Augmented gradient descent (LR = 0.200, σ=[None, 3, 7, 11])
prompt = "a coffee mug on the floor of an office"
(CLIP version = ViT-L/14)

initial s = -1.01
(s = 0.09, loss = -0.24)

initial s = -1.00
(s = 0.14, loss = -0.24)

Gradient values $\frac{\partial L}{\partial s}$

s (x-translation)