mathematics

MDPI

Article

KDiscShapeNet: A Structure-Aware Time Series Clustering Model with Supervised Contrastive Learning

Xi Chen [1], Yufan Jiang [1], Yingming Zhang [1] and Chunhe Song [2,*]

[1] School of Software, Shenyang University of Technology, Shenyang 110870, China; chenxi0914@sut.edu.cn (X.C.); jyf0104@smail.sut.edu.cn (Y.J.); zhangyingming@smail.sut.edu.cn (Y.Z.)
[2] Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110870, China
* Correspondence: songchunhe@sia.cn; Tel.: +86-180-4139-9056

Abstract

Time series clustering plays a vital role in various analytical and pattern recognition tasks by partitioning structurally similar sequences into semantically coherent groups, thereby facilitating downstream analysis. However, building high-quality clustering models remains challenging due to three key issues: (i) capturing dynamic shape variations across sequences, (ii) ensuring discriminative cluster structures, and (iii) enabling end-to-end optimization. To address these challenges, we propose KDiscShapeNet, a structure-aware clustering framework that systematically extends the classical k-Shape model. First, to enhance temporal structure modeling, we adopt Kolmogorov–Arnold Networks (KAN) as the encoder, which leverages high-order functional representations to effectively capture elastic distortions and multi-scale shape features of time series. Second, to improve intra-cluster compactness and inter-cluster separability, we incorporate a dual-loss constraint by combining Center Loss and Supervised Contrastive Loss, thus enhancing the discriminative structure of the embedding space. Third, to overcome the non-differentiability of traditional K-Shape clustering, we introduce Differentiable k-Shape, embedding the normalized cross-correlation (NCC) metric into a differentiable framework that enables joint training of the encoder and the clustering module. We evaluate KDiscShapeNet on nine benchmark datasets from the UCR Archive and the ETT suite, spanning healthcare, industrial monitoring, energy forecasting, and astronomy. On the Trace dataset, it achieves an ARI of 0.916, NMI of 0.927, and Silhouette score of 0.931; on the large-scale ETTh1 dataset, it improves ARI by 5.8% and NMI by 17.4% over the best baseline. Statistical tests confirm the significance of these improvements ($p < 0.01$). Overall, the results highlight the robustness and practical utility of KDiscShapeNet, offering a novel and interpretable framework for time series clustering.

Keywords: time series data; clustering; data mining

MSC: 68W99

check for updates

Academic Editor: Pedro Carpena

Received: 28 July 2025
Revised: 25 August 2025
Accepted: 29 August 2025
Published: 1 September 2025

Citation: Chen, X.; Jiang, Y.; Zhang, Y.; Song, C. KDiscShapeNet: A Structure-Aware Time Series Clustering Model with Supervised Contrastive Learning. *Mathematics* **2025**, *13*, 2814. https://doi.org/10.3390/math13172814

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The widespread and systematic collection of time series data across various domains has enabled the characterization of dynamic processes evolving over time [1]. With the increasing volume of time series data, the embedded information becomes increasingly complex, thereby posing new challenges for effective analysis and model design. As a fundamental data mining task, time series clustering plays a crucial role in diverse

real-world applications, such as healthcare monitoring, financial risk control, predictive maintenance, and user behavior modeling [2,3]. Early studies predominantly adopted distance-based clustering methods, such as Global k-means [4] and k-Shape [5]. The former relies on an iterative, incremental clustering strategy in Euclidean space, while the latter employs normalized cross-correlation (NCC) to ensure phase invariance in sequence alignment. Although these approaches are computationally efficient, they exhibit limited expressive power when dealing with complex sequences characterized by nonlinear temporal structures or multi-scale dynamics. In response to these limitations, recent efforts have increasingly focused on deep representation learning has been increasingly applied to time series clustering, leading to a series of model advancements. AutoShape [6] pioneered the integration of shapelet mechanisms with autoencoders, enabling automatic discovery of discriminative local subsequences and enhancing sensitivity to fine-grained sequence structures. TS2Vec [7] adopts contrastive learning with sliding windows to extract multi-scale contextual information, exhibiting strong transferability and stability across downstream tasks. Voice2Series [8], by reprogramming pretrained speech recognition models, demonstrates the potential of cross-modal representations for time series modeling, thereby promoting generalization in deep clustering frameworks.

Despite recent advances, deep representation learning alone often fails to prevent issues such as ambiguous cluster boundaries and cluster collapse. To address this, Zhong et al. [9] introduced Deep Temporal Contrastive Clustering, which incorporates a temporal contrastive mechanism to enforce intra-class similarity and inter-class separability within local windows, effectively mitigating unclear clustering boundaries. Ma et al. [10] further enhanced clustering discriminability by jointly optimizing the clustering and representation learning objectives, leveraging pseudo-labels to dynamically guide the clustering process. SOM-Times [11] extends self-organizing maps to temporal clustering tasks, resulting in improved topological structure and stability, particularly demonstrating strong practical utility in clinical applications. As end-to-end optimization becomes a prevailing trend, traditional non-differentiable metrics such as Dynamic Time Warping (DTW) increasingly reveal their limitations in complex tasks, restricting the trainability of clustering models. To overcome this constraint, Boniol et al. proposed k-Graph [12], a structure-aware clustering model based on graph embedding, which enhances the expressiveness for complex temporal structures while maintaining interpretability. TCGAN [13] leverages generative adversarial networks to construct differentiable sequence generators and discriminators, enabling dynamic augmentation of clustering objectives in an end-to-end fashion. Explainable TN-ODE [14], on the other hand, employs tensorized neural ordinary differential equations to accurately model temporal dynamics at arbitrary time steps, combining continuous modeling capabilities with interpretability. Collectively, these studies reflect an evolutionary trajectory in time series clustering research-from static distance-based methods toward differentiable, structure-aware models—demonstrating notable improvements in shape modeling, inter-class separability, and structural integration. Future work may further benefit from integrating multi-scale structural learning, cross-modal alignment, and self-supervised graph augmentation strategies, thereby advancing the generalization and explainability of clustering models in large-scale heterogeneous temporal data.

While recent methods have significantly enhanced the representational capacity of unsupervised models, several critical limitations remain:

(1) Insufficient modeling of temporal shape: Time series often exhibit varying forms due to differences in sampling frequency, amplitude perturbations, or nonlinear phase shifts. As a result, feature extraction methods based on Euclidean distance or fixed filters fail to accurately capture the underlying semantic consistency across sequences, making them ineffective in modeling local shape variations and elastic alignments.

(2) Ambiguous inter-cluster boundaries and cluster collapse: In unsupervised settings, imbalanced cluster distributions, scattered intra-class representations, and degraded embeddings frequently arise, leading to unstable clustering performance.
(3) Non-differentiable distance metrics hinder end-to-end learning: Metrics such as the normalized cross-correlation (NCC) used in k-Shape are non-Euclidean and inherently non-differentiable, making them incompatible with neural network optimization frameworks. This restricts joint training between the feature encoder and the clustering objective. Without a differentiable formulation, the learned representations cannot be dynamically adjusted to accommodate shape alignment needs, resulting in suboptimal and unstable clustering outcomes.

To address the aforementioned challenges, this paper proposes KDiscShapeNet, a novel time series clustering framework that integrates supervised contrastive learning with structure-aware modeling. At its core, KDiscShapeNet employs a highly expressive encoder based on Kolmogorov–Arnold Networks (KAN), which captures complex morphological variations through nonlinear functional representations. To retain shape alignment capabilities while enabling end-to-end optimization, we introduce a differentiable k-Shape clustering mechanism, embedding normalized cross-correlation within a differentiable framework. Furthermore, a dual discriminative enhancement strategy—combining supervised contrastive loss with center loss—enforces stronger intra-cluster compactness and inter-cluster separability, thereby improving boundary clarity and clustering robustness under unsupervised settings.

The main contributions of this work are summarized as follows:

(1) We propose KDiscShapeNet, a unified time series clustering framework that integrates discriminative learning with structure-aware modeling. By jointly incorporating Center Loss and Supervised Contrastive Loss, the model enforces intra-cluster compactness and inter-cluster separability. Meanwhile, a Differentiable k-Shape mechanism is introduced to preserve temporal shape alignment, enabling end-to-end optimization within a cohesive training paradigm.
(2) We leverage Kolmogorov–Arnold Networks (KAN) as the clustering encoder. With its high-order nonlinear functional representation and inherent interpretability, KAN significantly enhances the model's ability to capture complex temporal deformations and dynamic patterns, while extending its applicability to unsupervised learning settings.
(3) We conduct extensive experiments on benchmark datasets, including UCR and ETT, encompassing both comparative evaluation and ablation studies. KDiscShapeNet consistently outperforms state-of-the-art time series clustering baselines in handling nonlinear and multi-scale sequences. Visualization and statistical significance analyses further confirm the model's robustness and interpretability during the learning process.

The remainder of this article is structured as follows. Section 2 reviews the related literature. Section 3 introduces the proposed KDiscShapeNet framework. Section 4 presents the experimental evaluation and ablation analysis. Finally, Section 5 concludes the study and outlines future research directions.

2. Related Work

2.1. Time Series Clustering

Time series clustering aims to partition a set of temporally ordered observations into distinct groups, such that sequences within the same cluster exhibit consistent behavioral

patterns or shape dynamics, while those in different clusters present significant structural divergence [15]. We consider a collection of time series defined as:

$$\mathcal{X} = \left\{\mathbf{x}^{(i)}\right\}_{i=1}^{N}, \quad \mathbf{x}^{(i)} = \left(x_1^{(i)}, \ldots, x_{T_i}^{(i)}\right) \tag{1}$$

where N denotes the number of time series; $\mathbf{x}^{(i)}$ is the i-th univariate time series; and T_i is its length. For the multivariate case, $\mathbf{x}^{(i)} \in \mathbb{R}^{T_i \times D}$, where D is the number of variables.

The objective is to jointly learn a representation and clustering assignment:

$$\mathbf{z}^{(i)} = f_\theta\left(\mathbf{x}^{(i)}\right) \in \mathbb{R}^d, \; y^{(i)} = g_\phi\left(\mathbf{z}^{(i)}\right) \in \{1, \ldots, K\} \tag{2}$$

where f_θ is the encoder parameterized by θ, which maps a time series to a d-dimensinal embedding $\mathbf{z}^{(i)}$; g_ϕ is the clustering function parameterized by ϕ, assigning each embedding to one of K clusters; and $y^{(i)}$s the predicted cluster index.

Optionally, when using soft assignments, we define:

$$\mathbf{p}^{(i)} = \text{softmax}\left(s\left(\mathbf{z}^{(i)}, \mathbf{C}\right)/\tau\right) \tag{3}$$

where $\mathbf{C} = [\mathbf{c}_1, \ldots, \mathbf{c}_K]$ are cluster prototypes, $s(\cdot, \cdot)$ is a similarity function, and $\tau > 0$ is a temperature parameter. Based on the underlying strategy used to assign samples to clusters, time series clustering methods can be broadly categorized into the following four types, as illustrated in Figure 1:

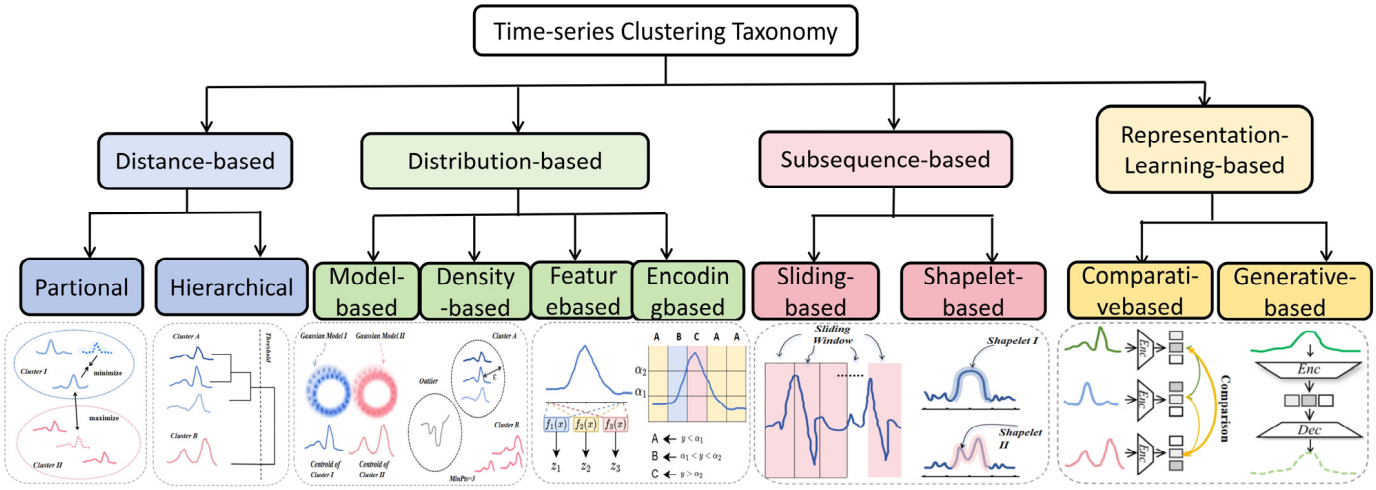


Figure 1. Taxonomy of time series clustering algorithms.

(1) Distance-based clustering: These methods directly partition raw time series based on explicitly defined similarity measures, and crisp partitional algorithms represent the most widely used category in this group. Representative approaches include k-Means, k-Medoids, and k-Shape, where each sample is deterministically assigned to a single cluster, emphasizing partition consistency and intra-cluster compactness. To improve robustness under uncertainty, nonlinearity, or multi-scale temporal variations, various extensions have been proposed, such as fuzzy k-Shape [16] and kernelized k-Shape [17]. These methods introduce soft assignment strategies or kernel-based similarity measures, enhancing the model's adaptability to ambiguous cluster boundaries and nonlinear structures.

(2) Distribution-based clustering: These methods model time series as samples generated from underlying probabilistic processes. Common approaches include Gaussian

Mixture Models (GMMs) and Hidden Markov Models (HMMs), which capture the statistical generative characteristics of the data. Due to their strong modeling assumptions, such methods are typically best suited for domain-specific applications, such as financial time series analysis or biomedical signal interpretation.

(3) Subsequence-based clustering: These methods aim to identify recurring, reusable pattern fragments from long time series and are widely used in applications such as activity recognition and event detection. Representative approaches include Symbolic Aggregate approXimation (SAX) [18], Matrix Profile [19], and Shapelet Transformation [20]. However, most of these techniques rely on heuristic search and sliding window mechanisms, which hinder their scalability to large-scale datasets.

(4) Representation-learning-based clustering: Leveraging the expressive power of deep neural networks, these methods automatically extract latent representations from time series and perform clustering in the learned embedding space. They often employ self-supervised pretraining, contrastive learning, or autoencoding objectives to construct well-structured feature spaces. Common architectures include autoencoders, convolutional neural networks (CNNs) [21], recurrent neural networks (RNNs), and the recently popular Transformer models [22]. Motivated by this line of research, we introduce the Kolmogorov–Arnold Network (KAN) into the time series clustering framework, aiming to enhance both the quality of embeddings and downstream clustering performance by combining strong representational capacity with functional interpretability.

2.2. Differentiable Clustering and End-to-End Optimization

Lafabregue et al. [23] systematically compared various end-to-end time series clustering frameworks and emphasized the importance of differentiable clustering mechanisms in enabling joint optimization between feature extraction and structural partitioning. By incorporating differentiable clustering objectives, models can simultaneously learn both the embedding space and the clustering structure during training, improving clustering discriminability, generalization, and training efficiency. This has become a critical trend in designing high-performance clustering systems.

The Deep Embedded Clustering (DEC) model introduces a soft assignment strategy and Kullback-Leibler (KL) divergence loss to gradually refine the clustering structure in the latent space [24]. Improved DEC (IDEC) enhances training stability by incorporating a reconstruction term from autoencoders [25]. Joint Unsupervised Learning (JULE) leverages an iterative agglomerative strategy, offering improved interpretability while maintaining differentiability throughout the training process [26].

In the context of shape-aware clustering, k-Shape remains a strong baseline due to its ability to capture morphological consistency. However, its reliance on the non-differentiable normalized cross-correlation (NCC) metric makes it incompatible with deep learning frameworks. To address this limitation, researchers have proposed differentiable alternatives—such as Soft-DTW [27] and DTWNet [28]—or have designed network architectures that approximate shape-driven clustering objectives within a differentiable framework.

In recent years, Transformer-based and contrastive learning frameworks have driven progress in end-to-end time series clustering. SCOTT integrates Transformer layers with temporal convolution and supervised contrastive loss, yielding strong results on UCR datasets; however, its dependence on labeled data restricts its use in unsupervised clustering [29]. RandomNet employs untrained deep neural networks to generate diverse representations and achieves competitive clustering performance across the entire UCR collection, yet the randomness of its features raises concerns regarding interpretability and

stability [30]. CDCC leverages cross-domain contrastive learning to enhance generalization on heterogeneous benchmarks, though it relies on auxiliary domains and introduces extra computational cost [31].

These studies highlight both the promise and the limitations of recent Transformer-based and contrastive methods, motivating the design of models such as KDiscShapeNet, which integrates shape-awareness, structural supervision, and differentiable optimization to provide a more balanced solution for time series clustering.

3. KDiscShapeNet

To enhance the representational capacity and structural preservation in time series clustering, we propose KDiscShapeNet, a structure-aware clustering framework. The model integrates three key components: a Kolmogorov–Arnold Network (KAN) for expressive encoding, a differentiable k-Shape clustering mechanism to retain temporal shape alignment, and a discriminative loss design that enforces clear inter-cluster separation. Together, these components form a unified, end-to-end trainable framework with strong clustering discriminability and interpretability.

3.1. Overall Architecture

The overall architecture of KDiscShapeNet is illustrated in Figure 2. The model takes raw time series as input and encodes each sequence using a multi-layer Kolmogorov–Arnold Network (KAN)–based encoder, referred to as KANEncoder, which captures high-order nonlinear representations. The resulting feature vectors are then passed to a Soft k-Shape Matching module, where they are aligned with a set of learnable prototype cluster centers. The similarity between each sample and the cluster prototypes is computed via a Differentiable NCC Similarity function, followed by a softmax layer that produces the final probabilistic cluster assignments. The model supports end-to-end training by jointly optimizing representation learning and clustering through a composite loss function integrating structural, discriminative, and alignment objectives. The entire framework is differentiable and optimized via backpropagation, offering both expressive power and strong clustering discriminability.

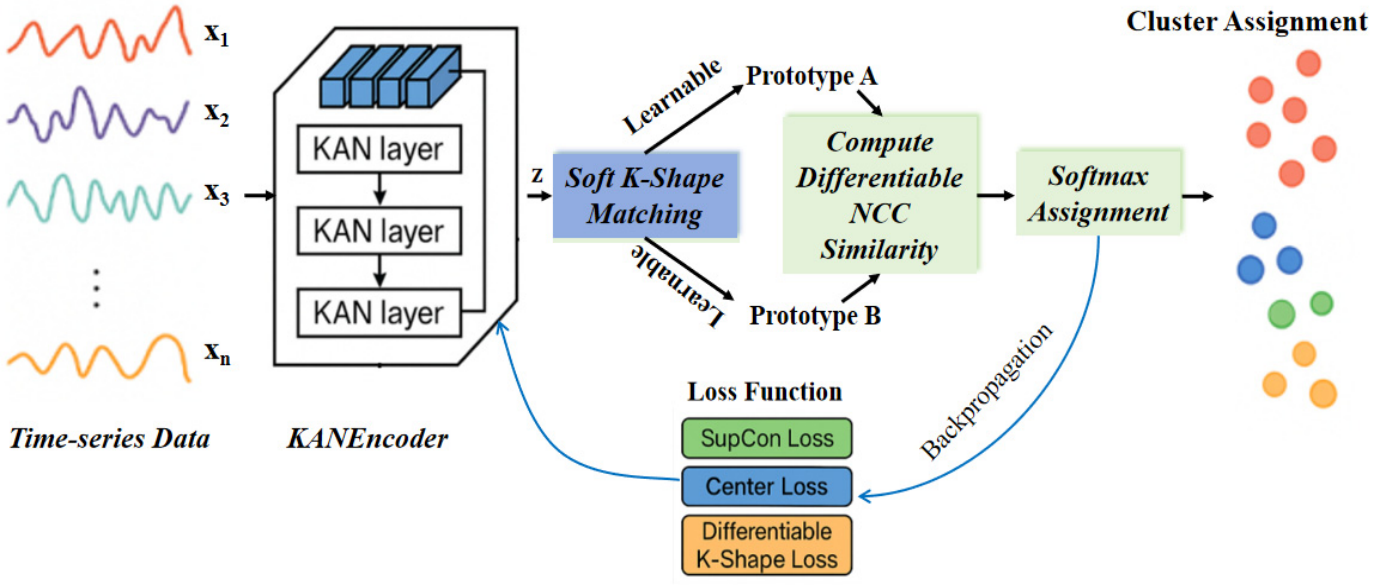


Figure 2. The overall architecture of KDiscShapeNet.

3.2. KANEncoder: High-Order Nonlinear Encoding

To enhance functional approximation and nonlinear expressiveness, KDiscShapeNet adopts Kolmogorov–Arnold Networks (KAN) as its feature encoder. The design of KAN is grounded in the Kolmogorov–Arnold representation theorem [32], which states that any multivariate continuous function can be expressed as a finite superposition of continuous univariate functions. Departing from conventional neural networks that rely on fixed activation functions, KAN introduces learnable one-dimensional kernel spline functions to model nonlinear transformations at each layer. This design improves representational flexibility and ensures numerical smoothness across the network. The general formulation of a KAN layer is expressed as:

$$f^{(l)}(x) = \sum_{j=1}^{m} w_j^{(l)} \cdot \phi_j^{(l)}(x_j) \tag{4}$$

where $x \in R^{d_l}$ is the input vector of the l-th layer with dimension d_l; the activation function $\phi_j^{(l)}(x_j)$ represents the learnable univariate transformation on the j-th input channel, and $w_j^{(l)}$ is the corresponding learnable weight, $f^{(l)}(x)$ represents the output of the l-th layer after aggregating all transformed input channels.

To enhance the expressiveness of learnable functions, each activation function $\phi(x)$ is parameterized using a basis function expansion with B-splines, following an additive and interpretable form:

$$\phi(x) = w_b \cdot b(x) + w_s \cdot \sum_i c_i B_i(x) \tag{5}$$

where $B_i(x)$ denote the predefined B-spline basis functions, and c_i are the corresponding learnable coefficients. The parameters w_b and w_s are scalar weights that control the contribution of the bias term $b(x)$ and the B-spline expansion term, respectively. This formulation allows the activation function $\phi(x)$ to flexibly capture nonlinear transformations by combining interpretable basis functions with learnable parameters, thereby enhancing the expressive power of the KAN encoder.

This architecture endows KAN with greater nonlinear expressiveness and function approximation capability compared to traditional multilayer perceptrons (MLPs). The transformation is formally defined as:

$$z = KANLayer_L \circ \ldots \circ KANLayer_1(x) \tag{6}$$

where $x \in R^T$ denotes the input time series with length T. Each $KANLayer_1(x)$ represents the transformation applied by the l-th KAN layer, and the composition of multiple layers progressively maps the input into a higher-dimensional latent representation. The final output $z \in R^d$ denotes the encoded feature vector in the representation space of dimension d, capturing both local nonlinear variations and global structural dependencies within the sequence.

This design not only preserves the temporal structure of the input but also substantially enhances the model's capacity to capture nonlinear dynamics, making it particularly well-suited for modeling time series characterized by high variability and complex morphological dynamics. The architecture of the KANEncoder is illustrated in Figure 3.

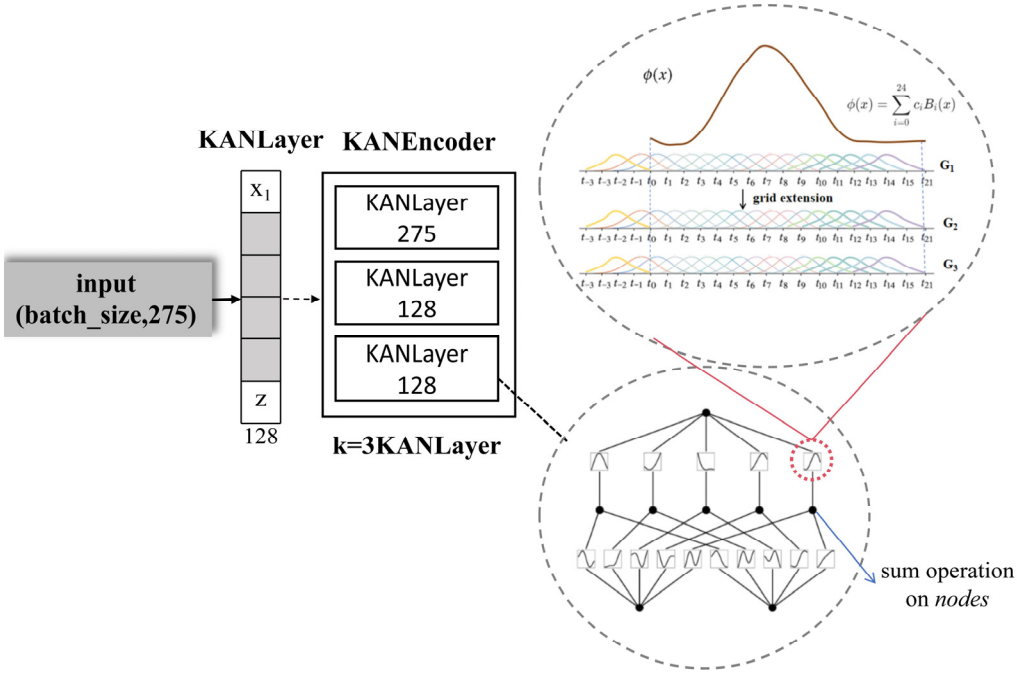


Figure 3. Architecture of the KANEncoder module.

3.3. Differentiable k-Shape Clustering with Soft Assignment

The Differentiable k-Shape clustering module preserves the core idea of the original k-Shape algorithm by maintaining the use of normalized cross-correlation (NCC) as the primary similarity measure. To support end-to-end training with neural network encoders, it introduces a soft assignment strategy and a differentiable clustering loss, enabling seamless gradient propagation during optimization. Specifically, the similarity between a time series representation z_i and a cluster prototype μ_k is computed using NCC. A softmax-based assignment mechanism is employed to transform these similarity scores into probabilistic cluster memberships, ensuring differentiability.

The normalized cross-correlation is formally defined as:

$$\mathrm{NCC}(z,\mu) = \frac{(z-\overline{z})^{\top}(\mu-\overline{\mu})}{||z-\overline{z}||\cdot||\mu-\overline{\mu}||} \tag{7}$$

where $z \in R^d$ denotes the encoded feature vector of the input time series, and $\mu \in R^d$ represents the corresponding prototype vector. The terms $\overline{z}$ and $\overline{\mu}$ indicate the mean values of z and μ, respectively, ensuring a centered operation. The numerator computes the inner product between the centered vectors, while the denominator normalizes by their Euclidean norms, yielding a similarity score in the range $[-1,1]$.

The soft assignment mechanism is defined as:

$$q_{ik} = \frac{exp(NCC\,(z_i,\mu_k)\,/\,\tau)}{\sum_{j=1}^{K} exp(NCC\,(z_i,\mu_k)\,/\,\tau)} L_{kshape} \tag{8}$$

where q_{ik} is the soft assignment probability of the i-th series to the k-th prototype μ_k; $NCC(\cdot,\cdot)$ is the normalized cross-correlation from Equation (7); $\tau > 0$ is the temperature parameter controlling distribution smoothness; and L_{kshape} is the scaling factor from the k-Shape objective.

The corresponding Differentiable k-Shape clustering loss is defined as:

$$L_{kshape} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K} q_{ik}\cdot NCC(z_i,\mu_k) \tag{9}$$

where L_{kshape} is the Differentiable k-Shape loss; q_{ik} is the soft assignment probability of time series z_i to cluster k; $NCC(z_i, \mu_k)$ is the normalized cross-correlation between the sample z_i and cluster prototype μ_k; N and K are the total number of samples and clusters. This loss ensures that time series are aligned in the feature space and the prototypes are refined through optimization, leading to better clustering accuracy.

3.4. Discriminative Enhancement Module

To further enhance the separability between clusters in the embedding space, KDiscShapeNet introduces two discriminative losses: Center Loss [33] and the Supervised Contrastive Loss (SupCon Loss) [34]. Center Loss encourages intra-class compactness by minimizing the distance between sample features and their corresponding class centers in the embedding space. It is formally defined as:

$$L_{center} = \frac{1}{2N}\sum_{i=1}^{N} \parallel z_i - \mu_{y_i} \parallel_2^2 \tag{10}$$

where $z_i \in R^d$ represents the output of the KANEncoder module, μ_{y_i} is the center of the class to which sample i belongs, determined by the label y_i; and N is the total number of samples in the batch; and $\| \cdot \|_2$ denotes the Euclidean norm.

SupCon Loss is employed to enhance local discriminability by pulling similar samples closer and pushing dissimilar samples further apart in the feature space. It is defined as:

$$L_{supcon} = \sum_{i \in I} \frac{-1}{|P(i)|} \sum_{p \in P(i)} log \frac{exp(z_i \cdot z_p / \tau)}{\sum\limits_{a \in A(i)} exp(z_i \cdot z_a / \tau)} \tag{11}$$

where $P(i)$ denotes the set of positive samples for the i-th sample in the batch, $A(i)$ represents the set of all candidate samples, and τ is the temperature parameter.

3.5. Joint Optimization Objective

The overall loss function of KDiscShapeNet integrates the three components described above and is defined as:

$$L_{total} = \lambda_1 \cdot L_{kshape} + \lambda_2 \cdot L_{center} + \lambda_3 \cdot L_{supcon} \tag{12}$$

where $\lambda_1, \lambda_2, \lambda_3$ are hyperparameters used to balance the contribution of each loss term.

The model is trained in an end-to-end fashion, where backpropagation is used to jointly optimize both the encoder parameters and the clustering prototypes. Benefiting from the high-order approximation capacity of KAN and its structure-preserving mechanism, KDiscShapeNet is capable of achieving robust and structure-consistent clustering results on complex and nonlinear time series data.

4. Experimental Evaluation

To comprehensively evaluate the performance of the proposed KDiscShapeNet model in time series clustering tasks, we design a systematic experimental setup covering a wide range of publicly available benchmark datasets. The model is compared against representative clustering methods across different categories, and its effectiveness is quantitatively analyzed using multiple evaluation metrics.

4.1. Settings

All experiments are conducted on a Linux server equipped with an Intel Xeon Gold 6238R CPU, 128 GB RAM, and an NVIDIA A100 GPU. The software environment includes Ubuntu 22.04, Python 3.11, PyTorch 2.1.2, and CUDA 12.1. We set the learning rate to

1×10^{-4}, batch size to 64, and weight decay to 1×10^{-5}, following common practices in related work. The temperature parameter τ in Equations (8) and (11) was fixed to 0.2, as it yielded stable performance across validation experiments. The loss weights $\lambda_1, \lambda_2, \lambda_3$ in Equation (12) were empirically set to (1.0, 0.5, 0.5), which balances the contributions of shape-based clustering, intra-class compactness, and contrastive learning. We found that variations within ± 0.2 around these values did not significantly affect the results, suggesting robustness of the chosen configuration.

We evaluate clustering performance on nine representative time series datasets, as listed in Table 1. These datasets span diverse domains and sequence types, including healthcare, industrial monitoring, activity recognition, and energy systems. The datasets are sourced from the UCR Archive [35] and the ETT benchmark suite. Prior to training, all time series are standardized via Z-score normalization to remove amplitude-induced bias and facilitate fair comparison.

Table 1. Summary of Benchmark Datasets Used for Evaluation.

Dataset	Classes	Train	Test	Description
Beef	5	30	30	Food industry application with limited sample size.
ECG200	2	100	100	Medical diagnostics involving short time series.
GunPoint	2	50	150	Video-based sequences with medium to long durations.
ItalyPowerDemand	2	67	1029	Energy demand forecasting with extremely short sequences.
SonyAIBORobotSurface1	2	20	601	Recognition under noisy environments with high task difficulty.
Trace	4	100	100	Simulated control signals consisting of short sequences.
StarLightCurves	2	10,000	8000	Astronomical observations with long sequences and large sample size.
ETTh1	-	12,288	4096	Power load data with hourly-resolution long sequences.
ETTm1	-	69,120	17,280	Power load data with ultra-long sequences at minute-level granularity.

The nine benchmark datasets summarized in Table 1 were selected to ensure diversity in domain, sequence length, and task characteristics. The UCR datasets (Beef, ECG200, GunPoint, ItalyPowerDemand, SonyAIBORobotSurface1, Trace, and StarLightCurves) encompass applications ranging from food quality control and medical diagnostics to human activity recognition, energy forecasting, robotic sensing, control simulations, and astronomical observations, thereby covering short, medium, and long sequences under both clean and noisy conditions. The ETT datasets (ETTh1 and ETTm1) further represent large-scale energy load forecasting with long and ultra-long horizons, reflecting industrial and utility practices. Collectively, this selection provides a comprehensive foundation for evaluating the proposed method across heterogeneous temporal patterns and real-world application scenarios.

During model training, we adopt a time-window-based view augmentation strategy. For example, the sliding window size on the Trace dataset is set to 24, while for other datasets it is dynamically adjusted based on the sequence length. The KDiscShapeNet model uses a three-layer KAN-based encoder with a network architecture of [275, 128, 64], where each layer is followed by a Rectified Linear Unit (ReLU) activation function. Training is performed using the Adam optimizer with an initial learning rate of 1×10^{-3}, a batch size of 64, and a total of 100 epochs. The number of augmented views is set to 4, the

temperature coefficient for contrastive learning is 0.1, and the learning rate is scheduled using cosine annealing.

To evaluate clustering performance, we adopt three widely used unsupervised clustering metrics: Adjusted Rand Index (ARI),Normalized Mutual Information (NMI) and Silhouette Coefficient (SC). These metrics assess clustering quality from different perspectives: label agreement, information-theoretic consistency, and spatial distribution compactness, respectively. The corresponding formulas are defined as follows:

(1) ARI

$$\mathrm{ARI} = \frac{\mathrm{TP}+\mathrm{TN}-\mathbb{E}[\mathrm{TP}+\mathrm{TN}]}{\binom{n}{2}-\mathbb{E}[\mathrm{TP}+\mathrm{TN}]} \tag{13}$$

In this context, True Positives (TP) refer to the number of time series pairs that are correctly assigned to the same cluster, while True Negatives (TN) represent the number of pairs correctly assigned to different clusters. The term $\binom{n}{2}$ denotes the total number of possible unique pairs among n samples.

(2) NMI

$$\mathrm{NMI} = \frac{\sum_{i=1}^{M}\sum_{j=1}^{M} N_{ij}\log\left(\frac{N\cdot N_{ij}}{|G_i|\cdot|A_j|}\right)}{\sqrt{\left(\sum_{i=1}^{M}|G_i|\log\frac{|G_i|}{N}\right)\left(\sum_{j=1}^{M}|A_j|\log\frac{|A_j|}{N}\right)}} \tag{14}$$

where N represents the total number of time series, $|G_i|$ and $|A_j|$ represent the number of elements in sets G_i and A_j, and $N_{ij} = |G_i \cap A_j|$ denotes the number of time series that belong to both G_i and A_j.

(3) SC

The Silhouette Coefficient is defined as:

$$s(i) = \frac{b(i)-a(i)}{\max\{a(i),b(i)\}} \tag{15}$$

where $a(i)$ denotes the average distance from sample i to all other samples in the same cluster, and $b(i)$ denotes the average distance from sample i to all samples in the nearest cluster. The value $s(i) \in [-1,1]$ indicates how well the sample fits into its own cluster. A higher value of $s(i)$ indicates better cluster cohesion and separation. The overall Silhouette Score (S) is the average of all $s(i)$ values [36].

$$S = \frac{1}{n}\sum_{i=1}^{n} s(i) \tag{16}$$

4.2. Comparative Experiment

To validate the effectiveness of KDiscShapeNet, we systematically compare KDiscShapeNet with representative time series clustering methods from three widely adopted paradigms. This evaluation aims to assess the model's performance advantages and adaptability across different modeling frameworks.

(1) Traditional time series clustering methods

These rely on handcrafted distance measures and are valued for simplicity and interpretability. We select three classical baselines: KMeans, a general-purpose algorithm using Euclidean distance; DBA-KMeans, which incorporates DTW to align non-synchronized sequences, suitable for industrial and biomedical tasks; and k-Shape, a shape-based method using normalized cross-correlation for phase invariance and global structural modeling.

(2) Representation learning combined with clustering

These approaches use deep encoders for feature extraction followed by conventional clustering. We consider TS2Vec + KMeans, where TS2Vec captures multi-scale shape variations via hierarchical contrastive learning, and TNC + KMeans, where Temporal Neighborhood Contrast enhances local discriminability through temporal proximity views.

(3) Deep clustering methods

These approaches integrate clustering mechanisms with deep networks in an end-to-end fashion and represent the cutting edge of unsupervised learning research. We consider the following methods: DTCR, which combines autoencoder-based reconstruction with clustering optimization to enforce compact and separable latent representations. TCGAN, which introduces adversarial training to enhance modeling of rare structures and inter-cluster boundaries in time series. k-Graph, a graph-based framework that learns topology-aware embeddings for interpretable time series clustering, serving as a state-of-the-art example of graph-driven unsupervised learning.

All selected baselines are representative of their respective paradigms and have been widely used on standard benchmarks such as UCR and ETT. We report Accuracy (ACC), Adjusted Rand Index (ARI), Normalized Mutual Information (NMI), and Silhouette Coefficient (SC). The experimental results are summarized in Tables 2 and 3.

In Figure 4a, the black dashed lines indicate the mean performance of each model, while the horizontal red dashed line denotes the best overall mean across all methods. Figure 4b presents the Critical Difference diagrams for three key performance metrics, where the numbers on the horizontal axis represent the average rank of each method, and the values above the lines correspond to the mean scores for the respective metric.

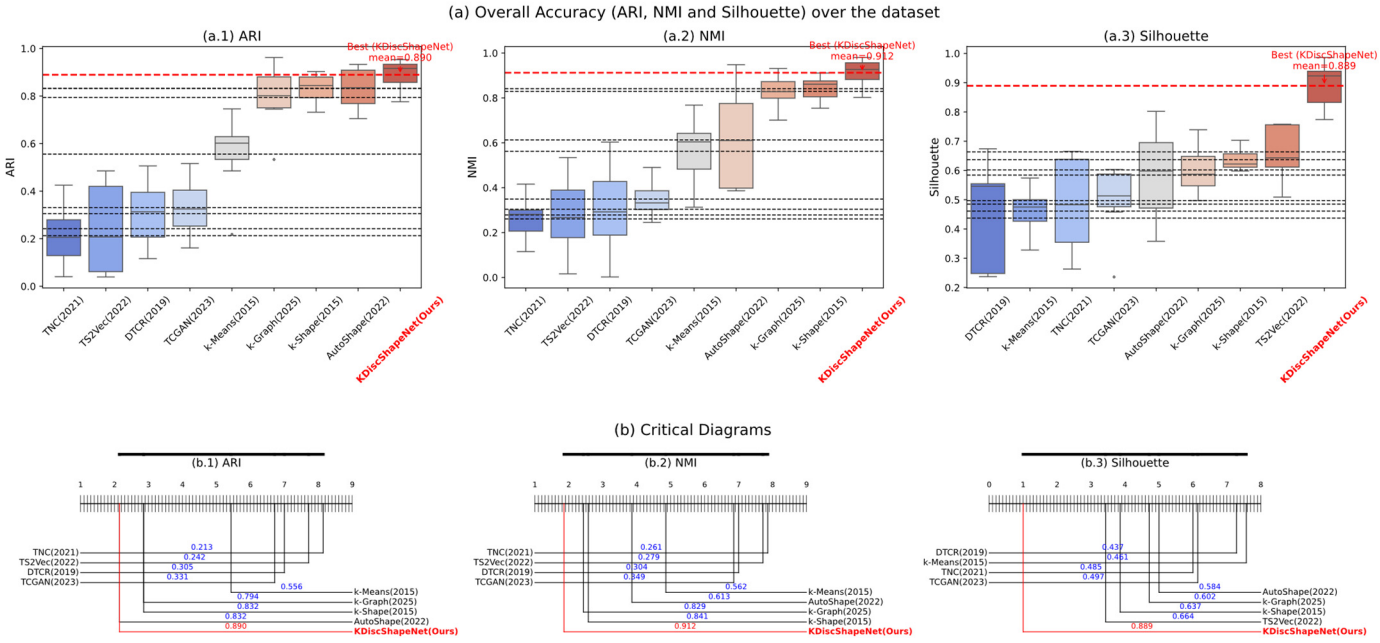


Figure 4. Comparison of KDiscShapeNet with other models. (**a**) Box plots of ARI, NMI, and Silhouette scores across datasets. (**b**) Critical difference diagrams for the three metrics, comparing representative methods including k-Shape [5], AutoShape [6], TS2Vec [7], DTCR [10], k-Graph [12], TCGAN [13], and TNC [9].

Table 2. Experimental Results of KDiscShapeNet on Labeled Datasets.

Data	Standard Metrics	k-Shape (2015)	DTCR (2019)	TNC (2021)	TS2Vec (2022)	AutoShape (2022)	TCGAN (2023)	CDCC (2024)	k-Graph (2025)	KDiscShapeNet (Ours)
Beef	ARI	0.786	0.116	0.425	0.077	0.778	0.426	0.781	0.859	0.776
	NMI	0.798	0.267	0.194	0.323	0.386	0.352	0.792	0.842	0.962
	Silhouette	0.598	0.251	0.665	0.613	0.358	0.603	0.511	0.612	0.986
ECG200	ARI	0.903	0.192	0.311	0.039	0.759	0.325	0.764	0.533	0.946
	NMI	0.881	0.292	0.219	0.267	0.393	0.245	0.775	0.821	0.802
	Silhouette	0.621	0.546	0.612	0.509	0.426	0.574	0.498	0.563	0.947
GunPoint	ARI	0.844	0.462	0.206	0.416	0.705	0.516	0.845	0.962	0.921
	NMI	0.861	0.110	0.115	0.016	0.403	0.420	0.853	0.931	0.862
	Silhouette	0.673	0.556	0.664	0.757	0.598	0.601	0.572	0.739	0.786
ItalyPowerDemand	ARI	0.799	0.221	0.248	0.046	0.835	0.281	0.621	0.756	0.953
	NMI	0.812	0.102	0.292	0.108	0.635	0.331	0.633	0.701	0.980
	Silhouette	0.641	0.674	0.483	0.755	0.517	0.513	0.401	0.497	0.774
SonyAIBORobotSurface1	ARI	0.864	0.313	0.076	0.207	0.895	0.161	0.693	0.745	0.875
	NMI	0.870	0.347	0.416	0.247	0.610	0.274	0.702	0.828	0.951
	Silhouette	0.622	0.237	0.378	0.610	0.676	0.494	0.455	0.587	0.879
Trace	ARI	0.896	0.328	0.182	0.424	0.922	0.225	0.872	0.801	0.916
	NMI	0.911	0.508	0.281	0.455	0.948	0.332	0.881	0.777	0.927
	Silhouette	0.903	0.192	0.311	0.039	0.759	0.325	0.596	0.533	0.946
StarLightCurves	ARI	0.703	0.245	0.331	0.643	0.802	0.458	0.804	0.532	0.931
	NMI	0.732	0.506	0.040	0.485	0.933	0.382	0.816	0.902	0.840
	Silhouette	0.754	0.603	0.308	0.534	0.915	0.490	0.538	0.903	0.902

Table 3. Experimental Results of KDiscShapeNet on Unlabeled Datasets.

Data	Standard Metrics	T-Loss (2019)	DTS-Cluster (2021)	TNC + KMeans (2021)	TS2Vec + KMeans (2022)	DTCC (2022)	AutoShape (2022)	KDiscShapeNet (Ours)
ETTh1	ACC	0.762	0.802	0.756	0.771	0.788	0.795	0.926
	ARI	0.690	0.733	0.672	0.685	0.712	0.721	0.751
	NMI	0.710	0.754	0.695	0.709	0.738	0.743	0.917
	Silhouette	0.414	0.451	0.390	0.403	0.428	0.436	0.619
ETTm1	ACC	0.754	0.798	0.747	0.763	0.779	0.785	0.904
	ARI	0.678	0.729	0.658	0.671	0.701	0.715	0.985
	NMI	0.702	0.744	0.682	0.695	0.723	0.731	0.981
	Silhouette	0.398	0.438	0.375	0.391	0.412	0.427	0.741

Figure 5 illustrates the clustering process of the KDiscShapeNet model on four representative datasets: Beef, GunPoint, StarLightCurves, and Trace, respectively. Figure 6 shows the clustering results of KDiscShapeNet on the unlabeled ETT dataset, highlighting its applicability in fully unsupervised settings.

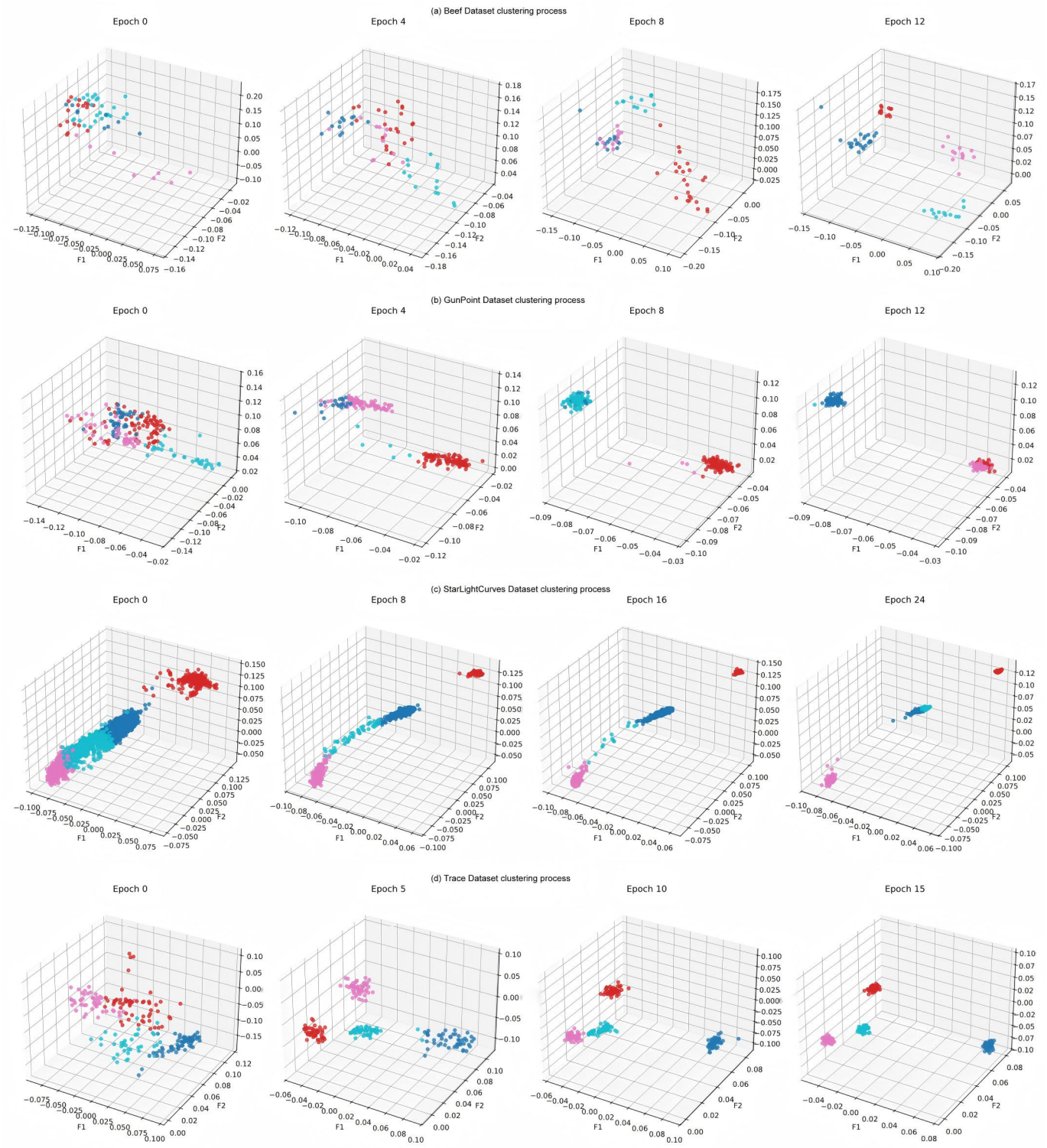


Figure 5. Clustering process of KDiscShapeNet on the unlabeled dataset.

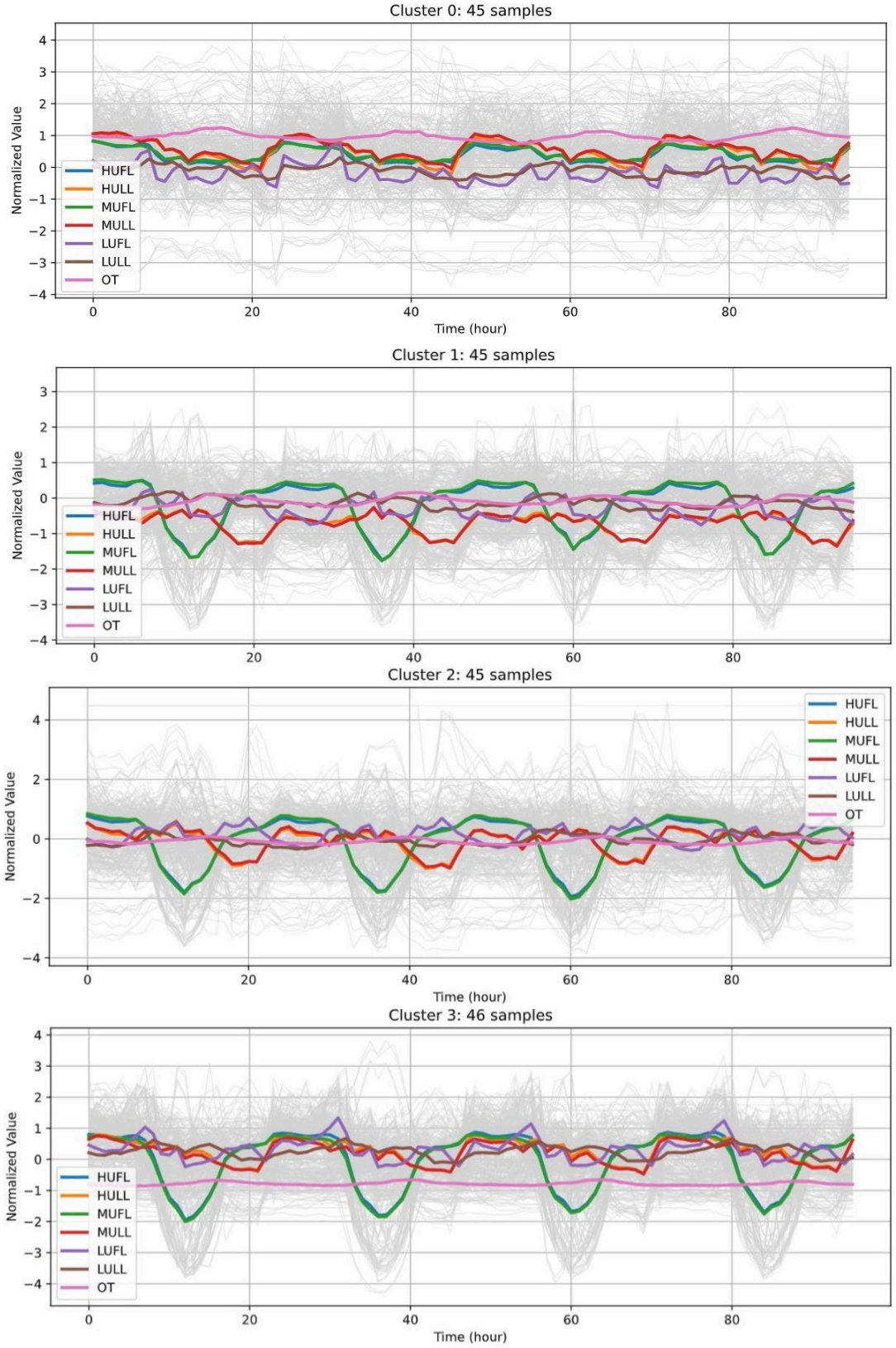


Figure 6. Clustering results of KDiscShapeNet on ETTh1 dataset.

Based on the experimental results, KDiscShapeNet demonstrates superior clustering performance across multiple typical time series clustering tasks. In the box plot of average scores shown in Figure 4a, KDiscShapeNet outperforms the comparison methods across all metrics with a smaller range of variation, indicating better robustness. The Critical Differ-

ence diagram in Figure 4b provides a statistical comparison of the rankings across several datasets, confirming the statistical significance of KDiscShapeNet's leading performance. KDiscShapeNet achieves high scores even in tasks with ultra-long sequences and noisy backgrounds. Its structure-aware design and discriminative subsequence modeling mechanism offer significant advantages when dealing with complex dynamic patterns. In cases where the sequences have extremely small sample sizes and highly localized change patterns, the model performs well overall but shows slightly higher sensitivity to initialization and the modeling of positive-negative samples compared to TS2Vec and AutoShape.

After validating the performance superiority of KDiscShapeNet through comparative experiments, we further employed Wilcoxon signed-rank tests to assess the robustness of the improvements from a statistical perspective. Figure 7 illustrates the results on ARI and NMI, reported as $-\log_{10}(p)$, with the red dashed line indicating the significance threshold of $p = 0.05$. The findings demonstrate that KDiscShapeNet achieves significant or highly significant improvements over most baseline methods, particularly against deep clustering models such as TNC, TS2Vec, and TCGAN, where significance values exceed 10, highlighting consistent and robust advantages. By contrast, only a few traditional methods (e.g., kGraph and kShape) exhibit marginal significance on certain metrics, suggesting residual statistical uncertainty. The Wilcoxon test results confirm that the improvements of KDiscShapeNet are not incidental but represent a statistically reliable and consistent advantage.

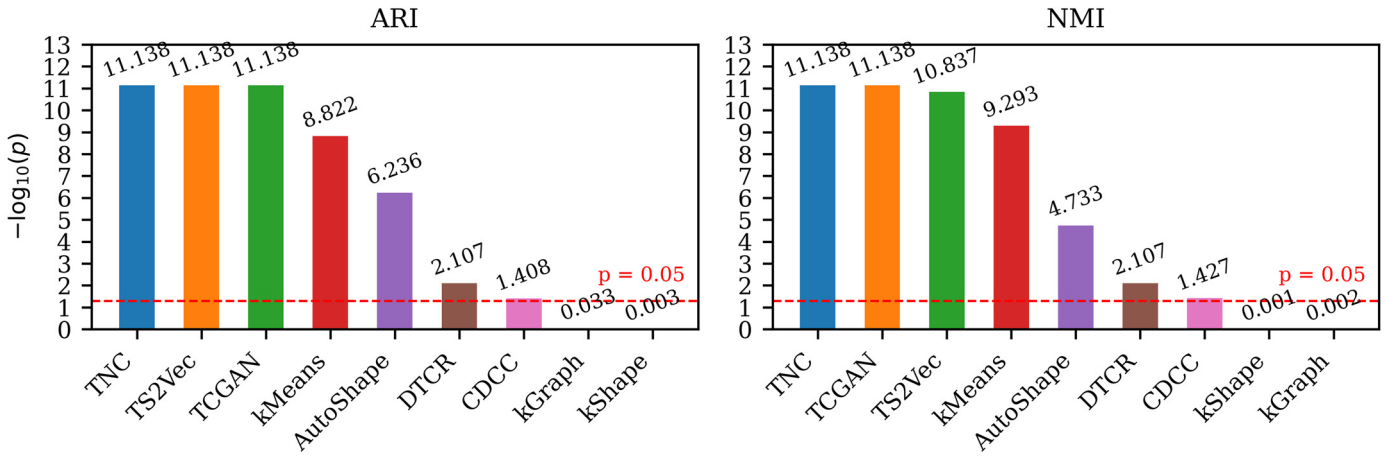


Figure 7. Wilcoxon test results of KDiscShapeNet against baselines on ARI and NMI (red line denotes $p = 0.05$).

Overall, KDiscShapeNet excels in most time series clustering scenarios, particularly those involving medium-to-long sequences, multi-pattern dynamics, and distinct structural boundaries. It strikes a good balance between traditional and deep learning models, offering both stability and strong generalization capabilities.

4.3. Ablation Study

To comprehensively evaluate the effectiveness of key components in the proposed KDiscShapeNet model, we conduct a systematic ablation study on the Trace dataset. The study investigates different combinations of the following modules: the KAN encoder, Supervised Contrastive Loss (SupCon Loss), and the prototype compactness constraint (Center Loss). The detailed experimental settings and module configurations are summarized in Table 4, ✔ indicates that the module is included, while ✗ indicates that the module is not included.

Table 4. Ablation Study Design on the Trace Dataset.

Experiment	KAN Encoder	SupCon Loss	Center Loss	ARI	NMI	Silhouette
A	✗	✗	✗	0.594	0.689	0.307
B	✔	✗	✗	0.541	0.509	0.321
C	✔	✔	✗	0.651	0.735	0.464
D	✔	✗	✔	0.934	0.940	0.879
E	✗	✔	✔	0.950	0.958	0.781
F	✗	✔	✗	0.912	0.925	0.613
G	✗	✗	✔	0.650	0.735	0.846
H	✔	✔	✔	0.916	0.927	0.931

Experimental results indicate that the Supervised Contrastive Loss (SupConLoss) is a critical component for enhancing clustering discriminability. Even in the absence of any structural information (Experiment F), SupConLoss alone elevates ARI to 0.912 and NMI to 0.925, significantly outperforming the baseline model (Experiment A). In contrast, Center Loss primarily contributes to improving cluster compactness and separability. When applied independently (Experiment G), it achieves a Silhouette Coefficient (SC) of 0.846, demonstrating strong capability in reinforcing inter-cluster boundaries.

The structure-aware encoder (KANEncoder), when used alone (Experiment B), actually leads to performance degradation (ARI drops to 0.541), likely due to the absence of targeted supervisory signals. However, when combined with auxiliary supervision (e.g., SupCon in Experiment C), ARI improves to 0.651, suggesting a moderate synergistic effect.

When SupConLoss and Center Loss are jointly applied (Experiment E), the model achieves the best overall performance across all three metrics—ARI: 0.950, NMI: 0.958, SC: 0.781—highlighting the strong complementarity between semantic discriminative enhancement and inter-cluster distribution regularization. Another notable configuration, Experiment D (KAN + Center), achieves the highest SC score of 0.879, further demonstrating that structural information, when guided by center constraints, can effectively sharpen cluster boundaries. Figure 8. summarizes these results in a radar chart, offering a visual comparison of each module's contribution to the evaluation metrics.

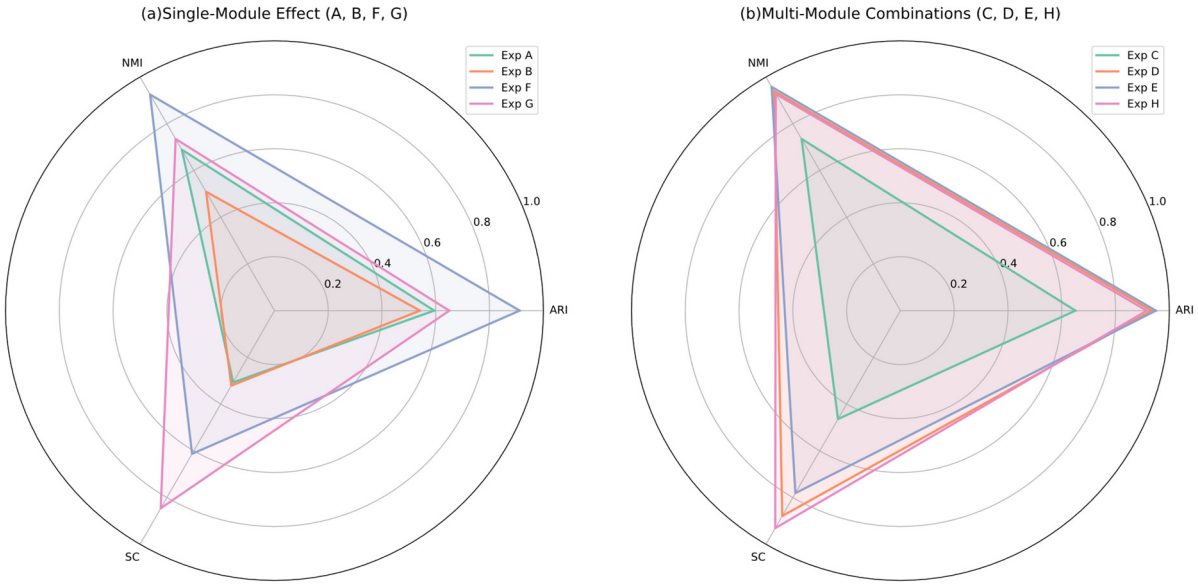


Figure 8. Radar chart of ablation study results. (**a**) Single-module performance comparison. (**b**) Multi-module combination analysis with radar plots.

From the single-module analysis shown in Figure 8a, we observe that introducing the KAN encoder alone (Experiment B) does not lead to significant improvements in ARI and NMI scores. In contrast, incorporating Supervised Contrastive Loss (SupCon) and Center Loss results in substantial gains in Silhouette Coefficient (SC), achieving values of 0.613 and 0.846, respectively. These results confirm the complementary roles of the two loss functions in enhancing both the discriminability and structural compactness of the learned representations.

The ablation study indicates that the KAN encoder alone (Experiment B) leads to performance degradation. This phenomenon can be attributed to its strong nonlinear approximation ability, which, without auxiliary supervision, tends to overfit local fluctuations and cause gradient instability. Unlike CNNs or RNNs that incorporate locality or sequential biases, KAN operates in a highly flexible functional space, making it more sensitive to noise and initialization. Consequently, additional regularization is required to stabilize training and ensure reliable clustering performance. Possible remedies include the use of lightweight regularization methods (e.g., dropout or spectral normalization), hybrid encoder designs that combine KAN with convolutional or attention modules to introduce inductive bias, and self-supervised pretraining strategies to provide robust initialization. These directions highlight the necessity of regularization when deploying KAN encoders and suggest avenues for improving their independent applicability in unsupervised settings.

In the multi-module combinations illustrated in Figure 8b, performance improvements become more pronounced. Specifically: Experiment D (KAN + Center) and Experiment E (SupCon + Center) outperform their corresponding single-module variants across all evaluation metrics, indicating a positive synergistic effect between structural modeling and discriminative loss design. The best overall performance is achieved in Experiment H (full module combination), reaching 0.916 (ARI), 0.927 (NMI), and 0.931 (SC). The corresponding radar plot demonstrates near-complete coverage, validating the complementary benefits of combining the KAN encoder with both contrastive constraints.

These findings confirm that the three components in KDiscShapeNet—namely discriminative representation learning, structure-aware encoding, and cluster separation—contribute collaboratively to unsupervised clustering. The ablation results highlight the importance of multi-objective optimization in structure-aware time series clustering and further underscore the interpretable complementarity among the proposed modules.

4.4. Complexity and Efficiency Analysis

The models exhibit distinct trade-offs between computational efficiency and clustering performance. Traditional methods such as k-Shape and k-Graph are lightweight ($O(n^2T)$time,$O(nT)$space) but scale poorly with large datasets. Deep clustering frameworks including DTCR, TNC, TS2Vec, AutoShape, and TCGAN significantly improve accuracy but incur higher cost ($O(nTd^2)$time,$O(nd)$space), with AutoShape adding an autoencoder branch and TCGAN suffering the heaviest adversarial training burden. CDCC reduces redundancy, achieving moderate complexity ($O(nTd)$time,$O(nd)$space). KAN, emphasizing structural representation with graph propagation, shows the largest overhead ($O(nTd^2)$time, high memory for adjacency storage). By contrast, the proposed KDiscShapeNet maintains controlled complexity ($O(nTd + nk)$time, $O(nd + k^2)$space), comparable to CDCC but far more efficient than KAN, while achieving state-of-the-art clustering performance, where n is the number of sequences and T is sequence length.

For the complete UCR collection (single GPU, batch size = 64, hidden width = 128, two KAN layers, and spline grid $G = 32$), KDiscShapeNet exhibits an average peak GPU memory usage of approximately 0.55 GB, with the worst-case peak reaching 1.35 GB.

Specifically, the memory requirement is around 0.15 GB for short sequences ($T < 500$), around 0.50 GB for medium-length sequences ($500 < T < 1000$), and up to 1.35 GB for long sequences ($1500 \leq T \leq 3000$). Under the same configuration, the KAN baseline achieves the shortest runtime (405.8 s) with a comparable memory footprint, which primarily reflects its simplified training procedure and should therefore be regarded as an empirical lower bound. In contrast, the classical k-Shape model incurs the longest runtime (5412.1 s) on CPU and does not utilize GPU memory, primarily due to computationally expensive prototype alignment and repeated correlation operations. The relatively longer runtime of KDiscShapeNet compared to the KAN baseline can be attributed to the integration of contrastive losses and structure-aware encoding, which introduce additional tensor operations while improving the quality of learned representations. KDiscShapeNet demonstrates a favorable trade-off: although its runtime is moderately longer than that of the simplified KAN baseline, it maintains compact GPU memory usage and significantly reduces training time relative to k-Shape, thereby achieving both computational efficiency and superior clustering performance.

5. Conclusions and Future Work

This paper addresses three key challenges in time series clustering—limited representation learning, ambiguous cluster boundaries, and lack of end-to-end optimization—by proposing KDiscShapeNet, a structure-aware clustering model with supervised contrastive mechanisms. The main innovations are as follows:

(1) Unified end-to-end framework: KDiscShapeNet integrates a Kolmogorov–Arnold Network (KAN) encoder with a Differentiable k-Shape clustering module, achieving joint optimization between feature representation and cluster assignment.
(2) Dual discriminative enhancement: The combination of Center Loss and Supervised Contrastive Loss enforces intra-cluster compactness and inter-cluster separability, strengthening discriminative capability.
(3) Structure-aware modeling: The design explicitly captures dynamic shape variations and local structural heterogeneity, improving adaptability to diverse temporal patterns.

Extensive experiments on diverse public time series benchmarks, including UCR and ETT, demonstrate that KDiscShapeNet achieves superior clustering performance in terms of accuracy, stability, and generalization. The model performs particularly well on medium-to-long sequences with clear structural boundaries and heterogeneous dynamics. Moreover, ablation studies confirm the complementary contributions of the KAN encoder, SupCon loss, and Center loss, validating the synergistic effect of multi-objective joint optimization. Visualization and statistical significance analyses further verify the robustness and interpretability of the framework, underscoring its reliability for real-world applications.

Despite the encouraging results, several limitations should be acknowledged. First, the proposed framework exhibits sensitivity to initialization when applied to ultra-short sequences, which may reduce its robustness in scenarios with highly limited temporal context. Potential mitigation strategies include adopting pretraining on large-scale auxiliary datasets to provide more stable initial representations, designing sequence augmentation schemes (e.g., temporal warping or subsequence recomposition) to alleviate variance, and incorporating lightweight convolutional or recurrent front-end modules to enrich local temporal cues before KAN encoding. From an architectural perspective, hybrid designs that integrate attention mechanisms or graph neural networks may further enhance stability by explicitly modeling inter-sequence dependencies and improving the calibration of positive and negative samples in contrastive learning. Moreover, adaptive initialization strategies such as meta-learning–based warm starts could provide more consistent convergence across heterogeneous datasets. Exploring these directions would not only address the sensitivity

issue but also broaden the applicability of KDiscShapeNet in real-world scenarios with short, noisy, or resource-constrained time series.

Author Contributions: Conceptualization, X.C. and C.S.; methodology, Y.J.; software, Y.J.; validation, X.C., C.S. and Y.Z.; formal analysis, X.C.; investigation, Y.J.; resources, C.S.; data curation, C.S.; writing—original draft preparation, Y.J.; writing—review and editing, Y.Z. and C.S.; visualization, Y.J.; supervision, X.C. and C.S.; project administration, C.S.; funding acquisition, C.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key R&D Program Project of China, grant number 2024YFB4709100; the National Natural Science Foundation of China, grant number 62273337; the Natural Science Foundation of Liaoning Province, China, grant number 2025-BS-0297; the Department of Education of Liaoning Province, China, grant number LJ212410142147; and the Fundamental Research Funds for the Central Universities, China, grant number N2523029. The APC was funded by the Department of Education of Liaoning Province, China.

Data Availability Statement: Data are contained within the article. The data presented in this study can be requested from the authors.

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

References

1. Palpanas, T.; Beckmann, V. Report on the first and second interdisciplinary time series analysis workshop (ITISA). *ACM SIGMOD Rec.* **2019**, *48*, 36–40. [CrossRef]
2. Bagnall, A.J.; Cole, R.L.; Palpanas, T.; Zoumpatianos, K. Data series management (Dagstuhl Seminar 19282). *Dagstuhl Rep.* **2019**, *9*, 24–39. [CrossRef]
3. Huang, Z.; Hao, H.; Du, L. Exploring the explainability of time series clustering: A review of methods and practices. In Proceedings of the 18th ACM International Conference on Web Search and Data Mining (WSDM 2025), Singapore, 10–14 March 2025; pp. 1005–1007.
4. Likas, A.; Vlassis, N.; Verbeek, J. The global k-means clustering algorithm. *Pattern Recognit.* **2003**, *36*, 451–461. [CrossRef]
5. Paparrizos, J.; Gravano, L. k-shape: Efficient and accurate clustering of time series. *ACM SIGMOD Rec.* **2016**, *45*, 69–76. [CrossRef]
6. Li, G.; Choi, B.; Xu, J.; Bhowmick, S.; Mah, D.; Wong, G. AUTOSHAPE: An autoencoder-shapelet approach for time series clustering. *arXiv* **2022**, arXiv:2208.04313. [CrossRef]
7. Yue, Z.; Wang, Y.; Duan, J.; Yang, T.; Huang, C.; Tong, Y.; Xu, B. TS2Vec: Towards universal representation of time series. In Proceedings of the AAAI Conference on Artificial Intelligence (AAAI 2022), Virtual Event, 22 February–1 March 2022; pp. 8980–8987.
8. Yang, C.H.H.; Tsai, Y.-Y.; Chen, P.-Y. Voice2Series: Reprogramming acoustic models for time series classification. *arXiv* **2021**, arXiv:2104.09577.
9. Zhong, Y.; Huang, D.; Wang, C.-D. Deep temporal contrastive clustering. *Neural Process. Lett.* **2023**, *55*, 7869–7885. [CrossRef]
10. Ma, Q.; Zheng, J.; Li, S.; Cottrell, G.W. Learning representations for time series clustering. In Proceedings of the 33rd International Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, BC, Canada, 8–14 December 2019; Curran Associates Inc.: Red Hook, NY, USA, 2019; p. 339.
11. Javed, A.; Rizzo, D.M.; Lee, B.S.; Gramling, R. Somtimes: Self organizing maps for time series clustering and its application to serious illness conversations. *Data Min. Knowl. Discov.* **2024**, *38*, 813–839. [CrossRef] [PubMed]
12. Boniol, P.; Tiano, D.; Bonifati, A.; Palpanas, T. k-graph: A graph embedding for interpretable time series clustering. *IEEE Trans. Knowl. Data Eng.* **2025**, *37*, 2680–2694. [CrossRef]
13. Huang, F.; Deng, Y. TCGAN: Convolutional generative adversarial network for time series classification and clustering. *Neural Netw.* **2023**, *165*, 868–883. [CrossRef] [PubMed]
14. Gao, P.; Yang, X.; Zhang, R.; Huang, K.; Goulermas, J.Y. Explainable tensorized neural ordinary differential equations for arbitrary-step time series prediction. *IEEE Trans. Knowl. Data Eng.* **2023**, *35*, 5837–5850. [CrossRef]
15. Esling, P.; Agon, C. Time-series data mining. *ACM Comput. Surv.* **2012**, *45*, 12. [CrossRef]
16. Yang, J.; Ning, C.; Deb, C.; Zhang, F.; Cheong, D.; Lee, S.E.; Sekhar, C.; Tham, K.W. k-Shape clustering algorithm for building energy usage patterns analysis and forecasting model accuracy improvement. *Energy Build.* **2017**, *146*, 27–37. [CrossRef]
17. Zhang, X.; Zhu, Y.; Cheng, W.; Wang, Y. Kernelized k-shape clustering for time series. In Proceedings of the 33rd AAAI Conference on Artificial Intelligence (AAAI 2019), Honolulu, HI, USA, 27 January–1 February 2019; pp. 5683–5690.

18. Lin, J.; Keogh, E.; Lonardi, S.; Chiu, B. A symbolic representation of time series, with implications for streaming algorithms. In Proceedings of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery (DMKD 2003), San Diego, CA, USA, 13 June 2003; pp. 2–11.
19. Yeh, C.C.M.; Zhu, Y.; Ulanova, L.; Begum, N.; Ding, Y.; Dau, H.A.; Silva, D.F.; Mueen, A.; Keogh, E. Matrix profile I: All pairs similarity joins for time series: A unifying view that includes motifs, discords and shapelets. In Proceedings of the 2016 IEEE International Conference on Data Mining (ICDM 2016), Barcelona, Spain, 12–15 December 2016; pp. 1317–1322.
20. Hills, J.; Lines, J.; Baranauskas, E.; Mapp, J.; Bagnall, A. Classification of time series by shapelet transformation. *Data Min. Knowl. Discov.* **2014**, *28*, 851–881. [CrossRef]
21. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [CrossRef]
22. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In Proceedings of the 31st International Conference on Neural Information Processing Systems (NeurIPS 2017), Long Beach, CA, USA, 4–9 December 2017; pp. 6000–6010.
23. Lafabregue, B.; Weber, J.; Gançarski, P.; Forestier, G. End-to-end deep representation learning for time series clustering: A comparative study. *Data Min. Knowl. Discov.* **2022**, *36*, 29–81. [CrossRef]
24. Xie, J.; Girshick, R.; Farhadi, A. Unsupervised deep embedding for clustering analysis. In Proceedings of the 33rd International Conference on Machine Learning (ICML 2016), New York, NY, USA, 19–24 June 2016; pp. 478–487.
25. Guo, X.; Gao, L.; Liu, X.; Yin, J. Improved deep embedded clustering with local structure preservation. In Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI 2017), Melbourne, Australia, 19–25 August 2017; pp. 1753–1759.
26. Yang, B.; Fu, X.; Sidiropoulos, N.D.; Hong, M. Towards k-means-friendly spaces: Simultaneous deep learning and clustering. In Proceedings of the 34th International Conference on Machine Learning (ICML 2017), Sydney, Australia, 6–11 August 2017; pp. 3861–3870.
27. Cuturi, M.; Blondel, M. Soft-DTW: A differentiable loss function for time-series. In Proceedings of the 34th International Conference on Machine Learning (ICML 2017), Sydney, Australia, 6–11 August 2017; pp. 894–903.
28. Cai, X.; Xu, T.; Yi, J.; Huang, J.; Rajasekaran, S. DTWNet: A dynamic time warping network. In Proceedings of the 33rd International Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, BC, Canada, 8–14 December 2019; pp. 10512–10522.
29. Liu, Y.; Wijewickrema, S.; Bester, C.; Song, Y.; Kasmarik, K.; Zhou, H. Time Series Representation Learning with Supervised Contrastive Temporal Transformer. In Proceedings of the 2024 International Joint Conference on Neural Networks (IJCNN), Yokohama, Japan, 30 June–5 July 2024; IEEE: Piscataway, NJ, USA, 2024; pp. 1–8.
30. Li, X.; Xi, W.; Lin, J. RandomNet: Clustering Time Series Using Untrained Deep Neural Networks. *Data Min. Knowl. Disc.* **2024**, *38*, 3473–3502. [CrossRef]
31. Peng, F.; Luo, J.; Lu, X.; Wu, F.; Xu, Y.; Liu, H.; Wang, L. Cross-Domain Contrastive Learning for Time Series Clustering. In Proceedings of the AAAI Conference on Artificial Intelligence, Vancouver, BC, Canada, 20–27 February 2024; AAAI Press: Palo Alto, CA, USA, 2024; Volume 38, pp. 8921–8929.
32. Liu, Z.; Wang, Y.; Vaidya, S. KAN: Kolmogorov–Arnold networks. *arXiv* **2024**, arXiv:2404.19756. [CrossRef]
33. Wen, Y.; Zhang, K.; Li, Z.; Qiao, Y. A discriminative feature learning approach for deep face recognition. In Proceedings of the 14th European Conference on Computer Vision (ECCV 2016), Amsterdam, The Netherlands, 11–14 October 2016; Leibe, B., Matas, J., Sebe, N., Welling, M., Eds.; Springer: Cham, Switzerland, 2016; pp. 499–515.
34. Khosla, P.; Teterwak, P.; Wang, C.; Sarna, A.; Tian, Y.; Isola, P.; Maschinot, A.; Liu, C.; Krishnan, D. Supervised contrastive learning. In Proceedings of the 34th International Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, BC, Canada, 6–12 December 2020; Article 1567.
35. Dau, H.A.; Bagnall, A.; Kamgar, K.; Yeh, C.C.M.; Zhu, Y.; Gharghabi, S.; Ratanamahatana, C.A.; Keogh, E. The UCR time series archive. *IEEE/CAA J. Autom. Sin.* **2019**, *6*, 1293–1305. [CrossRef]
36. Zhang, Q.; Wu, J.; Zhang, P.; Long, G.; Zhang, C. Salient subsequence learning for time series clustering. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *41*, 2193–2207. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.