

Expansive Networks: Linear Parameter Injections At Train-Time

Dmitri Tkatch

June 29, 2025

Abstract

While mathematically equivalent and contrary to modern wisdom, two consecutive matrices of sizes 100×10000 and 10000×100 with no non-linearity inbetween them actually learn more easily than one 100×100 matrix. We call the act of replacing the 100×100 matrix by the former two matrices "expansion". We propose Expansive Networks – a family of neural networks that are trained with expansions which are then collapsed at test-time via matrix multiplication. Only one layer is expanded at a time for computational memory savings. In practice we find it is more fruitful to expand depth-wise rather than width-wise. On MNIST, expanding a layer to 20 layers results in 27 more correct classifications (a +0.27% improvement in accuracy).