Supplementary Materials

Vsevolod Kleshchenko[1], Vladimir Igoshin[1], Costantino De Angelis[2], and Mihail Petrov [1]

[1]*Department of Physics and Engineering, ITMO University, 191002, Lomonosova St. 9, St. Petersburg, Russia and*

[2]*Department of Information Engineering, University of Brescia, Via Branze 38, 25123 Brescia, Italy*

A. Total field formulation

It should be mentioned, that all the results presented so far were obtained using scattered-field formulation. It means that only the scattered field is collected by the detectors, and the incident field was not included during training. In practice, however, the total field is usually measured, which includes both the incident and scattered contributions. For total field training, the detector response is given by integrating the total intensity over a set of detector regions Ω_k:

$$I_k = \int_{\Omega_k} \left| \mathbf{E}^{\infty}_{\mathrm{tot}}(r_{\mathrm{det}}; \theta, \phi) \right|^2 d\Omega, \tag{1}$$

Despite the more complex interference effects between incident and scattered fields, the total-field formulation achieves comparable performance to the scattered-field case since the T-matrix layer can learn to exploit interference for classification, as shown in Figure 1.

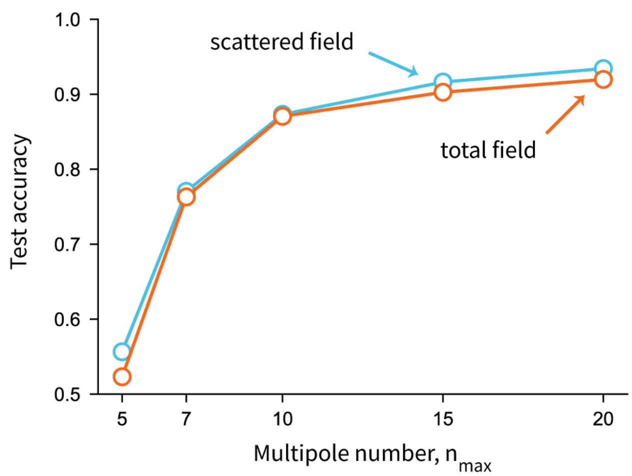

FIG. 1. MNIST classification accuracy as a function of the number of multipole modes $n_{\max}$ for the scattered and total-field formulation with reciprocity, passivity, and axial symmetry constraints (NA = 0.9).

B. Performance for all constraint combinations

The comparison of all constraint combinations in terms of multipole numbers and trainable parameters are presented at Figure 2.

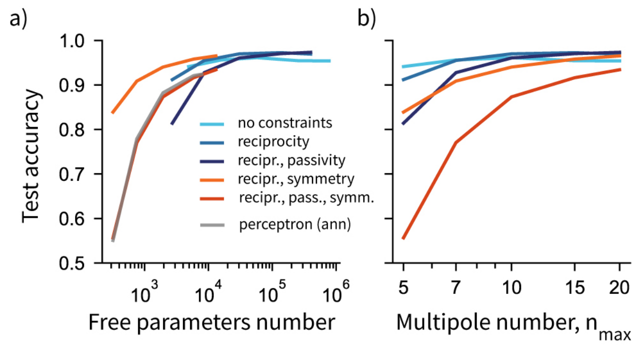

FIG. 2. Test accuracy of the T-matrix-based classifier for different combinations of physical constraints as a function of the free parameters number (a) and the number of multipole modes (b).

C. Dataset representation

In this work, the input for the T-matrix-based network is the set of multipole coefficients obtained by projecting the input image onto the spherical harmonic basis. This representation is compared with the standard pixel basis in Figure 3a in terms of the Fischer Discriminant ratio (FDR) trace and the Silhouette score, which are commonly used metrics for evaluating the separability of classes in a given feature space. The FDR trace shows the ratio of between-class variance to within-class variance, with higher values indicating better class separability:

$$\mathrm{FDR} = \frac{\sum_{k=1}^{K} N_k \|\mu_k - \mu\|^2}{\sum_{k=1}^{K} \sum_{i=1}^{N_k} \|x_i^{(k)} - \mu_k\|^2},$$

where K is the number of classes, N_k is the number of samples in class k, μ_k is the mean vector of class k, μ is the overall mean vector, and $x_i^{(k)}$ are the individual samples in class k. The Silhouette score measures how similar an object is to its own cluster compared to other clusters, with values ranging from -1 to 1, where higher values indicate better clustering:

$$\mathrm{Silhouette} = \frac{1}{N} \sum_{i=1}^{N} \frac{b(i) - a(i)}{\max\{a(i), b(i)\}},$$

where N is the total number of samples, $a(i)$ is the average distance from the i-th sample to all other samples in the same class, and $b(i)$ is the average distance from the i-th sample to all samples in the nearest different class. The results show that the multipole representation of the input data has a higher FDR and Silhouette score compared to the pixel basis, indicating that multipole expansion provides a more discriminative feature space for classification. The

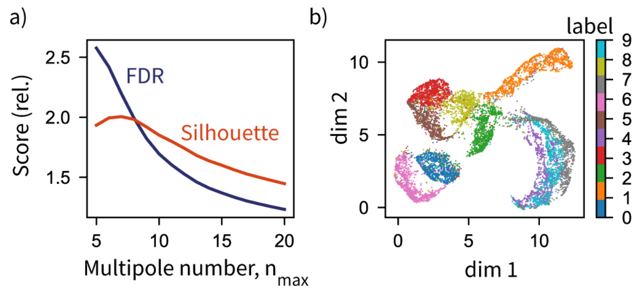

FIG. 3. (a) Fischer Discriminant ratio (FDR) and Silhouette score of the multipole representation of the MNIST dataset. The values are normalized by the corresponding values obtained for the original MNIST dataset. (b) UMAP projection of the multipole representation of MNIST test dataset for $n_{\max} = 20$, $NA = 0.9$, colored by class label.

projection via UMAP (Uniform Manifold Approximation and Projection) demonstrates the multipole representation of MNIST digits, as shown in Figure 3b. The value of the FDR as well as the Silhouette score decrease with increasing $n_{\max}$. It can be attributed to the fact that higher-order multipole modes capture important details of the input image, however, they also introduce more noise in the feature space because of the larger number of modes with higher m values.

D. Relation to artificial neural networks

The proposed scattering-based model can be directly interpreted in terms of a standard artificial neural network (ANN) layer of the form

$$\mathbf{y} = \mathbf{W}\mathbf{x} + \mathbf{b}. \tag{2}$$

In our case, the input vector $\mathbf{x}$ corresponds to the set of incident multipole coefficients $(\mathbf{a}, \mathbf{b})$, while the output vector $\mathbf{y}$ is given by the scattered coefficients $(\mathbf{p}, \mathbf{q})$. The role of the weight matrix $\mathbf{W}$ is played by the T-matrix,

$$\begin{pmatrix} \mathbf{p} \\ \mathbf{q} \end{pmatrix} = \mathbf{T} \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix}. \tag{3}$$

Thus, the electromagnetic scatterer implements a linear transformation in a modal feature space, formally equivalent to a fully connected layer.

A key distinction from conventional neural networks is the absence of an explicit bias term. In the present formulation, the mapping is strictly linear and homogeneous, reflecting the linearity of Maxwell's equations for passive media. Any effective bias can only arise indirectly, for example through interference with a fixed reference field or through detector calibration, but is not an intrinsic degree of freedom of the scatterer.

Another important difference concerns the nature of the nonlinearity. In standard ANNs, nonlinear activation functions are applied element-wise to $\mathbf{y}$. In contrast, the optical system remains linear up to the measurement stage, and nonlinearity is introduced through intensity detection,

$$I_k \sim |\mathbf{E}_{\mathrm{sc}}|^2 , \tag{4}$$

which depends quadratically on the modal coefficients $(\mathbf{p}, \mathbf{q})$. The overall mapping therefore takes the form

$$(\mathbf{a}, \mathbf{b}) \xrightarrow{\mathbf{T}} (\mathbf{p}, \mathbf{q}) \xrightarrow{|\cdot|^2} \{I_k\}, \tag{5}$$

which can be viewed as a linear layer followed by a fixed quadratic activation and spatial pooling.

Finally, the parameter space of $\mathbf{T}$ is not arbitrary but constrained by physical laws, including reciprocity, passivity, and symmetry. As a result, the effective weight matrix belongs to a restricted manifold of physically realizable operators. This distinguishes the present system from conventional ANNs and implies that learning is performed under strong inductive bias imposed by electromagnetic theory.

E. Passivity effect

We consider a simplified single-layer intensity model with temperature τ

$$z_k = |w_k^\dagger x|^2, \qquad p_k = \frac{\exp(z_k/\tau)}{\sum_j \exp(z_j/\tau)}$$

where x denotes the incident coefficients and w_k is the effective total linear map to the k-th detector. The full detector-integrated formulation has the same quadratic scaling with the scattered-field amplitude. For the correct class y, *margin* and *loss* are:

$$m_{yj} = z_y - z_j, \qquad \mathcal{L} = \log\left(1 + \sum_{j \neq y} e^{-m_{yj}/\tau}\right).$$

If the effective optical gain is bounded by $\|W\|_2 \le s$, then $z_k \le s^2\|x\|_2^2$, and $m_{yj} \le s^2\|x\|_2^2$. Thus, when

$$s^2\|x\|_2^2/\tau \ll 1$$

the loss remains close to $\log K$, and margin scale is limited. The *gradients* $\nabla_{w_k}\mathcal{L} \propto \frac{1}{\tau}\left(p_k - \mathbf{1}[k=y]\right)(w_k^\dagger x)x$ for the spectrally bounded layer can therefore produce very small parameter updates since

$$\|\nabla_{w_k}\mathcal{L}\| \lesssim \frac{s}{\tau}\|x\|_2^2$$

However, *temperature* allow to rescale the margins seen by the loss

$$m_{\text{eff}} = \frac{m}{\tau},$$

Thus, decreasing τ can compensate for the small logit and gradient scale produced by passivity.

Robustness is discussed here in terms of perturbation $\widetilde{w}_k = w_k + \delta w_k$, where $\|\delta w_k\|_2 \le \epsilon$. The corresponding change of an intensity score and margins:

$$|\delta z_k| \le (2s\epsilon + \epsilon^2)\|x\|_2^2, \qquad |\delta m_{yj}| \le 2(2s\epsilon + \epsilon^2)\|x\|_2^2$$

Thus, a smaller spectral norm reduces the sensitivity of the logits and margins to perturbations.