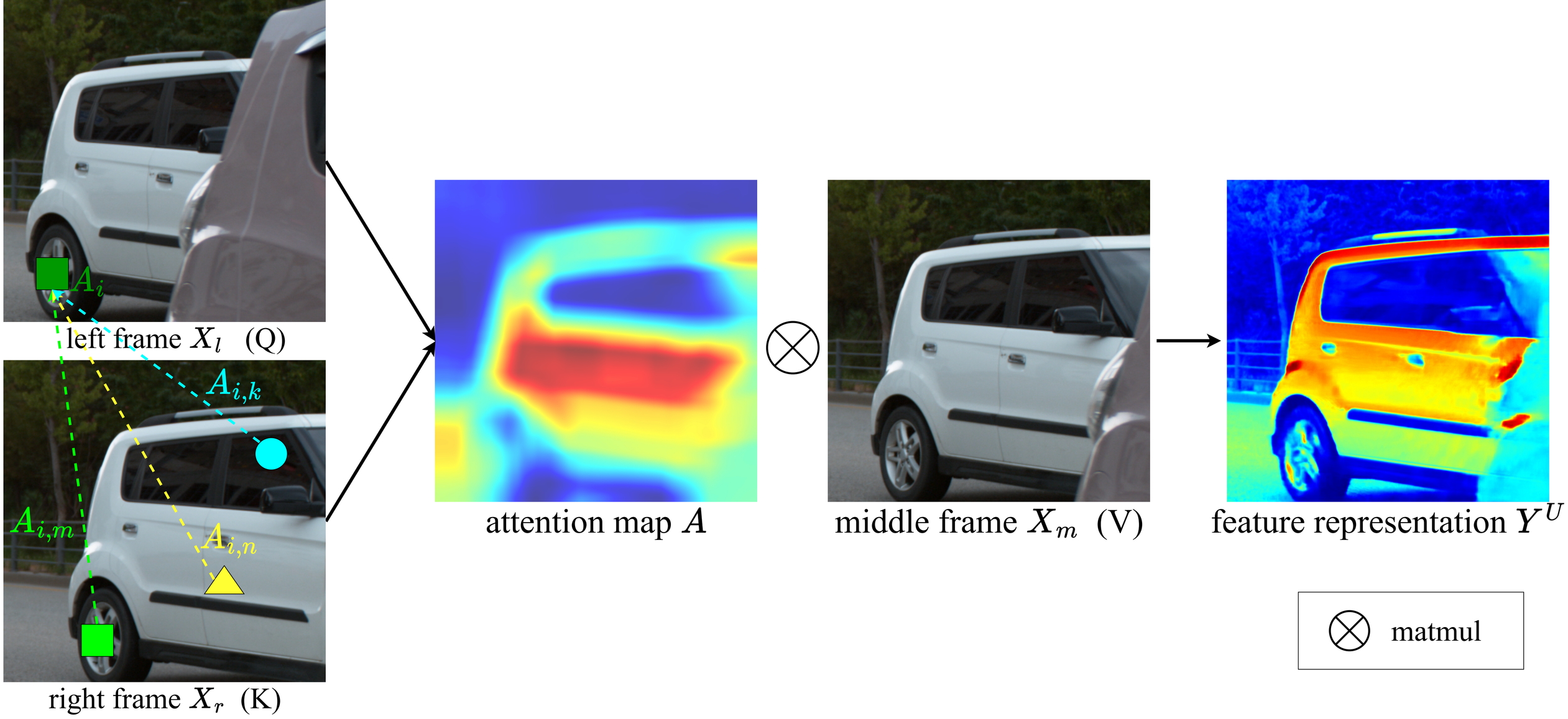
A_i
left frame X_l (Q)
$A_{i,k}$
$A_{i,m}$
$A_{i,n}$
right frame X_r (K)
attention map A
middle frame X_m (V)
feature representation Y^U
matmul