Novelty Assessment Report

Paper: Understanding the Learning Phases in Self-Supervised Learning via Critical Periods
PDF URL: https://openreview.net/pdf?id=UxIRc97ecL
Venue: ICLR 2026 Conference Submission
Year: 2026
Report Generated: 2025-12-29

Abstract

Self-supervised learning (SSL) has emerged as a powerful pretraining strategy to learn transferable representations from unlabeled data. Yet, it remains unclear how long SSL models should be pretrained for such representations to emerge. Contrary to the prevailing heuristic that longer pretraining translates to better downstream performance, we identify a transferability trade-off: across diverse SSL settings, intermediate checkpoints often yield stronger out-of-domain (OOD) generalization, whereas additional pretraining primarily benefits in-domain (ID) accuracy. From this observation, we hypothesize that SSL progresses through learning phases that can be characterized through the lens of critical periods (CP). Prior work on CP has shown that supervised learning models exhibit early phases of high plasticity, followed by a consolidation phase where adaptability declines but task-specific performance keeps increasing. Since traditional CP analysis depends on supervised labels, for SSL we rethink CP in two ways. First, we inject deficits to perturb the pretraining data and measure the quality of learned representations via downstream tasks. Second, to estimate network plasticity during pretraining we compute the Fisher Information matrix on pretext objectives, quantifying the sensitivity of model parameters to the supervisory signal defined by the pretext tasks. We conduct several experiments to demonstrate that SSL models do exhibit their own CP, with CP closure marking a sweet spot where representations are neither underdeveloped nor overfitted to the pretext task. Leveraging these insights, we propose CP-guided checkpoint selection as a mechanism for identifying intermediate checkpoints during SSL that improve OOD transferability. Finally, to balance the transferability trade-off, we propose CP-guided self-distillation, which selectively distills layer representations from the sweet spot (CP closure) checkpoint into their overspecialized counterparts in the final pretrained model.

Core Task Landscape

This paper addresses: **Learning Phases and Transferability in Self-Supervised Pretraining**

A total of **50 papers** were analyzed and organized into a taxonomy with **19 categories**.

Taxonomy Overview

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

- **Theoretical Foundations and Mechanisms**
- **Pretraining Methodologies and Architectures**
- **Transfer Learning Strategies and Adaptation**
- **Application Domains and Empirical Studies**

Complete Taxonomy Tree

- Learning Phases and Transferability in Self-Supervised Pretraining Survey Taxonomy
- Theoretical Foundations and Mechanisms
  - Learning Phase Characterization ★ (1 papers)
  - [0] Understanding the Learning Phases in Self-Supervised Learning via Critical Periods (Anon et al., 2026) View paper
  - Transferability Analysis and Measurement (3 papers)
  - [11] A Broad Study on the Transferability of Visual Representations with Contrastive Learning (Ashraful Islam, 2021) View paper
  - [38] GSTBench: A Benchmark Study on the Transferability of Graph Self-Supervised Learning (Yu Song, 2025) View paper
  - [40] Exploring Model Transferability through the Lens of Potential Energy (Xiaotong Li, 2023) View paper
  - Representation Learning Principles (2 papers)
  - [2] Self-supervised learning: A succinct review (V. Rani, 2023) View paper
  - [45] Self-Supervised Representation Learning: Introduction, advances, and challenges (Linus Ericsson, 2021) View paper
- Pretraining Methodologies and Architectures
  - Contrastive Learning Approaches (4 papers)
  - [3] Big self-supervised models are strong semi-supervised learners (Chen Ting, 2020) View paper
  - [19] Boost supervised pretraining for visual transfer learning: Implications of self-supervised contrastive representation learning (Jinghan Sun, 2022) View paper
  - [29] MoCo Pretraining Improves Representation and Transferability of Chest X-ray Models (Hari Sowrirajan, 2020) View paper
  - [35] Look, Listen, and Attend: Co-Attention Network for Self-Supervised Audio-Visual Representation Learning (Ying Cheng, 2020) View paper
  - Generative and Predictive Methods (3 papers)
  - [6] Maefe: Masked autoencoders family of electrocardiogram for self-supervised pretraining and transfer learning (Huaicheng Zhang, 2022) View paper
  - [26] DDAE: Towards Deep Dynamic Vision BERT Pretraining (Honghao Chen, 2024) View paper
  - [32] Self-supervised pretraining of transformers for satellite image time series classification (Yuan Yuan, 2020) View paper
  - Model Architecture Innovations (2 papers)

  - [4] The lottery tickets hypothesis for supervised and self-supervised pre-training in computer vision models (Tianlong Chen, 2021) View paper
  - [16] SFT: Few-Shot Learning via Self-Supervised Feature Fusion With Transformer (Jit Yan Lim, 2024) View paper
  - Multimodal and Cross-Modal Pretraining (3 papers)
  - [9] Grounded Language-Image Pre-training (Li, 2021) View paper
  - [13] Multi-modal representation learning in retinal imaging using self-supervised learning for enhanced clinical predictions (Emese SÃ¼kei, 2024) View paper
  - [39] A Multimodal Protein Representation Framework for Quantifying Transferability Across Biochemical Downstream Tasks (F. Hu, 2023) View paper
- Transfer Learning Strategies and Adaptation
  - Finetuning and Adaptation Techniques (3 papers)
  - [5] Improved transferability of self-supervised learning models through batch normalization finetuning (Kirill Sirotkin, 2024) View paper
  - [47] DRAFT: A novel framework to reduce domain shifting in self-supervised learning and its application to children's ASR (Fan, 2022) View paper
  - [50] Recognizing surgical phases anywhere: Few-Shot Test-Time adaptation and Task-Graph guided refinement (Yuan Kun, 2025) View paper
  - Knowledge Transfer and Distillation (2 papers)
  - [41] Class Incremental Learning with Self-Supervised Pre-Training and Prototype Learning (Liu Wenzhuo, 2023) View paper
  - [46] Boosting self-supervised learning via knowledge transfer (Noroozi, 2018) View paper
  - Domain Shift and Robustness (2 papers)
  - [36] PEAMATL: A strategy for developing near-infrared spectral prediction models under domain shift using self-supervised transfer learning (Yu Yang, 2023) View paper
  - [44] When Does Contrastive Learning Preserve Adversarial Robustness from Pretraining to Finetuning? (Fan Lijie, 2021) View paper
  - Few-Shot and Low-Resource Transfer (3 papers)
  - [21] FDSP-HRID: Few-Shot Detector With Self-Supervised Pretraining for High-Speed Rail Infrastructure Defects (Zhaorui Hong, 2024) View paper
  - [25] On-Device Constrained Self-Supervised Learning for Keyword Spotting via Quantization Aware Pre-Training and Fine-Tuning (Gene-Ping Yang, 2024) View paper
  - [27] â¦ by Supervised or Selfâ Supervised Learning for Chest Radiograph Classification: A Comparative Study Against ImageNet Counterparts in Coldâ Start Active Learning (H Yuan, 2025) View paper
- Application Domains and Empirical Studies
  - Medical and Biomedical Imaging (4 papers)
  - [10] Self-supervised learning for skin cancer diagnosis with limited training data (Chandra, 2024) View paper
  - [18] Improving the classification of veterinary thoracic radiographs through inter-species and inter-pathology self-supervised pre-training of deep learning models (Weronika Celniak, 2023) View paper
  - [24] Spatial-Temporal Pre-Training for Embryo Viability Prediction Using Time-Lapse Videos (Shi Zhiyi, 2025) View paper
  - [42] BarlowTwins-CXR: enhancing chest X-ray abnormality localization in heterogeneous data with cross-domain self-supervised learning (Haoyue Sheng, 2024) View paper
  - Biological and Molecular Data Analysis (2 papers)
  - [1] Reusability report: Exploring the transferability of self-supervised learning models from single-cell to spatial transcriptomics (Chuangyi Han, 2025) View paper
  - [49] MPCD: A Multitask Graph Transformer for Molecular Property Prediction by Integrating Common and Domain Knowledge. (Xixi Yang, 2024) View paper
  - Remote Sensing and Earth Observation (3 papers)
  - [23] Consecutive pre-training: A knowledge transfer learning strategy with relevant unlabeled data for remote sensing domain (Zhang, 2022) View paper
  - [31] A Self-Supervised Learning Pretraining Framework for Remote Sensing Image Change Detection (Ling Wan, 2025) View paper
  - [48] S3L: Spectrum Transformer for Self-Supervised Learning in Hyperspectral Image Classification (Hufeng Guo, 2024) View paper
  - Computer Vision Applications (3 papers)
  - [14] Learning Transferable Representations for Image Anomaly Localization Using Dense Pretraining (Haitian He, 2024) View paper
  - [30] Self-supervised Pre-training for Mirror Detection (Jiaying Lin, 2023) View paper
  - [37] Visual Reinforcement Learning With Self-Supervised 3D Representations (Yanjie Ze, 2022) View paper
  - Speech and Audio Processing (2 papers)
  - [8] Leveraging in-the-wild Data for Effective Self-supervised Pretraining in Speaker Recognition (Shuai Wang, 2023) View paper
  - [34] Self-Supervised Convolutional Audio Models are Flexible Acoustic Feature Learners: A Domain Specificity and Transfer-Learning Study (Ogg, 2025) View paper
  - Natural Language Processing (1 papers)
  - [22] Open-Domain Response Generation in Low-Resource Settings using Self-Supervised Pre-Training of Warm-Started Transformers (Tarek Naous, 2023) View paper
  - Graph and Structured Data (2 papers)
  - [7] GraphCLIP: Enhancing Transferability in Graph Foundation Models for Text-Attributed Graphs (Yun Zhu, 2024) View paper
  - [20] Redundancy is Not What You Need: An Embedding Fusion Graph Auto-Encoder for Self-Supervised Graph Representation Learning (Mengran Li, 2024) View paper
  - Specialized Domain Applications (6 papers)
  - [12] Cost-effective image recognition of water leakage in metro tunnels using self-supervised learning (Yining Gu, 2024) View paper
  - [15] Transfer learning application of self-supervised learning in ARPES (S. Ekahana, 2023) View paper
  - [17] Improving Seismic Fault Recognition with Self-Supervised Pre-Training: A Study of 3D Transformer-Based with Multi-Scale Decoding and Fusion (Zeren Zhang, 2024) View paper
  - [28] Advancing animal behavior recognition with self-supervised pre-training on unlabeled data (Axiu Mao, 2025) View paper
  - [33] A Universal Feature Extractor Based on Self-Supervised Pre-Training for Fault Diagnosis of Rotating Machinery under Limited Data (Zitong Yan, 2023) View paper

◦ [43] Self-supervised transfer learning of physiological representations from free-living wearable data (Spathis, 2021) View paper

Narrative

Core task: understanding learning phases and transferability in self-supervised pretraining. The field has organized itself around four major branches. Theoretical Foundations and Mechanisms investigates the underlying principles governing how self-supervised models learn and generalize, including phase transitions and critical periods during training. Pretraining Methodologies and Architectures encompasses the diverse algorithmic strategies—contrastive methods like MoCo Chest Xray[29], masked modeling approaches, and hybrid techniques—that enable models to extract useful representations from unlabeled data. Transfer Learning Strategies and Adaptation focuses on how pretrained representations are fine-tuned or adapted to downstream tasks, exploring questions of domain shift, few-shot learning as in Surgical Phases Few-Shot[50], and parameter-efficient adaptation methods such as those studied in BatchNorm Finetuning Transfer[5]. Application Domains and Empirical Studies documents the breadth of real-world deployments, from medical imaging in SSL Skin Cancer[10] and Retinal Multimodal SSL[13] to remote sensing in Consecutive Pretraining Remote Sensing[23] and specialized domains like Seismic Fault Transformer[17].

A particularly active line of work examines the temporal dynamics of pretraining: when and how representations become useful, and whether certain learning windows are more critical than others. Critical Periods SSL[0] sits squarely within this theoretical inquiry, characterizing distinct phases during self-supervised training and their impact on downstream transferability. This contrasts with more application-driven studies that take pretrained models as given and focus on adaptation strategies, such as BatchNorm Finetuning Transfer[5], which explores efficient fine-tuning by selectively updating normalization layers. Meanwhile, works like Big SSL Semi-Supervised[3] bridge pretraining and semi-supervised learning, highlighting trade-offs between label efficiency and representation quality. By situating learning phase characterization within the broader theoretical branch, Critical Periods SSL[0] complements empirical transfer studies and offers mechanistic insights into why certain pretraining regimes yield more robust or adaptable features.

Related Works in Same Category

No sibling papers were found in the same taxonomy leaf. A taxonomy-subtopic-level comparison will be produced instead.

Taxonomy-Level Summary

The original leaf focuses on the temporal dynamics and developmental stages during pretraining itself, examining how learning unfolds over time. The sibling subtopics address complementary aspects: one examines the fundamental mechanisms underlying what is learned (representation principles), while the other quantifies how well learned representations transfer across contexts. Together, these form a progression from understanding learning dynamics to characterizing learned features to measuring their utility.

Similarities: - All three subtopics analyze self-supervised pretraining from analytical rather than purely empirical perspectives - Each involves understanding properties of learned representations beyond final task performance - All three exclude purely methodological contributions focused on improving pretraining techniques - Each subtopic requires theoretical or analytical frameworks to characterize aspects of self-supervised learning

Differences: - Learning Phase Characterization focuses on temporal/developmental aspects (when and how learning occurs), while Representation Learning Principles focuses on mechanistic aspects (what enables effective learning), and Transferability Analysis focuses on outcome measurement (how well learning generalizes) - The original leaf examines dynamics during training, Representation Learning Principles examines fundamental properties of the learning process itself, while Transferability Analysis examines post-training characteristics - Learning Phase Characterization emphasizes critical periods and stages, Representation Learning Principles emphasizes feature quality and generalization mechanisms, while Transferability Analysis emphasizes quantification and cross-domain measurement - The original leaf's temporal focus is process-oriented, while Transferability Analysis is outcome-oriented, and Representation Learning Principles bridges both by examining underlying mechanisms

Suggested Search Directions: - Papers examining phase transitions or critical learning periods during self-supervised pretraining - Studies on curriculum effects or training dynamics specific to self-supervised objectives - Research on early stopping criteria based on learning phase identification - Work analyzing how different pretraining stages contribute to downstream transferability

Sibling Subtopics

- **Representation Learning Principles** (leaves: 1, papers: 2)
- Scope: Research on fundamental mechanisms of feature learning, representation quality, and generalization in self-supervised frameworks.
- Exclude: Excludes architecture-specific implementations; see Pretraining Methodologies or Model Architecture Innovations.
- **Transferability Analysis and Measurement** (leaves: 1, papers: 3)
- Scope: Studies quantifying and characterizing transfer capability across domains, tasks, or modalities using analytical frameworks.
- Exclude: Excludes empirical transfer applications without theoretical analysis; see Empirical Transfer Studies.

Contributions Analysis

Overall novelty summary. The paper investigates temporal dynamics of self-supervised learning pretraining, identifying a transferability trade-off where intermediate checkpoints yield stronger out-of-domain generalization while extended pretraining benefits in-domain accuracy. It resides in the 'Learning Phase Characterization' leaf under 'Theoretical Foundations and Mechanisms', which contains only this single paper among 50 total papers across 19 leaf nodes. This placement indicates a relatively sparse research direction focused specifically on temporal phase analysis during SSL pretraining, distinguishing it from the more populated methodological and application-oriented branches.

The taxonomy reveals neighboring theoretical work in 'Transferability Analysis and Measurement' (3 papers) and 'Representation Learning Principles' (2 papers), which examine transfer capability and feature learning mechanisms but without explicit temporal phase characterization. The broader 'Pretraining Methodologies' branch contains 13 papers across contrastive, generative, and architectural innovations, while 'Transfer Learning Strategies' encompasses 11 papers on adaptation techniques. The paper's focus on learning phases during pretraining positions it at the intersection of theoretical analysis and practical transfer concerns, bridging mechanistic understanding with downstream performance implications.

Among 27 candidates examined across three contributions, no clearly refuting prior work was identified. The transferability trade-off analysis examined 10 candidates with 0 refutations, the critical period reformulation for SSL examined 7 candidates with 0 refutations, and the checkpoint selection intervention examined 10 candidates with 0 refutations. This limited search scope suggests that within the top semantic matches and citation expansions, no prior work explicitly documents the same temporal trade-off phenomenon or applies critical period analysis to self-supervised settings, though the search does not claim exhaustive coverage of all potentially relevant literature.

Based on examination of 27 semantically related candidates, the work appears to occupy a distinct position within SSL research by explicitly characterizing learning phases and their differential impact on in-domain versus out-of-domain transfer. The sparse population of its taxonomy leaf and absence of refuting candidates among examined papers suggest novelty in this specific analytical framing, though the limited search scope means potentially relevant work outside the top-K semantic neighborhood may exist but was not captured in this analysis.

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

Contribution 1: Identification of transferability trade-off in SSL pretraining

Description: The authors demonstrate that extended SSL pretraining creates a trade-off where intermediate checkpoints achieve better out-of-domain generalization, whereas longer pretraining primarily benefits in-domain accuracy. This challenges the prevailing heuristic that longer pretraining always improves downstream performance.

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. Cross-domain pre-training with language models for transferable time series representations

URL: View paper

Brief Assessment

Language Models Time Series[56] focuses on cross-domain pre-training for time series data using language models, not on the transferability trade-off between in-domain and out-of-domain performance during SSL pretraining phases as studied in the original paper.

---

2. Health Assessment of Rotating Equipment With Unseen Conditions Using Adversarial Domain Generalization Toward Self-Supervised Regularization Learning

URL: View paper

Brief Assessment

Rotating Equipment Adversarial[59] focuses on domain generalization for rotating equipment health assessment under unseen operating conditions, not on analyzing SSL pretraining dynamics or transferability trade-offs across pretraining duration.

---

3. How well do self-supervised models transfer?

URL: View paper

Brief Assessment

SSL Transfer Evaluation[57] focuses on comparing different SSL methods' transfer performance across tasks, not on analyzing how pretraining duration affects the trade-off between in-domain and out-of-domain performance. The candidate does not examine intermediate checkpoints or temporal dynamics during pretraining.

---

4. Improving generalization for ai-synthesized voice detection

URL: View paper

Brief Assessment

Synthesized Voice Detection[51] focuses on AI-synthesized voice detection using disentanglement learning and domain-agnostic features, not on self-supervised learning pretraining dynamics or transferability trade-offs between in-domain and out-of-domain performance in SSL models.

---

5. Disentangled graph self-supervised learning for out-of-distribution generalization

URL: View paper

Brief Assessment

Disentangled Graph OOD[54] focuses on graph-structured data and out-of-distribution generalization through disentangled representations, not on the temporal dynamics of SSL pretraining duration or the trade-off between intermediate vs. extended pretraining checkpoints in vision models.

---

6. GraphCLIP: Enhancing Transferability in Graph Foundation Models for Text-Attributed Graphs

URL: View paper

Brief Assessment

GraphCLIP[7] focuses on graph foundation models for text-attributed graphs, not general self-supervised learning (SSL) pretraining dynamics. The candidate addresses cross-domain transferability in graph neural networks through contrastive graph-summary pretraining, which is a different domain and methodology from the original paper's analysis of SSL pretraining phases across vision models.

---

7. Self-supervised learning for generalizable out-of-distribution detection

URL: View paper

Brief Assessment

SSL OOD Detection[52] focuses on out-of-distribution detection using self-supervised learning for image classification, not on analyzing the temporal dynamics of SSL pretraining or the trade-off between in-domain and out-of-domain performance across different pretraining durations.

---

8. Robust and data-efficient generalization of self-supervised machine learning for diagnostic imaging

URL: View paper

Brief Assessment

Diagnostic Imaging Generalization[53] focuses on medical imaging applications with limited discussion of SSL pretraining dynamics. The provided context fragments are insufficient to assess claims about transferability trade-offs across pretraining duration.

---

9. SelfReg: Self-supervised Contrastive Regularization for Domain Generalization

URL: View paper

Brief Assessment

SelfReg Domain Generalization[55] focuses on domain generalization through contrastive regularization across multiple source domains, not on analyzing SSL pretraining duration or the transferability trade-off between intermediate and final checkpoints during self-supervised pretraining.

---

10. Decoding Musical Neural Activity in Patients With Disorders of Consciousness Through Self-Supervised Contrastive Domain Generalization

URL: View paper

Brief Assessment

Musical Neural Consciousness[58] focuses on self-supervised learning for EEG classification in medical diagnosis (disorders of consciousness), not on analyzing transferability trade-offs during SSL pretraining phases or comparing in-domain versus out-of-domain performance dynamics.

Contribution 2: Reformulation of critical period analysis for SSL

Description: The authors adapt critical period analysis from supervised learning to SSL by injecting deficits into pretraining data and computing Fisher Information on pretext objectives rather than class labels. This reformulation enables tracking plasticity phases during SSL pretraining without requiring downstream supervision.

This contribution was assessed against **7 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. Privacy-Aware Continual Self-Supervised Learning on Multi-Window Chest Computed Tomography for Domain-Shift Robustness

URL: View paper

Brief Assessment

Privacy Continual SSL[63] focuses on continual self-supervised learning for medical imaging with privacy preservation through latent replay, not on critical period analysis or Fisher Information dynamics during SSL pretraining.

2. Visual Reinforcement Learning With Self-Supervised 3D Representations

URL: View paper

Brief Assessment

Visual RL 3D[37] focuses on 3D representation learning for visual reinforcement learning in robotics, not on critical period analysis in self-supervised learning. The paper does not address critical periods, Fisher Information dynamics during SSL pretraining, or plasticity phases in SSL.

3. Rethinking Evaluation Protocols of Visual Representations Learned via Self-supervised Learning

URL: View paper

Brief Assessment

Rethinking SSL Evaluation[61] focuses on evaluation protocols (linear probing, transfer learning) and hyperparameter sensitivity in SSL methods, not on critical periods or plasticity phases during pretraining. The paper does not address Fisher Information dynamics or deficit injection experiments to track learning phases.

4. Augmentation-aware Self-supervised Learning with Conditioned Projector

URL: View paper

Brief Assessment

Augmentation Aware SSL[62] focuses on augmentation-aware conditioning in self-supervised learning to preserve augmentation information in representations, not on critical period analysis or plasticity phases during SSL pretraining.

5. Self-Supervised Representation Learning for Quasi-Simultaneous Arrival Signal Identification Based on Reconnaissance Drones

URL: View paper

Brief Assessment

Quasi-Simultaneous Signal Identification[60] focuses on signal identification for reconnaissance drones using masked autoencoder pretraining, not on critical period analysis or plasticity dynamics in SSL.

6. Making Self-supervised Learning Robust to Spurious Correlation via Learning-speed Aware Sampling

URL: View paper

Brief Assessment

Learning Speed Sampling[65] focuses on mitigating spurious correlations in SSL through learning-speed aware sampling, not on analyzing critical periods or plasticity phases during SSL pretraining.

7. Prediction of Pea Yield and Nodulation from Proximal Field and Root Imaging

URL: View paper

Brief Assessment

Pea Yield Prediction[64] focuses on agricultural yield prediction using proximal field and root imaging with self-supervised features (DINOv2). It does not address critical period analysis, Fisher Information dynamics, or SSL pretraining phases.

Contribution 3: CP-guided checkpoint selection and self-distillation interventions

Description: The authors introduce two practical methods leveraging critical period insights: CP-guided checkpoint selection identifies intermediate checkpoints at CP closure for improved OOD transfer, and CP-guided self-distillation selectively distills early-layer representations from CP checkpoints into final models to balance the transferability trade-off.

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. Online distilling from checkpoints for neural machine translation

URL: View paper

Brief Assessment

Online Checkpoint Distillation[69] focuses on using checkpoints as teacher models during NMT training to prevent overfitting, not on identifying critical periods or analyzing out-of-domain transferability in self-supervised learning.

2. Where to Begin: Efficient Pretraining via Subnetwork Selection and Distillation

URL: View paper

Brief Assessment

Subnetwork Selection Distillation[71] focuses on efficient pretraining of small language models through subnetwork initialization and distillation from larger teachers, not on identifying critical periods or checkpoint selection for OOD robustness in self-supervised visual learning.

3. Neural Machine Translation Transfer Model Based on Mutual Domain Guidance

URL: View paper

Brief Assessment

Mutual Domain Guidance[74] focuses on neural machine translation domain transfer using distillation and pretraining models, not on checkpoint selection or self-distillation for improving OOD robustness in self-supervised learning pretraining.

4. From Large to Small Distillation in the Age of Large Language Models

URL: View paper

Brief Assessment

Large to Small Distillation[70] focuses on knowledge distillation from large to small language models, not on checkpoint selection during pretraining or self-distillation for OOD robustness in SSL.

5. A Simple Recipe for Improving Out-of-Domain Retrieval in Dense Encoders

URL: View paper

Brief Assessment

Dense Encoder Retrieval[75] focuses on checkpoint selection and distillation for dense retrieval models in out-of-domain settings, not on self-supervised learning pretraining phases or critical periods. The technical domains and objectives differ fundamentally.

6. Latent syntax weaving in large language model representations: A novel mechanism for self-referential consistency in neural architectures

URL: View paper

Brief Assessment

Latent Syntax Weaving[68] focuses on self-referential consistency mechanisms in LLM representations, not on checkpoint selection or self-distillation for improving out-of-domain robustness in pretraining as described in the original contribution.

7. Capture the key in reasoning to enhance CoT distillation generalization

URL: View paper

Brief Assessment

CoT Distillation Generalization[66] focuses on distilling chain-of-thought reasoning from LLMs to smaller models by identifying key reasoning steps through dual CoTs and minimum edit distance. This is fundamentally different from CP-guided checkpoint selection for SSL pretraining, which uses Fisher information dynamics to identify optimal checkpoints during self-supervised learning.

8. Progressive distillation induces an implicit curriculum

URL: View paper

Brief Assessment

Progressive Distillation Curriculum[67] focuses on knowledge distillation using intermediate teacher checkpoints to accelerate student training through implicit curriculum learning. The original paper's CP-guided methods address out-of-domain robustness in self-supervised learning by identifying critical periods via Fisher information and perturbation experiments—a fundamentally different problem domain and mechanism.

9. LATERAL: Learning Automatic, Transfer-Enhanced, and Relation-Aware Labels

URL: View paper

Brief Assessment

LATERAL[73] mentions checkpoint selection and self-distillation in passing but does not present these as systematic methods for improving OOD robustness through critical period analysis. The candidate focuses on cross-domain few-shot learning rather than the original paper's framework of using critical periods to guide checkpoint selection and self-distillation for balancing transferability trade-offs.

10. Efficient Knowledge Distillation from Model Checkpoints

URL: View paper

Brief Assessment

Checkpoint Knowledge Distillation[72] focuses on using intermediate checkpoints as teachers for knowledge distillation to improve student model performance, not on identifying checkpoints for OOD transfer or balancing transferability trade-offs in self-supervised pretraining contexts.

Appendix: Text Similarity Detection

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

References

- [0] Understanding the Learning Phases in Self-Supervised Learning via Critical Periods View paper
- [1] Reusability report: Exploring the transferability of self-supervised learning models from single-cell to spatial transcriptomics View paper
- [2] Self-supervised learning: A succinct review View paper
- [3] Big self-supervised models are strong semi-supervised learners View paper
- [4] The lottery tickets hypothesis for supervised and self-supervised pre-training in computer vision models View paper
- [5] Improved transferability of self-supervised learning models through batch normalization finetuning View paper
- [6] Maefe: Masked autoencoders family of electrocardiogram for self-supervised pretraining and transfer learning View paper
- [7] GraphCLIP: Enhancing Transferability in Graph Foundation Models for Text-Attributed Graphs View paper
- [8] Leveraging in-the-wild Data for Effective Self-supervised Pretraining in Speaker Recognition View paper
- [9] Grounded Language-Image Pre-training View paper
- [10] Self-supervised learning for skin cancer diagnosis with limited training data View paper
- [11] A Broad Study on the Transferability of Visual Representations with Contrastive Learning View paper
- [12] Cost-effective image recognition of water leakage in metro tunnels using self-supervised learning View paper
- [13] Multi-modal representation learning in retinal imaging using self-supervised learning for enhanced clinical predictions View paper
- [14] Learning Transferable Representations for Image Anomaly Localization Using Dense Pretraining View paper

- [15] Transfer learning application of self-supervised learning in ARPES View paper
- [16] SFT: Few-Shot Learning via Self-Supervised Feature Fusion With Transformer View paper
- [17] Improving Seismic Fault Recognition with Self-Supervised Pre-Training: A Study of 3D Transformer-Based with Multi-Scale Decoding and Fusion View paper
- [18] Improving the classification of veterinary thoracic radiographs through inter-species and inter-pathology self-supervised pre-training of deep learning models View paper
- [19] Boost supervised pretraining for visual transfer learning: Implications of self-supervised contrastive representation learning View paper
- [20] Redundancy is Not What You Need: An Embedding Fusion Graph Auto-Encoder for Self-Supervised Graph Representation Learning View paper
- [21] FDSP-HRID: Few-Shot Detector With Self-Supervised Pretraining for High-Speed Rail Infrastructure Defects View paper
- [22] Open-Domain Response Generation in Low-Resource Settings using Self-Supervised Pre-Training of Warm-Started Transformers View paper
- [23] Consecutive pre-training: A knowledge transfer learning strategy with relevant unlabeled data for remote sensing domain View paper
- [24] Spatial-Temporal Pre-Training for Embryo Viability Prediction Using Time-Lapse Videos View paper
- [25] On-Device Constrained Self-Supervised Learning for Keyword Spotting via Quantization Aware Pre-Training and Fine-Tuning View paper
- [26] DDAE: Towards Deep Dynamic Vision BERT Pretraining View paper
- [27] â□¦ by Supervised or Selfâ□□Supervised Learning for Chest Radiograph Classification: A Comparative Study Against ImageNet Counterparts in Coldâ□□Start Active Learning View paper
- [28] Advancing animal behavior recognition with self-supervised pre-training on unlabeled data View paper
- [29] MoCo Pretraining Improves Representation and Transferability of Chest X-ray Models View paper
- [30] Self-supervised Pre-training for Mirror Detection View paper
- [31] A Self-Supervised Learning Pretraining Framework for Remote Sensing Image Change Detection View paper
- [32] Self-supervised pretraining of transformers for satellite image time series classification View paper
- [33] A Universal Feature Extractor Based on Self-Supervised Pre-Training for Fault Diagnosis of Rotating Machinery under Limited Data View paper
- [34] Self-Supervised Convolutional Audio Models are Flexible Acoustic Feature Learners: A Domain Specificity and Transfer-Learning Study View paper
- [35] Look, Listen, and Attend: Co-Attention Network for Self-Supervised Audio-Visual Representation Learning View paper
- [36] PEAMATL: A strategy for developing near-infrared spectral prediction models under domain shift using self-supervised transfer learning View paper
- [37] Visual Reinforcement Learning With Self-Supervised 3D Representations View paper
- [38] GSTBench: A Benchmark Study on the Transferability of Graph Self-Supervised Learning View paper
- [39] A Multimodal Protein Representation Framework for Quantifying Transferability Across Biochemical Downstream Tasks View paper
- [40] Exploring Model Transferability through the Lens of Potential Energy View paper
- [41] Class Incremental Learning with Self-Supervised Pre-Training and Prototype Learning View paper
- [42] BarlowTwins-CXR: enhancing chest X-ray abnormality localization in heterogeneous data with cross-domain self-supervised learning View paper
- [43] Self-supervised transfer learning of physiological representations from free-living wearable data View paper
- [44] When Does Contrastive Learning Preserve Adversarial Robustness from Pretraining to Finetuning? View paper
- [45] Self-Supervised Representation Learning: Introduction, advances, and challenges View paper
- [46] Boosting self-supervised learning via knowledge transfer View paper
- [47] DRAFT: A novel framework to reduce domain shifting in self-supervised learning and its application to children's ASR View paper
- [48] S3L: Spectrum Transformer for Self-Supervised Learning in Hyperspectral Image Classification View paper
- [49] MPCD: A Multitask Graph Transformer for Molecular Property Prediction by Integrating Common and Domain Knowledge. View paper
- [50] Recognizing surgical phases anywhere: Few-Shot Test-Time adaptation and Task-Graph guided refinement View paper
- [51] Improving generalization for ai-synthesized voice detection View paper
- [52] Self-supervised learning for generalizable out-of-distribution detection View paper
- [53] Robust and data-efficient generalization of self-supervised machine learning for diagnostic imaging View paper
- [54] Disentangled graph self-supervised learning for out-of-distribution generalization View paper
- [55] SelfReg: Self-supervised Contrastive Regularization for Domain Generalization View paper
- [56] Cross-domain pre-training with language models for transferable time series representations View paper
- [57] How well do self-supervised models transfer? View paper
- [58] Decoding Musical Neural Activity in Patients With Disorders of Consciousness Through Self-Supervised Contrastive Domain Generalization View paper
- [59] Health Assessment of Rotating Equipment With Unseen Conditions Using Adversarial Domain Generalization Toward Self-Supervised Regularization Learning View paper
- [60] Self-Supervised Representation Learning for Quasi-Simultaneous Arrival Signal Identification Based on Reconnaissance Drones View paper
- [61] Rethinking Evaluation Protocols of Visual Representations Learned via Self-supervised Learning View paper
- [62] Augmentation-aware Self-supervised Learning with Conditioned Projector View paper
- [63] Privacy-Aware Continual Self-Supervised Learning on Multi-Window Chest Computed Tomography for Domain-Shift Robustness View paper
- [64] Prediction of Pea Yield and Nodulation from Proximal Field and Root Imaging View paper
- [65] Making Self-supervised Learning Robust to Spurious Correlation via Learning-speed Aware Sampling View paper
- [66] Capture the key in reasoning to enhance CoT distillation generalization View paper
- [67] Progressive distillation induces an implicit curriculum View paper
- [68] Latent syntax weaving in large language model representations: A novel mechanism for self-referential consistency in neural architectures View paper
- [69] Online distilling from checkpoints for neural machine translation View paper

- [70] From Large to Small Distillation in the Age of Large Language Models View paper
- [71] Where to Begin: Efficient Pretraining via Subnetwork Selection and Distillation View paper
- [72] Efficient Knowledge Distillation from Model Checkpoints View paper
- [73] LATERAL: Learning Automatic, Transfer-Enhanced, and Relation-Aware Labels View paper
- [74] Neural Machine Translation Transfer Model Based on Mutual Domain Guidance View paper
- [75] A Simple Recipe for Improving Out-of-Domain Retrieval in Dense Encoders View paper