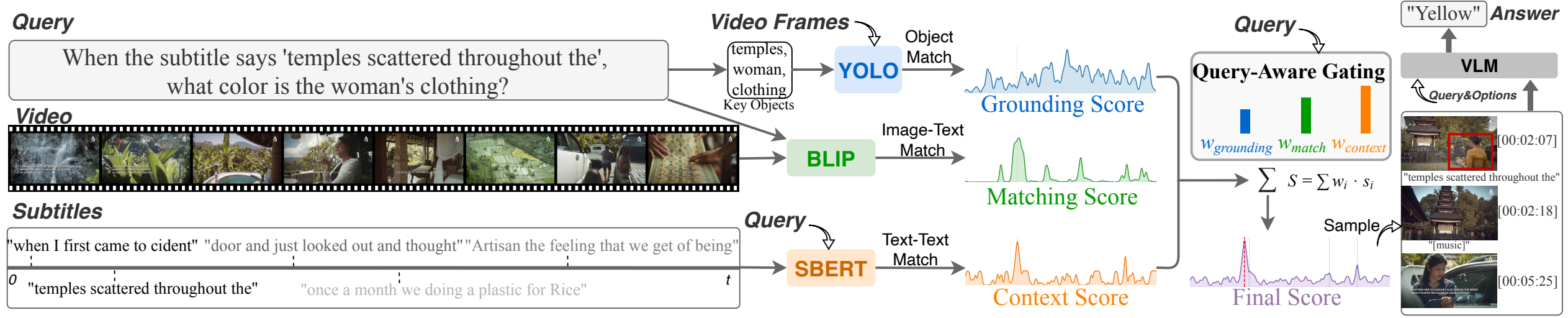

Query
When the subtitle says 'temples scattered throughout the', what color is the woman's clothing?
Video
Subtitles
"when I first came to cident"
"door and just looked out and thought"
"Artisan the feeling that we get of being"
0
"temples scattered throughout the"
"once a month we doing a plastic for Rice"
t
Video Frames
temples, woman, clothing
Key Objects
YOLO
Object Match
Grounding Score
BLIP
Image-Text Match
Matching Score
Query
SBERT
Text-Text Match
Context Score
Query
Query-Aware Gating
$w_{grounding}$ w_{match} $w_{context}$
$\sum$ $S = \sum w_i \cdot s_i$
Final Score
Sample
"Yellow"
Answer
VLM
Query&Options
[00:02:07]
"temples scattered throughout the"
[00:02:18]
"[music]"
[00:05:25]