

Interleaved Convolution in Eyeriss V1 PE Array

Setup

Consider a scenario with an ifmap row of size 5 and two filter rows of size 3 each. This setup will illustrate how two filter rows can be interleaved and then convolved with the same ifmap row using the Eyeriss V1's row-stationary dataflow architecture.

Ifmap row of size 5: $[x_0, x_1, x_2, x_3, x_4]$

Two filter rows of size 3 each:

Filter 1 row: $[w_{10}, w_{11}, w_{12}]$

Filter 2 row: $[w_{20}, w_{21}, w_{22}]$

Step-by-Step Explanation

1. Convolution Step by Step

The PE performs the convolution by sliding the filter across the ifmap row, performing multiply-accumulate operations for each position. This is done simultaneously for both filters using the interleaved weights.

Outputs are generated as follows:

Position 0 using $[x_0, x_1, x_2]$:

- $Psum1_0 = x_0 * w_{10} + x_1 * w_{11} + x_2 * w_{12}$
- $Psum2_0 = x_0 * w_{20} + x_1 * w_{21} + x_2 * w_{22}$

Position 1 using $[x_1, x_2, x_3]$:

- $Psum1_1 = x_1 * w_{10} + x_2 * w_{11} + x_3 * w_{12}$
- $Psum2_1 = x_1 * w_{20} + x_2 * w_{21} + x_3 * w_{22}$

Position 2 using $[x_2, x_3, x_4]$:

- $Psum1_2 = x_2 * w_{10} + x_3 * w_{11} + x_4 * w_{12}$
- $Psum2_2 = x_2 * w_{20} + x_3 * w_{21} + x_4 * w_{22}$