Poster for Paper ID 43

Azad Singh, Deepak Mishra
Department of Computer Science and Engineering
Indian Institute of Technology Jodhpur
{singh.63, dmishra} @iitj.ac.in

BYOL leads migration of self-supervised learning techniques from contrastive to non-constative paradigm. Non-constative techniques do not require negative pairs to learn meaningful representations from the data. In this paper we extend BYOL from a single target network branch to multiple branches, which enables simultaneous analysis of multiple augmented views of an input image. Increasing branches of the target networks marginally increase the computational cost as each branch is updated only using exponential moving average of online network's parameters. Number of loss pairs depends upon number of target branches employed in the model and defines as $k^2(k+1)$, where k is number or target networks and 1 refers to single online network. Similarity score is than calculated between the l_2 normalized $\overline{q}_\theta(z_\theta)$ and $\overline{z}_{\xi_i}$, using the loss defined in equation eq1

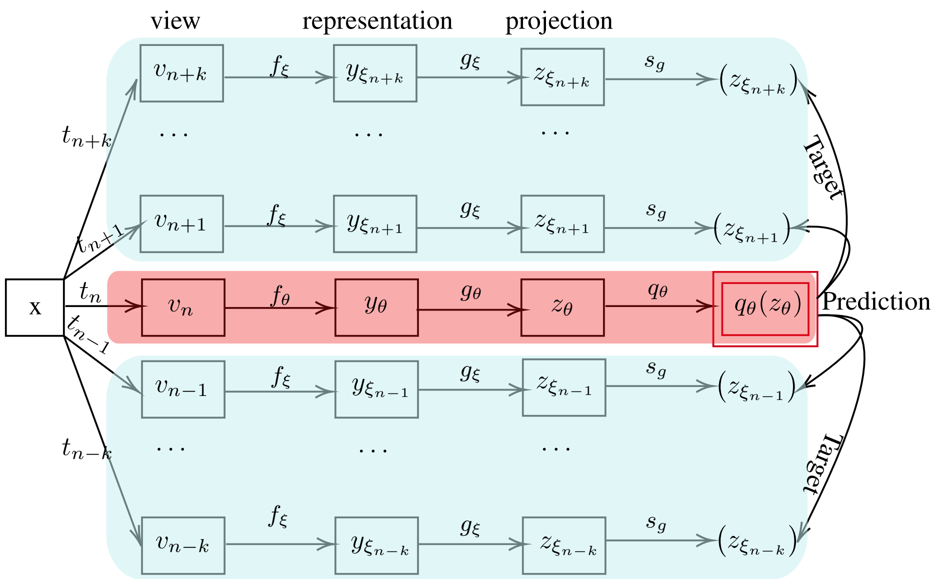

Figure 1: Overview of the proposed MT-BYOL, having multiple target network branches for an online network. $(t_{n+k}, ..., t_{n+1}, t_n, t_{n-1}, ..., t_{n-k})$ are the stochastic compositions of augmentations which create $(v_{n+k}, ..., v_{n+1}, v_n, v_{n-1}, ..., v_{n-k})$ augmented views of image x. θ are the trainable parameters of online network while ξ is updated as EMA of θ. MT-BYOL minimizes the cross-model similarity loss between representations generated by online network and target networks. s_g represents the stop-gradient operator.

$$L_\theta^{MT} = \frac{1}{2k} \sum_{i \in n-k, i \neq n}^{n+k} \left\| \overline{q}_\theta({z_\theta}^n) - {\overline{z}_\xi}^i \right\|_2^2 \tag{1}$$

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**

Table 1: Test-set classification accuracy of linear classifier evaluated on embeddings generated by frozen encoder for datasets