Multi-Modal Deep Clustering: Unsupervised Partitioning of Images

Guy Shiran, Daphna Weinshall

School of Computer Science and Engineering, The Hebrew University of Jerusalem

Background

Data Clustering

- Partition data points into groups such that points in each group are more similar to each other than to points in the other groups.

Image Clustering

- Traditional methods (e.g. K-means, GMM, DBSCAN, etc..) do not work well with raw images in pixel-space.
- More meaningful data representation is required for effective clustering.

Deep Clustering

- Solve image representation learning and clustering jointly in a unified framework.

Unsupervised Clustering Algorithm

- Given images $\{x_i\}_{i=1}^n$, fix n targets randomly sampled from a Gaussian Mixture Model:

$$Y = \{y_i | y_i \in \mathbb{R}^d, \|y_i\|_2 = 1\}, \quad |Y| = n$$

- Learn model $f_\theta : X \to \mathbb{R}^d$ and mapping $P : [n] \to [n]$:

$$min_{P,\theta} \frac{1}{n} \sum_i \ell(f_\theta(x_i), y_{P(i)})$$

Optimization:

1. Obtain batch of images b and targets c
2. Compute $f_\theta(X_b)$
3. Compute P* by minimizing L w.r.t P
4. Compute $\nabla_\theta L(\theta)$ using P*
5. Update $\theta \leftarrow \lambda \nabla_\theta L(\theta)$

(3) Is solved with the Hungarian algorithm

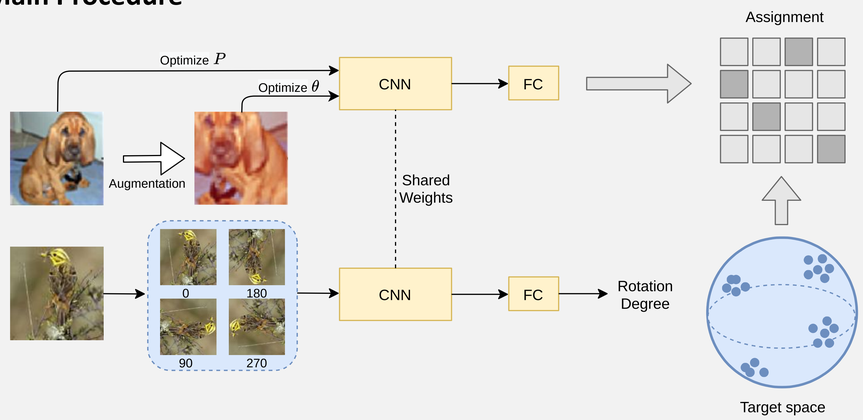
- Assign clusters by mixture component association:

$$c_i = argmin_{j \in [k]} \ell(f_\theta(x_i), \mu_j)$$

where $\{\mu_j\}_{j=1}^k$ are GMM mean vectors.

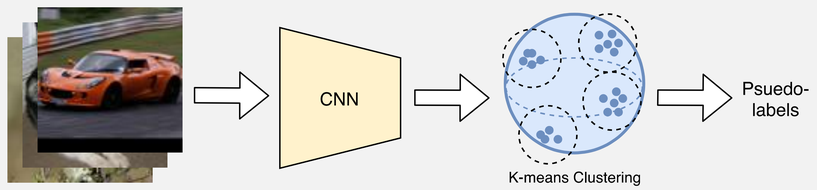
Full Method

Main Procedure



Enhance image features, to facilitate a better clustering, by solving an additional auxiliary self-supervised task of predicting image rotations.

Refinement Stage



In final stage we relax equally sized mixture component assumption. Discard fixed targets, iteratively apply K-means on $f_\theta(X)$ and use cluster assignments as pseudo-targets.

Experimental Results

Clustering results using a ResNet-18 backbone for natural image datasets and a 4-layer CNN for MNIST.

	MNIST		CIFAR-10		CIFAR-100		STL-10		ImageNet-10		Tiny-ImageNet	
	NMI	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI	ACC
k-means	0.499	0.572	0.087	0.228	0.083	0.129	0.124	0.192	0.119	0.241	0.065	0.025
SC	0.663	0.696	0.103	0.247	0.090	0.136	0.098	0.159	0.151	0.274	0.063	0.022
AE	0.725	0.812	0.239	0.313	0.100	0.164	0.249	0.303	0.210	0.317	0.131	0.041
DEC	0.772	0.843	0.257	0.301	0.136	0.185	0.276	0.359	0.282	0.381	0.115	0.037
JULE	0.913	0.964	0.192	0.272	0.103	0.137	0.182	0.277	0.175	0.300	0.102	0.033
DAC	0.935	0.978	0.396	0.522	0.185	0.238	0.249	0.303	0.394	0.527	0.190	0.066
IIC	**0.978**	**0.992**	0.512	0.617	0.224	0.257	0.431	0.499	-	-	-	-
DCCM	-	-	0.496	0.623	0.285	0.327	0.376	0.482	0.608	0.710	0.224	0.108
Ours (avg.)	0.971	0.990	**0.703**	**0.820**	**0.418**	**0.446**	**0.593**	**0.694**	**0.719**	**0.811**	**0.274**	**0.119**
Ours (ste)	±.000	±.000	±.011	±.019	±.003	±.006	±.005	±.013	±.008	±.012	±.001	±.001
Ours (best)	0.973	0.991	0.720	0.843	0.423	0.464	0.609	0.741	0.732	0.830	0.277	0.121

Image Features Evaluation

Evaluate image features of trained ConvNet using a linear evaluation protocol and by applying K-means.

	CIFAR-10			CIFAR-100		
	K-means		Linear	K-means		Linear
	NMI	ACC	ACC	NMI	ACC	ACC
ImNet Labels	0.321	0.407	0.782	0.247	0.281	0.646
NAT	0.044	0.162	0.315	0.037	0.095	0.177
RotNet	0.329	0.349	0.740	0.261	0.284	0.543
NAT+RotNet	0.413	**0.511**	0.764	0.190	0.232	0.499
Ours	**0.428**	0.397	**0.869**	**0.395**	**0.347**	**0.662**

Ablation Study

- Sobel filters are often used as pre-processing to discourage clustering based on trivial cues such as color.
- With a rotation loss this is not only unnecessary, it harms clustering quality.

Sobel	Rotations	ACC	NMI
		0.492	0.428
✔		0.560	0.463
	✔	**0.820**	**0.703**
✔	✔	0.725	0.610

Experiments on CIFAR-10.

Summary

- For the clustering of images we desire meaningful and effective representations.
- We propose a clustering framework that trains a ConvNet by learning cluster assignments alongside model parameters by solving a linear assignment problem using the Hungarian algorithm.
- Random image transformations insert prior knowledge of invariance within clusters into model.
- Auxiliary rotation loss is very effective in helping model learn better image features that produce a quality clustering.