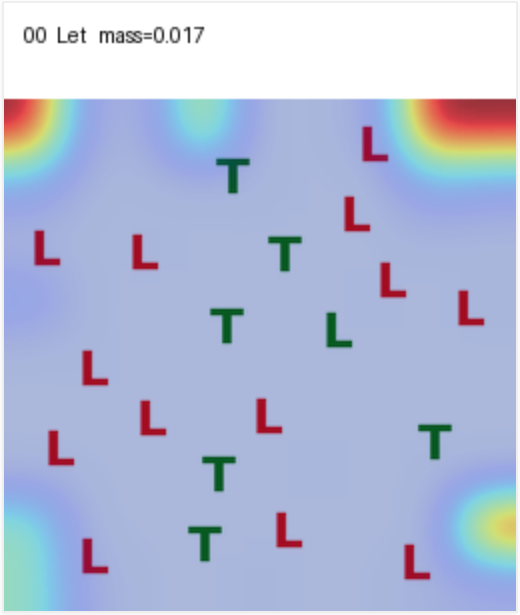
Image: incongruent-20_0
Mode: energy
Generated token (index — text): 00 — Let

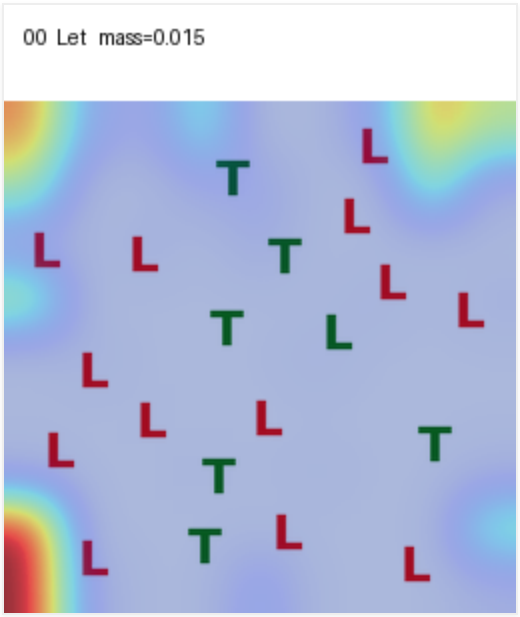
[00] Let

prob = L1-normalized over visual tokens (sum=1; shows allocation). energy = prob × total image attention (shows allocation + magnitude).

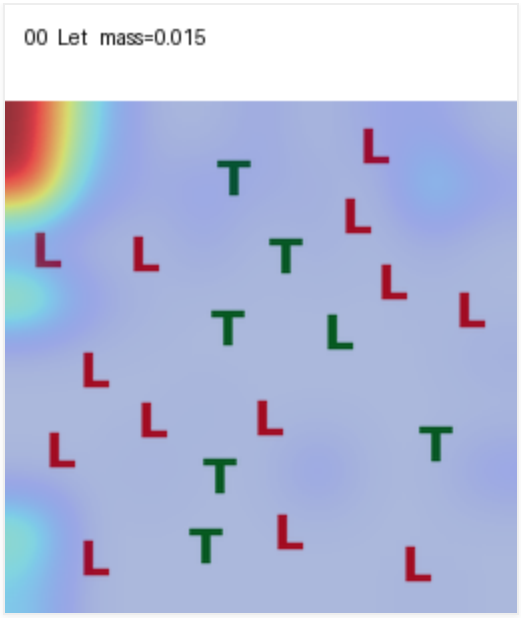
Layer 0



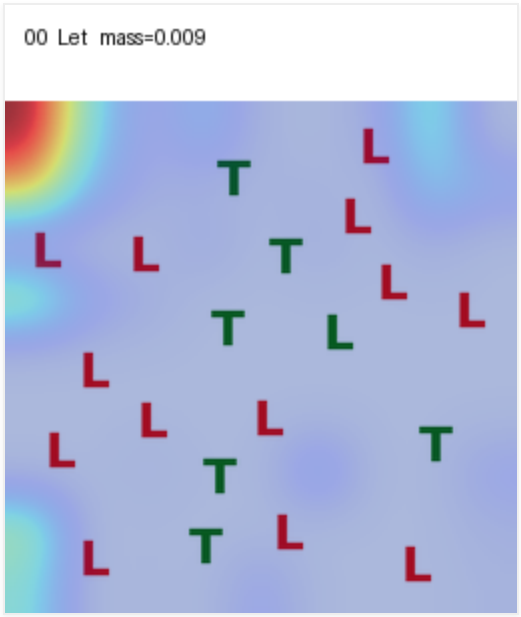
Layer 1



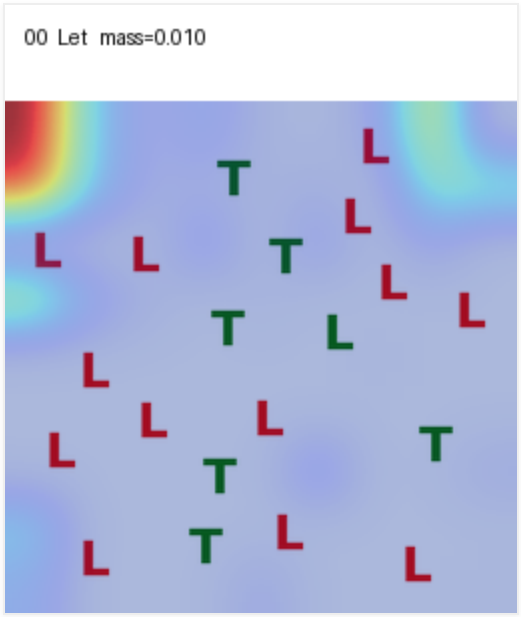
Layer 10



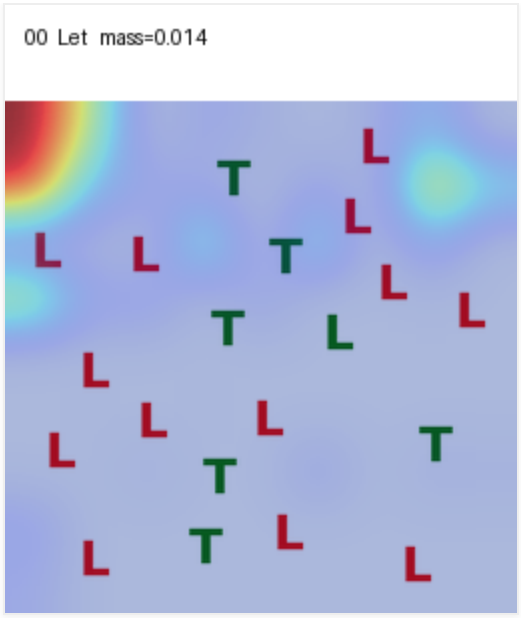
Layer 11



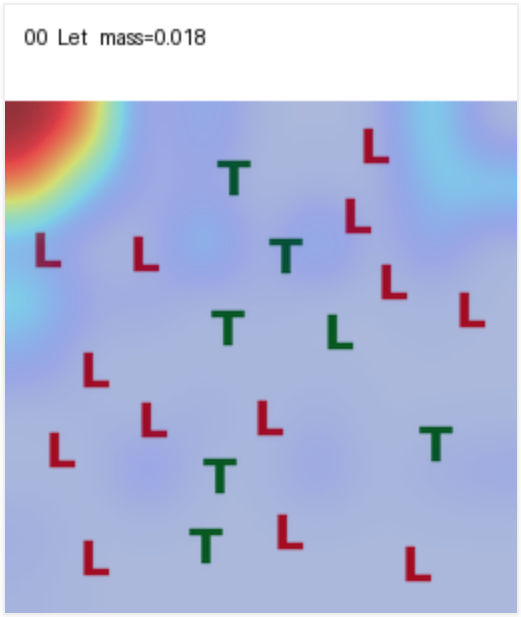
Layer 12



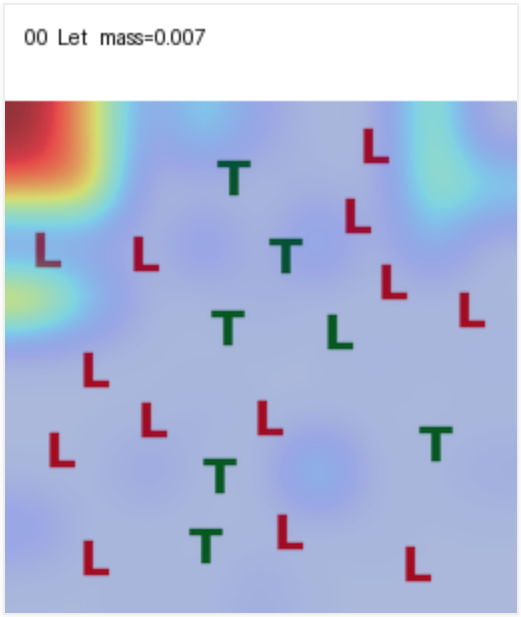
Layer 13



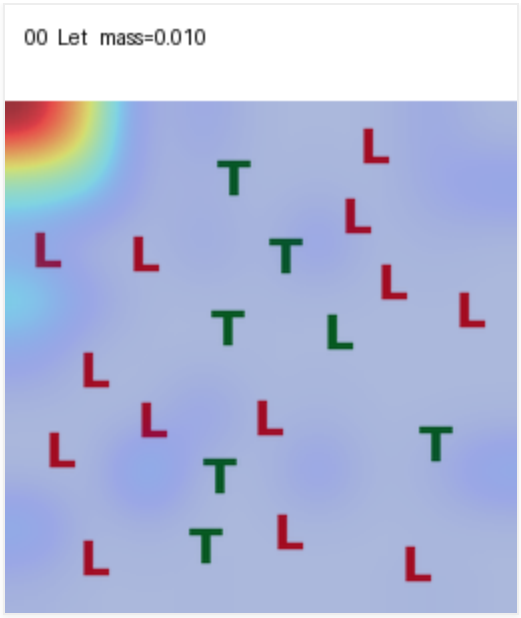
Layer 14



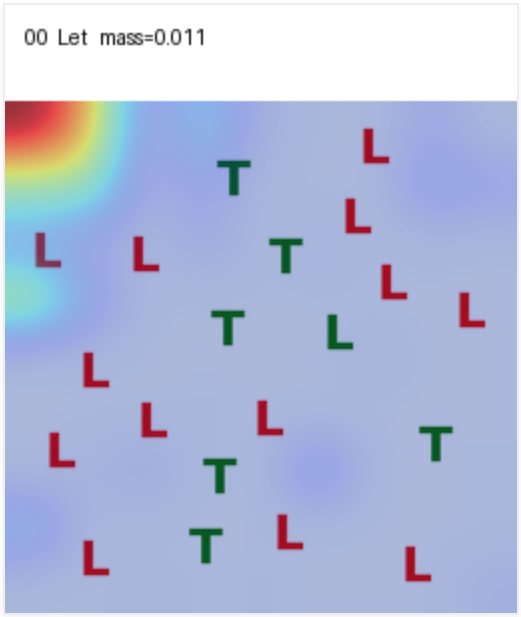
Layer 15



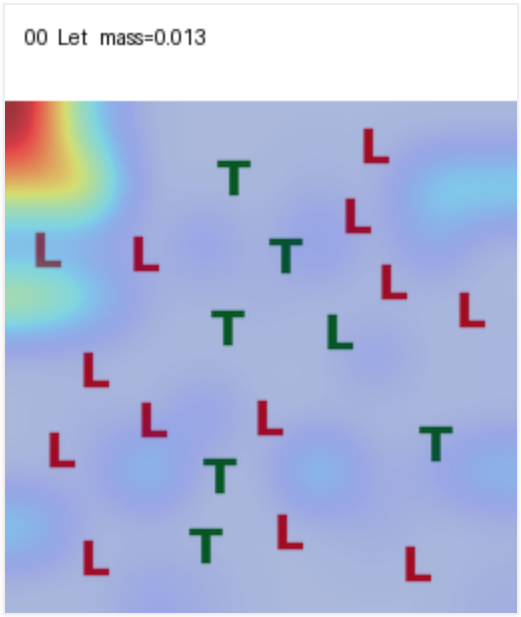
Layer 16



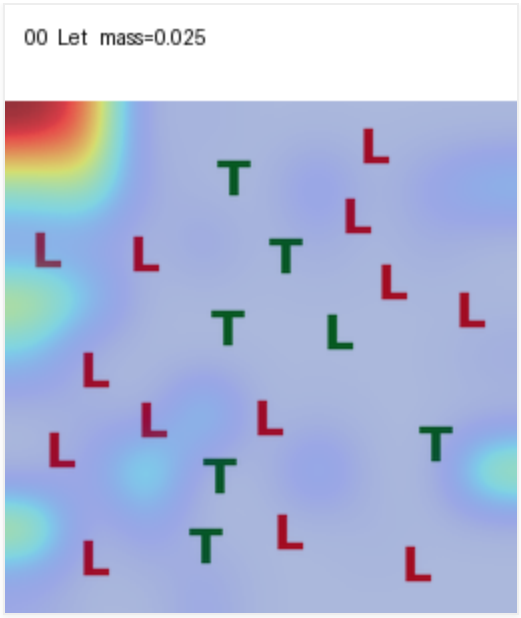
Layer 17



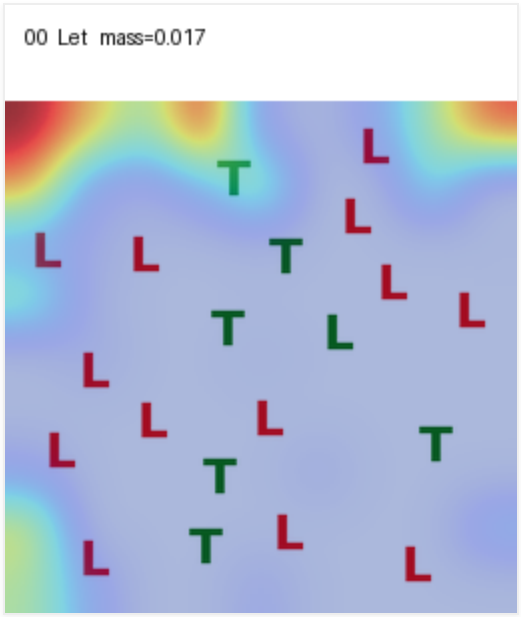
Layer 18



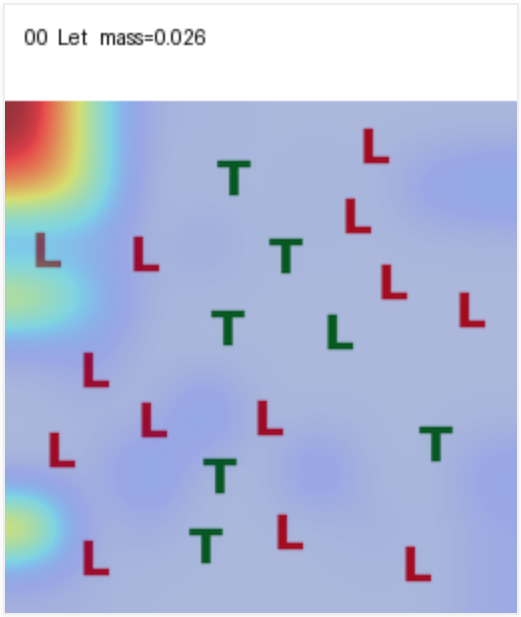
Layer 19



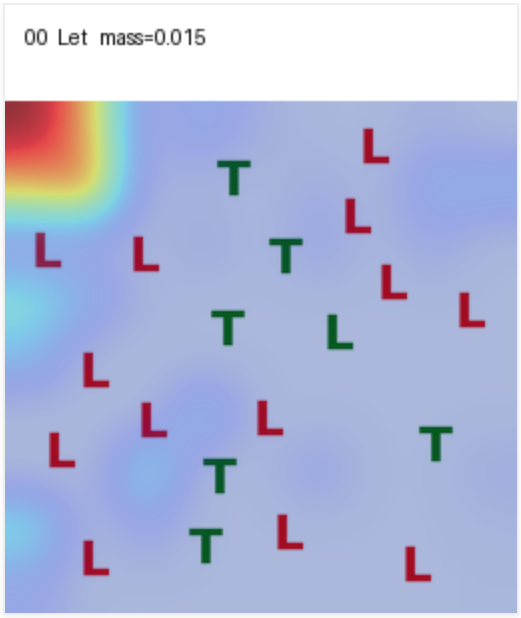
Layer 2



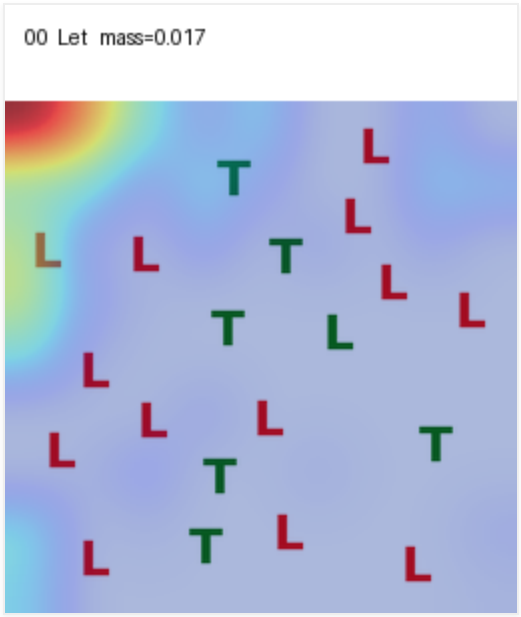
Layer 20



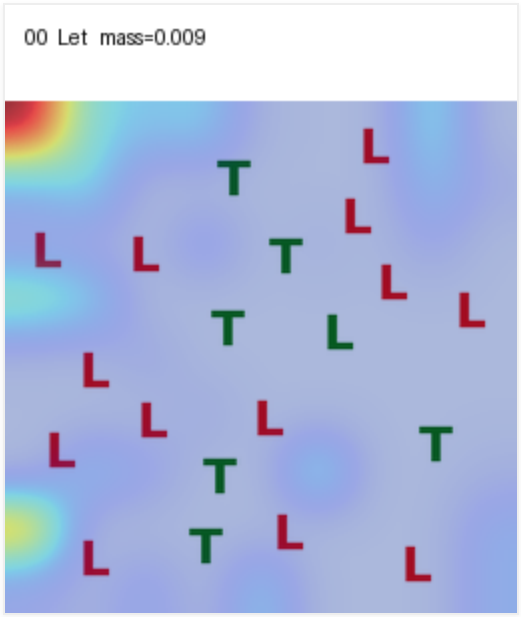
Layer 21



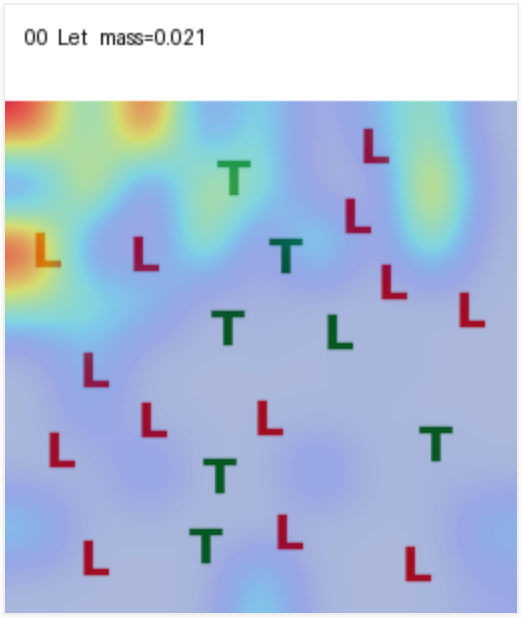
Layer 22



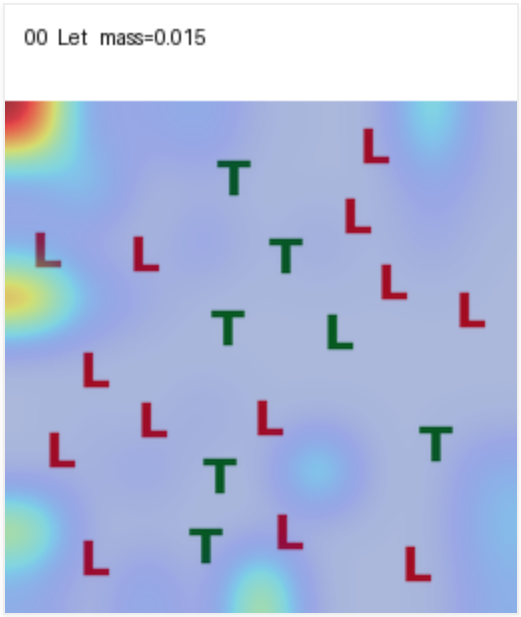
Layer 23



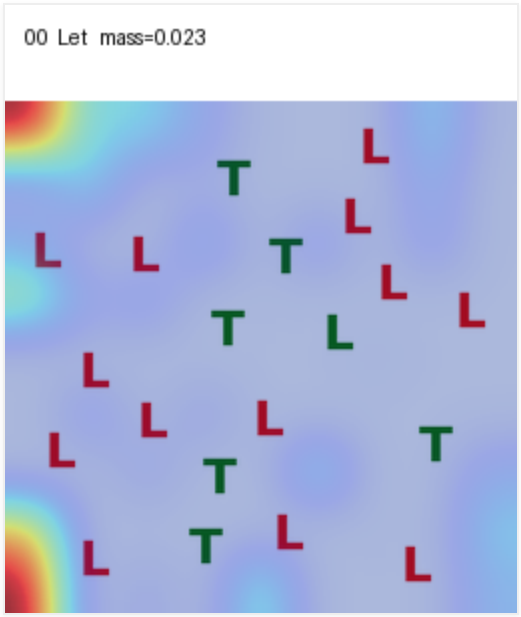
Layer 24



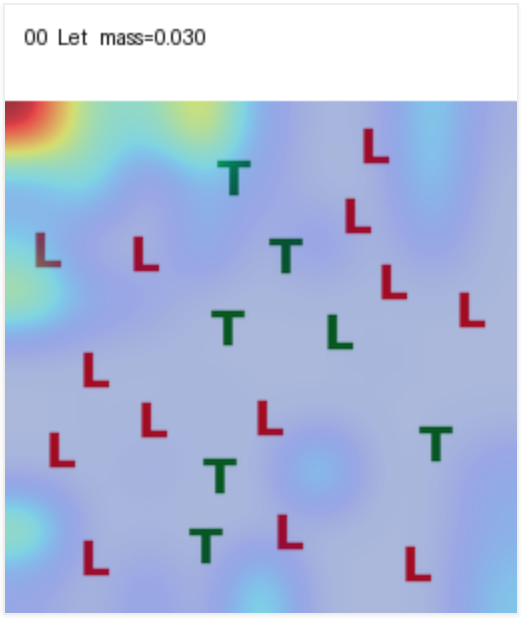
Layer 25



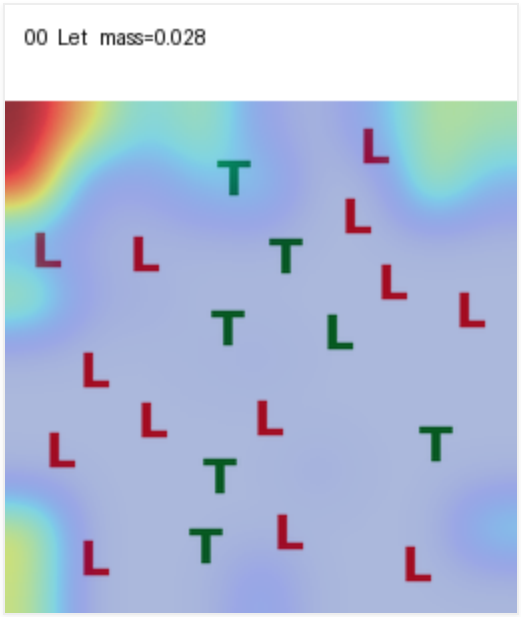
Layer 26



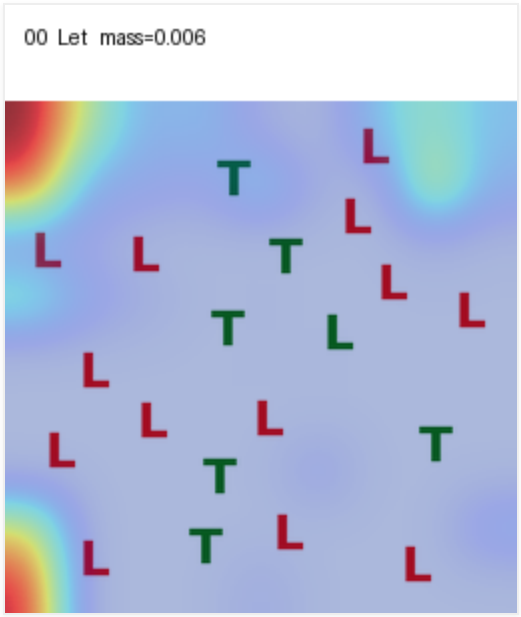
Layer 27



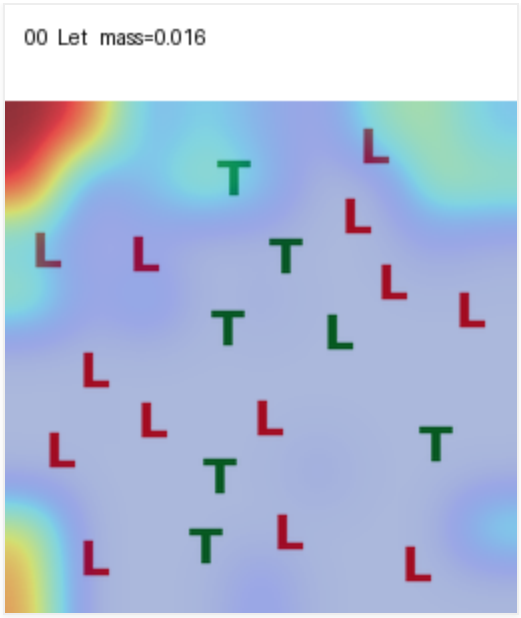
Layer 3



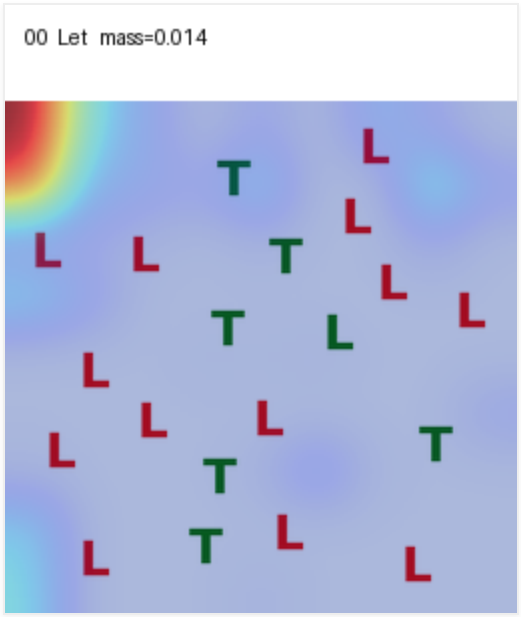
Layer 4



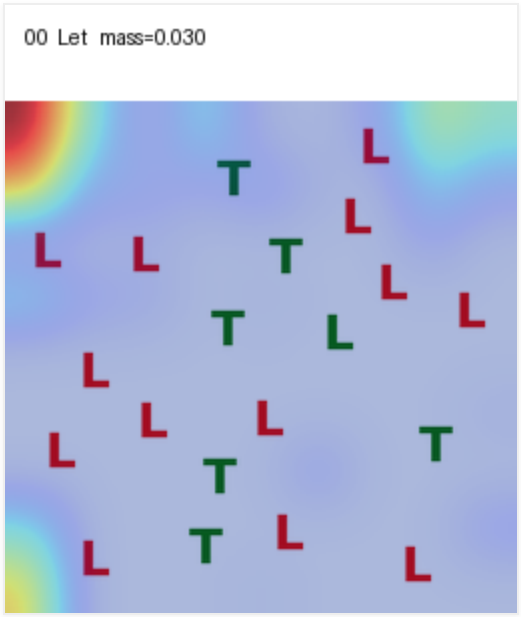
Layer 5



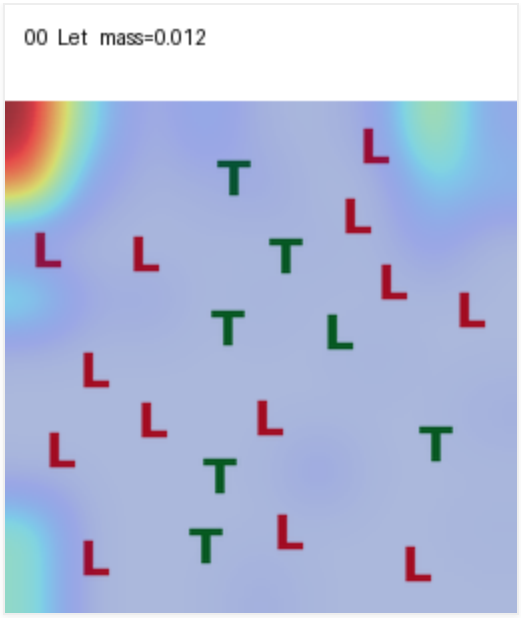
Layer 6



Layer 7



Layer 8



Layer 9