NSF REU 2026
Week 7

Tiffany Marcus

9 July 2026

CholecT45 + SSG-VQA

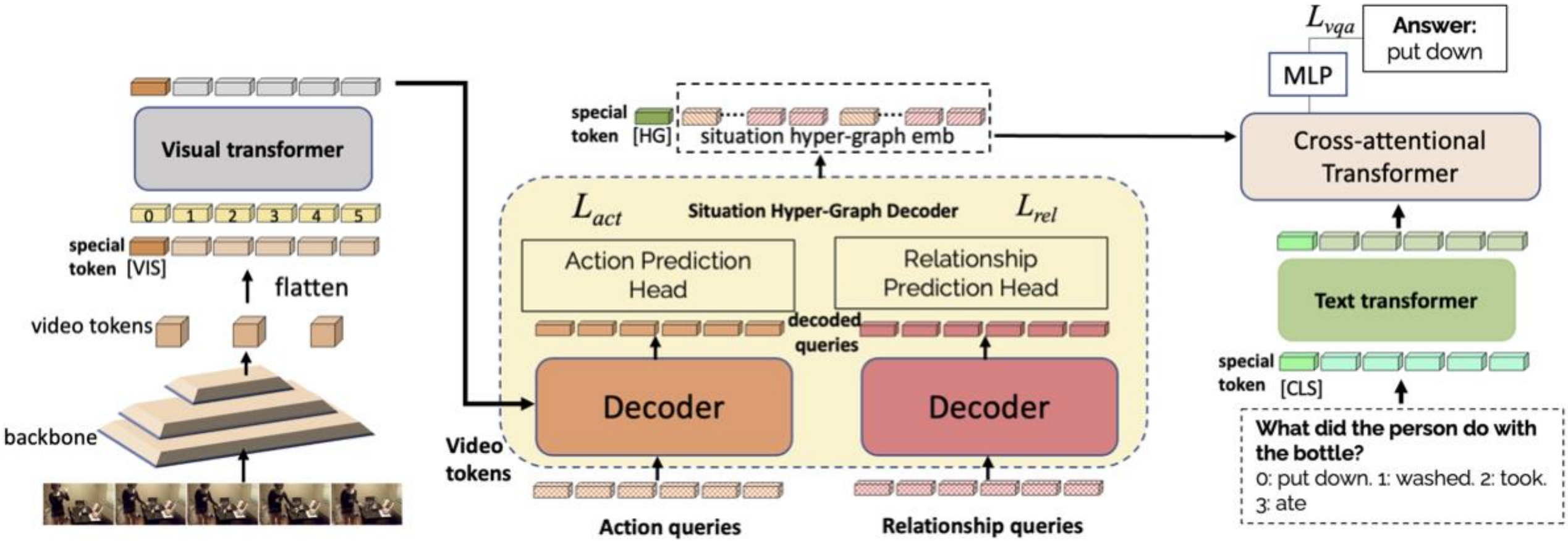
CholecT45

- **Videos**
- triplets: <instrument, verb, target>
- Action and object labels

SSG-VQA

- Scene graphs
- **Bounding boxes**
- **Action and spatial relations**
- VQA questions and answers

Hyper-graphs



Khan, A. U., Kuehne, H., Wu, B., Chheu, K., Bousselham, W., Gan, C., Lobo, N., & Shah, M. (2023, April 18). *Learning Situation Hyper-Graphs for video question answering*. arXiv.org. https://arxiv.org/abs/2304.08682

Why hyper-graphs?

- Do not require object detection as input or object level supervision
- Predicts all object-relation-object and actor-relation-object triples with associated actions for all video frames.

Hyper-graphs

Nodes

- Frame
- Object
- Action event

Edges

- Frame –contains--> object/event
- Object –spatial relation--> object
- Object –action relation--> object
- Event –uses--> object
- Frame/event x –temporal relation--> frame/event y

Hyper-graphs

Hyper-Edges

- Event
- Object nodes involved
- Time/frame occurred

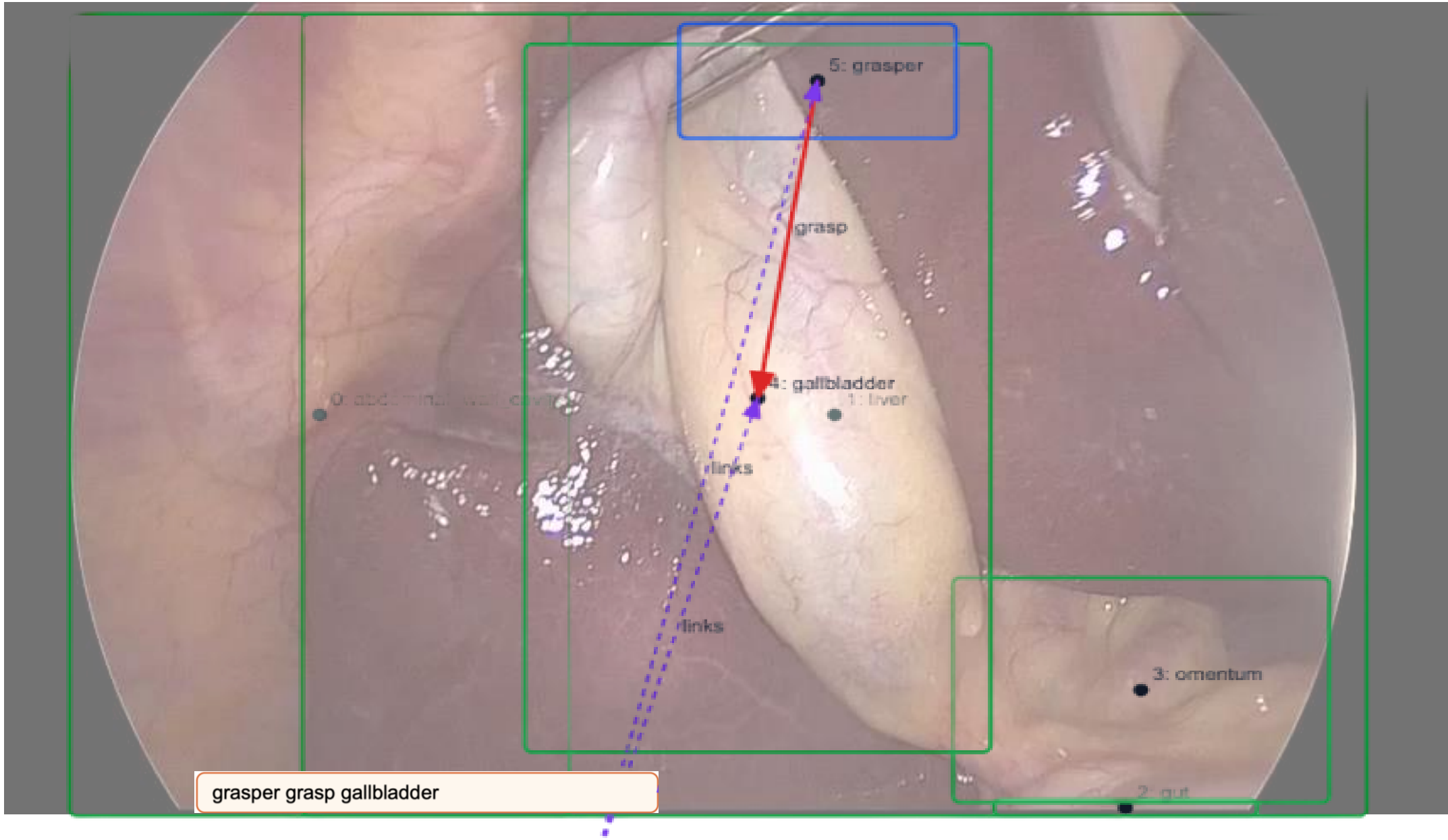
Green: Anatomy

Blue: Instrument

Orange: Action event

Red: Action relation

Purple: Hyperedge



Nodes

- - frame: VID01 frame 0
- - object: object 0: abdominal_wall_cavity (anatomy) at top-left
- - object: object 1: liver (anatomy) at top-mid
- - object: object 2: gut (anatomy) at bottom-right
- - object: object 3: omentum (anatomy) at bottom-right
- - object: object 4: gallbladder (anatomy) at top-mid
- - object: object 5: grasper (instrument) at top-mid
- - event: grasper grasp gallbladder

Semantic edges

- VID01 frame 0 --contains_object--> object 0: abdominal_wall_cavity
- VID01 frame 0 --contains_object--> object 1: liver
- VID01 frame 0 --contains_object--> object 2: gut
- VID01 frame 0 --contains_object--> object 3: omentum
- VID01 frame 0 --contains_object--> object 4: gallbladder
- VID01 frame 0 --contains_object--> object 5: grasper

- VID01 frame 0 --contains_event--> grasper grasp gallbladder
- grasper grasp gallbladder --uses_instrument--> object 5: grasper
- grasper grasp gallbladder --acts_on_target--> object 4: gallbladder

Spatial edges

- object 0: abdominal_wall_cavity --above--> object 5: grasper
- object 1: liver --above--> object 5: grasper
- object 2: gut --above--> object 0: abdominal_wall_cavity
- object 2: gut --above--> object 1: liver
- object 2: --above--> object 4: gallbladder
- object 2: gut --above--> object 5: grasper
- object 3: omentum --above--> object 0: abdominal_wall_cavity
- object 3: omentum --above--> object 1: liver
- object 3: omentum --above--> object 4: gallbladder
- object 3: omentum --above--> object 5: grasper
- object 4: gallbladder --above--> object 5: grasper

Hyperedge

Event node: grasper grasp gallbladder ->

[

action edge: grasper grasp gallbladder;

object 5: grasper; object 4: gallbladder

]

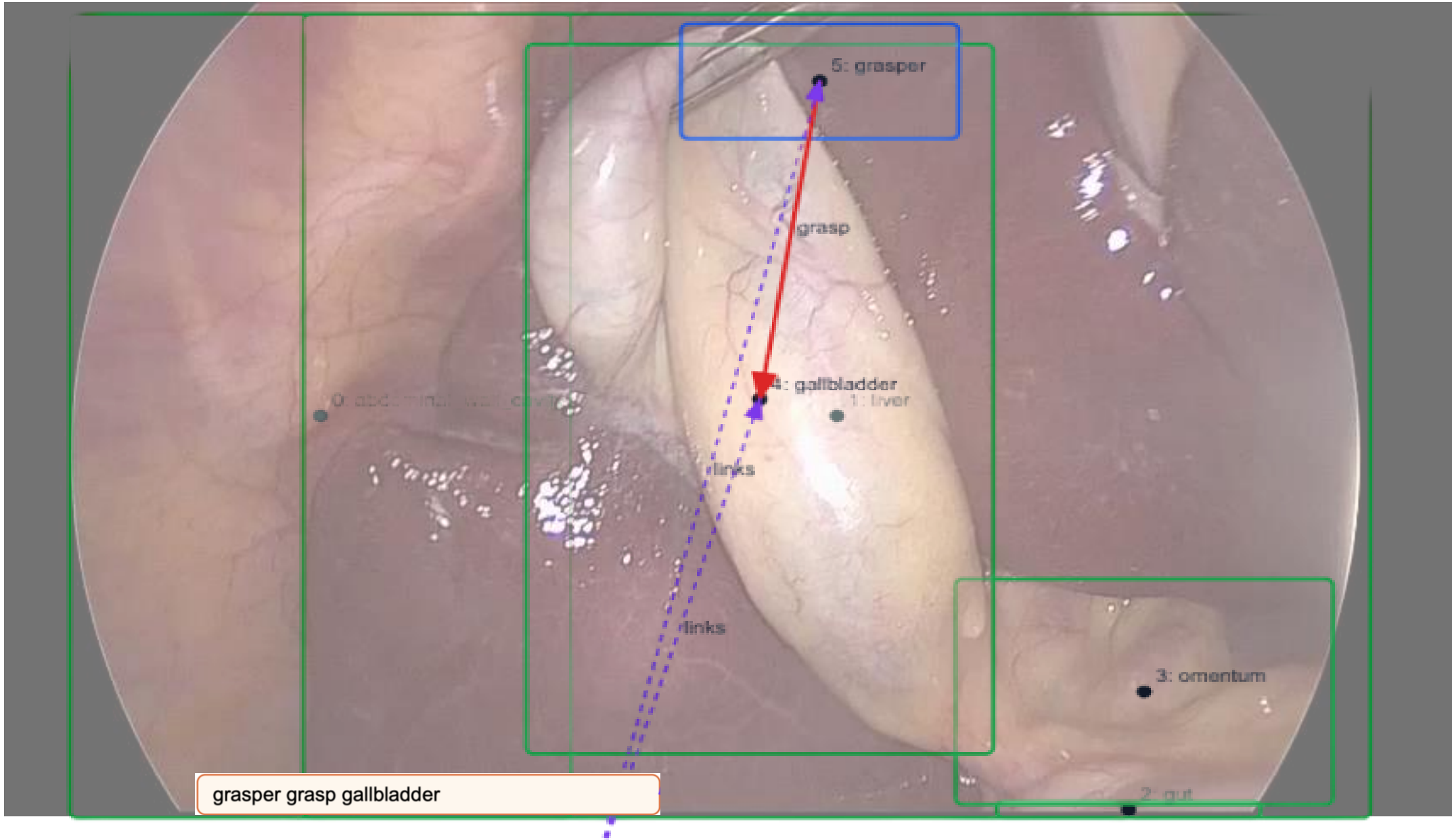
Green: Anatomy

Blue: Instrument

Orange: Action event (node)

Red: Action relation (edge)

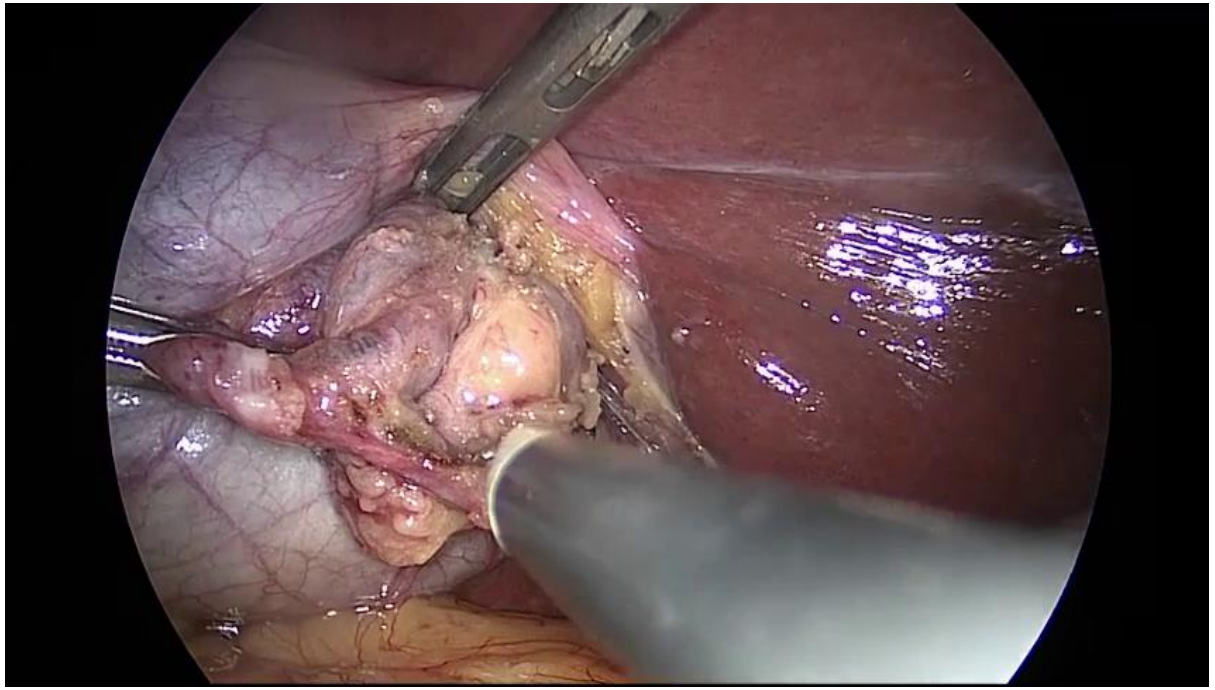
Purple: Hyperedge



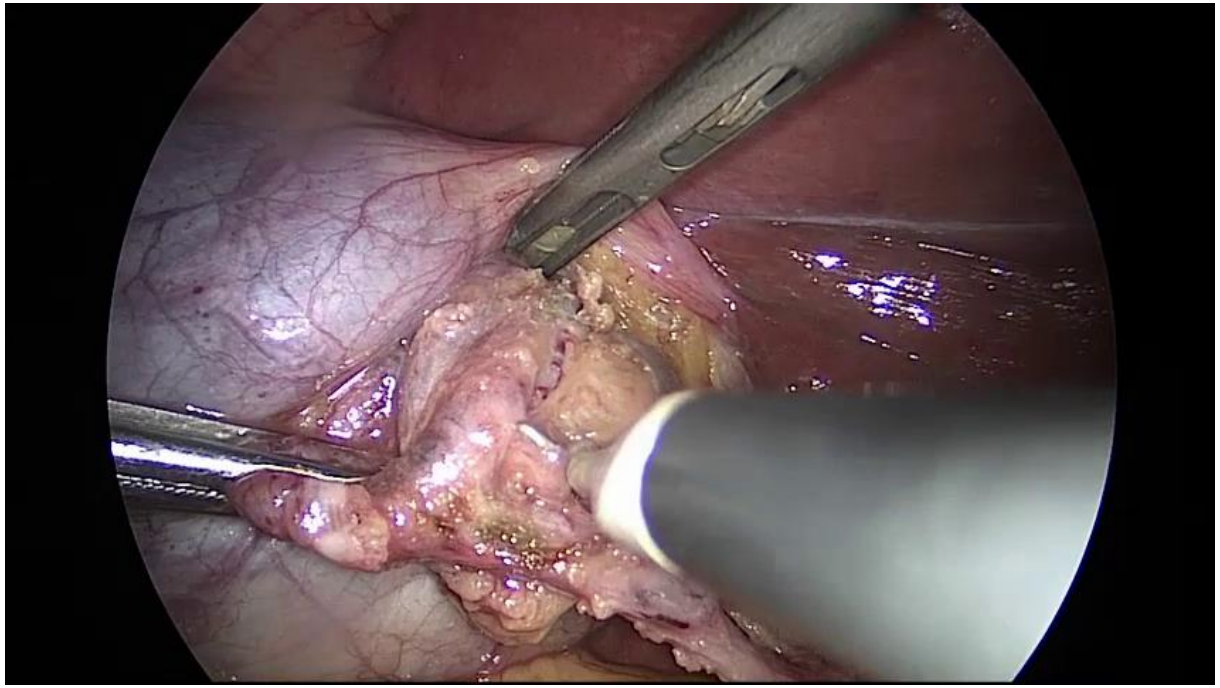
VQA example

Question: What happens to the gallbladder after it is retracted by the hook?

Answer: It is dissected by the hook



Frame 838: Hook retracts gallbladder



Frame 841: Hook dissects gallbladder

Next Steps

- Add sub-models: backbone (SlowR50 and ResNext101) and visual transformer (multilayer transformer model with self-attention heads)
- Implement the Situation Hyper-graph decoder
- Evaluate against graph only and video encoding only