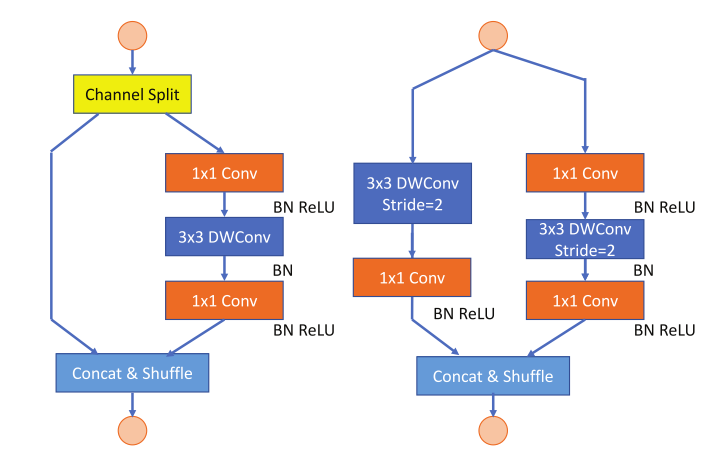
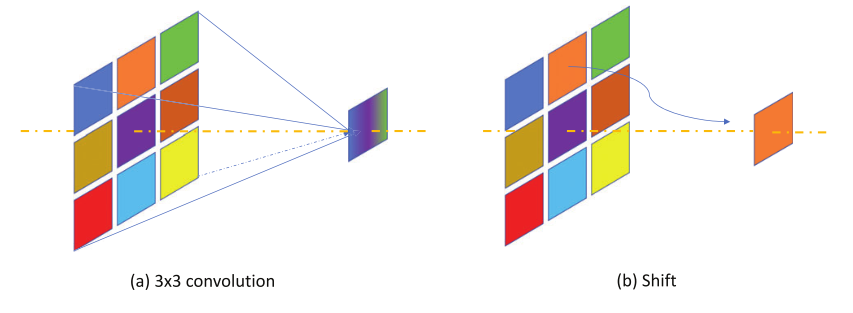
**Synetgy: Algorithm-hardware Co-design for ConvNet Accelerators on Embedded FPGAs**

1. Contributions
   1. A novel CNN model named DiracDeltaNet which achieves 89% top 5 accuracy on ImageNet with 48x fewer parameters and 65x operations.
   2. A CNN accelerator named Synetgy for DiracDeltaNet on an Ultra96 board through HLS, which achieves an inference speed of 96.5 FPS on the ImageNet classification task.
2. DiracDeltaNet(based on ShuffleNetV2)
   1. ShuffleNetV2



ShuffleNetV2 basic block and down-sampling block

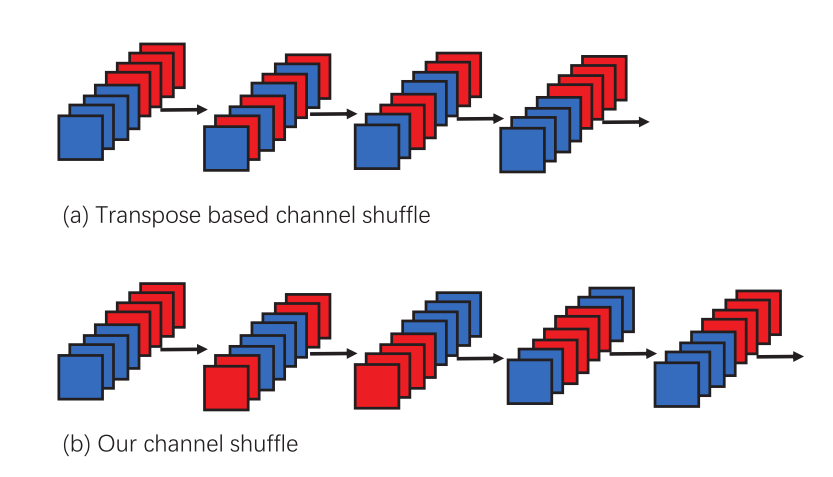
* 1. Replace 3x3 depth-wise convolution by shift operation.



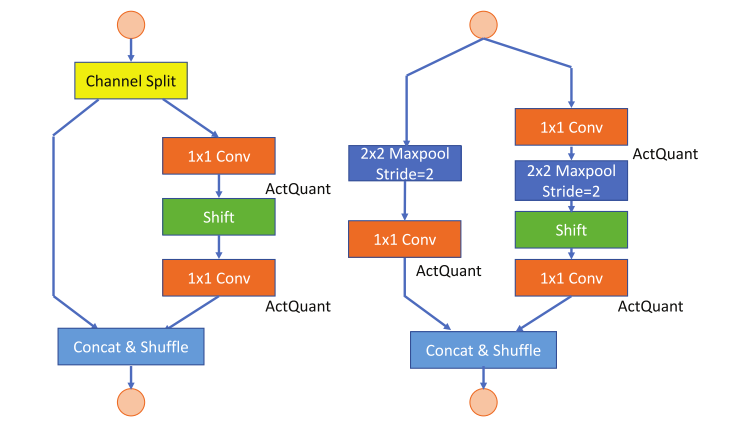
3x3 convolution vs shift operation

* 1. Double the output filter number of the first 1×1 convolution on the non-skip branch.
  2. Modify the channel shuffle’s order to make it more

hardware efficient.

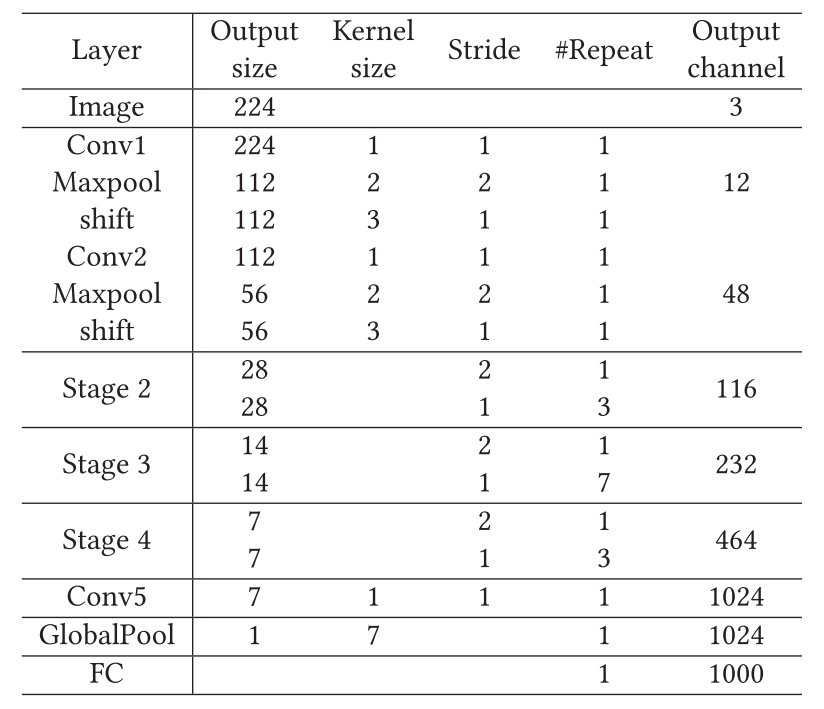


* 1. Modified ShuffleNetV2:

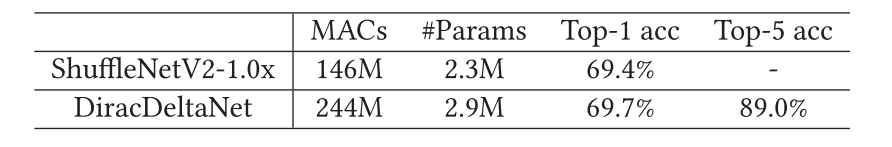


Basic blocks

* 1. Quantization
  2. Macro-structure of DiracDeltaNet

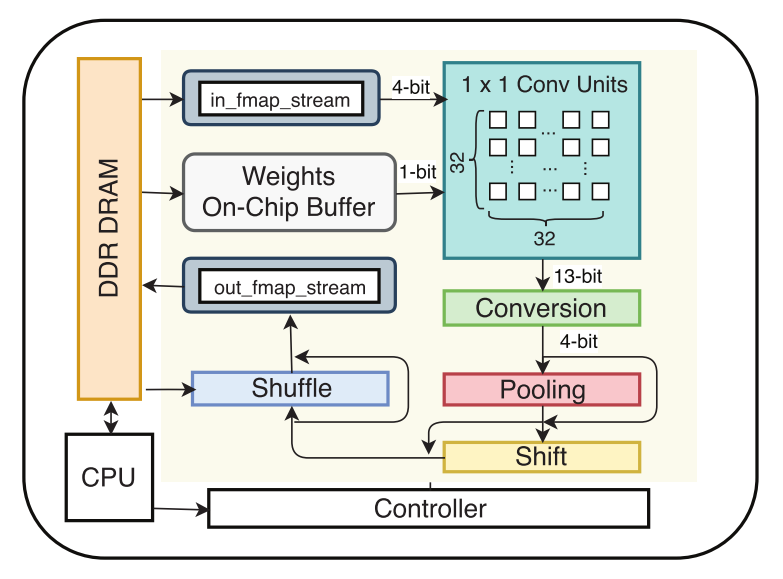


Structure

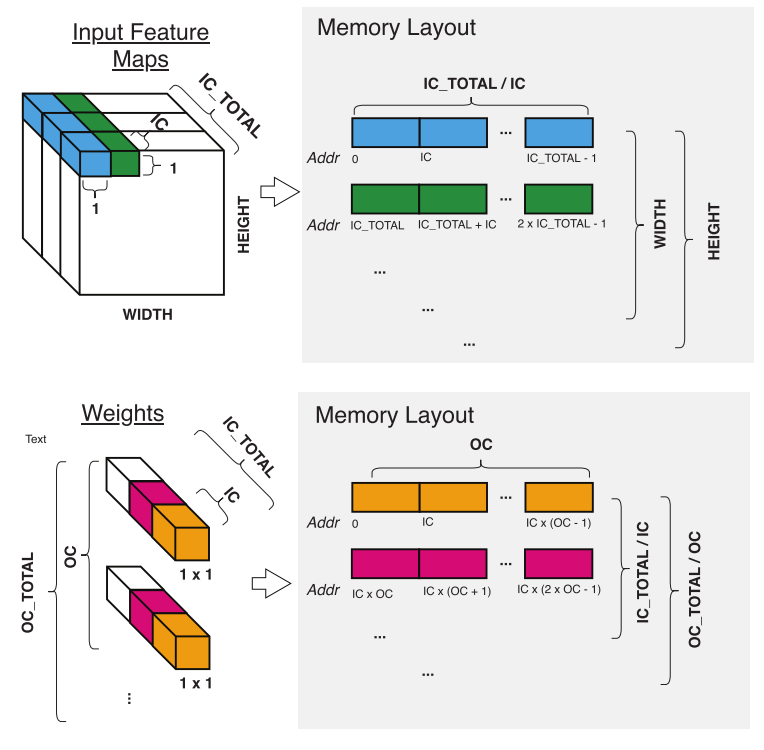


Comparison

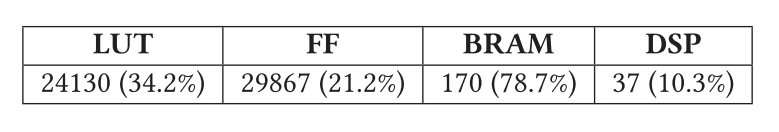
1. Synetgy
   1. Overview



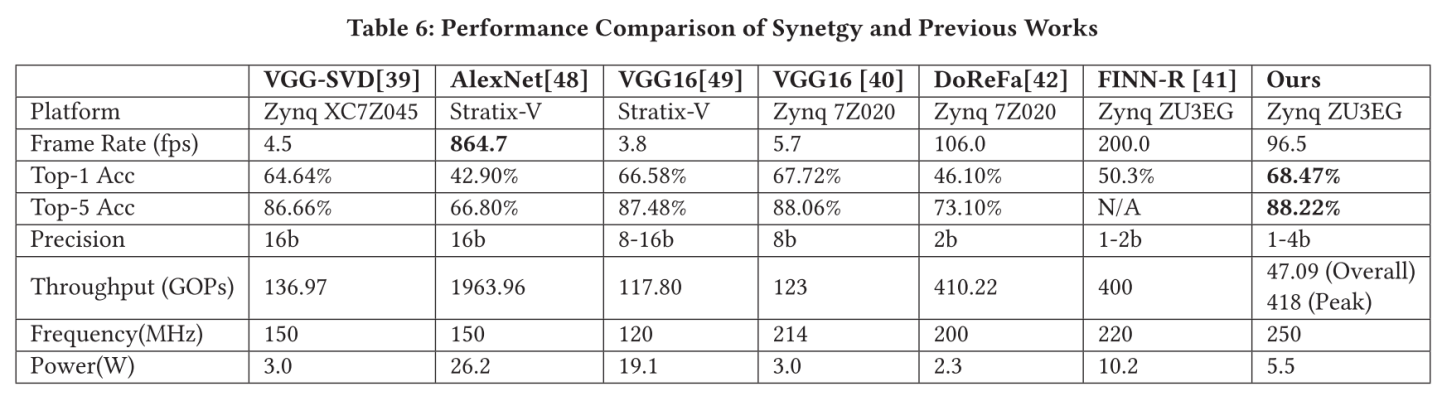
* 1. Input Layout in DRAM



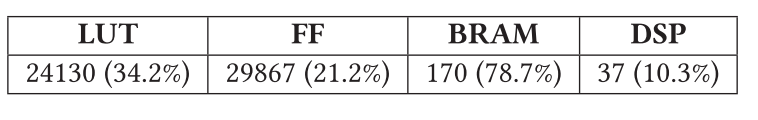
* 1. Resource usage



1. Results



Comparison



Resource Utilization