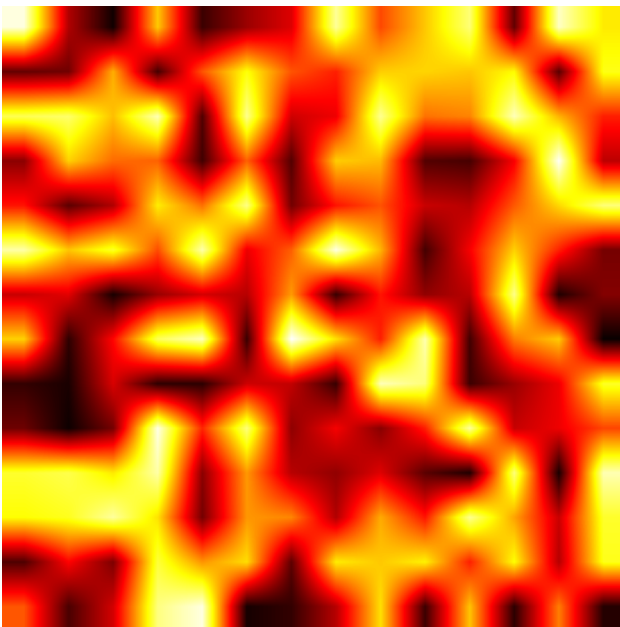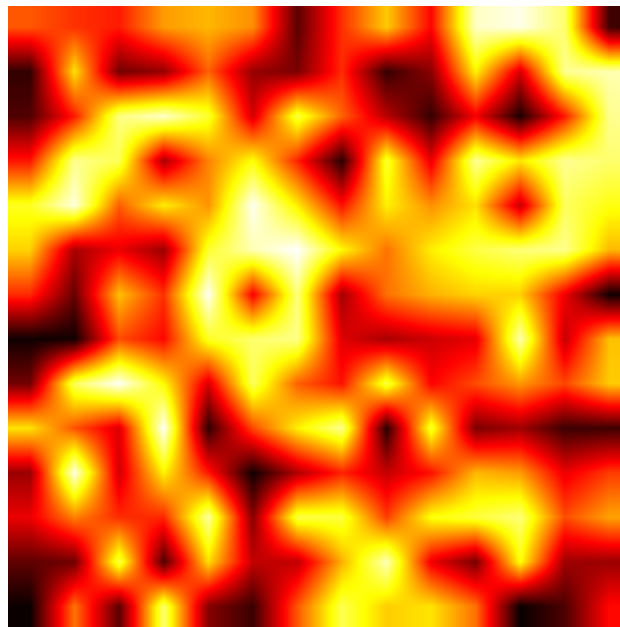VLM Cross-Attention Visualization: Text→Ima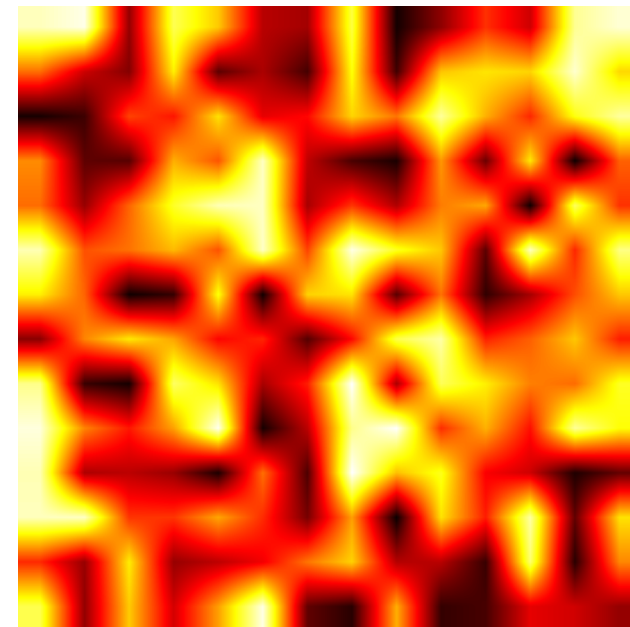ge