Clustering Latent Representations for Semi-Supervised Learning

Serkan Karakulak
Aakriti Gupta
David Martuscello

May 13, 2019

Observation on Semi-Supervised Learning Setting

- Both labeled and unlabeled data are generated by the same latent distribution
- Neural Networks map inputs to higher level latent representations
- If there is no overfitting, we can expect to have consistent latent representations between the labeled and unlabeled samples

Idea

Having consistent clusters of latent representations between the labeled and unlabeled data at deeper layers of the network would regularize it without loss of high level information.

Proposed Solution

- We propose using clustering in the latent vector space and regularize the network by deviation from cluster centroids
- Cluster centroids and network parameters can simultaneously be learned during the training
- Cost function is differentiable

$$\ell_c(x) = \sum_i \min_{C_{ij}} (f_\theta^i(x) - C_{ij})^2$$

$$\mathcal{L} = \min_\theta \frac{1}{N} \sum_{k=1}^{N} \Big(\ell(f_\theta(x_k^L), y_k) + \beta \ell_c(x_k^L) \Big) + \frac{1}{M} \sum_{k=1}^{M} \beta \ell_c(x_k^U)$$

where x_k^L and x_k^U are the labeled and unlabeled samples respectively.

Proposed Solution

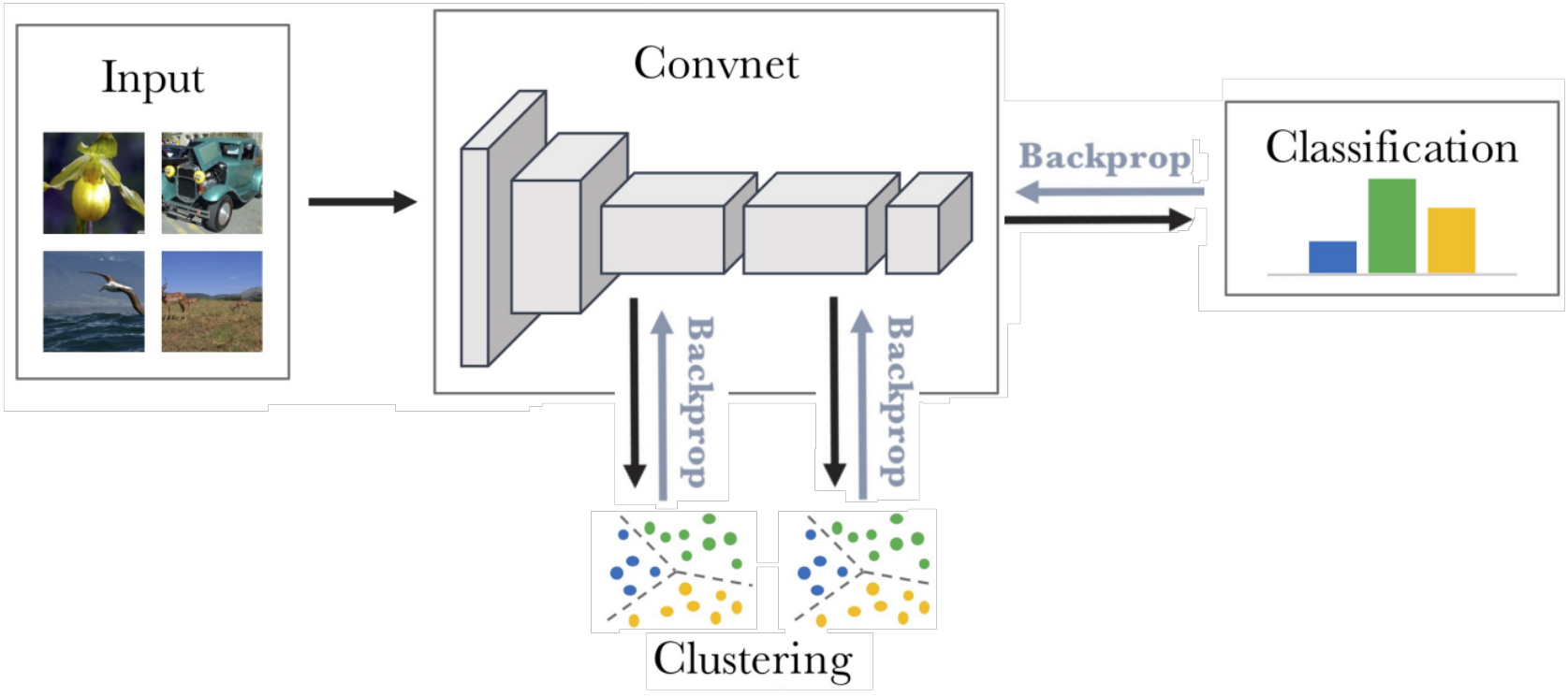


Figure 1: Model architecture.

Observed Clusters

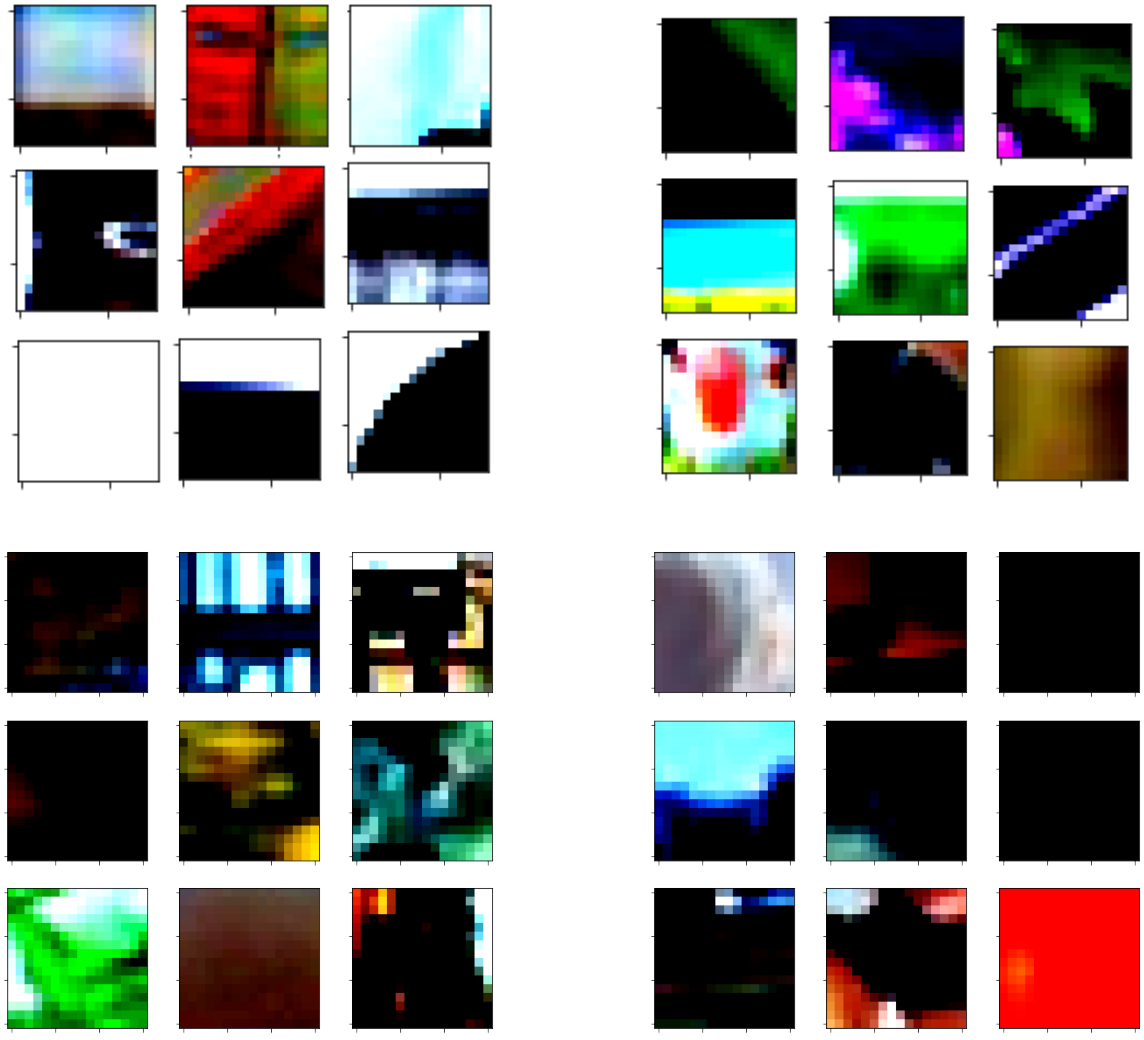
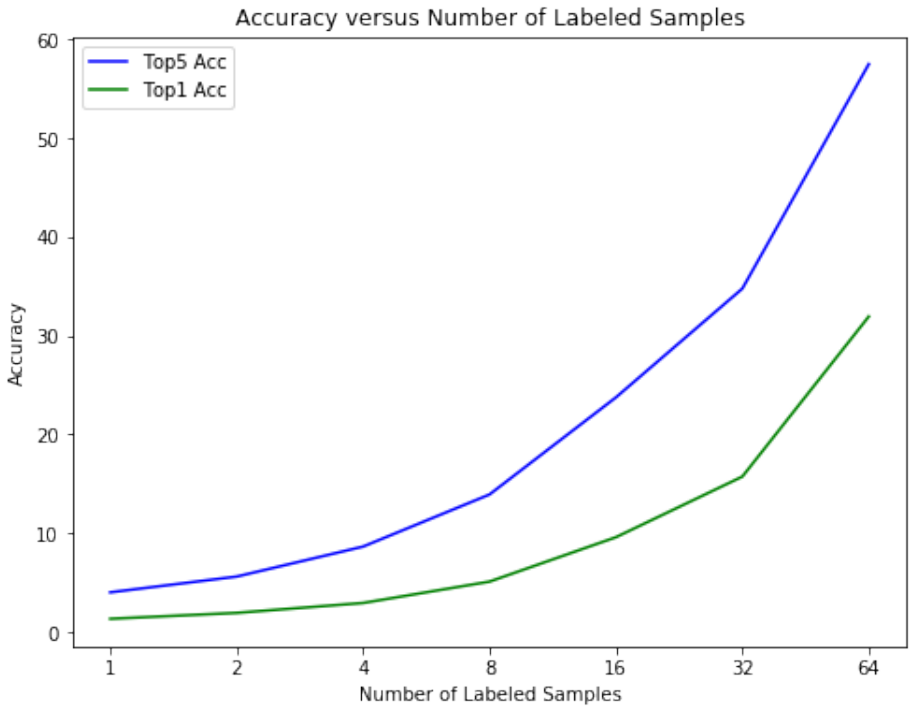
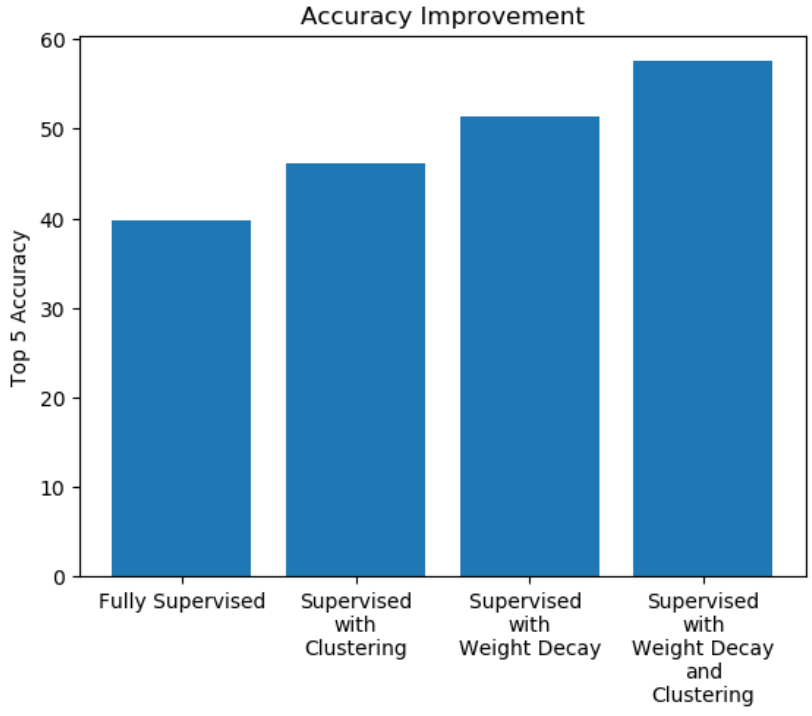


Figure 2: Sampled images from 4 different clusters.

Accuracy


Accuracy versus Number of Labeled Samples
Top5 Acc
Top1 Acc
Accuracy
Number of Labeled Samples
0
10
20
30
40
50
60
1
2
4
8
16
32
64


Accuracy Improvement
Top 5 Accuracy
0
10
20
30
40
50
60
Fully Supervised
Supervised with Clustering
Supervised with Weight Decay
Supervised with Weight Decay and Clustering