Contrastive Quant: Quantization Makes Stronger Contrastive Learning

Yonggan Fu
Rice University
yf22@rice.edu

Qixuan Yu
Rice University
qy12@rice.edu

Meng Li
Meta Inc.
meng.li@fb.com

Xu Ouyang
Rice University
xo2@rice.edu

Vikas Chandra
Meta Inc.
vchandra@fb.com

Yingyan Lin
Rice University
yingyan.lin@rice.edu

ABSTRACT

Contrastive learning learns visual representations by enforcing feature consistency under different augmented views. In this work, we explore contrastive learning from a new perspective. Interestingly, we find that quantization, when properly engineered, can enhance the effectiveness of contrastive learning. To this end, we propose a novel contrastive learning framework, dubbed Contrastive Quant, to encourage feature consistency under both differently augmented inputs via various data transformations and differently augmented weights/activations via various quantization levels. Extensive experiments, built on top of two state-of-the-art contrastive learning methods SimCLR and BYOL, show that Contrastive Quant consistently improves the learned visual representation.

ACM Reference Format:

Yonggan Fu, Qixuan Yu, Meng Li, Xu Ouyang, Vikas Chandra, and Yingyan Lin. 2022. Contrastive Quant: Quantization Makes Stronger Contrastive Learning . In *Proceedings of the 59th ACM/IEEE Design Automation Conference (DAC) (DAC '22), July 10–14, 2022, San Francisco, CA, USA.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3489517.3530419

1 INTRODUCTION

Contrastive learning has emerged as the state-of-the-art (SOTA) unsupervised representation learning from images. For example, Momentum Contrast (MoCo) [1] shows that unsupervised pre-training can surpass its ImageNet-supervised counterpart in multiple detection and segmentation tasks, while SimCLR further reduces the gap in linear classifier accuracy between unsupervised and supervised pre-training representations. As such, there has been a growing interest in further boosting its achievable performance and developing improved contrastive learning pipelines.

In this work, we explore contrastive learning from a new perspective, inspired by recent works showing that properly designed weight perturbations or quantization help the models learn a smoother loss landscape [2, 3]. For example, [2] adversarial perturbations on both inputs and weights help smooth the loss landscape of model weights and thus narrow the robust generalization gap and [3] shows that a properly designed precision schedule helps DNN converge to a better local optima, as a low precision helps the optimization space exploration in a similar way a high learning rate does. We are thus motivated to ask an intriguing question: "*Can quantization, which itself can boost the model efficiency, be leveraged to develop improved contrastive learning pipelines*"? If the answer is positive, it can not only lead to more accurate contrastive learning techniques on top of existing methods, but also open up a new understanding in the role of quantization on contrastive learning, potentially inspiring and motivating more contrastive learning innovations.

Interestingly, we find that quantization, when properly engineered, can enhance the effectiveness of contrastive learning. Specifically, we make the following contributions:

- We are the first to study the role of quantization in the context of contrastive learning pipelines, and show that quantization can be leveraged to enhance the performance of contrastive learning. We believe that this view can open up a new perspective for future contrastive learning innovations.
- We propose a novel contrastive learning framework, dubbed Contrastive Quant, to encourage the feature consistency under both (1) different augmented inputs via various data transformations and (2) different augmented weights and activations via various quantization levels. In particular, the feature consistency under injected noises via quantization in Contrastive Quant can be viewed as augmentations on both model weights and intermediate activations, serving as a complement to the input augmentations.
- Extensive experiments, built on top of two SOTA contrastive learning methods SimCLR and BYOL, show that our Contrastive Quant consistently improves the learned visual representation, especially with limited labeled data under semi-supervised scenarios. For example, our Contrastive Quant achieves a 8.69% and 10.27% higher accuracy on ResNet-18 and ResNet-34, respectively, on ImageNet when fine-tuning with 10% labeled data.

2 RELATED WORK

DNN quantization. Quantization is one of the most promising DNN compression techniques which trims down the model complexity from the most fine-grained bit level. In particular, existing quantization methods represent model weights, activations, and gradients using lower floating-point precision [4] or fixed-point precision [5]. The accuracy degradation after quantization can be

DAC '22, July 10–14, 2022, San Francisco, CA, USA

ACM ISBN 978-1-4503-9142-9/22/07…$15.00
https://doi.org/10.1145/3489517.3530419

minimized through quantization-aware training [5] which explicitly considers quantization noise in the training process.

Contrastive learning. Self-supervised learning, which leverages the input data themselves for supervision to benefit the downstream tasks, has achieved great progress. Early works [6] adopt generative models to recover the original data distributions without making any assumptions for the downstream tasks to learn good representations. Later works shed light on the potential of discriminative models for representation learning. Recently, contrastive learning has gained increased popularity thanks to its excellent performance. The kernel spirit behind contrastive learning is to learn invariant features by maximizing the mutual information of the latent representations of differently augmented views of the images. In particular, instance discrimination [7] makes the first attempt to discriminate between different instances via a Noise-Contrastive Estimation (NCE) loss and following works consider different strategies to construct the different views. For example, CMC [8] converts RGB images to the Lab color space and maximizes the mutual information between different color channel views, and SimCLR [9] adopts different augmentations as different views and maximizes the consistency between different views. Readers are referred to [10] for more details about unsupervised learning and contrastive learning. In this work, we explore contrastive learning from a new perspective, i.e., the potential positive role of quantization in contrastive learning, and find that quantization, when properly engineered, can enhance the effectiveness of contrastive learning, potentially inspiring and motivating more contrastive learning innovations.

3 THE PROPOSED CONTRASTIVE QUANT FRAMEWORK

In this section, we introduce our Contrastive Quant framework, which for the first time explores quantization's positive effects on contrastive learning, in addition to merely boosting the model efficiency. We start with the motivation for and inspiration leading to our framework in Sec. 3.1 and then introduce the key concept in Sec. 3.2. Next, we discuss potential designs to enhance the positive effect of quantization on contrastive learning and the implementation details for applying Contrastive Quant on top of existing contrastive learning frameworks in Sec. 3.3 and Sec. 3.4.

3.1 Motivation

The spirit of contrastive learning. Self-supervised learning methods aim at learning representations with semantic priors that can generally benefit their downstream tasks. The core spirit behind contrastive learning [1, 7–9, 11, 12], one of the most effective self-supervised learning methods, is to learn invariant features by maximizing the mutual information between the latent representations of differently augmented image views. Such a feature consistency is known to be beneficial for both the standard generalization [13] and robust generalization [14, 15], while new perspectives for enforcing such feature consistency in addition to data augmentations are still under-explored.

Inspiration from recent works. A recent work [2] implies another view of encouraging feature consistency in the context of adversarial training. In particular, it shows that training with properly generated perturbations onto the weights can serve as a complement to adversarial perturbations onto the inputs, which helps smooth the loss landscape and narrow the robust generation gap, i.e., improves the feature consistency under adversarial attacks. In parallel, [3] shows that a low precision has a similar effect as a high learning rate, favoring the training space exploration, and proposes a properly designed precision schedule to help DNN converge to a better local optima with a more smoothed loss landscape. Considering quantization can naturally serve as perturbations onto both model weights and intermediate feature maps, these prior works inspire us to leverage quantization to learn better representations by encouraging feature consistency under differently augmented weights/activations via various quantization levels, as a complement to enforcing consistency via data augmentations.

3.2 The key concept

Contrastive learning, as a kind of instance discrimination method, encourages feature consistency via maximizing the mutual information, approximated by minimizing an NCE loss [7], between the extracted features from different perspectives of the same instance which can be formulated as:

$$\max\ I(f, f^{+}) \approx \min\ NCE(f, f^{+}) \tag{1}$$

$$= \min\ \mathbb{E}\left[-log \frac{exp(f \cdot f^{+}/\tau)}{\sum_{i=0}^{K} exp(f \cdot f^{-}/\tau) + exp(f \cdot f^{+}/\tau)}\right] \tag{2}$$

where f and f^{-} are the extracted features of the same instance under different perspectives, named positive pairs, of the encoder F, $I(f, f^{+})$ denotes their mutual information, K is the number of negative samples, and τ is a temperature parameter that controls the concentration level of the distribution.

In previous data augmentation based contrastive learning methods, the positive pairs f and f^{+} are generated by different augmentation combinations:

$$f = F(Aug_1(x), \theta)\ ,\ f^{+} = F(Aug_2(x), \theta) \tag{3}$$

where x is the given instance, θ is the model weight, and Aug_1 and Aug_1 denote two different augmentations. To further enhance the feature consistency from a new perspective, we propose the Contrastive Quant framework to augment both the model weights and intermediate activations in addition to the input augmentations via injecting quantization noise of different levels into the weights and activations, as formulated below:

$$f = F_{q_1}(x, \theta_{q_1})\ ,\ f^{+} = F_{q_2}(x, \theta_{q_2}) \tag{4}$$

where f and f^{+} are generated by the encoder quantized to different precisions q_1 and q_2, which can be randomly selected from a precision set during training. The detailed quantization scheme we adopt is discussed in Sec. 3.4.

Nevertheless, we empirically find that merely enhancing the feature consistency via quantization augmented weights/activations without input augmentations will lead to inferior contrasting learning, indicating that the structured priors learned from the consistency between different augmented views are essential to the downstream tasks and it is necessary to apply our Contrastive Quant framework on top of existing input augmentation based contrasting learning methods. A natural question is thus "how to effectively combine augmented inputs and augmented weights/activations via quantization to boost the performance of contrastive learning"? To

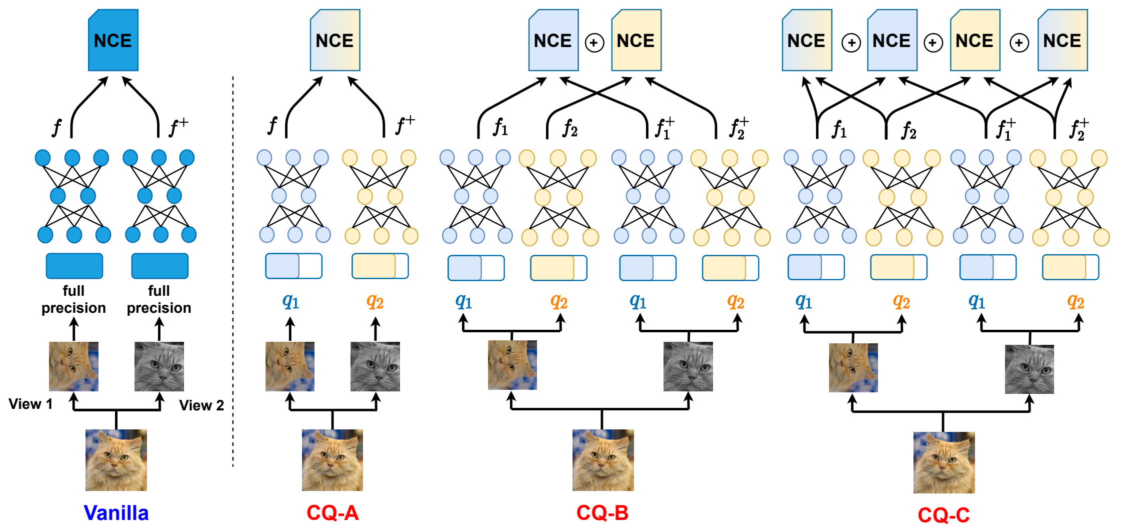

Figure 1: An overview of the proposed Contrastive Quant design pipelines.

answer this question, we explore the potential designs in Sec. 3.3 towards an effective Contrastive Quant framework.

3.3 The design pipeline

As shown in Fig. 1, we propose three candidate designs, denoted as CQ-A, CQ-B, and CQ-C, respectively, to exploit the potential of applying augmented weights/activations via quantization towards on top of the commonly used augmented inputs towards better contrastive learning pipelines.

Analysis of CQ-A. CQ-A (see the second column of Fig. 1) is one of the most intuitive designs for applying augmented weights/activations on top of existing input augmentation based contrastive learning methods, which can be formulated as:

$$f = F_{q_1}(Aug_1(x), \theta_{q_1})\ ,\ f^+ = F_{q_2}(Aug_2(x), \theta_{q_2})\ ,\ Loss = NCE(f, f^+) \tag{5}$$

where *Loss* is the final objective to be minimized. Eq. 5 shows that CQ-A views the quantization precision as an additional augmentation parameter similar to the rotation degree or noise level in data augmentation operators [9], which are randomly selected during inference with different views for the same instance. In this way, the executions of input augmentations and weight/activation augmentation are combined in a sequential manner, which can potentially strengthen the overall augmentation magnitude. As validated in Sec. 4, such strengthened augmentations will aggressively benefit large-scale datasets like ImageNet while may not be very helpful on small-scale datasets as strong augmentations may distort the images' structures [16], which is more likely to happen on small-scale datasets.

Analysis of CQ-B. CQ-B (see the third column of Fig. 1) is a variant with more mild augmentations. Instead of explicitly constraining the feature consistency under differently augmented weights/activations via various quantization levels, CQ-B only enforces the feature consistency under differently augmented inputs with the same randomly selected precision, which can be formulated as:

$$f_1 = F_{q_1}(Aug_1(x), \theta_{q_1}),\ f_2 = F_{q_2}(Aug_1(x), \theta_{q_2}) \tag{6}$$

$$f_1^+ = F_{q_1}(Aug_2(x), \theta_{q_1}),\ f_2^+ = F_{q_2}(Aug_2(x), \theta_{q_2}) \tag{7}$$

$$Loss = NCE(f_1, f_1^+) + NCE(f_2, f_2^+) \tag{8}$$

The final objective is averaged over two randomly selected precisions, which implicitly encourages the feature consistency under different quantization levels. As such, this design can potentially mitigate the potential risk of distorting the images' structures with augmentations that are too strong, especially on small-scale datasets, while the newly introduced priors by CQ-B over input augmentation based methods are not rich enough, limiting its achievable performance improvement.

Analysis of CQ-C. Combining the advantages of both CQ-A and CQ-B, CQ-C (see the rightmost column of Fig. 1) explicitly encourages (1) the feature consistency under differently augmented views with the same quantization level, and (2) the feature consistency under differently augmented weights/activations via various quantization levels within the same view, which is formulated as:

$$Loss = NCE(f_1, f_1^+) + NCE(f_2, f_2^+) + NCE(f_1, f_2) + NCE(f_1^+, f_2^+) \tag{9}$$

where f_1, f_2, f_1^+, and f_2^+ follow the definition in Eq. 6 and 7. Different from CQ-A which sequentially augments the inputs and weights/activations, CQ-C decouples them and enforces the feature consistency from the two perspectives separately, which can potentially mitigate the risk of introducing augmentations that are too strong while introducing sufficient new priors on top of existing contrastive learning methods. As validated in Sec. 4, CQ-C can consistently improve the performance on both small-scale and large-scale datasets.

3.4 More implementation details

The adopted quantization scheme. We adopt the commonly adopted linear quantizer [5] to quantize both weights and activations in our Contrastive Quant, i.e.,

$$A_q = S_a \lfloor \frac{A}{S_a} \rceil\ ,\quad where\ S_a = \frac{A_{range}}{2^q - 1} \tag{10}$$

where A here can denote the model weights or activations, q is the quantization bit-width, and A_{range} is the dynamic range of A, i.e., the difference between the maximum and minimum of A values.

Since the encoder will be quantized to different values during training, learnable quantizers with trainable quantization parameters are found to be unstable here, so we directly adopt the linear quantizer.

Applying on top of SimCLR. Adapting our Contrastive Quant to the SOTA SimCLR framework [9] is simple and direct via (1) modifying the NCE loss in Eq. 1 to the NT-Xent one [9], and (2) adding a projection head after the encoder to learn a better representation.

Applying on top of BYOL. BYOL [11] relies on two networks, i.e., the online and target networks, to learn from each other based on the feature consistency under different views, where the target network is updated via the moving average of the online network instead of the gradients, and this training process does not involve any negative pair. We adapt our Contrastive Quant in a natural manner in that (1) we modify the NCE loss in Eq. 1 to the Mean Square Error (MSE) loss adopted by [11]; (2) we add a projection head and prediction head after the encoder following [11]; and (3) we stop the gradient propagation along the target network and apply both views of the same instance into the online network/target network alternatively to improve the data reusability following [11].

4 EXPERIMENT RESULTS

In this section, we first introduce our experiment setup, benchmarking experiment results over SOTA contrastive learning methods, and then ablation studies of Contrastive Quant for better understanding its effectiveness and design pipelines.

4.1 Experiment setup

Networks, datasets, and evaluation settings. We consider six networks on two datasets, i.e., ResNet-18/34/74/110/156/MobileNetV2 on CIFAR-100 and ResNet-18/34 on ImageNet, featuring diverse DNN models and data statistics for a solid evaluation. We also transfer the pretrained models on ImageNet to the downstream detection task Pascal VOC [17]. In this work, we consider both the fine-tuning, linear evaluation, and transfer learning settings. As we quantize the model to different precisions during the contrastive learning processes, we mainly consider the fine-tuning setting under a fixed precision, i.e., full precision (denoted as FP) or 4-bit, with a limited amount of labeled data (10% or 1%) to stabilize the weight/activation distribution under the precision choice.

Training finetuning settings. For experiments on ImageNet, we follow all the training and finetuning settings in [9]. For experiments on CIFAR-100, we follow the optimizer settings, augmentation choices, and the projection head design in [9] and train the models with a batch size of 512 for 1000 epochs. For both fine-tuning and linear evaluation settings, we adopt an SGD optimizer with a momentum of 0.9 and a cosine learning rate decay with an initial learning rate of 0.1 to fine-tune the models for 50 epochs.

Precision sets. As our Contrastive Quant framework randomly selects two precision q_1 and q_2 from a pre-defined precision set in each training iteration, the precision set may influence its training optimality. We adopt *4-16* (every precision between 4-bit and 16-bit), *6-16*, and *8-16* as the potential precision sets.

4.2 Benchmark on ImageNet

We first apply our Contrastive Quant framework on top of SimCLR to train ResNet-18/34 on ImageNet and benchmark with the vanilla SimCLR in both fine-tuning and linear evaluation settings under both FP and 4-bit precisions. In particular, we validate the effectiveness of all the three designs in Fig. 1 and adopt two precision sets *6-16* and *8-16* considering 4-bit may significantly degrade the final accuracy on large-scale datasets like ImageNet.

Table 1: Benchmark Constrastive Quant against SimCLR on top of ResNet-18/34 on ImageNet. Here we adopt the fine-tuning settings on 10%/1% labeled data with the two precision sets.

Network	Method	Precision Set	Fine-tune Acc. (FP)		Fine-tune Acc. (4-bit)	
			10% labels	1% labels	10% labels	1% labels
ResNet-18	SimCLR	-	42.44	19.18	39.12	17.24
	CQ-A	6-16	**51.39**	28.87	**48.80**	**27.13**
		8-16	51.13	**28.97**	48.63	26.66
	CQ-C	6-16	44.97	20.83	42.01	18.63
		8-16	45.10	20.98	41.90	18.72
ResNet-34	SimCLR	-	47.53	23.43	44.65	21.69
	CQ-A	6-16	**55.76**	33.37	53.32	31.30
		8-16	55.72	**33.70**	**53.33**	**31.64**
	CQ-C	6-16	50.45	26.32	47.65	24.53
		8-16	50.22	26.21	47.70	24.74

Fine-tuning results. As shown in Tab. 1, we can see that (1) both CQ-A and CQ-C achieve a consistent improvement over the vanilla SimCLR on both ResNet-18/34 and the two precision sets, indicating enforcing feature consistency under differently augmented weights/activations via various quantization levels indeed benefits the downstream tasks, which could provide a new prior; (2) CQ-C achieves a 1.39%∼2.89% and 2.69%∼3.05% higher accuracy over SimCLR on ResNet-18 and ResNet-34, respectively, while CQ-A achieves an even surprising at of 8.69%∼9.89% and 8.19%∼10.27% on ResNet-18 and ResNet-34, respectively, indicating the sequentially applied augmentations onto the inputs and weights/activations, which together lead to a stronger augmentation, greatly benefit the training process on large-scale datasets; and (3) we find that the training process of CQ-B can easily fail, which suffers from severe gradient explosion before learning a good representation, indicating the importance of adopting the feature consistency loss of differently augmented weights/actions under the same view as CQ-C, which is the only difference between the two designs.

Table 2: Comparing CQ-A and CQ-C with SimCLR under the linear evaluation setting on ImageNet.

Network	SimCLR	CQ-C	**CQ-A**
ResNet-18	29.31	31.90	**44.91**
ResNet-34	34.96	36.14	**47.88**

Linear evaluation results. As observed from Tab. 2, we can see that CQ-C and CQ-A achieve a 2.59%/1.18% and 15.60%/12.92% higher accuracy on ResNet-18/34, respectively. Furthermore, CQ-A still achieves the most aggressive improvements over SimCLR, aligning with our assumption about CQ-A's stronger augmentation effects which benefits ImageNet training.

Transfer to downstream detection tasks. We further transfer the pretrained ResNet-18/34 trained with different methods on ImageNet to a downstream detection task on top of YOLOv4 [18] following [19]. In particular, we follow [1, 19] to train the models on the combined training/validation set of Pascal VOC 2007 and Pascal VOC 2012 [17], and evaluate on the test set of Pascal VOC

Table 3: Comparing CQ-A and CQ-C with vanilla SimCLR via transferring the ImageNet pretrained models to the detection task.

Network	Method	AP	AP50	AP75
ResNet-18	Vanilla SimCLR	25.09	49.2	22.74
	CQ-C	32.94	63.96	29.28
	CQ-A	**36.39**	**69.08**	**32.64**
ResNet-34	Vanilla SimCLR	35.58	67.51	31.88
	CQ-C	36.54	68.77	34.17
	CQ-A	**38.77**	**72.13**	**35.85**

Table 4: Benchmark against SimCLR on six network with CIFAR-100 with the fine-tuning setting.

Network	Method	Fine-tune Acc. **(FP)** 10% labels	Fine-tune Acc. **(FP)** 1% labels	Fine-tune Acc. **(4-bit)** 10% labels	Fine-tune Acc. **(4-bit)** 1% labels
ResNet-18	SimCLR	61.51	42.51	59.78	40.73
	CQ-C	**61.75**	**43.80**	**60.12**	**42.59**
ResNet-34	SimCLR	63.05	45.11	61.44	43.63
	CQ-C	**63.58**	**48.05**	**61.47**	**45.75**
ResNet-74	SimCLR	51.93	30.40	50.37	28.56
	CQ-C	**52.52**	**31.39**	**51.12**	**29.70**
ResNet-110	SimCLR	52.78	31.16	51.69	30.11
	CQ-C	**54.47**	**33.17**	**52.28**	**32.66**
ResNet-152	SimCLR	53.57	32.93	52.14	31.06
	CQ-C	**55.44**	**34.98**	**53.04**	**33.54**
MobileNetV2	SimCLR	49.73	24.18	46.47	18.98
	CQ-C	**51.59**	**26.12**	**49.82**	**20.82**

2007. As observed from Tab. 3, we can see that CQ-A and CQ-C consistently outperform the vanilla SimCLR after being transferred to the downstream detection task in terms of all the three metrics. Specifically, CQ-A achieves a 11.30%/3.19% higher AP compared with the vanilla SimCLR on top of ResNet-18/34, respectively.

Insights. [16] finds that stronger augmentations can distort the images' structures [16], thus can be harmful to learn a good representation, while they mainly discuss within the domain of input augmentations. Based on the success of CQ-A in Tab. 1, our Contrastive Quant can potentially serve as a new direction to explore stronger augmentations which consistently benefit the downstream tasks. A future direction is to explore other kinds of perturbations on weights/activations in addition to our Contrastive Quant to build more effective augmentations.

4.3 Benchmark on CIFAR-100

We benchmark Contrastive Quant on top of SimCLR [9]/BYOL [11] with the vanilla SimCLR/BYOL. As shown in Sec. 4.4, CQ-C outperforms CQ-A and CQ-B on small-scale datasets like CIFAR-100, which is consistent with the analysis in Sec. 3.3. Therefore, in this subsection we adopt CQ-C with a precison set of *6-16*.

Benchmark against SimCLR with fine-tuning settings. As shown in Tab. 4, we can see that (1) our CQ-C consistently outperforms the vanilla SimCLR on all the six networks, e.g., an accuracy improvement of 0.24%∼3.35% and 0.99%∼2.94% when fine-tuning with 10% and 1% data with FP, respectively; and (2) our CQ-C achieves more notable improvements over SimCLR on top of larger models and less labeled data, indicating its scalability and practicality in real-world applications.

Table 5: Benchmark against SimCLR on six DNNs & CIFAR-100 under the linear evaluation setting.

Method	ResNet-18	ResNet-34	ResNet-74	ResNet-110	ResNet-152	MobileNetV2
SimCLR	**64.91**	65.92	52.96	53.53	53.97	52.53
CQ-C	64.78	**66.54**	**54.06**	**54.76**	**55.12**	**53.97**

Table 6: Benchmark against BYOL on three network with CIFAR-100 under the fine-tuning setting.

Network	Method	Precision Set	Fine-tune Acc. (FP) 10% labels	Fine-tune Acc. (FP) 1% labels	Fine-tune Acc. **(4-bit)** 10% labels	Fine-tune Acc. **(4-bit)** 1% labels
ResNet-18	BYOL	-	55.26	34.22	53.44	32.93
	CQ-C	6-16	**58.84**	**39.21**	**56.74**	**37.54**
ResNet-34	BYOL	-	65.83	50.95	64.00	49.37
	CQ-C	6-16	**66.77**	**51.91**	**65.21**	**50.55**
MobileNetV2	BYOL	-	49.85	23.32	44.65	19.58
	CQ-C	6-16	**54.59**	**31.96**	**50.97**	**26.60**

Benchmark against SimCLR with linear evaluation settings. Tab. 5 shows that our CQ-C still achieves a better accuracy (except ResNet-18) than the vanilla SimCLR.

Benchmark against BYOL with fine-tuning settings. A consistent improvement can be observed when benchmarking against BYOL in Tab. 6, i.e., CQ-C achieves a 0.94%∼6.32% and 0.96%∼8.64% higher accuracy when fine-tuning with 10% and 1% data with FP, respectively.

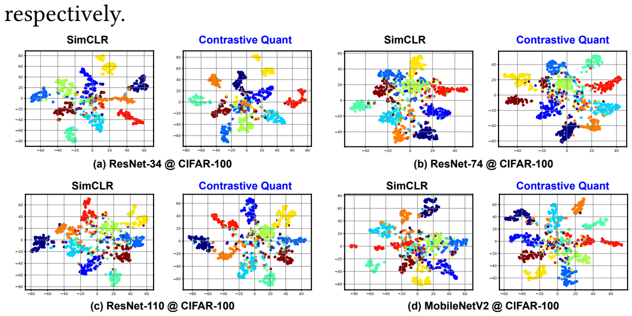

Figure 2: Visualizing the learned representations of Contrastive Quant and SimCLR using t-SNE [20].

Visualizing the learned representations. We adopt t-SNE [20] to visualize the learned representations of the models trained by Contrastive Quant and SimCLR in Fig. 2. We can see that the representations learned by Contrastive Quant show a better linear separability, especially under larger models.

4.4 Ablation studies: Benchmark CQ variants

As shown in Tab. 7, we can observe that (1) CQ-C achieves an overall better performance than the other two variants, especially when being fine-tuned with less labeled data, and (2) CQ-A achieves marginally better or comparable results over the vanilla SimCLR, different from its superior performance on ImageNet, aligning with our analysis in Sec. 3.3 showing augmentations that are too strong may distort the images' structures [16] on small-scale datasets.

4.5 Ablation studies: Augment with different quantization levels only

Setup. To justify whether different quantization levels in our Contrastive Quant play a similar role as data augmentations, we design a new variant of Contrastive Quant named CQ-quant, where

Table 7: Ablation studies of Constrastive Quant's variants with a precision set of *6-16* on CIFAR-100.

Network	Method	Fine-tune Acc. (FP) 10% labels	Fine-tune Acc. (FP) 1% labels	Fine-tune Acc. (4-bit) 10% labels	Fine-tune Acc. (4-bit) 1% labels
ResNet-34	SimCLR	63.05	45.11	61.44	43.63
	CQ-A	**63.63**	45.60	**61.77**	43.56
	CQ-B	63.57	45.26	61.76	43.60
	CQ-C	63.58	**48.05**	61.47	**45.75**
ResNet-74	SimCLR	51.93	30.40	50.37	28.56
	CQ-A	51.89	29.95	**51.45**	28.99
	CQ-B	52.36	30.48	51.20	29.28
	CQ-C	**52.52**	**31.39**	51.12	**29.70**
MobileNetV2	SimCLR	49.73	24.18	46.47	18.98
	CQ-A	49.93	24.57	46.01	19.38
	CQ-B	**51.78**	25.21	47.81	20.81
	CQ-C	51.59	**26.12**	**49.82**	**20.82**

Table 8: Evaluating CQ-Quant augmented by different quantization levels only on CIFAR-100.

Network	Precision Set	Fine-tune Acc. (FP) 1% labels	Fine-tune Acc. (FP) 10% labels	Linear evaluation
ResNet-74	6-16	**7.64**	**29.14**	**15.79**
	8-16	4.64	21.37	10.98
	No SSL Training	2.90	20.76	3.69
ResNet-110	6-16	**7.43**	**27.69**	**14.10**
	8-16	6.41	21.58	11.83
	No SSL Training	2.21	20.56	3.15

each input is only augmented by different quantization levels without being augmented by different data augmentation methods. To be more specific, the loss function is modified from Eq. 9 to be: $Loss = NCE(f_1, f_2)$. We benchmark CQ-Quant with the baseline without SSL training, i.e., training from scratch during evaluation, on top of ResNet-74 and ResNet-110.

Observations. As shown in Tab. 8, we can observe that (1) CQ-Quant with different precision sets can consistently outperform the baseline without SSL training, indicating that the different quantization levels can indeed create a contrastive task with a similar effect as different data augmentations on the inputs; (2) CQ-Quant with more diverse precision settings can achieve a better fine-tuning/linear-evaluation accuracy, which aligns with our intuition that the diversity of augmentations contributes to the success of contrastive learning; and (3) data augmentations are still necessary towards decent contrastive learning performances.

5 CONCLUSION

In this work, we for the first time explore quantization's positive effects on boosting contrastive learning's accuracy in addition to merely boosting the model efficiency, and propose a novel contrastive learning framework, dubbed Contrastive Quant, to enhance the feature consistency under both (1) differently augmented inputs via various data transformations and (2) differently augmented weights/activations via various quantization levels. Extensive experiments on top of two state-of-the-art contrastive learning methods, SimCLR and BYOL, show that Contrastive Quant consistently improves the learned visual representation, especially with limited labeled data under semi-supervised scenarios.

ACKNOWLEDGEMENT

The work is supported by the National Science Foundation (NSF) through the EPCN program (Award number: 1934767) and an NSF CAREER award (Award number: 2048183).

REFERENCES

[1] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9729–9738, 2020.

[2] Dongxian Wu, Shu-Tao Xia, and Yisen Wang. Adversarial weight perturbation helps robust generalization. *Advances in Neural Information Processing Systems*, 33, 2020.

[3] Yonggan Fu, Han Guo, Meng Li, Xin Yang, Yining Ding, Vikas Chandra, and Yingyan Lin. Cpt: Efficient deep neural network training via cyclic precision. *arXiv preprint arXiv:2101.09868*, 2021.

[4] Naigang Wang, Jungwook Choi, Daniel Brand, Chia-Yu Chen, and Kailash Gopalakrishnan. Training deep neural networks with 8-bit floating point numbers. In *Advances in neural information processing systems*, pages 7675–7684, 2018.

[5] Benoit Jacob, Skirmantas Kligys, Bo Chen, Menglong Zhu, Matthew Tang, Andrew Howard, Hartwig Adam, and Dmitry Kalenichenko. Quantization and training of neural networks for efficient integer-arithmetic-only inference. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2704–2713, 2018.

[6] Aaron Van Oord, Nal Kalchbrenner, and Koray Kavukcuoglu. Pixel recurrent neural networks. In *International Conference on Machine Learning*, pages 1747–1756. PMLR, 2016.

[7] Zhirong Wu, Yuanjun Xiong, Stella X Yu, and Dahua Lin. Unsupervised feature learning via non-parametric instance discrimination. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3733–3742, 2018.

[8] Yonglong Tian, Dilip Krishnan, and Phillip Isola. Contrastive multiview coding. *arXiv preprint arXiv:1906.05849*, 2019.

[9] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PMLR, 2020.

[10] Xiao Liu, Fanjin Zhang, Zhenyu Hou, Zhaoyu Wang, Li Mian, Jing Zhang, and Jie Tang. Self-supervised learning: Generative or contrastive. *arXiv preprint arXiv:2006.08218*, 1(2), 2020.

[11] Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre H Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pires, Zhaohan Daniel Guo, Mohammad Gheshlaghi Azar, et al. Bootstrap your own latent: A new approach to self-supervised learning. *arXiv preprint arXiv:2006.07733*, 2020.

[12] Xinlei Chen and Kaiming He. Exploring simple siamese representation learning. *arXiv preprint arXiv:2011.10566*, 2020.

[13] Osman Semih Kayhan and Jan C van Gemert. On translation invariance in cnns: Convolutional layers can exploit absolute spatial location. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14274–14285, 2020.

[14] Minseon Kim, Jihoon Tack, and Sung Ju Hwang. Adversarial self-supervised contrastive learning. *arXiv preprint arXiv:2006.07589*, 2020.

[15] Ziyu Jiang, Tianlong Chen, Ting Chen, and Zhangyang Wang. Robust pre-training by adversarial contrastive learning. *arXiv preprint arXiv:2010.13337*, 2020.

[16] Xiao Wang and Guo-Jun Qi. Contrastive learning with stronger augmentations. *arXiv preprint arXiv:2104.07713*, 2021.

[17] Mark Everingham, SM Ali Eslami, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes challenge: A retrospective. *International journal of computer vision*, 111(1):98–136, 2015.

[18] Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao. Yolov4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*, 2020.

[19] Tianlong Chen, Jonathan Frankle, Shiyu Chang, Sijia Liu, Yang Zhang, Michael Carbin, and Zhangyang Wang. The lottery tickets hypothesis for supervised and self-supervised pre-training in computer vision models. *arXiv preprint arXiv:2012.06908*, 2020.

[20] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008.