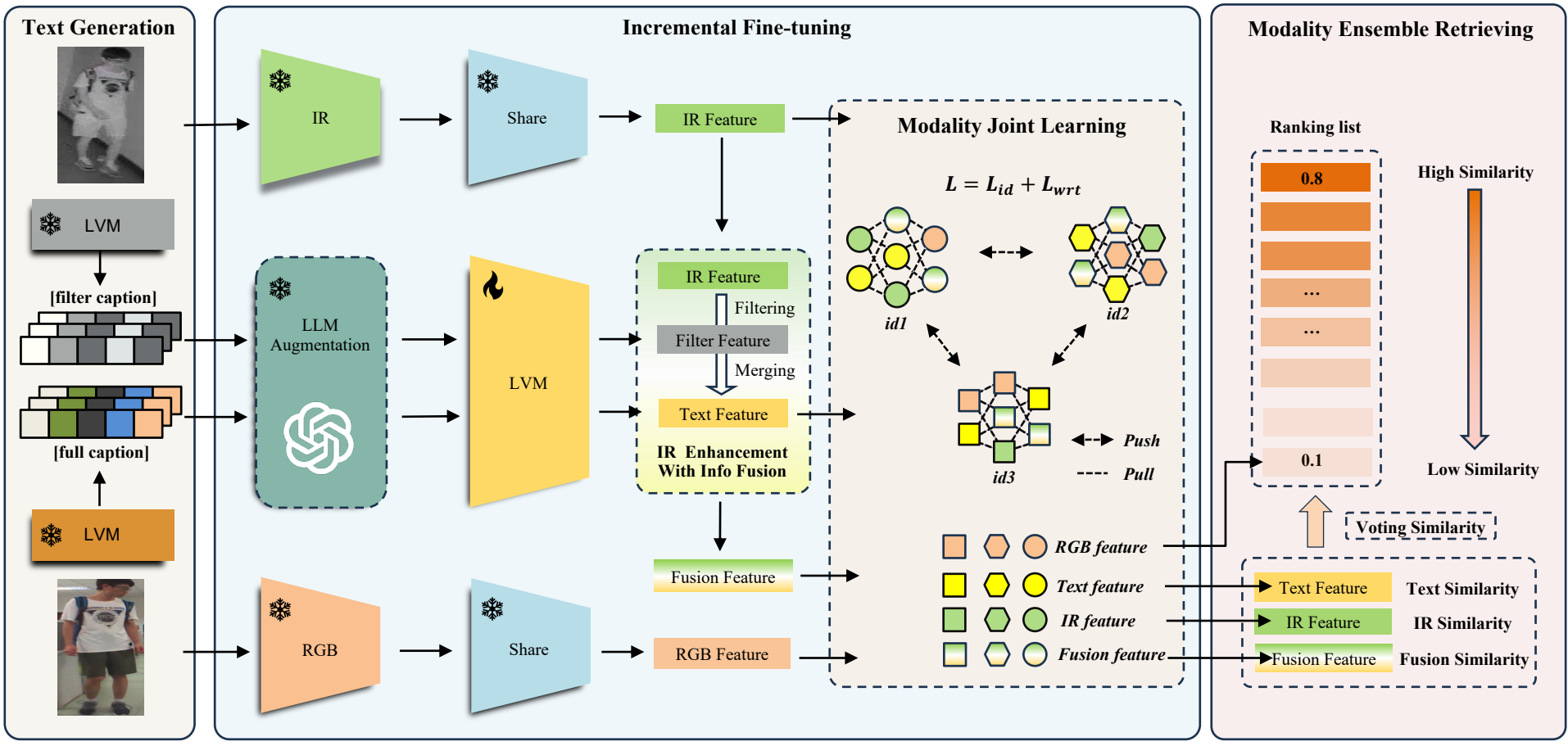
Text Generation
LVM
[filter caption]
[full caption]
LVM
Incremental Fine-tuning
IR
Share
IR Feature
LLM Augmentation
LVM
IR Feature
Filtering
Filter Feature
Merging
Text Feature
IR Enhancement With Info Fusion
Fusion Feature
RGB
Share
RGB Feature
Modality Joint Learning
$L = L_{id} + L_{wrt}$
id1
id2
id3
Push
Pull
RGB feature
Text feature
IR feature
Fusion feature
Modality Ensemble Retrieving
Ranking list
0.8
...
...
0.1
High Similarity
Low Similarity
Voting Similarity
Text Feature
Text Similarity
IR Feature
IR Similarity
Fusion Feature
Fusion Similarity