Temp-SCONE: Temporal OOD Detection and Generalization for Wild Data

SDSU San Diego State University

Aditi Naiknaware* Sanchit Singh* Hajar Homayouni Salimeh Sekeh

Department of Computer Science, San Diego State University

{anaiknaware7153, ssingh1949, hhomayouni, ssekeh}@sdsu.edu

*Equal contribution

NEURAL INFORMATION PROCESSING SYSTEMS

Introduction

- Open-world learning (OWL) requires models that can **adapt to evolving environments** while **reliably detecting OOD inputs**, rather than assuming a single, static distribution.
- In practice, deployed models face:
  - **Covariate shifts**: input distribution changes but labels stay the same (e.g., weather, noise, corruptions).
  - **Semantic shifts**: entirely new classes at test time that were unseen during training.

Original image   Covariate shift   Semantic shift

- Existing OWL/OOD frameworks (e.g., SCONE) handle **covariate + semantic OOD** under a **static** wild distribution, but do **not** explicitly model **temporal drift**.
- Real-world systems (traffic, user behavior, faces, etc.) exhibit **temporal shifts**: the wild data distribution **changes over time**, causing gradual performance degradation if not controlled.
- We aim to build a framework that:
  - **Retains SCONE's energy-based margin separation** for semantic OOD detection.
  - **Improves robustness under temporal drift** by enforcing **confidence stability** across time steps.
  - **Supports OWL on wild, evolving data** with both covariate and semantic shifts.

Hypothesis: Exploiting temporal-based confidence in SCONE improves OOD generalization in downstream time steps and controls performance shocks during domain transitions in dynamic environments, taking a step toward reliable OWL.

Problem Setup & Notations

- Let (x, y) denote an input–label pair and $f_\theta(x)$ denote the model's predictive logits.
- At each timestep t, the data consist of ID samples, covariate OOD samples, and semantic OOD samples.
- Temporal drift causes these distributions to vary across timesteps.
- Objective: maintain accurate ID classification, separate covariate vs. semantic OOD, and enforce temporally stable confidence.
- **Classifier:** $f_\theta : \mathcal{X} \to \mathbb{R}^K$, with prediction $\hat{y}(x) = \arg\max_y f_\theta^y(x)$.
- **Energy function:**

$$E_\theta(x) = -\log \sum_{y=1}^{K} e^{f_\theta^y(x)}.$$

- **OOD detector** $g_\theta(x)$ with decision rule

$$D_\theta(x) = \begin{cases} \text{ID}, & g_\theta(x) > \lambda, \\ \text{OOD}, & g_\theta(x) \le \lambda. \end{cases}$$

- **In-distribution (ID):** $(x, y) \sim P_{X,Y}$ with marginal P_{in}.
- **OOD types:**
  - Covariate OOD: $P_{\text{out}}^{\text{cov}}$ (shifted inputs, same labels).
  - Semantic OOD: $P_{\text{out}}^{\text{sem}}$ (novel classes unseen in training).
- **Static wild distribution:**

$$P_{\text{wild}} = \Big(1 - \sum_{\text{type}} \pi_{\text{type}}\Big) P_{\text{in}} + \sum_{\text{type}} \pi_{\text{type}} P_{\text{out}}^{\text{type}}, \quad \text{type} \in \{\text{covariate}, \text{semantic}\}.$$

- **Temporal wild distribution:**

$$P_{\text{wild},t} = \left(1 - \sum_{\text{type}} \pi_{\text{type},t}\right) P_{\text{in}} + \sum_{\text{type}} \pi_{\text{type},t} P_{\text{out},t}^{\text{type}}.$$

- **Max-confidence:**

$$s_{\max}(f_\theta(x)) = \max_j f_\theta^j(x).$$

$$\text{ATC}(s) = \mathbb{E}_{P_{\text{in}}}\Big[\mathbf{1}\{s(f_\theta(x)) < \delta\}\Big].$$

$$\text{AC}(s) = \mathbb{E}_{P_{\text{in}}}\big[s(f_\theta(x))\big].$$

The model should not undergo sharp confidence changes across timesteps. To express this requirement,

$$\left|\mathbb{E}_{P_{\text{in}}^{t+1}} \mathbf{1}\{s(f_\theta(x)) < \delta\} - \mathbb{E}_{P_{\text{in}}^{t}} \mathbf{1}\{s(f_\theta(x)) < \delta\}\right| \le \varepsilon.$$

Methodology

Overview. SCONE is designed to jointly improve (1) ID accuracy on P_{in}, (2) covariate OOD generalization on $P_{\text{out}}^{\text{cov}}$, and (3) semantic OOD detection on $P_{\text{out}}^{\text{sem}}$.

Energy Margin. It introduces a negative energy margin $\eta < 0$ that builds on the WOODS framework. The margin enforces:

$$E_\theta(x) < \eta \quad \text{for ID and covariate OOD,}$$

while semantic OOD samples are pushed into the high-energy region

$$E_\theta(x) > 0.$$

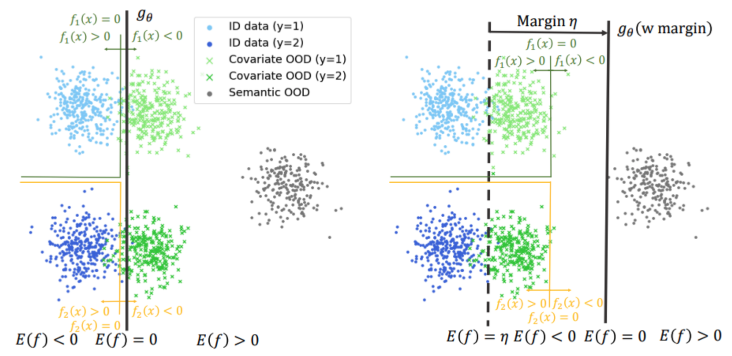
Resulting Separation.

- $E_\theta(x_{\text{ID}}) < \eta$ (semantically valid: ID + covariate OOD)
- $E_\theta(x_{\text{OOD}}) > 0$ (semantic OOD / novel classes)

Training Objective. The training objective combines:

- ID classification loss,
- energy-based OOD losses,
- augmented Lagrangian constraints

to enforce the desired margin structure while maintaining ID performance.



(a) No margin ($\eta = 0$)   (b) Energy margin ($\eta < 0$)

Source: Bai et al., ICML 2023.

Extension to Dynamic Environments. Temp-SCONE extends SCONE to dynamic environments by incorporating temporal confidence statistics (ATC and AC). Temporal drift is captured by changes in confidence between consecutive timesteps, and large deviations indicate instability under evolving distributions.

Temporal Constraint. To control this drift, Temp-SCONE imposes a temporal constraint using both ID and covariate-shifted samples. Because indicator functions are non-differentiable, differentiable surrogates are used to estimate drift.

Directional Drifts.

- ID drift:

$$d_{\text{ID}} = [p_{\text{in}}^{t-1} - s_{\text{in}}^{t}]_+$$

penalizes ID confidence drops.

- Covariate drift:

$$d_{\text{cov}} = [s_{\text{cov}}^{t} - p_{\text{cov}}^{t-1}]_+$$

penalizes confidence increases on covariate OOD.

Temporal Loss.

$$L_{\text{temp},t} = w_{\text{temp}}(d_{\text{ID}} + d_{\text{cov}}),$$

where w_{temp} is an adaptive temporal weight that increases when drift becomes large, ensuring smoother confidence behavior across timesteps.

At each timestep t, Temp-SCONE optimizes:

$$L_{\text{total}}^t = L_{\text{CE}}^t + \lambda_{\text{out}} L_{\text{out}}^t + \lambda_{\text{temp}} L_{\text{temp}}^t + L_{\text{ALM}}^t.$$

Theoretical Analysis

Let $G\text{Err}_t(f)$ denote the OOD generalization error at timestep t under covariate shift. Theorem 5.1 provides a lower bound on the change in generalization error between consecutive timesteps:

$$G\text{Err}_{t+1}(f) - G\text{Err}_t(f) \;\ge\; -\tilde{\kappa}\,\Delta_{t\to t+1}^{\text{cov,sem}} - \tilde{\kappa}\,\Xi_{t\to t+1}^{\text{sem}} - \delta_t^2\,\mathbb{E}_{P_{\text{out}}^{t,\text{cov}}}[I_F(\theta)] + C_{t\to t+1} + \text{Conf}_t - \text{Conf}_{t+1}.$$

Interpretation

- $\Delta_{t\to t+1}^{\text{cov,sem}}$: total-variation shift discrepancy between covariate and semantic OOD across timesteps.
- $\Xi_{t\to t+1}^{\text{sem}}$: contribution of the semantic OOD loss L_{reg} (MSP-based).
- $\mathbb{E}_{P_{\text{out}}^{t,\text{cov}}}[I_F(\theta)]$: Fisher information under covariate OOD at timestep t.
- $\text{Conf}_t, \text{Conf}_{t+1}$: model confidence drift across time.

Implications for Temp-SCONE

- Decreasing MSP-based OOD loss can **impair** temporal OOD generalization without proper regularization.
- The temporal OOD error gap depends on: OOD loss and distribution shift, Fisher information, Confidence drift $\text{Conf}_t - \text{Conf}_{t+1}$.
- Temp-SCONE's temporal loss controls confidence drift, improving stability under temporal shifts.
- Provides theoretical justification for confidence-based temporal regularization.

Results

Result 1: Yearbook (7 Timesteps).

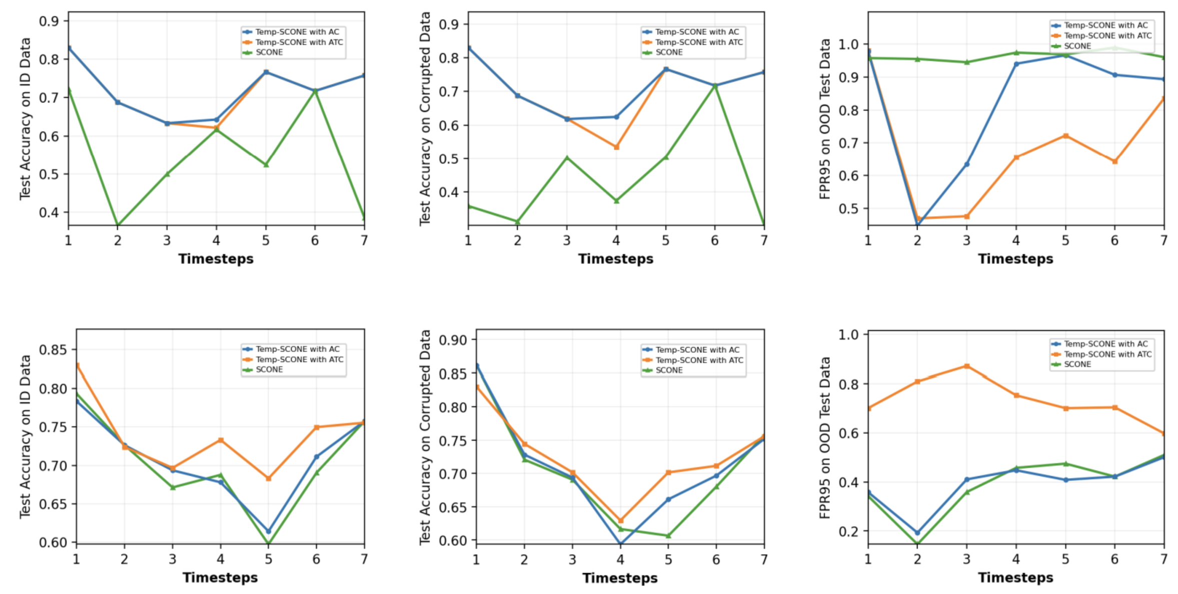


Figure 1: Dynamic Data (YearBook - 7 timesteps), FairFace is OOD data, (top row WRN, bottom row ViT). Columns show ID Acc.↑, OOD Acc.↑, FPR95 ↓.

Temp-SCONE achieves higher ID accuracy, OOD accuracy, and lower FPR95 than SCONE across all timesteps, with reduced confidence and performance variability.

Result 2: CLEAR (10 Timesteps).

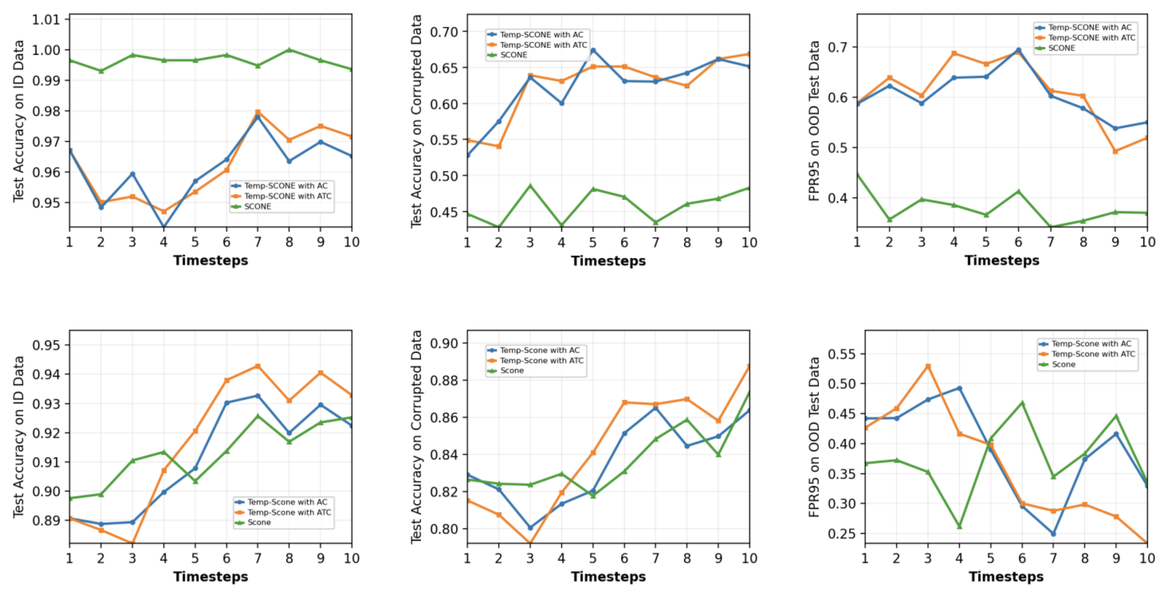


Figure 2: Dynamic Data (CLEAR - 10 timesteps), Places365 is OOD data, (top row WRN, bottom row ViT). Columns show ID Acc.↑, OOD Acc.↑, FPR95 ↓.

Adding the temporal stability term (AC/ATC) improves robustness under temporal shift, yielding higher clean and corrupted accuracy and lower late-timestep FPR95 for both WRN and ViT.

Conclusion

- Temp-SCONE improves corrupted-data accuracy and reduces FPR95 under temporal drift.
- AC/ATC regularization stabilizes confidence across timesteps.
- Vision transformers gain the largest ID/OOD improvements under temporal shift.
- Temp-SCONE enhances OWL by improving robustness and calibration under evolving distributions.

Acknowledgements

This work was supported by NSF CAREER CCF-2451457 and the Microsoft AFMR program; the views expressed are solely those of the authors.

References

1. H. Bai et al., *'Feed Two Birds with One SCONE: Exploiting Wild Data for Both Out-of-Distribution Generalization and Detection.'* Proc. ICML, 2023.
2. S. Garg and S. Balakrishnan, *'Leveraging Unlabeled Data to Predict Out-of-Distribution Performance.'* ICLR, 2022.
3. Z. Lin et al., *'The CLEAR Benchmark: Continual Learning on Real-World Imagery.'* NeurIPS Datasets & Benchmarks, 2021.
4. S. Ginosar et al., *'A Century of Portraits: A Visual Historical Record of American High School Yearbooks.'* ICCV Workshops, 2015.

Paper Link   Code Repository