POSTECH
INDUSTRIAL AND MANAGEMENT ENGINEERING, POSTECH

# Self-supervised Learning of Pretext-Invariant Representation

포항공과대학교 산업경영공학과

Stochastic Systems Lab

Jongwon Kim

April 4, 2022

Ishan Misra, Laurens van der Maaten

*"Computer Vision and Pattern Recognition"* (*CVPR*), 2020

STOCHASTIC SYSTEMS LAB.

# Contents

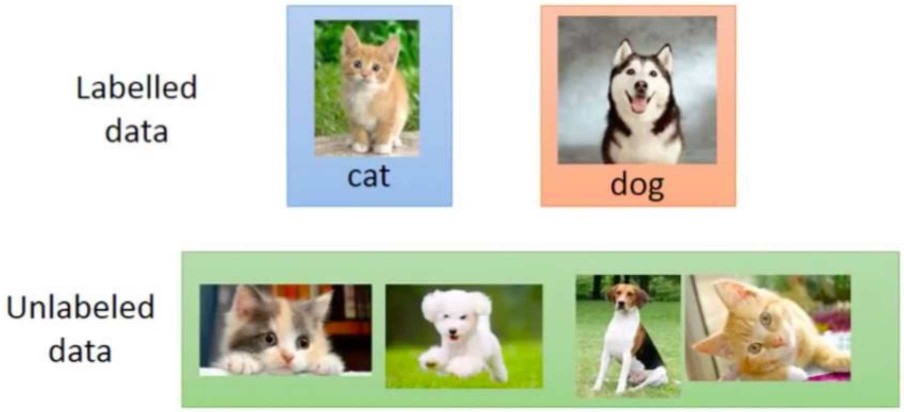


STOCHASTIC SYSTEMS LAB.

# Problem statement

Why we use self-supervised learning?

## Object detection or classification on deep learning model




**Massive amount of unlabeled data**

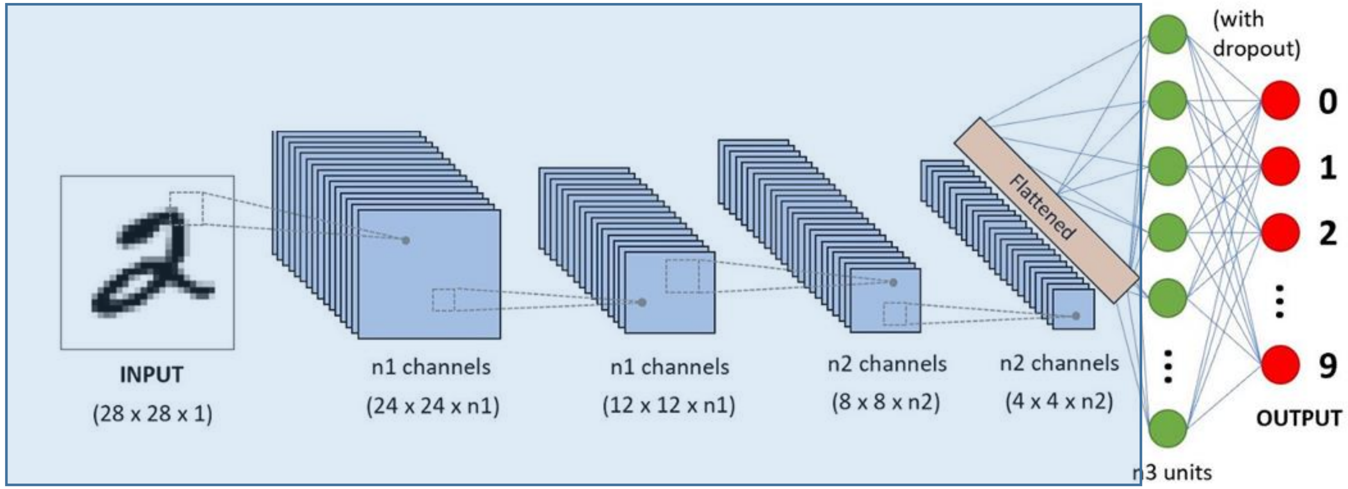
**Self-supervised learning** is the most promising way to approximate a form of **common sense** from **unlabeled data** in AI system

# Problem statement

Why we use self-supervised learning?

How to use **common sense** in deep learning model? → **Transfer learning**

- ✓ Fix **encoder**
- ✓ Fine-tune **encoder** with 1~10% labelled data by linear classifier
- ✓ Fine-tune **encoder** with 1~10% labelled data by real task




Encoder

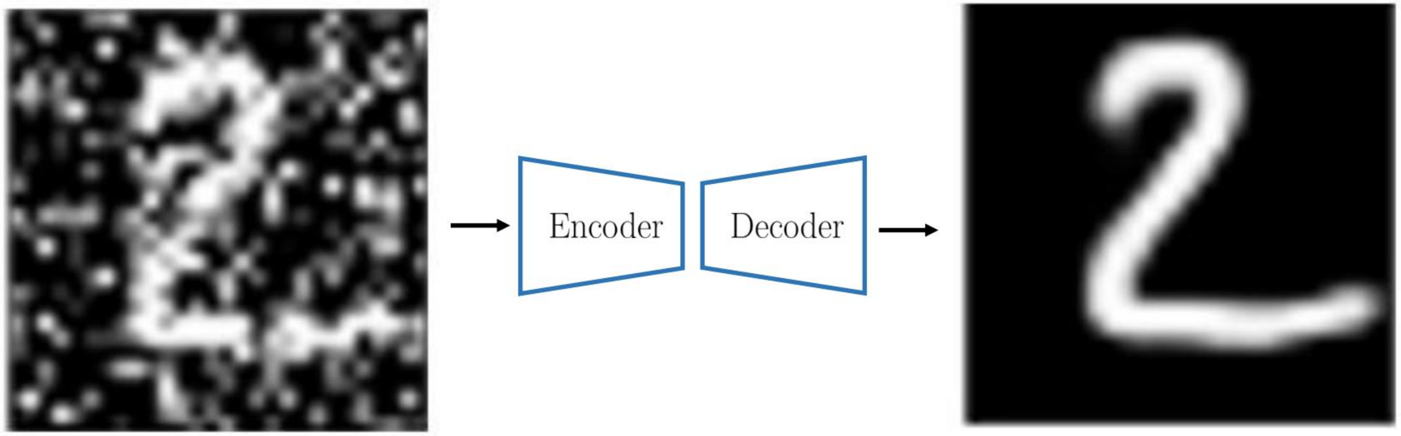
# Previous studies

Auto-encoder

**Auto-encoder** is the one of the most simple classical **self-supervised learning**.

**Auto-encoder** does **not perform well** for empirically high resolution images.

- ✓ Lots of effort spent on **“useless” details**: exact color, good boundary.



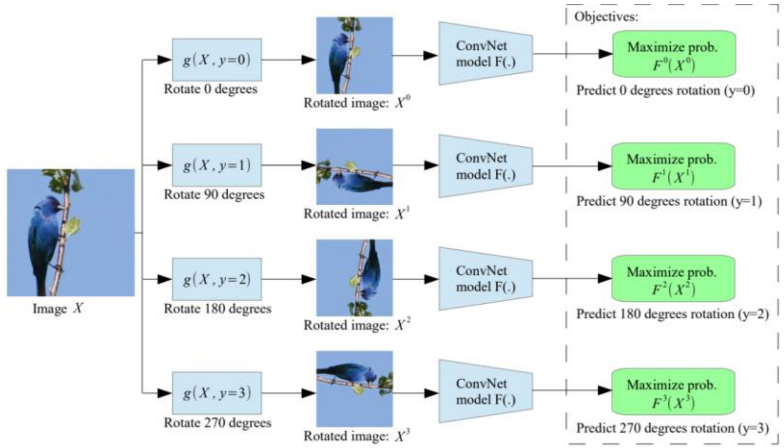
# Previous studies

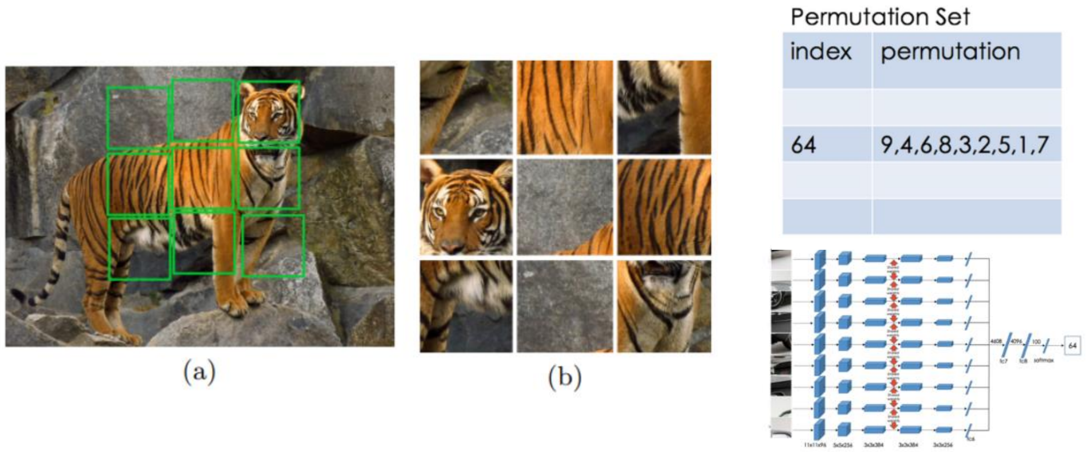
Self-supervised learning

**Create a pretext classification problem** from unlabeled data.

- ✓ RotNet [1]: Create a model to classify the **degree of rotation.**
- ✓ Jigsaw [2]: Turn one image into a **jigsaw puzzle.**

**Limitation: the encoder represents the pretext feature.**



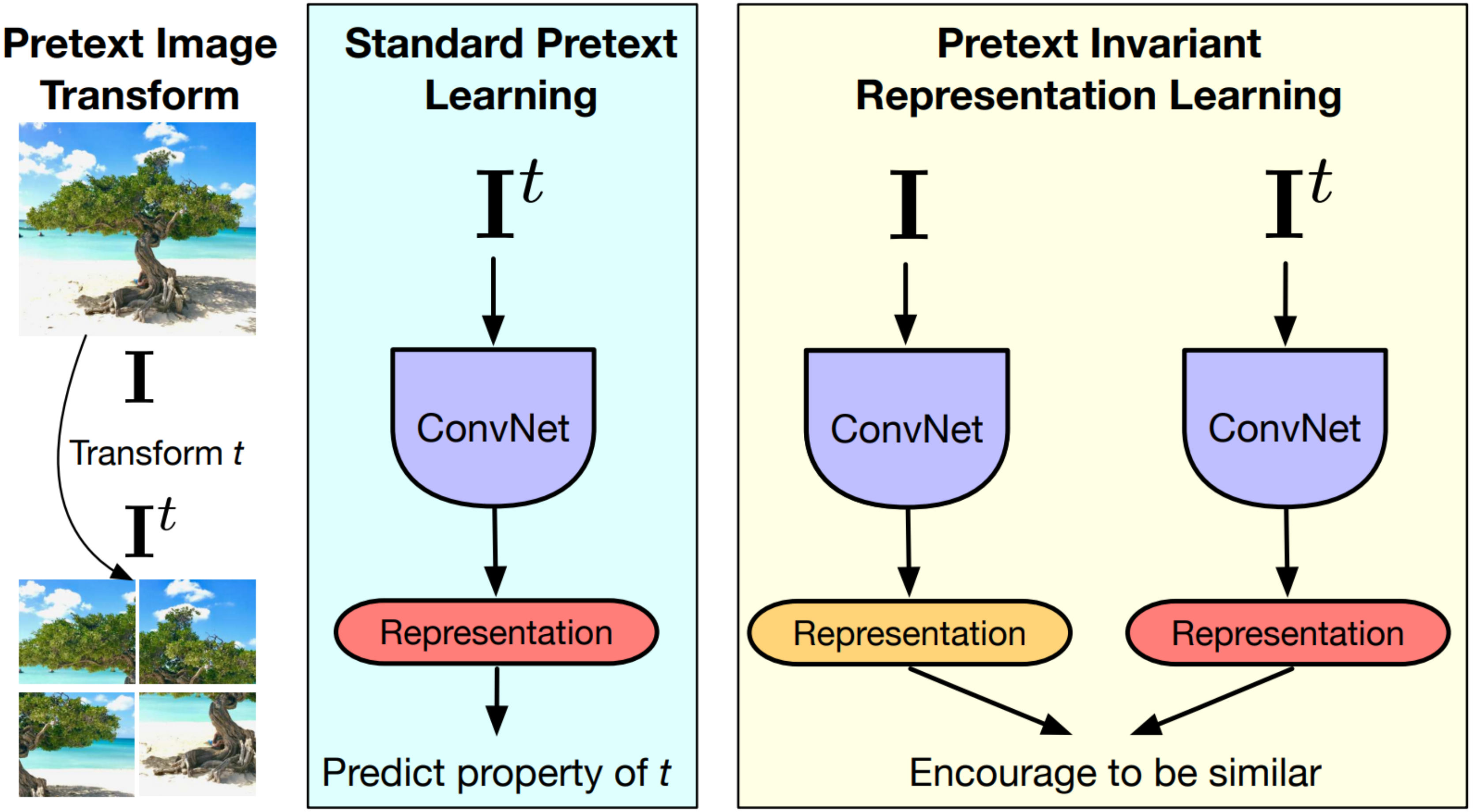


[1] Gidaris, Spyros, Praveer Singh, and Nikos Komodakis. "*Unsupervised representation learning by predicting image rotations*", ICLR, 2018
[2] Mehdi Noroozi, Paolo Favaro, "*Unsupervised Learning of Visual Representations by Solving Jigsaw Puzzles*", CVPR 2016


STOCHASTIC SYSTEMS LAB.

# Idea

Self-supervised Learning of Pretext-Invariant Representation



# Result

Classification on ImageNet dataset

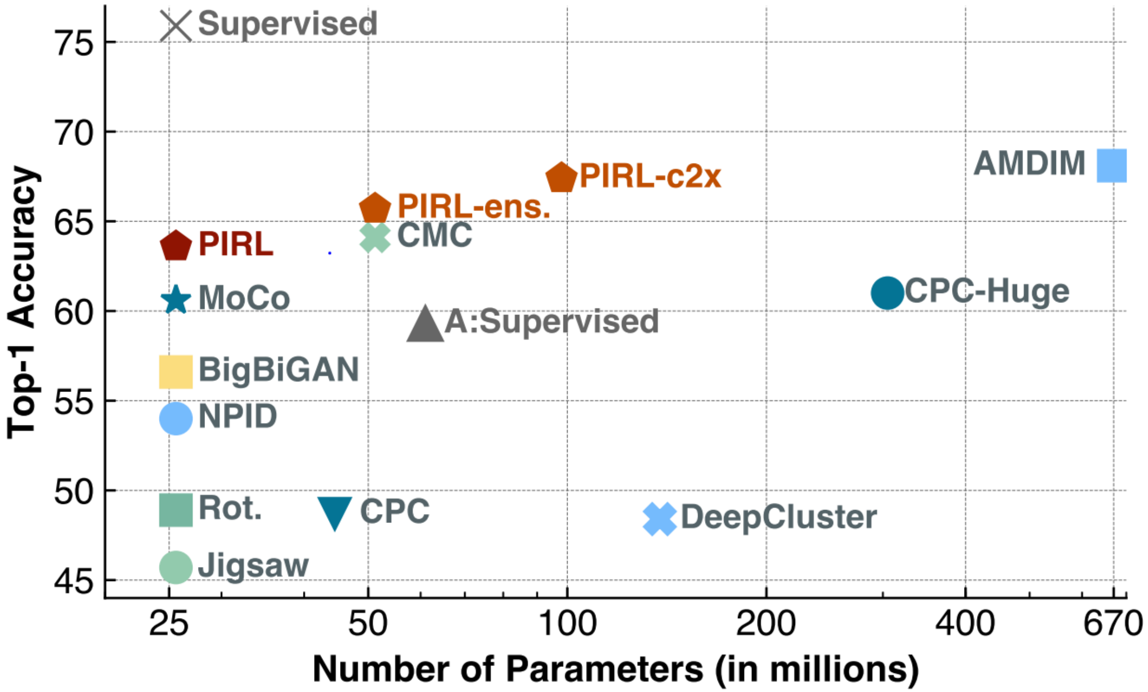



**Figure 2: ImageNet classification with linear models.** Single-crop top-1 accuracy on the ImageNet validation data as a function of the number of parameters in the model that produces the representation ("A" represents AlexNet). Pretext-Invariant Representation Learning (PIRL) sets a new state-of-the-art in this setting (red marker) and uses significantly smaller models (ResNet-50). See Section 4.2 for more details.

Problem: Classification
Dataset: ImageNet
Model: ResNet-50
Transfer learning: Fixed encoder

STOCHASTIC SYSTEMS LAB.

# Result

Object detection on ImageNet dataset

| Method | Network | $AP^{all}$ | $AP^{50}$ | $AP^{75}$ | $\Delta AP^{75}$ |
|---|---|---|---|---|---|
| Supervised | R-50 | 52.6 | **81.1** | 57.4 | =0.0 |
| Jigsaw [19] | R-50 | 48.9 | 75.1 | 52.9 | -4.5 |
| Rotation [19] | R-50 | 46.3 | 72.5 | 49.3 | -8.1 |
| NPID++ [72] | R-50 | 52.3 | 79.1 | 56.9 | -0.5 |
| PIRL **(ours)** | R-50 | **54.0** | 80.7 | **59.7** | **+2.3** |
| CPC-Big [26] | R-101 | – | 70.6* | – | |
| CPC-Huge [26] | R-170 | – | 72.1* | – | |
| MoCo [24] | R-50 | 55.2*† | 81.4*† | 61.2*† | |

**Table 1: Object detection on VOC07+12 using Faster R-CNN.** Detection AP on the VOC07 test set after finetuning Faster R-CNN models (keeping BatchNorm fixed) with a ResNet-50 backbone pre-trained using self-supervised learning on ImageNet. Results for supervised ImageNet pre-training are presented for reference. Numbers with * are adopted from the corresponding papers. Method with † finetunes BatchNorm. PIRL significantly outperforms supervised pre-training without extra pre-training data or changes in the network architecture. Additional results in Table 6.

Problem: Object detection

Dataset: VOC07

Model: ResNet-50 + Faster R-CNN

Transfer learning: Fine tuning

# Thank you