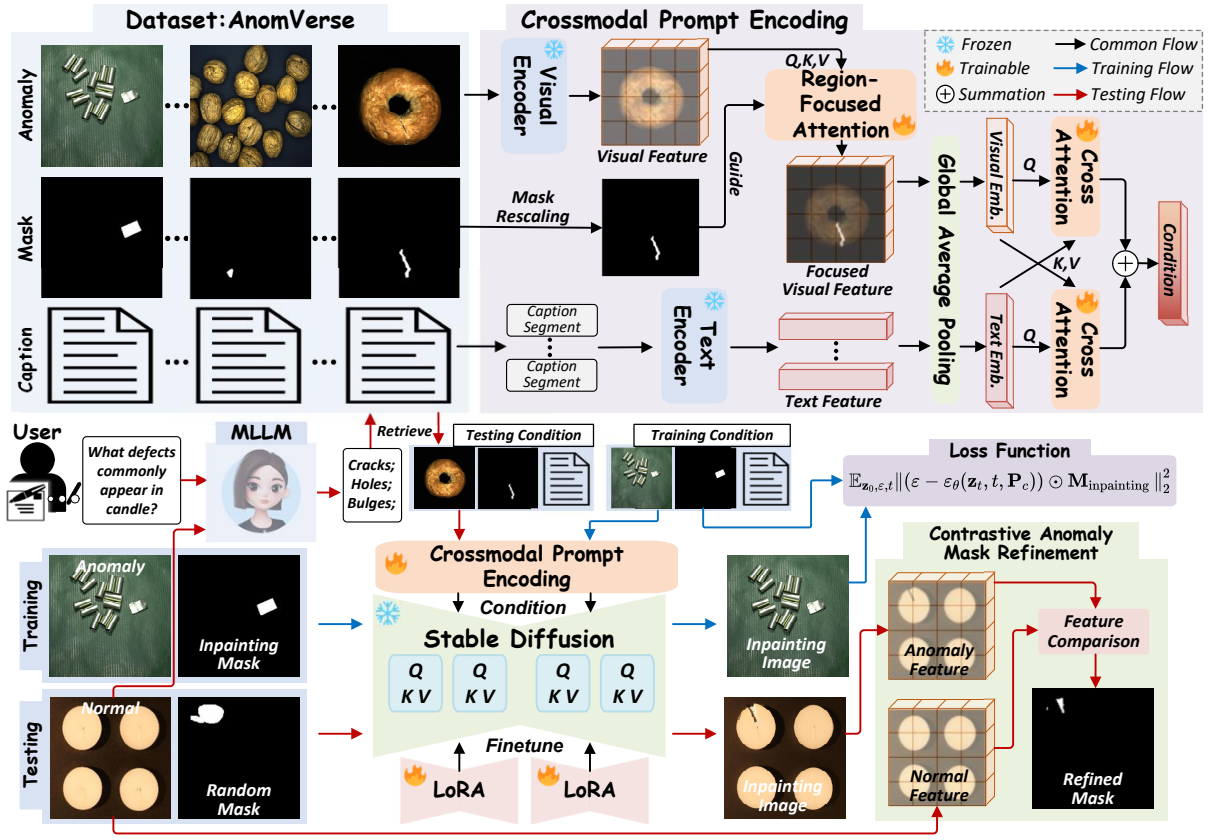
Dataset: AnomVerse
Anomaly
Mask
Caption
Crossmodal Prompt Encoding
Visual Encoder
Visual Feature
Q,K,V
Region-Focused Attention
Guide
Mask Rescaling
Focused Visual Feature
Global Average Pooling
Visual Emb.
Text Emb.
Q
K,V
Cross Attention
Cross Attention
Condition
Caption Segment
Caption Segment
Text Encoder
Text Feature
Frozen
Trainable
Summation
Common Flow
Training Flow
Testing Flow
User
What defects commonly appear in candle?
MLLM
Retrieve
Cracks; Holes; Bulges;
Testing Condition
Training Condition
Loss Function
$\mathbb{E}_{\mathbf{z}_0,\varepsilon,t}\|(\varepsilon - \varepsilon_\theta(\mathbf{z}_t, t, \mathbf{P}_c)) \odot \mathbf{M}_{\text{inpainting}}\|_2^2$
Training
Anomaly
Inpainting Mask
Testing
Normal
Random Mask
Crossmodal Prompt Encoding
Condition
Stable Diffusion
Q K V
Finetune
LoRA
LoRA
Inpainting Image
Inpainting Image
Contrastive Anomaly Mask Refinement
Anomaly Feature
Normal Feature
Feature Comparison
Refined Mask