Semi-Supervised Learning of Visual Features by Non-Parametrically Predicting View Assignments with Support Samples

School of Industrial and Management Engineering, Korea University

Jaehoon Kim

Contents

Research Purpose

Predicting View **A**ssignments **w**ith **S**upport Samples (PAWS)

- Semi-Supervised Learning of Visual Features by Non-Parametrically Predicting View Assignments with Support Samples (ICCV, 2021)
  - 2022년 11월 4일 기준 50회 인용됨

Semi-Supervised Learning of Visual Features by Non-Parametrically Predicting View Assignments with Support Samples

Mahmoud Assran, Mathilde Caron, Ishan Misra, Piotr Bojanowski, Armand Joulin,
Nicolas Ballas*, and Michael Rabbat*
Facebook AI Research
{massran, mathilde, imisra, bojanowski, ajoulin, ballasn, mikerabbat}@fb.com

Abstract

This paper proposes a novel method of learning by predicting view assignments with support samples (PAWS). The method trains a model to minimize a consistency loss, which ensures that different views of the same unlabeled instance are assigned similar pseudo-labels. The pseudo-labels are generated non-parametrically, by comparing the representations of the image views to those of a set of randomly sampled labeled images. The distance between the view representations and labeled representations is used to provide a weighting over class labels, which we interpret as a soft pseudo-label. By non-parametrically incorporating labeled samples in this way, PAWS extends the distance-metric loss used in self-supervised methods such as BYOL and SwAV to the semi-supervised setting. Despite the simplicity of the approach, PAWS outperforms other semi-supervised methods across architectures, setting a new state-of-the-art for a ResNet-50 on ImageNet trained with either 10% or 1% of the labels, reaching 75.5% and 66.5% top-1 respectively. PAWS requires 4× to 12× less training

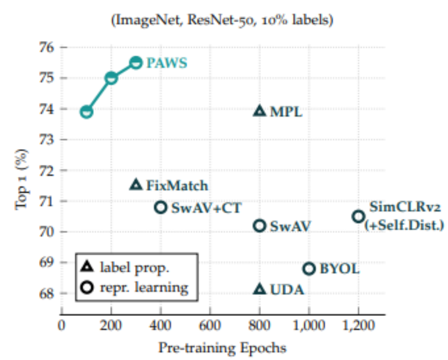


Figure 1: Training a ResNet-50 on ImageNet when only 10% of the training set is labeled. The figure shows top-1 validation accuracy as a function of the number of training epochs. The proposed method, PAWS, achieves higher accuracy than previous work while requiring significantly fewer training epochs. Concretely, 100 epochs of PAWS training takes less than 8.5 hours using 64 NVIDIA V100-16G GPUs.

Research Purpose

Predicting View **A**ssignments **w**ith **S**upport Samples (PAWS)

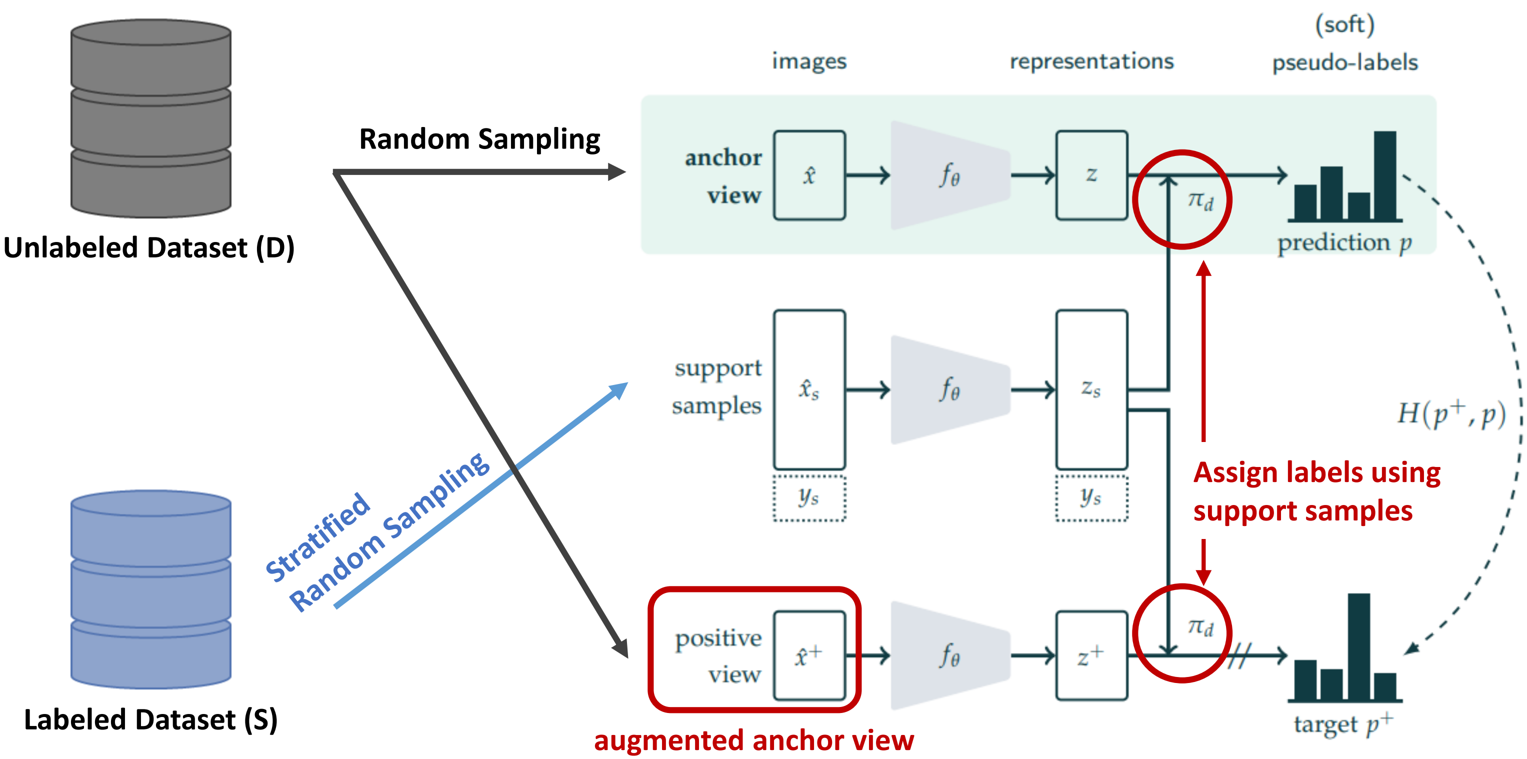
❖ Semi-Supervised Learning of Visual Features by Non-Parametrically Predicting View Assignments with Support Samples (ICCV, 2021)

- Self-supervised learning 방법론을 semi-supervised learning에 활용함
  - ✓ BYOL/SwAV에서 사용한 distance-metric loss를 semi-supervised learning에 적용함 ← Self-supervised learning
  - ✓ 레이블 데이터를 활용하여 pseudo-labeling을 진행함 ← Semi-supervised learning
- Self-supervised learning이 상당한 연산량을 요구하는 단점을 semi-supervised 방식으로 해소했다고 볼 수 있음 (개인 의견)

Proposed Method

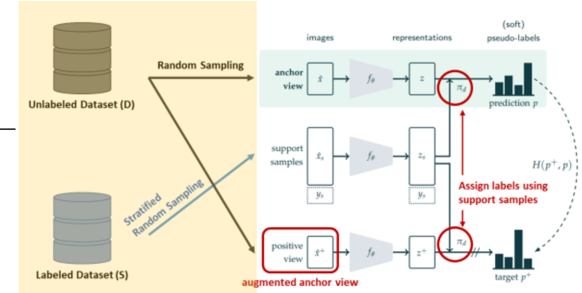
❖ Overview of Learning Framework

- 레이블 유/무에 따른 데이터셋이 나뉘며 각각 미니 배치를 형성하는 샘플링 방법론이 다름
- Anchor와 positive view가 동일한 pseudo-label을 갖도록 학습을 진행



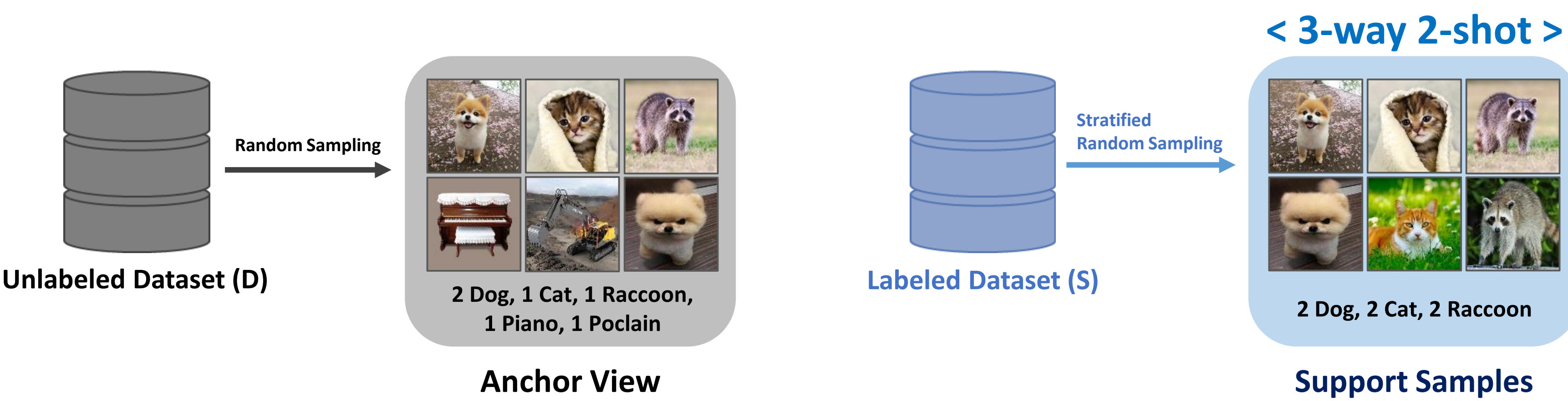
Assran, Mahmoud, et al. "Semi-supervised learning of visual features by non-parametrically predicting view assignments with support samples." *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021.

Proposed Method

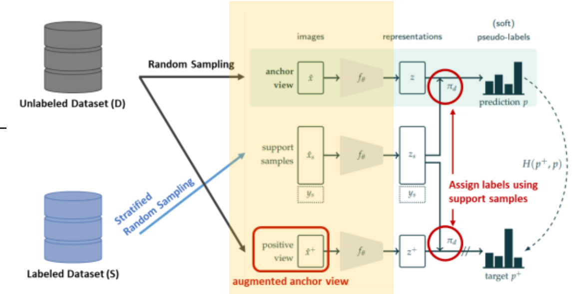


❖ Step1. Sampling Data

- Unlabeled dataset에서는 random sampling을 진행함
- Labeled dataset에서는 stratified random sampling을 진행함
- Stratified random sampling은 선택한 임의의 클래스마다 동일한 수의 데이터를 추출
- 이는 few-shot learning에서 미니 배치를 구성할 때 사용하는 k-way n-shot 방식과 동일함
  (ex. 3-way 2-shot : 세 가지 클래스에서 데이터 두 개씩 추출)



Proposed Method



❖ Step 2. Augmentation 적용 및 Positive view 생성

- Augmentation으로 positive view를 생성함
- Support samples에도 동일한 augmentation이 적용됨
- 실제 실험에서 augmentation 종류는 SimCLR를 참고하였으며,

  SwAV 방식을 참고하여 두 개의 large crop과 여섯 개의 small crops를 생성함

  ✓ SimCLR에서 사용한 augmentation: random crop, horizontal flip, color distortion, Gaussian blur

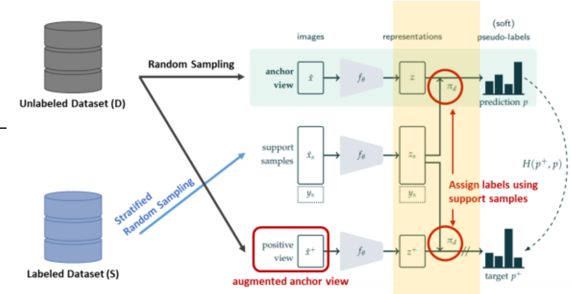
Anchor View

Augmentation

Crop, Rotation, Jitter

Positive View

Proposed Method



❖ Step 3. Assign labels using support samples (pseudo-labeling)

- Similarity Classifier로 클래스 별 support samples와의 거리를 나타낸 soft pseudo-label을 생성
- 거리(d)는 exponential temperature-scaled cosine으로 계산함 → $\exp(a^T b / \|a\| \|b\| \tau)$

$$\pi_d(z_i, \mathbf{z}_S) = \sum_{(z_{s_j}, y_j) \in \mathbf{z}_S} \left(\frac{d(z_i, z_{s_j})}{\sum_{z_{s_k} \in \mathbf{z}_S} d(z_i, z_{s_k})} \right) y_j$$

View (Anchor/Positive)

Support samples

j^{th} support sample

Label of j^{th} support sample (one-hot vector)

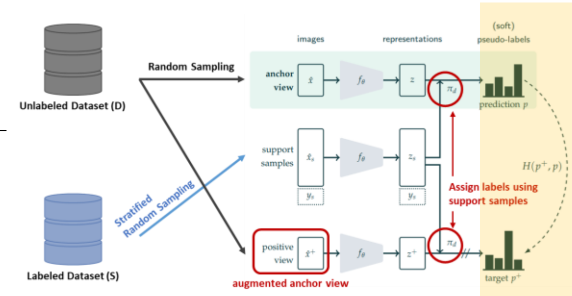
Softmax of distance between view and support sample

0.2	강아지
0.3	고양이
0.5	너구리

Support sample이 3-way 2-shot인 경우
View와 support samples의 너구리 데이터 두 개와의 거리에 대한 합이 됨

< Similarity Classifier (π_d) >

Proposed Method



❖ Step 4. Sharpening the labels and compute the loss

- Cross entropy loss를 두 번 사용하여 anchor와 positive view 상호간의 pseudo-label을 맞추도록 학습
- 이 때 trivial solution이 발생하는 것을 방지하기 위하여 한 쪽에 sharpening을 적용함
- Mean entropy maximization은 예측이 하나의 클래스로 집중되는 것을 억제하는 역할을 함
  - ✓ 실험에서는 배치 크기를 매우 크게 하기 때문에 random sampling 시 클래스 분포가 uniform 할 것이라고 가정하는 듯 (개인 의견)

Positive view   Anchor view

$$\frac{1}{2n}\sum_{i=1}^{n}\left(H(\rho(p_i^+),p_i)+H(\rho(p_i),p_i^+)\right)-H(\bar{p})$$

Cross Entropy Loss (*H*)   Sharpening Function (*ρ*)   Mean Entropy Maximization

$$\rho\left(\begin{array}{|c|c|}\hline 0.2 & \text{강아지} \\ \hline 0.3 & \text{고양이} \\ \hline 0.5 & \text{너구리} \\ \hline\end{array}\right) \rightarrow \begin{array}{|c|c|}\hline 0.1 & \text{강아지} \\ \hline 0.15 & \text{고양이} \\ \hline 0.75 & \text{너구리} \\ \hline\end{array}$$

Regularization Term

< Objective Function >

Experiment

vs. Baselines (semi-supervised / self-supervised learning)

❖ Result (fine-tuning w/ 1%, 10% labels)

- ResNet-50을 사용하여 ImageNet 데이터를 학습 및 성능 검증
- PAWS가 보다 적은 epoch을 학습하더라도 더 높은 성능을 보임
- PAWS-NN은 별도의 fine-tuning 없이 사전 학습된 표현 벡터로 최근접 이웃 분류를 수행한 것
- PAWS-NN 역시 baseline 보다 뛰어난 성능을 보여주며 이를 통해 사전 학습이 효과적으로 이루어졌음을 알 수 있음

ResNet-50

Method	Epochs	Top 1 1%	Top 1 10%
Methods using label propagation:			
UDA [15]	800	–	68.1
FixMatch [11]	300	–	71.5
MPL [6]	*800	–	73.9
Methods using only representation learning:			
BYOL [4]	1000	53.2	68.8
SwAV [3]	800	53.9	70.2
SwAV+CT [40]	400	–	70.8
SimCLRv2 [1]	800	57.9	68.4
SimCLRv2 (+Self.Dist.) [1]	1200	60.0	70.5
PAWS	**100**	**63.8**	**73.9**
PAWS	**200**	**66.1**	**75.0**
PAWS	**300**	**66.5**	**75.5**
Non-parametric classification (no fine-tuning):			
PAWS-NN	100	61.5	71.0
PAWS-NN	200	63.2	71.9
PAWS-NN	300	64.2	73.1

* **Epochs**는 사전학습에서의 epoch를 의미함

DMQA hcai

Experiment

Ablation study

❖ Result (support samples에서 k-way n-shot 값을 조정하여 실험)

- Stratified random sampling 시 선택하는 클래스의 종류가 많아질수록 성능이 상승하는 것 확인
- 이는 view에서 pseudo-label에서 해당되는 클래스가 없는 경우를 제거하기 때문으로 추정
- 1000-way 16-shot 실험은 labeled 데이터가 총 12,811개 존재하여 진행 불가

Classes	Imgs. per Class	Top 1 1%	Top 1 10%
1000	16	–	74.5
1000	12	63.9	74.2
960	7	63.8	73.9
960	4	63.7	72.0
448	8	61.8	70.1

Table 3: **Support Set.** Ablating the composition of the sampled support mini-batches when training a ResNet-50 on ImageNet for 100 epochs. Our default setup is shaded in green. Increasing the size of the support set improves performance. However, when sampling a fixed number of instances, it is preferable to sample many classes with a few images per class, rather than few classes with many images per class.

Conclusion

❖ PAWS는 self-supervised learning과 semi-supervised learning의 교집합으로서 제안된 방법론

- 예측 값을 예측 하는 것(Bootstrapping, BYOL), 사전 학습 후 조정 학습을 진행하는 것 (two stage learning)과 같이 전체적인 학습 프레임워크는 self-supervised learning 방법론을 따름
- 다만, 각 view의 예측 값을 구할 때 support sample(labeled data)을 활용하여 pseudo-labeling을 진행하는 것은 semi-supervised 방법론을 따름
- 두 학습 방법론을 결합하여 기존의 방법론 보다 **매우 효율적인 학습 횟수**와 **효과적인 학습 성능**을 달성할 수 있었음

Thank You