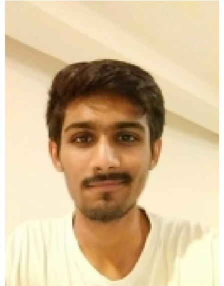
CVPR VIRTUAL JUNE 19-25

Binary Networks for Computer Vision
CVPR 2021 Workshop

BNAS-v2: A Summary with Empirical Improvements

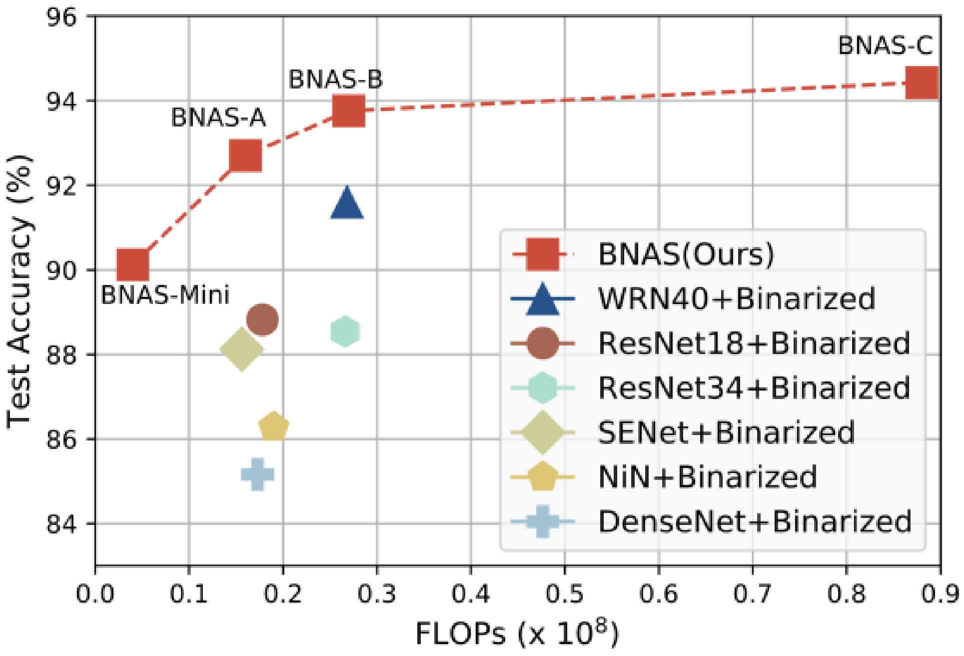
Dahyun Kim[1] Kunal P. Singh[2] Jonghyun Choi[1]



[1] GIST [2] AI2

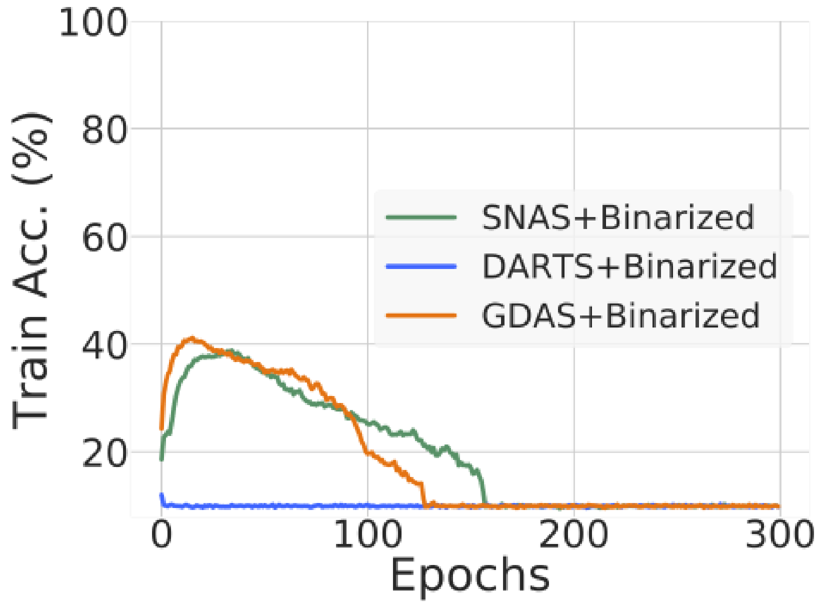
Introduction

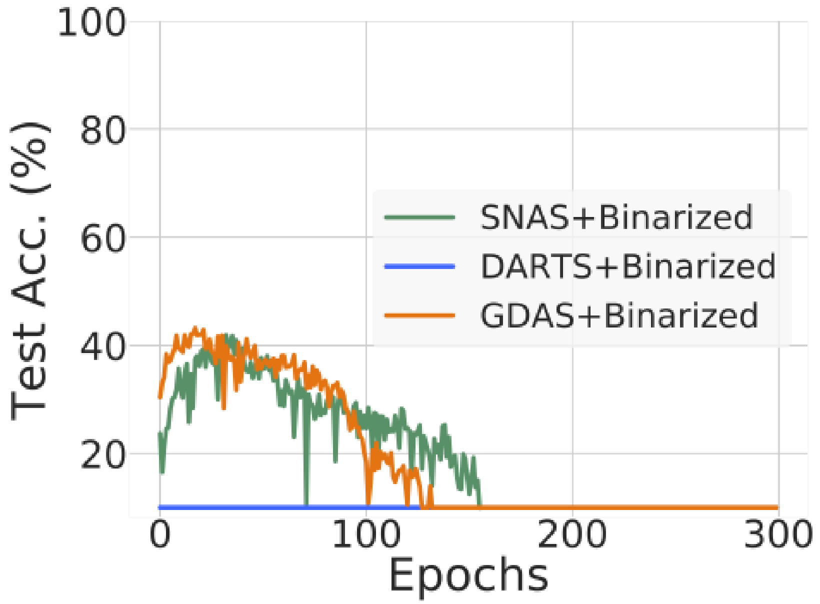
- Conjecture binarizing well-known FP networks is sub-optimal
- Want to search for architectures that are specialized for the binary domain
- Use cell-based NAS method with:
  - New binary search space
  - New cell template
  - New search objective



Want to Search Binary Networks

- We use cell based differentiable search methods
- Tried SNAS[1], DARTS[2], GDAS[3]





They don't work (∵ quantization error)

[1] Xie et al., "**SNAS: stochastic neural architecture search**," ICLR 2019
[2] Liu et al., "**DARTS: Differentiable architecture search**," ICLR 2019
[3] Dong et al., "**Searching for A Robust Neural Architecture in Four GPU Hours**," CVPR 2019

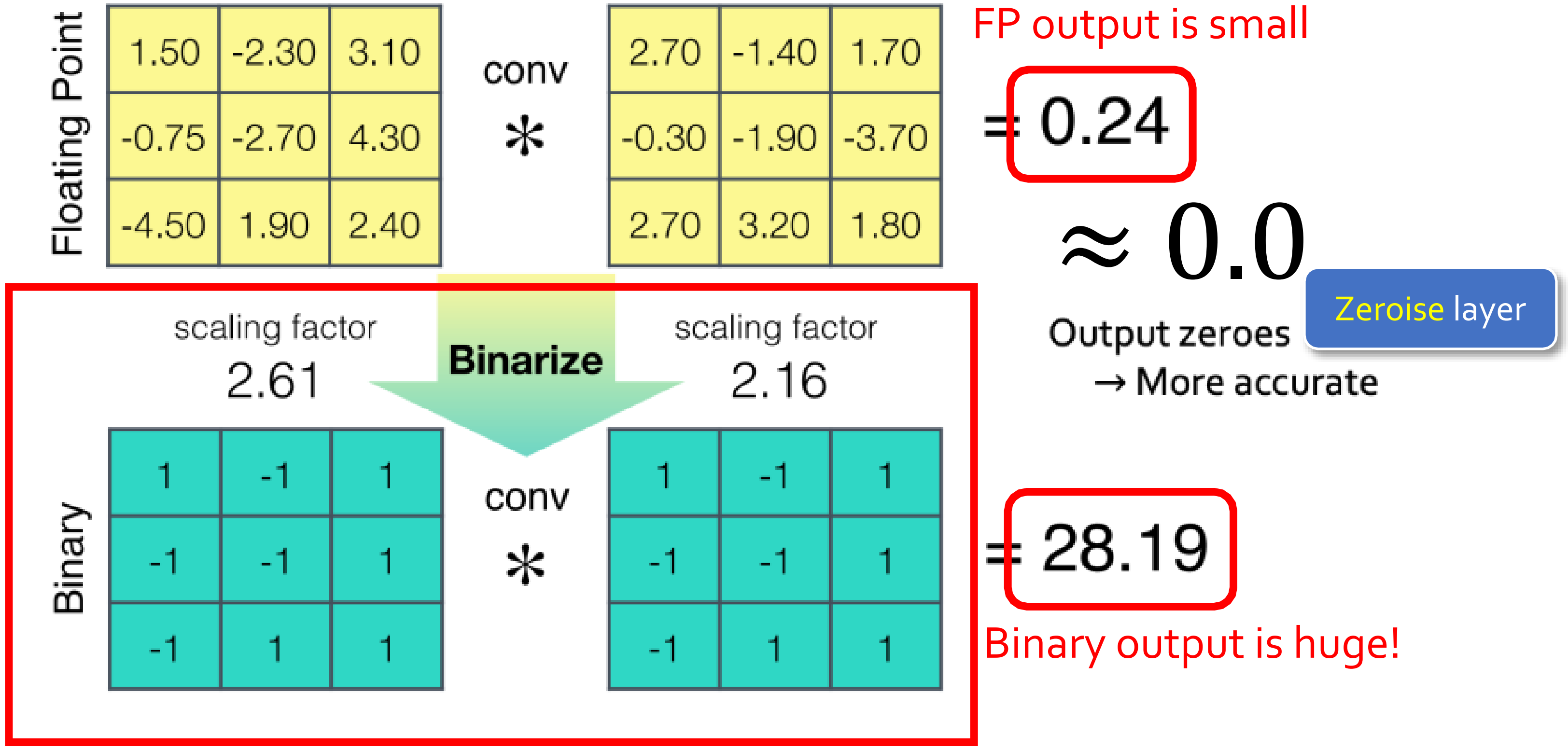
Reduce Quantization Error by **New Search Space**

- Construct binarize network with each layer type alone
  - Evaluate inference accuracy on CIFAR10: random guess is 10%
- To see which types of convolution is resilient to quantization error

Layer Type	Conv		Dil. Conv		Sep. Conv	
Kernel Size	3 × 3	5 × 5	3 × 3	5 × 5	3 × 3	5 × 5
FP Acc. (%)	61.78	60.14	56.97	55.17	56.38	57.00
Bin. Acc. (%)	46.15	42.53	41.02	37.68	10.00	10.00

Same as random guess!

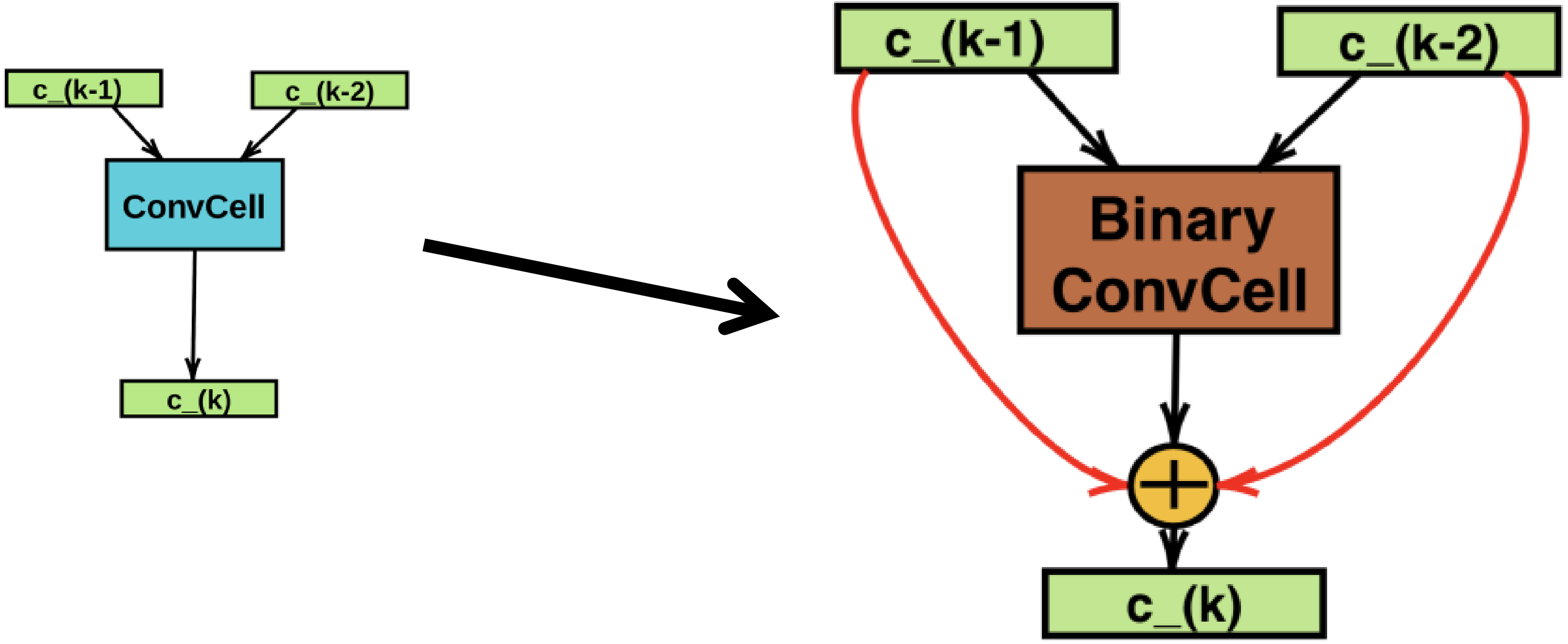
Reduce Quantization Error by **New Search Space** (cont'd)



Reduce Quantization Error by **New Cell Template**

CVPR VIRTUAL JUNE 19-25
Binary Networks for Computer Vision
CVPR 2021 Workshop

New cell template



Add residual connection from a previous cell
→ Gradient information is better preserved

GIST · Ai2

Diversify Early Search by **New Search Objective**

Searched network's learnable parameters

"Search" parameters

Annealing hyper-param.

$$\widetilde{\mathcal{L}}_S(D;\theta_{\alpha_B}) = \mathcal{L}_S(D;\theta,p) - \lambda H(p)e^{(-t/\tau)}$$

Original search objective

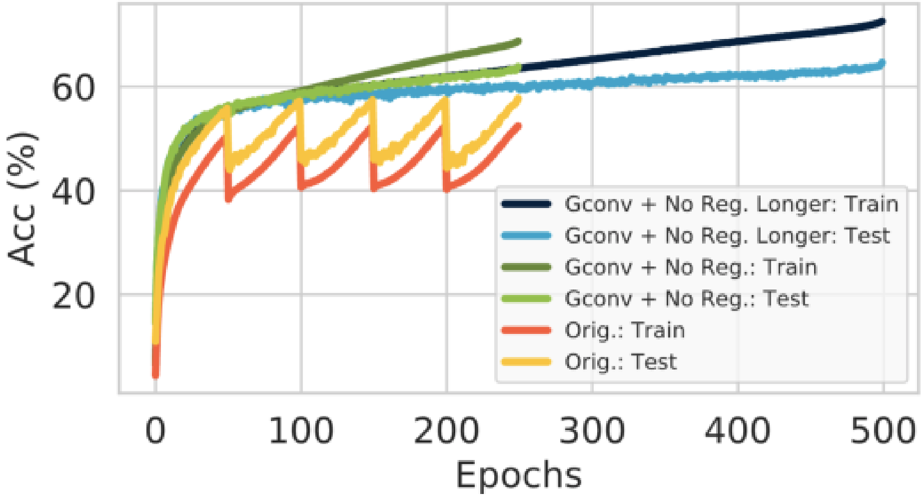
Diversity regularizer

Epoch number

- Undertrained convolutions layers < pooling layers early on
- Propose to enforce exploration in the early search phase
  - By an entropy maximizing regularizer

Empirical Improvements

- The first convolution (or stem) is kept FP
  - Use group convolutions in the stem to reduce FLOPs
- Standard training schemes may contain excess regularization
  - For binary networks, this may actually cause under-fitting
  - Minimize regularization while training



Comparison with SOTA Binary Networks

- Outperforms or performs on par with many SOTA binary nets

FLOPs ($\times 10^8$)	Method (Backbone Arch.)	Binarization Scheme	Pretraining	Top-1 Acc. (%)	Top-5 Acc. (%)
~ 1.48	BinaryNet (ResNet18) [3]	Sign	✗	42.20	67.10
	ABC-Net (ResNet18) [9]	Clip + Sign	✗	42.70	67.60
	BNAS-D	Sign + Scale	✗	57.69	79.89
	BNAS-D-No-Reg	Sign + Scale	✗	61.60	82.91
	BNAS-D v2†	Sign + Scale	✗	63.82	84.25
	BNAS-D v2 Multi-Stage†	Sign + Scale	✓	66.03	85.42
	BATS† [1]	Sign + Scale	✓	66.10	87.00
~ 1.63	Bi-Real (Bi-Real Net18) [12]	Sign + Scale	✓	56.40	79.50
	XNOR-Net++ (ResNet18) [2]	Sign + Scale*	✗	57.10	79.90
	PCNN (ResNet18) [4]	Projection	✓	57.30	80.00
	BONN (Bi-Real Net18) [5]	Bayesian	✗	59.30	81.60
	BinaryDuo (ResNet18) [7]	Decoupled	✓	60.40	82.30
~ 1.78	ABC-Net (ResNet34) [9]	Clip + Scale	✗	52.40	76.50
~ 1.93	Bi-Real (Bi-Real Net34) [12]	Sign + Scale	✓	62.20	83.90
~ 6.56	CBCN (Bi-Real Net18) [10]	Sign + Scale	✓	61.40	82.80

Thank You

https://github.com/gistvision/bnas