A PyTorch-like Pseudocode

Algorithm 1: PyTorch-like pseudocode for MYOL loss computation

Input: Input views `v1, v2`

```
p1, p2 = pred(encoder(v1)), pred(encoder(v2)) # online network
z1, z2 = encoder_t(v1), encoder_t(v2) # target network
# input mixup
lam = Beta(alpha, alpha).sample()
randidx = randperm(len(v1))
mixed_v = lam * v1 + (1-lam) * v2[randidx]

# norm_mse: normalized mean squared error
loss_byol = (norm_mse(p1, z2) + norm_mse(p2, z1)) / 2

# projection mixup
mixed_z = lam * z1 + (1-lam) * z2[randidx]
loss_myol = norm_mse(pred(encoder(mixed_v)), mized_z)

loss = loss_byol + loss_myol
```

B Self-Supervised Learning Methods Comparison

In this section, we compare BYOL (Grill *et al.* 2020) with other self-supervised learning methods SimCLR (Chen *et al.* 2020) and MoCo (He *et al.* 2020). SimCLR is a simple contrastive learning method that uses a large batch size to ensure a sufficient number of negative pairs. MoCo, on the other hand, addresses the negative sample challenge by utilizing a memory bank. While both SimCLR and MoCo serve as baseline methods for Un-Mix (Shen *et al.* 2022) and i-Mix (Lee *et al.* 2021), our experiments show that BYOL is a sufficient baseline for achieving efficient and robust self-supervised learning on small images.

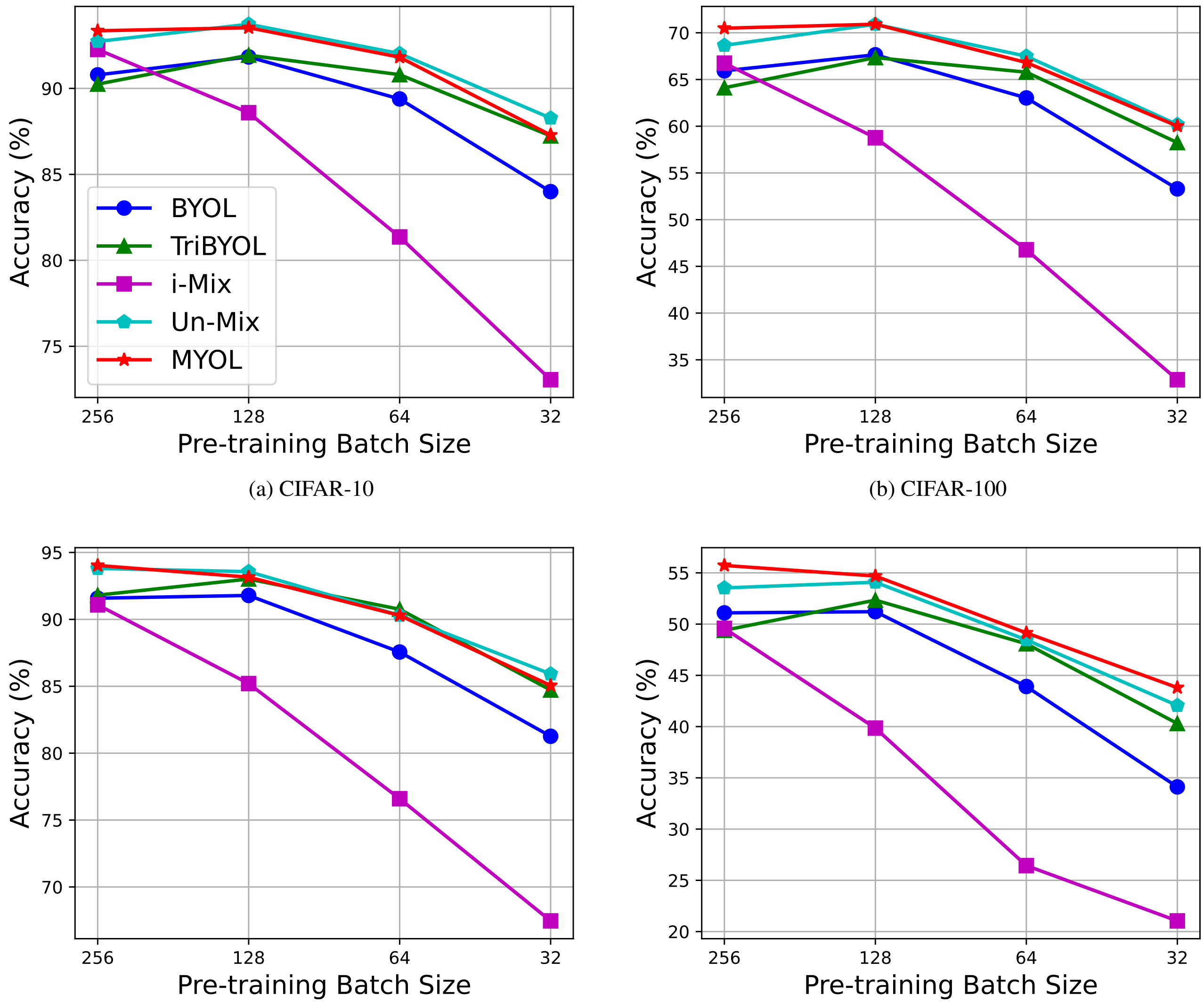
Table 1: Comparison of SimCLR, MoCo, and BYOL in terms of linear evaluation performance with varying pre-training batch sizes in terms of classification accuracy (%).

Pre-training Batch Size	Method	CIFAR-10	CIFAR-100	STL-10	Tiny-ImageNet
256	SimCLR	87.14±0.15	62.96±0.34	83.21±0.12	45.92±0.13
	MoCo	88.08±0.28	64.07±0.37	85.17±0.07	48.37±0.16
	BYOL	**90.79±0.12**	**65.93±0.22**	**91.58±0.32**	**51.10±0.08**
128	SimCLR	84.49±0.16	58.99±0.33	80.02±0.22	42.82±0.40
	MoCo	85.61±0.05	61.42±0.42	82.60±0.03	46.43±0.06
	BYOL	**91.83±0.14**	**67.65±0.34**	**91.79±0.03**	**51.21±0.13**
64	SimCLR	80.05±0.43	54.01±0.55	76.15±0.39	38.41±0.25
	MoCo	82.67±0.34	58.35±0.31	79.53±0.16	42.91±0.30
	BYOL	**89.39±0.24**	**63.03±0.19**	**87.56±0.57**	**43.91±0.24**
32	SimCLR	75.60±0.27	46.83±0.50	70.23±0.17	31.19±0.19
	MoCo	79.22±0.01	**54.37±0.03**	75.89±0.42	**39.35±0.34**
	BYOL	**84.00±0.47**	53.30±0.24	**81.26±0.29**	34.12±1.53

C Plots for the Entire Datasets

In this section, we show the plots for the entire datasets that were used in the experiments, as well as the plots for the Tiny-ImageNet dataset.

C.1 Pre-training Batch Sizes Comparison

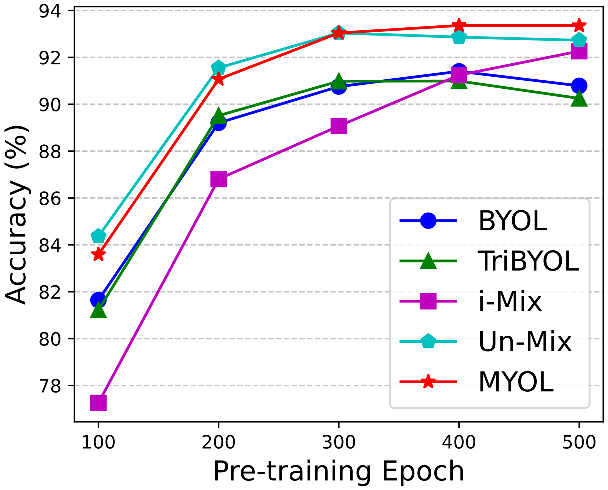


(a) CIFAR-10 (b) CIFAR-100

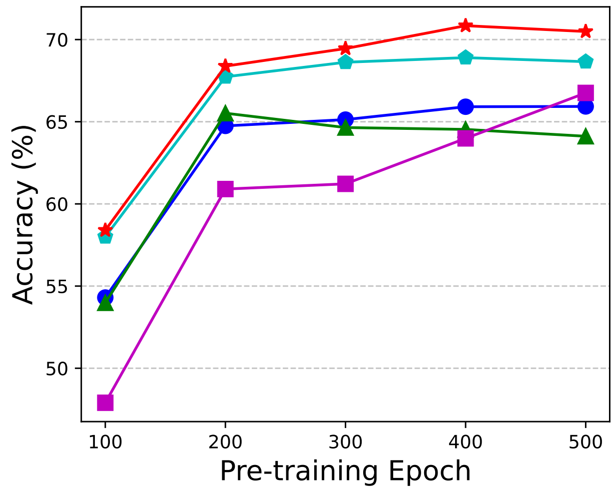
(c) STL-10 (d) Tiny-ImageNet

Figure 1: Comparison of linear evaluation performance according to the pre-training batch sizes in terms of classification accuracy (%).

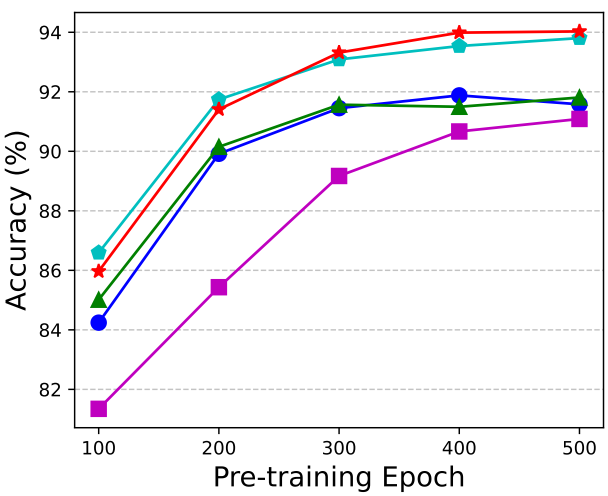
C.2 Pre-training Epochs Comparison



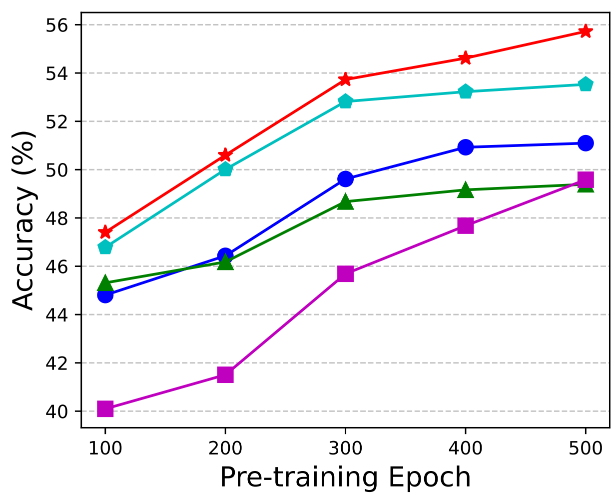
(a) CIFAR-10



(b) CIFAR-100



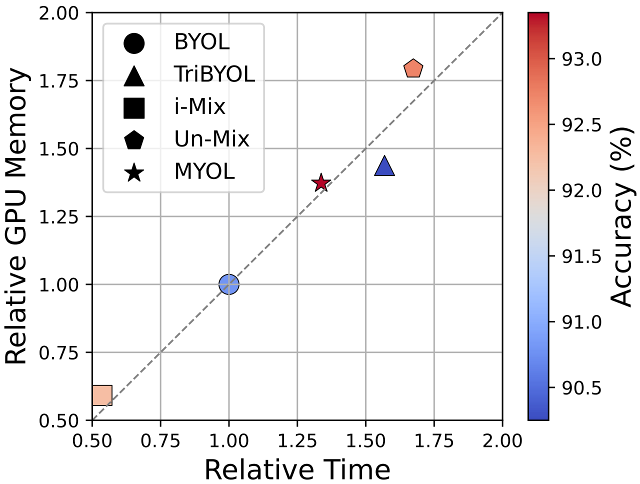
(c) STL-10



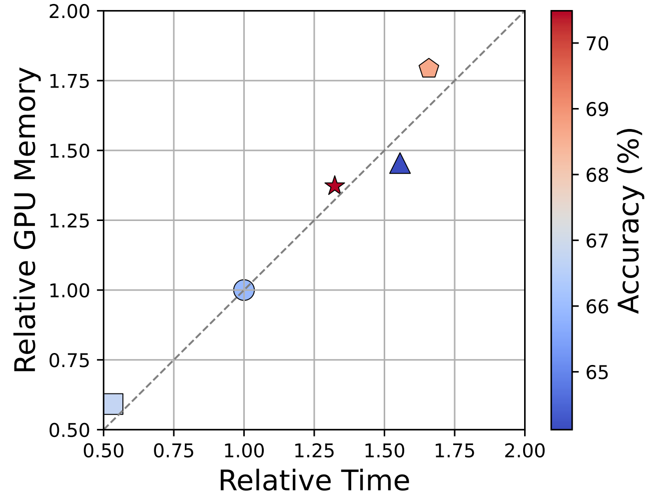
(d) Tiny-ImageNet

Figure 2: Comparison of linear evaluation performance according to the number of pre-training epochs in terms of classification accuracy (%).

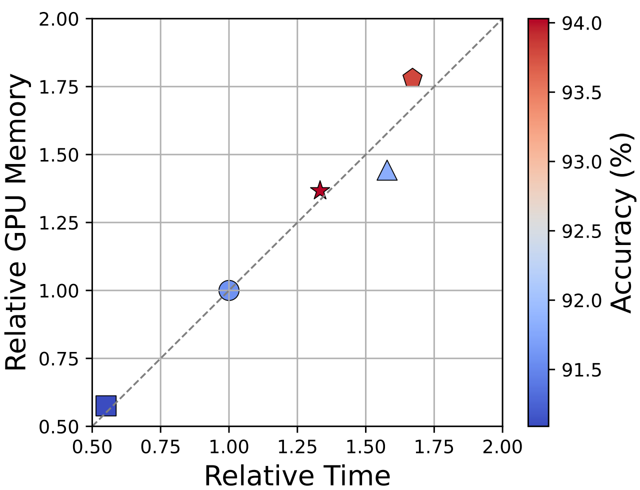
C.3 Time and GPU Memory Comparison



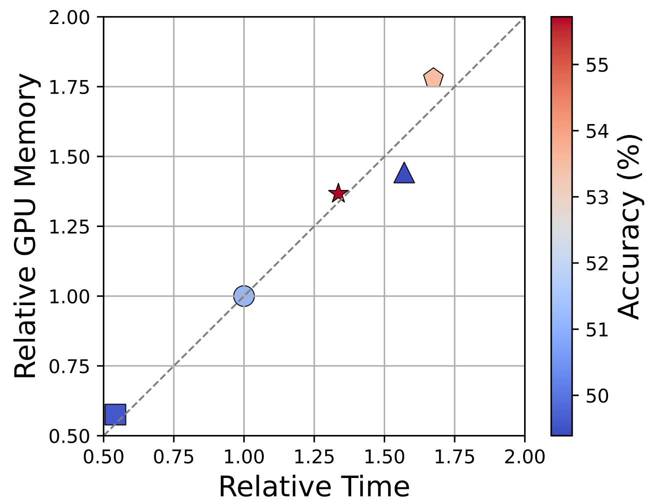
(a) CIFAR-10



(b) CIFAR-100



(c) STL-10



(d) Tiny-ImageNet

Figure 3: Comparison of time and GPU memory required for pre-training on a relative scale, with marker color indicating classification accuracy.

References

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *Proceedings of International Conference on Machine Learning*, volume 119, pages 1597–1607, 2020.

Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre H. Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pires, Zhaohan Daniel Guo, Mohammad Gheshlaghi Azar, Bilal Piot, Koray Kavukcuoglu, Rémi Munos, and Michal Valko. Bootstrap your own latent-a new approach to self-supervised learning. In *Advances in Neural Information Processing Systems*, volume 33, pages 21271–21284, 2020.

Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9729–9738, 2020.

Kibok Lee, Yian Zhu, Kihyuk Sohn, Chun-Liang Li, Jinwoo Shin, and Honglak Lee. i-Mix: A domain-agnostic strategy for contrastive representation learning. In *Proceedings of International Conference on Learning Representations*, 2021.

Zhiqiang Shen, Zechun Liu, Zhuang Liu, Marios Savvides, Trevor Darrell, and Eric Xing. Un-Mix: Rethinking image mixtures for unsupervised visual representation learning. In *Proceedings of AAAI Conference on Artificial Intelligence*, volume 36, pages 2216–2224, 2022.