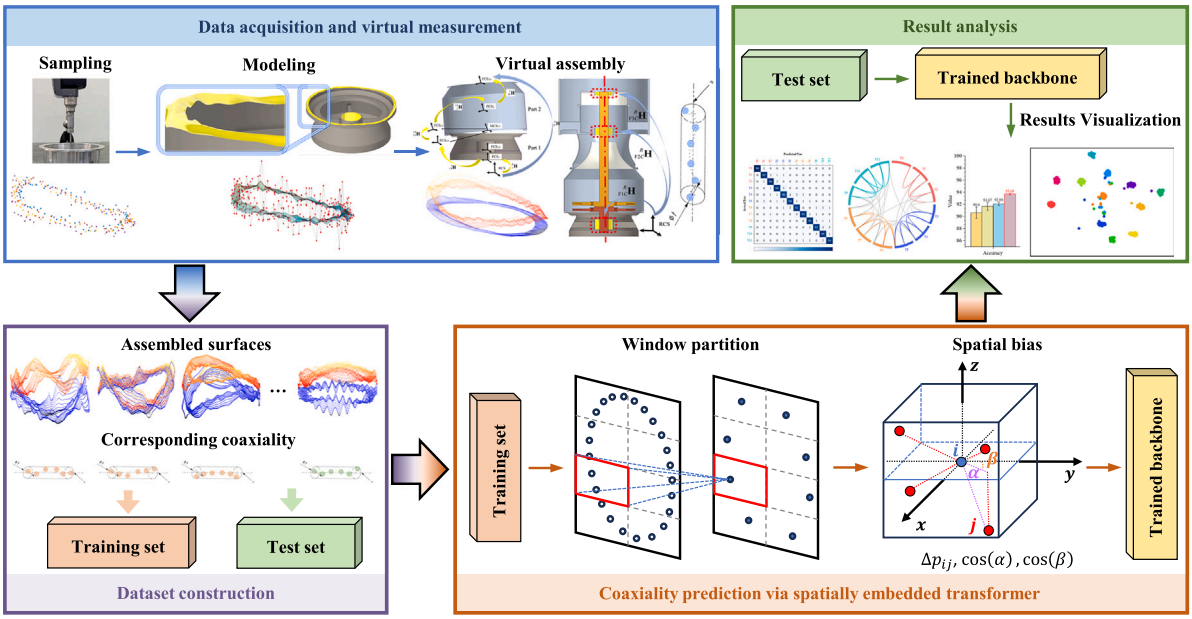


Fig. 2. Framework overview.

cloud inputs but at the cost of increased computational demand. To end this, PT V1 [29] incorporates vector attention [36] into the transformer architecture, reducing the computational load between feature maps and alleviating memory issues. Building on this, PT V2 [30] simplifies vector attention further, enhancing computational efficiency. Despite its advancements, PT V2 struggles with collecting local information, leading to challenges in training convergence with limited data. Additionally, current point cloud transformers are generally designed for general tasks and are seldom utilized in precision assembly fields.

To bridge these gaps, this work introduces a novel transformer backbone, SETrans, which integrates spatial information from neighboring point clouds into the self-attention block to enhance local information aggregation. Two datasets consisting of aero-engine components have been constructed to evaluate the usability and generalizability of SETrans in industrial assembly contexts.

3. Method overview

This study develops an end-to-end point cloud deep learning framework to realize high-precision and high-efficiency coaxiality prediction. The proposed coaxiality prediction system consists of three integrated modules, as shown in Fig. 2.

In the first step, the virtual measurement method is designed to construct the relationship between unassembled surface point clouds and their corresponding coaxialtiy. This involves sampling high-precision points from simulated real aero-engine parts and reconstructing them through a non-uniform rational B-splines (NURBS) surface model for upsampling. These upsampled point clouds are then utilized within a virtual assembly model that constructs the assembled relationships and enables various combinations for effective data augmentation.

Then, in the second step, all these augmented data are randomly split into a training set and test set to construct a point cloud dataset for subsequent deep learning training.

In the third step, SETrans is designed to realize better performance on the point clouds feature extraction task. This transformer integrates spatial information as a bias to the feature map, optimizing the use of angle and distance information between neighboring point clouds and demonstrating the enhanced capabilities of the proposed architecture in managing point domain data.

Finally, in the fourth step, the trained model establishes a mapping relationship between the point cloud data and assembled coaxiality, facilitating an end-to-end prediction from points input in unassembled conditions. The performance and generalizability of SETrans are evaluated using two different types of simulated aero-engine flange components. Additionally, the online prediction results of SETrans and the virtual measurement models are visually displayed using t-distributed stochastic neighbor embedding (t-SNE) and a chord diagram, providing an intuitive evaluation of the model's effectiveness. This comprehensive approach leverages advanced virtual modeling and deep learning techniques to enhance accuracy and efficiency in coaxiality prediction for industrial applications.

4. Virtual measurement for aero-engines

In aero-engine assembly, directly measuring the coaxiality of internal assembled components such as bearing mounting holes and rotor shafts is impractical due to the enclosed assembly structure. The virtual measurement method in this paper provides a solution for coaxiality prediction by integrating real part measurements with virtual coaxiality calculations. This method enhances point density through GDE reconstruction while preserving the accuracy of the original high-precision point clouds, thus reducing the costs associated with individual surface measurements. Additionally, the assembly dataset created through this virtual measurement method is used to train SETrans, an end-to-end coaxiality prediction model that streamlines the process by eliminating the need for frequent adjustments to the virtual assembly model.

4.1. Process of virtual measurement

The virtual measurement for aero-engine coaxiality is illustrated in Fig. 3. After parts measurement for high-precision point clouds in physical space, dense point clouds are obtained by geometric distributed error reconstruction in digital space, and coaxiality is calculated by the virtual assembly. The mapping between the high-precision point clouds and the coaxiality of a single aero-engine is established. Moreover, for aero-engine coaxiality prediction in mass production, large volumes of high-precision dense point clouds are obtained by data augmentation, and large volumes of aero-engine coaxiality are calculated by the single aero-engine virtual measurement and Monte Carlo method. The assembly dataset that maps between the point clouds and the coaxiality for SETrans model training is established.

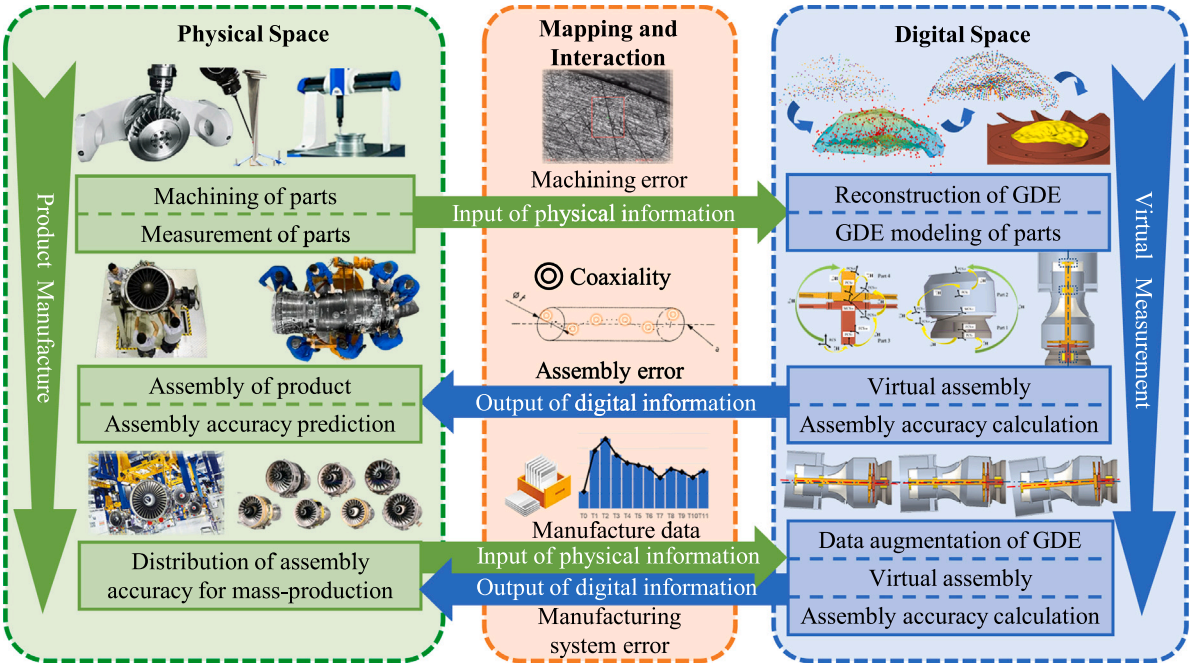


Fig. 3. Process of virtual measurement.

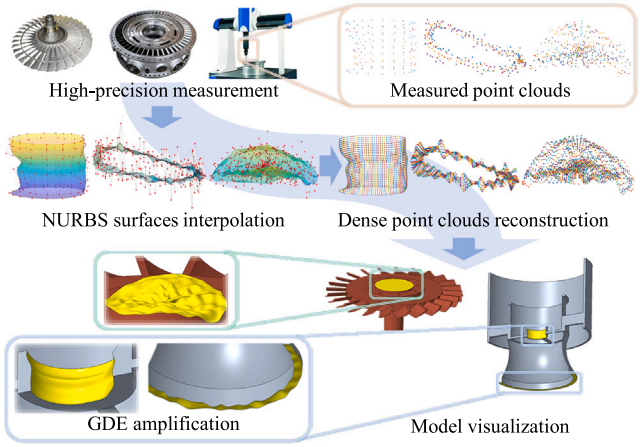


Fig. 4. Geometric distributed error of aero-engine modeling.

4.2. Geometric distributed error reconstruction

To maintain data accuracy, the NURBS method is used to interpolate and model the measured point cloud in GDE reconstruction. Compared to analytic models, NURBS can reconstruct complex geometric shapes with greater accuracy. The NURBS surface is a continuous free-form surface that accurately and comprehensively describes the GDE of the aero-engine surface [37]. A NURBS surface with p degree in the u direction and q degree in the v direction is defined as Eq. (1).

$$S(u,v) = \frac{\sum_{i=0}^{m}\sum_{j=0}^{n} N_{i,p}(u)N_{j,q}(v)w_{i,j}P_{i,j}}{\sum_{i=0}^{m}\sum_{j=0}^{n} N_{i,p}(u)N_{j,q}(v)w_{i,j}} = \sum_{i=0}^{m}\sum_{j=0}^{n} N_{i,p}(u)N_{j,q}(v)P_{i,j}^{w} \quad ,0 \leqslant u,v \leqslant 1 \tag{1}$$

Where, u and v represent knot parameters, $N_{i,p}(u)$ and $N_{j,q}(v)$ are the B-spline basis functions determined by knot vectors $\boldsymbol{U}$ and $\boldsymbol{V}$. $w_{i,j}$ is the control point weight factor and $P_{i,j}$ is the surface control point. GDE reconstruction for aero-engine part surfaces is illustrated in Fig. 4.

The measured point cloud is interpolated by the double cubic NURBS surface to establish the GDE surface model. Then, the GDE surface model is discretized into the dense point cloud for virtual assembly and coaxiality calculation. As shown in Fig. 4, the reconstructed dense point cloud maintains the GDE information of the initial measured point cloud while predicting the GDE information of the surface. The visualized GDE solid part model of the aero-engine part is also illustrated in Fig. 4.

4.3. Virtual assembly and coaxiality calculation

According to the assembly structure of the aero-engine, coaxiality is calculated using the coordinate values of the shaft and hole point clouds within the Reference Coordinate System (RCS). These coordinate values are derived from the initial coordinates in the Feature Coordinate System (FCS), taking into account the position and posture of the shaft and hole surfaces as shown in Eq. (2). In the virtual assembly process, the positioning and orientation of one part are determined based on two parameters: the position and posture of the assembly datum, which is a surface on the preceding assembly part, and the minor translational and rotational errors induced by the contact between two GDE surfaces. Virtual assembly and coaxiality calculation for a single aero-engine, as illustrated in Fig. 5, involves calculating three contact points and the position and posture of the part within the Assembly Motion Coordinate System (MCS) using the difference surface method. The position and posture of each part, along with the coordinate values of the shaft and hole point clouds in the RCS, are computed using the HTM method. Finally, the coaxiality of both the casing system and rotor system is determined using these point clouds.

${}^{1}_{P}v$ is the unknown coordinate value of the point cloud in RCS. ${}^{n}_{P}v$ is the known coordinate value of the point cloud in FCS. ${}^{1}_{P}v$ can be calculated with ${}^{n}_{P}v$ and ${}^{1}_{n}\mathbf{H}$ by HTM method, as shown in Eq. (2).

$${}^{1}_{P}v = {}^{1}_{n}\mathbf{H}^{n}_{P}v = \left(\prod_{k=1}^{n} {}^{k}_{k+1}\mathbf{H}\right){}^{n}_{P}v \tag{2}$$

Where, ${}^{1}_{n}\mathbf{H}$ represents the position and posture of FCS in RCS. ${}^{k}_{k+1}\mathbf{H}$ represents the position and posture of (k+1)th coordinate system in kth coordinate system. Considering assembly limit structures, such as rabbet and dowel pin, degrees of freedom of X, Y, and C in MCS are

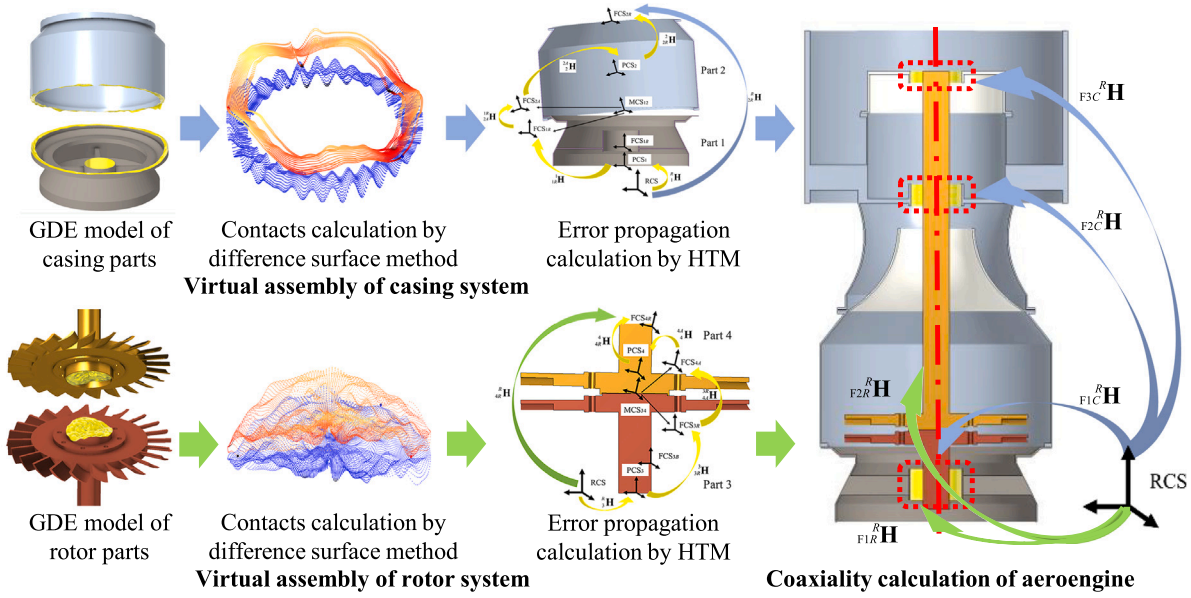


Fig. 5. Virtual measurement of coaxiality.

limited. Equivalently, $dx = dy = 0$, $\sin\theta z = 0$, and $\cos\theta z = 1$. Therefore, HTM of MCS can be simplified as shown in Eq. (3). Where, $\sigma = {}^{k}_{k+1}\theta x$, $\varsigma = {}^{k}_{k+1}\theta y$ and $\tau = {}^{k}_{k+1}dz$ represent translation and rotation of (k+1)th coordinate system in kth coordinate system.

$$ {}^{k}_{k+1}\mathbf{H} = \begin{bmatrix} \cos\varsigma & 0 & \sin\varsigma & 0 \\ \sin\sigma\sin\varsigma & \cos\sigma & -\sin\sigma\cos\varsigma & 0 \\ -\cos\sigma\sin\varsigma & \sin\sigma & \cos\sigma\cos\varsigma & \tau \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{3} $$

In the context of mass production, measuring coaxiality presents significant cost challenges. To address this, a virtual measurement approach combined with the Monte Carlo method is utilized to predict coaxiality in large-scale production scenarios, as illustrated in Fig. 6. Initially, high-precision real measurements of a small sample set are used to obtain GDE point clouds. These initial point clouds are then augmented using production files, which include recorded information on part characteristics such as flatness and cylindricity, to generate a sufficient quantity of augmented GDE point clouds. In simulating parts selection during mass production, the Monte Carlo method is employed, allowing for the calculation of coaxiality through virtual measurement for an individual aero-engine. Ultimately, this process enables the calculation of coaxiality across the mass production of aero-engines and facilitates the establishment of an assembly dataset.

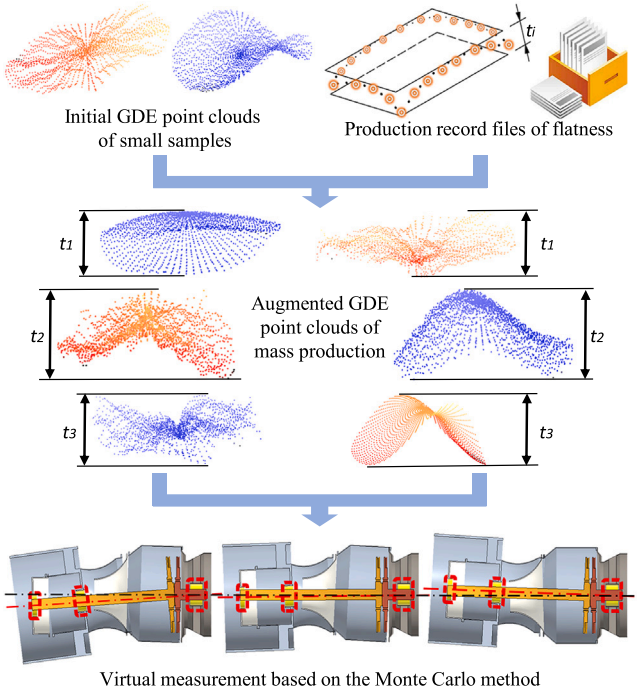


Fig. 6. Virtual measurement of mass production.

5. Spatially embedded transformer model

Following the virtual measurement method, the massive assembly surfaces are prepared. However, the calculation process is time-consuming and requires experienced engineers to build 3D models for every workpiece. To save the modeling costs, deep learning is introduced to realize end-to-end coaxiality prediction with input point clouds instead of tedious math calculations. For better performance, a new point cloud deep learning backbone, SETrans, has been developed. This model is designed specifically for point cloud data, incorporating spatial information to enhance feature extraction and efficiency, thus facilitating more accurate coaxiality assessments in aero-engine assembly.

5.1. Total architecture of SETrans

The total architecture of SETrans is illustrated in Fig. 7. The model establishes a projection relationship between the input point clouds P and the corresponding coaxility C. The input $P = \{p_i \in \mathbb{R}^3,\ i = 1, 2, \ldots, N\}$ comprises N points for a pair of assembly surfaces, where each point p_i is represented by its 3D coordinates (x_i, y_i, z_i). These coordinates are presented with a channel size of three and are processed through a stem that utilizes two multilayer perceptrons (MLP) to extend the data to 32 channels. Guided by the principles outlined in transformer diagrams [33,38], SETrans incorporates layer normalization in place of batch normalization and employs the Gaussian Error Linear Unit (GeLU) as the activation function, replacing the traditional Rectified Linear Unit (ReLU). After the stem, the point clouds undergo four stages of feature map rescaling, where the channel size is expanded to 512, and the number of points is reduced to one in 256. Each stage consists of a transition down layer and a feature extractor layer.

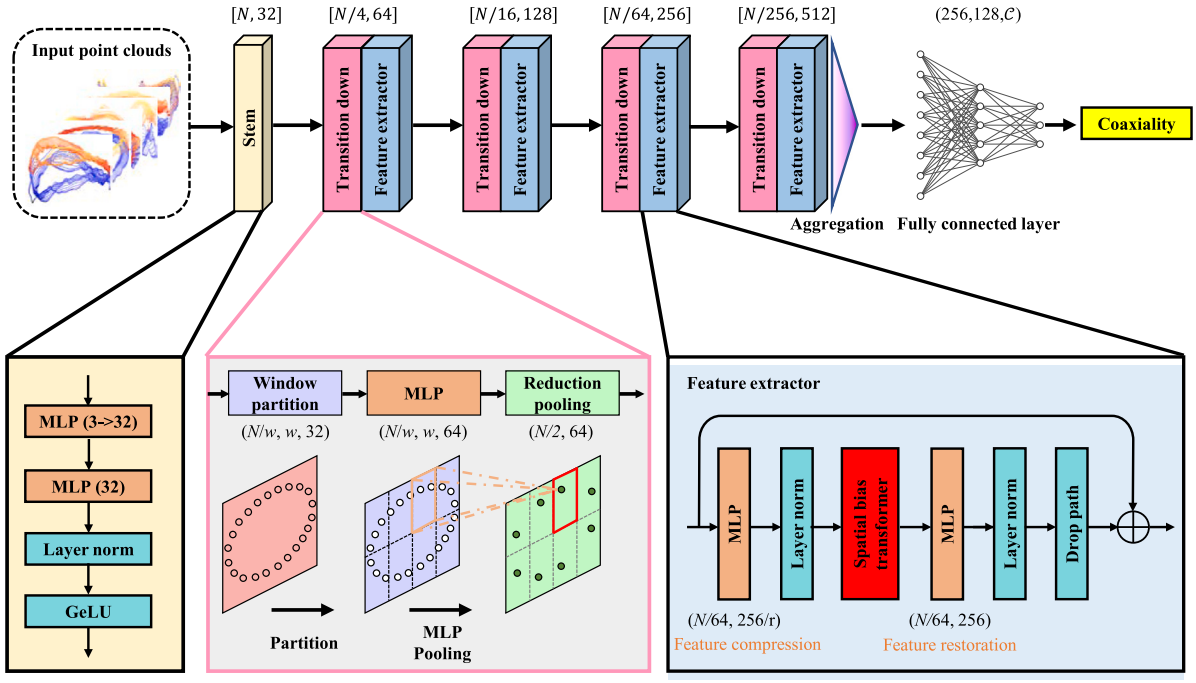


Fig. 7. The total architecture of SETrans.

The transition down layer doubles the receptive field and establishes a global correlation of information among point cloud neighborhoods. It segments the input features into N/w groups, each managing w examples. To prevent repeated sampling in dense points and missing sampling in sparse areas, a window partition strategy is proposed to replace traditional KNN grouping. This approach uses the spliced 3D space for partitioning, moving away from reliance solely on point-to-point distances. After the window partition, an MLP is inserted to realize map extension and a reduction layer is used to reshape the feature map to three dimensions. This approach uses the spliced 3D space for partitioning, moving away from reliance solely on point-to-point distances.

A novel design, a spatial bias transformer block, is proposed to realize high-precision prediction. To manage computational demands, a bottleneck design is implemented, featuring two MLP layers positioned on either side of the spatial bias transformer block. The previous MLP compresses the feature with r ratio, and the subsequent MLP restores the feature size after the transformer block. Notably, r is set to one in the initial stages to ensure robust information interaction due to the lower receptive fields and is increased to two in later stages to align with channel expansion. Similar to the stem, the layer norm realizes the normalization, and GeLU works as the activation function. A skip link and a drop path are also integrated to enhance model stability and convergence, helping to prevent the vanishing gradient problem.

After the feature extraction, a symmetry layer is introduced to aggregate the number channel, realized by max pooling in this work. Finally, three fully connected layers, configured as [256, 128, C], link the classifier and are equipped with a sigmoid function and cross-entropy loss function. The layer norm and dropout layer are inserted into the fully connected layers to prevent gradient vanishing and guarantee the model converges.

5.2. Spatial bias transformer block

Transformer, pivotal in revolutionizing downstream tasks in 2D image processing, adapts effectively to point cloud inputs by aggregating global information. Unlike regular image data, disordered point clouds challenge traditional vision transformer blocks due to their lack of inherent order. To address this, spatial information about point neighbors is integrated into the transformer block, enhancing the relationships within disordered point clouds. This integrated information acts as a bias to enrich the Query, Key, and Value feature maps. A new module, the spatial bias transformer, has been designed, as detailed in Fig. 8.

The input feature map p_i is denoted as $\mathbb{R}^{B\times N\times C}$, where B is the batch size, N presents the number of points, and C is the current feature map size. Drawing from PCT V2 [30], the attention feature map is generalized by input p_i and its sampled neighbor $p_j = S(p_i) \in \mathbb{R}^{B\times K\times C}$. The Query, Key and Value are generalized by feature transformations, denoted as: $\mathcal{F}_Q = l_1(p_i)$, $\mathcal{F}_{K,V} = l_{2,3}(p_j)$, where $\mathcal{F}_{Q,K,V} \in \mathbb{R}^{C^{Q,K,V}}$. To balance the bottleneck design and reduce computation, the Q, K and V are maintained at the same output channel size C. Due to the point clouds having inherent information between the point space distance, the scalar attention feature map design in the vanilla transformer is replaced with vector attention in this work. The vector attention is generated by the Query and Key using a subtraction function rather than a scaled dot product, formulated as follows:

$$\mathcal{F}_{SETrans} = \alpha\left(\lambda \cdot \psi\left(\mathcal{F}_Q, \mathcal{F}_K\right)\right) \odot \mathcal{F}_V \tag{4}$$

Here, ψ represents the relation function for vector calculation between $\mathcal{F}_Q$ and $\mathcal{F}_K$, implemented as matrix subtraction. λ is a learnable parameter to realize feature fusion, and α is the softmax for attention feature map generalization. $\odot$ denotes the Hadamard product for the final feature map. To utilize the 3D position information between the input p_i and its neighbor p_j, spatial information is calculated and added as a bias in Eq. (4). The bias is defined as:

$$\delta_{dis} = \Delta p_{ij} = \left(\Delta x_{ij},\ \Delta y_{ij},\ \Delta z_{ij}\right) \tag{5}$$

$$\delta_{ang} = \cos\alpha + \cos\beta \tag{6}$$

$$\cos\alpha = \sqrt{\left(\Delta x_{ij}\right)^2 + \left(\Delta y_{ij}\right)^2} \Big/ \left|\Delta p_{ij}\right| \tag{7}$$

$$\cos\beta = \left|\Delta y_{ij}\right| \Big/ \sqrt{\left(\Delta x_{ij}\right)^2 + \left(\Delta y_{ij}\right)^2} \tag{8}$$

$$\delta = \theta_1\left(\delta_{dis}\right) + \mu \cdot \theta_2\left(\delta_{ang}\right) \tag{9}$$

δ_{dis} is the bias integrating the spatial distance information. Δp_{ij} is the Euclidean distance set calculated by the input points p_i and their

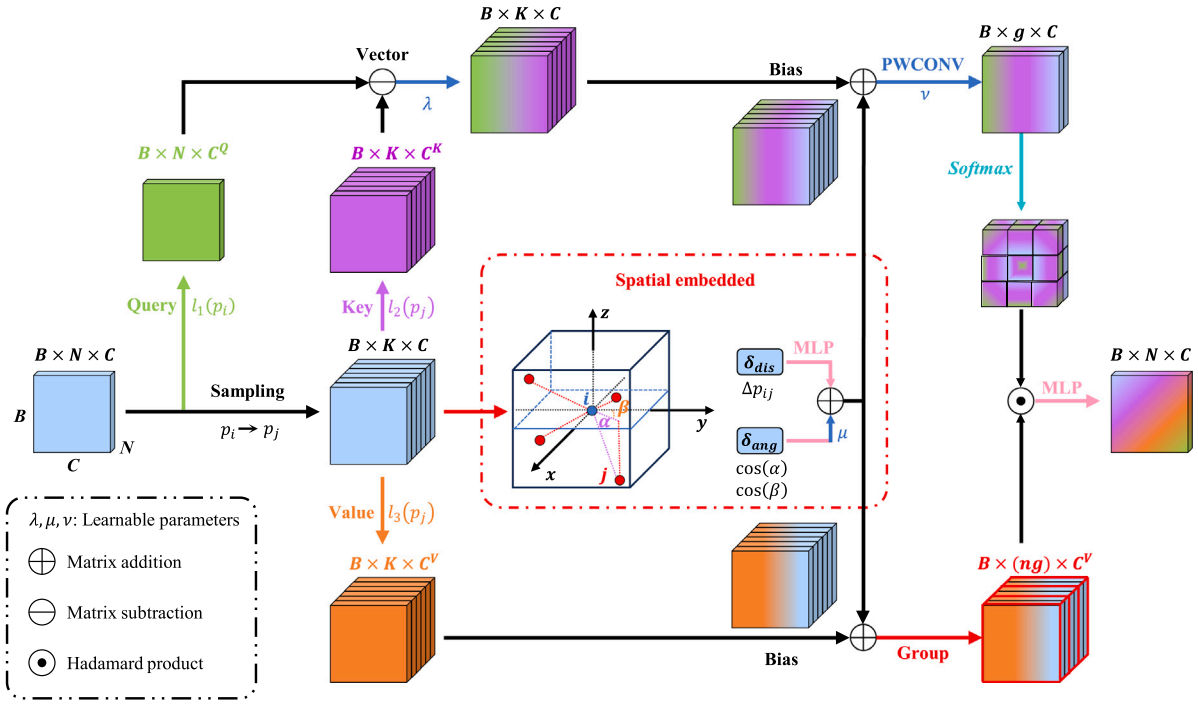


Fig. 8. Spatial bias transformer block.

sampling neighbor p_j. δ_{ang} is the concatenation of the altitude angle $\cos\alpha$ and the azimuth angle $\cos\beta$ for enriching the geometric features. The spatially embedded bias δ is the integration of distance bias δ_{dis} and angle bias δ_{ang}. To ensure the module convergence, two MLP layers θ_1 and θ_2 are employed to modify the δ_{dis} and δ_{ang}. μ is a learnable parameter normalizing the data dimensions to ensure balanced bias fusion. Finally, the bias is inserted into the attention feature map, rewriting Eq. (4) as follows:

$$\mathcal{F}_{SETrans} = \alpha\left(\lambda \cdot \psi\left(\mathcal{F}_Q, \mathcal{F}_K\right) + \delta\right) \odot \left(\mathcal{F}_V + \delta\right) \tag{10}$$

The spatially embedded information enhances comprehensive geometric information capture and coheres point cloud feature map cooperation, crucial for handling complex point cloud data. However, this combination increases the feature encoding channels and raises the risk of computational redundancy. To overcome the weight increase with the expanding of receptive fields, a group design is introduced in the attention feature map fusion process. The feature of Value $\mathcal{F}_V$ is split into g groups and reshaped to $\mathbb{R}^{B\times(ng)\times C^V}$. The n is the number of group records as $n = K/g$. Each group shared the attention feature map generated by Query and Key. To fit the group size, the Value feature map is resized to $\mathbb{R}^{B\times g\times C}$ with a pointwise convolution (recorded as PWCONV in Fig. 8) and all the QK feature maps share the same Value attention in each group. With the group process, the $\mathcal{F}_{SETrans}$ is rewritten mathematically,

$$\mathcal{F}_{SETrans} = \sum_{n=1}^{K/g} \alpha\left(\lambda \cdot \psi\left(\mathcal{F}_Q, \mathcal{F}_K\right) + \delta\right) \odot Sm(\nu\left(\mathcal{F}_V + \delta\right)) \tag{11}$$

The ν is a learnable parameter to help adjust the size of $\mathcal{F}_V$ and Sm indicates the $Softmax$ layer. The final feature map is the mixture of n groups. After the Hadamard product, the $\mathcal{F}_{SETrans}$ is fed to an MLP with GeLU activation and LayerNorm. All the attention informations are fused and transported to the next layer.

5.3. Window partition point clouds sampling

The point cloud deep learning diagrams, such as PointNet [18], DGCNN [22], and PointNext [39], leverage sampling and encoding methods to realize the pooling procedure, as shown in Fig. 9. The farthest point [19] and grid sampling [40] are often chosen to sample the center point for the following neighborhood encoding process. After initial sampling, KNN grouping is introduced to gather the neighboring area points to aggregate the spatial information. Subsequently, the pooling layer resamples the point clouds and reduces the data dimension for network training. In the traditional sampling process, point clouds may be duplicated or omitted due to uncontrollable information density and overlap in the input. Such oversampling not only diminishes the prediction accuracy but also increases computational complexity. To address these issues, the window partition point clouds sampling method is introduced for enhanced efficiency.

As shown in Fig. 9, window partition point cloud sampling splits the input point clouds $P = (p_i, f_i),\ i \in (1, N)$ to K subsets $[P_1, P_2, P_3, \ldots, P_K]$. Each subset $P_K = (p_k, f_k)$, contains the position information p_k and its corresponding feature map f_k. Different from KNN grouping, the window partition directly separates the space instead of calculating the distance between the center point and other sampling points. The non-overlapping spatial separation ensures that points are neither repeatedly sampled nor missed. Additionally, omitting spatial distance calculation ehances calculation efficiency. After the window partition, each group functions similarly to a KNN group, and a pooling layer is employed to condense the information of related points. The position p_k passes through a mean pool layer to derive the new sampled position p'_k, while, the feature f_k undergoes a combination of Maxpool and MLPs to produce the transformed feature f'_k. The MLP is specifically designed to adjust feature sizes to ensure compatibility with subsequent network stages.

6. Experiment

This section evaluates the SETrans with the point cloud datasets established from aero-engine simulated workpieces, comparing it against other state-of-the-art (SOTA) point cloud deep learning methods. Datasets for casing and rotor assembly scenarios are created to test the proposed system thoroughly. To assess the model's generalization capabilities, these two datasets are sampled using different devices, ensuring diversity across data domains.

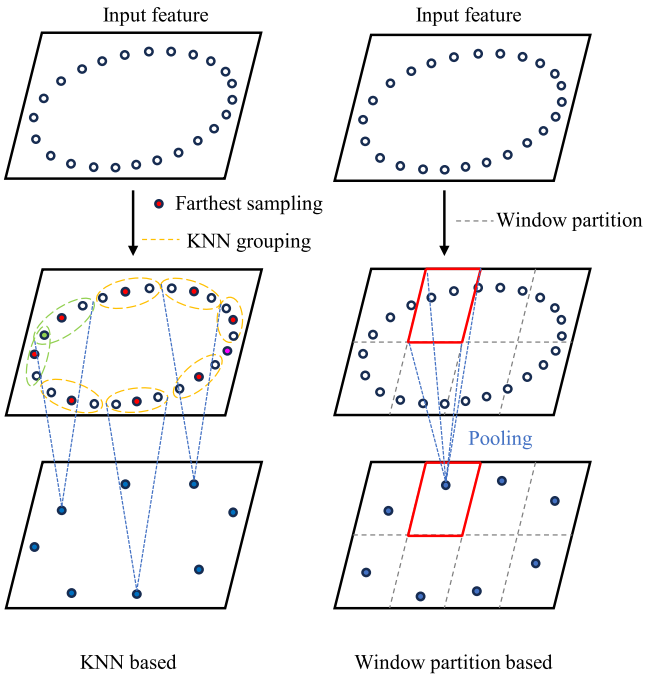


Fig. 9. Point clouds down sampling comparison.

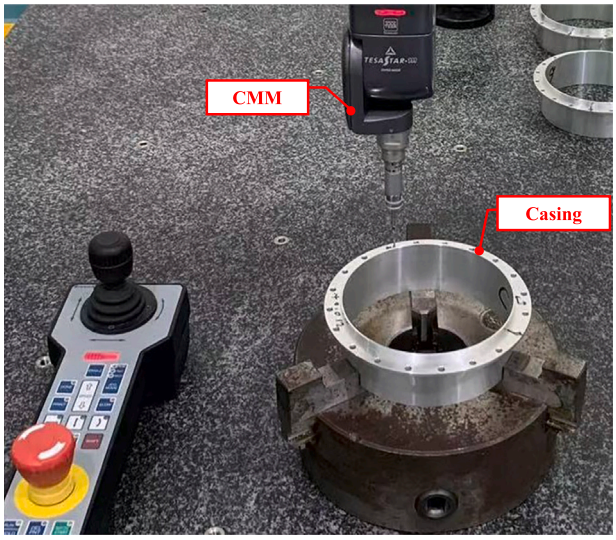


Fig. 10. CMM contact sampling.

6.1. Case 1: engine casing assembly with contact sampling

In this experiment, the focus is on an engine casing assembly dataset sampled using CMM contact techniques. The SETrans model, along with seven other backbone architectures, is tested using this high-precision input data.

6.1.1. Data description

The point clouds are measured by the Hexagon Leitz PMM-XI12107 CMM with $\pm(0.5 + L/500)$ μm, as shown in Fig. 10. Fourteen aero-engine simulated parts are designed and processed to generate the assembly casing surface. The original point clouds are modeled by the GDE modeling illustrated in Section 3 and are used to calculate coaxiality values for the assembly. Each surface contains 1600 point clouds. The generated surfaces are augmented with weight combinations, rotation, function addition, error scaling, mirroring, noise signal addition, and filtering to match more actual assembly situations. The dataset is categorized into twelve classes based on different coaxiality values, ranging from T0 to T11, as detailed in Table 1. The classes are further grouped into four grades, reflecting the assembly quality in real-world scenarios. The entire dataset comprises 4800 examples, with 80% (3840 examples) designated for training and the remaining 20% (960) reserved for testing. Each class is evenly represented with 400 samples. Visualizations of the assembly surface point clouds for different classes are shown in Fig. 11.

Table 1
Details of dataset.

Tier	Coaxiality value (mm)	Grade
T0	0	
T1	0.01	Premium Grade
T2	0.02	
T3	0.03	
T4	0.04	Qualified Grade
T5	0.05	
T6	0.06	
T7	0.07	Reprocessed Grade
T8	0.08	
T9	0.09	
T10	0.1	Substandard Grade
T11	>0.1	

Table 2
Hyperparameter configuration.

Config	Value
Batch size	24
Epoch	500
Drop rate	0.2
Learning rate	0.002
Optimizer	Adam, betas = (0.9,0.999)
StepLR	Step size = 20, gamma = 0.7
Decay rate	0.0002
n_points	1024

6.1.2. Experiment setting

The SETrans are programmed in Pytorch (1.9.0) with Python (3.8.11). The model is trained on a workstation with a CPU of Intel Xeon Platinum 8375C @2.90 GHz and an NVIDIA GeForce RTX 3090 GPU with 24 GB memory using the PyCharm. The model configuration for training includes a batch size of 24 and an input size of 1024 points per sample. Adam optimizer is chosen with a 0.002 learning rate and 0.9 β_1, 0.999 β_2. A learning rate decay strategy is incorporated to enhance both training speed and model convergence accuracy. The learning rate declines in 0.0002 rate and renews in every 20 epochs with 0.7 γ. The training process lasts 500 epochs with a 0.2 drop rate. Specific values of the hyperparameters are detailed in Table 2.

To verify the high precision of spatially embedded design and calculation simplification of window patching, the training process of SETrans and modification baseline PCT are compared in Fig. 12. The training loss and accuracy exhibit swift alterations and retain their initial trajectory during the first 20 epochs, a phenomenon attributable to the substantial learning rate at the onset of the training process. The learning rate decay strategy is applied in the later epochs, promoting stable training loss and successful model convergence. Benefitting from the window partition, the SETrans has a significantly more rapid decline than PCT in the first 50 epochs training stages (0.089 compared to 0.283), highlighting the effectiveness of this design in helping to reduce computing costs. By the 500th epoch, both models achieve stable training losses, indicating good convergence. At the completion of training, SETrans shows a lower training loss (0.014 vs. 0.063), confirming its superior performance and the benefits of its design.

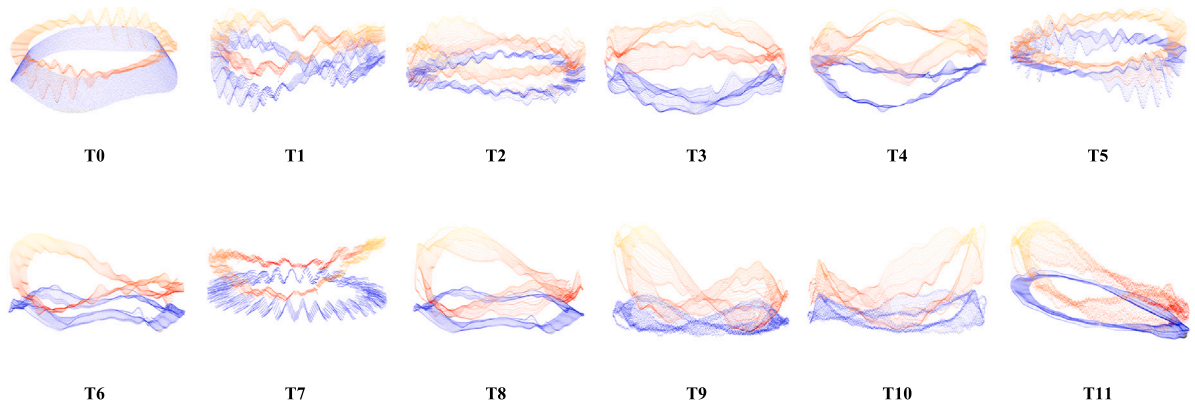


Fig. 11. Visualization of casing samples in different tiers.

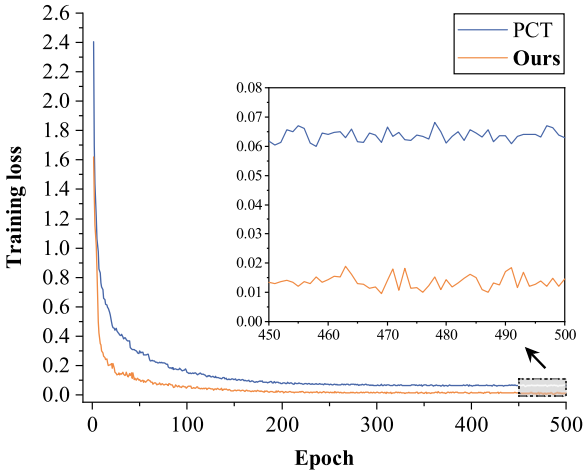


Fig. 12. Training process.

6.1.3. Experiment results

In this section, accuracy (Acc), recall (Rec), precision (Pre), and F1-score (F1) are used to assess the coaxiality prediction performance of the proposed SETrans. These four criteria are defined as

$$\begin{aligned} Acc &= \frac{N_{TP}+N_{TN}}{N_{TP}+N_{TN}+N_{FP}+N_{FN}} \\ Rec &= \frac{N_{TP}}{N_{TP}+N_{FN}} \\ Pre &= \frac{N_{TP}}{N_{TP}+N_{FP}} \\ F1 &= \frac{2\times Pre\times Rec}{Pre+Rec} \end{aligned} \tag{12}$$

Where N_{TP}, N_{FP}, N_{TN}, N_{FN} denote the number of true positives, false positives, true negatives, and false negatives, respectively. SETrans is benchmarked against multiple models, including MLP-based point cloud deep learning baseline PointNet and its modification model, the PointNet++, and PointNext . Additionally, the SOTA method of graph neural network, the DGCNN, is tested in the engine coaxiality dataset. Due to the proposed SETrans being implemented with the transformer mechanism, the transformer-based point clouds deep learning baseline, PT V1, PT V2, and PCT [32] are also tested in this dataset with these four criteria. The test results from Table 3 indicate that SETrans outperformed all other models in the engine sleeve assembly task with a top accuracy of 93.65%. PointNet, limited by its shared MLP structure, scored the lowest accuracy at 87.50%, and PointNet++ also fell short of reaching 90% accuracy. Among the MLP-based models, PointNext performed the best with an accuracy of 91.888%. In the transformer

Table 3
Test result.

Model	Acc	Recall	Precision	F1-score
PointNet	0.8750	0.8722	0.8688	0.8702
PointNet++	0.8927	0.8854	0.8891	0.8863
DGCNN	0.9010	0.8987	0.8916	0.9028
PointNext	0.9188	0.9116	0.9159	0.9092
PT V1	0.8958	0.8852	0.8933	0.8996
PT V2	0.9083	0.9107	0.9024	0.8869
PCT	0.9157	0.9153	0.9006	0.9183
This paper	**0.9365**	**0.9306**	**0.9279**	**0.9348**

category, PCT led with a 91.57% accuracy, benefiting from its global aggregation architecture. SETrans topped the accuracy metric and led in recall, precision, and F1-score, showing gains of 1.53%, 1.2%, and 1.65% points, respectively over the next best performer, PointNext. These results underscore the effectiveness and reliability of SETrans in coaxiality prediction tasks, which are crucial for high-precision assembly.

The test results of SETrans, as detailed in Fig. 13 and Fig. 14, provide a comprehensive analysis of its performance in coaxiality prediction across different tiers of aero-engine assembly. The confusion matrix visualized in Fig. 13 indicates that all tiers achieve high accuracy, exceeding 90%, with T0 showing exceptional performance at 98.75% accuracy. The remaining four tiers also register accuracies above 95%, underscoring the reliability of SETrans in this task. In the context of aero-engine assembly, it is crucial to manage coaxiality errors within acceptable limits, as categorized in Table 1. Evaluation of the prediction results based on error grades, as depicted in the chord diagram (Fig. 14), shows distinct connections. Colored lines link misjudgments within the same error grades, whereas gray lines connect misjudgments across different grades. Significantly, there are no misjudgments between premium and substandard, highlighting the stability of SETrans and its ability to prevent severe misclassifications that could lead to the scrapping of high-quality parts. Considering misjudgments within the same error grade as acceptable, T9 sees a prediction accuracy increase from 92.5% to 98.8%, a 6.3 percentage point improvement. Similarly, the overall accuracy of SETrans across the dataset improves by 3.02 percentage points, from 93.65% to 96.67%. The detailed visualization shows that SETrans realizes stable predictions in each tier and has the potential for further enhancements in real-world applications.

The Average Precision (AP) curve, depicted in Fig. 15, highlights the performance of SETrans compared to seven other baseline models in the coaxiality prediction task. In the MLP-based models, PointNext leads with an AP of 92.1, the highest among the baselines. The transformer-based model, PCT, closely follows with an AP of 91.9. SETrans surpasses both, achieving the highest AP in the dataset at

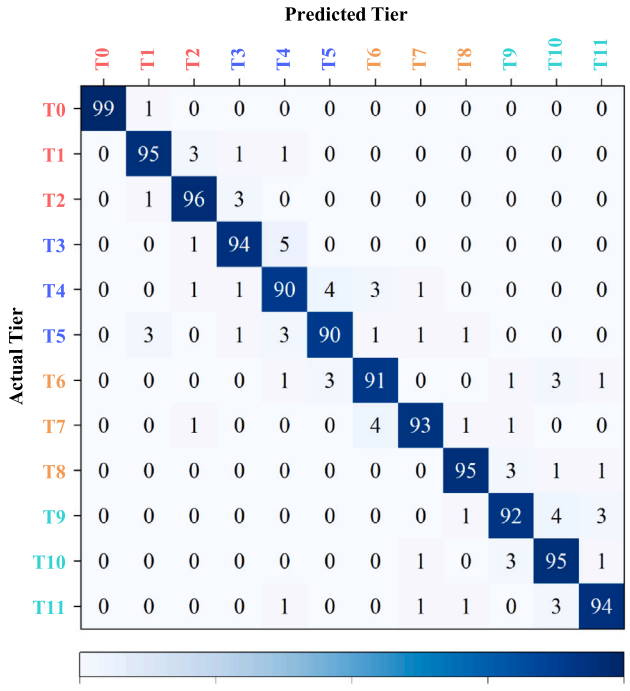


Fig. 13. Confusion matrix visualization for case 1.

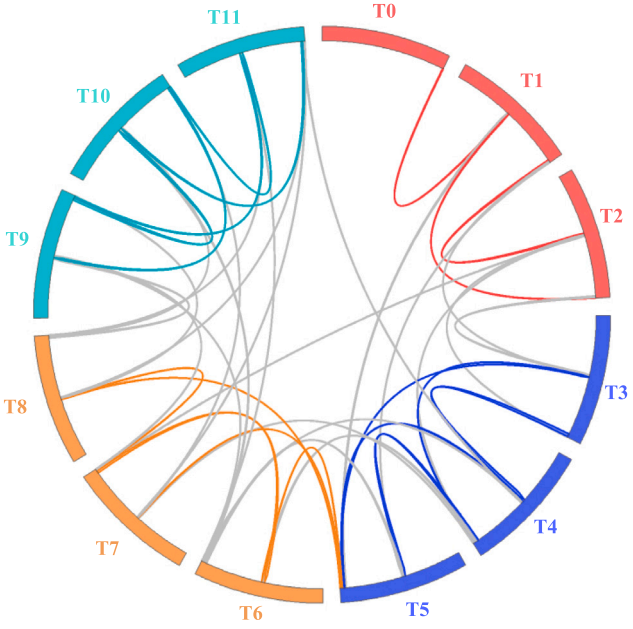


Fig. 14. Chord diagram based on casing assembly.

94.2, indicating a 2.1 point gain over PointNext. This superior AP value underscores the robust performance of SETrans, characterized by high accuracy and stability.

SETrans integrates spatially embedded information into the transformer feature map to enhance efficiency in information utilization. Controlled experiments were conducted to evaluate the effectiveness of this spatially embedded design. The spatial bias (δ) was tested in different configurations: linked solely to the attention map branch generated by Query and Key, connected only to the feature transformation branch generated by Value, and implemented in both branches. A baseline

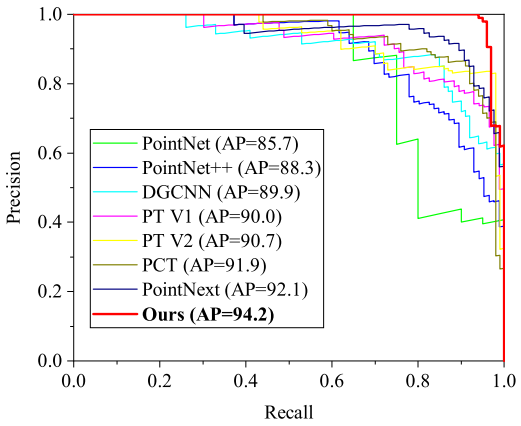


Fig. 15. AP comparison with other baselines in case 1.

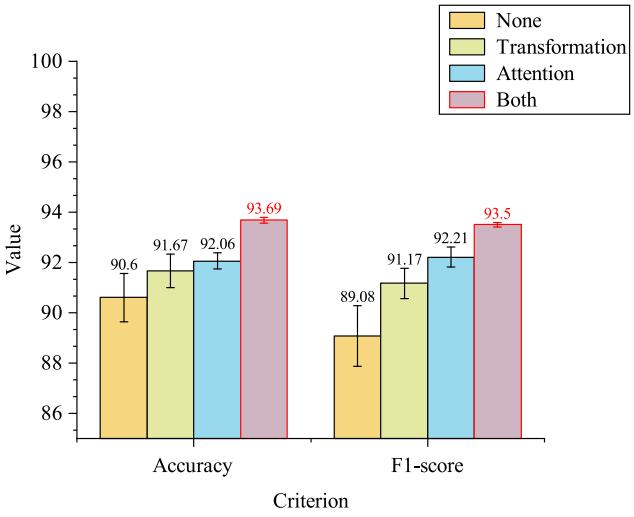


Fig. 16. Ablation study in casing assembly.

transformer without any spatial bias (None) was also evaluated. As shown in Fig. 16, the absence of spatial bias led to a notable performance decline, with accuracy dropping to 90.6% and a higher standard deviation (STD) of 0.97. The best results were achieved when δ was applied to both branches, reaching an accuracy of 93.69% and an F1-score of 93.5. Adding δ exclusively to one branch resulted in lower accuracies of 92.06% and 91.67%, respectively. The ablation study indicated that implementing spatially embedded information in both branches is essential.

Further analysis of the effectiveness of SETrans is presented in Fig. 17, where the extracted point cloud features are visualized using t-distributed stochastic neighbor embedding (t-SNE). This visualization demonstrates that SETrans more distinctly separates the points representing different tiers compared to the other baselines, reflecting stronger recognition capabilities. This improvement is attributed to the strategic integration of spatially embedded bias and the global information processing capabilities of the transformer block. These findings collectively affirm the advanced performance of SETrans in handling complex spatial data, making it a potent tool for tasks requiring precise geometric predictions.

6.2. Case 2: Engine rotor assembly with noncontact sampling

In Case 2, a high-precision point cloud dataset is constructed based on the engine rotor assembly situation, employing a different sampling

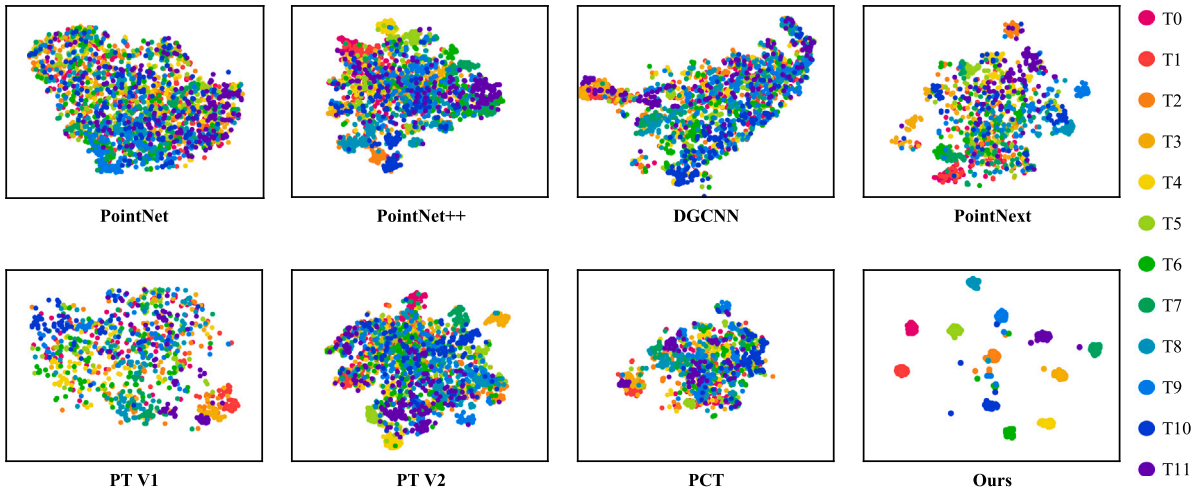


Fig. 17. T-SNE for case 1.

method to demonstrate the generalization capabilities of the proposed SETrans model. The OGP is used as a noncontact sampling device to gather the point clouds.

6.2.1. Data description

Point clouds are gathered by an OGP optical scanner with $\pm(1.2 + L/250)$ μm measurement precision, as shown in Fig. 18. Eighteen aero-engine simulated rotor parts are sampled and used to construct surface assembly point clouds. The CMM and OGP employ different sampling densities, capturing 180 points per surface and 936 points per surface, respectively. Additionally, the casing and rotor are constructed from different materials, specifically steel for the casing and aluminum for the rotor. This variation in sample data distribution further demonstrates that SETrans can be deployed in various assembly situations. The sampled points are processed using the GDE modeling technique outlined in Section 4 to model the relationship between assembled pairs and their corresponding coaxiality. The labeling approach remains consistent with the first case study. The entire dataset comprises 3600 examples, with 80% (2880 examples) designated for training and the remaining 20% (720) reserved for testing. Each class is evenly represented with 300 samples. Similar model designs and hyperparameters from the first case study are also applied in the second case to validate the robustness and adaptability of SETrans. Visualizations of the engine plane assembly situation across 12 classes are shown in Fig. 19.

6.2.2. Experiment results

Similar to the engine sleeve assembly experiments in case 1, the engine rotor assembly is evaluated against seven other baseline models using four criteria. The results indicate that SETrans achieves the highest performance, with an accuracy of 94.31%, which is 2.64 percentage points higher than the second-best model, PCT (91.67%). In terms of precision and F1 score, SETrans surpasses all other models, achieving 93.17% in precision and 94.25% in F- score, with gains over the next best model of 1.23% and 2.9%, respectively. Compared to case 1, the SETrans has a better training effect improvement compared to the second position in case 2. The accuracy improvement in case 2 is 2.64%, 0.56% higher than the improvement in case 1. The improvement in the F1 score is most significant, with a 1.25% points gain (2.9% compared to 1.65%). The experiment demonstrates that SETrans is capable of adapting to diverse assembly contexts and across different domains. The improved performance observed in case 2 underscores the effectiveness of the spatially embedded design of the model. This design enables SETrans to more effectively handle point clouds that are sparse and unevenly distributed, showcasing its robustness and versatility in processing complex spatial data (see Table 4).

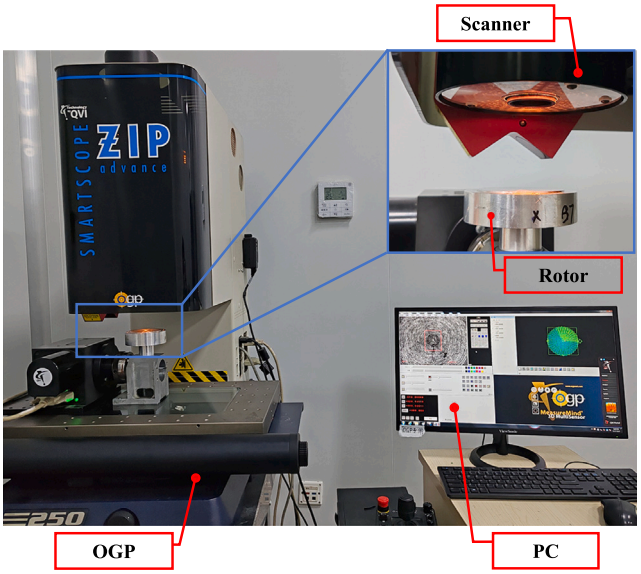


Fig. 18. OGP noncontact sampling.

Table 4
Test result.

Model	Acc	Recall	Precision	F1
PointNet	0.8569	0.8634	0.8618	0.8599
PointNet++	0.8833	0.8796	0.8851	0.8742
DGCNN	0.8917	0.8871	0.8994	0.8768
PointNext	0.8986	0.8867	0.8932	0.9001
PT V1	0.9056	0.9026	0.9194	0.9010
PT V2	0.9111	0.9077	0.9003	0.9026
PCT	0.9167	0.9125	0.9086	0.9135
This paper	**0.9431**	**0.9382**	**0.9317**	**0.9425**

The experimental findings from case 2 are depicted in Fig. 20, where the confusion matrix illustrates that all classes achieve a prediction accuracy exceeding 91%, with the lowest at 91.67% and the highest at 98.33%. This data indicates that SETrans has improved by one percentage point over case 1, suggesting that the transformer is more effective with sparsely distributed point clouds. The chord diagram exhibits trends similar to those observed in case 1, with no misjudgments between substandard and premium grades (as shown in Fig. 21), thus underscoring the reliability of the proposed system in aero-engine assembly scenarios. If misjudgments within the same

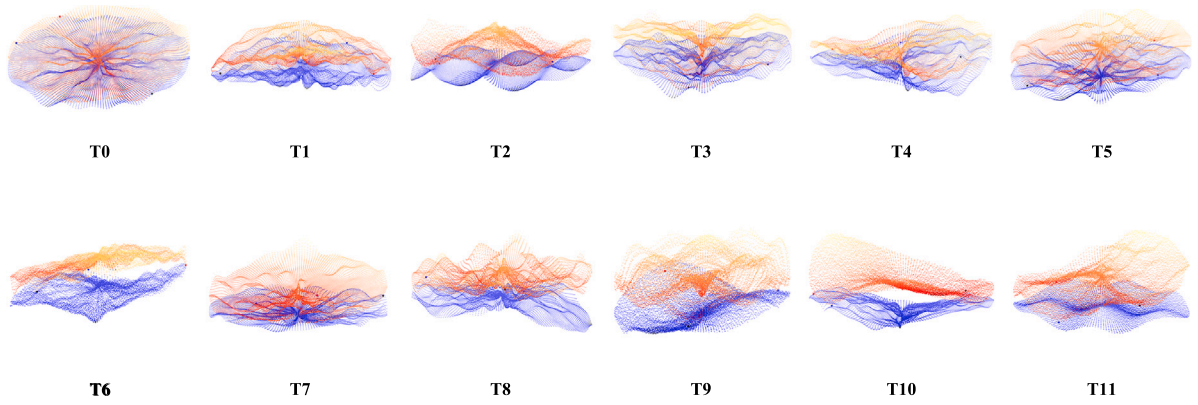


Fig. 19. Visualization of rotor samples in different tiers.

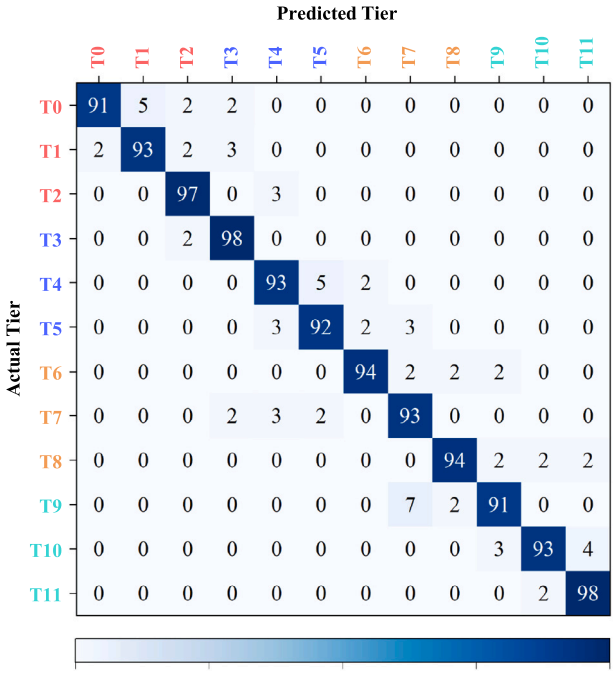


Fig. 20. Confusion matrix visualization for case 2.

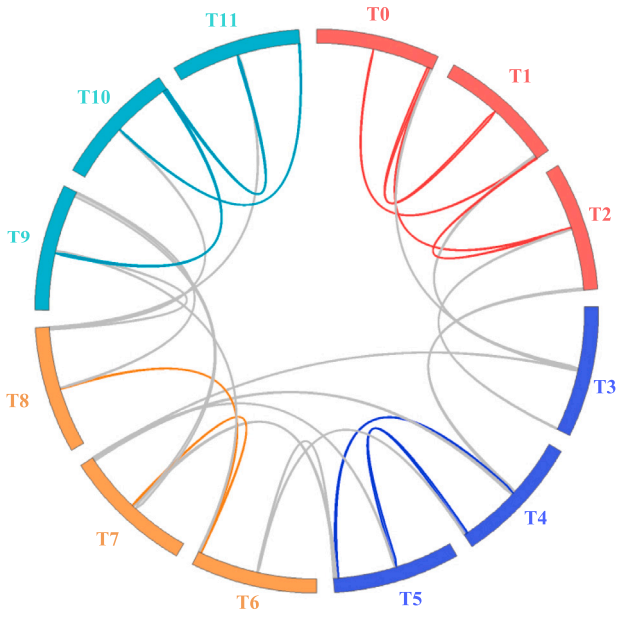


Fig. 21. Chord diagram based on rotor assembly.

grade are considered accurate, SETrans shows an enhancement of 2.4 percentage points, increasing from 94.3% to 96.7%.

Fig. 22 provides a detailed view of the AP curves, highlighting SETrans' superior performance in engine rotor assembly scenarios. The SETrans AP curve, positioned closest to the top right corner of the recall/precision coordinate system, achieves top performance with a 95.2 AP score. Among other models, PointNext leads the MLP-based category with an 88.7 AP score, and PCT stands out among transformer-based baselines with a 92.9 AP score. PT V1, although the least effective among transformer-based networks, still surpasses the best MLP-based score by 2.7 points (91.4 compared to 88.7), demonstrating that transformers aggregate global information more effectively and better suit the sparse engine plane point clouds.

Similar to the experiment design in case 1, the influence of spatial bias is also tested in the rotor assembly context, as shown in Fig. 23. The addition of bias to different transformer feature branches reveals that models without bias adjustment perform the worst, achieving 90.36% accuracy and a 90.06% F1-score. This suggests that spatial information significantly influences transformer training in coaxiality prediction. Among the configurations, attention branch bias adjustment

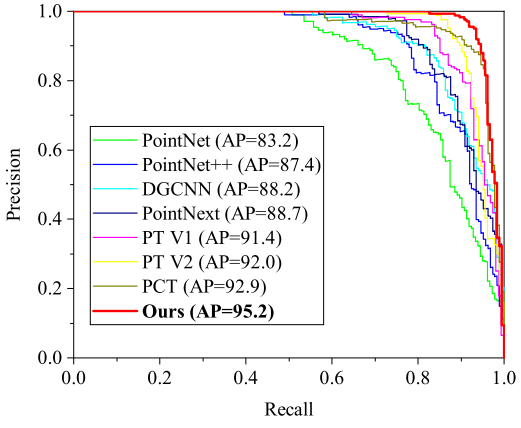


Fig. 22. AP comparison with other baselines in case 2.

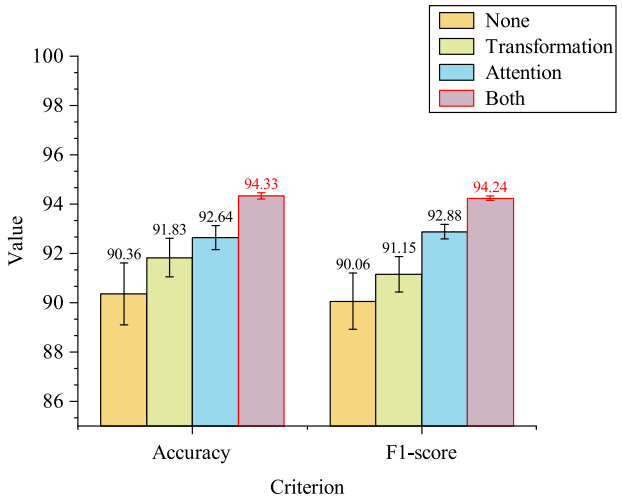


Fig. 23. Ablation study in rotor assembly.

outperforms feature transformation by 0.81% in accuracy (92.64% vs. 91.83%), indicating greater sensitivity to spatial bias. Both branch bias adjustment achieves the highest performance with 94.33% accuracy and a 0.116 standard deviation, confirming the efficiency and robustness of the proposed design.

The t-SNE visualization in Fig. 24 showcases the distinct separation of point features by SETrans, mirroring trends observed in case 1. Only minimal feature fusion occurs between T4 and T5 classes, aligning with the confusion matrix observations. This visualization further confirms the effectiveness and adaptability of SETrans across diverse data distributions.

6.3. Ablation study for SETrans module design

To illustrate the effectiveness of the proposed SETrans design, the influence of embedded bias, window partition sampling, and Value feature grouping were analyzed. The ablation study was implemented, with results presented in Table 5. The baseline results of the vanilla transformer are labeled as "basic"; results with window partition sampling are labeled as "window", and those with Value feature grouping are labeled as "group". Specifically, three scenarios of embedded bias were discussed: embedding the angle (labeled as "ang"), embedding the distance (labeled as "dis"), and incorporating both (labeled as "ang + dis"). The results show that compared to the baseline ("basic"), window partition sampling and Value feature grouping achieved accuracy improvements of 3.19% and 1.38% in Case 1, and 2.8% and 2.52% in Case 2, respectively. Both angle and distance embeddings as spatial biases improved test accuracy in different cases. The combined design of angle and distance biases performed the best, achieving 92.09% accuracy in Case 1 and 92.74% in Case 2, demonstrating the module architecture's effectiveness. SETrans, integrating embedded bias, window partition sampling, and Value feature grouping, showed the best performance with gains of 5.19% points over the vanilla transformer in Case 1 and 5.3% points in Case 2. The ablation study confirms that all three modular designs of SETrans enhance testing performance and are applicable across various industrial scenarios.

6.4. Discussion

In the conducted experiments, SETrans was evaluated using point clouds derived from simulated aero-engine parts. The high-precision point cloud datasets are established utilizing the virtual measurement method. The SETrans was compared against seven other models from MLP-based, GNN-based, and transformer-based categories, ultimately

Table 5
Ablation study of modular designs of SETrans.

	Case 1 casing		Case 2 rotor	
	Acc	F1	Acc	F1
Basic	0.8846	0.8802	0.8901	0.8876
Ang	0.8913	0.8957	0.9019	0.0925
Dis	0.9073	0.8992	0.9115	0.9092
Ang + dis	0.9209	0.9231	0.9274	0.9253
Window	0.9165	0.9071	0.9181	0.9224
Group	0.8984	0.9014	0.9153	0.9187
SETrans	**0.9365**	**0.9348**	**0.9431**	**0.9425**

achieving the highest performance metrics. The proposed SETrans also can be extended to other tolerances, such as parallelism. To underscore the generalizability of SETrans, two distinct datasets representing different assembly scenarios, the casing assembly and the rotor assembly, were utilized. These datasets employed varied sampling methods: contact CMM and noncontact OGP, ensuring a broad representation of domain diversity. SETrans achieved top accuracies of 93.65% and 94.31% across these datasets. Compared to the casing, the point cloud distribution of the rotor is more comprehensive due to the absence of hollow parts. This structural feature enables the transformer's global feature extraction functionality to be fully utilized, leading to better training performance on the rotor dataset with a smaller amount of data. The method's end-to-end prediction capability, which requires no additional modeling or parameter tuning, enhances its universality and offers significant labor cost savings.

To further evaluate the generalizability of SETrans quantitatively, the data distribution was enriched by constructing two additional datasets: casings sampled by OGP and rotors sampled by CMM. The sample and point numbers in these two datasets are similar to the setting in case 1 and case 2. Together with the two datasets from case 1 and case 2, there are now a total of four datasets. The proposed method was tested on these four datasets and evaluated using the area under the receiver operating characteristic curve (AUC) as the metric. The experimental results, shown in Fig. 25, indicate that SETrans achieved the highest AUC, with scores of 0.9832, respectively. It outperformed the second-best model, PCT, by 0.022 scores. Additionally, SETrans exhibited the most minor standard deviation, with 0.00856, respectively. These results demonstrate that SETrans performs optimally across different input data domains and offers the best prediction stability. The experiments quantitatively confirm that SETrans has superior generalization capabilities compared to other deep learning baselines.

The SETrans is specifically designed for point cloud inputs and achieves top performance across various data domains. It not only excels in assembly tasks but also shows potential for extension to other generic tasks that require point cloud input. In the future, efforts will focus on enhancing the model to reduce the volume of training data required and lower overall training costs. Currently, the model focuses on the assembly of only two key components of actual aero-engines. Future research could explore the effects of multiple components assembly scenarios and their impact on coaxiality error. This expansion could potentially enhance the model's applicability and accuracy in complex assembly tasks.

7. Conclusions

This paper introduces an aero-engine coaxiality prediction method that combines the virtual measurement with the SETrans model. This method is designed to provide high precision end-to-end prediction of aero-engine coaxiality, a parameter traditionally unmeasurable directly on production lines. By employing virtual measurement, this method reconstructs the complete GDE at the micron scale and establishes assembly datasets for mass production. Additionally, SETrans, a

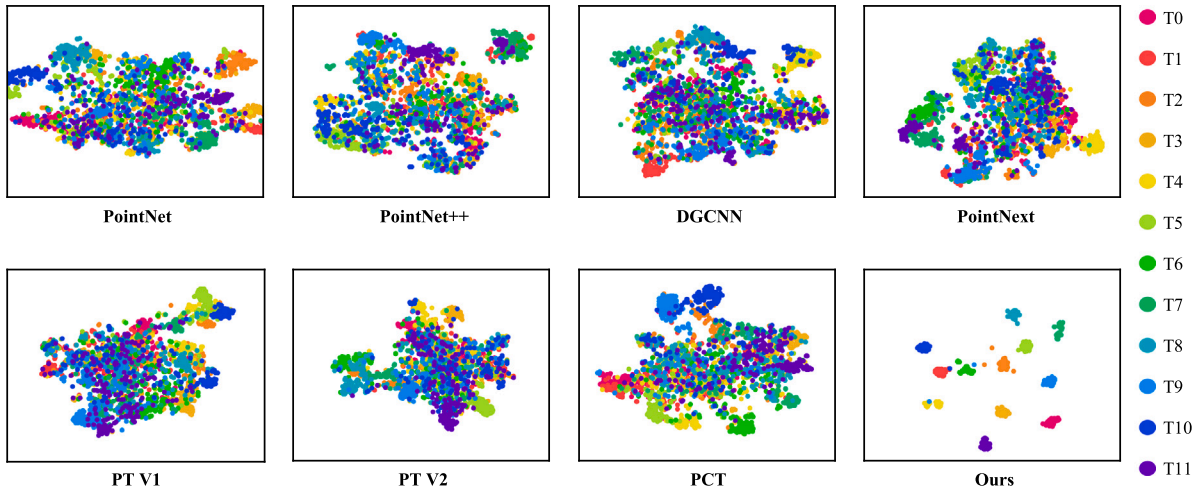


Fig. 24. T-SNE for case 2.

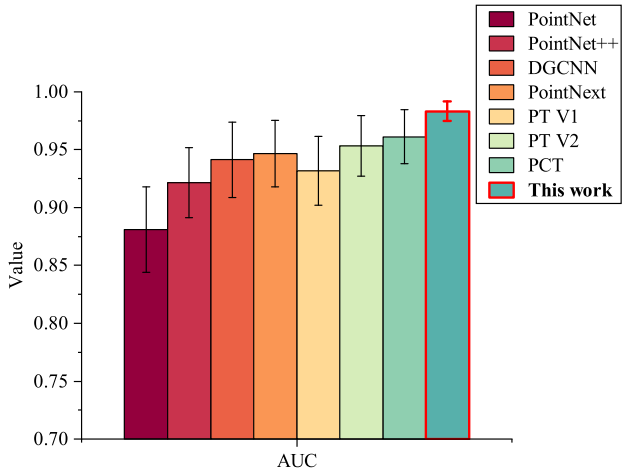


Fig. 25. AUC evaluation with different data domains.

transformer-based deep learning backbone for point clouds, streamlines the process by eliminating the need to construct specific aero-engine models for each target and removes the requirement for parameter tuning in each virtual measurement scenario. This end-to-end deep learning-based approach enhances the efficiency of coaxiality prediction and significantly reduces labor costs in the modeling process. The main contributions of this work are summarized as follows:

(1) This work represents the first attempt to apply the transformer mechanism, particularly its self-attention operator, to the field of aero-engine coaxiality prediction with complex point cloud input. The approach is naturally suited to handle invariant and disorderly input elements, thereby ensuring high prediction accuracy.

(2) The virtual measurement method combines real part measurements with virtual coaxiality calculations to align discrete point clouds with the assembly coaxiality of aero-engine components. Based on micron-scale point cloud measurements of surfaces, the GDE of aero-engine parts is reconstructed using the NURBS method, enabling the calculation of assembled coaxiality. Additionally, GDE point clouds are augmented to enhance coaxiality prediction in mass production and to establish an assembly dataset.

(3) A novel point cloud deep learning backbone, SETrans, is introduced for effective spatial information aggregation. Enhancements such as embedding the distance and angle of neighboring points within the transformer blocks and designing spatially embedded biases are pivotal. These tailored adjustments for point cloud spatial relationships facilitate improved performance even with limited input data.

(4) To mitigate the computational redundancy typically associated with complex transformer architectures, a group strategy is integrated into the SETrans block. Additionally, a spatial window partition method is proposed to prevent missing samples, enabling better aggregation of neighboring point information and reducing computation costs from repeated sampling.

(5) SETrans is tested on an aero-engine point clouds dataset with an error level of one μ m in sampling precision. The experiments demonstrates that SETrans could achieve 0.01 mm precision in coaxiality prediction and outperform other point cloud deep learning baselines. To confirm its generalization capabilities, SETrans is evaluated using two datasets with different aero-engine assembly scenarios and sampling devices, achieving 93.61% and 94.21% prediction accuracy, respectively. These results indicate its adaptability to diverse data domains and potential for deployment in various industrial scenarios.

The proposed method achieves the prediction of aero-engine coaxiality for different assembly components that cannot be directly measured on the production line. The virtual measurement method aligns the aero-engine coaxiality with point clouds and establishes an assembly dataset. SETrans addresses the challenge of independent modeling for each coaxiality prediction case in aero-engine assembly, facilitating end-to-end coaxiality prediction directly from input point clouds without the need for part-specific modeling. The superior performance in coaxiality prediction tasks, as compared to other point cloud deep learning baselines, underscores the effectiveness of the spatial embedding design. The robust performance across different datasets further confirms the generalization and robustness of SETrans, highlighting its potential for widespread industrial application.

CRediT authorship contribution statement

Wu Tianyi: Writing – original draft, Visualization, Software, Methodology, Data curation, Conceptualization. **Shang Ke:** Writing – original draft, Visualization, Software, Methodology, Data curation, Conceptualization. **Jin Xin:** Writing – review & editing, Supervision, Resources, Funding acquisition. **Zhang Zhijing:** Writing – review & editing, Supervision, Resources, Funding acquisition. **Li Chaojiang:** Writing – review & editing, Supervision, Resources, Funding acquisition. **Steven Wang:** Resources, Supervision. **Liu Jun:** Writing – review & editing, Supervision, Resources, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (project No. U22B2088), the Research Grant Council (RGC) of Hong Kong under Grant 11217922, 11212321 and Grant ECS-21212720, and the Science and Technology Innovation Committee of Shenzhen under Grant Type-C SGDX20210823104001011. Additionally, we appreciate the support from the Beijing Advanced Innovation Discipline for Science and Technology of Opto-Electromechanical Micro-Nano Manufacturing.

Data availability

Data will be made available on request.

References

[1] Li Li-li, Chen Kun, Gao Jian-min, Liu Jun-kong, Gao Zhi-yong, Dai Hong-wei, Research on optimizing-assembly and optimizing-adjustment technologies of aero-engine fan rotor blades, Adv. Eng. Inform. 51 (2022) 101506.
[2] Yingjie Mei, Chuanzhi Sun, Chengtian Li, Yongmeng Liu, Jiubin Tan, Research on intelligent assembly method of aero-engine multi-stage rotors based on SVM and variable-step AFSA-BP neural network, Adv. Eng. Inform. 54 (2022) 101798.
[3] Maria G. Juarez, Vicente J. Botti, Adriana S. Giret, Digital twins: Review and challenges, J. Comput. Inf. Sci. Eng. 21 (3) (2021) 030802.
[4] M. Eswaran, M.V.A. Raju Bahubalendruni, Challenges and opportunities on AR/VR technologies for manufacturing systems in the context of industry 4.0: A state of the art review, J. Manuf. Syst. 65 (2022) 260–278.
[5] Siyi Ding, Xiaohu Zheng, Variation analysis considering the partial parallel connection in aero-engine rotor assembly, Energies 15 (12) (2022) 4451.
[6] Alain Desrochers, Walid Ghie, Luc Laperriere, Application of a unified Jacobian—torsor model for tolerance analysis, J. Comput. Inf. Sci. Eng. 3 (1) (2003) 2–14.
[7] Maowei Zhang, Yongmeng Liu, Chuanzhi Sun, Xiaoming Wang, Jiubin Tan, Measurements error propagation and its sensitivity analysis in the aero-engine multistage rotor assembling process, Rev. Sci. Instrum. 90 (11) (2019).
[8] Benjamin Schleich, Sandro Wartzack, Approaches for the assembly simulation of skin model shapes, Comput. Aided Des. 65 (2015) 18–33.
[9] Ci He, Shuyou Zhang, Lemiao Qiu, Xiaojian Liu, Zili Wang, Assembly tolerance design based on skin model shapes considering processing feature degradation, Appl. Sci. 9 (16) (2019) 3216.
[10] Yulan Guo, Hanyun Wang, Qingyong Hu, Hao Liu, Li Liu, Mohammed Bennamoun, Deep learning for 3d point clouds: A survey, IEEE Trans. Pattern Anal. Mach. Intell. 43 (12) (2020) 4338–4364.
[11] Yulan Guo, Hanyun Wang, Qingyong Hu, Hao Liu, Li Liu, Mohammed Bennamoun, Deep learning for 3d point clouds: A survey, IEEE Trans. Pattern Anal. Mach. Intell. 43 (12) (2020) 4338–4364.
[12] Hang Su, Subhransu Maji, Evangelos Kalogerakis, Erik Learned-Miller, Multi-view convolutional neural networks for 3d shape recognition, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 945–953.
[13] Tan Yu, Jingjing Meng, Junsong Yuan, Multi-view harmonized bilinear network for 3d object recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 186–194.
[14] Songle Chen, Lintao Zheng, Yan Zhang, Zhixin Sun, Kai Xu, Veram: View-enhanced recurrent attention model for 3d shape classification, IEEE Trans. Vis. Comput. Graphics 25 (12) (2018) 3244–3257.
[15] Daniel Maturana, Sebastian Scherer, Voxnet: A 3d convolutional neural network for real-time object recognition, in: 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS, IEEE, 2015, pp. 922–928.
[16] Truc Le, Ye Duan, Pointgrid: A deep network for 3d shape understanding, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 9204–9214.
[17] Gernot Riegler, Ali Osman Ulusoy, Andreas Geiger, Octnet: Learning deep 3d representations at high resolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 3577–3586.
[18] Charles R. Qi, Hao Su, Kaichun Mo, Leonidas J. Guibas, Pointnet: Deep learning on point sets for 3d classification and segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 652–660.
[19] Charles Ruizhongtai Qi, Li Yi, Hao Su, Leonidas J. Guibas, Pointnet++: Deep hierarchical feature learning on point sets in a metric space, Adv. Neural Inf. Process. Syst. 30 (2017).
[20] Yangyan Li, Rui Bu, Mingchao Sun, Wei Wu, Xinhan Di, Baoquan Chen, Pointcnn: Convolution on x-transformed points, Adv. Neural Inf. Process. Syst. 31 (2018).
[21] Jiageng Mao, Xiaogang Wang, Hongsheng Li, Interpolated convolutional networks for 3d point cloud understanding, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 1578–1587.
[22] Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E. Sarma, Michael M. Bronstein, Justin M. Solomon, Dynamic graph cnn for learning on point clouds, ACM Trans. Graph. 38 (5) (2019) 1–12.
[23] Guohao Li, Matthias Muller, Ali Thabet, Bernard Ghanem, Deepgcns: Can gcns go as deep as cnns? in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 9267–9276.
[24] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, 2018, arXiv preprint arXiv:1810.04805.
[25] Zihang Dai, Zhilin Yang, Yiming Yang, Jaime Carbonell, Quoc V. Le, Ruslan Salakhutdinov, Transformer-xl: Attentive language models beyond a fixed-length context, 2019, arXiv preprint arXiv:1901.02860.
[26] Xuran Pan, Tianzhu Ye, Zhuofan Xia, Shiji Song, Gao Huang, Slide-transformer: Hierarchical vision transformer with local self-attention, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 2082–2091.
[27] Xinyu Liu, Houwen Peng, Ningxin Zheng, Yuqing Yang, Han Hu, Yixuan Yuan, Efficientvit: Memory efficient vision transformer with cascaded group attention, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 14420–14430.
[28] Dezun Zhao, Wenbin Cai, Lingli Cui, Adaptive thresholding and coordinate attention-based tree-inspired network for aero-engine bearing health monitoring under strong noise, Adv. Eng. Inform. 61 (2024) 102559.
[29] Hengshuang Zhao, Li Jiang, Jiaya Jia, Philip H.S. Torr, Vladlen Koltun, Point transformer, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 16259–16268.
[30] Xiaoyang Wu, Yixing Lao, Li Jiang, Xihui Liu, Hengshuang Zhao, Point transformer v2: Grouped vector attention and partition-based pooling, Adv. Neural Inf. Process. Syst. 35 (2022) 33330–33342.
[31] Xinhai Liu, Zhizhong Han, Yu-Shen Liu, Matthias Zwicker, Point2sequence: Learning the shape representation of 3d point clouds with an attention-based sequence to sequence network, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 33, No. 01, 2019, pp. 8778–8785.
[32] Meng-Hao Guo, Jun-Xiong Cai, Zheng-Ning Liu, Tai-Jiang Mu, Ralph R. Martin, Shi-Min Hu, Pct: Point cloud transformer, Comput. Vis. Media 7 (2021) 187–199.
[33] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al., An image is worth 16x16 words: Transformers for image recognition at scale, 2020, arXiv preprint arXiv:2010.11929.
[34] Juho Lee, Yoonho Lee, Jungtaek Kim, Adam Kosiorek, Seungjin Choi, Yee Whye Teh, Set transformer: A framework for attention-based permutation-invariant neural networks, in: International Conference on Machine Learning, PMLR, 2019, pp. 3744–3753.
[35] Jiancheng Yang, Qiang Zhang, Bingbing Ni, Linguo Li, Jinxian Liu, Mengdie Zhou, Qi Tian, Modeling point clouds with self-attention and gumbel subset sampling, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 3323–3332.
[36] Hengshuang Zhao, Jiaya Jia, Vladlen Koltun, Exploring self-attention for image recognition, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 10076–10085.
[37] Zhongqing Zhang, Zhijing Zhang, Xin Jin, Qiushuang Zhang, A novel modelling method of geometric errors for precision assembly, Int. J. Adv. Manuf. Technol. 94 (2018) 1139–1160.
[38] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, Baining Guo, Swin transformer: Hierarchical vision transformer using shifted windows, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 10012–10022.
[39] Guocheng Qian, Yuchen Li, Houwen Peng, Jinjie Mai, Hasan Hammoud, Mohamed Elhoseiny, Bernard Ghanem, Pointnext: Revisiting pointnet++ with improved training and scaling strategies, Adv. Neural Inf. Process. Syst. 35 (2022) 23192–23204.
[40] Hugues Thomas, Charles R. Qi, Jean-Emmanuel Deschaud, Beatriz Marcotegui, François Goulette, Leonidas J. Guibas, Kpconv: Flexible and deformable convolution for point clouds, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 6411–6420.