1 Optical device and simulation

Simulation setup: optical device and electrical fields
- Simulation software: Rigorous Coupled-Vector Analysis (RCWA).
- Optical device: Silicon (Si) nanostructures represented by a binary vector.

Input signal: Normal incident light.
Output signal: Electric field distributed in a 2D space.

Data set for learning:
- 200,000 training samples, 50,000 test samples.
- Si structure: $y \in \{-1, +1\}^{20}$
- Output electrical field collected at near-field and far-field locations with $x \in \mathbb{C}^{61}$

![Figure 1 Schematic of 1D Si structure with simulation procedure.](image)

2 Neural networks

Neurons
- Activation function: Rectified Linear Unit (ReLU)
  \[ f(z) = \max(0, z) \]
- Fully connected neural networks
  - Parameters at layer $l$
    - Neuron weights: $W^{(l)} \in \mathbb{R}^{H_l \times H_{l-1}}$
    - Neuron bias: $b^{(l)} \in \mathbb{R}^{H_l}$
  - Forward-feeding
    \[ z^{(l)} = f_l \left(W^{(l)} z^{(l-1)} + b^{(l)} \right) \]
  - For all $j$, the activation function is ReLU in this project

Prediction
\[ \hat{y} = W^{(2)} z^{(1)} + b^{(2)} \]

Training process
- Cost function: Mean Square Error (MSE) function
  \[ J(W, b; y) = \frac{1}{m} \sum_{i=1}^{m} \| y^{(i)} - \hat{y}^{(i)} \|_2^2 \]

Minimization procedure
- Backward propagation to find the derivatives of the cost function with respect to parameters $W^{(l)}, b^{(l)}$ between each layer
- Stochastic gradient descent (SGD) to find the optimal solution
- Constant learning rate $\alpha = 0.001$

![Figure 3 Output electrical field given light normal on Si structures shown in Figure 1.](image)

3 Training results and discussion

Training configurations and definitions
- Neural networks configuration
  - The magnitude and phase of the transmitted electrical field are used as the input to the neural networks, i.e. $x \in \mathbb{C}^{23 \times 51 \times 102}$
  - The magnitude and phase are normalized before feeding into the neural network
  - Two hidden layers with $H_1 = H_2 = 51$

Error definitions
- Training and testing error:
  \[ e_{\text{train/test}} = \frac{\sum_{i=1}^{m} \| y^{(i)} - \hat{y}^{(i)} \|_2}{\sum_{i=1}^{m} \| W^{(1)} \|_2} \]
- Structure prediction error
  - Structure prediction
    \[ \hat{y} = \begin{cases} +1, & \hat{y} > 0 \\ -1, & \hat{y} < 0 \end{cases} \]
  - Error
    \[ e_{\text{prediction}} = \frac{1}{m} \sum_{i=1}^{m} \sum_{k=1}^{K} I(\hat{y}_k \neq y_k) \]

![Figure 4 Neural networks with 2 hidden layers.](image)

4 Application

Physical implication: Deviation efficiency
We achieved 43.4% efficiency as a 45° beam deflector. The Si nanostructures were predicted by our deep learning algorithm using ideal 100% deflected electric fields as input in the test.

![Figure 7 Comparison of 45° deflected transmitted light electric field of the predicted Si structure and the ideal 100% efficiency electric field](image)

1. kzhang3@stanford.edu
2. shiyuliu@stanford.edu
3. wangqing@stanford.edu