C3: Cross-instance guided Contrastive Clustering

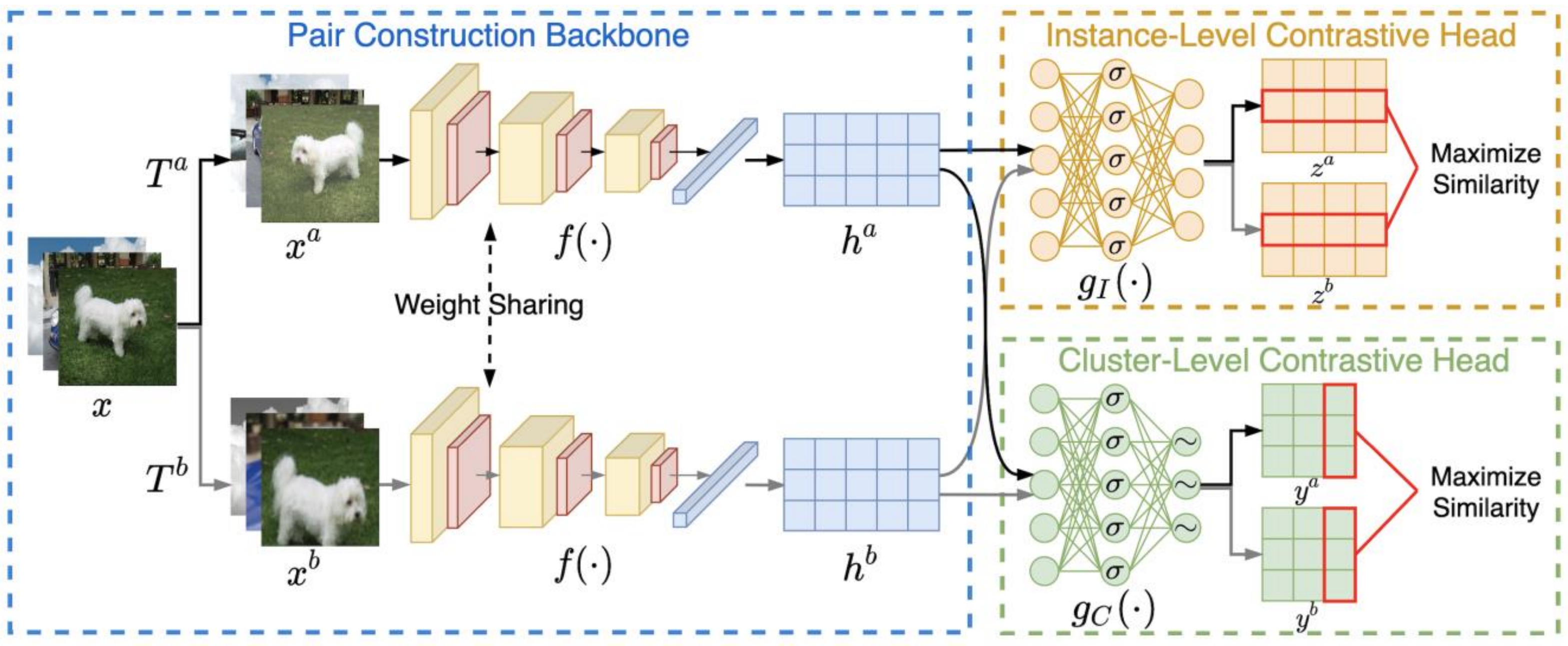
Mohammadreza Sadeghi, Hadi Hojjati, Narges Armanfard

Department of Electrical and Computer Engineering, McGill University Montreal, QC, Canada
Mila - Quebec AI Institute Montreal, QC, Canada

Problem Statement and Related work

- **Problem Statement:** Given an unlabelled dataset $X = \{x_1, x_2, \ldots, x_N\}$ and a predefined cluster number parameter M the goal of the clustering problem is to partition X into M disjoint groups.

- **Contrastive clustering (CC) [1]** is a method that combines ideas from both clustering and contrastive learning. It aims to group similar data points together while pushing dissimilar data points apart in an unsupervised manner. The core idea is to use a contrastive loss function that encourages embeddings of data points from the same cluster to be closer while pushing apart those from different clusters.



$$\ell_i^a = -\log \frac{\exp(s(z_i^a, z_i^b)/\tau_I)}{\sum_{j=1}^{N}[\exp(s(z_i^a, z_j^a)/\tau_I) + \exp(s(z_i^a, z_j^b)/\tau_I)]}$$

$$\hat{\ell}_i^a = -\log \frac{\exp(s(y_i^a, y_i^b)/\tau_C)}{\sum_{j=1}^{M}[\exp(s(y_i^a, y_j^a)/\tau_C) + \exp(s(y_i^a, y_j^b)/\tau_C)]}$$

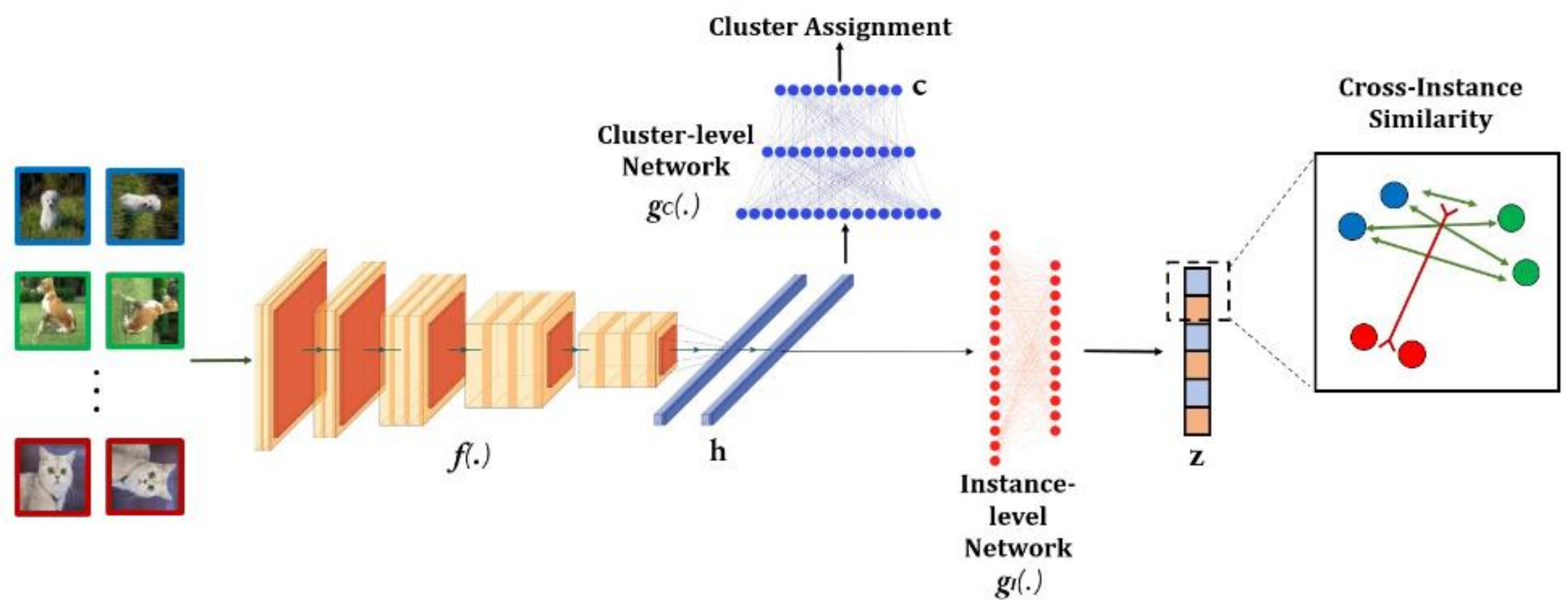
Proposed Method

Drawbacks of existing Contrastive Clustering methods:

1. ignore cross-instance patterns which carry essential information for finding potential positive pairs.
2. consider all other data points in a batch as negative sample even if they are the same objects in two different images.

Our Contribution:

1. We propose a groundbreaking technique to discover similarities between samples.
2. we propose a novel weighting scheme that aims to separate more challenging data samples. With this weighting scheme, we aim at reducing the effect of false negative, noisy, and anomaly samples.



Our Proposed Loss function:

$$\tilde{\ell}_i^a = -\log \frac{\sum_{k\in\{a,b\}}\sum_{j=1}^{N} \mathbb{1}\{z_i^{a\top} z_j^k \geq \zeta\}\exp(z_i^{a\top} z_j^k)}{\sum_{k\in\{a,b\}}\sum_{j=1}^{N} w_{ij}^k \exp(z_i^{a\top} z_j^k)},$$

1. If, for a pair of instances, the similarity is greater than or equal to a threshold ζ, we consider those samples to be similar and pull them closer together
2. Weights w_{ij}^k aims to reduce the effect of false negative, noisy and anomaly samples.

- **Weighting Scheme:**

To obtain optimum value for weights, we Propose to solve the following optimization problem:

$$\min_{w_{ij}^k} \sum_{k\in\{a,b\}} \sum_{j=1}^{N} -w_{ij}^k(1 - |z_i^{a\top} z_j^k|) - \frac{1}{\Gamma}\mathrm{H}(W_i) \qquad s.t. \sum_{k\in\{a,b\}} \sum_{j=1}^{N} w_{ij}^k = 1$$

1. By considering the first term, we assign lower weights to samples that are too close and too far from each other.
2. We add the second term to avoid trivial solution.
3. We solve this optimization problem using Lagrange multiplier technique The final close-form solution is as follow.

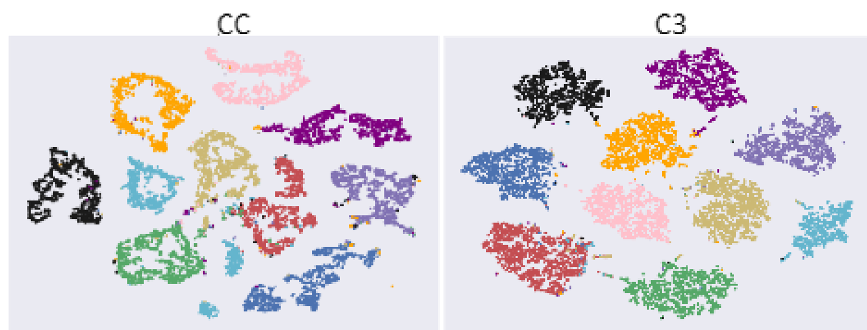
$$w_{ij}^k = \frac{\exp(\Gamma(1 - |z_i^{a\top} z_j^k|))}{\sum_{k\in\{a,b\}}\sum_{j=1}^{N}\exp(\Gamma(1 - |z_i^{a\top} z_j^k|))}$$

Experiments and Results

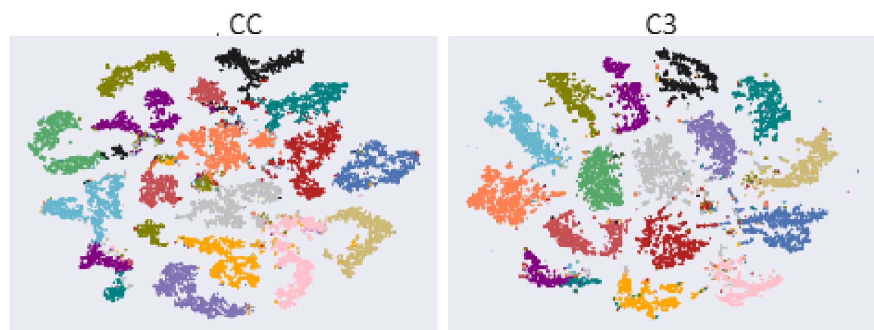
Table 1: Clustering performance of different methods.

Algorithm	CIFAR-10 NMI	CIFAR-10 ACC	CIFAR-10 ARI	CIFAR-100 NMI	CIFAR-100 ACC	CIFAR-100 ARI	ImageNet-10 NMI	ImageNet-10 ACC	ImageNet-10 ARI	ImageNet-Dogs NMI	ImageNet-Dogs ACC	ImageNet-Dogs ARI	Tiny-ImageNet NMI	Tiny-ImageNet ACC	Tiny-ImageNet ARI
K-means [24]	0.087	0.229	0.049	0.084	0.130	0.028	0.119	0.241	0.057	0.055	0.105	0.020	0.065	0.025	0.005
SC [41]	0.103	0.247	0.085	0.090	0.136	0.022	0.151	0.274	0.076	0.038	0.111	0.013	0.063	0.022	0.004
AC [9]	0.105	0.228	0.065	0.098	0.138	0.034	0.138	0.242	0.067	0.037	0.139	0.021	0.069	0.027	0.005
NMF [2]	0.081	0.190	0.034	0.079	0.118	0.026	0.132	0.230	0.065	0.044	0.118	0.016	0.072	0.029	0.005
AE [1]	0.239	0.314	0.169	0.100	0.165	0.048	0.210	0.317	0.152	0.104	0.185	0.073	0.131	0.041	0.007
DAE [34]	0.251	0.297	0.163	0.111	0.151	0.046	0.206	0.304	0.138	0.104	0.190	0.078	0.127	0.039	0.007
DCGAN [27]	0.265	0.315	0.176	0.120	0.151	0.045	0.225	0.346	0.157	0.121	0.174	0.078	0.135	0.041	0.007
DeCNN [40]	0.240	0.282	0.174	0.092	0.133	0.038	0.186	0.313	0.142	0.098	0.175	0.073	0.111	0.035	0.006
VAE [18]	0.254	0.291	0.167	0.108	0.152	0.040	0.193	0.334	0.168	0.107	0.179	0.079	0.113	0.036	0.006
JULE [39]	0.192	0.272	0.138	0.103	0.137	0.033	0.175	0.300	0.138	0.054	0.138	0.028	0.102	0.033	0.006
DEC [37]	0.275	0.301	0.161	0.136	0.185	0.050	0.282	0.381	0.203	0.122	0.195	0.079	0.115	0.037	0.007
DAC [4]	0.396	0.522	0.306	0.185	0.238	0.088	0.394	0.527	0.302	0.219	0.275	0.111	0.190	0.066	0.017
ADC [12]	-	0.325	-	-	0.160	-	-	-	-	-	-	-	-	-	-
DDC [5]	0.424	0.524	0.329	-	-	-	0.433	0.577	0.345	-	-	-	-	-	-
DCCM [35]	0.496	0.623	0.408	0.285	0.327	0.173	0.608	0.710	0.555	0.321	0.038	0.182	0.224	0.108	0.038
IIC [17]	-	0.617	-	-	0.257	-	-	-	-	-	-	-	-	-	-
PICA [15]	0.591	0.696	0.512	0.310	0.337	0.171	0.802	0.870	0.761	0.352	0.352	0.201	0.277	0.098	0.040
GATCluster [25]	0.475	0.610	0.402	0.215	0.281	0.116	0.609	0.762	0.572	0.322	0.333	0.200	-	-	-
CC [22]	0.678*	0.770*	0.607*	0.421*	0.423*	0.261*	0.850*	0.893*	0.811*	0.436*	0.421*	0.268*	0.331*	0.137*	0.062*
EDESC [3]	0.627	0.464	-	0.385	0.370	-	-	-	-	-	-	-	-	-	-
C3 (Ours)	**0.743**	**0.836**	**0.703**	**0.435**	**0.456**	**0.274**	**0.905**	**0.943**	**0.860**	**0.447**	**0.434**	**0.280**	**0.335**	**0.140**	**0.064**

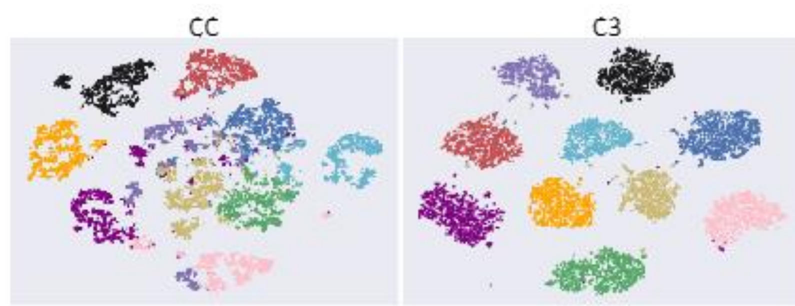
- **TSNE Visualization**



(a) ImageNet-10



(b) ImageNet-Dogs



CIFAR-10

Conclusion

In this paper, we proposed C3, an algorithm for contrastive data clustering that incorporates the similarity between different instances to form a better representation for clustering. We experimentally showed that our method could significantly outperform the state-of-the-art on five challenging computer vision datasets. In addition, through additional experiments, we evaluated different aspects of our algorithm and provided several intuitions on how and why our proposed scheme can help in learning a more cluster-friendly representation.

References

[1] Li, Yunfan, et al. "Contrastive clustering." *Proceedings of the AAAI conference on artificial intelligence*. Vol. 35. No. 10. 2021.