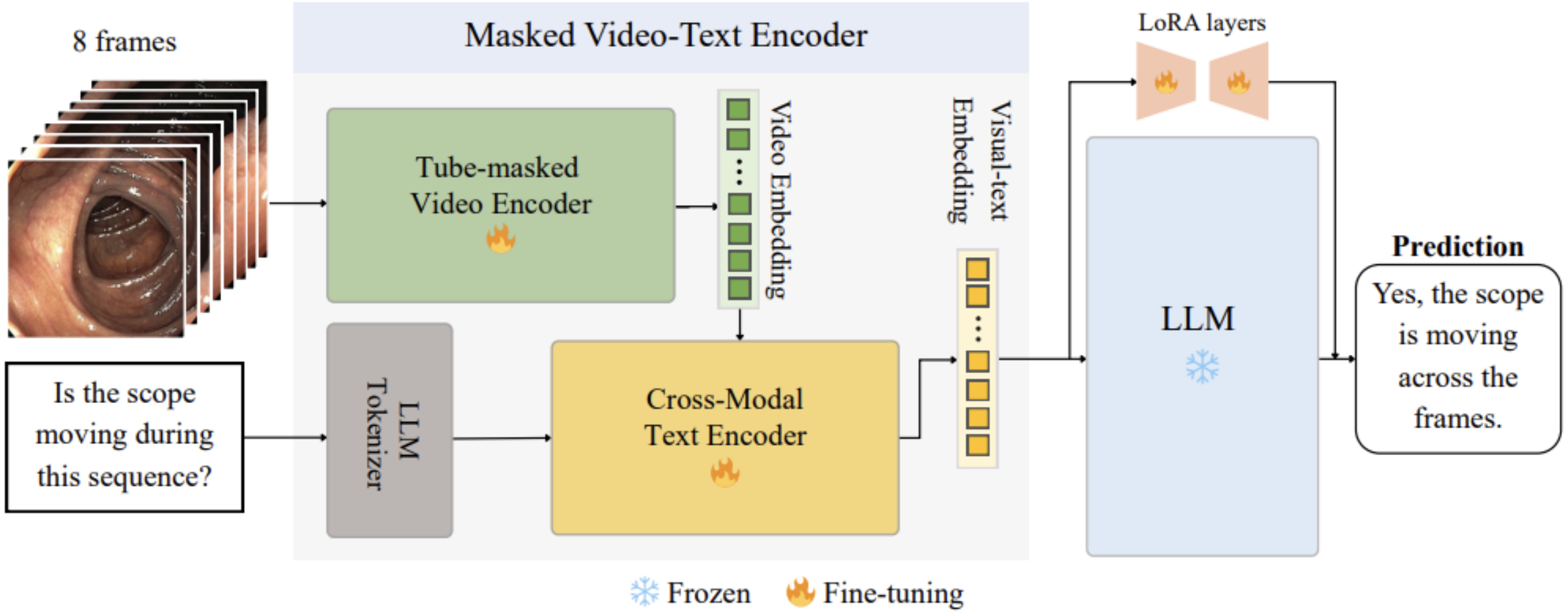
NSF REU 2026
Week 4

Tiffany Marcus

18 June 2026

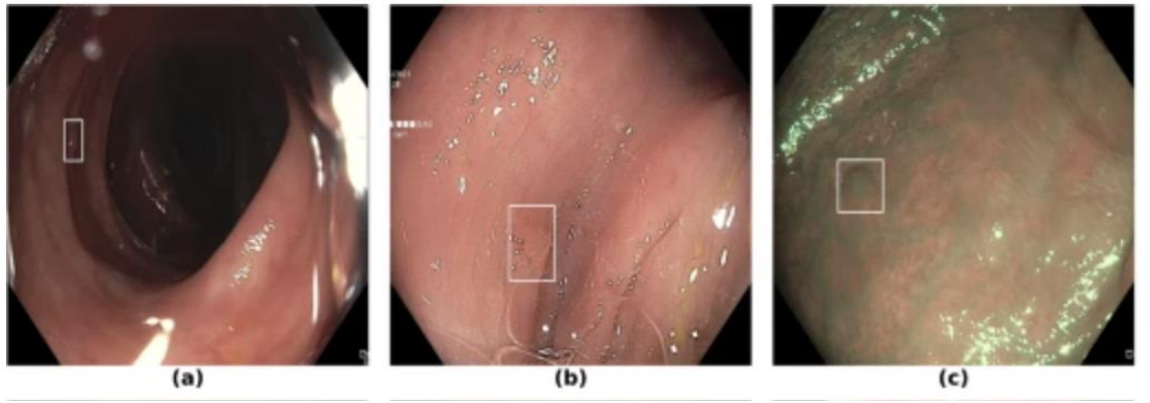
SurgViVQA



Surgvivqa: Temporally-grounded video question answering for surgical scene. (n.d.). https://arxiv.org/pdf/2511.03325

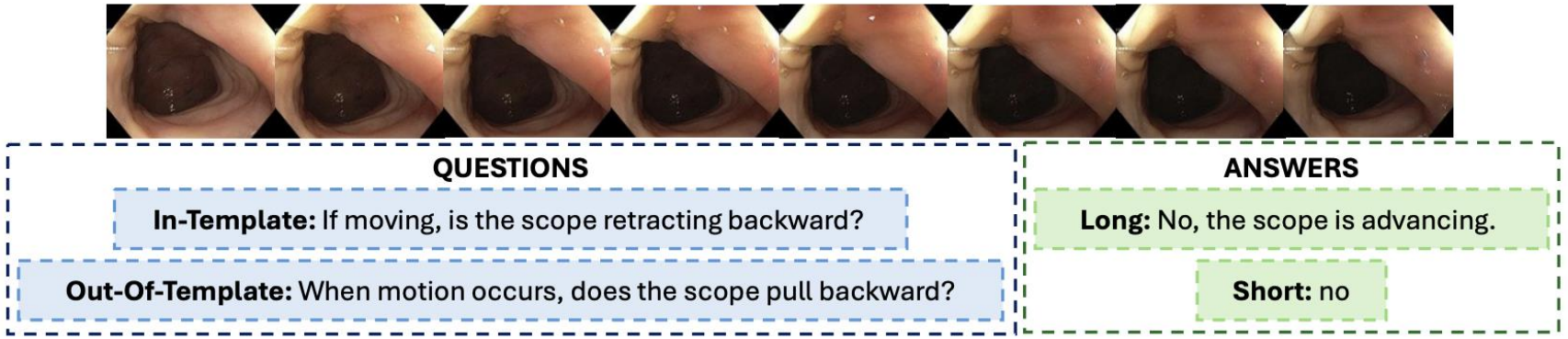
Original REAL-Colon Dataset

- Real colonoscopy procedures
- High resolution: 1920 × 1080
- 60 recordings of different colonoscopies
- >2.75 million video frames
- 30 frames/second
- Bounding boxes surrounding polyps



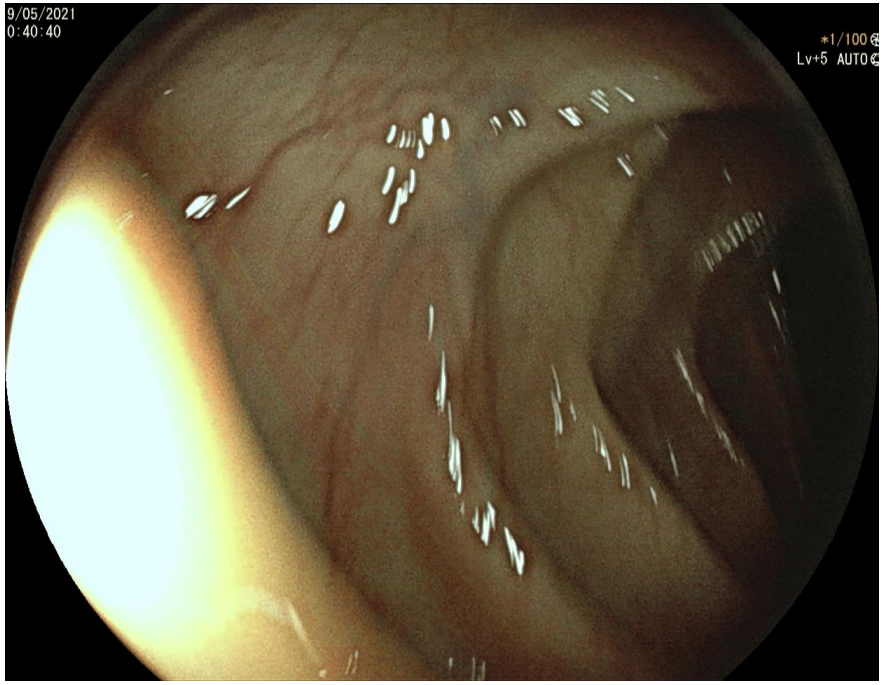
REAL-Colon-VQA Test Set

- 6 videos for testing
- Annotations extended to include:
  - Tools
  - Motion
  - Visibility
  - Occlusions
  - Flushing
  - Illumination
- ~1 frame/second
- Video clips made up of 8 frames
- QA pairs for each clip, 750 total
- In-template & out-of-template questions
- Long and short answers



Prediction Example

"Is the scope primarily advancing or withdrawing in this sequence?

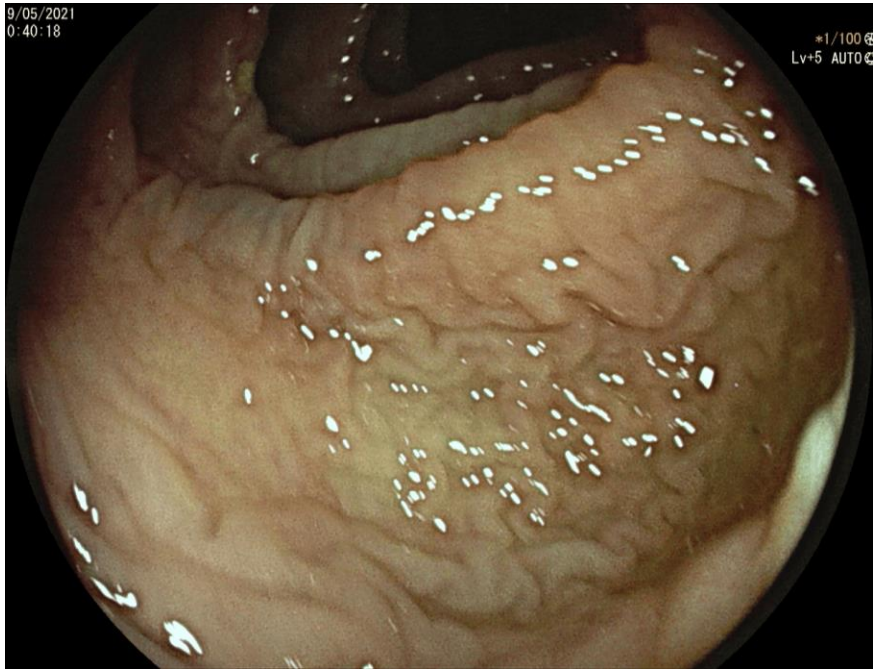


Ground truth: The scope is primarily advancing

Prediction: The scope is primarily advancing

Example of Motion Misidentification

Question: When motion occurs, is the scope pushing forward?



Ground truth: Yes, the scope is advancing
Keyword: "Yes"

Prediction: No, the scope is withdrawing

Exact Match: False
Keyword Match: False

BLEU

How much of the prediction is in the reference answer?

Reference: The scope is advancing

BLEU-1: The scope is withdrawing = ¾ = 0.75

BLEU-2: The scope is withdrawing = 2/3 = 0.667

BLEU-3: The scope is withdrawing = ½ = 0.5

BLEU-4: The scope is withdrawing = 0/1 = 0

Example of Tool Misidentification

Question: Which instrument is present in this clip?

Ground truth: Forceps are visible on screen
Keyword: "Forceps"







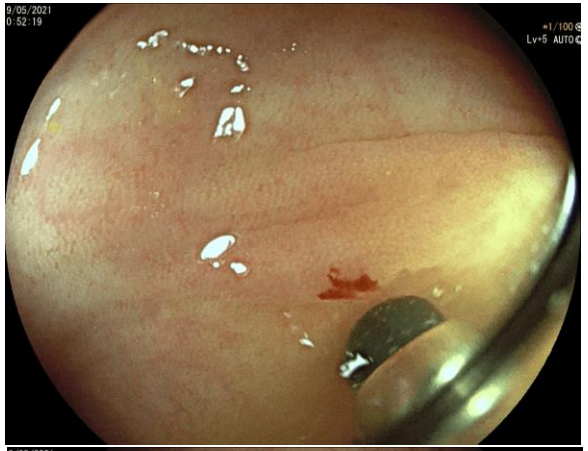
Prediction 1: Forceps are visible on screen
Exact Match: True
Keyword Match: True

Prediction 2: Snare is visible on screen
Exact Match: False
Keyword Match: False

Example of Keyword Accuracy

Question: Which instrument is present in this clip?

Ground truth: Forceps are visible on screen
Keyword: "Forceps"





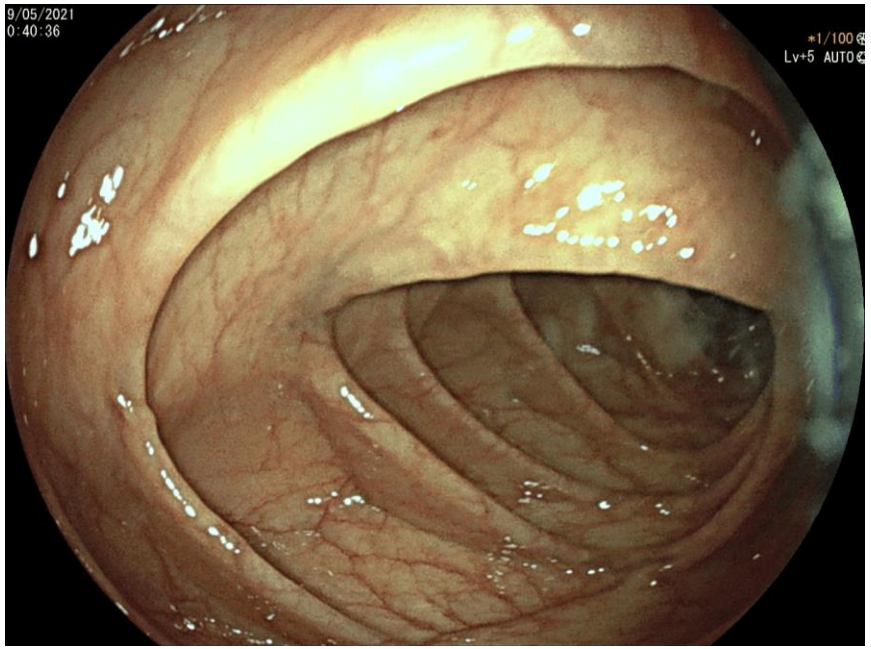


Prediction 1: **Forceps** are visible on screen
Exact Match: True
Keyword Match: True

Prediction 2: **Snare** is visible on screen
Exact Match: False
Keyword Match: False

Example of False Positive

Question: Is the scope positioned outside the patient?



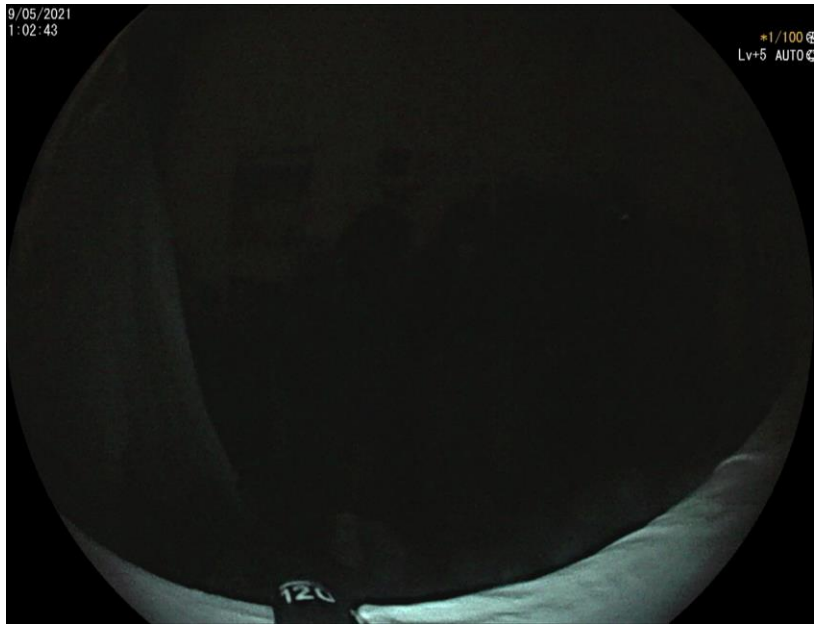
Ground truth: No, the scope stays within the patient.
Keyword: "No"

Prediction: No, the scope is not in the patient.

Exact Match: False
Keyword Match: True

Example of Semantic Difference

Question: Is the scope positioned outside the patient?



Ground truth: Yes, the scope is outside the patient.
Keyword: "Yes"

Prediction: No, the scope is not in the patient.

Exact Match: False
Keyword Match: False

ROUGE & METEOR

How much of the reference answer is in the prediction?

- ROUGE:
  - ROUGE-1: Counts number of common words
  - ROUGE-2: Counts number of common word pairing
  - ROUGE-L: Measures longest common subsequence

- METEOR: Measures text similarity

Current Results on Test Set

- Single keyword accuracy: 48.40%
- BLEU: 34.40%
- ROUGE-L: 53.64%
- METEOR: 52.69%

Next Steps

- Add a scene graph layer (e.g., PyTorch Geometric Temporal Libray)
- Expand to other datasets