LEARNING STYLE-INVARIANT TEMPORAL SEMANTICS FOR SOURCE-FREE TIME-SERIES DOMAIN ADAPTATION

Anonymous authors
Paper under double-blind review

ABSTRACT

We study source-free time-series domain adaptation (SFTSDA), where a source-trained predictor is transferred to an unlabeled target domain while source samples are unavailable. Existing approaches improve transfer through imputation, time–frequency modeling, and target self-training, yet target-specific temporal style may remain entangled with class-relevant semantics, causing style variation to be interpreted as class evidence and subsequently reinforced through pseudo-label adaptation. We propose style-invariant temporal semantics (SITS), which formulates semantic invariance through target-style orbit learning. Target-only channel and spectral statistics define closed-form style transformations, while evolving temporal semantic neighborhoods determine which transformations are admissible for each sample, forming a target-style orbit. SITS then learns relation-aware temporal semantics by maximizing the worst-case class margin over each orbit. Progressive pseudo-posterior refinement updates the semantic neighborhoods used to construct the next orbit. We further show that a positive orbit margin preserves the predicted class over the covered target-style family, providing a direct link between the orbit-wise objective and style-invariant temporal semantics. Experiments on three real-world time-series datasets evaluate the proposed SITS framework against representative source-free and test-time adaptation baselines. The code is available at `https://github.com/Sky-byte-box/SITS_code`.

1 INTRODUCTION

Time-series models are increasingly deployed across diverse environments where variations in subjects, sensors, acquisition conditions, and operating states inevitably introduce distribution shifts between training and deployment domains. Such shifts can alter temporal and spectral patterns, leading to degraded prediction performance when models are directly transferred to unseen domains. Conventional time-series domain adaptation mitigates this issue by leveraging labeled source data through temporal alignment, adversarial learning, self-supervised adaptation, and time–frequency representation transfer (Cai et al., 2021; Liu & Xue, 2021; Ragab et al., 2022; He et al., 2023; Liu et al., 2024a). However, in many real-world applications, source sequences cannot be accessed after deployment due to privacy, storage, or data-ownership constraints. Source-free time-series domain adaptation (SFTSDA) therefore aims to adapt a pretrained source model using only unlabeled target sequences, providing a practical solution for cross-domain time-series analysis in applications such as human activity recognition, industrial monitoring, and physiological signal interpretation.

Recent SFTSDA research addresses domain shifts through temporal structure modeling and target prediction refinement. The former exploits sequential and spectral dependencies, whereas the latter improves the supervision obtained from unlabeled target sequences. For temporal modeling, the mask and impute (MAPU) framework uses masked temporal imputation to retain transferable dynamics without revisiting source samples (Ragab et al., 2023). Time-frequency domain adaptation (TFDA) jointly exploits temporal and spectral representations, while wavelet-based multi-scale temporal imputation (LEMON) combines multi-scale imputation with energy-aware fusion (Furqon et al., 2025; Zhong et al., 2025). Position regularization-based temporal-spectral imputation (PRTSI) further preserves key-pattern positions and long-range dependencies (Lin et al., 2026). For target

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

prediction refinement, evidential MAPU (E-MAPU) estimates uncertainty to reduce overconfident predictions, and classifier ensemble-based source-free domain adaptation (CE-SFDA) combines fixed and learnable classifiers for pseudo-label generation (Ragab et al., 2026; Pei et al., 2025). Prior evaluation of model salience with posterior adaptation (PrEPoA) uses salience evaluation and prototype-based updates, while class-aware temporal imputation (CATI) introduces semantic guidance for discriminative temporal details (Cai et al., 2026; Zhong et al., 2026). These developments improve temporal modeling and target supervision, motivating a closer examination of how class evidence responds to target-specific style variation.

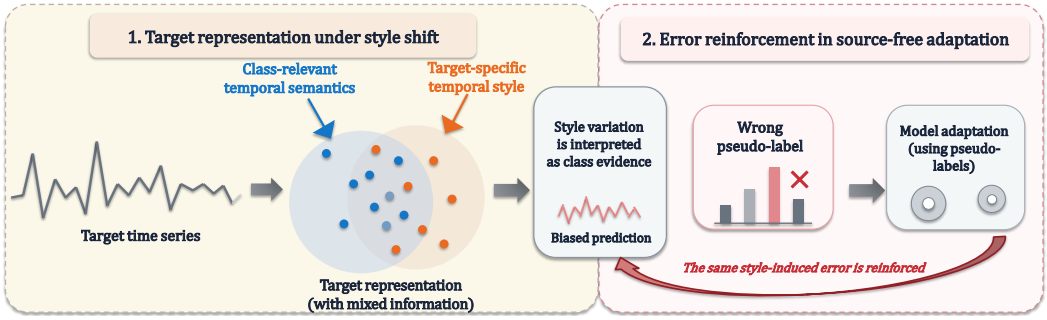


Figure 1: Target-specific temporal style may be interpreted as class evidence and subsequently reinforced through pseudo-label-guided adaptation.

We should note that improvements in temporal modeling and target prediction refinement do not by themselves ensure that class evidence is preserved under target-specific style variation. Acquisition-related changes in amplitude, channel offset, or spectral emphasis can affect a source model's predictions even when the underlying class remains unchanged. These factors can also carry task-relevant information, so their variation cannot be treated as universally label-preserving. The challenge is to identify plausible style changes that remain compatible with each sample's temporal semantics and to retain class evidence across those changes. If style-induced errors enter the initial target predictions, subsequent confidence filtering or pseudo-label updating may reinforce them because target predictions provide adaptation supervision. Addressing this failure mode therefore requires a connection between the selection of admissible style variations and the learning of class-relevant temporal representations. Fig. 1 illustrates how style-sensitive predictions can be reinforced through pseudo-label-guided adaptation.

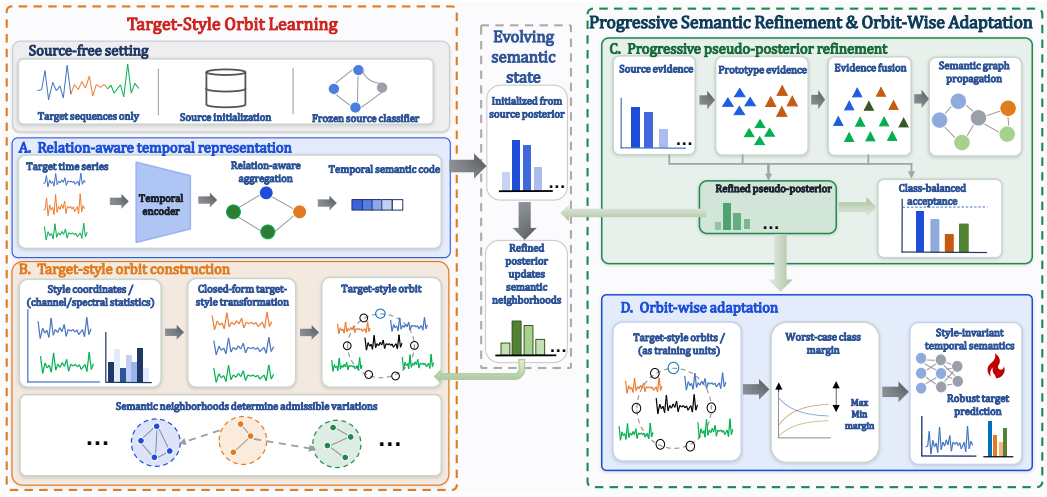


Figure 2: Overview of SITSfor source-free time-series domain adaptation.

To address this issue, we propose style-invariant temporal semantics (SITS) for source-free time-series domain adaptation, as illustrated in Fig. 2. Source predictions initialize the semantic state used to guide subsequent target adaptation. Relation-aware temporal representations first capture local transitions from unlabeled target sequences. Target-only channel and spectral statistics then define

closed-form style transformations, while the current semantic neighborhood identifies compatible style donors for constructing each sample's target-style orbit. Progressive pseudo-posterior refinement combines source and prototype evidence with semantic graph propagation to update both adaptation supervision and the neighborhoods used for orbit construction. This gives refinement a direct role in selecting the style variations against which class evidence is evaluated. The representation is subsequently updated by maximizing the weakest class margin across each admissible orbit. The updated semantic geometry guides the next refinement round, coupling target prediction refinement with style-invariant representation learning. Section 3 presents the formulation and optimization procedure.

Our contributions are summarized as follows. (1) We propose SITS, a framework for learning relation-aware temporal semantics that preserve class evidence across target-supported style variations during source-free time-series adaptation. (2) We develop target-only closed-form style transformations and an evolving semantic neighborhood to identify compatible style donors and construct sample-specific target-style orbits. (3) We couple progressive pseudo-posterior refinement with worst-case orbit-margin learning and show, under explicit coverage and margin assumptions, when the predicted class is preserved throughout the covered target-style family. (4) We evaluate SITS on three real-world time-series benchmarks against representative adaptation baselines through cross-domain comparisons, ablation studies, parameter sensitivity analysis, and representation visualization.

2 Related Work

2.1 Test-time adaptation

Existing test-time adaptation (TTA) methods predominantly update a deployed model using unlabeled test data, typically through entropy minimization, classifier adjustment, or reliability-aware model updates. Representative approaches include test-time entropy minimization (TENT) (Wang et al., 2021a), the test-time template adjuster (T3A) (Iwasawa & Matsuo, 2021), efficient anti-forgetting test-time adaptation (EATA) (Niu et al., 2022), and sharpness-aware and reliable entropy minimization (SAR) (Niu et al., 2023). While effective for improving test-time predictions, direct adaptation can become more challenging when the test stream is non-stationary or exhibits temporal dependence. Continual test-time adaptation (CoTTA) and robust adaptation against temporal correlation therefore consider evolving test streams (Wang et al., 2022; Gong et al., 2022), whereas robust test-time adaptation (RoTTA) and non-parametric classifier adaptation (AdaNPC) exploit dynamic memory or neighborhood structure (Yuan et al., 2023; Zhang et al., 2023). The training-free dynamic adapter (TDA) further introduces lightweight caches (Karmanov et al., 2024), and augmented contrastive clustering with uncertainty-aware prototyping (ACCUP) extends TTA to time-series data through uncertainty-aware prototypes and contrastive clustering (Gong et al., 2025). Target prediction refinement also commonly uses confidence-based sample selection, consistency regularization, noise-robust learning, and prototype- or neighborhood-based updates (Sohn et al., 2020; Zhang & Sabuncu, 2018; Diamant et al., 2024). Nevertheless, these strategies primarily improve target predictions themselves. In SITS, the refined semantic state additionally determines the admissible target-style neighborhoods over which invariance is learned.

2.2 Style and disentangled time-series representations

Style-related representation learning has been widely explored in domain generalization, where feature statistics and synthesized variations are used to reduce reliance on domain-specific cues. MixStyle perturbs feature statistics, learning to diversify expands the source distribution, and progressive random convolutions generate appearance variations (Zhou et al., 2021; Wang et al., 2021b; Choi et al., 2023). Related studies further consider min–max stylization, frequency characteristics, contrastive style regularization, and seasonal–trend decomposition to separate semantic information from style-like variation in structured representations (Liu et al., 2024b; He et al., 2024; Matsun et al., 2025; Woo et al., 2022). These observations motivate the use of target-side temporal and spectral statistics in time-series adaptation. In SITS, such statistics are not modeled by an explicit style branch; instead, they define a transformation family over which temporal semantics are required to remain invariant.

3 METHODOLOGY

3.1 PROBLEM FORMULATION

Source-free adaptation cannot revisit source observations to reconstruct or explicitly align the source and target distributions. The transferable supervision is therefore confined to the deployed source hypothesis, while the unlabeled target set provides the evidence needed to characterize the new domain. We begin from the source model's response to each target sequence and use it as the initial semantic state for subsequent adaptation. Let a source-trained classifier be $g_0 = h_{\phi_0} \circ f_{\theta_0}$, where f is a time-series encoder, h is a linear classifier, and θ_0 and ϕ_0 are their source-trained parameters, respectively. During adaptation, only unlabeled target samples $\mathcal{D}_t = \{x_i\}_{i=1}^{N_t}$ are observed. For a K-class problem, the source hypothesis provides the initial target posterior $p_i^{(0)} = \mathrm{softmax}(h_{\phi_0}(f_{\theta_0}(x_i))) \in \mathbb{R}^K$, which may be biased by target-specific style. Our objective is to learn a temporal semantic representation c_i that preserves class-relevant temporal structure while remaining insensitive to plausible target-specific style variation. We refer to this representation as *style-invariant temporal semantics*. In our formulation, style is not represented by a second learned branch. It is described by target-only statistics that generate a family of closed-form transformations against which the semantic representation is tested.

We use only weak supervision derived from the target batch. No target class annotation or source example is introduced. For an input $x_i \in \mathbb{R}^{C\times L}$, where C and L denote the number of channels and sequence length, the encoder produces T temporal tokens $H_i \in \mathbb{R}^{T\times d_h}$ with token dimension d_h. To retain local temporal transitions during aggregation, each token is weighted according to both its current state and its first-order change as follows:

$$z_i = \sum_{t=1}^{T} \frac{\exp\big(w_a^\top \varphi([H_{i,t},\, H_{i,t} - H_{i,t-1}])\big)}{\sum_{\ell=1}^{T} \exp(w_a^\top \varphi([H_{i,\ell},\, H_{i,\ell} - H_{i,\ell-1}]))} H_{i,t}, \tag{1}$$

where $H_{i,0} = H_{i,1}$, w_a denotes the learnable scoring vector, and $\varphi(\cdot)$ is a lightweight temporal relation projector. The normalized coefficient therefore favors tokens with informative states and local temporal changes.

3.2 LEARNING STYLE-INVARIANT TEMPORAL SEMANTICS

3.2.1 TARGET-STYLE COORDINATES AND CLOSED-FORM ACTION

Although Eq. (1) preserves local temporal transitions, the aggregated representation can still inherit acquisition-dependent variation that should not determine class identity. We therefore characterize target style directly from each unlabeled sequence using low-order channel statistics and coarse spectral energy. For channel u, let μ_{iu} and σ_{iu} be the temporal mean and standard deviation, and let E_{ib} be the normalized energy in frequency band b. We define:

$$d_i = \Big[\mu_{i1:C},\ \log(\sigma_{i1:C} + \epsilon),\ \log(E_{i1:B} + \epsilon)\Big], \tag{2}$$

where B is the number of coarse frequency bands and $\epsilon > 0$ is a small numerical constant. The vector d_i provides a target-only description of sequence style rather than a class label or reconstruction target. To transfer these factors without introducing a learned style decoder, we normalize channel u as $\nu_i^u(t) = (x_i^u(t) - \mu_{iu})/(\sigma_{iu} + \epsilon)$ and define the band-wise transfer ratio $\zeta_{ijb} = \sqrt{(E_{jb} + \epsilon)/(E_{ib} + \epsilon)}$, where t indexes time and $b(\omega)$ denotes the coarse frequency band containing frequency ω. The ratio ζ_{ijb} transfers the spectral energy of band b from sample j to sample i. Based on these quantities, the temporal content of x_i is expressed with the target statistics of x_j as:

$$[\mathcal{T}_{i\leftarrow j}(x_i)]^u = \mu_{ju} + \sigma_{ju}\, \mathcal{F}^{-1}\big[\zeta_{ij,b(\omega)}\, \mathcal{F}(\nu_i^u)(\omega)\big], \tag{3}$$

where $\mathcal{F}$ and $\mathcal{F}^{-1}$ denote the Fourier and inverse Fourier transforms, respectively. The resulting transformation is fully determined by target observations and introduces no trainable style generator. To evaluate whether class-relevant temporal information is preserved under these style changes, the aggregated feature in Eq. (1) is mapped to the classifier-facing semantic code $c_i = c_\Theta(x_i) = \mathrm{norm}(W_c z_i)$, with semantic posterior $p_i^{\mathrm{sem}} = \mathrm{softmax}(h_{\phi_0}(c_i))$. Here, W_c is the trainable semantic projection, $\mathrm{norm}(\cdot)$ denotes ℓ_2 normalization, and Θ collects the target-adapted encoder parameters and W_c, while h_{ϕ_0} remains frozen.

3.2.2 EVOLVING TEMPORAL SEMANTIC NEIGHBORHOOD

The semantic code c_i provides the representation on which style invariance is evaluated, but it does not yet specify which target samples can exchange style without changing class meaning. Since target labels are unavailable, this compatibility must be inferred from local semantic structure and prediction agreement. We therefore generate two mild views x_i^w and x_i^a of each target sample using amplitude-preserving jitter and local masking. For $v \in \{w, a\}$, let $p_i^v = \mathrm{softmax}(h_{\phi_0}(c_\Theta(x_i^v)))$. Their prediction agreement defines below:

$$q_i = \frac{1}{2}\left(p_i^w + p_i^a\right), \qquad \kappa_i = 1 - \frac{H(q_i)}{\log K}, \tag{4}$$

where $H(q) = -\sum_{k=1}^{K} q_k \log q_k$ is Shannon entropy and $\kappa_i \in [0, 1]$ is prediction confidence. Let $\mathbf{C}_\Theta = [c_\Theta(x_i)]_{i=1}^{N_t}$ and $Q = [q_i]_{i=1}^{N_t}$. Mutual proximity and posterior agreement are encoded directly in the semantic affinity, *i.e.,*

$$\begin{aligned} A_{ij} = \mathbb{I}[j \in \mathcal{N}_{K_{\mathrm{nn}}}^{\mathbf{C}_\Theta}(i)]\mathbb{I}[i \in \mathcal{N}_{K_{\mathrm{nn}}}^{\mathbf{C}_\Theta}(j)]\mathbb{I}[q_i^\top q_j \geq \xi]\sqrt{\kappa_i \kappa_j}\,(q_i^\top q_j) \\ \times \frac{1 + \cos(c_i, c_j)}{2}, \end{aligned} \tag{5}$$

where $\mathcal{N}_{K_{\mathrm{nn}}}^{\mathbf{C}_\Theta}(i)$ is the K_{nn}-nearest-neighbor set in the current semantic space, $\xi \in [0, 1]$ is the semantic-compatibility threshold, and $\mathbb{I}[\cdot]$ is the indicator function. The admissible neighborhood is then $\mathcal{N}_i(Q, \mathbf{C}_\Theta) = \{j : A_{ij} > 0\}$, and the resulting graph is denoted by $A_\Theta(Q) = \mathcal{G}(\mathbf{C}_\Theta, Q)$. Thus, changes in either the semantic code or the posterior state directly update the styles that can be exchanged.

3.2.3 TARGET-STYLE ORBIT

The neighborhood induced by Eq. (5) identifies target samples whose styles can be transferred without relying on class annotations. To evaluate whether the semantic code remains discriminative under these admissible changes, we collect the original sequence and its neighborhood-induced transformations into a sample-specific set. For each target sample, this set is defined as:

$$\mathcal{O}_i(Q) = \{x_i\} \cup \{\mathcal{T}_{i \leftarrow j}(x_i) : j \in \mathcal{N}_i(Q, \mathbf{C}_\Theta)\}, \tag{6}$$

where $\mathcal{O}_i(Q)$ contains the original sequence and style-transformed realizations supported by semantically compatible neighbors. We refer to it as the *target-style orbit* of x_i, over which semantic invariance is evaluated while preserving its temporal content. Let $F(c) = h_{\phi_0}(c)$ denote the frozen source logits and $\widehat{y}_i(Q) = \arg\max_k q_{ik}$. Rather than introducing a separate margin for each orbit element, we directly measure the weakest class evidence over the orbit as:

$$\underline{m}_i(\Theta, Q) = \inf_{\widetilde{x} \in \mathcal{O}_i(Q)} \left[F_{\widehat{y}_i(Q)}(c_\Theta(\widetilde{x})) - \max_{k \neq \widehat{y}_i(Q)} F_k(c_\Theta(\widetilde{x})) \right], \tag{7}$$

where $F_k(\cdot)$ denotes the kth frozen-source logit. A positive $\underline{m}_i$ means that every admissible target-style realization preserves the same class decision.

3.3 PROGRESSIVE TEMPORAL SEMANTIC REFINEMENT

The target-style orbit depends on the semantic neighborhood, but the initial source hypothesis may provide an unreliable neighborhood under domain shift. We therefore estimate the latent target semantic state progressively. For each sample, we summarize the two-view uncertainty from Eq. (4) by $u_i = H(q_i)/\log K + \frac{1}{2}\|p_i^w - p_i^a\|_1$ and convert it to the reliability weight $w_i = \exp(-u_i/\tau_u)$, where $\tau_u > 0$ controls how strongly uncertain samples are down-weighted. Given $Q^{(r)} = [q_i^{(r)}]_{i=1}^{N_t}$ at outer round r, class k is represented by the reliability-weighted prototype $\mu_k^{(r)} = (\sum_i w_i q_{ik}^{(r)} c_i)/(\sum_i w_i q_{ik}^{(r)} + \epsilon)$. Its prototype posterior is $\pi_{ik}^{(r)} \propto \exp(\cos(c_i, \mu_k^{(r)})/\tau_p)$, normalized over the K classes, where $\tau_p > 0$ is the prototype temperature. We then fuse source-hypothesis and prototype evidence with a normalized product of experts,

$$\bar{q}_{ik}^{(r)} = \frac{(p_{ik}^{(0)} + \epsilon)^{1-\beta_r}(\pi_{ik}^{(r)} + \epsilon)^{\beta_r}}{\sum_{\ell=1}^{K}(p_{i\ell}^{(0)} + \epsilon)^{1-\beta_r}(\pi_{i\ell}^{(r)} + \epsilon)^{\beta_r}}, \tag{8}$$

where $\beta_r \in [0, 1]$ is the round-dependent interpolation coefficient that gradually shifts the prediction from source guidance toward target semantic structure. The resulting evidence is then propagated

over the current semantic graph. With $A = [A_{ij}]$, $D_{ii} = \sum_j A_{ij}$, and $S = D^{-1/2}AD^{-1/2}$, the inner diffusion starts from $Q_0^{(r)} = \bar{Q}^{(r)}$, where $\bar{Q}^{(r)} = [\bar{q}_i^{(r)}]_{i=1}^{N_t}$, and proceeds as

$$Q_{t+1}^{(r)} = (1-\gamma)Q_0^{(r)} + \gamma S Q_t^{(r)}, \qquad 0 < \gamma < 1, \tag{9}$$

where t is the inner diffusion index and $\gamma \in (0,1)$ is the graph-diffusion coefficient. After T_g inner steps, rows are normalized to obtain $Q^{(r+1)}$. Let $s_i^{(r+1)} = \max_\ell q_{i\ell}^{(r+1)}$ and $\widehat{y}_i^{(r+1)} = \arg\max_\ell q_{i\ell}^{(r+1)}$ denote the confidence and predicted class of sample i. To avoid a curriculum dominated by majority classes, for each class k we set $\delta_k^{(r)}$ to the $(1-\rho_r)$-quantile of $\{s_i^{(r+1)} : \widehat{y}_i^{(r+1)} = k\}$, where $\rho_r \in (0,1)$ increases monotonically from $\rho_{\min}$ to $\rho_{\max}$. Samples satisfying $s_i^{(r+1)} \geq \delta_{\widehat{y}_i^{(r+1)}}^{(r)}$ form $\mathcal{M}^{(r)}$, and $\mathcal{M}(Q)$ denotes the corresponding class-balanced acceptance operator. For fixed $(\bar{Q}, S)$ within an outer round, Eq. (9) induces $\Psi_\Theta(Q; \bar{Q}, S) = (1-\gamma)\bar{Q} + \gamma SQ$. Because $0 < \gamma < 1$ and $\|S\|_\infty \leq 1$, this map is contractive and converges to the unique fixed point satisfying $Q^\star = \Psi_\Theta(Q^\star; \bar{Q}, S)$. We then define $\widehat{y}_i^\star = \arg\max_k q_{ik}^\star$, $\mathcal{O}_i^\star = \mathcal{O}_i(Q^\star)$, and $\mathcal{M}^\star = \mathcal{M}(Q^\star)$, where $q_{ik}^\star$ is the kth entry of $Q^\star$. These quantities provide the refined semantic state used by the orbit-wise adaptation step described next.

3.4 Orbit-wise source-free adaptation

Given the refined state $Q^\star$, each accepted sample is associated with a pseudo-class and a target-style orbit determined by its current semantic neighborhood. The representation should preserve the corresponding class evidence across all realizations in that orbit rather than only for the observed sequence. Accordingly, the complete source-free adaptation criterion is

$$\Theta^\star = \arg\max_\Theta \frac{1}{|\mathcal{M}^\star|} \sum_{i \in \mathcal{M}^\star} \inf_{\widetilde{x} \in \mathcal{O}_i^\star} \left[F_{\widehat{y}_i^\star}(c_\Theta(\widetilde{x})) - \max_{k \neq \widehat{y}_i^\star} F_k(c_\Theta(\widetilde{x})) \right], \tag{10}$$

where $F_k(\cdot)$ is the kth frozen-source logit and $|\mathcal{M}^\star|$ is the number of accepted target samples. Eq. (10) maximizes the weakest class-relevant temporal margin over the admissible target-style orbit of each accepted sample. Consequently, adaptation favors semantic representations whose class evidence is preserved across target-supported style variations.

This orbit-wise criterion also yields a direct preservation property beyond the finite orbit itself.

Theorem 1 (Orbit-wise class preservation)**.** *Let $\mathcal{S}_i$ be the plausible target-style family of sample i. Assume $F = h_{\phi_0}$ is L_c-Lipschitz in the semantic code and that $\mathcal{O}_i^\star$ is an ε_i-cover of $\mathcal{S}_i$ after semantic encoding: for every $x' \in \mathcal{S}_i$, there exists $\widetilde{x} \in \mathcal{O}_i^\star$ such that $\|c_\Theta(x') - c_\Theta(\widetilde{x})\|_2 \leq \varepsilon_i$. If $\underline{m}_i(\Theta, Q^\star) > 2L_c\varepsilon_i$, then*

$$\arg\max_k F_k(c_\Theta(x')) = \widehat{y}_i^\star, \qquad \forall x' \in \mathcal{S}_i, \tag{11}$$

where F_k is the kth source logit and $\widehat{y}_i^\star$ is the pseudo-class induced by the converged semantic state $Q^\star$.

Proof. By definition of $\underline{m}_i$, every orbit element has margin at least $\underline{m}_i$. For any $x' \in \mathcal{S}_i$, choose its covered orbit element $\widetilde{x}$. Lipschitz continuity changes the correct-class logit and any competing logit by at most $L_c\varepsilon_i$, so the margin at x' is at least $\underline{m}_i - 2L_c\varepsilon_i > 0$. Hence the predicted class is preserved. □

For gradient-based optimization, the inner infimum is replaced by the differentiable soft minimum $\operatorname{smin}_{\tau_m}(a_1, \ldots, a_J) = -\tau_m \log \sum_{j=1}^J \exp(-a_j/\tau_m)$ over the finite orbit, where J is the orbit cardinality and $\tau_m > 0$ controls the approximation to the worst case. We then perform gradient ascent on the resulting smooth surrogate of Eq. (10) with step size $\eta_\Theta > 0$. The next refinement round is computed in the updated semantic geometry, allowing Q, $\mathcal{N}_i$, and $\mathcal{O}_i$ to evolve with the representation. The complete target-adaptation procedure, from semantic refinement and orbit construction to representation update, is summarized in Algorithm 1.

Algorithm 1 SITS orbit-wise source-free target adaptation

Require: source model g_0, target data $\mathcal{D}_t$, rounds R, graph steps T_g

```
Initialize target adapter and Q^(0) from g_0
Compute target style coordinates d_i by Eq. (2)
for r = 0 to R − 1 do
    Encode target sequences by Eq. (1) and obtain c_i through the semantic projection
    Build A_Θ and derive N_i from Eq. (5)
    Refine Q^(r) using prototype evidence, Eqs. (8)–(9), and class-balanced acceptance
    Construct O_i(Q^(r+1)) with the closed-form style action in Eqs. (3) and (6)
    Update Θ by gradient ascent on the smooth surrogate of Eq. (10)
end for
return refined target predictions
```

4 Experiments

4.1 Experimental Preparation

Datasets. We conduct experiments on UCIHAR, MFD, and Sleep-EDF, covering human activity recognition, machine fault diagnosis, and sleep-stage classification (Anguita et al., 2013; Lessmeier et al., 2016; Goldberger et al., 2000). UCIHAR contains 9-channel signals from six activity classes, with 128 time points per sample and 2,300 source and 990 target samples in the adopted split. MFD contains three fault classes represented by single-channel vibration sequences of length 5,120, with 7,312 source and 3,604 target samples. Sleep-EDF contains five sleep stages represented by single-channel EEG segments of length 3,000, with 14,280 source and 6,130 target samples. Each subject or operating condition is treated as one domain. We consider two source-to-target transfers for each benchmark, namely $9 \to 18$ and $12 \to 16$ for UCIHAR, $1 \to 2$ and $1 \to 0$ for MFD, and $12 \to 5$ and $9 \to 14$ for Sleep-EDF.

Baseline Methods. We compare SITS with representative test-time adaptation methods and recent source-free time-series adaptation methods using the same backbone and training configurations. The comparison includes Source Only, TENT (Wang et al., 2021a), SHOT-IM (Liang et al., 2020), EATA (Niu et al., 2022), SAR (Niu et al., 2023), RoTTA (Yuan et al., 2023), ACCUP (Gong et al., 2025), CATI (Zhong et al., 2026), and LEMON (Zhong et al., 2025). Source Only is included as the non-adapted reference, while CATI and LEMON provide recent comparisons from the perspectives of semantic enhancement and domain-invariant temporal dynamics. To account for potential class imbalance, we use Macro-F1 (MF1) as the evaluation metric. We report the mean and standard deviation over three runs for each source-to-target transfer.

Implementation Details. To ensure a controlled comparison, all methods employ the same three-block one-dimensional convolutional encoder with 64, 128, and 128 filters. Specifically, the source model is pretrained for 40 epochs using Adam with a batch size of 32 and a learning rate of 10^{-3}. During source-free adaptation, the source classifier of SITS is kept frozen. We set the adaptation learning rate to 3×10^{-4} for UCIHAR and MFD and 10^{-5} for Sleep-EDF. The remaining hyperparameters are fixed across the corresponding source-to-target transfers, with $K_{\mathrm{nn}} = 8$, $\xi = 0.7$, $\tau_u = 0.1$, $\tau_p = 0.1$, $\gamma = 0.8$, $T_g = 10$, five refinement rounds, an acceptance ratio increasing from 0.25 to 0.75, and $\tau_m = 0.1$. This common backbone and training configuration allows the comparison to focus on differences among the adaptation strategies.

4.2 Comparative Experiments

Table 1 presents the quantitative comparison across six source-to-target transfers on UCIHAR, MFD, and Sleep-EDF. Bold and underline denote the best and second-best results, respectively. SITS achieves the highest MF1 in all six cases, exceeding the second-best results by average margins of 0.69, 0.45, and 0.62 percentage points on UCIHAR, MFD, and Sleep-EDF, respectively. Compared with ACCUP, SITS improves MF1 by 2.80, 1.56, 1.18, and 1.68 points on transfers $12 \to 16$, $1 \to 2$, $1 \to 0$, and $9 \to 14$, respectively. Its average improvement over Source Only is 12.09 points across all six transfers, showing the benefit of adapting the pretrained model to the target distribution. The

consistent gains across activity recognition, fault diagnosis, and sleep-stage classification support the effectiveness of the proposed framework on heterogeneous time-series tasks.

Table 1: Macro-F1 (%) on the selected transfer pairs across the three time-series datasets.

	UCIHAR		MFD		Sleep-EDF	
Method	$9 \to 18$	$12 \to 16$	$1 \to 2$	$1 \to 0$	$12 \to 5$	$9 \to 14$
Source Only	75.38±2.57	58.69±1.31	78.45±2.12	72.68±2.93	53.59±0.75	58.67±3.60
TENT	81.95±3.30	60.15±1.14	88.02±0.90	88.35±0.68	57.66±1.59	63.43±1.83
SHOT-IM	79.09±2.19	61.15±3.15	86.17±2.94	67.87±5.70	60.00±1.11	66.66±2.30
EATA	82.71±3.78	60.36±1.30	87.93±0.92	88.30±0.54	58.38±1.42	64.07±1.46
SAR	82.49±3.32	60.37±1.31	88.06±0.92	88.32±0.73	57.74±1.56	63.42±1.87
RoTTA	79.92±2.32	59.10±1.32	85.83±1.84	81.10±4.33	51.64±0.87	56.31±2.76
ACCUP	83.45±3.39	64.60±4.92	89.69±0.53	88.87±5.16	63.99±1.57	70.02±1.17
CATI	84.18±2.74	66.35±3.58	90.82±0.58	89.58±2.11	64.52±1.41	71.05±1.14
LEMON	83.92±2.96	65.94±3.81	90.63±0.61	89.42±2.24	64.11±1.52	70.76±1.26
SITS	**84.50±2.85**	**67.40±3.76**	**91.25±0.64**	**90.05±1.73**	**65.10±1.32**	**71.70±1.05**

4.3 Ablation studies

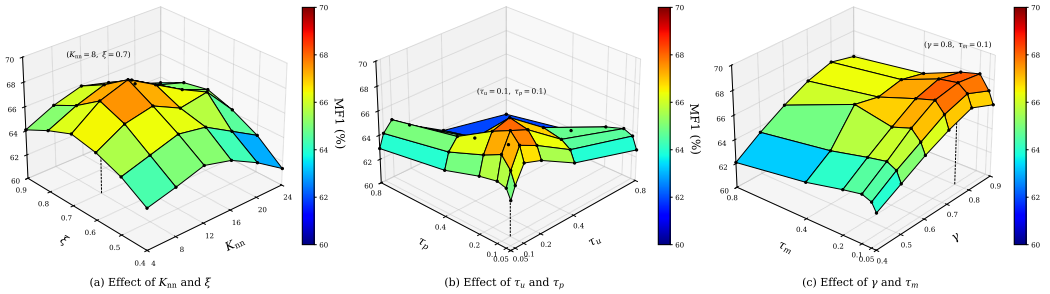
Table 2 presents the ablation results averaged over the two selected transfers on each dataset. Removing target-style orbit construction produces the largest MF1 decreases of 7.30, 6.75, and 8.60 points on UCIHAR, MFD, and Sleep-EDF, respectively. Removing progressive semantic refinement reduces MF1 by 6.50, 6.10, and 7.70 points, while decoupling style transfer from the evolving semantic neighborhood leads to decreases of 5.70, 5.25, and 6.90 points. These results support coupling the selection of admissible style variations with progressive refinement of the target semantic state. Removing relation-aware aggregation or replacing the weakest-margin criterion with mean-orbit optimization also reduces performance, although by smaller margins. Sleep-EDF exhibits the largest decrease for most variants. Across the three datasets, target-style orbit construction and progressive semantic refinement account for the largest gains among the evaluated components.

Table 2: Mechanism-focused ablation in macro-F1 (%) averaged over the two selected transfers of each dataset. Results are reported as mean ± standard deviation over three runs.

Variant	UCIHAR	MFD	Sleep-EDF
Target-style orbit construction			
w/o target-style orbit	68.65±3.10	83.90±1.55	59.80±1.72
Independent style transfer + refinement	70.25±2.94	85.40±1.40	61.50±1.55
Temporal semantic representation and refinement			
w/o relation-aware aggregation	72.15±2.72	87.10±1.21	63.90±1.24
w/o progressive semantic refinement	69.45±3.02	84.55±1.49	60.70±1.68
Orbit-wise optimization			
Mean-orbit objective	72.85±2.65	87.75±1.14	64.70±1.10
Full SITS	**75.95±2.36**	**90.65±0.92**	**68.40±0.84**

4.4 Parameter sensitivity analysis

We investigate the sensitivity of SITS to six key hyperparameters associated with semantic neighborhood construction, progressive refinement, and orbit-wise optimization. Fig. 3 presents the MF1 scores on Sleep-EDF under three pairwise parameter settings. The results demonstrate that SITS maintains strong performance over relatively broad parameter regions, while inappropriate settings lead to noticeable degradation. For K_{nn} and ξ, the performance generally improves as the semantic neighborhood becomes more informative and then decreases when either parameter moves beyond a suitable range. Besides, the τ_u–τ_p surface exhibits a more asymmetric variation, where large temperature values produce a clearer performance drop because uncertainty weighting and prototype calibration become less discriminative. Notably, the γ–τ_m surface forms a relatively broad high-performance region, indicating that moderate variations in graph diffusion and soft-minimum approximation have a limited influence on the final prediction. Local optima can also be observed in Fig. 3(a) and Fig. 3(c), suggesting that the interaction between paired parameters is not strictly monotonic. Overall, the adopted settings $K_{\mathrm{nn}} = 8$, $\xi = 0.7$, $\tau_u = 0.1$, $\tau_p = 0.1$, $\gamma = 0.8$, and $\tau_m = 0.1$ provide consistently strong performance across the three parameter combinations.

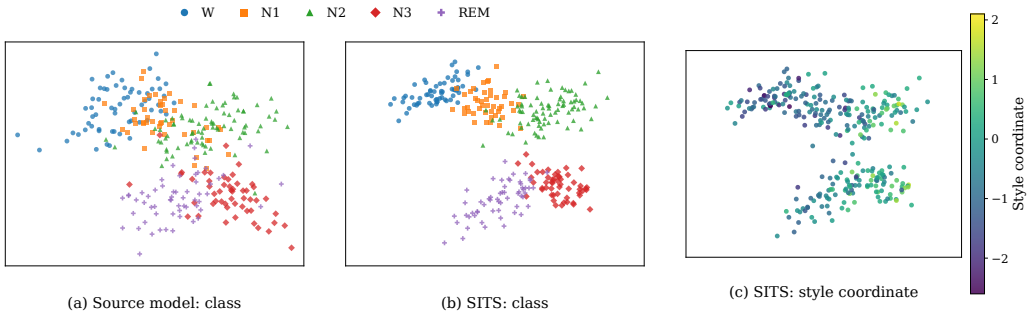


(a) Effect of K_{nn} and ξ (b) Effect of τ_u and τ_p (c) Effect of γ and τ_m

Figure 3: Pairwise parameter sensitivity on Sleep-EDF for (a) K_{nn} and ξ, (b) τ_u and τ_p, and (c) γ and τ_m.

4.5 Visualization Analysis of Style-Invariant Temporal Semantics

We further examine whether SITS reorganizes target representations around class-relevant semantics rather than target-specific style using t-SNE. Fig. 4 visualizes the same target samples from the source model and the adapted model. Compared with the source representation in Fig. 4(a), the features produced by SITS in Fig. 4(b) exhibit more compact class clusters and clearer separation between different classes, indicating that adaptation improves the discriminative semantic structure on the target domain. Moreover, when the same adapted features are colored by the target-only style coordinate in Fig. 4(c), the style values are broadly mixed within the learned class structure and do not form a pronounced style-oriented partition. This observation suggests that the adapted representation is less dominated by target-specific style variation while preserving class-relevant temporal semantics. Overall, the visualization provides qualitative evidence that SITS learns more discriminative and less style-dependent target representations.



(a) Source model: class (b) SITS: class (c) SITS: style coordinate

Figure 4: Visualization analysis of style-invariant temporal semantics on the target domain. The same target samples are visualized by t-SNE for (a) the source model colored by class, (b) SITS colored by class, and (c) SITS colored by the target-only style coordinate.

5 Conclusion

In this study, we propose a novel source-free time-series domain adaptation framework for learning style-invariant temporal semantics without accessing source data. More specifically, a target-style orbit construction strategy is first introduced to characterize the target-domain variations and preserve class-relevant temporal structures under domain shifts. Second, we develop a progressive temporal semantic refinement mechanism, which exploits target neighborhood consistency and pseudo-posterior guidance to progressively improve the discriminative capability of target representations. Third, we theoretically analyze the relationship between style variations and semantic representations, and establish the optimization objective to constrain the influence of target-specific styles during adaptation. Finally, based on the refined temporal semantics, the proposed framework generates more transferable target representations for final prediction. Experiments on three time-series benchmarks demonstrate that SITS achieves competitive performance against the considered adaptation baselines.

REFERENCES

Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra, and Jorge Luis Reyes-Ortiz. A public domain dataset for human activity recognition using smartphones. In *Proceedings of the European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, pp. 437–442, 2013.

Rongyao Cai, Ming Jin, Qingsong Wen, Kexin Zhang, and Yong Liu. Source-free time-series domain adaptation with prior evaluation of model salience. *IEEE Transactions on Neural Networks and Learning Systems*, 37(6):2884–2898, 2026.

Ruichu Cai, Jiawei Chen, Zijian Li, Wei Chen, Keli Zhang, Junjian Ye, Zhuozhang Li, Xiaoyan Yang, and Zhenjie Zhang. Time series domain adaptation via sparse associative structure alignment. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pp. 6859–6867, 2021.

Seokeon Choi, Debasmit Das, Sungha Choi, Seunghan Yang, Hyunsin Park, and Sungrack Yun. Progressive random convolutions for single domain generalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10312–10322, 2023.

Idit Diamant, Amir Rosenfeld, Idan Achituve, Jacob Goldberger, and Arnon Netzer. De-confusing pseudo-labels in source-free domain adaptation. In *Proceedings of the European Conference on Computer Vision*, 2024.

Muhammad Tanzil Furqon, Mahardhika Pratama, Ary Shiddiqi, Lin Liu, Habibullah Habibullah, and Kutluyil Dogancay. Time and frequency synergy for source-free time-series domain adaptations. *Information Sciences*, 695:121734, 2025.

Ary L. Goldberger, Luis A. N. Amaral, Leon Glass, Jeffrey M. Hausdorff, Plamen Ch. Ivanov, Roger G. Mark, Joseph E. Mietus, George B. Moody, Chung-Kang Peng, and H. Eugene Stanley. Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220, 2000.

Peiliang Gong, Mohamed Ragab, Min Wu, Zhenghua Chen, Yongyi Su, Xiaoli Li, and Daoqiang Zhang. Augmented contrastive clustering with uncertainty-aware prototyping for time series test time adaptation. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.1*, pp. 390–401, 2025.

Taesik Gong, Jongheon Jeong, Taewon Kim, Yewon Kim, Jinwoo Shin, and Sung-Ju Lee. Note: Robust continual test-time adaptation against temporal correlation. In *Proceedings of the 36th Conference on Neural Information Processing Systems (NeurIPS)*, volume 35, pp. 27253–27266, 2022.

Huan He, Owen Queen, Teddy Koker, Consuelo Cuevas, Theodoros Tsiligkaridis, and Marinka Zitnik. Domain adaptation for time series under feature and label shifts. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 12746–12774, 2023.

Xilin He, Jingyu Hu, Qinliang Lin, Cheng Luo, Weicheng Xie, Siyang Song, Muhammad Haris Khan, and Linlin Shen. Towards combating frequency simplicity-biased learning for domain generalization. In *Proceedings of the 38th Conference on Neural Information Processing Systems (NeurIPS)*, volume 37, 2024.

Yusuke Iwasawa and Yutaka Matsuo. Test-time classifier adjustment module for model-agnostic domain generalization. In *Proceedings of the 35th Conference on Neural Information Processing Systems (NeurIPS)*, volume 34, pp. 2427–2440, 2021.

Adilbek Karmanov, Dayan Guan, Shijian Lu, Abdulmotaleb El Saddik, and Eric Xing. Efficient test-time adaptation of vision-language models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14162–14171, 2024.

Christian Lessmeier, James Kuria Kimotho, Detmar Zimmer, and Walter Sextro. Condition monitoring of bearing damage in electromechanical drive systems by using motor current signals of electric motors: A benchmark data set for data-driven classification. In *Proceedings of the PHM Society European Conference*, volume 3, 2016.

Jian Liang, Dapeng Hu, and Jiashi Feng. Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation. In *Proceedings of the 37th International Conference on Machine Learning*, pp. 6028–6039, 2020.

Haojie Lin, Guangju Yang, and Tian-Jian Luo. Position regularization-based temporal-spectral imputation for source-free time-series domain adaptation. *Neurocomputing*, 661:131981, 2026.

Mingyang Liu, Xinyang Chen, Yang Shu, Xiucheng Li, Weili Guan, and Liqiang Nie. Boosting transferability and discriminability for time series domain adaptation. In *Proceedings of the 38th Conference on Neural Information Processing Systems (NeurIPS)*, volume 37, 2024a.

Qiao Liu and Hui Xue. Adversarial spectral kernel matching for unsupervised time series domain adaptation. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence*, pp. 2744–2750, 2021.

Songhua Liu, Xin Jin, Xingyi Yang, Jingwen Ye, and Xinchao Wang. Stydesty: Min-max stylization and destylization for single domain generalization. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 31293–31311, 2024b.

Aleksandr Matsun, Numan Saeed, Fadillah Adamsyah Maani, and Mohammad Yaqub. Condisr: Contrastive disentanglement and style regularization for single domain generalization. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 2881–2889, 2025.

Shuaicheng Niu, Jiaxiang Wu, Yifan Zhang, Yaofo Chen, Shijian Zheng, Peilin Zhao, and Mingkui Tan. Efficient test-time model adaptation without forgetting. In *Proceedings of the 39th International Conference on Machine Learning*, pp. 16888–16905, 2022.

Shuaicheng Niu, Jiaxiang Wu, Yifan Zhang, Zhiquan Wen, Yaofo Chen, Peilin Zhao, and Mingkui Tan. Towards stable test-time adaptation in dynamic wild world. In *Proceedings of the International Conference on Learning Representations*, 2023.

Ercheng Pei, Wangdong Zhao, Zhanxuan Hu, Lang He, Hailong Ning, and Haifeng Chen. Classifier ensemble based source-free domain adaptation for time series classification. *Knowledge-Based Systems*, 330:114584, 2025.

Mohamed Ragab, Emadeldeen Eldele, Zhenghua Chen, Min Wu, Chee-Keong Kwoh, and Xiaoli Li. Self-supervised autoregressive domain adaptation for time series data. *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

Mohamed Ragab, Emadeldeen Eldele, Min Wu, Chuan-Sheng Foo, Xiaoli Li, and Zhenghua Chen. Source-free domain adaptation with temporal imputation for time series data. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 1989–1998, 2023.

Mohamed Ragab, Peiliang Gong, Emadeldeen Eldele, Wenyu Zhang, Min Wu, Chuan-Sheng Foo, Daoqiang Zhang, Xiaoli Li, and Zhenghua Chen. Evidentially calibrated source-free time-series domain adaptation with temporal imputation. *IEEE Transactions on Knowledge and Data Engineering*, 38(1):290–306, 2026.

Kihyuk Sohn, David Berthelot, Chun-Liang Li, Zizhao Zhang, Nicholas Carlini, Ekin D. Cubuk, Alex Kurakin, Han Zhang, and Colin Raffel. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. In *Proceedings of the 34th Conference on Neural Information Processing Systems (NeurIPS)*, volume 33, pp. 596–608, 2020.

Dequan Wang, Evan Shelhamer, Shaoteng Liu, Bruno Olshausen, and Trevor Darrell. Tent: Fully test-time adaptation by entropy minimization. In *Proceedings of the International Conference on Learning Representations*, 2021a.

Qin Wang, Olga Fink, Luc Van Gool, and Dengxin Dai. Continual test-time domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7201–7211, 2022.

Zijian Wang, Yadan Luo, Ruihong Qiu, Zi Huang, and Mahsa Baktashmotlagh. Learning to diversify for single domain generalization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 834–843, 2021b.

Gerald Woo, Chenghao Liu, Doyen Sahoo, Akshat Kumar, and Steven Hoi. Cost: Contrastive learning of disentangled seasonal-trend representations for time series forecasting. In *Proceedings of the International Conference on Learning Representations*, 2022.

Longhui Yuan, Binhui Xie, and Shuang Li. Robust test-time adaptation in dynamic scenarios. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 15922–15932, 2023.

Yifan Zhang, Xue Wang, Kexin Jin, Kun Yuan, Zhang Zhang, Liang Wang, Rong Jin, and Tieniu Tan. Adanpc: Exploring non-parametric classifier for test-time adaptation. In *Proceedings of the 40th International Conference on Machine Learning*, pp. 41647–41676, 2023.

Zhilu Zhang and Mert Sabuncu. Generalized cross entropy loss for training deep neural networks with noisy labels. In *Proceedings of the 32nd Conference on Neural Information Processing Systems (NeurIPS)*, volume 31, 2018.

Yingyi Zhong, Wen'an Zhou, and Liwen Tao. Source-free time series domain adaptation with wavelet-based multi-scale temporal imputation. *Neural Networks*, 188:107428, 2025.

Yingyi Zhong, Wen'an Zhou, and Liwen Tao. Source-free time series domain adaptation with class-aware temporal imputation. *Information Processing & Management*, 63(5):104658, 2026.

Kaiyang Zhou, Yongxin Yang, Yu Qiao, and Tao Xiang. Domain generalization with mixstyle. In *Proceedings of the International Conference on Learning Representations*, 2021.