KOREA UNIVERSITY

Revisiting batch normalization in quantization of super-resolution networks

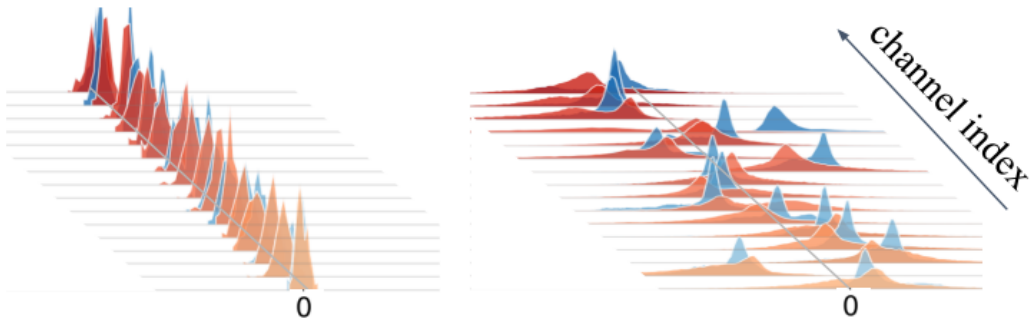
Myungjun Son, Dongjea Kang, Hongjae Lee, Jun-Sang Yoo and Seung-Won Jung

(submitted to IEIE AISP)

고려대학교 KOREA UNIVERSITY

손명준

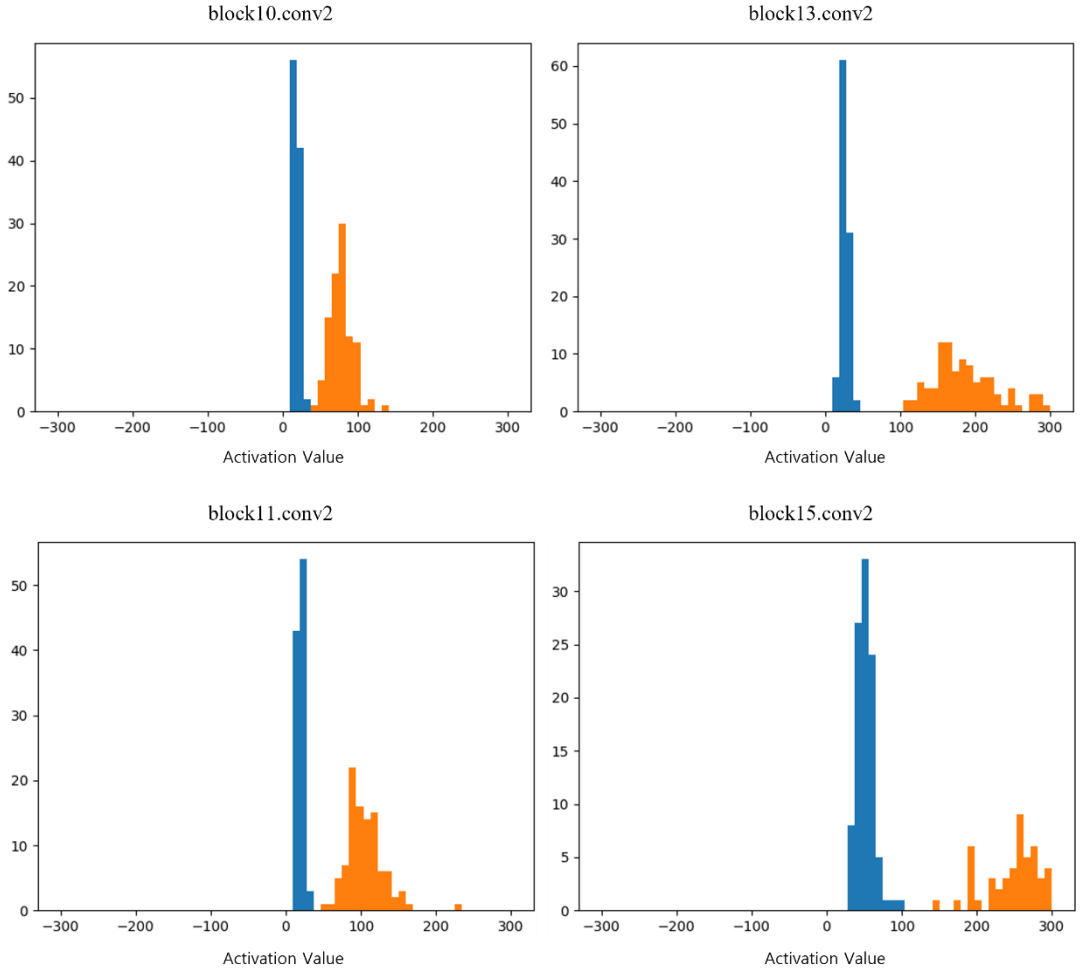
Introduction



(a) ResNet-18 (b) EDSR

Figure 2: Channel-wise feature map distributions of two distinct images (red and blue) in pre-trained (a) ResNet [20] on image classification task and (b) EDSR [32] on image SR task. In SR networks, channels present diverse non-zero distributions that also vary upon the input image.

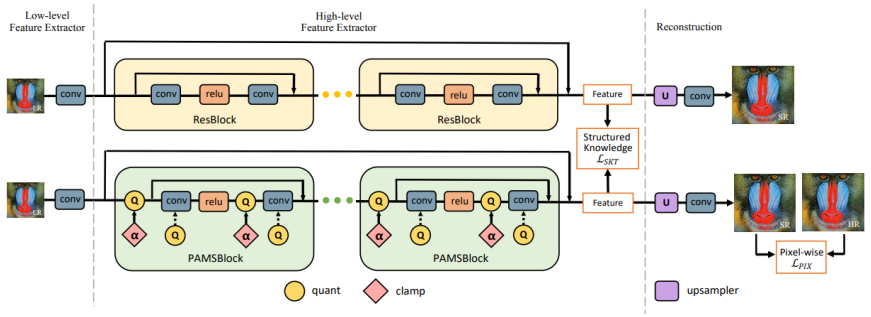
Hong, Cheeun, et al.



Introduction

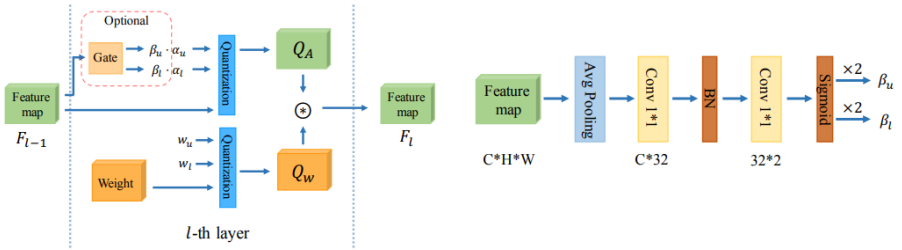
PAMS: Quantized Super-Resolution via Parameterized Max Scale

Huixia Li[1†], Chenqian Yan[1†], Shaohui Lin[2], Xiawu Zheng[1],
Baochang Zhang[3], Fan Yang[4], Rongrong Ji[15⋆]



Dynamic Dual Trainable Bounds for Ultra-low Precision Super-Resolution Networks

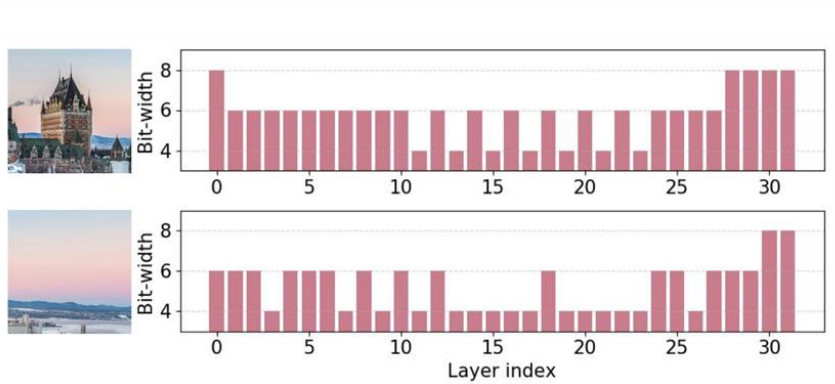
Yunshan Zhong[1,2], Mingbao Lin[3], Xunchao Li[2], Ke Li[3],
Yunhang Shen[3], Fei Chao[1,2], Yongjian Wu[3], Rongrong Ji[1,2⋆]



CADyQ: Content-Aware Dynamic Quantization for Image Super-Resolution

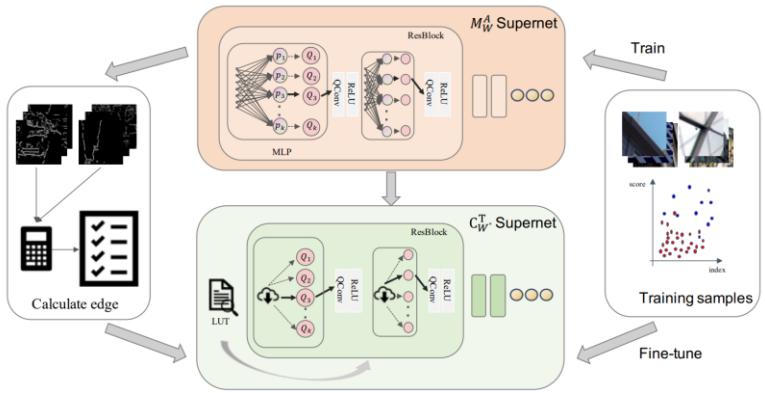
ECCV 2022

Cheeun Hong[1], Sungyong Baik[3], Heewon Kim[1],
Seungjun Nah[1,4], and Kyoung Mu Lee[1,2]

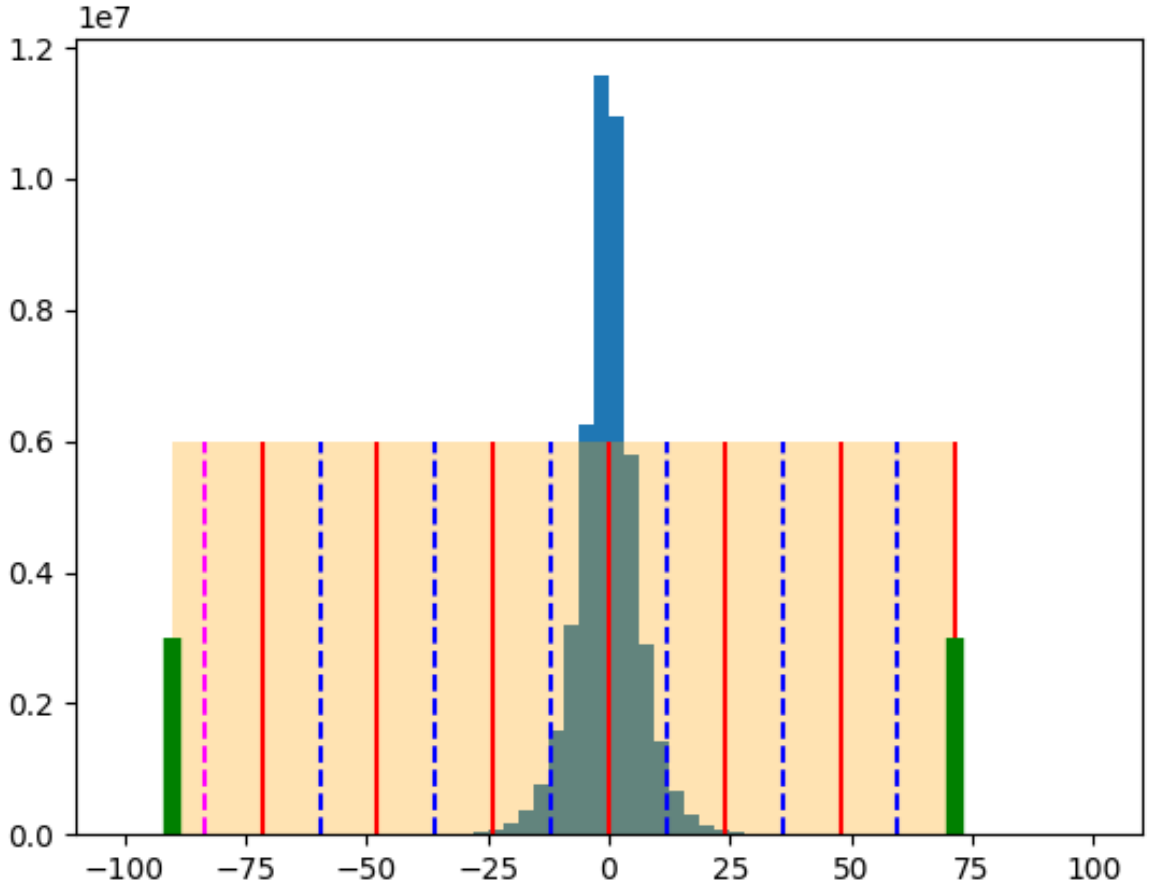


CABM: Content-Aware Bit Mapping for Single Image Super-Resolution Network with Large Input

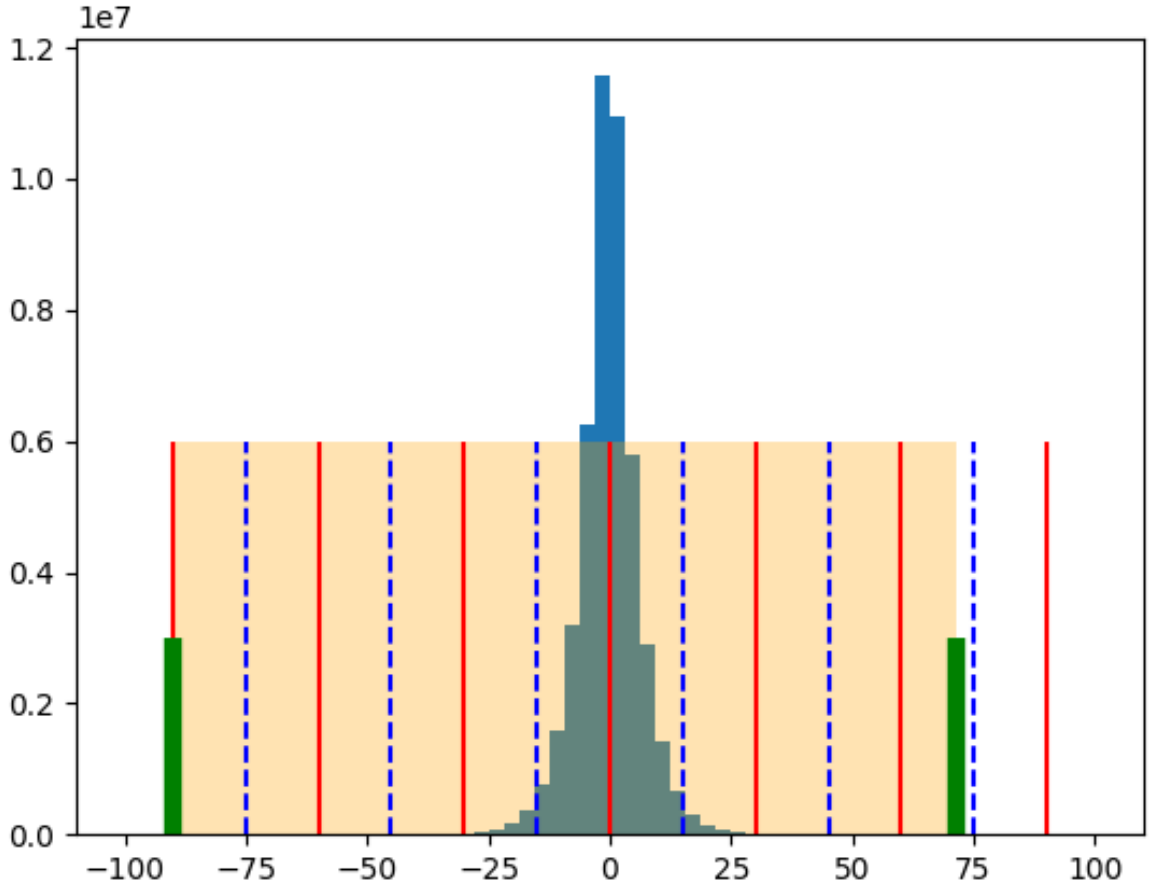
Senmao Tian[1,2] Ming Lu[3] Jiaming Liu[2,4] Yandong Guo[5]
Yurong Chen[3] Shunli Zhang[1*]



Quantization unfitness



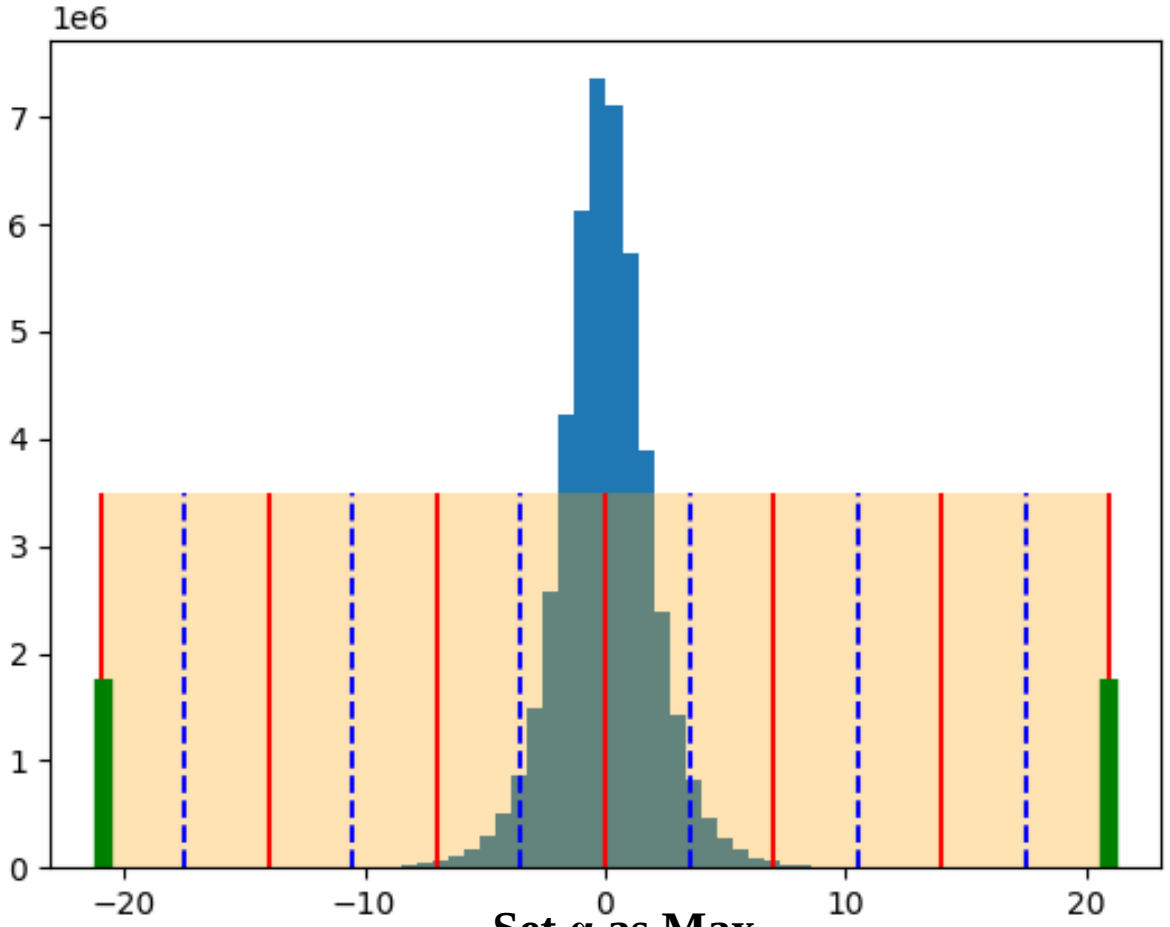
Set α as Max



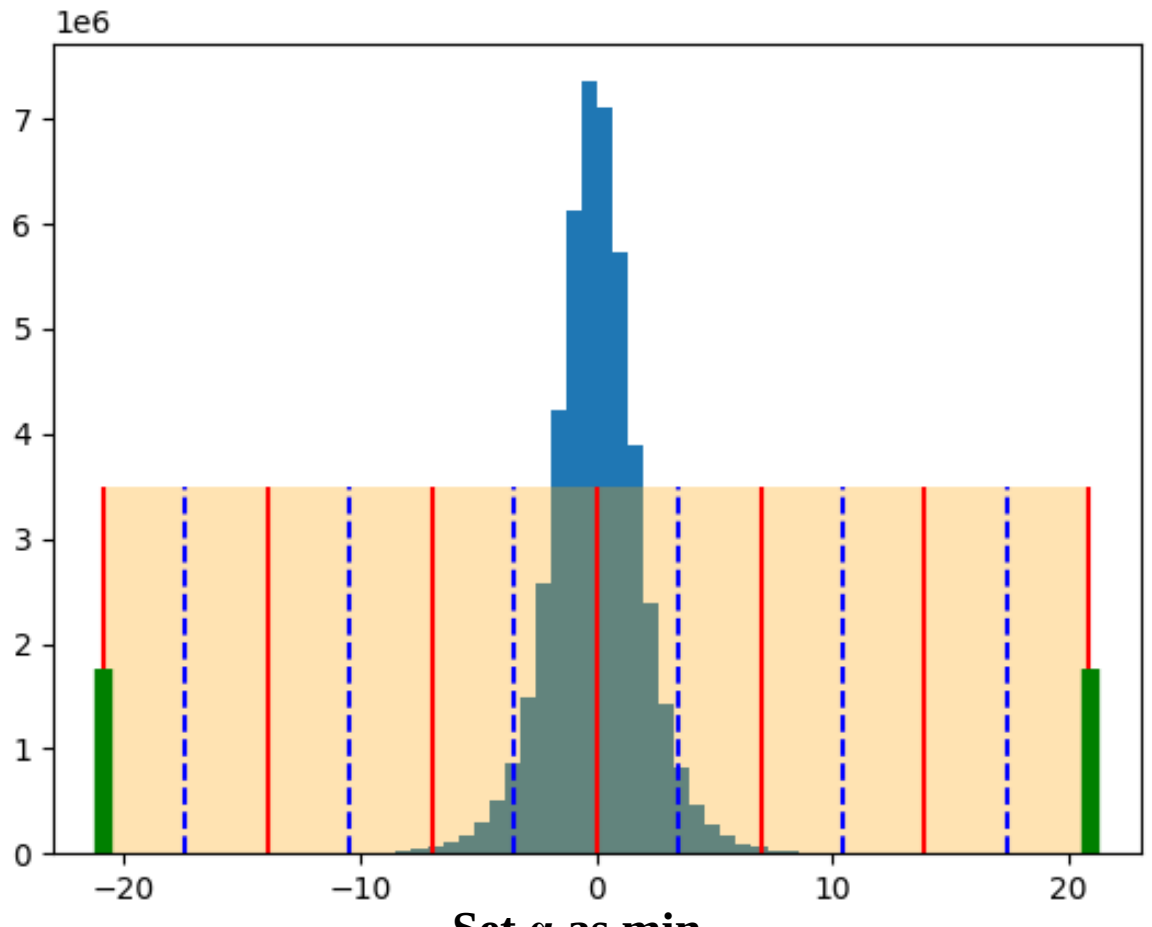
Set α as min

Red line : quantization level
Blue dotted line : quantization edge
Green bar : min / Max of activation

Quantization unfitness



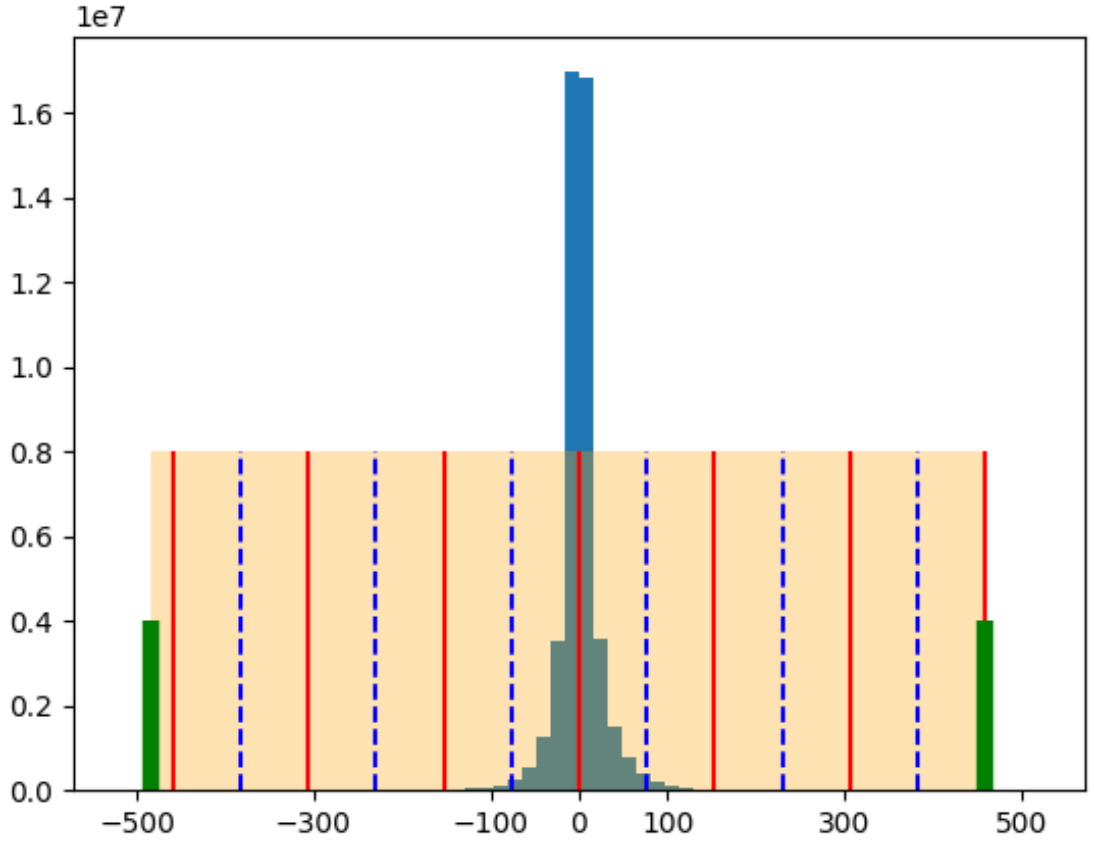
Set α as Max



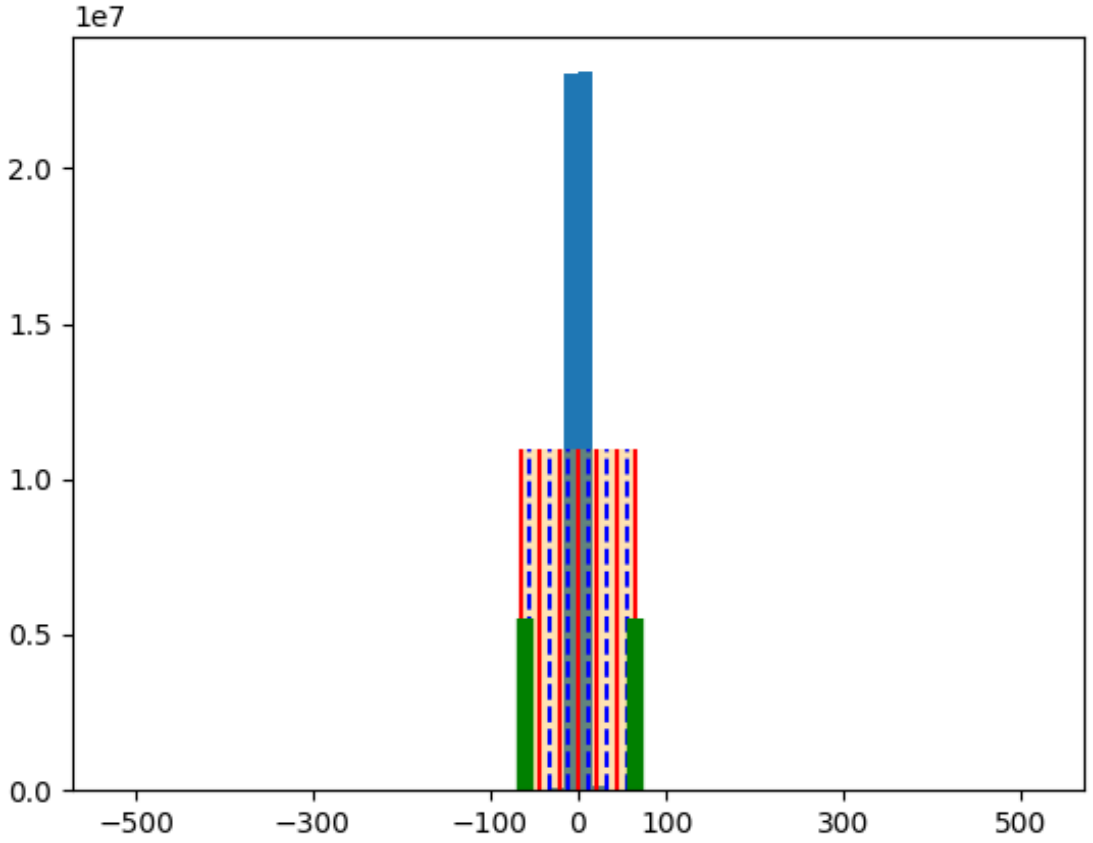
Set α as min

Red line : quantization level
Blue dotted line : quantization edge
Green bar : min / Max of activation

Quantization range



w/o BN



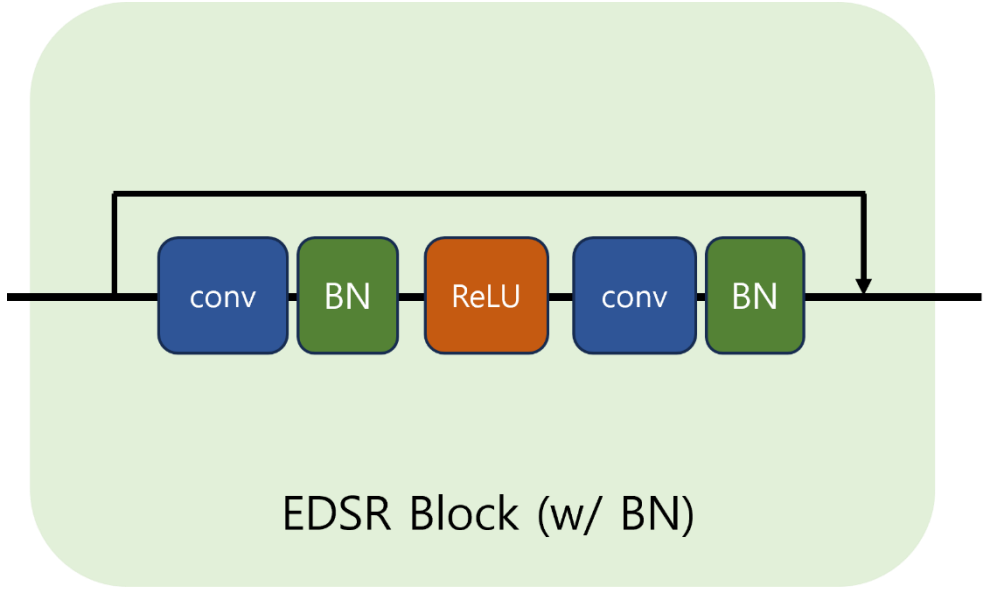
w BN

Red line : quantization level
Blue dotted line : quantization edge
Green bar : min / Max of activation

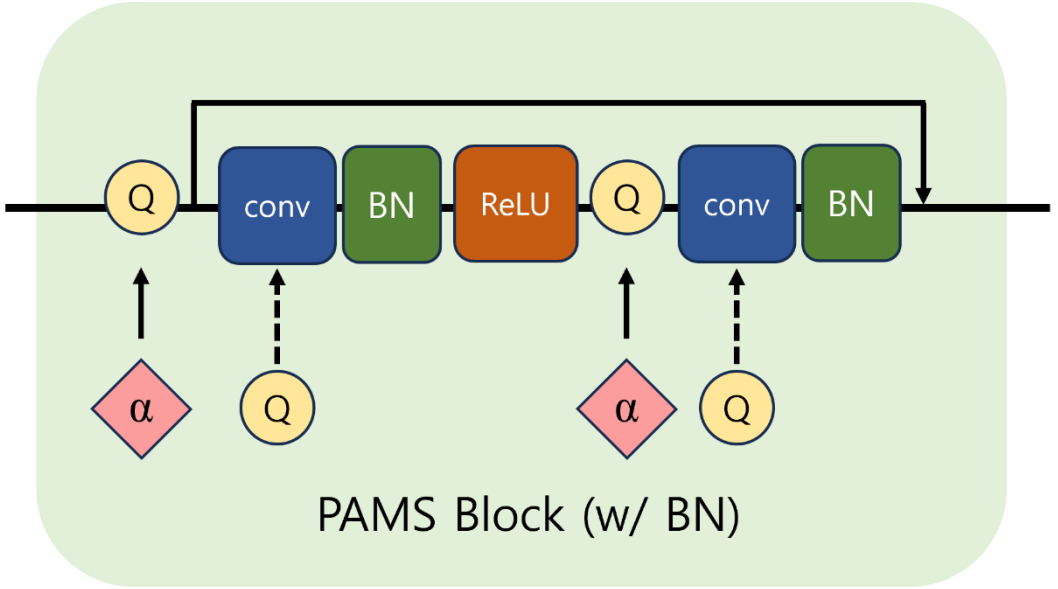
Quantization range

레이어	α 설정 기준	양자화 간격
block11.conv2	Max	75.90
	Min	57.72
block11.conv2†	Max	9.73
	Min	8.38
block13.conv2	Max	96.43
	Min	91.52
block13.conv2†	Max	12.19
	Min	11.94
block15.conv2	Max	153.34
	Min	161.82
block15.conv2†	Max	21.81
	Min	20.01

Experiment



(a)



(b)

Experiment

KOREA UNIVERSITY

PAMS

	Precision	Urban100	Test2k	Test4k
EDSR	-	26.02	27.59	28.99
EDSR†	-	25.84	27.54	28.92
EDSR+PAMS	8bits	26.03	27.59	28.98
+ w/ BN (ours)	8bits	25.85	27.54	28.92
EDSR+PAMS	4bits	25.30	27.40	28.73
+ w/ BN (ours)	4bits	**25.40**	27.39	28.73
EDSR+PAMS	3bits	23.50	26.59	27.69
+ w/ BN (ours)	3bits	**25.00**	**27.31**	**28.60**
IDN	-	25.50	27.40	28.74
IDN†	-	25.53	27.42	28.76
IDN+PAMS	8bits	25.46	27.39	28.73
+ w/ BN (ours)	8bits	**25.50**	**27.41**	**28.75**
IDN+PAMS	4bits	24.35	26.91	28.11
+ w/ BN (ours)	4bits	**24.64**	**27.03**	**28.27**
IDN+PAMS	3bits	23.24	26.36	27.44
+ w/ BN (ours)	3bits	**23.99**	**26.71**	**27.89**

Experiment

KOREA UNIVERSITY

DDTB

	Precision	Urban100	Test2k	Test4k
EDSR	-	26.02	27.59	28.99
EDSR†	-	25.84	27.54	28.92
EDSR+PAMS	8bits	26.03	27.60	28.99
+ w/ BN (ours)	8bits	25.81	27.54	28.92
EDSR+PAMS	4bits	25.66	27.51	28.86
+ w/ BN (ours)	4bits	25.55	28.49	28.85
EDSR+PAMS	3bits	25.34	27.41	28.72
+ w/ BN (ours)	3bits	25.31	**27.42**	**28.75**
IDN	-	25.50	27.40	28.74
IDN†	-	25.53	27.42	28.76
IDN+PAMS	8bits	25.02	27.26	28.53
+ w/ BN (ours)	8bits	**25.03**	**27.27**	**28.54**
IDN+PAMS	4bits	24.73	27.12	28.34
+ w/ BN (ours)	4bits	**24.74**	**27.13**	**28.37**
IDN+PAMS	3bits	23.73	26.66	27.81
+ w/ BN (ours)	3bits	**24.24**	**26.90**	**28.09**

Conclusion

1. DCNN 기반 SR 모델 연구는 **BN을 제거**하여 모델의 성능을 높이는 흐름으로 진행됨
2. 4bit 이하의 **낮은 비트 정밀도**에서 BN을 추가하면 양자화 오류가 **감소**함
3. BN을 포함하는 과정에서 **메모리 요구량 증가**와 **고주파 요소의 손실** 등의 문제가 발생함
4. BN에 의해 발생하는 문제를 줄이는 **양자화에 특화**된 **새로운 정규화 알고리즘**의 연구가 필요할 것으로 보임

KOREA UNIVERSITY

Q&A