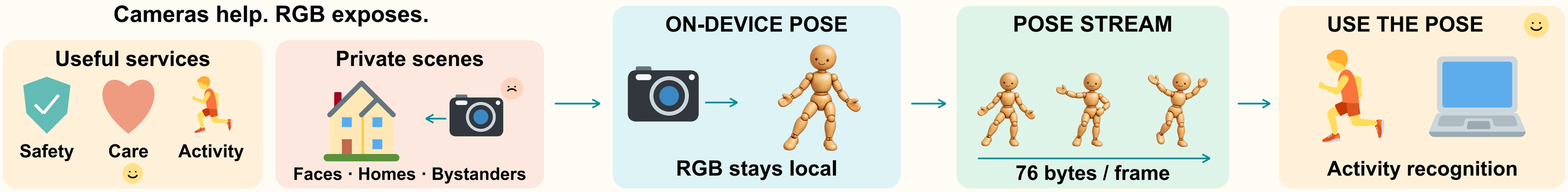
WiP: Privacy-First Pose Streams for IoT Cameras

Aveline Xiaoyuan Han Chenshu Wu
The University of Hong Kong

A camera that shares pose, not video.



Useful Cameras, Private Lives

Smart cameras support security, assisted living and activity-aware services in everyday spaces.

Many activity-aware services need motion cues, yet RGB captures much more: **faces, clothing, bystanders and private surroundings**. Streaming or storing video extends that exposure beyond the camera.

The Gap: RGB or Fixed Labels

Capture-side protection reduces visual detail through optical or image transformations.

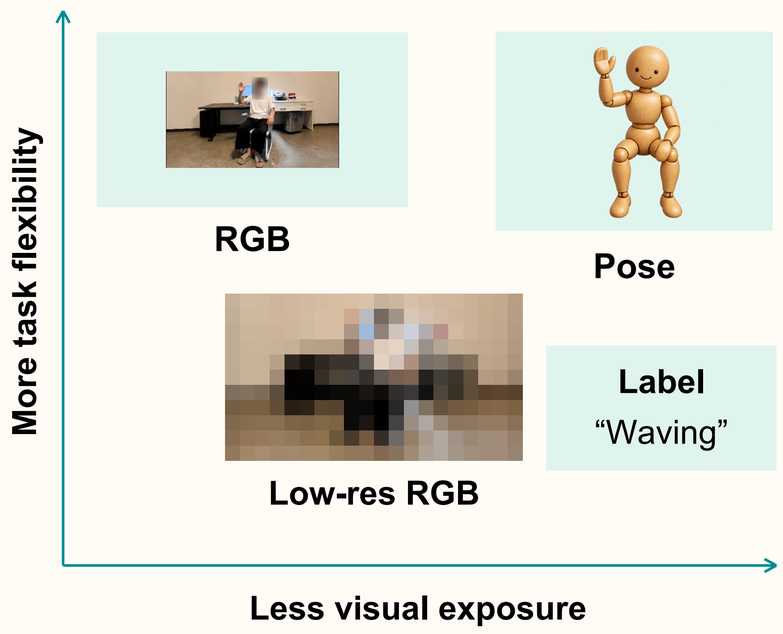
System-side protection controls video storage, sharing and query release.

Video often remains the interface. Labels reveal less, but bind the camera to a fixed task.

Research gap

Can a constrained MCU camera expose reusable motion while keeping appearance-rich RGB local?

Pose: Utility with Less Exposure



Representation comparison: face blurred, pose illustrated, axes conceptual.

Challenges

1. Fit the whole camera. Early features need spatial detail, yet share scarce SRAM with runtime state and communication buffers.

One SRAM budget, several demands

Pose features	Runtime	Buffers

Schematic: shared space, not measured proportions

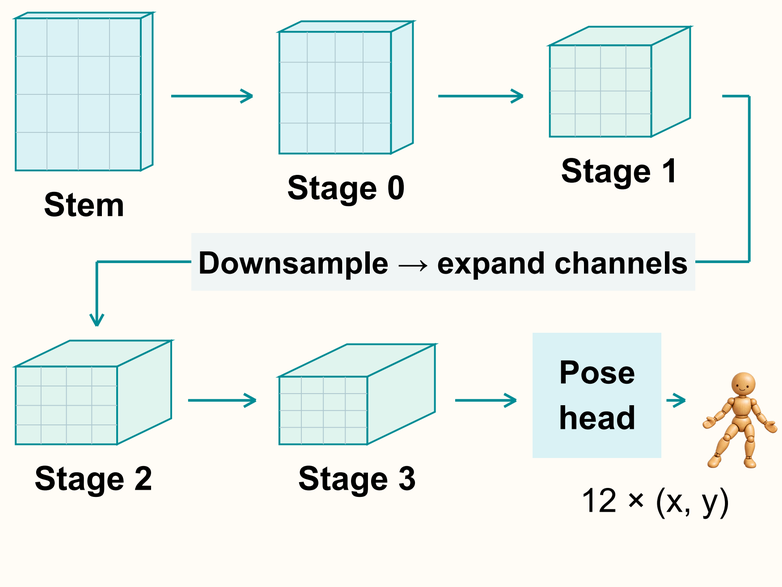
2. Preserve motion that matters. Good frame-level pose accuracy does not guarantee useful activity recognition. Wrist trajectories, joint velocity and short movements also matter—motivating structural and temporal refinement.

1. Activation-Aware Pose

Insight. Wide early features consume scarce MCU memory and compute.

Design. μStage keeps early stages narrow, then expands channels as spatial resolution falls.

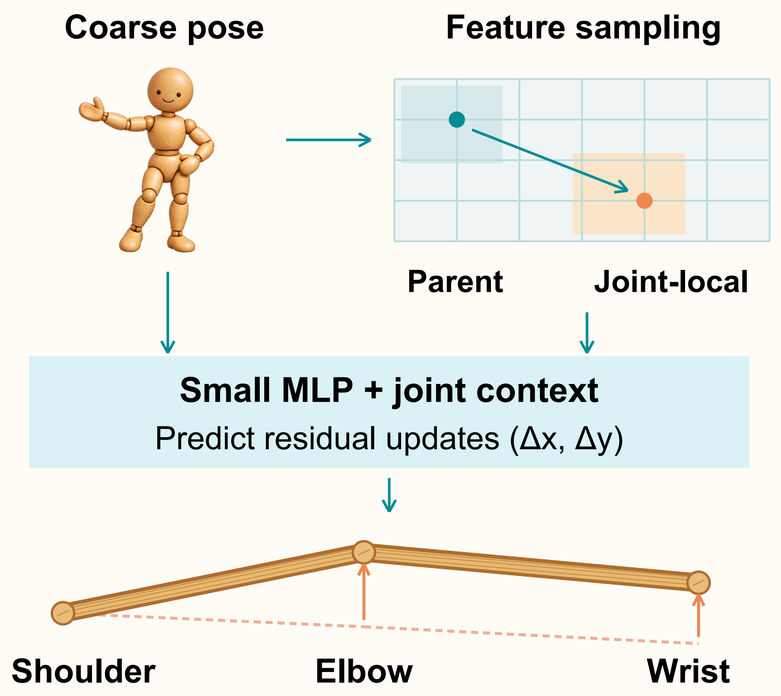
Keep high-resolution features narrow



2. Structural Arm Refinement

Insight. Arm structure helps correct wrist errors.

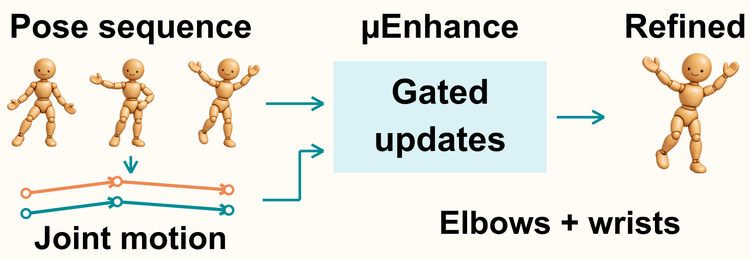
Design. ArmChain uses local and parent-joint features to refine elbows, then wrists.



Dashed: initial arm; wood: illustrated update.

3. Motion-Aware Refinement

Insight. Activity recognition needs consistent motion across frames.



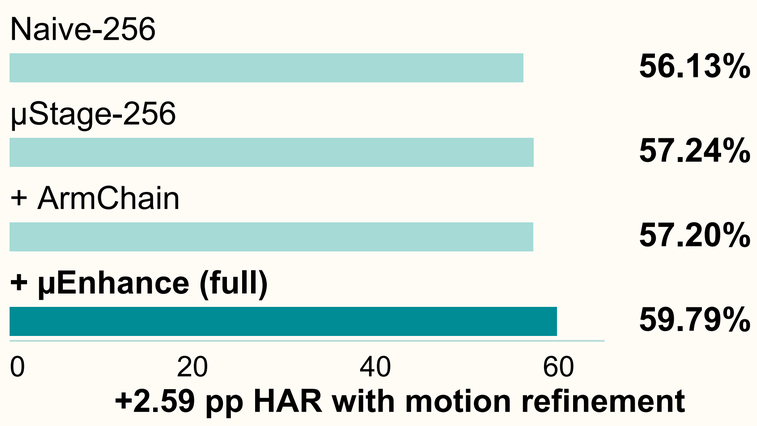
Useful Pose, Measurable Gains

20-class NTU60 X-Sub; source-adapted CTR-GCN. Preliminary WiP results.

59.79% HAR Top-1

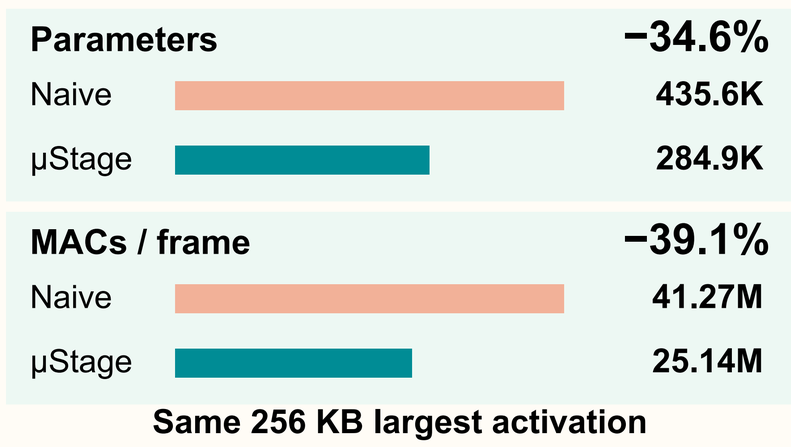
69.70% Pose PCK@0.10

Activity accuracy across the pipeline



Less Compute, Useful Pose

Naive-256 → μStage-256



Same 256 KB largest activation

From MCU to Live Demo

STM32H753



WiP prototype

INT8: 1.290 s/frame

M5Stack CoreS3



Live demo

Pose-stream interface

More Information

Homepage

xiaoyuanh@connect.hku.hk