Improving the Energy Efficiency and Robustness of tinyML Computer Vision Using Log-Gradient Input Images

Qianyun "Savy" Lu, Boris Murmann
Stanford University, CA 94305, United States

Conventional Computer Vision Pipeline

Image Sensor → RAW → Image Signal Processor (ISP) → JPEG → CNN or Other Algorithms

Mono or Bayer → Multi stages, e.g. • Demosaicking • White balance → • Classification • Detection • Segmentation

Energy hungry, designed for human perception
Mismatches: tiny → power ML → machine

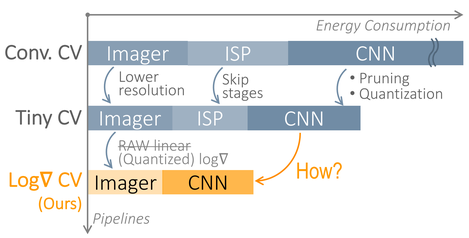
Solution: Improve each part

Lower resolution → Skip ISP stages → • Pruning • Quantization • Architecture search

Proposed Log∇ Computer Vision Pipeline

Log∇ Imager → log∇ → CNN

Hardware-algorithm codesign
• log∇ imager (prior work [1])
• no ISP (end-to-end learning)



0 log∇ images ~~log∇ sensor~~ compute log∇ from RAW
1 smaller CNN
A Architecture search → model size, memory usage, MACs
B Fixed architectures → filter redundancy, prunability, illumination robustness

0 Compute log∇ images from RAW

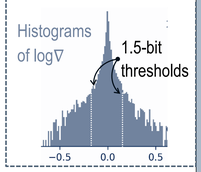
Image $P \in \mathbb{R}^{H \times W}$

- Take log $P' = \log(P + 1)$
- Take gradients: $\log \nabla = P' * f$

$$f = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

- Quantization: Empirical thresholds

For 16-bit RAW, log∇ ∈ (-22.2 22.2)



Hardware efficiency: Use a ratio-to-digital converter (RDC), drop 'log':

$$(\log \nabla)_{j,k} = \log P_{j,k-1} - \log P_{j,k+1} = \log\left(\frac{P_{j,k-1}}{P_{j,k+1}}\right) \approx Q\left(\frac{P_{j,k-1}}{P_{j,k+1}}\right)$$

Datasets

PASCAL RAW 2014

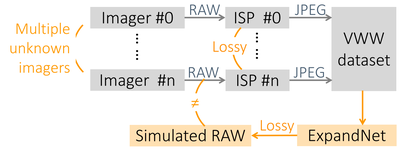
6550 images, demosaicked grayscale
3 classes: {bicycle, car, person}
One sensor, Nikon D3200 DSLR

Nikon DSLR D3200 → 16-bit RAW → PASCAL RAW 2014

Visual Wake Words (VWW)

123k+ RGB images in 2 classes: {background, person}

Why not VWW? Lossy JPEG images & large variance in (unknown) camera characteristics

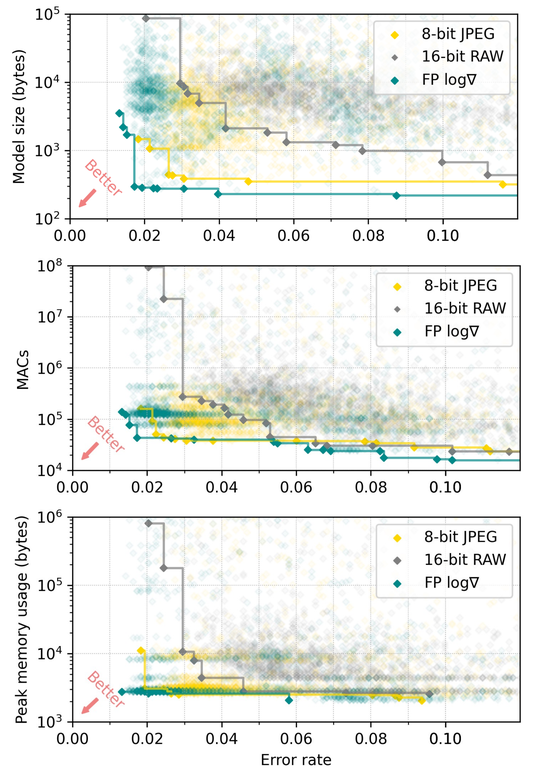


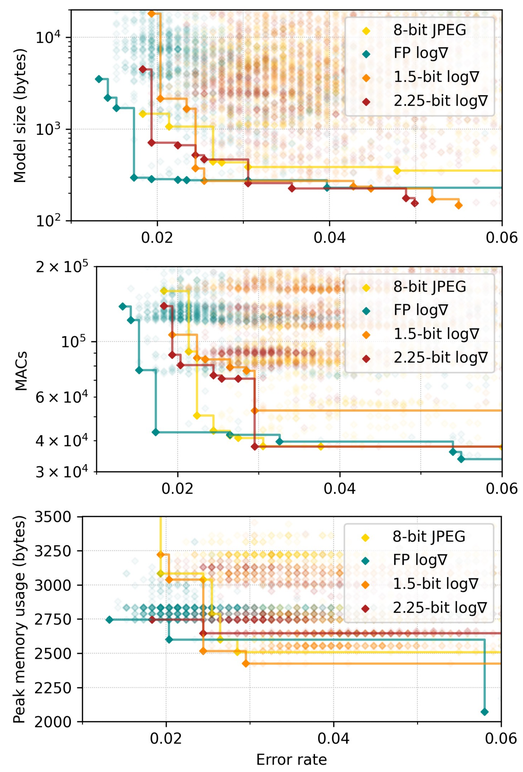
Experiment A CNN Architecture Search: µNAS

µNAS → for microcontrollers

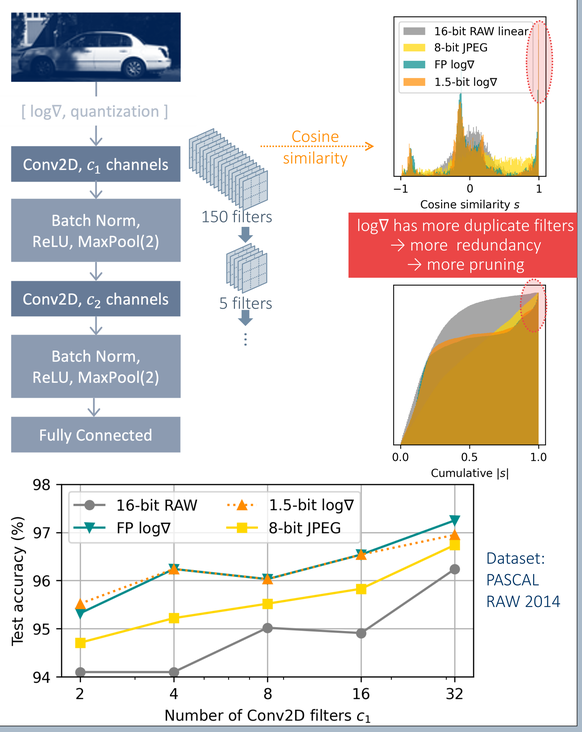
Set bounds: model size, memory usage, MACs

8-bit JPEG → µNAS
log∇ → µNAS





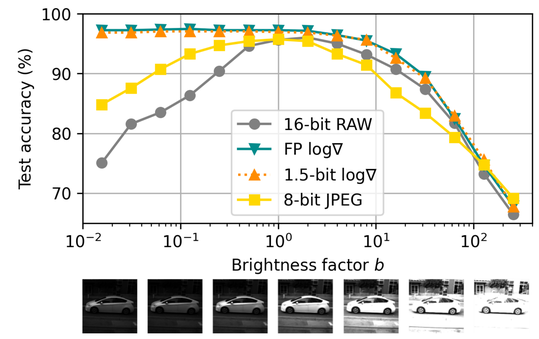
Experiment B Fixed CNN Architectures



log∇ has more duplicate filters
→ more redundancy
→ more pruning

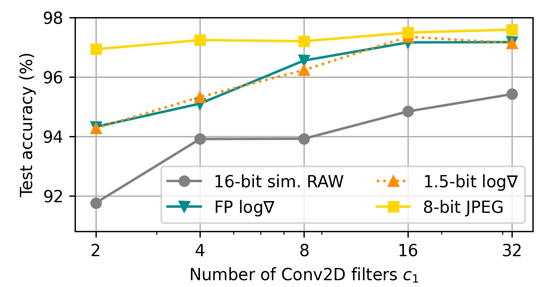
Dataset: PASCAL RAW 2014

Robustness Against Illumination Changes



VWW Results

Use simulated RAW images produced by ExpandNet.



Simulated RAW (lossy) ≠ RAW

Future Work

- Quantized training
- Threshold search
- Adversarial robustness

Acknowledgements & References

This work was supported in part by the ACCESS AI Chip Center for Emerging Smart Systems, sponsored by InnoHK funding, Hong Kong SAR. Qianyun Lu was supported by Stanford Graduate Fellowship in Science & Engineering.

[1] C. Young, A. Omid-Zohoor, P. Lajevardi and B. Murmann. "A data-compressive 1.5/2.75-bit log-gradient QVGA image sensor with multi-scale readout for always-on object detection," IEEE JSSC, vol. 54, no. 11, pp. 2932-2946, Nov. 2019.

[2] A. Omid-Zohoor, D. Ta, and B. Murmann. 2014-2015. PASCALRAW: Raw Image Database for Object Detection.

[3] A. Chowdhery, P. Warden, J. Shlens, A. Howard, and R. Rhodes. 2019. Visual Wake Words Dataset. arXiv:1906.05721 [cs.CV]

[4] D. Marnerides, T. Bashford-Rogers, J. Hatchett, and K. Debattista. 2018. ExpandNet: A deep convolutional neural network for high dynamic range expansion from low dynamic range content. In Computer Graphics Forum, Vol. 37. Wiley Online Library, 37–49.

[5] E. Liberis, Ł. Dudziak, and N. D Lane. 2021. uNAS: Constrained neural architecture search for microcontrollers. In Proceedings of the 1st Workshop on Machine Learning and Systems. 70–79.