TRIBYOL: TRIPLET BYOL FOR SELF-SUPERVISED REPRESENTATION LEARNING

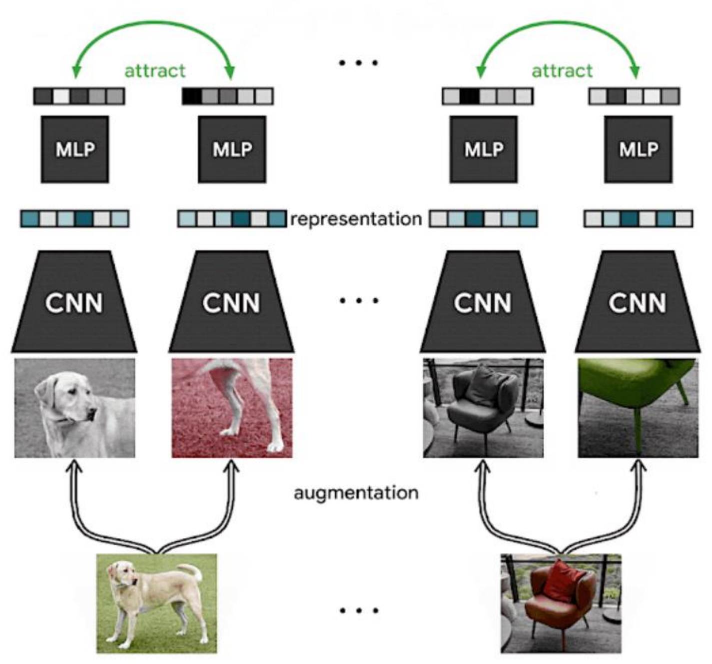
Guang Li, *Ren Togo, Takahiro Ogawa and Miki Haseyama* *Hokkaido University*

icassp 2022 Singapore

Paper ID: 1917

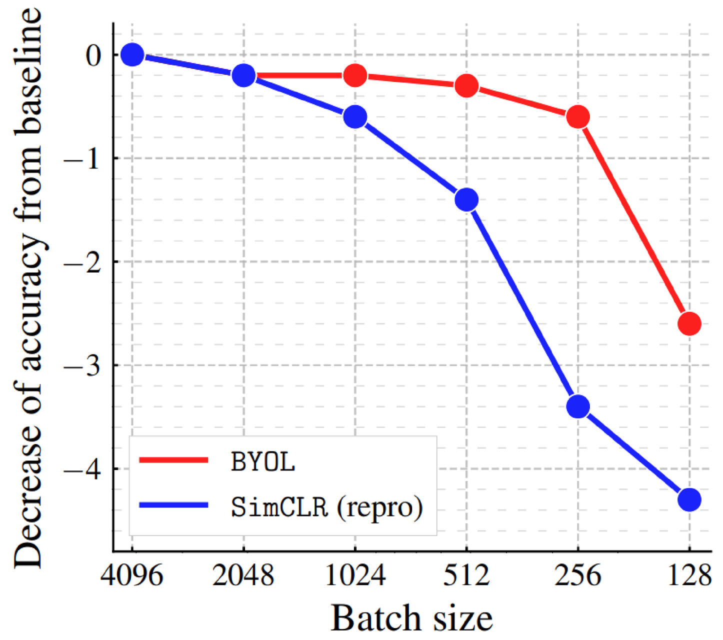
INTRODUCTION

Self-supervised learning based on data augmentation



- Self-supervised learning is a means for pre-training networks to learn good representations without human providing labeled data.
- Self-supervised learning based on data augmentation is the process of training a classifier to distinguish between "similar" and "dissimilar" input data.
- SimCLR and BYOL are two state-of-the-art self-supervised learning methods with this scheme.

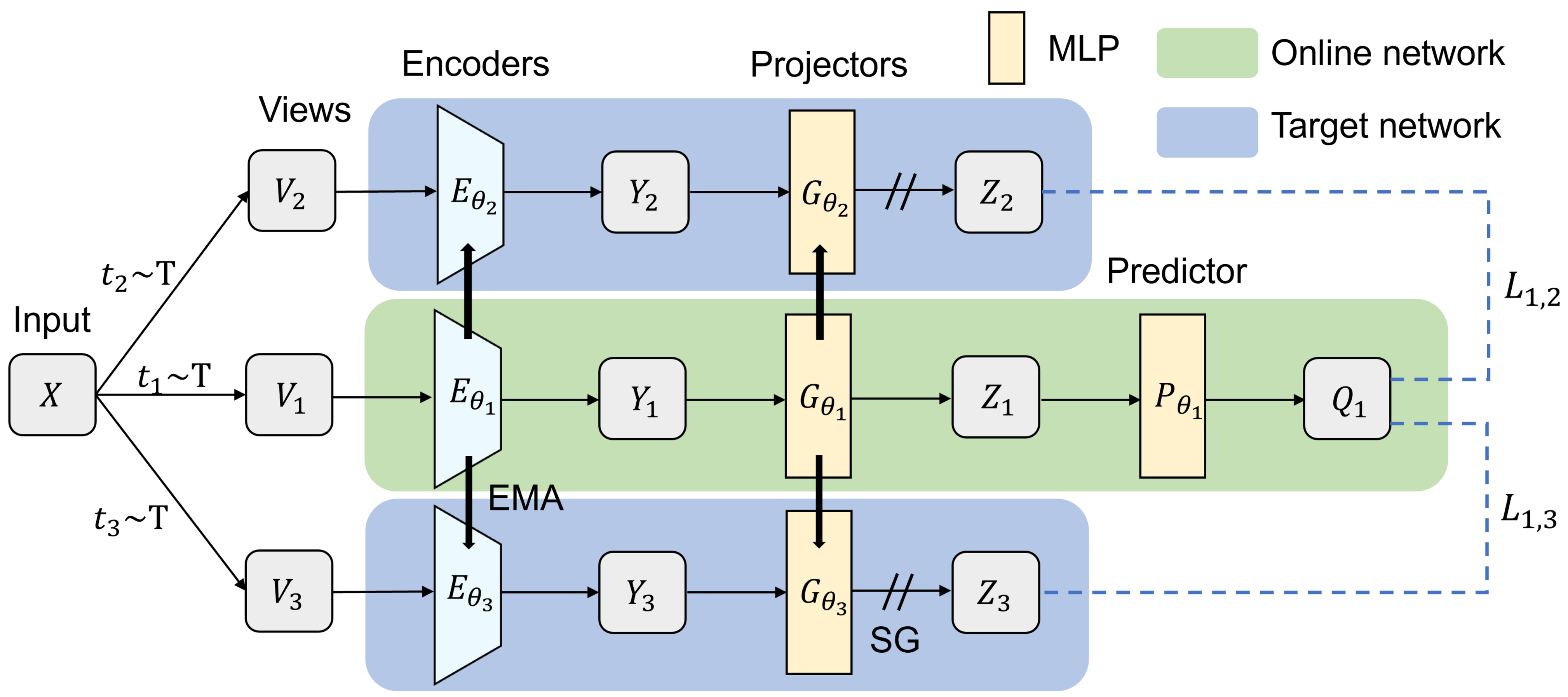
Accuracy degradation in small-batch cases



- Phenomenon: The accuracy of SimCLR and BYOL drastically decreases as the batch size decreases.
- Reason: When batch size decreases, these methods can not learn enough semantic information from limited views.
- Problem: Some real-world images, such as medical and remote sensing images, are high-resolution and can only train in small-batch cases.

High-accuracy self-supervised learning in small-batch cases is needed.

PROPOSED METHOD



- L: MSE loss of normalized predictions and projections
- //: stop-gradient
- : exponential moving average
- SG: stop gradient
- EMA: exponential moving average
- Q: predictions (downscaled image features)
- P: predictor (multilayer perceptron)
- Z: projections (downscaled image features)
- G: projectors (multilayer perceptron)
- Y: extracted image features
- E: encoders (backbone)
- V: augmented views
- t: transformations randomly sampled from distribution T
- X: input chest X-ray image

- Different from BYOL which uses the Siamese network, we propose the triplet network combined with a triple-view loss for learning better representations with small batch sizes.
- **Novelty:** The addition of augmented views can increase mutual information and encourage a more transformation-invariant representation in small-batch cases.
- We confirm that our method can drastically outperform state-of-the-art self-supervised learning methods on several datasets in small-batch cases.

Our method can learn sufficient semantic information from images in small-batch cases.

EXPERIMENTAL RESULTS

Dataset

Eight benchmark datasets.

- MNIST
- FashionMNIST
- KMNIST
- USPS
- SVHN
- CIFAR-10
- CIFAR-100
- STL-10

Settings

Self-supervised learning:

- Encoder: ResNet50
- MLP hidden size: 512
- MLP projection size: 128
- Batch size: 32, 64, 128
- Moving average: 0.996
- Training epoch: 80

Comparison methods

Six state-of-the-art (SOTA) self-supervised learning methods and two supervised learning methods.

- Cross, BYOL, SimSiam, PIRL-Jigsaw, PIRL-Rotation, SimCLR
- Supervised transfer learning from ImageNet, Supervised learning from scratch

Linear evaluation results with different batch sizes

	CIFAR-10			CIFAR-100			STL-10		
Method	b32	b64	b128	b32	b64	b128	b32	b64	b128
TriBYOL	**79.09**	**85.35**	**87.31**	**49.07**	**59.90**	**63.05**	**75.41**	**83.16**	**88.19**
Cross	76.01	82.06	83.50	48.04	54.65	58.79	69.66	78.38	83.79
BYOL	68.67	81.47	83.79	41.21	49.68	58.34	49.60	80.09	84.88
SimSiam	58.42	71.25	75.58	1.00	37.06	49.21	10.00	65.20	71.78
PIRL-rotation	-	-	55.78	-	-	31.55	-	-	50.26
PIRL-jigsaw	-	-	49.94	-	-	27.36	-	-	48.55
SimCLR	-	-	52.58	-	-	21.26	-	-	44.50
ImageNet		82.37			60.72			91.76	

Verified that our method was effective in small-batch cases.

Fine-tuning results with different numbers of labels (b128)

	CIFAR-10			CIFAR-100			STL-10		
Method	1%	10%	100%	1%	10%	100%	1%	10%	100%
TriBYOL	**56.60**	**71.73**	**87.07**	**9.50**	**23.57**	**58.92**	**56.66**	**67.72**	**97.34**
Cross	50.88	67.34	86.03	6.81	20.96	57.23	42.80	59.22	93.28
BYOL	56.28	70.94	86.87	9.38	22.51	58.17	53.96	65.98	97.26
SimSiam	43.16	63.08	84.76	4.86	14.76	54.70	40.38	49.96	88.38
From Scratch	32.29	57.24	83.87	5.95	17.47	56.70	20.38	39.10	83.96
ImageNet	69.99	84.27	91.29	27.48	52.41	70.80	81.24	86.34	98.00

Verified that our method was effective even using few training data.

Transfer learning results on different datasets (b128)

	MNIST	FashionMNIST	KMNIST	USPS	SVHN	CIFAR-10	CIFAR-100	STL-10
TriBYOL	**98.74**	**91.76**	**92.00**	**96.61**	**75.23**	**80.09**	**55.88**	**79.11**
Cross	98.54	91.28	90.33	96.21	71.29	77.55	51.53	76.26
BYOL	98.41	90.77	89.88	96.06	68.75	75.31	48.51	74.04
SimSiam	97.58	89.25	83.31	94.02	58.70	64.51	35.63	63.44
ImageNet	98.58	90.85	90.77	96.56	77.34	82.37	60.72	91.76

Verified that our method was effective for transfer learning.

This study was supported in part by AMED Grant Number JP21zf0127004.
This study was conducted on the Data Science Computing System of Education and Research Center for Mathematical and Data Science, Hokkaido University.