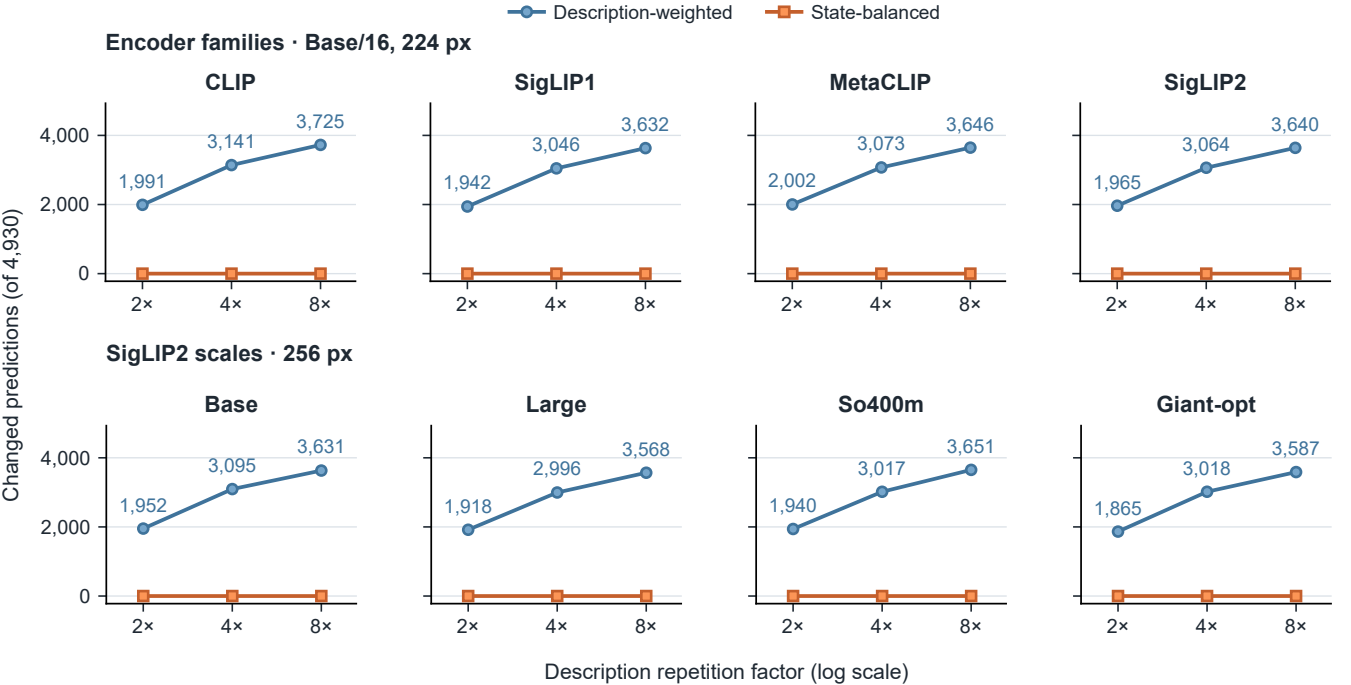

Description-weighted
State-balanced
Encoder families · Base/16, 224 px
CLIP
SigLIP1
MetaCLIP
SigLIP2
SigLIP2 scales · 256 px
Base
Large
So400m
Giant-opt
1,991
3,141
3,725
1,942
3,046
3,632
2,002
3,073
3,646
1,965
3,064
3,640
1,952
3,095
3,631
1,918
2,996
3,568
1,940
3,017
3,651
1,865
3,018
3,587
4,000
2,000
0
2×
4×
8×
Changed predictions (of 4,930)
Description repetition factor (log scale)