

# Deep image clustering

## 1. Clustering은 어떨 때 필요한가?

---

우리가 일반적으로 알고있는 clustering은 classification 테스트에 가까움

일례로, ImageNet challenge의 경우 1000개의 label을 가지는 이미지를 classification하고, 정확도를 대결

하지만 label이 존재하는 데이터를 이용하는 classification의 경우, supervised learning의 한계를 가지고 있음

- label이 주어지지 않은 이미지는 어떻게 분류해야 할까?
- 좀 더 General 한 분류작업을 할 수 있지 않을까?
- label 데이터에 bias가 존재한다면?
  - e.g. 농구 라벨을 가지는 이미지에는 흑인이 많이 포함돼 있음 → 흑인을 농구로 분류 함

→ 이러한 한계는 unsupervised learning을 이용한 clustering이 해결해 줄 수 있음  
라벨(bias)을 제거하고, 직접 만들자!

## 2. Concept of Deep image clustering

---

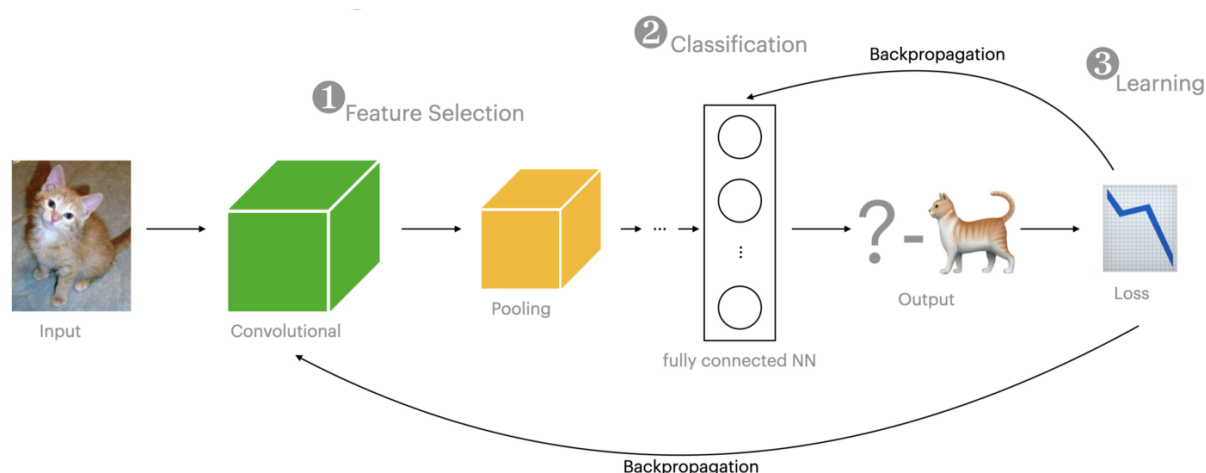
ECCV 2018 에서 발표된,

**“Deep Clustering for Unsupervised Learning of Visual Features”** 논문을 통해 Deep image clustering이 어떻게 이루어지는지 살펴보자!

Convolution network는 이미지 내 특정 속성, 정보들을 정량화 하여 표현할 수 있음

- 입력 이미지를 아무런 가공 없이 분류하는 것은 불가능
- 입력 이미지에서 무언가 공통된 특성을 가지는 패턴 혹은 정보들을 정량적으로 표현이 필요
- convnet은 이미지를 고정된 차원의 벡터 공간으로 맵핑해 주는 역할을 함 (feature vector)

## 일반적인 Image classifier



- 일반적인 Image classifier는 convnet을 통해 image feature vector를 추출하고, fully-connected NN을 통하여 입력 이미지가 특정 class에 속할 확률을 계산
- convnet의 각 convolution layer의 kernel 들은, 이미지 내 특정한 패턴을 추출해 주는 역할을 함

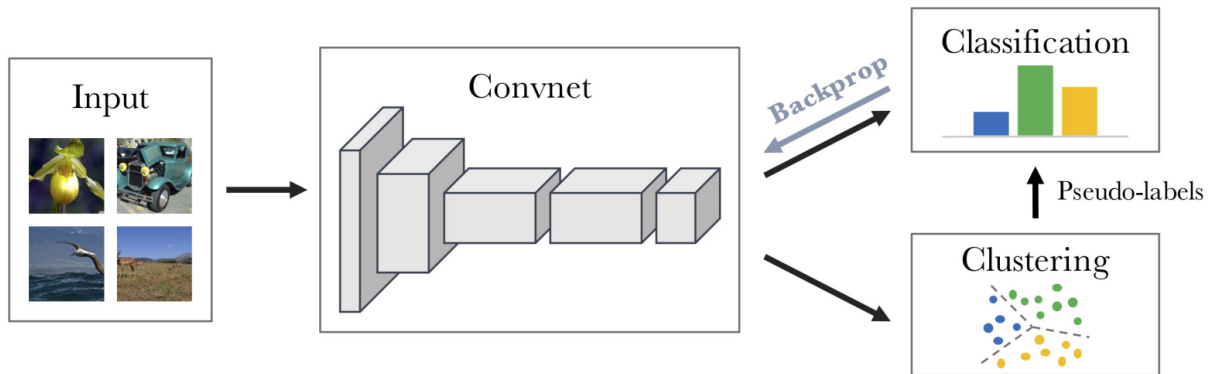
**classification을 위한 convnet의 학습은 어떻게 이루어 질까?**

$$\min_{\theta, W} \frac{1}{N} \sum_{n=1}^N \ell(g_W(f_{\theta}(x_n)), y_n)$$

- $x_n : X = \{x_1, x_2, \dots, x_N\}$ 의 원소로  $N$ 개의 이미지를 가진 데이터셋에 속함
- $y_n$  : 이미지의 label로 클래스가  $K$ 개 있다면 각 원소가  $[0, 1]$ 에 속하는  $K$ 차원 벡터
- $\ell$  : negative log-softmax로 cross-entropy와 동일함
- $f_{\theta}$  : 입력 이미지를  $d$ 차원 벡터로 맵핑.  $\theta$ 는 convnet의 파라미터를 뜻함
- $g_W$  :  $d$ 차원 벡터를  $K$ 차원 벡터로 맵핑.  $W$ 는 분류기의 파라미터

- 이미 주어진 label을 사용하여 학습을 하기 때문에 supervised learning 임

## 논문에서 제안하는 Image clustering 방법



### 과정

- Image classification과 마찬가지로 입력이미지를 convnet에 통과시켜 feature vector를 획득
- Convnet에 의해 생성된 feature vector들에 대해 K-means clustering을 수행하여 학습 이미지들을 k개의 pseudo-labeling 부여 및 학습

### K-means 를 사용하여, convnet 출력 결과들의 clustering

$$\min_{C \in \mathbb{R}^{d \times k}} \frac{1}{N} \sum_{n=1}^N \min_{y_n \in \{0,1\}^k} \|f_{\theta}(x_n) - Cy_n\|_2^2$$

- $C$  : centroid matrix로  $f_{\theta}(x_n)$ 의 차원이  $d$ 라면,  $d \times K$ 차원의 행렬.  $K$ 는 centroid의 수.
- $y_n$  : 구하고자 하는 pseudo-label
- Clustering을 통해  $C$ 를 구하고, 이를 통해  $y_n$ 을 구함

학습은 어떻게..?

- clustering 결과로 얻은 K개의 pseudo-label의 개수 분포와, 학습 데이터셋의 class 별 분포를 cross-entropy loss를 이용하여 일치시킴
- (의견) 입력 이미지 하나하나에 대해 소속 class에 대한 loss를 계산하는 것이 아닌, 전체 데이터의 분포를 GT와 일치시키는 것 → unsupervised learning?

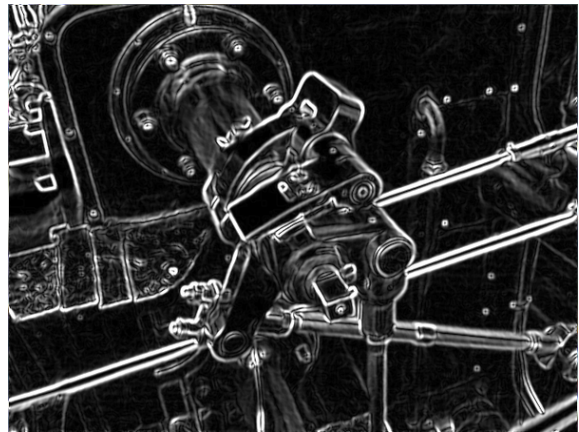
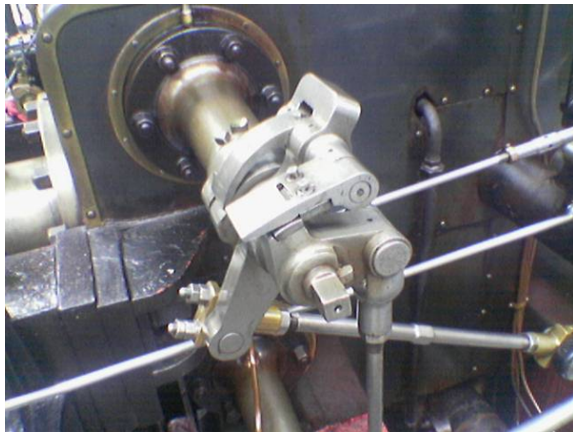
### 특징

- convnet의 weight들을 normal distribution에서 무작위로 추출해서 구성하였을 때, 12%의 clustering 성능을 보여줌
  - 즉 무작위한 kernel의 특성을 기반으로 나쁜 feature vector를 추출 했다는 것
- 학습 시간의 1/3 가량을 K-means clustering 연산에 사용

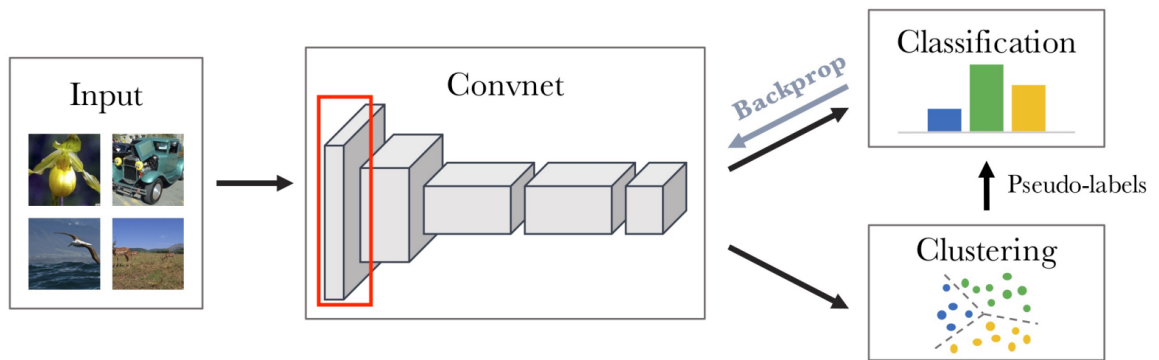
## 3. 꿀팁 of Deep image clustering

### Image preprocessing

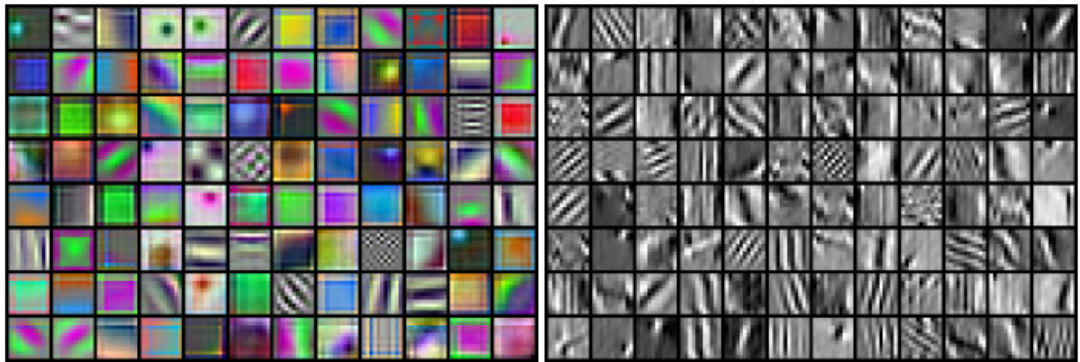
- Unsupervised learning을 이용한 분류 방법에서는 color 요소가 특징 추출에 방해가 된다고 함
  - 입력 이미지에 sobel filter를 적용하여 color를 제거하고 contrast를 향상시킴



- RGB 입력 vs Sobel filtered 입력



- 두 입력에 대해서 각각 Convnet을 학습 시켰을 때, 첫번째 layer의 conv kernel을 시각화 해보면,



RGB 입력에 대한 kernel(좌)은 패턴보다 색상 위주 정보에 대한 학습 결과를 알 수 있고,  
 Sobel filtered 입력에 대한 kernel(우)은 패턴에 치중된 정보들을 학습함을 알 수 있음

## 4. Performance evaluation

Rank	Model	Accuracy↑	NMI	ARI	Train set	Backbone	Extra Training Data	Paper	Code	Result	Year	Tags
1	<b>SPICE*</b>	0.918	0.850	0.836	Train	ResNet-18	×	<a href="#">SPICE: Semantic Pseudo-labeling for Image Clustering</a>			2021	
2	<b>RUC</b>	0.903				ResNet-18	✓	<a href="#">Improving Unsupervised Image Clustering With Robust Learning</a>			2020	
3	<b>IMC-SwAV (Best)</b>	0.897	0.818	0.8	Train	ResNet-18	×	<a href="#">Information Maximization Clustering via Multi-View Self-Labeling</a>			2021	
4	<b>IMC-SwAV (Avg+-)</b>	0.891	0.811	0.79	Train	ResNet-18	×	<a href="#">Information Maximization Clustering via Multi-View Self-Labeling</a>			2021	
5	<b>SCAN</b>	0.883	0.797	0.772	Train	ResNet-18	×	<a href="#">SCAN: Learning to Classify Images without Labels</a>			2020	
6	<b>SCAN (Avg)</b>	0.876	0.787	0.758	Train	ResNet-18	×	<a href="#">SCAN: Learning to Classify Images without Labels</a>			2020	
7	<b>Single-Noun Prior</b>	0.853	0.731	0.702	Train+Test	ViT-B-32	✓	<a href="#">The Single-Noun Prior for Image Clustering</a>			2021	
8	<b>ConCURL</b>	0.846	0.762	0.715	Train		×	<a href="#">Representation Learning for Clustering via Building Consensus</a>			2021	
9	<b>MMDC</b>	0.820	0.703			ResNet18	×	<a href="#">Multi-Modal Deep Clustering: Unsupervised Partitioning of Images</a>			2019	
10	<b>IDFD</b>	0.815	0.711	0.663	Train+Test	ResNet-18	✓	<a href="#">Clustering-friendly Representation Learning via Instance Discrimination and Feature Decorrelation</a>			2020	
11	<b>TSUC</b>	0.81	-	-	Train	ResNet-18	✓	<a href="#">Mitigating Embedding and Class Assignment Mismatch in Unsupervised Image Classification</a>			2020	
12	<b>CC</b>	0.79	0.705	0.637	Train+Test	ResNet34	×	<a href="#">Contrastive Clustering</a>			2020	
13	<b>DHOG</b>	0.666	0.585	0.492	Train+Test	ResNet-18	✓	<a href="#">DHOG: Deep Hierarchical Object Grouping</a>			2020	
14	<b>DCCM</b>	0.623	0.496	0.408	Train+Test	AlexNet	✓	<a href="#">Deep Comprehensive Correlation Mining for Image Clustering</a>			2019	
15	<b>IIC</b>	0.617	0.511	0.411	Train+Test	ResNet-34	✓	<a href="#">Invariant Information Clustering for Unsupervised Image Classification and Segmentation</a>			2018	
16	<b>DAC</b>	0.522	0.4	0.301	Train+Test	ConvNet	✓	<a href="#">Deep Adaptive Image Clustering</a>			2017	
17	<b>DeepCluster</b>	0.374	-	-	Train+Test	ResNet-34	✓	<a href="#">Deep Clustering for Unsupervised Learning of Visual Features</a>			2018	

## Reference

- Caron, M., Bojanowski, P., Joulin, A., & Douze, M. (2018). Deep clustering for unsupervised learning of visual features. In Proceedings of the European conference on computer vision (ECCV) (pp. 132-149).
- O'Mahony, N., Campbell, S., Carvalho, A., Harapanahalli, S., Hernandez, G. V., Krpalkova, L., ... & Walsh, J. (2019, April). Deep learning vs. traditional computer

vision. In *Science and information conference* (pp. 128-144). Springer, Cham.ISO 690