# Task 1E: Confusion hi confusion hai !!

## Problem Description

This task involved implementing the previously discussed optimizations for a neural network to evaluate the performance gain of such optimizations on a real application. Both matrix multiplication and transpose functions were optimized for this task.

## Performance Comparison

### Matrix Multiplication

The matrix multiplication operation was optimized using unrolling, reordering, tiling and SIMD functions. The speedup comparisons between these techniques is shown in 1

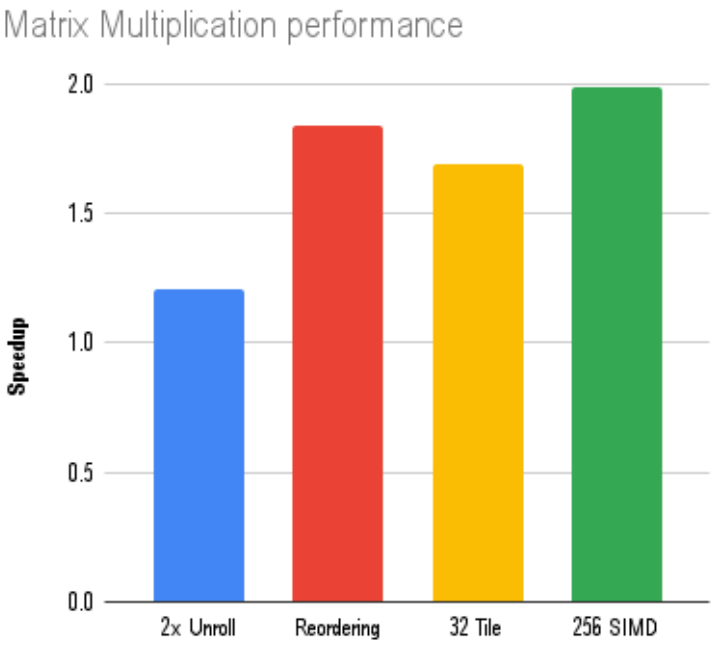


Figure 1: Performance comparison of optimization techniques for Matrix Multiplication

We notice that maximum speedup is achieved by using SIMD functions which perform vector operations. A significant speedup is also achieved by simply reordering the loops for better reuse.

Speedups for different unrolling lengths can be seen in 2. There is a continuously decreasing trend suggesting that the performance gained by unrolling is quickly nullified by the increase in the number of instructions for longer unrolls.

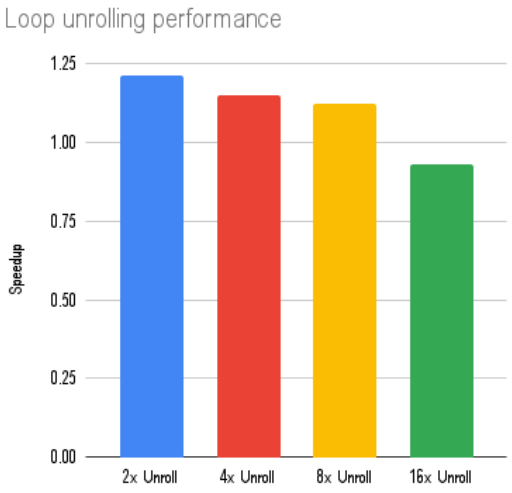


Figure 2: Performance comparison of unrolling lengths for Matrix Multiplication

Speedups for different tile sizes can be seen in 3. The highest speedup is achieved at tile size of 32, which indicates that the locality and reuse of cache blocks is most effective for this tile size.

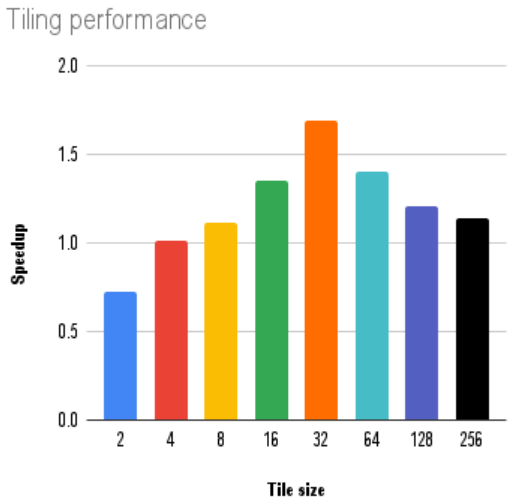


Figure 3: Speedup comparison of different tile sizes for Matrix Multiplication

### Transpose

The transpose option was optimized by implementing tiling. SIMD was also tried but resulted in negligible improvement in performance. With tiling, a reduction of 5-10 ms in execution time was found across all runs. The performance comparison can be seen in 4.

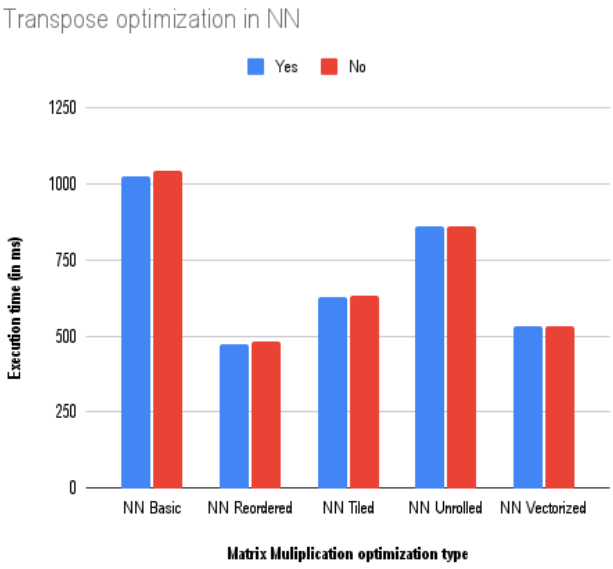


Figure 4: Transpose optimize comparison