TGCD: A Framework for Generalized Category Discovery in Time-Series Data

Chandan Gautam[1*], **Lew Choon Hean**[3], **Ankit Das**[2], **Xiaoli Li**[1], **Savitha Ramasamy**[1]

[1]Institute for Infocomm Research (I[2]R), Agency for Science, Technology and Research (A*STAR), Singapore
[2]Institute of High Performance Computing (IHPC), Agency for Science, Technology and Research (A*STAR), Singapore
[3]National University of Singapore, Singapore
gautamc@a-star.edu.sg, e0958179@u.nus.edu, dasak@a-star.edu.sg, xlli@a-star.edu.sg, ramasamysa@a-star.edu.sg

Abstract

Generalized Category Discovery (GCD) aims to classify labeled instances from known categories while discovering novel categories from unlabeled data. Despite recent progress in GCD for computer vision, existing GCD approaches largely rely on static final-step representations (in the visual domain), overlooking the temporally evolving nature of time-series data. In this paper, we introduce TGCD, the first framework specifically designed for GCD in time-series data. TGCD leverages both the dynamics of latent representations and the heterogeneity of predictions across multiple temporal segments to disover unknown (i.e., novel) categories, based on a pre-trained time-series foundation model. We propose a unified learning objective for TGCD that integrates the following three components: (i) a Stochastic Temporal Segment Dropout (STeSD) objective that regularizes the model by selectively penalizing high-entropy segments to encourage confident predictions on uncertain regions of the time-series, and (ii) a Known–Unknown Temporal Discriminability (KUTD) objective that promotes representational separation between known and unknown categories within unlabeled data and (iii) a margin-aware classification objective to improve generalization. Empirical evaluation on six multivariate time-series datasets demonstrates that the TGCD substantially outperforms existing GCD methods, particularly in discovering unknown categories. We further conduct ablation studies to highlight the individual contributions of each component. Additionally, we provide the first comprehensive benchmarking of recent GCD approaches on time-series data, revealing the limitations of naive transfer and underscoring the benefits of temporal modeling.

Introduction

Conventional supervised learning paradigms assume that the complete label space is known and fail to detect, discover, and generalize effectively when faced with novel patterns or previously unseen category instances. This challenge underscores the importance of Generalized Category Discovery (GCD) (Vaze et al. 2022a), also known as semi-supervised open-world learning (Cao, Brbic, and Leskovec 2022), that not only recognizes and classifies known categories, but also identifies unknown categories (i.e., novel categories). There is a growing recent research interest in developing GCD for computer vision (Vaze et al. 2022a; Cao, Brbic, and Leskovec 2022; Sun and Li 2023; Ye et al. 2024; Ma et al. 2025). On the other hand, time-series data is ubiquitous in several applications, including sensor outputs in industrial IoT applications, analyzing physiological signals in healthcare, and tracking activities in smart environments. In these application domains, new categories keep emerging with continued utility. For example, the number of failure modes in IoT applications or the number of disease phenotypes in healthcare applications are perpetually expanding. However, GCD methods developed for image are not directly applicable to time-series data as it offers unique challenges due to the lack of clear semantics or visually indistinguishable patterns that restricts its generalizability even within similar application domains. Moreover, owing to the need for multiple sensors or signals to monitor or analyze a system, the data is often multi-variate due to multiple channels that are sampled at distinct frequencies. Therefore, there is a need to develop a GCD method for time-series data.

Developing GCD methods for time-series is challenging due to the temporal nature of data, which is generated as long sequences resulting from the system's continuous operation. Owing to this long sequence, there is inherent heterogeneity in predictions across multiple segments of the time-series, and significant variability in marginal distributions. Furthermore, as the signal patterns between multiple categories of time-series data are visually indistinguishable, the challenge of distinguishing between known and unknown categories of data becomes more prominent. This paper addresses these challenges by developing a GCD framework for time-series data, termed TGCD, which incorporates three key objectives: (i) Stochastic Temporal Segment Dropout (STeSD), which aims to circumvent the heterogeneity in predictions across multiple segments of the time-series through a stochastic dropout mechanism that directs the model to learn from segments of the time-series with high heterogeneity while stochastically discarding segments with less heterogeneity. The stochastic dropout is guided by the entropy of the individual segments in the time-series; (ii) Known–Unknown Temporal Discriminability (KUTD), which helps to enhance the indiscriminable patterns between known and unknown categories based on the temporal trajectory of intermediate feature embeddings. By promot-

*Corresponding Author

ing smoother temporal representations for labeled samples and allowing greater variation for unlabeled ones, KUTD guides the model towards better separation between known and novel (or unknown) categories; and (iii) a margin-aware classification objective, which calibrates category separation margins using softmax confidence from unlabeled data, enabling balanced learning across all categories.

To the best of our knowledge, we present TGCD (Time-series Generalized Category Discovery), the first framework specifically designed for GCD in time-series data, incorporating the three key objectives outlined above. We argue that temporal modeling provides a unique opportunity to improve unknown category discovery by leveraging two complementary characteristics: (1) Temporal evolution of predictive uncertainty, and (2) Temporal discriminability of learned representations. In particular, TGCD is based on the *hypothesis that samples from known categories exhibit more consistent predictions and stable latent trajectories over time, whereas samples from unknown categories are characterized by higher uncertainty and more volatile representational dynamics.* By integrating these temporal insights, TGCD effectively separates known and unknown categories, addressing challenges overlooked by prior GCD methods that rely mainly on static, final-step features.

We evaluate the TGCD extensively on six datasets across multiple domains (facilities, human body, nature etc), with varying sequence lengths and wide range of total number of categories (15 to 60). It is observed from the results that TGCD consistently outperforms the state-of-the-art GCD methods in classifying the known categories, and discovering unknown categories. Specifically, the TGCD outperforms the state-of-the-art GCD methods by an average margin of $\approx 3\%$. We also conduct several ablation studies to study the performance of the proposed TGCD under the influence of varying unknown data ratio, varying labeled data ratio, distinct choice of time-series foundation models, and also analyze the temporal stability. The results from the ablation studies support our choice of strategies for TGCD. More importantly, our framework is backbone-agnostic and compatible with any temporal encoder, including recent foundation models such as MOMENT (Goswami et al. 2024).

In summary, this paper makes the following contributions:

1. First formulation of GCD for time-series: While prior GCD approaches have primarily targeted the computer vision domain, our work is the first to address GCD in multivariate time-series data.
2. Development of the TGCD framework: We propose a novel framework with multiple objectives tailored for GCD in time-series data.
3. Stochastic Temporal Dropout (STeSD): We introduce a stochastic dropout mechanism that encourages the model to learn from segments exhibiting higher heterogeneity by discarding those with lower heterogeneity.
4. Temporal Discriminability Modeling: We propose the Known–Unknown Temporal Discriminability (KUTD) loss, which encourages the separation of known and novel categories by modeling their temporal dynamics in the latent space.
5. Benchmarking: We conduct the first benchmark evaluation of GCD methods on time-series datasets.

Related Work

We present the related works in generalized category discovery and temporal uncertainty.

Generalized Category Discovery

Generalized Category Discovery (GCD) aims to simultaneously classify labeled samples from known categories and discover unknown categories from unlabeled data. This problem is especially relevant where labeled data is only available for a subset of the categories. While GCD has been extensively explored in the computer vision domain, its adaptation to time-series domain remains underexplored. Early methods such as DTC (Han, Vedaldi, and Zisserman 2019), UNO (Fini et al. 2021), GCD(Vaze et al. 2022a), RankStats (Han et al. 2021), and OpenMix (Zhong et al. 2021) laid the groundwork for generalized category discovery by integrating deep embedding clustering with supervised classification to separate known and unknown classes. Subsequent methods like ORCA (Cao, Brbic, and Leskovec 2022) and (Rizve et al. 2022) have introduced a pairwise similarity loss that implicitly clusters unlabeled data into known and unknown categories. More recent advances have improved discovery through contrastive learning objectives (Sun and Li 2023), learning pace synchronization (Ye et al. 2024), and prototype-based learning (Zheng et al. 2024; Ma et al. 2025) strategies. Additional advances have been proposed by ConceptGCD (Weng, Xiao, and Jiang 2023), FlipClass (Lin et al. 2024), Active-GCD (Ma et al. 2024), and SPTNet (Wang, Vaze, and Han 2024) through concept-aware modeling, student-teacher-learning, active querying, and attention alignment, respectively. Despite this progress, most GCD methods operate on static final-step representations, ignoring temporal cues that are prominent in sequential time-series data-a gap that motivates the present work.

Temporal Uncertainty

Measuring uncertainty progression in time-series has been explored in forecasting and safety-critical domains (Malinin and Gales 2018; Ovadia et al. 2019; Malinin and Gales 2021; Patharkar et al. 2024). Furthermore, recent works have evaluated temporal uncertainty calibration and distribution shifts in time-series data under a closed-world setting (Fan et al. 2023; Clevert et al. 2024). However, the role of temporal uncertainty in the context of Generalized Category Discovery (i.e., open-world semi-supervised learning) has been largely overlooked. To the best of our knowledge, our framework is the first to leverage both uncertainty in distinguishing known and unknown categories, as well as uncertainty arising from heterogeneity across multiple temporal segments, for Generalized Category Discovery in time-series data.

Method

In this section, we formalize the problem setup and introduce our proposed method for Generalized Category Dis-

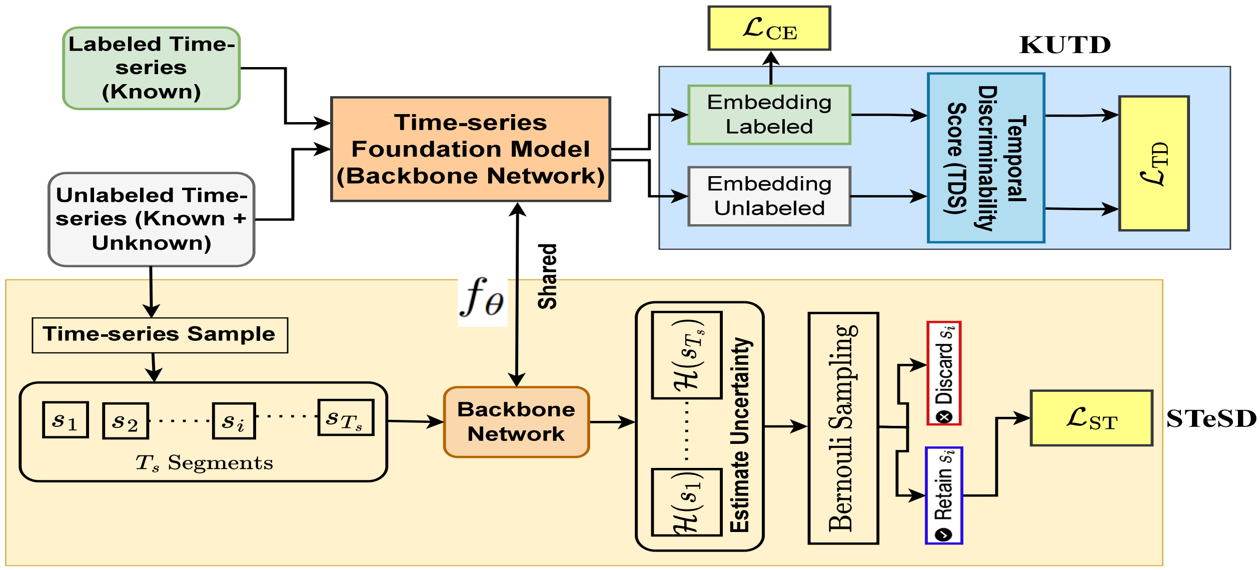

Figure 1: Overview of the proposed TGCD framework. A shared time-series foundation model extracts embeddings from labeled and unlabeled data. Segment-level entropy guides Stochastic Temporal Segment Dropout (STeSD) to focus learning on uncertain regions. The Known–Unknown Temporal Discriminability (KUTD) module promotes separation of known and unknown categories using temporal dynamics. A margin-aware classification head ($\mathcal{L}_{\mathrm{CE}}$) computes the final predictions. The framework jointly optimizes all components for effective generalized category discovery in time-series data.

covery (GCD) in time-series data, where only a subset of categories have labeled samples (i.e., known categories), while the unlabeled data contains instances from both known and unknown categories.

Problem Setup

We consider the problem of Generalized Category Discovery (GCD) for multi-variate time series data. The training dataset comprises labeled and unlabeled samples:

$$\mathcal{D}_L^{tr} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N_L}, \quad \mathcal{D}_U^{tr} = \{x_j\}_{j=1}^{N_U},$$

where each $\mathbf{x} \in \mathbb{R}^{C\times T}$ is a multivariate time series with C channels and T time steps. The labeled dataset $\mathcal{D}_L$ contains samples from the known set of categories $\mathcal{K}_{\mathrm{k}}$, while the unlabeled set $\mathcal{D}_U$ contains samples from both $\mathcal{K}_{\mathrm{k}}$ and a disjoint set of novel (unknown) categories $\mathcal{K}_{\mathrm{uk}}$, i.e.,

$$\mathcal{K}_{\mathrm{k}} \cap \mathcal{K}_{\mathrm{uk}} = \emptyset.$$

We define a predictive model $f_\theta(x) \in \Delta^{\mathcal{K}}$ mapping to a $\mathcal{K}$-dimensional categorical distribution, where $\mathcal{K} = |\mathcal{K}_{\mathrm{k}}| + |\mathcal{K}_{\mathrm{uk}}|$. The objective of this predictive model is twofold: (1) to perform accurate classification on samples from $\mathcal{K}_{\mathrm{k}}$ in the unlabeled data, and (2) to discover and separate unknown categories ($\mathcal{K}_{\mathrm{uk}}$) within the unlabeled data.

Overall Framework of TGCD

The temporal progression and predictive uncertainty in multivariate time-series data due to heterogeneity in predictions across multiple segments presents unique challenges that remain underexplored in GCD. Motivated by this, we propose the **Time-series Generalized Category Discover (TGCD)** framework by leveraging uncertainty-aware and temporally-evolving representations for GCD in time-series data. The proposed TGCD framework (as shown in Figure 1) leverages a pretrained *time-series foundation model*, typically a Transformer-based architecture trained with self-supervised objectives (e.g., masked time series modeling), as the representational backbone for extracting features ($f_\theta^{emb} : \mathbb{R}^{C\times T} \rightarrow \mathbb{R}^d$; *emb*: embedding). A pre-trained time-series foundation model is chosen because they are good representations of complex temporal dependencies and generate transferable representations for both labeled and unlabeled samples of the time-series data.

In addition to the pre-trained foundation model, a classification head and a softmax are used to map the embeddings to the class labels, parametrized by $W : \mathbb{R}^d \rightarrow \mathbb{R}^{\mathcal{K}}$. The classification head consists of a single linear (fully connected) layer that comprises of a first set of known categories followed by anticipated number of unknown categories $|\mathcal{K}| = |\mathcal{K}_{\mathrm{k}}| + |\mathcal{K}_{\mathrm{uk}}|$. The prediction for a given instance $\mathbf{x}_i$ is obtained by computing $\arg\max(W^\top \cdot f_\theta^{emb}(\mathbf{x_i}))$, where W is the classification weight matrix. As the number of unknown categories is not known *a priori*, we follow the similar approach as (Cao, Brbic, and Leskovec 2022), where some heads that corresponds to an anticipated unknown category remain redundant until the model is presented with samples corresponding to that unknown category.

TGCD achieves GCD through three key learning objectives: (i) a Stochastic Temporal Segment Dropout (STeSD) objective ($\mathcal{L}_{\mathrm{ST}}$), which regularizes the model by encouraging confident predictions on the most heterogeneous temporal segments retained through a Bernoulli-based uncertainty-aware selection process; (ii) a Known–Unknown Temporal Discriminability (KUTD) objective ($\mathcal{L}_{\mathrm{TD}}$), which encourages representational divergence for unlabeled samples; and (iii) a margin-aware classification objective ($\mathcal{L}_{\mathrm{CE}}$), which improves the discriminability of known and unknown categories by incorporating prediction confidence. To the best of our knowledge, this is the first formulation that jointly models temporal heterogeneity and representational dynamics to facilitate the discovery of unknown categories in time-series data. Thus, the unified learning objective of TGCD is represented by:

$$\mathcal{L} = \mathcal{L}_{\mathrm{ST}} + \mathcal{L}_{\mathrm{TD}} + \mathcal{L}_{\mathrm{CE}}.$$

Next, we describe each objective in detail.

Stochastic Temporal Segment Dropout

There is an inherent predictive uncertainty in long time-series sequences due to heterogeneity in predictions across multiple segments of the multi-variate time-series. Hence, we leverage the predictive uncertainty for the individual segments to stochastically decide which temporal segments of a sequence to retain or drop during training. To this end, we develop a novel regularization technique called *Stochastic Temporal Segment Dropout (STeSD)* to enhance the generalization of TGCD under partial supervision and inter-segment heterogeneity.

Given an input time-series sequence, we divide it into T_s non-overlapping temporal segments. For each segment s_i, we compute the predictive class distribution $\hat{p}_i$ from the model's output logits and estimate uncertainty using entropy ($\mathcal{H}(s_i)$):

$$\mathcal{H}(s_i) = -\sum_{\kappa=1}^{\mathcal{K}} \hat{p}_i^{(\kappa)} \log \hat{p}_i^{(\kappa)}, \tag{1}$$

where $\mathcal{K}$ is total number of categories. These entropy values are normalized across all segments of the sequence:

$$\tilde{\mathcal{H}}(s_i) = \frac{\mathcal{H}(s_i)}{\sum_{j=1}^{T_s} \mathcal{H}(s_j)}. \tag{2}$$

We then define the probability of retaining a segment in the sequence via an exponentially decaying function:

$$\mathcal{P}_{\text{retain}}(s_i) = 1 - \exp(-\beta \cdot \tilde{\mathcal{H}}(s_i)), \tag{3}$$

where β is a temperature hyperparameter controlling the sensitivity of the dropout to segment-level uncertainty. During training, each segment is retained according to a Bernoulli trial:

$$\text{retain}(s_i) \sim \text{Bernoulli}(\mathcal{P}_{\text{retain}}(s_i)). \tag{4}$$

This stochastic dropout mechanism selectively omits segments with lower uncertainty while retaining segments with higher uncertainty, thus directing learning toward challenging temporal patterns. The retained segments are used to train the model through an uncertainty-based regularization loss as follows:

$$\mathcal{L}_{\text{ST}} = \frac{1}{|\mathcal{S}_{\text{retain}}|} \sum_{s_i \in \mathcal{S}_{\text{retain}}} \mathcal{H}(s_i), \tag{5}$$

where $\mathcal{S}_{\text{retain}}$ denotes the set of retained segments in a sequence. This loss encourages confident predictions over the retained uncertain segments, thereby enhancing generalization for the TGCD.

STeSD is inspired by uncertainty-guided learning (Kendall and Gal 2017), which emphasizes the utility of predictive uncertainty in model optimization, and curriculum-based dropout strategies (Morerio et al. 2017), which adaptively modulates difficulty during training. However, unlike these prior works, STeSD introduces a stochastic, segment-level dropout mechanism tailored for time-series data, making it well-suited for applications involving partial labels, distributional shift, and novel category emergence.

Known–Unknown Temporal Discriminability

The *Known–Unknown Temporal Discriminability (KUTD)* objective is designed to capture the representational dynamics of time-series data and facilitate the effective separation of known and unknown categories. It leverages intermediate feature embeddings extracted from the model f_θ^{emb} for both labeled and unlabeled data to distinguish between known and unknown categories within the unlabeled set. Specifically, KUTD minimizes temporal discriminability among samples from known categories, while maximizing discriminability between samples from known and unknown categories. This encourages temporal stability in the representations of known-category samples, while promoting representational divergence in potentially unknown-category samples.

Let $\mathbf{x} = (x_1, x_2, \ldots, x_T) \in \mathbb{R}^{C\times T}$ be a multivariate time series, and let $\phi_t(x) \in \mathbb{R}^d$ denote the latent representation (embedding) produced by an intermediate layer of the model f_θ when given partial input $\mathbf{x}_{1:t}$. That is,

$$\phi_t(x) = f_\theta^{\text{emb}}(x_{1:t}), \tag{6}$$

where $f_\theta^{\text{emb}}(x_{1:t})$ refers to the d-dimensional latent embedding (e.g., obtained using average-pooled features from MOMENT backbone (Goswami et al. 2024)).

We define the *Temporal Discriminability Score* (TDS) of a sample x as the average pairwise Euclidean distance between embeddings at selected normalized time checkpoints $\mathcal{T} = \{\tau_1, \tau_2, \ldots, \tau_\nu\} \subset (0, 1]$:

$$\text{TDS}(x) = \frac{2}{\nu(\nu-1)} \sum_{1\le i<j\le \nu} \left\| \phi_{\tau_i T}(x) - \phi_{\tau_j T}(x) \right\|_2, \tag{7}$$

where the factor $\frac{2}{\nu(\nu-1)}$ denotes the number of unique pairs among k embeddings, which ensures that the TDS value is scale-independent with respect to how many time steps you evaluate. To leverage this score for known-versus-unknown partitioning, we divide each mini-batch into labeled ($\mathcal{D}_L$) and unlabeled ($\mathcal{D}_U$) sets. We then compute:

$$\mathcal{L}_{\text{TD}} = \mathbb{E}_{x\in\mathcal{D}_L}\left[\text{TDS}(x)\right] - \lambda_{\text{tds}} \cdot \mathbb{E}_{x\in\mathcal{D}_U}\left[\text{TDS}(x)\right], \tag{8}$$

where $\lambda_{\text{tds}} > 0$ controls the strength of the drift promotion for unlabeled samples $\mathcal{D}_U$. Although the unlabeled set $\mathcal{D}_U$ contains both known and unknown samples, their identities remain hidden during training. We assume that

Dataset	# Timeseries	$\lvert\mathcal{D}_L^{tr}\rvert$	$\lvert\mathcal{D}_U^{tr}\rvert$	$\lvert\mathcal{D}^{ts}\rvert$	$\lvert\mathcal{Y}_{\mathrm{k}}\rvert$	$\lvert\mathcal{Y}_{\mathrm{uk}}\rvert$	$\mathcal{K}$	C	T	Type	Domain
CharacterTrajectories	2858	160	1983	715	15	5	20	3	119	Motion	Facilities
DSA	9120	504	6336	2280	14	5	19	45	125	Motion	Human Body
Fungi	204	15	138	51	14	4	18	1	201	Other	Nature
GestureMidAirD1	338	20	233	85	20	6	26	1	360	HAR	Human Body
ShapesAll	1200	90	810	300	45	15	60	1	512	Image	Generated
SwedishLeaf	1125	66	777	282	11	4	15	1	128	Image	Nature

Table 1: Statistics of datasets. Each dataset is partitioned into three disjoint subsets: $\mathcal{D}_L^{tr}$ (labeled training data), $\mathcal{D}_U^{tr}$ (unlabeled training data containing both known and unknown categories), and $\mathcal{D}^{ts}$ (test data). The label space is partitioned into known categories $\mathcal{Y}_{\mathrm{k}}$ and unknown categories $\mathcal{Y}_{\mathrm{uk}}$

.

unknown samples exhibit higher TDS values due to their misalignment with the temporal dynamics of known categories. Thus in Eq. (8), minimizing $\mathbb{E}_{x\in\mathcal{D}L}[\mathrm{TDS}(x)]$ enforces temporal smoothness for known categories, while maximizing $\mathbb{E}x \in \mathcal{D}_U[\mathrm{TDS}(x)]$ preserves temporal variability among unlabeled samples. This contrast facilitates separation of known and novel categories by promoting stability in known-category representations and allowing divergence in potentially unknown ones. A theoretical justification of this loss behavior is provided in the supplementary material (Theorem 1).

Margin-aware Classification Objective

GCD problem faces the challenge of preventing overfitting to known categories while discovering novel (i.e., unknown) ones. To address this, we leverage confidence-aware decision boundary adaptation strategies commonly used in GCD methods (Scheirer et al. 2013; Neal et al. 2018; Vaze et al. 2022b) to adjust classification margins adaptively. By estimating the overall prediction confidence from unlabeled samples, we compute a margin that is applied consistently across all classes (i.e., categories) during training. This margin encourages correct classification on labeled data while widening decision boundaries in a confidence-aware manner, thereby reducing bias towards known categories and aiding discovery of unknown ones. Specifically, we define the margin δ as: $\delta = \gamma \cdot \left(1 - \frac{1}{K}\sum_{j=1}^{K}\hat{p}_j\right)$ where $\hat{p}_j$ denotes the average softmax confidence for class j computed over the unlabeled dataset $\mathcal{D}_U$, and $\gamma > 0$ is a margin scaling hyperparameter controlling the magnitude of the margin. The margin-regularized cross-entropy loss is formulated as:

$$\mathcal{L}_{CE} = -\log\left(\frac{\exp(z_y+\delta)}{\exp(z_y+\delta)+\sum_{k\neq y}\exp(z_k)}\right), \quad (9)$$

where, z_y is the logit for the ground-truth class y, z_k is the outputs of the linear classification layer before the softmax function for class k. This margin-aware formulation reduces the margin for categories with high confidence and increases it for less certain ones, promoting balanced learning for both known and unknown categories in GCD.

Overall, our complete training objective integrates uncertainty-aware regularization through entropy dynamics, temporal discriminability-based separation and supervised learning via margin-aware classification of known and unknown categories. Together, these components leverage both labeled supervision and unlabeled temporal structure for GCD in multivariate time-series data.

Experiments

Experimental Setup

Datasets. We evaluate our method on six publicly available time-series classification datasets (Dau et al. 2018; Altun, Barshan, and Tunçel 2010; Dau et al. 2019) spanning diverse temporal characteristics, data size and application domains. Table 1 summarizes key statistics and configurations for each dataset under the GCD setting.

Baselines. We evaluate the TGCD in comparison with several state-of-the-art Generalized Category Discovery (GCD) approaches originally developed for the computer vision domain: ORCA (Cao, Brbic, and Leskovec 2022), NACH (Guo et al. 2022), GCD (Vaze et al. 2022a), ProtoGCD (Ma et al. 2025), OpenCon (Sun and Li 2023), and LPS (Ye et al. 2024). *Since none of these methods are designed for time-series data, we adapt and extend their publicly available implementations to construct strong baselines for our setting.* To ensure a fair comparison, we replace the backbone network of these existing state-of-the-art methods with the same MOMENT (Goswami et al. 2024) model as ours.

Evaluation Metrics. To evaluate the performance of both the proposed and existing methods, we calculate the accuracy between the ground-truth labels and the model's cluster assignments on the test set using the Hungarian algorithm (Kuhn 1955) as follows: (1) η_{all}: accuracy for all instances, i.e., with known and unknown categories. (2) η_k: accuracy for instances in the known categories. (3) η_{uk}: accuracy for instances in the unknown categories.

Performance Study

We evaluate the performance of our proposed method, TGCD, across six diverse time-series datasets under the GCD setting and presented results in Table 2. Our method consistently outperforms state-of-the-art baselines across almost all datasets and metrics. Notably, TGCD achieves the highest overall accuracy (η_{all}) on all the six datasets, with significant margins over the closest baselines. For instance, on CharacterTrajectories, TGCD attains an η_{all} of 73.01%,

Methods	CharacterTrajectories			DSA			Fungi			GestureMidAirD1			ShapesAll			SwedishLeaf		
	η_{all}	η_k	η_{uk}	η_{all}	η_k	η_{uk}	η_{all}	η_k	η_{uk}	η_{all}	η_k	η_{uk}	η_{all}	η_k	η_{uk}	η_{all}	η_k	η_{uk}
ORCA	69.23	90.70	64.36	65.92	88.21	56.17	76.47	92.11	61.54	51.76	61.54	55	51	67.11	37.33	57.09	75.24	53.95
NACH	68.45	91.25	65.20	64.38	87.20	55.10	75.12	91.75	58.10	53.10	60	56.25	53.42	67.90	39.25	59.12	76.83	54.9
GCD	66.97	80.55	51.21	30.45	31.48	29.13	75	84.72	90	48.22	59.13	58.46	50.83	56.96	56.07	58.16	75.73	78.95
ProtoGCD	28.39	35.10	50.33	44.82	60.42	45	33.33	34.21	61.54	30.59	30.77	50	3.67	4.89	12	42.20	51.46	77.63
OpenCon	54.27	66.22	76.70	54.67	65.28	57.63	60.78	65.79	61.54	56.47	67.69	65	45.67	53.78	53.33	53.9	58.74	68.42
LPS	67.69	89.75	74.47	59.25	78.81	51.5	74.51	**94.74**	61.54	45.88	46.15	60	45.33	58.67	41.33	62.06	82.04	85.83
TGCD (Ours)	**73.01**	**95.26**	**80.85**	**67.24**	**91.01**	**59.87**	**80.39**	92.11	**92.31**	**60**	**68.32**	**66.87**	**56**	**72.89**	**56.67**	**64.54**	**83.98**	**89.47**

Table 2: Comparison of the proposed TGCD with SOTA methods. The top-performing result is shown in bold red color.

Setting	η_{all}	η_k	η_{uk}
Full (All losses)	**73.01**	**95.26**	**80.85**
No $\mathcal{L}_{\text{TD}}$	71.28	95.01	76.32
No $\mathcal{L}_{\text{ST}}$	70.12	94.89	74.70
$\mathcal{L}_{\text{CE}}$ only	66.85	94.45	68.25

Table 3: Ablation study on the CharacterTrajectories dataset, analyzing the contribution of each temporal component.

outperforming the second-best (ORCA) by 3.78%. While TGCD yields consistently strong results across all metrics, we acknowledge that in one case, LPS attains slightly higher known category accuracy (η_k) on *Fungi*, TGCD compensates with a much stronger η_{uk}, resulting in the best η_{all}. overall. This trend highlights TGCD's ability to maintain a strong trade-off between accurately classifying known categories and discovering unknown ones. These improvements are attributed to the proposed **STeSD** objective, which adaptively drops the segments based on entropy to regularize the model, and the **KUTD** objective that enhances temporal discrimination between known and unknown categories. Together, these components enable **TGCD** to discover unknown categories more reliably without sacrificing performance on known categories.

Ablation Study

We conduct experiments on the CharacterTrajectories dataset to evaluate the effectiveness of each individual strategy and the impact of hyperparameter variations in TGCD.

Effect of the individual Objective functions: As shown in Table 3, we systematically remove each loss term from our full model to study the impact of each loss term. It is pertinent to note that all three objectives ($\mathcal{L} = \mathcal{L}_{\text{CE}} + \mathcal{L}_{\text{ST}} + \mathcal{L}_{\text{TD}}$) are critical for the highest performance, confirming the synergistic effect of them. Using only the classification objective adversely affects the overall accuracy and accuracy of unknown categories. This emphasizes the need for $\mathcal{L}_{\text{ST}}$ and $\mathcal{L}_{\text{TD}}$. Furthermore, while ignoring the KUTD objective leads to drop in accuracy of both overall and unknown category, ignoring the Stochastic Temporal Segment Dropout (STeSD) objective has strong effects on the accuracies of unknown categories. These observations validate our hypothesis that enforcing temporal uncertainty strengthens the model's generalization under GCD setting for time-series data. Overall, both temporal losses ($\mathcal{L}_{\text{ST}}$ and $\mathcal{L}_{\text{TD}}$) contributes significantly to the GCD performance of the proposed TGCD framework.

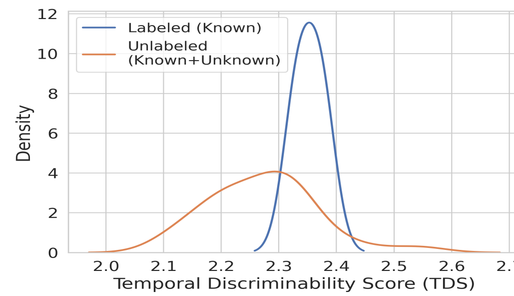

Figure 2: Analysis of Temporal Discriminability Score on CharacterTrajectories Dataset

Analysis of Temporal Discriminability Score. We study the effect of the Temporal Discriminability Score (TDS) distributions for labeled (known) and unlabeled (known + unknown) samples on the CharacterTrajectories dataset using the final trained model. Fig. 2 demonstrates that the known categories exhibit more consistent temporal representations compared to unknown ones, which supports our hypothesis. Specifically, the labeled data exhibits a sharp, narrow distribution centered around a lower TDS value, indicating high temporal stability and consistency in their intermediate feature embeddings over time. On the other hand, the unlabeled data has a wider, more spread-out TDS distribution, indicating more temporal variation. The difference in the densities between these two groups is consistent with our theoretical assumption that unknown categories, which the model hasn't seen before, have more unstable temporal patterns. These results confirm that our KUTD loss effectively uses this difference to separate unknown categories from known ones based on their temporal behavior. To quantify this separation between known and unknown temporal dynamics, we compute the symmetric KL divergence between their TDS distributions as 3.9733. The clear margin between the two distributions validates the discriminative power of temporal signals in our proposed framework TGCD.

Effect of Known Category Ratio (KCR). Next, we demonstrate the effect of the Known Category Ratio (KCR) (the proportion of known (labeled) categories) affects Time-

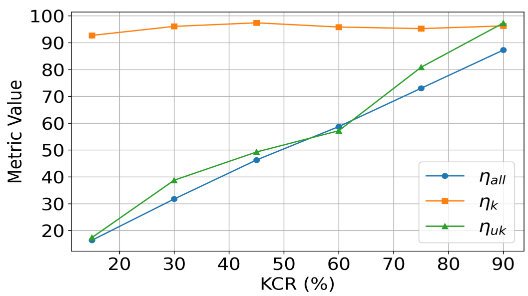

Figure 3: Analysis of KCR on CharacterTrajectories Dataset

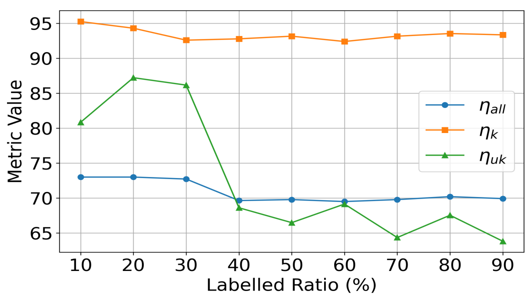

Figure 4: Effect of labeled sample ratio on CharacterTrajectories Dataset

Series GCD performance. As shown in Fig. 3, increasing KCR leads to consistent improvements in both overall and unknown category accuracy. The proposed method maintains high accuracy on known categories across all KCR values. However, performance on unknown categories is notably lower when the number of known categories is small (i.e., low KCR). This highlights the inherent difficulty of generalized category discovery under limited supervision. Nonetheless, the steady improvement in unknown accuracy with more labeled categories suggests that our method scales well with supervision and leverages labeled information effectively when available.

Effect of labeled sample ratio. We investigate the impact of the labeled sample ratio (i.e., the percentage of labeled data used for training among all known category samples) on GCD performance for time-series data. Fig. 4 summarizes the results. Interestingly, performance is strongest when only 10–30% of known samples are labeled, indicating that the model effectively leverages limited supervision. As the labeled ratio increases beyond this point, overall and unknown accuracy gradually decline. This trend suggests that a larger labeled set may overfit the model to known categories, reducing its ability to generalize to unknown ones. These results underscore the need for a careful balance between labeled supervision and unknown-category discovery in GCD.

Analysis of various backbones for TGCD. Finally, we study the effectiveness of the proposed **TGCD** framework under different temporal representations by evaluating the framework with various backbone architectures commonly

Backbone	η_{all}	η_k	η_{uk}
TST	65.17	83.87	62.77
TS-TCC	72.17	92.41	66.49
TS2Vec	61.40	79.89	58.51
TimesNet	68.67	92.22	44.68
MOMENT (NO Fine-Tuning)	11.47	14.80	38.30
MOMENT	**73.01**	**95.26**	**80.85**

Table 4: Impact of different welll-known backbones on the Proposed Framework TGCD

used in time-series representation learning.

As shown in Table 4, our method consistently benefits from stronger temporal representations, with **MOMENT** achieving the highest overall accuracy ($\eta_{all} = 73.01$), known category accuracy ($\eta_k = 95.26$), and unknown category accuracy ($\eta_{uk} = 80.85$). In contrast, pre-trained MOMENT without fine-tuning performs poorly, indicating the necessity of task-specific adaptation. Other backbones like TS-TCC and TimesNet show strong performance on seen categories but comparatively lower generalization to unknowns. This highlights that while certain backbones encode discriminative features for known categories, they may lack the flexibility to generalize under GCD setting. Overall, the superior performance of MOMENT within TGCD confirms its suitability for time-series GCD task.

We also provide additional ablation studies in the supplementary, including the impact of the number of segments in STeSD and the sensitivity to other hyper-parameters.

Conclusion

This paper presents TGCD (Time-series Generalized Category Discovery), a novel framework addressing the underexplored problem of generalized category discovery in time-series data. To our knowledge, this is the first work to formulate GCD specifically for time-series and to propose an associated framework and benchmark. Our framework targets the prominent challenge of distinguishing between known and unknown categories due to long sequences of time-series data presented with visually indistinguishable patterns. To this end, we introduce TGCD, which achieves three key objectives: (i) Stochastic Temporal Segment Dropout (STeSD), designed to circumvent heterogeneity in predictions across multiple segments of the time-series; (ii) Known–Unknown Temporal Discriminability (KUTD), which enhances discriminability between known and unknown categories; and (iii) a margin-aware classification objective to enable unbiased learning across all categories. Performance studies on a diverse set of publicly available datasets demonstrate the effectiveness of TGCD compared to state-of-the-art methods. The proposed framework consistently outperforms existing GCD methods, achieving gains in overall accuracy ranging from 1.32% to 3.92%. Furthermore, ablation studies highlight the importance and effectiveness of each individual strategy.

Acknowledgments

This work was supported by the Agency for Science, Technology and Research (A*STAR) through its DesCartes programme and is supported by the National Research Foundation, Prime Minister's Office, Singapore under its Campus for Research Excellence and Technological Enterprise (CREATE) programme.

References

Altun, K.; Barshan, B.; and Tunçel, O. 2010. Comparative study on classifying human activities with miniature inertial and magnetic sensors. *Pattern Recognit.*, 43: 3605–3620.

Cao, K.; Brbic, M.; and Leskovec, J. 2022. Open-World Semi-Supervised Learning. In *International Conference on Learning Representations (ICLR)*.

Clevert, D.; Svensson, E.; Svensson, J.; et al. 2024. Temporal Evaluation of Uncertainty Quantification Under Distribution Shift. In *In Odour-Based Deep Drug Discovery Workshop, AIDD*. Springer.

Dau, H. A.; Bagnall, A.; Kamgar, K.; Yeh, C.-C. M.; Zhu, Y.; Gharghabi, S.; Ratanamahatana, C. A.; and Keogh, E. 2019. The UCR time series archive. *IEEE/CAA Journal of Automatica Sinica*, 6(6): 1293–1305.

Dau, H. A.; Keogh, E.; Kamgar, K.; Yeh, C.-C. M.; Zhu, Y.; Gharghabi, S.; Ratanamahatana, C. A.; Yanping; Hu, B.; Begum, N.; Bagnall, A.; Mueen, A.; Batista, G.; and Hexagon-ML. 2018. The UCR Time Series Classification Archive. https://www.cs.ucr.edu/~eamonn/time_series_data_2018/.

Fan, W.; Wang, P.; Wang, D.; Wang, D.; Zhou, Y.; and Fu, Y. 2023. Dish-ts: a general paradigm for alleviating distribution shift in time series forecasting. In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, 7522–7529.

Fini, E.; Sangineto, E.; Lathuilière, S.; Zhong, Z.; Nabi, M.; and Ricci, E. 2021. A unified objective for novel class discovery. In *Proceedings of the IEEE/CVF international conference on computer vision*, 9284–9292.

Goswami, M.; Szafer, K.; Choudhry, A.; Cai, Y.; Li, S.; and Dubrawski, A. 2024. MOMENT: A Family of Open Time-series Foundation Models. In *International Conference on Machine Learning*, 16115–16152. PMLR.

Guo, L.-Z.; Zhang, Y.-G.; Wu, Z.-F.; Shao, J.-J.; and Li, Y.-F. 2022. Robust semi-supervised learning when not all classes have labels. *Advances in neural information processing systems*, 35: 3305–3317.

Han, K.; Rebuffi, S.-A.; Ehrhardt, S.; Vedaldi, A.; and Zisserman, A. 2021. Autonovel: Automatically discovering and learning novel visual categories. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(10): 6767–6781.

Han, K.; Vedaldi, A.; and Zisserman, A. 2019. Learning to discover novel visual categories via deep transfer clustering. In *Proceedings of the IEEE/CVF international conference on computer vision*, 8401–8409.

Kendall, A.; and Gal, Y. 2017. What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision? In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 30.

Kuhn, H. W. 1955. The Hungarian method for the assignment problem. *Naval research logistics quarterly*, 2: 83–97.

Lin, H.; An, W.; Wang, J.; Chen, Y.; Tian, F.; Wang, M.; Wang, Q.; Dai, G.; and Wang, J. 2024. Flipped classroom: Aligning teacher attention with student in generalized category discovery. *Advances in Neural Information Processing Systems*, 37: 60897–60935.

Ma, S.; Zhu, F.; Zhang, X.-Y.; and Liu, C.-L. 2025. Protogcd: Unified and unbiased prototype learning for generalized category discovery. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Ma, S.; Zhu, F.; Zhong, Z.; Zhang, X.-Y.; and Liu, C.-L. 2024. Active generalized category discovery. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 16890–16900.

Malinin, A.; and Gales, M. 2018. Predictive uncertainty estimation via prior networks. *Advances in neural information processing systems*, 31.

Malinin, A.; and Gales, M. 2021. Uncertainty Estimation in Autoregressive Structured Prediction. In *International Conference on Learning Representations*.

Morerio, P.; Cavazza, J.; Volpi, R.; Vidal, R.; and Murino, V. 2017. Curriculum Dropout. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*.

Neal, L.; Olson, M.; Fern, A.; and Li, W.-K. 2018. Open set learning with counterfactual images. In *ECCV*, 613–628.

Ovadia, Y.; Fertig, E.; Ren, J.; Nado, Z.; Sculley, D.; Nowozin, S.; Dillon, J. V.; Lakshminarayanan, B.; and Snoek, J. 2019. Can You Trust Your Model's Uncertainty? Evaluating Predictive Uncertainty Under Dataset Shift. In *Advances in Neural Information Processing Systems*, 13969–13980. NeurIPS.

Patharkar, A.; Huang, J.; Wu, T.; Forzani, E.; Thomas, L.; Lind, M.; and Gades, N. 2024. Eigen-entropy based time series signatures to support multivariate time series classification. *Scientific reports*, 14(1): 16076.

Rizve, M. N.; Kardan, N.; Khan, S.; Shahbaz Khan, F.; and Shah, M. 2022. Openldn: Learning to discover novel classes for open-world semi-supervised learning. In *European Conference on Computer Vision*, 382–401. Springer.

Scheirer, W. J.; de Rezende Rocha, A.; Sapkota, A.; and Boult, T. E. 2013. Toward open set recognition. *IEEE TPAMI*, 35(7): 1757–1772.

Sun, Y.; and Li, Y. 2023. OpenCon: Open-world Contrastive Learning. *Transactions on Machine Learning Research*.

Vaze, S.; Han, K.; Vedaldi, A.; and Zisserman, A. 2022a. Generalized category discovery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7492–7501.

Vaze, S.; Pang, G.; Kirillov, A.; Yuille, A.; and Lin, T.-Y. 2022b. Open-set recognition using intra-class splitting. In *CVPR*.

Wang, H.; Vaze, S.; and Han, K. 2024. SPTNet: An Efficient Alternative Framework for Generalized Category Discovery with Spatial Prompt Tuning. In *International Conference on Learning Representations (ICLR)*.

Weng, T.; Xiao, J.; and Jiang, H. 2023. Decompose novel into known: Part concept learning for 3d novel class discovery. *Advances in Neural Information Processing Systems*, 36: 54282–54299.

Ye, B.; Gan, K.; Wei, T.; and Zhang, M.-L. 2024. Bridging the gap: learning pace synchronization for open-world semi-supervised learning. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, 5362–5370.

Zheng, H.; Pu, N.; Li, W.; Sebe, N.; and Zhong, Z. 2024. Prototypical hash encoding for on-the-fly fine-grained category discovery. *Advances in Neural Information Processing Systems*, 37: 101428–101455.

Zhong, Z.; Zhu, L.; Luo, Z.; Li, S.; Yang, Y.; and Sebe, N. 2021. OpenMix: Reviving Known Knowledge for Discovering Novel Visual Categories in an Open World. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 9462–9470.