Transferable Contrastive Network for Generalized Zero-Shot Learning

Huajie Jiang[1,2,3,4], Ruiping Wang[1,2], Shiguang Shan[1,2], Xilin Chen[1,2]
[1] Key Laboratory of Intelligent Information Processing of Chinese Academy of Sciences (CAS), Institute of Computing Technology, CAS, Beijing, 100190, China
[2] University of Chinese Academy of Sciences, Beijing, 100049, China
[3]Shanghai Institute of Microsystem and Information Technology, CAS, Shanghai, 200050, China
[4]School of Information Science and Technology, ShanghaiTech University, Shanghai, 200031, China
huajie.jiang@vipl.ict.ac.cn, {wangruiping, sgshan, xlchen}@ict.ac.cn

This document gives more details about the influences of different fusion approaches.

1. Influences of Fusion Approaches

Information fusion is an important part for TCN. In order to demonstrate the effectiveness of the proposed approach, we explore different network structures to fuse the information of images and class semantics:

a) Element-wise product (EP): $\boldsymbol{z}_{ij} = f(\boldsymbol{x}_i) \otimes g(\boldsymbol{a}_j)$

b) Concatenation (C): $\boldsymbol{z}_{ij} = [f(\boldsymbol{x}_i), g(\boldsymbol{a}_j)]$

c) Difference (D): $\boldsymbol{z}_{ij} = |f(\boldsymbol{x}_i) - g(\boldsymbol{a}_j)|$

Table 1 shows the recognition results on AWA1 and CUB. From the comparisons of 'Base' and 'TCN', we can see that TCN achieves more balanced performance for the source and target classes, which indicates that knowledge transfer plays an important role in preventing the model from overfitting the source classes. Moreover, our approach is general to different fusion approaches, among which the element-wise product (EP) shows more robust performance. The superiority of EP may result from that it makes a feature selection for the images, which enhances useful features and weakens useless ones for classification.

Dataset	Fusion	Method	ZSL	GZSL		
				ts	tr	H
AWA1	EP	Base	70.15	9.22	64.78	16.14
		TCN	**70.34**	49.40	76.48	**60.03**
	C	Base	60.76	15.53	55.94	24.31
		TCN	63.63	49.19	71.24	58.20
	D	Base	66.34	11.18	67.74	20.21
		TCN	67.78	47.21	68.65	55.95
CUB	EP	Base	56.62	24.70	64.90	37.84
		TCN	59.54	52.58	52.03	**52.30**
	C	Base	50.91	23.47	56.50	33.17
		TCN	53.53	46.95	43.29	45.05
	D	Base	58.28	27.31	61.55	37.83
		TCN	**60.43**	44.35	55.17	49.17

Table 1. Comparisons of different network structures. 'EP' means element-wise product, which is used in the paper; 'C' means concatenation, where the features are concatenated; 'D' stands for difference, where subtraction operation between two features is performed. 'Base' denotes the model without knowledge transfer.