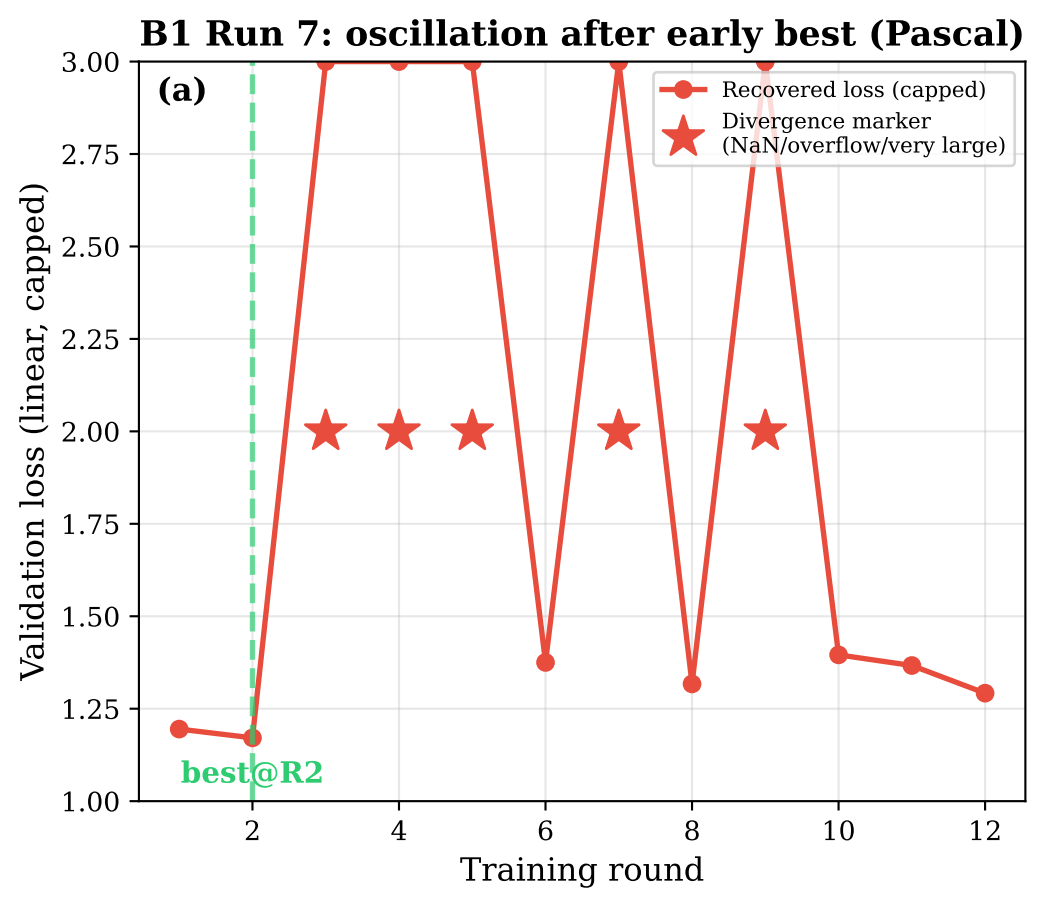
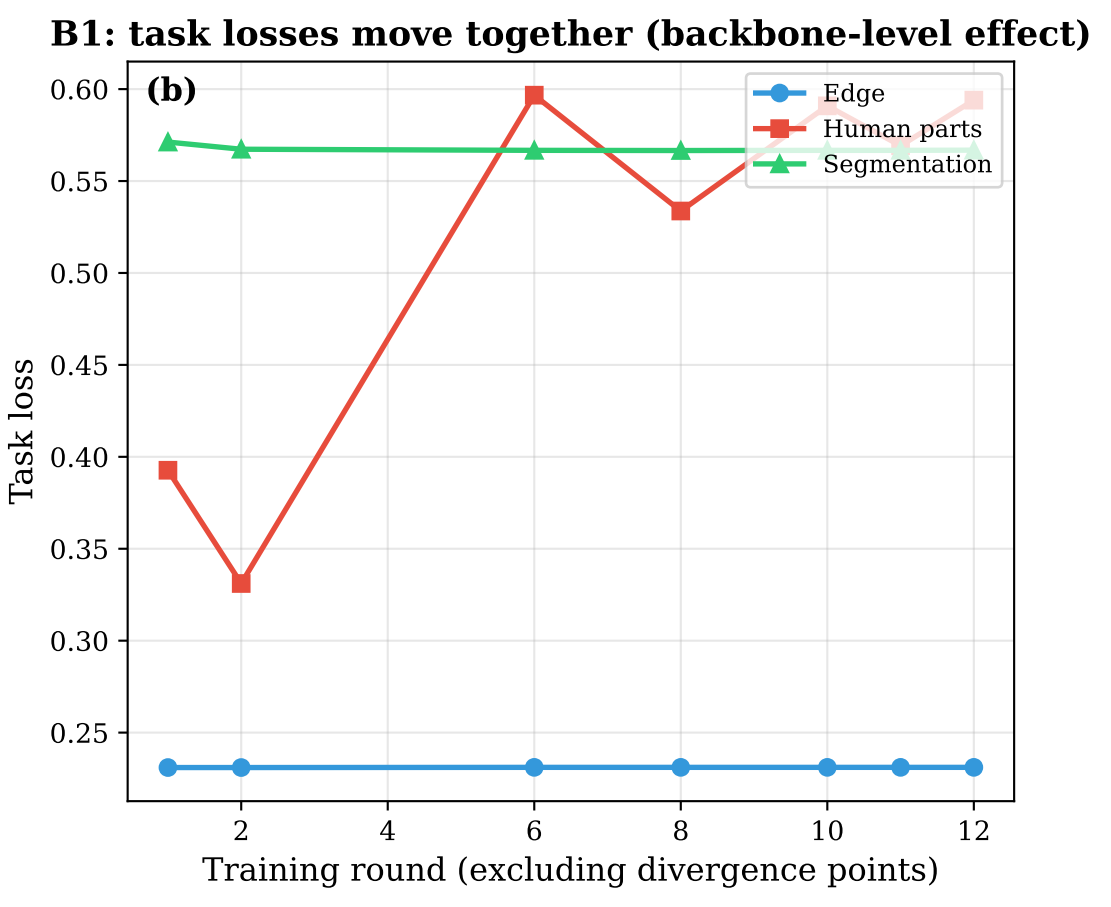

B1 Run 7: oscillation after early best (Pascal)
(a)
Recovered loss (capped)
Divergence marker
(NaN/overflow/very large)
best@R2
Validation loss (linear, capped)
Training round
3.00
2.75
2.50
2.25
2.00
1.75
1.50
1.25
1.00
2
4
6
8
10
12


B1: task losses move together (backbone-level effect)
(b)
Edge
Human parts
Segmentation
Task loss
Training round (excluding divergence points)
0.60
0.55
0.50
0.45
0.40
0.35
0.30
0.25
2
4
6
8
10
12