Deep Learning for Computer Vision

Fall 2022

https://cool.ntu.edu.tw/courses/189345 (NTU COOL)

http://vllab.ee.ntu.edu.tw/dlcv.html (Public website)

Yu-Chiang Frank Wang 王鈺強, Professor

Dept. Electrical Engineering, National Taiwan University

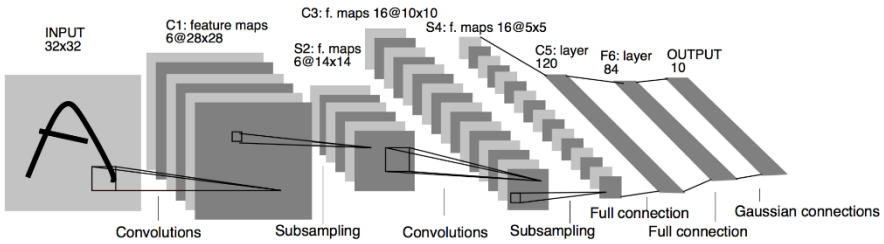
2022/12/6

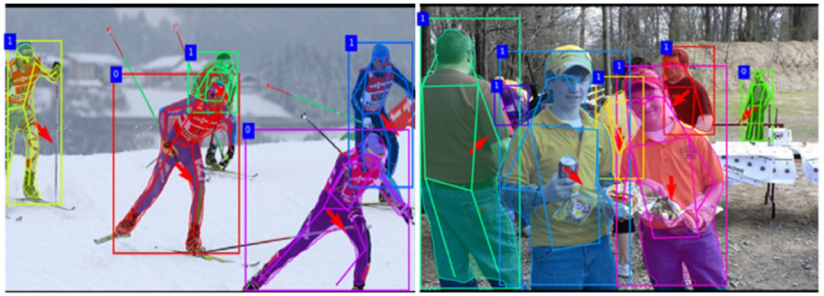
What to Cover Today…

- Self-Supervised Learning (SSL)
  - Pretext Tasks vs. Contrastive Learning
  - SSL Beyond Images
- Invited Talk
  - 10 Secrets You Need to Know About Software Engineering & Career Planning

Supervised Learning

- Deep learning plus supervised learning are rocking the world ...





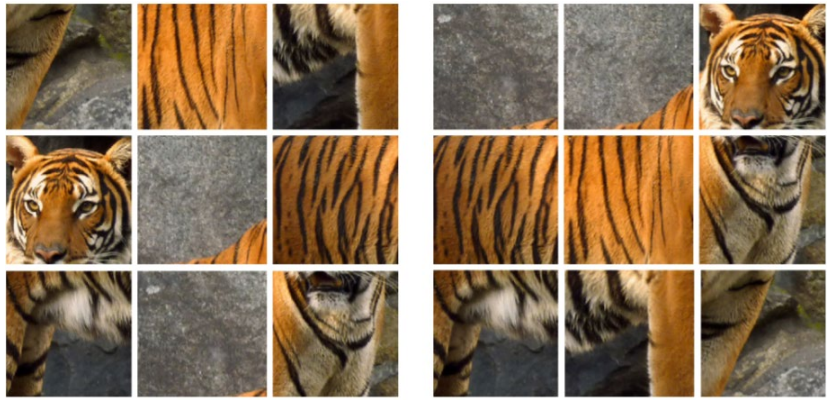
- In real world scenarios, data-annotation is quite **time-consuming**
- Could one exploit supervised signals from **unlabeled** data?

Self-Supervised Learning (SSL)

- Learning discriminative representations from **unlabeled** data
- Create self-supervised tasks via **data augmentation**

Colorization

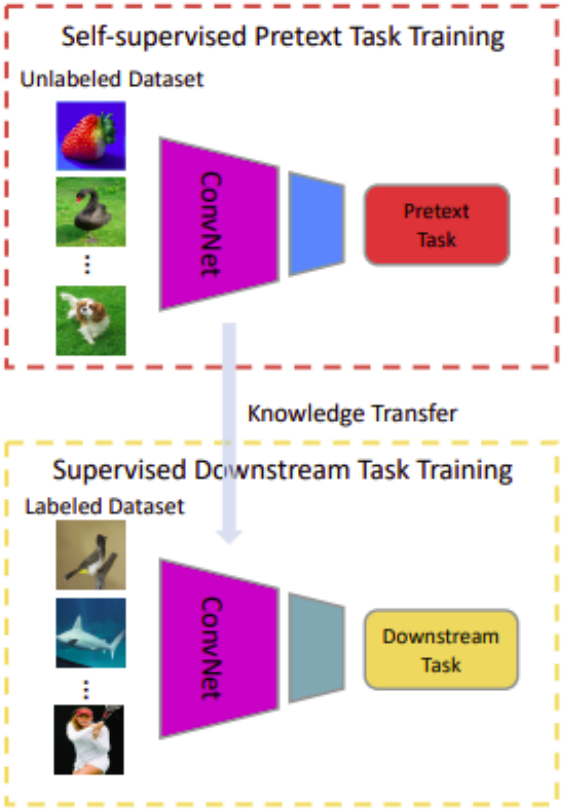
Rotation



Jigsaw Puzzle

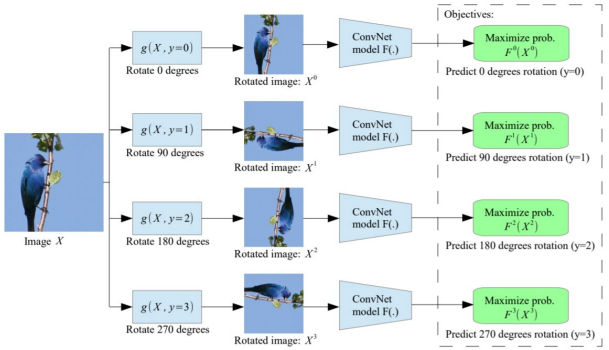
Self-Supervised Learning (SSL)

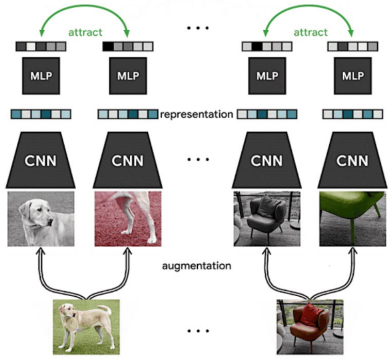
- Self-Supervised Pretraining
- Supervised Fine-tuning

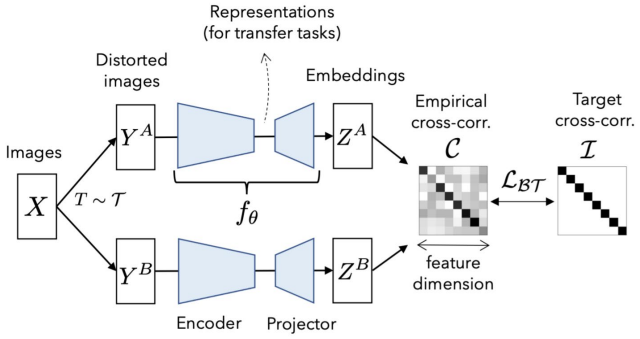


Self-Supervised Learning (SSL)

- Pretext Tasks
  - Jigsaw (ECCV'16)
  - RotNet (ICLR'18)
- Contrastive Learning
  - CPC (ICML'20)
  - SimCLR (ICML'20)
- Learning w/o negative samples
  - BYOL (NeurIPS'20)
  - Barlow Twins (ICML'21)

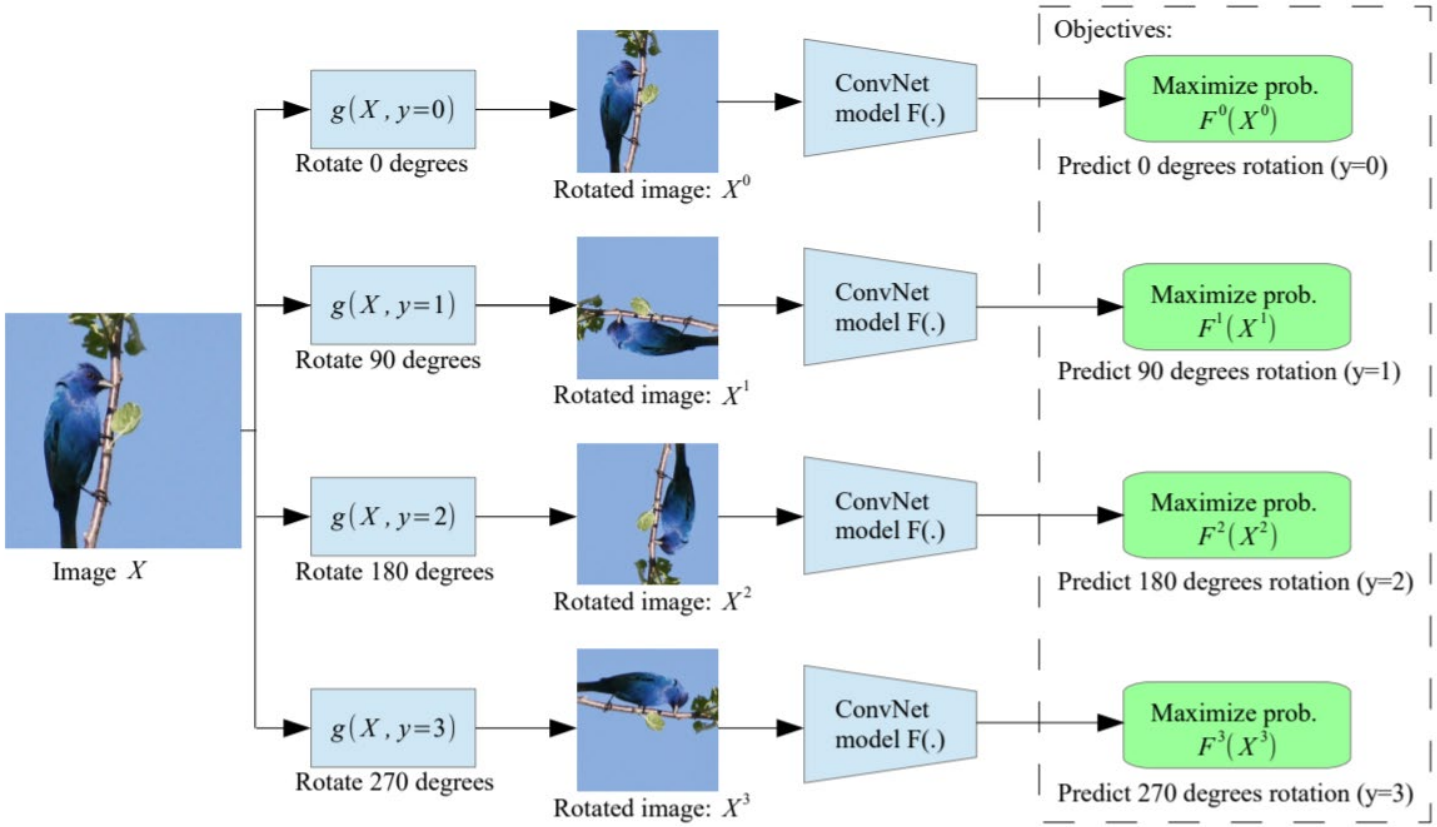






RotNet

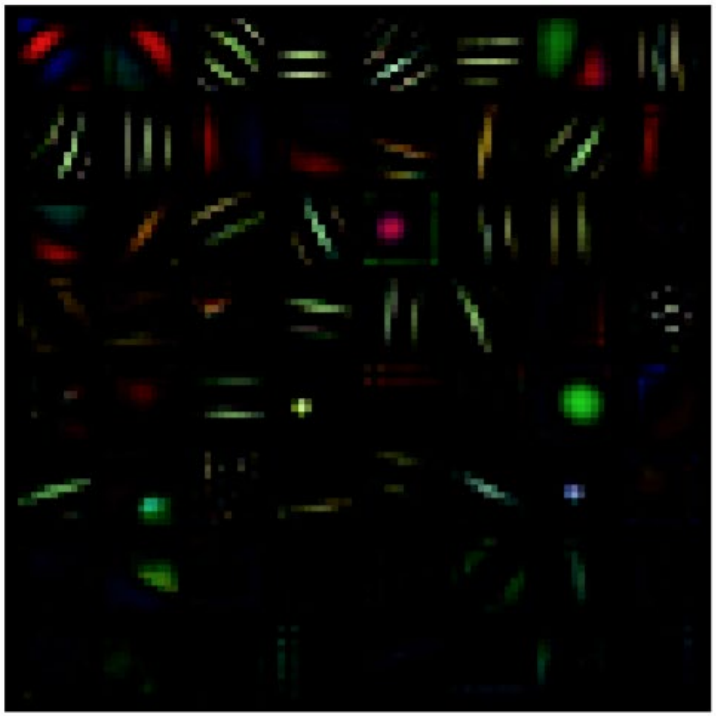
- Learning to predict the **rotation** angle



Gidaris et al. "Unsupervised Representation Learning by Predicting Image Rotations." ICLR 2018

RotNet

- Filters learned with SSL exhibit more **variety**

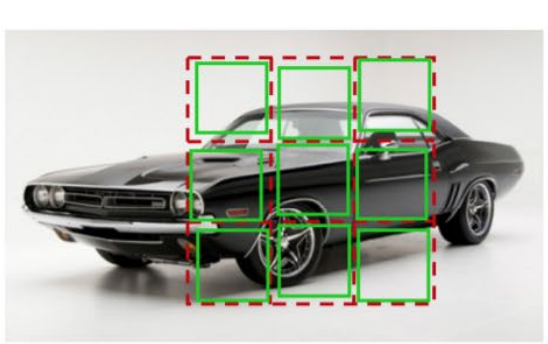


(a) **Supervised**

(b) **Self-supervised to recognize rotations**

Jigsaw Puzzle

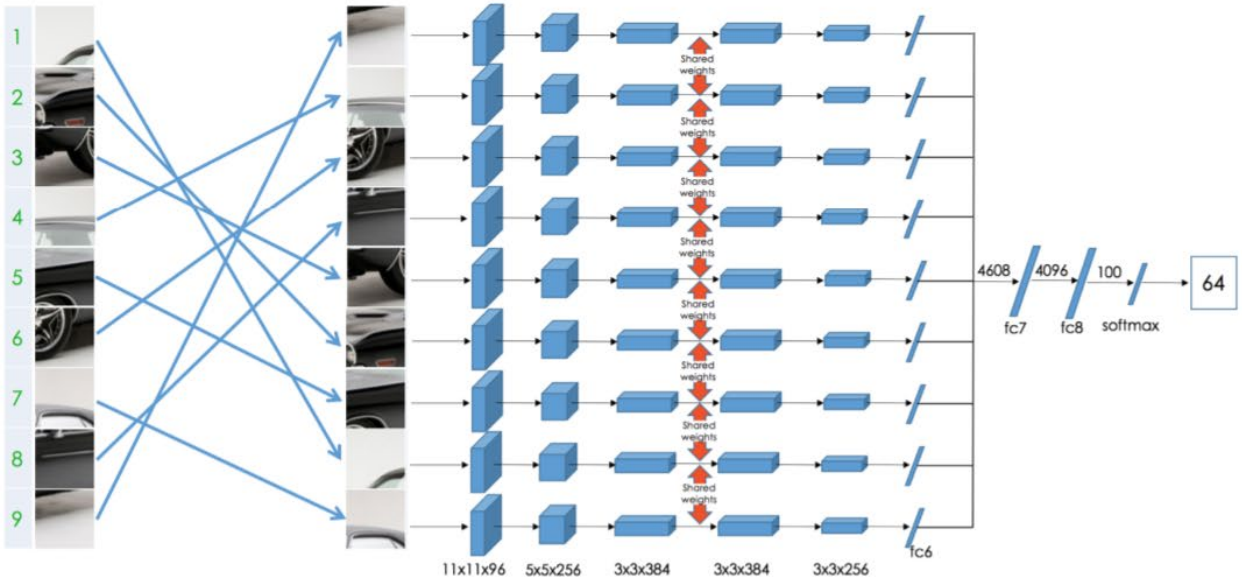
- Assign the **permutation index** and perform augmentation
- Solve jigsaw puzzle by predicting the permutation index



Permutation Set

index	permutation
64	9,4,6,8,3,2,5,1,7

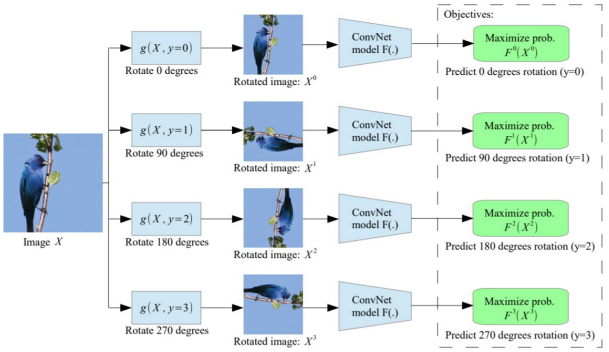
Reorder patches according to the selected permutation

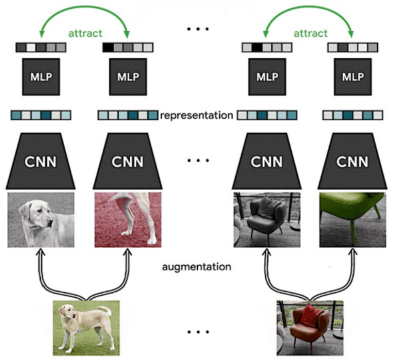


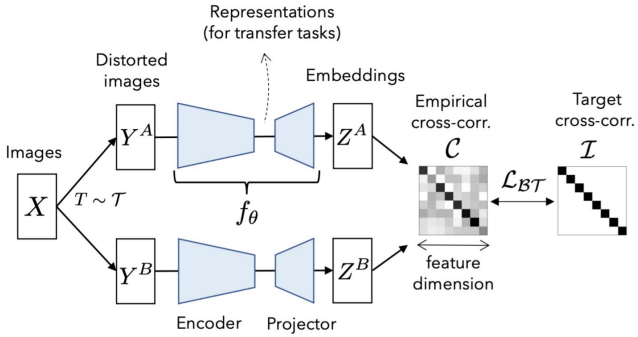
Noroozi et al. "Unsupervised learning of visual representations by solving jigsaw puzzles." ECCV 2016

Self-Supervised Learning (SSL)

- Pretext Tasks
  - Jigsaw (ECCV'16)
  - RotNet (ICLR'18)
- Contrastive Learning
  - CPC (ICML'20)
  - SimCLR (ICML'20)
- Learning w/o negative samples
  - BYOL (NeurIPS'20)
  - Barlow Twins (ICML'21)

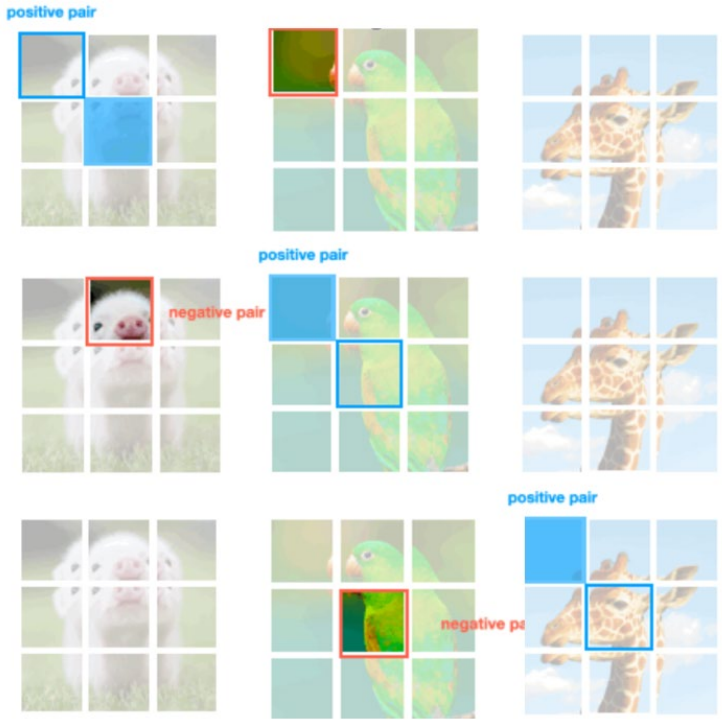






Contrastive Predictive Coding (CPC)

- Sample **positive** patches from itself and **negative** patches from other images
- **Maximize** positive similarities and **minimize** negative ones



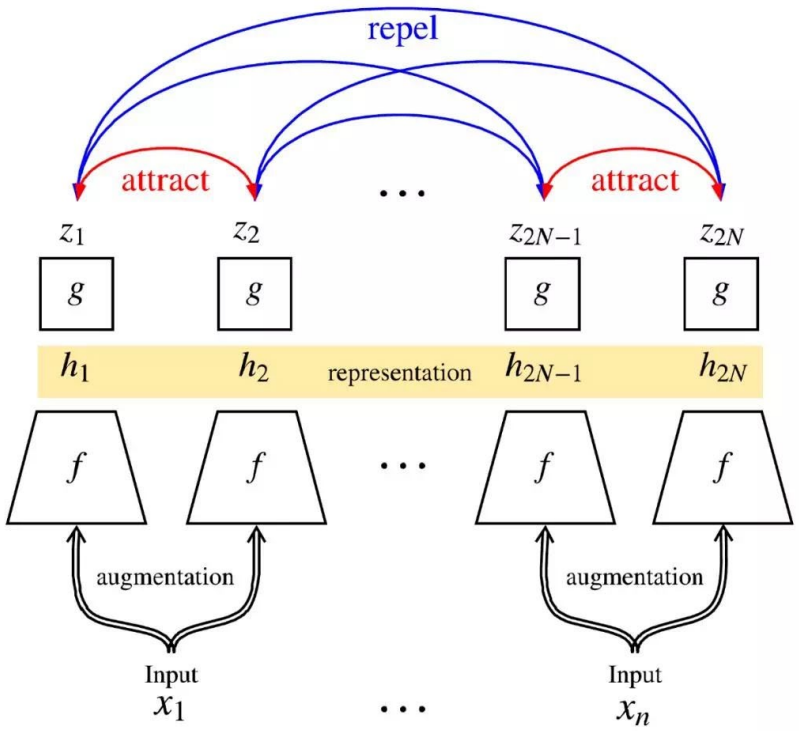
$$\mathcal{L}_{\text{CPC}} = -\sum_{i,j,k} \log \frac{\exp(\hat{\boldsymbol{z}}_{i+k,j}^{T} \boldsymbol{z}_{i+k,j})}{\exp(\hat{\boldsymbol{z}}_{i+k,j}^{T} \boldsymbol{z}_{i+k,j}) + \sum_{l} \exp(\hat{\boldsymbol{z}}_{i+k,j}^{T} \boldsymbol{z}_{l})}$$

positive (numerator) · positive (denominator) · negative (denominator)

Henaff et al. “Data-efficient image recognition with contrastive predictive coding.“ ICML 2020

SimCLR

- **Attract** augmented images and **repel** negative samples
- Improve the representation quality with **projection heads** (*g*)…why?



Chen et al. "A simple framework for contrastive learning of visual representations." ICML 2020

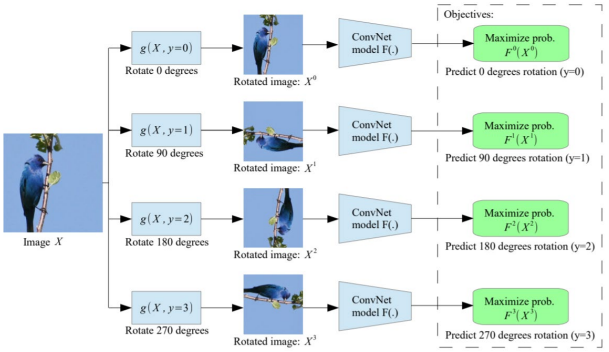
SimCLR

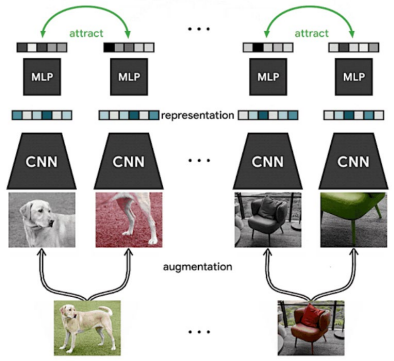
- Experiments on semi-supervised settings

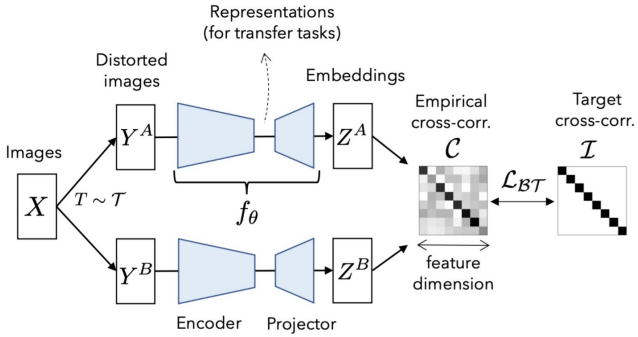
Method	Architecture	Label fraction (Top 5) 1%	10%
Supervised baseline	ResNet-50	48.4	80.4
Methods using other label-propagation:			
Pseudo-label	ResNet-50	51.6	82.4
VAT+Entropy Min.	ResNet-50	47.0	83.4
UDA (w. RandAug)	ResNet-50	-	88.5
FixMatch (w. RandAug)	ResNet-50	-	89.1
S4L (Rot+VAT+En. M.)	ResNet-50 (4×)	-	91.2
Methods using representation learning only:			
InstDisc	ResNet-50	39.2	77.4
BigBiGAN	RevNet-50 (4×)	55.2	78.8
PIRL	ResNet-50	57.2	83.8
CPC v2	ResNet-161(∗)	77.9	91.2
SimCLR (ours)	ResNet-50	75.5	87.8
SimCLR (ours)	ResNet-50 (2×)	83.0	91.2
SimCLR (ours)	ResNet-50 (4×)	**85.8**	**92.6**

Self-Supervised Learning (SSL)

- Pretext Tasks
  - Jigsaw (ECCV'16)
  - RotNet (ICLR'18)
- Contrastive Learning
  - CPC (ICML'20)
  - SimCLR (ICML'20)
- Learning w/o negative samples
  - BYOL (NeurIPS'20)
  - Barlow Twins (ICML'21)

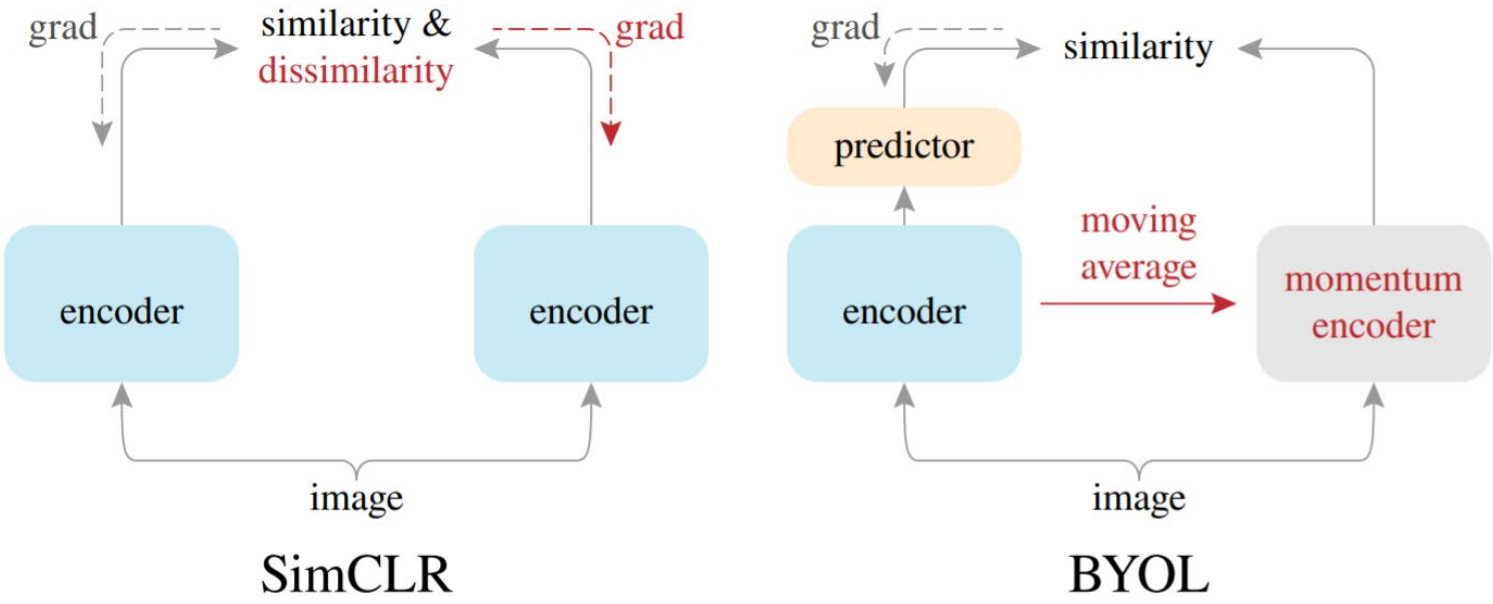






Bootstrap Your Own Latent (BYOL)

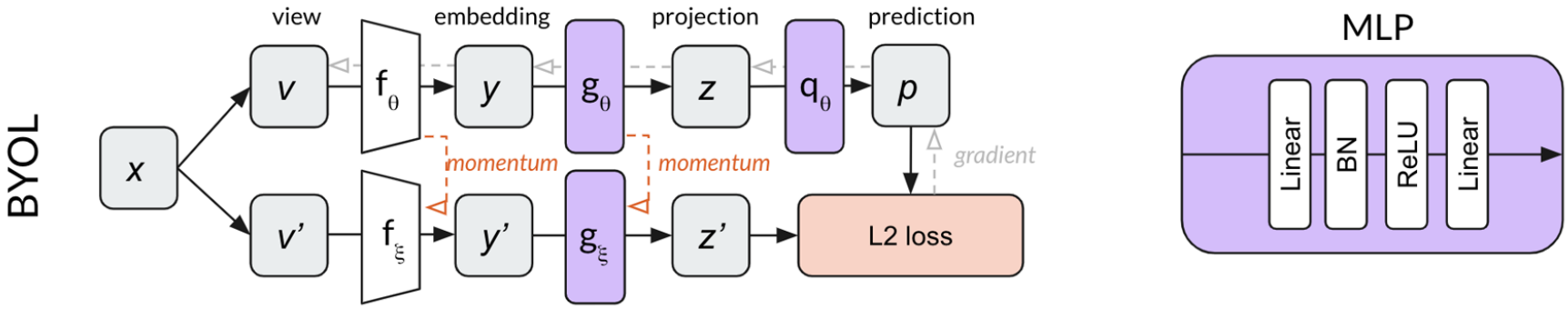
- No need of negative pairs
- Introduce the **predictor** for asymmetry to avoid collapse
- Exponential Moving Average (**EMA**) $\theta_T \leftarrow \tau\theta_T + (1-\tau)\theta_S$



Grill et al. "Bootstrap your own latent: A new approach to self-supervised learning." NeurIPS 2020

BYOL

- No need of negative pairs
- Introduce the **predictor** for asymmetry to avoid collapse
- Exponential Moving Average (**EMA**)

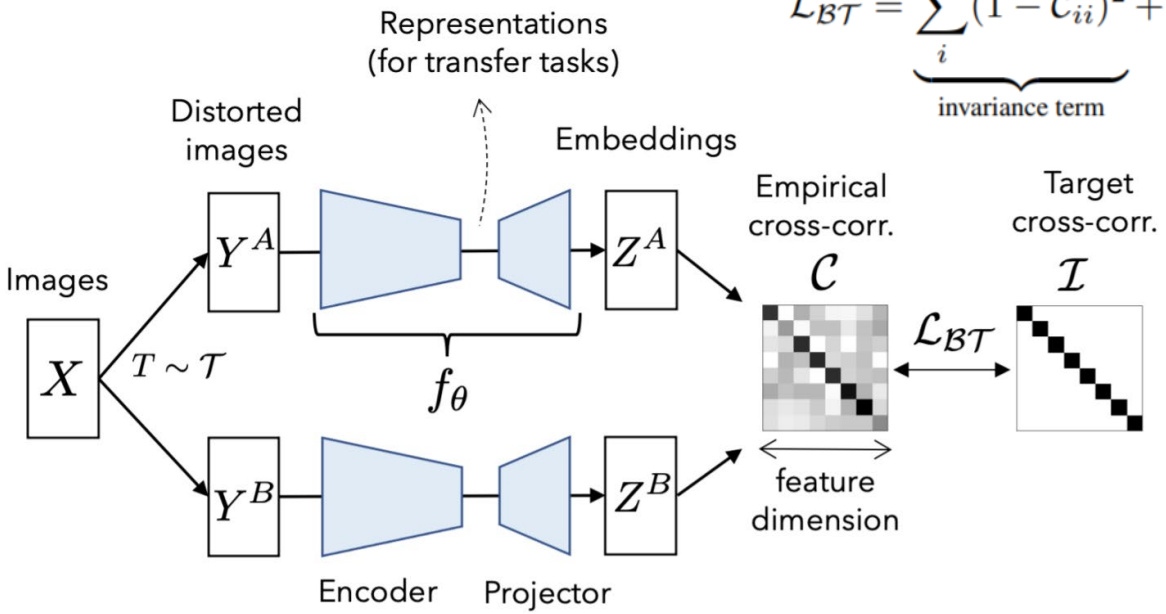


Grill et al. "Bootstrap your own latent: A new approach to self-supervised learning." NeurIPS 2020

Barlow Twins

- Enforce **diversity** among **feature dimensions**
- Maximize diagonal terms and minimize off-diagonal ones
- No need of negative pairs, predictor network, gradient stopping or moving average techniques

$$\mathcal{L}_{\mathcal{BT}} \triangleq \underbrace{\sum_i (1 - \mathcal{C}_{ii})^2}_{\text{invariance term}} + \lambda \underbrace{\sum_i \sum_{j \neq i} \mathcal{C}_{ij}{}^2}_{\text{redundancy reduction term}}$$



Zbontar et al. "Barlow twins: Self-supervised learning via redundancy reduction." ICML 2021

Barlow Twins

- Experiments on classification

Method	Top-1		Top-5	
	1%	10%	1%	10%
Supervised	25.4	56.4	48.4	80.4
PIRL	-	-	57.2	83.8
SIMCLR	48.3	65.6	75.5	87.8
BYOL	53.2	68.8	78.4	89.0
SWAV	53.9	**70.2**	78.5	**89.9**
BARLOW TWINS (ours)	**55.0**	69.7	**79.2**	89.3

Barlow Twins

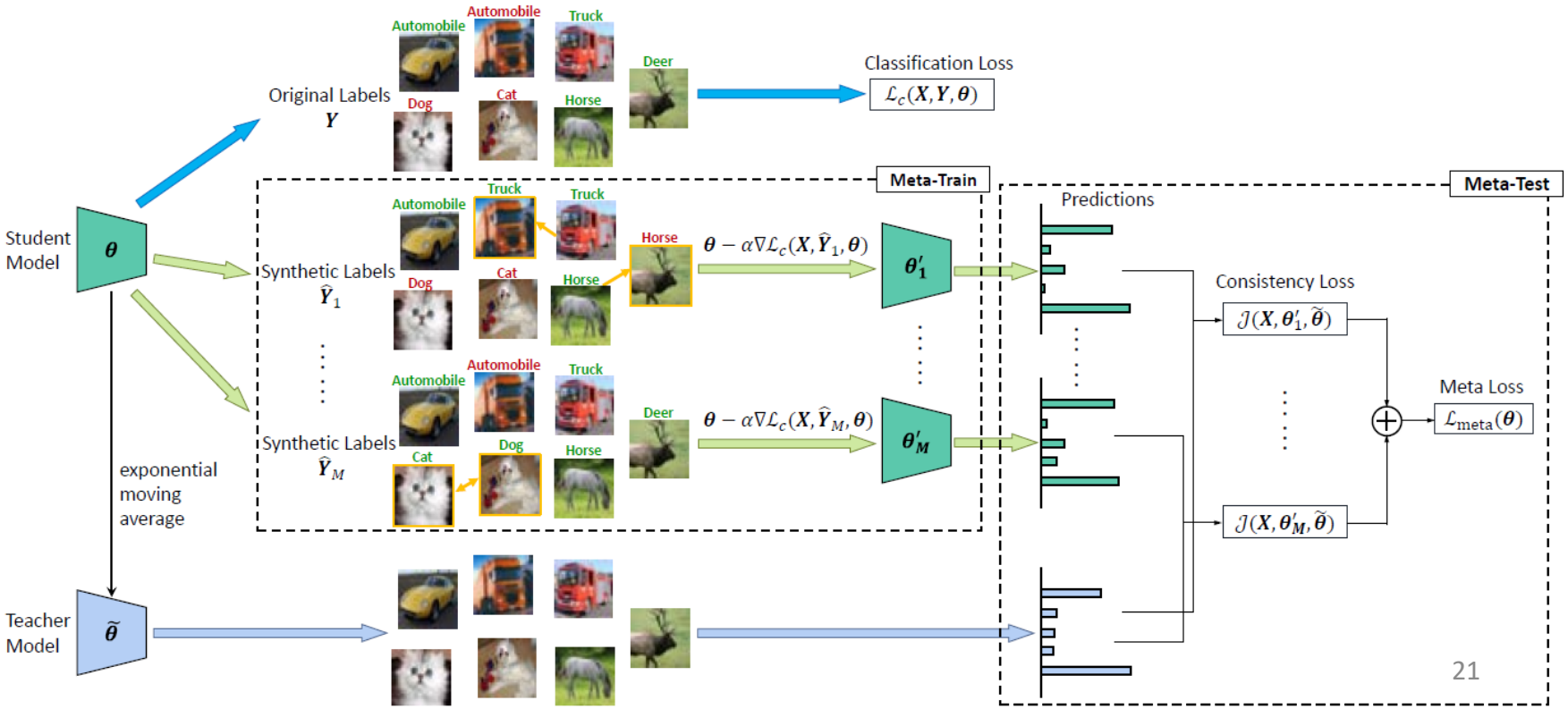
- Experiments on detection and segmentation

Method	VOC07+12 det			COCO det			COCO instance seg		
	AP_{all}	AP_{50}	AP_{75}	AP^{bb}	AP^{bb}_{50}	AP^{bb}_{75}	AP^{mk}	AP^{mk}_{50}	AP^{mk}_{75}
Sup.	53.5	81.3	58.8	38.2	58.2	41.2	33.3	54.7	35.2
MoCo-v2	**57.4**	82.5	**64.0**	**39.3**	58.9	**42.5**	**34.4**	55.8	36.5
SwAV	56.1	**82.6**	62.7	38.4	58.6	41.3	33.8	55.2	35.9
SimSiam	57	82.4	63.7	39.2	**59.3**	42.1	**34.4**	**56.0**	**36.7**
BT (ours)	56.8	**82.6**	63.4	39.2	59.0	**42.5**	34.3	**56.0**	36.5

SSL Beyond Image Data

- What about videos?

- What about noisy data? J. Li et al., Learning to Learn from Noisy Labeled Data, *CVPR* 2019



What's Next?

- Self-Supervised Learning (SSL)
  - Pretext Tasks vs. Contrastive Learning
  - SSL Beyond Images
- Invited Talk
  - 10 Secrets You Need to Know About Software Engineering & Career Planning