

Supplementary Information for: Self-optimizing multichannel optical computing

Fatma Nur Kılınç¹ and Uğur Teğın^{1,2,*}

¹Department of Computational Sciences and Engineering, Koç University, İstanbul, Türkiye

²Department of Electrical and Electronics Engineering, Koç University, İstanbul, Türkiye

*Corresponding author: utegin@ku.edu.tr

1 Benchmarking against digital methods

HAM10000 dermatoscopic lesion classification

To contextualize our results on the HAM10000, we compare our multichannel optical computing performance with recent state-of-the-art digital deep learning methods (Supplementary Table 1). Our multichannel MPLC with RGB encoding, combined with a simple Ridge classifier achieves 98% accuracy without any data augmentation, performing competitively against the strongest recent digital approaches that rely on extensive augmentation strategies.

Notably, the deep ensemble model by Thwin et al. reports performance both with and without augmentation (96.00% vs 90.00%), clearly demonstrating the substantial impact of data augmentation on digital baselines. That our optical preprocessing approach attains 98% accuracy without augmentation suggests that the nonlinear optical transformations implemented by the cascaded phase masks yield inherently discriminative feature representations. Our grayscale MPLC system achieves 92% accuracy, which also exceeds the non-augmented digital baseline (90.00%), further underscoring the value of optical processing.

STL-10 natural image classification

Supplementary Table 2 compares our STL-10 results with recent digital methods. STL-10 is challenging for optical computing because it contains only 13,000 labeled images across 10 classes, while most state-of-the-art approaches additionally exploit 100,000 unlabeled images in a semi-supervised

Supplementary Table 1: HAM10000 Dataset Benchmark.

Method	Year	Acc. (%)	Augment.
SkinLesion-STN [1]	2024	98.00	Yes
Enhanced CNN [2]	2024	97.78	Yes
Deep ensemble [3]	2024	96.00	Yes
Fuzzy ensemble [4]	2025	95.14	Yes
Deep ensemble [3]	2024	90.00	No
MC-MPLC (Gray) + Ridge	2025	92%	No
MC-MPLC (RGB) + Ridge	2025	98%	No

settings. Most state-of-the-art methods leverage these unlabeled samples, whereas our approach uses only the labeled images in a fully supervised regime. Despite this constraint, our optimized multichannel MPLC system with Bayesian-optimized channel mixing achieves 86% accuracy, placing it competitively alongside recent semi-supervised approaches.

Supplementary Table 2: STL-10 Dataset Benchmark.

Method	Year	Acc. (%)	Type
SS-RML-CNN [5]	2020	94.10	Semi-sup.
IIC [6]	2019	88.80	Semi-sup.
CoMatch [7]	2021	79.80	Semi-sup.
MC-MPLC (Gray) + Ridge	2025	68%	Sup.
MC-MPLC (RGB) + Ridge	2025	79%	Sup.
MC-MPLC (RGB) + C.M.+Ridge	2025	86%	Sup.

Flowers-17 image classification

We further evaluate our approach on the Flowers-17 dataset to test generalization to fine-grained natural image recognition (Supplementary Table 3). Flowers-17 contains 17 classes with relatively few images per category, and most recent digital methods rely on strong CNN back-ends and heavy data augmentation in this setting. In contrast, our multichannel MPLC operates on only half of the available labeled images (the remaining half is used for system optimization) and does not employ any data augmentation for training or testing.

Supplementary Table 3: Flowers-17 Dataset Benchmark.

Method	Year	Acc. (%)	Augment.
VA-ResNet50 [8]	2021	85.7	Yes
ResNet L + R [9]	2023	98.03	Yes
DCNN features + ML classifiers [10]	2019	99.8	Yes
MC-MPLC (RGB) + Ridge	2025	64.00	No
MC-MPLC (RGB) + Mixing + Ridge	2025	76.00	No

Abalone age regression

For the Abalone age regression task, we defined a standardized benchmark protocol to compare optical and digital approaches. Using the same normalized features and train–test split, we evaluated common digital regressors (linear, ridge, lasso, random forest, gradient boosting, and a one-hidden-layer MLP), which achieved RMSEs of approximately 2.26–2.29 years. Under this protocol, our multichannel MC-MPLC + Ridge model, trained on optically encoded features, attains an RMSE of 1.99 years, outperforming all digital baselines.

Supplementary Table 4: Abalone Age Regression Benchmark.

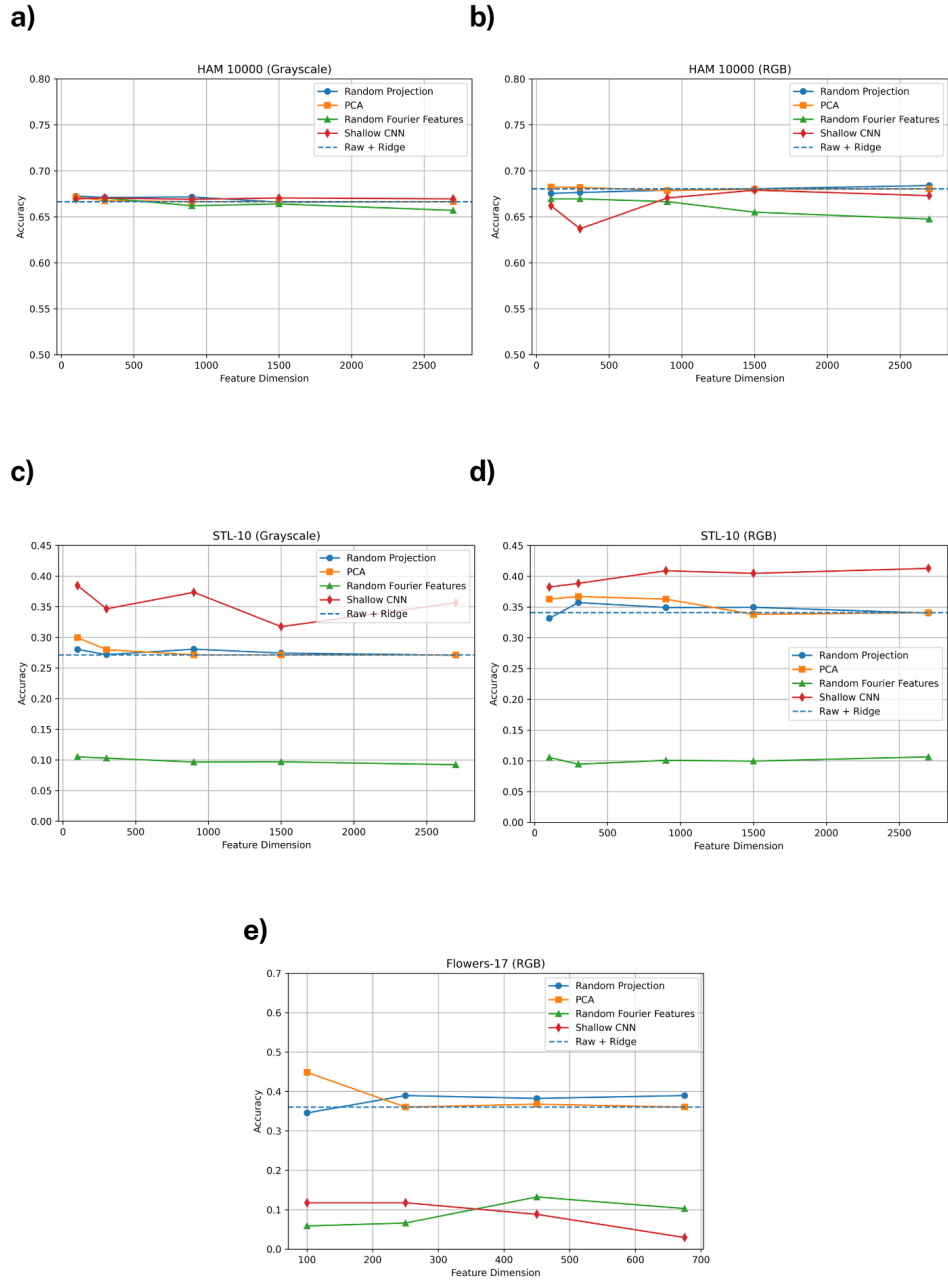
Method	RMSE (years)	Norm. RMSE	R^2
Linear regression	2.27	0.081	0.52
Ridge regression	2.27	0.081	0.52
Lasso regression	2.27	0.081	0.52
Random forest regressor	2.29	0.082	0.52
Gradient boosting	2.26	0.081	0.53
MLP regressor	2.26	0.081	0.53
MC-MPLC (RGB) + Ridge	1.99	0.071	0.64

2 Additional Comparison with Feature-matched Digital Baselines

To assess whether the observed gains can be explained by generic digital feature expansion alone, we compared the proposed MPLC-based features against several feature-matched digital baselines, including random projection, PCA, random Fourier features, and a shallow CNN.

As shown in Supplementary Fig. 1, these digital baselines provide only limited improvements over the raw-pixel Ridge baseline and remain below the MPLC-based optical results reported in the main text. This suggests that the observed gains are not explained solely by increased feature dimensionality, but also reflect the structured transformation performed by the optical system.

Although phase modulation and free-space propagation are linear in the optical field, the overall pipeline is not purely linear in the input data because the input is phase-encoded and the detected output is optical intensity. This introduces an effective nonlinearity prior to the final linear readout.



Supplementary Fig. 1: Comparison with feature-matched digital baselines across datasets. Random projection, Principal component analysis (PCA), random Fourier features, and a shallow CNN are compared against the raw-pixel Ridge baseline at different feature dimensions. (a) HAM10000 dataset in grayscale format. (b) HAM10000 dataset in RGB format. (c) STL-10 dataset in grayscale format. (d) STL-10 dataset in RGB format. (e) Flowers-17 dataset in RGB format.

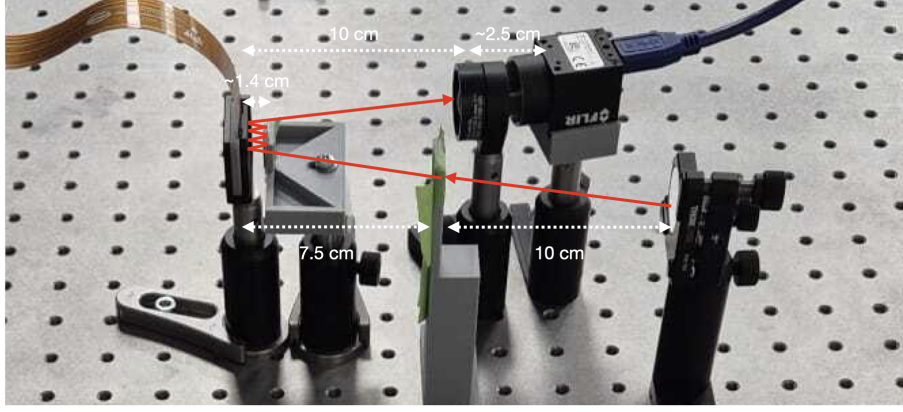
3 Sensitivity to Random Phase-Mask Initialization

To assess the sensitivity of the proposed approach to random phase-mask initialization, we performed a simulation study on the Flowers-17 dataset in which optical propagation was modelled using the Angular Spectrum Method (ASM). Fifty independent random seeds were used to generate distinct phase-mask realisations while all other system parameters were held fixed. For each realisation, the full optical inference pipeline was evaluated and the classification accuracy was recorded. Across all 50 runs, the mean accuracy was 41.51% with a standard deviation of 2.60%, indicating that performance is only weakly sensitive to the choice of initialisation seed and that the reported results are representative of the typical behaviour of the system rather than the outcome of a particularly favourable phase-mask configuration.

It should be noted that this analysis was performed in simulation rather than experiment. As discussed in the main text, the simulation model does not fully capture all experimental imperfections and therefore should not be interpreted as an exact predictor of experimental performance. Nevertheless, it provides a useful estimate of the relative sensitivity of the proposed method to initialization randomness.

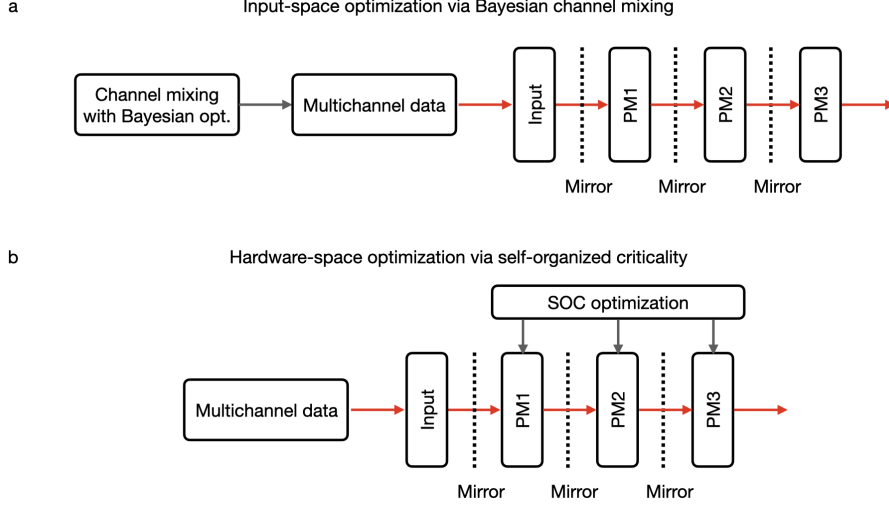
4 Experimental setup and data flow

To make the implementation easier to follow, we provide here both an image of the physical setup and a schematic of the corresponding data flow. Supplementary Figure 2 shows the main optical components in the experiment: light from the source first passes through a polarizer, is guided by mirrors and a slit into the MPLC system, and the output is then collected by a 30 mm lens and recorded by the camera. This figure is intended to clarify the physical layout of the system and the optical propagation path.



Supplementary Fig. 2: Physical experimental setup. Light from the source passes through a polarizer, is directed by mirrors and a slit into the Multiplane Light Conversion (MPLC) system, and the output is collected by a 30 mm lens and recorded by the camera.

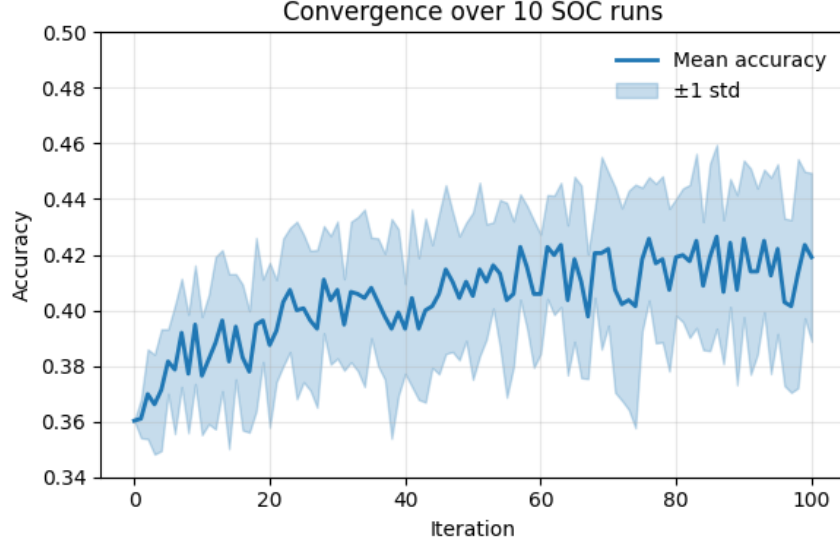
Figure 3 shows the logical data flow for the two optimization strategies. In the input-space optimization scheme, channel mixing of the multichannel input is optimized directly in a hardware-in-the-loop manner and then encoded at the input plane, while the randomly initialized phase masks remain fixed. In the hardware-space optimization scheme, the input is kept fixed and the phase masks are optimized directly. Together, these two figures clarify both the physical implementation of the system and the stage at which optimization is performed in each case.



Supplementary Fig. 3: Data-flow schematic of input-space and hardware-space optimization in the Multi-plane Light Conversion (MPLC) system. (a) Multichannel data are optimized in the input space, while the randomly initialized phase masks remain fixed. (b) The input is fixed and the phase masks are optimized directly. Red arrows denote optical propagation and gray arrows denote optimization.

5 Robustness across independent SOC runs

To quantify robustness with respect to the stochastic SOC updates, we performed 10 independent simulation optimisation runs using the same initial random phase masks. For each run, the input image and optical settings were fixed, while the position of the added sand grain on the Abelian sand-pile grid was re-sampled with a different random seed at every iteration. Supplementary Figure 4 shows the resulting mean accuracy and the corresponding ± 1 standard deviation envelope as a function of iteration. The mean accuracy increases from approximately 0.36 at iteration 0 to approximately 0.42 by iteration 100, and all runs converge to this same accuracy range. To summarise this behaviour quantitatively, we report the mean and standard deviation of the initial and final accuracies across runs in Table 5.



Supplementary Fig. 4: Mean test accuracy (solid line) and ± 1 standard deviation (shaded region) over 10 independent SOC optimisation runs as a function of iteration. Each run uses the same initial random phase masks but a randomly re-sampled sand-grain perturbation location at each iteration. All runs converge to an accuracy plateau of approximately 0.42 demonstrating robustness of the SOC procedure to stochastic perturbations.

Supplementary Table 5: Robustness of SOC optimisation across 10 independent runs. Values shown are mean $\pm$ standard deviation of test accuracy at the initial (iteration 0) and final (iteration 100) states.

	Initial accuracy	Final accuracy
SOC (10 runs)	0.3603	0.4191 ± 0.0303

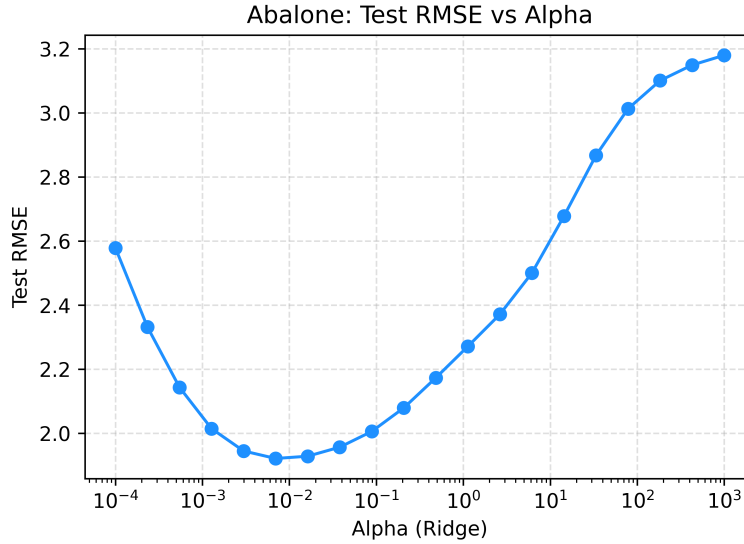
6 Ridge regression hyperparameter selection

Ridge regression introduces an L2 regularization term to standard linear regression to prevent overfitting, particularly when dealing with multicollinearity or high-dimensional feature spaces. The regularization penalizes large coefficient values, controlled by a hyperparameter $\alpha \geq 0$. The objective function minimized in Ridge regression is:

$$\min_{\mathbf{w}} \{ \|\mathbf{y} - \mathbf{X}\mathbf{w}\|_2^2 + \alpha \|\mathbf{w}\|_2^2 \} \quad (1)$$

where $\mathbf{X}$ is the feature matrix extracted from optical intensity measurements, $\mathbf{y}$ is the target vector (class labels for classification or continuous values for regression), $\mathbf{w}$ is the coefficient vector, and α controls the regularization strength. When $\alpha = 0$, Ridge regression reduces to ordinary least squares; as α increases, coefficients are increasingly shrunk toward zero.

For the Abalone regression task, we conducted a more detailed hyperparameter search to illustrate the sensitivity of performance to α . Supplementary Figure 5 shows the root mean square error (RMSE) on the test set as a function of α . The plot reveals a characteristic U-shaped curve.



Supplementary Fig. 5: Root mean square error (RMSE) versus Ridge regression regularization parameter α for the Abalone regression task.

References

- [1] Suma, K. G. *et al.* Skinlesion-stn: Enhancing skin lesion classification using spatial transformer networks. In *2024 2nd International Conference on Artificial Intelligence Trends and Pattern Recognition (ICAITPR)*, 1–6 (2024).

- [2] Musthafa, M. M., Mahesh, T. R., Vinoth Kumar, V. & Guluwadi, S. Enhanced skin cancer diagnosis using optimized cnn architecture and checkpoints for automated dermatological lesion classification. *BMC Medical Imaging* **24**, 201 (2024).
- [3] Thwin, S. M. & Park, H.-S. Skin lesion classification using a deep ensemble model. *Applied Sciences* **14**, 5599 (2024).
- [4] Halder, A. *et al.* A fuzzy rank-based deep ensemble methodology for multi-class skin cancer classification. *Scientific Reports* **15**, 6268 (2025).
- [5] Cevikalp, H., Benligiray, B. & Gerek, O. N. Semi-supervised robust deep neural networks for multi-label image classification. *Pattern Recognition* **100**, 107164 (2020).
- [6] Ji, X., Henriques, J. F. & Vedaldi, A. Invariant information clustering for unsupervised image classification and segmentation. In *Proceedings of the IEEE International Conference on Computer Vision*, 9865–9874 (2019).
- [7] Li, J., Xiong, C. & Hoi, S. C. H. Comatch: Semi-supervised learning with contrastive graph regularization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 9475–9484 (2021).
- [8] Cao, S. & Song, B. Visual attentional-driven deep learning method for flower recognition. *Mathematical Biosciences and Engineering* **18**, 1981–1991 (2021).
- [9] Pereira-Ferrero, V. H., Valem, L. P. & Pedronette, D. C. G. Feature augmentation based on manifold ranking and lstm for image classification. *Expert Systems with Applications* **213**, 118995 (2023).
- [10] Mete, B. R. & Ensari, T. Flower classification with deep cnn and machine learning algorithms. In *2019 3rd International Symposium on Multidisciplinary Studies and Innovative Technologies*, 1–5 (2019).