Samsung Research

GENIE: Show Me the Data for Quantization

(WED-AM-366)

Yongkweon Jeon*, Chungman Lee*, Ho-young Kim*

Samsung Research

Correspondence to: *dragwon.jeon@samsung.com*

Presenter: Ho-young Kim

June, 21st, 2023

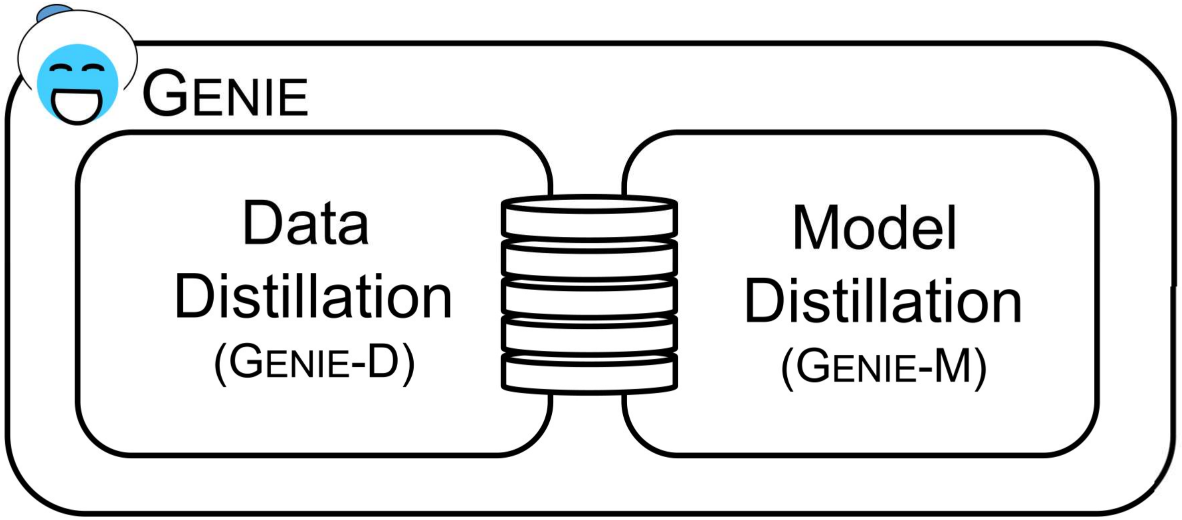
JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

GENIE: The novel approach for Zero-shot Quantization

Samsung Research

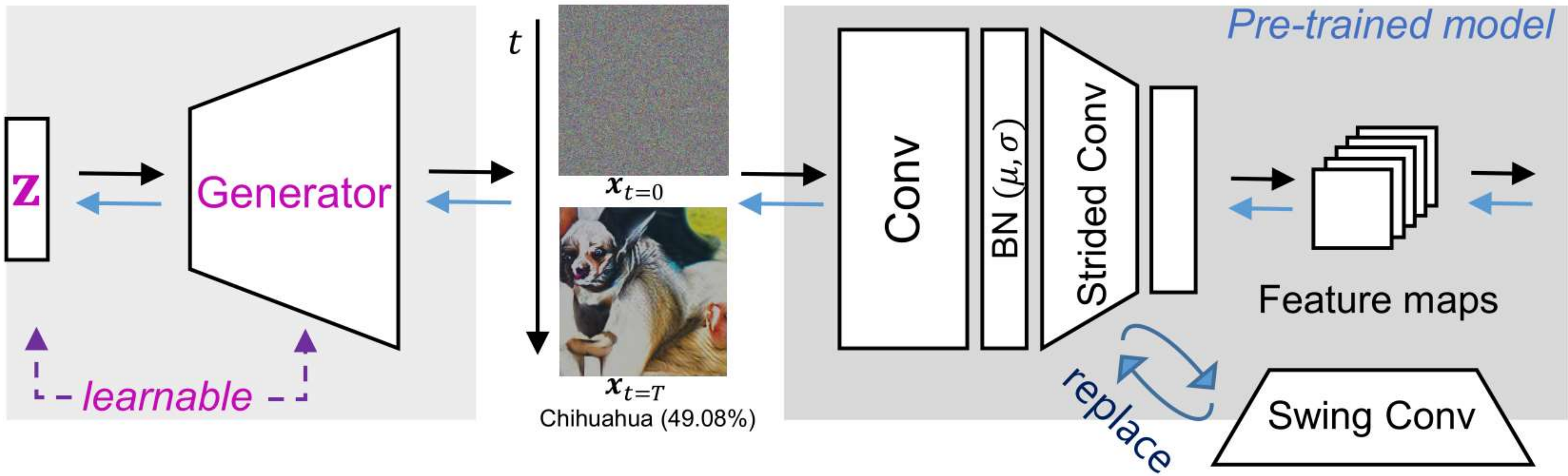
- Zero-shot quantization (ZSQ): Quantization method using only synthetic data instead of the real data
  - Distill-based approach (DBA)
  - Generator-based approach (GBA)

- Unlike most former approaches, we adopt PTQ rather than QAT as a quantization scheme, and it improves ZSQ performance significantly within much shorter time.



JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

GENIE-D overview



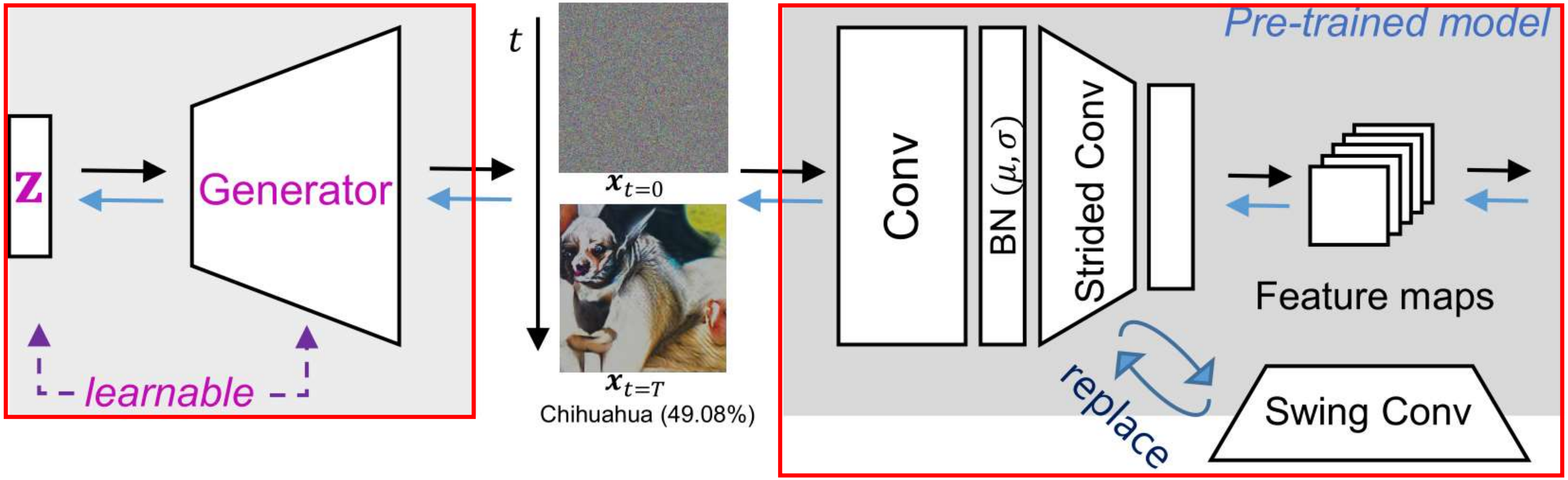
- Distill fake data which meets batch normalization statistics (BNS) $\boldsymbol{\mu}_l$ and $\boldsymbol{\sigma}_l$ of the pretrained model

$$\mathcal{L}_{BNS}^{D} = \sum_{l=0}^{L} (\|\boldsymbol{\mu}_l^s - \boldsymbol{\mu}_l\|^2 + \|\boldsymbol{\sigma}_l^s - \boldsymbol{\sigma}_l\|^2)$$

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

New Features in GENIE-D

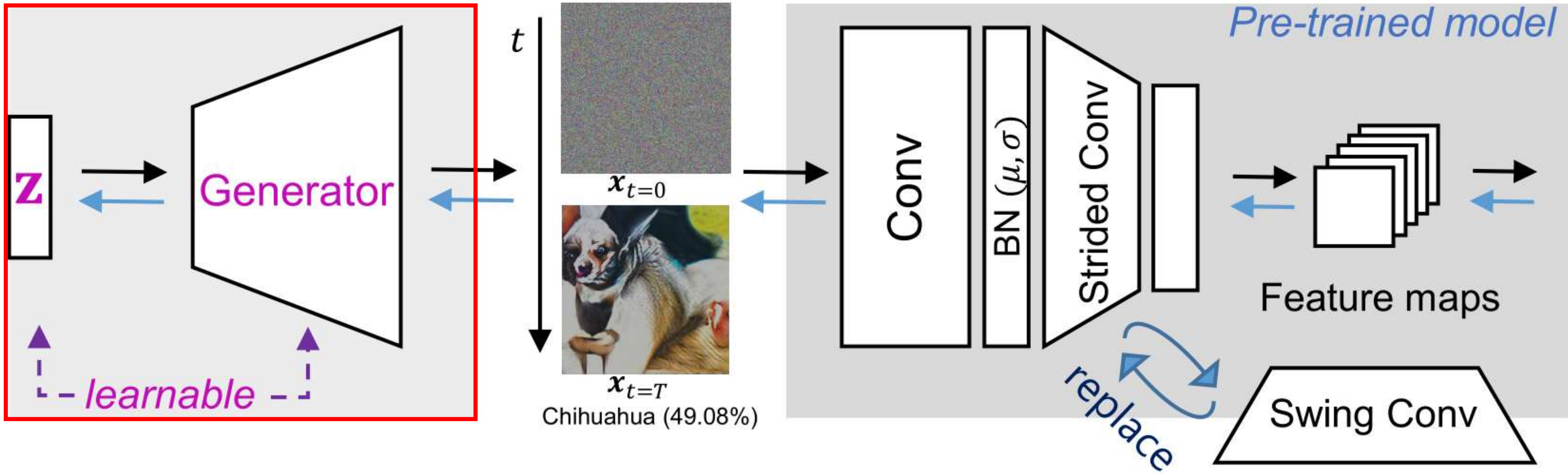
Samsung Research



- ➢ Learnable latent vector $\mathbf{z}$
- ➢ Swing Convolution

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

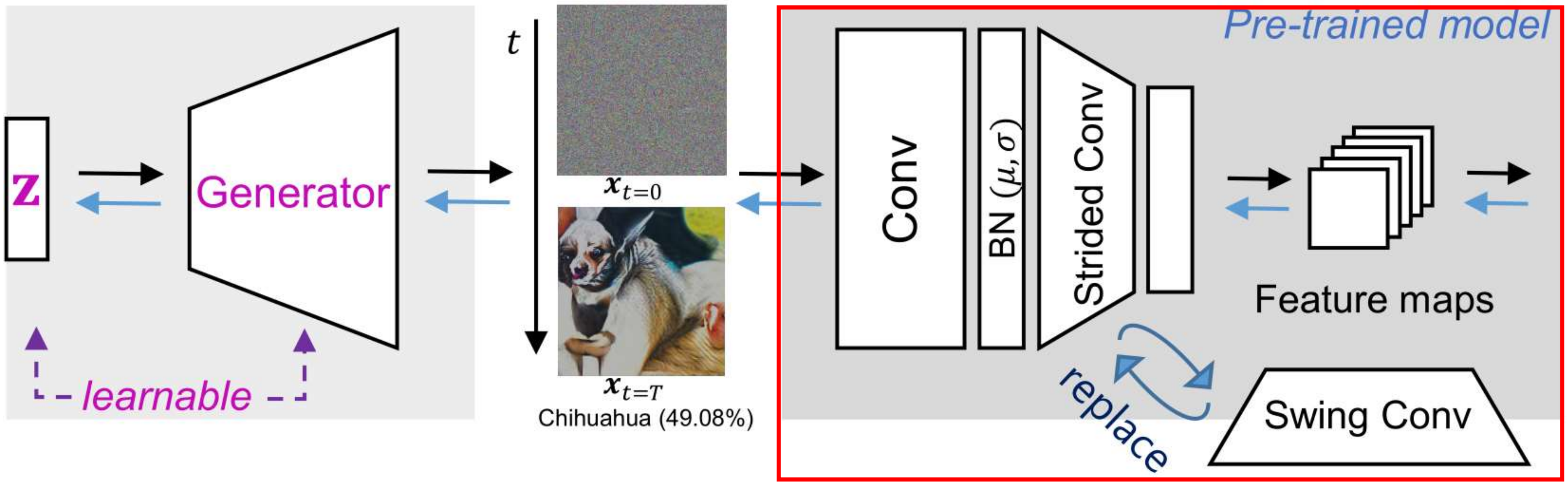
Learnable latent vector $\mathbf{z}$



- Inspired by Generative Latent Optimization (GLO) (Bojanowski et al., 2018)
  - No need to fit to random noises → Stable convergence of generator (see Fig A5)
  - Exploring in the latent space → Efficient distillation of the pretrained model's knowledge

Swing Convolution

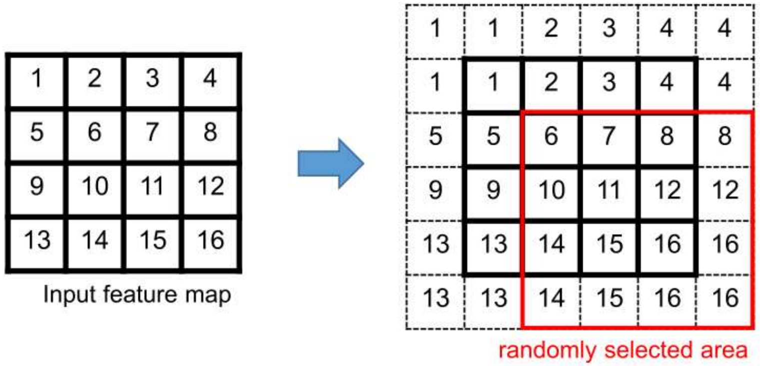
Samsung Research



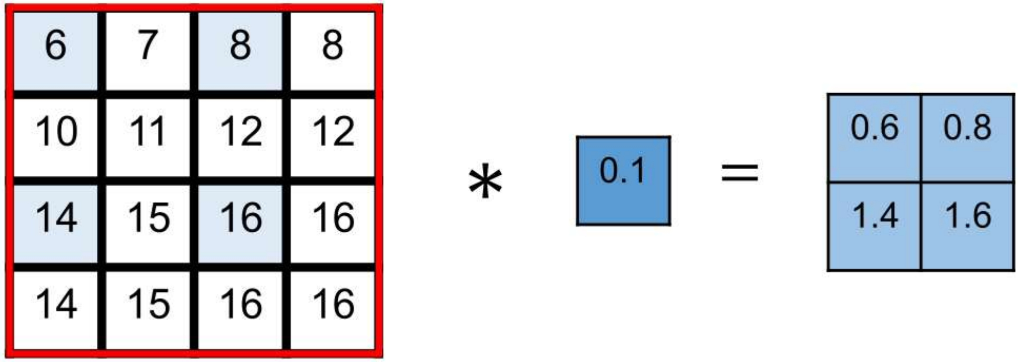
- Replace all n-strided convolution layers with swing convolution layers of same stride when only synthesizing the dataset
  - Decreasing information loss
  - Reducing checkerboard artifact

(a) Distilling without *swing conv* (b) Distilling with *swing conv*

The Mechanism of Swing Convolution



(a) Reflection padding & random crop.



(b) 2-stride convolution (`conv2d(kernel_size=1, stride=2)`).

- Swing convolution
  - Randomly select feature maps to be convolved at each step
    - Padding is required for the margin of randomness
  - Every pixel can deliver information due to the stochasticity.
  - Since random selection is done uniformly, all pixels are updated evenly after enough steps.

- Normal n-strided convolution
  - Convolve only the information of a fixed feature map in any step
  - There are unreachable pixels, which never provide information for data distillation.
  - Pixels are updated unevenly, and this incurs checkerboard artifacts (Odena et al., 2016)

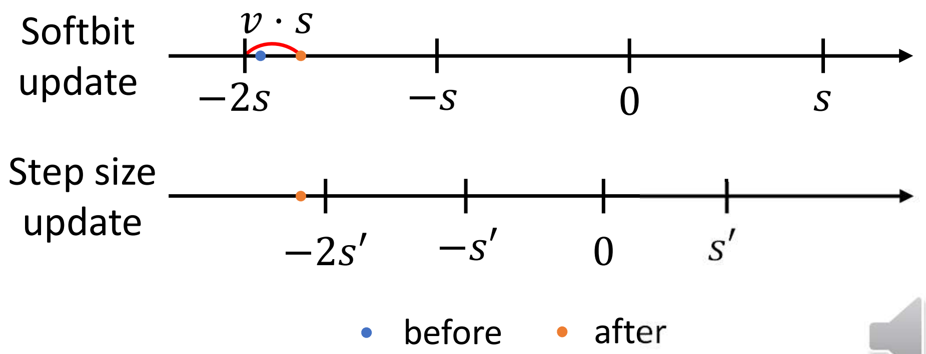
GENIE-M: Sub-module for PTQ

Samsung Research

- Quantization is a task that maps parameters to proper grid points on the range set by a step size s with the minimal performance loss.

- In AdaRound (Nagel et al., 2020), a PTQ scheme on which GENIE-M is based, the authors optimize only softbit $v \in [\mathbf{0}, \mathbf{1}]$ to find a mapping for higher accuracy, but use a fixed step size at initialized.
  - They pointed out that the joint optimization of s and v is not trivial.

- Example. Conflict of updates

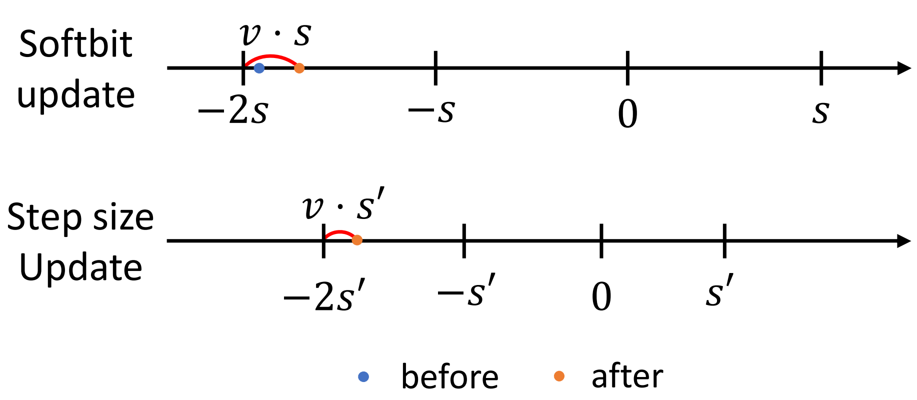


JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

The Algorithm of GENIE-M

- Enable joint optimization by releasing the dependency between s and v (line 3)

- Example. Resolution of the conflict



Algorithm 2 CLASS GENIE-M

1: **def**: __INIT__(self, $\boldsymbol{W}$, $bits$)

2: self.s $\leftarrow$ SetStepSize($\boldsymbol{W}$, $bits$)

3: self.$\boldsymbol{B}$ $\leftarrow$ clip($\left\lfloor \frac{\boldsymbol{W}}{\text{self}.s} \right\rfloor$, n, p).detach()

4: self.$\boldsymbol{V}$ $\leftarrow \frac{\boldsymbol{W}}{self.s} - self.\boldsymbol{B}$

5: **def**: FORWARD(self)

6: return self.s $\times$ (self.$\boldsymbol{B}$+self.$\boldsymbol{V}$)

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Experimental Results: Ablation study

Samsung Research

Table 2. Result of the ablation study on CNN Models (top-1 accuracy (%))

	#Bits (W/A)	Swing	Generator	z	Genie-M	ResNet-18	ResNet-50	MobileNetV2	MnasNet-1.0
FP	32/32					71.08	77.00	72.49	73.52
M1	4/4					69.19	74.87	66.22	58.52
M2					✓	69.25	74.94	66.25	58.82
M3		✓				69.49	75.43	67.80	63.98
M4			✓			69.17	74.96	66.41	64.63
M5			✓	✓		69.58	75.39	67.92	66.15
M6		✓	✓	✓		69.62	75.47	68.28	66.55
M7		✓	✓	✓	✓	**69.66**	**75.59**	**68.38**	**66.94**
M1	2/4					61.96	66.72	36.58	31.22
M2					✓	62.62	66.95	37.12	32.45
M3		✓				63.74	69.44	44.00	34.64
M4			✓			60.13	65.28	34.92	35.50
M5			✓	✓		64.06	70.16	47.96	45.47
M6		✓	✓	✓		64.34	69.87	49.89	47.34
M7		✓	✓	✓	✓	**65.10**	**69.99**	**53.38**	**48.21**

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Experimental Results: CNN

Samsung Research

Table 3. Evaluation of CNN Models I (top-1 accuracy (%))

	Methods	#Bits (W/A)	ResNet-18	ResNet-50	MobileNetV2	MnasNet-1.0
	Full Prec.	32/32	71.08	77.00	72.49	73.52
Single Model	ZeroQ+BRECQ‡	4/4	69.32	73.73	49.83	52.04
	KW+BRECQ‡		69.08	74.05	59.81	55.48
	IntraQ†+BRECQ		68.77	68.16	63.78	-
	Qimera+BRECQ		67.86	72.90	58.33	-
	GENIE-D+BRECQ [ours]		69.70	74.89	64.68	55.42
	GENIE [ours]		**69.66**	**75.59**	**68.38**	**66.94**
Mix*	MixMix+BRECQ‡		69.46	74.58	64.01	57.87
	GENIE-D+BRECQ [ours]		69.71	74.89	64.97	51.25
	GENIE [ours]		**69.77**	**75.41**	**68.70**	**67.45**
Real	QDROP§		69.62	75.45	68.84	-
	GENIE-M [ours]		**69.81**	**75.61**	**69.23**	**68.29**
Single Model	ZeroQ+BRECQ	2/4	61.63	64.16‡	34.39	13.83
	KW+BRECQ‡		-	57.74	-	-
	IntraQ†+BRECQ		55.39	44.78	35.38	-
	Qimera+BRECQ		47.80	49.13	3.73	-
	GENIE-D+BRECQ [ours]		64.24	69.38	45.28	29.72
	GENIE [ours]		**65.10**	**69.99**	**53.38**	**48.21**
Mix*	MixMix+BRECQ‡		-	66.49	-	-
	GENIE-D+BRECQ [ours]		64.91	69.96	42.19	31.22
	GENIE [ours]		**65.44**	**70.62**	**53.36**	**49.65**
Real	QDROP§		65.25	70.65	54.22	-
	GENIE-M [ours]		**66.23**	**71.06**	**57.74**	**55.57**

Table 4. Evaluation of CNN Models II (top-1 accuracy (%))

Methods		ResNet-18	ResNet-50	MobileNetV2
Full Prec.		71.47	77.73	73.03
GDFQ+AIT*	4/4	65.51	64.24	65.39
Qimera+AIT*		66.83	67.63	66.81
ARC+AIT*		65.73	68.27	66.47
ZAQ†		-	70.06	-
IntraQ‡		66.47	-	65.10
GENIE-D+AIT		66.91	-	-
GENIE [ours]		**68.69**	**74.21**	**69.59**
GDFQ+AIT	2/4	0.10	0.10	0.11
Qimera+AIT		0.10	0.10	0.12
ARC+AIT		0.11	0.10	0.13
IntraQ		0.14	-	0.17
GENIE-D+AIT		**0.50**	-	-
GENIE [ours]		**58.73**	**54.83**	**45.84**

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Experimental Results: GENIE-M performance on real data

Samsung Research

Table 5. Performance comparison using ***real samples*** $(1K)$ (top-1 Accuracy (%))

Methods	#Bits (W/A)	ResNet-18	ResNet-50	MobileNetV2	RegNetX-600M	RegNetX-3.2G	MnasNet-2.0
Full Prec.	32/32	71.08	77.00	72.49	73.71	78.36	76.68
AdaRound+QDROP†	4/4	69.10	75.03	67.89	70.62	76.33	72.39
GENIE-M+No Drop **[ours]**	4/4	69.13	74.93	68.22	70.87	76.50	72.68
GENIE-M+QDROP [ours]	4/4	**69.35**	**75.21**	**68.65**	**71.13**	**76.75**	**73.37**
AdaRound+No Drop†	2/4	64.16	69.60	51.61	61.52	70.29	60.00
AdaRound+QDROP†	2/4	64.66	70.08	52.92	63.10	70.95	62.36
GENIE-M+No Drop **[ours]**	2/4	65.27	70.39	55.55	63.66	71.79	62.76
GENIE-M+QDROP [ours]	2/4	**65.77**	**70.51**	**56.38**	**64.55**	**72.35**	**64.10**
AdaRound+QDROP†	3/3	65.56	71.07	54.27	64.53	71.43	63.47
GENIE-M+No Drop **[ours]**	3/3	65.50	71.08	55.28	64.37	72.05	62.17
GENIE-M+QDROP [ours]	3/3	**66.16**	**71.61**	**57.54**	**65.68**	**72.72**	**64.80**
AdaRound+No Drop†	2/2	46.64	47.90	4.55	25.52	39.76	9.51
AdaRound+QDROP†	2/2	51.14	54.74	8.46	38.90	52.36	22.70
GENIE-M+No Drop **[ours]**	2/2	50.52	51.80	12.63	34.03	40.97	19.60
GENIE-M+QDROP [ours]	2/2	**53.71**	**56.71**	**17.10**	**42.00**	**55.31**	**28.56**

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Conclusion

Samsung Research

- We propose a novel zero-shot quantization approach, both image distillation method and PTQ scheme, for CNN, called GENIE.
  - GENIE-D successfully synthesizes the meaningful data by adopting GLO and swing convolution
  - GENIE-M Jointly optimizes both quantization parameters as learnable parameters

- We have achieved a new state-of-the-art accuracy of zero-shot quantization on various CNN models.

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Samsung Research

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Thank You

Correspondence to: *dragwon.jeon@samsung.com*