

Efficient self-supervised learning using dataset distillation

Master/Eng internship

General information

- **Duration** : 6 months (standard stipend). To start between February and March 2026.
- **Institute** : Université Paris Cité, Laboratoire d'Informatique Paris Descartes (LIPADE), team [Systèmes Intelligents de Perception](#)
- **Location** : 45 rue des Saints-Pères, 75006, Paris
- **Supervision** : Ayoub Karine¹, Camille Kurtz², Laurent Wendling³
- **Keywords** : computer vision, deep learning, self-supervised learning, efficiency, dataset distillation
- **Application** : send an email with CV to ayoub.karine@u-paris.fr, camille.kurtz@u-paris.fr and laurent.wendling@u-paris.fr with subject "[DD-SSL]" (the application through message in LinkedIn will not be considered)

Proposed topic

Context

The performance of supervised deep learning methods in computer vision heavily depends on the availability of labeled data, whose annotation is both time-consuming and requires expert knowledge. To overcome this limitation, Self-Supervised Learning (SSL) has emerged as a promising alternative to address the challenge of limited annotations. In this paradigm, models learn from unlabeled data by generating their own supervisory signals. The resulting pre-trained models can then be fine-tuned on various downstream tasks such as image classification, object detection, and semantic segmentation. However, achieving performance comparable to supervised learning often requires large-scale datasets and high training costs, which significantly increase computational and storage demands. This internship aims to alleviate these constraints by exploring data distillation techniques to make SSL training more efficient.

Work to be done

Dataset Distillation (DD) [1] aims to condense a large-scale training dataset into a much smaller synthetic one such that models trained on the distilled data achieve performance comparable to those trained on the original dataset (see [figure 1](#)). Most existing DD methods are designed for efficient supervised learning and can be broadly classified into three main categories [2] : (1) Performance Matching, which minimizes the loss on the synthetic dataset by aligning the performance of models trained on real and synthetic data, (2) Parameter Matching, which trains two neural networks respectively on real and synthetic data and encourages similarity in their parameters and (3) Distribution Matching, which generates synthetic data that closely mimics the distribution of the original dataset. In this internship, we will focus on the Parameter Matching approach. Building upon the work of Cazenavette et al. [3], the authors of [4] extended this concept to SSL using knowledge distillation [5, 6, 7], particularly employing SSL methods such as Barlow Twins and SimCLR. In the same vein, this internship will explore the DINO (self-Distillation with NO labels, MetaAI) SSL method [8], which naturally produces teacher–student parameter trajectories that can be leveraged for Parameter Matching. The different steps of the internship are :

- ▷ **Step 1 – Literature review** : Review recent dataset distillation methods applied to computer vision, with a focus on parameter matching and SSL-based approaches.
- ▷ **Step 2 – Trajectory Observation** : Analyze and visualize the teacher–student parameter trajectories generated by DINO during SSL training.
- ▷ **Step 3 – Integration into Data Distillation Frameworks** : Design a trajectory matching loss based on DINO's teacher–student dynamics and train a student model on synthetic data guided by these trajectories.
- ▷ **Step 4 – Test on down-stream computer vision tasks** : Assess the effectiveness of the proposed approach on tasks such as image classification

1. <https://www.ayoub-karine.com/>

2. <https://www.camille-kurtz.com/>

3. <https://helios2.mi.parisdescartes.fr/~lwendlin/>

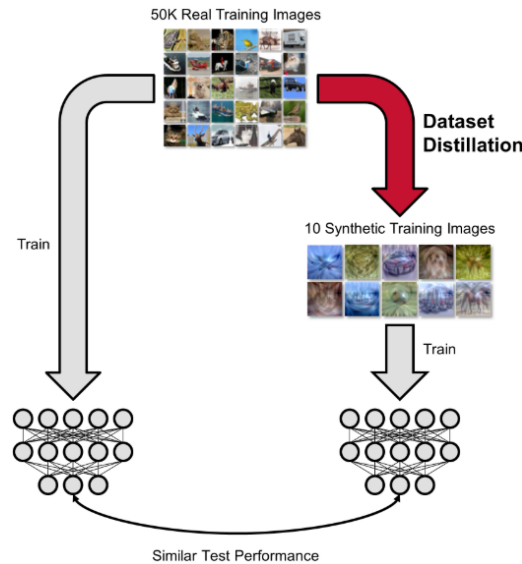


FIGURE 1 – Principle of dataset distillation [3].

Desired background for the candidate

We are looking for a Master 2 student or final year of MSc, or engineering school in computer science. The ideal candidate should have knowledge in deep learning, computer vision, Python programming and an interest in efficient machine/deep learning.

Bibliography

- [1] Tongzhou WANG et al. "Dataset distillation". In : *arXiv preprint arXiv :1811.10959* (2018).
- [2] Ruonan YU, Songhua LIU et Xinchao WANG. "Dataset distillation : A comprehensive review". In : *IEEE transactions on pattern analysis and machine intelligence* 46.1 (2023), p. 150-170.
- [3] George CAZENAVETTE et al. "Dataset distillation by matching training trajectories". In : *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022, p. 4750-4759.
- [4] Siddharth JOSHI, Jiayi NI et Baharan MIRZASOLEIMAN. "Dataset Distillation via Knowledge Distillation : Towards Efficient Self-Supervised Pre-training of Deep Networks". In : *The Thirteenth International Conference on Learning Representations*. 2025. URL : <https://openreview.net/forum?id=c61unr33XA>.
- [5] Geoffrey HINTON, Oriol VINYALS et Jeff DEAN. "Distilling the knowledge in a neural network". In : *arXiv preprint arXiv :1503.02531* (2015).
- [6] Ayoub KARINE, Thibault NAPOLÉON et Maher JRIDI. "I2CKD : Intra- and inter-class knowledge distillation for semantic segmentation". In : *Neurocomputing* 649 (oct. 2025), p. 130791. URL : <https://hal.science/hal-05144692>.
- [7] Ayoub KARINE, Thibault NAPOLÉON et Maher JRIDI. "Channel-spatial knowledge distillation for efficient semantic segmentation". In : *Pattern Recognition Letters* 180 (avr. 2024), p. 48-54. URL : <https://hal.science/hal-04488459>.
- [8] Oriane SIMÉONI et al. "Dinov3". In : *arXiv preprint arXiv :2508.10104* (2025).