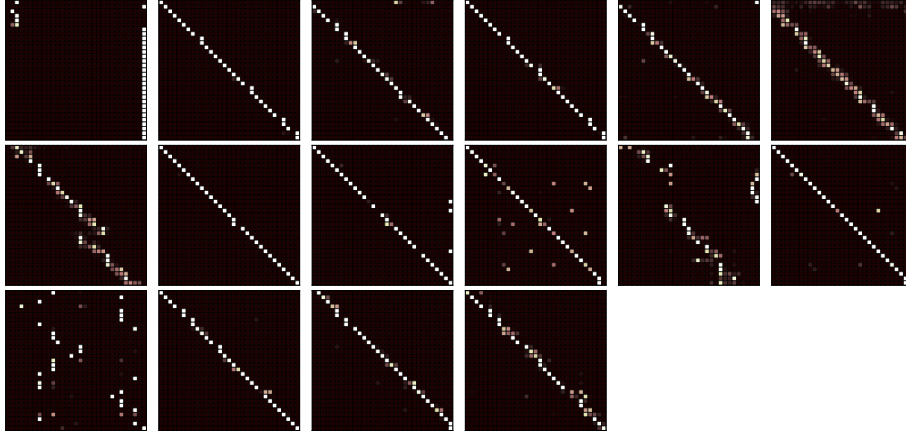
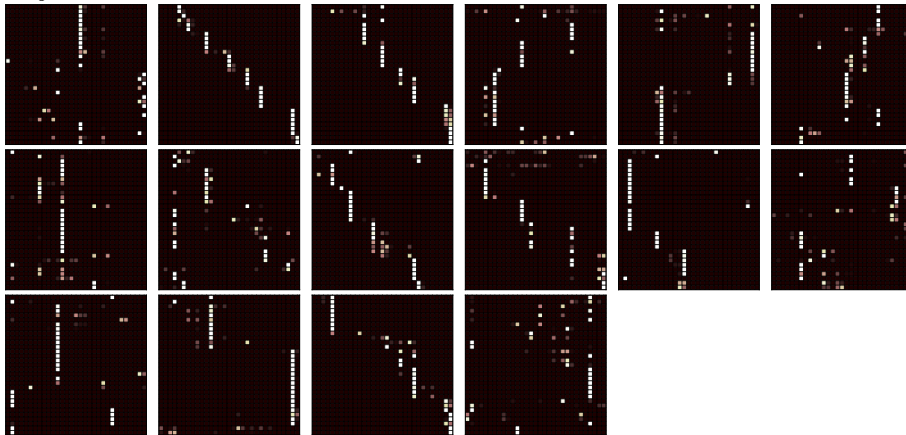


Appendix: Visualization of all 96 self-attention heads

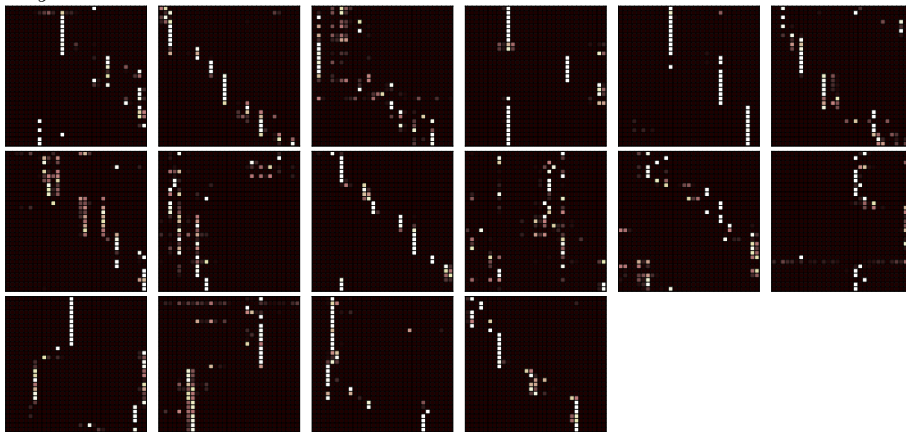
Layer 1:



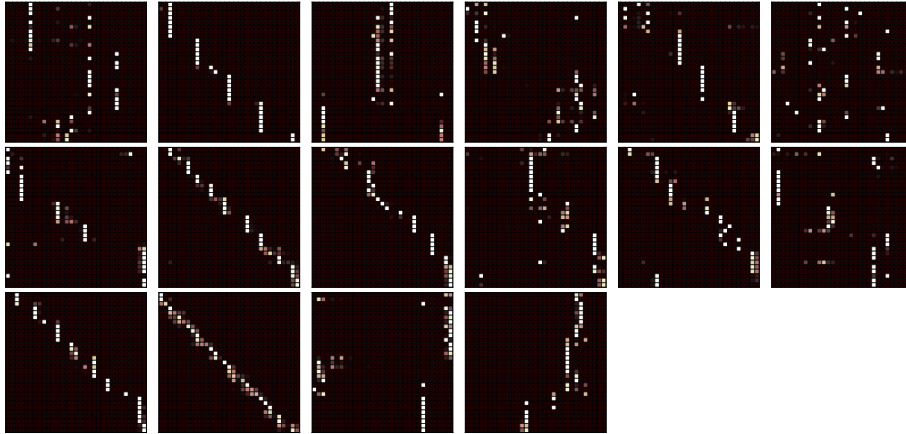
Layer 2:



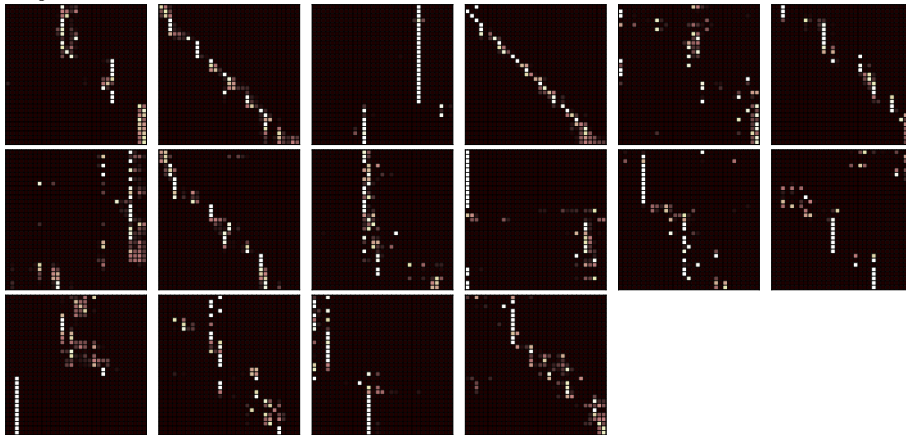
Layer 3:



Layer 4:



Layer 5:



Layer 6:

