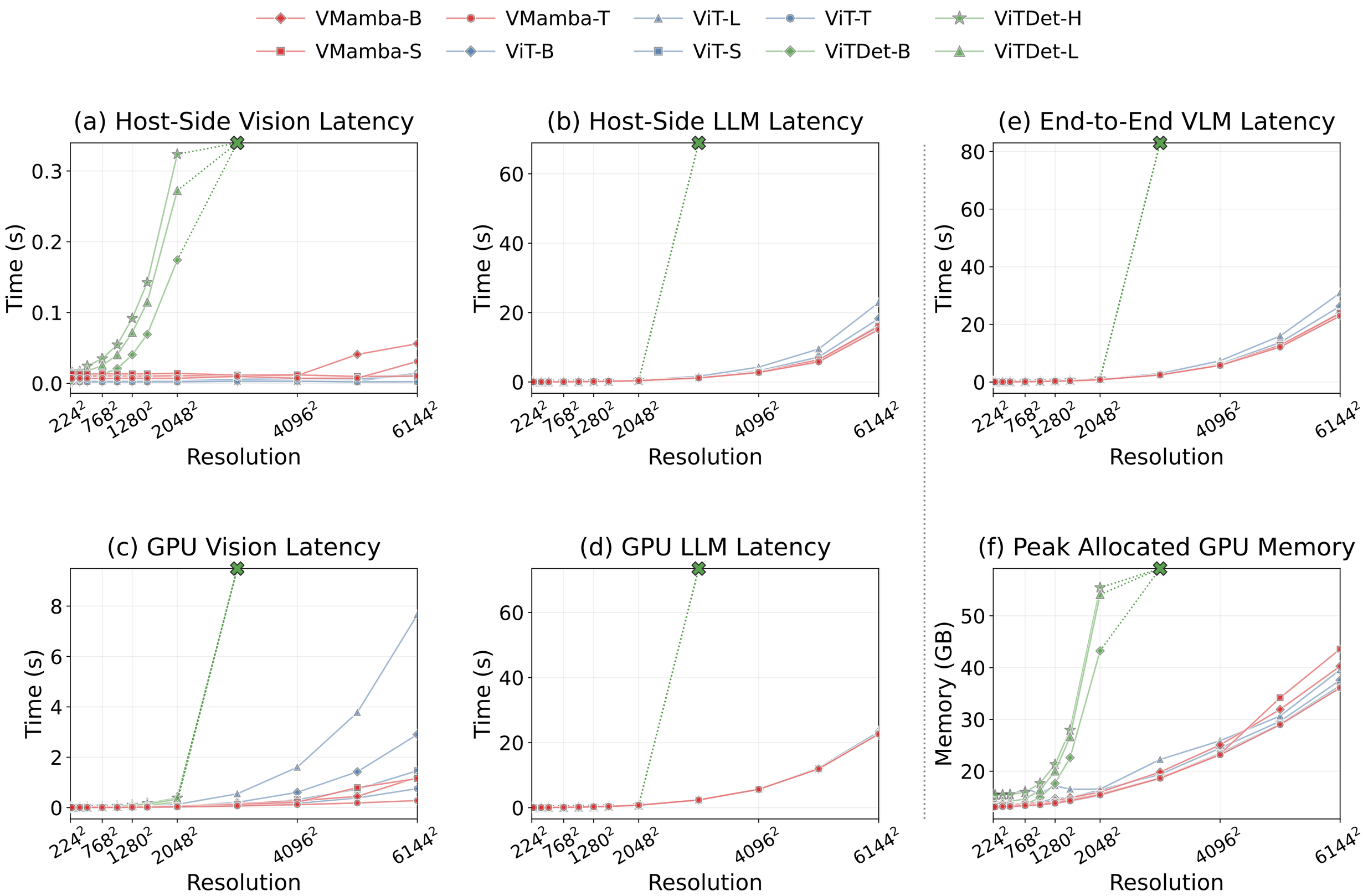
VMamba-B
VMamba-S
VMamba-T
ViT-B
ViT-L
ViT-S
ViT-T
ViTDet-B
ViTDet-H
ViTDet-L
(a) Host-Side Vision Latency
(b) Host-Side LLM Latency
(e) End-to-End VLM Latency
(c) GPU Vision Latency
(d) GPU LLM Latency
(f) Peak Allocated GPU Memory
Time (s)
Memory (GB)
Resolution
224²
768²
1280²
2048²
4096²
6144²