Branching Out for Better BYOL

Azad Singh, Deepak Mishra
Indian Institute of Technology Jodhpur

NeurIPS21 SSL Workshop

BYOL[1] leads migration of SSL techniques from contrastive to non-constative paradigm.

Non-contrastive approach that mitigates the inherent computational constraint imposed by contrastive methods.

Key factors of recent success in SSL are

- Stochastic data augmentation techniques
- Siamese configuration of deep neural networks

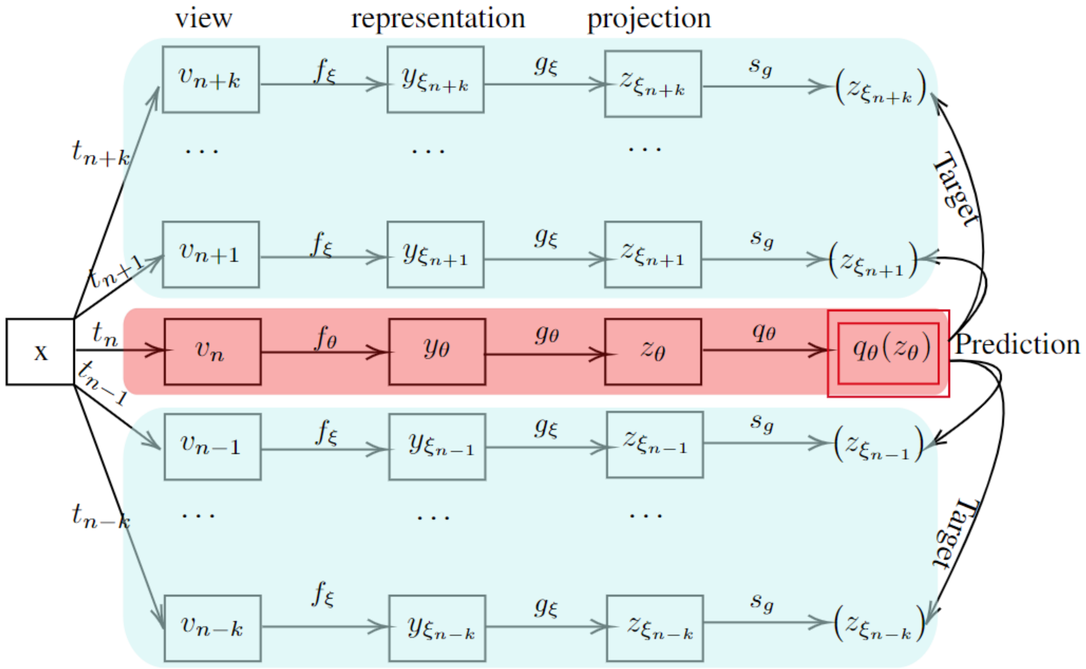
BYOL in its original form is limited to only two augmented views per training cycle.

This motivates us to extend BYOL from a single target network branch to multiple branches.

CONTRIBUTIONS

- We extend BYOL from a single target network for an online network to multiple target networks for simultaneous processing of multiple augmented views of an image.
- We show that MT-BYOL achieves considerably better performance as compared to BYOL with only marginally increasing of total computational cost
- We empirically show that Multi-Target BYOL is relatively more resilient to changes in batch size.
- We evaluate the representations learned by MT-BYOL under the linear evaluation protocols on various computer vision datasets and report the corresponding results.

METHOD



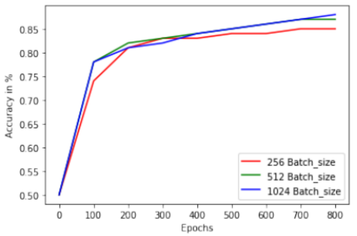
RESULTS

Method	BYOL			MT-BYOL(2)			MT-BYOL(3)		
Batchsize	256	512	1024	256	512	1024	256	512	1024
CIFAR10	85.46	87.59	88.34	90.29	90.51	91.31	90.64	91.19	**91.56**
CIFAR100	62.21	63.29	64.78	66.11	66.72	67.21	66.38	67.47	**67.58**
STL10	87.31	88.48	89.72	91.11	92.23	92.37	91.73	92.67	**92.71**
Tiny-ImageNet	54.46	55.79	56.63	56.54	56.78	57.12	56.73	57.03	**57.43**

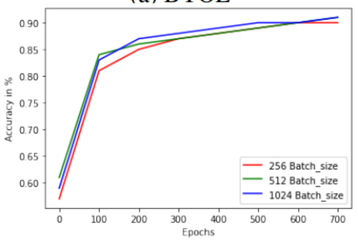
- Test set classification accuracy of linear classifier evaluated on embeddings generated by the frozen encoder for different datasets.

DISCUSSION

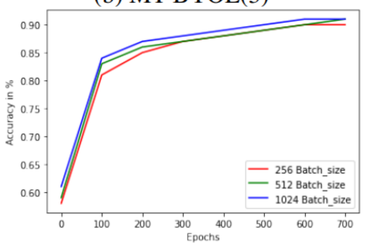
- Multiple target network branches considerably improve BYOL's performance across all the datasets.
- Marginal improvements in performance with MT-BYOL(3) over MT-BYOL(2).
- With 2 target network branches, 12 cross-model views are generate which provide enough regularization to the online network



(a) BYOL



(b) MT-BYOL(3)



(c) MT-BYOL(3)

- Plots shows the comparison of performance of BYOL with MT-BYOL(3) with different values of η.
- (a) Shows the convergence plot of BYOL.
- (b) Shows the performance of MT-BYOL(3) with identical initial value of η for all the target network branches.
- (c) Have the plot for different values η 's corresponding to different branches of target networks.
- Effect of variation in initial values of η is marginal which provide empirical evidences, that multiple augmented views are the major factor in the performance of MT-BYOL.
- Multiple branches of target network cost marginal computational overhead as their parameters are estimated by EMA of online network.
- In future we will explore the effect of further increasing the branches and other augmentation techniques.