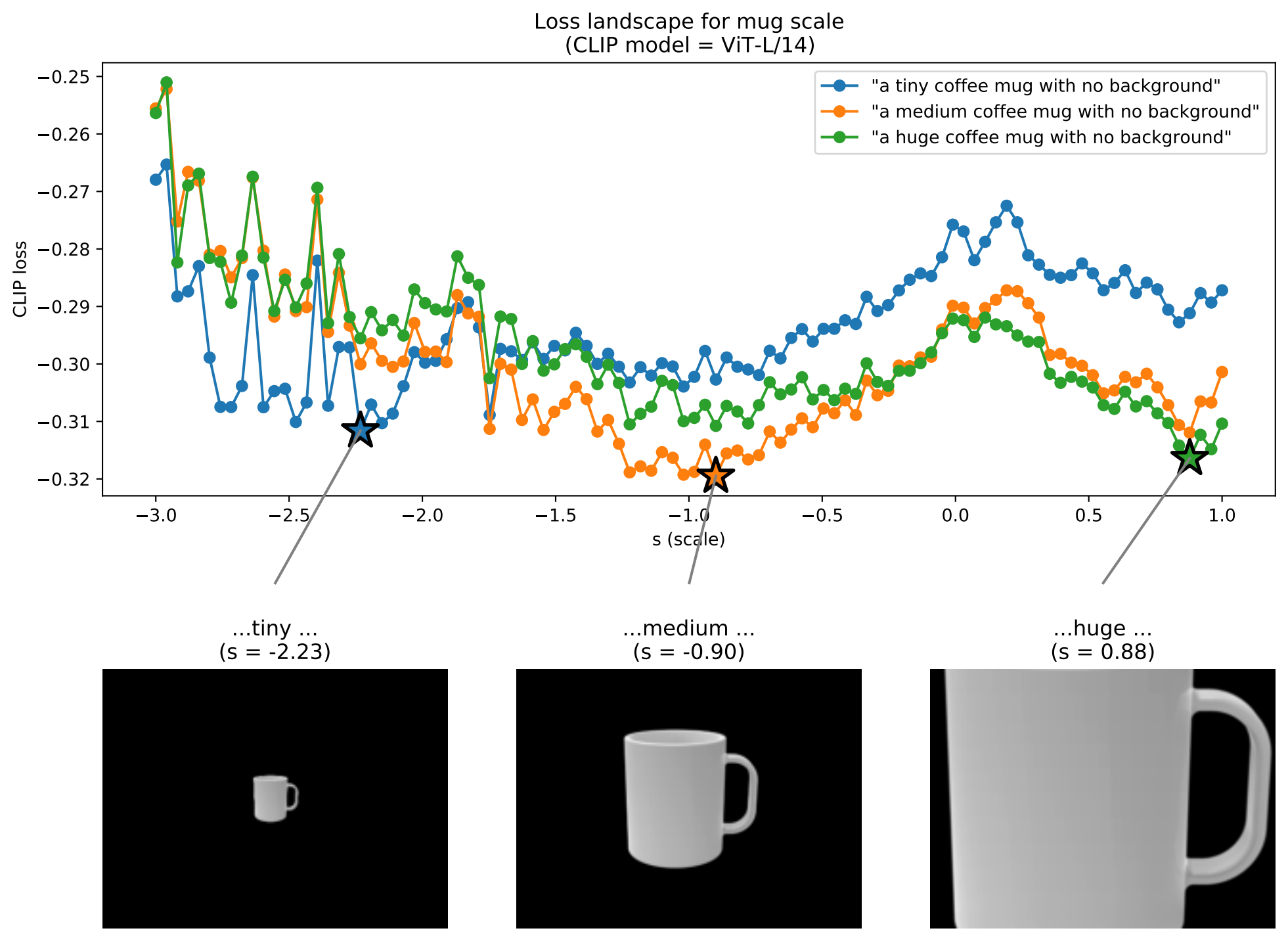
Loss landscape for mug scale
(CLIP model = ViT-L/14)
"a tiny coffee mug with no background"
"a medium coffee mug with no background"
"a huge coffee mug with no background"
CLIP loss
−0.25
−0.26
−0.27
−0.28
−0.29
−0.30
−0.31
−0.32
−3.0
−2.5
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
s (scale)
...tiny ...
(s = -2.23)
...medium ...
(s = -0.90)
...huge ...
(s = 0.88)