Geometry-Driven Diffusion Model

Rendered condition

S-Encoder M-Encoder N-Encoder

Encode conditions

Text Embedding

Use pretrained LLM to extract text prompt

LLaVA

"A hand is holding dispenser on a table covered with green cloth"
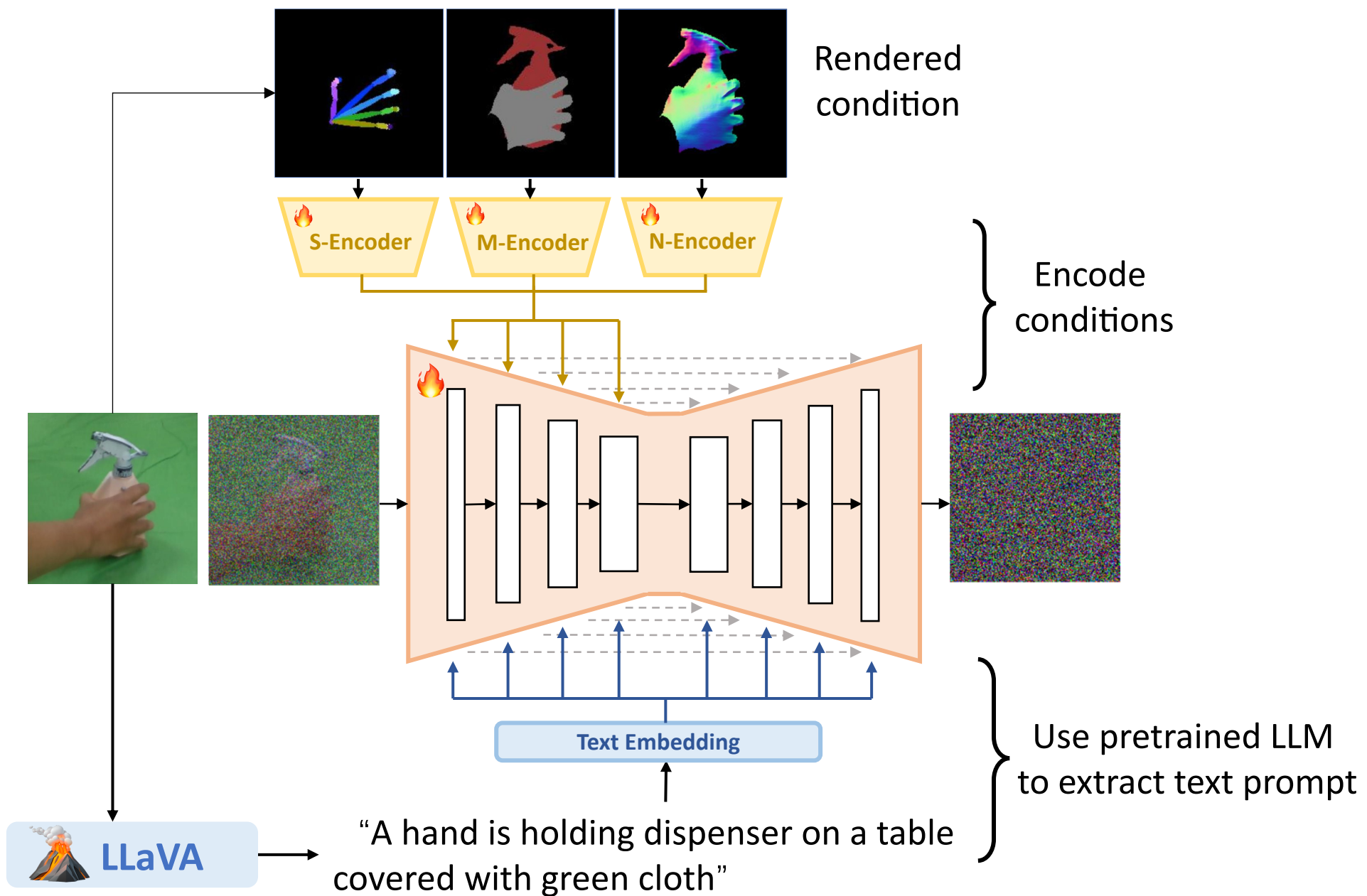
Appearance Regularization

HOI Dataset

Regularize Dataset

Pretrained Stable Diffusion

Diffusion Model

Diffusion Model

$$L = L_{HOI} + wL_{reg}$$

Prompt
"A village with a cobblestone street."
"A futuristic city on the moon."
"A pristine, frozen mountain lake."
...