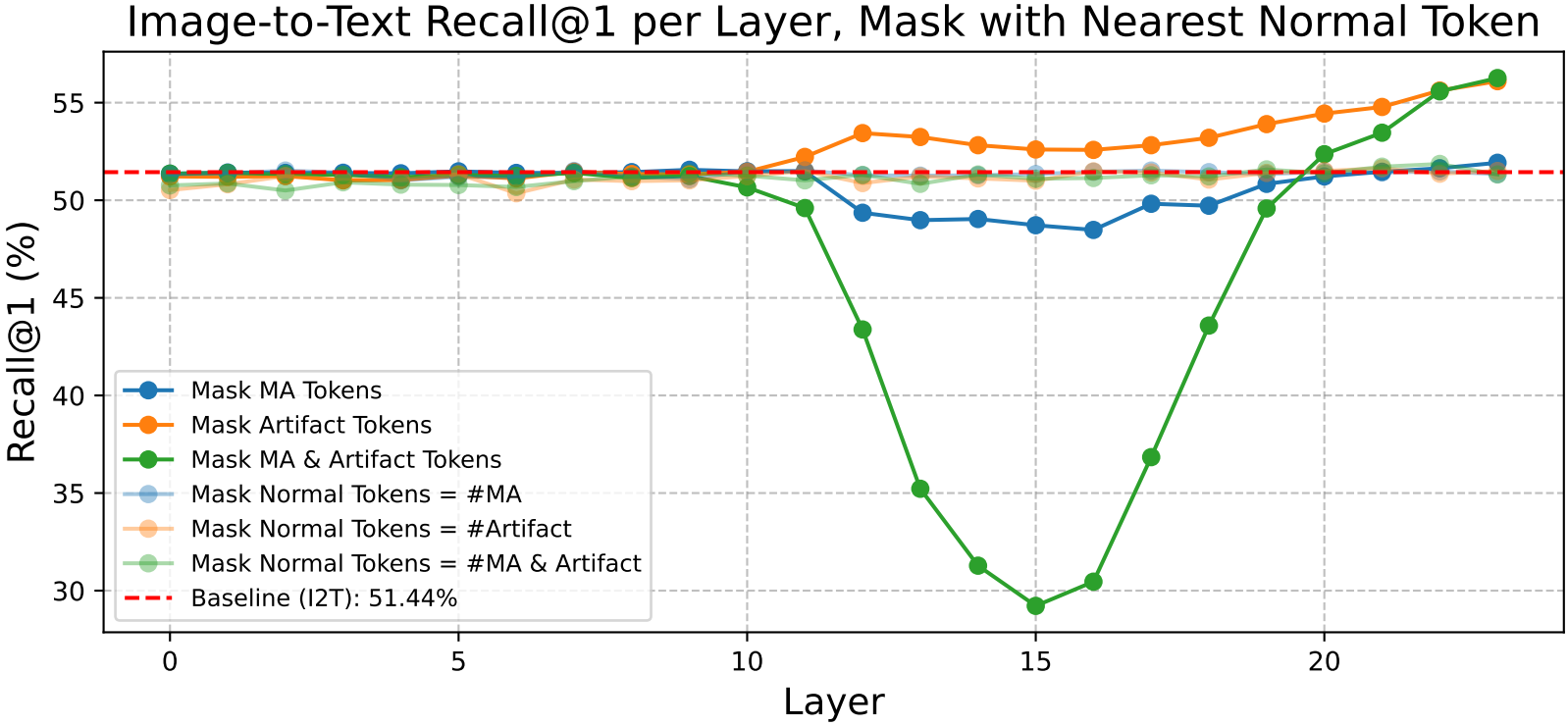
Image-to-Text Recall@1 per Layer, Mask with Nearest Normal Token
Recall@1 (%)
Layer
55
50
45
40
35
30
0
5
10
15
20
Mask MA Tokens
Mask Artifact Tokens
Mask MA & Artifact Tokens
Mask Normal Tokens = #MA
Mask Normal Tokens = #Artifact
Mask Normal Tokens = #MA & Artifact
Baseline (I2T): 51.44%