



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

$\widehat{M}_2 = \text{hungarian\_matching}(Xp1, Xp2)$

# Step 2: Align Activation Regions

# Initialize Frame 2's activation map

$A_2 = \text{init\_activation\_map}(\text{Frame2})$

# Apply the matching sequence  $\widehat{M}_2$  to Frame 2's activation map

$A'_2 = \text{align\_activation\_map}(A_2, \widehat{M}_2)$

# Return the aligned activation map for Frame 2

return  $A'_2$