Transductive Zero-Shot Learning with Visual Structure Constraint

Ziyu Wan Dongdong Chen Yan Li Xingguang Yan Junge Zhang Yizhou Yu Jing Liao
ziyuwan2-c@my.cityu.edu.hk

Motivation

• Discriminativity of pre-trained CNN: Visual features could be separated automatically.
• Domain Shift: Synthetic centers deviate from real centers.
• Learn a better mapping: Since no point-wise constraint could be directly utilized, how to capture the structure information to alleviate the domain shift problem?

Method

1. VCL (Visual Center Learning)

Semantic Space to Visual Space: MSE Loss of Source Domain:

2. CDVSc (Chamfer-Distance-based VSC)

Inspired by 3D point clouds task, we propose use the symmetric Chamfer distance to solve the structure matching problem.

3. BMVSc (Bipartite-Matching-based VSC)

Consider one important prior for ZSL: The matching relation should conform to one-to-one principle. To address the existed many-to-one matching in CDVSc, we propose to optimize the min-weight perfect matching problem using KM algorithm.

4. WDVSc (Wasserstein-Distance-based VSC)

Global optimal matching is not always valid especially when the approximated visual centers are not accurate enough. From the perspective of discrete distributions, we use the Wasserstein distance to measure the distance.

The Sinkhorn iterations could be written as

Final objective:

Experiment

Experimental results on AwA1, AwA2, CUB, SUN72 and SUN10

Reference