Novelty Assessment Report

Paper: Specialization after Generalization: Towards Understanding Test-Time Training in Foundation Models
PDF URL: https://openreview.net/pdf?id=1c6Ao3CpKt
Venue: ICLR 2026 Conference Submission
Year: 2026
Report Generated: 2025-12-29

Abstract

Recent empirical studies have explored the idea of continuing to train a model at test-time for a given task, known as test-time training (TTT), and have found it to yield significant performance improvements. However, there is limited understanding of why and when TTT is effective. Earlier explanations mostly focused on the observation that TTT may help when applied to out-of-distribution adaptation or used with privileged data. However, the growing scale of foundation models with most test data being in-distribution questions these explanations. We instead posit that foundation models remain globally underparameterized, with TTT providing a mechanism for specialization after generalization—focusing capacity on concepts relevant to the test task. Specifically, under the linear representation hypothesis, we propose a model in which TTT achieves a substantially smaller in-distribution test error than global training. We empirically validate our model's key assumptions by training a sparse autoencoder on ImageNet, showing that semantically related data points are explained by only a few shared concepts. Finally, we perform scaling studies across image and language tasks that confirm the practical implications of our model, identifying the regimes where specialization is most effective.

Core Task Landscape

This paper addresses: **Test-Time Training for In-Distribution Prediction in Foundation Models**
A total of **50 papers** were analyzed and organized into a taxonomy with **36 categories**.

Taxonomy Overview

The research landscape has been organized into the following main categories:

- **Core Test-Time Training Mechanisms and Theory**
- **Vision-Language Foundation Model Adaptation**
- **Computer Vision Task-Specific Adaptation**
- **Medical and Healthcare Applications**
- **Time Series Foundation Models**
- **Speech and Audio Foundation Models**
- **Reinforcement Learning and Behavioral Foundation Models**
- **Continual and Online Test-Time Adaptation**
- **Cross-Domain and Multimodal Generalization**
- **Uncertainty Quantification and Reliability**
- ... and 1 more categories

Complete Taxonomy Tree

- Test-Time Training for In-Distribution Prediction in Foundation Models Survey Taxonomy
- Core Test-Time Training Mechanisms and Theory
  - Theoretical Foundations and Specialization Mechanisms ★ (3 papers)
  - [0] Specialization after Generalization: Towards Understanding Test-Time Training in Foundation Models (Anon et al., 2026) View paper
  - [28] Test time training enhances in-context learning of nonlinear functions (Suzuki, 2025) View paper
  - [33] Test-Time Training Provably Improves Transformers as In-context Learners (Ildiz, 2025) View paper
  - Forward-Only and Backpropagation-Free Optimization (2 papers)
  - [6] Test-time model adaptation with only forward passes (Niu, 2024) View paper
  - [34] E-BATS: Efficient Backpropagation-Free Test-Time Adaptation for Speech Foundation Models (Jia Hong, 2025) View paper
  - Retrieval-Augmented and Nearest-Neighbor Approaches (2 papers)
  - [14] Ts-rag: Retrieval-augmented generation based time series foundation models are stronger zero-shot forecaster (Pan Zijie, 2025) View paper
  - [49] Reproducibility Report: Test-Time Training on Nearest Neighbors for Large Language Models (Boyang Zhou, 2025) View paper
- Vision-Language Foundation Model Adaptation
  - Prompt-Based and Zero-Shot Adaptation (3 papers)
  - [5] Robust test-time adaptation for zero-shot prompt tuning (Ding-Chu Zhang, 2024) View paper
  - [15] A lost opportunity for vision-language models: a comparative study of online test-time adaptation for vision-language models (DÃ¶bler, 2024) View paper
  - [31] Words Matter: Leveraging Individual Text Embeddings for Code Generation in CLIP Test-Time Adaptation (Silva-RodrÃguez, 2024) View paper
  - Distribution Shift and Out-of-Distribution Robustness (3 papers)
  - [1] Test-time linear out-of-distribution detection (Ke Fan, 2024) View paper
  - [9] Noisy Test-Time Adaptation in Vision-Language Models (Cao, 2025) View paper

    - [41] Negation-Aware Test-Time Adaptation for Vision-Language Models (Han, 2025) View paper
    - Collaborative and Federated Test-Time Adaptation (1 papers)
    - [10] Latte: Collaborative test-time adaptation of vision-language models in federated learning (Bao, 2025) View paper
    - Bayesian and Probabilistic Adaptation Frameworks (1 papers)
    - [20] Bayesian Test-time Adaptation for Object Recognition and Detection with Vision-language Models (Zhou Lihua, 2025) View paper
- Computer Vision Task-Specific Adaptation
    - Object Detection and Recognition (1 papers)
    - [17] Test-Time Adaptive Object Detection with Foundation Model (Gao Yingjie, 2025) View paper
    - Canonicalization and Geometric Transformation (1 papers)
    - [7] Test-time canonicalization by foundation models for robust perception (Singhal, 2025) View paper
    - 3D Vision and Point Cloud Processing (2 papers)
    - [8] TTT-KD: Test-Time Training for 3D Semantic Segmentation Through Knowledge Distillation From Foundation Models (Lisa Weijler, 2024) View paper
    - [36] Adapt-As-You-Walk Through the Clouds: Training-Free Online Test-Time Adaptation of 3D Vision-Language Foundation Models (Mehran Tamjidi, 2025) View paper
    - Diffusion Model-Based Adaptation (1 papers)
    - [3] Test-time adaptation with diffusion models (M Prabhudesai, 2023) View paper
    - Monocular Depth Estimation (1 papers)
    - [23] A Simple yet Effective Test-Time Adaptation for Zero-Shot Monocular Metric Depth Estimation (Chapoutot, 2024) View paper
- Medical and Healthcare Applications
    - Medical Image Segmentation (2 papers)
    - [4] AutoMiSeg: Automatic Medical Image Segmentation via Test-Time Adaptation of Foundation Models (Li XingJian, 2025) View paper
    - [13] TTA-FM: Patient-Specific Test-Time Adaptation Using Foundation Models for Improved Prostate Segmentation in Magnetic Resonance Images (Hariharan Ravishankar, 2024) View paper
    - Surgical Workflow Understanding (1 papers)
    - [24] Recognizing Surgical Phases Anywhere: Few-Shot Test-time Adaptation and Task-graph Guided Refinement (Yuan Kun, 2025) View paper
    - Medical Image-to-Image Translation (1 papers)
    - [37] Sample-Aware Test-Time Adaptation for Medical Image-to-Image Translation (Di Feola, 2025) View paper
    - Anti-Forgetting and Robustness Under Shift (1 papers)
    - [39] Anti-forgetting Test-Time Adaptation for Robust Medical Image Analysis Under Distribution Shift (Zhenyu Wu, 2025) View paper
- Time Series Foundation Models
    - In-Context Learning and Few-Shot Forecasting (2 papers)
    - [19] In-Context Fine-Tuning for Time-Series Foundation Models (Das, 2024) View paper
    - [35] In-Context and Few-Shots Learning for Forecasting Time Series Data based on Large Language Models (Saroj Gopali, 2025) View paper
    - Covariate-Aware and Multivariate Adaptation (2 papers)
    - [18] Time Series Foundation Models for Multivariate Financial Time Series Forecasting (Marconi, 2025) View paper
    - [40] TFMAdapter: Lightweight Instance-Level Adaptation of Foundation Models for Forecasting with Covariates (Afrin Dange, 2025) View paper
    - Domain-Specific Time Series Applications (3 papers)
    - [12] In-context learning of temporal point processes with foundation inference models (Berghaus, 2025) View paper
    - [43] Adapting Large Language Models to Forecast in Frequency Domain (Yungeng Zhang, 2025) View paper
    - [44] Zero-shot forecasting of optical network telemetry using large language models (Abdelli, 2025) View paper
- Speech and Audio Foundation Models
    - Acoustic Domain Shift and Robustness (1 papers)
    - [29] Advancing Test-Time Adaptation in Wild Acoustic Test Settings (Hongfu Liu, 2023) View paper
    - Speaker-Deficiency and Dysarthric Speech Adaptation (1 papers)
    - [26] Structured Speaker-Deficiency Adaptation of Foundation Models for Dysarthric and Elderly Speech Recognition (Hu, 2024) View paper
    - Efficient Domain Adaptation for Speech (1 papers)
    - [50] Efficient Domain Adaptation for Speech Foundation Models (Bo Li, 2023) View paper
- Reinforcement Learning and Behavioral Foundation Models
    - Zero-Shot Policy Retrieval and Fast Adaptation (1 papers)
    - [2] Fast Adaptation with Behavioral Foundation Models (Sikchi, 2025) View paper
    - Dynamics Adaptation and Partial Observability (1 papers)
    - [11] Zero-Shot Adaptation of Behavioral Foundation Models to Unseen Dynamics (Zisman, 2025) View paper
    - Optimistic Task Inference Through Interaction (1 papers)
    - [22] Optimistic Task Inference for Behavior Foundation Models (Bagatella, 2025) View paper
- Continual and Online Test-Time Adaptation
    - Continual Domain Adaptation and Anti-Forgetting (1 papers)
    - [21] Fast and Accurate Continual Test Time Domain Adaptation (Haihang Wu, 2024) View paper
- Cross-Domain and Multimodal Generalization
    - Multimodal Domain Adaptation Frameworks (1 papers)
    - [25] Advances in Multimodal Adaptation and Generalization: From Traditional Approaches to Foundation Models (Dong Hao, 2025) View paper
    - Causality-Driven Distribution Shift Handling (1 papers)
    - [38] Navigating Distribution Shifts in ML through Causality and Foundation Models (Guanglin, 2024) View paper
    - Hierarchical and Structured Domain Adaptation (1 papers)
    - [30] Efficient Hierarchical Domain Adaptation for Pretrained Language Models (Alexandra Chronopoulou, 2021) View paper

- Equivariant Canonicalization for Adaptation (1 papers)
  - [46] Equivariant Adaptation of Large Pretrained Models (Mondal, 2023) View paper
- Cross-Domain Adaptation in Specialized Tasks (2 papers)
  - [47] Cross-domain Rumor Detection via Test-Time Adaptation and Large Language Models (Yuxia Gong, 2025) View paper
  - [48] FloodDAN: Unsupervised Flood Forecasting based on Adversarial Domain Adaptation (Ruizhi Zhou, 2022) View paper
- Uncertainty Quantification and Reliability
  - Uncertainty-Aware Fine-Tuning and Calibration (1 papers)
  - [16] Fine-Tuning with Uncertainty-Aware Priors Makes Vision and Language Foundation Models More Reliable (TGJ Rudner, 2024) View paper
  - Reward Model Robustness Under Shift (1 papers)
  - [27] A Baseline Analysis of Reward Models' Ability To Accurately Analyze Foundation Models Under Distribution Shift (LeVine, 2023) View paper
- Specialized Domain Foundation Models
  - Protein and Molecular Foundation Models (1 papers)
  - [32] AMix-1: A Pathway to Test-Time Scalable Protein Foundation Model (Zhou Jiang, 2025) View paper
  - Astrophysical and Stellar Foundation Models (1 papers)
  - [42] SpectraFM: Tuning into Stellar Foundation Models (Bovy, 2024) View paper
  - Neuroscience and EEG Foundation Models (1 papers)
  - [45] NeuroTTT: Bridging Pretraining-Downstream Task Misalignment in EEG Foundation Models via Test-Time Training (Wang Suli, 2025) View paper

Narrative

Core task: test-time training for in-distribution prediction in foundation models. This field explores how large pre-trained models can be adapted at inference time to improve performance on specific test instances or distributions without requiring extensive offline retraining. The taxonomy reveals a rich landscape organized around several major themes. At the foundation lie Core Test-Time Training Mechanisms and Theory, which develop the mathematical principles and algorithmic frameworks—such as self-supervised auxiliary tasks, parameter-efficient updates, and specialization strategies—that enable models to learn from test data on the fly. Branching outward, the taxonomy encompasses modality-specific adaptations: Vision-Language Foundation Model Adaptation addresses multimodal alignment challenges (e.g., Robust Prompt Tuning[5], Noisy VLM Adaptation[9]), while Computer Vision Task-Specific Adaptation and Medical and Healthcare Applications tackle domain-specific requirements in imaging and clinical settings (e.g., AutoMiSeg[4], TTA-FM Prostate[13]). Parallel branches cover Time Series Foundation Models (TS-RAG[14], Financial Time Series[18]), Speech and Audio Foundation Models (Dysarthric Speech Adaptation[26]), and Reinforcement Learning and Behavioral Foundation Models (Behavioral Foundation Adaptation[2], Zero-Shot Behavioral Adaptation[11]). Cross-cutting concerns appear in Continual and Online Test-Time Adaptation, Cross-Domain and Multimodal Generalization, Uncertainty Quantification and Reliability (Uncertainty-Aware Priors[16]), and Specialized Domain Foundation Models.

Within the theoretical core, a particularly active line of work investigates how foundation models can specialize after broad pre-training, balancing generalization with instance-specific refinement. Specialization after Generalization[0] sits squarely in this theoretical branch alongside TTT Nonlinear Functions[28] and TTT Transformers ICL[33], which explore architectural mechanisms for test-time learning in transformer-based systems. These works contrast with more application-driven approaches: while Diffusion Test Adaptation[3] and Forward Pass Adaptation[6] emphasize lightweight, single-pass adjustments for efficiency, the theoretical studies examine deeper questions about what representations and learning rules enable effective specialization without catastrophic forgetting. The original paper's emphasis on theoretical foundations and specialization mechanisms positions it as a conceptual anchor for understanding when and why test-time training succeeds, complementing empirical investigations across the taxonomy's diverse application domains.

Related Works in Same Category

The following **2 sibling papers** share the same taxonomy leaf node with the original paper:

1. Test time training enhances in-context learning of nonlinear functions

Authors: Suzuki, Taiji, Kento Kuwataka, Taiji Suzuki | **Year/Venue**: 2025 | **URL**: View paper

Abstract

Test-time training (TTT) enhances model performance by explicitly updating designated parameters prior to each prediction to adapt to the test data. While TTT has demonstrated considerable empirical success, its theoretical underpinnings remain limited, particularly for nonlinear models. In this paper, we investigate the combination of TTT with in-context learning (ICL), where the model is given a few examples from the target distribution at inference time. We analyze this framework in the setti...

Relationship Analysis

Both papers belong to the theoretical foundations category, establishing models explaining when and why test-time training improves in-distribution prediction. They share overlapping focus on theoretical analysis of TTT mechanisms, with both examining specialization and adaptation at test-time. The key difference is that the original paper proposes a specialization-after-generalization framework under the linear representation hypothesis for foundation models broadly, while the candidate paper specifically analyzes TTT combined with in-context learning for single-index models in transformers, focusing on nonlinear function learning with explicit convergence rates.

---

2. Test-Time Training Provably Improves Transformers as In-context Learners

Authors: Ildiz, M. Emrullah, Halil Alperen Gozeten, Zhang Xuechen, M. E. Ildiz, et al. (15 authors total) | **Year/Venue**: 2025 | **URL**: View paper

Abstract

Test-time training (TTT) methods explicitly update the weights of a model to adapt to the specific test instance, and they have found success in a variety of settings, including most recently language modeling and reasoning. To demystify this success, we investigate a gradient-based TTT algorithm for in-context learning, where we train a transformer model on the in-context demonstrations provided in the test prompt. Specifically, we provide a comprehensive theoretical characterization of linear ...

Relationship Analysis

Both papers belong to the Theoretical Foundations and Specialization Mechanisms category, establishing theoretical models for test-time training in foundation models. They share overlapping focus on explaining why TTT improves in-distribution prediction through specialization mechanisms, with both analyzing how TTT adapts models to local task-specific concepts. The key difference is that the original paper focuses on the "specialization after generalization" framework under the linear representation hypothesis with sparse concept spaces and superposition, while the candidate paper specifically analyzes transformers as in-context learners, characterizing TTT through gradient-based updates on linear attention models with explicit sample complexity bounds.

Contributions Analysis

Overall novelty summary. The paper proposes a theoretical framework explaining test-time training (TTT) as 'specialization after generalization' under the linear representation hypothesis, arguing that foundation models remain globally underparameterized and benefit from focusing capacity on test-relevant concepts. It resides in the 'Theoretical Foundations and Specialization Mechanisms' leaf, which contains only three papers total, including this one. This represents a sparse research direction within the broader taxonomy of 50 papers across 36 topics, suggesting the theoretical underpinnings of in-distribution TTT remain relatively underexplored compared to application-driven work.

The taxonomy reveals that most TTT research concentrates on modality-specific adaptations (vision-language, medical imaging, time series, speech) and practical mechanisms (prompt tuning, retrieval-augmented methods, diffusion-based adaptation). The paper's theoretical leaf sits within 'Core Test-Time Training Mechanisms and Theory,' which also includes forward-only optimization and retrieval-augmented approaches—these neighboring leaves focus on algorithmic efficiency rather than foundational explanations. The sibling papers (TTT Nonlinear Functions, TTT Transformers ICL) examine architectural mechanisms for test-time learning, whereas this work addresses the more fundamental question of why TTT improves in-distribution prediction through a specialization lens.

Among 15 candidates examined across three contributions, no refutable prior work was identified. The 'specialization after generalization' framework examined 10 candidates with zero refutations; the linear representation hypothesis model examined 1 candidate; and the empirical validation through sparse autoencoders examined 4 candidates. These statistics reflect a limited search scope (top-K semantic search plus citation expansion), not an exhaustive literature review. The absence of refutable candidates suggests that among the examined papers, none directly anticipated the theoretical model linking global underparameterization to in-distribution TTT benefits, though the small sample size limits strong conclusions.

Based on the limited search scope of 15 candidates, the theoretical framing appears novel within the examined literature, particularly the claim that TTT addresses underparameterization rather than distribution shift. However, the sparse population of the theoretical leaf (3 papers) and the modest search scale mean substantial related work may exist outside the top-K semantic matches. The analysis covers conceptual positioning but cannot definitively assess novelty against the full field.

This paper presents **3 main contributions**, each analyzed against relevant prior work:

Contribution 1: Specialization after generalization framework for test-time training

Description: The authors propose a conceptual framework where test-time training enables foundation models to specialize by temporarily reallocating capacity to concepts relevant to the immediate test task, rather than requiring out-of-distribution data or privileged information. This mechanism addresses global underparameterization by locally adapting the model.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Rewards-in-context: Multi-objective alignment of foundation models with dynamic preference adjustment

URL: View paper

Brief Assessment

Rewards-in-Context[54] addresses multi-objective alignment of foundation models through supervised fine-tuning with reward conditioning, not test-time training for specialization after generalization as described in the original paper.

---

2. Transformers as Statisticians: Provable In-Context Learning with In-Context Algorithm Selection

URL: View paper

Brief Assessment

Transformers as Statisticians[57] focuses on in-context learning mechanisms where transformers learn to select and apply different algorithms during inference on new task instances. This differs fundamentally from the original paper's test-time training framework, which involves temporarily fine-tuning foundation models on local neighborhoods to reallocate capacity for immediate test tasks.

---

3. Test-Time Learning for Large Language Models

URL: View paper

Brief Assessment

Test-Time LLM Learning[53] focuses on test-time adaptation for language models using input perplexity minimization on unlabeled test data, while the original paper proposes a broader conceptual framework of 'specialization after generalization' that applies to foundation models generally (vision and language) through temporarily reallocating capacity to relevant concepts. The candidate's approach is specific to LLMs and uses a different mechanism (perplexity minimization) rather than the original's concept-based capacity reallocation under the linear representation hypothesis.

---

4. FedDG-MoE: Test-Time Mixture-of-Experts Fusion for Federated Domain Generalization

URL: View paper

Brief Assessment

FedDG-MoE[55] focuses on federated domain generalization using mixture-of-experts adapters with test-time fusion based on feature similarity. This differs from the original paper's framework of test-time training for foundation models that temporarily reallocates capacity to relevant concepts through local adaptation on in-distribution data.

---

5. Bayesian test-time adaptation for vision-language models

URL: View paper

Brief Assessment

Bayesian VLM Test[58] focuses on test-time adaptation for vision-language models (CLIP) using Bayesian methods to adapt likelihood and prior distributions. The original paper addresses test-time training for foundation models more broadly, proposing a conceptual framework where models specialize by reallocating capacity to relevant concepts. These are distinct technical approaches to different problem settings.

---

6. Time Series Foundation Models for Multivariate Financial Time Series Forecasting

URL: View paper

Brief Assessment

Financial Time Series[18] focuses on time series foundation models for financial forecasting tasks (treasury yields, volatility, equity spreads) through pretraining and fine-tuning. It does not address test-time training mechanisms, specialization after generalization frameworks, or the linear representation hypothesis that are central to the original paper's contribution.

---

7. Dual-personalizing adapter for federated foundation models

URL: View paper

Brief Assessment

Dual-personalizing Adapter[60] focuses on federated learning with test-time personalization for handling distribution shifts across clients, not on the specialization-after-generalization mechanism for foundation models proposed in the original paper.

---

8. Contrastive adapters for foundation model group robustness

URL: View paper

Brief Assessment

Contrastive Adapters[59] focuses on improving group robustness in foundation models through adapter training with contrastive learning, not on test-time training for specialization. The candidate addresses a different problem (subpopulation shifts) with a different mechanism (contrastive adapters trained once, not per-test-sample).

---

9. Dynamic adaptation of lora fine-tuning for efficient and task-specific optimization of large language models

URL: View paper

Brief Assessment

Dynamic LoRA Adaptation[52] focuses on adaptive fine-tuning mechanisms for task-specific optimization of LLMs using dynamic weight allocation and input feature-based strategies. It does not address test-time training for specialization after generalization in foundation models, nor does it discuss the conceptual framework of temporarily reallocating capacity to relevant concepts during test time.

---

10. A foundation model for generalized brain MRI analysis

URL: View paper

Brief Assessment

Brain MRI Foundation[56] focuses on self-supervised pretraining of foundation models for brain MRI analysis with downstream task adaptation, not on test-time training mechanisms or the specialization after generalization framework proposed in the original paper.

Contribution 2: Theoretical model under the linear representation hypothesis

Description: The authors develop a theoretical model based on the linear representation hypothesis where test-time training can achieve lower in-distribution test error than globally trained models. The model formalizes how TTT efficiently recovers the local meaning of superimposed concepts in underparameterized feature spaces.

This contribution was assessed against **1 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Test-time adaptation induces stronger accuracy and agreement-on-the-line

URL: View paper

Brief Assessment

Agreement-on-the-Line[51] focuses on test-time adaptation strengthening linear correlations between ID and OOD accuracy/agreement, not on developing theoretical models for TTT under the linear representation hypothesis to reduce in-distribution test error.

Contribution 3: Empirical validation through sparse autoencoders and scaling studies

Description: The authors train sparse autoencoders on ImageNet to validate that local neighborhoods are supported by few concepts and that TTT implicitly finds sparse solutions. They conduct scaling studies across vision and language tasks demonstrating that TTT provides the largest performance gains in the underparameterized regime.

This contribution was assessed against **4 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. SAUCE: Selective Concept Unlearning in Vision-Language Models with Sparse Autoencoders

URL: View paper

Brief Assessment

SAUCE[63] focuses on concept unlearning in vision-language models using sparse autoencoders, not on validating test-time training through scaling studies on vision-language tasks as described in the original contribution.

---

2. Scaling Vision with Sparse Mixture of Experts

URL: View paper

Brief Assessment

Sparse Mixture Experts[64] focuses on sparse mixture-of-experts architectures for vision transformers with routing mechanisms, not on validating test-time training through sparse autoencoders or analyzing TTT's effectiveness across underparameterized regimes in vision-language tasks.

---

3. Training Vision-Language Process Reward Models for Test-Time Scaling in Multimodal Reasoning: Key Insights and Lessons Learned

URL: View paper

Brief Assessment

Vision-Language Process Rewards[61] focuses on process reward models for vision-language reasoning with test-time scaling via MCTS and outcome reward models, not on sparse autoencoders validating test-time training through concept sparsity analysis on vision-language tasks.

---

4. Efficient Test-Time Scaling for Small Vision-Language Models

URL: View paper

Brief Assessment

Small VLM Scaling[62] focuses on test-time scaling for vision-language models through augmentation and adaptation techniques, not on validating test-time training through sparse autoencoders or analyzing concept sparsity in local neighborhoods as described in the original contribution.

Appendix: Text Similarity Detection

No high-similarity text segments were detected across any compared papers.

References

- [0] Specialization after Generalization: Towards Understanding Test-Time Training in Foundation Models View paper
- [1] Test-time linear out-of-distribution detection View paper
- [2] Fast Adaptation with Behavioral Foundation Models View paper
- [3] Test-time adaptation with diffusion models View paper
- [4] AutoMiSeg: Automatic Medical Image Segmentation via Test-Time Adaptation of Foundation Models View paper
- [5] Robust test-time adaptation for zero-shot prompt tuning View paper
- [6] Test-time model adaptation with only forward passes View paper
- [7] Test-time canonicalization by foundation models for robust perception View paper
- [8] TTT-KD: Test-Time Training for 3D Semantic Segmentation Through Knowledge Distillation From Foundation Models View paper
- [9] Noisy Test-Time Adaptation in Vision-Language Models View paper
- [10] Latte: Collaborative test-time adaptation of vision-language models in federated learning View paper
- [11] Zero-Shot Adaptation of Behavioral Foundation Models to Unseen Dynamics View paper
- [12] In-context learning of temporal point processes with foundation inference models View paper
- [13] TTA-FM: Patient-Specific Test-Time Adaptation Using Foundation Models for Improved Prostate Segmentation in Magnetic Resonance Images View paper
- [14] Ts-rag: Retrieval-augmented generation based time series foundation models are stronger zero-shot forecaster View paper
- [15] A lost opportunity for vision-language models: a comparative study of online test-time adaptation for vision-language models View paper
- [16] Fine-Tuning with Uncertainty-Aware Priors Makes Vision and Language Foundation Models More Reliable View paper
- [17] Test-Time Adaptive Object Detection with Foundation Model View paper
- [18] Time Series Foundation Models for Multivariate Financial Time Series Forecasting View paper
- [19] In-Context Fine-Tuning for Time-Series Foundation Models View paper
- [20] Bayesian Test-time Adaptation for Object Recognition and Detection with Vision-language Models View paper
- [21] Fast and Accurate Continual Test Time Domain Adaptation View paper
- [22] Optimistic Task Inference for Behavior Foundation Models View paper
- [23] A Simple yet Effective Test-Time Adaptation for Zero-Shot Monocular Metric Depth Estimation View paper
- [24] Recognizing Surgical Phases Anywhere: Few-Shot Test-time Adaptation and Task-graph Guided Refinement View paper
- [25] Advances in Multimodal Adaptation and Generalization: From Traditional Approaches to Foundation Models View paper
- [26] Structured Speaker-Deficiency Adaptation of Foundation Models for Dysarthric and Elderly Speech Recognition View paper
- [27] A Baseline Analysis of Reward Models' Ability To Accurately Analyze Foundation Models Under Distribution Shift View paper
- [28] Test time training enhances in-context learning of nonlinear functions View paper
- [29] Advancing Test-Time Adaptation in Wild Acoustic Test Settings View paper
- [30] Efficient Hierarchical Domain Adaptation for Pretrained Language Models View paper
- [31] Words Matter: Leveraging Individual Text Embeddings for Code Generation in CLIP Test-Time Adaptation View paper
- [32] AMix-1: A Pathway to Test-Time Scalable Protein Foundation Model View paper
- [33] Test-Time Training Provably Improves Transformers as In-context Learners View paper
- [34] E-BATS: Efficient Backpropagation-Free Test-Time Adaptation for Speech Foundation Models View paper
- [35] In-Context and Few-Shots Learning for Forecasting Time Series Data based on Large Language Models View paper
- [36] Adapt-As-You-Walk Through the Clouds: Training-Free Online Test-Time Adaptation of 3D Vision-Language Foundation Models View paper
- [37] Sample-Aware Test-Time Adaptation for Medical Image-to-Image Translation View paper
- [38] Navigating Distribution Shifts in ML through Causality and Foundation Models View paper
- [39] Anti-forgetting Test-Time Adaptation for Robust Medical Image Analysis Under Distribution Shift View paper
- [40] TFMAdapter: Lightweight Instance-Level Adaptation of Foundation Models for Forecasting with Covariates View paper
- [41] Negation-Aware Test-Time Adaptation for Vision-Language Models View paper
- [42] SpectraFM: Tuning into Stellar Foundation Models View paper
- [43] Adapting Large Language Models to Forecast in Frequency Domain View paper
- [44] Zero-shot forecasting of optical network telemetry using large language models View paper
- [45] NeuroTTT: Bridging Pretraining-Downstream Task Misalignment in EEG Foundation Models via Test-Time Training View paper
- [46] Equivariant Adaptation of Large Pretrained Models View paper
- [47] Cross-domain Rumor Detection via Test-Time Adaptation and Large Language Models View paper
- [48] FloodDAN: Unsupervised Flood Forecasting based on Adversarial Domain Adaptation View paper
- [49] Reproducibility Report: Test-Time Training on Nearest Neighbors for Large Language Models View paper
- [50] Efficient Domain Adaptation for Speech Foundation Models View paper
- [51] Test-time adaptation induces stronger accuracy and agreement-on-the-line View paper
- [52] Dynamic adaptation of lora fine-tuning for efficient and task-specific optimization of large language models View paper
- [53] Test-Time Learning for Large Language Models View paper
- [54] Rewards-in-context: Multi-objective alignment of foundation models with dynamic preference adjustment View paper
- [55] FedDG-MoE: Test-Time Mixture-of-Experts Fusion for Federated Domain Generalization View paper
- [56] A foundation model for generalized brain MRI analysis View paper
- [57] Transformers as Statisticians: Provable In-Context Learning with In-Context Algorithm Selection View paper
- [58] Bayesian test-time adaptation for vision-language models View paper
- [59] Contrastive adapters for foundation model group robustness View paper
- [60] Dual-personalizing adapter for federated foundation models View paper
- [61] Training Vision-Language Process Reward Models for Test-Time Scaling in Multimodal Reasoning: Key Insights and Lessons Learned View paper
- [62] Efficient Test-Time Scaling for Small Vision-Language Models View paper
- [63] SAUCE: Selective Concept Unlearning in Vision-Language Models with Sparse Autoencoders View paper
- [64] Scaling Vision with Sparse Mixture of Experts View paper