Supplemental Document

Optics EXPRESS

Multiscale diffractive U-Net: a robust all-optical deep learning framework modeled with sampling and skip connections: supplement

YIMING LI,[1,2,3] ZEXI ZHENG,[4] RAN LI,[5] QUAN CHEN,[2] HAITAO LUAN,[1,3] HUI YANG,[2,6] QIMING ZHANG,[1,3,*] AND MIN GU[1,3,7]

[1] *Institute of Photonic Chips, University of Shanghai for Science and Technology, Shanghai 200093, China*
[2] *Shanghai Key Lab of Modern Optical System, School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China*
[3] *Centre for Artificial-Intelligence Nanophotonics, School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China*
[4] *School of Mechanical Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China*
[5] *School of Health Science and Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China*
[6] *yanghui@usst.edu.cn*
[7] *gumin@usst.edu.cn*
[*] *qimingzhang@usst.edu.cn*

Supplement DOI: https://doi.org/10.6084/m9.figshare.21102574

Parent Article DOI: https://doi.org/10.1364/OE.468648

Multiscale Diffractive U-Net: a robust all-optical deep learning framework modeled with sampling and skip connections: supplemental document

1. Network modeling and parameter definition.

The product of the input wave and the *i*-th pixel transmission coefficient (b_i) determine the amplitude and relative phase of this secondary wave. Based on this, at the output layer, the output function $y_{out_i}(x,y,z)$ of the *i*-th pixel located at (x_i, y_i, z_i) position can be written as:

$$y_{out_i}(x,y,z) = w_i(x,y,z) \cdot t_i(x_i, y_i, z_i) \cdot y_{in_i}(x_i, y_i, z_i) \tag{S1}$$

where:

$$y_{in_i}(x_i, y_i, z_i) = \sum_k (w_k(x,y,z) \cdot y_{0_k}(x_i, y_i, z_i)) \tag{S2}$$

$$t_i(x_i, y_i, z_i) = a_i(x_i, y_i, z_i) \cdot \exp(j\varphi_i(x_i, y_i, z_i)) \tag{S3}$$

$y_{0_i}(x,y,z)$ the field distribution at the input field, $a_i(x_i, y_i, z_i)$ is the amplitude coefficient, λ is the wavelength, and φ_i is the phase value of each pixel.

In the training phase, a_i and φ_i are modelled as follows [37]:

$$a_i = \frac{ReLU(a_i)}{\max_{0<i\leq M\{ReLU(a_i)\}}} \tag{S4}$$

$$\varphi_i = 2\pi \times \beta_i \tag{S5}$$

where *ReLU* refers to Rectified Linear Unit, and *M* is the number of neurons per layer. Based on Eq. (S6), the phase term of each neuron, φ_i, becomes unbounded, but since the $\exp(j\varphi_i(x_i, y_i, z_i))$ term is periodic (and bounded) with respect to φ_i the error back-propagation algorithm is able to find a solution for the task in hand. The amplitude term a_i , is kept within the interval (0,1) by using an explicit normalization step shown in Eq. (S5).

2. Features of different scales of MDUNet

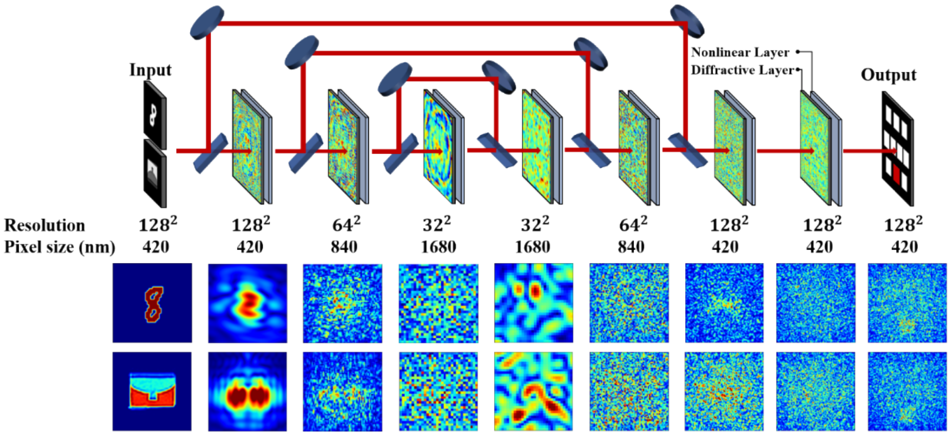
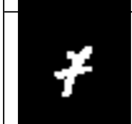
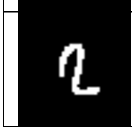


Fig. S1. The details of diffractive layers of MDUNet and the features of different scales after sampling: MNIST and Fashion-MNIST.

3. Comparison of prediction results with different sampling depths.

Table S1. Comparison of prediction results with different sampling depths

Object	Label	MDUNet / Layers				Object	Label	MDUNet / Layers			
		5	7	9	11			5	7	9	11
	6	6	8	6	8		3	3	3	8	8
	5	5	6	5	6		6	6	4	4	6
	8	8	8	3	3		4	0	4	0	4
	4	7	4	9	4		3	3	5	5	3
	6	5	6	8	6		8	8	8	5	5
	9	9	9	1	1		5	8	5	8	5
	2	2	2	3	3		1	1	3	3	1
	4	4	9	9	4		8	8	9	9	8
	7	7	7	2	2		2	8	2	8	2
	7	7	7	1	1		9	7	9	5	9
	2	2	8	8	1		8	3	8	3	8