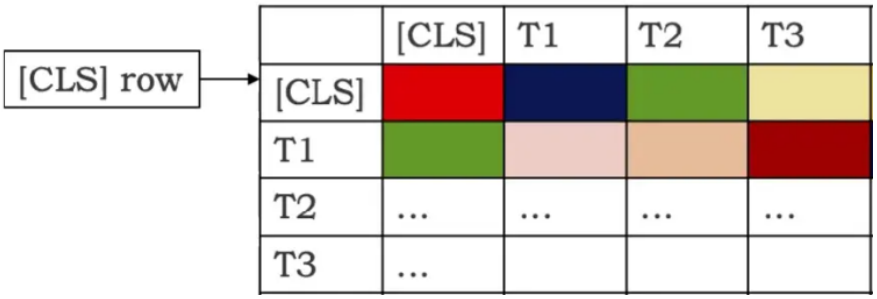
Attention 模块中，可以提取 attention_weight 矩阵。将矩阵转置，可得到如下的矩阵，矩阵的第 i 行表示所有 token 对输出 token i 的贡献



ViT 是分类模型，用包含了全局特征的[CLS] token 去做预测，所以我们关注 attention matrix 中的第 0 行，这一行就可以当作所有输入 token 对模型决策的贡献程度。取 Transformer 最后一层的 attention matrix，提取对应[CLS]的那一行，得到一个 197 维的向量，然后再剔除第 0 个元素，就得到了最后一层的 196 个输入 token，对最终的[CLS] token 的贡献。如果我们认为最后一层的输入 token i，就代表了 patch i，则我们可以用这个向量画出 attention 的热力图，表示[CLS]更关注哪些 patch。（见下面的 raw_attn_weights 和 raw_attn）

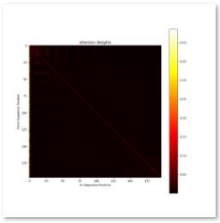
但由于 token 经过了多层自注意力模块，最后一层的 token i 不完全代表 patch i，于是尝试采用 attention rollout 的方法来分析注意力的流动。假设层与层之间的 self-attention 的 token 信息是线性传递的，self-attention 层是线性组合的，而其它的 FFN 影响不大可以忽略。所以我们就可以用矩阵相乘来“聚合”多层 attention matrix 的信息。在计算第 j 层的输入 token 对第 i 层的输出 token 的贡献时（j<=i），可以用以下公式进行计算，其中~A 代表 rollout attn

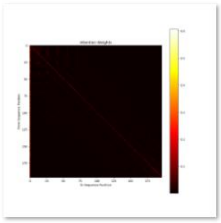
$$\tilde{A}(l_i) = \begin{cases} A(l_i)\tilde{A}(l_{i-1}) & \text{if } i > j \\ A(l_i) & \text{if } i = j \end{cases}$$

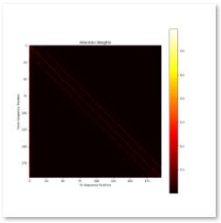
其中矩阵 A=0.5*原始注意力矩阵+0.5*单位矩阵（建模残差连接）

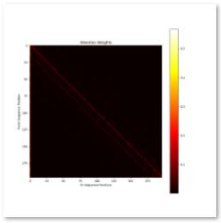
$$A = 0.5W_{att} + 0.5I,$$

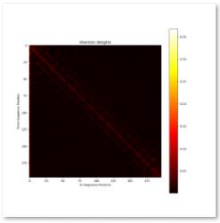
以下内容展示原图，raw attention map（横轴是 to token，纵轴是 from token），raw attention 在 patch level 上的可视化，rollout attention map，rollout attention 在 patch level 上的可视化。

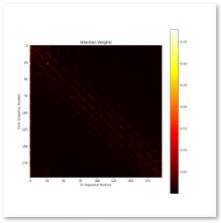

raw_attn_weights_layer_0

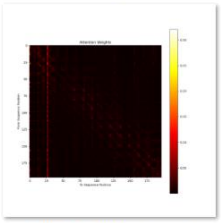

raw_attn_weights_layer_1

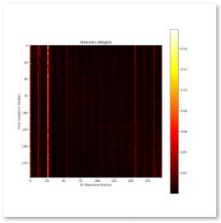

raw_attn_weights_layer_2

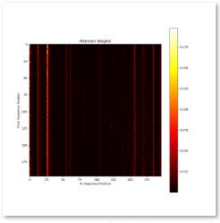

raw_attn_weights_layer_3

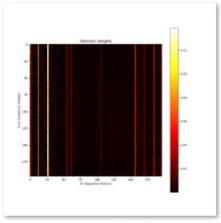

raw_attn_weights_layer_4

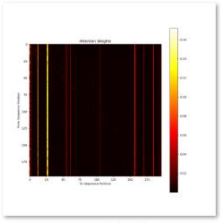

raw_attn_weights_layer_5

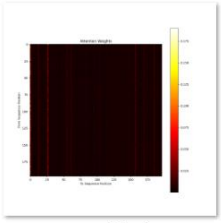

raw_attn_weights_layer_6

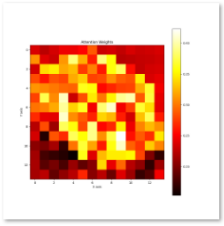

raw_attn_weights_layer_7

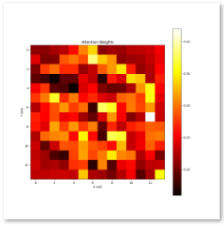

raw_attn_weights_layer_8

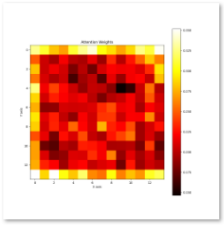

raw_attn_weights_layer_9

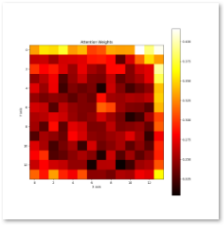

raw_attn_weights_layer_10

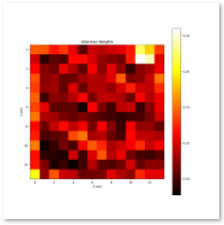

raw_attn_weights_layer_11

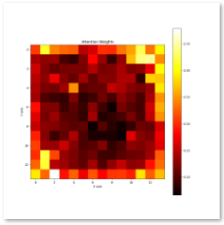

raw_attn_layer_0

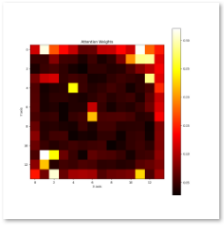

raw_attn_layer_1

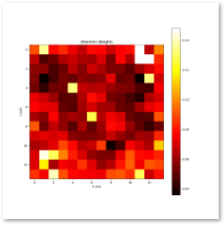

raw_attn_layer_2

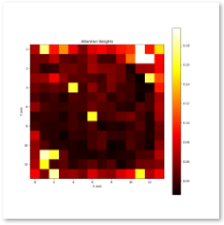

raw_attn_layer_3

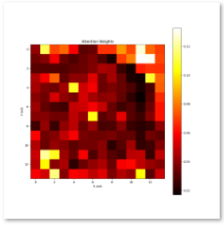

raw_attn_layer_4

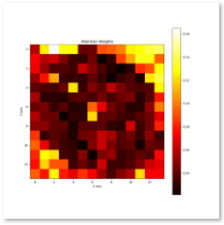

raw_attn_layer_5

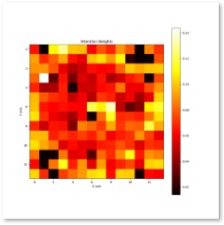

raw_attn_layer_6

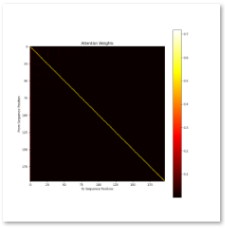

raw_attn_layer_7

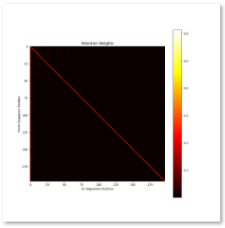

raw_attn_layer_8

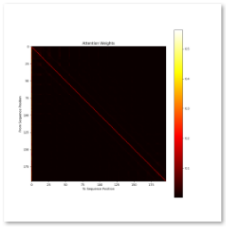

raw_attn_layer_9

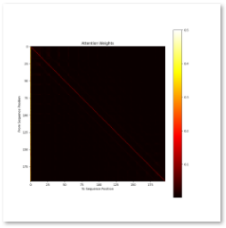

raw_attn_layer_10

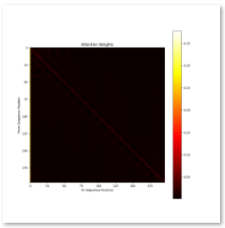

raw_attn_layer_11

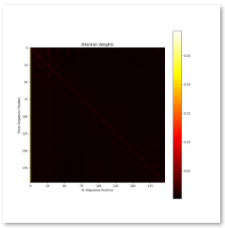

rollout_attn_weights_layer_0

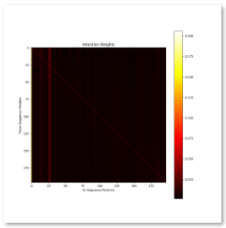

rollout_attn_weights_layer_1

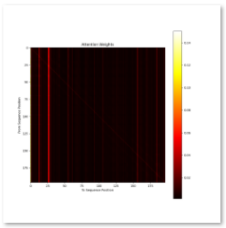

rollout_attn_weights_layer_2

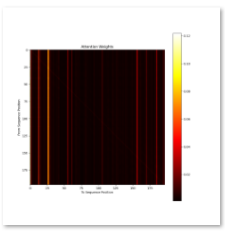

rollout_attn_weights_layer_3

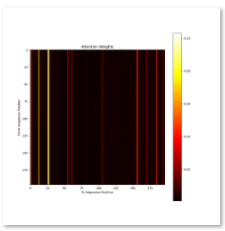

rollout_attn_weights_layer_4

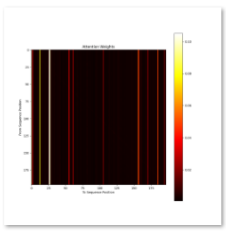

rollout_attn_weights_layer_5

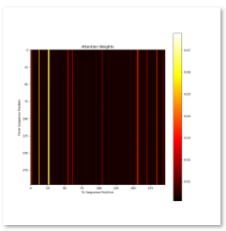

rollout_attn_weights_layer_6

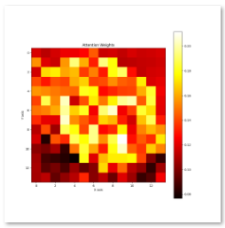

rollout_attn_weights_layer_7

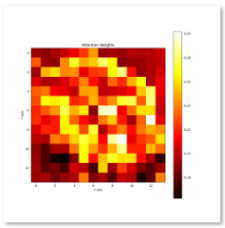

rollout_attn_weights_layer_8

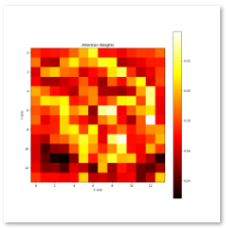

rollout_attn_weights_layer_9

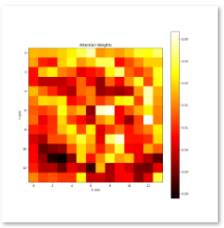

rollout_attn_weights_layer_10

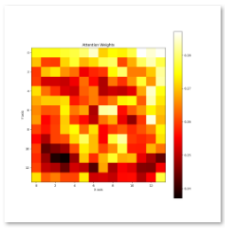

rollout_attn_weights_layer_11

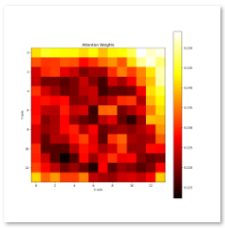

rollout_attn_layer_0

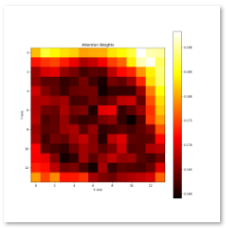

rollout_attn_layer_1

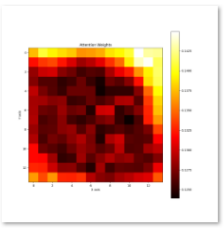

rollout_attn_layer_2

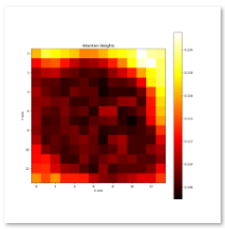

rollout_attn_layer_3

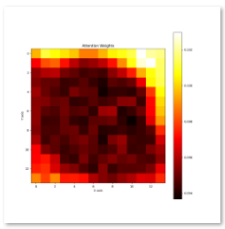

rollout_attn_layer_4

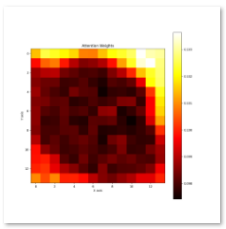

rollout_attn_layer_5

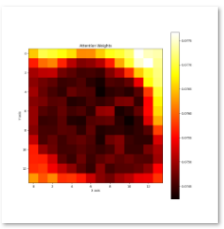

rollout_attn_layer_6

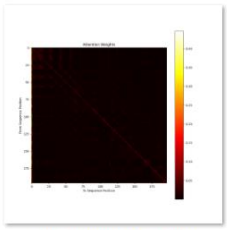

rollout_attn_layer_7

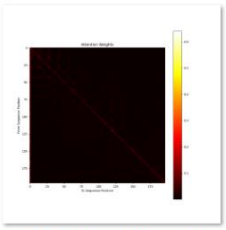

rollout_attn_layer_8

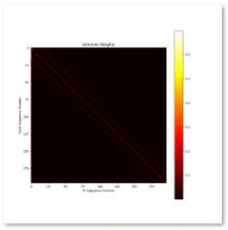

rollout_attn_layer_9

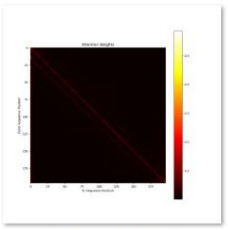

rollout_attn_layer_10

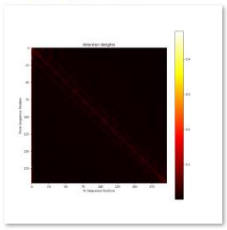

rollout_attn_layer_11

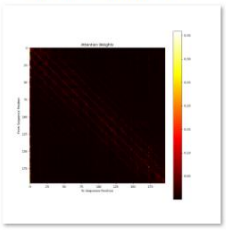

raw_attn_weights_layer_0

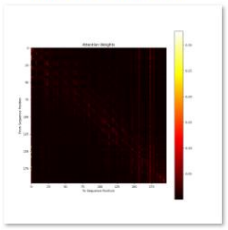

raw_attn_weights_layer_1

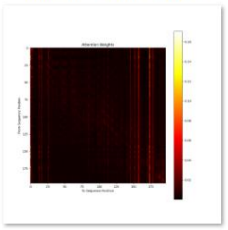

raw_attn_weights_layer_2

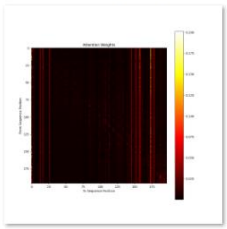

raw_attn_weights_layer_3

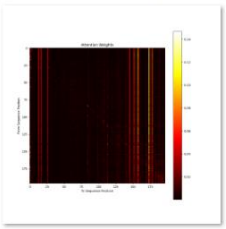

raw_attn_weights_layer_4

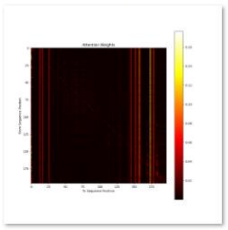

raw_attn_weights_layer_5

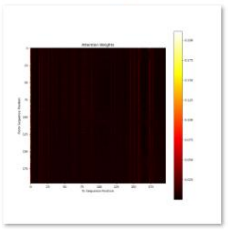

raw_attn_weights_layer_6

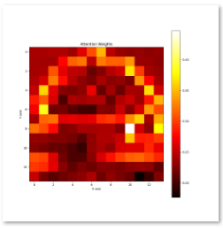

raw_attn_weights_layer_7

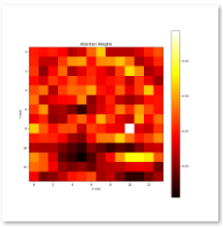

raw_attn_weights_layer_8

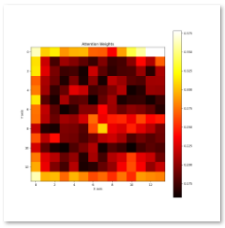

raw_attn_weights_layer_9

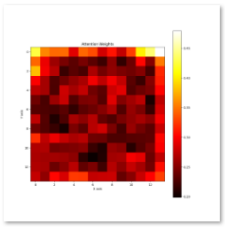

raw_attn_weights_layer_10

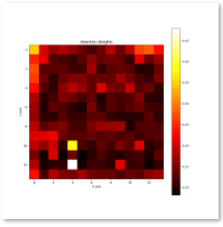

raw_attn_weights_layer_11

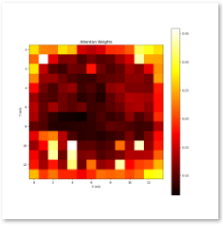

raw_attn_layer_0

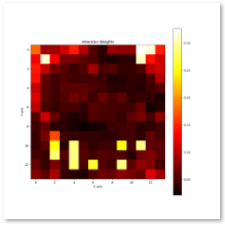

raw_attn_layer_1

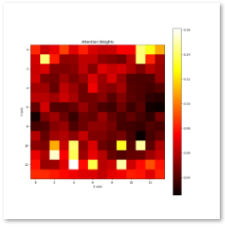

raw_attn_layer_2

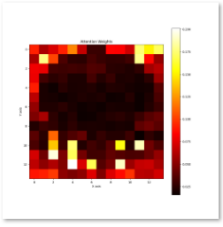

raw_attn_layer_3

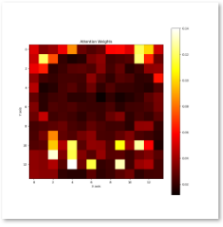

raw_attn_layer_4

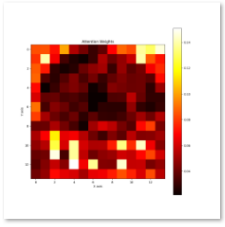

raw_attn_layer_5

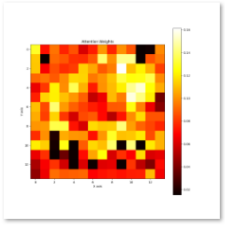

raw_attn_layer_6

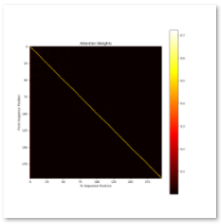

raw_attn_layer_7

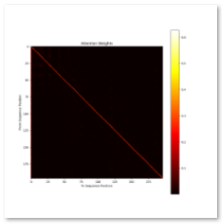

raw_attn_layer_8

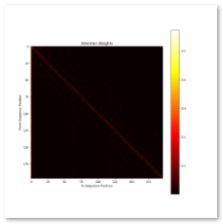

raw_attn_layer_9

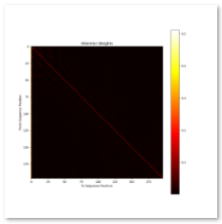

raw_attn_layer_10

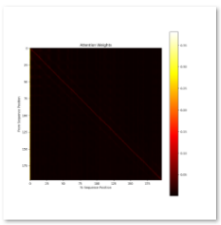

raw_attn_layer_11

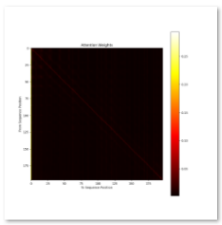

rollout_attn_weights_layer_0

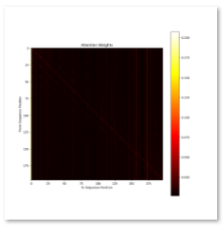

rollout_attn_weights_layer_1

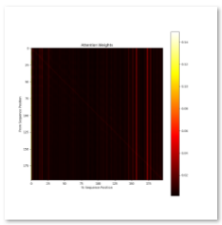

rollout_attn_weights_layer_2

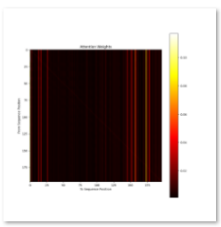

rollout_attn_weights_layer_3

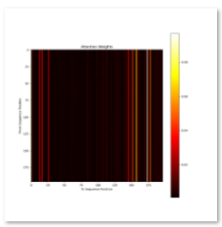

rollout_attn_weights_layer_4

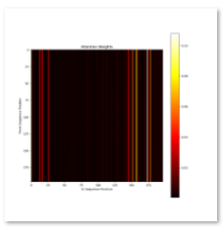

rollout_attn_weights_layer_5

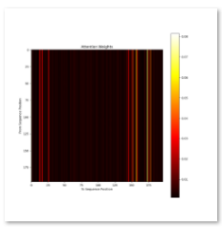

rollout_attn_weights_layer_6

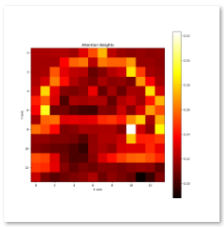

rollout_attn_weights_layer_7

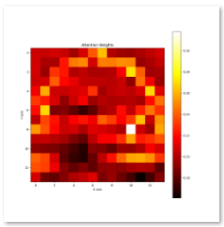

rollout_attn_weights_layer_8

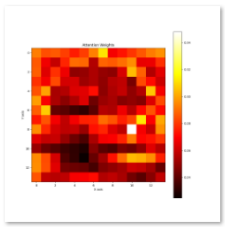

rollout_attn_weights_layer_9

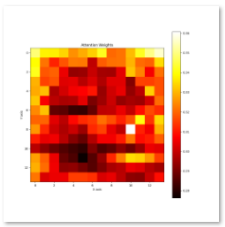

rollout_attn_weights_layer_10

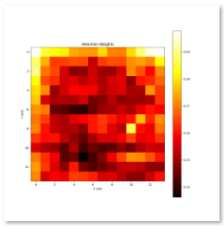

rollout_attn_weights_layer_11

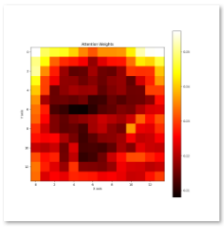

rollout_attn_layer_0

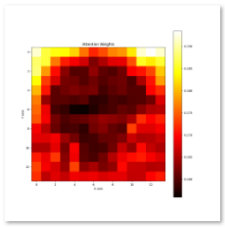

rollout_attn_layer_1

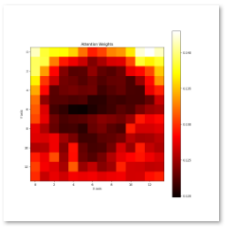

rollout_attn_layer_2

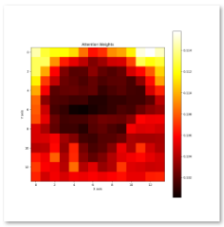

rollout_attn_layer_3

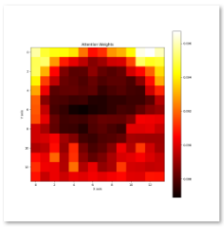

rollout_attn_layer_4

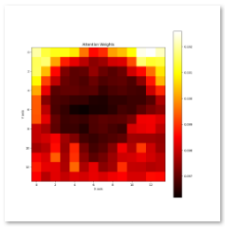

rollout_attn_layer_5

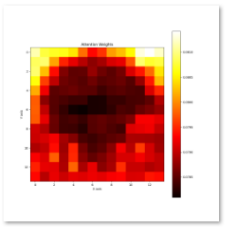

rollout_attn_layer_6

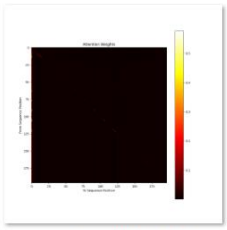

rollout_attn_layer_7

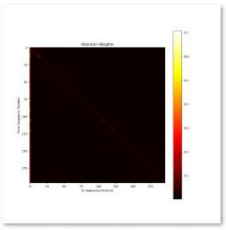

rollout_attn_layer_8

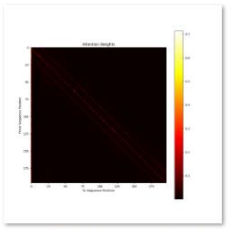

rollout_attn_layer_9

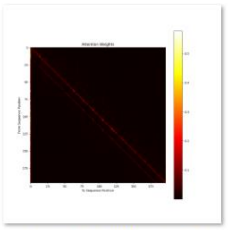

rollout_attn_layer_10

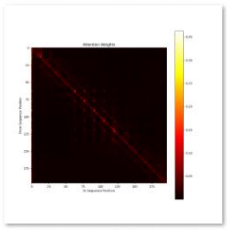

rollout_attn_layer_11

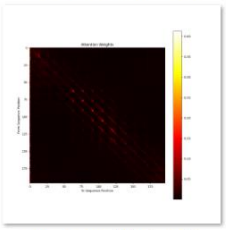

raw_attn_weights_layer_0

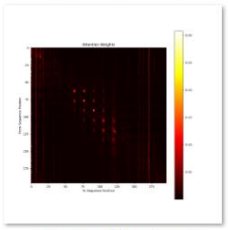

raw_attn_weights_layer_1

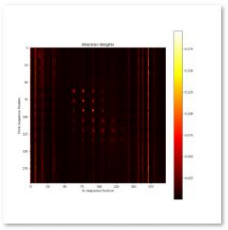

raw_attn_weights_layer_2

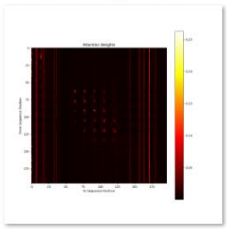

raw_attn_weights_layer_3

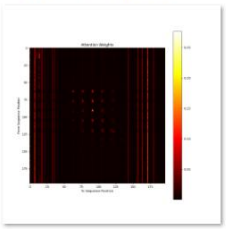

raw_attn_weights_layer_4

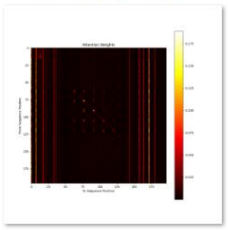

raw_attn_weights_layer_5

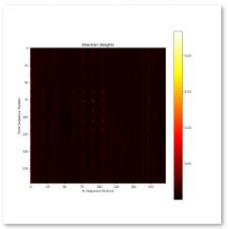

raw_attn_weights_layer_6

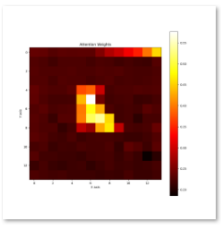

raw_attn_weights_layer_7

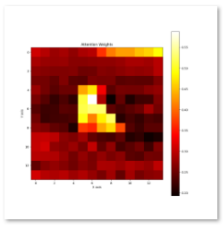

raw_attn_weights_layer_8

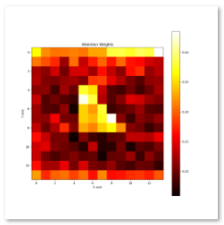

raw_attn_weights_layer_9

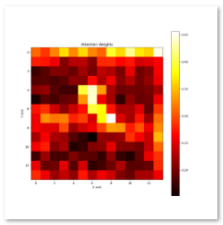

raw_attn_weights_layer_10

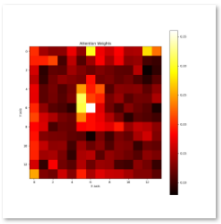

raw_attn_weights_layer_11

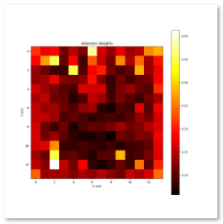

raw_attn_layer_0

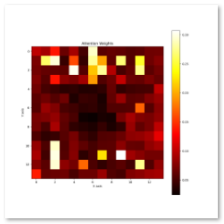

raw_attn_layer_1

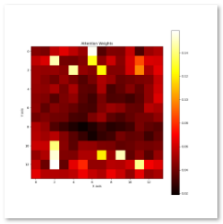

raw_attn_layer_2

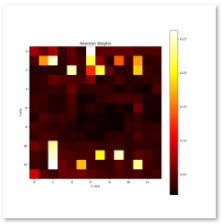

raw_attn_layer_3

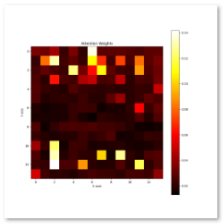

raw_attn_layer_4

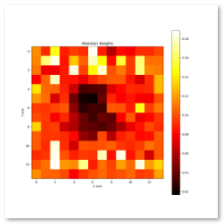

raw_attn_layer_5

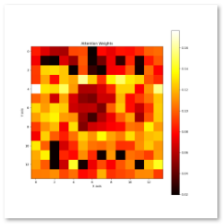

raw_attn_layer_6

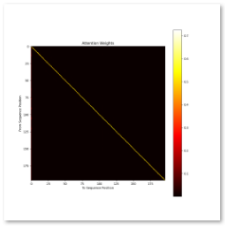

raw_attn_layer_7

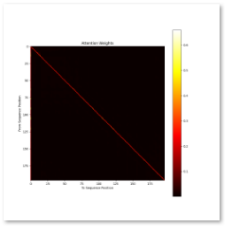

raw_attn_layer_8

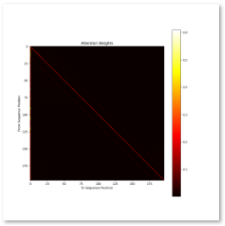

raw_attn_layer_9

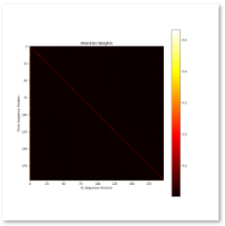

raw_attn_layer_10

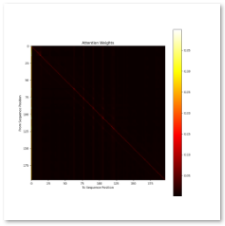

raw_attn_layer_11

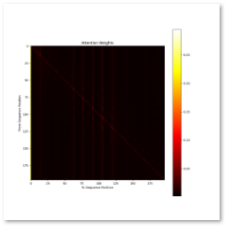

rollout_attn_weights_layer_0

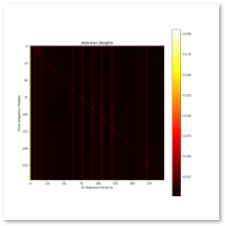

rollout_attn_weights_layer_1

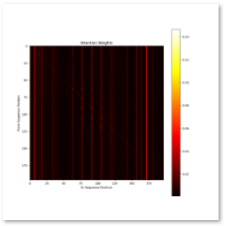

rollout_attn_weights_layer_2

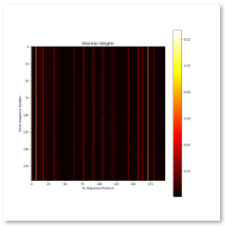

rollout_attn_weights_layer_3

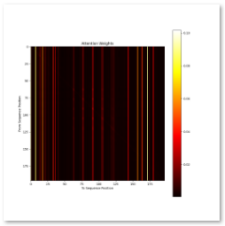

rollout_attn_weights_layer_4

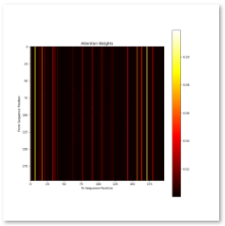

rollout_attn_weights_layer_5

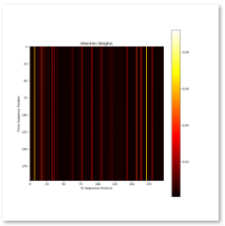

rollout_attn_weights_layer_6

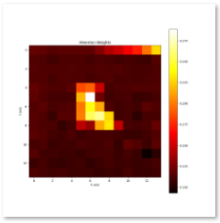

rollout_attn_weights_layer_7

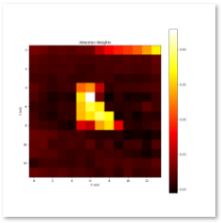

rollout_attn_weights_layer_8

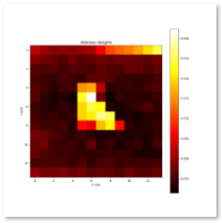

rollout_attn_weights_layer_9

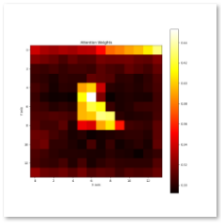

rollout_attn_weights_layer_10

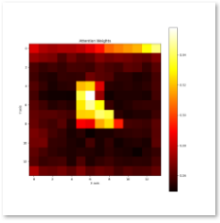

rollout_attn_weights_layer_11

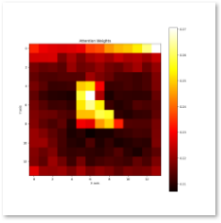

rollout_attn_layer_0

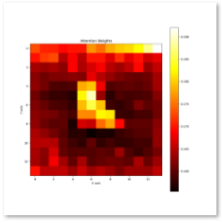

rollout_attn_layer_1

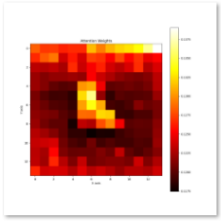

rollout_attn_layer_2

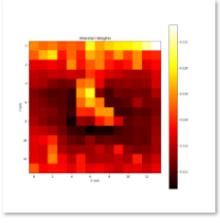

rollout_attn_layer_3

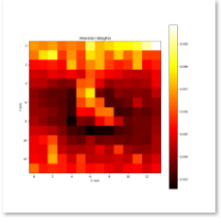

rollout_attn_layer_4

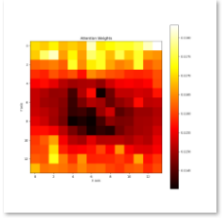

rollout_attn_layer_5

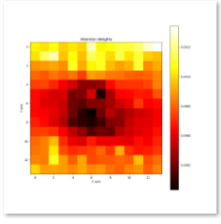

rollout_attn_layer_6

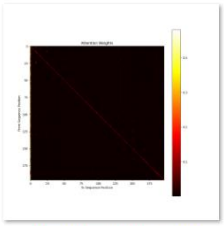

rollout_attn_layer_7

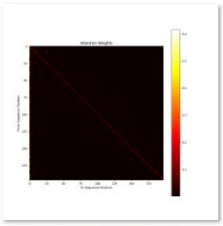

rollout_attn_layer_8

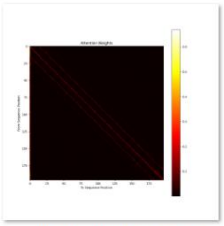

rollout_attn_layer_9

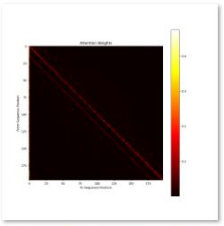

rollout_attn_layer_10

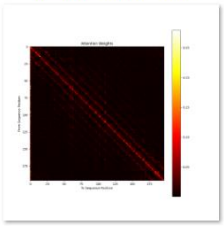

rollout_attn_layer_11

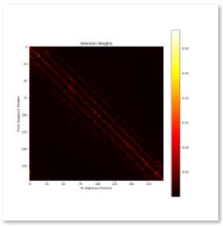

raw_attn_weights_layer_0

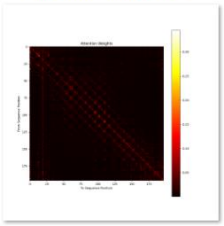

raw_attn_weights_layer_1

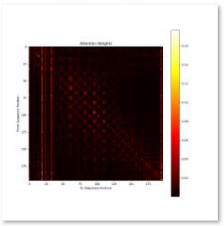

raw_attn_weights_layer_2

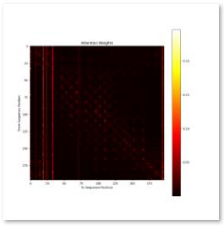

raw_attn_weights_layer_3

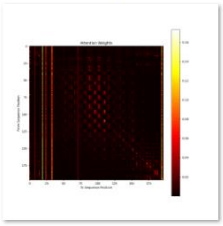

raw_attn_weights_layer_4

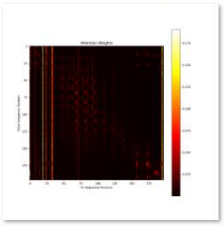

raw_attn_weights_layer_5

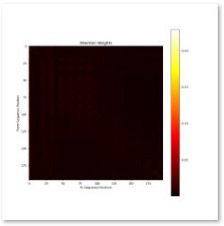

raw_attn_weights_layer_6

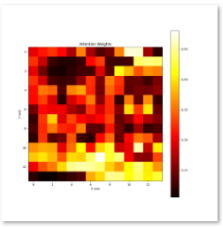

raw_attn_weights_layer_7

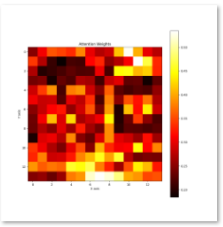

raw_attn_weights_layer_8

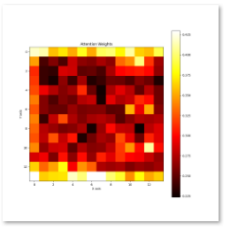

raw_attn_weights_layer_9

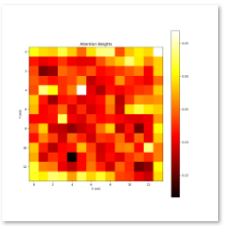

raw_attn_weights_layer_10

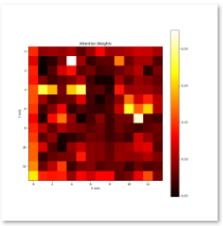

raw_attn_weights_layer_11

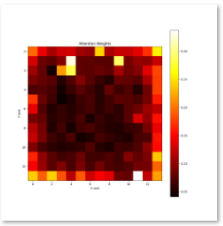

raw_attn_layer_0

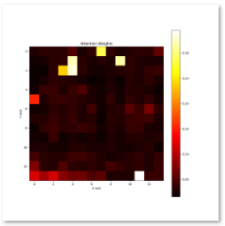

raw_attn_layer_1

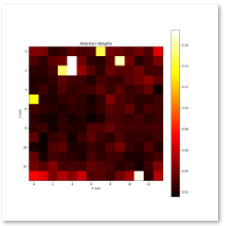

raw_attn_layer_2

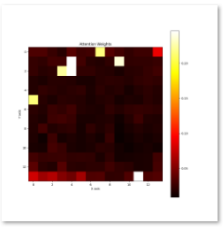

raw_attn_layer_3

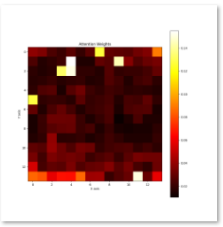

raw_attn_layer_4

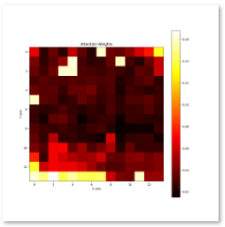

raw_attn_layer_5

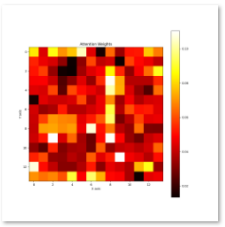

raw_attn_layer_6

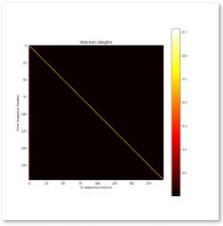

raw_attn_layer_7

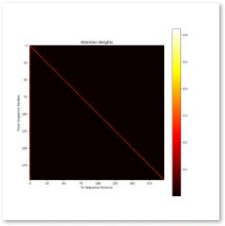

raw_attn_layer_8

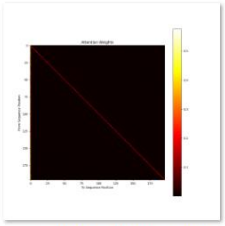

raw_attn_layer_9

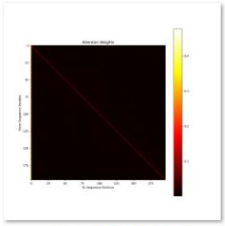

raw_attn_layer_10

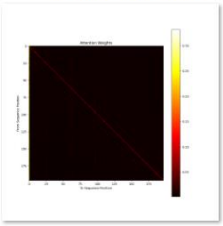

raw_attn_layer_11

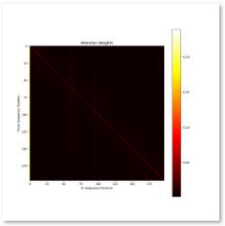

rollout_attn_weights_layer_0

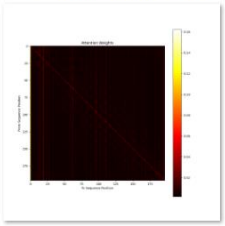

rollout_attn_weights_layer_1

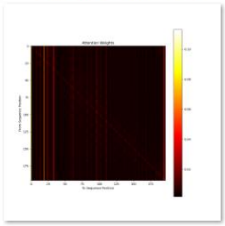

rollout_attn_weights_layer_2

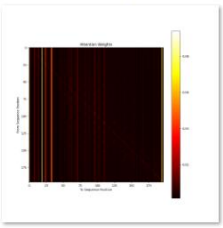

rollout_attn_weights_layer_3

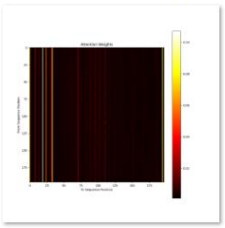

rollout_attn_weights_layer_4

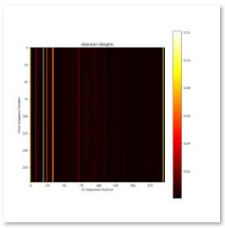

rollout_attn_weights_layer_5

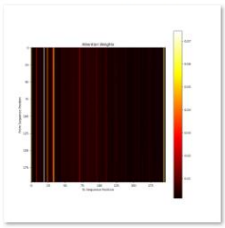

rollout_attn_weights_layer_6

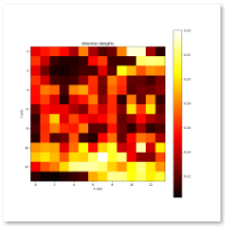

rollout_attn_weights_layer_7

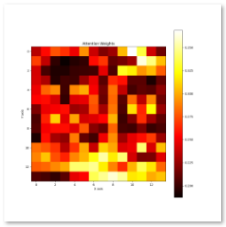

rollout_attn_weights_layer_8

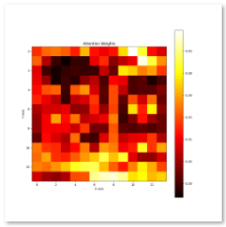

rollout_attn_weights_layer_9

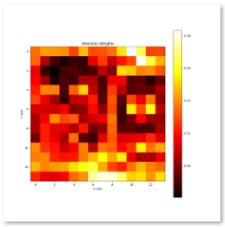

rollout_attn_weights_layer_10

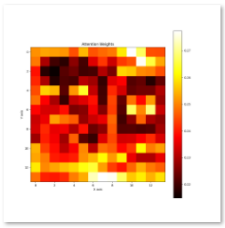

rollout_attn_weights_layer_11

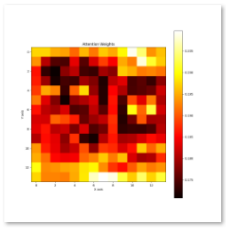

rollout_attn_layer_0

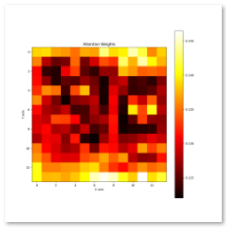

rollout_attn_layer_1

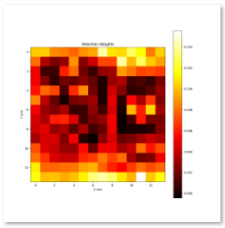

rollout_attn_layer_2

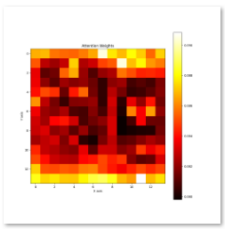

rollout_attn_layer_3

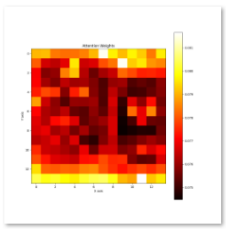

rollout_attn_layer_4

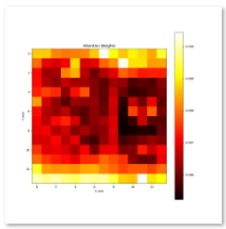

rollout_attn_layer_5

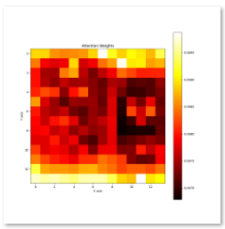

rollout_attn_layer_6

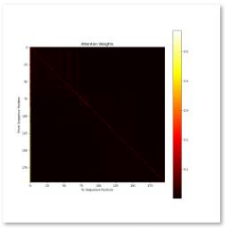

rollout_attn_layer_7

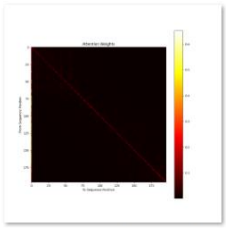

rollout_attn_layer_8

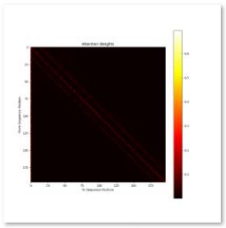

rollout_attn_layer_9

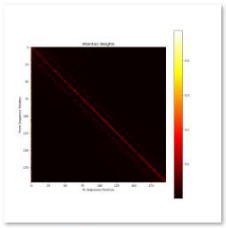

rollout_attn_layer_10

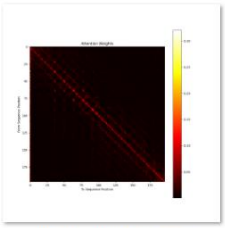

rollout_attn_layer_11

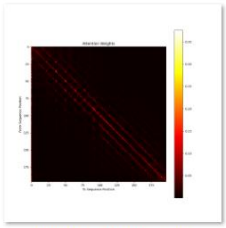

raw_attn_weights_layer_0

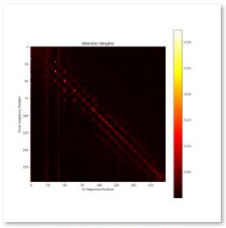

raw_attn_weights_layer_1

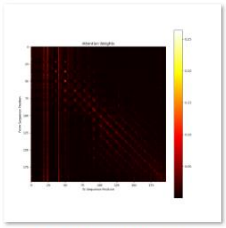

raw_attn_weights_layer_2

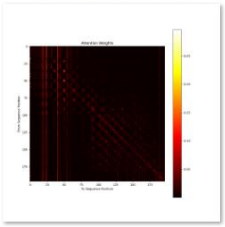

raw_attn_weights_layer_3

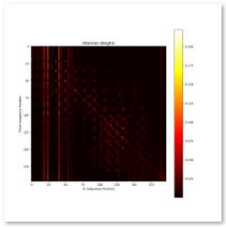

raw_attn_weights_layer_4

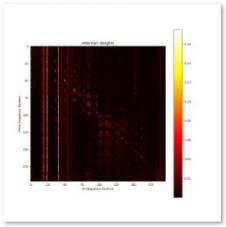

raw_attn_weights_layer_5

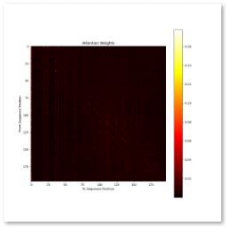

raw_attn_weights_layer_6

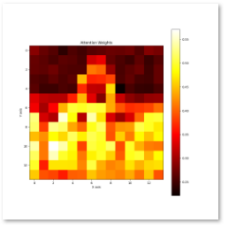

raw_attn_weights_layer_7

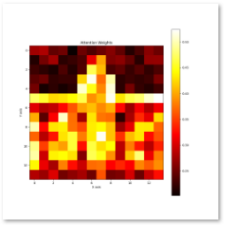

raw_attn_weights_layer_8

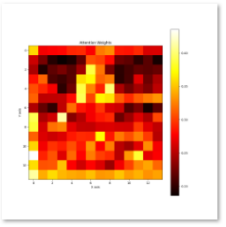

raw_attn_weights_layer_9

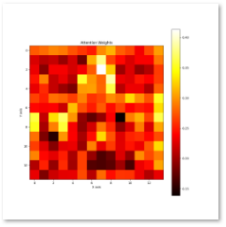

raw_attn_weights_layer_10

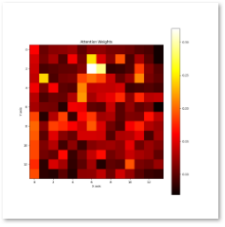

raw_attn_weights_layer_11

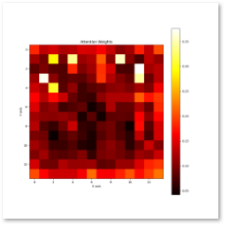

raw_attn_layer_0

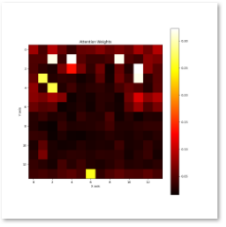

raw_attn_layer_1

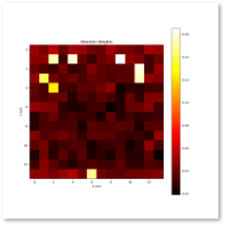

raw_attn_layer_2

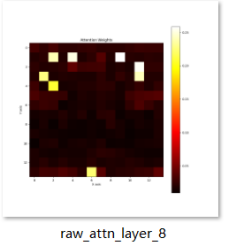

raw_attn_layer_3

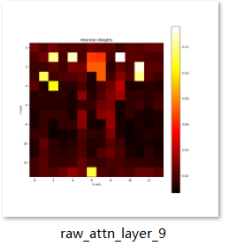

raw_attn_layer_4

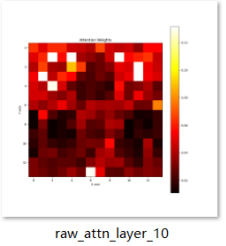

raw_attn_layer_5

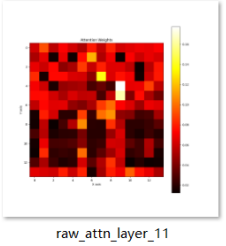

raw_attn_layer_6

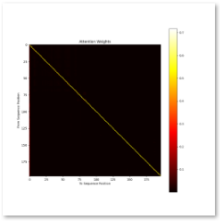

raw_attn_layer_7

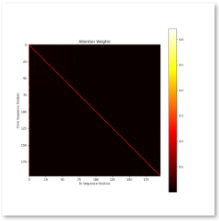

raw_attn_layer_8

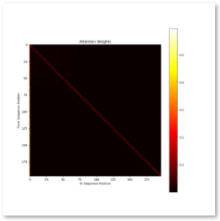

raw_attn_layer_9

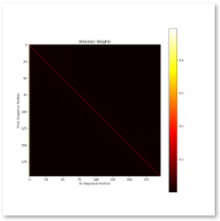

raw_attn_layer_10

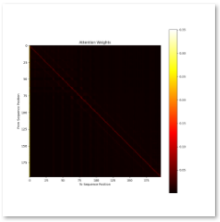

raw_attn_layer_11

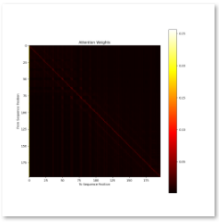

rollout_attn_weights_layer_0

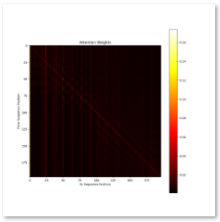

rollout_attn_weights_layer_1

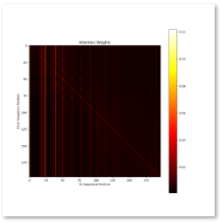

rollout_attn_weights_layer_2

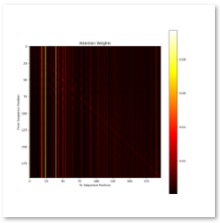

rollout_attn_weights_layer_3

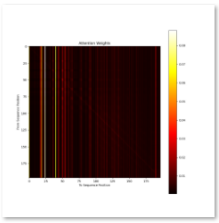

rollout_attn_weights_layer_4

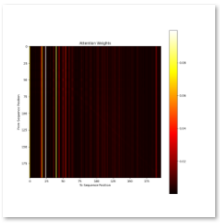

rollout_attn_weights_layer_5

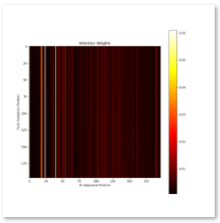

rollout_attn_weights_layer_6

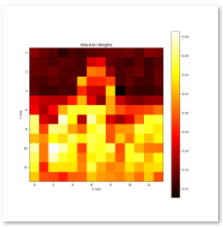

rollout_attn_weights_layer_7

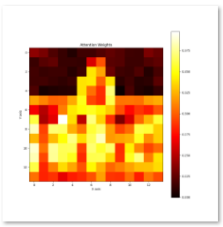

rollout_attn_weights_layer_8

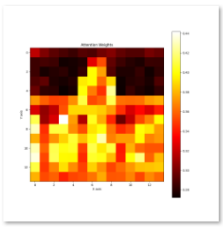

rollout_attn_weights_layer_9

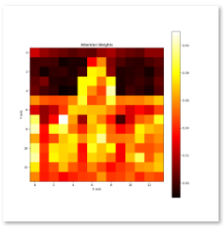

rollout_attn_weights_layer_10

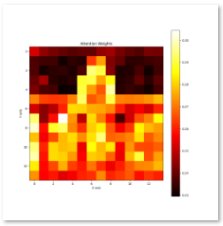

rollout_attn_weights_layer_11

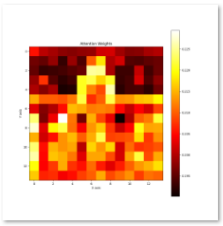

rollout_attn_layer_0

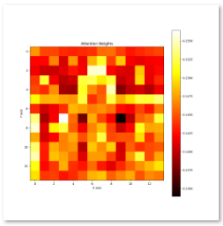

rollout_attn_layer_1

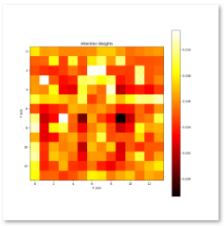

rollout_attn_layer_2

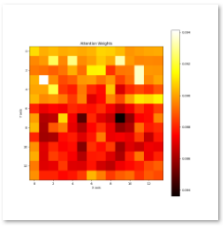

rollout_attn_layer_3

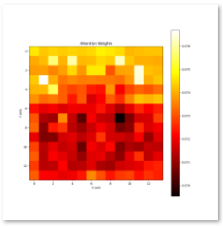

rollout_attn_layer_4

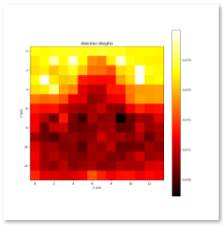

rollout_attn_layer_5

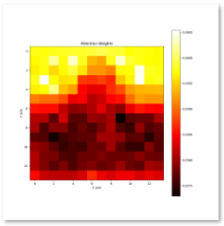

rollout_attn_layer_6

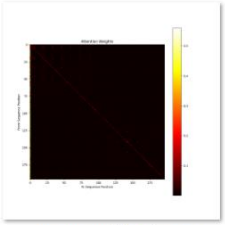

rollout_attn_layer_7

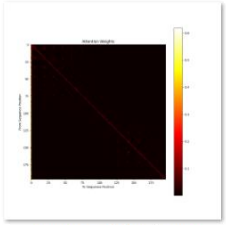

rollout_attn_layer_8

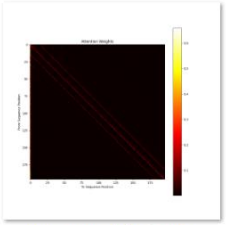

rollout_attn_layer_9

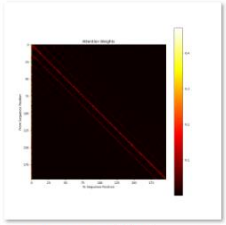

rollout_attn_layer_10

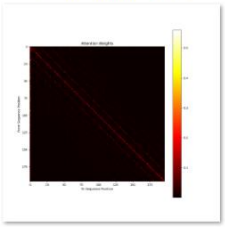

rollout_attn_layer_11

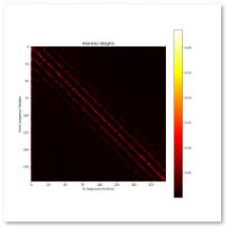

raw_attn_weights_layer_0

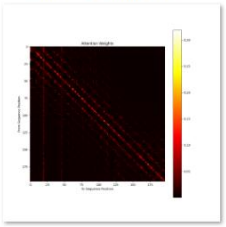

raw_attn_weights_layer_1

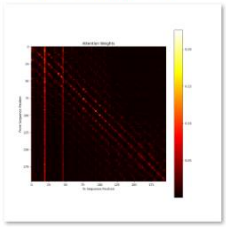

raw_attn_weights_layer_2

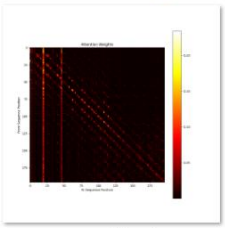

raw_attn_weights_layer_3

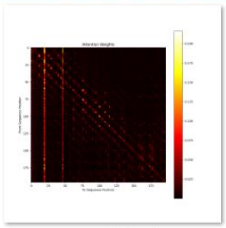

raw_attn_weights_layer_4

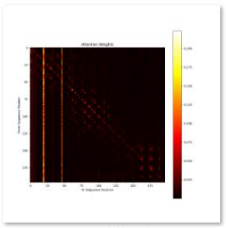

raw_attn_weights_layer_5

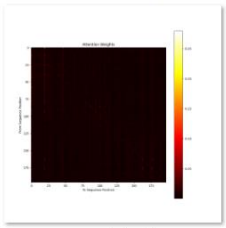

raw_attn_weights_layer_6

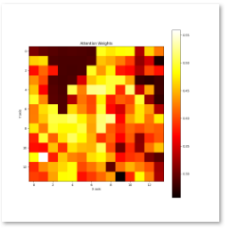

raw_attn_weights_layer_7

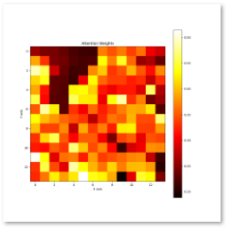

raw_attn_weights_layer_8

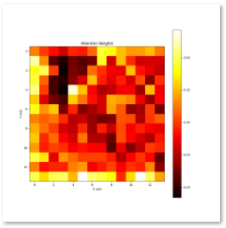

raw_attn_weights_layer_9

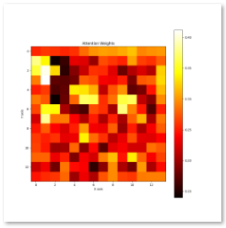

raw_attn_weights_layer_10

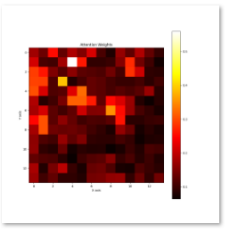

raw_attn_weights_layer_11

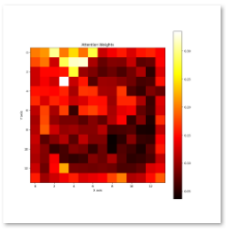

raw_attn_layer_0

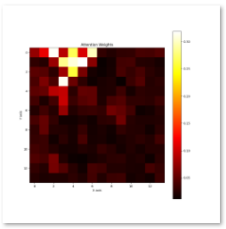

raw_attn_layer_1

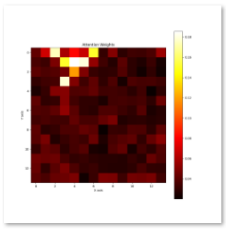

raw_attn_layer_2

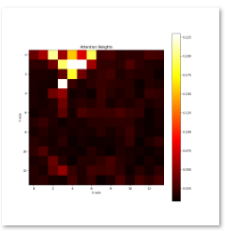

raw_attn_layer_3

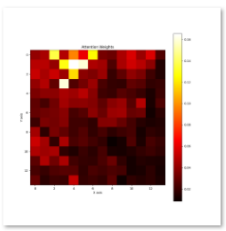

raw_attn_layer_4

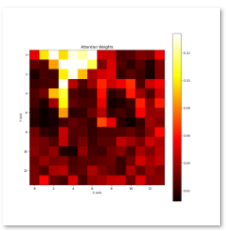

raw_attn_layer_5

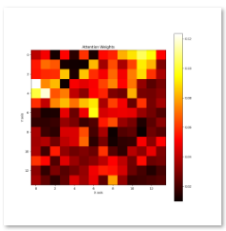

raw_attn_layer_6

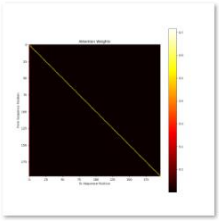

raw_attn_layer_7

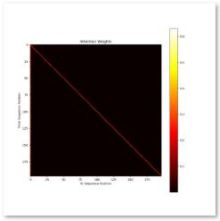

raw_attn_layer_8

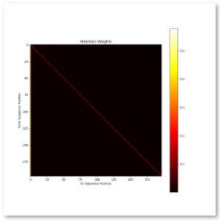

raw_attn_layer_9

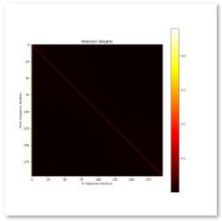

raw_attn_layer_10

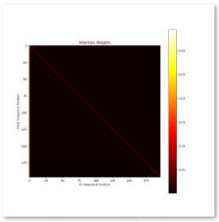

raw_attn_layer_11

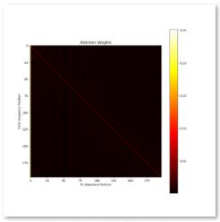

rollout_attn_weights_layer_0

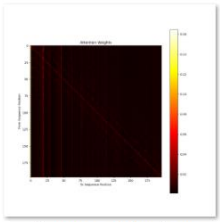

rollout_attn_weights_layer_1

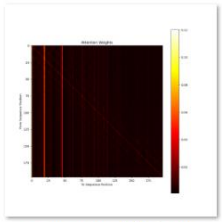

rollout_attn_weights_layer_2

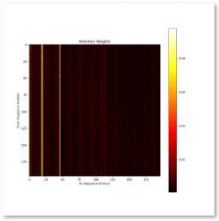

rollout_attn_weights_layer_3

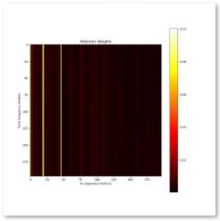

rollout_attn_weights_layer_4

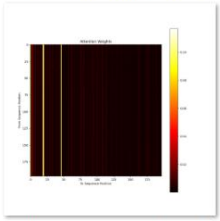

rollout_attn_weights_layer_5

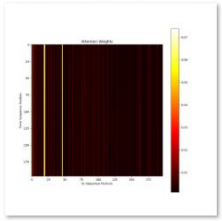

rollout_attn_weights_layer_6

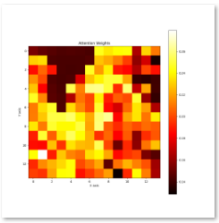

rollout_attn_weights_layer_7

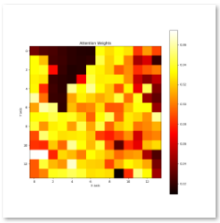

rollout_attn_weights_layer_8

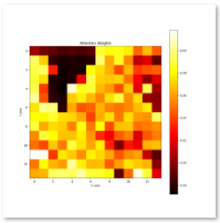

rollout_attn_weights_layer_9

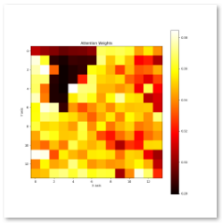

rollout_attn_weights_layer_10

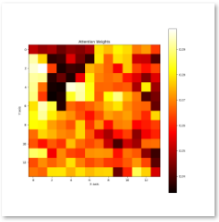

rollout_attn_weights_layer_11

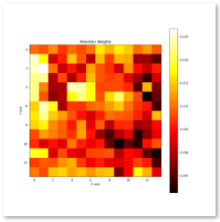

rollout_attn_layer_0

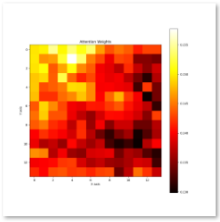

rollout_attn_layer_1

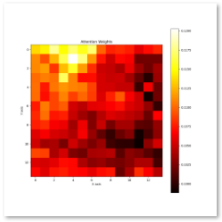

rollout_attn_layer_2

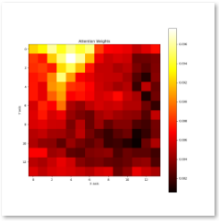

rollout_attn_layer_3

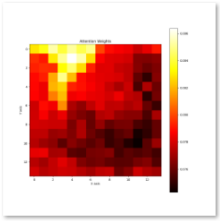

rollout_attn_layer_4

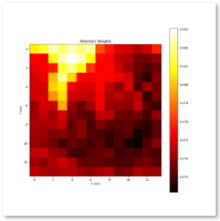

rollout_attn_layer_5

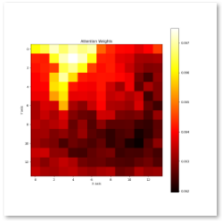

rollout_attn_layer_6


rollout_attn_layer_7


rollout_attn_layer_8


rollout_attn_layer_9


rollout_attn_layer_10


rollout_attn_layer_11