Intro

You will develop a self-supervised learning (SSL) method to learn visual features from a large unlabeled image corpus (∼500k images). Train your model however you like (objective, architecture, optimizer, pipeline). We will not provide labels for the **pretraining set**; students are encouraged to read and study image-SSL methods. Your backbone must have fewer than **100M parameters** and must be **randomly initialized**.

Overview

- **Data.** We provide (i) an unlabeled pretraining set (`pretrain/`) and (ii) a public downstream dataset (`eval_public/`) with labeled `train/test` splits. Staff also maintain a private held-out set (`eval_private/`). Data repository: Hugging Face dataset (tsbpp/fall2025_deeplearning). We currently provide (approximately) ∼500k images; if there is strong demand, we may release more. We also encourage students to collect their own data as part of their research.
- **Evaluation (Linear Eval: Linear Probe or k-NN).** After SSL pretraining, the learned encoder must remain *frozen at all times*. Evaluate features using either (a) a *linear probe* trained on `train` or (b) *k-NN* with a feature bank built on `train`.
- **Competition.** Teams compete on performance. The public leaderboard uses `eval_public/test` with the provided evaluation scripts (linear probe and k-NN) under a frozen encoder. Final ranking will include the staff-run evaluation on the private held-out set. **Do not train or adapt on any test images.**
- **Compute resources.** Please use Greene, Google Colab, and other available resources to train. We are negotiating with NYU HPC and will announce additional compute support in the coming days.

Dates

- **Sanity-check phase**: Before releasing public tests, we will use CIFAR-10/100 as a temporary sanity check to validate pipelines and evaluation.
- **Initial test + Kaggle platform release**: November 18, 2025.
- **Final public test release**: November 25, 2025.
- **Final submission deadline**: December 2, 2025 (11:59pm local time).
- **Report**: Due during the exam period. 4 pages (excluding references), with citations, using the CVPR template.
- *Note on deadline flexibility*: The final submission deadline may be pushed back by a few days; for now, plan for the listed date.

Rules

1. **Do not train on test images.** If you are caught training or adapting on any test images (public or private), you will receive a **0** and be reported for academic dishonesty.

2. **Your model must be randomly initialized.**
3. **Model size cap.** Backbone parameters must be strictly < 100M at train time.
4. **Image resolution.** All images are **96 px**. Do not change the resolution.

Potentially Helpful FAQs

What models/optimizers/methods are allowed?
Any backbone (must be < 100M parameters), any optimizer, and any SSL method are allowed. Document your choices and rationale.

Do you provide labels for the pretraining set?
No. The pretraining set is unlabeled by design. Labels exist only in `eval_public/train` and `test` for evaluation (k-NN uses `train` labels to build the feature bank).

Can we use additional data?
Yes. You may incorporate additional data as part of your research. Clearly document sources and preprocessing.

Can we adapt on test images without labels?
No. Test images are for inference only.

How are we evaluated?
Top-1 accuracy using k-NN on `eval_public/test` with features from a frozen backbone; staff will run the same on a private held-out set for final ranking.

What about linear probing?
You may report linear-probe results in your write-up, but the leaderboard uses k-NN for apples-to-apples comparison.

Are we competing?
Yes. There is a public leaderboard and final ranking including the private held-out set.

Should we tune to the public test?
We do not recommend overfitting to the public test. The private held-out set may differ; the goal is to learn a universal representation that transfers.

Where should we post logistics questions?
Please post logistics and administrative questions *publicly* on Campuswire so the whole class benefits from the answers.

How should we ask staff research questions?
If your question involves your research ideas or approach, post on Campuswire with the audience set to **staff only** to avoid leaking ideas to other teams.

Potentially Helpful Papers

- **SimCLR: A Simple Framework for Contrastive Learning of Visual Representations** (2020). Contrastive learning with strong augmentations.
- **MoCo v2: Improved Baselines with Momentum Contrastive Learning** (2020). Momentum encoder and queue for contrastive learning.

- **MoCo v3: An Empirical Study of Training Self-Supervised Vision Transformers** (2021). SSL with ViTs.
- **BYOL: Bootstrap Your Own Latent** (2020). Non-contrastive learning with a target network.
- **SwAV: Unsupervised Learning of Visual Features by Contrasting Cluster Assignments** (2020). Online clustering with multi-crop.
- **Barlow Twins: Self-Supervised Learning via Redundancy Reduction** (2021). Alignment with cross-correlation objective.
- **VICReg** (2022). Variance-Invariance-Covariance regularization.
- **VICRegL** (2023). Local features and region-level learning.
- **DINO: Emerging Properties in Self-Supervised Vision Transformers** (2021). Self-distillation with ViTs.
- **DINOv2: Learning Robust Visual Features without Supervision** (2023). Strong ViT features at scale.
- **DINOv3: Scalable Self-Supervised Vision Models** (2025). Further improvements over DINO/DINOv2 for ViT-based SSL (e.g., Gram anchoring).
- **MAE: Masked Autoencoders Are Scalable Vision Learners** (2021). Masked image modeling with lightweight decoder.
- **iBOT: Image BERT Pre-Training with Online Tokenizer** (2022). Token-level SSL with ViTs.
- **EMP-SSL: Extreme-Multi-Patch Self-Supervised Learning** (2023). One-epoch SSL via many image patches/crops.
- **WebSSL: Scaling Language-Free Visual Representation Learning** (2025). Self-supervised learning at web scale without language supervision.