[Video Link]

Electrical and Computer Engineering B.S. 2026
Tyler Lash
Worcester Polytechnic Institute

Evaluating Privacy-Oriented Image Encoding Methods for Embedded Vision

Motivation

Embedded Constraints

1

Compute

- -Small models
- -Shallow pipelines

2

Memory

- -Limited SRAM/Flash
- -Compact inputs
- -Limited features

3

Power

- -Energy per inference
- -Batteries

4

Bandwidth

- -Near-sensor
- -Compression

Question:
How much visual information can be removed?

Raw Images: Issues

1

Dimensionality

Raw pixels scale poorly with resolution and color channels:

-More compute
-More memory
-Sample complexity

2

Privacy

Full-fidelity images expose unnecessary visual information:

-Harmful detail
-Identifiable info

3

Processing

Learning from raw pixels with tight resource budgets:

-Unstructured
-Learn from scratch
-Not efficient

Motivation:
Reduce input complexity while preserving task-relevant structure

Related Works

	Near-Sensor	Classic encodings	Privacy vs. Fidelity
Objectives	→ Reduce data movement and on-device compute → Enable real-time inference on constrained hardware	→ Preserve task-relevant structure → Reduce input dimensionality and redundancy	→ Limit exposure of sensitive visual detail → Balance recognition accuracy and privacy
Experiments	→ Early image transformations (gradient, downsampling) → Sensor-adjacent preprocessing for tinyML	→ Downsampling, quantization, smoothing → Edge and gradient-based representations	→ Resolution reduction and obfuscation → Input degradation as a privacy mechanism
Deliverables	→ Lower latency and power consumption → Improved efficiency without model scaling	→ Compact feature representations → Shape- and structure-focused inputs	→ Reduced identifiability → Quantified privacy–utility trade-offs

Methodology

1

Fixed Model + Pipeline

Same CNN architecture, optimizer, epochs, and data splits across all experiments: isolates effect of input encoding

2

Varying Inputs

Only input representation changes (raw, downsampled, quantized, blurred, noisy, edge-based)

3

Multiple Datasets

Benchmarks + real embedded data:
-MNIST
-CIFAR-10
-ESP32-S3[1] hand gestures

4

Generalization

Analyze accuracy **and** training dynamics (train vs. validation loss, stability, overfitting)

[1] Seeed Studio XIAO ESP32-S3 Sense equipped w/ 8MB flash, 8MB PSRAM, OV3660 camera module, 1GB MicroSD

Experimental Pipeline

A standardized pipeline is used:

1 Image

- -MNIST / CIFAR-10
- -ESP32-S3 hand gestures
- -Raw captured images

No task-specific preprocessing

2 Encoding

- -Downsampling
- -Quantization
- -Blur / Noise
- -Sobel edges

Each applied offline sequentially

3 CNN

- -Fixed architecture
- -Fixed training parameters
- -Same epochs, same optimizer

No per-encoder tuning

4 Prediction

- -Accuracy
- -Loss
- -Training dynamics

Direct comparison across encodings

Encodings

Quantization

Decreases precision

Source: Wikimedia Commons, "Dithering example (undithered 16-color palette)", CC BY-SA 3.0

Blur

Smoothes details

Source: CorelDRAW, "Gaussian Blur"

Noise

Random variations

Source: Tudor Barbu, "Variational Image Denoising Approach with Diffusion Porous Media Flow"

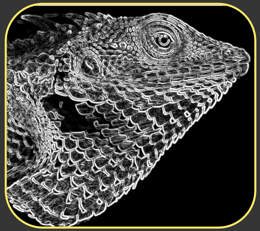
Downsampling

Reduces resolution

Source: GIASSA.NET, "Downsampling"

Sobel Edges

Shapes/boundaries



Source: Wikimedia Commons, "Lizard Canny Edge Detector Intensity Gradient", CC BY-SA 3.0

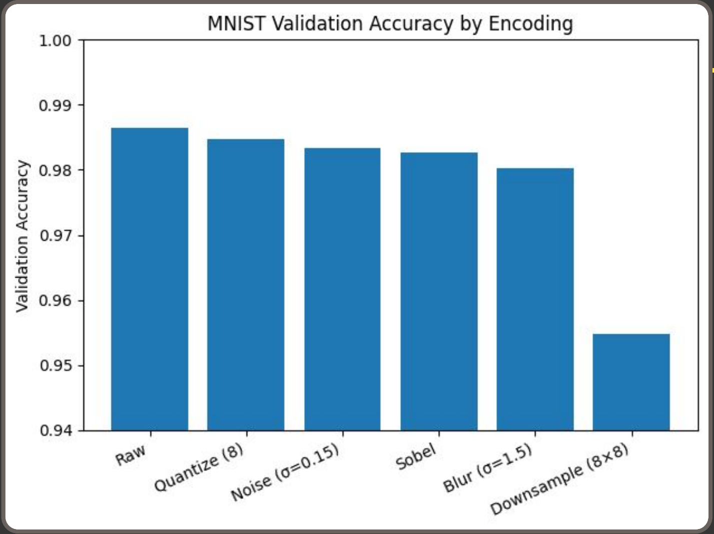
Datasets

	MNIST	CIFAR-10	Hand Gestures
What:	→ Grayscale handwritten digits (28×28) → Low visual complexity	→ Natural RGB images → 10 object classes (32×32)	→ Custom dataset captured on ESP32-S3 Sense → Limited resolution, small dataset size
Why:	→ Baseline / best-case scenario for aggressive encodings	→ Standard benchmark for texture / color-dependent machine vision	→ Realistic embedded vision scenario
Tests:	→ Shape preservation despite extreme information reduction	→ Sensitivity to loss of texture, color, fine detail	→ Shape-dominated tasks under data & hardware constraints

Results

MNIST

- All encodings preserve very high accuracy
- Shape dominates digit recognition
- Downsampling causes the largest drop

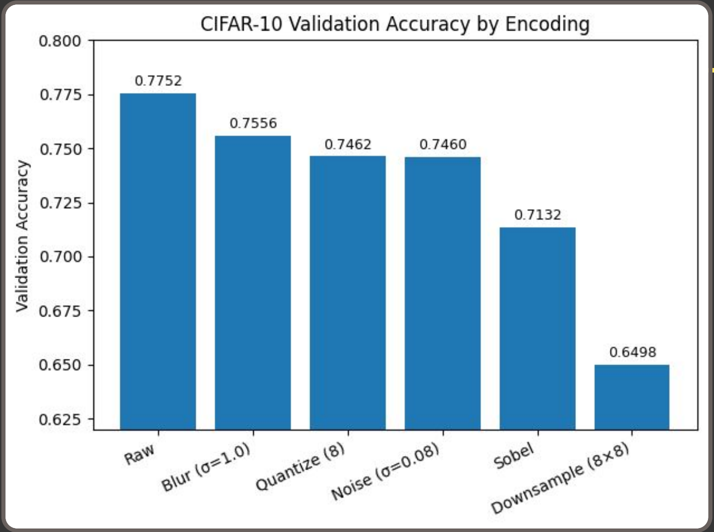


Encoding	Val. Accuracy	Val. Loss
Raw	0.9865	0.0460
Quantize (8)	0.9847	0.0550
Noise (σ=0.15)	0.9833	0.0518
Sobel	0.9827	0.0539
Blur (σ=1.5)	0.9803	0.0636
Downsample (8×8)	0.9548	0.1439

Results

CIFAR-10

- All encodings reduce accuracy vs. raw
- Texture and color removal has accuracy costs
- Sobel + downsampling are worst performers

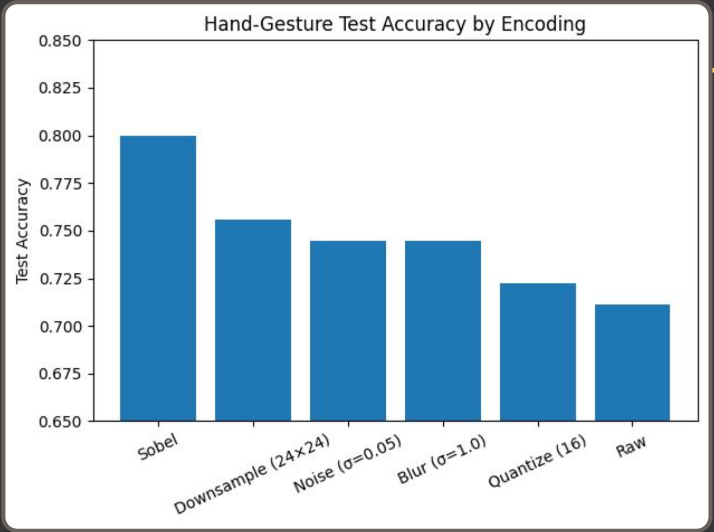


Encoding	Val. Accuracy	Val. Loss
Raw	0.7752	0.8384
Blur (σ=1.0)	0.7556	0.8557
Quantize (8)	0.7462	1.0120
Noise (σ=0.08)	0.7460	0.7802
Sobel edges	0.7132	1.1098
Downsample (8×8)	0.6498	1.0899

Results

Hand Gestures

- Averaged across 3 seeds
- Edge-based encodings (Sobel, Downsampling) perform best
- Raw images perform WORST
- Shape dominates appearance

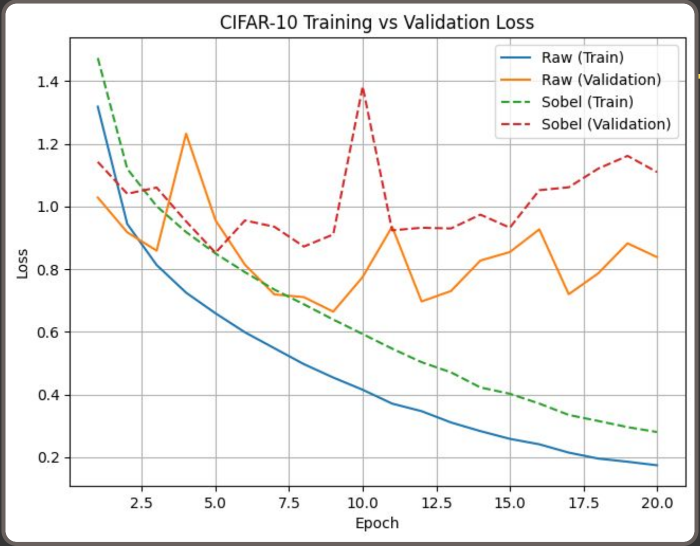


Encoding	Test Accuracy	Test Loss
Sobel edges	0.8000±0.0667	1.0644±0.5890
Downsample (24x24)	0.7556±0.0839	1.2495±0.6690
Noise (σ=0.05)	0.7444±0.0839	0.9233±0.2275
Blur (σ=1.00)	0.7444±0.1018	1.1643±0.6148
Quantize (16)	0.7222±0.0839	1.3394±0.6255
Raw	0.7111±0.0839	0.8521±0.5229

Results

Training vs. Validation Loss

- CIFAR-10 validation loss plateaus early while training loss continues to decrease
- Large generalization gap seen in both raw **and** encoded inputs
- **Encoding affects how models fail or succeed, not *just* final accuracy**



Encoding	Final Train Loss	Final Val Loss	Generalization Gap
Raw	0.17	0.84	0.67
Sobel	0.28	1.11	0.83

Key Findings

1	Reduced-Fidelity Inputs Can Match or Exceed Raw Images	**Efficiency without accuracy loss:** Several encodings achieved accuracy comparable to raw images, indicating that full pixel fidelity is often unnecessary for classification.
2	Encodings Preserving Structure Generalize Best	**Edges over appearance:** Sobel edge inputs consistently performed well, especially for hand gestures, by emphasizing shapes while suppressing confusing features.
3	Dataset Complexity Informs Encoding Effectiveness	**Effectiveness is dependent on dataset:** Simple datasets like MNIST handle most encodings well, whereas more complex or noisy datasets can *benefit* from information reduction.

Conclusion

Privacy, efficiency, and accuracy can be optimized simultaneously.

Raw images

are not strictly needed for embedded vision tasks.

Encoding is a design choice and a practical consideration.

Carefully choosing

encodings can preserve predictive features while reducing cost.

Preprocessing enables practical deployment on microcontrollers.

Structure-focused

representations are particularly effective given constrained data.

Future Work

1	**On-Device Deployment**	Implement selected encodings directly on the ESP32-S3 and evaluate end-to-end inference behavior.
2	**Efficiency Evaluation**	Measure latency, memory usage, and energy consumption for each encoding to quantify real embedded tradeoffs.
3	**Adaptive & Learned Encodings**	Explore task-aware or learned encoding schemes that aim to dynamically balance the issues of privacy, efficiency, and accuracy.

References

[1] Lu, Q., & Murmann, B. (2023). Enhancing the energy efficiency and robustness of tinyML computer vision using coarsely-quantized log-gradient input images. ACM Transactions on Embedded Computing Systems, 22(5), 1–23. https://doi.org/10.1145/3591466.

[2] Fabre, A., et al. (2024). From near-sensor to in-sensor: A state-of-the-art review of embedded AI vision systems. Sensors, 24(16), 5446. https://doi.org/10.3390/s24165446.

[3] Gonzalez, R. C., & Woods, R. E. (2018). Digital image processing (4th ed.). Pearson.

[4] Lindeberg, T. (1994). Scale-space theory: A basic tool for analysing structures at different scales. Journal of Applied Statistics, 21(2), 225–270. https://doi.org/10.1080/757582976.

[5] Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 886–893). https://doi.org/10.1109/CVPR.2005.177.

[6] Ryoo, M. S., Rothrock, B., Fleming, C., & Yang, H. J. (2017). Privacy-preserving human activity recognition from extreme low resolution. In Proceedings of the AAAI Conference on Artificial Intelligence (pp. 4255–4262).

[7] Wang, Y., Cheng, Z., Yi, X., Kong, Y., Wang, X., Xu, X., Yan, Y., Yu, C., Patel, S., & Shi, Y. (2023). Modeling the trade-off of privacy preservation and activity recognition on low-resolution images. arXiv preprint arXiv:2303.10435.

[8] McPherson, R., Shokri, R., & Shmatikov, V. (2016). Defeating image obfuscation with deep learning. arXiv preprint arXiv:1609.00408.

[9] Ra, M.-R., Govindan, R., & Ortega, A. (2013). P3: Toward privacy-preserving photo sharing. In Proceedings of the USENIX Symposium on Networked Systems Design and Implementation (NSDI).