MLMI17 Advanced Computer Vision:

Coursework #1: Representation Learning and Generative Models

Ayush Tewari

Due: Feb 23rd, 2026 online via moodle

Your answers should contain an explanation of what you do. You must also give an *interpretation* of what the numerical values and graphs you provide *mean* – why are the results the way they are? Each question should be labelled and answered separately and distinctly. Total combined length of answers must not exceed 5 sides of a4 (plus cover page), minimum 11pt font, 1 inch margins.

Part 1: Representation Learning This part of the coursework uses two files: `representations.py`, which contains the model definitions and training logic, and `part1.py`, which is the main script for running experiments. You are required to complete the missing code blocks marked with `### YOUR CODE STARTS HERE ###` in `representations.py`. All code is hosted at https://github.com/CambridgeCVCourses/CW1/tree/main.

(a) **Feature Extraction and t-SNE Analysis:** Extract features from the provided dataset using CLIP, DINO, and MAE backbones. Visualize these using t-SNE. Compare how these three pre-training objectives (contrastive, self-distillation, and masked auto-encoding) cluster the fruit classes in the embedding space.

(b) **Linear Probing:** Train a linear head on top of the frozen backbones. Report validation accuracies and provide an interpretation of the training loss curves. Does the ranking of models in the t-SNE visualization correlate with their linear probing performance? (HINT: Accuracy with CLIP should be around 90%.)

(c) **Fine-tuning:** Initialise a model with your best linear probe and fine-tune the entire backbone. Report validation accuracies and provide an interpretation of the training loss curves. Compare these results with the linear probe results. Explain the performance gain (or lack thereof).

Part 2: Generative Models (Will be added by the 6th when lecture material is final. Please check back then.)