# Automatically Discovering NCD with adaptive prototype learning

Annotated Data $D^l = \{(x_i^l, y_i^l), i = 1 \dots N^l\}$

$\hookrightarrow \in C^u$

$C^l \cap C^u = \phi$ // Disjoint condition.

## Method

$\boxed{\text{stage 1}}$ prototypical representation learning.

loss: $L_{s1} = \underbrace{L_{ins}}_{\text{instance loss}} + \underbrace{L_{cat}}_{\text{categorical loss.}}$

### 1. Instance Discrimination:

DINO scheme

(ADAboost update of DINO)

$V = \{\overbrace{x_1^g, x_2^g}^{\text{Augmented global view (no rotation)}}, \underbrace{x_3, \dots}_{\text{rotated images}}\}$ // Augmented $x$ image.

$$L_{ins} = \frac{1}{N} \sum_{x \in \{x_1^g, x_2^g\}} \sum_{\substack{x' \in V \\ x' \neq x}} H(P_t(x), P_s(x'))$$

DINO head

## 2.1 Category Discrimination :

- unified classifier for both $c^l$ & $c^u$

(semi-sup setting)

### 2.1.1 online prototype learning

initial prototype:

$$P_c^{init} = w_c^{init} / \| w_c^{init} \|_2$$

[single layers weight uniform Dist]

total $\underline{\underline{c}}$ of them. $P_1 \ldots P_{c^u}$

pairwise seperation

Assigning Online pseudo labels

for the max c

pseudo label

$$y_i^u = \arg\max_c \cos\theta\left(P_c, f_\phi(x_i)\right)$$

[think of binary head for each class]

Each head gives a class score.

$\{c^l \cup c^u\}$

all classifier head.

$n-1$ $c-1$

$$L_{cls} = -\frac{1}{N} \sum_{i=0} \sum_{c=0} p_{c,i} \log(y_{c,i})$$

Binary    Probability.

## Updating Online Prototype:

$$\boxed{p_c \leftarrow \beta p_c + (1-\beta) f_\phi(x_i) \quad s.t. \quad x_i \in D^u}$$

Take prototype & use them to pseudo label

[Remove the Sinkhorn-knopp]

[Randomization prevents trivial Solution]

## Pairwise prototype loss:

(Another regularization)

$$L_{pas} = \frac{1}{K} \sum_{i=1}^{K} \max_{j \in C^u} M_{ij}$$

$; \; M = PP^T - 2I$

$\in \mathbb{R}^{K \times D}$

Regularize on Cross prototype Similarity

pairwise

Similarity

I guess should be 1

Joint optimization

$$\boxed{\mathcal{L}_{cat} = L_{cls} + \lambda L_{pas}}$$

**Stage 2**

prototypical Self-training

↳ improves more

claim: online pseudo-label are bad.

- Discard classifier of Stage 1
- Retrain a parametric classifier. (both base & novel classes)

**2.1 pseudo-label**

Spectral clustering the pseudo-label.

**2.2 Label rectification:**

Select example from $N^u$ with higher Confidence.

previous prototype.

$$\mathcal{L}_{rect} = - \sum^{N^u - 1} \cos \theta_{(p, f_i(x))} \sum^{C-1} p_{c,i} \log(y_{c,i})$$

$$i=0 \qquad \qquad c=0$$

$$\underbrace{\qquad\qquad\qquad}_{\text{weights}} \; ; \; \underbrace{x_i \in D^u}_{\text{CE loss.}}$$

## 2.3 Joint Optimization

$$\mathcal{L}_{CE} = -\sum_{i=0}^{N^l-1} \sum_{c=0}^{C-1} p_{c,i} \log(y_{c,i}) \; ; \; x_i \in D^l$$

$$\boxed{\mathcal{L}_{S2} = \frac{1}{N^u + N^l} \left( \mathcal{L}_{CE} + \mathcal{L}_{rect} \right)}$$