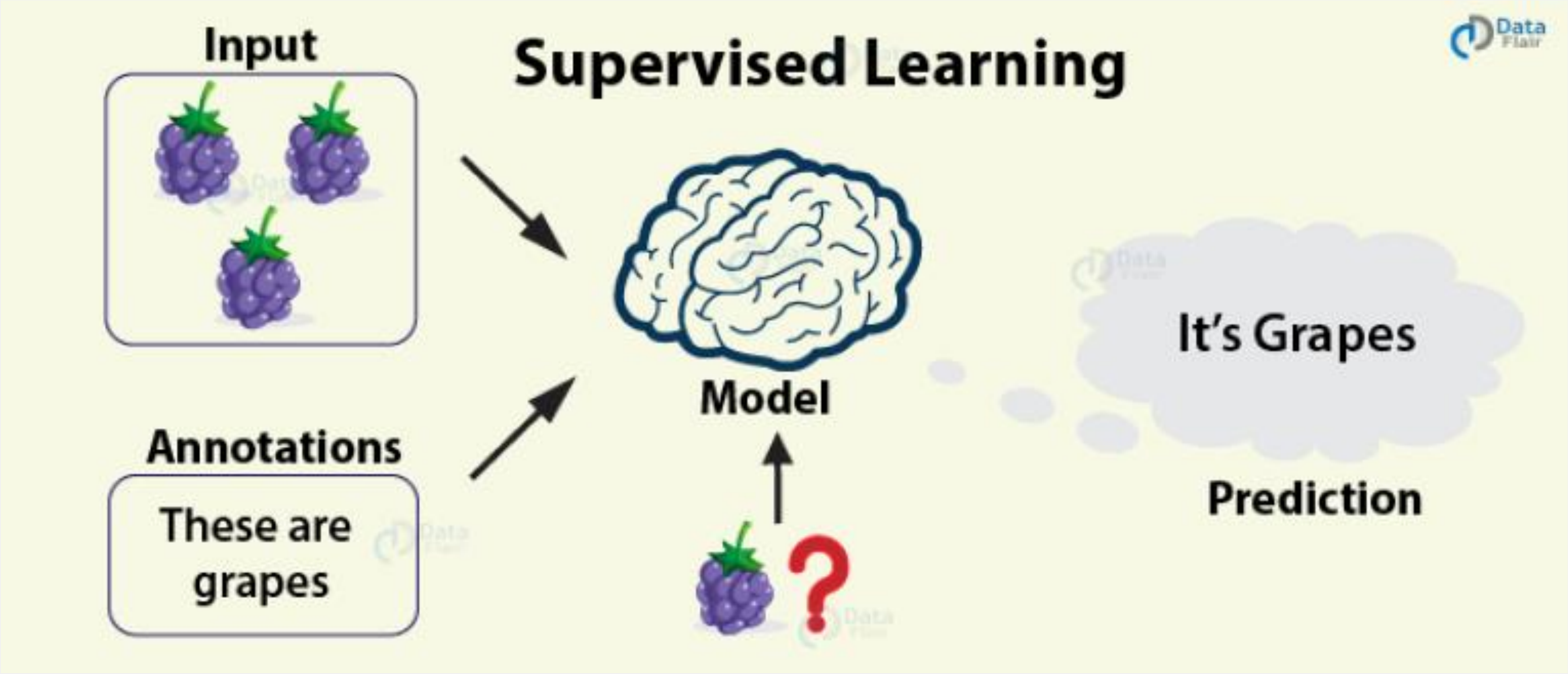
Self-supervised Learning

Self-supervised Learning
Introduction

Why research and development in AI have skyrocketed?

- The focus was largely on supervised learning methods that need huge amounts of labeled data to train system for specific use cases



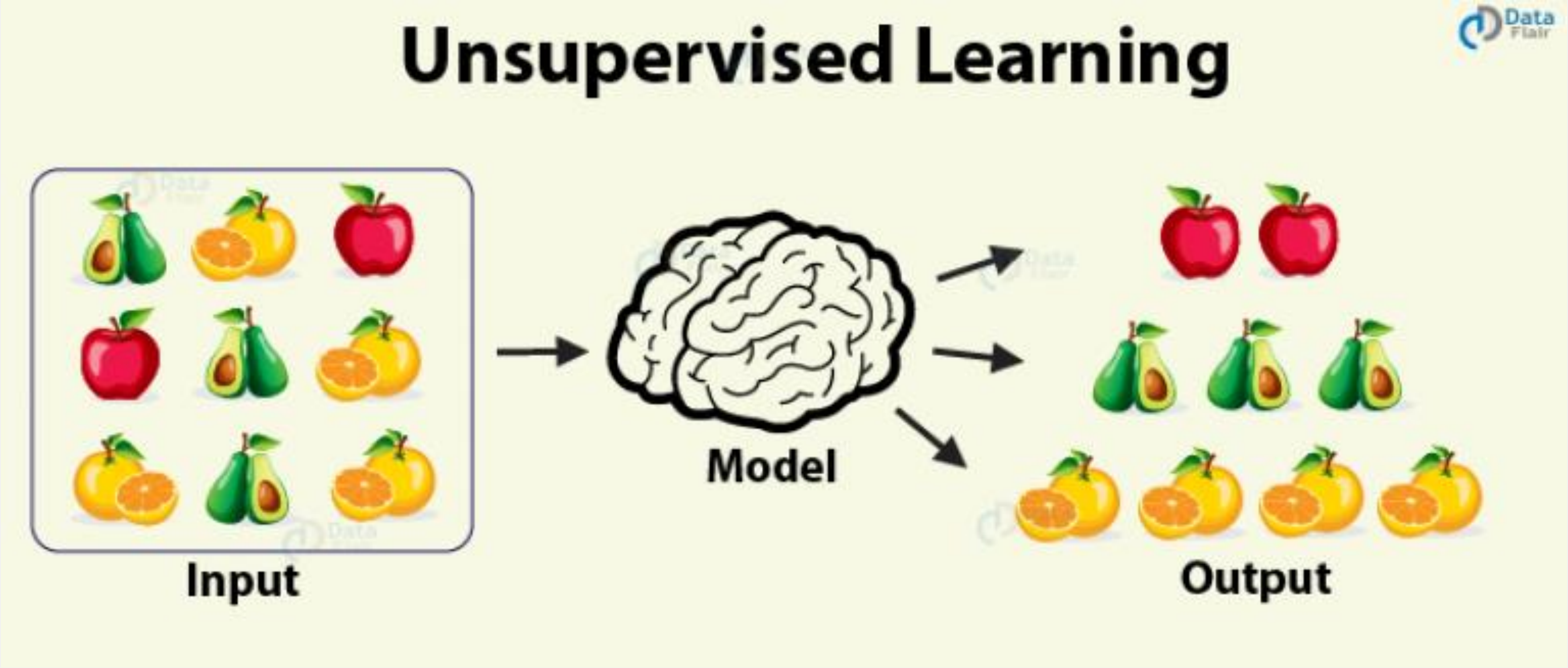
[Reference]

Self-supervised Learning

Introduction

Unsupervised learning

- It is a deep learning technique used to find implicit patterns of data without being trained on labeled data

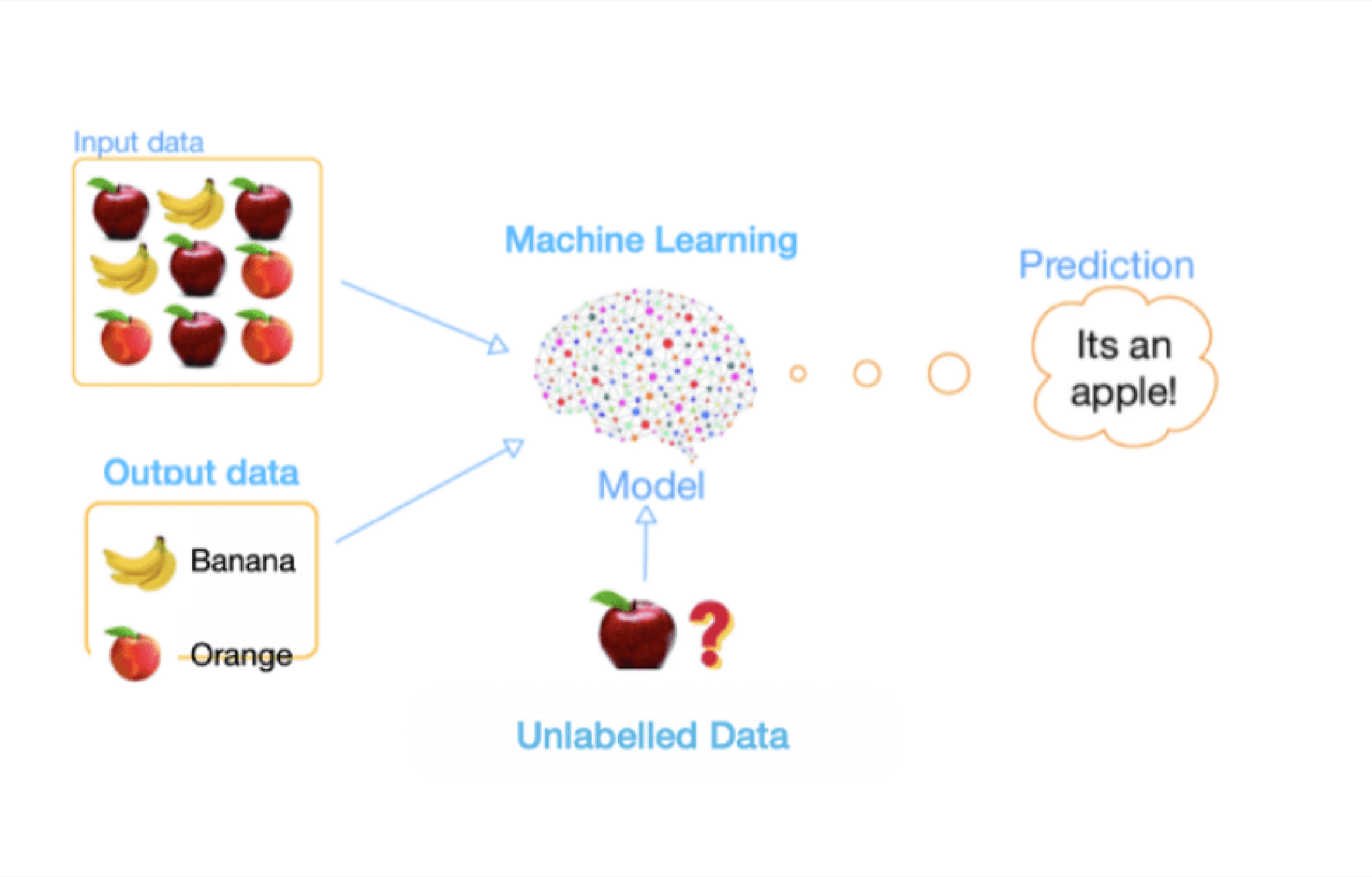


[Reference](#)

Self-supervised Learning
Introduction

Semi-supervised learning

- ➢ We have input data and a fraction of input data is labeled as the output



Reference

Self-supervised Learning

Introduction

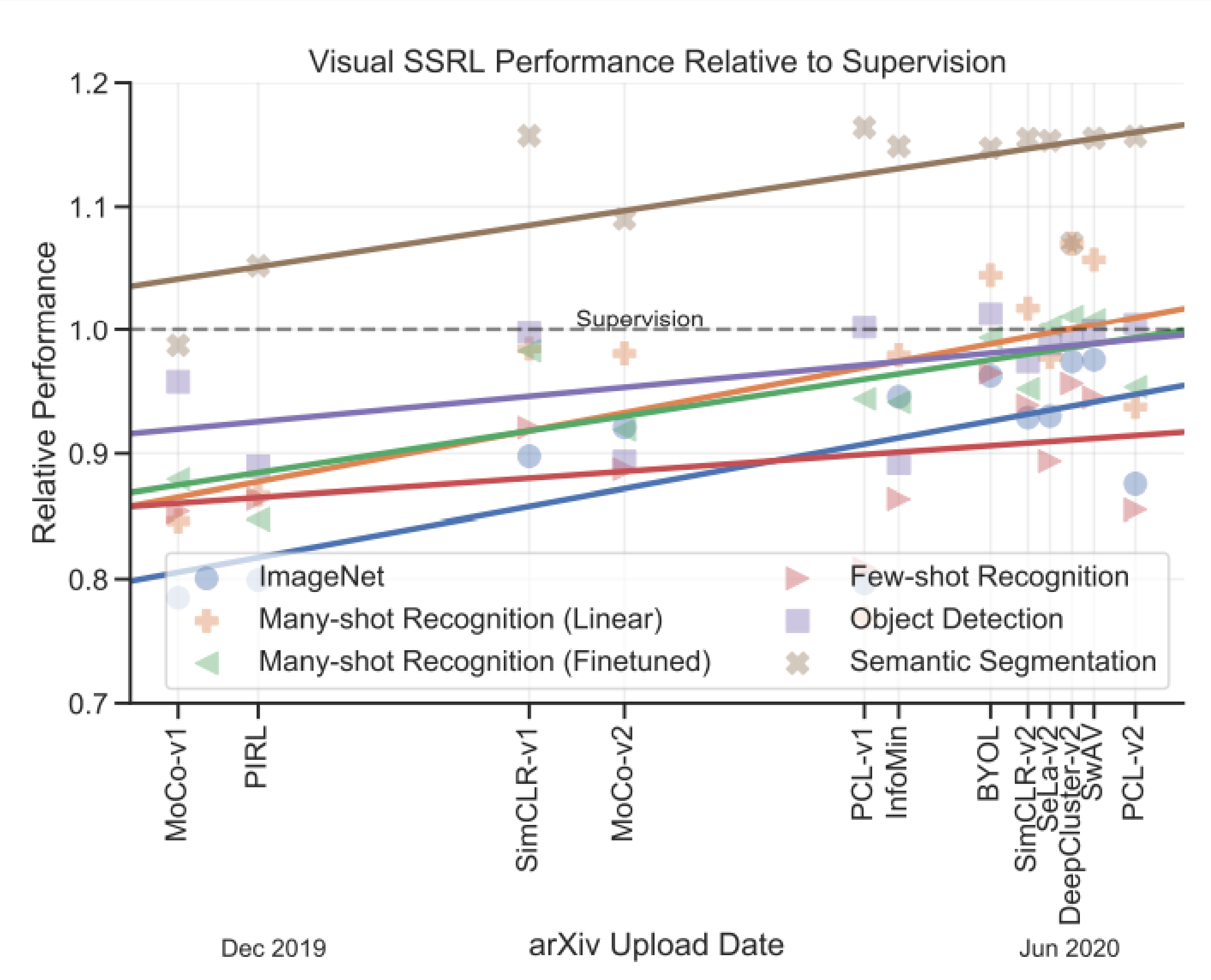
Self-supervised learning (SSL)

- It is an evolving machine learning technique to solve the challenges posed by the over-dependence of labeled data
- A special type of representation learning via unlabeled data
- Model trains itself to learn one part of the input from another part of the input

Why do we need SSL?

- High cost - The cost of good quality labeled data is very high in terms of time and money
- Lengthy lifecycle - The preparation lifecycle is a long process including data clean, annotation, review, and reconstruction

Self-supervised Learning
Introduction



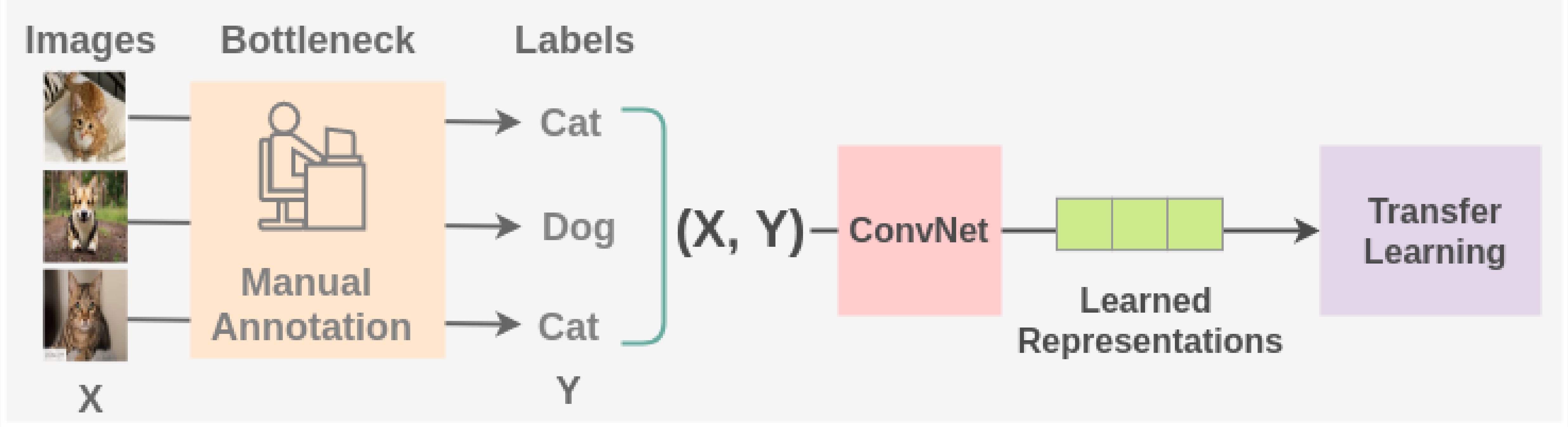
Reference

Self-supervised Learning

Introduction

The workflow of SSL

- ➢ Training with unlabeled data to obtain a general representation
- ➢ Fine-tuning with few labeled data



Reference

Self-supervised Learning

Introduction

Approaches

- Generative
- Predictive
- Contrastive
- Bootstrapping
- Regularization

01

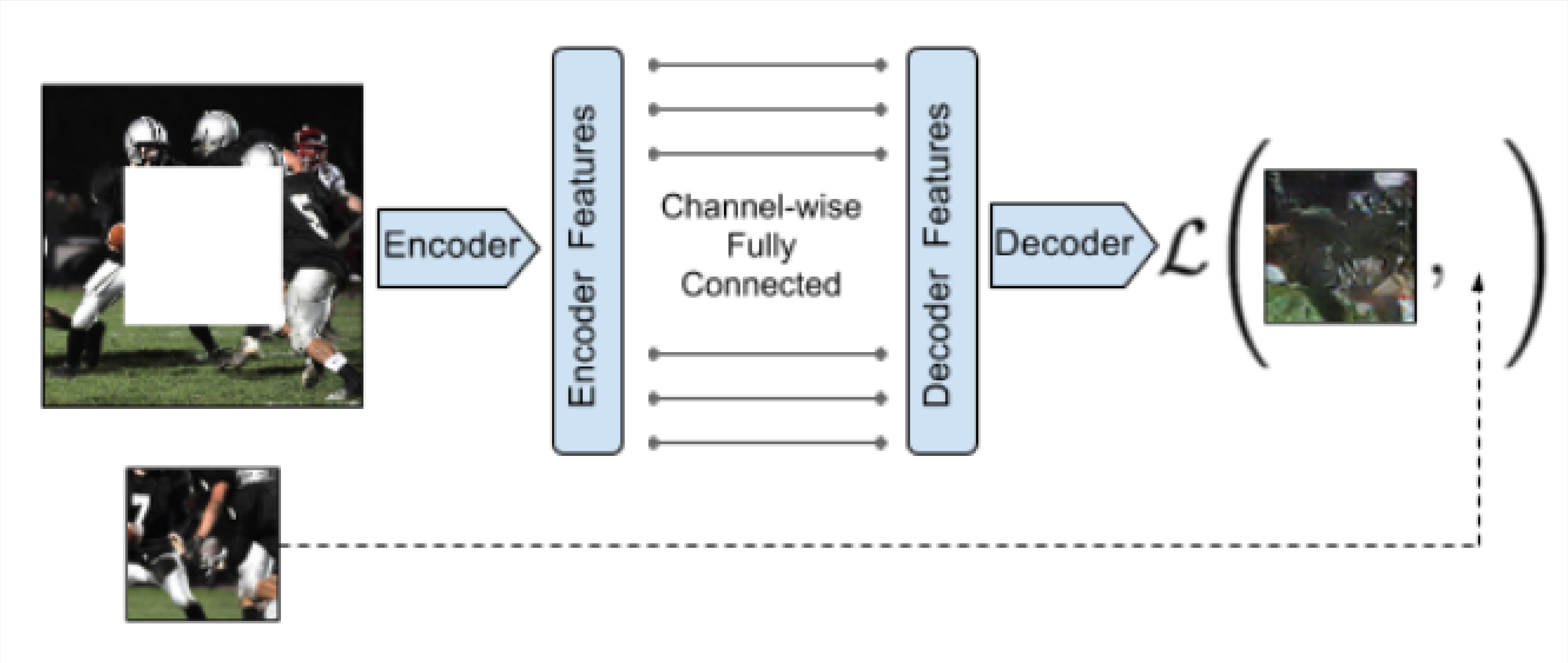
Generative

Self-supervised Learning

Generative

Generative

- Training model to reconstruct the pixel space
  - Image inpainting

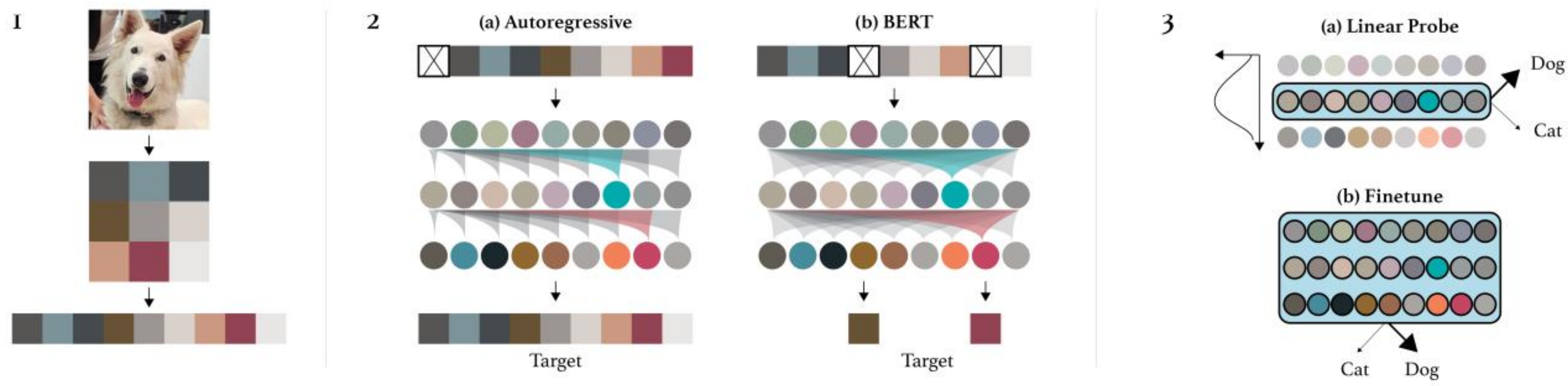


[Reference]

Self-supervised Learning
Generative

Generative

- ➢ Training model to reconstruct the pixel space
  - Image inpainting



Reference

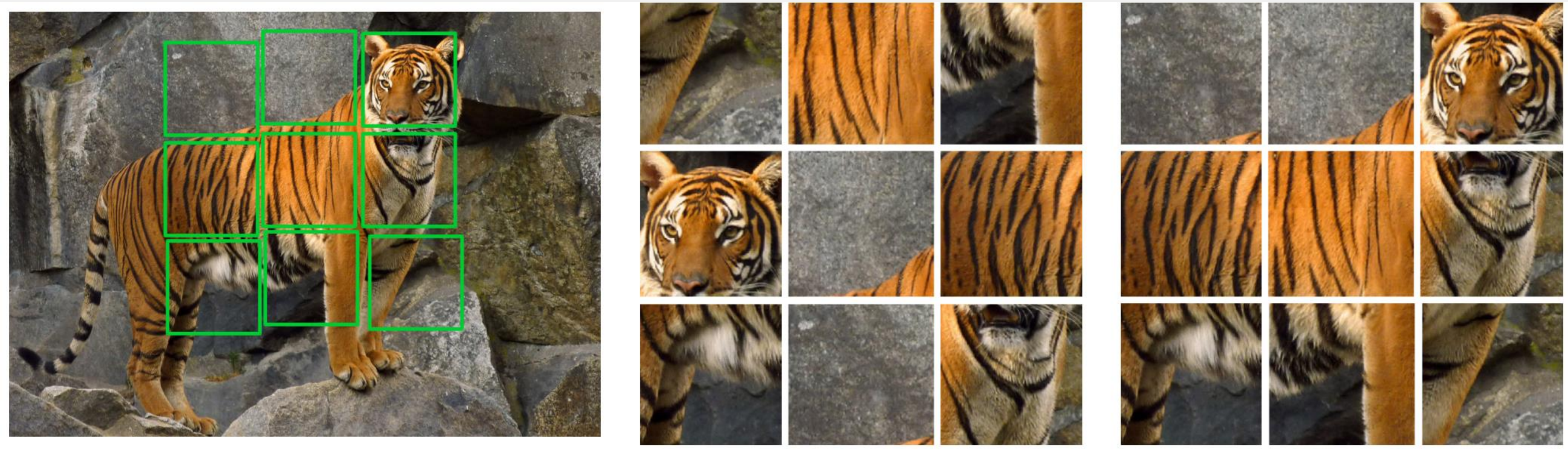
02

Predictive

Self-supervised Learning
Predictive

Predictive

- "Change" and "recovery" image without pixel generation
  - High-level representation generation based on pixel is a hard task
  - Context prediction



(a) (b) (c)

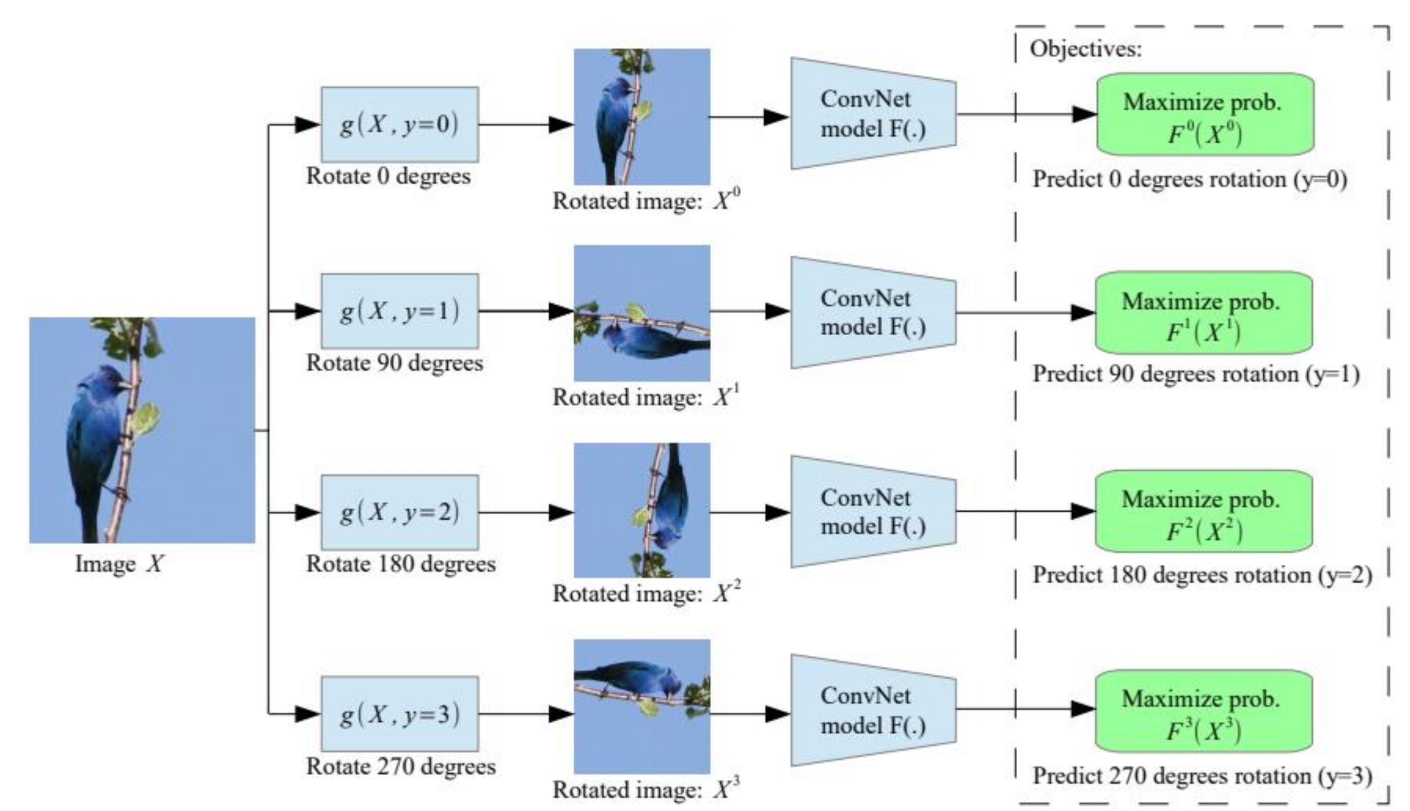
Reference

Self-supervised Learning

Predictive

Predictive

- "Change" and "recovery" image without pixel generation
  - High-level representation generation based on pixel is a hard task
  - Context prediction

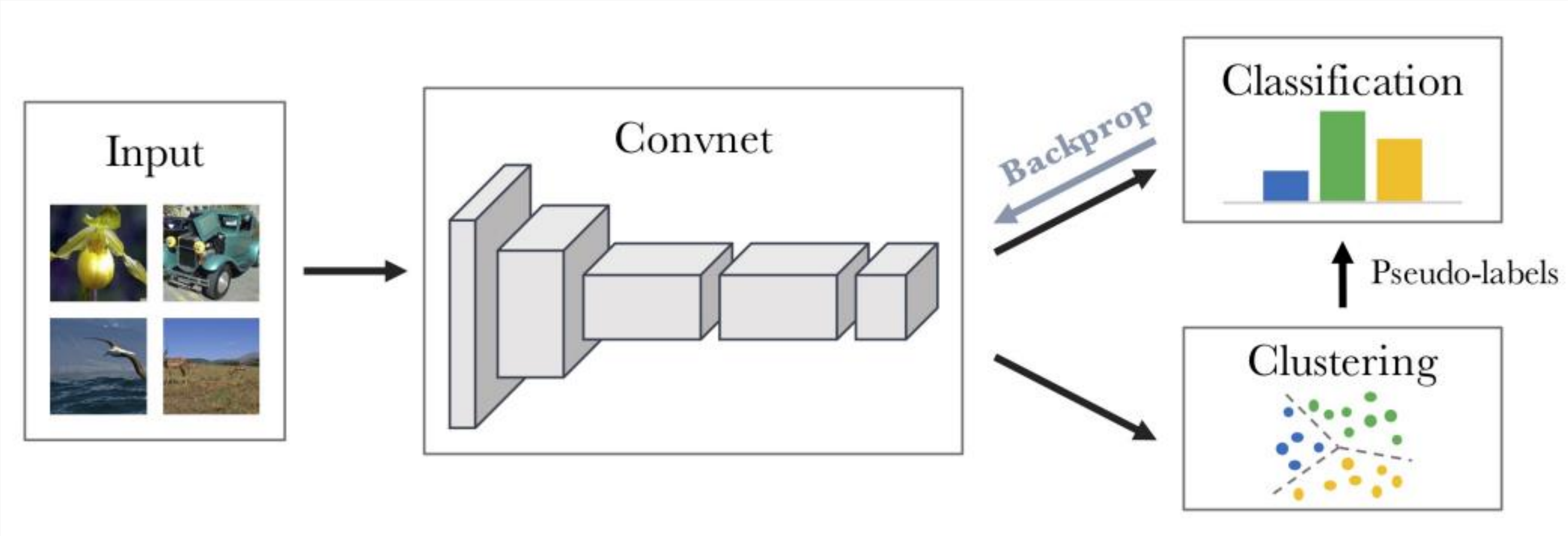


Reference

Self-supervised Learning

Predictive

- "Change" and "recovery" image without pixel generation
  - High-level representation generation based on pixel is a hard task
  - Context prediction



Reference

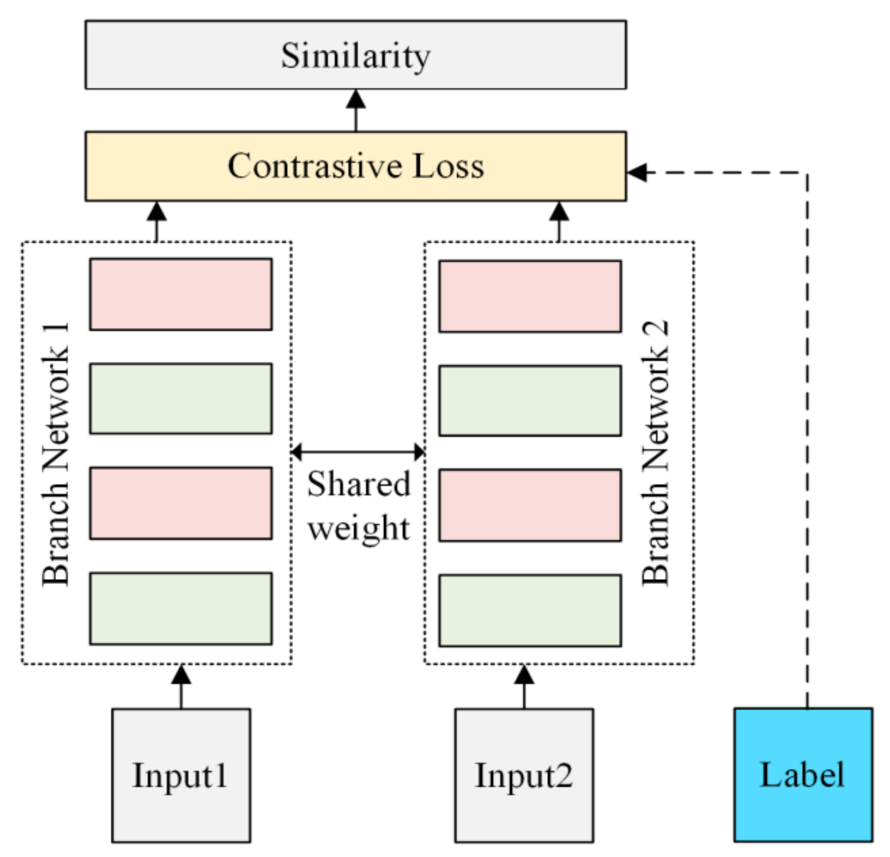
03

Contrastive

Self-supervised Learning
Contrastive

Contrastive

- A widely used approach in SSL
- The higher similarity between images of same class is the better
  - Siamese network



Reference

Self-supervised Learning

Contrastive

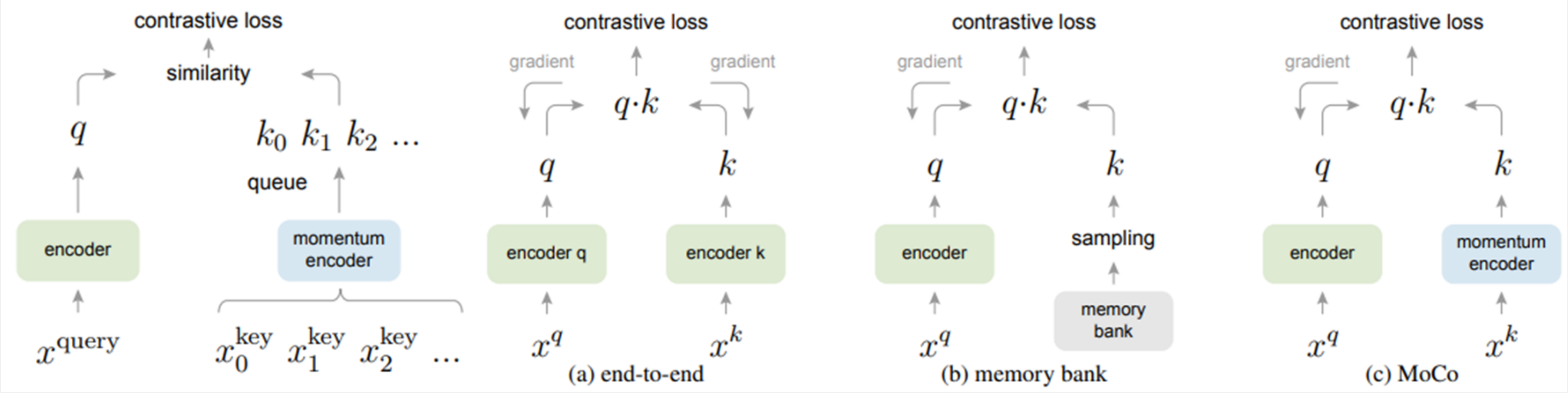
Contrastive

[Reference](#)

Self-supervised Learning
Contrastive

Contrastive - MoCov1

- Dictionary as a queue
  - Enqueue a batch representation and dequeue the oldest representation
- Momentum encoder
  - Keep queue dictionary data consistent
- Shuffling BN
  - shuffle the data order before training and recovery the order after extracting representation

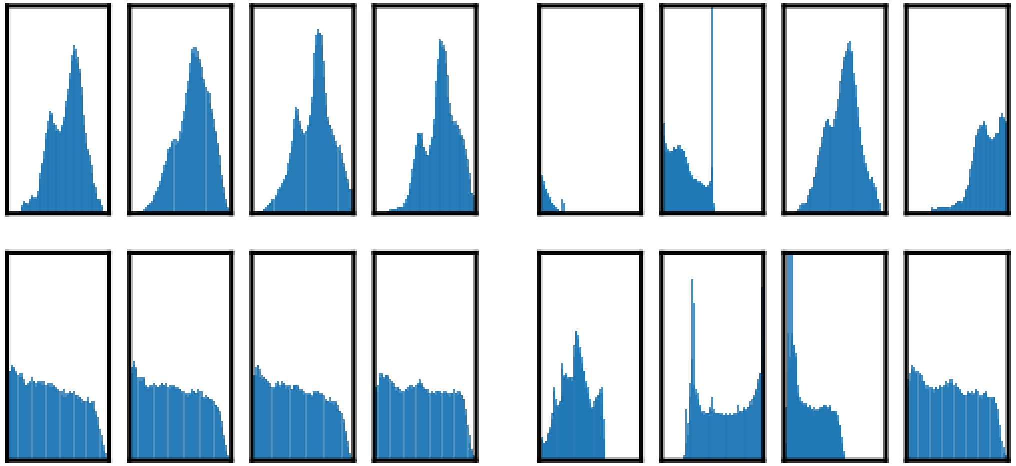


[Reference](#)

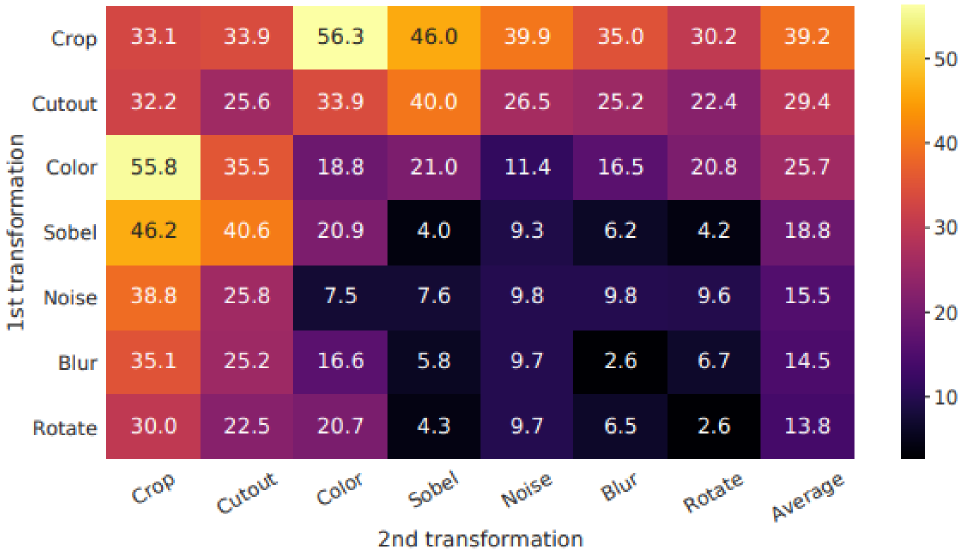
Self-supervised Learning
Contrastive

Contrastive - SimCLRv1

- Data augmentation combination
- Projection head
- NT-Xent loss function



(a) Without color distortion. (b) With color distortion.

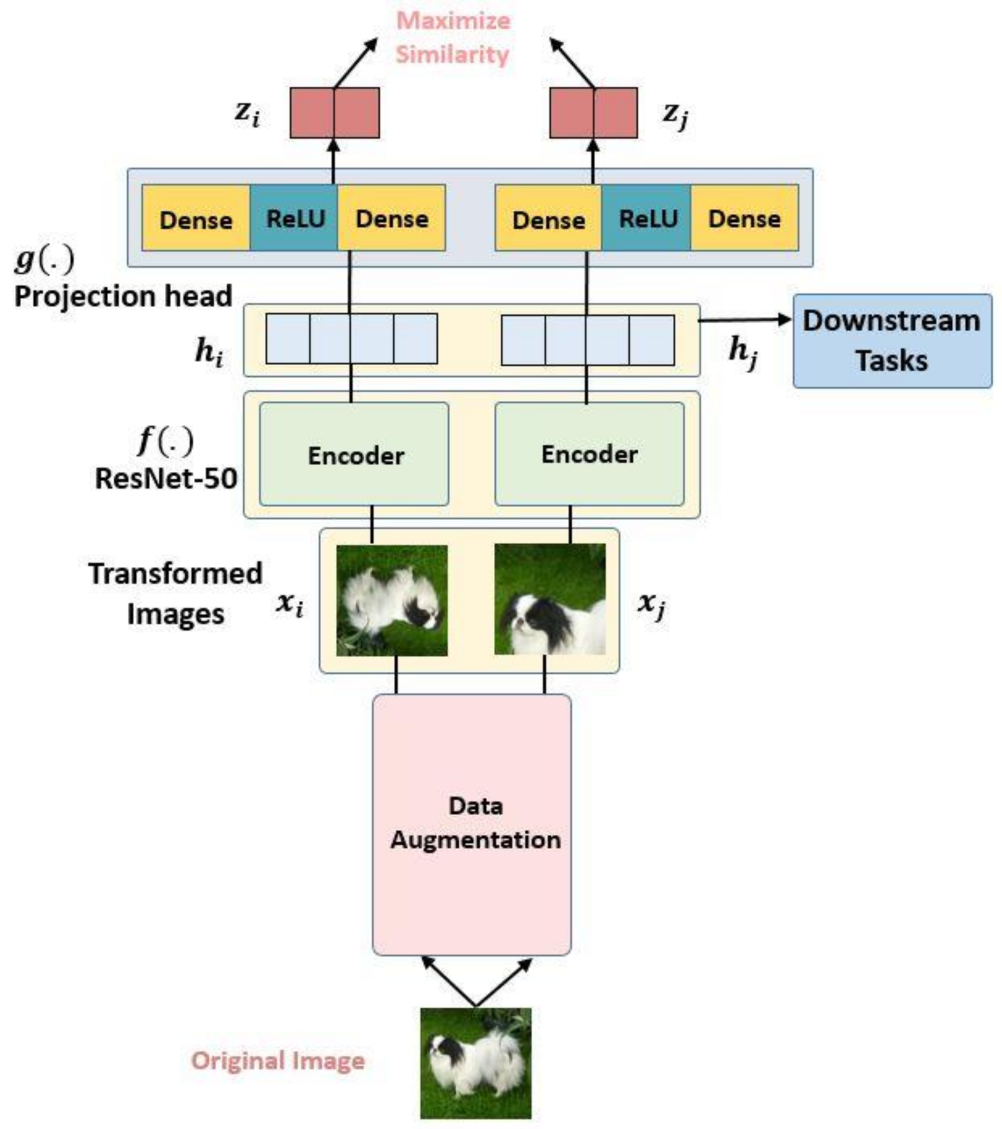


Reference

Self-supervised Learning
Contrastive

Contrastive - SimCLRv1

- Projection head



[Reference]

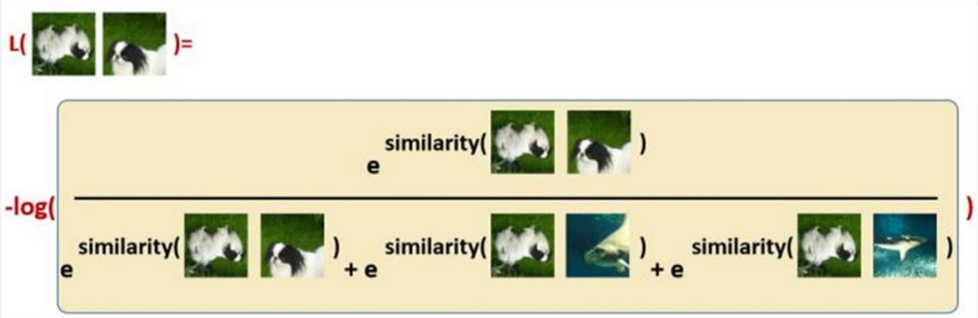
Self-supervised Learning
Contrastive

Contrastive - SimCLRv1

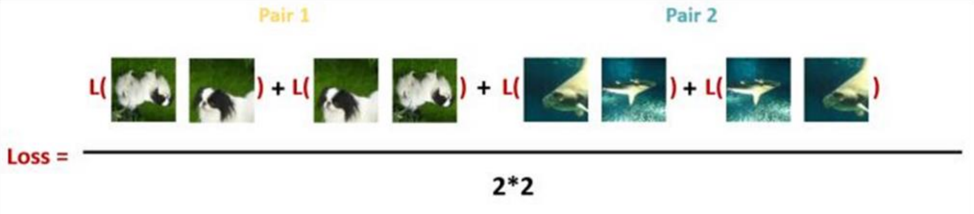
- NT-Xent loss function

$$s_{i,j} = \boldsymbol{z}_i^\top \boldsymbol{z}_j / (\|\boldsymbol{z}_i\| \|\boldsymbol{z}_j\|) \qquad \text{\# pairwise similarity}$$

$$\ell(i,j) \textbf{ as } \ell(i,j) = -\log \frac{\exp(s_{i,j}/\tau)}{\sum_{k=1}^{2N} \mathbb{1}_{[k \neq i]} \exp(s_{i,k}/\tau)}$$



$$\mathcal{L} = \frac{1}{2N} \sum_{k=1}^{N} [\ell(2k-1, 2k) + \ell(2k, 2k-1)]$$



Reference

04

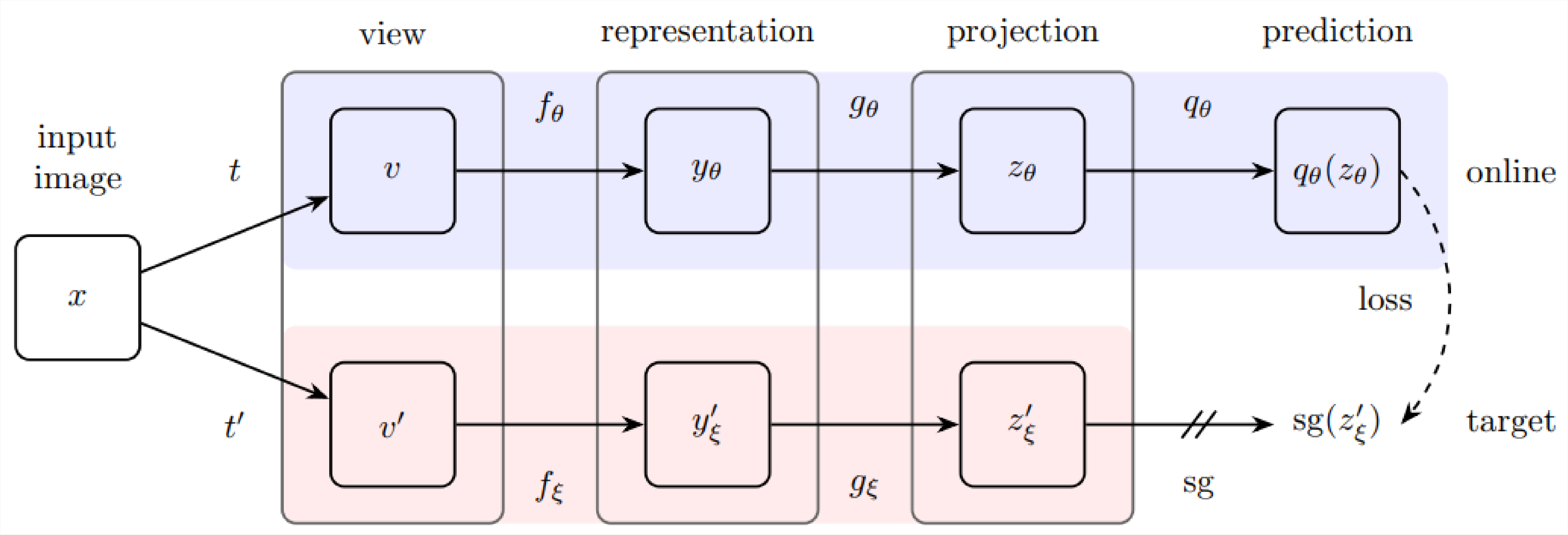
Bootstrapping

Bootstrapping

- ➢ In contrastive methods, the negative samples selection is a hard problem
  - The contribution of negative samples is to avoid model collapse
- ➢ How to training without negative samples?
  - BYOL
  - SimSiam

Self-supervised Learning
Bootstrapping

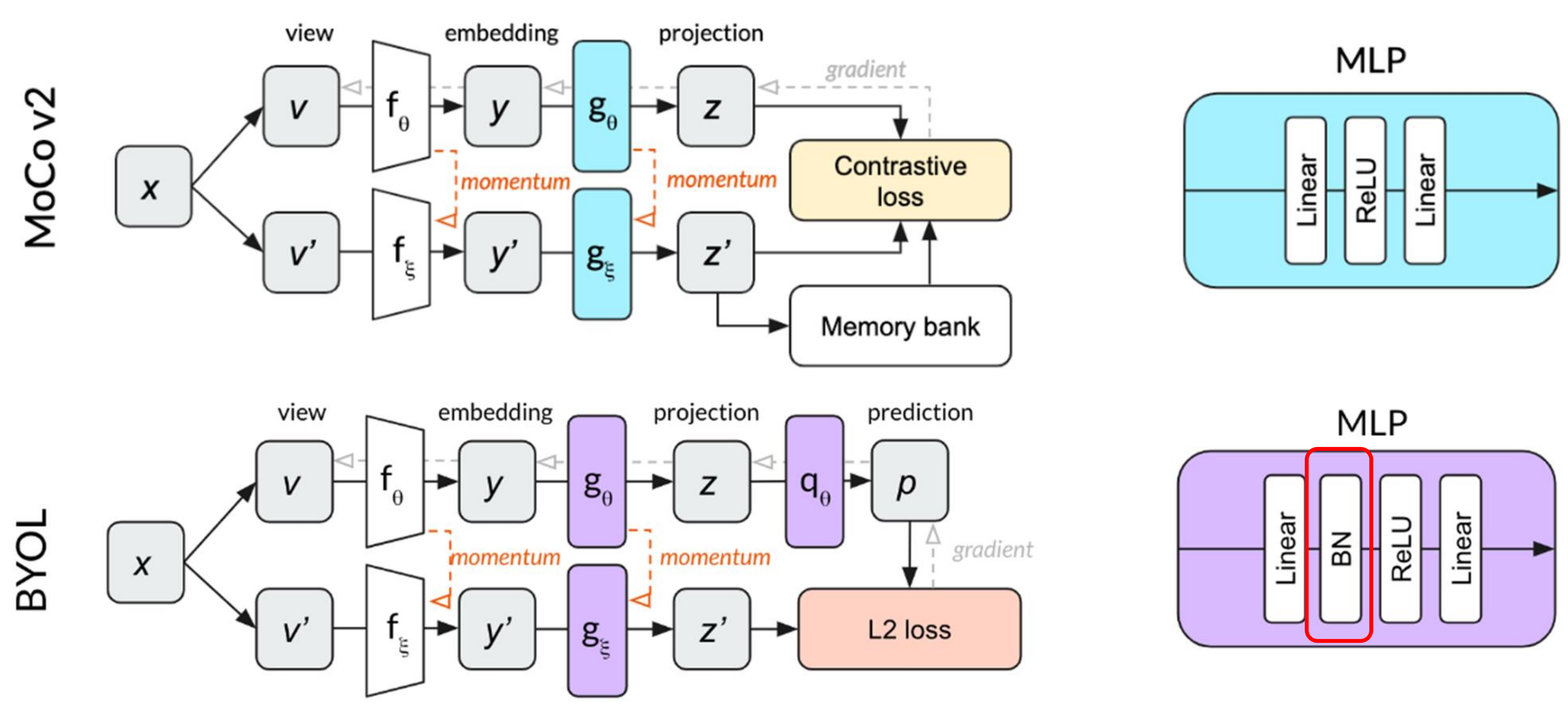
BYOL (Bootstrap your own latent)



Reference

Self-supervised Learning
Bootstrapping

BYOL (Bootstrap your own latent)

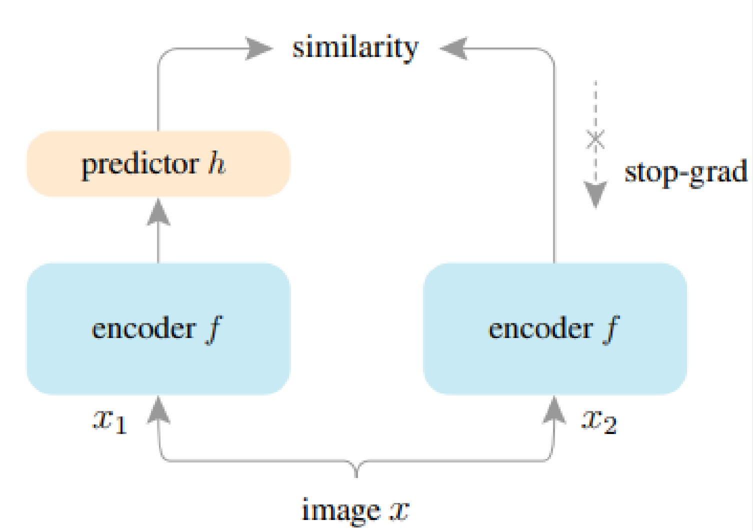


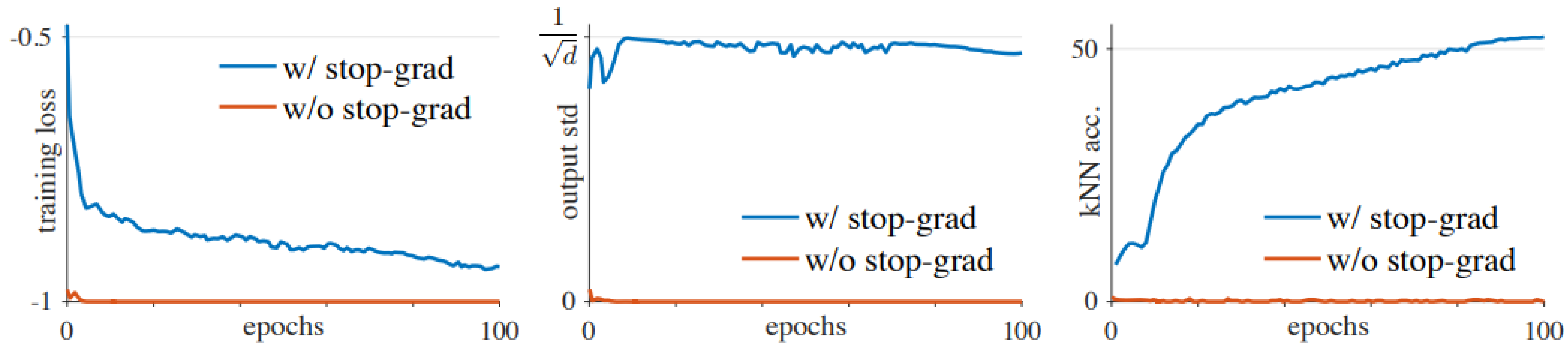
Reference

Self-supervised Learning

Bootstrapping

SimSiam (Simple siamese)





	acc. (%)
w/ stop-grad	67.7±0.1
w/o stop-grad	0.1

[Reference](#)

05

Regularization

Self-supervised Learning

Regularization

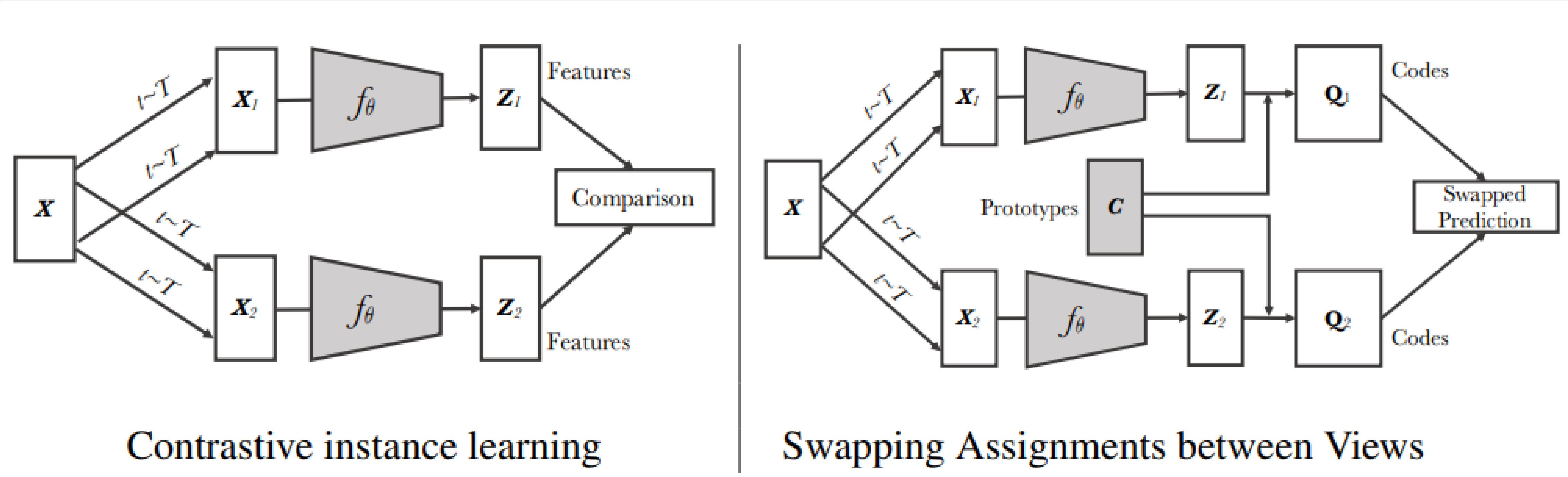
Simple extra regularization

- Training also without negative samples
- Representation mining with regularization while training
    - SwAV
    - Barlow twins

Self-supervised Learning

Regularization

SwAV (Swapping assignments between views)

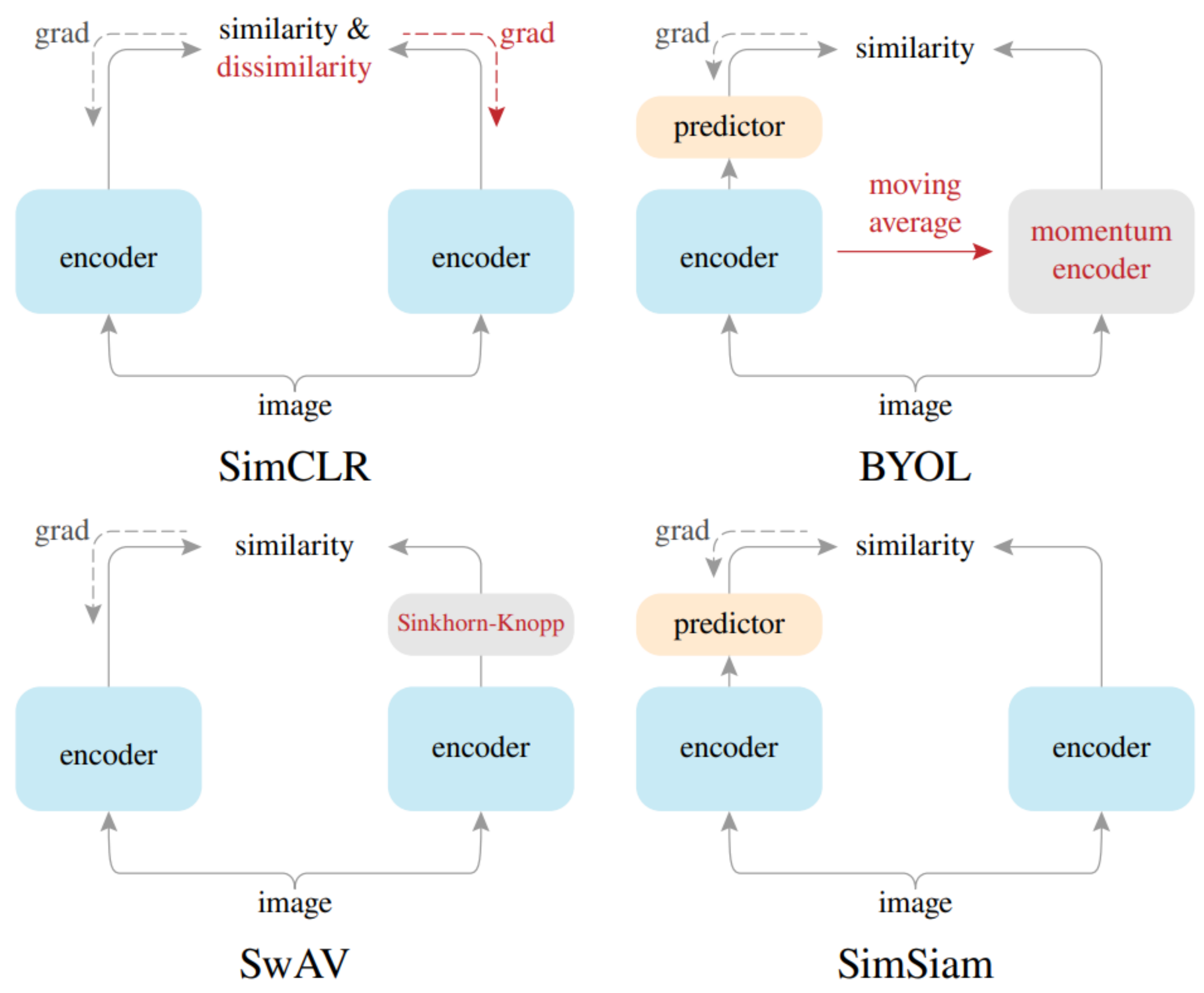


Reference

Self-supervised Learning

Regularization

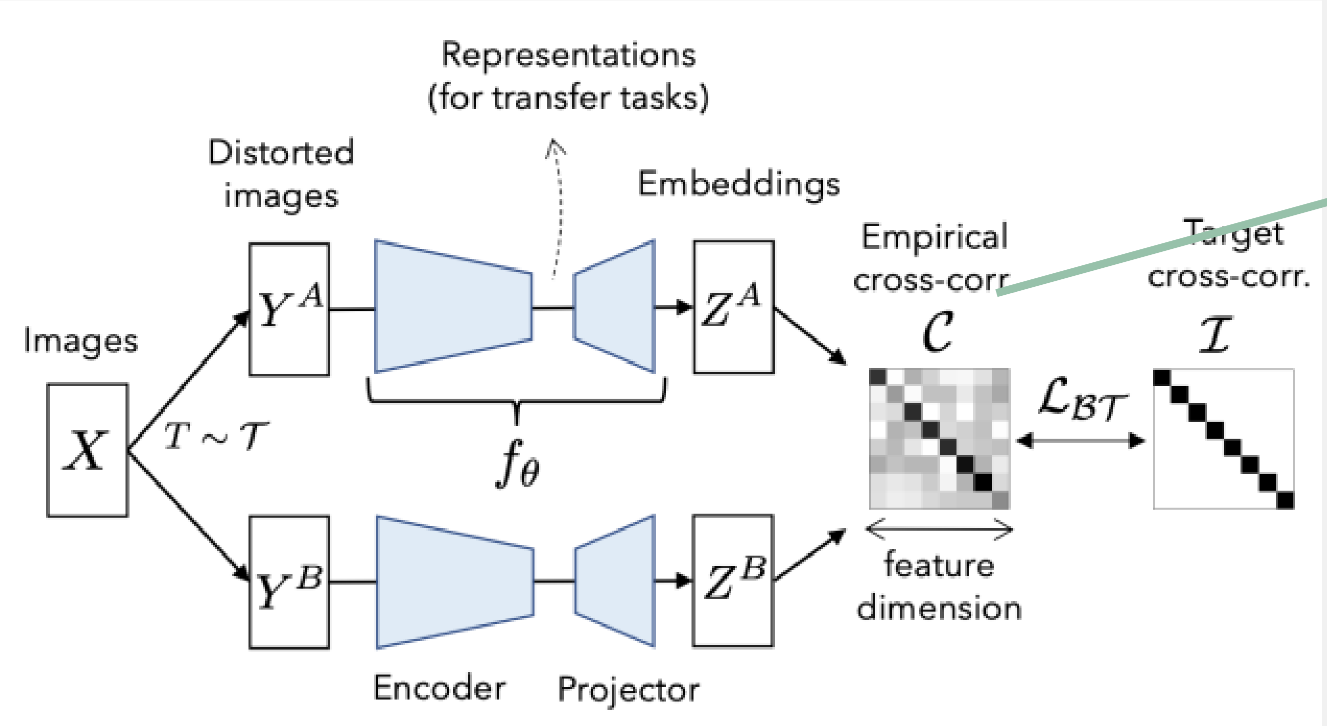
SwAV (Swapping assignments between views)



[Reference](#)

Self-supervised Learning
Regularization

Barlow Twins



$$\mathcal{C}_{ij} \triangleq \frac{\sum_b z^A_{b,i} z^B_{b,j}}{\sqrt{\sum_b \left(z^A_{b,i}\right)^2} \sqrt{\sum_b \left(z^B_{b,j}\right)^2}}$$

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