# Multimodal Dataset Distillation for Efficient Learning

May 2024

## Project description

Dataset distillation compresses a large dataset into a smaller, synthetic version that maintains its effectiveness for training models on various tasks. This process is useful for understanding the structure of the datasets better and has practical applications in areas such as continual learning and data privacy (1).

This project aims to integrate infrared and visual imagery to develop an advanced multimodal data distillation method. This approach will enhance the accuracy of detection models and enable the creation of a comprehensive dataset for diverse applications.

**Motivation** The increasing size of datasets in many domains poses a significant challenge for model training due to the substantial computational resources required. Dataset distillation, which integrates both infrared and visual inputs, can address these challenges by enhancing data richness and reducing the complexity typically associated with large single-modality datasets. This integration promises to overcome environmental limitations such as variable lighting and occlusions, providing a robust dataset that is easier to manage and more effective in practical applications. The ultimate goal is to produce a dataset that supports efficient model training for a broad ranges of applications, paving the way for advancements in fields such as environmental monitoring, autonomous navigation, and public safety.

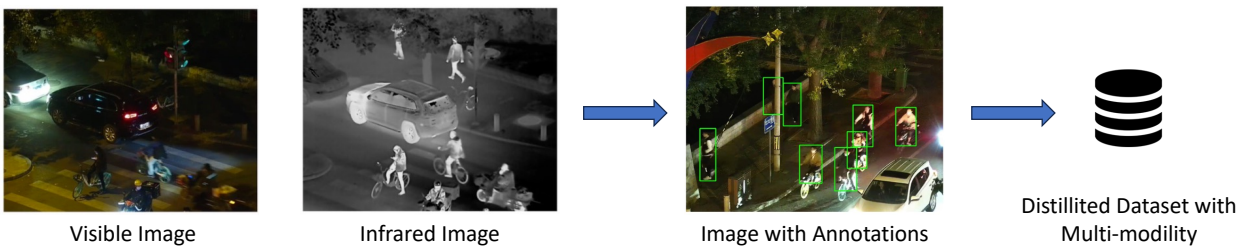


Figure 1: The overview of the multimodal dataset distillation.

**Goal** The goal of this project is to develop an advanced multimodal dataset distillation method by integrating infrared and visual imagery. The distilled dataset will support the efficient training of machine learning models for various tasks, including object detection, pattern recognition, and anomaly detection. By providing diverse, rich, and robust training data that reflect a wide range of environmental conditions, we aim to enhance model accuracy and performance in these applications.

In this project you will mainly experiment with the publicly available datasets such as LLVIP: A Visible-infrared Paired Dataset for Low-light Vision (2) and KAIST Multispectral Pedestrian Detection Benchmark (3), and compare your multimodal data distillation method to several state-of-the-art (SOTA) single-modal data distillation baselines.

## Additional information

- **What will you learn?** The student will work on the SOTA transformer architectures for representation learning, and gain hands-on experience on foundation models when evaluating the performance of the developed methodology.
- **Requirements:** Deep learning knowledge (experience with self-supervised learning is a big plus), good Python and PyTorch skills, version control (git).
- **Supervisors:** Dr. Yun Cheng (yun.cheng@sdsc.ethz.ch),
  Prof. Dr. Olga Saukh (saukh@tugraz.at), Prof. Dr. Fernando Perez-Cruz

## References

[1] Guang, H.: Awesome dataset distillation. `https://github.com/Guang000/Awesome-Dataset-Distillation` (2024)

[2] Jia, X., Zhu, C., Li, M., Tang, W., Zhou, W.: Llvip: A visible-infrared paired dataset for low-light vision. In: Proceedings of the IEEE/CVF international conference on computer vision, 3496–3504 (2021)

[3] Hwang, S.: Rgbt pedestrian detection. `https://soonminhwang.github.io/rgbt-ped-detection/` (2021)