Introduction to Deep Clustering

발표자 : 백인성

2023.02.17.

Data Mining
Quality Analytics

발표자 소개

Insung Baek

❖ 백인성 (Insung Baek)

- Korea University
- Data Mining & Quality Analytics Lab
- Ph.D. Candidate (2018. 9 ~ Present)

❖ Research Interest

- Explainable Artificial Intelligence algorithms (Explainable A.I.)
- Game Artificial Intelligence (Game A.I.)
- Application of deep learning and machine learning algorithms.
- Deep learning algorithms for image.

❖ Contact

- E-mail: insung_baek01@korea.ac.kr

Copyright ⓒ 2023, All rights reserved.

목차

1. Introduction

2. Deep Clustering

3. **C**ontrastive **C**lustering (CC)

4. **Pro**totype Scattering and **Pos**itive Sampling (ProPos)

5. Conclusions

Copyright ⓒ 2023, All rights reserved.

Data Mining
Quality Analytics

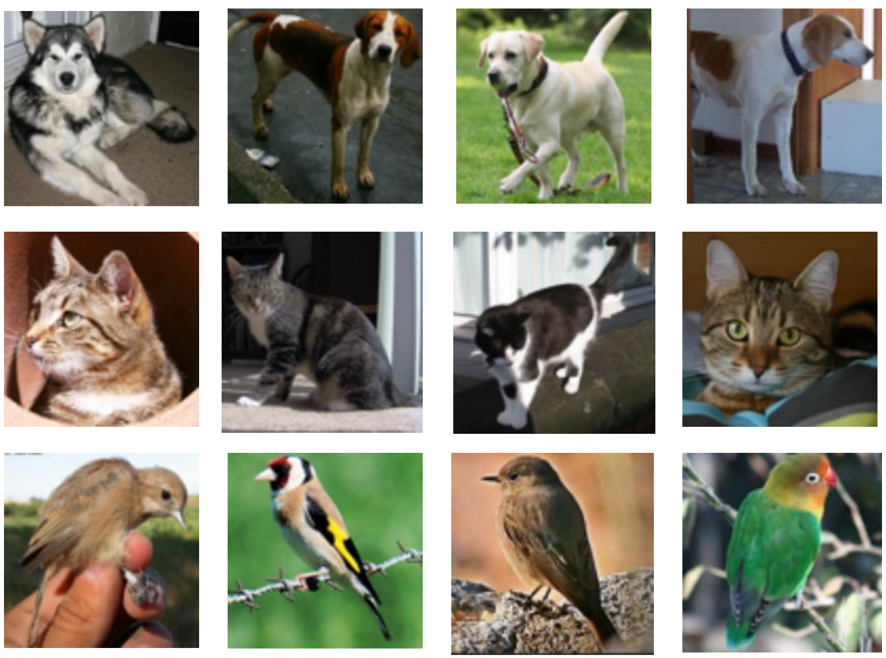
1. Introduction

Data Mining
Quality Analytics

Introduction

군집화(Clustering) 소개

❖ 군집화 (Clustering): Label을 사용하지 않고 데이터가 가지고 있는 특성만을 활용해 유사한 특성을 가진 개체끼리 모으는 것 (군집화 하는 것)을 의미함

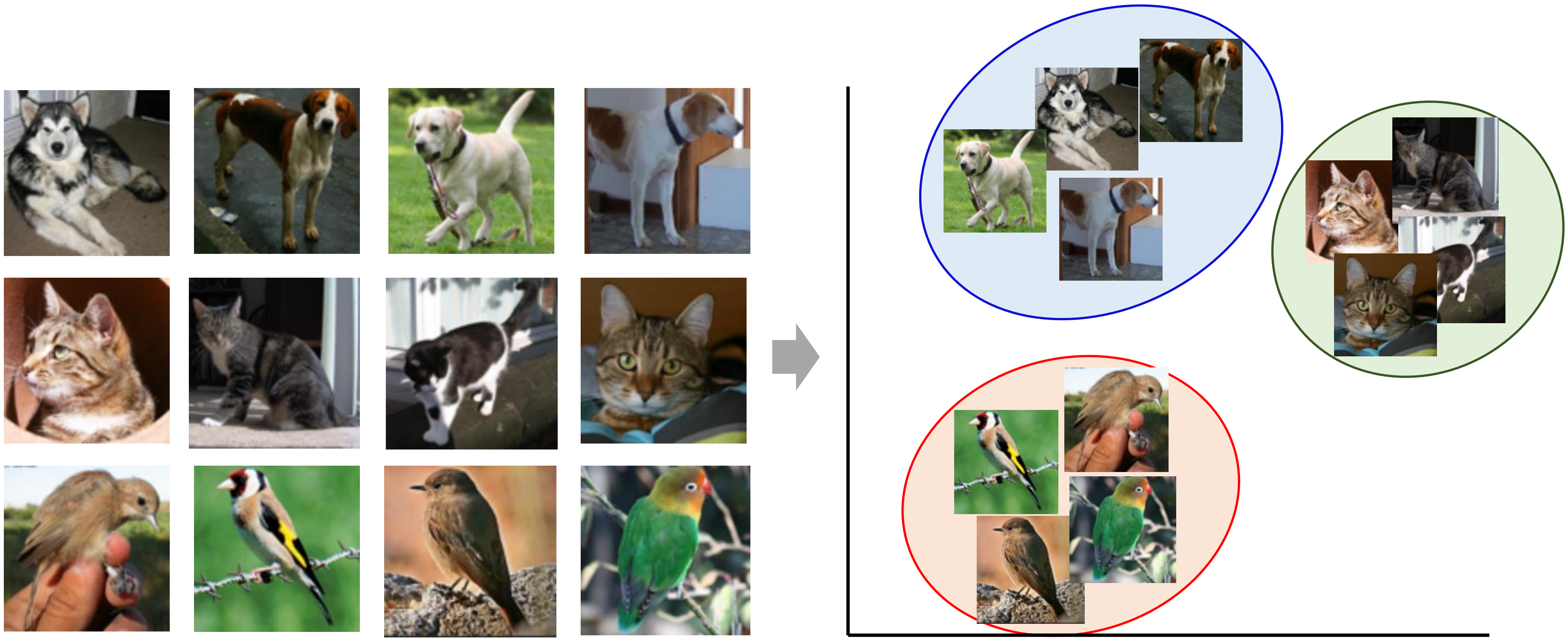


Copyright ⓒ 2023, All rights reserved.

Introduction

군집화(Clusterin g) 소개

❖ 군집화(Clustering): Label을 사용하지 않고 데이터가 가지고 있는 특성만을 활용해 유사한 특성을 가진 개체끼리 모으는 것(군집화 하는 것)을 의미함

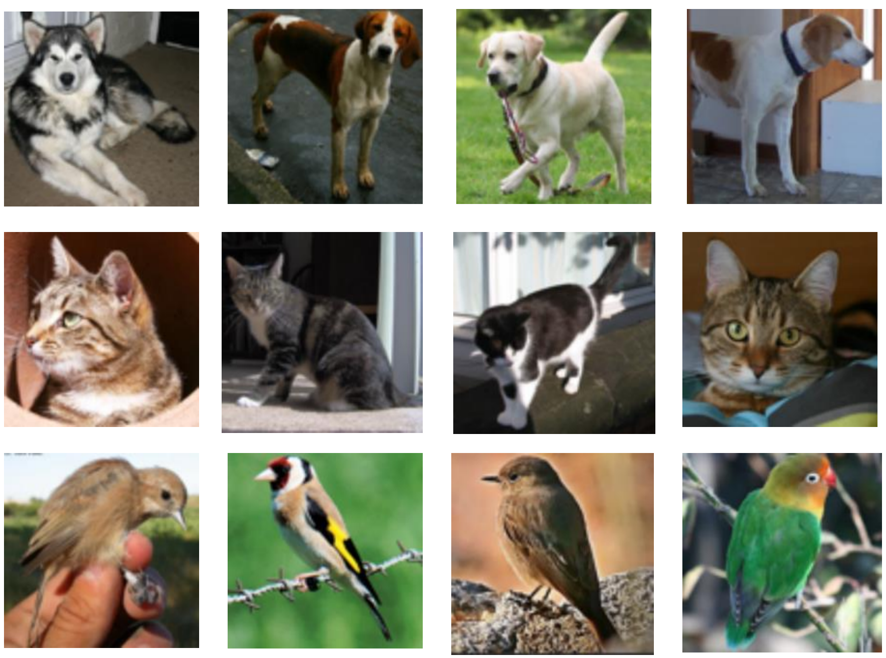


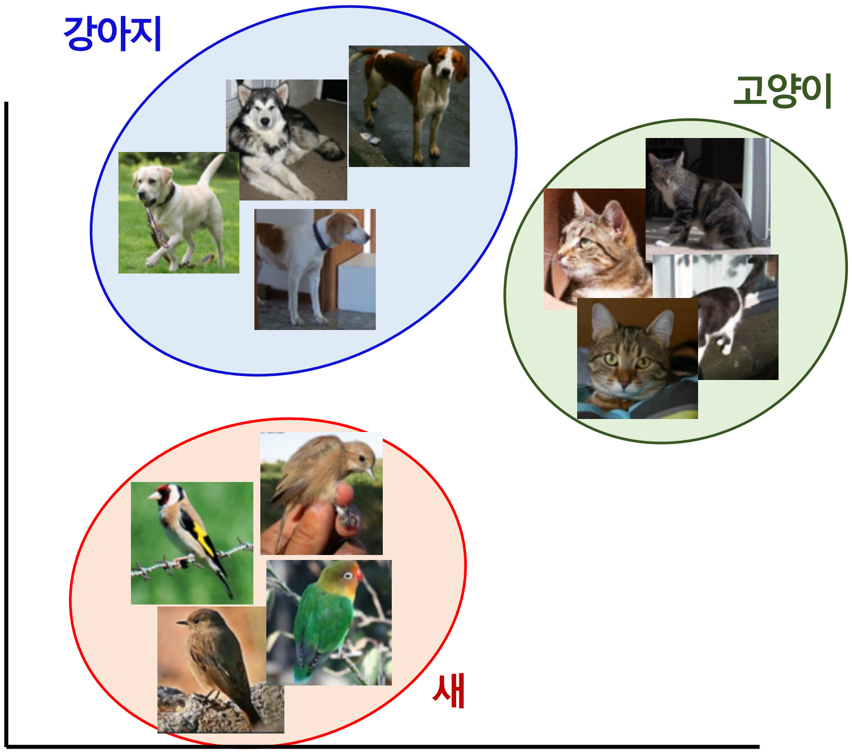
Copyright ⓒ 2023, All rights reserved.

Introduction

군집화(Clustering) 소개

❖ 군집화 (Clustering): Label을 사용하지 않고 데이터가 가지고 있는 특성만을 활용해 유사한 특성을 가진 개체끼리 모으는 것 (군집화 하는 것)을 의미함





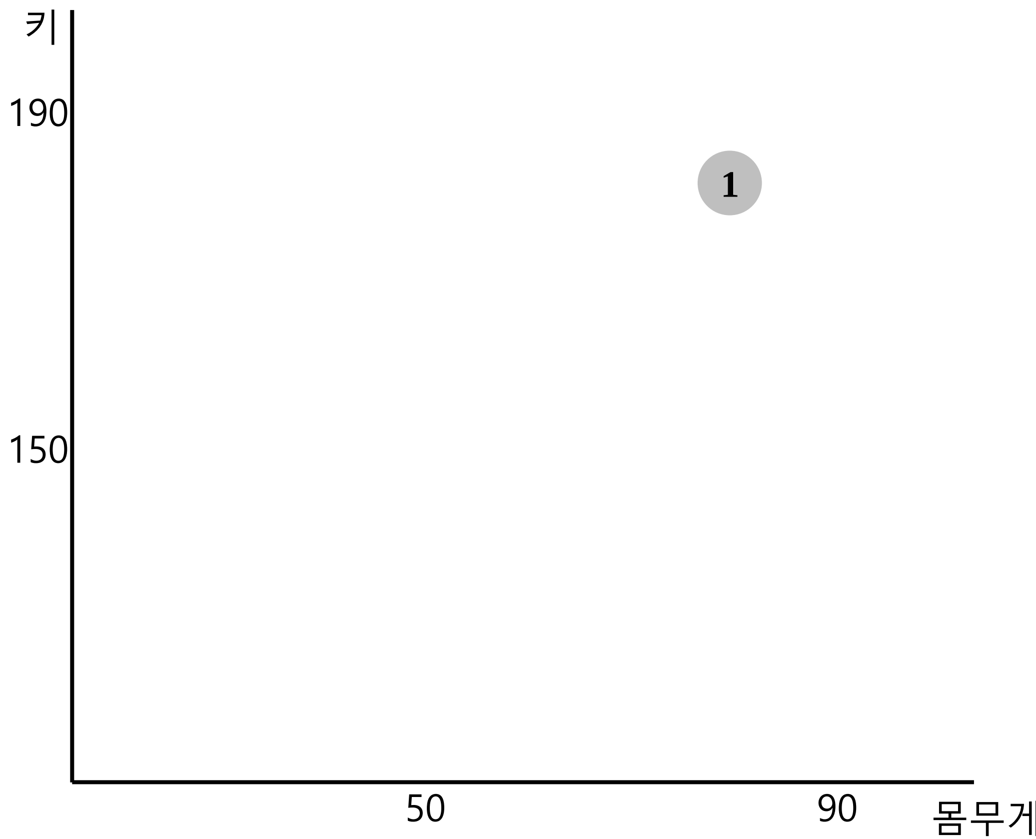
Copyright ⓒ 2023, All rights reserved.

Introduction

일반적인 군집화(Clustering) 과정

- ❖ 11명의 키와 몸무게 데이터를 가지고 군집화를 진행해보고자 함
- ❖ 키와 몸무게 값을 우측에 존재하는 좌표평면에 표시함

No	키	몸무게
1	180	80
2	175	78
3	150	45
4	160	51
5	158	48
6	188	85
7	177	72
8	163	55
9	165	52
10	179	81
11	151	50



Copyright ⓒ 2023, All rights reserved.

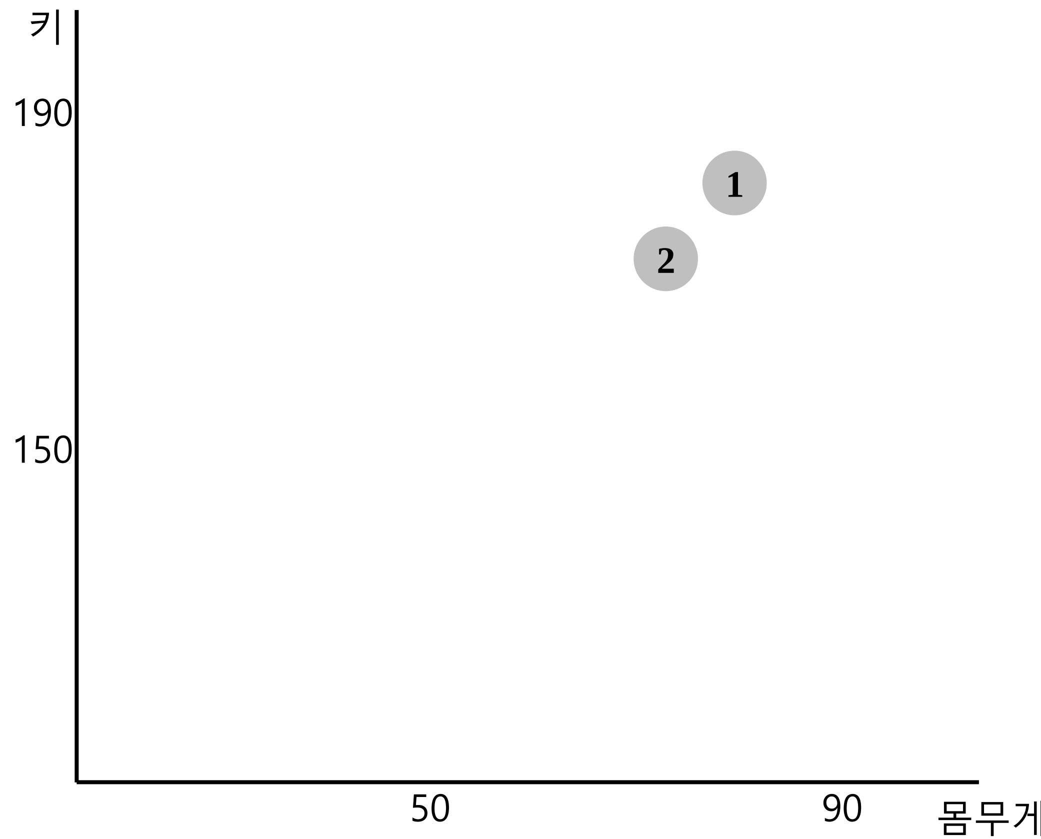
Data Mining Quality Analytics

Introduction

일반적인 군집화(Clustering) 과정

- ❖ 11명의 키와 몸무게 데이터를 가지고 군집화를 진행해보고자 함
- ❖ 1번 사람부터 11번까지 순차적으로 표시를 진행

No	키	몸무게
1	180	80
2	175	78
3	150	45
4	160	51
5	158	48
6	188	85
7	177	72
8	163	55
9	165	52
10	179	81
11	151	50



Copyright ⓒ 2023, All rights reserved.

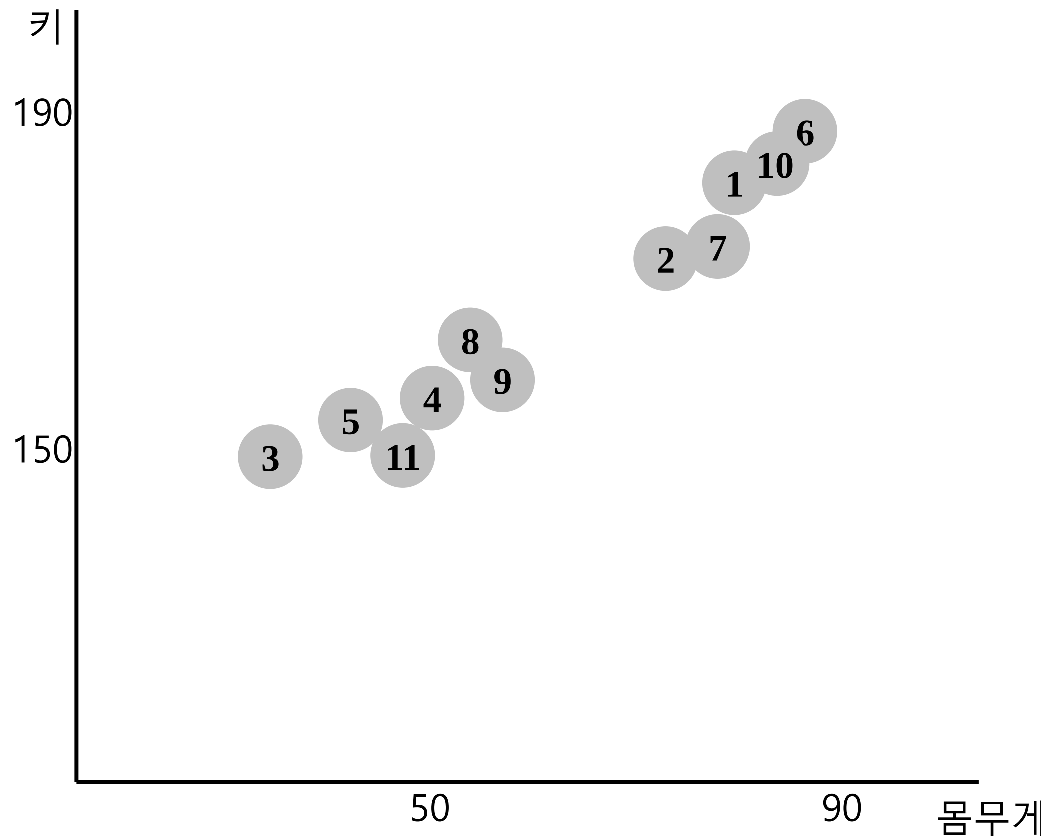
Data Mining Quality Analytics

Introduction

일반적인 군집화(Clustering) 과정

❖ 모든 사람 정보를 좌표 평면에 그리게 되면 우리는 육안으로 군집 2개를 형성할 수 있음

No	키	몸무게
1	180	80
2	175	78
3	150	45
4	160	51
5	158	48
6	188	85
7	177	72
8	163	55
9	165	52
10	179	81
11	151	50



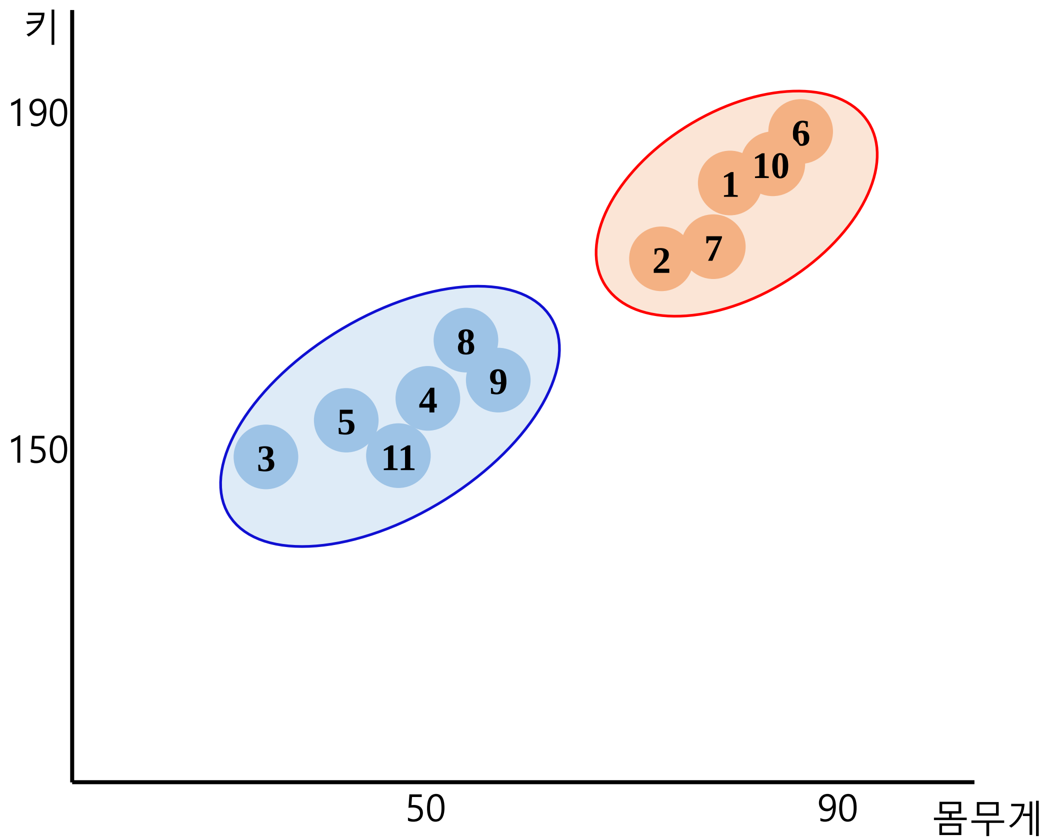
Copyright ⓒ 2023, All rights reserved.

Introduction

일반적인 군집화(Clustering) 과정

- ❖ 빨간색 군집 = 키가 175이상이면서 몸무게가 72kg 이상인 그룹
- ❖ 파랑색 군집 = 키가 165이하이면서 몸무게가 55kg 이하인 그룹

No	키	몸무게
1	180	80
2	175	78
3	150	45
4	160	51
5	158	48
6	188	85
7	177	72
8	163	55
9	165	52
10	179	81
11	151	50



Reference: https://www.datamaker.io/posts/17/

Copyright ⓒ 2023, All rights reserved.

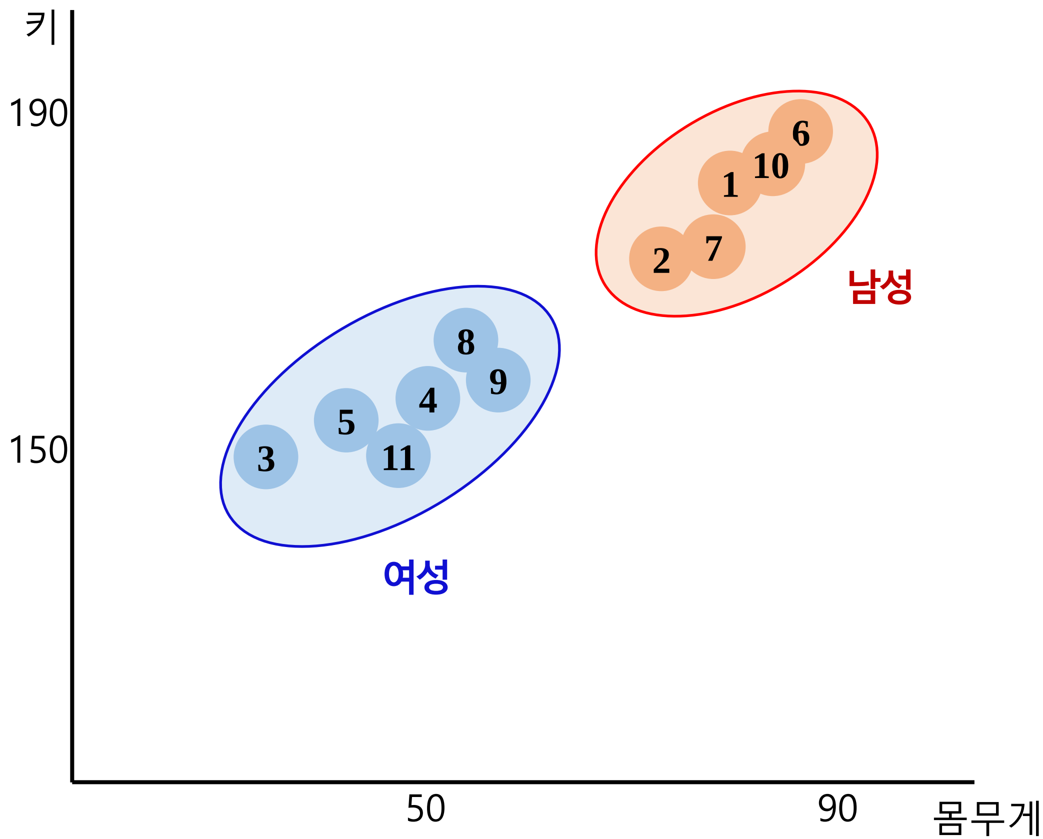
Data Mining Quality Analytics

Introduction

일반적인 군집화(Clustering) 과정

- ❖ 키와 몸무게 2개 특성만 가지고 군집을 나눴을 때 2그룹으로 나뉘게 됨
- ❖ 추가 수집한 데이터에서 성별을 확인한 결과 남성, 여성 그룹으로 나누어짐을 확인

No	키	몸무게	성별
1	180	80	남
2	175	78	남
3	150	45	여
4	160	51	여
5	158	48	여
6	188	85	남
7	177	72	남
8	163	55	여
9	165	52	여
10	179	81	남
11	151	50	여



Reference: https://www.youtube.com/watch?v=8zB-_LrAraw&list=PLpIPLT0Pf7IoTxTCi2MEQ94MZnHaxrP0j&index=14 (군집분석, 김성범 교수님)

Copyright ⓒ 2023, All rights reserved.

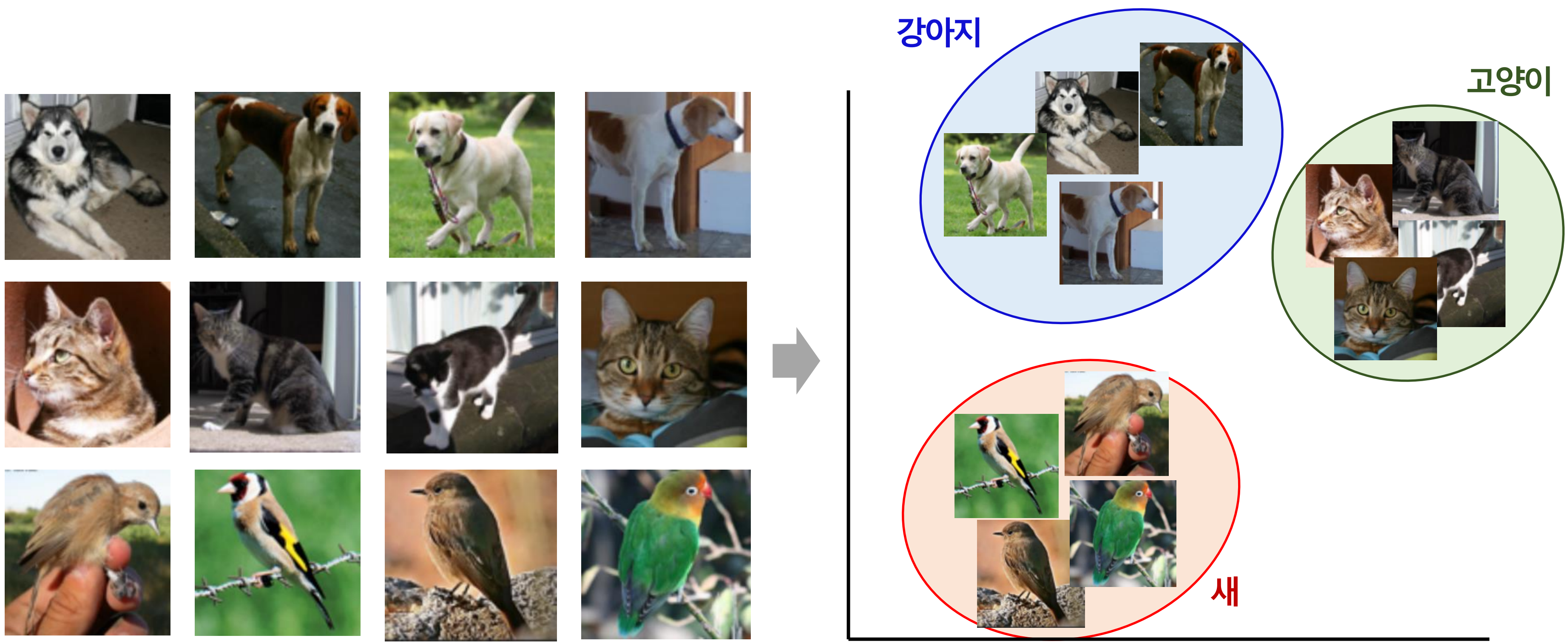
Data Mining Quality Analytics

Introduction

고차원 데이터에서 군집화(Clustering) 어려움

❖ 하지만 X 데이터가 2개인 단순한 데이터와 달리 고차원 데이터는 군집을 쉽게 하기 어려움

Q. 이미지와 같은 고차원 데이터에서는 어떻게 해야 군집을 잘할 수 있을까?



Copyright ⓒ 2023, All rights reserved.

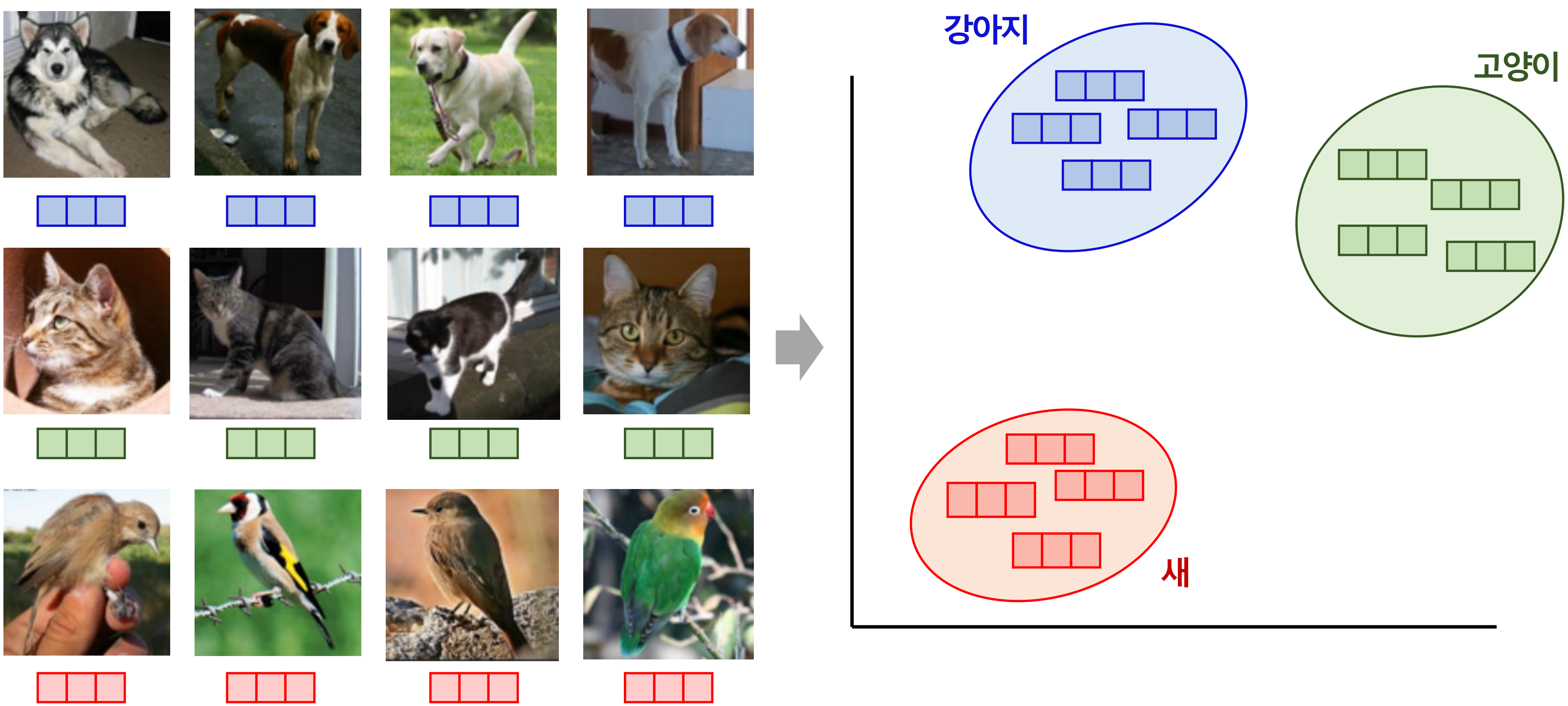
Introduction

고차원 데이터에서 군집화(Clustering) 어려움

❖ 하지만 X 데이터가 2개인 단순한 데이터와 달리 고차원 데이터는 군집을 쉽게 하기 어려움

Q. 이미지와 같은 고차원 데이터에서는 어떻게 해야 군집을 잘할 수 있을까?

A. 이미지 특징을 잘 추출하는 vector를 만들어 군집을 진행하자!



Copyright ⓒ 2023, All rights reserved.

2. Deep Clustering

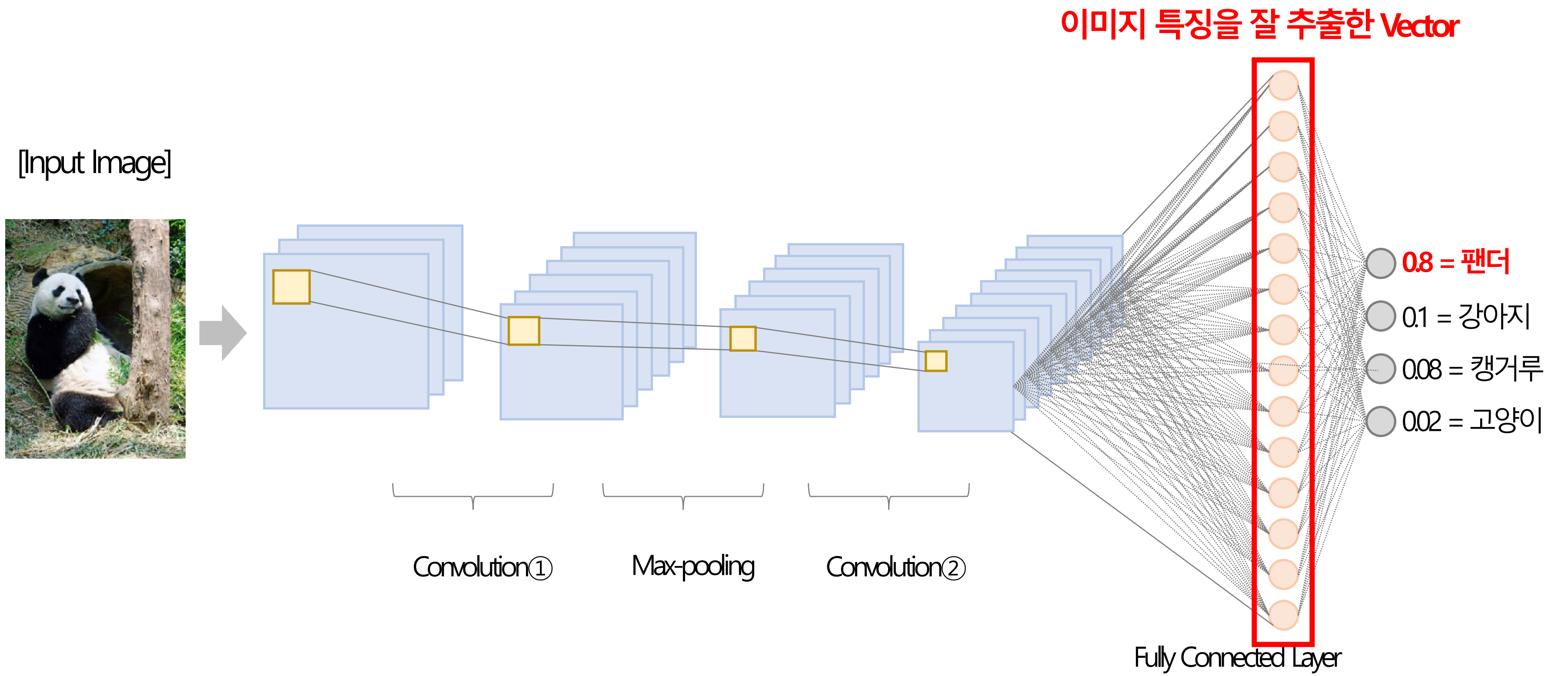
Data Mining
Quality Analytics

Deep Clustering

Deep Learning

- ❖ 일반적인 딥러닝 모델은 다양한 class 분류, 수치 예측 등 다양한 task를 풀 수 있음
- ❖ 이 때, 이미지 같이 고차원 데이터의 특징을 잘 추출해야만 다양한 task를 정확하게 풀 수 있음

  → 결국, 군집화를 위해 딥러닝 모델을 활용해 고차원 이미지의 특징을 잘 추출해보고자 함



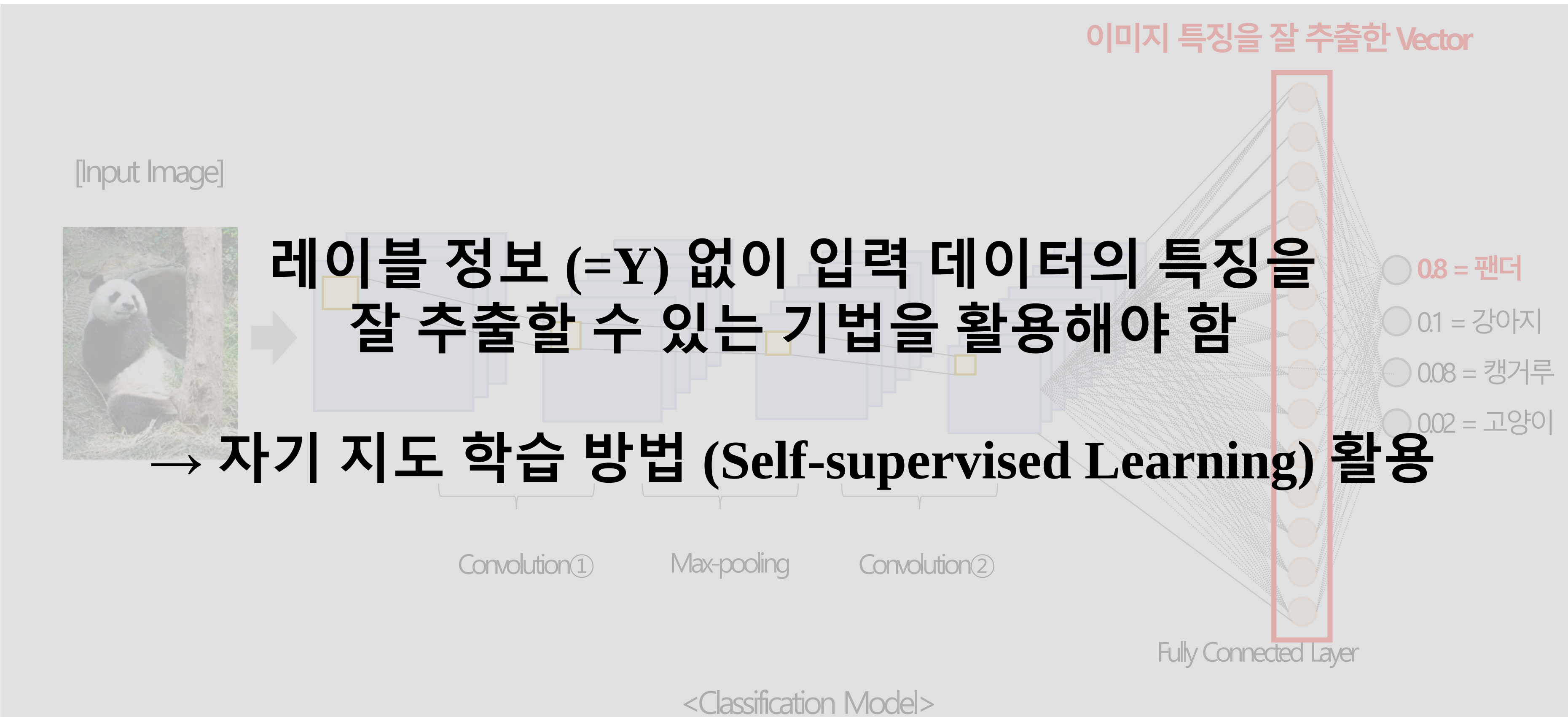
<Classification Model>

Copyright ⓒ 2023, All rights reserved.

Deep Clustering

Deep Learning

- ❖ 하지만 일반적인 딥러닝 모델은 이미지에 대한 정답(=y)인 레이블 정보가 필요함
- ❖ 이미지의 특징만을 활용하는 군집화 기법에서는 정답 정보 없이도 데이터 특징을 잘 추출할 수 있는 기법들을 활용해야 함



레이블 정보 (=Y) 없이 입력 데이터의 특징을 잘 추출할 수 있는 기법을 활용해야 함

→ 자기 지도 학습 방법 (Self-supervised Learning) 활용

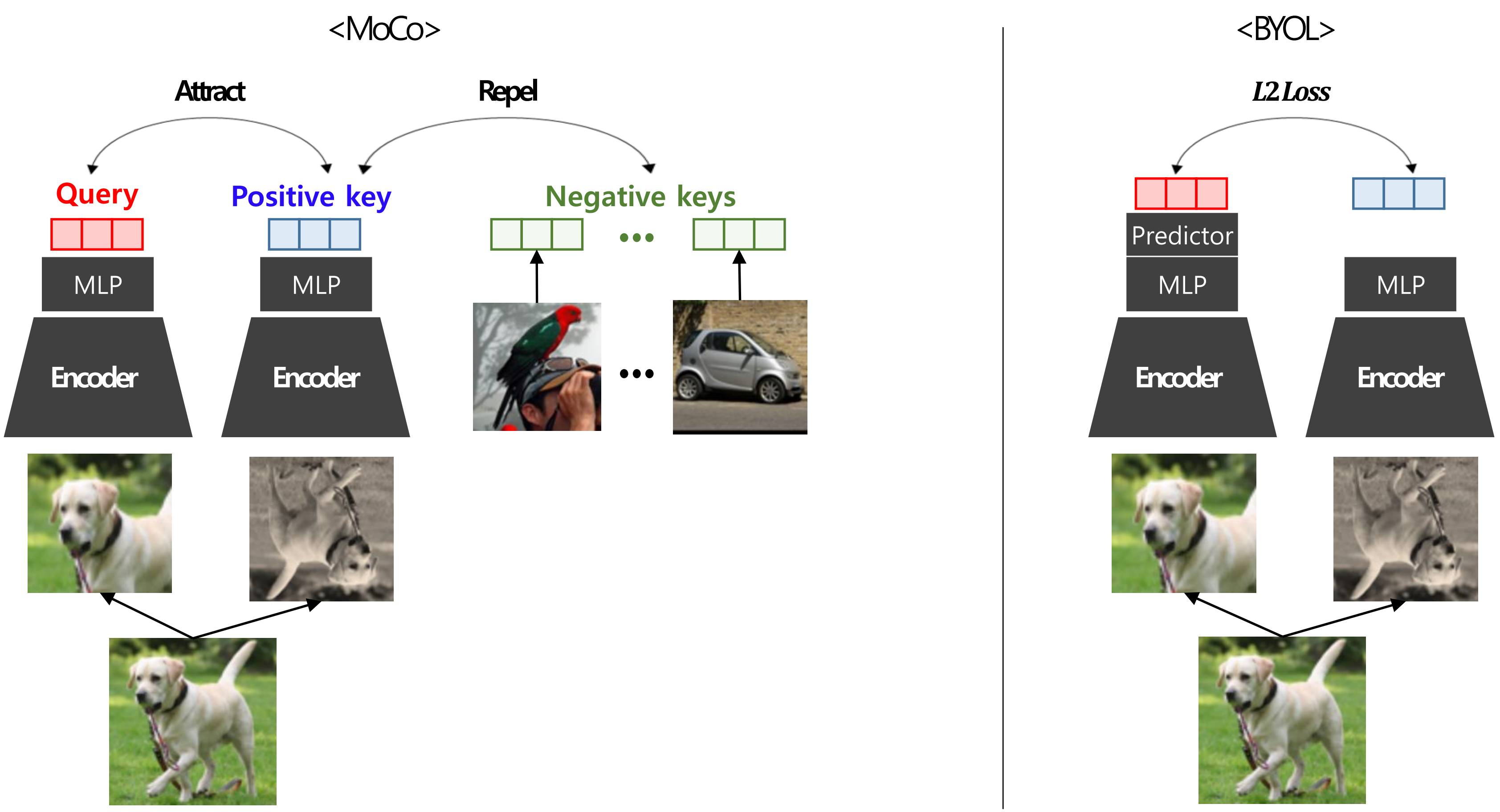
Copyright ⓒ 2023, All rights reserved.

Data Mining Quality Analytics

Deep Clustering

Self-supervised Learning

❖ 자기 지도 학습 (Self-supervised learning)은 레이블 (label)이 없는 데이터에 대해서 데이터가 지니고 있는 특성 자체를 잘 학습해 좋은 representation을 얻게 해주는 방법론임



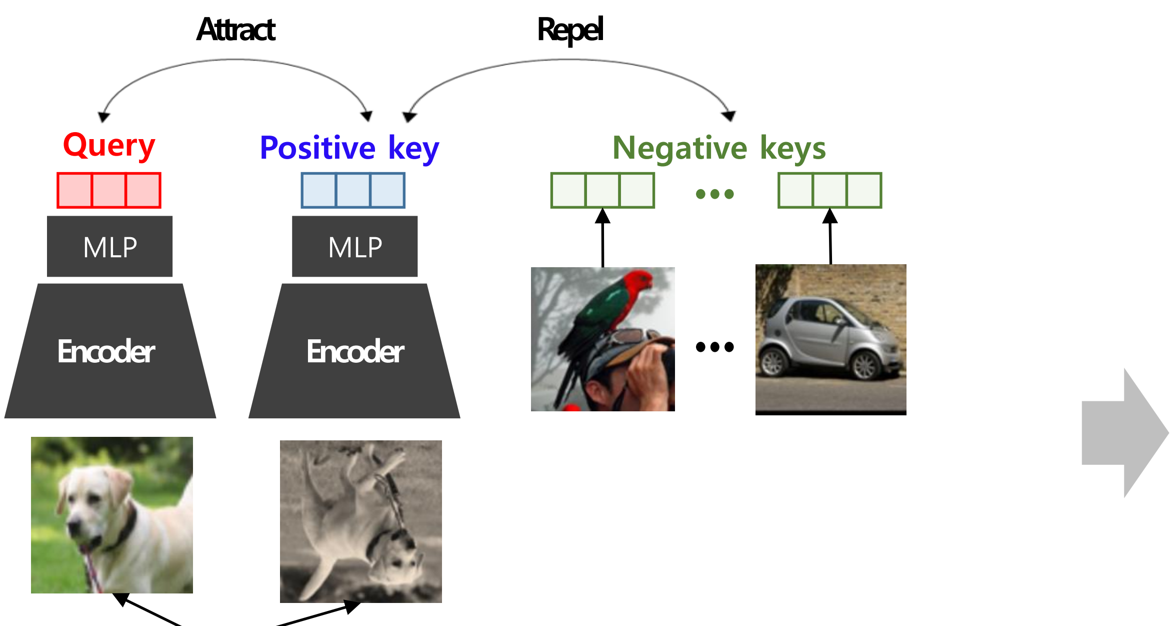
Copyright ⓒ 2023, All rights reserved.

Deep Clustering

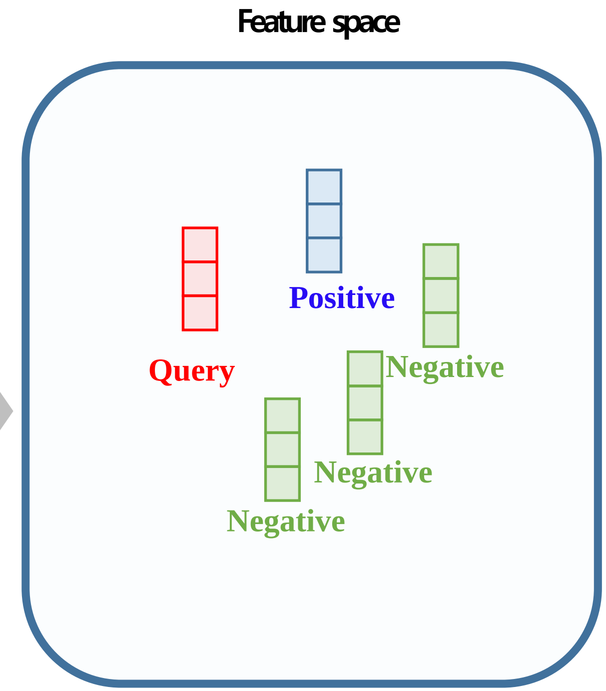
Self-supervised Learning

❖ MoCo (Momentum contrast for unsupervised visual representation learning)는 negative sample을 활용한 contrastive learning으로 데이터 내 특징을 잘 추출할 수 있는 특징 공간을 만들 수 있음

<MoCo>



$$\text{MoCo Loss} = -\log\frac{\exp\left(f(x)^T f(x^+)\right)}{\exp\left(f(x)^T f(x^+)\right) + \sum_{j=1}^{N-1}\exp\left(f(x)^T f(x_j)\right)}$$



Query와 **Positive key**는 가까워지도록 학습
Query와 **Negative keys**는 멀어지도록 학습

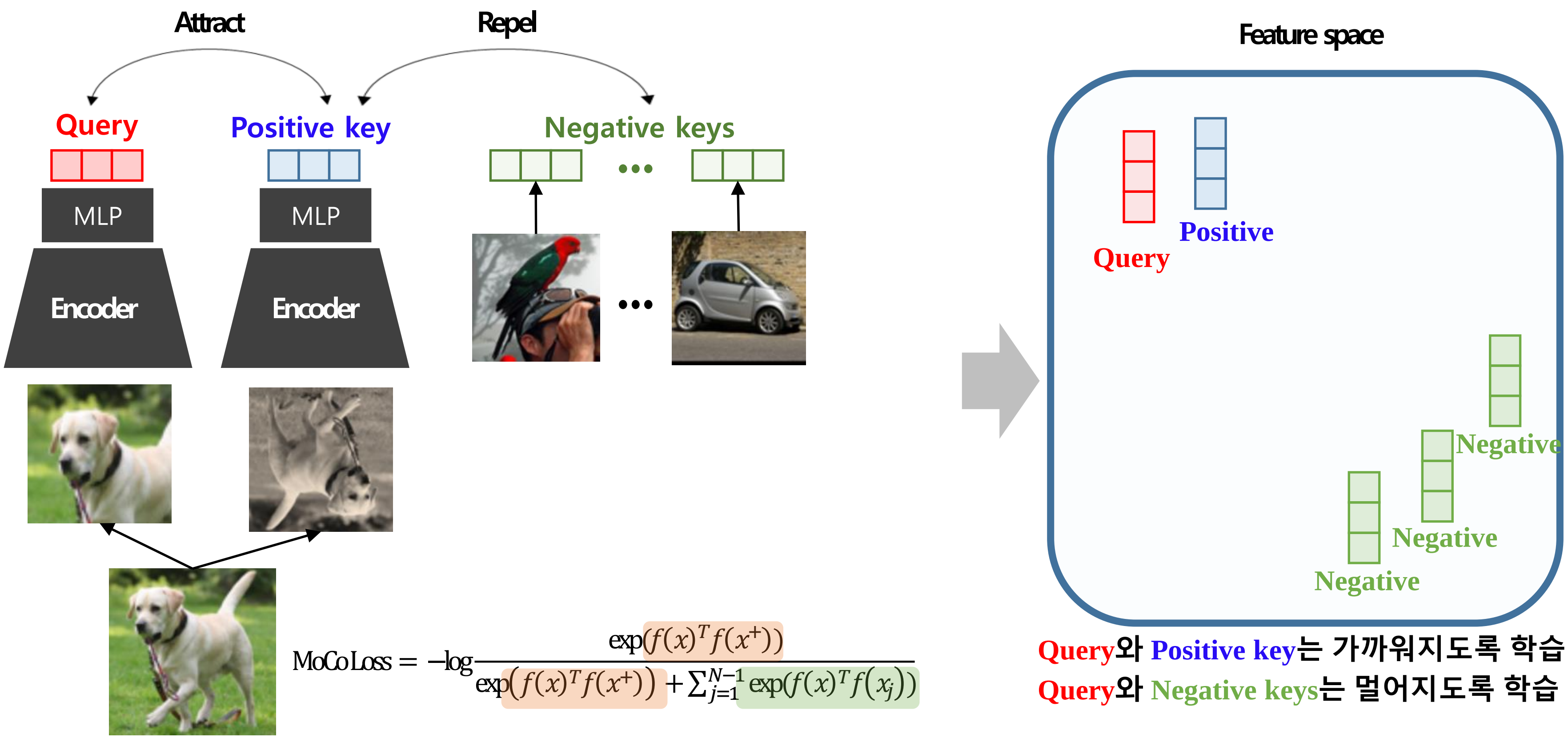
Copyright ⓒ 2023, All rights reserved.
Data Mining Quality Analytics

Deep Clustering

Self-supervised Learning

❖ 원본 이미지에서 각각 augmentation해 추출된 query와 positive key는 가까워지게 학습 진행하고 다른 이미지인 negative keys는 멀어지도록 학습을 진행함 (Contrastive Learning)

<MoCo>



$$\text{MoCo Loss} = -\log \frac{\exp\left(f(x)^T f(x^+)\right)}{\exp\left(f(x)^T f(x^+)\right) + \sum_{j=1}^{N-1} \exp\left(f(x)^T f(x_j)\right)}$$

Query와 Positive key는 가까워지도록 학습
Query와 Negative keys는 멀어지도록 학습

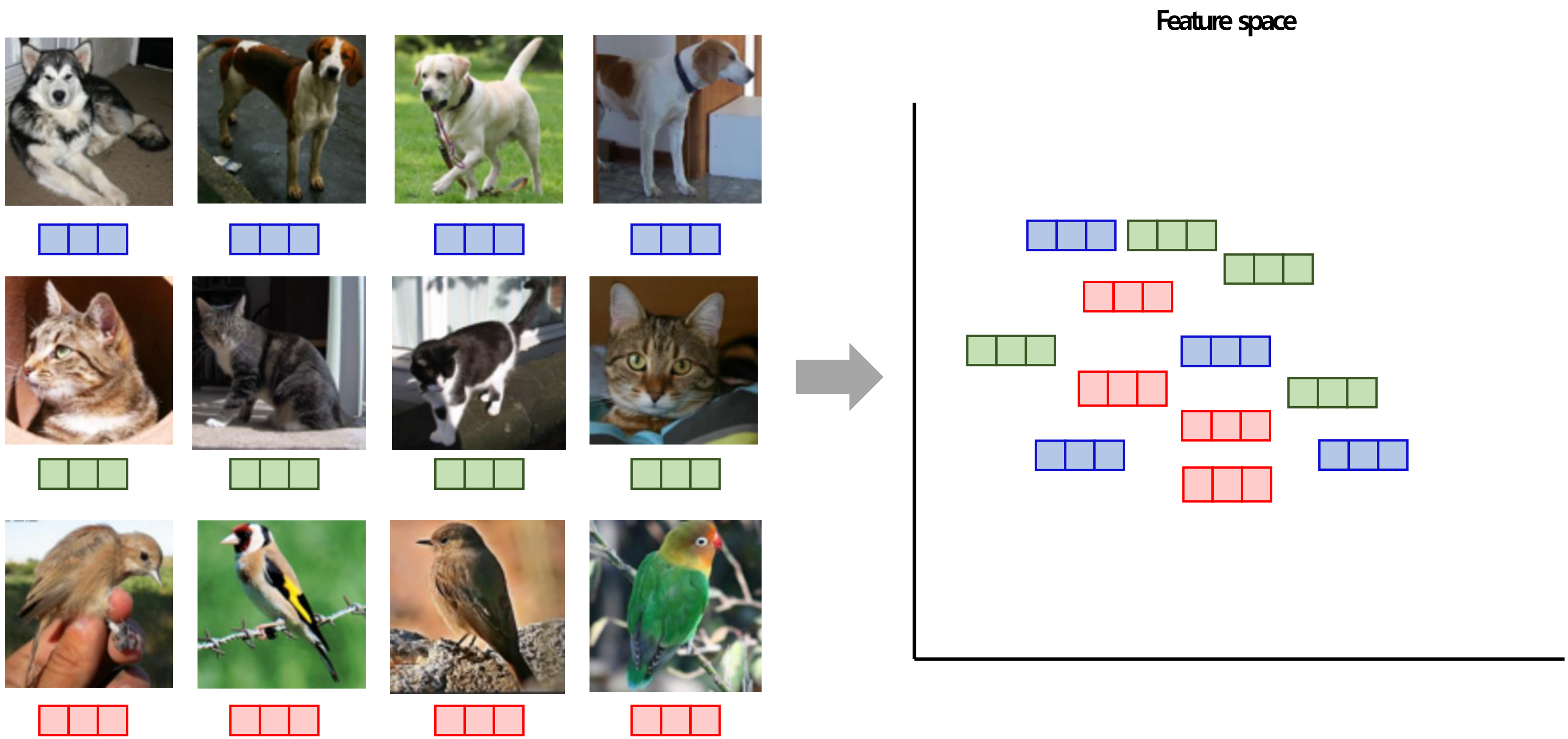
Copyright ⓒ 2023, All rights reserved.

Deep Clustering

Deep clustering

- ❖ 만약 MoCo를 학습하지 않고 clustering을 진행한다면?
- ❖ 먼저 학습을 진행하기 전에는 특징 공간 (feature space)에서 군집이 잘 형성되지 않을 것임

<MoCo로 학습을 진행하기 전>



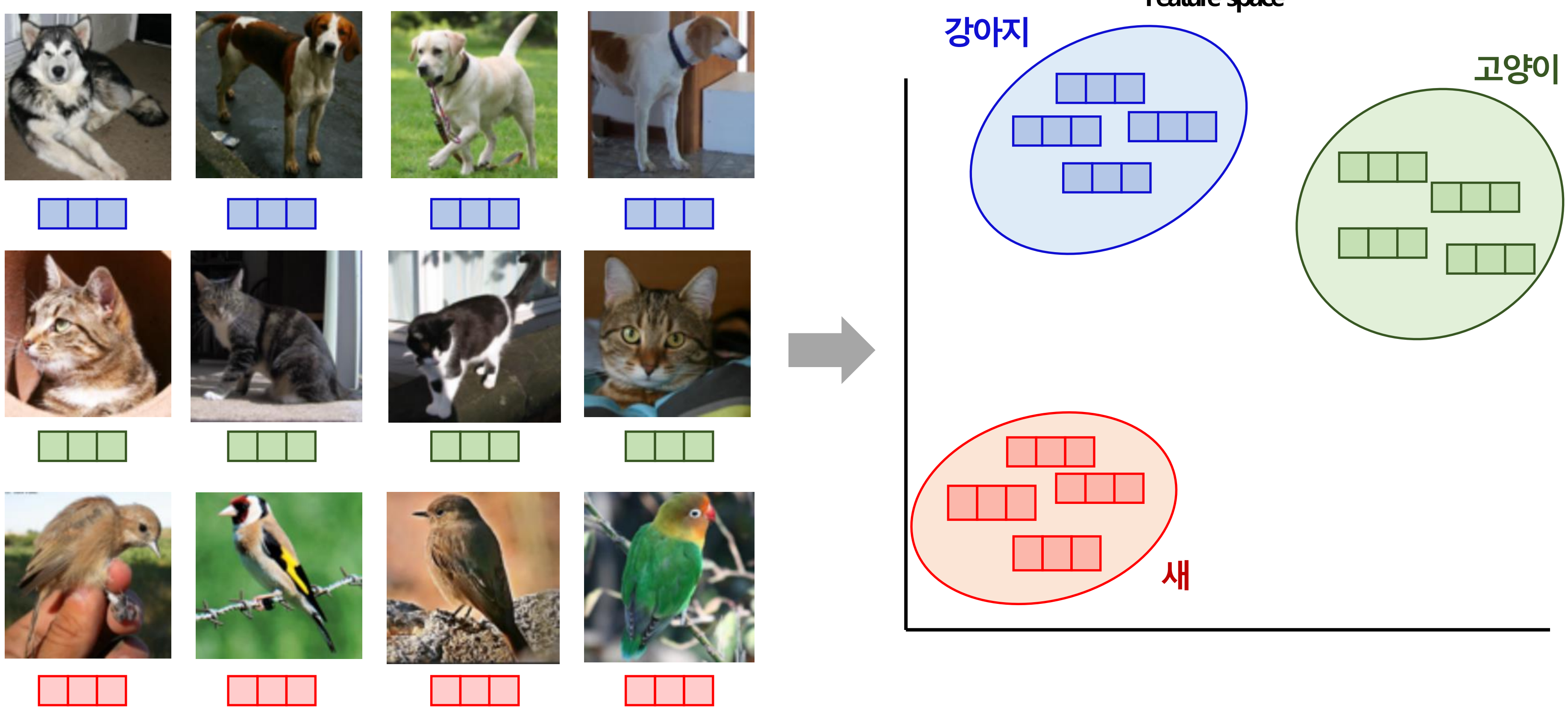
Copyright ⓒ 2023, All rights reserved.

Deep Clustering

Deep clustering

- ❖ 하지만 MoCo로 학습을 진행한 후 clustering을 진행한다면?
- ❖ 유사한 특징을 지닌 데이터끼리 잘 모여서 각각의 cluster를 잘 구분할 수 있을 것임

<MoCo로 학습을 진행한 후>



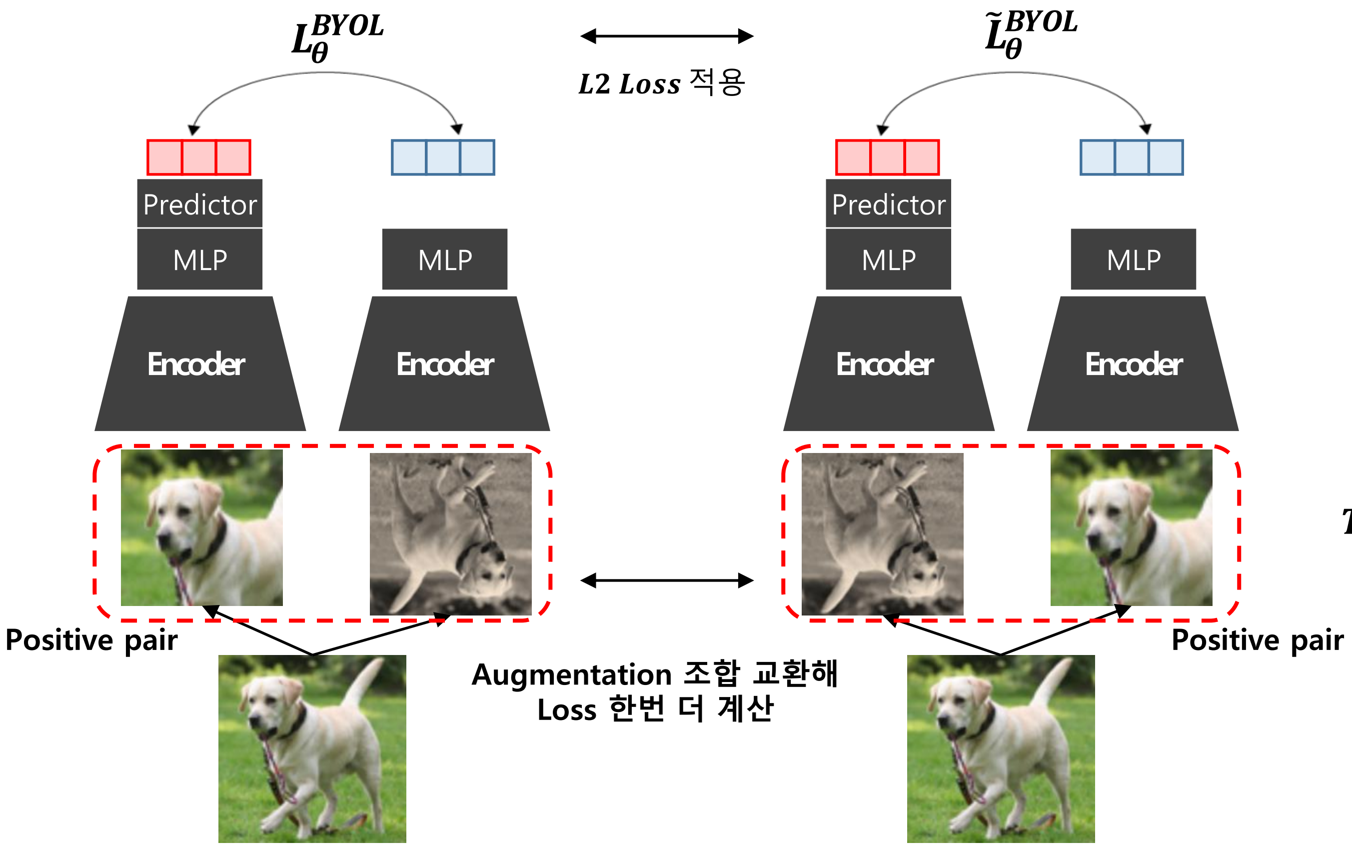
Copyright ⓒ 2023, All rights reserved.

Deep Clustering

Deep clustering

❖ BYOL (Bootstrap Your Own Latent)은 negative sample을 활용하지 않고 positive pair만 활용해 데이터 내 특징을 잘 추출할 수 있도록 학습하는 방법론임 (Non-contrastive Learning)

<BYOL>



$$Total\ Loss = L_\theta^{BYOL} + \tilde{L}_\theta^{BYOL}$$

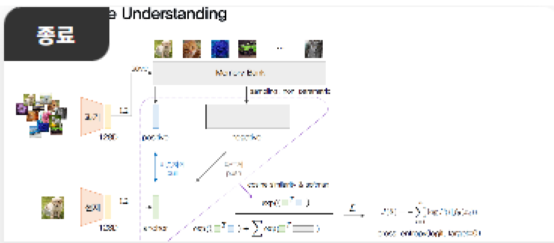
Copyright ⓒ 2023, All rights reserved.

기존 연구실 세미나

기존 연구실 세미나 (Self-supervised Learning)

❖ 본 세미나는 Self-supervised Learning 방법론을 clustering에 활용한 연구 중심으로 설명 진행

❖ Self-supervised Learning 방법론에 대해 자세하게 알고 싶으신 분은 아래 세미나 참고



Towards Contrastive Learning

발표자: 곽민구

2021년 1월 29일

오후 1시 ~

온라인 비디오 시청 (YouTube)

세미나 정보 보기 ⟶

Dive into BYOL

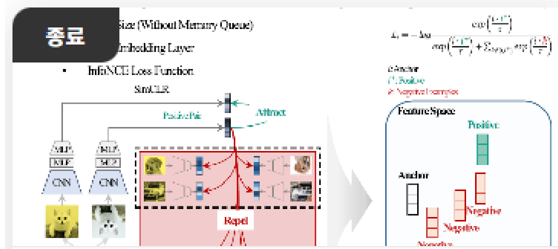
발표자: 김재훈

2021년 2월 19일

오후 1시 ~

온라인 비디오 시청 (YouTube)

세미나 정보 보기 ⟶



Applications of Self-Supervised Learning

발표자: 이영재

2022년 3월 4일

오전 3시 ~

온라인 비디오 시청(YouTube)

온라인 비디오 시청 (YouTube)

세미나 정보 보기 ⟶

Reference: http://dmqa.korea.ac.kr/activity/seminar/308
http://dmqa.korea.ac.kr/activity/seminar/310
http://dmqa.korea.ac.kr/activity/seminar/355

Copyright ⓒ 2023, All rights reserved.

2. Contrastive Clustering (CC)

Data Mining Quality Analytics

Contrastive Clustering (CC)

CC 논문

❖ Contrastive Clustering (CC) (2021)

- 데이터 간 contrastive learning과 군집 간 contrastive learning을 진행해 clustering하는 연구
- 2021년도 AAAI Conference on Artificial Intelligence에서 발표됨
- 23년 2월 16일 기준 218회 인용

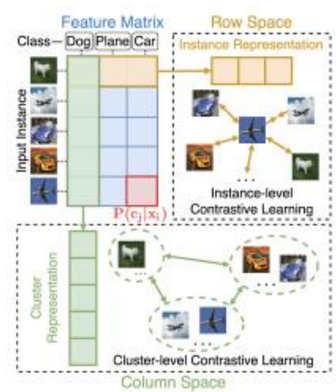
Contrastive Clustering

Yunfan Li[1], Peng Hu[1], Zitao Liu[2], Dezhong Peng[1,4,5], Joey Tianyi Zhou[3], Xi Peng[1*]

[1]College of Computer Science, Sichuan University, China
[2]TAL Education Group, China
[3]Institute of High Performance Computing, A*STAR, Singapore
[4]Shenzhen Peng Cheng Laboratory, China
[5]College of Computer & Information Science, Southwest University, China
yunfanli.gm@gmail.com, penghu.ml@gmail.com, zitao.jerry.liu@gmail.com, pengdz@scu.edu.cn,
joey.tianyi.zhou@gmail.com, pengx.gm@gmail.com

Abstract

In this paper, we propose an online clustering method called Contrastive Clustering (CC) which explicitly performs the instance- and cluster-level contrastive learning. To be specific, for a given dataset, the positive and negative instance pairs are constructed through data augmentations and then projected into a feature space. Therein, the instance- and cluster-level contrastive learning are respectively conducted in the row and column space by maximizing the similarities of positive pairs while minimizing those of negative ones. Our key observation is that the rows of the feature matrix could be regarded as soft labels of instances, and accordingly the columns could be further regarded as cluster representations. By simultaneously optimizing the instance- and cluster-level contrastive loss, the model jointly learns representations and cluster assignments in an end-to-end manner. Besides, the proposed method could timely compute the cluster assign-



Reference: Li, Y., Hu, P., Liu, Z., Peng, D., Zhou, J. T., & Peng, X. (2021, May). Contrastive clustering. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 35, No. 10, pp. 8547-8555).

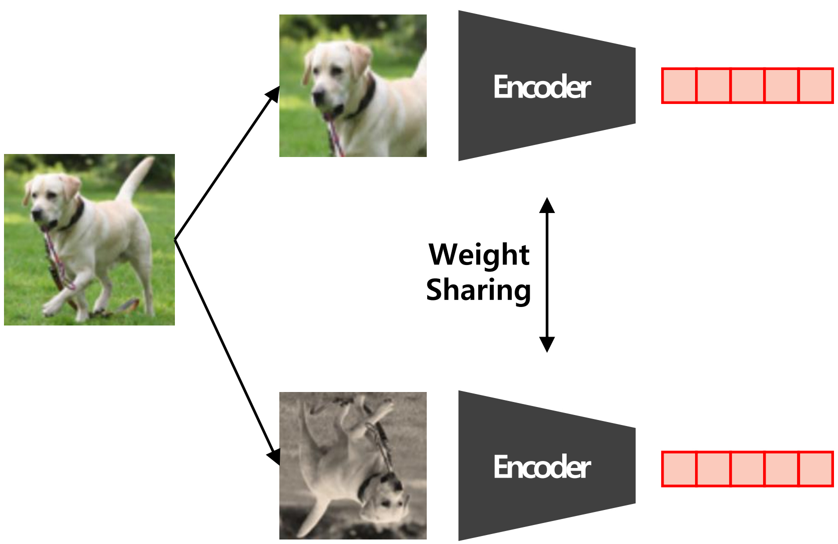
Copyright ⓒ 2023, All rights reserved.

Data Mining Quality Analytics

Contrastive Clustering (CC)

Pari Construction Backbone

❖ 앞서 보았던 self-supervised learning 방식과 같이 동일 샘플에서 각각 다른 augmentation을 취한 이미지를 생성해 각각의 특징을 요약하는 vector를 생성하는 구조로 모델을 구축



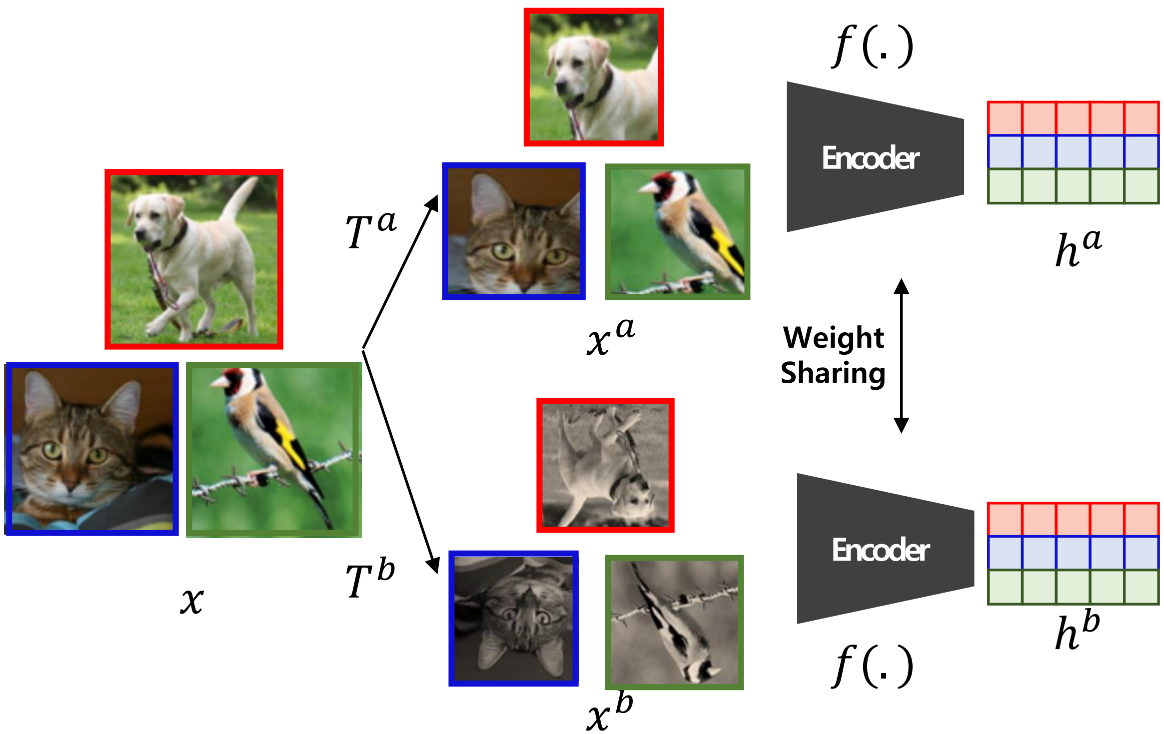
Reference: Li, Y., Hu, P., Liu, Z., Peng, D., Zhou, J. T., & Peng, X. (2021, May). Contrastive clustering. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 35, No. 10, pp. 8547-8555).

Copyright ⓒ 2023, All rights reserved.

Contrastive Clustering (CC)

Pair Construction Backbone

- ❖ 실제 학습시에는 이미지가 batch 사이즈만큼 한번에 들어가기 때문에 아래와 같이 확장함
- ❖ Batch size = 3으로 둔 예시임



Reference: Li, Y., Hu, P., Liu, Z., Peng, D., Zhou, J. T., & Peng, X. (2021, May). Contrastive clustering. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 35, No. 10, pp. 8547-8555).

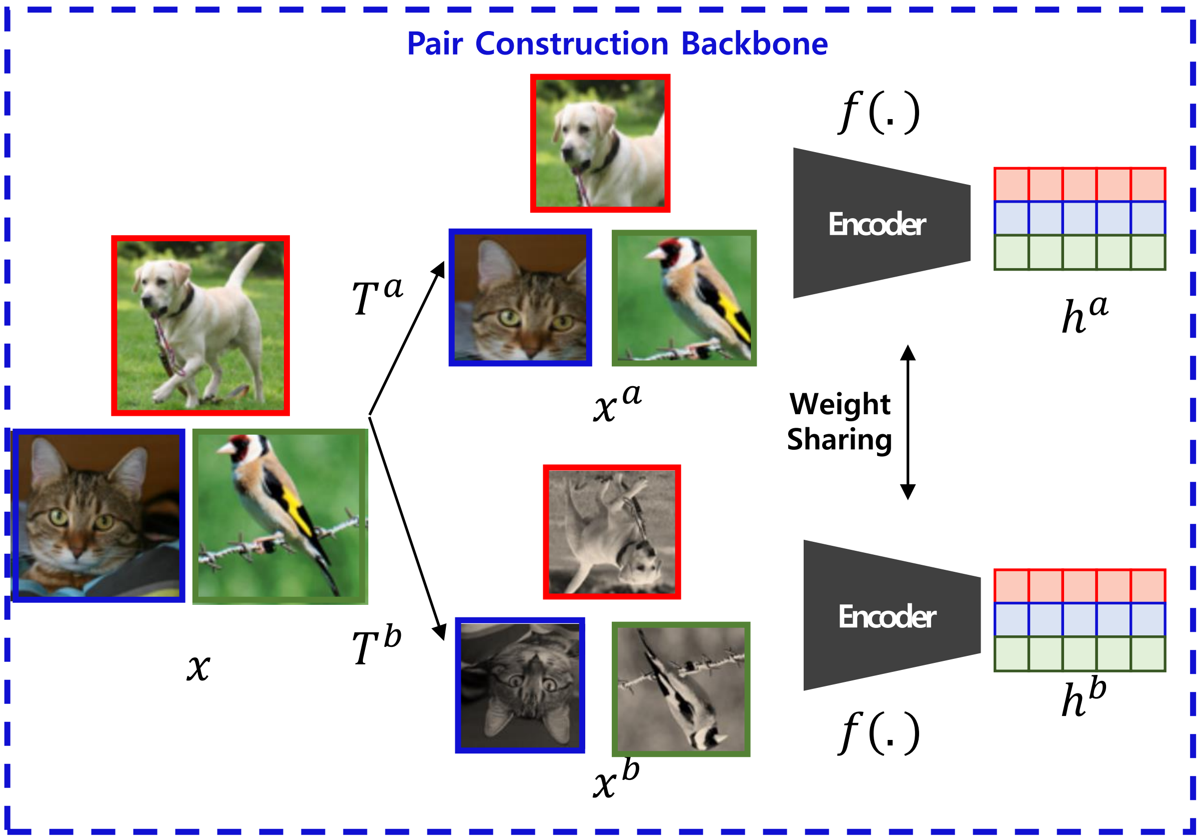
Copyright ⓒ 2023, All rights reserved.

Data Mining Quality Analytics

Contrastive Clustering (CC)

Pair Construction Backbone

❖ Batch size만큼 특징 벡터를 요약할 수 있는 부분을 Pair Construction Backbone이라 지칭함



Reference: Li, Y., Hu, P., Liu, Z., Peng, D., Zhou, J. T., & Peng, X. (2021, May). Contrastive clustering. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 35, No. 10, pp. 8547-8555).

Copyright ⓒ 2023, All rights reserved.

Contrastive Clustering (CC)

$$\text{MoCo Loss} = -\log \frac{\exp\left(f(x)^T f(x^+)\right)}{\exp\left(f(x)^T f(x^+)\right) + \sum_{j=1}^{N-1} \exp\left(f(x)^T f(x_j)\right)}$$

N: 샘플 수
τ_I: $hyper\ parameter$

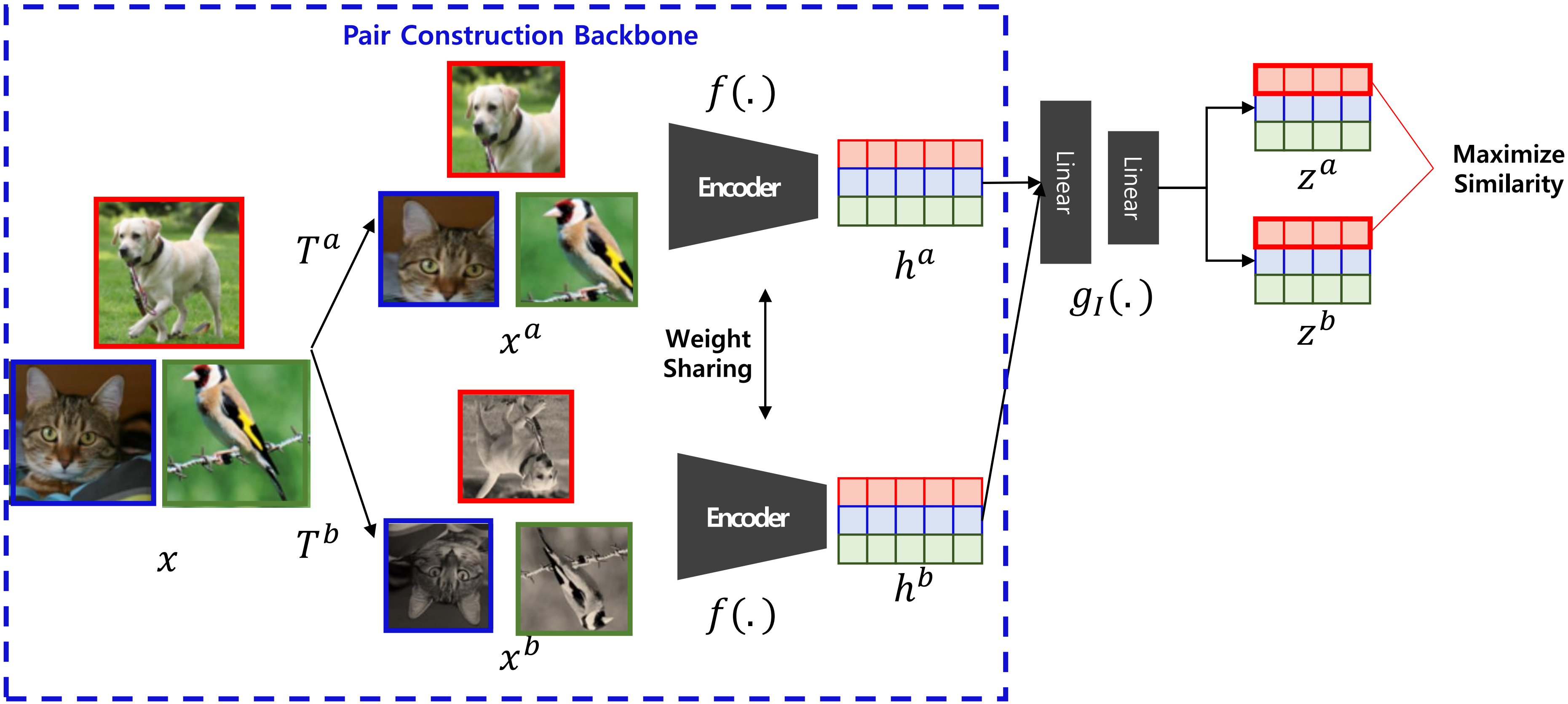
Instance-Level Contrastive Head

같은 샘플을 augmentation한 데이터는 가깝게

$$s(z_i^a, z_j^a) = \frac{(z_i^a)(z_i^b)^T}{\|z_i^a\| \|z_i^b\|}, \qquad l_i^a = -\log \frac{\exp\left(\frac{s(z_i^a, z_j^a)}{\tau_I}\right)}{\sum_{j=1}^{N} \left[\exp\left(\frac{s(z_i^a, z_j^a)}{\tau_I}\right) + \exp\left(\frac{s(z_i^a, z_j^b)}{\tau_I}\right)\right]}$$

$$L_{ins} = \frac{1}{2N} \sum_{i=1}^{N} (l_i^a + l_i^b)$$

Batch내 다른 데이터는 멀어지게



Reference: Li, Y., Hu, P., Liu, Z., Peng, D., Zhou, J. T., & Peng, X. (2021, May). Contrastive clustering. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 35, No. 10, pp. 8547-8555).

Copyright ⓒ 2023, All rights reserved.

Contrastive Clustering (CC)

$$\text{MoCo Loss} = -\log \frac{\exp\left(f(x)^T f(x^+)\right)}{\exp\left(f(x)^T f(x^+)\right) + \sum_{j=1}^{N-1} \exp\left(f(x)^T f(x_j)\right)}$$

N: 샘플 수

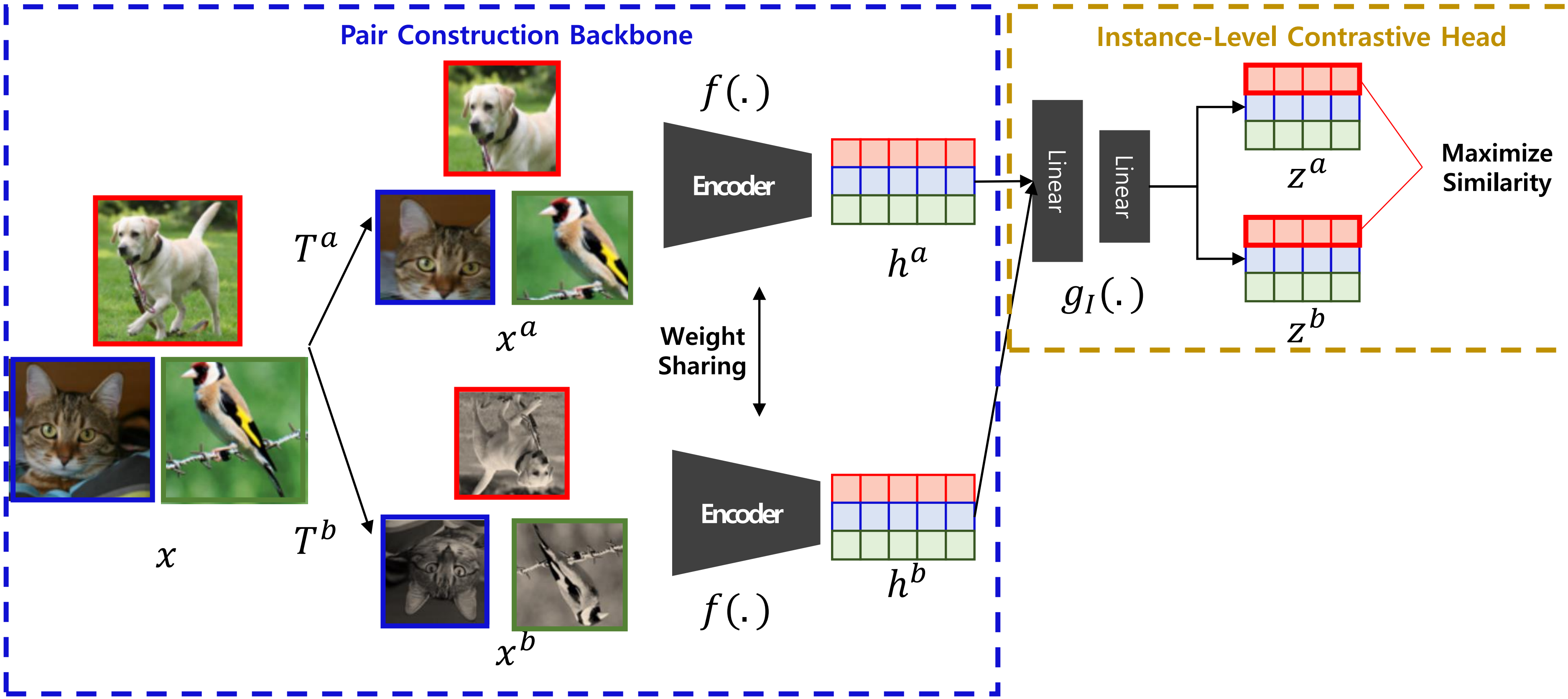
τ_I: $hyper\ parameter$

Instance-Level Contrastive Head

같은 샘플을 augmentation한 데이터는 가깝게

$$s(z_i^a, z_j^a) = \frac{(z_i^a)(z_i^b)^T}{\|z_i^a\|\|z_i^b\|}, \quad l_i^a = -\log \frac{\exp\left(\frac{s(z_i^a, z_j^a)}{\tau_I}\right)}{\sum_{j=1}^{N} \left[\exp\left(\frac{s(z_i^a, z_j^a)}{\tau_I}\right) + \exp\left(\frac{s(z_i^a, z_j^b)}{\tau_I}\right)\right]} \qquad L_{ins} = \frac{1}{2N} \sum_{i=1}^{N} (l_i^a + l_i^b)$$

Batch내 다른 데이터는 멀어지게



Reference: Li, Y., Hu, P., Liu, Z., Peng, D., Zhou, J. T., & Peng, X. (2021, May). Contrastive clustering. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 35, No. 10, pp. 8547-8555).

Copyright ⓒ 2023, All rights reserved.

Contrastive Clustering (CC)

$$\text{MoCo Loss} = -\log \frac{\exp(f(x)^T f(x^+))}{\exp(f(x)^T f(x^+)) + \sum_{j=1}^{N-1} \exp(f(x)^T f(x_j))}$$

M: 군집 개수
τ_C: *hyper parameter*
$H(Y)$: 제약식

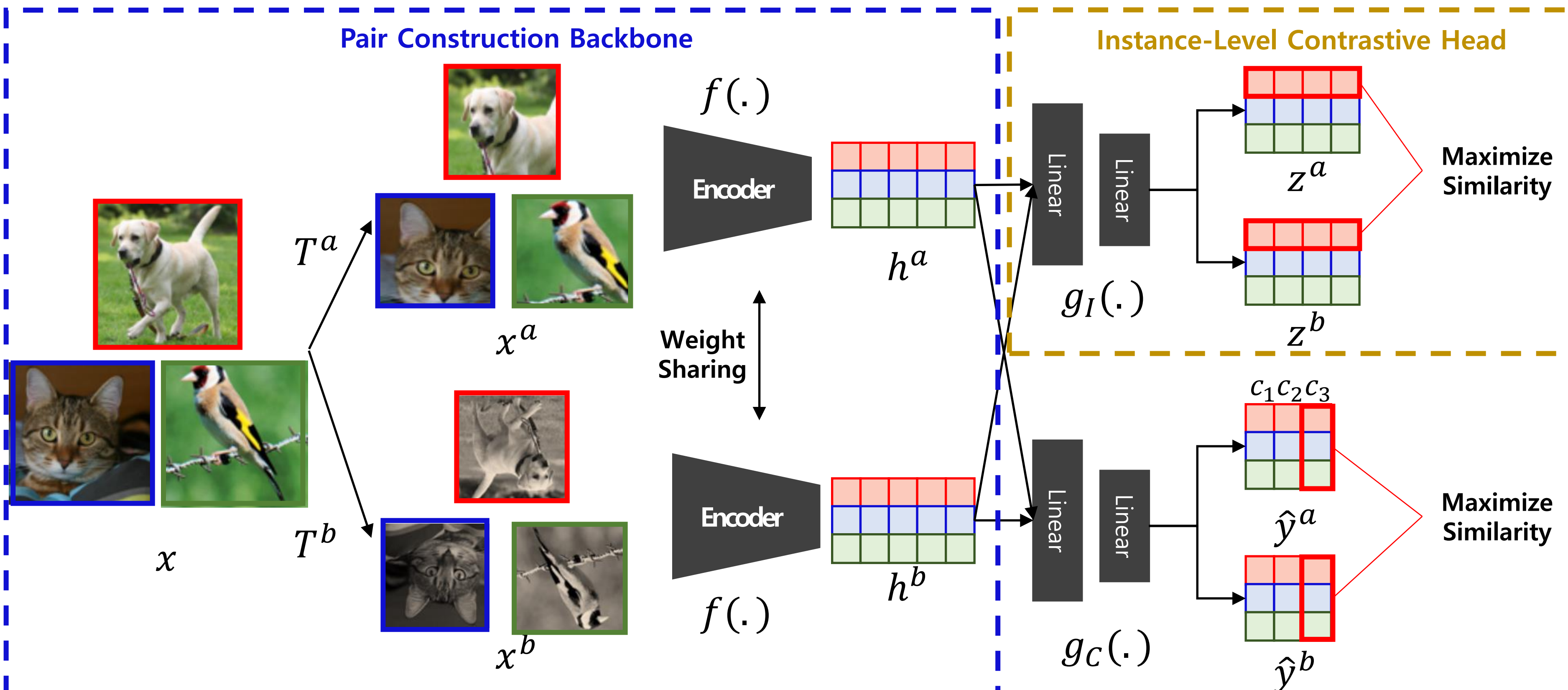
Cluster-Level Contrastive Head

같은 cluster는 가깝게

$$s(\hat{y}_i^a, \hat{y}_j^a) = \frac{(\hat{y}_i^a)(\hat{y}_i^b)^T}{\|\hat{y}_i^a\|\|\hat{y}_i^b\|}, \quad \hat{l}_i^a = -\log \frac{\exp\left(\frac{s(\hat{y}_i^a, \hat{y}_j^a)}{\tau_C}\right)}{\sum_{j=1}^{N} \left[\exp\left(\frac{s(\hat{y}_i^a, \hat{y}_j^a)}{\tau_C}\right) + \exp\left(\frac{s(\hat{y}_i^a, \hat{y}_j^b)}{\tau_C}\right)\right]}$$

$$L_{clu} = \frac{1}{2M} \sum_{i=1}^{N} (\hat{l}_i^a + \hat{l}_i^b) - H(Y)$$

Batch내 다른 cluster는 멀어지게



Reference: Li, Y., Hu, P., Liu, Z., Peng, D., Zhou, J. T., & Peng, X. (2021, May). Contrastive clustering. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 35, No. 10, pp. 8547-8555).

Copyright ⓒ 2023, All rights reserved.

Contrastive Clustering (CC)

$$\text{MoCo Loss} = -\log \frac{\exp\left(f(x)^T f(x^+)\right)}{\exp\left(f(x)^T f(x^+)\right) + \sum_{j=1}^{N-1} \exp\left(f(x)^T f(x_j)\right)}$$

M: 군집 개수
τ_C: $hyper\ parameter$
$H(Y)$: 제약식

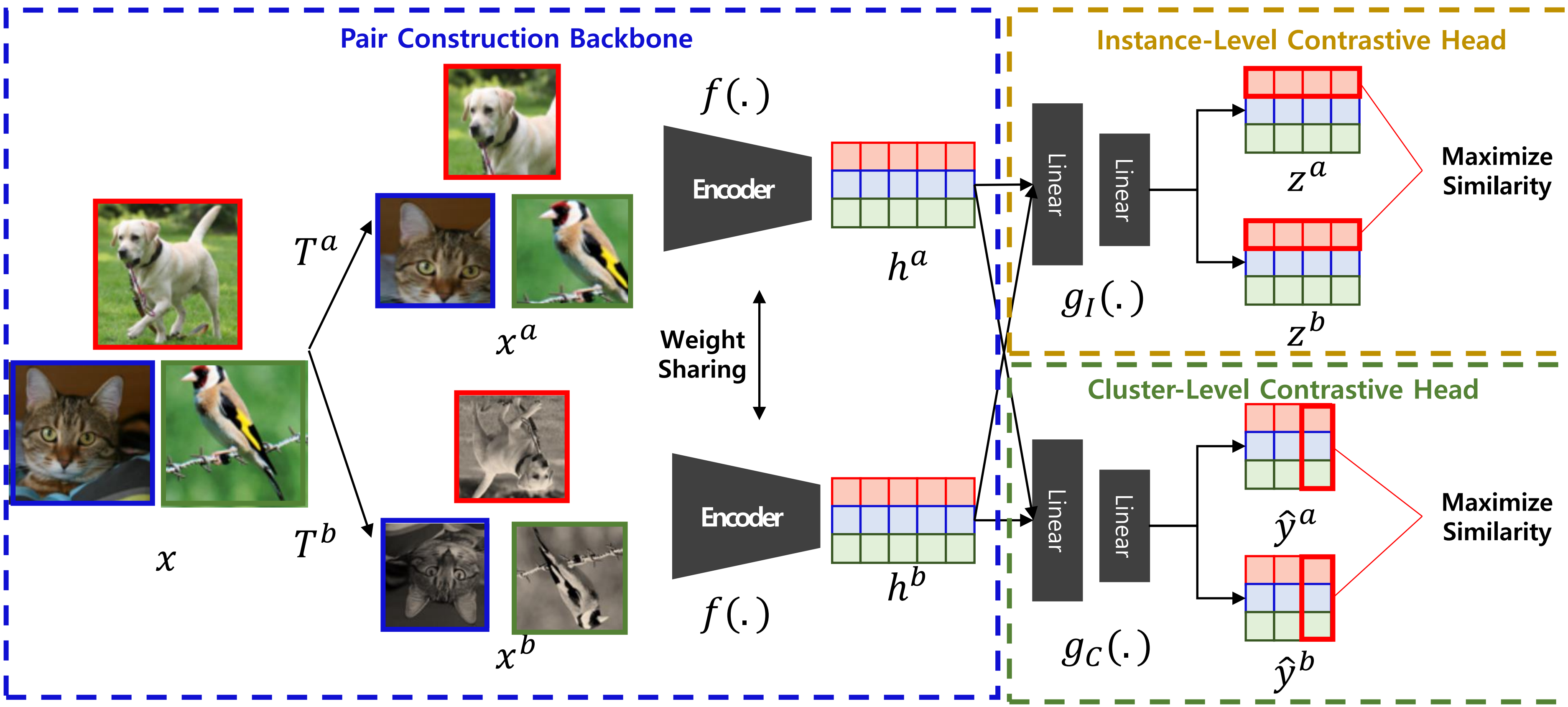
Cluster-Level Contrastive Head

같은 cluster는 가깝게

$$s(\hat{y}_i^a, \hat{y}_j^a) = \frac{(\hat{y}_i^a)(\hat{y}_i^b)^T}{\|\hat{y}_i^a\|\|\hat{y}_i^b\|}, \quad \hat{l}_i^a = -\log \frac{\exp\left(\frac{s(\hat{y}_i^a, \hat{y}_j^a)}{\tau_C}\right)}{\sum_{j=1}^{N} \left[\exp\left(\frac{s(\hat{y}_i^a, \hat{y}_j^a)}{\tau_C}\right) + \exp\left(\frac{s(\hat{y}_i^a, \hat{y}_j^b)}{\tau_C}\right)\right]}$$

$$L_{clu} = \frac{1}{2M} \sum_{i=1}^{N} (\hat{l}_i^a + \hat{l}_i^b) - H(Y)$$

Batch내 다른 cluster는 멀어지게



Reference: Li, Y., Hu, P., Liu, Z., Peng, D., Zhou, J. T., & Peng, X. (2021, May). Contrastive clustering. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 35, No. 10, pp. 8547-8555).

Copyright ⓒ 2023, All rights reserved.

Contrastive Clustering (CC)

정량 평가

- ❖ NMI (normalized mutual information), ACC (accuracy), ARI (adjusted rand index) 모두 clustering에 대한 정답이 있을 때 평가할 수 있는 지표 → 모두 1에 가깝게 클수록 좋은 결과임
- ❖ CC가 다양한 데이터셋에서 가장 좋은 성능을 보임

Dataset	CIFAR-10			CIFAR-100			STL-10			ImageNet-10		
Metrics	NMI	ACC	ARI	NMI	ACC	ARI	NMI	ACC	ARI	NMI	ACC	ARI
K-means	0.087	0.229	0.049	0.084	0.130	0.028	0.125	0.192	0.061	0.119	0.241	0.057
SC	0.103	0.247	0.085	0.090	0.136	0.022	0.098	0.159	0.048	0.151	0.274	0.076
AC	0.105	0.228	0.065	0.098	0.138	0.034	0.239	0.332	0.140	0.138	0.242	0.067
NMF	0.081	0.190	0.034	0.079	0.118	0.026	0.096	0.180	0.046	0.132	0.230	0.065
AE	0.239	0.314	0.169	0.100	0.165	0.048	0.250	0.303	0.161	0.210	0.317	0.152
DAE	0.251	0.297	0.163	0.111	0.151	0.046	0.224	0.302	0.152	0.206	0.304	0.138
DCGAN	0.265	0.315	0.176	0.120	0.151	0.045	0.210	0.298	0.139	0.225	0.346	0.157
DeCNN	0.240	0.282	0.174	0.092	0.133	0.038	0.227	0.299	0.162	0.186	0.313	0.142
VAE	0.245	0.291	0.167	0.108	0.152	0.040	0.200	0.282	0.146	0.193	0.334	0.168
JULE	0.192	0.272	0.138	0.103	0.137	0.033	0.182	0.277	0.164	0.175	0.300	0.138
DEC	0.257	0.301	0.161	0.136	0.185	0.050	0.276	0.359	0.186	0.282	0.381	0.203
DAC	0.396	0.522	0.306	0.185	0.238	0.088	0.366	0.470	0.257	0.394	0.527	0.302
ADC	–	0.325	–	–	0.160	–	–	0.530	–	–	–	–
DDC	0.424	0.524	0.329	–	–	–	0.371	0.489	0.267	0.433	0.577	0.345
DCCM	0.496	0.623	0.408	0.285	0.327	0.173	0.376	0.482	0.262	0.608	0.710	0.555
IIC	–	0.617	–	–	0.257	–	–	0.610	–	–	–	–
PICA	0.591	0.696	0.512	0.310	0.337	0.171	0.611	0.713	0.531	0.802	0.870	0.761
CC(Ours)	**0.705**	**0.790**	**0.637**	**0.431**	**0.429**	**0.266**	**0.764**	**0.850**	**0.726**	**0.859**	**0.893**	**0.822**

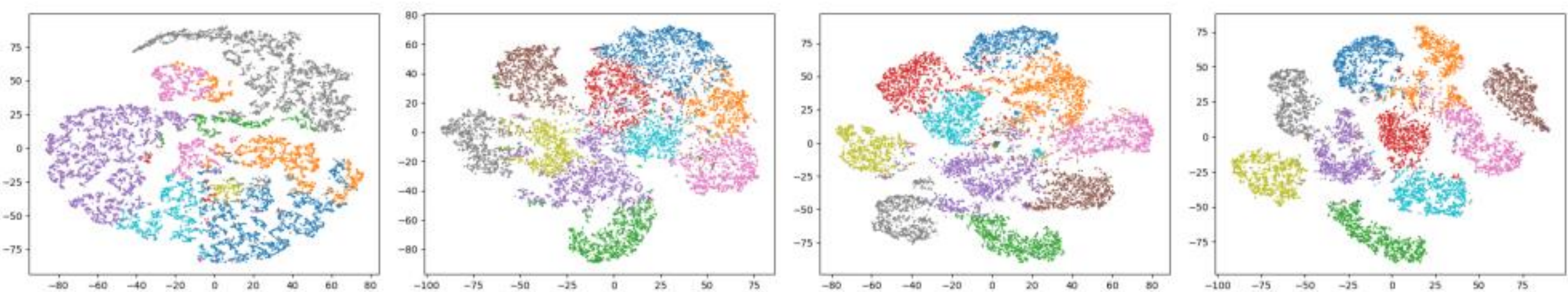
Reference: Li, Y., Hu, P., Liu, Z., Peng, D., Zhou, J. T., & Peng, X. (2021, May). Contrastive clustering. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 35, No. 10, pp. 8547-8555).

Copyright ⓒ 2023, All rights reserved.

Contrastive Clustering (CC)

정성 평가

- ❖ 학습 Epoch 수가 증가함에 따라 cluster별로 데이터가 잘 나뉘고 있는지를 t-SNE로 시각화
- ❖ Epoch 수가 증가(좌→우)함에 따라 cluster가 잘 분리되고 있음을 확인



(a) 0 epoch (NMI = 0.183) (b) 20 epoch (NMI = 0.472) (c) 50 epoch (NMI = 0.628) (d) 100 epoch (NMI = 0.737)

Figure 4: The evolution of instance features and cluster assignments across the training process on ImageNet-10. The colors indicate the cluster assignment obtained from CCH and the features for t-SNE are computed from ICH.

Reference: Li, Y., Hu, P., Liu, Z., Peng, D., Zhou, J. T., & Peng, X. (2021, May). Contrastive clustering. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 35, No. 10, pp. 8547-8555).

Copyright ⓒ 2023, All rights reserved.

3. Prototype Scattering and Positive Sampling (ProPos)

Data Mining
Quality Analytics

Prototype Scattering and Positive Sampling (ProPos)

ProPos논문

❖ Prototype Scattering and Positive Sampling (ProPos) (2022. 10)

- 군집은 contrastive learning을 하고 데이터는 non-contrastive learning 해서 clustering하는 연구
- 2022년도 IEEE **T**ransactions on **P**attern **A**nalysis and **M**achine **I**ntelligence에 개제됨
- 23년 2월 16일 기준 2회 인용

This article has been accepted for publication in IEEE Transactions on Pattern Analysis and Machine Intelligence. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/TPAMI.2022.3216454

IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. XX, NO. XX, XX 2022 1

Learning Representation for Clustering via Prototype Scattering and Positive Sampling

Zhizhong Huang, *Graduate Student Member, IEEE*, Jie Chen, Junping Zhang, *Senior Member, IEEE*, and Hongming Shan, *Senior Member, IEEE*

Abstract—Existing deep clustering methods rely on either contrastive or non-contrastive representation learning for downstream clustering task. Contrastive-based methods thanks to negative pairs learn uniform representations for clustering, in which negative pairs, however, may inevitably lead to the class collision issue and consequently compromise the clustering performance. Non-contrastive-based methods, on the other hand, avoid class collision issue, but the resulting non-uniform representations may cause the collapse of clustering. To enjoy the strengths of both worlds, this paper presents a novel end-to-end deep clustering method with prototype scattering and positive sampling, termed ProPos. Specifically, we first maximize the distance between prototypical representations, named prototype scattering loss, which improves the uniformity of representations. Second, we align one augmented view of instance with the sampled neighbors of another view—assumed to be truly positive pair in the embedding space—to improve the within-cluster compactness, termed positive sampling alignment. The strengths of ProPos are avoidable class collision issue, uniform representations, well-separated clusters, and within-cluster compactness. By optimizing ProPos in an end-to-end expectation-maximization framework, extensive experimental results demonstrate that ProPos achieves competing performance on moderate-scale clustering benchmark datasets and establishes new state-of-the-art performance on large-scale datasets. Source code is available at https://github.com/Hzzone/ProPos.

Index Terms—Contrastive learning, deep clustering, representation learning, self-supervised learning, unsupervised learning

Reference: Huang, Z., Chen, J., Zhang, J., & Shan, H. (2022). Learning Representation for Clustering Via Prototype Scattering and Positive Sampling. IEEE Transactions on Pattern Analysis and Machine Intelligence.

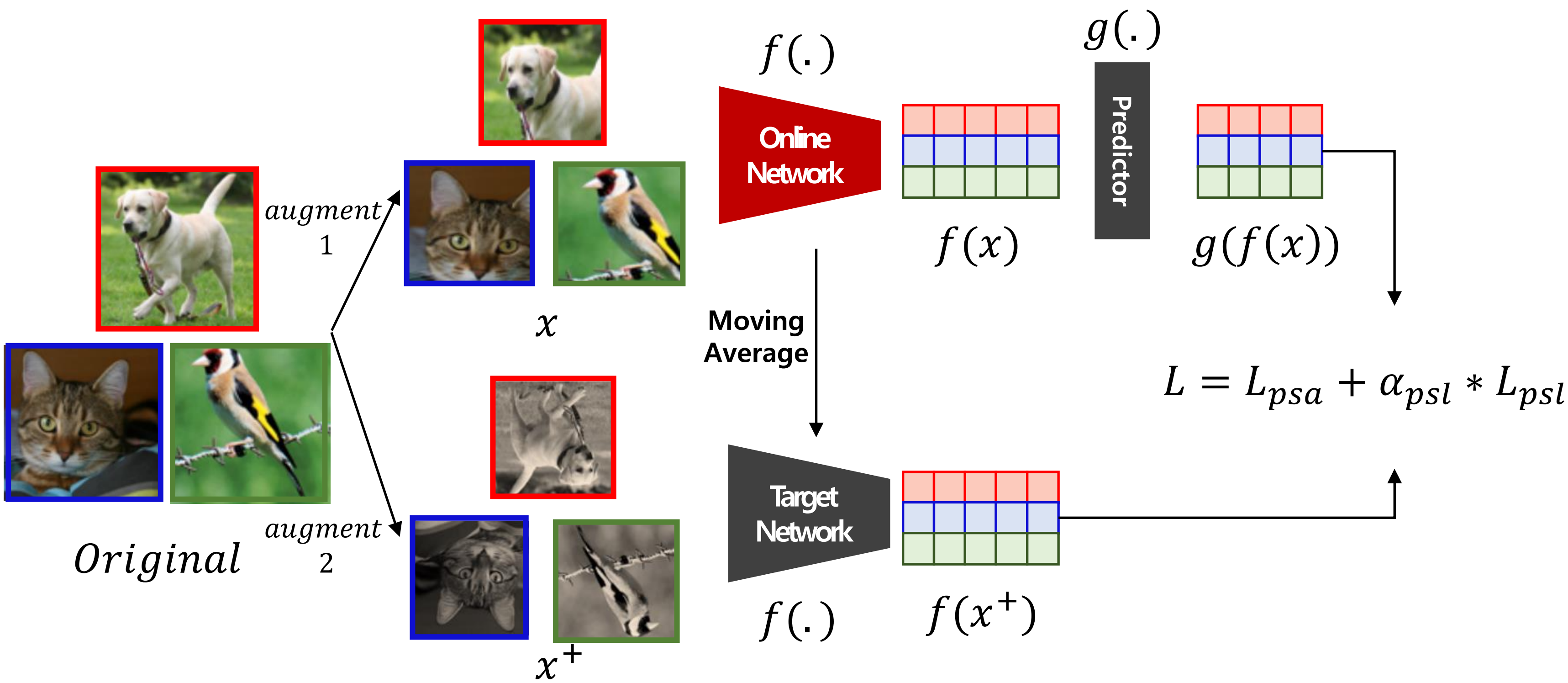
Copyright ⓒ 2023, All rights reserved.

Prototype Scattering and Positive Sampling (ProPos)

$m, \alpha_{psl}: hyper\ parameter$

ProPos 간략 구조

- ❖ CC 논문에서의 Pair Construction Backbone 구조와 유사한 구조를 지님
- ❖ 하지만 online network와 target network가 weigh를 공유하는 형태가 아님
- ❖ Target network는 momentum update 방식 사용 $(\theta_{target} = m\theta_{target} + (1-m)\theta_{online})$

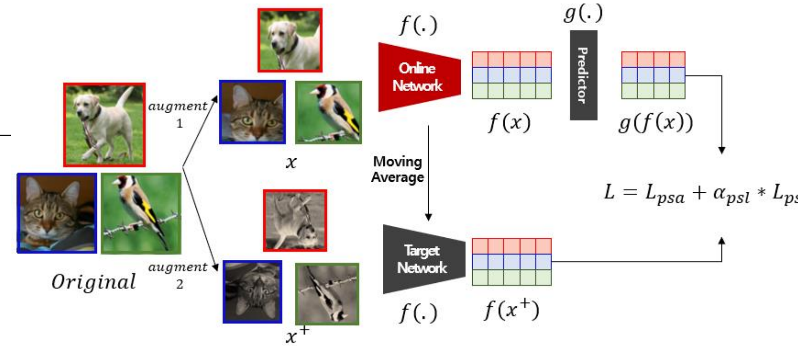


Reference: Huang, Z., Chen, J., Zhang, J., & Shan, H. (2022). Learning Representation for Clustering Via Prototype Scattering and Positive Sampling. IEEE Transactions on Pattern Analysis and Machine Intelligence.

Copyright ⓒ 2023, All rights reserved.

ProPos

$$\text{MoCo Loss} = -\log \frac{\exp\left(f(x)^T f(x^+)\right)}{\exp\left(f(x)^T f(x^+)\right) + \sum_{j=1}^{N-1} \exp\left(f(x)^T f(x_j)\right)}$$

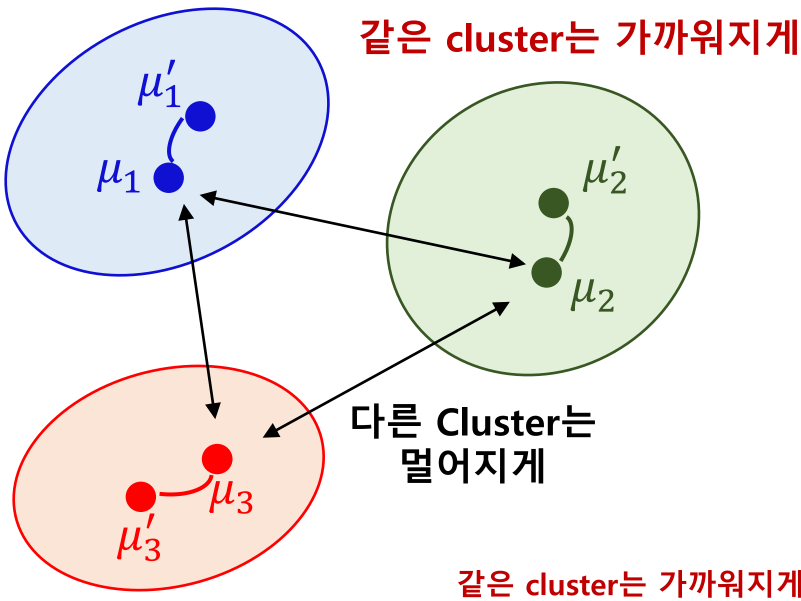


PSL Loss와 & PSA Loss

❖ PSL Loss: Prototype Scattering Loss (Cluster)

❖ PSA Loss: Positive Sampling Alignment Loss (개체)

<PSL Loss>

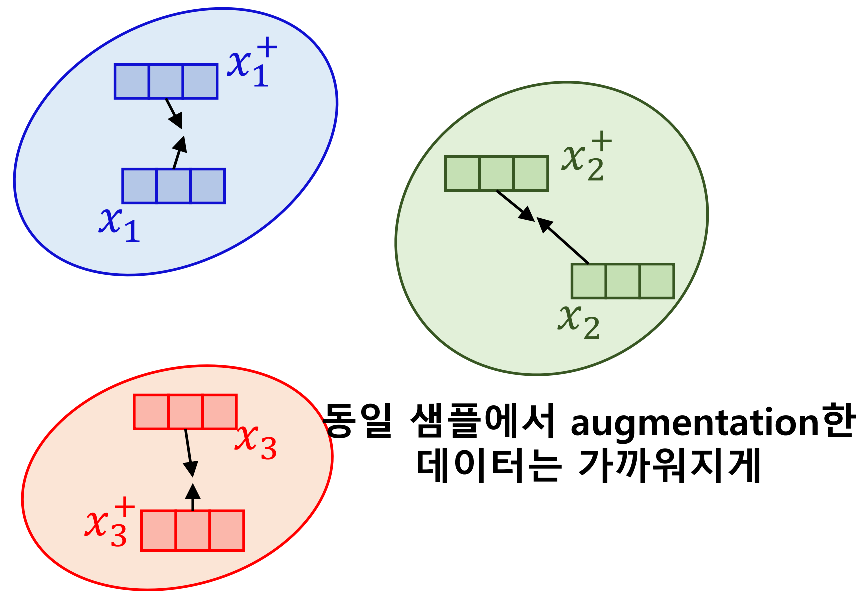


같은 cluster는 가까워지게

$$L_{psa} = \frac{1}{K}\sum_{k=1}^{K} -log \frac{\exp\left(\frac{\mu_k^T, \mu_k'}{\tau}\right)}{\exp\left(\frac{\mu_k^T, \mu_k'}{\tau}\right) + \sum_{j\neq k}^{K} exp\left(\frac{\mu_k^T, \mu_j}{\tau}\right)}$$

다른 cluster는 멀어지게

<PSA Loss>



BYOL L2 Loss와 유사

$$L_{psl} = \|g(f(x) + \sigma\varepsilon) - f'(x^+) - \|_2^2$$

$$\varepsilon \sim N(0, I)$$

Reference: Huang, Z., Chen, J., Zhang, J., & Shan, H. (2022). Learning Representation for Clustering Via Prototype Scattering and Positive Sampling. IEEE Transactions on Pattern Analysis and Machine Intelligence.

Copyright ⓒ 2023, All rights reserved.

Data Mining Quality Analytics

Prototype Scattering and Positive Sampling (ProPos)

ProPos 상세 구조

- ❖ ProPos에서는 매 epoch별로 k-means를 활용해 pseudo clustering label을 생성함
- ❖ PSA loss와 PSL loss를 기반으로 pseudo clustering label이 더 정확해질 수 있도록 학습 진행

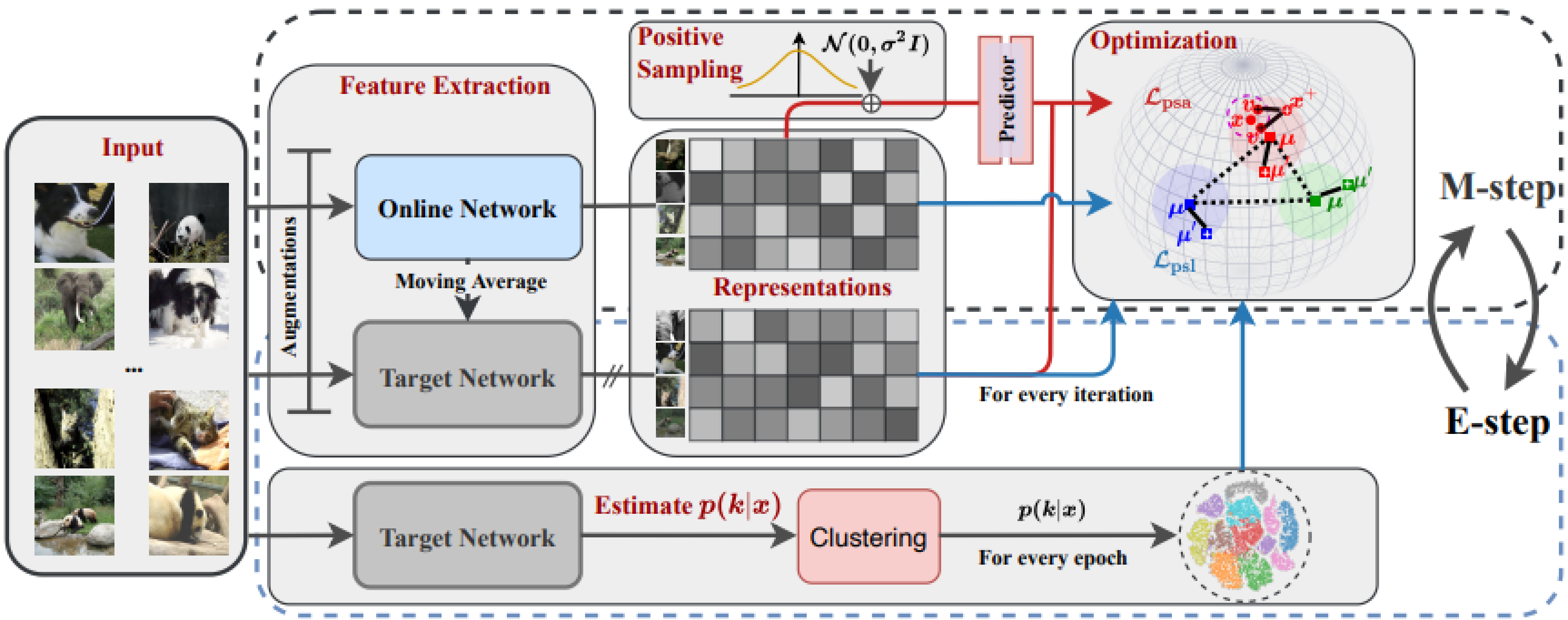


Fig. 2. The overall framework of the proposed ProPos in an EM framework.

Reference: Huang, Z., Chen, J., Zhang, J., & Shan, H. (2022). Learning Representation for Clustering Via Prototype Scattering and Positive Sampling. IEEE Transactions on Pattern Analysis and Machine Intelligence.

Copyright ⓒ 2023, All rights reserved.

Prototype Scattering and Positive Sampling (ProPos)

정량 평가

- ❖ 제안한 ProPos가 다양한 데이터셋에서 가장 좋은 성능을 보임
- ❖ 특히, CIFAR-10, ImageNet-10처럼 class별로 데이터 수도 충분하고 균형 있을 때 더 큰 성능 보임

TABLE 2
Clustering results (%) of various methods on five benchmark datasets. The best and second best results are shown in bold and underline, respectively. We split different methods according to different training paradigms. The works most related to our method are IDFD and PCL that improve the representations for clustering.

Method	CIFAR-10			CIFAR-20			STL-10			ImageNet-10			ImageNet-Dogs		
	NMI	ACC	ARI	NMI	ACC	ARI	NMI	ACC	ARI	NMI	ACC	ARI	NMI	ACC	ARI
IIC [36]	51.3	61.7	41.1	-	25.7	-	43.1	49.9	29.5	-	-	-	-	-	-
DCCM [35]	49.6	62.3	40.8	28.5	32.7	17.3	37.6	48.2	26.2	60.8	71.0	55.5	32.1	38.3	18.2
PICA [55]	56.1	64.5	46.7	29.6	32.2	15.9	-	-	-	78.2	85.0	73.3	33.6	32.4	17.9
SCAN [6]	79.7	88.3	77.2	48.6	50.7	33.3	69.8	80.9	64.6	-	-	-	-	-	-
NMM [56]	74.8	84.3	70.9	48.4	47.7	31.6	69.4	80.8	65.0	-	-	-	-	-	-
CC [8][1]	70.5	79.0	63.7	43.1	42.9	26.6	76.4	85.0	72.6	85.9	89.3	82.2	44.5	42.9	27.4
MiCE [10]	73.7	83.5	69.8	43.6	44.0	28.0	63.5	75.2	57.5	-	-	-	42.3	43.9	28.6
GCC [41]	76.4	85.6	72.8	47.2	47.2	30.5	68.4	78.8	63.1	84.2	90.1	82.2	49.0	52.6	36.2
TCL [38][1]	81.9	88.7	78.0	52.9	53.1	35.7	**79.9**	**86.8**	**75.7**	87.5	89.5	83.7	62.3	64.4	51.6
TCC [39]	79.0	90.6	73.3	47.9	49.1	31.2	73.2	81.4	68.9	84.8	89.7	82.5	55.4	59.5	41.7
MoCo [1]	66.9	77.6	60.8	39.0	39.7	24.2	61.5	72.8	52.4	-	-	-	34.7	33.8	19.7
SimSiam [2]	78.6	85.6	73.6	52.2	48.5	32.7	65.9	71.6	57.2	83.1	92.1	83.3	58.3	67.4	50.1
BYOL [3]	81.7	89.4	79.0	55.9	56.9	39.3	71.3	82.5	65.7	86.6	93.9	87.2	63.5	69.4	54.8
IDFD [9]	71.1	81.5	66.3	42.6	42.5	26.4	64.3	75.6	57.5	**89.8**	95.4	90.1	54.6	59.1	41.3
PCL [7]	80.2	87.4	76.6	52.8	52.6	36.3	71.8	41.0	67.0	84.1	90.7	82.2	44.0	41.2	29.9
ProPos (ours)	**88.6**	**94.3**	**88.4**	**60.6**	**61.4**	**45.1**	75.8	86.7	73.7	89.6	**95.6**	**90.6**	**69.2**	**74.5**	**62.7**

[1] CC and TCL use a large image size of 224 × 224 for all datasets.

Reference: Huang, Z., Chen, J., Zhang, J., & Shan, H. (2022). Learning Representation for Clustering Via Prototype Scattering and Positive Sampling. IEEE Transactions on Pattern Analysis and Machine Intelligence.

Copyright ⓒ 2023, All rights reserved.

Prototype Scattering and Positive Sampling (ProPos)

정성 평가

- ❖ ProPos 결과에 대한 t-SNE 시각화 결과를 보여줌
- ❖ BYOL보다 더 clustering이 잘 되었음을 확인할 수 있음
- ❖ 또한, ProPos에서 제안한 PSL, PSA loss를 모두 사용했을 때 결과가 가장 좋았음

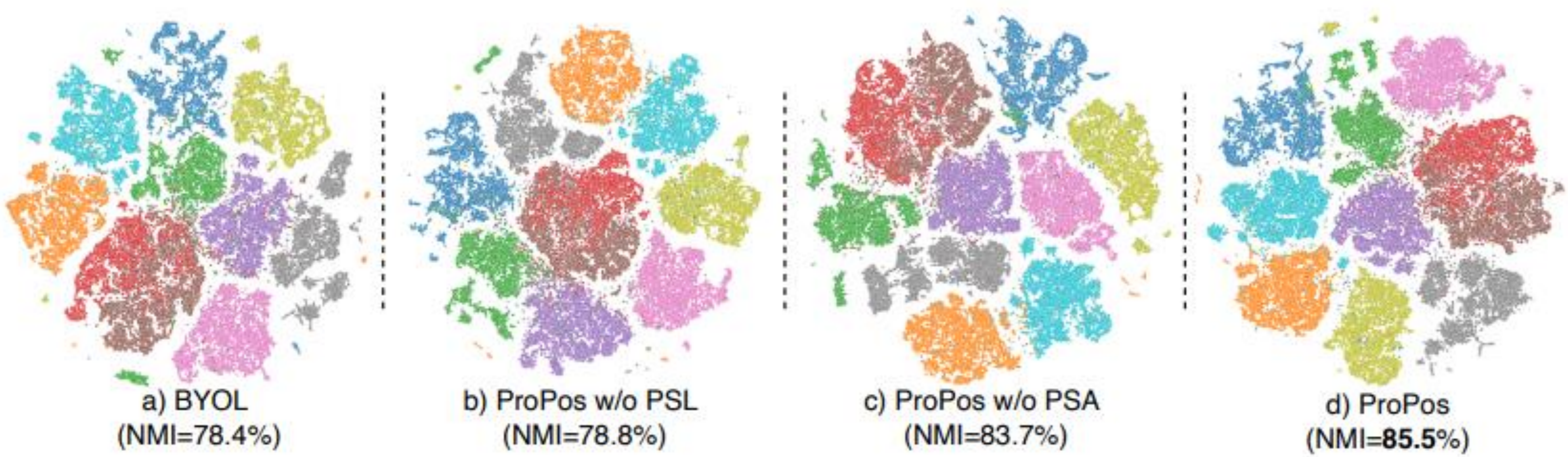


Fig. 5. Visualization of feature representations learned by different representation learning frameworks and our proposed ProPos on CIFAR-10 with t-SNE. Zoom in for better view.

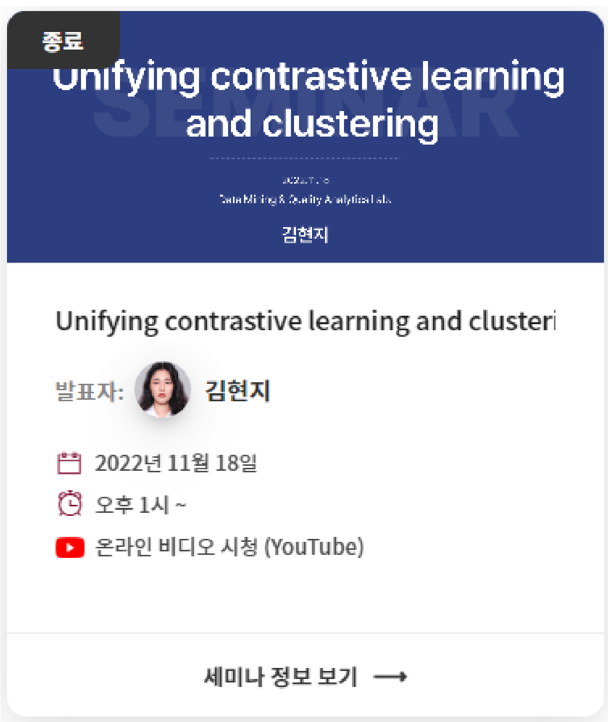
Reference: Huang, Z., Chen, J., Zhang, J., & Shan, H. (2022). Learning Representation for Clustering Via Prototype Scattering and Positive Sampling. IEEE Transactions on Pattern Analysis and Machine Intelligence.

Copyright ⓒ 2023, All rights reserved.

기존 연구실 세미나

이전 연구실 세미나

- ❖ 본 세미나에서 소개하지 않은 Deep Clustering 기법을 소개한 기존 연구실 세미나를 추천
- ❖ PCL, SwAV 연구에 대해 자세한 내용을 들을 수 있음



PROTOTYPICAL CONTRASTIVE LEARNING OF UNSUPERVISED REPRESENTATIONS

(ICLR 2021)

Published as a conference paper at ICLR 2021

PROTOTYPICAL CONTRASTIVE LEARNING OF UNSUPERVISED REPRESENTATIONS

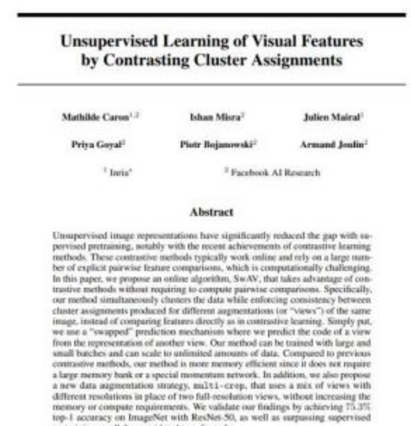
Junnan Li, Pan Zhou, Caiming Xiong, Steven C.H. Hoi
Salesforce Research
{junnan.li,pzhou,cxiong,shoi}@salesforce.com

ABSTRACT

This paper presents Prototypical Contrastive Learning (PCL), an unsupervised representation learning method that bridges contrastive learning with clustering. PCL not only learns low-level features for the task of instance discrimination, but more importantly, it encodes semantic structures discovered by clustering into the learned embedding space. Specifically, we introduce prototypes as latent variables to help find the maximum-likelihood estimation of the network parameters in an Expectation-Maximization framework. We iteratively perform E-step as finding the distribution of prototypes via clustering and M-step as optimizing the network via contrastive learning. We propose ProtoNCE loss, a generalized version of the InfoNCE loss for contrastive learning, which encourages representations to be closer to their assigned prototypes. PCL outperforms state-of-the-art instance-wise contrastive learning methods on multiple benchmarks with substantial improvement in low-resource transfer learning. Code and pretrained models are available at https://github.com/salesforce/PCL.

Unsupervised Learning of Visual Features by Contrasting Cluster Assignments

(NeurIPS 2020)



Unsupervised Learning of Visual Features by Contrasting Cluster Assignments

Mathilde Caron[1,2] Ishan Misra[2] Julien Mairal[1]
Priya Goyal[2] Piotr Bojanowski[2] Armand Joulin[2]
[1] Inria* [2] Facebook AI Research

Abstract

Unsupervised image representations have significantly reduced the gap with supervised pretraining, notably with the recent achievements of contrastive learning methods. These contrastive methods typically work online and rely on a large number of explicit pairwise feature comparisons, which is computationally challenging. In this paper, we propose an online algorithm, SwAV, that takes advantage of contrastive methods without requiring to compute pairwise comparisons. Specifically, our method simultaneously clusters the data while enforcing consistency between cluster assignments produced for different augmentations (or "views") of the same image, instead of comparing features directly as in contrastive learning. Simply put, we use a "swapped" prediction mechanism where we predict the code of a view from the representation of another view. Our method can be trained with large and small batches and can scale to unlimited amounts of data. Compared to previous contrastive methods, our method is more memory efficient since it does not require a large memory bank or a special momentum network. In addition, we also propose a new data augmentation strategy, multi-crop, that uses a mix of views with different resolutions in place of two full-resolution views, without increasing the memory or compute requirements. We validate our findings by achieving 75.3% top-1 accuracy on ImageNet with ResNet-50, as well as surpassing supervised pretraining on all the considered transfer tasks.

Reference: http://dmqa.korea.ac.kr/activity/seminar/386

Copyright ⓒ 2023, All rights reserved.

4. Conclusions

Conclusions

❖ Conclusions

✓ 데이터 내 좋은 특징을 잘 추출할 수 있는 딥러닝 모델 이를 활용해 clustering을 진행하려는 Deep Clustering 방법론이 많이 발전되고 있음

✓ 특히, 그 중에서도 자기도학습 방법론 (Self-supervised learning)이 발전되면서 이를 활용해 clustering을 진행하는 연구들이 활발히 진행되고 있음

✓ Contrastive Clustering (CC) 방법론은 군집 간 contrastive learning과 각 데이터 개체 간 contrastive learning을 사용해 군집화 성능을 향상 시켰음

✓ Prototype Scattering and Positive Sampling (ProPos) 방법론은 군집 사이는 contrastive learning을 사용하고 데이터 개체는 non-contrastive learning을 사용해 군집하는 방법론으로 CC 대비해서도 군집화 성능을 크게 향상 시킨 알고리즘임

✓ 군집 간 거리는 멀어지고 유사한 특징을 지닌 개체는 가까워지도록 학습을 유도할 수 있다면 더 좋은 성능의 Deep Clustering 알고리즘을 제안할 수 있을 것으로 기대함

Copyright ⓒ 2023, All rights reserved.

감사합니다.

Data Mining
Quality Analytics

Appendix

Data Mining
Quality Analytics

❖ Reference (Paper)

- Li, Y., Hu, P., Liu, Z., Peng, D., Zhou, J. T., & Peng, X. (2021, May). Contrastive clustering. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 35, No. 10, pp. 8547-8555).
- Huang, Z., Chen, J., Zhang, J., & Shan, H. (2022). Learning Representation for Clustering Via Prototype Scattering and Positive Sampling. IEEE Transactions on Pattern Analysis and Machine Intelligence.
- Ren, Y., Pu, J., Yang, Z., Xu, J., Li, G., Pu, X., ... & He, L. (2022). Deep clustering: A comprehensive survey. arXiv preprint arXiv:2210.04142.

❖ Reference (Website)

- https://www.lgresearch.ai/blog/view/?seq=193
- http://dmqa.korea.ac.kr/activity/seminar/308
- http://dmqa.korea.ac.kr/activity/seminar/310
- http://dmqa.korea.ac.kr/activity/seminar/355
- http://dmqa.korea.ac.kr/activity/seminar/386