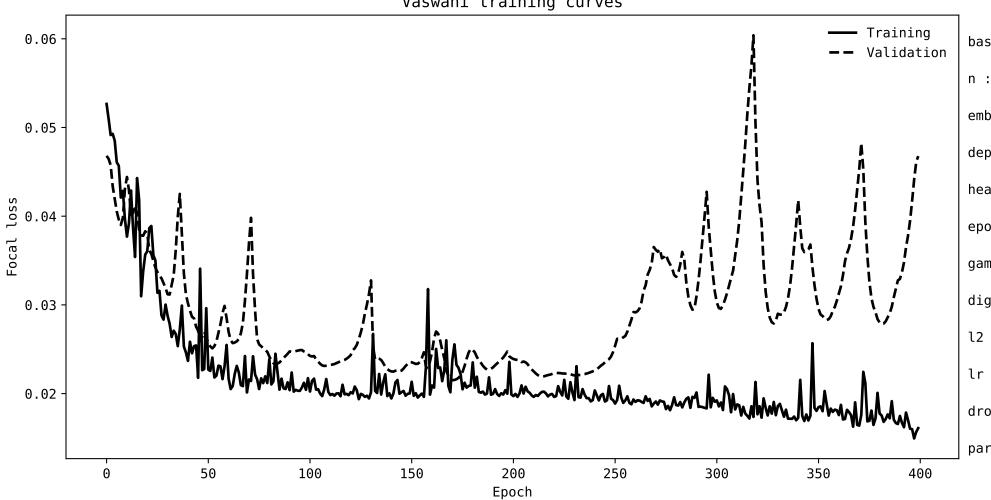
Vaswani training curves



base : 12

n : 1024

emb : 32

depth : 3

heads : 4

epochs: 400

gamma : 2

digits : 3

l2 : 1e-05

lr : 0.001

dropout : 0.1

params: 3824