



Long-Distance Gesture Recognition using Dynamic Neural Networks

Shubhang Bhatnagar^{1,2}, Sharath Gopal², Narendra Ahuja¹, Liu Ren²

¹ University of Illinois Urbana-Champaign

² Bosch Research North America, Bosch Center for Artificial Intelligence

Oct 2nd 2023

Motivation

Gesture Recognition for Human-Machine Interaction

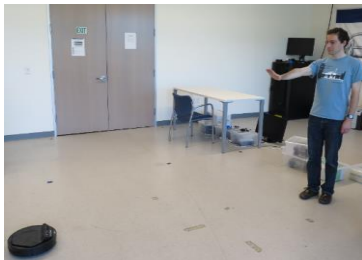
- Gestures form a natural interface for communication between humans and machines
- Provide a convenient and contact-less way to communicate with robots



Mobile Robots
and Drones



Home
Automation



AR/VR



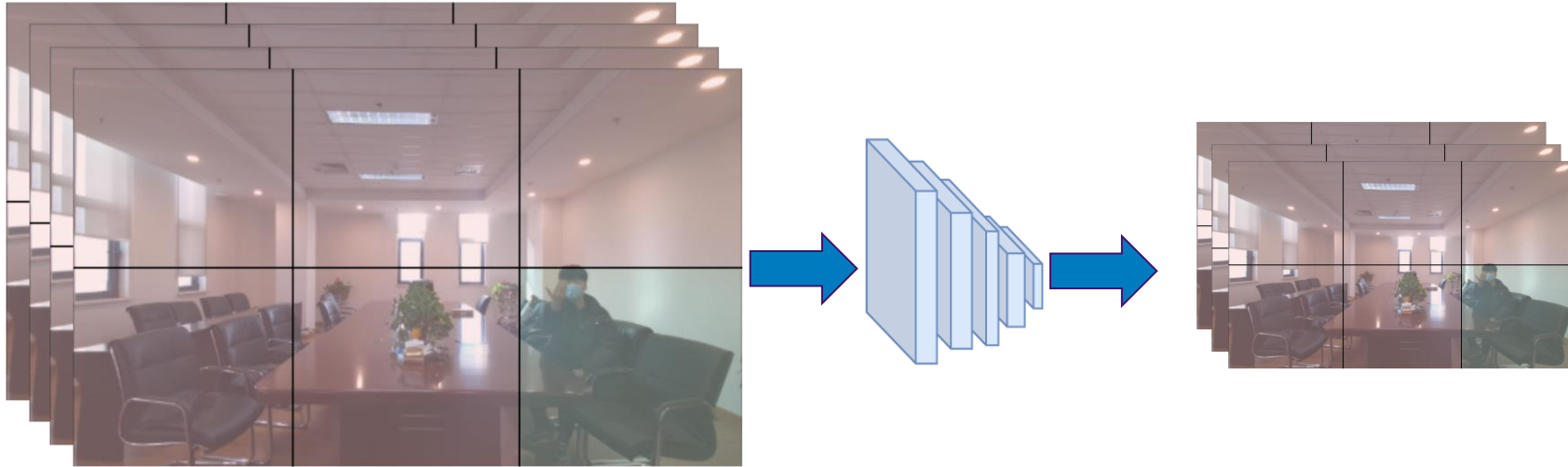
Conference
Rooms



No Gesture Detected

Challenges

Long Distance Gesture Recognition



Gesturing
subject is small

3D CNNs down
sample input

Features might
lose too much
gesture
information

Higher resolution
video requires
more compute,
bandwidth!

Method

Spatially Dynamic Neural Network

Spatially Dynamic Neural networks can adapt computational graph to input at run-time

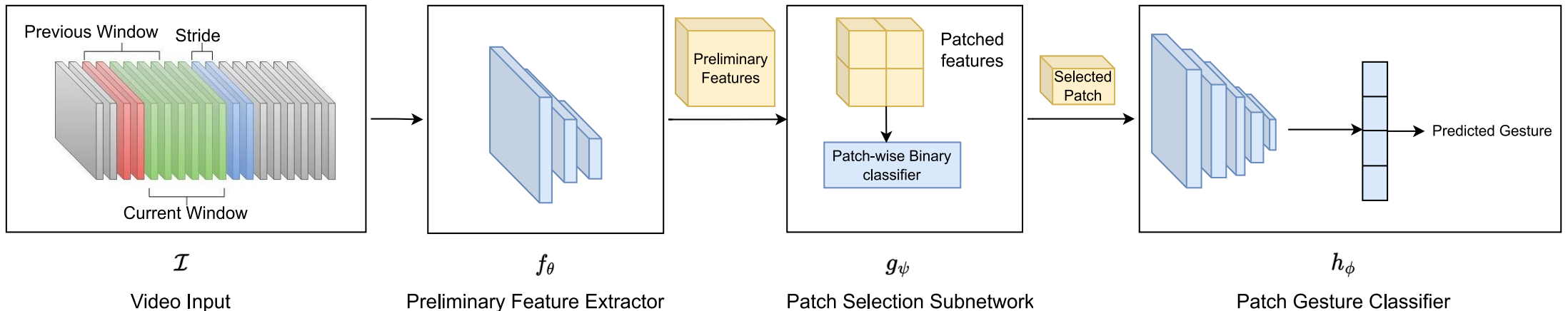


Selectively Discard **background** features

Selectively Preserve **gesturing subject** features

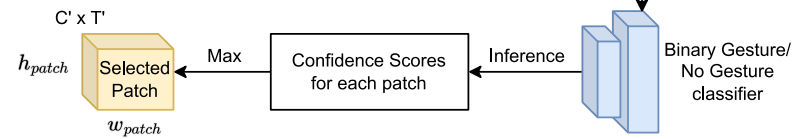
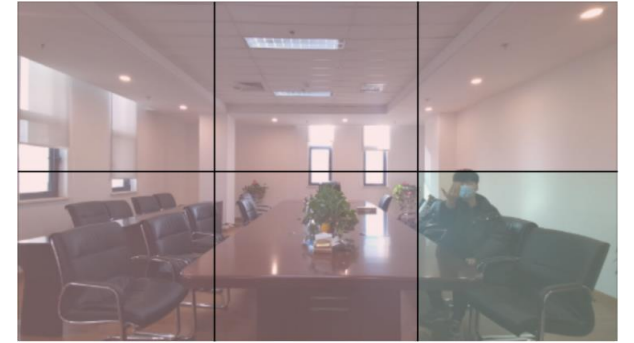
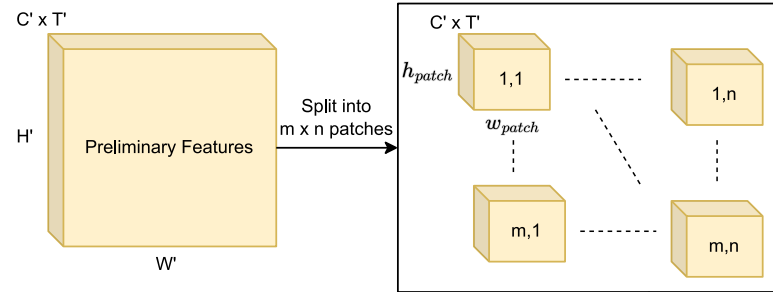
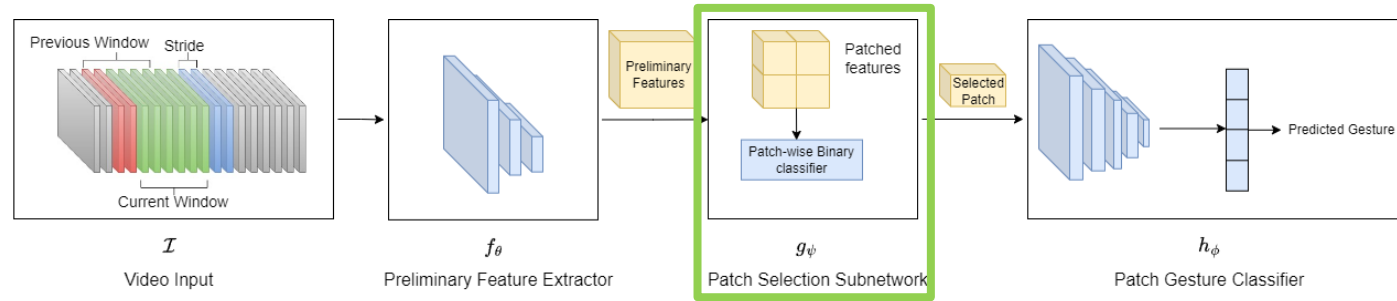
Our Pipeline

3 blocks made of 3D CNN subnetworks

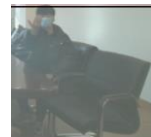


Method

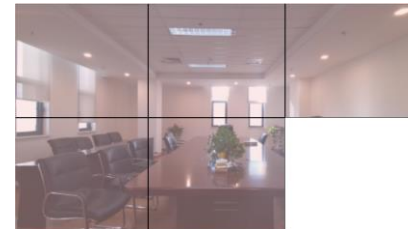
Patch Selection Subnetwork



Selected Patch



Discarded Patches

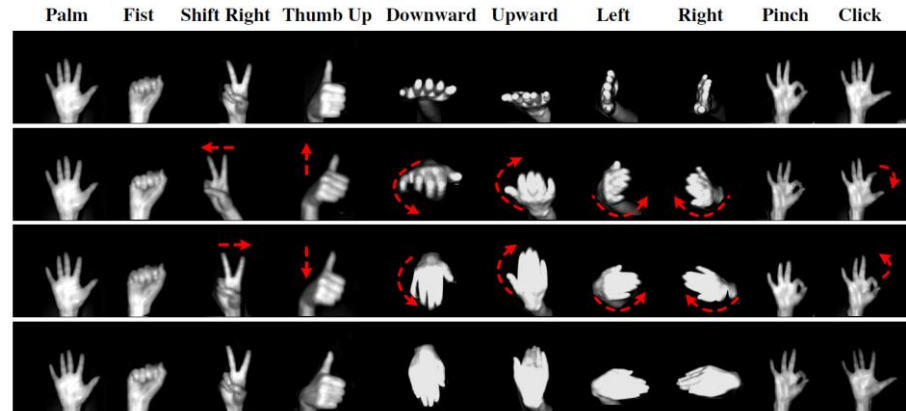


Benchmark

Long Distance Gesture Recognition

Measure Top-1 Accuracy on LDConGR [1]
long range gesture recognition dataset

- subjects seated 1-4 meters away
- 10 gestures (static and dynamic)



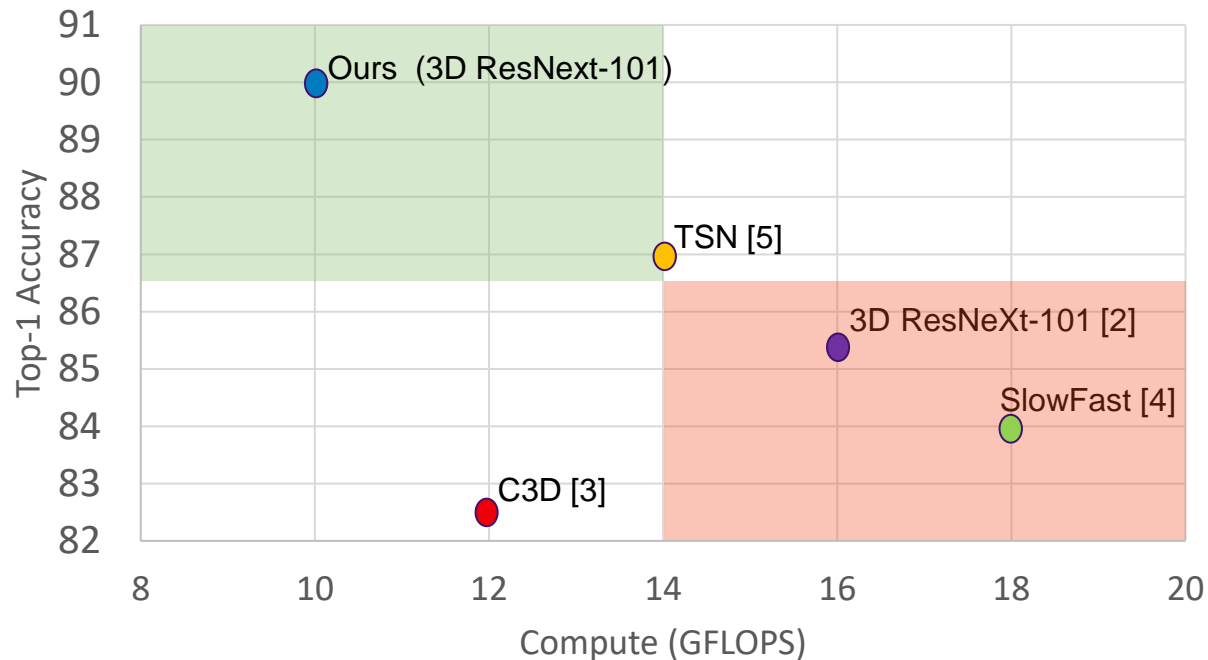
Results

Long Distance Gesture Recognition

Lower compute required compared to state-of-the-art

Better performance compared to state-of-the-art

Significant improvements in resource constrained settings

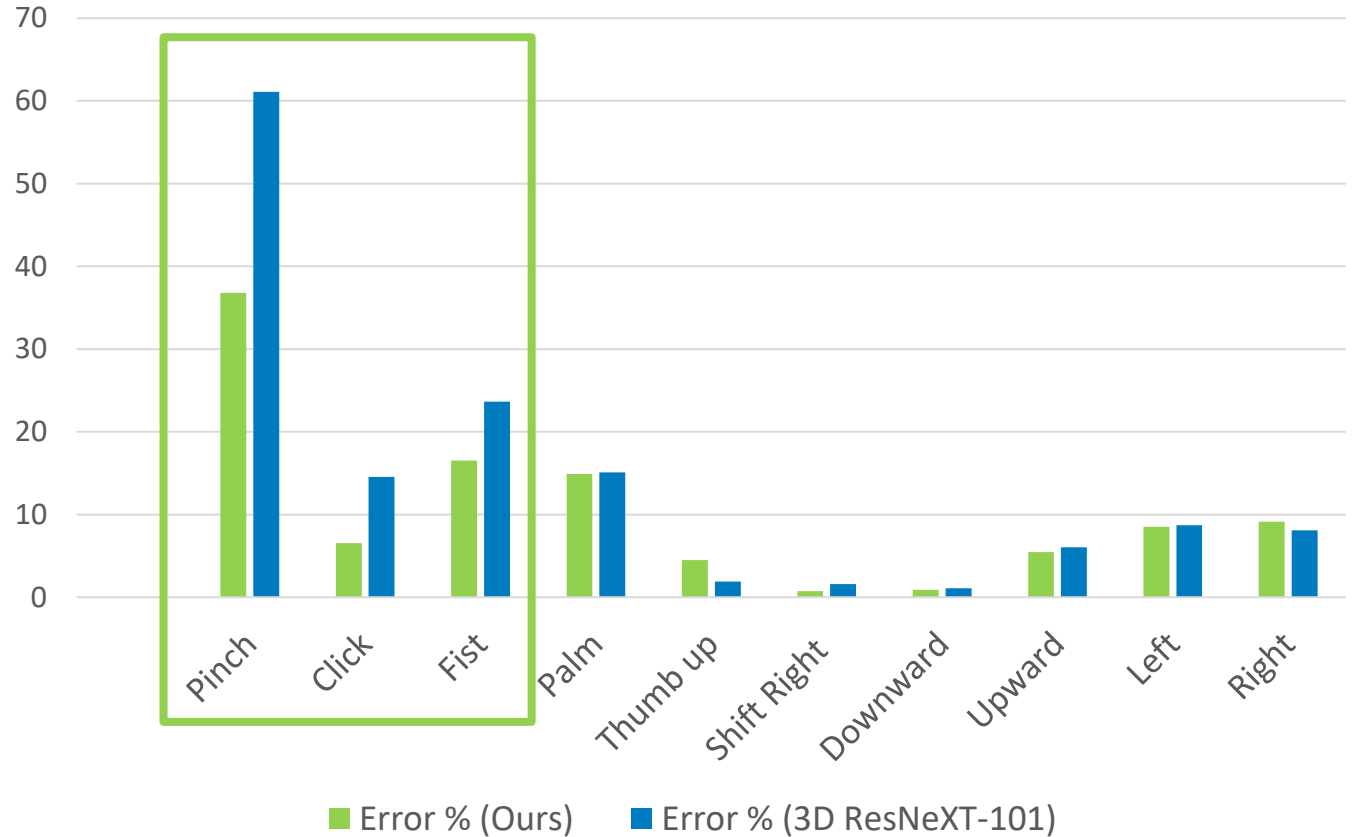


Method	Compute (GFLOPS)	Top-1 Accuracy
Ours (3D MobileNet)	1.5	76.68
3D MobileNet [6]	1.5	65.33

Results

Detecting Finer Gestures

Error : Lower is Better



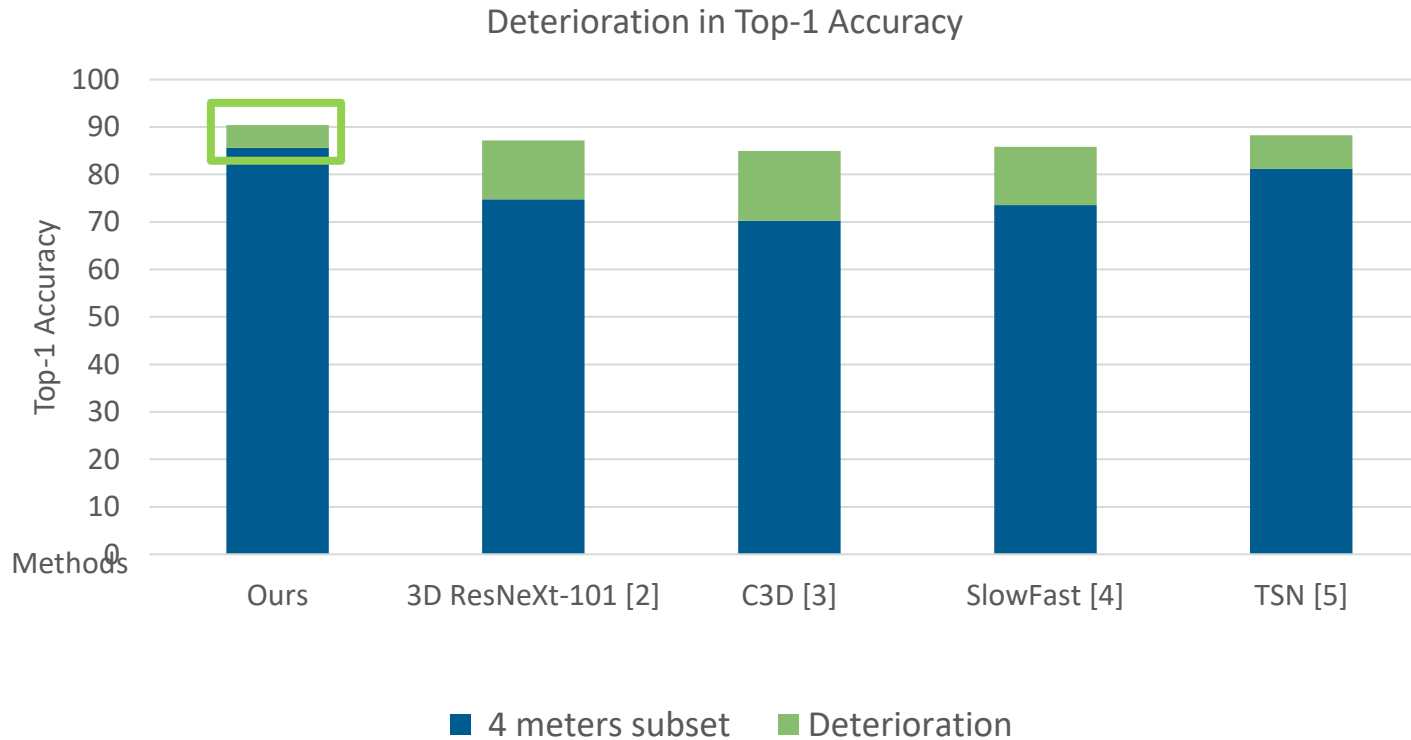
Better Performance on **finer gestures** due to utilization of higher resolution features

- 25% lower error on Pinch
- 8% lower error on Click
- 7% lower error on Fist

Results

Deterioration with Distance

Lower deterioration in performance with distance



Conclusion

- We propose a novel dynamic neural network for long distance gesture recognition
- Proposed network can recognize gestures more **accurately** and **efficiently** by discarding background features early on
- Shows state-of-the-art performance for long distance recognition, especially for resource constrained devices



Visit us at 10 am @ Track 13 (Visual Learning) , Slot 9



Thank You!
Questions ?