Novel Class Discovery

Trends and Applications of Computer Vision

Noemi Canovi (230487), Kevin Depedri (229358),
Jacopo Donà (229369), Mostafa Haggag (229674)
University of Trento

21st December 2022

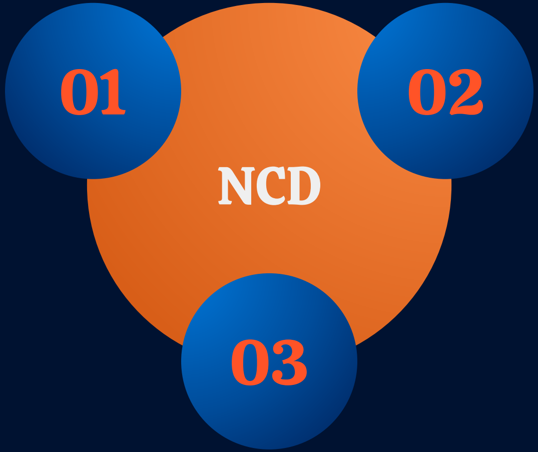
Outline

01

Introduction

Introduction to Novel Class Discovery (NCD) task

Putting together related tasks



Semi-Supervised Learning

Tackle the problem of missing labels

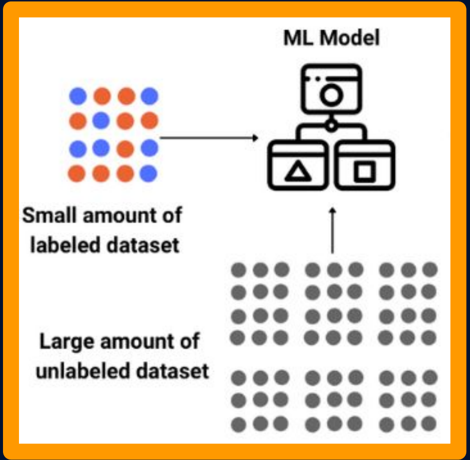
Transfer Learning

Leverage features from pre-trained models

Unsupervised Clustering

Find similarities between samples

Semi-Supervised Learning (SSL)

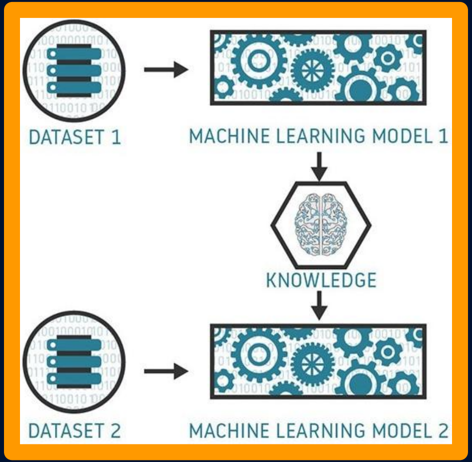


The goal is to solve a classification problem in which part of the data is labeled and the rest is not.

We assume that the classes in the labeled and unlabeled sets are the same.

In NCD this assumption does not hold. The classes between the two sets are disjoint. The unlabeled classes does not have any corresponding sample in the labeled dataset.

Transfer Learning

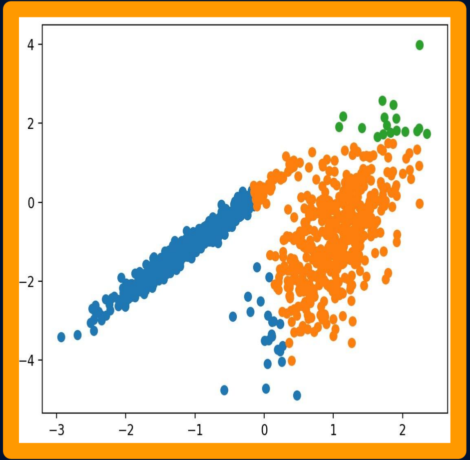


The goal is to take a model pre-trained over another (possibly bigger) dataset and leverage its features representation.

This model is then fine-tuned on task specific labeled data.

In NCD this is not fully possible. No labeled dataset is available for the new classes.

Unsupervised Clustering

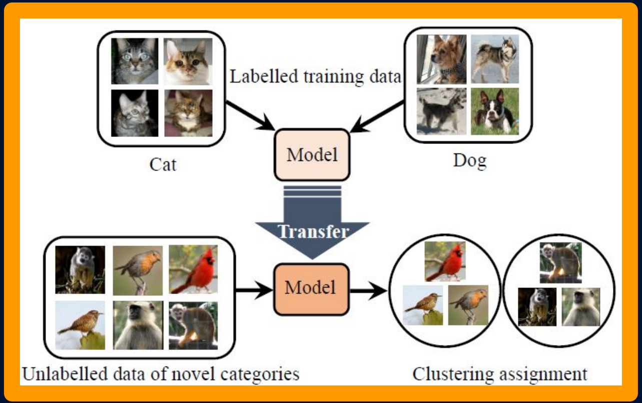


The goal is to partition unlabeled samples into different categories.

No labels are given. Therefore, samples are aggregated into clusters depending on their similarities.

In NCD this assumption is relaxed, since, labeled source data allows to learn features and pattern to distinguish between different classes.

Formal definition of NCD task



In Novel Class Discovery we assume to have two disjoint dataset:

1. Labeled dataset $D^l = \{(x^l_i, y^l_i), i = 1, ..., N\}$
2. Unlabeled dataset $D^u = \{x^u_i, i = 1, ..., M\}$

We know:

1. The $y^l_i \in \{1, ..., C^l\}$ labels of D^l
2. Number C^u of classes in the unlabeled dataset

We want to train a model that is able to classify images from both the datasets.

02

AutoNovel

K. Han, S. Rebuffi, S. Ehrhardt, A. Vedaldi, A. Zisserman, ICLR 2020

Automatically Discovering and Learning New Visual Categories with Ranking Statistics

Three Stage Pipeline

Pre-training with self-supervision

Supervised Fine-tuning

Joint Objective function minimization

Pre-training phase

Labeled data initialization introduces representational bias.

Replace supervised pre-training with self-supervision.

RotNet

$\{0°, 90°, 180°, 270°\}$

Training on Labeled Data

- Fine-tune the pre-trained model Φ on the labeled data to learn a classifier for known samples
- Compute gradients only for the last macro-block of Φ and classification head to avoid overfitting
- Uses standard Cross-Entropy Loss

Transfer Learning with Rank Statistics

Goal

Compute pseudo-labels to use as ground truth for the novel classes

- Define a similarity notion between pair of samples x_i , x_j
- Rank the feature activations of an image
- If the first k rankings of a pair i , j are the same, images are deemed similar.

$$s_{ij} = \mathbb{1}\left\{\mathrm{top}_k(\Phi(x_i^u)) = \mathrm{top}_k(\Phi(x_j^u))\right\}$$

Transfer Learning with Rank Statistics

- Compute pairwise-similarity for all unlabelled samples

- Apply a second head for the image representation of the unlabelled data:
  number of output neurons = number of novel classes

- To calculate the loss, they compute BCE on the s_{ij} pseudo-labels multiplied by the softmax activations of the unlabelled head

Joint Training

- Model Architecture: shared backbone Φ and two heads
- Fine-tune last block of Φ with the two heads combining the two losses

$$L = L_{CE} + L_{BCE} + \omega(t) L_{MSE}$$

- Consistency component to limit pseudo-labels moving target problem
- Incremental Learning improves accuracy, single classifier

03

UNO

E. Fini, E. Sangineto, S. Lathuili, Z. Zhong, M. Nabi, E. Ricci, ICCV 2021

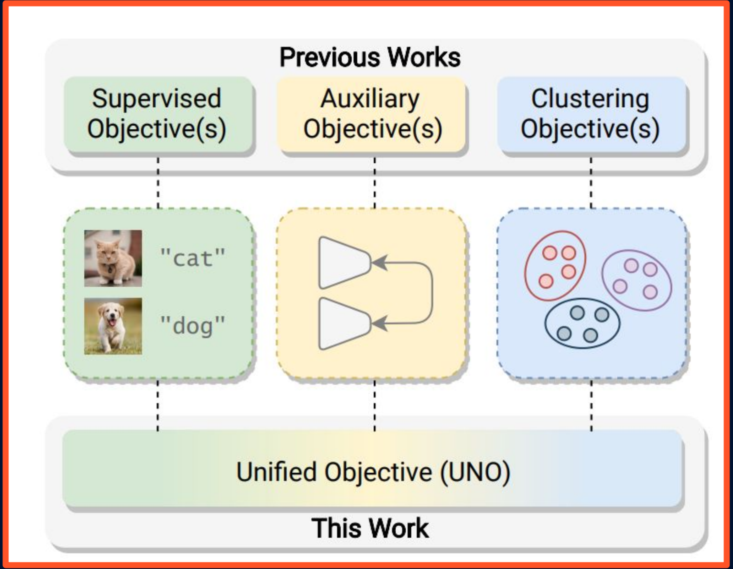
A Unified Objective for Novel Class Discovery

UNO

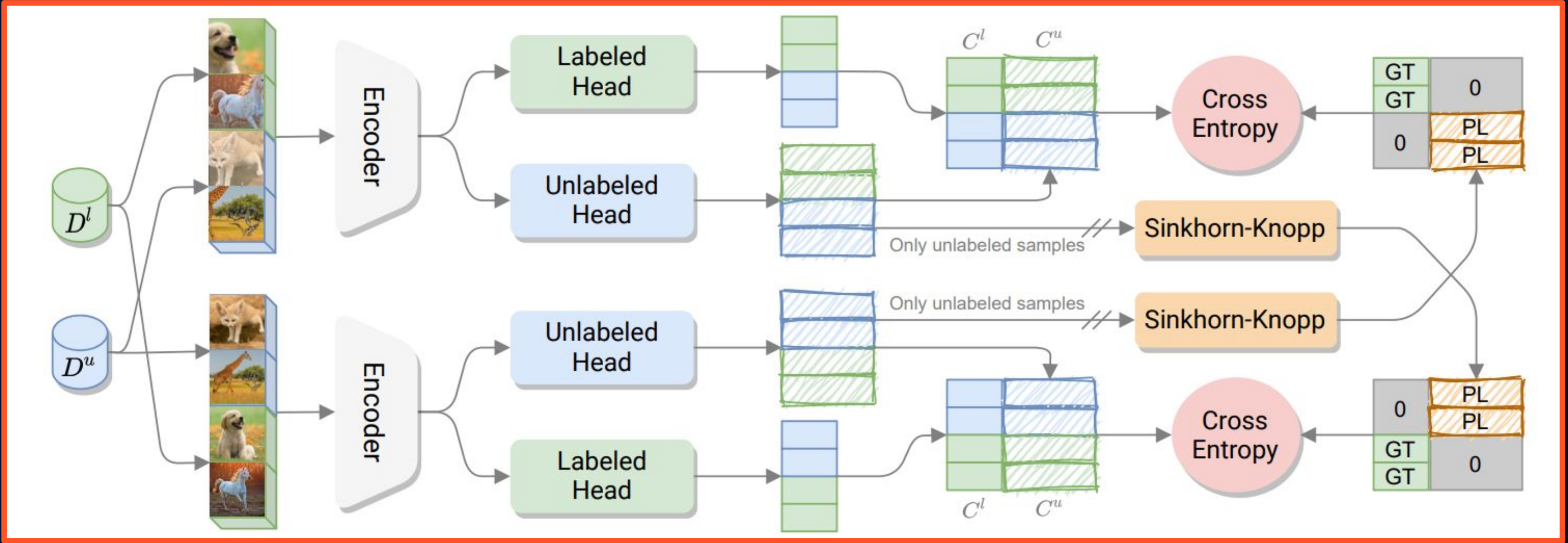
Idea:

**3 Stage Pipeline
is complex and expensive**

Unified Objective

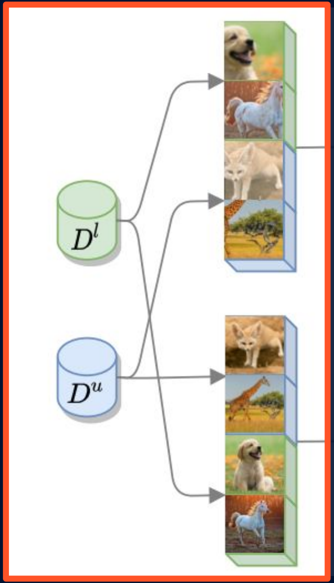


UNO



D^l
D^u
Encoder
Encoder
Labeled Head
Unlabeled Head
Unlabeled Head
Labeled Head
C^l
C^u
Cross Entropy
GT
GT
0
0
PL
PL
Only unlabeled samples
Sinkhorn-Knopp
Only unlabeled samples
Sinkhorn-Knopp
Cross Entropy
0
PL
PL
GT
GT
0
C^l
C^u

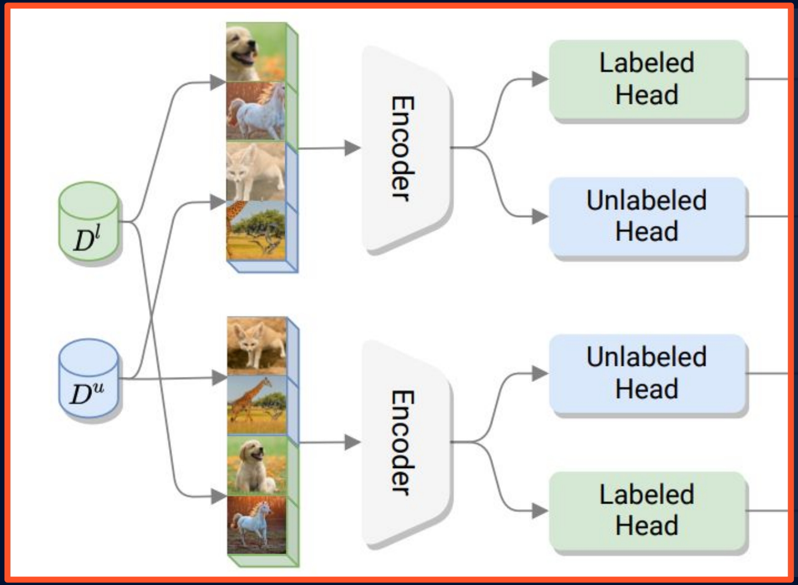
UNO



From each image:

- Two different views v1 and v2 are created
- Done by using random transformations such as cropping and color jittering

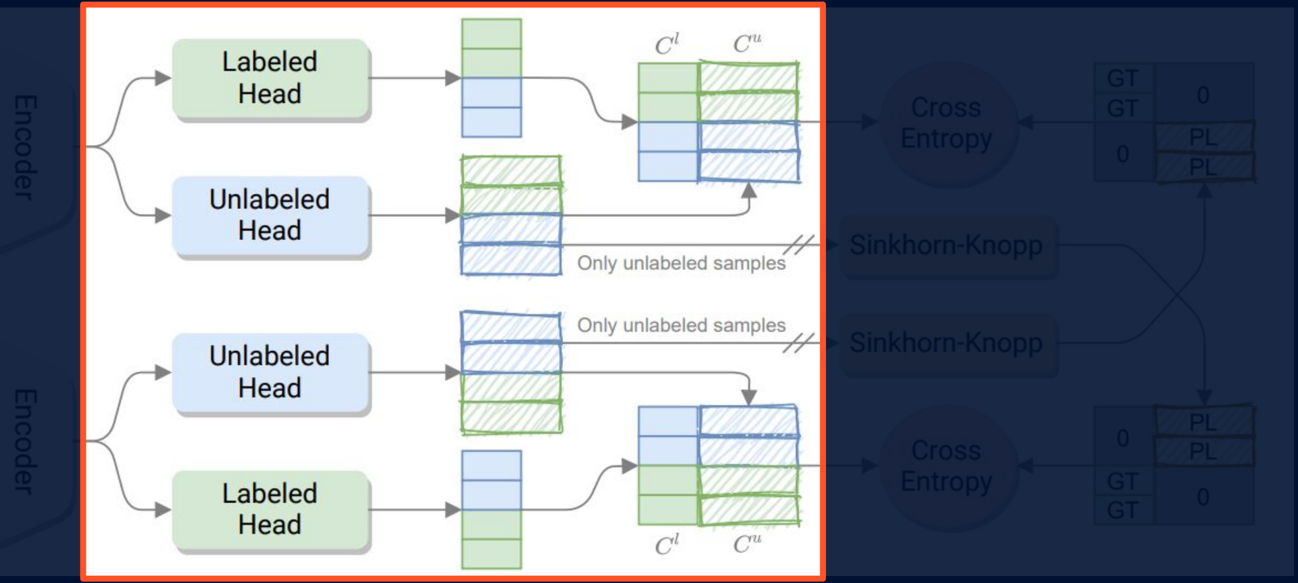
UNO



- Common encoder
- Two different heads:
  - Labeled Head
  - Unlabeled Head

UNO

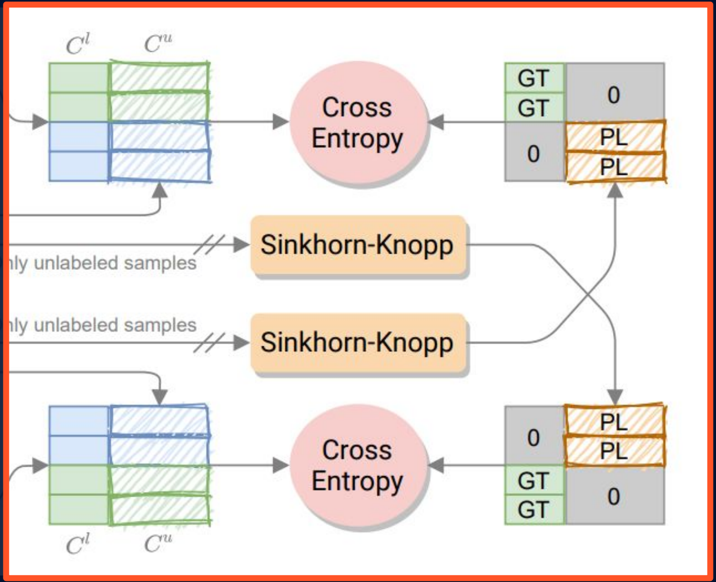
The two heads outputs two logits that are concatenated



UNO

Cross-entropy loss

- For labeled classes:
  ground-truth is associated with the two views
- For unlabeled classes:
  The two views are used to compute two corresponding pseudo-labels that are swapped



04

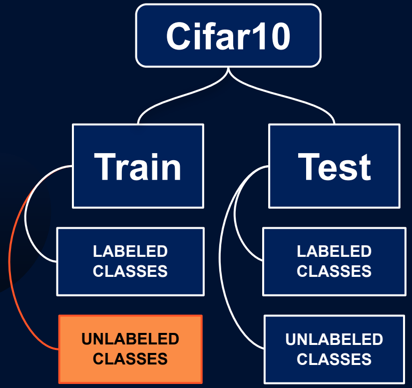
AutoNovel Experiments

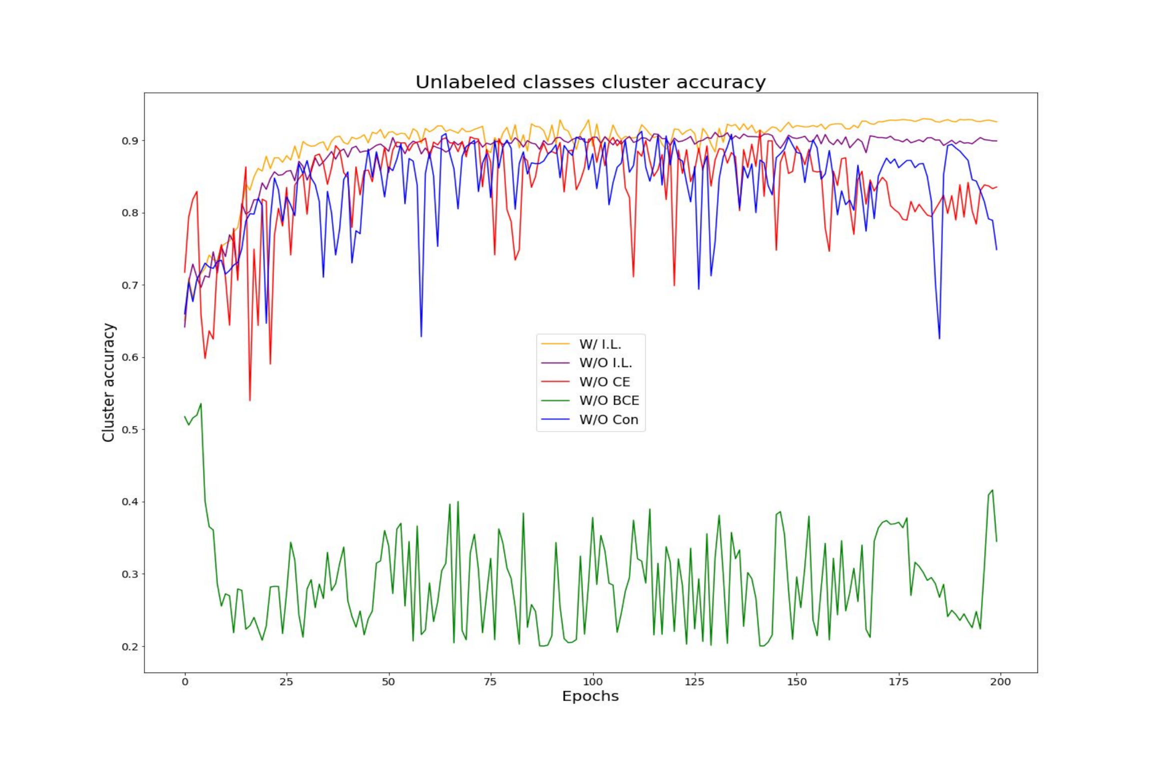
Experiment 1

Removing different loss terms

$$L_{total} = L_{BCE} + L_{MSE} + L_{CE} + L_{IL}$$

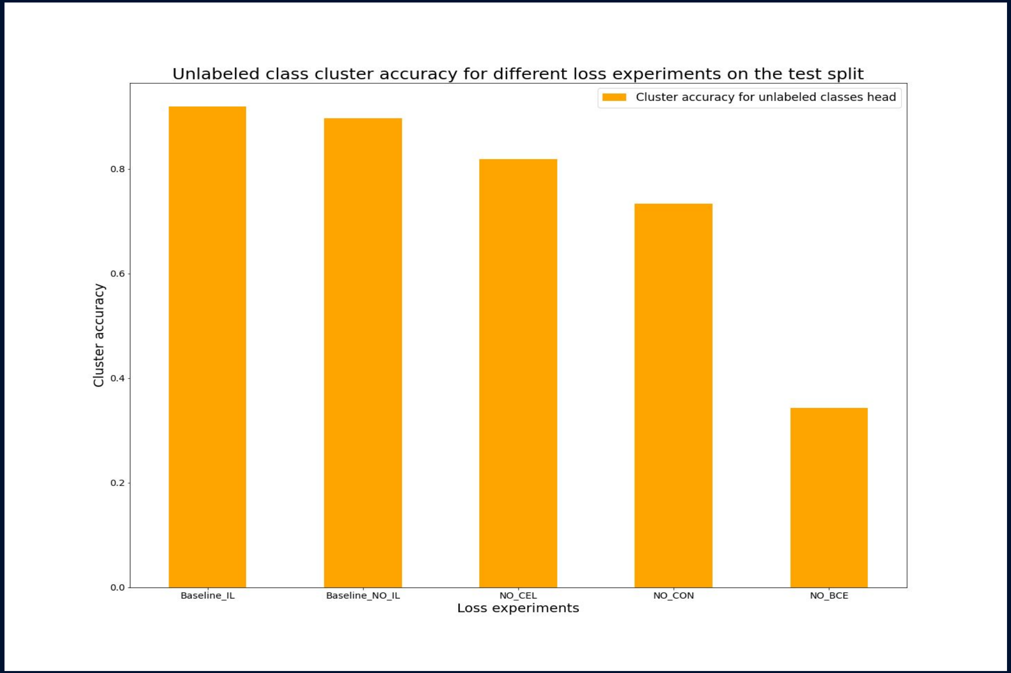
Cluster accuracy





Evaluation on different models

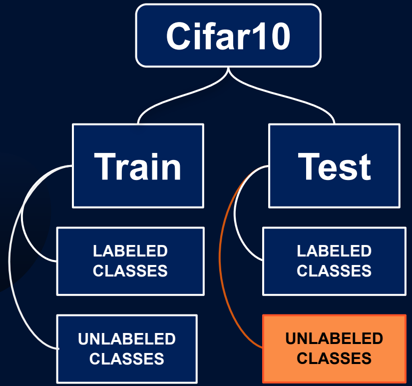
Loss Exp.	All classes cluster accuracy (test set)	Unlabeled classes cluster accuracy (train set)	Unlabeled classes cluster accuracy (test set)
Baseline IL	0.8952	0.92788	0.9196
NO CEL	0.441	0.83656	0.8234
NO BCE	0.563	0.35368	0.3496
No CON	0.822	0.76776	0.7618
No IL	-	0.906	0.9008

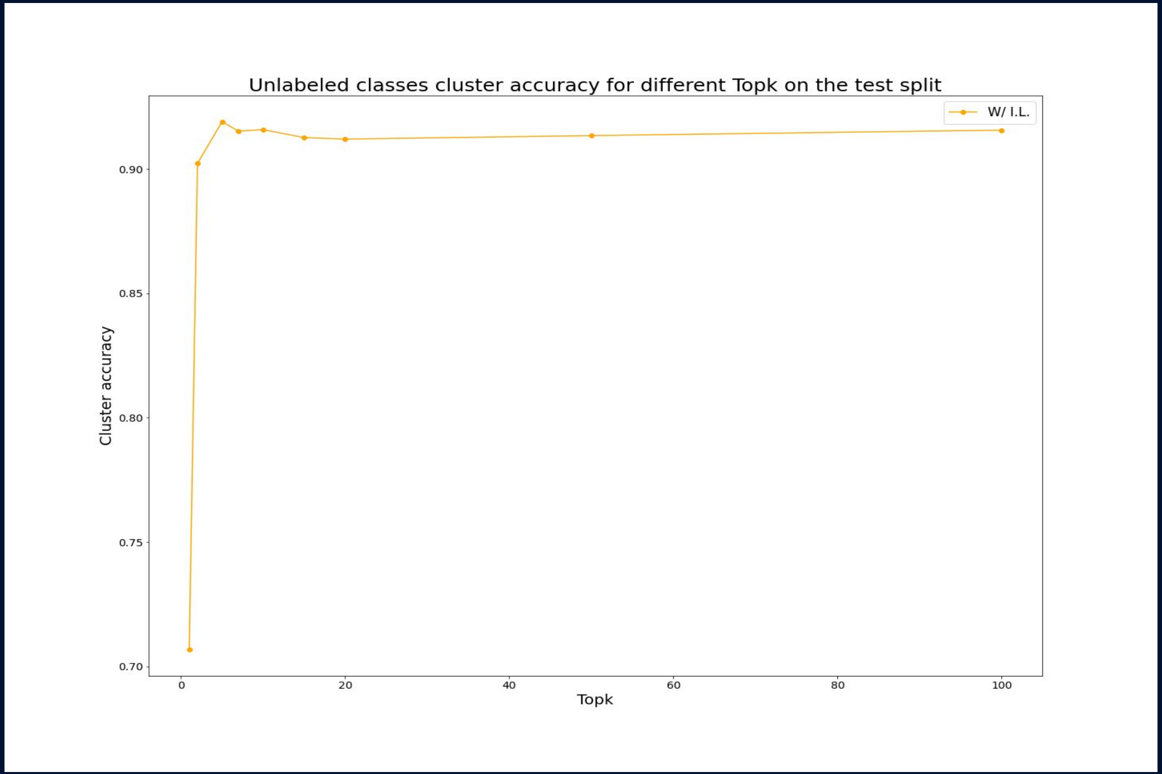


Experiment 2

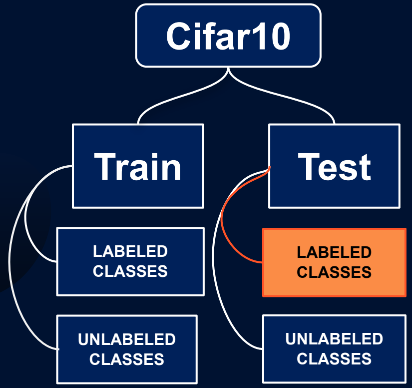
Trying different Topk parameters

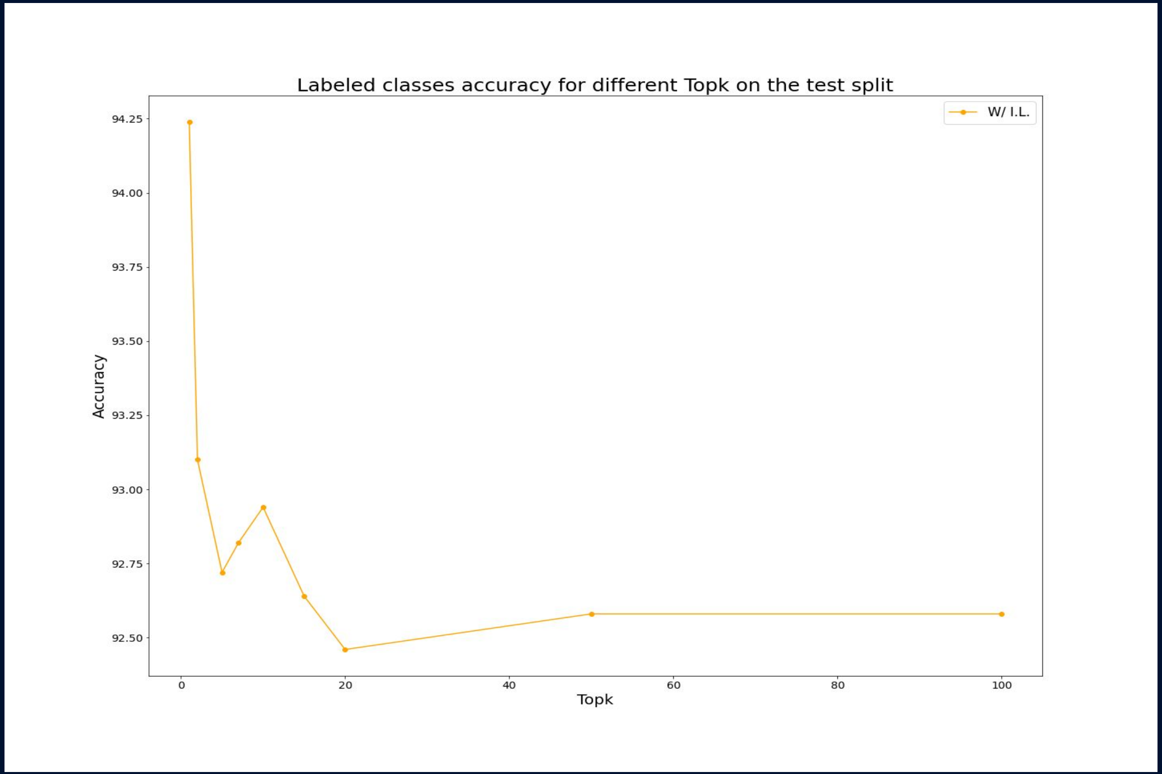
Cluster accuracy





Accuracy

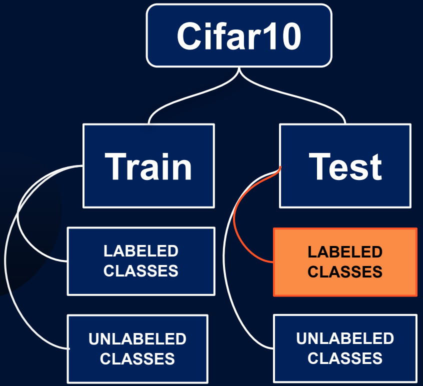

Cifar10
Train
Test
LABELED CLASSES
UNLABELED CLASSES
LABELED CLASSES
UNLABELED CLASSES

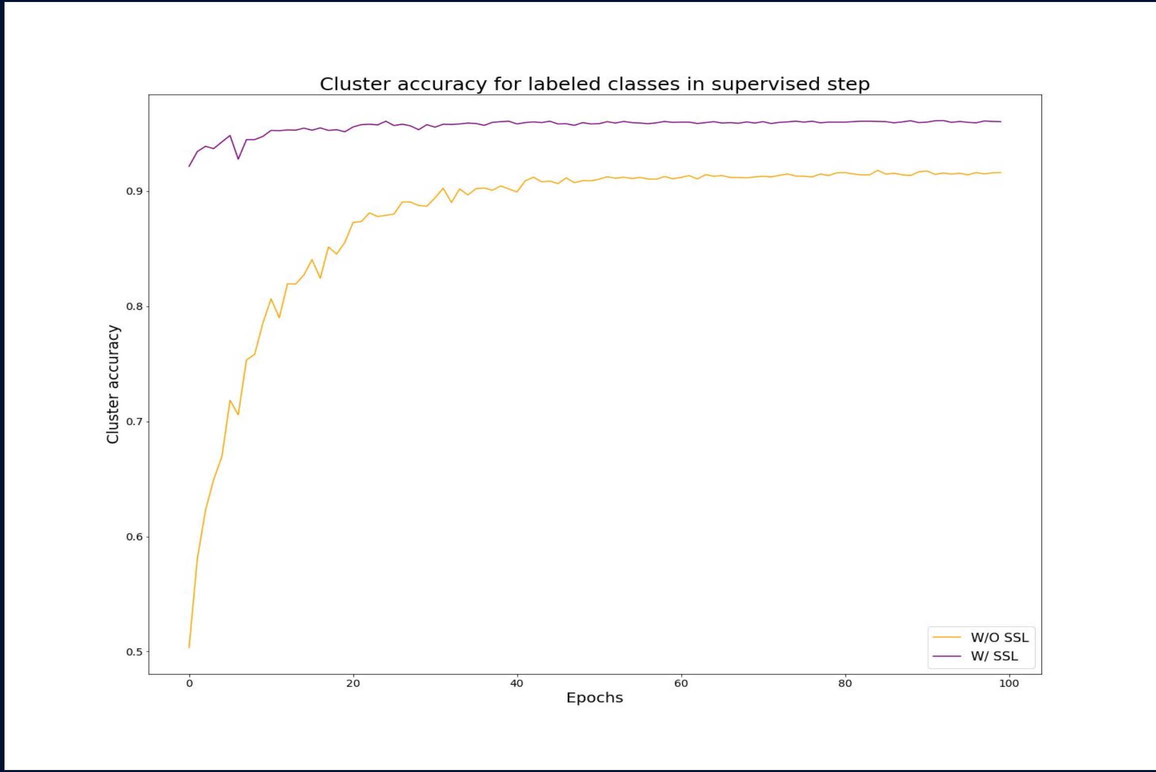

Labeled classes accuracy for different Topk on the test split
W/ I.L.
Accuracy
94.25
94.00
93.75
93.50
93.25
93.00
92.75
92.50
Topk
0
20
40
60
80
100

Experiment 3

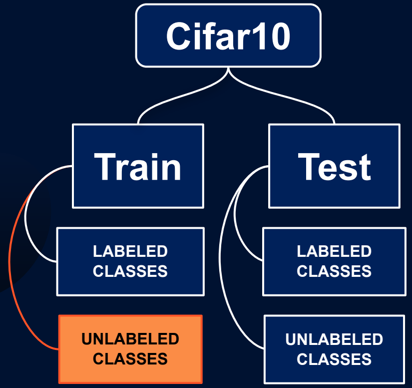
Removing Self-Supervised Learning

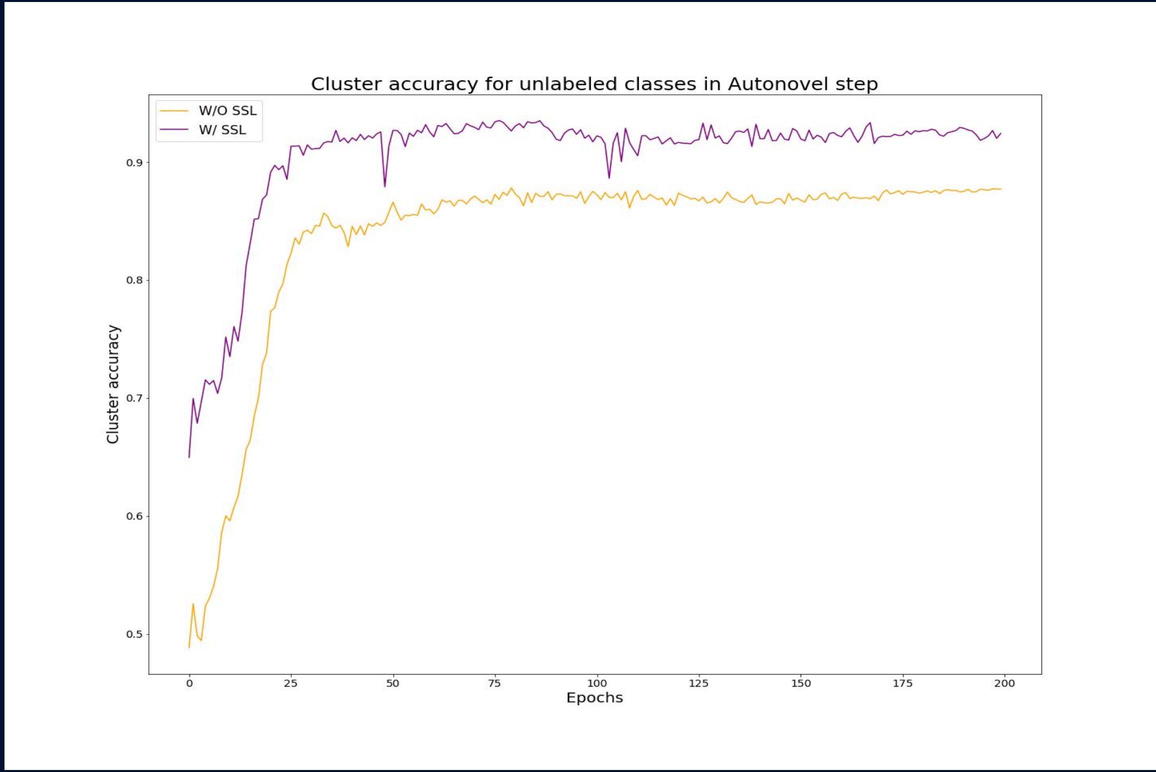
Cluster accuracy in supervised step





Cluster accuracy in Autonovel step





Experiment 4

Using different self supervised Learning

solo-learn: A Library of Self-supervised Methods for Visual Representation Learning

tests passing | docs failing | codecov 88%

solo-learn

Change in Architecture of ResNet-18

Barlow Twins: Self-Supervised Learning via Redundancy Reduction

Jure Zbontar [*1] Li Jing [*1] Ishan Misra [1] Yann LeCun [1 2] Stéphane Deny [1]

Exploring Simple Siamese Representation Learning

Xinlei Chen Kaiming He
Facebook AI Research (FAIR)

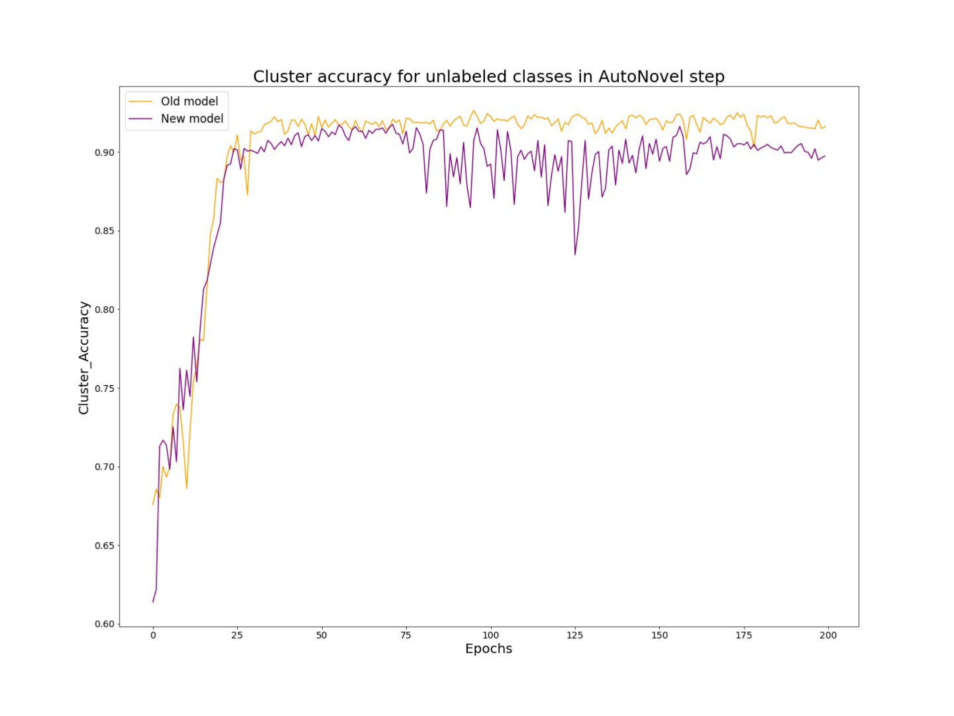
Unsupervised Learning of Visual Features by Contrasting Cluster Assignments

Mathilde Caron[1,2] Ishan Misra[2] Julien Mairal[1]
Priya Goyal[2] Piotr Bojanowski[2] Armand Joulin[2]
[1] Inria* [2] Facebook AI Research

Supervised Contrastive Learning

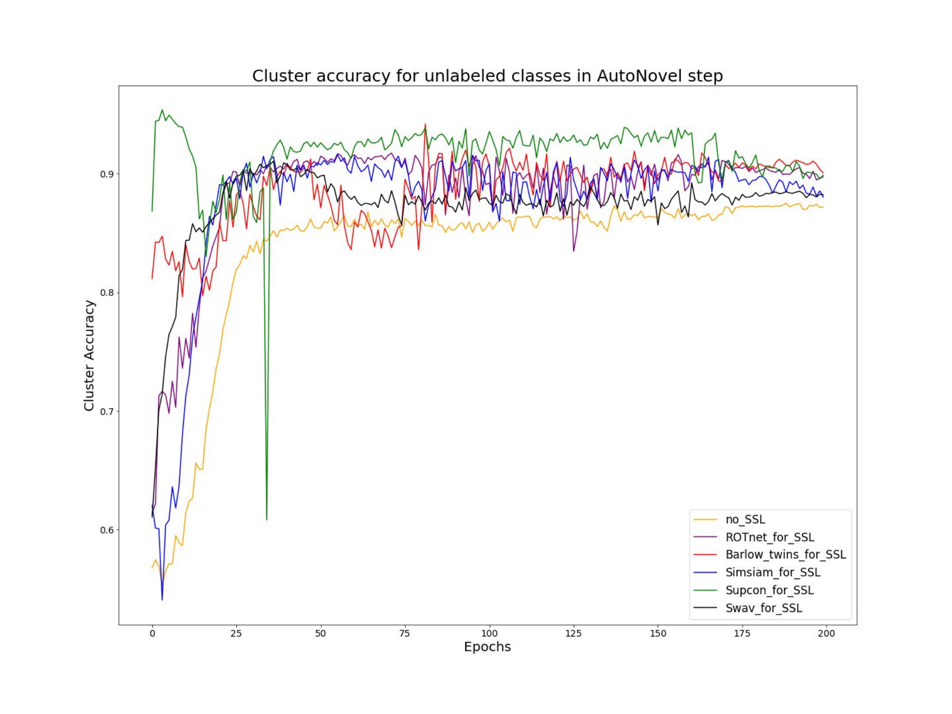
Prannay Khosla [*] Google Research
Piotr Teterwak [*†] Boston University
Chen Wang [†] Snap Inc.
Aaron Sarna [‡] Google Research
Yonglong Tian [†] MIT
Phillip Isola [†] MIT
Aaron Maschinot Google Research
Ce Liu Google Research
Dilip Krishnan Google Research

Difference in architecture



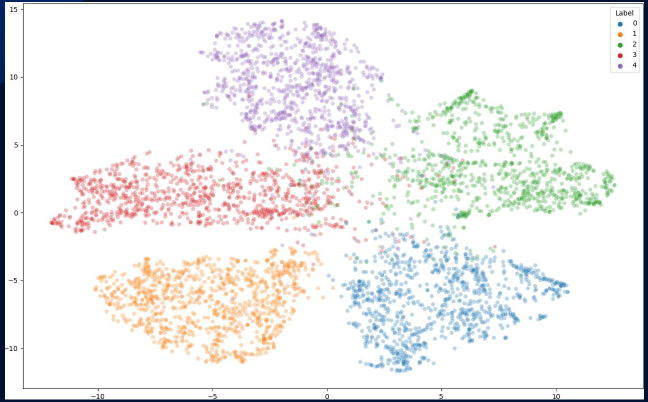
Step	Metrics	Orginal Auto Novel resnet-18	New resnet 18
Self supervised learning	Test accuracy (Test set)	0.9608	0.9535
Supervised learning	Cluster accuracy known classes (Test set)	0.9585	0.9476
Novel class discovery	Cluster accuracy Novel casses (Test set)	0.9146	0.897

Evaluation of different methods



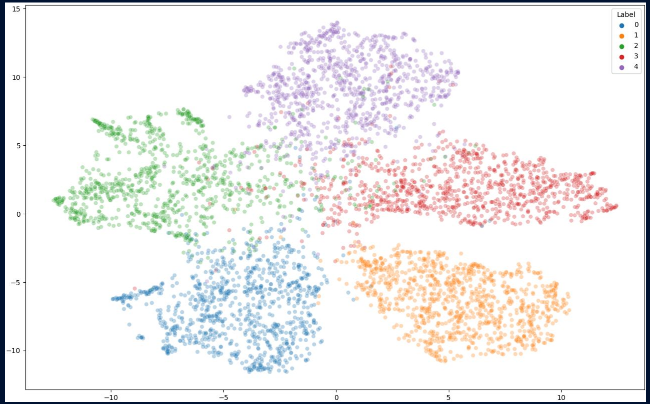
SSL method	All classes cluster accuracy (test set)	Unlabeled classes cluster accuracy (train set)	Unlabeled classes cluster accuracy (test set)
No SSl	0.8192	0.872	0.867
Rot-net	0.8853	0.903	0.897
Barlow twins	0.8964	0.901	0.899
Simsiam	0.8774	0.880	0.874
Supcon	0.8917	0.899	0.887
Swav	0.8659	0.882	0.873

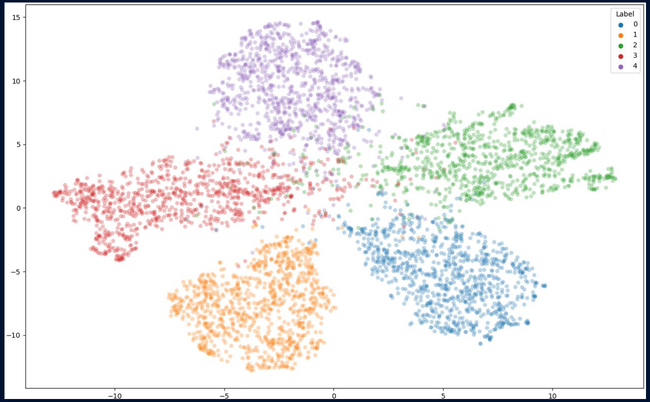
Results of t-SNE – head 1 – labeled



← Barlow Twins

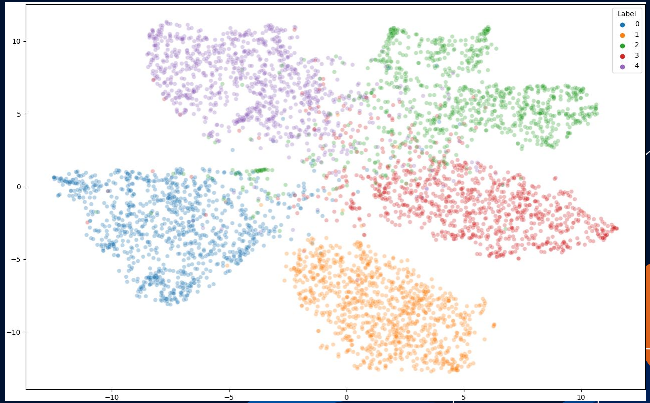
SimSam→



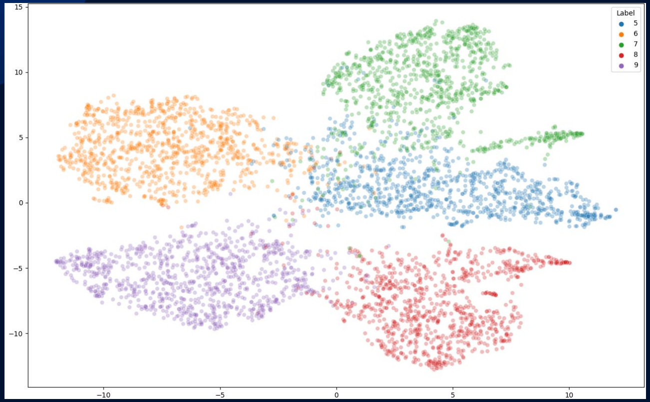


←SupCon

Swav→

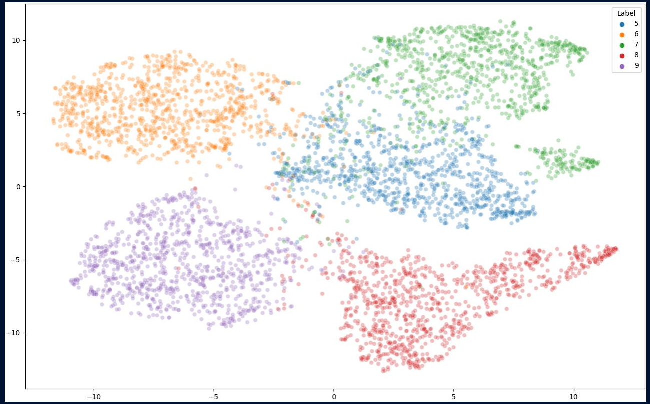


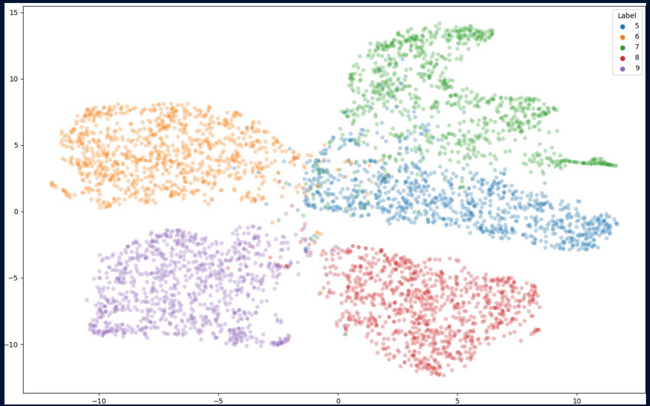
Results of t-SNE - head 2 - unlabeled



← Barlow Twins

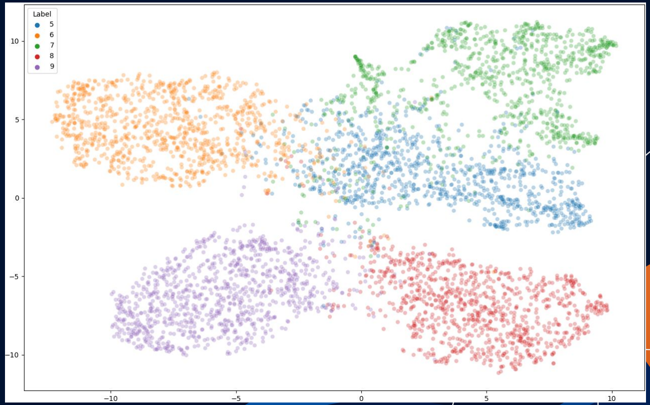
SimSam→





←SupCon

Swav→



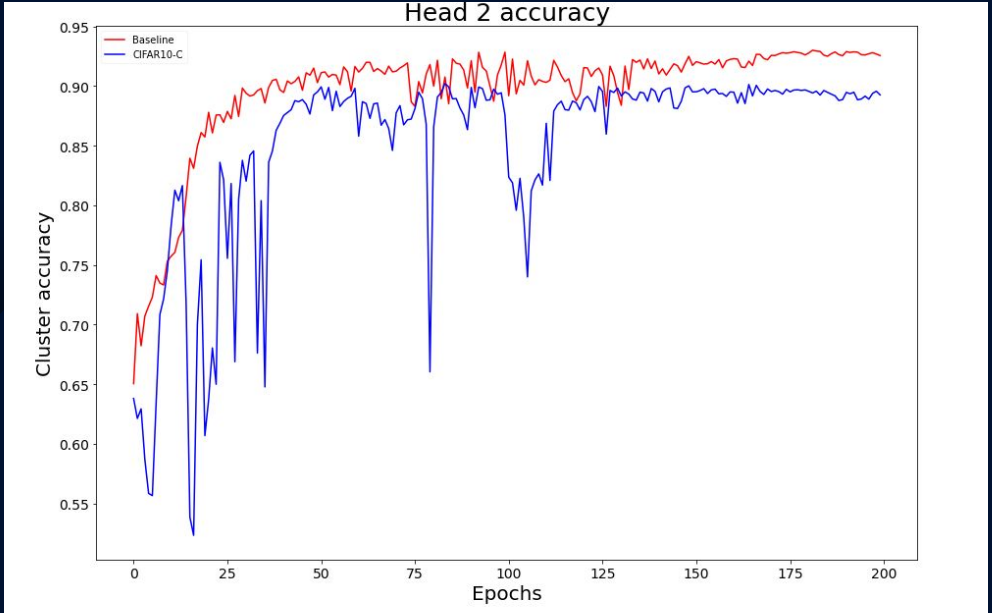
Experiment 5

Domain Shift

Cifar10 Corrupted

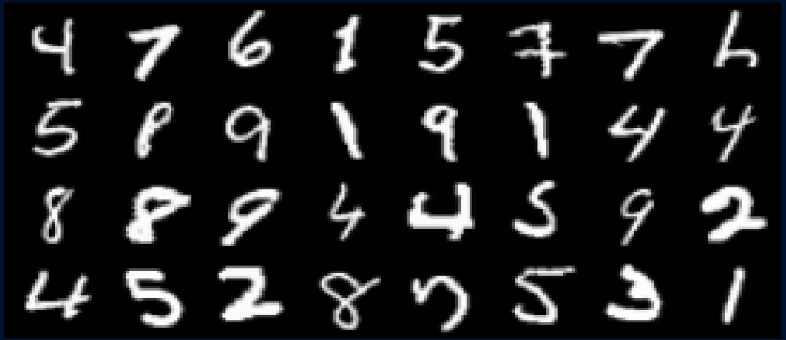
Applied Gaussian Noise to Cifar10 novel classes

Cifar10 -> Cifar10-C

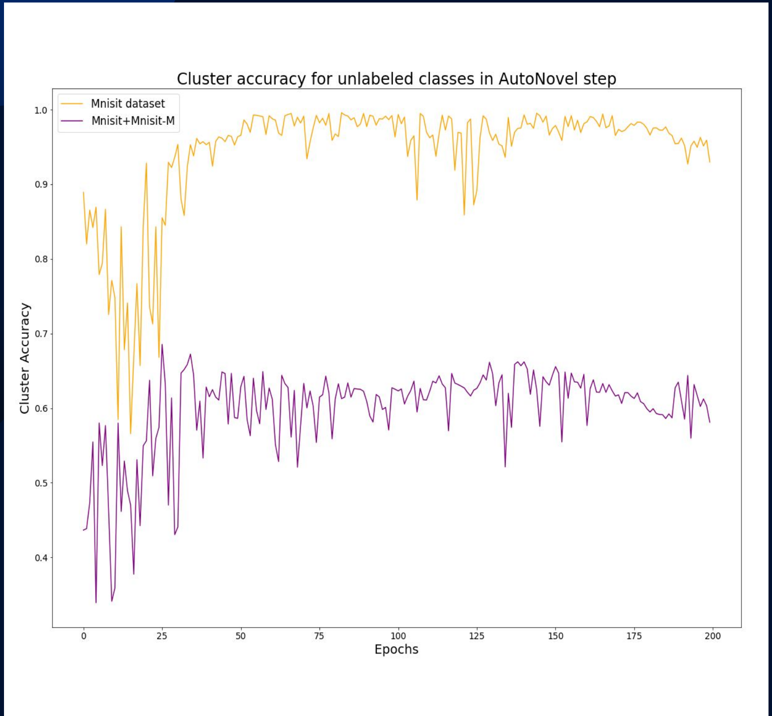


Dataset	All class accuracy (test set)	Unlabeled Classes cluster accuracy (train set)	Unlabeled Classes cluster accuracy (test set)
Cifar10	0.895	0.927	0.919
Cifar10 + Cifar10-C	0.919	0.892	0.879

MNIST-M Dataset



MNIST -> MNIST-M

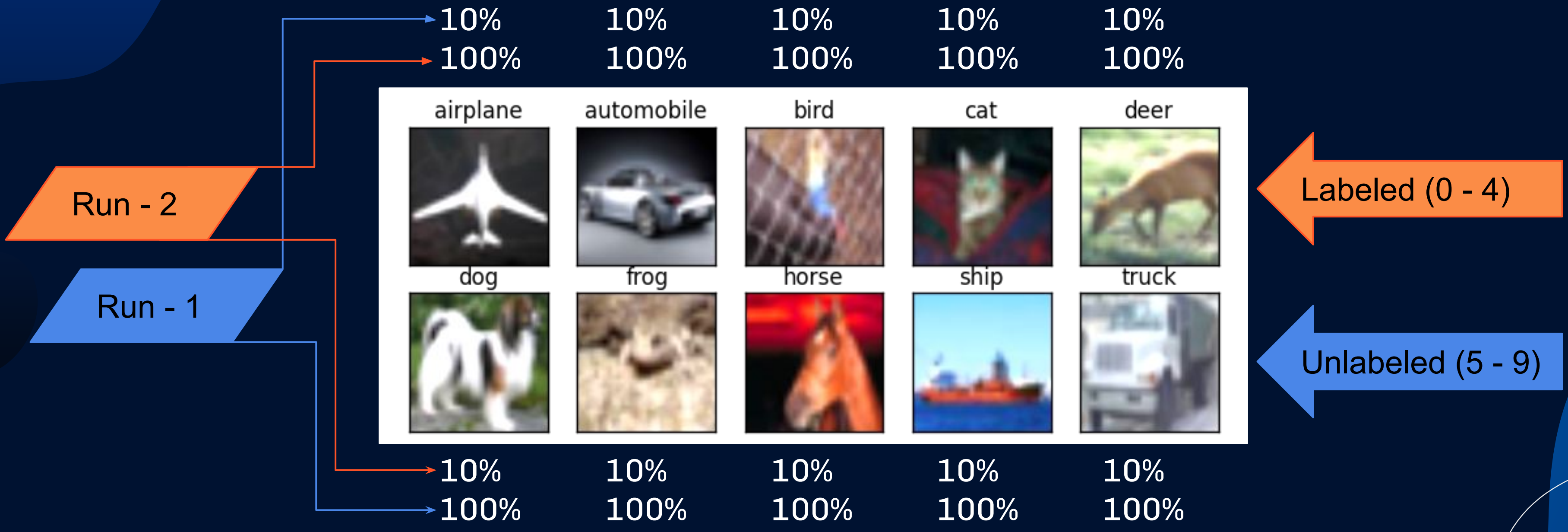


Dataset	All classes cluster accuracy (test set)	Unlabeled classes cluster accuracy (train set)	Unlabeled classes cluster accuracy (test set)
Mnist (0-9)	0.9721	0.930	0.924
Mnist (0-4) + Mnist-M (5-9)	0.795	0.581	0.573

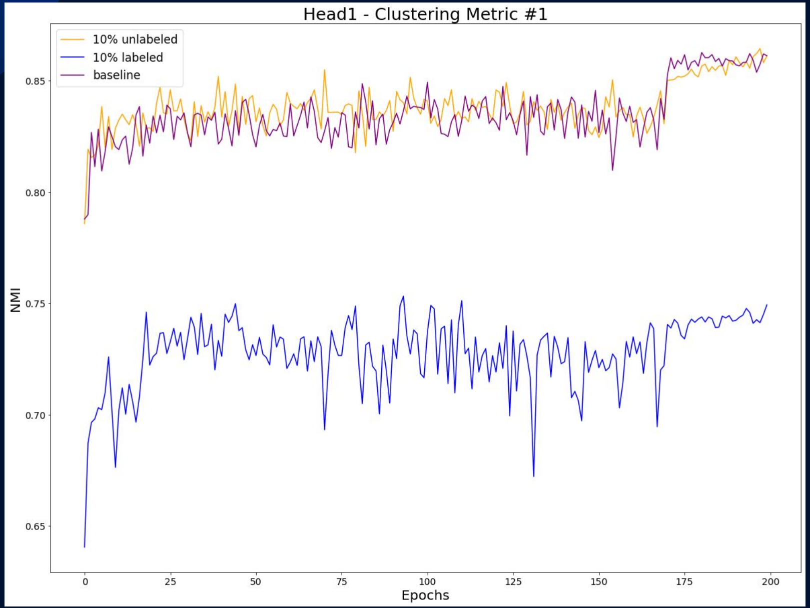
Experiment 6

Unbalanced classes

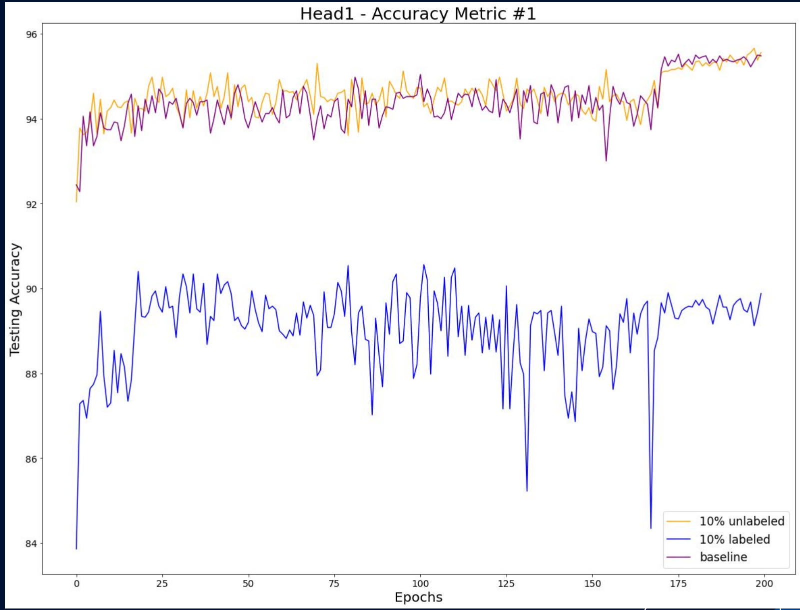
Slicing CIFAR-10



Accuracy comparison – Head 1

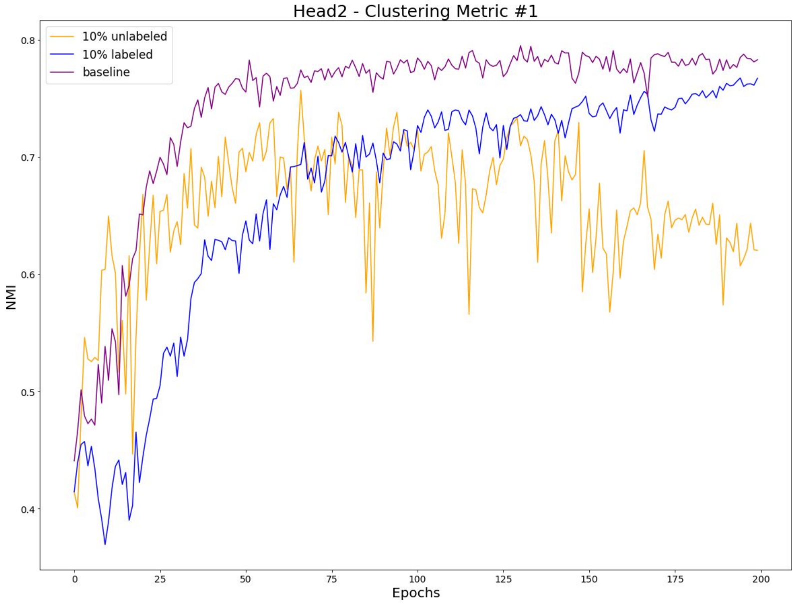


On head 1 working with less unlabeled samples has no important negative effects wrt. working with less labeled ones

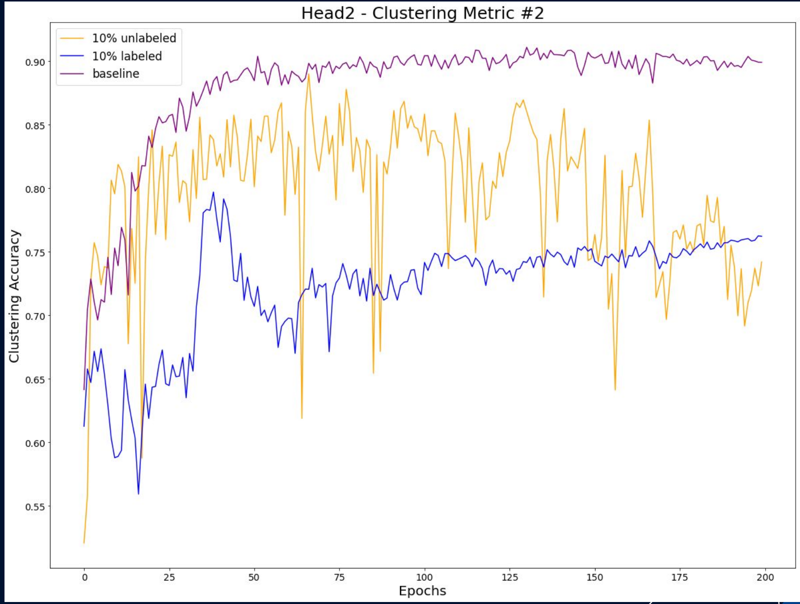


Labeled samples are particularly important to train head1 correctly

Accuracy comparison – Head 2

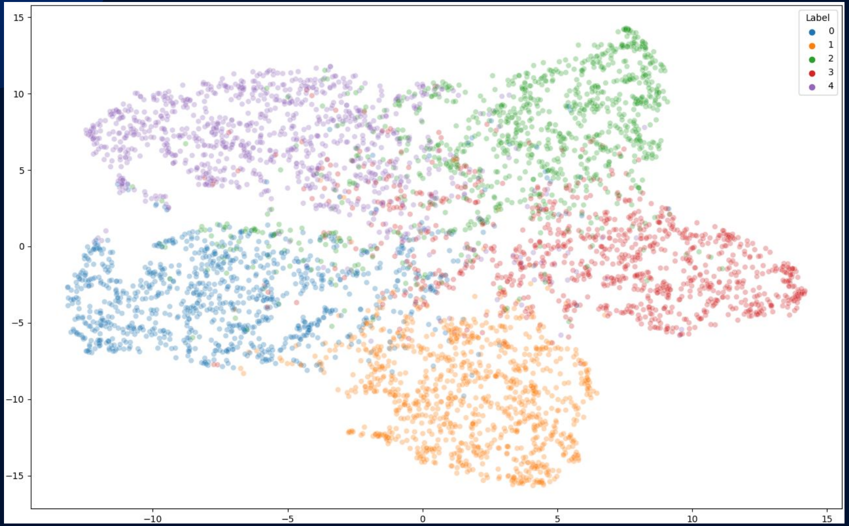


On head 2 working with less labeled or unlabeled samples has a medium/major negative effect



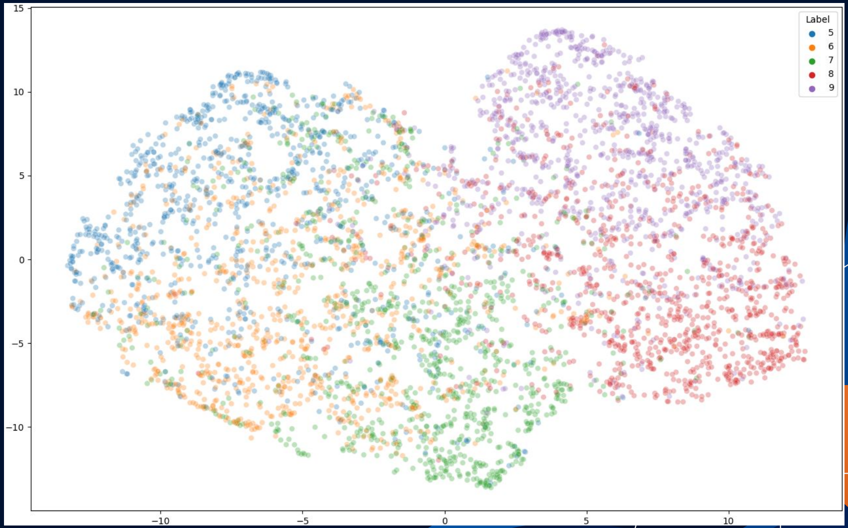
Both the type of samples are important to learn to discriminate between new and old classes

Results of t-SNE – Run 1 – 10% labeled



← Head 1, labeled classes test split

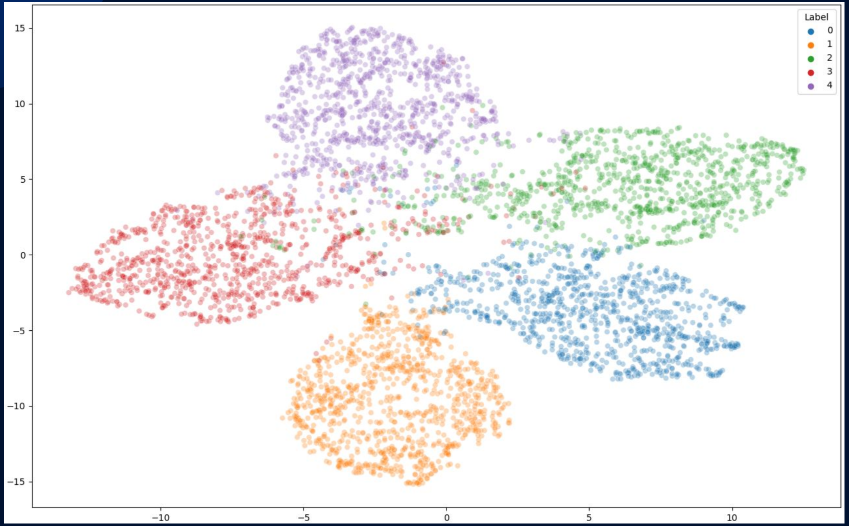
- Only 10% of the labeled samples have been used in the training



- All the unlabeled samples have been used in the training

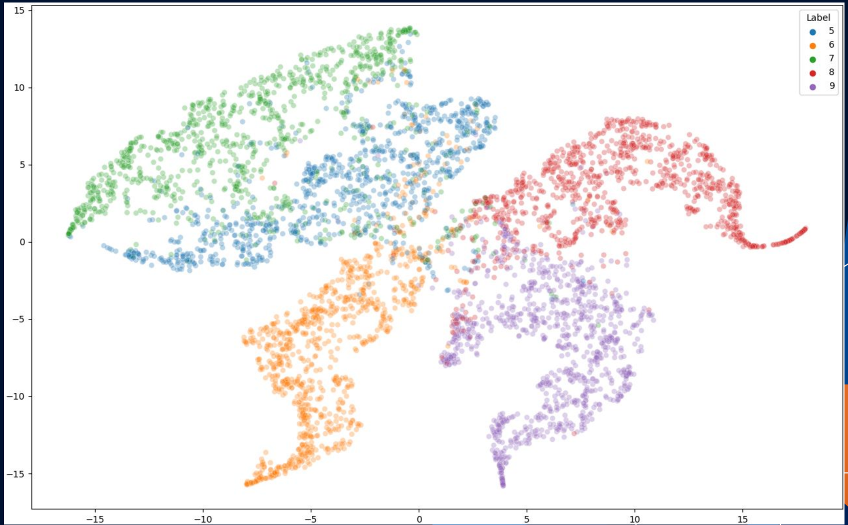
Head 2, unlabeled classes test split →

Results of t-SNE - Run 2 - 10% unlabeled



← Head 1, labeled classes test split

- All the labeled samples have been used in the training



- Only 10% of the unlabeled samples have been used in the training

Head 2, unlabeled classes test split →

Experiment 7

Different number of Lab-Unlab classes

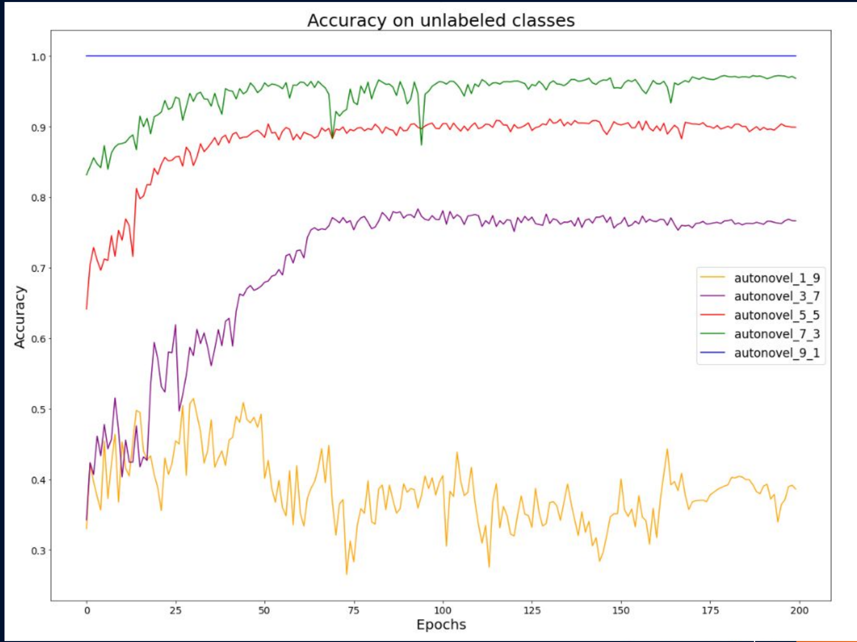
Unbalance on the number of classes

We tried different values for the number of labeled and unlabeled classes.

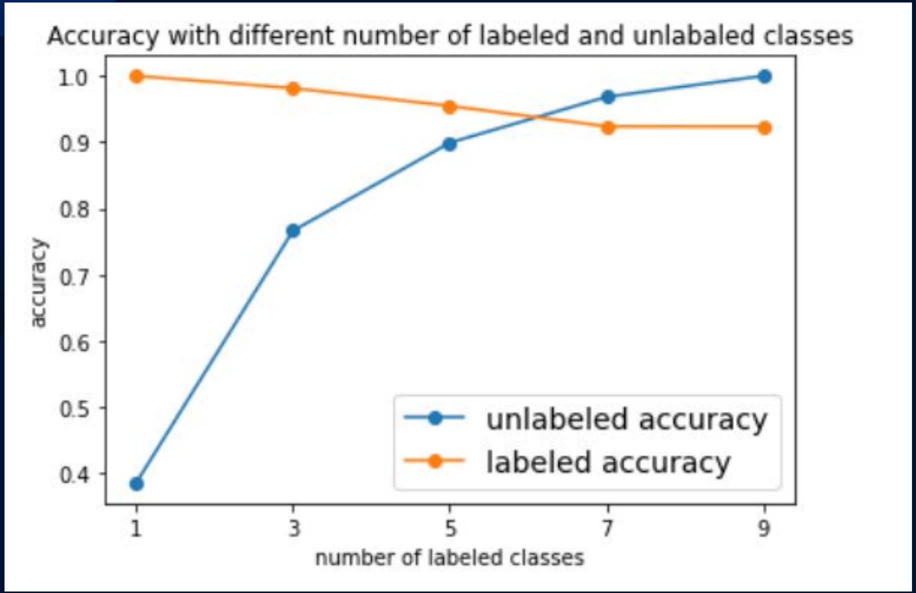
CIFAR10 has 10 classes → 5 different setting:

Labeled Classes	Unlabeled Classes
1	9
3	7
5	5
7	3
9	1

Unbalance on the number of classes



Unbalance on the number of classes



		Labeled Classes Accuracy	Unlabeled Classes Accuracy
Labeled and Unlabeled Classes	1_9	1	0.386
	3_7	0.982	0.766
	5_5	0.955	0.899
	7_3	0.924	0.968
	9_1	0.924	1

05

UNO Experiments

Experiment 1

Different number of Lab-Unlab classes

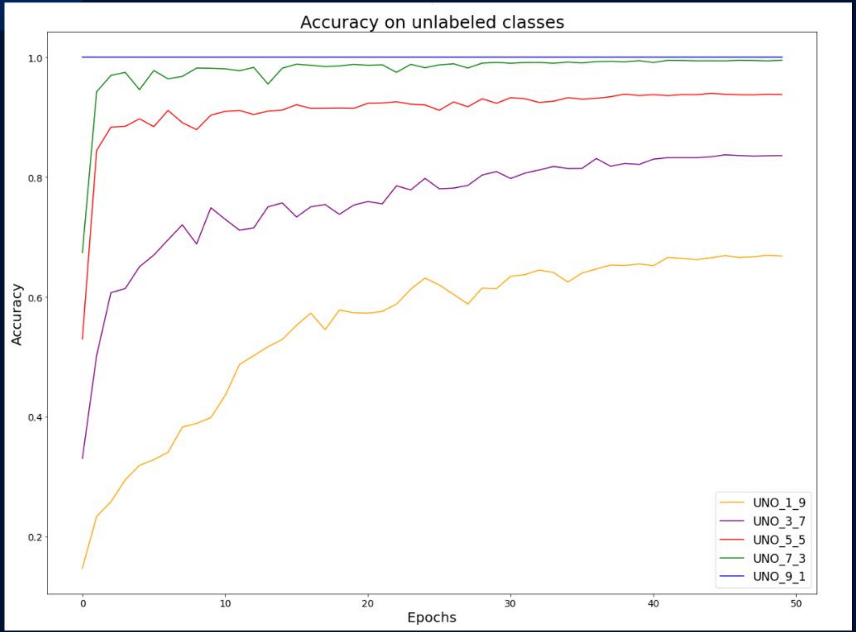
Unbalance on the number of classes

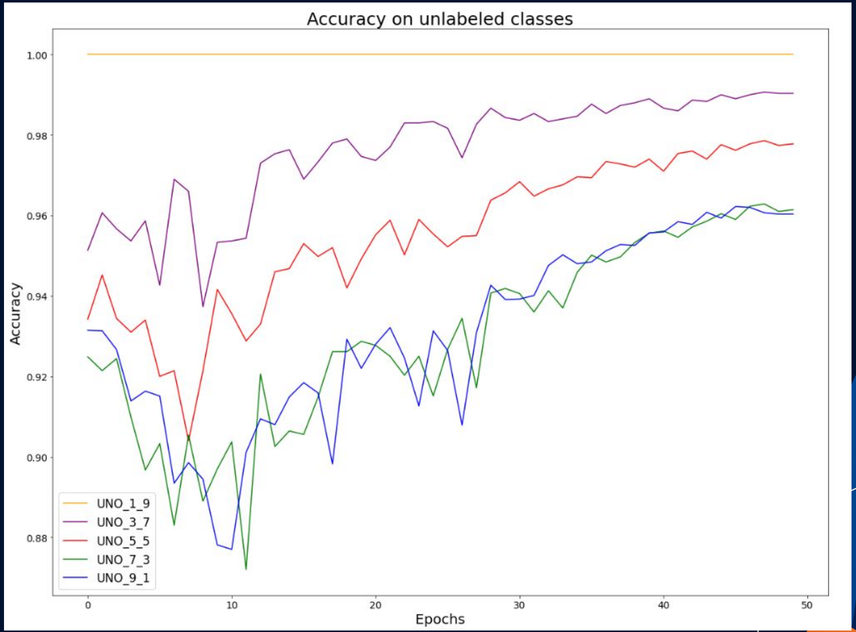
Same as autoNovel,
we tried different values for the number of labeled and unlabeled classes.

CIFAR10 has 10 classes → 5 different setting:

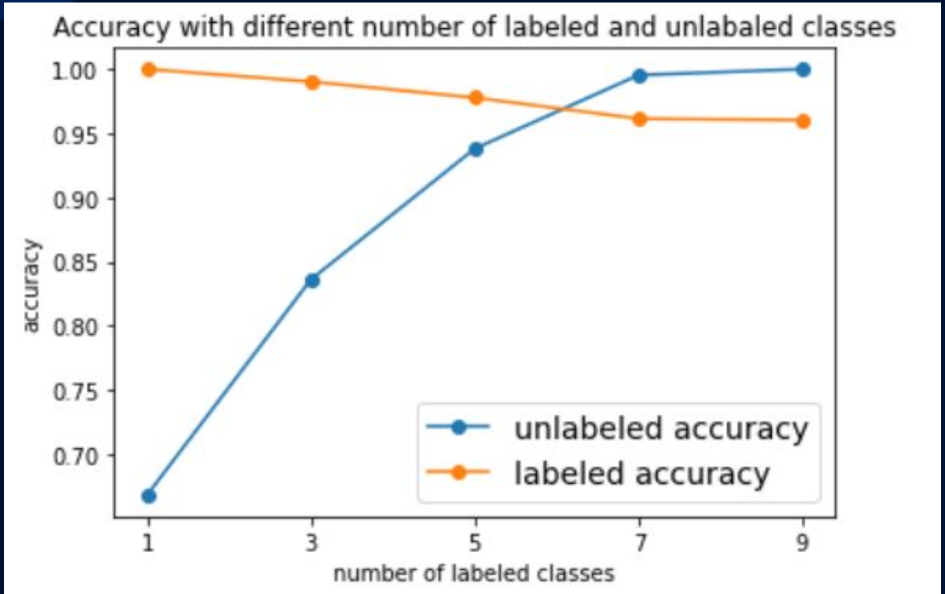
Labeled Classes	Unlabeled Classes
1	9
3	7
5	5
7	3
9	1

Unbalance on the number of classes





Unbalance on the number of classes



		Labeled Classes Accuracy	Unlabeled Classes Accuracy
Labeled and Unlabeled Classes	1_9	1	0.668
	3_7	0.990	0.836
	5_5	0.977	0.938
	7_3	0.961	0.995
	9_1	0.960	1

Unbalance on the number of classes: Comparison between autoNovel and UNO

	Labeled and Unlabeled Classes	Labeled Classes Accuracy	Unlabeled Classes Accuracy
1_9	autoNovel	1	0.386
	UNO	1	0.668
3_7	autoNovel	0.981	0.766
	UNO	0.990	0.836
5_5	autoNovel	0.955	0.899
	UNO	0.977	0.938
7_3	autoNovel	0.924	0.968
	UNO	0.961	0.995
9_1	autoNovel	0.924	1
	UNO	0.960	1

		Time per Epoch (avg)	
		Google Colab	Azure Virtual Machine
autoNovel	self-supervision	6 min	12 min
	supervised pretraining	14 sec	28 sec
	joint objective	1 min	2 min
UNO	supervised pretraining	1 min	2 min
	unified objective	4 min 30 sec	9 min

06

Outlook

Conclusion

- Different SSL techniques lead to similar results
- Domain shift is problematic
- Labeled samples are the most important to learn when working on datasets with unbalanced classes

NCD is a relevant and useful task

BUT, much more work is needed to increase performance and reduce training time

Thank you for the attention!