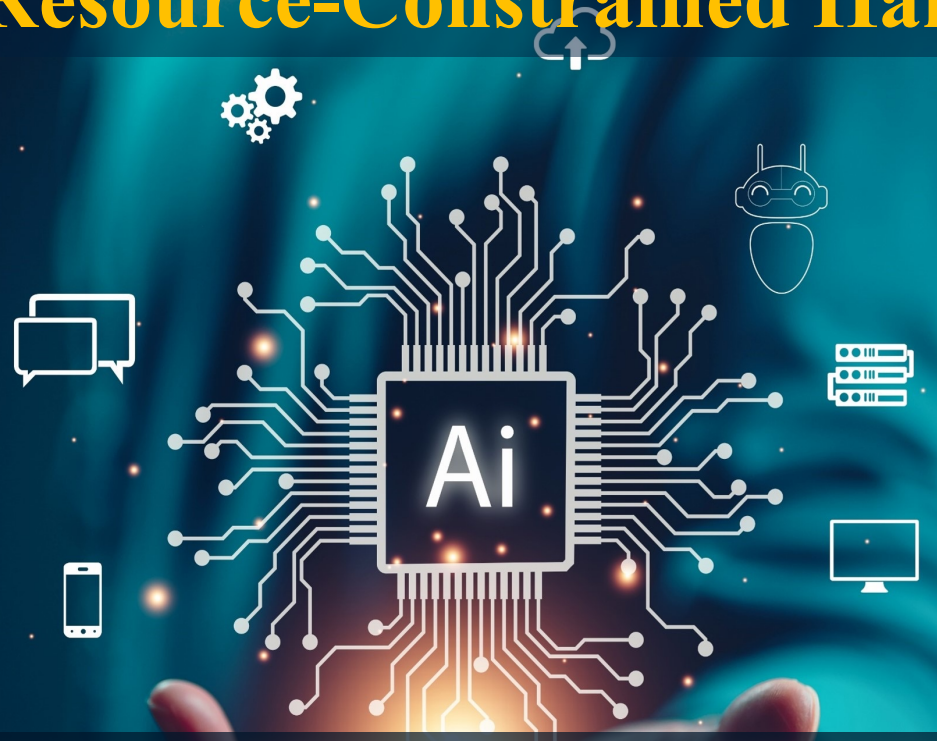
tinyML® Research Symposium

Enabling Ultra-low Power Machine Learning at the Edge

April 22, 2024

www.tinyML.org

TinyVQA: Compact Multimodal Deep Neural Network for Visual Question Answering on Resource-Constrained Hardware



Hasib-Al Rashid*[1,3], Argho Sarkar*[1], Aryya Gangopadhyay[1], Maryam Rahnemoonfar[2] and Tinoosh Mohsenin[3]

1. University of Maryland, Baltimore County
2. Lehigh University
3. Johns Hopkins University

*Both authors contributed equally to this research.

Presenter Details

Hasib-Al Rashid

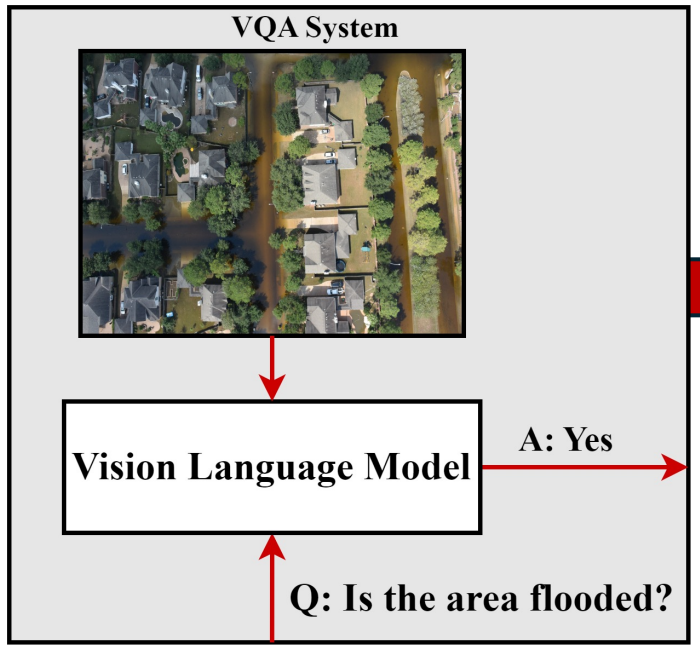
Ph.D. in Computer Engineering,
University of Maryland, Baltimore County (UMBC),
Graduation: June 2024

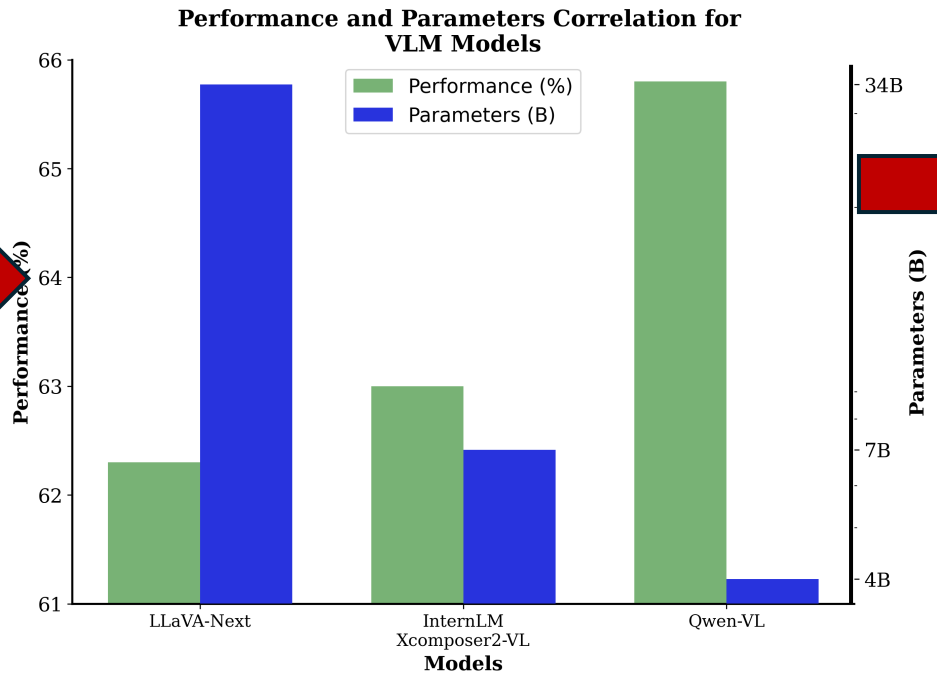
Visiting Research Assistant,
Johns Hopkins University

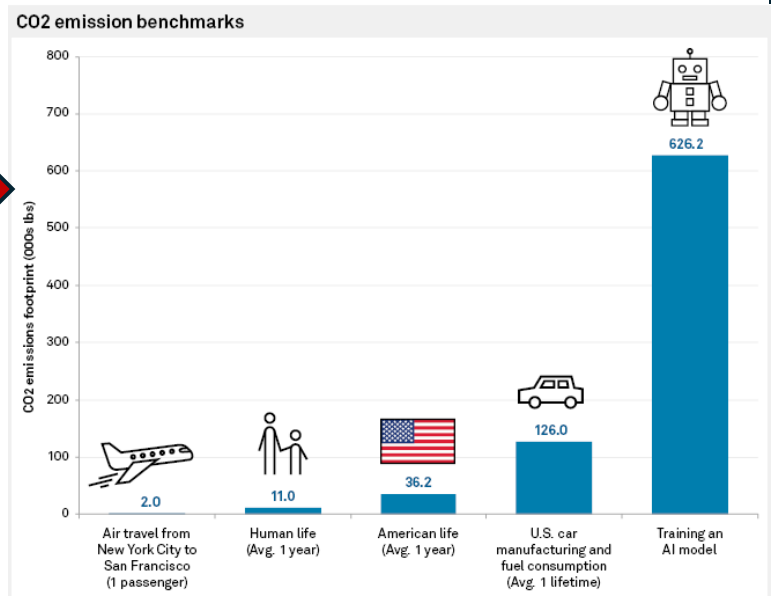
- ***Advisor's name:***
  - Dr. Tinoosh Mohsenin, Johns Hopkins University
- ***Research area/interests:***
  - Multimodal Deep Neural Networks,
  - Model Compressions for Edge Hardware,
  - Software-Hardware Co-design.
- ***Email:*** hrashid4@jhu.edu

Vision Language Models (VLMs)

- Vision language models are a type of generative models that take image and text inputs, and generate text outputs
- Example: Visual Question Answering (VQA)
- State-of-the-arts VLMs have billions of parameters
- Training these Large VLMs are not sustainable

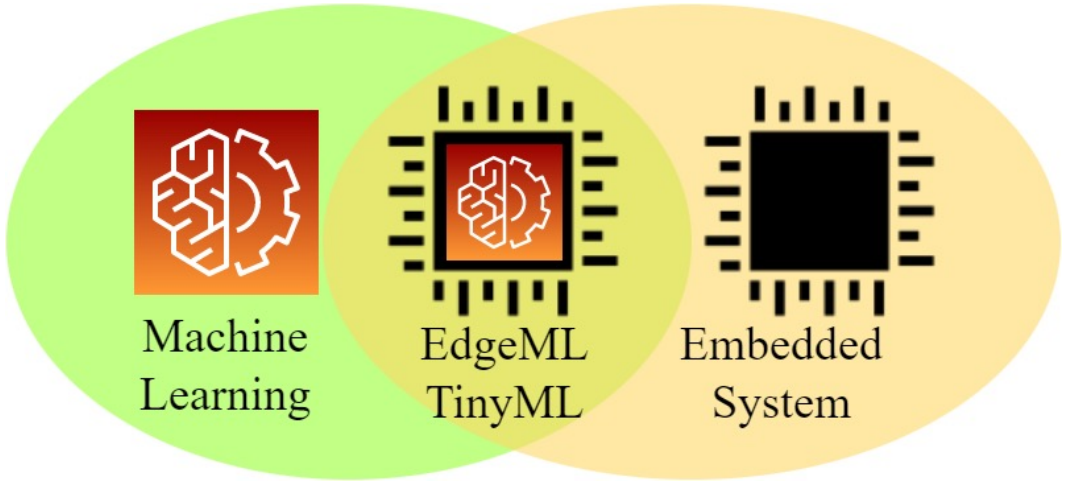


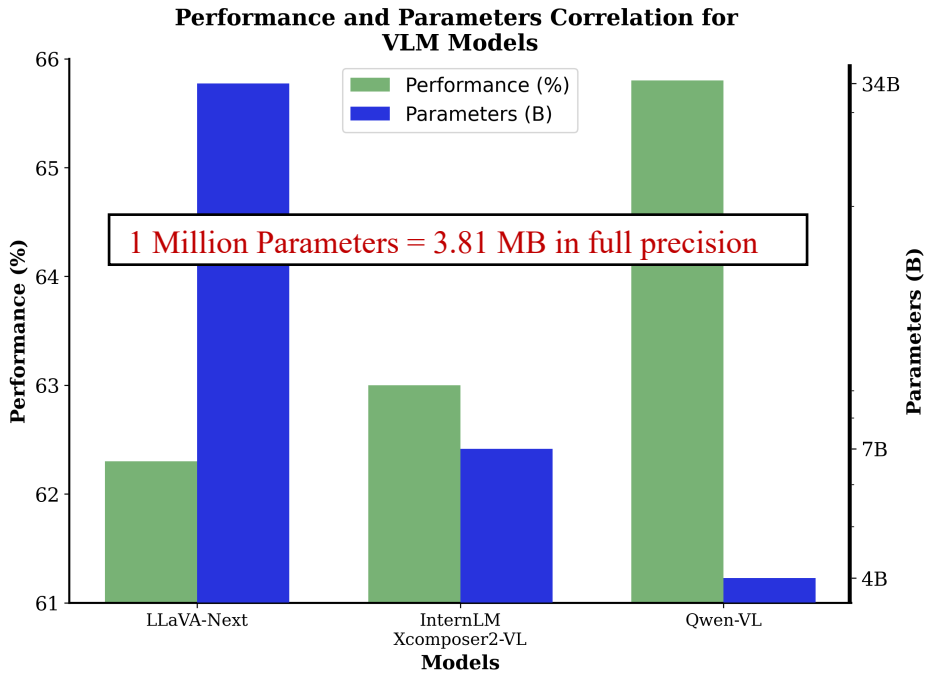




https://www.spglobal.com/marketintelligence/en/news-insights/trending/HyvwuXMO9YgqHfj7J6tGlA2

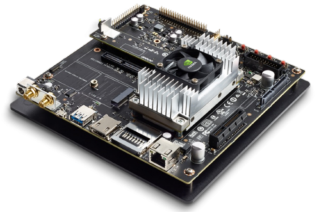
Motivation Towards tinyML

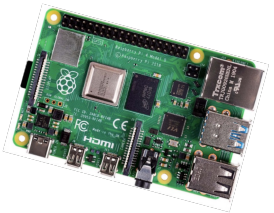




Research Gap


FPGA

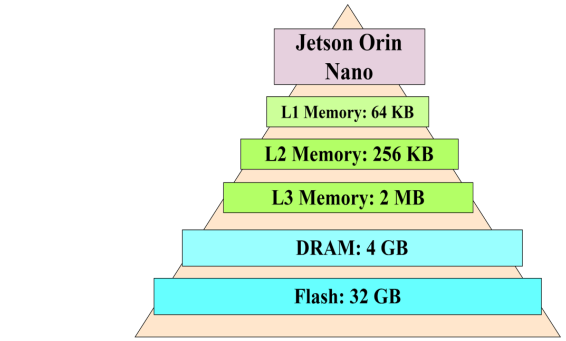

Nvidia TX2

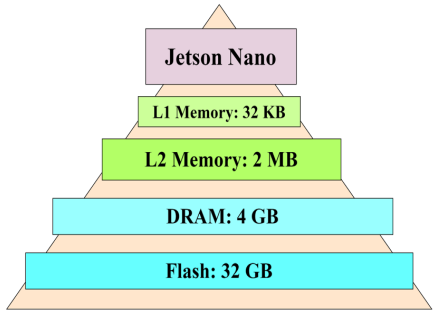

Raspberry Pi

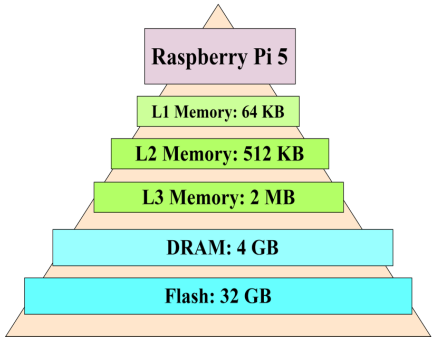

Arduino

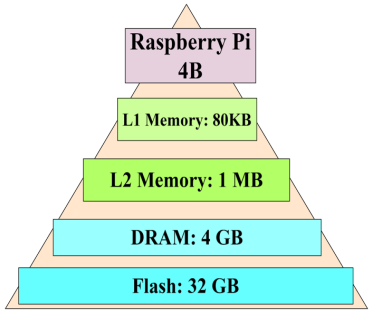

Nvidia Jetson Nano

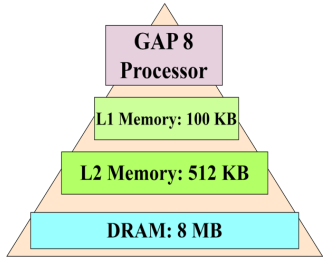

GAPuino







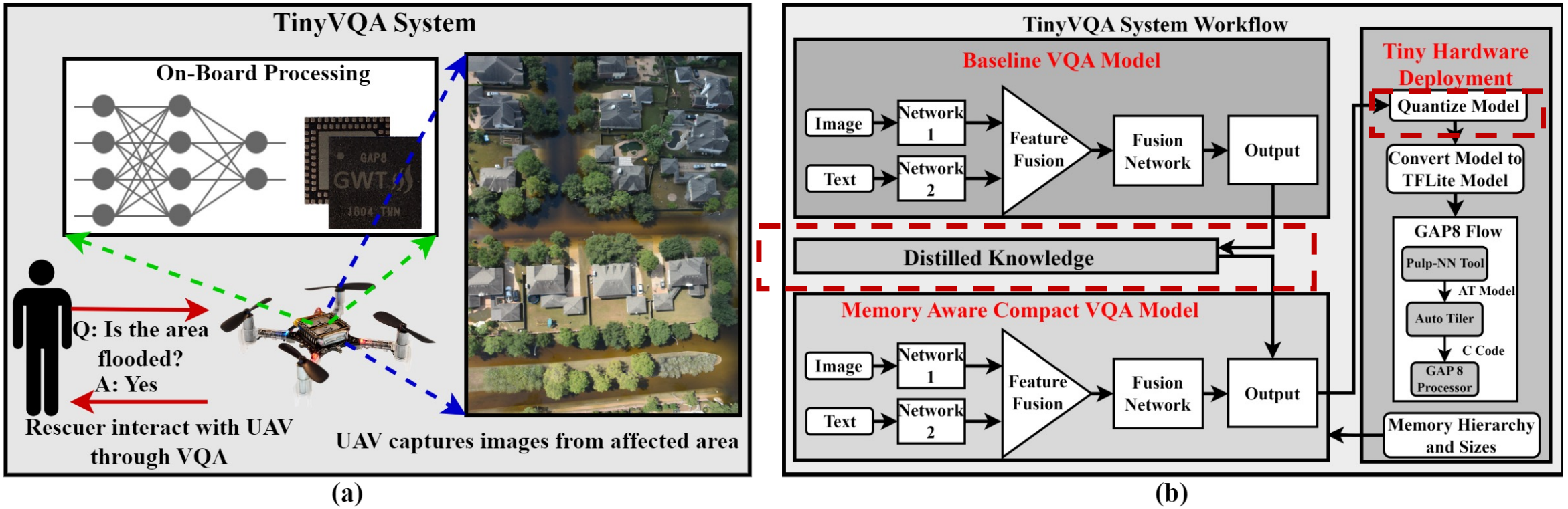




Memory Hierarchies of Different Off-the-Shelf Edge Devices

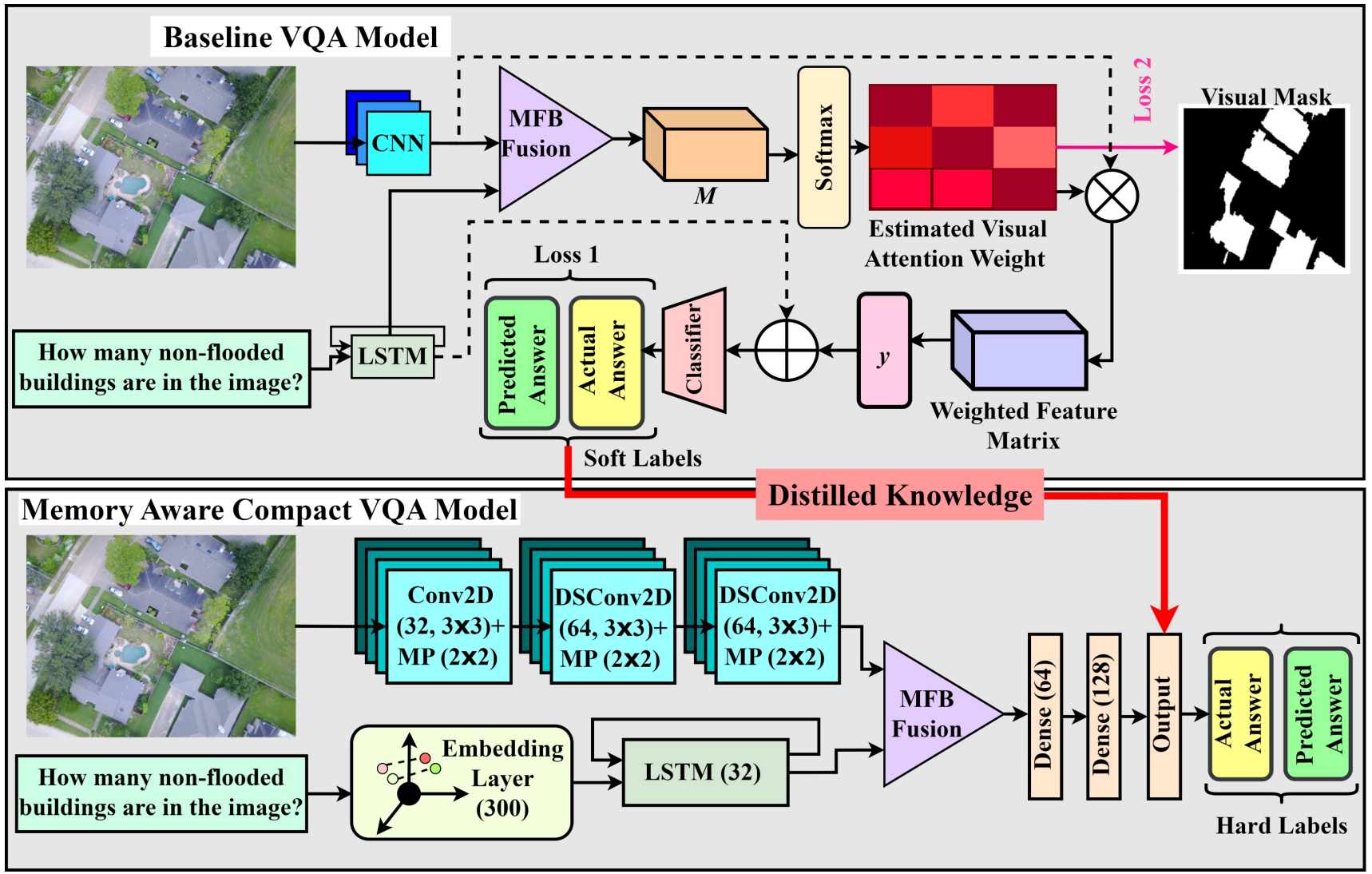
TinyVQA: Enabling On-device Visual Question Answering

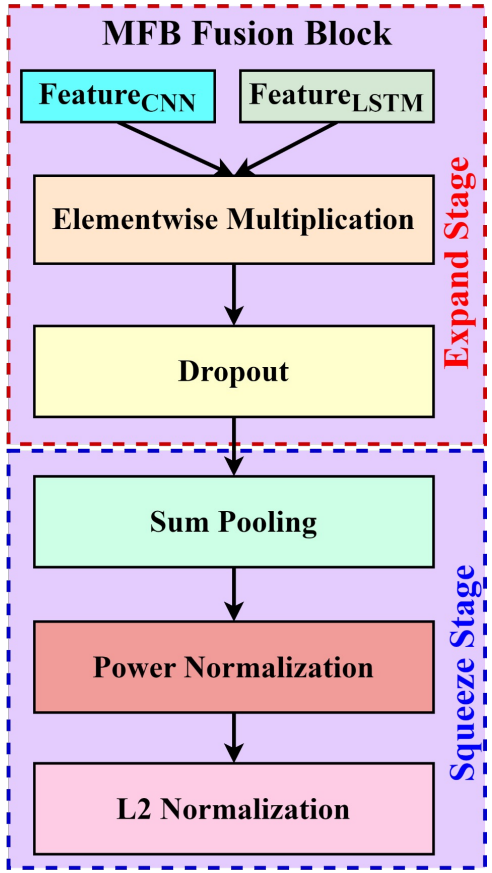
Overview of TinyVQA System



- Highlevel overview of proposed TinyVQA system.
  - Rescuer can acquire effective information about the affected area by asking questions when a drone coupled with a VQA system captures images from the hurricane-stricken area from a high altitude.
- The flow diagram of proposed TinyVQA system.
  - Proposed TinyVQA is the sequential combination of the steps shown in the diagram.

TinyVQA: Proposed Compact Model Architecture





Reference: Sarkar, Argho, et al. "Sam-vqa: Supervised attention-based visual question answering model for post-disaster damage assessment on remote sensing imagery." IEEE Transactions on Geoscience and Remote Sensing (2023).

Evaluation Dataset

DATA COLLECTION

Hurricane Harvey

Ford Bent County, TX

1. IMAGES COLLECTED FROM A HEIGHT OF **200FT**

2. SPATIAL RESOLUTION IS **1.5 CM**

	STATISTICS
#IMAGES	2345
#QUESTIONS	10,480
#QUESTION-TYPES	7
HIGHEST LENGTH OF QUESTION	11
#ANSWRS	49

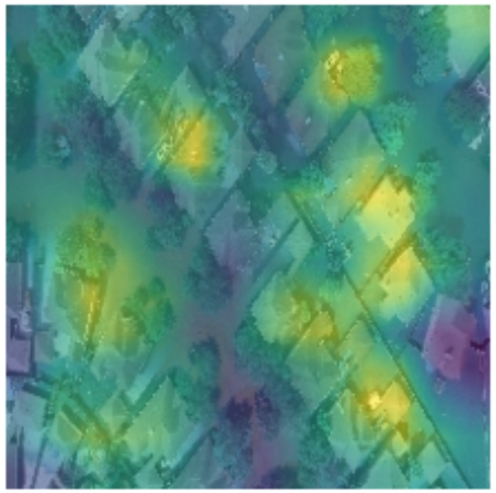
FloodNet-VQA Dataset Generation

Maryam Rahnemoonfar, Tashnim Chowdhury, **Argho Sarkar** et.al., FloodNet: A High-Resolution Aerial Imagery Dataset for Post Flood Scene Understanding, IEEE Access 2021

Evaluation Results

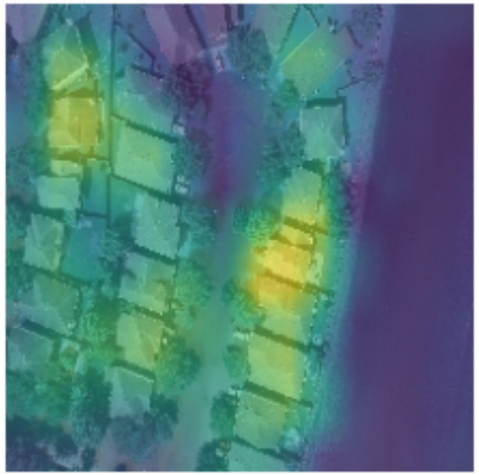
Inputs

TinyVQA



Q: What is the level of density in this image?

Ground-truth: High | Predicted: High



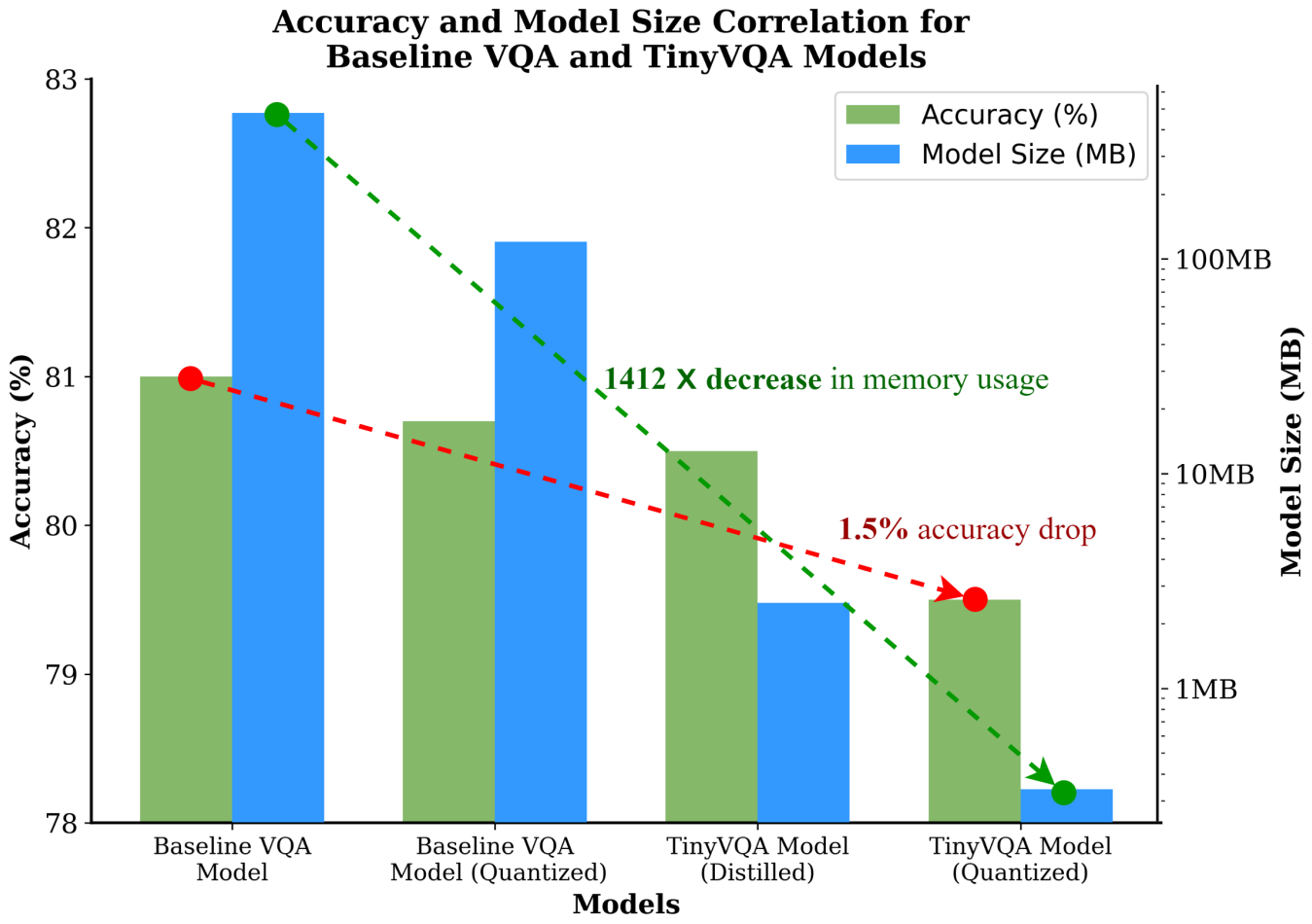
Q: Do the rescuers need to provide help urgently?

Ground-truth: Yes | Predicted: Yes

- Derived visual attentions for given questions from TinyVQA model.
- The yellowish tone in the image denotes higher attention weight.
- Attention learned with visual supervision emphasizes the relevant image portions (buildings and roads in this case) to address the questions from the top and bottom images.

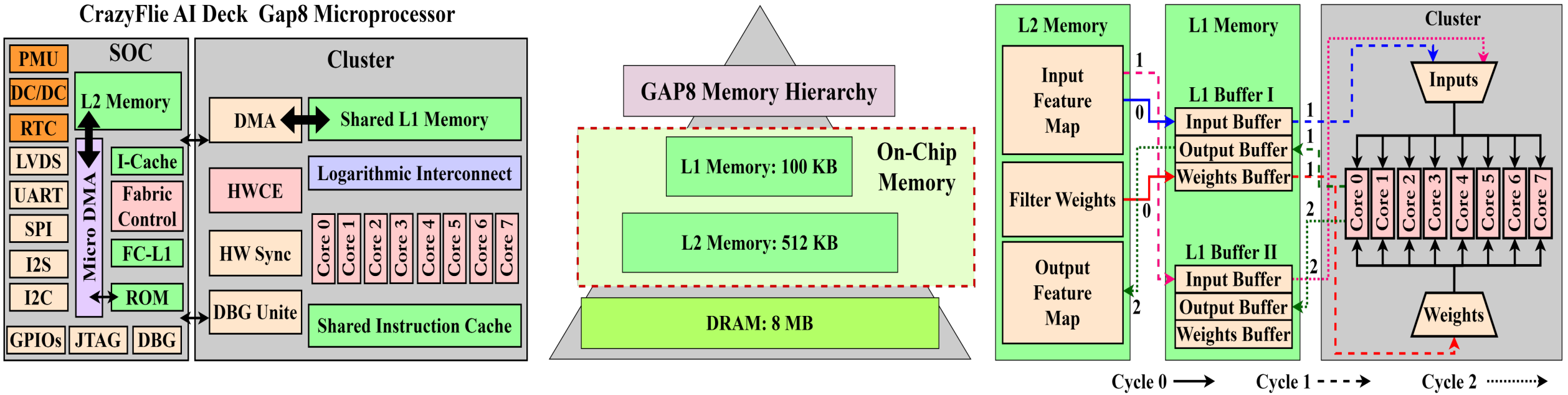
Reference: Sarkar, Argho, et al. "Sam-vqa: Supervised attention-based visual question answering model for post-disaster damage assessment on remote sensing imagery." IEEE Transactions on Geoscience and Remote Sensing (2023).

Evaluation Results



- Accuracy and Model Size Correlation for Baseline VQA and TinyVQA for FloodNet dataset.
  - Baseline model achieved 81% accuracy with 479 MB model size
  - TinyVQA model achieved 79.5% accuracy with 339 KB model size.

Tiny Hardware of Choice: Gap8 Processor



- Detailed block diagram of Crazyflie AI-deck powered by GAP8 microprocessor
- Memory Hierarchy for GAP8 microprocessor:
  - GAP 8 microprocessor has L1 Memory of 100 KB
    - 80 KB shared in compute engine
    - 20 KB for low power MCU
  - L2 memory of 512 KB and 8MB of DRAM
- TinyVQA flow:
  - Left: the DMA manages L2 -L1 communication using double-buffering.
  - Right: the cluster executes PULP-NN on tile stored in one of the L1 buffers.

TinyVQA Deployment on Tiny Drone

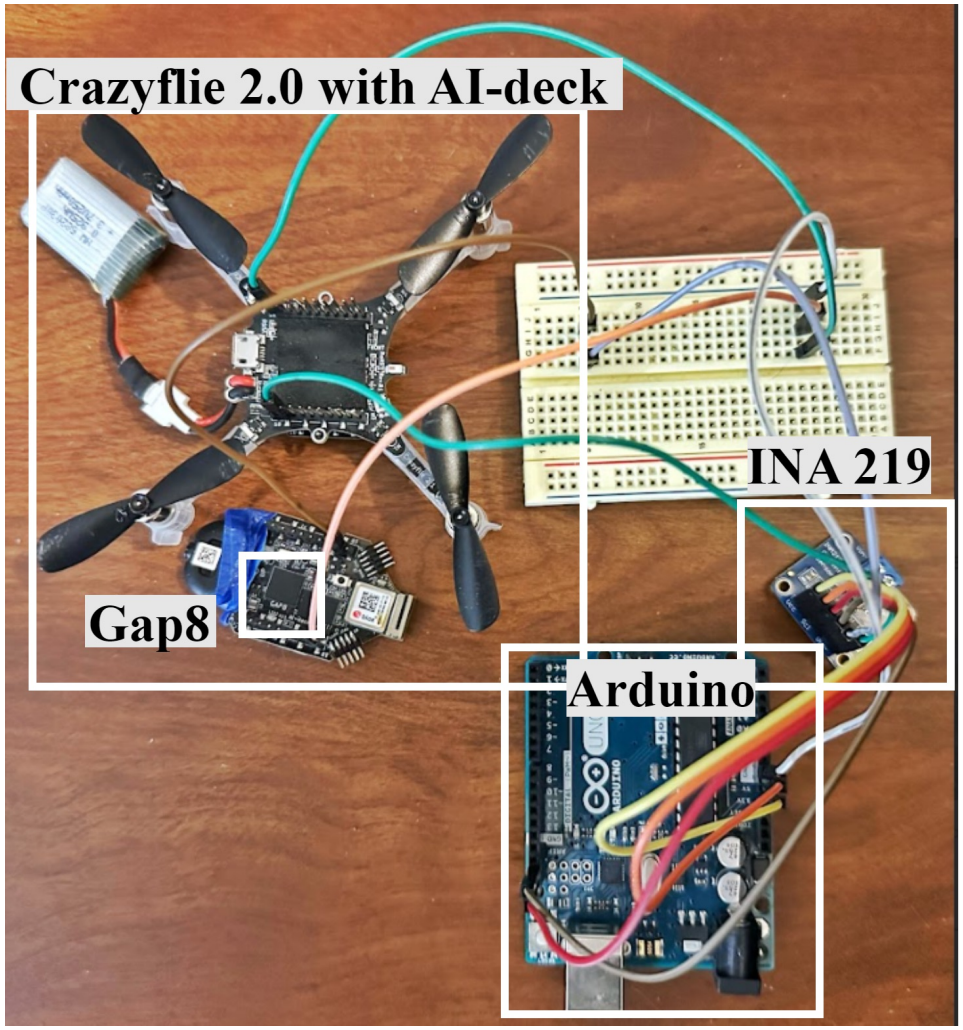


Table 1: Resource utilization data of TinyVQA implemented on GAP8 Processor

Resources	L1 Memory	L2 Memory	DRAM
Available for Use (KB)	52.7	400	8000
TinyVQA Utilization (KB)	49 (93%)	290 (73%)	0

Table 2: Implementation Results of the proposed TinyVQA and Comparison with Previous Work.

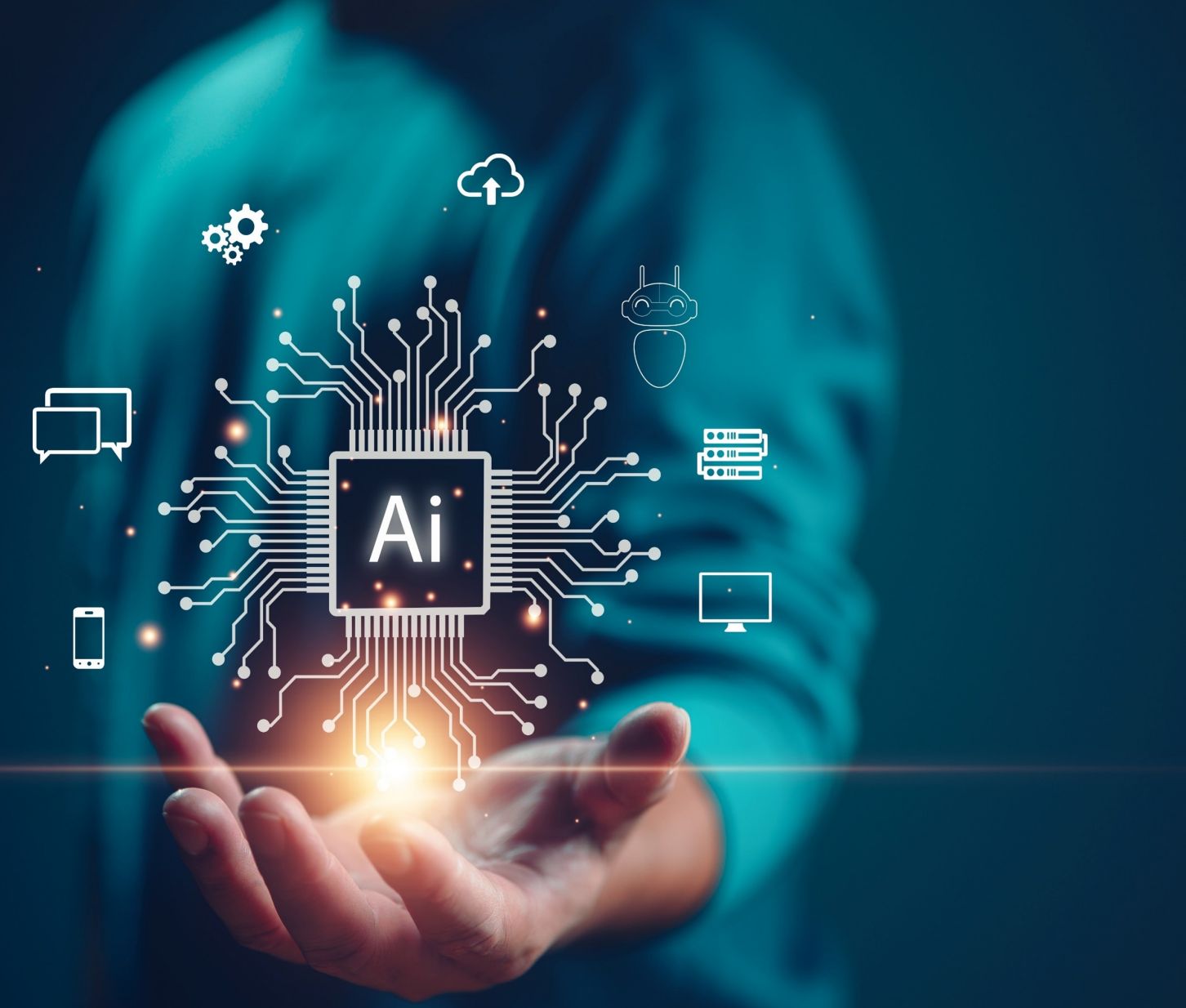
Architecture	TinyVQA (this work)	MobiVQA [2]
Dataset	FloodNet [6]	VQAv2 [3]
Modality Used	Image + Text	Image + Text
Deployment Devices	Gap8 Processor	Nvidia TX2 Board
Frequency (MHz)	175	–
Latency (ms)	56	213
Power (W)	0.7	–
Energy (J)	0.2	5.6

Conclusion

- TinyVQA integrates multiple data modalities into compact, low-power devices.
- Achieved 79.5% accuracy on FloodNet dataset, proving effectiveness in post-disaster scenarios.
- Deployment on power-efficient Crazyflie 2.0 drone with AI deck and GAP8 microprocessor.
- Operational proficiency exemplified with 56 ms latency and 0.7 W power consumption.
- Opens new avenues for autonomous, intelligent systems in critical, resource-limited environments.

thank you!

Contact: hrashid4@jhu.edu



TINY ML

Copyright Notice

www.tinyml.org