

Rethinking Generalizability and Discriminability of Self-Supervised Learning from Evolutionary Game Theory Perspective

Jiangmeng Li^{1, *} · Zehua Zang^{1, 2, *} · Qirui Ji^{1, 2, *} · Chuxiong Sun¹ · Wenwen Qiang¹ ✉ · Junge Zhang³ · Changwen Zheng¹ · Fuchun Sun⁴ · Hui Xiong^{5 6}

Received: date / Accepted: date

Abstract Representations learned by self-supervised approaches are generally considered to possess sufficient *generalizability* and *discriminability*. However, we disclose a nontrivial *mutual-exclusion* relationship between such critical representation properties by conducting the exploratory demonstration on self-supervised learning. State-of-the-art self-supervised methods are inclined to solely enhance one of the generalizability and discriminability. Thus, learning representations *jointly* possessing strong generalizability and discriminability is a specific challenge for self-supervised learning. To this end, we revisit the learning paradigm of self-supervised learning from the evolutionary game theory (EGT) perspective and elaborate on the theoretical roadmap to achieve a desired trade-off between the representation properties. EGT performs well in analyzing the trade-off point in a *two-player* game by utilizing dynamic system modeling. However, the EGT analysis requires sufficient annotated data, which is antithetical to the principle of self-supervised learning, i.e., the EGT analysis cannot be performed without the annotations of the specific target domain for self-supervised learning. Thus, to en-

hance the methodological generalization, we propose a novel self-supervised learning method, which leverages the advances of reinforcement learning to jointly benefit from the general guidance of EGT and sequentially optimize the model to chase the consistent improvement of generalizability and discriminability for specific target domain during pre-training. On top of this, we provide a benchmark to evaluate the generalizability and discriminability of learned representations integrally. Theoretically, we establish that the proposed method tightens the generalization error upper bound of self-supervised learning. Empirically, our method achieves state-of-the-art on various benchmarks. Our implementation is available at <https://github.com/ZangZehua/essl>.

Keywords Self-supervised learning · Generalizability · Discriminability · Evolutionary Game Theory · Reinforcement learning

1 Introduction

As Moore’s Law is slowing down Shalf (2020), the Connectionism-based data-driven learning paradigm hits a developmental bottleneck since improving the model performance by simply increasing the neural network parameters is unachievable. Additionally, empowering the model’s capacity by feeding large-scale annotated data to the model Deng et al. (2009) suffers from the excessive annotating workloads. To this end, self-supervised learning (SSL) emerges, which builds a well-designed self-supervised task by leveraging the guidance of a specific knowledge foundation, e.g., contrastive learning is based on the knowledge of multi-view learning assumption Sridharan and Kakade (2008); Tsai et al. (2021); Wang et al. (2022), and masked language/image modeling (M-M) is based on

✉ Wenwen Qiang
E-mail: qiangwenwen@iscas.ac.cn

* Equal Contributions.

¹ Institute of Software, Chinese Academy of Sciences, Beijing, China

² University of Chinese Academy of Sciences, Beijing, China

³ Institute of Automation, Chinese Academy of Science, Beijing, China

⁴ Tsinghua University, Beijing, China

⁵ The Hong Kong University of Science and Technology (Guangzhou), Guangdong, China

⁶ The Hong Kong University of Science and Technology, Hong Kong SAR, China

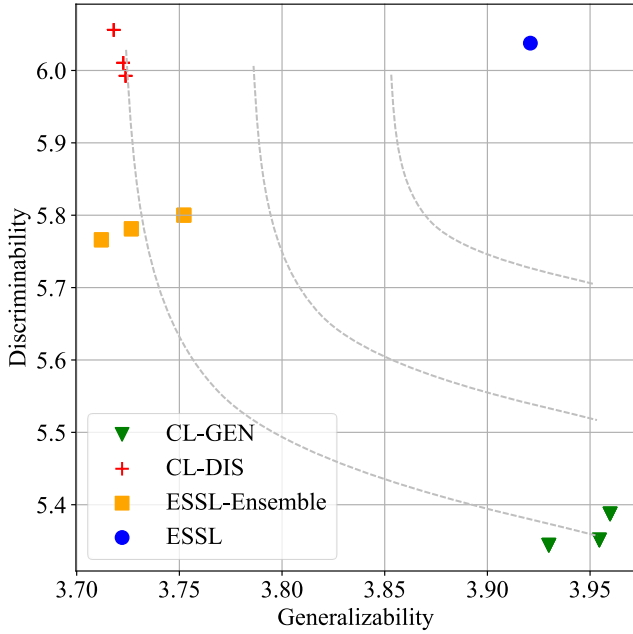


Fig. 1 Negative correlation between generalizability and discriminability. Under the benchmark experimental settings Zbontar et al. (2021), each model uses a particular loss function, e.g., InfoNCE. Each single point presents trade-off point about model’s generalizability and discriminability. See Section 4.1 for the proposed metric of calculating these trade-off points, and the experiments are conducted on four image benchmark datasets, STL-10, CIFAR-10, CIFAR-100, Tiny-ImageNet-200. The dashed lines denote the *potential* correlation linear regression results. As a conclusive observation, ESSL achieves the best joint performance. Results are multiplied by 100 to facilitate observations and avoid ambiguous decimals.

the knowledge of masking-and-reconstruction paradigm Al-Mubaid (2007); Wang et al. (2022); Elkins (2008). SSL further trains the model to achieve the self-supervised task in a completely unsupervised manner. Benefited from the knowledge-guided data-driven learning paradigm, SSL approaches achieve empirical successes in various fields, such as computer vision Hjelm et al. (2019); Bao et al. (2022); He et al. (2022a); Xie et al. (2022) and natural language processing Brown et al. (2020); Devlin et al. (2019).

In the field of SSL, the learned representations are generally regarded as having sufficient *generalizability* (or *transferability*) and *discriminability*, and thus recent works draw limited attention to researching the intrinsic relationship between the designed self-supervised task and the properties of the representations learned by SSL methods. To this end, we investigate such representation properties for various self-supervised tasks. Specifically, we recap the preliminaries of the conventional contrastive learning paradigm: bringing views of the *same* image together and pushing views of the *dif-*

ferent images farther apart. Theoretical proofs Wang et al. (2022) demonstrate that such a learning paradigm encourages the model to learn the discriminative information from the inputs, but the guarantees and evidence provided by Wang et al. (2022); Tian et al. (2020a) further demonstrate that the contrastive objective improves the *domain-specific* discriminability while degenerates the *cross-domain* generalizability of the learned representations. On the contrary, specific SSL methods are inclined to boost the generalizability yet undermine the discriminability of the representations since such approaches only require representations to contain information-decoupled dimensions without the intuition to capture the discriminative information Zbontar et al. (2021); Nakamura (2000); Liu et al. (2022). To profoundly understand the relationship between generalizability and discriminability, we conduct a motivating experiment in Figure 1, which demonstrates the existence of a mutual-exclusion relationship between such properties in SSL.

Does the *juste-milieu* exist to learn a self-supervised representation with strong generalizability and discriminability?

To acquire this aspiration, we rethink the SSL paradigm from EGT perspective. The intuition behind our behavior is that EGT, which benefits from dynamic system modeling, can predict the rational scenarios in the game of two players or populations jointly cooperating and competing. Indirect EGT is commonly applied to examine the interaction mechanism of complex behaviors between *populations* Huck and Oechssler (1999); feng Ji et al. (2019); Bester and Güth (1998). On this basis, we hold an intuition that fusing an SSL model with strong generalizability and another SSL model with strong discriminability in an appropriate manner can derive the desired SSL model with jointly strong generalizability and discriminability. Thus, following EGT, we treat the generalizability and discriminability models as separated populations and assign the corresponding *proportions* Huck and Oechssler (1999); feng Ji et al. (2019); Bester and Güth (1998), i.e., the tuning hyper-parameters. Our objective is to learn appropriate tuning hyper-parameters to boost the joint performance of generalizability and discriminability for SSL models from the perspective of EGT. Following the empirical evolutionary game analysis method Gatenby and Vincent (2003); Bester and Güth (1998); Huck and Oechssler (1999); feng Ji et al. (2019), we define the ideal event and analyze the impacts of tuning hyper-parameters on the dynamic evolution process, thereby deriving appropriate hyper-parameters.

However, performing the complete EGT analysis on each target downstream dataset requires the over-

multitudinous corresponding annotations on the target downstream dataset, which is antithetical to the principle of SSL. Thus, we propose the *Evolutionary game-guided Self-Supervised Learning* approach, namely *ESSL*. Specifically, we first select several annotated datasets as the *representative priori* dataset, which is treated as the available inputs for the EGT analysis to generate instructive tuning hyper-parameters for the fusion of generalizability and discriminability models. Note that the selected datasets do *not* include the target dataset for SSL. The intuition behind such a behaviour is that the trade-off hyper-parameters derived by the EGT analysis on the representative priori datasets hold general effectiveness on various datasets. Yet, directly adopting such firm tuning hyper-parameters cannot consistently boost the joint performance of the learned representations on specific datasets due to the distribution gaps between domains.

To this end, to learn the tuning hyper-parameters for the model fusion by jointly leveraging the guidance of EGT and dynamically adapting the hyper-parameters for the specific target dataset, the proposed ESSL introduces the reinforcement learning (RL) paradigm to enable the fitting of hyper-parameters towards the target dataset. The reason behind the above process is that tuning the hyper-parameters in SSL is a *sequential optimization* process, which adheres to the property of the Markov Decision Process, and the guidance of EGT, i.e., the instructive tuning hyper-parameters generated by the EGT analysis, can be introduced in the *reward function* of RL with ease. Concretely, to comprehensively evaluate the generalizability and discriminability of learned representations, we further provide a novel self-supervised benchmark. Then, we conduct extensive comparisons on the proposed benchmark and the conventional self-supervised benchmark to sufficiently demonstrate the performance superiority of ESSL over candidate self-supervised methods. The significant contributions of this work are four-fold:

- We disclose the mutual-exclusion relationship between the generalizability and discriminability of the representations learned by SSL methods.
- By revisiting the SSL paradigm from the EGT, we propose an innovative SSL method, namely ESSL, which jointly leverages the guidance of the EGT-based analyses and the sequential optimization of the RL approach to empower representations to acquire strong generalizability and discriminability in a balanced manner.
- We provide theoretical analyses to demonstrate that ESSL obtains the tighter generalization error upper

bound on downstream tasks compared with conventional SSL methods.

- Under the self-supervised benchmarks and the proposed benchmark comprehensively measuring generalizability and discriminability, the results prove the effectiveness of ESSL.

2 Related Works

Self-supervised learning. In the unsupervised learning setting, self-supervised contrastive learning [Chuang et al. \(2020\)](#); [Robinson et al. \(2021\)](#); [Elkan and Noto \(2008\)](#); [du Plessis et al. \(2014\)](#); [Sridharan and Kakade \(2008\)](#); [Tsai et al. \(2021\)](#); [Grill et al. \(2020\)](#); [Ermolov et al. \(2021\)](#); [Zbontar et al. \(2021\)](#); [Verma et al. \(2021\)](#); [Xiao et al. \(2021\)](#) has achieved great successes. Specifically, CMC [Tian et al. \(2020a\)](#) and AMDIM [Bachman et al. \(2019\)](#) use contrastive learning on the multi-view data. SimCLR [Chen et al. \(2020\)](#) and MoCo [He et al. \(2020\)](#) employ large batches or memory banks to enlarge the available negative features to learn better representations. SwAV [Caron et al. \(2020\)](#) compares the cluster assignments under different views instead of directly comparing features by using more views. BYOL [Grill et al. \(2020\)](#) and BarlowTwins [Zbontar et al. \(2021\)](#) present a crucial issue: insufficient self-supervision may lead to feature collapse. Benefiting from the well-designed architectures, benchmark SSL methods provide relatively sufficient self-supervision to improve the discriminability performance. Orthogonal to existing methods, we focus on *jointly* improving the generalizability and discriminability of representations, thereby boosting the SSL model performance.

Evolutionary game theory. Considering the bounded rationality and learning mechanisms, EGT [Osborne and Rubinstein \(1994\)](#); [Easley and Kleinberg \(2010\)](#); [Vincent \(1985\)](#) focuses on the decision-making process and solves multiple equilibriums well in biology [Gatenby and Vincent \(2003\)](#) and economics [Bester and Güth \(1998\)](#); [Huck and Oechssler \(1999\)](#); [feng Ji et al. \(2019\)](#). [Gatenby and Vincent \(2003\)](#) uses EGT to frame the tumor-host interface, which elucidates key biological parameters controlling the advance of tumor tissue into the surrounding host tissue. [Bester and Güth \(1998\)](#); [Huck and Oechssler \(1999\)](#) elaborate whether altruism and fair distribution are evolutionarily stable in some instances. [feng Ji et al. \(2019\)](#) applies EGT to examine the interaction mechanism of complex behaviors between local governments and auto manufacturers. Inspired by the practical successes of EGT, the guidance of EGT is expected to be promising in our case.

Reinforcement learning. Recently, RL has achieved great successes in many real-world domains, such as Game AI [Silver et al. \(2017, 2018\)](#); [Vinyals et al. \(2019\)](#), Robotics [Kalashnikov et al. \(2018\)](#); [Hüttenrauch et al. \(2017\)](#), autonomous vehicles [Kiran et al. \(2021\)](#) and so on. Considering the strong sequential decision and policy optimization ability, RL can be regarded as a promising solution for Artificial General Intelligence (AGI). Furthermore, RL has been successfully applied in many Computer Vision (CV) and Natural Language Processing (NLP) tasks for hyper-parameters tuning [Li et al. \(2019\)](#); [Yu et al. \(2018\)](#); [Kumar et al. \(2023\)](#); [Sun et al. \(2023\)](#). Hence, we utilize Proximal Policy Optimization (PPO) [Schulman et al. \(2017\)](#), a well-studied policy gradient RL method, to search the optimal hyper-parameters that balance ESSL’s generalizability and discriminability.

3 Preliminaries

To elaborate on the detailed implementation of ESSL, we recap necessary preliminaries in this section.

3.1 Preliminaries of Self-Supervised Learning

Contrastive learning. Given $X = \{x_i^j | i \in \{1, \dots, N^S\} \text{ and } j \in \{1, \dots, M^V\}\}$, as a multi-view dataset, self-supervised contrastive learning (SS-CL) [Tian et al. \(2020a\)](#); [Chen et al. \(2020\)](#) learns representations by maximizing agreements between the views of the same samples $\{x_i^j, x_i^{j'}\} | j, j' \in \{1, \dots, M^V\}$ and $j \neq j'$ (*positive pairs*), while minimizing agreements between the views of different samples $\{x_i^j, x_{i'}^{j'}\} | j, j' \in \{1, \dots, M^V\}$ and $i \neq i'$ (*negative pairs*). The input data x_i^j is fed into the encoder $f_\theta(\cdot)$ to learn a representation h_i^j , and h_i^j is mapped into a feature z_i^j by a projection head $g_\vartheta(\cdot)$, where θ and ϑ are the network parameters, respectively. $f_\theta(\cdot)$ and $g_\vartheta(\cdot)$ are trained by using a contrastive loss, i.e., InfoNCE [van den Oord et al. \(2018\)](#):

$$\mathcal{L}_{InfoNCE} = -\mathbf{E}_{X_S} \left[\log \frac{d(\{z^+\})}{d(\{z^+\}) + \sum_{k=1}^K d(\{z^-\}_k)} \right], \quad (1)$$

where X_S is *i.i.d.* sampled from X , z^+ denotes a positive pair, z^- denotes a negative pair, and $d(\cdot)$ is a similarity measurement formula, e.g., cosine similarity.

Feature constraint learning. As another representative SSL paradigm, feature constraint learning achieves impressive performance superiority. Specifically, BarlowTwins [Zbontar et al. \(2021\)](#) proposes a

novel self-supervised feature constraint learning loss, which is formulated as follows:

$$\mathcal{L}_{BarlowTwins} = \sum_i (1 - C_{ii})^2 + \epsilon \sum_i \sum_{i' \neq i} C_{ii'}^2, \quad (2)$$

$$C_{ii'} = \frac{\sum_b z_{b,i}^j z_{b,i'}^{j'}}{\sqrt{\sum_b (z_{b,i}^j)^2} \sqrt{\sum_b (z_{b,i'}^{j'})^2}},$$

where ϵ is a positive constant, trading off the importance of the first and second terms of the loss, b indexes batch samples. $C_{ii'}$ is the cross-correlation matrix computed between $z_{b,i}^j$ and $z_{b,i'}^{j'}$ with values in the range between -1 and 1 , and i, i', j, j' are indexes.

3.2 Preliminaries of Evolutionary Game Theory

we recap the necessary introductions of EGT [Weibull \(1997\)](#); [Easley and Kleinberg \(2010\)](#). The key insight of EGT is that many behaviors involve the interaction of multiple organisms in a population, and the success of any one of these organisms depends on how its behavior interacts with that of others. So we evaluate the fitness of an individual organism in the context of the full population in which it lives. In EGT, fitness is like a payoff, which depends on the strategies of the organisms with which it interacts. Evolutionary Stable Strategy (ESS) and Replicator Dynamics (RD) are two critical concepts in EGT [Weibull \(1997\)](#); [Easley and Kleinberg \(2010\)](#), which are introduced as follows.

Evolutionary stable strategy. ESS refers to a strategy in a population that any other strategy cannot replace with higher fitness once it has reached a certain proportion. In other words, an ESS is a strategy that, once it comes to a certain proportion of the population, will remain stable and maintain that proportion through the evolutionary process. This concept was proposed by the biologist John Maynard Smith [Maynard Smith \(1974\)](#) and is now widely used in various evolutionary game models [Gatenby and Vincent \(2003\)](#); [Bester and Güth \(1998\)](#); [Huck and Oechssler \(1999\)](#); [feng Ji et al. \(2019\)](#).

Replicator dynamics. RD is a dynamic model that describes changes in the proportion of strategies in evolutionary games. It is based on the assumption that individuals in each generation will “replicate” their strategy to the next generation with a certain probability, and the payoff of each strategy determines this probability. Using RD, we can predict how the proportion of strategies in evolutionary games will change over time under different initial conditions.

The basic form of the constructed RD equation is:

$$\frac{dx_i(t)}{dt} = [f(s_i, x) - f(x, x)]x_i, \quad (3)$$

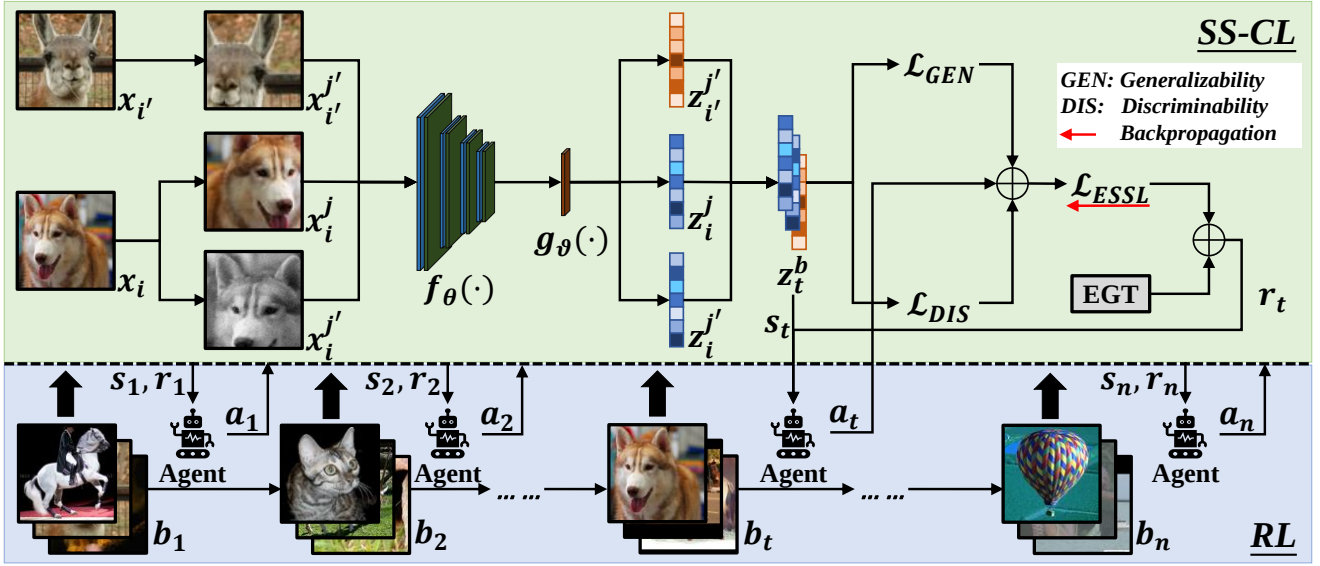


Fig. 2 The framework of ESSL. The module in green denotes the training procedure of the SSL model. The module in blue is the modeled Markov Decision Processes of RL to adjust weights for the *generalizability* loss, i.e., $\mathcal{L}_{GEN}$, and the *discriminability* loss, i.e., $\mathcal{L}_{DIS}$, in a self-paced manner.

where s_i is one strategy in strategy sets, x_i denotes what proportion of individuals in the game group select the strategy s_i at time t , $f(s_i, x)$ denotes the expected payoff of the individual s_i , $f(x, x)$ denotes the expected payoff for the entire population, and $dx(t)/dt$ represents the change rate of individuals that select the strategy s_i per unit time. This equation can be explained as follows: the rate of change in individual strategy proportion is directly proportional to the difference between the current strategy's payoff and the average group payoff. When $f(s_i, x) - f(x, x) > 0$, the proportion of strategies will increase. On the contrary, when $f(s_i, x) - f(x, x) < 0$, its strategy proportion will decrease.

4 Methodology

In this section, we elaborate the proposed ESSL, and the approach framework is illustrated in Figure 2. The detailed nomenclature is provided in **Appendix A**. We emphasize the roadmap of ESSL as follows: 1) we transform the target objective of achieving a SSL representation with strong generalizability and discriminability into the target objective of tuning the hyper-parameters of two decoupled SSL methods and further combining the candidate methods to build the desired SSL method in an ensemble manner; 2) we perform the EGT analysis for the generalizability and discriminability of SSL methods on *certain* representative priori datasets to derive the average hyper-parameters, which

is treated as the guidance for the following RL-based hyper-parameter tuning process; 3) we impose the RL-based approach to jointly leverage the guidance of the EGT analysis and tackle the sequential optimization of the hyper-parameter tuning. Therefore, the derived ESSL can be well generalized into various SSL fields.

4.1 Revisiting the Learning Paradigm of the Evolutionary Game Theory Perspective

As the motivating exploration in Figure 1, we demonstrate the negative correlation between the generalizability and discriminability of a representation learned by SSL. Concretely, to chase the trade-off between the generalizability and discriminability, we intuitively revisit the learning paradigm of self-supervised approaches from the EGT perspective and thus treat the generalizability and discriminability models as two *populations* working against and cooperate. Precisely, the process of tuning the hyper-parameters, i.e., the impacts of the generalizability and discriminability models to the fusion of such models *exactly fits* the case of population evolution in EGT, and the dynamic property of the hyper-parameter tuning process adheres the requirement of EGT, such that we state that adopting EGT to derive the desired trade-off between generalizability and discriminability is theoretically sound.

EGT model assumption. The current benchmark cannot comprehensively evaluate the generalizability and discriminability of a self-supervised model, which

Table 1 Payoff matrix for both players, i.e., the generalizability and discriminability models, and the payoffs of players are different. Refer to **Section 4.1** for the denotations.

		Discriminability model	
		A	U
Generalizability model	A	$\omega_1 \cdot (G_1 - N_2) + (D_2 - N_1);$ $(G_1 - N_2) + \omega_2 \cdot (D_2 - N_1)$	$\omega_1 \cdot G_1 + D_1;$ $G_1 + \omega_2 \cdot D_1$
	U	$\omega_1 \cdot G_2 + D_2; G_2 + \omega_2 \cdot D_2$	0; 0

vary due to different pretext tasks or loss functions, such that following the procedure of EGT, we propose an *integrated benchmark*, and such a benchmark further is the foundational *assumption* of EGT model. Specifically, assuming that supervised learning (SL) achieves lower classification errors than SSL on the downstream tasks, we propose a benchmark to measure the generalizability and discriminability of self-supervised models. Given two datasets, $\mathcal{D}$ and $\mathcal{D}'$, and a self-supervised model M , based on the pre-training on $\mathcal{D}$, G denoted the generalizability of M on $\mathcal{D}'$, and D denotes the discriminability of M on $\mathcal{D}$:

$$G = \frac{1}{ACC_{SL}(\mathcal{D}') - ACC_M(\mathcal{D} \rightarrow \mathcal{D}')}, \quad (4)$$

$$D = \frac{1}{ACC_{SL}(\mathcal{D}) - ACC_M(\mathcal{D} \rightarrow \mathcal{D})},$$

where $ACC_{SL}(\mathcal{D})$ and $ACC_{SL}(\mathcal{D}')$ denote the classification accuracies on $\mathcal{D}$ and $\mathcal{D}'$, respectively. $ACC_M(\mathcal{D} \rightarrow \mathcal{D})$ and $ACC_M(\mathcal{D} \rightarrow \mathcal{D}')$ denote the accuracy on $\mathcal{D}$ and $\mathcal{D}'$ pre-trained on $\mathcal{D}$.

Given a generalizability model M_g and a discriminability model M_d , M_{ens} denotes the ensemble model of M_g and M_d . We denote N_1 as the negative impact of M_g on the discriminability of M_d , and N_2 denotes the negative impact of M_d on the generalizability of M_g :

$$N_1 = ACC_{SL}(\mathcal{D}') - ACC_{M_{ens}}(\mathcal{D} \rightarrow \mathcal{D}'), \quad (5)$$

$$N_2 = ACC_{SSL}(\mathcal{D} \rightarrow \mathcal{D}') - ACC_{M_{ens}}(\mathcal{D} \rightarrow \mathcal{D}'),$$

where $ACC_{M_{ens}}(\mathcal{D} \rightarrow \mathcal{D}')$ denotes the accuracy on $\mathcal{D}'$ pre-trained by the ensemble model on $\mathcal{D}$.

EGT model framework. We develop a *group-double population indirect evolutionary model* [Huck and Oechssler \(1999\)](#); [feng Ji et al. \(2019\)](#); [Bester and Güth \(1998\)](#), where the generalizability and discriminability models present the sufficient and finite populations, respectively, of the chosen pure strategies. The strategy spaces of both players are $S_G = \{A, U\}$ and $S_D = \{A, U\}$, where A denotes “Adopted” and U denotes “Unadopted”. The indirect EGT can be expressed as follows: the combined model selects models based

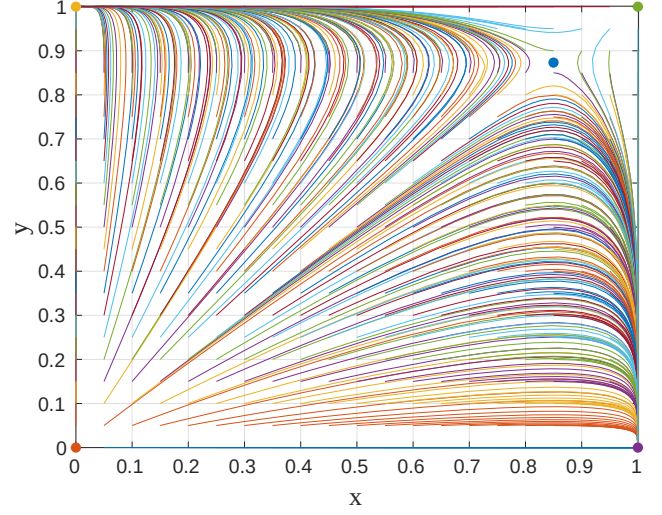


Fig. 3 Phase diagram for the dynamic system. x and y represent the proportion of discriminability and generalizability models, respectively. Every line denotes the evolving strategy of the candidate models under different initializations.

on individual utility maximization. Note that the utility contains both generalizability and discriminability. The system state changes along with the time, and the model utility changes accordingly, resulting in sequential and dynamical adjustments in the strategies chosen by generalizability and discriminability models.

Based on the earlier assumption, Table 1 illustrates a payoff matrix for players across different strategies. In this table, G_1 and D_1 represent the generalizability and discriminability metrics for M_g , and G_2 and D_2 correspond to M_d . Additionally, ω_1 and ω_2 indicate the weights assigned to generalizability and discriminability.

EGT model analysis. Let x ($0 \leq x \leq 1$) represent the proportion of discriminability models, and y ($0 \leq y \leq 1$) represents the proportion of generalizability models which select the A strategy. On the contrary, $1-x$ and $1-y$ represent the proportion of discriminability and generalizability models selecting the U strategy, respectively. $\mathbf{H}$ denotes the income matrix of the gen-

eralizability model, which is defined by

$$\mathbf{H} = \begin{pmatrix} \omega_1 \cdot (G_1 - N_2) + (D_2 - N_1) & \omega_1 \cdot G_1 + D_1 \\ \omega_1 \cdot G_2 + D_2 & 0 \end{pmatrix}. \quad (6)$$

When the generalizability model selects the strategy A , given $\mathbf{e}$, $\mathbf{x}$ and $\mathbf{y}$ denoting the proportion metrics of the chosen strategy, respectively, the expected utility is defined by

$$U_{G1} = \mathbf{eHx}^\top = (1 \ 0) \cdot \begin{pmatrix} \omega_1 \cdot (G_1 - N_2) + (D_2 - N_1) & \omega_1 \cdot G_1 + D_1 \\ \omega_1 \cdot G_2 + D_2 & 0 \end{pmatrix} \begin{pmatrix} x \\ 1 - x \end{pmatrix}, \quad (7)$$

and the corresponding average expected utility of the generalizability models is implemented by

$$\overline{U_G} = \mathbf{yHx}^\top = (y \ 1 - y) \cdot \begin{pmatrix} \omega_1 \cdot (G_1 - N_2) + (D_2 - N_1) & \omega_1 \cdot G_1 + D_1 \\ \omega_1 \cdot G_2 + D_2 & 0 \end{pmatrix} \begin{pmatrix} x \\ 1 - x \end{pmatrix}. \quad (8)$$

According to the Malthusian equation [Malthus \(1872\)](#), the growth of A action of the generalizability model is equivalent to $U_{G1} - \overline{U_G}$. Therefore, the RD equation is defined as follows:

$$F(y) = \frac{dy}{dt} = y[\mathbf{eHx}^\top - \mathbf{yHx}^\top] = y(1 - y) \cdot [\omega_1 \cdot G_1 + D_1 - (D_1 + \omega_1 \cdot G_2 + \omega_1 \cdot N_2 + N_1)x], \quad (9)$$

The first derivative of Equation 9 is detailed by

$$\frac{dF(y)}{dt} = (1 - 2y) \cdot [\omega_1 \cdot G_1 + D_1 - (D_1 + \omega_1 \cdot G_2 + \omega_1 \cdot N_2 + N_1)x]. \quad (10)$$

It can be easily demonstrated that $y = 0$ or $y = 1$, and $x = x^*$ are the numerical solutions of the equation $F(y) = dy/dt = 0$, where

$$x^* = \frac{\omega_1 \cdot G_1 + D_1}{D_1 + \omega_1 \cdot G_2 + \omega_1 \cdot N_2 + N_1}. \quad (11)$$

Under the firm condition that $0 \leq x^* \leq 1$, based on the stability theorem [Friedman \(1991\)](#), we can derive that when $F(x) = 0$ and $F'(x) \leq 0$, the proposition that x is the ESS holds. For any y , if $x = x^*$, $F(y) = 0$, and $F'(y) = 0$ are satisfied, axis y is in a stable state. With the conditions that $x < x^*$, $F'(y)|_{y=0} > 0$, and $F'(y)|_{y=1} < 0$, we indicate that $y = 1$ is the only ESS. Accordingly, if $x > x^*$, $F'(y)|_{y=0} < 0$, and $F'(y)|_{y=1} > 0$ holds, we derive that $y = 0$ is the only ESS.

Adhering to the aforementioned principle, we can obtain the income matrix $\mathbf{K}$ of the discriminability

Table 2 Local stability analyses of equilibrium points.

point	$\det(J)$	$\text{tr}(J)$	Result
(0,0)	+	+	unstable point
(0,1)	+	-	stable point
(1,0)	+	-	stable point
(1,1)	+	+	unstable point
(x^*, y^*)	-	0	saddle point

model and the corresponding RD equation, which are defined as follows:

$$\mathbf{K} = \begin{pmatrix} (G_1 - N_2) + \omega_2 \cdot (D_2 - N_1) & G_1 + \omega_2 \cdot D_1 \\ G_2 + \omega_2 \cdot D_2 & 0 \end{pmatrix}, \quad (12)$$

and

$$F(x) = \frac{dx}{dt} = x[\mathbf{eK}^\top \mathbf{y}^\top - \mathbf{xK}^\top \mathbf{y}^\top] = x(1 - x) \cdot [G_2 + \omega_2 \cdot D_2 - (\omega_2 \cdot D_1 + G_2 + N_2 + \omega_2 \cdot N_1)y]. \quad (13)$$

The first derivative of Equation 13 is detailed as follows:

$$\frac{dF(x)}{dt} = (1 - 2x) \cdot [G_2 + \omega_2 \cdot D_2 - (\omega_2 \cdot D_1 + G_2 + N_2 + \omega_2 \cdot N_1)y], \quad (14)$$

where $x = 0$ or $x = 1$, and $y = y^*$ are the numerical solutions of the equation $F(x) = dx/dt = 0$, and

$$y^* = \frac{G_2 + \omega_2 \cdot D_2}{\omega_2 \cdot D_1 + G_2 + \omega_2 \cdot N_1 + N_2}. \quad (15)$$

With the condition that $0 \leq y^* \leq 1$, we obtain that for any x , if $y = y^*$, $F(x) = 0$, and $F'(x) = 0$ holds, axis x is in a stable state. If $y < y^*$, $F'(x)|_{x=0} > 0$, and $F'(x)|_{x=1} < 0$ are satisfied, $x = 1$ is the only ESS. Accordingly, if $y > y^*$, $F'(x)|_{x=0} < 0$, and $F'(x)|_{x=1} > 0$ are satisfied, $x = 0$ is the only ESS.

According to Equation 9 and Equation 13, we obtain a 2-dim nonlinear dynamic system for the generalizability and the discriminability models, which formulates as follows:

$$\begin{cases} F(x) = \frac{dx}{dt} = x(1 - x)[G_2 + \omega_2 \cdot D_2 \\ \quad - (\omega_2 \cdot D_1 + G_2 + N_2 + \omega_2 \cdot N_1)y] \\ F(y) = \frac{dy}{dt} = y(1 - y)[\omega_1 \cdot G_1 + D_1 \\ \quad - (D_1 + \omega_1 \cdot G_2 + \omega_1 \cdot N_2 + N_1)x] \end{cases}, \quad (16)$$

Local stability analyses of equilibrium points.

Based on Equation 16, the equilibrium points of the dynamic system are (0,0), (0,1), (1,0), (1,1),

(x^*, y^*) . According to Friedman’s proposal, i.e., “describe (co)evolution of population(s) with dynamics defined by differential equations” [Friedman \(1991\)](#), we obtain the Jacobian matrix J of Equation 16 as follows:

$$J = \begin{bmatrix} \frac{\partial F(x)}{\partial x} & \frac{\partial F(x)}{\partial y} \\ \frac{\partial F(y)}{\partial x} & \frac{\partial F(y)}{\partial y} \end{bmatrix} = \begin{bmatrix} (1-2x) \cdot [G_2 + \omega_2 \cdot D_2 - (\omega_2 \cdot D_1 + G_2 + N_2 + \omega_2 \cdot N_1)y] & -x(1-x) \cdot (\omega_2 \cdot D_1 + G_2 + N_2 + \omega_2 \cdot N_1) \\ -y(1-y) \cdot (D_1 + \omega_1 \cdot G_2 + \omega_1 \cdot N_2 + N_1) & (1-2y) \cdot [\omega_1 \cdot G_1 + D_1 - (D_1 + \omega_1 \cdot G_2 + \omega_1 \cdot N_2 + N_1)x] \end{bmatrix}. \quad (17)$$

From Table 2, we can get results that the equilibrium points $(0, 0)$, $(1, 1)$ are unstable points; $(0, 1)$, $(1, 0)$ are stable points; (x^*, y^*) is a saddle point. Figure 3 demonstrates the dynamic system’s phase diagram. Intuitively, we reckon that (x^*, y^*) can be a compromised point for the combined model to obtain the satisfactory trade-off between the generalizability and discriminability. Such a compromised point (x^*, y^*) can be derived by adopting the approach, i.e., Equation 11 and Equation 15. The experiments in Section 6.5 empirically prove the inference behind our behavior.

4.2 Chasing the Trade-off between Generalizability and Discriminability

In practice, we notice that conducting the EGT-based analyses on each downstream task incurs excessive workloads, which require sizeable human annotation and computational costs. Additionally, solely generalizing the hyper-parameters derived by the EGT analysis on several representative downstream tasks cannot achieve desired model performance improvement. The corresponding empirical proofs are provided in Section 6. Therefore, we first perform the EGT analysis on the representative prior datasets to derive the average hyper-parameters, serving as the guidance of EGT. Then, we propose to leverage a practical RL approach to adjust the hyper-parameters dynamically within the guidance of EGT. Refer to Section 6 for the setting of representative priori datasets in practice.

Performing the reinforcement learning. Given a multi-view dataset X , we propose to learn a self-supervised representation with the desired trade-off between the generalizability and discriminability. In practice, we intuitively leverage an *ensemble* manner, which

can be formally described as follows:

$$\mathcal{L}_{ESSL} = \alpha \mathcal{L}_{GEN} + \beta \mathcal{L}_{DIS}, \quad (18)$$

where α and β are the hyper-parameters controlling the balance between the $\mathcal{L}_{DIS}$ and $\mathcal{L}_{GEN}$, i.e., the losses present the generalizability and discriminability, respectively. In the implementation, we adopt $\mathcal{L}_{InfoNCE}$ as $\mathcal{L}_{GEN}$ and $\mathcal{L}_{BarlowTwins}$ as $\mathcal{L}_{DIS}$. To chase the best trade-off, we employ an effective RL method, PPO [Schulman et al. \(2017\)](#). At each time step t , the RL agent receives a state s_t , which contains information the agent can observe. In our formulation, the state is formulated by

$$s_t = F_s(z_t^b), \quad (19)$$

where $F_s(\cdot)$ is the mean function, and z_t^b is the features of batch b . Under state s_t , the agent takes an action $a_t = (a_t^\alpha, a_t^\beta)$. Then, the parameters α and β at the time step t can be calculated as:

$$\alpha_t = V + a_t^\alpha, \quad \beta_t = V + a_t^\beta, \quad (20)$$

where V is the hyper-parameters.

Employing the guidance of EGT. Given a set of representative datasets $S_d = \{\mathcal{D}^1, \mathcal{D}^2, \dots, \mathcal{D}^n\}$, we primarily derive the average result of the SL accuracies on such datasets, and refer to [Appendix C.2](#) for the detailed results. Accordingly, $\bar{G}$ denotes the average SSL accuracy measuring the generalizability, $\bar{D}$ denotes the average SSL accuracy measuring the discriminability, $\bar{N}_1$ denotes the average effect on the discriminability model by the generalizability model, and $\bar{N}_2$ denotes the average effect on the generalizability model by the discriminability model.

In our framework, the *reward function* is composed of two key components:

$$r_t = G(\alpha_t, \beta_t, x^*, y^*) + \xi \frac{1}{|\mathcal{L}_{ESSL_t} - \phi \mathcal{L}_{ESSL_{t-1}}|}, \quad (21)$$

The first term is defined by $G(\cdot, \cdot, \cdot, \cdot)$, which represents the cosine similarity function. Here, (x^*, y^*) denotes the *a priori* balanced point between generalizability and discriminability, as determined by EGT. The hyper-parameter ξ is utilized to balance different reward signals, while ϕ is another hyperparameter designed to the exploration term. Overall, the first term of the reward function allows RL to dynamically pursue a balance between generalizability and discriminability based on prior knowledge acquired from general datasets. However, relying solely on this general prior knowledge often falls short in uncovering the latent characteristics specific to a dataset. Hence, we have devised a second term in the reward function, aimed at encouraging agents to explore the most suitable balance point for the specific dataset.

5 Theoretical Analyses on Generalization Error Bounds

Benchmarking models on classification tasks is a standard evaluation approach for self-supervised approaches. Therefore, we present the generalization error bounds of the proposed ESSL on classification tasks. The benchmarking is implemented by training a softmax classifier for minimizing the traditional cross-entropy loss [Zhang and Sabuncu \(2018\)](#) as follows:

$$\mathcal{L}_{softmax}(f; T) = \inf_W \mathcal{L}_{cross-entropy}(W \cdot f; T), \quad (22)$$

where f denotes the fixed encoder pre-trained by self-supervised methods, W is the linear-probing classifier, and T is the label on the downstream task. Holding Equation 22, we present the generalization error for the feature representation $f(X)$ learned from X by

$$\mathcal{L}_{GE}^f = \mathbf{E}_X \left[\inf_W \mathcal{L}_{cross-entropy}(W \cdot f(X); T) \right]. \quad (23)$$

For the encoder f , we can further bridge the generalization error $\mathcal{L}_{GE}^f$ with the conventional contrastive learning objective $\mathcal{L}_{InfoNCE}^f$, defined in Equation 1, by introducing the *Rademacher complexity* [Saunshi et al. \(2019\)](#); [Qiang et al. \(2022\)](#) as follows:

Theorem 1 Suppose f^* is a sufficiently trained neural network by following the proposed ESSL objective and following Equation 18, f^* is acquired as follows:

$$f^* = \arg \min_f \alpha \mathcal{L}_{GEN}^f + \beta \mathcal{L}_{DIS}^f. \quad (24)$$

Given the condition that α is frozen, with the probability at least $1 - \delta$, we hold that the upper bound of generalization error can be approximated by

$$\begin{aligned} \left| \mathcal{L}_{GE}^{f^*} - \mathcal{L}_{InfoNCE}^{f^*} \right| \leq & O \left(\frac{\sqrt{1 + 1/N^{batch}} \cdot \mathcal{R}_{\mathcal{H}}(\beta)}{N^S} \right. \\ & \left. + \sqrt{\frac{\log(1/\delta) \cdot \log^2(N^S)}{N^S}} \right), \end{aligned} \quad (25)$$

where N^S denotes the number of training samples, and N^{batch} denotes the size of a mini-batch. $\mathcal{R}_{\mathcal{H}}(\beta)$ presents the Rademacher complexity, which establishes a constant property that $\mathcal{R}_{\mathcal{H}}(\beta)$ is monotonically decreasing w.r.t. β .

We provide proof to attest to the correctness and integrity of the proposed theoretical analyses. To demonstrate the validity of the corresponding Theorem, we provide a Lemma as follows:

Lemma 1 Let $f \in \mathcal{H}$, where $\mathcal{H}$ is a restricted function hypothesis space, and f^* be a neural network sufficiently trained by following Equation 24. When α is fixed, with the probability at least $1 - \delta$ over the training data $X = \{X_1, X_2, \dots, X_{N^S}\}$, the following inequity holds:

$$\begin{aligned} \mathcal{L}_{GE}^{f^*} \leq \mathcal{L}_{InfoNCE}^{f^*} + & O \left(\frac{\sqrt{1 + 1/N^S} \cdot \mathcal{R}_{\mathcal{H}}(\beta)}{N^S} \right. \\ & \left. + \sqrt{\frac{\log(1/\delta) \cdot \log^2(N^S)}{N^S}} \right). \end{aligned} \quad (26)$$

The Rademacher Complexity is defined as

$$\mathcal{R}_{\mathcal{H}}(\beta) = \mathbf{E}_{\sigma \in \{\pm 1\}^{3 \cdot d \cdot N^S}} \left[\sup_{f \in \mathcal{H}_{\beta}} \langle \sigma, f \rangle \right], \quad (27)$$

where d denotes the dimensionality of the representations learned by f , and the restricted function hypothesis set $\mathcal{H}_{\beta}$ corresponding to the coefficient β of the proposed ESSL is determined by

$$\mathcal{H}_{\beta} = \{f | f \in \mathcal{H}, \text{ and } \mathcal{R}_1(f) \leq 4/\beta\}. \quad (28)$$

Holding Lemma 1, we can thus derive

Corollary 1 Following the conventional contrastive principle, for f^* , with the probability at least $1 - \delta$, we have that

$$\begin{aligned} \left| \mathcal{L}_{GE}^{f^*} - \mathcal{L}_{InfoNCE}^{f^*} \right| \leq & O \left(\frac{\sqrt{1 + 1/N^{neg}} \cdot \mathcal{R}_{\mathcal{H}}(\beta)}{N^S} \right. \\ & \left. + \sqrt{\frac{\log(1/\delta) \cdot \log^2(N^S)}{N^S}} \right), \end{aligned} \quad (29)$$

where N^S is the total number of training samples, and N^{neg} is the size of negative pairs, which is greatly close to N^S for the integrated dataset due to the instance-level contrastive principle. $\mathcal{R}_{\mathcal{H}}(\beta)$ is the Rademacher complexity, and $\mathcal{R}_{\mathcal{H}}(\beta)$ is monotonically decreasing w.r.t. β .

Proof Holding Equation 22 and Equation 24, for the generalization error $\mathcal{L}_{GE}^f$, we have that:

$$\begin{aligned}
& \mathcal{L}_{GE}^f \\
&= \mathbf{E}_X \left[\inf_W \mathcal{L}_{cross-entropy}(W \cdot f, T) \right] \\
&= \mathbf{E}_X \left[-\log \frac{e^{f(X)^T u_{c^{pos}}}}{e^{f(X)^T u_{c^{pos}}} + \sum e^{f(X)^T u_{c^{neg}}}} \right] \\
&= \mathbf{E}_X \left[-\log \left(e^{f(X)^T u_{c^{pos}}} \right) \right. \\
&\quad \left. + \log \left(e^{f(X)^T u_{c^{pos}}} + \sum e^{f(X)^T u_{c^{neg}}} \right) \right] \\
&= \mathbf{E}_X \left[\log \left(e^{f(X)^T \mathbf{E}_{X^{pos}}[f(X^{pos})]} \right) \right. \\
&\quad \left. + N^S \mathbf{E} \left[e^{f(X)^T E_{X^{neg}}[f(X^{neg})]} \right] \right] \\
&\quad \left. - \log \left(e^{f(X)^T \mathbf{E}_{X^{pos}}[f(X^{pos})]} \right) \right] \\
&\geq \mathbf{E}_X \left[\log \left(e^{f(X)^T \mathbf{E}_{X^{pos}}[f(X^{pos})]} \right) \right. \\
&\quad \left. + N^S \mathbf{E}_{X^{neg}} \left[e^{f(X)^T \mathbf{E}_{X^{neg}}[f(X^{neg})]} \right] \right] \\
&\quad \left. - \log \left(e^{f(X)^T \mathbf{E}_{X^{pos}}[f(X^{pos})]} \right) \right] \\
&= \mathcal{L}_{InfoNCE}^f.
\end{aligned} \tag{30}$$

Then, we can intuitively obtain

$$\begin{aligned}
& \mathcal{L}_{GE}^{f^*} - \mathcal{L}_{InfoNCE}^f \\
&\leq \mathcal{L}_{InfoNCE}^{f^*} - \mathcal{L}_{InfoNCE}^f \\
&\leq O \left(\frac{\sqrt{1 + 1/N^{neg}} \cdot \mathcal{R}_{\mathcal{H}}(\beta)}{N^S} \right. \\
&\quad \left. + \sqrt{\frac{\log(1/\delta) \cdot \log^2(N^S)}{N^S}} \right).
\end{aligned} \tag{31}$$

Therefore, Corollary 1 holds.

Concretely, we conduct Corollary 1 in the practical scenario, where only a mini-batch of training samples is available for the neural network encoder f in a gradient back-propagating step, such that a tractable surrogate variable is introduced in Equation 29, i.e., N^{neg} is converted into N^{batch} . Therefore, we theoretically substantiate the correctness and integrity of Theorem 1.

Considering Equation 25, we observe that the upper bound of generalization error gradually decreases with the increase of the training sample size N^S .

Table 3 Top-1 accuracies (in %) with linear probing on STL10, dubbed by *S10*, CIFAR10, dubbed by *C10*, CIFAR100, dubbed by *C100*, and Tiny-ImageNet, dubbed by *Tiny*. The compared methods are pre-trained and then evaluated on conventional benchmark datasets in an iterative manner.

PT	Model	Eval			
		S10	C10	C100	Tiny
S10	SimCLR	76.4	56.0	28.0	24.6
	MoCo-v2	80.1	67.3	37.0	27.8
	BarlowTwins	85.5	70.0	41.5	29.8
	BYOL	83.7	58.7	32.2	29.2
	ESSL(Ours)	86.1	73.4	45.4	33.7
C10	SimCLR	62.2	73.4	42.1	23.8
	MoCo-v2	65.3	65.7	37.2	21.8
	BarlowTwins	73.1	83.0	43.8	23.1
	BYOL	74.8	50.6	24.0	21.3
	ESSL(Ours)	72.0	83.6	44.1	22.8
C100	SimCLR	62.2	73.4	42.0	25.6
	MoCo-v2	62.0	60.4	43.2	19.7
	BarlowTwins	69.5	73.8	51.1	23.7
	BYOL	66.5	50.7	26.2	23.5
	ESSL(Ours)	70.0	75.1	51.1	23.2
Tiny	SimCLR	69.8	60.4	35.6	31.1
	MoCo-v2	67.3	57.6	27.6	21.3
	BarlowTwins	76.0	70.0	39.9	32.7
	BYOL	73.0	58.0	31.6	32.3
	ESSL(Ours)	76.8	73.4	40.5	32.0

Such an observation is consistent with the conventional SL methods. On top of this, for the relatively non-trivial mini-batch size N^{batch} , the first term of generalization error $\sqrt{1 + 1/N^{batch}} \cdot \mathcal{R}_{\mathcal{H}}(\beta)$ is effectively reduced. Thus the generalization error upper bound is tightened, consistent with the empirical induction in SS-CL approaches. Connecting the proposed approach with the generalization error, we attest that the enlargement of the coefficient β of ESSL directly results in the reduction of the Rademacher complexity $\mathcal{R}_{\mathcal{H}}$. Concretely, due to the non-triviality of β , we can theoretically substantiate that performing SSL in the ESSL manner derives a relatively tighter generalization error upper bound, such that ESSL holds stronger generalizability than the conventional methods.

6 Experiments

In this section, we conduct a series of experiments to evaluate the proposed ESSL from multiple perspectives comprehensively. In this regard, we bootstrap the evaluation by answering the following research questions:

1. Is ESSL consistently competitive with other baselines in conventional benchmark datasets (Section: 6.1) and large-scale benchmark datasets (Section: 6.2)?
2. How about the generalizability and discriminability of self-supervised methods with different loss functions on various datasets (Section: 6.3)?
3. How can we combine the two loss functions as best as possible from the network structure perspective (Section: 6.4)?
4. How much performance improvement does the imposed EGT and RL bring (Section: 6.5)?
5. How does RL shift the balance achieved by the EGT analysis (Section: 6.6)?

Experimental schemes. In practice, the generalizability is measured by conducting the pre-training and testing on *heterogeneous* datasets, which is determined as the generalization task. In contrast, the discriminability is measured by conducting the pre-training and testing on the *homogeneous* dataset, which is treated as the discrimination task.

Note that the representative priori datasets include STL10 Coates et al. (2011), CIFAR10 Krizhevsky (2009), CIFAR100 Krizhevsky (2009), and Tiny-ImageNet-200 Krizhevsky (2009) for the detailed implementation of the EGT analysis.

Benchmark baselines. The compared baselines include SimCLR Chen et al. (2020), MoCo He et al. (2020), CMC Tian et al. (2020a), CPC Hénaff (2020), InfoMin Aug Tian et al. (2020b), MoCo-v2 He et al. (2020), BarlowTwins Zbontar et al. (2021), BYOL Grill et al. (2020), SSL-HSIC Li et al. (2021b), MAE He et al. (2022b), CAE Chen et al. (2023), and VICReg Bardes et al. (2022). Note that, for conventional benchmark experiments, the compared methods consistently leverage ResNet-18 He et al. (2016) as the backbone networks; for large-scale benchmark experiments, most candidate methods, including the proposed ESSL, adopt standard ResNet-50 He et al. (2016) as the backbone network, while MAE and CAE adopt ViT-B-32 Dosovitskiy et al. (2021) as the backbone network. In detail, ResNet-50 possesses 25M parameters, and ViT-B possesses 86M parameters. On top of this, MAE and CAE impose 1600 epochs during training, while the rest of the methods impose 1000 epochs.

6.1 Performance of Classification Task on Conventional Benchmark Datasets

Experimental settings. Through an image classification task, we evaluate the generalizability and discriminability of ESSL on four conventional benchmark

datasets, i.e., STL10 Coates et al. (2011), CIFAR10 Krizhevsky (2009), CIFAR100 Krizhevsky (2009), and Tiny-ImageNet-200 (or Tiny-ImageNet) Krizhevsky (2009). Note that the computation of generalizability and discriminability requires the detailed implementation of baselines, such that we compare the proposed ESSL with the most representative benchmark methods in the field of SSL to the joint evaluation of generalizability and discriminability. We pre-train ResNet-18 as the backbone and three linear layers as the projection head for 500 epochs on each dataset for each method as the pre-training phase. Then, we use the fixed backbone pre-trained above to train a linear layer on every dataset as the linear probing phase. More training details are in Appendix C.3. The top-1 classification accuracy of each method on each dataset is reported in Table 3. All the experimental results are achieved using the code in Appendix C.1, and we collect the results of 10 trials and derive the average results for comparisons.

Generally, our proposed ESSL achieves state-of-the-art performance to both generalizability and discriminability. In this regard, ESSL can simultaneously learn the representations with better generalizability and discriminability. To better visualize the generalizability and discriminability, we follow the proposed metric in Section 4.1, and the results are shown in Figure 4. We multiplied the experimental results’ numerical values by 100 to facilitate observations and avoid ambiguous decimals. These results are consistent with the motivating experiments, and ESSL can still obtain the best overall performance.

6.2 Performance of Classification Task on Large-Scale Benchmark Dataset

Discriminability evaluation on ImageNet. To further show the method’s performance, we evaluate ESSL on the large-scale benchmark dataset, ImageNet-1000 (or ImageNet) Deng et al. (2009), by imposing 1000 epochs pre-training and 100 epochs linear probing. The backbone network consists of benchmark baselines, i.e., a ResNet-50 is treated as the backbone, and three linear layers are treated as the projection head. From the discriminability perspective, we report our experimental results in Table 4. More implementation details can be found in Appendix C. Note that all the experimental baseline results are directly referred from the published papers. As shown in Table 4, our proposed ESSL outperforms all the baselines on the image classification task. That is to say, ESSL can learn the representations with strong discriminability on the large-scale benchmark dataset.

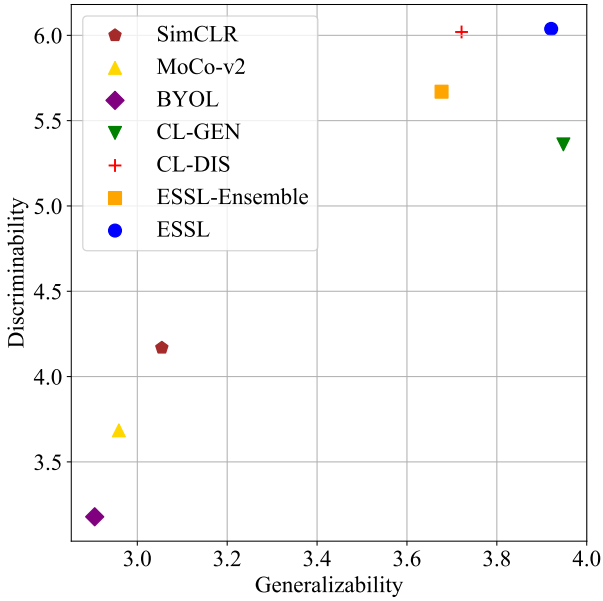


Fig. 4 Evaluating generalizability and discriminability on the proposed EGT-based benchmark.

Generalizability evaluation with imposing the transfer learning tasks on ImageNet. We evaluate the learned representation of ESSL on other classification datasets to determine whether the representations, learned on ImageNet, can be generalized into heterogeneous domains, which demonstrates the generalizability of ESSL. We perform the linear evaluation and fine-tuning on the same sets of classification tasks used in [Chen et al. \(2020\)](#); [Kornblith et al. \(2019\)](#) by carefully following their evaluation protocol. Performance is reported in Table 5. ESSL outperforms SimCLR on CIFAR10 and CIFAR100, and BYOL on CIFAR10. We further report the results of transfer training on STL10, while the benchmark methods do not provide the corresponding results. The reason behind that we select SimCLR and BYOL as the compared baselines is that only SimCLR and BYOL report the corresponding transfer learning results on the conducted benchmark datasets, including ImageNet, STL10, CIFAR10, CIFAR100, and Tiny-ImageNet. The transfer learning results comprehensively demonstrate the strong generalizability achieved by ESSL on the large-scale benchmark dataset.

6.3 Generalizability and Discriminability of Introducing Different Loss Functions on Different Datasets

We conduct experiments to illustrate this phenomenon to clarify the generalizability and discriminability of

introducing two heterogeneous loss functions, i.e., in-fNCE and Barlow Twins. Following the settings in Appendix C, we denote CL-GEN as the model with loss function $\mathcal{L}_{GEN}$ and CL-DIS as the model with loss function $\mathcal{L}_{DIS}$. Note that the only difference between CL-GEN and CL-DIS is the loss function implemented. Then, we evaluate their generalizability and discriminability using the same training strategy as Section 6.1 on the same datasets, STL-10, CIFAR-10, CIFAR-100 and Tiny-ImageNet-200. The results are reported in Figure 4.

From the generalizability perspective, CL-GEN can achieve better results than CL-DIS on these non-pre-trained datasets. In contrast, CL-DIS outperforms CL-GEN on the pre-trained datasets. CL-GEN has better generalizability than CL-DIS, but CL-DIS has better discriminability than CL-GEN. Without loss of generalization, $\mathcal{L}_{GEN}$ can improve the model in the generalizability perspective, while $\mathcal{L}_{DIS}$ can improve the model in the discriminability perspective.

6.4 Ablation: Influence of the Number of Projection Head

When we combine the advantages of the two loss functions, i.e., $\mathcal{L}_{GEN}$ and $\mathcal{L}_{DIS}$, we intend to avoid the performance degeneration incurred by the variance of the number of projection heads. To this end, we conduct experiments to choose the best number of projection heads. We follow the settings in Section 6.1 and compare the generalizability and discriminability of two pre-trained models, where the model, with only one projection head, shares the projection head with two loss functions. The model, with two projection heads, back-propagates the gradients of the two loss functions to the backbone network, respectively. The experimental results are reported in Table 6, where we can observe a significant performance gap between the model with one projection head and the model with two projection heads. Following the experimental results, we implement the one projection head structure as the network structures of our proposed ESSL and the corresponding ablation models.

6.5 Ablation: Influence of EGT Module

Does the conducted EGT module really improve ESSL’s generalizability and discriminability? To answer this question, we set up an ablation study to analyze the influence of the proposed EGT analysis upon the generalizability and discriminability of ESSL model. Experimental settings are the same as in Section 6.1 on

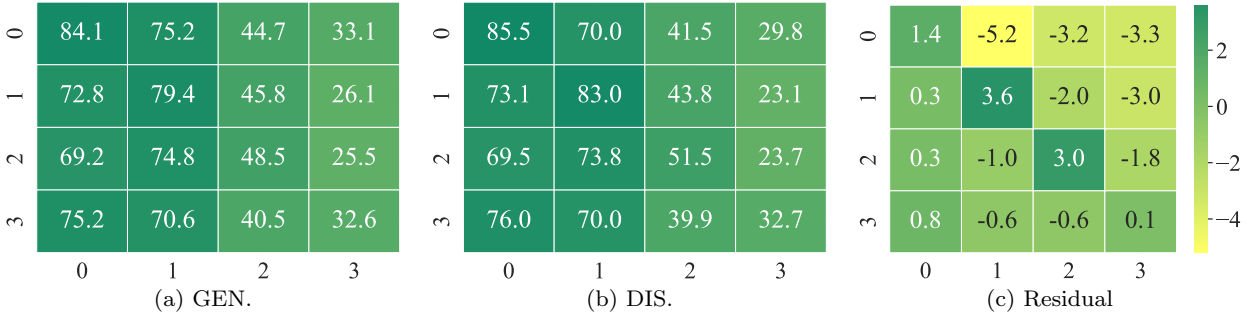


Fig. 5 Generalizability and discriminability evaluations with different self-supervised losses. (a) shows the results pre-trained with $\mathcal{L}_{GEN}$ for stronger generalizability. (b) shows the results pre-trained $\mathcal{L}_{DIS}$ for stronger discriminability. (c) demonstrates the residuals by subtracting (a) from (b). All the results are under the same training paradigm.

Table 4 Linear probing results on the large-scale benchmark.

Methods	Architecture	Param	epoch	Top-1	Top 5
SimCLR	ResNet-50(1x)	25M	1000	69.3	89.0
MoCo	ResNet-50(1x)	25M	1000	71.1	-
CMC	ResNet-50(1x)	25M	1000	66.2	87.0
CPC	ResNet-50(1x)	25M	1000	63.8	85.3
InfoMin Aug	ResNet-50(1x)	25M	1000	73.0	91.1
BarlowTwins	ResNet-50(1x)	25M	1000	73.5	91.0
BYOL	ResNet-50(1x)	25M	1000	74.3	91.6
SSL-HSIC	ResNet-50(1x)	25M	1000	72.2	90.7
VICReg	ResNet-50(1x)	25M	1000	73.2	91.1
MAE	ViT-B-32	86M	1600	68.6	-
CAE	ViT-B-32	86M	1600	71.4	-
ESSL(Ours)	ResNet-50(1x)	25M	1000	74.5	91.9

Table 5 Transfer learning results on large-scale benchmark with the standard ResNet-50 architecture.

Methods	STL10	CIFAR10	CIFAR100
Supervised	-	93.6	78.3
SimCLR	-	90.6	71.6
BYOL	-	91.3	78.4
ESSL(Ours)	97.4	91.7	78.0

the same datasets, STL-10, CIFAR-10, CIFAR-100 and Tiny-ImageNet-200. The experimental results are reported in Table 7. In details, ESSL-Ensemble denotes ESSL imposing the plain ensemble strategy to combine the heterogeneous loss functions, e.g., $\alpha = 1$ and $\beta = 1$. ESSL-EGT denotes ESSL imposing the EGT static strategy (in our implementation is $\alpha = 0.85, \beta = 0.87$) to combine the heterogeneous loss functions. ESSL-EGT+RL denotes ESSL imposing the RL-guided EGT dynamic strategy to combine the heterogeneous loss functions, as shown in Figure 6, which is the final version of our proposed ESSL.

By observing the performance gap between ESSL-Ensemble and the ablation models, e.g., CL-GEN and CL-DIS, we find that arbitrarily combining heterogeneous loss functions may *degenerates* the performance of ESSL to both the generalizability and discriminability, such that introducing the EGT analysis-based guidance is crucial in this regard.

By observing the performance gap between ESSL-EGT and ESSL-Ensemble, we derive that when pre-training ESSL with the guidance of the EGT analysis, the generalizability and discriminability of ESSL are widely improved. We argue that when chasing better generalizability and discriminability, the position of the trade-off point to the EGT analysis is pivotal, and simply changing the position of the trade-off point or changing the direction of the trade-off point from the origin may incur undesired degeneration on the generalizability and discriminability of ESSL.

We further study the significance of introducing the RL-based optimization of ESSL. Table 7 shows the results of pre-training the model with the guidance of the static EGT trade-off point, i.e., ESSL-EGT, and pre-

Table 6 Results on conventional benchmark datasets with different projection head numbers.

Pre-training	Head Num	Evaluation			
		STL10	CIFAR10	CIFAR100	Tiny-ImageNet
STL10	1	86.0	73.3	41.4	29.7
	2	83.6	72.4	41.7	27.5
CIFAR10	1	71.1	82.0	44.0	22.4
	2	62.8	67.3	36.9	16.3
CIFAR100	1	68.3	73.0	49.9	21.7
	2	63.5	67.4	35.9	15.7
Tiny-ImageNet	1	76.4	72.8	43.1	35.1
	2	69.2	63.7	33.5	21.3

Table 7 Results on conventional benchmark datasets with the proposed ESSL and various ablation models, where CL-GEN denotes the model pre-trained only with $\mathcal{L}_{GEN}$. CL-DIS denotes the model pre-trained only with $\mathcal{L}_{DIS}$. ESSL-Ensemble denotes a model pre-trained with the simple addition of $\mathcal{L}_{GEN}$ and $\mathcal{L}_{DIS}$. ESSL-EGT denotes a model pre-trained with the addition of $\mathcal{L}_{GEN}$ and $\mathcal{L}_{DIS}$ guided by the static EGT saddle point. ESSL-EGT+RL denotes a model pre-trained with the addition of $\mathcal{L}_{GEN}$ and $\mathcal{L}_{DIS}$ guided by RL-based dynamic trade-off points.

Pre-training	Model	Evaluation			
		STL10	CIFAR10	CIFAR100	Tiny-ImageNet
STL10	CL-GEN	84.1	75.2	44.7	33.1
	CL-DIS	85.5	70.0	41.5	29.8
	ESSL-Ensemble	84.0	74.8	44.3	27.5
	ESSL-EGT	86.0	73.3	41.4	29.7
	ESSL-EGT+RL	86.1	73.4	45.4	33.7
CIFAR10	CL-GEN	72.8	79.4	45.8	26.1
	CL-DIS	73.1	83.0	43.8	23.1
	ESSL-Ensemble	71.2	79.0	41.9	22.1
	ESSL-EGT	71.1	82.0	44.0	22.4
	ESSL-EGT+RL	72.0	83.6	44.1	22.8
CIFAR100	CL-GEN	69.2	74.8	48.5	25.5
	CL-DIS	69.5	73.8	51.1	23.7
	ESSL-Ensemble	63.5	71.4	47.1	25.2
	ESSL-EGT	68.3	73.0	49.9	21.7
	ESSL-EGT+RL	70.0	75.1	51.1	23.2
Tiny-ImageNet	CL-GEN	75.2	70.6	40.5	32.6
	CL-DIS	76.0	70.0	39.9	32.7
	ESSL-Ensemble	74.1	70.0	40.2	30.1
	ESSL-EGT	76.4	72.8	43.1	35.1
	ESSL-EGT+RL	76.8	73.4	40.5	32.0

training the model with the guidance of the RL-based dynamic trade-off points, i.e., ESSL-EGT+RL. We observe that ESSL-EGT+RL consistently outperforms ESSL-EGT on the generalizability and discriminability comparisons. We can conclude that the proposed RL-based optimization further improves the generalizability and discriminability of the ESSL model. The RL-based optimization primarily promotes the generalizability of ESSL by a significant margin.

6.6 The Improvement Derived by Introducing Reinforcement Learning upon the EGT Analysis

We conduct visualization experiments to understand further the improvement derived by introducing the RL-based optimization upon ESSL. In detail, the proposed EGT analysis provides a significant initial trade-off point of the generalizability and discriminability for ESSL, and RL further empowers ESSL to dynamically adapt the trade-off point to the training domain, i.e., the training dataset, which is achieved by leveraging the Markov Decision Processes.

For the ease of observation of the process and analysis towards the experimental results, we sample 2,000 points uniformly from all 194,560 points records of the hyper-parameters, i.e., α and β , derived by the RL module of ESSL during the pre-training under the same settings as Section 6.1 on STL10. The visualization is shown in Figure 6, where the horizontal axis is α , i.e., *Alpha*, and the vertical axis is β , i.e., *Beta*, and the points produced in the early pre-training stage are inclined to purple. The points produced in the late pre-training stage are inclined to yellow. Specifically, the early pre-training stage denotes that the network has just been initialized or pre-trained with a few epochs, such as 50 or 100. In contrast, the late pre-training stage denotes that the network has been pre-trained after several epochs, such as 900 or 1000.

From the visualization results, we can observe that the points in the early stage of pre-training are distributed in various parts of the area. However, as pre-training is processed, the points come closer to the saddle point derived by the EGT analysis. At the end of pre-training, RL-based ESSL performs dynamic modifications based on the fixed saddle point derived by the EGT analysis, thereby achieving the ultimate trade-off point. We argue that on the one hand, the EGT analysis can successfully guide ESSL to chase the trade-off between generalizability and discriminability, and the guidance of the fixed saddle point, derived by the EGT analysis, is solid; on the other hand, introducing the RL-based optimization can indeed dynamically update the trade-off point within leveraging the guidance of the EGT analysis in a self-paced manner. As shown in Table 7, the empirical proofs demonstrate that the generalizability and discriminability of the ultimate ESSL model, i.e., ESSL+EGT+RL, widely outperform that of the ablation mode, i.e., ESSL+EGT.

7 Limitations and Broader Impacts

The optimization potential exists for ESSL’s training process for complexity reduction. Accordingly, acquiring the guidance of EGT requires specific representative downstream datasets with annotations, such that performing EGT analysis in an unsupervised or semi-supervised manner is promising scientific research for the community.

8 Conclusion

We disclose a mutual-exclusion relationship between the generalizability and discriminability of learned representations by conducting empirical explorations,

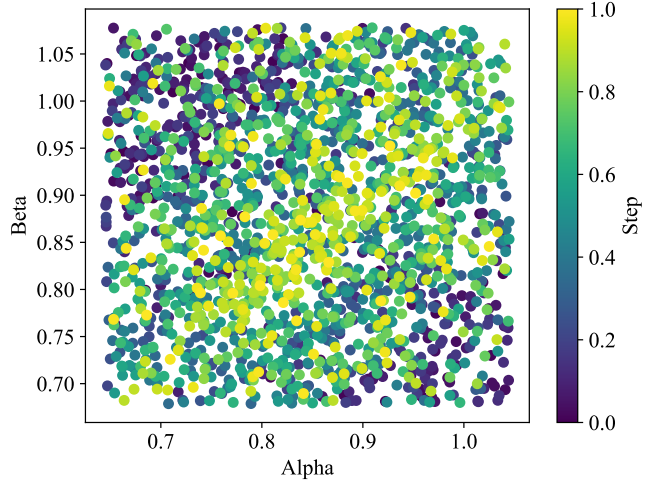


Fig. 6 Trade-off points at different stages of pre-training. As the pre-training processes, the points produced by ESSL gradually change from purple to yellow. Note that Alpha and Beta in the figure represent α and β in Equation 18, respectively.

which undermines the development of state-of-the-art self-supervised methods. To this end, we revisit the SSL paradigm from the EGT perspective and explore achieving a desired trade-off between generalizability and discriminability. Imposing the guidance of the EGT and the RL, we propose a novel SSL method called ESSL. To comprehensively evaluate the representation quality, we provide a novel self-supervised benchmark. The theoretical evidence demonstrates that ESSL tightens the generalization error bound of SSL. The experimental results prove that our method achieves state-of-the-art on multiple benchmarks.

Data Availability

The datasets used in this paper are publicly available. Their names and links are as follows:

1. STL10 <https://cs.stanford.edu/~acoates/stl10/>.
2. CIFAR10 <https://www.cs.toronto.edu/~kriz/cifar.html>.
3. CIFAR100 <https://cs.stanford.edu/~acoates/stl10/>.
4. Tiny-ImageNet-200 <https://www.kaggle.com/c/tiny-imagenet>.
5. ImageNet <https://image-net.org/index.php>.

Acknowledgments

The authors would like to thank the anonymous reviewers for their valuable comments. This work is supported by the Fundamental Research Program, Grant

No. JCKY2022130C020, and the Youth Innovation Promotion Association CAS.

References

- Al-Mubaid H (2007) A learning-classification based approach for word prediction. *Int Arab J Inf Technol* 4(3):264–271, URL <http://www.iajit.org/Vol.%204%2c%20No.%2c%20July%202007/July2007-Abstract.htm#11>
- Bachman P, Hjelm RD, Buchwalter W (2019) Learning representations by maximizing mutual information across views. In: Wallach HM, Larochelle H, Beygelzimer A, d’Alché-Buc F, Fox EB, Garnett R (eds) *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, pp 15509–15519, URL <https://proceedings.neurips.cc/paper/2019/hash/ddf354219aac374f1d40b7e760ee5bb7-Abstract.html>
- Bao H, Dong L, Piao S, Wei F (2022) Beit: BERT pre-training of image transformers. In: *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022, OpenReview.net*, URL <https://openreview.net/forum?id=p-BhZSz59o4>
- Bardes A, Ponce J, LeCun Y (2022) Vicreg: Variance-invariance-covariance regularization for self-supervised learning. In: *The Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022, OpenReview.net*, URL <https://openreview.net/forum?id=xm6YD62D1Ub>
- Bester H, Güth W (1998) Is altruism evolutionarily stable? *Journal of Economic Behavior Organization* 34(2):193–209, DOI [https://doi.org/10.1016/S0167-2681\(97\)00060-7](https://doi.org/10.1016/S0167-2681(97)00060-7), URL <https://www.sciencedirect.com/science/article/pii/S0167268197000607>
- Brown TB, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P, Neelakantan A, Shyam P, Sastry G, Askell A, Agarwal S, Herbert-Voss A, Krueger G, Henighan T, Child R, Ramesh A, Ziegler DM, Wu J, Winter C, Hesse C, Chen M, Sigler E, Litwin M, Gray S, Chess B, Clark J, Berner C, McCandlish S, Radford A, Sutskever I, Amodei D (2020) Language models are few-shot learners. In: *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*
- Caron M, Misra I, Mairal J, Goyal P, Bojanowski P, Joulin A (2020) Unsupervised learning of visual features by contrasting cluster assignments. In: Larochelle H, Ranzato M, Hadsell R, Balcan M, Lin H (eds) *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, URL <https://proceedings.neurips.cc/paper/2020/hash/70feb62b69f16e0238f741fab228fec2-Abstract.html>
- Chen T, Kornblith S, Norouzi M, Hinton GE (2020) A simple framework for contrastive learning of visual representations. In: *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event, PMLR, Proceedings of Machine Learning Research*, vol 119, pp 1597–1607, URL <http://proceedings.mlr.press/v119/chen20j.html>
- Chen X, Ding M, Wang X, Xin Y, Mo S, Wang Y, Han S, Luo P, Zeng G, Wang J (2023) Context autoencoder for self-supervised representation learning. URL <https://openreview.net/forum?id=Gb2RNdY5595>
- Chuang C, Robinson J, Lin Y, Torralba A, Jegelka S (2020) Debaised contrastive learning. In: Larochelle H, Ranzato M, Hadsell R, Balcan M, Lin H (eds) *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, URL <https://proceedings.neurips.cc/paper/2020/hash/63c3ddcc7b23daa1e42dc41f9a44a873-Abstract.html>
- Coates A, Ng AY, Lee H (2011) An analysis of single-layer networks in unsupervised feature learning. In: Gordon GJ, Dunson DB, Dudík M (eds) *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, AISTATS 2011, Fort Lauderdale, USA, April 11-13, 2011, JMLR.org, JMLR Proceedings*, vol 15, pp 215–223, URL <http://proceedings.mlr.press/v15/coates11a/coates11a.pdf>
- Deng J, Dong W, Socher R, Li L, Li K, Fei-Fei L (2009) Imagenet: A large-scale hierarchical image database. In: *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2009)*, 20-25 June 2009, Miami, Florida, USA, IEEE Computer Society, pp 248–255, DOI 10.1109/CVPR.2009.5206848, URL <https://doi.org/10.1109/CVPR.2009.5206848>
- Devlin J, Chang M, Lee K, Toutanova K (2019) BERT: pre-training of deep bidirectional transform-

- ers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers), Association for Computational Linguistics, pp 4171–4186, DOI 10.18653/v1/n19-1423, URL <https://doi.org/10.18653/v1/n19-1423>
- Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, Uszkoreit J, Houlsby N (2021) An image is worth 16x16 words: Transformers for image recognition at scale. In: 9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021, OpenReview.net, URL <https://openreview.net/forum?id=YicbFdNTTy>
- Easley DA, Kleinberg JM (2010) Networks, Crowds, and Markets - Reasoning About a Highly Connected World. Cambridge University Press, DOI 10.1017/CBO9780511761942, URL http://www.cambridge.org/gb/knowledge/isbn/item2705443/?site_locale=en_GB
- Elkan C, Noto K (2008) Learning classifiers from only positive and unlabeled data. In: Li Y, Liu B, Sarawagi S (eds) Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Las Vegas, Nevada, USA, August 24-27, 2008, ACM, pp 213–220, DOI 10.1145/1401890.1401920, URL <https://doi.org/10.1145/1401890.1401920>
- Elkins J (2008) Six stories from the end of representation: images in painting, photography, astronomy, microscopy, particle physics, and quantum mechanics, 1980-2000. Stanford University Press, URL <http://www.sup.org/books/title/?id=1341>
- Ermolov A, Siarohin A, Sangineto E, Sebe N (2021) Whitening for self-supervised representation learning. In: Meila M, Zhang T (eds) Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18-24 July 2021, Virtual Event, PMLR, Proceedings of Machine Learning Research, vol 139, pp 3015–3024, URL <http://proceedings.mlr.press/v139/ermolov21a.html>
- Friedman D (1991) Evolutionary games in economics. *Econometrica* 59(3):637–666, URL <http://www.jstor.org/stable/2938222>
- Gatenby RA, Vincent TL (2003) Application of quantitative models from population biology and evolutionary game theory to tumor therapeutic strategies. *Molecular cancer therapeutics* 2(9):919–927
- Gesmundo A (2022) A continual development methodology for large-scale multitask dynamic ML systems. CoRR abs/2209.07326, DOI 10.48550/arXiv.2209.07326, URL <https://doi.org/10.48550/arXiv.2209.07326>, 2209.07326
- Grill J, Strub F, Altché F, Tallec C, Richemond PH, Buchatskaya E, Doersch C, Pires BÁ, Guo Z, Azar MG, Piot B, Kavukcuoglu K, Munos R, Valko M (2020) Bootstrap your own latent - A new approach to self-supervised learning. In: Larochelle H, Ranzato M, Hadsell R, Balcan M, Lin H (eds) Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual, URL <https://proceedings.neurips.cc/paper/2020/hash/f3ada80d5c4ee70142b17b8192b2958e-Abstract.html>
- He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016, IEEE Computer Society, pp 770–778, DOI 10.1109/CVPR.2016.90, URL <https://doi.org/10.1109/CVPR.2016.90>
- He K, Fan H, Wu Y, Xie S, Girshick RB (2020) Momentum contrast for unsupervised visual representation learning. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, June 13-19, 2020, Computer Vision Foundation / IEEE, pp 9726–9735, DOI 10.1109/CVPR42600.2020.00975, URL <https://doi.org/10.1109/CVPR42600.2020.00975>
- He K, Chen X, Xie S, Li Y, Dollár P, Girshick RB (2022a) Masked autoencoders are scalable vision learners. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022, New Orleans, LA, USA, June 18-24, 2022, IEEE, pp 15979–15988, DOI 10.1109/CVPR52688.2022.01553, URL <https://doi.org/10.1109/CVPR52688.2022.01553>
- He K, Chen X, Xie S, Li Y, Dollár P, Girshick RB (2022b) Masked autoencoders are scalable vision learners. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022, New Orleans, LA, USA, June 18-24, 2022, IEEE, pp 15979–15988, DOI 10.1109/CVPR52688.2022.01553, URL <https://doi.org/10.1109/CVPR52688.2022.01553>
- Hénaff OJ (2020) Data-efficient image recognition with contrastive predictive coding. In: Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event,

- PMLR, Proceedings of Machine Learning Research, vol 119, pp 4182–4192, URL <http://proceedings.mlr.press/v119/henaff20a.html>
- Hjelm RD, Fedorov A, Lavoie-Marchildon S, Grewal K, Bachman P, Trischler A, Bengio Y (2019) Learning deep representations by mutual information estimation and maximization. In: 7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019, Open-Review.net, URL <https://openreview.net/forum?id=Bk1r3j0cKX>
- Huck S, Oechssler J (1999) The indirect evolutionary approach to explaining fair allocations. *Games and Economic Behavior* 28(1):13–24, DOI <https://doi.org/10.1006/game.1998.0691>, URL <https://www.sciencedirect.com/science/article/pii/S0899825698906911>
- Hüttenrauch M, Šošić A, Neumann G (2017) Guided deep reinforcement learning for swarm systems. arXiv preprint arXiv:170906011
- feng Ji S, Zhao D, juan Luo R (2019) Evolutionary game analysis on local governments and manufacturers’ behavioral strategies: Impact of phasing out subsidies for new energy vehicles. *Energy* 189:116064, DOI <https://doi.org/10.1016/j.energy.2019.116064>, URL <https://www.sciencedirect.com/science/article/pii/S0360544219317591>
- Kalashnikov D, Irpan A, Pastor P, Ibarz J, Herzog A, Jang E, Quillen D, Holly E, Kalakrishnan M, Vanhoucke V, et al. (2018) Scalable deep reinforcement learning for vision-based robotic manipulation. In: Conference on Robot Learning, PMLR, pp 651–673
- Kiran BR, Sobh I, Talpaert V, Mannion P, Al Sallab AA, Yogamani S, Pérez P (2021) Deep reinforcement learning for autonomous driving: A survey. *IEEE Transactions on Intelligent Transportation Systems*
- Kolesnikov A, Beyer L, Zhai X, Puigcerver J, Yung J, Gelly S, Houlsby N (2020) Big transfer (bit): General visual representation learning. In: Vedaldi A, Bischof H, Brox T, Frahm J (eds) *Computer Vision - ECCV 2020 - 16th European Conference, Glasgow, UK, August 23-28, 2020, Proceedings, Part V*, Springer, Lecture Notes in Computer Science, vol 12350, pp 491–507, DOI [10.1007/978-3-030-58558-7_29](https://doi.org/10.1007/978-3-030-58558-7_29), URL https://doi.org/10.1007/978-3-030-58558-7_29
- Kornblith S, Shlens J, Le QV (2019) Do better imagenet models transfer better? In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019, Computer Vision Foundation / IEEE, pp 2661–2671, DOI [10.1109/CVPR.2019.00277](https://doi.org/10.1109/CVPR.2019.00277), URL http://openaccess.thecvf.com/content_CVPR_2019/html/Kornblith_Do_Better_ImageNet_Models_Transfer_Better_CVPR_2019_paper.html
- Krizhevsky A (2009) Learning multiple layers of features from tiny images pp 32–33, URL <https://www.cs.toronto.edu/~kriz/learning-features-2009-TR.pdf>
- Kumar A, Pratap A, Singh AK (2023) Generative adversarial neural machine translation for phonetic languages via reinforcement learning. *IEEE Trans Emerg Top Comput Intell* 7(1):190–199, DOI [10.1109/TETCI.2022.3209394](https://doi.org/10.1109/TETCI.2022.3209394), URL <https://doi.org/10.1109/TETCI.2022.3209394>
- Li D, Wu H, Zhang J, Huang K (2019) Fast A3RL: aesthetics-aware adversarial reinforcement learning for image cropping. *IEEE Trans Image Process* 28(10):5105–5120, DOI [10.1109/TIP.2019.2914360](https://doi.org/10.1109/TIP.2019.2914360), URL <https://doi.org/10.1109/TIP.2019.2914360>
- Li S, Liu Z, Wu D, Liu Z, Li SZ (2021a) Boosting discriminative visual representation learning with scenario-agnostic mixup. *CoRR* abs/2111.15454, URL <https://arxiv.org/abs/2111.15454>, [2111.15454](https://arxiv.org/abs/2111.15454)
- Li Y, Pogodin R, Sutherland DJ, Gretton A (2021b) Self-supervised learning with kernel dependence maximization. In: Ranzato M, Beygelzimer A, Dauphin YN, Liang P, Vaughan JW (eds) *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*, pp 15543–15556, URL <https://proceedings.neurips.cc/paper/2021/hash/83004190b1793d7aa15f8d0d49a13eba-Abstract.html>
- Liu X, Wang Z, Li Y, Wang S (2022) Self-supervised learning via maximum entropy coding. *CoRR* abs/2210.11464, DOI [10.48550/arXiv.2210.11464](https://doi.org/10.48550/arXiv.2210.11464), URL <https://doi.org/10.48550/arXiv.2210.11464>, [2210.11464](https://arxiv.org/abs/2210.11464)
- Malthus TR (1872) *An Essay on the Principle of Population*.
- Maynard Smith J (1974) The theory of games and the evolution of animal conflicts. *Journal of Theoretical Biology* 47(1):209–221, DOI [https://doi.org/10.1016/0022-5193\(74\)90110-6](https://doi.org/10.1016/0022-5193(74)90110-6), URL <https://www.sciencedirect.com/science/article/pii/0022519374901106>
- Nakamura TK (2000) Statistical mechanics of a collisionless system based on the maximum entropy principle. *The Astrophysical Journal* 531(2):739
- van den Oord A, Li Y, Vinyals O (2018) Representation learning with contrastive predictive coding. *CoRR* abs/1807.03748, URL <https://arxiv.org/abs/1807.03748>

- [org/abs/1807.03748](#), 1807.03748
- Osborne MJ, Rubinstein A (1994) A course in game theory. MIT press
- du Plessis MC, Niu G, Sugiyama M (2014) Analysis of learning from positive and unlabeled data. In: Ghahramani Z, Welling M, Cortes C, Lawrence ND, Weinberger KQ (eds) Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada, pp 703–711, URL <https://proceedings.neurips.cc/paper/2014/hash/35051070e572e47d2c26c241ab88307f-Abstract.html>
- Qiang W, Li J, Zheng C, Su B, Xiong H (2022) Interventional contrastive learning with meta semantic regularizer. In: Chaudhuri K, Jegelka S, Song L, Szepesvári C, Niu G, Sabato S (eds) International Conference on Machine Learning, ICML 2022, 17-23 July 2022, Baltimore, Maryland, USA, PMLR, Proceedings of Machine Learning Research, vol 162, pp 18018–18030, URL <https://proceedings.mlr.press/v162/qiang22a.html>
- Robinson JD, Chuang C, Sra S, Jegelka S (2021) Contrastive learning with hard negative samples. In: 9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021, OpenReview.net, URL <https://openreview.net/forum?id=CR1XOQOUTH->
- Saunshi N, Plevrakis O, Arora S, Khodak M, Khandeparkar H (2019) A theoretical analysis of contrastive unsupervised representation learning. In: Chaudhuri K, Salakhutdinov R (eds) Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA, PMLR, Proceedings of Machine Learning Research, vol 97, pp 5628–5637, URL <http://proceedings.mlr.press/v97/saunshi19a.html>
- Schulman J, Wolski F, Dhariwal P, Radford A, Klimov O (2017) Proximal policy optimization algorithms. CoRR abs/1707.06347, URL <http://arxiv.org/abs/1707.06347>, 1707.06347
- Shalf J (2020) The future of computing beyond moore’s law. Philosophical Transactions of the Royal Society A 378(2166):20190061
- Silver D, Schrittwieser J, Simonyan K, Antonoglou I, Huang A, Guez A, Hubert T, Baker L, Lai M, Bolton A, et al. (2017) Mastering the game of go without human knowledge. nature 550(7676):354–359
- Silver D, Hubert T, Schrittwieser J, Antonoglou I, Lai M, Guez A, Lanctot M, Sifre L, Kumaran D, Graepel T, et al. (2018) A general reinforcement learning algorithm that masters chess, shogi, and go through self-play. Science 362(6419):1140–1144
- Sridharan K, Kakade SM (2008) An information theoretic framework for multi-view learning. In: Servedio RA, Zhang T (eds) 21st Annual Conference on Learning Theory - COLT 2008, Helsinki, Finland, July 9-12, 2008, Omnipress, pp 403–414, URL <http://colt2008.cs.helsinki.fi/papers/94-Sridharan.pdf>
- Sun Y, Yuan B, Xiang Q, Zhou J, Yu J, Dai D, Zhou X (2023) Intelligent decision-making and human language communication based on deep reinforcement learning in a wargame environment. IEEE Trans Hum Mach Syst 53(1):201–214, DOI 10.1109/THMS.2022.3225867, URL <https://doi.org/10.1109/THMS.2022.3225867>
- Tian Y, Krishnan D, Isola P (2020a) Contrastive multiview coding. In: Vedaldi A, Bischof H, Brox T, Frahm J (eds) Computer Vision - ECCV 2020 - 16th European Conference, Glasgow, UK, August 23-28, 2020, Proceedings, Part XI, Springer, Lecture Notes in Computer Science, vol 12356, pp 776–794, DOI 10.1007/978-3-030-58621-8_45, URL https://doi.org/10.1007/978-3-030-58621-8_45
- Tian Y, Sun C, Poole B, Krishnan D, Schmid C, Isola P (2020b) What makes for good views for contrastive learning? URL <https://proceedings.neurips.cc/paper/2020/hash/4c2e5eaae9152079b9e95845750bb9ab-Abstract.html>
- Tsai YH, Wu Y, Salakhutdinov R, Morency L (2021) Self-supervised learning from a multi-view perspective. In: 9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021, OpenReview.net, URL https://openreview.net/forum?id=-bdp_8Itjwp
- Verma V, Luong T, Kawaguchi K, Pham H, Le QV (2021) Towards domain-agnostic contrastive learning. In: Meila M, Zhang T (eds) Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18-24 July 2021, Virtual Event, PMLR, Proceedings of Machine Learning Research, vol 139, pp 10530–10541, URL <http://proceedings.mlr.press/v139/verma21a.html>
- Vincent T (1985) Evolutionary games. Journal of Optimization Theory and Applications 46(4):605–612
- Vinyals O, Babuschkin I, Czarnecki WM, Mathieu M, Dudzik A, Chung J, Choi DH, Powell R, Ewalds T, Georgiev P, et al. (2019) Grandmaster level in starcraft ii using multi-agent reinforcement learning. Nature 575(7782):350–354
- Wang Y, Zhang Q, Wang Y, Yang J, Lin Z (2022) Chaos is a ladder: A new theoretical understanding of contrastive learning via augmentation overlap. In: The

- Tenth International Conference on Learning Representations, ICLR 2022, Virtual Event, April 25-29, 2022, OpenReview.net, URL <https://openreview.net/forum?id=ECvgmYVyeUz>
- Weibull JW (1997) Evolutionary game theory. MIT press
- Xiao T, Wang X, Efros AA, Darrell T (2021) What should not be contrastive in contrastive learning. In: 9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021, OpenReview.net, URL <https://openreview.net/forum?id=CZ8Y3NzuVz0>
- Xie Z, Zhang Z, Cao Y, Lin Y, Bao J, Yao Z, Dai Q, Hu H (2022) Simmim: a simple framework for masked image modeling. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022, New Orleans, LA, USA, June 18-24, 2022, IEEE, pp 9643–9653, DOI 10.1109/CVPR52688.2022.00943, URL <https://doi.org/10.1109/CVPR52688.2022.00943>
- Yu K, Dong C, Lin L, Loy CC (2018) Crafting a toolchain for image restoration by deep reinforcement learning. In: 2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018, Computer Vision Foundation / IEEE Computer Society, pp 2443–2452, DOI 10.1109/CVPR.2018.00259, URL http://openaccess.thecvf.com/content_cvpr_2018/html/Yu_Crafting_a_Toolchain_CVPR_2018_paper.html
- Zbontar J, Jing L, Misra I, LeCun Y, Deny S (2021) Barlow twins: Self-supervised learning via redundancy reduction. In: Meila M, Zhang T (eds) Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18-24 July 2021, Virtual Event, PMLR, Proceedings of Machine Learning Research, vol 139, pp 12310–12320, URL <http://proceedings.mlr.press/v139/zbontar21a.html>
- Zhang Z, Sabuncu MR (2018) Generalized cross entropy loss for training deep neural networks with noisy labels. In: Bengio S, Wallach HM, Larochelle H, Grauman K, Cesa-Bianchi N, Garnett R (eds) Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3-8, 2018, Montréal, Canada, pp 8792–8802, URL <https://proceedings.neurips.cc/paper/2018/hash/f2925f97bc13ad2852a7a551802feea0-Abstract.html>

A Nomenclature

For ease of understanding the detailed usage of abbreviations and symbols, we provide the nomenclature in Table 8 and Table 9.

B Datasets

We evaluate our method on five datasets, STL10 Coates et al. (2011), CIFAR10 Krizhevsky (2009), CIFAR100 Krizhevsky (2009), Tiny-ImageNet-200 Krizhevsky (2009) and ImageNet Deng et al. (2009). The details of each dataset are shown in Table 10.

C Implementation Details

C.1 Code Repositories Duplicated

To evaluate each model’s generalizability and discriminability, we use open-source code repositories to train base-lines and implement our method. For BarlowTwins Zbontar et al. (2021), the repository is <https://github.com/facebookresearch/barlowtwins>, SimCLR Chen et al. (2020) is <https://github.com/sthalles/SimCLR>, MoCo He et al. (2020) is <https://github.com/facebookresearch/moco> and BYOL Grill et al. (2020) is <https://github.com/sthalles/PyTorch-BYOL>. We implement our method based on the settings of BarlowTwins Zbontar et al. (2021). We implement our work on $4 \times$ NVIDIA Tesla A100 for 72 hours. Since the GPUs are shared with other groups, the training time is just an approximate time rather than an exact one.

C.2 The Supervised Learning Results Used for Calculating EGT

For calculating EGT, we use the performance of existing supervised learning on conventional benchmark datasets, which are 99.60% Gesmundo (2022) on STL10, 99.37% Kolesnikov et al. (2020) on CIFAR10, 93.51% Kolesnikov et al. (2020) on CIFAR100 and 72.18% Li et al. (2021a) on Tiny-ImageNet-200.

C.3 Hyper-Parameters

In our EGT model, we set the value of ω_1 and ω_2 to 1, which corresponds to our objective that we hope the model can acquire strong generalizability and discriminability in a balanced manner. The details of training hyper-parameters are shown in Table 11.

D Pseudo-code and Time Complexity

The details of the training process are shown in Algorithm 1. Related symbols can refer to Table 8.

Algorithm 1 ESSL: Evolutionary game-guided Self-Supervised Learning

Input: Multi-view dataset X with NV views of each sample, minibatch size n , and hyper-parameters α, β, δ .
Initialize The neural network parameters: θ and ϑ for view-specific encoders $f_\theta(\cdot)$ and projection heads $g_\vartheta(\cdot)$, $AGENT$ for agent network. l for learning rate of contrastive learning, l_a for learning rate of RL, train agent per k step.
repeat
 for t -th training iteration (T epochs) **do**
 Iteratively sample minibatch $\tilde{X} = \{X_i\}_{i=(t-1)n}^{tn}$. (time cost t_0)
 $z_t = g_\vartheta(f_\theta(\tilde{X}))$ (time cost t_1)
 # reinforcement learning training step
 $s_t = F_s(z_t)$ (time cost t_2)
 $\alpha_t, \beta_t = a_t = AGENT(s_t)$ (time cost t_3)
 $r_t = G(\alpha_t, \beta_t, x^*, y^*) + \xi \frac{1}{|L_t - \phi L_{t-1}|}$ (time cost t_4)
 Store transition (s_t, a_t, r_t) (time cost t_5)
 if $t \% k == 0$ **then**
 $AGENT = AGENT + l_a \delta_{AGENT} \mathcal{L}_{PPO}$ (time cost t_6)
 end if
 # contrastive learning training step (time cost t_7)
 $\mathcal{L}_{ESSL} = \alpha_t \mathcal{L}_{InfoNCE} + \beta_t \mathcal{L}_{BarlowTwins}$ (time cost t_8)
 $\theta \leftarrow \theta - \ell \Delta_\theta \mathcal{L}_{ESSL}$ (time cost t_9)
 end for
until θ, ϑ converge.

Following the pseudo-code above, we further calculate the time complexity of ESSL.

$$\begin{aligned} T(episode, t) &= t_0 + t_1 + t_2 + t_3 + t_4 + t_5 + t_8 + t_9 \times T \\ &= t \times T \\ &= T \end{aligned} \tag{32}$$

Note that ESSL does not change the order of time complexity, which is still a linear complexity related to epoch numbers.

Table 8 Nomenclature of preliminary and method

Nomenclature	
Variables	
$\mathcal{D}$	dataset
M	model
M_g	generalizability model
M_d	discriminability model
M_{ens}	the ensemble model of M_g and M_d
G	the gain of generalizability by M
G_1	the gain of generalizability by M_g
G_2	the gain of generalizability by M_d
D	the gain if discriminability by M
D_1	the gain of discriminability by M_g
D_2	the gain of discriminability by M_d
N_1	the negative impact of M_g on the discriminability of M_d
N_2	the negative impact of M_d on the generalizability of M_g
S_G	the strategy spaces of generalizability model
S_D	the strategy of discriminability model
A	"Adopted" strategy
U	"Unadopted" strategy
$\mathbf{H}$	the income matrix of the generalizability model
$\mathbf{K}$	the income matrix of the discriminability model
U_{G1}	the expected utility when the generalizability model selects the strategy A
$\bar{U}_G$	the corresponding average expected utility of the generalizability models
$F(\cdot)$	replicator dynamic equation
$\mathcal{L}_{InfoNCE}$	InfoNCE loss
$\mathcal{L}_{BarlowTwins}$	BarlowTwins loss
$\mathcal{L}_{ESSL}$	ESSL loss
s_t	state in step t of agent
a_t	action in step t that agent select
r_t	reward in step t that agent receives
Parameters	
ω_1	the preference for generalizability
ω_2	the preference for discriminability
x	the proportion of discriminability models
y	the proportion of generalizability models
ξ	hyper-parameter in the reward function
V	hyper-parameter of action
a_t^α	the hyper-parameter controlling generalizability at time step t
b_t^β	the hyper-parameter controlling discriminability at time step t
Acronyms	
SSL	self-supervised learning
M-M	masked language/image modeling
EGT	evolutionary game theory
ESSL	evolutionary game-guided self-supervised learning
RL	reinforcement learning
SS-CL	self-supervised contrastive learning
AGI	artificial general intelligence
CV	computer vision
NLP	natural language processing
PPO	Proximal Policy Optimization
SL	supervised learning
ESS	evolutionary stable strategy
RD	replicator dynamic
PT	pre-trained dataset
S10	STL10
C10	CIFAR10
C100	CIFAR100
Tiny	Tiny-ImageNet-200

Table 9 Nomenclature of theoretical analyses

Nomenclature	
Variables	
$\mathcal{L}_{softmax}$	softmax loss
f	the fixed encoder pre-trained by self-supervised methods
W	linear-probing classifier
T	the label on the downstream tasks
$\mathcal{L}_{GE}^f$	the generalization error for encoder f
$\mathcal{L}_{InfoNCE}^f$	the conventional contrastive learning objective for encoder f
$\mathcal{L}_{BarlowTwins}^f$	the BarlowTwins loss for encoder f
N^S	the number of training samples
N^{batch}	the size of a mini-batch
$\mathcal{R}_H(\beta)$	the Radermacher complexity
J	Jacobian matrix of the RD equations
$\mathcal{H}$	a restricted function hypothesis space
X	training data
d	the dimensionality of the representations learned by f
N^{neg}	the size of negative pairs
Parameters	
α	the hyper-parameter to denote the weight of $\mathcal{L}_{InfoNCE}^f$
β	the hyper-parameter to denote the weight of $\mathcal{L}_{BarlowTwins}^f$
δ	the probability hyper-parameter to hold the upper bound of generalization error

Table 10 Characteristics of image datasets used.

Dataset	Classes	OTE ¹	TE ²	VE ³	TE ⁴	Accuracy Measure
STL10	10	113000	100000	5000	8000	Top-1 accuracy
CIFAR10	10	50000	45000	5000	10000	Top-1 accuracy
CIFAR100	100	50000	44933	5067	10000	Top-1 accuracy
Tiny-ImageNet-200	200	120000	100000	10000	10000	Top-1 accuracy
ImageNet	1000	1281167	1271158	10009	50000	Top-1 accuracy

¹ Original train examples² Train Examples³ Valid Examples⁴ Test Examples**Table 11** Hyper-Parameters.

Hyper-Parameter	Value
Epoch	500 ¹ , 1000 ²
Batch Size	256
Learning Rate Weights	0.2
Learning Rate Biases	0.0048
Weight Decay	0.0004
Projector Size	8192-8192-8192
lambda	0.0051
Temperature	0.07
N views	2
Z	1.5
V	0.5
k	200

¹ For STL10, CIFAR10, CIFAR100, Tiny-ImageNet-200² For ImageNet-1000