Enhancing Visual Encoder Self-Supervised Pre-training with Out-of-Distribution Detection

Lake Wang, Ziyue (Tom) Zhou, Patrick Su
Mentor: Mengye Ren, Chris Hoang

Introduction

Self-Supervised Learning (SSL) struggles with long-tailed datasets, where rare critical samples are overshadowed by frequent ones. This limits performance in safety-critical tasks like autonomous driving. We address this by integrating **Out-of-Distribution (OOD) detection** to SSL frameworks. Our contributions in this study are:

- An OOD detection module to balance rare samples.
- A dynamic memory buffer for rare sample integration.
- Improved representation learning in long-tailed datasets.

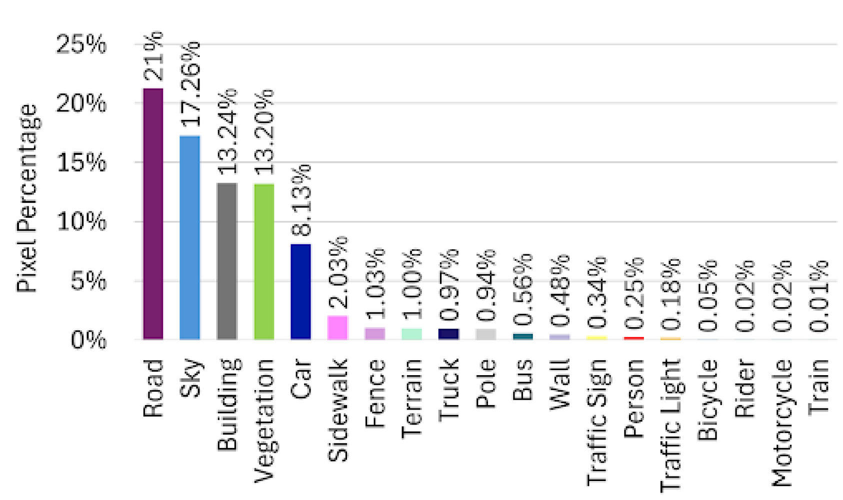


Figure 1. Long-Tailed Distribution

Figure 2. Autonomous Driving Dataset Sample

Visual Analysis and Proof of Concept

To validate the OOD detection module, we analyzed feature embeddings and identified rare samples.

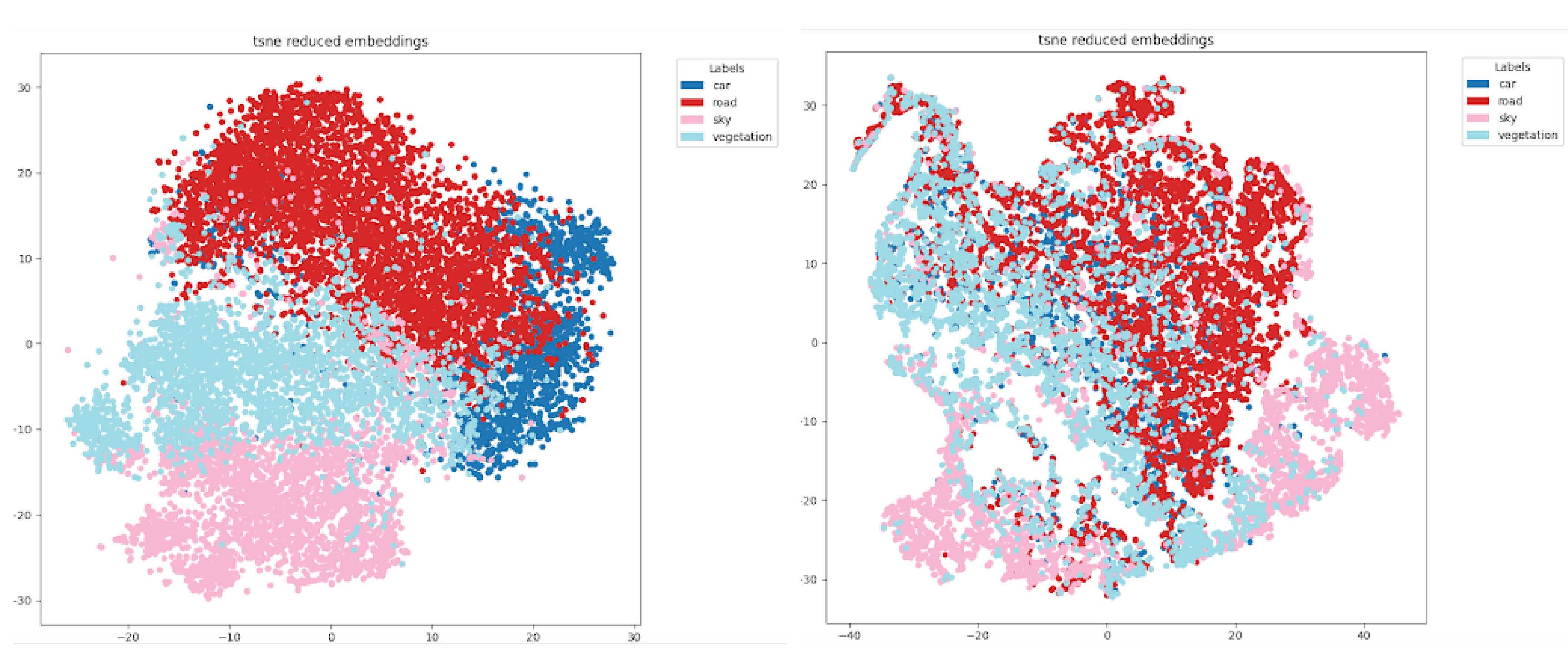
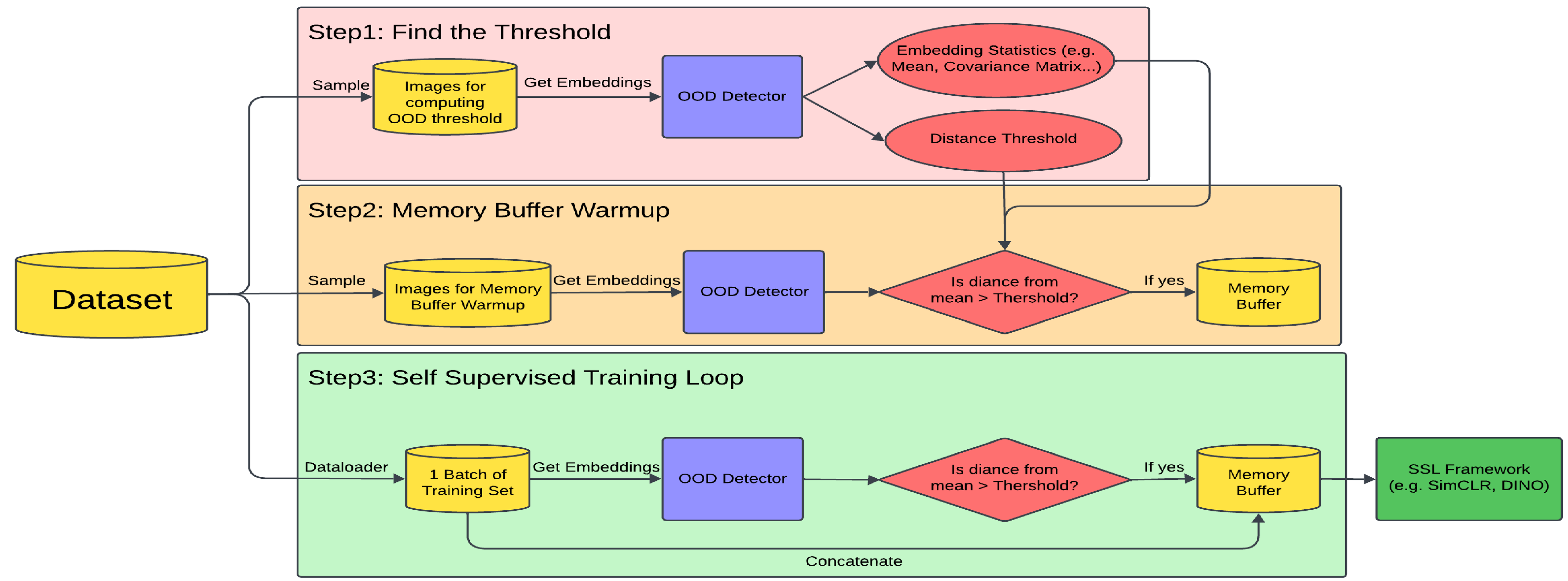


Figure 3. Clustering of embeddings from ImageNet (left) and BDD weights (right)

Key Observations: Common classes form tight, distinct clusters in the embedding space, demonstrating the model's ability to group semantically similar objects effectively.



Out-of-Distribution Detection Pipeline

- **Augmentations**: Random resized crops including global crops and local crops are combined to generate training samples.
- **Threshold Calculation**: Embedding statistics determine OOD thresholds for identifying rare samples.
- **Memory Buffer**: A FIFO buffer dynamically manages rare samples, balancing underrepresented instances.
- **Batch Combination**: Rare samples from the buffer are combined with regular batches for balanced SSL training.

Future Work

- Explore segmentation models as the OOD detector to improve feature extraction and rare object detection.
- Implement real-time threshold updates using online learning to adapt to evolving data distributions.
- Address labeling inaccuracies by refining pixel-level labels with robust networks.

Experiment Setup and Results

The experiments are conducted using the BDD10K dataset with per-pixel semantic labels to evaluate the OOD detector.

Precision (the proportion of true rare samples among detected OOD samples) is the most critical metric in evaluating performance for imbalanced datasets. It ensures that the memory buffer is populated with truly rare samples.

Pre-trained Model	Training Set Size	Threshold Percentile	Accuracy	Precision	Recall	F1
Random Detector	–	–	0.1920	0.1580	–	–
DINO-ImageNet	256	0.99	0.7868	0.1861	0.0338	0.0571
DINO-ImageNet	256	0.95	0.7601	0.1628	0.0612	0.0890
DINO-ImageNet	256	0.90	0.7299	**0.1925**	0.1287	0.1543
DINO-ImageNet	256	0.80	0.6483	0.1780	0.2316	0.2013
DINO-ImageNet	1024	0.99	0.7982	**0.2101**	0.0196	0.0359
DINO-ImageNet	1024	0.95	0.7733	0.1916	0.0573	0.0882
DINO-ImageNet	1024	0.90	0.7372	0.1829	0.1075	0.1354
DINO-ImageNet	1024	0.80	0.6714	0.1841	0.2088	0.1957
DINO-BDD100K	1024	0.99	0.7975	0.1696	0.0149	0.0274
DINO-BDD100K	1024	0.95	0.7643	0.1959	0.0746	0.1080
DINO-BDD100K	1024	0.90	0.7296	0.2011	0.1389	0.1643
DINO-BDD100K	1024	0.80	0.6708	**0.2021**	0.2441	0.2211