Barlow Twins

Friday, 8. October 2021 11:41 PM

- Based on Efficient Coding Principle by Barlow => present
  - Produces 2 distorted views of all Images
  - 2 Data Augmentations are performed on all samples :
    random cropping, resizing to 224 × 224, horizontal flipping, color jittering, converting to grayscale, Gaussian blurring, and solarization, The first two transfor- mations (cropping and resizing) are always applied, while the last five are applied randomly, with some probability. This probability is different for the two distorted views in the last two transformations (blurring and solarization).
  - Loss Function : Correlation Matrix should be identity (part 1) and the redundancy term aims to whiten the matrix to remove redundant informatio
  -

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

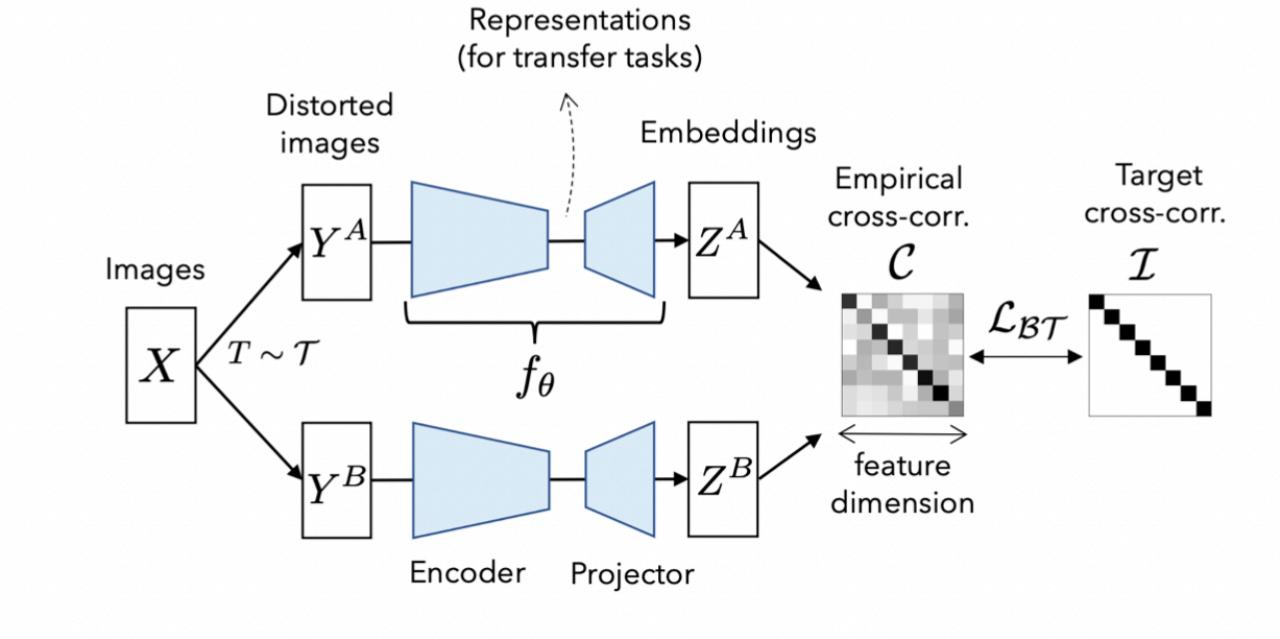
  -

$$\mathcal{L}_{\mathcal{BT}} = \underbrace{\sum_i \left(1 - \frac{\langle z^A_{\cdot,i}, z^B_{\cdot,i} \rangle_b}{\left\| z^A_{\cdot,i} \right\|_2 \left\| z^B_{\cdot,i} \right\|_2}\right)^2}_{\text{invariance term}} + \lambda \underbrace{\sum_i \sum_{j \neq i} \left(\frac{\langle z^A_{\cdot,i}, z^B_{\cdot,j} \rangle_b}{\left\| z^A_{\cdot,i} \right\|_2 \left\| z^B_{\cdot,j} \right\|_2}\right)^2}_{\text{redundancy reduction term}}$$

  -
  - Lambda : Tradeoff param in the information bottleneck framework (variance vs rendundancy term)
  - Benefits from very high dimensional embeddings
    Barlow Twins does not :
    - require large batch size (unlike infoNCE)

Comparions to InfoCe :

- InfoCe maximizes pariwise distances whereas BT decorrelates embedding vectors
  => less data required
  => BT benefits from large embeddings



- Encoder ResNet-50 : without final classification layer
- Projector Net (Decoder): 3 layer net
- Follows BYOL optimization scheme using LARS optimizer
- Robust to small batch sizes
- Very dependent on augmentations (unlike BYOL)

Table 2. **Semi-supervised learning on ImageNet** using 1% and 10% training examples. Results for the supervised method are from (Zhai et al., 2019). Best results are in **bold**.

Method	Top-1		Top-5	
	1%	10%	1%	10%
Supervised	25.4	56.4	48.4	80.4
PIRL	-	-	57.2	83.8
SIMCLR	48.3	65.6	75.5	87.8
BYOL	53.2	68.8	78.4	89.0
SWAV	53.9	**70.2**	78.5	**89.9**
BARLOW TWINS (ours)	**55.0**	69.7	**79.2**	89.3