Supervision Accelerates Pre-training in Contrastive Semi-Supervised Learning of Visual Representations

School of industrial and Management Engineering, Korea University

Young Jae Lee

KOREA UNIVERSITY

DMQA

Contents

Research Purpose

❖ Supervision Accelerates Pre-training in Contrastive Semi-Supervised Learning of Visual Representations (NeurIPS, 2022)

- McGill/Montreal 대학, Facebook AI Research에서 연구하였고 2022년 10월 28일 기준 약 9회 인용

Supervision Accelerates Pre-training in Contrastive Semi-Supervised Learning of Visual Representations

Mahmoud Assran[1,2,3], **Nicolas Ballas**[3], **Lluis Castrejon**[1,3,4], **and Michael Rabbat**[1,3]
[1]Mila – Quebec AI Institute
[2]Department of Electrical and Computer Engineering, McGill University
[3]Facebook AI Research
[4]Department of Computer Science and Operations Research, University of Montreal
`{massran, ballasn, lcastrejon, mikerabbat}@fb.com`

Abstract

We investigate a strategy for improving the efficiency of contrastive learning of visual representations by leveraging a small amount of supervised information during pre-training. We propose a semi-supervised loss, SuNCEt, based on noise-contrastive estimation and neighbourhood component analysis, that aims to distinguish examples of different classes in addition to the self-supervised instance-wise pretext tasks. On ImageNet, we find that SuNCEt can be used to match the semi-supervised learning accuracy of previous contrastive approaches while using less than half the amount of pre-training and compute. Our main insight is that leveraging even a small amount of labeled data during pre-training, and not only during fine-tuning, provides an important signal that can significantly accelerate contrastive learning of visual representations. Our code is available online at github.com/facebookresearch/suncet.

Research Purpose

❖ Supervision Accelerates Pre-training in Contrastive Semi-Supervised Learning of Visual Representations (NeurIPS, 2022)

- Visual Representations에 대한 대조 학습 효율성을 향상시키기 위해 사전 학습 동안 소량의 데이터를 활용할 수 있는 Semi-Supervised Loss인 Supervised Noise Contrastive Estimation (SuNCEt) 제안
- SuNCEt는 Noise-Contrastive Estimation과 Neighborhood Component Analysis 기반으로 함

Supervised Noise Contrastive Estimation (SuNCEt)

❖ Supervised Noise Contrastive Estimation (SuNCEt)

- Background
  - ✓ Minimize the Normalized Temperature-Scaled Cross Entropy Loss for Instance-Wise Discrimination

$$\ell_{\text{inst}}(z_{i,1}) = -\log \frac{\exp(\text{sim}(z_{i,1}, z_{i,2})/\tau)}{\sum_{z \in \mathcal{Z} \setminus \{z_{i,1}\}} \exp(\text{sim}(z_{i,1}, z)/\tau)}, \quad (1)$$

where $\text{sim}(a, b) = \frac{a^T b}{\|a\| \|b\|}$ denotes the cosine similarity and $\tau > 0$ is a temperature parameter.

DMQA

Supervised Noise Contrastive Estimation (SuNCEt)

❖ Supervised Noise Contrastive Estimation (SuNCEt)

- SuNCEt
  - ✓ 대조 학습의 Computational Efficiency를 향상시키기 위한 제안 손실 함수
  - ✓ S: Labeled Samples Set
  - ✓ $Z_S(\theta) = \left(f_\theta(\hat{x})\right)_{\hat{x} \in S}$: Associated Set of Parameterized Embeddings
  - ✓ $\hat{x} \in S$: Anchor Image
  - ✓ $z = f_\theta(\hat{x})$: Representation Vector / y: Class Label

$$\ell(z) = -\log \frac{\sum_{z_j \in \mathcal{Z}_y(\theta)} \exp(\mathrm{sim}(z, z_j)/\tau)}{\sum_{z_k \in \mathcal{Z}_S(\theta) \setminus \{z\}} \exp(\mathrm{sim}(z, z_k)/\tau)}, \quad (2)$$

which is then averaged over all anchors $\frac{1}{|S|} \sum_{z \in \mathcal{Z}_S(\theta)} \ell(z)$.

Supervised Noise Contrastive Estimation (SuNCEt)

❖ Supervised Noise Contrastive Estimation (SuNCEt)

- SuNCEt
  - ✓ Motivation: Relationship Between Contrastive Representation Learning and Distance-Metric Learning
  - ✓ Neighborhood Component Analysis Form
  - ✓ $d(z, z_j)$: Temperature-Scaled Cosine Similarity Metric with Query Embedding z와 다른 labeled images z_j
  - ✓ $z_j \in Z_S(\theta)\{z\}$: Selects Point as its neighbor
  - ✓ Predicts Class Label $\hat{y} = c$, Given a Query Embedding z

$$d(z, z_j) = z^T z_j / (\|z\| \|z_j\| \tau)$$

$$p(z_j | z) = \frac{\exp(d(z, z_j))}{\sum_{z_k \in Z_{S(\theta)} \setminus \{z\}} \exp(d(z, z_k))}.$$

$$p(\hat{y} = c | z) = \sum_{z_j \in Z_c(\theta)} p(z_j | z) = \frac{\sum_{z_j \in Z_c(\theta)} \exp(d(z, z_j))}{\sum_{z_k \in Z_S(\theta) \setminus \{z\}} \exp(d(z, z_k))}, \quad (3)$$

Experiments

❖ Experiments – Sample & Computational Efficiency

- Top-5 Validation Accuary 기준으로 SuNCEt를 추가로 적용했을 때 빠르게 수렴

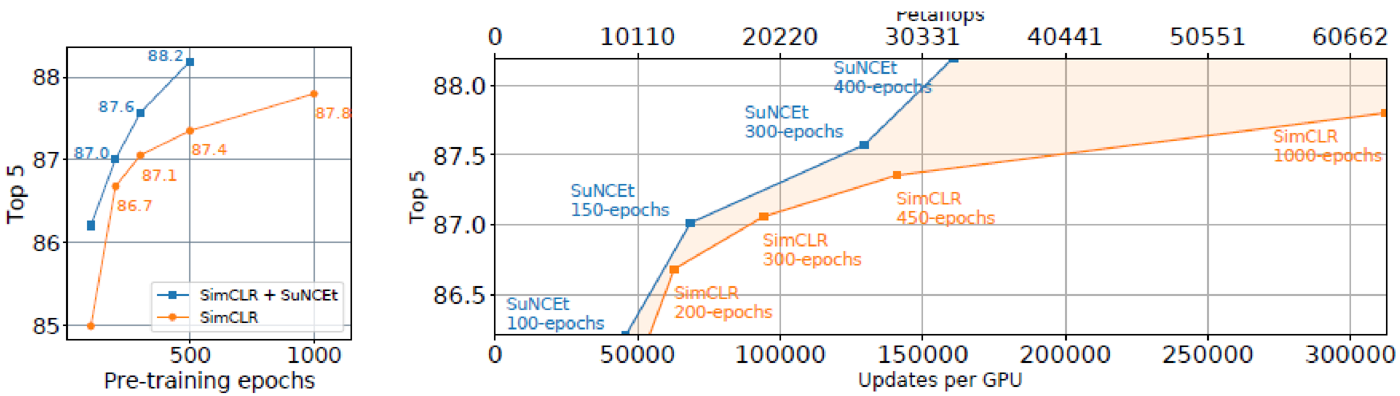


Figure 1: Top-5 validation accuracy of a ResNet50 pre-trained on ImageNet with access to 10% of the labels. Orange markers depict SimCLR self-supervised pre-training followed by fine-tuning. Blue markers depict the combination of SimCLR + SuNCEt. Using SuNCEt to leverage available labels during pre-training (not only fine-tuning), **(i)** accelerates convergence and produces better models (left sub-figure); and **(ii)** can match the semi-supervised learning accuracy of SimCLR whith much less pre-training (right sub-figure). Orange shading in the right sub-figure depicts compute saved. We train all methods using 64 V100 GPUs. One SimCLR epoch corresponds to 312 updates per GPU.

DMQA

Experiments

❖ Experiments – Sample & Computational Efficiency

- ImageNet에서 Top-1 Accuracy 기준 비교 결과
- SuNCEt를 추가로 적용했을 때 더 빠르게 수렴

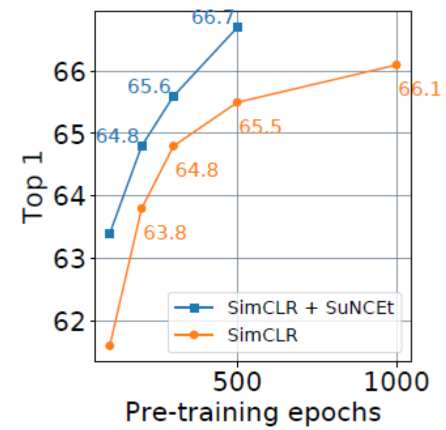
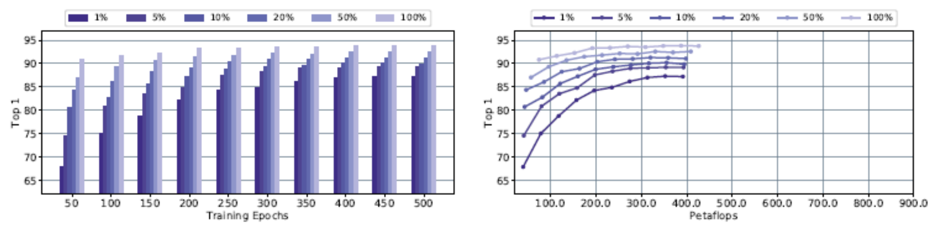


Figure 2: Top-1 accuracy of a ResNet50 pre-trained on ImageNet with access to 10% of the labels. Orange markers depict SimCLR self-supervised pre-training followed by fine-tuning. Blue markers depict the combination of SimCLR + SuNCEt. Using SuNCEt to leverage available labels during pre-training (not only fine-tuning) accelerates convergence and produces better models.

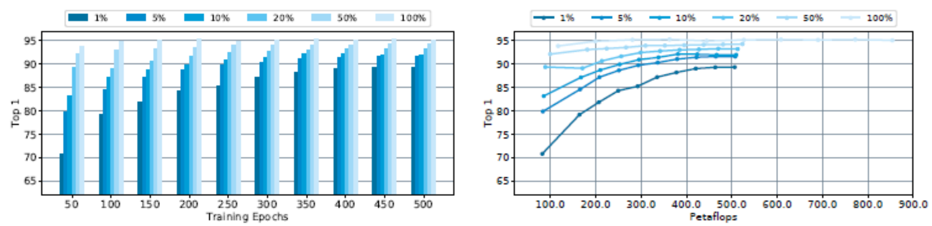
Experiments

❖ Experiments – Sample & Computational Efficiency

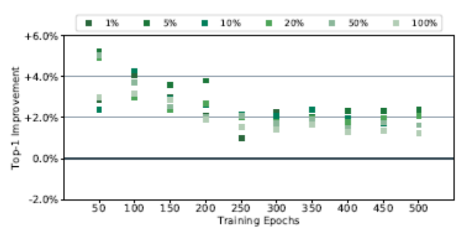
- CIFAR 10에서 SimCLR / SimCLR + SuNCEt 비교 결과



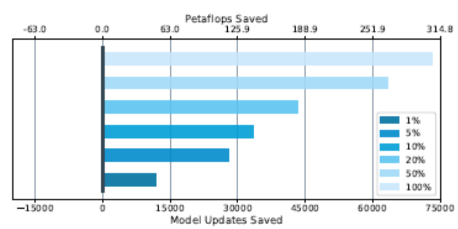
(a) SimCLR test-set convergence with fine-tuning on various percentages of labeled data.



(b) SimCLR + SuNCEt test-set convergence with fine-tuning on various percentages of labeled data.



(c) SuNCEt improvement in test-set convergence with fine-tuning on various percentages of labeled data.



(d) Computation saved by SuNCEt in reaching the best SimCLR test accuracy with fine-tuning on various percentages of labeled data.

DMQA

Experiments

❖ Experiments – Sample & Computational Efficiency

- SimCLR 외 SwAV에 SuNCEt 추가 적용했을 때 비교 결과

Table 3: Validation accuracy of a ResNet50 pre-trained on ImageNet with access to 10% of labels. Contrastive methods like SimCLR [7] and SwAV [5] can leverage SuNCEt during pre-training to surpass their baseline semi-supervised accuracy in half the number of pre-training epochs. SuNCEt+SwAV is also competitive with other semi-supervised approaches and outperforms FixMatch+RandAugment in terms of top-5 accuracy.

Method	Epochs	Top-1	Top-5
Supervised [55]	200	56.4	80.4
NPID++ [52, 35]	800	–	81.5
PIRL [35]	800	–	83.8
UDA + RandAugment [53]	–	68.8	88.5
FixMatch + RandAugment [45]	300	**71.5**	89.1
SimCLRv2 [8]	1200	68.4	89.2
SimCLR [7]	1000	65.6	87.8
SimCLR+SuNCEt (**ours**)	500	66.7	88.2
SwAV [5]	800	70.2	**89.9**
SwAV+SuNCEt (**ours**)	400	70.8	**89.9**

DMQA

Conclusion

- 대조 학습의 샘플 및 계산 효율성을 향상시키기 위해 Supervised Noise Contrastive Estimation (SuNCEt) 제안
- 사전 학습 동안 소량의 데이터를 활용할 수 있는 Semi-Supervised Loss Function
- SuNCEt는 Noise-Contrastive Estimation과 Neighborhood Component Analysis 기반으로 함

- Reference

  - Assran, M., Ballas, N., Castrejon, L., & Rabbat, M. (2020). Supervision Accelerates Pre-training in Contrastive Semi-Supervised Learning of Visual Representations. arXiv preprint arXiv:2006.10803.


Thank you