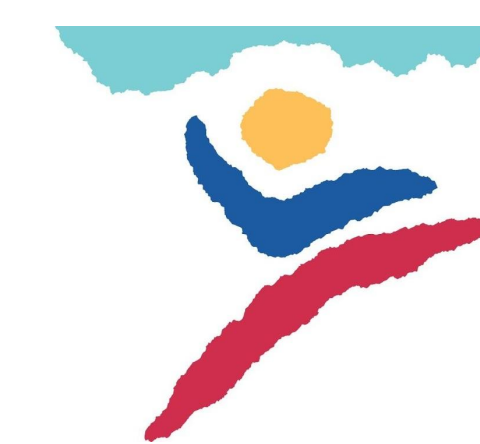# Vision-Based Hand Hygiene Monitoring in Hospitals

Zelun Luo, Boya Peng, Zuozhen Liu
Stanford University

Lucile Packard Children's Hospital AT STANFORD

Introduction

- Visual data capturing patient experience and environment during health care episodes is a valuable yet underexplored source of clinical data.

- Such data can contain rich information about both patient condition (e.g. distress and mobility) and clinical care activities (e.g. bundle compliance and hand hygiene).

- We propose to deploy depth sensors in hospitals that capture the physical space near hand hygiene dispensers, and use computer vision methods to detect when a person performs a hand hygiene action.

- In this work, we present several approaches for machine learning-based interpretation of visual clinical data. We first explored a SVM based baseline model, followed by a CNN-based hand hygiene detection model, and a posed-based model[1,2].
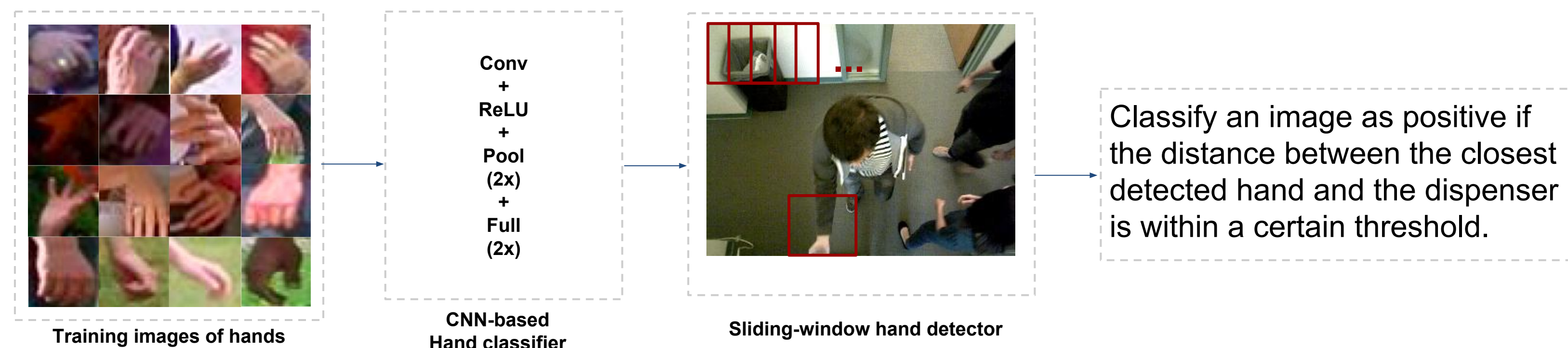
Approach

CNN-based hand hygiene detection
- Convolutional neural networks (CNNs) have achieved state-of-the-art performance for a wide range of visual recognition tasks.
- We trained a CNN to classify the hand hygiene action in video frames.
- Our network architecture consists of 2 convolutional layers, each followed by a max pooling layer, and 2 fully connected layers.
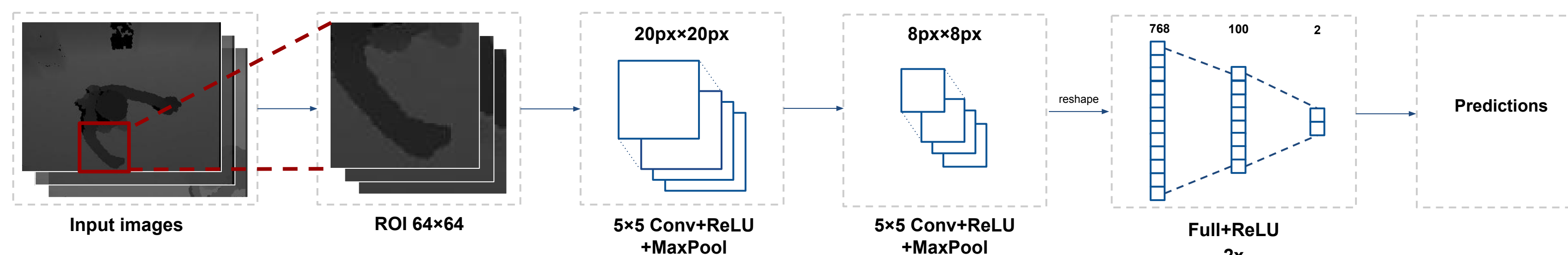
Pose-based approach
- We also explored a pose-based approach that provides additional information of which person is performing the hand hygiene. In this approach we first detect each person's hands and then label hand hygiene if the hand is directly under a soap dispenser.

Classify an image as positive if the distance between the closest detected hand and the dispenser is within a certain threshold.

Hospital Sensor Setup

Depth sensors have been deployed across a full unit of Lucile Packard Children's Hospital Stanford. Sensors are placed in both corridors and patient rooms to capture all hand hygiene dispensers and potential moments of hand hygiene.


Sensor mounted on wall

Corridor views

Pilot Dataset

- For an initial pilot study, we collected data from a depth sensor mounted in a lab environment.

Dataset Statistics	
Time elapsed (seconds)	5,093
Number of frames	25,465
Number of positive hand-hygiene instances	63
Number of positve hand-hygiene frames	630

- Examples of depth images from the dataset: from left to right, the first two are positive instances of hand hygiene, and the last two are challenging negative instances.
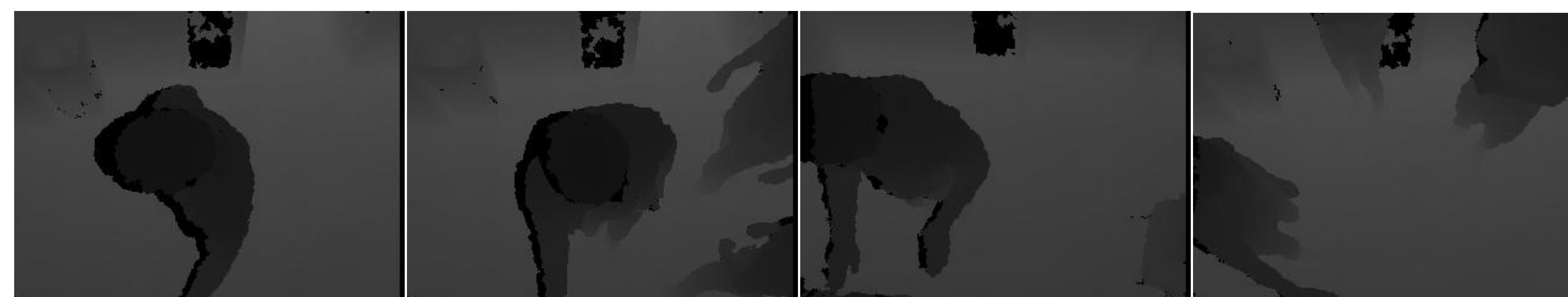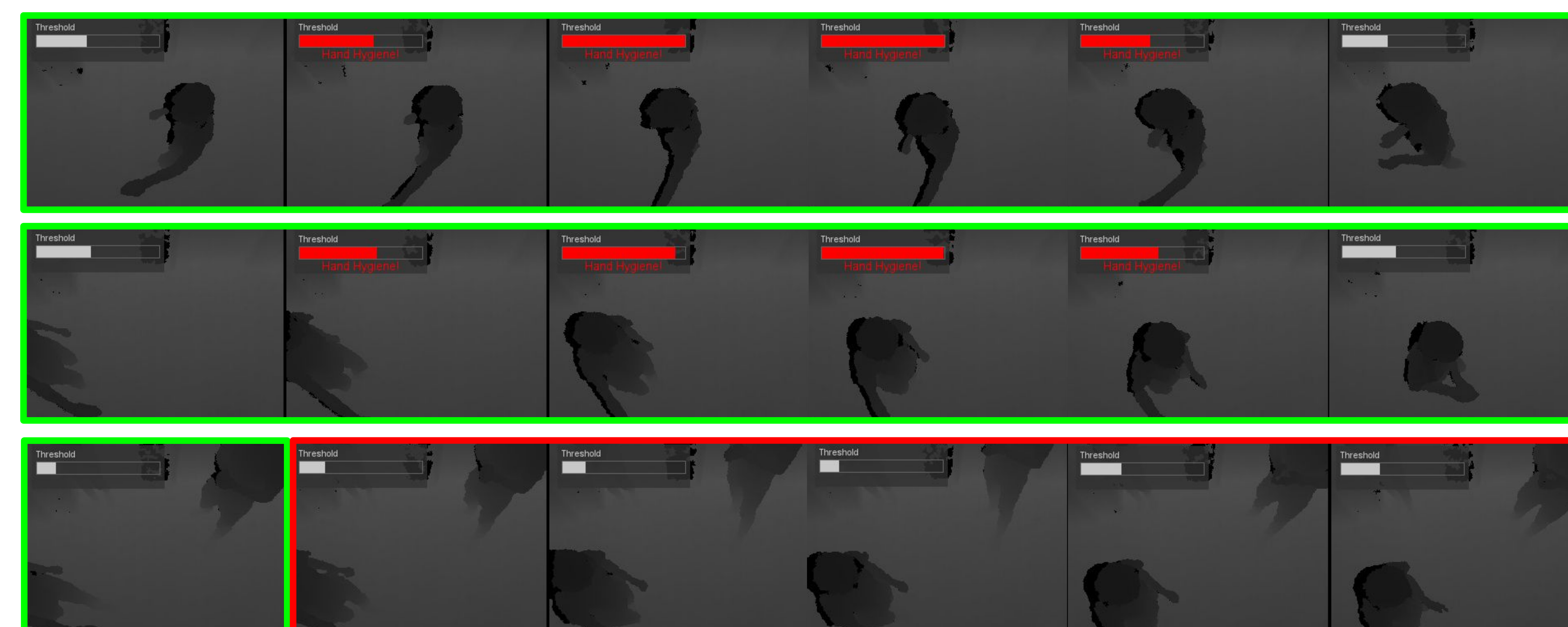

Results

- **Mean Average Precision (MAP):**

	RGB	Depth
SVM baseline over dispenser region	0.561	0.889
CNN over full image	0.695	0.450
CNN over dispenser region	**0.956**	**0.937**
Pose-based approach	0.807	-

- **Examples of detection results (green for correct labelings; red for incorrect labelings):**

References

[1] J. Tompson, A. Jain, Y. LeCun, C. Bregler. "Joint Training of a Convolutional Network and a Graphical Model for Human Pose Estimation" in proceedings of Neural Information Processing Systems (NIPS 2014)
[2] J.Shotton et al. "Real-Time Human Pose Recognition in Parts from Single Depth Images", in CVPR 2011