QReg: On Regularization Effects of Quantization

MohammadHossein AskariHemmat[1], Reyhane Askari Hemmat[2], Alex Hoffman[3], Ivan Lazarevich[3], Sudhakar Sah[3], Ehsan Saboori[3], Olivier Mastropietro[3], Yvon Savaria[1], Jean-Pierre David[1]

[1]Ecole Polytechnique Montreal, [2]Mila, Universite de Montreal, [3]Deeplite Inc. Montreal

ICML 2022 HAET Workshop
July 23 2022

Outline:

- Related Works
- Modeling Quantization as an Additive Noise
- Experiments and Results
- Conclusion
- Acknowledgement

Related Works:

Regularization effect of quantization has been studied before:

1. Effect of Quantization on Accuracy Improvement.
2. Analytical Studies.
3. Using Quantization for its Regularization Effect.

Related Works (2):

Regularization effect of quantization has been studied before:

1. Effect of Quantization on Accuracy Improvement.
2. Analytical Studies.
3. Using Quantization for its Regularization Effect.

Our Contribution:

1. Studying relationship between quantization level and regularization effect.
2. Providing empirical study over different quantization levels and methods and different models, datasets and tasks.

Modeling Quantization as an Additive Noise (1):

Weight quantization can be modeled as weight perturbation:

$$f(x, w_q) = f(x, w + \delta)$$

$f(x, w_q)$: Is the predicted target of the network f parameterized by w

δ : We assumed, quantization noise follows a normal distribution: $\delta \sim \mathcal{N}(0, \sigma I)$

Modeling Quantization as an Additive Noise (2):

For simplicity, we consider a regression problem:

$$\mathcal{L} = \frac{1}{m} \sum_{i=1}^{m} \|\hat{y}_i - y_i\|_2^2$$

Applying a first-order Taylor approximation around the weights of the full precision model:

$$\tilde{\mathcal{L}} \approx \mathcal{L} + \frac{\sigma\delta^2}{m} \sum_{i=1}^{m} \|\nabla_w \hat{y}_i\|_2^2$$

Modeling Quantization as an Additive Noise (3):

For simplicity, we consider a regression problem:

$$\mathcal{L} = \frac{1}{m}\sum_{i=1}^{m} \|\hat{y}_i - y_i\|_2^2$$

New loss is penalized by quantization level

Applying a first-order Taylor approximation around the weights of the full precision model:

$$\tilde{\mathcal{L}} \approx \mathcal{L} + \frac{\sigma\delta^2}{m} \sum_{i=1}^{m} \|\nabla_w \hat{y}_i\|_2^2$$

Experiments and Results (1):

We tested regularization effect of quantization over different:

1. Models:
   - Resnet18, Resnet20, Resnet50, Mobilenet V1, Yolo5n.
2. Datasets:
   - CIFAR10, CIFAR100, VOC.
3. Quantization methods:
   - LSQ, PACT, DoReFa.
4. Quantization levels:
   - 2 bits, 4 bits, 8 bits, FP32

Experiments and Results (2):

For each test, we used different augmentations on original dataset:

Test Set	Brightness	Contrast	Defocus Blur	Elastic Transform
Fog	Frost	Gaussian Blur	Gaussian Noise	Glass Blur
Impulse Noise	Jpeg Compression	Motion Blur	Pixelate	Saturate
Shot Noise	Snow	Spatter	Speckle Noise	Zoom Blur

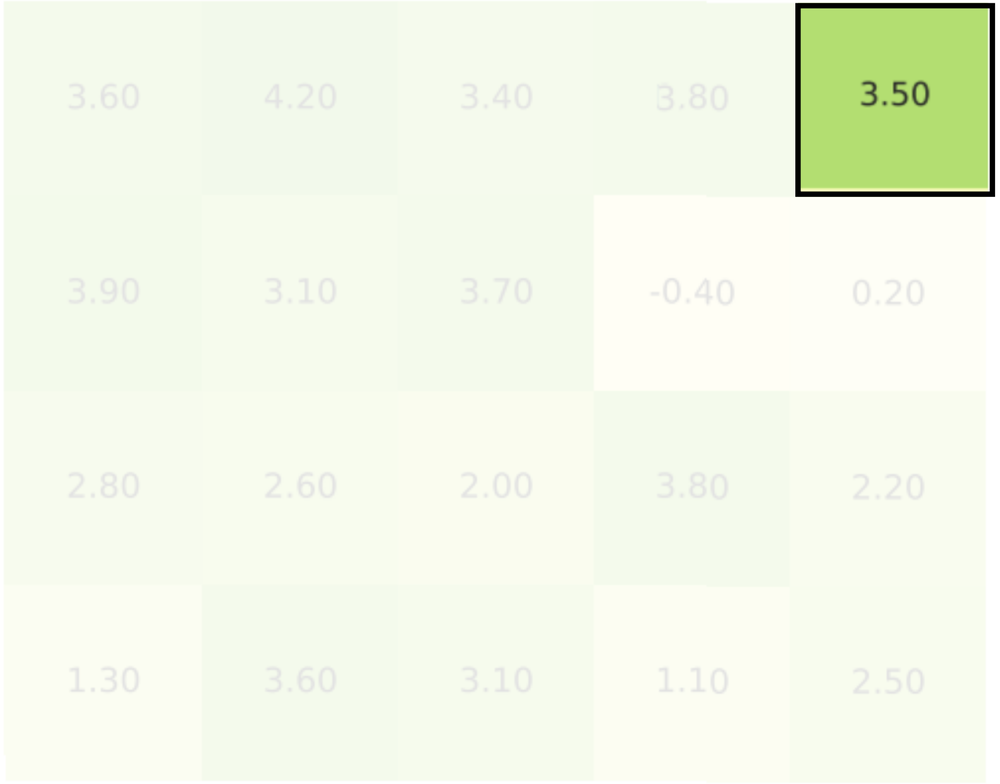
Experiments and Results (3):

The value in each cell corresponds to difference between quantized and FP32 model:

3.60	4.20	3.40	3.80	3.50
3.90	3.10	3.70	-0.40	0.20
2.80	2.60	2.00	3.80	2.20
1.30	3.60	3.10	1.10	2.50

Experiments and Results (3):

The value in each cell corresponds to difference between quantized and FP32 model:



Quantized model performed better

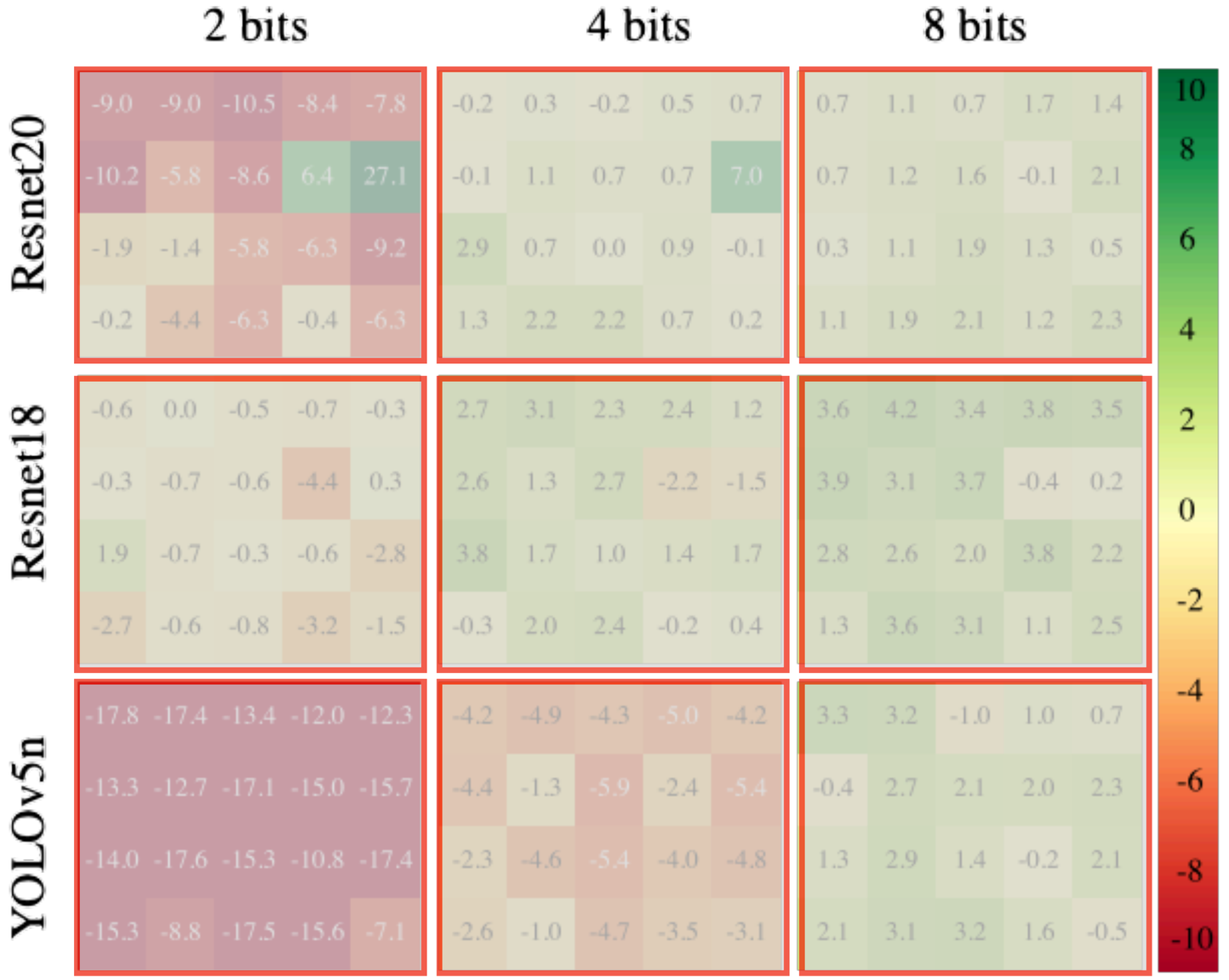
Experiments and Results (3):

The value in each cell corresponds to difference between quantized and FP32 model:

FP32 model performed better

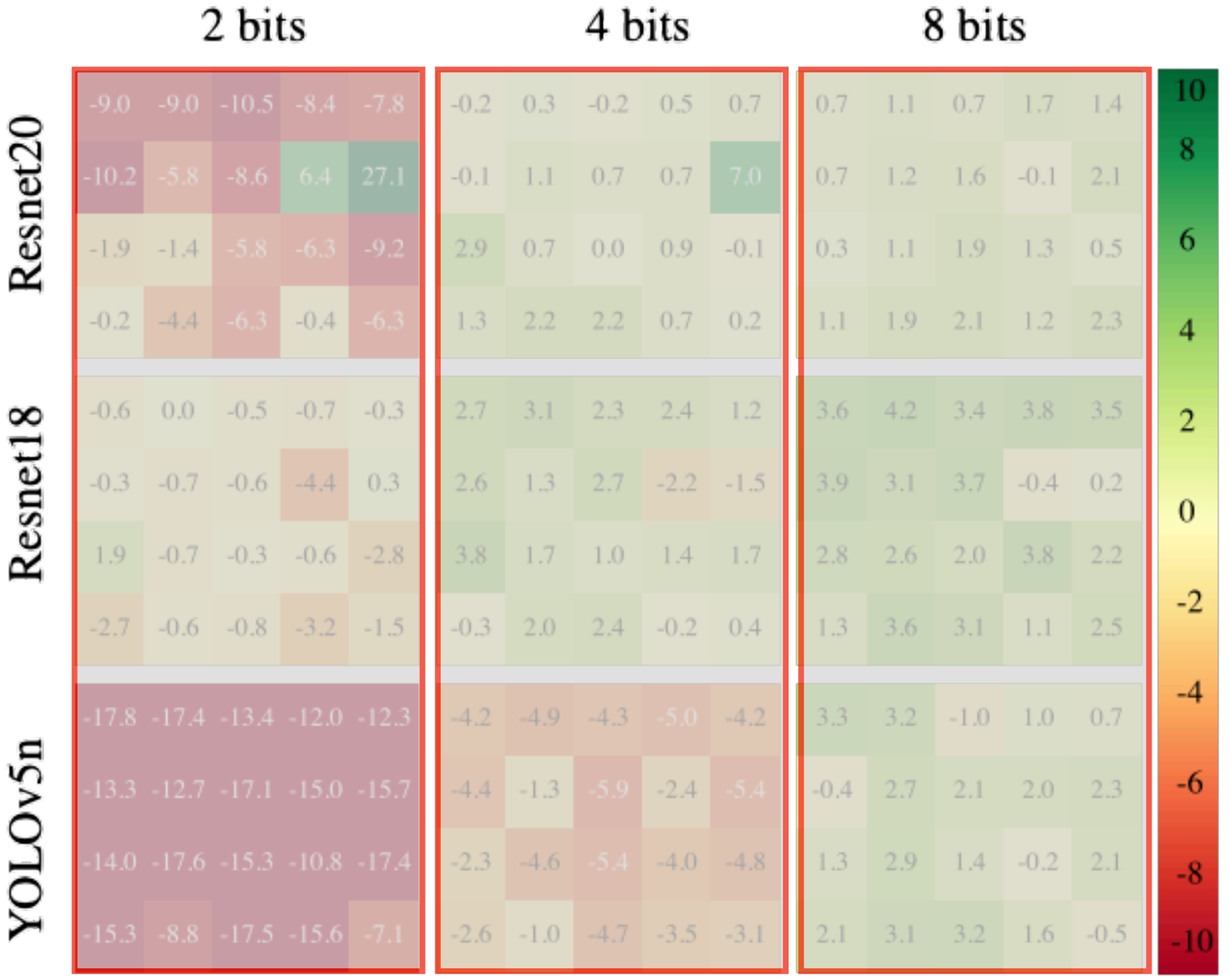
Experiments and Results (4) LSQ Quantization:

1. Nine different experiments.



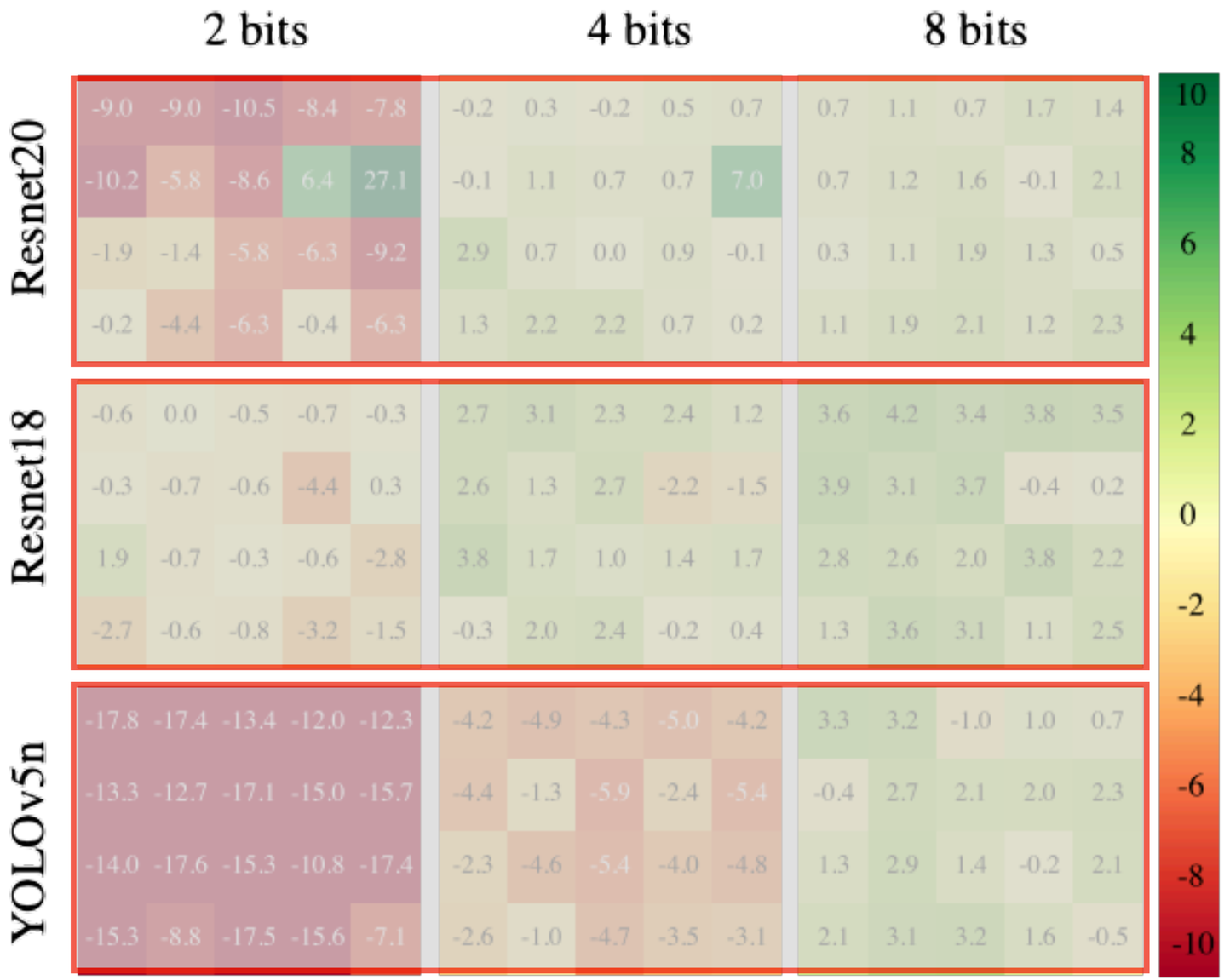
Experiments and Results (4) LSQ Quantization:

1. Nine different experiments.
2. Three different quantization levels.



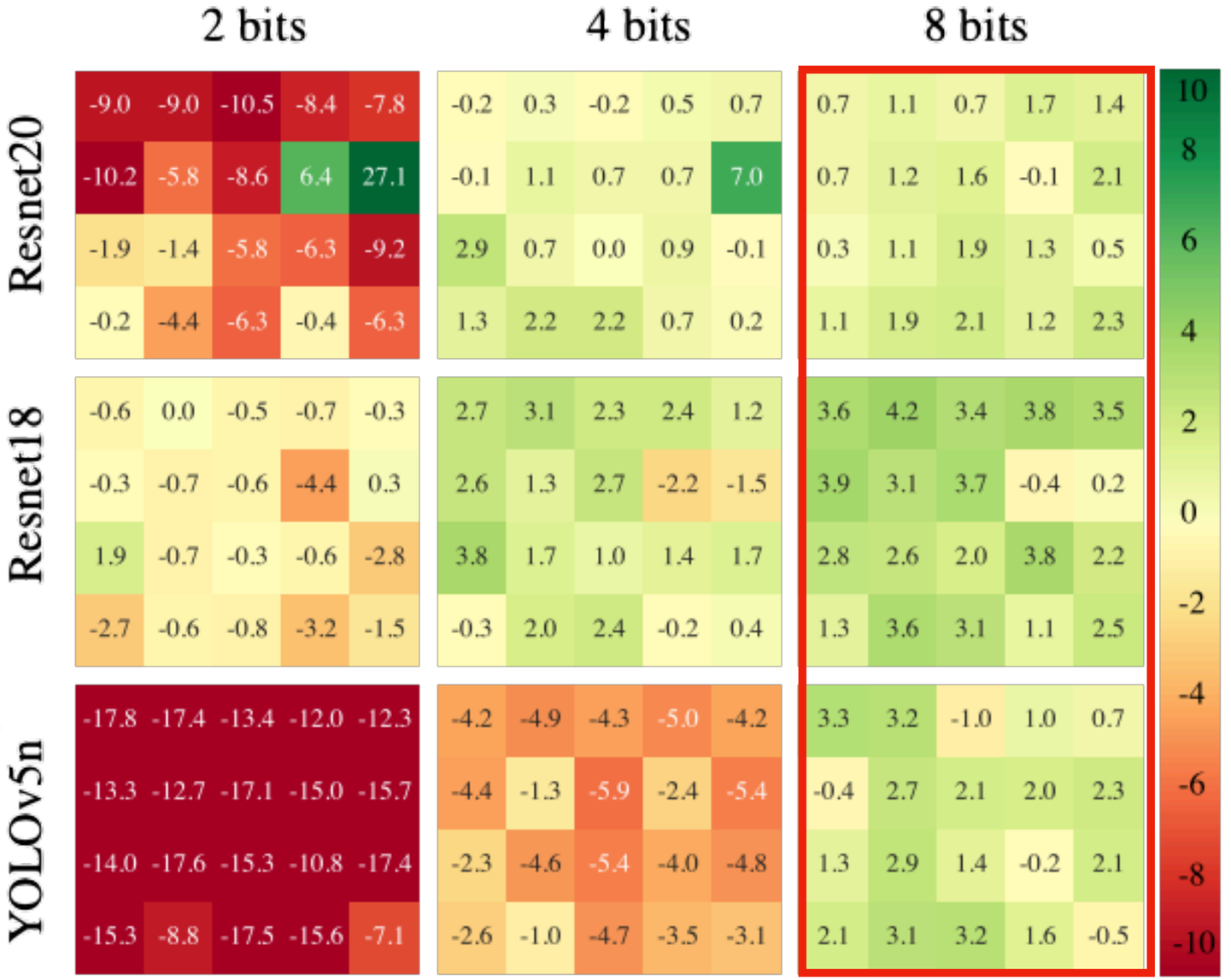
Experiments and Results (4) LSQ Quantization:

1. Nine different experiments.
2. Three different quantization levels.
3. Three different models:
    1. Resnet20: CIFAR10
    2. Resnet18: CIFAR100
    3. YOLOv5n: VOC



Experiments and Results (4) LSQ Quantization:

1. Nine different experiments.
2. Three different quantization levels.
3. Three different models:
    1. Resnet20: CIFAR10
    2. Resnet18: CIFAR100
    3. YOLOv5n: VOC
4. 8-bit models perform consistently better.



Experiments and Results (5) Relative Improvement:

1. Relative improvement score:

$$\textit{Error Improvement} = log(\frac{100 - fval}{100 - qval}) * 100$$

Experiments and Results (5) Relative Improvement:

1. Relative improvement score.

2. 4-bit and 8-bit models have better generalization.

	Quantization Level	Avg Accuracy Augmented Data	Relative Error Improvement Eq.5
Resnet20 Cifar10	2 bits	79.17	-8.32
	4 bits	83.94	2.98
	8 bits	84.07	3.33
	FP32	82.80	0.00
Resnet18 Cifar100	2 bits	49.43	-0.84
	4 bits	51.76	1.21
	8 bits	53.05	2.39
	FP32	50.40	0.00
YOLOv5n VOC	2 bits	14.66	-7.86
	4 bits	24.90	-2.31
	8 bits	30.34	0.96
	FP32	28.78	0.00

Conclusion:

1. We formalized quantization noise and study how it effects training.

2. We showed how quantization level is correlated to the regularization term.

3. We provided a extensive list of experiments where we tested our hypothesis on different models, tasks, quantization methods and levels.

4. Based on our study, we propose 8-bit quantization provides a reliable form of regularization in different vision tasks and models.

Acknowledgements

The authors would like to thank:

1. Mohammad Pezeshki and Anush Sankaran for helpful discussion about designing empirical studies presented in this paper

2. FRQNT and NSERC for providing financial support for this project.