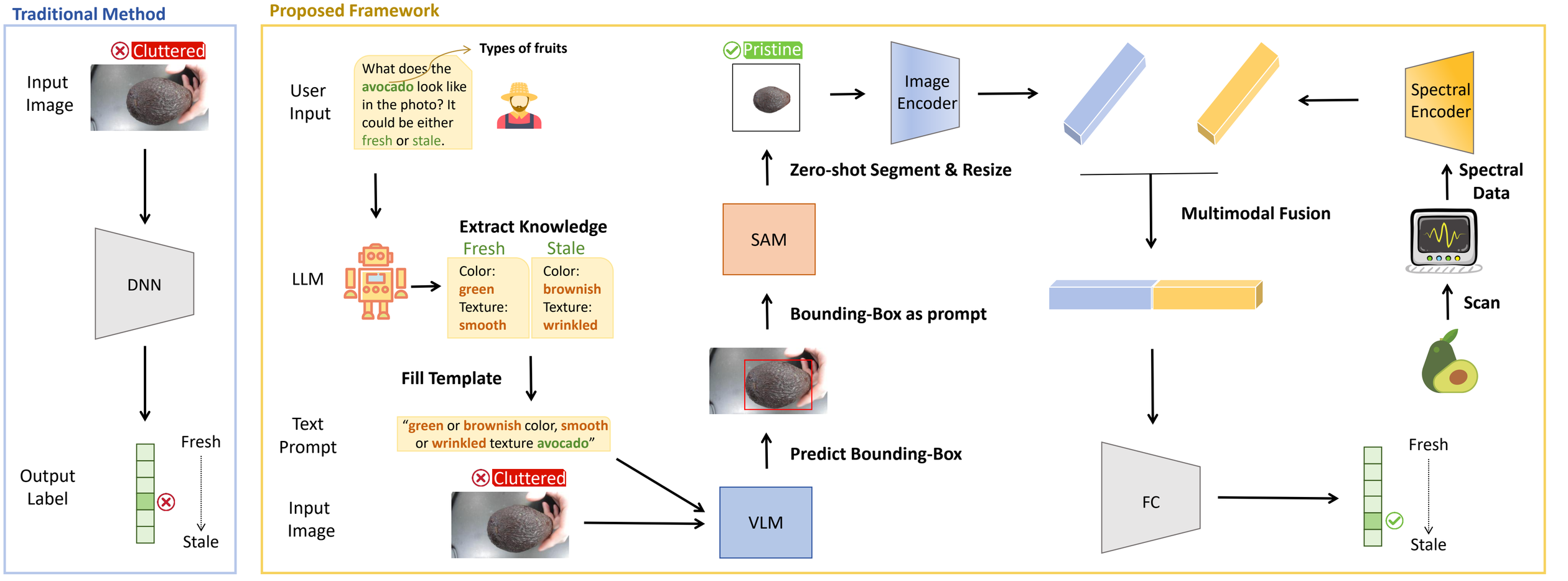
Traditional Method
Cluttered
Input Image
DNN
Output Label
Fresh
Stale
Proposed Framework
User Input
Types of fruits
What does the avocado look like in the photo? It could be either fresh or stale.
LLM
Extract Knowledge
Fresh
Color: green
Texture: smooth
Stale
Color: brownish
Texture: wrinkled
Fill Template
Text Prompt
"green or brownish color, smooth or wrinkled texture avocado"
Cluttered
Input Image
VLM
Predict Bounding-Box
Bounding-Box as prompt
SAM
Zero-shot Segment & Resize
Pristine
Image Encoder
Spectral Encoder
Spectral Data
Scan
Multimodal Fusion
FC
Fresh
Stale