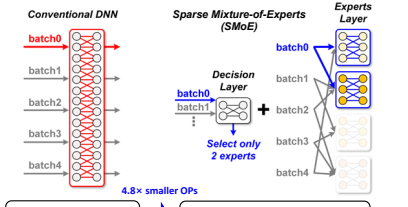
Space-Mate: A 303.5mW Real-Time NeRF SLAM Processor with Sparse-Mixture-of-Experts-based Acceleration

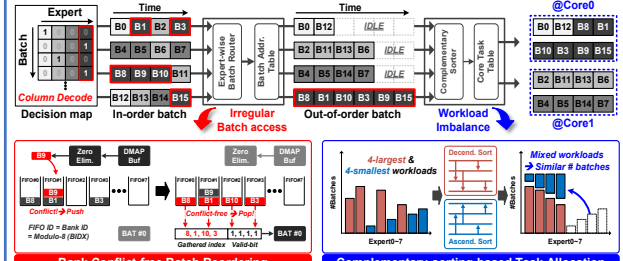
Gwangtae Park, Seokchan Song, Haoyang Sang, Dongseok Im, Donghyeon Han, Sangyeob Kim, Hongseok Lee, and Hoi-Jun Yoo

Motivation



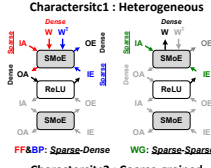
- NeRF-based SLAM for robotic applications face computation barrier
- Accuracy depends on DNN capacity (= # Params)
- Multiple expert networks to keep high accuracy
- Conditional computing for small computation cost

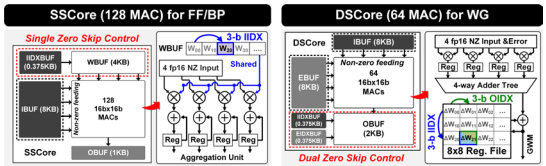
Proposed Sparse Mixture-of-Experts Acceleration



- Expert-stationary dataflow on each core to remove redundant expert parameter fetch
- Challenge1: Irregular batch access
- Challenge2: expert workload imbalance
- 6.5× speed-up compared to the non-expert-stationary dataflow

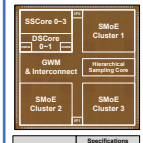
Proposed Heterogeneous Coarse-grained Sparse Core





- Heterogeneous core to accelerate different DNN training stages
- Coarse-grained zero-skipping controller to minimize logic overhead
- 32.4~55.6% higher energy efficiency than homogeneous core

Implementation Result



Specifications	
Technology	Samsung 28 nm
Die Area	4.5 mm × 4.5 mm
SRAM [KB]	1348 KB
Supply Voltage	0.7 ~ 0.9V
Max. Frequency	200 MHz
Data Type	DNN: FP16
	3D Embed.: FP8
Sensor Type	VGA, RGB-D
Frame Rate [FPS]	13.1 ~ 52.6
Tracking / Mapping Performance	
Operating Condition	0.7V 50MHz / 0.75V 100MHz / 0.9V 200MHz
Power [mW]	191.7 / 118.4 / 226 / 283.1 / 570.2 / 655.7
Energy /Frame [mJ/Frame]	2.44 / 6.34 / 2.71 / 7.03 / 3.42 / 8.76

Mobile Robot NeRF-SLAM System w/ Space-Mate

SLAM Result & Comparison w/ Previous SLAM Processor