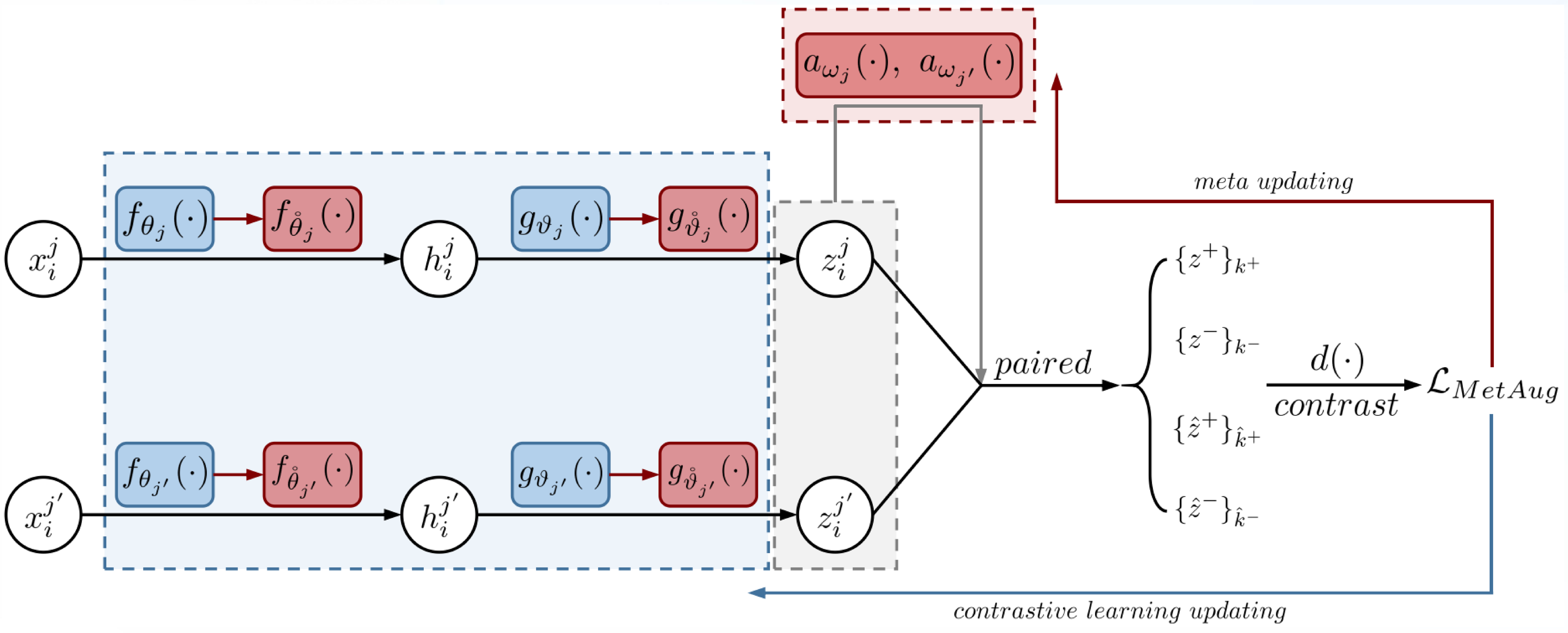
中国科学院软件研究所学术年会'2022
暨计算机科学国家重点实验室开放周
学术论文

MetAug: Contrastive Learning via Meta Feature Augmentation

Jiangmeng Li*, Wenwen Qiang*, Changwen Zheng, Bing Su, Hui Xiong

Proceedings of the 39th International Conference on Machine Learning, in Proceedings of Machine Learning Research 162:12964-12978

李江梦（15633670625，jiangmeng2019@iscas.ac.cn）



Contrastive Learning Preliminaries

- Contrastive loss formulation

$$\mathcal{L} = -\underset{X_S}{\mathbb{E}}\left[\log\frac{d\left(\{z^+\}\right)}{d\left(\{z^+\}\right)+\sum_{k=1}^{K} d\left(\{z^-\}_k\right)}\right]$$

- X_S : a set of pairs randomly sampled from X
- $\{z^+\}$: a positive pair
- $\{z^-\}_k$: K negative pairs, $k \in \{1,\dots,K\}$
- $d(\cdot)$: a discriminating function

- The contrastive loss guides the learned features to bring positive pairs together and push negative pairs farther apart.

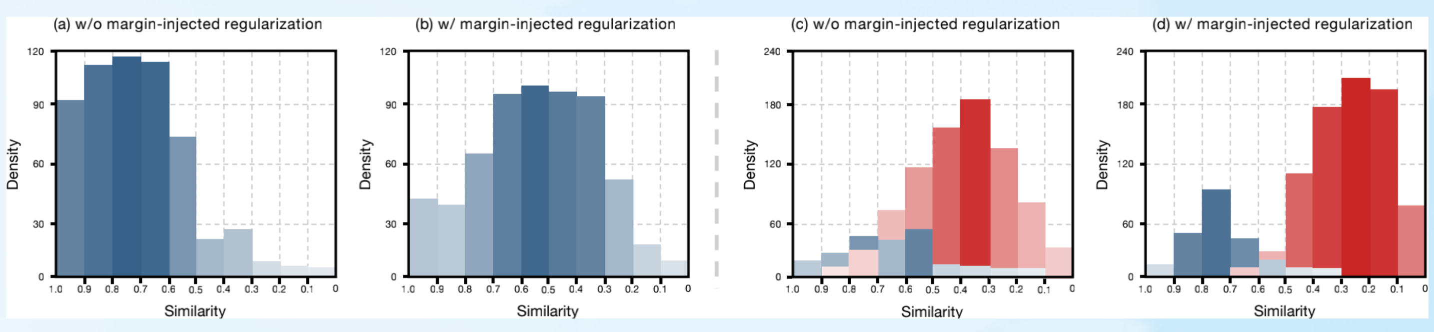
Motivation

- Contrastive learning heavily relies on informative features, or "hard" (positive or negative) features
  - Early works include informative features by applying complex data augmentations or adopting large batch size or memory bank
  - Recent works design elaborate sampling approaches to explore informative features
- Learning anti-collapsed feature augmentation

To this end, we propose to directly augment the features in latent space by using the anti-colapsed feature augmentation, thereby learning discriminative representations without a large amount of input data.

Margin-Injected Meta Feature Augmentation

- Meta feature augmentation generator (MAG)
  - Leverages second-derivative technique to update the parameters with respect to the improvement of the contrastive learning
- Margin-injected regularization
  - Injects a margin to encourage MAGs to generate anti-collapsed augmented features



Optimization-Driven Unified Contrast

- Jointly contrasts all features in one gradient back-propagation step
- Emphasizes the weight to the similarity that deviates from the optimum and decrases the weight to the similarity having close proximity with the optimum

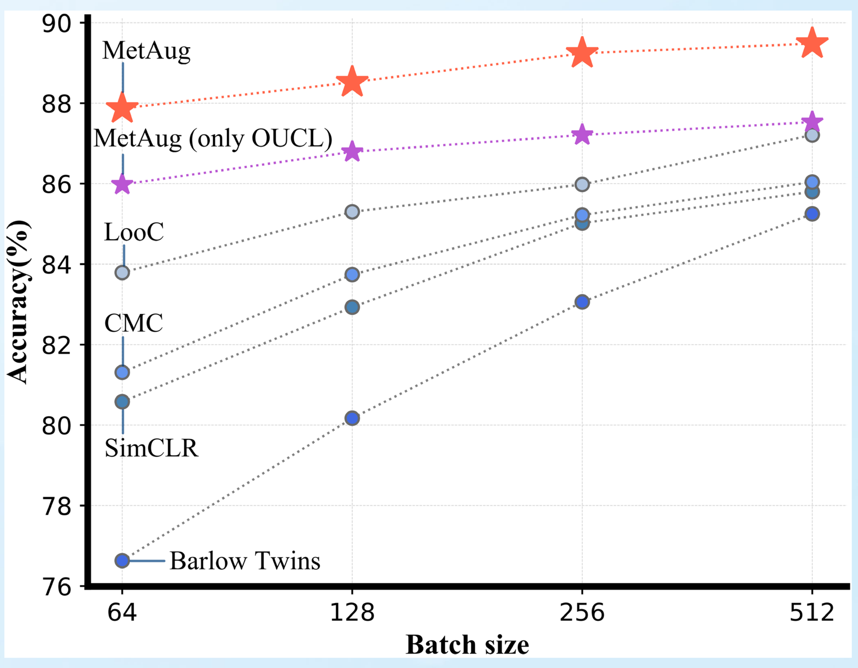
$$\mathcal{L}_{OUCL} = \frac{1}{\beta}log\left\{1+\sum_{k^-=1}^{K^-}\sum_{k^+=1}^{K^+} exp\left[\beta\left(\left(d(\{z^+\}_{k^+})-1\right)^2+\left(d(\{z^-\}_{k^-})\right)^2-2\gamma^2\right)\right]\right\}$$

Experimental Results

- Comparison with self-supervised learning methods

Model	Tiny ImageNet		STL-10		CIFAR10		CIFAR100	
	conv	fc	conv	fc	conv	fc	conv	fc
Fully supervised	36.60		68.70		75.39		42.27	
BiGAN	24.38	20.21	71.53	67.18	62.57	62.74	37.59	33.34
NAT	13.70	11.62	64.32	61.43	56.19	51.29	29.18	24.57
DIM	33.54	36.88	72.86	70.85	73.25	73.62	48.13	45.92
SplitBrain‡	32.95	33.24	71.55	63.05	77.56	76.80	51.74	47.02
SwAV	39.56 ± 0.2	38.87 ± 0.3	70.32 ± 0.4	71.40 ± 0.3	68.32 ± 0.2	65.20 ± 0.3	44.37 ± 0.3	40.85 ± 0.3
SimCLR	36.24 ± 0.2	39.83 ± 0.1	75.57 ± 0.3	77.15 ± 0.3	80.58 ± 0.2	80.07 ± 0.2	50.03 ± 0.2	49.82 ± 0.3
CMC‡	41.58 ± 0.1	40.11 ± 0.2	83.03	85.06	81.31 ± 0.2	83.28 ± 0.2	58.13 ± 0.2	56.72 ± 0.3
MoCo	35.90 ± 0.2	41.37 ± 0.2	77.50 ± 0.2	79.73 ± 0.3	76.37 ± 0.3	79.30 ± 0.2	51.04 ± 0.2	52.31 ± 0.2
BYOL	41.59 ± 0.2	41.90 ± 0.1	81.73 ± 0.3	81.57 ± 0.2	77.18 ± 0.2	80.01 ± 0.2	53.64 ± 0.2	53.78 ± 0.2
Barlow Twins	39.81 ± 0.3	40.34 ± 0.2	80.97 ± 0.3	81.43 ± 0.3	76.63 ± 0.3	78.49 ± 0.2	52.80 ± 0.2	52.95 ± 0.2
DACL	40.61 ± 0.2	41.26 ± 0.1	80.34 ± 0.2	80.01 ± 0.3	81.92 ± 0.2	80.87 ± 0.2	52.66 ± 0.2	52.08 ± 0.3
LooC	42.04 ± 0.1	41.93 ± 0.2	81.92 ± 0.2	82.60 ± 0.2	83.79 ± 0.2	82.05 ± 0.2	54.25 ± 0.2	54.09 ± 0.2
SimCLR + Debiased	38.79 ± 0.2	40.26 ± 0.2	77.09 ± 0.3	78.39 ± 0.2	80.89 ± 0.2	80.93 ± 0.2	51.38 ± 0.2	51.09 ± 0.2
SimCLR + Hard	40.05 ± 0.3	41.23 ± 0.2	79.86 ± 0.2	80.20 ± 0.2	82.13 ± 0.2	82.76 ± 0.1	52.69 ± 0.2	53.13 ± 0.2
CMC‡ + Debiased	41.64 ± 0.2	41.36 ± 0.1	83.79 ± 0.3	84.20 ± 0.2	82.17 ± 0.2	83.72 ± 0.2	58.48 ± 0.2	57.16 ± 0.2
CMC‡ + Hard	42.89 ± 0.2	42.01 ± 0.2	83.16 ± 0.3	85.15 ± 0.2	83.04 ± 0.2	86.22 ± 0.2	58.97 ± 0.3	59.13 ± 0.2
MetAug (only OUCL)‡	42.02 ± 0.1	42.14 ± 0.2	84.09 ± 0.2	84.72 ± 0.3	85.98 ± 0.2	87.13 ± 0.2	59.21 ± 0.2	58.73 ± 0.2
MetAug‡	**44.51 ± 0.2**	**45.36 ± 0.2**	**85.41 ± 0.3**	**85.62 ± 0.2**	**87.87 ± 0.2**	**88.12 ± 0.2**	**59.97 ± 0.3**	**61.06 ± 0.2**

- Comparison under multiple batch sizes



- Comparisons with different data augmentations

ıset. ID	Data augmentations						Methods		
	horizontal flip	rotate	random crop	random grey	color jitter	mixup	DACL	LooC	MetAug
1	✓	✓					-	80.73	**87.05**
2			✓				-	81.16	**87.53**
3				✓			-	80.70	**86.81**
4					✓		-	81.64	**87.79**
5	✓		✓				-	82.05	**88.12**
6		✓			✓		-	82.16	**88.01**
7	✓		✓			✓	80.87	82.21	**88.22**
8	✓	✓	✓	✓	✓	✓	82.09	83.17	**88.65**