Article

Self-Supervised Learning Principles Challenges and Emerging Directions

Jenifer Nadine *

Posted Date: 24 February 2025

doi: 10.20944/preprints202502.1894.v1

Keywords: Self-supervised learning; Representation learning; Contrastive learning; Clustering-based learning; Generative models; Deep learning; Unsupervised learning; Transfer learning; Machine learning; Artificial intelligence

 doi:10.20944/preprints202502.1894.v1

Review

Self-Supervised Learning Principles Challenges and Emerging Directions

Jenifer Nadine

Warwick Mathematics Institute, University of Warwick, United Kingdom; jenifer.nadine@warwick.ac.uk

Abstract: Self-supervised learning (SSL) has emerged as a transformative paradigm in machine learning, enabling models to learn meaningful representations from vast amounts of unlabeled data. By leveraging pretext tasks that generate supervisory signals intrinsically from data, SSL has significantly reduced the need for costly human annotations and has demonstrated remarkable performance across diverse domains, including computer vision, natural language processing, speech processing, robotics, and healthcare. This survey provides a comprehensive overview of self-supervised learning, covering its fundamental principles, major methodological approaches, and real-world applications. We categorize SSL into four primary paradigms: contrastive learning, clustering-based learning, generative modeling, and predictive learning. We discuss the theoretical underpinnings of these approaches, highlight their strengths and limitations, and analyze their impact on downstream tasks. Additionally, we explore the integration of SSL with deep learning architectures and its role in improving model generalization, robustness, and efficiency. Despite its successes, SSL faces several challenges, including the computational cost of large-scale training, sensitivity to domain shifts, difficulties in designing optimal pretext tasks, and a lack of theoretical understanding. We outline open research questions and promising future directions, such as multimodal SSL, efficient pretraining techniques, self-supervised reinforcement learning, and fairness-aware SSL. As self-supervised learning continues to evolve, it holds the potential to redefine machine learning by enabling more scalable, efficient, and generalizable models. This survey aims to provide researchers and practitioners with a comprehensive understanding of SSL, facilitating further advancements in this rapidly growing field.

Keywords: self-supervised learning; representation learning; contrastive learning; clustering-based learning; generative models; deep learning; unsupervised learning; Transfer learning; machine learning; artificial intelligence

1. Introduction

In recent years, self-supervised learning (SSL) has emerged as a powerful paradigm for learning representations from data without the need for extensive human-labeled annotations [1]. Traditional supervised learning methods heavily rely on large-scale labeled datasets, which are expensive and time-consuming to acquire [2]. This dependency poses significant challenges in domains where labeling is difficult, ambiguous, or requires expert knowledge, such as medical imaging, natural language processing, and scientific research [3]. In contrast, self-supervised learning mitigates this challenge by leveraging vast amounts of unlabeled data, defining auxiliary (pretext) tasks that enable models to learn useful features without explicit supervision [4]. Self-supervised learning has gained significant traction due to its ability to produce high-quality representations that can be effectively transferred to downstream tasks such as classification, segmentation, object detection, and reinforcement learning. By exploiting inherent structures and relationships in the data, SSL enables models to learn rich and meaningful representations that generalize well across different domains [5]. The rapid advancements in SSL techniques have been driven by breakthroughs in contrastive learning, clustering-based approaches, generative modeling, and predictive learning. The fundamental idea behind self-supervised learning is to create surrogate tasks that provide meaningful supervision signals [6]. These tasks are designed to

 doi:10.20944/preprints202502.1894.v1

capture essential properties of data distributions and encourage the model to learn representations that encode semantic, syntactic, or structural relationships [7]. For instance, in computer vision, contrastive methods such as SimCLR and MoCo maximize agreement between different augmentations of the same image while minimizing agreement with different images [8]. In natural language processing, methods like BERT and GPT employ masked language modeling and next-sentence prediction to learn deep contextual representations [9]. Meanwhile, in speech processing, approaches such as wav2vec2.0 have demonstrated remarkable success by leveraging self-supervised pre-training on large-scale speech corpora [10]. The impact of self-supervised learning has been particularly profound in deep learning, where the ability to leverage unlabeled data is crucial for scaling models to unprecedented levels of complexity and performance. Recent studies have shown that self-supervised pretraining not only enhances performance on benchmark datasets but also improves sample efficiency, robustness, and generalization in low-data regimes [11]. Furthermore, SSL has shown promise in reducing the biases inherent in supervised learning, as models trained in a self-supervised manner are less dependent on human-annotated labels that might introduce noise and inconsistencies [12]. This survey aims to provide a comprehensive overview of the field of self-supervised learning, highlighting key methodologies, theoretical underpinnings, and practical applications [13]. We categorize SSL approaches into major paradigms, including contrastive learning, clustering-based methods, generative approaches, and predictive learning techniques [14]. Additionally, we explore the integration of SSL with various deep learning architectures, discuss its applicability in diverse domains, and outline emerging trends and future directions. The remainder of this survey is organized as follows: Section 2 provides a historical perspective and foundational principles of self-supervised learning [15]. Section 3 delves into the major methodologies, discussing their advantages, limitations, and theoretical motivations [16]. Section 4 explores applications of SSL across computer vision, natural language processing, speech processing, and robotics. Section 5 highlights ongoing challenges and potential research directions [17]. Finally, Section 6 concludes with a discussion on the future of self-supervised learning and its broader impact on artificial intelligence [18]. With the growing importance of data-efficient learning paradigms, self-supervised learning is poised to become a cornerstone of modern machine learning, paving the way for more robust, scalable, and generalizable AI systems. Through this survey, we aim to equip researchers and practitioners with a thorough understanding of SSL, providing insights into both its current capabilities and its promising future.

2. Historical Perspective and Foundations

Self-supervised learning (SSL) has its roots in the broader fields of unsupervised learning and representation learning, both of which have been central to artificial intelligence research for decades [19]. The goal of unsupervised learning has traditionally been to discover meaningful patterns in data without explicit labels, a challenge that has been explored through clustering, density estimation, and dimensionality reduction techniques [20]. Early attempts at learning representations without supervision include principal component analysis (PCA), independent component analysis (ICA), and autoencoders, all of which sought to extract useful features from raw data [21]. The emergence of deep learning in the 2010s led to a paradigm shift in representation learning, with self-supervised learning playing an increasingly important role. Early deep learning models, such as deep belief networks (DBNs) and stacked autoencoders, demonstrated the power of unsupervised pretraining, but they lacked a structured way to define auxiliary tasks that could guide feature learning [22]. The introduction of contrastive learning and predictive learning mechanisms provided a more systematic framework for self-supervised learning, enabling models to leverage vast amounts of unlabeled data to learn generalizable representations [23]. One of the earliest successes of self-supervised learning came from natural language processing (NLP), where word embedding techniques such as Word2Vec, GloVe, and FastText utilized co-occurrence statistics to learn semantic representations of words. These methods laid the foundation for modern self-supervised pretraining approaches, such as BERT and GPT, which employ masked language modeling and autoregressive prediction tasks to learn deep

 doi:10.20944/preprints202502.1894.v1

contextual representations. In the domain of computer vision, early self-supervised techniques focused on tasks such as image inpainting, colorization, and context prediction. However, the breakthrough came with the development of contrastive learning methods such as SimCLR, MoCo, and BYOL, which demonstrated that instance discrimination tasks could lead to representations rivaling those learned through fully supervised methods [24]. These methods introduced the concept of leveraging positive and negative pairs, where models maximize agreement between augmented views of the same image while ensuring separation from different images [25]. Another foundational approach in self-supervised learning is clustering-based methods, which rely on grouping similar representations together without explicit labels [26]. Methods such as DeepCluster and SwAV utilize iterative clustering mechanisms to refine learned representations and have been particularly effective in scenarios with large-scale unlabeled data [27]. Meanwhile, generative approaches, such as Variational Autoencoders (VAEs) and Generative Adversarial Networks (GANs), have contributed to self-supervised learning by training models to generate realistic samples from learned distributions [28]. Despite the significant advancements in self-supervised learning, several challenges remain, including the design of effective pretext tasks, the need for large-scale computational resources, and the difficulty of evaluating learned representations. Nevertheless, the field continues to evolve rapidly, with new methods and theoretical insights driving progress in areas ranging from robotics to bioinformatics [29]. The next section explores the major methodologies of self-supervised learning in greater depth, categorizing approaches based on their core principles and training objectives [30].

3. Major Methodologies in Self-Supervised Learning

Self-supervised learning (SSL) encompasses a diverse range of techniques designed to learn useful data representations without relying on labeled annotations [31]. These methodologies are broadly categorized into four main paradigms: contrastive learning, clustering-based approaches, generative methods, and predictive learning techniques. Each of these approaches defines a different type of pretext task that serves as a learning signal, enabling models to acquire rich and generalizable features [32]. In this section, we provide a comprehensive overview of these methodologies, discussing their core principles, strengths, and limitations.

3.1. Contrastive Learning

Contrastive learning has emerged as one of the most effective and widely adopted self-supervised learning techniques [33]. The central idea behind contrastive learning is to maximize the similarity between representations of positive pairs (i.e., augmented views of the same instance) while minimizing the similarity between negative pairs (i.e., representations of different instances). By enforcing this objective, contrastive learning ensures that representations capture discriminative features that generalize well to downstream tasks [34]. Several influential methods have been developed in this paradigm, including:

- **SimCLR** : This method leverages data augmentations and a contrastive loss function to learn robust representations [35]. It introduces a learnable projection head and temperature-scaled similarity functions to improve performance [36].
- **MoCo** : Momentum Contrast (MoCo) utilizes a momentum-based encoder to maintain a dynamic dictionary of negative samples, thereby stabilizing contrastive learning and enabling learning with large-scale datasets [37].
- **BYOL** : Unlike traditional contrastive learning methods, Bootstrap Your Own Latent (BYOL) eliminates the need for negative samples by employing a teacher-student framework with momentum updates [38].
- **SimSiam** : This method further refines contrastive learning by demonstrating that even without negative samples or momentum encoders, meaningful representations can be learned using a stop-gradient mechanism [39].

 doi:10.20944/preprints202502.1894.v1

Contrastive learning has been particularly successful in computer vision applications, outperforming supervised learning in several benchmark tasks [40]. However, it remains computationally expensive due to the need for large batch sizes and complex memory bank mechanisms [41].

3.2. Clustering-Based Approaches

Clustering-based self-supervised learning methods leverage the idea that similar data points should be assigned to the same cluster, even in the absence of explicit labels [42]. These approaches iteratively refine feature representations by optimizing clustering assignments [43]. Notable methods in this category include:

- **DeepCluster** : This method performs iterative k-means clustering in feature space, updating both feature representations and cluster assignments in an alternating fashion [44].
- **SwAV** : Swapping Assignments between Views (SwAV) introduces a novel clustering-based approach where representations are learned by solving a swapped prediction problem across different augmentations [45].
- **SeLa** : Self-Labelling (SeLa) uses optimal transport techniques to assign cluster labels, improving the stability and effectiveness of clustering-based learning [46].

Clustering-based methods provide a natural way to learn high-level semantic representations and are particularly useful in scenarios with diverse and complex data distributions [47]. However, they can be sensitive to the choice of clustering algorithms and hyperparameters [48].

3.3. Generative Approaches

Generative modeling plays a crucial role in self-supervised learning by training models to generate, reconstruct, or transform data in meaningful ways. These methods encourage representations that capture the underlying structure of data distributions. Key generative approaches include:

- **Autoencoders** : Autoencoders and their variants, such as Variational Autoencoders (VAEs), learn compressed representations by encoding data into a latent space and reconstructing it from that space [49].
- **Generative Adversarial Networks (GANs)** : GANs train a generator and discriminator in an adversarial framework, leading to the learning of high-quality generative representations [50].
- **Masked Image Modeling** : Methods such as MAE (Masked Autoencoder) extend BERT-like pretraining ideas to computer vision by masking image patches and training models to reconstruct them.

Generative methods have been widely applied in image synthesis, denoising, and anomaly detection [51]. However, they often require extensive computational resources and may struggle with stability during training [52].

3.4. Predictive Learning Techniques

Predictive learning methods involve designing pretext tasks where the model must predict missing or transformed parts of the input data. These methods provide a simple yet effective way to learn meaningful representations [53]. Prominent examples include:

- **BERT** : The Bidirectional Encoder Representations from Transformers (BERT) model introduced masked language modeling (MLM), where certain tokens are masked and the model is trained to predict them.
- **GPT** : The Generative Pretrained Transformer (GPT) series follows an autoregressive approach, predicting the next token given previous tokens [54].
- **wav2vec** : Applied to speech processing, wav2vec pretrains models by learning to predict masked audio segments from their context [55].
- **Self-Supervised Learning for Robotics** : Predictive tasks in robotics involve learning future states or dynamics to enable better control policies [56].

 doi:10.20944/preprints202502.1894.v1

Predictive learning techniques are highly flexible and have demonstrated significant success in language, vision, and speech domains [57]. However, designing effective pretext tasks remains a challenge, as some tasks may lead to trivial or suboptimal representations [58].

3.5. Comparison of SSL Approaches

Each self-supervised learning methodology offers distinct advantages and trade-offs, as summarized in Table 1 [59].

Table 1. Comparison of different self-supervised learning methodologies.

Method	Key Idea	Strengths and Weaknesses
Contrastive Learning	Instance discrimination	Highly effective, but computationally expensive
Clustering-Based	Iterative clustering refinement	Captures high-level semantics, but sensitive to hyperparameters
Generative Approaches	Data reconstruction/generation	Produces detailed representations, but costly to train
Predictive Learning	Predict missing data	Flexible and effective, but task design is crucial

In the next section, we explore the diverse applications of self-supervised learning across various domains, highlighting its impact in computer vision, natural language processing, speech processing, and robotics [60].

4. Applications of Self-Supervised Learning

Self-supervised learning (SSL) has demonstrated remarkable success across a wide range of domains, providing significant improvements in representation learning without the need for large amounts of labeled data [61]. From computer vision to natural language processing, speech recognition, and robotics, SSL has enabled breakthroughs in efficiency, performance, and scalability [62]. This section explores the diverse applications of SSL, highlighting key advancements and real-world use cases [63].

4.1. Computer Vision

The field of computer vision has greatly benefited from SSL, with many state-of-the-art models leveraging self-supervised pretraining to enhance feature extraction and generalization [64]. SSL has been successfully applied in the following areas:

- **Image Classification**: Self-supervised models such as SimCLR , MoCo , and SwAV have achieved performance on par with fully supervised models on image classification tasks by leveraging large-scale unlabeled datasets such as ImageNet.
- **Object Detection and Segmentation**: Pretrained self-supervised models provide strong feature representations that transfer well to downstream tasks like object detection (e.g., Faster R-CNN, YOLO) and segmentation (e.g., Mask R-CNN, DeepLab) [65].
- **Medical Imaging**: SSL has significantly improved medical image analysis, where labeled data is scarce [66]. Methods such as contrastive learning and masked image modeling have been used in MRI and CT scan analysis [67].
- **Video Understanding**: Self-supervised video models use temporal consistency and frame prediction to learn representations for action recognition, video retrieval, and anomaly detection in surveillance systems [68].

By reducing dependence on labeled data, SSL has unlocked new possibilities in computer vision applications, making deep learning more practical for real-world deployment [69,70].

4.2. Natural Language Processing

Self-supervised learning has revolutionized natural language processing (NLP) by enabling large-scale pretraining of deep language models [71]. Key applications include:

- **Text Understanding and Generation**: Transformer-based models such as BERT , RoBERTa , and GPT utilize self-supervised objectives like masked language modeling and autoregressive text

 doi:10.20944/preprints202502.1894.v1

prediction to achieve state-of-the-art performance in tasks such as sentiment analysis, machine translation, and text summarization [72].

- **Question Answering and Chatbots**: Self-supervised models have significantly improved natural language understanding in conversational AI systems, powering advanced chatbots and virtual assistants [73].
- **Semantic Search and Information Retrieval**: SSL has enhanced document ranking, question-answering retrieval systems, and knowledge extraction in search engines [74].
- **Biomedical and Legal Text Analysis**: SSL models have been fine-tuned for specialized domains such as bioinformatics (BioBERT) and legal document processing, where labeled data is scarce [75].

Self-supervised pretraining has drastically reduced the need for human-labeled datasets in NLP, making it feasible to train large-scale models that generalize across multiple tasks [76].

4.3. Speech and Audio Processing

SSL has transformed speech and audio processing by enabling models to learn from large amounts of unlabeled audio data. Key applications include:

- **Speech Recognition**: Models like wav2vec 2.0 use contrastive pretraining on raw audio waveforms, achieving significant improvements in automatic speech recognition (ASR) without requiring transcribed speech data [77].
- **Speaker Identification and Verification**: SSL-based embeddings such as HuBERT enhance speaker recognition and verification systems in real-world applications like voice authentication.
- **Music and Sound Classification**: SSL models have been applied to tasks such as music genre classification, environmental sound recognition, and audio event detection, improving generalization across diverse datasets [78].
- **Speech Enhancement and Denoising**: Self-supervised representations help in denoising speech signals, making speech processing models more robust in noisy environments [79].

By leveraging vast amounts of unlabeled speech data, SSL has significantly improved the efficiency and performance of modern speech and audio systems [80].

4.4. Robotics and Reinforcement Learning

In robotics and reinforcement learning (RL), SSL has been instrumental in enabling robots to learn meaningful representations from sensory data without explicit supervision. Applications include:

- **Robot Perception**: Self-supervised vision models allow robots to understand and interpret their surroundings, facilitating object recognition, depth estimation, and scene understanding [81].
- **Control and Policy Learning**: SSL is used in RL to pretrain representations for state estimation and action prediction, improving sample efficiency in robotic control tasks [82].
- **Autonomous Navigation**: Self-supervised methods enable robots and autonomous vehicles to learn from unlabeled sensor data, enhancing obstacle avoidance and motion planning [83].
- **Grasping and Manipulation**: SSL has been applied to robotic grasping, allowing robots to learn object interactions through self-generated experiences .

Self-supervised learning has enabled robots to operate more autonomously by reducing their reliance on costly expert demonstrations and labeled datasets [84].

4.5. Healthcare and Biomedical Applications

The healthcare domain has increasingly benefited from SSL, where labeled medical data is often scarce and expensive to obtain. Applications include:

- **Disease Diagnosis and Prognosis**: SSL-based models have been applied to X-ray, MRI, and histopathology image analysis, improving diagnostic accuracy while reducing the need for labeled medical images .

 doi:10.20944/preprints202502.1894.v1

- **Drug Discovery**: SSL has been used to predict molecular properties, accelerating drug discovery by leveraging vast amounts of unlabeled chemical compound data [85].
- **Genomics and Bioinformatics**: Self-supervised techniques have enabled better representation learning in genomics, leading to improvements in gene sequence analysis and protein structure prediction .

By making better use of unlabeled medical data, SSL has the potential to advance precision medicine and accelerate research in healthcare [86].

4.6. Finance and Anomaly Detection

SSL has also been applied in the financial sector and security-related fields, where labeled fraud and anomaly data are limited. Applications include:

- **Fraud Detection**: Self-supervised models can learn transaction patterns and identify anomalies in financial transactions, improving fraud detection in banking and e-commerce [87].
- **Stock Market Prediction**: SSL has been used to pretrain models for financial forecasting, leveraging unlabeled market data to identify trends and patterns.
- **Cybersecurity**: Self-supervised anomaly detection methods help detect unusual network activity and security breaches with minimal labeled data [88].

The ability of SSL to learn from vast amounts of unstructured financial data makes it highly valuable for risk assessment and anomaly detection [89].

4.7. Scientific Research and Other Domains

Beyond traditional AI applications, SSL has been applied to various scientific disciplines, including:

- **Astronomy**: SSL models help analyze vast amounts of unlabeled astronomical data for galaxy classification and cosmic event detection.
- **Climate Science**: Self-supervised models are being used to improve climate predictions and analyze satellite imagery for environmental monitoring.
- **Material Science**: SSL assists in predicting material properties, accelerating discoveries in chemistry and physics [90].

4.8. Summary

Table 2 summarizes the key applications of self-supervised learning across different domains [91].

Table 2. Applications of self-supervised learning in various domains.

Domain	Key Applications
Computer Vision	Classification, object detection, segmentation
NLP	Language modeling, chatbots, search
Speech	ASR, speaker identification, audio classification
Robotics	Perception, navigation, manipulation
Healthcare	Medical imaging, drug discovery, genomics
Finance	Fraud detection, stock market prediction

The next section explores the challenges and future directions in self-supervised learning [92].

5. Challenges and Future Directions

Despite its remarkable progress, self-supervised learning (SSL) still faces several challenges that limit its widespread adoption and effectiveness [93]. These challenges span theoretical, computational, and practical considerations. Addressing these limitations will be crucial for the future development of SSL. In this section, we discuss the major challenges and outline potential directions for future research [94].

 doi:10.20944/preprints202502.1894.v1

5.1. Challenges in Self-Supervised Learning

5.1.1. Designing Effective Pretext Tasks

One of the fundamental challenges in SSL is the design of pretext tasks that encourage the learning of useful and transferable representations [95]. While contrastive learning, clustering, and generative methods have achieved significant success, the effectiveness of these approaches varies depending on the data domain [96]. For example:

- Contrastive learning relies on negative samples, which can be difficult to define optimally [97].
- Clustering-based methods require careful initialization and hyperparameter tuning [98].
- Predictive modeling approaches may lead to trivial solutions where the model learns to exploit shortcuts instead of meaningful representations [99].

Developing more generalizable and domain-agnostic pretext tasks remains an open research problem.

5.1.2. Computational Costs and Scalability

Many SSL methods, especially contrastive learning, require large batch sizes and extensive computational resources [100]. Training models such as SimCLR or MoCo can be prohibitively expensive, requiring specialized hardware such as TPUs or multi-GPU setups [101]. Challenges include:

- The need for large-scale negative sampling, which increases memory consumption [102].
- Training instability due to the complexity of optimization [103].
- The high computational burden of data augmentations in vision-based SSL.

Research into more efficient training paradigms, such as memory-efficient contrastive learning and self-distillation, is crucial for making SSL accessible to a broader range of users [104].

5.1.3. Evaluation and Benchmarking

Unlike supervised learning, where model evaluation is straightforward using labeled test sets, evaluating SSL models is more challenging. Issues include:

- The lack of standardized benchmarks across different domains [105].
- The reliance on downstream tasks for evaluation, which may not always reflect the quality of learned representations [106].
- The difficulty of comparing different SSL methods due to variations in experimental setups [107].

Developing more robust evaluation metrics that do not rely on extensive labeled data would significantly benefit SSL research.

5.1.4. Domain Adaptation and Generalization

SSL models often struggle with domain shifts, meaning that representations learned from one dataset may not transfer well to another [108]. For instance:

- An SSL model trained on natural images may perform poorly on medical images [109].
- Language models trained on one domain (e.g., news articles) may not generalize well to another (e.g., scientific texts) [110].

Future research should focus on improving domain adaptation techniques in SSL, potentially through meta-learning, domain-invariant representation learning, or hybrid SSL-supervised approaches [111].

5.1.5. Robustness to Noisy and Biased Data

SSL models, like supervised models, can inherit biases present in the training data. Since they rely on learning patterns from raw data, they may amplify existing biases or learn spurious correlations [112]. Additionally, self-supervised pretraining on noisy or low-quality data may lead to suboptimal representations [113]. Addressing these issues requires:

- Methods for detecting and mitigating biases in self-supervised models [114].
- Robust SSL approaches that can handle noise and outliers effectively [115].

 doi:10.20944/preprints202502.1894.v1

- Techniques for fairness-aware self-supervised learning [116].

5.1.6. Lack of Theoretical Understanding

While SSL has demonstrated empirical success, its theoretical foundations remain underexplored. Questions such as why certain pretext tasks lead to better representations and how SSL relates to human learning are still open [117]. Future research directions include:

- Developing mathematical frameworks to analyze self-supervised learning.
- Understanding the role of mutual information and information bottlenecks in SSL [118].
- Exploring the connections between SSL and cognitive science to draw insights from human learning processes [119].

5.2. Future Directions in Self-Supervised Learning

To overcome the aforementioned challenges, several promising research directions are emerging in SSL [120].

5.2.1. Beyond Contrastive Learning: Towards More Efficient Methods

Recent developments such as BYOL and SimSiam have demonstrated that meaningful representations can be learned without negative samples [121]. Future research could explore:

- Novel architectures that reduce the dependency on contrastive loss.
- Hybrid SSL approaches that combine contrastive, clustering, and generative methods.
- Self-distillation techniques for improving SSL efficiency [122].

5.2.2. Self-Supervised Learning for Multimodal Data

Most SSL research has focused on single modalities such as images or text [123]. However, real-world applications often involve multimodal data (e.g., vision and language, speech and text) [124]. Future work could explore:

- Cross-modal contrastive learning to align different data modalities [125].
- Joint representations that can be leveraged for multimodal tasks such as video understanding and robotics [126].
- Applications of SSL to emerging fields such as bioinformatics and autonomous systems [127].

5.2.3. Self-Supervised Learning for Small Data Regimes

While SSL is often associated with large-scale datasets, recent efforts are exploring its potential for small-data regimes [128]. Key directions include:

- Few-shot and meta-learning approaches that integrate SSL [129].
- Self-supervised learning tailored for low-data domains such as medical imaging and remote sensing.
- Personalized SSL models that adapt to individual users in applications such as healthcare and recommender systems [130].

5.2.4. Integrating SSL with Supervised and Reinforcement Learning

SSL is often viewed as an alternative to supervised learning, but integrating the two can lead to more powerful models. Future directions include:

- Semi-supervised learning that combines self-supervised pretraining with limited labeled data.
- Self-supervised reinforcement learning for efficient exploration in RL environments [131].
- Continual self-supervised learning for lifelong adaptation in dynamic environments [132].

5.2.5. Towards More Human-Like Learning

A long-term goal of SSL is to develop learning paradigms that more closely resemble human intelligence [133]. Potential research areas include:

 doi:10.20944/preprints202502.1894.v1

- Self-supervised learning that incorporates reasoning and abstraction [134].
- SSL models that actively seek information, similar to human curiosity-driven learning.
- The integration of self-supervised learning with neuromorphic computing for brain-inspired AI [135].

5.3. Summary

Table 3 summarizes the major challenges and future directions in self-supervised learning.

Table 3. Challenges and future directions in self-supervised learning.

Challenges	Future Directions
Pretext task design	Hybrid and domain-agnostic SSL tasks
High computational cost	Efficient self-distillation and model compression
Evaluation difficulties	Standardized SSL benchmarks and metrics
Domain generalization issues	Transfer learning and domain adaptation methods
Bias and robustness concerns	Fairness-aware and noise-resistant SSL
Theoretical limitations	Mathematical frameworks for SSL understanding

In the next section, we conclude by summarizing the key insights from our survey on self-supervised learning.

6. Conclusion

Self-supervised learning (SSL) has emerged as a powerful paradigm for learning meaningful representations from unlabeled data, significantly advancing the fields of computer vision, natural language processing, speech recognition, robotics, and beyond. By leveraging large amounts of unlabeled data, SSL has reduced the reliance on costly labeled datasets and demonstrated remarkable performance across a wide range of applications.

In this survey, we have provided a comprehensive overview of self-supervised learning, discussing its fundamental principles, various methodological approaches, and key applications. We have explored different categories of SSL, including contrastive learning, clustering-based methods, generative models, and predictive learning approaches. Additionally, we examined the impact of SSL in real-world domains and highlighted its effectiveness in scenarios where labeled data is scarce.

Despite its success, SSL faces several challenges, including high computational costs [136], difficulties in designing optimal pretext tasks, domain adaptation issues, and a lack of theoretical understanding. Addressing these challenges will require further research into more efficient SSL frameworks, robust evaluation metrics, and better generalization strategies. Future work in SSL should also focus on multimodal learning, low-data regimes, and integrating SSL with supervised and reinforcement learning to develop more adaptable and human-like AI systems.

As self-supervised learning continues to evolve, its potential for revolutionizing artificial intelligence is becoming increasingly evident. By enabling models to learn from vast amounts of unlabeled data, SSL paves the way for more scalable, efficient, and versatile AI systems. With ongoing advancements, self-supervised learning is poised to play a central role in the next generation of intelligent systems, driving innovation across numerous scientific and industrial fields.

References

1. Sun, F.; Liu, J.; Wu, J.; Pei, C.; Lin, X.; Ou, W.; Jiang, P. BERT4Rec: Sequential recommendation with bidirectional encoder representations from transformer. In Proceedings of the CIKM, 2019, pp. 1441–1450.
2. Bach, F.; Jenatton, R.; Mairal, J.; Obozinski, G. Optimization with sparsity-inducing penalties. *Foundations and Trends in Machine Learning* **2012**, *4*, 1–106.
3. Denton, E.L.; Chintala, S.; Fergus, R.; et al. Deep generative image models using a laplacian pyramid of adversarial networks. In Proceedings of the Neural Inf. Process. Syst., 2015, pp. 1486–1494.
4. Tran, D.; Ranganath, R.; Blei, D.M. Hierarchical Implicit Models and Likelihood-Free Variational Inference. In Proceedings of the Neural Inf. Process. Syst., 2017, pp. 2794–2802.

 doi:10.20944/preprints202502.1894.v1

5. Xu, H.; Zhou, Z.; Qiao, Y.; Kang, W.; Wu, Q. Self-supervised Multi-view Stereo via Effective Co-Segmentation and Data-Augmentation. In Proceedings of the AAAI Conf.Artif. Intell., 2021, pp. 3030–3038.
6. Hoang, Q.; Nguyen, T.D.; Le, T.; Phung, D. MGAN: Training Generative Adversarial Nets with Multiple Generators. In Proceedings of the Int. Conf. Learn. Represent., 2018, pp. 1–24.
7. Wu, H.; Zheng, S.; Zhang, J.; Huang, K. Gp-gan: Towards realistic high-resolution image blending. In Proceedings of the ACM Int. Conf. Multimedia, 2019, pp. 2487–2495.
8. Li, C.L.; Chang, W.C.; Cheng, Y.; Yang, Y.; Póczos, B. Mmd gan: Towards deeper understanding of moment matching network. In Proceedings of the Neural Inf. Process. Syst., 2017, pp. 2203–2213.
9. Wang, X.; Tang, X. Face photo-sketch synthesis and recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, *31*, 1955–1967.
10. Frey, B.J.; Hinton, G.E.; Dayan, P. Does the wake-sleep algorithm produce good density estimators? In Proceedings of the Neural Inf. Process. Syst., 1996, pp. 661–667.
11. Zhou, X.; Zhou, H.; Liu, Y.; Zeng, Z.; Miao, C.; Wang, P.; You, Y.; Jiang, F. Bootstrap latent representations for multi-modal recommendation. In Proceedings of the WWW, 2023, pp. 845–854.
12. Liu, Z.; Gui, J.; Luo, H. Good helper is around you: Attention-driven Masked Image Modeling. In Proceedings of the AAAI Conf.Artif. Intell., 2023, pp. 1799–1807.
13. Liu, M.; Ding, Y.; Xia, M.; Liu, X.; Ding, E.; Zuo, W.; Wen, S. STGAN: A Unified Selective Transfer Network for Arbitrary Image Attribute Editing. In Proceedings of the IEEE Conf. Comput. Vis. Pattern Recognit., 2019, pp. 3673–3682.
14. Wei, C.; Xie, L.; Ren, X.; Xia, Y.; Su, C.; Liu, J.; Tian, Q.; Yuille, A.L. Iterative reorganization with weak spatial constraints: Solving arbitrary jigsaw puzzles for unsupervised representation learning. In Proceedings of the IEEE Conf. Comput. Vis. Pattern Recognit., 2019, pp. 1910–1919.
15. Heusel, M.; Ramsauer, H.; Unterthiner, T.; Nessler, B.; Hochreiter, S. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In Proceedings of the Neural Inf. Process. Syst., 2017, pp. 6626–6637.
16. Liang, X.; Lee, L.; Dai, W.; Xing, E.P. Dual motion gan for future-flow embedded video prediction. In Proceedings of the IEEE Int. Conf. Comput. Vis., 2017, pp. 1744–1752.
17. Vapnik, V. *The nature of statistical learning theory*; Springer Science & Business Media, 2013.
18. Chen, Y.; Lai, Y.K.; Liu, Y.J. Cartoongan: Generative adversarial networks for photo cartoonization. In Proceedings of the IEEE Conf. Comput. Vis. Pattern Recognit., 2018, pp. 9465–9474.
19. Wang, G.; Manhardt, F.; Shao, J.; Ji, X.; Navab, N.; Tombari, F. Self6D: Self-Supervised Monocular 6D Object Pose Estimation. In Proceedings of the Eur. Conf. Comput. Vis., 2020.
20. Yan, S.; Xu, X.; Xu, D.; Lin, S.; Li, X. Image Classification with Densely Sampled Image Windows and Generalized Adaptive Multiple Kernel Learning. *IEEE Transactions on Cybernetics* **to be published**.
21. Chen, N.; Zhu, J.; Xing, E.P. Predictive subspace learning for multi-view data: a large margin approach. In Proceedings of the NIPS, 2010, pp. 361–369.
22. Wu, X.D.; Yu, K.; Ding, W.; Wang, H.; Zhu, X.Q. Online Feature Selection with Streaming Features. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2013**, *35*, 1178–1192.
23. Meng, Y.; Xiong, C.; Bajaj, P.; Bennett, P.; Han, J.; Song, X.; et al. Coco-lm: Correcting and contrasting text sequences for language model pretraining. *Neural Inf. Process. Syst.* **2021**, *34*, 23102–23114.
24. Thomas, M.; Cover, T. Elements of information theory. *Wiley-Interscience, 2nd edition* **2006**.
25. Li, X.; Gui, J.; Li, P. Random Fourier Features for Kernel Multi-view Discriminant Analysis. In Proceedings of the European Conference on Artificial Intelligence, 2020.
26. Xia, L.; Huang, C.; Xu, Y.; Zhao, J.; Yin, D.; Huang, J. Hypergraph contrastive collaborative filtering. In Proceedings of the SIGIR, 2022, pp. 70–79.
27. Zhang, Y.; Yeung, D.Y. A convex formulation for learning task relationships in multi-task learning. In Proceedings of the Conference on Uncertainty in Artificial Intelligence, 2010, pp. 733–742.
28. Kim, J.; Monteiro, R.D.; Park, H. Group sparsity in nonnegative matrix factorization. In Proceedings of the SIAM International Conference on Data Mining, 2012.
29. Girdhar, R.; Fouhey, D.F.; Rodriguez, M.; Gupta, A. Learning a predictable and generative vector representation for objects. In Proceedings of the Eur. Conf. Comput. Vis., 2016, pp. 484–499.
30. Li, N.; Guo, G.D.; Chen, L.F.; Chen, S. Optimal subspace classification method for complex data. *International Journal of Machine Learning and Cybernetics* **2013**, *4*, 163–171.
31. Yu, J.; Yin, H.; Xia, X.; Chen, T.; Li, J.; Huang, Z. Self-Supervised Learning for Recommender Systems: A Survey. *arXiv preprint arXiv:2203.15876* **2022**.

 doi:10.20944/preprints202502.1894.v1

32. Arora, S.; Ge, R.; Liang, Y.; Ma, T.; Zhang, Y. Generalization and equilibrium in generative adversarial nets (gans). In Proceedings of the Int. Conf. Mach. Learn., 2017, pp. 224–232.
33. Jia, W.; Hu, R.X.; Zhao, Y.; Gui, J.; Zhu, Y.H. Palmprint Recognition Using Band-Limited Minimum Average Correlation Energy Filter. In Proceedings of the International Conference on Hand-Based Biometrics, 2011, pp. 1–6.
34. Zhou, D.; Burges, C.J. Spectral clustering and transductive learning with multiple views. In Proceedings of the Int. Conf. Mach. Learn., 2007, pp. 1159–1166.
35. Bishop, C.M. *Pattern recognition and machine learning*; Springer, New York, 2006.
36. El-Nouby, A.; Sharma, S.; Schulz, H.; Hjelm, D.; El Asri, L.; Kahou, S.E.; Bengio, Y.; Taylor, G.W. Tell, Draw, and Repeat: Generating and modifying images based on continual linguistic instruction. In Proceedings of the IEEE Int. Conf. Comput. Vis., 2019, pp. 10303–10311.
37. Tschannen, M.; Djolonga, J.; Ritter, M.; Mahendran, A.; Houlsby, N.; Gelly, S.; Lucic, M. Self-Supervised Learning of Video-Induced Visual Invariances. In Proceedings of the IEEE Conf. Comput. Vis. Pattern Recognit., 2020, pp. 13806–13815.
38. Xia, L.; Kao, B.; Huang, C. OpenGraph: Towards Open Graph Foundation Models. *arXiv preprint arXiv:2403.01121* **2024**.
39. Hu, K.; Shao, J.; Liu, Y.; Raj, B.; Savvides, M.; Shen, Z. Contrast and order representations for video self-supervised learning. In Proceedings of the IEEE Int. Conf. Comput. Vis., 2021, pp. 7939–7949.
40. Wang, J.; Kumar, S.; Chang, S.F. Semi-supervised hashing for large-scale search. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 2393–2406.
41. Haykin, S. *Neural networks and learning machines*; Prentice Hall, 2008.
42. Li, J.; Liang, X.; Wei, Y.; Xu, T.; Feng, J.; Yan, S. Perceptual generative adversarial networks for small object detection. In Proceedings of the IEEE Conf. Comput. Vis. Pattern Recognit., 2017, pp. 1222–1230.
43. Maaten, L.v.d.; Hinton, G. Visualizing data using t-SNE. *Journal of Machine Learning Research* **2008**, *9*, 2579–2605.
44. Pan, J.; Dong, J.; Liu, Y.; Zhang, J.; Ren, J.; Tang, J.; Tai, Y.W.; Yang, M.H. Physics-based generative adversarial models for image restoration and beyond. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *43*, 2449–2462.
45. Wang, W.Y.; Mazaitis, K.; Cohen, W.W. A Soft Version of Predicate Invention Based on Structured Sparsity. In Proceedings of the Int. Joint Conf. Artif. Intell., 2015, pp. 3918–3924.
46. Chen, T.; Luo, C.; Li, L. Intriguing Properties of Contrastive Losses. In Proceedings of the Neural Inf. Process. Syst. Curran Associates, Inc., 2021, Vol. 34, pp. 11834–11845.
47. Naikal, N.; Yang, A.Y.; Sastry, S.S. Informative feature selection for object recognition via sparse PCA. In Proceedings of the IEEE Int. Conf. Comput. Vis., 2011, pp. 818–825.
48. Dai, Z.; Yang, Z.; Yang, F.; Cohen, W.W.; Salakhutdinov, R.R. Good semi-supervised learning that requires a bad gan. In Proceedings of the Neural Inf. Process. Syst., 2017, pp. 6510–6520.
49. Hu, Z.; Dong, Y.; Wang, K.; Chang, K.W.; Sun, Y. GPT-GNN: Generative Pre-Training of Graph Neural Networks. In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2020, pp. 1857–1867.
50. Berthelot, D.; Schumm, T.; Metz, L. Began: Boundary equilibrium generative adversarial networks. *arXiv preprint arXiv:1703.10717* **2017**.
51. Feichtenhofer, C.; Fan, H.; Xiong, B.; Girshick, R.; He, K. A large-scale study on unsupervised spatiotemporal representation learning. In Proceedings of the Proceedings of the IEEE Conf. Comput. Vis. Pattern Recognit., 2021, pp. 3299–3309.
52. Chua, T.S.; Tang, J.; Hong, R.; Li, H.; Luo, Z.; Zheng, Y.T. NUS-WIDE: A Real-World Web Image Database from National University of Singapore. In Proceedings of the ACM Conference on Image and Video Retrieval, 2009, pp. 1–9.
53. Carlini, N.; Wagner, D. Towards evaluating the robustness of neural networks. In Proceedings of the IEEE Symposium on Security and Privacy, 2017, pp. 39–57.
54. Peng, H.C.; Long, F.H.; Ding, C. Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2005**, *27*, 1226–1238.
55. Tao, Y.; Gao, M.; Yu, J.; Wang, Z.; Xiong, Q.; Wang, X. Predictive and contrastive: Dual-auxiliary learning for recommendation. *TCSS* **2022**.
56. Liu, Z.Q.; Lin, S.L.; Tan, M.T. Sparse Support Vector Machines with L_p Penalty for Biomarker Identification. *IEEE-ACM Transactions on Computational Biology and Bioinformatics* **2010**, *7*, 100–107.

 doi:10.20944/preprints202502.1894.v1

57. Frey, B.J.; Brendan, J.F.; Frey, B.J. *Graphical models for machine learning and digital communication*; MIT press, 1998.
58. Liu, W.; Wang, J.; Ji, R.; Jiang, Y.G.; Chang, S.F. Supervised hashing with kernels. In Proceedings of the IEEE Conf. Comput. Vis. Pattern Recognit., 2012, pp. 2074–2081.
59. Spurr, A.; Aksan, E.; Hilliges, O. Guiding infogan with semi-supervision. In Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Springer, 2017, pp. 119–134.
60. Cai, X.; Wang, C.; Xiao, B.; Chen, X.; Zhou, J. Regularized Latent Least Square Regression for Cross Pose Face Recognition. In Proceedings of the IJCAI, 2013, pp. 1247–1253.
61. Xie, Y.; Wang, Z.; Ji, S. Noise2Same: Optimizing A Self-Supervised Bound for Image Denoising. In Proceedings of the Neural Inf. Process. Syst., 2020.
62. Wu, Y.; Xie, R.; Zhu, Y.; Ao, X.; Chen, X.; Zhang, X.; Zhuang, F.; Lin, L.; He, Q. Multi-view multi-behavior contrastive learning in recommendation. In Proceedings of the DASFAA. Springer, 2022, pp. 166–182.
63. Lin, K.; Li, D.; He, X.; Zhang, Z.; Sun, M.T. Adversarial ranking for language generation. In Proceedings of the Neural Inf. Process. Syst., 2017, pp. 3155–3165.
64. Rusu, A.A.; Rabinowitz, N.C.; Desjardins, G.; Soyer, H.; Kirkpatrick, J.; Kavukcuoglu, K.; Pascanu, R.; Hadsell, R. Progressive neural networks. *arXiv preprint arXiv:1606.04671* **2016**.
65. Goodfellow, I. NIPS 2016 tutorial: Generative adversarial networks. *arXiv preprint arXiv:1701.00160* **2017**.
66. Zhai, D.; Liu, X.; Ji, X.; Zhao, D.; Satoh, S.; Gao, W. Supervised distributed hashing for large-scale multimedia retrieval. *IEEE Transactions on Multimedia* **2017**, *20*, 675–686.
67. Jia, W.; Cai, H.Y.; Gui, J.; Hu, R.X.; Lei, Y.K.; Wang, X.F. Newborn footprint recognition using orientation feature. *Neural Computing and Applications* **2012**, *21*, 1855–1863.
68. Zhao, J.; Xiong, L.; Li, J.; Xing, J.; Yan, S.; Feng, J. 3d-aided dual-agent gans for unconstrained face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *41*, 2380–2394.
69. Truong, Q.T.; Salah, A.; Lauw, H.W. Bilateral variational autoencoder for collaborative filtering. In Proceedings of the WSDM, 2021, pp. 292–300.
70. Zniyed, Y.; Nguyen, T.P.; et al. Enhanced network compression through tensor decompositions and pruning. *IEEE Transactions on Neural Networks and Learning Systems* **2024**.
71. Gretton, A.; Borgwardt, K.M.; Rasch, M.J.; Schölkopf, B.; Smola, A. A kernel two-sample test. *Journal of Machine Learning Research* **2012**, *13*, 723–773.
72. Bolón-Canedo, V.; Sánchez-Maroño, N.; Alonso-Betanzos, A. A review of feature selection methods on synthetic data. *Knowledge and information systems* **2013**, *34*, 483–519.
73. Li, J.; Tao, D. Simple exponential family PCA. *IEEE Transactions on Neural Networks and Learning Systems* **Mar. 2013**, *24*, 485–497.
74. Yan, X.; Misra, I.; Gupta, A.; Ghadiyaram, D.; Mahajan, D. ClusterFit: Improving Generalization of Visual Representations. In Proceedings of the IEEE Conf. Comput. Vis. Pattern Recognit., 2020, pp. 6509–6518.
75. Duan, R.; Ma, X.; Wang, Y.; Bailey, J.; Qin, A.K.; Yang, Y. Adversarial camouflage: Hiding physical-world attacks with natural styles. In Proceedings of the IEEE Conf. Comput. Vis. Pattern Recognit., 2020, pp. 1000–1008.
76. Zhu, Y.; Elhoseiny, M.; Liu, B.; Peng, X.; Elgammal, A. A generative adversarial approach for zero-shot learning from noisy texts. In Proceedings of the IEEE Conf. Comput. Vis. Pattern Recognit., 2018, pp. 1004–1013.
77. Bian, W.; Tao, D. Constrained Empirical Risk Minimization Framework for Distance Metric Learning. *IEEE Transactions on Neural Networks and Learning Systems* **Aug. 2012**, *23*, 1194–1205.
78. Zhang, M.; Ding, C.; Zhang, Y.; Nie, F. Feature Selection at the Discrete Limit. In Proceedings of the AAAI Conf.Artif. Intell., 2014, pp. 1355–1361.
79. Arora, S.; Risteski, A.; Zhang, Y. Do GANs learn the distribution? Some theory and empirics. In Proceedings of the Int. Conf. Learn. Represent., 2018, pp. 1–16.
80. Moosavi-Dezfooli, S.M.; Fawzi, A.; Fawzi, O.; Frossard, P. Universal adversarial perturbations. In Proceedings of the IEEE Conf. Comput. Vis. Pattern Recognit., 2017, pp. 1765–1773.
81. Chae, D.K.; Kang, J.S.; Kim, S.W.; Choi, J. Rating augmentation with generative adversarial networks towards accurate collaborative filtering. In Proceedings of the WWW, 2019, pp. 2616–2622.
82. Cao, J.; Hu, Y.; Zhang, H.; He, R.; Sun, Z. Learning a high fidelity pose invariant model for high-resolution face frontalization. In Proceedings of the Neural Inf. Process. Syst., 2018, pp. 2867–2877.

 doi:10.20944/preprints202502.1894.v1

83. Li, T.; Ogihara, M. Toward intelligent music information retrieval. *IEEE Transactions on Multimedia* **2006**, *8*, 564–574.
84. Yang, J.; Kannan, A.; Batra, D.; Parikh, D. Lr-gan: Layered recursive generative adversarial networks for image generation. In Proceedings of the Int. Conf. Learn. Represent., 2017, pp. 1–21.
85. Yuan, F.; He, X.; Karatzoglou, A.; Zhang, L. Parameter-efficient transfer from sequential behaviors for user modeling and recommendation. In Proceedings of the SIGIR, 2020, pp. 1469–1478.
86. Nutt, C.L.; Mani, D.R.; Betensky, R.A.; Tamayo, P.; Cairncross, J.G.; Ladd, C.; Pohl, U.; Hartmann, C.; McLaughlin, M.E.; Batchelor, T.T.; et al. Gene expression-based classification of malignant gliomas correlates better with survival than histological classification. *Cancer Research* **2003**, *63*, 1602–1607.
87. Chen, X.; Duan, Y.; Houthooft, R.; Schulman, J.; Sutskever, I.; Abbeel, P. Infogan: Interpretable representation learning by information maximizing generative adversarial nets. In Proceedings of the Neural Inf. Process. Syst., 2016, pp. 2172–2180.
88. Lu, C.; Tang, J.; Lin, M.; Lin, L.; Yan, S.; Lin, Z. Correntropy Induced L2 Graph for Robust Subspace Clustering. In Proceedings of the IEEE Int. Conf. Comput. Vis., 2013.
89. Iyer, G.; Krishna Murthy, J.; Gupta, G.; Krishna, M.; Paull, L. Geometric consistency for self-supervised end-to-end visual odometry. In Proceedings of the CVPR Workshops, 2018, pp. 267–275.
90. Gui, J.; Sun, Z.; Ji, S.; Tao, D.; Tan, T. Feature Selection Based on Structured Sparsity: A Comprehensive Study. *IEEE Transactions on Neural Networks and Learning Systems* **2017**, *28*, 1490–1507.
91. Yan, S.; Wang, H. Semi-supervised learning by sparse representation. In Proceedings of the SIAM International Conference on Data Mining, 2009, pp. 792–801.
92. An, Y.; Xue, H.; Zhao, X.; Zhang, L. Conditional Self-Supervised Learning for Few-Shot Classification. In Proceedings of the Int. Joint Conf. Artif. Intell., 2021, pp. 2140–2146.
93. Wang, X.; Liu, N.; Han, H.; Shi, C. Self-supervised heterogeneous graph neural network with co-contrastive learning. In Proceedings of the Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, 2021, pp. 1726–1736.
94. Wang, X.; Yu, K.; Dong, C.; Change Loy, C. Recovering realistic texture in image super-resolution by deep spatial feature transform. In Proceedings of the IEEE Conf. Comput. Vis. Pattern Recognit., 2018, pp. 606–615.
95. Wang, T.C.; Liu, M.Y.; Zhu, J.Y.; Tao, A.; Kautz, J.; Catanzaro, B. High-resolution image synthesis and semantic manipulation with conditional gans. In Proceedings of the IEEE Conf. Comput. Vis. Pattern Recognit., 2018, pp. 8798–8807.
96. Zhao, W.X.; Zhou, K.; Li, J.; Tang, T.; Wang, X.; Hou, Y.; Min, Y.; Zhang, B.; Zhang, J.; Dong, Z.; et al. A survey of large language models. *arXiv preprint arXiv:2303.18223* **2023**.
97. Pascual, S.; Bonafonte, A.; Serra, J. SEGAN: Speech enhancement generative adversarial network. In Proceedings of the Interspeech, 2017, pp. 3642–3646.
98. Lester, B.; Al-Rfou, R.; Constant, N. The power of scale for parameter-efficient prompt tuning. *arXiv preprint arXiv:2104.08691* **2021**.
99. Wang, X.; Chen, W.; Wang, Y.F.; Wang, W.Y. No metrics are perfect: Adversarial reward learning for visual storytelling. In Proceedings of the Annual Meeting of the Association for Computational Linguistics, 2018, pp. 1–15.
100. Song, Y.; Ma, C.; Wu, X.; Gong, L.; Bao, L.; Zuo, W.; Shen, C.; Lau, R.W.; Yang, M.H. Vital: Visual tracking via adversarial learning. In Proceedings of the IEEE Conf. Comput. Vis. Pattern Recognit., 2018, pp. 8990–8999.
101. Alemi, A.A.; Fischer, I. GILBO: one metric to measure them all. In Proceedings of the Neural Inf. Process. Syst., 2018, pp. 7037–7046.
102. Zhang, Z.Y.; Li, T.; Ding, C. Non-negative tri-factor tensor decomposition with applications. *Knowledge and information systems* **2013**, *34*, 243–265.
103. Chen, Z.; Ye, X.; Du, L.; Yang, W.; Huang, L.; Tan, X.; Shi, Z.; Shen, F.; Ding, E. AggNet for Self-supervised Monocular Depth Estimation: Go An Aggressive Step Furthe. In Proceedings of the ACM Int. Conf. Multimedia, 2021, pp. 1526–1534.
104. Cai, D.; He, X.; Han, J. Semi-supervised discriminant analysis. In Proceedings of the IEEE Int. Conf. Comput. Vis., 2007, pp. 1–7.
105. Athalye, A.; Carlini, N.; Wagner, D. Obfuscated gradients give a false sense of security: Circumventing defenses to adversarial examples. In Proceedings of the Int. Conf. Mach. Learn., 2018, pp. 274–283.
106. Hu, H.; Cui, J.; Wang, L. Region-Aware Contrastive Learning for Semantic Segmentation. In Proceedings of the IEEE Int. Conf. Comput. Vis., 2021, pp. 16291–16301.

 doi:10.20944/preprints202502.1894.v1

107. Gui, J.; Wang, C.; Zhu, L. Locality preserving discriminant projections. In Proceedings of the International Conference on Intelligent Computing, 2009, pp. 566–572.
108. Yang, M.; Liao, M.; Lu, P.; Wang, J.; Zhu, S.; Luo, H.; Tian, Q.; Bai, X. Reading and Writing: Discriminative and Generative Modeling for Self-Supervised Text Recognition. *arXiv preprint arXiv:2207.00193* **2022**.
109. Belhumeur, P.N.; Hespanha, J.P.; Kriegman, D.J. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Trans. Pattern Anal. Mach. Intell.* **Jul. 1997**, *19*, 711–720.
110. Uesaka, T.; Morino, K.; Sugiura, H.; Kiwaki, T.; Murata, H.; Asaoka, R.; Yamanishi, K. Multi-view Learning over Retinal Thickness and Visual Sensitivity on Glaucomatous Eyes. In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2017, pp. 2041–2050.
111. Boureau, Y.; Le Roux, N.; Bach, F.; Ponce, J.; LeCun, Y. Ask the locals: multi-way local pooling for image recognition. In Proceedings of the IEEE Int. Conf. Comput. Vis., 2011, pp. 2651–2658.
112. Tenenbaum, J.; De Silva, V.; Langford, J. A global geometric framework for nonlinear dimensionality reduction. *Science* **2000**, *290*, 2319–2323.
113. Nguyen, T.T.; Chang, K.; Hui, S.C. Supervised term weighting centroid-based classifiers for text categorization. *Knowledge and information systems* **2013**, *35*, 61–85.
114. Wang, X.; Gupta, A. Generative image modeling using style and structure adversarial networks. In Proceedings of the Eur. Conf. Comput. Vis., 2016, pp. 318–335.
115. Guyon, I.; Elisseeff, A. An introduction to variable and feature selection. *Journal of Machine Learning Research* **2003**, *3*, 1157–1182.
116. You, Z.H.; Lei, Y.K.; Gui, J.; Huang, D.S.; Zhou, X. Using manifold embedding for assessing and predicting protein interactions from high-throughput experimental data. *Bioinformatics* **2010**, *26*, 2744–2751.
117. Zou, H.; Hastie, T. Regularization and variable selection via the Elastic Net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **2005**, *67*, 301–320.
118. Wei, W.; Huang, C.; Xia, L.; Zhang, C. Multi-Modal Self-Supervised Learning for Recommendation. In Proceedings of the WWW, 2023, pp. 790–800.
119. Hegde, C.; Indyk, P.; Schmidt, L. A Nearly-Linear Time Framework for Graph-Structured Sparsity. In Proceedings of the Int. Conf. Mach. Learn., 2015, pp. 928–937.
120. Hu, R.X.; Jia, W.; Zhang, D.; Gui, J.; Song, L.T. Hand shape recognition based on coherent distance shape contexts. *Pattern Recognition* **2012**, *45*, 3348–3359.
121. Amodio, M.; Krishnaswamy, S. TraVeLGAN: Image-to-image Translation by Transformation Vector Learning. In Proceedings of the IEEE Conf. Comput. Vis. Pattern Recognit., 2019, pp. 8983–8992.
122. Lin, G.; Gao, C.; Li, Y.; Zheng, Y.; Li, Z.; Jin, D.; Li, Y. Dual contrastive network for sequential recommendation. In Proceedings of the SIGIR, 2022, pp. 2686–2691.
123. Liu, Z.; Ma, Y.; Schubert, M.; Ouyang, Y.; Xiong, Z. Multi-Modal Contrastive Pre-training for Recommendation. In Proceedings of the ICMR, 2022, pp. 99–108.
124. Larsen, A.B.L.; Sønderby, S.K.; Larochelle, H.; Winther, O. Autoencoding beyond pixels using a learned similarity metric. In Proceedings of the Int. Conf. Mach. Learn., 2016, pp. 1558–1566.
125. Tibshirani, R.; Saunders, M.; Rosset, S.; Zhu, J.; Knight, K. Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **2005**, *67*, 91–108.
126. Silberman, N.; Hoiem, D.; Kohli, P.; Fergus, R. Indoor segmentation and support inference from rgbd images. In Proceedings of the Eur. Conf. Comput. Vis. Springer, 2012, pp. 746–760.
127. Qin, X.; Yuan, H.; Zhao, P.; Liu, G.; Zhuang, F.; Sheng, V.S. Intent Contrastive Learning with Cross Subsequences for Sequential Recommendation. In Proceedings of the WSDM, 2024, pp. 548–556.
128. Zhu, F.; Zhu, Y.; Chang, X.; Liang, X. Vision-language navigation with self-supervised auxiliary reasoning tasks. In Proceedings of the IEEE Conf. Comput. Vis. Pattern Recognit., 2020, pp. 10012–10022.
129. Li, W.J.; Wang, S.; Kang, W.C. Feature learning based deep supervised hashing with pairwise labels **2016**. pp. 1711–1717.
130. Petzka, H.; Fischer, A.; Lukovnicov, D. On the regularization of Wasserstein GANs. In Proceedings of the Int. Conf. Learn. Represent., 2018, pp. 1–24.
131. Shetty, R.; Rohrbach, M.; Anne Hendricks, L.; Fritz, M.; Schiele, B. Speaking the same language: Matching machine to human captions by adversarial training. In Proceedings of the IEEE Int. Conf. Comput. Vis., 2017, pp. 4135–4144.
132. Yu, J.; Gao, M.; Yin, H.; Li, J.; Gao, C.; Wang, Q. Generating reliable friends via adversarial training to improve social recommendation. In Proceedings of the ICDM. IEEE, 2019, pp. 768–777.

 doi:10.20944/preprints202502.1894.v1

133. Qiao, T.; Zhang, J.; Xu, D.; Tao, D. MirrorGAN: Learning Text-to-image Generation by Redescription. In Proceedings of the IEEE Conf. Comput. Vis. Pattern Recognit., 2019, pp. 1505–1514.
134. Gui, J.; Liu, T.; Sun, Z.; Tao, D.; Tan, T. Fast supervised discrete hashing. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *40*, 490–496.
135. Berman, D.; Avidan, S.; Avidan, S. Non-local image dehazing. In Proceedings of the IEEE Conf. Comput. Vis. Pattern Recognit., 2016, pp. 1674–1682.
136. Zniyed, Y.; Nguyen, T.P.; et al. Efficient tensor decomposition-based filter pruning. *Neural Networks* **2024**, *178*, 106393.