Review Series of Recent Deep Learning Papers:

Parameter Prediction Paper: METRIC LEARNING WITH ADAPTIVE DENSITY DISCRIMINATION

Reviewed by : Arshdeep Sekhon

[1]Department of Computer Science, University of Virginia
https://qdata.github.io/deep2Read/

Good Representations and Classification

- classification algorithms often serve as convenient feature extractors
- training a network for classification on a large dataset, and retaining the outputs of the last layer as feature inputs for other tasks
- But, classification maps each to a single, scalar prediction : dispose of all information but class label

Motivation

- d be able to construct a representation which is amenable to classification, while still maintaining more fine-grained information
- Use Distance Metric Learning Approaches(DML)

Distance Metric Learning Approaches(DML)

- DML learns a transformation to a representation space where distance is in correspondence with a notion of similarity.

Issues with Previous Metric Learning approaches

In Brief:

- Semantic Similarity based on class :
  - destroys intra-class variation and inter class similarity
- To deal with the previous issue, Local Similarity– each example similar to only a few neighbors
  - issue: neighbors fixed a priori in input space: contradiction!
- Triplet loss
  - short sighted: penalizing individual pairs or triplets of examples

Magnet Loss: Key idea

- use a neighborhood, instead of only triplets
- similarity defined in representation space
- New loss that penalises overlap of clusters, and reduces intracluster distances

Magnet Loss: Key idea

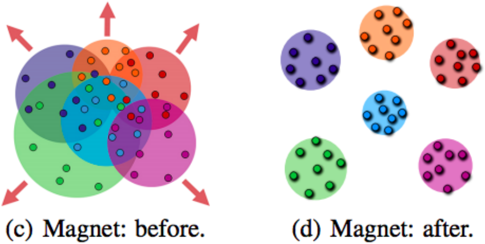

Figure: an entire local neighbourhood of nearest clusters is retrieved, and their overlaps are penalized

- parametrized map: $f(\dot{;}\Theta)$
- representation: $r_n = f(x_n, \Theta)$
- Assignment of clusters: to reduce intra cluster distance

r is representation in new space

$$
\begin{aligned}
\mathcal{I}_1^c, \ldots, \mathcal{I}_K^c &= \arg \min_{I_1^c, \ldots, I_K^c} \sum_{k=1}^{K} \sum_{\mathbf{r} \in I_k^c} \|\mathbf{r} - \boldsymbol{\mu}_k^c\|_2^2 \,, \\
\boldsymbol{\mu}_k^c &= \frac{1}{|I_k^c|} \sum_{\mathbf{r} \in I_k^c} \mathbf{r} \,.
\end{aligned}
$$

Figure: assignment of clusters

Magnet Loss: Formulation

- $C(r)$ is class of r
- $\mu(r)$ is cluster assignment

$$\mathscr{L}(\boldsymbol{\Theta}) = \frac{1}{N}\sum_{n=1}^{N}\left\{-\log\frac{e^{-\frac{1}{2\sigma^2}\|\mathbf{r}_n-\boldsymbol{\mu}(\mathbf{r}_n)\|_2^2-\alpha}}{\sum_{c\neq C(\mathbf{r}_n)}\sum_{k=1}^{K}e^{-\frac{1}{2\sigma^2}\|\mathbf{r}_n-\boldsymbol{\mu}_k^c\|_2^2}}\right\}_+$$

Figure: Loss

Experiments

- Stanford Dogs : Fine Grained Classification
- Object Attributes Dataset: 25 attribute annotations for 90 classes of an updated version of ImageNet, with about 25 annotated examples per class.