REPRESENTATION LEARNING ALGORITHMS

Mansi Tomer, Nirbhay Sharma

ABSTRACT

Learning rich feature representations is crucial for vision tasks. We investigate various self-supervised representation learning algorithms. Additionally, we propose Distribution Alignment Regularizer (DARe), a novel self-supervised contrastive learning (CL) framework based on VAE. DARe outperforms other baseline works in unsupervised settings. The code is available at *Code*

1. PROBLEM DEFINITION AND MOTIVATION

Representation learning is a popular ML paradigm which is motivated by the importance of rich representations which are crucial to achieve better performance on tasks such as image classification.

2. PRIOR WORK

The prior work includes SimCLR [1], SupCon [2], BYOL [3], SimSiam [4], Barlow Twins [5], and Triplet Loss [6], The prior work primarily focus on learning rich representations via varieties of loss functions including contrastive loss, knowledge distillation, Correlation loss etc.

3. PROGRESS

We successfully implement SOTA algorithms, SimCLR, SupCon, Triplet Margin Loss, Barlow Twins, SimSiam, and BYOL from scratch in Pytorch. We additionally, propose a novel VAE inspired loss (Fig. 1), which explicitly aligns the embedding distributions, creating more robust features.

4. EXPERIMENTS AND RESULTS

The Experiments are done on CIFAR100/10 datasets with Resnet18/50 encoder and presented in Table 1

5. SUMMARY OF CONTRIBUTIONS AND NOVELTY

The encoder f get embedding, g predicts $\mu, \log(\sigma^2)$, the projections z_i are computed using reparameterization. DARe is a combination of CL + JSD between the distributions.

$$\mathcal{L}_{DARe} = \mathcal{L}_{con}(z_1, z_2) + \lambda \mathcal{L}_{JSD}(\mathcal{N}(\mu_1, \sigma_1^2 I), \mathcal{N}(\mu_2, \sigma_2^2 I))$$

Algorithm	CIFAR10 (R50)	CIFAR100 (R50)	CIFAR10 (R18)	CIFAR100 (R18)
SimCLR	87.5	57.7	85.9	55.0
SupCon	**94.0**	74.7	**93.5**	**70.4**
Triplet	83.4	**76.3**	86.0	64.5
Barlow Twins	81.2	47.7	80.3	45.8
BYOL	83.0	47.0	84.8	54.8
SimSiam	76.5	34.5	88.6	62.3
DARe (Ours)	**89.4**	-	**87.3**	**61.6**

Table 1. Comparison of algorithms on CIFAR10/100 datasets using ResNet-50/18 (R50/18) encoder

$$\mathcal{L}_{con}(z_1, z_2) = -\sum_{i=1}^{n} \log\left(\frac{e^{sim(z_{1i}, z_{2i})/\tau}}{\sum_{j \neq i} e^{sim(z_{1i}, z_{2j})/\tau}}\right)$$

$$\mathcal{L}_{KL}(\mathcal{N}(\mu_1, \sigma_1^2 I), \mathcal{N}(\mu_2, \sigma_2^2 I)) = \frac{1}{2}\left(\sum_i \left(\frac{\sigma_{1i}^2}{\sigma_{2i}^2} + \log\left(\frac{\sigma_{2i}^2}{\sigma_{1i}^2}\right) + \frac{(\mu_{2i} - \mu_{1i})^2}{\sigma_{2i}^2}\right)\right) \quad (1)$$

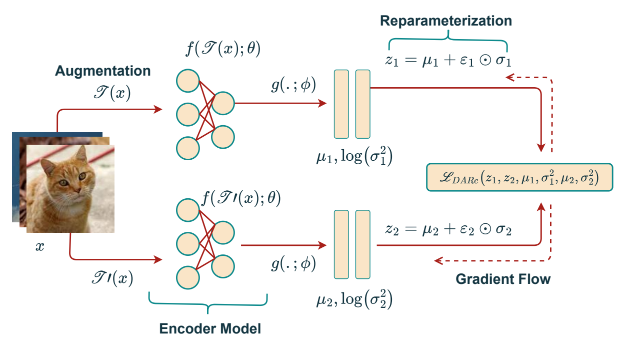

Fig. 1. DARe Framework. We predict $\mu, \log(\sigma^2)$ for augmented views of the image. We then use reparameterization and apply DARe Loss (CL + JSD loss combined).

6. RESPONSIBILITY OF INDIVIDUAL MEMBERS

- *Mansi Tomer (24014)*: Worked on SimSiam, Barlow Twins, Triplet
- *Nirbhay Sharma (24806)*: Worked on SimCLR, SupCon, BYOL, Novelty (DARe)

7. REFERENCES

[1] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton, "A simple framework for contrastive learning of visual representations," in *International conference on machine learning*. PmLR, 2020, pp. 1597–1607.

[2] Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan, "Supervised contrastive learning," *Advances in neural information processing systems*, vol. 33, pp. 18661–18673, 2020.

[3] Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pires, Zhaohan Guo, Mohammad Gheshlaghi Azar, et al., "Bootstrap your own latent-a new approach to self-supervised learning," *Advances in neural information processing systems*, vol. 33, pp. 21271–21284, 2020.

[4] Xinlei Chen and Kaiming He, "Exploring simple siamese representation learning," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 15750–15758.

[5] Jure Zbontar, Li Jing, Ishan Misra, Yann LeCun, and Stéphane Deny, "Barlow twins: Self-supervised learning via redundancy reduction," in *International conference on machine learning*. PMLR, 2021, pp. 12310–12320.

[6] Florian Schroff, Dmitry Kalenichenko, and James Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 815–823.