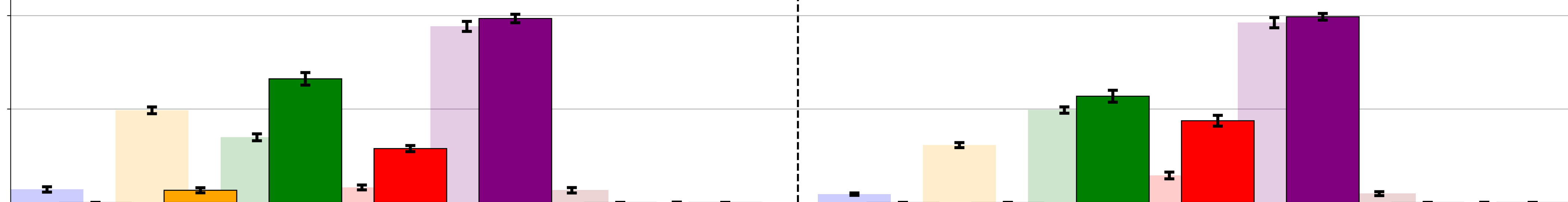
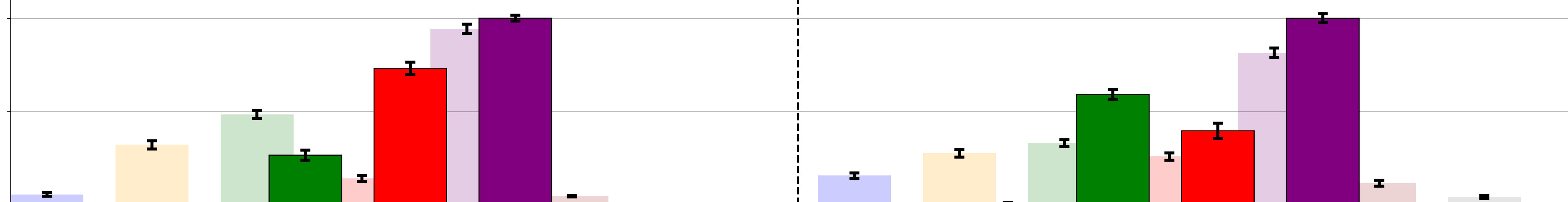


WikiText Norm Att

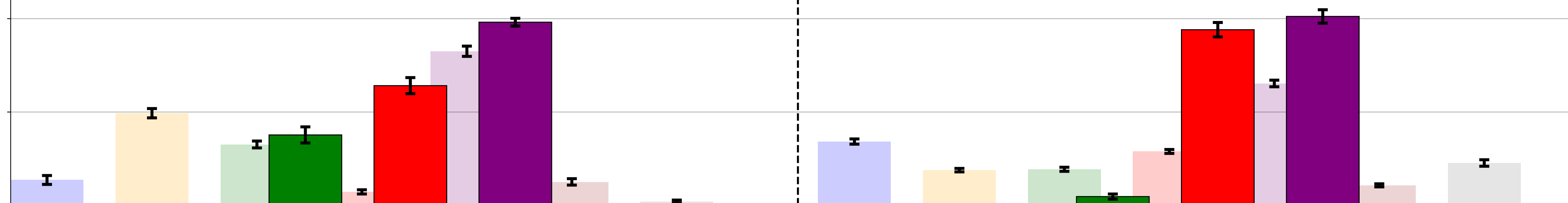
layer 1



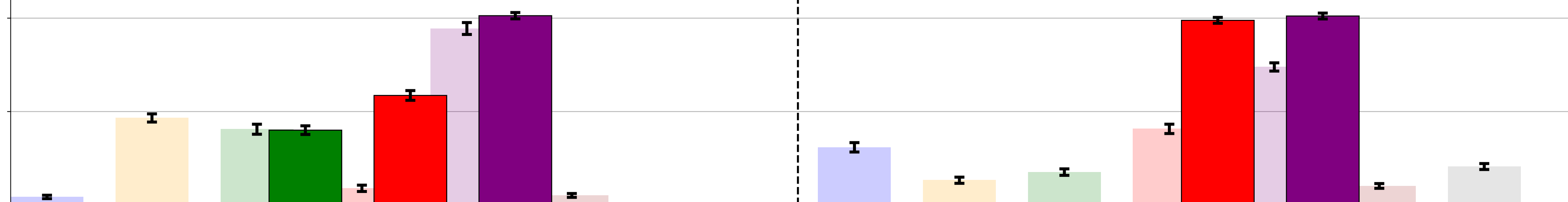
layer 2



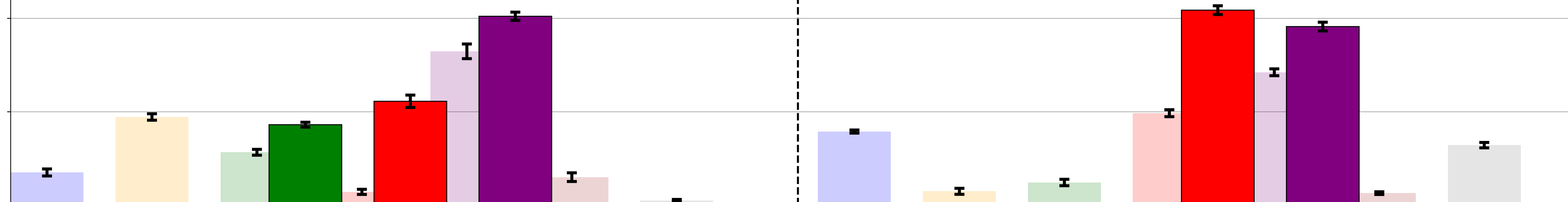
layer 3



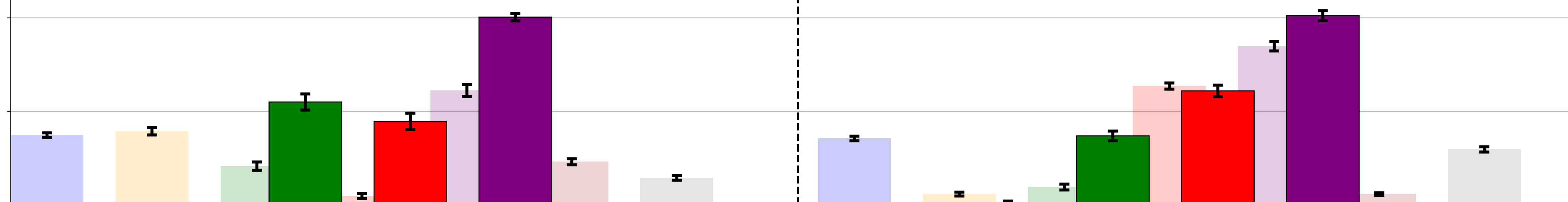
layer 4



layer 5



layer 6



with conv
(42.6)

with gating
(47.9)