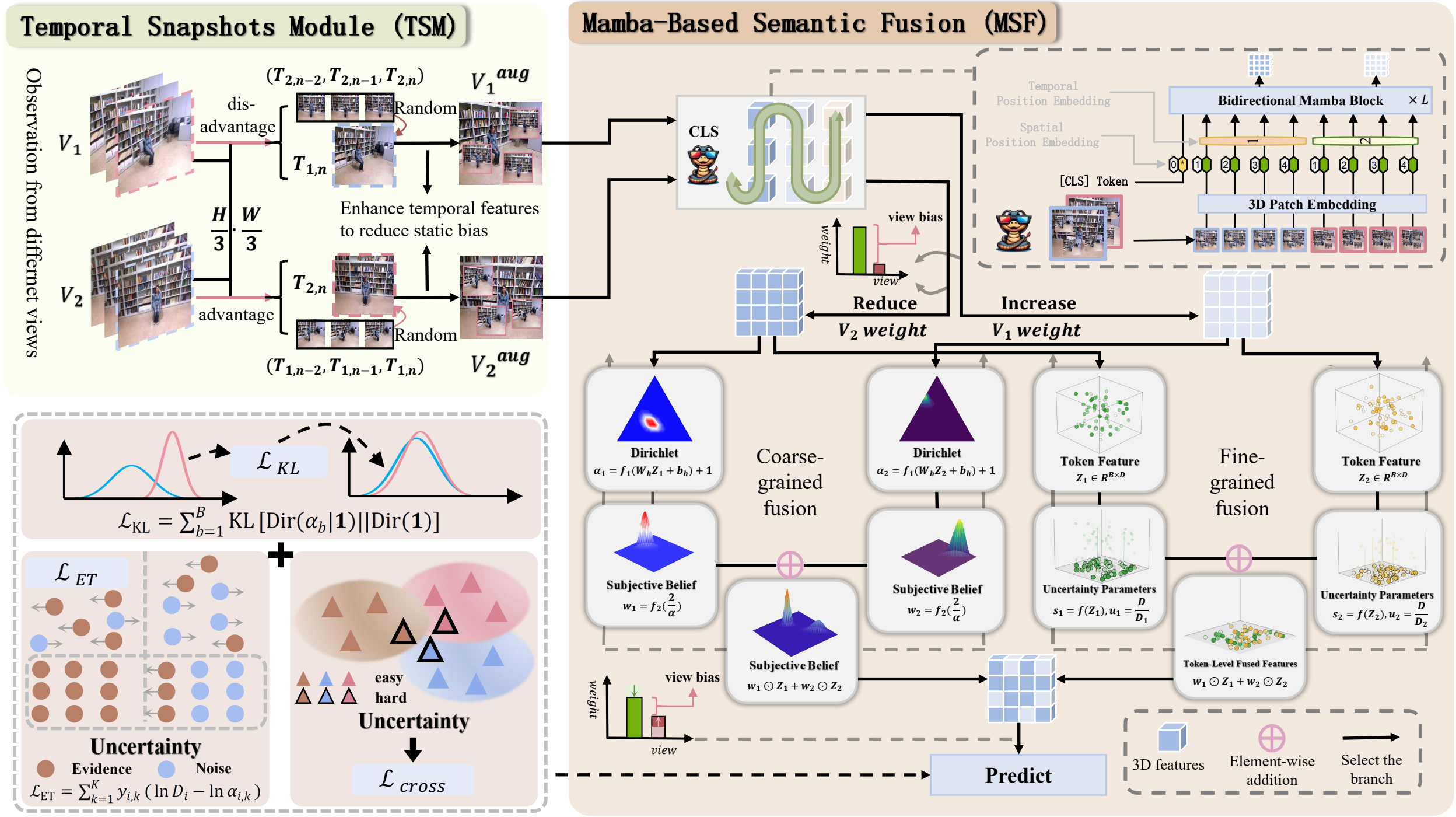
Temporal Snapshots Module (TSM)
Observation from different views
V_1
V_2
dis-advantage
advantage
$\frac{H}{3} \cdot \frac{W}{3}$
$(T_{2,n-2}, T_{2,n-1}, T_{2,n})$
$T_{1,n}$
Random
V_1^{aug}
Enhance temporal features to reduce static bias
$T_{2,n}$
Random
$(T_{1,n-2}, T_{1,n-1}, T_{1,n})$
V_2^{aug}
$\mathcal{L}_{KL}$
$\mathcal{L}_{\mathrm{KL}} = \sum_{b=1}^{B} \mathrm{KL}\left[\mathrm{Dir}(\alpha_b|\mathbf{1})||\mathrm{Dir}(\mathbf{1})\right]$
$\mathcal{L}_{ET}$
Uncertainty
Evidence
Noise
$\mathcal{L}_{\mathrm{ET}} = \sum_{k=1}^{K} y_{i,k}\left(\ln D_i - \ln \alpha_{i,k}\right)$
easy
hard
Uncertainty
$\mathcal{L}_{cross}$
Mamba-Based Semantic Fusion (MSF)
CLS
Temporal Position Embedding
Spatial Position Embedding
[CLS] Token
Bidirectional Mamba Block
$\times L$
3D Patch Embedding
view bias
weight
view
Reduce
V_2 weight
Increase
V_1 weight
Dirichlet
$\alpha_1 = f_1(W_h Z_1 + b_h) + 1$
Dirichlet
$\alpha_2 = f_1(W_h Z_2 + b_h) + 1$
Token Feature
$Z_1 \in R^{B\times D}$
Token Feature
$Z_2 \in R^{B\times D}$
Coarse-grained fusion
Fine-grained fusion
Subjective Belief
$w_1 = f_2(\frac{2}{\alpha})$
Subjective Belief
$w_2 = f_2(\frac{2}{\alpha})$
Uncertainty Parameters
$s_1 = f(Z_1), u_1 = \frac{D}{D_1}$
Uncertainty Parameters
$s_2 = f(Z_2), u_2 = \frac{D}{D_2}$
Subjective Belief
$w_1 \odot Z_1 + w_2 \odot Z_2$
Token-Level Fused Features
$w_1 \odot Z_1 + w_2 \odot Z_2$
Predict
3D features
Element-wise addition
Select the branch