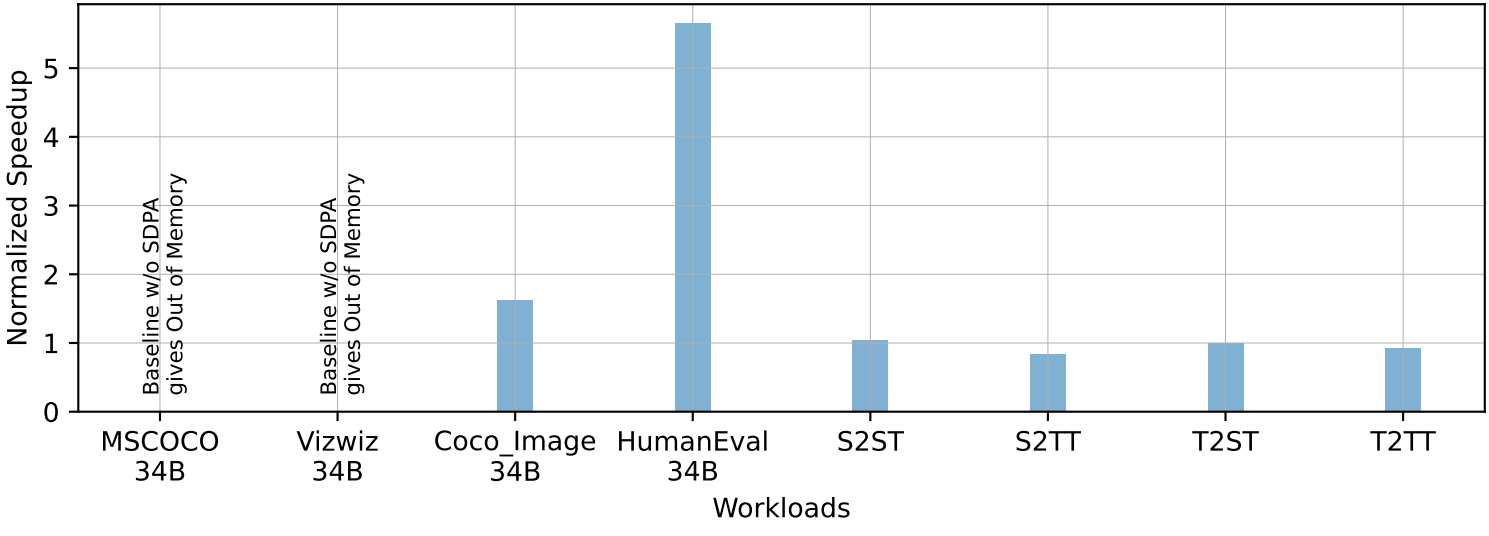
Normalized Speedup
0
1
2
3
4
5
Baseline w/o SDPA
gives Out of Memory
Baseline w/o SDPA
gives Out of Memory
MSCOCO
34B
Vizwiz
34B
Coco_Image
34B
HumanEval
34B
S2ST
S2TT
T2ST
T2TT
Workloads