



Here's a pseudocode representation of the process described:

Extract patch tokens from the frames

$Xp1 = \text{extract_patch_tokens}(\text{Frame1})$

$Xp2 = \text{extract_patch_tokens}(\text{Frame2})$

Step 1: Patch Token Matching

Use Hungarian algorithm to find the optimal correspondence between patches

$M_hat2 = \text{hungarian_matching}(Xp1, Xp2)$

Step 2: Align Activation Regions

Initialize Frame 2's activation map

$A_2 = \text{init_activation_map}(\text{Frame2})$

Apply the matching sequence M_hat2 to Frame 2's activation map

$A'_2 = \text{align_activation_map}(A_2, M_hat2)$

Return the aligned activation map for Frame 2

$\text{return } A'_2$