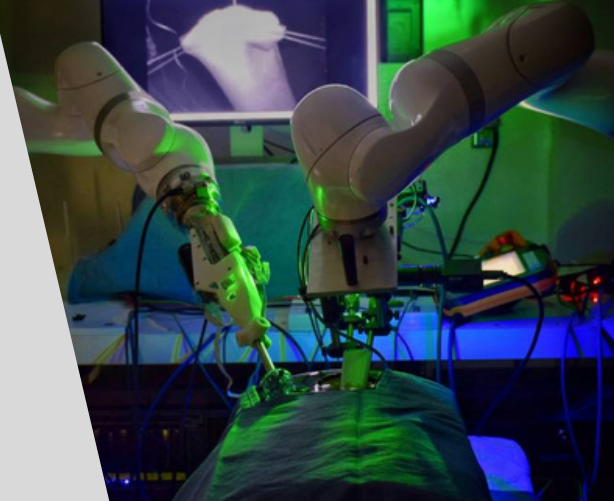




Platform Goals

We detect instrument, verb, and target triplets in the operating room in real time. That capability enables safety critical alerts, richer workflow analytics, and context aware assistance. The system runs on device with less than 10 milliseconds per frame at less than 1 watt and is robust to smoke, occlusions, motion blur, and lighting changes. We have built a neuromorphic backbone that favors sparsity, preserves temporal microstructure, and limits memory traffic, enabling private offline use at the edge.

Use Case

We evaluated the CholecT45 (CC BY-NC-SA): 45 cholecystectomy videos, 90,489 frames at 1 fps. We used a curated set of 100 triplet classes from the $6 \times 10 \times 15$ space, with multi-label binary targets per frame. We follow the official cross-validation folds and report AP_i, mAP_{it}, and mAP_{ivt} using ivtmetrics. Augmentations include vertical and horizontal flips, contrast shifts, and 90-degree rotations. This evaluation is non-commercial.

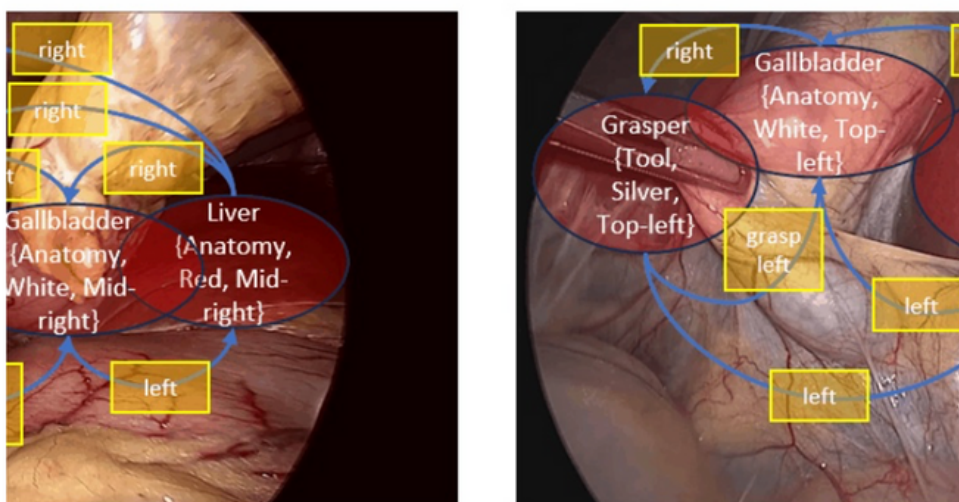


Figure 1. Dataset Overview.

Architecture

Our neuromorphic multitask recognizer comprises three component heads for instrument, verb, and target, together with a joint triplet head on a sparse, causal spatiotemporal backbone. The core uses leaky integrate and fire neurons with learnable leak and threshold. Residual connections stabilize training and enable deeper temporal context. A lightweight temporal attention block improves sequence disambiguation. An adaptive Poisson event encoder exploits sparsity by emphasizing salient pixel changes. Training uses mixed precision FP16 and FP32 and a multitask objective that jointly optimizes component predictions and the fused triplet output.

The deployed accelerator processes sequences of eight time steps in order. Each neuron maintains a 16 bit membrane potential. Weights and activations follow our quantization specification. The 11.2 million parameter model maps cleanly to FPGA using HLS with output stationary tiling, line buffered feature storage, and double buffered DMA.

The top module exposes AXI Stream for frame ingress, AXI4 MM for weights and feature buffers, and AXI Lite for control and status. Postprocessing fuses component scores into calibrated triplet posterior probabilities and runs entirely on chip to avoid host transfers and interface stalls.

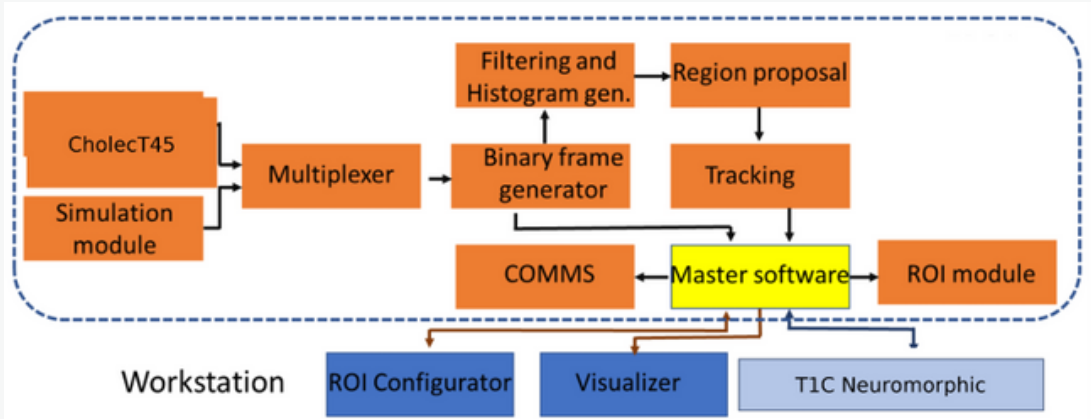


Figure 2. End-to-end CholecT45 pipeline with ROI-driven neuromorphic inference. Frames from the dataset are multiplexed and converted to binary frames, then filtered with histogram features to generate region proposals. A tracker maintains temporal coherence while the Master software orchestrates I/O via COMMS, dispatches ROIs to the T1C Neuromorphic accelerator, and updates the Visualizer. The ROI Configurator lets the operator define/update regions of interest, and the ROI module applies those settings for on-device triplet recognition.

Technical Performance

On an AMD/Xilinx Artix UltraScale+ (XCAU15P), the accelerator sustains ~125 GOP/s at ~0.68 W (~**184 GOP/s/W**). With identical preprocessing and post-processing, **we observe efficiency advantages of ~593× vs Intel i9-12900H** (0.31 GOP/s/W), **~48× vs NVIDIA RTX 3060** (3.81 GOP/s/W), and **~23× vs Jetson Nano** (8.0 GOP/s/W). **Measured latency is under 3.3 ms per frame**, which supports clinical frame rates with margin.

Results	
Processor	Power Efficiency GOP/s/W
T1C Neuromorphic Hardware	184.00
NVIDIA Jetson Nano	8.00
NVIDIA RTX 3060 GPU	3.81
Intel i9-12900H CPU	0.31

Roadmap

Currently our priorities are focusing on characterizing the baseline gesture recognition model on our custom neuromorphic hardware, despite the current challenges posed by hardware capabilities and configuration within the HLS framework. This initial step will allow us to establish reliable measurements of the model’s performance and energy efficiency directly on neuromorphic hardware. Building on these results, the next phase will focus on improving recognition accuracy and energy efficiency by exploring alternative model architectures and training mechanisms, such as more advanced neuron models, optimized learning algorithms, and specialized training strategies tailored for spiking neural networks (SNN).