

Self-Supervised Learning

(Algorithm & application)

Seokho Moon

Nov 20, 2020





❖ 문석호 (Seokho Moon)

- 고려대학교 산업경영공학과 대학원 재학 중
- Data Mining & Quality Analytics Lab (김성범 교수님)
- 석사과정 (2019.09 ~)

❖ 관심 연구 분야

- Self-supervised learning
- Semi-supervised learning
- Anomaly detection

❖ E-mail

- danny232@korea.ac.kr

관련 분야를 공부할 때 큰 도움이 되었던 PR12 및 블로그 그리고
DMQA 연구실의 강현구 연구원님에게 감사 인사를 전합니다.

<https://lilianweng.github.io/lil-log/2019/11/10/self-supervised-learning.html>

<https://hoya012.github.io/blog/Self-Supervised-Learning-Overview/>

PR-208: Unsupervised Visual Representation Learning Overview: Toward Self-Supervision by Hoseong Lee

PR-260: Momentum Contrast for Unsupervised Visual Representation Learning by Taeh Kim

<https://2-chaе.github.io/category/2.papers/20>



Contents

1. Introduction

- Background
- Research timeline

2. Algorithm paper

- NPID, CVPR 2018
- MoCo, CVPR 2020
- PIRL, CVPR 2020

3. Application paper

- WaPIRL, IEEE T SEMICONDUCT M 2020 (preprint)
- Subject-Aware Contrastive Learning for Biosignals, CoRR 2020 (preprint)

4. Conclusion

- Comments

Introduction

Background

딥러닝 계열 모델 특성상 잘 학습시키기 위해서는
많은 양의 labeling 작업이 된 데이터가 필요!

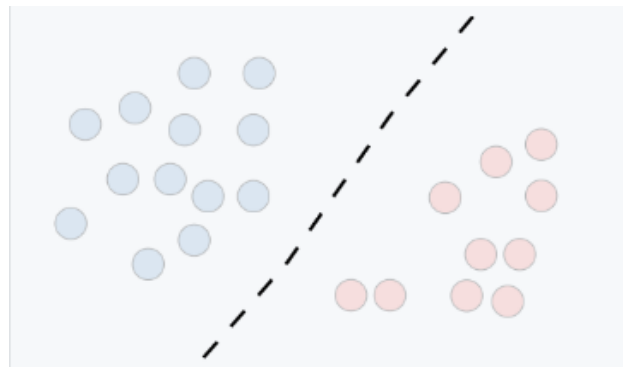
Introduction

Background

딥러닝 계열 모델 특성상 잘 학습시키기 위해서는
많은 양의 labeling 작업이 된 데이터가 필요!

Discriminative model

→ Input과 output의 관계(함수)를 알아내기 위해서



→ Cat

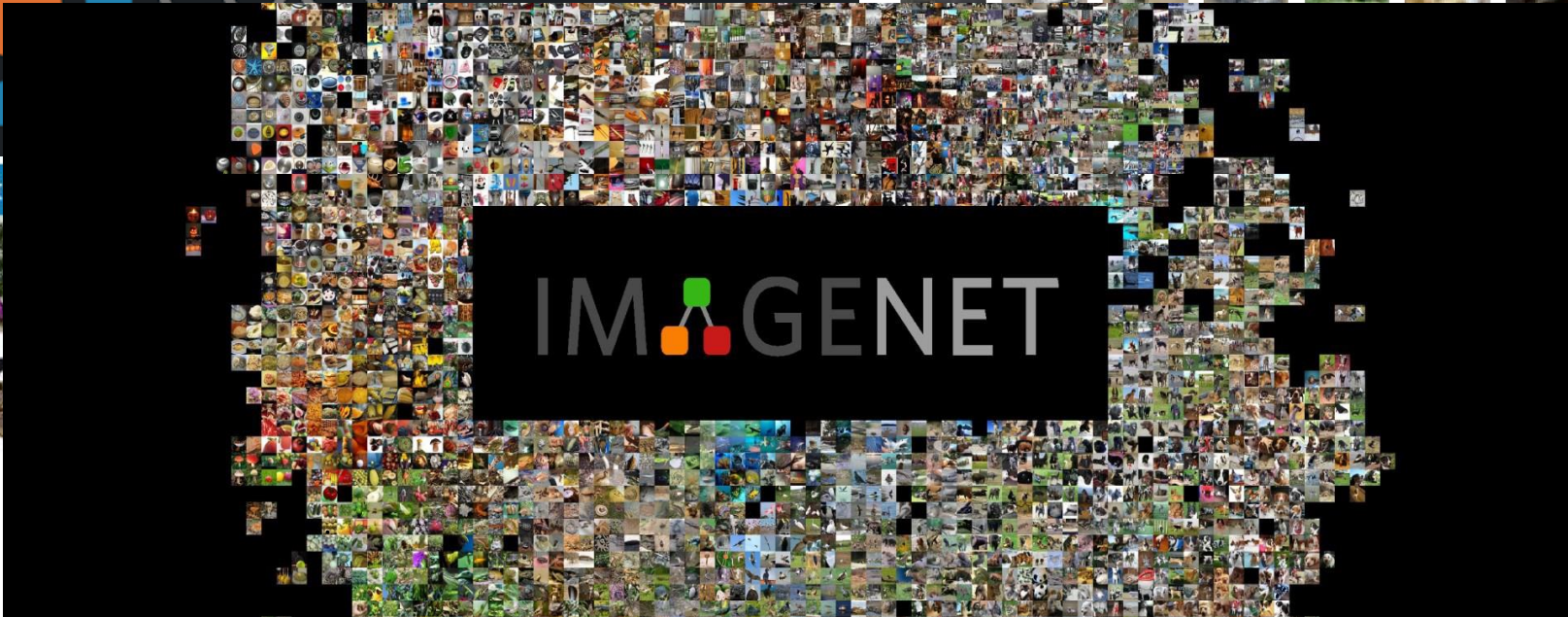
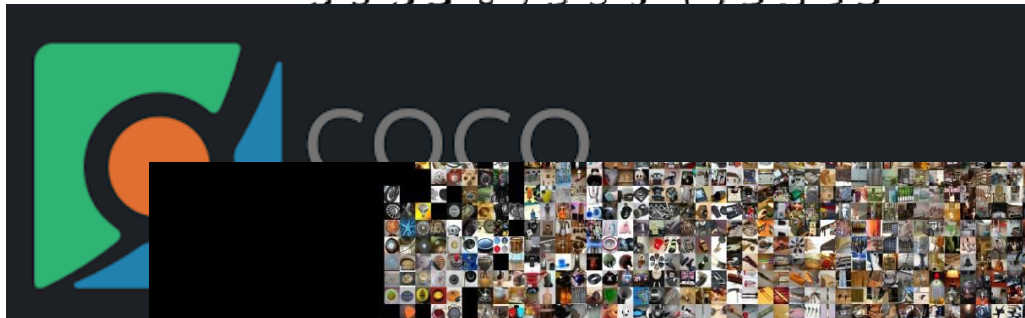
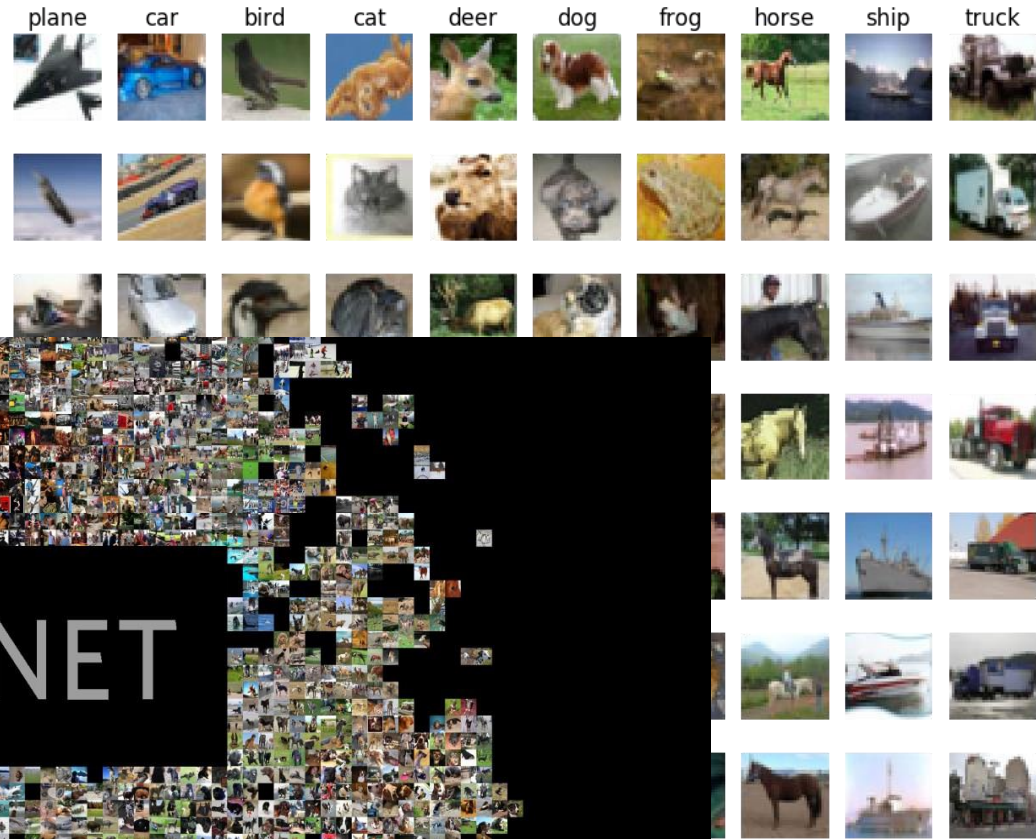


→ Dog

Introduction

Background

0000000000000000
1111111111111111
2222222222222222
3333333333333333



ImageNet

Introduction

Background

Labeling 작업에는 많은 노력과 비용이 요구됨

Labeling 작업에는 많은 노력과 비용이 요구됨



이러한 문제에 도전하는 방법들

Transfer Learning

Domain Adaptation

Semi-Supervised learning

Weakly-supervised learning

Self-supervised learning

Labeling 작업에는 많은 노력과 비용이 요구됨



이러한 문제에 도전하는 방법들

Transfer Learning

Domain Adaptation

Semi-Supervised learning

Weakly-supervised learning

Self-supervised learning

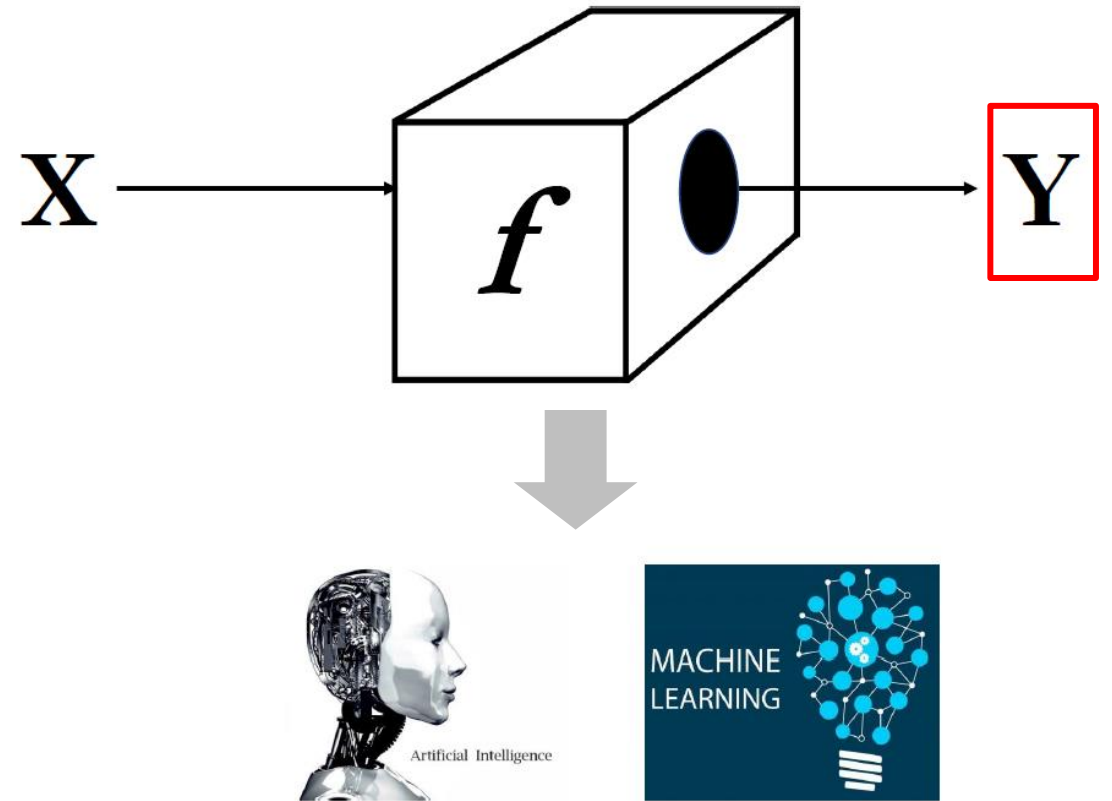


Unsupervised learning 방법론

Introduction

Background

Supervision(지도, 감독)이란?



Introduction

Background

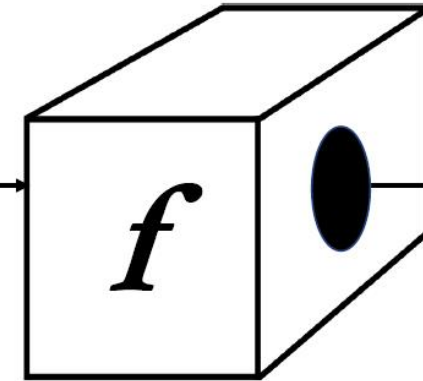
Supervision(지도, 감독)이란?



우리가 흔히 아는 label이라고 생각하면 좋음

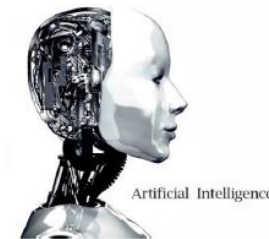


X



Cat

Y



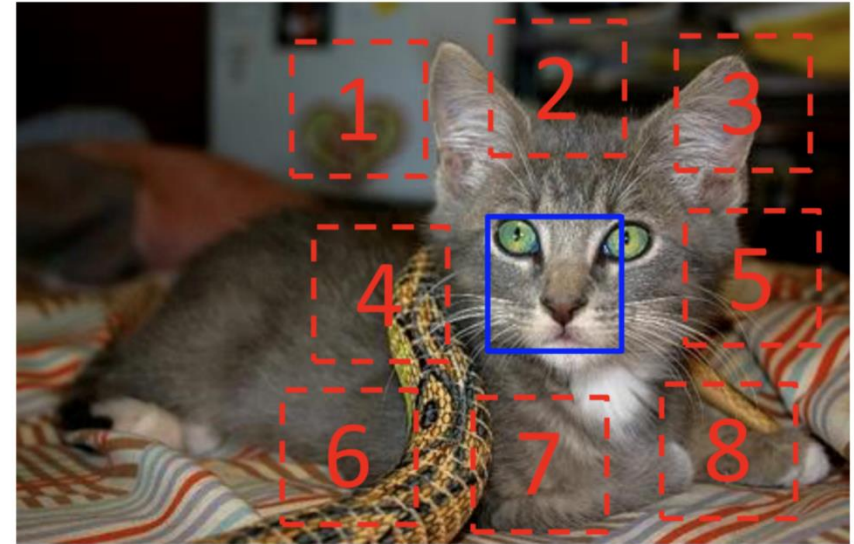
Introduction

Background

Label이 없는 데이터 상황에서
Self-supervision 이란?



Image의 feature를 추출하는 모델이 있다고 할 때
Input 데이터의 한 부분이 다른 부분의 supervision
역할을 하게 하면 이것이 self-supervision 이다



$$X = \left(\begin{array}{c} \text{[Kitten Face]} \\ \text{[Kitten Ear]} \end{array} \right); Y = 3$$

Introduction

Research timeline

❖ Pretext task (2014 ~ 2018)

- Pretext task를 잘 정의해서 주어진 입력 이미지들에 대한 정보를 잘 추출하는 방식
- Exemplar – 2014 NIPS, context prediction – 2015 ICCV, jigsaw puzzle – 2016 ECCV, count – 2017 ICCV, rotation – 2018 ICLR ...

❖ Contrastive learning (2018 ~)

- Contrastive learning을 활용하여 주어진 입력 이미지들에 대한 정보를 추출하는 방식
- NPID – 2018 CVPR, CPC – 2018 Arxiv, MoCo – 2020 CVPR, SimCLR – 2020 ICML ...

❖ New approach (2020 ~)

- Network 구조나 clustering 개념을 도입하여 주어진 입력 이미지들에 대한 정보를 추출하는 방식
- BYOL – 2020 NIPS, PCL – 2020 ArXiv ...

Introduction

Research timeline

❖ Pretext task (2014 ~ 2018)

- Pretext task를 잘 정의해서 주어진 입력 이미지들에 대한 정보를 잘 추출하는 방식
- Exemplar – 2014 NIPS, context prediction – 2015 ICCV, jigsaw puzzle – 2016 ECCV, count – 2017 ICCV, rotation – 2018 ICLR ...

❖ Contrastive learning (2018 ~)

- Contrastive learning을 활용하여 주어진 입력 이미지들에 대한 정보를 추출하는 방식
- NPID – 2018 CVPR, CPC – 2018 Arxiv, MoCo – 2020 CVPR, SimCLR – 2020 ICML ...

❖ New approach (2020 ~)

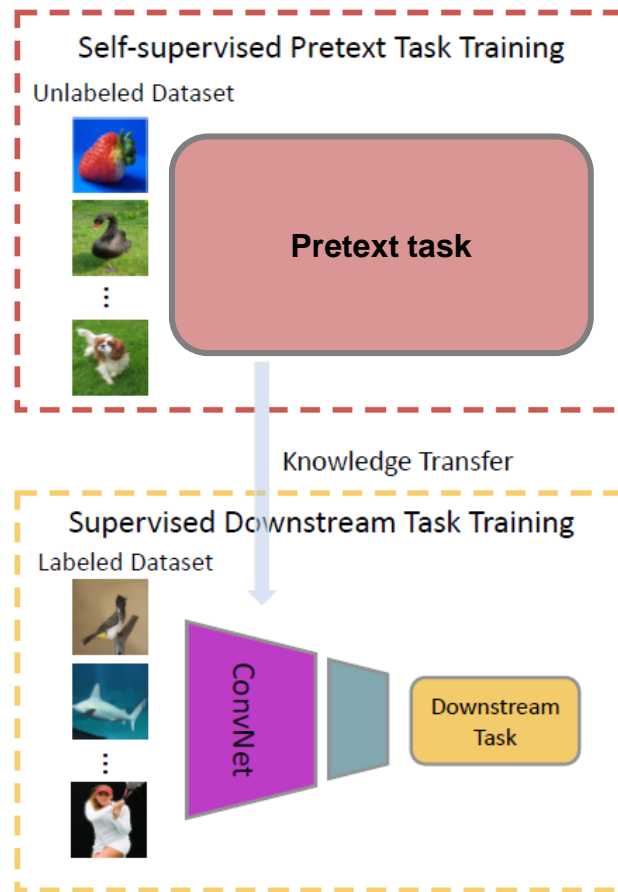
- Network 구조나 clustering 개념을 도입하여 주어진 입력 이미지들에 대한 정보를 추출하는 방식
- BYOL – 2020 NIPS, PCL – 2020 ArXiv ...

Introduction

Research timeline

❖ Pretext task (2014 ~ 2018)

- Pretext task를 잘 정의해서 주어진 입력 이미지들에 대한 정보를 잘 추출하는 방식



Unlabeled dataset를 input으로 받아 사용자가 정의한 문제(pretext task)를 network가 학습하게 하여 데이터 자체에 대한 이해도를 높이고자 함

↑

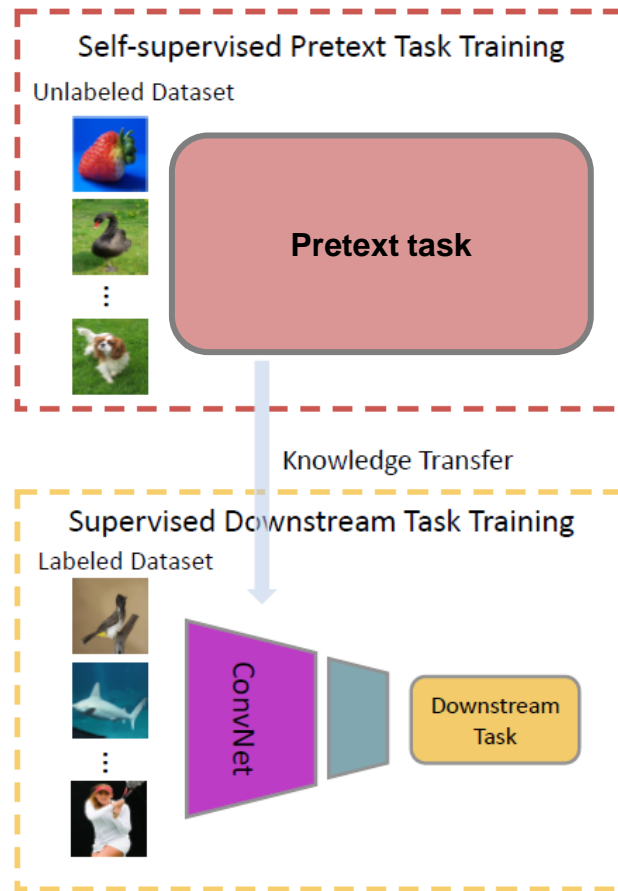
Pretext task가 잘 짜여졌다면 input을 효과적으로 representation할거라는 가정

Introduction

Research timeline

❖ Pretext task (2014 ~ 2018)

- Pretext task를 잘 정의해서 주어진 입력 이미지들에 대한 정보를 잘 추출하는 방식



Pretext task에서 나온 pre-training되어진 network를 적극적으로 사용자가 풀고자 하는 문제(downstream task, e.g., classification)에 transfer learning하는 방법

Introduction

Research timeline

❖ Pretext task (2014 ~ 2018)

- Pretext task를 잘 정의해서 주어진 입력 이미지들에 대한 정보를 잘 추출하는 방식

Exemplar, 2014 NIPS

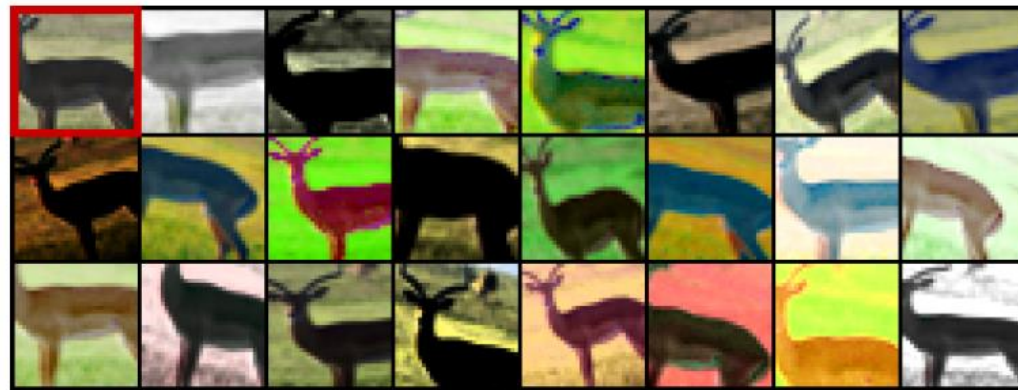


Fig. 2. Several random transformations applied to one of the patches extracted from the STL unlabeled dataset. The original ('seed') patch is in the top left corner.

Rotation, 2018 ICLR

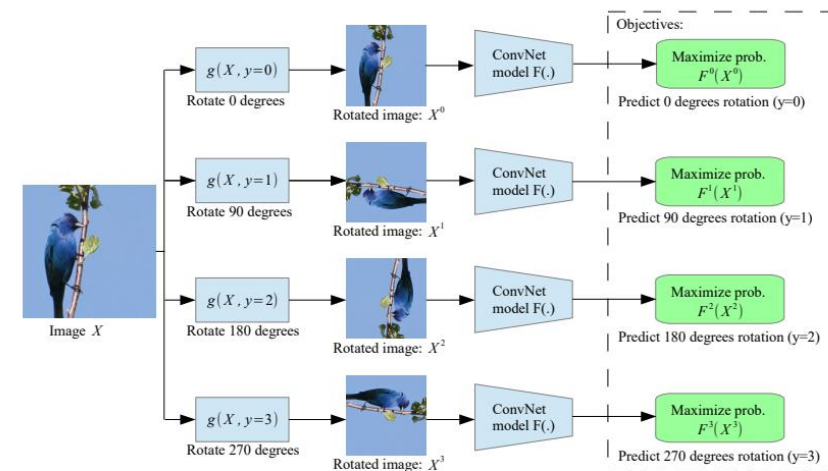


Figure 2: Illustration of the self-supervised task that we propose for semantic feature learning. Given four possible geometric transformations, the 0, 90, 180, and 270 degrees rotations, we train a ConvNet model $F(\cdot)$ to recognize the rotation that is applied to the image that it gets as input. $F^y(X^{y^*})$ is the probability of rotation transformation y predicted by model $F(\cdot)$ when it gets as input an image that has been transformed by the rotation transformation y^* .

→ 각 이미지마다 학습이 진행되므로 데이터셋이 커질수록 연산량이 급속도로 증가하여 성능 향상이 어려운 단점 존재

Introduction

Research timeline

❖ Pretext task (2014 ~ 2018)


- 더 자세한 설명은 기존 세미나 자료를 참고해주시면 감사하겠습니다.




종료

Self-Supervised Representation Learning

Seokho Moon
May 1, 2020

Self-Supervised Representation Learning

발표자:  문석호


 2020년 5월 1일
 오후 1시 ~
 화상 프로그램 이용(Zoom)

[세미나 정보 보기 →](#)




Self-Supervised Representation Learning

2020년 4월 27일 오후 5:56 / 조회수: 2910

REFERENCES

 200501_Open_Seminar_Self_Supervised_Representation_Learning_Seokho.pdf

INFORMATION

 2020년 5월 1일  오후 1시 ~  화상 프로그램 이용(Zoom)

발표자:  문석호

TOPIC

Self-Supervised Representation Learning

OVERVIEW

최근까지 supervised learning 방법을 통해 다양한 domain에서 neural network를 이용한 모델들이 우수한 성능을 보여주고 있는 상황이다. 다만, neural network 기반의 모델들은 label이 정확하게 있는 학습 데이터가 많이 필요하다는 문제점이 있다. 따라서 label이 없거나 클래스가 불균형 문제가 있는 데이터의 경우 학습이 어렵게 된다. 또한 현실에서는 비용 등의 문제로 label이 없는 데이터가 더 많다. 이러한 문제점을 극복하기 위해 데이터 간의 특성을 representation을 이용하는 self supervised learning 방법론을 사용한다. 이번 세미나에서는 위 방법에 관해 대표적인 몇 가지를 소개하고자 한다.

Introduction

Research timeline

❖ Pretext task (2014 ~ 2018)

- Pretext task를 잘 정의해서 주어진 입력 이미지들에 대한 정보를 잘 추출하는 방식
- Exemplar – 2014 NIPS, context prediction – 2015 ICCV, jigsaw puzzle – 2016 ECCV, count – 2017 ICCV, rotation – 2018 ICLR ...

❖ Contrastive learning (2018 ~)

- Contrastive learning을 활용하여 주어진 입력 이미지들에 대한 정보를 추출하는 방식
- NPID – 2018 CVPR, CPC – 2018 Arxiv, MoCo – 2020 CVPR, SimCLR – 2020 ICML ...

❖ New approach (2020 ~)

- Network 구조나 clustering 개념을 도입하여 주어진 입력 이미지들에 대한 정보를 추출하는 방식
- BYOL – 2020 NIPS, PCL – 2020 ArXiv ...

Contents

1. Introduction

- Background
- Research timeline

2. Algorithm paper

- NPID, CVPR 2018
- MoCo, CVPR 2020
- PIRL, CVPR 2020

3. Application paper

- WaPIRL, IEEE T SEMICONDUCT M 2020 (preprint)
- Subject-Aware Contrastive Learning for Biosignals, CoRR 2020 (preprint)

4. Conclusion

- Comments

Algorithm paper

NPID, CVPR 2018 (Unsupervised Feature Learning via Non-Parametric Instance Discrimination)

❖ NPID, CVPR 2018

- 2020년 11월 20일 기준 289회 인용
- 세계적인 컨퍼런스 중에 하나인 CVPR(IEEE Conference on Computer Vision and Pattern Recognition)에서 발표된 논문

Unsupervised Feature Learning via Non-Parametric Instance Discrimination

Zhirong Wu^{*†}
^{*}UC Berkeley / ICSI

Yuanjun Xiong^{†‡} Stella X. Yu^{*}
[†]Chinese University of Hong Kong

Dahua Lin[†]
[‡]Amazon Rekognition

Abstract

Neural net classifiers trained on data with annotated class labels can also capture apparent visual similarity among categories without being directed to do so. We study whether this observation can be extended beyond the conventional domain of supervised learning: Can we learn a good feature representation that captures apparent similarity among instances, instead of classes, by merely asking the feature to be discriminative of individual instances?

We formulate this intuition as a non-parametric classification problem at the instance-level, and use noise-contrastive estimation to tackle the computational challenges imposed by the large number of instance classes.

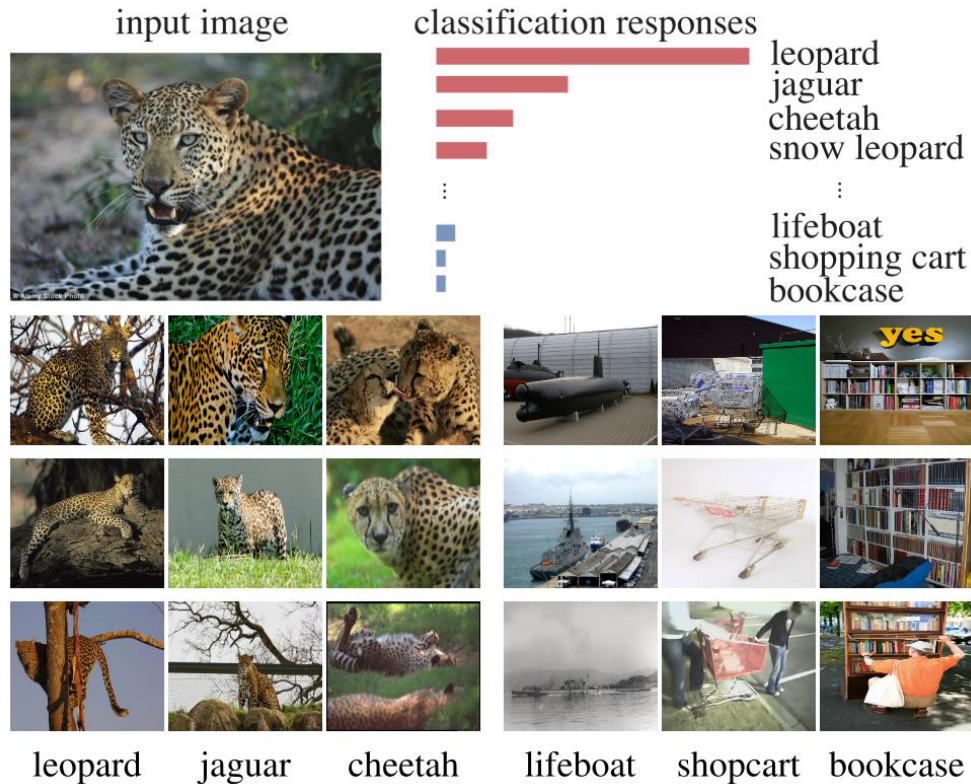
Our experimental results demonstrate that, under unsupervised learning settings, our method surpasses the state-of-the-art on ImageNet classification by a large margin. Our method is also remarkable for consistently improving test performance with more training data and better network architectures. By fine-tuning the learned feature, we further obtain competitive results for semi-supervised learning and object detection tasks. Our non-parametric model is highly compact: With 128 features per image, our method requires only 600MB storage for a million images, enabling fast nearest neighbour retrieval at the run time.

Algorithm paper

NPID, CVPR 2018 (Unsupervised Feature Learning via Non-Parametric Instance Discrimination)

❖ Motivation

- 지도학습 기반의 이미지 분류 모델 결과, 비슷한 이미지일 때 확률값이 높은 것을 보고 동기를 얻음



[지도학습 기반 이미지 분류 모델 결과]

지도학습을 통해 **input image**와 **label** 쌍으로 이루어진 데이터셋을 학습시킨 후

input image를 표범(leopard)으로 하여 나오는 모델의 **output**을 보면

재규어(jaguar)나 치타(cheetah)로 예측하는 확률값이 다른 클래스

(e.g., lifeboat, shopping cart)보다 높음



이를 통해 모델이 따로 **instances** 간의 구분(유사성)을 학습하지 않아도,

잘 추출된 특징값은 instances 간의 유사도 정보를 가지고 있을 것이라 생각

Algorithm paper

NPID, CVPR 2018 (Unsupervised Feature Learning via Non-Parametric Instance Discrimination)

❖ Framework

- Instances 간의 유사도를 contrastive learning 방식으로 학습하여 모든 입력 이미지를 128차원의 특징벡터로 변환
- 이 특징벡터 space는 모든 이미지가 들어오므로 기존의 이미지(class로 간주)가 많아지면서 생긴 연산 문제 또한 해결

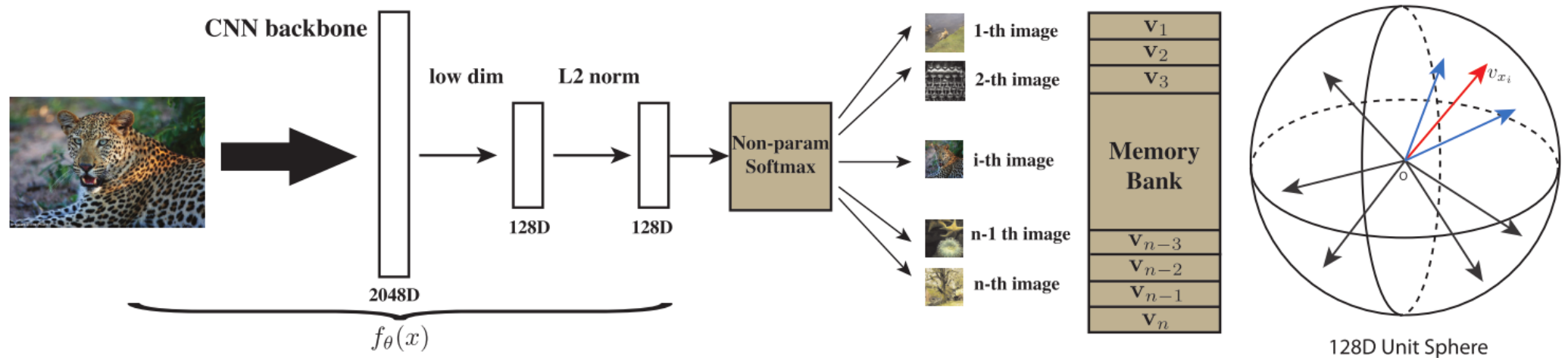


Figure 2: The pipeline of our unsupervised feature learning approach. We use a backbone CNN to encode each image as a feature vector, which is projected to a 128-dimensional space and L2 normalized. The optimal feature embedding is learned via instance-level discrimination, which tries to maximally scatter the features of training samples over the 128-dimensional unit sphere.

Algorithm paper

NPID, CVPR 2018 (Unsupervised Feature Learning via Non-Parametric Instance Discrimination)

❖ Method → Non-Parametric Classifier

- n개의 image : n개의 class를 가지는 x_1, x_2, \dots, x_n 로 표현
- 인코더 f_θ 에서 나온 이미지 x_i 의 특징 벡터 : $f_\theta(x_i) = v_i$
- w_j 는 일반적으로 class j를 구별하기 위해 학습되는 parameter

- Parametric classifier →
$$P(i|\mathbf{v}) = \frac{\exp(\mathbf{w}_i^T \mathbf{v})}{\sum_{j=1}^n \exp(\mathbf{w}_j^T \mathbf{v})}$$

128차원의 특징 벡터에 softmax 개념의 classifier를 적용하게끔 변경

- Non-parametric classifier →
$$P(i|\mathbf{v}) = \frac{\exp(\mathbf{v}_i^T \mathbf{v} / \tau)}{\sum_{j=1}^n \exp(\mathbf{v}_j^T \mathbf{v} / \tau)}$$

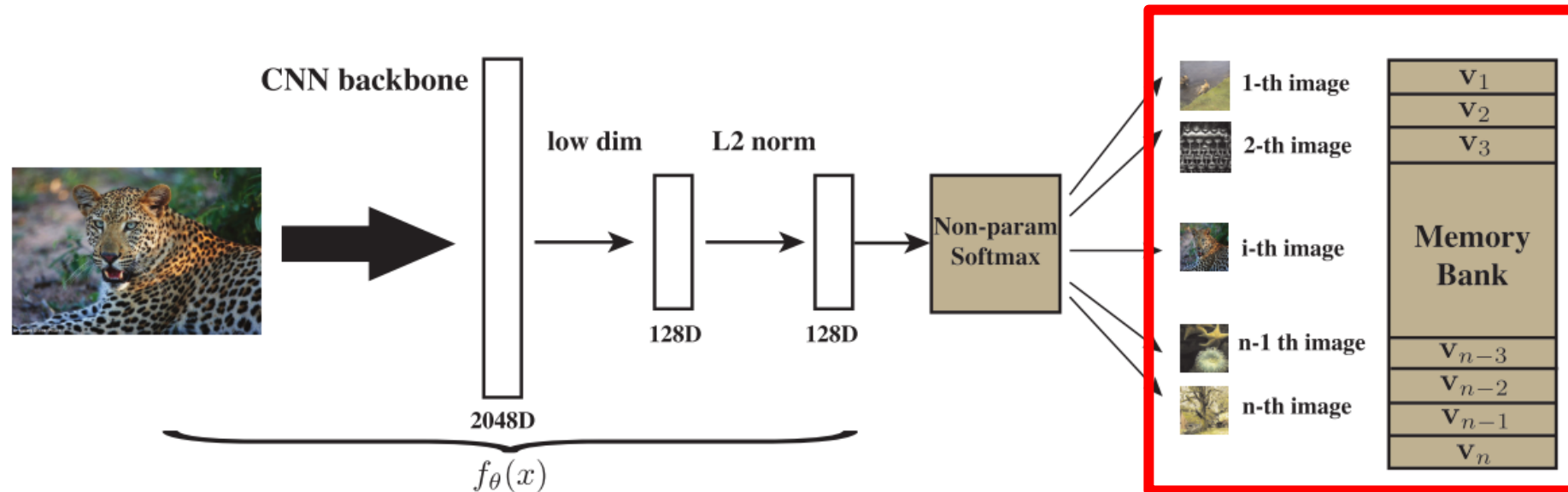
→ 기존 학습 데이터에만 의존하지 않게 하고 instances 간의 비교를 수월하게 함

Algorithm paper

NPID, CVPR 2018 (Unsupervised Feature Learning via Non-Parametric Instance Discrimination)

❖ Method → Memory Bank

- 인코더 f_θ 에서 나온 이미지 x_i 의 특징 벡터 v_i 를 전부 계산해야 하는데 이미지 개수가 증가하면 메모리 문제가 발생
- 이를 해결하기 위해 memory bank V 를 도입하여 계산된 특징벡터를 저장하여 사용
- Learning iteration(epoch)이 진행될 때 θ 와 동시에 업데이트



Algorithm paper

NPID, CVPR 2018 (Unsupervised Feature Learning via Non-Parametric Instance Discrimination)

❖ Method → NCE(Noise-Contrastive Estimation)

- NCE 방식은 주로 multi class 분류 문제에서 class가 너무 많아질 경우, data sample과 noise sample을 구분하는 이진분류 문제로 변환하여 주는 것
- 이 방법을 적용하는 이유는 n의 개수가 많아질 경우 계산이 너무 복잡해지기 때문

$$P(i|\mathbf{v}) = \frac{\exp(\mathbf{v}_i^T \mathbf{v} / \tau)}{\sum_{j=1}^n \exp(\mathbf{v}_j^T \mathbf{v} / \tau)} \quad \longrightarrow \quad \begin{aligned} P(i|\mathbf{v}) &= \frac{\exp(\mathbf{v}^T \mathbf{f}_i / \tau)}{Z_i} \\ Z_i &= \sum_{j=1}^n \exp(\mathbf{v}_j^T \mathbf{f}_i / \tau) \end{aligned}$$

- 최종적으로 나온 확률을 통해 negative log-posterior distribution을 minimize하는 목적 함수를 사용

$$h(i, \mathbf{v}) := P(D = 1 | i, \mathbf{v}) = \frac{P(i|\mathbf{v})}{P(i|\mathbf{v}) + mP_n(i)} \quad \longrightarrow \quad \begin{aligned} J_{NCE}(\boldsymbol{\theta}) &= -E_{P_d} [\log h(i, \mathbf{v})] \\ &\quad - m \cdot E_{P_n} [\log(1 - h(i, \mathbf{v}'))] \end{aligned}$$

Algorithm paper

NPID, CVPR 2018 (Unsupervised Feature Learning via Non-Parametric Instance Discrimination)

❖ Experiment & result

- 기존 pretext task 기반의 방법론들과 비교하였을 때 우수한 성능 확인

Image Classification Accuracy on ImageNet							
method	conv1	conv2	conv3	conv4	conv5	kNN	#dim
Random	11.6	17.1	16.9	16.3	14.1	3.5	10K
Data-Init [16]	17.5	23.0	24.5	23.2	20.6	-	10K
Context [2]	16.2	23.3	30.2	31.7	29.6	-	10K
Adversarial [4]	17.7	24.5	31.0	29.9	28.0	-	10K
Color [47]	13.1	24.8	31.0	32.6	31.8	-	10K
Jigsaw [27]	19.2	30.1	34.7	33.9	28.3	-	10K
Count [28]	18.0	30.6	34.3	32.5	25.7	-	10K
SplitBrain [48]	17.7	29.3	35.4	35.2	32.8	11.8	10K
Exemplar[3]			31.5			-	4.5K
Ours Alexnet	16.8	26.5	31.8	34.1	35.6	31.3	128
Ours VGG16	16.5	21.4	27.6	33.1	37.2	33.9	128
Ours Resnet18	16.0	19.9	26.3	35.7	42.1	40.5	128
Ours Resnet50	15.3	18.8	24.4	35.3	43.9	42.5	128

Table 2: Top-1 classification accuracies on ImageNet.

Image Classification Accuracy on Places							
method	conv1	conv2	conv3	conv4	conv5	kNN	#dim
Random	15.7	20.3	19.8	19.1	17.5	3.9	10K
Data-Init [16]	21.4	26.2	27.1	26.1	24.0	-	10K
Context [2]	19.7	26.7	31.9	32.7	30.9	-	10K
Adversarial [4]	17.7	24.5	31.0	29.9	28.0	-	10K
Video [44]	20.1	28.5	29.9	29.7	27.9	-	10K
Color [47]	22.0	28.7	31.8	31.3	29.7	-	10K
Jigsaw [27]	23.0	32.1	35.5	34.8	31.3	-	10K
SplitBrain [48]	21.3	30.7	34.0	34.1	32.5	10.8	10K
Ours Alexnet	18.8	24.3	31.9	34.5	33.6	30.1	128
Ours VGG16	17.6	23.1	29.5	33.8	36.3	32.8	128
Ours Resnet18	17.8	23.0	30.3	34.2	41.3	36.7	128
Ours Resnet50	18.1	22.3	29.7	34.1	42.1	38.7	128

Table 3: Top-1 classification accuracies on Places, based directly on features learned on ImageNet, without any fine-tuning.

Algorithm paper

MoCo, CVPR 2020 (Momentum Contrast for Unsupervised Visual Representation Learning)

❖ MoCo, CVPR 2020

- 2020년 11월 20일 기준 380회 인용
- 세계적인 컨퍼런스 중에 하나인 CVPR(IEEE Conference on Computer Vision and Pattern Recognition)에서 발표된 논문

Momentum Contrast for Unsupervised Visual Representation Learning

Kaiming He Haoqi Fan Yuxin Wu Saining Xie Ross Girshick

Facebook AI Research (FAIR)

Code: <https://github.com/facebookresearch/moco>

Abstract

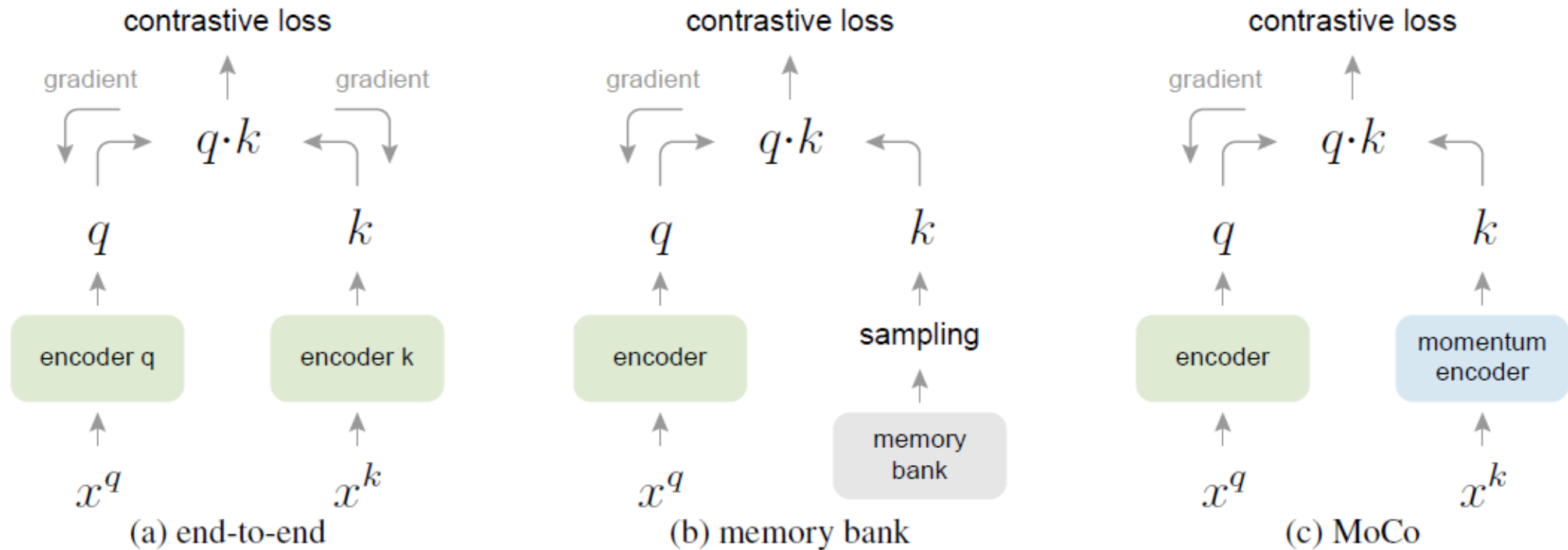
*We present Momentum Contrast (MoCo) for unsupervised visual representation learning. From a perspective on contrastive learning [29] as dictionary look-up, we build a dynamic dictionary with a queue and a moving-averaged encoder. This enables building a large and consistent dictionary on-the-fly that facilitates contrastive unsupervised learning. MoCo provides competitive results under the common linear protocol on ImageNet classification. More importantly, the representations learned by MoCo transfer well to downstream tasks. MoCo can **outperform** its supervised pre-training counterpart in 7 detection/segmentation tasks on PASCAL VOC, COCO, and other datasets, sometimes surpassing it by large margins. This suggests that the gap between unsupervised and supervised representation learning has been largely closed in many vision tasks.*

Algorithm paper

MoCo, CVPR 2020 (Momentum Contrast for Unsupervised Visual Representation Learning)

❖ Contribution

- 큰 dictionary를 통해 많은 negative sample의 수를 확보하여 성능 향상
- Key들의 consistency를 최대한 맞춰주어 학습이 안정적으로 진행되도록 함

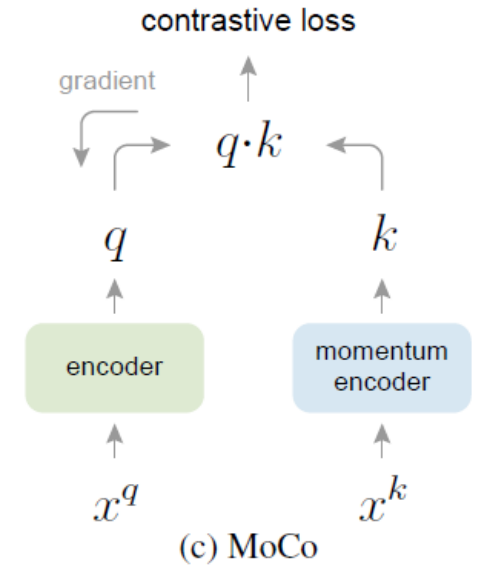


Algorithm paper

MoCo, CVPR 2020 (Momentum Contrast for Unsupervised Visual Representation Learning)

❖ Method – contrastive learning

- NPID에서 사용한 것처럼 infoNCE 개념의 loss function을 사용
- q 는 인코딩을 통과한 query이고 k 는 dictionary에 있는 key
- 여기서 k_+ 가 query의 positive sample의 개념으로 볼 수 있음
- 직관적으로 볼 때, k_+ 을 분류하는 $K+1$ 개의 softmax과 유사한 classifier라고 저자들은 말함



$$\mathcal{L}_q = -\log \frac{\exp(q \cdot k_+ / \tau)}{\sum_{i=0}^K \exp(q \cdot k_i / \tau)}$$

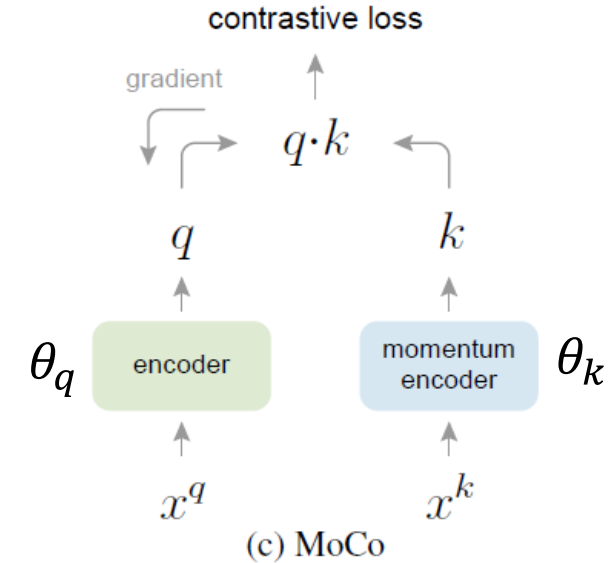
Algorithm paper

MoCo, CVPR 2020 (Momentum Contrast for Unsupervised Visual Representation Learning)

❖ Method – Momentum contrast

- Dictionary as a queue
 1. 여기서 queue 개념을 도입하여 기존의 memory bank와의 차별점을 둠
 2. Batch size에 크게 의존하지 않게 dictionary의 크기를 구성할 수 있음
 3. Queue 구조 상 가장 처음 들어온 특징벡터가 가장 먼저 나가게 되면서 consistency가 유지
- Momentum update
 1. Queue에 있는 key의 특징벡터가 consistent하려면 기존 key가 가지는 parameter θ_k 를 많이 가져가야 함
 2. 본 논문에서는 실험적으로 $m=0.999$ 일 때 가장 학습이 안정적이고 우수한 성능을 보여줌

$$\theta_k \leftarrow m\theta_k + (1 - m)\theta_q$$

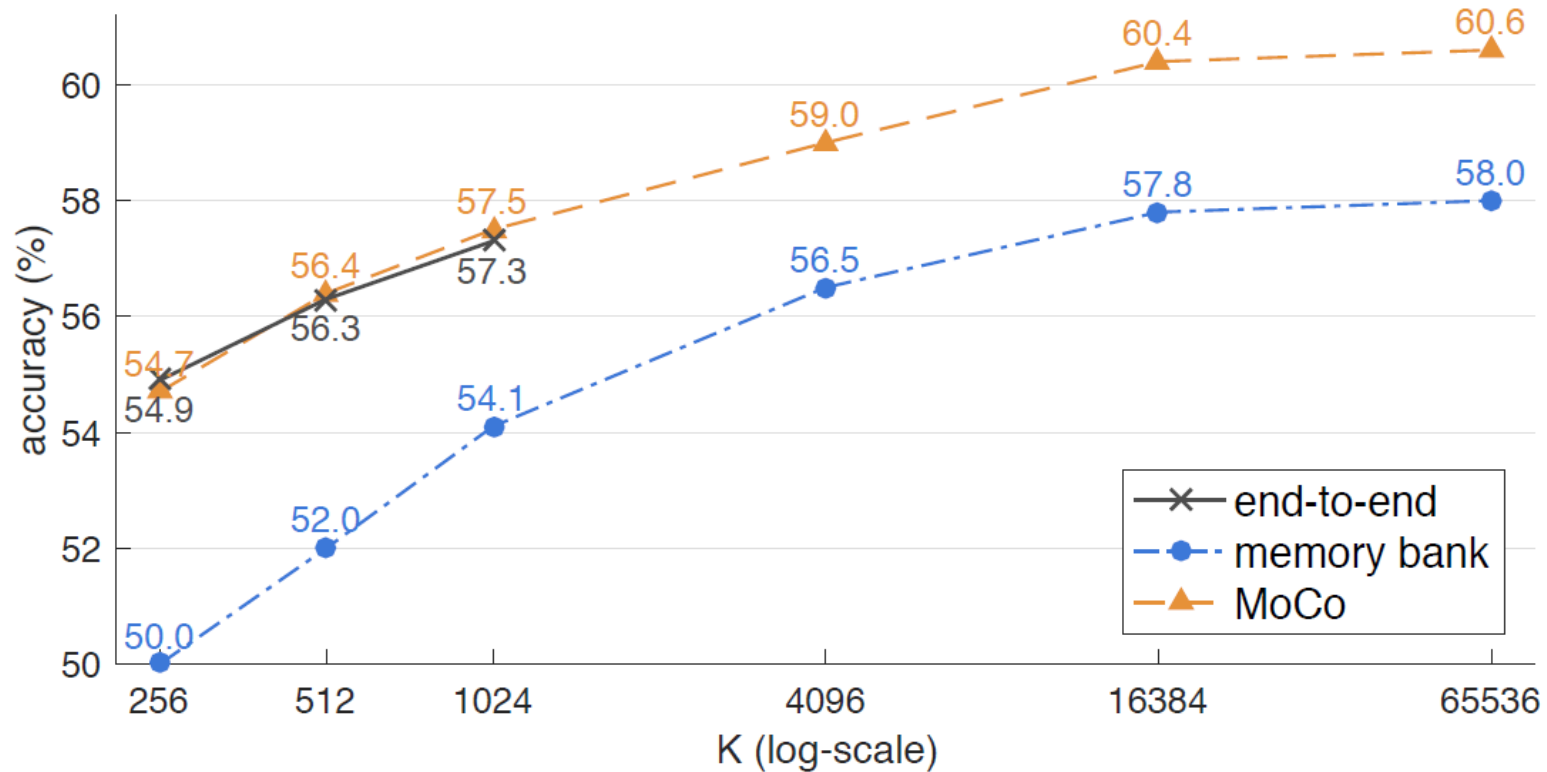


Algorithm paper

MoCo, CVPR 2020 (Momentum Contrast for Unsupervised Visual Representation Learning)

❖ Experiment & result

- K(negative sample)의 크기를 증가시킬수록 성능이 향상됨을 확인
- End-to-end 모델은 batch size에 의존하기 때문에 일정 사이즈 이상 커지지 못함



Algorithm paper

PIRL, CVPR 2020 (Self-Supervised Learning of Pretext-Invariant Representations)

❖ PIRL, CVPR 2020

- 2020년 11월 20일 기준 107회 인용
- 세계적인 컨퍼런스 중에 하나인 CVPR(IEEE Conference on Computer Vision and Pattern Recognition)에서 발표된 논문

Self-Supervised Learning of Pretext-Invariant Representations

Ishan Misra Laurens van der Maaten
Facebook AI Research

Abstract

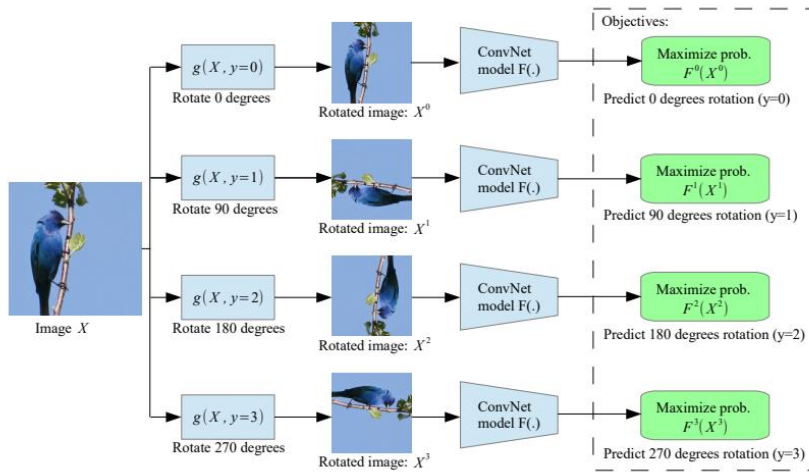
*The goal of self-supervised learning from images is to construct image representations that are semantically meaningful via pretext tasks that do not require semantic annotations. Many pretext tasks lead to representations that are covariant with image transformations. We argue that, instead, semantic representations ought to be invariant under such transformations. Specifically, we develop Pretext-Invariant Representation Learning (PIRL, pronounced as “pearl”) that learns invariant representations based on pretext tasks. We use PIRL with a commonly used pretext task that involves solving jigsaw puzzles. We find that PIRL substantially improves the semantic quality of the learned image representations. Our approach sets a new state-of-the-art in self-supervised learning from images on several popular benchmarks for self-supervised learning. Despite being unsupervised, **PIRL outperforms supervised pre-training** in learning image representations for object detection. Altogether, our results demonstrate the potential of self-supervised representations with good invariance properties.*

Algorithm paper

PIRL, CVPR 2020 (Self-Supervised Learning of Pretext-Invariant Representations)

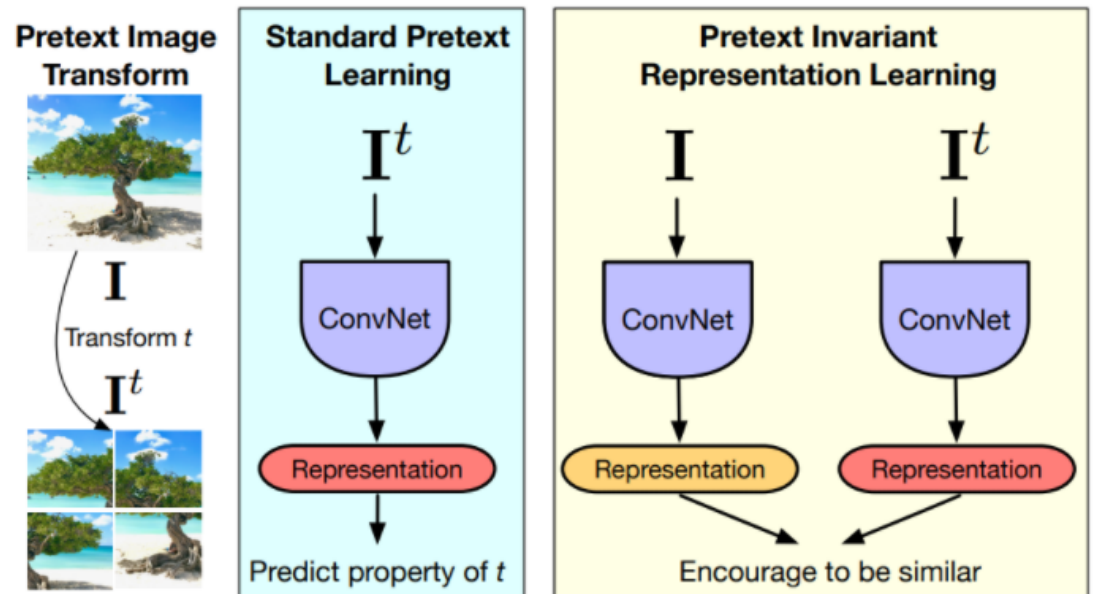
❖ Idea

- 어떤 pretext task를 선정하는지에 상관없이 같은 이미지면 transformation 후에도 visual semantic을 변화시키지 않는다는 아이디어에서 출발



[Rotation 기본 개념]

Standard pretext task 예시



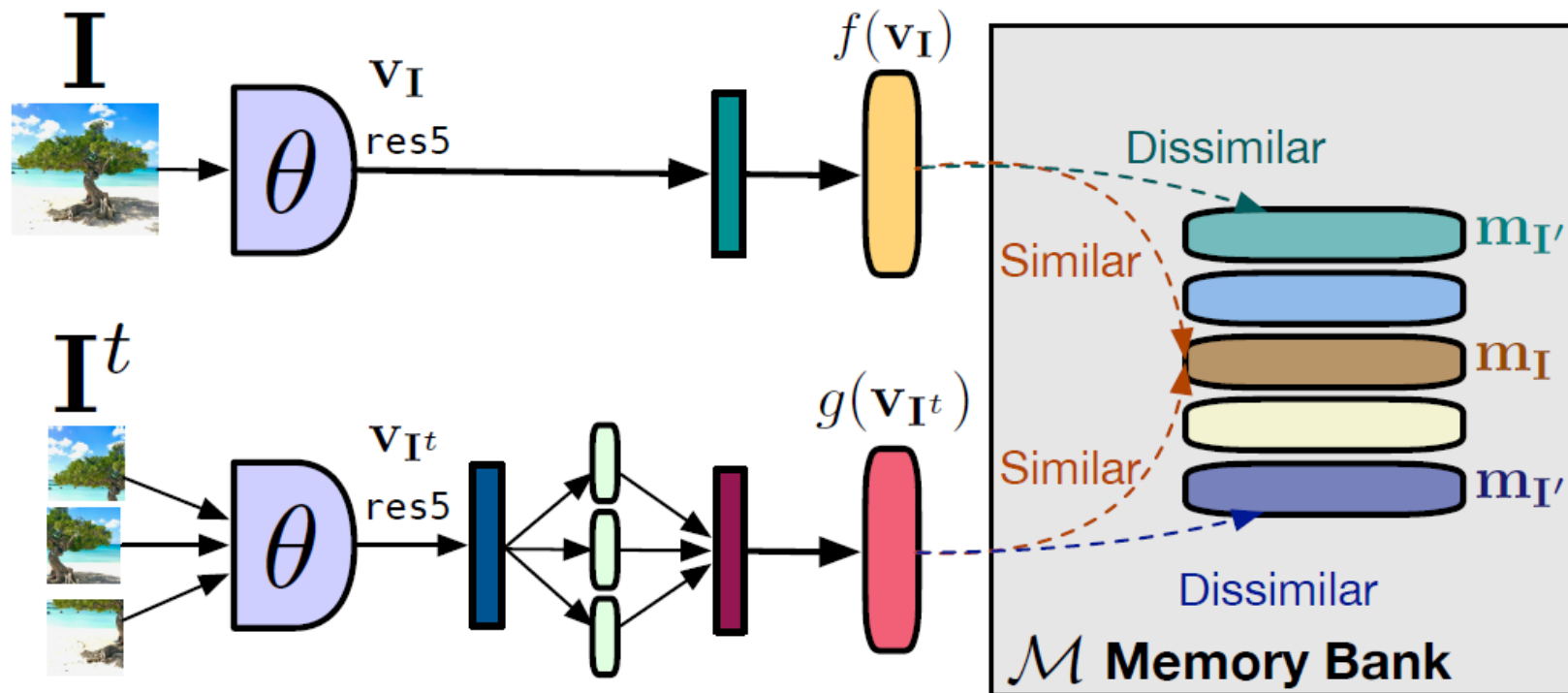
[PIRL 기본 개념]

Algorithm paper

PIRL, CVPR 2020 (Self-Supervised Learning of Pretext-Invariant Representations)

❖ Framework

- 같은 이미지에서 나온 representation은 transformation과 무관하게 유사해야 하고, 다른 이미지와 겹치지 않아야 함



Algorithm paper

PIRL, CVPR 2020 (Self-Supervised Learning of Pretext-Invariant Representations)

❖ Framework & method

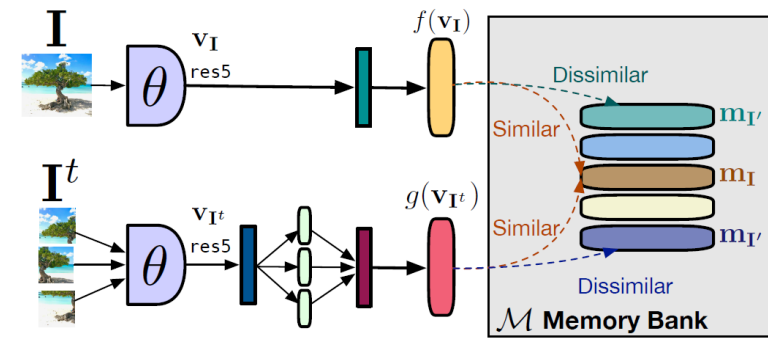
- 같은 이미지에서 나온 representation은 transformation과 무관하게 유사해야 하고, 다른 이미지와 겹치지 않아야 함
- Cosine similarity s 를 사용하고 NCE 적용하면 아래와 같음

$$h(\mathbf{v}_I, \mathbf{v}_{I^t}) = \frac{\exp\left(\frac{s(\mathbf{v}_I, \mathbf{v}_{I^t})}{\tau}\right)}{\exp\left(\frac{s(\mathbf{v}_I, \mathbf{v}_{I^t})}{\tau}\right) + |\mathcal{D}_N|/|\mathcal{D}|}$$

- 최종적으로 사용되는 loss function



$$L_{\text{NCE}}(\mathbf{I}, \mathbf{I}^t) = -\log [h(f(\mathbf{v}_I), g(\mathbf{v}_{I^t}))] \\ - \sum_{\mathbf{I}' \in \mathcal{D}_N} \log [1 - h(g(\mathbf{v}_{\mathbf{I}'}^t), f(\mathbf{v}_{\mathbf{I}'}))]$$



Algorithm paper

PIRL, CVPR 2020 (Self-Supervised Learning of Pretext-Invariant Representations)

❖ Experiment & result

- 본 논문 출판 당시에는 가장 우수한 성능을 보여줌
- 창의적인 아이디어만을 가지고 뛰어난 결과를 보여준 논문이고 다양한 도전이 가능하도록 함

Method	Parameters	Transfer Dataset			
		ImageNet	VOC07	Places205	iNat.
ResNet-50 using evaluation setup of [21]					
Supervised	25.6M	75.9	87.5	51.5	45.4
Colorization [21]	25.6M	39.6	55.6	37.5	–
Rotation [20]	25.6M	48.9	63.9	41.4	23.0
NPID++ [81]	25.6M	59.0	76.6	46.4	32.4
MoCo [26]	25.6M	60.6	–	–	–
Jigsaw [21]	25.6M	45.7	64.5	41.2	21.3
PIRL (ours)	25.6M	63.6	81.1	49.8	34.1
Different architecture or evaluation setup					
NPID [81]	25.6M	54.0	–	45.5	–
BigBiGAN [13]	25.6M	56.6	–	–	–
AET [85]	61M	40.6	–	37.1	–
DeepCluster [6]	61M	39.8	–	37.5	–
Rot. [35]	61M	54.0	–	45.5	–
LA [89]	25.6M	60.2 [†]	–	50.2 [†]	–
CMC [73]	51M	64.1	–	–	–
CPC [59]	44.5M	48.7	–	–	–
CPC-v2 [28]	305M	61.0	–	–	–
BigBiGAN-Big [13]	86M	61.3	–	–	–
AMDIM [4]	670M	68.1	–	55.1	–

Table 2: Image classification with linear models.

Method	Data fraction → Backbone	1%	10%
		Top-5 Accuracy	
Random initialization [81]	R-50	22.0	59.0
NPID [81]	R-50	39.2	77.4
Jigsaw [21]	R-50	45.3	79.3
NPID++ [81]	R-50	52.6	81.5
VAT + Ent Min. [22, 52]	R-50v2	47.0	83.4
S ⁴ L Exemplar [84]	R-50v2	47.0	83.7
S ⁴ L Rotation [84]	R-50v2	53.4	83.8
PIRL (ours)	R-50	57.2	83.8
Colorization [39]	R-152	29.8	62.0
CPC-v2 [28]	R-170 and R-11	64.0	84.9

Table 3: Semi-supervised learning on ImageNet.

Contents

1. Introduction

- Background
- Research timeline

2. Algorithm paper

- NPID, CVPR 2018
- MoCo, CVPR 2020
- PIRL, CVPR 2020

3. Application paper

- WaPIRL, IEEE T SEMICONDUCT M 2020 (preprint)
- Subject-Aware Contrastive Learning for Biosignals, CoRR 2020 (preprint)

4. Conclusion

- Comments

Algorithm paper

Self-Supervised Representation Learning for Wafer Bin Map Defect Pattern Classification, IEEE T SEMICONDUCT M 2020 (preprint)

❖ Motivation & framework

- Kahng and Kim이 반도체 분야에 최초로 self-supervised learning(PIRL)을 적용한 논문
- 반도체 산업에서 중요한 WBM(wafer bin map)의 불량을 탐지하기 위해 unlabeled WBMs을 이용하여 모델 구축

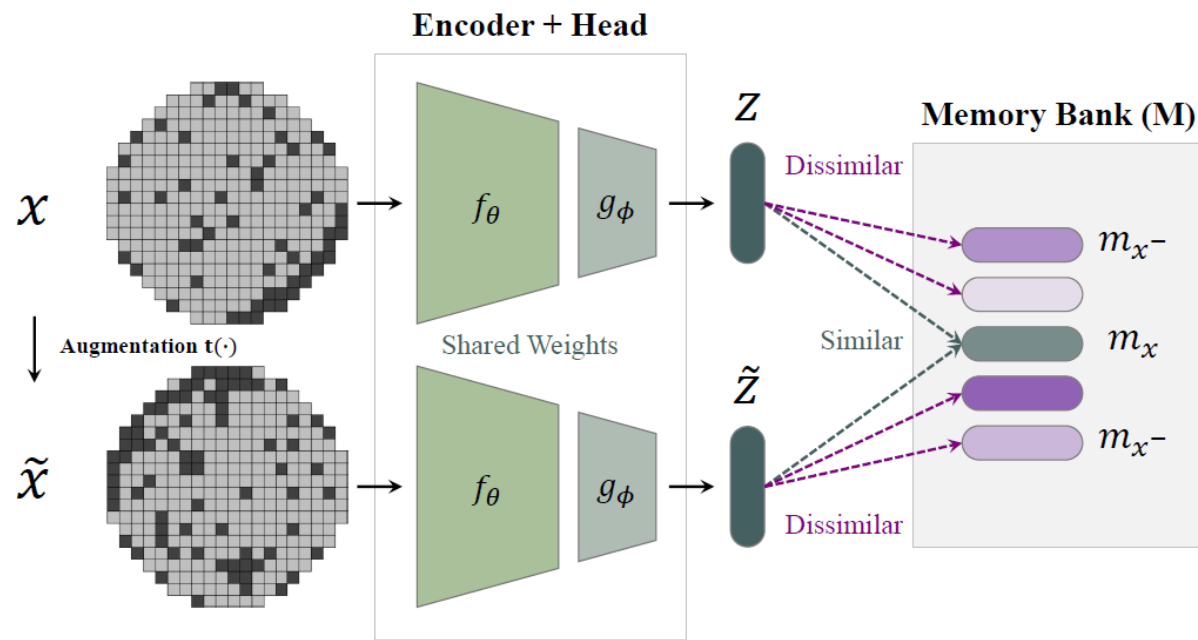
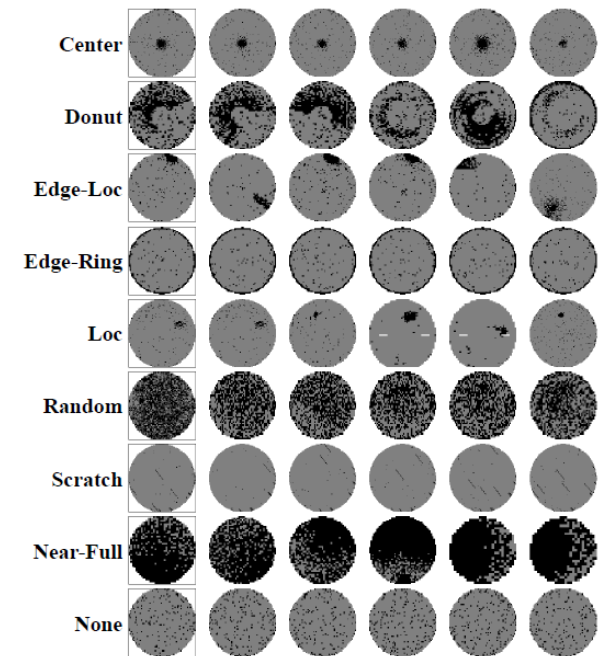


Fig. 1: A schematic overview of our proposed pre-training method, WaPIRL.



Algorithm paper

Self-Supervised Representation Learning for Wafer Bin Map Defect Pattern Classification, IEEE T SEMICONDUCT M 2020 (preprint)

❖ Experiment & result

- 5가지 augmentation 기법을 사용하였으며, crop의 효과가 가장 뛰어남
- 기존 모델들보다 본 논문에서 제안한 방법이 가장 우수한 성능을 보임

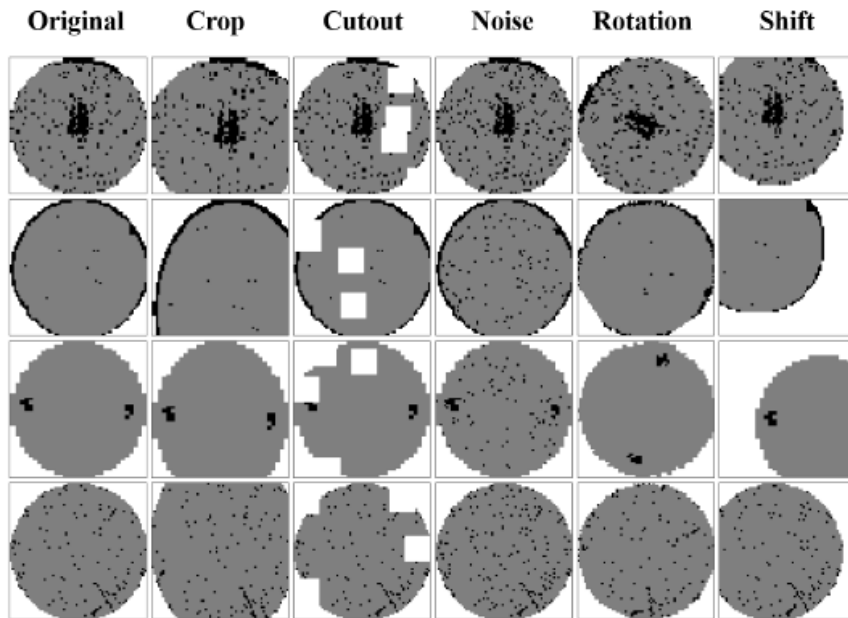


Fig. 2: Five different augmentations used in our work.

TABLE III: Comparisons of different classification models, in terms of the macro F_1 score. The best values for each proportion of labeled data is in **bold**. We report the median values over ten repeated trials with different random seeds.

Model	Labeled Data					
	1,383 (1%)	6,918 (5%)	13,836 (10%)	34,590 (25%)	69,180 (50%)	138,630 (100%)
WMFPR	0.498	0.692	0.717	0.726	0.749	0.767
WMDPI + RF	0.489	0.674	0.721	0.784	0.805	0.829
WMDPI + GB	0.412	0.606	0.623	0.667	0.704	0.727
WMDPI + SVE	0.572	0.696	0.717	0.757	0.785	0.801
CNN-WDI	0.630	0.790	0.839	0.851	0.858	0.877
d-CAE (ResNet-18)	0.403	0.706	0.745	0.830	0.850	0.869
d-CAE (ResNet-50)	0.411	0.856	0.788	0.845	0.861	0.859
WaPIRL+Crop (AlexNet)	0.708	0.783	0.831	0.849	0.870	0.885
WaPIRL+Crop (VGG16)	0.741	0.815	0.839	0.864	0.880	0.897
WaPIRL+Crop (ResNet-18)	0.693	0.792	0.821	0.857	0.879	0.895
WaPIRL+Crop (ResNet-50)	0.684	0.794	0.823	0.857	0.873	0.892

Algorithm paper

Subject-Aware Contrastive Learning for Biosignals, CoRR 2020 (preprint)

❖ Motivation & framework

- 생체 신호(biosignals)를 통해 인간의 활동을 예측하는 모형은 잠재가치가 크나, 사람을 대상으로 하는 데이터인 만큼 데이터의 수가 적고 labeling 작업도 값비쌈
- 적은 수의 EEG(뇌파), ECG(심전도) 데이터를 이용하여 self-supervised learning을 이용한 신체 활동 예측 모형 구축

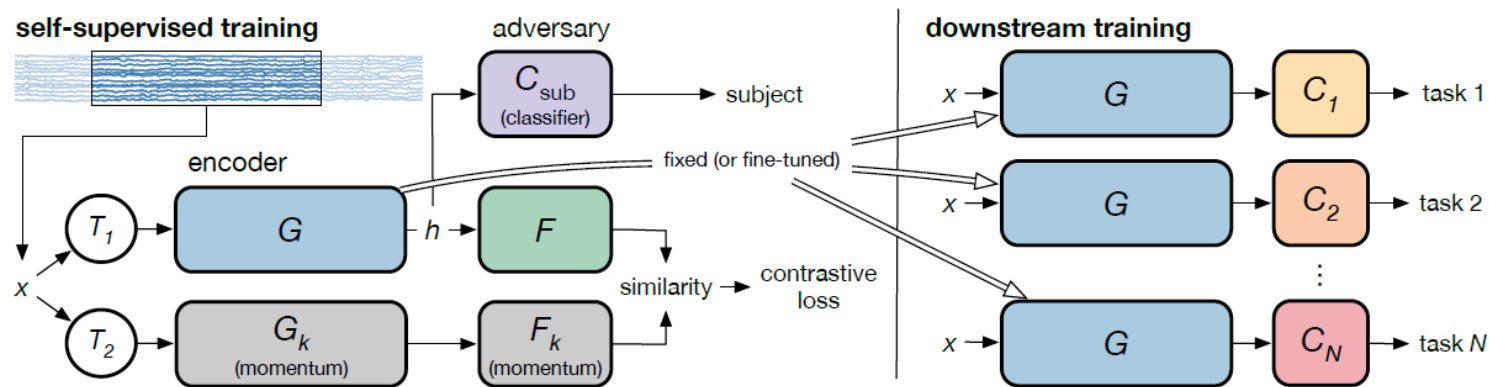


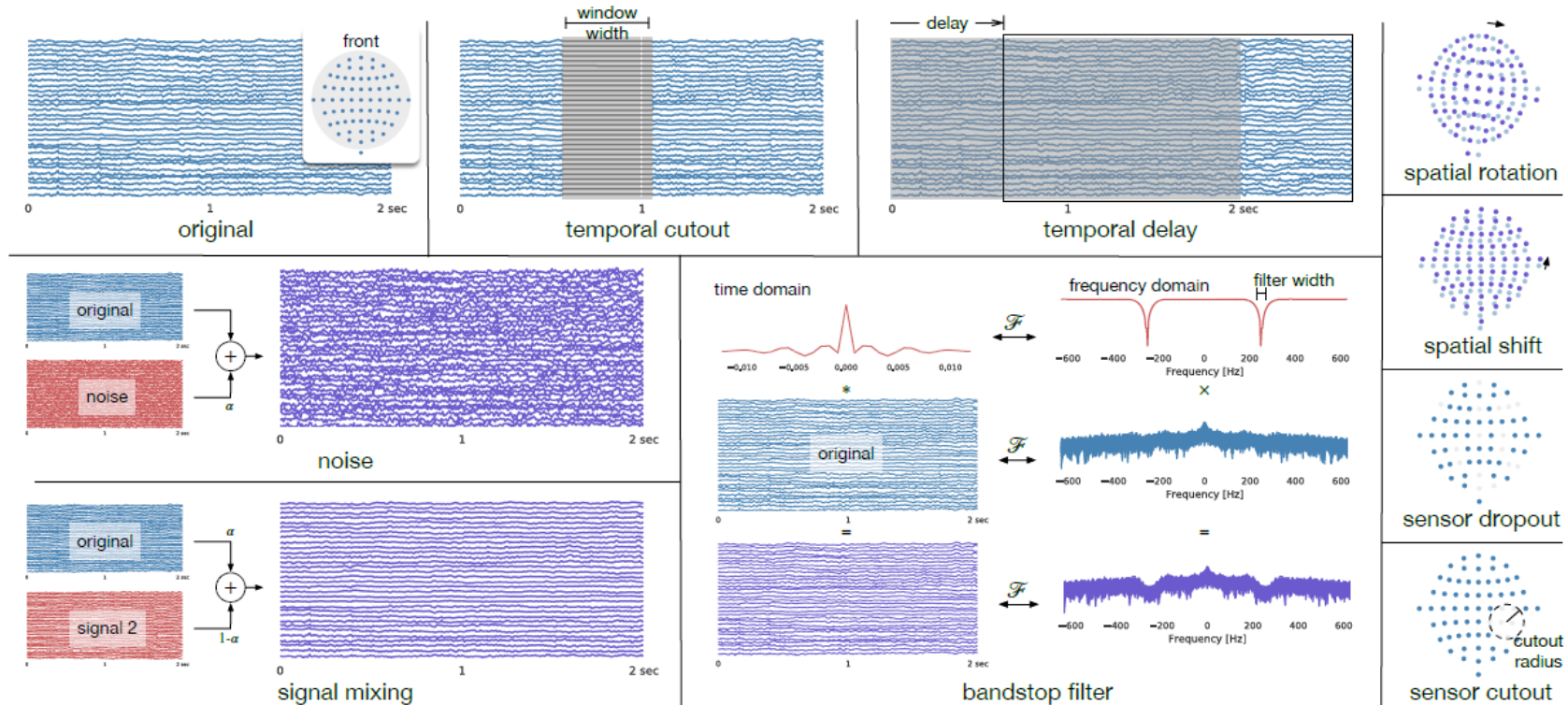
Figure 1: Method overview. An encoder G is trained using self-supervised training (left) with an adversarial subject identifier to minimize subject-specific content. Alternatively, subject-awareness can be introduced through subject-specific contrastive loss. The resulting encoder can then be applied for different downstream tasks by attaching a classification model (right).

Algorithm paper

Subject-Aware Contrastive Learning for Biosignals, CoRR 2020 (preprint)

❖ Method – augmentation

- Self-supervised learning 방법에서 가장 중요한 부분이 augmentation 기법
- 이미지 데이터와는 다르게 시그널 데이터는 아래와 같은 augmentation 기법을 사용하여 contrastive learning 진행



Algorithm paper

Subject-Aware Contrastive Learning for Biosignals, CoRR 2020 (preprint)

❖ Experiment & result

- EEG, ECG 데이터를 통해 신체 활동 예측에 기존 방법론보다 우수한 성능을 보임
- 적은 수의 데이터셋에 강건하고 효과적인 모델이라는 장점이 있음

Table 3: EEG classification accuracy [%]. General representations were learned through self-supervised training, and fine-tuning from these models (last three rows) improved accuracies compared to end-to-end (ours) training from random initialization.

	Intersubject		Intrasubject	
	2 class ^a	4 class ^b	2 class ^a	4 class ^b
Kim et al., Random forest [31]	80.1	-	-	-
Dose et al., CNN [32]	80.1	-	-	-
End-to-end (ours)	81.0±0.9	50.6±1.0	76.3±2.4	44.0±1.6
<i>Fine-tuned from:</i>				
Base SSL	81.1±0.6	52.6±0.7	78.6±2.5	50.8±1.4
Subject-specific	81.6±0.8	53.9±0.4	79.3±2.0	50.5±1.6
Subject-invariant	81.2±0.9	52.8±0.8	79.6±2.3	49.8±1.5

^a2 class: right fist and left fist; ^b4 class: right fist, left fist, both fists, and both feet.

Table B.1: Accuracies for intersubject beat classification: the training and testing were performed on separate set of subjects. General representations were learned through self-supervised training, and fine-tuning from these models improved accuracies (last six rows). For conciseness, only the standard deviation of the overall accuracies and F1 scores are reported. Our best results are bolded.

	Overall	SVEB		VEB	
	Acc	Acc / Se / +P	F1	Acc / Se / +P	F1
Luo et al. [48]	89.3	96.2 / 15.4 / 47.3	23.2	95.5 / 60.4 / 66.8	63.4
Huang et al. [40]	93.8	95.1 / 91.1 / 42.2	57.7	99.0 / 93.9 / 90.9	92.4
Niu et al. [43]	96.4	- / 76.5 / 76.6	76.6	- / 85.7 / 94.1	89.7
End-to-end (ours)	91.0±2.0	95.9 / 25.8 / 41.8	31.4±10.3	97.8 / 95.6 / 78.9	85.8±8.2
End-to-end ^R (ours)	91.9±1.8	96.3 / 44.5 / 52.2	46.7±14.9	98.4 / 96.4 / 83.4	89.2±4.3
<i>Fine-tuned from:</i>					
Base SSL ^R	91.7±1.8	95.6 / 30.2 / 39.2	33.3±8.1	98.5 / 96.8 / 83.7	89.6±3.9
Subject-specific ^R	93.2±1.6	96.0 / 42.8 / 50.1	43.8±10.0	98.9 / 96.6 / 88.7	92.4±3.3
Subject-invariant ^R					
λ = 1.0	89.4±3.0	96.3 / 38.8 / 48.9	42.5±13.8	97.8 / 95.8 / 77.4	85.4±4.4
λ = 0.1	91.8±2.2	96.2 / 35.5 / 48.8	40.8±14.6	98.0 / 95.3 / 79.2	86.3±5.3
λ = 0.01	90.4±2.2	95.7 / 35.2 / 42.2	37.3±16.8	98.1 / 96.4 / 80.0	87.2±5.2
λ = 0.001	91.3±2.7	95.8 / 26.5 / 40.1	31.6±13.5	98.1 / 96.2 / 80.4	87.2±7.1

^RIncluding RR-interval information as input to the classifier.

Abbreviations: SVEB = supraventricular ectopic beat; VEB = ventricular ectopic beat; Acc = accuracy [%]; Se = sensitivity [%]; +P = positive predictability; F1 = F1 score.

Contents

1. Introduction

- Background
- Research timeline

2. Algorithm paper

- NPID, CVPR 2018
- MoCo, CVPR 2020
- PIRL, CVPR 2020

3. Application paper

- WaPIRL, IEEE T SEMICONDUCT M 2020 (preprint)
- Subject-Aware Contrastive Learning for Biosignals, CoRR 2020 (preprint)

4. Conclusion

- Comments

Conclusion

❖ Comments

- Self-supervised learning 분야가 컴퓨터 비전 쪽에서 굉장히 많이 연구되고 있고 SOTA 모델이 금방 갱신되고 있는 상황 → 많은 연구자들이 흥미롭게 생각하고 있는 분야라는 생각
- 이번 세미나는 Contrastive learning을 다룬 논문들을 주로 살펴보았는데 최근 BYOL(Bootstrap Your Own Latent) 이나 PCL(Prototypical Contrastive Learning) 등 다른 아이디어를 통해 성능을 확보하는 알고리즘들도 나오고 있음 → BYOL은 기존 SOTA보다 우수한 성능을 보이고 있어 추후 기회가 된다면 리뷰할 예정
- 컴퓨터 비전쪽에서 시작한 연구지만 다양한 산업 도메인에 적용을 고민해볼 여지가 충분한 것 같다고 생각 → 특히 시그널 쪽 데이터에 augmentation 기법을 잘 개선하여 적용하면 좋은 성능을 기대해볼 수 있을 거라 느낌

Thank you

