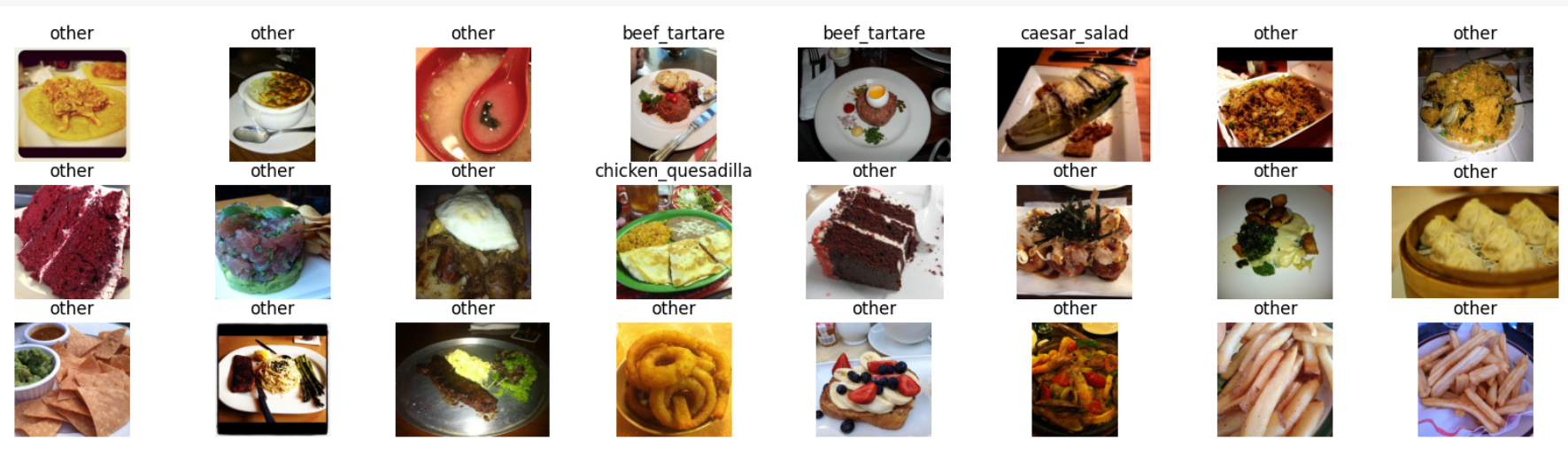
**DLP PROJECT REPORT**



**Deep Self-Learning From Noisy Labels**

****

**Group Members:**

**21k-4696 Sahil Kukreja**

**21k-3410 Rohan Kumar**

**21k-4821 Lov Kumar**

**Objective:**

The objective of this project is to implement and evaluate a noise-robust deep learning framework for image classification on real-world noisy datasets, specifically Clothing1M and Food-101. By integrating the Semantic Mean Prototypes (SMP) strategy within a Deep Self-Learning (DSL) framework, the model aims to automatically refine its training labels through pseudo-labeling based on dense feature representations. This self-correcting mechanism enhances label reliability without requiring manually cleaned annotations, leading to improved classification performance in the presence of web-sourced or weakly supervised noisy labels.

**Problem Statement:**

Large image datasets like Clothing1M and Food-101 often have incorrect or noisy labels because they are collected from the web or labeled automatically. This noise makes it hard for deep learning models to learn correctly, as they tend to memorize these wrong labels, which hurts their performance. Many existing methods try to fix this but require clean data, knowledge of how much noise there is, or complicated training processes. Our goal is to create a simple and effective way to train models that can handle noisy labels on their own, without needing clean labels or extra information. We aim to do this by using a strategy that finds the most common features for each class (called prototypes) to help correct the noisy labels and improve training.

**Methodology:**

**Datasets**

* **Clothing1M**:
  + 1 million training images with real-world noisy labels.
  + For efficient experimentation, 20,000 randomly selected training samples were used.
  + Evaluation was conducted on 10,000 clean test images.
* **Food-101**:
  + 75,750 training images across 101 food categories with weak supervision and noise.
  + 25,250 clean test images were used for evaluation.

#### **Data Preprocessing**

* **Resizing**: All images were resized to 256×256 pixels.
* **Cropping**: A center crop of 224×224 was applied.
* **Normalization**: Images were normalized using mean = [0.5, 0.5, 0.5] and std = [0.5, 0.5, 0.5].
* **Augmentation (DSL only)**:
  + Random Horizontal Flip
  + Color Jitter
  + Random Crop (optional)

#### **Base Model Setup**

A ResNet architecture pretrained on ImageNet was fine-tuned for both datasets:

* **Clothing1M**: Final FC layer modified to output 14 logits.
* **Food-101**: Final FC layer outputs 101 logits.

model = torchvision.models.resnet50(pretrained=True)

model.fc = nn.Linear(model.fc.in\_features, num\_classes)

The penultimate feature vector (2048-dimensional) was extracted to compute semantic prototypes for each class.

#### **Training Strategy (SMP + DSL Fusion)**

Training proceeds in two phases:

##### Phase A: Warm-up

* Epochs 1–3: Train using only original noisy labels.
* Loss function: Standard CrossEntropy.

##### Phase B: Label Fusion Training

* Epochs 7–10: Combine noisy and pseudo-labels using a label fusion loss

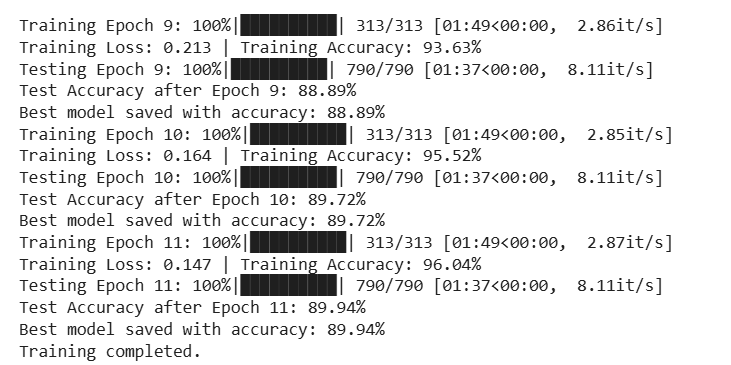
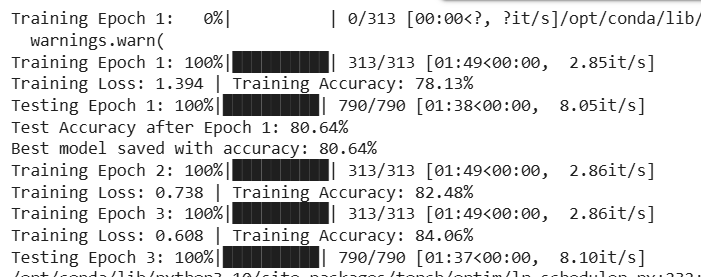
**Optimization**:

Optimizer: SGD, Learning Rate: 0.01, Momentum: 0.9, Weight Decay: 5×10e-4

**Results**

**For Food101N:**

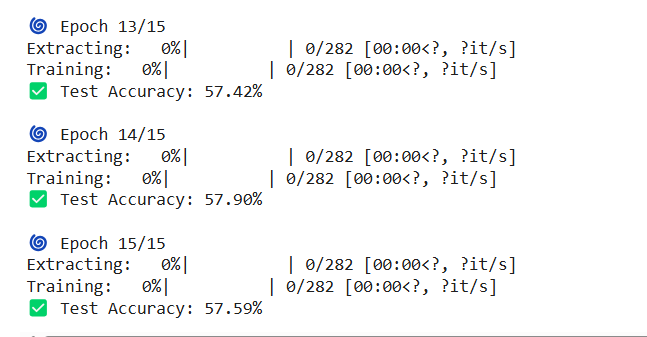
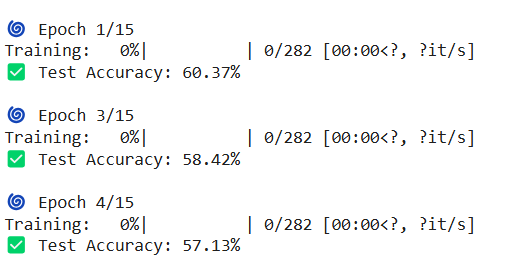
* The model demonstrated strong generalizationdespite noisy labels, improving steadily in both training and testing.
* Peak test accuracy reached 89.94% at Epoch 11, reflecting high noise robustness, likely due to the large dataset size and effective optimization.
* The training loss consistently dropped while accuracy rose, indicating stable convergence.

****

**For Clothing1M:**

The model was trained using 20,000 noisy training images and evaluated on 10,000 clean test samples. The experiment followed a 15-epoch schedule, with the first 5 epochs using only noisy labels (α = 0), and the remaining epochs incorporating SMP-based pseudo-labels (α = 0.5).

* Warm-up phase showed a slight decline in accuracy, potentially due to noisy supervision.
* After SMP-based label correction began (Epoch 6), accuracy initially dropped due to unstable pseudo-labeling, but gradually improved with more refined predictions.
* Final accuracy (**57.59%**) reflects the ability of SMP and DSL to partially recover from label noise, despite using a small subset (20k of 1M) for training.

****

**References:**

[1] Giorgio Patrini, Alessandro Rozza, Aditya Krishna Menon, Richard Nock, and Lizhen Qu. Making deep neural networks robust to label noise: A loss correction approach. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 1944–1952, 2017.

[2] Kaur, A., Kumar, M., & Kaur, P. (2021). "FoodINN: A Food Image Classification Approach Using Deep Convolutional Neural Networks." *Elsevier Journal on Computational Vision*.

[3] Han, Bowen, et al, "Deep Self-Learning From Noisy Labels." Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 5138–5147.

[4] Kuang-Huei Lee, Xiaodong He, Lei Zhang, and Linjun Yang. Cleannet: Transfer learning for scalable image classi f ier training with label noise. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 5447–5456, 2018.

[5] Mengye Ren, Wenyuan Zeng, Bin Yang, and Raquel Urta sun. Learning to reweight examples for robust deep learning. arXiv preprint arXiv:1803.09050, 2018.