

本文引用格式: 陈泓达,陈培钦.基于交叉分布对齐的深度自监督多视图聚类方法[J].自动化与信息工程,2022,43(4):12-17.

CHEN Hongda, CHEN Peiqin. Deep self supervision multi view clustering based on cross distribution alignment[J]. Automation & Information Engineering, 2022,43(4):12-17.

基于交叉分布对齐的深度自监督多视图聚类方法

陈泓达 陈培钦

(广东工业大学, 广东 广州 510006)

摘要: 为高效利用多视图数据的一致性和互补性信息, 提高多视图聚类效果, 提出一种基于交叉分布对齐的深度自监督多视图聚类方法。首先, 采用交叉对齐策略学习视图间的潜在联系, 得到多视图共享潜在表示; 然后, 执行聚类操作, 并将聚类结果作为伪标签信息, 建立一条自监督路径; 接着, 在统一的框架中联合学习优化; 最后, 在 3 个公共多视图数据上进行实验, 结果表明, 本文提出方法的聚类评价指标均表现出较好性能。

关键词: 多视图聚类; 变分自编码器; 交叉分布对齐; 自监督

中图分类号: TP391

文献标识码: A

文章编号: 1674-2605(2022)04-0003-06

DOI: 10.3969/j.issn.1674-2605.2022.04.003

Deep Self Supervision Multi View Clustering Based on Cross Distribution Alignment

CHEN Hongda CHEN Peiqin

(Guangdong University of Technology, Guangzhou 510006, China)

Abstract: In order to efficiently use the consistency and complementarity information of multi view data to improve the multi view clustering effect, a deep self supervised multi view clustering method based on cross distribution alignment is proposed. First, the cross alignment strategy is used to learn the potential relationship between views and obtain the potential representation shared by multiple views; Then, the clustering operation is performed, and the clustering result is taken as pseudo label information to establish a self supervised path; Then, joint learning optimization in a unified framework; Finally, experiments are carried out on three public multi view data, and the results show that the clustering evaluation indexes of the proposed method show good performance.

Keywords: multi view clustering; variational autoencoder; cross distribution alignment; self supervision

0 引言

聚类作为一种无监督学习方法备受研究者关注, 并在计算机视觉和机器学习等领域被广泛研究^[1-2]。随着互联网和传感器技术的迅速发展, 相关领域的数据每天都呈指数增长, 且这些数据已从单一视图描述演变成各种类型的多视图描述, 从而形成多视图数据。多视图数据从不同角度描述同一对象的多种信息, 提供丰富互补信息, 使不同视图之间既存在联系又存在差异。

多视图聚类 (multi-view clustering, MVC) 是将多

视图特征信息相似的对象分到同一类, 不相似的对象分到不同类, 且可跨不同视图搜索其一致性的集群。自 1967 年以来, 研究人员已提出许多聚类算法, 如 K-means^[3]、谱聚类^[4]和子空间聚类^[5-8]等。但这些算法仅适用于单视图数据, 无法扩展应用于多视图数据。

从 21 世纪初开始, 多视图聚类方法得到多方面研究。NGIAM 等^[9]通过训练双峰深度自动编码器获取视图共享表示; WANG 等^[10]扩展典型相关性分析, 引入自编码器正则化, 提出一种深度典型性相关自动编码器 (deep canonically correlation autoencoder,

DCCAE), 但上述方法仅适用于 2 个视图的情况。2012 年, HUANG 等^[11]提出谱聚类的亲和度聚合 (affinity aggregation for spectral clustering, AASC) 算法, 将谱聚类扩展应用于多视图数据, 并取得较好效果。2014 年, XIA 等^[12]利用各视图转移概率矩阵获取共享低秩转移概率矩阵, 并通过低秩和稀疏分解的鲁棒多视图光谱聚类 (robust multiview spectral clustering, RMSC) 方法, 取得令人满意的聚类效果。考虑到不同视图对全局聚类能力的差异, NIE 等^[13]提出多视图自适应 Procrustes 加权 (adaptively weighted Procrustes, AWP) 方法, 对 Procrustes 均值进行加权改进, 并应用于多视图数据。随着图学习的快速兴起, 许多研究者将图学习应用于多视图聚类。ZHAN 等^[14]提出的多视图共识图聚类 (multiview consensus graph clustering, MCGC), 通过多视图学习具有 k 个连通分量的共识图, 可直接从共识图中获取聚类结果。LIANG 等^[15]在统一的目标函数中, 同时

对多视图一致性和不一致性进行建模, 提出相似图融合 (similarity graph fusion, SGF) 和相异图融合 (dissimilarity graph fusion, DFG) 2 种方法, 前者通过相似图对多个视图信息进行融合, 后者利用距离进行视图信息融合。

虽然现有的多视图聚类研究取得了一些进展, 但多视图数据内在的有用信息尚未被充分利用, 主要难点在于如何挖掘多视图数据的一致性和互补性信息来提升聚类效果。为此, 本文将深度学习应用于多视图学习, 并受自监督思想启发, 提出一种基于交叉分布对齐的深度自监督多视图聚类 (deep self-supervision multi-view clustering based on cross-distribution alignment, DSMVCCDA) 方法, 可有效提升多视图数据聚类效果。

1 基本架构

本文提出的基于交叉分布对齐的深度自监督多视图聚类 (DSMVCCDA) 网络架构如图 1 所示。

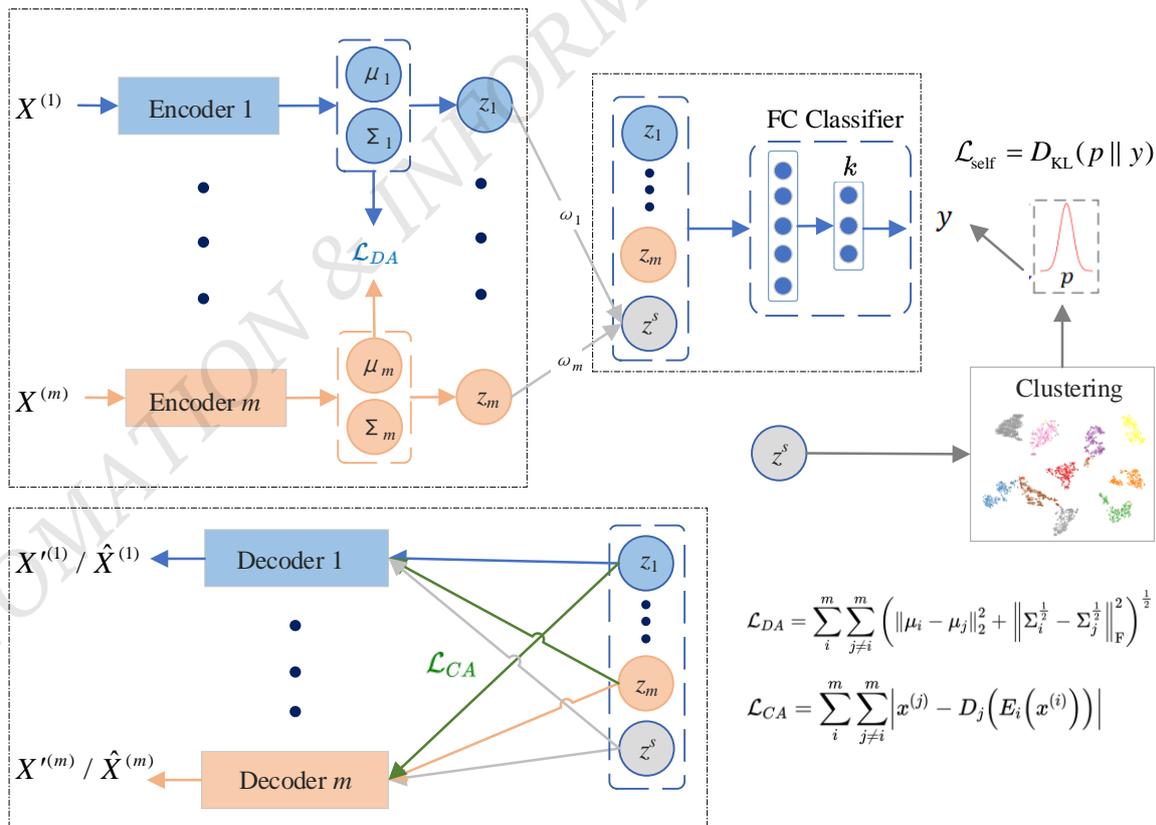


图 1 基于交叉分布对齐的深度自监督多视图聚类网络架构

由图 1 可知, DSMVCCDA 主要由 3 部分构成:

- 1) 交叉分布对齐 \mathcal{L}_{DA} 和 \mathcal{L}_{CA} , 可有效学习多视图数据的一致性和互补性信息;
- 2) 生成共享潜在表示 z^s , 将变分自编码器 (variational autoencoder, VAE) [16] 学习到的每个视图潜在表示进行加权融合, 得到多视图的共享潜在表示;
- 3) 自监督聚类 \mathcal{L}_{self} , 将共享潜在表示 z^s 执行聚类操作后得到的聚类标签作为伪标签, 监督优化分类器网络, 通过神经网络反向传播联合优化网络参数, 提高网络特征学习能力和聚类效果。

1.1 交叉分布对齐

交叉分布对齐是指利用 VAE 网络中编码器 (encoder) 对视图潜在分布进行对齐, 同时为减少分布对齐损失的信息, 在解码器 (decoder) 重构过程中引入交叉对齐来平衡多视图数据的学习。

由于 DSMVCCDA 的目标是学习多个视图间的交叉互补信息的组合形式, 故 VAE 网络结构包含 m 个视图编码器, 且每一个视图编码器对应一个解码器, 以映射到潜在表示空间。为使 VAE 网络在学习多个视图交叉互补信息的同时, 能最大限度地减少信息丢失, 需通过解码器网络重建原始数据, 所以 DSMVCCDA 模型的基本 VAE 损失是 m 个视图的 VAE 损失之和:

$$\mathcal{L}_{VAE} = \sum_i^m \mathbb{E}_{q_\phi(z|x)} \left[\log p_\theta(x^{(i)}|z) \right] - \delta D_{KL} \left(q_\phi(z|x^{(i)}) \| p_\theta(z) \right) \quad (1)$$

式中: 第一项为解码网络 $p_\theta(x^{(i)}|z)$ 重构误差; 第二项为编码网络 $q_\phi(z|x^{(i)})$ 和先验分布 $p_\theta(z)$ 的 KL 散度。

为使 DSMVCCDA 学习到多视图的一致性信息, 采用 Wasserstein 距离 [17] 最小化不同视图间的距离, 对齐潜在分布。设 2 个分布 i 和 j 之间的 W 距离 Y_{ij} 的闭合形式解为

$$Y_{ij} = \left(\|\mu_i - \mu_j\|_2^2 + \|\Sigma_i^{\frac{1}{2}} - \Sigma_j^{\frac{1}{2}}\|_F^2 \right)^{\frac{1}{2}} \quad (2)$$

式中:

- μ ——多元高斯分布的均值;
- Σ ——多元高斯分布的方差。

对具有 m 组分布对齐的多视图数据, 分布对齐损失函数可表示为

$$\mathcal{L}_{DA} = \sum_i^m \sum_{j \neq i}^m Y_{ij} \quad (3)$$

考虑到分布对齐给学习多视图一致性信息带来的较强约束, 引入交叉对齐可以平衡 DSMVCCDA 模型对多视图一致性和互补性信息的学习。交叉对齐利用其他视图在 VAE 学习到的潜在表示, 加入自身视图特定的解码器进行交叉学习, 可以在保留特定视图信息的情况下, 利用其他视图信息提高模型的学习能力。交叉对齐损失函数为

$$\mathcal{L}_{CA} = \sum_i^m \sum_{j \neq i}^m \left| x^{(j)} - D_j \left(E_i \left(x^{(i)} \right) \right) \right| \quad (4)$$

式中:

- $x^{(j)}$ ——第 j 个视图的原始数据;
- D_j ——第 j 个视图的解码器;
- $E_i(x^{(i)})$ ——第 i 个视图通过编码器生成的潜在表示。

1.2 共享潜在表示 z^s

考虑到不同视图的重要性差异问题, 利用视图权重参数 $W = [\omega_1, \omega_2, \dots, \omega_m]^T$ ($\omega_i \geq 0, \sum \omega_i = 1$), 对每个视图的交叉潜在特征进行加权融合, 得到共享潜在表示 z^s , 作为后续聚类层的输入实现聚类优化。

$$z^s = \sum_i^m \omega_i z_i \quad (5)$$

式中:

- ω_i ——第 i 个视图的权重系数;
- z_i ——第 i 个视图的权重潜在变量。

在生成多视图共享潜在表示的融合层加入重构损失, 使视图权重参数利用神经网络反向传播 [18] 自动更新, 共享潜在表示 z^s 的重构损失表示为

$$\mathcal{L}_{\text{Fusion}} = \sum_i^m \|x^{(i)} - \hat{x}^{(i)}\|^2 \quad (6)$$

式中：

$x^{(i)}$ ——第 i 个视图的原始数据；

$\hat{x}^{(i)}$ ——利用 z^s 解码得到的重构数据。

1.3 自监督聚类

本文借鉴深度嵌入式聚类 (deep embedding for clustering, DEC)^[19] 定义基于质心的软分配分布，最小化其与辅助目标分布的 Kullback-Leibler (KL) 散度差异，在 DSMVCCDA 网络联合进行优化学习。首先，DEC 将嵌入特征 z_i 进行 K-means 运算，获得网络参数 f_θ 的初始估计，并得出初始簇质心 μ_j ；然后，将学生 t 分布作为软分配，测量嵌入特征 z_i 与簇质心 μ_j 之间的相似性：

$$q_{ij} = \frac{(1 + \|z_i - \mu_j\|^2 / a)^{-\frac{a+1}{2}}}{\sum_j (1 + \|z_i - \mu_j'\|^2 / a)^{-\frac{a+1}{2}}} \quad (7)$$

式中：

a —— t 分布自由度，由于无法在无监督方式下对其进行交叉验证^[19]，不失一般性，故设置为 1；

q_{ij} ——第 i 个数据属于第 j 类的概率。

DEC^[19] 定义的辅助目标分布 p 为

$$p_{ij} = \frac{q_{ij}^2 / f_j}{\sum_j q_{ij}^2 / f_j} \quad (8)$$

式中：

$f_j = \sum_i q_{ij}$ ——软集群的频率。

上述目标分布可通过提高软分配的高置信度得分来指导聚类^[19]，故聚类损失为

$$\mathcal{L}_c = KL(p \| q) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}} \quad (9)$$

由于多视图聚类是不受监督学习的过程，无法得知学习的多视图特征信息的有效贡献程度。因此，为进一步挖掘多视图数据内在信息，提出了自监督学习方法。与无监督学习相比，自监督学习可利用数据集

本身信息构造伪标签信息，监督网络的训练过程。本文提出的 DSMVCCDA，将 VAE 网络学习到的多视图共享潜在表示执行聚类操作后得到的具有标签信息的分布 p 作为伪标签，设计一个具有两层全连接层的分类器作为下游任务，该分类器的输入由所有视图的潜在表示和共享潜在表示组合而成，输出层则使用 Softmax 函数得到预测分类的概率分布。

本文采用 KL 散度来衡量 2 个概率分布的差异，并将其作为损失函数来优化分类器网络参数。自监督分类器损失函数为

$$\mathcal{L}_{\text{self}} = D_{\text{KL}}(p \| y) = \sum p(x_i) \log \frac{p(x_i)}{y(x_i)} \quad (10)$$

式中：

$p(x_i)$ ——样本在伪标签中的概率分布；

$y(x_i)$ ——样本通过分类器得到的概率分布。

结合多视图特征学习和聚类损失，DSMVCCDA 模型的总体目标损失函数为

$$\mathcal{L}_{\text{DSMVCCDA}} = \mathcal{L}_{\text{VAE}} + \alpha \mathcal{L}_{\text{DA}} + \varepsilon \mathcal{L}_{\text{CA}} + \beta \mathcal{L}_{\text{Fusion}} + \gamma \mathcal{L}_c + \eta \mathcal{L}_{\text{self}} \quad (11)$$

式中：

$\alpha, \varepsilon, \beta, \gamma, \eta$ ——分布对齐损失、交叉对齐损失、加权融合损失、聚类损失、自监督损失的平衡参数。

2 数据集与参数设置

为验证本文提出 DSMVCCDA 方法的有效性，选取 3 个公开数据集 MSRC-v1^[20]、NUS-WIDE^[21] 和 Caltech101 进行对比实验，数据集统计信息如表 1 所示。

表 1 3 个数据集统计信息

数据集	视图数/ 个	类别数/ 类	样本数/ 个
MSRC-v1	5	7	210
NUS-WIDE	5	31	2 000
Caltech101	6	102	9 144

实验中自编码器采用三层完全连接的神经网络

dv-500-500-1024- z^s , 解码器和编码器的网络结构对称。 z^s 的维度从侧面反映网络模型对视图数据特征信息的学习能力。为实现以更少的特征维度信息达到更好的聚类效果, 经过调参实验, 维度设置为 10。在 3 个数据集实验中, 公式(11)的各个参数设置最优值如表 2 所示。

表 2 3 个数据集的最优参数设置

数据集	α	ϵ	β	γ	η
MSRC-v1	0.54	0.04	10	10	10
NUS-WIDE	0.54	0.04	10	10	20
Caltech101	0.54	0.05	20	10	15

3 实验结果与分析

3.1 实验环境配置

本文实验基于 Python3.6 编程语言, 采用深度学习框架 Pytorch1.1.0 搭建环境, 显卡配置为 GTX 1080Ti 和 CUDA10.0。

3.2 实验结果与分析

实验选用 AASC、RMSC、AWP、MCGC、SGF、DFG 等多视图聚类算法, 在 3 个数据集上进行对比实验。聚类评价指标使用聚类准确性 (accuracy, ACC) 和纯度 (purity)。其中, ACC 用于测量聚类算法获得的实际标签和预测标签之间的准确性; purity 则计算正确聚类数占总数的比例。2 个指标值越大说明聚类算法性能越好。ACC 和 purity 结果取 10 次实验平均值和标准差作为最终的聚类结果, 结果分别如表 3 和表 4 所示。

表 3 多视图聚类算法在不同数据集上的 ACC 值 %

对比方法	MSRC-v1	NUS-WIDE	Caltech101
AASC	77.33±0.25	15.70±0.19	23.80±0.77
RMSC	71.05±1.70	15.49±0.62	22.77±0.93
AWP	76.19±0.00	14.60±0.00	26.22±0.00
MCGC	84.76±0.00	12.75±0.00	23.00±0.00
SGF	80.43±0.15	16.22±0.45	24.04±0.45
DFG	87.14±0.00	16.80±0.12	23.53±0.39
DSMVCCDA	94.29±0.31	21.30±0.25	30.08±0.05

表 4 多视图聚类算法在不同数据集上的 purity 值 %

对比方法	MSRC-v1	NUS-WIDE	Caltech101
AASC	77.33±0.25	23.92±0.40	40.11±0.39
RMSC	76.14±1.80	24.48±0.30	38.98±0.34
AWP	79.52±0.00	22.85±0.00	42.79±0.00
MCGC	84.76±0.00	22.40±0.00	43.12±0.00
SGF	83.81±0.00	25.86±0.43	46.26±0.21
DFG	87.14±0.00	26.77±0.31	46.28±0.34
DSMVCCDA	94.29±0.31	29.55±0.15	50.35±0.04

由表 3、表 4 可知: DSMVCCDA 方法在聚类指标 ACC 上, 数据集 MSRC-v1、NUS-WIDE 和 Caltech101 分别比次优值高 7.15%、4.5% 和 3.86%; 在聚类指标 purity 上, 数据集 MSRC-v1、NUS-WIDE 和 Caltech101 分别比次优值高 7.15%、2.78% 和 4.07%。

从不同数据集聚类效果的数值和稳定性来看, 同一种多视图聚类方法在 3 个数据集上表现不尽相同, 主要原因是多视图数据集本身结构的复杂性。主要表现在:

1) 视图间相关性不强, 如 NUS-WIDE 是由 5 个低级特征组成的多视图, 其中颜色直方图 (color histogram, CH)、块方式颜色矩 (color moment of block mode, CM) 和颜色相关图 (color correlation, CORR) 具有一定的相关性, 而与边缘方向直方图 (edge direction histogram, EDH) 和小波纹理 (wavelet texture, WT) 相关性不强, 可能造成聚类结果偏低;

2) 低质量分类, 数据集需要聚类的类别数量是按照人为拟定的标准给定, 标准可能过于苛刻, 且由于网络没有监督信息, 也可能造成聚类结果偏低;

3) 不平衡分类, 即数据集的聚类类别数量和样本数不平衡, 给网络训练带来一定的学习偏差。

得益于神经网络强大的拟合非线性能力, 本文提出的 DSMVCCDA 方法采用深度生成式模型变分自编码器, 可通过学习共享潜在表示来挖掘视图间的潜在相关性, 且引入自监督学习方法可使网络减弱数据集本身结构带来的影响, 充分挖掘多视图数据集内在信息, 因此, 本文方法在 3 个数据集上表现稳定且优异。

4 结语

为高效挖掘多视图数据的一致性和互补性信息, 本文提出一种基于交叉分布对齐的深度自监督多视图聚类方法。该方法在变分自编码器中使用交叉分布对齐策略学习多视图共享潜在表示, 执行聚类操作, 并将聚类结果标签信息作为伪标签来监督网络的学习, 共同优化网络参数。实验结果表明, 本文方法在 3 个公共数据集上的聚类效果比其他多视图聚类方法表现更优异稳定, 证明了该方法的有效性。

参考文献

- [1] CARON M, BOJANOWSKI P, JOULIN A et al. Deep clustering for unsupervised learning of visual features[C]// Proceedings of the European conference on computer vision (ECCV), 2018:132-149.
- [2] CHANG J, WANG L, MENG G, et al. Deep adaptive image clustering[C]//Proceedings of the IEEE International Conference on Computer Vision, 2017:5879-5887.
- [3] HARTIGAN J A, WONG M A. A K-means clustering algorithm[J]. Journal of the Royal Statistical Society(Applied Statistics), 1979,28(1):100-108.
- [4] GUO C, ZHENG S, XIE Y, et al. A survey on spectral clustering[C]//World Automation Congress. IEEE, 2012:53-56.
- [5] VIDAL R. Subspace clustering[J]. IEEE Signal Processing Magazine, 2011,28(2):52-68.
- [6] GAO H, NIE F, LI X, et al. Multi-view subspace clustering [C]//IEEE International Conference on Computer Vision. IEEE, 2015:4238-4246.
- [7] ZHANG C, FU H, LIU S, et al. Low-rank tensor constrained multiview subspace clustering[C]//IEEE International Conference on Computer Vision. IEEE, 2015:1582-1590.
- [8] YIN M, GAO J, XIE S, et al. Multiview subspace clustering via tensorial t-product representation[J]. IEEE Transactions on Neural Networks and Learning Systems, 2018,30(3):851-864.
- [9] NGIAM J, KHOSLA A, KIM M, et al. Multimodal deep learning[C]//ICML, 2011.
- [10] WANG W, ARORA R, LIVESCU K, et al. On deep multi-view representation learning[C]//International Conference on Machine Learning. PMLR, 2015:1083-1092.
- [11] HUANG H C, CHUANG Y Y, CHEN C S. Affinity aggregation for spectral clustering[C]//2012 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2012: 773-780.
- [12] XIA R, PAN Y, Du L, et al. Robust multi-view spectral clustering via low-rank and sparse decomposition[C]//Proceedings of the AAAI conference on artificial intelligence, 2014, 28(1).
- [13] NIE F, TIAN L, LI X. Multiview clustering via adaptively weighted Procrustes[C]//Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining, 2018:2022-2030.
- [14] ZHAN K, NIE F, WANG J, et al. Multiview consensus graph clustering[J]. IEEE Transactions on Image Processing, 2018, 28(3):1261-1270.
- [15] LIANG Y, HUANG D, WANG C D, et al. Multi-view graph learning by joint modeling of consistency and inconsistency [J]. arXiv preprint arXiv, 2008:10208,2020.
- [16] KINGMA D P, WELLMING M. Auto-encoding variational bayes y[J]. arXiv preprint arXiv;1312.6114,2013.
- [17] GIVENS C R, SHORTT R M. A class of Wasserstein metrics for probability distributions[J]. The Michigan Mathematical Journal,1984,31(2):231-240.
- [18] RUMELHART D E, HINTON G E, WILLIAMS R J. Learning representations by back-propagating errors[J]. Nature, 1986,323(6088):533-536.
- [19] XIE J, GIRSHICK R, FARHADI A. Unsupervised deep embedding for clustering analysis[C]//Proceedings of the 33rd International Conference on International Conference on Machine Learning-Volume 48, 2016:478-487.
- [20] WINN J, JOJIC N. LOCUS: learning object classes with unsupervised segmentation[C]//Tenth IEEE International Conference on Computer Vision (ICCV'05), 2005,1:756-763.
- [21] CHUA T S, TANG J, HONG R, et al. Nus-wide: a real-world web image database from national university of Singapore [C]//Proceedings of the ACM international conference on image and video retrieval, 2009:1-9.

作者简介:

陈泓达, 男, 1995 年生, 硕士研究生, 主要研究方向: 深度多视图聚类、机器视觉等。E-mail: chenrongdaCHD@163.com
陈培钦, 男, 1996 年生, 硕士研究生, 主要研究方向: 模式识别、机器视觉等。