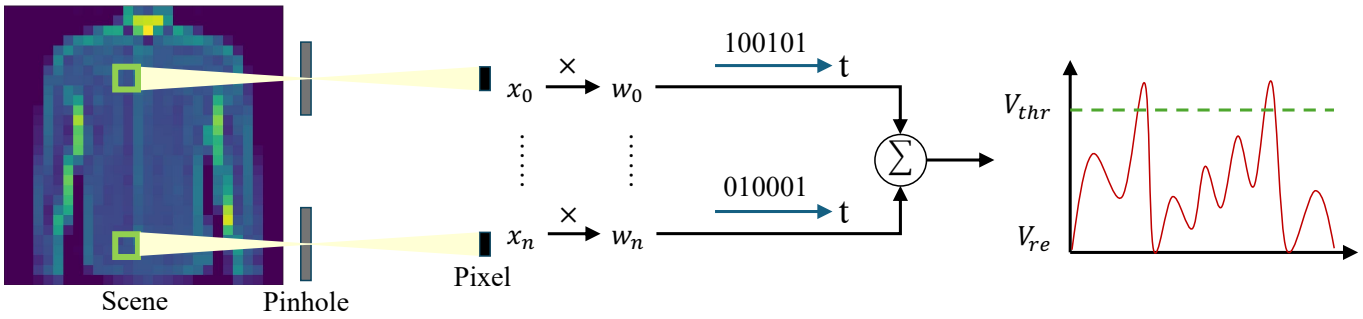
# Laconic Neuromorphic Vision

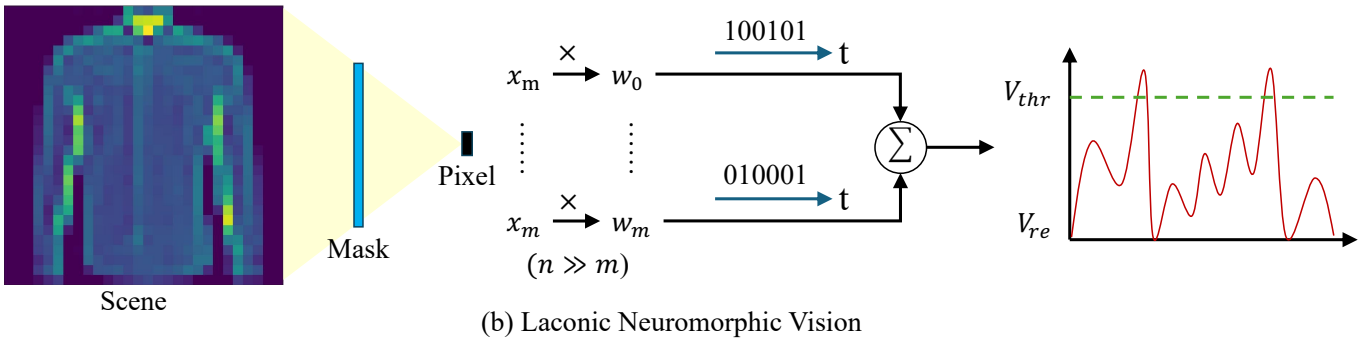
Yaping Zhao
*Department of Electrical and Electronic Engineering*
*The University of Hong Kong*
Hong Kong SAR of China
ORCID: 0000-0002-3482-0942
Email: zhaoyp@connect.hku.hk

Edmund Y. Lam*
*Department of Electrical and Electronic Engineering*
*The University of Hong Kong*
Hong Kong SAR of China
ORCID: 0000-0001-6268-950X
Email: elam@eee.hku.hk

2025 IEEE International Symposium on Circuits and Systems (ISCAS) | 979-8-3503-5683-0/25/$31.00 ©2025 IEEE | DOI: 10.1109/ISCAS56072.2025.11043673




(a) Traditional Neuromorphic Vision




(b) Laconic Neuromorphic Vision

Fig. 1: Comparison between (a) traditional neuromorphic vision, where the full scene is captured with sensors, and (b) laconic neuromorphic vision, where sparse encoding by masks is used to reduce the number of input spikes while retaining essential information for task performance.

***Abstract*—In this paper, we introduce laconic neuromorphic vision, a novel visual intelligence system that encodes scene dynamics into sparse spike representations and processes them using spiking neural networks, mimicking the sparse activation patterns of biological neurons. This approach drastically reduces the input data required while maintaining high accuracy in image classification. Experiments on the MNIST and Fashion-MNIST datasets demonstrate that our system achieves comparable results to traditional methods with significantly fewer input pixels and spikes. This work provides a fresh perspective on energy-efficient, biologically inspired vision systems, with potential applications in edge computing and autonomous systems. Code and data will be released upon manuscript acceptance.**

***Index Terms*—neuromorphic imaging, neuromorphic vision, computer vision, compressed sensing, computational imaging**

*Corresponding author.

This work is supported in part by the Research Grants Council (GRF 17201822 and TRS T45-701/22-R) of Hong Kong SAR, China.

## I. Introduction

Neuromorphic vision refers to imaging systems and neural networks that emulate the structure and function of biological brains [1]–[4]. A key approach in this area is the use of spiking neural networks (SNN) [5], [6], which are increasingly recognized for their potential in various applications. Unlike traditional artificial neural networks (ANN), where all neurons are activated across multiple layers during computations, SNN better mimics the brain's neural dynamics through sparse activation patterns. In an SNN, only a fraction of neurons are active at any given time, with neurons firing spikes in response to specific stimuli. This mirrors how biological neurons process information, where only certain neurons respond to particular inputs, making SNN much more energy-efficient and biologically realistic compared to fully-activated ANN.

The energy-efficient, event-driven nature of SNN offers distinct advantages, particularly in applications requiring low

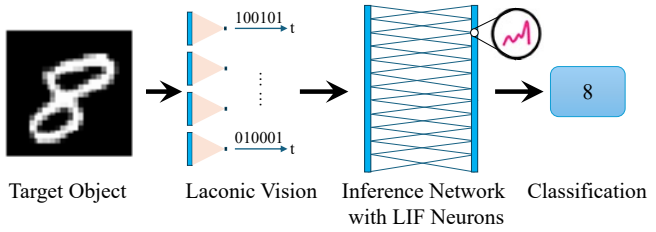



Fig. 2: Overview of the laconic neuromorphic vision. The target scene is sparsely encoded into spikes and processed by a spiking neural network (SNN) with leaky integrate-and-fire (LIF) neurons, leading to the final output.

power and real-time processing, such as mobile and edge computing environments [7]–[9]. ANN, with its large, densely connected networks, demands high computational power and energy to operate, as every neuron is involved in each computation. This leads to inefficiencies, especially when compared to SNN, which is more selective and activates neurons only when necessary. However, despite these advantages, one challenge in SNN lies in handling large-scale, high-resolution input data. Efficient encoding mechanisms are critical to prevent the computational load from overwhelming the energy-saving benefits of SNN.

While much research has focused on the sparse neural activation within SNN, relatively little attention has been paid to the sparsity of the input data itself. Current neuromorphic systems often mirror traditional vision systems by processing large, high-resolution images, requiring cumbersome architectures to manage high-dimensional inputs. This approach overlooks the possibility that the input neurons—those initially triggered by sensory information—could themselves be more sparsely encoded. Consequently, while SNN excels at energy-efficient processing, the inefficiency of handling large, redundant inputs limits its full potential in reducing computational complexity.

In this work, we propose **laconic neuromorphic vision**, a novel approach that leverages sparse input encoding to complement the sparse neuron activation of SNN. Our approach drastically reduces the number of input neurons activated by encoding the input into sparse spike representations, significantly lowering the computational burden from the outset, as shown in Figure 1. This method not only aligns closely with biological principles but also minimizes input data size without compromising task performance. As depicted in Figure 2, the system encodes the target scene into spikes, which are processed by a spiking neural network with leaky integrate-and-fire (LIF) neurons to generate the final output.

Through extensive experiments on MNIST and Fashion-MNIST, we demonstrate that even with fewer initial neurons firing, tasks such as image classification can be performed successfully, preserving energy efficiency and enhancing the scalability of neuromorphic systems for real-world applications, including autonomous systems [10] and edge computing devices [7]–[9].

## II. Related Work

Compressive imaging, such as single-pixel imaging [11]–[16] and snapshot compressive imaging [17]–[23], focuses on reconstructing dense images using computational methods like compressed sensing [24]–[26]. Our approach bypasses reconstruction, directly inferring task outputs from sparse encoding to reduce computational load while maintaining accuracy.

Neuromorphic imaging [1]–[4], often using event cameras [27]–[29], can capture scene changes asynchronously for low-latency data acquisition. We propose laconic neuromorphic vision by integrating sparse spike representations into SNN, reducing input redundancy and improving biological plausibility and efficiency. By optimizing both inputs and activations, we push the limits of energy efficiency in neuromorphic systems.

Minimalist vision [30], [31] pursues extreme data reduction, often sacrificing robustness to simpler and controlled tasks like workspace monitoring or occupancy detection. Zhao et al. [32] used learnable masks to capture essential information, focusing on traditional neural networks and overlooking the potential of SNN. We address this gap by applying laconic principles to neuromorphic systems, leveraging the event-driven, sparse activations of SNN.

## III. Method

This section describes laconic neuromorphic vision in a mathematical manner, which consists of: 1) sparse encoding with learnable masks, 2) spikes generation from encoded pixels, 3) inference using an SNN, and 4) training with loss.

Traditional vision systems often capture redundant data. To address this, we use learnable masks to sparsely encode essential information. Each mask $M(x, y)$, where $0 \leq M(x, y) \leq 1$, acts as a filter that compresses the scene into encoded pixels:

$$p = \int \int_{x,y} I(x, y) M(x, y)\, dx\, dy, \tag{1}$$

where $I(x, y)$ is the scene intensity. The masks learn to capture relevant features, reducing the dimensionality of the input.

In practice, event-based sensors can directly provide spike outputs. For simulation purpose, the probability of generating a spike for each pixel $p_i$ is given by:

$$P(\text{spike at time } t) = 1 - e^{-\lambda p_i}, \tag{2}$$

where $\lambda$ controls the spike rate.

The spikes pass through the SNN, where spiking neurons process the input over time. The class prediction $\hat{y}$ is based on the neuron with the highest spike count over time $T$:

$$\hat{y} = \arg\max_c \left( \frac{1}{T} \sum_{t=1}^{T} s_c(t) \right), \tag{3}$$

where $s_c(t)$ is the spike output of neuron $c$ at time $t$. The SNN uses Leaky Integrate-and-Fire (LIF) neurons, where the membrane potential $V(t)$ follows:

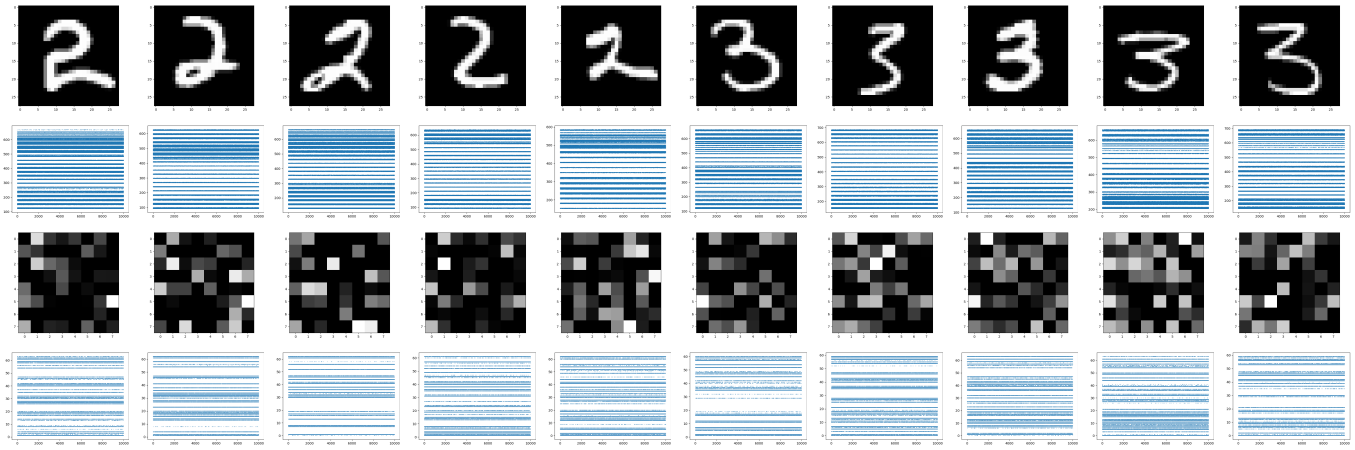


Fig. 3: Visualization of the original $28 \times 28$ MNIST images (1st row), the Poisson spike trains generated from these images (2nd row), the $8 \times 8$ encoded pixels (3rd row), and the Poisson spike trains corresponding to the encoded pixels (4th row).

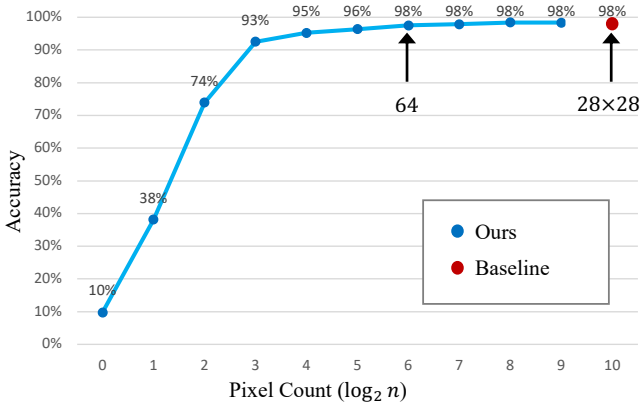



Fig. 4: Performance vs. pixel count (log scale) on MNIST.

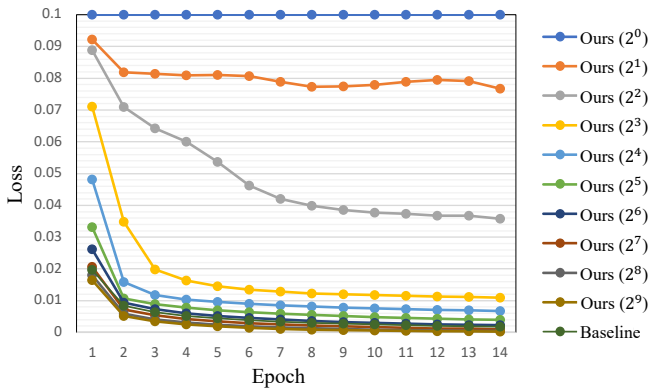



Fig. 5: Training loss over the number of epochs on MNIST.

$$\tau \frac{dV(t)}{dt} = -V(t) + RI(t), \tag{4}$$

where $\tau$ is the membrane time constant, $R$ is the membrane resistance, and $I(t)$ is the input current generated by incoming spikes. When $V(t)$ exceeds a threshold $V_{thr}$, a spike is generated, and $V(t)$ resets to its resting potential.

We train the SNN using Mean Squared Error (MSE) loss, aiming for the correct class neuron to spike most frequently:

$$L = \frac{1}{C} \sum_{c=1}^{C} \left( y_c - \frac{1}{T} \sum_{t=1}^{T} s_c(t) \right)^2, \tag{5}$$

where $y_c$ is the target spike rate, and $s_c(t)$ is the output spike rate of class $c$ at time $t$.

## IV. Experiment

We evaluate the performance of our laconic neuromorphic vision system using the MNIST [33] and Fashion-MNIST [34] datasets. Each image is compressed using learnable masks, converted into Poisson spikes, and processed by an SNN. The datasets have a fixed resolution of 28x28 pixels, with the full-resolution images serving as the baseline. Figure 3 and 7 provide visualization for intuitive understanding.

The results for MNIST are presented in Figures 4, 5, and 6.

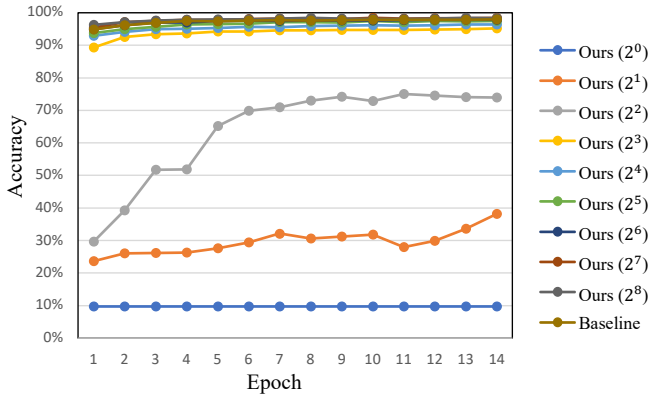



Fig. 6: Testing accuracy over the number of epochs on MNIST.

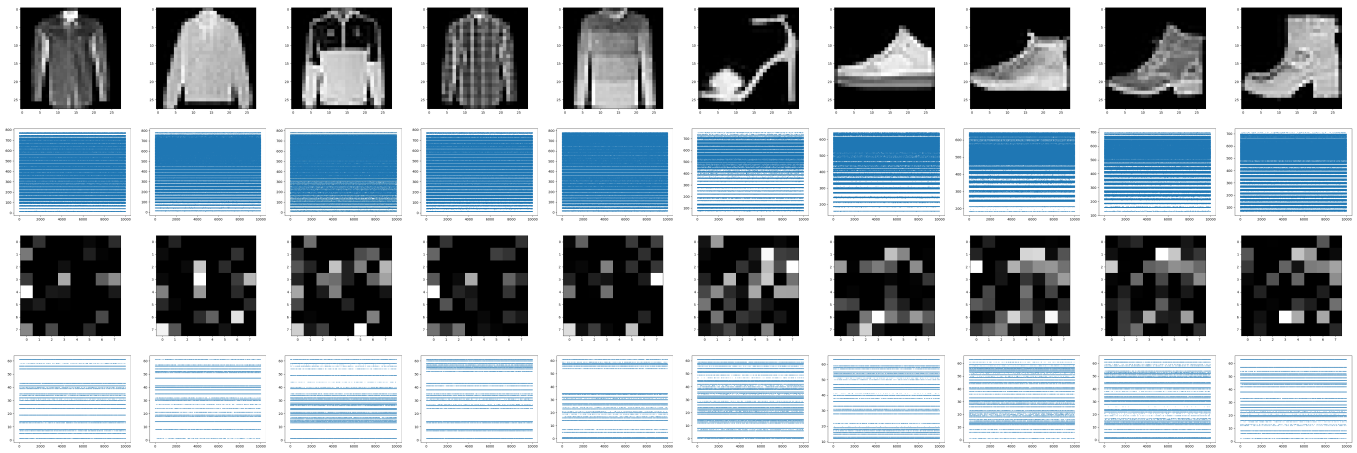


Fig. 7: Visualization of the $28 \times 28$ Fashion-MNIST images (1st row), the Poisson spike trains generated from these images (2nd row), the $8 \times 8$ encoded pixels (3rd row), and the Poisson spike trains corresponding to the encoded pixels (4th row).

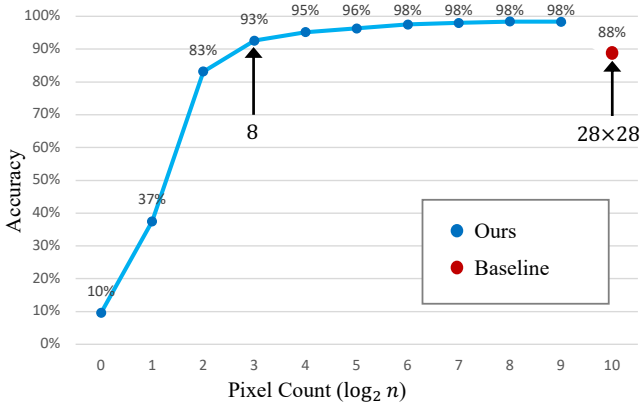



Fig. 8: Performance vs. pixel count on Fashion-MNIST.

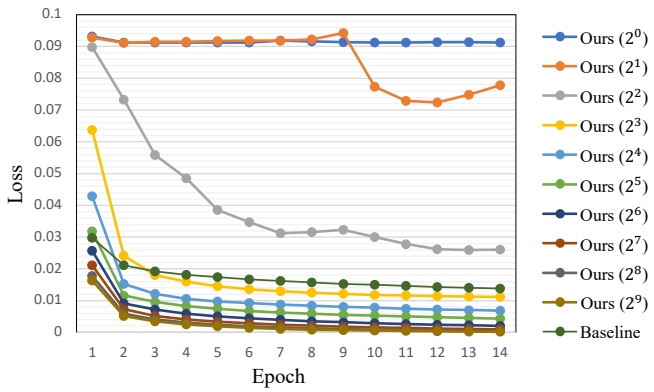



Fig. 9: Training loss over the epochs on Fashion-MNIST.

In Figure 4, with only 64 encoded pixels, the SNN achieves 98% accuracy, comparable to using all $28 \times 28$ pixels. Even with just 8 encoded pixels, 93% accuracy is achieved, demonstrating the efficiency of the laconic neuromorphic approach. As Figures 5 and 6 show, once the number of encoded pixels exceeds 16, the training loss and testing accuracy converge similarly, highlighting the effectiveness of sparse inputs.

Regrading results on Fashion-MNIST, Figure 8 shows that with only 8 encoded pixels, the SNN reaches 93% accuracy, comparable to using all $28 \times 28$ pixels, despite the complexity of the clothing structures. Figures 9 and 10 depict the training loss and testing accuracy over epochs, with both converging after a small number of epochs when more than 8 encoded pixels are used. This confirms that even with minimal spike inputs, the system effectively learns and performs the task.

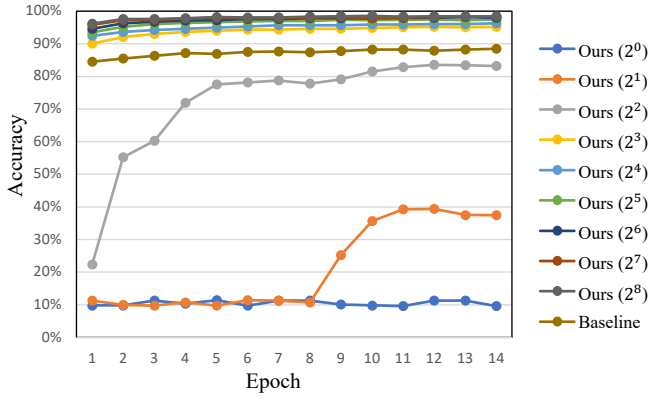



Fig. 10: Testing accuracy over the epochs on Fashion-MNIST.

## V. Conclusion

In this work, we introduce laconic neuromorphic vision. Our approach leverages the inherent sparsity and energy efficiency of SNN while further reducing input complexity through sparse encoding. Experiments demonstrate that our system achieves high classification accuracy with fewer pixels and spike inputs. These results highlight the potential of laconic neuromorphic vision for efficient, low-power, and biologically inspired vision systems.

## REFERENCES


[1] G. Indiveri and R. Douglas, "Neuromorphic vision sensors," *Science*, vol. 288, no. 5469, pp. 1189–1190, 2000.

[2] Q.-B. Zhu, B. Li, D.-D. Yang, C. Liu, S. Feng, M.-L. Chen, Y. Sun, Y.-N. Tian, X. Su, X.-M. Wang *et al.*, "A flexible ultrasensitive optoelectronic sensor array for neuromorphic vision systems," *Nature communications*, vol. 12, no. 1, p. 1798, 2021.

[3] S. Zhu, C. Wang, H. Liu, P. Zhang, and E. Y. Lam, "Computational neuromorphic imaging: principles and applications," in *Computational Optical Imaging and Artificial Intelligence in Biomedical Sciences*, vol. 12857. SPIE, 2024, pp. 4–10.

[4] P. Zhang, S. Zhu, C. Wang, Y. Zhao, and E. Y. Lam, "Neuromorphic imaging with super-resolution," *IEEE Transactions on Circuits and Systems for Video Technology*, 2024.

[5] S. Ghosh-Dastidar and H. Adeli, "Spiking neural networks," *International journal of neural systems*, vol. 19, no. 04, pp. 295–308, 2009.

[6] A. Tavanaei, M. Ghodrati, S. R. Kheradpisheh, T. Masquelier, and A. Maida, "Deep learning in spiking neural networks," *Neural networks*, vol. 111, pp. 47–63, 2019.

[7] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, "Edge computing: Vision and challenges," *IEEE internet of things journal*, vol. 3, no. 5, pp. 637–646, 2016.

[8] N. Abbas, Y. Zhang, A. Taherkordi, and T. Skeie, "Mobile edge computing: A survey," *IEEE Internet of Things Journal*, vol. 5, no. 1, pp. 450–465, 2017.

[9] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, "A survey on mobile edge computing: The communication perspective," *IEEE communications surveys & tutorials*, vol. 19, no. 4, pp. 2322–2358, 2017.

[10] G. Chen, H. Cao, J. Conradt, H. Tang, F. Rohrbein, and A. Knoll, "Event-based neuromorphic vision for autonomous driving: A paradigm shift for bio-inspired visual sensing and perception," *IEEE Signal Processing Magazine*, vol. 37, no. 4, pp. 34–49, 2020.

[11] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE signal processing magazine*, vol. 25, no. 2, pp. 83–91, 2008.

[12] M. S. Asif, A. Ayremlou, A. Sankaranarayanan, A. Veeraraghavan, and R. G. Baraniuk, "FlatCam: Thin, lensless cameras using coded aperture and computation," *IEEE Transactions on Computational Imaging*, vol. 3, no. 3, pp. 384–397, 2016.

[13] N. Antipa, G. Kuo, R. Heckel, B. Mildenhall, E. Bostan, R. Ng, and L. Waller, "DiffuserCam: lensless single-exposure 3d imaging," *Optica*, vol. 5, no. 1, pp. 1–9, 2017.

[14] V. Boominathan, J. K. Adams, J. T. Robinson, and A. Veeraraghavan, "PhlatCam: Designed phase-mask based thin lensless camera," *IEEE transactions on pattern analysis and machine intelligence*, vol. 42, no. 7, pp. 1618–1629, 2020.

[15] G. M. Gibson, S. D. Johnson, and M. J. Padgett, "Single-pixel imaging 12 years on: a review," *Optics Express*, vol. 28, no. 19, pp. 28 190–28 208, 2020.

[16] M. P. Edgar, G. M. Gibson, and M. J. Padgett, "Principles and prospects for single-pixel imaging," *Nature photonics*, vol. 13, no. 1, pp. 13–20, 2019.

[17] X. Yuan, D. J. Brady, and A. K. Katsaggelos, "Snapshot compressive imaging: Theory, algorithms, and applications," *IEEE Signal Processing Magazine*, vol. 38, no. 2, pp. 65–88, 2021.

[18] Y. Zhao, S. Zheng, and X. Yuan, "Deep equilibrium models for video snapshot compressive imaging," *arXiv preprint arXiv:2201.06931*, 2022.

[19] Q. Yang and Y. Zhao, "Revisit dictionary learning for video compressive sensing under the plug-and-play framework," in *SPIE Proceedings*, vol. 12166, 2022, pp. 2018–2025.

[20] Y. Zhao, S. Zheng, and X. Yuan, "Deep equilibrium models for snapshot compressive imaging," in *AAAI Conference on Artificial Intelligence*, vol. 37, no. 3, 2023, pp. 3642–3650.

[21] Y. Zhao and E. Y. Lam, "SASA: saliency-aware self-adaptive snapshot compressive imaging," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2024, pp. 2370–2374.

[22] Y. Zhao, "Mathematical cookbook for snapshot compressive imaging," *arXiv preprint arXiv:2202.07437*, 2022.

[23] Y. Zhao and E. Y. Lam, "Optimizing object detection in electro-optical systems with snapshot compressive imaging," in *SPIE Proceedings*, 2024.

[24] D. L. Donoho, "Compressed sensing," *IEEE Transactions on information theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[25] Z. Y. Chong, Y. Zhao, Z. Wang, and E. Y. Lam, "Solving inverse problems in compressive imaging with score-based generative models," in *IEEE International Conference on Data Science and Advanced Analytics*, 2023, pp. 1–10.

[26] Y. Zhao, Q. Zeng, and E. Y. Lam, "Adaptive compressed sensing for real-time video compression, transmission, and reconstruction," in *IEEE International Conference on Data Science and Advanced Analytics*, 2023, pp. 1–10.

[27] G. Gallego, T. Delbrück, G. Orchard, C. Bartolozzi, B. Taba, A. Censi, S. Leutenegger, A. J. Davison, J. Conradt, K. Daniilidis *et al.*, "Event-based vision: A survey," *IEEE transactions on pattern analysis and machine intelligence*, vol. 44, no. 1, pp. 154–180, 2020.

[28] Y. Zhao, R. Chen, C. Wang, and E. Y. Lam, "esports broadcasts with event cameras," in *International Conference on Neural Information Processing*. Springer, 2024.

[29] Y. Zhao, P. Zhang, C. Wang, and E. Y. Lam, "Controllable unsupervised event-based video generation," in *IEEE International Conference on Image Processing*, 2024, pp. 2278–2284.

[30] J. Klotz and S. K. Nayar, "Minimalist vision with freeform pixels," in *European Conference on Computer Vision*, 2024.

[31] P. Pooj, M. Grossberg, P. N. Belhumeur, and S. K. Nayar, "The minimalist camera," in *The British Machine Vision Conference*, 2018, p. 141.

[32] Y. Zhao and E. Y. Lam, "Laconic vision for image classification," *under review*, 2024.

[33] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[34] H. Xiao, K. Rasul, and R. Vollgraf, "Fashion-MNIST: a novel image dataset for benchmarking machine learning algorithms," *arXiv preprint arXiv:1708.07747*, 2017.