HoloCode: Hybrid Optical-Electronic Edge Encoding for Privacy-Preserving Cloud Training

Ruofan Xing[1] Arman Akbari[1] Weikai Lin[2] Adith Boloor[3] Alexander Montes McNeil[1,4]
Michael Moebius[5] Yongmin Liu[1] Yuhao Zhu[2] Xuan Zhang[1]
[1]Northeastern University [2]University of Rochester
[3]Washington University in St. Louis [4]Draper Scholar, Draper [5]Draper
{xing.ruo, akbari.ar, montesmcneil.a, y.liu, xuan.zhang}@northeastern.edu
{wlin33, yzhu}@rochester.edu, adith@wustl.edu, mmoebius@draper.com

Abstract

*Privacy-preserving machine learning defends against adversaries without sacrificing task accuracy. In latency-critical, resource-constrained settings, existing cryptographic and encoding approaches incur heavy overheads, causing intolerable delays and energy costs. We present **HoloCode**, a hybrid optical–electronic pipeline delivering strong privacy with sub-5ms latency at a fraction of state-of-the-art energy. HoloCode encodes task-relevant signals, shields sensitive features, resists inversion attacks, and locks models with a private key preventing misuse. It builds on an edge–cloud framework pushing inference to the edge to cut latency, at the cost of higher edge energy. To break this, we leverage zero-energy optical processing to reduce latency and energy simultaneously. Against strong baselines, HoloCode achieves **10** × faster inference and **50%** lower edge energy, preserving accuracy while resisting leakage and reconstruction attacks.*

1. Introduction

Edge devices generate data improving models but contain sensitive info—medical readings, faces, or private locations. Cloud training leverages this, yet transmitting raw inputs risks privacy amidst tighter regulations and breaches.

Collaborative systems risk leakage, man-in-the-middle attacks [3, 4], and untrusted providers. Risks are acute in latency-critical applications: autonomous robots need 33–100ms/frame (10–20W) [19]; drones sub-2ms [2]; AR/VR 5–20ms (1.59mJ/inference) [17].

HE [7] and MPC [16] offer guarantees but amplify overhead [13, 23]. Instance encoding [3, 11] explores linear transformations [15, 22] but fails to hide private features) and neural encoders [10] but imposes edge overhead.

Privacy-preserving optics [1] offer light-speed, near-zero-energy hardware privacy. Traditional filtering [20, 21] contrasts with end-to-end designs [18] optimizing parameters. Yet, optics are inverted given parameter access [4].

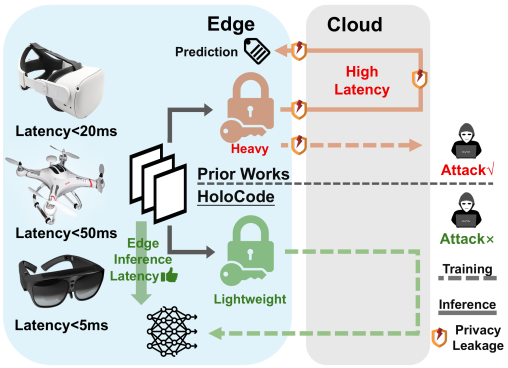

Figure 1. Motivation for HoloCode: lightweight, private edge inference.

Figure 1 illustrates existing failures. We propose HoloCode, a hybrid optical–electronic framework. Zero-energy optics [14] handle intensive operations; the digital domain handles mixing. This locally encodes utility-aligned, noninvertible, and key-protected models.

Evaluation demonstrates superior trade-offs. It achieves sub-5ms latency (85% reduction) and 70% energy savings with $< 3\%$ accuracy degradation. Contributions include: (1) a method achieving noninvertibility and key protection; (2) a hybrid protocol for secure collaboration; (3) evaluation showing superior trade-offs (SSIM=0.25, PSNR=9.78, LPIPS=0.64); and (4) system characterization showing **5**× **faster inference** and **8**× **better efficiency**.

2. System Model & Characterization

We detail our system (§2.1) and threat models (§**??**), followed by characterization of limitations (§2.2).

2.1. System Overview

We assume trusted operator manages edge devices, while the cloud and network are exposed to adversaries.

Edge-only inference serves utility tasks locally using pre-trained models. While preserving privacy, it lacks domain adaptation on fresh data, degrading performance (baseline).

Collaborative inference (HE, instance encoding) assumes constrained edges. Local data is encrypted/encoded and transmitted to the cloud, with labels returned to edge [1, 5].

Separated training/inference, common in linear obfuscation [22], splits operations: ***Offline Training*** initializes a model on public data in a trusted environment. ***Online Training*** sends encoded data to an untrusted cloud to update the utility model, returning it to the edge. ***Local Inference*** executes the updated model on encoded data locally.

Threat Model. Edge devices capture dataset D and locally encode it to D'. Adversaries may access D' and utility model $\mathcal{N}$. We encode D' to prevent D's reconstruction or sensitive inference. Furthermore, accessing $\mathcal{N}$ must not enable inference on raw D, preventing exploitation.

2.2. Systems Characterization

We characterize latency and energy for four solutions: Poseidon (HE) [23], Enc^2 (hybrid) [5], learnable obfuscation (LO) [22], and FCRL (ARL) [9].

We encode a single 224×224 grayscale image for ResNet-18. FHE, Enc^2, and ARL use collaborative inference; FCRL runs the first residual block locally. Latency sums computation and communication. LO performs local mixing, projection, noise injection, and inference.

Local computations are estimated via a systolic array simulator [6] . Cloud execution uses a dual Xeon/L40S server. Total latency sums edge encoding (chip cycles), transmission, and cloud inference. Energy derives from simulator operation counts. For encryption (Poseidon, Enc^2), we estimate performance using relative ratios [5].

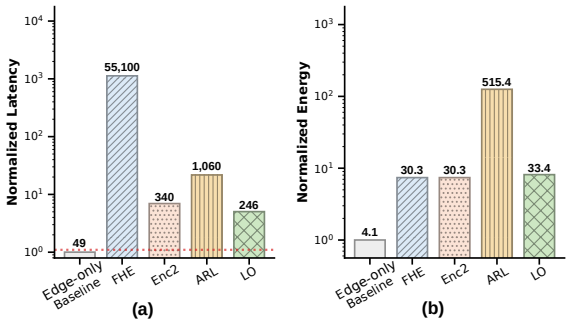

Figure 2. System characterization. (a) Edge inference latency; (b) Encoding and inference energy (log scale).

Figure 2 summarizes results. Plain inference offers lowest overhead, whereas privacy approaches incur significant costs: HE is computation-bound, Enc^2 is transmission-bound, and LO/ARL incur substantial edge encoding costs. Current solutions violate latency-critical constraints (≈50ms), motivating HoloCode's lightweight design.

3. HoloCode Framework

3.1. Framework Formulation

Consider training set $X = \{x_1, \ldots, x_n\}$, $x_i \in \mathbb{R}^d$ with labels $Y = \{y_1, \ldots, y_n\}$, $y_i \in \{\mathbf{1}_1, \ldots \mathbf{1}_c\}$, and utility model $\mathcal{N}$ with parameters θ. We optimize θ^* for utility while designing transformations $\mathcal{F}(X)$ and $\mathcal{F}(Y)$ achieving a favorable privacy–utility trade-off. As shown in Figure 3, this includes a DONN encoder, private-key optical projection, digital data mixing, and noise injection.

Following [22], we generate a per-user linear projection matrix $K \in \mathbb{R}^{d \times d_0}$, a data mixing matrix $M \in \mathbb{R}^{m \times m_0}$, a private encoder with learnable weights $W \in \mathbb{R}^{d \times d}$, and Gaussian noise δ. The transformation is $\mathcal{F}(X) = MXWK + \delta$ and $\mathcal{F}(Y) = MY$. W is the only trainable component; others are randomly generated per user.

$\mathcal{F}(X)$ and $\mathcal{F}(Y)$ transmit to the cloud to train $\mathcal{N}$, and the model returns to the edge. Since encoding combines dimensionality reduction, mixing, permutation, and noise, the mapping from (X, Y) to released data is randomized and many-to-one, preventing unique recovery without keys.

3.2. Private Encoding Pipeline

Private Encoder. We learn W maximizing the privacy–utility trade-off via adversarial learning, where encoded data $\mathcal{F}(X)$ are consumed by utility model $\mathcal{N}$ and privacy model $\mathcal{P}$. Parameters W, $\theta_{\mathcal{N}}$, and $\theta_{\mathcal{P}}$ are trained with cross-entropy losses $\mathcal{L}_{\mathcal{N}}$ and $\mathcal{L}_{\mathcal{P}}$. Balancing losses preserves utility while suppressing leakage.

We co-train $\mathcal{F}$, $\mathcal{N}$, and $\mathcal{P}$ via adversarial min–max:

$$\min_{W, \theta_{\mathcal{N}}} \max_{\theta_{\mathcal{P}}} J(W, \theta_{\mathcal{N}}, \theta_{\mathcal{P}}), \quad J = \lambda_{\mathcal{N}} \mathcal{L}_{\mathcal{N}} + \lambda_{\mathcal{P}} \mathcal{L}_{\mathcal{P}}. \quad (1)$$

where maximization strengthens $\mathcal{P}$, and minimization updates W and $\theta_{\mathcal{N}}$. Through alternating updates, W converges to effectively balance privacy and utility.

Linear Projection. After cloud training, model $\mathcal{N}$ (parameters θ^*) returns for edge inference. Adversaries obtaining W and θ^* could apply the model to unauthorized inputs. To prevent misuse, we incorporate a *private key* following the Johnson–Lindenstrauss lemma [12].

Digitally, the key is a Gaussian random projection matrix K_S, with seed S determining the user's key. [Image of Johnson-Lindenstrauss lemma random projection concept] The same key applies during training and inference; mismatched keys prevent meaningful predictions. However, digital projection's high cost for high-resolution images is addressed via optical implementation.

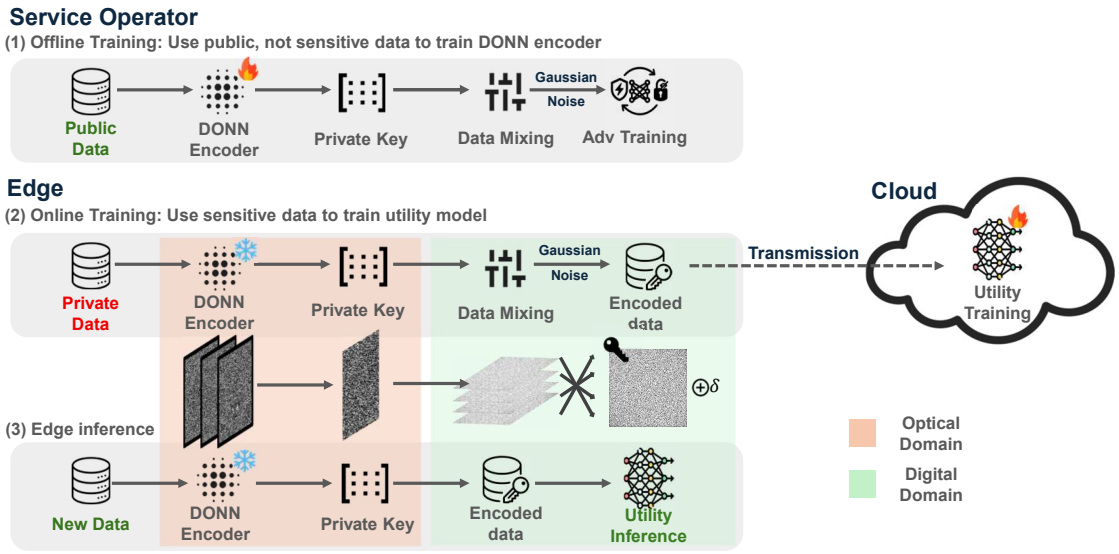

Figure 3. HoloCode overview. (1) DONN weights are trained adversarially on public data. (2) The frozen DONN and private transformations encode data for cloud training. (3) The trained model performs edge inference using encoded inputs to minimize leakage.

3.3. Hardware Implementation

DONN-based Encoder. We adopt end-to-end Lightridge framework [14] for optical modeling and hardware-aware compilation. Using Fresnel approximation, we formulate differentiable phase masks and free-space propagators, enabling detector-to-source backpropagation. This differentiability permits co-training the DONN with adversarial heads, yielding stronger privacy–utility trade-offs while preserving inherent latency and energy advantages.

Metasurface-based Linear Projection. We employ a metasurface to impose random, spatially varying phase delays on the DONN output. A fixed random seed generates a unique phase pattern (optical key K_S), scrambling spatial information before the wavefront reaches the sensor. Simulating this via the Angular Spectrum Method (ASM) [8], we implement linear projection K_S entirely in the optical domain, avoiding costly digital matrix operations.

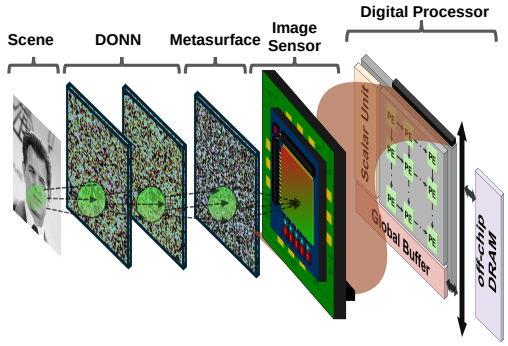

Figure 4. Hardware implementation of HoloCode

Digital-domain Data Mixing. While optical encoding suppresses sensitive features, known point spread functions (PSF) render transformations invertible [4]. To ensure non-invertibility, we introduce digital data mixing. We randomly interpolate image–label pairs for each sample. Since inputs are i.i.d., the mixed sample matches the original distribution, preventing reconstruction without mixing indices.

System Integration. As shown in Figure 4, the DONN (W) removes privacy-correlated features while retaining utility statistics. A metasurface applies optical key K_S, and the scrambled wavefront is captured by a CMOS sensor. Sensor tiles stream to on-chip SRAM, then to a digital processor with systolic PE array performing: (i) data mixing loaded from off-chip DRAM and (ii) classifier inference. Double buffering overlaps sensing, computing, and transfers.

4. Evaluation

4.1. Experiment Setup

Models. We evaluate digital and optical HoloCode on CelebA (gender/privacy, smile/utility). Digital encoding uses UNet; optical employs a DONN and metasurface. Downstream, we use ResNet–18. For reconstruction, we finetune a pretrained GAN-based Pix2Pix.

System Characterization. We evaluate latency and energy using the setup in §2.2 and a hardware simulator [6].

Baselines. We compare against Enc^2 [5] (ARL with 4 ResNet-18 layers), PrivateEye [1], and learnable obfuscation [22] (via matrix masking and mixing).

Table 1. Privacy–utility trade-off. T1 (smiling) is utility (↑), T2 (gender) is privacy (↓). Bold/underlined denote best/second best.

Method	T1 ↑	T2 ↓	PUTS ↑
Baseline	93.5	98.0	/
Ideal	93.5	57.0	1.91
Digital Methods			
Enc^2(ARL)	**92.6**	<u>58.0</u>	<u>1.62</u>
LearnableObfus	91.6	88.4	1.06
Ours (UNet-tiny)	<u>92.5</u>	**57.4**	**1.65**
Optical Methods			
PrivateEye	85.4	62.7	1.30
Ours (DONN-3)	**89.0**	**62.6**	**1.39**

4.2. Privacy-preserving Performance

Privacy-Utility Trade-off. Table 1 reports digital trade-offs. *Baseline* denotes raw accuracy; *Ideal* represents baseline utility with random-guess privacy (57%). We define Privacy–Utility Trade-off Score (PUTS) as $\frac{\mathrm{AUC}_{\mathrm{T}_p}}{\mathrm{AUC}_{\mathrm{T}_u}}$; higher values indicate better preservation.

Table 1 shows Enc^2 and our digital implementation approach the ideal, confirming adversarial learning aligns utility and privacy. Conversely, learnable obfuscation applies uniform transformations, preserving utility but failing to obscure private tasks. UNet-tiny outperforms DONN due to architectural complexity, yet DONN's small deviation indicates strong privacy. Crucially, HoloCode achieves better trade-offs than PrivateEye.

Reconstruction Attack. We quantify invertibility via SSIM, PSNR, and LPIPS. Visualizations (Figure 5) show PrivateEye and ARL are easily inverted, while Learnable Obfuscation and our method remain robust. Table 2 confirms ARL performs worst due to neural encoder susceptibility. Leveraging mixing ambiguity, Learnable Obfuscation and HoloCode achieve comparable robustness.

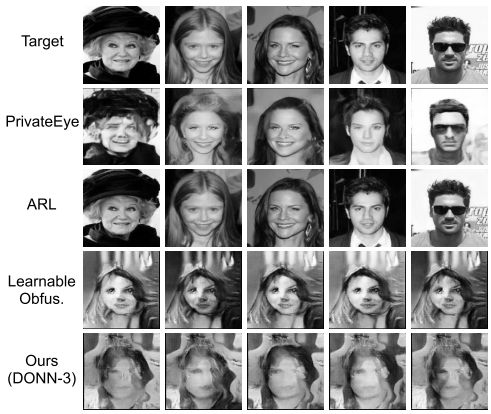

Figure 5. Visualization of reconstruction-based attacks.

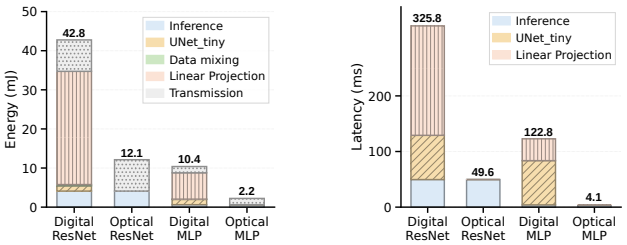

Figure 6. System latency and energy characterization breakdown and comparison with different classifiers.

Model Exploitation. We evaluate threats where adversaries intercept utility models. Assuming known architecture without keys, replacing projection K_s with K_{adv} drops utility from **92.5%** to **52.0%**, showing robustness.

4.3. System Performance

Following §2.2, we characterize edge overhead for grayscale inference. Figure 6a shows digital projection dominates energy via memory-intensive operations; optical implementation drastically reduces this. Figure 6b compares latencies; optical transformation yields **7**× speedup for intensive operations. Switching the downstream head to a compact MLP reduces both metrics. Optical–MLP stays below 50ms with lowest energy, whereas digital variants violate budgets. This confirms HoloCode flexibly tunes complexity for savings, with optical–MLP being most efficient.

4.4. Discussion

We test HoloCode on other dataset, e.g. LFWA, demonstrating identical strong performance. We perform ablation study by removing data mixing, indicating it substantially increases ambiguity, improving reconstruction resistance from 0.69 to 0.25 SSIM and 23.30 to 9.78 PSNR.

5. Conclusion

We propose a hybrid optical–electronic privacy-preserving encoding for latency-critical, resource-constrained scenarios. Using encoder, linear projection and data mixing, ensuring utility alignment, noninvertibility, and key protection. HoloCode achieves privacy-utility trade-offs comparable to SOTA methods, maintaining robustness against various attacks.

Table 2. Reconstruction attack results (SSIM↓, PSNR↓, LPIPS↑)

Method	SSIM ↓	PSNR (dB) ↓	LPIPS ↑
PrivateEye	0.60	19.02	0.32
ARL (Enc^2)	0.89	31.87	0.06
LearnableObfus	**0.23**	**9.64**	**0.64**
Ours (DONN-3)	<u>0.25</u>	<u>9.78</u>	**0.64**

References

[1] Adith Boloor, Weikai Lin, Tianrui Ma, Yu Feng, Yuhao Zhu, and Xuan Zhang. Privateeye: In-sensor privacy preservation through optical feature separation. In *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 2357–2367. IEEE, 2025. 1, 2, 3

[2] Pietro Bonazzi, Christian Vogt, Michael Jost, Lyes Khacef, Federico Paredes-Vallés, and Michele Magno. Towards low-latency event-based obstacle avoidance on a fpga-drone. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 4938–4946, 2025. 1

[3] Nicholas Carlini, Samuel Deng, Sanjam Garg, Somesh Jha, Saeed Mahloujifar, Mohammad Mahmoody, Abhradeep Thakurta, and Florian Tramèr. Is private learning possible with instance encoding? In *2021 IEEE Symposium on Security and Privacy (SP)*, pages 410–427. IEEE, 2021. 1

[4] Jiacheng Cheng, Xiang Dai, Jia Wan, Nick Antipa, and Nuno Vasconcelos. Learning a dynamic privacy-preserving camera robust to inversion attacks. In *European Conference on Computer Vision*, pages 349–367. Springer, 2024. 1, 3

[5] Hao-Jen Chien, Hossein Khalili, Amin Hass, and Nader Sehatbakhsh. Enc2: Privacy-preserving inference for tiny iots via encoding and encryption. In *Proceedings of the 29th Annual International Conference on Mobile Computing and Networking*, pages 1–16, 2023. 2, 3

[6] Yu Feng, Paul Whatmough, and Yuhao Zhu. Asv: Accelerated stereo vision system. In *Proceedings of the 52nd Annual IEEE/ACM International Symposium on Microarchitecture*, pages 643–656, 2019. 2, 3

[7] Ran Gilad-Bachrach, Nathan Dowlin, Kim Laine, Kristin Lauter, Michael Naehrig, and John Wernsing. Cryptonets: Applying neural networks to encrypted data with high throughput and accuracy. In *International conference on machine learning*, pages 201–210. PMLR, 2016. 1

[8] Joseph W Goodman. *Introduction to Fourier optics*. Roberts and Company publishers, 2005. 3

[9] Umang Gupta, Aaron M Ferber, Bistra Dilkina, and Greg Ver Steeg. Controllable guarantees for fair outcomes via contrastive information estimation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 7610–7619, 2021. 2

[10] Chong Huang, Peter Kairouz, Xiao Chen, Lalitha Sankar, and Ram Rajagopal. Generative adversarial privacy. *arXiv preprint arXiv:1807.05306*, 2018. 1

[11] Yangsibo Huang, Zhao Song, Kai Li, and Sanjeev Arora. Instahide: Instance-hiding schemes for private distributed learning. In *International conference on machine learning*, pages 4507–4518. PMLR, 2020. 1

[12] Kasper Green Larsen and Jelani Nelson. Optimality of the johnson-lindenstrauss lemma. In *2017 IEEE 58th annual symposium on foundations of computer science (FOCS)*, pages 633–638. IEEE, 2017. 2

[13] Junyi Li and Heng Huang. Faster secure data mining via distributed homomorphic encryption. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2706–2714, 2020. 1

[14] Yingjie Li, Ruiyang Chen, Minhan Lou, Berardi Sensale-Rodriguez, Weilu Gao, and Cunxi Yu. Lightridge: An end-to-end agile design framework for diffractive optical neural networks. In *Proceedings of the 28th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 4*, pages 202–218, 2023. 1, 3

[15] Zhijian Liu, Zhanghao Wu, Chuang Gan, Ligeng Zhu, and Song Han. Datamix: Efficient privacy-preserving edge-cloud inference. In *European Conference on Computer Vision*, pages 578–595. Springer, 2020. 1

[16] Payman Mohassel and Yupeng Zhang. Secureml: A system for scalable privacy-preserving machine learning. In *2017 IEEE symposium on security and privacy (SP)*, pages 19–38. IEEE, 2017. 1

[17] Julian Moosmann, Pietro Bonazzi, Yawei Li, Sizhen Bian, Philipp Mayer, Luca Benini, and Michele Magno. Ultra-efficient on-device object detection on ai-integrated smart glasses with tinyissimoyolo. In *European Conference on Computer Vision*, pages 262–280. Springer, 2024. 1

[18] Vincent Sitzmann, Steven Diamond, Yifan Peng, Xiong Dun, Stephen Boyd, Wolfgang Heidrich, Felix Heide, and Gordon Wetzstein. End-to-end optimization of optics and image processing for achromatic extended depth of field and super-resolution imaging. *ACM Transactions on Graphics (TOG)*, 37(4):1–13, 2018. 1

[19] Amir Taherin, Juyi Lin, Arash Akbari, Arman Akbari, Pu Zhao, Weiwei Chen, David Kaeli, and Yanzhi Wang. Cross-platform scaling of vision-language-action models from edge to cloud gpus, 2025. 1

[20] Zihao W Wang, Vibhav Vineet, Francesco Pittaluga, Sudipta N Sinha, Oliver Cossairt, and Sing Bing Kang. Privacy-preserving action recognition using coded aperture videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019. 1

[21] Zhenyu Wu, Haotao Wang, Zhaowen Wang, Hailin Jin, and Zhangyang Wang. Privacy-preserving deep action recognition: An adversarial learning framework and a new dataset. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(4):2126–2139, 2020. 1

[22] Hanshen Xiao, G Edward Suh, and Srinivas Devadas. Formal privacy proof of data encoding: The possibility and impossibility of learnable encryption. In *Proceedings of the 2024 on ACM SIGSAC Conference on Computer and Communications Security*, pages 1834–1848, 2024. 1, 2, 3

[23] Yinghao Yang, Huaizhi Zhang, Shengyu Fan, Hang Lu, Mingzhe Zhang, and Xiaowei Li. Poseidon: Practical homomorphic encryption accelerator. In *2023 IEEE International Symposium on High-Performance Computer Architecture (HPCA)*, pages 870–881. IEEE, 2023. 1, 2