

# Transformer

Axel Pontén

May 14, 2025

Code is found at <https://github.com/axel-ponten/advanced-deep-learning-VT25/tree/main/transformer>

I implemented a transformer with 8 heads, 5 layers, comp dim 128 and feedforward dim 512. I used early stopping with patience 20 and learning rate of 1e-3, and all predictions are done with the model of the lowest val loss.

The total number of trainable parameters is roughly 1,000,000. The network converged quickly and the predictions are good.