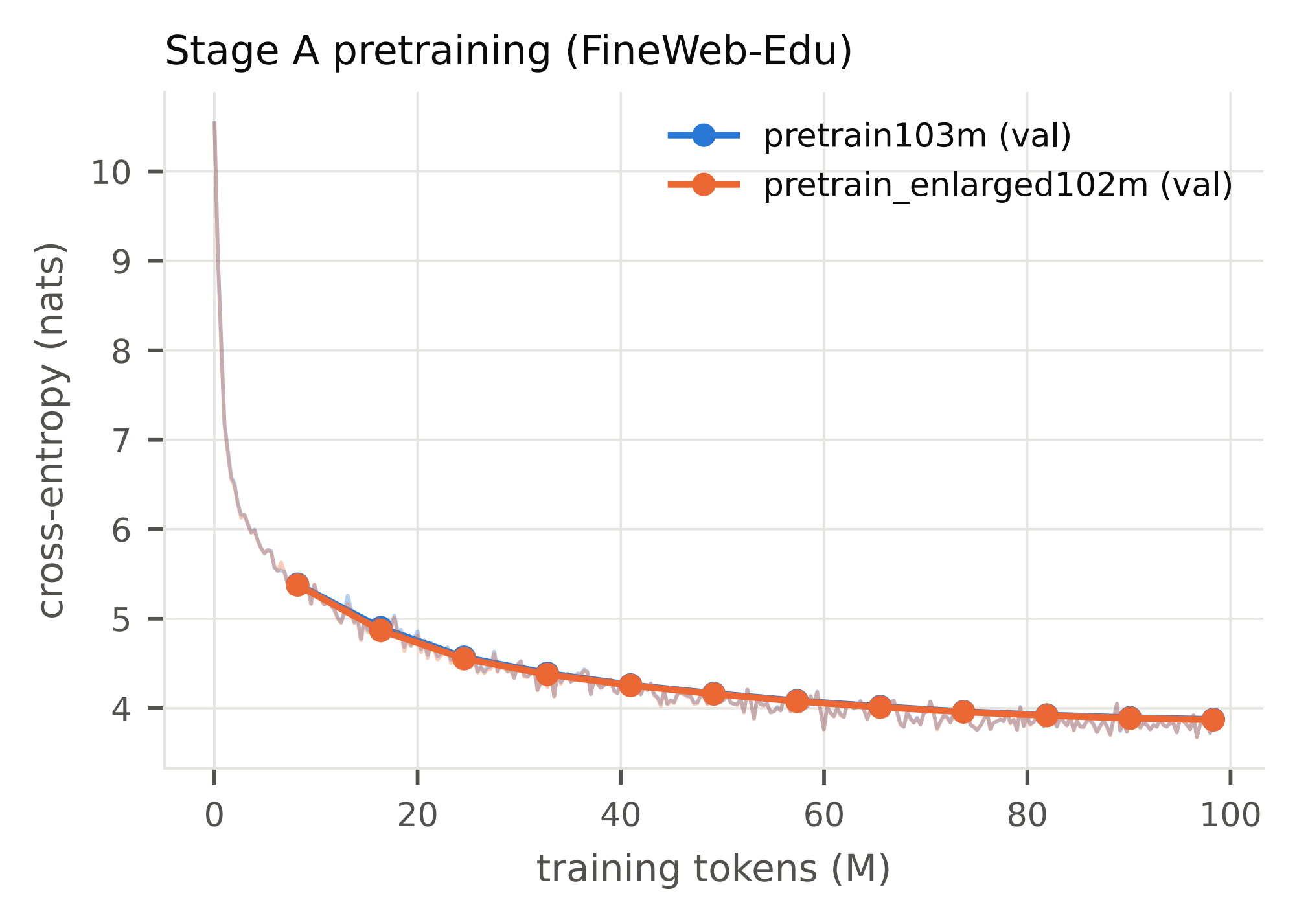

Stage A pretraining (FineWeb-Edu)
pretrain103m (val)
pretrain_enlarged102m (val)
cross-entropy (nats)
10
9
8
7
6
5
4
0
20
40
60
80
100
training tokens (M)