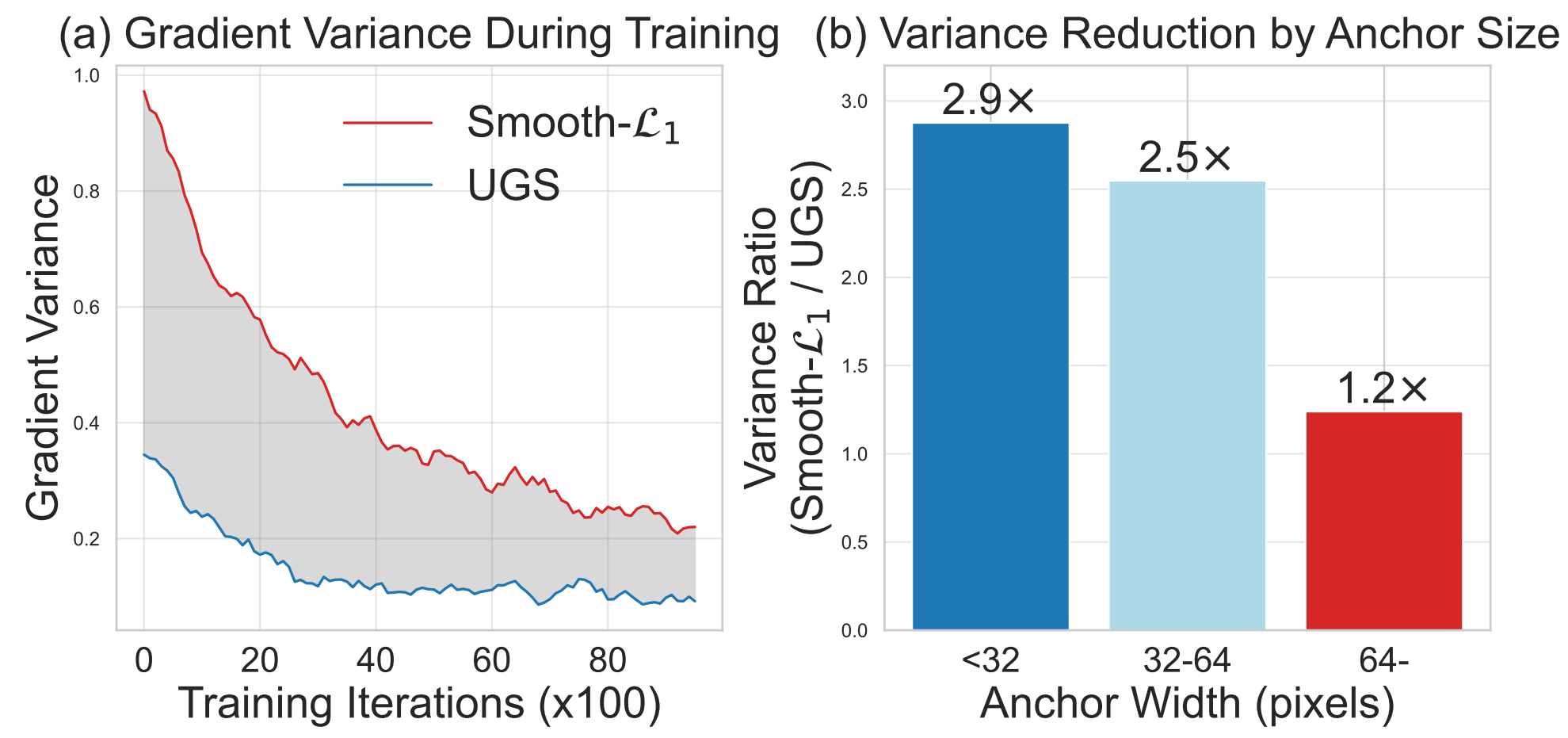

(a) Gradient Variance During Training
Gradient Variance
1.0
0.8
0.6
0.4
0.2
Smooth-$\mathcal{L}_1$
UGS
0
20
40
60
80
Training Iterations (x100)
(b) Variance Reduction by Anchor Size
Variance Ratio
(Smooth-$\mathcal{L}_1$ / UGS)
3.0
2.5
2.0
1.5
1.0
0.5
0.0
2.9×
2.5×
1.2×
<32
32-64
64-
Anchor Width (pixels)