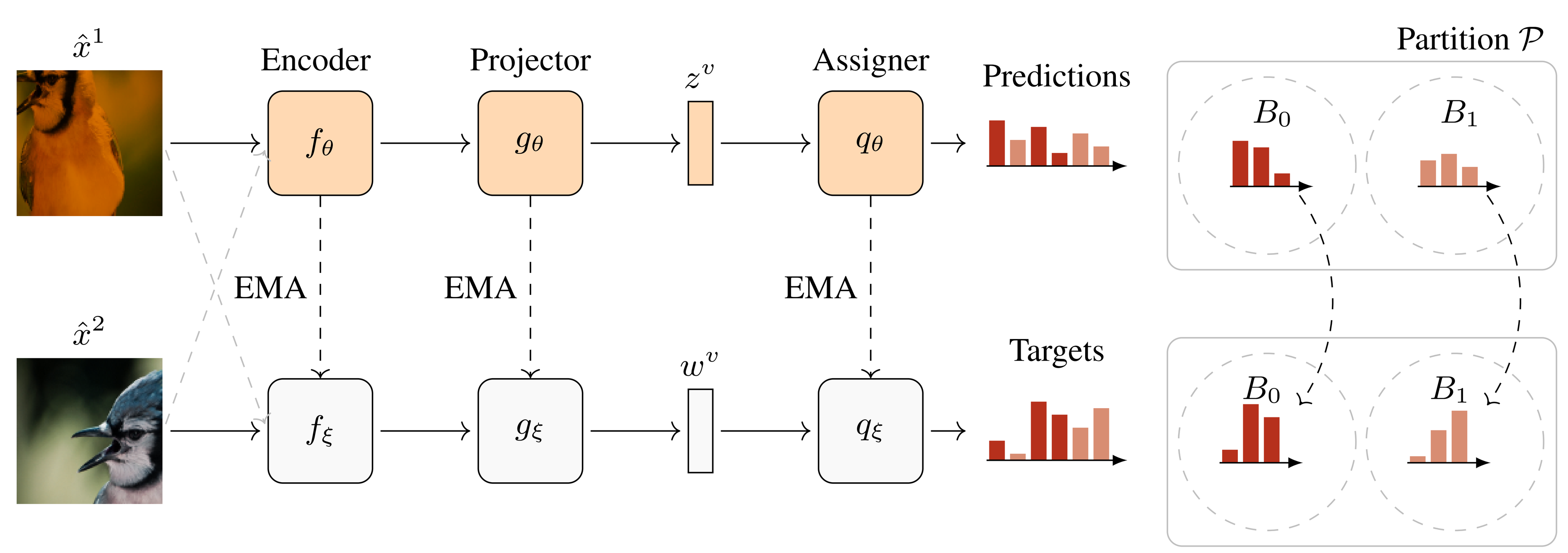
Representation Learning via Consistent Assignment of Views over Random Partitions

Thalles Silva ⊛ Adín Ramírez Rivera ❄
⊛ University of Campinas ❄ University of Oslo

Goal Devise multiple pseudo-classification tasks by arranging the prototypes into random partitions.

Methodology



Motivation

Solving the dimensionality problem of SSL clustering methods.

- Current methods use non-differentiable modules, e.g., Sinkhorn-Knopp, centering, and k-means, to stabilize training and solve the cluster assignment problem.

We propose:

- A method to solve the **dimensionality problem** that arises when we seek to learn a large number of prototypes.
- A **pretext task based on the divide and conquer** approach to tackle unsupervised distribution learning.

The random partition pretext task

Follow the standard SSL pipeline.

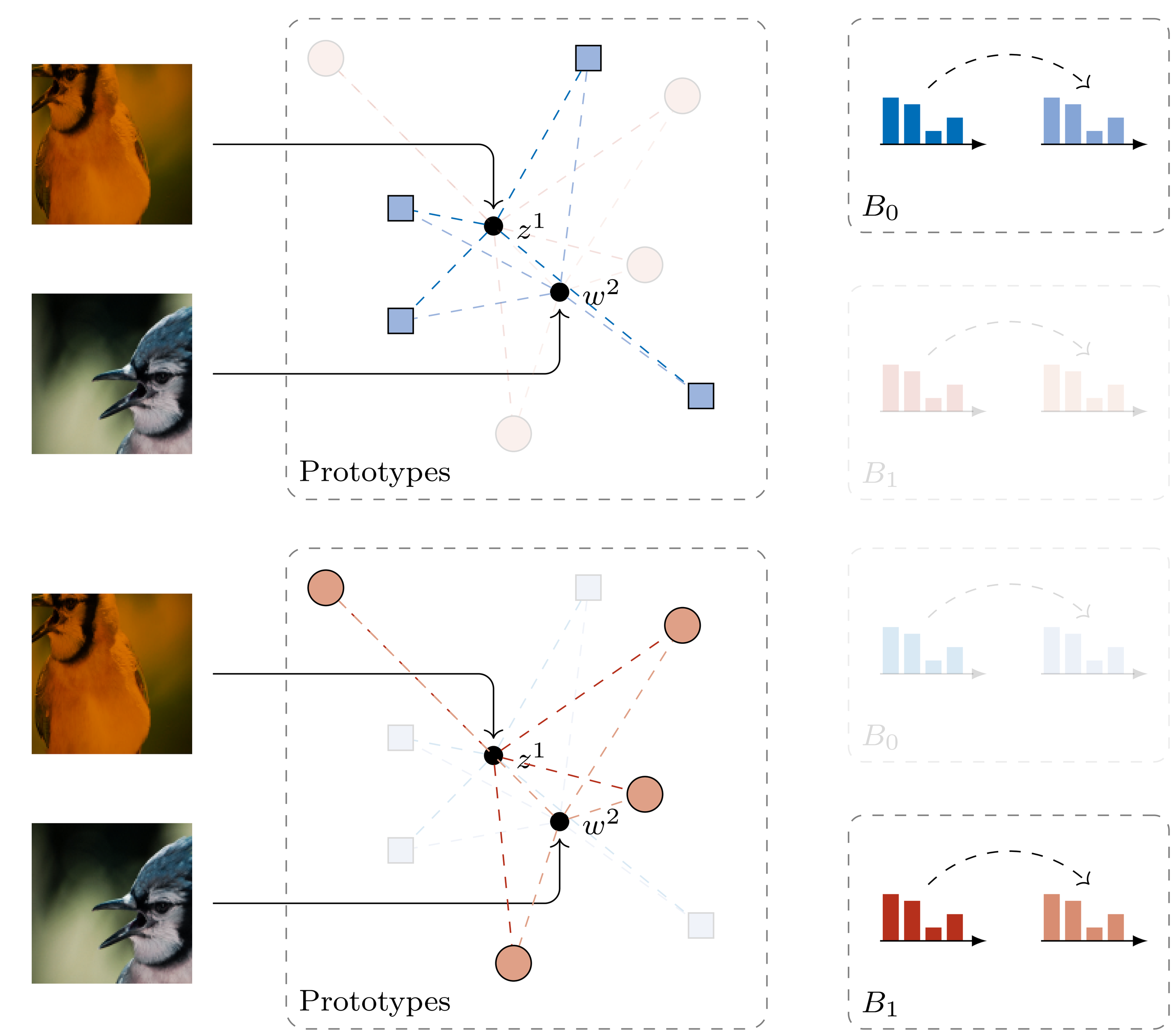
- Create views from an image using random augmentations.
- Define **two encoder streams in a teacher-student setup**, where each stream consumes a different view.
- Pass the features to student and teacher projector heads and **receive individual vector representations**.
- Pass the views' representations to the assigner function and obtain **probability distributions** relating each view to a set of learnable prototypes.

Should I optimize my loss function over all prototypes?

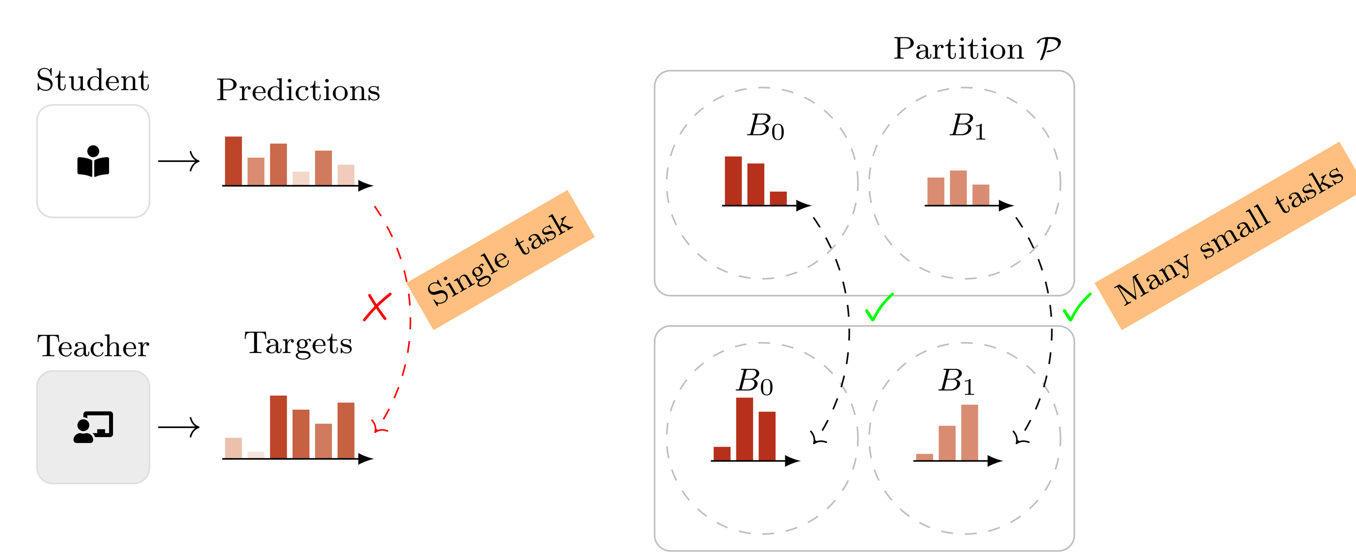
- For a large number of prototypes, it will be **a highly unstable task**.

Instead, solve the dimensionality problem!

- Assign views to **subsets of prototypes**.



The random partition strategy



Avoiding collapsed solutions

Keep in mind!

- Partitions need to be recreated **randomly** at each iteration.

How can I avoid training collapse?

- Just maximize the **entropy of the mean predictions** over a batch.

In math, you minimize this!

$$\mathcal{L} = \frac{1}{N N_{\mathcal{P}}} \sum_{i}^{N} \sum_{j}^{N_{\mathcal{P}}} \left(\mathcal{L}_c(s_{i,j}^1, t_{i,j}^2) + \mathcal{L}_c(s_{i,j}^2, t_{i,j}^1) \right) - H(\bar{p}_j), \qquad (1)$$

where,

$$\mathcal{L}_c(a, b) = -\log \langle a, b \rangle . \qquad (2)$$

Results

Transfer learning (k-NN)

Methods	Ep	Pets	Flowers	Aircraft	Cars	Country	Food	STL	GTSRB	Avg @k 10	20	100	200
		results for $k = 20$											
oBoW (mc)	200	57.3	61.9	18.1	11.5	12.0	47.4	96.6	50.6	44.3	44.4	43.5	43.0
DeepC-v2 (mc)	800	78.3	76.3	32.0	25.0	13.6	62.3	95.6	63.4	56.0	55.8	54.4	53.3
SwAV (mc)	800	77.0	75.2	29.0	22.7	13.8	59.1	95.2	63.2	54.5	54.4	53.1	52.2
DINO (mc)	800	80.9	**81.6**	35.3	30.1	**14.4**	62.0	95.6	62.9	57.9	57.8	56.9	56.0
Triplet	980	83.5	77.7	33.4	25.2	14.1	61.5	95.6	63.5	56.5	56.8	56.2	55.6
BarlowT	1000	82.9	78.8	32.7	26.3	13.3	61.4	94.8	65.6	56.8	57.0	56.4	55.7
MoCo-v3	1000	86.4	79.0	36.9	29.3	12.4	60.0	**96.7**	72.8	59.2	59.2	58.4	57.8
CARP	400	**86.8**	80.0	**42.1**	**33.5**	12.3	58.4	95.9	**75.3**	**60.4**	**60.5**	**59.7**	**59.2**
CARP (mc)	400	83.9	80.3	34.8	27.1	14.2	**62.9**	95.5	62.8	57.6	57.7	56.8	56.0

Clustering evaluation

	ImageNet-1M			CIFAR-10			CIFAR-100			GTSRB		
Method	NMI	AMI	ARI	NMI	AMI	ARI	NMI	AMI	ARI	NMI	AMI	ARI
PCL v2	69.7	47.5	22.2	46.7	46.6	34.8	49.1	42.2	17.1	44.1	44.1	13.0
SeLa-v2	68.7	45.5	21.3	42.0	41.9	30.6	49.7	42.9	18.2	45.7	43.2	12.0
DeepC-v2	69.7	47.1	22.4	47.0	46.9	35.5	53.2	46.7	21.8	48.1	45.7	13.5
SwAV	68.5	45.1	20.5	46.8	46.8	37.0	52.1	45.5	20.0	51.0	48.8	15.0
DINO	69.2	46.2	21.7	39.6	39.5	28.0	47.6	40.4	16.2	52.0	49.8	15.4
MIRA	68.9	45.7	21.2	39.5	39.4	28.8	49.0	42.1	17.6	51.6	49.4	15.8
CARP	**70.3**	**48.0**	**23.9**	**49.0**	**48.9**	**38.7**	**54.5**	**48.2**	**23.1**	**54.8**	**52.7**	**19.6**

Few-shot classification on Pascal VOC07 (mAP) and INat2018 (acc)

		Pascal VOC07						INat2018					
Method	Ep	n=1	n=2	n=4	n=8	n=16	full	n=1	n=2	n=4	n=8	n=16	full
PCL v2	200	**47.9**	59.6	66.2	74.5	78.3	85.4	1.4	1.6	2.3	2.9	4.8	2.1
DINO (mc)	800	45.6	58.4	66.6	74.8	79.6	88.2	6.5	12.0	20.4	29.6	35.9	30.4
Triplet	980	43.6	56.2	64.6	73.8	79.6	**88.3**	11.4	19.1	28.9	37.6	44.0	41.4
MoCo-v3	1000	46.6	59.6	67.0	75.4	**80.2**	87.4	8.1	12.2	18.5	27.2	33.5	28.0
CARP (mc)	200	46.0	58.3	66.5	75.5	79.5	88.0	8.6	14.4	23.6	32.7	38.2	33.9
	400	47.1	**59.8**	**67.3**	**75.8**	80.0	88.2	**11.5**	**19.6**	**29.6**	**39.1**	**45.1**	**42.6**

Image retrieval

		$\mathcal{R}$Ox		$\mathcal{R}$par	
Method	ep	M	H	M	H
Sup.	100	49.8	8.5	74.0	52.1
DINO	800	35.4	11.1	55.9	27.5
Triplet	980	35.3	12.0	58.2	28.7
MoCo-v3	1000	33.1	10.9	**59.1**	**31.3**
CARP	200	**38.8**	**15.5**	58.8	30.4

Copy detection

Method	Ep	mAP
Rnd	–	25.7
DINO	800	78.8
Triplet	980	81.7
VICReg	1000	83.7
MoCo-v3	1000	80.6
CARP (mc)	400	**84.0**

Thirty-seventh Conference on Neural Information Processing Systems
Acknowledgements: Sigma2 NN8104K, RCN 309439, & CAPES 001.