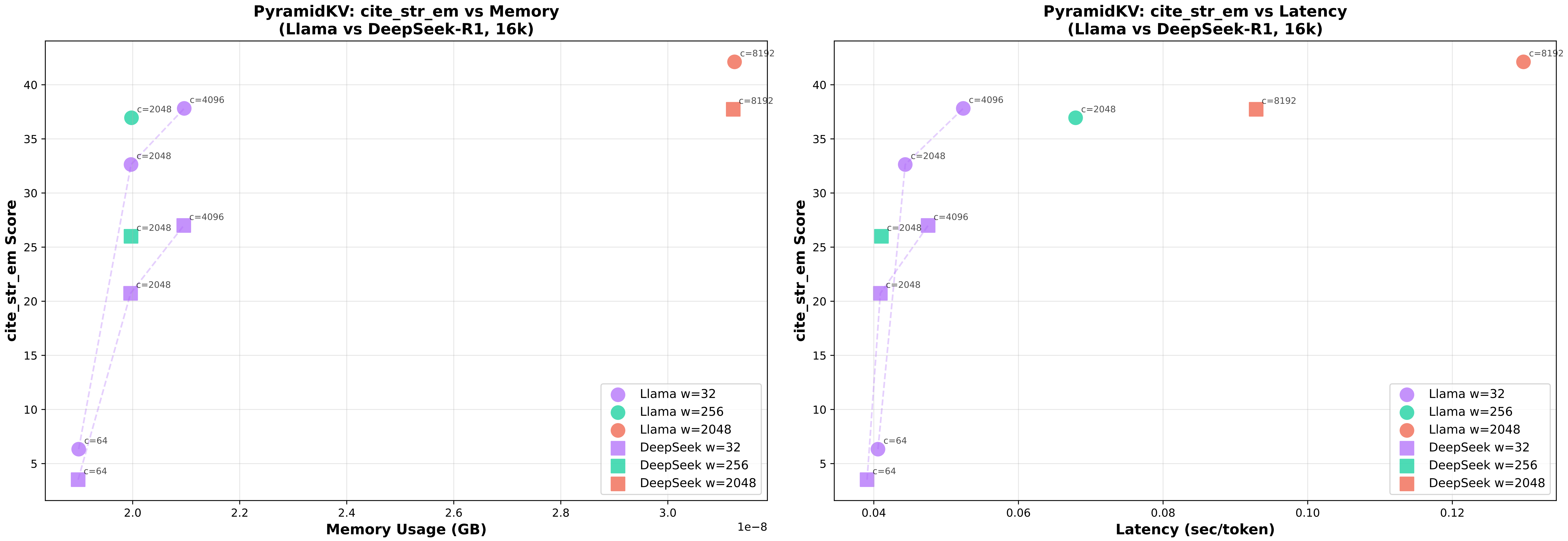

PyramidKV: cite_str_em vs Memory
(Llama vs DeepSeek-R1, 16k)

PyramidKV: cite_str_em vs Latency
(Llama vs DeepSeek-R1, 16k)