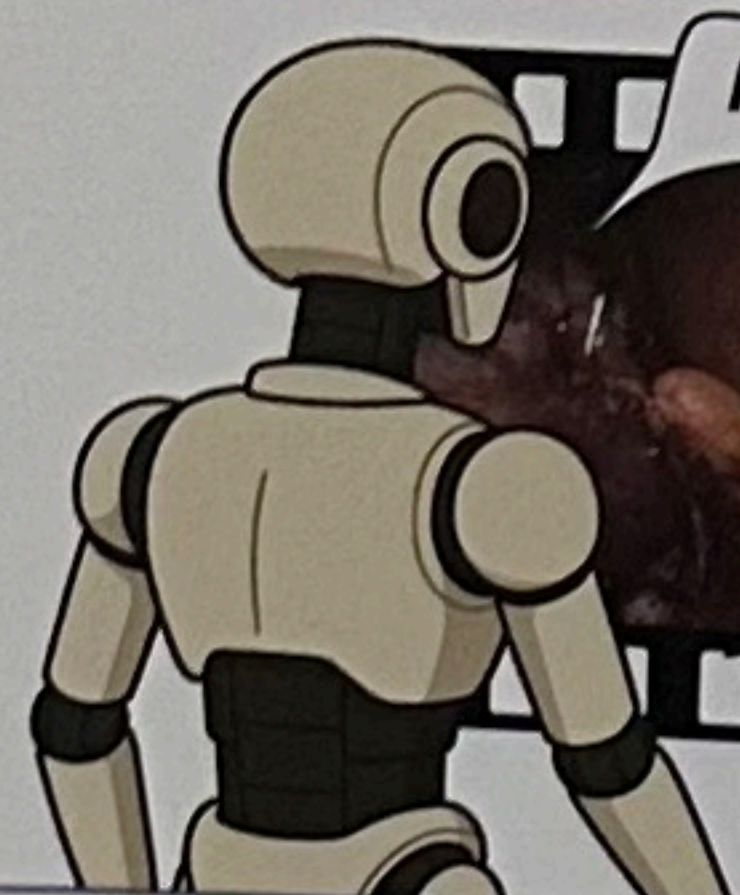# SurgX: Neuron-Concept Association for Explainable Surgical Phase Recognition

Ka Young Kim, Hyeon Bae Kim, Seong Tae Kim
Kyung Hee University, Republic of Korea

KYUNG HEE UNIVERSITY
AUGMENTED INTELLIGENCE LAB
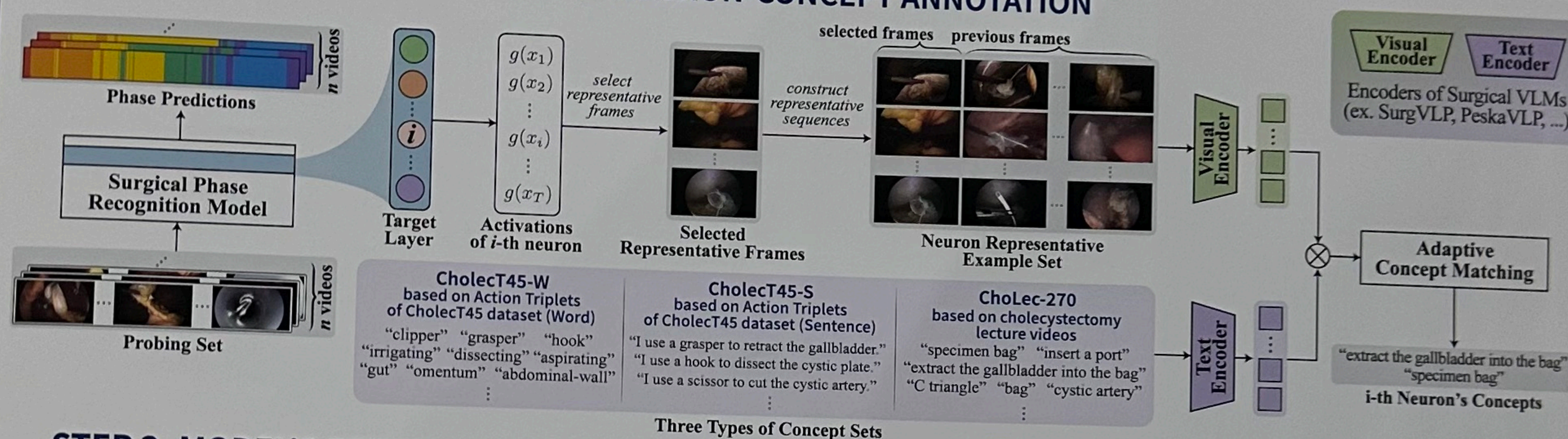MICCAI 2025 — PROJECT PAGE

It's Clipping and Cutting phase!

Why so?

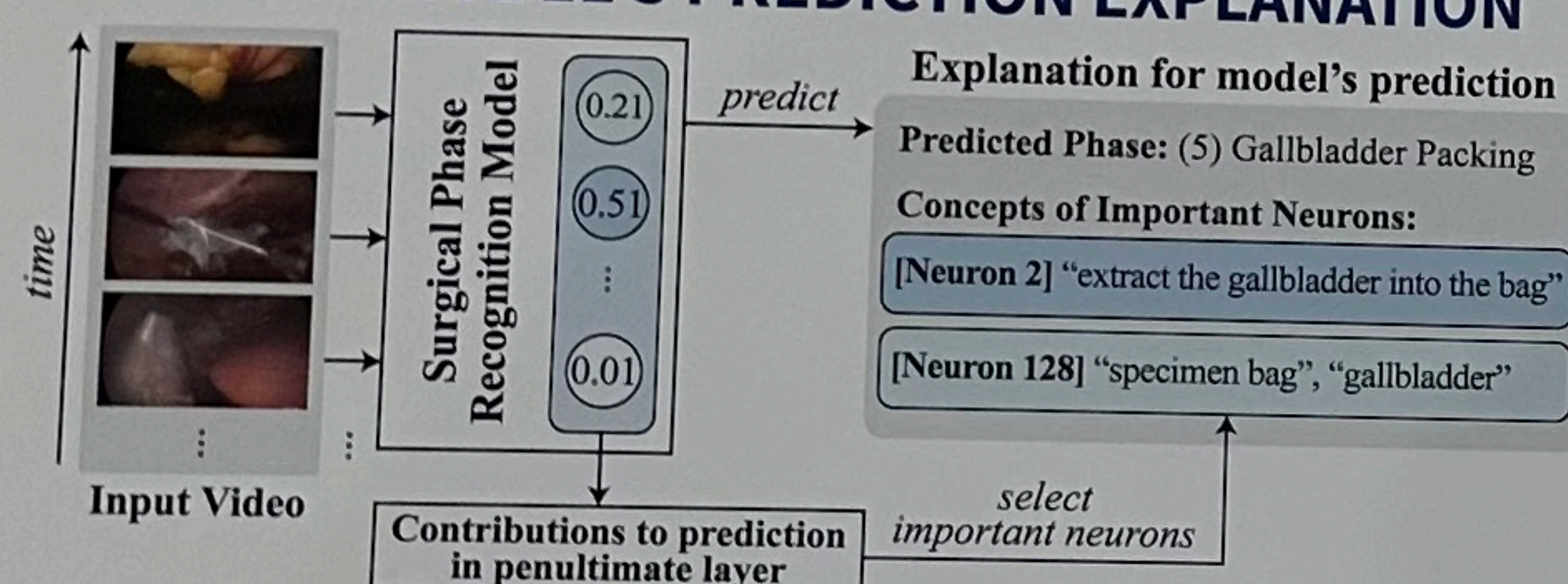*The concept of HEPATOCYSTIC TRIANGLE and CYSTIC ARTERY BETWEEN CLIPS in the input strongly drove my decision!*

SurgX, A Framework to Explain Surgical Phase Recognition Model

STEP 1. NEURON-CONCEPT ANNOTATION

Phase Predictions — Surgical Phase Recognition Model — Probing Set (n videos)

Target Layer — Activations of i-th neuron — $g(x_1), g(x_2), \ldots, g(x_i), \ldots, g(x_T)$ — select representative frames — Selected Representative Frames — construct representative sequences — Neuron Representative Example Set (selected frames, previous frames)

Visual Encoder — Text Encoder — Encoders of Surgical VLMs (ex. SurgVLP, PeskaVLP, ...)

Adaptive Concept Matching — "extract the gallbladder into the bag", "specimen bag" — i-th Neuron's Concepts

Three Types of Concept Sets

CholecT45-W based on Action Triplets of CholecT45 dataset (Word): "clipper" "grasper" "hook" "irrigating" "dissecting" "aspirating" "gut" "omentum" "abdominal-wall"

CholecT45-S based on Action Triplets of CholecT45 dataset (Sentence): "I use a grasper to retract the gallbladder." "I use a hook to dissect the cystic plate." "I use a scissor to cut the cystic artery."

CholLec-270 based on cholecystectomy lecture videos: "specimen bag" "insert a port" "extract the gallbladder into the bag" "C triangle" "bag" "cystic artery"

STEP 2. MODEL'S PREDICTION EXPLANATION

Input Video — Surgical Phase Recognition Model — (0.21)(0.51)...(0.01) — predict

Contributions to prediction in penultimate layer — select important neurons

Explanation for model's prediction

Predicted Phase: (5) Gallbladder Packing

Concepts of Important Neurons:
[Neuron 2] "extract the gallbladder into the bag"
[Neuron 128] "specimen bag", "gallbladder"

Neuron Representative Sequence Selection

Method	Concept Alignment Word	Sentence	Avg	Prediction Interpretability Word	Sentence	Avg
Single-Frame	0.3576	0.4569	0.4073	0.6426	0.5513	0.5970
Contiguous-Sequence	0.3641	0.4711	0.4176	0.6458	0.5487	0.5973
Dilated-Sequence (3)	0.3534	0.4920	0.4227	0.6545	0.5499	0.6022
Dilated-Sequence (5)	0.4153	0.4796	0.4475	0.6532	0.5452	0.5992
Dilated-Sequence (10)	0.3684	0.4004	0.3844	0.6856	0.5548	0.6202

Ablation Studies

- **Concept Alignment**: Cosine similarity between key neuron's concepts and the groundtruth class.
- **Prediction Interpretability**: Cosine similarity between key neuron's concepts and predicted class.

Concept Sets

Concept Set	Concept Alignment Word	Sentence	Avg	Prediction Interpretability Word	Sentence	Avg	Unique Concept
CholecT45-W	0.4078	0.3681	0.3880	0.5978	0.5099	0.5539	7
CholecT45-S	0.3000	0.4538	0.3769	0.4679	0.5564	0.5122	11
CholLec-270	0.4153	0.4796	0.4475	0.6532	0.5452	0.5992	34

Neuron Representative Frame Selection

| Method	Concept Alignment Word	Sentence	Avg	Prediction Interpretability Word	Sentence	Avg
Global Threshold	0.3222	0.4621	0.3922	0.5586	0.4819	0.5203
Global TopK	0.3099	0.4622	0.3861	0.5369	0.4531	0.4950
Video-wise Threshold	0.4153	0.4796	0.4475	0.6532	0.5452	0.5992
Video-wise TopK	0.4146	0.4824	0.4485	0.6454	0.5403	0.5929

Qualitative Result with Explanation Examples

Explanations for Correct Predictions
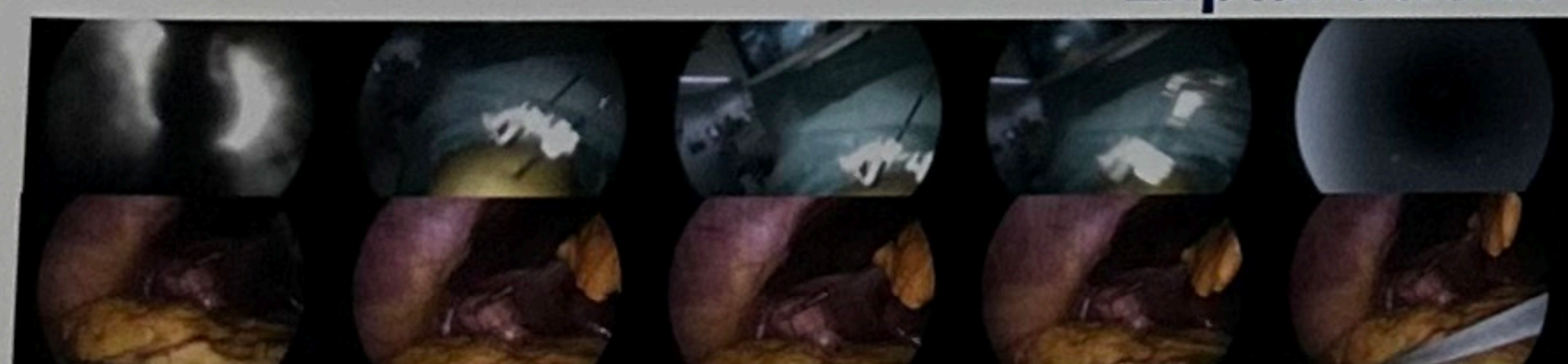
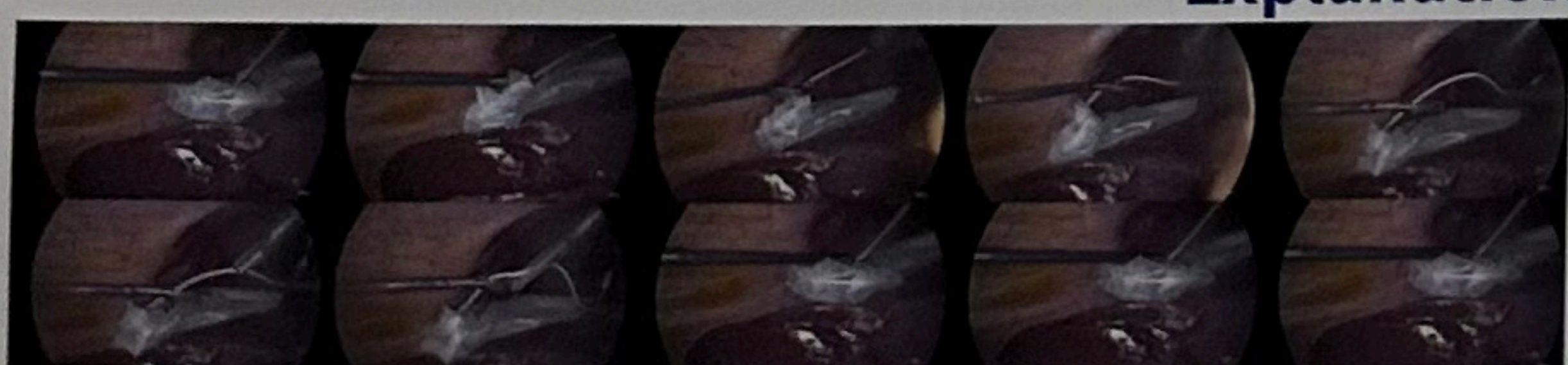- **GT:** (1) Preparation **Pred:** (1) Preparation
Key Neuron: #91
"grasping the gallbladder," "empty the gallbladder", "insert a port"
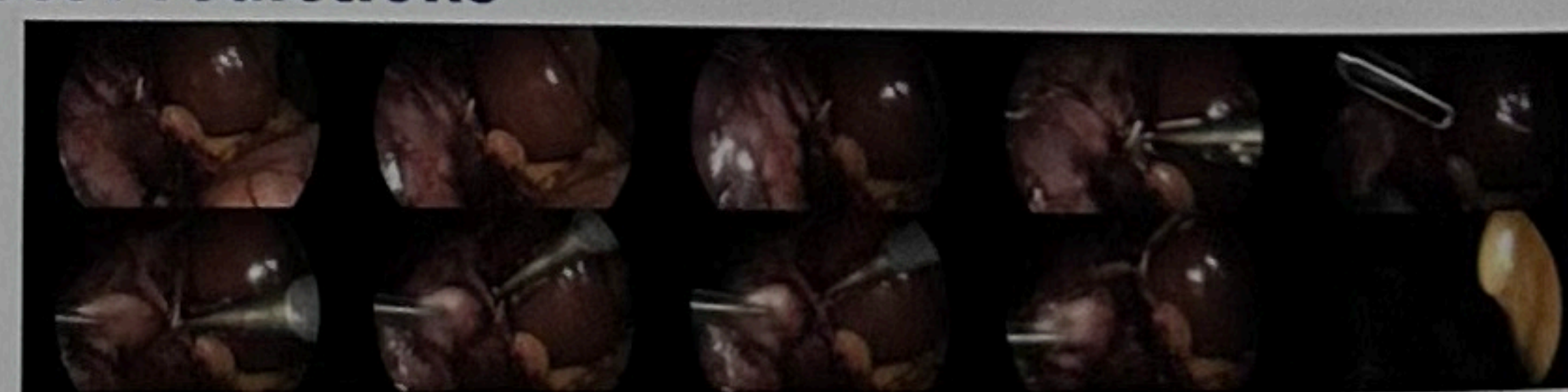
- **GT:** (3) Clipping and Cutting **Pred:** (3) Clipping and Cutting
Key Neuron: #182
"hepatocystic triangle", "cystic artery is isolated between clips"

Explanations for Mispredictions

- **GT:** (6) Cleaning and Coagulation **Pred:** (5) Gallbladder Packing
Key Neuron: #250
"extract the gallbladder into the bag"

- **GT:** (4) Gallbladder Dissection **Pred:** (3) Clipping and Cutting
Key Neuron: #182
"hepatocystic triangle", "cystic artery is isolated between clips"