Memory SAM: Detailed Mathematical Pipeline Flow

Complete End-to-End Pipeline with Mathematical Formulations

Input Image
$I \in \mathbb{R}^{H_0 \times W_0 \times 3}$

Feature Extraction
$F = \Phi(I) \in \mathbb{R}^{C \times H \times W}$
$X = \text{flat}(F) \in \mathbb{R}^{N \times C}$

Normalization
$\tilde{X}_i = \frac{X_i}{\|X_i\|_2 + \varepsilon}$

Memory Search
$\{I^{(m)}, M^{(m)}, F^{(m)}\}$

Sparse Matching
$S_{ij} = \tilde{X}_i \cdot \tilde{X}_j^{(m)}$

Coordinate Mapping
$(u, v) \rightarrow (x, y)$

Foreground Selection
$I_{\text{fg}} = \{i \mid M[u(i), v(i)] = 1\}$

Background Selection
$_{\text{bg}} = \{i \mid M[u(i), v(i)] = 0\}$

Matching Filtering
$\mathcal{M}_{\text{fg}} = \{(i, j) \mid S_{ij} \geq \tau_{\text{fg}}\}$

Candidate Collection
$\mathcal{P}_{\text{fg}} = \{(x_i, y_i)\}$

K-Means Clustering
$\min \sum_i \|Z_i - \mu_{c(i)}\|_2^2$

Representative Points
$\hat{\mathcal{P}}_{\text{fg}}$

Prompt Construction
$\mathcal{U} = \{(p, y_p)\}$
$y_p = 1$ **for foreground**

SAM2 Input
$g(I, \mathcal{U})$

Mask Prediction
$\{\hat{M}_\ell\}_{\ell=1}^L$

Score Calculation
$s_\ell \in [0, 1]$

Best Mask Selection
$\ell^* = \arg\max_\ell s_\ell$

Final Output
$\hat{M} = \hat{M}_{\ell^*}$

Memory Update
Store $(I, \hat{M}, F)$

Quality Assessment
IoU, Dice Score

Visualization
Overlay, Sparse Matches

Result Storage
Structured Output

Performance Metrics
Speed, Accuracy

System Optimization
Memory, Computation

Final Similarity: $\text{sim} = \frac{1}{|I_{\text{fg}}|} \sum_{i \in I_{\text{fg}}} \max_{j \in \mathcal{J}_{\text{fg}}} S_{ij}$

Complexity: $O(|\mathcal{I}||\mathcal{J}|C)$ *for matching,* $O(KTn)$ *for K-Means*