Randomized Quantization: A Generic Augmentation for Data Agnostic Self-supervised Learning

Huimin Wu[1*] Chenyang Lei[2*] Xiao Sun[4] Peng-Shuai Wang[3]
Qifeng Chen[1] Kwang-Ting Cheng[1] Stephen Lin[5] Zhirong Wu[5]
[1]HKUST [2]CAIR, HKISI_CAS [3]Peking University [4]Shanghai AI Lab [5]Microsoft Research Asia

1. Ablation study

Channel-wise augmentation. We investigate channel-wise quantization against its instance-wise counterpart. Under the data-agnostic assumption, the value range may vary significantly for different channels. Applying randomized quantization in a channel-wise fashion respects the unique physical property of each data channel. In addition, it adds greater complexity to the augmentations. This allows to create more diverse tasks and facilitates contrastive learning. Empirically, we find that channel-wise quantization is better than instance-wise quantization by a margin of 3.6% (67.9% vs. 64.3%).

2. Compactibility with fully augmented MoCo-v3

We append the randomized quantization to the full augmentation of MoCo-v3, leading to a marginal improvement from 68.9% to 69.0%.

3. Compactibility with fully supervised learning

We follow the ResNet50 supervised training recipe on ImageNet from TorchVision. We report the average performance over three runs. The baseline result is 75.56%. With additional randomized quantization augmentation, the result improves to 75.83%. This shows that randomized quantization is effective in supervised learning as well. However, its effect is not as strong as in self-supervised learning.

4. Detailed results on the DABS benchmark

References

[1] Alex Tamkin, Vincent Liu, Rongfei Lu, Daniel Fein, Colin Schultz, and Noah Goodman. Dabs: A domain-agnostic benchmark for self-supervised learning. *arXiv preprint arXiv:2111.12062*, 2021. 2

*Equal contribution. Work done during an internship at MSRA.

Dataset	Domain	Metric	None	e-Mix [1]	Ours
CIFAR-10	Images	Accuracy	24.20	39.43	**47.70**
Birds	Images	Accuracy	1.62	3.86	**4.16**
VGG Flower	Images	Accuracy	9.03	25.96	**30.20**
DTD (Textures)	Images	Accuracy	7.39	8.83	**10.90**
GTSRB (Traffic)	Images	Accuracy	14.33	65.07	**86.80**
FGVC-Aircraft	Images	Accuracy	2.70	10.15	**12.60**
LibriSpeech Sp. ID	Speech	Accuracy	17.12	60.18	**62.70**
VoxCeleb Sp. ID	Speech	Accuracy	0.59	2.43	**2.69**
AudioMNIST	Speech	Accuracy	33.13	80.35	**82.80**
Google Speech	Speech	Accuracy	4.87	19.22	**26.00**
Fluent Locations	Speech	Accuracy	62.09	60.93	**65.20**
Fluent Actions	Speech	Accuracy	26.15	29.87	**31.40**
Fluent Objects	Speech	Accuracy	30.13	39.89	**40.80**
COLA	English Text	Pearson Corr.	0.00	**8.40**	8.27
MNLI_Matched	English Text	Accuracy	35.80	**37.80**	36.70
MNLI_Mismatched	English Text	Accuracy	36.60	**37.50**	37.00
MRPC	English Text	Accuracy	68.40	66.20	**68.90**
QNLI	English Text	Accuracy	57.70	**57.90**	57.40
QQP	English Text	Accuracy	65.10	64.30	**65.50**
RTE	English Text	Accuracy	**54.50**	51.30	52.70
SST2	English Text	Accuracy	57.00	**58.10**	55.80
STSB	English Text	Accuracy	4.20	11.40	**13.70**
WNLI	English Text	Accuracy	43.60	47.90	**50.70**
PAWS-X EN	Multilingual Text	Accuracy	**57.85**	54.85	56.20
PAWS-X FR	Multilingual Text	Accuracy	**57.80**	55.90	55.90
PAWS-X ES	Multilingual Text	Accuracy	**58.55**	55.50	54.80
PAWS-X DE	Multilingual Text	Accuracy	**58.85**	56.50	55.50
PAWS-X ZH	Multilingual Text	Accuracy	**57.35**	55.35	54.20
PAWS-X JP	Multilingual Text	Accuracy	**57.55**	57.35	56.70
PAWS-X KO	Multilingual Text	Accuracy	**58.80**	57.70	56.60
PAMAP2	Sensor	Accuracy	69.81	79.48	**84.90**
CheXpert	Chest X-Rays	Avg. AUROC	68.14	72.40	**73.40**
ChestX-ray8	Chest X-Rays	Avg. AUROC	57.00	63.00	**64.70**
VQA	Vision/Language	Accuracy	**57.50**	48.90	54.40

Table 1. Detailed comparisons with e-Mix [1] on the DABS benchmark.