Unsupervised Embedding Learning via Invariant and Spreading Instance Feature

Ye, M., Zhang, X., Yuen, P. C., & Chang, S. F. (2020). *Unsupervised Embedding Learning via Invariant and Spreading Instance Feature.* Paper presented at the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).

Primary Problem

unsupervised embedding learning problem, which requires an **effective similarity measurement** between samples in low-dimensional embedding space

Hypotheses

- Motivated by the positive concentrated and negative separated properties
- utilize the **instance-wise supervision** to approximate these properties, which aims at learning data augmentation invariant and instance spread-out features.
- **instance based softmax embedding method**

Methods and Measures used

- Instance-wise Softmax Embedding
    - Softmax Embedding with Classifier Weights.

$$P(i|\mathbf{x}_j) = \frac{\exp(\mathbf{w}_i^T \mathbf{f}_j)}{\sum_{k=1}^{n} \exp(\mathbf{w}_k^T \mathbf{f}_j)}. \tag{1}$$

    - Softmax Embedding with Memory Bank.

$$P(i|\mathbf{x}_j) = \frac{\exp(\mathbf{v}_i^T \mathbf{f}_j/\tau)}{\sum_{k=1}^{n} \exp(\mathbf{v}_k^T \mathbf{f}_j/\tau)}, \tag{2}$$

- **Softmax Embedding on 'Real' Instance Feature**
    - To achieve the goal that features of the same instance under different data augmentations are invariant, while the features of different in- stances are spread-out.
    - In particular, for instance x_i, the augmented sample $\hat{x}_i$ should be classified into instance i, and other instances x_j , $j \neq i$ shouldn't be classified into instance i.

$$P(i|\hat{\mathbf{x}}_i) = \frac{\exp(\mathbf{f}_i^T \hat{\mathbf{f}}_i/\tau)}{\sum_{k=1}^{m} \exp(\mathbf{f}_k^T \hat{\mathbf{f}}_i/\tau)}. \tag{3}$$

$$P(i|\mathbf{x}_j) = \frac{\exp(\mathbf{f}_i^T \mathbf{f}_j/\tau)}{\sum_{k=1}^{m} \exp(\mathbf{f}_k^T \mathbf{f}_j/\tau)}, \; j \neq i \tag{4}$$

$$J = -\sum_i \log P(i|\hat{\mathbf{x}}_i) - \sum_i \sum_{j \neq i} \log(1 - P(i|\mathbf{x}_j)). \quad (7)$$

- Siamese network

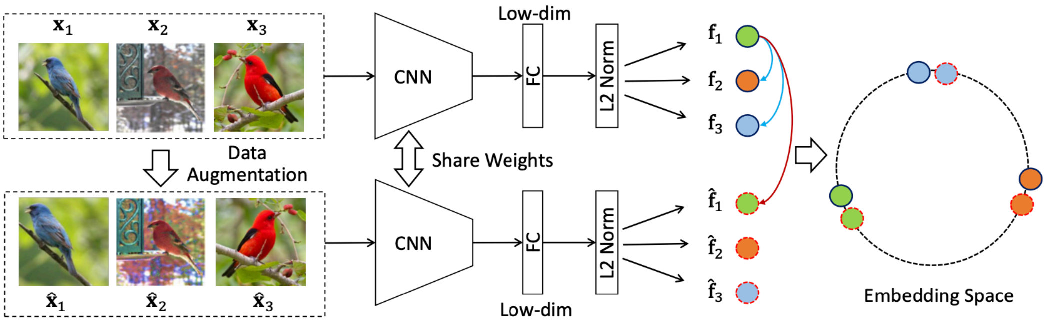

Figure 2: The framework of the proposed unsupervised learning method with Siamese network. The input images are projected into low-dimensional normalized embedding features with the CNN backbone. Image features of the same image instance with different data augmentations are invariant, while embedding features of different image instances are spread-out.

Key Findings

主要是设计了一个**基于实例的softmax嵌入方法/损失函数(instance based softmax embedding method)**

并且分析了公式是如何pull正例对的similarity,以及push负例对的similarity

Dataset

CIFAR-10 [23]

STL-10