Use All the Labels: A Hierarchical Multi-Label Contrastive Learning Framework
(CVPR 2022)

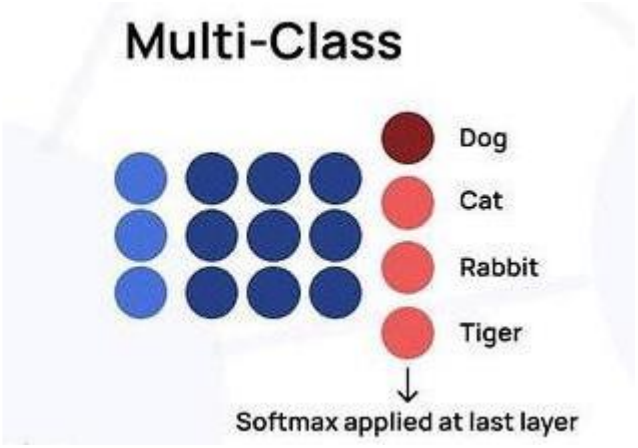
Shu Zhang, Rang Xu, Caiming Xiong, Chetan Ramaiah

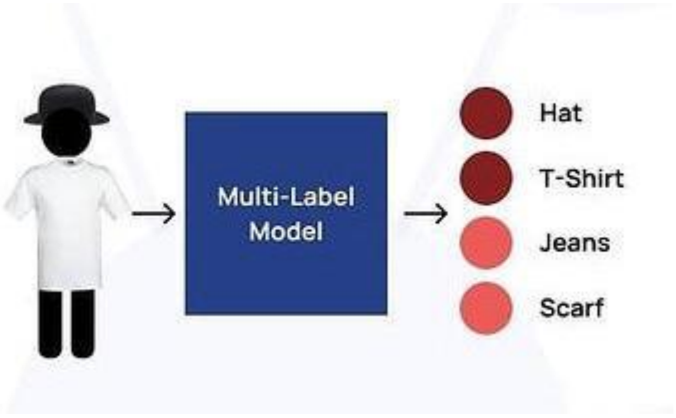
정 시 열

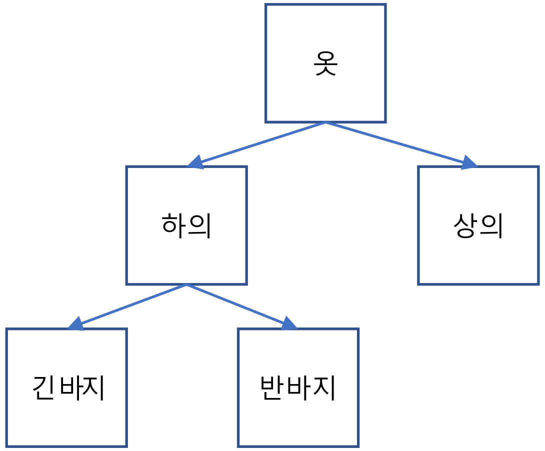
목차

Introduction

- 현실 세계에서는 Data가 Hierarchical Multi-labels인 경우가 많음.
- 하지만 현재 Contrastive Learning Framework는 Single Supervisory Signal에 중점
- Single Task Learning에서는 Single Supervisory Signal에 집중해도 괜찮지만 Representation Learning에서는 데이터의 모든 정보를 활용하는 것이 중요하다. 왜냐하면 임의의 downstream task에서 데이터를 일관되고 정확하게 표현해야 하기 때문이다.

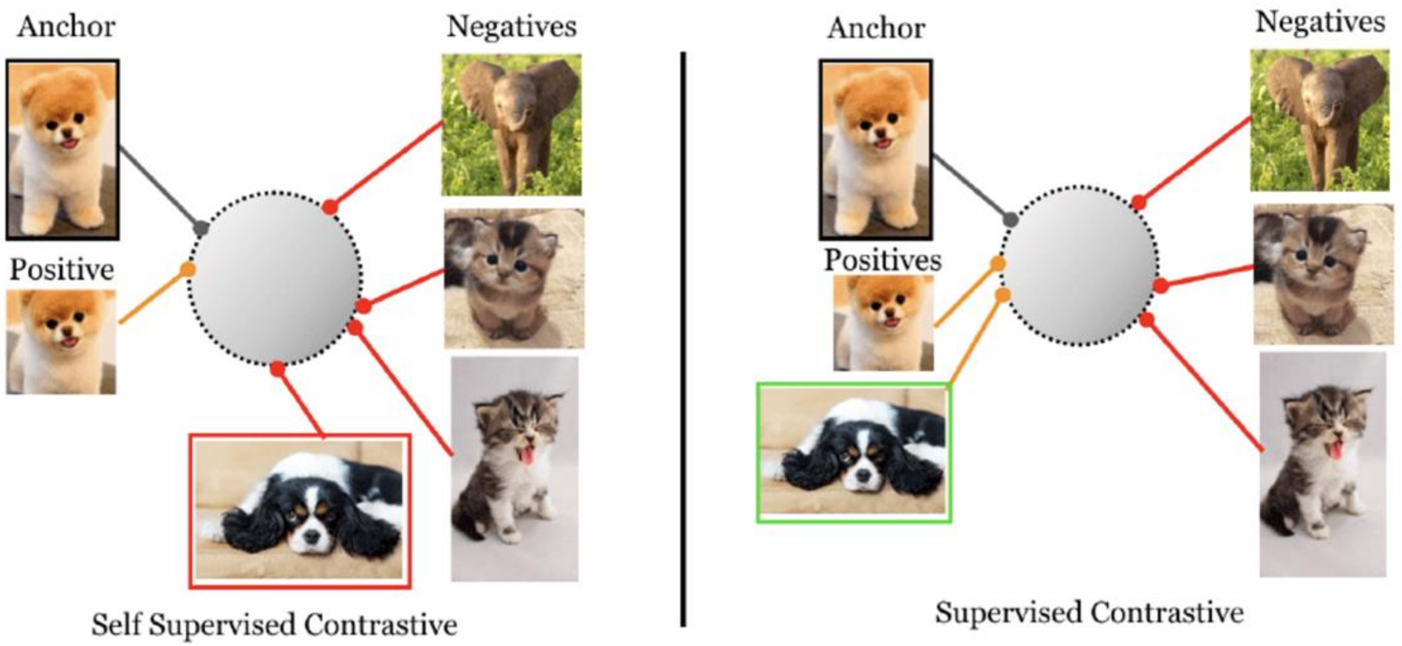






Introduction

- Contrastive Learning의 Unsupervised 와 Supervised 방식은 Positive를 구성하는 방식에서 차이를 가짐.
  Unsupervised : different view of the same image
  Supervised : same class and their augmentation



Introduction

- 기존의 접근 방법들은 Multi-label learning을 지원하지 않고 label들 간의 관계 이용하지 못한다는 단점이 있음
- 이에 모든 label을 사용하며, Embedding space내부에 label들의 관계성을 담아낼 수 있는 3가지의 loss를 제시함.

Methodology

- **Self-Supervised Contrastive Loss**

$$\mathcal{L}^{self} = \sum_{i=1}^{2N} \mathcal{L}_i^{self}$$

$$\mathcal{L}_i^{self} = -\log \boxed{\frac{\exp\left(\boldsymbol{z}_i \cdot \boldsymbol{z}_{j(i)}/\tau\right)}{\sum_{k=1}^{2N} \mathbb{1}_{i \neq k} \cdot \exp\left(\boldsymbol{z}_i \cdot \boldsymbol{z}_k/\tau\right)}}$$

Inner product between positives **maximize**

Inner product between negatives **minimize**

minimize → **maximize**

- i : anchor image
- j : positive sample

- **Supervised Contrastive Loss**

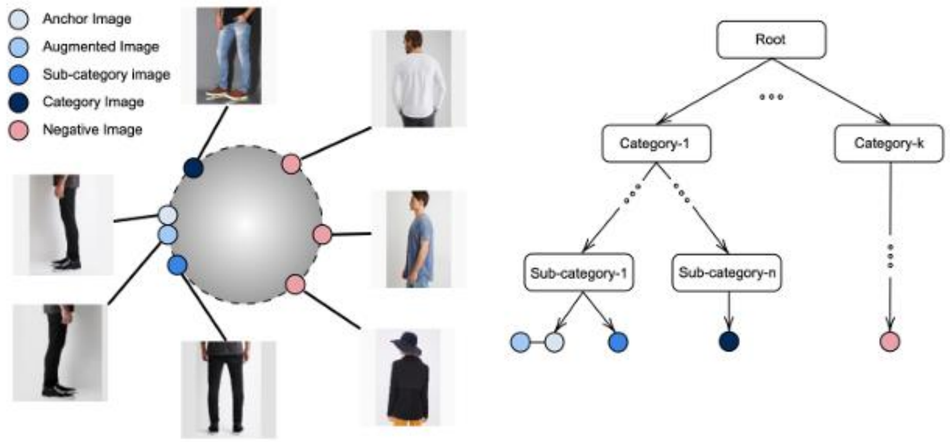
한계점: Single Label에서만 적용 가능.

$$L^{\text{sup}} = \sum_{i \in I} \frac{-1}{|P(i)|} \sum_{p \in P} \log \frac{\exp(f_i \cdot f_p/\tau)}{\sum_{a \in A \setminus i} \exp(f_i \cdot f_a/\tau)}$$

- i : anchor image
- p : positive sample
- P(i) : positive sample 개수
- f : feature vector

Methodology

- **Terminology**
  - ✓ Hierarchical multi–label의 구조를 Tree로 형상화 함
    Leaf node : unique image identifier
    Non–Leaf node : Broader Category
  - ✓ Level (Depth)
    (High) : correspond to specific categories
  - ✓ Positive Pair
    레벨 l에서 공통의 부모 노드를 가진 pair



(a) (b)

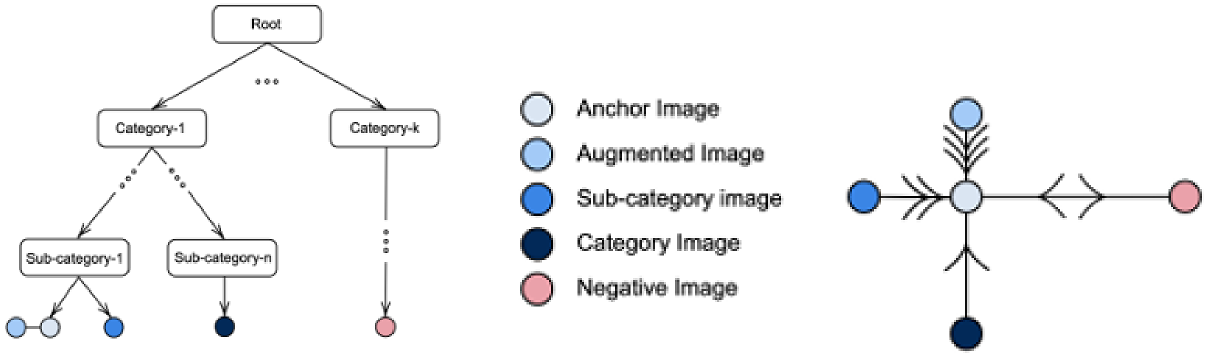
Methodology

- **Hierarchical Multi-label Contrastive Loss (HMC)**
  - 높은 레벨에서 구성된 Pair에게는 보다 더 많은 가중치를 부여하여 더욱 가깝게 만든다

$$L^{\text{pair}}(i, p_l^i) = \log \frac{\exp(f_i \cdot f_p^l / \tau)}{\sum_{a \in A \setminus i} \exp(f_i \cdot f_a / \tau)}$$

$$L^{\text{HMC}} = \sum_{l \in L} \frac{1}{|L|} \sum_{i \in I} \frac{-\lambda_l}{|P_l(i)|} \sum_{p_l \in P_l} L^{\text{pair}}(i, p_l^i)$$

- i : anchor image
- p : positive sample
- l : Level
- f : feature vector
- $\lambda_l = f(l)$: fixed penalty for each level



(a)

(b)

Approach	Classification Accuracy
HiMulCon, $\lambda_l = 1$	73.1
HiMulConE, $\lambda_l = 1$	73.8
HiMulCon, $\lambda_l = \frac{1}{l}$)	70.8
HiMulCon, $\lambda_l = exp(\frac{1}{l})$)	70.29
HiMulCon, $\lambda_l = 2^{\frac{1}{l}}$)	69.4
SupCon	73.6
HiMulCon, $\lambda_l = exp(\frac{1}{\lvert L\rvert - l})$	74.29
HiMulConE, $\lambda_l = exp(\frac{1}{\lvert L\rvert - l})$	**76.07**

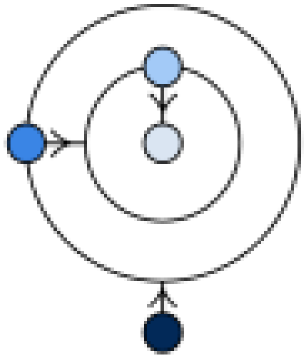
Methodology

- **Hierarchical Constraint Enforcing Loss (HiConE)**
  - 계층 구조를 지키기 위해선 Data가 특정 Class에 속하면, 상위 Class에도 속해야함
  - Hierarchical Constraint를 "**높은 레벨에서 구성된 pair의 loss는 낮은 레벨에서 구성된 pair의 loss보다 클 수 없다**."고 정의.

$$L^{\mathrm{pair}}(i, p_l^i) = \log \frac{\exp(f_i \cdot f_p^l / \tau)}{\sum_{a \in A \setminus i} \exp(f_i \cdot f_a / \tau)}$$

$$L^{\mathrm{pair}}_{\max}(l) = \max_{(i, p_l^i)} L^{\mathrm{pair}}(i, p_l^i)$$

$$\sum_{l \in L} \frac{1}{|L|} \sum_{i \in I} \frac{-1}{|P(i)|} \sum_{p_l \in P_l} \max(L^{\mathrm{pair}}(i, p_l^i), L^{\mathrm{pair}}_{\max}(l-1))$$

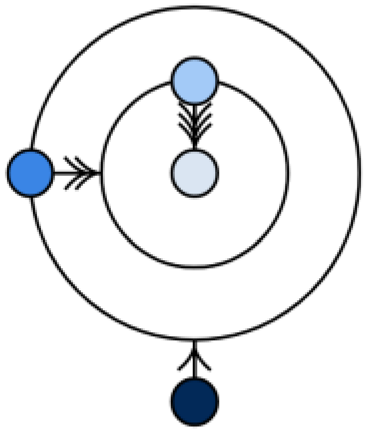


(b)

Methodology

- **Hierarchical Multi-label Constraint Enforcing Contrastive Loss (HiMulConE)**
  - HMC Loss 성질과 HiConE Loss 성질 결합

$$\sum_{l\in L}\frac{1}{|L|}\sum_{i\in I}\frac{-\lambda_l}{|P(i)|}\sum_{p_l\in P_l}\max(L^{\text{pair}}(i,p_l^i),L^{\text{pair}}_{\max}(l-1))$$



(c)

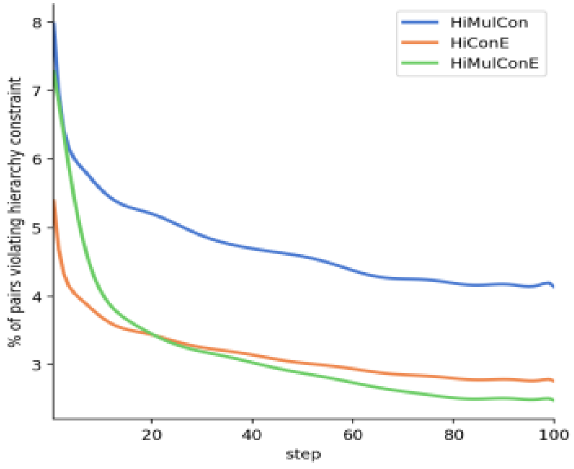
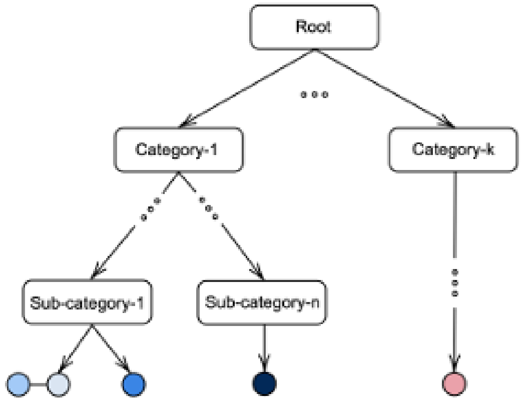


Figure 4. Hierarchical Violations during training

Loss Function	% of pairs violating constraint
HiMulCon	7.45
HiConE	4.26
HiMulConE	**3.74**
SupCon	9.4
SimClr	14.95
Cross Entropy	27.43

Methodology

- **Hierarchical Batch Sampling Strategy**
  - 사전 연구들을 통해서, representation learning에서 sampling이 중요하며 배치 내부에 많은 양의 hard positives/negatives를 구성하는 것이 중요하다고 밝혀짐.
  - Hierarchical Multi-label setting에서도 batch내부에 전 Level에 걸친 충분한 표현이 담겨야 하기에 Sampling 방법을 제안함.
    1. Random하게 Anchor image를 선별
    2. Multi-label의 각 label별 구성하는 sub-tree에서 이미지를 샘플링 한다. (모든 라벨에서 추출한다)
    3. Batch size만큼 반복



(b)

Experiments

- 총 3가지 Downstream task를 통하여 제안하는 Loss를 검증함.
  (image classification, image retrieval accuracy on sub-categories, NMI for clustering quality)
- Dataset을 seen과 unseen으로 나눠서, seen data를 기반으로 Encoder를 학습시키고 frozen 시킨 이후에, unseen data를 활용한 Downstream task 수행
  (예를 들어 40개의 카테고리가 있다면, 22개의 카테고리는 seen으로 18개의 카테고리는 unseen으로 설정)

Classification accuracy

	ImageNet [29]	DeepFashion [21]	iNaturalist [36]	ModelNet40 [45]
SimCLR [5]	69.53	70.38	54.02	79.26
Cross Entropy	77.60	72.44	56.86	81.31
SupCon [18]	78.70	72.82	57.28	81.60
Guided [10]	76.60	72.61	57.33	83.49
HiMulConE (Ours)	**79.14**	**73.21**	**59.40**	**88.46**

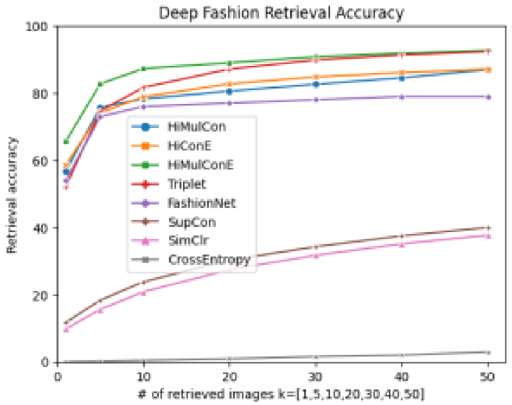
Table 1. Top-1 classification accuracy on the full datasets. Standard datasets and splits as described in the original papers are used here. For ImageNet and iNaturalist datasets, the task is classification at the finest sub-category level, whereas super-category level classification is performed for DeepFashion and ModelNet40. All baseline results were reproduced.

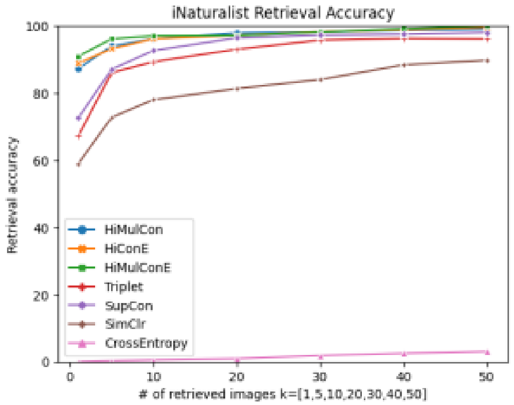
	DeepFashion		ModelNet40	
	Seen	Unseen	Seen	Unseen
SimCLR	70.26	68.12	77.09	72.26
Cross Entropy	77.81	71.94	85.17	79.77
SupCon	**81.46**	73.93	88.33	79.28
Guided	79.34	74.04	89.01	82.22
HiMulCon	80.54	74.88	89.28	84.44
HiConE	80.67	75.28	89.09	84.40
HiMulConE	80.52	**75.29**	**89.45**	**85.37**

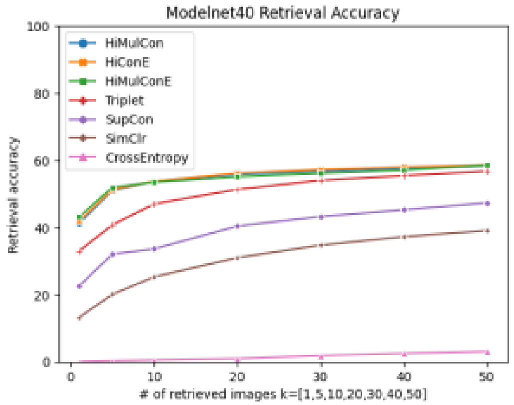
Table 2. Top-1 classification accuracy on the seen / unseen splits of DeepFashion In-Shop and ModelNet40.

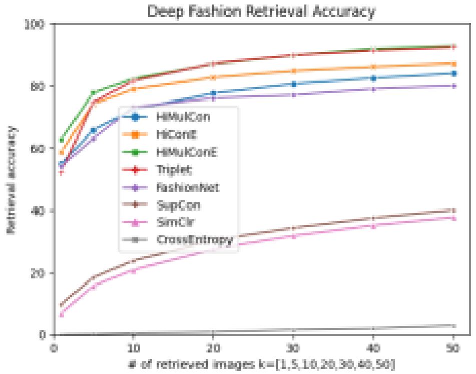
Experiments

Image Retrieval Accuracy









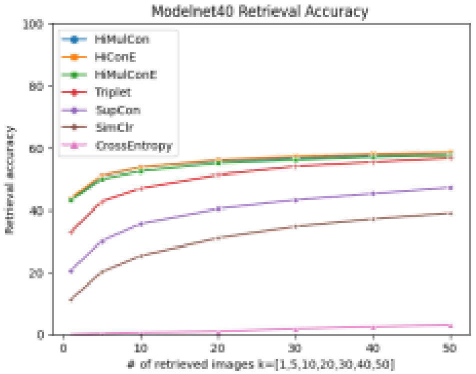


Table 4. Retrieval results using MAP evaluation metrics.

	DeepFashion	iNaturalist	ModelNet40
SupCon	31.5	61.5	21.6
HiMulConE	**35.6**	**66.9**	**26.0**

Experiments

Clustering quality

	DeepFashion		ModelNet40	
	Category NMI	Product NMI	Category NMI	Product NMI
SimCLR	0.15	0.73	0.31	0.52
CE	0.1	0.66	0.12	0.4
SupCon	0.57	0.68	0.57	0.69
HiMulCon	0.57	0.8	**0.62**	**0.88**
HiConE	0.58	0.78	0.61	**0.88**
HiMulConE	**0.59**	**0.81**	**0.62**	**0.88**

Table 3. NMI on DeepFashion In-Shop and ModelNet40. CE represents cross entropy. Product NMI represents its mean NMI.

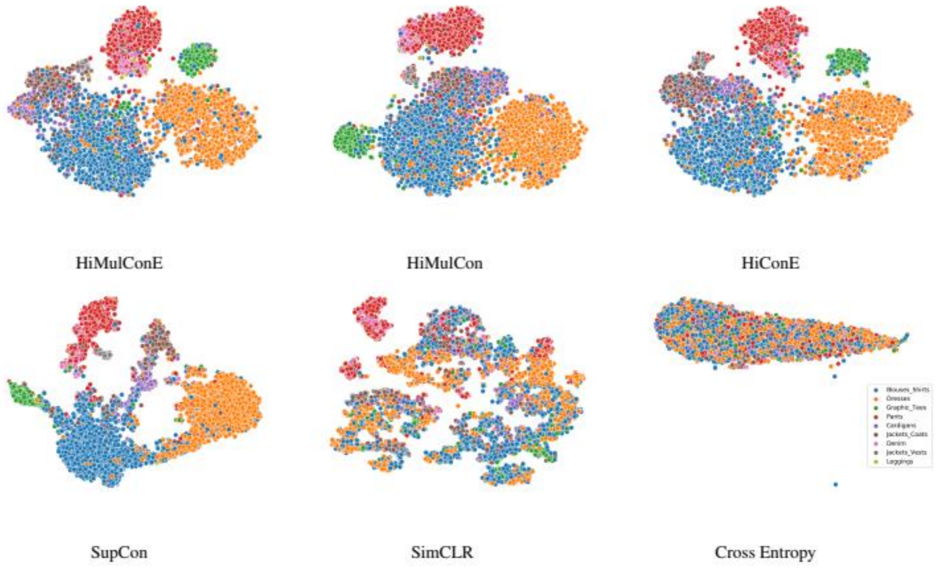


Figure 2. t-sne visualizations of the Deep Fashion dataset

Sampling strategy

Approach	Seen	Unseen
Hierarchical batch sampling	**80.52**	**75.29**
Category level sampling	79.81	72.63
Random Sampling	77.96	71.59

Table 1. Ablation study on effectiveness of the batch sampler

Conclusion

기여
- 계층구조를 가진 Multi-label에서 활용가능한 Framework를 제시
- 일반화 성능이 우수함

한계
- Encoder학습을 위해서는 계층 구조를 지닌 Multi-label을 데이터셋을 활용
- 각 노드는 오직 하나의 부모 노드를 가짐

End