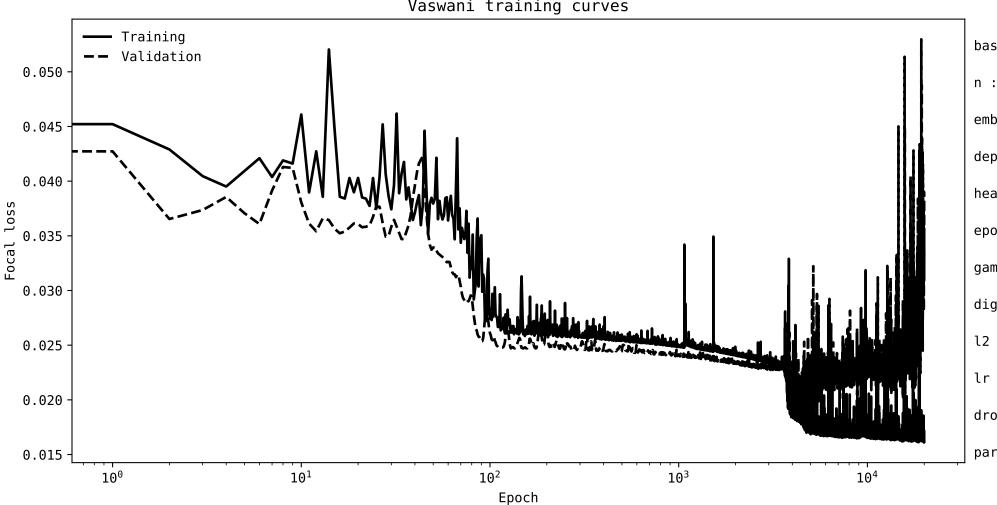
Vaswani training curves



base : 16

n: 16384

emb : 64

depth : 2

heads : 8

epochs: 2000

gamma : 2

digits : 4

l2 : 1e-05

lr : 0.001

dropout : 0.5

params: 1008