HONGIK UNIVERSITY

2025/01/07

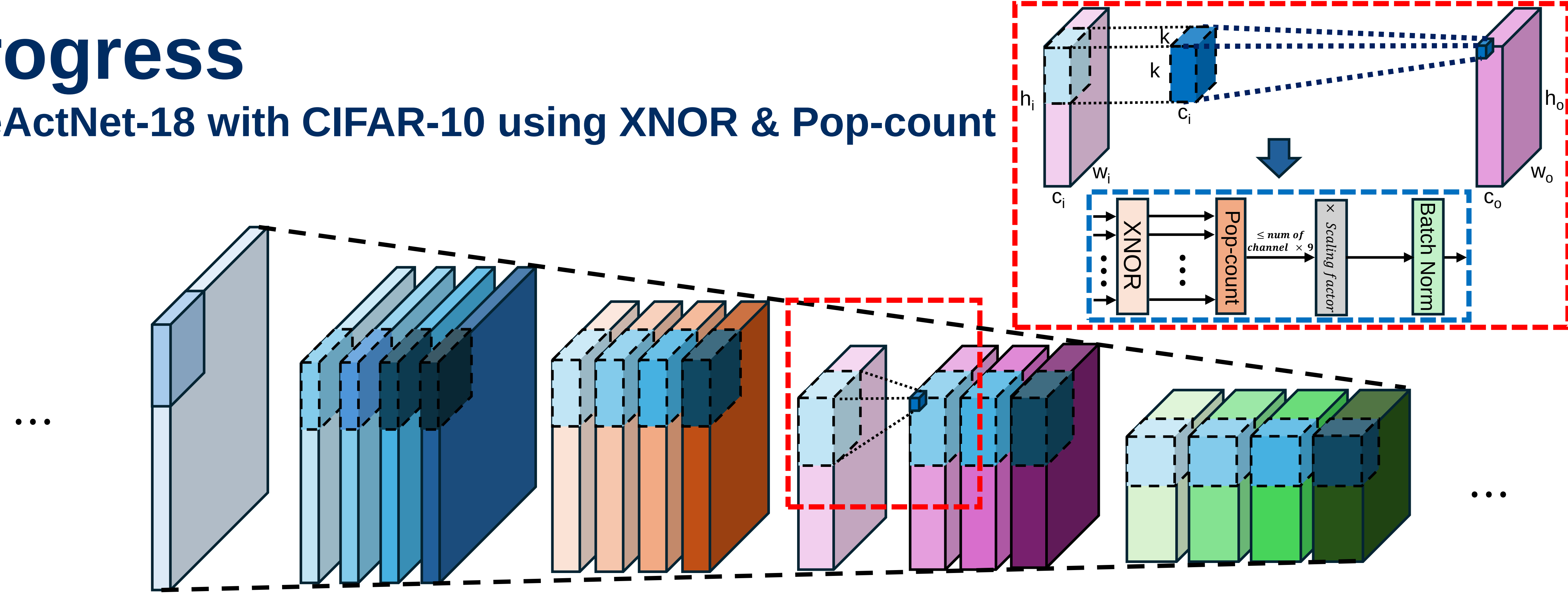
Bin-Count Network

Hyungdong Park, Inguk Yeo
Department of Computer Engineering

Contents

Progress

- ReActNet-18 with CIFAR-10 using XNOR & Pop-count

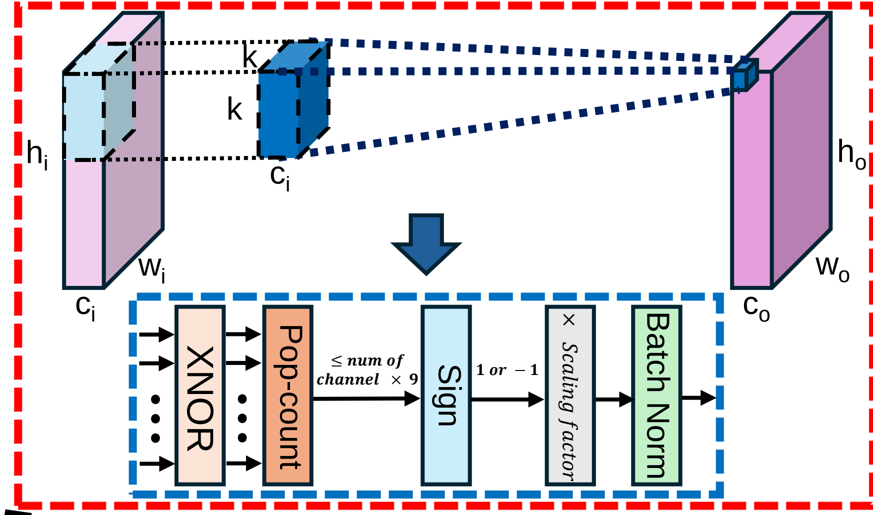


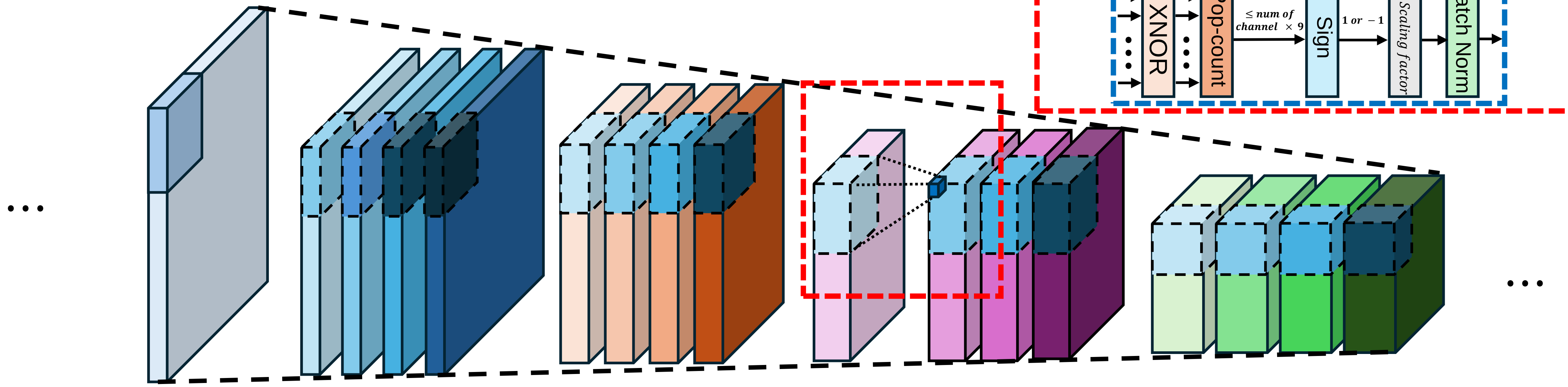
Num of channels	64	64	128	256	512	Pooling & FC
Num of layers	1	4	4	4	4	
Image size	64 × 32 × 32	64 × 32 × 32	128 × 16 × 16	256 × 8 × 8	512 × 4 × 4	
Operations	⊛	XNOR & Pop-Count & Multiplication & Batch Norm				
Activations and weights	ℝ	ℝ (After BN) & Binarized values (1 or -1)				ℝ
# units of Mul		c_o x h_o x w_o				

Implementing binarized counting on baseline model

Progress

- Software Implementation of Bin-Count Network with CIFAR-10

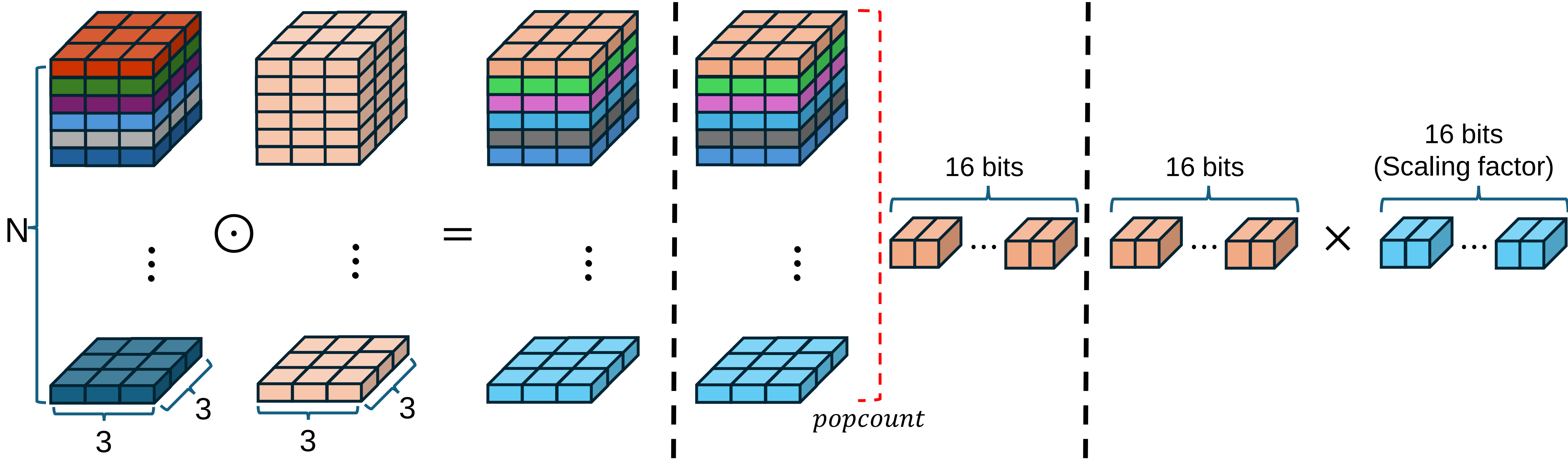




Num of channels	64	64	128	256	512	Pooling & FC
Num of layers	1	4	4	4	4	
Image size	$64 \times 32 \times 32$	$64 \times 32 \times 32$	$128 \times 16 \times 16$	$256 \times 8 \times 8$	$512 \times 4 \times 4$	
Operations	$\circledast$	XNOR & Pop-Count & Multiplication & Batch Norm				
Activations and weights	$\mathbb{R}$	$\mathbb{R}$ (After BN) & Binarized values (1 or -1)				$\mathbb{R}$
# units of Mul		None				

Progress

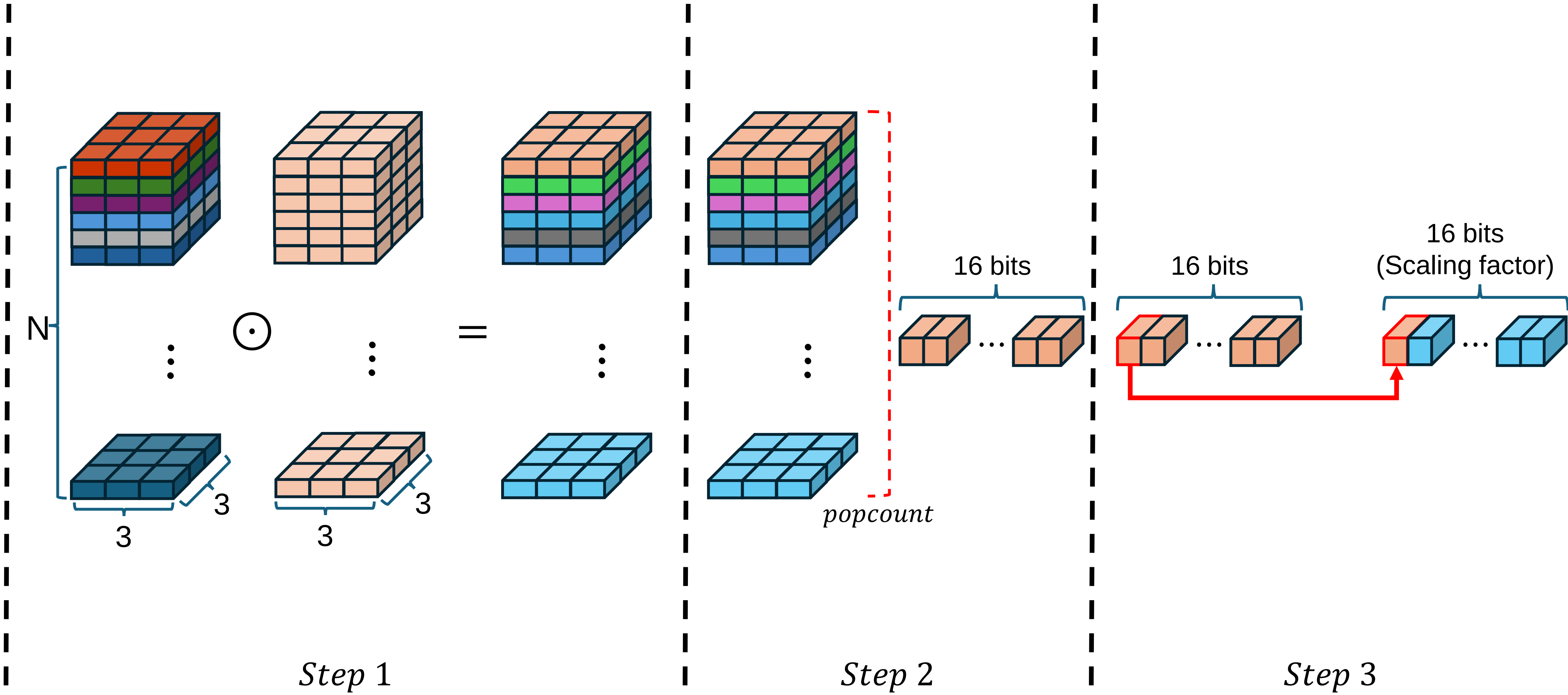
- Hardware computations per a xnor-popcount for ReActNet-18 (our baseline)



	Step 1	*Step* 2	*Step* 3
Operations	*Xnor* (⊙)	*Popcount*	*Integer multiplication* & *Bit shift*
# operations	$N \times 3^2$	$(N \times 3^2) - 1$	1 *Integer multiplication* & 1 *Bit shift*

Progress

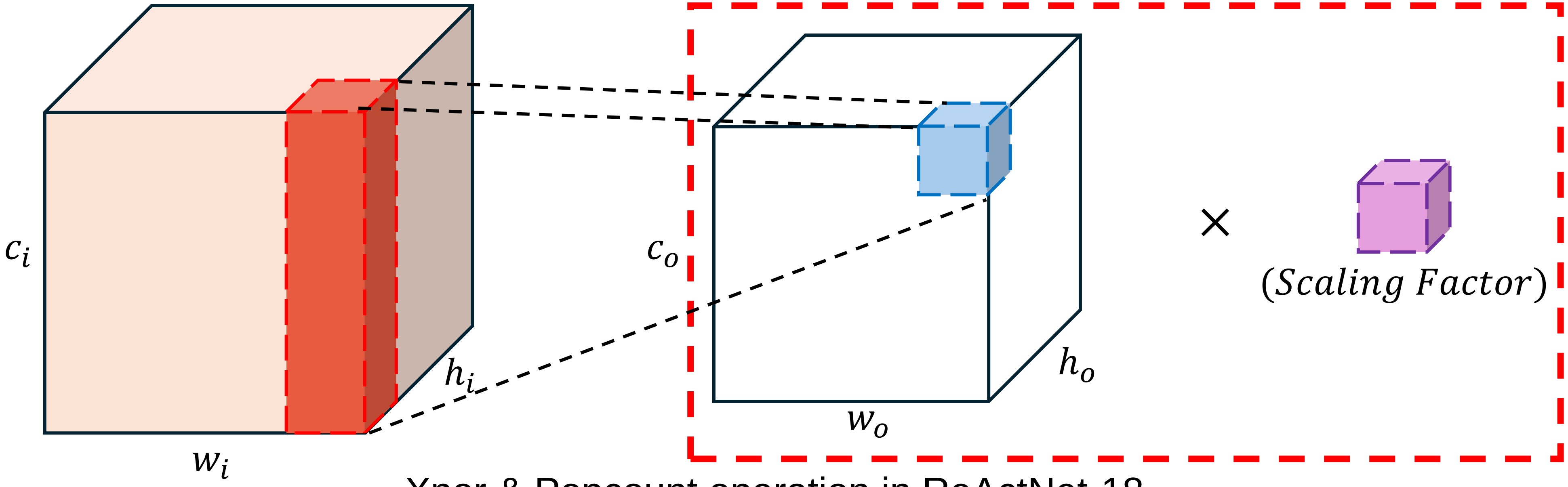
- Hardware computations per a xnor-popcount for our model



	Step 1	Step 2	Step 3
Operations	$Xnor\ (\odot)$	$Popcount$	$None$
# operations	$N \times 3^2$	$(N \times 3^2) - 1$	$None$

Progress

- Hardware computations per a layer

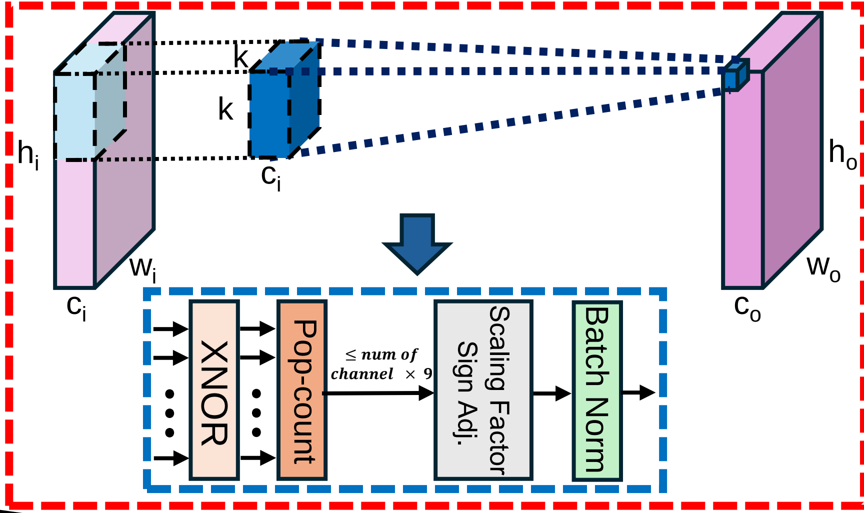


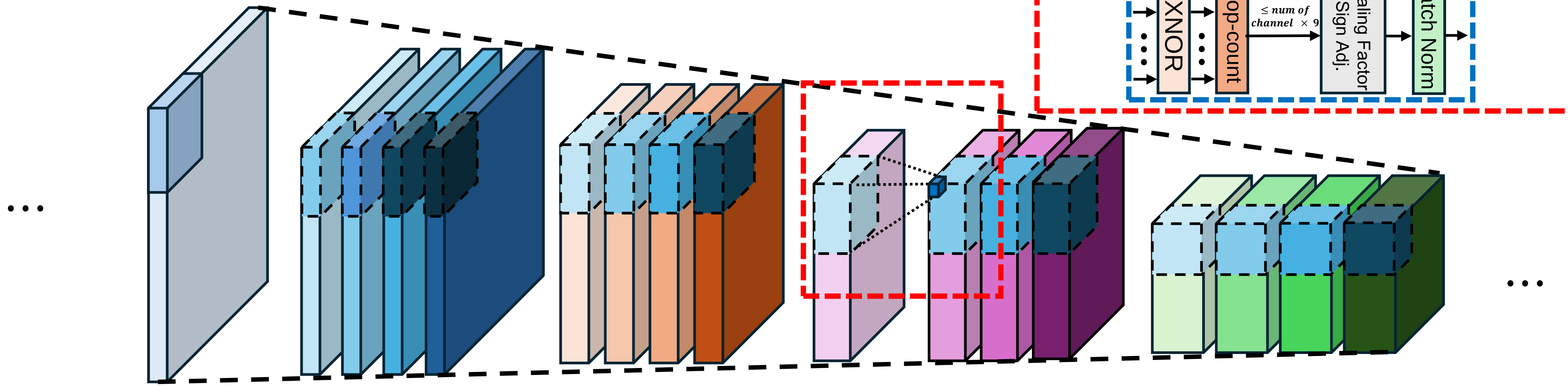
Xnor & Popcount operation in ReActNet-18

Models	Operations	# Operations per a layer
ReActNet-18	Integer Multiplication & Bit shift	$c_o \times w_o \times h_o$ $Integer\ Multiplications\ \&\ Bit\ shifts$
QAT-binarized counting ReActNet-18	None	None

Progress

- Hardware computations of model-wide perspective





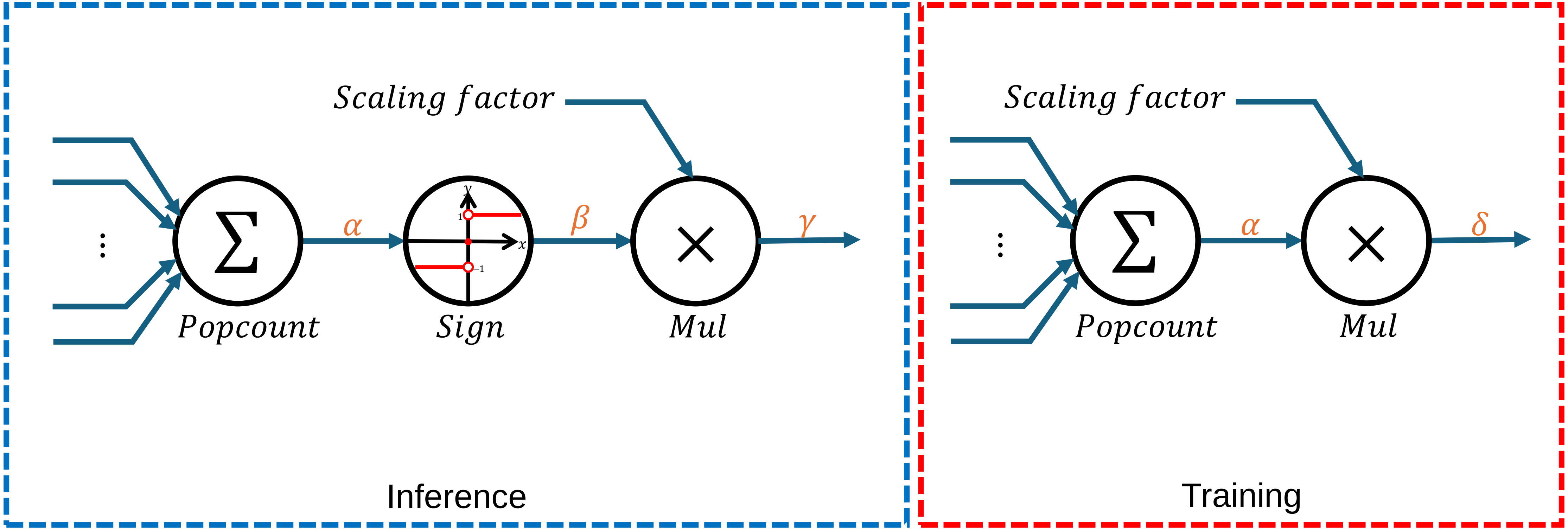
QAT-binarized counting ReActNet-18

Models	Operations	# Operations
ReActNet-18	Integer Multiplication & Bit shift	557,056 Integer Multiplications & Bit shifts
QAT-binarized counting ReActNet-18	None	None

Binarized counting techniques

Progress

- Structure for the PTQ-binarized counting



$|\alpha| \leq (channel\_num \times kernel^2)$

$\beta = \pm 1$

$\gamma = \pm scaling\_factor$

$\delta = \pm (scaling\_factor \times channel\_num \times kernel^2)$

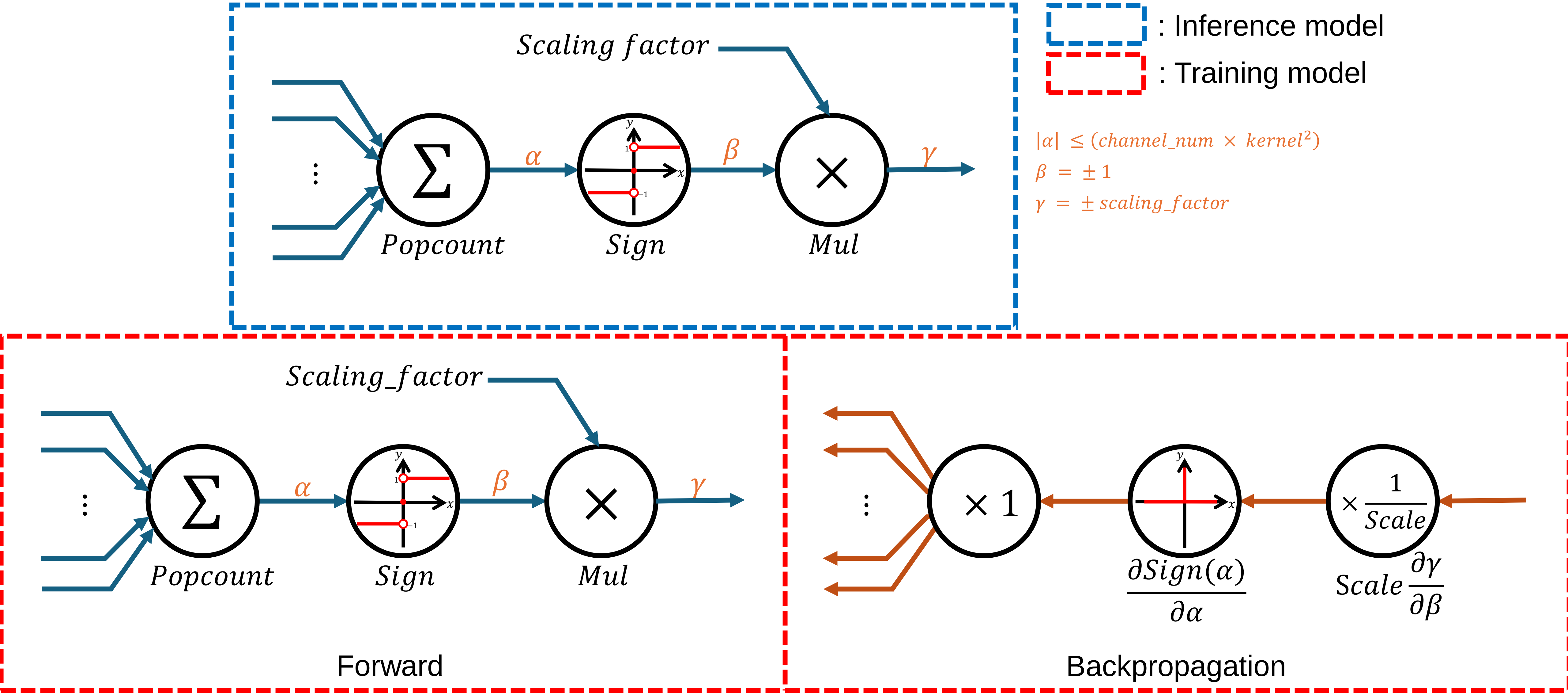
Progress

- PTQ-binarized counting's results with CIFAR-10

Models	Top-1 Accuracy (%)	Top-5 Accuracy (%)
ReActNet-18	93.380	99.800
PTQ-binarized counting ReActNet-18	10.000	52.040
Bi-RealNet-18	88.770	98.250
PTQ-binarized counting Bi-RealNet-18	10.000	50.000

Progress

- Structure for the PTQ-binarized counting



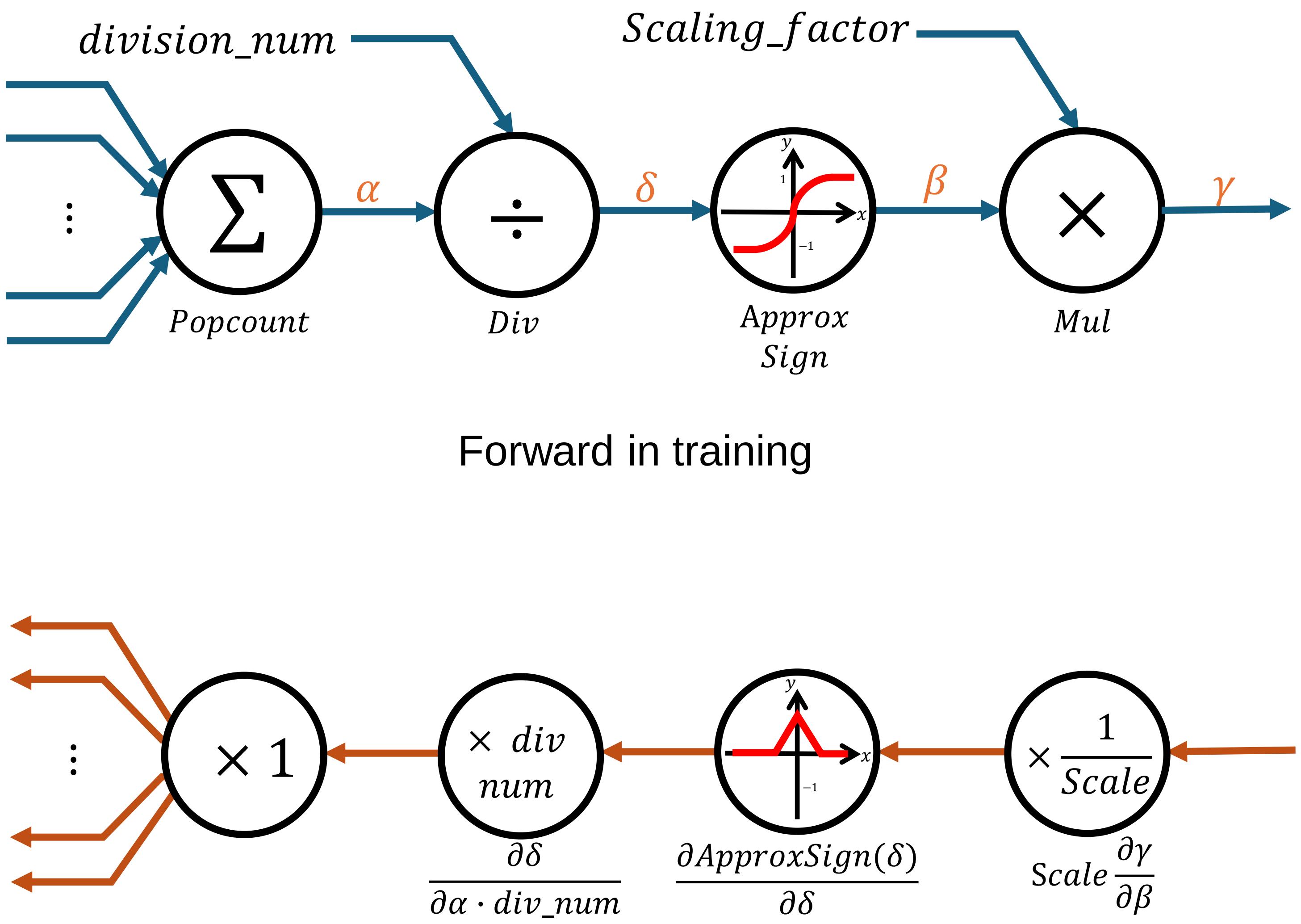
Progress

- Simple QAT-binarized counting's results with CIFAR-10

Models	Top-1 Accuracy (%)	Top-5 Accuracy (%)
ReActNet-18	93.380	99.800
Simple QAT-binarized counting ReActNet-18	84.930	99.250
Bi-RealNet-18	88.770	98.250
Simple QAT-binarized counting Bi-RealNet-18	30.070	79.690

Progress

- Structure for the PTQ-binarized counting



Forward in training

Backpropagation in training

$|\alpha| \le (channel\_num \times kernel^2)$

$\beta = \pm 1$

$\gamma = \pm\, scaling\_factor$

$\delta = \pm\, (channel\_num \times kernel^2 \div division\_num)$

Progress

- QAT-binarized counting ReActNet-18's results with CIFAR-10 along with division num

Division num	Top-1 Accuracy (%)	Top-5 Accuracy (%)
$channel\ num + \alpha$	92.150	99.640
$(channel\ num \times kernel^2) + \alpha$	89.580	99.460
$channel\ num \times \alpha$	92.510	99.640
$(channel\ num \times kernel^2) \times \alpha$	92.160	99.660
Min-Max Normalization $(channel\ num \times kernel^2)$	89.230	99.390

Progress

- QAT-binarized counting's results with CIFAR-10

Models	Top-1 Accuracy (%)	Top-5 Accuracy (%)
ReActNet-18	93.380	99.800
Simple QAT-binarized counting ResNet-18	84.930	99.250
QAT-binarized counting ReActNet-18	92.510	99.640
Bi-RealNet-18	88.770	98.250
Simple QAT-binarized counting Bi-RealNet-18	30.070	79.690
QAT-binarized counting Bi-RealNet-18	87.660	98.720

Thank you